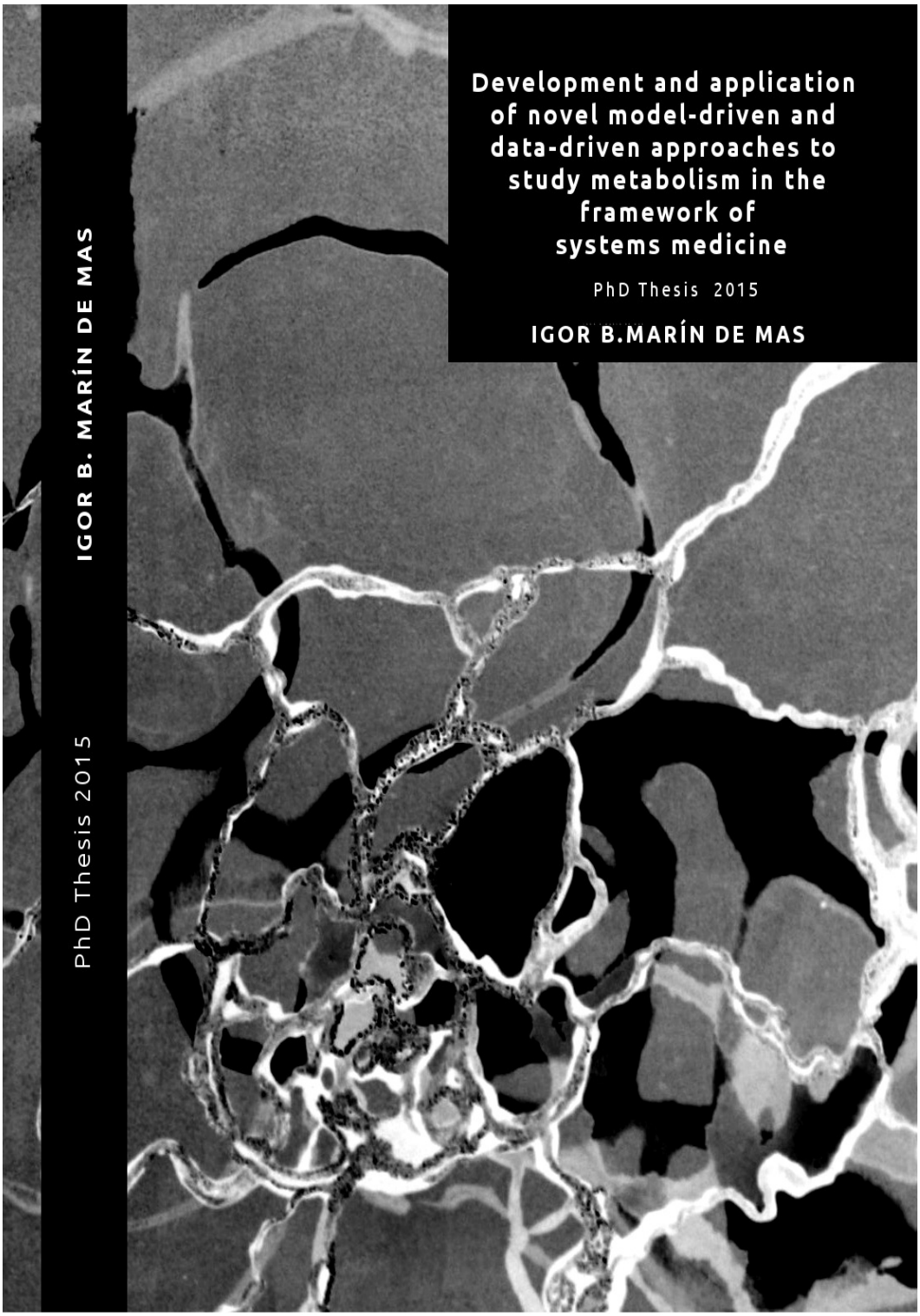
# DNA Methylation Dynamics during Myogenesis

Elvira Carrió Gaspar


IGOR B. MARÍN DE MAS
PhD Thesis 2015
Development and application of novel model-driven and data-driven approaches to study metabolism in the framework of systems medicine
PhD Thesis 2015
IGOR B.MARÍN DE MAS

# Development and application of novel model-driven and data-driven approaches to study the metabolism in the framework of systems medicine

Igor B. Marín de Mas

PhD Thesis

2015


Universitat de Barcelona

Programa de Doctorado en Biotecnología
Departament de Bioquímica I Biologia Molecular
Facultat de Biologia

**Development and application of novel model-driven and data-driven approaches to study metabolism in the framework of systems medicine**

Memoria presentada por **Igor Bartolomé Marín de Mas** para optar al grado de Doctor por la Universidad de Barcelona

Marta Cascante Serratosa
Codirectora y tutora

Balázs Papp
Codirector

Igor Bartolomé Marín de Mas
Doctorando

# Index

# 1. Introduction

## 1.1. Cell and metabolism

Cell is considered as the basic unit of live. It is delimited by a lipidic membrane enclosing an heterogeneous environment in which the different types of bio-molecules

interact and perform different biological processes.

Cells contain a large number and variety of biochemical constituent being the most representative: i) nucleic acids, ii) proteins and iii) metabolites

The nucleic acids are polymeric chains consisting of up to millions of monomers known as nucleotides. Each nucleotide has three components: a 5-carbon sugar, a phosphate group, and a nitrogenous base. Based on the carbon sugar, we can distinguish two families of nucleic acids, if the sugar is a deoxyribose, the polymer is DNA and if it's a ribose, the polymer is RNA. They also differ in one of their building blocks or monomers, while adenine, guanine and cytosine are common in both, thymine and uracil are exclusive of DNA and RNA respectively. DNA encodes the genetic instructions used to produce and regulate proteins.

The information is stored in the DNA in the form of sequences of nucleotides that encode for protein sequence. RNA molecules express the activity of the genes that are transcribed in a given scenario. In particular mRNA (messenger RNA) molecules are an intermediate product of the protein production process. The final product of the trancription-translation process that involves DNA and RNA are the proteins (Figure 1). These macromolecules are formed by binding amino acids with peptide bonds and perform a variety of biological roles ranging from regulatory (signaling pathways) or structural functions (contractile proteins in striated muscle) to the catalysis of biochemical reactions (enzymes) among many others.

Finally, metabolites are the smallest molecules including a big amount of different families of compounds such as sugars, fatty acids, lipids, … etc. The bio-transformation of these molecules are involved in several essential biological functions among which can be highlighted the production of building block necessary to create complex molecules or to satisfy cellular energy demand.

The set of chemical bio-transformations that consume or produce metabolites is known as metabolism. The metabolic reactions are catalyzed by a set of proteins named enzymes and as a whole is described in term of the metabolic fluxes (units of substrate metabolized per unit of time). These enzyme-catalyzed reactions allow organisms to

grow and reproduce, maintain their structures, and respond to their environments.

Based on its function the metabolism can be divided in two categories: catabolism and anabolism. Catabolism is aimed to produce energy by degrading substrates into simpler molecules. This energy is stored in different compounds such as Adenosine triphosphate (ATP). In this process the substrates are oxidized generating reductive power (i.e. NADPH) that in turns can be used to produce ATP. The anabolic processes synthesize the building blocks for complex cellular components. Unlike catabolic reactions, the anabolic processes are energy-demanding which results in a high inter-connected dependence between the anabolic and catabolic reactions.

Metabolism has been studied for decades due to its close relationship with cellular phenotype and consequently with several biological processes.

In particular, the analysis of cellular metabolism has allowed the diagnosis of diseases [1], novel drug target discovery [2] and improvement of biotechnological processes [3] among others.

Metabolism can be studied by gathering the metabolic reactions so as to meet a specific biological role. These groups of metabolic reactions that operate together are known as metabolic pathways. These pathways consume or produce specific compound through a combination of enzyme-based catalyzed reactions to satisfy a particular biological demand. A number of pathways are well described in the literature: pentose phosphate pathways, glycoysis, tricarboxylic acid (TCA) cycle, etc, which are refereed as canonical metabolic pathways.

Metabolic pathways don't operate in isolation and within an organism many different pathways work together to produce an overall global flux (reaction/compound) distribution. Consequently, emergent behaviors may arise from the interaction between the different metabolic pathways. In addition, based on the paradigm of the molecular biology, the metabolism is the final result of the different processes occurring within the cell: DNA → RNA → Protein + Metabolite → Metabolism.

Thus, based on this rationale, in order to study the complex behavior of metabolism is necessary the use of techniques that permit to measure the different constituents of the

cell massively. In this sense the development and use of automation equipment with classical cell biology techniques (high-throughput technologies) permit the rapid and highly parallel measurement of the different types of biomolecules to address biological questions that are otherwise unattainable using conventional methods. The set of techniques measuring massively the different constituents of the cells are known as the different -omics [4].

## 1.2. Omic data in Systems Medicine

In the last decades, we have witnessed the development and the increasing use of high-throughput technologies. These new technologies allow to analyze and quantify massively the different biochemical constituents existing within the cell increasing the amount, quality and variety of molecular data (the so-called omic data) [4].

Current omic technologies permit to identify and quantify molecules at different layers of the cascade of events connecting the genotype with the phenotype (figure 1).

The different experimental -omic approaches are high-throughput, data-driven and attempt to understand the cell metabolism as an integrated system [5].

Thus, some of the most widely used omics in systems medicine include: transcriptomic, proteomic, metabolomics and fluxomic . Following, some of the better established ‘omics’ being used in the approach of systems medicine are briefly described:

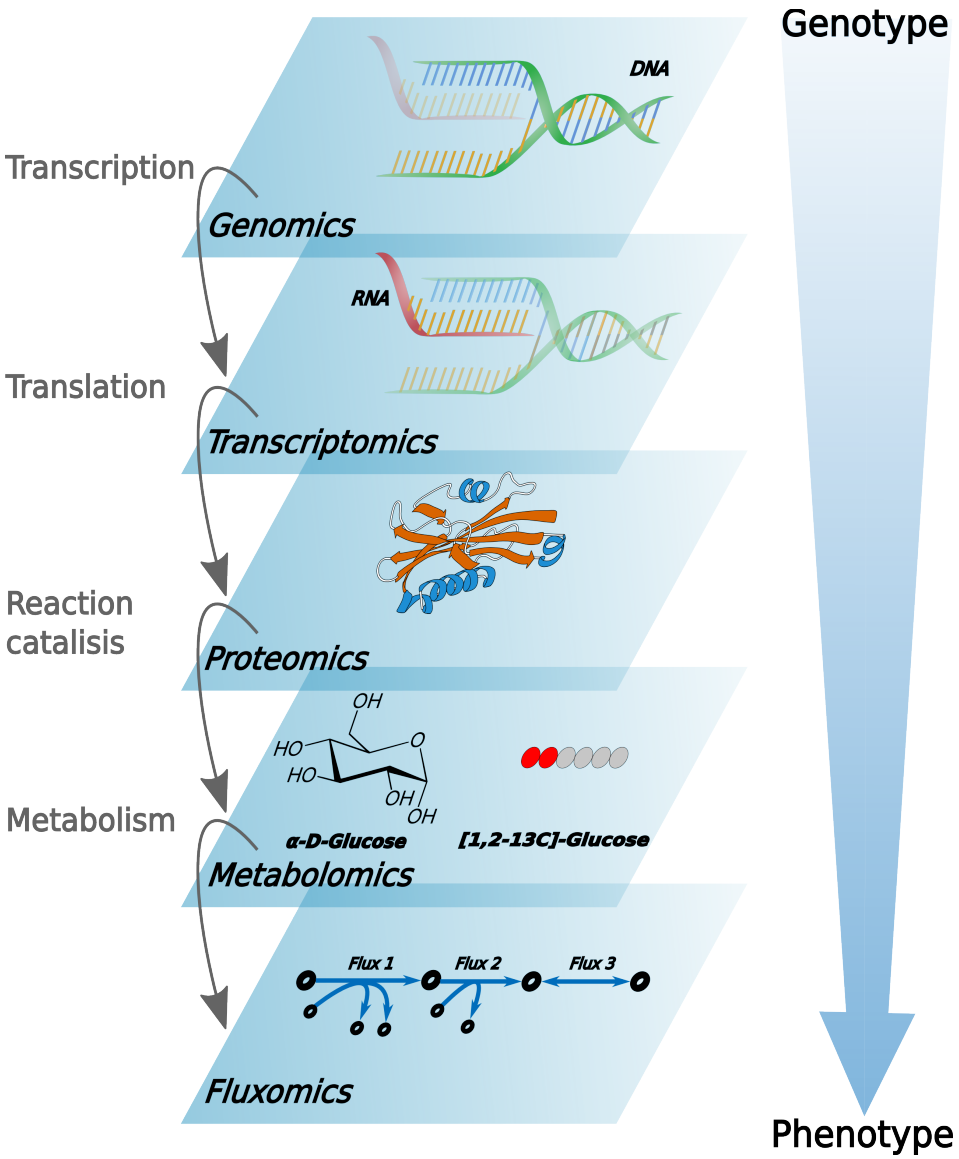



Figure 1. The most relevant omic data in integrated omic studies, represented as the different layers of biological information

### 1.2.1 Transcriptomics

This discipline aims to identify and quantify the whole set of RNA molecules such as messenger RNA (mRNA or transcripts) or non-coding RNA [6]. Transcriptomics

enable the quantitative measurement of the dynamic expression of mRNA molecules and their variation in different scenarios at the genome scale, thus reflecting the genes that have been expressed at any given time. This technology is based on the artificial hybridization of mRNA that takes place in a matrix enclosing the corresponding complementary sequence (probes). One of the most widely used microarray technologies is the high-intensity oligonucleotide microarray. This technique measures the absolute abundance levels of mRNA by individually hybridizing each sample. This hybridization produces a fluorescent dye whose intensity is directly related to the amount of mRNA in the biological sample. Next, these intensities are preprocessed in order to account for affinity differences between probes, non-equal quantities of mRNA in each microarray or noisy intensity [7]

This technology allows the measurement of thousands of mRNAs simultaneously providing a vast amount of information with a relative low cost. Particularly repositories like Gene Expression Omnibus [8] or Array Express [9] stores thousands of microarray datasets that are freely available.

### 1.2.2 Metabolomics

For most organisms, there is no direct relationship between gene expression or protein abundance and metabolites [10]. Metabolomics provides a global characterization and quantification of the metabolic profile of a given sample. The metabolic composition of a a cell gives a closer picture of the metabolic processes, and hence of the phenotype, than other omic approaches such as transcriptomics or proteomics. However, metabolomic approaches are intricate due to three main reasons: i) thousands of metabolites could be present at time in a sample, some of them are almost identical in terms of molecular composition and abundance, ii) the concentration of two metabolites present in the same sample may differ in several orders of magnitude [11] and iii) among the biological constituents of the cell, metabolites present the largest heterogeneity in term of chemical and physical properties [12]. Metabolomics analysis is typically performed by employing mass spectrometry techniques. These techniques including time-of-flight mass spectrometry (GC-TOF), high-performance liquid

chromatography-mass spectrometry (LC-MS) or capillary electrophoresis-mass spectrometry (CE-MS), among others, are able to measure hundreds of metabolites simultaneously allowing the qualitative and quantitative determination of the whole metabolome of a given cell. In brief, these techniques convert the individual molecules into ions that can be moved about and manipulated by external electric and magnetic fields. It allows the measurement of the mass-to-charge ratio and determine the abundance of a given molecule. In addition to mass spectrometry techniques, Nuclear Magnetic Resonance (NMR) also enables the detection and quantification of metabolites. This technique measures the magnetic response of the atomic nucleus of a sample to an external magnetic field. In both techniques, the response is measured generating a particular spectrum that is used to identify the metabolites in the sample. Despite the advances in NMR, the sensitivity of this technique still require further improvements to achieve the same sensitivity as mass spectrometry techniques. Additionally, in order to achieve higher coverage and better identification, metabolomics analysis can also be performed through a combined application of several technologies [13].

In contrast to transcriptomics or proteomics, the availability of metabolomic data is still scarce and limited. One of the few repositories of metabolomic data is the Human Metabolome Databse [14], that freely provides structural information of thousands of metabolites presents in the human cells.

### 1.2.3 Fluxomics

The different omics explained above provides snapshots of a set of components present in a cell at a given time. Metabolic fluxes are the final result of the interplay of gene expression, protein concentration, enzyme kinetics, regulation and metabolite concentrations. By analogy to other ‘omics-type analyses this technology was termed ‘fluxomics’. Metabolomics and proteomics provides quantitative measurements of metabolites and proteins, respectively. However, inferring metabolic fluxes directly from these omics may lead to erroneous conclusions due the nonlinear nature of the interactions between enzymes and metabolites. Fluxomics gathers the set of

experimental and computational tools enabling to determine the metabolic reactions fluxes [12]. This omic integrates in vivo measurements of metabolic fluxes with mathematical models to allow the determination of absolute flux through a metabolic network. Some fluxes defining the overall rate of nutrients consumption or production (exchange reactions) can be easily measured (e.g., consumption of Glucose or secretion of Lactate). By contrary, intracellular fluxes are more difficult to determine and more detailed considerations are required [15]. In this sense the use of metabolomic data from isotope labeling experiments is emerging as a powerful tool to determine the internal metabolic fluxes.

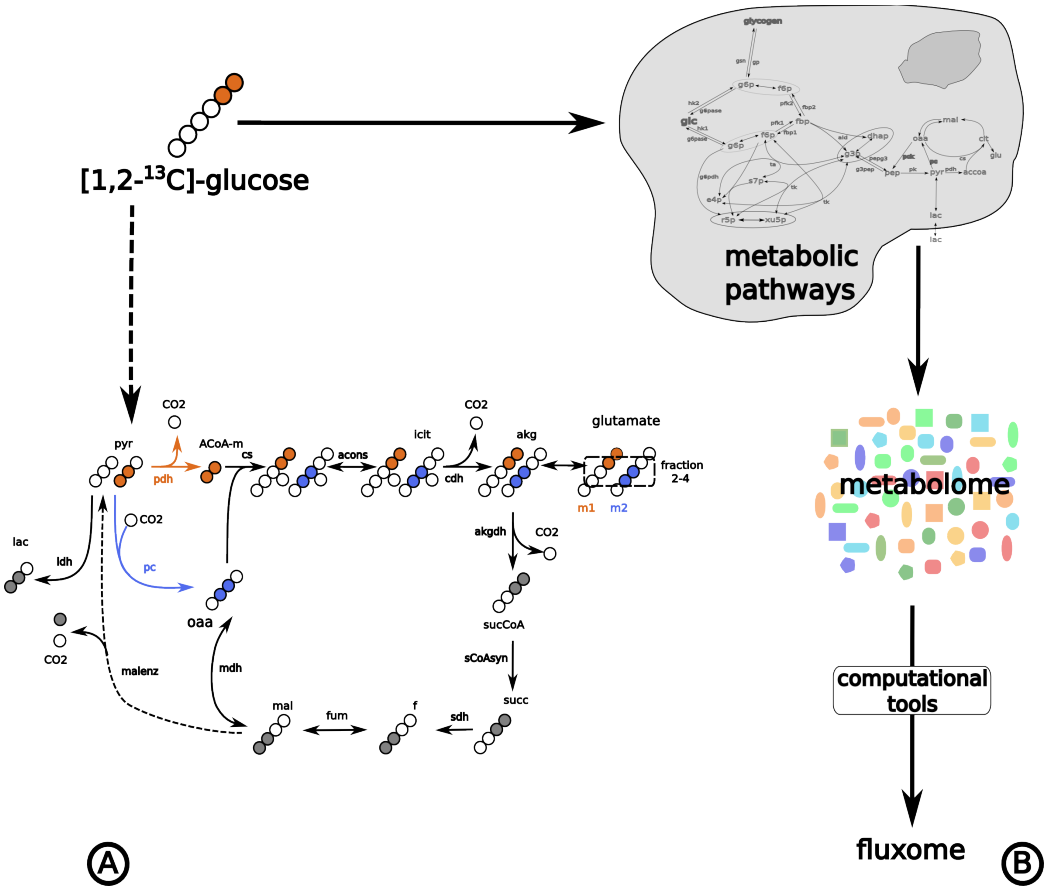


Figure 2. A Representation of how the metabolic transformations of one molecule of glucose labeled in the first and second carbon (orange circles) can produce different labeling patterns of glutamate depending on the metabolic reactions involved in the process. Thus, we can analyze glutamate labeling using NMR or the fraction 2-4 using mass spectrometry techniques in order to determine the metabolic reactions involved in its metabolism. B Represents a work-flow to determine metabolic fluxes from metabolomic data based on labeled substrates. Thus, using computational tools that integrate a previous knowledge of the metabolic pathways, we can analyze data from experiments with labeled substrates in order to determine the internal metabolic fluxes.

These experiments are based on the use of substrates labeled with isotopic tracers, mainly labeled with $^{13}C$ (stable isotope), introduced at specific locations. As is shown in the figure 2A, the inter-conversion of substrates via intracellular enzyme-catalyzed reactions alters the labeling pattern of the carbon skeleton in the metabolites [16]. The experimental measurements of $^{13}C$ enrichment of metabolites allows the mathematical tools to fit intracellular fluxes.

The isotopic enrichment of metabolites is represented with the concept of positional isomers (Isotopomers) or mass isomers (Isotopologues). Isotopomers can be measured by Nuclear Magnetic Resonance while isotopologues are measured by mass spectrometry techniques. Both techniques are based on $^{13}C$ (a non-radioactive isotope of $^{12}C$),, whose natural abundance is around 1.1%. The nuclear magnetic properties of this isotope as well as its atomic mass are different from $^{12}C$ allowing to NMR, in the first case and mass spectrometry techniques, in the second, to detect and quantify those metabolites with labeled carbons.

### 1.3. Computational approaches in Systems Medicine

The advent of high-throughput technologies in the 21th century has revolutionized the study of biological systems transforming molecular biology into a data-rich discipline. These, the so-called, omic technologies have provided quantitative data for thousands of cellular components across a wide variety of scales. It moved the study of biological systems from a reductionist approach, in which all molecular components (proteins, DNA, metabolites, etc.) were studied separately, to a scenario where holistic

approaches are required in order to understand how the biological actors interact in highly interconnected networks where discrete biological function can only rarely be attributed to an individual molecule [17]. In this context Systems Medicine has emerged as an integrative discipline for the study of complex interactions between biological systems components to explain the functional organization of cells within the context of organs, organisms and their physiology [18]. However, extraction of 'knowledge' from this ocean of omic data is not trivial [19]. In order to cope with this challenging task a wide range of computational methods has been developed. Then, based on how the experimental data is used, two main general approaches can be distinguished: i) data-driven approaches based on statistical methods and ii) model-driven approaches which include an explicit description of the biological network.

### 1.3.1 Data-driven approaches

Data-driven approaches aim to extract knowledge and generate new biological understanding directly from the huge volume and diversity of bioscience data that is now available and so underpin and enable biological research. A large variety of statistical methods are available to determine the most significant differences between two groups of biological interest (e.g. treated and control) [20]. Since high throughput genomics is one of the most widely used technique to extract data from biological samples, most of the data-driven approaches are focused on the analysis of gene expression data. Next, are briefly explained the data-driven approaches used in this thesis, which are among the most relevant and widely tested methods used in systems medicine

#### 1.3.1.1 Identification of genes differentially expressed by Rank Product

The rank product method is a widely accepted non-parametric technique for identifying differentially expressed genes in replicated microarray experiments under two experimental conditions or groups [21]. It has also been widely applied to other omic datasets, for example in proteomics and metabolomics [22,23]. This method uses any measure that compares two conditions like the logarithm based two fold change (log2

FC), to entail ranking genes according to their differential expression within each replicate experiment and subsequently calculating the product of the ranks across replicates. This approach assumes that i) only a minority of genes are affected by relevant expression changes, ii) the measurements are independent between replicate arrays, iii) most changes are independent of each other and iv) the variance of the measurements is similar for all genes.

Thus, for an experiment examining n genes in k replicates, assuming that the genes are randomly distributed in the k lists, the probability that a given gene was at the top of each list (rank 1) is $1/^{n}k$. Generally, for each gene g, one can calculate the corresponding combined probability to be at the top of each list (rank 1) as a rank product ($RP_{up}$):

$$RP_g^{up} = \prod_{i=1}^{k} \left(\frac{r_{i,g}^{up}}{n_i}\right) \quad (1)$$

Where k is total number of replicates, $r^{up}_{ig}$ is the position of gene g in the list of genes in the $i^{th}$ replicate sorted by decreasing FC that is normalized by the length of the $i^{th}$ replicate ($n_i$). The most strongly up-regulated gene would have a $r^{up}=1$. Likewise $r^{down}=1$ corresponds to the most strongly down-regulated gene. The RP values are calculated over all possible pairwise comparisons. Thus, the genes with smallest RP ($RP^{up}$ or $RP^{down}$) values are those that are more differentially expressed. Next, in order to correct for the fact that the pairwise comparisons between samples are not independent is necessary to determine the significance of gene expression changes and which genes are likely to be truly differentially expressed. Then, for an experiment with k replicates and n genes, is defined the statistic p'. It is calculated by multiplying RP by a factor F representing the number of possible products of k numbers smaller than n that are equal to the numerator of the RP value:

$$p_g^{'} = \mathrm{RP}_g . F \quad (2)$$

The F factor can be determined by using a simple permutation-based estimation procedure [24]. Other approaches propose a continuous gamma distribution approximation for the log-transformed rank products [25] or an exact distribution of

rank products [26]. In this manner we can estimate how likely it is to observe a given RP value or better in a random experiment. In addition, is necessary to determine the conservative estimate of the percentage of false-positives (PFP) or q, for each gene g, that is calculated as follows:

$$q_g = \mathrm{E} \cdot \frac{\mathrm{RP}_g}{\mathrm{rank}(g)} \quad (3)$$

Here, rank(g) gives the position of gene g in a list of all genes sorted by increasing RP value. This estimates the PFP [27] providing a significance level to the observed changes on each gene. Now, one can decide how large a PFP would be acceptable and extend the list of accepted genes up to the gene with this $q_g$ value.

Thus, RP method offers reliable results in highly noisy data. This approach is powerful for identifying biologically relevant expression changes and can lead to a sharp reduction in the number of replicate experiments needed to obtain reproducible results.

1.3.1.2 Define pathways differentially activated and cell-specific gene signature by Gene-set enrichment analysis

Typically, the strategies for gene expression analysis are focused on identifying individual genes that vary significantly between two states of interest. Although useful, these methods fail to detect biological processes, such as metabolic pathways, transcriptional programs, and stress responses, that are distributed across an entire network of genes and subtle at the level of individual genes. Gene Set Enrichment Analysis (GSEA) is a powerful analytical method for interpreting gene expression data [28]. This method allows to define the sets of genes that are combinatorially selected by exogenous and endogenous environmental changes and are associated to a complex polygenetic phenotypes (stage of differentiation, disease state, responsiveness to exogenous perturbations, ...) or cellular process (gene signature) and determine if they are over-expressed in one of the states of interest [29]. This approach requires a combination of analytical methods and high performance experimental for identifying related sets of genes (e.g. genes in pathways or functional classifications) associated with phenotypic changes. Identification generally means the determination of which

genes sets among a known collection are related or the discovery of gene sets that were not previously known to be related [30].

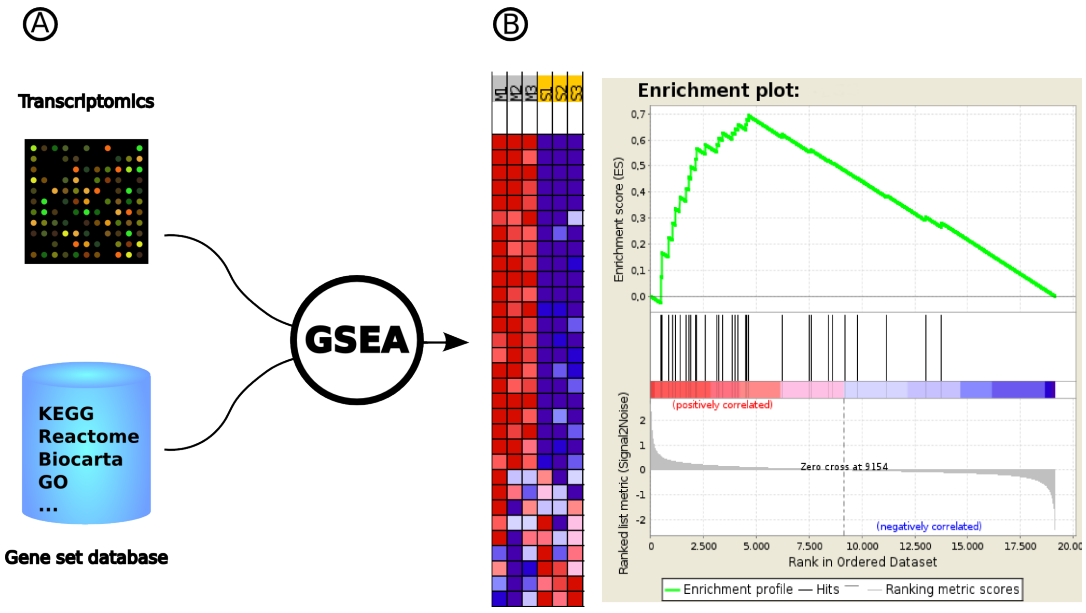



Figure 3. GSEA: The input of a GSEA are a predefined list of genes extracted from a gene set database or build based on the case of study and at least two gene expression profiles that will be compared. GSEA compares the gene expression profiles (by FC, correlation, signal to noise, ….), ranks it and finally determines if the genes defined in the list are mainly grouped in the top of the list (over-expressed in one group), in the bottom (over-expressed in the other group) or randomly distributed (the list of genes is not differentially expressed). As is represented in the figure, this is quantitatively evaluated by defining an enrichment score (ES) that is calculated based on a weighted Kolmogorov Smirnov (WKS) test.

Gene-set enrichment analysis uses as input a predefined list of genes extracted from a gene set database or build based on the case of study and at least two gene expression profiles that will be compared (Figure 3A). Briefly, the work-flow consists of the following steps:

(i) rank all genes by the magnitudes of their differential expression and select a window in the ranked list, i.e. a contiguous run of some number of genes starting at any rank

(ii) define an enrichment score (ES) based on a weighted Kolmogorov Smirnov (WKS) test that is defined by the following expression:

$$ES = P_{hit} - P_{miss} \quad (4)$$

Where:

$$P_{miss}(S,i) = \sum_{g_i \notin S;\, j \le i} \frac{1}{(N - N_H)} \quad (5)$$

$$P_{hit}(S,i) = \sum_{g_i \in S;\, j \le i} \frac{|r_j|^p}{N_R} \text{ ; here } N_R = \sum_{g_i \in S} |r_j|^p \quad (6)$$

$N_H$ is the total amount of genes. The exponent p controls the weight of each step, then if p=0 the analysis is reduced to a standard Kolmogorov Smirnov analysis. if p=1, the analysis weights the genes by the magnitude of their differential expression (it can be higher than 1). This equation measures the difference between the number of genes in a pre-specified gene set that are observed in the window, and the number of occurrences if the genes in the set were uniformly distributed in the list. The ES is normalized for each gene set to account for the size of the set, yielding a normalized enrichment score (NES).

(iii) simulate a background distribution of the enrichment score by shuffling samples and estimate the statistical significance of the gene set [31]

(iv) repeat steps i–iii for all pre-specified gene sets (hypotheses) and for various window sizes

(v) correct for multiple hypotheses testing.

Thus, for a given gene list L and based on the list of genes based on their differential expression M, the method calculates an enrichment score that reflects the degree to which a set L is over-represented at the extremes (top or bottom) of the entire ranked list M (Figure 3B). Thus, GSEA approach allows to determine altered metabolic pathways and processes by using predefined gene sets or discover new processes and gene cross-correlations by defining case specific gene subsets (gene signatures). In addition, this method can be applied to other data sets such as proteomics, genotyping information, or metabolite profiles [32].

1.3.1.3 Gene expression interdependency discovery by correlation analysis

Correlation analysis are focused on the study of a broad class of statistical relationships involving dependence between two random variables or two sets of data (correlation does not imply causation). Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient, or simply "correlation coefficient" [33]. This coefficient is defined as follows:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\left(\frac{X_i-\overline{X}}{s_X}\right)\left(\frac{Y_i-\overline{Y}}{s_Y}\right)\right) \quad (7)$$

Here $X_i$ and $Y_i$ are the ith measurement of the random variables X and Y along n samples where i = 1, 2, ..., n, $\overline{X}$ , Sx and $\overline{Y}$ , Sy are the mean value and the standard deviation of the variables X and Y respectively and are defined by the following expressions:,

$$\overline{X} = \frac{1}{n}\sum_{i=0}^{n} X_i \quad \text{, and} \quad s_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2} \quad (8)$$

The equation 7 represents a symmetric function with values between -1 and 1. In cases that exist a perfect direct (increasing) linear relationship or correlation between the variables the correlation is +1, inversely, the correlation is -1 when exist a perfect decreasing (inverse) linear relationship (anticorrelation), the intermediate values indicate different degrees of linear dependence between the variables finally as it approaches zero there is less of a relationship (closer to uncorrelated)[34]. Thus, the closer the coefficient is to either −1 or 1, the stronger the correlation between the variables. The degree of dependence between variables *X* and *Y* does not depend on the scale on which the variables are expressed. Thus, having a set of samples that represent the gene expression levels in a time course, the exposure to different concentrations of a certain compound (i.e. drug, ROS levels, ...), stages of a process (i.e. metastatic

progression), …, we can determine the correlation or anticorrelation of the same variable along the different samples or how the correlations between variables evolve in a time course series samples. By determining the correlation between *n* random variables *X*1, ..., *Xn* of a given set of samples we can construct a *n*×*n* correlation matrix whose *i*,*j* entry is corr(*Xi*,*Xj*). This matrix is symmetric because the correlation between *Xi* and *Xj* is the same as the correlation between *Xj* and *Xi*. The information extracted from these matrix can be represented graphically in an interaction graph in which the nodes represents the biological components (genes, proteins, …) and the rows connecting the nodes the degree of correlation between them (Bayesian networks).

### 1.3.2 Model-driven approaches

As was exposed previously, data-driven methods are extremely efficients to extract knowledge from a massive amount of data. However, in some situations these approaches are not suitable to analyze biological phenomena. For example, the amount of experimental data required to determine patterns, correlations and mechanisms are not available or simply, it is necessary to add new knowledge extracted from the literature in order to constraint the space of solutions to be analyzed. Typically, this information is introduced in the form of models. A model is an abstraction of a biological system representing the interactions and interdependencies between the components of a given system (metabolites, genes, proteins, etc.).

These models can be represented in the form of graphs, nevertheless, these representations are not precise, usually cannot represent all the dynamic behavior of the biological systems and are useless to handle large networks. On the other hand, computational models provide a precise mathematical representation of knowledge allowing to interpret and evaluate measured data, analyze system's behavior (e.g. identify important parts for a particular behavior, etc.), generate and test hypothesis. Additionally, these computational tools allow us to improve the consistency between model and data in an iterative fashion [35-37]. Consequently, mathematical modeling has become an essential tool for the comprehensive understanding of cell metabolism

and its interactions with the environmental and process conditions.

There exist a large variety of mathematical models to represents different aspects of the cellular metabolism. However, since among the main omic technologies, fluxomics provides a better representation of the phenotypes, those models that represents somehow the metabolic processes occurring within the cell, are specially useful to study the aberrant metabolic adaptations in diseases with strong metabolic components such as cancer or chronic diseases [12].

Thus, the metabolism can be studied through models representing a specific metabolic pathway, a whole metabolic network or a gene regulatory network regulating the metabolism (e.g. by the up or down-regulation of key enzymes).

Following are briefly summarized some of the most relevant model-driven approaches to study the metabolism:

#### 1.3.2.1 Kinetic modeling & $^{13}C$ Fluxomics

Biological systems are not static entities. They change over time in response to a variety of perturbations. Kinetic modeling methods are aimed to integrate the corresponding data sets, allowing one to infer the kinetics and dynamics of the reactions between all the chemical entities in a cell [38].

By this computational tool one can represent a metabolic network that incorporates the stoichiometry and thermodynamic of the system (direction of metabolic reactions) and a detailed knowledge of the enzyme regulation, kinetic rate laws and their associated parameter values to formulate kinetic models that can accurately capture the dynamic responses of the metabolic networks [39].

The mathematical formalization is performed by representing the metabolic concentrations in a system of ordinary differential equations (ODEs) that can be solved by Flux Balance Analysis (FBA).

Thus, the concentration of the ith metabolite in the model is described by a variable $S_i$. As a result of mass balances (no matter can appear or disappear), the change of this variable over time ($dS_i/dt$) is given by the sum of the rates of the enzymes synthesizing the metabolite minus the sum of the rates of the enzymes utilizing the metabolite as is

expressed in the following equation:

$$\frac{dS_i}{dt} = R_{input} - R_{output} \quad (12)$$

Here $R_{input}$ and $R_{output}$ are the set of reactions in which the ith metabolite is a product or a substrate respectively. In kinetic modeling approaches, the rate of each enzymatic reaction can be described by kinetic law equations, the following equation represents an example of a Michaelis-Menten kinetic equation that is one of the best-known models of enzyme kinetics:

$$R_m = \frac{V_{max} * S_i}{K_m + S_i} \quad (13)$$

The equation above represents the mth reaction that uses the ith metabolite as substrate. Here $S_i$ is the concentration of the ith metabolite that is a substrate of the reaction, $V_{max}$ and $K_m$ are the kinetic parameters of the equation, the first represents the maximum rate achieved by the system, at maximum (saturating) substrate concentrations and $K_m$ is the substrate concentration at which the reaction rate is half of $V_{max}$.

This process yields a system of ODEs in which $dS_i/dt$ is on one side and the metabolite-dependent rate laws are on the other side of the equations. With this system of differential equation, the metabolic network can be simulated, and by solving the system of ODEs the steady state can be calculated, in which all reaction rates and metabolite concentrations are constant. The figure 4 illustrates this process.

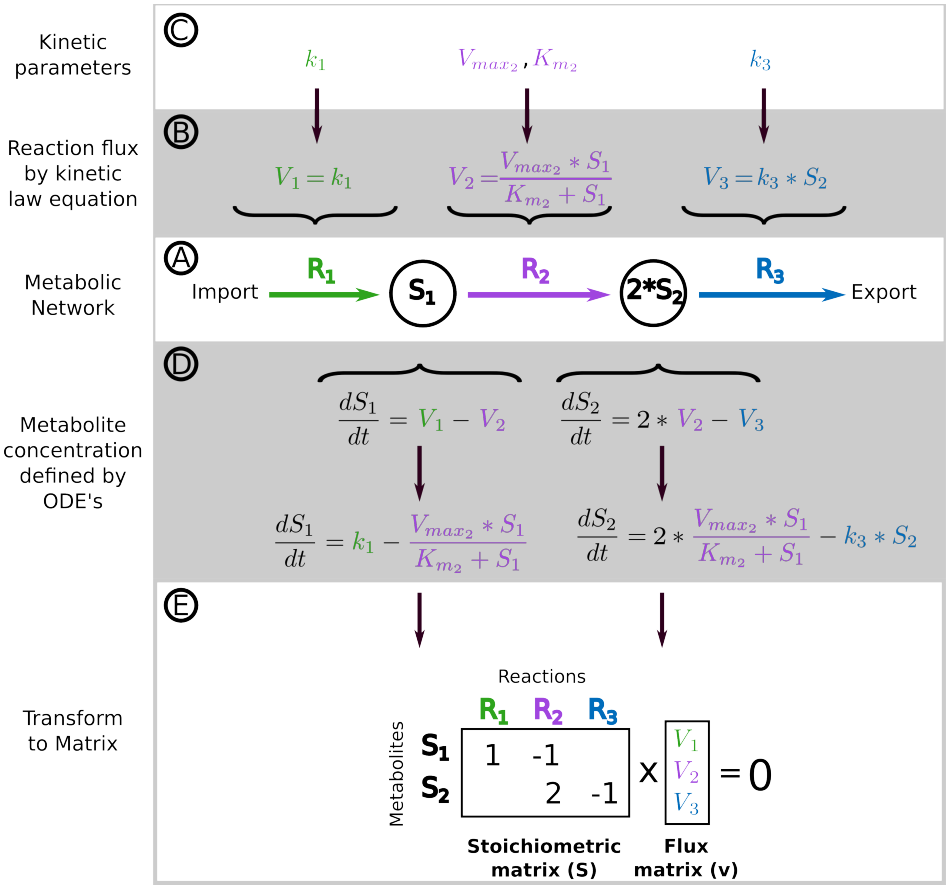



Figure 4: Kinetic metabolic model. A: Toy network representing a metabolic process in which the nodes are the metabolites involved in the process and the arrows the biochemical reactions that transform the substrates in to the final products. Metabolite $S_1$ is imported into the toy network by reaction $R_1$, the reaction $R_2$ produces two molecules of $S_2$ from 1 molecule of $S_1$, finally the reaction $R_3$ exports the metabolite $S_2$ outside the network. B: Equations describing the metabolic fluxes; $V_1$ is the flux of reaction $R_1$ and represents a linear equation that depends on the kinetic parameter $k_1$, $V_2$ is the flux of reaction $R_2$ that is based on a Michaelis-Menten kinetic equation finally $V_3$ represents the metabolic flux through reaction $R_3$ it is based on a linear equation that depends on the kinetic parameter $k_3$ and the concentration of the metabolite $S_3$. C: kinetic parameter that determine the metabolic fluxes trough the reaction in the network. D: System of ODE's that summarizes the kinetic information from B and the stoichiometric and thermodynamic information embed into the metabolic network. This system of ODE's determines the metabolic concentrations of the metabolites S1 and S2. E: Transforming the system of ODE's to a stoichiometric matrix regarding defines the concentrations.

Figure 4 shows how to express the information embed into a metabolic network and the kinetic law equations into a system of ODE's and finally in a stoichiometric matrix that can capture the dynamics of the metabolism. Based on the kinetic laws governing the metabolic fluxes through the reactions (Fig 4B and 4C) and the stoichiometric and thermodynamic information embedded in the network one can construct a system of ODE's that defines how the metabolic concentrations varies along the time (Fig 4D). This system of ODE's can be represented in the form of stoichiometric matrix (S) (Fig 4E). This matrix contains the stoichiometric relation between the metabolites and reactions, this matrix is multiplied by the flux matrix (v) that contains the metabolic fluxes that are function of the kinetic parameters. If we are analyzing the system at steady state, then S x v = 0 because, by definition, the metabolic concentrations don't vary. This method is widely used in systems biology existing a large number of softwares developed to simulate biological networks based on kinetic models [40] .

In order to improve the predictive capabilities of classical kinetic modeling, new approaches has been developed to incorporate experimental measurements that allows a better understanding of systems dynamics. Stable isotope tracing using [1,2- $^{13}$C]-glucose as a source of carbon, has been described as a very powerful tool for metabolic flux profiling [41-46]. The metabolite labeling proceeds during cell incubation with substrates contained stable isotope tracer, usually $^{13}$C. The specific pattern of various $^{13}$C isotopic isomers (isotopomers) fractions measured using mass spectrometry or nuclear magnetic resonance techniques characterized the distribution of metabolic fluxes in the cells under the studied conditions. To evaluate the flux distribution from measured isotopomer distribution special software tools are necessary, therefore various tools were created for this purpose [47]. Thus, simulating isotopic isomer (isotopomer or isotopologue) distributions allows a more in-deep analysis of the kinetic and regulatory characteristics of metabolome.

However, for a long time, the detailed analysis of isotopomer distribution was restricted to isotopic steady state [48]. The analysis of metabolic fluxes not restricted by isotopic steady state is very important even if in the experiment metabolic fluxes do

not change; although intracellular metabolites could reach steady state in a range of minutes, the existence of intracellular stores, such as glycogen, which intensively exchange with intermediates of central carbohydrate metabolism, essentially delay the time necessary for establishing isotopic steady state. Amino acids and lipids are also linked with central metabolism and could prolongate the pre-steady state phase for isotopomers. These internal stores and external metabolites could be far from isotopic steady state during hours of labeling experiment. Of course, always there is a possibility of measuring the labeling of such stores and apply classical $^{13}$C-FBA for the "fast" intermediates of central metabolism considering that they are in quasi-steady state. However, the simulation and comparison with experiment the time course of such "slow" variables provides additional restrictions that helps to evaluate the fluxes. Moreover, there is another reason for using non-stationary analysis based on a kinetic model of considered pathways: it has a potency to perform a more profound analysis of kinetic characteristics and regulation in the pathway, if experimental data are sufficient for that. Such advantages stimulated the development of other bioinformatic tools for non-stationary flux analysis [49]. Thus, the isotopic non-steady state analysis, can be used for the study of cellular metabolism under normal conditions as well as under the conditions of stress or pathology. The tools applicable for analysis of non-steady state conditions appeared relatively recently. Then, for example, the software Isodyn has been developed to perform $^{13}$C non-stationary flux analysis [49,50]. This software simulates $^{13}$C redistribution in metabolites by automatically constructing and solving large systems of differential equations for isotopomers. The basis of Isodyn is a kinetic model of metabolic pathway. Such a model simulates the metabolic fluxes, which transfer the $^{13}$C label from substrates, externally supplied, to the intermediates and products of cellular metabolism. Using the metabolic fluxes simulated by the kinetic model specific module of Isodyn simulates the distribution of label, which could be compared with experimental distribution. By adjusting the parameters of kinetic model to fit experimental label distribution Isodyn defines the characteristics of analyzed metabolic pathway. This methodology allows to simulate the metabolic fluxes, which transfer the $^{13}$C label from substrates, externally supplied, to the intermediates and

products of cellular metabolism. Currently, the stable isotope tracing of metabolites has been used to identify the adaptive changes of fluxes in human in normal and diseased states [51], in isolated cells [52], cancer cell cultures [53], and organisms such as fungi [54], yeast [54,55], etc.

Deterministic, ODE-based enzyme-kinetic models have the longest history in the area of metabolic pathway modeling. These approaches and specially those that permit to perform non-stacionary $^{13}$C-FBA offer a highly curated representation of the metabolic and regulatory processes and are able to predict specific metabolic fluxes with a high precession. However, the local complexity of kinetic models leads to the fact that for reasons of feasibility usually single pathways or even just some reactions are modeled. These facts restrict the kinetic models to a relatively small size networks, being able to describe small metabolic networks with high precision but not large networks like GSMM.

1.3.2.2 Constraint-based methods & Genome-scale Metabolic Models

Genome-scale Metabolic Models (GSMMs) are emerging as a potential solution to decipher the molecular mechanisms underlying a variety of metabolic diseases such as cancer or diabetes in the context of systems biology [56]. GSMMs represent the metabolic reaction complement encoded by an organism's genome. These models are built based on the literature and databases and enable to summarize and codify information known about the metabolism of an organism.

Recent reconstructions of human metabolism such as Recon1 [57], Edinburg Human Metabolic Network (EHMN) [58] or the most recent reconstructions of human metabolism, Recon2 [59] are widely used to study the mechanism of diseases with a strong metabolic component, such as cancer or diabetes [60-63]. This systems biology tool enables the mathematical representation of bio-transformations and metabolic processes occurring within the organism and offers an appropriate framework to integrate the increasing amount of "omic" data generated by the different high-throughput technologies. The transformation into a mathematical formulation is mostly

driven by constraint-based modeling (CBM) [64] and allows the systematic simulation of different phenotypes, environmental conditions, gene deletion, etc. This approach allows modeling the complexity of cancer metabolism and tackling more problematic biological questions such as the role of metabolism in cancer disease [1].

Genome-scale constraint-based metabolic models have been used for a variety of applications, involving studies on evolution [65], metabolic engineering [66-68], genome annotation [69] or drug discovery [2] with a high relevance in cancer research. Indeed, GSMMs can efficiently capture the complexity of metabolic diseases in a holistic manner and permit to improve existing therapies or develop new ones [70].

GSMMs include stoichiometric details for the set of known reactions in a given organism. These large scale metabolic models require computational methods to be qualitatively analyzed.

Traditionally, approaches based on ordinary differential equation have been used for characterization of dynamic cell states. However, this full-scale dynamic modeling is frequently infeasible for large scale networks because of a paucity of necessary parameter values. Constraint-based methods (CBM) permit the analysis of large-scale biochemical systems under conditions where kinetic parameters need not be defined (steady state). Genome-scale constraint-based metabolic models can be used to predict or describe cellular behaviors, such as growth rates, uptake/secretion rate or intracellular fluxes [64]. Flux Balance Analysis (FBA) is one of the most widely used CBM for the study of biochemical networks. Unlike FBA in kinetic modeling where the variables were the kinetic parameters, here the variables used in FBA include the fluxes through transport and metabolic reactions and model parameters include reaction stoichiometry, biomass composition, ATP requirements, and the upper and lower bounds for individual fluxes which define the maximum and minimum allowable fluxes of the reactions.

The first step in FBA is the mathematical representation of the metabolic reactions in the form of a numerical matrix, with stoichiometric coefficients of each reaction (stoichiometric matrix), where the metabolites are represented in rows and reactions in columns. FBA employs mass actions formalism for the mathematical representation of

the metabolic networks: **dC/Dt=S.v**., where v and C are vectors of reaction fluxes and metabolite concentration respectively, t is time and S is the stoichiometric matrix [Figure 5A]. Next step is to impose constraints to the metabolic network. Constraints are fundamentally represented in two ways: i) steady-state mass-balance imposes constraints on stoichiometry and network topology on the metabolic fluxes through the network. Additionally, steady state assumption also imposes constraints that narrow the space of solutions. By definition, the change in the concentration of a certain metabolite over time at steady state is 0: **dC/Dt=0,** thus: **S.v=0.** These constraints ensure that for each metabolite in the network the net production rate equals the net consumption rate and ii) inequalities that impose bounds on the system: every reaction can also be given upper and lower bounds. These restrictions are based on measured rates (e.g. metabolite uptake/secretion rates) or reaction reversibility (e.g. irreversible fluxes have a zero lower bound) and are used to define the environmental conditions in a given simulation such as nutrient or $O_2$ availability that can be related with a specific tumor microenvironment or stages in the tumor progression. Finally it is necessary to define a phenotype in the form of a biological objective that is relevant to the problem being studied (objective function). Typically, objective functions are related with growth rate prediction. GSMMs define this phenotype by an artificial biomass production reaction, that is, the rate at which metabolic compounds are converted into biomass constituents (nucleic acids, lipid, proteins, etc.). The biomass reaction is based on experimental measurements of biomass composition and is unique for each organism or cell type. Thus, an objective function could be the maximization of growth rate that can be accomplished by calculating the set of metabolic fluxes that result in the maximum flux through biomass production reaction. Since the uncontrolled cell growth is the basis of tumor progression, this approach is widely used in the simulation of cancer cell metabolism. The objective function can be adapted to the specific cell type or organism, however the objective that better defines our case of study is not always obvious, especially in multicellular organisms [71].

Taken together, the mathematical representation of the metabolic reactions and of the objective function, is defined as a system of linear equations that are solved by a

number of algorithms and software developed for this purpose [72]. Predictions of values for these fluxes are obtained by optimizing for an objective function, while simultaneously satisfying constraint specifications.

GSMMs offer an advantageous platform for the integration of omic data (e.g. ref [73]) [Figure 5B]. In this framework cellular and molecular phenotypes are simulated allowing the development of biological hypotheses and discoveries [74]. Metabolic reconstruction of the human metabolism have been successfully used for a variety of analyses of omic data, including applications in data visualization [75], deducing regulatory rules [76], network medicine [77], constructing tissue-specific models [78] or multi-cellular modeling [79]. Thus, omic data can be used to further constrain the non-uniqueness of constraint-based solutions space and thereby enhance the precision and accuracy of model prediction [74] [Figure 5A-C]. To achieve this aim a number of FBA-driven algorithms that integrate omic data into GSMMs have been developed [62,76,78,80-88]. It has enabled to gain further biological and mechanistic understanding of how cancer benefits from metabolic modifications [1]. This model-driven approach allows the discovery of potential biomarkers and drug targets [62,89]. The identification of new biomarkers is of major importance to biomedical research for early diagnosis and monitoring treatments efficiently. The identification of cancer biomarkers is possible due to aberrant metabolism of tumors that alters the profile of absorption and nutrients secretion.

Omic data of clinical samples (mainly transcriptomics data) can be used to infer the exchange rates of different metabolites for each individual sample via GSMM analysis (alterations in exchange reactions in the model).

Thus, those metabolites that significantly differ between two clinical groups in their exchange rates are then considered as potential biomarkers. However, this task is especially challenging in the case of cancer due to metabolic abnormalities resulting from complex and elaborate genetic and epigenetic alterations that modify the expression of a variety of cancer-associated isoenzymes.

In order to determine potential biomarkers in cancer, several computational approaches has been developed. For example metabolic phenotypic analysis (MPA) method uses

GPR association to integrate transcriptomics and proteomics data within a GSMM to infer metabolic phenotypes [63].

MPA was used to study breast cancer metabolism and predict potential biomarker. These predictions that include amino acid and choline-containing metabolites are supported by a number of experimental evidences [90]. Another recently developed algorithm is mCADRE which has been used to systematically simulate the metabolic function of 26 cancer cell types (among other cell types) [61]. This algorithm has been able to identify several pathways such as folate metabolism, eicosanoid metabolism, fatty acid activation and nucleotide metabolism that are enriched in tumor tissue compared to their corresponding normal tissue.

Many enzymes involved in these pathways are already used as chemotherapy targets. Other approaches such as flux variability analysis [91] or sampling analysis [92] are also suitable to predict metabolic biomarker candidates by integrating omic data into a GSMM. The novel drug discovery is based on the abnormalities existing in various reactions/pathways of cancer metabolism. These differences can be used as drug targets to attack specific weaknesses of the tumor and hence compromising its viability, but not that of non-cancerous cells [93]. For example INIT method [62] was used to identify characteristic metabolic features of cancer cells by inferring the active metabolic network of 16 different cancer types and compare them with the healthy cell types where they come from. These metabolic differences may play an important role in proliferation of cancer cells and could be potential drug targets. This method found significant differences in polyamines metabolism, the isoprenoid biosynthesis and the prostaglandins and leukotrienes pathways in cancer cells compared with healthy cells.

Some of the reactions that were found that have different activity in cancer cell, are already used in the clinical practice as therapeutic targets [94,95].

Based on the rationale that the differences between normal and tumoral cells can be potential therapeutic targets, several approaches have been developed, that considers different aspects of cancer metabolism for the discovery of new drug targets:

-Antimetabolite: One of the most common anticancer drugs are antimetabolites. An antimetabolite is structurally similar to a certain metabolite but it cannot be used to

produce any physiologically important molecule. Antimetabolite-based drugs act on key enzymes preventing the use of endogenous metabolites, resulting in the disruption of the robustness of cancer cells and reduction or suppression of cell growth. For example, antimetabolites such as antifolates or antipurines mimic folic acid and purines [96]. GSMMs approach can be used to systematically simulate the effect of potential antimetabolites in cancer research. To achieve this, methods such as tINIT (Task-driven Integrative Network Inference for Tissues) algorithm have been developed [70]. This method has been used to i) reconstruct personalized GSMMs for six Hepatocellular carcinoma patients based on proteomics data and the Human Metabolic Reaction database [62] and ii) identify anticancer drugs that are structural analogs to targeted metabolites (antimetabolites). tINIT algorithm was able to identify 101 antimetabolites, 22 of which are already used in cancer therapies and the remaining can be considered as new potential anticancer drugs.

-Synthetic lethal: The genetic lesions occurring in cancer not only promote oncogenic state but are also associated with dependencies that are specific to these lesions and absent in non-cancer cells. Two genes are considered “synthetic lethal” if the isolated mutation on either of them is compatible with the cell viability but simultaneous mutation is lethal [97]. Analogously, two genes are considered to interact in a “synthetic sick” fashion, if simultaneous mutation reduces cell fitness below a certain threshold without being lethal [97]. Enzymes encoded by genes that are in synthetic lethal or sick interactions with known, non-druggable cancer-driving mutations can be potential anti-cancer drug targets. This approach has two main advantages: first, we can indirectly target non-druggable cancer-promoting lesions by inhibiting druggable synthetic lethal interactors and secondly we can achieve a high selectivity by exploiting true synthetic lethal interactions for anti-cancer therapy. This is especially remarkable in the case of cancer-specific isoenzymes which are emerging as one of the most promising anti-cancer drug targets. GSMMs provide an excellent tool for the systematic simulation of specific pairs of gene knock-out (KO) to unveil those combinations that compromise the viability of cancer cells (synthetic lethal). By definition, gene KO is simulated by giving value zero to gene expression and the effect

of gene deletion is transferred to the metabolic reaction level by GPR association. Thus, for instance, the flux through a reaction that is associated only to one knocked-out gene would be zero. If the reaction is catalyzed by isoenzymes or complexes, the effect of a gene deletion is more complex. However, predicting the metabolic state of a cell after a gene KO is a challenging task because after the gene knock-out the system evolves into a new steady-state that tends to be as close as possible to the original steady-state [98]. To overcome these difficulties several algorithms have been developed. For example, MOMA algorithm minimizes the euclidean norm of flux differences between metabolic states of the knock-out compared with the wild type [99]. ROOM method minimizes the total number of significant flux changes from the wild type flux distribution [99]. In other words, MOMA minimizes the changes in the overall flux distribution while ROOM minimizes the number of fluxes to be modified after the gene knock-out [Fig. 3D]. As an example of employing the concept of synthetic lethality in cancer, a GSMM approach has been used to develop a genome scale network model of cancer metabolism [89]. The model predicted 52 cytostatic drug targets (40% of which were known) and further predicted combinations of synthetic lethal drug targets those that were validated using NCI-60 cancer cell collection. In a remarkable example, synthetic lethality between Haem-oxygenase and fumarate-hydratase was predicted by the GSMM approach and was also experimentally validated [100]. The number and the quality of these predictions prove the capabilities of this approach to identify synthetic lethal pairs of genes as potential novel drug target in cancer.

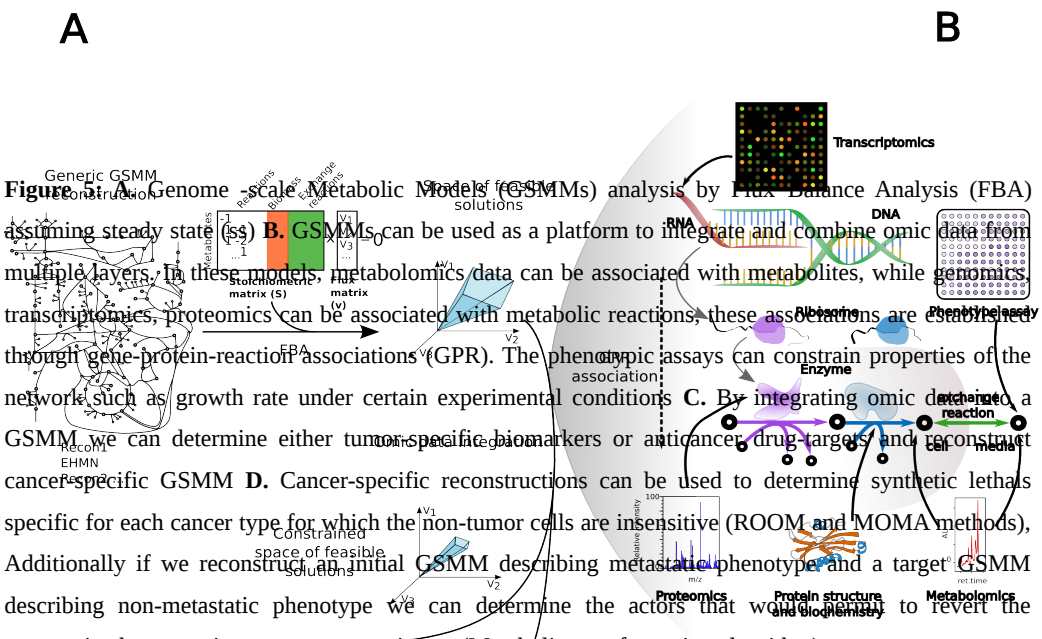
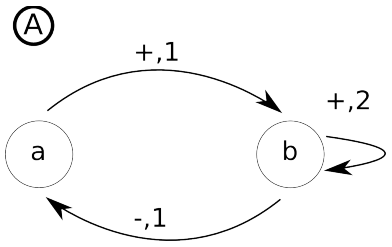



**Figure 5: A.** Genome -scale Metabolic Models (GSMMs) analysis by Flux Balance Analysis (FBA) assuming steady state (ss) **B.** GSMMs can be used as a platform to integrate and combine omic data from multiple layers. In these models, metabolomics data can be associated with metabolites, while genomics, transcriptomics, proteomics can be associated with metabolic reactions, these associations are established through gene-protein-reaction associations (GPR). The phenotypic assays can constrain properties of the network such as growth rate under certain experimental conditions **C.** By integrating omic data into a GSMM we can determine either tumor-specific biomarkers or anticancer drug-targets and reconstruct cancer-specific GSMM **D.** Cancer-specific reconstructions can be used to determine synthetic lethals specific for each cancer type for which the non-tumor cells are insensitive (ROOM and MOMA methods). Additionally if we reconstruct an initial GSMM describing metastatic phenotype and a target GSMM describing non-metastatic phenotype we can determine the actors that would permit to revert the metastatic phenotype into a non-metastatic one (Metabolic transformation algorithm).

1.3.2.3 Discrete modeling & Gene-regulatory network

An increasing number of evidences point the fact that there is no one-to-one mapping between the genotype and the phenotype [101]. In order to elucidate the fundamental principles that govern how genomic information translates into organism complexity, is necessary to overcome the current habit of ad hoc explanations and instead embrace novel holistic approaches that involve computer modeling. Most modeling approaches aim at recreating a living system via computer simulation, by including as much details as possible. In contrast, boolean and discrete network models represent an abstraction and a coarse-graining of a network, such that it can serve as a simple, efficient tool for the extraction of the very basic design principles of molecular regulatory networks (such as a gene regulatory network (GRN) or a signaling pathway), without having to deal with all the biochemical details [35,102]. Typically, the mathematical description of GRN using discrete modeling approach is based on the asynchronous logical description proposed by R. Thomas [103]. This approach allows a semi-qualitative

analysis of complex biological networks that describes the dynamics of the system [104]. GRN are often represented in the form of an interaction graph, where nodes represent genes and arrows represent interactions between genes. The mathematical equations expressing these relationships are represented in the form of focal equations in the figure 6.b. These equations determine the trend towards which each node evolves depending of the current state of the network.



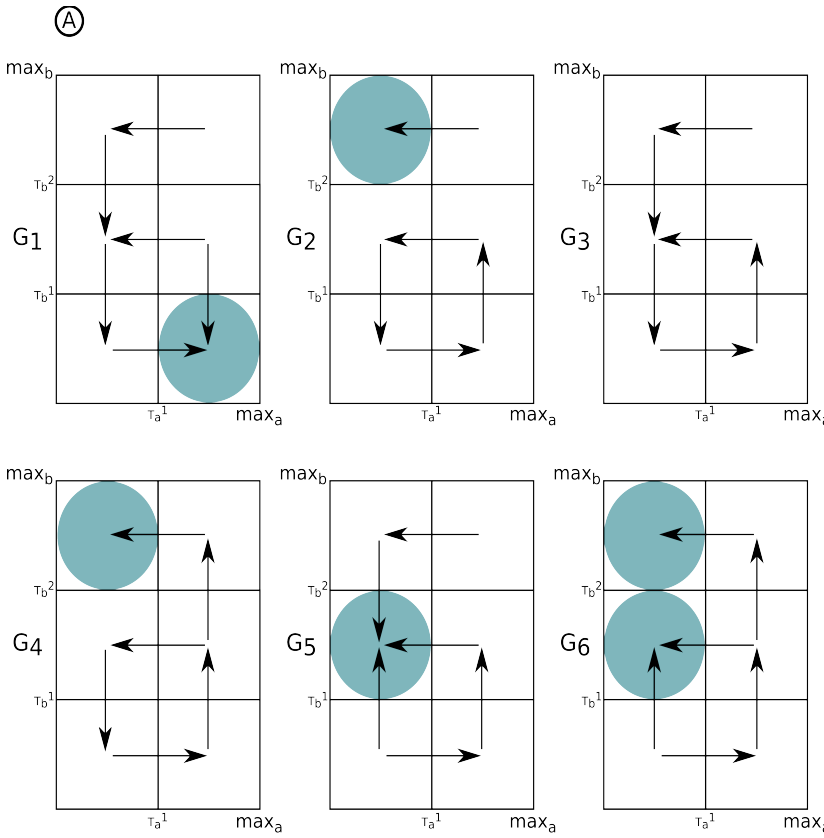

B

$$X_a = K_a * s^-(x_b, \theta_b^1) + K_a^b * s^+(x_b, \theta_b^1)$$

$$X_b = K_b * s^-(x_a, \theta_a^1)s^-(x_b, \theta_b^2) + K_b^a * s^+(x_a, \theta_a^1)s^-(x_b, \theta_b^2) + K_b^b * s^-(x_a, \theta_a^1)s^+(x_b, \theta_b^2) + K_b^{ab} * s^+(x_a, \theta_a^1)s^+(x_b, \theta_b^2)$$

Figure 6: Interaction graph. A: The figure illustrates a network with two genes represented as node “a” and “b” and the interactions are represented by arrows where + and - represents activations and inhibitions respectively and the number represents the thresholds above which the interaction is active (i.e. a activates b when a is above threshold 1). The concentrations take discrete values, each one representing an interval between two consecutive thresholds. Then, taking for instance the node a (that can represent whether a protein or a transcript) if its logical variable $x_a$=0 then the concentration of “a” is bellow the threshold $\theta^1_a$, whereas $x_a$=1 means that the concentration of “a” is higher than (or equal to) $\theta^1_a$. Here the threshold $\theta^1_a$ determines the minimum amount of “a” necessary to activate b. B: Focal equations relating a state characterized by the vector of protein concentrations [$x_a$,$x_b$]and its focal state [$X_a$,$X_b$].

Thus these equations defines the dynamic behavior of the network that can be represented as a state in a transition graph. This graphical representation shows how the network goes from one state to the next depending on its structure. Each state of the network is represented by a vector of protein concentrations that take discrete values representing the interval between two consecutive thresholds. These states are represented in figure 7 where the arrows are the transition between each state and its possible successors.




Ⓑ

$$K_a > K_a^b,$$
$$((K_b^b > K_b) \wedge (K_b^b \geq 2)) \vee$$
$$((K_b^{ab} > K_b^a) \wedge (K_b^{ab} \geq 2)),$$
$$(K_b^a > K_b) \vee (K_a^{ab} > K_b^b)$$

Ⓒ

$$K_a \geq K_a^b,$$
$$(K_b \leq K_b^b) \wedge (K_b \leq K_b^a),$$
$$(K_b^b \leq K_b^{ab}) \wedge (K_b^a \leq K_a^{ab})$$

Figure 7: Transition graph. A: Transition graphs $G_1$,...,$G_6$ of the instantiated models satisfying consistency e(Ce). Ce is the set of (b) observability and (c) additivity constraints for the example in figure 6

A specific attractor value called focal state ,[$X_1$,...,$X_n$], is associated to a given state S. It represents the expression levels toward which the genes tend to evolve. A successor of S in the graph,S'=[$x'_1$,...,$x'_n$], is deduced by comparing the value of a logical variable with that of its focal state. The transition of S to S' is assumed to be asynchronous, in

the sense that at most one logical variable xi is updated at a time. If the variable xi is updated, the formal relationship between these states is expressed as follows: $x'_i=x_i+1$ if $X_i > x_i$ and $x'_i=x_i-1$ if $X_i < x'_i$. If no logical variable $x_i$ is updated then the focal state of S is equal to S and S is its own successor: it is said steady (or stationary).

The focal state value $X_i$ of gene I depends on the state S of the network and in particular, on a set of conditions regarding the presence or absence of activators and inhibitors of gene I . For the simple example in Fig. 6, the focal state value $X_a$ of gene a depends on the influence of B on a, that is, whether the concentration of B is below ($x_b=0$) or above its first threshold value. Such interactions are expressed by means of products of step functions of the form:

$$s^{+}(x_j,\theta_j^r)=1 \quad \text{if } x_j \geq r \quad \text{else } 0$$

$$s^{-}(x_j,\theta_j^r)=1 \quad \text{if } x_j < r \quad \text{else } 0$$

Discrete network models can help to examine how gene regulatory interactions generate the coherent, rule-like behavior of a cell – the first level of integration in the multi-scale complexity of the living organism. Hereby the various cell fates, such as differentiation, proliferation and apoptosis, are treated as attractor states of the network.

This modeling language allows us to integrate qualitative gene and protein interaction data to explain a series of hitherto non-intuitive cell behaviors. Thus, discrete modeling offers a solution to sort the limitations of the current simplistic ‘one gene - one function - one target’ paradigm, allowing to develop of conceptual tools to increase our understanding of how the intricate interplay of genes gives rise to a global biological phenotype.

## 1.4 Limitations and challenges in the state of the art of computational systems medicine

In previous sections we have presented a brief introduction to cell metabolism and the different omics data used in this thesis. We have also overview the state of the art of some of the most relevant computational approaches in the study of metabolism. In this thesis we have developed and apply novel computational approaches in order to overcome the existing limitations of the computational tools for the study of metabolism at different levels: i) metabolic pathways , ii) metabolic network and iii) the cross-talk between metabolic and gene regulatory networks.

We identified important limitations in the computational tools currently used to study the metabolism in each of the levels mentioned above, these are : i) computational approaches based on non-stationary $^{13}$C-FBA does not account for the possibility that metabolite channeling exist in metabolic pathways ii) metabolic drug-target discovery based genome-scale metabolic models (GSMM) in cancer research is limited to homogeneous cell populations, being tumors heterogeneous ecosystems iii) the study of mechanism underlying the behavior of metabolic network from the point of view of crosstalk between metabolic and gene regulatory networks is limited in size and complexity due the limitations of the current computational approaches.

As we will describe following, in some cases, these limitations compromise the results of the analysis, in others it makes the analysis is unmanageable. Following these challenges are presented in more detail

### 1.4.1 Limitations and challenges in the determination of metabolite channeling: glycolytic channeling in hepatocytes as a case of concept

Metabolic pathways are collections of enzyme-mediated reactions. Due the nonlinear nature of the interactions between enzymes and metabolites, computational methods are required to describe their complex and dynamic behavior [105-107].

Kinetic metabolic models are widely used in the study of metabolic pathways [108]. This computational approach describes the metabolic reactions at the enzymatic level

incorporating the kinetic rate laws and their associated parameter values [109-111], allowing the study of important behaviors in biological networks such as oscillations [112,113 ] or bi-stability [114,115]. This approach permits to infer the rate of turnover of molecules through the metabolic reactions providing information about the activity state of metabolic pathways.

The integration of stable isotope-labeled precursor of metabolic pathways, mainly $^{13}C$-labeld substrates, is the most commonly used experimental technique to infer internal metabolic fluxes [12]. These intracellular fluxes cannot be measured directly, but can be estimated through interpretation of stable isotope patterns in metabolites using kinetic models. It is used to understand the adaptive changes of fluxes in various experimental conditions with evident applications in biomedicine [51]. $^{13}C$ labeling is applied to study metabolism in primary cultures [52], cultured cancer cells [53], and organisms such as fungi [54] or yeast [54,55], but also *in vivo* studies in animals [116]. Various methods of optimization were developed for metabolic flux estimation from isotopomer distributions [48].

However, the existing tools were designed mainly for the evaluation of canonical metabolic pathways without taking into account the specificity of network topology. Ignoring the network topology, or in other words its compartmental structure, can compromise the results of metabolic flux analysis [117]. Thus, in a compartmentalized network, the same compound existing in different sub-cellular spaces likely possesses compartment-specific $^{13}C$ signatures, although metabolite extraction before mass spectrometry or nuclear magnetic resonance analysis allows the measurement of an average $^{13}C$-pattern [19-21]. Compartmentalization of cellular metabolism not only appears as a well known membrane-separated organelles, but also as metabolite channeling. A metabolite is channeled through a chain of sequential reactions if a molecule, produced by one of co-localized enzymes, has a higher probability of being used in the next reaction than a molecule of the same species produced in another cellular location. Thus, when the channeled metabolites, although not separated by a membrane, have a restricted capacity of mixing with molecules situated out of the

channel. The existence of such not investigated and unexpected compartments significantly complicate the tracer-based flux analysis. Experimentally, channeling could be determined by the measurement of i) utilization of externally added metabolites versus that produced in the chain of consecutive reactions, or ii) diffusion of intermediates to the outside medium. However, it can be expected that channeling occur only in intact cells. Moreover, one cannot exclude the possibility that the metabolic channeling and compartmentalization differ between the cell from various tissues and this can unlimitedly multiply the number of experiments necessary for defining the structure of metabolism in cells.

The measurements of this phenomenon can be performed by tracing the products from isotropically labeled substrates. Hardin and Finder [118] showed that the metabolites of glycolysis downstream of fructose 1,6-bisphosphate are predominantly channeled in vascular smooth muscle cells and likewise, Cascante [119] showed that in permeabilized hepatocytes in the presence of saturating concentrations of fructose 2,6-bisphosphate there is channeling in glycolysis between glucose and fructose-1,6-bisphosphate. These and other similar studies using isotope tracer support the hypothesis of channeling in glycolysis. In general, when isotope tracers are used, the restriction of metabolite mixing in intracellular space by channeling could significantly affect the pattern of isotopic isomer distribution in metabolites.

The experimental difficulties and the computational limitations exposed above, point out the necessity to develop new computational tools allowing the evaluation of metabolic topology within the pathways, and specially metabolic channeling due the existent technical difficulties to be measured. Among the different kinetic modeling approaches, those based on dynamic $^{13}$C-FBA are specially suitable for the study of internal metabolic fluxes [53]. A software based on this approach, such as Isodyn [53], can be further developed to account for metabolic channeling in order to evaluate and quantify this phenomenon. This is the basis of the work developed in the first chapter of this thesis, using as a case of concept the metabolic channeling in hepatocytes.

### 1.4.2 Limitations and challenges in metabolic drug-target discovery based on GSMMs due tumor heterogeneity: prostate cancer as a case of concept

As is discussed above, GSMMs offer an excellent platform for the analysis of the aberrant metabolism in diseases with a strong metabolic component such as cancer. This systems biology tool integrates the complete set of chemical reactions that occur in living organisms enabling the study of the metabolism in a holistic manner [69].

The abnormalities of cancer metabolism provide a basis for novel drug discovery. These differences can be used as drug target to attack specific weaknesses of the tumor and hence compromise its viability, but not that of non-cancerous cells [120]. Based on this rationale several approaches have been developed that consider different aspects of cancer metabolism for the discovery of new drug targets. A variety of methods and algorithms has been developed for the systematic evaluation of potential metabolic drug targets such as ROOM or MOMA [121]. This approaches has been applied successfully in a variety drug targeting discover analysis in the scope of cancer research [2,100]. These studies has identified numerous single drug target or combinations of metabolic drug targets that separately have no effect on tumor viability but in combination are lethal (synthetic lethal) [122]
However these analysis usually considers tumor as a homogeneous group of non differentiated cells. This can lead to erroneous conclusions and limits the capability to estimate potential drug targets.

The intratumoral heterogeneity confers an extreme flexibility and adaptation capability to cancer cells that enhances tumor progression and represents a challenge for target-directed therapies [123]. The intratumoral heterogeneity is driven by two main processes i) epithelial-mesenchymal-transition, by which epithelial cells gain invasive properties and lose at least part of their epithelial phenotypes [124] and ii) mesenchymal-epithelial-transition by which mesenchymal cells can revert to an epithelial gene program displaying strong self-renewal and survival properties [125-127].

Then, for example, it has been characterized two clonal subpopulations of a the prostate cancer cell line PC-3 presenting opposite cancer stem cell features in their phenotypes: PC-3/M (Cancer stem Cell-enriched) and PC-3/S (non-Cancer Stem Cell) [124]. It has been also hypothesized that these subpopulations can present a collaborative relationship, leading to an enhanced tumor malignancy [127].

Tumor heterogeneity represents a hurdle that must be overcome in order to develop new and more efficient anti-cancer therapies. Better understanding of how tumor heterogeneity protects cancer cells is imperative to design new therapies aimed at targeting this tumor-protective niche [128,129].

The study of the metabolism in heterogeneous cellular populations to identify potential drug targets must be overcome from an holistic perspective by developing novel computational approaches based on GSMMs. This is rationale has motivated the work that is explained in detail in chapter 2.

### 1.4.3 Limitations and challenges integrating probabilistic and deterministic approaches to study the crosstalk between GRN and metabolism: abnormal adaptation to training in COPD patients as a case of concept

Systems medicine aims to explain physiology, development, and pathology based on modular networks of expression, interaction, regulation, and metabolism [130,131]. Gene regulatory networks (GRNs) control the adaptation to environmental perturbations via cell type-specific gene expression and interactions between transcription factors (TFs) and regulatory promoter regions [132]. These mechanisms may affect the metabolic network by regulating the activity of key enzymes. One of the major challenges in systems medicine is the study of the crosstalk between metabolic and gene regulatory networks and its role in cellular response to environmental conditions and external perturbations [133]. This is especially crucial in the case of multi-factorial diseases, in which multiple regulatory mechanisms lead the cell to an aberrant metabolism and abnormal adaptation to environment conditions.

In this sense Chronic obstructive pulmonary disease (COPD) is affected by a variety of factors leading to alterations in lung function such as expiratory flow limitation (low forced expiratory volume during the first second, $FEV_1$) and abnormal pulmonary gas exchange (arterial hypoxemia) [134]. This disease, is a prevalent chronic condition in approximately 9% of the adult population above 45 years (Vestbo J GOLD 2013), imposing a high burden on health systems worldwide being the third leading cause of death in the United States [135]. Skeletal muscle dysfunction is a characteristic systemic effect of the disease that affects approximately one-third of patients [136]. It has a multi-factorial nature [136], but it is acknowledged that high reactive oxygen species (ROS) generation in the mitochondrial electronic chain (EC), as well as nitroso-redox disequilibrium, are likely play a relevant role in both pulmonary and systemic manifestations of the disease affecting the GRN [137]. Moreover, altered skeletal muscle redox status plays a central role in the abnormal adaptive changes observed in COPD patients after endurance training [138]. It has been hypothesized that the systemic effects of the disease might share abnormal regulation of key metabolic pathways [139]. Thus, the state of the mitochondrial EC depends on carbohydrate metabolism and transport of adenine nucleotides, which, in turn, depends on the mitochondria-cytoskeleton interactions. All these factors can have a functional impact on the respiratory chain and on mitochondrial ROS production through complex III [140].

A large number of data-driven approaches has been applied in the study of COPD [141,142]. An example of this approaches are Bayesian networks based statistical analysis such as correlation analysis [143]. These approaches although being able to study causalities and give a rough idea of the processes associated with a given disease, are not suitable to study the dynamic of complex mechanisms. On the other hand, mechanistic models approaches including kinetic models are based on a detailed representation of the system and their parameters. This approaches are frequently infeasible for large scale and highly interconnected networks because of a paucity of necessary parameter values. Finally discrete models representing an abstraction and a

coarse-graining of a network, can serve as a simple, efficient tool for the extraction of basic design principles of molecular regulatory networks, without having to deal with all the biochemical details [102]. However the study of large and highly inter-connected networks becomes unworkable due the exponential increase of parameters that cannot be compensated by an usually limited amount of reliable experimental data. This lack of suitable computational tools makes that, in spite the efforts done in this area, our knowledge of the mechanisms underlying the abnormal metabolic adaptation to training in COPD patients remains still limited.

Thus, development a novel computational approach that overcomes the limitation exposed above needs to be carried out in order to handle large and complex networks such as those involved in the abnormal adaptation to training in COPD patients.

To face this challenging task is necessary to develop an approach that integrates the capability of probabilistic methods to analyze big data and the the possibility to analyze complex regulatory mechanism with the minimum amount of data provided by discrete model-driven approaches. This idea is developed in the third chapter of this thesis where it is explained in detail.

# 2. Objectives

The general aim of this thesis is to develop and apply new computational tools to overcome existing limitations in the analysis of metabolism. This thesis is focused on developing new computational strategies to overcome the following identified limitations: i) the existing metabolic flux analysis tools does not account for the existence of metabolic channeling ii) Metabolic drug-target discovery based on GSMM does not consider the different cell subpopulations existing within the tumor and iii) current mechanistic and probabilistic computational approaches are not suitable to study the complexity of the crosstalk between metabolic and gene regulatory networks. More specifically, this thesis is divided in three chapters, each of them aimed to overcome one of the limitations summarized above. The main aims of this thesis are summarized in the following bullet list:

- To develop a new computational tool based on non-stationary $^{13}$C-FBA to evaluate different models reflecting different topologies of intracellular metabolism, using the channeling in hepatocytes as case of concept. To achieve this specific aim the following objectives have been defined:
    - To develop, test and implement a new function for Isodyn software package, which performs consistent changes in the basic kinetic model and the in module for simulation of isotopomer distribution.
    - To determine the existence of channeling at the level of hexoses 6-phosphate in hepatocytes using the developed function

- To develop and apply a new method to perform a comparative Genome-scale metabolic model reconstruction analysis accounting for intra-tumoral heterogeneity, using prostate cancer (pc) as a case of concept. To achieve this specific aim the following objectives have been defined:
    - To develop a method that integrate trancriptomic data into a comparative genome-scale metabolic network reconstruction analysis

in the context of intra-tumoral heterogeneity

- To determine subpopulation-specific drug targets
- To Determine a metabolic gene signature associated to tumor progression in pc and explore correlations with other types of cancer.

- To develop a novel computational method combining probabilistic and mechanistic approaches to integrate multi-level omic data into a discrete model-based analysis. This method is aimed to analyze the mechanism underlying the crosstalk between metabolism and gene regulation, using as case of concept the study of the abnormal adaptation to training in COPD patients. To achieve this aim specific the following objectives have been defined:
  - To reconstruct a COPD and muscle specific GRN based on transcriptomic and literature-based data
  - To build a discrete model based on the GRN previously reconstructed and constraints based on probabilistic approaches
  - To determine the mechanism underlying the abnormal adaptation to endurance training in COPD patient

# 3. Inform of the directors:

The work developed by Igor Bartolomé Marín de Mas in this PhD thesis has resulted in two publication and other three that are currently in process to be published in international scientific journals. The PhD student is the first author of three of the five articles that integrate this thesis. Additionally other six publications are included in the appendix. These publications include one book chapter one scientific paper and four reviews from which one the PhD student is the first author. All the results included in this thesis have not been presented in any other thesis.

Following is described the list of all the articles included in the thesis, indicating the impact factor of the journal in which the article was published or has been submitted and specifying the tasks of the PhD student in each one of them:

1. **Igor Marin de Mas** , Vitaly A Selivanov , Silvia Marin, Josep Roca , Matej Oresic, Loranne Agius and Marta Cascante. **Compartmentation of glycogen metabolism revealed from 13C isotopologue distributions.** *BMC Systems Biology* 2011, 5:175 doi:10.1186/1752-0509-5-175.
   Impact factor: 3.148
   URL: http://www.biomedcentral.com/1752-0509/5/175
   Igor Marín developed the metabolic network model and performed all the computational analysis. He also participated in the development of a computational algorithm to adapt Isodyn to the analysis of alternative metabolic pathway topologies.. He also wrote the first version of the manuscript.

2. **Igor Marin de Mas**, Esther Aguilar, Erika Zodda, Timothy M Thomson, Balázs Papp, and Marta Cascante. **Network model-driven discovery of calcitriol metabolism and fatty acid oxidation as therapeutic targets in metastatic prostate cance.** *Submitted to Mol. Syst. Biol. January 2015.*
   *Impact factor:* 14.099

Igor Marín performed all the analysis and the development of the computational tool to integrate transcriptomic data into a subpopulation-specific Genome-scale metabolic network reconstruction. He has written the first version of the manuscript and he is has been in charge to refine the article with the feedback of the co-authors.

3. Esther Aguilar, **Igor Marín de Mas**, Erika Zodda, Silvia Marín, Fionnuala Morrish, Vitaly Selivanov, Óscar Meca-Cortés, Hossain Delowar, Mònica Pons, Inés Izquierdo, Antoni Celià-Terrassa, Pedro de Atauri, Josep J Centelles, David Hockenbery, Timothy M Thomson and Marta Cascante **Metabolic landscape and vulnerabilities of prostate metastatic epithelial cancer stem cell independent of epithelial-mesenchymal transition.** *Submitted to Cell.Metab. Biol. January 2015.*
   *Impact factor:* 16.747
   Igor Marín performed all the statistical analysis to determine a gene signature related with the tumoral progression in prostate cancer and its comparison with other cancer types. He also participated in the interpretation of the results and in writing the part of results and discussions of the manuscript related with these data analysis.

4. Dieter Maier , Wenzel Kalus , Martin Wolff , Susana G Kalko , Josep Roca , **Igor Marin de Mas** , Nil Turan , Marta Cascante , Francesco Falciani , Miguel Hernandez , Jordi Villa-Freixa and Sascha Losko. **Knowledge management for Systems Biology a general and visually driven framework applied to translational medicine.** *BMC Systems Biology* 2011, 5**:**38doi:10.1186/1752-0509-5-38. Published: 5 March 2011.
   Impact factor: 3.148
   *URL: http://www.biomedcentral.com/1752-0509/5/38*

Igor Marín performed a data-mining analysis on a set of relevant publications and incorporated all the relevant information into BioMax database. He also participated in writing the article

5. **Igor Marin de Mas**, Fanchon E., Vitaly A. Selivanov, Josep Roca and Marta Cascante. **A novel discrete model-driven approach unveils abnormal metabolic adaptation to training in COPD.** *Submitted to Mol. Syst. Biol. January 2015.*

   *Impact factor: 14.099*

   Igor Marín performed all the computational analysis, the manual curation of the Gene regulatory Network and the development of the computational tools . He has been in charge of writing the first version of the article and has been also in charge of refining it with the feedback of the co-authors.

6. Través PG, de Atauri P, Marín S, Pimentel-Santillana M, Rodríguez-Prados JC, **Igor Marín de Mas**, Selivanov VA, Martín-Sanz P, Boscá L, Cascante M.. **Relevance of the MEK/ERK Signaling Pathway in the Metabolism of Activated Macrophages: A Metabolomic Approach.** J Immunol. 2012 doi: 10.4049*/jimmunol.1101781 . Published 2011 Dec 21.*

   *Impact factor:* 5.520

   URL: http://www.ncbi.nlm.nih.gov/pubmed/22190182

   Igor Marín participated in the computational analysis. He also participated in writing the article

7. Marta Cascante, Adrián Benito, **Igor Marín de Mas**, Josep J. Centelles, Anibal Miranda and Pedro de Atauri. **A Systems Biology Approach to Study Metabolic Syndrome.** *Springer, Switzerland, 2013, (6):237-50. Published 2013*

URL: HYPERLINK "http://www.springer.com/biomed/book/978-3-319-01007-6"http://www.springer.com/biomed/book/978-3-319-01007-6

Igor Marín participated in the writing this review article

8. **Marín de Mas Igor**, Aguilar Esther, Jayaraman Anusha, Polat Ibrahim H., Martín-Bernabé Alfonso, Bharat Rohit, Foguet Carles, Milà Enric, Papp Balázs, Josep J. Centelles and Cascante Marta. **Cancer cell metabolism as new targets for novel designed therapies.** Future Med Chem. 2014;6(16):1791-810.

   *Impact factor: 4*

   *URL*: http://www.future-science.com/doi/abs/10.4155/fmc.14.119

   Igor Marín participated in writing this review article and in the coordination of the work of the different co-authors

9. Roca Josep, Vargas Claudia, Cano Isaac, Selivanov Vitaly, Barreiro Esther, Maier Dieter, Falcian Francesco, Wagner Peter, Cascante Marta, Garcia-Aymerich Judith, Kalko Susana, **Marin Igor**, Tegner Jesper, Escarrabill Joan, Agustí Alvar, Gomez-Cabrero David. **Chronic Obstructive Pulmonary Disease Heterogeneity. Challenges for Health Risk Assessment, Stratification and Management**. J Transl Med. 2014 Nov 28;12 Suppl 2:S3.

   *Impact factor:* 3.991

   URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4255905/

   Igor Marín participated in writing this review article

10. David Gomez-Cabrero, Jörg Menche, Isaac Cano, Imad Abugessaisa, Mercedes Huertas-Migueláñez, Akos Tenyi, **Igor Marin de Mas**, Narsis Kiani, Francesco Marabita, Francesco Falciani, Kelly Burrowes, Dieter Maier, Peter Wagner, Vitaly Selivanov, Marta Cascante, Josep Roca, Albert-László

Barabási, Jesper Tegner. **Systems Medicine: from molecular features and models to the clinic in COPD.** J Transl Med. 2014 Nov 28;12 Suppl 2:S4.

*Impact factor:* 3.991

URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4255907/

Igor Marín participated in writing this review article

11. Cascante Marta,de Atauri P, Gomez-Cabrero David, Wagner Peter, Centelles JJ, Marin S, Cano Isaac, Velickovski Felip, **Marin de Mas Igor**, Maier Dieter, Roca Josep, Sabatier Philippe. **The Biohealth Computing Model for Master and PhD students.** J Transl Med. 2014 Nov 28; 12(Suppl 2): S11.

    *Impact factor:* 3.991

    URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4255883/

    Igor Marín participated in the design of Biohealth computing thematic described in this manuscript

# 04. Summary of the results, global discussion and final remarks

## 4.1 Summary of the results

In this thesis we have developed and apply novel computational strategies to overcome various limitations in some of the most relevant computational approaches necessaries for the study the metabolism from the perspective of systems medicine.

As it is exposed in previous sections, we can divided our achievements in three main parts that corresponds with the three chapters in which is divided this thesis: i) development of a new computational method to study the topology of metabolic pathways taking as a proof of concept the study of channeling in the upper part of glycolysis in hepatic cells, ii) developing new strategies based on GSMM analysis to study metabolic vulnerabilities associated to intra-tumoral heterogeneity in prostate cancer and analysis of tumor progression in prostate cancer by combining different data-driven approaches and finally iii) the development of a novel computational method that combines probabilistic and mechanistic approaches to integrate multilevel omic data into a discrete model-based analysis, using as case of concept the abnormal adaptation to trainin observed in COPD patients.

Following are summarized the results obtained this thesis, divided by their corresponding chapters:

### 4.1.1 (Chapter 01) Study of the metabolic channeling in hepatocytes by developing a novel model-driven approach based on non-stationary $^{13}$C-FBA

In this chapter we overcome the existing limitations in the analysis of the topology of metabolic pathways.

More specifically we developed a computational method to evaluate and discriminate different metabolic topologies based on their consistency with the experimental isotopic distribution of substrates . This chapter summarizes the result from these analysis which resulted in a publication

Chapter 1a. **Compartmentalization of glycogen metabolism revealed from 13C isotopologue distributions.**

The dynamics of all possible isotopic isomers in glucose, lactate and glutamate from the incubation medium and glucose from glycogen in cell pellets accumulated by two hours of incubation of liver cells with [1,2-$^{13}C_2$]D-glucose [33] was simulated with Isodyn [53] using two different schemes that included either one or two pools of hexose phosphates as is shown in Figure 1.

The model, which does not consider channeling (further referred as model A), accounts for only one well mixed common hexose 6-phosphate pool (Figure 1A). In accordance with the definition given in the introduction, channeling assumes the existence of metabolite compartments, which could have different isotopologue composition and does not freely mix by diffusion with a pool of the same metabolite outside of the channel. The presence of two compartments with different isotopomer composition indicates metabolite channeling (Figure 8B), and the respective model is referred further as model B.

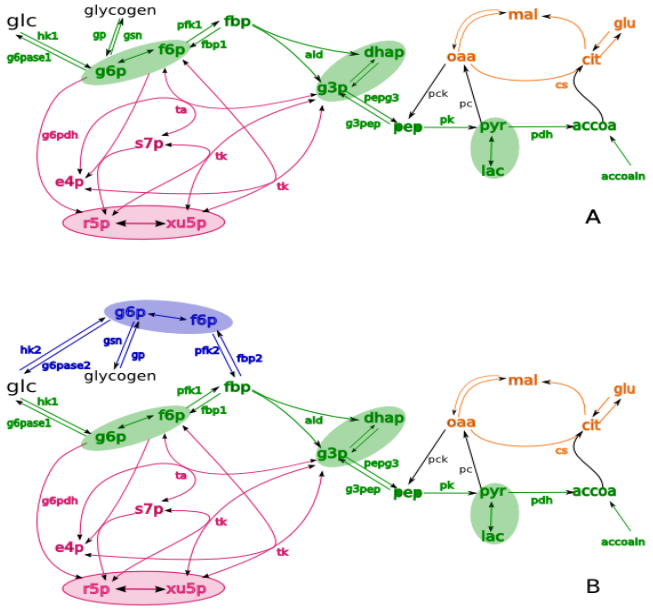


Figure 8 - The schemes of kinetic models used as a base for simulation of isotopologue distribution. Metabolites are connected by biochemical reactions represented by arrows, Various colors indicate metabolites and reactions of specific pathways: green, glycolysis/gluconeogenesis; red, pentose phosphate pathways; orange, TCA cycle. The metabolites enclosed in ellipses are considered to be in fast equilibrium. (A), the basic model that includes one pool of hexose phosphates common for glycolysis and gluconeogenesis. (B), the model that includes also the additional pool of hexose phosphates (blue) that represents channeling in gluconeogenesis.

We used $\chi^2$ criteria to evaluate the accuracy of our model predictions. Additionally, since we were comparing two schemes with different number of reactions and consequently different degrees of freedom, we use the statistic Q. This statistic is defined by the inverse gamma Q function that gives the probability that a model with F degrees of freedom is correct and $\chi^2$ by chance could exceed a determined value.

Thus, the model A that does not assume metabolite channeling in glycogen synthesis fitted the experimental isotopologue distribution with $\Sigma_1\chi 2=38.28$ (Table 1). 22 experimental points and 16 essential parameters, calculated as described in the chapter 1, defined number of degrees of freedom F=6. The given values of $\Sigma_1\chi^2$ and the number of degrees of freedom define an extremely low value of incomplete gamma function $Q=9.9\cdot10^{-7}$, that unambiguously indicates that the model that does not account for channeling should be rejected [34]. Conversely, the model B that assumes channeling, fits the measured isotopologue distribution much better, $\Sigma_1\chi^2=3.13$ (Table 1). This model has four degrees of freedom, deduced from the same number of experimental points and 18 essential parameters. These values defined Q=0.536, which allowed the acceptance of the model.

| | Experiment | | | Simulated | |
|---|---|---|---|---|---|
| | mean | | sd | **Channeling** | **Mixed** |
| **Glucose** | | | $\chi^2$= | **0,442** | **0,406** |
| m0 | 0.512 | ± | 0.0069 | 0.511 | 0.511 |
| m1 | 0,00913 | ± | 0.002 | 0.0084 | 0.00839 |
| m2 | 0,478 | ± | 0.00652 | 0.481 | 0.481 |
| [mM] | 19.7 | ± | 1.92 | 20.4 | 20.2 |
| **Lactate** | | | $\chi^2$= | **1,43** | **7,32** |
| m0 | 0.86 | ± | 0.0482 | 0.839 | **0.81** |
| m1 | 0,0235 | ± | 0.00802 | 0.0237 | **0.0178** |
| m2 | 0,0946 | ± | 0.0388 | **0.133** | **0.17** |
| m3 | 0,022 | ± | 0.0438 | 0.00381 | **0.00145** |
| [mM] | 0,81 | ± | 0.51 | 0.959 | **1.48** |
| **Glutamate C2-C5** | | | $\chi^2$= | **0,0564** | **0,0424** |
| m0 | 0,912 | ± | 0,0343 | 0,912 | 0,912 |
| m1 | 0,0299 | ± | 0,0116 | 0,0301 | 0,0298 |
| m2 | 0,0523 | ± | 0.0217 | 0,0574 | 0,0567 |
| **Glutamate C2-C4** | | | $\chi^2$= | **0,0049** | **0,00437** |
| m0 | 0,919 | ± | 0,0339 | 0,919 | 0,919 |
| m1 | 0,0365 | ± | 0,00175 | 0,0356 | 0,355 |
| m2 | 0,0446 | ± | 0,0166 | 0,0454 | 0,0451 |
| **Glycogen** | | | $\chi^2$= | **1,2** | **30,5** |
| m0 | 0.608 | ± | 0,0388 | 0.598 | **0.658** |
| m1 | 0,0162 | ± | 0.0033 | 0.0151 | **0.0271** |
| m2 | 0.362 | ± | 0.0351 | 0.375 | **0.299** |
| m3 | 0,00399 | ± | 0.0011 | 0.00422 | **0.00791** |
| m4 | 0,00961 | ± | 0.0026 | 0.00748 | 0.00749 |
| m5 | 0,000464 | ± | 0,00016 | 0.000432 | 0.000533 |
| mg/mL | 0.355 | ± | 0,112 | 0.313 | **0.232** |
| $\Sigma_1 \chi^2$ | | | | **3.13** | **38,28** |
| **Glycogen C1-C4** | | | $\chi^2$= | **1,42** | **6,68** |
| m0 | 0.613 | ± | 0.0448 | 0.627 | **0.679** |
| m1 | 0.0224 | ± | 0.00834 | **0.0133** | 0.0297 |
| m2 | 0.357 | ± | 0.0425 | 0.358 | **0.289** |
| **Glycogen C3-C6** | | | $\chi^2$= | **7,97** | **30** |
| m0 | 0.952 | ± | 0.00767 | 0.952 | 0.951 |
| m1 | 0.00743 | ± | 0.00211 | **0.0131** | **0.018** |
| m2 | 0.0371 | ± | 0.00467 | 0.0333 | **0.0279** |
| $\Sigma_2 \chi^2$ | | | | **9,21** | **36,68** |
| $\Sigma_t \chi^2 = \Sigma_1 \chi^2 + \Sigma_2 \chi^2$ | | | | **12,52** | **74,95** |

Table 1- Measured and simulated fractions of isotopologues and total concentrations of metabolites. Isotopologues (m0, non-labeled; m1, containing one 13C isotope; m2, two 13C isotopes, etc) produced by isolated hepatocytes from glucose as the only substrate contained 50% of [1,2-13C2]D-glucose were measured in glucose from medium, glucose from glycogen and its fragments, lactate, and fragments of glutamate after two hours of incubation. The measurements are presented as mean ± standard deviation. The data were simulated using two models that either accounted for channeling or suggested a single "mixed" pool of hexose phospates in accordance with the schemes presented in Figure 8. The fitting was performed using a stochastic algorithm based on linear anealing algorithm.

In addition the model was further validated by performing new analysis that included the quantification of the fragments 1-4 and 3-6 of of glycogen (Table 1 $\Sigma_2\chi 2$). It permitted to evaluate more in detail the effect of channeling in the isotopic distribution of glycogen. Another validation came from a series of two experiments where hepatocytes were incubated in the presence of glucose and lactate.

The conditions in the two experiments were identical with the exception that in one of these experiments only glucose was labeled [144], and in the other one only lactate was labeled. Thus, despite the labeling in the substrates is different, the metabolic fluxes are equal in both conditions.

All these validations provided the same conclusions, rejecting the model that did no assume metabolite channeling in glycogen synthesis (model A) and accepting the model that included metabolic channeling (model B).

