Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE ZOOTECNICHE

CicloXXIV

Settore scientifico-disciplinare di afferenza: AGR19

STUDY OF CANDIDATE GENES FOR PRODUCTION AND CARCASS TRAITS IN ITALIAN HEAVY PIGS: IDENTIFICATION OF NEW DNA MARKERS, EXPRESSION STUDIES AND ASSOCIATION ANALYSIS USING DIFFERENT EXPERIMENTAL DESIGNS

Presentata da**: FRANCESCA BERTOLINI**

Coordinatore Dottorato
Chiar.mo Prof. Paolo Bosi

Relatore
Chiar.mo Prof. Vincenzo Russo

Correlatore
Chiar.mo Prof Luca Fontanesi

esame finale anno 2012

CONTENTS

RIASSUNTO

L'allevamento del suino pesante italiano è principalmente orientato verso la produzione di prodotti stagionati di alto pregio. Particolarmente importante è la produzione del prosciutto crudo, che è strettamente regolata e richiede specifiche caratteristiche della carcassa correlate con le caratteristiche della coscia fresca. Inoltre dato che i suini vengono macellati ad un peso di circa 160kg, il settore dell'allevamento del suino pesante italiano si trova a dover far fronte a numerosi problemi di efficienza di produzione che sono correlati a tutti gli aspetti biologici legati a crescita, conversione alimentare, deposizione di grasso e così via. E' ben noto che caratteri produttivi e caratteristiche della carcassa sono in parte geneticamente determinate. Perciò come primo step per comprendere le basi genetiche di attributi che possono avere un impatto diretto o indiretto nella produzione del prosciutto crudo, l'approccio del gene candidato può essere usato per identificare marcatori a DNA associati a caratteri di importanza economica.

In questa tesi abbiamo investigato tre geni candidati per caratteri produttivi e relativi alla carcassa (in particolare deposizione di grasso e crescita) in razze suine usate per la produzione del prosciutto crudo, utilizzando differenti approcci sperimentali al fine di trovare marcatori molecolari associati con questi caratteri.

I primi due geni, descritti nel capitolo uno e due (*TRIB3* e *PCSK1* rispettivamente), sono stati scelti in base all'importante ruolo che hanno nello sviluppo dell'obesità umana, e poi considerati come geni candidati per il deposito di grasso nel suino.

Il capitolo uno riporta il sequenziamento di una porzione del gene *TRIB3* suino. Abbiamo identificato due polimorfismi (tra cui una mutazione missenso) nel primo esone codificante che erano in completo linkage disequilibrium. L'analisi *in silico* della mutazione missenso ha suggerito

che questa mutazione potrebbe avere putativi effetti funzionali. Abbiamo poi dimostrato attraverso studi di associazione che questo marcatore genetico era associato con spessore lardo dorsale in suini Large White Italiana Duroc Italiana in due differenti disegni sperimentali. Abbiamo analizzato anche l'espressione di questo gene, mostrando che *TRIB3* è espresso in vari tessuti, incluso grasso e muscolo scheletrico.

Nel capitolo due abbiamo riportato il risequenziamento di circa 5.1 Kb del gene *PCSK1* suino in differenti razze. Diverse SNPs sono state identificate ed utilizzate per la costruzione di aplotipi e per l'analisi di relazioni filogenetiche. Questo gene è stato ri-mappato nel cromosoma suino 2, fornendo informazioni che posso essere integrate nella versione del genoma suino Sscrofa10.2. L'espressione genica è stata valutata in diversi tessuti ed è stata usata come attributo nelle analisi di associazione. E' stata condotta un'analisi di associazione tra SNP (Single nucleotide polimorphism) di *PCSK1* e svariati caratteri, tra cui anche quelli produttivi e relativi alla carcassa, in 5 gruppi di suini provenienti da tre diversi disegni sperimentali. I risultati hanno indicato che le SNP analizzate erano associate a numerosi attributi tra cui spessore lardo dorsale e grasso intermuscolare visibile nella razza Duroc Italiana (DI) e performance di crescita nella Large White Italiana (LWI). Comunque gli effetti stimati nella LWI erano opposti rispetto agli effetti ripostati nella razza Duroc. Una suggestiva associazione ($P<0.10$) è stata osservata con l'attività della catepsina B muscolare che è un parametro importante della qualità della carne per la produzione del prosciutto crudo.

Nel capitolo tre abbiamo analizzato una SNP nel gene *MUC4* chè è già noto essere in stretto linkage disequilibrium col locus *F4bcR* in differenti popolazioni di suini. Questo locus è coinvolto in modo dominante nello sviluppo della suscettibilità all'infezione da ETEC (Enterotoxigenic *Escherichia coli*) ed è una delle maggiori cause di mortalità nei suinetti in fase di pre-svezzamento. Questo locus

sembra essere anche associato con il tasso di crescita. Dato che questo marcatore molecolare è usato per identificare animali suscettibili, abbiamo analizzato questo SNP in razze italiane locali ed abbiamo applicato un approccio di selective genotyping nelle tre principali razze commerciali di suino pesante (LWI, DI, e Landrace Italiana). Abbiamo osservato un'associazione dell'allele suscettibile con un più alto incremento medio giornaliero e spessore lardo dorsale nei suini LWI ed un più alto incremento medio giornaliero nella Landrace Italiana. Questo locus è un buon esempio della complessità nell'applicare selezione assistita da marcatori nelle razze suine.

ABSTRACT

Heavy pig breeding in Italy is mainly oriented for the production of high quality processed products. Of particular importance is the dry cured ham production, which is strictly regulated and requires specific carcass characteristics correlated with green leg characteristics. Furthermore, as pigs are slaughtered at about 160 kg live weight, the Italian pig breeding sector faces severe problems of production efficiency that are related to all biological aspects linked to growth, feed conversion, fat deposition and so on. It is well known that production and carcass traits are in part genetically determined. Therefore, as a first step to understand genetic basis of traits that could have a direct or indirect impact on dry cured ham production, a candidate gene approach can be used to identify DNA markers associated with parameters of economic importance.

In this thesis, we investigated three candidate genes for carcass and production traits (particularly fat deposition and growth) in pig breeds used for dry cured ham production, using different experimental approaches in order to find molecular markers associated with these parameters.

The first two genes, described in chapter one and two, (*TRIB3* and *PCSK1* respectively), were chosen according to their important role on human obesity development, and considering them as candidate genes for fat deposition in pigs.

Chapter one reports on the sequencing of a portion of the porcine *TRIB3* gene. We identified two polymorphisms (one was a missense mutation) in the first coding exon which were in complete linkage disequilibrium. *In silico* analysis of the missense mutation suggested that it could have a putative functional effects. Then, we demonstrated through association studies, that this gene marker was associated with back fat thickness in Italian Large White and Italian Duroc pigs in two different experimental designs. We analysed also the expression of this gene in different porcine tissues showing that *TRIB3* is expressed in several tissues, including fat and skeletal muscle.

In Chapter two we report on the resequencing of about 5.1 kb of the porcine *PCSK1* gene in different breeds. Several SNPs were identified and used for haplotype construction and analysis of phylogenetic relationships. This gene was re-mapped on porcine chromosome 2 providing information that could be integrated in the Sscrofa10.2 genome version. Gene expression was evaluated in different tissues and used as a trait in association analyses. Association analysis between *PCSK1* single nucleotide polymorphisms (SNPs) and production, carcass and several other traits were conducted on five groups of pigs from three different experimental designs. Results indicated that the analysed SNPs were associated with several traits including back fat thickness and visible intermuscular fat in Italian Duroc (ID) and growth performances in Italian Large White (ILW) and in ILW x Italian Landrace pigs. However, the effects estimated in the ILW were opposite to the effects reported in the ID pigs. Suggestive association (P<0.10) was observed with muscle cathepsin B activity that is an important meat quality parameter for the production of dry-cured hams.

In Chapter three, we analysed a SNP on the *MUC4* gene which is already known to be in close linkage disequilibrium with the *F4bcR* locus in different pig populations. This locus is involved in the development of susceptibility of ETEC (Enterotoxigenic *Escherichia coli*) infection in a dominant way and it is one of the major cause of mortality in pre-weaned piglets. This locus seems also to be associated with growth rate. Since this marker is used as a marker to identify susceptible animals, we analysed this SNP in Italian local breeds and applied a selective genotyping approach in the three main commercial Italian heavy pig breeds (ILW, ID and Italian Landrace). We observed an association of the susceptible allele with higher average daily gain and back fat thickness in ILW pigs and higher average daily gain in Italian Landrace. This locus is a good example of the complexity of applying marker assisted selection in pig breeding.

General introduction

GENERAL INTRODUCTION

Heavy pig breeding is a particular characteristic of the Italian pig industry. The first big difference but not unique, comparing with those of other countries, is the weight of slaughtered animals:160-180 kg instead of the common 100-130 kg of live weight of other pig production industries mainly oriented for the production of fresh meat. The reason for this weight is that these heavy pigs are destined to high quality processed products. The most important production is the dry-cured ham, which is safeguard and regulated by its POD ("Denominazione di origine protetta") status, geographic indicator for the quality of product (Reg. CEE 2081/92 substituted by Reg. CEE 510/2006; General Disciplinary of Parma ham; DOP Disciplinary of San Daniele ham). Therefore Italian heavy pig production is strictly regulated by rules fixed from Consortia, based on the experience of producers and scientific research. The main rules that breeders have to follow to produce POD dry-cured hams concern feeding, age and live weight of slaughtering and breeds in order to satisfy specific requirements. Dry-cured ham does not require any additive; the production is based only on salt addiction and through the monitoring of humidity and ambient temperature. It is clear that in this contest meat and carcass characteristics are extremely important. Above all, two main factors are considered: salting loss and back fat thickness. Salting loss can be defined as the "measure of the water-holding capacity under the salting condition"; Back fat thickness is fundamental to have enough fat coverage of the ham, following the rules of Consortia of Parma and San Daniele hams because this guarantees a correct seasoning without loosing excessive quantity of water and a conservation of the typical organoleptic components of the dry-cured ham (Bosi and Russo, 2004). Other important traits are also intramuscular and intermuscolar fat content which can reduce the seasoning loss; however an excess of this deposition is not positive, because of the possible "grassinatura" defect know as nut-like depots (Ufficio Tecnico ANAS, 2003).

In 2008, in Italy were slaughtered about 9.1 million of heavy pigs. The valorization of heavy pig for high quality season product is added by the effort to give value not only to POD products but also to the whole carcass for the fresh meat circuit, with the birth of the POD Gran Suino Padano label, which is produced according with rules for the production of dry-cured hams. Gran Suino Padano Consortium has been recognized by the Italian Ministry of Agriculture in 2007 (Piasentier et al., 2009)

CANDIDATE GENE ANALYSIS, A FEW EXAMPLES AND APPLICATIONS IN PIG BREEDING

During the last few years, the combination of molecular and quantitative genetics have modified traditional selection schemes in livestock. .

DNA-based technology and the genetic markers used for selection have accelerated the improvement of animal performances. It is well known that carcass traits are controlled by an unknown number of quantitative trait loci (QTL) and molecular markers can be used to identify individual genes which may have an effect on the regulation of productive and economically important traits. Marked Assisted Selection (MAS) was the first applied into selection programs for economically important traits even if at present Genomic Selection is substituting this approach. However, the first step is the identification of markers associated with production traits. A possible strategy is to analyse candidate genes. Candidate genes are genes whose product might be directly or indirectly involved in the biological processes that could affect production traits. Variability in these genes could explain a quote of variability for the targeted traits. Their biological roles should already be known, together with other information like genomic position and expression level, not necessary all in the species of interest. The candidate gene approach can be considered as a shortcut to identify markers associated with production traits even if sometimes it requires a large amount of time to perform significant analysis (Andersson, 2001).

Anyway, the candidate gene approach has been successfully applied in pigs. Several markers associated with production traits in pigs have been identified using this approach or a combination of this approach with QTL mapping.

To optimize this approach in pigs it could be worth to use information coming from humans. In particular many studies in human have already identified genes affecting obesity and fat deposition that could be investigated also in pig to identify markers of interested in pig breeding.

Here, we reports a few examples of the most important genes already shown to affect carcass and meat quality traits identified using a candidate gene approach or combining information from QTL mapping and other approaches.

RYR1

Ryanodine receptor 1 (RYR1) is an ion channel involved in regulation of the release of Ca2+ in skeletal muscle (Fujii et al., 1991). The polymorphism c.1843C>T is the causative mutation of one of the most important meat defect, the PSE (Pale, Soft, Exudative), an acronym that resumes aspects and characteristics of defective meat. PSE meat is for a detrimental defect for seasoned products therefore animals carrying this mutations have been eliminated from heavy pig breeds that are virtually considered free of this mutation.

.

PRKAG3

5'-AMP-activated protein kinase subunit gamma-3 is an isoform of the regulatory γ-subunit of AMPK (adenosine monofosphate activate protein kinase). The R225Q mutation determines what it was generally referred to as the RN (Rendement Napole)-phenotype (Milan et al., 2000). The RN phenotype, the so-called "acid meat", was identified in the Hampshire breed and it is caused by low

pH and high post-mortem degradation of muscle glycogen. Therefore this defect should be excluded for the production of dry cured hams. For this reason the Hampshire breed is not allowed in the genetic lines used for the production of the terminal fattened pigs whose legs are processed by the ham Consortia.

MC4R

This gene codes for the melanocortine 4 receptor, a G transmembrane protein receptor involved in regulation of metabolism. In human, mutations in this gene cause the most frequent form of monogenic childhood obesity, with more than 90 polymorphisms detected in various cohorts. It is inherited as dominant but its penetrance is normally incomplete and variable (Tao et al., 2006; Lubrano-Berthelier et al., 2006). In pig, a missense mutation c.1426G>A (Kim et al., 2000) in a conserved sequence of the seventh transmembrane region of the receptor is associated with back fat, growth rate, feed intake in a variety of commercial and experimental pig lines (for example see: Bruun et al., 2006; Piórkowska et al., 2010).

FTO

FTO codes for 2-oxoglutarate-dependent nucleic acid demethylase that probably plays a role in oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA (Jia et al., 2008). This gene was associated with body mass index and it is considered the most important genetic factor for common obesity in children and adults (Dina et al., 2007). In pigs, polymorphisms in this gene have been associated with several carcass and production traits (Fan et al., 2009; Fontanesi et al., 2009, 2010). In particular, Fontanesi et al. (2009) have shown that a polymorphism in this gene is associated with intermuscular fat content in Italian Duroc pigs providing a marker that could be used to improve this trait in this heavy pig breed.

IGF2

This gene is imprinted in several species including the pig. (de Chiara et al., 1991; Van Laere et al., 2003; Vu et al., 1994). A polymorphism in intron 3 of the porcine *IGF2* gene is the causative mutation of an important QTL localized on porcine chromosome 2 for lean meat and fat (Van Laere et al., 2003). This mutation has been shown to have an important effect on these traits, on growth rate and feed efficiency also in Italian heavy pigs (Fontanesi et al., 2010) with putative indirect effects on sow prolificacy (Stinckens A., et al., 2010).

CANDIDATE GENES STUDIED IN THIS THESIS

During the last few years many other studies have investigated candidate genes in pigs. However, considering traits that could be important in heavy pig production it was possible to select and analyse three other genes that are introduced below.

TRIB3

Tribbles homolog 3 (TRIB3) is a pseudokinase which is involved in the control of metabolic processes, stress response and cell viability. It has been shown that TRIB3 affects insulin action (Du et al., 2003) and suppresses adipocyte differentiation (Takahashi, et al., 2008). A missense mutation in the human *TRIB3* gene (p.Q84R) is associated with insulin resistance, cardiovascular disease, diabetes and obesity (Gong et al.,2009; Prudente et al., 2005; Prudente et al., 2009; Shi et al., 2009). Based on these evidences *TRIB3* could be considered a candidate gene for fat deposition traits in pigs.

PCSK1

The pro hormone convertase 1/3 enzyme (PCSK1) is involved in the maturation of several hormones which are important in regulation of central and peripherical metabolism. Several studies have demonstrated its involvement in the development of monogenic and polygenic obesity in humans (Benzinou *et al.*, 2008; Chang *et al.*, 2010; Farooqi *et al.*, 2007; Jackson *et al.*, 1997; Kilpeläinen *et al.*, 2009). For this reason it has been considered as a candidate gene to identify DNA markers associated with fat deposition traits in Italian heavy pigs.

MUC4

MUC4 encodes for a membrane-bound-O-glycoprotein which has a function of protecting and lubricating the epithelial surfaces. A SNP localized in exon 7 (DQ848681: g.8227C>G) of this gene is in very close linkage disequilibrium with the *F4bcR* locus (Jørgensen et al., 2004) that is responsible for the susceptibility to ETEC (Enterotoxigenc E.coli) infection in pre-weaned pigs (Bijlsma et al., 1982; Gibbons et al., 1977; Sellwood et al., 1975). This gene may also play many other important roles in growth, fetal development, epithelial renewal and differentiation, epithelial integrity, carcinogenesis, and metastasis (Corfield et al., 2001; Moniaux et al., 2001). This suggests that polymorphisms in the *MUC4* gene could be associated with other important production traits. In addiction a few studies of the *F4bcR* locus using villous based tests (Edfors-Lilja et al., 1986; Yan et al., 2009) have evidenced that the presence F4bcR could be associated with higher growth rate in pigs during the fattening period. All these elements suggest *MUC4* as a candidate gene for growth in pigs.

REFERENCES

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. Nature Reviews Genetics, 2:130-138.

Benzinou, M., Creemers, J.W., Choquet, H., Lobbens, S., Dina, C., Durand, E., Guerardel, A, Boutin, P., Jouret, B., Heude, B., Balkau, B., Tichet, J., Marre, M., Potoczna, N., Horber, F., Le Stunff, C., Czernichow, S., Sandbaek, A., Lauritzen, T., Borch-Johnsen, K., Andersen, G., Kiess, W., Körner, A., Kovacs, P., Jacobson, P., Carlsson, L. M., Walley, A. J., Jørgensen, T., Hansen, T., Pedersen, O., Meyre, D. and Froguel, P. (2008). Common nonsynonymous variants in PCSK1 confer risk of obesity. Nature Genetics 40, 943-945.

Bijlsma, I. G., de Nijs, A., van der Meer, C., and Frik, J. F. (1982). Different pig phenotypes affect adherence of Escherichia coli to jejunal brush borders by K88ab, K88ac, or K88ad antigen. Infection Immunity.;37:891–894.

Bosi, P., and Russo, V. (2004). The production of the heavy pig for high quality processed products. Italian Journal of Animal Science, 3, 309-321.

Briskey EJ Etiological status and associated studies of pale, soft and exudative porcine musculature. Advances in Food Research 1964, 13, 89-178.

Bruun, C. S., Jørgensen, C. B., Nielsen, V. H., Andersson, L., and Fredholm, M. (2006). Evaluation of the porcine melanocortin 4 receptor (MC4R) gene as a positional candidate for a fatness QTL in a cross between Landrace and Hampshire. Animal Genetics, 37(4), 359–362.

Chang, Y. C., Chiu, Y. F., Shih, K. C., Lin, M. W., Sheu, W. H, Donlon, T., Curb, J. D., Jou, Y. S., Chang, T. J., Li, H. Y. and Chuang, L. M. (2010). Common PCSK1 haplotypes are associated with obesity in the Chinese population. Obesity (Silver Spring) 18, 1404-1409.

Corfield, A. P., Carroll, D., Myerscough, and N., Probert, C. S. (2001) Mucins in the gastrointestinal tract in health and disease. Front Biosci.;6:D1321–D1357.

DeChiara, T. M. Robertson, and E.J. Efstratiadis, Parental imprinting of the mouse insulin-like growth factor II gene, Cell 64 (1991) 849±859.

Dina, C., Meyre, D.,Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L. M., Kiess, W., Vatin, V., Lecoeur, C., Delplanque, J., Vaillant, E., Pattou, F., Ruiz, J. Weill, J., Levy-Marchal, C., Horber, F., Potoczna, N., Hercberg, S., Le Stunff, C., Bougnères, P., Kovacs P., Marre, M., Balkau, B., Cauchi, S., Chèvre, J. C., and Froguel, P.. (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. Nature Genetics, 39, 724-726.

Du, K., Herzig, S., Kulkarni, R. N., and Montminy, M. (2003). TRB3: A tribbles homolog that inhibits Akt/PKB activation by insulin in liver. Science, 300, 1574−1577.

Edfors-Lilja, I., Petersson, H., Gahne, B.. (1986) Perfomance of pigs with or without the intestinal receptor for Escherichia coli K88. Anim Prod;42:381–387.

Farooqi, I. S., Volders, K., Stanhope, R., Heuschkel, R., White, A., Lank, E., Keogh, J., O'Rahilly, S. and Creemers, J. W. (2007). Hyperphagia and early-onset obesity due to a novel homozygous missense mutation in prohormone convertase 1/3. Journal of Clinical Endocrinology and Metabolism 92, 3369-3373.

Fan, B., Du, Z. Q., and Rothschild, M. F. (2009). The fat mass and obesity-associated (FTO) gene is associated with intramuscular fat content and growth rate in the pig. Anim Biotechnol. 2009;20(2):58-70.

Fontanesi, L., Scotti, E., Buttazzoni, L., Davoli, R., and Russo, V. (2009). The porcine fat mass and obesity associated (FTO) gene is associated with fat deposition in Italian Duroc pigs. Anim Genet. 2008, 40(1):90-93.

Fontanesi, L.,Speroni, C., Buttazzoni, L., Scotti, E., Dall'Olio, S., Nanni Costa, L., Davoli, R, and Russo,V. (2010). The insulin-like growth factor 2 (IGF2) gene intron3-g.3072G>A polymorphism is not the only Susscrofa chromosome 2p mutation affecting meat production and carcass traits in pigs: evidence from the effects of a cathepsin D (CTSD) gene polymorphism. J Anim Sci. 88(7):2235-45.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V. K., Weiler, J. E., O'Brien, P. J. and MacLennan, D.H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science 1991, 253(5018), 448-451.

Gibbons, R. A., Sellwood, R., Burrows, M., Hunter, P. A.. (1977) Inheritance of resistance to neonatal Escherichia coli diarrhea in pig: examination of the genetic system. Theor Appl Genet.; 51:65–70

Gong, H.P., Wang, Z.H., Jiang, H., Fang, N.N., Li, J.S., Shang, Y.Y., Zhang, Y., Zhong, M.,& Zhang, W. (2009). TRIB3 functional Q84R polymorphism is a risk factor for metabolic syndrome and carotid atherosclerosis. Diabetes Care, 32, 1311-1333.

Jackson, R. S., Creemers, J. W., Ohagi, S., Raffin-Sanson, M. L, Sanders, L., Montague, C. T., Hutton, J. C and O'Rahilly, S. (1997). Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. Nature Genetics 16, 303-306.

Jia G., Yang, C. G., Yang, S., Jian, X., Yi, C., Zhou, Z., He, C.: (2008) Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human. FEBS Lett.,15; 582(23-24): 3313–3319

Jørgensen, C. B., Cirera, S., Archibald, A. L., Andersson, L., Fredholm, M., Edfors-Lilja, I.(2004) Porcine polymorphisms and methods for detecting them. Patent number WO2004048606.

Kilpeläinen, T. O., Bingham, S. A., Khaw, K. T., Wareham, N. J. and Loos, R. J. (2009). Association of variants in the PCSK1 gene with obesity in the EPIC-Norfolk study. Human Molecular Genetics 18, 3496-3501.

