OULU 2014
D 1251

ACTA
UNIVERSITATIS OULUENSIS

Johanna Huusko

GENETIC BACKGROUND OF SPONTANEOUS PRETERM BIRTH AND LUNG DISEASES IN PRETERM INFANTS

STUDIES OF POTENTIAL SUSCEPTIBILITY GENES AND POLYMORPHISMS

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU,
FACULTY OF MEDICINE,
INSTITUTE OF CLINICAL MEDICINE,
DEPARTMENT OF PAEDIATRICS;
MEDICAL RESEARCH CENTER OULU;
OULU UNIVERSITY HOSPITAL

ACTA UNIVERSITATIS OULUENSIS
D Medica 1251

JOHANNA HUUSKO

GENETIC BACKGROUND OF SPONTANEOUS PRETERM BIRTH AND LUNG DISEASES IN PRETERM INFANTS

Studies of potential susceptibility genes and polymorphisms

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Health and Biosciences of the University of Oulu for public defence in Auditorium 12 of the Department of Paediatrics, on 6 June 2014, at 12 noon

UNIVERSITY OF OULU, OULU 2014

Copyright © 2014
Acta Univ. Oul. D 1251, 2014

Supervised by
Professor Mikko Hallman
Docent Ritva Saastamoinen (*ex* Haataja)
Doctor Minna Karjalainen

Reviewed by
Professor Johanna Schleutker
Docent Eero Kajantie

Opponent
Professor Tarja Laitinen

ISBN 978-952-62-0473-4 (Paperback)
ISBN 978-952-62-0474-1 (PDF)

ISSN 0355-3221 (Printed)
ISSN 1796-2234 (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2014

Huusko, Johanna, Genetic background of spontaneous preterm birth and lung diseases in preterm infants. Studies of potential susceptibility genes and polymorphisms
University of Oulu Graduate School; University of Oulu, Faculty of Medicine, Institute of Clinical Medicine, Department of Paediatrics; Medical Research Center Oulu; Oulu University Hospital
Acta Univ. Oul. D 1251, 2014
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Each year in Finland, approximately 5.7% of infants are born preterm, i.e., before 37 completed weeks of gestation. Preterm birth is a major cause of mortality and several neonatal morbidities, especially the respiratory diseases. Infants born very preterm (<32 wk) are at higher risk of developing a chronic lung disease called bronchopulmonary dysplasia (BPD). The genetic factors predisposing to spontaneous preterm birth (SPTB) and BPD are incompletely known.

The aims of this thesis project were to identify genetic factors that affect susceptibility to SPTB and BPD. Genetic case-control association studies were performed in mothers and infants of northern Finnish origin (SPTB study), or in multiple populations of very preterm infants of Finnish or European origin (BPD study). The candidate genes were selected based on their proposed roles in inflammation which is involved in both SPTB and BPD susceptibility. Additionally, the aim was to study the possible functional role of polymorphisms in the gene encoding surfactant protein B (SP-B) that have been shown previously to associate with pulmonary function.

An association between Met31Thr polymorphisms in the gene encoding SP-D *(SFTPD)* and SPTB infants was found. The other collectin genes that were studied, encoding SP-A and mannose-binding lectin, did not associate with SPTB in mothers or infants.

An intronic polymorphism in the gene encoding Kit ligand *(KITLG)* was associated with the risk of BPD in the northern Finnish and in the combined population that originated from Finland, Canada and Hungary. The role of *KITLG* in BPD was further supported by biomarker data, which showed higher concentrations of Kit ligand at the time of birth in infants that later developed BPD. The genes encoding interleukin 6 (IL-6), its receptors, IL-10, tumor necrosis factor alpha or glucocorticoid receptor did not associate with BPD susceptibility.

Finally, a genetic variant 131Thr in the gene encoding SP-B *(SFTPB)* was associated with lower SP-B levels *in vivo* and delayed secretion *in vitro*.

To date, there is no effective method to prevent SPTB, and especially the extremely preterm infants are at an increased risk of developing serious respiratory diseases. Better understanding of the mechanisms underlying both SPTB and BPD could help in the successful prediction of risk groups as well as in the design of new preventive and treatment strategies.

Keywords: bronchopulmonary dysplasia, genetic association studies, genetic polymorphism, inflammation, Kit ligand, premature birth, respiratory disease, surfactant proteins

Huusko, Johanna, Ennenaikaisen syntymän ja keskosten keuhkosairauksien perinnöllisen taustan tutkiminen.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Lääketieteellinen tiedekunta, Kliinisen lääketieteen laitos, Lastentaudit; Medical Research Center Oulu; Oulun yliopistollinen sairaala
Acta Univ. Oul. D 1251, 2014
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Noin 5,7 % lapsista syntyy Suomessa ennenaikaisesti, eli ennen kuin raskaus on kestänyt täydet 37 viikkoa. Ennenaikainen syntymä altistaa vastasyntyneen lapsen vakaville pitkäaikaissairauksille. Erityisesti hyvin pienillä keskosilla, jotka ovat syntyneet ennen 32. raskausviikkoa, on suurempi riski sairastua vakavaan hengitysvaikeuteen eli bronkopulmonaaliseen dysplasiaan, joka tunnetaan myös nimellä BPD-tauti. Perinnölliset tekijät vaikuttavat niin spontaanin ennenaikaisen syntymän (SEAS) kuin BPD-taudinkin taustalla, mutta nämä tekijät tunnetaan huonosti.

Tässä väitöskirjatyössä pyrittiin tunnistamaan perinnöllisiä tekijöitä, jotka vaikuttavat SEAS:in ja BPD-taudin taustalla. Perinnöllisen taustan selvittämisessä ehdokasgeenien sisältämien muuntelevien kohtien esiintyvyyttä verrattiin terveiden verrokkien ja tautitapausten välillä. SEAS-tutkimuksessa tutkimusväestö koostui suomalaisista äideistä ja heidän lapsistaan. BPD-tutkimuksessa oli mukana hyvin ennenaikaisesti syntyneitä lapsia Suomesta, Kanadasta ja Unkarista. Tämän lisäksi kokeellisten tutkimusten avulla tutkittiin aiemmin keuhkosairauksiin liittyneen geenin muuntelevien kohtien osuutta sen koodaaman surfaktanttiproteiini (SP) B:n toiminnassa.

Tutkimuksissa havaittiin SP-D:tä koodaavan geenin Met31Thr-polymorfismin olevan mahdollinen riskitekijä SEAS:lle lapsilla, mutta se ei selittänyt SEAS-riskiä äideissä. SP-A:ta ja mannoosia sitovaa lektiiniä koodaavilla geeneillä ei ollut yhteyttä SEAS-riskiin.

Kit-ligandia koodaavan geenin intronissa sijaitseva polymorfismi selitti BPD-tautiriskiä pohjoissuomalaisessa sekä yhdistetyssä tutkimusväestössä. Lisäksi lapsilla, jotka myöhemmin sairastuivat BPD-tautiin, havaittiin suurempia Kit-ligandipitoisuuksia syntymähetkellä. Interleukiini 6:ta (IL-6), sen reseptoreita, IL-10:ta, tuumorinekroosifaktori-alfaa tai glukokortikoidireseptoria koodaavien geenien polymorfismien ja BPD-taudin välillä ei ollut yhteyttä.

SP-B:tä koodaavan geenin Ile131Thr-polymorfismin Thr-variaatio liittyi alhaisempaan SP-B:n pitoisuuteen lapsivedessä sekä hidastuneeseen proteiinin tuottoon kokeellisessa solumallissa.

Tulokset antavat uutta tietoa SEAS:n ja BPD-taudin perinnöllisestä taustasta. Tämä tieto voi auttaa synnytyksen käynnistymiseen sekä BPD-alttiuteen johtavien biologisten mekanismien selvittämisessä ja uusien hoitokeinojen kehittämisessä.

Asiasanat: bronkopulmonaalinen dysplasia, ennenaikainen synnytys, geenipolymorfismi, geneettiset assosiaatiotutkimukset, interleukiinit, Kit-ligandi, surfaktanttiproteiinit, vastasyntyneen hengitysvaikeus

Acknowledgements

This work was carried out at the Department of Pediatrics and Medical Research Center Oulu, University of Oulu, during the years 2009–2014.

I wish to express my deepest gratitude to my supervisor Professor Mikko Hallman for providing the opportunity to work with this interesting and highly important topic, and for the great support, continuous ideas and optimism throughout the study. I am grateful also for my other supervisors Docent Ritva Saastamoinen (*ex* Haataja) and Doctor Minna Karjalainen. Docent Saastamoinen is acknowledged for her excellent guidance since the beginning of my research work when I started as a Master's student, until today, and her expertise in the functional analyses along with the genetic studies has been invaluable. I am very grateful to Doctor Karjalainen for the excellent guidance on the genetic studies and during the writing period of this thesis. I also wish to thank her for the great company at the scientific meetings and for the invaluable support and friendship during some busy times. A key person in our research group, Professor Mika Rämet is acknowledged for his valuable comments regarding the study designs and manuscripts.

I am grateful for the reviewers of this thesis, Professor Johanna Schleutker and Docent Eero Kajantie, for providing valuable comments to improve the content of the thesis. Deborah Kaska is acknowledged for revising the English language of the thesis, and Anna Vuolteenaho is acknowledged for revising the language in the Finnish abstract. Docent Katri Pylkäs and Doctor Reetta Vuolteenaho are acknowledged for the good comments and advices during the doctoral thesis committee meetings.

Tuula Kaukola, Aino Luukkonen, Mari Mahlman, Riitta Marttila and Riitta Vikeväinen are acknowledged for collecting the study subjects. I wish to thank all the co-authors and collaborators for their contribution to the original articles included in this thesis, especially Pascal Lavoie is acknowledged for his valuable advice. Sture Andersson, Susan Guttentag, Anneli Kari, Stephen Kingsmore, Liisa Lehtonen, Reija Paananen, Ulla Petäjä-Repo, Ulla Sankilampi, Miklós Szabó, Outi Tammela, Saija Taponen, Gergely Toldi, and Anu Tuohimaa are all acknowledged. Especially Saija Taponen is acknowledged for her important contribution to the *in vitro* studies.

I wish to thank all the current and former members of our research group for the many great moments during these years. New members of the group, Antti Haapalainen and Anu Pasanen, are acknowledged for bringing interesting

discussion topics to the coffee breaks. Johanna Ulvila is thanked for the friendship that has lasted for many years. I warmly thank Marjatta Paloheimo for the friendly and invaluable help in many practical issues during all these years. I want to express my gratitude to Maarit Haarala and Mirkka Ovaska for providing me a great help in the laboratory work. I also wish to thank you for many great moments at work and free-time, and for your friendship during these years.

I am grateful to laboratory manager Tiina Hurskainen for help in practical issues in CRC, and I wish to thank all the members in the neighboring groups in CRC for the friendly atmosphere and for many joyful moments during the coffee breaks. Especially my current roommate Elina Malo is acknowledged for answering my numerous questions during the writing process of this thesis. I wish to thank Antti Viklund who has always been a great help in computer related issues. Risto Bloigu is acknowledged for the help in statistical analyses.

I wish to thank my old and new friends for adding fun and joy to my life, and for the many memorable moments. Without you this journey would have been dull, and especially Jenni Häli, Sinikka Kanniainen, Saara Keränen and Aino Odedeyi are thanked for long-lasting friendships.

I wish to express my love and gratitude to my family; my parents Arja and Kalevi Huusko, my brothers Petri and Jarkko, and sister Miia, as well as their families, for the love and support in every stage of my life. Tuija is especially thanked for the help during my postgraduate studies. I am grateful to my Scottish family Karen, Jim, Sandie, Craig, Nathan, and Ronnie, for their love and support.

I own my deepest love and gratitude to my partner Ross Gillespie for his continuous love, support and encouragement during these years. I also wish to thank you for the many joyful moments and the great Scottish humor that you have brought into my life.

This work was financially supported by Emil Aaltonen Foundation, Olvi Foundation, and the Alma och K.A Snellman Foundation.

Oulu, April 2014 Johanna Huusko

Abbreviations

Δi4	Intron four length variation in *SFTPB* gene
AEI	Allelic expression imbalance
Asp	Aspartic acid
ATII	Alveolar type II
BPD	Bronchopulmonary dysplasia
cDNA	Complementary DNA
CHO	Chinese hamster ovary
CI	Confidence interval
CRD	Carbohydrate-recognition domain
ER	Endoplasmic reticulum
FU	Fluorescence unit
GA	Gestational age
Gln	Glutamine
Gly	Glycine
GR	Glucocorticoid receptor
GWAS	Genome-wide association study
IL	Interleukin
IL-1β	Interleukin 1 beta
IL6	Interleukin 6 gene
IL6R	Interleukin 6 receptor alpha gene
IL6ST	Interleukin 6 signal transducer gene (i.e., glycoprotein 130 gene)
IL10	Interleukin 10 gene
Ile	Isoleucine
IUGR	Intrauterine growth restriction
KITLG	Kit ligand gene
LD	Linkage disequilibrium
Leu	Leucine
Lys	Lysine
MAF	Minor allele frequency
MBL	Mannose-binding lectin
MBL2	Mannose-binding lectin gene
Met	Methionine
mRNA	Messenger RNA
NR3C1	Glucocorticoid receptor gene
OR	Odds ratio

PCR	Polymerase chain reaction
PMA	Postmenstrual age
PPROM	Preterm premature rupture of membranes
PROM	Premature rupture of membranes
ProSP-B	Precursor of surfactant protein B
RDS	Respiratory distress syndrome
RFLP	Restriction fragment length polymorphism
ROC	Receiver operating characteristic
RSV	Respiratory syncytial virus
SFTPA1	Surfactant protein A1 gene
SFTPA2	Surfactant protein A2 gene
SFTPB	Surfactant protein B gene
SFTPD	Surfactant protein D gene
SNP	Single nucleotide polymorphism
SP	Surfactant protein
SPTB	Spontaneous preterm birth
Thr	Threonine
TNF-α	Tumor necrosis factor alpha
TNFA	Tumor necrosis factor alpha gene
tSNP	Tagging single nucleotide polymorphism
Val	Valine

List of original articles

This thesis is based on the following articles, which are referred to in the text by their Roman numbers.

I Karjalainen MK, Huusko JM, Tuohimaa A, Luukkonen A, Hallman M* & Haataja R* (2012) A study of collectin genes in spontaneous preterm birth reveals an association with a common surfactant protein D gene polymorphism. Pediatr Res 71: 93–99.

II Huusko JM, Mahlman M, Karjalainen MK, Haataja R, Marttila R, Kaukola T, Toldi G, Szabó M, Kingsmore SF, Rämet M, Lavoie PM & Hallman M on behalf of Gen-BPD Study Group (2014) Polymorphisms of the gene encoding Kit ligand are associated with bronchopulmonary dysplasia. Pediatr Pulmonol. DOI: 10.1002/ppul.23018.

III Huusko JM, Karjalainen MK, Mahlman M, Haataja R, Kari MA, Andersson S, Toldi G, Tammela O, Rämet M, Lavoie PM & Hallman M on behalf of Gen-BPD Study Group (2014) A study of genes encoding cytokines (*IL6*, *IL10*, *TNFA*), cytokine receptors (*IL6R*, *IL6ST*), and glucocorticoid receptor (*NR3C1*) and susceptibility to bronchopulmonary dysplasia. Manuscript.

IV Taponen S*, Huusko JM*, Petäjä-Repo UE, Paananen R, Guttentag SH, Hallman M & Haataja R (2013) Allele-specific N-glycosylation delays human surfactant protein B secretion *in vitro* and associates with decreased protein levels *in vivo*. Pediatr Res 74: 646–651.

*Equal contribution

Contents

Abstract
Tiivistelmä
Acknowledgements 7
Abbreviations 9
List of original articles 11
Contents 13
1 Introduction 15
2 Review of the literature 17
2.1 Introduction to genetics of complex disorders 17
2.2 Prematurity and its consequences 18
2.2.1 Definition, outcomes and prevention of preterm birth 19
2.2.2 Factors associated with the risk of spontaneous preterm birth (SPTB) 20
2.3 Acute defense mechanisms 24
2.3.1 Innate immunity 25
2.3.2 Pulmonary surfactant 26
2.3.3 Surfactant protein A and D, and mannose-binding lectin 27
2.3.4 Surfactant protein B 29
2.4 Respiratory disorders in preterm infants 32
2.4.1 Respiratory distress syndrome 32
2.4.2 Bronchopulmonary dysplasia 33
2.5 Studying the role of genetic factors in predisposition to preterm birth and respiratory outcomes 39
2.5.1 Methods to study genetic predisposition 39
2.5.2 Genetic studies of spontaneous preterm birth 43
2.5.3 Genetic studies of bronchopulmonary dysplasia 44
2.5.4 Discovering the functional role of genetic variation in disease predisposition 45
3 Aims of the study 49
4 Materials and methods 51
4.1 Ethical considerations 51
4.2 Human sample collection and study populations 51
4.2.1 Study I 51
4.2.2 Studies II and III 53
4.2.3 Study IV 55
4.3 DNA sample preparation 55

4.4 Selection of polymorphisms and genotyping 56
4.4.1 Study I 56
4.4.2 Studies II and III 57
4.4.3 Study IV 57
4.5 Analysis of Kit ligand concentrations in umbilical cord blood specimens (II) 58
4.6 Protein quantitation, secretion kinetics and allelic imbalance assay (IV) 58
4.6.1 Protein quantitation from amniotic fluid samples 59
4.6.2 Generation of stably transfected proSP-B–expressing CHO cells and proSP-B secretion kinetics 59
4.6.3 Allelic expression imbalance assay 60
4.7 Statistical analysis 60
5 Results **63**
5.1 Association of *SFTPD* Met31Thr polymorphism and SPTB (I) 63
5.2 Association of *KITLG* and bronchopulmonary dysplasia (II) 63
5.3 No association of *IL6*, *IL6R*, *IL6ST*, *IL10*, *TNFA* and *NR3C1* genes with bronchopulmonary dysplasia (III) 66
5.4 Studies of *SFTPB* polymorphisms (IV) 66
5.4.1 Association of *SFTPB* Ile131Thr and SP-B levels *in vivo* 66
5.4.2 Secretion kinetics and allelic expression imbalance assay 67
6 Discussion **69**
6.1 Main findings 69
6.1.1 Role of *SFTPD* Met31Thr polymorphism in SPTB (I) 69
6.1.2 *KITLG* as a novel candidate for bronchopulmonary dysplasia susceptibility (II) 70
6.1.3 Role of *SFTPB* 131Thr in SP-B levels and secretion (IV) 72
6.2 Methodological considerations 73
6.2.1 Study populations 73
6.2.2 Candidate gene study as a method to find disease predisposing variants 74
6.2.3 Choice of protein quantitation method and use of CHO cells (IV) 75
6.3 Significance of the findings and future perspectives 76
7 Conclusions **79**
References **81**
Original articles **95**

1 Introduction

Preterm birth is one of the major causes of infant mortality and morbidity in the neonatal period and in later infancy (Goldenberg *et al.* 2008). Each year, approximately 5.7% of infants in Finland are born preterm, i.e., prior to 37 completed weeks of pregnancy (THL 2013). According to current understanding, both maternal and fetal factors may influence the risk of spontaneous preterm birth (SPTB). These risk factors include environmental (pregnancy related) risk factors, as well as a genetic predisposition (Plunkett & Muglia 2008). However, despite numerous genetic studies, knowledge about the genetic background predisposing to preterm birth still remains incomplete.

Preterm birth is the leading cause of neonatal deaths in high-income countries, accounting for approximately one fifth of total childhood deaths (Liu *et al.* 2012). Additionally, infants born preterm are at increased risk of short and long term consequences, most commonly neurodevelopmental and respiratory defects such as respiratory distress syndrome (RDS) and the chronic lung disease bronchopulmonary dysplasia (BPD) (Fanaroff *et al.* 2007). The incidence of mortality and degree of morbidity increase as the gestational age (GA) decreases; the most vulnerable are the extremely preterm (GA < 28 wk) and very preterm (GA < 32 wk) infants (Saigal & Doyle 2008).

BPD is the major cause of respiratory morbidity among the very preterm infants. However, improved neonatal care has altered the disease incidence, and now, more immature infants survive, although they are at increased risk of developing BPD (Bhandari & Bhandari 2009, Fanaroff *et al.* 2007). Twin studies indicate a strong genetic influence, with genetic factors accounting for up to 79% of the variance in liability for moderate-to-severe BPD (Lavoie *et al.* 2008). Despite numerous candidate gene studies and two recent genome-wide association studies (GWAS), the actual set of genes and specific pathways associated with the predisposition to BPD are not well known (Hadchouel *et al.* 2011, Lavoie & Dube 2010, Shaw & O'Brodovich 2013, Wang *et al.* 2013).

In this work, a candidate gene approach was applied to study variants within the genes that potentially predispose to spontaneous preterm birth and to BPD. In addition, a functional approach was applied to study a pulmonary disease-associated genetic variant in the gene encoding surfactant protein B (SP-B) in more detail, and to clarify its role in the respiratory outcome.

2 Review of the literature

2.1 Introduction to genetics of complex disorders

Complex disorders in humans (such as type 2 diabetes) are multifactorial in nature; both environmental and genetic factors influence the overall susceptibility to the disease phenotype. Many of these complex disorders cluster in families, which provides evidence of an important genetic component. With the help of family and twin studies, the heritability (h^2), i.e., the additive genetic factors that explain a fraction of the variation in the phenotype, can be estimated; h^2 is usually between 30–50% for common disorders or quantitative traits. Often, the specific pattern of inheritance is not a simple Mendelian inheritance where a single gene affects the phenotype. Complex disorders are affected by multiple common and rare genetic variants that have incomplete penetrance; the effect of one variant alone is not sufficient to cause the phenotype. The genetic contribution in complex phenotypes is most often a cumulative effect of a large number of common and rare variants; the common variants are likely to have only a low or modest effect on their own. Evolutionary selection pressure has also been lower for common disorders than for rare single-gene diseases that may even cause early death or decreased reproduction. (Hirschhorn 2005, Robinson *et al.* 2014)

The most common class of genetic variation is a substitution of one nucleotide in the genome, called a single nucleotide polymorphism (SNP), which can either be located within the protein coding region (either altering or not altering the amino acid composition) or in non-coding regions such as introns. In addition, insertions or deletions of one or more nucleotides, copy number variations, or large structural variations such as duplications or rearrangements of a larger segment of DNA are found frequently within the genome (Frazer *et al.* 2009, Marian 2012). SNPs located in the same genomic region are often correlated with each other, i.e., are in linkage disequilibrium (LD). The LD structure varies among the genomic areas and between different populations (Frazer *et al.* 2009). Genetic variants are classified as common, uncommon (i.e., low-frequency variants) or rare when they occur with a minor allele frequency (MAF) of > 5%, 1–5% or < 1% in the population, respectively. For complex traits, the causal genetic variants are often common but have less deleterious effects on survival and reproduction, i.e., these variants are not so deleterious that they are eliminated by natural selection (Hirschhorn 2005). According to recent reports,

rare variants are also thought to play a role in complex disorders (Panoutsopoulou *et al.* 2013). Although evolutionary selection removes some of the deleterious rare variants, new variants arise constantly. The specific methods used to discover genes involved in complex disorder traits are discussed in chapter 2.5.1.

Despite the vast effort and thousands of genetic studies, current knowledge of the genetic contribution to many complex traits remains limited. However, remarkable progress in understanding the molecular pathways that underlie complex disorders has been attributed with genetic studies, which can further provide intervention or drug targets for treatment. With the help of genetics, disease susceptibility groups and clinical subphenotypes will eventually be identified, which may aid in the development of tailored personalized treatments (Frazer *et al.* 2009). Examples of complex disorders and their genetic backgrounds; SPTB and serious pulmonary diseases affecting preterm infants, are discussed in this thesis.