In addition, we determined the distribution of metabolic fluxes in the accepted model, the set of metabolic fluxes is shown in table 2

| | Glucose as the only substrate | | | Glucose with lactate | | | |
|---|---|---|---|---|---|---|---|
| | | 99% confidence interval | | | 99% confidence interval | | |
| | bestfit | ( min - max ) | | bestfit | ( min - max ) | | model A |
| hk1 | 0.0026894 | ( 0,002150 - 0,003080 ) | | 0,1523 | ( 0,0748 - 0,3774 ) | | 0,0299 |
| hk2 | 0.0021436 | ( 0,001710 - 0,002480 ) | | 0,0057 | ( 0,0042 - 0,0076 ) | | |
| g6pase1 | 5,50E-05 | ( 2,70E-05 - 7,80E-05 ) | | 0,1506 | ( 0,0731 - 0,3756 ) | | 0,0285 |
| g6pase2 | 2,13E-05 | ( 0,0 - 0,000066 ) | | 0,0052 | ( 0,0029 - 0,0072 ) | | |
| pfk1 | 0.003048 | ( 0,002370 - 0,003530 ) | | 0,0024 | ( 0,0015 - 0,0047 ) | | 0,0058 |
| pfk2 | 0.0004816 | ( 0,0 - 0,000950 ) | | 0,0008 | ( 0,0006 - 0,0013 ) | | |
| fbpase1 | 0.0004446 | ( 0,000270 - 0,000560 ) | | 0,0013 | ( 0,0006 - 0,0024 ) | | 0,0059 |
| fbpase2 | 0.0005496 | ( 0,000330 - 0,000690 ) | | 0,0023 | ( 0,0016 - 0,0032 ) | | |
| gp | 2,56E-06 | ( 1,70E-06 - 5,10E-06 ) | | 0,0001 | ( 0,0001 - 0,0002 ) | | 2,33E-005 |
| gs | 0.0021929 | ( 0,001660 - 0,002680 ) | | 0,0022 | ( 0,0019 - 0,0027 ) | | 0,0016 |
| aldf | 0.0192413 | ( 0,007480 - 0,022580 ) | | 0,0159 | ( 0,0153 - 0,0164 ) | | 0,0156 |
| aldr | 0.016706 | ( 0,004480 - 0,020410 ) | | 0,0164 | ( 0,0153 - 0,0167 ) | | 0,0157 |
| aldex | 0.0424386 | ( 0,010420 - 0,056150 ) | | 0,0137 | ( 0,0104 - 0,0196 ) | | 0,0201 |
| g3pep | 0.0064103 | ( 0,004490 - 0,007700 ) | | 0,2583 | ( 0,1678 - 0,3394 ) | | 0,1627 |
| pepg3 | 0.0013391 | ( 0,000140 - 0,002370 ) | | 0,2587 | ( 0,1663 - 0,3396 ) | | 0,163 |
| pk | 0.0050704 | ( 0,003910 - 0,006230 ) | | 0,02 | ( 0,0157 - 0,0239 ) | | 0,015 |
| lacin | 1,71E-07 | ( 1,20E-07 - 2,60E-07 ) | | 0,2317 | ( 0,1715 - 0,2686 ) | | 0,1375 |
| lacout | 0.00507 | ( 0,003910 - 0,006230 ) | | 0,2223 | ( 0,1621 - 0,259 ) | | 0,1251 |
| pc | 1,20E-07 | ( 5,70E-08 - 1,90E-07 ) | | 0,0204 | ( 0,0155 - 0,0233 ) | | 0,0154 |
| pepck | 1,16E-08 | ( 5,30E-9 - 2,80E-8 ) | | 0,0204 | ( 0,0155 - 0,0233 ) | | 0,0153 |
| maloa | 1,85E-07 | ( 9,60E-8 - 3,00E-7 ) | | 0,0836 | ( 0,0489 - 0,1119 ) | | 0,0341 |
| oamal | 3,80E-08 | ( 1,30E-8 - 8,10E-8 ) | | 0,0747 | ( 0,0368 - 0,0998 ) | | 0,0249 |
| cs | 2,55E-07 | ( 1,30E-7 - 4,00E-7 ) | | 0,0089 | ( 0,007 - 0,0148 ) | | 0,0092 |
| citmal | 1,47E-07 | ( 8,40E-8 - 2,30E-7 ) | | 0,0089 | ( 0,0069 - 0,0147 ) | | 0,0091 |
| pdh | 2,75E-07 | ( 7,70E-8 - 5,00E-7 ) | | 0,0089 | ( 0,007 - 0,0148 ) | | 0,0092 |
| g6pdh | 3,87E-06 | ( 2,80E-6 - 7,50E-6 ) | | 0,0019 | ( 0,0014 - 0,0021 ) | | 6,35E-005 |
| p5p->s7p | 0.0011849 | ( 0,000880 - 0,002270 ) | | 0,0006 | ( 0,0004 - 0,0007 ) | | 0,0005 |
| s7p->r5p | 0.0011925 | ( 0,000890 - 0,002280 ) | | 4,35E-6 | ( 2,56E-6 - 2,98E-5 ) | | 0,0005 |
| f6p->p5p | 9,59E-06 | ( 4,20E-6 - 3,40E-5 ) | | 4,02E-5 | ( 1,17E-5 - 9,84E-5 ) | | 6,11E-006 |
| p5p->f6p | 4,93E-06 | ( 2,80E-6 - 2,80E-5 ) | | 0,0007 | ( 0,0005 - 0,0008 ) | | 1,17E-005 |
| f6p->s7p | 1,85E-05 | ( 9,20E-6 - 7,30E-5 ) | | 1,40E-5 | ( 4,56E-6 - 2,45E-5 ) | | 1,15E-005 |
| s7p->f6p | 9,56E-06 | ( 4,80E-6 - 6,60E-5 ) | | 1,65E-6 | ( 9,03E-7 - 8,56E-6 ) | | 2,08E-005 |
| p5p<->g3p | 0.0006152 | ( 2,70E-4 - 2,83E-3 ) | | 0,0018 | ( 0,0008 - 0,0033 ) | | 0,0003 |
| f6p<->s7p | 7,69E-08 | ( 2,10E-8 - 7,70E-7 ) | | 1,53E-5 | ( 4,62E-6 - 2,78E-5 ) | | 2,44E-007 |
| p5p<->s7p | 0.0022967 | ( 1,37E-3 - 7,06E-3 ) | | 1,51E-6 | ( 8,37E-7 - 8,16E-6 ) | | 0,001 |
| f6p->s7p | 0.0015144 | ( 0,000850 - 0,001940 ) | | 0,0009 | ( 0,0003 - 0,0021 ) | | 0,0017 |
| s7p->f6p | 0.0015015 | ( 0,000850 - 0,001920 ) | | 0,0015 | ( 0,0008 - 0,0027 ) | | 0,0017 |
| f6p<->g3p | 0.0075338 | ( 0,002170 - 0,013330 ) | | 0,0164 | ( 0,0085 - 0,0352 ) | | 0,0041 |
| s7p<->e4p | 0.0003018 | ( 0,000140 - 0,000550 ) | | 0,0001 | ( 2,01E-5 - 0,0002 ) | | 0,0007 |
| | | | | | | | |
| $\chi^2$ | 12,52 | | | 36 | | | |

Table 2: Metabolic fluxes corresponding to the best fit of experimental data and their 99% confidence intervals in model B. The name of fluxes correspond to the abbreviation used in Figure 8

### 4.1.2 (Chapter 02) Study of the vulnerabilities associated to tumor heterogeneity and tumor progression in prostate cancer by using Genome-scale metabolic network models and data-driven approaches

In this chapter we developed and applied a comparative Genome-scale metabolic model reconstruction analysis. This analysis was performed on a cellular model consisting of two clonal subpopulation of the prostate cancer cell line PC-3 in order to account for intra-tumoral heterogeneity, We also determined a metabolic gene signature associated to tumor progression and metastasis by performing a differential expression analysis on these subpopulations. This chapter summarizes the results of these two analysis included in two different articles that has been recently submitted to international scientific journals

**Chapter 2a.** Network model-driven discovery of mechanisms underlying calcitriol and fatty acid oxidation as therapeutic targets in metastatic prostate cancer

We developed and implemented an approach based on CBM in Large-scale metabolic networks assessing intratumoral heterogeneity in order to discover new drug targets. More specifically, we focused on the study of the metabolic activity profiles of two clonal sub-populations isolated from an established prostate cancer cell line (PC-3): PC-3/S (like non-Cancer stem cells (CSC) c) and PC-3/M (like CSC). These sub-populations coexist within the same tumor and represents an excellent cellular model to study how the intra-tumoral heterogeneity provides advantages to the tumor in terms of metastatic capability and drug resistance. We characterized the metabolic profile of both subpopulations by performing a comparative GSMM Analysis. To this aim we inferred the activity state of the metabolic network of PC-3/M and PC-3/S cells by integrating their transcriptomic data in the most recent reconstruction of human metabolism (Recon2) [59] using one of the most widely tested CBM [81]. Brief, this method seeks the maximal similarity between the activity state of the metabolic

network and the gene expression profile. This procedure allowed us to infer the metabolic activity state profile of each PC-3 subpopulation. These results were validated by comparing the predicted activity state of the exchange reactions with available measurements of uptake /secretion rates [145]. In table 3 are summarized those pathways with more significant differences

| Reactions active in PC-3/M & inactive in PC-3/S | | Reactions active in PC-3/S & inactive in PC-3/M | |
|---|---|---|---|
| **Pathway** | **nº Active reactions** | **Pathway** | **nº Active reactions** |
| Fatty acid oxidation | 30 | Cholesterol metabolism | 5 |
| Fatty acid synthesis | 4 | Citric acid cycle | 1 |
| Miscellaneous | 2 | Eicosanoid metabolism | 6 |
| N-glycan degradation | 6 | Fatty acid synthesis | 2 |
| N-glycan synthesis | 1 | Nucleotide interconversion | 3 |
| Pyrimidine catabolism | 1 | Sphingolipid metabolism | 3 |
| Sphingolipid metabolism | 1 | Transport, extracellular | 1 |
| Transport, extracellular | 1 | Transport, peroxisomal | 1 |
| Transport, lysosomal | 1 | Tryptophan metabolism | 1 |
| Transport, mitochondrial | 7 | Unassigned | 2 |
| | | Vitamin D metabolism | 4 |

Table 3: The table on the left represents those reactions active in PC-3/M and inactive in PC-3/S. The table on the right represents those reactions active in PC-3/S and inactive in PC-3/M. In both tables, the pathways are in the first column and the number of reactions that are active in one of the subpopulations are in the second column

We observed that the activity of fatty acid oxidation predicted by the model is higher in PC-3/M than in PC-3/S cells. The oxidation of fatty acids in the mitochondria produces NADH, FADH2 and AcetilCoA that directly fuels TCA cycle and the electron transport chain to produce energy in the form of ATP molecules. This strategy can be used by PC-3/M cells to sustain their higher proliferation rate.

Most of the reactions differentially activated in this pathway involve Carnitine palmitoil transferase 1 (CPT1). This mitochondrial membrane protein actively transports long-chain fatty acids (LCFA) from cytosol into the mitochondria. The substrates of CPT1 in the cytosol, exclusive of PC-3/M, are cervonyl coenzyme A, eicosatetranoyl coenzyme A, arachidyl coenzyme A, trans-2-octadecenoyl-CoA(4-), palmitate, Malonyl-CoA, linoelaidyl coenzyme A and vaccenyl coenzyme A.

Interestingly, it has been reported that five of these eight LCFA have antiproliferative effects [146-150].

The model predicts a higher activity of eicosanoid metabolism in PC-3/S cells and more specifically in those reactions involved in Arachidoic acid metabolism (AA). This molecule is metabolized from Arachidoil CoA, a long-chain fatty acid (LCFA). In addition, AA is the precursor of the Eicosanoid metabolism that produces some prostaglandines with antiproliferative effects like prostaglandine J2 (PJ2) [151]. AA is also the precursor of the cannabinoid 2-Arachidonoylglycerol (2-AG) that activates the cannabinoid receptor type 1 (CB1) [152] that, in turns, reduces tumor progression in prostate cancer cells [153]. This fact could explain why PC-3/S proliferates more slowly than PC-3/M cells. These metabolic processes are predicted to be more active in PC-3/S and the metabolism of 2-AG exclusively in this cell line.

Vitamin D3 metabolism is also predicted to be exclusively active in PC-3/S subpopulation. This molecule controls the proliferation in prostate cells [154] and has antiproliferative effects on a number of cancer cell types, being PC3 cells one of the few cell lines insensitive to this drug [155]. This prediction is consistent with the lower proliferation rate observed in PC-3/S cells.

Finally the predictions related with the differential activity of CPT1, calcitriol metabolism and AA metabolism were validated using different experimental techniques that supported the model predictions.

Chapter 2b. **Contrasting glycolytic flux and mitochondrial metabolism profiles define prostate epithelial cancer stem cells vs. non-cancer stem cells**

We aimed to determine the potential significance in cancer progression of metabolic pathways differentially active between PC-3 subpopulations.

Here we used the transcriptomic information [124] of PC-3 subpopulations to extract a set of metabolic genes significantly enriched in PC-3M (cancer stem cell -CSC-phenotype) cells relative to PC-3S (non-CSC) cells by using Rank Product method

[21]. This gene set includes enzymes involved in the metabolism of serine and glycine (GLDC, CBS), branched-chain amino acids (BCAT1), glutamate and proline (ASS1, AGMAT), which support the metabolic flexibility accompanying the CSC phenotype. It also includes genes that participate in the synthesis of purine nucleotides (GUCY1A3, PDE3B and ADCY7) as part of reactions that converge on substrates that are actively metabolized in PC-3M cells, such as glycine and one-carbon units from tetrahydrofolate.

By applying Gene Set Enrichment Analysis [28], we found that the PC-3M-enriched metabolic gene set is significantly enriched along malignant progression in prostate cancer (Figure 9) and other 11 types of human tumors, including ovarian, bladder, adrenocortical, head and neck, stomach or rectal carcinoma, melanoma and mesothelioma (Figure 9).

These observations support the proposal that the metabolic pathways and reactions that we have found differentially active in our CSC vs. non-CSC cell model are relevant for tumor progression in human tumors.

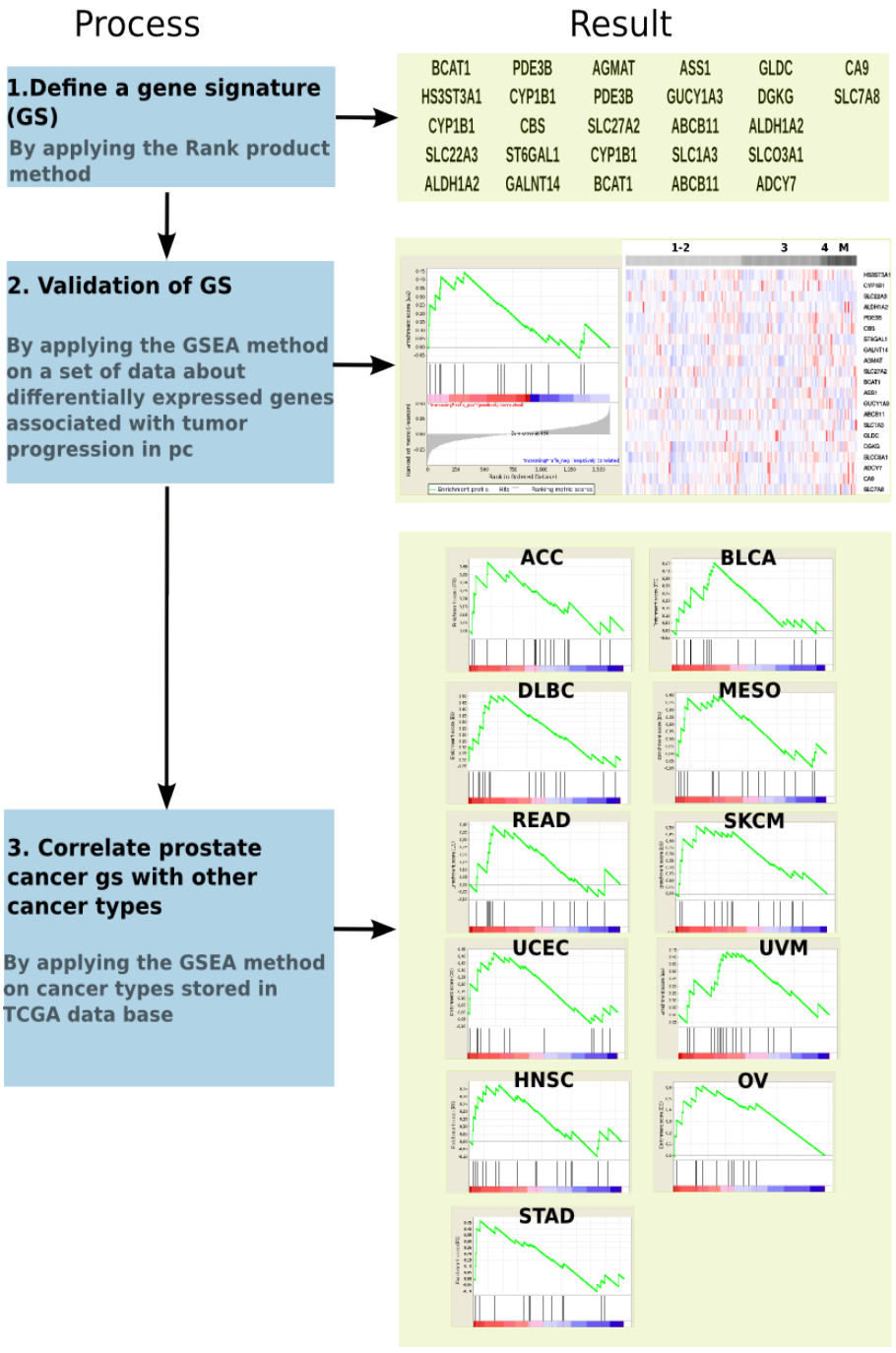

Process
Result
1.Define a gene signature (GS)
By applying the Rank product method
BCAT1 PDE3B AGMAT ASS1 GLDC CA9
HS3ST3A1 CYP1B1 PDE3B GUCY1A3 DGKG SLC7A8
CYP1B1 CBS SLC27A2 ABCB11 ALDH1A2
SLC22A3 ST6GAL1 CYP1B1 SLC1A3 SLCO3A1
ALDH1A2 GALNT14 BCAT1 ABCB11 ADCY7
2. Validation of GS
By applying the GSEA method on a set of data about differentially expressed genes associated with tumor progression in pc
1-2
3
4
M
3. Correlate prostate cancer gs with other cancer types
By applying the GSEA method on cancer types stored in TCGA data base
ACC
BLCA
DLBC
MESO
READ
SKCM
UCEC
UVM
HNSC
OV
STAD

Figure 9: Gene signature associated to tumor progression in prostate cancer. The scheme describes the procedure for defining and validating a gene signature associated with tumor progression. 1. The gene signature was defined by applying the Rank product method [21]. 2. Gene Set Enrichment Analysis (GSEA) [28] was applied to analyse datasets looking for gene sets that are regulated together. These datasets were computed RMA expression values. A normalized enrichment score (NES) was used to rank the datasets, showing a high concordance of our gene signature with gene sets associated with tumour progression in pc. The box in the right shows the enrichment score returned and the heat-map. Finally, by applying again GSEA, our gene signature correlated with gene expression of other eleven cancer types from the TCGA database: ACC (Adrenocortical carcinoma); BLCA (Bladder Urothelial Carcinoma); HNSC (Head and Neck squamous cell carcinoma); DLBC (Lymphoid Neoplasm Diffuse Large B-cell Lymphoma); MESO (Mesothelioma); OV (Ovarian serous cystadenocarcinoma); READ (Rectum adenocarcinoma); SKCM (Skin Cutaneous Melanoma); STAD (Stomach adenocarcinoma); UCEC (Uterine Corpus Endometrial Carcinoma); UVM (Uveal Melanoma).

### 4.1.3 (Chapter 03) Study of gene regulatory mechanisms underlying the abnormal metabolic adaptation to training in COPD by combining probabilistic and mechanistic approaches to integrate multilevel omic data into a discrete model-based analysis

In this chapter we developed a novel computational method combining probabilistic and mechanistic approaches to integrate multi-level omic data into a discrete model-based analysis. Part of the data that was integrated was previously obtained by data-mining and manual curation of relevant publications. Here, we analyzed the mechanism underlying the crosstalk between metabolism and gene regulation, using as case of concept the study of the abnormal adaptation to training in COPD patients. This chapter summarizes the results of these two analysis. The data-mining analysis is included in an article that is already published, the development of the computational tool generated an article that was recently submitted to an international scientific journal.

Chapter 3a. **Knowledge management for systems biology a general and visually driven framework applied to translational medicine**

In this work we implemented the literature curation for COPD specific enzyme and compound data. It included, 54 inflammation and tissue specific pathways and 122 COPD and exercise specific metabolite and enzyme concentrations and activities. The pathway curation followed a standard text-mining supported process as described for example in [156] while the enzyme concentration and activity curation was fully manual due to the small set of available relevant publications. This information was integrated into BioXM that integrates with a large number of public and private databases. This together with the fact that has been developed around the concept of object-oriented semantic integration makes this tool extremely potent to extract meaningful information that was used in later works.

Chapter 3b. **Knowledge management for systems biology a general and visually driven framework applied to translational medicine**

Here we focused on the anomalous muscle adaptation to training of COPD patients, more specifically we investigated the different effects of ROS on the GRN that governs the energy metabolism in COPD patients compared with healthy people.

For this aim we developed a computational approach to i) first, reconstruct a mitochondria-specific GRN using transcriptomic and bibliography-based data and ii) integrate the reconstructed GRN and multi-level omic data into a discrete model based on R. Thomas formalism [103].

We used trancriptomic data extracted from vastus lateralis biopsies of three groups: COPD patients with normal BMI, COPD patients with low BMA and control group, before and after an 8-weeks training program (six groups in total) (Fig 4.1).

In order to design a COPD-specific GRN we first choose the nodes of the network by determining which genes were more differentially expressed between conditions

(Figure 10.2). Next we used IPA Software [157] and DroID database [158] to define the interaction between the nodes(Fig 4.3 and 4).

Finally the sign and the direction of the interactions were manually curated using literature based data (Figure 10.5).

The resulting GRN reconstruction is shown in the figure 10.5. Next we integrated the information embedded in the GRN previously reconstructed into a discrete model. For this aim we used SysBiOX framework [104]. This software is implemented in ASP (Answer Set Programming) and allows to search for solutions of a system of constraints bearing on a network described in the multilevel logical formalism developed by R. Thomas et al. [103].

Additionally, in order to reduce the number of models that satisfy the constraints imposed by the topology of the network we integrate constraints from different sources [7,160,161] using different probabilistic methods such as rank product or correlation analysis [21,33] (Figure 10.6).

Once the GRN was integrated into a discrete model together with the constraints previously mentioned, we analyzed the system in order to infer the activity state of the nodes of the network in the six different conditions. For this aim we impose an objective function that minimize the discrete value of the nodes while satisfying the constraints imposed by the topology of the network and the constraints defined in the probabilistic analysis.

From this analysis, we obtained 36 possible models that were consistent with the constraints and the network topology. Finally, these models were analyzed in order to determine how the discrete value of each gene evolved in response to the training program.

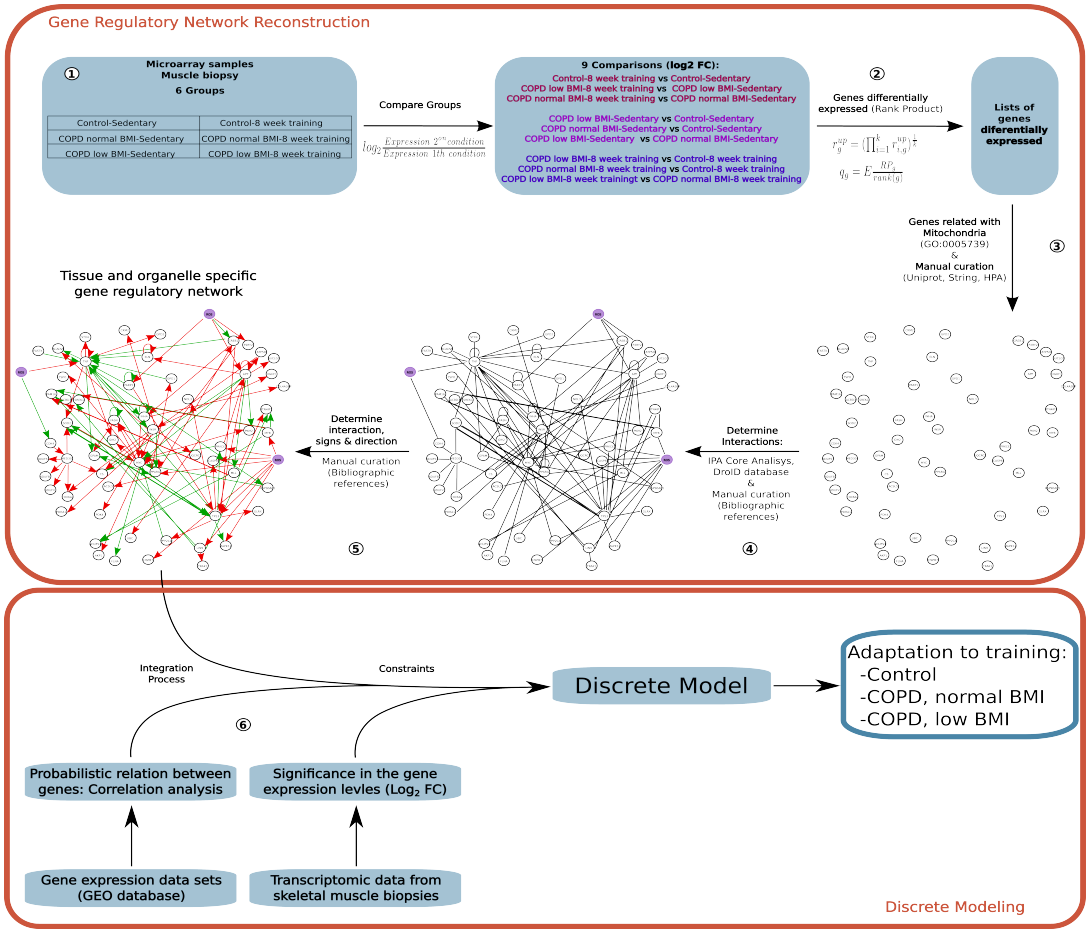



Figure 10 Here is represented i) the process of GRN reconstructions based on gene expression data, ii) the integration of constraints based on probabilistic approaches and iii) build a discrete model by integrating the GRN reconstruction and the constraints

According to model predictions COPD patients have an abnormal adaptation to training compared with control group. Table 4 summarizes the changes predicted by the model in the three groups of study in response to the training program:

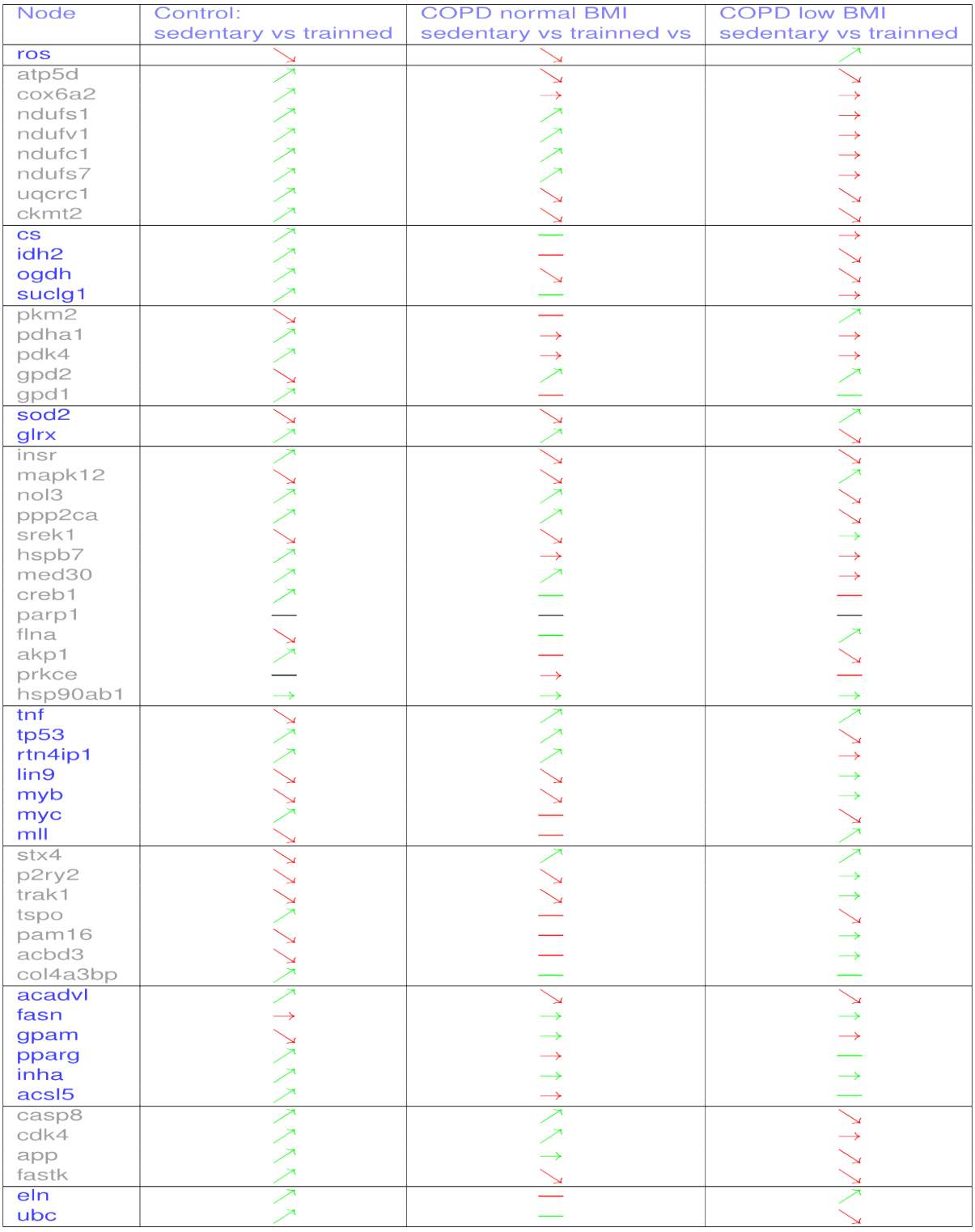

Node
Control: sedentary vs trainned
COPD normal BMI sedentary vs trainned vs
COPD low BMI sedentary vs trainned
ros
atp5d
cox6a2
ndufs1
ndufv1
ndufc1
ndufs7
uqcrc1
ckmt2
cs
idh2
ogdh
suclg1
pkm2
pdha1
pdk4
gpd2
gpd1
sod2
glrx
insr
mapk12
nol3
ppp2ca
srek1
hspb7
med30
creb1
parp1
flna
akp1
prkce
hsp90ab1
tnf
tp53
rtn4ip1
lin9
myb
myc
mll
stx4
p2ry2
trak1
tspo
pam16
acbd3
col4a3bp
acadvl
fasn
gpam
pparg
inha
acsl5
casp8
cdk4
app
fastk
eln
ubc

Table 4: In the first column are the gene names, from the second to fourth column are represented the adaptation to the training program of the three different groups using the following code: the red arrows pointing down describe a reduction in the discrete value of the node in response to training predicted by all the models, the green arrows pointing up represent an increment in the discrete value of the node in response to training predicted by all the models, red right arrows represent nodes with no adaptation to training remaining the value of the node equal to 0 predicted by all the models, green right arrows represent nodes with no adaptation to training remaining the node in its maximum discrete value predicted by all the models, green lines represent an increment in the discrete value of the node in response to training predicted by between the 99 and the 70 % of the models, red lines represent a decrease in the discrete value of the node in response to training predicted by between the 99 and the 70 % of the models, black lines represent nodes with no consistent prediction

In table 4 is represented the different adaptation to training in the three groups of study. As opposite to healthy group, the model predicted a down-regulation in the activity of EC, TCA cycle and creatine kinase as response to an 8-weeks training program in COPD patients with low BMI. Interestingly, for COPD patients with normal BMI, the model predicts an up-regulation of complex I (as it occurs in healthy patients) and a slight increase in Citrate synthase and Succinyl-CoA ligase. However, the model also predicts that complexes III and V, other key enzymes of TCA cycle, and creatine kinase activities decrease after exercise which suggests a defective energetic metabolism adaptation to training COPD patients with normal BMI. The adaptation to training in healthy group provides a higher capability to consume O2 and produce energy. The model also predicts that healthy group up-regulates cdk4 and down-regulate tnf in response to exercise whereas COPD normal BMI are able also to up-regulate cdk4 but by contrary increase tnf in response to the exercise. COPD with low BMI are not able to respond at all to exercise with respect cdk4 that remains low and also present the abnormal increase of tnf displayed by normal BMI COPD patients in response to exercise. This facts indicate defective muscle remodeling pathways in COPD patients. The fact that both cdk4 and tnf presents opposite patterns of regulation in response to exercise in low BMI COPD patients suggests that muscle remodeling is more defective

in these patients. Interestingly Glycerol-3-phosphate dehydrogenase cytosolic isoform is over-expressed in control group and under-expressed in COPD groups, we can observe the opposite pattern in the mitochondrial isoform. The mitochondrial isoform catalyze a reaction where NADH is produced, probably mitochondrial Glycerol-3-phosphate dehydrogenase mitochondrial isoform is over-expressed in COPD groups after training to partially compensate the reduction in E.C. activity. The insulin receptor factor (insr) is predicted to have a different response to training in COPD patients compared with healthy group that up-regulates its expression after training. The insulin receptor factor triggers a number of signaling events, which one of the most important effects is to produce a translocation of the glucose transporter SLC2A4/GLUT4 from cytoplasmic vesicles to the cell membrane, it facilitates glucose transport increasing the energy metabolism. Then, the down-regulation of insr predicted in both COPD groups would reduce the capability of the muscle to use glucose as source of energy. The model also predicted differences in acadvl between healthy group and both COPD groups. The acadvl (Very long-chain specific acyl-CoA dehydrogenase) is a key player in beta-oxidation that produces Acetyl-CoA that fuels TCA Cycle and EC to finally produce energy in the form of ATP molecules. This enzyme is up-regulated after training in healthy group and down-regulated in both COPD groups. Then, unlike the control group, COPD patient reduce the expression of this key enzyme in the beta-oxidation that in turns deprives the energy metabolism. We also observe differences in the nodes related with glycolysis. Thus, the model predicts that pdha1 is up-regulated in control group, the same protein is inactive in both COPD groups before and after the training. This enzyme plays a key role in glycolysis and TCA cycle metabolism catalyzing the reaction that decarboxylates pyruvate to AcoA fueling the TCA cycle. Finally, the model was able to predict a reduction of ROS in control group after training, this prediction is consistent with previously reported observations [161], it also predicted a reduction of ROS levels in COPD group with normal BMI, on the other hand the levels of ROS in COPD group with low BMI was

predicted to increase after training.

Finally model predictions were validated using two different strategies. First the model was asked for the state of one specific gene when the constraints related with that gene were removed. In all the analysis the activity state of the asked gene was the same than in the original analysis that considered all the constraints, validating our model predictions. Secondly, we used experiments where one of the genes of the network was experimentally knocked-out [GEO Database Reference Series: GSE7162, GSE25072, GSE2236 and GSE36041], we integrated the corresponding gene expression in the form of constraints into the model excepting those constraints related with the knocked-down gene. In all the cases the model was able to predict the knocked-down gene, which supports our model predictions.

Gathering the model predictions we have an overview in which healthy people adapts to training by increasing the activity of those pathways involved in energy production, by contrary, this machinery has a defective adaptation to training in COPD patients with normal BMI and is more severe in COPD group with low BMI. The abnormal adaptation also affects genes and proteins related with proteolysis, it could explain the lose of muscle mass observed in the most severe COPD patients.

Here we have developed different strategies that allowed us to reconstruct a highly curated COPD-specific GRN and the integration of the mechanism described by this network with information extracted from probabilistic approaches. It has permitted us to unlock important mechanisms underlying the abnormal adaptation to training in COPD patients and the role of ROS in this disease. All this shows the potential of this method and opens new possibilities in the study of complex diseases such as COPD or other chronic diseases providing new therapeutic targets or better tools to design more efficient strategies to mitigate the effects of the diseases.

## 4.2 Global discussion

In the last decades, the advent of high-throughput technologies has transformed molecular biology into a data-rich discipline. It has promoted the development of a large variety of computational tools to unveil the molecular mechanism underling the different cellular processes. In the scope of systems medicine, are of especial interest those methods focused on the study of metabolism due to its close relationship with phenotype and consequently with several biological processes. However, existing limitations and pitfalls in these computational approaches can compromise the reliability of the results and in some cases can make the problem unmanageable. Thus, this thesis is aimed to develop and apply new computational methods in order to overcome pitfalls and limitations of some of the most relevant computational approaches used in the study of metabolism at different levels these are briefly: i) computational approaches based on non-stationary 13C-FBA does not account for metabolic topology of the metabolic pathways ii) metabolic drug-target discovery based Genome-scale metabolic models (GSMM) in cancer research is limited to homogeneous populations, being tumors heterogeneous ecosystems iii) the study of mechanism underlying the behavior of metabolic network from the point of view of crosstalk between metabolic and GRN is limited in size and complexity due the limitations of the current computational approaches.

In chapter 1 we developed a new computational method based on non-stationary $^{13}$C-FBA to study the metabolic channeling in glycogen synthesis occurring in hepatocite. More specifically we designed and test new functionalities for the software Isodyn in order to perform the discrimination of various models reflecting different topologies of intracellular metabolism and choose the model most consistent with $^{13}$C tracing experiments.

In order to test this computational tool, we evaluated two different models reflecting two different schemes, one defining a well-mixed pool of hexose 6-phosphate and other simulating the effect of channeling as two separated pools of hexose 6-phosphate

which did not freely mix by diffusion. In order to evaluate the reliability of model predictions we used $\chi^2$ criterion. However, since we were evaluating two models with different degrees of freedom, we also used the parameter Q, this statistic gives the probability that a model with F degrees of freedom is correct and $\chi^2$ by chance could exceed a dtermined value. Despite the analyzed dataset was restricted by extracellular metabolites and glycogen, the model that did not account for the channeling failed to fit experimental data, and application of $\chi^2$ and Q criterion clearly rejected this model. On the other hand, this criterion indicated acceptability of the model that assumes channeling in glycogen synthesis. According to this assumption the labeling of glycogen glucose is not so strictly linked with the labeling of hexose phosphates of glycolytic pathways.

In this situation the step of model validation by the consistency with isotopologues of glycogen glucose fragments was important for the verifications of simulated positions of $^{13}C$ isotopes. Simultaneous fitting of isotopologues of whole glycogen glucose and fragments containing carbons 1-4 and 3-6, in fact, specifies fractions of $^{13}C$ isotopes in positions 1-2, 3-4, and 5-6.

The channeling, revealed in the model, requires interaction of at least five enzymes: hexokinase, phosphoglucomutase, uridiltransferase, glycogen synthase and fructose bisphosphatase. The interaction between these enzymes can be understood as a result of co-localization and limitation of diffusion of substrates/products, confined between the bodies of co-localized enzymes. The indications that diffusion in biological structures can be essentially limited appear in various studies of cell physiology. The diffusion of cAMP is possibly extremely restricted in proximity of cyclic nucleotide-gated channels [161]. There are various indications that the diffusion of ATP is restricted in the proximity of $K_{ATP}$ channels [49] and in myofibrils [162]. The findings described here are in line with the other indications of limited diffusion in various biological structures.

Since the fundamental work of Atkinson [1977], many indications of channeling in

glycolysis were found [163], which are in line with the presented study. In general, the colocalization of several enzymes resulting in the formation of spaces restricted for diffusion from outside, probably, could be a reason of apparent diffusion limitation for small metabolites.

The presented work shows that the analysis of the functional structure of metabolic networks can be studied based on isotopologue spectra. The approach proposed here to determine metabolite compartmentalization from $^{13}C$ distribution in metabolites does not require specific experiments. Instead, it requires a specific analysis related with the implementation of various schemes and application of model discrimination analysis to define the compatibility of the schemes with the data. This approach can be extrapolated to larger datasets, being able to reveal new information about the network topology. It opens a perspective to examine the compartmentalization and metabolic flux profile in various cells under physiological and patho-physiological conditions.

In chapter 2 we focused on the development of computational approaches that allowed us to study of the metabolic profile of two clonal sub-populations isolated from an established prostate cancer cell line (PC-3): PC-3/S (like epithelial-mesenchymal transition cell line) and PC-3/M (like tumor-initiating cells cell line). These sub-populations coexist within the same tumor and represents an excellent cellular model to study how the intra-tumoral heterogeneity provides advantages to the tumor in terms of metastatic capability and drug resistance.

Here we have used a CBM that allowed us to determine the metabolic activity state of PC-3 subpopulations by integrating transcriptomic data into a GSMM reconstruction. This method seeks to maximize the similarity between the gene expression pattern and the activity state of the metabolic network. Consequently it allow us to take our study further than a classical gene expression analysis. It's worth noting that some of the main metabolic differences that were predicted between PC-3/M and PC-3/S did not fit the gene expression patterns. Then, for example, the gene expression associated to the reaction that catalyzes the hidroxilation of Vitamin D3 (part of the Vitamin D3

metabolism) is higher in PC-3/M cells, however our computational analysis predicted that this reaction was active only PC-3/S subpopulation. As is explained in Results section this prediction was experimentally validated and demonstrates the importance of consider possible post-transcriptional effects in order to capture the metabolic processes underlying certain phenotypes.

By integrating gene expression data into a GSMM reconstruction we predicted differences in the metabolic activity state of the two PC-3 subpopulations. We observed that the activity of fatty acid oxidation predicted by the model is higher in PC-3/M than in PC-3/S cells. More specifically most of the reactions differentially activated in this pathway involve Carnitine palmitoil transferase 1 (CPT1). This mitochondrial membrane protein actively transports long-chain fatty acids (LCFA) from cytosol into the mitochondria and plays an inportant role in the mitochondrial beta-oxidation and consequently in energy metabolism. The model predicts a higher activity of eicosanoid metabolism in PC-3/S cells and more specifically in those reactions involved in Arachidoic acid metabolism (AA). This molecule is metabolized from Arachidoil CoA, a long-chain fatty acid (LCFA) and is the precursor of a variety of molecules with antiproliferative effects [151-153]. Vitamin D3 metabolism was also predicted to be exclusively active in PC-3/S subpopulation. This molecule controls the proliferation in prostate cells [154] and has antiproliferative effects on a number of cancer cell types, being PC3 cells one of the few cell lines insensitive to this drug [155]. These predictions were validated experimentally. These predictions were experimentally validated supporting the model predictions.

Based on our findings we have proposed a mechanism of metabolic reprogramming where, the higher CPT1 and β-oxidation activity in PC-3/M can accomplish two roles. First, and probably the most evident, is to maintain the energetic requirements imposed by PC-3/M high growth rate and secondly to eliminate those LCFA with antiproliferative effects. Most of these LCFA are precursors of prostaglandines that are metabolized in the eicosanoid metabolism. Then, for example, AA is metabolized from

one of these LCFA, the arachidyl coenzyme A and as we mentioned previously, AA is a precursor of 2-AG with antiproliferative effects [152,153]. In addition, 2-AG is a precursor of Prostaglandine J (PGJ) that inhibits androgen receptor signaling which over-expression is critical for the growth and progression of prostate cancer [164]. Our computational analysis predicted that 2-AG was metabolized only in PC-3/S and experimental measurements demonstrated that the levels of its precursor (AA) and one of its products (PGJ) were higher in PC-3/S. On the other hand we demonstrated that, unlike PC-3/M, PC-3/S cells have the capability to uptake Calcitriol into the nucleus and consequently are sensitive the antiproliferative effects of this compound. Calcitriol also regulates the homeostasis of calcium increasing its intracellular concentration [165] that is necessary in the metabolization of some of the previously prostaglandines with antiproliferative effects such as PGJ or PGA1 [166]. These evidences drive us to prose a mechanism in which the metabolic role of LCFA changes in the two PC-3 subpopulations, which can explain their different phenotypes. The figure 4 illustrate this mechanism:

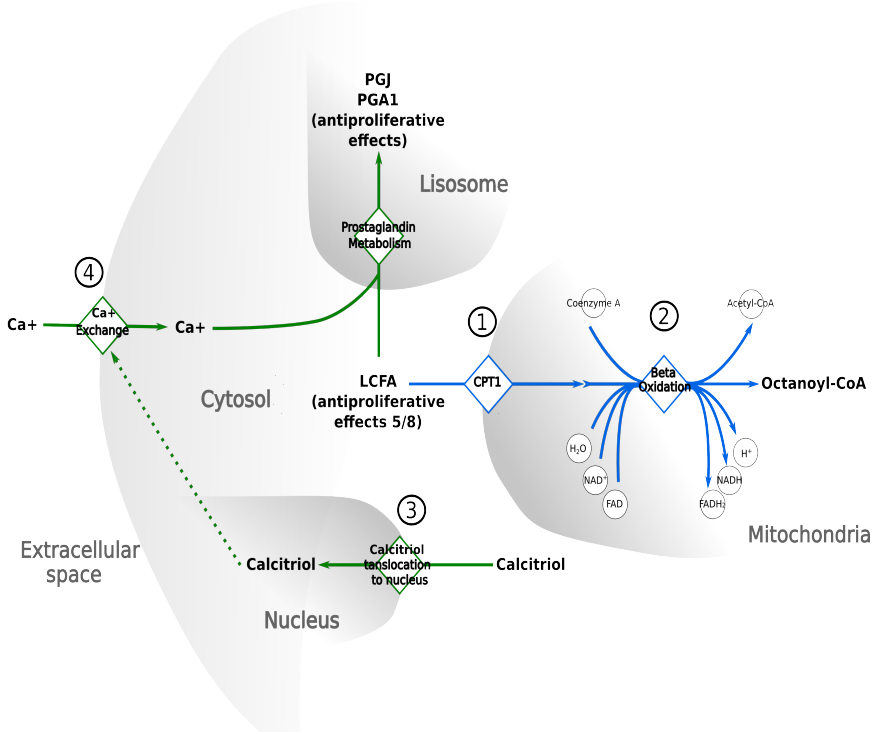


Figure 11: Proposed mechanism of metabolic reprogramming between PC-3/M and PC-3/S. Green arrows represents pathways or groups of metabolic reactions mainly active in PC-3/S subpopulations while the ones highlighted in blue are those mainly active in PC-3/M.

Thus, the over-activity of CPT1 in PC-3/M cells increases the entry of LCFA into the mitochondria (1) to be oxidized (2) and produce energy to sustain the highly proliferative activity of PC-3/M and eliminate the LCFA reducing their antiproliferative effects. On the other hand CPT1 activity in PC-3/S cells is lower and consequently LCFA are accumulated in the cytosol. Additionally only this subpopulation can translocate Calcitriol into the nucleus (3) that increases the intracellular Ca2+ (4). Ca2+ is necessary in some steps of the prostaglandine metabolization (4). Thus, the higher Ca2+ concentration and the increased availability of cytosolic LCFA in PC-3/S (due a lower activity of CPT1), would increase the activity of prostaglandine metabolic pathway and more specifically the metabolization of those molecules with antimetastatic effects. Summing up, the different metabolic reprogramming involving LCFA utilization drives PC-3/M cells to be more metastatic and highly proliferative than PC-3/S subpopulation.

We demonstrated that the resistance of PC-3 cells towards Calcitriol was provided by the resistance of PC-3/M subpopulation to this compound and the lower efficacy of Etomoxir reducing cell growth on PC-3 compared with other prostate cell lines was due the low metabolic dependency of PC-3/S subpopulation to CPT1.

In addition, we defined a metabolic gene signature associated to the dual model of prostate cancer used in this study (PC-3/S (like non-CSC) and PC-3/M (like CSC)). We found that it was correlated with tumor progression in prostate cancer and in other cancer types. This analysis demonstrates that the mechanism observed in PC-3/M and PC-3/S can be extrapolated to the study of tumor progression in prostate cancer and in other types of cancer. Furthermore, these evidences suggest that the mechanism of metabolic reprogramming between PC-3/M and PC-3/S exposed above, could be a new feature of tumor progression and metastasis in prostate cancer and in other tumor

types. This new understanding of the metabolic mechanism underlying tumor progression open new perspectives in developing new anti-metastatic therapies that takes advantage from weakness associated to tumor heterogeneity.

In chapter 3 we have developed a novel computational method that integrates mechanistic and probabilistic approaches to integrate multiple layers of biological data into a discrete model-based analysis. As a case of concept, have investigated the mechanisms underlying the abnormal adaptation to training associated to COPD disease. More specifically, it has been focused on determine how ROS affects the gene regulatory mechanisms associated to the energetic metabolism before and after 8-weeks training program. It has been carried out in two steps, first we reconstruct a tissue, organelle and condition specific GRN that combined the effects of training and of COPD on the mechanisms that regulate gene expression. It has been achieved by integrating transcriptomic and literature-based data. This reconstruction integrates the most relevant mechanism that regulates gene expression in response to training in healthy people and COPD patients. Next we integrated the GRN reconstruction into a discrete model offering a platform to integrate omic data from different sources as well as simulate the evolution of the regulatory mechanism due to perturbations produced by the levels of ROS in response to training. Next we performed a correlation analysis on an extensive number of gene expression data-sets. We developed an strategy that allowed us to integrate the information extracted from the probabilistic analysis into the mechanism described in our GRN reconstruction in the form of constraints. Once we integrated the GRN reconstruction and the constraints from the correlation analysis into a discrete model, we determined the activity state of the GRN of the three groups of study in response to training.

Gathering the model predictions we have an overview In which healthy people adapts to training by increasing the activity of those pathways involved in energy production, by contrary, this machinery has a defective adaptation to training in COPD patients with normal BMI  and is more severe in COPD group with low BMI. The abnormal

adaptation also affects genes and proteins related with proteolisis, it could explain the lose of muscle mass observed in the most severe COPD patients.
Summing up, in this chapter we have presented the development of different strategies that allowed us to reconstruct a highly curated COPD-specific GRN and the integration of the mechanism described by this network with information extracted from probabilistic approaches. It has permitted us unlock important mechanisms underlying the abnormal adaptation to training in COPD patients and the role of ROS in this disease. All this shows the potential of this method and opens an avenue of possible strategies to study complexity of COPD and other chronic diseases providing new therapeutic targets or better tools to design more efficient strategies to mitigate the effects of the disease.

## 4.2 Conclusions

1. The development of a new computational function for the software Isodyn allow to use the analysis of isotopomers to determine the functional structure of the metabolic network, more specifically the structure associated to metabolite channeling. This approach is a conceptually novel, and the work presented in this thesis illustrates the potential of this application.
2. The analysis of isotopomer distribution of metabolites produced by cells incubated with $^{13}C$ labeled glucose allows to determine the existence of channeling in the metabolism of glycogen in hepatocytes. It has been determined by using the $\chi$ 2 and Q value as criteria to evaluate and compare models with and without channeling
3. The comparative genome-scale metabolic reconstruction analysis developed in this thesis and applied on PC-3/M and PC-3/S cells allow to determine

important metabolic differences between subpopulations that are supported by the literature and fits with their corresponding phenotypes and experimental observations

4. CPT1 activity is higher in PC-3/M resulting in a more active beta-oxidation in this subpopulation. Those substrates of CPT1, exclusive of PC-3/M subpopulation are LCFA with antiproliferative effects. It suggest that the over-activity of CPT1 in PC-3/M has a double role: produce energy by oxidizing LCFA into the mitochondria and reduce their antiproliferative effects. This findings fits with the more proliferative phenotype of PC-3/M. On the other hand Eicosanoid metabolism and Calcitriol uptake into the nucleus are more active in PC-3/S. Both processes result in antiproliferative effects which fits with the less proliferative phenotype of PC-3/S.
5. PC-3M and PC-3S subpopulations present different sensitivity to treatment with calcitriol and etomoxir. While PC3M cells are sensitive only to the inhibition of CPT1 by etomoxir, calcitriol reduces the proliferation mainly in PC-3S cells but not in PC-3M.
6. The set of genes significantly up-regulated in PC-3M compared with PC-3S correlates with prostate cancer progression and metastasis and with other 11 cancer type.
7. The integration of a GRN describing the mechanisms of the interactions between nodes with the constraints obtained from probabilistic approaches into a discrete model-based analysis has provided a detailed estimation of how the GRN and the associated metabolism are adapted to training in both COPD groups compared with control group.
8. The analysis of the GRN shows that healthy people adapts to training by increasing the activity of those pathways involved in energy production, by contrary, this machinery has a defective adaptation to training in COPD patients with normal BMI and is more severe in COPD group with low BMI.

The abnormal adaptation also affects genes and proteins related with proteolisis which explains the lose of muscle mass observed in the most severe COPD patients.

# 05. References

1.Rajcevic U, Knol J, Piersma S et al. Colorectal cancer derived organotypic spheroids maintain essential tissue characteristics but adapt their metabolism in culture. Proteome Sci. 12(1), 39 (2014)

2.Kim HU, Kim SY, Jeong H et al. Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. Mol. Syst. Biol. 7(1), 460 (2011.)

3.Matsuda F, Furusawa C, Kondo T, Ishii J, Shimizu H, Kondo A. Engineering strategy of yeast metabolism for higher alcohol production. Microb. Cell Fact. 10(1), 70 (2011).

4.Ram PT, Mendelsohn J, Mills GB. Bioinformatics and systems biology. Mol Oncol. 2012 Apr;6(2):147-54.

5.Zhang W, Li F, Nie L. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. Microbiology. 2010 Feb;156(Pt 2):287-301.

6.Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995 Oct;270(5235):467-70.

7.Smyth GK, Speed T. Normalization of cDNA microarray data. Methods. 2003 Dec;31(4):265-73.

8.Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013 Jan;41(Database issue):D991-5.

9.Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M, Mani R,

Rayner T, Sharma A, William E, Sarkans U, Brazma A. ArrayExpress--a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 2007 Jan;35(Database issue):D747-50

10.ter Kuile BH, Westerhoff HV. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway.FEBS Lett. 2001 Jul 6;500(3):169-71.

11.van den Berg RA1, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006 Jun 8;7:142.

12.Cascante M, Marin S. Metabolomics and fluxomics approaches. Essays Biochem. 2008;45:67-81.

13.Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. Curr Opin Microbiol 7, 296–307.

14.Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res. 2013 Jan;41(Database issue):D801-7.

15.Sauer U. Metabolic networks in motion: 13C-based flux analysis.Mol Syst Biol. 2006;2:62.

16.Gregory N. Stephanopoulos, Aristos A. Aristidou and Jens Nielsen. Metabolic Engineering. Academyc press; 1998

17.Albert-László Barabási & Zoltán N. Oltvai. Network biology: understanding the cell's functional organization. Nature Reviews Genetics 2004 February 5, 101-113

18.Auffray C, Chen Z, Hood L. Systems medicine: the future of medical

genomics and healthcare. Genome Med. 2009 Jan 20;1(1):2.

19.Palsson B, Zengler K: The challenges of integrating multi-omic data sets. Nat. Chem. Biol. 6(11), 787-789 (2010).

20.Aerts JM, Haddad WM, An G, Vodovotz Y. From data patterns to mechanistic models in acute critical illness. J Crit Care. 2014 Aug;29(4):604-10.

21.Breitling R, Armengaud P, Amtmann A, Herzyk P.: Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 2004 Aug 27;573(1-3):83-92.

22.S. Smit, M.J. van Breemen, H.C.J. Hoefsloot, A.K. Smilde, J.M.F.G. Aerts, C.G. de Koster. Assessing the statistical validity of proteomics based biomarkers. Anal. Chim. Acta, 592 (2007), pp. 210–217

23.A. Fukushima, M. Kusano, H. Redestig, M. Arita, K. Saito. Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. BMC Syst. Biol., 5 (2011), p. 1

24.Joshua Millstein and Dmitri Volfson. Computationally efficient permutation-based confidence interval estimation for tail-area FDR. Front Genet. 2013; 4: 179.

25.J.A. Koziol Comments on the rank product method for analyzing replicated experiments FEBS Lett., 584 (2010), pp. 941–944

26.Eisinga, R., Breitling, R., and Heskes, T. (2013). The exact probability distribution of the rank product statistics for replicated experiments. FEBS Letters, 587:677—682

27.Storey JD, Tibshirani R.: Statistical methods for identifying differentially expressed genes in DNA microarrays. Methods Mol Biol. 2003;224:149-57

28.Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

29.Jui-Hung Hung, Tun-Hsiang Yang, Zhenjun Hu, Zhiping Weng and Charles DeLisi. Gene set enrichment analysis: performance evaluation and usage guidelines. Brief Bioinform. 2012 May;13(3):281-91.

30.Molecular Signatures Database v3.0. Available from: http://www.broadinstitute.org/gsea/msigdb/index.jsp (1th December 2014, date last accessed).

31. Bradley Efron. Second Thoughts on the Bootstrap. Statist. Sci. Volume 18, Issue 2 (2003), 135-140.

32.Wu X, Hasan MA, Chen JY. Pathway and network analysis in proteomics. J Theor Biol. 2014 Dec 7;362:44-52.

33.J. L. Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. The American Statistician, 42(1):59–66, February 1988

34.Dowdy, S. and Wearden, S. (1983). "Statistics for Research", Wiley. ISBN 0-471-08602-9 pp 230

35.Fisher, J., Henzinger, T., 2007. Executable cell biology. Nat. Biotechnol. 25 (11), 1239–1249

36.van Riel, N., 2006. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. Brief Bioinform. 7 (4), 364–374;

37.Kell, D., Knowles, J., 2006. The role of modeling in systems biology. In: System Modeling in Cellular Biology: From Concepts to Nuts and Bolts.

MIT Press, Cambridge, MA, pp. 3–18.

38.Resat H, Petzold L, Pettigrew MF. Kinetic modeling of biological systems.Methods Mol Biol. 2009;541:311-35.

39.Chakrabarti A, Miskovic L, Soh KC, Hatzimanikatis V. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. Biotechnol J. 2013 Sep;8(9):1043-57.

40.Alves R, Antunes F, Salvador A.. Tools for kinetic modeling of biochemical networks. Nat Biotechnol. 2006 Jun;24(6):667-72.

41.Vizan P,Sanchez-Tena S, Alcarraz-Vizan G, Soler M, Messeguer R, Pujol MD, Lee WP, Cascante M. Characterization of the metabolic changes underlying growth factor angiogenic activation: identification of new potential therapeutic targets. Carcinogenesis 2009, 30:946-952.,

42.Vizan P, Boros LG, Figueras A, Capella G, Mangues R, Bassilian S, Lim S, Lee WP, Cascante M. K-ras codon-specific mutations produce distinctive metabolic phenotypes in NIH3T3 mice [corrected] fibroblasts. Cancer Res 2005, 65:5512-5515.

43.Marin S, Lee WP, Bassilian S, Lim S, Boros LG, Centelles JJ,FernAndez-Novell JM, Guinovart JJ, Cascante M. Dynamic profiling of the glucose metabolic network in fasted rat hepatocytes using [1,2- 13 C2]glucose. Biochem J 2004, 381:287-294.

44.Boros LG, Lapis K, Szende B, Tomoskozi-Farkas R, Balogh A, Boren J, Marin S, Cascante M, Hidvegi M. Wheat germ extract decreases glucose uptake and RNA ribose formation but increases fatty acid synthesis in MIA pancreatic adenocarcinoma cells. Pancreas 2001, 23:141-147.

45.Boros LG, Cascante M, Lee WNP: Metabolic profiling of cell growth and death in cancer. applications in drug discovery. Drug Discov Today 2002,

7:364-372.

46.Boren J, Cascante M, Marin S, Comin-Anduix B, Centelles JJ, Lim S, Bassilian S, Ahmed S, Lee WN, Boros LG. Gleevec (STI571) influences metabolic enzyme activities and glucose carbon flow toward nucleic acid and fatty acid synthesis in myeloid tumor cells. J Biol Chem 2001, 276:37747-37753.

47.Nargund S, Sriram G. Mathematical modeling of isotope labeling experiments for metabolic flux analysis. Methods Mol Biol. 2014;1083:109-31.

48.Wiechert W: 13C metabolic flux analysis. Metab Eng 2001, 3:195-206.

49.Selivanov VA, Puigjaner J, Sillero A, Centelles JJ, Ramos-Montoya A, Lee PW, Cascante M. An optimized algorithm for flux estimation from isotopomer distribution in glucose metabolites. Bioinformatics 2004, 20:3387-3397

50.Selivanov VA, Meshalkina LE, Solovjeva ON, Kuchel PW, Ramos-Montoya A, Kochetov GA, Lee PW, Cascante M. Rapid simulation and analysis of isotopomer distributions using constraints based on enzyme mechanisms: an example from HT29 cancer cells. Bioinformatics. 2005 Sep 1;21(17):3558-64.

51.Fan TWM, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, Miller DM. Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM). Mol Cancer 2009, 8: 41.

52.Amaral AI, Teixeira AP, Martens S, Bernal V, Sousa MF, Alves PM. Metabolic alterations induced by ischemia in primary cultures of astrocytes. merging 13C NMR spectroscopy and metabolic flux analysis. J Neurochem 2010, 113:735-748.

53.Selivanov VA, Vizán P, Mollinedo F, Fan TW, Lee PW, Cascante M. Edelfosine-induced metabolic changes in cancer cells that precede the overproduction of reactive oxygen species and apoptosis. BMC Syst Biol 2010, 4:135.

54.Jouhten P, Pitkänen E, Pakula T, Saloheimo M, Penttilä M, Maaheimo H. 13C-metabolic flux ratio and novel carbon path analyses confirmed that Trichoderma reesei uses primarily the respirative pathway also on the preferred carbon source glucose. BMC Syst Biol 2009, 3:104.

55.Aboka FO, Heijnen JJ, van Winden WA. Dynamic 13C-tracer study of storage carbohydrate pools in aerobic glucose-limited Saccharomyces cerevisiae confirms a rapid steady-state turnover and fast mobilization during a modest stepup in the glucose uptake rate. FEMS Yeast Res 2009, 9:191-201.

56.Mardinoglu A, Nielsen J. Systems medicine and metabolic modelling. J. Intern. Med. 271(2), 142-154 (2012).

57.Duarte NC, Becker SA, Jamshidi N et al.. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Nat. Acad. Sci. 104(6), 1777-1782 (2007).

58.Ma H1, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I. The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol. 2007;3:135.

59.Thiele I, Swainston N, Fleming RM et al.. A community-driven global reconstruction of human metabolism. Nat. Biotech. 31(5), 419-425 (2013).

60.Bordbar A, Feist AM, Usaite-Black R, Woodcock J, Palsson BO, Famili I. A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. BMC Syst. Biol. 5(1), 180 (2011).;

61.Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst. Biol. 6(1), 153 (2012).;

62.Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS Comput Biol 8(5), e1002518 (2012).

63.Jerby L, Wolf L, Denkert C et al.. Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. Cancer Res. 72(22), 5712-5720 (2012).

64.Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat. Biotech. 28(3), 245-248 (2010).

65.Papp B, Notebaart RA, Pál C. Systems-biology approaches for predicting genomic evolution. Nat. Rev. Genet. 12(9), 591-602 (2011).

66.Matsuda F, Furusawa C, Kondo T, Ishii J, Shimizu H, Kondo A. Engineering strategy of yeast metabolism for higher alcohol production. Microb. Cell Fact. 10(1), 70 (2011).

67.Park JM, Song H, Lee HJ, Seung D. Genome-scale reconstruction and in silico analysis of Klebsiella oxytoca for 2,3-butanediol production. Microb. Cell Fact. 12(1), 20 (2013).

68.Lütke-Eversloh T, Bahl H. Metabolic engineering of Clostridium acetobutylicum: recent advances to improve butanol production. Curr. Opin. Biotechnol. 22(5), 634-647 (2011).

69.Kumar VS, Maranas CD. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. PLoS Comput. Biol. 5(3), e1000308 (2009).

70.Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J:

Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol. Syst. Biol. 10(3), (2014).

71.Wintermute E, Lieberman T, Silver P. An objective function exploiting suboptimal solutions in metabolic networks. BMC Syst. Biol. 7(1), 98 (2013).

72.Schellenberger J, Que R, Fleming RM et al.. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nat. Protoc. 6(9), 1290-1307 (2011).

73.Bordbar A, Mo ML, Nakayasu ES et al. Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. Mol. Syst. Biol. 8(1), (2012).

74.Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10(4), 291-305 (2012).

75.Jensen PA, Papin JA. MetDraw: automated visualization of genome-scale metabolic network reconstructions and high-throughput data. Bioinformatics, btt758 (2014)

76.Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc. Nat. Acad. Sci. 107(41), 17845-17850 (2010).

77.Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat. Rev. Genet. 12(1), 56-68 (2011).

78.Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Molecular Systems Biology 6, (2010).

79.

80.Lewis NE, Schramm G, Bordbar A et al.. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. Nat. Biotech. 28(12), 1279-1285 (2010).

81.Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. Bioinformatics 26(24), 3140-3142 (2010).

82.Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. Network-based prediction of human tissue-specific metabolism. Nature Biotechnology 26(9), 1003-1010 (2008).

83.Papin JA, Colijn C, Brandes A et al.. Interpreting Expression Data with Metabolic Flux Models: Predicting Mycobacterium tuberculosis Mycolic Acid Production. PLoS Computational Biology 5(8), e1000489 (2009).

84.Jensen PA, Papin JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics 27(4), 541-547 (2010).

85.Gonçalves E, Pereira R, Rocha I, Rocha M. Optimization Approaches for theIn SilicoDiscovery of Optimal Targets for Gene Over/Underexpression. Journal of Computational Biology 19(2), 102-114 (2012).

86.Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC Syst Biol 6, 153 (2012).

87.Galhardo M, Sinkkonen L, Berninger P, Lin J, Sauter T, Heinaniemi M. Integrated analysis of transcript-level regulation of metabolism reveals disease-relevant nodes of the human metabolic network. Nucleic Acids Res 42(3), 1474-1496 (2014).

88.Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BO, Hyduke DR.

GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics 29(22), 2900-2908 (2013).

89.Maranas CD, Dreyfuss JM, Zucker JD et al.. Reconstruction and Validation of a Genome-Scale Metabolic Model for the Filamentous Fungus Neurospora crassa Using FARM. PLoS Computational Biology 9(7), e1003126 (2013).

90.Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. Mol. Syst. Biol. 7(1), (2011).

91.Davis VW, Bathe OF, Schiller DE, Slupsky CM, Sawyer MB. Metabolomics and surgical oncology: Potential role for small molecule biomarkers. J. Surg. Oncol. 103(5), 451-459 (2011).

92.Murabito E, Simeonidis E, Smallbone K, Swinton J. Capturing the essence of a metabolic network: a flux balance analysis approach. J. Theor. Biol. 260(3), 445-452 (2009).

93.Schellenberger J, Palsson BØ. Use of randomized sampling for analysis of metabolic networks. J. Biol. Chem. 284(9), 5457-5461 (2009).

94.Rohit Duggal BM, Ulrike Geissinger, Huiqiang Wang, Nanhai G. Chen, Prasad S. Koka, and Aladar A. Szalay. Biotherapeutic Approaches to Target Cancer Stem Cells. J. Stem Cells 8(3-4), 135-149 (2013).

95.Philips MR, Cox AD. Geranylgeranyltransferase I as a target for anti-cancer drugs. J. Clin. Invest. 117(5), 1223-1225 (2007).

96.Dudakovic A, Tong H, Hohl R. Geranylgeranyl diphosphate depletion inhibits breast cancer cell migration. Invest. New Drugs 29(5), 912-920 (2011).

97.Hebar A, Valent P, Selzer E. The impact of molecular targets in cancer

drug development: major hurdles and future strategies. Expert Rev. Clin. Pharmacol., (2013).

98.Conde-Pueyo N, Munteanu A, Solé RV, Rodríguez-Caso C. Human synthetic lethal inference as potential anti-cancer target gene detection. BMC Syst. Biol. 3(1), 116 (2009).

99.Barbash DA, Lorigan JG. Lethality in Drosophila melanogaster/Drosophila simulans species hybrids is not associated with substantial transcriptional misregulation. J. Exp. Zool., Part B 308(1), 74-84 (2007).

100.Ren S, Zeng B, Qian X. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. BMC bioinformatics 14(Suppl 2), S17 (2013).

101.Frezza C, Zheng L, Folger O et al.: Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. Nature 477(7363), 225-228 (2011).

102.Huang S. Genomics, complexity and drug discovery: insights from Boolean network models of cellular regulation. Pharmacogenomics. 2001 Aug;2(3):203-22.

103.de Jong, H., Ropers, D. Qualitative approaches to the analysis of genetic regulatory networks. In: System Modeling in Cellular Biology: From Concepts to Nuts and Bolts. 2006 MIT Press, Cambridge, MA, pp. 125–148

104.Thomas R, Kaufman M. Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. Chaos. 2001 Mar;11(1):170-179.

105.Corblin F, Tripodi S, Fanchon E, Ropers D, Trilling L. A declarative constraint-based method for analyzing discrete genetic regulatory

networks. Biosystems. 2009 Nov;98(2):91-104

106.Lazebnik Y. Can a biologist fix a radio? Or, what I learned while studying apoptosis. Cancer Cell 2:179–182, Mendes P, Kell D (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 2002 14:869–883

107.Szallasi Z, Stelling J, Periwal V (2006) System modeling in cellular biology: from concepts to nuts and bolts. MIT Press, Boston

108.Mendes P, Kell D (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 14:869–883

109.Tummler K, Lubitz T, Schelker M, Klipp E. New types of experimental data shape the use of enzyme kinetics for dynamic network modeling. FEBS J. 2014 Jan;281(2):549-71.

110.Hynne F, Danø S, Sørensen PG. Full-scale model of glycolysis in Saccharomyces cerevisiae . Biophys Chem, 2001, 94 : 121 − 163.

111.Yang K, Ma W, Liang H, et al . Dynamic simulations on the arachidonic acid metabolic network. PLOS. Comput Biol , 2007, 3 : e55.

112.Snoep JL. The silicon cell initiative: working towa rds a detailed kinetic description at the cellular level. Curr Opin Biotech , 2005, 16 : 336 − 343.

113.Fung E, Wong WW, Suen JK, et al . A synthetic gene- metabolic oscillator. Nature , 2005, 435 : 118 − 122.

114.Goldbeter A, Gonze D, Houart G, et al . From simple to complex oscillatory behavior in metabolic and genet ic control networks. Chaos , 2001, 11 : 247 − 260.

115.Craciun G, Tang YZ, Feinberg M. Understanding bistability in complex

enzyme-driven reaction netwo rks. PNAS, 2006, 103 : 8697 − 8702.

116.Zhu X-M, Yin L, Hood L, et al . Calculating robustness of epigenetic states in phage λ life cycle. Funct Integr Genomics , 2004, 4 : 188 − 195.

117.Fendt SM, Sauer U: Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. BMC Syst Biol 2010, 4:12.

118.van Winden W, Verheijen P, Heijnen S: Possible pitfalls of flux calculations based on (13)C-labeling. Metab Eng 2001, 3:151-162.

119.Hardin CD, Finder DR: Glycolytic flux in permeabilized freshly isolated vascular smooth muscle cells. Am J Physiol 1998, 274, C88-C96

120.Cascante M, Centelles JJ, Agius L: Use of alpha-toxin from Staphylococcus aureus to test for channelling of intermediates of glycolysis between glucokinase and aldolase in hepatocytes. Biochem J, 2000, 352, 899-905

121.Butcher, E. C., Berg, E. L. & Kunkel, E. J. Systems biology in drug discovery. Nat. Biotech. 22, 1253–1259 (2004).

122.Shlomi, Tomer, Omer Berkman, and Eytan Ruppin. "Regulatory On/off Minimization of Metabolic Flux Changes After Genetic Perturbations." Proceedings of the National Academy of Sciences of the United States of America 102, no. 21 (May 24, 2005): 7695–7700.

123.T. Kaelin WG Jr. The concept of synthetic lethality in the context of anticancer therapy. Nat Rev Cancer 2005; 5:689-98. U. Lucchesi JC. Synthetic lethality and semi-lethality among functionally related mutants of Drosophila melanfgaster. Genetics 1968; 59:37-44

124.Peinado H, Olmeda D, Cano A: Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat. Rev. Cancer 7(6), 415-428 (2007).

125.Celià-Terrassa T, Meca-Cortés Ó, Mateo F et al.: Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. J. Clin. Invest. 122(5), 1849 (2012).

126.Korpal M, Ell BJ, Buffa FM et al.: Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. Nat. med. 17(9), 1101-1108 (2011).

127.Ocaña OH, Córcoles R, Fabra Á et al.: Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. Cancer cell 22(6), 709-724 (2012).

128.Celià-Terrassa T, Meca-Cortés O, Mateo F, de Paz AM, Rubio N, Arnal-Estapé A, Ell BJ, Bermudo R, Díaz A, Guerra-Rebollo M, Lozano JJ, Estarás C, Ulloa C, Álvarez-Simón D, Milà J, Vilella R, Paciucci R, Martínez-Balbás M, de Herreros AG, Gomis RR, Kang Y, Blanco J, Fernández PL, Thomson TM. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. J Clin Invest. 2012 May 1;122(5):1849-68.

129.Vaupel P: Tumor microenvironmental physiology and its implications for radiation oncology. In: Seminars in radiation oncology. 2004.

130.Tannock IF, Lee CM, Tunggal JK, Cowan DS, Egorin MJ: Limited penetration of anticancer drugs through tumor tissue a potential cause of resistance of solid tumors to chemotherapy. Clin. Cancer Res. 2002 8(3)

131.Long TA, Brady SM, Benfey PN (2008) Systems approaches to identifying gene regulatory networks in plants. 24: 81–103

132.Wellmer F, Riechmann JL (2010) Gene networks controlling the initiation of flower development. 26: 519–527

133.Tian C, Zhang X, He J, Yu H, Wang Y, Shi B, Han Y, Wang G, Feng X, Zhang C, Wang J, Qi J, Yu R, Jiao Y. An organ boundary-enriched gene

regulatory network uncovers regulatory hierarchies underlying axillary meristem initiation. Mol Syst Biol. 2014 Oct 30;10:755

134.Marín de Mas I, Aguilar E, Jayaraman A, Polat IH, Martín-Bernabé A, Bharat R, Foguet C, Milà E, Papp B, Centelles JJ, Cascante M. Cancer cell metabolism as new targets for novel designed therapies. Future Med Chem. 2014;6(16):1791-810.

135.Vestbo J, Wedzicha J.NICE and GOLD response. Lancet Respir Med. 2013 Aug;1(6):442.

136.Murray CJ, Lopez AD. Measuring the global burden of disease. N Engl J Med. 2013 Aug 1;369(5):448-57.

137.Maltais F, Decramer M, Casaburi R, Barreiro E, Burelle Y, Debigaré R, Dekhuijzen PN, Franssen F, Gayan-Ramirez G, Gea J, Gosker HR, Gosselink R, Hayot M, Hussain SN, Janssens W, Polkey MI, Roca J, Saey D, Schols AM, Spruit MA, Steiner M, Taivassalo T, Troosters T, Vogiatzis I, Wagner PD; ATS/ERS Ad Hoc Committee on Limb Muscle Dysfunction in COPD. An official American Thoracic Society/European Respiratory Society statement: update on limb muscle dysfunction in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2014 May 1;189(9):e15-62

138.Mishra PK, Raghuram GV, Jain D, Jain SK, Khare NK, Pathak N. Mitochondrial oxidative stress-induced epigenetic modifications in pancreatic epithelial cells. Int J Toxicol. 2014 Mar-Apr;33(2):116-29.)

139.Sala E, RocaJ, Marrades RM, Alonso J, Gonzalez De Suso JM, Moreno A, Barberá JA, Nadal J, de Jover L, Rodriguez-Roisin R, Wagner PD.: Effects of endurance training on skeletal muscle bioenergetics in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1999;159:1726-34

140.Roca J, Vargas C, Cano I, Selivanov V, Barreiro E, Maier D, Falciani F, Wagner P, Cascante M, Garcia-Aymerich J, Kalko S, De Mas I, Tegnér J, Escarrabill J, Agustí A, Gomez-Cabrero D; Synergy-COPD consortium. Chronic Obstructive Pulmonary Disease heterogeneity: challenges for health risk assessment, stratification and management. J Transl Med. 2014 Nov 28;12 Suppl 2:S3.

141.SelivanovVA, Votyakova TV, Zeak JA, Trucco M, Roca J, Cascante M.: Bistability of mitochondrial respiration underlies paradoxical reactive oxygen species generation induced by anoxia. PLoS Comput Biol. 2009;5:e1000619.

142.Morrow JD, Qiu W, Chhabra D, Rennard SI, Belloni P, Belousov A, Pillai SG, Hersh CP., Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. BMC Med Genomics. 2015 Jan 13;8(1):1.,

143.Zhang WJ, Hubbard Cristinacce PL, Bondesson E, Nordenmark LH, Young SS, Liu YZ, Singh D, Naish JH, Parker GJ. MR Quantitative Equilibrium Signal Mapping: A Reliable Alternative to CT in the Assessment of Emphysema in Patients with Chronic Obstructive Pulmonary Disease. Radiology. 2015 Jan 7:132953.

144.Wright, S. (1921). "Correlation and Causation" (PDF). Journal of Agricultural Research 20 (7): 557–585.

145.Marin S, Lee WP, Bassilian S, Lim S, Boros LG, Centelles JJ, FernAndez-Novell JM, Guinovart JJ, Cascante M: Dynamic profiling of the glucose metabolic network in fasted rat hepatocytes using [1,2-13C2]glucose. Biochem J 2004, 381:287-294.

146.Esther Aguilar, Igor Marín, Erika Zodda, Vitaly Selivanov, Silvia Marín, Pedro de Atauri, Hossain Delowar, Mònica Pons, Antoni Celià-Terrassa,

Óscar Meca-Cortés, Fionnuala Morrish, David Hockenbery, Josep J Centelles, Timothy M Thomson and Marta Cascante. Metabolic landscape and vulnerabilities of prostate metastatic epithelial cancer stem cell independent of epithelial-mesenchymal transition. Submited to Cancer Cell January 2015

147.Muir K, Hazim A, He Y, Peyressatre M, Kim DY, Song X, Beretta L. Proteomic and lipidomic signatures of lipid metabolism in NASH-associated hepatocellular carcinoma. Cancer Res. 2013 Aug 1;73(15):4722-31.

148.Koontongkaew S, Monthanapisut P, Saensuk T. Inhibition of arachidonic acid metabolism decreases tumor cell invasion and matrix metalloproteinase expression. Prostaglandins Other Lipid Mediat. 2010 Nov;93(3-4):100-8.

149.Liu SY, Zhang RL, Kang H, Fan ZJ, Du Z. Human liver tissue metabolic profiling research on hepatitis B virus-related hepatocellular carcinoma. World J Gastroenterol. 2013 Jun 14;19(22):3423-32.

150.Flowers M, Fabriás G, Delgado A, Casas J, Abad JL, Cabot MC. C6-ceramide and targeted inhibition of acid ceramidase induce synergistic decreases in breast cancer cell growth. Breast Cancer Res Treat. 2012 Jun;133(2):447-58.

151.Fritz V, Benfodda Z, Henriquet C, Hure S, Cristol JP, Michel F, Carbonneau MA, Casas F, Fajas L. Metabolic intervention on lipid synthesis converging pathways abrogates prostate cancer growth. Oncogene. 2013 Oct 17;32(42):5101-10.

152.Pierron A, Le Pape E, Montaudié H, Castela E, De Donatis GM, Allegra M, Bertolotto C, Rocchi S, Cheli Y, Ballotti R, Passeron T. PGJ2 restores RA sensitivity in melanoma cells by decreasing PRAME and EZH2. J

Dermatol Sci. 2014 Mar;73(3):258-61.

153.Kojima S, Sugiura T, Waku K, Kamikawa Y.. Contractile response to a cannabimimetic eicosanoid, 2-arachidonoylglycerol, of longitudinal smooth muscle from the guinea-pig distal colon in vitro. Eur J Pharmacol. 2002 May 31;444(3):203-7.

154.Nithipatikom K, Gomez-Granados AD, Tang AT, Pfeiffer AW, Williams CL, Campbell WB. Cannabinoid receptor type 1 (CB1) activation inhibits small GTPase RhoA activity and regulates motility of prostate carcinoma cells. Endocrinology. 2012 Jan;153(1):29-41.

155.Munetsuna E, Kawanami R, Nishikawa M, Ikeda S, Nakabayashi S, Yasuda K, Ohta M, Kamakura M, Ikushiro S, Sakaki T. Anti-proliferative activity of 25-hydroxyvitamin D3 in human prostate cells. Mol Cell Endocrinol. 2014 Feb 15;382(2):960-70.

156.Trump DL, Muindi J, Fakih M, Yu WD, Johnson CS. Vitamin D compounds: clinical development as cancer therapy and prevention agents. Anticancer Res. 2006 Jul-Aug;26(4A):2551-6.

157.Losko S, Wenger K, Kalus W, Ramge A, Wiehler J, Heumann K: Knowledge Networks of Biological and Medical Data: An Exhaustive and Flexible Solution to Model Life Science Domains. In Data Integration in the Life Sciences. Volume 4075/2006. Springer Berlin/Heidelberg; 2006:232-239.

158.Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014 Feb 15;30(4):523-30.

159.Murali T, Pacifico S, Yu J, Guest S, Roberts GG 3rd, Finley RL Jr. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. Nucleic

Acids Res. 2011 Jan;39(Database issue):D736-43.

160.Puente-Maestu L, Tejedor A, Lázaro A, de Miguel J, Alvarez-Sala L, González- Aragoneses F, Simón C, Agustí A.: Site of Mitochondrial ROS Production in Skeletal Muscle of COPD and its Relationship with Exercise Oxidative Stress. Am J Respir Cell Mol Biol. 2012

161.Dieter Maier, Wenzel Kalus, Martin Wolff, Susana G Kalko, Josep Roca, Igor Marin de Mas, Nil Turan, Marta Cascante, Francesco Falciani, Miguel Hernandez, Jordi Villà-Freixa, Sascha Losko. Knowledge management for systems biology a general and visually driven framework applied to translational medicine. BMC Syst Biol. 2011; 5: 38. Published online 2011 March 5.

162.Karpen JW, Rich TC. The fourth dimension in cellular signaling. Science. 2001 Sep 21;293(5538):2204-5.

163.SelivanovVA, Krause S, Roca J, CascanteM. Modeling of spatial metabolite distributions in the cardiac sarcomere. Biophys J. 2007 May 15;92(10):3492-500.

164.Clegg JS1, Jackson SA. Glucose metabolism and the channeling of glycolytic intermediates in permeabilized L-929 cells. Arch Biochem Biophys. 1990 May 1;278(2):452-60.

165.Kaikkonen S, Paakinaho V, Sutinen P, Levonen AL, Palvimo J, Prostaglandin 15d-PGJ(2) inhibits androgen receptor signaling in prostate cancer cells. Mol Endocrinol. 2013 Feb;27(2):212-23.

166.Wranicz J, Szostak-Węgierek D. Health outcomes of vitamin D. Part I. Characteristics and classic role. Rocz Panstw Zakl Hig. 2014;65(3):179-84.

167.Kasem Nithipatikom and William B. Campbell. Roles of Eicosanoids in Prostate Cancer. Future Lipidol. Aug 1, 2008; 3(4): 453–467.

# 06. Publications

# Chapter 1:
# Study of the metabolic channeling in hepatocytes by developing a novel model-driven approach based on non-stationary $^{13}$C-FBA

## Summary

In this chapter we overcome the existing limitations in the analysis of the topology of metabolic pathways. More specifically we developed a computational method to evaluate and discriminate different metabolic topologies based on their consistency with the experimental isotopic distribution of substrates . This chapter summarizes the result from these analysis which resulted in a publication

# Chapter 1a

## Compartmentalization of glycogen metabolism revealed from 13C isotopologue distributions.

**Igor Marin de Mas**[1], Vitaly A. Selivanov[1, 2, §], Silvia Marin[1], Josep Roca[3], Matej Orešič[4], Loranne Agius[5] and Marta Cascante[1§]

[1]Department of Biochemistry and Molecular Biology, Faculty of Biology, Institute of Biomedicine of University of Barcelona (IBUB) and IDIBAPS, Unit Associated with CSIC, Diagonal 645, 08028-Barcelona, Spain

[2]A.N.Belozersky Institute of Physico-Chemical Biology, MSU, Moscow 199899, Russia

[3]Hospital Clínic, IDIBAPS, CIBERES; Universitat de Barcelona, Barcelona 08028, Spain

[4]Technical Research Centre of Finland, Espoo, and Institute for Molecular Medicine, Helsinki, Finland

[5]Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, UK

Marin de Mas et al. BMC Systems Biology 2011, 5:175
http://www.biomedcentral.com/1752-0509/5/175

BMC Systems Biology

METHODOLOGY ARTICLE Open Access

# Compartmentation of glycogen metabolism revealed from $^{13}C$ isotopologue distributions

Igor Marin de Mas[1,2], Vitaly A Selivanov[1,2,3*], Silvia Marin[1,2], Josep Roca[4], Matej Orešič[5], Loranne Agius[6] and Marta Cascante[1,2*]

**Abstract**

**Background:** Stable isotope tracers are used to assess metabolic flux profiles in living cells. The existing methods of measurement average out the isotopic isomer distribution in metabolites throughout the cell, whereas the knowledge of compartmental organization of analyzed pathways is crucial for the evaluation of true fluxes. That is why we accepted a challenge to create a software tool that allows deciphering the compartmentation of metabolites based on the analysis of average isotopic isomer distribution.

**Results:** The software Isodyn, which simulates the dynamics of isotopic isomer distribution in central metabolic pathways, was supplemented by algorithms facilitating the transition between various analyzed metabolic schemes, and by the tools for model discrimination. It simulated $^{13}C$ isotope distributions in glucose, lactate, glutamate and glycogen, measured by mass spectrometry after incubation of hepatocytes in the presence of only labeled glucose or glucose and lactate together (with label either in glucose or lactate). The simulations assumed either a single intracellular hexose phosphate pool, or also channeling of hexose phosphates resulting in a different isotopic composition of glycogen. Model discrimination test was applied to check the consistency of both models with experimental data. Metabolic flux profiles, evaluated with the accepted model that assumes channeling, revealed the range of changes in metabolic fluxes in liver cells.

**Conclusions:** The analysis of compartmentation of metabolic networks based on the measured $^{13}C$ distribution was included in Isodyn as a routine procedure. The advantage of this implementation is that, being a part of evaluation of metabolic fluxes, it does not require additional experiments to study metabolic compartmentation. The analysis of experimental data revealed that the distribution of measured $^{13}C$-labeled glucose metabolites is inconsistent with the idea of perfect mixing of hexose phosphates in cytosol. In contrast, the observed distribution indicates the presence of a separate pool of hexose phosphates that is channeled towards glycogen synthesis.

## Background

$^{13}C$ isotope tracing, aimed in the evaluation of metabolic fluxes in living cells has been developing during last decades [1]. This experimental technique required a specific mathematical analysis, and it was created [2]. Currently, the stable isotope tracing of metabolites has been refined and is used to identify the adaptive changes of fluxes in man in normal and diseased states [3], in isolated cells [4], cancer cell cultures [5], and organisms such as fungi [6], yeast [7,8], etc. $^{13}C$ tracer fluxomics can be combined with the analysis of gene and protein expressions to provide insight into multilevel regulation of cellular processes [9].

However, the rapidly developing experimental $^{13}C$ tracer metabolomics surpasses the theoretical analysis of measured data. For a long time the detailed analysis of isotopomer distribution was possible only for isotopic steady state [10]. The tools applicable for analysis of non-steady state conditions appeared relatively recently [11-14], and the methodology of rule-based modeling used in some of these tools expanded to different areas of analysis of complex biological systems [15]. Although the analysis of $^{13}C$ tracer data could result in the discovery of unknown metabolic pathways [16], the existing tools were designed mainly for the evaluation of metabolic fluxes assuming certain established topology of

* Correspondence: selivanov@ub.edu; martacascante@ub.edu
[1]Department of Biochemistry and Molecular Biology, Faculty of Biology, Universitat de Barcelona, Av Diagonal 643, 08028 Barcelona, Spain
Full list of author information is available at the end of the article

reaction network. However, ignoring the specificity of topology of particular reaction network, or in other words its compartmental structure, can compromise the results of metabolic flux analysis [17].

The topology of metabolic network could be complicated by substrate channeling [18-24], which could be seen as metabolite compartmentation. The latter follows from the definition, which says that a pathway intermediate is 'channeled' when, a product just produced in the pathway has a higher probability of being a substrate for the next enzyme in the same pathway, compared to a molecule of the same species produced in a different place [23,25].

Usually, studies designed for the analysis of channeling require invasive experiments, such as permeabilization of cells and determination of diffusion of labeled metabolites from or into the presumable channel [22-24]. However, it can be expected that experimental procedures destroy some kinds of channeling that occur in intact cells. Moreover, one cannot exclude the possibility that the metabolic channeling and compartmentation differ between various tissues and this could increase indefinitely the number of experiments necessary for defining the structure of metabolism in cells. Here, we propose a solution for such a problem: to determine the metabolic compartmentation by analyzing $^{13}C$ isotopic isomer distributions in products of metabolism of labeled substrates; i.e. in the same study, which is designed for the evaluation of metabolic flux profile, thereby, not recruiting additional experiments.

Thus, the objective of the presented work is to create and implement a tool assessing the compartmentation based on $^{13}C$ distribution. The challenge here is that, although the same compound, located in different subcellular spaces, likely possesses compartment-specific $^{13}C$ signatures, the measurements average out the compartment-specificity [17,26,27]. The tool must help decipher the compartment-specific distribution of metabolic fluxes, consistent with the measured average labeling. Such deciphering is based on a simple idea that the compartment-specific simulation better fit $^{13}C$ data, if the really existing compartments are taken into account. To estimate the goodness of data fit by various schemes of metabolic compartmentation we implement model discrimination analysis.

Two out of three experiments analyzed were described elsewhere [28], and metabolic fluxes were evaluated based on the application of simple formulas directly to experimental data. Such simple analysis (the only achievable then) does not account for all possible exchange of isotopes and their recycling. Moreover, these formulas imply that the network topology is known and can only give a formal ratio of main fluxes without its verification. Whereas the simulation of isotopic isomer distribution using the predicted fluxes can be compared with experiments and thus verify the predictions. Here, we describe the use of such simulations for the analysis of network topology, which is absolutely impossible by using simple formulas. The new analytical tool provides the opportunity to re-evaluate previously generated experimental data gaining new insights into the topology of the studied metabolic network, and assessing metabolic flux profile in detail in various physiological and pathological conditions.

## Results

### Accounting for channeling in the reaction scheme of model

The dynamics of all possible isotopic isomers in glucose, lactate and glutamate from the incubation medium and glucose from glycogen in cell pellets, accumulated by two hours of incubation of liver cells with [1,2-$^{13}C_2$]D-glucose [28], were simulated with Isodyn using two different schemes that included either one or two pools of hexose phosphates, as shown in Figure 1. The conditions of incubation do not assume to reach steady state for the labeling of measured external metabolites, and the dynamic simulations correspond to non-steady state conditions of the experiment. The model, which does not consider channeling (referred to as model A), accounts for a single, well-mixed, common hexose 6-phosphate pool (Figure 1A). In accordance with the definition given in introduction, channeling assumes the existence of metabolite compartments, which could have had a different isotopomer (and isotopologue) composition and does not freely mix by diffusion with a pool of the same metabolite outside the channel. The presence of two compartments with different isotopomer compositions indicates metabolite channeling (Figure 1B), and the respective model is referred to as model B.

### Fitting the measured isotopologue distribution

The same stochastic algorithm of minimization of normalized deviations between the measured and computed data ($\chi^2$) in the global space of parameters (described in Methods) was applied to each model for data fitting. The fractions of isotopologues measured in glucose, lactate, fragments of glutamate and glycogen, and their best fit using the two models are shown in Table 1. $\chi^2$ is shown for each metabolite separately and $\Sigma_1\chi^2$ sums all the individual $\chi^2$.

Model A that does not assume metabolite channeling in glycogen synthesis fitted the experimental isotopologue distribution with $\Sigma_1\chi^2$ = 38.28 (Table 1). 22 experimental points and 16 essential parameters, calculated as described in Methods, defined the number of degrees of freedom, F = 6. The given values of $\Sigma_1\chi^2$ and the number of degrees of freedom defining an extremely low

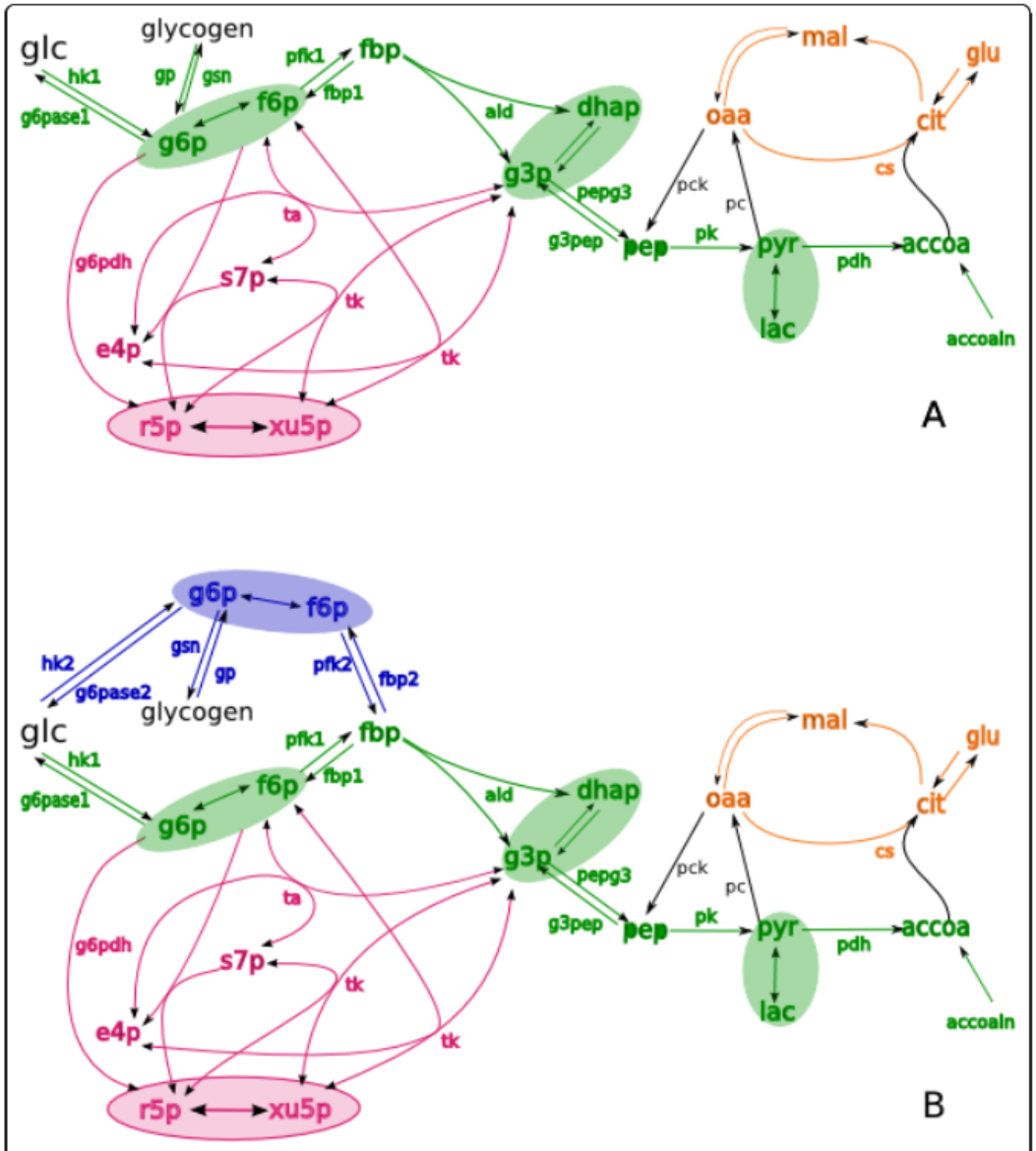



**Figure 1 The schemes of kinetic models used as a base for simulation of isotopologue distribution**. Metabolites are connected by biochemical reactions represented by arrows. Various colors indicate metabolites and reactions of specific pathways: green, glycolysis/ gluconeogenesis; red, pentose phosphate pathways; orange, TCA cycle. The metabolites enclosed in ellipses are considered to be in fast equilibrium. (A), the basic model that includes one pool of hexose phosphates common for glycolysis and gluconeogenesis. (B), the model that includes also the additional pool of hexose phosphates (blue) that represents channeling in gluconeogenesis. Abbreviations are explained in the list of abbreviations in the text.

**Table 1 Measured and simulated fractions of isotopologues and total concentrations of metabolites.**

| | Experiment | Simulated | |
|---|---|---|---|
| | mean sd | Channeling | Mixed |
| **Glucose** | | **$\chi^2$ = 0.442** | **0.406** |
| m0 | 0.512 ± 0.0069 | 0.511 | 0.511 |
| m1 | 0.00913 ± 0.002 | 0.0084 | 0.00839 |
| m2 | 0.478 ± 0.00652 | 0.481 | 0.481 |
| [mM] | 19.7 ± 1.92 | 20.4 | 20.2 |
| **Lactate** | | **$\chi^2$ = 1.43** | **7.32** |
| m0 | 0.86 ± 0.0482 | 0.839 | **0.81** |
| m1 | 0.0235 ± 0.00802 | 0.0237 | **0.0178** |
| m2 | 0.0946 ± 0.0388 | **0.133** | **0.17** |
| m3 | 0.022 ± 0.0438 | 0.00381 | **0.00145** |
| [mM] | 0.81 ± 0.51 | 0.959 | **1.48** |
| **Glutamate C2-C5** | | **$\chi^2$ = 0.0564** | **0.0424** |
| m0 | 0.912 ± 0.0343 | 0.912 | 0.912 |
| m1 | 0.0299 ± 0.0116 | 0.0301 | 0.0298 |
| m2 | 0.0523 ± 0.0217 | 0.0574 | 0.0567 |
| **Glutamate C2-C4** | | **$\chi^2$ = 0.0049** | **0.00437** |
| m0 | 0.919 ± 0.0339 | 0.919 | 0.919 |
| m1 | 0.0365 ± 0.00175 | 0.0356 | 0.355 |
| m2 | 0.0446 ± 0.0166 | 0.0454 | 0.0451 |
| **Glycogen** | | **$\chi^2$ = 1.2** | **30.5** |
| m0 | 0.608 ± 0.0388 | 0.598 | **0.658** |
| m1 | 0.0162 ± 0.0033 | 0.0151 | **0.0271** |
| m2 | 0.362 ± 0.0351 | 0.375 | **0.299** |
| m3 | 0.00399 ± 0.0011 | 0.00422 | **0.00791** |
| m4 | 0.00961 ± 0.0026 | 0.00748 | 0.00749 |
| m5 | 0.000464 ± 0.00016 | 0.000432 | 0.000533 |
| mg/mL | 0.355 ± 0.112 | 0.313 | **0.232** |
| **$\Sigma_1 \chi^2$** | | **3.13** | **38.28** |
| **Glycogen C1-C4** | | **$\chi^2$ = 1.42** | **6.68** |
| m0 | 0.613 ± 0.0448 | 0.627 | **0.679** |
| m1 | 0.0224 ± 0.00834 | **0.0133** | 0.0297 |
| m2 | 0.357 ± 0.0425 | 0.358 | **0.289** |
| **Glycogen C3-C6** | | **$\chi^2$ = 7.97** | **30** |
| m0 | 0.952 ± 0.00767 | 0.952 | 0.951 |
| m1 | 0.00743 ± 0.00211 | **0.0131** | **0.018** |
| m2 | 0.0371 ± 0.00467 | 0.0333 | **0.0279** |
| **$\Sigma_2 \chi^2$** | | **9.21** | **36.68** |
| **$\Sigma_t \chi^2 = \Sigma_1 \chi^2 + \Sigma_2 \chi^2$** | | **12.52** | **74.95** |

Isotopologues (m0, non-labeled; m1, containing one 13C isotope; m2, two 13C isotopes, etc) produced by isolated hepatocytes from glucose as the only substrate contained 50% of [1,2-13C2]D-glucose were measured in glucose from medium, glucose from glycogen and its fragments, lactate, and fragments of glutamate after two hours of incubation. The measurements are presented as mean ± standard deviation. The data were simulated using two models that either accounted for channeling or suggested a single "mixed" pool of hexose phosphates in accordance with the schemes presented in Figure 1. The fitting was performed using a stochastic algorithm described in Methods. The difference between the best fit and experimental data ($\chi$2, see Methods) are shown for each metabolite and summarized for the whole set of data.

value of incomplete gamma function Q = $9.9 \cdot 10^{-7}$ unambiguously indicates that the model which does not account for channeling should be rejected [29].

Conversely, model B that assumes channeling fits the measured isotopologue distribution much better, with $\Sigma_1 \chi^2$ = 3.13 (Table 1). This model has four degrees of freedom, deduced from the same number of experimental points, and 18 essential parameters. These values defined Q = 0.536, which allowed the acceptance of the model.

Thus, the comparative study of two schemes based on the goodness of fit of the experimental data allowed rejection of the model that assumes a single common pool of hexose phosphates and acceptance of the alternative model, which accounts for channeling of intermediates in glucose metabolism.

### Model validation

Electron impact ionization used in mass spectrometry often splits molecules into fragments. Since the localization of such fragments in the molecule is known, the fact that a $^{13}C$ atom belongs to a given fragment restricts the possible positions of this isotope in the molecule. This information can further restrict the possible set of solutions. The fractions of isotopologues from glycogen were measured not only in whole glucose molecules, but also in their fragments containing carbon atoms either 1-4 [28] or 3-6. For model B, the best fit described above fits also the whole set of data including the fragments of glucose from glycogen as Table 1 shows, giving $\Sigma_t \chi^2$ = 12.52. This value and 8 degrees of freedom deduced for this model from 29 experimental points and 21 essential parameters define the value Q = 0.129. This value indicates that the model is acceptable, thus confirming the conclusion based on the simulation of $^{13}C$ distribution in the whole molecule of glucose from glycogen without accounting for the fragments.

As Table 1 shows, model A fits the whole set of data with $\Sigma_t \chi^2$ = 74.9. With number of degrees of freedom of 11 (29 experimental points and 18 essential parameters), the value of Q was $1.42 \cdot 10^{-11}$. This value further indicates that the model of a homogeneous pool of hexose phosphates should be rejected.

Another validation of the channeling came from a series of two experiments where hepatocytes were incubated in the presence of glucose and lactate (as described in Methods). The conditions in the two experiments were virtually identical with the exception that glucose was labeled in one of these experiments [28] and lactate in the other. Cells consumed lactate,

and the label from lactate ascended up to extracellular glucose as shown in Table 2. In these two experiments, the incubations with liver cells resulted in different isotopic isomer distribution since different substrates were labeled at the beginning. However, the same conditions for cell metabolism suggest that the same set of metabolic fluxes must fit the isotopologue distribution measured in both the experiments. Respectively, the algorithm subsequently performed the simulations of two experiments for the same set of model parameters and summed $\chi 2$ for both simulations. The algorithm for parameter optimization had searched for the set of parameters that minimized $\Sigma\chi^2$ and Table 2 shows the best fit of the two experiments performed using each of the two models.

Model B fits both the experiments with $\Sigma\chi^2$ = 36.1 (22.52 when label is in glucose and 13.56 when label is in lactate). Both the experiments together provided 46 points. With 24 essential parameters, this problem has 22 degrees of freedom that corresponds to Q = 0.09, thus indicating that the model is acceptable. If this set of parameters which gives the best fit for model B is used for model A (where the same flux of glycogen synthesis is directed from the common pool of

**Table 2 Isotopologue distribution produced by isolated hepatocytes in the presence of glucose and lactate.**

| | Experiment 1 label in glucose | Simulations B | A | Experiment 2 label in lactate | Simulations B | A |
|---|---|---|---|---|---|---|
| **glucose:** | | **$\chi^2$ = 7.21** | **5.72** | | **$\chi^2$ = 5.52** | **0.576** |
| m0 | 0.532 ± 0.0098 | 0.514 | 0.514 | 0.979 ± 0.01 | 0.976 | 0.984 |
| m1 | 0.00846 ± 0.0022 | 0.0114 | 0.00918 | 0.0063 ± 0.0055 | 0.00484 | 0.00354 |
| m2 | 0.459 ± 0.0103 | 0.473 | 0.475 | 0.0055 ± 0.0064 | 0.00891 | 0.00469 |
| m3 | – – | – | – | 0.0064 ± 0.0016 | 0.00996 | 0.00628 |
| [mM] | 20.6 ± 2.91 | 21 | 21.1 | 20.9 ± 2.22 | 20.8 | 20.9 |
| **glycogen:** | | **$\chi^2$ = 6.46** | **27.3** | | **$\chi^2$ = 2.22** | **5.73** |
| m0 | 0.681 ± 0.032 | 0.683 | 0.57 | 0.909 ± 0.026 | 0.907 | 0.9 |
| m1 | 0.0119 ± 0.031 | 0.00767 | 0.0131 | 0.017 ± 0.0066 | 0.0166 | 0.0225 |
| m2 | 0.302 ± 0.031 | 0.308 | 0.409 | 0.038 ± 0.01 | 0.0313 | 0.0298 |
| m3 | 0.0017 ± 0.001 | 0.000136 | 0.00208 | 0.0273 ± 0.0071 | 0.0363 | 0.04 |
| m4 | 0.0032 ± 0.0016 | 0.000593 | 0.00565 | 0.0036 ± 0.0016 | 0.00329 | 0.00317 |
| m5 | – – | – | – | 0.003 ± 0.014 | 0.00335 | 0.00271 |
| mg/mL | 0.263 ± 0.084 | 0.256 | 0.196 | 0.262 ± 0.0691 | 0.256 | 0.196 |
| **gln14:** | | **$\chi^2$ = 5.31** | **19.2** | | **$\chi^2$ = 2.23** | **4.59** |
| m0 | 0.678 ± 0.032 | 0.692 | 0.588 | 0.93 ± 0.024 | 0.924 | 0.921 |
| m1 | 0.016 ± 0.0046 | 0.00561 | 0.0112 | 0.033 ± 0.009 | 0.0343 | 0.0472 |
| m2 | 0.3 ± 0.032 | 0.302 | 0.399 | 0.019 ± 0.0079 | 0.0178 | 0.0128 |
| m3 | – – | – | – | 0.014 ± 0.005 | 0.0209 | 0.0163 |
| m4 | – – | – | – | 0.004 ± 0.003 | 0.00267 | 0.0264 |
| **gln36:** | | **$\chi^2$ = 1.48** | **14.2** | | **$\chi^2$ = 1.97** | **4.48** |
| m0 | 0.98 ± 0.0101 | 0.987 | 0.968 | 0.924 ± 0.024 | 0.923 | 0.911 |
| m1 | 0.00408 ± 0.0018 | 0.0051 | 0.0101 | 0.0265 ± 0.007 | 0.0332 | 0.0348 |
| m2 | 0.0139 ± 0.0078 | 0.00766 | 0.216 | 0.027 ± 0.0081 | 0.0188 | 0.0219 |
| m3 | – – | – | – | 0.021 ± 0.0067 | 0.0221 | 0.0292 |
| **lactate:** | | **$\chi^2$ = 2.06** | **2.73** | | **$\chi^2$ = 1.62** | **9.76** |
| m0 | 0.974 ± 0.026 | 0.991 | 0.985 | 0.636 ± 0.017 | 0.621 | 0.608 |
| m1 | 0.0026 ± 0.0019 | 0.000974 | 0.00135 | 0.0166 ± 0.0025 | 0.0167 | 0.0172 |
| m2 | 0.0094 ± 0.0037 | 0.00773 | 0.0141 | 0.0318 ± 0.0035 | 0.0302 | 0.0245 |
| m3 | 0.00136 ± 0.023 | 0.00000227 | 0.0000436 | 0.316 ± 0.0213 | 0.332 | 0.35 |
| [mM] | 6.18 ± 0.75 | 6.8 | 6.73 | 3.18 ± 0.43 | 6.29 | 6.22 |
| **$\Sigma\chi^2$** | | **22.52** | **69.15** | | **13.56** | **25.136** |

Before incubation the medium contained either 50% of [1,2-13C2]D-glucose and unlabeled lactate (experiment 1) or 50% uniformly 13C-labeled lactate and unlabeled glucose (experiment 2). The measurements are presented as mean ± standard deviation. The data were fit by two models (A and B). The conditions of incubation and measurements, and data fitting are described in Methods. The difference between the best fit and experimental data ($\chi 2$, see Methods) are shown for each metabolite and summarized for the whole set of data.

hexoses), $\Sigma\chi^2$ increases up to 331, where glycogen gave the greatest $\Sigma\chi^2$. The fitting procedure reduced $\Sigma\chi^2$ for model A down to 94.286 (69.15 when label was in glucose and 25.13 when label was in lactate) with the isotopologue distribution shown in Table 2. For this best fit, the same number of experimental points (46), and 24 essential parameters (22 degrees of freedom) give $Q = 6.325 \cdot 10^{-11}$, which indicates that model A is incorrect.