Kim, K. S., Larsen, N., Short, T., Plastow, G., & Rothschild, M. F. (2000). A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. Mammalian Genome: Official Journal of the International Mammalian Genome Society, 11(2), 131–135.

Le Roy, P., Naveau, J., Elsen, J. M., and Sellier ,P. (1990). Evidence for a new major gene influencing meat qualityin pigs. Genetical research, 55(1), 33–40

Lubrano-Berthelier, C, Dubern, B., Lacorte, J. M., Picard, F., Shapiro, A., Zhang, S., Bertrais, S., Hercberg, S., Basdevant, A., Clement, K., and Vaisse, C. (2006). Melanocortin 4 receptor mutations in a large cohort of severely obese adults: prevalence, functional classification, genotype-phenotype relationship, and lack of association with binge eating. J Clin Endocrinol Metab. 9,1811–1818.

Milan, D., Jeon, J .T., Looft, C., Amarger, V., Robic, A., Thelander, M., et al. (2000). A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science, 288(5469), 1248–1251.

Moniaux, N., Escande, F., Porchet, N., Aubert, J. P., and Batra, S. K.. (2001). Structural organization and classification of the human mucin genes. Front Biosci.;6:D1192–D1206.

Piasentier, E., Favotto, S., Saccà, E., Sepulcri, A., and Vitale M. (2009). qualitá della carne del suino pesantein rapporto al genotipo e allo stato d'ingrassamento. Qualità e sicurezza nella fiiera del prosiutto. 23-29.

Piórkowska, K., Tyra, M., Rogoz, M., Ropka-Molik, K., Oczkowicz, M., & Rózycki, M. (2010). Association of the melanocortin-4 receptor (MC4R) with feed intake, growth, fatness and carcass composition in pigs raised in Poland. Meat Science, 85(2), 297–301.

Prudente, S., Hribal, M.L., Flex, E., Turchi, F., Morini, E., De Cosmo, S., et al. (2005). The functional Q84R polymorphism of mammalian Tribbles homolog TRB3 is associated with insulin resistance and related cardiovascular risk in Caucasians from Italy. Diabetes, 54, 2807-2811.

Prudente, S., Scarpelli, D., Chandalia, M., Zhang, Y.Y., Morini, E., Del Guerra, S., et al. (2009).The TRIB3 Q84R polymorphism and risk of early-onset type 2 diabetes. Journal of Clinical Endocrinology and Metabolism, 94, 190-196

Sellwood, R., Gibbons, R. A., Jones, G. W., and Rutter, J. M. (1975). Adhesion of enteropathogenic. Escherichia coli to pig intestinal brush borders: the existence of 2 pig phenotypes. J Med Microbiol.;8:405–411.

Shi, Z., Liu, J., Guo, Q., Ma, X., Shen, L., Xu, S., Gao, H., Yuan, X., and Zhang, J. (2009). Association of TRB3 gene Q84R polymorphism with type 2 diabetes mellitus in Chinese population. Endocrine, 35, 414-419

Stinckens, A., Mathur, P., Janssens, S., Bruggeman, V., Onagbesan, O. M. , Schroyen, M., Spincemaille, G., Decuypere, E., Georges, M., Buys, N. (2010). Indirect effect of IGF2 intron3 g.3072G>A mutation on prolificacy in sows. Anim Genet. 41(5):493-8.

Takahashi, Y., Ohoka, N., Hayashi, H., and Sato, R. (2008). TRB3 suppresses adipocytes differentiation by negatively regulating PPARγ transcriptional activity. Journal of Lipid Research, 49, 880-892.

Tao, Y. X. (2009). Chapter 6 mutations in melanocortin-4 receptor and human obesity. Progress in Molecular Biology and Translational Science, 88, 173–204

UFFICIO TECNICO ANAS, 2003. Il contributo dellaselezione italiana alla filiera del suino pesante.Inf. Agr. 59 (13, suppl.): 7-8.

Van Laere, A.-S., M. Nguyen, M. Braunschweig, C. Nezer, C. Collette, L. Moreau, A. L. Archibald, C. S. Haley, N. Buys, M. Tally, G. Andersson, M. Georges, and L. Andersson. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425:832–836.

Vu, T. H.. and Hoffman, A. R. (1994). Promoter-specifc imprinting of the human insulin-like growth factor-II gene, Nature 371 714±717.

Yan, X. M., Ren, J., Huang, X., Zhang, Z. Y, Ouyang, J., Zeng, W. H., Zou, Z. Z, Yang, S. J., Yang, B., and Huang, L. S. (2009). Comparison of production traits between pigs with and without the Escherichia coli F4 receptors in a White Duroc x Erhualian intercross F2 population. J Anim Sci.87:334–339.

AIM

AIM

The aim of this thesis was to apply a candidate gene approach to identify DNA markers associated with carcass and production traits in Italian heavy pigs. Three genes were investigated: *TRIB3, PCSK1* and *MUC4*.

Different experimental designs and approaches were used to this purpose: we sequenced parts of these genes, identified and analysed polymorphisms, performed expression studies in different pig tissues, and used identified markers in association analyses in different populations applying a selective genotyping approach or investigating random purebred or crossbred populations.

Chapter one

The porcine *tribbles homolog 3* (*TRIB3*) gene: identification of a missense mutation and association analysis with meat quality and production traits in Italian heavy pigs

L. Fontanesi [a], M. Colombo [a], E. Scotti [a], L. Buttazzoni [b], F. Bertolini [a], S. Dall'Olio [a], R. Davoli [a], V. Russo [a]

[a] DIPROVAL, Sezione di Allevamenti Zootecnici, Faculty of Agriculture, University of Bologna, Via F.lli Rosselli 107, 42123 Reggio Emilia, Italy

[b] Associazione Nazionale Allevatori Suini (ANAS), Via Lazzaro Spallanzani 4/6, 00161 Roma, Italy

Published in ***Meat Science*** 2010 86(3):808-13

Abstract

TRIB3 plays an important role in the energy metabolism. This work aimed to study the porcine tribbles homolog 3 (*TRIB3*) gene and to evaluate its association with meat quality and carcass traits in pigs. By sequencing a portion of the porcine *TRIB3* gene we identified two single nucleotide polymorphisms (SNPs) in the first coding exon (one synonymous SNP: c.132T>C; and one missense mutation: c.146C>T, p.P49L). The two polymorphisms were in complete linkage disequilibrium. In silico analysis of the p.P49L mutation suggested that it could have functional effects. Association studies in four groups of pigs (651 animals in total) indicated that this gene marker was associated with back fat thickness in Italian Large White and Italian Duroc pigs in two different experimental designs ($P<0.1$ and $P<0.05$). This polymorphism tended to be associated with lactate content of *semimembranosus* muscle ($P<0.1$). Among several other tissues, *TRIB3* is expressed in fat and skeletal muscle.

Keywords: association study; back fat thickness; Italian heavy pigs; missense mutation; SNP; TRIB3

.

1 **Introduction**

DNA markers in several obesity-related genes have been associated with fat deposition and carcass traits in different pig populations. For example, following evidences that the fat mass and obesity associated (*FTO*) gene explains an important quote of the genetic variability of common obesity in humans (Dina et al., 2007; Frayling et al., 2007), polymorphisms in the porcine *FTO* gene have been identified and association analyses have shown that this gene might be an important source of variability for fat deposition traits in several pig breeds and populations (Fan, Du, & Rothschild, 2009; Fontanesi, Scotti, Buttazzoni, Davoli, & Russo, 2009; Fontanesi et al., 2010). Similar other studies in pigs have successfully investigated candidate genes chosen according to their function and role on the target physiological or production traits.

Tribbles homolog 3 (TRIB3) is a pseudokinase implicated in the control of metabolic processes, stress response and cell viability. TRIB3 affects insulin action by binding to and inhibiting Akt phosporylation, that is a key step in insulin signalling (Du et al., 2003). In addition, TRIB3 suppresses adipocyte differentiation by down regulating PPARγ transcriptional activity (Takahashi, Ohoka, Hayashi, & Sato, 2008). A missense mutation in the human *TRIB3* gene (p.Q84R) is associated with insulin resistance, cardiovascular disease, diabetes and obesity (Gong et al., 2009; Shi et al., 2009; Prudente et al., 2005; Prudente et al., 2009).

Based on these evidences *TRIB3* could be considered an important candidate gene for fat deposition traits in pigs. Moreover, its role in energy metabolism, and particularly on muscle glycogen content, could bring consequences also on meat quality traits in pigs.

Here we investigated the porcine *TRIB3* gene and identified polymorphisms (including a missense mutation) that were used in association studies with several meat quality, meat production and carcass traits in different Italian heavy pig populations.

2. Material and methods

2.1. Animals

The association study was conducted on 4 groups of heavy pigs (3 groups of Italian Large White animals and one group of Italian Duroc pigs) individually performance tested at the Test Station of the National Pig Breeder Association (ANAS). These animals are structured on triplets of siblings from the same litter (two females and one castrated male) and their data are used for the genetic evaluation of a boar from the same litter (sib-testing). An Italian Large White group was made by 266 pigs of this breed (177 females and 89 castrated males, from 79 different sires), not selected by any phenotypic criteria (random group) and already described (Fontanesi et al., 2008b; a few animals analysed in the previous study were not included here because it was not possible to obtain a genotype for the *TRIB3* marker). This group of pigs was slaughtered on 6 different days in 2002. The other two Italian Large White groups were made by 100 animals each chosen among 3591 Italian Large White pigs evaluated in the period 1996-1999, according to a selective genotyping approach based on two different estimated breeding values (EBVs; see below for methods of calculation and details about the traits). One group of 100 pigs (69 females and 31 castrated males from 71 different sires) had extreme divergent EBVs (50 with the highest and 50 with the lowest EBV) for back fat thickness (BFT). The second group of 100 pigs (69 females and 31 castrated males from 73 different sires) had extreme divergent EBVs (50 with the highest and 50 with the lowest EBV) for lean cuts (LC). While 200 extreme EBVs were considered for the two traits (BFT and LC), only 178 different pigs were analysed because 22 animals presented extreme values for both traits. The group of Italian Duroc pigs was made by 208 animals (131 females and 77 castrated males, from 105 different sires) slaughtered in the years 1995-2003 in 59 different days and unselected by any phenotypic criteria.

Another panel of unrelated pigs of different breeds with no phenotypic record, was used for allele frequency evaluation (Table 1).

Blood, hair root and/or muscle samples collected from all these pigs were used for DNA extraction carried out with standard protocols.

2.2. Performance, carcass and meat quality traits

The test period began when animals were 30 – 45 days old and it ended when they reached 155 ± 5 kg live weight. The nutritive level was *quasi ad libitum*. Feed intake was recorded daily and body weight was measured bimonthly, then daily gain and feed:gain ratio (FGR) were calculated. At the end of the test, animals from two batches on trial (the remaining from the older and the faster growing from the more recent batch) were mixed at loading and transported to a commercial slaughterhouse located 24.5 km from the Test Station. After unloading, pigs were immediately stunned by CO_2 (concentration 87%) using a dip lift system (Butina, Denmark) and bled in a lying position. Within 3 hours after slaughter, BFT at the level of *Musculus gluteus medius*, weight of LC (necks and loins) and weight of hams (only in Italian Large White pigs) were collected. Moreover, only for the Italian Duroc pigs visual evaluation of intermuscular fat content (VIF) was obtained on leg muscles based on a binary scale (presence/absence).

In addition, for the group of 266 Italian Large White pigs measures of pH_1 (at 2 h *post mortem*) and pH_u (at 24 h *post mortem*) on *Musculus semimembranosus* were taken by a Crison pH-meter equipped with an Ingold Xerolite electrode (Mettler Toledo, Udorf, Switzerland). Samples from the same muscle were collected at 30 min *post mortem* and immediately frozen in liquid nitrogen and later freeze-dried for glycolytic potential (GP) determination. This parameter was measured according to Monin, Mejenes-Quijano, Talmant, & Sellier (1987) as described in Fontanesi et al. (2008b), separately determining lactate content and the sum of glycogen, glucose

and glucose-6-phosphate. Glycolytic potential was calculated as the sum of: 2[glycogen + glucose + glucose-6-phoshate] + [lactate] according to Monin & Sellier (1985) and expressed as μmol of lactic acid equivalent per g of fresh muscle.

2.3. Identification and analysis of DNA polymorphisms

BLASTN queries with a human *TRIB3* cDNA (GenBank accession number NM_021158) against porcine sequences deposited in DNA databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi) were used to retrieve pig *TRIB3* gene sequences. *In silico* mapping of the identified porcine *TRIB3* sequences was obtained using the ENSEMBL BLAT search tool (http://www.ensembl.org/Multi/blastview) against the Sscrofa9 genome assembly. PCR primers (forward: 5'-CACTGGCAGAGGAAGAGAGG-3'; reverse: 5'-AATATCAGGACCCCCTCAGC-3') were designed to amplify a fragment of a 456 bp including exon 4.

Identification of single nucleotide polymorphisms (SNPs) in the porcine *TRIB3* gene was carried out by sequencing a 456 bp fragment from 40 pigs of different breeds (16 Italian Large White, 16 Italian Duroc, 1 Italian Landrace, 4 Belgian Landrace, 1 Hampshire and 2 Meishan). PCR was carried out using a T-Gradient (Biometra, Goettingen, Germany) thermal cycler in a final volume of 20 μL that included 10 pmol of each primer 1.5 mM $MgCl_2$, 2.5 mM each dNTP, 1 U AmpliBioTherm Taq DNA polymerase (Fisher Molecular Biology, Trevose, PA, USA). The PCR profile was the following: an initial step of denaturation for 5 min at 95 °C; 35 cycles of 30 s at 95 °C, 30 s at 63 °C and 30 s at 72 °C; the final extension step was for 5 min at 72 °C. Sequencing reactions were produced for ExoSAP-IT® (USB Corporation, Cleveland, Ohio, USA)-treated PCR products using the same PCR primers and the Big Dye v3.1 kit (Applied Biosystems, Foster City, CA, USA). Sequencing products, after purification steps, were loaded

on an ABI3100 Avant sequencer (Applied Biosystems). Sequencing electropherograms were analysed using CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA).

A polymorphism identified in the *TRIB3* gene was genotyped by PCR-RFLP in all animals used for the association studies and for allele frequency evaluation. Briefly, 5 µL of PCR products was digested overnight at 37 °C with 3 U of *Hinf*I (Fermentas, Vilnius, Lithuania) restriction enzyme in a final volume of 25 µL containing 1X enzyme reaction buffer. PCR-RFLP products were resolved on 10% polyacrylamide/bis-acrylamide 29:1 gels and visualized with 1X GelRed Nucleid Acid Gel Stain (Biotium Inc., Hayward, CA, USA). PCR-RFLP patterns for this polymorphism were: allele c.132T resulted in two fragments of 259 and 197 bp while allele c.132C was not cut (a fragment of 456 bp). Pigs used in the association studies were also genotyped, as described in Russo et al. (1993), for the *RYR1* c.1843C>T mutation, causing the pale, soft, exudative defect in the meat (Fujii et al., 1991). All animals resulted with genotype c.1843CC.

2.4. In silico *functional analysis of the missense mutation*

In silico functional analysis of the novel identified missense mutation was obtained using the SIFT (Sorting Intolerant From Tolerant) program (Ng & Henikoff, 2003) and the evolutionary analysis of coding SNPs tool of PANTHER (Protein ANalysis THrough Evolutionary Relationships; Thomas et al. 2003; Thomas & Kejariwal, 2004), whose predictions have been experimentally validated (Brunham et al., 2005). SIFT is a sequence homology-based tool that sorts intolerant from tolerant amino acid substitutions and predicts whether an amino acid substitution at a particular position in a protein will affect protein function and hence, potentially alter the phenotype. Positions with normalized probabilities <0.05 are predicted to be deleterious. To evaluate this prediction, SIFT calculates the median conservation value, which measures the

diversity of the sequences in the alignment. Conservation is calculated for each position in the alignment and the median of these values is obtained. Conservation ranges from $\log_2 20 = 4.32$, when a position is completely conserved and only one amino acid is observed, to zero, when all 20 amino acids are observed at a position. Predictions based on sequence alignments with median conservation values higher than 3.25 are less diverse and will have a higher false positive error. SIFT analysis was carried out against NCBI non redundant database (Dec. 2008) using default options. PANTHER estimates the likelihood of a particular non-synonymous (amino-acid changing) coding SNP to cause a functional impact on the protein. It calculates the substitution position-specific evolutionary conservation (subPSEC) score based on an alignment of evolutionarily related proteins (Thomas et al. 2003; 2006; Thomas & Kejariwal, 2004). The probability that a given variant will cause a deleterious effect on protein function is estimated by $P_{deleterious}$, such that a subPSEC score of -3 corresponds to a $P_{deleterious}$ of 0.5 (Brunham et al., 2005). The subPSEC score is the negative logarithm of the probability ratio of the wild-type and mutant amino acids at a particular position. PANTHER subPSEC scores are continuous values from 0 (neutral) to about -10 (most likely to be deleterious).

2.5 RNA isolation and reverse transcription PCR

Isolation of total RNA from back fat, brain, heart, kidney, liver, lung, skeletal muscle, spleen and thyroid specimens (about 100 mg), collected after slaughtering from a castrated male pig, was carried out using the RNeasy® Lipid Tissue kit (Qiagen, Duesseldorf, Germany) (for back fat and brain) or the FastPure™ RNA kit (TaKaRa Bio Inc., Shiga, Japan) (for all other tissues), following the manufacturers' instructions. After RNA extraction, about 1 µg of total RNA for each tissue was treated by RNase-Free DNase set (Qiagen) and 40 ng were retrotranscribed with Improm II Reverse Transcription system (Promega Corporation, Madison, WI) using oligo(dT)

primers and following the manufacturer's protocol. A fragment of 211 bp of the *TRIB3* cDNA was amplified with the primers designed on exon 4 (forward: 5'-AGGAGCACATAGGGTCCAAG-3'; reverse: 5'-CACCTCTGGCTGTTTCCACT-3') using the same conditions reported above except that annealing temperature was set at 57 °C. *GAPDH* cDNA amplification, used as reference, was obtained with primer pair GAPDH_437 reported in Fontanesi, Colombo, Beretti, & Russo (2008a). *TRIB3* and *GAPDH* amplified cDNA fragments were electrophoresed as reported above. All analyses were repeated three times.

2.6. Association studies

Breeding values for average daily gain (ADG, expressed in g), LC (expressed in kg), BFT (expressed in mm), ham weight (HW, expressed in kg), FGR and VIF were predicted by a BLUP-multiple trait animal model including the fixed factors of age at the beginning of test, body weight at slaughter, age at slaughter, day of slaughtering and inbreeding coefficient, besides the random factors of animal and litter.

For the random group of performance tested Italian Large White and Italian Duroc pigs, associations between the *TRIB3* genotypes and EBVs for the different traits were assessed by the GLM procedure of SAS, release 8.02 (SAS Institute Inc., Cary, NC). The model included only fixed effects of the marker genotypes, because all other factors contributing to the variability of the investigated traits were already corrected in the calculation of EBVs. For the same group of Italian Large White pigs, the MIXED procedure (SAS Institute Inc.) was used to evaluate associations between the same genotypes and meat quality parameters (pH_1, pH_u, glycogen and lactate content and GP). This model included sire as random effect and the fixed effects of day of slaughter and sex besides the genotype at the analysed DNA marker. Additive genetic effect in the Italian Large White population was estimated as half of the difference between the EBVs (or

between phenotypic traits) of the two homozygous groups. Dominance effect was estimated as the difference between the EBV (or between phenotypic traits) of the heterozygous group and the average of the two homozygous groups. Differences from zero of the estimates of additive and dominance effects were tested by *t*-test (Russo et al., 2008).

For the two groups of Italian Large White pigs chosen for selective genotyping based on extreme values of EBVs for BFT and LC, respectively, Fisher's exact test of significance (two-tailed) of differences in allele frequency between the positive and negative groups was used.

3. Results and discussion

3.1. Identification and analysis of mutations

A few porcine sequences (one expressed sequence tag, FD628621; and two draft sequences of chromosome 17 BAC clones, CU856069 and CU606854) including portions of the *TRIB3* gene were identified by BLASTN analysis (e-value <1e-53). A fragment of 456 bp corresponding to exon 4 (the first coding exon) and including parts of intron 3 and intron 4 (according to the organization of the human *TRIB3* gene, Ensembl ENSG00000101255 release 37, March 2010) was selected for further analyses. *In silico* mapping of this porcine sequence confirmed its assignment to porcine chromosome (SSC) 17 (position 36752125-36752540 nucleotides), that in the meantime was assembled and partially annotated. As the human *TRIB3* gene is located on chromosome 20p13-p12.2, this assignment is in agreement with the human-porcine comparative mapping data (Lahbib-Mansais et al., 2005; Hart et al., 2007). Resequencing of the 456 bp fragment in 40 pigs of different breeds revealed two SNPs, a synonymous polymorphism (c.132T>C) and a nonsynonymous mutation (c.146C>T) causing the p.P49L amino acid substitution (EMBL accession no. FN677934). This missense mutation is in a quite good

conserved position of the TRIB3 protein (Figure 1). *In silico* evaluation of its effects obtained with SIFT indicated that the p.P49L substitution affect protein function with a score of 0.01. Median sequence conservation was 3.73 probably due to the low number (n. = 20) of sequences available in databases used for the SIFT analysis represented at this position. Therefore this prediction, even if highly significant should be considered with caution and may need other evidences. Another *in silico* analysis of the p.P49L amino acid change was obtained with PANTHER that produced a subPSEC = -3.82231 and a $P_{deleterious}$ = 0.69473, supporting the indication obtained with SIFT.

According to our sequencing results based mainly on Italian Large White and Italian Duroc pigs, the two identified polymorphisms resulted in complete linkage disequilibrium within and across breeds, therefore a convenient PCR-RFLP protocol was used to genotype the c.132T>C SNP that captured information of the two haplotypes we identified in the porcine *TRIB3* gene ([T:C] and [C:T], considering the two polymorphic sites). Haplotype frequencies obtained by genotyping the c.132T>C polymorphism have been analysed in 7 different pig breeds (Table 1). Haplotype [C:T] resulted the most frequent in all studied breeds except that in Italian Large White and Belgian Landrace.

3.2. Expression of the porcine **TRIB3** *gene in different tissues*

Few data on the expression of the *TRIB3* gene are available in other species (Du et al., 2003) and no information has been reported in pigs yet. Therefore, we analysed different porcine tissues to have a first overview on *TRIB3* gene expression (Figure 2). Expression of this gene was evidenced in back fat, brain, heart, liver, skeletal muscle and spleen whereas it was not detected in kidney, lung and thyroid. These results could represent a first step in understanding the

biological roles of the *TRIB3* gene related to its effects on production traits as shown by the association studies reported below.

3.3. Association analyses

Table 2 reports the results of association studies carried out in the Italian Large White and Italian Duroc populations unselected by any phenotypic criteria. It was interesting to note that in both groups of pigs *TRIB3* diplotypes were suggestively associated ($P<0.1$) with BFT. The diplotype that was homozygous for the p.49L amino acid ([T:C]) showed higher BFT (positive EBV) in both breeds. In addition, mainly for Italian Duroc pigs, the direction of the effects on the EBVs of the other traits, even if not significant, was in agreement with the well established correlation between those traits. In the Italian Large White pigs, *TRIB3* diplotypes showed effects on lactate content ($P = 0.0673$) while its effects on glycolytic potential, even if not significant, showed the same tendency of lactate content. For these traits the alternative diplotype ([C:T]/[C:T]) homozygous for the p.49P amino acid, showed lower values. Additive and dominance effects for traits resulting with $P<0.20$ are reported in Table 3. Additive effects ($P<0.05$) were observed for BFT in both breeds, for ADG and LC EBVs in the Italian Duroc pigs, and for glycolytic potential in Italian Large White animals only. In the latter breed dominance effect ($P<0.10$) was obtained for lactate content.