2.2 Prematurity and its consequences

Each year in Finland, approximately 3,400 infants are born prematurely which is ~5.7% of all live births, and this rate has remained relatively constant during the last two decades (THL 2013). The incidence of preterm birth is approximately 8.6% in high-income countries and 11.1% worldwide; ranging from 5% in several European countries to 18% in some African countries (Blencowe *et al.* 2012). Birth is classified as preterm when it occurs before 37 completed weeks of gestation. Preterm birth is the leading cause of perinatal mortality and morbidity; it accounts for up to 75% of perinatal deaths (i.e., death before birth or within 7 days after birth) and 50% of long-term disabilities (Goldenberg *et al.* 2008). Preterm birth and its complications are one of the leading causes of neonatal deaths (i.e., death within 28 days after birth), and account for approximately one million annual neonatal deaths worldwide (Blencowe *et al.* 2012, Liu *et al.* 2012). The majority of preterm births occur within 34 to 36 weeks of gestation, and approximately 15% of preterm infants are born with less than 32 weeks of gestation in high-income countries (Blencowe *et al.* 2012). Preterm birth can be divided into two clinical subtypes: indicated, which results from an elective induction or Caesarean section delivery due to maternal or fetal indications (e.g. preeclampsia, intrauterine growth restriction [IUGR]), or spontaneous (Goldenberg *et al.* 2008).

2.2.1 Definition, outcomes and prevention of preterm birth

In spontaneous preterm births, labor (as indicated by regular uterine contraction and changes in the cervix before 37 weeks of gestation) starts with intact fetal membranes or after a spontaneous premature rupture of the fetal membranes (PROM). Approximately 40–50% of all preterm births are due to a labor with intact membranes and one fourth of them start with PROM (Goldenberg *et al.* 2008). The factors influencing the onset of preterm labor are not well known, but both maternal and fetal roles in the timing of parturition are proposed (Mendelson 2009).

Preterm infants (especially those that are born very preterm) are more susceptible to various short and long term neurodevelopmental, behavioral and respiratory defects than infants born at term (Doyle & Anderson 2010, Saigal & Doyle 2008). An example of a poor neurodevelopmental outcome is cerebral palsy, a neuromuscular defect that is manifested as spasticity or dystonia (Allen 2008, Hagberg *et al.* 2001). Furthermore, defects in vision (e.g. retinopathy of prematurity) and hearing, mental retardation, and other chronic problems with cognitive skills and behavior (e.g. learning and educational skills, attention deficit) also tend to be most common in extremely preterm infants (Allen 2008, Doyle & Anderson 2010, Saigal & Doyle 2008). Although most of the preterm infants survive without serious health problems, very preterm infants have a higher incidence of risk factors (e.g. high blood pressure, impaired glucose tolerance and lower physical activity), which may predispose to cardiometabolic disorders in adulthood (Kajantie & Hovi 2014). In addition, very preterm infants might have lower educational status and minor deficits in social interactions in adulthood (Moster *et al.* 2008). Furthermore, infants born preterm are particularly prone to respiratory disorders due to lung immaturity and are susceptible to lung damage. Preterm infants are at high risk of developing respiratory distress syndrome shortly after birth (Hallman & Saarela 2012), and very preterm infants are more likely to be later diagnosed with chronic respiratory dysfunction called bronchopulmonary dysplasia (Jobe & Bancalari 2001). RDS and BPD are discussed in detail in chapters 2.4.1 and 2.4.2, respectively. Although late preterm infants (GA 34–36 weeks) are considered relatively mature and are less affected by the most serious consequences, the proportion of these infants among all preterm infants is large, and they have an increased risk of respiratory morbidities compared to term born infants (Colin *et al.* 2010, Loftin *et al.* 2010).

Currently, no effective method exists to prevent preterm delivery. The length of the cervix and a fetal fibronectin test have been proposed to predict preterm birth (Goldenberg *et al.* 2005), but biomarkers that effectively predict SPTB and are useful in clinical practice remain to be identified (Kacerovsky *et al.* 2013). Specific interventions to reduce the risk of preterm delivery may include elimination of risk factors (e.g. smoking). Screening for infections and treating low-risk mothers with antibiotics is not beneficial. Instead, monitoring and treating women with known risk factors, such as previous preterm deliveries or spontaneous abortions together with short cervix and a positive fetal fibronectin test, might prove to be efficient to reduce the risk (Iams *et al.* 1998, Iams *et al.* 2008). Progesterone administration has been reported to increase the duration of pregnancy in very high-risk cases (i.e., short cervix in early pregnancy). Treatment with tocolytic agents may delay preterm birth for 24–48 h, which gives time to initiate an antenatal glucocorticoid treatment that is used to improve the postnatal outcome of the preterm fetus (Iams *et al.* 2008). The predisposing factors vary during the pregnancy, and thus defining a biomarker that predicts mothers at risk is difficult (Goldenberg *et al.* 2005). Genetic factors could provide more efficient predictors of risk because they represent markers that are stable throughout pregnancy (Plunkett & Muglia 2008).

2.2.2 Factors associated with the risk of spontaneous preterm birth (SPTB)

Normal pregnancy requires a complex interplay of fetal and maternal factors that maintain uterine quiescence, as well as actively inhibit the initiation of uterine contractions, until the optimal time for the onset of parturition occurs. In human pregnancy, maternal and placental progesterone levels are maintained high throughout the pregnancy. It has been suggested that inactivation or suppression of the progesterone receptor may be involved in the onset of parturition. In addition, the onset of parturition is preceded by the maturation of the fetal hypothalamic-pituitary-adrenal-axis that increases the levels of corticotrophin-releasing hormone. This leads to increased production of cortisol from the fetal adrenal gland near term (Mendelson 2009). Glucocorticoids promote surfactant synthesis, which promotes fetal lung maturity and increases the surfactant secretion required for air-breathing (Bolt *et al.* 2001). At term, an elevated level of cortisol may also influence the onset of labor by promoting the production of prostaglandins, placental estrogen and cervical connective tissue matrix

breakdown, which leads to stimulation of uterine contractility and eventually to the onset of labor. In term pregnancies, it is proposed that mechanical stretch caused by fetal growth together with fetal hormonal signaling initiate the labor–producing pathways; these include macrophage migration to the uterus, breakdown of the cervical matrix and generation of inflammatory cytokines (Mendelson 2009).

Preterm labor and preterm birth can be the result of complex interactions between several environmental and genetic risk factors (Fig. 1). Several maternal risk factors are associated with preterm birth such as psychological and social stress, heavy smoking or use of drugs and alcohol, chronic medical disorders, low socioeconomic or educational status, young age of the mother, too low or high body mass index, as well as ethnicity. These environmental exposures account for most of the variation in preterm birth prevalence worldwide and contribute significantly to the preterm birth risk in individuals of African ancestry (York *et al.* 2010, York *et al.* 2013). Furthermore, excessive intrauterine mass, especially in multiple pregnancies, predisposes strongly to preterm labor; twin pregnancies have a 5- to 8-fold risk of preterm birth (Goldenberg *et al.* 2008). Additionally, a short pregnancy interval associates with increased risk of preterm birth (Conde-Agudelo *et al.* 2006), or placental dysfunction (Conde-Agudelo *et al.* 2007), whereas a long interval increases the risk of preeclampsia (Conde-Agudelo *et al.* 2007).

Inflammation affecting the placenta and fetal membranes (i.e., chorioamnionitis), is more common in preterm compared to term births. Chorioamnionitis is especially related to cases of preterm PROM (PPROM) (French & McGregor 1996). The lower the gestational age, the more likely the preterm birth has been affected by chorioamnionitis. The cause of inflammation is thought to be an infection that is mostly asymptomatic but may cause symptoms (e.g. fever, abdominal pain) in clinical chorioamnionitis. An inflammatory response is thought to be a common immediate cause of preterm labor (Romero *et al.* 2006).

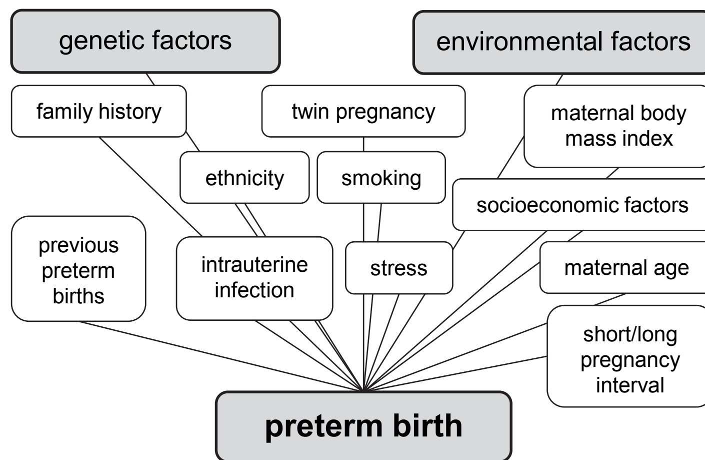

Fig. 1. The major risk factors predisposing to preterm birth.

Infection and inflammation in preterm birth

Infection and inflammation are associated with preterm birth, as well as fetal injury (Goldenberg *et al.* 2000, Romero *et al.* 2006). The association of intrauterine infection with the risk of preterm delivery is GA dependent; most often the intrauterine infection is observed in pregnancies that result in very preterm delivery (GA < 30 wk), whereas infections are less commonly associated with late preterm deliveries (Goldenberg *et al.* 2000). Animal studies have shown that intrauterine infection or administration of microbial products to pregnant animals results in the onset of preterm labor (Agrawal & Hirsch 2012).

Preterm deliveries are mostly associated with ascending infection or inflammation from the cervix to the decidual space and eventually into the fetal membranes and to the amniotic cavity by crossing through the chorioamniotic membranes. Sometimes, ongoing infection can cross from mother to infant through the placenta or *via* migration from the abdominal cavity through the fallopian tubes, or by inadvertent pathogen transfer into the amniotic cavity during amniocentesis (Goldenberg *et al.* 2000, Romero *et al.* 2006). The most commonly detected microorganism from the amniotic fluid is *Ureaplasma urealyticum*, which is most commonly associated with extremely preterm delivery (GA ~24 wk) (Goldenberg *et al.* 2000).

Although bacteria are often present with preterm labor with intact membranes and PROM, they are also detected in the membranes from elective term deliveries (Romero *et al.* 2006). Many factors protect the uterus and the fetus from microorganisms and from the inflammatory response. The first line of defense is the physical barrier formed by the cervix, as well as the decidual and fetal membranes which produce antimicrobial peptides such as surfactant proteins A and D (SP-A and SP-D) that enhance killing and clearance of the bacteria *via* macrophage phagocytosis. Innate immunity is also an early defense mechanism, which recognizes the presence of microorganisms, prevents tissue invasion, and finally triggers the host response to limit proliferation but also to adjust the magnitude of inflammation (Romero *et al.* 2006). Inflammation maintains tissue homeostasis, but dysregulation of the inflammatory response may lead to an unfavorable outcome (Speer 2006). Induction of the inflammatory cascade by receptors that recognize patterns of microbial membranes (e.g. Toll-like receptors) results in activation of inflammatory transcription factors (e.g. nuclear factor-κB and activator protein-1) that induce inflammatory cytokines, migration of inflammatory cells, and further activates both proteases and prostaglandin synthesis. This leads to cervical ripening and increased uterine contractility (Mendelson 2009).

Genetic predisposition to preterm birth

There is considerable individual and family based evidence that inherited factors together with environmental factors play important roles in the predisposition to preterm birth. Preterm birth is often seen to aggregate within families. A genetic contribution is also indicated by the fact that different ethnic groups show different incidences of preterm birth, even after the potential confounding effects (e.g. socioeconomic factors) have been taken into account (Plunkett & Muglia 2008).

Pregnancy history is an important predictor of preterm birth; mothers who have delivered preterm are at an increased risk of delivering preterm in a subsequent pregnancy (Boyd *et al.* 2009). Mothers with previous spontaneous preterm deliveries have approximately a 3-fold increased risk for both spontaneous and indicated preterm deliveries in the subsequent pregnancy compared to mothers with previous uncomplicated term deliveries (Ananth *et al.* 2006, Mercer *et al.* 1999). Additionally, mothers with indicated preterm delivery at low gestational ages are at increased risk of subsequent preterm delivery due to

either indicated or spontaneous onset (Ananth *et al.* 2006). Recurrent preterm delivery tends to occur around the same GA as the previous pregnancy, and especially mothers delivering at lower gestational ages have the highest risk of delivering at early GA in the subsequent pregnancy (Ananth *et al.* 2006, Mercer *et al.* 1999).

Furthermore, mothers themselves who have been born preterm have an increased risk of delivering preterm and this risk increases if the mother was born very preterm (Porter *et al.* 1997, Wilcox *et al.* 2008). In addition, women whose mothers, older sisters or half-sisters of a shared mother have delivered preterm also have increased risk of preterm delivery (Boyd *et al.* 2009, Plunkett *et al.* 2008, Winkvist *et al.* 1998) and PROM (Plunkett *et al.* 2008). In contrast, the paternal contribution to preterm birth is thought to be relatively small; fathers born preterm do not have an increased risk of having a preterm child (Wilcox *et al.* 2008), and a father's family history of preterm birth does not increase the risk of subsequent preterm birth (Boyd *et al.* 2009).

Although determining the exact genetic effect separately from shared or individual environmental factors is difficult, previous twin studies have estimated the heritability of gestational length or preterm birth to be approximately 31–36% (Clausson *et al.* 2000). The individual birth timing is relatively constant across pregnancies and maternal genetic factors have been estimated to account for 30–40% of the variation in the birth timing, when gestation length is considered as a quantitative trait (Kistka *et al.* 2008, Plunkett *et al.* 2009). Both maternal (14%) and fetal (11%) genetic factors contribute to the variation in gestation length together with the shared environment of the full-siblings (Lunde *et al.* 2007). It is therefore suggested that the maternal genome or maternally inherited (with less effect of paternally inherited) genes acting in the fetus contribute to the birth timing (Plunkett *et al.* 2009). Moreover, the etiology of preterm birth is complex because the etiology may differ between late and very preterm births. It is noteworthy that most of the studies mentioned above did not separate the preterm births occurring due to spontaneous onset from the medically indicated deliveries.

2.3 Acute defense mechanisms

The candidate genes that were investigated in this thesis are all involved in host defense. This selection of candidate genes was based on experimental and translational studies of the phenotypes of interest. In the following review,

relevant features of the innate immunity and the surfactant systems are briefly explained.

2.3.1 Innate immunity

The innate immune system is the first line of defense; it responds immediately to pathogens in a non-specific manner and is present at all times, but lacks memory and therefore does not improve over time despite repeated exposure to a pathogen (Haagsman *et al.* 2008, Melville & Moss 2013). Innate immunity is critical for newborn infants who have not yet developed a more sophisticated adaptive immune system that offers long lasting and protective immunity to the host (Melville & Moss 2013). Innate immunity not only facilitates a rapid killing and clearance of pathogens, but also has an important role in fine tuning the response to inflammation. It is sufficient to respond to an infection or injury but not so excessive that it causes damage in the inflamed tissue (Haagsman *et al.* 2008).

The innate immune system of preterm infants is more immature than that of term born infants, as is the adaptive immune system that develops later. Therefore, the response to infection in preterm infants is reduced and they often have deficient numbers of inflammatory cells (e.g. neutrophils and monocytes). Neutrophils and monocytes (that differentiate into macrophages at the target tissue) are in the first line of defense against pathogens. They release antimicrobial proteins and peptides that bind to the microorganisms and destroy them (Melville & Moss 2013). These cells are recruited from the vasculature into the target tissue with the help of several molecules (selectins, adhesion molecules). After capture from the blood flow, they start to roll and adhere to the vessel walls where they transmigrate into the surrounding tissue and finally to the site of infection or injury as guided by the chemokines (Nussbaum & Sperandio 2011). At the target site, the inflammatory cells contribute to pathogen clearance by phagocytosis. In addition, macrophages secrete cytokines and chemokines that further activate other immune cells (e.g. adaptive immune B and T cells). Due to a reduction of pathogen killing and clearance, as well as the production of cytokines, preterm infants are at increased risk of infections. Moreover, intrauterine inflammation causes premature activation of the immune system and cytokine production, which could lead to immune tolerance reducing the postnatal immune function (Melville & Moss 2013).

2.3.2 Pulmonary surfactant

The postnatal gas exchange takes place in alveoli; structures that provide a large surface area and a very short diffusion pathway between the air and blood (Creuwels *et al.* 1997). In addition, it is vital to have functioning defense mechanisms against inhaled pathogens and a well-functioning system to reduce surface tension at the air-liquid interface of the alveoli. This is especially vital during the transition from the fluid-filled to gas-filled lung at birth (Creuwels *et al.* 1997, Orgeig *et al.* 2011, Whitsett 2010).

The alveolar epithelium consists of type I and type II epithelial cells; type I cells form the alveolar wall and type II cells produce surfactant, a lipid-rich complex on the inner surface of the alveoli. Alveolar type II (ATII) epithelial cells differentiate between 24–34 weeks of gestation in humans (Nkadi *et al.* 2009). Pulmonary surfactant is a complex mixture of phospholipids, neutral lipids (mainly cholesterol), and proteins; the composition is 90% lipids and 10% proteins. Surfactant proteins are divided into hydrophilic SP-A and SP-D and hydrophobic SP-B and SP-C (Rooney *et al.* 1994). The main function of surfactant is to lower the surface tension of the alveoli, maintain the lung volume and prevent alveolar collapse at the end of expiration. Additionally, surfactant has an important role in immunity; it protects against infections and inappropriate inflammatory responses (Creuwels *et al.* 1997, Johansson & Curstedt 1997). Previously, surfactant proteins were thought to be lung specific, but the proteins are expressed outside the lung as well. This suggests that the primitive role of surfactant proteins was in host-defense and that surface activity related functions evolved gradually during vertebrate evolution with the adjustment to air-breathing (Daniels & Orgeig 2003, Orgeig *et al.* 2011).

Surfactant is constantly synthesized, packaged and secreted by ATII cells and recycled back to the ATII cells from the alveolar space (Hawgood & Clements 1990). Surfactant is present in different structural forms; as lamellar bodies, tubular myelin and a monolayer at the air-liquid interphase. The newly synthesized surfactant components are transported from the endoplasmic reticulum (ER) (lipids), or through the ER and Golgi (proteins), to the lamellar bodies that are intracellular multilayer storage granules (Johansson & Curstedt 1997, Perez-Gil & Weaver 2010). Surfactant metabolism (synthesis, intracellular transport and storage, exocytosis and degradation) is controlled by mechanical (e.g. stretch) and neurohormonal factors (e.g. glucocorticoids). The phospholipid and protein concentrations increase with advancing GA (Orgeig *et al.* 2011, Wert

et al. 2009). Surfactant maturation occurs before birth, and the onset of air-breathing promotes surfactant secretion. Furthermore, glucocorticoids (the most studied hormones) accelerate surfactant maturation and synthesis (Orgeig *et al.* 2011). A well-functioning surfactant turnover, where inactive surfactant is removed from the alveolar space and newly synthesized or recycled active surfactant is secreted into the alveolar space, is vital (Perez-Gil & Weaver 2010). Surfactant turnover is slower in newborns (especially in preterm newborns) than in adults (Nkadi *et al.* 2009). Surfactant abnormalities often contribute to severe lung dysfunction, and altered surfactant protein content has been observed in a variety of lung diseases in both infants and adults (Nkadi *et al.* 2009, Wert *et al.* 2009, Whitsett 2010).

2.3.3 Surfactant protein A and D, and mannose-binding lectin

Collectins, a family of collagen-type lectins, include SP-A, SP-D and mannose-binding lectin (MBL). They are pattern-recognition molecules that bind to the surface of a variety of microorganisms (bacteria, viruses, yeast and fungi). Additionally, collectins bind to the surface of immune cells (e.g. macrophages) and activate them (Jack *et al.* 2001, Johansson & Curstedt 1997, Kingma & Whitsett 2006). Collectins recognize, aggregate, and opsonize the pathogen structures for further disposal. Collectins also recruit and activate phagocytic cells (e.g. neutrophils and monocytes) at the site of inflammation, thereby enhancing phagocytosis, i.e., the killing and clearance of pathogens. Activation of immune cells leads to a release of cytokines and chemokines that recruit additional inflammatory cells (Haagsman *et al.* 2008, Kingma & Whitsett 2006, Kishore *et al.* 2006, Lawson & Reid 2000). Both SP-A and SP-D have antimicrobial properties with direct microbial growth inhibition, and they also function in resistance to allergen challenge and in pulmonary hypersensitivity, as well as in the induction of helper-T cell polarization, clearance of apoptotic and necrotic cells, and control of tissue remodeling (Kishore *et al.* 2006). In addition, SP-A and SP-D have roles in the surfactant system (Johansson & Curstedt 1997, Kingma & Whitsett 2006, Kishore *et al.* 2006).

Collectins share a common primary (i.e., peptide chain) structure that consists of an N-terminal domain, collagenous domain, α-helical neck domain and carbohydrate-recognition domain (CRD) (Haagsman *et al.* 2008, Kishore *et al.* 2006). The CRD functions in pathogen recognition and binding with high affinity. The different higher order structures of collectins are controlled by the spacing of

each trimeric CRD, and this determines the efficacy of pathogen binding (Fig. 2) (Haagsman *et al.* 2008, Johansson & Curstedt 1997). In the lung, SP-A and SP-D are synthesized and secreted by ATII cells and bronchial Clara cells (Johansson & Curstedt 1997, Kishore *et al.* 2006), and MBL is synthesized in liver hepatocytes and secreted into the blood stream (Jack *et al.* 2001). The synthesis of SP-A is developmentally regulated; it is detectable in the lung only during the last third of human gestation and detectable in amniotic fluid from the 34th wk of gestation (Hawgood & Clements 1990). Increased production of SP-A by fetal lung near term may serve as an important hormonal signal for activation of an inflammatory cascade to initiate labor (Mendelson 2009).

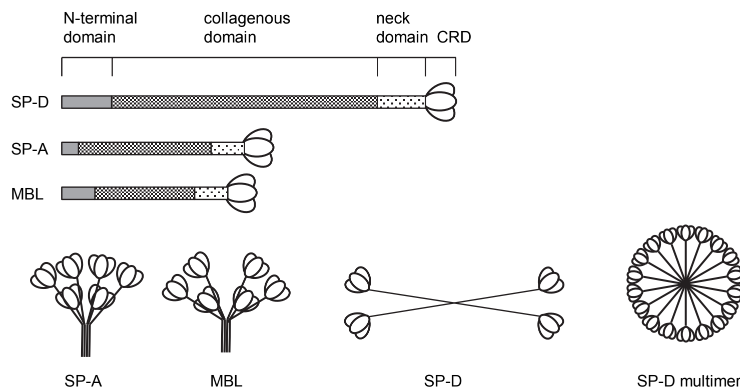

Fig. 2. Schematic representation of SP-A, SP-D and MBL primary (above) and higher order (below) structures (modified from Haagsman *et al.* 2008, Wright 2005). SP-A and MBL are octadecamers consisting of six trimeric subunits, whereas SP-D is a dodecamer consisting of four trimeric subunits; CRD, carbohydrate-recognition domain.

The genes encoding the human collectins are located in a cluster on the long arm of chromosome 10. The human SP-A is encoded by two highly homologous functional genes (~4.5 kb in size) evolved by gene duplication, *SFTPA1* and *SFTPA2*; the genes are in LD and located in opposite transcriptional orientation. Several common coding and splice variants have been identified for both of the genes. The products of both the *SFTPA1* and *SFTPA2* genes are required for the structural formation of the stable mature SP-A protein (Silveyra & Floros 2013). SP-D is encoded by a single gene (*SFTPD*, 44.9 kb), which is in partial LD with the SP-A locus. Coding and non-coding SNPs have been found within the gene (Silveyra & Floros 2012). Variations in *SFTPA1*, *SFTPA2* and *SFTPD* genes have

been associated with several pulmonary diseases (e.g. asthma, respiratory syncytial virus [RSV] infection) as well as non-pulmonary diseases in children and adults (Silveyra & Floros 2012). RDS is the most studied neonatal disease, and in Finnish preterm infants, *SFTPA1* variants and *SFTPA1*/*SFTPA2* haplotypes are associated with RDS (Haataja *et al.* 2001, Marttila *et al.* 2003a, Marttila *et al.* 2003c, Rämet *et al.* 2000). Variations of the two *SFTPA* genes and *SFTPD* are associated with RDS and BPD in other populations (Silveyra & Floros 2012).