### Metabolic flux distribution

If model discrimination analysis indicates that a model should be rejected, as in the case of model A, the distribution of metabolic fluxes obtained with such a model cannot be reliable. In contrary, if the analysis suggests accepting a model, as in the case of model B, there is much more confidence that it evaluates true metabolic fluxes. Therefore we analyzed the distribution of metabolite fluxes computed with model B. Table 3 shows

**Table 3 Metabolic fluxes corresponding to the best fit of experimental data and their 99% confidence intervals**

| | Glucose as the only substrate 99% confidence interval bestfit (min - max) | Glucose with lactate 99% confidence interval bestfit (min - max) | Model A |
|---|---|---|---|
| **hk1** | 0.0026894(0.002150-0.003080) | 0.152266(0.074815-0.377413) | 0.01981 |
| **hk2** | 0.0021436(0.001710-0.002480) | 0.0056939(0.0041563-0.0076078) | |
| **g6pase1** | 5.50E-05(2.70E-5-7.80E-5) | 0.150611(0.073125-0.375565) | 0.01809 |
| **g6pase2** | 2.13E-05(0.0-0.000066) | 0.005166(0.0029064-0.0071982) | |
| **pfk1** | 0.0030480(0.002370-0.003530) | 0.0023558(0.0015063-0.0046942) | 0.00345 |
| **pfk2** | 0.0004816(0.0-0.000950) | 0.0007975(0.0006197-0.0012778) | |
| **fbpase1** | 0.0004446(0.000270-0.000560) | 0.0013469(0.0005781-0.0023709) | 0.00363 |
| **fbpase2** | 0.0005496(0.000330-0.000690) | 0.0023329(0.00161-0.0032427) | |
| **gp** | 2.56E-06(1.70E-6-5.10E-6) | 0.000143(5.12E-05-0.0001883) | 0.00000 |
| **gs** | 0.0021929(0.001660-0.002680) | 0.0022087(0.0018593-0.0027097) | 0.00188 |
| **aldf** | 0.0192413(0.007480-0.022580) | 0.0159083(0.0153142-0.0163931) | 0.01608 |
| **aldr** | 0.016706(0.004480-0.020410) | 0.0164349(0.0153054-0.016708) | 0.01625 |
| **aldex** | 0.0424386(0.010420-0.056150) | 0.0137425(0.0103922-0.0195718) | 0.01565 |
| **g3pep** | 0.0064103(0.004490-0.007700) | 0.258278(0.1677705-0.3394195) | 0.12224 |
| **pepg3** | 0.0013391(0.000140-0.002370) | 0.258701(0.166264-0.339572) | 0.12257 |
| **pk** | 0.0050704(0.003910-0.006230) | 0.0199774(0.0157079-0.0239355) | 0.01497 |
| **lacin** | 1.71E-07(1.20E-7-2.60E-7) | 0.231716(0.1714965-0.2686475) | 0.12266 |
| **lacout** | 0.00507(0.003910-0.006230) | 0.222326(0.1621185-0.2590095) | 0.11074 |
| **pc** | 1.20E-07(5.70E-8-1.90E-7) | 0.0204457(0.0155337-0.0233495) | 0.01542 |
| **pepck** | 1.16E-08(5.30E-9-2.80E-8) | 0.020401(0.0154933-0.0232863) | 0.01530 |
| **maloa** | 1.85E-07(9.60E-8-3.00E-7) | 0.0835934(0.048937-0.111892) | 0.03396 |
| **oamal** | 3.80E-08(1.30E-8-8.10E-8) | 0.0747156(0.0368199-0.0997713) | 0.02261 |
| **cs** | 2.55E-07(1.30E-7-4.00E-7) | 0.008922(0.0069631-0.0147641) | 0.01147 |
| **citmal** | 1.47E-07(8.40E-8-2.30E-7) | 0.0088778(0.0069227-0.0146991) | 0.01135 |
| **pdh** | 2.75E-07(7.70E-8-5.00E-7) | 0.0089212(0.0069623-0.0147633) | 0.01147 |
| **g6pdh** | 3.87E-06(2.80E-6-7.50E-6) | 0.0018986(0.0013506-0.0021492) | 0.00000 |
| **p5p > s7p** | 0.0011849(0.000880-0.002270) | 0.0006229(0.0004459-0.0007215) | 0.00051 |
| **s7p > r5p** | 0.0011925(0.000890-0.002280) | 4.35E-06(2.56E-06-2.98E-05) | 0.00047 |
| **f6p > p5p** | 9.59E-06(4.20E-6-3.40E-5) | 4.02E-05(1.17E-05-9.84E-05) | 0.00000 |
| **p5p > f6p** | 4.93E-06(2.80E-6-2.80E-5) | 0.0006793(0.0004937-0.000751) | 0.00000 |
| **f6p > s7p** | 1.85E-05(9.20E-6-7.30E-5) | 1.40E-05(4.56E-06-2.45E-05) | 0.00000 |
| **s7p > f6p** | 9.56E-06(4.80E-6-6.60E-5) | 1.65E-06(9.03E-07-8.56E-06) | 0.00000 |
| **p5p-g3p** | 0.0006152(2.70E-4-2.83E-3) | 0.0017868(0.0007706-0.0033249) | 0.00036 |
| **f6p-s7p** | 7.69E-08(2.10E-8-7.70E-7) | 1.53E-05(4.62E-06-2.78E-05) | 0.00000 |
| **p5p-s7p** | 0.0022967(1.37E-3-7.06E-3) | 1.51E-06(8.37E-07-8.16E-06) | 0.00067 |
| **f6p > s7p** | 0.0015144(0.000850-0.001940) | 0.000862(0.0002796-0.0021263) | 0.00117 |
| **s7p-f6p** | 0.0015015(0.000850-0.001920) | 0.0014894(0.0008194-0.0026708) | 0.00119 |
| **f6p-g3p** | 0.0075338(0.002170-0.013330) | 0.0164459(0.0085357-0.0352201) | 0.00230 |
| **s7p-e4p** | 0.0003018(0.000140-0.000550) | 7.81E-05(2.01E-05-0.0002102) | 0.00061 |

The names of fluxes are given in the list of abbreviations in the text.

metabolic fluxes corresponding to the best fit of the measured data by model B, together with their 99% confidence intervals. The metabolic fluxes of the two experiments, performed in the presence of lactate using two different substrate labelings, were the same (Table 2). Thus, Table 3 shows two sets of fluxes with their confidence intervals evaluated by model B. One of the sets corresponds to the best fit of isotopologue distribution measured only in the presence of glucose (as shown in Table 1), and the other in the presence of glucose and lactate (as shown in Table 2).

At first glance, a notable difference is seen between metabolic fluxes under the two different conditions. The fluxes for the best fit indicate that the presence of lactate had perturbed the entire central carbohydrate metabolism of hepatocytes. Without lactate, almost half of the glucose consumed(hk) was used to replenish the glycogen store (glgsn) that was exhausted during starvation, and the rest was mainly converted to lactate except a small part that was burned in TCA cycle. Although net consumption of glucose did not change much by the addition of lactate, the fluxes of glucose input (hk) and output (g6ph) taken separately are increased by almost two orders of magnitude. Thus, recycling of metabolites increased without affecting the net influx of glucose. The addition of lactate increased recycling in many other points downstream of glucose entrance. This refers to the flux through fructose bisphosphatase (fbpase), which forms a futile cycle with phosphofructokinase (pfk). The increase of flux transforming glyceraldehyde 3-phosphate into phosphoenolpyruvate (g3pep, it lumped a set of reactions) is accompanied by the increase of reactions in the reverse direction (pepg3). Essentially there is an increased futile recycling through pyruvate kinase (pk), pyruvate carboxylase (pc) and phosphoenolpyruvate carboxykinase (pepck). This recycling is accompanied by an increase of flux through the TCA cycle (pdh, cs, citmal) linked with enhancing of energy production. Some changes took place in the pentose phosphate pathway, but they were not as impressive as in glycolysis and TCA cycle.

The confidence intervals for some of the fluxes (e.g. hkI) were large. However, many intervals for the two studied conditions do not overlap, and so the changes described above for the best fit remain qualitatively the same for the whole intervals.

For each fit (as it can be seen in best fit fluxes presented in Table 3) the ratio of the two inputs is different for the two hexose phosphate pools. The hkII is less than hkI, and, in contrary, the contribution of fbpaseII is higher than fbpaseI. This results in a different isotopomer distribution in glycogen relative to the glucose in the medium. Moreover, the reactions of the pentose phosphate pathway that interchange various sugar phosphates with fructose phosphates introduce additional differences to the isotopomer content between glycogen and hexose phosphates fueling glycolysis.

Among the fluxes of pentose phosphate pathway, the most essential are the exchange between triose and pentose phosphates, and fructose and sedoheptulose phosphates. These exchanges also contribute to the difference between isotopomer content of glycogen and hexose phosphates fueling glycolysis.

Although model A was rejected, the fluxes corresponding to the best fit in the presence of lactate are shown in the last column of Table 3. They are different from those computed with model B. This difference shows the possible error in the results if the applied model does not properly account for the compartmentation of metabolites.

## Discussion

### Possible sources of errors and the implemented way of avoiding them

Stable isotope tracing is a promising sensitive technique for the study of metabolism in living cells. However, it is sensitive to various flaws and incompleteness in the data analysis. That is why tools for tracer data analysis must take into account the possible sources of errors. In particular, omitting some isotope exchange reactions may lead to significant errors in the calculated flux distributions [17]. The model used here simplifies some sequential reactions of the pentose phosphate pathways, TCA cycle, and glycolysis, which are grouped together. However, all the possible splitting and re-formations of carbon skeleton in the considered pathways are taken into account [30,31]. This gives a confidence in avoiding potential errors related with the simulation of incomplete set of fluxes.

Not accounting for compartmentation due to metabolic channeling is another pitfall that can result in incorrect estimation of metabolic fluxes [17]. Table 3 shows that model A, which does not account for channeling, gives remarkably different results. However, based on the current state of the art in fluxomics, it is difficult to include into consideration all the possible microcompartments *a priori*. Although there are studies that confirm the associations of enzymes and channeling [18-24], it is not clear how general the studied cases are with regard to various organisms and tissues, and whether the metabolite compartmentation studied *in vitro* still persist *in vivo*. In modeling the metabolic networks, considering a single well mixed pool for each metabolite still remains commonly accepted (e.g. [32]). The method proposed here to determine metabolite compartmentation from $^{13}C$ distribution in metabolites does not require specific experiments. Instead, it requires a specific analysis related with the

The net glucose consumed as well as lactate produced are burned through the TCA cycle thus producing energy necessary for the recycling. Thus, in the presence of only glucose, its essential part is used to replenish glycogen, whereas in the presence of both glucose and lactate the cultured liver cells apparently burns these substrates (preferentially lactate).

The presented analysis of the entire set of experiments characterized the capacity of hepatocytes to modify metabolic state under extreme conditions. The characterization of metabolism of hepatocytes is inseparable from the detection of the real compartmentation of considered pathways. Application of this methodology to larger datasets will reveal new information about the network topology. It opens a perspective to examine the compartmentation and metabolic flux profile in various cells under physiological and patho-physiological conditions.

## Conclusions

Compartmentation of intracellular metabolism, appeared as a general phenomenon, results that the analysis of metabolic flux distribution should be inseparable from the analysis of compartmental structure of studied pathways. Here we proposed a methodology implemented in our software to reveal compartmental structure and metabolic flux distribution from the distribution of $^{13}C$ isotopomers measured in the products of cells incubated with $^{13}C$ labeled substrates. This methodology is based on varying the schemes for simulation of measured data and applying the model discrimination analysis. The application of this methodology to the analysis of $^{13}C$ isotopomer distributions measured in metabolites of isolated liver cells revealed a separate compartment of hexose phosphates related with substrate channeling in glycogen metabolism. This analysis provided the distribution of metabolic fluxes in central carbohydrate metabolism of the cells incubated with $^{13}C$ labeled glucose, and revealed the changes of fluxes that were induced by addition of lactate in the incubation media.

## Methods

To analyze cellular metabolic flux profiles for specific conditions *in situ* and various schemes of metabolic reactions we used the software tool "Isodyn" (from "isotopomer dynamics") [11,12]. It simulates isotopomer distributions in the same way as classical kinetic models simulate the time-course of metabolite concentrations.

### Models and data fitting

The systems of differential equations corresponding to the schemes presented in Figure 1 and expressions for the rates of individual reactions are given in Additional File 1. The metabolite fluxes and concentrations were obtained as a numerical solution of the differential equations using the BDF method implemented in our software Isodyn. The software then uses these values of total concentrations and fluxes to construct and solve differential equations for all isotopomers of metabolites presented in Figure 1. The algorithms for constructing equations for isotopomers are described elsewhere [11,12].

The concentrations of isotopologues needed to be compared with experimental data were calculated as a sum of the respective isotopomer concentrations computed by Isodyn. Fitting of the experimental data was performed by minimizing $\chi^2$, the square of deviations between measured isotopologue fractions ($y_i$) and values ($y(x_i, a)$) computed for the set of parameters, a, as fractions of isotopologues $x_i$, normalized by experimental standard deviations ($\sigma_i$):

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - y(x_i; a)}{\sigma_i}\right]^2$$

The minimization was performed in the global space of parameters using our implementation of simulated annealing algorithm supplemented by coordinate descent in local area [5]. The sets of parameters, thus optimized, defined the sets of metabolic fluxes resulting in simulated isotopologue distributions and the value of $\chi^2$. Multiple application of the optimization resulted in multiple sets of fluxes characterized by different values of $\chi^2$. Application of $\chi^2$ threshold to the obtained sets of fluxes [29] defined confidence intervals for the metabolic fluxes. The implementation of this procedure in Isodyn was described elsewhere [5].

### Modification of reaction scheme

A change of equations of the basic kinetic model could be usually performed easily, even graphically: there are algorithms that transform a drawn scheme into a system of ordinary differential equations (ODE) [46]. However, in the case of Isodyn the removal or addition of reactions in the basic kinetic model must be followed by the respective changes in calculation of the isotopomer distribution. The consistent change of both the modules (basic kinetic model and calculation of respective isotopomer distributions) was performed automatically and the sources of respective programs are available for free at http://www.bq.ub.es/bioqint/label_distribution/tutorial.tar.gz..

### Aldolase reaction

An essential restriction, which helps to distinguish between models, is including the interdependency of fluxes catalyzed by the same enzyme, as we have shown for transketolase and transaldolase [12]. Here, similar

interdependency of various isotope-exchange fluxes is considered for the aldolase-catalyzed reaction [30], which normally are not included in classical kinetic models.

The scheme in Figure 2 shows the possible isotope-exchange fluxes in the aldolase reaction, accounted for in the model. The flux, shown in Figure 2A with green lines, transforms the whole molecule of fbp in the pool of trioses. It constitutes only a part of the flux $v_3$, because another part of $v_3$ produces dhap originated not from fbp, but from the same pool of trioses, bound to the enzyme through the reaction $v_{-3}$. The steady state fraction of $v_3$ that produces dhap originated from fbp (that equals to the fraction of bound dhap originated from fbp (PfE-dhap)) can be expressed as the ratio of input of molecules originated from fbp to the total input in E-dhap:

$$P^f_{E-dhap} = \frac{v_2 . P^f_{E-fbp}}{v_2 + v_{-3}} \quad (1)$$

Here $Pf_{E\text{-}fbp}$ is the fraction of $v_2$, which brings the carbons originated from fbp to E-dhap (or the fraction of bound fbp originated from fbp). Another fraction of $v_2$, $(1\text{-}P^f_{E\text{-}fbp})$ brings carbons originated from triose pool, which were bound through reactions $v_{-3}$ and $v_{-2}$. $Pf_{E\text{-}fbp}$, in turn, can be expressed as the ratio of input of molecules originated from fbp to total input in E-fbp:

$$P^f_{E-fbp} = \frac{v_{-2} . P^f_{E-dhap} + v_1}{v_1 + v_{-2}} \quad (2)$$

The solution of equations (1) and (2) is

$$P^f_{E-dhap} = \frac{v_2 . v_1}{v_2 . v_1 + v_{-3} . v_1 + v_{-3} . v_{-2}} \quad (3)$$

where the rates vi could be expressed through the rate constants and substrate concentrations. The forward flux through the whole cycle (indicated by green lines in Figure 2A) is

$$v_f = P^f_{E-dhap} . v_3 = \frac{v_1 . v_2 . v_3}{v_2 . v_1 + v_{-3} . v_1 + v_{-3} . v_{-2}} \quad (4)$$

The reverse flux of fbp formation (red lines in Figure 2A) from triose phosphates ($P^t_{E\text{-}fbp}$) could be described similar to (4):

$$v_r = P^t_{E-fbp} . v_{-1} = \frac{v_{-1} . v_{-2} . v_{-3}}{v_2 . v_1 + v_{-3} . v_1 + v_{-3} . v_{-2}} \quad (5)$$

Figure 2B shows two additional fluxes, which only exchange half a molecule of fbp with g3p pool. As in

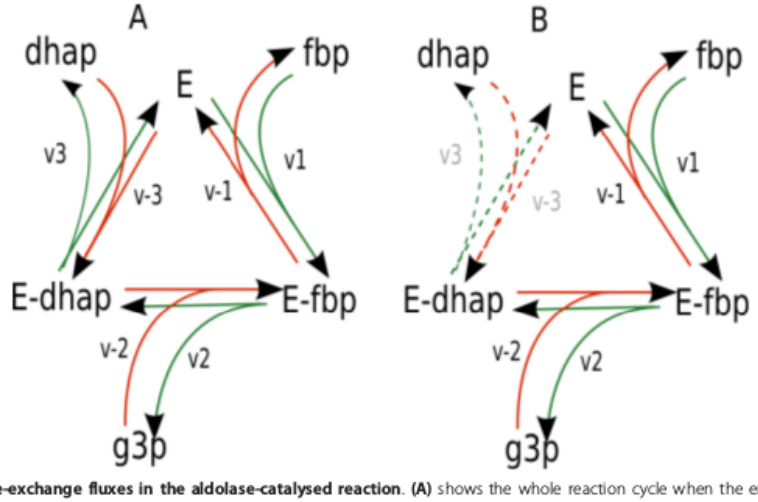



**Figure 2 Isotope-exchange fluxes in the aldolase-catalysed reaction. (A)** shows the whole reaction cycle when the enzyme (E) forms a complex with fbp (E-fbp), releases g3p keeping dhap (E-dhap), and finally releases dhap returning to (E). Forward flux (green lines) through the whole reaction cycle brings isotopes originated from the fbp pool into the pools of dhap and g3p, and reverse flux (red lines) brings isotopes originated from the pools of dhap and g3p into fbp pool. **(B)** shows forward (green lines) and reverse (red lines) fluxes that only exchange isotopes of upper part of fbp molecule with g3p pool without releasing dhap. $v_i$ designate the respective rates of elementary steps of reaction mechanism.

the cases described above, the fraction of the first three carbons originated from fbp ($P^{fg}_{E\text{-}dhap}$) in (E-fbp) is:

$$P^{fg}_{E-fbp} = \frac{v_1}{v_1 + v_{-2}} \quad (6)$$

and the forward (thin black in Figure 2B) flux:

$$v_{fg} = \frac{v_1 . v_2}{v_1 + v_{-2}} \quad (7)$$

The flux in the opposite direction (indicated by thick gray lines) is described likewise.

Thus, the model accounts for three isotope-exchange fluxes related with aldolase activity: forward and reverse flux through the whole cycle of enzyme reaction, and pure isotope exchange flux between f6p and g3p, without the change of total concentrations of these metabolites. They are not used in classical kinetic simulations, where only the net flux is important, but they are necessary for the subsequent simulation of isotopologue distribution. The isotope-exchange fluxes of transketolase and transaldolase were implemented similarly as described elsewhere [12].

### $\chi^2$ criterion for the acceptance or rejection of model

To analyze the structure of metabolic networks, model discrimination analysis was used to test various kinetic models of the same pathways and reject the ones inconsistent with experimental data. Isodyn implements criteria for acceptance or rejection of a model based on the values of normalized square of difference between experimental data and simulation ($\chi^2$) and numbers of degrees of freedom [29].

The fitting algorithm implemented in Isodyn identifies the global minimum for the function $\chi^2$ and the respective set of parameters and fluxes. If the model is acceptable, the estimated fluxes are also acceptable as a model prediction. The value of $\chi^2$ is used in Isodyn as a criterion for acceptance or rejection of model as it is described in [29]. The probability that a model with F degrees of freedom is correct and $\chi^2$ by chance could exceed a determined value, is given as an incomplete gamma function (Q(a,x), where a = F/2 (F is the number of degrees of freedom), and x = $\chi^2$/2):

$$Q(a,x) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt \quad (8)$$

Here Γ(a) is gamma function:

$$\Gamma(z) = \int t^{z-1} e^{-t} dt \quad (9)$$

The model is acceptable if Q value is larger than 0.05. It can be acceptable even with Q value larger than 0.001, if the errors are not normal or have been moderately underestimated. But if the Q value is lower than 0.001, the model must be rejected as inconsistent with experimental data.

### Estimation of number of degrees of freedom

Formally, the number of degrees of freedom (F) is calculated as the difference between number of data points (N, which in our case is the number of fractions of isotopologues for all measured metabolites and total metabolite concentrations) and parameters (P) in the model:

$$F = N - P \quad (10)$$

However, in the case if the model is underdetermined, it could happen that the fit of the given data is insensitive to some parameters or there are ambiguous combinations equally affecting the fit, so that the parameters could not be distinguished. The presence of such parameters does not improve the fit and thus do not decrease the number of degrees of freedom. Both situations result in the fact that the Hessian matrix (the matrix of second derivatives of objective function $\chi^2$ with respect to $[\alpha_{kl}] = [\frac{\partial^2 \chi^2}{\partial a_k \partial a_l}]$ sparameters,) is singular, or numerically close to singularity. In such situations the real number of degrees of freedom is higher than the formal difference (10). The maximal set of parameters that could be defined by the given experimental data could be estimated based on the analysis of Hessian matrix.

This matrix is calculated as follows. First derivative of $\chi^2$ with respect to the parameters is:

$$\frac{\partial \chi^2}{\partial \alpha_k} = -2 \sum_{i=1}^{N} \frac{[y_i - \gamma(x_i; a)]}{\sigma_i^2} \frac{\partial \gamma(x_i; a)}{\partial \alpha_k} \quad (11)$$

where k = 1,2,...M (number of parameters)
Differentiation of these functions gives

$$\frac{\partial^2 \chi^2}{\partial \alpha_k \partial \alpha_l} = -2 \sum_{i=1}^{N} \frac{1}{\sigma_i^2} [a - b] \quad (12)$$

Where:

$$a = \frac{\partial \gamma(x_i; a)}{\partial \alpha_k} \frac{\partial \gamma(x_i; a)}{\partial \alpha_l}$$

$$b = [y_i - \gamma(x_i; a)] \frac{\partial^2 \gamma(x_i; a)}{\partial \alpha_k \partial \alpha_l}$$

The second term under the sum, which contains the second derivative of fitting function y(xi,a) is usually ignored [27]:

$$\alpha_{kl} = -2\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left[\frac{\partial y(x_i;a)}{\partial a_k}\frac{\partial y(x_i;a)}{\partial a_l}\right] \quad (13)$$

As it is indicated above, the singularity of this matrix means that some parameters of the model could not be defined in principle, given the specific dataset analyzed. To find out what characteristics of the studied system (parameters of the model) could be revealed, given the model and a specific dataset, the standard procedure of singular value decomposition of Hessian matrix was used [29].

Following this standard routine the Hessian matrix A is decomposed to the product of orthogonal matrix U, the vector (or diagonal matrix) of singular values W, and the orthogonal matrix V. The ratio of maximal and minimal values in the vector W, called condition number, characterizes the singularity of A. If some values of W are zeros, matrix A is strictly singular, but it could be numerically close to singularity, or ill-conditioned, if the condition number is close to machine precision. The covariance matrix C = A-1 was found from singular value decomposition as C = (V·W-1·UT), where W-1 = [diag(1/wi)]. Diagonal elements of C are the variances of parameters and the other elements are covariances. If matrix A is ill-conditioned, its inverse C cannot be defined and the failure of finding the inverse indicates that the number of parameters is excessive.

Isodyn finds the maximal set of parameters of the model, which, being considered as a subject for fitting, give a non-singular Hessian matrix. The size of this set could be considered as the number of parameters, which affects the number of degrees of freedom in the model with regard to given experimental data. If in fact the model has more parameters, the other parameters are not distinguishable by the given experimental data and must be considered as constants. The number of degrees of freedom (F) is defined as a difference between the numbers of experimental data and effective model parameters, and the value of incomplete gamma function Q(F/2, χ2/2), defined as described in [29], indicates the acceptance of the model.

## Experimental methods

### Materials

[1,2-$^{13}C_2$]D-glucose (> 99% enriched) and [U-$^{13}C_3$]L-lactate (> 99% enriched) were purchased from Isotec (Miamisburg, OH), and other reagents from Sigma-Aldrich Company (St. Louis, MO).

### Animals

180-200 g male Wistar rats were used. They were maintained in a 12h:12h light-dark cycle with free access to standard laboratory rat chow pellets (Panlab) and water. Animals were deprived of food 24 hours prior to hepatocyte isolation. Experiments were conducted according to guidelines accepted by the University Animal Care and Use Committee. Appropiate measures were taken to minimize pain or discomfort of animals.

### Preparation of cells and incubation

Suspensions of isolated parenchymal liver cells were prepared from 24-h starved animals as previously described [47]. Cells were resuspended in Krebs-Ringer bicarbonate buffer, pH 7.4. Preparations with viability below 90%, established by the trypan blue exclusion method, were not used. Samples (6 ml) of these suspensions, containing 2.3 × $10^6$ cells/ml, were incubated at 37°C with gassing and continuous shaking (160 strokes/min, which is the minimum shaking that assures total suspension of cells) for 2 h, as it is the optimum time to ensure maximum glycogen synthesis without diminishing cell viability. Conditions for cell incubation were: a) 20 mM glucose, with glucose enriched 50% in [1,2-$^{13}C_2$]-glucose, b) 20 mM glucose + 9 mM lactate + 1 mM pyruvate (10 mM lactate/pyruvate (9:1)), with glucose enriched 50% in [1,2-$^{13}C_2$]-glucose and c) 20 mM glucose + 10 mM lactate/pyruvate (9:1), with lactate enriched 50% in [U-$^{13}C_3$]-lactate.

### Measurement of metabolites

At the beginning and end of incubations, cells were centrifuged (3000 g, 20 s), and incubation medium and cell pellets were obtained. For glycogen determination, cell pellets were immediately homogenized with 30% (w/v) KOH using a modification of Chan et al. methodology [48], where we have used 3 mM paper to precipitate glycogen. Glucose and lactate incubation medium concentrations were determined using spectrophotometric methods as described in [49,50].

### Gas Chromatography/Mass Spectrometry sample processing and analysis

At the end of incubations, cells were centrifuged so that the incubation medium and cell pellet are separated and everything was frozen in liquid nitrogen and stored at -80°C until processing for GC/MS analysis. Incubation media were processed for isolation of lactate, glucose, and glutamate using previously established methods [51,52]. Glycogen was isolated from cell pellets after ethanol precipitation of glycogen over 3 mM paper, and then treated with

amyloglucosidase, and the hydrolyzed glucose was isolated using ion exchange chromatography [6]. Immediately after that, glucose from the medium or from hydrolyzed glycogen, as well as lactate and glutamate were derivatized for GC/MS analysis [51,53,54]. A mass selective detector HP 5973 equipment coupled to a gas chromatograph HP 6890 was used for all the metabolites as described elsewhere [51,53,54]. Chemical ionization was used to give the molecular ion (C1-C6) of the glycogen or medium glucose molecules at m/z 328, and the same for the lactate molecule (C1-C3) at m/z 328. Electron impact ionization was used to characterize isotopologues of C1-C4 (m/z 242) and C3-C6 (m/z 187) glycogen glucose fragments, as well as C2-C4 (m/z 152) and C2-C5 (m/z 198) glutamate fragments.

Results of the mass isotopologues in glucose, lactate and glutamate are reported as molar fractions of m0, m1, m2, etc, where m0, m1, m2... indicate the number of $^{13}C$ atoms in the molecule [55]. The data for each independent experiment obtained after subtraction of natural $^{13}C$ isotope enrichment are presented in Additional File 2.

## Additional material

**Additional file 1: Differential equations of the used kinetic models.** Kinetic models were used to simulate the total fluxes and concentrations of metabolites. Based on these calculated total values Isodyn further simulates the distribution of isotopic isomers.

**Additional file 2: Measured distributions of $^{13}C$ isotopomers in metabolites.** The table presents the data for each independent experiment obtained as described in Methods after subtraction natural $^{13}C$ isotope enrichment.

### List of abbreviations

Metabolites: glc: glucose; glu: glutamate; lac: lactate; glgn: glycogen; g6p: glucose 6 phosphate; f6p: fructose 6-phosphate; fbp: fructose 1,6-bisphosphate; dhap: dihydroxyacetone phosphate; g3p: glyceraldehyde 3-phosphate; pep: phosphoenolpyruvate; pyr: pyruvate; accoa: acetyl coenzyme A; e4p: erythrose 4-phosphate; s7p: sedoheptulose 7-phosphate; r5p: ribose 5-phosphate; xu5p: xylulose 5-phosphate; mal: malate; oaa: oxaloacetic acid; cit: citrate. Enzymes: g6pase: glucose 6-phosphatase; gs: glycogen synthase; gp: glycogen phosphorylase; hk1 & hk2: hexokinase; fbp1 and fbp2: fructose 1,6-bisphosphatase; pfk1 and pfk2: phosphofructokinase; ald: aldolase; tk: transketolase; ta: transaldolase; g6pdh: glucose 6-phosphate dehydrogenase; pck: phosphoenolpyruvate carboxykinase; pk: pyruvate kinase; pdh: pyruvate dehydrogenase complex; cs: citrate synthase; pc: pyruvate carboxylase.

### Acknowledgements and Funding

This work was supported by the European Commission Seventh Framework Programme FP7 (Diaprepp Health-F2-2008-202013, Etherpaths KBBE-grant n° 222639, Synergy-COPD project grant agreement n° 270086); the Spanish Government and the European Union FEDER funds (SAF2011-25726) and Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación of Spanish Government & European Regional Development Fund (ERDF) "Una manera de hacer Europa" ISCIII-RTICC (RD06/0020/0046); Generalitat de Catalunya-AGAUR, (2009SGR1308 and 2009 CTP 00026); Foundation Marato TV3-042010.

MC acknowledges the support received through the prize "ICREA Academia" for excellence in research, funded by ICREA foundation-Generalitat de Catalunya.

Authors thank Dr J.J. Guinovart and A. Adrover (Universitat de Barcelona and Institute for Research of Barcelona, Spain) for their financial and technical support in hepatocyte incubation, Dr W.N. P. Lee (Harbor-UCLA Medical Center, CA, USA) for his help in GC/MS analysis, and Anusha Jayaraman for the important assistance in the preparation of manuscript.

### Author details

[1]Department of Biochemistry and Molecular Biology, Faculty of Biology, Universitat de Barcelona, Av Diagonal 643, 08028 Barcelona, Spain. [2]Institute of Biomedicine of Universitat de Barcelona (IBUB) and CSIC-Associated Unit, Spain. [3]A.N.Belozersky Institute of Physico-Chemical Biology, MSU, Moscow 199899, Russia. [4]Hospital Clínic, IDIBAPS, CIBERES; Universitat de Barcelona, Barcelona 08028, Spain. [5]Technical Research Centre of Finland, Espoo, and Institute for Molecular Medicine, Helsinki, Finland. [6]Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle, UK.

### Authors' contributions

IMM performed the analysis and wrote the paper, VAS developed the algorithms and wrote the paper, SM made the experiments and analyzed the data, JR analyzed the data, MO analyzed the data, LA analyzed the data and wrote the paper, MC analyzed the data and wrote the paper. All authors read and approved the final manuscript.

Received: 16 April 2011 Accepted: 28 October 2011
Published: 28 October 2011

### References

1. Irving CS, Wong WW, Shulman RJ, Smith EO, Klein PD: **[13C]bicarbonate kinetics in humans: intra- vs interindividual variations.** *Am J Physiol* 1983, **245**:R190-202.
2. Chance EM, Seeholzer SH, Kobayashi K, Williamson JR: **Mathematical analysis of isotope labeling in the citric acid cycle with applications to 13C NMR studies in perfused rat hearts.** *J Biol Chem* 1983, **258**:13785-94.
3. Fan TWM, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, Miller DM: **Altered regulation of metabolic pathways in human lung cancer discerned by (13)C stable isotope-resolved metabolomics (SIRM).** *Mol Cancer* 2009, **8**:41.
4. Amaral AI, Teixeira AP, Martens S, Bernal V, Sousa MF, Alves PM: **Metabolic alterations induced by ischemia in primary cultures of astrocytes: merging 13C NMR spectroscopy and metabolic flux analysis.** *J Neurochem* 2010, **113**:735-748.
5. Selivanov VA, Vizán P, Mollinedo F, Fan TW, Lee PW, Cascante M: **Edelfosine-induced metabolic changes in cancer cells that precede the overproduction of reactive oxygen species and apoptosis.** *BMC Syst Biol* 2010, **4**:135.
6. Jouhten P, Pitkänen E, Pakula T, Saloheimo M, Penttilä M, Maaheimo H: **13C-metabolic flux ratio and novel carbon path analyses confirmed that Trichoderma reesei uses primarily the respirative pathway also on the preferred carbon source glucose.** *BMC Syst Biol* 2009, **3**:104.
7. Jouhten P, Rintala E, Huuskonen A, Tamminen A, Toivari M, Wiebe M, Ruohonen L, Penttilä M, Maaheimo H: **Oxygen dependence of metabolic fluxes and energy generation of Saccharomyces cerevisiae CEN.PK113-1A.** *BMC Syst Biol* 2008, **2**:60.
8. Aboka FO, Heijnen JJ, van Winden WA: **Dynamic 13C-tracer study of storage carbohydrate pools in aerobic glucose-limited Saccharomyces cerevisiae confirms a rapid steady-state turnover and fast mobilization during a modest stepup in the glucose uptake rate.** *FEMS Yeast Res* 2009, **9**:191-201.
9. Fendt SM, Sauer U: **Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates.** *BMC Syst Biol* 2010, **4**:12.
10. Wiechert W: **13C metabolic flux analysis.** *Metab Eng* 2001, **3**:195-206.
11. Selivanov VA, Puigjaner J, Sillero A, Centelles JJ, Ramos-Montoya A, Lee PW, Cascante M: **An optimized algorithm for flux estimation from isotopomer distribution in glucose metabolites.** *Bioinformatics* 2004, **20**:3387-3397.
12. Selivanov VA, Meshalkina LE, Solovjeva ON, Kuchel PW, Ramos-Montoya A, Kochetov GA, Lee PWN, Cascante M: **Rapid simulation and analysis of isotopomer distributions using constraints based on enzyme**

mechanisms: an example from HT29 cancer cells. *Bioinformatics* 2005, **21**:3558-3564.
13. Schaub J, Mauch K, Reuss M: **Metabolic flux analysis in Escherichia coli by integrating isotopic dynamic and isotopic stationary 13C labeling data.** *Biotechnol Bioeng* 2008, , **99**: 1170-85.
14. Wahl SA, Nöh K, Wiechert W: **13C labeling experiments at metabolic nonstationary conditions: an exploratory study.** *BMC Bioinformatics* 2008, **9**:152.
15. Selivanov VA, Votyakova TV, Pivtoraiko VN, Zeak J, Sukhomlin T, Trucco M, Roca J, Cascante M: **Reactive oxygen species production by forward and reverse electron fluxes in the mitochondrial respiratory chain.** *PLoS Comput Biol* 2011, **7**:e1001115.
16. Zamboni N, Sauer U: **Novel biological insights through metabolomics and 13C-flux analysis.** *Curr Opin Microbiol* 2009, **12**:553-558.
17. van Winden W, Verheijen P, Heijnen S: **Possible pitfalls of flux calculations based on (13)C-labeling.** *Metab Eng* 2001, **3**:151-162.
18. Cornish-Bowden A, Cardenas ML: **Channeling can affect concentrations of metabolic intermediates at constant net flux - artifact or reality.** *Eur J Biochem* 1993, **213**:87-92.
19. Mendes P, Kell DB, Westerhoff HV: **Channelling can decrease pool size.** *Eur J Biochem* 1992, **204**:257-266.
20. Atkinson D: **Conservation of solvent capacity. Cellular Energy Metabolism and its Regulation.** *Academic Press NY* 1977, 13-31.
21. Meyer FM, Gerwig J, Hammer E, Herzberg C, Commichau FM, Völker U, Stülke J: **Physical interactions between tricarboxylic acid cycle enzymes in Bacillus subtilis: evidence for a metabolon.** *Metab Eng* 2011, **13**:18-27.
22. Hardin CD, Finder DR: **Glycolytic flux in permeabilized freshly isolated vascular smooth muscle cells.** *Am J Physiol* 1998, **274**:C88-C96.
23. Shearer G, Lee JC, Koo J, Kohl DH: **Quantitative estimation of channeling from early glycolytic intermediates to CO in intact Escherichia coli.** *FEBS J* 2005, **272**:3260-3269.
24. Cascante M, Centelles JJ, Agius L: **Use of alpha-toxin from Staphylococcus aureus to test for channelling of intermediates of glycolysis between glucokinase and aldolase in hepatocytes.** *Biochem J* 2000, **352**:899-905.
25. Huang X, Holden HM, Raushel FM: **Channeling of substrates and intermediates in enzyme-catalyzed reactions.** *Annual Review of Biochemistry* 2001, **70**:149-180.
26. Kleijn RJ, van Winden WA, van Gulik WM, Heijnen JJ: **Revisiting the 13C-label distribution of the non-oxidative branch of the pentose phosphate pathway based upon kinetic and genetic evidence.** *FEBS J* 2005, **272**:4970-4982.
27. Zamboni N: **13C metabolic flux analysis in complex systems.** *Curr Opin Biotechnol* 2011, **22**:103-108.
28. Marin S, Lee WP, Bassilian S, Lim S, Boros LG, Centelles JJ, FernAndez-Novell JM, Guinovart JJ, Cascante M: **Dynamic profiling of the glucose metabolic network in fasted rat hepatocytes using [1,2-13C2]glucose.** *Biochem J* 2004, **381**:287-294.
29. Press WH, Flannery BP, Teukolsky SA, Vetterling WT: **Numerical Recipes in C: The Art of Scientific Computing.** *Cambridge University Press NY* 2002.
30. Flanigan I, Collins JG, Arora KK, MacLeod JK, Williams JF: **Exchange reactions catalyzed by group-transferring enzymes oppose the quantitation and the unravelling of the identify of the pentose pathway.** *Eur J Biochem* 1998, **213**:477-485.
31. Rose IA, Warms JV, Kuo DJ: **Concentration and partitioning of intermediates in the fructose bisphosphate aldolase reaction Comparison of the muscle and liver enzymes.** *J Biol Chem* 1987, **262**:692-701.
32. Nikerel IE, van Winden WA, Verheijen PJ, Heijnen JJ: **Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics.** *Metab Eng* 2009, , **11**: 20-30.
33. Xiong W, Liu L, Wu C, Yang C, Wu Q: **13C-tracer and gas chromatography-mass spectrometry analyses reveal metabolic flux distribution in the oleaginous microalga Chlorella protothecoides.** *Plant Physiol* 2010, **154**:1001-11.
34. Tang Y, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, Keasling JD: **Pathway confirmation and flux analysis of central metabolic pathways in Desulfovibrio vulgaris hildenborough using gas chromatography-mass spectrometry and Fourier transform-ion cyclotron resonance mass spectrometry.** *J Bacteriol* 2007, **189**:940-9.
35. Richardson AD, Yang C, Osterman A, Smith JW: **Central carbon metabolism in the progression of mammary carcinoma.** *Breast Cancer Res Treat* 2008, **110**:297-307.
36. Rich TC, Fagan KA, Nakata H, Schaack J, Cooper DM, Karpen JW: **Cyclic nucleotide-gated channels colocalize with adenylyl cyclase in regions of restricted cAMP diffusion.** *J Gen Physiol* 2000, **116**:147-161.
37. Karpen JWT, Rich C: **The fourth dimension in cellular signaling.** *Science* 2001, **293**:2204-2205.
38. Meléndez-Hevia E, Guinovart JJ, Cascante M: **The role of channelling in glycogen metabolism. In: Chanelling in intermediary metabolism.** *Portland Press London* 1997, 269-291.
39. Selivanov VA, Alekseev AE, Hodgson DM, Dzeja PP, Terzic A: **Nucleotide-gated KATP channels integrated with creatine and adenylate kinases: amplification, tuning and sensing of energetic signals in the compartmentalized cellular environment.** *Mol Cell Biochem* 2004, **256-257**:243-256.
40. Abraham MR, Selivanov VA, Hodgson DM, Pucar D, Zingman LV, Wieringa B, Dzeja PP, Alekseev AE, Terzic A: **Coupling of cell energetics with membrane metabolic sensing. Integrative signaling through creatine kinase phosphotransfer disrupted by M-CK gene knock-out.** *J Biol Chem* 2002, **277**:24427-24434.
41. Selivanov VA, Krause S, Roca J, Cascante M: **Modeling of spatial metabolite distributions in the cardiac sarcomere.** *Biophys J* 2007, **92**:3492-2500.
42. Brown KS, Sethna JP: **Statistical mechanical approaches to models with many poorly known parameters.** *Phys Rev* 2003, E 68.
43. Battogtokh D, Asch DK, Case ME, Arnold J, Schuttler B: **An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of Neurospora crassa.** *Proc Natl Acad Sci USA* 2002, **99**:16904-16909.
44. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP: **Universally sloppy parameter senitivities in systems biology models.** *PloS Comput Biol* 2007, **3**:1871-1878.
45. Gomez-Cabrero D, Compte A, Tegner J: **Workflow for generating competing hypothesis from models with parameter uncertainty.** *Interface Focus* 2011.
46. Fritzson PA: **Principles of Object-Oriented Modeling and Simulation with Modelica.** *IEEE Press NY* 2004, 21.
47. Fernández-Novell JM, Ariño J, Guinovart JJ: **Effects of glucose on the activation and translocation of glycogen synthase in diabetic rat hepatocytes.** *Eur J Biochem* 1994, **226**:665-67188.
48. Chan TM, Exton JH: **A rapid method for the determination of glycogen content and radioactivity in small quantities of tissue or isolated hepatocytes.** *Anal Biochem* 1976, **71**:96-105.
49. Kunst A, Draeger B, Ziegenhorn J: **D-Glucose; UV-methods with Hexolinase and Glucose-6-phosphate Dehydrogenase.** *In Methods of enzymatic analysis* 1984, **6**:163-172.
50. Passoneau JV, Lowry OH: **Enzymatic analysis: a practical guide.** *The Humana Press Inc* 1993.
51. Tserng KY, Gilfillan CA, Kalhan SC: **Determination of carbon-13 labeled lactate in blood by gas chromatography/mass spectrometry.** *Anal Chem* 1984, **56**:517-523.
52. Katz J, Lee WN, Wals PA, Bergner EA: **Studies of glycogen synthesis and the Krebs cycle by mass isotopomer analysis with [U-13C]glucose in rats.** *J Biol Chem* 1989, **264**:12994-13004.
53. Kurland IJ, Alcivar A, Bassilian S, Lee WNP: **Loss of [13C]glycerol carbon via the pentose cycle. Implications for gluconeogenesis measurement by mass isotoper distribution analysis.** *J Biol Chem* 2000, **275**:36787-36793.
54. Lee WP, Edmond J, Bassilian S, Morrow J: **Mass isotopomer study of glutamine oxidation and synthesis in primary culture of astrocytes.** *Develop Neurosci* 1996, **18**:469-477.
55. Lee W-NP, Byerley LO, Bergner EA, Edmond J: **Mass isotopomer analysis: theoretical and practical considerations.** *Biol Mass Spectrom* 1991, **20**:451-458.

doi:10.1186/1752-0509-5-175
**Cite this article as:** Marin de Mas *et al*: Compartmentation of glycogen metabolism revealed from $^{13}$C isotopologue distributions. *BMC Systems Biology* 2011 **5**:175.

# Chapter 2:
# Study the vulnerabilities associated to tumor heterogeneity and tumor progression in prostate cancer by using Genome-scale metabolic network models and data-driven approaches

## Summary

In this chapter we developed and applied a comparative Genome-scale metabolic model reconstruction analysis. This analysis was performed on a cellular model consisting of two clonal subpopulation of the prostate cancer cell line PC-3 in order to account for intra-tumoral heterogeneity, We also determined a metabolic gene signature associated to tumor progression and metastasis by performing a differential expression analysis on these subpopulations. This chapter summarizes the results of these two analysis included in two different articles that has been recently submitted to international scientific journals

# Chapter 2a

## Network model-driven discovery of calcitriol metabolism and fatty acid oxidation as therapeutic targets in metastatic prostate cancer

**Igor Marin de Mas**[1,2,3], Esther Aguilar[1,2], Erika Zodda[1,2,4], Timothy M Thomson[4], Balázs Papp[3,§,] and Marta Cascante[1,2,§,]

[1]Department of Biochemistry and Molecular Biology, Faculty of Biology, Institute of Biomedicine of University of Barcelona (IBUB) and IDIBAPS, Unit Associated with CSIC

[2]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[3]Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Szeged, Hungary.

[4]Department of Cell Biology, Barcelona Institute for Molecular Biology (IBMB), National Research Council (CSIC), Barcelona, Spain

[§]Corresponding author

# Network model-driven discovery of calcitriol metabolism and fatty acid oxidation as therapeutic targets in metastatic prostate cancer

**Igor Marin de Mas**[1,2,3], Esther Aguilar[1,2], Erika Zodda[1,2,4], Timothy M Thomson[4], Balázs Papp[3,§,] and Marta Cascante[1,2,§,]

[1]Department of Biochemistry and Molecular Biology, Faculty of Biology, Institute of Biomedicine of University of Barcelona (IBUB) and IDIBAPS, Unit Associated with CSIC

[2]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[3]Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Szeged, Hungary.

[4]Department of Cell Biology, Barcelona Institute for Molecular Biology (IBMB), National Research Council (CSIC), Barcelona, Spain

[§]Corresponding author

Papp Balázs; Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, H-6726 Szeged, Temesvári krt. 62., Hungary. Phone: +36 62 / 599 – 661; Fax: +36 62 / 433 – 506; e-mail address: pappb@brc.hu

Marta Cascante; Department of Biochemistry and Molecular Biology, Faculty of Biology, Universitat de Barcelona, Av Diagonal 643, 08028 Barcelona, Spain. Phone: +34-934021593; Fax: +34-934021559; e-mail address: martacascante@ub.edu

**Abstract**

Tumors undergoing Epithelial-Mesenchymal Transition (EMT) enhance their invasiveness initiating metastasis and drug resistance by promoting intra-tumoral heterogeneity. However, the mechanism underlying this process remains poorly understood. Here we built comparative genome-scale metabolic models (GSMM) based on transcriptomic data generated for a dual-cell prostate cancer cell model as a tool to unveil key metabolic vulnerabilities specific to distinct cancer cell subpopulations. In this model, PC-3M subpopulation displays CSC properties and high metastatic potential, while PC-3S subpopulation expresses EMT markers but displays low metastatic capacity.

Comparative GSMM predicted for PC-3M cells a metabolic impediment to the nuclear entry of calcitriol and high long-chain fatty acids entry into the mitochondria via CPT1 while contribute poorly to prostaglandin synthesis through eicosanoid metabolism. This metabolic reprogramming is predicted to endow the highly proliferative PC-3M cells by enhancing energy metabolism. In contrast, GSMM predicted for PC-3S cells a robust nuclear entry of calcitriol and fatty acid metabolic fluxes that mirrored those predicted for PC-3M cells. These predictions suggest a subpopulation-specific differential sensitivity to calcitriol and inhibition of fatty acid oxidation that was confirmed experimentally.

Metabolic rewiring of cancer cells favours their growth in adverse conditions. The same adaptive reprogramming that fosters cancer cell growth can originate de novo vulnerabilities that may be exploited therapeutically.

**Background**

Prostate cancer (PC) is the most commonly diagnosed non-cutaneous malignancy among Western men and accounts for the second leading cause of cancer-related death [1]. At the beginning of its development, PC is dependent on androgens, which explains the classical therapeutic approach towards the blockade of androgens, resulting in decreased neoplastic growth. Nevertheless, sooner or later, PC becomes independent of androgens, resuming growth in a more aggressive form [2].

Additionally, the coexistence within the same tumor of a variety of subpopulations, featuring different phenotypes (intra-tumoral heterogeneity) associated to tumor evolution and progression, confers an extreme flexibility and adaptation capability to cancer cells. This diversity is reached through genetic evolution of neoplastic cells and epigenetic and metabolic reprogramming of neoplastic and non-neoplastic tumor components that enhances tumor progression and represents a challenge for target-directed therapies [3,4].

Thus, intra-tumoral heterogeneity represents a hurdle that must be overcome in order to develop new and more efficient anti-cancer therapies. Thus, study of the metabolism in these heterogeneous cellular populations must be approached from an global perspective integrating the whole metabolism and accounting for different subpopulations.

In this context, Genome-scale Metabolic Models (GSMMs) have emerged as a potential solution to decipher the complexity of the molecular mechanisms underlying cancer in the context of systems biology [5,6]. This model-driven approach has been successfully used for the systematic analysis of cancer metabolic network in order to identify vulnerabilities in the form of one reaction or pairs of reactions (synthetic lethals) whose targeting compromise the viability of the cancer cell and can be potentially used to develop novel therapies [7,8]. However, the differences in metabolic physiology between intra-tumoral subpopulations have not yet been taken into account in these computational approaches.

Here we focused on the study of the metabolic activity profiles of two clonal sub-populations isolated from an established prostate cancer cell line (PC-3): PC-3/S (non-Cancer Stem Cells) and PC-3/M (Cancer Stem Cells -CSC-) [9]. These sub-populations were derived from the same tumor cell line and represent an excellent cellular model to study how intra-tumoral heterogeneity provides advantages to the tumor in terms of metastatic capability and drug resistance.

Thus, we have built comparative genome-scale metabolic network models (GSMM) based on transcriptomic data generated for a dual-cell PC cell model, derived from the PC cell line PC-3 [9], as a tool to unveil key metabolic nodes and vulnerabilities specific to distinct cancer cell subpopulations. In this cell model, the PC-3/M subpopulation displays CSC properties and high metastatic potential, while the PC-3/S subpopulation expresses Epithelial-Mesenchimal Transition (EMT) markers but displays low metastatic capacity or expression of other CSC properties. By systematic computational analysis of PC-3/M and PC-3/S metabolic network profiles we have identified subpopulation-specific -targets involving calcitriol metabolism and carnitine palmitoyl transferase 1 (CPT1) activity.

Here we present model predictions that were experimentally validated demonstrating that the low efficacy of both calcitriol and etomoxir in metastatic PC tumors is conferred by the different sensitivities of PC-3/M and PC-3/S subpopulations towards these drugs and suggests the combinatorial use of calcitriol and etomoxir as a strategy to treat metastatic prostate tumor. The approach presented hereby opens new avenues in the development of more specific and efficient anti-tumoral therapies.

**Results**

**i) Characterization of the subpopulations metabolic profile by a comparative GSMM Analysis**

Here, we have focused on the study of the metabolic activity profiles of two clonal sub-populations isolated from an established prostate cancer cell line (PC-3): PC-3/S

(like epithelial-mesenchymal transition cell line) and PC-3/M (like tumor-initiating cells cell line) [9].

To infer the activity state of the metabolic network in each subpopulation, we have used previously generated transcriptomic data for PC-3/S and PC-3/M cells using microarray technology [9] and integrated these data into a genome-scale reconstruction of human metabolic network [10]. We applied one of the most widely tested constraint-based methods (CBM) [11] to integrate the normalized gene expression data into the most recent reconstruction of human metabolism (Recon2) [10]. In brief, this method defines an upper threshold above which the genes are considered as highly expressed and a lower threshold below which the genes are considered as lowly expressed and seeks a network activity state in which the number of highly expressed genes that are active and the number of lowly expressed genes that are inactive are maximized. We note that the activity state cannot be uniquely determined for all reactions owing to the presence of potential alternative activity states that display the same overall consistency with the expression data. To identify such cases of non-uniqueness, we performed a sensitivity analysis (See Methods) in which we defined, for each metabolic reaction, if it was active, inactive or undetermined. This procedure allowed us to infer the metabolic activity state profile of both, particular reactions and metabolic pathways, in each PC-3 subpopulation (see Supplementary Material 1 and 3).

These results were validated by comparing the predicted activity state of the exchange reactions with available measurements of uptake /secretion rates [12]. These showed a highly significant overlap with the predictions obtained by integrating gene expression and GSSM, for which the contingency table can be found in Supplementary Material 3.

In order to define a set of reactions whose activity state was unambiguously different between PC-3/M and PC-3/S cells, we performed a robustness analysis to determine the activity state of the metabolic reactions using different sets of thresholds. Next, by comparing the results obtained in this analysis we excluded those reactions which activity state prediction was not consistent along the different set of thresholds (predicted to be active or inactive depending on the thresholds) as well as those

reactions that were not predicted to have different activity state between subpopulations within the same analysis at least once. Thus, we defined a set of reactions with constant and robust predicted activity states and independent of the parameters used in the analysis with different activity states between subpopulations. The results of this analysis are summarized in Table 1.

| Reactions active in PC-3/M & inactive in PC-3/S | |
|---|---|
| **Pathway** | **nº Active reactions** |
| Fatty acid oxidation | 30 |
| Fatty acid synthesis | 4 |
| Miscellaneous | 2 |
| N-glycan degradation | 6 |
| N-glycan synthesis | 1 |
| Pyrimidine catabolism | 1 |
| Sphingolipid metabolism | 1 |
| Transport, extracellular | 1 |
| Transport, lysosomal | 1 |
| Transport, mitochondrial | 7 |

| Reactions active in PC-3/S & inactive in PC-3/M | |
|---|---|
| **Pathway** | **nº Active reactions** |
| Cholesterol metabolism | 5 |
| Citric acid cycle | 1 |
| Eicosanoid metabolism | 6 |
| Fatty acid synthesis | 2 |
| Nucleotide interconversion | 3 |
| Sphingolipid metabolism | 3 |
| Transport, extracellular | 1 |
| Transport, peroxisomal | 1 |
| Tryptophan metabolism | 1 |
| Unassigned | 2 |
| Vitamin D metabolism | 4 |

Table 1. Pathways differentially activated between subpopulations: This table shows the number of reactions active only in one of the subpopulations and the corresponding pathway. Left, reactions active in PC-3/M and inactive in PC-3/S. Right, reactions active in PC-3/S and inactive in PC-3/M.

Table 1 shows the results of the robustness analysis. A more detailed table specifying all the reactions can be found in Supplementary Material 5. We can observe that the reactions differentially activated in the two subpopulations are clustered in specific pathways defined in the Comparative GSMM Analysis.

We observed that the activity of fatty acid oxidation predicted by the model is higher in PC-3/M than in PC-3/S cells. The oxidation of fatty acids in the mitochondria produces NADH, FADH2 and AcetylCoA that fuels the production of energy via TCA cycle and the electron chain. This strategy can be used by PC-3/M cells to sustain their higher proliferation rate. Most of the reactions differentially activated in this pathway involve carnitine palmitoil transferase 1 (CPT1). This mitochondrial membrane protein actively transports long-chain fatty acids (LCFA) from cytosol into the mitochondria. The analysis provided a set of eight cytosolic LCFA that were predicted to be substrates of

CPT1 exclusively in PC-3/M. Interestingly, it has been reported that five of these eight LCFA have antiproliferative effects [12-17].

The analysis also predicted a higher activity of eicosanoid metabolism in PC-3/S cells and more specifically in those reactions involved in arachidoic acid metabolism (AA). This molecule is metabolized from arachidoil CoA, one of the eight LCFA previously mentioned. In addition, AA is the precursor of eicosanoid metabolism that produces several prostaglandins with antiproliferative effects such as prostaglandin J2 (PJ2) [18]. AA is also the precursor of the cannabinoid 2-Arachidonoylglycerol (2-AG) that activates the cannabinoid receptor type 1 (CB1) [19,20] that, in turns, reduces tumor progression in prostate cancer cells [21]. This fact could explain, at least in part, why the proliferative rate of PC-3/S cells is significantly lower than that of PC-3/M cells. These metabolic processes are predicted to be more active in PC-3/S cells and the metabolization of 2-AG would take place exclusively in this cell line. Vitamin D3 metabolism is also predicted to be exclusively active in the PC-3/S subpopulation. This molecule controls proliferation in prostate cells [22] and has antiproliferative effects on a number of cancer cell lines, being PC3 cells one of the few cell lines insensitive to this drug [23]. This prediction is consistent with the lower proliferative ration observed in PC-3/S cells.

**ii) Difference in calcitriol metabolism explains drug tolerance**

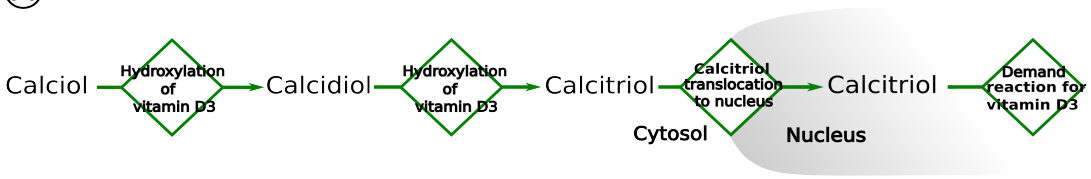
One of the most relevant metabolic differences we have found between the two PC-3 subpopulations is in calcitriol metabolism: it is predicted to be active in PC-3/S but inactive in PC-3/M (Figure 1.A). Calcitriol is the most active form of vitamin D and triggers several cellular mechanisms that regulate calcium homeostasis when entered in the nucleus. Most importantly, calcitriol has antiproliferative activity in a variety of human cancer cells by inducing apoptosis and G0-G1 cell cycle arrest [23]. However PC3 cells are one of the few cancer cell lines that are insensitive to calcitriol treatment.

We hypothesize that the resistance of PC-3 cells towards the antiproliferative effects of calcitriol [23,24] could be explained by the metabolic impediment of PC-3/M cells to the nuclear entry of calcitriol. To test this hypothesis we designed an experiment in

which PC-3/S and PC-3/M cells were cultured in parallel in the presence of increasing concentrations of calcitriol and cell proliferation was monitored (Figure 1.B).

In line with the above hypothesis, we found that PC-3/S cells are significantly more #sensitive# than PC-3/M cells to the inhibitory activity of calcitriol on proliferation.d. Together with the fact that the growth rate of PC-3/S cells is half of that of PC-3/M cells [9] these findings may provide an explanation why the global inhibitory effect of calcitriol on heterogeneous PC-3 cells is weaker than on other cancer cell lines.

Ⓐ




Ⓑ

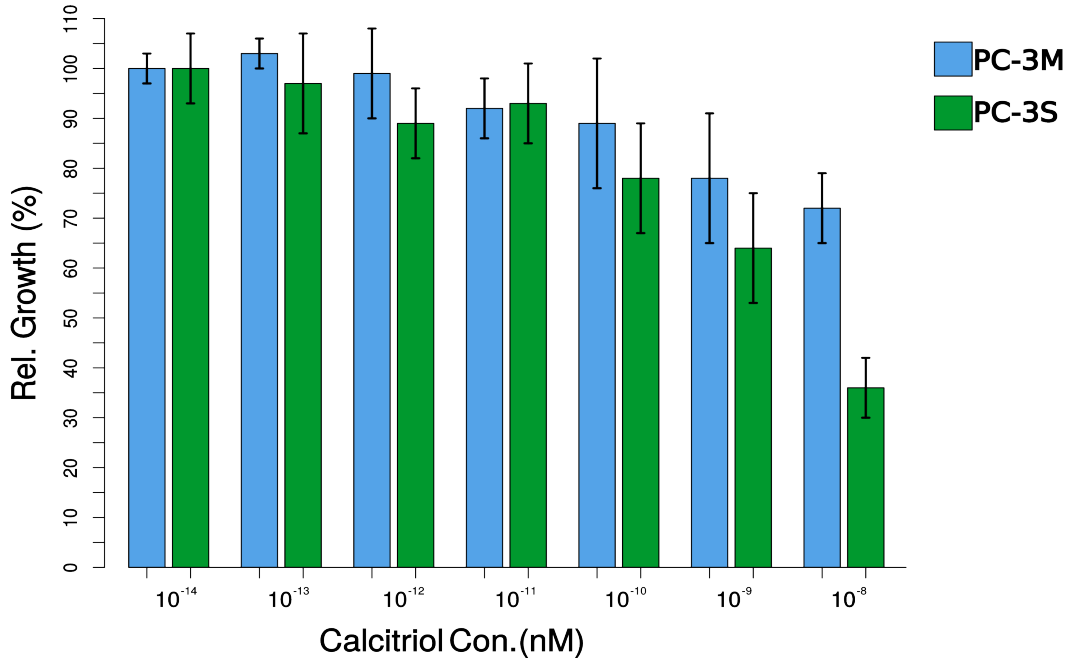


**Figure 1: A:** Vitamin D3 metabolic pathway and translocation into the nucleus is predicted to be more active in PC-3S cells **B:** Experimental validation of model predictions. Response of proliferation in PC-3M (blue bars) and PC-3S (green bars) to an increasing concentration of calcitriol relative to a control without calcitriol. All the proliferations are normalized by the proliferation rate measured in the absence of calcitriol. Green and blue bars represent the response of PC-3S and PC-3M cells respectively, the error bars represents the means ± SD

### v) Difference in fatty acid uptake into mitochondria explains the higher etomoxir sensitivity of PC-3/M

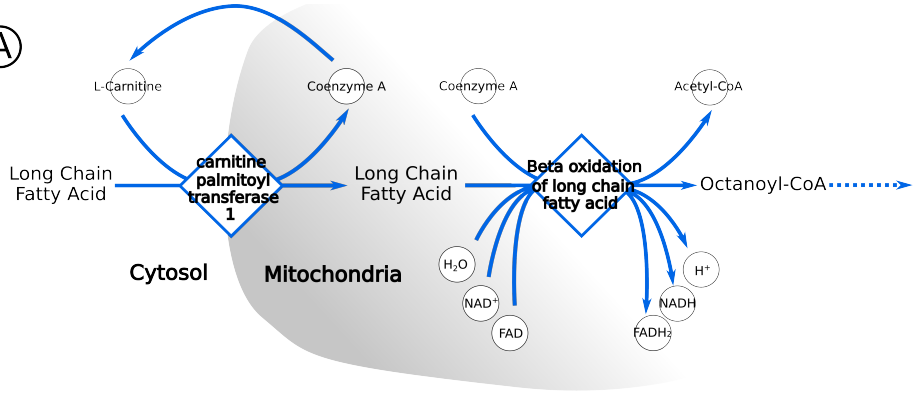
Our computational analysis predicts both a higher long-chain fatty acids (LCFA) entry into the mitochondria via CPT1 and a more active LCFA β-oxidation in PC-3/M cells. LCFA must be imported into mitochondria to be degraded via β-oxidation and CPT1, mitochondrial membrane enzyme that plays a critical role in its transport into the mitochondrial matrix [24]. To experimentally verify that CPT1 protein level differs between PC-3/S and PC-3/M, we compared CPT1 levels in PC-3/S and PC-3/M subpopulations by Western blotting (Figure XX, see Methods). In line with the computational inference, we found that PC-3M cells express 30% higher levels of CPT1 PC-3/S cells.

Etomoxir is a CPT1 inhibitor that consequently inhibits β-oxidation and its associated $O_2$ consumption [24]. Since in many cancer types, tumor onset and progression relies on lipid fuel more than on aerobic glycolysis, this compound is widely used in anti-cancer therapies [24]. However, as stated above, its antiproliferative efficacy on PC-3 cells is the lowest among the different PC cell lines studied[24]. Based these evidences and the results of our analysis, we hypothesized that tthis could be explained by a low activity of CPT1 in the PC-3/S subpopulation.

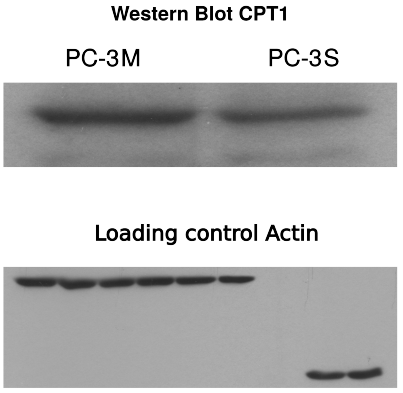
In order to experimentally test this hypothesis, we measured the oxygen consumption rate (OCR) before and after exposure to etomoxir of both subpopulations.

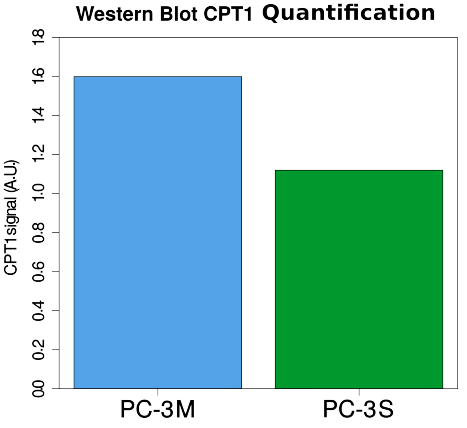
We found that PC3/M cells show a 30% higher sensitivity to CPT1 inhibition than PC-3/S cells, implying that CPT1 plays a more relevant role in beta oxidation in the PC-3/M subpopulation.

Ⓐ


L-Carnitine
Coenzyme A
Coenzyme A
Acetyl-CoA
Long Chain Fatty Acid
carnitine palmitoyl transferase 1
Long Chain Fatty Acid
Beta oxidation of long chain fatty acid
Octanoyl-CoA
Cytosol
Mitochondria
$H_2O$
$NAD^+$
FAD
$H^+$
NADH
$FADH_2$

Ⓑ


Western Blot CPT1
PC-3M
PC-3S
Loading control Actin


Western Blot CPT1 Quantification
CPT1signal (A.U.)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
PC-3M
PC-3S

Ⓒ

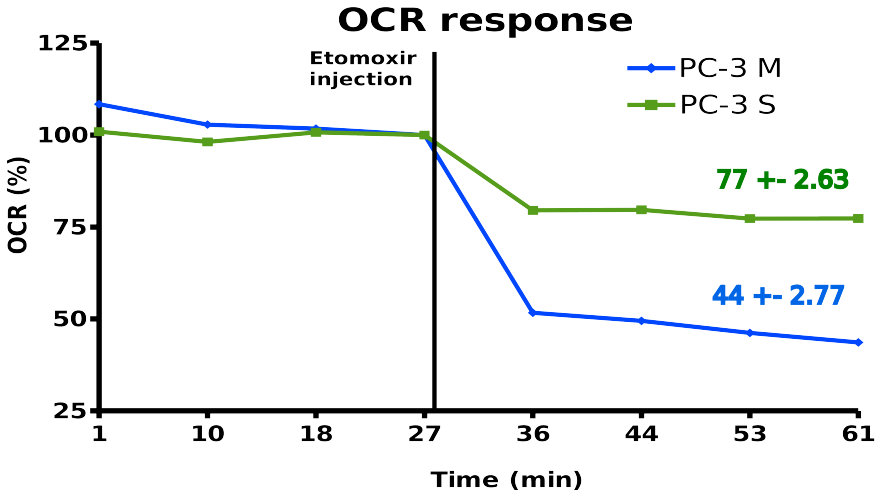

OCR response
Etomoxir injection
PC-3 M
PC-3 S
77 +- 2.63
44 +- 2.77
OCR (%)
125
100
75
50
25
1
10
18
27
36
44
53
61
Time (min)

Figure 2 A: Long chain fatty acid transport (LCFA) from cytosol into the mitochondria via CPT1 and LCFA β-oxidation oxidation until Octanoyl-CoA is predicted to be more active in PC-3M cells. B: Validation of model prediction. On the left, Western blot of CPT1 in PC-3/S on the right and in PC-3/M on the left and the loading control done with actine. On the right, quantification of wetern blot by using ImageJ software [25]. Blue bar represents the quantification of CPT1 in PC-3M and the green bar represents the quantification of CPT1 in PC-3S, both in arbitrary units (A.U.). C: Validation of model predictions. Oxygen consumption rate (OCR) before and after inhibit CPT1 with Etomoxir. The measurements are normalized by fixing the OCR before the inhibition as the 100%. Green line represents the OCR associated to PC-3/S subpopulation and blue line the OCR associated to PC-3/M subpopulation. The final values are represents as means ± SD

## Discussion

Intratumoral heterogeneity is key to understanding the hierarchical and functional relationships between different neoplastic cell populations within a given tumor, with direct implications on tumor dynamics and progression. Here we have focused on the study of the metabolic profiles of two clonal sub-populations isolated from an established prostate cancer cell line (PC-3): PC-3/S (presenting stable traits of epithelial-mesenchymal transition) and PC-3/M (enriched in tumor-initiating cells) [9]. These sub-populations were derived from the same PC cell line and thus subpopulations with related phenotypes can be assumed to coexist within the same tumor and represent an excellent model to study how intra-tumoral heterogeneity provides advantages to the tumor in terms of metastatic capability and drug resistance, and also to study relationships of gene and metabolic networks to mesenchymal-like or TIC attributes of neoplastic cells.

Here we have applied a CBM that allowed us to predict the metabolic activity states of these PC-3 subpopulations by integrating transcriptomic data into a GSMM reconstruction. This method treats gene expression levels merely as cues for the likelihood that their associated reactions carry metabolic fluxes, taking into consideration possible post-transcriptional modifications. As such, a reaction associated with a highly expressed gene does not necessarily carry out the corresponding metabolic flux at proportionate rates. As a consequence, the method

allows us to propose predictions that go beyond conventional gene expression analysis. We found that the enrichment provided by our model-driven analysis in terms of active reactions or associated with a highly expressed gene in one subpopulation and with inactive or associated with a lowly expressed gene in the other was 0.0020% before and 0.0087% after the analysis with an associated level of significance of 5.686E-6 (calculated by T test). It is worth noting that some of the main metabolic differences that were predicted between PC-3/M and PC-3/S cells did not correlate with the expression patterns of relevant genes. One example is provided by the expression levels of genes associated with the reaction that catalyzes the hydroxylation of Vitamin D3, higher in PC-3/M cells, which our computational analysis predicted to be active only in the PC-3/S subpopulation (See Supplementary material 5). This prediction was experimentally validated as a higher sensitivity of PC-3S cells to calcitriol, thus demonstrating the importance of considering possible post-transcriptional effects in order to capture the metabolic processes underlying specific cellular phenotypes.

Additional metabolic activity state patterns predicted for our PC-3 subpopulations were consistent with bibliographical evidences consistent with phenotypes similar to our cell model. Thus, the synthesis of purines is predicted to be more active in PC-3/M cells. This correlation has long been known to apply in highly proliferative cancer cells where purines are used as building blocks required for high DNA duplication rates [26]. On the other hand, the catabolism of purines is predicted to be more active in PC-3/S cells, correlating with their lower rates of DNA synthesis proliferation. The predicted high purine metabolism activity in PC-3/M is also consistent with the higher activity state of glutamate metabolism predicted for the same cells, necessary to provide amino groups for the synthesis of purines and pirimidines [27]. Glycolisys is also predicted to be more active in PC-3/M cells, a characteristic shared by most highly proliferative cancer cells [28]. Another example is provided by N-glycan degradation pathways, predicted to be more active in PC-3/M cells, while N-glycan synthesis pathways are predicted to be mainly active in PC-3/S cells.

Our analysis further predicted acid ceramidase (AC) ASAH1 to be active only in PC-3/M cells. As a consequence of their lipogenic phenotype, cancer cells can accumulate

proapoptotic ceramides [29]. Possibly as an adaptive response, many tumor cells expressed high levels of AC to catalyze the hydrolysis of ceramide into sphingosine thus reducing their proapoptotic effects [30]. This prediction is corroborated by our previous finding that the ASAH1 enzymatic activity is significantly higher in PC-3/M cells than in PC-3/S cells [31].

Our model has also predicted a higher CPT1 activity in PC-3/M cells than PC-3/S cells. The substrates of CPT1 in the cytosol, exclusive of PC-3/M, are cervonyl coenzyme A, eicosatetranoyl coenzyme A, arachidyl coenzyme A, trans-2-octadecenoyl-CoA(4-), palmitate, Malonyl-CoA, linoelaidyl coenzyme A and vaccenyl coenzyme A. Interestingly, has been reported that five of these eight LCFAs have antiproliferative effects [13-17]. Thus, the higher CPT1 and β-oxidation activities in PC-3/M cells could perform two roles. First, and probably the most evident, they would maintain the energetic requirements imposed by the high growth rates of PC-3/M cells and, second, they would be used to eliminate LCFAs with antiproliferative effects. The majority of these LCFAs are prostaglandin precursors of eicosanoid metabolism. Thus, AA is metabolized from arachidyl coenzyme A and AA is a precursor of the antiproliferative 2-AG [19,20]. In addition, 2-AG is a precursor of prostaglandin J (PGJ), at inhibisor of androgen receptor signaling whose overexpression is critical for the growth and progression of prostate cancer [21]. On the other hand we demonstrated that, unlike PC-3/M cells, PC-3/S cells are sensitive to the anti-proliferative effects of calcitriol, that can be given by their predicted capability to uptake calcitriol into the nucleus and consequently are more sensitive than PC-3M cells to the antiproliferative effects of this compound. Calcitriol also regulates the homeostasis of calcium increasing its intracellular concentration [32], necessary for the metabolization of some prostaglandins with antiproliferative effects such as PGJ or PGA1 [21]. These observations lead us to prose a mechanism by which the differential metabolism of LCFAs in the two PC-3 subpopulations would explain at least part of their opposing phenotypes (Fig 3).

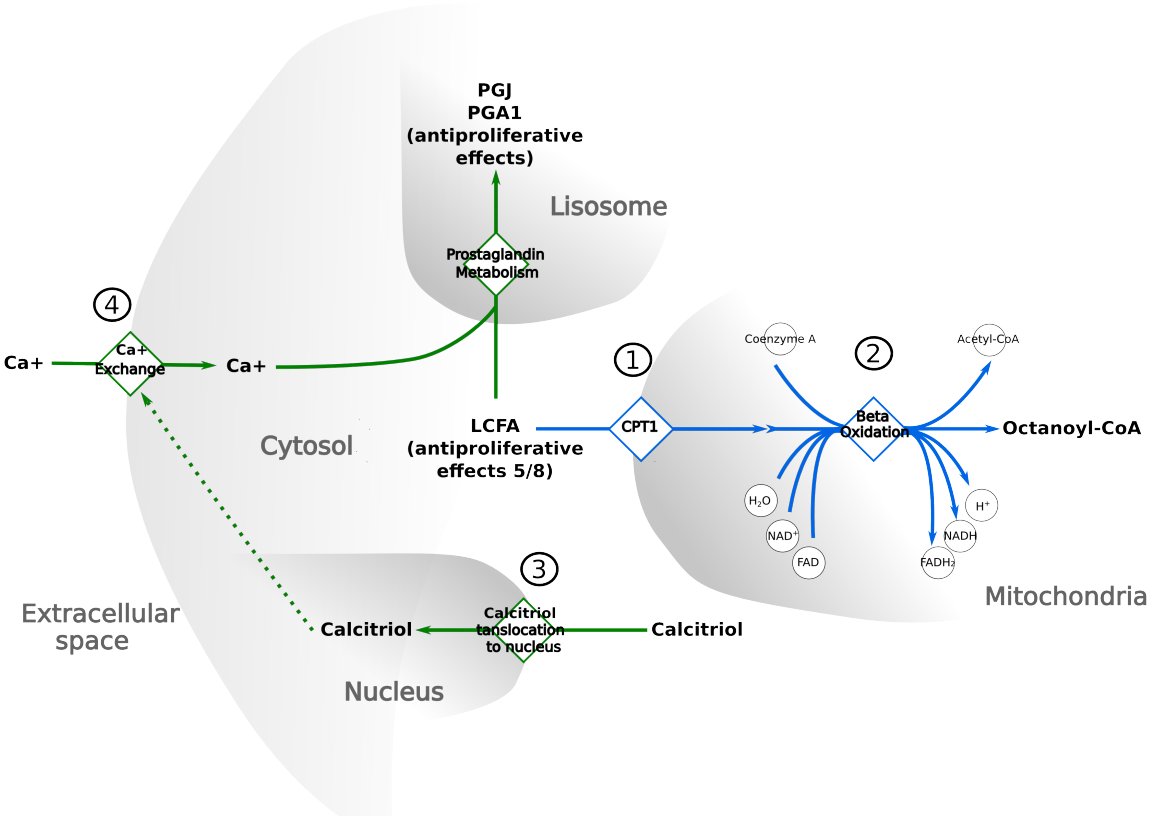



Figure 3: Proposed mechanism of metabolic reprogramming between PC-3/M and PC-3/S. Green solid arrows represents pathways or groups of metabolic reactions mainly active in PC-3/S subpopulations, blue solid arrows represents pathways or groups of metabolic reactions active in PC-3/M, discontinue green arrow represent the calcitriol signaling pathway that regulates calcium homeostasis

The relatively highactivity of CPT1 in PC-3/M cells increases the entry of LCFAs into the mitochondria (1) to be oxidized (2) to produce energy to sustain a highly proliferative activity. This causes decreased levels of LCFAs, thus preventing their antiproliferative effects. In contrast, the lower CPT1 activity in PC-3/S cells would lead to an accumulation of LCFAs in the cytosol. Additionally only PC-3S cells are predicted to possess efficient mechanisms for nuclear translocation of calcitriol (3) which results in increased levels of intracellular Ca2+ (4). Ca2+ is necessary in some

steps of the prostaglandin metabolim (4), and more specifically those with antimetastatic effects. Summing up, we propose that the metabolic reprogramming involving LCFA utilization drives PC-3/M cells to be more metastatic and highly proliferative than the PC-3/S subpopulation.

Our analysis of a dual-cell model representing distinct and opposing neoplastic phenotypes allows us to propose subpopulation-specific and complementary therapeutic interventions. Thus, the resistance of PC-3 cells to calcitriol may be explained by the observed resistance of the PC-3/M subpopulation to this compound. On the other hand, and the poor performance of etomoxir at inhibiting the growth of PC-3 cells compared to other prostate cell lines may be explained by the low metabolic dependence of the PC-3/S subpopulation on CPT1. In other words, androgen-independent prostate cancer cells with CSC attributes similar to PC-3M cells would be sensitive to etomoxir, while tumor cell subpopulations with EMT attributes similar to PC-3S cells would be sensitive to calcitriol. Therefore a combined therapeutic strategy that uses calcitriol and etomoxir would be expected to target two major tumor cell subpopulations, CSCs and non-CSCs with mesenchymal-like characteristics.

Our study represents a novel approach to discern metabolic vulnerabilities associated with heterogeneous tumor cell populations and opens new avenues for the development of new anti-cancer therapies.

## Methods

### Experimental data

In this work we have studied the metabolic differences of two different clonal subpopulations of a prostate cancer cell line (PC-3) [9]. For this aim we have used two types of experimental data:

Transcriptomic data: We determined the gene expression levels of each cell subpopulation by microarray analysis (*Affymetrix* genechip u133a 2.0). Next, the gene expression data was normalized by RMA [33].

Consumption and production of metabolites: Additionally, we use the measured consumption and production of some metabolites [12] to asses the reliability of model predictions (See Supplementary material 3). These metabolites were: glucosa, lactate, pyruvate, glutamate and aminoacids.

**Metabolic model**

To obtain an accurate cell-specific genome-scale metabolic models of the PC-3 subpopulations, we performed a subpopulation-specific genome-scale network reconstruction analysis by integrating the transcriptomic data into the most recent reconstruction of human metabolism (Recon2) [10]. Recon2 is a genome-scale stoichiometric model that represents the entire network of human metabolic reactions and accounts for 1,789 enzyme-encoding genes, 7,440 reactions and 2,626 unique metabolites distributed over eight cellular compartments. This generic genome-scale metabolic model provides the appropriate transcript-protein-reaction associations that permit the integration of the previously mentioned transcriptomic data for which we used a widely tested constraint-based method [11].

**Model reduction**

In order to reduce the computational time necessary to perform the analysis, the metabolic model (Recon2) was reduced. The reduction was done by removing the blocked reactions in the model. These reactions are those incapable of carrying any metabolic flux [34]. The reduction was done in two steps:

Determine the blocked reactions: In order to determine the blocked reactions in the model, we perform a Flux Variability Analysis (FVA) [35-37] using Fasimu software [38]. This analysis computes minimal and maximal flux in each reaction. Each analysis evaluates the feasibility of the simulation. The reactions in which their maximization and minimization simulations was not obtained a feasible solution were considered as blocked reactions [34].

Remove blocked reactions: Once the blocked reactions were determined we removed

them from the model. Next, those metabolites that were neither as product nor substrate of any reactions were eliminated from the model. Finally, we checked if it was some compartment without metabolites in order to be also removed (no compartment was removed).

**Transcriptomic data integration**

We integrated the transcriptomic data into Recon2 by using the gene-protein-reaction (GPR) associations included in the model. These associations are "and/or" logical sentences that establish a relation between the metabolic reactions and the genes encoding the enzymes that catalyze them. GPR associations include information related with isoenzymes (using the logical "or"), complexes (using the logical "and") or direct gene-reaction relations (i.e. the activity of Reaction1 depends on: "(geneA and geneB) or (geneC and geneD)" ). It lead us to integrate the gene expression data from PC-3/M and PC-3/S subpopulations into Recon2 and determine the gene expression level associated to the metabolic reactions in each subpopulation. For this aim we substituted the logical "and" and "or" by "minimum" and "maximum". Thus, for example if the activity of a given reaction depends on the expression of different genes and it is defined by the following logical expression "(geneA and geneB) or (geneC and geneD)", and the expression of the gene A, B, C and D are 0.5, 3, 1 and 0.1 respectively. Then, by integrating the gene expression levels into the logical sentence and replacing the logical operators by "minimum" and "maximum" we obtain the following expression: "max(min(0.5,3),min(1,0.1))". Thus, based on the transcriptomic data and the GPR association, the gene expression associated to the reaction is 0.5.

**Expression-based activity prediction**

Next, we used gene expression levels associated to the metabolic reactions to infer the activity states of reactions in the network by using a recently developed constraint-based method [11].

Optimization algorithm: This method solves a mixed integer linear programming

(MILP) problem to obtain a flux distribution in which the number of reactions associated with highly expressed genes is maximized ($R_H$), and the number of reactions associated with lowly expressed genes is minimized ($R_L$) while satisfies the thermodynamic and stoichiometric constrains imposed by the model:

$$\max_{v,y^+,y^-}\left(\sum_{i\in R_H}\left(y_i^+ + y_i^-\right)+\sum_{i\in R_L} y_i^+\right)$$

(1) $S * v = 0$

The mass balance constraint: where v is the flux vector and S is a n x m stoichiometric matrix, in which n is the number of metabolites and m is the number of reactions.

(2) $v_{min} \le v \le v_{max}$

Thermodynamic constraints, that restrict flow direction, are imposed by setting $v_{min}$ and $v_{max}$ as lower and upper bounds respectively.

(3) $y_1^+, y_i^- \in [0,1]$

The Boolean variables y+ and y–. In $R_H$ reactions represent whether the reaction is active or not respectively. In $R_L$ y+ represents the reaction is not active.

(4) $v_i + y_i^+\left(v_{min,i} - \varepsilon\right) \ge v_{min,i}, i \in R_H$

A highly expressed reaction is considered to be active if it carries a significant positive flux that is greater than a positive threshold $\varepsilon$. In our study $\varepsilon$=1. Consequently the ith reaction is active if: $v_i \ge 1$

$$v_i + y_i^-\left(v_{min,i} + \varepsilon\right) \le v_{max,i}, i \in R_H$$

or has a significant negative flux < – $\varepsilon$ (as our model didn't consider reversible reactions it cannot occur)

(5) $v_{min,i}\left(1 - y_i^+\right) \le v_i \le v_{max,i}\left(1 - y_i^+\right), i \in R_L$

Lowly expressed reactions are considered to be inactive if they carry zero metabolic flux, though changing equation (5) to enable these reactions to carry a low metabolite flux (that is, with an upper bound lower than $\varepsilon$) and still be considered inactive provides qualitatively similar results. The following figure illustrates the process:

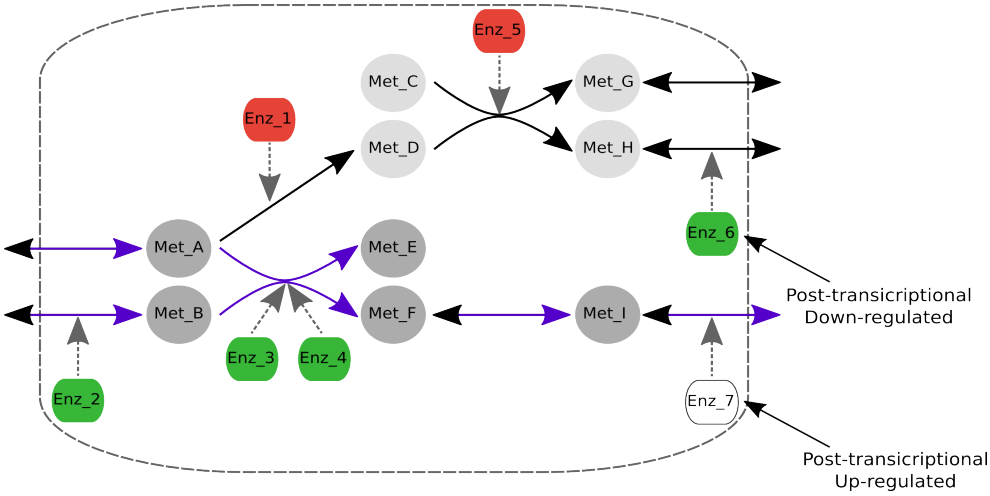



Figure 4: This figure represents a toy metabolic network. Here the nodes from Met_A to Met_I represents the metabolites involved in this network, the metabolic reactions are represented by the continuous arrows, the nodes from Enz_1 to Enz_7 represent the enzymes that catalyze the metabolic reactions and the discontinuous arrows indicate to which enzyme is associated each metabolic reaction. The enzymes associated to highly expressed genes are highlighted in green, the ones associated to lowly expressed genes are highlighted in red and in white are those enzymes associated to normally expressed genes. The algorithm penalizes the use of reactions associated with lowly expressed genes and prices the use of those associated with highly expressed genes. Thus, based on the expression of the genes associated to this metabolic network, the algorithm will predict that the reactions highlighted in purple will be active while the reactions in black will be inactive. Consequently the metabolite Met_I will be secreted but not Met_G neither Met_H

Parameters: This method defines an upper threshold above which the expression of a certain gene is considered high and other bellow which the gene expression is considered low. In our study the chosen upper and lower threshold were those symmetric percentiles that maximize the cases where the number of reactions associated to highly express genes in one subpopulation were associated with lowly expressed gene in the other subpopulation and vice versa. Thus, we defined the upper threshold at the $66^{th}$ percentile and the lower threshold at the 33th percentile. The method also uses the parameter $\varepsilon$ that represents the flux above which a given reaction

is considered to carries a significant metabolic flux. As is defined in [11] we gave to $\varepsilon$ a value of 1. Once the thresholds were fixed, we perform the expression-based activity prediction analysis by using Fasimu software [38].

**Sensitivity analysis**

In the Expression-based activity prediction analysis we found an optimal solution in terms of the objective function maximization, but this solution may not be unique. Actually, it may exist a space of alternative optimal solutions that represents alternative steady-state flux distributions obtaining the same similarity with the gene expression data (the same objective function value).

To account for these alternative solutions, we employed Sensitivity analysis [11].

This is performed by solving two MILP problems (as is described in Expression-based activity prediction) for each reaction to find the maximal attainable similarity with the expression data when the reaction is: (i) forced to be activated and (ii) forced to be inactivated.

Thus, a reaction is considered to be active if a higher similarity with the expression data is achieved when the reaction is forced active than when it is inactive (the objective function is higher when the reaction is active). Conversely, it is considered to be inactive if the similarity is higher when the reactions is forced to be inactive. If the similarities with the experimental data are equal in both cases the activity state of the reaction is considered to be undetermined. From this analysis we could infer which pathways were more active in each subpopulation.

**Reliability of model predictions**

By analyzing the predicted activity state of the exchange reaction we can infer which metabolites are consumed and/or produced. In order to determine the goodness of our model predictions we compare qualitatively the consumption and production of some measured metabolites (see experimental data) with the corresponding model predictions. This comparison was done by constructing a 2x2 confusion matrix and the levels of significance were determined using Fisher exact test (See results).

**Robustness analysis**

The algorithm used to integrate the information from gene expression levels into a Genome-scale metabolic network reconstruction defines a threshold above which the expression of the genes is considered high and other bellow which the expression of the genes is considered low. It raises the necessity to perform a robustness analysis in order to demonstrate the lack of dependency of our predictions to the thresholds used in the analysis. In order to determine the robustness of our prediction we performed the analysis previously described in sensitivity analysis defining different sets of thresholds:

- upper threshold: 25$^{th}$ percentile; lower threshold: 75$^{th}$ percentile
- upper threshold: 33$^{th}$ percentile; lower threshold: 66$^{th}$ percentile
- upper threshold: 40$^{th}$ percentile; lower threshold: 60$^{th}$ percentile

Thereby, we defined a set of reactions that were predicted to be active, inactive or undetermined (the method cannot predict their activity state) independently of the thresholds.

**Oxygen consumption rate (OCR)**

A XF24 Extracellular Flux Analyzer (Seahorse Bioscience) was used to measure the rate change of dissolved oxygen in media immediately surrounding adherent cells cultured in a XF24-well microplate (Seahorse Bioscience). PC-3M and PC-3S cells were seeded in XF24-well microplates at $4.5 \cdot 10^4$ cells/well and $9.0 \cdot 10^4$ cells/well, respectively, in 100 μL of growth medium, adding 100 μL more after 3-5 h, and then incubated at 37 °C with 5% $CO_2$ overnight. After overnight incubation and 1 h before the Seahorse assay, growth media was replaced by basal media (unbuffered DMEM; Sigma-Aldrich) with 3 mM glucose and 5 mM carnitine. The sensor cartridge was loaded with etomoxir and calibrated prior to the start of the Seahorse experiment. Response to the etomoxir treatment (final concentration 30 μM) is expressed as LOG2 to indicate the fold change comparing the measured point immediately after and before

the corresponding injection. The drug etomoxir was purchased from Sigma-Aldrich.

**EXPERIMENTOS CALCITRIOL – PROLIFERACIÓN**

Cell proliferation was assessed by Hoechst staining (HO33342; Sigma-Aldrich). Briefly, cells were seeded in 96-well plates and media was replaced after 24 h with complete fresh media containing ethanol vehicle or the drug calcitriol. At the end of the experiment, media was removed, cells were washed with PBS, the supernatant was aspirated and 100 L of 0.01% SDS was added to each well. Plates were then frozen at -20°C until analyzed. To analyze the samples, plates were thawed at 37°C until fully liquid and 100 L of HO33342 stain solution were added to each well, minimizing light exposure. HO33342 stain solution was prepared by diluting HO33342 stock solution (10 mg/mL in $H_2O$) with Assay Buffer containing 1 M NaCl, 0.1 M EDTA, 1 M Tris, pH 7.4. Plates, covered with tinfoil, were placed on a shaker and incubated at 37°C for 1 h in the dark. Finally, fluorescence was measured in a fluorescence plate reader at 355 nm excitation and 460 nm emmission.

**Acknowledgements**

This work was supported by the European Commission Seventh Framework Programme FP7; the Spanish Government and the European Union FEDER funds (SAF2011-25726); Generalitat de Catalunya-AGAUR

**Author Contributions**

IM performed all the computational analysis , wrote the draft, edited versions of the manuscript and final versions, generated the illustrations and participated in the design of specific experiments; EA performed and analyzed most experiments and participated in the design of the experiments, edited versions of the manuscript; EZ performed additional experiments; TMT designed and supervised specific experiments and wrote the draft and final versions of the manuscript; PB designed and supervised

the overall study and specific experiments including data analysis and wrote and approved manuscript drafts and final versions; MC designed and supervised the overall study and specific experiments including data analysis and wrote and approved manuscript drafts and final versions.

**References**


1. Altekruse SF, Huang L, Cucinelli JE, McNeel TS, Wells KM, Oliver MN. Spatial patterns of localized-stage prostate cancer incidence among white and black men in the southeastern United States, 1999-2001. Cancer Epidemiol Biomarkers Prev. 2010 Jun;19(6):1460-7.

2. Karantanos T, Evans CP, Tombal B, Thompson TC, Montironi R, Isaacs WB. Understanding the Mechanisms of Androgen Deprivation Resistance in Prostate Cancer at the Molecular Level. Eur Urol. 2014 Oct 8.

3. Renovanz M, Kim EL. Intratumoral heterogeneity, its contribution to therapy resistance and methodological caveats to assessment. Front Oncol. 2014 Jun 10;4:142.

4. Janiszewska M, Beca F, Polyak K.. Tumor heterogeneity: the lernaean hydra of oncology? Oncology (Williston Park). 2014 Sep;28(9). pii: 201367.

5. Mardinoglu A, Nielsen J. Systems medicine and metabolic modelling. J. Intern. Med. 2012. 271(2), 142-154.

6. Marín de Mas Igor, Aguilar Esther, Jayaraman Anusha, Polat Ibrahim H., Martín-Bernabé Alfonso, Bharat Rohit, Foguet Carles, Milà Enric, Papp Balázs, Josep J. Centelles, Cascante Marta. Cancer cell metabolism as new targets for novel designed therapies. Future Medicinal Chemistry 2014.

7. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T: Predicting selective drug targets in cancer through metabolic networks. Mol. Syst. Biol. 2011. 7(1).

8. Frezza C, Zheng L, Folger O et al.: Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. Nature 2011. 477(7363), 225-228.

9. Celià-Terrassa T, Meca-Cortés O, Mateo F, de Paz AM, Rubio N, Arnal-Estapé A, Ell BJ, Bermudo R, Díaz A, Guerra-Rebollo M, Lozano JJ, Estarás C, Ulloa C, Álvarez-Simón D, Milà J, Vilella R, Paciucci R, Martínez-Balbás M, de Herreros AG, Gomis RR, Kang Y, Blanco J, Fernández PL, Thomson TM. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. J Clin Invest. 2012 May 1;122(5):1849-68.

10. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novère N, Malys N, Mazein A, Papin JA, Price ND, Selkov E Sr, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, van Beek JH, Weichart D, Goryanin I, Nielsen J, Westerhoff HV, Kell DB, Mendes P, Palsson BØ. A community-driven global reconstruction of human metabolism. Nat Biotechnol. 2013 May;31(5):419-25.

11. Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. Network-based prediction of human tissue-specific metabolism. Nat Biotechnol. 2008 Sep;26(9):1003-10.

12. Esther Aguilar, Igor Marín, Erika Zodda, Vitaly Selivanov, Silvia Marín, Pedro de Atauri, Hossain Delowar, Mònica Pons, Antoni Celià-Terrassa, Óscar Meca-Cortés, Fionnuala Morrish, David Hockenbery, Josep J Centelles, Timothy M Thomson and Marta Cascante. Metabolic landscape and vulnerabilities of prostate metastatic epithelial cancer stem cell independent of epithelial-mesenchymal transition. Submited to Cancer Cell January 2015

13. Muir K, Hazim A, He Y, Peyressatre M, Kim DY, Song X, Beretta L. Proteomic and lipidomic signatures of lipid metabolism in NASH-associated hepatocellular carcinoma. Cancer Res. 2013 Aug 1;73(15):4722-31

14. Koontongkaew S, Monthanapisut P, Saensuk T. Inhibition of arachidonic acid metabolism decreases tumor cell invasion and matrix metalloproteinase expression. Prostaglandins Other Lipid Mediat. 2010 Nov;93(3-4):100-8

15. Liu SY, Zhang RL, Kang H, Fan ZJ, Du Z. Human liver tissue metabolic profiling research on hepatitis B virus-related hepatocellular carcinoma. World J Gastroenterol. 2013 Jun 14;19(22):3423-32.

16. Flowers M, Fabriás G, Delgado A, Casas J, Abad JL, Cabot MC. C6-ceramide and targeted inhibition of acid ceramidase induce synergistic decreases in breast cancer cell growth. Breast Cancer Res Treat. 2012 Jun;133(2):447-58.

17. Fritz V, Benfodda Z, Henriquet C, Hure S, Cristol JP, Michel F, Carbonneau MA, Casas F, Fajas L. Metabolic intervention on lipid synthesis converging pathways abrogates prostate cancer growth. Oncogene. 2013 Oct 17;32(42):5101-10.

18. Kaikkonen S, Paakinaho V, Sutinen P, Levonen AL, Palvimo J, Prostaglandin 15d-PGJ(2) inhibits androgen receptor signaling in prostate cancer cells. Mol

Endocrinol. 2013 Feb;27(2):212-23.

19. Kojima S, Sugiura T, Waku K, Kamikawa Y.. Contractile response to a cannabimimetic eicosanoid, 2-arachidonoylglycerol, of longitudinal smooth muscle from the guinea-pig distal colon in vitro. Eur J Pharmacol. 2002 May 31;444(3):203-7.

20. Nithipatikom K, Gomez-Granados AD, Tang AT, Pfeiffer AW, Williams CL, Campbell WB. Cannabinoid receptor type 1 (CB1) activation inhibits small GTPase RhoA activity and regulates motility of prostate carcinoma cells. Endocrinology. 2012 Jan;153(1):29-41.

21. Kasem Nithipatikom and William B. Campbell. Roles of Eicosanoids in Prostate Cancer. Future Lipidol. Aug 1, 2008; 3(4): 453–467.

22. Munetsuna E, Kawanami R, Nishikawa M, Ikeda S, Nakabayashi S, Yasuda K, Ohta M, Kamakura M, Ikushiro S, Sakaki T. Anti-proliferative activity of 25-hydroxyvitamin D3 in human prostate cells. Mol Cell Endocrinol. 2014 Feb 15;382(2):960-70

23. Trump DL, Muindi J, Fakih M, Yu WD, Johnson CS. Vitamin D compounds: clinical development as cancer therapy and prevention agents. Anticancer Res. 2006 Jul-Aug;26(4A):2551-6.

24. Pierron A, Le Pape E, Montaudié H, Castela E, De Donatis GM, Allegra M, Bertolotto C, Rocchi S, Cheli Y, Ballotti R, Passeron T. PGJ2 restores RA sensitivity in melanoma cells by decreasing PRAME and EZH2. J Dermatol Sci. 2014 Mar;73(3):258-61.

25. Collins TJ (July 2007). "ImageJ for microscopy". BioTechniques 43 (1 Suppl):

25–30.

26. Green DR, Galluzzi L, Kroemer G. Cell biology. Metabolic control of cell death. Science. 2014 Sep 19;345(6203):1250256.

27. Kim MH, Kim H. Oncogenes and tumor suppressors regulate glutamine metabolism in cancer cells. J Cancer Prev. 2013 Sep;18(3):221-6.

28. Granchi C, Fancelli D, Minutolo F. An update on therapeutic opportunities offered by cancer glycolytic metabolism. Bioorg Med Chem Lett. 2014 Sep 20. pii: S0960-894X(14)00983-4.

29. Hirsch HA, Iliopoulos D, Joshi A, Zhang Y, Jaeger SA, Bulyk M, Tsichlis PN, Shirley Liu X, Struhl K. A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. Cancer Cell. 2010 Apr 13;17(4):348-61.

30. Ogretmen B, Hannun YA. Biologically active sphingolipids in cancer pathogenesis and treatment. Nat Rev Cancer. 2004 Aug;4(8):604-16.

31. Camacho L, Meca-Cortés O, Abad JL, García S, Rubio N, Díaz A, Celià-Terrassa T, Cingolani F, Bermudo R, Fernández PL, Blanco J, Delgado A, Casas J, Fabriàs G, Thomson TM. Acid ceramidase as a therapeutic target in metastatic prostate cancer. J Lipid Res. 2013 May;54(5):1207-20.

32. Wranicz J, Szostak-Węgierek D. Health outcomes of vitamin D. Part I. Characteristics and classic role. Rocz Panstw Zakl Hig. 2014;65(3):179-84.

33. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density

oligonucleotide array probe level data. Biostatistics. 2003 Apr;4(2):249-64.

34. Joshua J. Hamilton, Jennifer L. Reed: Identification of Functional Differences in Metabolic Networks Using Comparative Genomics and Constraint-Based Models. PLoS ONE 2012, 7(4): e34670

35. Francisco Llaneras and Jesus Pico: A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. BMC Bioinformatics 2007, 8:421.

36. Jennifer L Reed and Bernhard Ø Palsson. Genome-scale in silico models of e. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res 2004, 14(9):1797–1805.

37. R. MahadevanÃ and C.H. Schilling: The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic Engineering 5 2003, 264–276.

38. Andreas Hoppe, Sabrina Hoffmann, Andreas Gerasch, Christoph Gille, Hermann-Georg Holzhütter: FASIMU: flexible software for flux-balance computation series in large metabolic networks. BMC Bioinformatics 2011, 12:28.

**Additional files**

Supplementary material 1: Metabolic reaction activity state

Supplementary material 2: Pathway activity state

Supplementary material 3: List of metabolites predicted to be consume and/or produced by PC-3 subpopulations

Supplementary material 4: Table of contingencies comparing predicted consumptions and productions with experimental measurements

Supplementary material 5: List of reactions active only in one subpopulation predicted by the robustness analysis

# Chapter 2b

## Metabolic landscape and vulnerabilities of prostate metastatic epithelial cancer stem cell independent of epithelial-mesenchymal transition

Esther Aguilar[1], **Igor Marín de Mas**[1], Erika Zodda[1,2], Vitaly Selivanov[1], Silvia Marín[1], Pedro de Atauri[1], Hossain Delowar[2], Mònica Pons[2], Antoni Celià-Terrassa[2], Óscar Meca-Cortés[2], Fionnuala Morrish[3], David Hockenbery[3], Josep J Centelles[1], Timothy M Thomson[2] and Marta Cascante[1]

**1** Department of Biochemistry and Molecular Biology, Faculty of Biology, University of Barcelona, Barcelona, Spain

**2** Department of Cell Biology, Molecular Biology Institute, National Research Council (IBMB-CSIC), Barcelona, Spain

**3** Fred Hutchinson Cancer Research Center, Seattle, WA, USA

# Metabolic landscape and vulnerabilities of prostate metastatic epithelial cancer stem cell independent of epithelial-mesenchymal transition

Esther Aguilar[1], **Igor Marín de Mas**[1], Erika Zodda[1,2], Vitaly Selivanov[1], Silvia Marín[1], Pedro de Atauri[1], Hossain Delowar[2], Mònica Pons[2], Antoni Celià-Terrassa[2], Óscar Meca-Cortés[2], Fionnuala Morrish[3], David Hockenbery[3], Josep J Centelles[1], Timothy M Thomson[2]* and Marta Cascante[1]*

**1** Department of Biochemistry and Molecular Biology, Faculty of Biology, University of Barcelona, Barcelona, Spain

**2** Department of Cell Biology, Molecular Biology Institute, National Research Council (IBMB-CSIC), Barcelona, Spain

**3** Fred Hutchinson Cancer Research Center, Seattle, WA, USA

* Corresponding authors: Marta Cascante, University of Barcelona, Diagonal 643, 08028, Barcelona, Spain, e-mail martacascante@ub.edu; and Timothy M Thomson, IBMB-CSIC, Baldiri Reixac 15-21, 08028 Barcelona, Spain, e-mail titbmc@ibmb.csic.es

**Summary**

In solid tumors, cancer stem cells (CSCs) can arise independently of epithelial-mesenchymal transition (EMT) but the metabolic reprogramming associated with independent CSC and EMT programs is poorly understood. Here we identify distinctive metabolic flux profiles associated with either metastatic prostate epithelial CSCs (e-CSCs) or non-CSCs expressing a stable EMT. e-CSCs displayed a marked Warburg effect, carbon source flexibility and vigorous serine, glycine, one-carbon (SGOC) metabolism. In contrast, non-CSCs showed stronger dependence on mitochondrial respiration, limited metabolic flexibility and high oxidative stress. Optimal growth of e-CSCs required the proton buffering capacity conferred by glutamine metabolism, while non-CSCs were more dependent on oxidative stress protective mechanisms. Importantly, a e-CSC-associated metabolic gene set correlates with metastatic progression in prostate cancer and 10 other tumor types.