The results obtained for BFT were confirmed in the groups of Italian Large White pigs selected according to the extreme values of their EBV (Table 4). Haplotype [T:C] (containing the p.49P amino acid) was more frequent in pigs with higher BFT EBVs (two tailed Fisher exact tests: $P<0.05$ considering all animals; $P<0.01$, considering only two-generation unrelated pigs).

It is interesting to note that other studies have reported several QTL for meat quality, growth and carcass traits on SSC17, in regions overlapping (in some cases) the position of the *TRIB3* gene

(Malek et al., 2001; Pierzchala et al., 2003; Rohrer, Thallman, Shackelford, Wheeler, &, Koohmaraie 2006). Moreover, analyses of SSC17 candidate gene markers have confirmed, at least in part, the effects observed in these QTL studies (Fan, Glenn, Geiger, Mileham, & Rothschild, 2008; Fan et al., 2010; Ramos, Bastiaansen, Plastow, & Rothschild, 2009; Ramos, Glenn, Serenius, Stalder, & Rothschild, 2008; Russo et al., 2008). In particular, several gene markers have been studied to analyse meat colour and related quality traits and their pleiotropic effects on other production traits, but it appears that investigated polymorphisms were not associated with BFT (Fan et al., 2008; Ramos et al., 2008). On SSC17, only a missense mutation in the cathepsin Z (*CTSZ*) gene was shown to affect fat deposition in Italian Large White pigs, but not in Italian Duroc pigs (Fontanesi et al., submitted; Russo et al., 2008), and this polymorphism in other commercial populations was associated with meat colour and growth rate (Ramos et al., 2009). Therefore the effects of some SSC17 markers could depend on the analysed population, suggesting they might be in different phases with the causative mutations. However, *TRIB3* polymorphism might affect BFT as evidenced in Italian heavy pigs analysing two different breeds and considering two different experimental designs.

4. Conclusions

We have identified a missense mutation in the porcine *TRIB3* gene that *in silico* analyses indicated to probably exert a functional effect on the biological role of the TRIB3 protein. This protein regulates insulin signalling at the level of Akt-2, a key modulator of insulin action in target cells, with important effects on energy metabolism and, via the regulatory action on adipogenenetic factors, on fat deposition. Gene expression analysis indicated *TRIB3* to be transcribed in tissues relevant for phenotypic traits whose variability could be explained, at least in part, by the polymorphisms in the *TRIB3* gene. In particular, this gene is expressed in adipose

and skeletal muscle tissues that are directly involved in determining production traits (BFT and lactate content) that resulted associated with the *TRIB3* genotypes. In addition, the effects on BFT seems consistent across breeds, so making *TRIB3* a particularly interesting marker for application in marker assisted selection programs in Italian heavy pig populations. Based on these evidences, it is quite tempting to speculate that the *TRIB3* p.P49L substitution could be involved in determining the evidenced effects. However, additional studies are needed to confirm this hypothesis.

Acknowledgments

We thank ANAS for providing samples and data. This work was supported by University of Bologna FAGenomicH and Italian MiPAAF SELMOL funds.

References

Andreozzi, F., Formoso, G., Prudente, S., Hribal, M.L., Pandolci, A., Bellacchio, E., Di Silvestre, S., Trischitta, V., Consoli, A., Sesti, G. (2008). TRIB3 R84 variant is associated with impaired insulin-mediated nitric oxide production in human endothelial cells. *Arteriosclerosis Thrombosis and Vascular Biology, 28,* 1355-1360.

Brunham, L.R., Singaraja, R.R., Pape, T.D., Kejariwal, A., Thomas, P.D., & Hayden, M.R. (2005). Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the *ABCA1* gene. *PLoS Genetics, 1,* e83.

Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., et al. (2007). Variation in *FTO* contributes to childhood obesity and severe adult obesity. *Nature Genetics, 39,* 724-726.

Du, K., Herzig, S., Kulkarni, R.N., & Montminy, M. (2003). TRB3: a tribbles homolog that inhibits Akt/PKB activation by insulin in liver. *Science, 300,* 1574-1577.

Fan, B., Du ,Z.Q., & Rothschild, M.F. (2009). The fat mass and obesity-associated (*FTO*) gene is associated with intramuscular fat content and growth rate in the pig. *Animal Biotechnology, 20,* 58-70.

Fan, B., Glenn, K.L., Geiger, B., Mileham, A., & Rothschild, M.F. (2008). Investigation of QTL regions on Chromosome 17 for genes associated with meat color in the pig. *Journal of Animal Breeding and Genetics, 125,* 240-247.

Fan, B., Lkhagvadorj, S., Cai, W., Young, J., Smith, R.M., Dekkers, J.C.M., Huff-Lonergan, E., Lonergan, S.M., & Rothschild, M.F. (2010). Identification of genetic markers associated with residual feed intake and meat quality traits in the pig. *Meat Science, 84,* 645-650.

Fontanesi, L., Colombo, M., Beretti, F., & Russo, V. (2008a). Evaluation of *post mortem* stability of porcine skeletal muscle RNA. *Meat Science, 80,* 1345-1351.

Fontanesi, L., Davoli, R., Nanni Costa, L., Beretti, F., Scotti, E., Tazzoli, M., Tassone, F., Colombo, M., Buttazzoni, L., & Russo, V. (2008b). Investigation of candidate genes for glycolytic potential of porcine skeletal muscle: association with meat quality and production traits in Italian Large White pigs. *Meat Science, 80,* 780-787.

Fontanesi, L., Scotti, E., Buttazzoni, L., Davoli, R., & Russo, V. (2009). The porcine fat mass and obesity associated (*FTO*) gene is associated with fat deposition in Italian Duroc pigs. *Animal Genetics, 40,* 90-93.

Fontanesi L, Scotti E, Buttazzoni L, Dall'Olio S, Bagnato, A., Lo Fiego, D.P., Davoli, R., & Russo, V. (2010). Confirmed association between a single nucleotide polymorphism in the *FTO* gene and obesity-related traits in heavy pigs. *Molecular Biology Reports, 37,* 461-466.

Frayling, T.M., Timpson, N.J,, Weedon, M.N, Zeggini, E., Freathy, R.M,, Lindgren, C.M., et al. (2007). A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science, 316,* 889-894.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V. K., Weiler, J. E., O'Brien, P. J., & MacLennan, D. H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science, 253,* 448-451.

Gong, H.P., Wang, Z.H., Jiang, H., Fang, N.N., Li, J.S., Shang, Y.Y., Zhang, Y., Zhong, M.,& Zhang, W. (2009). TRIB3 functional Q84R polymorphism is a risk factor for metabolic syndrome and carotid atherosclerosis. *Diabetes Care, 32,* 1311-1333.

Hart, E.A., Caccamo, M., Harrow, J.L., Humphray, S.J., Gilbert, J.G., Trevanion, S., Hubbard, T., Rogers, J., & Rothschild, M.F. (2007). Lessons learned from the initial sequencing of the pig genome: comparative analysis of an 8 Mb region of pig chromosome 17. *Genome Biology, 8,* R168.

Lahbib-Mansais, Y., Karlskov-Mortensen, P., Mompart, F., Milan, D., Jørgensen, C.B., Cirera, S., Gorodkin, J., Faraut. T., Yerle, M., & Fredholm, M. (2005). A high-resolution comparative map between pig chromosome 17 and human chromosomes 4, 8, and 20: identification of synteny breakpoints. *Genomics, 86,* 405-413.

Malek, M., Dekkers, J.C., Lee, H.K., Baas, T.J., Prusa, K., Huff-Lonergan, E., & Rothschild, M.F. (2001). A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. *Mammalian Genome, 12,* 637-645.

Monin, G., & Sellier P. (1985). Pork of low technological quality with a normal rate of muscle pH fall in the immediate post-mortem period: the case of the Hampshire breed. *Meat Science, 13,* 49-63.

Monin, G., Mejenes-Quijano, A., Talmant, A., & Sellier, P. (1987). Influence of breed and muscle metabolic type on muscle glycolytic potential and meat pH in pigs. *Meat Science, 20,* 149-158.

Nanni Costa, L., Tassone F., Davoli, R., Fontanesi, L., Dall'Olio, S., Colombo, M., & Russo, V. (2009). Glycolytic potential in *semimembranosus* muscle of Italian Large White pigs. *Journal of Muscle Food, 20,* 392-400.

Ng, P.C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research, 31,* 3812-3814.

Pierzchala, M., Cieslak, D., Reiner, G., Bartenschlager, H., Moser, G, & Geldermann, H. (2003). Linkage and QTL mapping for Sus scrofa chromosome 17. *Journal of Animal Breeding and Genetics, 120 (Suppl. 1),* 132-137.

Prudente, S., Hribal, M.L., Flex, E., Turchi, F., Morini, E., De Cosmo, S., et al. (2005). The functional Q84R polymorphism of mammalian Tribbles homolog TRB3 is associated with insulin resistance and related cardiovascular risk in Caucasians from Italy. *Diabetes, 54,* 2807-2811.

Prudente, S., Scarpelli, D., Chandalia, M., Zhang, Y.Y., Morini, E., Del Guerra, S., et al. (2009). The TRIB3 Q84R polymorphism and risk of early-onset type 2 diabetes. *Journal of Clinical Endocrinology and Metabolism, 94,* 190-196.

Ramos, A.M., Bastiaansen, J.W., Plastow, G.S., & Rothschild, M.F. (2009). Genes located on a SSC17 meat quality QTL region are associated with growth in outbred pig populations. *Animal Genetics, 40,* 774-778.

Ramos, A.M., Glenn, K.L., Serenius, T.V., Stalder, K.J., & Rothschild, M.F. (2008). Genetic markers for the production of US country hams. *Journal of Animal Breeding and Genetics, 125,* :248-527.

Rohrer, G. A., Thallman, R. M., Shackelford, S., Wheeler, T., & Koohmaraie, M. (2006). A genome scan for loci affecting pork quality in a Duroc-Landrace F2 population. *Animal Genetics, 37*, 17-27.

Russo, V., Davoli, R., Tagliavini, J., Dall'Olio, S., Bigi, D., Costosi, E., & Fontanesi, L. (1993). Identificazione del genotipo dei suini per la sensibilità all'alotano a livello di DNA mediante PCR. *Zootecnia e Nutrizione Animale, 19,* 89-93.

Russo, V., Fontanesi, L., Scotti, E., Beretti, F., Davoli, R., Nanni Costa, L., Virgili, R., & Buttazzoni, L. (2008). Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs. *Journal of Animal Science, 86*, 3300-3314.

Shi, Z., Liu, J., Guo, Q., Ma, X., Shen, L., Xu, S., Gao, H., Yuan, X., & Zhang, J. (2009). Association of TRB3 gene Q84R polymorphism with type 2 diabetes mellitus in Chinese population. *Endocrine, 35,* 414-419.

Takahashi, Y., Ohoka, N., Hayashi, H., & Sato, R. (2008). TRB3 suppresses adipocyte differentiation by negatively regulating PPARγ transcriptional activity. *Journal of Lipid Research, 49,* 880-892.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., & Narechania, A. (2003). PANTHER: A library of protein families and subfamilies indexed by function. *Genome Research, 13,* 2129-2141.

Thomas, P.D., & Kejariwal, A. (2004). Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proceedings of the National Academy of Sciences of the USA, 101,* 15398-15403.

Thomas, P.D., Kejariwal, A., Guo, N., Mi, H., Campbell, M.J., Muruganujan, A., & Lazareva-Ulitsky, B. (2006). Applications for protein sequence-function evolution data:

mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research, 34*, W645-W650.

Table 1. Haplotype frequencies at the *TRIB3* gene.

Breeds	No. of pigs	Haplotype frequencies[b]	
		[C:T]	[T:C]
Italian Large White	266 [a]	0.457	0.543
Italian Duroc	207 [a]	0.725	0.275
Italian Landrace	29	0.845	0.155
Belgian Landrace	14	0.321	0.679
Hampshire	10	1.000	0.000
Pietrain	9	0.944	0.056
Meishan	12	0.708	0.292

[a] The animals used for allele frequency evaluation in the Italian Large White and Italian Duroc breeds are those of the random groups of pigs used in the association study.

[b]Haplotypes are considered according to the genotyping results of the c.132T>C SNP. Haplotypes are defined with the two linked SNPs: c.132T>C and c.146C>T.

Table 2. Association analysis between the *TRIB3* diplotypes (haplotype combinations) estimated breeding values (EBVs) and phenotypic measures for meat quality parameters in Italian Large White and Italian Duroc pigs. Least square means are reported with their standard errors in parenthesis.

Breeds	Traits[a]	Haplotype combinations[b]			*P*[c]
		[C:T]/[C:T]	[C:T]/[T:C]	[T:C]/[T:C]	
Italian Large White		No. = 56	No. = 131	No. = 79	
	ADG (g)	32.786 (3.576)	33.298 (2.338)	33.772 (3.010)	0.978
	BFT (mm)	-3.045 (0.505)	-1.982 (0.330)	-1.537 (0.425)	0.070
	FGR	-0.153 (0.021)	-0.152 (0.013)	-0.146 (0.017)	0.936
	HW (kg)	0.544 (0.081)	0.562 (0.053)	0.646 (0.069)	0.543
	LC (kg)	2.199 (0.248)	1.981 (0.162)	1.881 (0.209)	0.612
	pH_1	5.948 (0.033)	5.921 (0.023)	5.922 (0.030)	0.761
	pH_2	5.652 (0.028)	5.651 (0.020)	5.672 (0.026)	0.761
	lactate	53.269 (2.063)	58.867 (1.387)	56.121 (1.821)	0.067
	glycogen	47.784 (3.127)	47.871 (2.253)	51.370 (2.848)	0.528
	GP	100.550 (3.142)	106.710 (2.246)	107.870 (2.854)	0.145
Italian Duroc		No. = 106	No. = 88	No. = 13	
	ADG (g)	34.897 (2.894)	30.534 (3.191)	19.616 (8.303)	0.182
	BFT (mm)	-2.715 (0.356)	-2.102 (0.393)	-0.500 (1.022)	0.098
	FGR	-0.1813 (0.015)	-0.164 (0.017)	-0.102 (0.044)	0.211
	LC (kg)	2.376 (0.199)	1.989 (0.219)	1.219 (0.571)	0.109

[a]ADG = average daily gain EBV; LC = lean cuts EBV; BFT = back fat thickness EBV; HW = ham weight EBV; FGR = feed:gain ratio EBV; GP = glycolytic potential. Lactate, glycogen and GP are expressed as μmol of lactic acid equivalent per g of fresh muscle.
[b]Haplotypes are considered according to the genotyping results of the c.132T>C SNP. Haplotypes are defined with the two linked SNPs: c.132T>C and c.146C>T.
[c]$P<0.10$ are underlined.

Table 3. Additive and dominance effects (with standard errors in parenthesis) obtained for the *TRIB3* marker. Results are reported only for association analyses with $P<0.20$.

Breeds	Traits[a]	Additive	P[b]	Dominance	P[b]
Italian Large White	BFT (mm)	0.754 (0.330)	<u>0.023</u>	0.308 (0.467)	0.510
	lactate	1.4260 (1.360)	0.296	4.172 (1.915)	<u>0.031</u>
	GP	3.658 (2.016)	<u>0.071</u>	2.501 (2.756)	0.365
Italian Duroc	ADG (g)	7.641 (4.396)	<u>0.084</u>	3.278 (5.432)	0.547
	BFT (mm)	-1.107 (0.541)	<u>0.042</u>	-0.495 (0.669)	0.460
	LC (kg)	0.577 (0.302)	<u>0.057</u>	0.192 (0.373)	0.608

[a]ADG = average daily gain EBV; LC = lean cuts EBV; BFT = back fat thickness EBV; GP = glycolytic potential. Lactate and GP are expressed as μmol of lactic acid equivalent per g of fresh muscle.
[b]$P<0.10$ are underlined.

Table 4. Haplotype frequencies and probability from Fisher's two-tailed exact test (*P*) of equal frequency in the negative (Ne) *vs* positive (Po) groups of pigs selected according to the extreme and divergent estimated breeding values (EBVs) for back fat thickness (BFT) and lean cuts (LC).

Traits [a]	Groups	No. of pigs [f]	Haplotype [C:T]	Haplotype [T:C]	*P* [g]
BFT	Ne [b]	50	0.610	0.390	<u>0.016</u>
	Po [b]	50	0.430	0.570	
	u Ne[c]	36	0.653	0.347	<u>0.007</u>
	u Po[c]	35	0.414	0.586	
LC	Ne [d]	50	0.470	0.530	0.252
	Po [d]	50	0.380	0.620	
	u Ne [e]	42	0.464	0.536	0.616
	u Po [e]	31	0.525	0.475	

[a] BFT = back fat thickness; LC = lean cuts.

[b] Animals with negative EBV (Ne) and positive (Po) EBV of the groups of pigs selected according to the extreme and divergent EBV for BFT.

[c] Unrelated (u) pigs at the second generation with negative EBV (Ne) and positive (Po) EBV of the groups of pigs selected according to the extreme and divergent EBV for BFT. These animals were among the 50 + 50 animals included in the rows above.

[d] Animals with negative EBV (Ne) and positive (Po) EBV of the groups of pigs selected according to the extreme and divergent EBV for LC.

[e] Unrelated (u) pigs at the second generation with negative EBV (Ne) and positive (Po) EBV of the groups of pigs selected according to the extreme and divergent EBV for LC. These animals were among the 50 + 50 animals included in the rows above.

[f] Number of pigs of each group.

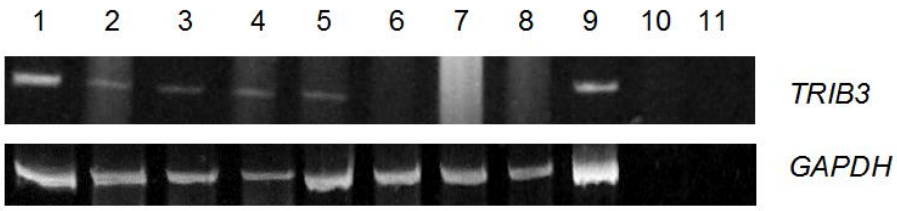
[g] Significant results are underlined. The tests compare the allele frequencies of the groups adjacent to the *P* value.

Figure 1. Alignment of the TRIB3 protein in different species around the p.P49L substitution.

Protein sequences of other species available in databases were: *Bos taurus*, NP_001069571; *Equus caballus*, XP_001499331; *Homo sapiens*, NP_066981; *Macaca mulatta*, XP_001111877; *Canis familiaris*, XP_542943; *Mus musculus*, NP_780302. Dots indicate the presence of the same amino acid of the first *Sus scrofa* sequence. Numbers indicate amino acid positions.

```
                    3        4        5
                    9        9        9
Sus scrofa p.49P    QPRPTPCPLPLSPPPAPTH
Sus scrofa p.49L    .........L.........
Bos taurus          ...LPS...T.N.....V.
Equus caballus      ...LP....A.......VR
Homo sapiens        ...LP..L.....T..DR
Macaca mulatta      ...LP..L.....T..DR
Canis familiaris    .SKLP...P.......AR
Mus musculus        E.G.L.SL..P.....SDL
```

Figure 2. Expression of the *TRIB3* gene in different tissues. *GAPDH* transcript amplification was included as control. 1 = back fat; 2 = brain; 3 = spleen; 4 = liver; 5 = heart; 6 = kidney; 7 = thyroid; 8 = lung; 9 = skeletal muscle; 10 = control genomic DNA; 11 = negative control.



CHAPTER TWO

Polymorphisms in an obesity related gene (*PCSK1*) are associated with fat deposition and production traits in Italian heavy pigs

L. Fontanesi[1], F. Bertolini[1], E. Scotti[1], P. Trevisi[1], L. Buttazzoni[2], S. Dall'Olio[1], R. Davoli[1], P. Bosi[1] and V. Russo[1]

[1]DIPROVAL, Sezione di Allevamenti Zootecnici, Faculty of Agriculture, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy

[2]CRA, Centro di Ricerca per la Produzione delle Carni e il Miglioramento Genetico, Via Salaria 31, 00015 Monterotondo Scalo, Roma, Italy

Submitted to ***Animal***, accepted for publication

Abstract

The proprotein convertase subtilisin/kexin type 1 (*PCSK1*) gene encodes the prohormone convertase 1/3 enzyme that processes prohormones into functional hormones that, in turn, regulate central and peripheral energy metabolism. Rare and common mutations in the human *PCSK1* gene cause severe monogenic or confer risk of obesity. We herein investigated the porcine *PCSK1* gene with the aim to identify polymorphisms associated with fat deposition and production traits in Italian heavy pigs. By resequencing about 5.1 kb of this gene in 21 pigs of different breeds we discovered 14 polymorphisms that were organized in 9 haplotypes, clearly distributed in two clades of putative European and Asian origin. Then we re-mapped this gene on porcine chromosome 2 and analysed its expression in several tissues including gastric oxyntic mucosa of weanling pigs in which PCSK1 processes the pre-pro-ghrelin into ghrelin, that in turn is involved in the control of feed intake and energy metabolism. Association analyses between *PCSK1* single nucleotide polymorphisms (SNPs) and production, carcass and several other traits were conducted on five groups of pigs from three different experimental designs, for a total of 1221 animals. Results indicated that the analysed SNPs were associated ($P<0.01$ or $P<0.05$) with several traits including back fat thickness and visible intermuscular fat in Italian Duroc (ID) and growth performances in Italian Large White (ILW) and in ILW x Italian Landrace pigs. However, the effects estimated in the ILW were opposite to the effects reported in the ID pigs. Suggestive association ($P<0.10$) was observed with muscle cathepsin B activity, opening, if confirmed, potential applications to reduce the excessive softness defect of the green hams that is of particular concern for the processing industry. The results obtained supported the need to further investigate the *PCSK1* gene to fully exploit the value of its variability and apply this information in pig breeding programs.

Running title: *PCSK1* and association with pig production traits

Keywords *association study; heavy pigs*; PCSK1*; production traits; SNP*

Implications

Polymorphisms in the porcine *PCSK1* gene are associated with several economic traits including obesity related traits, like back fat thickness and visible intermuscular fat, and cathepsin B activity. Exploitation of *PCSK1* variants in pig breeding programs is promising, providing that breed specific effects are considered and further investigated.

Introduction

Fat deposition is a key issue in pig breeding as it affects production efficiency, consumer acceptance of pork, pork quality and suitability of pork for processing (Russo and Nanni Costa, 1995). A large number of studies have reported QTL for different fat deposition traits in pigs (Hu and Reecy, 2007; Rothschild *et al.*, 2007). Furthermore, analyses of candidate genes for fat deposition derived from studies in humans and mice have allowed the identification of several gene polymorphisms associated with fatness in different pig lines and breeds. Mutations in several human genes have shown important roles in determining or conferring risks of obesity and obesity-related traits (like diabetes, cardiovascular diseases, etc.). Polymorphisms in the homologous pig genes have been associated with fatness and several other production and carcass traits. For example, among these genes it is worth to mention a missense mutation in the melanocortin 4 receptor (*MC4R*) gene that is associated with feed intake and backfat thickness in several pig lines/breeds (Kim *et al.*, 2000; Houston *et al.*, 2004; Bruun *et al.*, 2006; Fan *et al.*, 2009c). More recently, polymorphisms in the *FTO* (Fan *et al.*, 2009a; Fontanesi *et al.*, 2009; Fontanesi *et al.*, 2010b), *CTSK* (Fontanesi *et al.*, 2010c), *TBC1D1* (Fontanesi *et al.*, 2011a),

TCF7L2 (Du *et al*., 2009) and *TRIB3* (Fontanesi *et al.*, 2010a) genes among several others (Fan *et al.*, 2009b; Switonski *et al*., 2010) are associated with fat deposition traits in different pig breeds.