MBL is encoded by a single gene (*MBL2*, 6.3 kb) located in chromosome 10. The major genetic determinants for MBL deficiency are the nonsynonymous structural variants in exon 1 at codons 52, 54, and 57, which generate the minor alleles named as D, B, and C, respectively. The wild-type alleles are designated A. In addition, promoter region polymorphisms (H/L, X/Y and P/Q), which form four major haplotypes (LXP, LYP, LYQ, HYP) are common. The HYP haplotype is associated with the highest serum MBL levels, whereas the LXP haplotype is associated with the lowest. Together, these structural and promoter region variants (in LD) form a total of seven common haplotypes HYPA, LYQA, LYPA, LXPA, LYPB, LYQC, and HYPD showing different frequencies in different populations (Madsen *et al.* 1995, Madsen *et al.* 1998). MBL deficiency is associated with overall immunodeficiency, and the protein level might modulate the severity of the disease (Jack *et al.* 2001). Previously, *MBL2* variants have been associated with preterm birth in mothers (Annells *et al.* 2004) and in infants (Bodamer *et al.* 2006). Additionally, *MBL2* gene variations leading to low MBL concentrations in preterm infants have been associated with BPD (Hilgendorff *et al.* 2007) and an increased risk of sepsis (Dzwonek *et al.* 2008).

2.3.4 Surfactant protein B

Surfactant protein B is a small highly hydrophobic protein that constitutes only ~1% of the total weight of the pulmonary surfactant. SP-B plays a major role in surfactant function and homeostasis, and therefore it is crucial for normal respiratory structure and function (Johansson & Curstedt 1997, Pryhuber 1998). SP-B functions to minimize the work of breathing and to maintain the alveolar integrity in the lung during the rapid respiratory cycles of inhalation and exhalation. During expiration, the surface area of the alveoli decreases. This increases the surface pressure of the monolayer, whereas SP-B together with phospholipids contributes to lowering this surface tension at the air-liquid interface and prevents the alveoli from collapsing. Additionally, SP-B accelerates

the absorption and spreading of phospholipids (that are in a gel state, thus not freely mobile to spread from the storage pools) to the surface film to maintain the integrity of the expanding alveoli during inhalation (Creuwels *et al.* 1997, Pryhuber 1998). Furthermore, SP-B is important for surfactant metabolism, especially when recycling surfactant from the alveolar space (Pryhuber 1998). Besides these surfactant related properties, SP-B has a function in protecting the lung from the adverse effects of inflammation (Epaud *et al.* 2003, Ikegami *et al.* 2005).

SP-B is critical for normal pulmonary function and gas exchange. A decrease in SP-B levels below 50% of the normal alveolar content results in abnormal lung function and may increase the risk of respiratory failure in the presence of hyperoxia (Clark *et al.* 1997, Tokieda *et al.* 1999). Further reduction of SP-B levels to less than 25% of normal is lethal (Melton *et al.* 2003). Mature SP-B can be detected in developing human lungs at 24 weeks of gestation, but the level is only 3% of the adult level (Beers *et al.* 1995). In the lung, SP-B expression is restricted to ATII cells and Clara cells, and increases in late gestation, which indicates lung maturation (Wert *et al.* 2009). In addition, glucocorticoids increase the expression of SP-B (Pryhuber 1998).

The structure and processing of SP-B

The bioactive SP-B is a ~18-kDa homodimer, which is the end product of extensive post-translational modification and proteolytic processing of a ~42-kDa preproprotein (containing 381 amino acids) to a mature 8-kDa monomer (79 amino acids) (Hawgood 1989, Pryhuber 1998). Human proSP-B contains two sites for N-linked glycosylation, Asn129 at the NH_2 terminus and Asn311 at the COOH terminus. Neither of these sites is present in the mature protein (Fig. 3). The N-terminal cleavage of the propeptide, which results in a ~25-kDa intermediate is independent of the cell type, while the COOH-terminal cleavage and final processing steps are restricted to ATII cells and are performed by cell type–specific proteases (Guttentag 2008, Pryhuber 1998). The NH_2-terminal glycosylation is specific to humans and is dependent on a nonsynonymous SNP Ile131Thr (rs1130866) in exon 4 of the *SFTPB* gene. This polymorphism results in the substitution of isoleucine with threonine creating a consensus sequence for glycosylation (Guttentag 2008, Haataja *et al.* 2000, Wang *et al.* 2003).

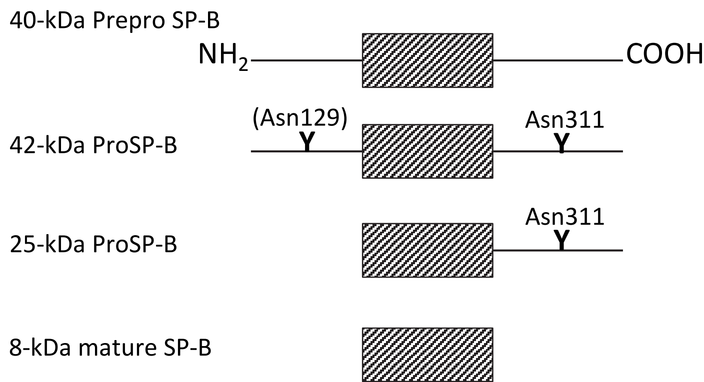

Fig. 3. The stages of processing SP-B in the lung: the N-glycosylation at Asn129 is specific to humans and dependent on an *SFTPB* Ile131Thr polymorphism, and therefore, is presented in parenthesis (modified from Hamvas 2006).

The SFTPB gene and previous associations to disease

The 11.4-kb *SFTPB* gene is located in chromosome 2, and encompasses 11 exons and 10 introns (Pryhuber 1998). More than 40 distinct *SFTPB* gene mutations have been identified that result in the total absence of SP-B or abnormal protein and phospholipid content, which may cause progressive life-threatening respiratory distress (Wert *et al.* 2009). Furthermore, targeted deletion of the *Sftpb* gene causes disturbed surfactant homeostasis, which leads to fatal respiratory failure in newborn mice (Clark *et al.* 1995, Melton *et al.* 2003).

The most extensively studied common polymorphism, the glycosylated *SFTPB* 131Thr variant, has been associated with several pulmonary diseases; RDS in infants (Haataja *et al.* 2000, Marttila *et al.* 2003b), acute RDS in adults (Lin *et al.* 2000), idiopathic pulmonary fibrosis among smokers (Selman *et al.* 2003), and pneumonia (Quasney *et al.* 2004). Other potential disease associating *SFTPB* variations that could have effect on transcriptional activity, messenger RNA (mRNA) splicing or stability (and thus could contribute to individual differences in mRNA and protein levels) include two SNPs in the promoter region, G/A−384 (rs3024791) (Thomas *et al.* 2006) and C/A−18 (rs2077079) (Steagall *et al.* 2007), and a length variation in intron 4 (Δi4) (Lin *et al.* 2005, Rova *et al.* 2004, Yang *et al.* 2013).

2.4 Respiratory disorders in preterm infants

Preterm infants are at increased risk of several respiratory morbidities. The features of the two common and serious disorders affecting preterm infants, RDS and BPD, are reviewed here. In addition, preterm infants are at risk of developing common and less serious disorders (e.g. transient tachypnea and apnea of prematurity), or rare and serious lung diseases (e.g. early-onset of pneumonia or malformations).

2.4.1 Respiratory distress syndrome

Neonatal RDS remains a common respiratory disorder among preterm infants. The incidence varies from ~5% in late preterm to ~50% among the extremely preterm infants. RDS is a result of lung immaturity together with a deficiency in the pulmonary surfactant system during the transition to air-breathing. The newer and advanced prevention and treatment methods (such as antenatal glucocorticoids, early surfactant administration and continuous distending pressure ventilation after birth) have decreased the incidence and transformed the disease from nearly always fatal respiratory failure to mostly mild-to-moderate respiratory distress (Hallman & Saarela 2012).

Besides the degree of prematurity and insufficient surfactant production, risk factors for RDS include, e.g. male sex, maternal diabetes, and the birth order of twins; the presenting twin has a lower risk of RDS. Furthermore, a genetic predisposition is evident in the risk of developing RDS together with the environmental stressors (Hallman & Haataja 2007, Hallman & Saarela 2012). Previous candidate gene studies performed to identify susceptibility genes for RDS have focused on surfactant proteins, as well as proteins associated with surfactant homeostasis. Genetic studies have shown associations to RDS and other neonatal respiratory diseases with genes encoding SP-A, SP-B, SP-C and ABCA3 (Hallman & Haataja 2006, Hallman & Haataja 2007, Karjalainen *et al.* 2012, Silveyra & Floros 2012). One of the variations associated with RDS, *SFTPB* Ile131Thr, is analyzed in detail in this study.

The introduction of surfactant and antenatal glucocorticoids into clinical practice has dramatically increased the survival of extremely preterm infants. This has exposed these preterm infants to an increased risk of BPD (Fanaroff *et al.* 2007).

2.4.2 Bronchopulmonary dysplasia

Bronchopulmonary dysplasia is a chronic lung disease of the newborn that remains a major cause of morbidity and mortality among infants born very preterm (GA < 32 wk) (Fanaroff *et al.* 2007). Northway and colleagues in 1967 were the first to describe the chronic lung disease of the newborn (i.e., the old form of BPD) that occurred in preterm infants but also in near-term and term infants that were treated with inappropriate mechanical ventilation (Northway *et al.* 1967). BPD is now known to affect only the very preterm infants and the etiology has become more complex. The clinical outcome of BPD is referred to here as the "new BPD" defined by the National Institute of Child Health and Human Development workshop in 2001 (Jobe & Bancalari 2001).

Definition, incidence, outcomes and prevention of bronchopulmonary dysplasia

BPD is defined as a disruption and dysregulation of distal lung growth and the pulmonary vasculature. Often, infants that develop BPD have mild-to-moderate respiratory distress shortly after birth but continue to require oxygen for at least several weeks (Bancalari *et al.* 2003, Laughon *et al.* 2009a). This deterioration may be triggered by certain antenatal and postnatal exposures that remain incompletely understood (Jobe 2003).

For infants born at GA < 32 wk, the diagnosis of BPD is defined as the requirement for supplemental oxygen or ventilation at 36 wk postmenstrual age (PMA). The severity of the disease for infants that require supplemental oxygen for at least 28 days is graded; as mild BPD (no supplemental O_2 or ventilation at 36 wk PMA), moderate BPD (need for supplemental O_2 but $< 30\%$ at 36 wk PMA) or severe BPD (need for supplemental $O_2 \geq 30\%$, or ventilation at 36 wk PMA) (Jobe & Bancalari 2001).

Improvements in noninvasive life-saving therapies (e.g. surfactant replacement therapy, prenatal steroids and gentler ventilator strategies) have improved the survival of extremely preterm infants. However, these infants are at increased risk of developing BPD. Therefore, the incidence of BPD has not decreased substantially and the prevalence is related to the degree of prematurity. The incidence of BPD in preterm infants varies between 5–40%; it is infrequent in infants born at gestations > 30 weeks and most common in infants born at < 28 weeks of gestation (Fanaroff *et al.* 2007).

Infants surviving with BPD have a long-term risk of lung and brain injury (Bhandari & Bhandari 2009). The effects of respiratory morbidity and decreased lung function of infants born very preterm and affected with BPD may persist for long periods. Infants with BPD have more rehospitalizations, usually due to respiratory disorders (e.g. RSV infection) during the first two years of life than preterm infants without BPD. In school age, the occurrence of wheezing and need for respiratory medications is higher in BPD infants (Bhandari & Bhandari 2009, Bhandari & McGrath-Morrow 2013, Greenough 2012). Additionally, at school age, the children born preterm and affected by mild-to-severe or moderate-to-severe BPD had seemingly lower pulmonary function (e.g. reduced airflow; percentage forced expiratory volume in 1 s; −16.2% or −18.9%, respectively) compared to children born at term. This provides evidence of impaired pulmonary function in infants with more severe forms of BPD, and the effects may persist into later childhood and adulthood (Kotecha *et al.* 2013b). In addition, the health of BPD survivors may decline faster with ageing and this might increase the risk of subsequent development of pulmonary disorders such as chronic obstructive pulmonary disease (Bhandari & McGrath-Morrow 2013, Kotecha *et al.* 2013b).

The current methods for preventing moderate-to-severe BPD among extremely preterm infants are limited. Preventive strategies include; appropriate dosage and time of antenatal steroids, minimally invasive ventilator modes, and early use of continuous distending airway pressures combined with exogenous surfactant treatment (Jobe 2011, Laughon *et al.* 2009b, Sweet *et al.* 2013). The use of supplemental glucocorticoids to treat infants at very high risk of BPD is controversial, since a balance between short-term gain and the risk of long-term side effects needs to be considered (de Benedictis & Bush 2012). In addition, early treatment of patent ductus arteriosus and infections, sufficient nutritional support with supplemental protein intake, vitamin A, or caffeine (used to treat apnea), are used to decrease the occurrence or severity of BPD (Bhandari & Bhandari 2009, Laughon *et al.* 2009b, Sweet *et al.* 2013). The most effective method to diminish BPD would be the prevention of preterm birth, especially a reduction in the number of extremely premature births (Bhandari & Bhandari 2009).

Risk factors of bronchopulmonary dysplasia

The major environmental risk factors that have been implicated to predispose to BPD include the degree of prematurity, fetal growth restriction and lung

inflammation (Bose *et al.* 2009, Fanaroff *et al.* 2007, Speer 2006). In addition, mechanical ventilation and excessive supplemental oxygen are life-saving but may cause inflammatory injury to the premature lung (Jobe 2011). Furthermore, inflammatory responses to infection or to other adverse effects may increase the susceptibility to BPD (Speer 2006). Finally, genetic factors together with environmental exposure are involved in the etiology of BPD (Lavoie & Dube 2010). The risk factors are summarized in Fig. 4.

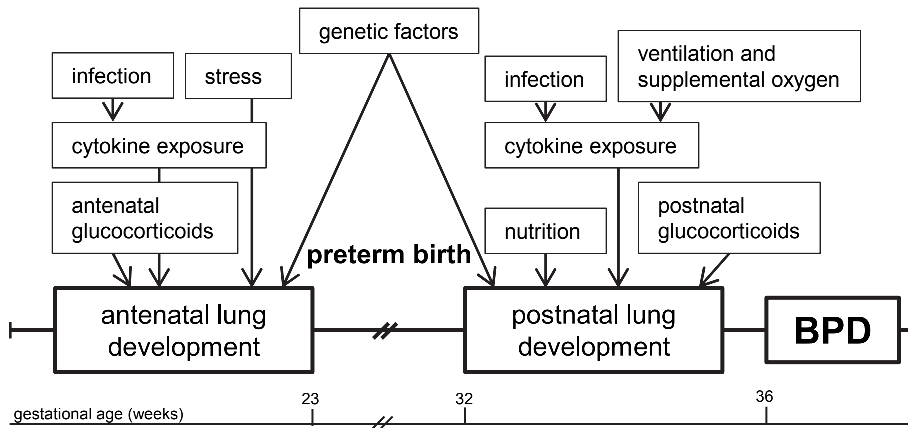

Fig. 4. The major risk factors predisposing to BPD (modified from Jobe 1999). Double slash illustrates preterm birth occurring at gestational weeks 23–32, which is a risk factor for BPD (diagnosed at 36 wk postmenstrual age).

Disturbed lung development in bronchopulmonary dysplasia

BPD occurs more often in infants born ≤ 28 weeks of gestation when the lungs of these infants are in the vulnerable early stages of lung development (Whitsett *et al.* 2004). Many antenatal (e.g. growth restriction, serious fetal inflammatory response) or postnatal (hypoxia, hyperoxia, mechanical ventilation, glucocorticoids, nutritional defects and proinflammatory mediators) insults may adversely alter the developmental program of the lung, especially inhibiting the secondary septation and formation of vascular networks (Albertine 2013, Whitsett *et al.* 2004). Impaired alveolarization and angiogenesis, showing fewer, simplified and larger alveoli and dysmorphic pulmonary vasculature, are the typical pathological findings of the new BPD (Coalson 2006).

The formation of the respiratory system begins as early as 28 days of gestation, but the branching occurs later (Morrisey & Hogan 2010). There is a complex interplay of cellular and molecular mechanisms (e.g. growth factors and transcriptional factors) that control the development, and the balance between stimulatory (e.g. glucocorticoids) and inhibitory effects (e.g. transforming growth factor-beta) is crucial at the different stages of lung morphogenesis (Maeda *et al.* 2007, Whitsett *et al.* 2004). Embryonic, pseudoglandular and canalicular periods of the lung development occur from 3 to around 26 weeks of gestation. The human lung is in the saccular stage of development from 24 to 38 weeks of gestation, when precursors of the alveoli emerge (Maeda *et al.* 2007). In addition, the distal saccules thin, pulmonary capillaries form the gas exchange region, and secondary septation further subdivides the alveolar saccules into emerging alveoli (beginning around 32 weeks of gestation); these modifications increase the surface area for gas exchange (Albertine 2013). The formation of alveoli continues beyond infancy; alveolar septation increases and continued thinning of the vascular mesenchyme creates thin alveolar-capillary structures (Maeda *et al.* 2007, Whitsett *et al.* 2004). Adult lung is comprised of approximately 300 million alveoli, enabling effective gas exchange (Whitsett *et al.* 2004).

Several transcription factors (e.g. glucocorticoid receptor [GR]), as well as SP-B and surfactant phospholipids are critical for the respiratory adaptation at birth (Maeda *et al.* 2007). Glucocorticoids have important roles in both prenatal and postnatal pulmonary development, as well as in vascular and gastrointestinal adaptation that are also critical for intact survival of the very preterm infants. In contrast, excessive antenatal and postnatal glucocorticoid administration suppresses growth and is further associated with a number of adverse effects. Furthermore, many inflammatory cells respond to glucocorticoids, and glucocorticoids have been shown to prevent the expression of cytokines such as interleukin-1 beta (IL-1β), tumor necrosis factor-alpha (TNF-α), IL-6 and IL-8, and thus glucocorticoids are often used as anti-inflammatory agents in various inflammatory conditions (Bolt *et al.* 2001).

The GR mediates the actions of endogenous and synthetic glucocorticoids. The human GR is encoded by the *NR3C1* gene located in chromosome 5 (Yang *et al.* 2012). Alternative splicing generates two highly homologous receptor isoforms: GR-α, the classic active form that is expressed in almost all tissues and cells and GR-β, an inactive minor form. There is a difference in cell and tissue type–specific regulation of GR expression and transcriptional activity. This enables target tissues to respond differently to circulating glucocorticoid

concentrations (Nicolaides *et al.* 2010, Zhou & Cidlowski 2005). Several familial and sporadic mutations of the *NR3C1* gene have been described, that might cause impairment of human GR action and lead to resistance or decreased sensitivity to glucocorticoids in tissues. In addition, common polymorphisms within *NR3C1* have been suggested to result in an altered response to glucocorticoids (Nicolaides *et al.* 2010, Yang *et al.* 2012)

Infection and inflammation in bronchopulmonary dysplasia

Infection and inflammation are some of the major contributors to the pathogenesis of BPD. Intrauterine infection and fetal exposure to inhaled pathogens from the amniotic cavity interferes with lung development. Exposure to infection leading to a fetal inflammatory response *in utero* often accelerates lung maturation processes, which protect the fetus from RDS, but may increase the risk of subsequent development of BPD. The fetal inflammatory response to intrauterine infection depends on the timing and duration of the exposure but also on the nature (severity, organisms) of the inflammatory stimuli (Kramer *et al.* 2009, Viscardi 2012). Infants exposed to intrauterine (antenatal) inflammation may be more prone to postnatal environmental exposures (e.g. mechanical ventilation, supplemental oxygen) that dysregulate or prolong the postnatal pulmonary inflammation and alter gene expression of the molecules involved in lung development (Bose *et al.* 2008, Kramer *et al.* 2009, Viscardi 2012).

Regardless of the fetal inflammatory response *in utero*, postnatal inflammatory lung injuries due to hyperoxia, mechanical ventilation or infections increase the risk of BPD. The initiation of the inflammatory cascade includes: chemokines attracting more inflammatory cells to the site, adhesion molecules facilitating migration of more inflammatory cells from blood vessels, and proteins that promote tissue damage (proinflammatory cytokines) and those that modulate the process (e.g. anti-inflammatory cytokines). Tissue damage (together with impairment in repair mechanisms) can lead to inhibition of alveolarization and vascular development, or interfere with the remodeling events during lung development (Bose *et al.* 2008, Speer 2006).

Preterm infants who develop BPD are suggested to have an imbalance between pro- and anti-inflammatory mediators, oxidant and antioxidants, as well as between proteases and their inhibitors (Bose *et al.* 2008, Speer 2006). Elevated proinflammatory cytokine levels of TNF-α, IL-1β, IL-8, IL-6 and glycoprotein 130, as well as increased numbers of inflammatory cells (neutrophils,

macrophages, mast cells, eosinophils) have been observed in cord blood, amniotic fluid or airways of infants that subsequently develop BPD (Ambalavanan *et al.* 2009, Kaukola *et al.* 2009, Speer 2006, Watterberg *et al.* 1996, Yoon *et al.* 1997, Yoon *et al.* 1999). Additionally, variable levels of anti-inflammatory cytokines IL-4 and IL-10 are observed in infants who subsequently develop BPD (Speer 2006). Recently, a genome-wide expression profiling of pulmonary mRNA, revealed a differential expression of mast cell–specific markers between infants with BPD and controls. Additionally, connective tissue mast cells were abundant in the lungs of infants dying from BPD (Bhattacharya *et al.* 2012). Mast cells are immune cells that are involved in allergy, wound healing and in antimicrobial defense (Abraham & Malaviya 1997).

Although, increased levels of several chemokines, adhesion molecules, pro- and anti-inflammatory cytokines, proteinases and growth factors have been observed in infants subsequently developing BPD (Bose *et al.* 2008), thus far, no biomarker has been identified that could accurately predict the risk of BPD development. Biomarker studies are often performed with biological samples collected from ventilated infants (tracheal aspirates or broncho alveolar lavage), which might have limitations due to the lack of non-ventilated healthy controls (Thompson & Bhandari 2008). Recently, microRNA expression profiles in infants subsequently developing BPD *vs.* non-BPD infants were studied to identify blood-based miRNAs as potential biomarkers for BPD development (Wu *et al.* 2013).

The Kit ligand (encoded by the *KITLG* gene in chromosome 12) is one of the principal growth factors for mast cells and thus could have a role in BPD. Additionally, IL-6 and its receptors, alpha (IL-6R) and glycoprotein 130, IL-10, TNF-α (encoded by the *IL6*, *IL6R*, *IL6ST*, *IL10*, *TNFA* genes in chromosomes 7, 1, 5, 1, 6, respectively), and GR could have a role in differential responses to inflammation or modify growth and thus influence the risk of BPD development. Therefore, the genes encoding these inflammatory mediators were investigated in this thesis for the predisposition to moderate-to-severe BPD.

Genetic predisposition to bronchopulmonary dysplasia

Along with environmental factors, genetic factors play an important role in predisposition to BPD. Identification of the genetic component by twin studies and genetic studies has underlined the importance of genetic factors in the predisposition to BPD (Lavoie & Dube 2010, Shaw & O'Brodovich 2013).

Previously, BPD was considered to result entirely from life-saving neonatal intensive care but is now thought to be a result of complex gene–environment interactions. Furthermore, not all infants that are born at the same GA, and thus have the same degree of immaturity of the lungs and are exposed to same environmental risk factors, develop BPD. There is only little to improve in clinical care practice in order to minimize the environmental stressors. Therefore, it is crucial to identify the molecular basis of BPD development in order to lower the incidence of the disease (Bhandari & Gruen 2006); understanding the genetic background of BPD may help in this.

Parker *et al.* in 1996 were the first to describe the familial tendency for BPD; after adjusting for potential risk factors, the BPD status of the first born twin was a significant predictor of BPD susceptibility for the second twin (Parker *et al.* 1996). Bhandari *et al.* in 2006 further estimated the heritability of BPD and showed that genetic factors account for 53% of the variance in the liability for BPD after adjusting for risk factors. The concordance ratio (observed/expected) of BPD among monozygotic twins was significantly higher than in dizygotic twins (Bhandari *et al.* 2006). Lavoie *et al.* in 2008 reported in their twin study a higher 79% heritability for moderate-to-severe BPD in infants with GA $\leq$ 30wk. They reported less heritability for the mild form of BPD, which is likely to be affected by environmental stressors rather than a genetic influence (Lavoie *et al.* 2008). To summarize, the heritability of BPD is estimated to be 50–80% among very preterm infants.