**Introduction**

Tumors represent heterogeneous collections of neoplastic and non-neoplastic cells displaying divergent phenotypes that engage in complex interactions among them (Merlo et al., 2006). In any given tumor, two major functional categories of cancer cells are represented by cancer stem cells (CSCs), endowed with self-renewal and tumor repopulation potentials, and non-CSCs (Hanahan and Weinberg, 2011). These two categories are not mutually exclusive in that extensive plasticity allows phenotypic switching between CSC and non-CSC states (Chaffer et al., 2011). In order to consider tumor subpopulation-specific therapeutic strategies, it is important to define molecular wirings and biochemical networks underlying these two phenotypic categories, including their prevalent metabolic states.

Cancer cells predominantly produce energy via enhanced glycolysis regardless of whether they are under normoxic or hypoxic condition, reflective of a metabolic adaptation that has been designated as the Warburg effect (Koppenol et al., 2011; Warburg, 1956). In highly proliferating cells, glycolysis provides selective advantages over mitochondrial oxidation as a readily available source of energy and carbon units for of nucleotide, lipid and protein biosynthesis (de Souza et al., 2011; Deberardinis et al., 2008; Locasale and Cantley, 2010; Pfeiffer et al., 2001). The Warburg effect is also accompanied with high lactate production, allowing the regeneration of $NAD^+$, essential for glycolysis to proceed.

Differences in metabolic states between neoplastic and non-neoplastic tumor compartments and interactions among them have been extensively studied (Martinez-

Outschoorn et al., 2014). Recent evidence also suggests a metabolic homogeneity among all cancer cells within a given tumor, such that distinct cancer cell subpopulations display distinct metabolic states. Thus, it has been found that aerobic glycolysis and a concomitant reduction of mitochondrial activity is preferentially exhibited by CSCs (Loureiro et al., 2013). Others have emphasized the importance of oxidative phosphorylation (OXPHOS) and -oxidation in tumor-initiating cells (Viale et al., 2014). On the other hand, the induction of an epithelial-mesenchymal transition (EMT), which can lead to the acquisition of CSC properties by some cancer cells (Mani et al., 2008), has been associated with a reduced mitochondrial metabolism through the suppression of fructose-1,6-bisphosphatase (FBP1) expression leading to enhanced glycolysis and decreased OXPHOS (Dong et al., 2013) or alternative mechanisms (Lee et al., 2012; Lin et al., 2012), or. Others have shown that circulating cancer cells expressing mesenchymal and EMT traits activate mitochondrial biogenesis and OXPHOS through a marked upregulation of PGC-1 (LeBleu et al., 2014).

Therefore, a consensus is still lacking on the key characteristics that underlie cancer cell metabolic heterogeneity, in particular with regards to metabolic states distinctive of CSC and non-CSCs. In order to address this issue, we have performed a comprehensive metabolic characterization of a dual-cell prostate cancer cell model comprising one subpopulation (PC-3M) with features of epithelial CSCs (e-CSCs), and a second isogenic neoplastic subpopulation (PC-3S) of non-CSCs that display features of EMT (Celia-Terrassa et al., 2012). From this analysis, we conclude that the e-CSC subpopulation in our model exhibits a strong Warburg effect, in contrast to a stronger

dependence on mitochondrial respiration of our mesenchymal-like non-CSC subpopulation. In addition, our analysis uncovers selective metabolic dependencies and vulnerabilities that may open new avenues to the design of therapeutic strategies aimed at targeting specific cancer cell subpopulations displaying distinct CSC and non-CSC phenotypes.

**Results**

Prostate e-CSCs display an enhanced glycolytic metabolism

Recently, we have described a dual-cell model derived from the PC-3 prostate cancer cell line in which a highly metastatic subpopulation (PC-3M) is enriched in e-CSC features, whereas a non-metastatic but highly invasive subpopulation (PC-3S) lacks features of CSC and displays a mesenchymal-like phenotype and gene program (Celia-Terrassa et al., 2012). This model allows to interrogate whether metabolic phenotypes associated with CSC states require the engagement of an EMT, as found for other models (Dong et al., 2013; Lee et al., 2012; Lin et al., 2012; Viale et al., 2014).

We first compared the activity of the glycolytic pathway in our cell model. The extracellular acidification rate (ECAR), which predominantly reflects the production of lactic acid derived from glycolysis, was significantly higher in PC-3M than in PC-3S cells (Fig. 1A). Consistently, PC-3M cells consumed more glucose and produced more lactate than PC-3S cells (Fig. 1B) and exhibited a significantly higher lactate dehydrogenase (LDH) activity (Fig. 1C). Incubation of cells with [1,2-$^{13}C_2$]-glucose allowed to determine that both glycolysis and the pentose phosphate pathway (PPP)

contributed to the higher production of lactate  glycolysis in PC-3M cells (Fig. 1D and Table S1). We further assessed the importance of glycolysis in both cell types by studying the inhibition of mitochondrial cell respiration in the presence of high levels of glucose (Crabtree effect), the amplitude of which is an indicator of the degree of preference for glycolysis over oxidative phosphorylation (OXPHOS) as a pathway to obtain energy and metabolic precursors (Crabtree, 1929). A significantly stronger reduction in cell respiration was observed in PC-3M cells than PC-3S cells (Fig. 1E), indicating that the e-CSC subpopulation has a greater preference than the non-CSC subpopulation to metabolize glucose through the glycolytic pathway when this substrate is available. Glucose deprivation and treatment with the glycolytic inhibitor 2-deoxyglucose (2-DG) decreased the proliferation of PC-3M cells significantly more than that of PC-3S cells (Fig. 1F), partly contributed by cell death (Fig. S1A) and a block at the G1 phase of the cell cycle (Fig. S1B). On the other hand, treatment with the LDH inhibitor oxamate significantly affected both cell subpopulations (Fig. 1G) probably because of the disruption of the cytosolic $NAD^+$/NADH balance. Treatment with 2-DG decreased cellular ATP levels in both cell subpopulations but more significantly in PC-3M cells (Fig. 1H), whereas oxamate treatment tended to increase ATP production (Fig. 1H). The latter result suggests that LDH inhibition directs pyruvate towards energetic pathways, such as mitochondrial respiration, characterized by a high ATP yield. Indeed, we observed an additive effect when cells were co-treated with oxamate and dichloroacetate (DCA), an inhibitor of pyruvate dehydrogenase kinases (PDHKs) that promotes pyruvate dehydrogenase (PDH) activity (Fig. 1H),

which led to a 10-20% increase in ATP levels.

When PC-3M cells were induced to acquire a partial EMT through forced overexpression of Snai1 (PC-3M/Snai1 cells), or when their self-renewal properties were inhibited through knockdown of the pluripotency factors SOX2, KLF4 and MYC (PC-3M/SKMkd cells) (Celia-Terrassa et al., 2012), they showed a reduced glycolytic phenotype (Fig. 1I). Likewise, exposure of cells to 2-DG affected to a lesser extent PC-3M/Snai1 and PC-3M/SKMkd cells as compared to control PC-3M cells (Fig. 1J). These observations reinforce the conclusion that the marked glycolytic phenotype and glucose dependence observed in PC-3M cells is associated with e-CSC properties which in our model requires the maintenance of a robust epithelial gene program.

Mesenchymal-like non-CSCs exhibit higher rates of mitochondrial respiration than e-CSCs

The above results suggest the occurrence of a marked Warburg effect in PC-3M cells. For a more formal conclusion in this regard, we explored the state of mitochondrial respiration in our cell model by using four mitochondrial respiratory chain inhibitors (Fig. S2). Oligomycin, an ATP synthase inhibitor, caused a more significant reduction in OCR in PC-3S cells (Fig. 2A), suggesting that they rely more on mitochondrial metabolism for bioenergetics than the e-CSC subpopulation. Oligomycin also dose-dependently inhibited the proliferation of PC-3S cells significantly more than that of PC-3M cells (Fig. 2B) and elicited increased ATP levels in PC-3S cells but not in PC-3M cells (Fig. 2C). Thus, in spite of its strong dependence on mitochondrial ATP, PC-

3S cells readily activated compensatory ATP biosynthetic pathways, such as glycolysis, in response to oligomycin.

We next compared the maximal mitochondrial respiratory capacity of these cells by measuring the OCR response to trifluorocarbonylcyanide phenylhydrazone (FCCP), a mitochondrial uncoupling agent. To better assess the mitochondrial respiratory capacity, we injected 2 mM pyruvate (Pyr) together with FCCP. Exposure to FCCP+Pyr boosted OCR significantly more in PC-3S cells than in PC-3M cells (Fig. 2D), suggesting that PC-3S cells are endowed with a significantly higher mitochondrial respiratory capacity (Fig. 2E). Exposure to the mitochondrial complex I and III inhibitors rotenone and antimycin reduced OCR in both cell subpopulations, although significantly more in PC-3S cells (Fig. 2F).

Consistent with the known correlation of mitochondrial metabolic activity with the generation of reactive oxygen species (ROS) (Starkov, 2008), PC-3S cells contained higher basal levels of ROS than PC-3M cells (Fig. 2G). The proton leak rate across the mitochondrial membrane completes the proton circuit and behaves as an antioxidant by reducing ROS production. Accordingly, PC-3S cells presented higher rates of proton leak than PC-3M cells (Fig. 2H), likely allowing them to counter their higher intracellular ROS levels.

In additional support of the importance of the epithelial and pluripotency gene programs for these metabolic readouts, PC-3M/Snai1 and PC-3M/SKMkd variants were more sensitive to all mitochondrial drugs (Fig. 2I). We thus conclude that PC-3M

cells, enriched in e-CSCs, are characterized by a high glycolytic flux and relatively low mitochondrial respiration, for which these cells depend on the expression of epithelial and pluripotency gene networks.

These cell subpopulations showed contrasting mitochondrial morphologies and organizations (Fig. S3). PC-3M cells exhibited mitochondrial morphologies consistent with extensive mitochondrial fragmentation or fission, which has been associated with impaired OXPHOS and a higher reliance on anaerobic glycolysis to generate energy (Sauvanet et al., 2010). In contrast, the mitochondrial morphologies in PC-3S cells were more compact, consistent with active fusion events (Westrate et al., 2014), a process coupled to a robust mitochondrial function and bioenergetics (Mishra et al., 2014). Thus, the non-CSC subpopulation displays a mitochondrial organization predicted to support a robust OXPHOS activity for the generation of ATP.

e-CSCs use alternative substrates to feed into the mitochondrial metabolism

Glucose deprivation and 2-DG treatment enhanced glutamine consumption more significantly in PC-3M cells than in PC-3S cells (Fig. 3A), indicating that PC-3M cells are endowed with a greater flexibility in the use of carbon sources other than glucose. This was further explored by determining OCR and ECAR profiles in full media (glucose plus glutamine) and in restricted media lacking either glucose or glutamine or both (minimal media). Although significant differences were not detected in the OCR baseline profile in full media conditions, PC-3M showed higher OCR values than PC-3S cells in all restricted media conditions (Fig. 3B), suggesting that they more readily

use alternative carbon sources to feed into mitochondrial metabolism. This also indicates that, although PC-3M cells are more dependent on glycolysis, their mitochondrial respiration is functional and enhanced when required. Likewise, ECAR (and thus glycolysis) was strongly dependent on glucose availability (Fig. 3C), supporting the importance of a switch to mitochondrial metabolism under glucose deprivation. To test the ability of PC-3M cells to use alternative substrates contributing to OCR, we analyzed the consumption of ketogenic amino acids, which can be degraded into acetyl-CoA. PC-3M cells consumed more glutamine (Fig. 3A) and the ketogenic amino acids leucine, isoleucine, lysine, threonine, tyrosine, tryptophan and phenylalanine than PC-3S cells (Fig. 3D). Additionally, we analyzed fatty acid -oxidation as a contributor to mitochondrial respiration by exposing cells to etomoxir (a carnitine palmitoyltransferase 1 (CPT1) inhibitor), which impairs fatty acid -oxidation and causes a reduction in OCR. We found that treatment with etomoxir affected the mitochondrial respiration of PC-3M cells significantly more than that of PC-3S cells (Fig. S4A). Consistently, higher protein levels of the fatty acid transporter CPT1 were found in PC-3M cells than in PC-3S cells (Fig. S4B).

In agreement with the observed OCR profiles, the sensitivity of the cells to oligomycin, FCCP + Pyr and rotenone + antimycin varied as a function of the available carbon source. With glucose as the only carbon source, PC-3S cells were relatively less sensitive than PC-3M cells to inhibition of mitochondrial respiration in response to oligomycin (Fig. 3E) and also compared to full media conditions (Fig. 2A), likely reflecting a decrease in their mitochondrial capacity in the absence of glutamine (Fig.

3B). Neither cell subpopulation showed significant differences in their responses to FCCP + Pyr and rotenone + antimycin under glucose or glutamine deprivation (Fig. 3E). Together, these observations indicate that PC-3M cells can use alternative substrates to fuel mitochondrial respiration in the absence of glucose, while PC-3S cells are highly dependent on glucose and glutamine to maintain a proper mitochondrial function.

Glucose and glutamine differentially contribute to TCA intermediates synthesis in e-CSCs and mesenchymal-like non-CSCs

The above evidences of differential mitochondrial metabolism were further investigated by analyzing the TCA metabolism through a $^{13}C$-tracer based approach. PC-3M and PC-3S cells were cultured with [1,2-$^{13}C_2$]-glucose and the mass isotopomer distribution of the intracellular TCA intermediates citrate, glutamate, fumarate, malate and aspartate was analyzed by GC/MS (Table S2). PC-3S cells incorporated more efficiently than PC-3M cells labeling from glucose to TCA intermediates, indicating that the CSC subpopulation, but not the non-CSC subpopulation, may divert most of the pyruvate derived from glycolysis away from mitochondria, consistent with a marked Warburg effect.

We also determined the mass isotopomer distribution of intracellular pyruvate and calculated the ratio of m2 labeling of TCA intermediates to m2 pyruvate as an indication of the relative rates of pyruvate oxidation in the TCA cycle. In this approach, we do not distinguish the enzyme (e.g. pyruvate dehydrogenase (PDH), pyruvate

carboxylase or malic enzyme) responsible for the incorporation of pyruvate to the TCA cycle (Fig. 4A). We found lower ratios of m2 TCA intermediates to m2 pyruvate in PC-3M cells (Fig. 4B), indicative of a decreased fraction of pyruvate entering the TCA cycle. Furthermore, entry of a second m2 acetyl-CoA from m2 pyruvate into the TCA cycle, specifically by the action of PDH, results in m4 citrate, glutamate and m3 fumarate, malate and aspartate (Fig. 4C). As with the m2/m2 pyruvate ratio, m4 (citrate and glutamate) and m3 (fumarate, malate and aspartate) to m2 pyruvate ratios were also decreased in PC-3M cells relative to PC-3S cells (Fig. 4D). These results show that the flux from glucose to the oxidative TCA cycle is stronger in PC-3S cells than PC-3M cells.

The contribution of glutamine to the TCA cycle was determined by analysis of the labeling profile of TCA intermediates after incubation of the cells with [U-$^{13}C_5$]-glutamine (Table S3), comparing m4 mass isotopomers (m5 in the case of glutamate) and considering the oxidative pathway (Fig S5A). We found a more significant contribution of this substrate to the synthesis of TCA intermediates in PC-3M cells than in PC-3S cells (Fig. 4E). We also considered whether glutamine was metabolized through reductive carboxylation reactions (Fig. S5B). Analysis of the levels of m5 citrate and m3 aspartate, malate and fumarate indicated that this metabolic pathway is more active in PC-3M cells (Fig 4F). These and the above results support the notion that the e-CSCs of our cell model compensate their reduced mitochondrial function by enhancing alternative metabolic pathways to sustain their needs and rapidly adapt to changes in nutrient availability. Interestingly, we also noticed that PC-3M cells

accumulated proline in the extracellular media (Fig. S6A) and that glutamine contributes significantly to the synthesis of this amino acid, especially in PC-3M cells (Fig. S6B).

PDH is a key enzyme mediating the entry of pyruvate to mitochondria. Treatment of cells with the PDHKs inhibitor DCA and the LDH inhibitor oxamate resulted in a more pronounced increase in mitochondrial respiration in PC-3M cells, particularly in response to DCA (Fig. 4G-H), indicating that activating PDH (through PDHKs inhibition) is more effective than increasing pyruvate levels (through LDH inhibition) to stimulate the mitochondrial entry and metabolism of pyruvate. These treatments boosted the respiration of PC-3S cells significantly less than that of PC-3M cells, suggesting a higher PDH baseline activity in PC-3S. Indeed, PDH was markedly more phosphorylated (i.e., more inactive) in PC-3M than PC-3S cells (Fig. 4I, inset) and transcript levels for PDHK1 were higher and the PDH phosphatase PDP2 lower in PC-3M cells than in PC-3S cells (Fig. 4I).

We next used Isodyn (Selivanov et al., 2006) to integrate the above biochemical and $^{13}C$-based mass isotopomer distributions and estimate metabolic flux distributions (relative to glucose uptake) in central carbon metabolism (Fig. S7). The resulting metabolic flux simulations lend further support to the occurrence of an enhanced entry of pyruvate *via* PDH into the TCA cycle, increased TCA cycle fluxes (CS, citakg and akgfum) and a higher flux associated with mitochondrial respiration (resp) in PC-3S cells, consistent with a more active mitochondrial metabolism (Fig. 4J).

Prostate e-CSCs are more dependent than mesenchymal-like non-CSCs on non-anaplerotic metabolic contributions from glutaminolysis

In light of the essential contribution of glutamine to the TCA cycle in PC-3M cells, we further explored the effects of inhibiting glutaminolysis on cell proliferation, viability and metabolic adaptations. Glutaminase (GLS1), a MYC transcription target frequently expressed at high levels in rapidly proliferating cells and CSCs (Gao et al., 2009; Liu et al., 2012), was expressed at significantly higher levels in PC-3M cells than PC-3S cells, in particular the glutaminase C (GAC) isoform (Fig. 5A and B). The expression of GAC in PC-3M cells was partly dependent on the maintenance of epithelial and pluripotent gene programs, as shown by a diminished GAC/KGA ratio in PC-3M/Snai1 and PC-3M/SKMkd cells (Fig. 5B). To assess the importance of glutaminase in cell proliferation, we incubated cells with the glutaminase inhibitor bis-2-(5-phenylacetamido-1,3,4-thiadiazol-2-yl)ethyl sulfide (BPTES). Cell proliferation was more significantly inhibited in PC-3M cells (Fig. 5C), with a decline of cells in the S phase of the cell cycle (Fig. S8). The sensitivity of PC-3M cells to growth inhibition by BPTES was attenuated by overexpression of Snai1 or knock down of SOX2, KLF4 and MYC (Fig. 5C), illustrating that the strong dependence on glutaminolysis of PC-3M cells relies on the maintenance of epithelial and pluripotency gene programs.

Both PC-3M and PC-3S cells adapted their metabolism to BPTES treatment by increasing glucose consumption (Fig. 5D) and, more significantly, lactate production (Fig. 5E). This prompted us to test the effects on viability of a combined inhibition of

glutaminolysis and glycolysis. We observed a marked additive effect on cell death by combining BPTES and 2-DG treatments in PC-3M cells but not in PC-3S cells (Fig. 5F), indicating that glycolysis and glutaminolysis are indeed two major metabolic pathways essential for the optimal growth and survival of PC-3M cells.

Because glutamine is anaplerotically used to supply the TCA cycle (DeBerardinis et al., 2008), we tested if the inhibition of cell proliferation by BPTES could be rescued with the cell-permeable □-ketoglutarate (□-KG) analog dimethyl □-ketone (DMK). Remarkably, cell growth was not rescued in BPTES-treated PC-3M or PC-3S cells after incubation with DMK (Fig. 5G). This suggests that, in these cells, glutamine exerts an activity independent of its anaplerotic function, essential for optimal cell proliferation. In addition to anaplerotic reactions, glutamine is involved in numerous metabolic pathways, including synthesis of glutathione, NADPH production and pH homeostasis (Huang et al., 2013; Son et al., 2013). In their basal states, PC-3S cells contained higher levels of ROS than PC-3M cells (Fig. 2G) and total intracellular glutathione (Fig. 5H), suggesting that the non-CSC subpopulation counteracts higher levels of ROS with an accumulation of glutathione. Although ROS production increased in both cell subpopulations after BPTES treatment (Fig. 5I), glutathione content was affected only in PC-3S cells (Fig. 5J), showing that, in these cells, the glutamine metabolism is indeed important for the synthesis of glutathione. Because both cell lines showed comparable increments in ROS levels in response to BPTES, these results fail to explain the observed differential sensitivity of PC-3M cells to glutaminase inhibition.

Recent evidences indicate that glutamine metabolism provides cellular resistance to acidic conditions through the release of ammonia (Huang et al., 2013). After incubation in pH 7.1 media (acidic conditions), the proliferation of PC-3M cells decreased significantly more than PC-3S cells, compared to cells grown in standard pH 7.9 media (Fig. 5K). We observed an additive growth inhibitory effect between low pH and BPTES, more pronounced in the case of PC-3M cells (Fig. 5K). Absence of sodium bicarbonate ($NaHCO_3$) in the culture medium was also deleterious to cell proliferation under acidic conditions (Fig. 5K). The importance of the epithelial and pluripotency gene programs in conferring sensitivity to glutaminase inhibition under relatively acidic conditions was shown by the enhanced resistance to BPTES treatment at low pH exhibited by PC-3M/Snai1 and PC-3M/SKMkd cells relative to control PC-3M cells (Fig. 5L). These observations suggest that the glutaminase reaction contributes to the buffering of surplus protons produced by PC-3M cells through their marked Warburg effect and consequent lactate production.

e-CSCs are characterized by a strong serine, glycine, one-carbon (SGOC) metabolism

Changes in cellular demands and preferences for amino acids result from the metabolic reprogramming associated with tumor progression and are closely linked to the Warburg effect (Hsu and Sabatini, 2008; Jain et al., 2012). The extensive differences found between our two subpopulations in glycolysis, mitochondrial metabolism and glutaminolysis led us to expect equally significant differences in amino acid metabolism. We analyzed the consumption and production profiles of alanine, glutamate, serine and glycine and their mass isotopomer distributions after incubating

cells with [1,2-$^{13}C_2$]-glucose (Table S2) or [U-$^{13}C_6$]-glucose (Table S4). PC-3M cells secreted more alanine than PC-3S cells (Fig. 6A), m2 alanine being the most abundant isotopologue (Fig. 6B), indicating that it is mainly of glycolytic origin. Consistently, PC-3M cells showed higher alanine transaminase (ALT) enzyme activity than PC-3S cells (Fig. 6C). In contrast, PC-3S cells secreted significantly more glutamate to the extracellular media than PC-3M cells (Fig. 6D) comprising both m2 glutamate (derived from glycolysis + 1$^{st}$ turn of the TCA cycle) and m3 glutamate (glycolysis + 2$^{nd}$ turn of the TCA cycle) (Fig. 6E). Figure S9 schematically illustrates the label distributions incorporated in glutamate from [U-$^{13}C_6$]-glucose.

The glycolytic flux can also be diverted from the glycolytic intermediate 3-phosphoglycerate towards the synthesis of serine, which in turn may be converted to glycine (Locasale, 2013). Both amino acids provide essential precursors for the synthesis of proteins, nucleic acids and lipids (Fan et al., 2014). PC-3M cells consumed more extracellular serine than PC-3S cells, while glycine was consumed by PC-3M cells and produced by PC-3S cells (Fig. 6F). After incubation of cells with [U-$^{13}C_6$]-glucose (Fig. 6G), m3 serine was predominant in PC-3M cells, whereas more evenly distributed representations of m1, m2 and m3 serine were found in PC-3S cells (Fig. 6G). The largely homogeneous m3 labeling of serine in PC-3M cells, together with a paucity of m1 and m2 labeled species, suggests that serine does not accumulate in these cells and is efficiently used as a building block for protein synthesis and/or funneled to catabolic reactions that rapidly transform serine to glycine. In line with this, the failure to secrete glycine (Fig. 6F) and the relatively low levels of m2 glycine

in PC-3M cells (Fig. 6H) suggest that glycine derived from serine is also rapidly used in these cells for biosynthetic purposes or cleaved to further contribute to one-carbon metabolic reactions. In fact, PC-3M cells expressed significantly higher levels than PC-3S cells of glycine decarboxylase (GLDC), the key enzyme catalyzing the glycine cleavage (Fig. 6I). In addition to GLDC, PC-3M cells also cells expressed higher levels than PC-3S cells of many of the other enzymes involved in serine and one-carbon metabolism (Fig. 6I) (Locasale, 2013; Tedeschi et al., 2013; Vazquez et al., 2011).

NADPH-generating reactions and fatty acid synthesis are favored in mesenchymal-like non-CSCs

The oxidative branch of the PPP uses glucose-6-phosphate as a substrate to generate NADPH, thus providing reducing power for other biosynthetic pathways (e.g. fatty acid synthesis) and to counter free radicals and oxidative stress. The non-oxidative branch recycles pentose phosphates to glycolytic intermediates and generates *de novo* ribose-5-phosphate for nucleotide synthesis. Two key PPP enzymes are glucose-6-phosphate dehydrogenase (G6PDH) in the oxidative branch and transketolase (TKT) in the non-oxidative branch. The metabolism of [1,2-$^{13}C_2$]-glucose through the oxidative and the non-oxidative branches of PPP results mainly in m1 and m2 labeled ribose, respectively. Label incorporation to ribose was significantly higher in PC-3M cells than in PC-3S cells (Fig. 7A), suggestive of a higher demand for nucleotide biosynthesis to sustain their high proliferation rate. Analysis of ribose mass isotopomer distribution revealed an increase in m1 and m2 ribose in PC-3M cells (Fig. 7B). The m1/m2 ribose ratio also indicated a differential contribution of the oxidative and non-oxidative

branches of the PPP (Fig. 7C), suggesting that the highly glycolytic PC-3M cells redirect back part of the glucose-based PPP intermediates to glycolysis through the non-oxidative branch. Although direct glycolysis was the main source of lactate in both cell subpopulations, lactate can be derived from a combination of the PPP and glycolysis, and this contribution was indeed higher in PC-3M cells (Fig. 1D). The differential use of the PPP branches was further supported by the observation of significantly higher enzymatic activity and expression levels of TKT in PC-3M cells and G6PDH in PC-3S cells (Fig. 7D-F).

The above results suggest a higher demand of NADPH in PC-3S cells, likely required to face higher levels of oxidative stress derived from a strong mitochondrial metabolism. Isodyn analysis indicated that the fluxes for NADPH-producing reactions, such as cytosolic malic enzyme (ME) and isocitrate dehydrogenase (participating in the citakg flux), are more robust in PC-3S cells (Fig. 4J). Isodyn also predicted an enhanced efflux of citrate from the mitochondrial to the cytosol compartment (citdmc flux) (Fig. 4J). NADPH reducing equivalents actively participate in many biosynthetic reactions, such as fatty acid synthesis. After incubation of cells with [1,2-$^{13}C_2$]-glucose, a significantly greater yield of m2 and m4 labeled palmitate (Fig. 7G) and stearate (Fig. 7H) was detected in PC-3S cells than PC-3M cells, suggesting that glucose-derived carbons are more efficiently routed towards fatty acid synthesis in PC-3S cells. We also found higher protein expression levels in PC-3S cells of ATP citrate lyase (ACLY), the primary enzyme responsible for cytosolic fatty acid biosynthesis (Fig. 7I). Together, these observations suggest that PC-3M and PC-3S cells differ in the balance

between fatty acid synthesis and oxidation, the former being favored in the non-CSC subpopulation and the latter in the e-CSC subpopulation.

A metabolic gene set overrepresented in PC-3M cells is associated with the malignant progression of several cancers

To determine the potential significance in cancer progression of metabolic pathways differentially active in our CSC *vs*. non-CSC model, we used our previously generated transcriptomic data (Celia-Terrassa *et al.*, 2012) to extract a set of metabolic genes significantly enriched in PC-3M cells relative to PC-3S cells (Table S5). This gene set includes enzymes involved in the metabolism of serine and glycine (GLDC, CBS), branched-chain amino acids (BCAT1), glutamate and proline (ASS1, AGMAT), which support the metabolic flexibility accompanying the CSC phenotype. It also includes genes that participate in the synthesis of purine nucleotides (GUCY1A3, PDE3B and ADCY7) as part of reactions that converge on substrates that are actively metabolized in PC-3M cells, such as glycine and one-carbon units from tetrahydrofolate. By applying Gene Set Enrichment Analysis (Subramanian *et al.*, 2005), we found that the PC-3M-overrepresented metabolic gene set is significantly enriched along malignant progression in prostate cancer (Fig 7J, Table S6) and 10 other types of human tumors, including ovarian, bladder, adrenocortical, head and neck, stomach or rectal carcinoma, melanoma and mesothelioma (Table S6). These observations support the relevance for human tumor progression of the metabolic pathways and reactions that we have found differentially active in our CSC *vs*. non-CSC cell model.

**Discussion**

The cancer stem cell hypothesis implies intratumoral phenotypic heterogeneity, to which epigenetic mechanisms are major contributors (Easwaran et al., 2014). Neoplastic cells can reversibly switch between CSC and non-CSC states (Chaffer et al., 2011), presenting a hurdle to efforts at targeting specific tumor subpopulations that may thus escape from therapeutic pressure (Easwaran et al., 2014). A successful strategy to forestall these escape mechanisms should contemplate the characterization of biochemical and gene networks and uncover vulnerabilities pertaining to both CSC and non-CSC neoplastic tumors components. Previous studies have failed to discriminate whether metabolic phenotypes observed in tumor cells with CSC features were due to the EMT (Dong et al., 2013; Le et al., 2012; LeBleu et al., 2014; Lee et al., 2012; Lin et al., 2012) or to CSC states, which recent studies have found that may occur in cancer cells unlinked to EMT (Beck et al., 2015; Celia-Terrassa et al., 2012). In order to approach this issue, we have undertaken a systematic comparative metabolic characterization of a dual-cell prostate cancer model in which CSC (PC-3M) and EMT (PC-3S) phenotypes and gene programs are clearly uncoupled. Our comparative analysis sheds light on major metabolic pathways and vulnerabilities accompanying CSC and non-CSC phenotypes, as schematically summarized in Figures S10 and S11

PC-3-derived e-CSCs are under an enhanced Warburg effect and use a diversity of

carbon sources for energy production

Our analysis shows that the PC-3M e-CSC subpopulation is highly dependent on aerobic glycolysis for cell proliferation and bioenergetics, largely uncoupled from the TCA cycle and OXPHOS. In contrast, the mesenchymal-like non-CSC PC-3S subpopulation relies more on mitochondrial respiration for the generation of energy and metabolizes significant amounts of glucose through the TCA cycle. However, in nutrient-restricted media, mitochondrial metabolism was compromised in PC-3S cells whereas PC-3M cells showed higher respiration rates thanks to their greater metabolic flexibility and ability to use alternative metabolic substrates to feed into the TCA cycle.

On these grounds, a promising strategy to target CSC subpopulations would be either to block glycolysis (e.g. 2-DG) or to promote OXPHOS via PDH activation with DCA to switch from glycolysis to OXPHOS and dramatically alter their bioenergetic and biosynthetic resources and their redox balance (ROS) (Ito and Suda, 2014). However, due to the metabolic flexibility displayed by PC-3M cells, simultaneous inhibition of glycolysis and alternative carbon sources, such as glutamine, may be required to efficiently target CSCs and thus substantially reduce their potential to proliferate and survive *in vitro* and limit their metastatic potential *in vivo* (Berridge et al., 2010). In contrast, PC-3S cells are more vulnerable to the inhibition of mitochondrial respiration by several mitochondrial drugs.

Together with published evidences, our observations lead us to conclude that cancer cells with properties of CSCs engage a Warburg effect, independently of whether their

phenotype is associated with an epithelial (our model) or a mesenchymal (other models) gene program. High glycolytic fluxes likely benefit cells with elevated proliferation rates through the production of glycolytic intermediates for *de novo* biosynthesis of nucleotides, lipids and proteins (Hamanaka and Chandel, 2012). Previous studies have suggested that glycolytic metabolism may be a broadly conserved stem cell property (Feng et al., 2014) and associated with undifferentiated states (Pacini and Borziani, 2014; Rehman, 2010).

Prostate e-CSCs exhibit a very active SGOC metabolism

We have also found that part of the glycolytic carbons was diverted toward serine and glycine biosynthesis, and PC-3M cells failed to accumulate these amino acids, suggestive of a high demand for serine and glycine for biosynthesis and one-carbon metabolism (Vazquez et al., 2011) (Tedeschi et al., 2013). PC-3S cells may present a more limited use of serine and glycine as shown by accumulation of m1, m2 and m3 serine and excretion of surplus glycine. A hypothesis of the reactions that drive serine and glycine metabolism is graphically summarized in Figure S12.

Non-anaperotic functions of glutamine metabolism support optimal proliferation of prostate epithelial CSCs

Glutaminolysis normally serves to anaplerotically fill the TCA, in particular in cells that shut down the entry of pyruvate into mitochondria as a consequence of PDH inhibition (DeBerardinis et al., 2008). However, the compromised growth of PC-3M cells treated with BPTES was not rescued by incubation with a cell-permeable -KG

analog, which suggests that the dependence of these cells on glutaminolysis is not explained solely through the anaplerotic functions of glutamine.

Of the different non-anaplerotic functions investigated for glutamine metabolism, its protective role from acidity was the most remarkable in PC-3M cells. We found that the growth of PC-3M cells was significantly more affected than PC-3S cells in relatively low pH conditions. Furthermore, we observed a synergistic effect between glutaminase inhibition by BPTES and low pH. We conclude that protection from the acidic conditions caused by a marked Warburg effect is a critical survival factor for PC-3M cells. Figure S13 illustrates other cellular reactions that also release ammonia, such as those driven by glutamate dehydrogenase (GLDH) and the glycine cleavage system. Targeting these enzymes or multienzyme complexes could additionally compromise the proliferation and viability of PC-3M cells and, by extension, CSCs, not only because of their specific role in metabolic networks but also of the consequent compromise of the cells pH buffering power. In contrast, glutamine metabolism in PC-3S cells was more closely linked to glutathione synthesis that confers protection from oxidative stress and therefore reasonable strategies to target non-CSCs with similar metabolic phenotypes might include interventions to limit antioxidant adaptive responses.

Enhanced flux of NADPH-generating reactions and fatty acid synthesis in prostate neoplastic mesenchymal-like non-CSCs

Our comparative analysis of the PPP has led us to uncover a differential use of the

oxidative and non-oxidative branches of the PPP, as summarized in Figures S10 and S11. PC-3M cells had a more active non-oxidative branch of the PPP, which may present an advantage for their high glycolytic phenotype, allowing a reversible and dynamic connection between glycolysis and the PPP. In contrast, PC-3S cells showed a preference for the oxidative branch, an important source of NADPH to support fatty acid synthesis and redox homeostasis maintenance. PC-3S cells also had a more pronounced fatty acid synthetic activity than PC-3M cells, in which fatty acid oxidation was more prevalent. From these observations, we predict a higher sensitivity of the non-CSC PC-3S cell subpopulation to perturbations in lipogenesis than the e-CSC PC-3M subpopulation.

In summary, we conclude that metabolic programs underlying CSC and EMT phenotypes can be mutually uncoupled and that each associates with preferred metabolic dependencies required for optimal cell proliferation and survival. The recognition of such dependencies leads to uncover metabolic vulnerabilities that can be targeted through tumor subpopulation-specific interventions. Our observations may establish new grounds for tumor subpopulation-specific and combined therapeutic strategies with potential impacts on the management of metastatic disease. Importantly, the relevance for tumor progression of metabolic pathways related to those detailed for our cell model is supported by a significant association with the malignant progression of several types of human tumors of sets of metabolic genes enriched in PC-3M cells.

**Experimental Procedures** (additional Experimental Procedures are described in online Supplemental Information)

Cells and Cell culture. PC-3M and PC-3S were clonally derived from the human cell line PC-3 and PC-3M cell variants used in several experiments were obtained as described (Celia-Terrassa et al., 2012). All cell subpopulations were cultured at 37°C in a 5% $CO_2$ atmosphere in RPMI 1640 media (Sigma-Aldrich or Biowest) with 10 mM glucose and 2 mM glutamine supplemented with 10% Fetal Bovine Serum (FBS) (PAA Laboratories), 1% pyruvate (1 mM) (Biological Industries), 1% streptomycin (100 µg/mL) / penicillin (100 units/mL) (Gibco), 1% nonessential amino acids (Biological Industries) and 200 µg/mL geneticin (Gibco).

Oxygen consumption rate (OCR) and extracellular acidification rate (ECAR). A XF24 Extracellular Flux Analyzer (Seahorse Bioscience) was used to measure the rate change of dissolved oxygen and protons in media immediately surrounding adherent cells cultured in a XF24-well microplate (Seahorse Bioscience). Cells were seeded in XF24-well microplates and incubated at 37 °C with 5% $CO_2$ overnight. One h before assays, growth media was replaced by basal media (unbuffered DMEM; Sigma-Aldrich) with or without supplements. The sensor cartridge was loaded with the test agents (Sigma-Aldrich) and calibrated prior to the start of the assays. Responses to the different treatments are expressed as $Log_2$ of the fold change comparing the measured point immediately after and before the corresponding injection.

$^{13}$C-tracer-based metabolomics. Culture media was replaced by fresh media containing either 10 mM 100% or 10 mM 50% [1,2-$^{13}C_2$]-glucose (Sigma-Aldrich), 10 mM 100%

[U-$^{13}C_6$]-glucose (Sigma-Aldrich) or 2 mM 100% [U-$^{13}C_5$]-glutamine (Sigma-Aldrich) and media, pellets and cell cultured plates collected and stored frozen until analysis. Analysis of $^{13}C$-labeled extracellular and intracellular metabolites for mass isotopomer distribution was done by gas chromatography coupled to mass spectrometry (GC/MS) on an Agilent 7890A GC instrument equipped with a HP5 capillary column connected to an Agilent 5975C MS. Fatty acids were analyzed with a GCMS-QP 2012 Shimadzu instrument equipped with a bpx70 (SGE) column. In all cases, 1 μL of sample was injected at 250 °C using helium as a carrier gas at a 1 mL/min flow rate.

Statistical analysis. Two-tailed Student's t-test for independent samples was applied where appropriate. In figures, bars represent mean ± standard deviation (SD), * indicates significance at p-value < 0.05, ** at $p < 0.01$ and *** at $p < 0.001$.

**Author Contributions**

EA designed, performed and analyzed most experiments, wrote the draft and edited versions of the manuscript and generated illustrations for the figures; IM, VS, PdA and JJC analyzed metabolic and expression data and participated in the design of specific experiments; EZ, HD and MP performed additional experiments; AC-T and OM-C supplied the cell model and variants and assisted in their characterization; FM and DH were instrumental with Seahorse experiments; TMT designed and supervised the overall study and specific experiments, analyzed the data, prepared the figures and wrote the draft and final versions of the manuscript; MC designed and supervised the overall study and specific experiments including data analysis and wrote and approved

manuscript drafts and final versions.

**Acknowledgments**

We are grateful to..... EA was the recipient of..., IM...., EZ....This work was supported by grants to MC from xxxxx, yyy and zzz, and to TMT from MICINN (SAF2011-24686), MINECO (SAF2012-40017-C02-01), Catalan Agència d'Ajuts Universitaris i de Recerca (AGAUR; 2009SGR1482), and Xarxa de Referència en Biotecnologia.

**References**

Beck, B., Lapouge, G., Rorive, S., Drogat, B., Desaedelaere, K., Delafaille, S., Dubois, C., Salmon, I., Willekens, K., Marine, J.C., *et al.* (2015). Different levels of twist1 regulate skin tumor initiation, stemness, and progression. Cell Stem Cell *16*, 67-79.

Celia-Terrassa, T., Meca-Cortes, O., Mateo, F., de Paz, A.M., Rubio, N., Arnal-Estape, A., Ell, B.J., Bermudo, R., Diaz, A., Guerra-Rebollo, M., *et al.* (2012). Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. J Clin Invest *122*, 1849-1868.

Chaffer, C.L., Brueckmann, I., Scheel, C., Kaestli, A.J., Wiggins, P.A., Rodrigues, L.O., Brooks, M., Reinhardt, F., Su, Y., Polyak, K., *et al.* (2011). Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A *108*, 7950-7955.

Crabtree, H.G. (1929). Observations on the carbohydrate metabolism of tumours. Biochem J *23*, 536-545.

DeBerardinis, R.J., Lum, J.J., Hatzivassiliou, G., and Thompson, C.B. (2008). The

biology of cancer: metabolic reprogramming fuels cell growth and proliferation. Cell Metab *7*, 11-20.

Dong, C., Yuan, T., Wu, Y., Wang, Y., Fan, T.W., Miriyala, S., Lin, Y., Yao, J., Shi, J., Kang, T., *et al.* (2013). Loss of FBP1 by Snail-mediated repression provides metabolic advantages in basal-like breast cancer. Cancer Cell *23*, 316-331.

Easwaran, H., Tsai, H.C., and Baylin, S.B. (2014). Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Mol Cell *54*, 716-727.

Fan, J., Ye, J., Kamphorst, J.J., Shlomi, T., Thompson, C.B., and Rabinowitz, J.D. (2014). Quantitative flux analysis reveals folate-dependent NADPH production. Nature *510*, 298-302.

Gao, P., Tchernyshyov, I., Chang, T.C., Lee, Y.S., Kita, K., Ochi, T., Zeller, K.I., De Marzo, A.M., Van Eyk, J.E., Mendell, J.T., *et al.* (2009). c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature *458*, 762-765.

Hsu, P.P., and Sabatini, D.M. (2008). Cancer cell metabolism: Warburg and beyond. Cell *134*, 703-707.

Huang, W., Choi, W., Chen, Y., Zhang, Q., Deng, H., He, W., and Shi, Y. (2013). A proposed role for glutamine in cancer cell growth through acid resistance. Cell Res *23*, 724-727.

Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A.L., Kafri, R., Kirschner, M.W., Clish, C.B., and Mootha, V.K. (2012). Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science *336*, 1040-1044.

Le, A., Lane, A.N., Hamaker, M., Bose, S., Gouw, A., Barbi, J., Tsukamoto, T., Rojas, C.J., Slusher, B.S., Zhang, H., *et al.* (2012). Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. Cell Metab *15*, 110-121.

LeBleu, V.S., O'Connell, J.T., Gonzalez Herrera, K.N., Wikman, H., Pantel, K., Haigis, M.C., de Carvalho, F.M., Damascena, A., Domingos Chinen, L.T., Rocha, R.M., *et al.* (2014). PGC-1alpha mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. Nat Cell Biol *16*, 992-1003, 1001-1015.

Lee, S.Y., Jeon, H.M., Ju, M.K., Kim, C.H., Yoon, G., Han, S.I., Park, H.G., and Kang, H.S. (2012). Wnt/Snail signaling regulates cytochrome C oxidase and glucose metabolism. Cancer Res *72*, 3607-3617.

Lin, C.C., Cheng, T.L., Tsai, W.H., Tsai, H.J., Hu, K.H., Chang, H.C., Yeh, C.W., Chen, Y.C., Liao, C.C., and Chang, W.T. (2012). Loss of the respiratory enzyme citrate synthase directly links the Warburg effect to tumor malignancy. Sci Rep *2*, 785.

Liu, W., Le, A., Hancock, C., Lane, A.N., Dang, C.V., Fan, T.W., and Phang, J.M. (2012). Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. Proc Natl Acad Sci U S A *109*, 8983-8988.

Locasale, J.W. (2013). Serine, glycine and one-carbon units: cancer metabolism in full circle. Nat Rev Cancer *13*, 572-583.

Selivanov, V.A., Marin, S., Lee, P.W., and Cascante, M. (2006). Software for dynamic analysis of tracer-based metabolomic data: estimation of metabolic fluxes and their

statistical analysis. Bioinformatics *22*, 2806-2812.

Son, J., Lyssiotis, C.A., Ying, H., Wang, X., Hua, S., Ligorio, M., Perera, R.M., Ferrone, C.R., Mullarky, E., Shyh-Chang, N., *et al.* (2013). Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. Nature *496*, 101-105.

Starkov, A.A. (2008). The role of mitochondria in reactive oxygen species metabolism and signaling. Ann N Y Acad Sci *1147*, 37-52.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.

**Figure legends**

**Figure 1. Higher glycolytic flux and dependence of PC-3M cells.** (**A**) Basal ECAR of PC-3M and PC-3S cells, normalized to cell numbers. (**B**) Glucose consumption and lactate production rates normalized to cell numbers. (**C**) LDH activity normalized to intracellular protein content. (**D**) Relative contributions of glycolysis or PPP to lactate production determined by combining mass isotopomer distribution from [1,2-$^{13}C_2$]-glucose incorporation and lactate concentrations. (**E**) Assessment of the Crabtree as the decrease in OCR after exposure of cells to 18.75 mM glucose. (**F**) Effect of glucose deprivation and 5 mM 2-DG treatment on cell proliferation, determined after 48 h of incubation. Results are shown as percentage of proliferation relative to cells cultured in full media and without 2-DG (Control). (**G**) Effect of 30 mM oxamate treatment on

cell proliferation, determined after 48 h of incubation. Results are shown as a percentage of proliferation relative to untreated cells. (**H**) Intracellular ATP levels, analyzed after 24 h of incubation with 5 mM 2-DG, 30 mM oxamate or 30 mM oxamate + 10 mM DCA, and values normalized to DNA content. (**I**) Normalized rates of glucose consumption and lactate production. (**J**) Effect of 5 mM 2-DG treatment on cell proliferation, determined after 48 h of incubation. Results are shown as percentages of proliferation relative to untreated control PC-3M, PC-3M/Snai1 or PC-3M/SKMkd cells (100% proliferation). Data are means ± SD. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$. All experiments were performed at least in triplicate.

**Figure 2. Enhanced mitochondrial respiration of PC-3S cells relative to PC-3M cells.** (**A**) Higher sensitivity of PC-3S mitochondrial respiration to oligomycin. OCR was measured after exposure of cells to 2 µM oligomycin, expressed as fold change ($Log_2$). (**B**) Higher sensitivity of PC-3S cell proliferation to oligomycin, Cells were incubated with the indicated oligomycin concentrations for 24 h and cell proliferation assessed. Results are shown as percentages of proliferation relative to cells cultured in the absence of oligomycin (Control). (**C**) Enhanced total ATP synthesis by PC-3S cells in response to mitochondrial ATP synthesis inhibition with oligomycin. Intracellular ATP levels, analyzed after cells were incubated with 2 M oligomycin for 24 h. (**D**) Higher mitochondrial respiratory capacity of PC-3S cells. OCR profiles generated after exposure of cells to oligomycin, FCCP + Pyr and rotenone + antimycin. (**E**) OCR fold change ($Log_2$) after FCCP + Pyr injection. (**F**) Higher sensitivity of PC-3S cells to mitochondrial complex I and III inhibitors. OCR fold change ($Log_2$) after exposure of

cells to rotenone + antimycin. (**G**) Intracellular ROS levels in PC-3M and PC-3S cells. (**H**) Proton leak rates in PC-3M and PC-3S cells. (**I**) Enhanced sensitivity to mitochondrial drugs by induction of EMT or knockdown of pluripotency genes in PC-3M cells. OCR fold change ($Log_2$) after oligomycin, FCCP + Pyr and rotenone + antimycin injections. Data are means ± SD. *p < 0.05, **p < 0.01 and ***p < 0.001. All experiments were performed at least in triplicate.

**Figure 3. PC-3M cells use alternative substrates to feed into the mitochondrial metabolism.** (**A**) Enhanced glutamine consumption in PC-3M cells in response to glucose deprivation. Glutamine consumption rate was determined after 48 h incubation in standard media conditions (Control) or without glucose or with 5 mM 2-DG. (**B, C**) Dependence of PC-3M and PC-3S cells on glucose and glutamine availability for mitochondrial respiration (**B**) and glycolysis (**C**). Basal OCR and ECAR levels were determined in full and restricted media conditions. Values normalized to cell number. (**D**) Ketogenic amino acids consumption profile in PC-3M and PC-3S cells. (**E**) Mitochondrial function under restrictive media conditions. OCR fold change ($Log_2$) following oligomycin, FCCP+Pyr and rotenone+antimycin injections in glucose- or glutamine-deprived media. Bars represent mean ± SD. **p < 0.01 and ***p < 0.001. All experiments were performed at least in triplicate.

**Figure 4. Glucose and glutamine contribute differentially to the synthesis of TCA intermediates in PC-3M and PC-3S cells.** (**A**) Representation of the labeling distribution in TCA intermediates from [1,2-$^{13}C_2$]-glucose in the first turn of the TCA cycle. (**B**) Decreased entry of pyruvate to the TCA cycle in PC-3M compared to PC-3S

cells. Ratios of m2 citrate, glutamate, fumarate, malate and aspartate normalized to m2 pyruvate labeling. (**C**) Representation of the labeling distribution in TCA intermediates from [1,2-$^{13}C_2$]-glucose with the entry of a second m2 acetyl-CoA into the TCA cycle, considering the PDH reaction as the main contributor. (**D**) Ratios of m4 citrate, glutamate and m3 fumarate, malate and aspartate normalized to m2 pyruvate labeling. (**E**) Enhanced contribution of glutamine to the TCA cycle in PC-3M cells relative to PC-3S cells. m4 (m5 in the case of glutamate) labeling of TCA intermediates after 24 h of incubation with 2 mM 100% [U-$^{13}C_5$]-glutamine and GC/MS analysis. (**F**) m5 labeling of citrate and m3 labeling of aspartate, malate and fumarate after 24 h of incubation with 2 mM 100% [U-$^{13}C_5$]-glutamine and GC/MS analysis. (**G**, **H**) Higher sensitivity of PC-3M cell OCR to DCA and oxamate relative to PC-3S cells. Sequential DCA and oxamate injections (1$^{st}$, 2$^{nd}$, 3$^{rd}$ and 4$^{th}$) were performed at 10, 30, 50 and 70 mM oxamate, and 10, 20, 30 and 40 mM DCA. (**I**) Higher phosphorylation levels of PDH in PC3-M cells. Gene expression levels of PDHK1 and PDP2, determined by real-time RT-PCR. Inset: PDH and PDH-P protein levels, determined by Western blotting. Actin was used as a protein loading and transfer control. (**J**) Isodyn predictions of increased TCA cycle fluxes and mitochondrial respiration in PC-3S cells. Left, schematic representation of reactions participating in the TCA cycle metabolism. Right, metabolic fluxes in PC-3M and PC-3S cells estimated with Isodyn. Metabolites: Pyr, pyruvate, AcCoA, acetyl-CoA; Cit, citrate; OAA, oxalacetate; Mal, malate; □-KG, □-ketoglutarate; Glu, glutamate; Fum, fumarate. Subscripts indicate m, mitochondrial and c, cytosolic. Fluxes: PDH, pyruvate dehydrogenase; PC, pyruvate

carboxylase; ME, malic enzyme; CS, citrate synthase; citdmc, flux of citrate from mitochondria to cytosol; citakg, transformation from citrate to -ketoglutarate; akgfum, transformation from -ketoglutarate to fumarate; resp, mitochondrial respiration (NADH-linked respiratory consumption rate). Data are means ± SD. *p < 0.05, **p < 0.01 and ***p < 0.001. All experiments were performed at least in triplicate.

**Figure 5 PC-3M cells are more dependent than PC-3S cells on non-anaplerotic metabolic contributions of glutaminolysis.** (**A**) PC-3M cells express higher levels of glutaminase than PC-3S cells. Protein levels of total GLS1 and the glutaminase isoforms GAC and KGA, determined by Western blotting. Actin was used as a protein loading and transfer control. (**B**) Dependence of GAC expression in PC-3M cells on epithelial and pluripotent gene programs. Gene expression levels of the ratio GAC/KGA, expressed as fold change relative to PC-3M cells. (**C**) Sensitivity of PC-3M cell growth to glutaminase inhibition by BPTES. PC-3M, PC-3S, PC-3M/Snai1 and PC-3M/SKMkd cells were incubated with 10 M BPTES for 48 h and the effect on cell proliferation determined. Results are shown as percentages of proliferation relative to cells cultured without BPTES (Control). (**D**) Enhanced glucose consumption in response to BPTES. Glucose consumption was determined after 48 h of incubation without (Control) or with 10 M BPTES (**E**) Enhanced lactate production in response to BPTES. Lactate levels were determined after 48 h of incubation without (Control) or with 10 M BPTES. (**F**) Additive effects of BPTES and 2-DG on PC-3M cell death. Cell death was determined by the annexin V – propidium iodide assay in PC-3M and PC-3S cells cultured in control media or media containing 5 mM 2-DG, 10 M BPTES

or 5 mM 2-DG + 10 μM BPTES for 48 h. Plots depict variations in the percentages of dead cells (apoptotic + necrotic). (**G**) Failure of DMK to rescue growth of BPTES-treated PC-3M or PC-3S cells. Cells were incubated with 10 μM BPTES in the presence or absence of 2 mM DMK for 48 h and cell proliferation assessed. (**H**) Higher levels of glutathione in PC-3S cells. Levels of total intracellular glutathione, normalized to protein concentration. (**I**) Enhanced ROS production in response to BPTES. Intracellular ROS levels after 48 h of incubation without (Control) and with 10 μM BPTES, shown as percentage of mean fluorescent intensity (MnX) relative to control cells. (**J**) Reduced glutathione levels in PC-3S cells in response to BPTES. Intracellular glutathione levels in cells untreated or treated with 10 μM BPTES for 48 h, normalized to protein concentration. (**K**) Effect on cell proliferation of acidic media, and the combination of low pH with BPTES (pH 7.1 + BPTES) or without bicarbonate buffering (pH 7.1 – $NaHCO_3$) after 48 h of incubation. Results are shown as percentage of proliferation relative to control conditions (pH 7.9). (**M**) Effect of acidic media (pH 7.0) and acidic media + 10 μM BPTES (pH 7.0 + BPTES) on cell proliferation. Cells were cultured for 48 h under the indicated conditions and cell proliferation assessed. Results are shown as the percentages of proliferation relative to standard pH conditions (pH 7.9). Bars represent mean ± SD. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$. All experiments were performed at least in triplicate.

**Figure 6. PC-3M cells are characterized by a strong serine, glycine, one-carbon (SGOC) metabolism.** (**A**) Higher rates of alanine secretion by PC-3M cells. (**B**) Glycolytic origin of secreted alanine. Mass isotopomer distribution of alanine after 24

h of incubation with 10 mM 100% [1,2-$^{13}C_2$]-glucose. (**C**) Higher ALT activity in PC-3M cells. ALT enzyme activity in PC-3M and PC-3S cells, normalized to protein concentration. (**D**) Higher secretion of glutamate by PC-3S cells. Glutamate production and glutamine consumption rates in PC-3M and PC-3S cells. (**E**) Mass isotopomer distribution of glutamate after 48 h of incubation with 10 mM 100% [U-$^{13}C_6$]-glucose. (**F**) Serine and glycine consumption/production profiles in PC-3M and PC-3S cells after 96 h of culture. Negative values indicate consumption whereas positive values refer to production. (**G**) Predominance of m3 serine in PC-3M cells. Extracellular serine mass isotopomer distribution after 48 h of incubation with 10 mM 100% [U-$^{13}C_6$]-glucose. (**H**) Low levels of m2 glycine in PC-3M cells. Extracellular glycine mass isotopomer distribution after 48 h of incubation with 10 mM 100% [U-$^{13}C_6$]-glucose. (**I**) Analysis of genes involved in the glycine cleavage system, serine and one-carbon metabolism, determined by transcriptomic analysis. Relative transcripts levels are represented as the $Log_2$ of gene expression ratios between PC-3M and PC-3S cells. Asterisk (*) indicates those genes that are differentially expressed at $p < 0.01$ and $Log_2$ fold-change > 0.5 or < -0.5. Data are means ± SD. **$p < 0.01$ and ***$p < 0.001$. All experiments were performed at least in triplicate.

**Figure 7. Enhanced NADPH-producing fluxes and fatty acid synthetic reactions in PC-3S cells.** (**A**) Higher ribose synthesis in PC-3M cells. Total $^{13}$C-ribose labeling, represented as SIGm (m1+m2+m3+m4+m5 labeled ribose). (**B**) Ribose mass isotopomer distribution. (**C**) More active oxidative branch of PPP in PC-3S cells than PC-3M cells. Oxidative *vs*. non-oxidative branch of the PPP calculated as m1/m2

ribose ratio. (**D**) Higher G6PDH activity in PC-3S cells. Total G6PDH activity, normalized to protein content. (**E**) Higher TKT activity in PC-3M cells. TKT enzyme activity, normalized to protein content. (**F**) G6PDH and TKT gene expression levels, determined by real-time RT-PCR. Shown are fold levels relative to PC-3M cells. (**G**) Greater synthesis of palmitate and (**H**) stearate from glucose in PC-3S cells than PC-3M cells. Mass isotopomer distribution of palmitate and stearate after 24 h of incubation with 50% [1,2-$^{13}C_2$]-glucose. (**I**) Higher ATP citrate lyase (ACLY) protein levels in PC-3S cells, assessed by Western blotting. Actin was used as a protein loading and transfer control. (**J**) Left, geneset enrichment analysis (GSEA) for an expression dataset for 150 prostate cancer samples showing a significant enrichment of the PC-3M metabolic geneset in metastases (M) relative to primary tumors, and in T3 and T4 stage primary tumors relative to T1 and T2 stage primary tumors. A Pearson correlation was applied to determine linear relationships between gene profiles and four phenotypes (class 1: metastatic; class 2: T4 stage primary; class 3: T3 stage primary; class 4: T1 and T2 stage primary) taken as continuous variables. Right, heatmap illustrating the relative expression levels of the 20 genes of the PC-3M metabolic geneset, highlighting selected genes (see text). Data are means ± SD. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$. All experiments (except in panel J) were performed at least in triplicate.

# Chapter 3:
# Study the vulnerabilities associated to tumor heterogeneity and tumor progression in prostate cancer by using Genome-scale metabolic network models and data-driven approaches

**Summary**

In this chapter we developed a novel computational method combining probabilistic and mechanistic approaches to integrate multi-level omic data into a discrete model-based analysis. Part of the data that was integrated was previously obtained by data-mining and manual curation of relevant publications. Here, we analyzed the mechanism underlying the crosstalk between metabolism and gene regulation, using as case of concept the study of the abnormal adaptation to training in COPD patients. This chapter summarizes the results of these two analysis. The data-mining analysis is included in an article that is already published, the development of the computational tool generated an article that was recently submitted to an international scientific journal.

# Chapter 3a

# Knowledge management for systems biology a general and visually driven framework applied to translational medicine

Dieter Maier [1*] , Wenzel Kalus [1] , Martin Wolff [1] , Susana G Kalko [2] , Josep Roca [2] , **Igor Marin de Mas** [4] , Nil Turan [3] , Marta Cascante [4] , Francesco Falciani [3] , Miguel Hernandez [5] , Jordi Villà-Freixa [5] , Sascha Losko [1]

Maier et al. BMC Systems Biology 2011, 5:38
http://www.biomedcentral.com/1752-0509/5/38

BMC Systems Biology

SOFTWARE Open Access

# Knowledge management for systems biology a general and visually driven framework applied to translational medicine

Dieter Maier[1*], Wenzel Kalus[1], Martin Wolff[1], Susana G Kalko[2], Josep Roca[2], Igor Marin de Mas[4], Nil Turan[3], Marta Cascante[4], Francesco Falciani[3], Miguel Hernandez[5], Jordi Villà-Freixa[5], Sascha Losko[1]

**Abstract**

**Background:** To enhance our understanding of complex biological systems like diseases we need to put all of the available data into context and use this to detect relations, pattern and rules which allow predictive hypotheses to be defined. Life science has become a data rich science with information about the behaviour of millions of entities like genes, chemical compounds, diseases, cell types and organs, which are organised in many different databases and/or spread throughout the literature. Existing knowledge such as genotype - phenotype relations or signal transduction pathways must be semantically integrated and dynamically organised into structured networks that are connected with clinical and experimental data. Different approaches to this challenge exist but so far none has proven entirely satisfactory.

**Results:** To address this challenge we previously developed a generic knowledge management framework, BioXM™, which allows the dynamic, graphic generation of domain specific knowledge representation models based on specific objects and their relations supporting annotations and ontologies. Here we demonstrate the utility of BioXM for knowledge management in systems biology as part of the EU FP6 BioBridge project on translational approaches to chronic diseases. From clinical and experimental data, text-mining results and public databases we generate a chronic obstructive pulmonary disease (COPD) knowledge base and demonstrate its use by mining specific molecular networks together with integrated clinical and experimental data.

**Conclusions:** We generate the first semantically integrated COPD specific public knowledge base and find that for the integration of clinical and experimental data with pre-existing knowledge the configuration based set-up enabled by BioXM reduced implementation time and effort for the knowledge base compared to similar systems implemented as classical software development projects. The knowledgebase enables the retrieval of sub-networks including protein-protein interaction, pathway, gene - disease and gene - compound data which are used for subsequent data analysis, modelling and simulation. Pre-structured queries and reports enhance usability; establishing their use in everyday clinical settings requires further simplification with a browser based interface which is currently under development.

## Background

In biological or clinical research the creation of knowledge, here defined as *"the realisation and understanding of patterns and their implications existing in information"* relies on data mining. This in turn requires the collection and integration of a diverse set of up-to-date data and the associated context i.e. *information*. These sets include unstructured information from the literature, specifically extracted information from the multitude of available databases, experimental data from "-omics" platforms as well as phenotype information and clinical data. Although a large amount of information is stored in numerous different databases (the 2010 NAR database issue listing more than 1200 [1]) even more is still embedded in unstructured free text. Over the last 15 years a large number of methods and software tools have been developed to

* Correspondence: dieter.maier@biomax.com
[1]Biomax Informatics AG, Planegg, Germany
Full list of author information is available at the end of the article

integrate aspects of biological knowledge such as signalling pathways or functional annotation with experimental data. However, it has proven extremely difficult to couple true semantic integration (i.e. the mapping of equivalent meaning and objects) across all information types relevant in a life science project with a flexible and extendible data model, robustness against structural changes in services and data, transparent usage, and low set-up and maintenance requirements (see [2] for an excellent recent review). In principle this difficulty arises from the high complexity of life science data, which is partly an artefact of the fragmented landscape of data sources but also stems from reasons integral to the life sciences. The ever extending "parts-list of life" itself already offers an astounding number of object classes, from the molecular to the organism, even if common naming/identifier and definitions could be agreed upon. In addition experimental data can only be interpreted in the context of the exact identity of the experimental sample, the samples environment, the samples processing and the processing and quality of the generated data. Even more than the occasional extension of the "parts-list" from our growing knowledge, technical development continually generates new data types, processing methods and experimental conditions. While life science projects in general will (hopefully) share some concepts, almost each one will require some individual adjustment to integrate and view the relevant information. Therefore an optimal data integration approach will ensure that the data model can be based on existing concepts (ideally ontological i.e. controlled, structured vocabulary) yet remains flexible and extendible by the advanced user. In this respect today's most successful (i.e. widely used) data integration approaches such as SRS [3] or Entrez [4] show only weak, cross-reference based data integration without semantic mapping to a common concept (categorised as link/index integration by Köhler [5] and Stein [6]). They depend on pairwise mappings between individual database entries provided by the data source e.g. from a protein sequence entry to the corresponding transcript, the mappings lack semantic meaning i.e. the notion that a protein is expressed from a gene can not be stated or queried. Additional processing and data mapping is required to answer even simple questions such as "which molecular mechanisms are known to be involved in the pathology of chronic obstructive pulmonary disease ?". Currently custom-developed data warehouses such as Atlas [7], BIOZON [8] or BioGateway [9], are the most common technical concept to achieve full semantic integration (in public and industry projects). While these are ideally suited to answer complex queries their inflexible and pre-determined data model and the necessary, often difficult, data synchronisation result in high set-up and maintenance costs. Further, adaptation of such data warehouses structure to an ever changing environment or requirements are difficult at best [6]. Fortunately, as more data sources start to adopt semantic web representations such as OWL [10] and RDF [11] maintenance for semantic mappings becomes less of an issue as concomitantly to adopting a common language to transport semantics many data sources also standardise the semantics they provide such as using common entity references and ontologies.

An optimisation, at least regarding data synchronisation, has been to present a semantically fully integrated view of the data while the underlying data is assembled on-the-fly from distributed sources using a coherent data model and semantic mappings [12,13] (categorised as federation/view integration by Köhler and Stein [5,6]). Details of this approach vary widely. The ad-hoc data assembly process can be provided by home made scripts or, more recently, using workflow engines such as Taverna [14]. The data model can be programmed with a specific language as in Kleisli [15] or may make use of standard ontologies as with TAMBIS [16]. Semantic mapping to the common concept can be produced by a view providing environment, such as BioMediator [17] and the Bio2RDF project [18], or can come from individual integrated data sources. In the latter case the data sources either provide such mappings voluntarily, working for the common good of the "semantic web" [19] or are forced to do so by a closed application environment such as caBIG [20], Gen2Phen/PaGE-OM [12,21] or GMOD [22]. While conceptually elegant, these approaches have some disadvantages: the start-up costs are quite high (e.g. [13,23]), the performance is determined by the slowest, least stable of the integrated resources, complex queries result in large joins which are hard to optimise, and data models are often hard to extend. Ad-hoc desktop data integration and visualisation tools such as Cytoscape [24], Osprey [25] or ONDEX [26] on the other hand combine excellent flexibility with good performance due to local data storage, however they do not allow large scale knowledge bases to be collaboratively generated, managed and shared.

Another issue, which is only partially addressed by current data integration solutions, is the need to organise not only public information but project-specific knowledge and data, keep it private or partially private for some time, store and connect experimental results and corresponding metainformation about materials and methods and, if eventually verified, merge it into the pool of common knowledge. This may for example take the form of an existing signal transduction pathway which is privately extended with new members or connections. The extension is then published and discussed within a specific project until it is accepted as common knowledge. While data resources such as GEO provide the option to keep submitted data private for some

time, they generally do not allow existing knowledge to be extended as described above or allow existing data to be annotated with private or public comments.

Our challenge was to develop a knowledge management environment that achieves several goals: focus on the management of project-specific knowledge; ease data model generation and extension; provide completely flexible data integration and reporting methods combined with intuitive visual navigation and query generation; and address the issues of set-up and maintenance cost.

To do so we chose to apply different aspects of the approaches described above. In the next sections we describe the creation of a knowledge base for chronic diseases based on the BioXM software platform that efficiently models complex research environments with a flexible management, query and reporting interface which automatically adapts to the conceptualisation of the modelled information.

## Implementation

### The BioXM rationale

BioXM has been developed around the concept of object-oriented semantic integration. In this concept semantically identical objects, which represent information about the same real world object, and the meanings of associations between these objects are identified and mapped based on data and descriptive metainformation [27]. In the life sciences this mostly concerns the mapping of biological entities and descriptive data from literature and databases to common instances of objects like *genes, phenotypes or patients.*, Associations between the entities are mapped as relations (e.g. *compound*X *inhibits protein*A) and object - relation information is contextually structured (e.g. *gene*B *expressed in tissue*Z *at time*T *after application of compound*X). Based on objects as nodes (in BioXM called "elements") and relations as edges a "semantic network" which provides semantic information about the connection between participating object instances can be generated.

To enable the extraction of knowledge from integrated information the definition of a protein complex through a series of associations (i.e. "Protein A *participates_in_complex* Complex B") should be supplemented by evidence *why, when* and *where* Protein A participates in Complex B and how we know about it. In BioXM any associated information not represented directly as object or relation can be stored as "annotation" or as a special, performance optimised type "experiment" for high-throughput data. Such evidence may constrain the association to certain conditions e.g. "ProteinA *interacts with* ProteinB during cell cycle phase M". Sub-sections of the semantic network which are true only in specific conditions are modelled as "context". These elementary concepts, "elements", "relations", "annotations" and "contexts" can be defined freely or use existing structured, controlled vocabularies (ontologies). Due to this flexibility when configuring the data model, it becomes possible to formalise huge networks of knowledge (see Figure 1 and Data model configuration) even when concepts as diverse as molecular processes, disease phenotypes or clinical information about patients are concerned. These *complex semantic networks of relationships* allow one to detect connections, extract patterns and answer complex questions like: "Which drugs are known to interact with the over-expressed genes of patient A and have been shown to influence the patients cancer type in a clinical trial?" or "retrieve all inflammation related human genes which interact (physically or functionally) with the metabolic enzymes in skeletal muscle which are affected by COPD".

Strong usability is a pre-requisite for user acceptance of any knowledge management system and is approached in several ways in BioXM. The data model itself can be generated graphically and, based on project requirements and processes, may start as a simple sketch converted into a concept map. Parts of the concept map may develop into complex ontologies as discussion and collaboration with the user community extend quick ad-hoc sketched concepts iteratively into precisely defined and structured vocabularies. While not providing an environment for formal ontology development and evaluation such as Protege [28], BioXM allows ontologies to be used for knowledge representation and supports importing, integrating, editing and exporting ontologies. To optimise performance, storage and maintenance resources, data can be managed alternatively within an underlying relational database or be seamlessly integrated from external data sources (i.e. from a users point of view no difference is made between internal and external data). This combines the advantages of a data-warehouse-based deep integration with the low maintenance costs of a federated environment. External information can be incorporated by using the embedded BioRS™Integration and Retrieval System [29] or by directly accessing external objects from e.g. a relational database, a web service or a software application. Integration of the external resource is achieved by XML based registration of the application interface into BioXM. The registration must contain information about the actual programmatic access to the application and the attributes made available. Registered objects become directly available within the graphical data model viewer where they can be mapped in the same way as internal objects (see Data model configuration). External information can serve either as "virtual" semantic objects, or as "read-only" annotation of semantic objects. Read-only annotations add information to any

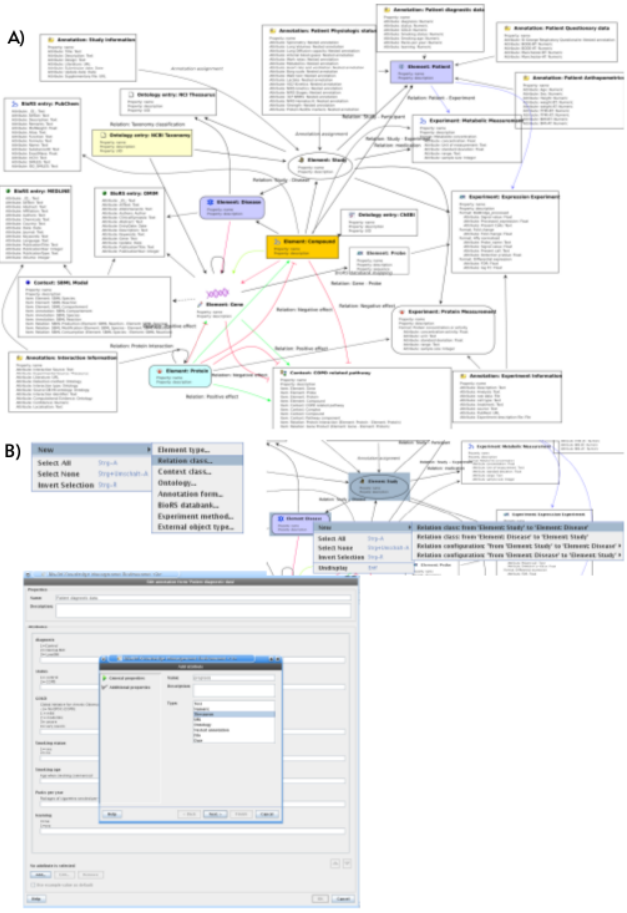



**Figure 1 BioBridge specific BioXM data model**. Partial visualisation of the BioBridge BioXM data model. A) Only the central objects and relations of the data model, such as "Gene", "Protein", "Expression Experiment" or "COPD pathway" are displayed. B) Data model configuration using the data model viewer context menu. New semantic objects are generated and edited directly within the graph viewer.

object, relation or other annotation; the information remains within the external data source and is read by the BioXM system on-the-fly but not modified. "Virtual semantic objects" can be organized in the project tree, can become part of a network, and can be annotated by the user like any other object. However, in contrast to BioXM internal objects they can not be changed e.g. renamed and the available search functions and performance depend on the implemented interface and features of the external source. Any change in content in

the original data source will become visible in BioXM immediately while full control of the business logic of publishing the objects and associated information remains with the external source. To access the data BioXM provides visual browsing of the knowledge network (Figure 2) as well as "natural language like" data mapping and query wizards (Figure 3) which automatically adapt to the changes of a given data model. The system thus enables scientists to create knowledge networks with flexible workflows for handling data and

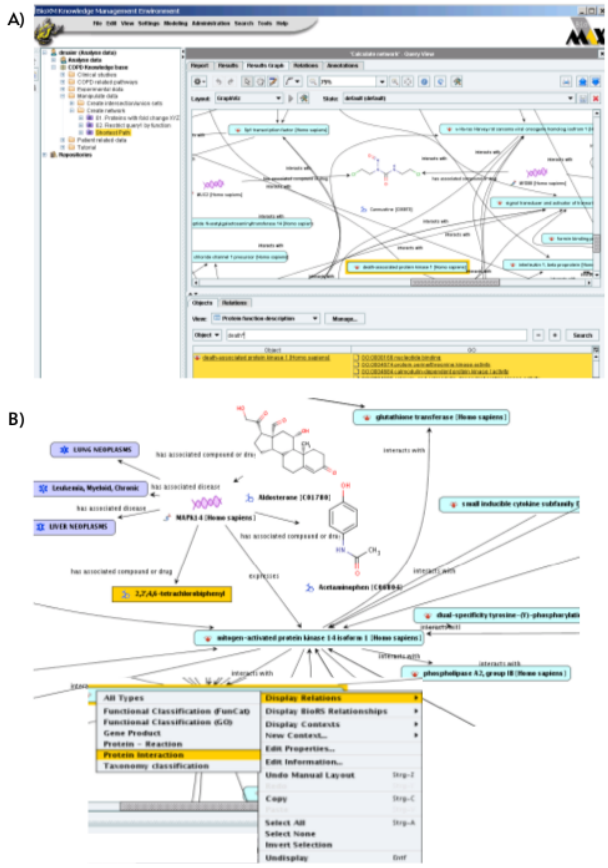



**Figure 2 Querying and visualising the knowledge network**. Visualisation of the knowledge network retrieved by querying for all genes simultaneously showing fold change <-1.0, >1.0 in a COPD expression experiment set, associated with functional information inferred by GO "signal transduction" and connected to each other using the "shortest path" network search. A) Provides an overview of the full retrieved network. B) Focuses on a specific part of the interaction network and the context menu to add gene - disease and gene - compound information for the MAPK14 gene.

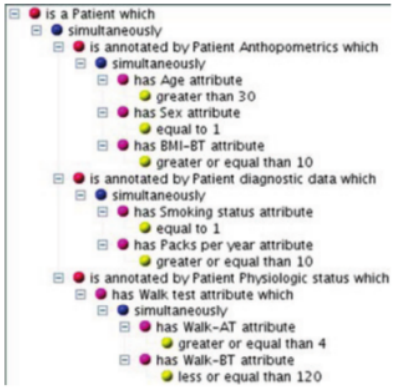



**Figure 3 Natural language like query wizard**. Formulating a natural language query and following the query building steps in the wizard provides a simple way to generate complex queries which can be saved as a smart folder for re-use and other users. Query: "Find all Patients which simultaneously have anthropometric attributes **Age >30**, **Sex male** and **BMI-BeforeTraining >= 10**, are diagnosed with **Smoking status = 1** and **Packs per >= 10** and whose Physiology has been measured with **Walk-AfterTraining >= 120** and **Walk-BeforeTraining <= 100**."

information types, including annotation and ontologies. In this way research projects can be modelled and extended dynamically to ultimately gain an adequate understanding of whole biological systems. As we are aware of the many issues associated with metainformation based identification of semantically identical objects we do not aim to provide an overall algorithmic solution to these but rather a framework which supports different mapping methods (see Populating the data model) and solves the organisation and use of the resulting semantic network.

#### Software Implementation

Details about the technical implementation of the BioXM software have been published elsewhere [30] and are only briefly summarised here.

BioXM is implemented as a platform-independent Java client-server application. The client Java application downloads as a Webstart package and can be started from any Browser (requires pre-installed Java 6.0.4 or newer). The server application has a modular architecture (see Figure 4). The resource and user management module allows control and restriction of user rights to view or modify certain data or parts of the data model. A complete audit trail for all entities and their relations is logged and supports secure data management. The application uses a relational database management system as backend (currently supports MySQL and Oracle). A Hibernate layer is employed for object-relation modelling. Therefore the database scheme is generic, containing the information about the object oriented data model as content. The command line and the SOAP web service application-programming interfaces (API) allow to integrate BioXM into larger bioinformatic infrastructures. Based on these APIs external applications such as BLAST [31] or network search algorithms are integrated. A plugin for the R statistics language http://www.r-project.org/ allows to access BioXM data directly from within the R interface and also to integrate any R method transparently as native BioXM views or analyses.

## Results

The EU FP6 BioBridge Systems Medicine project http://www.biobridge.eu focused on the integration of genomics and chronic disease phenotype data with modelling and simulation tools for clinicians to support understanding, diagnosis and therapy of chronic diseases. We have configured and extended the generic BioXM knowledge management environment to create the knowledge base for this translational system biology approach, focusing on chronic obstructive pulmonary disease (COPD) as an initial use case.

#### Data model configuration

In general within BioXM a particular scientific area of interest is semantically modelled as a network of related elements (see Table 1 for a list of fundamental semantic concepts available in BioXM). While there is some agreement throughout the life sciences regarding a number of semantic objects such as *gene* or *phenotype* the different communities such as clinical research, virology, plant research or synthetic biology differ on the concepts and definitions they use. Not only will a plant related knowledge base require the object *plant* instead of *patient*, a *vector* in virology might describe an infectious agent while in synthetic biology it will more likely define a DNA expression shuttle. Different ontology development initiatives (e.g. http://www.obofoundry.org/, http://bioportal.bioontology.org/) try to develop a consensus on these issues but currently no overarching "life science" data model can be defined. Therefore one of the main features of BioXM is the ability to dynamically create a data model specific to the project for which it is used and, while a project develops, easily adapt the data model based on the consensus between the project stakeholders.

Within BioBridge we had initial discussions between the project partners (clinicians, experimentalists and

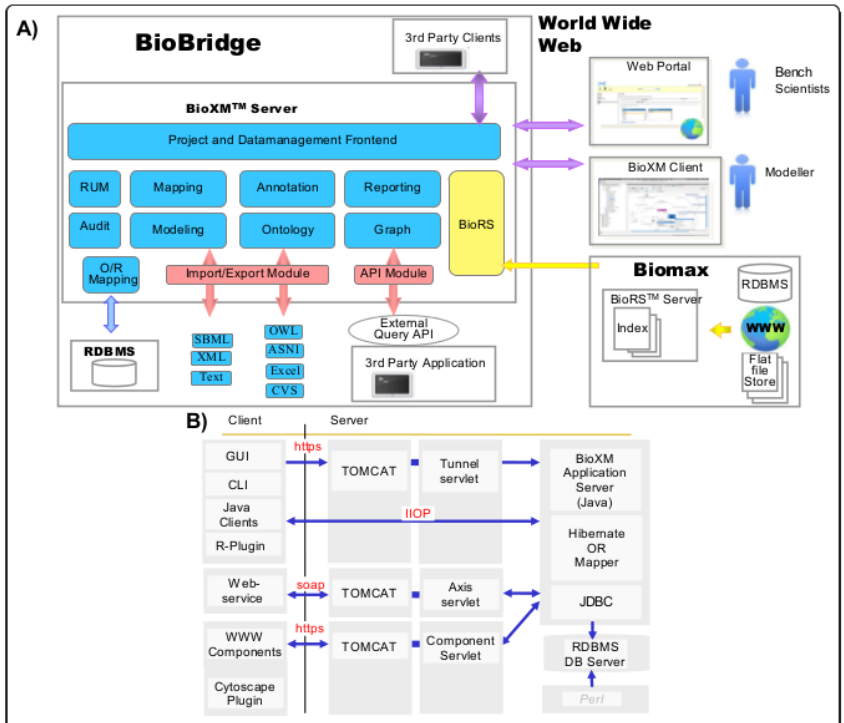



**Figure 4 BioXM architectural concept**. A) The BioXM Java client-server application implements a modular architectural concept on the server side. Tasks such as resource and user management, reporting or graph layouting are implemented as individual modules. B) Inter-module communication and interaction with external applications, web services, browser and Java clients is based on standard protocols.

modellers) about the kind of knowledge that needed to be represented (see Table 2 for the resulting list) as well as how this representation should be conceptualised. This is a typical first step for a knowledge management project which usually provides only an initial, rough idea of the final goal that will need many iterations to reach. Setting-up and changing the data model is done directly within the data model graph (Figure 1) using a context menu to create and edit the fundamental semantic objects (Figure 1B). From the "new" function the basic semantic concepts (element, relation, annotation, experiment, context, ontology, external object) become available to generate for example an element *gene*. By selecting existing object definitions they can be edited, connected by new or existing relations, assigned an annotation or assigned into a context. We created elements such as "gene", "protein", "patient", "compound" and a total of 15 element types sufficed to describe the clinical and biological knowledge relevant to chronic disease (see additional file 1 for a full data model XML export). Information about the interconnection of entities is captured in relations such as "is medicated by" for the compound a patient receives as medication or "regulates" for a protein that regulates expression of a gene. The initial data model was configured within one week and subsequently many iterations extended and adapted the model while the knowledge base was already being populated and in productive use. In total within the current BioBridge COPD knowledge base we used 82 relation types to capture the details of

**Table 1 Fundamental semantic objects**

| Semantic object | Description | Example |
|---|---|---|
| Element | Represents a basic unit of a knowledge model | "Gene" element type can be used to create the "STAT3" gene element "Disease term" element type can be used to create the "pancreatic tumor" disease term element |
| Relation | Describes a relationship between semantic objects | "Gene-disease" relation class can be used to create the "STAT3 is associated with disease pancreatic tumor" relation |
| Annotation | Extends the properties of a semantic object by a set of attributes | Gene report<br>Patient record<br>Protein entry<br>Literature abstract<br>Experimental data (evidence) |
| Experiment | Performance optimized extension of an element by a set of attributes | Expression data<br>Metabolomics data<br>Proteomics data |
| Ontology | Classifies semantic objects according to a defined hierarchical nomendature of concepts | "3.2.2.21 DNA-3-methaladenine glycosidase II" entry is part of the "EC numbers" ontology<br>Gene Ontology to classify biological function<br>NCI Thesaurus of disease terms taxonomy |
| Context | Represents sets of semantic objects | Metabolic pathways Protein complexes A disease process or pattern |
| Database/ external object | A basic unit of a knowledge model populated from an external application/ database | dbSNP Sequence Variant Genome feature |

The fundamental semantic objects used in BioXM which allow a descriptive model of the world to be formulated and data resources to be related to that model. The semantic objects allow an extendable model to be defined from a set of well-defined building blocks (adapted from [30]).

semantic connections between element types, ontology terms, experiments or sub-networks. To describe assemblies of entities or relations with some common feature e.g. a signalling pathway we defined sub-networks as *contexts*. 16 types of contexts where used to capture for example SBML-based simulation models [32], KEGG pathways [33] or inflammatory processes involved in COPD derived by literature mining. All "semantic concepts" (such as *elements*, *relations*, *contexts* or *ontology* ) can be associated by annotations with information such as age, weight and gender for a patient, function for a gene or experimental evidence for a protein-protein interaction. Annotations are based on freely definable forms and support hierarchical organisation of information (nested annotation forms). Multiple semantic objects can share annotation to imply relationships. Within BioBridge we defined 61 annotation forms with 892 attributes to provide for example, electronic case report forms for anthropometric, diagnostic, physiologic and questionnaire data. Experimental data is seen conceptually as special, performance optimised annotation. We defined seven experiment formats covering data types such as transcription, metabolite or enzyme kinetics. BioXM supports the conceptualisation of entire areas of interest by using ontologies, which can be used to infer facts and construct abstract queries. 19 different pre-existing ontologies such as GO or the NCI-thesaurus were integrated into the BioBridge data model. The graphical data model builder and the context menu used for the configuration of the BioBridge specific data model is shown in Figure 1. Within BioBridge we focus on the configuration of a data model suited to knowledge, clinical and experimental date around COPD, other instances of BioXM have been configured to cover different diseases or indeed fields of life science research such as enzyme biotechnology or synthetic biology, underscoring the general applicability of the semantic data model configuration process.

### Populating the data model

The model described above enabled us to semantically integrate existing public databases and information derived from the literature with clinical and experimental data created during the BioBridge project.

To populate the knowledge base with data from public databases with large sizes and regular updates we mainly use virtual objects with a manual mapping of the object concept into the data model. For sources without appropriate interfaces or weak performance and for project internal data and knowledge we manually generated mappings in import-templates. A graphical wizard for import-template generation provides a selection of possible import options and mappings. Only applicable objects of the data model are presented for example after defining an object as *gene* only relations enabled for *gene* are available for the next mapping step. The available selection automatically adapts to any change in the data model configuration. From this selection the import operations are assembled by drag-and-drop to provide the mapping for a given data source forming an import script which can be saved and re-used (see Figure 5). For sources which are imported and provide regular updates a scheduling system is used to define automatic execution of the corresponding access and import methods.

Mapping a resource to the data model requires expertise about the semantic concept of the resource and the configured BioXM data model. To integrate the individual entities of a data source semantically the mapping method for the entities need to be defined. If available, BioXM makes use of namespace based standard identifiers, existing cross-references and ontologies for the

**Table 2 COPD specific knowledge base**

| Source database | Information type | Current statistics | Level of curation | Updates/ Version |
|---|---|---|---|---|
| **BIND** | Protein interaction<br>Molecular complexes<br>Pathways | 6256 Interactions | High throughput data submission and manually curated from literature | last public version 20.3.07 |
| **BioGrid** | Protein interaction | 19 707 interactions | Manually curated from literature<br>Different evidence codes | updated monthly |
| **BRENDA** | Enzyme kinetics | 4 729 | Manually curated from literature | updated monthly |
| **ChEBI** | Compound information | 15 367 | Curated from different data sources | updated weekly |
| **Comparative Toxicogenomics Database (CTD)** | Compound-gene, Compound-disease and Gene-disease relationships | 259 898 relations | Manually curated from literature | updated monthly |
| **EntrezGene** | Gene functional information | 80 793 human, mouse and rat genes | Curated information integrated from different databases, based on RefSeq genomes | updated weekly |
| **Enzyme** | Enzyme related functional information | 4 833 | Manually curated from literature | updated weekly |
| **GEO** | Functional genomics data (expression, ChIP-chip etc.) | >400 000 individual experiments | User submission | updated weekly |
| **IntAct** | Protein interaction | 21 584 binary interactions | Literature curation<br>User submission | updated weekly |
| **KEGG** | Pathways | 418 pathways | Manually curated from literature | updated monthly |
| **LIGAND** | Compound information | 15 185 | Manually curated from the published literature | updated monthly |
| **MIPS Mammalian** | Protein Interaction | 410 interactions | Manually curated from literature | current release 31.10.07 |
| **OMIM** | Gene - disease relations | 20 823 | Curated from literature | updated weekly |
| **Pfam** | Protein family information | 10 340 families | Manually curated from sequence alignments | 23.0 |
| **ProLinks** | Interaction | >1.4 million | Automatic inference | last release 18.10.04 |
| **PubChem** | Compound information | >26 million | Automatic collection | updated weekly |
| **PubMed** | Literature abstracts | >19 million | Automatic collection with manual curation | updated weekly |
| **Reactome** | Interactions and pathways | >600 pathways, >24 000 Inteactions | Manual curation | updated monthly |
| **RefSeq** | DNA and protein sequences | >11 million | Automatic processing and manual curation | updated weekly |
| **Unigene** | Transcript sequences | >2 million | User submission followed by automatic clustering | updated monthly |
| **UniProt** | Protein sequences | >10 million | Automatic processing, Swissprot subsection manual curation | updated bi-weekly |

Public data integrated into BioXM in the current version of the BioBridge COPD knowledge base.

population of the data model. In most cases the semantics of a given data source are not (yet) described in machine readable form and the initial mapping template needs to be generated manually. The BioXM core framework is extended with pre-defined semantic mappings currently existing for about 70 public data resources and formats (see additional file 2). In addition text-mining and sequence similarity (BLAST) based mappings are enabled, however users need to be aware of the pitfalls of these methods as no automatic conflict resolution is attempted.

For the COPD knowledge base we use entities, references and ID mappings provided by EntrezGene [34], Genbank [35], RefSeq [36], HGNC [37], ENSEMBL [38], UniProt [39] and EMBL [40] to populate the system with instances of *genes* and *proteins* from human, mouse

**Figure 5 Populating the data model**. Based on the given data model, the import wizard provides the selection of available import operations in the left frame. These are moved by drag-and-drop into the right frame where they form the import script which provides the mapping information between a data source and the data model. Here two elements from the data source are defined as type "Protein" and are referenced by their UniProt IDs. The relation between the two proteins is a "Protein interaction" from the Reactome data source and the associated evidence is stored as annotation.

and rat. Starting with EntrezGene we create gene instances and map entities from the other sources iteratively by reference. For each database the quality of the references to external sources needs to be judged individually and correspondingly be constrained against ambiguous connections. UniProt protein entries for example provide references to DNA databases, with some references pointing to mRNA which allows the corresponding gene to be identified uniquely, and others pointing to contigs and whole chromosomes with multiple gene references. Use of references in this case therefore is constrained to the target entry type "mRNA". A new instance is generated for each database entry from the corresponding organisms which can not be mapped to an existing instance by ID reference; no name based mapping or name conflict resolution is attempted at this stage. As the knowledge base develops iterative rounds of extension occur with additional data sources. Based on non-ambiguous identifiers we map additional information from the sources described below with mappings being extended, removed and remapped during each updating round.

Generating an import template using the import wizard requires no software development knowledge and for many sources only takes minutes (e.g. for protein-protein interaction data which uses UniProt accessions to unambiguously identify the protein entities and the Molecular Interaction Ontology [41] to describe the interaction type and evidence). However, integration can also take up to a week of software development if extensive parsing and transformation of a complex data source such as ENSEMBL is required.

Naming conflicts and lack of descriptive, structured metainformation are the main reasons for the lack of semantic integration in the life sciences, issues that are as much technological as sociological. The use of a structured knowledge management tool within BioBridge ensured all newly produced data makes use of unique identifiers and provides extensive, structured metainformation. This semantic integration and standardisation fostered data exchange as well as social interactions within the project, which are a pre-requisite for translational systems biology projects and their highly diverse multi-subject expert teams. In addition the semantic integration greatly simplifies the future sharing of the produced data as it is immediately available in semantic form.

For the import of data several formats are supported from simple manually mapped delimiter formats such as tab-delimited to XML formats with potentially fully automatic semantic mapping like Pedro [42], SBML or OWL. If machine readable metainformation is provided, such as MIRIAM references in SBML, they are used to automatically map the imported entities to existing instances of semantic objects. In the current version, the knowledge base integrates more than 20 different public databases (see Table 2) representing a total of 80 793 relevant genes (30 246 human, 27 237 mouse and 23 310 rat), 1 307 pathways, 78 528 compounds with related gene/disease information, 1 525 474 protein interactions and the entire Gene Expression Omnibus and PubChem databases resulting in a total of 3 666 313 connections within the knowledge network. In addition two BioBridge specific datasets, 54 inflammation and tissue specific pathways and 122 COPD and exercise specific metabolite and enzyme concentrations and activities were manually curated from the literature within the project. The pathway curation followed a standard text-mining supported process as described for example in [43] while the enzyme concentration and activity curation was fully manual due to the small set of available relevant publications. To our knowledge BioBridge thus provides the first semantically integrated knowledge base of public COPD-specific information. In addition the resource will be continuously extended as more COPD specific data becomes publicly available e.g. the experimental data generated within BioBridge (160 pre- and post-training expression, metabolite and

proteomics data sets) will become publicly available as soon as the consortium has generated an initial analysis of the data. Currently the COPD knowledge base contains almost 10 million experimental result data of which almost 6 million come from public data. In other projects we are currently using BioXM with several hundred million data points on networks with tens of million edges and nodes, showing that the approach scales for at least two more orders of magnitude (unpublished data).

### Browse, query and retrieve

Here we provide an overview of the available functionality, for a detailed step-by-step tutorial of the knowledge base please see the additional file 3: Step_by_Step_Tutorial_BioXM.pdf. When accessing the BioBridge portal http://www.biobridge.eu/bio/ you will have to register to access the knowledge network (the registration is required to sustain funding support and enable personalisation of the interface, no further use of personal data will be made). Then access the knowledge base by following the BioXM links. The BioXM user interface (Figure 6) provides a visually driven query system with which information can be browsed from a network graph (see Figure 2B), based on pre-defined queries (purple *smart folders* in the navigation tree, see Figure 2A) or by interactive query generation (a detailed application example is described below).


**Figure 6 BioXM graphical user interface overview**. The BioXM graphical user interface (GUI) consists of three frames. A Navigation bar provides the functions for importing, managing, reporting and searching data. A project and repositories frame to the left, allows all data available to a user to be accessed in the repositories section and the data to be organised in a user and project specific way in the projects section. A right frame, is used to display detailed information about any object selected in the left frame.

Users visually browse and query the network simply by right-clicking on any focus of interest (e.g. a gene, a patient or a protein-protein interaction) so that associated entities can be added to the existing network visualisation. The corresponding context menu is dynamic, offering all those entities for selection which, based on the data model, are directly associated with the initial focus (i.e. one step in the network). A researcher could, for example, expand from a gene to include its relationships with diseases. From the disease association it may be of interest to identify patients represented in the gene expression database who share that particular diagnosis. Entities distanced by more than one step in the data model can be associated with each other by complex queries which transverse several nodes within the graph and aggregate information to decide whether a connection is valid. These complex queries are transparent to the user who executes them as part of the graphical navigation when asking for "associated objects" (see below for query construction). Graph based navigation will become difficult in terms of visualisation layout and performance beyond several thousand objects.

The intuitive graphic query system therefore is supplemented by a more complex wizard that allows dynamic networks to be created by in depth, structured searches, which combine semantic terms that are dynamically pre-defined by the data model (see Figure 3 for the query wizard and Figure 2A for a resulting network. The additional files 3 and 4 provide details and example data on how to create a query). The query construction is natural language like and thus allows to generate complex searches without knowledge of special query languages such as SQL or SPARQL but some knowledge about the data model must be acquired to work efficiently with the wizard. A search for all patients diagnosed with COPD severity grade above 2 but no cancer which have low body mass index for example would read: "*Object to find is a Patient which simultaneously is annotated by Patient diagnostic data which has GOLD attribute greater than 2 and is annotated by Patient Anthropometrics which has BMI-BT attribute less than 18 and never is diagnosed with a NCI Thesaurus entry which is inferred by ontology entry which has name like '*cancer*'*". A query can be saved as a "smart folder" or query template for re-use and thus allows experienced users to share their complex queries with less frequent users. For saved queries "Query variables" can be defined so generic smart folder queries can be adapted to specific question. In the example above the actual parameters for COPD severity grade, BMI and

diagnosed disease might be set as variables for other users to change. Normal folders (yellow) allow users to organise data manually in their private space by drag-and-drop e.g. to create a permanent list of "favourite genes" or a specific pathway. In contrast the content of "smart folders" is dynamic, as it is actually a query result, which immediately updates whenever changes in the content of the knowledge base occur. Defining a query takes between seconds for simple "search all compounds used as medication" type questions to tens of minutes for complex questions which traverse the full connectivity of the semantic network. In the same way performance of query execution directly depends on the complexity of the query. Queries traversing many connections in the semantic network with entwined constrains may take several minutes to execute while simple queries even with millions of results return within seconds.

All individual entries, contents of folders and query results are organized in configurable reports presenting the initial element and any desired associated information. To reflect differences in interest and focus for different users multiple reports can be defined for any object for example providing a quick overview of patient laboratory data for the clinician while another patient report provides the expression data for the data analyst. Reports can be exported both in tabular and XML format. Reports can transparently integrate external applications (e.g. the statistical programming environment R) in order to derive visualisation and/or further analysis of the data (Figure 7). The report is based on individual "view items" defined with the same type of wizard as for query generation and import template generation.

Therefore configuring reports does require no software development skill but is based on an understanding of the data model configuration. As with the query and import template wizards, view items are drawn from the data model using functions such as "related object", "assigned annotation" or "query result" which can be further restricted to specific types such as "relation of type protein expression". Configuring a new report on average takes only minutes but, as reports can contain query results, can also take tens of minutes if a new complex query needs to be defined. Reports defined for an object like *gene* can be re-used as "nested reports" wherever a *gene* type object is included as view item in another report allowing complex reports to be assembled from simple units.

Report display performance directly depends on the configuration and takes between seconds and several minutes. Simple, fast reports depend on directly related information e.g. a gene report which brings together Sequence Variant, gene-disease and gene-compound information. Complex, slow reports integrate queries to traverse the semantic network and pull together distant information e.g. the medication for the patients for which a given gene was upregulated.

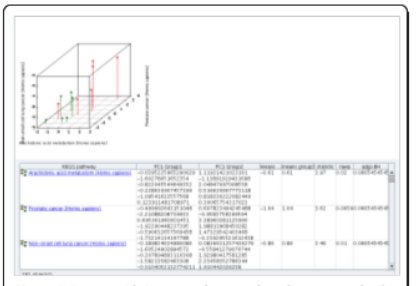


**Figure 7 Report with integrated external application result**. This result view shows the first three significantly different pathways between sedentary and trained healthy people. The 3 D scatterplot on the top visualises the PC1 for each experiment as spot in the 3 D space with the KEGG pathways as dimensions. Green (experiment group 1, here pre-training) and red (experiment group 2, here post-training) spots clearly occupy two different regions of the plot, indicating differences. The significance of the differences is visible in the tabular report where the first column provides the name of the pathway. Column 2 and 3 list the PC1 values for each of the associated experiments in group 1 and 2. Columns 4 and 5 show the overall PC1 mean of the pre- and post-training data. The following columns list the t-, p- and adjusted p-value respectively.

View items can also be used within "information layers" which visualise the information directly on top of a network graph by changing size and colour of the displayed objects. To define the information layer ranges of expected numerical or nominal values in the view item are assigned to colour and size ranges for the graphical object display. In a simple case this is used to display expression data on top of a gene network but based on using query results as view items it can also be used to display the number of publications associated with a gene - phenotype association. Within the graph information layers are executed for every suitable object displayed and depending on the complexity of the defined view item the generation of the overlay can take between seconds and several minutes.

### Application case

The BioBridge knowledge base implementation enables the integrative analysis of clinical data e.g. questionnaires, anthropometric and physiologic data with gene expression and metabolomics data and literature derived molecular knowledge. The knowledge base is currently used by data analysis and modelling groups within BioBridge to extend literature-derived, COPD-specific

molecular networks with probabilistic networks derived from expression data (method described in [44]). Output of the probabilistic networks together with expression and metabolomics data is then used to tune mathematical models of the central metabolism ([45]) for COPD specific simulations.

As an example use case we briefly describe the initial search for connections between molecular sub-networks affected by exercise, which is generated as a starting point of the BioBridge investigation (for detailed descriptions of this and further use cases see the associated *Step-by-step tutorial*). To this end we searched the expression data in the COPD knowledge base to retrieve studies involving patients diagnosed with diseases affecting muscle tissue or involving "exercise" as treatment. Based on these experiments we used the R interface to conduct a principal component analysis to extract the KEGG pathways most strongly affected by expression changes (R scripts provided in additional files 5 and 6). Interestingly the affected pathways are mainly associated with tissue remodelling and signal transduction pathways. Using the enzyme and compound concentration and kinetic measurements extracted from key manuscripts on muscle dysfunction in COPD and training effects we find that a number of compounds and proteins involved in key pathways with altered expression derived from the principal component analysis show significant changes in concentration/activity in the published literature. Therefore independent support from existing knowledge is gained from the integrated COPD knowledge resource for the statistical analysis results of expression data. For these compounds and proteins we use the network search algorithm to search the entire COPD knowledge network but restrict the allowed connections to human genetic interaction, protein-protein interaction, gene-compound or gene-disease interaction (see Table 2 for the individual sources mapped to each of these relation types). The resulting network connects inflammatory and metabolic processes affected in COPD patients (see Figure 8A and additional file 7: Knowledge_network_table.pdf). Visualising the number of disease specific pathways associated with each of the nodes as information layer in this network immediately provides one aspect of the potential weight of the individual nodes for future investigation (see Figure 8B) with TNF-receptor associated factor 2 and 6 showing the highest relevance in this respect for all newly connected nodes. The definition of the network search can be changed to include multiple additional information types e.g. drug-relations. Weighing of evidence is achieved by specifying penalty weights in the network search parameter set for each type of relation searched. Additionally query results provide further evidence directly visualised as information layer on top of the network graph to change size and colour of the displayed objects. Within the query all attributes available in the knowledge base can be used as described in the previous section, from the number of independent literature occurrences supporting a gene-disease relation to the category of the experimental evidence supporting a protein-protein interaction to derive an informed weighted list of further investigation targets. Different information layers can be defined to overlay different types of information such as quantitative experimental data from gene expression or qualitative text-mining results. The decision about what kind of evidence should be weighted in which way is depending on the individual type of data considered, we do currently not provide automatic scoring or ranking algorithms. However, based on the R plug-in or the API, R-scripts or external algorithms can be integrated to calculate corresponding scores. Due to the fact that information layer can integrate complex queries for many objects this is a performance critical function and may take several minutes to execute.