Another interesting candidate gene is the proprotein convertase subtilisin/kexin type 1 (*PCSK1*) gene. This gene encodes the prohormone convertase 1/3 enzyme that belongs to the subtilisin-like proprotein convertase family. *PCSK1* is mainly expressed in neuroendocrine cells and its main function is to process prohormones into functional hormones that, in turn, regulate central and peripheral energy metabolism. In particular, this enzyme cleaves proinsulin, proopiomelanocortin, prorenin, proenkephalin, prodynorphin, prosomatostatin and progastrin generating their corresponding bioactive peptides (Steiner *et al.*, 1996; Muller and Lindberg, 1999; Zhu *et al*., 2002; Scamuffa *et al*., 2006). In addition, PCSK1 is required for processing the pre-pro-ghrelin into ghrelin in the gastric mucosa (Zhu *et al*., 2006). Ghrelin is involved in the control of feed intake and energy metabolism. This polypeptide is secreted from the stomach and circulates in the bloodstream under fasting conditions, transmitting a hunger signal from the periphery to the central nervous system (Koijma and Kangawa, 2005). In pigs, forced weaning practises of most modern production systems cause a period of voluntary feed deprivation and weight loss (Forbes, 1995) that represents an interesting model to evaluate the endocrine adaptation to the food deprivation response that involves the PCSK1-ghrelin control and regulation (Salfen *et al.*, 2003).

As a proof of the important role of *PCSK1* for the endocrine system, rare mutations in the human *PCSK1* gene cause severe monogenic obesity (Jackson *et al.*, 1997; Farooqi *et al*., 2007). Common nonsynonymous variants in this gene confer risk of obesity in different human cohorts (Benzinou *et al*., 2008; Kilpeläinen *et al*., 2009; Chang *et al*., 2010; Heni *et al*., 2010; Qi *et al*., 2010). Similar effects have been observed in mice in which a missense mutation leads to obesity, abnormal proinsulin processing and multiple endocrinological defects (Lloyd *et al*., 2006).

On the basis of these lines of evidence reported in humans and mice, we selected the porcine *PCSK1* gene as a candidate gene for fat deposition and production traits in pigs. We re-mapped and re-sequenced this gene, identified polymorphisms, elaborated sequence information, analysed gene expression and showed that the porcine *PCSK1* gene is associated with fat deposition and several other production traits in different Italian heavy pig breeds/lines

Materials and methods

Animals

All procedures involving animals followed Italian and European Union regulations for animal care and slaughtering or, if requested by the Italian legislation, were approved by the Ethical Commission of the University of Bologna for animal experiments.

The association study was conducted on five groups (1-5) of pigs from three different experimental approaches (a, b, and c), as defined and listed below, for a total of 1221 animals.

The first four groups (1-4) included performance tested pigs grown at the Test Station of the National Pig Breeder Association (ANAS). Performance test was conducted on triplets of siblings from the same litter (two females and one castrated male). Data were used for the genetic evaluation of a male from the same litter (sib-testing).

Experimental design “a”. Of these four groups of animals, two were not selected by any phenotypic or genotypic criteria (random groups).

1-a) One of them was made up of 271 Italian Large White pigs (180 females and 91 castrated males, from 78 different sires; Fontanesi *et al.*, 2008b). Pigs of this group were slaughtered over 6 different days in 2002.

2-a) The other random group was made up by 197 Italian Duroc pigs (128 females and 69 castrated males, from 91 different sires) slaughtered at 33 different days in the years 1995-2003.

Experimental design “b”. Two other groups came from the most extreme and divergent estimated breeding values (EBVs) for: 3-b) back fat thickness (BFT; Italian Large White); and 4-b) visible intermuscular fat (VIF; Italian Duroc) among performance tested pigs.

3-b) Italian Large White pigs were all females (no. = 560) with no common parents, selected among ~12,000 performance tested pigs of this breed in the period 1995-2007 (280 with the lowest BFT EBV, mean and s.d = –9.40 ± 1.60 mm; and 280 with the highest BFT EBV, mean and s.d =+8.00 ± 5.95 mm).

4-b) The group of Italian Duroc consisted of 100 pigs (58 females and 42 castrated males from 62 different sires) selected among 1225 pigs of that breed (evaluated in same period 1996-1999; Fontanesi *et al.,* 2009, 2011b): 50 with the most negative (-2.35 ± 0.27) and 50 with the most positive VIF EBV (+2.17 ± 0.34). Details about traits and method of EBV calculation are reported below.

Experimental design “c”. 5-c) The fifth group was made up by 93 weaning pigs (Italian Large White x Italian Landrace; 22 females and 71 castrated males) for which growth performances and feed intake were recorded during a trial of tryptophan supplementation in the diet with or without oral challenge with *Escherichia coli* K88 (Bosi *et al.*, 2004; Trevisi *et al.*, 2009; Trevisi *et al.*, 2010). This latter group of pigs was also used for quantitative *PCSK1* gene expression analysis in the gastric oxyntic mucosa, considering the particular physiological state in this stressing period in which animals are voluntary feed deprived.

Another group of 21 unrelated pigs of different breeds for which no phenotypic data were available was used for sequence analysis (sequencing panel; see below). Additional pigs of different breeds (20 Italian Landrace, 22 Belgian Landrace, 18 Hampshire, 30 Casertana, and 12 Meishan) were used for allele frequency analysis of three tag SNPs. Phenotypic data were not

available for these pigs. A castrated male pig of ~160 kg live weight was used for collection of different tissues for qualitative *PCSK1* gene expression analysis.

Genomic DNA was extracted from muscle, blood or hair roots using standard protocols.

Traits

Performance testing started when pigs were 30-45 days old and ended when animals reached 155 ± 5 kg live weight. The nutritive level was *quasi ad libitum*. Feed intake was recorded daily and body weight was measured bimonthly, then daily gain and feed:gain ratio were calculated. At the end of test, selected animals from two contiguous batches on trial were mixed at loading and transported to a commercial slaughterhouse located at 24.5 km away from the Test-Station. After unloading, pigs were immediately stunned and bled in a lying position. Within 3 hours *post mortem* at the slaughterhouse, back fat thickness at the level of *Musculus gluteus medius*, weight of lean cuts (necks and loins), and weight of hams were measured. Only for the Italian Duroc pigs visual evaluation of intermuscular fat content (VIF) was obtained on leg muscles based on a binary scale (presence/absence). These phenotypic traits were available for performance tested pigs from the random groups and for the Italian Duroc used in the selective genotyping experiment. Only for the 271 Italian Large White pigs of the random group, measures of pH_1 (at 2 h *post mortem*), pH_u (at 24 h *post mortem*), glycolytic potential (GP), including glycogen and lactate content (30 min *post mortem*) and cathepsin B activity (Catb) (24 h *post mortem*) were obtained on *Musculus semimembranosus* as previously described (Fontanesi *et al*., 2008b; Russo *et al.*, 2008). For the Italian Large White of the selective genotyping experiment, only BFT measures were available.

Weaning Italian Large White x Italian Landrace pigs were raised in single boxes at the experimental farm of the University of Bologna. Body weight and feed intake were measured

daily for two weeks starting at weaning (21-24 days) and the values were averaged for the first and second week, respectively (Trevisi *et al.*, 2009; Trevisi *et al.*, 2010).

Sequencing and identification of polymorphisms

When we started this study, no complete porcine *PCSK1* gene sequence was available in DNA databases. Therefore BLASTN queries (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with a porcine *PCSK1* cDNA sequence (GenBank accession number NM_214038) against porcine sequences (including trace records) deposited in DNA databases were used to retrieve pig *PCSK1* genomic sequences corresponding to all recognized exons (according to the comparative information available in human) and parts of intronic regions (data not shown). PCR primers were designed with PRIMER3 (http://fokker.wi.mit.edu/primer3/input.htm) in intronic regions (except for the last gene region) to amplify 14 fragments encompassing exonic sequences (Table S1) from 21 pigs of the sequencing panel (Table 1). PCR was carried out using a PTC-100 (MJ Research, Watertown, MA, USA) thermal cycler in a final volume of 20 µL that included 10 pmol of each primer 2.0 mM $MgCl_2$, 2.5 mM each dNTP, 1 U AmpliBioTherm Taq DNA polymerase (Fisher Molecular Biology, Trevose, PA, USA). The PCR profile was the following: an initial step at 95 °C for 5 min; 35 cycles of 30 s at 95 °C, 30 s at the appropriate annealing temperature (Table S1) and 30 s at 72 °C; the final extension step was for 5 min at 72 °C. Sequencing reactions were produced for ExoSAP-IT® (USB Corporation, Cleveland, Ohio, USA)-treated PCR products using the same PCR primers and the Big Dye v3.1 kit (Applied Biosystems, Foster City, CA, USA). Sequencing products were electrophoresed on an ABI3100 Avant sequencer (Applied Biosystems). Using CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA) all sequence chromatograms were aligned and visually inspected to identify polymorphisms.

Sequence analyses

Phases including all detected polymorphisms in the sequenced panel were reconstructed with PHASE program v. 2.1.1 (Stephens *et al.*, 2001). Nucleotide diversity (π), haplotype diversity and Tajima's *D* were estimated using DNASP v. 4.10.9 (Rozas *et al.*, 2003). Tajima's *D* measures whether the observed frequencies of segregating mutations are compatible with the frequencies expected under a standard neutral model. Positive selection or the presence of weakly deleterious mutations as well as population growth tend to give an excess of low frequency variants, resulting in negative test values. Balancing selection or population contraction may cause an excess of intermediate-variants and positive *D* values. Significance of the *D* tests was determined by coalescent simulations (Rozas *et al.*, 2003). Phylogenetic analysis of the identified haplotypes of the *PCSK1* gene was conducted with MEGA4 software v. 4.0.2 (Tamura *et al.* 2007) with Kimura's two parameter model and Neighbour Joining (NJ) tree. Standard errors were obtained from 1,000 bootstrap replicates. The HAPLOVIEW program v.4.1 was used to determine the presence and length of haplotype blocks (using the four gamete rule option), and potential tag SNPs (Barrett *et al.*, 2005).

Genotyping of SNPs

Three *PCSK1* SNPs were analysed by PCR-RFLP. DNA extracted as reported above was amplified as previously described (Table S1). Digestion was carried out overnight at 37 or 65 °C in a final volume of 25 µl containing 1X enzyme reaction buffer, 5 µl of PCR product and either 2 U of *Alu*I (MBI Fermentas, Vilnius, Lithuania) for SNP g.5182A>T or 2 U of *Tai*I (MBI Fermentas) for SNP g.1736C>T or either 2 U of *Taa*I (MBI Fermentas) for SNP g.1966C>A (reference sequence GenBank/EMBL HE599222). Digested DNA fragments were electrophoresed on 1X TBE 10% polyacrylamide/bis-acrylamide 29:1 gels and stained with

ethidium bromide or 1X GelRed Nucleid Acid Gel Stain (Biotium Inc., Hayward, CA, USA). To evaluate precision of the three genotyping protocols, sequenced animals were genotyped. Both sequencing and PCR-RFLP analyses showed the same results.

Gene expression

Qualitative gene expression analysis: Isolation of total RNA from back fat, brain, heart, kidney, liver, lung, skeletal muscle, spleen, thyroid, and gastric oxyntic mucosa (collected from a weaning pig) specimens (50-100 mg) was carried out using the RNeasy® Lipid Tissue kit (Qiagen, Duesseldorf, Germany) (for backfat and brain), the RNeasy® Midi kit (Qiagen) (for gastric oxyntic mucosa) or the FastPure™ RNA kit (TaKaRa Bio Inc., Shiga, Japan) (for all other tissues), following the manufacturers' instructions. The purity and concentration of the total RNA extracted were checked using the Nanodrop ND 1000 spectrophotometer (Nanodrop Technologies Inc., Wilmington, DE, USA) and RNA integrity was controlled by agarose gel electrophoresis analysis. Then, about 1 µg of total RNA for each tissue was treated by RNase-Free DNase set (Qiagen) and 40 ng were retrotranscribed with Improm II Reverse Transcription system (Promega Corporation, Madison, WI) using oligo(dT) primers and following the manufacturer's protocol. A fragment of 122 bp of the *PCSK1* cDNA was amplified with the primers designed on the 3'-untranslated region (3'-UTR) of the *PCSK1* cDNA (GenBank accession number NM_214038; Table S1) using the same PCR cycling conditions as reported above. *GAPDH* cDNA amplification, used as reference, was obtained with primer pair GAPDH_437 reported in Fontanesi *et al.* (2008a). *PCSK1* and *GAPDH* amplified cDNA fragments were electrophoresed as reported above. All analyses were repeated three times.

Quantitative gene expression analysis (qPCR): This analysis was carried out on gastric oxyntic mucosa RNA extracted as described above from specimens collected from the weaning pigs. Total RNA was evaluated and RNase-Free DNase treated as reported above. After these

steps, 1 μg of total RNA was reverse transcribed using the ImProm-II Reverse Transcription System (Promega). External PCR primers were used to amplify a fragment which served as the external homologous DNA standard of a known copy number (Table S1). This product was purified using the QIAquick PCR Purification Kit (Qiagen). The DNA quality and concentration were evaluated by the Nanodrop ND 1000 spectrophotometer (Nanodrop Technologies Inc.) and agarose gel electrophoresis. Then the fragment was serially diluted in 1:10 steps and a standard curve was created to perform an absolute quantitative analysis using internal primers (Table S1). The quantification reactions were performed in a Light Cycler instrument (Roche, Mannheim, Germany). The amplification was carried out in a 10 μL volume containing 2 μL of cDNA, 8 pmol of each primer, and 5 μL of SYBR® Premix Ex Taq™ II (Perfect Real Time; Takara Bio Inc., Japan). The fast protocol was 40 cycles at 95°C for 5 s and 62 °C for 20 s. The detection of the fluorescent product was set at the last step of each cycle. The specificity of each amplification was determined by melting curve analysis and electrophoresis on 1.2% agarose gels stained as previously reported. All amplifications were repeated three times and data were expressed as gene transcript copies /μl cDNA.

Radiation hybrid mapping

The INRA-Minnesota 7000 rads radiation hybrid (RH) panel (IMpRH panel; Yerle *et al.*, 1998) consisting of 118 rodent-porcine hybrid cell lines was screened by means of PCR using primer pair indicated in Table S1. No PCR fragment was obtained from the control rodent genomic DNA. The PCR reactions were visualized on 10% polyacrylamide/bis-acrylamide 29:1 or 2% agarose gels. The results of RH PCR products were analysed with the IMpRH mapping tool accessible through the http://imprh.toulouse.inra.fr/ web address (Milan *et al.*, 2000).

Association analyses

EBVs for average daily gain (ADG, expressed in g), lean cuts (LC, expressed in kg), back fat thickness (BFT, expressed in mm) and feed:gain ratio (FGR) were calculated for both random groups of sib-tested Italian heavy pigs (1-a and 2-a). EBV for ham weight (expressed in kg, HW) was calculated for Italian Large White pigs only (1-a). EBV for visible intermuscular fat (VIF) was calculated for Italian Duroc pigs only (2-a). EBVs for BFT and VIF were calculated for the two groups of pigs (Italian Large White and Italian Duroc, respectively) used in the selective genotyping analysis (3-b and 4-b). EBVs were calculated using a BLUP-Multiple Trait-Animal Model with different models for each trait. Depending on the trait, models included the fixed effects of sex, batch on trial, inbreeding coefficient of the animal, interaction of sex by age at slaughtering, date of slaughtering and the random effects of litter and animal. Random Residuals (RRs) were calculated for all considered traits in the random groups (1-a and 2-a) of performance tested Italian Large White and in all Italian Duroc pigs (ADG, LC, BFT, FGR and HW in Italian Large White; ADG, LC, BFT, FGR, HW, and VIF in Italian Duroc). RRs were obtained by using liner fixed models including the same factors used for each trait in the Best Linear Unbiased Prediction (BLUP)-Multiple Trait Animal Model (Fontanesi *et al.*, 2010d).

Association analyses were carried out independently for the two random groups of sib-tested Italian pigs (1-a, Italian Large White; and 2-a, Italian Duroc). Associations between the two polymorphic *PCSK1* SNPs (g.1736C>T and g.5182A>T) and EBVs or RRs were assessed by using the general linear model (GLM) procedure of SAS, release 9.2 (SAS Institute Inc. Cary, NC, USA). The models included only the fixed effects of individual marker genotypes. All other factors contributing to variability of the investigated traits were already considered in the calculation of EBV or RR. For meat quality traits in the performance tested Italian Large White pigs (1-a), the procedure MIXED of SAS was applied to a model that included date of slaughtering, sex and *PCSK1* genotype for g.1736C>T and g.5182A>T SNPs. EBV and RR

distribution (and their residuals) for BFT, ADG, LC, and FGR traits in the sub-group of Italian Duroc animals pigs with the extreme values of EBVs for VIF (4-b), did not differ from the normal distribution (Shapiro-Wilk test for these traits was $P > 0.15$). Therefore, to test the association between the genotypes at the *PCSK1* gene and the calculated EBVs and RRs for BFT, ADG, LC and FGR, this sub-group of pigs (100 animals) was merged with the sub-group of pigs of the same breed not pre-selected by any phenotypic or genotypic criteria (197 animals), thus forming a group of 297 Italian Duroc pigs (2-a + 4-b). Association between VIF and *PCSK1* genotypes was carried out using only the 197 Italian Duroc pigs not pre-selected in which VIF EBVs were normally distributed (2-a). In addition, haplotypes between the two *PCSK1* genotyped SNPs were inferred using the PHASE program v. 2.0 (Stephens *et al.*, 2001). Evaluation of the haplotype substitution effect on EBVs and RRs was obtained separately for the same two performance tested Italian heavy pig groups (1-a and 2-a) using the PROC REG of SAS with a model including the number (0, 1, 2) of the haplotypes identified. As discussed in Russo *et al.* (2008), to overcome the effects of multiple tests on nominal comparison-wise error rate *P*-values, significant threshold (independently for the single marker analysis and for the haplotype analyses across the two random populations) was identified using the Proportion of False Positive (PFP) approach (Fernando *et al.*, 2004). PFP thresholds were calculated as described in Bagnato *et al.* (2008). As reported below, the excess of significant tests caused an almost coincidence of *P*-nominal value and P_{PFP} for single marker tests and a higher P_{PFP} than the *P*-nominal value in haplotype association analysis, similarly as reported by Tal-Stein *et al.* (2010). Therefore in both cases we adopted *P*-nominal value = 0.05 as the threshold for significant association and *P*-nominal value = 0.10 as the threshold to indicate suggestive association. In addition, additive genetic and dominance effects of the two *PCSK1* SNPs were estimated for the two random performance tested pig populations as it follows: additive effect, $a = 1/2(BB-AA)$; dominance effect, $d = AB-1/2(AA+BB)$. Estimates of the effects were tested by *t*-test for significant deviation from zero. Ratio between the absolute values of *d* and *a* ($|d/a|$) were

used to indicate possible gene effects: $|d/a|<0.2$, additive; $0.2<|d/a|<0.8$, partial dominance; $0.8<|d/a|<1.2$, dominance; $|d/a|>1.2$, overdominance (Stuber *et al.*, 1987).

For the two groups of pigs chosen for selective genotyping based on extreme values of EBVs for BFT (3-b, Italian Large White) and VIF (4-b, Italian Duroc), respectively, Fisher's exact test of significance (two-tailed) or Chi square test (where appropriate) of differences in allele frequency between the positive and negative groups were used.

For the weaning Italian Large White x Italian Landrace pigs (5-c), association between *PCSK1* gene expression in gastric oxyntic mucosa, growth performances and feed intake, and *PCSK1* SNPs was analyzed using the PROC GLM procedure of SAS with a factor design, including the *PCSK1* g.1736C>T or the g.5182A>T SNPs, the diet, the susceptibility to *E. coli* K88, the trial, and the litter within the trial. In the case of the g.5182A>T polymorphism, only two pigs carried the AA genotype, therefore this genotypic class was excluded from the analysis.

In these two later studies (extreme divergent groups and weaning pigs), results with *P*-nominal value ≤ 0.05 were considered significant. Other *P*-nominal values $0.05 < P < 0.10$ were considered as suggestive for associations.

Results

Gene expression analyses

As information of the range of tissues in which *PCSK1* is expressed was not available in pigs, we first tested its expression in 10 different tissues (Fig. S1). *PCSK1* expression was evident in most of the analysed tissues with some differences in signal intensity, except in liver and spleen in which we did not obtain any amplified fragment. As expected from reports in other species (i.e. Gagnon *et al.*, 2009) and according to the fundamental function of PCSK1 in

processing key-gut/intestine prohormones (Zhu *et al.*, 2006), gastric oxyntic mucosa showed a qualitatively higher level of *PCSK1*gene expression compared to the other tissues (Fig. S1). Therefore qPCR was used to precisely evaluate the level of expression of this gene in weaning pigs and to evaluate its association with *PCSK1* SNPs (see below). Considering the data obtained in the analysed pigs, gastric oxyntic mucosa contained on average 996 ± 65 *PCSK1* gene transcript copies per µl of cDNA.

Radiation hybrid mapping of the porcine PCSK1 gene

The porcine *PCSK1* gene has been already assigned to porcine chromosome (SSC) 2 by RH analysis (Shimogiri *et al.*, 2006). In addition, a partial sequence with several gaps and assembling problems of the *PCSK1* gene has been included in the Sscrofa 9.2 version of SSC2, between nucleotides 92,001,906 and 92,024,579 (Ensembl entry: ENSSSCG00000014169). This is in agreement with comparative mapping between SSC2 and human chromosome (HSA) 5 in which this gene is localized (GRCh37: position 95.73-95.77 Mb), as indicated by several studies (Rink *et al.*, 2006; Hamasima *et al.*, 2008). However, this gene has not been included in the Sscrofa10.2 version of SSC2 as BLAST analysis with the porcine *PCSK1* cDNA sequence (NM_214038.1) did not reveal any significant match with any region in this updated SSC2 version, but only with an un-assembled scaffold (LOCUS NW_003540493, chrU_scaffold4254; data not shown). Therefore we confirmed the previous assignments to SSC2 by RH mapping using the IMpRH panel. The retention fraction of the amplified *PCSK1* fragment was 22% and the closest marker obtained by two-point analysis was *S0226* (distance = 39 cR; LOD = 10.31), already placed on the RH map of SSC2 (Hawken *et al.*, 1999). This microsatellite is the same that resulted the closest marker in the two-point analysis reported by Shimogiri *et al.* (2006), confirming the previous evidences. Multipoint mapping localized the *PCSK1* gene between microsatellites

SW1602 and SW1320 (74.8 cM and 76.9 cM, respectively, in the linkage map of the USDA database, http://www.marc.usda.gov/genome/swine/swine.html).