2.5 Studying the role of genetic factors in predisposition to preterm birth and respiratory outcomes

2.5.1 Methods to study genetic predisposition

The most common ways to study genes that contribute to a phenotype is by population based candidate gene or genome-wide association analyses or by family based linkage analysis. Features of these three methods are summarized in Table 1. Before selecting the specific methods to study genetic predisposition, it is extremely important to carefully define the phenotype, and also the collection of the study material should be well designed. Although the phenotype definition has been performed according to strict criteria, the spectrum of potential confounding genetic factors that affect the background can be broad.

A candidate gene association study compares the frequency differences among cases (i.e., affected individuals) and controls (i.e., healthy individuals). The study populations are non-related individuals. It is a hypothesis-based study and it may be used when there is some previous knowledge or hypotheses regarding the molecular mechanisms that underlie the disease pathophysiology. Its limitation is that the human genome contains approximately 21,000 protein-coding genes (Harrow *et al.* 2012) and without previous knowledge, finding and identifying the correct disease predisposing gene(s) and finally the causal polymorphisms and variants, is problematic. A large part of the genome contains segments of strong LD and variants in the segment are strongly correlated with each other, and thus one variation (i.e., tagging SNPs [tSNP]) captures the variation of the other SNPs in the LD block. Thus, the number of SNPs to study can be decreased by selecting tSNPs that cover overall variation among larger genomic areas with less effort. Additionally, nonsynonymous SNPs, i.e., variations that alter amino acid composition in the gene product and thus might directly affect protein function, are often selected for study. A candidate gene study is applicable also for small study populations, when the sample size does not exceed the number needed for a genome-wide approach. Studies with small sample sizes should always be replicated in separate populations. Its limitation is that the analysis is limited to the selected gene(s). (Hattersley & McCarthy 2005)

Linkage analysis is an optimal method to use when there is large family data available. No prior knowledge of a predisposing gene or its location is needed. However, its limitation is the pre-assumption of the inheritance model, which can cause bias when misinterpreted. Using linkage analysis, it is possible to study large genomic areas with a less dense set of markers (compared to population based analyses), presuming that the marker that flanks the disease gene segregates with the disease in families. It is best suited for single-gene disorders and has less power to detect modest effects of common variation on the complex disorders or quantitative traits. Population based association analyses have more power to detect a modest effect of common variants in a complex phenotype, but a denser set of markers is needed to capture overall variation in the genomic area. Linkage analysis using family data has less risk of bias caused by population stratification. The region of linkage needs to be studied more thoroughly to reveal the causative genes and variants (Hirschhorn & Daly 2005, Hirschhorn 2005).

A genome-wide association study is a hypothesis-free way to study association because there is no need for prior knowledge of the disease predisposing genes or pathways (Frazer *et al.* 2009). This method requires large

sample sizes to obtain sufficient statistical power to detect mild-to-moderate effects of common genetic variations, which often have, e.g. an additive effect on the phenotype (Ward & Kellis 2012). GWAS often provides new insights into the disease pathways by novel findings (i.e., finding genes/pathways that were not previously thought to be involved in the etiology, and the possible effect of subphenotypes) (Frazer *et al.* 2009, Ward & Kellis 2012). A challenge of GWAS is the difficulty of finding the link between a statistically significant association and its functionally (biologically) relevant role in the etiology of the disorder. Often, the associating variant (mostly non-coding) is located in a genomic region far from any gene; thus, it is likely to be regulatory rather than a coding variant or it may be in LD with the actual causal variant. Furthermore, the association involved in one population might not be observed in another population (due to differences in MAF and LD patterns, as well as other underlying factors that cause population-specific stratification) so that the results are not reproducible. GWAS does not identify rare variants and has limited power to detect gene–gene and gene–environment interactions; thus, the results only explain a fraction of the total genetic variation that contributes to the phenotype (Frazer *et al.* 2009, Ward & Kellis 2012). The number of GWASs has increased dramatically, and to date approximately 1,900 publications showing over 13,000 common SNPs that are associated with hundreds of phenotypes (divided into 17 common traits) have been published by the National Human Genome Research Institute database (http://www.genome.gov/gwastudies) (Hindorff *et al.* 2014). Despite numerous studies and associations, a large amount of the heritability of the traits remains undefined. This missing heritability can be due to phenotype associating common genetic variants with small effects as well as interactions that have yet to be discovered, but also to a lack of discovery of the rare and novel variants with moderate to high penetrance. In addition, undetected epistasis can explain the low proportion of detected heritability (Cordell 2002, Frazer *et al.* 2009).

Rare variants in common diseases can be further identified by next-generation sequencing (Ward & Kellis 2012). Next-generation sequencing is a high-throughput technology for exome or whole genome sequencing, which has the advantage of obtaining an enormous amount of data within a few weeks and at relatively low cost compared to the traditional Sanger–based sequencing approach. However, interpreting which variant is causal is difficult (Mardis 2013).

Table 1. Characteristics of typically used genetic analysis methods (summarized from Frazer *et al.* 2009, Hirschhorn & Daly 2005, Hirschhorn 2005, Mardis 2013, Marian 2012, Ward & Kellis 2012).

Feature	Candidate gene association study	Linkage analysis	Genome-wide association study	Next-generation sequencing
Study population	General population / selected cases and controls	Family members, both affected and non-affected	General population / selected cases and controls	Family members / General population / selected cases and controls
Required population size	> 100	Several affected individuals / family	> 1,000	A few or several affected / non-affected individuals
Risk of population stratification	Moderate	No risk	Moderate, can be controlled	Moderate, can be controlled
Need of prior hypothesis of potentially predisposing genes	Yes	No	No	No
Genetic variation coverage	Only selected genes	Whole genome or a set of chromosomes	Whole genome	Whole exome / genome
Number of SNPs studied	Limited number	100 – thousands	500,000 – millions	All
Power to detect association with common variants with a modest effect	Low–moderate	Low–moderate	Moderate–Good	Moderate–Good
Accuracy of the location of the associating/causative SNP	Moderate–high	Low	Moderate	Very high
Detection of rare variants	No	Yes	No	Yes

2.5.2 Genetic studies of spontaneous preterm birth

Preterm birth is a complex disorder where genetic, epigenetic and environmental factors play a role (Hallman 2012). It is likely an outcome of complex additive interactions between genes and the environment. A combination of polymorphisms of different genes in the context of a particular environmental risk factor could lead to preterm birth because a single factor is not sufficient to cause the onset of preterm delivery by itself. To further amplify the complexity of preterm birth, both maternal and fetal genetic factors are estimated to contribute to the onset of parturition and should be taken into account; often, both the mother and fetus have been studied as affected in the case-control setting. In addition, the onset of delivery can be analyzed both as a quantitative (GA as continuous variable) and a dichotomous trait (Plunkett & Muglia 2008).

In many of the previous candidate gene association studies, the focus has been on the genes that contribute to the different pathways linked to parturition, genes involved infection and inflammation (e.g. *TNFA*, *IL4*, *IL10*, *IL6*, *IL6R*, *MBL2*), response to stress and environmental toxins or diet (e.g. *PON1*, *PON2*), remodeling of connective tissue (e.g. *MMP1*, *MMP9*), factors contributing to uterine contractions (*ADRB2*) and overall placental function (e.g. *VEGF*, *F5*, *F7*) (Plunkett & Muglia 2008). More recent candidate gene studies have also shown associations with several genes, including, e.g. *CSF2*, *IL12A*, and *PTGER3* (Harmon *et al.* 2013, Ryckman *et al.* 2010). One of the more recent larger scale candidate gene studies in a Hispanic population from Chile with 190 genes and a total of 775 SNPs, showed significant associations to SPTB with fetal genes involved in inflammatory responses (*IL6R*, *IGF2*, *IL2*) (Romero *et al.* 2010b), and variations within maternal genes involved in extracellular matrix metabolism (*TIMP2*, *COL4A3*) were associated with SPTB and PPROM (Romero *et al.* 2010a, Romero *et al.* 2010b). However, previous candidate gene studies have shown inconsistent results with a lack of replication by other investigators, probably due to small sample sizes, differences in phenotype definition and exclusion criteria, as well as population stratification (mixed ethnicities) (Plunkett & Muglia 2008).

In addition to candidate gene studies, recent more comprehensive approaches to study SPTB have suggested associations with several genes involved in multiple pathways, underlining the complexity of the SPTB phenotype. Linkage analyses have identified the genes encoding insulin-like growth factor 1 receptor (*IGF1R*) and androgen receptor (*AR*) genes as potential SPTB susceptibility genes (Haataja *et al.* 2011, Karjalainen *et al.* 2012). An evolutionary based approach

was used to identify genes showing accelerated evolution in humans and implicating a gene encoding follicle stimulating hormone receptor (*FSHR*) as a susceptibility gene for SPTB (Plunkett *et al.* 2011). Additionally, a whole-exome sequencing study implicated a gene encoding complement receptor 1 (*CR1*) in SPTB susceptibility (McElroy *et al.* 2013). However, to date, no GWAS of SPTB has been published and the actual predisposing genes remain to be discovered.

2.5.3 Genetic studies of bronchopulmonary dysplasia

Most of the reported candidate gene association studies of BPD have focused on the genes involved in innate immunity, especially the genes encoding the pro- and anti-inflammatory cytokines. Additionally, genes involved in processes of lung development (vascular and lung remodeling, antioxidant defense system, cell and tissue injury or repair) and also the genes encoding lung surfactant proteins have been widely studied (Bhandari & Gruen 2006, Lavoie & Dube 2010, Shaw & O'Brodovich 2013). In addition to the genes that are assumed to have a direct role in the etiology of BPD, several genes that are involved primarily in the etiology of the known predisposing risk factors for BPD have been widely studied, including genes involved in the predisposition to preterm birth, patent ductus arteriosus and sepsis (Bokodi *et al.* 2007, Parton *et al.* 2006).

A list of previous candidate gene studies using the new definition of the BPD phenotype is shown in Table 2. Despite the numerous genetic studies performed previously, the actual set of genes and pathways involved in the etiology of BPD are not well known. Several reviews have summarized the previous candidate gene association studies, but most of those studies show weak or no association and are often not replicated (Bhandari & Gruen 2006, Bokodi *et al.* 2007, Lavoie & Dube 2010, Parton *et al.* 2006, Shaw & O'Brodovich 2013). The explanation for the lack of association or replication of the previous results is often the population heterogeneity and the sample size. Studies having an extremely small number of individuals affected with BPD, but also with a low number of controls, have only limited power to detect modest genetic effects. This also creates a chance of false-positive results, which can be diminished by adjusting the threshold for statistical significance for multiple comparisons. In addition, many of the previously reported genetic association studies have been performed using a variable definition of BPD (mild BPD included in the cases), unmatched cases and controls (e.g. preterm cases and term controls), or the study populations have shown ethnic admixture (Lavoie & Dube 2010, Shaw & O'Brodovich 2013).

Thus far, two genome-wide association studies of BPD have been published. The first GWAS suggested the *SPOCK2* gene, which encodes testican-2, as a susceptibility gene for BPD, and this association to BPD was replicated in an independent study population. The new susceptibility gene was further studied in a rat model that showed increased *Spock2* mRNA levels when rats were exposed to hyperoxia (Hadchouel *et al.* 2011). This association was not replicated in the second GWAS, nor did they identify any other susceptibility loci for BPD at the statistical significance level, but several suggestive findings were reported for further research (Wang *et al.* 2013). For the identification of rare variants in disease predisposition, next-generation sequencing of exomes from 15 severe cases of BPD is ongoing (Somaschini *et al.* 2012).

Table 2. Previously suggested pathways and genes associated with (moderate-to-severe) BPD (according to Shaw & O'Brodovich 2013).

Pathway	Associated gene[1]	References
Lung development:	*MMP16*	Hadchouel *et al.* 2008
alveolarization, vasculogenesis	*SPOCK2*	Hadchouel *et al.* 2011
	F7	Hartel *et al.* 2006
	ACE	Kazzi & Quasney 2005
	VEGF	Mailaparambil *et al.* 2010
Innate immune system &	*NFKBIA*	Ali *et al.* 2013
Inflammatory responses	*MBL2*	Hilgendorff *et al.* 2007
	TNFA	Kazzi *et al.* 2004, Mailaparambil *et al.* 2010
	TLR4,TLR5	Lavoie *et al.* 2012, Sampath *et al.* 2012
	TLR10,TLR6	Mailaparambil *et al.* 2010, Winters *et al.* 2013
	MIF	Prencipe *et al.* 2011
	HLA	Rocha *et al.* 2011

*1*genetic association studies using the new definition of BPD (need for supplemental oxygen at 36 wk PMA). Genotypic or allelic associations are reported with significance level of $p < 0.05$, whereas information about *p* value correction for multiple comparisons or whether the effect of potential risk factors (e.g. GA or birth weight) was taken into account was not always available.

2.5.4 Discovering the functional role of genetic variation in disease predisposition

Genetic variation can result in, e.g. increased or reduced expression, even complete absence of the gene product, or a malfunctioning or non-functional protein. To understand the biological role of the suggested disease associating polymorphism(s), a number of further studies besides the genetic analyses are

needed. First, it is necessary to determine if the observed association implicates true causality, because the associating variant may be in LD with the actual causative variant. Therefore, it is important to capture the other variants in the loci that were not yet genotyped, and this can be done by mapping the genomic area of interest more densely, e.g. by sequencing or imputation (in GWAS). After this, the newly discovered variants can be further investigated as candidates for disease predisposition (Edwards *et al.* 2013, Ward & Kellis 2012). In addition, it is unlikely that the causal variant acts alone; therefore, the role of the associating variant together with the potentially interacting variant(s) in different genomic loci affecting the predisposition (i.e., epistasis) should be studied (Cordell 2009).

As a preliminary step to connect an associating SNP or other type of polymorphism to a disease process, the possible function of the associated variant(s) (and especially the role of a non-coding variant) can be assessed by investigating whether it is located within any functional non-coding elements, i.e., genetic regulatory elements (e.g. transcription factor binding sites, enhancers, promoters, or regulatory RNAs). This can be studied by exploring, e.g. different SNP functional annotation databases and by applying a variety of bioinformatics approaches that use computational predictions to interpret molecular mechanisms of the disease associated loci (Ward & Kellis 2012). Specific tools, such as expression quantitative trait loci (eQTLs) analysis, can be used to predict the allele-specific effect of non-coding variants on gene expression and regulation in different cells or tissues (Nica & Dermitzakis 2013).

Finally, the function of the proposed candidate gene in disease predisposition can be established by applying *in vitro* and *in vivo* techniques and model systems. For example, the role of the gene can be investigated using *in vivo* animal studies (e.g. mouse, zebra fish) with overexpressing or knock-out models of the target gene. In addition, the effect of specific variations and genes can be studied by targeting the alternative allelic variants in animal models or in appropriate cell lines. To further characterize the role of variations in disease predisposition, specific protein level (e.g. quantitation, localization) and RNA expression level (e.g. microarray, RNAi) assays can be performed. Computer simulations and structural biology can be useful in predicting the effect of the candidate polymorphism on the molecular structure. If the variation in question is a nonsynonymous SNP, exploring the potential consequence of the amino acid substitution in the protein structure can provide clues for further studies to find out, e.g. effects on differential ligand binding or on enzyme activity, depending on the function of the encoded protein. Genetic variants may also affect gene

regulation by epigenetic mechanisms (i.e., affect changes in gene expression that are caused by mechanisms other than changes affecting the DNA sequence; e.g. DNA methylation, and histone modification), and the epigenetic effects may play a role in disease predisposition (Edwards *et al.* 2013, Marian 2012, Peters & Musunuru 2012).

To link the genetic variation to a specific phenotype is complicated; the different animal or tissue models that are used do not always represent the complex disorder in humans, and the regulatory variants are mostly not evolutionary conserved in different species. Additionally, the expression of candidate genes might have different tissue or cell specificity or the expression is limited to certain conditions (Peters & Musunuru 2012). Therefore, integrative functional genomics approaches combining genetic association studies with, e.g. interaction analyses, bioinformatics, epigenetics, transcriptomics, proteomics or metabolomics, further improves the interpretation of the initial discovery (Peters & Musunuru 2012, Ward & Kellis 2012). Overall, unraveling the molecular role of the potential disease associating variant to understand when and how it influences the overall biological system is far from simple and despite a large effort, the exact molecular mechanisms will in most cases remain partly unknown. At their best, however, functional genomic and computational approaches are tools that can help in identifying essential molecular disease-predisposing mechanisms at cellular and molecular levels and can enable the discovery of new variant-specific drug targets.

3 Aims of the study

The aims of this thesis were to identify maternal and fetal genes that could predispose to spontaneous preterm birth, as well as to identify genetic polymorphisms that may influence the risk of developing bronchopulmonary dysplasia in very preterm infants. In addition, the functional role of a nonsynonymous SNP that has been previously associated to respiratory outcomes was studied. The working hypotheses were based on evidence from previous studies. The specific aims were:

1. To investigate collectin genes (those encoding SP-A, SP-D and MBL) as candidates for predisposition to SPTB; these genes were selected as candidates due to their important role in infection and inflammation.
2. To investigate *KITLG*, *IL6*, *IL6R*, *IL6ST*, *IL10*, *TNFA*, *NR3C1* genes as candidates for predisposition to BPD; the investigated genes were selected on the basis of previous biomarker studies or available biomarker data (*IL6ST* and *KITLG*, respectively), as well as their potential role in differentiation (*NR3C1*), or in infection and inflammatory responses, which are central features in BPD.
3. To investigate the functional role of four *SFTPB* polymorphisms, with specific focus on the Ile131Thr polymorphism; this variation has been associated with several pulmonary outcomes and the aim was therefore to provide evidence for its functionality and causality *in vivo* and *in vitro*.

4 Materials and methods

The contents of substudies are summarized in Table 3.

4.1 Ethical considerations

The studies I–IV were approved by the ethics committees of the participating centers; the Ethics Committee of Oulu University Hospital, the University of British Columbia Clinical Research Ethics Board, the University of Alberta Ethics Board, and the Semmelweis University Ethical Committee. Written informed consent was obtained from the participants or from their parents.

4.2 Human sample collection and study populations

4.2.1 Study I

The case population included mothers with spontaneous preterm delivery (< 37 completed weeks of gestation) and their singleton preterm infants. These mothers had either recurrent (preterm delivery at least twice) or sporadic (one preterm delivery) occurrence of SPTB. A total of 94 mothers with recurrent SPTB and their preterm infants (n = 189), as well as 214 mothers with sporadic SPTB and their preterm infants (n = 217) were included. The cases were retrospectively selected from the 1973–2003 birth diaries and prospectively during 2003–2005 in Oulu University Hospital. A detailed description of the exclusion criteria and clinical characteristics are in the original article I. The controls were collected prospectively in Oulu University Hospital during 2004–2007, including 201 mothers with exclusively at least three term (GA > 37 wk) deliveries and their singleton term born infant (n = 201). Study populations are summarized in Fig 5.

All mothers and infants included in the study were collected in Oulu University Hospital; the population in this area (i.e., northern Finland) is known to represent relatively strong genetic homogeneity (Jakkula *et al.* 2008). Both mother and fetus were analyzed as affected, i.e., the phenotypes studied were giving birth spontaneously preterm or being delivered spontaneously preterm, respectively.

Table 3. Summary of the substudies in this thesis.

Study	Scope	Approaches	Study population / material used	Rationale for candidate gene selection	No. of analyzed genes (SNPs)
I	Candidate gene study of collectin (SP-A, SP-D, MBL) polymorphisms in SPTB	Case-control association analysis	Northern Finnish SPTB mothers and SPTB infants, and mothers delivering at term and term infants (Fig. 5).	Role in infection and inflammation	4 (6)
II	Candidate gene study of *KITLG* polymorphism in BPD	Case-control association analysis and protein quantitation	Very preterm infants (GA < 31 wk) (Fig. 6) from Finland, Canada, and Hungary	Role in infection and inflammation	1 (8)
III	Candidate gene study of *IL6*, *IL6R*, *IL6ST*, *IL10*, *TNFA*, and *NR3C1* polymorphism in BPD	Case-control association analysis and study of epistasis	Very preterm infants (GA < 31 wk) (Fig. 6) from Finland, Canada, and Hungary	Role in infection and inflammation	6 (44)
IV	Functional study of *SFTPB* polymorphism(s) *in vivo* and *in vitro*	Genotype–phenotype correlation analysis, studies on allele-specific cell lines	Amniotic fluid and DNA sample pairs from term deliveries, Chinese hamster ovary cells, adult lung samples	Previous associations to preterm birth related pulmonary outcomes	1 (4)

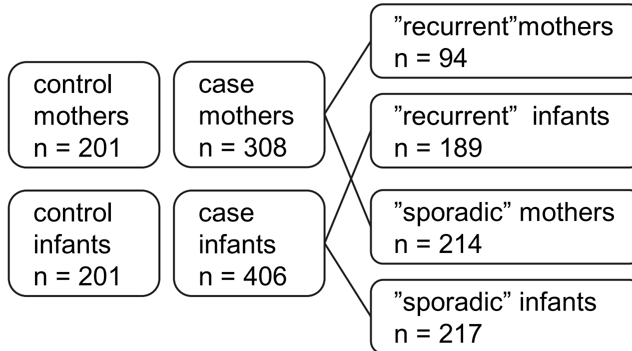

Fig. 5. Populations included in the study I. Controls are mothers with exclusively term deliveries and their term born infant, and cases are mothers with one (sporadic) or more (recurrent) preterm deliveries and their preterm infant. Both mother and infant were studied as affected.

4.2.2 Studies II and III

Studies II and III included cohorts of very preterm infants (GA < 31 wk) with or without moderate-to-severe BPD, originating from northern Finland (n = 253) and Canada (n = 126), as well as a population originating from southern Finland (n = 111) in study II. For replication studies, additional populations from Finland (n = 282 or n = 227, study II or study III, respectively) and Hungary (n = 69) were included. The study populations are summarized in Fig. 6, and the clinical characteristics of the cases and controls are described in detail in the original articles II and III.

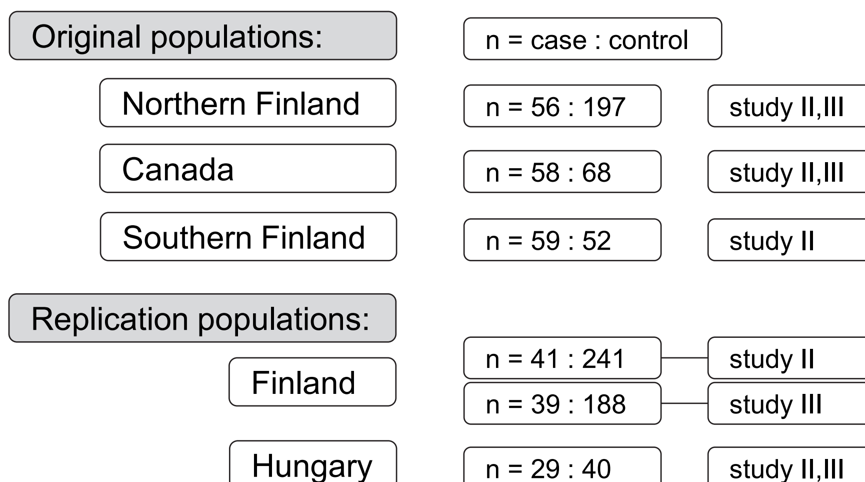

Fig. 6. Populations included in the studies II and III. Cases were infants with moderate-to-severe BPD and controls were infants with none-to-mild BPD. The case-control numbers vary between studies II and III due to technical reasons.

The northern Finnish population was prospectively recruited at Oulu University Hospital in 1997–2010. Samples from infants in the southern Finnish population born in 1997–2003 were retrospectively collected at Helsinki University Hospital and prospectively collected at Tampere University Hospital and Seinäjoki Central Hospital. This population was not included in study III due to genotyping failure for a high number of individuals. Initially in study II, the northern and southern Finnish populations were analyzed separately because the two regions are known to represent slightly different patterns of genetic variation (Jakkula *et al.* 2008). The Canadian population included infants born in neonatal intensive care units in Vancouver and Edmonton in 2006–2008, and all infants were of European descent. Populations were also combined for analyses.

The Finnish replication population included infants born in University Hospitals of Oulu, Kuopio, Tampere, and Helsinki in 2010–2012, and in Turku University Hospital in 2001–2006 and 2011–2012. All infants were of the Finnish origin. These populations were not included in the original analyses, because the samples were prospectively collected during or after the original analyses. Another replication population included infants born in Budapest, Hungary in 1995–2002. These infants were of European descent and the samples were collected in Semmelweis University Hospital.