## Conclusions

While the promises of the Semantic Web continue to creep slowly into existence [46] individual projects need immediate, adaptable solutions which allow project specific knowledge conceptualisations to be set-up with low start-up cost and the flexibility to extend, standardise and exchange their data and knowledge. Using a generic knowledge management framework we were able to configure and populate a productively used, project specific systems biology knowledge base within 6 month with similar, software development based integration projects being reported to take between 2-5 years [13,23]. The COPD knowledge base, set-up as the central knowledge management resource of the BioBridge project, provides a free, comprehensive, easy to use resource for all COPD related clinical research and will be continuously extended aiming to generate the definitive resource on clinical research in COPD. More broadly our configuration based approach to semantic integration is generally applicable to close the knowledge management gap between public and project specific data affecting a large number of current systems biology and high-throughput data dependent clinical research projects. To bridge the gap between the current user interface, tuned to suit experienced, frequent users, and everyday clinical application the BioBridge project developed a simplified web portal interface for a number of use-cases. Based on the feedback from clinical users we recently developed a Foswiki [47] plug in for BioXM which unifies the simple set-up of a Wiki with the knowledge management functions. From this we will develop a browser based portal as the primary access to the COPD knowledge base. Future directions include: support of

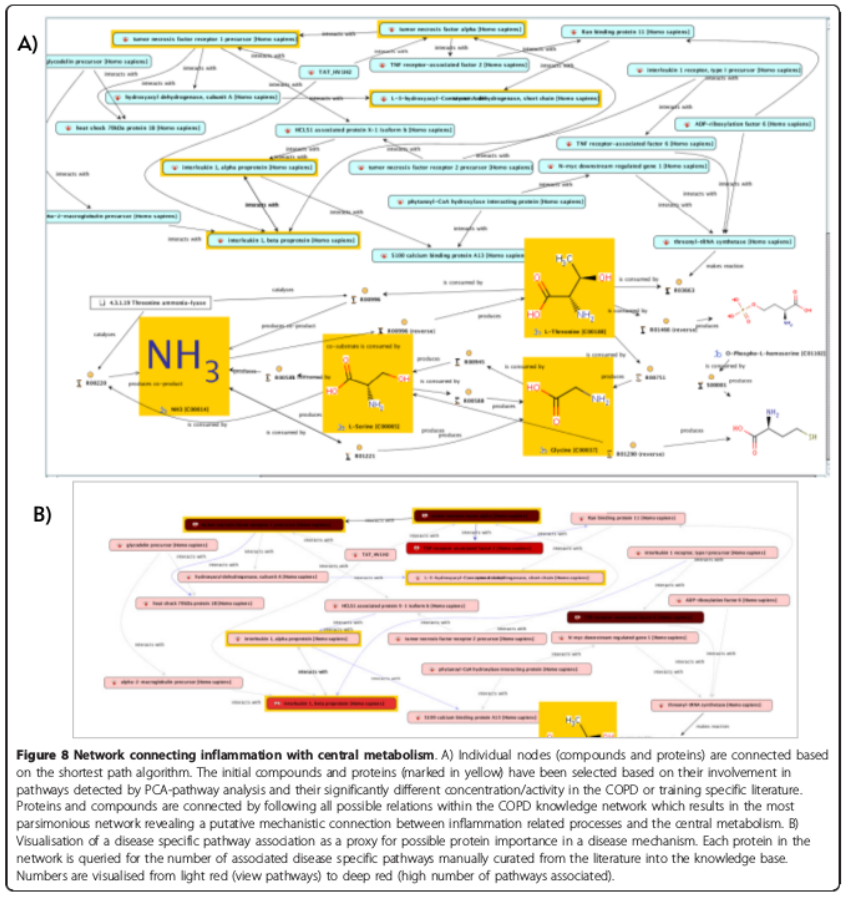



**Figure 8 Network connecting inflammation with central metabolism**. A) Individual nodes (compounds and proteins) are connected based on the shortest path algorithm. The initial compounds and proteins (marked in yellow) have been selected based on their involvement in pathways detected by PCA-pathway analysis and their significantly different concentration/activity in the COPD or training specific literature. Proteins and compounds are connected by following all possible relations within the COPD knowledge network which results in the most parsimonious network revealing a putative mechanistic connection between inflammation related processes and the central metabolism. B) Visualisation of a disease specific pathway association as a proxy for possible protein importance in a disease mechanism. Each protein in the network is queried for the number of associated disease specific pathways manually curated from the literature into the knowledge base. Numbers are visualised from light red (view pathways) to deep red (high number of pathways associated).

additional structured languages for import and export such as BioPAX [48]and CellML [49]; development of a workflow framework for data analysis and integration of algorithmic methods for semantic mapping.

## Availability and requirements

The BioBridge COPD knowledge base as a data resource is freely available to academic users (requires pre-installed Java 6.0.4 or newer). Upon request to DM the BioXM software application itself can be made available for academic PhD research projects within the Biomax BioXM PhD collaboration programme.

## Additional material

**Additional file 1: XML export of the full COPD knowledge base data model containing all semantic objects, annotation and experiment definitions**.

**Additional file 2: Table of existing mapping scripts and wrappers.**

**Additional file 3: Step-by-step tutorial to access and query the COPD specific BioXM instance set-up as part of the BioBridge project.**

**Additional file 4: A set of Affymetrix probes used within the Step-by-step tutorial.**

**Additional file 5: R script for pathway level principal component analysis of expression data integrated into BioXM and used as part of the analysis described in the Step-by-step tutorial.**

**Additional file 6: R script for significance testing of differentially regulated pathways identified by principal component analysis integrated into BioXM and used as part of the analysis described in the Step-by-step tutorial.**

**Additional file 7: Analysis of the sub-network connecting inflammation to central metabolism which is derived from the overall COPD knowledge network based on shortest path network search.**

**Acknowledgements**
We thank all the members of the BioBridge team for valuable discussions on the data model and their evaluation of BioXM. This work was supported by the European Commission (FP6) BioBridge LSHG-CT-2006-037939. We thank the unknown reviewers for their very valuable suggestions and criticism which helped to clarify and structure the manuscript.

**Author details**
[1]Biomax Informatics AG, Planegg, Germany. [2]Hospital Clinic-IDIBAPS-CIBERES, Universitat de Barcelona, Barcelona, Spain. [3]School of Biosciences and Institute of Biomedical Research (IBR), University of Birmingham, Birmingham, UK. [4]Departament de Bioquimica i Biologia Molecular, Institut de Biomedicina at Universitat de Barcelona IBUB and IDIBAPS-Hospital Clinic, Barcelona, Spain. [5]Computational Biochemistry and Biophysics lab, Research Unit on Biomedical Informatics (GRIB) of IMIM/UPF, Parc de Recerca Biomdica de Barcelona (PRBB); Barcelona, Spain.

**Authors' contributions**
DM developed and populated the BioBridge specific data model, provided pre-structured queries and reports and drafted the manuscript. WK devised the BioXM software architecture. MW devised and implemented the BioXM R interface and integration. SGK co-developed the experimental part of the BioBridge data model and provided data analysis. JR conceived the integrative BioBridge approach and provided input to the BioBridge specific data model. FF conceived the data analysis workflow and co-drafted the manuscript. NT implemented the data analysis workflow and co-drafted the manuscript. MC co-developed the BioBridge data model, conceived the literature curation process and co-drafted the manuscript. IMM implemented the literature curation for COPD specific enzyme and compound data. MH implemented the BioBridge portal integration with BioXM. JFV conceived the BioBridge portal architecture. SL devised the BioXM concept and managed its implementation. All authors read and approved the final manuscript.

**Competing interests**
DM, WK, MW and SL are employed at Biomax Informatics AG and will therefore be affected by any effect of this publication on the commercial version of the BioXM software. JR, FF, NT, MC, IMM SK and JVF are employed at academic institutes and will therefore be affected by any effect of this publication on their publication records.

**Received: 14 June 2010 Accepted: 5 March 2011**
**Published: 5 March 2011**

**References**

1. Cochrane GR, Galperin MY: **Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2010.** *Nucl Acids Res* 2010, **38**:D1-4.
2. Goble C, Stevens R: **State of the nation in data integration for bioinformatics.** *J Biomed Inform* 2008, **41**:687-93.
3. Etzold T, Ulyanov A, Argos P: **SRS: information retrieval system for molecular biology data banks.** *Methods Enzymol* 1996, **266**:114-28.
4. Schuler GD, Epstein JA, Ohkawa H, Kans JA: **Entrez: molecular biology database and retrieval system.** *Methods Enzymol* 1996, **266**:141-62.
5. Köhler J: **Integration of life science databases.** *Drug Discovery Today: BIOSILICO* 2004, **2**:61-69.
6. Stein LD: **Integrating biological databases.** *Nat Rev Genet* 2003, **4**:337-345.
7. Shah SP, Huang Y, Xu T, Yuen MMS, Ling J, Ouellette BFF: **Atlas - a data warehouse for integrative bioinformatics.** *BMC Bioinformatics* 2005, **6**:34.
8. Birkland A, Yona G: **BIOZON: a system for unification, management and analysis of heterogeneous biological data.** *BMC Bioinformatics* 2006, **7**:70.
9. Antezana E, Blondé W, Egaña M, Rutherford A, Stevens R, De Baets B, Mironov V, Kuiper M: **BioGateway: a semantic systems biology tool for the life sciences.** *BMC Bioinformatics* 2009, **10(Suppl 10)**:S11.
10. **Web Ontology Language OWL/W3C Semantic Web Activity.** [http://www.w3.org/2004/OWL/].
11. **RDF - Semantic Web Standards.** [http://www.w3.org/RDF/].
12. Thorisson GA, Muilu J, Brookes AJ: **Genotype-phenotype databases: challenges and solutions for the post-genomic era.** *Nat Rev Genet* 2009, **10**:9-18.
13. Hu M, Mural R, Liebman M: *Biomedical Informatics in Translational Research* Artech House Publishers; 2008.
14. Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, Li P, Oinn T: **Taverna: a tool for building and running workflows of services.** *Nucleic Acids Res* 2006, **34**:W729-732.
15. Davidson SB, Wong L: **The Kleisli Approach to Data Transformation and Integration.** 2001, 135-165.
16. Stevens R, Baker P, Bechhofer S, Ng G, Jacoby A, Paton NW, Goble CA, Brass A: **TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources.** *Bioinformatics* 2000, **16**:184-186.
17. Mork P, Halevy A, Tarczy-hornoch P: **A model for data integration systems of biomedical data applied to online genetic databases.** *Proceedings of the Symposium of the American Medical Informatics Association* 2001, 473-477.
18. Belleau F, Nolin M, Tourigny N, Rigault P, Morissette J: **Bio2RDF: towards a mashup to build bioinformatics knowledge systems.** *J Biomed Inform* 2008, **41**:706-716.
19. Berners-Lee T, Hendler J, Lassila O: **The Semantic Web.** *Scientific American* 2001, 34-43.
20. Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragoso G, Sahni H, Gustafson S, Buetow KH: **caCORE: a common infrastructure for cancer informatics.** *Bioinformatics* 2003, **19**:2404-12.
21. Brookes AJ, Lehvaslaiho H, Muilu J, Shigemoto Y, Oroguchi T, Tomiki T, Mukaiyama A, Konagaya A, Kojima T, Inoue I, Kuroda M, Mizushima H, Thorisson GA, Dash D, Rajeevan H, Darlison MW, Woon M, Fredman D, Smith AV, Senger M, Naito K, Sugawara H: **The phenotype and genotype experiment object model (PaGE-OM): a robust data structure for information related to DNA variation.** *Hum Mutat* 2009, **30**:968-977.
22. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S: **The generic genome browser: a building block for a model organism system database.** *Genome Res* 2002, **12**:1599-1610.
23. Post LJG, Roos M, Marshall MS, van Driel R, Breit TM: **A semantic web approach applied to integrative bioinformatics experimentation: a biological use case with genomics data.** *Bioinformatics* 2007, **23**:3080-3087.
24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**:2498-2504.
25. Breitkreutz B, Stark C, Tyers M: **Osprey: a network visualization system.** *Genome Biol* 2003, **4**:R22.
26. Köhler J, Baumbach J, Taubert J, Specht M, Skusa A, Rüegg A, Rawlings C, Verrier P, Philippi S: **Graph-based analysis and visualization of experimental results with ONDEX.** *Bioinformatics* 2006, **22**:1383-1390.
27. Bornhövd C, Buchmann A: **A Prototype for Metadata-Based Integration of Internet Sources.** In *Advanced Information Systems Engineering. Volume 1626*. Springer Berlin/Heidelberg; 1999:439-445.

28. Noy NF, Crubezy M, Fergerson RW, Knublauch H, Tu SW, Vendetti J, Musen MA: **Protege-2000: an open-source ontology-development and knowledge-acquisition environment.** *AMIA Annu Symp Proc* 2003, 953.
29. Kaps A, Dyshlevoi K, Heumann K, Jost R, Kontodinas I, Wolff M, Hani J: **The BioRS(TM) Integration and Retrieval System: An open system for distributed data integration.** *JB* 2006, 3.
30. Losko S, Heumann K: **Semantic data integration and knowledge management to represent biological network associations.** *Methods Mol Biol* 2009, **563**:241-258.
31. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-402.
32. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.** *Bioinformatics* 2003, **19**:524-31.
33. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
34. Maglott D, Ostell J, Pruitt KD, Tatusova T: **Entrez Gene: gene-centered information at NCBI.** *Nucleic Acids Res* 2005, **33**:D54-8.
35. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2008, **36**:D25-30.
36. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35**:D61-D65.
37. Wain HM, Lush M, Ducluzeau F, Povey S: **Genew: the human gene nomenclature database.** *Nucleic Acids Res* 2002, **30**:169-171.
38. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M: **The Ensembl genome database project.** *Nucleic Acids Res* 2002, **30**:38-41.
39. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LL: **UniProt: the Universal Protein knowledgebase.** *Nucleic Acids Res* 2004, **32**:D115-9.
40. Kneale GG, Kennard O: **The EMBL nucleotide sequence data library.** *Biochem Soc Trans* 1984, **12**:1011-1014.
41. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stümpflen V, Ceol A, Chatr-aryamontri A, Armstrong J, Woollard P, Salama JJ, Moore S, Wojcik J, Bader GD, Vidal M, Cusick ME, Gerstein M, Gavin A, Superti-Furga G, Greenblatt J, Bader J, Uetz P, Tyers M, Legrain P, Fields S, Mulder N, Gilson M, Niepmann M, Burgoon L, De Las Rivas J, Prieto C, Perreau VM, Hogue C, Mewes H, Apweiler R, Xenarios I, Eisenberg D, Cesareni G, Hermjakob H: **The minimum information required for reporting a molecular interaction experiment (MIMIx).** *Nat Biotechnol* 2007, **25**:894-898.
42. Jameson D, Garwood K, Garwood C, Booth T, Alper P, Oliver SG, Paton NW: **Data capture in bioinformatics: requirements and experiences with Pedro.** *BMC Bioinformatics* 2008, **9**:183.
43. Losko S, Wenger K, Kalus W, Ramge A, Wiehler J, Heumann K: **Knowledge Networks of Biological and Medical Data: An Exhaustive and Flexible Solution to Model Life Science Domains.** In *Data Integration in the Life Sciences. Volume 4075/2006*. Springer Berlin/Heidelberg; 2006:232-239.
44. Sameith K, Antczak P, Marston E, Turan N, Maier D, Stankovic T, Falciani F: **Functional modules integrating essential cellular functions are predictive of the response of leukaemia cells to DNA damage.** *Bioinformatics* 2008, **24**:2602-7.
45. Selivanov VA, de Atauri P, Centelles JJ, Cadefau J, Parra J, Cussó R, Carreras J, Cascante M: **The changes in the energy metabolism of human muscle induced by training.** *J Theor Biol* 2008, **252**:402-410.
46. Good BM, Wilkinson MD: **The Life Sciences Semantic Web is full of creeps!** *Brief Bioinform* 2006, **7**:275-86.
47. **Foswiki.** [http://foswiki.org/].
48. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, Schacherer F, Martinez-Flores I, Hu Z, Jimenez-Jacinto V, Joshi-Tope G, Kandasamy K, Lopez-Fuentes AC, Mi H, Pichler E, Rodchenkov I, Splendiani A, Tkachev S, Zucker J, Gopinath G, Rajasimha H, Ramakrishnan R, Shah I, Syed M, Anwar N, Babur O, Blinov M, Brauner E, Corwin D, Donaldson S, Gibbons F, Goldberg R, Hornbeck P, Luna A, Murray-Rust P, Neumann E, Reubenacker O, Samwald M, van Iersel M, Wimalaratne S, Allen K, Braun B, Whirl-Carrillo M, Cheung K, Dahlquist K, Finney A, Gillespie M, Glass E, Gong L, Haw R, Honig M, Hubaut O, Kane D, Krupa S, Kutmon M, Leonard J, Marks D, Merberg D, Petri V, Pico A, Ravenscroft D, Ren L, Shah N, Sunshine M, Tang R, Whaley R, Letovksy S, Buetow KH, Rzhetsky A, Schachter V, Sobral BS, Dogrusoz U, McWeeney S, Aladjem M, Birney E, Collado-Vides J, Goto S, Hucka M, Le Novère N, Maltsev N, Pandey A, Thomas P, Wingender E, Karp PD, Sander C, Bader GD: **The BioPAX community standard for pathway data sharing.** *Nat Biotechnol* 2010, **28**:935-942.
49. Lloyd CM, Halstead MDB, Nielsen PF: **CellML: its future, present and past.** *Prog Biophys Mol Biol* 2004, **85**:433-450.

doi:10.1186/1752-0509-5-38
**Cite this article as:** Maier *et al.*: **Knowledge management for systems biology a general and visually driven framework applied to translational medicine.** *BMC Systems Biology* 2011 **5**:38.

# Chapter 3b

## A novel discrete model-driven approach unveils abnormal metabolic adaptation to training in COPD

**Igor Marin de Mas**[1,2], Erich Fanchon[3], Vitaly A. Selivanov[1,2], Josep Roca[4] and Marta Cascante[1,2,§]

[1]Department of Biochemistry and Molecular Biology, Faculty of Biology, Institute of Biomedicine of University of Barcelona (IBUB) and IDIBAPS, Unit Associated with CSIC, Diagonal 645, 08028-Barcelona, Spain

[2]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[3]Faculté de Médecine de Grenoble, BCM team, UJF-Grenoble 1, CNRS, Laboratoire TIMC-IMAG UMR 5525, 38041, Grenoble, France.

[4]Hospital Clínic, IDIBAPS, CIBERES; Universitat de Barcelona, Barcelona 08028, Spain

[§]Corresponding author

# A novel discrete model-driven approach unveils abnormal metabolic adaptation to training in COPD

**Igor Marin de Mas**[1,2], Erich Fanchon[3], Vitaly A. Selivanov[1,2], Josep Roca[4] and Marta Cascante[1,2,§]

168.

[1]Department of Biochemistry and Molecular Biology, Faculty of Biology, Institute of Biomedicine of University of Barcelona (IBUB) and IDIBAPS, Unit Associated with CSIC, Diagonal 645, 08028-Barcelona, Spain

[2]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[3]Faculté de Médecine de Grenoble, BCM team, UJF-Grenoble 1, CNRS, Laboratoire TIMC-IMAG UMR 5525, 38041, Grenoble, France.

[4]Hospital Clínic, IDIBAPS, CIBERES; Universitat de Barcelona, Barcelona 08028, Spain

[§]Corresponding author

**Abstract**

Skeletal muscle dysfunction is a systemic effect seen in approximately one-third of patients with Chronic Obstructive Pulmonary Disease (COPD). It has a multifactorial nature and shows association with high burden on health care and poor patient prognosis. Cell hypoxia, high reactive-oxygen-species (ROS) production and altered endurance training-induced adaptive changes are well recognized traits of the phenomenon. The current study describes a discrete model-driven method combining mechanistic and probabilistic approaches to decipher the role of ROS in the activity state of the gene regulatory network assessed before and after an 8-weeks endurance training program. The model predictions indicate abnormal training-induced responses at several levels in COPD patients, namely: TCA cycle, mitochondrial electronic chain, creatine kinase and insulin receptor factor leading to less efficient energy metabolism. The computational approach also predicts abnormal regulation of skeletal muscle cytokines (CDK4 and TNF-α) suggesting defective tissue remodeling. The model outcomes contribute to enhanced knowledge of altered skeletal muscle training-induced adaptations seen in COPD patients and they help to design new avenues for a cost-effective preventive therapies in patients with chronic diseases. Moreover, the novel modelling tools generated in the current research overcome limitations of previous computational approaches and shows high potential for exploring relevant characteristics and effects of interventions on highly interconnected complex biological networks.

## Background

Systems biology aims to explain physiology, development, and pathology based on modular networks of expression, interaction, regulation, and metabolism [1,2]. Gene regulatory networks (GRNs) [3] control the adaptation to environmental perturbations via cell type-specific gene expression and interactions between transcription factors (TFs) and regulatory promoter regions [4]. These mechanisms may affect the metabolic network by regulating the activity of key enzymes. One of the major challenges of the application of systems biology into the medical field is the study of the crosstalk between metabolic and gene regulatory networks and its role in cellular response to environmental conditions and external perturbations [5]. This is especially crucial in the case of multi-factorial diseases [6], in which multiple regulatory mechanisms lead the cell to an aberrant metabolism and abnormal adaptation to environment conditions that explain heterogeneities of disease progress and clinical manifestations among patients.

Chronic obstructive pulmonary disease (COPD) is a complex chronic disease caused by inhalation of irritants by susceptible individuals. Pulmonary manifestations of the disease are characterized by alterations in lung function such as expiratory flow limitation (low forced expiratory volume during the first second, $FEV_1$) and abnormal pulmonary gas exchange (arterial hypoxemia) [7]. COPD is a highly prevalent chronic condition in approximately 9% of the adult population above 45 years [7], imposing a high burden on health systems worldwide being the third leading cause of death in the United States [8]. Skeletal muscle dysfunction is a characteristic systemic effect of the disease that affects approximately one-third of patients [9]. It has a multi-factorial nature [9], but it is acknowledged that high reactive oxygen species (ROS) generation in the mitochondrial electronic chain (EC), as well as nitroso-redox disequilibrium, are likely play a relevant role in both pulmonary and systemic manifestations of the disease affecting the GRN [10]. Moreover, altered skeletal muscle redox status plays a

central role in the abnormal adaptive changes observed in COPD patients after endurance training [11]. It has been hypothesized that the systemic effects of the disease might share abnormal regulation of key metabolic pathways. Thus, the state of the mitochondrial EC depends on carbohydrate metabolism and transport of adenine nucleotides, which, in turn, depends on the mitochondria-cytoskeleton interactions. All these factors, as well as cellular oxygenation [12], can have a functional impact on the respiratory chain and on mitochondrial ROS production through complex III [13]. Different modelling approaches have been proposed to explore underlying mechanisms of complex biological processes, but none of them seems suitable for a proper characterization of highly inter-connected biological networks as those likely involved in abnormal skeletal muscle metabolic adaptations to training in COPD patients.

**Limitations of current modeling approaches**

A large number of data-driven approaches has been applied in the study of COPD [14, , 15]. Among these approaches are, Bayesian networks based correlation analysis [16.]. However, although Bayesian networks are able to assess causalities and may provide a rough idea of the processes associated with a given disease, they are not suitable to study complex mechanisms dynamics.

On the other hand, mechanistic modelling approaches, including kinetic or discrete models, are based on detailed representations of the system and their parameters. However, these approaches do not seem appropriate for exploring large scale and highly interconnected networks because of the paucity of precise knowledge on necessary parameter values. It is of note that discrete models, representing an abstraction and a coarse-graining of a network, can serve as a simple but efficient tool for the extraction of basic design principles of molecular regulatory networks, without having to deal with all the biochemical details [17.]. It is concluded that the study of large and highly inter-connected networks becomes unworkable due the exponential increase of parameters that cannot be compensated by the usually limited amount of

reliable experimental data. This lack of suitable computational tools makes that, in spite the efforts in this area, our knowledge of the mechanisms underlying the abnormal metabolic adaptation to training in COPD patients remains still limited.

The current research aims to develop a novel computational approach to overcome the modelling limitations alluded to above. To this end, we have reconstructed a discrete model that represents the mechanism of skeletal muscle-specific GRN associated to energy metabolism and used this model as platform to integrate constraints based on statistical analysis such as gene correlation analysis or gene differentially expression analysis

As a use case to explore the potential of the novel modelling approach, the current study analyzes how perturbation in ROS levels dynamically affects the mitochondrial GRN and metabolism and its effects on the abnormal adaptation to endurance training in COPD patients. To this aim, we have assessed three different groups of individuals s: i) healthy sedentary subjects (controls), ii) COPD patients with preserved skeletal muscle mass (normal Body Mass Index, BMI) and iii) COPD with skeletal muscle mass wasting (low BMI), before and after an 8-week high intensity endurance training program [18,19] using trancriptomic measurements carried out in skeletal muscle (vastus lateralis) biopsies representative of the 6 different states.That is, three groups of subjects studied before and after training.

Briefly, the novel computational approach has been carried out in two parts, as displayed in detail in **Figure 1**: i) GRN reconstructions and ii) integration of the reconstructed GRN including constraints based on gene correlation and gene differentially expression analysis into a discrete model-based analysis [31]

In order to reconstruct a GRN specific for COPD patients, we first performed a differential gene expression analysis on the 6 different states to determine a set of genes that will be the nodes of the network, next we used IPA software [20], DroID [21] and BioXm databases [22] to define the interactions between nodes and a manual curation to determine the sign and direction of the interactions. Next we integrated the

reconstructed GRN a into a discrete model. In this process the size of the GRN was reduced in a way that the mechanistic information remained in the model as is explained in [paper]. This model-driven approach enables to capture the dynamic processes underlying the behavior of complex systems such as GRN [23] or signaling pathways [24]. Finally, in order to constraint the space of feasible solutions, we integrated constraints in the model. These constraints were based on probabilistic analysis performed on transcriptomic data extracted from publicly available data-sets from GEO [25].

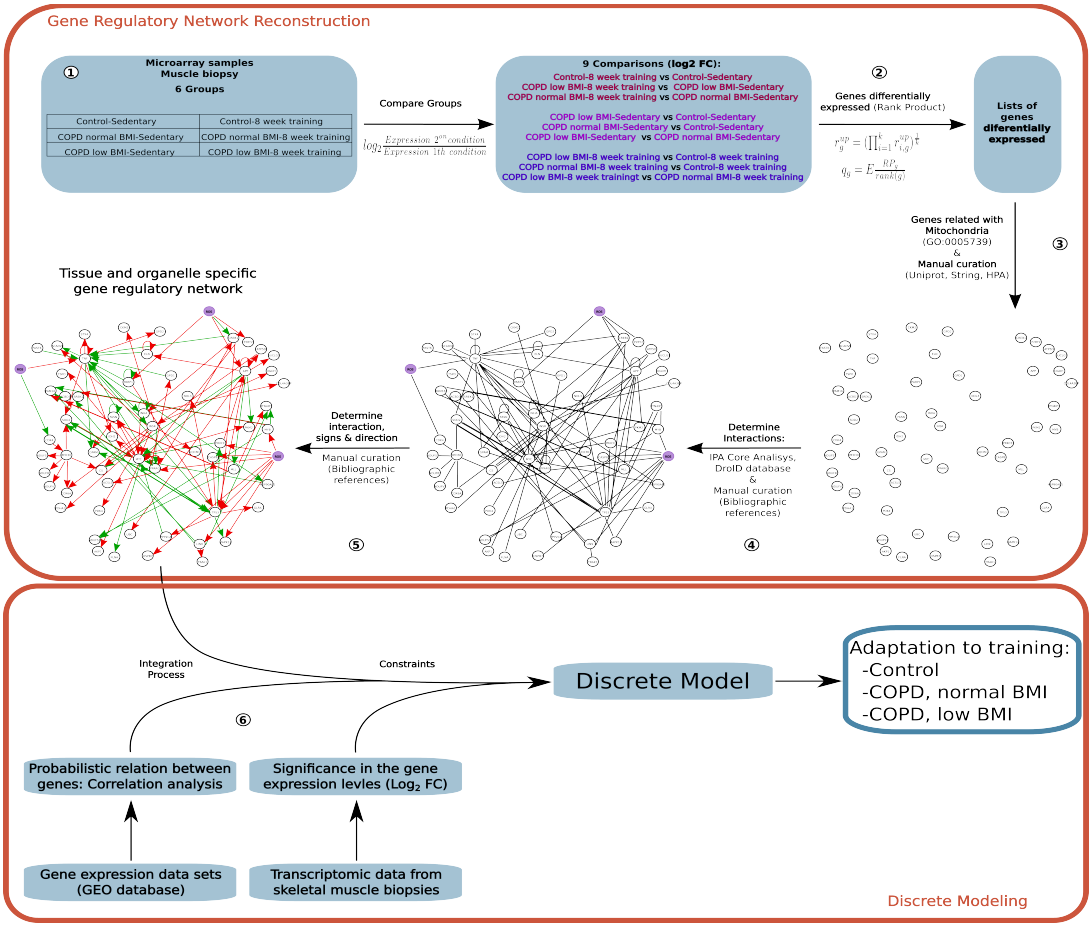


Fig. 1: Here are represented the process of reconstructing a  i) the process of GRN reconstructions based on gene expression data, ii) the integration of constraints based on probabilistic approaches and iii) build a discrete model by integrating the GRN reconstruction and the constraints. 1. The GRN reconstruction started by the generation of a draft, automated reconstruction based on the trancriptomic data previously normalized using RMA [26] of the six different states [18,19]. Since both, the energy metabolism and the processes responsible of ROS production occur mainly in the mitochondria, we the GRN reconstruction was focused on those genes differentially expressed between conditions and associated to some mitochondrial process. 2. This analysis was achieved by Rank Product method (RP) [27] to determine the gene candidates and 3. Gene Ontology and Human Proteins Atlas databases [28,29] to select the genes associated to some mitochondrial process. 4. Next, to determine gene associations we used IPA software and DroID [20,21]. 5. Finally, we manually curate the GRN reconstruction using a large number of bibliomic data sources. The main objective of curation was to identify and correct incomplete or erroneous annotation and identify the direction and the sign of the interactions. Once the GRN reconstruction was finished we converted it into a mathematical discrete model based on R. Thomas formalism [30] that describe mechanistically the interactions between those mitochondrial-associated genes that were differentially expressed between states.6. In order to refine the accuracy of our model predictions, we used public available muscle-related gene expression data-sets from Gene Expression Omnibus (GEO) database [25] to impose constraints to our model. We integrated these constraints in the form of inequalities based on probabilistic approaches. Thus, if we observed a strong correlation between two non-connected genes, their expression values were forced to evolve in the same direction. The rationale was just the opposite in the case of a strong anti-correlation.

Thus, in this work, we propose a method by which the interactions between genes are determined by performing a tissue and organelle specific GRN reconstruction and the constraints are defined using probabilistic approaches, finally both, the GRN reconstruction and the constraints are integrated into a discrete model in order to capture the mechanism triggering the abnormal metabolism observed in COPD patients.

Summing up, we have developed a new methodology that combines both probabilistic and deterministic analysis. This new approach represents a qualitative step forward in the study of complex and multi-factorial diseases like COPD and that can be extrapolated to other multi-factorial diseases.

**Results and discussion**

169. We focused on the anomalous muscle adaptation to training of COPD patients; more specifically, we investigated the different effects of ROS on the GRN that governs the energy metabolism in COPD patients compared with healthy people.

170. To this aim, we developed a computational approach to i) first, reconstruct a mitochondria-specific GRN using transcriptomic and bibliography-based data; and, ii) integrate the reconstructed GRN and data from public data sets into a discrete model based on R. Thomas formalism.

171. In this section we present i) the results of the mitochondria-specific GRN reconstruction analysis that captures the essential actors of the mitochondrial energy metabolism , ii) the building of a discrete model [23] and the integration of constraints from gene expression data-sets based on probabilistic approaches to determine the mechanisms governing the abnormal adaptation to training of COPD patients iii) the application of this computational method to predict the mechanism underlying the abnormal adaptation to training of COPD patients compared with healthy people; and, finally iv) the validations of the model predictions.

**i) Mithochondrial-specific Gene Regulatory Network reconstruction of COPD adaptation to training**

The reconstruction of a GRN that integrates information of the regulatory mechanisms underlying the different adaptation to training in COPD was done in three steps: i) Gene expression analysis, ii) Automated network reconstruction and iii) Manual curation of the network reconstruction.

Gene expression analysis: In this study we used transcriptomic data (Affymetrix genechip u133a 2.0) from skeletal muscle biopsies (vastus lateralis) of the six states previously mentioned. We used this data to reconstruct a GRN that integrates the effects of both training and COPD disease.

First, the gene expression data was normalized by robust multi-array average (RMA) [26]. Next we compared the gene expression levels of each group before and after training as well as the differences between groups before and after the training program (9 comparisons) (see Methods). Thus, for each gene we obtained nine $\log_2$ Fold changes (FC) corresponding to each comparisons. This measurements provided a numerical value describing how much up or down regulated is a gene in one group compared with the same gene in other group. Based on the $\log_2$ FCs and using Rank Product method (see Methods) we determined which genes were differentially expressed [27]. We considered those genes with p' and q parameters below 0.001 and 0.01 respectively as differentially expressed (see Methods). Since skeletal muscle mitochondrial dysfunction is a central actor in COPD [31] and ROS is mainly a sub-product of the mitochondrial energy metabolism, we focused our network reconstruction on mitochondrial processes. To this aim, we used Gene Ontology Database (GO) [25] to filter those genes related with some mitochondrial process (GO:0005743). Finally we obtained a list of 105 genes that changed significantly between the six states. In addition, the list was manually curated by using String [33], Uniprot [34] and Human Protein Atlas [29] databases in order to ensure that all the genes and the proteins that they encoded were related with some mitochondrial activity and were expressed in human skeletal muscle. These genes were those used to perform the GRN reconstruction. The list of genes differentially expressed can be found in the Supplementary material 1.

Automatic GRN reconstruction: We used the genes defined in the previous section as differentially expressed between states as nodes in the GRN reconstruction. The reconstructed interaction graph integrates the changes and the regulatory mechanisms related with the adaptation to training in both, COPD and healthy groups, as well as the effects of the disease. To this aim, we used the list of genes to perform an Ingenuity Core Analysis (see Methods). This analysis is included in Ingenuity IPA software [20] and allows the integration of omic data together with a large number of databases to explore the molecular interactions focused on the list of genes differentially expressed

previously defined. We performed this analysis including not only reported interactions in human skeletal muscle but also in other human excitable tissues such as other types of muscle, neuronal tissue, .., etc. In order to connect those genes from which was no reported interaction in human, we included putative interaction described in Drosophila by using DroID database [21]. In addition, we included ROS as a new node in the network and its interactions based on the literature [34-52].The graphical representation of this analysis is shown in the figure 2A and in more detail in the supplementary material. The figure represents an interaction network by connecting nodes representing the genes differentially expressed and nodes included in the analysis in order to integrate all the gens in a single regulatory network.

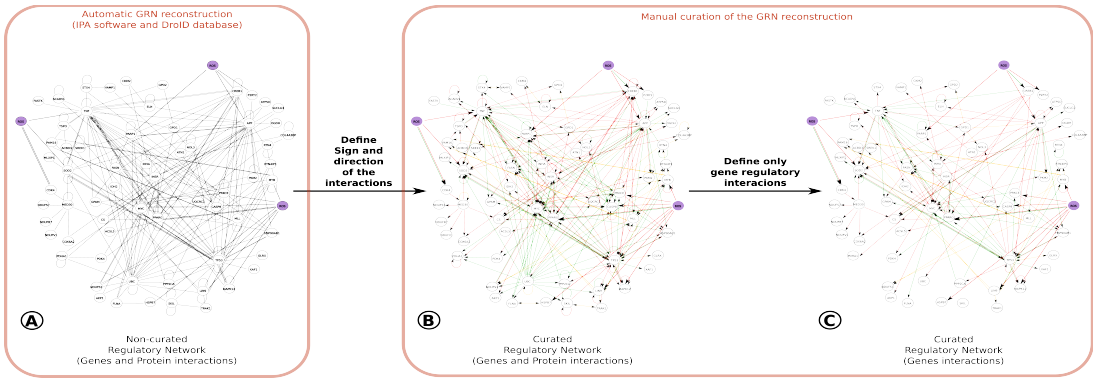



Fig. 2: This figure represents the different steps of the reconstruction of a GRN. The nodes represents the genes with significant differences in the gene expression between groups or genes included to construct a single GRN, the violet nodes represent ROS, the lines between nodes represent the interactions between them, he green arrows are negative interactions (inhibition of the gene expression, ubiquitination, ...), the positive interactions are represented with red arrows, finally those interactions in which the direction and/or the sign is unknown are colored in yellow. A: Automated interaction network reconstruction generated by IPA software and DroID database of the list of genes differentially expressed between condictions. B: Manual curation of the automated reconstructed interaction network. Here are defined the direction and the sign of the interactions based on a large set of bibliiographical references. C: Interaction network considering only gene regulatory interactions. This network was mathematically represented as a discrete model in order to determine the mechanism associated to the abnormal adaptation to training in COPD patientes.

**(i) Curation of the GRN reconstruction**: In order to correct incomplete or erroneous annotations and identify the direction and the sign of the interactions, we manually curated the GRN reconstruction based on a large number of bibliographic data sources [53-179]. The resulting interaction network is represented in the figure 2B and in more detail in the supplementary material.

We can observe that is possible that exists more than one interaction between two nodes with the same direction but with different sign. This is because the effect of the expression of a certain affector gene on a target gene or protein can have different sign depending on the concentration of the affector. This means that depending on the protein or RNA levels the interaction can be an activation or an inhibition. Some examples of this dual behavior are the effect of ROS on PPARG and CDK4 or the influence of MYC on CASP8. Thus, the resulting gene regulatory interaction network accounted for 65 nodes (64 genes + ROS) and 334 interactions. The detailed list of interactions and nodes can be found in the Supplementary material (Supplementary material 2).

As was previously explained, the interaction network includes gene regulatory interactions and protein-protein interactions such as phosphorilations, ubiquitinations, …, etc. However, our experimental data is based on transcriptomic data. Thus, we only consider the gene regulatory interactions to be integrated into the discrete model analysis. The graphical representation of the resulting interaction network can be found in the figure 2C and in more detail in the Supplementary material. The resulting interaction network allowed us explain all the changes in the gene expression levels that were observed experimentally.

Thus, we performed the reconstruction of a mitochondria-specific GRN. In this process we defined the nodes of the network based on those genes that changed more drastically between conditions and their corresponding regulatory interactions were defined based on information extracted from databases (IPA analysis and DroID database) and a thorough manual curation. Since this GRN capture the essential genes

and interactions related with the response to an endurance training and with COPD disease, it offers an excellent platform to integrate any sort of omic data into a discrete model analysis allowing the identification of the dynamics governing the abnormal adaptation to training in COPD patients compared with a control group.

**ii) Discrete modeling building and probabilistic-based constraint integration**

GRN integration into a discrete model. Next, we integrated the information embedded in the GRN previously reconstructed into a discrete model. For this aim we used SysBiOX framework [180]. This software is implemented in ASP (Answer Set Programming) and allows to integrate the interactions occurring in a transition graph into a discrete model based on the multilevel logical formalism developed by R. Thomas et al. [30]. By using this discrete model-driven approach we inferred the gene expression and protein concentration levels using ROS as source of perturbation in the different conditions. In order to simplify the GRN conserving the equilibrium and stabilities of the initial network we performed a reduction transformation that permitted to simplify the analysis (See methods) and reduce the calculation time [181].

Integration of constraints based on probabilistic approaches: In order to reduce the number of models that satisfy the constraints imposed by the topology of the network we integrate constraints from different sources:

a) Constraints based on the transcriptomic data from muscle biopsies: We constrained the space of feasible solution by using the FC calculated in the previous section in the form of inequalities. Then, for example these inequalities can express that the expression of a certain gene is higher or equal (≥) in one condition than in other. Next example illustrates how these FC were used. Considering the FC of the gene GPAM and comparing the three groups in sedentary conditions we have the following $\log_2$FC values: “COPD with low BMI vs Control”: 0.95044, “COPD with normal BMI vs Control”: 0.47991, “COPD with normal BMI vs COPD with low BMI”: 0.77446. Then this information was introduced in the model as follows: $GPAM_{\text{COPD with low BMI-}}$

sedentary ≥ $GPAM_{COPD}$ with normal BMI-sedentary ≥ $GPAM_{Control\text{-}sedentary}$. In this manner excluded from our analysis all the solutions in which $GPAM_{COPD}$ with normal BMI-sedentary was higher than $GPAM_{COPD}$ with low BMI-sedentary or in which $GPAM_{Control\text{-}sedentary}$ was higher than $GPAM_{COPD}$ with normal BMI-sedentary or than $GPAM_{COPD}$ with low BMI-sedentary.

b) Constraints based on correlation analysis using gene expression data set: In order to improve the accuracy of model predictions we introduced constraint based on the correlation analysis done on gene expression data sets extracted from Gene Omnibus (GEO) database [25]. We include in this analysis datasets from [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array (GPL570), [HG-U133A] Affymetrix Human Genome U133A Array (GPL96) and Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (GPL6480), more specifically those datasets related with muscle sample measurements. We also integrated information from the correlation analysis performed on GSE3078 dataset that corresponds to experiments measuring the transcriptome changes in response to hydrogen peroxide that simulates different levels of ROS.

We determined the correlation between the genes in order to introduce this information as constraints in our model (see Methods). We consider as significant those correlations with a r value higher than 0.85 and a corrected p value lower than 0,01. GSE3078 dataset was used to establish correlations that constraints states with reported differences between their ROS levels, i.e. Control group and COPD groups before the training program [31], as the other datasets contains expression levels from healthy people, we use them to impose constraints in control group before and after training. These results of the correlation analysis can be found in the Supplementary material (Supplementary material 4).

**iii) Discrete model combining gene regulatory mechanism and probabilistic-based constraints predicts abnormal metabolic adaptation to training in COPD patientg**

Once the GRN was reduced and integrated into a discrete model together with the

constraints previously mentioned, we analyzed the system in order infer the activity state of the nodes of the network in the six different conditions. For this aim we impose an objective function that minimize the discrete value of the nodes while satisfying the constraints embed in the network (by the network topology) and the constraints defined in the probabilistic analysis.

Finally we obtained 36 possible models that were consistent with the network topology and the constraints that we had imposed. These models were analyzed in order to determine how the discrete value of each gene evolved in response to the training program.

According to model predictions COPD patients have an abnormal adaptation to training compared with control group. In the figure below are represented the changes predicted by the model in the three groups of study in response to the training program:

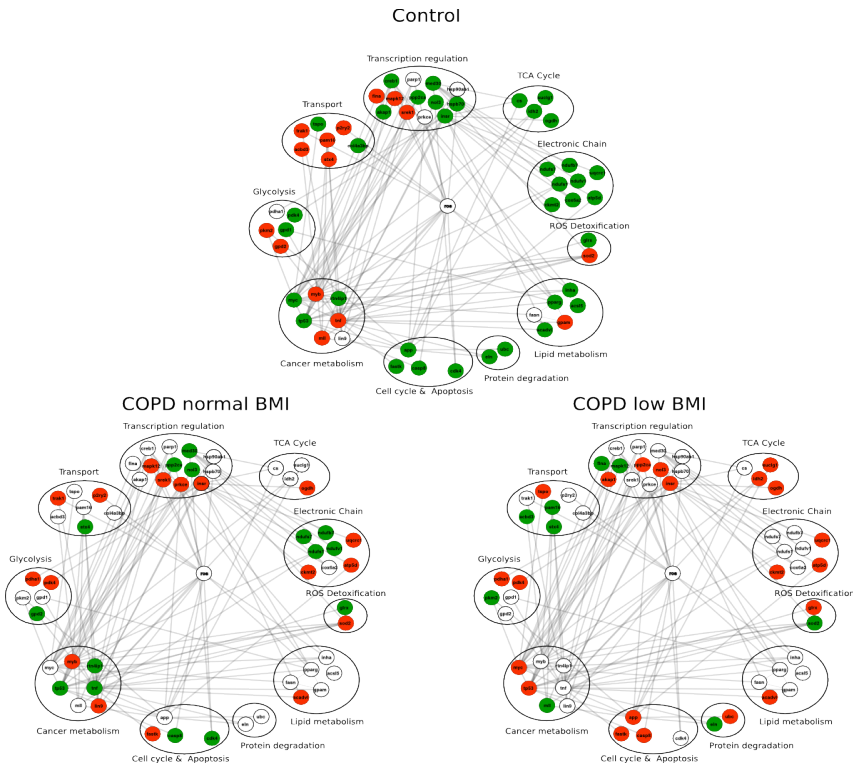



Fig. 3: In this figure is represented the same network than in figure 2A but centered on ROS. The nodes

are grouped based on their molecular function. Green nodes are the nodes that is predicted to increases their expression after training, red nodes are predicted to decrease their expression after training and white nodes are the nodes that is not predicted to change their expression after training or the the model is not able to predict the behavior of the node (except ROS).

In the figure above, we can observe that the exercise provoke more changes at transcriptomic level in the control group than in both COPD groups. The changes represented in the figure correspond only to those nodes in which response to training was constant among all the solutions. Then, for example all the genes encoding subunits of electronic chain complexes increase their activity in control group and this prediction is supported by the 36 models.

The following table summarizes the results obtained in the analysis:

| Node | Control: sedentary vs trainned | COPD normal BMI sedentary vs trainned vs | COPD low BMI sedentary vs trainned |
|---|---|---|---|
| ros |  |  |  |
| atp5d |  |  |  |
| cox6a2 |  |  |  |
| ndufs1 |  |  |  |
| ndufv1 |  |  |  |
| ndufc1 |  |  |  |
| ndufs7 |  |  |  |
| uqcrc1 |  |  |  |
| ckmt2 |  |  |  |
| cs |  |  |  |
| idh2 |  |  |  |
| ogdh |  |  |  |
| suclg1 |  |  |  |
| pkm2 |  |  |  |
| pdha1 |  |  |  |
| pdk4 |  |  |  |
| gpd2 |  |  |  |
| gpd1 |  |  |  |
| sod2 |  |  |  |
| glrx |  |  |  |
| insr |  |  |  |
| mapk12 |  |  |  |
| nol3 |  |  |  |
| ppp2ca |  |  |  |
| srek1 |  |  |  |
| hspb7 |  |  |  |
| med30 |  |  |  |
| creb1 |  |  |  |
| parp1 |  |  |  |
| flna |  |  |  |
| akp1 |  |  |  |
| prkce |  |  |  |
| hsp90ab1 |  |  |  |
| tnf |  |  |  |
| tp53 |  |  |  |
| rtn4ip1 |  |  |  |
| lin9 |  |  |  |
| myb |  |  |  |
| myc |  |  |  |
| mll |  |  |  |
| stx4 |  |  |  |
| p2ry2 |  |  |  |
| trak1 |  |  |  |
| tspo |  |  |  |
| pam16 |  |  |  |
| acbd3 |  |  |  |
| col4a3bp |  |  |  |
| acadvl |  |  |  |
| fasn |  |  |  |
| gpam |  |  |  |
| pparg |  |  |  |
| inha |  |  |  |
| acsl5 |  |  |  |
| casp8 |  |  |  |
| cdk4 |  |  |  |
| app |  |  |  |
| fastk |  |  |  |
| eln |  |  |  |
| ubc |  |  |  |

Table 1: Adaptation to training: In the first column are the gene names, from the second to fourth column are represented the adaptation to the training program of the three different groups using the following code: the red arrows pointing down describe a reduction in the discrete value of the node in response to training predicted by all the models, the green arrows pointing up represent an increment in the discrete value of the node in response to training predicted by all the models, red right arrows represent nodes with no adaptation to training remaining the value of the node equal to 0 predicted by all the models, green right arrows represent nodes with no adaptation to training remaining the node in its maximum discrete value predicted by all the models, green lines represent an increment in the discrete value of the node in response to training predicted by between the 99 and the 70 % of the models, red lines represent a decrease in the discrete value of the node in response to training predicted by between the 99 and the 70 % of the models, black lines represent nodes with no consistent prediction

In the table above is represented the different adaptation to training in the three groups of study. As opposite to healthy group, the model predicted a down-regulation in the activity of EC, TCA cycle and creatine kinase as response to an 8-weeks training program in COPD patients with low BMI. Interestingly, for COPD patients with normal BMI, the model predicts an up-regulation of complex I (as it occurs in healthy patients) and an slight increase in Citrate synthase and Succinyl-CoA ligase. However, the model also predicts that complexes III and V, other key enzymes of TCA cycle, and creatine kinase activities decrease after exercise which suggests a defective energetic metabolism adaptation to training in in COPD patients with normal BMI. The adaptation to training in healthy group provides a higher capability to consume O2 and produce energy. The model also predicts that healthy group up-regulates cdk4 and down-regulate tnf in response to exercise whereas COPD normal BMI are able also to up-regulate cdk4 but by contrary increase tnf in response to the exercise. COPD with low BMI are not able to respond at all to exercise with respect cdk4 that remains low and also present the abnormal increase of tnf displayed by normal COPD patients in response to exercise. This facts indicate defective muscle remodeling pathways in COPD patients. The fact that both cdk4 and tnf presents opposite patterns of regulation in response to exercise in low COPD patients suggest that muscle remodeling is more defective in these patients. Interestingly Glycerol-3-phosphate dehydrogenase cytosolic

isoform is over-expressed in control group and under-expressed in COPD groups, we can observe the opposite pattern in the mitochondrial isoform. The mitochondrial isoform catalyze a reaction where NADH is produced, probably mitochondrial Glycerol-3-phosphate dehydrogenase mitochondrial isoform is over-expressed in COPD groups after training to partially compensate the reduction in E.C. activity. The insulin receptor factor (insr) is predicted to have a different response to training in COPD patients compared with healthy group that up-regulates its expression after training. The insulin receptor factor triggers a number of signaling events, which one of the most important effects is to produce a translocation of the glucose transporter SLC2A4/GLUT4 from cytoplasmic vesicles to the cell membrane, it facilitates glucose transport increasing the energy metabolism. Then, the down-regulation of insr predicted in both COPD groups would reduce the capability of the muscle to use glucose as source of energy. The model also predicted differences in acadvl between healthy group and both COPD groups. The acadvl (Very long-chain specific acyl-CoA dehydrogenase) is a key player in beta-oxidation that produces Acetyl-CoA that fuels TCA Cycle and finally the EC to produce ATP. This enzyme is up-regulated after training in healthy group and down-regulated in both COPD groups. Then, unlike the control group, COPD patient reduce the expression of this key enzyme in the beta-oxidation that in turns deprives the energy metabolism. We also observe differences in the nodes related with glycolisys. Thus, the model predicts that pdha1 is up-regulated in control group, the same protein is inactive in both COPD groups before and after the training. This enzyme plays a key role in glycolisis and TCA cycle metabolism catalyzing the reaction that decarboxilates pyruvate to AcoA fueling TCA cycle. Finally, the model was able to predict a reduction of ROS in control group after training, this prediction is consistent with previously reported observations [31], it also predicted a reduction of ROS levels in COPD group with normal BMI, on the other hand the levels of ROS in COPD group with low BMI was predicted to increase after training.

**iv) Validation**

GRN reconstruction: In order to evaluate the consistency of the network topology with

the experimental observations imposed as constraints into the model, we transform the discrete model into a boolean model. Boolean models are more restrictive in the sense that they have a lower number of possible solutions. Thus, following this rationale if boolean network was consistent with the experimental data, then the discrete model would be consistent too. This transformation is achieved by forcing to all the thresholds of each node to be equal. The network was consistent with the experimental data giving at least one solution (data not shown) and this validates the GRN reconstruction.

ROS level predictions: In order to validate the levels of ROS predicted by the model in the different conditions we eliminated all the inequalities that directly constrained ROS. Next we performed the same analysis described in "ii". This non-constrained-ROS model was able to predict the same levels of ROS than the constrained-ROS model that were supported by several experimental observations [31,182] validating the evolution of ROS in the different groups predicted by the model.

Gene expession prediction: Additionally, we validated the predictions for the other nodes of the network by using two different strategies.

First, the model was asked for the state of one specific gene when the constraints related with that gene were removed. Here we expected that the evolution of the gene from sedentary to trained condition was the same than in the model considering for all the constraints. This analysis was performed with all the genes of the network excepting ros. In all the analysis, the evolution of the activity state of the asked gene, from sedentary to trained, was the same than in the analysis considering all the constraints, validating our model predictions (Supplementary material 5).

Secondly, we used three gene expression data set publicly available of experiments where one of the genes of the network was experimentally knocked-out in mutant mice skeletal muscle [GEO Database Reference Series: GSE7162, GSE25072 and GSE2236]. More specifically these knock-down genes were myc, tp53 and sod2, which represent some of the most highly connected nodes in the COPD-specific GRN.

We integrated the corresponding gene expression in the form of constraints into the model as is explained in methods, excepting those constraints related with the knocked-down gene. Here we expected that the discrete value of the knocked-down gene in the mutant mice was predicted to be lower than in control mice with a normal expression of the given gene. In the three cases, the model was able to predict the knocked-down gene, which supports our model predictions (Supplementary mateiral 5).

**Discussion**

In this paper, we have investigated the mechanisms underlying the abnormal adaptation to training associated to COPD disease. More specifically, it has been focused on determine how ROS affects the gene regulatory mechanisms associated to the energetic metabolism before and after 8 weeks of training program. It has been carried out in two steps, first we reconstruct a tissue, organelle and condition specific GRN. It has been achieved by integrating transcriptomic and literature-based data. The resulting reconstruction integrates the most relevant mechanisms that regulate gene expression in response to training in healthy people and COPD patients. Next we integrated the GRN reconstruction into a discrete model offering a platform to further integrate omic data from different sources as well as simulate the evolution of the regulatory mechanism due to perturbations produced by the levels of ROS in response to training. Next we performed a correlation analysis on an extensive number of gene expression data-sets. We developed an strategy that allowed us to integrate the information extracted from this probabilistic analysis into the mechanism described in the GRN reconstruction in the form of constraints. Once we integrated the GRN reconstruction and the constraints from the correlation analysis into a discrete model, we determined the activity state of the GRN of the three groups in response to training. Gathering the model predictions we have an overview in which healthy people adapts to training by increasing the activity of those pathways involved in energy production, by contrary, this machinery has a defective adaptation to training in COPD patients with normal

BMI and is more severe in COPD group with low BMI. The abnormal adaptation also affects genes and proteins related with proteolysis, it could explain the lose of muscle mass observed in the most severe COPD patients.

Summing up, in this paper we have presented the development of different strategies that allowed us to reconstruct a highly curated COPD-specific GRN and the integration of the mechanism described by this network with information extracted from probabilistic approaches. It has permitted us to unlock important mechanisms underlying the abnormal adaptation to training in COPD patients and the role of ROS in this disease. All this shows the potential of this method and opens an avenue of possible strategies to study complex diseases such as COPD or other chronic diseases providing new therapeutic targets or better tools to design more efficient strategies to mitigate the effects of the disease.

## Methods

### Experimental data

In this work we have studied the abnormal muscle adaptation to training associated to COPD disease and the role of ROS in the activity state of the GRN in six different states: Control grup (12 patients), COPD with normal BMI (13 patients) and COPD with low BMI (6 patients) [18] before and after a 8 weeks training program [19].
We determined the gene expression levels of each patient by microarray analysis (Affymetrix genechip u133a 2.0) of their skeletal muscle tissue biopsies (bastus lateralis). Next, the gene expression data was normalized by RMA [26].

### Rank Product

Fold change: Once the microarrays data was preprocessed as is described in [26], we determined the effect of training and the effect of COPD on the transcriptomic profile of the six groups by calculating the log2 FC from the following comparisons:

A) Trained Control vs Sedentary Control

B) Trained COPD with low BMI vs Sedentary COPD with low BMI

C) Trained COPD with normal BMI vs Sedentary COPD with normal BMI

D) Sedentary COPD with low BMI vs Sedentary Control

E) Sedentary COPD with normal BMI vs Sedentary Control

F) Sedentary COPD with low BMI vs Sedentary COPDwith normal BMI

G) Sedentary COPD with low BMI vs Sedentary Control

H) Sedentary COPD with normal BMI vs Sedentary Control

I) Sedentary COPD with low BMI vs Sedentary COPDwith normal BMI

This measures give a numerical value describing how up or down regulated is a gene in one group compared with the same gene in other group as is described by the following expression:

EMBED Microsoft Equation 3.0259 $$FC = \log_2\left(\frac{\text{Expression } 2^{nd} \text{ condition}}{\text{Expression } 1^{st} \text{ condition}}\right)$$

(1)

Thus, if the log2 FC is positive for a given gene, it is overexpressed in the second condition compared with the first condition and vice versa if it is negative.

Rank product: The rank product (RP) method is a widely accepted non-parametric technique for identifying differentially expressed genes in replicated microarray experiments under two experimental conditions or groups [27 ]. It has also been widely applied to other omic datasets, for example in proteomics and metabolomics [183,184]. This method uses any measure that compares two conditions like the logarithm based two fold change (log2 FC), to entail ranking genes according to their differential expression within each replicate experiment and subsequently calculating the product of the ranks across replicates. This approach assumes that i) only a minority of genes are affected by relevant expression changes, ii) the measurements are independent between replicate arrays, iii) most changes are independent of each other and iv) the variance of the measurements is similar for all genes.

This approach uses three parameters to identify genes that are differentially expressed between different conditions: i) RP gives a numerical value of how up or down regulated is a given gene between two different conditions, ii) p', this statistic quantify the statistical significance of the RP and iii) q that represents percentage of false-

positives.

Thus, for an experiment examining n genes in k replicates, assuming that the genes are randomly distributed in the k lists, the probability that a given gene was at the top of each list (rank 1) is $1/n^k$. Generally, for each gene g, one can calculate the corresponding combined probability to be at the top of each list (rank 1) as a rank product ($RP_{up}$):

$$RP_g^{up} = \prod_{i=1}^{k} \left( \frac{r_{i,g}^{up}}{n_i} \right) \qquad (1)$$

Where k is total number of replicates, $r^{up}_{ig}$ is the position of gene g in the list of genes in the $i^{th}$ replicate sorted by decreasing FC that is normalized by the length of the $i^{th}$ replicate ($n_i$). The most strongly up-regulated gene would have a $r^{up}=1$. Likewise $r^{down}=1$ corresponds to the most strongly down-regulated gene. The RP values are calculated over all possible pairwise comparisons. Thus, the genes with smallest RP ($RP^{up}$ or $RP^{down}$) values are those that are more differentially expressed. Next, in order to correct for the fact that the pairwise comparisons between samples are not independent is necessary to determine the significance of gene expression changes and which genes are likely to be truly differentially expressed. Then, for an experiment with k replicates and n genes, is defined the statistic p'. It is calculated by multiplying RP by a factor F representing the number of possible products of k numbers smaller than n that are equal to the numerator of the RP value:

$$p_g^{'} = \mathrm{RP}_g . F \quad (2)$$

The F factor can be determined by using a simple permutation-based estimation procedure [185]. Other approaches propose a continuous gamma distribution approximation for the log-transformed rank products [186] or an exact distribution of rank products [187]. In this manner we can estimate how likely it is to observe a given RP value or better in a random experiment. In addition, is necessary to calculates the conservative estimate of the percentage of false-positives (PFP) for each gene g by

calculating the statistic q as follows:

$$q_g = \text{E}.\frac{\text{RP}_g}{\text{rank}(g)} \quad (3)$$

Here, rank(g) gives the position of gene g in a list of all genes sorted by increasing RP value. This estimates the PFP [188] providing a significance level to the observed changes on each gene. Now, one can decide how large a PFP would be acceptable and extend the list of accepted genes up to the gene with this $q_g$ value.
Thus, RP method offers reliable results in highly noisy data. This approach is powerful for identifying biologically relevant expression changes and can lead to a sharp reduction in the number of replicate experiments needed to obtain reproducible results.

**Gene Regulatory Network reconstruction**

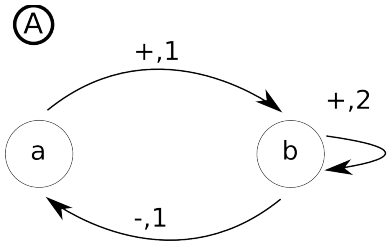
In order to reconstruct a GRN considering the effect of both training and the disease, we did a first draft automated GRN reconstruction. For this aim we used Ingenuity Pathway Analysis (IPA) software and the IPA database (Ingenuity® Systems, www.ingenuity.com) to establish the relationships between those genes defined as differentially expressed. This software provides a platform that allows to integrate different layers of information such as genomics or proteomics and its graphical representation in a single interaction network. IPA Core Analysis can be used to link genotypes (molecular profiling data) to phenotypes and molecular events (Cellular and Disease Processes, Signaling and Metabolic Pathways, Molecular Networks), defining the reported interactions between genes, DNA, proteins, etc [20].
Thus, by this analysis we reconstruct a first draft automated GRN. However, some genes still remained unconnected. It could be mainly because a paucity of reported regulatory interactions in human skeletal muscle tissue. In order to bridge the existing gaps, we imposed putative interaction by expanding our analysis first to other muscular tissues in human (smooth and cardiac), next to other excitable tissues in human (i.e. neurons) and finally to Drosophila's gene regulatory network (DroID database) [21].

For some of the new putative interactions was necessary to include new genes into the network. We choose these new nodes (genes) trying to use the minimum amount of new genes to connect the unconnected genes. Additionally for the genes extracted from Drosophila's interaction network we imposed that their human homologous were expressed in skeletal muscle ( by using Atlas protein database [29])

Finally, we manually curated the GRN reconstruction using data from a large number of bibliographic data sources [31,164]. The main objective of literature curation was to identify and correct incomplete or erroneous annotation and identify the direction and the sign of the interactions.

**Discrete modeling**

An increasing number of evidences point the fact that there is no one-to-one mapping between genotype and phenotype. In order to elucidate the fundamental principles that govern how genomic information translates into organismal complexity, is necessary to overcome the current habit of ad hoc explanations and instead embrace novel holistic approaches that involve computer modeling. Most modeling approaches aim at recreating a living system via computer simulation, by including as much details as possible. In contrast, boolean and discrete network models represent an abstraction and a coarse-graining of a network, such that it can serve as a simple, efficient tool for the extraction of the very basic design principles of molecular regulatory networks (such as a gene regulatory network (GRN) or a signaling pathways), without having to deal with all the biochemical details [17.]. Typically, the mathematical description of GRN using discrete modeling approach is based on the asynchronous logical description proposed by R. Thomas [30 ]. This approach allows a semi-qualitative analysis of complex biological networks that describes the dynamics of the system [180]. GRN are often represented in the form of an interaction graph, where nodes represent genes and arrows represent interactions between genes. The mathematical equations expressing these relationships are represented in the form of focal equations in the figure 6.b. These equations determine the trend towards which each node evolves depending of the current state of the network.




B

$$X_a = K_a * s^-(x_b, \theta_b^1) + K_a^b * s^+(x_b, \theta_b^1)$$

$$X_b = K_b * s^-(x_a, \theta_a^1)s^-(x_b, \theta_b^2) + K_b^a * s^+(x_a, \theta_a^1)s^-(x_b, \theta_b^2) + K_b^b * s^-(x_a, \theta_a^1)s^+(x_b, \theta_b^2) + K_b^{ab} * s^+(x_a, \theta_a^1)s^+(x_b, \theta_b^2)$$

Fig. 6: Interaction graph. A: The figure illustrates a network with two genes represented as node “a” and “b” and the interactions are represented by arrows where + and - represents activations and inhibitions respectively and the number represents the thresholds above which the interaction is active (i.e. a activates b when a is above threshold 1). The concentrations take discrete values, each one representing an interval between two consecutive thresholds. Then, taking for instance the node a (that can represent whether a protein or a transcript) if its logical variable $x_a$=0 then the concetration of “a” is bellow the threshold $\theta^1_a$, whereas $x_a$=1 means that the concentration of “a” is higher than (or equal to) $\theta^1_a$. Here the threshold $\theta^1_a$ determines the minimum amount of “a” necessary to activate b. B: Focal equations relating a state characterized by the vector of protein concentrations [xa,xb]and its focal state [Xa,Xb].

Thus these equations defines the dynamic behavior of the network that can be represented as a state in a transition graph. This graphical representation shows how the network goes from one state to the next depending on its structure. Each state of the network is represented by a vector of protein concentrations that take discrete values that represent the interval between two consecutive thresholds. These states are represented in Fig 7 where the arrows are the transition between each state and its possible successors.

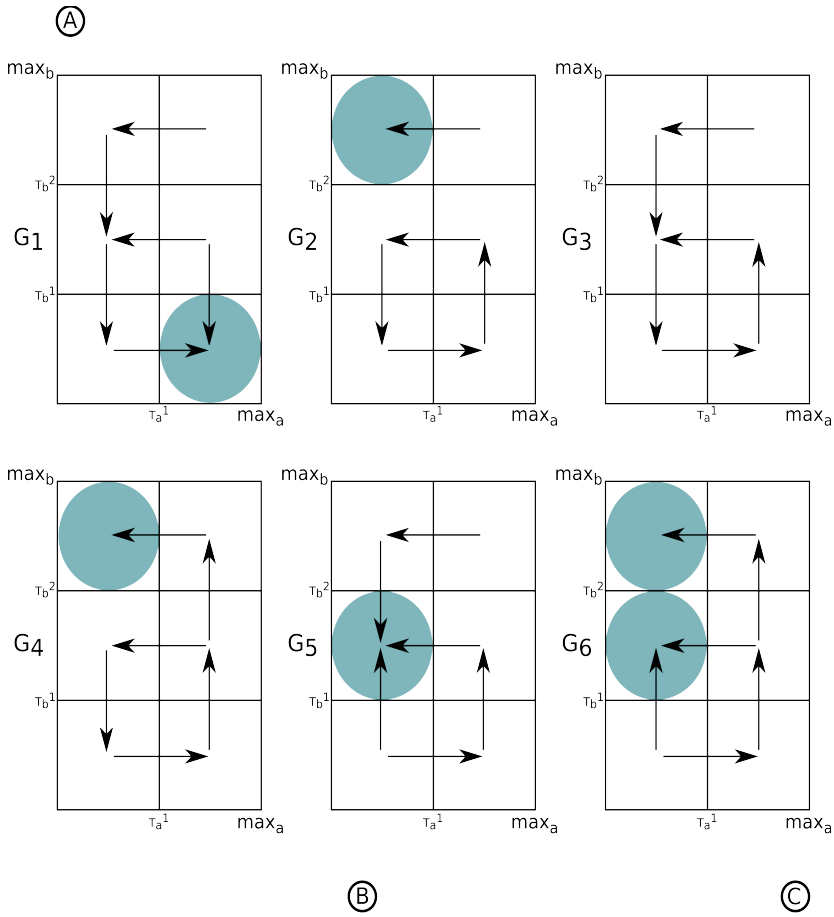



Fig. 7: Transition graph. A: Transition graphs G1,...,G6 of the instantiated models satisfying consistency e(Ce). Ce is the set of (b) observability and (c) additivity constraints for the example in Fig. 6

A specific attractor value called focal state ,[X1,...,Xn], is associated to a given state S. It represents the expression levels toward which the genes tend to evolve. A successor of S in the graph,S'=[x'1,...,x'n], is deduced by comparing the value of a logical

variable with that of its focal state. The transition of S to S' is assumed to be asynchronous, in the sense that at most one logical variable xi is updated at a time. If the variable xi is updated, the formal relationship between these states is expressed as follows: x'i=xi+1 if Xi>xi and x'i=xi−1 if Xi<x'i. If no logical variable xi is updated then the focal state of S is equal to S and S is its own successor: it is said steady (or stationary). The focal state value Xi of gene I depends on the state S of thenetwork and in particular, on a set of conditions regarding the presence or absence of activators and inhibitors of gene I . For the simple example in Fig. 6, the focal state value Xa of gene adepends on the influence of B on a, that is, whether the concentration of B is below (xb=0) or above its first threshold value. Such interactions are expressed by means of products of step functions of the form:

$$s^{+}(x_j, \theta_j^r) = 1 \text{ if } x_j \geq r \text{ else } 0 \quad (14)$$
$$s^{-}(x_j, \theta_j^r) = 1 \text{ if } x_j < r \text{ else } 0$$

Discrete network models can help to examine how gene regulatory interactions generate the coherent, rule-like behavior of a cell – the first level of integration in the multi-scale complexity of the living organism. Hereby the various cell fates, such as differentiation, proliferation and apoptosis, are treated as attractor states of the network. This modeling language allows us to integrate qualitative gene and protein interaction data to explain a series of hitherto non-intuitive cell behaviors. Thus, discrete modeling offers a solution to sort the limitations of the current simplistic ‘one gene - one function - one target’ paradigm, allowing to develop of conceptual tools to increase our understanding of how the intricate interplay of genes gives rise to a global biological phenotype.

**Gene Regulatory Networ reduction**

The more complex a GRN becomes, the more difficult it is to design and analyze the behavior of the system. This is mainly because of important physiological factors such as translation processes, signaling cascades, time delays,..., etc, that play important roles in the dynamics of the system. However, these factors cannot be ignored in order

to design a model providing correct predictions [189,190]. In order to reduce the dimensionality and complexity of the GRN conserving the equilibrium and stabilities of the initial network we performed a reduction transformation that permitted to simplify the analysis [181]. By using this approach we can eliminate the nodes without self-feedback loop one by one until the interaction graph becomes irreducible.

Sign of a path: In an interaction graph, given value 1 to the $m^{th}$ interaction if it is positive or -1 if it is negative and having a path of length l connecting two nodes, we can determine the sign of a specific path by using the following expression:

$$\prod_{m=1}^{l-1} s_{p_m+1 p_m} = 1 \text{ (or } -1) \quad (6)$$

Thus, if the sign of the path is positive the affector node is affecting positively the target node and negatively if the sign is negative.

Irreducibility: An interaction graph is irreducible only when there is at least one path from the j-th node to the i-th node for all i,j $\in$ N ($i \neq j$) where N is the total amount of components in the transition graph, $p_1,...p_l \in N$ and $e_{p_a p_b}$ is an edge from node $p_a$ to node $p_b$. In other words, an irreducible interaction graph is that in which exist a path connecting each node with all the others nodes and there is no nodes with only one input and one output (self-feedback loops are not included).

Feedback loop: A feedback loop is defined as a path from the i-th node of a network to the same i-th node, if it consist of a single edge then it is a self-feedback loop.

Reduction method: The reduction procedure is interpreted by the following operations on the interaction graph. As is described in the figure 6A, changes on node A can be propagated through node B affecting node C, thus the information contained in the node B as well as their inputs and outputs can be integrated into a direct interaction from A to C Having an affector node acting on a target node, the first step in this process is remove the target node. Next, we create new edges from the affector node to all the nodes from which an edge went out from the target node, the sign of this new interaction is determined by the equation (6). This process is repeated until the network is irreducible. This method can be done on feedback loops or linear path as is shown in

the figures 6A and 6B.

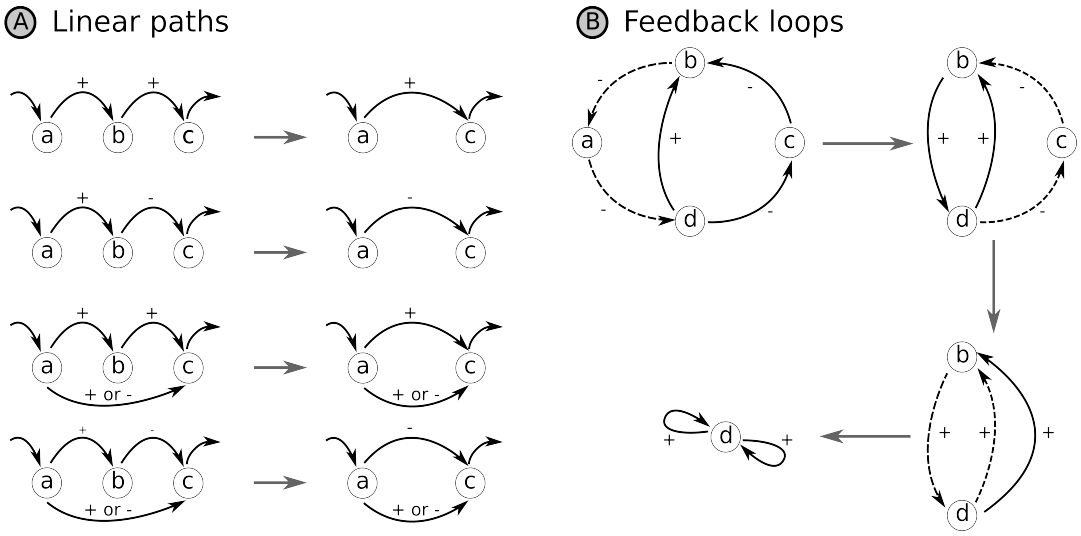



Fig. 6. Methods to reduce interaction graph. The nodes represent genes or proteins and the arrows represent the interactions between them. Each interaction can be either an inhibition (-) or an activation (+) of the target node. A: This graph shows the process to reduce the dimensionality of four possible linear path. Based on the equation 6 if a activates b and b activates c, the effect of a on c is also an activation as shown in the first and thirds interaction graphs, however, if one of the interactions is negative the effect of a on c is an inhibition as is shown in the second and fourth interaction graphs. This process is independent of possible direct effects of a on the target node c as is represented in the thirds and fourths interactions graphs. B: This graph represents an example of feedback loop and the process to reduce its size. This process starts by removing the node a and merging the edge b-a and a-d based on equation 6 obtaining consequently b activates d. The same process is repeated removing a node at each step until the network is irreducible.

Thus, according to this methodology all the nodes without any self-feedback loop and/or experimental data available can be eliminated one by one reducing the dimensionality of the system but conserving the equilibrium and stabilities of the original system.

**Constraint integration into the model**

The integration of constraints was held in two ways depending on the source of the

data.

Based on levels of significance: By analyzing the microarray data from skeletal muscle tissue biopsies, as is previously described , we could determine which genes were differentially expressed between the 6 different states studied in this paper by doing 9 different comparisons.

This information was integrated into the discrete model in the form of inequalities. Then, for example, if it was determined that the expression of gene “a” is significantly higher in “condition 1” than in “condition 2” ($\log_2 FC > 0$ and $p < 0.01$), the following constraint is introduced into the model: $Ga_{con1} \geq Ga_{con2}$, where $Ga_{con1}$ is the discrete value of gene a in the first condition and $Ga_{con2}$ is the discrete value of the same gene in the second condition. In this manner we defined correlations in the discrete values of the nodes between the six different states.

Based on correlation analysis: Additional constraints from public microarray gene expression data set were integrated into the discrete model. We performed a correlation analysis that was lately integrated as constraints in the discrete model in the form of inequalities. This analysis is based on the rationale that the discrete values of two genes that are significantly correlated cannot evolve in opposite directions (one increasing its expression and the other decreasing its expression), analogously two genes significantly anti-correlated cannot evolve in the same direction.

Then, taking for example the non-connected nodes A and B with 4 and 3 possible discrete values respectively it can exist an specific case as is represented in the table 2

| Genes\Conditions | Condition 1 | Condition 2 |
|---|---|---|
| A | 2 | 3 |
| B | 1 | BC2 |

Table 2.

If exist a significant correlation between gene A and B then there are only two possible

discrete values for $B_{C2}$, 1 or 2, this scenario excludes $B_{C2}$=0. However, if instead there is a significant anticorrelation between gene A and B, then $B_{C2}$ only can be 1 or 0 excluding $B_{C2}$=2.

In order to perform this analysis systematically, we have developed an algorithm that analyzes gene expression data sets performing a correlation analysis between the genes of different samples determining the correlations existing between them and their corresponding corrected significance. Next, those correlations with a level of significance above a certain threshold are converted into constraints based on the criteria previously described and integrated into the discrete model in the form of inequalities. This analysis is based on Pearson correlation test and the algorithm is written in R and Bash script (Supplementary material 5). This approach permits to integrate information from probabilistic analysis into a deterministic model in the form of constraint.

**Authors' contributions**

IMM performed all the computational analysis , wrote the draft, edited versions of the manuscript and final versions, generated the illustrations and participated in the design of analysis; EF coordinated all the computational analysis , participated versions of the manuscript and final versions, generated the illustrations and participated in the design of specific experiments, participated in the design of analysis ; VAS participated in the design of specific experiments, participated in the design of analysis, participated in edited versions of the manuscript and final versions ; JR designed and supervised the overall study and specific experiments including data analysis and wrote and approved manuscript drafts and final versions; MC designed and supervised the overall study and specific experiments including data analysis and wrote and approved manuscript drafts and final versions.

**Acknowledgements**

This work was supported by the European Commission Seventh Framework

Programme FP7 (Synergy-COPD project grant agreement n° 270086); the Spanish Government and the European Union FEDER funds (SAF2011-25726); Generalitat de Catalunya-AGAUR
IMM acknowledges the support received by all the members of the TIMC-IMAG BCM group.

**References**

1. Long TA, Brady SM, Benfey PN (2008) Systems approaches to identifying gene regulatory networks in plants. Annu Rev Cell Dev Biol 24: 81–10
2. Wellmer F, Riechmann JL (2010) Gene networks controlling the initiation of flower development. Trends Genet 26: 519–527
3. Barabasi, A.; Oltvai, Z. N. (2004). "Network biology: understanding the cells' functional organization". Nature Reviews Genetics 5 (2): 101–113.
4. Tian C, Zhang X, He J, Yu H, Wang Y, Shi B, Han Y, Wang G, Feng X, Zhang C, Wang J, Qi J, Yu R, Jiao Y. An organ boundary-enriched gene regulatory network uncovers regulatory hierarchies underlying axillary meristem initiation. Mol Syst Biol. 2014 Oct 30;10:755.
5. oward a road map for global -omics: a primer on -omic technologies. Coughlin SS. Am J Epidemiol. 2014 Dec 15;180(12):1188-95.
6. Synergy-COPD: a systems approach for understanding and managing chronic diseases.Gomez-Cabrero D, Lluch-Ariet M, Tegnér J, Cascante M, Miralles F, Roca J; Synergy-COPD consortium. J Transl Med. 2014 Nov 28;12 Suppl 2:S2.
7. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med. 2013 Feb 15;187(4):347-65. doi: 10.1164/rccm.201204-0596PP

8. Murray CJ1, Lopez AD. Measuring the global burden of disease. N Engl J Med. 2013 Aug 1;369(5):448-57 doi: 10.1056/NEJMra1201534.
9. Maltais F, Decramer M, Casaburi R, Barreiro E, et al. An official American Thoracic Society/European Respiratory Society statement: update on limb muscle dysfunction in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2014;189:e15-62.
10. Mitochondrial dysfunction in COPD patients with low body mass index. Rabinovich RA, Bastos R, Ardite E, Llinàs L, Orozco-Levi M, Gea J, Vilaró J, Barberà JA, Rodríguez-Roisin R, Fernández-Checa JC, Roca J. Eur Respir J. 2007 Apr;29(4):643-50.
11. Am J Respir Crit Care Med. 2001 Oct 1;164(7):1114-8. Reduced muscle redox capacity after endurance training in patients with chronic obstructive pulmonary disease. Rabinovich RA, Ardite E, Troosters T, Carbó N, Alonso J, Gonzalez de Suso JM, Vilaró J, Barberà JA, Polo MF, Argilés JM, Fernandez-Checa JC, Roca J
12. Oxygen pathway modeling estimates high reactive oxygen species production above the highest permanent human habitation. Cano I, Selivanov V, Gomez-Cabrero D, Tegnér J, Roca J, Wagner PD, Cascante M. PLoS One. 2014 Nov 6;9(11):e111068.
13. SelivanovVA, Votyakova TV, Zeak JA, Trucco M, Roca J, Cascante M.: Bistability of mitochondrial respiration underlies paradoxical reactive oxygen species generation induced by anoxia. PLoS Comput Biol. 2009;5:e1000619.
14. Morrow JD, Qiu W, Chhabra D, Rennard SI, Belloni P, Belousov A, Pillai SG, Hersh CP., Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. BMC Med Genomics. 2015 Jan 13;8(1):1.
15. Zhang WJ, Hubbard Cristinacce PL, Bondesson E, Nordenmark LH, Young SS, Liu YZ, Singh D, Naish JH, Parker GJ. MR Quantitative Equilibrium Signal Mapping: A Reliable Alternative to CT in the Assessment of Emphysema in Patients with Chronic Obstructive Pulmonary Disease.

Radiology. 2015 Jan 7:132953.

16. Wright, S. (1921). "Correlation and Causation" (PDF). Journal of Agricultural Research 20 (7): 557–585
17. Fisher, J., Henzinger, T., 2007. Executable cell biology. Nat. Biotechnol. 25 (11),1239–1249.; de Jong, H., Ropers, D., 2006. Qualitative approaches to the analysis of genetic regulatory networks. In: System Modeling in Cellular Biology: From Concepts to Nuts and Bolts. MIT Press, Cambridge, MA, pp. 125–148
18. Garcia-Aymerich J, Gómez FP, Benet M, Farrero E, Basagaña X, Gayete À, Paré C, Freixa X, Ferrer J, Ferrer A, Roca J, Gáldiz JB, Sauleda J, Monsó E, Gea J, Barberà JA, Agustí À, Antó JM; PAC-COPDStudy Group.: Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. Thorax. 2011 May;66(5):430-7.
19. Sala E, RocaJ, Marrades RM, Alonso J, Gonzalez De Suso JM, Moreno A, Barberá JA, Nadal J, de Jover L, Rodriguez-Roisin R, Wagner PD.: Effects of endurance training on skeletal muscle bioenergetics in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1999;159:1726-34.
20. Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014 Feb 15;30(4):523-30.
21. Murali T, Pacifico S, Yu J, Guest S, Roberts GG 3rd, Finley RL Jr. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. Nucleic Acids Res. 2011 Jan;39(Database issue):D736-43.
22. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, Turan N, Cascante M, Falciani F, Hernandez M, Villà-Freixa J, Losko S. Knowledge management for systems biology a general and visually driven framework applied to translational medicine. BMC Syst Biol. 2011 Mar 5;5:38.
23. Corblin F, Fanchon E, Trilling L. Applications of a formal approach to decipher discrete genetic networks. BMC Bioinformatics. 2010 Jul 20;11:385.

24. Fauré A, Vreede BM, Sucena E, Chaouiya C. A discrete model of Drosophila eggshell patterning reveals cell-autonomous and juxtacrine effects. PLoS Comput Biol. 2014 Mar 27;10(3):e1003527
25. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013 Jan;41(Database issue):D991-5.
26. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP.: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4:249-64.
27. Breitling R, Armengaud P, Amtmann A, Herzyk P.: Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 2004 Aug 27;573(1-3):83-92.
28. Berardini TZ, Li D, Huala E, Bridges S, Burgess S, McCarthy F, Carbon S, Lewis SE, Mungall CJ, Abdulla A, Wood V, Feltrin E, Valle G, Chisholm RL, Fey P, Gaudet P, Kibbe W, Basu S, Bushmanova Y, Eilbeck K, Siegele DA, McIntosh B, Renfro D, Zweifel A, Hu JC, Ashburner M, Tweedie S, Alam-Faruque Y, Apweiler R, Auchinchloss A, Bairoch A, Barrell D, Binns D, Blatter MC, Bougueleret L, Boutet E, Breuza L, Bridge A, Browne P, Chan WM, Coudert E, Daugherty L, et al. The Gene Ontology in 2010: extensions and refinements. Nucleic Acids Res. 2010 Jan;38(Database issue):D331-5.
29. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, Wernerus H, Björling L, Ponten F. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010 Dec;28(12):1248-50.
30. Thomas R, Kaufman M. Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. Chaos 2001, 11 :180-195.
31. Puente-Maestu L, Tejedor A, Lázaro A, de Miguel J, Alvarez-Sala L,

González- Aragoneses F, Simón C, Agustí A.: Site of Mitochondrial ROS Production in Skeletal Muscle of COPD and its Relationship with Exercise Oxidative Stress. Am J Respir Cell Mol Biol. 2012

32. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013 Jan;41(Database issue):D808-15.
33. Apweiler R, Bateman A, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, Alpi E, Antunes R, Arganiska J, Barrera Casanova et al. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014 Jan;42(Database issue):D191-8.
34. Prasanthi JR, Larson T, Schommer J, Ghribi O.: Silencing GADD153/CHOP gene expression protects against Alzheimer's disease-like pathology induced by 27-hydroxycholesterol in rabbit hippocampus. PLoS One. 2011;6(10):e26420.
35. Havens CG, Ho A, Yoshioka N, Dowdy SF.: Regulation of late G1/S phase transition and APC Cdh1 by reactive oxygen species. Mol Cell Biol. 2006;26:4701-11.
36. Macip S, Igarashi M, Fang L, Chen A, Pan ZQ, Lee SW, Aaronson SA.: Inhibition of p21-mediated ROS accumulation can rescue p21-induced senescence. EMBO J. 2002;21:2180-8.
37. Riemann A, Schneider B, Ihling A, Nowak M, Sauvant C, Thews O, Gekle M.: Acidic environment leads to ROS-induced MAPK signaling in cancer cells. PLoS One. 2011;6(7):e22445. Epub 2011 Jul 26.
38. Ki YW, Lee JE, Park JH, Shin IC, Koh HC.: Reactive oxygen species and mitogen-activated protein kinase induce apoptotic death of SH-SY5Y cells in response to fipronil. Toxicol Lett. 2012;211:18-28.
39. Schuchardt M, Prüfer J, Prüfer N, Wiedon A, Huang T, Chebli M, Jankowski V, Jankowski J, Schäfer-Korting M, Zidek W, van der Giet M, Tölle M.: The endothelium-derived contracting factor uridine adenosine tetraphosphate

induces P2Y(2)-mediated pro-inflammatory signaling by monocyte chemoattractant protein-1 formation. J Mol Med (Berl). 2011;89:799-810.

40. D'Andrea P, Romanello M, Bicego M, Steinberg TH, Tell G.: H(2)O(2) modulates purinergic-dependent calcium signalling in osteoblast-like cells. Cell Calcium. 2008;43:457-68.
41. Loor G, Kondapalli J, Schriewer JM, Chandel NS, Vanden Hoek TL, Schumacker PT.: Menadione triggers cell death through ROS-dependent mechanisms involving PARP activation without requiring apoptosis. Free Radic Biol Med. 2010;49:1925-36.
42. Lee YH, Lee NH, Bhattarai G, Yun JS, Kim TI, Jhee EC, Yi HK.: PPARgamma inhibits inflammatory reaction in oxidative stress induced human diploid fibloblast. Cell Biochem Funct. 2010;28:490-6.
43. Luciani A, Villella VR, Vasaturo A, Giardino I, Pettoello-Mantovani M, Guido S, Cexus ON, Peake N, Londei M, Quaratino S, Maiuri L.: Lysosomal accumulation of gliadin p31-43 peptide induces oxidative stress and tissue transglutaminase-mediated PPARgamma downregulation in intestinal epithelial cells and coeliac mucosa. Gut. 2010 Mar;59:311-9.
44. Costa AD, Garlid KD.: Intramitochondrial signaling: Interactions among mitoKATP, PKCepsilon, ROS, and MPT. Am J Physiol Heart Circ Physiol. 2008;295:H874-82.
45. Jung YS, Ryu BR, Lee BK, Mook-Jung I, Kim SU, Lee SH, Baik EJ, Moon CH.: Role for PKC-epsilon in neuronal death induced by oxidative stress. Biochem Biophys Res Commun. 2004;320:789-94.
46. Sarkey JP, Chu M, McShane M, Bovo E, Mou YA, Zima AV, de Tombe PP, Kartje GL, Martin JL.: Nogo-A knockdown inhibits hypoxia/reoxygenation-induced activation of mitochondrial-dependent apoptosis in cardiomyocytes. J Mol Cell Cardiol. 2011;50:1044-55.
47. Mi YJ, Hou B, Liao QM, Ma Y, Luo Q, Dai YK, Ju G, Jin WL.: Amino-Nogo-A antagonizes reactive oxygen species generation and protects immature primary cortical neurons from oxidative toxicity. Cell Death Differ.

2012;19:1175-86.

48. Mattiussi M, Tilman G, Lenglez S, Decottignies A.: Human telomerase represses ROS-dependent cellular responses to Tumor Necrosis Factor-α without affecting NF-κB activation. Cell Signal. 2012;24:708-17.
49. Tsutsumi K, Fujikawa H, Kajikawa T, Takedachi M, Yamamoto T, Murakami S.: Effects of L-ascorbic acid 2-phosphate magnesium salt on the properties of human gingival fibroblasts. J Periodontal Res. 2012;47:263-7
50. Ahmad I, Khan MI, Patil G, Chauhan LK.: Evaluation of cytotoxic, genotoxic and inflammatory responses of micro- and nano-particles of granite on human lung fibroblast cell IMR-90. Toxicol Lett. 2012 Feb 5;208:300-7.
51. Bensaad K, Cheung EC, Vousden KH.: Modulation of intracellular ROS levels by TIGAR controls autophagy. EMBO J. 2009;28:3015-26.
52. Liu B, Chen Y, St Clair DK.: ROS and p53: A versatile partnership. Free Radic Biol Med. 2008;44:1529-35.
53. Levine AJ.: 27-Hydroxycholesterol in Rabbit Hippocampus. Annu Rev Biochem. 1993;62:623-51.
54. Reittie JE, Yong KL, Panayiotidis P, Hoffbrand AV.: Interleukin-6 inhibits apoptosis and tumour necrosis factor induced proliferation of B-chronic lymphocytic leukaemia. Leuk Lymphoma. 1996;22:83-90
55. Jagadish MN, Fernandez CS, Hewish DR, Macaulay SL, Gough KH, Grusovin J, Verkuylen A, Cosgrove L, Alafaci A, Frenkel MJ, Ward CW.: Insulin-responsive tissues contain the core complex protein SNAP-25 (synaptosomal-associated protein 25) A and B isoforms in addition to syntaxin 4 and synaptobrevins 1 and 2. Biochem J. 1996;317:945-54.
56. Lu Z, Liu D, Hornia A, Devonish W, Pagano M, Foster DA.: Activation of protein kinase C triggers its ubiquitination and degradation. Mol Cell Biol. 1998;18:839-45.
57. Kinscherf R, Claus R, Wagner M, Gehrke C, Kamencic H, Hou D, Nauen O, Schmiedt W, Kovacs G, Pill J, Metz J, Deigner HP.: Apoptosis caused by oxidized LDL is manganese superoxide dismutase and p53 dependent. FASEB

J. 1998;12:461-7.

58. Souri M, Aoyama T, Hoganson G, Hashimoto T.: Very-long-chain acyl-CoA dehydrogenase subunit assembles to the dimer form on mitochondrial inner membrane. FEBS Lett. 1998;426:187-90.
59. Foster LJ, Yeung B, Mohtashami M, Ross K, Trimble WS, Klip A.: Binary interactions of the SNARE proteins syntaxin-4, SNAP23, and VAMP-2 and their regulation by phosphorylation. Biochemistry. 1998;37:11089-96.
60. Herdegen T, Leah JD.: Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. Brain Res Brain Res Rev.;28:370-490.
61. Scheid MP, Foltz IN, Young PR, Schrader JW, Duronio V.: Ceramide and cyclic adenosine monophosphate (cAMP) induce cAMP response element binding protein phosphorylation via distinct signaling pathways while having opposite effects on myeloid cell survival. Blood. 1999;93:217-25.
62. Cohen SB, Zheng G, Heyman HC, Stavnezer E.: Heterodimers of the SnoN and Ski oncoproteins form preferentially over homodimers and are more potent transforming agents. Nucleic Acids Res. 1999;27:1006-14.
63. Taylor CT, Fueki N, Agah A, Hershberg RM, Colgan SP.: Critical role of cAMP response element binding protein expression in hypoxia-elicited induction of epithelial tumor necrosis factor-alpha. J Biol Chem. 1999;274:19447-54.
64. Simbulan-Rosenthal CM, Haddad BR, Rosenthal DS, Weaver Z, Coleman A, Luo R, Young HM, Wang ZQ, Ried T, Smulson ME.: Chromosomal aberrations in PARP(-/-) mice: genome stabilization in immortalized cells by reintroduction of poly(ADP-ribose) polymerase cDNA. Proc Natl Acad Sci U S A. 1999;96:13191-6.
65. Simbulan-Rosenthal CM, Haddad BR, Rosenthal DS, Weaver Z, Coleman A, Luo R, Young HM, Wang ZQ, Ried T, Smulson ME.: Chromosomal aberrations in PARP(-/-) mice: genome stabilization in immortalized cells by reintroduction of poly(ADP-ribose) polymerase cDNA. Proc Natl Acad Sci U

S A. 1999;96:13191-6.

66. De Nadai C, Sestili P, CantoniO, Lièvremont JP, Sciorati C, Barsacchi R, Moncada S, Meldolesi J, Clementi E.: Nitric oxide inhibits tumor necrosis factor-alpha-induced apoptosis by reducing the generation of ceramide. Proc Natl Acad Sci U S A. 2000;97:5480-5.
67. Giebler HA, Lemasson I, Nyborg JK.: p53 recruitment of CREB binding protein mediated through phosphorylated CREB: a novel pathway of tumor suppressor regulation. Mol Cell Biol. 2000;20:4849-58.
68. Chung SH, Polgar J, Reed GL.: Protein kinase C phosphorylation of syntaxin 4 in thrombin-activated human platelets. J Biol Chem. 2000;275:25286-91.
69. Wu G, Toyokawa T, Hahn H, Dorn GW 2nd.: Epsilon protein kinase C in pathological myocardial hypertrophy. Analysis by combined transgenic expression of translocation modifiers and Galphaq. J Biol Chem. 2000;275:29927-30.
70. Way JM, Harrington WW, Brown KK, Gottschalk WK, Sundseth SS, Mansfield TA, Ramachandran RK, Willson TM, Kliewer SA.: Comprehensive messenger ribonucleic acid profiling reveals that peroxisome proliferator-activated receptor gamma activation has coordinate effects on gene expression in multiple insulin-sensitive tissues. Endocrinology. 2001;142:1269-77.
71. Kyriakis JM, Avruch J.: Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. Physiol Rev. 2001;81:807-69.
72. Tong L, Thornton PL, Balazs R, Cotman CW.: Beta -amyloid-(1-42) impairs activity-dependent cAMP-response element-binding protein signaling in neurons at concentrations in which cell survival Is not compromised. J Biol Chem. 2001;276:17301-6.
73. Strosznajder JB, Jeśko H, Strosznajder RP.: Effect of amyloid beta peptide on poly(ADP-ribose) polymerase activity in adult and aged rat hippocampus. Acta Biochim Pol.;47:847-54.
74. Drane P, Bravard A, Bouvard V, May E.: Reciprocal down-regulation of p53

and SOD2 gene expression-implication in p53 mediated apoptosis. Oncogene. 2001;20:430-9.

75. Klemm DJ, Leitner JW, Watson P, Nesterova A, Reusch JE, Goalstone ML, Draznin B.: Insulin-induced adipocyte differentiation. Activation of CREB rescues adipogenesis from the arrest caused by inhibition of prenylation. J Biol Chem. 2001;276:28430-5.
76. Bonni S, Wang HR, Causing CG, Kavsak P, Stroschein SL, Luo K, Wrana JL.: TGF-beta induces assembly of a Smad2-Smurf2 ubiquitin ligase complex that targets SnoN for degradation. Nat Cell Biol. 2001;3:587-95.
77. Lai Z, Ferry KV, Diamond MA, Wee KE, Kim YB, Ma J, Yang T, Benfield PA, Copeland RA, Auger KR.: Human mdm2 mediates multiple mono-ubiquitination of p53 by a mechanism requiring enzyme isomerization. J Biol Chem. 2001;276:31357-67.
78. Kolobova E, Tuganova A, Boulatnikov I, Popov KM.: Regulation of pyruvate dehydrogenase activity through phosphorylation at multiple sites. Biochem J. 2001;358:69-77.
79. Fulda S, Küfer MU, Meyer E, van Valen F, Dockhorn-Dworniczak B, Debatin KM.: Sensitization for death receptor- or drug-induced apoptosis by re-expression of caspase-8 through demethylation or gene transfer. Oncogene. 2001;20:5865-77.
80. Soldatenkov VA, Chasovskikh S, Potaman VN, Trofimova I, Smulson ME, Dritschilo A.: Transcriptional repression by binding of poly(ADP-ribose) polymerase to promoter sequences. J Biol Chem. 2002;277:665-70.
81. Trockenbacher A, Suckow V, Foerster J, Winter J, Krauss S, Ropers HH, Schneider R, Schweiger S.: MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. Nat Genet. 2001;29:287-94.
82. Li H, Degenhardt B, Tobin D, Yao ZX, Tasken K, Papadopoulos V.: Identification, localization, and function in steroidogenesis of PAP7: a peripheral-type benzodiazepine receptor- and PKA (RIalpha)-associated

protein. Mol Endocrinol. 2001;15:2211-28.

83. Casley CS, Canevari L, Land JM, Clark JB, Sharpe MA.: Beta-amyloid inhibits integrated mitochondrial respiration and key enzyme activities. J Neurochem. 2002;80:91-100.
84. Gannon-Murakami L, Murakami K.: Selective association of protein kinase C with 14-3-3 zeta in neuronally differentiated PC12 Cells. Stimulatory and inhibitory effect of 14-3-3 zeta in vivo. J Biol Chem. 2002;277:23116-22.
85. Valentini G, Chiarelli LR, Fortin R, Dolzan M, Galizzi A, Abraham DJ, Wang C, Bianchi P, Zanella A, Mattevi A.: Structure and function of human erythrocyte pyruvate kinase. Molecular basis of nonspherocytic hemolytic anemia. J Biol Chem. 2002;277:23807-14.
86. Gadéa G, Lapasset L, Gauthier-Rouvière C, Roux P.: Regulation of Cdc42-mediated morphological effects: a novel function for p53. EMBO J. 2002;21:2373-82.
87. Hu WH, Hausmann ON, Yan MS, Walters WM, Wong PK, Bethea JR.: Identification and characterization of a novel Nogo-interacting mitochondrial protein (NIMP). J Neurochem. 2002;81:36-45.
88. Ding L, Wang H, Lang W, Xiao L.: Protein kinase C-epsilon promotes survival of lung cancer cells by suppressing apoptosis through dysregulation of the mitochondrial caspase pathway. J Biol Chem. 2002;277:35305-13.
89. Basu A, Lu D, Sun B, Moor AN, Akkaraju GR, Huang J.: Proteolytic activation of protein kinase C-epsilon by caspase-mediated processing and transduction of antiapoptotic signals. J Biol Chem. 2002;277:41850-6.
90. Zhang H, Shi X, Zhang QJ, Hampong M, Paddon H, Wahyuningsih D, Pelech S.: Nocodazole-induced p53-dependent c-Jun N-terminal kinase activation reduces apoptosis in human colon carcinoma HCT116 cells. J Biol Chem. 2002;277:43648-58.
91. Yu S, Matsusue K, Kashireddy P, Cao WQ, Yeldandi V, Yeldandi AV, Rao MS, Gonzalez FJ, Reddy JK.: Adipocyte-specific gene expression and adipogenic steatosis in the mouse liver due to peroxisome proliferator-activated receptor

gamma1 (PPARgamma1) overexpression. J Biol Chem. 2003;278:498-505.

92. Le DA, Wu Y, Huang Z, Matsushita K, Plesnila N, Augustinack JC, Hyman BT, Yuan J, Kuida K, Flavell RA, Moskowitz MA.: Caspase activation and neuroprotection in caspase-3- deficient mice after in vivo cerebral ischemia and in vitro oxygen glucose deprivation. Proc Natl Acad Sci U S A. 2002;99:15188-93.
93. Mizuide M, Hara T, Furuya T, Takeda M, Kusanagi K, Inada Y, Mori M, Imamura T, Miyazawa K, Miyazono K.: Two short segments of Smad3 are important for specific interaction of Smad3 with c-Ski and SnoN. J Biol Chem. 2003;278:531-6.
94. Francis GA, Fayard E, Picard F, Auwerx J.: Nuclear receptors and the control of metabolism. Annu Rev Physiol. 2003;65:261-311.
95. Sato S, Tomomori-Sato C, Banks CA, Sorokina I, Parmely TJ, Kong SE, Jin J, Cai Y, Lane WS, Brower CS, Conaway RC, Conaway JW.: Identification of mammalian Mediator subunits with similarities to yeast Mediator subunits Srb5, Srb6, Med11, and Rox3. J Biol Chem. 2003;278:15123-7.
96. Suzawa M, Takada I, Yanagisawa J, Ohtake F, Ogawa S, Yamauchi T, Kadowaki T, Takeuchi Y, Shibuya H, Gotoh Y, Matsumoto K, Kato S.: Cytokines suppress adipogenesis and PPAR-gamma function through the TAK1/TAB1/NIK cascade. Nat Cell Biol. 2003;5:224-30.
97. Ivaska J, Bosca L, Parker PJ.: PKCepsilon is a permissive link in integrin-dependent IFN-gamma signalling that facilitates JAK phosphorylation of STAT1. Nat Cell Biol. 2003;5:363-9.
98. Kahns S, Kalai M, Jakobsen LD, Clark BF, Vandenabeele P, Jensen PH.: Caspase-1 and caspase-8 cleave and inactivate cellular parkin. J Biol Chem. 2003;278:23376-80.
99. Simonen M, Pedersen V, Weinmann O, Schnell L, Buss A, Ledermann B, Christ F, Sansig G, van der Putten H, Schwab ME.: Systemic deletion of the myelin-associated outgrowth inhibitor Nogo-A improves regenerative and plastic responses after spinal cord injury. Neuron. 2003;38(2):201-11.

100. Marques CA, Keil U, Bonert A, Steiner B, Haass C, Muller WE, Eckert A.: Neurotoxic mechanisms caused by the Alzheimer's disease-linked Swedish amyloid precursor protein mutation: oxidative stress, caspases, and the JNK pathway. J Biol Chem. 2003;278:28294-302.

101. Ruan H, Pownall HJ, Lodish HF.: Troglitazone antagonizes tumor necrosis factor-alpha-induced reprogramming of adipocyte gene expression by inhibiting the transcriptional regulatory functions of NF-kappaB. J Biol Chem. 2003;278:28181-92.

102. Yahagi N, Shimano H, Matsuzaka T, Najima Y, Sekiya M, Nakagawa Y, Ide T, Tomita S, Okazaki H, Tamura Y, Iizuka Y, Ohashi K, Gotoda T, Nagai R, Kimura S, Ishibashi S, Osuga J, Yamada N.: p53 Activation in adipocytes of obese mice. J Biol Chem. 2003;278:25395-400.

103. Giri RK, Selvaraj SK, Kalra VK.: Amyloid peptide-induced cytokine and chemokine expression in THP-1 monocytes is blocked by small inhibitory RNA duplexes for early growth response-1 messenger RNA. J Immunol. 2003;170:5281-94.

104. Kim YK, Han JW, Woo YN, Chun JK, Yoo JY, Cho EJ, Hong S, Lee HY, Lee YW, Lee HW.: Expression of p21(WAF1/Cip1) through Sp1 sites by histone deacetylase inhibitor apicidin requires PI 3-kinase-PKC epsilon signaling pathway. Oncogene. 2003;22:6023-31.

105. Wu JM, Xiao L, Cheng XK, Cui LX, Wu NH, Shen YF.: PKC epsilon is a unique regulator for hsp90 beta gene in heat shock response. J Biol Chem. 2003;278:51143-9.

106. Li J, Hawkins IC, Harvey CD, Jennings JL, Link AJ, Patton JG.: Regulation of alternative splicing by SRrp86 and its interacting proteins. Mol Cell Biol. 2003;23:7437-47.

107. Rego AC, Oliveira CR.: Mitochondrial dysfunction and reactive oxygen species in excitotoxicity and apoptosis: implications for the pathogenesis of neurodegenerative diseases. Neurochem Res. 2003;28:1563-74.

108. Koutnikova H, Cock TA, Watanabe M, Houten SM, Champy MF, Dierich A, Auwerx J.: Compensation by the muscle limits the metabolic consequences of lipodystrophy in PPAR gamma hypomorphic mice. Proc Natl Acad Sci U S A. 2003;100:14457-62.

109. Wilms H, Claasen J, Röhl C, Sievers J, Deuschl G, Lucius R.: Involvement of benzodiazepine receptors in neuroinflammatory and neurodegenerative diseases: evidence from activated microglial cells in vitro. Neurobiol Dis. 2003;14:417-24.

110. Zhang HG, Hyde K, Page GP, Brand JP, Zhou J, Yu S, Allison DB, Hsu HC, Mountz JD.: Novel tumor necrosis factor alpha-regulated genes in rheumatoid arthritis. Arthritis Rheum. 2004Eb;50:420-31.

111. Mandel S, Weinreb O, Amit T, Youdim MB.: Cell signaling pathways in the neuroprotective actions of the green tea polyphenol (-)-epigallocatechin-3-gallate: implications for neurodegenerative diseases. J Neurochem. 2004;88:1555-69.

112. Ha HC.: Defective transcription factor activation for proinflammatory gene expression in poly(ADP-ribose) polymerase 1-deficient glia. Proc Natl Acad Sci U S A. 2004;101:5087-92.

113. Hussain SP, Amstad P, He P, Robles A, Lupold S, Kaneko I, Ichimiya M, Sengupta S, Mechanic L, Okamura S, Hofseth LJ, Moake M, Nagashima M, Forrester KS, Harris CC.: p53-induced up-regulation of MnSOD and GPx but not catalase increases oxidative stress and apoptosis. Cancer Res. 2004;64:2350-6.

114. Gregory Powell J, Wang X, Allard BL, Sahin M, Wang XL, Hay ID, Hiddinga HJ, Deshpande SS, Kroll TG, Grebe SK, Eberhardt NL, McIver B.: The PAX8/PPARgamma fusion oncoprotein transforms immortalized human thyrocytes through a mechanism probably involving wild-type PPARgamma inhibition. Oncogene. 2004;23:3634-41.

115. Banno T, Gazel A, Blumenberg M.: Effects of tumor necrosis factor-alpha (TNF alpha) in epidermal keratinocytes revealed using global

transcriptional profiling. J Biol Chem. 2004;279:32633-42.

116. Komuro A, Imamura T, Saitoh M, Yoshida Y, Yamori T, Miyazono K, Miyazawa K.: Negative regulation of transforming growth factor-beta (TGF-beta) signaling by WW domain-containing protein 1 (WWP1). Oncogene. 2004;23:6914-23.