Polymorphisms, sequence statistics and haplotype diversity

As the porcine *PCSK1*gene sequence available on Sscrofa9.2 did not include sequence information from a few coding exons and a few others have been assembled in reverse complement position compared to most of the correctly defined Ensembl exons, we retrieved the lacking exons and defined the problematic ones mining trace records and aligning cDNA and genomic sequences. The resulting assembled sequence (EMBL accession no. HE599222) was compared to the sequence available in Sscrofa10.2 genome version that independently confirmed the correct assembly. Then, we resequenced 5,180 bp in 21 pigs of different breeds, encompassing almost all coding sequence (CDS; 2,062 bp, only a few bp of the last coding exon were not included) and parts of intronic and 5'-untranslated region (3,118 bp) (EMBL accession number HE599222). On the whole, we identified one indel (in a poly T stretch in intron 4) and 13 SNPs, 4 of which were synonymous polymorphisms in coding regions (one SNP every ~515 bp of CDS), whereas all others polymorphisms were in introns (one every ~312 bp of non coding sequences) (Table 1). *In silico* prediction of the effects of the identified polymorphisms did not report any strong signal of potential functional roles for the different variants (data not shown).

Computation of nucleotide diversity (π) for sequenced regions (Table 2) indicated a quite low level of variability in the Western pig breeds (ranging from 0.011% to 0.040%) compared to the Meishan (0.090%). Tajima's *D* (D_T) values were not significant in any breed (P>0.10) probably due to the low number of sequenced animals. However, it is interesting to note a positive value in Italian Large White (+1.680), whereas in Italian Duroc and Hampshire this parameter was negative (-1.278 and -1.337) (Table 2).

Haplotypes were inferred using all SNPs but excluding the indel for the difficulties in determining the genotypes of the animals from sequences data. Inferred haplotypes (n. = 9) with their occurrence in the sequenced pig panel are reported as a note to Table 1. Haplotype (H) 3 was the most frequent (17 out of 42 sequences) and it was almost fixed in Italian Duroc (12/14). Haplotype 1 was the second most frequent in the sequenced panel (24%), it was in all Western breeds and was the most frequent in Hampshire (5/6). Haplotype 6 was the most frequent in Italian Large White (4/10) but was observed also in one Meishan pig. Other Meishan sequences included haplotypes 7 and 8. Haplotype 9 was identified in one Hampshire pig. Neighbor-Joining tree of the haplotypes (Fig. S2) clearly indicated differences between two groups of haplotypes. One included the most divergent haplotypes (H8 and H9) that might be of Asian origin, considering that H8 was identified in one Meishan. However, H9 identified in a Western pig (Hampshire) might be derived by introgression of Asian haplotypes into Western originated breeds. Signs of putative introgression of Asian haplotypes in European breeds could be seen also for H7 that was shared by Meishan, Italian Large White, Belgian Landrace, and Italian Duroc.

Figure 1 reports the linkage disequilibrium patterns (r^2) between polymorphism pairs across the porcine *PCSK1*gene. Two main haplotype blocks can be observed from the obtained data. The largest one, that includes the 3' part of the gene, spans ~30 kb (without including several unknown not sequenced intronic regions of the Sscrofa10.2 chrU_scaffold4254). Using the aggressive tagging algorithm of HAPLOVIEW (Barrett *et al.*, 2005), 9 SNPs tests captured all variability (13 considered polymorphisms; mean max r^2 = 1.00).

Three tag SNPs detected by sequencing (g.1696C>A, g.1737T>C, and g.5182A>T) were chosen for genotyping as they could be easily analysed by PCR-RFLP and were among the tagged SNPs identified as reported above. Allele frequencies of these three SNPs in a larger number of pigs of the same breeds used for SNP discovery, including also 30 Casertana pigs, are reported in Table S2. Results confirmed those obtained from the sequencing panel for the same

polymorphic sites. For example, allele C of the g.1737T>C SNP was contained only in H9 identified in a Hampshire pig. This SNP resulted polymorphic only in the Hampshire breed (Table S2). For this reason it was not considered for further genotyping and association studies.

Association analyses

Results of the association analyses in the random groups of Italian Large White and Italian Duroc pigs (1-a and 2-a) between the two *PCSK1* SNPs of interest (g.1696C>A and g.5182A>T) and meat production, carcass traits, growth performances and meat quality traits are reported in Table 3. An excess of tests with *P*-nominal value <0.10 is evident (33 out of 54 tests), therefore *P*-nominal value of 0.05 corresponded to the P_{PFP} significant threshold of 0.05. Both EBVs and RRs were considered in this analysis for ADG, BFT, FGR, HW, LC and VIF: for most traits for which association analyses with EBV produced *P*-nominal value <0.10, results obtained using RR confirmed in most case the same significance levels. In particular, for g.5182A>T in the Italian Duroc population, 4 out of 5 significant EBVs resulted in a corresponding RR below *P*-nominal value of 0.10. Also HW RR (not available as EBV) showed a *P*-nominal value far below 0.10 (Table 3). For the other genotyped SNP (g.1696C>A), 5 significantly or suggestively associated trait/EBVs had two corresponding suggestively associated RR based values (ADG and LC) plus HW RR with $P < 0.05$. In Italian Large White pigs, the g.1696C>A genotypes with significant effects on ADG EBV ($P < 0.01$), FGR EBV ($P < 0.05$) and LC EBV ($P < 0.02$), were also suggestively associated, close to the suggestive threshold or significantly associated with the corresponding RR values (ADG RR, $P < 010$; FGR RR, $P = 0.107$; LC RR, $P < 0.05$). This polymorphic site was also suggestively associated with lactate content (Table 3). Effect on LC was also reported for the other marker (g.5182A>T) in both EBV ($P < 0.01$) and RR ($P < 0.02$). The genotypes of this SNP were suggestively associated or significantly associated with several meat quality parameters (pH_1, lactate content and cathepsin B activity: P <0.10; glycogen

content: $P < 0.05$). In the Italian Large White population, BFT did not result to be significantly associated with g.1696C>A or g.5182A>T genotypes.

By observing the effects of the two genotyped SNPs on the significantly or suggestively associated traits in the two considered breeds it appears an opposite effect (with some difference in additive or dominance effects) in Italian Large White as compared to Italian Duroc (Tables 3 and S3). For example, genotype CC of the g.1696C>A showed a positive effect in the Italian Large White whereas it was the less favourable in Italian Duroc for ADG (Tables 3 and S3). And this was evident for most traits analysed in both breeds. The same was true for the g.5182A>T SNP. For example, genotype AA was associated with lower ADG in Italian Large White whereas in Italian Duroc TT was the less favourable genotype (Tables 3 and S3).

Results of single marker analysis were confirmed using haplotypes of these two genotyped SNPs (Table S4). All four possible haplotypes were inferred in both pig breeds. One of which ([g.1696A:g.5182T], indicated as [A:T]) was carried by few animals and was excluded from the association analyses (Table S4). The most frequent haplotype in Italian Large White was [A:A] (~46%), whereas in Italian Duroc it was [C:T] (~52%) (Table S4). Considering the 63 tests reported in Table S4, more than 50% (33 out of 63) of the test showed a $P < 0.10$. This excess of tests in this bin class produced a $P_{PFD} = 0.05$ that corresponded to a *P*-nominal value of ~0.10. Therefore, also for haplotype analyses we took the *P*-nominal value of 0.05 as the threshold for significance and the *P*-nominal value of 0.10 as the threshold for suggestive association. Haplotype substitution effects were highly significant for ADG EBV ($P < 0.001$) and significant for ADG RR ($P < 0.05$) for haplotype [A:A] in both breeds, but, as expected from the results of the analysis of the two separated SNPs, in the opposite direction (Table S4). For haplotype [C:T], substitution effect for 11 out of 21 trait/breed combinations were significant in the Italian Duroc pigs. Haplotype [C:A] showed significant substitution effect in Italian Duroc only (Table S4). Significant results were for BFT and VIF for both EBV and RR in opposite direction as

compared to the effects of haplotype [C:T] on the same traits (Table S4). This result confirmed the role of the g.5182A>T SNP in affecting BFT and VIF in Italian Duroc breed.

An independent confirmation of the effect of this *PCSK1* polymorphic site in affecting VIF was provided by the selective genotyping experiment in which two extreme and divergent groups for VIF EBV were genotyped (4-b; Table S5). Significant differences in allele frequencies between the two tails were reported considering all 50 + 50 pigs ($P < 0.05$). Suggestive differences ($P < 0.10$) were maintained considering only two-generation unrelated pigs (Table S5). Allele T was the most frequent in the positive group (VIF with higher value, that means higher content of VIF) confirming the effect of this allele from the previous analysis (Tables 3 and S4). As expected, no significant allele frequency difference was reported for the g.1696C>A SNP (Table S5). Results for the selective genotyping experiment for BFT in the Italian Large White breed (3-b) confirmed the lack of effects of the g.1696C>A and g.5182A>T SNPs on this trait (Table S5).

To further evaluate the effects of *PCSK1*, we genotyped the g.1696C>A and g.5182A>T SNP in 93 weaning Italian Large White x Italian Landrace pigs (5-c) for which growth performances, feed intake and mRNA *PCSK1* level in gastric oxyntic mucosa were measured (Table 4). Significant association ($P < 0.05$) between the g.1696C>A genotypes and daily feed intake during the second week of trial was identified confirming the suggestive association ($P < 0.10$) with daily gain on the same period. Animals with genotype AA showed increased feed intake and growth rate, confirming, to some extent, the effect of this SNP on growth traits, already found in the performance tested Italian Large White and Italian Duroc pigs (Table 3). However, the direction of the effect seems similar to that reported in Italian Duroc in which animals with the same AA genotype had a higher ADG. No significant effects were evidenced for the g.5182A>T SNP on growth performance. Both genotyped SNP were not associated with *PCSK1* gene expression level in the analysed tissue.

Discussion

PCSK1 also known as PC1 or PC3 belongs to a family of serine proteinases that is responsible for processing secretory precursor proteins into their active forms, making this enzyme a key regulator of the secretory pathway in mammals. Several reports in humans have indicated that polymorphisms in this genes cause severe early-onset obesity or are associated with common obesity (e.g.: Jackson *et al.*, 1997; Benzinou *et al.*, 2008). Therefore, this gene is an interesting candidate to explain a quote of variability of fat deposition and other obesity-related traits in pigs.

Mapping of the porcine *PCSK1* gene confirmed its position on SSC2. A few studies have reported QTL for growth rate at different stages in the chromosome region overlapping the positions of the microsatellites that bracket *PCSK1* using a Berkshire x Yorkshire population (Malek *et al.* 2001; Thomsen *et al.*, 2004) and a Wild Boar x Pietrain reference family (Lee *et al.*, 2003). To our knowledge, no fat deposition QTL with a peak in this region have been reported, even if 95% confidence intervals are usually very large. However, statistical analysis of SSC2 QTL seems quite complicated by the complex pattern of effects in different populations and by the large effects of the p-arm region (Fontanesi *et al.*, 2010d, 2011b). In addition, other reports have localized another important QTL region for back fat thickness and, possibly, lean cut content, around 30-50 cM (Lee *et al.*, 2003; Liu *et al.*, 2007; Tortereau *et al.*, 2011).

The results of the association analyses indicated that two polymorphic sites in intronic regions of the porcine *PCSK1* gene (g.1696C>A and g.5182A>T) are associated with several meat production, growth efficiency, carcass and meat quality traits in Italian Large White, Italian Duroc and in a commercial cross (Tables 3, 4, S3, S4 and S5). The findings obtained in the Italian Large White and Italian Duroc random groups (1-a and 2-a) for fat deposition traits have been confirmed by the results obtained in the selective genotyping experiment (3-b and 4-b) set up for VIF (Italian Duroc) and BFT (Italian Large White). The results obtained in the weaning Italian Large White x Italian Landrace pigs confirmed the effects of the most polymorphic marker

(g.1696C>A) in this population for growth performances observed also in the random groups of the two performance tested breeds.

PCSK1 is required for processing of pre-pro-ghrelin into ghrelin in the gastric mucosa (Zhu *et al.*, 2006), that, in turn, plays a key role in the control of feed intake and energy metabolism (Wren *et al.*, 2001). Thus it was tempting to assume an association between the analysed polymorphisms and the level of expression of the *PCSK1* gene in the key tissue for ghrelin production (gastric oxyntic mucosa). This was tested in weaning pigs (5-c) in which feed intake is in general limited by the weaning stress. However the analysed polymorphisms were not associated with *PCSK1* mRNA level in this tissue. The complex interacting factors related to the particular condition that the piglets are suffering after the chease of suckling could act masking differences in *PCSK1* gene expression, if any.

Interesting results have been obtained for a few meat quality traits. In particular the g.5182A>T SNP was associated with glycogen content, and suggestively associated with pH_1, lactate content and cathepsin B activity in Italian Large White pigs (1-a). As *PCSK1* is expressed in skeletal muscle, even if at moderate level according to our raw qualitative analysis (Fig. S1), and considering that the encoded enzyme has proteolytic functions, it could be possible that PCSK1 is involved directly in some biological processes determining variability on these traits. Quantitative analysis of *PCSK1* mRNA expression in this tissue and its relationship with *PCSK1* variability could provide additional information to clarify this question. Moreover, it is also worth to point out that the g.5182A>T SNP is the first marker reported to have a possible effect on cathepsin B activity of the muscle. This trait is important for the production of dry-cured hams as an excessive cathepsin B activity is associated with excessive softness and other defects of the meat that are of particular concerns for the processing industry (Russo *et al.*, 2000; Virgili and Schivazappa, 2002). If this result is confirmed, it could provide important applications in this

sector by using marker assisted selection based on this SNP or, eventually, other *PCSK1* polymorphisms.

In the association analysis carried out in the two random groups of the performance tested Italian Large White and Italian Duroc pigs (1-a and 2-a) we used both EBVs and RRs for traits that are currently included in the national genetic evaluation (http://www.anas.it) of these breeds. It is worth to mention that these traits have high heritability (> 0.30) and that have high genetic correlations (Ciobanu *et al.*, 2011; Clutter, 2011). Several studies, mainly in dairy cattle, have evaluated the properties of EBV in association studies suggesting a lower, or at best equivalent, power in using EBV as compared to raw or adjusted phenotypic measurements (e.g. Israel and Weller, 2002; Thomsen *et al.*, 2001). Other simulation studies have reported that using EBVs could produce a higher level of type I errors compared to other approaches (Ekine *et al.*, 2010), even if evaluation for the level of type II errors are not reported. In general, tests that used EBVs gave lower *P* values than those based on RRs (Tables 3 and S4), confirming, to some extent a possible overestimation of the effects and higher rate of false positives in these analyses. However, this possible bias seems reduced for traits with high heritability, but additional studies should be carried out to evaluate this issue. The combined use of RRs could, at least in part, prevent this problem and might indicate results that should be considered with caution (Fontanesi *et al.*, 2010d). The use of RRs produced 12 out of 20 tests with *P*-nominal value < 0.10 (60%) and the use of EBVs produced 17 out of 24 tests with *P*-nominal value below this threshold (71%), considering the single marker tests (Table 3). In the association analysis with haplotypes 20/30 and 13/33 tests for EBVs and RRs, respectively, had *P*-nominal value < 0.10. In single marker and haplotype analyses RR were at least suggestively significant only when EBV were below the 0.10 suggestive threshold, but the contrary was not true as it can be easily deduced from what indicated above. However, even if we consider the few differences between EBV and RR it is clear that the analysed markers are associated with several production traits in both

Italian Large White and Italian Duroc pigs. Effects within breeds were consistent with the genetic correlation among recorded traits (Ciobanu *et al.*, 2011; Clutter, 2011). However, comparing the results in these two breeds, the effects of both polymorphisms (and their haplotypes) was clearly in the opposite direction for all traits available in both groups (Tables 3, S3 and S4). This means that these *PCSK1* markers are in different phases with (a) close causative mutation(s) affecting the analysed traits. By re-sequencing more than 5.1 kb of the porcine *PCSK1* gene in 21 pigs of different breeds (for a total of about 1.09 Mb) we did not identify any putative causative mutation. Therefore it could be possible that not identified regulatory mutations in this gene or in other close genes could be involved in determining the observed phenotypic effects.

Another striking difference between Italian Duroc and Italian Large White comes from the opposite values obtained calculating D_T in the sequenced panel (Table 2). Even if not significant, the negative D_T value in the Italian Duroc might indicate a tendency towards directional selection whereas the positive D_T value in the Italian Large White might indicate a tendency towards balancing selection. Confirmation of these values in a larger sample of pigs is needed before any conclusion but, in any case, it is tempting to match these preliminary indications with the results of the association study in which *PCSK1* SNPs have opposite effects in the same two breeds. The explanation of one of these two genetic aspects could provide useful information to understand the other. For this reason it is also important to consider that haplotype heterogeneity at his locus has been only in part considered in our study. A larger number of haplotypes than those actually used in the association studies have been identified (Tables 1 and 2, Fig. S2), suggesting that we might have missed information that are worth of further exploitation.

Results obtained in this study will guide further investigations to solve the case of different effects in different breeds enlarging the analysis of haplotype blocks by identifying and adding additional polymorphisms in the *PCSK1* gene or in close regions and evaluating them in additional pig populations. Other functional studies could provide other data that may complement the results we already obtained.

Acknowledgements

We thank Dr. M. Yerle (INRA) for providing the INRA-Minnesota 7000 rads RH panel and ANAS for providing samples and data. This work was supported by University of Bologna (FAGenomicH project) and MiPAAF (Selmol project) funds.

References

Bagnato A, Schiavini F, Rossoni A, Maltecca C, Dolezal M, Medugorac I, Sölkner J, Russo V, Fontanesi L, Friedman A, Soller M and Lipkin E 2008. Quantitative trait loci affecting milk yield and protein percentage in a three-country Brown Swiss population. Journal of Dairy Science 91, 767-783.

Barrett JC, Fry B, Maller J and Daly MJ 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263-265.

Benzinou M, Creemers JW, Choquet H, Lobbens S, Dina C, Durand E, Guerardel A, Boutin P, Jouret B, Heude B, Balkau B, Tichet J, Marre M, Potoczna N, Horber F, Le Stunff C, Czernichow S, Sandbaek A, Lauritzen T, Borch-Johnsen K, Andersen G, Kiess W, Körner A, Kovacs P, Jacobson P, Carlsson LM, Walley AJ, Jørgensen T, Hansen T, Pedersen O, Meyre D and Froguel P 2008. Common nonsynonymous variants in *PCSK1* confer risk of obesity. Nature Genetics 40, 943-945.

Bosi P, Gremokolini C, Trevisi P, Mazzoni M, Bonilauri P, Sarli G and Casini L 2004. Oral challenge with *E. coli* K88 as a tool to assess growth and health performance in feeding trials of weaned pigs. Journées Recherche Porcine 36, 125-132.

Bruun CS, Jørgensen CB, Nielsen VH, Andersson L and Fredholm M 2006. Evaluation of the porcine melanocortin 4 receptor (*MC4R*) gene as a positional candidate for a fatness QTL in a cross between Landrace and Hampshire. Animal Genetics 37, 359-362.

Chang YC, Chiu YF, Shih KC, Lin MW, Sheu WH, Donlon T, Curb JD, Jou YS, Chang TJ, Li HY and Chuang LM 2010. Common *PCSK1* haplotypes are associated with obesity in the Chinese population. Obesity (Silver Spring) 18, 1404-1409.

Ciobanu DC, Lonergan SM and Huff-Lonergan EJ 2011. Genetics of meat quality and carcass traits, in: Rothschild MF, Ruvinsky A (Eds.), The Genetics of the Pig. CAB International, New York, pp. 355-389.

Clutter AC 2011. Genetics of performance traits, in: Rothschild MF, Ruvinsky A (Eds.), The Genetics of the Pig. CAB International, New York, pp. 325-354.

Du ZQ, Fan B, Zhao X, Amoako R and Rothschild MF 2009. Association analyses between type 2 diabetes genes and obesity traits in pigs. Obesity (Silver Spring) 17, 323-329.

Ekine C, Rowe S, Bishop S and de Koning D-J 2010. What is the best phenotype for genome-wide association studies in data with defined pedigree? Proceedings of the 9th World Congress of Genetics Applied to Livestock Production, 1-6 August, Leipzig, Germany. Communication no. 263.

Fan B, Du ZQ and Rothschild MF 2009a. The fat mass and obesity-associated (*FTO*) gene is associated with intramuscular fat content and growth rate in the pig. Animal Biotechnology 20, 58-70.

Fan B, Onteru SK, Nikkilä MT, Stalder KJ and Rothschild MF 2009b. Identification of genetic markers associated with fatness and leg weakness traits in the pig. Animal Genetics 40, 967-970.

Fan B, Onteru SK, Plastow GS and Rothschild MF 2009c. Detailed characterization of the porcine *MC4R* gene in relation to fatness and growth. Animal Genetics 40, 401-409.

Farooqi IS, Volders K, Stanhope R, Heuschkel R, White A, Lank E, Keogh J, O'Rahilly S and Creemers JW 2007. Hyperphagia and early-onset obesity due to a novel homozygous missense mutation in prohormone convertase 1/3. Journal of Clinical Endocrinology and Metabolism 92, 3369-3373.

Fernando RL, Nettleton D, Southey BR, Dekkers JC, Rothschild MF and Soller M 2004. Controlling the proportion of false positives in multiple dependent tests. Genetics 166, 611-619.

Fontanesi L, Colombo M, Beretti F and Russo V 2008a. Evaluation of *post mortem* stability of porcine skeletal muscle RNA. Meat Science 80, 1345-1351.

Fontanesi L, Davoli R, Nanni Costa L, Beretti F, Scotti E, Tazzoli M, Tassone F, Colombo M, Buttazzoni L and Russo V 2008b. Investigation of candidate genes for glycolytic potential of porcine skeletal muscle: association with meat quality and production traits in Italian Large White pigs. Meat Science 80, 780-787.

Fontanesi L, Scotti E, Buttazzoni L, Davoli R and Russo V 2009. The porcine fat mass and obesity associated (*FTO*) gene is associated with fat deposition in Italian Duroc pigs. Animal Genetics 40, 90-93.

Fontanesi L, Colombo M, Scotti E, Buttazzoni L, Bertolini F, Dall'Olio S, Davoli R and Russo V 2010a. The porcine tribbles homolog 3 (*TRIB3*) gene: Identification of a missense mutation and association analysis with meat quality and production traits in Italian heavy pigs. Meat Science 86, 806-813.

Fontanesi L, Scotti E, Buttazzoni L, Dall'Olio S, Bagnato A, Lo Fiego DP, Davoli R and Russo V 2010b. Confirmed association between a single nucleotide polymorphism in the *FTO* gene and obesity-related traits in heavy pigs. Molecular Biology Reports 37, 461-466.

Fontanesi L, Scotti E, Buttazzoni L, Dall'Olio S, Davoli R and Russo V 2010c. A single nucleotide polymorphism in the porcine cathepsin K (*CTSK*) gene is associated with back fat thickness and production traits in Italian Duroc pigs. Molecular Biology Reports 37, 491-495.

Fontanesi L, Speroni C, Buttazzoni L, Scotti E, Dall'Olio S, Nanni Costa L, Davoli R and Russo V 2010d. The *IGF2 intron3-g.3072G>A* polymorphism is not the only *Sus scrofa* chromosome 2p mutation affecting meat production and carcass traits in pigs: evidences from the effects of a cathepsin D (*CTSD*) gene polymorphism. Journal of Animal Science 88, 2235-2245.

Fontanesi L, Colombo M, Tognazzi L, Scotti E, Buttazzoni L, Dall'Olio S, Davoli R and Russo V 2011a. The porcine *TBC1D1* gene: mapping, SNP identification, and association study with meat, carcass and production traits in Italian heavy pigs. Molecular Biology Reports 38, 1425-1431.

Fontanesi L, Scotti E, Speroni C, Buttazzoni L and Russo V 2011b. A selective genotyping approach identifies single nucleotide polymorphisms in porcine chromosome 2 genes associated with production and carcass traits in Italian heavy pigs. Italian Journal of Animal Science 10, e15.

Forbes JM 1995. *Voluntary Food Intake and Diet Selection in Farm Animals*, edn 2. Ed JM Forbes. Wallingford, Oxon, UK: CAB International.

Gagnon J, Mayne J, Mbikay M, Woulfe J and Chrétien M 2009. Expression of PCSK1 (PC1/3), PCSK2 (PC2) and PCSK3 (furin) in mouse small intestine. Regulatory Peptides 152, 54-60.

Hamasima N, Mikawa A, Suzuki H, Suzuki K, Uenishi H and Awata T 2008. A new 4016-marker radiation hybrid map for porcine-human genome analysis. Mammalian Genome 19, 51-60.

Hawken RJ, Murtaugh J, Flickinger GH, Yerle M, Robic A, Milan D, Gellin J, Beattie CW, Schook LB and Alexander LJ 1999. A first-generation porcine whole-genome radiation hybrid map. Mammalian Genome 10, 824-830.

Heni M, Haupt A, Schäfer SA, Ketterer C, Thamer C, Machicao F, Stefan N, Staiger H, Häring HU and Fritsche A 2010. Association of obesity risk SNPs in *PCSK1* with insulin sensitivity and proinsulin conversion. BMC Medical Genetics 11, 86.

Houston RD, Cameron ND and Rance KA 2004. A melanocortin-4 receptor (*MC4R*) polymorphism is associated with performance traits in divergently selected Large White pig populations. Animal Genetics 35, 386-390.

Hu ZL and Reecy JM 2007. Animal QTLdb: beyond a repository. A public platform for QTL comparisons and integration with diverse types of structural genomic information. Mammalian Genome 18, 1-4.

Israel C and Weller JI 2002. Estimation of quantitative trait loci effects in dairy cattle populations. Journal of Dairy Science 85, 1285-1297.

Jackson RS, Creemers JW, Ohagi S, Raffin-Sanson ML, Sanders L, Montague CT, Hutton JC and O'Rahilly S 1997. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. Nature Genetics 16, 303-306.

Kilpeläinen TO, Bingham SA, Khaw KT, Wareham NJ and Loos RJ 2009. Association of variants in the *PCSK1* gene with obesity in the EPIC-Norfolk study. Human Molecular Genetics 18, 3496-3501.

Kim KS, Larsen N, Short T, Plastow G and Rothschild MF 2000. A missense variant of the porcine melanocortin-4 receptor (*MC4R*) gene is associated with fatness, growth, and feed intake traits. Mammalian Genome 11, 131-135.

Kojima M and Kangawa K 2005. Ghrelin: structure and function. Physiological Reviews 85, 495-522.

Lee SS, Chen Y, Moran C, Cepica S, Reiner G, Bartenschlager H, Moser G and Geldermann H 2003. Linkage and QTL mapping for *Sus scrofa* chromosome 2. Journal of Animal Breeding and Genetics 120 (Suppl. 1), 11-19.

Liu G, Jennen DG, Tholen E, Juengst H, Kleinwächter T, Hölker M, Tesfaye D, Un G, Schreinemachers HJ, Murani E, Ponsuksili S, Kim JJ, Schellander K and Wimmers K. 2007. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. Animal Genetics 38, 241-252.

Lloyd DJ, Bohan S and Gekakis N 2006. Obesity, hyperphagia and increased metabolic efficiency in Pc1 mutant mice. Human Molecular Genetics 15, 1884-1893.

Malek M, Dekkers JC, Lee HK, Baas TJ and Rothschild MF 2001. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. Mammalian Genome 12, 630-606.

Milan D, Hawken R, Cabau C, Leroux S, Genet C, Lahbib Y, Tosser G, Robic A, Hatey F, Alexander L, Beattie C, Schook L, Yerle M and Gellin J 2000. IMpRH server: an RH mapping server available on the Web. Bioinformatics 16, 558-559.

Muller L and Lindberg I 1999. The cell biology of the prohormone convertases PC1 and PC2. Progress in Nucleic Acid Research and Molecular Biology 63, 69-108.

Qi Q, Li H, Loos RJ, Liu C, Hu FB, Wu H, Yu Z and Lin X 2010. Association of *PCSK1* rs6234 with obesity and related traits in a Chinese Han population. PLoS One 5, e10590.

Rink A, Eyer K, Roelofs B, Priest KJ, Sharkey-Brockmeier KJ, Lekhong S, Karajusuf EK, Bang J, Yerle M, Milan D, Liu WS and Beattie CW 2006. Radiation hybrid map of the porcine genome comprising 2035 EST loci. Mammalian Genome 17, 878-885.

Rothschild MF, Hu ZL and Jiang Z 2007. Advances in QTL mapping in pigs. International Journal of Biological Sciences 3, 192-197.

Rozas J, Sanchez-DelBarrio JC, Messeguer X and Rozas R 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496-2497.

Russo V, Buttazzoni L, Baiocco C, Davoli R, Nanni Costa L, Schivazappa C and Virgili R 2000. Heritability of muscular cathepsin B activity in Italian Large White pigs. Journal of Animal Breeding and Genetics 117, 37-42.

Russo V., Fontanesi L., Scotti E., Beretti F., Davoli R., Nanni Costa L., Virgili R. and Buttazzoni L. 2008. Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs. Journal of Animal Science 86, 3300-3314.

Russo V and Nanni Costa L 1995. Suitability of pig meat for salting and the production of quality processed products. Pig News & Information 16, 7N-26N.

Salfen BE, Carroll JA and Keisler DH 2003. Endocrine responses to short-term feed deprivation in weanling pigs. Journal of Endocrinology 178, 541-551.

Scamuffa N, Calvo F, Chrétien M, Seidah NG and Khatib AM 2006. Proprotein convertases: lessons from knockouts. FASEB Journal 20, 1954-1963.

Shimogiri T, Kiuchi S, Hiraiwa H, Hayashi T, Takano Y, Maeda Y and Yasue H 2006. Assignment of 117 genes from HSA5 to the porcine IMpRH map and generation of a dense human-pig comparative map. Animal Genetics 37, 503-508.

Steiner DF, Rouillé Y, Gong Q, Martin S, Carroll R and Chan SJ 1996. The role of prohormone convertases in insulin biosynthesis: evidence for inherited defects in their action in man and experimental animals. Diabetes and Metabolism 22, 94-104.

Stephens M, Smith NJ and Donnelly P 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68, 978-989.

Stuber CW, Edwards MD and Wendel JF 1987. Molecular-marker facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. Crop Science 27, 639-648.

Switonski M, Stachowiak M, Cieslak J, Bartz M and Grzes M 2010. Genetics of fat tissue accumulation in pigs: a comparative approach. Journal of Applied Genetics 51, 153-168.

Tal-Stein R, Fontanesi L, Dolezal M, Scotti E, Bagnato A, Russo V, Canavesi F, Friedmann A, Soller M, Lipkin E 2010. A genome scan for QTL affecting milk somatic cell count in Israeli and Italian Holstein cows by means selective DNA pooling with multiple marker mapping. Journal of Dairy Science 93, 4913-4927.

Tamura K, Dudley J, Nei M and Kumar S 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24, 1596-1599.

Thomsen H, Reinsch N, Xu N, Looft C, Grupe S, Kühn C, Brockmann GA, Schwerin M, Leyhe-Horn B, Hiendleder S, Erhardt G, Medjugorac I, Russ I, Förster M, Brenig B, Reinhardt F,

Reents R, Blümel J, Averdunk G and Kalm E 2001. Comparison of estimated breeding values, daughter yield deviations and de-regressed proofs within a whole genome scan for QTL. Journal of Animal Breeding and Genetics 118, 357-370.

Thomsen H, Lee HK, Rothschild MF, Malek M and Dekkers JC 2004. Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine. Journal of Animal Science 82, 2213-2228.

Tortereau F, Gilbert H, Heuven HC, Bidanel JP, Groenen MA and Riquet J 2011. Number and mode of inheritance of QTL influencing backfat thickness on SSC2p in Sino-European pig pedigrees. Genetics Selection Evolution 43, 11.

Trevisi P, Melchior D, Mazzoni M, Casini L, De Filippi S, Minieri L, Lalatta-Costerbosa G and Bosi P 2009. A tryptophan-enriched diet improves feed intake and growth performance of susceptible weanling pigs orally challenged with *E. coli* K88. Journal of Animal Science 87, 148-156.

Trevisi P, Corrent E, Messori S, Casini L and Bosi P 2010. Healthy newly weaned pigs require more tryptophan to maximize feed intake if they are susceptible to *Escherichia coli* K88. Livestock Science 134, 236-238.

Virgili R and Schivazappa C 2002. Muscle traits for long matured dried meats. Meat Science 62, 331-343.

Wren AM, Small CJ, Abbott CR, Dhillo WS, Seal LJ, Cohen MA, Batterham RL, Taheri S, Stanley SA, Ghatei MA and Bloom SR 2001. Ghrelin causes hyperphagia and obesity in rats. Diabetes 50, 2540-2547.

Yerle M, Pinton P, Robic A, Alfonso A, Palvadeau Y, Delcros C, Hawken R, Alexander L, Beattie C, Schook L, Milan D and Gellin J 1998. Construction of a whole-genome radiation

hybrid panel for high-resolution gene mapping in pigs. Cytogenetics and Cell Genetics 82, 182-188.

Zhu X, Zhou A, Dey A, Norrbom C, Carroll R, Zhang C, Laurent V, Lindberg I, Ugleholdt R, Holst JJ and Steiner DF 2002. Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. Proceedings of the National Academy of Science of the USA 99, 10293-10298.

Zhu X, Cao Y, Voogd K and Steiner DF 2006. On the processing of proghrelin to ghrelin. Journal of Biological Chemistry 281, 38867-38870.

Table 1. Polymorphisms identified in the porcine *PCSK1* gene.

Polymorphism no.	Polymorphism (EMBL/GenBank HE599222)[1]	Position on Sscrofa10 chrU_scaffold4254	Gene region[2]
1	g.1199G>T	7407	Intron 3
2	g.1200G>A	7408	Intron 3
3	g.1696C>A[3]	9508	Intron 4
4	g.1737T>C[4]	9549	Intron 4
5	g.1959T(18_21)	10831	Intron 4
6	g.2550T>C	16332	Exon 6
7	g.2705C>T	16424	Intron 6
8	g.3676C>G	21009	Intron 8
9	g.4114T>C	23609	Intron 9
10	g.5181A>T[3]	35312	Intron 11
11	g.5459T>C	35590	Intron 12
12	g.6270T>C	39896	Exon 14
13	g.6390T>C	40016	Exon 14
14	g.6459T>C	40085	Exon 14

[1]Haplotypes (H1-9) inferred from the sequencing panel were (in parenthesis are reported the number of haplotype copies observed in the analysed breeds: ILW = Italian Large White; IL = Italian Landrace; ID = Italian Duroc; BL = Belgian Landrace; H = Hampshire; M = Meishan; SNP order is as reported in the table excluding the indel, polymorphism no. 5): H1, GGCTTCCTATTTT (5 H, 2 ILW, 1 BL, 1 ID, 1 IL); H2, GGCTTCGTATTTT (2 IL); H3, GGCTTCCTTTTTT (12 ID, 3 ILW, 2 IL); H4, TACTCCGTATTTT (1 IL); H5, GGATTCCTATTTT (1 ILW); H6, GAATTCCTATTTT (4 ILW, 1 BL, 1 ID, 1 M); H7, GGCTTTCTATTTT (2 M); H8, GGCTTCGCACCCC (1 M); H9, GGCCTCGCATCCT (1 H).

[2] According to the structure of the human *PCSK1* gene (Ensembl ENSG00000175426).

[3] SNPs genotyped and used for association studies.

[4] SNP genotyped but not useful for association studies because not polymorphic in the analysed populations.

Table 2. *PCSK1* gene statistics in the sequenced panel. Statistics were calculated at the breed level and considering together all sequenced animals.

Breeds/groups	No. of sequences[1]	*S*	*h*	*hD* (s.d.)	π % (s.d.)	D_T
Italian Duroc	14	3	3	0.275 (0.148)	0.011 (0.007)	-1.278
Italian Large White	10	3	4	0.778 (0.091)	0.030 (0.004)	+1.680
Italian Landrace	6	5	4	0.867 (0.129)	0.040 (0.012)	-0.315
Hampshire	6	5	2	0.333 (0.215)	0.032 (0.021)	-1.337
Meishan	4	9	3	0.833 (0.222)	0.090 (0.032)	-0.492
All[2]	42	13	9	0.763 (0.045)	0.037 (0.007)	-1.135

S = segregating SNP sites; *h* = number of haplotypes; *hD* = haplotype diversity (standard deviation); π = nucleotide diversity (standard deviation); D_T = Tajima's *D*.

[1]Two sequences for each animal.

[2]Including the sequenced Belgian Landrace pig.

Table 3. Association analysis between the g.1696C>A and g.5182A>T *PCSK1* SNP genotypes and estimated breeding values (EBV) and random residuals (RR) for meat production and carcass traits in Italian Large White (ILW) and Italian Duroc (ID) pigs and phenotypic measures for meat quality parameters in ILW pigs. Least square means are reported with their standard errors in parenthesis. Only results with P <0.10 are reported.

Breed[1]/SNP	Trait[2]	Genotype[3]			$P^{4,5}$
		11	***12***	***22***	
ILW/g.1696C>A	ADG/EBV (g)	40.370 (2.614)	29.453 (2.818)	28.918 (2.835)	0.0035
	FGR/EBV	-0.176 (0.015)	-0.145 (0.016)	-0.119 (0.017)	0.041
	LC/EBV (kg)	2.405 (0.183)	1.740 (0.197)	1.766 (0.198)	0.019
	ADG/RR (g)	11.317 (7.775)	-7.622 (8.433)	-12.490 (8.433)	0.087
	LC/RR (kg)	0.597 (0.265)	-0.507 (0.288)	-0.266 (0.288)	0.012
	Lactate (µ mol/g)	53.843 (1.615)	58.189 (1.686)	58.839 (1.736)	0.072
ILW/g.5182A>T	ADG/EBV (g)	29.946 (1.822)	43.298 (3.448)	45.100 (8.231)	0.0012
	FGR/EBV	-0.137 (0.011)	-0.184 (0.020)	-0.188 (0.048)	0.090
	HW/EBV (kg)	0.543 (0.042)	0.666 (0.080)	0.920 (0.191)	0.081
	LC/EBV (kg)	1.788 (0.127)	2.623 (0.241)	2.597 (0.575)	0.0057
	LC/RR (kg)	-0.303 (0.186)	0.860 (0.351)	0.663 (0.839)	0.011
	pH_1	5.916 (0.019)	5.988 (0.034)	5.831 (0.076)	0.070
	Lactate (µ mol/g)	58.038 (1.148)	53.229 (2.107)	52.600 (4.901)	0.097
	Glycogen (µ mol/g)	46.937 (1.929)	51.824 (3.300)	64.008 (6.882)	0.029
	CatB (nmol/min/g)	1.1502 (0.019)	1.204 (0.033)	1.0451 (0.071)	0.083
ID/g.1696C>A	ADG/EBV (g)	26.024 (1.858)	35.944 (4.905)	57.714 (7.866)	0.0002
	BFT/EBV (mm)	-1.615 (0.256)	-1.681 (0.675)	-4.057 (1.082)	0.091
	FGR/EBV	-0.142 (0.010)	-0.187 (0.027)	-0.261 (0.043)	0.011
	LC/EBV (kg)	1.757 (0.128)	1.891 (0.338)	4.221 (0.542)	0.00007
	VIF/EBV	-0.134 (0.087)	0.030 (0.228)	-1.153 (0.366)	0.018
	ADG/RR (g)	-0.354 (4.343)	25.867 (11.468)	9.960 (18.390)	0.097
	HW/RR (kg)	0.001 (0.066)	0.078 (0.175)	0.782 (0.280)	0.026

	LC/RR (kg)	0.003 (0.155)	-0.197 (0.410)	1.401 (0.658)	0.097
ID/g.5182A>T	ADG/EBV (g)	33.870 (3.497)	33.821 (2.412)	17.867 (3.061)	0.00009
	BFT/EBV (mm)	-2.588 (0.482)	-2.068 (0.332)	-0.627 (0.422)	0.004
	FGR/EBV	-0.172 (0.019)	-0.176 (0.013)	-0.111 (0.017)	0.005
	LC/EBV (kg)	2.399 (0.246)	2.069 (0.170)	1.221 (0.215)	0.0007
	VIF/EBV	-0.616 (0.163)	-0.243 (0.113)	0.232 (0.143)	0.0005
	ADG/RR (g)	3.770 (8.394)	13.152 (5.790)	-8.429 (7.349)	0.071
	BFT/RR (mm)	-0.614 (0.616)	-0.130 (0.425)	1.037 (0.539)	0.099
	FGR/RR	-0.018 (0.034)	-0.048 (0.024)	0.048 (0.030)	0.045
	HW/RR (kg)	0.008 (0.127)	0.208 (0.088)	-0.146 (0.112)	0.042
	VIF/RR	-0.060 (0.054)	0.029 (0.037)	0.124 (0.0470)	0.036

[1] ILW = Italian Large White (random group); ID = Italian Duroc (random group + selective genotyping group, excluding for visible intermuscular fat).

[2] ADG = average daily gain; BFT = backfat thickness; FGR = feed:gain ratio; HW = ham weight; LC = lean cuts; VIF = visible intermuscular fat; pH_1 = pH measured at 2 h *post mortem* on *M. semimembranosus*; pH_2 = pH measured at 24 h *post mortem* on the same muscle; GP = glycolytic potential; Catb = cathepsin B activity; EBV = Estimated Breeding Value; RR = Random Residual.

[3] Allele 1 is C for g.1696C>A and A for g.5182A>T. The number of animals for each genotypic class was: ILW/g.1696C>A: CC, n = 100; CA, n = 85; AA, n = 86; ILW/g.5182A>T: AA, n = 204; AT, n = 57; TT, n = 10; ID/g.1696C>A: CC, n = 247; CA, n = 36; AA, n = 14 (for VIF: CC, n = 163; CA, n = 21; AA, n = 13); ID/g.5182A>T: AA, n = 67; AT, n = 142; TT, n = 88 (for VIF: AA, n = 57; AT, n = 97; TT, n = 43).

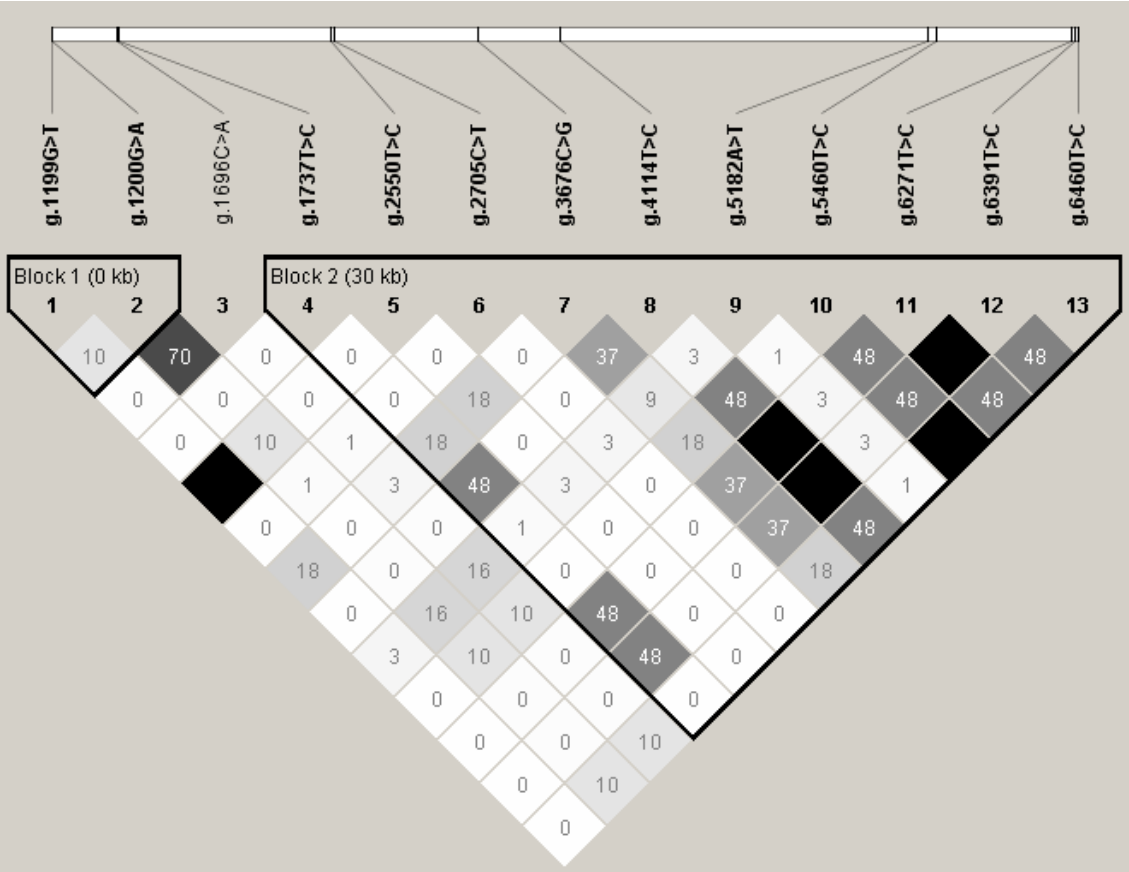
[4] Underlined when $P<0.05$. Of the 54 tests, only results with $P<0.10$ are reported. A complete list is available from the Authors.

Table 4. Effect of the g.1696C>A and g.5182A>T *PCSK1* genotypes on growth, feed intake and *PCSK1* mRNA expression level in oxyntic mucosa of weaning pigs. Least square means for each genotypic class are reported with their standard errors in parenthesis.

Trait[1]	**g.1696C>A genotypes**				**g.5182A>T genotypes**			
	CC (n. 22)	**CA (n. 54)**	**AA (n. 17)**	***P***[3]	**AA (n. 66)**	**AT (n. 25)**	**TT (n. 2)**	***P***
DG-1W (g)	41.1 (20.9)	23.8 (10.2)	27.9 (21.4)	0.765	29.4 (9.6)	24.9 (17.0)	n.e.[2]	0.735
DFI-1W (g)	130.0 (14.3)	123.3 (7.0)	141.4 (14.7)	0.546	132.5 (6.6)	113.3 (11.6)	n.e.	0.348
DG-2W (g)	228.8 (17.8)	273.5 (8.7)	295.3 (18.3)	0.084	269.3 (8.6)	268.7 (15.2)	n.e.	0.284
DFI-2W (g)	83.4 (28.0)	149.3 (13.6)	168.1 (28.4)	<u>0.038</u>	128.9 (13.2)	173.8 (23.1)	n.e.	0.966
PCSK1 expression	1267 (219)	1248 (108)	1057 (226)	0.749	1135 (99.2)	1415 (170.5)	n.e.	0.395

[1] DG-1W = Daily live weight gain during the first week; DFI-1W = Daily feed intake during the first week; DG-2W = Daily live weight gain during the second week; DFI-2W = Daily feed intake during the second week; PCSK1 expression = *PCSK1* mRNA quantification in gastric oxyntic mucosa (expressed in copies/µl cDNA).
[2] n.e.= not estimated.
[3] Significant results are underlined.

Figure 1. Linkage disequilibrium (r^2) plot at the porcine *PCSK1* gene. SNP were positioned on the Sscrofa10 chrU_scaffold4254 containing the porcine *PCSK1* gene. The polyT indel was not included in the analysis.



Supplementary material

Table S1. PCR primers, amplified *PCSK1* gene regions and analysis of the amplified fragments.

Primer pair	Primer forward (5'-3')	Primer reverse (5'-3')	Annealing temp.[1]	Amplified region (bp)[2]	Use[3]
1	TCCACTCAGCCGGGAGAC	GCTCAAGAGAGTGCAACCTG	58	Part of 5'-flanking region, exon 1, part of intron 1 (459)	Sequencing
2	CCTCTGAAATGTGTGAAACAGAA	TGGTTTGAAGACAAATGCAAA	61	Part of intron 1, exon 2, part of intron 2 (356)	Sequencing
3	GGAGCTAGTTAAAGGGAAGATGA	GCAACAACCCTTCCTCACAT	63	Part of intron 2, exon 3, part of intron 3 (357)	Sequencing
4	TGCATCAAGCAAATCCTGAG	TTTATGTGCACTGGCAGGAG	60	Part of intron 3, exon 4, part of intron 4 (311)	Sequencing, PCR-RFLP (*Taa*I and *Tai*I)[4]
5	GCAAGCTTTCCGGTTATCAG	CATTTGAGGGAAGCATTTCA	60	Part of intron 4, exon 5, part of intron 5 (393)	Sequencing
6	CTTGGGCCCTTCATCTGATA	TGCAGCAAATTTTGAAAGGA	58	Part of intron 5, exon 6, part of intron 6 (354)	Sequencing
7	ACAGGCATGTGGGACATACA	CCCCCATATTTAAACAGTCAAG	58	Part of intron 6, exon 7, part of intron 7 (324)	Sequencing
8	GCTGGAGTACCTGGAGTGGA	CCCCAACTGAGACATCAAGC	58	Part of intron 7, exon 8, part of intron 8 (426)	Sequencing

9	CTGTGGTTTGGAGGAGGAAG	TGCAATGTTAACAGGAAGAGAGG	61	Part of intron 8, exon 9, part of intron 9 (354)	Sequencing
10	TCTCTGAATGAAAATGCTTGTTTT	ACTCTGGCAAATGCCATCTC	59	Part of intron 9, exon 10, part of intron 10 (407)	Sequencing
11	CCTAATCCTGAACTGGGTTCT	ACCCTTAAAAATTTAAACACATGG	59	Part of intron 10, exon 11, part of intron 11 (300)	Sequencing
12	CCAAGGGGACAGAACTTGAA	CTCATTTCTCCCAGTTTCCA	58	Part of intron 11, exon 12, part of intron 12 (357)	Sequencing, PCR-RFLP (*Alu*I) [4], RH mapping
13	GCTTTTTGGAAGGAGTTTGTTT	CTAACTGCTGAGCCACGATG	58	Part of intron 12, exon 13, part of intron 13 (425)	Sequencing
14	TAATAACGGCTTGCCTCCTC	CCCTGTGCTTGTAAGGCTTC	60	Part of intron 13, part of exon 14 (356)	Sequencing
GE_15	GATGGGAAAACTGTGGAAGG	CAGCATACCAGGGGGATAGG	57	Part of exon 7 and 8 (177)	External primers for qPCR analysis
GE_16	GTCAAACAGGGGAGACAAGG	CTGATGGAGATGGTGTAGATGC	62	Exon 8 (80)	Internal primers for qPCR analysis; qualitative gene expression analysis

[1] Annealing temperature (°C).

[2] Amplified fragment size does not include primers.

[3] Analyses carried out with the amplified fragments.

[4] PCR-RFLP patterns were: g.1696C>A analysed with *Taa*I, allele C = 351 bp, allele A = 244 + 127 bp; g.1737T>C analysed with *Tai*I, allele T = 351 bp, allele C = 285 + 66 bp; g.5182A>T analysed with *Alu*I, allele A = 337 + 60 bp, allele T = 397 bp.

Table S2. Allele frequencies of three genotyped *PCSK1* SNPs (g.1696C>A, g.1737T>C, and g.5182A>T) in different pig breeds. Only the frequency of the first allele is included in the table.

Breeds	No. of pigs	g.1696C	g.1737T	g.5182A
Italian Duroc[1]	297	0.892	1.000[2]	0.465
Italian Large White[1]	271	0.526	1.000[2]	0.858
Italian Landrace	20	0.900	1.000	0.975
Hampshire	18	0.972	0.889	0.972
Belgian Landrace	22	0.977	1.000	0.519
Casertana	30	0.800	n.t.[3]	0.833
Meishan	12	0.667	1.000	1.000

[1] Allele frequencies are reported for pigs used for association analysis of Table 3.

[2] Based on the analysis of 100 animals for the two breeds.

[3] Not tested.

Table S3 Additive (*a*) and dominance (*d*) effects (with standard errors in parenthesis: s.e.) obtained for the *PCSK1* SNP analysed in the Italian Large White (ILW) and Italian Duroc (ID) pigs. Results are reported for association analyses (see Table 3) with $P < 0.10$.

Breed/SNP[1]	Trait[2]	Additive effect (s.e.)	*P*	Dominance effect (s.e)	*P*	|d/a|[3]
ILW/g.1696C>A	ADG/EBV (g)	-5.726 (1.928)	0.003	-5.190 (3.415)	0.130	0.906
	FGR/EBV	0.028 (0.011)	0.012	0.002 (0.020)	0.903	0.085
	LC/EBV (kg)	-0.319 (0.135)	0.018	-0.346 (0.239)	0.148	1.083
	ADG/RR (g)	-11.903 (5.735)	0.039	-7.036 (10.198)	0.491	0.591
	LC/RR (kg)	-0.431 (0.196)	0.028	-0.672 (0.348)	0.055	1.559
	Lactate (µ mol/g)	2.498 (1.194)	0.038	1.848 (1.999)	0.356	0.740
ILW/g.5182A>T	ADG/EBV (g)	7.577 (4.215)	0.073	5.775 (5.446)	0.290	0.763
	FGR/EBV	-0.026 (0.025)	0.304	-0.021 (0.032)	0.507	0.808
	HW/EBV (kg)	0.189 (0.098)	0.055	-0.066 (0.127)	0.604	0.349
	LC/EBV (kg)	0.405 (0.294)	0.171	0.430 (0.380)	0.259	1.063
	LC/RR (kg)	0.483 (0.430)	0.262	0.680 (0.555)	0.222	1.407
	pH_1	-0.042 (0.039)	0.275	0.114 (0.051)	0.025	2.714
	Lactate (µ mol/g)	-2.719 (2.508)	0.280	-2.090 (3.261)	0.522	0.769
	Glycogen (µ mol/g)	8.535 (3.469)	0.015	-3.649 (4.617)	0.430	0.428
	CatB (nmol/min/g)	-0.053 (0.036)	0.145	0.107 (0.048)	0.026	2.019
ID/g.1696C>A	ADG/EBV (g)	15.845 (4.041)	0.0001	-5.925 (6.355)	0.352	0.374
	BFT/EBV (mm)	-1.221 (0.556)	0.029	1.156 (0.874)	0.187	0.947
	FGR/EBV	-0.059 (0.022)	0.007	0.014 (0.034)	0.679	0.241
	LC/EBV (kg)	1.232 (0.278)	0.00003	-1.098 (0.438)	0.013	0.891
	VIF/EBV	-0.510 (0.188)	0.007	0.6736 (0.296)	0.024	1.322
	ADG/RR (g)	5.157 (9.448)	0.586	21.064 (14.859)	0.157	4.085
	HW/RR (kg)	0.390 (0.144)	0.007	-0.313 (0.226)	0.167	0.803
	LC/RR (kg)	0.699 (0.338)	0.039	-0.899 (0.532)	0.092	1.286

ID/g.5182A>T	ADG/EBV (g)	-8.001 (2.324)	0.0007	7.953 (3.349)	0.018	0.994
	BFT/EBV (mm)	0.981 (0.320)	0.002	-0.461 (0.461)	0.319	0.470
	FGR/EBV	0.031 (0.013)	0.015	-0.035 (0.018)	0.054	1.139
	LC/EBV (kg)	-0.589 (0.163)	0.0004	0.259 (0.236)	0.273	0.440
	VIF/EBV	0.424 (0.108)	0.0001	-0.051 (0.156)	0.744	0.120
	ADG/RR (g)	-6.099 (5.578)	0.275	15.481 (8.040)	0.055	2.538
	BFT/RR (mm)	0.825 (0.409)	0.044	-0.342 (0.590)	0.563	0.415
	FGR/RR	0.171 (0.171)	0.148	0.294 (0.294)	0.058	1.715
	HW/RR (kg)	-0.077 (0.085)	0.365	0.277 (0.122)	0.0239	3.605
	VIF/RR	0.092 (0.036)	0.011	-0.004 (0.052)	0.946	0.038

[1] ILW = Italian Large White (random group); ID = Italian Duroc (random group + selective genotyping group, excluding for visible intermuscular fat).

[2] ADG = average daily gain; BFT = backfat thickness; FGR = feed:gain ratio; HW = ham weight; LC = lean cuts; VIF = visible intermuscular fat; EBV = Estimated Breeding Value; RR = Random Residual.

[3] The ratio $|d/a|$ was considered to indicate possible gene effects (Stuber *et al.*, 1987): $|d/a|<0.2$, additive; $0.2<|d/a|<0.8$, partial dominance; $0.8<|d/a|<1.2$, dominance; $|d/a|>1.2$, overdominance.

Table S4. Estimated regression coefficients and standard errors (SE) of the haplotype substitution effect for estimated breeding values (EBV) and random residuals (RR) of production traits in the Italian Large White (ILW) and Italian Duroc (ID) pigs.

Breed[1]	Trait[2]	Haplotype [A:A][3]		Haplotype [C:A][3]		Haplotype [C:T][3]	
		Regression coefficient (SE)	*P*[4]	Regression coefficient (SE)	*P*[4]	Regression coefficient (SE)	*P*[4]
ILW	ADG/EBV (g)	-6.504 (1.934)	0.0009	2.409 (2.116)	0.256	10.604 (3.217)	0.001
	BFT/EBV (mm)	-0.075 (0.281)	0.791	0.378 (0.301)	0.211	-0.537 (0.467)	0.251
	FGR/EBV	0.031 (0.011)	0.006	-0.018 (0.012)	0.138	-0.036 (0.019)	0.056
	HW/EBV (kg)	-0.055 (0.045)	0.224	-0.009 (0.049)	0.845	0.171 (0.074)	0.022
	LC/EBV (kg)	-0.367 (0.135)	0.007	0.120 (0.147)	0.415	0.630 (0.225)	0.005
	ADG/RR (g)	-12.105 (5.755)	0.036	7.233 (6.219)	0.246	16.461 (9.598)	0.087
	BFT/RR (mm)	-0.201 (0.374)	0.591	0.471 (0.401)	0.241	-0.362 (0.622)	0.561
	FGR/RR	0.047 (0.024)	0.047	-0.038 (0.026)	0.138	-0.045 (0.040)	0.254
	HW/RR (kg)	-0.074 (0.096)	0.443	-0.016 (0.103)	0.880	0.247 (0.159)	0.122
	LC/RR (kg)	-0.488 (0.197)	0.014	0.159 (0.214)	0.460	0.874 (0.328)	0.008
ID	ADG/EBV (g)	14.609 (3.666)	<0.0001	3.244 (2.485)	0.193	-9.851 (2.355)	<0.0001
	BFT/EBV (mm)	-0.879 (0.507)	0.084	-0.786 (0.334)	0.019	1.140 (0.321)	0.0005
	FGR/EBV	-0.060 (0.020)	0.003	-0.013 (0.013)	0.332	0.039 (0.013)	0.003

LC/EBV (kg)	0.912 (0.256)	0.0004	0.285 (0.172)	0.099	-0.687 (0.163)	<0.0001
VIF/EBV	-0.327 (0.172)	0.058	-0.351 (0.113)	0.002	1.025 (0.412)	0.013
ADG/RR (g)	18.591 (8.657)	0.033	0.118 (5.778)	0.984	-6.660 (5.604)	0.236
BFT/RR (mm)	-0.091 (0.646)	0.888	-0.960 (0.424)	0.024	1.025 (0.412)	0.013
FGR/RR	-0.072 (0.035)	0.042	-0.012 (0.024)	0.608	0.037 (0.023)	0.103
HW/RR (kg)	0.334 (0.132)	0.012	-0.049 (0.088)	0.576	-0.096 (0.086)	0.263
LC/RR (kg)	0.567 (0.308)	0.067	0.076 (0.205)	0.709	-0.265 (0.199)	0.184
VIF/RR	-0.039 (0.055)	0.476	-0.083 (0.036)	0.024	0.081 (0.035)	0.022

[1] ILW = Italian Large White; ID = Italian Duroc.

[2] ADG = average daily gain; BFT = backfat thickness; FGR = feed:gain ratio; HW = ham weight; LC = lean cuts; VIF = visible intermuscular fat; EBV = Estimated Breeding Value; RR = Random Residual.

[3] Haplotypes are indicated as follows: [A:A] = g.1696A and g.5182A; [C:A] = g.1696C and g.5182A; [C:T] = g.1696C and g.5182T. The number of the [A:A], [C:A], and [C:T] haplotype copies were: 249, 218, and 70 in ILW; 57, 223, and 309 in ID. Haplotype [A:T] was inferred only in 7 (ILW) and 6 (ID) copies and was not considered in this analysis.

[4]Underlined when $P<0.05$.

Table S5. Allele frequencies for the polymorphisms at the *PCSK1* locus and probability from Fisher's two-tailed exact test or Chi square test (*P*) of equal frequency in the negative (Ne) *vs* positive (Po) groups of pigs selected according to the extreme and divergent estimated breeding values (EBVs) for two traits.

Breed/Trait[1]	Groups	No. of pigs[5]	Allele frequencies		*P*[6]	No. of pigs[5]	Allele frequencies		*P*[6]
			g.1696C	g.1696A			g.5182A	g.5182T	
ILW/BFT EBV	u Ne[2]	269	0.556	0.444	0.566	276	0.844	0.156	0.173
	u Po[2]	274	0.538	0.462		279	0.873	0.127	
ID/VIF EBV	Ne[3]	50	0.950	0.050	0.283	50	0.400	0.600	<u>0.034</u>
	Po[3]	50	0.900	0.100		50	0.250	0.750	
	u Ne[4]	29	0.931	0.069	0.749	29	0.431	0.569	0.089
	u Po[4]	33	0.909	0.091		33	0.273	0.727	

[1] ILW = Italian Large White; ID = Italian Duroc; BFT = back fat thickness; VIF = visible intermuscular fat; EBV = estimated breeding value.

[2] Unrelated (u) pigs at the second generation with negative EBV (Ne) and positive (Po) EBV of the groups of pigs selected according to the extreme and divergent EBV for BFT.

[3] Animals with negative EBV (Ne) and positive (Po) EBV of the groups of pigs selected according to the extreme and divergent EBV for VIF.

[4] Unrelated (u) pigs at the second generation with negative EBV (Ne) and positive (Po) EBV of the groups of pigs selected according to the extreme and divergent EBV for VIF. These animals were among the 50 + 50 animals included in the rows above.

[5] Number of pigs of each group. A few ILW pigs of the selected extreme tails (560 + 560) for BFT EBV have not been genotyped.

[6] Significant results are underlined. The tests compare the allele frequencies of the groups adjacent to the *P* value.

Figure S1. Expression of the porcine *PCSK1* gene in different tissues. *GAPDH* transcript amplification was included as control. 1 = brain; 2 = fat; 3 = spleen; 4 = heart; 5 = liver; 6 = kidney; 7 = thyroid; 8 = lung; 9 = skeletal muscle; 10 = gastric oxyntic mucosa.

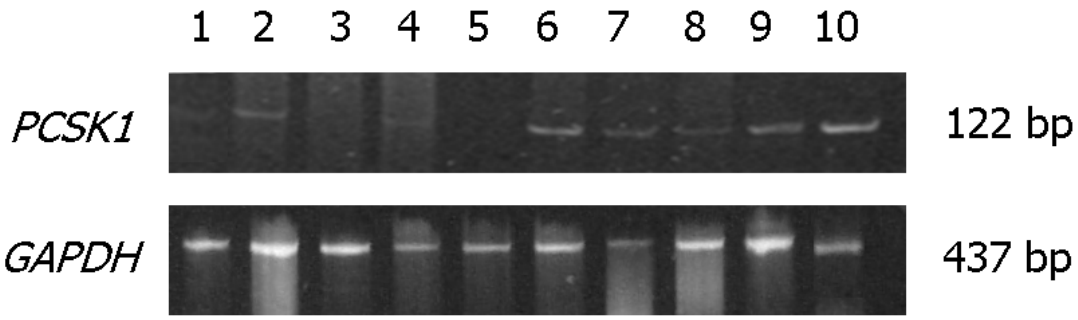
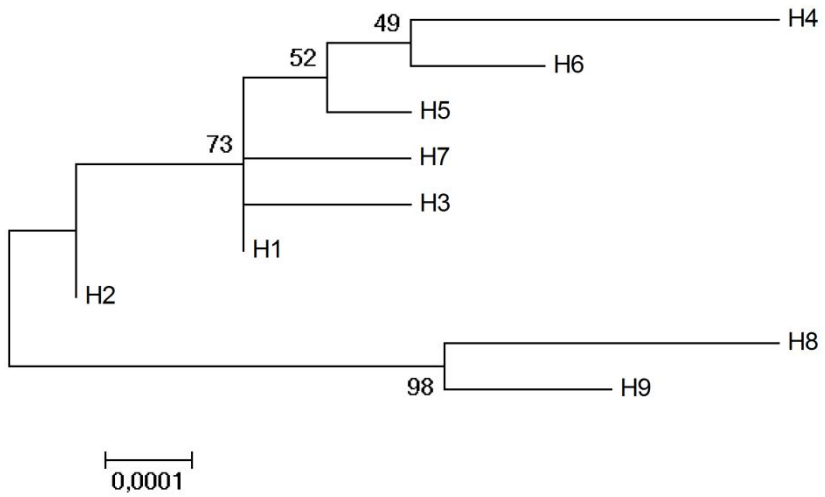


Figure S2. Neighbour-Joining tree obtained with the porcine *PCSK1* haplotypes. Haplotypes are defined as a note to Table 1. The frequency in the bootstrap resampling is reported at each node.



Chapter three

Analysis of association between the *MUC4* g.8227C>G polymorphism and production traits in italian heavy pigs using a selective genotyping approach

L. Fontanesi [1,4] Francesca Bertolini[1,4] 1,4, Stefania Dall'Olio[1] 1, Luca Buttazzoni[2], Maurizio Gallo[3], and Vincenzo Russo[1]

[1]*Dept. of Agro-Food Science and Technology, Sezione di Allevamenti Zootecnici, University of Bologna, Viale Fanin 48, 40127 Bologna, Italy*

[2]*CRA, Centro di Ricerca per la Produzione delle Carni e il Miglioramento Genetico, Via Salaria 31, 00015 Monterotondo Scalo, Roma, Italy*

[3]*Associazione Nazionale Allevatori Suini (ANAS), Via L. Spallanzani 4, 00161 Roma, Italy*

[4]These Authors contributed equally to this work

Submitted to ***Animal biotechnology***, accepted for publication.

ABSTRACT

Susceptibility to enterotoxigenic *Escherichia coli* (ETEC) K88 strains that express F4ab and F4ac fimbriae is inherited as a dominant autosomal locus (*F4bcR*), with the alternative recessive allele determining resistance. The susceptible allele appeared also to be associated with a favorable growth rate even if with discordant results. A single nucleotide polymorphism (SNP) in exon 7 of the mucin 4 (*MUC4*) gene (DQ848681: g.8227C>G), shown to be in close linkage disequilibrium with the *F4bcR* locus in different pig populations, has been used as a marker to identify susceptible animals, substituting invasive villous adhesion tests. We herein analysed this SNP in Italian local breeds and applied a selective genotyping approach in Italian Large White, Italian Landrace and Italian Duroc comparing allele frequency distribution in groups of pigs with extreme and divergent estimated breeding values (EBV) for average daily gain (ADG) and backfat thickness (BFT) to evaluate if this marker is associated with these traits in performance tested pigs. Allele C (associated with susceptibility to ETEC) was associated with higher ADG and BFT in Italian Large White pigs (P=6.66E-04 and P=0.012, respectively) and higher ADG in Italian Landrace (7.23E-12). This polymorphism was poorly informative in Italian Duroc pigs. Antagonistic associations of the *MUC4* g.8227C>G alleles on susceptibility to ETEC and growth evidence the complexity of applying marker assisted selection in pig breeding.

Running head: *MUC4* SNP and production traits in pigs

Key words: MUC4, Italian heavy pigs, polymorphism, growth rate, selective genotypying

INTRODUCTION

Enterotoxigenic *Escherichia coli* (ETEC) K88 strains that express F4ab and F4ac fimbriae represent the main cause of diarrhea and death of piglets before or just after weaning. Susceptibility or resistance to ETEC is determined by the presence or absence of receptors for these fimbriae on the brush border of pig intestinal epithelium.[1] Susceptibility (presence of receptors) is inherited as a dominant autosomal locus (*F4bcR*) with the alternative recessive allele (absence of receptors) determining resistance.[2-4] The F4bcR locus has been mapped on porcine chromosome 13, in a region where the mucin 4 (*MUC4*) gene is located [5-8]. *MUC4* encodes for a membrane-bound-O-glycoprotein that is widely expressed on the surface of gastrointestinal epithelial cells in which has a function of protecting and lubricating the epithelial surfaces. Therefore *MUC4* has been considered a strong positional and functional candidate gene for the *F4bcR* locus. A single nucleotide polymorphism (DQ848681: g.8227C>G) in exon 7 of this gene (known as *Xba*I polymorphism), shown to be in very close linkage disequilibrium with the *F4bcR* locus in different pig populations, has been used as a marker to identify susceptible animals, substituting invasive villous adhesion tests.[9-12] Allele C is associated with the absence of the receptors (resistant) whereas allele G is associated with the presence of the receptors (susceptible), even if *MUC4* does not seem the causative gene for the *F4bcR* locus.[13,14]

Mucins play also many other important roles in growth, fetal development, epithelial renewal and differentiation, epithelial integrity, carcinogenesis, and metastasis[15,16] suggesting that polymorphisms in the *MUC4* gene could be associated with other important production traits. Interestingly, a few studies of the *F4bcR* locus using villous based tests[17,18] have evidenced that the presence of F4ab and F4ac receptors could be associated with higher growth rate in pigs during the fattening period, even if this effect was not reported by Baker et al.[19]. Contrasting results about possible effects on growth efficiency were reported for the pre-weaning period.[17,18] However, the suggested antagonistic effects of F4bcR alleles on susceptibility to ETEC and

growth after weaning could result in complex interactions between natural and artificial selection, that might influence allele frequencies at this locus in different pig populations.[18]

We herein analysed the *MUC4* g.8227C>G SNP in Italian Large White, Italian Landrace and Italian Duroc as well as Italian local breeds and, using a selective genotyping approach, we showed that this marker is associated with growth rate and backfat thickness in Italian heavy pigs.

MATERIALS AND METHODS

Animals and traits

The association study was conducted using a selective genotyping approach with animals of three heavy pig breeds (Italian Large White, Italian Landrace and Italian Duroc) chosen according to the extreme and divergent estimated breeding values (EBVs) for average daily gain (ADG; all breeds) or for back fat thickness (Italian Large White only; see Table 1). Details of EBV calculation are reported below. All these animals were performance tested at the Test Station of the National Pig Breeder Association (ANAS) during the period 1996-2007. These pigs are those included in triplets of the same litter (two females and one castrated male) that are used for the evaluation of a candidate boar from the same litter (sib-test). Performance tested pigs start the evaluation period at 30 to 45 days of age and they end it at 155 ± 5 kg live weight. The nutritive level was *quasi ad libitum*. During the test period, body weight was measured bimonthly, then daily gain was calculated. At the end of test, animals were transported to a commercial abattoir where they were slaughtered. Backfat thickness was measured on the carcasses at the level of *Musculus gluteus medius.*

The Italian Large White pigs were all females (one female per triplet, and at least two-generation unrelated) chosen according to their ADG-EBV (200 with most negative and 200 with most positive EBV) or their BFT-EBV (280 with most negative and 280 with most positive EBV) within a performance tested population of ~12,000 pigs of this breed (Table 1). The Italian Landrace (141 females and 59 castrated males) were chosen according to their ADG-EBV (100

with most negative and 100 with most positive EBV) within a performance tested population of ~5,000 pigs of this breed (Table 1). The Italian Duroc (134 females and 66 castrated males) were chosen according to their ADG-EBV (100 with most negative and 100 with most positive EBV) within a performance tested population of ~7,000 pigs of this breed (Table 1).

In addition, minimum related pigs of four Italian local breeds (Calabrese, Casertana, Cinta Senese and Nero Siciliano), for which no phenotypic traits were available, were used for allele frequency analysis (Table 2).

Genotyping

Genomic DNA was extracted from blood using a standard protocol. After quality control, a few animals of the selective genotyping panels were excluded from genotyping or genotyping failed (Table 3). Genotyping of the *MUC4* g.8227C>G SNP was carried out by PCR-RFLP as carried out by Jørgensen et al.[9]. Briefly, PCR was carried out using a PTC-100 (MJ Research, Watertown, MA, USA) thermal cycler in a total volume of 20 µl that included 10 ng of genomic DNA, 10 pmol of each primer (forward 5'-GTGCCTTGGGTGAGAGGTTA-3, reverse 5'-CACTCTGCCGTTCTCTTTCC-3'), 2.0 mM $MgCl_2$, 2.5 mM each dNTP, 1 U of AmpliBioTherm DNA polymerase (Fisher Molecular Biology, Trevose, PA, USA). The amplification profile was the following: an initial step of denaturation for 5 min at 95 °C; 35 cycles of 30 s at 95°C, 30 s at 61.5 °C and 30 s at 72 °C; the final extension step was for 5 min at 72 °C. Then, 3-5 µl of the PCR product was used for digestion with 3 U of *Xba*I (MBI Fermentas, Vilnius, Lithuania). PCR-RFLP products were resolved on 10% polyacrylamide/ bis-acrylamide 29:1 gels and visualized with 1× GelRed Nucleic Acid Gel Stain (Biotium Inc., Hayward, CA, USA). Allele C resulted in an undigested product of 367 bp whereas allele G pattern was composed by two fragments of 151 bp and 216 bp.

Statistical analysis

Estimated breeding values for ADG (expressed in g) and BFT (expressed in mm) were predicted by a BLUP-multiple trait animal model including the fixed factors of age at the beginning of test, body weight at slaughter, age at slaughter, day of slaughtering and inbreeding coefficient, besides the random factors of animal and litter. Means and measures of variability of the considered EBVs are reported in Table 1. Chi square or two tailed Fisher's exact tests (where appropriate) of significance of allele frequency differences between the two extreme tails of genotyped pigs was calculated for each trait/breed combination considering all animals or only animals without common parents (for the Italian Landrace and Italian Duroc).

RESULTS AND DISCUSSION

Allele frequencies of the *MUC4* g.8227C>G SNP in Italian pig breeds including several local pig breeds are reported in Table 2. Allele G, indicated as susceptible to ETEC infection, has been identified in all breeds except in Cinta Senese. However, its frequency is low in all other local breeds and in the Italian Duroc breed, ranging from 0.05 in this latter breed to 0.28 in Nero Siciliano. This might indirectly support the higher rusticity of these breeds compared to other commercial breeds and could be, at least for the local breeds, derived by an adaptation to the extensive management operated in marginal conditions in which disease resistance is an important requisite. The frequency of the G allele was close to 0.50 in Italian Large White and Italian Landrace. These results for the Italian Duroc and Large White populations match, to some extent, those of the other reports that investigated allele frequencies for this marker in different pig breeds, including Danish and Swiss Duroc (0.06) Yorkshire (0.58), Swiss Large White (0.51) and Swiss Landrace (0.51), whereas were quite different for the Italian Landrace compared to the Danish Landrace population (0.96).[9,13]

Association analysis was carried out using a selective genotyping and comparing the distribution of the two *MUC4* g.8227C>G alleles in the extreme and divergent tails for the considered traits in three Italian heavy pig breeds (Table 2). In Italian Large White, allele G was more frequent in the tails with positive ADG-EBV (P=6.66E-04) and negative BFT-EBV (i.e. lower BFT; P=0.012), in agreement with the correlation between these two traits. The same highly significant difference in allele frequencies between the two tails for ADG-EBV was obtained in the Italian Landrace. Again, allele G was more frequent in the positive ADG-EBV group, both considering all selected animals (P<1.0E-20) or only two generation unrelated pigs (P=7.23E-12). In Italian Landrace, the two alleles in the two different tails were at opposite frequency, strongly indicating that the *MUC4* g.8227C>G SNP is associated with ADG in this breed and confirming the results obtained in the Italian Large White pigs. In Italian Duroc, allele G was observed at very low frequency, preventing any possibility to evidence difference between the two extreme ADG-EBV tails.

Results we obtained genotyping a marker in close linkage disequilibrium with the *F4bcR* locus are in agreement with results reported by Edfors-Lilja et al.[17] and Yan et al.[18], who investigated this locus using villous based tests to evidence the presence/absence of F4ab and F4ac receptors in Swedish Yorkshire x Swedish Landrace crosses and in a White Duroc x Erhulian intercross F2 population, respectively. In both studies pigs with indicated F4 receptors grew faster than pigs without receptors during the fattening period. Yan et al. [18] showed that animals with F4ab and F4ac receptors grew faster also during the pre-weaning period. The contrary was true in one of the two pig sets reported by Edfors-Lilja et al.[17] in which the incidence of diarrhoea was twice as high as in a second set of pigs in which there was no difference on growth rate between the piglets with or without receptors. A lower growth rate in the piglets with the receptors was also reported by Bosi et al.[20] but in this study animals were orally challenged with K88 *E. coli*. In untreated controls this difference disappeared. Similar evidences were reported in another challenging experiment where growth rate in pre-weaned

challenged pigs was lower than that of control piglets not differentiated according to their potential sensitivity or resistance to K88 *E. coli*.[21] Therefore, it could be possible that the contrasting results obtained by Edfors-Lilja et al.[17] and Yan et al.[18] during the suckling period might be derived by differences of diarrhoea incidence in the different experiments that largely influenced piglets growth early in life. It is worth to mention that all Italian heavy pigs we genotyped in this study were performance tested after weaning (from 30-45 days to a final weight of about 155 kg). For this reason we could not evaluate if there were differences between the two piglet life periods. In our animals, EBV for ADG does not directly include information about the pre-weaning period even if pre-selection of animals for the Test station takes into account health information. Summing up and considering the way in which Italian heavy pigs are performance tested, as the *MUC4* allele in close linkage disequilibrium with the *F4bcR* susceptible allele is also associated with higher growth rate, the susceptible allele is maintained in the population as a result of the selection for the pigs that have a favourable growth rate. On the other hand natural selection against susceptible genotypes (that might be in part eliminated during the pre-weaning period as animals die more frequently than those carrying the resistant genotype) could counterbalance in some way their positive effects on ADG. Therefore balanced natural and artificial selection might maintain the susceptible allele in the Italian Large White and Italian Landrace populations, as also suggested by Yan et al.[18] in other pig populations.

In addition, our study is the first that has reported an effect of the *MUC4* polymorphism that marks the *F4bcR* locus on BFT. This might essentially be due to the correlation between ADG and BFT as already mentioned above. Indeed, the significant allele frequency difference for BTF in Italian Large White was not as high as that reported for ADG. It will be interesting to confirm this issue in the Italian Landrace population for which the selective genotyping approach of study was carried out for ADG only. Comparing the results obtained for ADG in the Italian Large White and Italian Landrace breeds, the latter showed the largest difference in allele frequency suggesting a very strong association between allele G and higher growth rate. As the *MUC4*

g.8227C>G SNP seems only a marker in close linkage disequilibrium with *F4bcR* locus,[13,14] it remains to demonstrate if the same *F4bcR* causative mutation has pleiotropic effects on growth performances or if the effects on growth are due to another mutation in close linkage disequilibrium with that that should cause ETEC susceptibility. The difference observed between these two breeds could be interesting for this perspective, as at present, the mutation determining the *F4bcR* locus has not yet been identified. To this aim, haplotype analysis across populations has been already used to better define a region containing the *F4bcR*.[13,14,22] Targeted haplotype analysis in Italian heavy pigs, using SNPs already reported in this chromosome region,[22,23] including polymorphisms detected in other mucin genes,[24,25] coupled with villous based tests to evidence the presence/absence of F4ab and F4ac receptors, should provide additional information to disentangle this question that has important practical applications.

The *MUC4* g.8227C>G marker is at present used in marker assisted selection plans to eradicate or reduce the frequency of the susceptible allele of the *F4bcR* locus in commercial pig populations.[9] Considering its antagonistic associations on susceptibility to ETEC and growth its implementation in marker assisted selection plans could produce a lower improvement on growth rate in these populations that might be counterbalanced by the economic gain derived by the higher number of weaned piglets. This locus is a good example that shows the complexity of applying marker assisted selection in pig breeding.

ACKNOWLEDGEMENTS

We thank ANAS for providing samples and data. This work was funded by Italian Ministry of Agriculture (MiPAAF Selmol and INNOVAGEN projects) and University of Bologna (FAGenomicH project).

REFERENCES

1. Van den Broeck W, Cox E, Oudega B, Goddeeris BM. The F4 fimbrial antigen of *Escherichia coli* and its receptors. ***Vet Microbiol.*** 2000;71:223–244.

2. Sellwood R, Gibbons RA, Jones GW, Rutter JM. Adhesion of enteropathogenic *Escherichia coli* to pig intestinal brush borders: the existence of 2 pig phenotypes. *J Med Microbiol.* 1975;8:405–411.

3. Gibbons RA, Sellwood R, Burrows M, Hunter PA. Inheritance of resistance to neonatal Escherichia coli diarrhea in pig: examination of the genetic system. *Theor Appl Genet*. 1977; 51:65–70.

4. Bijlsma IG, de Nijs A, van der Meer C, Frik JF. Different pig phenotypes affect adherence of *Escherichia coli* to jejunal brush borders by K88ab, K88ac, or K88ad antigen. *Infection Immunity.* 1982;37:891–894.

5. Guérin G, Duval-Iflah Y, Bonneau M, Bertaud M, Guillaume P, Ollivier L. Evidence for linkage between K88ab, K88ac intestinal receptors to *Escherichia coli* and transferrin loci in pigs. *Anim Genet.* 1993;24:393–396.

6. Edfors-Lilja I, Gustafsson U, Duval-Iflah Y, Ellergren H, Johansson M, Juneja RK, Marklund L, Andersson L. The porcine intestinal receptor for *Escherichia coli* K88ab, K88ac: regional localization on chromosome 13 and influence of IgG response to the K88 antigen. *Anim Genet.* 1995;26:237-242.

7. Python P, Jorg H, Neuenschwander S, Hagger C, Stricker C, Burgi E ,Bertschinger HU, Stranzinger G, Vogeli P. Fine-mapping of the intestinalreceptor locus for enterotoxigenic *Escherichia coli* F4ac on porcinechromosome 13. *Anim Genet.* 2002;33:441–447.

8. Jørgensen CB, Cirera S, Anderson SI, Archibald AL, Raudsepp T, Chowdhary B, Edfors-Lilja I, Andersson L, Fredholm M. Linkage and comparativemapping of the locus controlling susceptibility towards *E. COLI* F4ab/ac diarrhoea in pigs. *Cytogenet Genome Res.* 2003;102:157–162.

9. Jørgensen CB, Cirera S, Archibald AL, Andersson L, Fredholm M, Edfors-Lilja I. Porcine polymorphisms and methods for detecting them. Patent number WO2004048606. 2004.

10. Jensen GM, Frydendahl K, Svendsen O, Jørgensen CB, Cirera S, Fredholm M, Nielsen JP, Møller K. Experimental infection with *Escherichia coli* O149:F4ac in weaned piglets. *Vet Microbiol.* 2006;115:243–249.

11. Rasschaert K, Verdonck F, Goddeeris BM, Duchateau L, Cox E. Screening of pigs resistant to F4 enterotoxigenic *Escherichia coli* (ETEC) infection. *Vet Microbiol.* 2007;123:249–253.

12. Trevisi P, Corrent E, Messori S, Casini L, Bosi P. *Healthy newly weaned pigs require more tryptophan to maximize feed intake if they are susceptible to Escherichia coli K88. Livest Sci.* 2010;134:236–238.

13. Joller D. Comparative molecular approaches to identify host determinants mediating adhesion of E. coli F4 strains in pigs. *Dissertation ETH number 18518.* 2009.

14. Rampoldi A, Jacobsen MJ, Bertschinger HU, Joller D, Bürgi E, Vögeli P, Andersson L, Archibald AL, Fredholm M, Jørgensen CB, Neuenschwander S. The receptor locus for *Escherichia coli* F4ab/F4ac in the pig maps distal to the *MUC4-LMLN* region. *Mamm Genome.* 2010;22:122–129.

15. Corfield AP, Carroll D, Myerscough N, Probert CS. Mucins in the gastrointestinal tract in health and disease. *Front Biosci.* 2001;6:D1321–D1357.

16. Moniaux N, Escande F, Porchet N, Aubert JP, Batra SK. Structural organization and classification of the human mucin genes. *Front Biosci.* 2001;6:D1192–D1206.

17. Edfors-Lilja I, Petersson H, Gahne B. Perfomance of pigs with or without the intestinal receptor for *Escherichia coli* K88. *Anim Prod* 1986;42:381–387.

18. Yan XM, Ren J, Huang X, Zhang ZY, Ouyang J, Zeng WH, Zou ZZ, Yang SJ, Yang B, Huang LS. Comparison of production traits between pigs with and without the Escherichia coli F4

receptors in a White Duroc x Erhualian intercross F2 population. J Anim Sci. 2009;87:334–339.

19. Baker DR, Billey LO, Francis DH. Distribution of K88 *Escherichia coli*-adhesive and nonadhesive phenotypes among pigs of four breeds. *Vet Microbiol.* 1997;54:123–132.
20. Bosi P, Gremokolini C, Trevisi P, Mazzoni M, Bonilauri P, Sarli G, Casini L. Oral challenge with *E. coli* K88 as a tool to assess growth and health performance in feeding trials of weaned pigs. *Journées de la Recherche Porcine.* 2004;36: 125–132.
21. *Madec F, Bridoux N, Bounaix S, Cariolet R, Duval-Iflah Y, Hampson DJ, Jestin A. Experimental models of porcine post-weaning colibacillosis and their relationship to post-weaning diarrhoea and digestive disorders as encountered in the field. Vet Microbiol. 2000;72: 295–310.*
22. Jacobsen M, Kracht SS, Esteso G, Cirera S, Edfors I, Archibald AL, Bendixen C, Andersson L, Fredholm M, Jørgensen CB. Refined candidate region specified by haplotype sharing for *Escherichia coli* F4ab/F4ac susceptibility alleles in pigs. *Anim Genet.* 2009;41:21–25.
23. Jacobsen M, Cirera S, Joller D, Esteso G, Kracht SS, Edfors I, Bendixen C, Archibald AL, Vogeli P, Neuenschwander S, Bertschinger HU, Rampoldi A, Andersson L, Fredholm M, Jørgensen CB. Characterisation of five candidate genes within the ETEC F4ab/ac candidate region in pigs. *BMC Res Notes.* 2011;4:225.
24. Zhang B, Ren J, Yan X, Huang X, Ji H, Peng Q, Zhang Z, Huang L. Investigation of the porcine *MUC13* gene: isolation, expression, polymorphisms and strong association with susceptibility to enterotoxigenic *Escherichia coli* F4ab/ac. *Anim Genet.* 2008;39:258–266.
25. Ji H, Ren J, Yan X, Huang X, Zhang B, Zhang Z, Huang L. The porcine *MUC20* gene: molecular characterization and its association with susceptibility to enterotoxigenic *Escherichia coli* F4ab/ac. *Mol Biol Rep.* 2011;38:1593–1601.

Table 1. Mean ± standard deviation (s.d.), minimum (Min.) and maximum (Max.) values for estimated breeding values (EBVs) in the two extreme tails for the four groups of Italian heavy pigs used in this study.

Breed/Trait[1]	**Tail**[2]	**No. of pigs**	**EBV Mean ± s.d.**	**Min. EBV**	**Max. EBV**
ILW/ADG (g)	negative	200	–30.67 ± 15.41	–92.00	–8.00
	positive	200	+82.11 ± 13.93	+69.00	+132.00
ILW/BFT (mm)	negative	280	–9.40 ± 1.60	–15.40	–7.40
	positive	280	+8.00 ± 5.95	+2.50	+14.40
IL/ADG (g)	negative	100[3]	–36.17 ± 11.98	–70.00	–21.00
	positive	100[3]	+110.83 ± 10.06	+96.00	+139.00
ID/ADG (g)	negative	100[3]	–27.50 ± 13.05	–105.00	–15.00
	positive	100[3]	+91.00 ± 7.65g	+80.00	+136.00

[1]ILW = Italian Large White; IL = Italian Landrace; ID = Italian Duroc; ADG = average daily gain; BFT = backfat thickness.
[2]Extreme divergent tails selected according to the lower (negative) or higher (positive) EBVs.
[3]IL/ADG negative tail: 73 females and 27 castrated males; IL/ADG positive tail: 68 females, 32 castrated males; ID/ADG negative tail: 67 females and 33 castrated males; ID/ADG positive tail: 67 females and 33 castrated males.

Table 2. Allele frequencies of the *MUC4* g.8227C>G polymorphism in different Italian pig breeds.

Breed	**No. of pigs**	**Allele frequency**	
		C	**G**
Italian Large White[1]	541	0.507	0.493
Italian Landrace[2]	65	0.540	0.460
Italian Duroc[3]	48	0.917	0.083
Calabrese	15	0.830	0.170
Casertana	27	0.910	0.090
Cinta Senese	22	1.000	0.000
Nero Siciliano	30	0.720	0.280

[1]Two generation unrelated pigs of the selective genotyping study using backfat thickness estimated breeding value.
[2]Two generation unrelated pigs of the selective genotyping study using average daily gain estimated breeding value.
[3]Two generation unrelated pigs of the selective genotyping study using average daily gain estimated breeding value.

Table 3. Differences of allele frequencies of the *MUC4* g.8227C>G polymorphism between the two extreme and divergent tails chosen using a selective genotyping approach based on estimated breeding values for production traits in Italian heavy pigs.

Breed/trait[1]	**Tail**[2]	**No. of pigs**[3]	**Allele frequencies**		***P***[4]
			C	**G**	
ILW/ADG	negative	186	0.524	0.476	6.66E-04
	positive	184	0.399	0.601	
ILW/BFT	negative	273	0.474	0.526	0.012
	positive	268	0.550	0.450	
IL/ADG (all)[5]	negative	93	0.740	0.260	<1.0E-20
	positive	95	0.080	0.920	
IL/ADG (unr)[6]	negative	45	0.733	0.267	7.23E-12
	positive	20	0.100	0.900	
ID/ADG (all)[5]	negative	92	0.960	0.040	0.590
	positive	87	0.940	0.060	
ID/ADG (unr)[6]	negative	29	0.910	0.090	1.00
	positive	19	0.920	0.080	

[1]ILW = Italian Large White; IL = Italian Landrace; ID = Italian Duroc; ADG = average daily gain; BFT = backfat thickness.
[2]Extreme divergent tails selected according to the lower (negative) or higher (positive) EBVs.
[3]After DNA quality control of extracted genomic DNA, a few animals of the selective genotyping panels (Table 1) were excluded from genotyping or genotyping failed.
[4]Chi square or two tailed Fisher's exact tests (where appropriate) of significance of allele frequency differences between the two extreme tails of the genotyped pigs. Tests refer to the comparisons between the negative and positive tails adjacent to the *P* value.
[5]Considering all selected pigs for the two tails.
[6]Considering only two-generation unrelated pigs for the two tails.

General conclusions

GENERAL CONCLUSIONS

In this thesis we conducted experiments of SNP association with carcass traits in Italian heavy pigs using different approaches to find DNA markers which could be useful to monitor and improve carcass characteristics and performance traits in pigs devoted to the production of POD dry-cured hams. Both selective genotyping approach and genotyping in random populations were used to confirm results.

For TRIB3 gene, the *in silico* analysis, using different tools agrees with the fact that p.P49L could be a mutation of functional relevance. This mutation has an effect on fat deposition (measured as EBV for back fat thickness) in both breeds that were taken into account in this experiment: Italian Large White and Italian Duroc. This association is confirmed both using a selective genotyping approach as well as using an independent group of randomly chosen animals. First data on *TRIB3* gene expression can be the starting point to further investigate the biological role of this gene.

The PCSK1 gene was first investigated in this thesis for association with production traits in pigs. Two intronic mutations were used in the association study showing significant effects on fat deposition and carcass traits in Italian Large White and Italian Duroc breeds but in different directions for the two breeds. Of particular interest is the association of the g.5182A>T SNP with cathepsin B activity. If this association will be confirmed, it will be the first marker to be associated with this parameter that is correlated with excessive softness and other characteristics that are a problem for dry cured ham production.

For the MUC4 polymorphism investigated in this study, the analysis of allele frequencies in local breeds indirectly confirmed the relation of the “resistant” allele of F4bcR locus with a higher rusticity. This also support the low “susceptible” allele frequency in the Italian Duroc population, that is considered, to some extent, most rustic commercial breed used in dry-cured ham

production. The less rustic breed which have a high frequency of the "susceptible" allele are Italian large White and Italian Landrace. In these breed it was particular interesting to note a very strong association with the putative susceptible-linked allele with higher average daily gain. This locus is a good example of the complexity of applying marker assisted selection in pig breeding.

Selective genotyping approach is a very useful approach to reduce the number of genotyped animals maintaining a high power in detecting associations between production traits and gene markers. It implies only the selection of the extremely high and low scoring animals from the continuous distribution of a quantitative trait. However, its efficiency is maximized for the particular trait for which tails have been chosen.

The use of candidate gene approach confirm its effectiveness to detect new molecular markers associated with production traits in pigs. In all three genes investigated we reported convincing association with production traits. However, further studies should be carried out to further confirm the effects of the analysed polymorphisms in other populations. Production traits in pigs are in general controlled by a large but unknown number of quantitative trait loci; for this reason the dissection of economically important traits would require a high number of genetic markers in order to develop high density and high throughput assays for the association studies. Recently new sequencing technologies, so called "next generation sequencing" allowed tremendous improvements on marker discovery. This, because they can generate large amounts of sequence data cutting down time and costs. Based on information coming from this new technology, Illumina produced a custom chip that can analyse more that 60,000 porcine SNP. This tool will complement a candidate gene approach and might be the basic source of information to apply to evaluate the potential and limits of the genomic selection in pigs.