The diagnosis of BPD was made according to the commonly used definition of supplemental oxygen requirement at 36 wk PMA. For infants that required supplemental O_2 for at least 28 days, the severity of BPD was graded as mild BPD, moderate BPD or severe BPD (no O_2 or ventilation at 36 wk PMA, $O_2 < 30\%$ at 36 wk PMA, or $O_2 \geq 30\%$ or ventilation at 36 wk PMA, respectively) (Jobe & Bancalari 2001). In addition, the oxygen reduction test (Walsh *et al.* 2004) was performed in moderate BPD for infants born in Finland after 2009, which eliminated < 10% of all cases of moderate BPD among these infants. Only one infant from each monozygotic twin pair was included and infants with malformations or those who died early were excluded. In case-control analyses, infants with no or mild BPD were considered controls and those with moderate or severe BPD were considered cases. This approach was based on twin studies that demonstrated genetic predisposition to moderate-to-severe BPD, but less to mild BPD (Lavoie *et al.* 2008). The possibly confounding effect of mild BPD included in the controls was taken into account in the analyses.

4.2.3 Study IV

To study the association between SP-B levels and the *SFTPB* polymorphisms, sample pairs of amniotic fluid and umbilical cord blood (n = 251) were collected from elective Cesarean section deliveries at Oulu University Hospital during 2000–2009. The SP-B protein in amniotic fluid originates from the fetal lung, and the umbilical cord blood was used to determine the fetal genotype. This study population included singleton infants with parents of Finnish origin born at term (GA 37–40 completed weeks). In addition, adult human lung tissue was obtained from patients undergoing lung surgery.

4.3 DNA sample preparation

For the samples collected in Finland, genomic DNA was extracted from whole-blood, buccal cells, dried paper blood spots, umbilical cord blood and tissue samples, or lung tissue specimens, as described in the original articles I–IV. For buccal cell and paper blood DNA samples, whole-genome amplification was performed, as described previously in detail (Karjalainen *et al.* 2012). DNA extraction for samples collected in Canada and Hungary are described in the original articles II and III.

4.4 Selection of polymorphisms and genotyping

Polymorphisms analyzed in the genetic studies I–IV are summarized in Table 4.

Table 4. Polymorphisms analyzed in the genetic studies I–IV.

Study	Gene	Polymorphism* (name with amino acid change if nonsynonymous)
I	*SFTPA1*	rs1136450 (Val50Leu)
	SFTPA2	rs1965708 (Gln223Lys)
	SFTPD	rs721917 (Met31Thr)
	MBL2	rs11003125, rs709620, rs1800450 (Gly54Asp)
II	*KITLG*	rs11104906, rs10858753, rs17424193, rs4842477, rs11104948, rs1907702, rs10506957, rs869408
III	*IL6*	rs2069827, rs1800797, rs1800795, rs2069832, rs2069840
	IL6R	rs4845617, rs1386821, rs4075015, rs4601580, rs4553185, rs4453032, rs4845374, rs4240872, rs4072391
	IL6ST	rs10471960, rs10940495, rs11739048, rs6875155
	IL10	rs3024493
	TNFA	rs1799964, rs1800629
	NR3C1	rs17287758, rs17209237, rs6196, rs10482689, rs10482682, rs6188, rs10482672, rs33383, rs2918417, rs6877893, rs4912905, rs2918415, rs17399352, rs2963155, rs9324924, rs7701443, rs4244032, rs4607376, rs13182800, rs12054797, rs4912910, rs12656106, rs12655166
IV	*SFTPB*	rs3024791, rs2077079, rs1130866 (Ile131Thr), Intron 4 length variation

*polymorphism accession number in the dbSNP database (http://ncbi.nih.gov/SNP).

4.4.1 Study I

Three nonsynonymous SNPs were selected for the study: *SFTPA1* Val50Leu (rs1136450), *SFTPA2* Gln223Lys (rs1965708), and *SFTPD* Met31Thr (rs721917). The selected *SFTPA* SNPs were tagging SNPs and located in functional domains and had previously been associated with pulmonary outcomes (Rämet *et al.* 2000). The *SFTPD* SNP has previously been associated with SP-D oligomerization and serum concentration (Heidinger *et al.* 2005, Leth-Larsen *et al.* 2005, Sorensen *et al.* 2006). In addition, three *MBL2* SNPs were included in the analyses: the promoter variations –550 (H/L) (rs11003125) and –221 (X/Y) (rs709620) and the exon 1 variation Gly54Asp (A/B) (rs1800450). These three SNPs are known to correlate with serum MBL levels (Frakking *et al.* 2006), and occur with moderate frequency in Caucasians. The SNPs in the study are listed in Table 4. The *SFTPA1*, *SFTPA2*, and *SFTPD* SNPs were genotyped by polymerase chain reaction-

restriction fragment length polymorphism (PCR-RFLP) analysis, as described previously (Rämet *et al.* 2000). The *MBL2* SNPs were genotyped using template-directed dye-terminator incorporation with fluorescence polarization detection, as described in the original article I.

4.4.2 Studies II and III

The genes were selected for the studies based on their proposed roles in inflammatory responses and thus they may have a potential role in the predisposition to BPD, and due to available biomarker data (e.g. *KITLG*). For each gene, the selection of tSNPs was performed using HapMap data (http://hapmap.ncbi.nlm.nih.gov/) with the release 24/phase I & II and the CEU population (CEPH; Utah residents with ancestry from northern and western Europe). For pairwise tagging, cut-off values of 0.1 for MAF and 0.9 for r^2 were used. For study II, a total of 8 *KITLG* (gene size 88.1 kb) tSNPs were chosen for genotyping. For study III, a total of 44 SNPs; 5 *IL6* (6.1 kb), 9 *IL6R* (64.3 kb), 4 *IL6ST* (59.9 kb), 1 *IL10* (4.9 kb), 2 *TNFA* (2.8 kb) and 23 *NR3C1* (157.6 kb), were selected for genotyping. The list of the SNPs is presented in Table 4. Genotyping of the tSNPs was performed with the Sequenom iPLEX Gold assay in the Technology Centre, Institute for Molecular Medicine Finland (FIMM), University of Helsinki. In the replication studies II and III, genotyping was performed by PCR-RFLP analysis, as described in the original articles II and III.

4.4.3 Study IV

The *SFTPB* Ile131Thr (rs1130866) variant was selected for the study due to its previous associations to respiratory diseases and its effect on proSP-B glycosylation (Marttila *et al.* 2003a, Wang *et al.* 2003). Other potential disease associating *SFTPB* variants that could have an effect on transcriptional activity or mRNA splicing, and thus could contribute to individual differences in mRNA and protein levels, included SNPs in the promoter region G/A−384 (rs3024791) (Thomas *et al.* 2006) and C/A−18 (rs2077079) (Steagall *et al.* 2007), and a length variation in intron 4 (Δi4) (Lin *et al.* 2005, Rova *et al.* 2004). Genotyping of *SFTPB* Ile131Thr and Δi4 polymorphisms were performed, as described previously (Haataja *et al.* 2000). The rs2077079 and rs3024791 SNPs were genotyped using PCR-RFLP analysis, as described in the original article IV.

4.5 Analysis of Kit ligand concentrations in umbilical cord blood specimens (II)

Kit ligand concentrations were measured from umbilical cord blood serum specimens that were collected from a cohort of very preterm infants (GA < 32 wk) born in Oulu University Hospital during 1998–2002. A total of 120 infants were included in the analysis: 35 infants subsequently developed moderate-to-severe BPD and 85 infants had no-to-mild BPD. The quantitation was performed using an antibody-based microarray, as described previously (Kaukola *et al.* 2009). The Kit ligand concentrations are reported as fluorescence units (FUs), which correspond to the amount of a sample protein. Additionally, the correlation between Kit ligand concentration and 8 *KITLG* SNPs was analyzed, including infants with a DNA sample available (n = 100).

4.6 Protein quantitation, secretion kinetics and allelic imbalance assay (IV)

In study IV, three main analyses were performed:

1. Measurement of relative SP-B protein levels in amniotic fluid samples, using western immunoblots and infrared fluorescence densitometry, and assessment of the SP-B levels for correlations with the four *SFTPB* polymorphisms.
2. Generation of stably transfected Chinese hamster ovary (CHO) cell lines expressing either proSP-B/131Ile or proSP-B/131Thr to examine the effect of glycosylation on proSP-B secretion kinetics; i.e., analysis of SP-B/Ile and SP-B/Thr expression in CHO cells, and secretion onto the medium using western blot analysis and a pulse-chase labeling experiment.
3. Determination of whether there is any preferential transcription between allelic variants of *SFTPB* Ile131Thr, using complementary DNA (cDNA) from heterozygous human lungs for an allelic expression imbalance (AEI) assay.

The methods used in these analyses are described in more detail in the supplemental material of the original article IV.

4.6.1 Protein quantitation from amniotic fluid samples

Proteins from amniotic fluid samples were separated under mildly reducing conditions using sodium dodecyl sulfate-polyacrylamide gel electrophoresis. A pooled amniotic fluid sample known to contain high levels of SP-B was used as a standardized reference in each gel. This preceded western immunoblotting, and direct fluorescence detection and quantitation of mature SP-B. In the densitometry analysis, the intensity of the sample compared to the reference was used to calculate the relative protein levels. The western analysis, the calculation of relative protein values as well as validity (reproducibility, linearity, level of detection and sensitivity) and rationale of the assay is described in more detail in the supplemental material of the original article IV.

4.6.2 Generation of stably transfected proSP-B–expressing CHO cells and proSP-B secretion kinetics

To generate stably transfected proSP-B–expressing CHO cells, RNA from human lung tissue (known to be heterozygous for *SFTPB* Ile131Thr) was reverse transcribed to cDNA. The 1,146-base pair reverse transcriptase-PCR products were subcloned into pGEM-T Easy vectors and subsequently into pcDNA5/FRT constructs. After identifying the correct clones for each variant by sequencing, the expression constructs containing either the Ile or Thr variant cDNA were further transfected into Flp-In-CHO cells. These cells were cultured and the SP-B–expressing clones were selected. Cells and culture media were assayed for SP-B/Ile and SP-B/Thr expression and secretion by western blot analysis. The presence of asparagine-linked oligosaccharides in the expressed proSP-B variants was confirmed with Peptide-N-glycosidase F digestion.

CHO cells stably expressing either proSP-B/Ile or proSP-B/Thr were selected for secretion kinetics. Nontransfected CHO cells were used as controls for the absence of SP-B. After a short period of starvation when the cells were depleted with Met/Cys free medium, the cells were labeled with [^{35}S]methionine/cysteine. After a 30 min pulse, both cells and media were harvested at 0, 30, 60, 120, and 240 min. Radiolabeled CHO cell pellets were lysed, and both lysates and media were used for immunoprecipitation. For secretion kinetics, radioactively labeled SP-B was detected and analyzed.

4.6.3 Allelic expression imbalance assay

RNA from adult human lung samples (n = 8) heterozygous for the *SFTPB* Ile131Thr polymorphism, was reverse transcribed to cDNA and each cDNA was amplified in 5–7 replicate PCR reactions. The PCR products were cleaned up before the SNaPshot reaction (duplicate reactions) that was used to detect potential differences in mRNA expression of the two Ile131Thr alleles. To correct for biases in the detection efficiency of different fluorophores in the SNaPshot assay, full-length SP-B Ile *vs.* Thr cDNA variant plasmid mix (molecular ratio of cDNA-Ile : cDNA-Thr; 50 : 50) was used as a reference sample in all of the AEI assays. To quantitate mRNA allelic expression, the cDNA ratio of Thr and Ile allele peak heights from the lung tissue samples were normalized with the values obtained from the reference plasmid sample used in the same reaction. Using this correction factor, a value of 1.0 represented an equal allelic ratio; i.e., no allelic imbalance. Deviation from this should demonstrate AEI; Thr : Ile ratios significantly below 1.0 indicate preferential expression of the Ile variant mRNA or reduced expression of the Thr variant mRNA, whereas a Thr : Ile ratio > 1.0 would mean the opposite, i.e., preferential expression of the Thr variant mRNA.

4.7 Statistical analysis

Statistical analyses were performed using SPSS Statistics 20.0 (IBM Corporation, Armonk, New York, USA). The nonparametric Mann–Whitney *U*-test and the Pearson χ^2 test were used for case-control comparisons of continuous and dichotomous clinical characteristics or genotype frequency comparisons. In studies II and III, clinical characteristics serving as possible risk factors for BPD (GA, birth weight, IUGR expressed as *Z*-score, gender, and proportion of singletons), were analyzed. In addition, Mann–Whitney *U*-test was used to analyze Kit ligand concentrations between cases and controls and among the *KITLG* genotypes (II). In study II, the sensitivity and specificity of Kit ligand in predicting moderate-to-severe BPD was calculated using receiver operating characteristic (ROC) statistics in SPSS. In addition, in the logistic regression model, dichotomous covariates of Kit ligand concentration (cut-off value defined on the basis of maximal predictive value in ROC), GA < 28 wk and IUGR (*Z*-score < −2 SD) were studied to predict the risk of moderate-to-severe BPD. In study IV, parametric ANOVA and non-parametric Kruskal–Wallis tests were used to analyze the correlation between SP-B protein levels and genotypes. Linear

regression analysis was used to study the effect of the independent variables (GA, respiratory distress and *SFTPB* 131Thr/Thr genotype) on SP-B protein levels. One-sample *t*-test was used for AEI analysis.

Haploview, v. 4.2 (Barrett *et al.* 2005) was used for case-control comparisons for allele and haplotype frequencies (χ^2 tests), Hardy–Weinberg equilibrium tests and to obtain pairwise LD values (D′ and r^2, where D′ refers to the strength of LD and r^2 describes the correlation coefficient between the two loci). PLINK 1.07 (Purcell *et al.* 2007) was used for logistic regression analyses (taking into account constitutional risk factors together with the genetic predisposition). In addition, the Cochran–Mantel–Haenszel test to control the possible effect of clustering in case-control analyses with different combined populations was performed with PLINK (II). Furthermore, the *epistasis* option in PLINK was used for analysis of epistasis, i.e., SNP–SNP interactions (III). The software uses logistic regression (for dichotomous phenotypes) to provide an odds ratio (OR) for the interaction of each pair of SNPs (i.e., whether the correlation between two alleles at the two loci is different between cases and controls), using pair-wise combinations of all of the SNPs. SNPHAP v. 1.3.1. (http://www-gene.cimr.cam.ac.uk/clayton/software) was used to construct *MBL2* phased haplotypes (I). *MBL2* haplogenotypes were classified according to their potential to induce high serum MBL concentrations into classes from 1 (lowest potential) to 10 (highest potential), as described previously (Pesonen *et al.* 2009), and analyzed with the nonparametric Mann–Whitney *U*-test using SPSS statistics.

In the analyses, a p value of < 0.05 was considered significant, and the p value was corrected for multiple testing when appropriate. In study I, p values were corrected for multiple testing using 10,000 permutations (Haploview); a corrected p value < 0.05 was considered significant. In studies II and III, the analyzed SNPs displayed high LD. Therefore, SNPSpD (Nyholt 2004), a method that takes into account LD between SNPs, was used to calculate the effective number of independent SNPs (i.e., that are not in LD with other SNPs) for each gene. According to SNPSpD, $p \leq 0.0144$ (II) and $p \leq 0.0016$ (III) were determined as the significance threshold for multiple comparisons. The Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc) (Purcell *et al.* 2003) was used to estimate the power of the study, as described in the original articles I–III.

5 Results

5.1 Association of *SFTPD* Met31Thr polymorphism and SPTB (I)

The allele frequency distribution of SNPs of four collectin genes *SFTPA1*, *SFTPA2*, *SFTPD*, and *MBL2* were analyzed between preterm infants (cases, n = 406) and term born infants (controls, n = 201). Additionally, the frequency distribution of these SNPs between mothers with preterm deliveries (n = 308) and mothers with exclusively term deliveries (n = 201) were compared.

The *SFTPD* Met31Thr Met allele was overrepresented in SPTB infants compared to controls (frequencies of 0.664 *vs.* 0.605, respectively, $p = 0.04$, OR 1.29, 95% confidence interval [CI] 1.01–1.66). This was apparent in infants from recurrent SPTB families (n = 189) compared to controls (frequencies of 0.717 *vs.* 0.605, respectively, $p = 0.001$, OR 1.65, 95% CI 1.22–2.22). This association remained significant after permutation correction for multiple testing ($p = 0.008$). MAF did not differ according to gender, and presence of PPROM or degree of prematurity (GA < 32 or > 32) did not have an effect on the results. No significant differences were observed in the *SFTPD* Met31Thr allele frequency distributions among the mothers.

No significant differences between the *SFTPA1*, *SFTPA2*, and *MBL2* gene polymorphisms (in comparisons of alleles, genotypes or haplotypes) and SPTB were observed in case-control analyses of infants or mothers. In addition, the *MBL2* haplotypes were given a score (1–10) of its relevance to MBL concentration, as predicted previously (Pesonen *et al.* 2009), but no significant differences were observed in the scores between case and control infants or mothers. Overall, the analyzed SNPs of two *SFTPA* genes and *MBL2* did not associate with SPTB.

5.2 Association of *KITLG* and bronchopulmonary dysplasia (II)

The frequency distribution of eight *KITLG* SNPs were analyzed between very preterm infants with no-to-mild BPD (controls) and moderate-to-severe BPD (cases) in populations originating from northern Finland (cases and controls, n = 253), southern Finland (n = 111) and Canada (n = 126), and SNP rs11104948 was further analyzed in replication populations originating from Finland (n = 282) and

Hungary (n = 69). Additional Kit ligand concentration analysis was performed for preterm infants with biomarker data available (n = 120).

Six *KITLG* SNPs were associated with moderate-to-severe BPD in the northern Finnish population (nominal $p < 0.05$), but only SNP rs11104948 (Table 5), as well as an AGGTCGCT-haplotype consisting of all eight SNPs ($p = 0.011$, OR 2.36, 95% CI 1.20–4.67) (Fig. 7) were significantly associated with BPD after considering the multiple testing ($p < 0.0144$). Thus, only the rs11104948 SNP was further studied in the replication populations. The rs11104948 SNP was not initially associated with BPD in the southern Finnish or Canadian populations (nor were the other *KITLG* SNPs or haplotypes associated with BPD in these two populations), or in the additional Finnish and Hungarian replication populations when the populations were analyzed separately (Table 5). When the populations were combined (all Finnish populations or all available populations), this SNP was significantly associated with BPD, and remained significant after taking into account multiple testing and a possible effect of population clusters (Table 5).

The results were controlled for possible confounding factors that differed between cases and controls (GA and intrauterine growth), but these did not have an effect on the initial results. Additionally, the effect of disease severity, birth year, multiple pregnancies, or degree of prematurity were studied, but these did not affect the result.

Table 5. Allele frequencies of the *KITLG* polymorphism rs11104948 in cases and controls within the different populations.

Study population[1]	Case, Control minor allele frequency	*p* value	OR (95% CI)[2]
Northern Finland, n = 253	0.136, 0.061	0.009c	2.42 (1.22–4.80)
Southern Finland, n = 111	0.098, 0.100	0.965	0.98 (0.40–2.42)
Canada, n = 126	0.112, 0.074	0.290	1.59 (0.67–3.78)
Replication Finland, n = 282	0.134, 0.083	0.135	1.71 (0.84–3.49)
Replication Hungary, n = 69	0.155, 0.138	0.771	1.15 (0.44–2.99)
Combined Finland, n = 639	0.122, 0.076	0.013[c]	1.69 (1.11–2.56)
Combined all populations, n = 834	0.123, 0.080	0.0099[c,d]	1.58 (1.11–2.24)[d]

[1]Included only infants with rs11104948 genotype data available; [2]Odds ratio (95% confidence interval); [c]significant at multiple-testing corrected significance threshold ($p < 0.0144$); [d]analysis was controlled for the effect of clusters, i.e., different populations (Finland, Canada and Hungary); the *p* values and OR are presented as a result of the Cochran–Mantel–Haenszel test.

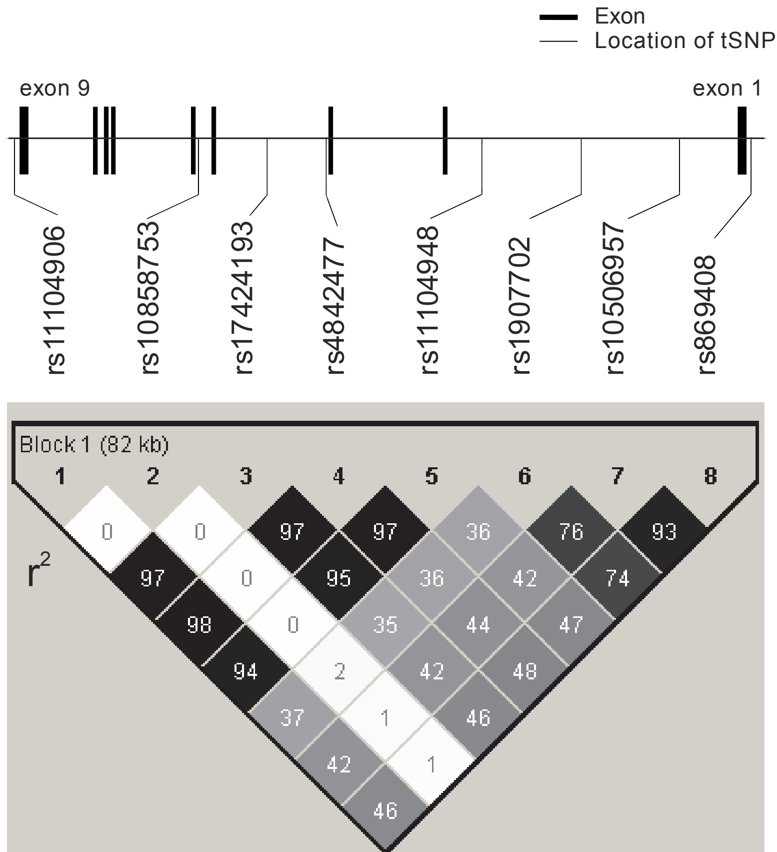

Fig. 7. The genotyped *KITLG* polymorphisms and their relative positions in the gene are shown on the top panel. Numbers in the squares are pairwise r^2 values (lower panel); the darker the square, the higher the correlation. All 8 polymorphisms resided in a single haplotype block. The third common AGGTCGCT-haplotype consisting of minor alleles (except for rs10858753) was overrepresented in BPD infants compared to controls (13.4% *vs.* 6.1%, $p = 0.011$). Due to relatively strong LD, only the rs11104948 SNP was selected for the replication study.

The Kit ligand concentrations were on average higher in the umbilical cord blood serum of very preterm infants who subsequently developed moderate-to-severe BPD (n = 35) than in those with no-to-mild BPD (n = 85) (mean / median ± SD 868 / 738 ± 383 FU *vs.* 632 / 555 ± 266 FU, respectively, $p < 0.001$). Kit ligand concentration predicted the risk of moderate-to-severe BPD (cut-off value 738 FU with a sensitivity of 49% and a specificity of 81%, area under ROC curve 0.72, *p*

= 0.0002). Furthermore, high Kit ligand concentration (≥ 738 FU), extremely low GA (< 28 wk) and IUGR (< −2 SD) together predicted the risk of moderate-to-severe BPD ($p = 0.010$, OR 3.46, 95% CI 1.34–8.90, $p = 0.010$, OR 3.44, 95% CI 1.34–8.82, and $p = 0.025$, OR 3.07, 95% CI 1.15–8.18, respectively). Kit ligand concentrations did not differ between cases and controls among the genotypes of the 8 *KITLG* SNPs.

5.3 No association of *IL6*, *IL6R*, *IL6ST*, *IL10*, *TNFA* and *NR3C1* genes with bronchopulmonary dysplasia (III)

A total of 44 SNPs from *IL6*, *IL6R*, *IL6ST*, *IL10*, *TNFA* and *NR3C1* genes were studied in populations of very preterm infants originating from northern Finland and Canada. None of studied SNPs in these genes associated with moderate-to-severe BPD. We further examined the effect of epistasis in the susceptibility to BPD. We detected a potential SNP–SNP interaction between *IL6ST* rs10471960 and *IL10* rs3024493 SNPs, for the BPD susceptibility in the northern Finnish population ($p = 0.0003$). This interaction did not reach the statistical significance threshold that was corrected for multiple testing ($p = 0.00022$; 231 tests performed in the epistasis analysis). Furthermore, this interaction was not evident in the Canadian population, nor was it detected in replication populations, or when the replication populations were combined with the initial northern Finnish and Canadian populations.