117. Fromenty B, Robin MA, Igoudjil A, Mansouri A, Pessayre D.: The ins and outs of mitochondrial dysfunction in NASH. Diabetes Metab. 2004;30:121-38.

118. Colland F, Jacq X, Trouplin V, Mougin C, Groizeleau C, Hamburger A, Meil A, Wojcik J, Legrain P, Gauthier JM.: Functional proteomics mapping of a human signaling pathway. Genome. 2 Jul;14(7):1324-32.

119. Patsouris D, Mandard S, Voshol PJ, Escher P, Tan NS, Havekes LM, Koenig W, März W, Tafuri S, Wahli W, Müller M, Kersten S.: PPARalpha governs glycerol metabolism. J Clin Invest. 2004;114:94-103.

120. Zhang Y, Wang JS, Chen LL, Zhang Y, Cheng XK, Heng FY, Wu NH, Shen YF.: Repression of hsp90beta gene by p53 in UV irradiation-induced apoptosis of Jurkat cells. J Biol 200t 8;279(41):42545-51.

121. He W, Lu Y, Qahwash I, Hu XY, Chang A, Yan R.: Reticulon family members modulate BACE1 activity and amyloid-beta peptide generation. Nat Me004p;10(9):959-65.

122. Scaife S, Brown R, Kellie S, Filer A, Martin S, Thomas AM, Bradfield PF, Amft N, Salmon M, Buckley CD.: Detection of differentially expressed genes in synovial fibroblasts by restriction fragment differential display. Rheumatology. 2004;43:1346-52.

123. Stein TD, Anders NJ, DeCarli C, Chan SL, Mattson MP, Johnson JA.: Neutralization of transthyretin reverses the neuroprotective effects of secreted amyloid precursor protein (APP) in APPSW mice resulting in tau phosphorylation and loss of hippocampal neurons: support for the amyloid hypothesis. J Neur200p 1;24(35):7707-17.

124. Zheng J, Devalaraja-Narashimha K, Singaravelu K, Padanilam BJ.:

Poly(ADP-ribose) polymerase-1 gene ablation protects mice from ischemic renal injury. Am J Physiol Renal Physiol. 2005;288:F387-98.

125. Kuratomi G, Komuro A, Goto K, Shinozaki M, Miyazawa K, Miyazono K, Imamura T.: NEDD4-2 (neural precursor cell expressed, developmentally down-regulated 4-2) negatively regulates TGF-beta (transforming growth factor-beta) signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. Biochem J. 2005;386:461-70.

126. Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV, Sun Y, Koo S, Perera RJ, Jain R, Dean NM, Freier SM, Bennett CF, Lollo B, Griffey R.: MicroRNA-143 regulates adipocyte differentiation. J Biol Chem. 2004;279:52361-5.

127. Wheeler DL, Martin KE, Ness KJ, Li Y, Dreckschmidt NE, Wartman M, Ananthaswamy HN, Mitchell DL, Verma AK.: Protein kinase C epsilon is an endogenous photosensitizer that enhances ultraviolet radiation-induced cutaneous damage and development of squamous cell carcinomas. Cancer200v 1;64(21):7756-65.

128. Izzotti A, Cartiglia C, Longobardi M, Bagnasco M, Merello A, You M, Lubet RA, De Flora S.: Gene expression in the lung of p53 mutant mice exposed to cigarette smoke. Cancer Res. 2004;64:8566-72.

129. Camden JM, Schrader AM, Camden RE, González FA, Erb L, Seye CI, Weisman GA.: P2Y2 nucleotide receptors enhance alpha-secretase-dependent amyloid precursor protein processing. J Biol Chem. 2005;280:18696-702.

130. Paris D, Ait-Ghezala G, Mathura VS, Patel N, Quadros A, Laporte V, Mullan M.: Anti-angiogenic activity of the mutant Dutch A(beta) peptide on human brain microvascular endothelial cells. Brain Res Mol Brain Res. 2005;136:212-30.

131. d'Abramo C, Massone S, Zingg JM, Pizzuti A, Marambaud P, Dalla Piccola B, Azzi A, Marinari UM, Pronzato MA, Ricciarelli R.: Role of peroxisome proliferator-activated receptor gamma in amyloid precursor

protein processing and amyloid beta-mediated cell death. Biochem J. 2005;391:693-8.

132. Youdim MB, Maruyama W, Naoi M.: Neuropharmacological, neuroprotective and amyloid precursor processing properties of selective MAO-B inhibitor antiparkinsonian drug, rasagiline. Drugs Today. 2005;41:369-91.

133. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M.: Towards a proteome-scale map of the human protein-protein interaction network. NatureOct (7062):1173-8.

134. Matsumoto M, Hatakeyama S, Oyamada K, Oda Y, Nishimura T, Nakayama KI.: Large-scale analysis of the human ubiquitin-related proteome. Proteomics. 2005;5:4145-51.

135. Tong XK, Nicolakakis N, Kocharyan A, Hamel E.: Vascular remodeling versus amyloid beta-induced oxidative stress in the cerebrovascular dysfunctions associated with Alzheimer's disease. J Neurosci. 2005;25:11165-74.

136. Vasilescu J, Smith JC, Ethier M, Figeys D.: Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. J Prot. 5 Nov-Dec;4(6):2192-200.

137. Lü X, de la Peña L, Barker C, Camphausen K, Tofilon PJ.: Radiation-induced changes in gene expression involve recruitment of existing messenger RNAs to and away from polysomes. Cancer006n 15;66(2):1052-61.

138. Hilbush BS, Morrison JH, Young WG, Sutcliffe JG, Bloom FE.: New prospects and strategies for drug target discovery in neurodegenerative

disorders. NeuroR00ct;2(4):627-37.

139. Hammond EM, Mandell DJ, Salim A, Krieg AJ, Johnson TM, Shirazi HA, Attardi LD, Giaccia AJ.: Genome-wide analysis of p53 under hypoxic conditions. Mol Cell Biol. 2006;26:3492-504.

140. Choi DS, Wang D, Yu GQ, Zhu G, Kharazia VN, Paredes JP, Chang WS, Deitchman JK, Mucke L, Messing RO.: PKCepsilon increases endothelin converting enzyme activity and reduces amyloid plaque pathology in transgenic mice. Proc Natl Acad Sci U S A. 2006;103:8215-20.

141. Dhar SK, Xu Y, Chen Y, St Clair DK.: Specificity protein 1-dependent p53-mediated suppression of human manganese superoxide dismutase gene expression. J Biol Chem. 2006;281:21698-709.

142. Alves da Costa C, Sunyach C, Pardossi-Piquard R, Sévalle J, Vincent B, Boyer N, Kawarai T, Girardot N, St George-Hyslop P, Checler F.: Presenilin-dependent gamma-secretase-mediated control of p53-associated cell death in Alzheimer's disease. J Neurosci. 2006;26:6377-85.

143. Sekiguchi K, Tian Q, Ishiyama M, Burchfield J, Gao F, Mann DL, Barger PM.: Inhibition of PPAR-alpha activity in mice with cardiac-restricted expression of tumor necrosis factor: potential role of TGF-beta/Smad3. Am J Physiol Heart Circ Physiol. 2007;292:1443-51.

144. Foo RS, Chan LK, Kitsis RN, Bennett MR.: Ubiquitination and degradation of the anti-apoptotic protein ARC by MDM2. J Biol Chem. 2007;282:5529-35.

145. Hurt EM, Thomas SB, Peng B, Farrar WL.: Integrated molecular profiling of SOD2 expression in multiple myeloma. Blood. 2007;109:3953-62.

146. Nagle CA, An J, Shiota M, Torres TP, Cline GW, Liu ZX, Wang S, Catlin RL, Shulman GI, Newgard CB, Coleman RA.: Hepatic overexpression of glycerol-sn-3-phosphate acyltransferase 1 in rats causes insulin resistance. J Biol Chem. 2007;282:14807-15.

147. Tsai CT, Wang DL, Chen WP, Hwang JJ, Hsieh CS, Hsu KL, Tseng CD, Lai LP, Tseng YZ, Chiang FT, Lin JL.: Angiotensin II increases expression

of alpha1C subunit of L-type calcium channel through a reactive oxygen species and cAMP response element-binding protein-dependent pathway in HL-1 myocytes. Circ Res. 2007;100:1476-85.

148. Saheki T, Iijima M, Li MX, Kobayashi K, Horiuchi M, Ushikai M, Okumura F, Meng XJ, Inoue I, Tajima A, Moriyama M, Eto K, Kadowaki T, Sinasac DS, Tsui LC, Tsuji M, Okano A, Kobayashi T.: Citrin/mitochondrial glycerol-3-phosphate dehydrogenase double knock-out mice recapitulate features of human citrin deficiency. J Biol Chem. 2007;282:25041-52.

149. Baranek T, Debret R, Antonicelli F, Lamkhioued B, Belaaouaj A, Hornebeck W, Bernard P, Guenounou M, Le Naour R.: Elastin receptor (spliced galactosidase) occupancy by elastin peptides counteracts proinflammatory cytokine expression in lipopolysaccharide-stimulated human monocytes through NF-kappaB down-regulation. J Immu007 1;179(9):6184-92

150. Wang T, Simbulan-Rosenthal CM, Smulson ME, Chock PB, Yang DC.: Polyubiquitylation of PARP-1 through ubiquitin K48 is modulated by activated DNA, NAD+, and dipeptides. J Cell Biochem. 2008;104:318-28.

151. Windheim M, Peggie M, Cohen P.: Two different classes of E2 ubiquitin-conjugating enzymes are required for the mono-ubiquitination of proteins and elongation by polyubiquitin chains with a specific topology. Biochem J. 2008;409:723-9.

152. Chen LH, Jiang CC, Watts R, Thorne RF, Kiejda KA, Zhang XD, Hersey P.: Inhibition of endoplasmic reticulum stress-induced apoptosis of melanoma cells by the ARC protein. Cancer Res. 2008;68:834-42.

153. Carlucci A, Adornetto A, Scorziello A, Viggiano D, Foca M, Cuomo O, Annunziato L, Gottesman M, Feliciello A.: Proteolysis of AKAP121 regulates mitochondrial activity during cellular hypoxia and brain ischaemia. EMBO J. 2008;27:1073-84.

154. Molenaar JJ, Ebus ME, Koster J, van Sluis P, van Noesel CJ, Versteeg R, Caron HN.: Cyclin D1 and CDK4 activity contribute to the undifferentiated

phenotype in neuroblastoma. Cancer008r 15;68(8):2599-609.

155. Walkey CJ, Spiegelman BM.: A functional peroxisome proliferator-activated receptor-gamma ligand-binding domain is not required for adipogenesis. J Biol 200p 5;283(36):24290-4.

156. Martin B, Brenneman R, Becker KG, Gucek M, Cole RN, Maudsley S.: iTRAQ analysis of complex proteome alterations in 3xTgAD Alzheimer's mice: understanding the interface between physiology and disease. PLoS One. 2008;3:e2750.

157. Kwong J, Hong L, Liao R, Deng Q, Han J, Sun P.: p38alpha and p38gamma mediate oncogenic ras-induced senescence through differential mechanisms. J Biol Chem. 2009;284:11237-46.

158. Wong RH, Chang I, Hudak CS, Hyun S, Kwan HY, Sul HS.: A role of DNA-PK for the metabolic gene regulation in response to insulin. Cell. R 206(6):1056-72.

159. Minig V, Kattan Z, van Beeumen J, Brunner E, Becuwe P.: Identification of DDB2 protein as a transcriptional regulator of constitutive SOD2 gene expression in human breast cancer cells. J Biol2009 22;284(21):14165-76.

160. Okamoto K, Taya Y, Nakagama H.: Mdmx enhances p53 ubiquitination by altering the substrate preference of the Mdm2 ubiquitin ligase. FEBS L009 3;583(17):2710-4.

161. Pan D, Zhu Q, Luo K.: SnoN functions as a tumour suppressor by inducing premature senescence. EMBO J. 2009;28:3500-13.

162. Jiang M, Fernandez S, Jerome WG, He Y, Yu X, Cai H, Boone B, Yi Y, Magnuson MA, Roy-Burman P, Matusik RJ, Shappell SB, Hayward SW.: Disruption of PPARgamma signaling results in mouse prostatic intraepithelial neoplasia involving active autophagy. Cell Death Differ. 2010;17:469-81

163. Lefkimmiatis K, Caratozzolo MF, Merlo P, D'Erchia AM, Navarro B, Levrero M, Sbisa' E, Tullo A.: p73 and p63 sustain cellular growth by transcriptional activation of cell cycle progression genes. Cancer Res.

2009;69:8563-71.

164. Luciani A, Villella VR, Vasaturo A, Giardino I, Pettoello-Mantovani M, Guido S, Cexus ON, Peake N, Londei M, Quaratino S, Maiuri L.: Lysosomal accumulation of gliadin p31-43 peptide induces oxidative stress and tissue transglutaminase-mediated PPARgamma downregulation in intestinal epithelial cells and coeliac mucosa. Gut. 2010;59:311-9.

165. Haas TL, Emmerich CH, Gerlach B, Schmukle AC, Cordier SM, Rieser E, Feltham R, Vince J, Warnken U, Wenger T, Koschny R, Komander D, Silke J, Walczak H.: Recruitment of the linear ubiquitin chain assembly complex stabilizes the TNF-R1 signaling complex and is required for TNF-mediated gene induction. Mol Cell. 2009;36:831-44.

166. Pandey AK, Munjal N, Datta M.: Gene expression profiling and network analysis reveals lipid and steroid metabolism to be the most favored by TNFalpha in HepG2 cells. PLoS One. 2010 4;5:e9063.

167.

168. Kim H, Nakamura F, Lee W, Hong C, Pérez-Sala D, McCulloch CA.: Regulation of cell adhesion to collagen via beta1 integrins is dependent on interactions of filamin A with vimentin and protein kinase C epsilon. Exp Cell Res. 2010;316:1829-44.

169. Zhang W, Guo Z, Jiang B, Niu L, Xia G, Wang X, Cheng T, Zhang Y, Wang J.: Identification of a functional p53 responsive element within the promoter of XAF1 gene in gastrointestinal cancer cells. Int J l. 0 Apr;36(4):1031-7.

170. Ozgen N, Lau DH, Shlapakova IN, Sherman W, Feinmark SJ, Danilo P Jr, Rosen MR.: Determinants of CREB degradation and KChIP2 gene transcription in cardiac memory. Heart Rhythm. 2010;7:964-70.

171. Simarro M, Giannattasio G, De la Fuente MA, Benarafa C, Subramanian KK, Ishizawar R, Balestrieri B, Andersson EM, Luo HR, Orduña A, Boyce J, Anderson P: Fas-activated serine/threonine phosphoprotein promotes immune-mediated pulmonary inflammation. J Immu010

1;184(9):5325-32.

172. Lima RT, Busacca S, Almeida GM, Gaudino G, Fennell DA, Vasconcelos MH.: MicroRNA regulation of core apoptosis pathways in cancer. Eur J er.11 Jan;47(2):163-74.

173. Kim MS, Ramakrishna S, Lim KH, Kim JH, Baek KH.: Protein stability of mitochondrial superoxide dismutase SOD2 is regulated by USP36. J Cellchem011 Feb;112(2):498-508.

174. Wiseman DA, Kalwat MA, Thurmond DC.: Stimulus-induced S-nitrosylation of Syntaxin 4 impacts insulin granule exocytosis. J Biol 201y 6;286(18):16344-54.

175. LeNoue-Newton M, Watkins GR, Zou P, Germane KL, McCorvey LR, Wadzinski BE, Spiller BW.: The E3 ubiquitin ligase- and protein phosphatase 2A (PP2A)-binding domains of the Alpha4 protein are both required for Alpha4 to inhibit PP2A degradation. J Biol2011 20;286(20):17665-71.

176. Chung YW, Kim HK, Kim IY, Yim MB, Chock PB.: Dual function of protein kinase C (PKC) in 12-O-tetradecanoylphorbol-13-acetate (TPA)-induced manganese superoxide dismutase (MnSOD) expression: activation of CREB and FOXO3a by PKC-alpha phosphorylation and by PKC-mediated inactivation of Akt, respectively. J Biol2011 26;286(34):29681-90.

177. Wang J, Huo K, Ma L, Tang L, Li D, Huang X, Yuan Y, Li C, Wang W, Guan W, Chen H, Jin C, Wei J, Zhang W, Yang Y, Liu Q, Zhou Y, Zhang C, Wu Z, Xu W, Zhang Y, Liu T, Yu D, Zhang Y, Chen L, Zhu D, Zhong X, Kang L, Gan X, Yu X, Ma Q, Yan J, Zhou L, Liu Z, Zhu Y, Zhou T, He F, Yang X.: Toward an understanding of the protein interaction network of the human liver. Mol Sy. 2011 Oct 11;7:536.

178. Krebs P, Fan W, Chen YH, Tobita K, Downes MR, Wood MR, Sun L, Li X, Xia Y, Ding N, Spaeth JM, Moresco EM, Boyer TG, Lo CW, Yen J, Evans RM, Beutler B.: Lethal mitochondrial cardiomyopathy in a hypomorphic Med30 mouse mutant is ameliorated by ketogenic diet. Proc Nad S S A. 2011 Dec 6;108(49):19678-82.

179. Faury G, Pezet M, Knutsen RH, Boyle WA, Heximer SP, McLean SE, Minkes RK, Blumer KJ, Kovacs A, Kelly DP, Li DY, Starcher B, Mecham RP.: Developmental adaptation of the mouse cardiovascular system to elastin haploinsufficiency. J Clinest.03 Nov;112(9):1419-28.

180. Papadopoulos V, Amri H, Li H, Boujrad N, Vidic B, Garnier M.: Targeted disruption of the peripheral-type benzodiazepine receptor gene inhibits steroidogenesis in the R2C Leydig tumor cell line. J Biol1997 19;272(51):32129-35.

181. Corblin F, Tripodi S, Fanchon E, Ropers D, Trilling L. A declarative constraint-based method for analyzing discrete genetic regulatory networks. Biosystems. 2009 Nov;98(2):91-104.

182. Kobayashi T, Chen L, Aihara K. Modeling genetic switches with positive feedback loops. J Theor Biol. 2003 Apr 7;221(3):379-99.

183. Powers SK, Jackson MJ. Exercise-induced oxidative stress: cellular mechanisms and impact on muscle force production. Physiol Rev. 2008 Oct;88(4):1243-76.

184. S. Smit, M.J. van Breemen, H.C.J. Hoefsloot, A.K. Smilde, J.M.F.G. Aerts, C.G. de Koster. Assessing the statistical validity of proteomics based biomarkers. Anal. Chim. Acta, 592 (2007), pp. 210–217, E. Wiederhold, T. Gandhi, H.P. Permentier, R. Breitling, B. Poolman, D.J. Slotboom. The yeast vacuolar membrane proteome Mol. Cell. Proteomics, 8 (2009), pp. 380–392

185. A. Fukushima, M. Kusano, H. Redestig, M. Arita, K. Saito. Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. BMC Syst. Biol., 5 (2011), p. 1

186. Joshua Millstein and Dmitri Volfson. Computationally efficient permutation-based confidence interval estimation for tail-area FDR. Front Genet. 2013; 4: 179.

187. J.A. Koziol Comments on the rank product method for analyzing replicated experiments FEBS Lett., 584 (2010), pp. 941–944

188. Eisinga, R., Breitling, R., and Heskes, T. (2013). The exact probability

distribution of the rank product statistics for replicated experiments. FEBS Letters, 587:677—682

189. Storey JD, Tibshirani R.: Statistical methods for identifying differentially expressed genes in DNA microarrays. Methods Mol Biol. 2003;224:149-57.

190. Wu FX. Delay-independent stability of genetic regulatory networks. IEEE Trans Neural Netw. 2011 Nov;22(11):1685-93. doi: 10.1109/TNN.2011.2165556.

191. Smolen P, Baxter DA, Byrne JH. Mathematical modeling of gene networks. Neuron. 2000 Jun;26(3):567-80.

# • Additional files

- Additional file 1 – List of genes obtained for the GRN reconstrution
- Additional file 2 – GRN reconstruction procedure
- Additional file 3 – List of genes differentially expressed and the number of comparisons in which are significantly different
- Additional file 4 – Correlation analysis

# Apendix I

The Journal of Immunology

# Relevance of the MEK/ERK Signaling Pathway in the Metabolism of Activated Macrophages: A Metabolomic Approach



Paqui G. Través, Pedro de Atauri, Silvia Marín, María Pimentel-Santillana, Juan-Carlos Rodríguez-Prados, Igor Marín de Mas, Vitaly A. Selivanov, Paloma Martín-Sanz, Lisardo Boscá and Marta Cascante

*J Immunol* 2012; 188:1402-1410; Prepublished online 21 December 2011;
doi: 10.4049/jimmunol.1101781
http://www.jimmunol.org/content/188/3/1402



*The Journal of Immunology* is published twice each month by
The American Association of Immunologists, Inc.,
9650 Rockville Pike, Bethesda, MD 20814-3994.

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

# Relevance of the MEK/ERK Signaling Pathway in the Metabolism of Activated Macrophages: A Metabolomic Approach

**Paqui G. Través,* Pedro de Atauri,[†,‡,1] Silvia Marín,[†,‡,1] María Pimentel-Santillana,* Juan-Carlos Rodríguez-Prados,[†] Igor Marín de Mas,[†,‡] Vitaly A. Selivanov,[†,‡] Paloma Martín-Sanz,[*,§] Lisardo Boscá,[*,§] and Marta Cascante[†,‡]**

**The activation of immune cells in response to a pathogen involves a succession of signaling events leading to gene and protein expression, which requires metabolic changes to match the energy demands. The metabolic profile associated with the MAPK cascade (ERK1/2, p38, and JNK) in macrophages was studied, and the effect of its inhibition on the specific metabolic pattern of LPS stimulation was characterized. A [1,2-[$^{13}C]_2$]glucose tracer-based metabolomic approach was used to examine the metabolic flux distribution in these cells after MEK/ERK inhibition. Bioinformatic tools were used to analyze changes in mass isotopomer distribution and changes in glucose and glutamine consumption and lactate production in basal and LPS-stimulated conditions in the presence and absence of the selective inhibitor of the MEK/ERK cascade, PD325901. Results showed that PD325901-mediated ERK1/2 inhibition significantly decreased glucose consumption and lactate production but did not affect glutamine consumption. These changes were accompanied by a decrease in the glycolytic flux, consistent with the observed decrease in fructose-2,6-bisphosphate concentration. The oxidative and nonoxidative pentose phosphate pathways and the ratio between them also decreased. However, tricarboxylic acid cycle flux did not change significantly. LPS activation led to the opposite responses, although all of these were suppressed by PD325901. However, LPS also induced a small decrease in pentose phosphate pathway fluxes and an increase in glutamine consumption that were not affected by PD325901. We concluded that inhibition of the MEK/ERK cascade interferes with central metabolism, and this cross-talk between signal transduction and metabolism also occurs in the presence of LPS.** ***The Journal of Immunology*****, 2012, 188: 1402–1410.**

Macrophages have important roles in innate and acquired immunity, as well as in tissue homeostasis (1, 2). Their activation is a complex process involving signaling events triggered by multiple inflammatory mediators, including exogenous factors, such as LPS, and endogenous mediators, such as cytokines and chemokines. Cytokines are major regulators of macrophage activation that limit the amount of inflammation and, thus, prevent toxicity and tissue damage (3, 4). Failure to induce an inflammatory response promotes unrestricted microbial proliferation and the development of serious infections, whereas excessive production of proinflammatory mediators may also be life threatening, as observed in patients with severe sepsis or septic shock. Therefore, immune responses must be tightly regulated (3, 5, 6).

NF-κB and MAPK signaling pathways (ERK, JNK, and p38) play a key role in the activation and regulation of innate and adaptive immune responses. For example, macrophages activate MEK/ERK cascade in response to bacterial infection. MEK/ERK signaling is involved in the activation of oxidative and nitrosative bursts, endosomal trafficking, and proinflammatory macrophage polarization (1, 3, 7–9). Therefore, MEK/ERK signaling is likely to enhance macrophage activity against intracellular pathogens (10–12). The MEK/ERK pathway in macrophages is one of the most widely studied intracellular signaling cascades involved in LPS-induced proinflammatory responses (10). In addition to this, the effect of inhibition of p38 and JNK with the selective inhibitors BIRB796 and BI78D3, respectively, has been evaluated (12, 13).

*Instituto de Investigaciones Biomédicas Alberto Sols, Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid, 28029 Madrid, Spain; [†]Department of Biochemistry and Molecular Biology, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain; [‡]Institute of Biomedicine, University of Barcelona, 08036 Barcelona, Spain; and [§]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, 08028 Barcelona, Spain

[1]P.d.A. and S.M. contributed equally to this work.

Received for publication June 17, 2011. Accepted for publication November 16, 2011.

This work was supported by Grants SAF2008-00164, BFU2011-24760, and PIB2010BZ-00540 from Spanish Ministry of Science and Innovation, Red Temática de Investigación Cooperativa en Cáncer, the Instituto de Salud Carlos III, Spanish Ministry of Science and Innovation and European Regional Development Fund "Una manera de hacer Europa" ISCIII-RTICC (RD6/0020/0046), and FIS-RECAVA (RD06/0014/0006) and CIBERehd founded by Instituto de Salud Carlos III, the European Commission (FP7) Etherpath KBBE Grant Agreement 222639, and by Agència de Gestió d'Ajuts Universitaris i de Recerca-Generalitat de Catalunya (Grant 2009SGR1308, 2009 CTP 00026, and Icrea Academia award 2010 to M.C.).

Address correspondence and reprint requests to Dr. Lisardo Boscá or Dr. Marta Cascante, Instituto de Investigaciones Biomédicas "Alberto Sols," Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid, Arturo Duperier 4, 28029 Madrid, Spain (L.B.) or Department of Biochemistry and Molecular Biology, Faculty of Biology, Universitat de Barcelona, Edifici Nou, Planta-2, Avinguda Diagona 1645, 08028 Barcelona, Spain (M.C.). E-mail addresses: lbosca@iib.uam.es (L.B.) and marta.cascante@ub.edu (M.C.)

Abbreviations used in this article: COX-2, cyclooxygenase 2; DAF-2DA, 4,5-diaminofluorescein diacetate; DCFH-DA, dichlorofluorescein diacetate; FBPase-2, fructose-2,6-bisphosphatase; Fru-1,6-$P_2$, fructose-1,6-bisphosphate; Fru-2,6-$P_2$, fructose-2,6-bisphosphate; G6PDH, glucose-6-phosphate dehydrogenase; L-PFK-2, liver-type-PFK-2; Mal, malate; NOS-2, NO synthase 2; Oaa, oxaloacetate; PDH, pyruvate dehydrogenase; PFK-1, 6-phosphofructo-1-kinase; PFK-2, 6-phosphofructo-2-kinase; 6PGDH, 6-phospho-D-gluconate dehydrogenase; PI, propidium iodide; PPP, pentose phosphate pathway; Pyr, pyruvate; ROS, reactive oxygen species; TCA, tricarboxylic acid; uPFK-2, PFKB3 isoenzyme of PFK-2.



www.jimmunol.org/cgi/doi/10.4049/jimmunol.1101781

Immune activation rapidly and substantially enhances metabolic outputs (14, 15). Macrophage activation is followed by rapid changes in nutrient flux, which also seems to be necessary for immune activation, indicating that signals produced by immune cells might directly regulate their metabolism. Indeed, studies have highlighted a key role for activated macrophages in controlling energy metabolism and insulin action (15–17). For example, low-grade chronic inflammation is associated with accumulation of macrophages in adipose tissue and predisposition to insulin resistance (15, 18).

In the current study, we aimed to characterize changes in the central carbon metabolic network induced by ERK inhibition and provide a tool to analyze the metabolic flux distribution in macrophages as cross-talk between signal transduction and metabolic events. For this purpose, we used LPS as a model of proinflammatory activation and PD325901 as a selective inhibitor of the MEK/ERK cascade (12). To determine the metabolic state of the cells, we used a tracer-based metabolomics approach with [1,2-[$^{13}$C]$_2$]glucose as the carbon source. Mass isotopomer distribution analysis of key metabolites has been described as a powerful tool to map metabolic flux distribution in several cellular models (19, 20). By tracking the changes in metabolic fluxes induced by ERK signaling modulators, we observed details of the cross-talk between inflammatory signal transduction and metabolic networks. Similar results on glycolytic metabolism were observed in a macrophage cell line in primary cultures of murine peritoneal macrophages and in human monocytes/macrophages.

## Materials and Methods

### *Materials*

The murine macrophage cell line RAW 264.7 was obtained from the American Type Culture Collection (Manassas, VA). RPMI 1640, FBS, cell culture, and chemical reagents were obtained from Lonza (Cologne, Germany); PD325901, BIRB796, and BI78D3 were from Calbiochem (San Diego, CA). [1,2-[$^{13}$C]$_2$]glucose (>99% enriched) was from Isotec (Miamisburg, OH). LPS and reagents for metabolite derivatization were from Sigma-Aldrich (St. Louis, MO). Abs were from Santa Cruz Biotech (Santa Cruz, CA), Cell Signaling (Danvers, MA), or Sigma-Aldrich.

### *Cell culture conditions*

RAW 264.7 cells were cultured in RPMI 1640 supplemented with glutamine (2 mM), 10% FBS, and antibiotics (100 U/ml penicillin, 100 μg/ml streptomycin, and 50 μg/ml gentamicin) at 37°C in 5% $CO_2$. When cells reached 80% subconfluency, the medium was replaced with a medium containing only 2% FBS. After overnight serum reduction, cell cultures were loaded with [1,2-[$^{13}$C]$_2$]glucose and treated with 0.5 μM PD325901 and 500 ng/ml LPS for the indicated periods of time. The same procedure was used for studies with p38 and JNK inhibitors but in the absence of labeled glucose. Following incubation, the medium was removed, and cells were scraped off the dishes and processed for RNA, proteins, and intracellular metabolites. Murine peritoneal macrophages and human monocyte/macrophages were prepared (14, 21) and were used as described for the RAW 264.7 cells.

### *Flow cytometry*

Cells were harvested and washed in PBS. After centrifugation at 4°C for 5 min and 1000 × *g*, cells were resuspended in Annexin V binding buffer (10 mM HEPES [pH 7.4], 140 mM NaCl, 2.5 mM $CaCl_2$) and labeled with Annexin V$^{FITC}$ solution and/or propidium iodide (PI) (100 μg/ml) for 15 min at room temperature in the dark. PI is impermeable to living and early apoptotic cells but stains necrotic and apoptotic dying cells with impaired membrane integrity in contrast to Annexin V, which stains early apoptotic cells.

### *6-Phosphofructo-2-kinase activity assay*

Cells (grown in 6-cm dishes) were homogenized in 1 ml a medium containing 20 mM potassium phosphate ([pH 7.4], 4°C), 1 mM DTT, 50 mM NaF, 0.5 phenylmethanesulfonyl fluoride, 10 μM leupeptin, and 5% poly (ethylene)glycol. After centrifugation in an Eppendorf centrifuge (15 min), poly(ethylene)glycol was added to the supernatant up to 15% (mass/vol) to fully precipitate the 6-phosphofructo-2-kinase (PFK-2). After resuspension of the pellet in the extraction medium, PFK-2 activity was assayed at pH 8.5 with 5 mM MgATP, 5 mM fructose-6-phosphate, and 15 mM glucose-6-phosphate. One unit of PFK-2 activity is the amount of enzyme that catalyzes the formation of 1 pmol Fru-2,6-bisphosphate (Fru-2,6-$P_2$)/min (22).

### *Metabolite assays*

Fru-2,6-$P_2$ was extracted from cells (cultured in 24-well plates) after homogenization in 100 μl 50 mM NaOH, followed by heating at 80°C for 10 min. The metabolite was measured by the activation of the pyrophosphate-dependent 6-phosphofructo-1-kinase (PFK-1) (22). Glucose and lactate were measured enzymatically in the culture medium (23). Glutamine was determined after deamination to glutamate, which was measured enzymatically using the enzyme glutamate dehydrogenase (23). NO release was determined spectrophotometrically by the accumulation of nitrite and nitrate in the medium (phenol red-free), as described before (14).

### *Preparation of cell extracts*

Cells (grown in six-well dishes) were washed twice with ice-cold PBS and homogenized in 0.2 ml buffer containing 10 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 1 mM EGTA, 10% glycerol, 0.5% CHAPS, 1 mM 2-ME, 0.1 mM PMSF, and a protease inhibitor mixture (Sigma-Aldrich). The extracts were vortexed for 30 min at 4°C and centrifuged for 10 min at 13,000 × *g*. The supernatants were stored at −20°C. Protein levels were determined using the Bio-Rad detergent-compatible protein reagent (Richmond, CA). All steps were carried out at 4°C.

### *Western blot analysis*

Samples of cell extracts containing equal amounts of protein (30 μg/lane) were boiled in 250 mM Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, and 2% 2-ME and separated in 10% SDS-PAGE. The gels were blotted onto a polyvinylidene fluoride membrane (GE Healthcare, Barcelona, Spain) and processed as recommended by the supplier of the Abs against the murine Ags: phospho-ERK1/2 (9101s), phospho-p38 (9211s), phospho-JNK (9251s), NO synthase 2 (NOS-2; sc-7271), cyclooxygenase 2 (COX-2; sc-1999), liver-type–PFK-2 (L–PFK-2) (sc-10096), and β-actin (A-5441). For PFKB3 isoenzyme of PFK-2 (uPFK-2), specific peptides of the isoenzyme were used to generate polyclonal Abs by immunizing rabbits (New Zealand White) with multiple intradermal injections of 300 μg Ag in 1 ml CFA, followed by boosters with 100 μg Ag in IFA. The blots were developed by the ECL protocol (Amersham), and different exposure times were used for each blot with a charged-coupling device camera in a luminescent-image analyzer (Molecular Imager, Bio-Rad) to ensure linearity of the band intensities.

### *RNA isolation and RT-PCR analysis*

One microgram of total RNA, extracted with TRIzol Reagent (Invitrogen) according to the manufacturer's instructions, was reverse transcribed using Transcriptor First Strand cDNA Synthesis Kit for RT-PCR, following the instructions of the manufacturer (Roche). Real-time PCR was conducted with SYBR Green on a MyiQ real-time PCR System (Bio-Rad), using the SYBR Green method. PCR thermocycling parameters (24) were 95°C for 10 min, 40 cycles of 95°C for 15 s, and 60°C for 1 min. All samples were analyzed for 36B4 expression in parallel. Each sample was run in duplicate and was normalized to 36B4. The replicates were then averaged, and fold induction was determined on ΔΔCt-based fold-change calculations. Primer sequences are available on request.

### *Measurement of reactive oxygen species and NO synthesis*

The generation of reactive oxygen species (ROS) was monitored using dichlorofluorescein diacetate (DCFH-DA). Cells were preincubated with 10 μM DCFH-DA for 15 min and fluorescence was measured using a cell cytometer. For fluorometric NO determination, the cell-permeable fluorophore 4,5-diaminofluorescein diacetate (DAF-2DA) was used. Cells were preincubated with 10 μM DAF-2DA for 15 min, and DAF-2DA fluorescence was measured in a cell cytometer.

### *Metabolite isolation and isotopologue analysis*

Glucose, lactate, and glutamate from the incubation medium were purified, derivatized, and analyzed, as previously described (19). Thus, glucose was purified from culture medium using a tandem set of Dowex-1X8/Dowex-50WX8 (Sigma-Aldrich) ion-exchange columns and converted to its aldonitrile pentaacetate derivative. The ion cluster around m/z 328 was

monitored (carbons 1 to 6 of glucose, chemical ionization). Lactate from the cell culture media was extracted by ethyl acetate after acidification with HCl. Lactate was derivatized to its propylamide-heptafluorobutyric form, and the cluster around m/z 328 (carbons 1 to 3 of lactate, chemical ionization) was monitored. Glutamate was separated from the medium using ion-exchange chromatography and converted to its n-trifluoroacetyl-n-butyl derivative. The ion clusters around m/z 198 (carbons 2 to 5 of glutamate, electron impact ionization) and m/z 152 (carbons 2 to 4 of glutamate, electron impact ionization) were monitored. RNA ribose was purified, derivatized, and analyzed, as previously described (20). In detail, RNA ribose was isolated by acid hydrolysis of cellular RNA after TRIzol purification of cell extracts. Ribose isolated from RNA was derivatized to its aldonitrile acetate form using hydroxylamine in pyridine and acetic anhydride, and the ion cluster around the m/z 256 (carbons 1 to 5 of ribose, chemical ionization) was monitored. Spectral data were corrected using regression analysis to extract natural [$^{13}$C] enrichment from the results (25). Measurement of [$^{13}$C] label distribution determined the different relative distribution percentages of isotopologues, and m0 (molecules without any [$^{13}$C] labels), m1 (molecules with one [$^{13}$C]), m2 (with two [$^{13}$C]), and so forth were reported as molar fractions.

### *Gas chromatography/mass spectrometry*

Mass spectral data were obtained on a QP2010 mass selective detector connected to a GC-2010 gas chromatograph (Shimadzu Scientific Instruments) using helium as the gas carrier and isobutane 0.0016 Pa as the reagent gas in chemical-ionization analysis. Settings were as follows: gas chromatograph inlet, 250°C for glucose, ribose, and glutamate and 200°C for lactate; transfer line, 250°C; and mass chromatography source, 200°C. A Varian VF-5 capillary column (30 m in length, 250 μm in diameter, and with a 0.25-μm film thickness) was used to analyze all of the compounds. In vitro experiments were carried out using duplicate cultures each time for each treatment regimen. Mass spectral analyses were carried out by three independent automated injections of 1 μl each sample and were accepted only if the standard sample deviation was <1% of the normalized peak intensity.

### *Estimation of internal fluxes based on the measured [$^{13}$C] redistribution*

Each $^{13}$C-labeled metabolite corresponds to a different isotopomer, which differs only in the labeling state of its individual atoms (26). For a specific metabolite, the number of possible isotopomers, $2^n$, depends on the number, $n$, of carbons for each metabolite. The relative abundance of product isotopomers depends on the labeled status of the substrates (50% [1,2-[$^{13}$C]$_2$]glucose) and the flux distribution throughout the metabolic network (27–29). Isotopomer abundances can be predicted by solving a system of isotopomer mass balance equations, where each equation describes the dependency of each isotopomer abundance on fluxes and isotopomer abundance of other metabolites (30). The space of solution for each condition (vehicle, LPS, PD325901, and PD325901+LPS) is scanned by solving the system of equations for feasible combinations of flux values for all reaction steps. All combinations satisfied the constraints associated with network topology described below, stoichiometry for each reaction, and measured fluxes for glucose consumption, lactate production, and glutamine consumption (23). Also, total [$^{13}$C] enrichment of ribose ($\sum$m = m1+m2+m3) was applied to fix the differential de novo RNA synthesis (step D in Fig. 5) among the analyzed conditions (Table I). For reversible reactions, exchange fluxes, which account for the cycle through the forward and backward reactions (29, 31), were considered in addition to the net reactions. Ratios m1/(m1+m2) and m2/(m1+m2) for lactate and glutamate C2–C4 and C2–C5 fragments and m1/(m1+m2+m3), m2/(m1+m2+m3), and m3/(m1+m2+m3) for ribose measured experimentally (Table I) and predicted for different flux distributions were compared (least squares). The 20 combinations of flux distributions with a best fitting were taken for each case (vehicle, PD325901, LPS, and PD325901+LPS).

### *Network structure*

The assumed network scheme corresponds to those in Fig. 5. Each solid arrow indicates a reversible or irreversible reaction step catalyzed by an enzyme (or transporter) or one block of enzymes. Dashed lines indicate regulatory connections (product inhibition by glucose-6-phosphate and activation of pyruvate kinase by fructose-1,6-bisphosphate (Fru-1,6-P$_2$). Letters correspond to the reaction steps. Reaction steps A–P account for glycolysis, as well as pentose phosphate pathway (PPP) and tricarboxylic acid (TCA) cycle enzyme-catalyzed reactions. Some reactions are neglected and grouped into blocks (e.g., reaction step F representing the block from GAPDH to pyruvate kinase), and others are assumed to be involved in rapid equilibriums (e.g., glucose-6-phosphate isomerase). Metabolites are combined into pools: a first pool for hexose-phosphates, including glucose-6-phosphate and fructose-6-phosphate; a second pool for pentose-phosphates accounting for ribose-5-phosphate, ribulose-5-phosphate, and xylulose-5-phosphate; and a third pool for oxaloacetate (Oaa) and malate (Mal). The rest of the metabolic intermediaries are Fru-1,6-P$_2$, dihydroxyacetonephosphate, GAPDH, sedoheptulose-7-phosphate, erythrose-4-phosphate, pyruvate (Pyr), acetyl-CoA, citrate, 2-oxoglutarate, and succinyl-CoA. Reaction steps A, D, T, U, X, Y, and Z represent the inputs and outputs of the metabolic system.

### *Estimation of flux dependencies on enzyme activities*

Estimation of flux dependencies on enzyme activities are based on the identification of control coefficients with fixed signs. The sign and magnitude of control coefficients depend on the topology of the network, the stoichiometry of the reactions, and the magnitudes of fluxes and of regulatory dependencies (enzyme–substrate affinity, inhibitions, and activations) (23). Magnitudes of regulatory dependencies are unknown, but the sign of some control coefficients are fixed, irrespective of these magnitudes. Others are sign indeterminate, meaning that they can be positive and negative, and some are always zero.

### *Statistical analysis*

The data shown are the means ± SD of three or four experiments. Statistical significance was estimated with the Student $t$ test for unpaired observations. Significance of isotopologue data (Table I) was analyzed using two-way ANOVA.

## Results

### *Characterization of macrophage activation after ERK1/2 inhibition*

To characterize the response of RAW 264.7 cells to the MEK/ERK selective inhibitor PD325901 (12) and LPS activation, several functional markers were used. Fig. 1*A* shows the dose-dependent inhibition of ERK1/2 phosphorylation by PD325901 in LPS-activated cells. The inhibitor significantly decreased LPS-induced NOS-2 and COX-2 protein levels (Fig. 1*B*), as well as nitrite plus nitrate accumulation in the medium (Fig. 1*C*). At the metabolic level, PD325901 decreased the basal levels of Fru-2,6-P$_2$, a potent activator of the glycolytic flux, and impaired its increase induced by LPS (Fig. 1*D*). This was associated with a decrease in the expression of the highly active uPFK-2 isoform induced by LPS and, concomitantly, a reduction in total PFK-2/fructose-2,6-bisphosphatase (FBPase-2) activity (Fig. 1*E*). Similar results in terms of ERK inhibition, NOS-2 and COX-2 expression, and changes in PFK-2 isoenzymes were observed with the MEK/ERK inhibitors SL327 and PD98059 (data not shown). Changes in mRNA correlated with those observed for protein levels of NOS-2, COX-2, uPFK-2, and L–PFK-2 (Fig. 1*F*). Moreover, to reinforce the specific effect of ERK1/2 inhibition on LPS activation, an increase in IL-12p40 (IL-12) and decrease in TNF-α mRNA levels were observed (Fig. 1*G*), as described before (32). PD325901 impaired LPS induction of IL-1β and IL-6 mRNA levels (Fig. 1*G*) but did not affect the levels of the chemokines CXCL-1 and CXCL-10 (Fig. 1*H*). Because cell activation might interfere with viability, the percentage of apoptotic cells was determined by measuring Annexin V and PI staining. PD325901 moderately influenced cell viability in resting macrophages but enhanced apoptosis in LPS-activated cells (Fig. 2*A*). Moreover, PD325901 decreased cell numbers at 18 h but did not significantly affect the percentage of cells gating at the S, G$_2$, and M phases of the cell cycle, which was <18% (Fig. 2*A*). The oxidation of DCFH-DA and DAF during LPS activation was measured at 18 h. PD325901 moderately increased the oxidation of both probes but impaired the large changes that accompany LPS activation (Fig. 2*B*). An image of cells after 18 h of treatment is shown in Fig. 2*C*.

To characterize the metabolic changes induced by ERK1/2 inhibition, RAW 264.7 cells were treated with 0.5 μM of PD325901

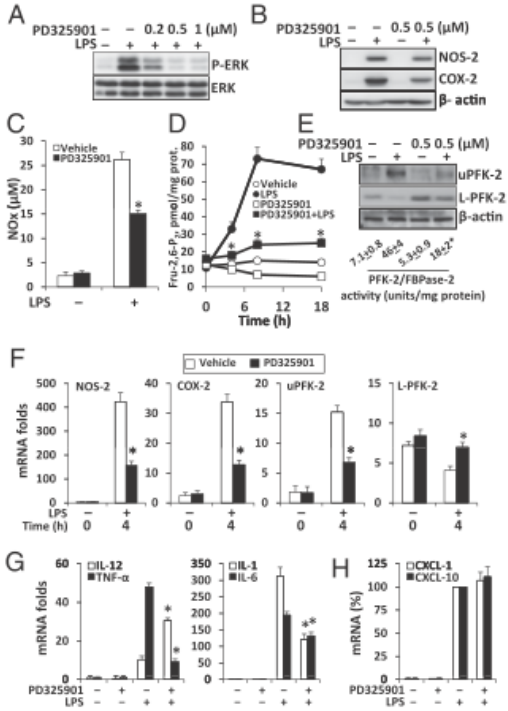



**FIGURE 1.** Effect of MEK/ERK inhibition on LPS activation of RAW 264.7 macrophages. Cells were maintained overnight in 2% FBS and treated with the indicated concentrations of PD325901 10 min before activation with 500 ng/ml LPS. The levels of phospho-ERK1/2 were determined at 30 min (*A*), and the levels of NOS-2, COX-2 (*B*), and nitrite plus nitrate in the medium (NOx) were determined after 18 h (*C*). *D*, The time course of the intracellular levels of Fru-2,6-$P_2$ was evaluated after treatment with 0.5 μM PD325901 and 500 ng/ml LPS. *E*, The protein levels of uPFK-2 and L-PFK-2 and the PFK-2/FBPase-2 activity were determined at 18 h. *F*, The mRNA levels of the indicated genes were determined at 0 and 4 h after LPS activation. The mRNA levels of IL-12, activated upon ERK1/2 inhibition, and TNF-α, IL-1β, IL-6 (*G*), and the chemokines CXCL-1 and CXCL-10 (*H*) were determined at 4 h after treatment. Results are representative blots for four experiments or the mean ± SD of four experiments. $^{*}p < 0.01$ versus no PD325901.

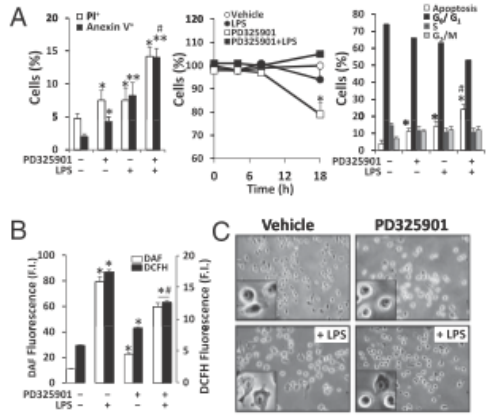



**FIGURE 2.** Effect of MEK/ERK inhibition on cell viability and oxidative stress. Cells were pretreated with 0.5 μM PD325901 10 min before activation with 500 ng/ml LPS. *A*, After 8 h of treatment, the percentage of cells positive for PI and Annexin V staining was determined (*left panel*). The time course of the cell density (*center panel*) and the cell cycle distribution at 18 h (*right panel*) were determined. *B*, The changes in fluorescence of DAF and DCFH-DA were determined at 18 h. *C*, A representative photograph of macrophages treated for 18 h with PD325901 and LPS at low cell density. Original magnification ×100; ×400 *inset*). Results show the mean ± SD of three experiments. $^{*}p < 0.05$, $^{**}p < 0.01$ versus the untreated condition. $^{\#}p < 0.01$ versus no PD325901.

and/or 500 ng/ml of LPS. Glucose and glutamine consumptions and lactate production after 1, 4, and 8 h of incubation are presented in Fig. 3*A*. Both glucose consumption and lactate production were lower in the presence of PD325901, whereas glutamine consumption was not affected. LPS stimulation increased glucose consumption and lactate production but did not induce these effects in the presence of PD325901. Interestingly, LPS increased glutamine consumption, regardless of the presence of PD325901. The ratios between lactate production and glucose consumption, as well as glucose/glutamine consumption are shown (Fig. 3*B*). Because glucose-6-phosphate dehydrogenase (G6PDH) and 6-phospho-D-gluconate dehydrogenase (6PGDH) activities might be affected by PD325901 in LPS-activated cells, the time course of their activity was measured, with a modest transient increase at 8 h, independent of PD325901 treatment (Fig. 3*C*). In addition to RAW 264.7 cells, the effect of the inhibition of ERK on LPS-dependent activation of glycolysis was investigated in peritoneal murine macrophages and in human monocyte/macrophages. As Fig. 3*D* shows, LPS challenge promoted uPFK-2 expression and an increase in Fru-2,6-$P_2$ levels in these macrophages. Treatment with PD325901 blunted the effect of LPS on both uPFK-2 expression and Fru-2,6-$P_2$ increase. In addition to this, good correlations between uPFK-2/Fru-2,6-$P_2$ levels and glucose consumption and lactate production were observed in the three types of macrophages analyzed (Fig. 3*E*).

Inhibition of p38 and JNK MAPKs with selective inhibitors was also evaluated in RAW 264.7 cells. The p38 inhibitor BIRB796 did not significantly affect cell viability at 0.5 μM (Fig. 4*A*; previous p38 inhibitors exhibited cytotoxic effects) and suppressed p38 phosphorylation (Fig. 4*B*). However, the selective JNK inhibitor BI78D3 significantly decreased cell viability at the minimal concentration required to suppress JNK phosphorylation in response to LPS (Fig. 4*A*, 4*B*). p38 inhibition did not influence the LPS-dependent uPFK-2 expression (Fig. 4*C*), the increase in Fru-2,6-$P_2$ levels, or the glycolytic flux in RAW 264.7 cells (Fig. 4*D*). With regard to JNK inhibition, it is difficult to draw conclusions about the effects on cell viability. Although treatment with BI78D3 decreased uPFK-2 levels at 8 h after LPS treatment (Fig. 4*C*), the Fru-2,6-$P_2$ levels at 8 h were 81% of those of LPS (Fig. 4*D*), which contrasts with the 69% inhibition observed after MEK/ERK inhibition (Fig. 1*D*).

### *Measured isotopologue distribution*

The metabolism of [1,2-[$^{13}$C]$_2$]glucose causes rearrangement, exchange, or loss of the [$^{13}$C] label, which is incorporated into the glucose metabolic intermediates in specific patterns. The [$^{13}$C] label enrichment of these intermediates also depends on the dilution of their unlabeled counterparts. Thus, a specific isotopologue distribution provides information on the flux of metabolites along the forward and reverse pathways of substrate cycles. RAW 264.7 cells treated with 0.5 μM of PD325901 and/or 500 ng/ml of LPS were incubated for 18 h with 10 mM glucose 50% enriched in [1,2-[$^{13}$C]$_2$]-D-glucose, and the isotopologue distributions were measured (Table I).

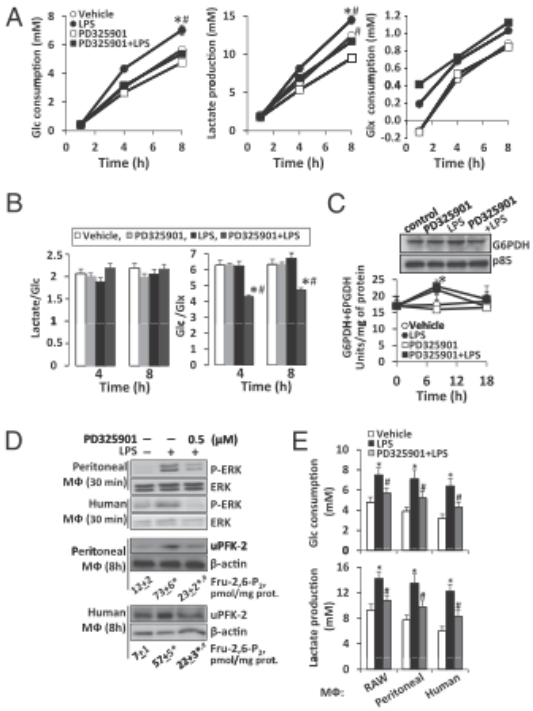


**FIGURE 3.** Effect of MEK/ERK inhibition on LPS-activation of metabolic fluxes in macrophages. RAW 264.7 cells were pretreated with 0.5 μM PD325901 10 min before activation with 500 ng/ml LPS. *A*, The glucose and glutamate consumption and lactate production were determined at the indicated times. *B*, The ratios between lactate and glucose and glucose and glutamate concentrations at 4 and 8 h after activation. *C*, G6PDH+6PGDH activities were determined at the indicated times, and a blot showing the protein levels at 18 h is shown. *D*, The effect of 0.5 μM PD325901 on LPS activation in peritoneal murine macrophages and human monocyte/macrophages was analyzed in terms of ERK phosphorylation (30 min), as well as uPFK-2 expression and Fru-2,6-$P_2$ levels (8 h). *E*, Glucose consumption and lactate release were determined at 8 h. Results show the mean ± SD of four experiments. $^{*}p < 0.01$ versus the untreated condition; $^{\#}p < 0.05$ versus no PD325901.

*Glucose and lactate in the medium.* Glucose enrichment was not significantly affected either by PD325901 or LPS treatment alone or in combination (data not shown), indicating that the macrophages did not release newly synthesized glucose into the medium. With regard to lactate, [$^{13}$C] incorporation through glycolysis results in the formation of lactate with two [$^{13}$C] (m2 lactate). m1 lactate mainly originates from the decarboxylation of [$^{13}$C] caused by the metabolism of [1,2-[$^{13}C_2$]glucose through the oxidative branch of the PPP and its subsequent recycling to glycolysis through the nonoxidative branch of PPP or by the action of Pyr cycling (mediated by phosphoenolpyruvate carboxykinase or malic enzyme). The parameter PPC (PPC = [m1/m2]/[3+(m1/m2)]) that represents the contribution of these last two pathways over glycolysis was lower after activation with LPS, regardless of the presence of PD325901. This suggested that MEK inhibition did not affect the relative contribution of these pathways to lactate formation.

*Ribose in RNA.* Pentose phosphates can be synthesized from glucose or glycolytic intermediates through two pathways: the oxidative and nonoxidative branches of the PPP. The ratio of m1/m2

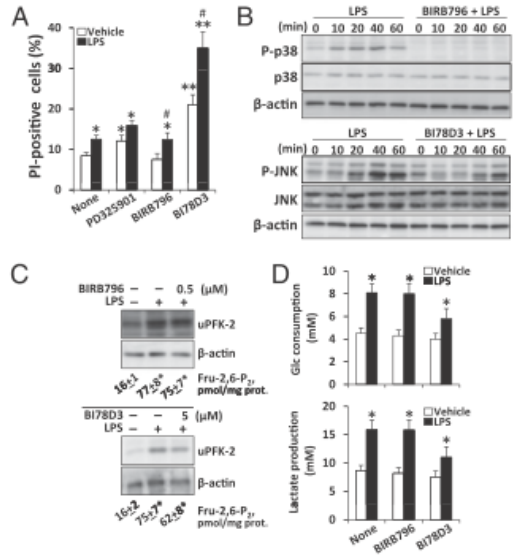


**FIGURE 4.** Effect of MAPK inhibition on LPS activation of metabolic fluxes in macrophages. RAW 264.7 cells were pretreated with 0.5 μM BIRB796 (p38 inhibitor) or 5 μM BI78D3 (JNK inhibitor) 10 min before activation with 500 ng/ml LPS. *A*, Cell viability was determined at 8 h by PI staining. *B*, MAPK inhibition was determined at the indicated times. The effect of MAPK inhibitors on LPS activation was analyzed in terms of uPFK-2 expression and Fru-2,6-$P_2$ levels (*C*), as well as glucose consumption and lactate release at 8 h (*D*). Results show the mean ± SD of three experiments. $^{*}p < 0.05$, $^{**}p < 0.01$ versus the untreated condition; $^{\#}p < 0.05$ versus no MAPK inhibitor.

among the different ribose isotopologue fractions represents the contribution of the oxidative versus the nonoxidative branch of PPP. This ratio changes from 1.29 in control to 1.10 in the presence of PD325901, 1.08 after LPS activation, and 1.13 in the presence of both, indicating a similar decrease in the oxidative branch of ribose synthesis in all cases. A part of RNA ribose was not synthesized de novo, because the nonlabeled nucleotides that existed before the incubation were reused in subsequent generations. This reused part contributed to the value of the nonlabeled fraction (m0) of defined RNA ribose. The lower m0 value found in control and LPS conditions suggested that PD325901 addition resulted in diminishing de novo synthesis of nucleotides.

*Glutamate in the medium.* Label distribution in glutamate allows us to estimate the relative contributions of pyruvate carboxylase and pyruvate dehydrogenase (PDH) to the TCA cycle (19). The fact that glutamate was mainly labeled at the fourth and fifth positions in all incubation conditions demonstrated that [$^{13}$C] from [1,2-[$^{13}C_2$]glucose entered the TCA cycle, mainly by PDH in RAW 264.7 cells, regardless of treatment. Furthermore, glutamate labeling increased in the presence of PD325901 and/or LPS, indicating that both stimuli and their combination increased the exchange between glutamate and α-ketoglutarate.

*Estimation of internal fluxes*

Mass isotopomer distribution analysis was completed with a numerical estimation of internal fluxes. To reveal the profiles of internal metabolic fluxes that underlie the isotopologue distributions corresponding to ERK1/2 inhibition in resting or activated cells, we analyzed the label distributions using the approach described in *Materials and Methods*. The metabolic network ana-

Table I. Isotopologue distribution in different metabolites

| Metabolite | Vehicle | LPS | PD325901 | PD325901+LPS |
|---|---|---|---|---|
| Lactate C1–C3 | | | | |
| m0 | 0.783 ± 0.0033 | 0.783 ± 0.0116 | 0.790 ± 0.004 | 0.772 ± 0.006*,# |
| m1 | 0.0200 ± 0.0033 | 0.0156 ± 0.0026** | 0.0184 ± 0.0016 | 0.0173 ± 0.0020 |
| m2 | 0.198 ± 0.004 | 0.212 ± 0.006** | 0.190 ± 0.004** | 0.211 ± 0.004* |
| PPC | 0.033 ± 0.006 | 0.024 ± 0.004** | 0.031 ± 0.003 | 0.027 ± 0.003* |
| Ribose C1–C5 | | | | |
| m0 | 0.752 ± 0.006 | 0.766 ± 0.002 | 0.801 ± 0.005** | 0.771 ± 0.003* |
| m1 | 0.121 ± 0.004 | 0.103 ± 0.003* | 0.092 ± 0.004** | 0.102 ± 0.001** |
| m2 | 0.0938 ± 0.0026 | 0.0954 ± 0.0012 | 0.0840 ± 0.0012* | 0.0905 ± 0.0017 |
| m3 | 0.0206 ± 0.0015 | 0.0212 ± 0.0006 | 0.010 ± 0.0086* | 0.0234 ± 0.0019 |
| m1/m2 | 1.29 ± 0.00 | 1.08 ± 0.05** | 1.10 ± 0.06** | 1.13 ± 0.02** |
| Glutamate C2–C5 | | | | |
| m0 | 0.974 ± 0.001 | 0.959 ± 0.002** | 0.960 ± 0.003** | 0.956 ± 0.002**,# |
| m1 | 0.0050 ± 0.0006 | 0.0111 ± 0.0008** | 0.0079 ± 0.0020** | 0.0099 ± 0.0008** |
| m2 | 0.0201 ± 0.0005 | 0.0287 ± 0.0009** | 0.0308 ± 0.0012** | 0.033 ± 0.0006**,## |
| Glutamate C2–C4 | | | | |
| m0 | 0.975 ± 0.001 | 0.960 ± 0.001** | 0.960 ± 0.003** | 0.956 ± 0.001**,## |
| m1 | 0.0245 ± 0.0007 | 0.0390 ± 0.0013** | 0.0390 ± 0.0026** | 0.0435 ± 0.0015**,## |
| m2 | 0.0007 ± 0.0003 | 0.0011 ± 0.0005 | 0.0007 ± 0.0012 | 0.0006 ± 0.0006 |
| Contributions to TCA cycle | | | | |
| Pyruvate carboxylase | 0.04 ± 0.01 | 0.04 ± 0.02 | 0.01 ± 0.04* | 0.02 ± 0.02 |
| PDH | 0.96 ± 0.01 | 0.96 ± 0.02 | 0.99 ± 0.04* | 0.98 ± 0.02 |

Isotopologue distribution of lactate (fragment C1–C3) and glutamate (fragments C2–C5 and C2–C4) secreted into the culture medium and RNA ribose (fragment C1–C5) after 18 h without LPS or PD325901 (vehicle) or with LPS and PD325901 individually or in combination. PPC parameter was estimated from the formula $(m1/m2)/\{3+(m1/m2)\}$ using lactate isotopologue fractions. Pyruvate carboxylase and PDH contributions to TCA cycle were estimated using $m2_{C2\text{-}C4}/m2_{C2\text{-}C5}$ and $(m2_{C2\text{-}C5} - m2_{C3\text{-}C4})/m2_{C2\text{-}C5}$, respectively. Values are expressed as mean ± SD.

*$p < 0.05$, **$p < 0.01$ versus the untreated condition; #$p < 0.05$, ##$p < 0.01$ versus no PD325901.

lyzed is depicted in Fig. 5, and the resulting numerical estimation of fluxes throughout the main steps in the metabolic network is presented in Fig. 6. The flux profile results indicated that RAW 264.7 cells under basal conditions were mainly glycolytic, having most of the consumed glucose (flux through A) converted into lactate (flux through T). The consumed glutamine in the TCA cycle (flux through U) was transformed to Oaa-Mal (fluxes through O and P), mainly recycled to Pyr (flux through R), and excreted into the medium as lactate. Flux through PDH (flux through L) was ~40–80 times lower than that from the triose phosphate pool to Pyr (flux through F), suggesting that glucose and glutamine are mainly rerouted to lactate, and only ~1.25–2.5% of the Pyr produced from glucose enters the TCA cycle in RAW 264.7 cells. The incubation of RAW 264.7 cells with PD325901 produced a clear decrease in almost all the analyzed fluxes. Furthermore, LPS increased the glycolytic flux, although this was inhibited by PD325901. With regard to PPP fluxes, observed differences in fluxes through B, G, H, and I showed a clear decrease in the presence of PD325901. A smaller decrease in the PPP fluxes was induced by LPS and when cells were coincubated with PD325901 and LPS. These differences in flux profiles are consistent with the different consumptions and productions of glucose, lactate, and glutamine and de novo synthesis of nucleotides.

### *Flux dependencies on enzyme activities*

At a specific network description of central carbon metabolism with a particular topology, reaction stoichiometry, flux values and sign

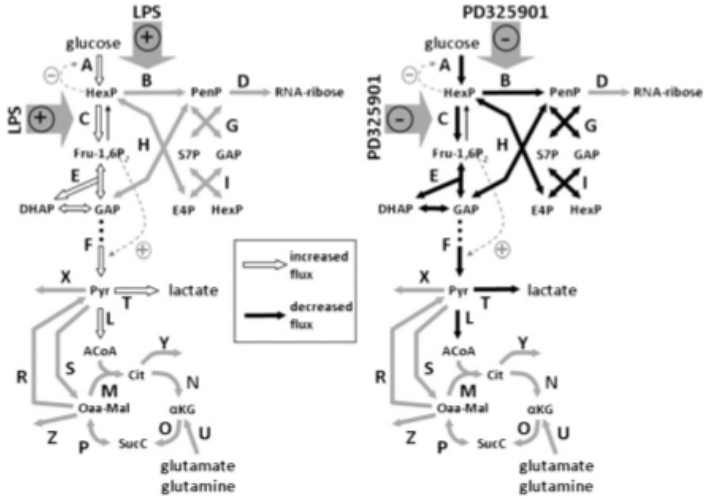



**FIGURE 5.** Cross-talk between MEK/ERK and key aspects of macrophage metabolism. Gray arrows represent the proposed activities that are regulated by signal transduction throughout MEK/ERK after incubation with PD325901 (*right panel*) or LPS (*left panel*). Positive (+) or negative (−) symbols predict activation or inhibition, respectively. ACoA, acetyl-CoA; Cit, citrate; DHAP, dihydroxyacetonephosphate; E4P, erythrose-4-phosphate; GAP, Fru-1,6$P_2$, glyceraldehyde-3-phosphate; HexP, hexose phosphates; αKG, 2-oxoglutarate; PenP, pentose phosphates; S7P, sedoheptulose-7-phosphate; SucC, succinyl-CoA.

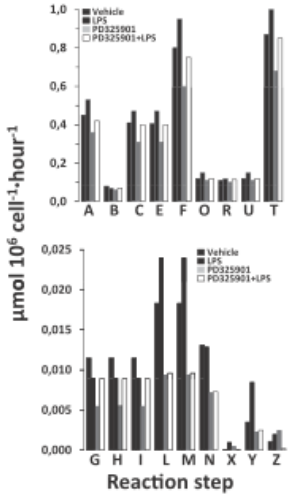



**FIGURE 6.** Metabolic fluxes in RAW 264.7 cells. Estimation of internal fluxes based on measured [$^{13}C$] redistribution. Bars are the median of the best 20 flux distributions corresponding to vehicle, LPS, PD325901, and PD325901 + LPS. Letters correspond to the reaction steps in the network schemes in Fig. 5.

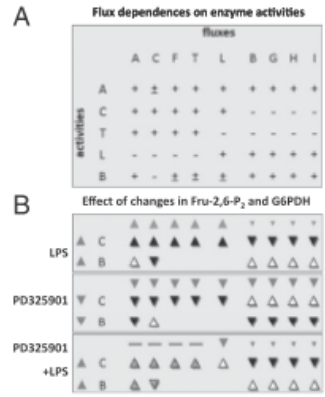



**FIGURE 7.** Flux dependencies on enzyme activities. Flux dependencies on enzyme activities (*A*) and compatibility of changes in fluxes with changes in enzyme activities (C, PFK-1; B, G6PDH+6PGDH) (*B*). "+," "−," and "±" predict the direction of changes in fluxes with respect to the direction of changes in enzyme activities: +, same direction; −, opposite direction; ±, indeterminate. Increase-decrease symbols ("▲", "▼", "–") refer to an increase or decrease in enzyme activity or flux: ▲, increase; ▼, decrease; –, no change. In *B*, gray symbols refer to observed changes in fluxes or activities. Black or white symbols identify compatibility in the direction of changes in fluxes and activities with the predicted dependencies in *A*: black, compatible or satisfied prediction; white, noncompatible or nonsatisfied prediction; filled, only compatible with very small (not observed) changes in fluxes.

of the regulatory dependencies (positive for enzyme-substrate dependencies and activations, and negative for inhibitions), dependencies among specific activities and the flux through a specific reaction depend on the relative magnitudes of the regulatory dependencies, which are unknown. However, some of these dependencies can be mainly positive or negative (23). A positive dependency indicates that a change in the enzyme activity is compatible or predicts a change in the flux that follows the same direction, irrespective of the magnitude of the regulatory dependencies. This means that an increase in the activity will induce an increase in the flux, whereas decreasing the activity will also decrease the flux. In contrast, a negative dependency indicates that changes in the activity will induce an inverse effect on the changes in the flux. Fig. 7*A* shows some of these sign-fixed dependencies for the main glycolytic and PPP fluxes with respect to changes in the activities of glucose uptake + hexokinase (reaction step A in Fig. 5), PFK-1 (reaction step C in Fig. 5), lactate dehydrogenase + lactate exchange (reaction step T in Fig. 5), PDH (reaction step L in Fig. 5), and G6PDH+6PGDH (reaction step B in Fig. 5).

The analysis of the compatibility of the measured changes in enzyme activities in the context of topology, stoichiometry, fluxes, and regulations affecting the central carbon metabolism provides fundamental information for interpreting the effects of LPS stimulation and PD325901 inhibition. Changes in glycolytic activity by regulating PFK-1 activity (reaction step C) are expected as a consequence of the changes in basal levels of Fru-2,6-$P_2$ (Fig. 1*D*), which is a potent activator of the glycolytic flux. More modest changes in G6PDH and 6PGDH activities (reaction step B) were recorded (Fig. 3*C*). Fig. 7*B* shows the compatibility of the direction of these changes in enzyme activities and the direction of changes in fluxes. In cells treated with PD325901, the decrease in PFK-1 (reaction step C) activity alone explains the decrease in the glycolytic fluxes (reaction steps A, C, F, T, and L) but not the changes in PPP fluxes (reaction steps B, G, H, and I). In contrast, a decrease in the activities of G6PDH+6PGDH (reaction step B) alone explains the observed changes in PPP fluxes but not all of the changes observed in glycolytic fluxes. Interestingly, this showed that changes in PFK-1 and G6PDH+6PGDH occur simultaneously, as has been experimentally observed, and could explain the changes in both glycolytic and PPP fluxes. In cells treated with LPS, the strong PFK-1 activation that follows the high levels of Fru-2,6-$P_2$ observed could qualitatively explain all of the changes in glycolytic and PPP fluxes. An increase in the activities of G6PDH+6PGDH alone will result in an increase in PPP fluxes, but this was not observed, given that the high levels of Fru-2,6-$P_2$ favored PFK-1 activation in the resulting flux profile. When cells were treated simultaneously with PD325901 and LPS, the slight increase in Fru-2,6-$P_2$ was not sufficient to activate the glycolytic flux profile characteristic of PFK-1 activation. The slight decrease in the PPP fluxes observed can be explained by the combined effect of changes in both PFK-1 and G6PDH+6PGDH activities.

## Discussion

A detailed [1,2-[$^{13}C$]$_2$]glucose tracer-based metabolomics approach, together with measured changes in glucose and glutamine consumption and lactate production, was used to characterize the effects of MEK/ERK inhibition on the basic metabolic response to LPS stimulation in macrophages. One of our previous studies showed that classic versus alternative macrophage activation involved the expression of specific sets of metabolic enzymes intended to cope with the energy demands of the activated cells (14). However, the finding that a single hit (i.e., MEK inhibition) might influence the LPS response in metabolic terms offers a new view on the cross-talk between cell activation and basic energy metabolism. Moreover, these effects on MEK/ERK inhibition were also observed in cultured peritoneal macrophages and in human monocytes differentiated to macrophages (21, 24). From a bioenergetics point of view, macrophages are essentially glycolytic cells (16, 33, 34) using anaerobic glycolysis to metabolize glucose. One of the regulators of glucose metabolism in macrophages is the increase in Fru-2,6-$P_2$ levels, which activates the flux through

PFK-1 (14, 35). In many glycolytic cells, Fru-2,6-$P_2$ levels are tightly regulated through balancing PFK-2/FBPase-2 activities. Four genes encode the PFK-2/FBPase-2 in mammals. The L-type is encoded by the *PFKB1* gene and is mainly expressed in the liver and muscle. The uPFK-2 is encoded by the *PFKB3* gene and has a predominantly kinase activity, with lower bisphosphatase activity. This gene is induced by hypoxia and regulated by phosphorylation, playing a role in the high glycolytic rate of various cell types, such as cancer cells (35, 36). In macrophages, innate and classic activation, but not the alternative IL-4/IL-13 stimulation, switches the expression of the PFK-2/FBPase-2 isoform from *PFKB1* prevailing in resting cells to *PFKB3*, resulting in an increase in Fru-2,6-$P_2$ levels and glycolytic flux (14). Interestingly, MEK/ERK inhibition impaired the LPS-dependent expression of uPFK-2, thus decreasing Fru-2,6-$P_2$ levels, PFK-2 activity, and, as expected, glucose consumption and lactate production but without changes in glutamine/glutamate consumption. The ability of the MEK/ERK pathway to prevent the switch from L–PFK-2 to uPFK-2 in response to LPS was unexpected and revealed fine tuning of macrophage activation. Other changes induced by LPS, such as a decrease in PPP fluxes, were not affected by PD325901. Indeed, using the same approach, a selective p38 inhibitor (12) did not interfere with the LPS enhancement of glycolytic flux, including the increase in uPFK-2/Fru-2,6-$P_2$ levels. However, the lack of a JNK inhibitor preserving cell viability complicates this study in these cells. Even though, analysis of lactate release and uPFK-2/Fru-2,6-$P_2$ levels in cells treated with BI78D3 and activated with LPS suggests a minor (if any) effect of JNK inhibition on carbon metabolism in RAW 264.7 cells.

The cross-talk between MEK/ERK and central carbon metabolism is summarized in Fig. 5. From an analytical point of view, macrophage activation with LPS is characterized by enhanced flux through PFK-1, via a Fru-2,6-$P_2$ increase, and explains the increases in the glycolytic pathway and the decrease in the reactions in the PPP. However, the transient (peak at 8h), but statistically significant, increase in activity through the G6PDH+6PGDH block should lead to changes in the opposite direction, which are likely to mediate the decrease in fluxes throughout the PPP via increased PFK-1 activity. Interestingly, the flux profile changed following PD325901 inhibition, with or without LPS, and could not be explained by the change in PFK-1 alone. Changes in both PFK-1 and in G6PDH+6PGDH are required to explain the observed flux profile. Indeed, an additional regulator of the cross-talk at the Fru-2,6-$P_2$ level is the expression of TIGAR, a p53-inducible enzyme that hydrolyzes Fru-2,6-$P_2$ to fructose-6-phosphate (37, 38). We investigated whether TIGAR was regulated by p53 levels in macrophages. However, p53 was only upregulated at the end of the activation process (data not shown), when there was a large increase in the synthesis of ROS and reactive nitrogen species. Interestingly, MEK/ERK inhibition decreased ROS production by LPS-activated macrophages, confirming an interference of this MAPK on the LPS-dependent activation program of the macrophage. However, at the same time, MEK/ERK inhibition moderately enhanced (8 h) or maintained (18 h) the metabolic flux through the G6PDH pathway, excluding a sequential dependence of these pathways during activation. In agreement with these results, p66Shc-deficient mice, which exhibit an attenuated ROS synthesis due to a defect in the activation of the NADPH oxidase complex, also exhibit a marked reduction in ERK activation (39).

Finally, ERK1/2 activation in macrophages under proinflammatory conditions has been associated with different pathophysiological situations, ranging from cancer to insulin resistance. For example, macrophage infiltration increases during tumor progression in mouse models of lung cancer, but the combined inhibition of MEK and PI3K ablated macrophage-mediated increases in epithelial growth, enhancing animal survival (40); in contrast, it was shown that the proinflammatory cytokine IL-1β reduces insulin receptor substrate 1 expression and prevents Akt activation, leading to insulin resistance through a mechanism that is partly mediated by ERK activation (41–43). Therefore, ERK1/2 regulation appears to be an important mediator of macrophage function.

In summary, the presented quantitative analysis revealed many more details about the metabolic effects of the signaling regulators studied, showing that the exploration of metabolic effects provides important details that cannot be shown by only qualitative analysis of experimental data. Our work is an example of quantitative analysis of the cross-talk between signal transduction and metabolism in RAW 264.7 cells.

## Acknowledgments

We thank Verónica Terrón for technical help.

## Disclosures

The authors have no financial conflicts of interest.

## References

1. Gordon, S., and F. O. Martinez. 2010. Alternative activation of macrophages: mechanism and functions. *Immunity* 32: 593–604.
2. Gordon, S. 2007. The macrophage: past, present and future. *Eur J. Immunol.* 37 (Suppl. 1): S9–S17.
3. Nathan, C. 2002. Points of control in inflammation. *Nature* 420: 846–852.
4. Hu, X., S. D. Chakravarty, and L. B. Ivashkiv. 2008. Regulation of interferon and Toll-like receptor signaling during macrophage activation by opposing feedforward and feedback inhibition mechanisms. *Immunol. Rev.* 226: 41–56.
5. Martinez, F. O., A. Sica, A. Mantovani, and M. Locati. 2008. Macrophage activation and polarization. *Front. Biosci.* 13: 453–461.
6. Pasare, C., and R. Medzhitov. 2004. Toll-like receptors: linking innate and adaptive immunity. *Microbes Infect.* 6: 1382–1387.
7. Mantovani, A., A. Sica, S. Sozzani, P. Allavena, A. Vecchi, and M. Locati. 2004. The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol.* 25: 677–686.
8. Nathan, C., and A. Ding. 2010. Nonresolving inflammation. *Cell* 140: 871–882.
9. Martinez, F. O., S. Gordon, M. Locati, and A. Mantovani. 2006. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. *J. Immunol.* 177: 7303–7311.
10. Rao, K. M. 2001. MAP kinase activation in macrophages. *J. Leukoc. Biol.* 69: 3–10.
11. Rao, K. M., T. Meighan, and L. Bowman. 2002. Role of mitogen-activated protein kinase activation in the production of inflammatory mediators: differences between primary rat alveolar macrophages and macrophage cell lines. *J. Toxicol. Environ. Health A* 65: 757–768.
12. Bain, J., L. Plater, M. Elliott, N. Shpiro, C. J. Hastie, H. McLauchlan, I. Klevernic, J. S. Arthur, D. R. Alessi, and P. Cohen. 2007. The selectivity of protein kinase inhibitors: a further update. *Biochem. J.* 408: 297–315.
13. Stebbins, J. L., S. K. De, T. Machleidt, B. Becattini, J. Vazquez, C. Kuntzen, L. H. Chen, J. F. Cellitti, M. Riel-Mehan, A. Emdadi, et al. 2008. Identification of a new JNK inhibitor targeting the JNK-JIP interaction site. *Proc. Natl. Acad. Sci. USA* 105: 16809–16813.
14. Rodríguez-Prados, J. C., P. G. Través, J. Cuenca, D. Rico, J. Aragonés, P. Martín-Sanz, M. Cascante, and L. Boscá. 2010. Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. *J. Immunol.* 185: 605–614.
15. Odegaard, J. I., R. R. Ricardo-Gonzalez, A. Red Eagle, D. Vats, C. R. Morel, M. H. Goforth, V. Subramanian, L. Mukundan, A. W. Ferrante, and A. Chawla. 2008. Alternative M2 activation of Kupffer cells by PPARdelta ameliorates obesity-induced insulin resistance. *Cell Metab.* 7: 496–507.
16. Wu, G. Y., C. J. Field, and E. B. Marliss. 1991. Glucose and glutamine metabolism in rat macrophages: enhanced glycolysis and unaltered glutaminolysis in spontaneously diabetic BB rats. *Biochim. Biophys. Acta* 1115: 166–173.
17. Rosa, L. F., Y. Cury, and R. Curi. 1992. Effects of insulin, glucocorticoids and thyroid hormones on the activities of key enzymes of glycolysis, glutaminolysis, the pentose-phosphate pathway and the Krebs cycle in rat macrophages. *J. Endocrinol.* 135: 213–219.
18. Zeyda, M., and T. M. Stulnig. 2007. Adipose tissue macrophages. *Immunol. Lett.* 112: 61–67.
19. Marin, S., W. N. Lee, S. Bassilian, S. Lim, L. G. Boros, J. J. Centelles, J. M. FernAndez-Novell, J. J. Guinovart, and M. Cascante. 2004. Dynamic profiling of the glucose metabolic network in fasted rat hepatocytes using [1,2-13C2]glucose. *Biochem. J.* 381: 287–294.
20. Ramos-Montoya, A., W. N. Lee, S. Bassilian, S. Lim, R. V. Trebukhina, M. V. Kazhyna, C. J. Ciudad, V. Noé, J. J. Centelles, and M. Cascante. 2006.

Pentose phosphate cycle oxidative and nonoxidative balance: A new vulnerable target for overcoming drug resistance in cancer. *Int. J. Cancer* 119: 2733–2741.
21. Prieto, P., J. Cuenca, P. G. Través, M. Fernández-Velasco, P. Martín-Sanz, and L. Boscá. 2010. Lipoxin A4 impairment of apoptotic signaling in macrophages: implication of the PI3K/Akt and the ERK/Nrf-2 defense pathways. *Cell Death Differ.* 17: 1179–1188.
22. Martin-Sanz, P., M. Cascales, and L. Boscá. 1989. Glucagon-induced changes in fructose 2,6-bisphosphate and 6-phosphofructo-2-kinase in cultured rat foetal hepatocytes. *Biochem. J.* 257: 795–799.
23. de Atauri, P., A. Benito, P. Vizán, M. Zanuy, R. Mangues, S. Marín, and M. Cascante. 2011. Carbon metabolism and the sign of control coefficients in metabolic adaptations underlying K-ras transformation. *Biochim. Biophys. Acta* 1807: 746–754.
24. Traves, P. G., S. Hortelano, M. Zeini, T. H. Chao, T. Lam, S. T. Neuteboom, E. A. Theodorakis, M. A. Palladino, A. Castrillo, and L. Bosca. 2007. Selective activation of liver X receptors by acanthoic acid-related diterpenes. *Mol. Pharmacol.* 71: 1545–1553.
25. Lee, W. N., L. O. Byerley, E. A. Bergner, and J. Edmond. 1991. Mass isotopomer analysis: theoretical and practical considerations. *Biol. Mass Spectrom.* 20: 451–458.
26. Schmidt, K., M. Carlsen, J. Nielsen, and J. Villadsen. 1997. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.* 55: 831–840.
27. Cascante, M., and S. Marin. 2008. Metabolomics and fluxomics approaches. *Essays Biochem.* 45: 67–81.
28. Sauer, U. 2006. Metabolic networks in motion: 13C-based flux analysis. *Mol. Syst. Biol.* 2: 62.
29. Wiechert, W., M. Möllney, S. Petersen, and A. A. de Graaf. 2001. A universal framework for 13C metabolic flux analysis. *Metab. Eng.* 3: 265–283.
30. Wiechert, W., M. Möllney, N. Isermann, M. Wurzel, and A. A. de Graaf. 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioeng.* 66: 69–85.
31. Wiechert, W. 2007. The thermodynamic meaning of metabolic exchange fluxes. *Biophys. J.* 93: 2255–2264.
32. Feng, G. J., H. S. Goodridge, M. M. Harnett, X. Q. Wei, A. V. Nikolaev, A. P. Higson, and F. Y. Liew. 1999. Extracellular signal-related kinase (ERK) and p38 mitogen-activated protein (MAP) kinases differentially regulate the lipopolysaccharide-mediated induction of inducible nitric oxide synthase and IL-12 in macrophages: *Leishmania* phosphoglycans subvert macrophage IL-12 production by targeting ERK MAP kinase. *J. Immunol.* 163: 6403–6412.
33. Bustos, R., and F. Sobrino. 1992. Stimulation of glycolysis as an activation signal in rat peritoneal macrophages. Effect of glucocorticoids on this process. *Biochem. J.* 282: 299–303.
34. Newsholme, P., S. Gordon, and E. A. Newsholme. 1987. Rates of utilization and fates of glucose, glutamine, pyruvate, fatty acids and ketone bodies by mouse macrophages. *Biochem. J.* 242: 631–636.
35. Bando, H., T. Atsumi, T. Nishio, H. Niwa, S. Mishima, C. Shimizu, N. Yoshioka, R. Bucala, and T. Koike. 2005. Phosphorylation of the 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase/PFKFB3 family of glycolytic regulators in human cancer. *Clin. Cancer Res.* 11: 5784–5792.
36. Calvo, M. N., R. Bartrons, E. Castaño, J. C. Perales, A. Navarro-Sabaté, and A. Manzano. 2006. PFKFB3 gene silencing decreases glycolysis, induces cell-cycle delay and inhibits anchorage-independent growth in HeLa cells. *FEBS Lett.* 580: 3308–3314.
37. Bensaad, K., E. C. Cheung, and K. H. Vousden. 2009. Modulation of intracellular ROS levels by TIGAR controls autophagy. *EMBO J.* 28: 3015–3026.
38. Li, H., and G. Jogl. 2009. Structural and biochemical studies of TIGAR (TP53-induced glycolysis and apoptosis regulator). *J. Biol. Chem.* 284: 1748–1754.
39. Tomilov, A. A., V. Bicocca, R. A. Schoenfeld, M. Giorgio, E. Migliaccio, J. J. Ramsey, K. Hagopian, P. G. Pelicci, and G. A. Cortopassi. 2010. Decreased superoxide production in macrophages of long-lived p66Shc knock-out mice. *J. Biol. Chem.* 285: 1153–1165.
40. Fritz, J. M., L. D. Dwyer-Nield, and A. M. Malkinson. 2011. Stimulation of neoplastic mouse lung cell proliferation by alveolar macrophage-derived, insulin-like growth factor-1 can be blocked by inhibiting MEK and PI3K activation. *Mol. Cancer* 10: 76.
41. Barbarroja, N., R. López-Pedrera, M. D. Mayas, E. García-Fuentes, L. Garrido-Sánchez, M. Macías-González, R. El Bekay, A. Vidal-Puig, and F. J. Tinahones. 2010. The obese healthy paradox: is inflammation the answer? *Biochem. J.* 430: 141–149.
42. Jager, J., T. Grémeaux, M. Cormont, Y. Le Marchand-Brustel, and J. F. Tanti. 2007. Interleukin-1beta-induced insulin resistance in adipocytes through down-regulation of insulin receptor substrate-1 expression. *Endocrinology* 148: 241–251.
43. Kopp, A., C. Buechler, M. Bala, M. Neumeier, J. Schölmerich, and A. Schäffler. 2010. Toll-like receptor ligands cause proinflammatory and prodiabetic activation of adipocytes via phosphorylation of extracellular signal-regulated kinase and c-Jun N-terminal kinase but not interferon regulatory factor-3. *Endocrinology* 151: 1097–1108.

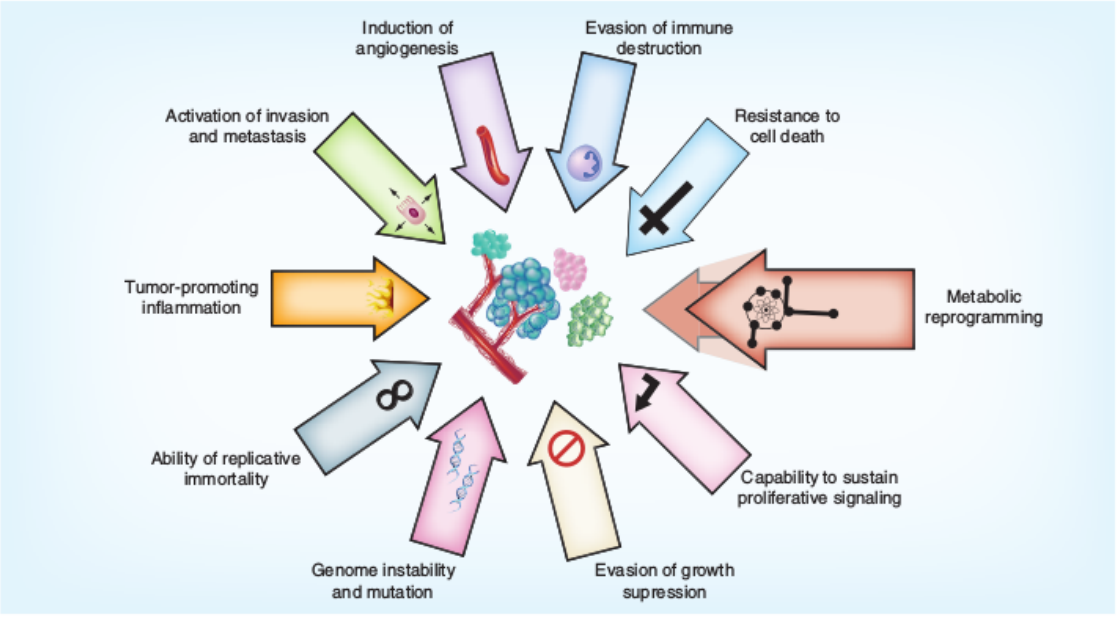



**Figure 1. Hallmarks of cancer.** The hallmarks of cancer comprise ten capabilities required during a multistep tumor pathogenesis to enable cancer cells to become tumorigenic and ultimately malignant. Metabolic reprogramming has been identified as an emerging hallmark and as a promising target for the treatment of cancer as there is a deregulation of bioenergetic controls and an abnormal use of metabolic pathways to sustain their biosynthetic and energetic needs.
Reproduced with permission from [2] © Elsevier.

developing new strategies and methods to drug and biomarker discovery, exploiting the reprogramming of metabolism that sustains cancer progression.

## Crosstalk between oncogenic signaling events & cancer cell metabolism

Through a better understanding of the complex networks of oncogenic signaling pathways, altered cellular metabolism emerges as one of the major routes through which oncogenes promote tumor formation and progression. Many key oncogenic signaling pathways converge to adapt tumor cell metabolism in order to support their growth and survival. The identification of new metabolic coordination mechanisms between altered metabolism and regulators of cell signaling networks, controlling both proliferation and survival, triggers the interest for new metabolism-based anticancer therapies. Several oncogenes, tumor suppressor genes and cell cycle regulators controlling cell proliferation and survival are intimately involved in modulating glycolysis, mitochondrial oxidative phosphorylation (OXPHOS), lipid metabolism, glutaminolysis and many other metabolic pathways (Figure 2). The accumulation of genetic abnormalities required for oncogenesis leads to changes in energetic and biosynthetic requirements that in turn affects the metabolic signature of cancer cells through interactions between enzymes, metabolites, transporters and regulators. High-throughput sequencing data reveals that the mutational events causing tumorigenesis are much more complex than previously thought and that the mutational range can vary even among tumors with identical histopathological features [5]. Some of the metabolic adaptations driven by oncogenic signaling events have been described as common to different tumors, but metabolic profiles can be significantly tissue/cell specific [6]. Here, we will highlight some of the most prevalent examples of crosstalks between oncogenic signaling events and pivotal metabolic pathways. *HIF-1* is a key regulator that initiates

**Key terms**

**Metabolic reprogramming:** Process in which the cellular metabolism evolves in order to adapt to new environmental conditions and perturbations. In the case of tumor, the energy metabolism is reprogrammed in order to sustain the high proliferative rate of cancer cells.

**Genome-scale metabolic models:** Those models that summarize and codify the information known about the metabolism of an organism based on the literature and databases. These models represent the metabolic reaction encoded by an organism's genome and can be transformed into a mathematical formulation in order to study the metabolic cell behavior.

# Apendix II

Review
Mini-Focus Issue: Systems biology



Future Medicinal Chemistry

# Cancer cell metabolism as new targets for novel designed therapies

Metabolic processes are altered in cancer cells, which obtain advantages from this metabolic reprogramming in terms of energy production and synthesis of biomolecules that sustain their uncontrolled proliferation. Due to the conceptual progresses in the last decade, metabolic reprogramming was recently included as one of the new hallmarks of cancer. The advent of high-throughput technologies to amass an abundance of omic data, together with the development of new computational methods that allow the integration and analysis of omic data by using genome-scale reconstructions of human metabolism, have increased and accelerated the discovery and development of anticancer drugs and tumor-specific metabolic biomarkers. Here we review and discuss the latest advances in the context of metabolic reprogramming and the future in cancer research.

Igor Marín de Mas[‡,1,2], Esther Aguilar[‡,1], Anusha Jayaraman[‡,1], Ibrahim H Polat[1], Alfonso Martín-Bernabé[1], Rohit Bharat[1], Carles Foguet[1], Enric Milà[1], Balázs Papp[2], Josep J Centelles[1] & Marta Cascante*[,1]

[1]Department of Biochemistry & Molecular Biology, Faculty of Biology, IBUB, Universitat de Barcelona & Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Unit Associated with CSIC, Diagonal 643, E-08028-Barcelona, Spain

[2]Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Temesvári krt. 62, H-6726 Szeged, Hungary

*Author for correspondence:
Tel.: +34 934021593
Fax: +34 934021559
martacascante@ub.edu

[‡]Authors contributed equally

Cancer is still one of the major causes of death worldwide and the statistics are devastating. According to the WHO the global burden of cancer has risen to 14.1 million new cases and 8.2 million cancer deaths in 2012 and the estimates predict that it could increase in its global incidence [1].

It was proposed 15 years ago by Hanahan and Weinberg that cancer development relies on the following basic biological capabilities, known as the 'hallmarks of cancer' that are acquired during the multistep process of tumor development: the capability to sustain proliferative signaling, resistance to cell death, evasion of growth suppression, ability of replicative immortality, tumor-promoting inflammation, genome instability and mutation, induction of angiogenesis and activation of invasion and metastasis. Owing to conceptual progress in the last decade, two new hallmarks, metabolic reprogramming and evasion of immune destruction, have been identified (Figure 1) [2].

Nowadays, it is widely recognized that metabolic reprogramming is essential to sustain tumor progression. Several metabolic adaptations described in cancer cells, such as the metabolization of glucose to lactate in the presence of oxygen (Warburg effect), are quite common among different cancer types. These changes are promoted by genetic and epigenetic alterations producing mutations or alterations in the expression of key metabolic enzymes that modify flux distributions in metabolic networks, providing advantages to cancer cells in terms of energy production and synthesis of biomolecules [3,4].

Understanding the mechanisms that trigger metabolic reprogramming in cancer cells and its role in tumoral progression is crucial, not only from a biological but also from a clinical stance, since this can be the basis towards improving existing cancer therapies or developing new ones.

In this review, we discuss the role of: the crosstalk between oncogenic signaling pathways and metabolism; the influence of non-genetic factors, such as tumor microenvironment, on metabolic reprogramming of cancer and stromal cells; the changes in isoenzymes patterns as potential therapeutic targets; and the new computational tools used by a systems biology approach in drug-target and biomarker discovery based on genome-scale metabolic models (GSMMs). Finally, we also discuss the future challenges in

FUTURE SCIENCE part of fsg

a coordinated transcriptional program activated by hypoxic stress (in response to low-oxygen conditions), to promote the metabolic shift from mitochondrial OXPHOS to glycolysis (Figure 2) through the induction of several genes, including glucose transporters and glycolytic enzymes, leading to an increased flux of glucose to lactate [7]. Additionally, *HIF-1* actively downregulates the OXPHOS flux by activation of PDK1, which inhibits the conversion of pyruvate to acetyl-CoA catalyzed by the tricarboxylic acid (TCA) cycle enzyme PDH.

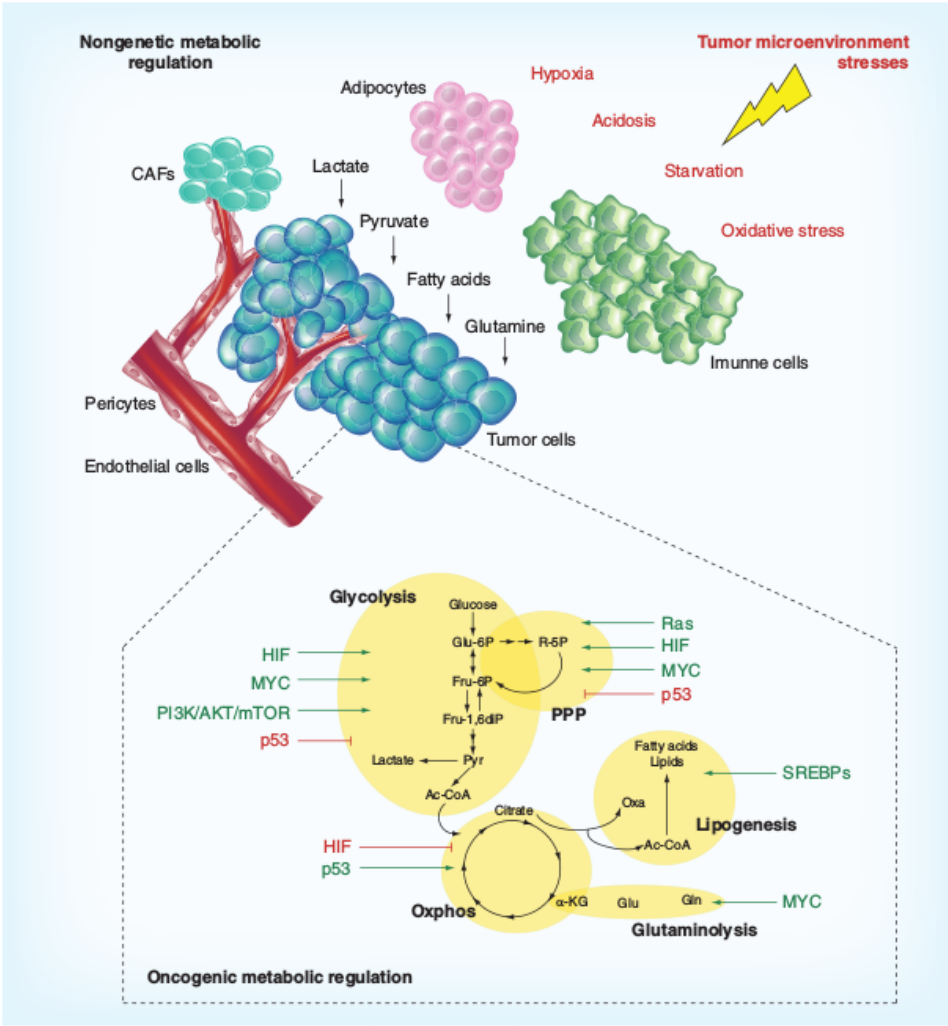



**Figure 2. Nongenetic and oncogenic influences on tumor metabolic reprogramming.** The nongenetic component (the tumor microenvironment) influences metabolic changes in tumor cells as a result of gradients of oxygenation and pH, nutrient availability, oxidative stress and the intercellular communication with stromal cells by means of metabolites such as lactate, pyruvate, fatty acids and glutamine. Combined with tumor microenvironment, the genetic component (oncogenes and tumor suppressors) plays a key role in metabolic reprogramming to ensure metabolites are shunted into pathways that support the energetic requirements and the biosynthesis of structural components, achieved by maintaining high rates of glycolysis and/or glutaminolysis, promoting the pentose phosphate pathway, slowing mitochondrial metabolism (oxidative phosphorylation) and utilizing tricarboxylic acid intermediates for biosynthetic precursors (e.g., fatty acids and lipids).
CAF: Cancer-associated fibroblastic cell; PPP: Pentose phosphate pathway; SREBP: Sterol regulatory element binding protein.

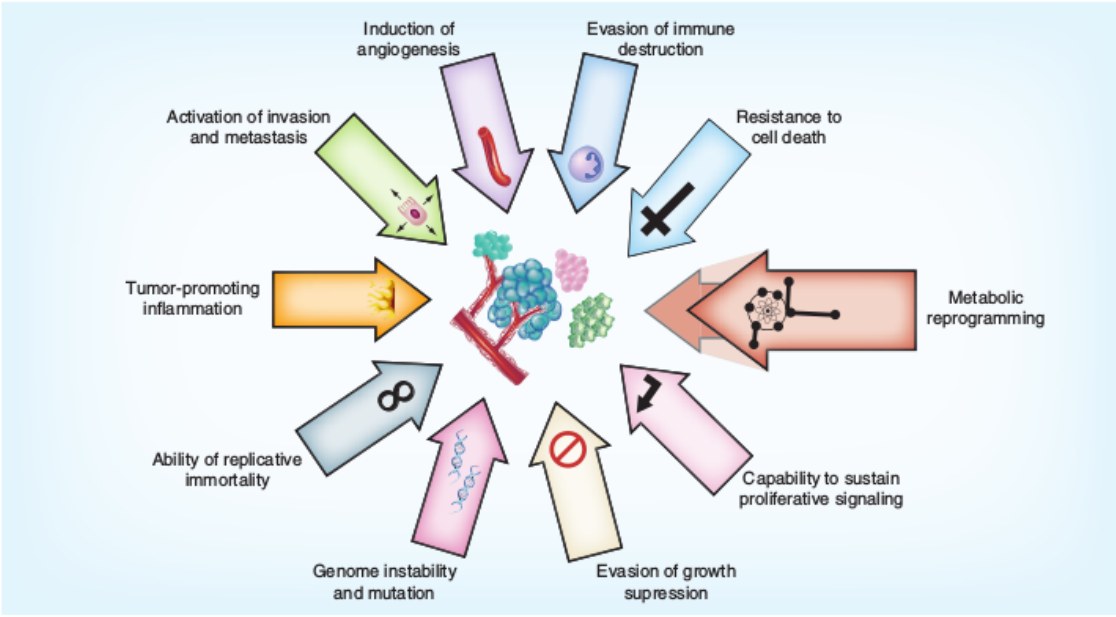



**Figure 1. Hallmarks of cancer.** The hallmarks of cancer comprise ten capabilities required during a multistep tumor pathogenesis to enable cancer cells to become tumorigenic and ultimately malignant. Metabolic reprogramming has been identified as an emerging hallmark and as a promising target for the treatment of cancer as there is a deregulation of bioenergetic controls and an abnormal use of metabolic pathways to sustain their biosynthetic and energetic needs.
Reproduced with permission from [2] © Elsevier.

developing new strategies and methods to drug and biomarker discovery, exploiting the reprogramming of metabolism that sustains cancer progression.

## Crosstalk between oncogenic signaling events & cancer cell metabolism

Through a better understanding of the complex networks of oncogenic signaling pathways, altered cellular metabolism emerges as one of the major routes through which oncogenes promote tumor formation and progression. Many key oncogenic signaling pathways converge to adapt tumor cell metabolism in order to support their growth and survival. The identification of new metabolic coordination mechanisms between altered metabolism and regulators of cell signaling networks, controlling both proliferation and survival, triggers the interest for new metabolism-based anticancer therapies. Several oncogenes, tumor suppressor genes and cell cycle regulators controlling cell proliferation and survival are intimately involved in modulating glycolysis, mitochondrial oxidative phosphorylation (OXPHOS), lipid metabolism, glutaminolysis and many other metabolic pathways (Figure 2). The accumulation of genetic abnormalities required for oncogenesis leads to changes in energetic and biosynthetic requirements that in turn affects the metabolic signature of cancer cells through interactions between enzymes, metabolites, transporters and regulators. High-throughput sequencing data reveals that the mutational events causing tumorigenesis are much more complex than previously thought and that the mutational range can vary even among tumors with identical histopathological features [5]. Some of the metabolic adaptations driven by oncogenic signaling events have been described as common to different tumors, but metabolic profiles can be significantly tissue/cell specific [6]. Here, we will highlight some of the most prevalent examples of crosstalks between oncogenic signaling events and pivotal metabolic pathways. *HIF-1* is a key regulator that initiates

**Key terms**

**Metabolic reprogramming:** Process in which the cellular metabolism evolves in order to adapt to new environmental conditions and perturbations. In the case of tumor, the energy metabolism is reprogrammed in order to sustain the high proliferative rate of cancer cells.

**Genome-scale metabolic models:** Those models that summarize and codify the information known about the metabolism of an organism based on the literature and databases. These models represent the metabolic reaction encoded by an organism's genome and can be transformed into a mathematical formulation in order to study the metabolic cell behavior.

Similar to *HIF-1*, oncogenic activation of *Myc* also triggers a transcriptional program that enhances glycolysis by directly inducing glucose transporters and glycolytic enzymes. Indeed, there is a crosstalk between *HIF-1* and *Myc*, whereby they cooperate to confer metabolic advantages to tumor cells by oxygen-dependent mechanisms, with a difference that, contrary to *HIF-1*, *Myc* upregulation has more significant consequences for many cells as it alters not only glycolysis but also glutaminolysis (Figure 2) and many other biosynthetic pathways [8]. The *Myc* oncogene stimulates glutamine uptake and glutaminolysis by inducing glutamine transporters directly and GLS, the enzyme that converts glutamine to glutamate, indirectly [9]. Besides glycolysis, glutaminolysis is another important metabolic pathway in cancer cells, which contributes not only as a source to replenish the TCA cycle, but also to control the redox potentials through generation of reductive equivalents, such as NADPH. In addition to glucose, a vast amount of glutamine is consumed by cancer cells. Glutamine is converted to glutamate and then to α-ketoglutarate (α-KG), which feeds the TCA cycle. Some tumors that show an upregulation of glutamine metabolism have been reported to exhibit 'glutamine addiction', that is, glutamine becomes essential during rapid growth. However, glutamine consumption and addiction are dependent on the metabolic profile of the cancer cells and in particular on the oncogene/tumor suppressor involved in tumor progression [10].

Activated PI3K/AKT/mTOR pathway is one of the most common signaling cascades altered in tumor cells and this pathway is one of the most heavily targeted to develop anticancer therapies. Many cancers are driven by aberrations in the PI3K/AKT/mTOR pathway promoting metabolic transformation through multiple metabolic pathways, including an increase in glucose and amino acid uptake (Figure 2), upregulation of glycolysis and lipogenesis and enhanced protein translation through Akt-dependent mTOR activation [11].

In cancer cells, the increased rate of *de novo* lipid biosynthesis is an important aspect of the metabolic reprogramming during oncogenesis. Lipid metabolism is regulated via activation of the sterol regulatory element binding proteins (SREBPs) (Figure 2), which are important regulators of the Akt/mTOR signaling pathway [12]. Indeed, various genes coding for enzymes involved in fatty acid and cholesterol biogenesis are targets of SREBPs, including ATP-citrate lyase, acetyl-CoA carboxylase and fatty acid synthase [13]. Lipogenesis is also controlled by the *RAS* oncogene through the action of *HIF-1*, which has been reported to induce the expression of fatty acid synthase in human breast cancer cell lines [14]. However, the *RAS* oncogene also modulates mitochondrial metabolism roughly increasing the activity of *Myc* and *HIF-1* [4], glycolysis and the pentose phosphate pathway (PPP) [15]. Proliferating cells, such as tumors, require high amounts of pentose phosphates for biosynthesis of macromolecules and NADPH for redox homeostasis maintenance [16]. Therefore, PPP plays a fundamental role in defining the metabolic phenotype of tumor cells. Hence, there are also examples of coordinated crosstalk between the main enzymes that control the PPP during oncogenesis and oncogenic signaling pathways. K-RAS and PI3K signaling have been shown to positively regulate G6PD, whereas p53, which is a transcription factor and regulator of the cell cycle and apoptosis, physically interacts with G6PD to negatively modulate its activity [17], and thereby downregulates PPP. On the other hand, active *HIF-1* signaling has been linked to both TKT and TKTL1, the enzymes catalyzing the rate-limiting step of the non-oxidative branch of the PPP [18].

In addition, alterations in *p53* are frequent events in tumorigenesis. The loss or inactivation of *p53* downregulates OXPHOS by inducing aerobic glycolysis through inhibiting glucose transporters and the glycolytic enzyme PGM and inducing TP53-induced glycolysis and apoptosis regulator, a negative regulator of glycolysis [19]. On the other hand PHF20 stabilizes and upregulates p53 resulting in a gain of functionality that drives the reprogramming of the metabolism of certain cancers cell lines, such as U87 (glioblastoma) or MCF7 (breast cancer) [20].

Other examples of oncogene-mediated metabolic reprogramming include mutations in genes encoding FH and succinate dehydrogenase, which are loss-of-function mutations and behave as tumor suppressor genes [21]. On the other hand, mutations in IDH-1 and IDH-2, do not result in inactivation of normal IDH enzymatic function but generation of novel gain-of-function mutation that enables the conversion of α-KG to D2-HG, which may act as an 'oncometabolite' by inhibiting multiple α-KG–dependent dioxygenases involved in epigenetic regulation [22].

Tumorigenesis occurs as a consequence, not only of the dysregulation of numerous oncogenic pathways, but also due to many nongenetic factors, including tumor microenvironment stresses, such as hypoxia, lactic acidosis and nutrient deprivation. The integration of these nongenetic factors within the genetic framework of cancer is the next logical step in understanding

**Key term**

**Tumor heterogeneity:** Variability among different tumors in the same organ (intertumoral heterogeneity) or the variability among cells in a tumor (intratumoral heterogeneity).

**tumor heterogeneity.** Research over the years has elucidated the cellular and molecular interactions (including metabolic reprogramming) occurring in the tumor microenvironment and are closely linked to the processes of angiogenesis and metastasis.

## Tumor microenvironment

Since the discovery of immune cells in tumor samples by Rudolf Virchow in 1863, various studies have shown the linkage of cancer to inflammation, vascularization and other conditions, which suggest that tumors do not act alone. Without its 'neighborhood' the survival of tumor cells could be a big question mark. The cellular heterogeneity in this microenvironment is complex and comprises of extracellular matrix, tumor cells and non-transformed normal cell types that co-evolve with the tumor cells (e.g., cancer-associated fibroblastic cells [CAFs], infiltrating immune cells and endothelial cells that constitute the tumor-associated vasculature) that are embedded within this matrix and nourished by the vascular network. In addition, there are many signaling molecules and chemicals, such as oxygen and protons, all of which can influence tumor cell proliferation, survival, invasion, metastasis and energy metabolism reprogramming. CAFs, one of the most abundant stromal cell types in different carcinomas, are activated fibroblasts that share similarities with fibroblasts, stimulated by inflammatory conditions or activated during wound healing. But, instead of suppressing tumor formation, CAFs can significantly promote tumorigenesis, invasion and *de novo* cancer initiation by some unique growth factors and cytokines secretion (e.g., EFG, FGF, IL6, IL8, VEGF etc), extensive tissue remodeling mediated by augmented expression of proteolytic enzymes (e.g., matrix metalloproteinases), deposition of extracellular matrix and pathogenic angiogenesis by liberating pro-angiogenic factors within the matrix [23]. Significant cell plasticity exists within this cell population, as both mesenchymal-to-epithelial and epithelial-to-mesenchymal transitions are known to occur, further enhancing stromal heterogeneity. Moreover, CAFs can enhance proliferation and invasion by inducing the epithelial-to-mesenchymal transitions on tumor cells [24,25]. Immune cell recruitment and localization in the tumor milieu vary widely in the lesions. Heterogeneity of tumor immune contexture is influenced by various factors, including those secreted by CAFs, the extension and permeability of the vasculature, and the tumor cells themselves. Importantly, macrophages comprise the most abundant immune population in the tumor microenvironment and are responsible for the production of cytokines, chemokines, growth factors, proteases and toxic intermediates, such as nitric oxide and reactive oxygen species [26]. Their contribution to tumor initiation, progression and metastasis can be attenuated by antioxidant treatments, such as butylated hydroxyanisole, as reactive oxygen species levels have been reported to regulate the differentiation and polarization state of macrophages. Endothelial cells that are 'hijacked' by the tumors play an important part in forming a transport system, although ineffective, but essential for its survival and growth. In addition, blood vessel formation needs a protein matrix for the endothelial cells to be attached to and also it needs pericytic cells to strengthen these vessels. But, since the pericytes are not known to function very well in tumor vessel formation, the vessels are always malformed and leaky [27].

In the last few years the concept of cancer stem cells (CSC), a small minority of cells in the tumor, has evolved to be a possible cause and source of tumor heterogeneity. Currently there are two models that describe tumor cell heterogeneity: the hierarchical CSC model, where self-renewing CSCs sustain the stem cell population while giving rise to progenitor cells that are not capable of self-renewal and can give rise to differentiating clones that contribute to overall tumor heterogeneity, and the stochastic (tumor microenvironment-driven). model in which cancer cells are clonally evolved, and virtually every single cell can self-renew and propagate tumors. In this model, the self-renewal capability of each cell is determined by distinct signals from the tumor microenvironment. Recent studies have suggested that tumor heterogeneity may exist in a model coordinating with both the CSC and the stochastic concepts [28].

## Metabolic reprogramming associated with cancer & stromal cell interaction

Recently, the relationship between tumor microenvironment and metabolic reprogramming has been highlighted and there has been extensive research about metabolic symbiosis between cancer and stromal cells. Among these interactions, it was shown that epithelial tumor cells induce oxidative stress in the normal stroma, inducing aerobic glycolysis in CAFs, as well as changes in inflammation, autophagy and mitophagy (Figure 2). As a consequence of this rewiring in CAFs metabolism, energy-rich metabolites (such as lactate, pyruvate and ketones) are secreted, feeding adjacent cancer cells. This tumor–stroma metabolic relationship is referred to as the 'reverse Warburg effect'. CSCs that are present within the tumor also rely more heavily on glycolysis, even in the presence of oxygen (Warburg effect), and decrease their mitochondrial activity in order to limit reactive oxygen species production. As these glycolytic and mitochondrial signatures help to maintain the CSC phenotype, recent studies have

focused their attention to these metabolic weaknesses to be combined with traditional chemotherapy that, alone, usually fails to target CSCs [29,30]. In addition, other stromal cells, such as adipocytes, are able to act as energy sources, transferring fatty acids that come from lipolysis to ovarian tumor cells for β-oxidation [31]. Deregulated lipogenesis has been shown to play an important role in the interactions between cancer cells and the surrounding stromal cells. Studies suggest that it affects the epithelial cell polarity during the early stages of cancer development [32], inducing cancer cell migration [33] and activation of angiogenesis involving signaling lipids (e.g., diacyl glycerides, lysophosphatidic acid and prostaglandins), fatty acid synthesis enzymes and overof the monoglyceride-lipase [34–36].

Loss of stromal caveolin-1 in CAFs has been associated with tumor progression and metastasis [37] and causes oxidative stress and induction of autophagy, which results in increased levels of glutamine and ammonia in the stromal microenvironment. This glutamine could be consumed by cancer cells for energy and anaplerotic reactions and ammonia acts as a potent inducer of autophagy, creating a vicious cycle [37]. The migration stimulating factor, a truncated isoform of fibronectin identified to be overexpressed by CAFs and other 'activated' fibroblasts, has been shown to increase lactate production in the stromal environment and decrease mitochondrial activity, suggesting a shift towards glycolysis during hypoxia in addition to promoting tumor growth without affecting tumor angiogenesis [38].

Angiogenesis has been long known to play a major role in supporting cancer cell growth in the tumor microenvironment. But since the newly formed blood vessels are mostly defective there is always a nutrition deficiency and acidosis in these areas (Figure 2). A biomarker study in the gastric cancer environment where a quantitative analysis of the organic acids that are the end products of metabolism, using GC–MS, showed an increase in glycolytic end-products, such as pyruvic and lactic acids, with respect to normal tissues [39]. The pattern of high acidification in the tumor microenvironment due to the accumulation of glycolytic end-products results in a nutrient-deficient environment. In addition, metabolic reprogramming of tumor-associated endothelial cells has been showing up wide interests. Upon tumor angiogenic activation, endothelial cells are pushed to a state of metabolic stress for increasing their proliferation rate to form new blood vessels, although the resulting network is abnormal and inefficient. These normal cells show higher glycolytic enzyme activities and lactate production, even in the presence of oxygen [40], and they continue proliferating even in the presence of hostile conditions and high nutrient deficiency [41]. Also it has been shown that endothelial cells, similar to tumor cells, have a high expression of monocarboxylate transporter 1 required for the lactate influx, revealing that these cells seek alternative metabolites in a nutrition-deficient environment [42]. Moreover, the inhibition of glycogenolysis in human umbilical vein endothelial cells has been shown to decrease cell viability and migration, elucidating the importance of glycogen for the survival of these cells [43]. The role of the PPP in cell viability has also been demonstrated, in that, the direct inhibition of G6PDH has been shown to decrease endothelial cell survival [43]. When tumor cells choose the less energy-efficient metabolic pathways, such as glycolysis and glutaminolysis, both leading to the production of lactic acid, the pH of the tumor microenvironment decreases. It has been shown that endothelial cells behave in a similar fashion while forming new tumor blood vessels. While this phenomenon is known, it has also been found that the decrease in pH in the surrounding microenvironment actually increases cancer survival by immune suppression. Loss of T-cell function has been reported under low pH environment, while restoring the pH to normal conditions has been found to restore T-cell function [44]. Similarly, the lactic acid generated has shown to increase the proliferation of endothelial cells by increased interleukin8/CXCL8 production [41,45]. From a therapeutic point of view, targeting the altered metabolic pathways leading to lactic acid accumulation in tumor microenvironment could inhibit tumor growth as this mechanism would restore the impaired immune response and also a combinatorial therapy with antiangiogenesis drugs could reduce the proliferation of endothelial cells and formation of new blood vessels [46].

An important event that occurs during the changes in tumor microenvironment, as the cancer progresses, is the metastasis of some selected cancer cells to distant sites. A receptive microenvironment is required for tumor cells to engraft distant tissues and metastasize. Although several studies have indicated the formation of a premetastatic niche in the secondary sites before the primary tumor metastasizes [47], we have to consider how metastatic cells are able to adapt to their new metabolic environment, which can differ to a greater or lesser extent with respect to its nutrient and oxygen availability. Metastatic cells should exhibit a remarkable and dynamic flexibility that enables them to rapidly switch between metabolic states [48]. In addition, the homeostasis of the sites for metastasis can be disrupted as consequence of the metabolic activity of metastatic cells. This has been observed in bone, where metastatic prostate cancer cells secrete glutamate into their extracellular environment as a side effect of cel-

lular oxidative stress protection, promoting the development of pathological changes in bone turnover [49]. Further studies are required to analyze these metabolic interplays between metastatic cells and tumor microenvironment in order to obtain more specific treatments and therapies.

## Isoenzymes: therapeutic targets in cancer

The technological advances that have occurred over the past decade and the increasing number of evidences that have emerged from previous studies show a wide array of metabolic rewiring in cancer cells. Many metabolic enzymes that are specific to important metabolic pathways and those altered in cancer cells have been identified. These enzymes have a key role in mediating the aberrant metabolism of cancer cells and could serve as a promising source of novel drug targets. Isoforms of many of these metabolic enzymes are found to be specifically expressed in tumor cells affecting important pathways of the energetic metabolism. The current research is being refocused on specifically targeting these isoforms that has shown to be a promising strategy to develop new anticancer treatments. In this part, we will highlight some of the most important, altered pathways and the specific isoenzymes, that could be used for drug targeting, in cancer disease.

### Glycolytic isoenzymes

Glycolytic pathway serves as the principal energetic source for a cell. The higher dependency of cancer cells upon glycolytic metabolism for the production of ATP provides a greater motive to target glycolytic enzymes (Figure 2). Many isoforms of these enzymes have been found to be specifically expressed in tumor cells and are being exploited as potential candidates to be used as drug targets. The transport of glucose across the plasma membrane is regulated by various isoforms of glucose transporters (GLUT1–14 or SLC2A1–14). GLUT1, -3 and -4 are found to be expressed at higher levels in cancer [50]. GLUT3 and other transporters could be targeted by the use of specific antibodies or drugs, such as phloretin or ritonavir, causing the cells to starve by blocking their nutrient uptake through these transporters.

Another important metabolic enzyme of the glycolytic pathway is HK, which regulates the first rate-limiting step of glucose metabolism. Cancer cells are heavily dependent on HK isoforms, such as HK2 [51]. The specific expression of HK2 in adipose tissue and skeletal muscles provides an opportunity to target this enzyme without having the risk of affecting other tissues. Compounds such as methyl jasmonate isolated from plants have been shown to disrupt the association between mitochondria and HKs (HK1 and -2). involved in regulating apoptosis [52] and have shown to be lethal to cancer cells *in vitro* [53].

Recent publications suggest a key role of PK isoenzyme – PKM2 – in mediating the Warburg effect in cancer cells [54], proving its prospective as an enzymatic anticancer drug target. The enzyme activity of PKM2 is inhibited downstream of cellular growth signals [55]. Cell proliferation and aerobic glycolysis in tumors are greatly dependent on this ability to inhibit the activity of the PKM2 enzyme. Many approaches using small-molecule inhibitors and small-hairpin RNA-based inhibition of *PKM2* have been shown to cause cell death and slow down cell proliferation *in vitro* [54,56]. The PFKFB3 isoform is shown to be important in RAS-mediated tumors and inhibition of PFKFB3 by small-molecule inhibitors has been shown to have cytostatic effect on the growth of cancer [57]. Inhibition of LDHA using FX11 or oxamate has been shown to induce oxidative stress and cause cell death in cancer cells [58,59]. Targeting LDHA combined with NAMPT inhibitors has been shown to slow down tumor regression and thus making it a potential candidate for drug targets [59].

### TCA isoenzymes/mitochondrial complex

PDK phosphorylates PDH and inhibits the conversion of pyruvate to acetyl-CoA, a key metabolite in the TCA cycle (Figure 2). Isoenzyme PDK3 is induced by upregulation of HIF-1α under hypoxic conditions and results in cells undergoing glycolysis instead of TCA for energy production. Inhibition of PDK3 increases the susceptibility of tumor cells towards anticancer drugs and causes inhibition of hypoxia-induced glycolysis [60]. Thus PDK3 could be used as a drug target to overcome drug resistance and improve chemotherapy.

Isoforms of IDH1 and -2 are found to be mutated in glioma and acute myeloid leukemia [61,62]. Mutations in IDH1 and -2 result in the overexpression of both of these enzymes and the production of 2-HG, which inhibits α-KG-dependent dioxygenase enzymes. Association between high levels of 2-HG and tumorigenicity is yet to be established, but interestingly the levels of several TCA metabolites remain unaltered, suggesting an alternate pathway that could be acting in normalizing the metabolite levels in cells with IDH1 mutations.

### Isoenzymes of the PPP

Cancer cells are in a constant demand for greater amounts of purines and pyrimidines to maintain their high proliferative nature (Figure 2). The key enzyme for the oxidative PPP, the G6PDH enzyme, is overexpressed in certain types of cancers and it has been shown to transform fibroblasts and help in tumor cell

proliferation [63]. On the other hand, the overexpression of TKTL1 in many forms of cancer could increase the concentration of glyceraldehyde-3-phosphate and help in mediating the Warburg effect in cancer cells [64]. Combinatorial approach of targeting G6PDH and TKTL1 can help overcome drug resistance and may cause cell death [65].

### Targeting isoenzymes of glutamine metabolism

Recent findings that point to the use of glutamine as a carbon source for the TCA cycle [66] in cancer cells encouraged researchers to consider enzymes of glutamine metabolism as potential therapeutic targets. 6-diazo-5-oxo-L-norleucine- or bis-2-(5-phenylacetamido-1,2,4-thiadiazol-2-yl)ethyl sulphide-mediated inhibition of GLS or siRNA-induced silencing of GLS and GDH have been shown to inhibit the activation of mTORC1 [67]. Thus, combinatorial targeting of GLS and GDH along with chemotherapy may prove to be more effective in cancer treatment. The differential expression of these cancer-associated isoenzymes can be used as potential biomarkers for early cancer prognosis or as **enzymatic drug targets**. However, the role and importance of these mutations in the reprogramming of the energetic metabolism observed in cancer cells is not always obvious. This makes it extremely difficult to evaluate the effects of these mutations in the cancer metabolism qualitatively or quantitatively. Additionally, the effects of these isoenzymes on metabolism can be attenuated or enhanced by compensatory and regulatory mechanisms. Taking into account these rationales, the need for a tool that permits a holistic analysis of the metabolic system is essential, in order to qualitatively evaluate the effects of a single or combination of different mutations within the whole metabolic network system. In the last few years, genome-scale metabolic network models have demonstrated their suitability for the integrated analysis of large and complex metabolic networks providing new clues for identifying drug targets.

## GSMMs as new tools emerging from systems biology approach to drug discovery

In the previous sections, we have presented evidences that support cancer onset and that the progression relies on metabolic abnormalities to balance energy demand and biomolecular synthesis (metabolic reprogramming) [68]. GSMMs are emerging as a potential solution to decipher the molecular mechanisms underlying cancer in the context of systems biology [69]. GSMMs represent the metabolic reaction complement encoded by an organism's genome. These models are built based on the literature and databases and enable one to summarize and codify information known about the metabolism of an organism.

Over 100 GSMMs have been built for different species, ranging from archea to mammals [70–84]. Reconstructions of human metabolism, such as Recon1 [81], Edinburgh Human Metabolic Network [82] or the most recent reconstructions of human metabolism, Recon2 [83], are widely used to study the mechanism of diseases with a strong metabolic component, such as cancer or diabetes [85–88].

This systems biology tool enables the mathematical representation of biotransformations and metabolic processes occurring within the organism and offers an appropriate framework to integrate the increasing amount of 'omic' data generated by the different high-throughput technologies.

The transformation into a mathematical formulation is mostly driven by constraint-based modeling (CBM) [89] and allows the systematic simulation of different phenotypes, environmental conditions, gene deletion and so on. This approach allows for modeling the complexity of cancer metabolism and tackling more problematic biological questions, such as the role of metabolism in cancer disease [90].

Genome-scale constraint-based metabolic models have been used for a variety of applications, involving studies on evolution [91], metabolic engineering [92–94], genome annotation [95] or drug discovery [96], with a high relevance in cancer research.

Indeed, GSMMs can efficiently capture the complexity of cancer metabolism in a holistic manner and permit to improve existing therapies or develop new ones [97].

In this chapter we discuss methods for building GSMMs and computational approaches to analyze and integrate 'omic' data into these large-scale metabolic network models. Finally, we introduce some of the most relevant softwares and algorithms developed for drug-target discovery that can be used in cancer research.

### GSMM reconstruction

Genome-scale metabolic reconstructions are created in a bottom-up manner based on genomic and bibliomic data and, thus, represent a biochemical, genetic and genomic knowledge base for the target organism [81–83]. However, to date we are still not able to completely and automatically reconstruct high-quality metabolic networks **(Figure 3A)** [98]. Genome-scale reconstruction starts with the generation of a draft,

**Key term**

**Enzymatic drug target:** A component in a metabolic pathway to which some other entity, such as a drug, is directed and/or binds.

automated reconstruction based on the genome annotation and biochemical databases of the target organism. This task can be achieved by using software tools, such as Pathways tool [99]. The genomic sequence of the targeted organism is coupled with the most recent annotations available from databases [100], such as GOLD or NCBI Entrez Gene databases [101,102].

Metabolic reactions can be associated with the annotated metabolic genes by using enzyme commission (E.C.), ID and biochemical reaction databases (e.g., KEGG [103] and BRENDA [104]). This process permits both linking metabolic genes with their corresponding encoded enzymes and determining the stoichiometric relationship of metabolic reactions with the metabolites and cofactors that they consume and/or produce.

The gene–protein-reaction association (GPR). is represented as Boolean relationships in which isoenzymes that catalyze the same reaction have an "OR" relation (only one of the genes that encode the different isoenzymes is required to have the reaction active) and the complexes that catalyze a reaction have an "AND" relation (all the genes that encode the different complex subunits are necessary to have the reaction active) [81]. GPR associations enable the mapping of transcriptomics or proteomics to the level of reactions.

Reactions can be located into different subcellular compartments based on protein location [81]. Reaction directionality can be determined from thermodynamic data. Additionally, artificial reactions, such as biomass reaction that define the ratio at which biomass constituents are produced (nucleic acids, lipid, proteins, etc) or exchange reactions that define the overall rate of nutrients consumption or production, are also defined in the reconstruction. These artificial reactions are necessary to predict or impose certain phenotypic conditions on the mathematical model.

Next, it is necessary to manually curate and refine the draft, automated reconstruction. The main objective of curation is to identify and correct incomplete or erroneous annotation, add reactions that occur spontaneously and remove gaps and metabolites that cannot be produced or consumed [81] through search on the literature and other databases.

Once the model is curated, it is evaluated and validated in an iterative fashion by using mathematical tools [105]. The aim of the validation process is to evaluate if the model is stoichiometrically balanced, find gaps in the network and search for candidate reactions for gap filling, quantitative evaluation of biomass precursor production and growth rate, compare predicted physiological properties with known properties and determine the metabolic capabilities of the model.

It is worth noting that once a GSMM has been constructed, it can be used in future reconstructions in order to expand and refine the model [81,83].

## Constraint-based methods as tools for tumor metabolism characterization

As was previously mentioned, GSMMs include stoichiometric details for the set of known reactions in a given organism. These large scale metabolic models require computational methods to be qualitatively analyzed. Traditionally, approaches based on ordinary differential equation have been used for characterization of dynamic cell states. However, this full-scale dynamic modeling is frequently infeasible for large-scale networks because of a paucity of necessary parameter values.

Constraint-based methods (CBMs) permit the analysis of large-scale biochemical systems under conditions where kinetic parameters need not be defined (steady state). Genome-scale constraint-based metabolic models can be used to predict or describe cellular behaviors, such as growth rates, uptake/secretion rate or intracellular fluxes [89]. Flux balance analysis (FBA) is one of the most widely used CBMs for the study of biochemical networks. The variables used in FBA include the fluxes through transport and metabolic reactions and model parameters include reaction stoichiometry, biomass composition, ATP requirements and the upper and lower bounds for individual fluxes, which define the maximum and minimum allowable fluxes of the reactions.

The first step in FBA is the mathematical representation of the metabolic reactions in the form of a numerical matrix, with stoichiometric coefficients of each reaction (stoichiometric matrix), where the metabolites are represented in rows and reactions in columns. FBA employs mass actions formalism for the mathematical representation of the metabolic networks: $dC/Dt = S.v.$, where v and C are vectors of reaction fluxes and metabolite concentration respectively, t is time and S is the stoichiometric matrix (Figure 3A).

The next step is to impose constraints to the metabolic network. Constraints are fundamentally represented in two ways:

- Steady-state mass-balance imposes constraints on stoichiometry and network topology on the metabolic fluxes through the network. Additionally, steady state assumption also imposes constraints that narrow the space of solutions. By definition, the change in the concentration of a certain metabolite over time at steady state is 0: $dC/Dt = 0$, thus: $S.v = 0$. These constraints ensure that for each metabolite in the network the net production rate equals the net consumption rate;

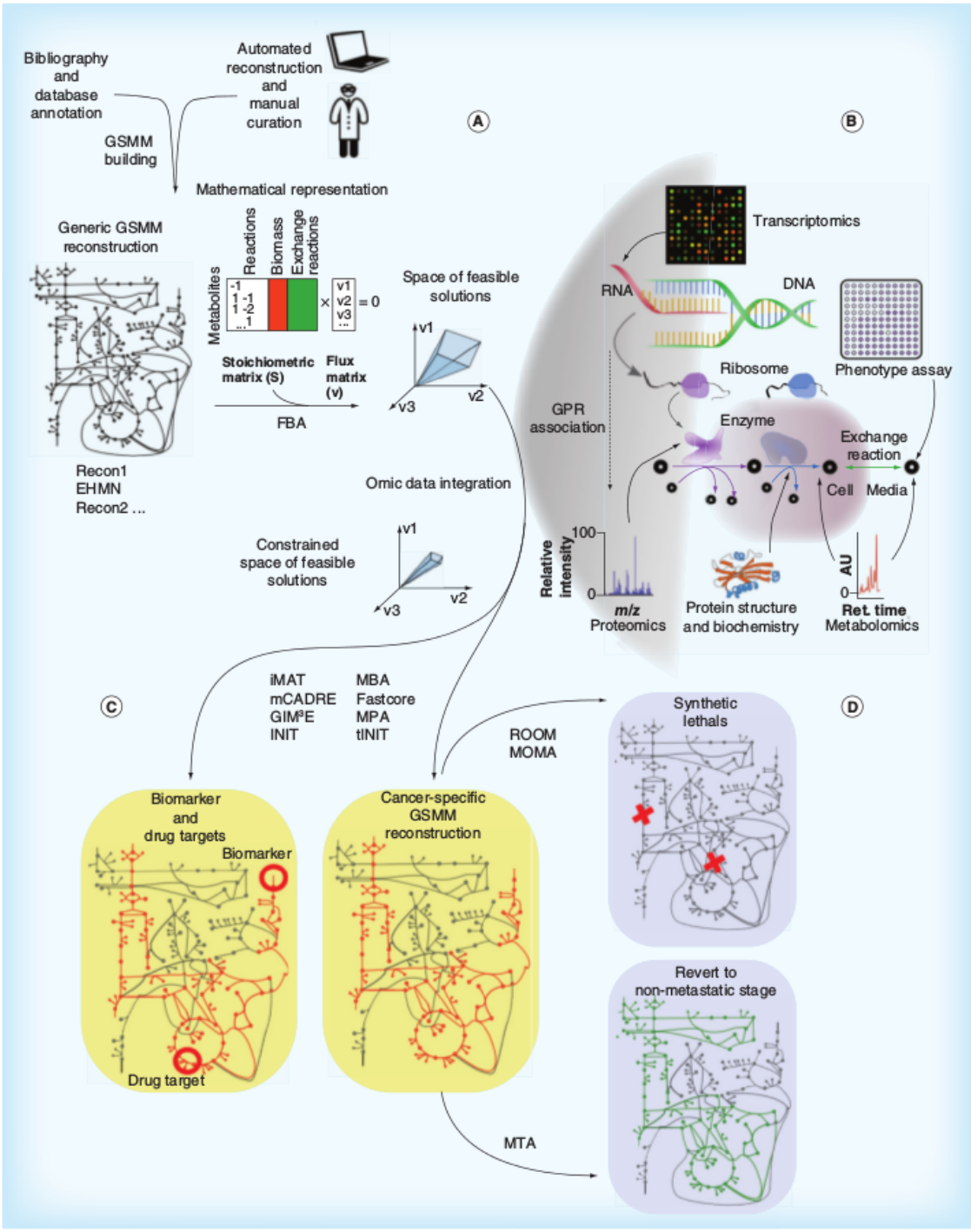

Bibliography and database annotation
Automated reconstruction and manual curation
GSMM building
A
B
Mathematical representation
Generic GSMM reconstruction
Reactions
Biomass
Exchange reactions
Metabolites
Stoichiometric matrix (S)
Flux matrix (v)
FBA
Space of feasible solutions
Recon1
EHMN
Recon2 ...
Omic data integration
Constrained space of feasible solutions
Transcriptomics
RNA
DNA
Ribosome
Phenotype assay
GPR association
Enzyme
Exchange reaction
Cell
Media
Relative intensity
m/z
Proteomics
Protein structure and biochemistry
AU
Ret. time
Metabolomics
C
iMAT
mCADRE
GIM3E
INIT
MBA
Fastcore
MPA
tINIT
D
ROOM
MOMA
Synthetic lethals
Biomarker and drug targets
Biomarker
Drug target
Cancer-specific GSMM reconstruction
Revert to non-metastatic stage
MTA

**Figure 3. Genome-scale metabolic model building and analysis (facing page). (A)** GSMM reconstruction starts with a draft automated version based on literature and databases, finally this version is manually curated in order to refine the model. Typically, these models are analyzed by using flux balance analysis, assuming steady state. **(B)** GSMMs can be used as a platform to integrate and combine omic data from multiple layers. In these models, metabolomics data can be associated with metabolites, while genomics, transcriptomics and proteomics can be associated with metabolic reactions, these associations are established through gene–protein-reaction associations. The phenotypic assays can constrain properties of the network, such as growth rate under certain experimental conditions. **(C)** By integrating omic data into a GSMM we can determine either tumor-specific biomarkers or anticancer drug-targets and reconstruct cancer-specific GSMM. **(D)** Cancer-specific reconstructions can be used to determine synthetic lethals specific for each cancer type for which the non-tumor cells are insensitive (ROOM and MOMA methods), Additionally if we reconstruct an initial GSMM describing metastatic phenotype and a target GSMM describing non-metastatic phenotype we can determine the actors that would permit to revert the metastatic phenotype into a non-metastatic one (MTA method).
FBA: Flux balance analysis; GSMM: Genome-scale metabolic model; Ret.: Retention.

- Inequalities that impose bounds on the system: every reaction can also be given upper and lower bounds. These restrictions are based on measured rates (e.g., metabolite uptake/secretion rates) or reaction reversibility (e.g., irreversible fluxes have a zero lower bound) and are used to define the environmental conditions in a given simulation, such as nutrient or $O_2$ availability, which can be related with a specific tumor microenvironment or stages in tumor progression.

Finally it is necessary to define a phenotype in the form of a biological objective that is relevant to the problem being studied (objective function). Typically, objective functions are related to growth rate prediction. GSMMs define this phenotype by an artificial biomass production reaction, that is, the rate at which metabolic compounds are converted into biomass constituents (nucleic acids, lipid, proteins, etc). The biomass reaction is based on experimental measurements of biomass composition and is unique for each organism or cell type. Thus, an objective function could be the maximization of growth rate that can be accomplished by calculating the set of metabolic fluxes that result in the maximum flux through biomass production reaction. Since uncontrolled cell growth is the basis of tumor progression, this approach is widely used in the simulation of cancer cell metabolism. The objective function can be adapted to the specific cell type or organism; however, the objective that better defines our case of study is not always obvious, especially in multicellular organisms [106].

Taken together, the mathematical representation of the metabolic reactions and of the objective function, is defined as a system of linear equations that are solved by a number of algorithms and software developed for this purpose [105]. Predictions of values for these fluxes are obtained by optimizing for an objective function, while simultaneously satisfying constraint specifications.

## Omic data integration

The advent of high-throughput technologies have transformed molecular biology into a data-rich discipline by providing quantitative data for thousands of cellular components across a wide variety of scales. However, extraction of 'knowledge' from this ocean of omic data has been challenging [107]. GSMMs have emerged as an advantageous platform for the integration of omic data (e.g., [108]; Figure 3B). In this framework cellular and molecular phenotypes are simulated allowing the development of biological hypotheses and discoveries [109]. Metabolic reconstruction of the human metabolism has been successfully used for a variety of analyses of omic data, including applications in data visualization [110], deducing regulatory rules [111], network medicine [112], constructing tissue-specific models [113] or multicellular modeling [114]. Thus, omic data can be used to further constrain the non-uniqueness of constraint-based solutions space and thereby enhance the precision and accuracy of model prediction (Figure 3A–C) [109]. To achieve this aim a number of FBA-driven algorithms that integrate omic data into GSMMs have been developed. Table 1 highlights some of the most relevant approaches recently developed to incorporate experimental omic data into GSMMs [86,87,113,115–117]

## Drug-target & biomarker discovery

Cancer cells maintain their high proliferation rate by adapting their metabolism based on the environmental conditions, such as pH, $O_2$ availability, vascularization or nutrient availability [118]. The elucidation of diverse metabolic alterations for the identification of biomarkers and novel drug targets has, therefore, been increased in recent years. An increasing number of methods and algorithms have been recently developed to integrate tumor-specific omic data into GSMMs. It has enabled the gain of further biological and mechanistic understanding of how cancer benefits from metabolic modifications [90]. This model-driven approach allows the discovery

**Key term**

**Omic data integration:** Computational process in which multi-omic data obtained from different high-throughput technologies, considering different aspects of the molecular biology, are integrated into genome-scale metabolic models in order to unveil emergent properties of the biological systems.

Table 1. Computation method for integrating omic data into global-scale metabolic models.

| Name | Input | Description | Ref. |
|---|---|---|---|
| iMAT | Gene expression data | Seeks to maximize the similarity between the gene expression and the metabolic profiles | [115] |
| mCADRE | Gene expression and metabolomic data | Uses tissue-specific data to identify a set of core reactions. Seeks to build a consistent network using all the core reactions and the minimum number of non-core reactions | [86] |
| GIM$^3$E | Gene expression and metabolomic data | Builds a network that satisfies an objective function while penalizing the inclusion of reactions catalyzed by genes with expression below a certain threshold. It can be further constrained to produce certain metabolites based on experimental evidences | [116] |
| INIT | Gene expression and metabolomic data | Seeks to build a model prioritizing the addition of reactions with strong evidence of their presence based on gene expression data. Can be forced to produce metabolites that have been detected experimentally | [87] |
| MBA | Transcriptomic, proteomic, metabolomic, bibliomic data | Uses tissue-specific data to identify high and moderate probability core reactions. Seeks to build a network with all the high-probability core reactions, the maximum moderate probability core reactions and the non-core reaction required to prevent gaps | [113] |
| Fastcore | Transcriptomic, proteomic, metabolomic, bibliomic data | Identify a set of core reactions based on tissue-specific data. Seeks to build a network that contains all reactions from the core set with the minimum set of additional reactions necessary | [117] |

of potential biomarkers and drug targets [87,97,119]. The identification of new biomarkers is of major importance to biomedical research for early diagnosis and monitoring treatments efficiently. The identification of cancer biomarkers is possible due to aberrant metabolism of tumors that alters the profile of absorption and nutrients secretion.

Omic data of clinical samples (mainly transcriptomics data) can be used to infer the exchange rates of different metabolites for each individual sample via GSMM analysis (alterations in exchange reactions in the model). Thus, those metabolites that significantly differ between two clinical groups in their exchange rates are then considered as potential biomarkers. However, this task is especially challenging in the case of cancer owing to metabolic abnormalities resulting from complex and elaborate genetic and epigenetic alterations that modify the expression of a variety of cancer-associated isoenzymes. In order to determine potential biomarkers in cancer, several computational approaches has been developed. For example, the metabolic phenotypic analysis (MPA) method uses GPR association to integrate transcriptomic and proteomic data within a GSMM to infer metabolic phenotypes [88]. MPA was used to study breast cancer metabolism and predict potential biomarkers. These predictions, wich include amino acid and choline-containing metabolites, are supported by a number of experimental evidences [120]. Another recently developed algorithm is mCADRE, which has been used to systematically simulate the metabolic function of 26 cancer cell types (among other cell types) [86]. This algorithm has been able to identify several pathways, such as folate metabolism, eicosanoid metabolism, fatty acid activation and nucleotide metabolism, that are enriched in tumor tissue compared with their corresponding normal tissue. Many enzymes involved in these pathways are already used as chemotherapy targets. Other approaches, such as flux variability analysis [121] or sampling analysis [122], are also suitable to predict metabolic biomarker candidates by integrating omic data into a GSMM. The novel drug discovery is based on the abnormalities existing in various reactions/pathways of cancer metabolism. These differences can be used as drug targets to attack specific weaknesses of the tumor and hence compromising its viability, but not that of non-cancerous cells [123]. For example, the INIT method [87] was used to identify characteristic metabolic features of cancer cells by inferring the active metabolic network of 16 different cancer types and compare them with the healthy cell types where they come from. These metabolic differences may play an important role in proliferation of cancer cells and could be potential drug targets. This method found

significant differences in polyamine metabolism, the isoprenoid biosynthesis and the prostaglandins and leukotrienes pathways in cancer cells compared with healthy cells. Some of the reactions that were found that have different activity in cancer cells, are already used in the clinical practice as therapeutic targets [124,125]. Based on the rationale that the differences between normal and tumoral cells can be potential therapeutic targets, several approaches have been developed that consider different aspects of cancer metabolism for the discovery of new drug targets:

### Antimetabolite

One of the most common anticancer drugs are antimetabolites. An antimetabolite is structurally similar to a certain metabolite but it cannot be used to produce any physiologically important molecule. Antimetabolite-based drugs act on key enzymes preventing the use of endogenous metabolites, resulting in the disruption of the robustness of cancer cells and reduction or suppression of cell growth. For example, antimetabolites, such as antifolates or antipurines, mimic folic acid and purines [126]. The GSMM approach can be used to systematically simulate the effect of potential antimetabolites in cancer research. To achieve this, methods such as the tINIT (Task-driven Integrative Network Inference for Tissues) algorithm have been developed [97]. This method has been used to reconstruct personalized GSMMs for six hepatocellular carcinoma patients based on proteomics data and the Human Metabolic Reaction database [87] and identify anticancer drugs that are structural analogs to targeted metabolites (antimetabolites). The tINIT algorithm was able to identify 101 antimetabolites, 22 of which are already used in cancer therapies and the remaining can be considered as new potential anticancer drugs.

### Synthetic lethal

The genetic lesions occurring in cancer not only promote the oncogenic state but are also associated with dependencies that are specific to these lesions and absent in non-cancer cells. Two genes are considered 'synthetic lethal' if the isolated mutation on either of them is compatible with cell viability but the simultaneous mutation is lethal [127]. Analogously, two genes are considered to interact in a 'synthetic sick' fashion, if simultaneous mutation reduces cell fitness below a certain threshold without being lethal [127].

Enzymes encoded by genes that are in synthetic lethal or sick interactions with known, non-druggable cancer-driving mutations can be potential anticancer drug targets. This approach has two main advantages: first, we can indirectly target non-druggable cancer-promoting lesions by inhibiting druggable synthetic lethal interactors and secondly we can achieve a high selectivity by exploiting true synthetic lethal interactions for anticancer therapy. This is especially remarkable in the case of cancer-specific isoenzymes, which are emerging as one of the most promising anticancer drug targets. GSMMs provide an excellent tool for the systematic simulation of specific pairs of gene knock-out (KO) to unveil those combinations that compromise the viability of cancer cells (synthetic lethal). By definition, gene KO is simulated by giving value zero to gene expression and the effect of gene deletion is transferred to the metabolic reaction level by GPR association. Thus, for instance, the flux through a reaction that is associated only to one knocked-out gene would be zero. If the reaction is catalyzed by isoenzymes or complexes, the effect of a gene deletion is more complex.

However, predicting the metabolic state of a cell after a gene KO is a challenging task, because after the gene KO the system evolves into a new steady-state that tends to be as close as possible to the original steady-state [128]. To overcome these difficulties several algorithms have been developed. For example, the MOMA algorithm minimizes the euclidean norm of flux differences between metabolic states of the KO compared with the wild type [129]. The ROOM method minimizes the total number of significant flux changes from the wild type flux distribution [129].

In other words, MOMA minimizes the changes in the overall flux distribution while ROOM minimizes the number of fluxes to be modified after the gene KO (Figure 3D). As an example of employing the concept of synthetic lethality in cancer, a GSMM approach has been used to develop a genome scale network model of cancer metabolism [119]. The model predicted 52 cytostatic drug targets (40% of which were known) and further predicted combinations of synthetic lethal drug targets, which were validated using NCI-60 cancer cell collection. In a remarkable example, synthetic lethality between heme oxygenase and fumarate hydratase was predicted by the GSMM approach and was also experimentally validated [130]. The number and the quality of these predictions prove the capabilities of this approach to identify synthetic lethal pairs of genes as potential novel drug target in cancer.

## Future perspective

Metabolism represents the essence of how cells interact with their environment to provide themselves with energy and the essential building blocks for life. In this review, we highlighted the role of a wide range of factors that trigger the malignant transformation of cancer metabolism as well as experimental and computational approaches to develop new therapies. Despite the encouraging achievements and improvements in cancer

research, there still exist limitations that need to be overcome in order to enhance the effectiveness of drug therapies in cancer disease.

One of the major challenges in targeting key metabolic pathways is the lack of clear understanding of how the cancer cell metabolic profile varies from a non-tumor proliferating cell and the potential toxicity risk associated with targeting metabolism. A better understanding of how the metabolism differs in a specific type of cancer or within the same type may help us predict and identify targets without affecting non-tumor cells. In this context, combination of metabolic and signaling pathway inhibitors has been proposed as one of the rational approaches [131]. Using computational approaches permits the systematic simulation of gene perturbations, either metabolic and/or non-metabolic, that could contribute to unveil novel key signaling nodes resulting in potential anticancer drug targets. Recently developed algorithms, such as PROM [111], allow the integration of transcriptomic data into GSMMs while considering the gene regulatory network structure of a given organism. This approach has been developed for predicting metabolic changes that result from genetic or environmental perturbation in *Escherichia coli*. However, it is obvious that algorithms accounting for both gene regulatory and metabolic networks could be used to analyze more precisely the effect of perturbations on oncogenes in cancer metabolism.

Tumor heterogeneity represents a hurdle that must be overcome in order to develop new and more efficient anticancer therapies. One of the factors triggering intratumoral heterogeneity is the tumor microenvironment, which interferes with the ability of drugs to penetrate tumor tissue and reach the entire tumor cells in a potentially lethal concentration. In addition, heterogeneity within the tumor microenvironment leads to marked gradients in the rate of cell proliferation and to regions of hypoxia and acidity, all of which can influence the sensitivity of the tumor cells to drug treatment. Better understanding of how tumor microenvironment protects cancer cells, during and immediately after chemo- or radiotherapy is imperative to design new therapies aimed at targeting this tumor-protective niche [132,133]. The use of drug delivery systems can improve the pharmacological properties of traditional chemotherapeutics by altering pharmacokinetics and biodistribution to overcome the harsh conditions of the tumor microenvironment. Moreover, the co-administration of chemotherapeutics and tumor-associated stromal-depleting drugs helps to target the fibrous structure of the modified extracellular matrix, which can result in a less penetrable tumor microenvironment [134].

Another interesting approach considers therapies that interfere in the metabolic co-operation between cancer cells and stromal cells in their microenvironment [135] or between intratumoral subpopulations. The study of the metabolic coupling between different cellular populations as potential drug targets can be achieved by reconstructing an artificial tumor microenvironment by using GSMMs approach. To date several algorithms have been developed that integrate omic data into a GSMM reconstruction that permit to compute the secretion and uptake rates of nutrients (Table 1) and hence study the complementary secretomes within a heterogeneous cellular community. However, test and validation of a metabolic model becomes more complex if it considers a heterogeneous cellular population. Nevertheless, recent studies on artificial microbial ecosystems have demonstrated the potential of this type of approach to study synergies in heterogeneous cellular communities [136] that could be extrapolated to the study of cancer to unveil the mechanisms underlying the cooperation between tumoral and stromal cells, as well as between intratumoral subpopulations.

The intratumoral microenvironment also confers an extreme flexibility and adaptation capability to cancer cells that enhances tumor progression and represents a challenge for target-directed therapies [137]. The intratumoral heterogeneity is driven by two main processes: epithelial-to-mesenchymal transitions, by which epithelial cells gain invasive properties and lose at least part of their epithelial phenotypes [138]; and mesenchymal-to-epithelial transitions, by which mesenchymal cells can revert to an epithelial gene program displaying strong self-renewal and survival properties [138–140]. Drug targets that repress these processes have been proposed to significantly reduce tumoral progression.

Anti-angiogenic therapy has been proposed for a long time as an interesting approach to reduce tumor growth. Tumor blood vessels are surrounded by a very hostile environment, with a high amount of acidosis, low oxygen regions, weak pericyte–endothelial cell interaction, leading to its tortuous and leaky vessels with gaps that allow easy escape of invading tumor cells [141,142]. Additionally, restoring the blood vessels to a 'normal' state would get the tumor vessels back on track to its proper functional form, reducing hypoxia-induced metastasis and improving the effects of chemotherapy [143,144]. Also it is expected to reduce the spreading of cancer cells, because pericytes that are required to strengthen blood vessels would be acting more efficiently and hence prevent the intravasation of the cancer cells through the gaps found in the normally leaky tumor vessels.

Therapies based on both metastatic targets arresting cancer cells in a non-metastatic stage and angiogenic targets normalizing tumor vessels are promising strategies to design new anticancer therapies. Coupling

this strategy with associated key metabolic pathways is a good approach in cancer treatment and requires computational tools to identify the putative targets. Recently developed methods, such as the 'metabolic transformation algorithm' allows the identification of the actors involved in metabolic transformations [145]. This methodology identifies targets that alter the metabolism retrieving the cells back from a given metabolic state to another metabolic state (Figure 3D). This method has been successfully used to find drug targets that revert disrupted metabolism focused on aging. However, this approach could be suitable to determine drug targets arresting tumor in a non-metastatic stage, normalize tumor vessels or prevent tumor intravasation, resulting in a reduction of tumor progression. Additionally, GSMM predictions could be refined by integrating information from dynamic $^{13}C$ FBA [146].

Moreover, combinatorial therapies, targeting angiogenesis and metastatic targets, have been proposed as a way to enhance anticancer therapies [27]. Traditionally, these approaches has been focused on targeting signaling pathways, such as the VEGF inhibition or VEGF receptors (R1/R2) blockade [147,148] and CXCR4 protein, which is involved in tumor colonization, or the cytokine PIGF, which prepares the metastatic niche in bone marrow for the cells invading from breast cancer [149]. However, studies on the metabolic reprogramming in endothelial cells have opened new avenues to explore the combinatorial therapies of targeting both tumors and their angiogenesis, in the context of metabolism.

**Key term**

**Combinatorial therapies:** Strategy that takes profit of the synergistic effects of two therapeutic treatments targeting different processes of the cellular biology.

The approaches reviewed here provide a guideline to improve the anticancer drug-target therapies focused on metabolic reprogramming. However, the lack of a proper model depicting the complete map of metabolic reactions, regulatory processes as well as tumor heterogeneity and synergistic cooperation between cellular

**Executive summary**

**Background**

- Nowadays, it is widely recognized that metabolic reprogramming is essential to sustain tumor progression. These changes are promoted by genetic and epigenetic alterations producing mutations in key metabolic enzymes that modify flux distributions in metabolic networks, providing advantages to cancer cells in terms of energy production and synthesis of biomolecules.

**Crosstalk between oncogenic signaling events & cancer cell metabolism**

- Many key oncogenic signaling pathways, such as *HIF*, *Myc*, PI3K/AKT/mTOR or SREBPs, converge to adapt tumor cell metabolism in order to support their growth and survival. They are intimately involved in modulating glycolysis, mitochondrial oxidative phosphorylation, lipid metabolism and glutaminolysis.

**Tumor microenvironment**

- The tumor microenvironment is complex and comprises the extracellular matrix, tumor and stromal cells (e.g., epithelial cells, fibroblasts and inflammatory cells) that are embedded within this matrix and nourished by vascular network. The tumor heterogeneity, signaling molecules and chemicals, such as oxygen and protons, can influence tumor cell proliferation, survival, invasion, metastasis and energy metabolism reprogramming.

**Isoenzymes: therapeutic targets in cancer**

- Isoforms of many of the enzymes specific to important metabolic pathways are found to be overexpressed in tumor cells affecting important pathways of the energetic metabolism. These isoforms have a key role in mediating the aberrant metabolism of cancer cells and could serve as a promising source of novel drug targets.
- These tumor-specific isoforms can be involved in important pathways, such as glycolysis, tricarboxylic acid cycle, pentose phosphate pathway and glutamine metabolism, among other important energetic pathways

**Genome-scale metabolic models as new tools emerging from systems biology approach to drug discovery**

- Genome-scale metabolic models are emerging as a potential solution to decipher the molecular mechanisms underlying cancer in the context of systems biology. These models represent the metabolic reactions encoded by an organism's genome and summarize and codify information known about the metabolism of that organism.
- These models use constraint-based methods for the mathematical representation of biotransformations and metabolic processes occurring within the organism and offer an appropriate framework to integrate the increasing amount of 'omic' data generated by the different high-throughput technologies.
- Genome-scale metabolic models approaches have allowed to identify a number of tumor-specific biomarkers, anticancer drug-target and synthetic lethal genes opening a promising avenue in the development of new anticancer therapies.

communities, makes selecting the best possible target combinations difficult. Thus, in order to develop more efficient anticancer therapies, more efforts need to be made in developing new methods to study tumor metabolism and obtain a better understanding of the molecular processes underlying tumor progression and invasion.

### Financial & competing interests disclosure

This work was supported by funds of European Commission METAFLUX (Marie Curie FP7-PEOPLE-2010 ITN-264780); Spanish Government and European Union FEDER Funds (SAF2011–25726); and Generalitat de Catalunya (2014SGR-1017 and Icrea Academia award 2010 granted to M Cascante). The authors have no other relevant affiliations orfinancial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

## References

1 Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J. Clin.* 62(1), 10–29 (2012).

▶2 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 144(5), 646–674 (2011).

▶3 Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell* 21(3), 297–308 (2012).

▶4 Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324(5930), 1029–1033 (2009).

▶5 Stratton MR. Exploring the genomes of cancer cells: progress and promise. *Science* 331(6024), 1553–1558 (2011).

▶6 Yuneva MO, Fan TW, Allen TD *et al.* The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab.* 15(2), 157–170 (2012).

▶7 Obacz J, Pastorekova S, Vojtesek B, Hrstka R. Cross-talk between HIF and p53 as mediators of molecular responses to physiological and genotoxic stresses. *Mol. Cancer* 12(1), 93 (2013).

▶8 Dang CV. Links between metabolism and cancer. *Genes Dev.* 26(9), 877–890 (2012).

▶9 Wise DR, Deberardinis RJ, Mancuso A *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl Acad. Sci. USA* 105(48), 18782–18787 (2008).

10 Pecqueur C, Oliver L, Oizel K, Lalier L, Vallette FM. Targeting metabolism to induce cell death in cancer cells and cancer stem cells. *Int. J. Cell Biol.* 2013, 805975 (2013).

▶11 Dan HC, Ebbs A, Pasparakis M, Van Dyke T, Basseres DS, Baldwin AS. Akt-dependent activation of mTORC1 Involves phosphorylation of mTOR by IKKalpha. *J. Biol. Chem.* 289(36), 25227–25240 (2014).

▶12 Porstmann T, Santos CR, Griffiths B *et al.* SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth. *Cell Metab.* 8(3), 224–236 (2008).

▶13 Alvarez MS, Fernandez-Alvarez A, Cucarella C, Casado M. Stable SREBP-1a knockdown decreases the cell proliferation rate in human preadipocyte cells without inducing senescence. *Biochem. Biophys. Res. Commun.* 447(1), 51–56 (2014).

▶14 Furuta E, Pai SK, Zhan R *et al.* Fatty acid synthase gene is up-regulated by hypoxia via activation of Akt and sterol regulatory element binding protein-1. *Cancer Res.* 68(4), 1003–1011 (2008).

▶15 Cascante M, Benito A, Zanuy M, Vizan P, Marin S, De Atauri P. Metabolic network adaptations in cancer as targets for novel therapies. *Biochem. Soc. Trans.* 38(5), 1302–1306 (2010).

▶16 Doherty JR, Yang C, Scott KEN *et al.* Blocking lactate export by inhibiting the MYC target MCT1 disables glycolysis and glutathione synthesis. *Cancer Res.* 74(3), 908–920 (2014).

▶17 Jiang P, Du W, Wang X *et al.* p53 regulates biosynthesis through direct inactivation of glucose-6-phosphate dehydrogenase. *Nat. Cell Biol.* 13(3), 310–316 (2011).

▶18 Bentz S, Cee A, Endlicher E *et al.* Hypoxia induces the expression of transketolase-like 1 in human colorectal cancer. *Digestion* 88(3), 182–192 (2013).

▶19 Sinthupibulyakit C, Ittarat W, St Clair WH, St Clair DK. p53 protects lung cancer cells against metabolic stress. *Int. J. Oncol.* 37(6), 1575–1581 (2010).

▶20 Cui G, Park S, Badeaux AI *et al.* PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nat. Struct. Mol. Biol.* 19(9), 916–924 (2012).

▶21 Xiao M, Yang H, Xu W *et al.* Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* 26(12), 1326–1338 (2012).

▶22 Xu W, Yang H, Liu Y *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* 19(1), 17–30 (2011).

▶23 Commandeur S, Ho SH, De Gruijl FR, Willemze R, Tensen CP, El Ghalbzouri A. Functional characterization of cancer-associated fibroblasts of human cutaneous squamous cell carcinoma. *Exp. Dermatol.* 20(9), 737–742 (2011).

▶24 Zhou B, Chen WL, Wang YY *et al.* A role for cancer-associated fibroblasts in inducing the epithelial-to-mesenchymal transition in human tongue squamous cell carcinoma. *J. Oral Pathol. Med.* 43(8), 585–592 (2014).

▶25 Yu Y, Xiao CH, Tan LD, Wang QS, Li XQ, Feng YM. Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through paracrine TGF-[beta] signalling. *Br. J. Cancer* 110(3), 724–732 (2014).

26 Zhang Y, Choksi S, Chen K, Pobezinskaya Y, Linnoila I, Liu ZG. ROS play a critical role in the differentiation of alternatively activated macrophages and the occurrence of tumor-associated macrophages. *Cell Res.* 23(7), 898–914 (2013).

27 De Bock K, Mazzone M, Carmeliet P. Antiangiogenic therapy, hypoxia, and metastasis: risky liaisons, or not? *Nat. Rev. Clin. Oncol.* 8(7), 393–404 (2011).

28 Wang W, Quan Y, Fu Q *et al.* Dynamics between cancer cell subpopulations reveals a model coordinating with both hierarchical and stochastic concepts. *PLoS ONE* 9(1), e84654 (2014).

29 Feng W, Gentles A, Nair RV *et al.* Targeting unique metabolic properties of breast tumor initiating cells. *Stem Cells* 32(7), 1734–1745 (2014).

30 Ciavardelli D, Rossi C, Barcaroli D *et al.* Breast cancer stem cells rely on fermentative glycolysis and are sensitive to 2-deoxyglucose treatment. *Cell Death Dis.* 5, e1336 (2014).

31 Nieman KM, Kenny HA, Penicka CV *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nat. Med.* 17(11), 1498–1503 (2011).

32 Willemarck N, Rysman E, Brusselmans K *et al.* Aberrant activation of fatty acid synthesis suppresses primary cilium formation and distorts tissue development. *Cancer Res.* 70(22), 9453–9462 (2010).

33 Park JB, Lee CS, Jang JH *et al.* Phospholipase signalling networks in cancer. *Nat. Rev. Cancer* 12(11), 782–792 (2012).

34 Nomura DK, Long JZ, Niessen S, Hoover HS, Ng SW, Cravatt BF. Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell* 140(1), 49–61 (2010).

35 Karnezis T, Shayan R, Caesar C *et al.* VEGF-D promotes tumor metastasis by regulating prostaglandins produced by the collecting lymphatic endothelium. *Cancer Cell* 21(2), 181–195 (2012).

36 Seguin F, Carvalho MA, Bastos DC *et al.* The fatty acid synthase inhibitor orlistat reduces experimental metastases and angiogenesis in B16-F10 melanomas. *Br. J. Cancer* 107(6), 977–987 (2012).

37 Pavlides S, Tsirigos A, Migneco G *et al.* The autophagic tumor stroma model of cancer. Role of oxidative stress and ketone production in fueling tumor cell metabolism. *Cell Cycle* 9(17), 3485–3505 (2010).

38 Carito V, Bonuccelli G, Martinez-Outschoorn UE *et al.* Metabolic remodeling of the tumor microenvironment: migration stimulating factor (MSF) reprograms myofibroblasts toward lactate production, fueling anabolic tumor growth. *Cell Cycle* 11(18), 3403–3414 (2012).

39 Hur H, Paik MJ, Xuan Y *et al.* Quantitative measurement of organic acids in tissues from gastric cancer patients indicates increased glucose metabolism in gastric cancer. *PLoS ONE* 9(6), e98581 (2014).

40 Peters K, Kamp G, Berz A *et al.* Changes in human endothelial cell energy metabolic capacities during *in vitro* cultivation. The role of "aerobic glycolysis" and proliferation. *Cell Physiol. Biochem.* 24(5–6), 483–492 (2009).

41 Harjes U, Bensaad K, Harris AL. Endothelial cell metabolism and implications for cancer therapy. *Br. J. Cancer* 107(8), 1207–1212 (2012).

42 Vegran F, Boidot R, Michiels C, Sonveaux P, Feron O. Lactate influx through the endothelial cell monocarboxylate transporter MCT1 supports an NF-kappaB/IL-8 pathway that drives tumor angiogenesis. *Cancer Res.* 71(7), 2550–2560 (2011).

43 Vizan P, Sanchez-Tena S, Alcarraz-Vizan G *et al.* Characterization of the metabolic changes underlying growth factor angiogenic activation: identification of new potential therapeutic targets. *Carcinogenesis* 30(6), 946–952 (2009).

44 Calcinotto A, Filipazzi P, Grioni M *et al.* Modulation of microenvironment acidity reverses anergy in human and murine tumor-infiltrating T lymphocytes. *Cancer Res.* 72(11), 2746–2756 (2012).

45 Polet F, Feron O. Endothelial cell metabolism and tumour angiogenesis: glucose and glutamine as essential fuels and lactate as the driving force. *J. Intern. Med.* 273(2), 156–165 (2013).

46 Choi SY, Collins CC, Gout PW, Wang Y. Cancer-generated lactic acid: a regulatory, immunosuppressive metabolite? *J. Pathol.* 230(4), 350–355 (2013).

47 Wels J, Kaplan RN, Rafii S, Lyden D. Migratory neighbors and distant invaders: tumor-associated niche cells. *Genes Dev.* 22(5), 559–574 (2008).

48 Cardenas-Navia LI, Mace D, Richardson RA, Wilson DF, Shan S, Dewhirst MW. The pervasive presence of fluctuating oxygenation in tumors. *Cancer Res.* 68(14), 5812–5819 (2008).

49 Sharma MK, Seidlitz EP, Singh G. Cancer cells release glutamate via the cystine/glutamate antiporter. *Biochem. Biophys. Res. Commun.* 391(1), 91–95 (2010).

50 Meneses AM, Medina RA, Kato S *et al.* Regulation of GLUT3 and glucose uptake by the cAMP signalling pathway in the breast cancer cell line ZR-75. *J. Cell. Physiol.* 214(1), 110–116 (2008).

51 Chen J, Zhang S, Li Y, Tang Z, Kong W. Hexokinase 2 overexpression promotes the proliferation and survival of laryngeal squamous cell carcinoma. *Tumor Biol.* 35(4), 3743–3753 (2014).

52 Robey RB, Hay N. Mitochondrial hexokinases, novel mediators of the antiapoptotic effects of growth factors and Akt. *Oncogene* 25(34), 4683–4696 (2006).

53 Galluzzi L, Kepp O, Tajeddine N, Kroemer G. Disruption of the hexokinase-VDAC complex for tumor therapy. *Oncogene* 27(34), 4633–4635 (2008).

54 Christofk HR, Vander Heiden MG, Harris MH *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 452(7184), 230–233 (2008).

55 Christofk HR, Vander Heiden MG, Wu N, Asara JM, Cantley LC. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* 452(7184), 181–186 (2008).

56 Vander Heiden MG, Christofk HR, Schuman E *et al.* Identification of small molecule inhibitors of pyruvate kinase M2. *Biochem. Pharmacol.* 79(8), 1118–1124 (2010).

57 Clem B, Telang S, Clem A *et al.* Small-molecule inhibition of 6-phosphofructo-2-kinase activity suppresses glycolytic flux and tumor growth. *Mol. Cancer Ther.* 7(1), 110–120 (2008).

58 Zhou M, Zhao Y, Ding Y *et al.* Warburg effect in chemosensitivity: targeting lactate dehydrogenase-A re-sensitizes Taxol-resistant cancer cells to Taxol. *Mol. Cancer* 9(1), 33 (2010).

59 Le A, Cooper CR, Gouw AM *et al.* Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression. *Proc. Nat. Acad. Sci. USA* 107(5), 2037–2042 (2010).

60 Lu CW, Lin SC, Chen KF, Lai YY, Tsai SJ. Induction of pyruvate dehydrogenase kinase-3 by hypoxia-inducible factor-1 promotes metabolic switch and drug resistance. *J. Biol. Chem.* 283(42), 28106–28114 (2008).

61 Yan H, Parsons DW, Jin G *et al.* IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* 360(8), 765–773 (2009).

62 Dang L, Jin S, Su SM. IDH mutations in glioma and acute myeloid leukemia. *Trends Mol. Med.* 16(9), 387–397 (2010).

63 El-Ashmawy N, El-Bahrawy H, Shamloula M, El-Feky O. Biochemical/metabolic changes associated with hepatocellular carcinoma development in mice. *Tumor Biol.* 35(6), 5459–5466 (2014).

64 Xu X, Zur Hausen A, Coy JF, Löchelt M. Transketolase-like protein 1 (TKTL1) is required for rapid cell growth and full viability of human tumor cells. *Int. J. Cancer* 124(6), 1330–1337 (2009).

65 Vizán P, Alcarraz-Vizán G, Díaz-Moralli S, Solovjeva ON, Frederiks WM, Cascante M. Modulation of pentose phosphate pathway during cell cycle progression in human colon adenocarcinoma cell line HT29. *Int. J. Cancer* 124(12), 2789–2796 (2009).

66 Dang CV. Glutaminolysis: supplying carbon or nitrogen or both for cancer cells? *Cell Cycle* 9(19), 3914–3916 (2010).

67 Durán RV, Oppliger W, Robitaille AM *et al.* Glutaminolysis activates Rag-mTORC1 signaling. *Mol. Cell* 47(3), 349–358 (2012).

68 Deberardinis RJ, Thompson CB. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* 148(6), 1132–1144 (2012).

69 Mardinoglu A, Nielsen J. Systems medicine and metabolic modelling. *J. Intern. Med.* 271(2), 142–154 (2012).

70 Gonnerman MC, Benedict MN, Feist AM, Metcalf WW, Price ND. Genomically and biochemically accurate metabolic reconstruction of *Methanosarcina barkeri fusaro*, iMG746. *Biotechnol. J.* 8(9), 1070–1079 (2013).

71 Benedict MN, Gonnerman MC, Metcalf WW, Price ND. Genome-scale metabolic reconstruction and hypothesis testing in the methanogenic archaeon *Methanosarcina acetivorans* C2A. *J. Bacteriol.* 194(4), 855–865 (2012).

72 Lee NR, Lakshmanan M, Aggarwal S *et al.* Genome-scale metabolic network reconstruction and *in silico* flux analysis of the thermophilic bacterium *Thermus thermophilus* HB27. *Microb. Cell Fact.* 13(1), 61 (2014).

73 Orth JD, Conrad TM, Na J *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism – 2011. *Mol. Syst. Biol.* 7(1), 535 ( 2011).

74 Nagarajan H, Sahin M, Nogales J *et al.* Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. *Microb. Cell Fact.* 12(1), 118 (2013).

75 Song C, Chiasson MA, Nursimulu N *et al.* Metabolic reconstruction identifies strain-specific regulation of virulence in *Toxoplasma gondii*. *Mol. Syst. Biol.* 9(1), 708 (2013).

76 Heavner B, Smallbone K, Barker B, Mendes P, Walker L. Yeast 5 – an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst. Biol.* 6(1), 55 (2012).

77 Caspeta L, Shoaie S, Agren R, Nookaew I, Nielsen J. Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and *in silico* evaluation of their potentials. *BMC Syst. Biol.* 6(1), 24 (2012).

78 Liu J, Gao Q, Xu N, Liu L. Genome-scale reconstruction and *in silico* analysis of *Aspergillus terreus* metabolism. *Mol. Biosyst.* 9(7), 1939–1948 (2013).

79 Arnold A, Nikoloski Z. Bottom-up metabolic reconstruction of arabidopsis and its application to determining the metabolic costs of enzyme production. *Plant Physiol.* 165(3), 1380–1391 (2014).

80 Dal'molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. C4GEM, a genome-scale metabolic model to study C4 plant metabolism. *Plant Physiol.* 154(4), 1871–1885 (2010).

81 Duarte NC, Becker SA, Jamshidi N *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA* 104(6), 1777–1782 (2007).

82 Hao T, Ma HW, Zhao XM, Goryanin I. The reconstruction and analysis of tissue specific human metabolic networks. *Mol. Biosyst.* 8(2), 663–670 (2012).

83 Thiele I, Swainston N, Fleming RM *et al.* A community-driven global reconstruction of human metabolism. *Nat. Biotech.* 31(5), 419–425 (2013).

84 Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst. Biol.* 4(1), 140 (2010).

85 Bordbar A, Feist AM, Usaite-Black R, Woodcock J, Palsson BØ, Famili I. A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst. Biol.* 5(1), 180 (2011).

86 Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* 6(1), 153 (2012).

87 Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* 8(5), e1002518 (2012).

88 Jerby L, Wolf L, Denkert C *et al.* Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. *Cancer Res.* 72(22), 5712–5720 (2012).

89 Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat. Biotech.* 28(3), 245–248 (2010).

90 Rajcevic U, Knol J, Piersma S *et al.* Colorectal cancer derived organotypic spheroids maintain essential tissue characteristics but adapt their metabolism in culture. *Proteome Sci.* 12(1), 39 (2014).

91 Papp B, Notebaart RA, Pál C. Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.* 12(9), 591–602 (2011).

92 Matsuda F, Furusawa C, Kondo T, Ishii J, Shimizu H, Kondo A. Engineering strategy of yeast metabolism for higher alcohol production. *Microb. Cell Fact.* 10(1), 70 (2011).

93 Park JM, Song H, Lee HJ, Seung D. Genome-scale reconstruction and *in silico* analysis of *Klebsiella oxytoca* for 2,3-butanediol production. *Microb. Cell Fact.* 12(1), 20 (2013).

94 Lütke-Eversloh T, Bahl H. Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. *Curr. Opin. Biotechnol.* 22(5), 634–647 (2011).

95 Kumar VS, Maranas CD. GrowMatch: an automated method for reconciling *in silico*/*in vivo* growth predictions. *PLoS Comput. Biol.* 5(3), e1000308 (2009).

96 Kim HU, Kim SY, Jeong H *et al.* Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Mol. Syst. Biol.* 7(1), 460 (2011).

97 Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* 10(3), 721 (2014).

98 Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7(2), 129–143 (2009).

99 Caspi R, Altman T, Dale JM *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 38(Suppl. 1), D473–D479 (2010).

100 Henry CS, Dejongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotech.* 28(9), 977–982 (2010).

101 Coordinators NR. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 41(D1), D8–D20 (2013).

102 Pagani I, Liolios K, Jansson J *et al.* The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 40(D1), D571–D579 (2012).

103 Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40(D1), D109–D114 (2012).

104 Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.* 37(Suppl. 1), D588–D592 (2009).

105 Schellenberger J, Que R, Fleming RM *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nat. Protoc.* 6(9), 1290–1307 (2011).

106 Wintermute E, Lieberman T, Silver P. An objective function exploiting suboptimal solutions in metabolic networks. *BMC Syst. Biol.* 7(1), 98 (2013).

107 Palsson B, Zengler K. The challenges of integrating multi-omic data sets. *Nat. Chem. Biol.* 6(11), 787–789 (2010).

108 Bordbar A, Mo ML, Nakayasu ES *et al.* Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Mol. Syst. Biol.* 8(1), 558 (2012).

109 Lewis NE, Nagarajan H, Palsson BØ. Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods. *Nat. Rev. Microbiol.* 10(4), 291–305 (2012).

110 Jensen PA, Papin JA. MetDraw: automated visualization of genome-scale metabolic network reconstructions and high-throughput data. *Bioinformatics* 30(9), 1327–1328 (2014).

111 Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* 107(41), 17845–17850 (2010).

112 Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12(1), 56–68 (2011).

113 Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* 6(1), 401 (2010).

114 Lewis NE, Schramm G, Bordbar A *et al.* Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain. *Nat. Biotech.* 28(12), 1279–1285 (2010).

115 Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. *Bioinformatics* 26(24), 3140–3142 (2010).

116 Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BØ, Hyduke DR. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics* 29(22), 2900–2908 (2013).

117 Galhardo M, Sinkkonen L, Berninger P, Lin J, Sauter T, Heinäniemi M. Integrated analysis of transcript-level regulation of metabolism reveals disease-relevant nodes of the human metabolic network. *Nucleic Acids Res.* 42(3), 1474–1496 (2013).

118 Lazar MA, Birnbaum MJ. De-meaning of metabolism. *Science* 336(6089), 1651–1652 (2012).

119 Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* 7(1), 501 (2011).

120 Davis VW, Bathe OF, Schiller DE, Slupsky CM, Sawyer MB. Metabolomics and surgical oncology: potential role for small molecule biomarkers. *J. Surg. Oncol.* 103(5), 451–459 (2011).

121 Murabito E, Simeonidis E, Smallbone K, Swinton J. Capturing the essence of a metabolic network: a flux balance analysis approach. *J. Theor. Biol.* 260(3), 445–452 (2009).

122 Schellenberger J, Palsson BØ. Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.* 284(9), 5457–5461 (2009).

123 Duggal R, Minev B, Geissinger U *et al.* Biotherapeutic approaches to target cancer stem cells. *J. Stem Cells* 8(3–4), 135–149 (2013).

124 Philips MR, Cox AD. Geranylgeranyltransferase I as a target for anti-cancer drugs. *J. Clin. Invest.* 117(5), 1223–1225 (2007).

125 Dudakovic A, Tong H, Hohl R. Geranylgeranyl diphosphate depletion inhibits breast cancer cell migration. *Invest. New Drugs* 29(5), 912–920 (2011).

126 Hebar A, Valent P, Selzer E. The impact of molecular targets in cancer drug development: major hurdles and future strategies. *Expert Rev. Clin. Pharmacol.* 6(1), 23–34 (2013).

127 Conde-Pueyo N, Munteanu A, Solé RV, Rodríguez-Caso C. Human synthetic lethal inference as potential anti-cancer target gene detection. *BMC Syst. Biol.* 3(1), 116 (2009).

128 Barbash DA, Lorigan JG. Lethality in *Drosophilamelanogaster/Drosophilasimulans* species hybrids is not associated with substantial transcriptional misregulation. *J. Exp. Zool. Part B Mol. Dev. Evol.* 308(1), 74–84 (2007).

129 Ren S, Zeng B, Qian X. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. *BMC Bioinformatics* 14(Suppl. 2), S17 (2013).

130 Frezza C, Zheng L, Folger O *et al.* Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* 477(7363), 225–228 (2011).

131 Zhao Y, Butler EB, Tan M. Targeting cellular metabolism to improve cancer therapeutics. *Cell Death Dis.* 4, e532 (2013).

132 Vaupel P. Tumor microenvironmental physiology and its implications for radiation oncology. *Sem. Radiat. Oncol.* 14(3), 198–206 (2004).

133 Tannock IF, Lee CM, Tunggal JK, Cowan DS, Egorin MJ. Limited penetration of anticancer drugs through tumor tissue a potential cause of resistance of solid tumors to chemotherapy. *Clin. Cancer Res.* 8(3), 878–884 (2002).

134 Olive KP, Jacobetz MA, Davidson CJ *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* 324(5933), 1457–1461 (2009).

135 Hulit J, Howell A, Gandara R, Sartini M, Arafat H, Bevilacqua G. Creating a tumor-resistant microenvironment. *Cell Cycle* 12(3), 480–490 (2013).

136 Ye C, Zou W, Xu N, Liu L. Metabolic model reconstruction and analysis of an artificial microbial ecosystem for vitamin C production. *J. Biotech.* 182, 61–67 (2014).

137 Peinado H, Olmeda D, Cano A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat. Rev. Cancer* 7(6), 415–428 (2007).

138 Celià-Terrassa T, Meca-Cortés Ó, Mateo F *et al.* Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J. Clin. Invest.* 122(5), 1849 (2012).

139 Korpal M, Ell BJ, Buffa FM *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat. Med.* 17(9), 1101–1108 (2011).

140 Ocaña OH, Córcoles R, Fabra Á *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* 22(6), 709–724 (2012).

141 Rundqvist H, Johnson RS. Hypoxia and metastasis in breast cancer. In: *Diverse Effects of Hypoxia on Tumor Progression*. Simon MC (Ed.) Springer, Berlin/Heidelberg, Germany, 121–139 (2010).

142 Nagy JA, Chang S-H, Shih S-C, Dvorak AM, Dvorak HF. Heterogeneity of the tumor vasculature. *Semin. Thromb. Hemost.* 36(3), 321–331 (2010).

143 Mazzone M, Dettori D, Leite De Oliveira R *et al.* Heterozygous deficiency of PHD2 restores tumor oxygenation and inhibits metastasis via endothelial normalization. *Cell* 136(5), 839–851 (2009).

144 Rolny C, Mazzone M, Tugues S *et al.* HRG inhibits tumor growth and metastasis by inducing macrophage polarization and vessel normalization through downregulation of PlGF. *Cancer Cell* 19(1), 31–44 (2011).

145 Yizhak K, Gabay O, Cohen H, Ruppin E. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat. Commun.* 4, 2632 (2013).

146 De Mas IM, Selivanov VA, Marin S *et al.* Compartmentation of glycogen metabolism revealed from 13C isotopologue distributions. *BMC Syst. Biol.* 5(1), 175 (2011).

147 Hassan S, Buchanan M, Jahan K *et al.* CXCR4 peptide antagonist inhibits primary breast tumor growth, metastasis and enhances the efficacy of anti-VEGF treatment or docetaxel in a transgenic mouse model. *Int. J. Cancer* 129(1), 225–232 (2011).

148 Coenegrachts L, Maes C, Torrekens S *et al.* Anti-placental growth factor reduces bone metastasis by blocking tumor cell engraftment and osteoclast differentiation. *Cancer Res.* 70(16), 6537–6547 (2010).

149 Hiratsuka S, Duda DG, Huang Y *et al.* CXC receptor type 4 promotes metastasis by activating p38 mitogen-activated protein kinase in myeloid differentiation antigen (Gr-1)-positive cells. *Proc. Natl Acad. Sci. USA* 108(1), 302–307 (2011).

# Apendix III

Roca et al. Journal of Translational Medicine 2014, 12(Suppl 2):S3
http://www.translational-medicine.com/content/12/S2/S3

JOURNAL OF TRANSLATIONAL MEDICINE

REVIEW Open Access

# Chronic Obstructive Pulmonary Disease heterogeneity: challenges for health risk assessment, stratification and management

Josep Roca[1,2*], Claudia Vargas[1], Isaac Cano[1], Vitaly Selivanov[1], Esther Barreiro[2,3], Dieter Maier[4], Francesco Falciani[5], Peter Wagner[6], Marta Cascante[1,2,7], Judith Garcia-Aymerich[8], Susana Kalko[1], Igor Marin De Mas[1], Jesper Tegnér[9], Joan Escarrabill[1,2], Alvar Agustí[1,2], David Gomez-Cabrero[9], the Synergy-COPD consortium

## Abstract

**Background and hypothesis:** Heterogeneity in clinical manifestations and disease progression in Chronic Obstructive Pulmonary Disease (COPD) lead to consequences for patient health risk assessment, stratification and management. Implicit with the classical *"spill over"* hypothesis is that COPD heterogeneity is driven by the pulmonary events of the disease. Alternatively, we hypothesized that COPD heterogeneities result from the interplay of mechanisms governing three conceptually different phenomena: 1) pulmonary disease, 2) systemic effects of COPD and 3) co-morbidity clustering, each of them with their own dynamics.

**Objective and method:** To explore the potential of a systems analysis of COPD heterogeneity focused on skeletal muscle dysfunction and on co-morbidity clustering aiming at generating predictive modeling with impact on patient management. To this end, strategies combining deterministic modeling and network medicine analyses of the Biobridge dataset were used to investigate the mechanisms of skeletal muscle dysfunction. An independent data driven analysis of co-morbidity clustering examining associated genes and pathways was performed using a large dataset (ICD9-CM data from Medicare, 13 million people). Finally, a targeted network analysis using the outcomes of the two approaches (skeletal muscle dysfunction and co-morbidity clustering) explored shared pathways between these phenomena.

**Results:** (1) Evidence of abnormal regulation of skeletal muscle bioenergetics and skeletal muscle remodeling showing a significant association with nitroso-redox disequilibrium was observed in COPD; (2) COPD patients presented higher risk for co-morbidity clustering than non-COPD patients increasing with ageing; and, (3) the on-going targeted network analyses suggests shared pathways between skeletal muscle dysfunction and co-morbidity clustering.

**Conclusions:** The results indicate the high potential of a systems approach to address COPD heterogeneity. Significant knowledge gaps were identified that are relevant to shape strategies aiming at fostering 4P Medicine for patients with COPD.

## Introduction

### COPD is a highly heterogeneous major chronic disease

Chronic Obstructive Pulmonary Disease (COPD) is a prevalent chronic disorder caused by the inhalation of irritants, mainly tobacco smoke. It affects approximately 9% of the adult population above 45 yrs [1]. The disease imposes a high burden on healthcare systems worldwide and it is currently the fourth leading cause of mortality [2]. In 2001, the World Health Organization (WHO) included COPD among the five major chronic disorders together with cardiovascular diseases, cancer, diabetes and mental disorders [3].

Current diagnosis of COPD is based on three concurrent criteria [4]: *i)* presence of respiratory symptoms, mainly shortness of breath and/or chronic cough/sputum; *ii)* history of inhalation of irritants; and, *iii)* forced

* Correspondence: jroca@clinic.ub.es
[1]IDIBAPS, Hospital Clinic. Facultat de Medicina, 08036, Barcelona, Catalunya, Spain
Full list of author information is available at the end of the article

spirometry testing indicating an obstructive ventilatory defect. The expiratory flow limitation observed in COPD patients is due to increased airways resistance and/or reduced lung elasticity caused by destruction of pulmonary parenchyma.

Approximately 15 to 20% of all tobacco smokers are prone to develop COPD and there is marked individual variability of both clinical manifestations and pulmonary disease progression [5-7] with relevant implications in terms of health risk assessment and patient management [6,8]. It is well established that COPD patients can present acute episodes of exacerbation with a negative impact on use of healthcare resources and prognosis [6,9]. Moreover, these patients can also show systemic effects of the disease - being skeletal muscle dysfunction/wasting [10] a characteristic one - and co-morbid conditions [6,11]. Highly prevalent chronic conditions such as cardiovascular disorders (CVD) and type 2 diabetes mellitus - metabolic syndrome (T2DM-MS) are often present as a co-morbidity cluster in COPD patients [12-14]. There is evidence that both skeletal muscle dysfunction/wasting and co-morbidity clustering are independently associated with poor prognosis [1,6,10].

### State of the art on COPD heterogeneity and main challenges

The Global initiative for Obstructive Lung Disease (GOLD) [1] has played a major role in raising COPD awareness and defining standards for treatment. GOLD has faced the challenge of COPD heterogeneity by evolving from an initial disease staging based only on the degree of airflow limitation ($FEV_1$, forced expiratory volume during the first second) [15] to the incorporation of symptoms and frequency of severe exacerbations into the scoring system (2011 GOLD update) (see Table 1 for details) and acknowledging the negative impact of co-morbid conditions on prognosis. Evidence-based data using the 2011 GOLD classification are currently emerging, but the results are not yet consolidated [16,17]. Alternative available options for COPD classification or prediction of survival [18-22] are also insufficient for *subject-specific* prediction and stratification of management.

ECLIPSE - The Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points study - has been a 3-year follow-up project of a large cohort of well characterized COPD GOLD II-IV patients [23] that has generated a relevant body of knowledge on several major aspects of the disease. A recent summary report (2014) on clinical implications of the project outcomes [6] stresses the impact of COPD heterogeneity observed in both the cross-sectional and the longitudinal study assessments.

All in all, there is a strong rationale for further research on subject-specific health risk prediction and stratification aiming at enhancing cost-effective management of COPD patients. However, there are several major challenges to be taken into account in the design of a systems approach to better understand COPD heterogeneity. Firstly, the overlap among different chronic obstructive airways diseases (COPD, asthma, bronchiectasis, bronchiolitis, etc...) requires novel disease taxonomies based on a better knowledge of underlying mechanisms that may likely result in a re-definition of COPD as a pulmonary disease [24]. Such an approach should solve three current deficits, namely: *i)* overlap between COPD and lung ageing; *ii)* availability of operational diagnostic criteria differentiating COPD from other obstructive pulmonary diseases; and, *iii)* a proper capture of pulmonary heterogeneity of the disease. Another major challenge is to clarify the current confusion between systemic effects of COPD and some co-morbid conditions due to the descriptive nature of the reporting. As an example, is anxiety-depression a systemic effect ?, a COPD complication? or a co-morbid condition?. Likely, a proper understanding of the mechanisms involved in the relationships between systemic inflammation in COPD and depression [25,26] may help to clarify this type of questions. To prevent further confusion, Synergy-COPD was focused on the analysis of skeletal muscle dysfunction/wasting as a well accepted paradigm of a systemic effect of COPD. Last, but not least, a major challenge is a proper understanding of the phenomenon of co-morbidity clustering. There is evidence indicating that co-morbidity clustering is only partly explained by shared

**Table 1 Risk classification of COPD patients according to the 2011 GOLD Update [1]**

| RISK *GOLD Classification* | | | | | RISK *Exacerbation History* |
|---|---|---|---|---|---|
| | 3-4 | C<br>High Risk,<br>Less Symptoms | D<br>High Risk,<br>More Symptoms | ≥2 | |
| | 1-2 | A<br>Low Risk,<br>Less Symptoms | B<br>Low Risk,<br>More Symptoms | 0-1 | |
| | mMRC 0-1<br>CAT < 10 | | | mMRC ≥ 2<br>CAT ≥ 10 | |

The 2011 COPD Update [1] defines four risk categories for COPD patients (A to D) depending upon: *i)* symptoms (modified dyspnea score from th Medical Research Council, mMRC) or CAT questionnaire; *ii)* spirometric classification: GOLD I: $FEV_1 \geq 80\%$ pred; GOLD II: $50\% \leq FEV_1 < 80\%$ pred; GOLD III: $30\% \leq FEV_1 < 50\%$ pred; and, GOLD IV: $FEV_1 < 30\%$ and/or $PaO_2 < 60$ mmHg breathing $F_iO_2$ 0.21); and, *iii)* frequency of exacerbations per year. Recent reports have assessed the predictive value of this classification [16,17]

risk factors among the concurrent diseases [12,27,28], namely: tobacco smoking, unhealthy diet and sedentarism.

### The Synergy-COPD project

As stated in [29] of the current Supplement, the Synergy-COPD project was conceived as a systems approach to explore the role of combined modeling strategies to better understand COPD heterogeneity. Within the core aim of the project was the purpose of transferring the acquired knowledge into healthcare by designing innovative strategies to effectively build-up 4P Medicine: predictive, preventive, personalized and participatory medicine for chronic patients. In this regard, COPD was identified as a proper use case to explore the potential for generalization of the approach to other chronic conditions.

The current chapter focuses on the biomedical dimensions of the project taking into consideration the implications on the healthcare scenario. Figure 1 depicts relevant bio-pathological processes involved in COPD displayed according to the classical *"spill over"* hypothesis [30] to explain the systemic effects of the disease. Two main limitations of this hypothesis are; *i)* its over-simplistic explanation of the phenomenon of systemic low-grade inflammation, not confirmed by ECLIPSE [6,11] and other studies [28] and, *ii)* lack of a proper consideration of the co-morbidity challenge. An implicit assumption of the hypothesis is that COPD heterogeneity is ultimately driven by the pulmonary events of the disease.

Alternatively, Synergy-COPD proposed that COPD heterogeneity is explained by the interplay of conceptually independent events occurring at three different levels: *i) Pulmonary disease* - determined by the effects of lung injury and local remodeling processes; *ii)* **Systemic effects of the disease** with different manifestations, such as skeletal muscle dysfunction/wasting [10]; and, *iii)* **Co-morbidity clustering** that refers to observed associations of different chronic disorders. Synergy-COPD targeted the analysis of associations among CVD, T2DM-MS and COPD. The project explored underlying mechanisms of COPD heterogeneity focusing on skeletal muscle dysfunction/wasting and co-morbidity clustering and it only marginally addressed pulmonary events of the disease.

## Method

### Planned strategies for assessing COPD heterogeneities

The overall biomedical strategy of the project and the specific input datasets have been reported in detail in [29]. Moreover, the different modeling tools and strategies are described in [31]. The three biomedical areas addressed in Synergy-COPD had specific study designs that are summarized below:

***Skeletal muscle dysfunction*** - The project explored three relevant aspects of skeletal muscle dysfunction and muscle wasting in COPD patients. Firstly, Synergy-COPD examined the degree of association between classical COPD GOLD stages (I to IV) and estimations of both

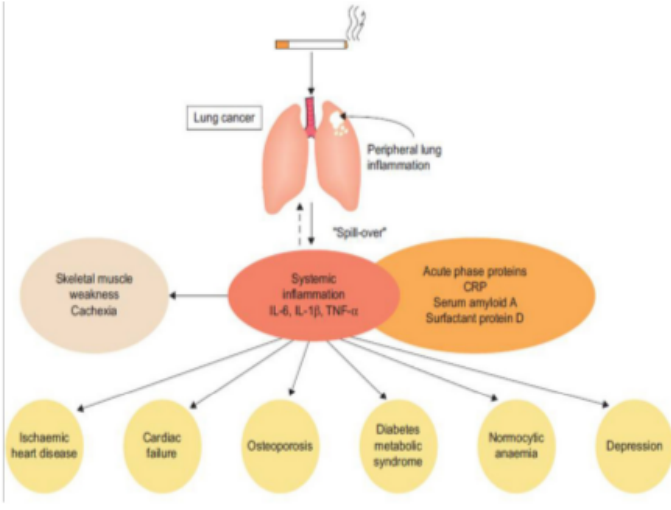



**Figure 1 Lung injury caused by inhaled irritants like tobacco smoking generates peripheral lung inflammation that may cause "spill over" of different types of cytokines into the systemic circulation**. According to this hypothesis, systemic inflammation causes skeletal muscle dysfunction and musde wasting, but it may also cause and worsen co-morbidities *(reproduced from* [30]*with permission)*

cellular oxygenation ($PmO_2$) and mitochondrial reactive oxygen species (ROS) levels in skeletal muscle exercising maximally ($VO_2max$). The study was done using a COPD dataset wherein $VO_2max$, cardiac output ($Q_T$), pulmonary ventilation-perfusion mismatching ($V_A/Q$ inequalities) and blood oxygenation, arterial and mixed venous blood, had been measured [32]. The analysis was carried out using the integrated deterministic model developed in the project wherein an integrated physiological $O_2$ pathway model was made interoperable with a biochemical model characterizing mitochondrial ROS generation, as reported in detail in [31]. Likewise, such an analysis was also done using the Biobridge dataset [33] wherein healthy subjects and COPD patients had a multilevel (omics, biochemical, physiological and clinical data) characterization of lower limb muscle, blood and whole-body changes from pre- to post- high intensity supervised resistance training during 8 weeks.

A second study in the project enriched the initial network medicine analysis [34-36]) from the Biobridge dataset [33] by including additional "omics" information [27,28], as well as an extended set of measurements on nitroso-redox balance carried out in blood and in skeletal muscle [28]. The purpose of the network approach described in detail in [31] was to compare healthy subjects and COPD patients in terms of the relationships among relevant metabolic pathways governing cellular bioenergetics, protein balance, and skeletal muscle remodeling paying particular attention to the role of nitroso-redox disequilibrium in the network modeling.

Finally, a third study undertaken within Synergy-COPD addressed the analysis of abnormal training adaptations comparing healthy subjects and COPD from the Biobridge database. Two different modeling strategies were undertaken: targeted probabilistic network modeling [34,37,38] and a Thomas network approach, as described in detail in [31]. In the former, associations between estimated skeletal muscle ROS levels obtained with the integrated deterministic model were compared with actual measurements pre- and post-training carried out both in blood and skeletal muscle.

**Co-morbidity analyses** - Two different studies were undertaken. Firstly, a data driven approach aimed at assessing different indices of relative risk for co-morbidity clustering in COPD patients aged +65 yrs compared to non-COPD patients. As reported in [29], the study was done using the Medicare dataset (13 million people) [29,31]. The research also identified genes and pathways associated with clusters of co-morbidities. A second independent study compared the outcomes of the data driven study with the pathway analysis of the co-morbidity clustering targeted in Synergy-COPD, namely: CVD, T2DM-MS and COPD as reported in [12,13]. The relevant pathways identified in the analysis of skeletal muscle dysfunction/wasting were compared with those seen in the co-morbidity clustering to explore commonalities.

**Pulmonary events** - In 2011, the PAC_COPD study [39] reported an unbiased cluster analysis identifying subtypes of COPD patients with clinical and prognostic implications. In the study [39], there was evidence of a dissociation between relatively low central airways resistance and high emphysema score in approximately one third of the patients. Because of the potential interest of the finding in terms of patient stratification, we used modeling techniques exploring spatial pulmonary heterogeneities to address the issue, as described in detail in [40]. The study was done in close collaboration with the AIRPROM project.

#### Health risk assessment and patient stratification

The design of strategies aiming at speeding the transfer of biomedical research achievements into clinical practice constitutes a core objective of Synergy-COPD. To this end, three specific goals were identified: *i)* to better understand the underlying biological mechanisms of the phenomena indicated above, namely: skeletal muscle dysfunction, co-morbidity clustering and dissociation between airway remodeling and emphysema score in COPD patients; *ii)* to identify combined markers with potential predictive power; and, *iii)* to construct patient-specific predictive modeling useful for the clinical decision making processes in primary care. The project outcomes from these three areas should help to generate rules feeding Clinical Decision Support Systems (CDSS) embedded into the clinical processes in primary care. The CDSS produced in the project targeted the areas displayed in Table 2. They were embedded into clinical processes supported by the Integrated Care Shared Knowledge Platform [41] deployed in the integrated healthcare district of Barcelona-Esquerra. Specific validation strategies were defined and executed, as reported in detail in [42].

## Results

#### Contributions to knowledge on COPD heterogeneities

Figure 2 summarizes four major aspects of the Synergy-COPD project: *(i)* main input data for the analyses, also described in detail in [29] and [43]; *(ii)* main biomedical analyses carried out during the project lifetime; *(iii)* novel resources generated from the developments done; and, finally, *(iv)* areas of impact from the project and recommendations to be done beyond the project life span.

**Skeletal muscle dysfunction in COPD patients**. Figure 3 displays the relationships between estimated skeletal muscle $PO_2$ ($PmO_2$) and mitochondrial ROS production for different levels of maximum $O_2$ transport and mitochondrial utilization capacity in a group of COPD patients with mild to severe disease [32], as explained in the figure legend. The central messages of

**Table 2 Clinical Decision Support Systems (CDSS) developed in Synergy-COPD for COPD management in an integrated care scenario.**

| |
|---|
| **1. Early diagnosis - COPD case finding program The suite of CDSS supports the regional deployment of a program of early COPD diagnosis targeting citizens at risk examined in pharmacy offices and non-diagnosed patients studied in primary care. Additional objectives of the program are to ensure high quality forced spirometry accessible across heathcare tiers, as well as prevention of over-diagnosis of COPD in elderly due to the GOLD diagnostic criteria [1].** |
| **2. Enhanced stratification of COPD patients It includes three families of CDSS with well differentiated objectives: *i)* enhance applicability of the 2011 GOLD Update criteria for COPD staging [1]; *ii)* facilitate off-line comparisons with other COPD staging criteria, namely: BODE, DOSE, ADO, etc...[18-21]; and, *iii)* enhanced stratification combining acquired knowledge in Synergy-COPD and consolidated findings from ECLIPSE [6].** |
| **3. Community-based integrated care program The suite of CDSS aims at supporting different integrated care services fostering the transfer of complexity from specialized care to the community with an active role of patients. The two programs being deployed are: *i)* sustainability of training-induced effects and promotion of physical activity; and, *ii)* management of patients under long-term oxygen therapy (LTOT). The two programs were assessed within NEXES [41,49], as part of the deployment of integrated care services in the health district of Hospital Clinic.** |

this analysis were: *i)* $PmO_2$ at maximal exercise is determined by the ratio between $O_2$ transport to mitochondrial and $O_2$ utilization capacity ($VO_2max/Vmax$ ratio), such that the lower the maximum $O_2$ transport potential for a given mitochondrial capacity, the lower $PmO_2$; *ii)* tissue oxygenation levels were not related with GOLD stages; and, *iii)* low $PmO_2$ values associated with abnormally high mitochondrial ROS production at peak exercise were predicted to occur in these patients. The analysis of the relative impact of the determinants of skeletal muscle oxygenation using the integrated model of $O_2$ pathway and mitochondrial ROS generation also indicated: (1) how functional heterogeneities of skeletal muscle $VO_2max/Vmax$ ratios may generate both very low and very high $PmO_2$ in the skeleletal muscle of these patients; and (2) the high impact of lung heterogeneity decreasing overall $O_2$ transport, as compared with the rather moderate role of skeletal muscle heterogeneity on mean $PmO_2$.

The Biobridge dataset [27,28,33] clearly indicated that COPD patients at rest, before training, showed nitroso-redox disequilibrium both in blood and skeletal muscle compared to healthy controls (Figure 4, upper panel). Moreover, a significant association of protein carbonylation levels between blood and skeletal muscle was observed in the patients (Figure 4, bottom panel) [28] in whom low-grade inflammation in peripheral blood, but not in skeletal muscle, was observed. The plasma

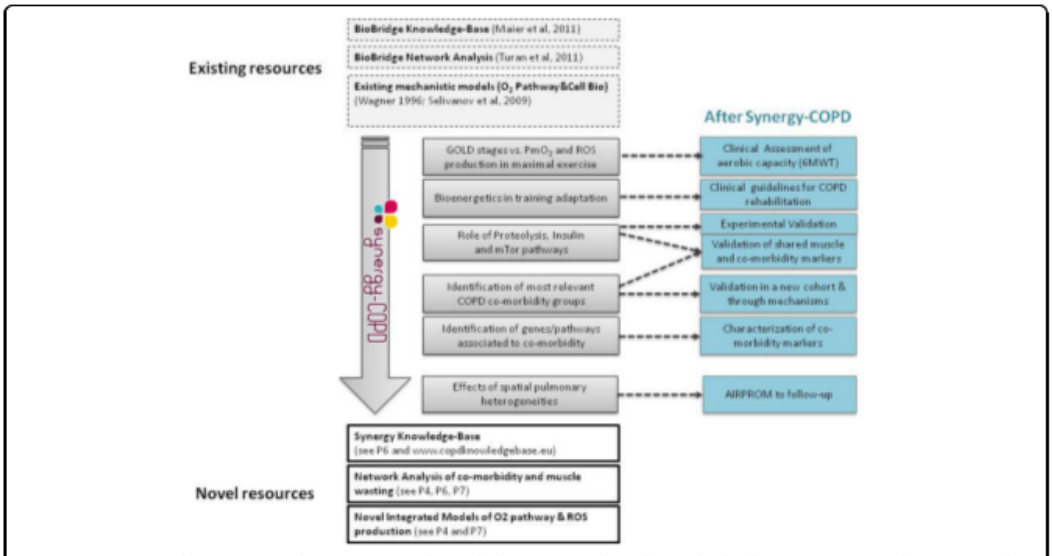



**Figure 2 Diagram indicating input data *(top rectangles with discontinuous line)*, biomedical achievements *(central grey rectangles)*, resources generated by the project *(bottom rectangles with continuous line)* and further developments to be considered after the Synergy-COPD project**. *(blue rectangles).*

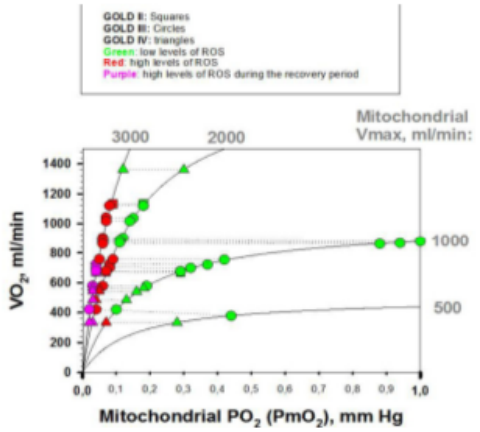



**Figure 3 Relationships between measured maximum $O_2$ transport ($VO_2$), y-axis; and, estimated cellular oxygenation ($PmO_2$), x-axis, in COPD patients**. The different symbols correspond to classical GOLD stages: squares, GOLD II; circles GOLD III; and, triangles, GOLD IV (measured $VO_2$ obtained from [32]). The symbols connected with discontinuous lines correspond to the same patient (same $VO_2$) with estimated $PmO_2$ values corresponding to different mitochondrial oxidative capacities (Vmax values and $VO_2$/Vmax ratios). For a given patient, the lower the $VO_2$/Vmax ratio, the lower was the estimated $PmO_2$. The colors correspond to the mitochondrial ROS generation: green, mitochondrial ROS levels similar to those seen in healthy subjects; red, abnormally high mitochondrial ROS levels; and, violet, high ROS levels that persist after exercise withdrawal. The lower the $PmO_2$, the higher were mitochondrial ROS levels.

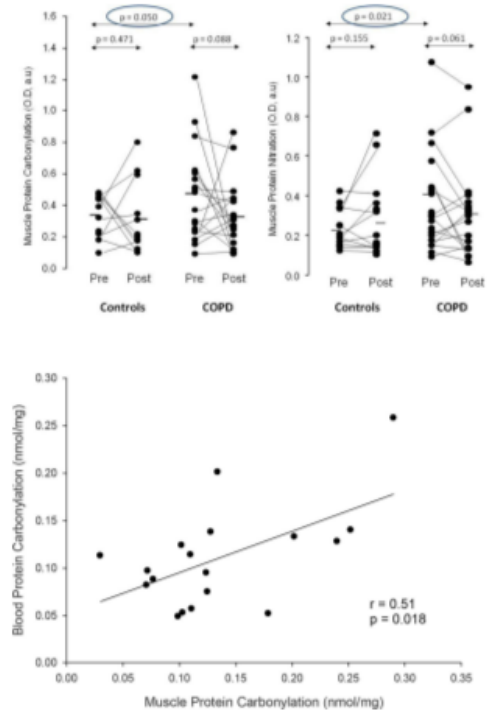



**Figure 4 Oxidative stress in COPD. Upper panel**: Muscle oxidative stress. Individual and mean group effects of an 8-week endurance training program on protein carbonylation (left) and protein nitration (right) in the vastus lateralis of healthy subjects (controls) and patients with COPD. At baseline (rest, pre-training measurements), COPD patients showed higher nitroso-redox disequilibrium than healthy subjects. A trend toward a decrease in oxidative stress was observed after training in COPD patients [28]. **Bottom panel**: Association between muscle and blood. COPD patients at baseline (rest, pre-training) showed an association of protein carbonylation levels between skeletal muscle and blood [28] (*reproduced from*[28]*with permission*).

metabolomic analysis in COPD was fully consistent with abnormal skeletal muscle abnormalities reported in these patients [27], namely: decreased oxidative capacity leading to abnormal ROS generation [28,44,45], up-regulation of glycolysis [46] and altered aminoacid metabolism [27,46] (Figure 5). The transcriptomic analysis showed lower and abnormal skeletal muscle gene expression at baseline in COPD patients compared to healthy subjects with clear differences between COPD patients with preserved skeletal muscle mass and those showing muscle wasting [33].

Most importantly, the network medicine approach assessing associations among three major pathways, namely bioenergetics, inflammation and skeletal muscle remodeling showed clear differences between the probabilistic models obtained in healthy subjects and those seen in COPD patients indicating failure to coordinately activate these pivotal metabolic pathways in the patients (Figure 6). Moreover, a sub-analysis carried out with few muscle samples identified a potential role of epigenetic changes contributing to the phenomenon of abnormal regulation of key metabolic pathways seen in COPD patients.

The study [27,28] demonstrated that high intensity 8-w endurance training significantly enhanced aerobic capacity both in healthy controls and in COPD patients without harmful effects on nitroso-redox equilibrium in severe COPD patients, but it confirmed abnormal training-induced skeletal muscle adaptations of the redox system in COPD indicated by a poor post-training increase of total gluthatione to oxidized gluthatione ratio in skeletal muscle seen in healthy subjects. Ongoing network analyses using the Biobridge dataset [27,28,33] further supports a pivotal role of nitroso-redox disequilibrium on skeletal muscle dysfunction in COPD patients.

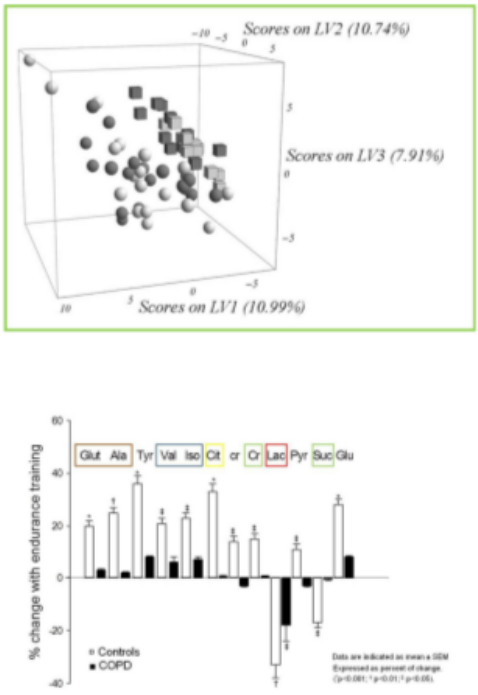



**Figure 5 Metabolic analysis**. **Upper panel**: Resting individual metabolic profiles in COPD patients (spheres) and in healthy sedentary subjects (cubes), including pre (black symbols) - and post -training data (grey symbols). The results are expressed by the three Latent Variables (LV1, 2 and 3) of the partial-least square discriminant analysis (PLS-DA). The percentages indicate the magnitude of the differences between the two groups of subjects for each dimension ($p<0.05$). **Bottom panel**: Endurance training responses of individual metabolites. Mean training-induced responses of individual metabolites. Data expressed as percent of change are indicated as mean ± SEM. (* $p<0.001$; † $p<0.01$; ‡ $p<0.05$) *(reproduced from [27]with permission)*

**Co-morbidity clustering** - The large data driven analysis of co-morbidities indicated that COPD patients showed a higher risk for co-morbidity clustering than non-COPD patients of the same age. Not surprisingly, the likelihood for co-morbidity occurrence in COPD patients significantly increased with age for most conditions. Moreover, it was shown that specific cytokines and variables associated to the redox system [28,46] presented significant relationships with co-morbidity clustering in COPD patients. The network medicine analysis of the targeted co-morbidity clustering including CVD and T2D-MS is currently underway.

**Dissociation between airway remodeling and pulmonary emphysema** - The modeling of spatial pulmonary heterogeneities was used to explore the characteristics of patients from the PAC-COPD study [39] in whom moderate to severe emphysema score, assessed by high resolution CT scan, was not accompanied by significant central airway remodeling. Consequently, these COPD patients showed mild $FEV_1$ impairment. Unfortunately, the maturity of the modeling developments did not allow completion of the analysis as initially planned, see [40] for further details.

### Tackling COPD heterogeneities to enhance health outcomes

As part of the strategies for transferring novel biomedical knowledge into the clinical arena, we developed three families of CDSS (Table 2) that were embedded into the clinical processes at primary care level using an Integrated Care Shared Knowledge Platform [41] as technological support to facilitate the management of chronic patients. The CDSS' rules combined existing and novel knowledge generated during the project's life-span, as explained in detail in [42]. The CDSS areas addressed in the project are briefly described below:

**Early diagnosis of COPD** - A COPD case-finding program for citizens with high risk for developing COPD was deployed at local level [41]. The program encompasses different aspects: *i)* remote support to automatic assessment of quality of forced spirometry both in pharmacy offices and primary care [47]; *ii)* support to coordination between informal care (pharmacy offices) and formal care (primary care and specialists); and, *iii)* enhanced 2011 GOLD-based COPD assessment with the use of recommended reference equations [48].

**Enhanced stratification of COPD patients**. A suite of CDSS supporting proposals for patient stratification has been built-up. The CDSS facilitates comparisons with available predictive indices for COPD patients. That is, classical GOLD staging (groups 1 to 4 based on $FEV_1$ % predicted) and 2011 GOLD update (groups A to D), taking into account symptoms score and frequency of severe exacerbations, can be automatically constructed and displayed in the primary care clinical workstation. Co-morbidities expressed as number&type and Charlson index are also considered together with novel proposals for decision algorithms based on the Synergy-COPD findings, as detailed in [42].

**Community-based integrated care management of COPD patients**. Deployment experiences of integrated care services [41] developed in parallel with Synergy-COPD have demonstrated positive health outcomes together with cost-containment through the transfer of healthcare complexity from specialized care to the community fostering an active and participatory role of both citizens at risk, patients and carers. In this scenario, the use of CDSS to support health professionals for chronic

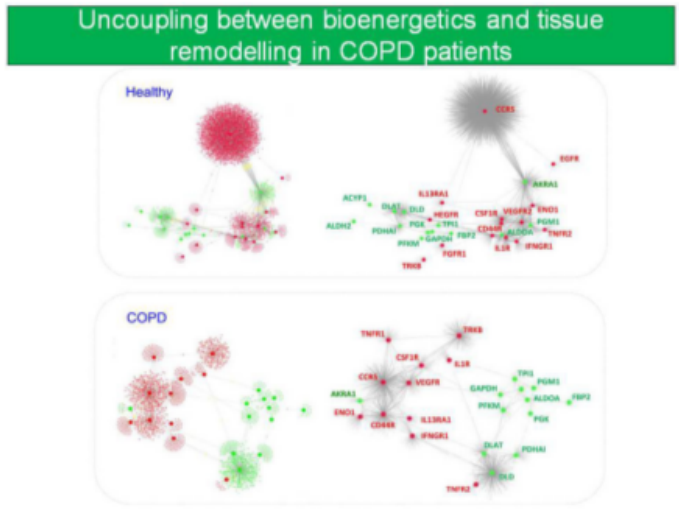



**Figure 6 Interaction networks using skeletal muscle expression profiling, plasma cytokines and physiological measurements.** Uncoupling between bioenergetics, inflammation and skeletal muscle remodeling was observed in COPD patients as compared to healthy subjects [33]

care management appears as an effective approach to transfer novel biomedical knowledge into healthcare. Such an approach was assessed in the validation work package of the project. Moreover, the parallel deployment experiences [41] carried out during the life time of the project identified the high potential of the Personal Health Folder (PHF) [49] for transferring different types of non-medical patient information, namely: life-styles, social frailty, adherence profile, etc... into formal healthcare, as detailed in [50].

#### Lessons learnt to foster 4P Medicine of chronic conditions

Synergy-COPD proposes a comprehensive strategy to foster the interplay between system research and predictive medicine for chronic conditions. This purpose led to the inception of the Digital Health Framework (DHF) extensively presented in [50] of the current Supplement, as the Synergy-COPD's proposal to favor iterations between informal care, formal care and biomedical research.

## Discussion

#### Biomedical contributions to COPD heterogeneity

The overall biological findings generated by the different studies associated with the project support the concept that the pulmonary events of the disease are not the only driver, and sometimes neither the main one, of the disturbances seen in these patients and do not explain by themselves COPD heterogeneity. Instead, a constellation of concurrent factors (systemic effects of tobacco, lack of physical activity, unhealthy diet, patient susceptibility to oxidative stress, cellular hypoxia, etc...) may lead to systemic nitroso-redox disequilibrium and low-grade blood inflammation seen in COPD patients with skeletal muscle dysfunction/wasting, as well as in those with co-morbidities.

There is evidence supporting that the combination of the subject's genetic susceptibility, changes in functional genomics, partly regulated by epigenetic events, together with changes in post-translational regulatory phenomena at several levels may modulate the impact of the factors triggering the disease and ultimately may determine the clinical manifestations, as well as the rate of progression of the different COPD components: pulmonary, systemic effects and co-morbidity clustering. Moreover, it can be speculated that the tight multilevel regulation of aerobic capacity in man [26], may constitute a relevant factor explaining the interdependence among the three drivers of COPD heterogeneity. As a simplified example, down-regulation of the skeletal muscle tricarboxilic acid cycle (TCA) metabolism together with up-regulation of glycolysis have been extensively reported in COPD patients with skeletal muscle dysfunction [10] as a leading cause of early lactate release during moderate exercise. The phenomenon overloads the ventilatory function facilitating air trapping and shortness of breath often contributing to sedentarism which, in turn, may constitute one of the contributing factors to insulin resistance, and T2DM, often seen in these patients. Obviously, the progression of pulmonary disease severity contributes to the vicious circle because it worsens

the underlying mechanisms involved in the systemic effects of COPD and the co-morbidity clustering.

The unifying hypothesis proposed by Synergy-COPD is supported by all the findings observed in the different studies carried out in the project. Moreover, it is consistent with different reports in the literature [26,44] indicating that preserved aerobic function seems to be associated with nitroso-redox equilibrium in the cardiovascular system and, consequently, with $O_2$ flow to $O_2$ uptake matching in peripheral tissues. All in all, the novel view of COPD heterogeneity proposed by the Synergy-COPD project has profound implications in terms of patient health risk assessment and stratification, but most importantly, it may deeply modulate both management and therapeutic approaches of these patients.

The 2014 report [6] analyzing the clinical implications of the ECLIPSE results strengthen the relevance of COPD heterogeneity on the clinical management of these patients. However, the highly valuable ECLIPSE's achievements favoring a new clinical vision of COPD essentially have a descriptive nature with well recognized limitations in terms of generating evidence on novel underlying disease mechanisms.

In summary, there are three relevant lessons learnt both from the project itself and from recent contributions like ECLIPSE:

- The characteristics of COPD heterogeneity clearly generate a mandate for the design of strategies aiming at individual health risk assessment and patient-oriented management stratification aiming at setting cost-effective preventive interventions to modulate disease progress.
- The unifying hypothesis for COPD heterogeneity explored in Synergy-COPD contains the core elements for the design of a coherent patient stratification strategy, as proposed in the current manuscript.
- The still on-going network analyses in Synergy-COPD may generate "in silico" hypothesis on the specifics of underlying mechanisms of COPD heterogeneity to be implemented in further refined versions of the proposed CDSS.

### Modeling tools and strategies

The project has generated important modeling outcomes and it has identified relevant challenges to be faced beyond the project lifetime. Moreover, Synergy-COPD has also shaped specific approaches to face those challenges, as extensively discussed in [31]. We acknowledge, however, limited achievements in terms of identification of combined biomarkers with predictive power, as well as in the development of subject-specific predictive modeling to feed the CDSS for the reasons discussed in other areas of the Supplement. We also recognized limited achievements in the targeted objectives relative to pulmonary events. There is no doubt that the positive interactions with AIR-PROM will contribute to a maturity of the modeling of pulmonary spatial heterogeneities described in [40] allowing to further undertake this specific challenge.

### Logistics for 4P medicine

The accepted limitations in terms of subject-specific predictive modeling did not preclude other relevant technological and organizational outcomes such as the developments of novel CDSS [42] and the formulation of the Digital Health Framework (DHF) [50]. We believe that the deployment of these tools within an integrated care scenario paves the way toward predictive, preventive, participatory and personalized (4P) medicine for these patients preventing fragmentation of care. It is important to note that the entire DHF [50] still requires a proof-of-concept validation before considering specific strategies for its deployment.

The transition toward a novel biomedical research scenario fostering 4P medicine has two major biomedical research goals, namely: *i)* to speed-up the transfer of biomedical knowledge, including novel therapies, into healthcare; and. *ii)* to generate operational feedback from healthcare and informal care into biomedical research. The last step shall generate two main added values. Firstly, biological knowledge will be enriched with information on different dimensions of the patient (adherence profile, frailty, life styles, socio-economical and environmental factors, etc...) and, secondly, it will facilitate an iterative process that shall result in progressive refinement of subject-specific predictive modeling. In this regard, the interoperability among the PHF, the Integrated Care Shared Knowledge Platform and the novel biomedical research platform proposed in [50], within the concept of the DHF, constitutes a major achievement of the project toward the consolidation of innovative biomedical research scenario that overcomes current limitations due to fragmentation of the information.

## Conclusions

The systems approach to COPD heterogeneity explored in the Synergy-COPD project has generated a novel view of the phenomenon wherein systemic effects of the disease and co-morbidity clustering, may all have a relevant role in the COPD patients, independently from pulmonary events. The interdependence between pulmonary and extra-pulmonary events was formulated. Moreover, the chapter assessed specific strategies for implementation of the system approach into integrated care management for chronic patients. Finally, the impact of the novel vision into biomedical research was explored.

**Competing interests**
The authors declare they have no competing interests.

**Acknowledgements**
We would like to thank all the members of the Synergy-COPD consortium for their support. Part of the data sets supporting this article are available in the GEO repository (e.g. http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27536) and in the HuDiNe repository (http://barabasilab.neu.edu/projects/hudine/resource/data/data.html). For the rest of data-sets we had permission to access and use the data.

**Declarations**
The research described in this paper is partly supported by the Synergy-COPD European project (FP7-ICT-270086). Publication of this article has been funded by the Synergy-COPD European project (FP7-ICT-270086).
This article has been published as part of *Journal of Translational Medicine* Volume 12 Supplement 2, 2014: Systems medicine in chronic diseases: COPD as a use case. The full contents of the supplement are available online at http://www.translational-medicine.com/supplements/12/S2.

**Authors' details**
[1]IDIBAPS, Hospital Clínic. Facultat de Medicina, 08036, Barcelona, Catalunya, Spain. [2]Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Bunyola, Balearic Islands. [3]Pulmonology Department-Muscle and Respiratory System Research Unit (URMAR), IMIM-Hospital del Mar, Health and Experimental Sciences Department (CEXS), Universitat Pompeu Fabra (UPF), Parc de Recerca Biomédica de Barcelona (PRBB), Barcelona, Catalonia, Spain. [4]Biomax Informatics AG, 282152 Planegg, Germany. [5]Centre for Computational Biology and Modeling (CCBM). Institute of Integrative Biology. University of Liverpool.Crown Street.L69 7ZB.UK. [6]Division of Physiology, Pulmonary and Critical Care Medicine, University of California, San Diego, La Jolla, CA, USA. [7]Departament de Bioquimica i Biologia Molecular i IBUB, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain. [8]Centre for Research in Environmental Epidemiology (CREAL). CIBER Epidemiología y Salud Pública (CIBERESP). Universitat Pompeu Fabra, Departament de Ciències Experimentals i de la Salut Barcelona, Spain. [9]Unit of Computational Medicine, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.

Published: 28 November 2014

**References**

1. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R: **Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary.** *American journal of respiratory and critical care medicine* 2013, **187**:347-365.
2. Murray CJL, Lopez AD: **Measuring the Global Burden of Disease.** *New England Journal of Medicine* 2013, **369**:448-457.
3. WHO: **2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases**.ISBN: 9789241597418 2008 [http://www.who.int/nmh/publications/9789241597418/en/], Accessed: 2013-08-20. (Archived by WebCite® at http://www.webcitation.org/6J0XGnHcJ.
4. Agusti A, Calverley P, Celli B, Coxson H, Edwards L, Lomas D, MacNee W, Miller B, Rennard S, Silverman E: **Characterisation of COPD heterogeneity in the ECLIPSE cohort.** *Respiratory Research* 2010, **11**:122.
5. Casanova C, de Torres JP, Aguirre-Jaime A, Pinto-Plata V, Marin JM, Cordoba E, Baz R, Cote C, Celli BR: **The progression of chronic obstructive pulmonary disease is heterogeneous: the experience of the BODE cohort.** *American journal of respiratory and critical care medicine* 2011, **184**:1015-1021.
6. Vestbo J, Agusti A, Wouters EF, Bakke P, Calverley PM, Celli B, Coxson H, Crim C, Edwards LD, Locantore N, Lomas DA, MacNee W, Miller B, Rennard SI, Silverman EK, Yates JC, Tal-Singer R, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints Study Investigators: **Should we view chronic obstructive pulmonary disease differently after ECLIPSE? A clinical perspective from the study team.** *American journal of respiratory and critical care medicine* 2014, **189**:1022-1030.
7. Soler-Cataluna JJ, Rodriguez-Roisin R: **Frequent chronic obstructive pulmonary disease exacerbators: how much real, how much fictitious?** *Copd* 2010, **7**:276-284.
8. Hurst J, Vestbo J, Anzueto A, Locantore N, Mullerova H, Tal-Singer R, Miller B, Lomas D, Agusti A, Macnee W: **Susceptibility to exacerbation in chronic obstructive pulmonary disease.** *The New England journal of medicine* 2010, **363**:1128-1138.
9. Suissa S, Dell'Aniello S, Ernst P: **Long-term natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality.** *Thorax* 2012, **67**:957-963.
10. Maltais F, Decramer M, Casaburi R, Barreiro E, Burelle Y, Debigare R, Dekhuijzen PN, Franssen F, Gayan-Ramirez G, Gea J, Gosker HR, Gosselink R, Hayot M, Hussain SN, Janssens W, Polkey MI, Roca J, Saey D, Schols AM, Spruit MA, Steiner M, Taivassalo T, Troosters T, Vogiatzis I, Wagner PD, ATS/ERS Ad Hoc Committee on Limb Muscle Dysfunction in COPD: **An official american thoracic society/european respiratory society statement: update on limb muscle dysfunction in chronic obstructive pulmonary disease.** *American journal of respiratory and critical care medicine* 2014, **189**:e15-62.
11. Miller J, Edwards LD, Agusti A, Bakke P, Calverley PM, Celli B, Coxson HO, Crim C, Lomas DA, Miller BE, Rennard S, Silverman EK, Tal-Singer R, Vestbo J, Wouters E, Yates JC, MacNee W, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators.: **Comorbidity, systemic inflammation and outcomes in the ECLIPSE cohort.** *Respiratory medicine* 2013, **107**:1376-1384.
12. Vanfleteren LE, Spruit MA, Groenen M, Gaffron S, van Empel VP, Bruijnzeel PL, Rutten EP, Op 't Roodt J, Wouters EF, Franssen FM: **Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease.** *American journal of respiratory and critical care medicine* 2013, **187**:728-735.
13. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, Zulueta J, Cabrera C, Zagaceta J, Hunninghake G, Celli B: **Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease.** *American journal of respiratory and critical care medicine* 2012, **186**:155-161.
14. Fabbri LM, Beghe B, Agusti A: **COPD and the solar system: introducing the chronic obstructive pulmonary disease comorbidome.** *American journal of respiratory and critical care medicine* 2012, **186**:117-119.
15. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J: **Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary.** *American journal of respiratory and critical care medicine* 2007, **176**:532-555.
16. Calverley PMA: **The ABCD of GOLD made clear.** *European Respiratory Journal* 2013, **42**:1163-1165.
17. Vogelmeier C, Vestbo J: **COPD assessment: I, II, III, IV and/or A, B, C, D.** *European Respiratory Journal* 2014, **43**:949-950.
18. Celli B, Cote C, Marin J, Casanova C, Montes De Oca M, Mendez R, Pinto Plata V, Cabral H: **The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease.** *The New England journal of medicine* 2004, **350**:1005-1012.
19. Puhan Ma, Garcia-Aymerich J, Frey M, ter Riet G, Antó JM, Agustí AG, Gómez FP, Rodríguez-Roisín R, Moons KG, Kessels AG: **Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index.** *The Lancet* 2009, **374**:704-711.
20. Puhan MA, Hansel NN, Sobradillo P, Enright P, Lange P, Hickson D, Menezes AM, Riet Gt, Held U, Domingo-Salvany A, Mosenifar Z, Antó JM, Moons KG, Kessels A, Garcia-Aymerich J, International COPD Cohorts Collaboration Working Group: **Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts.** *BMJ Open* 2012, **2**.
21. Jones RC, Donaldson GC, Chavannes NH, Kida K, Dickson-Spillmann M, Harding S, Wedzicha JA, Price D, Hyland ME: **Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index.** *American journal of respiratory and critical care medicine* 2009, **180**:1189-1195.
22. Motegi T, Jones RC, Ishii T, Hattori K, Kusunoki Y, Furutate R, Yamada K, Gemma A, Kida K: **A comparison of three multidimensional indices of**

**COPD severity as predictors of future exacerbations.** *International journal of chronic obstructive pulmonary disease* 2013, **8**:259-271.
23. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, Hagan G, Knobil K, Lomas DA, MacNee W, Silverman EK, Tal-Singer E, ECLIPSE investigators: **Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE).** *The European respiratory journal* 2008, **31**:869-873.
24. Wardlaw AJ, Silverman M, Siva R, Pavord ID, Green R: **Multi-dimensional phenotyping: towards a new taxonomy for airway disease.** *Clinical & Experimental Allergy* 2005, **35**:1254-1262.
25. Dustin ML: **Signaling at neuro/immune synapses.** *Journal of Clinical Investigation* 2012, **122**:1149-1155.
26. Poole DC, Hirai DM, Copp SW, Musch TI: **Muscle oxygen transport and utilization in heart failure: implications for exercise (in)tolerance.** *American journal of physiology Heart and circulatory physiology* 2012, **302**:H1050-1063.
27. Rodriguez DA GA-V, Diaz-Moralli S, Reed M, Gomez FP, Falciani F, Günther U, Roca J, Cascante M: **Plasma metabolic profile in COPD patients: effects of exercise and endurance training.** *Metabolomics* 2012, **8**:508-516.
28. Rodriguez DA, Kalko S, Puig-Vilanova E, Perez-Olabarria M, Falciani F, Gea J, Cascante M, Barreiro E, Roca J: **Muscle and blood redox status after exercise training in severe COPD patients.** *Free radical biology & medicine* 2012, **52**:88-94.
29. Gomez-Cabrero D L-AM, Tegner J, Cascante M, Miralles F, Roca J, the Synergy-COPD consortium: **Synergy-COPD: A systems approach for understanding and managing Chronic Diseases.** *BMC Journal of Translational Medicine* 2014.
30. Barnes PJ, Celli BR: **Systemic manifestations and comorbidities of COPD.** *European Respiratory Jornal* 2009, **33**:1165-1185.
31. Gomez-Cabrero D MJ, Cano I, Abugessaisa I, Huertas-Migueláñez M, Tenyi A, Marin de Mas I, Kiani N, Marabita F, Falciani F, Burrowes K, Maier D, Wagner P, Selivanov V, Cascante M, Roca J, Barabási A, Tegnér J: **Systems Medicine: from molecular features and models to the clinic in COPD.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S4.
32. Blanco I, Gimeno E, Munoz PA, Pizarro S, Gistau C, Rodriguez-Roisin R, Roca J, Barbera JA: **Hemodynamic and gas exchange effects of sildenafil in patients with chronic obstructive pulmonary disease and pulmonary hypertension.** *American journal of respiratory and critical care medicine* 2010, **181**:270-278.
33. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Curnow SJ, Rodriguez DA, Cascante M, O'Neill L, Egginton S, Roca J, Falciani F: **A Systems Biology Approach Identifies Molecular Networks Defining Skeletal Muscle Abnormalities in Chronic Obstructive Pulmonary Disease.** *PLoS Comput Biol* 2011, **7**:e1002129.
34. Handy DE, Loscalzo J, Leopold JA: **Systems analysis of oxidant stress in the vasculature.** *IUBMB life* 2013, **65**:911-920.
35. Zhou X, Menche J, Barabasi AL, Sharma A: **Human symptoms-disease network.** *Nature communications* 2014, **5**:4212.
36. Barabasi AL: **Network medicine--from obesity to the "diseasome".** *The New England journal of medicine* 2007, **357**:404-407.
37. Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease.** *Nat Rev Genet* 2011, **12**:56-68.
38. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabasi AL: **The implications of human metabolic network topology for disease comorbidity.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:9880-9885.
39. Garcia-Aymerich J, Gomez FP, Benet M, Farrero E, Basagana X, Gayete A, Pare C, Freixa X, Ferrer J, Ferrer A, *et al*: **Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes.** *Thorax* 2011, **66**:430-437.
40. Burrowes Kelly S DT, Brightling C: **Computational modeling of the obstructive lung diseases asthma and COPD.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S5.
41. NEXES: **Supporting Healthier and Independent Living for Chronic Patients and Elderly.**2008-2013, CIP-ICT-PSP-225025.
42. Velickovski FCL, Roca J, Burgos F, Galdiz J, Nueria M, Lluch Ariet M: **Clinical Decision Support Systems (CDSS) for preventive management of COPD patients.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S9.
43. Cano ITA, Schueller CH, Wolff M, Huertas M, Gomez-Cabrero D, Antczak PH, Roca J, Cascante M, Falciani F, Maier D: **The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S6.
44. Barreiro E, Rabinovich R, Marin-Corral J, Barberà JA, Gea J, Roca J: **Chronic endurance exercise induces quadriceps nitrosative stress in patients with severe COPD.** *Thorax* 2009, **64**:13-19.
45. Rabinovich R, Bastos R, Ardite E, Llinàs L, Orozco-Levi M, Gea J, Vilaró J, Barberà Ja, Rodríguez-Roisin R, Fernández-Checa JC, Roca J: **Mitochondrial dysfunction in COPD patients with low body mass index.** *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology* 2007, **29**:643-650.
46. van den Borst B, Slot IG, Hellwig VA, Vosse BA, Kelders MC, Barreiro E, Schols AM, Gosker HR: **Loss of quadriceps muscle oxidative phenotype and decreased endurance in patients with mild-to-moderate COPD.** *Journal of applied physiology* 2013, **114**:1319-1328, (Bethesda, Md: 1985).
47. Burgos F, Disdier C, de Santamaria EL, Galdiz B, Roger N, Rivera ML, Hervas R, Duran-Tauleria E, Garcia-Aymerich J, Roca J: **Telemedicine enhances quality of forced spirometry in primary care.** *European Respiratory Journal* 2012, **39**:1313-1318.
48. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MS, Zheng J, Stocks J: **Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations.** *The European respiratory journal* 2012, **40**:1324-1343.
49. Barberan-Garcia A, Vogiatzis I, Solberg HS, Vilaro J, Rodriguez DA, Garasen HM, Troosters T, Garcia-Aymerich J, Roca J: **Effects and barriers to deployment of telehealth wellness programs for chronic patients across 3 European countries.** *Respiratory medicine* 2014, **108**:628-637.
50. Cano I L-AM, Gomez-Cabrero D, Maier D, Kalko S, Cascante M, Tégner J, Miralles F, Herrera D, Roca J, the Synergy COPD consortium: **Biomedical Research in a Digital Health Framework.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S10.

**doi:10.1186/1479-5876-12-S2-S3**
**Cite this article as:** Roca *et al.*: **Chronic Obstructive Pulmonary Disease heterogeneity: challenges for health risk assessment, stratification and management.** *Journal of Translational Medicine* 2014 **12**(Suppl 2):S3.

# Apendix IV

Gomez-Cabrero et al. Journal of Translational Medicine 2014, 12(Suppl 2):S4
http://www.translational-medicine.com/content/12/S2/S4

JOURNAL OF TRANSLATIONAL MEDICINE

RESEARCH Open Access

# Systems Medicine: from molecular features and models to the clinic in COPD

David Gomez-Cabrero[1†], Jörg Menche[2,3†], Isaac Cano[4], Imad Abugessaisa[1], Mercedes Huertas-Migueláñez[5], Akos Tenyi[6], Igor Marin de Mas[6], Narsis A Kiani[1], Francesco Marabita[1], Francesco Falciani[7], Kelly Burrowes[8], Dieter Maier[9], Peter Wagner[10], Vitaly Selivanov[6], Marta Cascante[6], Josep Roca[4], Albert-László Barabási[2,3,11,12,13*†], Jesper Tegnér[1*†]

## Abstract

**Background and hypothesis:** *Chronic Obstructive Pulmonary Disease* (COPD) patients are characterized by heterogeneous clinical manifestations and patterns of disease progression. Two major factors that can be used to identify COPD subtypes are muscle dysfunction/wasting and co-morbidity patterns. We hypothesized that COPD heterogeneity is in part the result of complex interactions between several genes and pathways. We explored the possibility of using a Systems Medicine approach to identify such pathways, as well as to generate predictive computational models that may be used in clinic practice.

**Objective and method:** Our overarching goal is to generate clinically applicable predictive models that characterize COPD heterogeneity through a Systems Medicine approach. To this end we have developed a general framework, consisting of three steps/objectives: (1) feature identification, (2) model generation and statistical validation, and (3) application and validation of the predictive models in the clinical scenario. We used muscle dysfunction and co-morbidity as test cases for this framework.

**Results:** In the study of muscle wasting we identified relevant features (genes) by a network analysis and generated predictive models that integrate mechanistic and probabilistic models. This allowed us to characterize muscle wasting as a general de-regulation of pathway interactions. In the co-morbidity analysis we identified relevant features (genes/pathways) by the integration of gene-disease and disease-disease associations. We further present a detailed characterization of co-morbidities in COPD patients that was implemented into a predictive model. In both use cases we were able to achieve predictive modeling but we also identified several key challenges, the most pressing being the validation and implementation into actual clinical practice.

**Conclusions:** The results confirm the potential of the Systems Medicine approach to study complex diseases and generate clinically relevant predictive models. Our study also highlights important obstacles and bottlenecks for such approaches (e.g. data availability and normalization of frameworks among others) and suggests specific proposals to overcome them.

* Correspondence: alb@neu.edu; jesper.tegner@ki.se
† Contributed equally
[1]Unit of computational Medicine, Center for Molecular Medicine, Department of Medicine, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden
[2]Center for Complex Network Research, Northeastern University Physics Department, Boston, MA 02115, USA
Full list of author information is available at the end of the article

## Introduction

Recent years have seen a paradigm shift in Life Sciences: *"from a fragmented to a systems approach, linear to non-linear methodology and from genome to physiome based analysis"* [1]. Systems Medicine, as an adaptation and extension of *Systems Biology*, embraces this paradigm and is becoming a cornerstone in the study of complex diseases. A general introduction to Systems Medicine is provided in [2]. In this article, part of a Supplement dedicated to the Synergy-COPD project [2], we review and assess the Synergy-COPD's Systems Medicine approach to study Chronic Obstructive Pulmonary Disease (COPD), both a chronic and a complex disease.

While the characterization of COPD has been extensively investigated and there is a continuous refinement of guidelines (e.g. GOLD), there is yet no consensus on a phenotypic definition of the term "COPD patient". For instance in [3] several sub-types of COPD patients were identified; see also [4] within this Supplement for further details on the heterogeneity in COPD. Briefly, within Synergy-COPD we aim to characterize two sources of heterogeneity: first we investigated the systemic effects associated with skeletal muscle dysfunction in COPD patients (*MusclDYS*). Second, we aimed to characterize co-morbidity patterns of COPD patients (*CoMorb*). Finally, we also investigated the interplay between the different heterogeneities, which may provide a novel description of COPD as being driven by the interaction between several factors.

Using COPD as a case-study we explore the notion that Systems Medicine provides tools to investigate and characterize disease heterogeneity. To this end our analyses follow a general three-step procedure: (1) first, we need to identify the relevant biomarkers (or more generally *features of interest, FoI*) for each case of heterogeneity; (2) in a second step, predictive models with the potential to be applied in the clinic are developed and validated statistically. Third and final, (3) the usefulness of the models in a clinical scenario has to be validated. To achieve this goal we integrated a wide variety of available resources, such as prior domain knowledge, relevant data-bases and existing probabilistic and mechanistic models.

The rest of this article is organized as follows: the next section details the Systems Medicine framework used in Synergy-COPD. The third and fourth sections describe the application of this framework in the characterization of *MusclDYS* and *CoMorb*. These sections include a brief description of the questions, the obtained results and the limitations of the proposed methodology. The final section provides the conclusions and summarizes the identified remaining challenges.

## The Synergy-COPD's systems medicine approach

Systems Medicine provides a comprehensive and general framework to investigate the complex interactions implicated in human disease in an integrated fashion. Consequently, there is no single defined set of methodologies associated with Systems Medicine. Instead, any methodology useful to investigate the question under study "as a system" can be considered as relevant to explore and validate. While the concrete focus of Synergy-COPD lay on COPD, we aimed at developing a more general framework that may also be applicable to other complex diseases. We therefore started by defining a generic three-step objective plan that sets the goals in our studies of *MusclDys* and *CoMorb* (Figure 1). The plan was then concretized and adapted to each question accordingly.

Our final goal was a characterization of the disease heterogeneities that can be transferred to clinical practice (Figure 1, Objective 3), in particular by implementing it into a Clinical Decision Support system (CDSS, [5]). The first step (Objective 1) towards this goal is to identify the relevant, *i.e.* most predictive, biological features among the large amount of available data, *e.g.* genes, metabolites and clinical variables, among many others. The second step (Objective 2) is to integrate these features into predictive models and to validate them. In the following we briefly review each objective for the two case studies *MusclDys* and *CoMorb* and introduce the respective different resources and methodologies that were used.

**Objective 1**, (Biomarker identification): having defined a question of interest (e.g. *MusclDys*) we first need to identify the relevant associated features. The core of this Objective is formed by publicly available data-sets and knowledge (e.g. Gene Omnibus [6]) that were integrated into a user-friendly knowledge-base [7]. The different methodologies for *MusclDys* and *CoMorb* are detailed in two separate sections.

**Objective 2 (Predictive modeling)**: the identified features are now used in predictive models that may provide insights into the question of interest. For instance, in *MusclDys* we aim to predict the effects of muscle dysfunction in a given patient. In *CoMorb* we aim to compute the probability of developing a specific co-morbidity in COPD patients. Those quantitative models are question-specific and require both statistical (*e.g.* through cross-validation [8]) and biological validation.

**Objective 3 (Clinical application)**: bridging the gap between a predictive model and its use in clinical practice constitutes an important and challenging task: Beyond the basic statistical and biological validation of a model, it also needs to be clinically relevant in the context of personalized medicine. In this objective, predictive models are reviewed for their possible uses in a CDSS. Once a model is considered useful in principle, both a thorough clinical validation and an optimal CDSS implementation are required.

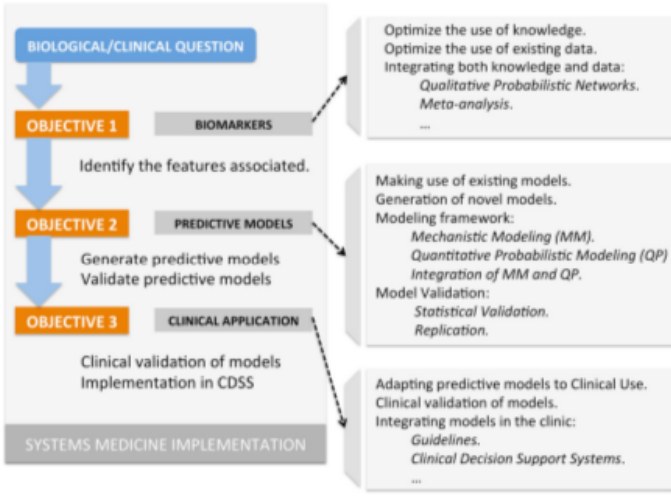



**Figure 1** Synergy-COPD's Systems Medicine framework

## Understanding COPD skeletal muscle dysfunction (*MusclDys*) through systems medicine

### Objective 1: Biomarker identification

To identify the relevant features associated with muscle dysfunction/wasting we used existing data and knowledge through (existing and novel) network-based methodologies. We considered the Biobridge clinical study [2] as the core of the data and extended it through publicly available data-sets from GEO [6]. Among the most relevant data-sets are the gene expression profiling of sputum in COPD ex-smokers (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22148) and Peripheral Blood Mononuclear Cell (PBMC) profiling of COPD patients by COPDgene [9]. For those data-sets used but not publicly available we had permission to access and analyze the data. Next we describe the interactome-based methodologies and results.

#### *The Interactome*

The etiology of COPD involves a multitude of intertwined molecular processes, many of which still remain unknown. These processes are embedded in the larger context of the *Interactome*, referring to a single comprehensive network integrating all molecular interactions, such as protein-protein interactions, regulatory protein-DNA interactions or metabolic interactions (see Figure 2). While ongoing efforts to systematically map the complete Interactome stand only at the beginning, currently available databases (Table 1) already include several hundreds of thousands of interactions. In order to explore such large networks, Systems Medicine has extensively adopted tools from network science [10-12]. Network approaches to human disease are based on the observation that the cellular components associated with a specific disease are not scattered randomly within the Interactome, but segregate in certain neighborhoods or *disease modules*. The identification of the specific disease modules is therefore an important step towards a holistic understanding of how molecular variations with small isolated effect sizes collectively give rise to a certain disease phenotype. The local agglomeration of disease-associated proteins within the Interactome can be used in this process, by extrapolating from the connectivity patterns of known disease-associated proteins to infer novel disease proteins [13,14]. Other applications of this principle include the identification of pathway members [15] or prioritization of weak GWAS loci [16]. In [17] a COPD specific protein interaction network was constructed around genes differentially expressed between healthy and COPD subjects. This network was then queried for potential drug targets that could reverse the expression changes. COPD specific gene expression is also the basis of a Systems Medicine method proposed in [18] that aims at identifying subgroups of COPD patients with different molecular signatures.

Typically, only direct *physical* (binding) interactions are considered in the Interactome. Another line of Systems Medicine network approaches uses *functional* networks, where links may also represent indirect associations, for example co-expression [19] or genetic interaction [20,21]. These networks are usually assembled from specific experimental data, rather than the more general

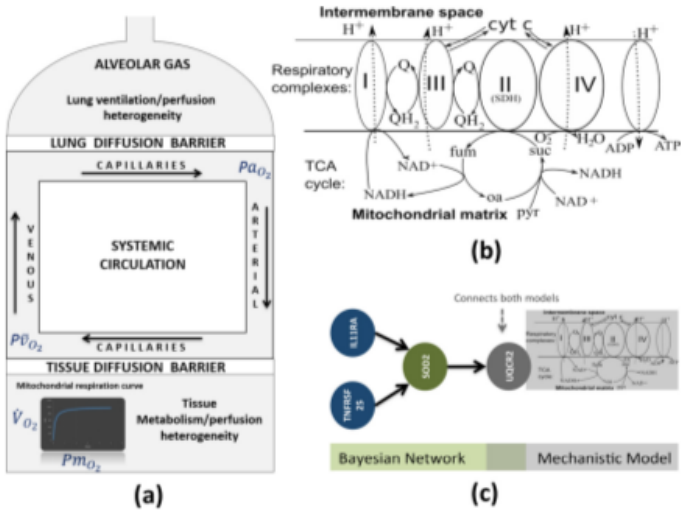



**Figure 2 Mechanistic Models and extensions**. (a) Oxygen transport and utilization model (M1). (b) Mitochondrial respiration and reactive-oxygen species generation model (M2). (3) Personalized model of M2: M2 model is personalized by a Bayesian network that predicts the values of UQCR2 (oxidative phosphorylation chain) by inflammation-associated measurements (IL11RA and TNFRSF25). All models can be simulated in the Simulation Environment [58] and the patient specific values can be obtained through a COPD Knowledge Based [7].

interaction data in public databases. An example for the use of functional networks in COPD is given by the study of [22]. In order to explore the molecular basis of muscle degeneration in COPD, they developed a network model integrating several types of relevant measurements, such as blood cytokine levels and muscle gene expression. They were thereby able to identify several tissue remodeling and bioenergetics pathways that fail to coordinate in COPD diseased muscles.

Both physical and functional-based networks are useful tools that allow the identification of de-regulated functional elements, and to zoom-in into the interactions driving that de-regulation. We considered this information to be relevant in the generation of predictive models addressing the characterization of heterogeneity in COPD.

***Methodologies and results***

To characterize skeletal muscle dysfunction in COPD patients before and after training we evaluated two hypotheses. In the first hypothesis, *MusclDys* is characterized by the de-regulated activity of a selected set of pathways; namely *transcription, proteolysis, immune activation* and/or *oxidative phosphorylation* (see [4] for more details). Using these pathways as a reference an initial list of associated genes was downloaded from the Synergy-COPD Knowledge-base SKB [7] and then filtered by clinical and biological experts. We followed a network approach similar to one described in the BioBridge analysis [22] and extended it by including differential expressed genes, metabolites, cytokines, and clinical variables in addition to the filtered list of genes. We generated a network for each combination of healthy *vs* COPD and untrained *vs* trained individuals, obtaining two main results. First, we identified a module (*i.e.* a network-based cluster of genes, *Mod*1) that is present in all networks and shows the interaction between *mRNA-translation, Insulin* and *mTOR* pathways. Interestingly, this module was also associated to immune

**Table 1 Major resources for protein interaction data:**

| Database / Interaction type | Reference |
|---|---|
| Databases integrating several sources | IntAct [59], MINT [60], BioGRID [61], HPRD [62], MIPS [63], STRING [64] |
| Protein complexes: | CORUM [65], [66] |
| Binary interactions (high-throughput) | CCSB-HI (CCSB), [67;68] |
| Regulatory interactions | TRANSFAC [69] |
| Kinase-substrate Interactions | PhosphositePlus [70] |

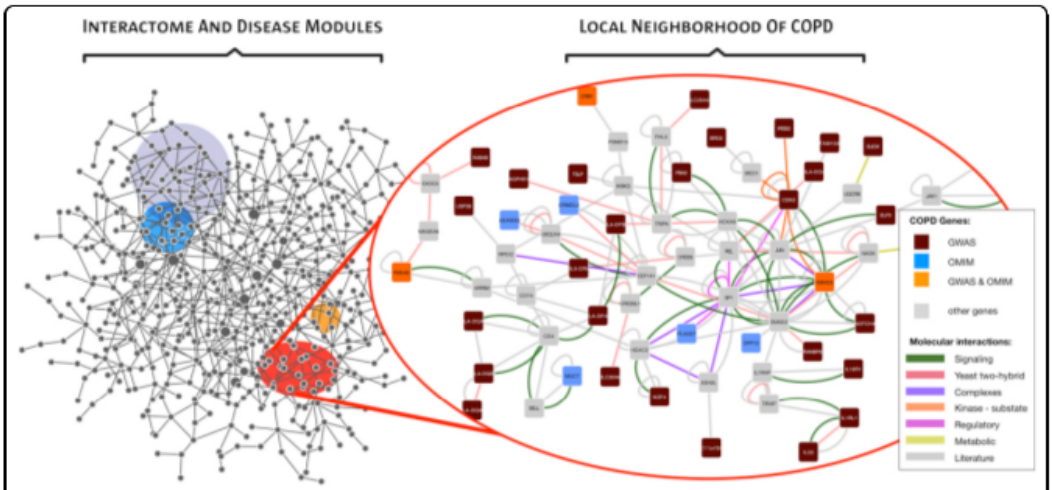



**Figure 3 The "Disease Interactome"**. The Interactome (left) represents the complex network of all molecular components (gene products, proteins, metabolites, RNAs etc) and their interactions. Diseases can be understood as local perturbations. The local neighborhood around genes reported to be associated with COPD (right) is only a very small subset of the full Interactome, yet already shows its enormous complexity.

markers such as *IL1B*. A second result is the general loss of co-regulation (including *Mod* 1 loss) in COPD patients' muscle after training, independently confirming [22]. These promising results require further corroborating data in order to obtain a robust predictive model.

In a second hypothesis, we considered that given the relevance of bioenergetics and immune markers in COPD patients, we could use them to explain the level of *Reactive Oxidative Species* (ROS, major markers of oxidative stress) in COPD patients' muscle. In order to identify the sub-network(s) linking immune and bioenergetics genes to ROS-associated genes (from [23;24]) in the context of COPD we developed a chain-based methodology [25] named ChainRank. Briefly, the methodology identifies relevant sub-networks by identifying and scoring chains of interactions that link specific targets. The type of interactions to include in the chain search are selected by the user; we selected among those interactome-networks included in the Synergy-COPD Knowledge-Base [7]. Scores are generated from the integration of multiple general and context specific measures. Finally, the algorithm allows the identification of genes that are over-represented in highly ranked chains as relevant features. This list was then used for generating personalized predictive models in Objective 3.

### Objective 2: Predictive models

The generation of predictive models in *MusclDys* involves three steps, beginning with the identification of existing mechanistic models that were then updated and adapted to better estimate particular features of interest (FoI). In this process we make use of Objective 1's results to either identify FoI or to personalize the models by adding disease-related parameters.

#### *The identification of mechanistic existing models*

Many phenomena in physiology and biology are of essentially nonlinear nature and therefore require quantitative descriptions in addition to qualitative ones [26]. We considered three well-described physiological models of interest to the characterization of COPD: (i) *Oxygen transport and utilization* [27,28] (*M-OX*); (ii) *Cell Bioenergetics, mitochondrial respiration and reactive-oxygen-species generation (ROS)* [23,24], (*M-ROS*); and, (iii) *Spatial heterogeneities of lung ventilation and perfusion* [29] (*M-HET*). The first two models (*M-OX* and *M-ROS*) are relevant for the characterization of the systemic effects of the disease in skeletal muscle, as they provide mechanistic description of both the oxygen pathway and ROS generation. The third model (*M-HET*) is relevant for the study of pulmonary events in a sub-set of COPD patients [3] with low pulmonary density (high emphysema score) and mild airway remodeling resulting in mild to moderate airflow limitation.

*Oxygen transport and utilization* [27,28]*(M-OS):* The model details the determinants of oxygen transport from air to mitochondria and characterizes oxygen utilization at mitochondrial level during maximum exercise. In summary, it constitutes the most complete integrative approach of the interplay among factors modulating oxygen transport (lungs, hearth, blood and skeletal

muscle) and oxygen utilization at muscle level (see Figure 2 (a)). In COPD patients, the oxygen transfer capacity from the atmosphere to the cell, as well as its utilization at mitochondrial level, can be limited. It is therefore of interest to observe the effects of such a limitation at all levels of the transport chain [30] and, in particular, its effects on muscle's mitochondria.

*Cell Bioenergetics, mitochondrial respiration and reactive-oxygen-species generation (M-ROS):* The model details mitochondrial respiration and its relation with the production of Reactive Oxidative Species (ROS). The model integrates two sub-models, the Electron Chain model and the TCA cycle module [23,24]. ROS production in the mitochondrial respiratory chain is a signal of cellular adaptation to the environment, but a sharp increase is incompatible with cell survival; therefore the predicting ROS production is a relevant task. Moreover, in smoking-related COPD patients the antioxidant capacity is severely reduced and further decreases after smoking cessation due to endogenous production of ROS [31,32]. The integrated model (M-OX + M-ROS) generated within the Synergy-COPD project allows estimating quantitatively the relationships between determinants of cell oxygenation and mitochondrial ROS generation. The interplay between ROS levels and the antioxidant capacity of the redox system ultimately determines tissue oxidative and nitrosative stress with important implications on pathway regulation and cell damage.

*Spatial heterogeneities of lung ventilation and perfusion* [29] (M-HET): The anatomy-based multi-scale model of the human pulmonary circulation allows for the study of pre- and post-occlusion flow and embolus-generated blow flow redistribution, among other features. It combines four independent simulations of model geometry, tissue mechanics, ventilation and blood flow allowing for a local description of alveolar ventilation and pulmonary blood flow. The lung modeling approaches are described in this Supplement in detail in a separate paper as interactive work with the FP7 EU Project *AirPROM*. A major relevance of lung modeling in Synergy-COPD is the characterization of patients with reduced lung density but without classical COPD symptoms of airway obstruction [3]. The inclusion of this modeling approach in Synergy-COPD had two main goals. Firstly, the analysis of the impact of spatial heterogeneities of lung ventilation and perfusion on blood oxygenation and, secondly, the study of the subset of COPD patients showing dissociation between high emphysema score and low intensity of airway remodeling, as indicated above and described in [3].

***Updating existing models***

In order to characterize skeletal muscle dysfunction in COPD (*MusclDys*) we modified the oxygen transport and utilization model (M-OX) and included the mitochondrial respiration [33]. The outcomes of this model were two fold: (1) it increased the physiological validity of the model by estimating the mitochondrial $P_{O2}$, (2) it allowed for the integration with bioenergetics models (M-ROS). A second extension of the model was the modeling of lung ventilation/perfusion heterogeneities and [33]; these extensions (see Figure 2(a)) allows better estimation and better personalization of the model in COPD.

In addition, we integrated models M-OX and M-ROS (IM) to model the relation between oxygen transport and ROS generation, which constitutes a major marker of skeletal muscle dysfunction. Parameter models were investigated again in the case of [23,24], to provide better ROS estimations. A major outcome of the integrated model analysis [34] is that it permits to estimate the effect of various states of oxygen supply and demand of mitochondrial $P_{O2}$ on ROS production; this is relevant in COPD patients with airway obstruction symptoms [22].

***Novel models***

In Synergy-COPD we generated novel models (Objective 2) by integrating existing mechanistic models, developed for the healthy individual, with features of interest (FoI) associated to *MusclDys* (obtained in Objective 1). Having identified ROS as a major marker of muscle dysfunction, we aimed to predict ROS status by surrogate variables identified in Objective 1. Initially those surrogate variables were immune markers from protein measurements in the blood as shown in [22]. To integrate the effect of the surrogate variables (SV) with the integrated model of oxygen transport and ROS generation we proposed to use SV values (which are more commonly used in clinical diagnostics) to estimate a subset of the integrated model parameter values. As a technical solution we proposed the used of Bayesian networks (see Figure 2(c), [35]) to connect immune markers and selected model parameters. This connection was also possible through the use of methodologies and results obtained during Objective 1 such as the ChainRank methodology briefly described elsewhere in the manuscript. However, the generation of accurate linking-by-Bayesian networks is limited by the requirement of a large number of samples. Therefore, to increase accuracy we made use of public available muscle-related data-sets in GEO [6] and included estimates of relations from other sources such as text-mining [36]) into the Bayesian network generation. While we considered that the proposed approach to be technically valid, still we need to increase the sample size to generate useful models, therefore the requirement of follow-up studies. The model can be run in the Synergy-COPD Simulation Environment and the patient specific values can be obtained through a COPD Knowledge Base [2,7].

As a second approach, we integrated transcriptomic data from muscle biopsies and literature-based data into a mathematical discrete model [37]. By this model-driven approach we aimed to determine the processes that lead the abnormal adaptation to training in COPD patients and the role of ROS in this process. Since skeletal muscle mitochondrial dysfunction is a central actor in COPD [38] this approach was based on those genes associated to selected mitochondrial processes from Objective 1's candidate biomarkers obtained in [22]. The modeling was achieved by inferring the activity state of a gene regulatory network (GRN) in six different states: Control group, COPD with normal body mass index (BMI) and COPD with low BMI before and after undergoing 8 weeks of training program [22]. We carried out this task in two parts: 1) GRN reconstructions and 2) Integration of GRN into a discrete model.

As a first step in the GRN reconstruction we curated the list of candidate biomarkers to be included by the re-analysis of the transcriptomic data of the six different states used previously in [22]. For this aim, we used statistical methods such as rank product [39] to determine the gene candidates and Gene Ontology and Human Proteins Atlas databases [40,41] to filter those genes associated with mitochondrial processes in skeletal muscle. Next, to determine gene associations we used IPA software and DroID [42,43]. Finally, In order to correct incomplete or erroneous annotations and identify the direction and the sign of the interactions, we manually curated the GRN reconstruction using a large number of bibliographic data sources.

The GRN reconstruction was then converted into a mathematical discrete model based on the *Thomas formalism* [44] by mechanistically describing the interactions between those mitochondrial-associated genes that were differentially expressed between states. In order to refine the accuracy of our model predictions, we used public available muscle-related data-sets in GEO [6] to impose constraints to our model. We integrated these constraints in the form of inequalities based on probabilistic approaches: if we observed a strong Pearson correlation ($rho > 0.9$) between two non-connected genes, their expression values were forced to evolve in the same direction. The rationale is just the opposite in the case of a strong anti-correlation. Then, summing up, we propose a method by which the interaction between genes are determined by performing a tissue and organelle specific GRN reconstruction and the constraints are defined using probabilistic approaches, finally both, the GRN reconstruction and the constraints are integrated into a discrete model in order to unveil the mechanisms governing the adaptation to training in the groups of study.

Together, both probabilistic approaches show a way forward to close the inherent under-determination gap of deterministic, quantitative models by coupling data driven and knowledge driven approaches.

#### Objective 3: Clinical application and limitations in Synergy-COPD

The models proposed to address *MusclDys* are still far from the clinical practice. We consider that the methodologies to achieve such goal do exist, but the data publicly available is limited. Several lessons can be learnt:

(1) While the use of mechanistic models is very valid to understand biological systems and diseases, they have serious limitations in the study of complex diseases. Complex diseases may be described as the combination of many factors, and mechanistic models will require too many parameters and consequently too much data to be *yet* clinically effective.
(2) However statistical predictive models (e.g. linear models, Bayesian networks, etc.), not necessarily mechanistically accurate, may provide a valid technical solution. The implementation of such technical solution requires the use of large amount of data to ensure accuracy and statistical validation. For this data to be obtained we consider necessary (1) to strengthen the policies promoting data-sharing (especially in the clinical context) and (2) the generation of large data-sets with proper experimental designs and clinically-driven hypotheses.
(3) Clinically driven research needs to be re-designed to align the different objectives described in Figure 1. We observed that the re-use of data is necessary but complex, and minimal modifications (such as extended questionnaires to patients providing samples) in existing bio-banks may be very useful.

## Understanding COPD co-morbidities through systems medicine

COPD has been associated with several diseases such as lung cancer [45], metabolic syndrome and cardiovascular diseases [46]. However, not all COPD patients share the same diseases or exhibit the same degree of co-morbidity. We hypothesized that the particular co-morbidities in a given COPD patient can be understood from his/her particular set of de-regulated pathways and genes. Identifying genes and pathways that are shared between COPD associated diseases could therefore allow for a more detailed characterization of COPD and its co-morbidities. In the first subsection we briefly describe our method to identify such potential biomarkers (pathways and/or genes). We then introduce our initial predictive models in the second subsection and finally discuss the clinical applications. Some of the data sets supporting this article are available in HuDiNe repository (http://barabasilab.neu.edu/projects/

hudine/resource/data/data.html); for the rest we had permission to access and analyze the data.

#### Biomarker and Co-morbidity identification

Using 13 million health records from U.S. Medicare [47], we identified 27 disease groups (DG) with significantly elevated risks to co-occur with COPD. These groups included both well-established associations like cardiovascular diseases or lung cancer, but also unexpected ones that could be interesting candidates for more focused follow-up investigations. In order to elucidate possible shared molecular origins between the disease groups and COPD, we considered their respective implicated pathways: for each disease group, we first constructed a comprehensive list of known associated genes from the literature (by pooling several sources of gene-disease associations such as OMIM, NIH Thesaurus and text-mining among others). We then performed a pathway enrichment analysis for each disease group. The results show that there are a number of pathways that are shared between different disease groups, suggesting that the observed co-morbidities are indeed rooted in shared molecular mechanisms. By further inspecting the genes within prevalent pathways we were able to identify a number of genes with the potential to characterize COPD co-morbidity. We are investigating if those markers may predict the level of co-morbidity. This could be of immediate relevance for the clinical practice, as co-morbidity has been associated to lower overall quality of life [48] and increased mortality [47,49]. To date, a number of interesting outcomes of this analysis remain to be validated in further studies. However, with currently available data we considered that the disease groups and co-factors such as age and gender could be used to generate predictive models.

#### Predictive models

We selected disease groups that are highly prevalent in COPD patients, such as heart and circulation associated diseases and digestive alterations. The observation that their prevalences vary with age prompted us to develop a first model (Objective 2) aiming to predict the probability for specific co-morbidities in COPD patients over different age strata; in this case we made use of the ranking of co-morbidities from Objective 1 to select those diseases of major interest in the generation of the predictive models. This model may be used as support information for clinicians in the daily practice (e.g. predictive medicine, or comparing observed symptoms with candidate co-morbidities). For a more robust clinical validation, however, follow-up studies in different cohorts will be required; for this reason we consider the comorbidity modeling as part of Objective 2, but closer to the Objective 3 that any other model presented.

#### Clinical application and limitations in Synergy-COPD

While co-morbidity is being generally accepted as a relevant clinical factor [47,50-55], co-morbid predictive models are rarely reaching the clinical practice. Our experiences gained throughout the Synergy-COPD project suggest several limitations:

(1) Incompatibilities in the medical nomenclature. While there are several large health registries available to investigate co-morbidities (e.g. Medicare, Swedish Registry and others), there is yet to agree a common diagnostic standard even for simplified administrative coding. In Sweden, for instance, nowadays ICD10 is being used, yet the registry also includes information coded in ICD7 to ICD9. In comparison, Medicare (as used in [47]) is mainly using ICD9 codes. Maps between ICD coding do exist, but they are not accurate, and every new ICD coding system may represent a different conceptual approach. ICD11 will represent a new challenge.

(2) There are many studies investigating specific co-morbidities, not only in COPD but in many other diseases. Yet, two major limitations inhibit the integration of these studies in larger meta-analyses: (1) diseases may be defined differently in each study and in many cases no official coding is followed; (2) the selection of diseases is biased towards well established and expected diseases. These two limitations reflect that most studies are developed to validate specific hypotheses. We believe that broader studies and normalized questionnaires will eventually facilitate meta-analyses and thereby increase the power of co-morbidity studies.

(3) Finally, previous large-scale studies are often limited to one type of data: either -omics data were collected, but no co-morbidity information, or the other way round. We believe that in the future it will be crucial to combine these two approaches.

## Conclusions

Despite the massive amounts of data collected in medical research throughout the last decades, our understanding of complex diseases still remains very limited. The fundamental shortcoming of our knowledge may be illustrated by a popular quote attributed to Ernest Rutherford: *all science is either physics or stamp collecting* [56]. Systems Medicine bears the promise of facilitating the transition from stamps to understanding. Indeed we are convinced that a systems perspective is necessary for the integration of all hitherto largely disconnected facts, thereby ultimately enabling clinically predictive tools. Synergy-COPD presents a large-scale case study of Systems Medicine applied to COPD, recognizing that both *Clinical Research*

and *Clinical Decision Systems* require the development of integrative quantitative models. Developing such models is a complex task which we addressed by adhering to a 3-step framework: (1) feature identification, (2) model generation and statistical validation, (3) clinical validation and implementation. We developed and used the framework targeting specifically the characterization of muscle-related systemic effects and co-morbidity as use-cases thus grounding the methodology in real-world applications. In both use-cases we were able to identify candidate biomarkers that may help characterizing COPD heterogeneity, and developed models with the potential to be considered in future Clinical Decision Support Systems (e.g. co-morbidity prevention and prognosis among other objectives). Throughout the project we identified several key factors that are currently limiting the clinical applicability of our approach: the most important ones were data availability, normalization of frameworks (*e.g.* ICD codes in co-morbidity) and the necessity of broader and optimized experimental designs (*e.g.* the inclusion of co-morbidity information in genomic studies).

In conclusion, we consider that the first steps to bridge the gap between basic research and clinical practice are built, however further steps are required to complete the path. To exploit the full potential of our results, future follow-ups are required for statistical and clinical validation, and once validated, predictive models (supported by longitudinal studies) will make a strong case for clinical applications. Further considerations on challenges and future are discussed in [57] on this Supplement.

We are at the juncture of a very exciting era, where Systems Medicine offers the possibility of a real connection between research and clinical applications. While Synergy-COPD may only represent a minor milestone along a long road, we are convinced it is a relevant and instructive case.

#### Competing interests
DM is part of Biomax Informatics AG. The rest of authors declare they have no competing interests.

#### Authors' contributions
DGC and JM defined an initial draft of the manuscript. DGC, JM, JR, JT, and LB reviewed and defined the final structure. DGC and JM wrote the manuscript. All authors first reviewed their specific sections in detail, then reviewed the full document, in both cases they proposed modifications; finally all authors agreed on the final version.

#### Acknowledgements
We would like to thank all the members of the Synergy-COPD consortium for their support.

#### Declaration
Publication of this article has been funded by the Synergy-COPD European project (FP7-ICT-270086). The opinions expressed in this manuscript are those of the authors and are not necessarily those of Synergy-COPD project's partners or the European Commission
This article has been published as part of *Journal of Translational Medicine* Volume 12 Supplement 2, 2014: Systems medicine in chronic diseases: COPD as a use case. The full contents of the supplement are available online at http://www.translational-medicine.com/supplements/12/S2.

#### Authors' details
[1]Unit of computational Medicine, Center for Molecular Medicine, Department of Medicine, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden. [2]Center for Complex Network Research, Northeastern University Physics Department, Boston, MA 02115, USA. [3]Department of Theoretical Physics, Budapest University of Technology and Economics, H-1111 Budafoki út. 8., Budapest, Hungary. [4]Hospital Clinic, IDIBAPS, CIBERES, Universitat de Barcelona, Barcelona, Catalunya, Spain. [5]Barcelona Digital Technology Centre Carrer Roc Boronat, 117 08018 Barcelona. [6]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. [7]Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool, UK. [8]Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK. [9]Biomax Informatics AG, Munich, Germany. [10]School of Medicine, University of California, San Diego, San Diego, CA 92093-0623A, USA. [11]Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Smith Bldg., Rm. 858A, 450 Brookline Ave, Boston, MA 02215, USA. [12]Center for Network Science, Central European University, Nadoru. 9, 1051 Budapest, Hungary. [13]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115 USA.

**Published: 28 November 2014**

#### References
1. Majumder D, Mukherjee A: **A passage through systems biology to systems medicine: adoption of middle-out rational approaches towards the understanding of therapeutic outcomes in cancer.** *The Analyst* 2011, **136(4)**:663-78.
2. Gomez-Cabrero , David , Lluch-Ariet , Magi , Tegner Jesper, Cascante Marta, Miralles Felip, Roca Josep, the Synergy-COPD consortium: **Synergy-COPD: A systems Approach for understanding and managing Chronic Diseases.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S2.
3. Garcia-Aymerich J, Gómez FP, Benet M, Farrero E, Basagaña X, Gayete À, Antó JM: **Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes.** *Thorax* 2011.
4. Josep Roca, Claudia Vargas, Isaac Cano, Vitaly Selivanov, Esther Barreiro, Dieter Maier, Francesco Falciani, Peter Wagner, Marta Cascante, Judith Garcia-Aymerich, Susana Kalko, Igor Marin, Jesper Tegner, Joan Escarrabill, Alvar Agustí, David Gomez-Cabrero, the Synergy-COPD consortium: **Chronic Obstructive Pulmonary Disease Heterogeneity: Challenges for Health Risk Assessment, Stratification and Management.** *BMC Journal of Translational Medicine, to appear* 2014, **12(Suppl 2)**:S3.
5. Filip Velickovski, Luigi Ceccaroni, Josep Roca, Felip Burgos, Galdiz Juan B, Marina Nueria, Magi Lluch-Ariet: **Clinical Decision Support Systems (CDSS) for preventive management of COPD patients.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S9.
6. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A: **NCBI GEO: archive for functional genomics data sets–update.** *Nucleic Acids Res* 2013, **41(Database)**:D991-5, Jan.
7. Isaac Cano, Ákos Tényi, Christine Schueller, Martin Wolff, Mercedes Huertas Migueláñez M, David Gomez-Cabrero, Philipp Antczak, Josep Roca, Marta Cascante, Francesco Falciani, Dieter Maier: **The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S6.
8. Picard , Richard , Cook , Dennis : **"Cross-Validation of Regression Models".** *Journal of the American Statistical Association* 1984, **79(387)**:575-583.
9. Bahr TM, Hughes GJ, Armstrong M, Reisdorph R, *et al*: **Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease.** *Am J Respir Cell Mol Biol* 2013, **49(2)**:316-23, Aug.
10. Vidal M, Cusick ME, Barabási AL: **Interactome networks and human disease.** *Cell* 2011, **144(6)**:986-98.
11. Schadt EE: **Molecular networks as sensors and drivers of common human diseases.** *Nature* 2009, **461**:218-223.
12. Barabási AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease.** In *Nature reviews. Volume 12.* Genetics; 2011(1):56-68.

13. Köhler S, et al: **Walking the interactome for pri- oritization of candidate disease genes.** *The American Journal of Human Genetics* 2008, **82**:949-958.
14. Oti M, Snel B, Huynen MA, Brunner HG: **Predicting disease genes using protein-protein interactions.** *Journal of medical genetics* 2006, **43**:691-698.
15. Navlakha S, Gitter A, Bar-Joseph Z: **A network-based approach for predicting missing pathway interactions.** *PLoS computational biology* 2012, **8**:e1002640.
16. Sharma A, Gulbahce N, Pezner SJ, Menche J, Ladenvall C, Folkersen L, Eriksson P, Orho-Melander M, Barabási AL: **Network-based Analysis of Genome Wide Association Data Provides Novel Candidate Genes for Lipid and Lipoprotein Traits.** *Molecular & Cellular Proteomics* **12.11(2013)**:3398-3408.
17. Bao H, Wang J, Zhou D, Han Z, Su L, Zhang Y, Li Q: **Protein-Protein Interaction Network Analysis in Chronic Obstructive Pulmonary Disease.** *Lung (2800)* 2013.
18. Menche J, Sharma A, Cho MH, Mayer RJ, Rennard SI, Celli B, Barabási AL: **A diVIsive Shuffling Approach (VIStA) for gene expression analysis to identify subtypes in Chronic Obstructive Pulmonary Disease.** *BMC Systems Biology* 2014, **8(Suppl 2)**:S8.
19. Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules.** *Science* 2003, **302**:249-255.
20. Davierwala , Armaity P, *et al*: **The synthetic genetic interaction spectrum of essential genes.** *Nature genetics* 2005, **37(10)**:1147-1152.
21. Franke L, van Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, Wijmenga C: **Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes.** *The American Journal of Human Genetics* 2006, **78**:1011-1025.
22. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Falciani F: **A Systems Biology Approach Identifies Molecular Networks Defining Skeletal Muscle Abnormalities in Chronic Obstructive Pulmonary Disease.** In *PLoS Computational Biology* S. Miyano 2011, **7(9)**:e1002129.
23. Selivanov Va, Votyakova TV, Pivtoraiko VN, Zeak J, Sukhomlin T, Trucco M, Cascante M: **Reactive oxygen species production by forward and reverse electron fluxes in the mitochondrial respiratory chain.** *PLoS Computational Biology* 2011, **7(3)**:e1001115.
24. Selivanov Va, Votyakova TV, Zeak Ja, Trucco M, Roca J, Cascante M: **Bistability of mitochondrial respiration underlies paradoxical reactive oxygen species generation induced by anoxia.** *PLoS Computational Biology* 2009, **5(12)**:e1000619.
25. Ákos Tényi, Pedro de Atauri, David Gomez-Cabrero, Isaac Cano Isaac, Francesco Falciani Francesco, Marta Cascante Marta, Josep Roca, Dieter Maier: **ChainRank, a chain search based method for prioritization and contextualisation of biological sub-networks.** 2014, Submitted.
26. Wolkenhauer O, Auffray C, Jaster R, Steinhoff G, Dammann O: **The road from systems biology to systems medicine.** *Pediatric research* 2013, **73(4 Pt 2)**:502-7.
27. Wagner PD: **A theoretical analysis of factors determining VO2max at sea level and altitude.** *Resp Physiol* 1996a, **106(3)**:329-43.
28. Wagner PD: **Determinants of maximal oxygen transport and utilization.** *Annual Review of Physiology* 1996b, **58**:21-50.
29. Burrowes KS, Clark aR, Tawhai MH: **Blood flow redistribution and ventilation-perfusion mismatch during embolic pulmonary arterial occlusion.** *Pulmonary Circulation* 2011, **1(3)**:365-76.
30. Heffner JE: **The story of oxygen.** 2013, **58(1)**:18-31.
31. Kirkham Pa, Barnes PJ: **Oxidative stress in COPD.** *Chest* 2013, **144(1)**:266-73.
32. Barnes PJ: **Cellular and Molecular Mechanisms of Chronic Obstructive Pulmonary Disease.** *Clinics in Chest Medicine* 2014, **35(1)**:71-86.
33. Cano I, Mickael M, Gomez-Cabrero D, Tegnér J, Roca J, Wagner PD: **Importance of Mitochondrial in Maximal O2 Transport and Utilization: a Theoretical Analysis.** *Respiratory Physiology & Neurobiology* 2013.
34. Selivanov Cano V, Gomez-Cabrero David, Tegnér Jesper, Roca J, Wagner PD, Cascante M: **Mathematical Modeling of Skeletal Muscle Reactive Oxygen Species Generation during Maximal Exercise as a function of Mitochondrial.** 2014, submitted.
35. Scutari M: **Learning Bayesian Networks with the bnlearn R Package.** 2010, **35(3)**.
36. Haibe-Kains B, Olsen C, Djebbari A, Bontempi G, Correll M, Bouton C, Quackenbush J: **Predictive networks: a flexible, open source, web application for integration and analysis of human gene networks.** *Nucleic Acids Research* 2012, **40(Database)**:D866-75.
37. de Mas Igor Marin, Fanchon E, Selivanov Vitaly A, Papp Balázs, Josep Roca, Marta Cascante: **A novel discrete model-driven approach unveils abnormal metabolic adaptation to training in COPD.**, Submitted.
38. Meyer A, Zoll J, Charles AL, Charloux A, de Blay F, Diemunsch P, Sibilia J, Piquard F, Geny B: **Skeletal muscle mitochondrial dysfunction during chronic obstructive pulmonary disease: central actor and therapeutic target.** *Exp Physiol* 2013, **98(6)**:1063-78.
39. Breitling R, Armengaud P, Amtmann A, Herzyk P: **Rank products: A simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments.** *(FEBS) Letters* 2004, **573(1-3)**:83-92.
40. Thomas PD, Mi H, Lewis S: **Ontology annotation: mapping genomic regions to biological function.** *Curr Opin Chem Biol* 2007, **11(1)**:4-11.
41. Uhlén M, Björling E, Agaton C, Szigyarto CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, Berglund L, Bergström K, Brumer H, Cerjan D, Ekström M, Elobeid A, Eriksson C, Fagerberg L, Falk R, Fall J, Forsberg M, Björklund MG, Gumbel K, Halimi A, Hallin I, Hamsten C, Hansson M, Hedhammar M, Hercules G, Kampf C, Larsson K, Lindskog M, Lodewyckx W, Lund J, Lundeberg J, Magnusson K, Malm E, Nilsson P, Odling J, Oksvold P, Olsson I, Oster E, Ottosson J, Paavilainen L, Persson A, Rimini R, Rockberg J, Runeson M, Sivertsson A, Sköllermo A, Steen J, Stenvall M, Sterky F, Strömberg S, Sundberg M, Tegel H, Tourle S, Wahlund E, Waldén A, Wan J, Wernérus H, Westberg J, Wester K, Wrethagen U, Xu LL, Hober S, Pontén F: **A human protein atlas for normal and cancer tissues based on antibody proteomics.** *Mol Cell Proteomics* 2005, **4(12)**:1920-32.
42. Murali T, Pacifico S, Yu J, Guest S, Roberts GG 3rd, Finley RL Jr: **DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila.** *Nucleic Acids Res* 2011, **39(Database)**:D736-43.
43. Krämer A, Green J, Pollard J Jr, Tugendreich S: **Causal analysis approaches in Ingenuity Pathway Analysis.** *Bioinformatics 15* 2014, **30(4)**:523-30.
44. Thomas R, Multistationarity Kaufman M: **the Basis of Cell Differentiation and Memory.** *II Logical Analysis of Regulatory Networks in Term of Feedback Circuits Chaos 2001* **11**:180-195.
45. Houghton aM: **Mechanistic links between COPD and lung cancer.** *Nature Reviews Cancer* 2013, **13(4)**:233-45.
46. Müllerova H, Agusti A, Erqou S, Mapel DW: **Cardiovascular comorbidity in COPD: systematic literature review.** *Chest* 2013, **144(4)**:1163-78.
47. Hidalgo Ca, Blumm N, Barabási AL, Christakis Na: **A dynamic network approach for the study of human phenotypes.** *PLoS Computational Biology* 2009, **5(4)**:e1000353.
48. Fabbri LM, Beghé B, Agustí A: **COPD and the solar system: introducing the chronic obstructive pulmonary disease comorbidome.** *American Journal of Respiratory and Critical Care Medicine* 2012, **186(2)**:117-9.
49. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, Celli B: **Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease.** *American Journal of Respiratory and Critical Care Medicine* 2012, **186(2)**:155-61.
50. Areias V, Carreira S, Anciães M, Pinto P, Bárbara C: **Co-morbidities in patients with gold stage 4 chronic obstructive pulmonary disease.** *Revista Portuguesa de Pneumologia* 2013.
51. Decramer M, Janssens W, Miravitlles M: **Chronic obstructive pulmonary disease.** *Lancet* 2012, **379(9823)**:1341-51.
52. Hernandez C, Jansa M, Vidal M, Nuñez M, Bertran MJ, Garcia-Aymerich J, Roca J: **The burden of chronic disorders on hospital admissions prompts the need for new modalities of care: a cross-sectional analysis in a tertiary hospital.** *QJM: Monthly Journal of the Association of Physicians* 2009, **102(3)**:193-202.
53. Inghammar M, Ekbom A, Engström G, Ljungberg B, Romanus V, Löfdahl C-G, Egesten A: **COPD and the risk of tuberculosis--a population-based cohort study.** *PloS One* 2010, **5(4)**:e10138.
54. Khdour MR, Hawwa AF, Kidney JC, Smyth BM, McElnay JC: **Potential risk factors for medication non-adherence in patients with chronic obstructive pulmonary disease (COPD).** *European Journal of Clinical Pharmacology* 2012, **68(10)**:1365-73.
55. Maclay JD, MacNee W: **Cardiovascular disease in COPD: mechanisms.** *Chest* 2013, **143(3)**:798-807.
56. Birks JB: **Rutherford at Manchester.** Heywood. London; 1962.
57. Felip Miralles, David Gomez-Cabrero, Magi Lluch-Ariet, Jesper Tegnér, Marta Cascante, Josep Roca, the Synergy-COPD consortium: **Predictive**

**Medicine: Outcomes, Challenges and Opportunities in the Synergy-COPD project.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S12.

58. Mercedes Huertas Migueláñez M, Daniel Mora, Isaac Cano, Dieter Maier, David Gomez-Cabrero, Magi Lluch-Ariet, Felip Miralles: **Simulation Environment and Graphical Visualization Environment: a COPD use-case.** *Journal of Translational Medicine* 2014, **12(Suppl 2)**:S7.
59. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K, Hermjakob H: **The intact molecular interaction database in 2010.** *Nucleic acids research* 2010, **38**:D525-D531, 2010.
60. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: **Mint, (2010) The molecular interaction database: 2009 update.** *Nucleic acids research* 2010, **38**:D532-D539.
61. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M: **The biogrid interaction database: 2011 update.** *Nucleic acids research* 2011, **39**:D698-D704, 2011.
62. Prasad TK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A: **Human protein reference database2009 update.** *Nucleic acids research* 2009, **37**:D767-D772.
63. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes HW, Ruepp A, Frishman D: **The MIPS mammalian protein-protein interaction database.** *Bioinformatics* 2005, **21(6)**:832-834.
64. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ: **STRING v9. 1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic acids research* 41 **D1(2013)**:D808-D815.
65. Ruepp A, Waegele B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW: **Corum: the comprehensive resource of mammalian protein complexes 2009.** *Nucleic acids research* 2010, **38**: D497-D501.
66. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, Wan C, Vlasblom J, Dar VU, Bezginov A, Clark GW, Wu GC, Wodak SJ, Tillier ER, Paccanaro A, Marcotte EM, Emili A: **A census of human soluble protein complexes.** *Cell* 2012, **150(5)**:1068-1081.
67. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksöz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE: **A human protein-protein interaction network: a resource for annotating the proteome.** *Cell* 2005, **122**:957-968.
68. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, , Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, *et al*: **Towards a proteome-scale map of the human protein-protein interaction network.** *Nature* 2005, **437**:1173-1178.
69. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, , Thiele S, Wingender E: **Transfac: transcriptional regulation, from patterns to profiles.** *Nucleic acids research* 2003, **31**:374-378.
70. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M: **Phosphositeplus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.** *Nucleic acids research* 2012, **40**:D261-D270.

doi:10.1186/1479-5876-12-S2-S4
**Cite this article as:** Gomez-Cabrero *et al.*: **Systems Medicine: from molecular features and models to the clinic in COPD.** *Journal of Translational Medicine* 2014 **12**(Suppl 2):S4.

# Appendix V

Cascante et al. Journal of Translational Medicine 2014, 12(Suppl 2):S11
http://www.translational-medicine.com/content/12/S2/S11

JOURNAL OF TRANSLATIONAL MEDICINE

REVIEW Open Access

# Workforce preparation: the Biohealth computing model for Master and PhD students

Marta Cascante[1,2,3*], Pedro de Atauri[1,2], David Gomez-Cabrero[4], Peter Wagner[5], Josep Joan Centelles[1], Silvia Marin[1], Isaac Cano[2], Filip Velickovski[6], Igor Marin de Mas[1], Dieter Maier[7], Josep Roca[2,3,8], Philippe Sabatier[3,9]

**Abstract**

The article addresses the strategic role of workforce preparation in the process of adoption of Systems Medicine as a driver of biomedical research in the new health paradigm. It reports on relevant initiatives, like CASyM, fostering Systems Medicine at EU level. The chapter focuses on the BioHealth Computing Program as a reference for multidisciplinary training of future systems-oriented researchers describing the productive interactions with the Synergy-COPD project.

## The new paradigm shift in healthcare

Discoveries in Health Sciences classically begin at "the bench", with basic research, and then progress to the clinical level, or patient's "bedside". Growing barriers and increasing complexities have made difficult to translate new knowledge to the bedside. But, scientists are increasingly aware that this "bench-to-bedside" approach shall really be a two-way system providing a fluent and efficient interplay between healthcare and basic biomedical research.

The purpose of the manuscript is to pull together the principal findings and recommendations of the various reports and publications concerning the paradigm shift in health taking into account an analysis of the changing nature of the interactions among health practice, research, and education. More specifically, we consider the implications for biomedical innovation from three perspectives: (i) Social context (Section: The social context); (ii) Scientific discipline (Section: Systems Medicine and workforce preparation); and (iii) Educational and training perspective (Section: Systems Medicine post-graduate education: the need for multidisciplinary collaboration). The latter addresses the question of what our global world should seek as objectives of biomedical education and innovation in the 21st Century, recognizing that these must change significantly to address rapidly changing needs and priorities.

* Correspondence: martacascante@ub.edu
[1]Departament de Bioquimica i Biologia Molecular i IBUB, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain
Full list of author information is available at the end of the article

Achievement of major breakthroughs in biomedicine depends on rapid adaptation of training schemes and on long-term investment in cutting-edge research. To this end, we analyze the Biohealth Computing program for Master and PhD students, as an example of educational and training program aiming to produce translational researchers aligned with the new health paradigm. Finally we suggest a roadmap to the future: a series of recommendations and actions aimed at transforming practice, research, and education, with the fundamental objective of sustaining and enhancing our capacity for biomedical innovation which constitutes a key element for economic prosperity and social well-being.

Systems thinking approach is not new to biomedicine, having been historically applied in the physiological sciences. However, it had to be abandoned because of the lack of necessary data and tools. Today, the revolutions in molecular biology and in Information and Communication Technologies (ICT) constitute key enablers to re-assess old problems under the new health paradigm.

## The social context

During few exhilarating decades, in the middle of the twentieth century, it seemed the world might have a reprieve from some major diseases. But, during last two decades of the last century, the concern about the epidemics of Non-Communicable Diseases (NCDs) increasingly grew [1-4]. The need for novel healthcare approaches for chronic patients triggered the Chronic Care model first formulated in early 2000 [5]. It was soon adopted by the World Health Organization (WHO) through the Innovative Care for

Chronic Conditions (ICCC) initiative [5-9]. In September 2011, the United Nations (UN) General Assembly formally acknowledged the impact of the NCDs epidemics and the need for reshaping health systems worldwide [1,4] toward adoption of the Chronic Care model. Further studies on health trends [10] confirm NCDs as: (1) leading cause of healthcare burden and mortality in the world; (2) increasing in prevalence; (3) extremely expensive with conventional healthcare approaches; and, (4) an under-appreciated cause of poverty and hindered economic development [1]. NCDs are multi-factorial in nature and involve complex gene-environment interactions [4]. Environmental factors, unhealthy life-styles and increased life expectancy (ageing), intrinsic host responses, such as local and systemic inflammation [11], epigenetic changes [12] and decoupling of basic regulatory mechanisms with impact on bioenergetics, immune responses and remodeling [12] may play significant roles in the initiation and persistence of NCDs and associated co-morbidities. Accordingly, the management of NCDs should move towards holistic approaches aiming at tackling all components of their complexities through deployment of integrated care services, as proposed by the Chronic Care model [5-9].

The on-going transition toward an integrated care approach in several EU regions *(European Innovation Partnership for Active and Healthy Ageing, EI-AHA)* [13] is fostered by three main driving forces: (i) the burden imposed by the epidemics of NCDs [10]; (ii) the need for generating efficiencies allowing further investments on innovation without increasing overall health costs; and, last but not least important, (iii) the paradigm change in the understanding of underlying mechanisms of NCDs [13-16]. A commonality of these three driving forces is the need for efficient and intensive use of Information and Communication Technologies (ICT) to support novel Integrated Care Services (ICS) for chronic patients [17]. The convergence between ICS and novel systems-oriented biomedical research is a pivotal unmet need for an effective deployment of 4P *(Predictive, Preventive, Personalized and Participatory)* medicine, as conceived in the Synergy-COPD project.

## Systems medicine and workforce preparation

### The concept of Systems Medicine

As defined by Federow and Gostin [18], stemming from Systems Biology, Systems Medicine focuses on biomedical research incorporating interactions between all components of health and disease. Systems Medicine is an approach that aims at understanding the multitude of non-linear dynamic interactions among components of a living system and their interplay with the environment. In this scenario, biomedical questions are addressed through integrating experiments in iterative cycles using computational modeling [13-16,19,20] to capture emergent properties that arise from dynamic interactions at systems level and cannot be inferred with a reductionist approach. Systems Medicine is proposing an innovative approach to our understanding of disease mechanisms that is contributing to build-up patient-based predictive medicine with important implications on future strategies for patient management.

### Holistic focus and use of models for system description and prediction of the system response to interventions

Under the new paradigm of Systems Medicine, a holistic approach requires the challenge of data integration (-omics, physiologic, clinical and environmental) as well as use of computer modeling for knowledge generation. This challenge can only be successfully achieved with the incorporation of tools traditionally restricted to physics, mathematics and bioinformatics to the classical arsenal of molecular biology and physiology with which biomedicine researchers and physicians have been mainly trained during the last decades [21-24].

### Systems Medicine of chronic disorders

It is well accepted that NCDs are caused by complex gene-environment interactions and they show clear socio-economic determinants. Moreover, it is well known that several chronic diseases are often present in a given patient as co-morbid disorders. Both analysis and management of NCDs with a reductionist approach does not appear to be efficient.

Recent advances in systems biology including network analysis are opening new avenues suggesting that commonly clustered co-morbid conditions may share abnormal regulation of pivotal metabolic pathways. The promising scenario generated by Systems Medicine may contribute to markedly enhance chronic patient's management based on better knowledge of bio-pathological mechanisms explaining disease occurrence and progress, as well as underlying factors determining clustering of co-morbid disorders.

The enhanced knowledge should facilitate future early prevention strategies with a personalized approach that may likely modulate chronic disease progress improving patient's prognosis. We understand that this constellation of forces should prompt a future convergence between deployment of integrated care and predictive medicine of NCDs wherein novel ICT applications can play an enabling role.

### A novel approach to disease taxonomy: from symptoms and individual biomarkers to dysfunctional biological networks

Disease diagnosis is traditionally based on assessment of characteristic symptoms/signs and individual biomarkers that lead to disease diagnosis. It is of note that major disease taxonomies, as well as therapies, are often organ-based [18]. Instead, Systems Medicine shows high potential to

build-up novel and more functional disease taxonomies based on the achievements of network analysis producing knowledge on underlying mechanisms of diseases. The novel approach routed on understanding dysfunctional biological networks as disease mechanisms may open the way to innovative preventive and therapeutic strategies [15,16,18,20]. In this context, the goal of 4P medicine is to identify biological/environmental markers in order to enrol individuals at high risk for developing a disease in early detection and prevention tracks.

## Systems medicine post-graduate education: the need for multidisciplinary collaboration

A major bottleneck for the development of Systems Medicine (and 4P Medicine) is lack of availability of experienced biomedical researchers with strong background in bioinformatics/mathematics that are able to cope with the current needs for both knowledge and data integration. Before the existence of multi-disciplinary educational programs there was not a consensus if it was "*better to start with people from a domain of science and try to give them data management skills, or start with people from a computer science background and try to give them domain skills*" [25]. However, in the last decade new multi-disciplinary educational programs are being developed opening the avenue for new Systems Medicine-driven educational and training profiles. It is our understanding that biomedical research can greatly benefit from the rapid emergence of the inter-disciplinary training programs bridging biological sciences, applied mathematics and medicine.

By training we consider both the *training as the development of skills* (e.g. learning as the use of tools and databases or a experimental technique) and the *training as the learning of knowledge and understanding of fundamental concepts* (e.g. through theoretical courses) [26]. The end-users of the training are future researchers aiming to work on the Systems Medicine paradigm. Among them, we may differentiate two profiles: those with major interest in the clinical/biological field and those more focused on the method-development. In both cases, training with a Systems Medicine approach, applies. Moreover, two important aspects must be considered in the training programs in Systems Medicine:

- Training across academic disciplines: Medicine, Physiology, Biotechnology, Mathematics, Physics, Bioinformatics, etc...
- Training new generations to be able to identify those areas where a systems approach will be most suitable to better address clinical questions, solve clinical problems or contribute to make health sector economically sustainable. For example, the Summer School's experience within the Biohealth Computing program has shown that Chronic Obstructive Pulmonary Disease (COPD) as a case study is one of these areas wherein a systems approach can be highly productive.

### The CASyM initiative

Recently, laudable efforts are being made worldwide to implement new multidisciplinary initiatives with a systems medicine approach. To mention just two examples: The Coordinated Action for Systems Medicine (CASyM) at EU level [27] and the MD/MS program in Systems Medicine recently launched by Georgetown University in USA [28]. CASyM was initiated in 2012, engaging 22 partners from 11 countries, with the mission of developing European wide implementation strategy for Systems Medicine that should integrate multiple elements: from clinical trials, drug development, public health and to medical economics initiatives.

The CASyM roadmap aims to identify clinical questions/problems that can be better addressed/solved using a systems approach. CASyM aims to engage relevant stakeholders in the implementation of research programs. It is also within CASyM's goals to contribute to the development of multidisciplinary training concepts toward a sustainable Systems Medicine implementation strategy for Europe.

The success of Systems Medicine will depend strongly on the extent to which the leaders accomplish the creation of the environment that researchers need to develop an understanding among different working cultures, and manage also to implement strategies that integrate these cultures into shared working practices. Senior academics can have the initial vision for a project in Systems Medicine, but it will inevitably be the junior researchers, and their networking with peers, who will develop the collaborative relations that eventually will deliver results. The infrastructure and working environment for the Systems Medicine must therefore be one that generates and fuels interaction at all levels. Junior researchers need to be encouraged and supported to commit time to such relations, and the familiarity and knowledge that they gain in the other discipline needs to be recognized.

### Origin and mission of the BioHealth Computing (BioHC) Post-Graduate School

The BHC Post-Graduate School (i.e. Masters and PhD programs) was conceived in 2010 to foster multidisciplinary training of researchers with a Systems Medicine approach aiming at fostering European competitiveness in biomedical sciences as well, as clinical research, and public health services in the 21$^{st}$ century. The program was designed to be outward looking, opened to world-class research, ranging from basic science to clinical practice and industrial products. Considering that a major educational initiative in Systems Medicine will require a broad base of disciplines

associating Medical and Life sciences but also Chemistry, Physics and Computational mathematics. Five European universities (Grenoble, Barcelona, Torino, Maastricht and Cluj-Napoca) decided to join efforts in the BioHealth Computing consortium (BioHC). The BioHC consortium associates a unique network of public-private organizations, including university hospital, bio-parks, industries and non-European academia from Asia and Latin-America, to create opportunities and to catalyze the development of Systems Medicine in Higher Education and Research. The BioHC consortium draws on the large expertise and technological resources available at the partner institutions to provide Master and PhD students with the tools necessary to address real world problems in a creative, interdisciplinary team setting.

The aim of the BioHC consortium is to bridge the double gap between bedside, bench and engineering researches, and between academic lab, industrial R&D and healthcare services, by implementing a multidisciplinary and inter-institutional Post-Graduate program (Master and PhD). The BioHC School has been implemented in 2011 at Master level and in 2014 at Doctoral level. The BioHC School trains students to extend medical technologies and advance pharmaceutical R&D and applications by leveraging advanced mathematical and computational methods in medical research and practice. The BioHC students gain a deeper understanding of the dynamics of pathophysiological mechanisms and facilitate their early diagnosis. The BioHC program implements interdisciplinary approaches in medical research and clinical practice. Systemic modeling and use of simulation tools and IT are a must to overcome technological obstacles described above and to develop a Product Lifecycle Management (PLM) approach in Medtech and Pharmatech which should have important consequences for the structure and functioning of the biopharmaceutical innovation system with contributions of BioHC in the increase of industrial competitiveness and reduction of overall healthcare costs.

### BioHC Post-Graduate program

Currently, the education and training of Systems Medicine researchers begins at postgraduate level. Since Systems Medicine is a new area of research, it is difficult to single out any one model of education and training as superior at this early stage. The tendency reflects the importance of mastering a parent discipline before moving on to a more wide ranging training. Nonetheless, rooting in a parent discipline allows the acquisition of a thorough understanding of the scientific method as well as providing a safe return path. The importance of this should not be underestimated, especially in light of the currently uncertainty of career opportunities in cross-discipline research. Systems researchers should be given opportunities to maintain their expertise in their parent discipline, for example by provision of protected time to pursue mono-disciplinary research. The BioHC Partner universities have introduced a 2$^{nd}$ year Master and 3-year PhD program in Systems medicine. In countries like Spain wherein most undergraduate studies have a 4-year length, the Master degree has only 1-year duration instead of 2-years. In this case, the graduate students can enter directly into the BioHC Program.

#### *Required profiles and parent disciplines*

As mentioned, the mastering a parent discipline before enrolling the BioHC Master Course program (2$^{nd}$ year) is required. Parent disciplines learned by the students (during the 1$^{st}$ year of the Ms) have to fit with one of the four dominant academic tracks described below. A mobility scheme provides for at least two locations in the same track. The four tracks are:

**Clinical research**. The objectives of this track are: (a) Enhance research skills for residents from the EU and other areas (PhD); (b) Prepare Allied Health Care Professionals for research activities; (c) Professional development for physicians and Allied Health Care Professionals.

**Molecular biotechnology**. The applicants have to master the techniques for the development of biotechnological applications. The Biotechnology is the use of microorganisms, eukaryotic cells or biological substances, such as enzymes, to perform specific industrial or manufacturing processes.

**Environmental health**. The applicants have to know the principles of exposure assessment, and its role in environmental hygiene, epidemiology, toxicology and risk assessment. After completing this course the student has gained understanding of the importance and complexity of exposure assessment..

**Computational mathematics**. The students must know the mathematical methods and computational tools available for the dynamic modeling, analysis and validation, of biological regulatory networks. Lectures focus on qualitative and analytical aspects of mathematical modeling, as well as on real applications.

The BioHC Erasmus Mundus Master Program (60 ECTS) encompasses four well defined periods: (i) Integrative period of Master students at the Summer School (6 ECTS), (ii) First semester in one of the partner universities (24 ECTS); (iii) Second semester in a different partner university (24 ECTS); and, (iv) Consolidation period at the Summer School (6 ECTS). The BioHC program anticipates that upon receiving their Master's degree, or after a PhD program, graduates will work R&D department in industries or academia for developing innovative diagnostic, therapeutic or healthcare products.

### Main principles of the BioHC program

Four major principles are guiding the BioHC program:

***Problem-based teaching*** is an approach in which small groups of students deal with previously constructed problems. Under supervision of BioHC Staff members and Invited Professors, the students build-up learning pathways to the point of producing syntheses and new knowledge. The problems are descriptions of a phenomenon or event to be analyzed by the group using prior knowledge. From this, the students seek to understand the underlying processes, and questions arise. This process consolidates the learning objectives and serve as the individual and group study content. Following this, the students check whether new information leads to an understanding of the problem, and discuss it again. The tutor takes on the role of facilitator, thereby stimulating the process and reflection on it.

***Student-centered learning***. It is well established that the teacher should provide guidance, tools and tasks, but the student needs to find for him/her self the solutions. This process generates self-confidence. The students show similar learning pattern despite the speed and outcomes may be different among them, depending upon skills. The research training is spread over two semesters, and the curriculum, also distributed over two semesters, is defined from the research topic assigned to the students. Research topics assigned to students addressed complex problems involving an interdisciplinary approach. The student's projects are led by two supervisors from two different scientific areas. The aim of this Joint Research Project is to give to BioHC student opportunity to develop a translational and multidisciplicary approach.

For the assignment of one subject to one student, it is anticipated that the topics will be relevant to the parent discipline of the student, and will address a question of importance for the student. Rather than dissecting complex disease processes and studying the individual parts of living systems, the BioHC program focuses on understanding the complete system and the underlying interactions of all the forces that make up that system. The curriculum of the students includes core courses related to the parent discipline of the student, and elective courses related to complementary disciplines. One student enrolled primarily in Clinical research track can follow secondarily courses given in the Computational mathematics track, to acquire skills needed for modeling biological interactions embedded in a pathological process. The main learning output of this student-centered learning is the students' ability to work on a joint research program. There was increased motivation, leadership development and team working from the students. This was translated through their written work, seminars and portfolio preparation. The evaluation process for these experiences presupposes well-founded practices that express the views of the subjects involved: self-assessment and observer assessment.

***Mobility path and double degrees.*** BioHC Partner universities have decided, in 2011, to seek to provide a solid grounding in Systems Medicine, for their students not only in the form of local education but also by actively encouraging them to broaden their horizons by mobility path. Based on their research topics, BioHC students follow a mobility scheme where they attend two labs of the participating universities. Each student's project is co-supervised by two PI from two Partner universities, and the students must spend at least 1 semester in each university. The projects can associate also an industry or a partner hospital and one university. The supervisors, from primary and secondary hosting Institutions, constitute the candidate's Supervisory Committee. The two supervisors have equal responsibilities and cooperate with student in the choice of training activities, including the definition of mobility choices, and supervise their research. Where possible, evaluation is held on the occasion of a conference or similar event organised by any of the BioHC Partners.

The BioHC Program meets the specifications of the Erasmus Mundus from the European Commission (EACEA): (i) mobility path between at least the two Partner Institutions; (ii) and giving rise to double degree after only one defence at one of the two universities of attachment, according to the national regulations for Academic Degrees (Master and PhD) and University's evaluations. Erasmus Mundus is open to both European and non-European students, and there are generous scholarships and funding opportunities available.

### BioHC Thematic Schools

The BioHC program has successfully developed the Summer School since 2012 and proposes additional Thematic Schools (TS) on specific ICT domains that may be a useful bridge between applied mathematicians and bio-scientists.

#### *Interactive Training Initiative (Synergy-BioHC Summer School)*

The Interactive Training Initiative (ITI) is proposed to the Ms students during the BioHC Summer School, at the opening of the BioHC Master Courses. This program brings together students, scientists and health professionals, in the International Campus of the Archamps Technopole (Geneva Lake area), to work on clinical findings in all phases of the "bench to bedside" process. The ITI provides students with concepts, methodology and tools to respond to challenging situation in Clinical and Life Sciences. Content covers the fundamentals of physiological and clinical processes, along with core medical principles, clinical research methods, and clinical trials design, as well as basics of applied mathematics and computing. The program culminates in a capstone design-project in which

students work in interdisciplinary teams co-advised by faculty members and investigators from industries and hospitals. Initial focus areas include healthcare, therapeutic design and delivery, construction of medical devices, tissue engineering and regenerative medicine. A title of example, in Chronic Obstructive Pulmonary Disease (COPD), BioHC Master projects are focussed on studying network perturbations and key players that result in fatal Lung Cancer and COPD progression. Coursework material includes translational aspects of therapeutic science and principles of engineering design.

Introductory conferences on Systems medicine and P4 Medicine are given during each morning, by internationally recognized scientists, and students are invited, during each afternoon and evening, to work together on one R&D project, by group of 5-6 persons coming from mixed educations and skills. This R&D project is relevant to modern-day translational research in the field of Public Health. The focus of the ITI is on problem-based approach through team working: (i) realizing a goal; (ii) on utilizing the available expertise of the team and of the professional and teaching staff. Through coursework and collaborative research, the ITI program aims to enable candidates to combine experimental and theoretical approaches to develop physical and quantitative models of biological processes.

***School initiatives currently in the design phase***

The high volumes of data, the variety of data sources and the time-criticality of the analyses of biomedical and clinical problems are opening an opportunity window for big-data analytics. The BioHC partner universities are developing decision-support technological tools for analysis of huge amounts of data to help professionals, healthcare providers, payers and biomedical industry. To this end, big data analytics and computational modeling have been identified as relevant topics for thematic open school activities (Spring or Winter programs) following the principles of the BioHC program that should be addressed to both Master and PhD students.

## Conclusions

Convergence between Systems Medicine research and adoption of Integrated Care Services constitutes a fundamental strategic move toward consolidation of the new health paradigm that will make 4P medicine a reality. However, a key pillar for the transition phase is the preparation of the workforce with a multidisciplinary orientation. The BioHC program described in the current chapter has evolved through a successful learning period constituting a practical example of a systems-oriented training initiative that has consolidated productive interactions during the life span of the Synergy-COPD project.

**Competing interests**

DM is employed at Biomax Informatics AG. The other authors declare that they have no competing interests.

**Authors' contributions**

MC, PdA, PW, DM, DG-C, JR and PS conceived and designed the manuscript. MC, JR and PS write the first draft of the manuscript. MC, PdA, DG-C, JJC, SM, JR and PS contributed to the writing of the manuscript. PW, IC, FV, IMdM and DM revised the article critically for important intellectual content and gave final approval of the version to be submitted and any revised version.

**Acknowledgements**

We would like to thank all the members of the Synergy-COPD consortium for their support in the research described.

**Declaration**

Publication of this article has been funded by the Synergy-COPD European project (FP7-ICT-270086). The opinions expressed in this paper are those of the authors and are not necessarily those of Synergy-COPD project's partners or the European Commission.

This article has been published as part of *Journal of Translational Medicine* Volume 12 Supplement 2, 2014: Systems medicine in chronic diseases: COPD as a use case. The full contents of the supplement are available online at http://www.translational-medicine.com/supplements/12/S2.

**Authors' details**

[1]Departament de Bioquimica i Biologia Molecular i IBUB, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain. [2]Hospital Clinic - Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS). Universitat de Barcelona, 08036 Barcelona, Spain. [3]BioHealth Computing Erasmus Mundus Master Program, University Joseph Fourier, Grenoble, France. [4]Unit of computational Medicine, Center for Molecular Medicine, Department of Medicine, Karolinska Institute and Karolinska University Hospital, SE-171 76 Stockholm, Sweden. [5]UCSD Division of Physiology, Pulmonary and Critical Care Medicine, University of California, San Diego, La Jolla, California, USA. [6]Department of eHealth, Barcelona Digital, Roc Boronat 117, 08017 Barcelona, Catalunya, Spain. [7]Biomax Informatics AG, Robert-Koch-Str. 2, Planegg, Germany. [8]Centro de Investigación en Red de Enfermedades Respiratorias (CibeRes), 07110 Palma de Mallorca, Spain. [9]Equipe environnement et prédiction de la santé des populations, Laboratoire TIMC (UMR 5525), CHU de Grenoble, Université Joseph Fourier, La Tronche, France.

**Published: 28 November 2014**

**References**

1. Rosenbaum L, Lamas D: **Facing a "Slow-Motion Disaster" - The UN Meeting on Non-communicable Diseases.** *N Engl J Med* 2011, , **365**: 2345-2348.
2. Murray CJ, Lopez AD: **Mortality by cause for eight regions of the world: Global Burden of Disease Study.** *Lancet* 1997, **349**:1269-1276.
3. Murray CJ, Lopez AD: **Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: global Burden of Disease Study.** *Lancet* 1997, **349**:1347-1352.
4. Bousquet J, Anto J, Sterk P, Adcock I, Chung K, Roca J, Agusti A, Brightling C, Cambon-Thomsen A, Cesario A, *et al*: **Systems medicine and integrated care to combat chronic noncommunicable diseases.** *Genome Medicine* 2011, **3**:43.
5. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A: **Improving chronic illness care: translating evidence into action.** *Health Aff (Millwood)* 2001, **20**:64-78.
6. Epping-Jordan JE, Galea G, Tukuitonga C, Beaglehole R: **Preventing chronic diseases: taking stepwise action.** *The Lancet* 2005, **366**:1667-1671.
7. Epping-Jordan JE, Pruitt SD, Bengoa R, Wagner EH: **Improving the quality of health care for chronic conditions.** *Quality & safety in health care* 2004, **13**:299-305.
8. Pruitt S, Canny J, Epping-Jordan JA: **Preparing A Health Care Workforce For The 21st Century: The Challenge Of Chronic Conditions.** *Noncommunicable Diseases and Mental Health Cluster, Chronic Diseases and*

*Health Promotion Department* World Health Organization; 2005 [http://www.who.int/chp/knowledge/publications/workforce_report.pdf].
9. **Action Plan for Global Strategy for the Prevention and Control of Noncommunicable Diseases.** World Health Organization; 2013 [http://www.who.int/nmh/publications/ncd-action-plan/en].
10. Murray CJ, Lopez AD: **Measuring the global burden of disease.** *N Engl J Med* 2013, **369**:448-457.
11. De Martinis M, Franceschi C, Monti D, Ginaldi L: **Inflamm-ageing and lifelong antigenic load as major determinants of ageing rate and longevity.** *FEBS Lett* 2005, **579**:2035-2039.
12. Turan N, Kalko S, Stincone A, Clarke K, Turan N, Sabah A, Howlett K, Curnow SJ, Rodriguez DA, Cascante M, O'Neill L, Egginton S, Roca J, Falciani F: **A Systems Biology Approach Identifies Molecular Networks Defining Skeletal Muscle Abnormalities in Chronic Obstructive Pulmonary Disease.** *PLoS Comput Biol* 2011, **7(9)**:e1002129.
13. Wollenhauer O, Auffray C, Baltrusch S, Blüthgen N, Byrne H, Cascante M, Ciliberto A, Dale T, Drasdo D, Fell D, Ferrell JE Jr, Gallahan D, Gatenby R, Günther U, Harms BD, Herzel H, Junghanss C, Kunz M, van Leeuwen I, Lenormand P, Levi F, Linnebacher M, Lowengrub J, Maini PK, Malik A, Rateitschak K, Sansom O, Schäfer R, Schürrle K, Sers C, Schnell S, Shibata D, Tyson J, Vera J, White M, Zhivotovsky B, Jaster R: **Systems biologists seek fuller integration of systems biology approaches in new cancer research programs.** *Cancer Res T* 2010, **70**:12-3.
14. Wolkenhauer O, Auffray C, Jaster R, Steinhoff G, Dammann O: **The road from systems biology to systems medicine.** *Pediatr Res* 2013, **73**:502-507.
15. Auffray C, Charron D, Hood L: **Predictive, preventive, personalized and participatory medicine: back to the future.** *Genome Medicine* 2010, **2**:57, doi: 10.1186/gm178.
16. Shublaq N: **Strategic Report for Translational Systems Biology and Bioinformatics in the European Union.** 2012 [http://www.inbiomedvision.eu/pdf/report-translationalbioinformatics-final.pdf].
17. Vogiatzis G, NEXES: **Supporting Healthier and Independent Living for Chronic Patients and Elderly (CIP-ICT-PSP-2007-225025).** 2012 [http://www.nexeshealth.eu/videos.html].
18. Federoff HJ, Gostin LO: **Evolving from reductionism to holism: is there a future for systems medicine?** *JAMA* 2009, **302**:994-996.
19. [http://www.erasysbio.net/datapool/page/2/ERASysBio_Systems_Biology_Strategy_Paper_25-Mar-2008.pdf].
20. Sansone SA, Rocca-Serra P, Field D, Sansone SA, Taylor C, Hofmann O, Fang H, Neumann S, Tong W, Amaral-Zettler L, Begley K, Booth T, Bougueleret L, Burns G, Chapman B, Clark T, Coleman LA, Das S, de Daruvar A, de Matos P, Copeland J, Edmunds S, Evelo CT, Forster MJ, Gaudet P, Gilbert J, Goble C, Griffin JL, Jacob D, Kleinjans J, Harland L, Haug K, Hermjakob H, Ho Su SJ, Laederach A, Liang S, Marshall S, McGrath A, Merrill E, *et al*: **Toward interoperable bioscience data.** *Nat Genet* 2012, **44**:121-126.
21. Loscalzo J, Barabasi AL: **Systems biology and the future of medicine.** *Wiley Interdiscip Rev Syst Biol Med* 2011, **3**:619-27.
22. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin d, I M, Turan N, Cascante M, Falciani F, Hernandez M, *et al*: **Knowledge management for systems biology a general and visually driven framework applied to translational medicine.** *BMC Syst Biol* 2011, **5**:38.
23. Baranzini SE: **Systems-based medicine approaches to understand and treat complex diseases. The example of multiple sclerosis.** *Autoimmunity* 2006, **39**:651-662.
24. Quintana FJ, Farez MF, Weiner HL: **Systems biology approaches for the study of multiple sclerosis.** *J Cell Mol Med* 2008, **12**:1087-1093.
25. Weidman S, Anrson T: **Steps Toward Large-Scale Data Integration in the Sciences: Summary of a Workshop.** *National Research Council of the National Academies 2010* [http://www.nap.edu/catalog.php?record_id = 12916].
26. Schneider MV, Watson J, Attwood T, Rother K, Budd A, McDowall J, Via A, Fernandes P, Nyronen T, Blicher T, Jones P, Blatter MC, De Las Rivas J, Judge DP, van der Gool W, Brooksbank C: **Bioinformatics training: a review of challenges, actions and support requirements.** *Briefings in Bioinformatics* 2010, **11(6)**:544-51.
27. **Coordinating Action on Systems Medicine Across Europe ([www.casym.eu/]): EU-CASyM Joint Workshop: European Systems Medicine roadmap discussions.** [https://www.casym.eu/casym-systems-medicine/], Brussels, 29 November 2012..
28. **Systems Medicine Program at the Georgetown University.** [http://gumc.georgetown.edu/SPI/systemsmedicine].

doi:10.1186/1479-5876-12-S2-S11
**Cite this article as:** Cascante *et al.*: **Workforce preparation: the Biohealth computing model for Master and PhD students.** *Journal of Translational Medicine* 2014 **12**(Suppl 2):S11.