5.4 Studies of *SFTPB* polymorphisms (IV)

5.4.1 Association of SFTPB Ile131Thr and SP-B levels in vivo

The relative SP-B levels in amniotic fluid were analyzed among the fetal genotypes of four *SFTPB* polymorphisms. SP-B in amniotic fluid originates from the fetal lung. The validity of the western analysis was investigated. Amniotic fluid SP-B levels differed significantly between the Ile131Thr genotypes ($p < 0.05$), but not among the other polymorphisms. The Thr/Thr genotypes showed lower SP-B levels compared to both the Ile/Thr and Ile/Ile genotypes (mean / median 49 / 41 *vs.* 66 / 61 *vs.* 61 / 53, respectively, $p < 0.05$). The difference between SP-B levels in the Ile/Ile and Ile/Thr genotypes was not statistically significant ($p > 0.1$).

In the linear regression model taking into account the effects of multiple variables simultaneously; Thr/Thr genotype (p = 0.012), GA (p = 0.001) and respiratory distress (p = 0.001), were significant predictors for SP-B levels in amniotic fluid. The other *SFTPB* polymorphisms and other factors (sex and birth weight) did not explain the SP-B levels, nor did they modify the effect of the *SFTPB* Thr/Thr genotype.

5.4.2 Secretion kinetics and allelic expression imbalance assay

Both CHO cell lines that were generated to express one of the two human *SFTPB* Ile131Thr allelic variants showed robust proSP-B production and secretion into the culture medium. A prominent band of approximately 42 kDa corresponding to proSP-B was detected in cell lysates and media. Both recombinant proteins, proSP-B/Ile and proSP-B/Thr, were sensitive to Peptide-N-glycosidase F due to COOH-terminal glycosylation at Asn311. However, the size shift was larger for proSP-B/Thr, consistent with the presence of an additional oligosaccharide in its NH_2 terminus at Asn129 that is absent from proSP-B/Ile. These data provided evidence for allele-specific differences in the N-glycosylation of the proSP-B in CHO cells.

In pulse-chase labeling experiments with six paired replicates, proSP-B accumulation was detected in the medium in a time-dependent manner lagging about 30–60 min behind detection in the cell lysate. The midpoint times of secretion, illustrated by the V50 values, were 51.9 ± 1.3 min and 58.2 ± 1.2 min for proSP-B/Ile and proSP-B/Thr, respectively. The results of the pulse-chase labeling experiments demonstrate that in CHO cells stably expressing proSP-B, the accumulation of NH_2-terminal N-glycosylated proSP-B/Thr in the culture medium was delayed by approximately 12% compared to nonglycosylated proSP-B/Ile (p = 0.0012).

The AEI assay, using RNA from Ile131Thr heterozygous lung samples (n = 8), was performed to determine if the observed differences in the previous experiments arose transcriptionally or post-transcriptionally. The expression ratio of the Ile and Thr variant mRNAs (multiple replicates), normalized to the reference sample ratio (plasmid DNA, 50:50), was 0.948 ± 0.111 (mean ± SD, range 0.792–1.165), which did not differ from the expected value of 1.0 (p = 0.229), indicating equal mRNA expression and lack of AEI between the two variants.

6 Discussion

In this thesis, the aim was to analyze candidate genes for associations with SPTB and BPD; the genetic background of both phenotypes is incompletely known. Other studies in this thesis were related to a previously identified genetic risk factor for various preterm birth related diseases (i.e., the *SFTPB* gene); here, the aim was to determine whether any presumably functional *SFTPB* polymorphisms correlated with SP-B levels and furthermore to perform functional studies of the Ile131Thr polymorphism within the *SFTPB* gene. The main findings of the studies are discussed below.

6.1 Main findings

6.1.1 Role of SFTPD Met31Thr polymorphism in SPTB (I)

Polymorphisms in four collectin genes *SFTPA1*, *SFTPA2*, *SFTPD* and *MBL2* were analyzed for association with SPTB. The *SFTPD* Met31Thr variation was associated with susceptibility to SPTB among preterm infants from recurrent SPTB families. The Met allele was overrepresented in preterm compared to term infants. Significant allele frequency differences were not observed among the mothers. The other genes were not associated with SPTB.

The *SFTPD* Met31Thr variation that was selected for the study was previously shown to affect SP-D oligomerization and protein concentration in the serum (Heidinger *et al.* 2005, Leth-Larsen *et al.* 2005, Sorensen *et al.* 2006). Furthermore, highly multimerized and lower molecular weight forms of SP-D have been shown to exhibit differential binding to microbial surface molecules (Leth-Larsen *et al.* 2005, Sorensen *et al.* 2009). Therefore, the observed association of *SFTPD* Met31Thr and SPTB is possibly due to the genotype-dependent differences in the binding of pathogens. In a recent study, the expression and oligomeric distribution of SP-D in bronchoalveolar lavage fluid was observed to differ between preterm and term infants; expression was lower on the first day of life in preterm infants subsequently developing chronic lung disease than in infants without lung dysfunction. Additionally, there was a defect in the binding affinity in preterm infants compared to term infants (Kotecha *et al.* 2013a). This is an interesting finding, given the association between the fetal *SFTPD* Met31Thr variation and SPTB.

In previous studies, the *SFTPD* 31Met allele was shown to predispose to RSV infection in term-born young children in the Finnish population (Lahti *et al.* 2002). In contrast, the 31Thr allele has been overrepresented in patients with tuberculosis in a Mexican population and allergic rhinitis in a Chinese population (Silveyra & Floros 2012). These results further suggest that the genetic susceptibility depends on the specific pathogen causing the inflammatory challenge, and that the effect might be variable in different populations.

The suggested role of the *SFTPD* polymorphism in the etiology of preterm labor remains to be investigated. SP-D is mainly produced in the lung, but it is present in other tissues as well, including the female reproductive tract, placenta, and fetal membranes (Kishore *et al.* 2006). The concentration of this protein increases in the amniotic fluid as gestation proceeds (Miyamura *et al.* 1994). SP-D is thought to play a dual inflammatory role (i.e., having both proinflammatory and anti-inflammatory effects) that is dependent on the binding orientation of the protein, which is affected by the degree of SP-D multimerization (Gardai *et al.* 2003). It is possible that the association of the *SFTPD* polymorphism with SPTB is related to this dual function, which may in part be defined by the alternative variants of the Met31Thr polymorphism.

The observed association of *SFTPD* Met31Thr was limited to families with recurrent SPTBs and less evident in the families with sporadic cases of SPTBs, which suggests a stronger genetic predisposition in the families with multiple SPTBs. The etiology of SPTB is extremely heterogeneous, but infection and inflammation are likely the most significant factors associated with SPTB. Our results suggest that the gene encoding SP-D may be one of the genes involved in inflammation that predisposes to SPTB. It may be that the role of this gene is particularly important in SPTBs involving infections.

6.1.2 KITLG as a novel candidate for bronchopulmonary dysplasia susceptibility (II)

For the genetic studies of BPD included in this thesis, several inflammatory related genes (*KITLG*, *IL6*, *IL6R*, *IL6ST*, *IL10*, *TNFA* and *NR3C1*) were studied. Only the *KITLG* polymorphisms were associated with the risk of BPD, the minor alleles of several SNPs being overrepresented in infants with moderate-to-severe BPD compared to controls. Furthermore, in infants born very preterm, the serum Kit ligand concentration at the time of birth was an independent risk factor for subsequent development of moderate-to-severe BPD.

The Kit ligand (also known as stem cell factor) is the principal growth factor for mast cells. It regulates mast cell homeostasis and function, but also has an important role in the early stages of hematopoiesis, i.e., formation of bone marrow-derived blood cells (Kassel *et al.* 2001, Kent *et al.* 2008). In addition, the Kit ligand acts in the production of proinflammatory cytokines and recruitment of fibroblasts. The Kit ligand is expressed by various structural and inflammatory cells in the airways and may play an important role in lung inflammation (Reber *et al.* 2006). Different airway cells produce variable ratios of the two isoforms that are generated by alternative splicing of the *KITLG* transcript (Kassel *et al.* 2001). Production of the Kit ligand is upregulated *in vitro* by proinflammatory stimuli (IL-1β, TNF-α) in vascular endothelial cells (Buzby *et al.* 1994) and lung fibroblasts (Da Silva *et al.* 2002). In contrast, treatment with glucocorticoids is shown to downregulate Kit ligand production (Reber *et al.* 2006). Previously, higher Kit ligand mRNA expression (Al-Muhsen *et al.* 2004) and serum levels (Makowska *et al.* 2009) have been observed in asthmatic patients with higher serum levels also correlated with disease severity (Makowska *et al.* 2009). After binding into its receptor (c-kit), the Kit ligand complex regulates mast cell proliferation, activation and the release of a number of inflammatory mediators (El-Agamy 2012). Excessive accumulation of mast cells in connective tissue has been observed in chronic obstructive pulmonary disease (Andersson *et al.* 2010) and in severe asthma (Balzar *et al.* 2011). A recent study, including a genome-wide transcriptional profiling, revealed a substantial accumulation of connective tissue mast cells in lungs of infants who died of BPD (Bhattacharya *et al.* 2012), and Kit ligand promotes the development of connective tissue–type mast cells *in vivo* (Tsai *et al.* 1991), further supporting the findings of our present study.

The mechanisms by which the associating *KITLG* polymorphisms could affect BPD susceptibility remain to be investigated. Thus far, no direct functional effects have been reported for the polymorphisms analyzed in this study. We hypothesize that the *KITLG* polymorphisms, likely together with some other BPD–predisposing genetic variants, influence the inflammatory responses possibly *via* the mast cell functions, and contribute to the etiology of BPD. The associating variants are common; thus, the effects of the minor alleles are likely to become detrimental only under certain disease predisposing conditions such as prematurity and ongoing intrauterine infection.

6.1.3 Role of SFTPB 131Thr in SP-B levels and secretion (IV)

For the study, we selected four *SFTPB* polymorphisms with previously suggested effects on SP-B mRNA or protein expression. Only Ile131Thr correlated with the SP-B levels *in vivo*. We observed lower concentrations of mature SP-B in amniotic fluid specimens collected from pregnancies with a Thr/Thr fetal genotype compared to other codon 131 genotypes. SP-B in amniotic fluid originates from the fetal lungs. Previously, this *SFTPB* nonsynonymous SNP Ile131Thr has been reported to affect the proSP-B protein N-glycosylation in humans (Wang *et al.* 2003), which may affect susceptibility to various pulmonary diseases such as RDS in neonates (Haataja *et al.* 2000) or acute RDS in adults (Lin *et al.* 2000).

This observation was further supported by *in vitro* experiments using transfected CHO cells that showed delayed secretion of the glycosylated proSP-B/Thr variant compared to the non-glycosylated proSP-B/Ile variant in the transfected cells. No allelic imbalance was evident in heterozygous Ile/Thr lung tissue samples; thus, this excludes preferential transcription of the Ile mRNA compared to the Thr mRNA variant, and suggests a post-transcriptional mechanism for these observations. This further suggests that the presence of the NH_2-terminal N-glycosylation site due to the amino acid substitution at proSP-B position 131 may be a genetic determinant of delayed passage of the protein through the secretory pathway and decreased extracellular SP-B concentrations. It is possible that the glycosylation status at Asn129 could have an effect on proSP-B stability, processing, secretion and/or folding, that results in altered quantities of mature SP-B among individuals carrying different Ile131Thr genotypes. This effect is not large enough to decrease the alveolar SP-B protein content by more than 50% of normal (Clark *et al.* 1997), which is the level required for normal pulmonary function under stable conditions. Additionally, this SNP has a high global frequency, and therefore it cannot be significantly harmful under normal conditions. However, it is possible that this SNP only becomes detrimental under certain disease predisposing conditions, e.g. when the immature lungs of preterm infants are exposed to an adverse environment at the time of birth. However, even a smaller decrease could serve as a trigger for transient respiratory distress during the critical first few minutes of life after preterm birth, when maintenance of adequate alveolar volumes for normal efficient breathing and gas exchange needs to start immediately with no delay in the optimal secretion of surfactant constituents. Our study provides evidence to support a functional role of the

SFTPB Ile131Thr polymorphism, which has previously been shown in several studies to associate with genetic susceptibility to pulmonary diseases.

6.2 Methodological considerations

6.2.1 Study populations

The strengths of our genetic studies of SPTB and BPD include; a rather small confounding effect of population stratification, well defined diagnosis and inclusion criteria, availability of replication populations, and consideration of the effect of multiple comparisons. The Finnish population is known to represent decreased genetic diversity as compared to other European populations (Sajantila *et al.* 1996), and especially the northern Finnish population, is known to be relatively genetically homogenous (Jakkula *et al.* 2008). Using more homogenous populations with less genetic diversity such as the Finnish population (generated by a small isolated founder population having subsequent rapid expansion and bottlenecks), can be beneficial when studying complex diseases (Peltonen *et al.* 2000). Additionally, the Finnish population largely share common environmental and cultural features, there are good records of family and medical data (national population registries) and diagnostics are standardized (due to uniform clinical training across the country) (Peltonen *et al.* 2000). Therefore, having less genetic and environmental heterogeneity, the confounding effect of population stratification is reduced, which increases the power of detecting small-to-moderate effects and makes it possible to detect associations with smaller sample sizes than in more heterogeneous populations (Peltonen *et al.* 2000).

Furthermore, in genetic studies, it is ideal to perform a replication study to confirm the initial results in additional populations. The replication populations only included infants of European origin, which was the main priority (to avoid population stratification) when selecting infants for the study, despite the decreased number of affected individuals (II and III). The effect of possible population heterogeneity was taken into account in the genetic analysis (II). There were no detectable differences among the populations of different geographical origins.

To avoid possible bias occurring from phenotypic differences, the individuals included in the SPTB study were carefully selected from families affected by one or many SPTBs, the pregnancy history has been taken into account, and analysis

included both maternal and fetal alleles to be investigated (I). Additionally, the diagnostic criterion for BPD was consistent across the participating hospitals using the same definition for mild-to-severe BPD (Jobe & Bancalari 2001). The case and control groups were uniformly defined between the different populations, i.e., the infants with a mild form of respiratory distress (mild BPD) were consistently included in the controls (II and III). Furthermore, the results were controlled for possible confounding (clinical) factors (I–IV), and to avoid type I errors, we used a limited number of polymorphisms and applied corrections for multiple testing for the genetic comparisons.

As in most of the genetic studies, the number of subjects is limited. The power of small sample size is not sufficient to detect small effects of common variants or statistically very significant associations. In our studies, the population sizes were also relatively small; even though our combined study population for BPD is among the largest populations used in genetic studies for BPD thus far. Furthermore, the Kit ligand biomarker analyses had limited power for statistical significance due to the small number of cases with moderate-to-severe BPD or individuals with the minor alleles of the SNPs. In addition, the sample size was too small for a reliable evaluation of the *IL6ST–IL10* epistasis effect, which was observed with a borderline statistical significance. Therefore, a replication of the studies should be performed in distinct and larger populations possibly with different ethnicity.

6.2.2 Candidate gene study as a method to find disease predisposing variants

One of the possible limitations in these studies is the use of the candidate gene approach with common variants. Due to our limited sample sizes, which do not provide adequate power to detect small-to-modest associations when studying a large number of polymorphisms, we could not perform a genome-wide association analysis or include rare variants in the study.

The candidate gene approach is hypothesis based and prior knowledge of the possible disease associating genes is desirable. The number of genes and number of variants within the gene to select can turn out to be problematic. In this study, the number of SNPs to be investigated (in order to encompass large genomic areas) was reduced by selecting tagging SNPs that capture most of the common variation within the investigated genes. However, when using tSNPs, the associated variant is not likely to be the causal variant but is in LD with the actual

causal variant that needs to be further identified (II). This can be avoided by selecting noncoding or previously associated variants, which have a relevant biological functional role and thus may be more likely to be disease predisposing. However, when only such polymorphisms are analyzed (I and IV), a smaller proportion of the total genetic variation within the gene is captured. This raises the possibility that other common and rare variants within the gene that are not analyzed could be disease predisposing. Candidate gene analyses always carry the risk of selecting incorrect genes to study and missing the true susceptibility genes. Additionally, gene–gene interactions where the effect of one variant can modify that of the other (that can be located in a different chromosome), might be missed when using a limited set of variants.

6.2.3 Choice of protein quantitation method and use of CHO cells (IV)

Western blot quantitation with infrared fluorescence can be considered a sufficient method for correlation studies that do not require accurate quantitation. The advantage of the western blot method (compared to e.g. enzyme-linked immunosorbent assay) is its ability to distinguish between all proSP-B intermediates and mature SP-B, and diminished background due to non-specific binding, which is typical for enzyme-linked immunosorbent assay–based SP-B quantitation methods. Western blot based quantitation was therefore justified as the method of choice, and also the validation experiments showed that it reproducibly yielded a consistent response and a semi-quantitative estimate of the relative protein levels. Additionally, the total volume of amniotic fluid varies considerably between pregnancies causing intrinsic variation in protein concentrations. Taking into account the limitations of the quantitation method and variability in amniotic fluid volumes, these analyses are not expected to provide accurate protein quantities for diagnostic use.

In addition, a potentially limiting feature is the use of CHO cells instead of ATII cells (which fail to maintain their surfactant related features in cell cultures). The type II cells are the only cell type known to be fully capable of completing all the steps in expressing SP-B, completely processing the proprotein to its mature form and secreting it outside of the cell. The benefits of using stably transfected cell lines (i.e., Flip-In CHO cells) include that the cDNA integration in the host cell genome is similar for both constructs reducing the background variability in the expression levels and secretion kinetic studies (typically present in transient transfections and in cell models with random DNA integration). However, type II

cells only secrete the mature SP-B, whereas all proSP-B is intracellular *in vivo*. Therefore, it is not straightforward to generalize the proSP-B expression and secretion data from CHO cells to SP-B secretion *in vivo*. Nevertheless, the initial NH_2-terminal cleavage is independent of cell type and occurs in the ER soon after translation. Therefore, in type II cells, any differences of the Ile131Thr allelic variants in processing, intracellular trafficking or secretion should originate from very early steps in the ER. Although the CHO cells, which are constitutively secreting cells with no regulated secretory pathway, are not indicative of the secretion kinetics in type II cells, any observed allele-specific differences in the absence of type II cell-specific regulated secretory pathway could provide supporting functional evidence for the role of the glycosylation status of Asn129 present in the Thr variant.

6.3 Significance of the findings and future perspectives

Our results suggest an important role for mediators involved in host defense and inflammatory response pathways in both SPTB, as well as in the respiratory morbidities of preterm birth. A specific *SFTPD* polymorphism, that influences the concentration and the binding affinity of SP-D associates with an increased risk of being born spontaneously preterm. SP-D may influence activation of labor processes when the pregnancy is challenged by inflammatory insult, mostly by ascending infection. Preterm infants are at increased risk of developing serious pulmonary dysfunction manifesting as BPD. A genetic variation within the *KITLG* gene was suggested to have a role in the pathogenesis of BPD. This result was further supported by biomarker data that showed higher concentrations of Kit ligand in infants that subsequently developed the serious forms of BPD than in infants that did not. Roles have been proposed for the Kit ligand and its receptor in lung inflammation and pulmonary growth. In addition, the Kit ligand promotes the differentiation of mast cells that accumulate in the lungs of infants with severe BPD. The hypotheses that arise from this thesis are summarized in Fig. 8.

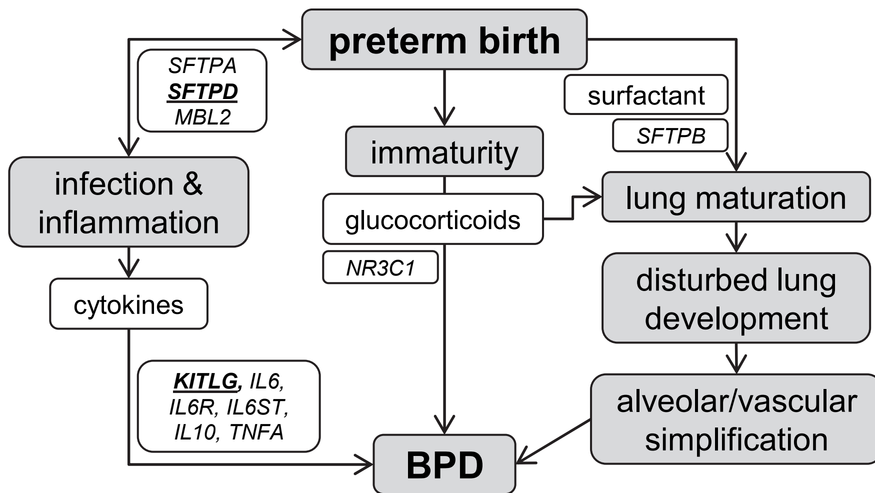

Fig. 8. A flow-chart showing some of the pathways and candidate genes proposed for the susceptibility to spontaneous preterm birth and BPD. The genes investigated in this study are shown, and those with significant associations, are highlighted.

Converting the results of both candidate gene and genome-wide association studies into precise molecular mechanisms and clinical applications is a slow process. Thus, performing further functional studies is essential for a better understanding of the role of the common genetic variation in predisposition to complex disorders. Previously, the *SFTPB* Ile131Thr polymorphism (which selectively alters proSP-B glycosylation) has been hypothesized to have a causal role for adverse pulmonary function such as RDS. In our study, this glycosylation delayed proSP-B secretion *in vitro* and was associated with decreased extracellular SP-B concentration *in vivo*, which supports the hypothesis that proSP-B glycosylation due to the Ile131Thr variation may have a causal role in genetic susceptibility to acute respiratory distress.

More functional studies of the suggested disease predisposing genes and their receptors, as well as further genetic studies, could provide clues about how they influence SPTB and BPD susceptibility. As for future perspectives, the hypothesis-free methods, e.g. GWAS and exome-sequencing (including the search for both common and rare variants, respectively), could be the best methods to investigate the genetic predisposition of SPTB and BPD. In addition, a focus on the genes with accelerated evolution in humans could result in increased power to detect susceptibility variants to SPTB. Human evolution (e.g. large brain, narrow

birth canal) has presumably resulted in selective pressure for earlier birth timing in humans compared to other primates (Hallman 2012, Plunkett *et al.* 2011). The "new" BPD has evolved recently and thus there has not been selective pressure that affects the susceptibility variants. Therefore, it is possible that the predisposing variants may occur at relatively high frequency in the population. Defining the role of the genes and their role in the pathway in predisposition to a disease trait helps not only to determine the etiology but also may eventually help in the design and development of new treatment strategies.

Preterm birth is a complex trait with a multifactorial etiology. Genetic studies contribute to a better understanding of the complex pathways leading to preterm birth and the common respiratory outcomes of infants born very preterm. Further research is necessary to define the role of inflammation and identify the factors involved in the initiation of the inflammatory cascade that leads to diverse outcomes, especially preterm birth and BPD.

7 Conclusions

The aim of the present study was to achieve better knowledge of the genetic factors predisposing to preterm birth and to the development of respiratory diseases in infants born preterm. The main results of these studies are as follows:

1. The *SFTPD* Met31Thr polymorphism is associated with SPTB in the infants.
2. Polymorphisms of *KITLG* associate with the risk for BPD. This novel finding was further supported by biomarker data showing higher concentrations of Kit ligand in infants that later develop BPD.
3. The *SFTPB* genetic variant 131Thr associated with lower SP-B levels *in vivo* and delayed secretion *in vitro*, and thus, could predispose to an increased risk of neonatal respiratory symptoms such as RDS.

Despite the remarkable progress in neonatal intensive care during the last decades, there are still no methods available to prevent preterm birth and its serious consequences, especially BPD. Better knowledge of the genetic background and the molecular mechanisms underlying these phenotypes is of specific interest in both Finland and globally. Knowledge about predisposing genetic factors could contribute to defining specific treatment strategies aimed at prevention of SPTB and BPD. Understanding some of the factors affecting individual susceptibility, and recognizing the causative molecular and cellular mechanisms underlying these events, might enable the identification of the high risk groups and eventually the development of new therapies. The mechanisms of how the disease associating polymorphisms affect the disease outcome are very rarely clear. Therefore, our functional study, which provided a mechanistic evidence to support a causal role of the *SFTPB* Ile131Thr polymorphism (previously shown in several studies to associate with genetic susceptibility to pulmonary outcomes, such as RDS), provides important information and serves as a starting point for future research.

References

Abraham SN & Malaviya R (1997) Mast cells in infection and immunity. Infect Immun 65(9): 3501–3508.

Agrawal V & Hirsch E (2012) Intrauterine infection and preterm labor. Semin Fetal Neonatal Med 17(1): 12–19.

Albertine KH (2013) Progress in understanding the pathogenesis of BPD using the baboon and sheep models. Semin Perinatol 37(2): 60–68.

Ali S, Hirschfeld AF, Mayer ML, Fortuno ES,3rd, Corbett N, Kaplan M, Wang S, Schneiderman J, Fjell CD, Yan J, Akhabir L, Aminuddin F, Marr N, Lacaze-Masmonteil T, Hegele RG, Becker A, Chan-Yeung M, Hancock RE, Kollmann TR, Daley D, Sandford AJ, Lavoie PM & Turvey SE (2013) Functional genetic variation in NFKBIA and susceptibility to childhood asthma, bronchiolitis, and bronchopulmonary dysplasia. J Immunol 190(8): 3949–3958.

Allen MC (2008) Neurodevelopmental outcomes of preterm infants. Curr Opin Neurol 21(2): 123–128.

Al-Muhsen SZ, Shablovsky G, Olivenstein R, Mazer B & Hamid Q (2004) The expression of stem cell factor and c-kit receptor in human asthmatic airways. Clin Exp Allergy 34(6): 911–916.

Ambalavanan N, Carlo WA, D'Angio CT, McDonald SA, Das A, Schendel D, Thorsen P, Higgins RD & Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network (2009) Cytokines associated with bronchopulmonary dysplasia or death in extremely low birth weight infants. Pediatrics 123(4): 1132–1141.

Ananth CV, Getahun D, Peltier MR, Salihu HM & Vintzileos AM (2006) Recurrence of spontaneous versus medically indicated preterm birth. Am J Obstet Gynecol 195(3): 643–650.

Andersson CK, Mori M, Bjermer L, Lofdahl CG & Erjefalt JS (2010) Alterations in lung mast cell populations in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 181(3): 206–217.

Annells MF, Hart PH, Mullighan CG, Heatley SL, Robinson JS, Bardy P & McDonald HM (2004) Interleukins-1, -4, -6, -10, tumor necrosis factor, transforming growth factor-beta, FAS, and mannose-binding protein C gene polymorphisms in Australian women: Risk of preterm birth. Am J Obstet Gynecol 191(6): 2056–2067.

Balzar S, Fajt ML, Comhair SA, Erzurum SC, Bleecker E, Busse WW, Castro M, Gaston B, Israel E, Schwartz LB, Curran-Everett D, Moore CG & Wenzel SE (2011) Mast cell phenotype, location, and activation in severe asthma. Data from the Severe Asthma Research Program. Am J Respir Crit Care Med 183(3): 299–309.

Bancalari E, Claure N & Sosenko IR (2003) Bronchopulmonary dysplasia: changes in pathogenesis, epidemiology and definition. Semin Neonatol 8(1): 63–71.

Barrett JC, Fry B, Maller J & Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263–265.

Beers MF, Shuman H, Liley HG, Floros J, Gonzales LW, Yue N & Ballard PL (1995) Surfactant protein B in human fetal lung: developmental and glucocorticoid regulation. Pediatr Res 38(5): 668–675.

Bhandari A & Bhandari V (2009) Pitfalls, problems, and progress in bronchopulmonary dysplasia. Pediatrics 123(6): 1562–1573.

Bhandari A & McGrath-Morrow S (2013) Long-term pulmonary outcomes of patients with bronchopulmonary dysplasia. Semin Perinatol 37(2): 132–137.

Bhandari V, Bizzarro MJ, Shetty A, Zhong X, Page GP, Zhang H, Ment LR, Gruen JR & Neonatal Genetics Study Group (2006) Familial and genetic susceptibility to major neonatal morbidities in preterm twins. Pediatrics 117(6): 1901–1906.

Bhandari V & Gruen JR (2006) The genetics of bronchopulmonary dysplasia. Semin Perinatol 30(4): 185–191.

Bhattacharya S, Go D, Krenitsky DL, Huyck HL, Solleti SK, Lunger VA, Metlay L, Srisuma S, Wert SE, Mariani TJ & Pryhuber GS (2012) Genome-wide transcriptional profiling reveals connective tissue mast cell accumulation in bronchopulmonary dysplasia. Am J Respir Crit Care Med 186(4): 349–358.

Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, Adler A, Vera Garcia C, Rohde S, Say L & Lawn JE (2012) National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 379(9832): 2162–2172.

Bodamer OA, Mitterer G, Maurer W, Pollak A, Mueller MW & Schmidt WM (2006) Evidence for an association between mannose-binding lectin 2 (MBL2) gene polymorphisms and pre-term birth. Genet Med 8(8): 518–524.

Bokodi G, Treszl A, Kovacs L, Tulassay T & Vasarhelyi B (2007) Dysplasia: a review. Pediatr Pulmonol 42(10): 952–961.

Bolt RJ, van Weissenbruch MM, Lafeber HN & Delemarre-van de Waal HA (2001) Glucocorticoids and lung development in the fetus and preterm infant. Pediatr Pulmonol 32(1): 76–91.

Bose C, Van Marter LJ, Laughon M, O'Shea TM, Allred EN, Karna P, Ehrenkranz RA, Boggess K, Leviton A & Extremely Low Gestational Age Newborn Study Investigators (2009) Fetal growth restriction and chronic lung disease among infants born before the 28th week of gestation. Pediatrics 124(3): e450–8.

Bose CL, Dammann CE & Laughon MM (2008) Bronchopulmonary dysplasia and inflammatory biomarkers in the premature neonate. Arch Dis Child Fetal Neonatal Ed 93(6): F455–461.

Boyd HA, Poulsen G, Wohlfahrt J, Murray JC, Feenstra B & Melbye M (2009) Maternal contributions to preterm delivery. Am J Epidemiol 170(11): 1358–1364.

Buzby JS, Knoppel EM & Cairo MS (1994) Coordinate regulation of Steel factor, its receptor (Kit), and cytoadhesion molecule (ICAM-1 and ELAM-1) mRNA expression in human vascular endothelial cells of differing origins. Exp Hematol 22(2): 122–129.

Clark JC, Weaver TE, Iwamoto HS, Ikegami M, Jobe AH, Hull WM & Whitsett JA (1997) Decreased lung compliance and air trapping in heterozygous SP-B-deficient mice. Am J Respir Cell Mol Biol 16(1): 46–52.

Clark JC, Wert SE, Bachurski CJ, Stahlman MT, Stripp BR, Weaver TE & Whitsett JA (1995) Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice. Proc Natl Acad Sci USA 92(17): 7794–7798.

Clausson B, Lichtenstein P & Cnattingius S (2000) Genetic influence on birthweight and gestational length determined by studies in offspring of twins. BJOG 107(3): 375–381.

Coalson JJ (2006) Pathology of bronchopulmonary dysplasia. Semin Perinatol 30(4): 179–184.

Colin AA, McEvoy C & Castile RG (2010) Respiratory morbidity and lung function in preterm infants of 32 to 36 weeks' gestational age. Pediatrics 126(1): 115–128.

Conde-Agudelo A, Rosas-Bermudez A & Kafury-Goeta AC (2006) Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. JAMA 295(15): 1809–1823.

Conde-Agudelo A, Rosas-Bermudez A & Kafury-Goeta AC (2007) Effects of birth spacing on maternal health: a systematic review. Am J Obstet Gynecol 196(4): 297–308.

Cordell HJ (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet 11(20): 2463–2468.

Cordell HJ (2009) Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 10(6): 392–404.

Creuwels LA, van Golde LM & Haagsman HP (1997) The pulmonary surfactant system: biochemical and clinical aspects. Lung 175(1): 1–39.

Da Silva CA, Kassel O, Mathieu E, Massard G, Gasser B & Frossard N (2002) Inhibition by glucocorticoids of the interleukin-1beta-enhanced expression of the mast cell growth factor SCF. Br J Pharmacol 135(7): 1634–1640.

Daniels CB & Orgeig S (2003) Pulmonary surfactant: the key to the evolution of air breathing. News Physiol Sci 18: 151–157.

de Benedictis FM & Bush A (2012) Corticosteroids in respiratory diseases in children. Am J Respir Crit Care Med 185(1): 12–23.

Doyle LW & Anderson PJ (2010) Adult outcome of extremely preterm infants. Pediatrics 126(2): 342–351.

Dzwonek AB, Neth OW, Thiebaut R, Gulczynska E, Chilton M, Hellwig T, Bajaj-Elliott M, Hawdon J & Klein NJ (2008) The role of mannose-binding lectin in susceptibility to infection in preterm neonates. Pediatr Res 63(6): 680–685.

Edwards SL, Beesley J, French JD & Dunning AM (2013) Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet 93(5): 779–797.

El-Agamy DS (2012) Targeting c-kit in the therapy of mast cell disorders: current update. Eur J Pharmacol 690(1–3): 1–3.

Epaud R, Ikegami M, Whitsett JA, Jobe AH, Weaver TE & Akinbi HT (2003) Surfactant protein B inhibits endotoxin-induced lung inflammation. Am J Respir Cell Mol Biol 28(3): 373–378.

Fanaroff AA, Stoll BJ, Wright LL, Carlo WA, Ehrenkranz RA, Stark AR, Bauer CR, Donovan EF, Korones SB, Laptook AR, Lemons JA, Oh W, Papile LA, Shankaran S, Stevenson DK, Tyson JE, Poole WK & NICHD Neonatal Research Network (2007) Trends in neonatal morbidity and mortality for very low birthweight infants. Am J Obstet Gynecol 196(2): 147.e1–147.e8.

Frakking FN, Brouwer N, Zweers D, Merkus MP, Kuijpers TW, Offringa M & Dolman KM (2006) High prevalence of mannose-binding lectin (MBL) deficiency in premature neonates. Clin Exp Immunol 145(1): 5–12.

Frazer KA, Murray SS, Schork NJ & Topol EJ (2009) Human genetic variation and its contribution to complex traits. Nat Rev Genet 10(4): 241–251.

French JI & McGregor JA (1996) The pathobiology of premature rupture of membranes. Semin Perinatol 20(5): 344–368.

Gardai SJ, Xiao YQ, Dickinson M, Nick JA, Voelker DR, Greene KE & Henson PM (2003) By binding SIRPalpha or calreticulin/CD91, lung collectins act as dual function surveillance molecules to suppress or enhance inflammation. Cell 115(1): 13–23.

Goldenberg RL, Culhane JF, Iams JD & Romero R (2008) Epidemiology and causes of preterm birth. Lancet 371(9606): 75–84.

Goldenberg RL, Goepfert AR & Ramsey PS (2005) Biochemical markers for the prediction of preterm birth. Am J Obstet Gynecol 192(5 Suppl): S36–46.

Goldenberg RL, Hauth JC & Andrews WW (2000) Intrauterine infection and preterm delivery. N Engl J Med 342(20): 1500–1507.

Greenough A (2012) Long term respiratory outcomes of very premature birth (<32 weeks). Semin Fetal Neonatal Med 17(2): 73–76.

Guttentag S (2008) Posttranslational Regulation of Surfactant Protein B Expression. Semin Perinatol 32(5): 367–370.

Haagsman HP, Hogenkamp A, van Eijk M & Veldhuizen EJ (2008) Surfactant collectins and innate immunity. Neonatology 93(4): 288–294.

Haataja R, Karjalainen MK, Luukkonen A, Teramo K, Puttonen H, Ojaniemi M, Varilo T, Chaudhari BP, Plunkett J, Murray JC, McCarroll SA, Peltonen L, Muglia LJ, Palotie A & Hallman M (2011) Mapping a new spontaneous preterm birth susceptibility gene, IGF1R, using linkage, haplotype sharing, and association analysis. PLoS Genet 7(2): e1001293.

Haataja R, Marttila R, Uimari P, Lofgren J, Rämet M & Hallman M (2001) Respiratory distress syndrome: evaluation of genetic susceptibility and protection by transmission disequilibrium test. Hum Genet 109(3): 351–355.

Haataja R, Rämet M, Marttila R & Hallman M (2000) Surfactant proteins A and B as interactive genetic determinants of neonatal respiratory distress syndrome. Hum Mol Genet 9(18): 2751–2760.

Hadchouel A, Decobert F, Franco-Montoya ML, Halphen I, Jarreau PH, Boucherat O, Martin E, Benachi A, Amselem S, Bourbon J, Danan C & Delacourt C (2008) Matrix metalloproteinase gene polymorphisms and bronchopulmonary dysplasia: identification of MMP16 as a new player in lung development. PLoS One 3(9): e3188.

Hadchouel A, Durrmeyer X, Bouzigon E, Incitti R, Huusko J, Jarreau PH, Lenclen R, Demenais F, Franco-Montoya ML, Layouni I, Patkai J, Bourbon J, Hallman M, Danan C & Delacourt C (2011) Identification of SPOCK2 as a susceptibility gene for bronchopulmonary dysplasia. Am J Respir Crit Care Med 184(10): 1164–1170.

Hagberg B, Hagberg G, Beckung E & Uvebrant P (2001) Changing panorama of cerebral palsy in Sweden. VIII. Prevalence and origin in the birth year period 1991–94. Acta Paediatr 90(3): 271–277.

Hallman M (2012) Premature birth and diseases in premature infants: common genetic background? J Matern Fetal Neonatal Med 25 Suppl 1: 21–24.

Hallman M & Haataja R (2006) Surfactant protein polymorphisms and neonatal lung disease. Semin Perinatol 30(6): 350–361.

Hallman M & Haataja R (2007) Genetic basis of respiratory distress syndrome. Front Biosci 12: 2670–2682.

Hallman M & Saarela T (2012) Respiratory Distress Syndrome: Predisposing Factors, Pathophysiology and Diagnosis. In Buonocore G (ed) Neonatology. A Practical Approach to Neonatal Diseases. Italia, Springer-Verlag: 441–454.

Hamvas A (2006) Inherited surfactant protein-B deficiency and surfactant protein-C associated disease: clinical features and evaluation. Semin Perinatol 30(6): 316–326.

Harmon QE, Engel SM, Olshan AF, Moran T, Stuebe AM, Luo J, Wu MC & Avery CL (2013) Association of polymorphisms in natural killer cell-related genes with preterm birth. Am J Epidemiol 178(8): 1208–1218.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R & Hubbard TJ (2012) GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22(9): 1760–1774.

Hartel C, Konig I, Koster S, Kattner E, Kuhls E, Kuster H, Moller J, Muller D, Kribs A, Segerer H, Wieg C, Herting E & Gopel W (2006) Genetic polymorphisms of hemostasis genes and primary outcome of very low birth weight infants. Pediatrics 118(2): 683–689.

Hattersley AT & McCarthy MI (2005) What makes a good genetic association study? Lancet 366(9493): 1315–1323.

Hawgood S (1989) Pulmonary surfactant apoproteins: a review of protein and genomic structure. Am J Physiol 257(2 Pt 1): L13–22.

Hawgood S & Clements JA (1990) Pulmonary surfactant and its apoproteins. J Clin Invest 86(1): 1–6.

Heidinger K, Konig IR, Bohnert A, Kleinsteiber A, Hilgendorff A, Gortner L, Ziegler A, Chakraborty T & Bein G (2005) Polymorphisms in the human surfactant protein-D (SFTPD) gene: strong evidence that serum levels of surfactant protein-D (SP-D) are genetically influenced. Immunogenetics 57(1–2): 1–7.

Hilgendorff A, Heidinger K, Pfeiffer A, Bohnert A, Konig IR, Ziegler A, Merz C, Frey G, Chakraborty T, Gortner L & Bein G (2007) Association of polymorphisms in the mannose-binding lectin gene and pulmonary morbidity in preterm infants. Genes Immun 8(8): 671–677.

Hindorff LA, MacArthur J, Morales J, Junkins HA, Hall PN, Klemm AK & Manolio TA (2014) A Catalog of Published Genome-Wide Association Studies. 2014. URL: www.genome.gov/gwastudies. Cited 16.04.2014.

Hirschhorn JN (2005) Genetic approaches to studying common diseases and complex traits. Pediatr Res 57(5 Pt 2): 74R-77R.

Iams JD, Goldenberg RL, Mercer BM, Moawad A, Thom E, Meis PJ, McNellis D, Caritis SN, Miodovnik M, Menard MK, Thurnau GR, Bottoms SE & Roberts JM (1998) The Preterm Prediction Study: recurrence risk of spontaneous preterm birth. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol 178(5): 1035–1040.

Iams JD, Romero R, Culhane JF & Goldenberg RL (2008) Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. Lancet 371(9607): 164–175.

Ikegami M, Whitsett JA, Martis PC & Weaver TE (2005) Reversibility of lung inflammation caused by SP-B deficiency. Am J Physiol Lung Cell Mol Physiol 289(6): L962–970.

Jack DL, Klein NJ & Turner MW (2001) Mannose-binding lectin: targeting the microbial world for complement attack and opsonophagocytosis. Immunol Rev 180: 86–99.

Jakkula E, Rehnstrom K, Varilo T, Pietilainen OP, Paunio T, Pedersen NL, deFaire U, Jarvelin MR, Saharinen J, Freimer N, Ripatti S, Purcell S, Collins A, Daly MJ, Palotie A & Peltonen L (2008) The genome-wide patterns of variation expose significant substructure in a founder population. Am J Hum Genet 83(6): 787–794.

Jobe AH (2003) Antenatal factors and the development of bronchopulmonary dysplasia. Semin Neonatol 8(1): 9–17.

Jobe AH (2011) The new bronchopulmonary dysplasia. Curr Opin Pediatr 23(2): 167–172.

Jobe AH & Bancalari E (2001) Bronchopulmonary dysplasia. Am J Respir Crit Care Med 163(7): 1723–1729.

Jobe AJ (1999) The new BPD: an arrest of lung development. Pediatr Res 46(6): 641–643.

Johansson J & Curstedt T (1997) Molecular structures and interactions of pulmonary surfactant components. Eur J Biochem 244(3): 675–693.

Kacerovsky M, Lenco J, Musilova I, Tambor V, Lamont R, Torloni MR, Menon R & PREBIC Biomarker Working Group 2012–2013 (2013) Proteomic Biomarkers for Spontaneous Preterm Birth: A Systematic Review of the Literature. Reprod Sci 21(3):283–295.

Kajantie E & Hovi P (2014) Is very preterm birth a risk factor for adult cardiometabolic disease? Semin Fetal Neonatal Med 19(2): 112–117.

Karjalainen MK, Huusko JM, Ulvila J, Sotkasiira J, Luukkonen A, Teramo K, Plunkett J, Anttila V, Palotie A, Haataja R, Muglia LJ & Hallman M (2012) A potential novel spontaneous preterm birth gene, AR, identified by linkage and association analysis of X chromosomal markers. PLoS One 7(12): e51378.

Kassel O, da Silva C & Frossard N (2001) The stem cell factor, its properties and potential role in the airways. Pulm Pharmacol Ther 14(4): 277–288.

Kaukola T, Tuimala J, Herva R, Kingsmore S & Hallman M (2009) Cord immunoproteins as predictors of respiratory outcome in preterm infants. Am J Obstet Gynecol 200(1): 100.e1–100.e8.

Kazzi SN, Kim UO, Quasney MW & Buhimschi I (2004) Polymorphism of tumor necrosis factor-alpha and risk and severity of bronchopulmonary dysplasia among very low birth weight infants. Pediatrics 114(2): e243–8.

Kazzi SN & Quasney MW (2005) Deletion allele of angiotensin-converting enzyme is associated with increased risk and severity of bronchopulmonary dysplasia. J Pediatr 147(6): 818–822.

Kent D, Copley M, Benz C, Dykstra B, Bowie M & Eaves C (2008) Regulation of hematopoietic stem cells by the steel factor/KIT signaling pathway. Clin Cancer Res 14(7): 1926–1930.

Kingma PS & Whitsett JA (2006) In defense of the lung: surfactant protein A and surfactant protein D. Curr Opin Pharmacol 6(3): 277–283.

Kishore U, Greenhough TJ, Waters P, Shrive AK, Ghai R, Kamran MF, Bernal AL, Reid KB, Madan T & Chakraborty T (2006) Surfactant proteins SP-A and SP-D: structure, function and receptors. Mol Immunol 43(9): 1293–1315.

Kistka ZA, DeFranco EA, Ligthart L, Willemsen G, Plunkett J, Muglia LJ & Boomsma DI (2008) Heritability of parturition timing: an extended twin design analysis. Am J Obstet Gynecol 199(1): 43.e1–43.e5.

Kotecha S, Davies PL, Clark HW & McGreal EP (2013a) Increased prevalence of low oligomeric state surfactant protein D with restricted lectin activity in bronchoalveolar lavage fluid from preterm infants. Thorax 68(5): 460–467.

Kotecha SJ, Edwards MO, Watkins WJ, Henderson AJ, Paranjothy S, Dunstan FD & Kotecha S (2013b) Effect of preterm birth on later FEV1: a systematic review and meta-analysis. Thorax 68(8): 760–766.

Kramer BW, Kallapur S, Newnham J & Jobe AH (2009) Prenatal inflammation and lung development. Semin Fetal Neonatal Med 14(1): 2–7.

Lahti M, Lofgren J, Marttila R, Renko M, Klaavuniemi T, Haataja R, Rämet M & Hallman M (2002) Surfactant protein D gene polymorphism associated with severe respiratory syncytial virus infection. Pediatr Res 51(6): 696–699.

Laughon M, Allred EN, Bose C, O'Shea TM, Van Marter LJ, Ehrenkranz RA, Leviton A & ELGAN Study Investigators (2009a) Patterns of respiratory disease during the first 2 postnatal weeks in extremely premature infants. Pediatrics 123(4): 1124–1131.

Laughon MM, Smith PB & Bose C (2009b) Prevention of bronchopulmonary dysplasia. Semin Fetal Neonatal Med 14(6): 374–382.

Lavoie PM & Dube MP (2010) Genetics of bronchopulmonary dysplasia in the age of genomics. Curr Opin Pediatr 22(2): 134–138.

Lavoie PM, Ladd M, Hirschfeld AF, Huusko J, Mahlman M, Speert DP, Hallman M, Lacaze-Masmonteil T & Turvey SE (2012) Influence of common non-synonymous Toll-like receptor 4 polymorphisms on bronchopulmonary dysplasia and prematurity in human infants. PLoS One 7(2): e31351.

Lavoie PM, Pham C & Jang KL (2008) Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the national institutes of health. Pediatrics 122(3): 479–485.

Lawson PR & Reid KB (2000) The roles of surfactant proteins A and D in innate immunity. Immunol Rev 173: 66–78.

Leth-Larsen R, Garred P, Jensenius H, Meschi J, Hartshorn K, Madsen J, Tornoe I, Madsen HO, Sorensen G, Crouch E & Holmskov U (2005) A common polymorphism in the SFTPD gene influences assembly, function, and concentration of surfactant protein D. J Immunol 174(3): 1532–1538.

Lin Z, Pearson C, Chinchilli V, Pietschmann SM, Luo J, Pison U & Floros J (2000) Polymorphisms of human SP-A, SP-B, and SP-D genes: association of SP-B Thr131Ile with ARDS. Clin Genet 58(3): 181–191.

Lin Z, Thomas NJ, Wang Y, Guo X, Seifart C, Shakoor H & Floros J (2005) Deletions within a CA-repeat-rich region of intron 4 of the human SP-B gene affect mRNA splicing. Biochem J 389(Pt 2): 403–412.

Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li M, Mathers C, Black RE & Child Health Epidemiology Reference Group of WHO and UNICEF (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 379(9832): 2151–2161.

Loftin RW, Habli M, Snyder CC, Cormier CM, Lewis DF & Defranco EA (2010) Late preterm birth. Rev Obstet Gynecol 3(1): 10–19.

Lunde A, Melve KK, Gjessing HK, Skjaerven R & Irgens LM (2007) Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. Am J Epidemiol 165(7): 734–741.

Madsen HO, Garred P, Thiel S, Kurtzhals JA, Lamm LU, Ryder LP & Svejgaard A (1995) Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. J Immunol 155(6): 3013–3020.

Madsen HO, Satz ML, Hogh B, Svejgaard A & Garred P (1998) Different molecular events result in low protein levels of mannan-binding lectin in populations from southeast Africa and South America. J Immunol 161(6): 3169–3175.

Maeda Y, Dave V & Whitsett JA (2007) Transcriptional control of lung morphogenesis. Physiol Rev 87(1): 219–244.

Mailaparambil B, Krueger M, Heizmann U, Schlegel K, Heinze J & Heinzmann A (2010) Genetic and epidemiological risk factors in the development of bronchopulmonary dysplasia. Dis Markers 29(1): 1–9.

Makowska JS, Cieslak M & Kowalski ML (2009) Stem cell factor and its soluble receptor (c-kit) in serum of asthmatic patients- correlation with disease severity. BMC Pulm Med 9: 27–2466–9–27.

Mardis ER (2013) Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto Calif) 6: 287–303.

Marian AJ (2012) Molecular genetic studies of complex phenotypes. Transl Res 159(2): 64–79.

Marttila R, Haataja R, Guttentag S & Hallman M (2003a) Surfactant protein A and B genetic variants in respiratory distress syndrome in singletons and twins. Am J Respir Crit Care Med 168(10): 1216–1222.

Marttila R, Haataja R, Rämet M, Lofgren J & Hallman M (2003b) Surfactant protein B polymorphism and respiratory distress syndrome in premature twins. Hum Genet 112(1): 18–23.

Marttila R, Haataja R, Rämet M, Pokela ML, Tammela O & Hallman M (2003c) Surfactant protein A gene locus and respiratory distress syndrome in Finnish premature twin pairs. Ann Med 35(5): 344–352.

McElroy JJ, Gutman CE, Shaffer CM, Busch TD, Puttonen H, Teramo K, Murray JC, Hallman M & Muglia LJ (2013) Maternal coding variants in complement receptor 1 and spontaneous idiopathic preterm birth. Hum Genet 132(8): 935–942.

Melton KR, Nesslein LL, Ikegami M, Tichelaar JW, Clark JC, Whitsett JA & Weaver TE (2003) SP-B deficiency causes respiratory failure in adult mice. Am J Physiol Lung Cell Mol Physiol 285(3): L543–549.

Melville JM & Moss TJ (2013) The immune consequences of preterm birth. Front Neurosci 7: 79 doi: 10.3389/fnins.2013.00079.

Mendelson CR (2009) Minireview: fetal-maternal hormonal signaling in pregnancy and labor. Mol Endocrinol 23(7): 947–954.

Mercer BM, Goldenberg RL, Moawad AH, Meis PJ, Iams JD, Das AF, Caritis SN, Miodovnik M, Menard MK, Thurnau GR, Dombrowski MP, Roberts JM & McNellis D (1999) The preterm prediction study: effect of gestational age and cause of preterm birth on subsequent obstetric outcome. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Am J Obstet Gynecol 181(5 Pt 1): 1216–1221.

Miyamura K, Malhotra R, Hoppe HJ, Reid KB, Phizackerley PJ, Macpherson P & Lopez Bernal A (1994) Surfactant proteins A (SP-A) and D (SP-D): levels in human amniotic fluid and localization in the fetal membranes. Biochim Biophys Acta 1210(3): 303–307.

Morrisey EE & Hogan BL (2010) Preparing for the first breath: genetic and cellular mechanisms in lung development. Dev Cell 18(1): 8–23.

Moster D, Lie RT & Markestad T (2008) Long-term medical and social consequences of preterm birth. N Engl J Med 359(3): 262–273.

Nica AC & Dermitzakis ET (2013) Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci 368(1620): 20120362.

Nicolaides NC, Galata Z, Kino T, Chrousos GP & Charmandari E (2010) The human glucocorticoid receptor: molecular basis of biologic function. Steroids 75(1): 1–12.

Nkadi PO, Merritt TA & Pillers DA (2009) An overview of pulmonary surfactant in the neonate: genetics, metabolism, and the role of surfactant in health and disease. Mol Genet Metab 97(2): 95–101.

Northway WH,Jr, Rosan RC & Porter DY (1967) Pulmonary disease following respirator therapy of hyaline-membrane disease. Bronchopulmonary dysplasia. N Engl J Med 276(7): 357–368.

Nussbaum C & Sperandio M (2011) Innate immune cell recruitment in the fetus and neonate. J Reprod Immunol 90(1): 74–81.

Nyholt DR (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74(4): 765–769.

Orgeig S, Morrison JL & Daniels CB (2011) Prenatal development of the pulmonary surfactant system and the influence of hypoxia. Respir Physiol Neurobiol 178(1): 129–145.

Panoutsopoulou K, Tachmazidou I & Zeggini E (2013) In search of low-frequency and rare variants affecting complex traits. Hum Mol Genet 22(R1): R16–21.

Parker RA, Lindstrom DP & Cotton RB (1996) Evidence from twin study implies possible genetic susceptibility to bronchopulmonary dysplasia. Semin Perinatol 20(3): 206–209.

Parton LA, Strassberg SS, Qian D, Galvin-Parton PA & Cristea IA (2006) The genetic basis for bronchopulmonary dysplasia. Front Biosci 11: 1854–1860.

Peltonen L, Palotie A & Lange K (2000) Use of population isolates for mapping complex traits. Nat Rev Genet 1(3): 182–190.

Perez-Gil J & Weaver TE (2010) Pulmonary surfactant pathophysiology: current models and open questions. Physiology (Bethesda) 25(3): 132–141.

Pesonen E, Hallman M, Sarna S, Andsberg E, Haataja R, Meri S, Persson K, Puolakkainen M, Ohlin H & Truedsson L (2009) Mannose-binding lectin as a risk factor for acute coronary syndromes. Ann Med 41(8): 591–598.

Peters DT & Musunuru K (2012) Functional evaluation of genetic variation in complex human traits. Hum Mol Genet 21(R1): R18–23.

Plunkett J, Borecki I, Morgan T, Stamilio D & Muglia LJ (2008) Population-based estimate of sibling risk for preterm birth, preterm premature rupture of membranes, placental abruption and pre-eclampsia. BMC Genet 9(44).

Plunkett J, Doniger S, Orabona G, Morgan T, Haataja R, Hallman M, Puttonen H, Menon R, Kuczynski E, Norwitz E, Snegovskikh V, Palotie A, Peltonen L, Fellman V, DeFranco EA, Chaudhari BP, McGregor TL, McElroy JJ, Oetjens MT, Teramo K, Borecki I, Fay J & Muglia L (2011) An evolutionary genomic approach to identify genes involved in human birth timing. PLoS Genet 7(4): e1001365.

Plunkett J, Feitosa MF, Trusgnich M, Wangler MF, Palomar L, Kistka ZA, DeFranco EA, Shen TT, Stormo AE, Puttonen H, Hallman M, Haataja R, Luukkonen A, Fellman V, Peltonen L, Palotie A, Daw EW, An P, Teramo K, Borecki I & Muglia LJ (2009) Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. Hum Hered 68(3): 209–219.

Plunkett J & Muglia LJ (2008) Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. Ann Med 40(3): 167–195.

Porter TF, Fraser AM, Hunter CY, Ward RH & Varner MW (1997) The risk of preterm birth across generations. Obstet Gynecol 90(1): 63–67.

Prencipe G, Auriti C, Inglese R, Devito R, Ronchetti MP, Seganti G, Rava L, Orzalesi M & De Benedetti F (2011) A polymorphism in the macrophage migration inhibitory factor promoter is associated with bronchopulmonary dysplasia. Pediatr Res 69(2): 142–147.

Pryhuber GS (1998) Regulation and function of pulmonary surfactant protein B. Mol Genet Metab 64(4): 217–228.

Purcell S, Cherny SS & Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19(1): 149–150.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ & Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3): 559–575.

Quasney MW, Waterer GW, Dahmer MK, Kron GK, Zhang Q, Kessler LA & Wunderink RG (2004) Association between surfactant protein B + 1580 polymorphism and the risk of respiratory failure in adults with community-acquired pneumonia. Crit Care Med 32(5): 1115–1119.

Reber L, Da Silva CA & Frossard N (2006) Stem cell factor and its receptor c-Kit as targets for inflammatory diseases. Eur J Pharmacol 533(1–3): 327–340.

Robinson MR, Wray NR & Visscher PM (2014) Explaining additional genetic variation in complex traits. Trends Genet 30(4): 124–132.

Rocha G, Proenca E, Areias A, Freitas F, Lima B, Rodrigues T, Alves H & Guimaraes H (2011) HLA and bronchopulmonary dysplasia susceptibility: a pilot study. Dis Markers 31(4): 199–203.

Romero R, Espinoza J, Goncalves LF, Kusanovic JP, Friel LA & Nien JK (2006) Inflammation in preterm and term labour and delivery. Semin Fetal Neonatal Med 11(5): 317–326.

Romero R, Friel LA, Velez Edwards DR, Kusanovic JP, Hassan SS, Mazaki-Tovi S, Vaisbuch E, Kim CJ, Erez O, Chaiworapongsa T, Pearce BD, Bartlett J, Salisbury BA, Anant MK, Vovis GF, Lee MS, Gomez R, Behnke E, Oyarzun E, Tromp G, Williams SM & Menon R (2010a) A genetic association study of maternal and fetal candidate genes that predispose to preterm prelabor rupture of membranes (PROM). Am J Obstet Gynecol 203(4): 361.e1–361.e30.

Romero R, Velez Edwards DR, Kusanovic JP, Hassan SS, Mazaki-Tovi S, Vaisbuch E, Kim CJ, Chaiworapongsa T, Pearce BD, Friel LA, Bartlett J, Anant MK, Salisbury BA, Vovis GF, Lee MS, Gomez R, Behnke E, Oyarzun E, Tromp G, Williams SM & Menon R (2010b) Identification of fetal and maternal single nucleotide polymorphisms in candidate genes that predispose to spontaneous preterm labor with intact membranes. Am J Obstet Gynecol 202(5): 431.e1–431.34.

Rooney SA, Young SL & Mendelson CR (1994) Molecular and cellular processing of lung surfactant. FASEB J 8(12): 957–967.

Rova M, Haataja R, Marttila R, Ollikainen V, Tammela O & Hallman M (2004) Data mining and multiparameter analysis of lung surfactant protein genes in bronchopulmonary dysplasia. Hum Mol Genet 13(11): 1095–1104.

Ryckman KK, Morken NH, White MJ, Velez DR, Menon R, Fortunato SJ, Magnus P, Williams SM & Jacobsson B (2010) Maternal and fetal genetic associations of PTGER3 and PON1 with preterm birth. PLoS One 5(2): e9040.

Rämet M, Haataja R, Marttila R, Floros J & Hallman M (2000) Association between the surfactant protein A (SP-A) gene locus and respiratory-distress syndrome in the Finnish population. Am J Hum Genet 66(5): 1569–1579.

Saigal S & Doyle LW (2008) An overview of mortality and sequelae of preterm birth from infancy to adulthood. Lancet 371(9608): 261–269.

Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C & Paabo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci USA 93(21): 12035–12039.

Sampath V, Garland JS, Le M, Patel AL, Konduri GG, Cohen JD, Simpson PM & Hines RN (2012) A TLR5 (g.1174C > T) variant that encodes a stop codon (R392X) is associated with bronchopulmonary dysplasia. Pediatr Pulmonol 47(5): 460–468.

Selman M, Lin HM, Montano M, Jenkins AL, Estrada A, Lin Z, Wang G, DiAngelo SL, Guo X, Umstead TM, Lang CM, Pardo A, Phelps DS & Floros J (2003) Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis. Hum Genet 113(6): 542–550.

Shaw GM & O'Brodovich HM (2013) Progress in understanding the genetics of bronchopulmonary dysplasia. Semin Perinatol 37(2): 85–93.

Silveyra P & Floros J (2012) Genetic variant associations of human SP-A and SP-D with acute and chronic lung injury. Front Biosci (Landmark Ed) 17: 407–429.

Silveyra P & Floros J (2013) Genetic complexity of the human surfactant-associated proteins SP-A1 and SP-A2. Gene 531(2): 126–132.

Somaschini M, Castiglioni E, Volonteri C, Cursi M, Ferrari M & Carrera P (2012) Genetic predisposing factors to bronchopulmonary dysplasia: preliminary data from a multicentre study. J Matern Fetal Neonatal Med 25 Suppl 4: 127–130.

Sorensen GL, Hjelmborg J, Kyvik KO, Fenger M, Hoj A, Bendixen C, Sorensen TI & Holmskov U (2006) Genetic and environmental influences of surfactant protein D serum levels. Am J Physiol Lung Cell Mol Physiol 290(5): L1010–1017.

Sorensen GL, Hoegh SV, Leth-Larsen R, Thomsen TH, Floridon C, Smith K, Kejling K, Tornoe I, Crouch EC & Holmskov U (2009) Multimeric and trimeric subunit SP-D are interconvertible structures with distinct ligand interaction. Mol Immunol 46(15): 3060–3069.

Speer CP (2006) Inflammation and bronchopulmonary dysplasia: a continuing story. Semin Fetal Neonatal Med 11(5): 354–362.

Steagall WK, Lin JP & Moss J (2007) The C/A(-18) polymorphism in the surfactant protein B gene influences transcription and protein levels of surfactant protein B. Am J Physiol Lung Cell Mol Physiol 292(2): L448–453.

Sweet DG, Carnielli V, Greisen G, Hallman M, Ozek E, Plavka R, Saugstad OD, Simeoni U, Speer CP, Vento M, Halliday HL & European Association of Perinatal Medicine (2013) European consensus guidelines on the management of neonatal respiratory distress syndrome in preterm infants--2013 update. Neonatology 103(4): 353–368.

THL (The Finnish National Institute for Health and Welfare) (2013) Perinatal statistics: parturients, deliveries and newborns 2012. Statistical report 24/2013. URL: http://urn.fi/URN:NBN:fi-fe201309276347.

Thomas KH, Meyn P & Suttorp N (2006) Single nucleotide polymorphism in 5'-flanking region reduces transcription of surfactant protein B gene in H441 cells. Am J Physiol Lung Cell Mol Physiol 291(3): L386–390.

Thompson A & Bhandari V (2008) Pulmonary Biomarkers of Bronchopulmonary Dysplasia. Biomark Insights 3: 361–373.

Tokieda K, Iwamoto HS, Bachurski C, Wert SE, Hull WM, Ikeda K & Whitsett JA (1999) Surfactant protein-B-deficient mice are susceptible to hyperoxic lung injury. Am J Respir Cell Mol Biol 21(4): 463–472.

Tsai M, Shih LS, Newlands GF, Takeishi T, Langley KE, Zsebo KM, Miller HR, Geissler EN & Galli SJ (1991) The rat c-kit ligand, stem cell factor, induces the development of connective tissue-type and mucosal mast cells *in vivo*. Analysis by anatomical distribution, histochemistry, and protease phenotype. J Exp Med 174(1): 125–131.

Viscardi RM (2012) Perinatal inflammation and lung injury. Semin Fetal Neonatal Med 17(1): 30–35.

Walsh MC, Yao Q, Gettner P, Hale E, Collins M, Hensman A, Everette R, Peters N, Miller N, Muran G, Auten K, Newman N, Rowan G, Grisby C, Arnell K, Miller L, Ball B, McDavid G & National Institute of Child Health and Human Development Neonatal Research Network (2004) Impact of a physiologic definition on bronchopulmonary dysplasia rates. Pediatrics 114(5): 1305–1311.

Wang G, Christensen ND, Wigdahl B, Guttentag SH & Floros J (2003) Differences in N-linked glycosylation between human surfactant protein-B variants of the C or T allele at the single-nucleotide polymorphism at position 1580: implications for disease. Biochem J 369(Pt 1): 179–184.

Wang H, St Julien KR, Stevenson DK, Hoffmann TJ, Witte JS, Lazzeroni LC, Krasnow MA, Quaintance CC, Oehlert JW, Jelliffe-Pawlowski LL, Gould JB, Shaw GM & O'Brodovich HM (2013) A Genome-Wide Association Study (GWAS) for Bronchopulmonary Dysplasia. Pediatrics 132(2): 290–297.

Ward LD & Kellis M (2012) Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol 30(11): 1095–1106.

Watterberg KL, Demers LM, Scott SM & Murphy S (1996) Chorioamnionitis and early lung inflammation in infants in whom bronchopulmonary dysplasia develops. Pediatrics 97(2): 210–215.

Wert SE, Whitsett JA & Nogee LM (2009) Genetic disorders of surfactant dysfunction. Pediatr Dev Pathol 12(4): 253–274.

Whitsett JA (2010) Review: The intersection of surfactant homeostasis and innate host defense of the lung: lessons from newborn infants. Innate Immun 16(3): 138–142.

Whitsett JA, Wert SE & Trapnell BC (2004) Genetic disorders influencing lung formation and function at birth. Hum Mol Genet 13 Spec No 2: R207–215.

Wilcox AJ, Skjaerven R & Lie RT (2008) Familial patterns of preterm delivery: maternal and fetal contributions. Am J Epidemiol 167(4): 474–479.

Winkvist A, Mogren I & Hogberg U (1998) Familial patterns in birth characteristics: impact on individual and population risks. Int J Epidemiol 27(2): 248–254.

Winters AH, Levan TD, Vogel SN, Chesko KL, Pollin TI & Viscardi RM (2013) Single nucleotide polymorphism in toll-like receptor 6 is associated with a decreased risk for ureaplasma respiratory tract colonization and bronchopulmonary dysplasia in preterm infants. Pediatr Infect Dis J 32(8): 898–904.

Wright JR (2005) Immunoregulatory functions of surfactant proteins. Nat Rev Immunol 5(1): 58–68.

Wu YT, Chen WJ, Hsieh WS, Tsao PN, Yu SL, Lai CY, Lee WC & Jeng SF (2013) MicroRNA expression aberration associated with bronchopulmonary dysplasia in preterm infants: a preliminary study. Respir Care 58(9): 1527–1535.

Yang N, Ray DW & Matthews LC (2012) Current concepts in glucocorticoid resistance. Steroids 77(11): 1041–1049.

Yang W, Ni L, Silveyra P, Wang G, Noutsios GT, Singh A, Diangelo SL, Sanusi O, Raval M & Floros J (2013) Motifs within the CA-repeat-rich region of Surfactant Protein B (SFTPB) intron 4 differentially affect mRNA splicing. J Mol Biochem 2(1): 40–55.

Yoon BH, Romero R, Jun JK, Park KH, Park JD, Ghezzi F & Kim BI (1997) Amniotic fluid cytokines (interleukin-6, tumor necrosis factor-alpha, interleukin-1 beta, and interleukin-8) and the risk for the development of bronchopulmonary dysplasia. Am J Obstet Gynecol 177(4): 825–830.

Yoon BH, Romero R, Kim KS, Park JS, Ki SH, Kim BI & Jun JK (1999) A systemic fetal inflammatory response and the development of bronchopulmonary dysplasia. Am J Obstet Gynecol 181(4): 773–779.

York TP, Eaves LJ, Neale MC & Strauss JF,3rd (2013) The contribution of genetic and environmental factors to the duration of pregnancy. Am J Obstet Gynecol doi: 10.1016/j.ajog.2013.10.001.

York TP, Strauss JF,3rd, Neale MC & Eaves LJ (2010) Racial differences in genetic and environmental risk to preterm birth. PLoS One 5(8): e12391.

Zhou J & Cidlowski JA (2005) The human glucocorticoid receptor: one gene, multiple proteins and diverse responses. Steroids 70(5–7): 407–417.

Original articles

I Karjalainen MK, Huusko JM, Tuohimaa A, Luukkonen A, Hallman M & Haataja R (2012) A study of collectin genes in spontaneous preterm birth reveals an association with a common surfactant protein D gene polymorphism. Pediatr Res 71: 93–99.

II Huusko JM, Mahlman M, Karjalainen MK, Haataja R, Marttila R, Kaukola T, Toldi G, Szabó M, Kingsmore SF, Rämet M, Lavoie PM & Hallman M on behalf of Gen-BPD Study Group (2014) Polymorphisms of the gene encoding Kit ligand are associated with bronchopulmonary dysplasia. Pediatr Pulmonol. DOI: 10.1002/ppul.23018.

III Huusko JM, Karjalainen MK, Mahlman M, Haataja R, Kari MA, Andersson S, Toldi G, Tammela O, Rämet M, Lavoie PM & Hallman M on behalf of Gen-BPD Study Group (2014) A study of genes encoding cytokines (*IL6*, *IL10*, *TNFA*), cytokine receptors (*IL6R*, *IL6ST*), and glucocorticoid receptor (*NR3C1*) and susceptibility to bronchopulmonary dysplasia. Manuscript.

IV Taponen S, Huusko JM, Petäjä-Repo UE, Paananen R, Guttentag SH, Hallman M & Haataja R (2013) Allele-specific N-glycosylation delays human surfactant protein B secretion *in vitro* and associates with decreased protein levels *in vivo*. Pediatr Res 74: 646–651.

Reprinted with permission from Nature Publishing Group (I and IV) and Wiley (II).

Original publications are not included in the electronic version of the dissertation.

1233. Mankinen, Katariina (2014) Neuropsychological performance and functional MRI findings in children with non-lesional temporal lobe epilepsy
1234. Karvonen, Henna (2014) Ultrastructural and functional characterization of myofibroblasts in lung diseases
1235. Sipola, Seija (2014) Colectomy in an ICU patient population : clinical and histological evaluation
1236. Lipponen, Kaija (2014) Potilasohjauksen toimintaedellytykset
1237. Jansson, Miia (2014) The effectiveness of education on critical care nurses' knowledge and skills in adhering to guidelines to prevent ventilator-associated pneumonia
1238. Turunen, Sanna (2014) Protein-bound citrulline and homocitrulline in rheumatoid arthritis : confounding features arising from structural homology
1239. Kuorilehto, Ritva (2014) Moniasiantuntijuus sosiaali- ja terveydenhuollon perhetyössä : monitahoarviointi Q-metodologialla
1240. Sova, Henri (2014) Oxidative stress in breast and gynaecological carcinogenesis
1241. Tiirinki, Hanna (2014) Näkyvien ja piilotettujen merkitysten rajapinnoilla : terveyskeskukseen liittyvät kulttuurimallit asiakkaan näkökulmasta
1242. Syrjänen, Riikka (2014) TIM family molecules in hematopoiesis
1243. Kauppila, Joonas (2014) Toll-like receptor 9 in alimentary tract cancers
1244. Honkavuori-Toivola, Maria (2014) The prognostic role of matrix metalloproteinase-2 and -9 and their tissue inhibitor-1 and -2 in endometrial carcinoma
1245. Pienimäki, Tuula (2014) Factors, complications and health-related quality of life associated with diabetes mellitus developed after midlife in men
1246. Kallio, Miki (2014) Muuttuuko lääketieteen opiskelijoiden käsitys terveydestä peruskoulutuksen aikana : kuusivuotinen seurantatutkimus
1247. Haanpää, Maria (2014) Hereditary predisposition to breast cancer – with a focus on *AATF*, *MRG15*, *PALB2*, and three Fanconi anaemia genes
1248. Alanne, Sami (2014) Musiikkipsykoterapia : Teoria ja käytäntö
1249. Nagy, Irina I. (2014) Wnt-11 signaling roles during heart and kidney development
1250. Prunskaite-Hyyryläinen, Renata (2014) Role of Wnt4 signaling in mammalian sex determination, ovariogenesis and female sex duct differentiation

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O. Box 8000 FI-90014 UNIVERSITY OF OULU FINLAND

A C T A U N I V E R S I T A T I S O U L U E N S I S

S E R I E S E D I T O R S

A SCIENTIAE RERUM NATURALIUM

Professor Esa Hohtola

B HUMANIORA

University Lecturer Santeri Palviainen

C TECHNICA

Postdoctoral research fellow Sanna Taskila

D MEDICA

Professor Olli Vuolteenaho

E SCIENTIAE RERUM SOCIALIUM

University Lecturer Veli-Matti Ulvinen

F SCRIPTA ACADEMICA

Director Sinikka Eskelinen

G OECONOMICA

Professor Jari Juga

EDITOR IN CHIEF

Professor Olli Vuolteenaho

PUBLICATIONS EDITOR

Publications Editor Kirsti Nurkkala

ISBN 978-952-62-0473-4 (Paperback)
ISBN 978-952-62-0474-1 (PDF)
ISSN 0355-3221 (Print)
ISSN 1796-2234 (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO