Mari Mahlman

GENETIC BACKGROUND AND ANTENATAL RISK FACTORS OF BRONCHOPULMONARY DYSPLASIA

UNIVERSITY OF OULU GRADUATE SCHOOL; UNIVERSITY OF OULU, FACULTY OF MEDICINE, MEDICAL RESEARCH CENTER, OULU UNIVERSITY HOSPITAL
ACTA UNIVERSITATIS OULUENSIS
D Medica 1472

MARI MAHLMAN

GENETIC BACKGROUND AND ANTENATAL RISK FACTORS OF BRONCHOPULMONARY DYSPLASIA

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 Oulu University Hospital (Kajaanintie 50), on 20 June 2018, at 12 noon

UNIVERSITY OF OULU, OULU 2018
Mahlman, Mari, Genetic background and antenatal risk factors of bronchopulmonary dysplasia.
University of Oulu Graduate School; University of Oulu, Faculty of Medicine; Medical Research Center; Oulu University Hospital
*Acta Univ. Oul. D 1472, 2018*
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

**Abstract**

Advances over the past few decades in ante- and neonatal care have led to the survival of a growing number of premature infants of extremely low gestational age. However, the occurrence of serious diseases, particularly those affecting the most immature infants, remains high. Bronchopulmonary dysplasia (BPD), a chronic lung disease of premature infants, is one such disease. Our current understanding of the molecular pathogenesis of BPD is incomplete; consequently, there are few preventive and therapeutic options for BPD. Moreover, it is challenging to predict the risk of BPD. Previous studies of BPD in twins revealed that the heritability of BPD is quite high. However, the individual genes that predispose premature infants to BPD are largely unknown.

The aim of this study was to identify and study genes associated with BPD in order to investigate its pathogenesis. An additional aim was to add to knowledge of the risk of BPD in newborn premature infants, with an emphasis on twins.

A candidate gene study found no consistent association between common polymorphisms of vascular endothelial growth factor receptor 2 and BPD. A second candidate gene study noted an association between the gene encoding Kit ligand and BPD. A genome-wide association study found a suggestive association between a locus close to the gene encoding C-reactive protein (CRP) and BPD, and in subsequent analyses, plasma levels of CRP during the first week of life predicted BPD. Finally, a nationwide register study found that the risk of BPD was lower in twins than in singletons.

The results of this study add to what is known of the genetics and pathogenesis of BPD. They also provide new data on the risk of BPD, which may be used to improve early identification of infants for whom the risk of developing BPD is high.

**Keywords:** bronchopulmonary dysplasia, C-reactive protein, candidate gene study, genetic association study, genetic polymorphism, genome-wide association study, Kit ligand, premature birth, premature infant, pulmonary inflammation, twins, vascular endothelial growth factor, vascular endothelial growth factor receptor 2
Mahlman, Mari, Bronkopulmonaalisen dysplasian perinnöllinen alttius ja syntymää edeltävät riskitekijät.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Lääketieteellinen tiedekunta; Medical Research Center; Oulun yliopistollinen sairaala
Acta Univ. Oul. D 1472, 2018
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Ennenaikaisen syntymän ja keskoslasten hoidon kehittymisen myötä yhä useammat huomattavan epäkypsinä syntyneet lapset jäävät henkiin. Samalla erityisesti juuri näitä lapsia uhkaavien sairauksien esiintyvyys on pysynyt korkeana. Bronkopulmonaalinen dysplasia (BPD, keskosen krooninen keuhkosairaus) on yksi näistä sairauksista.

BPD:n molekyylitasoinen tautimekanismi on vielä osin tuntematon, eikä BPD:tä tehokkaasti estävää tai siitä parantavaa hoitoa ole. Myös BPD riskin arvioiminen vastasyntyneen keskoslasten kohdalla on vaikeaa. BPD on huomattavan perinnöllinen tauti. BPD:lle altistavista geenistä on kuitenkin vasta vähän tietoa.

Tämän tutkimuksen tavoitteena oli lisätä tietoa BPD:n tautimekanismissa tutkimalla BPD:lle altistavia geenejä. Lisäksi tutkimuksessa tarkasteltiin BPD:n esiintyvyyttä ja syntymää edeltäviä riskitekijöitä erityisesti kaksosten osalta.


Tutkimuksen tulokset lisäävät tietoa BPD:n perinnöllisyydestä ja sitä kautta BPD:n tautimekanismissa. Tutkimus toi myös uutta tietoa BPD:n riskitekijöistä parantaa vastasyntyneen keskoslappmen BPD-riskin arviota.

Asiasanat: bronkopulmonaalinen dysplasia, C-reaktiivinen proteiini, ehdokasgeenitutkimus, ennenaikainen syntymä, geenipolymorfismi, geneettinen assosiaatiotutkimus, kaksoset, keskosen krooninen keuhkosairaus, keskosuus, keuhkojen tulehdusreaktio, Kit ligandi, koko genomin assosiaatiotutkimus, verisuonten endoteelikasvutestejä
To my family
Acknowledgements

This study was carried out at the Department of Pediatrics, University of Oulu, and Oulu University Hospital, during the years 2010–2018.

My deepest gratitude goes to my main supervisor and the father of this project, Professor emeritus Mikko Hallman. Mikko, your never-ending enthusiasm, energy and compassion towards science, it is incredible. Even more than that, I thank you for letting me learn from your humble, hard-working attitude and for all the cordiality and support you addressed me during the years of this project. I’m also grateful to my other supervisors, Professor Mika Rämet and Doctor Minna Karjalainen. Mika and Minna, I thank you for your expertise and invaluable guidance into the world of science. Without your efforts and contribution, I would not be writing these lines.

I thank Docent Päivi Tapanainen, the Head of the Department of Pediatrics at Oulu, for once welcoming me to the department and for creating not only a research friendly but a true friendly atmosphere there. I thank also Professor emeritus Matti Uhari, for letting your extraordinary skills in statistics shine also over me, no matter the modest absorption rate. I’m also grateful to Pediatric Research Center, Helsinki University Hospital, for providing me excellent facilities to finalize this thesis.

Professor Riitta Veijola and Professor Johanna Uusimaa, the members of my follow-up group; I thank you for your time and the friendly attitude you showed towards this project.

I would also like to acknowledge my pre-examiners Docent Samuli Rautava and Docent Krista Rantakari. Your wise comments and constructive criticism during the review process clearly improved this thesis. I’m also grateful to Docent Heikki Lukkarinen for your expertise in the field of pulmonology and for accepting to be my opponent.

For this thesis project, a DNA collection named GenBPD was established in all five Finnish University Hospitals. I thank Professor Sture Andersson and Docent Marjo Metsäranta, Helsinki University Hospital, Professor Liisa Lehtonen, Turku University Hospital, Docent Outi Tammela, Tampere University Hospital and Docent Ulla Sankilampi, Kuopio University Hospital, for welcoming GenBPD and me as a researcher to your institutions. I would also like to thank all the colleagues, nurses and midwives who participated in the studies of this thesis by collecting DNA-samples and clinical data, and by recruiting the study subjects. Especially I would like to thank all the children and families who participated in the studies; without you this work would not have been possible.
All my co-authors are acknowledged for their valuable contributions to the original publications in this thesis. In addition to the persons acknowledged elsewhere in this text, I thank Dirk Bassler, Claude Danan, Christophe Delacourt, Ritva Haataja, Alice Hadchouel, Tuula Kaukola, Stephen Kingsmore, Thierry Lacaze-Masmonteil, Pascal Lavoie, Riitta Marttila, Louis Muglia, Marja Ojaniemi, Aarno Palotie, Christian Poets, Tytti Pokka, Miklos Szabo and Gergely Toldi.

All the members of our research group, especially Johanna and Anu; I thank you for companionship and for practical advice when completing this project. Riitta Vikeväinen, The research nurse in our group; to you I give such a warm hug that you definitely feel it: You are the one who shared most of the pain with me with this journey and I’m positive; also the one enjoying the most leaving it behind. Laura, Elisa, Jenni and other great colleagues at the Researcher community at the sixth floor of Biomedicum; I thank you for your friendship and peer support towards writing these lines.

I would also like to thank all the colleagues, doctors as well as nurses and midwives, for the work we’ve done together in the clinics, and for the endless amount of opportunities I’ve been given. Especially I would like to acknowledge Docent Anna-Liisa Järvenpää and Docent Anneli Kari, for once introducing the world of neonatology to me. My friends and former colleagues at Oulu, with a special emphasis on the team of Young Investigators at the legendary Avokonttori; I thank you for our friendship, and for so many good moments during my stay at OYS. You do have a good pössis up there and I’m happy to once have been part of it! My clinical colleagues at Helsinki, – no-one mentioned, no-one forgotten – I thank you for welcoming me back and I look forward working together with you.

Our daughter Kerttu was seriously ill with streptococcus A empyema in February this year. I want to use this opportunity to thank all the staff at Children’s Hospital in Helsinki who participated in the care of her – nowadays Kerttu is fully recovered and the gratefulness of me as a mother goes beyond the words.

Touko, Visa and Ade, it has been a privilege to grow beside you. Thank you for being your beautiful selves. Petri, for sharing the parenthood and especially for being a team when Kerttu was sick. Rammu, for all the berries and rug socks. Hannu and Arja, for the bonus-mummila at Oulu.

Elbe, my best friend since I was a toddler. You are such an intelligent, beautiful and strong woman with an unbeatable sense of humor and an outstanding attitude towards life; gratefulness. I thank you for everything we’ve gone through during the years, and yes, we have. Hugs hugs and hugs, and onto the next, sis!
All my other friends; I thank you all for what you are to me: A joy in my everyday life, a warm hug in the middle of hurry and a faithful support in times of trouble. Saltsu, I thank you for being such a warm, wise and empathic friend, and for all the long and brilliant analyses of the state of things we’ve had during the years. Catharina, for the great moments and conversations at Med School at Karolinska and especially outside of it. Stefan, for all that it means to be a true cousin. Joao and the family, for the extended Portuguese-Finnish family filled with hugs and friendship. Maarit, for letting the goodness shine through you over my family. Kirsi, Susanna and other great mothers in my current neighbourhood; in our community you never stay awake alone!

Finally, my closest ones. My parents Tuula and Seppo: Nothing has ever been out of my reach, and that is because of you. The way you love, support and believe in me, no matter the situation – you are simply the best ever existed parents. My brother Sami, I thank you for always being there for me, and for all the fun we’ve had together. And for your all but constructive criticism towards my efforts here in life; you have given me a great training to defend this thesis! Then, Ari, my body. When you came into my life, that was such a gamechanger. Easy-going good mood, companionship, support and shared dreams; I thank you for all that we have, not to talk about what we will have! And then, finally, Kerttu and Kuutti, the most beloved children in the world. This book is for you, no, this universe is. Always remember this; what matters is what is inside of you, not what you reach and what you do not. I love you more than anything and I always will. And I’m blessed, grateful and extremely proud of being your mother.

This study was supported by the Alma and K. A. Snellman Foundation; the Foundation for Pediatric Research; the Medical Research Center Oulu, the National Graduate School of Clinical Investigation and the Scholarship Fund of the University of Oulu.

Helsinki, May 2018

Mari Mahlman
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW</td>
<td>birth weight</td>
</tr>
<tr>
<td>BPD</td>
<td>bronchopulmonary dysplasia</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variation</td>
</tr>
<tr>
<td>CP</td>
<td>cerebral palsy</td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
</tr>
<tr>
<td>FIMM</td>
<td>Institute for Molecular Medicine Finland</td>
</tr>
<tr>
<td>FiO2</td>
<td>fraction of inspired oxygen</td>
</tr>
<tr>
<td>GA</td>
<td>gestational age</td>
</tr>
<tr>
<td>GEE</td>
<td>generalized estimating equation</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association study</td>
</tr>
<tr>
<td>h²</td>
<td>heritability</td>
</tr>
<tr>
<td>IU</td>
<td>intra-uterine</td>
</tr>
<tr>
<td>IUGR</td>
<td>intra-uterine growth restriction</td>
</tr>
<tr>
<td>LD</td>
<td>linkage disequilibrium</td>
</tr>
<tr>
<td>MAF</td>
<td>minor allele frequency</td>
</tr>
<tr>
<td>MSC</td>
<td>mesenchymal stromal cell</td>
</tr>
<tr>
<td>NCPAP</td>
<td>nasal continuous positive airway pressure</td>
</tr>
<tr>
<td>NICHD</td>
<td>National Institute of Child Health and Human Development</td>
</tr>
<tr>
<td>NNT</td>
<td>number needed to treat</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>PMA</td>
<td>post-menstrual age</td>
</tr>
<tr>
<td>RCT</td>
<td>randomized controlled trial</td>
</tr>
<tr>
<td>RDS</td>
<td>respiratory distress syndrome</td>
</tr>
<tr>
<td>ROP</td>
<td>retinopathy of prematurity</td>
</tr>
<tr>
<td>RR</td>
<td>relative risk</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>sFlt-1</td>
<td>soluble fms-like tyrosine kinase 1</td>
</tr>
<tr>
<td>SGA</td>
<td>small for gestational age</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SNV</td>
<td>single nucleotide variant</td>
</tr>
<tr>
<td>tSNP</td>
<td>tagging SNP</td>
</tr>
<tr>
<td>VEGF</td>
<td>vascular endothelial growth factor</td>
</tr>
<tr>
<td>VEGFR</td>
<td>vascular endothelial growth factor receptor</td>
</tr>
</tbody>
</table>
List of original articles

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:


# Table of contents

**Abstract**  
9

**Tiivistelmä**  

**Acknowledgements**  
13

**Abbreviations**  
15

**List of original articles**  
17

**Table of contents**  
17

1  **Introduction**  
19

2  **Review of the literature**  
21

2.1  Bronchopulmonary dysplasia (BPD)  
21

2.1.1  Definition and diagnostics  
21

2.1.2  Incidence  
24

2.1.3  Pathogenesis and risk factors  
25

2.1.4  Prevention and treatment  
34

2.1.5  Outcome  
38

2.1.6  Twins and BPD  
40

2.2  BPD as a complex disease  
41

2.2.1  Basic concepts in studying complex diseases  
41

2.2.2  Methods to study complex diseases  
43

2.2.3  Twin studies of the heritability of BPD  
47

2.2.4  Candidate gene studies of BPD  
48

2.2.5  Genome-wide associations studies of BPD  
48

2.2.6  Other genomic studies of BPD  
50

2.2.7  Epigenetic studies of BPD  
52

3  **Aims of the study**  
55

4  **Methods**  
57

4.1  Design  
57

4.2  Study infants  
57

4.3  Selection of genotyped SNPs  
59

4.4  DNA sample preparation and genotyping  
60

4.5  Functional analyses of the associating genes  
60

4.6  Statistical analyses  
60

4.7  Definitions  
62

4.7.1  Definition of BPD in present studies  
62

4.7.2  Definition of IUGR/SGA in present studies  
63

4.8  Ethical considerations  
63
5 Results

5.1 *VEGFR2* – no consistent association with BPD (I) .............................................. 65

5.2 *KITLG* – an association with BPD (II) .............................................................. 65
   5.2.1 SNP rs11104948 associated with BPD ......................................................... 65
   5.2.2 Cord blood levels of Kit ligand predicted BPD .............................................. 67

5.3 CRP – an association with BPD (III) .................................................................. 67
   5.3.1 SNP rs11265269 associated suggestively with BPD ................................. 68
   5.3.2 SNP rs3093059 and haplotypes from the CRP region associated with BPD ................................................................. 70
   5.3.3 SNP rs3093059 associated with plasma CRP levels ................................. 70
   5.3.4 Plasma CRP levels predicted BPD ................................................................. 71
   5.3.5 Other SNPs with suggestive associations with BPD ................................. 71

5.4 The risk of BPD was lower in twins than in singletons (IV) .............................. 73
   5.4.1 The antenatal risk factors of BPD did not differ between twins and singletons ...................................................................................... 75

6 Discussion ........................................................................................................... 77

6.1 Methodological considerations .......................................................................... 77
   6.1.1 Study populations ....................................................................................... 77
   6.1.2 Use of candidate gene and genome-wide association studies to identify genes involved in BPD ................................................................. 78

6.2 Main Findings .................................................................................................... 81
   6.2.1 VEGF and BPD (I) .................................................................................. 81
   6.2.2 Kit ligand and BPD (II) ............................................................................. 82
   6.2.3 CRP and BPD (III) .................................................................................. 83
   6.2.4 Twins and BPD (IV) ................................................................................ 86
   6.2.5 Future studies ............................................................................................ 87

7 Conclusions ......................................................................................................... 89

References ............................................................................................................. 91

List of original publications .................................................................................. 115
1 Introduction

Human pregnancy lasts for an average of 40 weeks. When a birth occurs before 37 completed gestational weeks, it is defined as preterm. Preterm birth is a major health concern that affects 15 million infants yearly. Approximately 20% of preterm infants are born before 32 gestational weeks and are classified as very preterm. (Blencowe et al., 2013; Martin, Hamilton, & Drake, 2018).

The survival of very preterm infants has improved dramatically over the last few decades. In Finland, 44% of infants born very preterm had one or more diagnosed prematurity-related morbidity during the first 3 years of life (Korvenranta et al., 2009). Furthermore, with the increased survival of even the most immature infants born at 23–24 weeks gestation, the incidence of some prematurity-related morbidities, such as bronchopulmonary dysplasia (BPD), has increased (Stoll et al., 2015). There are approximately 10,000 new diagnoses of BPD each year in the United States alone (Jensen & Schmidt, 2014). In Finland, the yearly incidence of BPD is around 60 (Small Preterm Infants data file, Finnish National Institute for Health and Welfare).

BPD is a chronic lung disease of premature infants. It clinically manifests as a prolonged need for respiratory support. Abnormal neurological outcome (Natarajan et al., 2012) and impaired lung function persisting into adulthood (Baraldi & Filippone, 2007) are among the long-term consequences of BPD. Despite intensive research, the molecular pathogenesis of BPD remains mostly unknown and the means of prevention are scarce. However, BPD, based on twin studies, is a particularly hereditary disease (V. Bhandari et al., 2006; Lavoie, Pham, & Jang, 2008); consequently, genetic studies may yield much information about the pathogenesis of BPD. To date, however, knowledge of BPD susceptibility genes is limited.

The aim of the present study was to increase the understanding of the pathogenesis of BPD by studying common genetic variations associated with BPD. We performed both candidate gene and genome-wide association studies. In addition, we evaluated the risk of BPD and the nongenetic antenatal risk factors of BPD, with an emphasis on twins, in a population-based register study.
2 Review of the literature

2.1 Bronchopulmonary dysplasia (BPD)

2.1.1 Definition and diagnostics

BPD was first described in 1967 by Northway et al. (Northway, Rosan, & Porter, 1967). In their landmark article, the authors described BPD as a condition with clinical, pathological, and radiological characteristics that advanced from acute respiratory distress syndrome (RDS) to progressive chronic lung disease. The mean gestational age (GA) at birth of the survivors in Northway et al.’s study was 34 weeks, and the mean birth weight (BW) was 2311 g, which illustrates the high neonatal mortality of that era. Autopsies of the nonsurviving BPD infants identified diffuse fibroproliferative changes, distal lung inflammation, and hypertensive pulmonary vascular remodeling. Radiographic changes included lung hyperexpansion and bullae. Northway et al. proposed in their seminal paper that BPD might result from complex interactions among the following features: preterm birth; lung injury caused by hyperoxia, ventilator treatment, and inflammation; and efforts to heal the injury: today, the association of all of these elements with BPD is recognized. In addition, the authors recognized the important comorbidities of BPD: brain injury and retinopathy of prematurity (ROP).

Since the first description of BPD, the survival of preterm infants has improved dramatically, even for infants of 22–24 weeks GA. The improved survival is mainly the result of treatment with antenatal corticosteroids (Liggins, 1969; Liggins & Howie, 1972) and surfactant (Avery & Mead, 1959; Fujiwara et al., 1980; Hallman et al., 1985). Progress has also been made in respiratory care, with the development of continuous positive airway pressure (Gregory, Kitterman, Phibbs, Tooley, & Hamilton, 1971), less aggressive mechanical ventilation with lower inspired oxygen concentration, and more careful monitoring of oxygenation (Kinsey & Hemphill, 1955). In line with this progress, the clinical characteristics of BPD began to change, and the term “old BPD” was established for the disease described by Northway et al., while the term “new BPD” was adopted to refer to the disease in the post-steroid and post-surfactant era (A. J. Jobe, 1999). Compared to infants with “old BPD,” those with “new BPD” are more premature, had no or mild RDS that responded rapidly to surfactant therapy, and have in general a less intense but prolonged need for respiratory support. At autopsy, lung histology is predominantly
characterized by impaired alveolar and vascular growth. Radiographic changes are not apparent until later in the course and are less striking compared to those observed in infants with “old BPD” (Coalson, 2006). Thus, the course of BPD no longer follows the progressive clinical and radiographic stages originally described by Dr. Northway et al.

As the clinical and pathophysiological characteristics of BPD changed, the definition and diagnosis of BPD also evolved. In 2001, the United States National Institute of Child Health and Human Development (NICHD) published the diagnostic criteria for BPD (A. H. Jobe & Bancalari, 2001). These criteria are based on the need for respiratory support at 28 days of age and at 36 weeks post-menstrual age (PMA) (Table 1). In contrast to the older definitions, positive pressure support without supplemental oxygen is also now considered.

| Table 1. Diagnostic criteria for BPD according to NICHD (A. H. Jobe & Bancalari, 2001). |
|----------------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Age of the child | Mild BPD | Moderate BPD | Severe BPD |
| 28 days | FiO2 > 0.21 or continuous positive pressure | FiO2 > 0.21 or continuous positive pressure | FiO2 > 0.21 or continuous positive pressure |
| 36 weeks PMA | No respiratory support | FiO2 < 0.30, no continuous positive pressure | FiO2 > 0.30 or continuous positive pressure |

BPD, bronchopulmonary dysplasia; FiO2, fraction of inspired oxygen; NICHD, National Institute of Child Health and Human Development; PMA, post-menstrual age

The NICHD criteria became the standard and are used in current studies, including studies I–IV, presented in this thesis. In addition to being widely used, the advantage of these criteria is that they group BPD by severity. Furthermore, in some studies, BPD that was moderate-to-severe based on NICHD criteria was associated with later pulmonary and neurodevelopmental problems (Ehrenkranz et al., 2005; Hines et al., 2017; Islam, Keller, Aschner, Hartert, & Moore, 2015). Moreover, in twin studies, moderate-to-severe BPD is significantly affected by genetic factors, whereas mild BPD is not (Lavoie et al., 2008).

In 2003, Walsh et al. published a physiological definition of BPD (Walsh, Wilson-Costello, Zadell, Newman, & Fanaroff, 2003). This definition is based on a standardized diagnostic test, the so-called oxygen reduction test, in which infants with a fraction of inspired oxygen (FiO2) of < 0.3 at 36 weeks PMA are subject to a timed stepwise reduction of oxygen to the level of room air. If baseline saturation values drop below 88%, the infant has failed the test and is diagnosed with BPD.
This physiological definition decreased the overall rate of BPD and reduced variation among centers (Walsh et al., 2004).

Even with the oxygen reduction test, the current definition and subsequent diagnostics of BPD remain unsatisfactory. The main reason for this is that the definition is purely operational; that is, it is based on the respiratory support provided rather than on the pathogenesis of the disease. Proximal and distal airway disease, pulmonary vascular disease, abnormal respiratory drive, and factors that affect chest wall mechanisms are different conditions that, at different grades and combinations, may lead to the need for respiratory support at 36 weeks PMA and, as such, to a diagnosis of BPD. Thus, BPD as we know it today is not a single disease, nor even a spectrum of disease, in the sense that it would result from a single pathophysiological process. Instead, “new BPD” is a combination of several chronic lung diseases (C. V. Lal & Ambalavanan, 2017). This probably explains the other main limitation of the current definition, which is its low predictive value. Being diagnosed even with moderate-to-severe BPD reveals little about the risk of later respiratory illnesses (e.g., reactive airway disease, rehospitalization due to recurrent respiratory exacerbation, and exercise intolerance). There is also a growing recognition that preterm infants who did not receive a diagnosis of BPD at 36 weeks PMA are still at risk of respiratory morbidities and abnormal lung function during childhood and adolescence (Manuck, Levy, Gyamfi-Bannerman, Jobe, & Blaisdell, 2016). Moreover, again because of the operational character of the current definition, the diagnostics of BPD may vary among physicians and institutions, further confirming the need for more precise diagnostic criteria.

There are several ongoing efforts to develop a more precise definition of BPD. A recent study identified the need for respiratory support (i.e., supplemental oxygen and/or positive pressure) at 40 weeks PMA as the best predictor of serious respiratory morbidity and a good predictor of neurosensory morbidity at the age of 18–21 months (Isayama et al., 2017). An interdisciplinary collaborative has proposed dividing severe BPD into two subgroups in accordance with the severity of symptoms. According to this proposal, infants who require ongoing respiratory support beyond 36 PMA—and who thus are at the highest risk of severe complications including pulmonary hypertension, poor growth, and neurodevelopmental problems—would be diagnosed with type 2 severe BPD (Abman et al., 2017). There has even been a proposal to replace the name BPD with “persistent pulmonary insufficiency of prematurity.” This new designation would cover not only the initial prolonged respiratory support, but also the increased risk of respiratory illness later during childhood and the increased
susceptibility to chronic lung disease, even in adulthood (Abman, Bancalari, & Jobe, 2017). A more precise definition would facilitate improvements in the design not only of clinical trials, with the subsequent possibility of selective strategies in clinical care, but also of investigations into BPD pathogenesis that include genetics (Poindexter et al., 2015).

2.1.2 Incidence

Unlike most other neonatal morbidities, the overall incidence of BPD has not decreased in the last few decades. Instead, it has remained constant or even increased, reflecting the improved survival of the most immature infants (Costeloe et al., 2012; Latini, De Felice, Giannuzzi, & Del Vecchio, 2013; Stensvold et al., 2017; Stoll et al., 2015). In some studies, the incidence of BPD has however decreased (Ancel et al., 2015).

The incidence of BPD is inversely related to gestational age. The severity of BPD is also related to immaturity: Infants born at the youngest GA most frequently have the most severe form of the disease.

<table>
<thead>
<tr>
<th>GA at birth, weeks</th>
<th>Incidence of moderate-to-severe BPD (%)</th>
<th>Incidence of severe BPD (%)</th>
<th>n</th>
<th>Study period</th>
<th>Country</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>69</td>
<td>NA</td>
<td>801</td>
<td>2008–2012</td>
<td>USA</td>
<td>NICHD Neonatal Research Network, (Stoll et al., 2015)</td>
</tr>
<tr>
<td>22–26</td>
<td>68</td>
<td>41</td>
<td>705</td>
<td>2006</td>
<td>Sweden</td>
<td>Express study, (EXPRESS Group, 2010)</td>
</tr>
<tr>
<td>23–26</td>
<td>NA</td>
<td>26</td>
<td>512</td>
<td>2011</td>
<td>France</td>
<td>Epipage-2, (Ancel et al., 2015)</td>
</tr>
<tr>
<td>27–31</td>
<td>NA</td>
<td>5</td>
<td>2633</td>
<td>2011</td>
<td>France</td>
<td>Epipage-2, (Ancel et al., 2015)</td>
</tr>
<tr>
<td>22–26</td>
<td>NA</td>
<td>42</td>
<td>185</td>
<td>2013–2014</td>
<td>Norway</td>
<td>(Stensvold et al., 2017)</td>
</tr>
</tbody>
</table>

BPD, bronchopulmonary dysplasia; GA, gestational age; NA, information not available; NICHD, National Institute of Child Health and Human Development.
The studies in Table 2, as well as most of the other recent publications on the incidence and risk factors of BPD, included both singletons and twins. The literature about BPD in twins is discussed later in this review and investigated further in study IV, presented herein.

2.1.3 Pathogenesis and risk factors

The current understanding of the molecular pathogenesis of BPD is incomplete. This is probably due to its pathological heterogeneity, which is discussed in detail in 2.1.1. In addition, it is challenging to study BPD pathogenesis because of the lack of relevant clinical samples (i.e., a series of lung biopsies from preterm infants). Animal studies have increased our understanding of the relationship between the clinical risk factors observed in epidemiological studies and BPD. However, many of the animal models differ from humans in terms of the immune system and lung structure, and most of the models emphasize just one of the contributors to lung injury, such as hyperoxia or intrauterine infection (Hilgendorff, Reiss, Ehrhardt, Eickelberg, & Alvira, 2014). In addition, despite intensive research, there are still several gaps in our understanding of normal lung development (Hagood & Ambalavanan, 2013).

A generally accepted theory posits that BPD has a multifactorial pathogenesis; that is, BPD develops in a genetically susceptible individual as the result of multiple predisposing pre- and postnatal conditions and events (Figure 1). According to a growing pool of data, BPD development begins in utero, where, in addition to the fetal response to proinflammatory stimuli, angiogenetic factors affecting the function of the placenta, play a central role. Pre-eclampsia, other maternal hypertensive disorders, and maternal smoking are examples of these factors clinically commonly evident as restricted fetal growth. In the postnatal pathogenesis of BPD, the three main processes are 1) impaired normal growth and development of the lungs, including vasculature; 2) injury and prolonged inflammation in the lungs caused by mechanical ventilation, oxidative stress, and infections; and 3) repair processes associated with the injury. The inflammatory processes probably potentiate abnormal angiogenesis, and vice versa. Furthermore, genetics direct and modify all reactions in a living organism, and BPD, according to twin studies, is a particularly hereditary syndrome. Thus, the pathogenesis of BPD is further complicated by interindividual differences in response to stimuli involved in the processes.
The immature lung

Despite the variations in phenotype, there is at least one common factor in the development of BPD: immaturity. GA is one of the most established risk factors for BPD, and the relationship is inverse; the lower the GA, the higher the risk of developing BPD. Lung development is a highly complex process that can be divided into five partially overlapping stages: embryonic, pseudoglandular, canalicular, saccular, and alveolar (Figure 2). At 23–25 weeks GA, the lungs are in the late canalicular stage of development. At that stage, the epithelium is differentiating to form a primitive blood-gas interface and surfactant production in the type II alveolar cells starts. There is also an extensive vasculogenesis with marked capillary proliferation and remodeling of the interstitial extracellular matrix. In the following saccular stage of the development (at 24–36 weeks GA), the primitive terminal airspaces form via further thinning of the interstitium between airspaces and vessels and a continued vascular expansion. Immediately afterwards, the alveoli, the final gas-exchange unit of the lungs, begin to develop. Alveolarization, however, is immature in the beginning and matures via elastogenesis and secondary septation, forming new alveoli during the first 2–3 years of life. The capillary network also continues to grow and remodel during the
alveolar stage (Baker & Alvira, 2014; Coalson, 2006; Morrisey & Hogan, 2010; L. J. Smith, McKay, van Asperen, Selvadurai, & Fitzgerald, 2010). Thus, being born very premature (i.e., <32 weeks GA) interrupts the normal growth and development of both alveoli and distal pulmonary vasculature. In BPD, these processes are impaired, with alveolar simplification and dysmorphic vasculature as a result.

Dysregulation of growth factors

The normal growth and development of alveoli and pulmonary vasculature are closely synchronized and regulated by multiple growth factors. In fact, the endothelial-derived angiocrine factors in pulmonary vessels promote and sustain normal alveolar growth and development (Jakkula et al., 2000; G. Seedorf et al., 2016). This discovery, together with the finding that these processes are impaired in BPD, led to the so-called vascular hypothesis of the pathogenesis of BPD (Abman, 2001). Abnormal tone and reactivity in pulmonary vasculature are general features of BPD, even if pulmonary hypertension that can be measured by standard cardiac echography is common only in the most severe cases of BPD (Ambalavanan & Mourani, 2014).

Vascular endothelial growth factor (VEGF-A or simply VEGF) is one of the main mediators of angiocrine signaling between endothelial and epithelial cells in the developing lung. VEGF is expressed in distal airspace epithelial cells in both the fetal and postnatal lung. VEGF stimulates endothelial mitogenesis and migration and differentiation, enhances vascular permeability, and increases the mobilization of endothelial precursor cells from the basement membrane. The
dosage, timing, and spacing of VEGF are strictly regulated. Disturbances in the VEGF pathway lead to vascular disaster of varied grades, from embryonic lethality after disruption of even a single allele of *vegf* in mice (Carmeliet *et al*., 1996), to retinopathy of prematurity (ROP), where anti-VEGF (bevacizumab among others) therapy inhibits pathological intravitreal neovascularization, the leading cause of childhood blindness (Mintz-Hittner, Kennedy, Chuang, & BEAT-ROP Cooperative Group, 2011).

VEGF has two receptors, of which vascular endothelial growth factor receptor 2 (VEGFR-2) is the major mediator of angiogenetic effects. VEGFR-2 is a transmembrane tyrosine kinase receptor expressed in the fetal pulmonary arterial endothelium (Yancopoulos *et al*., 2000). *vegfr2* function is also crucial for survival of mouse fetuses (Shalaby *et al*., 1995). Soluble fms-like tyrosine kinase 1 (sFlt-1) is a shorter version of VEGFR-1, the other proangiogenic receptor of VEGF. sFlt-1 results from an alternative splicing of *VEGFR1* and lacks the transmembrane and intracytoplasmic domains (and thus function) of VEGFR1. sFlt-1 acts as a decoy receptor for VEGF, which results in an antiangiogenic effect (Liu, Afink, & Dijke, 2012).

There is a large body of evidence that confirms the role of VEGF in the pathogenesis of BPD. Briefly, in animal studies, VEGF inhibition disturbs microvascular development and alveolar septation (Le Cras, Markham, Tuder, Voelkel, & Abman, 2002), and recovery after pulmonary injury caused by hyperoxia can be improved by rhVEGF-therapy (Kunig *et al*., 2005). In humans, low cord blood levels of VEGF are associated with BPD (Mestan *et al*., 2017). In lung biopsies from infants who died of BPD, expression of *VEGF* and *VEGFR-1* (Bhatt *et al*., 2001) and VEGFR-2 protein levels (Janer, Andersson, Haglund, Karikoski, & Lassus, 2008) were lower compared to infants who died of nonpulmonary causes.

Although much less studied, there is also evidence that the growth factor Kit ligand (also known as stem cell factor or mast cell growth factor) is dysregulated in BPD. Kit ligand is also involved in hematopoiesis, and it is the main growth factor for mast cells (Roskoski, 2005). Binding of Kit ligand to its main receptor, c-kit, on the surface of mast cells leads to activation of multiple pathways that influence mast cell proliferation, differentiation, and activation, with consequent release of a number of proinflammatory cytokines and chemokines (El-Agamy, 2012; Tsai *et al*., 1991). Kit ligand is upregulated in inflammatory conditions; this finding has been observed in humans and in mice (Reber, Da Silva, & Frossard, 2006). In humans, Kit ligand is expressed in various pulmonary cells and has been
coupled with pulmonary inflammatory conditions, such as asthma (Da Silva, Reber, & Frossard, 2006). Mast cells have been found in excess in human infants dying of BPD (Bhattacharya et al., 2012; Lyle, Tryka, Griffin, & Taylor, 1995).

In addition to VEGF and Kit ligand, several other growth factors play important roles in normal lung development and have been shown to be involved in the pathogenesis of BPD, albeit mainly in animal studies. Platelet-derived growth factor (PDGF), a driver of alveolar septation, and transforming growth factor beta (TGF beta), associated with apoptosis, matrix remodeling, and inflammation, are examples (Bourbon, Boucherat, Chailley-Heu, & Delacourt, 2005). Beyond the individual effects of these and other growth factors, there is growing recognition of associated networks and reciprocal signaling among the factors (Oak & Hilgendorff, 2017). The reported interaction between nuclear factor kappa B (NF-κB) and VEGF is an example; inhibition of NF-κB leads to alveolar simplification and disruption of angiogenesis in a process carried out via the VEGF pathway (Josef et al., 2012). Another example is hepatocyte growth factor (HGF), also a downstream mediator of VEGF (G. J. Seedorf et al., 2016; Yamamoto et al., 2007). Administration of HGF increases microvasculature and improves alveolar simplification after lung injury in newborn mice (Ohki et al., 2009). In humans, HGF levels are decreased during the first 2 weeks of life in tracheal aspirates from infants who later developed BPD (Lassus, Heikkila, Andersson, von Boguslawski, & Andersson, 2003). Nitric oxide (NO), via endothelial nitric oxide synthase (eNOS), is yet another mediator of the VEGF pathway (Abman, 2010). Vascular density and airspace development are impaired and VEGF mRNA expression is reduced in the lungs of fetal mice deficient in eNOS (Han et al., 2004). Thus, according to the current understanding, normal pulmonary growth and development depends upon a network of several crucial growth factors. In BPD, many of these are disrupted.

Pulmonary inflammation as a response to oxidative stress, mechanical ventilation, and infection

There is comprehensive evidence of the presence of persistent inflammatory processes in the pathogenesis of BPD. In addition to animal studies, numerous studies on human neonates have shown increased levels of inflammatory cells and proinflammatory cytokines in infants subsequently developing BPD. The increased levels of cells (neutrophils, eosinophils and macrophages) and cytokines (interleukins 1 beta, 6 and 8; tumor necrosis factor alpha; interferon gamma;
granulocyte colony stimulating factor; monocyte chemoattractant proteins 1–3; and macrophage inflammatory proteins 1a and 1b) have in different human studies been detected in cord blood, plasma, tracheal aspirates and urine of infants later developing BPD (Balany & Bhandari, 2015; A. Bhandari & Bhandari, 2013). The recruitment of inflammatory cells leads to lung injury via release of tissue-damaging enzymes and cellular apoptosis. Several central signaling pathways in the lungs become dysregulated, such as the aforementioned NF-κB and TGF pathways. The result is impaired development of epithelial, mesenchymal, and endothelial cell structures (Hilgendorff et al., 2014).

Many injuries involved in the pathogenesis of BPD lead to pulmonary inflammation. Toll-like receptors (TLRs) are examples of mediators between the different inflammation-causing conditions and the inflammatory response. TLRs may be triggered by both endogenous (e.g., oxidized phospholipids) and exogenous (e.g., endotoxins) ligands (Satoh & Akira, 2016). Triggering of TLRs leads to activation of innate immunity, which, in turn, stimulates and reinforces adaptive immunity. The main causes of inflammation in the pathogenesis of BPD are oxidative stress, mechanical ventilation, and infection.

Very preterm birth exposes lungs to oxidative stress. The normal partial pressure of oxygen in the umbilical vein returning oxygenated blood from placenta is approximately 4.7 kPa, compared to 11–13.5 kPa in the arteries of healthy adults at normal atmospheric air pressure (Finnemore & Groves, 2015). Thus, even without a high FiO2 as part of the respiratory management strategy in the neonatal intensive care unit, all very preterm infants are exposed to relative hyperoxia and subsequent oxidative stress. In preterm infants, antioxidant mechanisms are inefficient and the balance between pro- and antioxidant forces tends to be towards pro-oxidant processes. Oxidative stress causes cellular damage and pulmonary fibrosis, with inflammation as one of its main mechanisms of action (V. Bhandari & Elias, 2006; Saugstad, 2003).

In addition to oxidative stress, pulmonary inflammation is promoted by lung injury caused by mechanical ventilation. Volutrauma (i.e., regional overdistension of alveoli and airways) is the major cause of ventilator-induced lung injury. Barotrauma and atelectrauma contribute (Donn & Sinha, 2003). As shown in healthy adults and in animal studies, large tidal volumes damage the pulmonary capillary endothelium, alveolar and airway epithelium, and basement membrane (Dreyfuss & Saumon, 1998; Hillman et al., 2011). This leads to activation of macrophages, with subsequent influx of proinflammatory cytokines and stimulation of vascular endothelial cells to express vascular adhesion molecules,
with subsequent transmigration of leukocytes from blood to interstitial and alveolar compartments (Attar & Donn, 2002; Hillman et al., 2011). The result is alveolar and interstitial edema, surfactant dysfunction, and a fibroproliferative response. In utero–ventilated fetal lambs, ventilation led to BPD-like changes in lung morphology (Allison et al., 2008). Gentle ventilation strategies with positive end-expiratory pressure (PEEP) slows down the development of edema and reduces the severity of tissue injury in rats (Dreyfuss, Soler, Basset, & Saumon, 1988). In human preterm neonates, early use of noninvasive nasal continuous positive airway pressure (NCPAP) reduces the occurrence of BPD (Fischer & Buhrer, 2013; Rigo, Lefebvre, & Broux, 2016; Schmolzer et al., 2013).

Infections, whether intrauterine, congenital, or postnatal, lead to cellular and cytokine inflammatory responses, resulting in histological features of impaired alveolar development in animal models (Ballard, Mallett, Pruszynski, & Cantey, 2016). *Ureaplasma* bacteria are one of the most studied infectious agents within the context of premature birth and BPD. *Ureaplasma* are detected in approximately 30% of infants with BW < 1500 g, and the colonization rate is inversely related to GA (Sung et al., 2011). According to experimental in vivo models, *ureaplasma*, among other mycoplasma species, elicit an inflammatory response in the lung, establish a chronic intrauterine infection, and contribute to the development of BPD (Viscardi & Kallapur, 2015). A recent meta-analysis found a significant association between *ureaplasma* colonization and BPD (odds ratio [OR] 2.22, 95% confidence interval [CI] 1.42–3.47) (Lowe et al., 2014). Furthermore, a systematic review of the use of azithromycin in neonates to prevent BPD found that azithromycin reduces the risk of BPD (relative risk [RR] = 0.83, 95% CI 0.71–0.97; number needed to treat [NNT] 10).

Chorioamnionitis (CA) is an inflammation of the chorion and amnion membranes of the placenta. CA is classified as either subclinical or clinical. In subclinical chorioamnionitis, there is a microscopically observable infiltration of inflammatory cells or a microbial invasion of the amniotic cavity. The signs of clinical chorioamnionitis include maternal fever, elevated white blood cell count, uterine tenderness, foul smelling vaginal discharge, and both maternal and fetal tachycardia (Hagberg, Wennerholm, & Savman, 2002). CA increases the incidence of preterm birth and several epidemiological studies have found an association with BPD (Hartling, Liang, & Lacaze-Masmonteil, 2012; Metcalfe, Lisonkova, Sabr, Stritzke, & Joseph, 2017). Other studies, however, found no such association (Ballard et al., 2016; Bersani, Thomas, & Speer, 2012; Torchin et al., 2017). Moreover, association analyses between cord blood cytokine levels and risk of BPD...
in infants born from pregnancies complicated by CA have yielded inconsistent results (Paananen et al., 2009). In one study, elevated cord blood concentrations of the proinflammatory cytokine interleukin 6 were more closely associated with BPD than elevated intra-amniotic interleukin 6 levels (Yoon et al., 1999). In animal studies, inoculation of the amniotic cavity with E. coli–derived lipopolysaccharide or live ureaplasma induces fetal inflammation and lung maturation. This inflammation leads to vascular remodeling and alveolar simplification. However, repetitive lipopolysaccharide exposure and/or chronic chorioamnionitis in experimental animals has been shown to lead to immune tolerance, with subsequent dampening of the inflammatory response and close to normal lung development (Kunzmann, Collins, Kuypers, & Kramer, 2013; Westover & Moss, 2012). Thus, the impact of chorioamnionitis in BPD pathogenesis may be strongly modified by the host response of the fetus, which may explain (at least in part) the inconsistency among epidemiological studies regarding the association between CA and BPD (Balany & Bhandari, 2015). The inconsistency may also be due to differences in definitions and whether potential confounding factors such as GA and intrauterine growth have been considered (Lacaze-Masmonteil, 2014).

Postnatal sepsis associates also with BPD (Stoll et al., 2002). Late-onset sepsis (i.e. sepsis occurring after 3 days of age) increased the risk of BPD (OR 2.74, 95% CI 2.54–2.94), and when the occurrence of infections decreased, the occurrence of BPD decreased as well (Lapcharoensap et al., 2015). In another recent study, one episode of late-onset sepsis increased the risk of BPD (adjusted OR 1.69, 95% CI 1.30–2.21), and two or more episodes of sepsis increased the risk further (adjusted OR 2.69, 95% CI 1.82–3.98) (Eriksson et al., 2015).

There are few studies of the airway microbiome in neonates. In one such study, there was an imbalance of the microbiota in the airways in newborns who subsequently developed BPD. The disturbance was speculated to prime the developing pulmonary immune system towards developing BPD and strengthens the hypothesis that airway colonization has already occurred by the time the infant is born (C. V. Lal et al., 2016). No studies on the microbiome in more distal lung tissue have been published to date.

In summary, persistent pulmonary inflammation is a typical response of premature infants to various BPD-predisposing conditions, of which hyperoxia, mechanical ventilation and infections are the most important. In susceptible individuals, inflammation sustains and leads to impaired alveolarization and simplified pulmonary microvasculature, the hallmarks of BPD.
Intrauterine growth restriction

Intrauterine growth restriction (IUGR), by definition, means that the growth of the fetus has been restricted by at least one pathological condition. By contrast, small for gestational age (SGA) also refers to a constitutionally small infant; that is, an infant whose low body BW in relation to the duration of gestation is not a result of a pathological condition but instead is due to below-average genetic growth potential. As such, these two definitions differ fundamentally from one another. Nevertheless, they are often (including in the studies presented in this thesis) used synonymously. To further obfuscate the picture, different definitions are used for IUGR/SGA. An international consensus statement defines SGA as an infant with BW < −2 standard deviations (SD), corresponding to the 2.3rd percentile of the mean BW. Another widely used definition is the one recommended by the World Health Organization (WHO), with a cut-off point of −1.28 SD, corresponding to the tenth percentile of the mean BW (Lee, Chernausek, Hokken-Koelega, Czernichow, & International Small for Gestational Age Advisory Board, 2003).

IUGR is one of the most established clinical antenatal risk factors for BPD (Bose et al., 2009; Eriksson et al., 2015; Henderson-Smart et al., 2006; M. K. Lal, Manktelow, Draper, Field, & Population-based study, 2003; Reiss, Landmann, Heckmann, Misselwitz, & Gortner, 2003; Torchin et al., 2016). IUGR commonly results from placental insufficiency, which is frequently caused by maternal conditions during pregnancy (e.g., pre-eclampsia or other hypertensive disorders, diabetes, smoking, use of alcohol, and viral infections). Placental dysfunction results in a limited supply of oxygen and nutrition to the fetus (Harding et al., 2000). In animal studies, protein deprivation leads to impaired alveolarization (Manuck et al., 2016). According to the well-known fetal origins hypothesis, fetal undernutrition may, via impaired programming, lead to persistent changes in postnatal growth and development (Barker & Osmond, 1986; Barker, 1998). In addition to the direct insufficiency of placenta, there is growing evidence of an association between placental dysangiogenesis and BPD. For example, maternal vascular underperfusion of placenta, which is a placental vascular abnormality related to maternal hypertensive disorders, was shown to predict the risk of BPD and pulmonary hypertension in BPD (Mestan et al., 2014). Thus, in addition to reduced function of the placenta in pregnancies complicated by pre-eclampsia and other maternal hypertensive disorders, the growth and development of fetal lungs in these pregnancies may also be impaired because of the anti-angiogenic intrauterine environment (possibly, among other factors, because of genetics).
Thebaud stated, “If your placenta doesn’t have it, chances are your lungs don’t have it either: The ‘vascular hypothesis’ of bronchopulmonary dysplasia starts in utero.” (Thebaud & Lacaze-Masmonteil, 2010).

2.1.4 Prevention and treatment

Until recently, caffeine was the only effective and safe means to prevent BPD (Schmidt et al., 2006). Respiratory management with early use of NCPAP (Fischer & Buhrer, 2013; Schmolzer et al., 2013) and less invasive surfactant administration (Aldana-Aguirre, Pinto, Featherstone, & Kumar, 2017; Rigo et al., 2016) was later shown to prevent BPD. Administration of systemic hydrocortisone also decreases the incidence of BPD (Baud et al., 2016) and has a promising safety profile (Baud et al., 2017). In addition, there are encouraging results from clinical trials that used mesenchymal stem cells to prevent BPD (Thebaud & Kourembanas, 2017). Other experimental therapies are under development. However, the most effective means of preventing BPD would probably be decreasing the prevalence of prematurity and IUGR, which are the most robust antenatal risk factors of BPD.

Caffeine

Because of the adverse outcomes associated with postnatal steroids, caffeine is the only effective and safe drug currently available to prevent BPD. Caffeine probably works by enhancing respiratory drive, but the compound also has anti-inflammatory and diuretic effects. In a RCT, 36% of infants who received caffeine developed BPD, compared to 47% in a placebo group (OR 0.63, 95% CI 0.52–0.76) (Schmidt et al., 2006). At follow-up at the corrected age of 18–21 months, infants who initially received caffeine had a lower prevalence of CP (cerebral palsy) and cognitive delay (Schmidt et al., 2007). At 5 years old, the difference between the groups was no longer significant (Schmidt et al., 2012), but at 11 years, children in the caffeine group had significantly less motor impairment (Schmidt et al., 2017). Current caffeine therapy starts substantially earlier than in the initial trial, and dosages are sometimes higher (Gentle, Travers, & Carlo, 2018).

Respiratory management

Early use of NCPAP, compared to intubation, prophylactic surfactant, and mechanical ventilation, decreases the incidence of BPD and the composite outcome.
of BPD or death (Fischer & Buhrer, 2013; Schmolzer et al., 2013). However, approximately 60% of very preterm infants develop RDS and thus would benefit from treatment with surfactant. Furthermore, so-called selective surfactant treatment should be given early in the course of RDS since, compared to treatment later on, early administration is associated with lower mortality and a lower incidence of BPD (Bahadue & Soll, 2012). This finding has led to the development of less invasive surfactant therapies. Of these, Less Invasive Surfactant Administration (LISA), in which surfactant is instilled intratracheally via a thin catheter to a spontaneously breathing infant under NCPAP has been shown to be slightly superior to the Intubate-Surfactant-Extubate (INSURE) strategy, in which the infant is intubated for endotracheal surfactant administration and, shortly thereafter, extubated back to NCPAP (Aldana-Aguirre et al., 2017; Rigo et al., 2016).

Hyperoxia is a robust clinical risk factor for BPD. The fraction of inspired oxygen is titrated according to the targeted oxygen saturation (spO2). The optimal level of spO2 has been investigated in large randomized controlled trials (RCTs) that included >5000 infants. In the SUPPORT (SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network et al., 2010) and BOOST-II (BOOST-II Australia and United Kingdom Collaborative Groups et al., 2016) studies, a lower spO2 level of 85–89% was associated with higher mortality despite a lower incidence of BPD compared to a higher spO2 level of 91–95%. However, the COT study did not find a difference in the rate of death or disability between the spO2 groups (Schmidt et al., 2013). In a meta-analysis of the studies, the lower target range for spO2 remained associated with higher mortality but without a lower incidence of BPD (Askie et al., 2017). Interestingly, in post-hoc analyses of the SUPPORT study data, the increased mortality in the lower target spO2 group was observed only in SGA infants (Walsh et al., 2016). These preliminary data peak towards development of personalized medicine, which, according to this example, could be performed even with the data that we already have.

**Corticosteroids**

Corticosteroids are administered to women who are at risk of premature delivery. A recent meta-analysis with almost 5000 infants confirmed the beneficial effects of corticosteroids: antenatal corticosteroids reduce the incidence of RDS, necrotizing enterocolitis (NEC), intraventricular hemorrhage (IVH), and neonatal death.
However, the study did not note any effect on the incidence of BPD (Roberts, Brown, Medley, & Dalziel, 2017). By contrast, postnatal corticosteroids effectively prevent BPD. However, especially when given during the first week of life, they are associated with both short- and long-term adverse effects. Cochrane database analyses found that dexamethasone given during the first week of life was associated with increased risks of CP and abnormal neurological examinations (Doyle, Ehrenkranz, & Halliday, 2014). In the same study, the adverse effects of hydrocortisone were weaker, but so were the beneficial effects. A recent RCT that compared low-dose, early administration (i.e., within the first week of life) of hydrocortisone with placebo found that 60% of infants who received hydrocortisone survived without BPD, compared to 51% of infants assigned to placebo (adjusted OR 1.48, 95% CI 1.02–2.16, NNT 12). No significant adverse events were noted, either during the initial period at the neonatal intensive care unit (Baud et al., 2016) or at 2-year follow-up (Baud et al., 2017). In the NEUROSIS RCT, early inhaled budesonide decreased the incidence of BPD. However, overall mortality in the budesonide group was higher than in the control group (Bassler et al., 2015). In the NEUROSIS follow-up study, mortality was still higher in the budesonide group at the corrected age of 18–22 months. Among the survivors, there were no differences in neurological status between the groups (Bassler et al., 2018). A recent meta-analysis of early inhaled corticosteroids compared with placebo found a lower incidence of BPD among survivors and a lower incidence of the composite outcome of BPD or death before 36 weeks PMA in the corticosteroid group (for the BPD outcome: RR 0.76, 95% CI 0.63–0.93; NNT to benefit one person [NNTB] 14). There were no statistically significant differences in short-term complications between groups and no differences in adverse events at long-term follow-up (V. S. Shah, Ohlsson, Halliday, & Dunn, 2017). Cochrane database analyses comparing inhaled and intravenously administrated corticosteroids, either before or after 8 days of age, in ventilated infants noted no significant differences in effectiveness or adverse event profiles (S. S. Shah, Ohlsson, Halliday, & Shah, 2017a; S. S. Shah, Ohlsson, Halliday, & Shah, 2017b). In another recent study, administration to infants with RDS of intratracheal surfactant and budesonide together compared to surfactant alone significantly decreased the incidence of BPD or death without immediate adverse effects (42% vs 66%, RR 0.58, 95% CI 0.44–0.77) (Yeh et al., 2016).
Stem cell and other experimental therapies

Stem cells have the capacity to self-renew and, depending on the type of stem cell, to differentiate into different kinds of cells. These properties may explain the ability of stem cells to take part in healing processes in organs after injury and are the rationale for studies of the therapeutic potential of stem cells in BPD, among other diseases.

Of the different types of multipotent stem cells, mesenchymal stromal cells (MSCs) have received the most attention within the context of BPD. MSCs have anti-inflammatory, antifibrotic, antiapoptotic, antioxidative, proangiogenic, and lung growth–promoting properties. Numerous studies in rodents have shown that MSCs prevent lung injury after a single intratracheal, intravenous, or intraperitoneal injection (Mobius & Thebaud, 2016). The therapeutic benefit of MSCs is mediated by paracrine activity (Fung & Thebaud, 2014), and recent preclinical studies have shown that cell-free conditioned media, probably because of the presence of extracellular vesicles or exosomes, have similar if not better lung protective effects than MSCs themselves (Kourembanas, 2015).

In humans, the first phase I study of MSCs comprised a single intratracheal injection within 5–14 days of life administered to nine infants at 23–29 weeks GA. This treatment was shown to be feasible and safe (Chang et al., 2014), and no adverse events were noted in the 2-year follow-up study (Ahn, Chang, Kim, Sung, & Park, 2017). The results of two more studies with a similar design are pending. In addition, there are ongoing investigations of therapies with other types of stem cells; for example, amnion epithelial cells, endothelial progenitor cells, and alveolar macrophages. Despite promising and potentially paradigm-shifting results, much remains to be learned about the biology and manufacturing processes pertaining to stem cell–based therapies before they can be used safely and effectively (Thebaud & Kourembanas, 2017).

Anti-sFlt-1 therapy is another experimental strategy that is currently under investigation. Administered both ante- and postnatally, anti-sFlt-1 therapy improved lung structure in an experimental rat model of BPD (Wallace et al., 2017). In addition, recombinant human insulin-like growth factor 1 (IGF-1) is also being investigated to treat or prevent BPD. IGF-1 is a major regulator of fetal growth and development of most organs, especially the central nervous system. Low serum concentrations of IGF-1 at postnatal days 3–21 correlate with BPD as well as several other neonatal morbidities (Hellstrom et al., 2016). According to preliminary data from an ongoing phase II study evaluating the effect of
recombinant human IGF-1 therapy, IGF-1 treatment significantly reduces the incidence of severe BPD (EAPS 2016, presented by David Ley et al.).

Vitamin A, inhaled nitric oxide, and azithromycin

Vitamin A and its metabolites are involved in lung development and repair of the respiratory epithelium. A recent meta-analysis compared vitamin A supplementation with placebo and found that vitamin A is associated with a small reduction in the incidence of BPD (RR 0.87, 95% CI 0.77–0.99, NNTB 11). There were no adverse effects, but the limited effect, together with the painful mode of administration of vitamin A (i.e., several intramuscular injections), makes its use less common in neonatal intensive care units (Darlow, Graham, & Rojas-Reyes, 2016).

Endogenous production of nitric oxide is needed for normal development of alveoli and pulmonary vasculature. Trials evaluating the effect of inhaled nitric oxide in preventing BPD have yielded inconsistent results. A systematic review found no effect of these therapies on the incidence of BPD (Donohue et al., 2011).

A systematic review of azithromycin treatment found that it significantly reduced the risk of BPD in preterm neonates (RR 0.83, 95% CI 0.71–0.98) (C. Smith et al., 2015). The review included three studies in which azithromycin was given intravenously during the first week of life followed by a 5-week therapy at a lower dosage. Four cases of infantile hypertrophic pyloric stenosis were noted; thus, further safety studies are warranted.

2.1.5 Outcome

Compared to their term-born peers, survivors of BPD have impaired respiratory function during early childhood, at school age, and during early adulthood. However, differences in lung function are less apparent and frequently insignificant between BPD survivors and prematurely born infants who did not have BPD (Caskey et al., 2016; Gough, Spence, Linden, Halliday, & McGarvey, 2012; Islam et al., 2015). This underlines the limited predictive value of the current diagnostic criteria for BPD. Interestingly, the main finding in lung function measurements in BPD survivors, even using the “new” BPD definition, has been airway obstruction. Restrictive impairment (i.e., reduction in lung volume and diffusion capacity) has been more modest and less consistent, suggesting that substantial catch-up growth occurs in the alveolar gas exchange surfaces of BPD infants (Islam et al., 2015). In
addition to impaired lung function, BPD survivors have recurrent respiratory infections. Rehospitalizations, particularly after infection with respiratory syncytial virus but also after infection with other common pathogens such as rhinovirus, are common during the first years of life (Chidekel, Rosen, & Bazzy, 1997; Kinsella, Greenough, & Abman, 2006).

Several studies have found associations between BPD and abnormal neurodevelopmental outcome (e.g., CP, cognitive delay, and educational and behavioural impairments). However, most of these studies were performed in the 1990s and used varying definitions of BPD. Furthermore, postnatal steroids, which were later found to be associated with neurodevelopmental problems, were widely used at that time (Doyle & Anderson, 2009). There are fewer recent studies, but most of them had similar findings: A study of >1000 infants (n = 603 with BPD, defined according to the physiological definition) with BW < 1000 g assessed the infants at 18–22 months corrected age. Moderate-to-severe CP, spastic diplegia, quadriplegia, and cognitive impairment were all significantly increased in BPD infants. In multivariable logistic regression analysis, BPD was independently associated with cognitive impairment (adjusted OR 2.4, 95% CI 1.40–4.13) (Natarajan et al., 2012). In another study using NICHD criteria for BPD and controlling for confounding factors, 75 infants with mild-to-severe BPD and 28 infants without BPD (mean GA 28 weeks, mean BW 1041 g) were evaluated several times before the corrected age of 12 months. Infants with severe BPD had higher incidences of neurodevelopmental delay throughout the study period, and after adjustment for confounding factors, BPD was found to have an independent adverse effect on developmental outcome (Jeng et al., 2008). Yet another study observed an association between the severity of BPD (diagnosed according to NICHD criteria) and CP and mental and psychomotor development (Ehrenkranz et al., 2005). However, a recent small retrospective study did not find that BPD had an independent effect on level of cognition (Brumbaugh, Colaizy, Patel, & Klein, 2018).

In summary, BPD is associated with abnormal lung function and neurodevelopmental impairment. However, at least in recent studies, there is not a great difference in the respiratory outcome between survivors of BPD and their premature peers who did not have BPD.
2.1.6 Twins and BPD

Twins comprise about 30% of very preterm infants (Anderson et al., 2016; Garg et al., 2010; Yeo et al., 2015). Twins differ from singletons in multiple aspects relevant to BPD. For example, they place a greater demand on placental function (Blickstein, 2005) and, in the case of the second-born twin, have less exposure to potential pathogens from the birth canal (Mazor et al., 1996; Phung et al., 2002; Romero et al., 1990). Furthermore, the incidence and risk factors of RDS, the other common respiratory disorder in preterm infants, differs between twins and singletons (Donovan et al., 1998; Marttila, Kaprio, & Hallman, 2004; Shinwell, Blickstein, Lusky, & Reichman, 2003).

Only a few studies have compared the occurrence of BPD in twins to that in singletons. One such study found no difference in the occurrence of BPD in twins and singletons (Donovan et al., 1998). In another study, the incidence of BPD in twins was lower than in singletons during one of the three study periods, while in the remaining periods there was no difference (Yeo et al., 2015). The third study found a lower occurrence of BPD in twins (Garg et al., 2010), but in the fourth study, the initial lower incidence of BPD in twins disappeared after controlling for relevant clinical factors (Papiernik et al., 2010).

For RDS, the incidence in first-born twins is lower than in second-born twins, at least among those born after 28 weeks gestation (Arnold, McLean, Kramer, & Usher, 1987; Hacking, Watkins, Fraser, Wolfe, & Nolan, 2001; Marttila et al., 2004). This is thought to be due to the accelerated lung maturity and the higher exposure to vaginal microbes for the first-born twin compared to the second-born twin (Mazor et al., 1996; Romero et al., 1990). No such difference between twins has been shown for BPD (Donovan et al., 1998; Hunter et al., 2017; Mei-Dan et al., 2017).

Knowledge of the antenatal nongenetic risk factors for BPD, also discussed in this review, comes mainly from studies that included both singletons and twins. Indeed, we found only one study in which the study subjects were exclusively singletons (Eriksson, Haglund, Odlind, Altman, & Kieler, 2014). The most established among the risk factors are low GA; low BW and low BW Z-score (i.e., low birth weight in relation to length of gestation); male gender; and, in most studies, maternal preeclampsia (Bi, Chen, & Huang, 2013; Bose et al., 2009; Eriksson et al., 2014; Eriksson et al., 2015; Gortner et al., 2011; Hansen, Barnes, Folkman, & McElrath, 2010; Henderson-Smart et al., 2006; Morrow et al., 2017; O'Shea, Davis, Doyle, & Victorian Infant Collaborative Study Group, 2012).
2.2 BPD as a complex disease

Genetic diseases and traits are traditionally divided into monogenic and polygenic types. In monogenic diseases, the disease results from modification in a single gene. In polygenic diseases, the disease is dependent upon modifications in several genes, with variable and sometimes very strong contributions from environmental factors. Because of this combination of multiple genetic susceptibility loci and the importance of environmental effects, polygenic diseases are also called multifactorial or complex diseases. BPD is a complex disease, as are most human diseases. (Hirschhorn, 2005).

2.2.1 Basic concepts in studying complex diseases

Common genomic variation

Genetically, humans are surprisingly similar; approximately 99.9% of our genomes are identical at the level of nucleotides. The remaining 0.1% varies, and the most common variation is replacement of a single nucleotide by a different nucleotide. If such a variant is common in a given population—that is, if the frequency at which the least common allele occurs within that population (known as the minor allele frequency, MAF) is > 0.01—the variant is called a single nucleotide polymorphism (SNP). Known SNPs are listed in databases (https://www.ncbi.nlm.nih.gov/projects/SNP/). Another type of common variation is insertions or deletions of small (< 50 nucleotides) segments of DNA. Yet another type is microsatellites, which are variations in the number of repeats of short DNA sequences in the large fraction of the human genome that consists of repetitive DNA sequences. In addition to these small-scale variations, moderately large-scale changes (e.g., copy-number variations, CNVs) in the DNA sequence are also common. (Lupski, 2007; Strachan, Goodship, & Chinnery, 2015).

Linkage disequilibrium

Linkage disequilibrium (LD) refers to a non-random association of alleles at two or more loci. These alleles are inherited together more often than would be expected by chance; that is, the frequency of the recombination rate between them is reduced. Generally, these alleles are in close physical proximity to each other. Because of LD, the genome consists of haplotype blocks, where the presence of one allele—
the so-called tagging allele or SNP (tSNP)—can be used to predict the presence of adjacent alleles. Thus, genotyping of a tSNP also reveals information about the other, nongenotyped SNPs in that haplotype. This property is used when designing candidate gene and genome-wide association (GWA) studies (Ardlie, Kruglyak, & Seielstad, 2002).

LD is closely linked to shared ancestral chromosomal segments: A new DNA variant is in tight disequilibrium with the adjacent alleles, but as time passes, recombination erodes the LD. Thus, the statistical reliability of the association between a given disease-susceptibility variant and the nearby genotyped marker SNP depends upon how many generations ago the mutation originated (Christensen & Murray, 2007). Newer populations, such as the population of Finland, display high LD (the high LD of the Finnish population is also attributable to its population history, which includes bottlenecks and small founder populations) (Varilo et al., 2000). Thus, the number of tSNPs required in studies based on common variants is lower for newer populations than when studying more ancient populations. Moreover, the chance that a specific variant associated with the disease is shared by different affected individuals is higher in a homogeneous population such as the Finnish population (Sajantila et al., 1996; Varilo et al., 2003; Varilo & Peltonen, 2004).

Role of common and rare variants in complex disease

Both common and rare variants likely play a role in complex disease. According to the common disease–common variant hypothesis (Lander, 1996), different combinations of common DNA variants increase the risk of complex diseases. The contribution of each of the DNA variants to the overall risk is small, and even a DNA variant that associates strongly with a complex disease is at best a susceptibility factor; the overall frequency of the variant is higher in affected individuals compared to unaffected individuals, but not all affected individuals have the variant while many unaffected individuals do have it (Strachan et al., 2015).

Because it takes time for a DNA variant to become common in a population, common variants are of ancient origin. One of the reasons why common variants have become common (i.e., were not eliminated by natural selection) is that many complex diseases are of late onset (i.e., they do not have an impact on reproduction). In addition, some of the common variants may have had beneficial effects in ancient environments. Common variants, typically, have very weak deleterious effects
Most of the candidate gene and GWA studies conducted are based on the common disease–common variant theory, since the arrays they use to genotype SNPs are composed of common variants.

In addition to the common disease–common variant hypothesis, the common disease–rare variant hypothesis (Pritchard, 2001) became established following the development of sequencing technologies in the late 2000s. According to the common disease–rare variant hypothesis, susceptibility to complex diseases is due mostly to a few DNA variants, each of which have a far greater impact on disease risk than previously estimated. Deletions, framshifts, and splice-site mutations are plausible effects of these variants. The frequency of the rare variants is much lower, with MAF < 0.01 for low-frequency variants and < 0.005 for rare variants (Cirulli & Goldstein, 2010). Hence, standard candidate gene or GWA studies would not capture them. There is a growing understanding that rare variants account for much of the functional variation in the human genome; thus, rare variants are of great importance to phenotypic variation and disease susceptibility (Fu et al., 2013; Tennessen et al., 2012). However, even acknowledging the importance of rare variants, it remains challenging to study the genetics of complex diseases. Because rare variants are rare and even population specific, very large sample sizes from single populations are required to capture them. For some diseases, such as age-related macular degeneration, sequencing of genes identified in GWASs has uncovered associated rare variants, but for other diseases such studies have not been successful (Saint Pierre & Genin, 2014; Strachan et al., 2015).

### 2.2.2 Methods to study complex diseases

The genetics of complex diseases are generally studied in case-control settings or in population cohorts. Candidate gene and GWA studies are the most commonly used methods. Both methods are population-based: that is, the included study individuals can be unrelated. Family-based linkage association studies, in which a genome-wide set of a few hundred or thousands of markers are genotyped in families with multiple affected individuals, have been less successful and are less commonly used to study complex diseases compared to monogenic diseases (Hirschhorn, 2005). No published study of BPD has used the linkage association method, probably due to a lack of families with several affected members. Rare variants can also be studied in candidate and GWA studies, but most of the studies conducted to date were designed to investigate only common variants.
In candidate gene and GWA studies, after genotyping, the frequencies of the minor alleles are compared between cases and controls. The independence of the associations is then tested in regression analyses with possibly confounding factors as covariates. Thereafter, the risk associated with each tested variant is calculated and expressed in ORs (i.e., the odds of being affected when possessing that certain variant divided by the odds of being affected when lacking that variant). Because of the risk of spurious associations, any associations found in a discovery population must be confirmed in independent replication populations. The impact of the association is strengthened if proof of the functionality of the associating gene can be found. The tested SNP is most reliably just a marker/locus that associates with the disease and with the disease-associating SNP (Figure 3).

![Fig. 3. Relationship between the tested SNP, the phenotype predisposing SNP and the phenotype under study. Modified from Terwilliger and Göring (Terwilliger & Goring, 2009)](image)

One of the challenges in association studies is thus how to link the associating SNP/locus with the true disease-associating variant and, most importantly, to the biologically relevant role of that variant (Lewis & Knight, 2012; Strachan et al., 2015). Pathway analyses are a method used for this purpose; they integrate and reduce the huge amount of genomic data into meaningful biological units (Kao, Leung, Chan, Yip, & Yap, 2017). Pathway analyses are particularly useful in studying complex diseases because they are more likely to identify variants in the same gene network than to discover variants in the same gene, especially if the variants to be discovered are different rare variants (Ramanan, Shen, Moore, & Saykin, 2012). Transcriptomics and proteomics are other approaches to study the function of specific genes in complex diseases. However, they are out of the scope of the present review.
**Twin studies**

Heritability ($h^2$) is the proportion of the variance of a phenotype that is attributable to genetic factors; that is, how much of the phenotype is due to genetics. Heritability values can be estimated from twin studies, and family or adoption studies can also be used. Twin studies compare the occurrence of the disease in monozygotic and dizygotic twins. This allows for separation of the genetic and environmental effects on disease development; monozygotic twins share all genetic information, while dizygotic twins share half of the genetic information, and they both share the intrauterine environment. This rationale is not without limitations; genes interact with the environment and, for example, post-zygotic mutations and epigenetic changes could affect only one monozygotic twin. However, greater disease concordance (i.e., greater within-pair similarity in monozygotic compared to dizygotic twin pairs) implies the presence of a genetic influence (Strachan et al., 2015).

The values of $h^2$ range from 0 to 1, with value 1 reflecting a situation in which the phenotype is due entirely to genetic factors. Typically, $h^2$ values for complex diseases are around 0.3–0.4. For BPD, the $h^2$ has been evaluated to be approximately 0.5–0.8 (V. Bhandari et al., 2006; Lavoie et al., 2008). The $h^2$ describes the genetic contribution to variance within a population and in a specific environment.

**Candidate gene studies**

Candidate gene studies are hypothesis driven; that is, the studied genes are selected because they are thought to be involved in the pathogenesis of the disease, either based on previous studies in humans and in experimental systems (e.g., cells and animals), or on theories of the pathobiology of the disease. In candidate gene studies, the genotyped SNPs are chosen from the HapMap dataset in the representative ethnic population. The SNPs are selected either because they are tSNPs (i.e., they also capture information about the other SNPs in the haplotype), because they are nonsynonymous (i.e., their variation changes the amino acid composition in the translated protein), or because they have some other known or anticipated biological function.
**Genome-wide association studies**

GWASs are association studies in which typically hundreds of thousands of common SNPs positioned throughout the genome are genotyped and compared individually between cases and controls. Microarray-based platforms are used for genotyping. In addition to genotyped SNPs, variants available from large-scale studies and from databases are also commonly tested; this procedure is called imputation. Because of the risk of false-positive associations due to multiple testing, a correction on the significance level needs to be made. By convention, the significance level is set to $5 \times 10^{-8}$ (Dudbridge & Gusnanto, 2008). Another common way to correct for multiple testing, which we used in the studies presented herein, is Bonferroni correction, in which the conventional $p$ value of 0.05 is divided by the number of variants tested.

**Whole-genome and exome-sequencing studies**

Sequencing determines the precise order of nucleotides in DNA. There are two main methods of sequencing: the standard dideoxy or Sanger sequencing, and the newer massively parallel DNA sequencing, which is also known as next-generation sequencing. In contrast to the Sanger method, next-generation sequencing can simultaneously sequence DNA fragments without electrophoresis. Next-generation sequencing is much faster and less expensive and has particularly enabled the widespread use of exome sequencing, including in clinical settings. (Grody, Thompson, & Hudgins, 2013).

Exome is a collective term for all exons in protein-coding genes. Exons are parts of the gene that encode the RNA molecule after introns have been removed by RNA splicing. Exome comprises 1–2% of the entire genome. Exome sequencing sequences exomes and, depending on the kit, additional regions of interest such as promoters and microRNA sites. Exome sequencing was first used to identify disease-associating genes in 2009. As costs have decreased, it has become widely used, even in clinics. Exome sequencing has successfully identified genes that underlie very rare autosomal recessive and congenital dominant disorders. It has also been used successfully in diseases with very heterogeneous phenotypes, such as neuropsychiatric diseases (Gilissen, Hoischen, Brunner, & Veltman, 2012; Majewski, Schwartzentruber, Lalonde, Montpetit, & Jabado, 2011).

With improved cost efficiency, whole-genome sequencing will probably replace exome sequencing, even in clinics. However, whole-genome sequencing...
generates 100 times more data compared to the already overwhelming amount of data obtained by exome sequencing. Thus, far more extensive bioinformatics analyses and storage facilities will be required (Grody et al., 2013; Majewski et al., 2011).

2.2.3 Twin studies of the heritability of BPD

Evidence for the strong heritability of BPD comes primarily from two twin studies. Earlier, however, Parker et al. studied 108 twin pairs of infants with BW <1500 g. BPD was defined as the need for supplemental oxygen at 28 days of age. The BPD status of the first-born twin was a significant predictor of the BPD status of the second-born twin (adjusted OR 12.3, \( p < 0.001 \)) (Parker, Lindstrom, & Cotton, 1996). Ten years later, in the first of the twin studies leading to the current heritability estimate of BPD, Bhandari et al. studied 450 twin pairs. Of these pairs, data on zygosity, determined by histopathology of the placenta, was available for 63 monozygotic and 189 dizygotic twin pairs. BPD was defined as the need for supplemental oxygen at 36 weeks PMA. The mean GA of the whole study population was 29.1 weeks, and the mean BW was 1290 g. In this retrospective study, the observed concordance for BPD among monozygotic twin pairs was significantly higher than the expected concordance; 12 out of 18 monozygotic twin pairs with one affected member had both members affected versus the 3.69 pairs expected. After controlling for significant risk factors (BW, male gender, RDS, and the treating institution), 53\% (95\% CI 16–89\%) of the observed variability in the incidence of BPD was directly attributable to genetic differences (i.e., the heritability of BPD was 0.53) (V. Bhandari et al., 2006). Two years later, Lavoie et al. compared the concordance of BPD (defined according to NICHD criteria) in a study of 478 twins. Of these, zygosity data, determined either by chorionicity in early fetal ultrasound scans or by placental histology, was available for 70 monozygotic and 89 dizygotic twin pairs. The mean GA for the whole study population was 28.0 weeks, and the mean BW was 1135 g. Also in this study, the intra-pair similarity in the incidence of moderate-to-severe BPD was greater among monozygotic twin pairs than among dizygotic twin pairs. According to model-fitting analyses, genetic effects accounted for 79\% (95\% CI 67–93\%) of the observed variance in moderate-to-severe BPD. By contrast, no genetic influence was observed for the occurrence of mild BPD (Lavoie et al., 2008).
2.2.4 Candidate gene studies of BPD

At date, around 70 candidate gene studies of BPD have been published. Most of them have investigated genes involved in inflammatory and antioxidant responses (V. Bhandari & Gruen, 2015; Lavoie & Dube, 2010; Shaw & O’Brodovich, 2013; Yu, Li, Snyder, Shaw, & O’Brodovich, 2016). TLR4 (Lavoie et al., 2012) and TLR6 (Winters et al., 2013) are examples of genes involved in the inflammatory response, as well as the genes encoding interleukin-18 receptor 1 (IL18R1) and its accessory protein (IL18RAP) (Floros et al., 2012). Examples from the antioxidant response system include NF-E2-related factor 2 (NFE2L2) and superoxide dismutases (SOD2 and SOD3) (Giusti et al., 2012; Sampath et al., 2015). Other genes associated with BPD in candidate gene studies are from processes linked to angiogenesis, such as the gene encoding nitric oxide (NO) and genes from the VEGF family (Fujioka et al., 2014; Kwinta et al., 2008). Some of the genes with positive associations in candidate genes studies were chosen not because of their plausible biological role in the pathogenesis of BPD, but because of a previous result in biomarker studies; for example, Kit ligand (KITLG) (Kaukola, Tuimala, Herva, Kingsmore, & Hallman, 2009).

2.2.5 Genome-wide associations studies of BPD

Before the present study III, three GWASs of moderate-to-severe BPD were published. Only one of these, the GWAS by Hadchouel et al., found an association at the genome-wide significance level. That study used a DNA pooling strategy in two discovery series, one with Caucasian neonates and the other with neonates of African ancestry. The total number of study infants in the discovery populations was 418 (43 BPD cases), the mean GA was 26.4 weeks, and the mean BW was 845 g. BPD was defined according to NICHD criteria. Fine mapping of the initial findings revealed an association between SNP rs1245560 in SPOCK2 and BPD: $p = 1.66 \times 10^{-7}$, with adjusted OR in Caucasian neonates 2.96 (95% CI 1.37–6.40) and adjusted OR in neonates of African ancestry 4.87 (95% CI 1.88–12.63). The association was confirmed in an independent Finnish replication population. Thereafter, SPOCK2 mRNA levels were found to be significantly increased during the alveolar stage of lung development in the lungs of newborn rats. In these rats, hyperoxia increased SPOCK2 expression compared to controls exposed to air. No significant association was found for mild BPD. In the Caucasian neonates, SNP rs1049269 in SPOCK2 was also associated with BPD (OR 3.21, 95% CI 1.51–6.82).
SPOCK2 is a member of the testican group of extracellular chondroitin and heparan sulphate proteoglycans. Its role in the lungs has not been explored, but since its expression is highest at the end of the alveolar stage, the authors speculate that it may be involved in termination of the septation process (Hadchouel et al., 2011).

In the second GWAS (Wang et al., 2013), one of the inclusion criteria was the requirement for mechanical ventilation during the first 3 days of life. This criterion was chosen, according to the authors, to minimize the effect of environmental factors on the development of BPD. However, and as also discussed by the authors, the requirement may not have increased the quality of the study, since some of the infants who later developed BPD did not need mechanical ventilation immediately after birth and the requirement makes the study different from other GWASs. The study population was mainly of Mexican–Hispanic origin, identified via genotyping. BPD was defined according to NICHD criteria. In the discovery set, the number of cases and controls was 899 and 827, respectively. The authors performed a GWAS, a subsequent imputation, an exome array, and a pathway analysis. No SNPs reached the formal significance level of $5 \times 10^{-8}$ in any of these analyses. Moreover, none of the previously BPD-associated SNPs showed an association. The SNP with the lowest $p$ value was intergenic SNP rs8089528 on chromosome 18 ($p = 8.64 \times 10^{-7}$). Another promising SNP was rs118078182 ($p = 1.30 \times 10^{-6}$), an intronic SNP in the gene encoding collagen type XXIII alpha 1 (COL23A1). The 5673 most promising SNPs were thereafter genotyped in a replication population consisting of 371 cases and 424 controls. No association was noted in the replication population either, or in the combined discovery and replication population (Wang et al., 2013).

The third GWAS of BPD was a combination of GWAS and subsequent pathway and gene expression analysis (Ambalavanan et al., 2014) that included 751 infants with a mean GA of 25.8 weeks and a mean BW of 758 g. Most of the study infants were of Caucasian or African-American ethnicity, as determined by genotyping. NICHD criteria for BPD was used. The authors ran three models: BPD or death vs survival without BPD, severe BPD or death vs survival without severe BPD, and severe BPD vs survival without BPD. No SNPs were associated with any of the outcomes at the genome-wide significance level. The top ten SNPs with $p$ values of $10^{-6}$ to $10^{-7}$ were thereafter mapped to genes, the genes were assigned to pathways ($n = 7650$), and the pathways were analysed by gene set enrichment analysis. Gene set enrichment analysis revealed that associating pathways (i.e., pathways with a false discovery rate of $<0.1$ [suggesting an approximately 10% rate of false positives] and $p < 0.001$) differed among the different outcomes to the
extent that only three pathways were shared by all outcomes. Thus, according to this study, the pathways involved in severe BPD differed from those involved in mild/moderate BPD. For the outcome BPD/death, the most associating pathway was miR-219 targets. MiR-219 is involved in the resolution of acute inflammation. For the other outcomes, severe BPD/death and severe BPD in survivors, the pathway with the lowest false discovery rate was phosphorous oxygen lyase activity, which is known to be involved in, for example, lung injury and development. In addition, differences by race/ethnicity were observed in the associating pathways. Gene expression data for the six most promising genes were thereafter extracted from the pre-existing dataset of Bhattacharya et al., which describes genome-wide gene expression in lung tissue obtained from human BPD cases and controls (Bhattacharya et al., 2012). Expression of miR-219 and CD44 were increased in the lungs of infants with BPD compared to the lungs of preterm infants without BPD. Finally, hyperoxia increased levels of miR-219 and CD44 in a rat model of BPD.

2.2.6 Other genomic studies of BPD

To date, one study has investigated CNV, another common variation of DNA, and BPD. This study included the same study individuals as the GWAS of Wang et al. (2013); that is, only infants who required mechanical ventilation during the first 3 days of life were included. After quality control, there were 848 cases and 783 controls in the study. No differences in the number of CNVs between cases and controls were noted. In addition, no CNV was associated with increased risk of BPD (Hoffmann et al., 2014).

There are currently two published exome-sequencing studies of BPD. Li et al. (2015) included 50 twin pairs, among whom 51 individuals had BPD defined according to NICHD criteria. This study was performed by the same group of investigators who performed the Wang et al. (2013) GWAS; thus, one of the inclusion criteria was the need for mechanical ventilation during the first 3 days of life. The other criterion was GA between 25 and 29 weeks. Only rare variants with high conservation were analysed. Exome sequencing identified 57,535 nonsynonymous (i.e., leading to an altered amino acid sequence) variants (missense or loss-of-function). These variants were located in 258 and 182 nonoverlapping genes in BPD and non-BPD infants, respectively. Pathway analyses revealed that the genes associated with BPD were enriched in processes involved in pulmonary structure and function, morphogenesis of embryonic epithelium, and regulation of
the Wnt signalling pathway. Furthermore, mice with mutations in the identified genes showed phenotypic similarities with human BPD. Finally, the affected genes were upregulated in human lungs and in rat pups exposed to hypoxia. Thus, this study provides new information on the molecular mechanisms of the pathogenesis of BPD, based on the plausible role of rare variants in the genetics of BPD (Li et al., 2015).

The exome-sequencing study of BPD by Carrera et al. was a pilot study (Carrera et al., 2015). This study also used NICHD criteria for BPD, but, because of the hypothesized stronger genetic component, only infants with severe BPD were included. Thus, there were no control infants in this study. The study infants were of Caucasian ancestry and had a mean BW of 778 g. After quality control, 21 of the 26 genotyped infants were analysed. The analyses focused on genes previously associated with BPD with a special emphasis on rare variants, and on new candidate genes based on pathway analyses performed after sequencing. Among the most promising genes was the gene encoding C-reactive protein (CRP). Other highlighted genes encoded NOS2 and TLR family members. Thus, the results of this pilot study reinforced the role of inflammation in the pathogenesis of BPD and the role of rare mutations in susceptibility to BPD (Carrera et al., 2015).

Finally, there are two published genome-wide gene-expression studies of BPD. Bhattacharya et al. (2012) analysed gene expression in samples of distal human lung tissue and then performed pathway analysis. The study population consisted of 11 infants who had died of BPD and who had had a GA of approximately 27 weeks at birth, and nine deceased control infants who had had a mild lung pathology (e.g., pneumonia or RDS). The infants were matched for GA at birth and at death. According to analyses of isolated RNA, 159 genes were expressed differentially in BPD lungs compared to control lungs. Expression of the most promising genes was confirmed by qPCR. In pathway analyses of these genes, processes significantly associated with BPD included, among others, cell cycle and immune cell regulation. Three of the top five most highly induced genes in BPD lungs were specific to mast cells. This suggested increase in mast cells in BPD lungs was confirmed by immunohistochemistry, which showed an accumulation of chymase-expressing connective tissue mast cells in the BPD lungs, as well as in a replication set. Finally, increased expression of markers specific to these mast cells were demonstrated in a mouse model of BPD (Bhattacharya et al., 2012). Thus, according to this study, an accumulation of connective tissue mast cells is a significant feature of the pathology of BPD.
In another genome-wide gene-expression study, Pietrzyk et al. (2013) compared gene expression in the peripheral blood mononuclear cells of 68 BPD infants and 43 control infants at the ages of 5, 14, and 28 days, separately. The study used the NICHD criteria for BPD. The mean GA was 27.8 weeks, and the mean BW was 1029 g. The number of differentially expressed genes was 2086 on day 5, 324 on day 14, and 3498 on day 28. The differences in gene expression were generally low: for most genes, it was between 1.0- and 1.5-fold. In BPD infants, the genes were generally downregulated. Pathway enrichment analysis found that in BPD infants, the cell cycle pathway was upregulated without an association with GA. Genes assigned to this pathway with the highest fold-change values were encoded members of the cyclin family (CYCA, CYCE, and CYCB) and cyclin related kinase 1 (CDK1). Four inflammation-related pathways were continuously downregulated in BPD infants; however, the differences were dependent upon GA and the severity of respiratory symptoms at the time of sampling. Of these pathways, the T cell receptor signaling pathway was the most significantly downregulated (Pietrzyk et al., 2013).

In summary, recent large-scale genomic studies have identified molecules and pathways that could plausibly be involved in the pathogenesis of BPD. However, the results of these studies, as well as the preceding candidate gene studies and GWASs, have been difficult to replicate.

2.2.7 Epigenetic studies of BPD

Epigenetic alterations affect the function of the genome without structural changes in DNA. DNA methylation, histone modification, and microRNA changes are examples of epigenetic mechanisms that regulate gene expression. Of these, methylation is the most studied within the context of lung development (Jaenisch & Bird, 2003). Generally, hypomethylation leads to gene expression and hypermethylation to gene silencing. Nutrition, oxidative stress, and infection influence the expression of genes involved in prenatal and postnatal lung development (McEvoy et al., 2014). SOD3, which encodes a member of the superoxide dismutase protein family, protects lung and other tissues from oxidative stress. Among other genes, SOD3 may be regulated by methylation during normal alveolar septation in mice and in humans and may be differentially regulated in infants with BPD (Cuna et al., 2015). Moreover, in utero tobacco exposure was shown to modify central biological pathways in newborns via an altered methylation pattern (Rotroff et al., 2016). Interestingly, epigenetic changes are
retained after mitosis in the daughter cells of the cell in which they originally occurred. Thus, epigenetic modifications caused by environmental changes can result in persistent (at least in the short-term) changes in cell function. There is a growing body of evidence that epigenetics, an important mechanism in cell differentiation and the development of tissues and organs, may also play a role in the pathogenesis of BPD (Jaenisch & Bird, 2003; Z. D. Smith & Meissner, 2013).
3 Aims of the study

The purpose of the present study was to provide improved understanding about both genetic and nongenetic predisposition to moderate-to-severe BPD. Given the multifactorial nature of BPD, our genetic studies were designed primarily to enhance the current knowledge of BPD pathogenesis. We also investigated the nongenetic, antenatal risk factors, particularly in twins, who comprise a substantial proportion of very preterm infants but who have been scarcely studied separately. The specific aims of the study were:

1. To determine if common polymorphisms of the genes encoding VEGF and its receptor VEGFR-2 are associated with BPD (study I).
2. To investigate whether common polymorphisms of the gene encoding Kit ligand are associated with BPD (study II).
3. To search for new genomic regions associated with BPD (study III).
4. To compare the risk of BPD between twins and singletons, and to determine the antenatal, nongenetic risk factors of BPD separately for twins and for singletons (study IV).
4 Methods

4.1 Design

Of the present studies, studies I–III are genetic. Studies I–II are candidate gene studies, study III is a GWAS. The genetic studies are prospective cohort studies. Study IV is a population-based epidemiologic study.

In the genetic studies, the initial genotyping, whether of SNPs of certain genes or at genome-wide level, was performed in so-called discovery populations. Thereafter, the most promising SNPs were analysed in additional cohorts of infants, the so-called replication populations. In studies with positive genetic association also in replication populations (II–III), the possible function of the most probably associated gene was tested by comparing the level of the gene product in infants later developing BPD and in those who would not.

In the epidemiologic study (study IV) the occurrence of BPD was calculated in singletons and in the following subgroups of twins: All twins, first-born twins and second-born twins, separately. The risk of BPD was thereafter compared between singletons and different subgroups of twins, respectively. In addition to moderate-to-severe BPD, also a composite outcome of BPD or death before 36 weeks PMA, was assessed. Finally, the plausible antenatal risk factors were assessed towards both outcomes in singletons and in different subgroups of twins.

4.2 Study infants

Study populations are summarized in Table 3. For the genetic studies in this thesis project, a DNA collection named GenBPD was established in all five Finnish University Hospitals in 2010. Also DNA-samples from infants born at the same centres in 1997–2010 were used. In addition, infants born in Canada, France and Hungary in 1995–2009 were included. Thus, the study individuals in the genetic studies were born between 1995 and 2015 (Table 3).
Table 3. Study populations in present studies I–IV.

<table>
<thead>
<tr>
<th>Study</th>
<th>Study population</th>
<th>Origin of the study population</th>
<th>N cases/controls</th>
<th>Study period</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Discovery</td>
<td>Northern Finnish</td>
<td>44/116</td>
<td>1997–2010</td>
</tr>
<tr>
<td></td>
<td>Replication</td>
<td>Finnish, Canadian, Hungarian</td>
<td>120/208</td>
<td>2006–2012</td>
</tr>
<tr>
<td>II</td>
<td>1st Discovery</td>
<td>Northern Finnish</td>
<td>56/197</td>
<td>1997–2010</td>
</tr>
<tr>
<td></td>
<td>3rd Discovery</td>
<td>Canadian</td>
<td>58/68</td>
<td>2006–2008</td>
</tr>
<tr>
<td></td>
<td>1st Replication</td>
<td>Finnish</td>
<td>41/241</td>
<td>2001–2012</td>
</tr>
<tr>
<td></td>
<td>2nd Replication</td>
<td>Hungarian</td>
<td>29/40</td>
<td>1995–2002</td>
</tr>
<tr>
<td>III</td>
<td>Discovery</td>
<td>Finnish</td>
<td>60/114</td>
<td>1997–2012</td>
</tr>
<tr>
<td></td>
<td>1st Replication</td>
<td>Finnish</td>
<td>105/221</td>
<td>1997–2015</td>
</tr>
<tr>
<td></td>
<td>2nd Replication</td>
<td>Canadian, French</td>
<td>123/265</td>
<td>1997–2009</td>
</tr>
<tr>
<td></td>
<td>3rd Replication</td>
<td>Finnish</td>
<td>31/198</td>
<td>1997–2015</td>
</tr>
<tr>
<td>IV</td>
<td>BPD outcome</td>
<td>Finnish</td>
<td>608/2839</td>
<td>2005–2013</td>
</tr>
<tr>
<td></td>
<td>BPD/death outcome</td>
<td>Finnish</td>
<td>1076/2839</td>
<td>2005–2013</td>
</tr>
</tbody>
</table>

1 Infants with moderate-to-severe BPD according to NICHD-criteria
2 Infants with no or mild BPD according to NICHD-criteria
3 Infants of Caucasian origin
4 Infants of African origin
5 Composite outcome of BPD or death before 36 weeks PMA
BPD, bronchopulmonary dysplasia; NICHD, National Institute of Child Health and Human Development; PMA, post-menstrual age

For the study IV, the data were drawn from Small Preterm Infants data file, which is part of the Medical Birth Register in Finland (Syntyneiden lasten rekisteri/Pienet keskoset) held by The Finnish National Institute for Health and Welfare (Terveyden ja hyvinvoinnin laitos, THL). The data file contains obstetrical and neonatal data from all live births in Finland with birth weight less than 1501g or with gestational age of less than 32 completed gestational weeks. The study infants in study IV were born between 2005 and 2013 (Table 3).

For practical reasons, inclusion criteria for GA varied between the studies; in study I it was set to 30 completed gestational weeks, in studies II–III to 31 and in study IV to 32 completed gestational weeks. The exclusion criteria for all the studies were major congenital malformations, since these could have had an impact on the development of lungs and thus the risk of BPD. In genetic studies, the study individuals in the discovery populations were of Finnish origin based on names (in GWAS, based on genetic markers), except from the Canadian discovery population in study II. In replication populations, the infants represented several ethnicities. The epidemiologic study was nation-wide and thus also infants with non-Finnish background were included. For GWAS i.e. the discovery population of the study
III, because of the limited sample size, the cases and controls were matched in regard to GA and IU-growth. In other stages of study III as well as in studies I and II, no matching was done.

The number of cases and controls in the studies were: Study I: 164/324; study II: 239/595; study III: 319/798 and study IV: 608/2839 for the BPD outcome and 1076/2839 for the composite outcome of BPD or death before 36 weeks PMA (Table 3). As expected, in all studies except for the abovementioned GWAS, infants with BPD were born with lower GA, with lower BW and more often as IUGR/SGA compared to the controls.

4.3 Selection of genotyped SNPs

VEGF/VEGFR2 and KITLG, the candidate genes in studies I and II, respectively, were chosen because of their plausible biological role in the pathogenesis of BPD. tSNPs were selected to capture the common variation within the investigated genes. In selection of SNPs, HapMap data (http://hapmap.ncbi.nlm.nih.gov/) with release 24/phase I & II for Utah residents i.e. for populations with ancestry from Northern and Western Europe was used. Cut-off values of 0.1 for MAF and 0.9 for $r^2$ (describing the LD) for SNPs were chosen. The number of SNPs selected for genotyping in respective discovery population was six for VEGFA, 25 for VEGFR2 and eight for KITLG. In candidate-gene studies, only SNPs associating with BPD in discovery populations, i.e. one SNP for each study, were further genotyped in replication populations.

In study III, after the quality control, 276 306 SNPs with MAF > 0.01 remained analysed in the discovery population. The sample size ($n = 60$ cases) in this initial GWAS was too small to expect statistically significant hits. Instead, the GWAS was used as hypothesis generating, and the most promising SNPs from it were selected for genotyping in the first internal replication population. Also SNPs associating, or showing a trend towards association with BPD in previous GWASs, were selected for first internal replication population. The study continued at this multistage approach i.e. only the most promising SNPs became genotyped in subsequent replication population. Finally, only four SNPs were left to be genotyped in the second and last internal replication population.
4.4 DNA sample preparation and genotyping

DNA was extracted mainly from umbilical cord blood (whole blood) or umbilical cord tissue according to standard protocols using commercial DNA extraction kits. If cord or cord blood was not available, DNA was extracted from saliva, buccal cells or tracheal aspirate specimens.

The SNP genotyping in studies I and II and in the replication populations in the study III, was performed either by using the Sequenom iPLEX Gold Assay (Sequenom, San Diego, CA, USA) at the Institute for Molecular Medicine Finland (FIMM) or by PCR-Restriction Fragment Length Polymorphism (RFLP) or by sequencing at University of Oulu. The genome wide genotyping in the study III was done by Illumina Infinium HumanCoreExome Bead Chip (Illumina, San Diego, CA, USA) at FIMM.

4.5 Functional analyses of the associating genes

Because of the observed association between BPD and SNPs from KITLG (II), and between BPD and SNPs close to the gene encoding CRP (III), the levels of Kit ligand and CRP, at serum and plasma, respectively, were determined and compared between infants later developing BPD and those surviving without. Also the possible association between the BPD-associating SNPs and Kit ligand and CRP, respectively, was tested.

The Kit ligand concentrations in umbilical cord blood were available from 120 very preterm Finnish infants (n = 35 cases, n = 85 controls). The quantitation had been done using an antibody-based microarray and reported as fluorescence units (Kaukola et al., 2009). For CRP, the plasma values from the first week of life were available from clinical charts to a subset of Finnish study infants (n = 73 cases, n = 202 controls). The maximum CRP level and the mean CRP level were defined for each study infant and used in analyses.

4.6 Statistical analyses

All studies in this thesis follow a general design of a case-control association study. Because of the known strong association between BPD and GA and IUGR, respectively (Laughon et al., 2011; Zeitlin et al., 2010), these variables were taken into account in the analyses, either as alternative phenotypes for the associating SNPs or as covariates in logistic regression analyses when calculating the OR:s for
the observed associations. In GWAS, the cases and controls were matched in regard to GA and IU-growth, and hence no adjustment was done.

In the genetic studies (I-III), the associations were determined as statistically significant differences in allele or haplotype frequencies between cases, i.e. infants who later will develop moderate-to-severe BPD, and controls, i.e. those who would develop only mild or no BPD at all. The associations were tested with Chi-square test or logistic regression in soft-wares designed to genetic analyses (Haplovlew version 4.2 and Plink version 1.07 or 1.09). Logistic regression in Plink or SPSS Statistics 20.0 was used to calculate ORs for the observed associations. Sample size in the candidate gene studies (I-II) were calculated with Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) (Purcell, Cherny, & Sham, 2003). Receiver operating characteristic (ROC) statistics was used to determine the BPD predictive value of Kit ligand concentration (II). The association between BPD and the level of Kit ligand in cord blood (as dichotomous Kit ligand concentration with cut off value defined as the maximal predictive value in ROC), and the association between BPD and KITLG genotypes were tested with logistic regression analyses in SPSS. In study III, two CRP markers were defined for each study individual: The highest CRP level and the mean CRP level during the first week of life. These markers were used as dichotomous (above/below median) variables in logistic regression analyses when testing the possible association between the levels of CRP and BPD, and between the levels of CRP and SNP genotypes.

In the epidemiologic study (IV), the risk of BPD was compared between singletons and twins with generalized estimating equation (GEE) with logit link function and with twin pairs clustered. The GEE method nestles the twin-pairs thus minimizing the impact of genetics on the risk of developing BPD. GEE was also chosen instead of traditional logistic regression analysis, because of the formal requirement of independency of the values in traditional logistic regression analysis. The risk of BPD between first- and second-born twins, and between first- and second-born twins and singletons, respectively, was compared with logistic regression analysis in SPSS. In the analyses of the potential risk factors, male gender, birth weight, birth weight Z-score, SGA and maternal preeclampsia were assessed in univariate analyses in singletons, in twins and in first- and second-born twins, separately. Also here, the analyses including both infants from a twin pair, were done with GEE-analysis, otherwise logistic regression was used. After univariate analyses, multivariate logistic or GEE analyses were performed. The composite outcome of BPD or death before 36 weeks PMA was assessed similarly.
In the genetic studies, several SNPs were analysed in the discovery populations (number of SNPs in the study I: 31; in the study II: 8; and in the study III: 276 306). Since this could have led to false-positive associations, a multiple testing correction was performed. In the studies I and III Bonferroni testing were used resulting in significance threshold of $p = 0.0016$ in the discovery population of the study I, and $p = 1.8 \times 10^{-7}$ (i.e., $-\log_{10}(p) > 6.74$) in the GWAS of the study III. In the study II, the analysed SNPs displayed such a high LD that they could not be considered as independent units. Thus, multiple testing correction in this study was done with SNPsPd (Nyholt, 2004), which considers LD between SNPs. According to this correction, a $p$-value for significance in the study II was determined to $p = 0.014$.

4.7 Definitions

4.7.1 Definition of BPD in present studies

In the present studies, BPD was defined as a moderate-to-severe BPD according to NICHD criteria (see about BPD definitions in chapter Review of the Literature). Also respiratory support in form of high-flow 5–6 litre per minute via nasal probes was considered as positive pressure support. This criterion was chosen because 1) in the twins studies, moderate-to-severe BPD has been significantly affected by genetic factors whereas mild BPD has not (Lavoie et al., 2008), and 2) moderate-to-severe BPD has shown to be associated with later pulmonary and neurodevelopmental problems (Ehrenkranz et al., 2005; Hines et al., 2017). The control subjects in the studies were infants with mild BPD or no BPD at all. In the epidemiologic study (study IV), also a composite outcome of BPD or death before 36 weeks PMA was assessed.

Because BPD is defined according to the given respiratory support (to achieve a certain goal in the oxygen saturation level in the peripheral blood), the diagnostics of BPD is dependent on the treatment practises. These may vary between the study centres and between the study periods. Because of this, the analyses were, when applicable, performed also with different definitions of cases and controls (by excluding the infants with moderate BPD from the cases and by excluding the infants with mild BPD from the controls). The results were also, when applicable, tested in different time periods.

The use of Oxygen reduction test started in Finland in April 2008 i.e. in the middle of the recruitment period for the studies of this thesis. Because this test has
been shown to shift the infants from the diagnosis of moderate-BPD to the diagnosis of mild-BPD, the analyses, when applicable, where done solely with infants with severe-BPD as cases.

4.7.2 Definition of IUGR/SGA in present studies

In present studies, IUGR/SGA was defined as BW less than minus two standard deviation (SD) corresponding to the 2.3th percentile of the mean BW. This definition was used instead of the other widely used definition of IUGR/SGA, i.e. BW below the 10th percentile (–1.28SD), because it probably encompasses most infants with disordered intrauterine growth (for further discussion on nomenclature around intrauterine growth, please see Review of the literature.) New Finnish growth curves were taken into use after being published in 2013 (Sankilampi, Hannila, Saari, Gissler, & Dunkel, 2013), i.e. in studies III and IV. In studies I and II, the previous Finnish growth data (Pihkala, Hakala, Voutilainen, & Raivio, 1989) were used.

4.8 Ethical considerations

All the present studies were approved by the Ethics Committees of the participating centres.

Written informed consent was obtained from the parents of the infants participating in the genetic studies (I–III). Preterm birth is always a challenging situation for the parents, and genetic studies, especially genome-wide and when taken from a new-born, can be loaded with ethical issues with all information revealed by them. To ease the parents’ role in the recruitment procedure, effort was made to recruit the patients to the studies by a person not being involved in the clinical care of the infant.
5 Results

5.1 VEGFR2 – no consistent association with BPD (I)

In study I, 25 tSNPs for VEGFR2 and six tSNPs for VEGF were genotyped in a discovery population consisting of 160 Finnish infants born in Oulu. Of those, 44 had moderate-to-severe BPD i.e., were cases (Table 3). In this population, VEGFR2 SNP rs4576072 associated with BPD with MAF of 23.9% in cases compared to 9.1% in controls (OR 3.15, 95% CI 1.62–6.12, \( p = 5.0 \times 10^{-4} \)). The SNP did not associate with GA or being born IUGR/SGA. In a logistic regression analysis with GA and IUGR/SGA as covariates, SNP rs4576072 remained a significant risk factor of BPD with an OR of 3.00 (95% CI 1.40–6.39, \( p = 0.005 \)). The other genotyped SNPs or haplotypes of VEGFR2 or VEGF did not associate with BPD (present study I, Table 3).

The SNP associating with BPD in the discovery population, VEGFR2 SNP rs4576072, was further genotyped in a replication population consisting of 328 infants of whom 120 had moderate-to-severe BPD (Table 3). This population was genetically more heterogeneous: 80 of the infants were Canadian with Caucasian origin, the rest of the infants were born in one of the five university hospitals in Finland, only 16 (4.9%) of infants were born in Oulu. In this population, the SNP rs4576072 did not associate with BPD (MAF 9.1% in cases compared to 13.2% in controls, \( p = 0.112 \)).

5.2 KITLG – an association with BPD (II)

In study II, eight tSNPs of the gene KITLG encoding Kit ligand were genotyped in discovery populations. The SNP rs11104948 associating with BPD in a discovery population, was thereafter genotyped in replication populations. Subsequently, concentrations of Kit ligand in the cord blood were compared between infants who later would develop BPD and those who would not.

5.2.1 SNP rs11104948 associated with BPD

The discovery population consisted of three different populations: 1) Northern Finland (n = 253, of which 56 infants with moderate-to-severe BPD i.e. cases, 2) Southern Finland (n = 111, of which 59 cases) and 3) Canada (n = 126, of which
58 cases) (Table 3 and present study II, Figure 1). In the Northern Finland cohort, of the eight SNPs genotyped, six SNPs showed nominal association with BPD. Of those, SNP rs11104948 remained significant after correction for multiple testing with MAF 13.6% in cases compared to 6.1% in controls (OR 2.42, 95% CI 1.22–4.80, \( p = 0.009 \)) (Table 4). No association between this SNP and birth weight, birth weight Z-score or IUGR/SGA was seen. In joint analysis of discovery populations, this SNP associated nominally with BPD (Table 4). In a logistic regression analysis with GA and birth weight Z-score as covariates, the associating SNPs remained significant risk factors of BPD \( (p < 0.05) \). When analysed separately, no SNP showed an association in the Southern Finnish or Canadian populations.

The replication population consisted of Finnish infants \((n = 282, \text{ of which 41 had moderate-to-severe BPD})\) and Hungarian infants \((n = 69, \text{ of which 29 were cases})\). Only SNP rs11104948 showing association in Northern Finnish population, was genotyped. In the Finnish replication population, the allele frequency distribution between cases and controls for this SNP was similar as in the Northern Finnish population; however, the difference was non-significant (Table 4). In the Hungarian population, there was no difference in the allele frequencies between cases and controls. However, when both replication populations were combined with the discovery populations and the possible confounding effect of clusters was considered, there was a significant association between this SNP rs11104948 and BPD \((\text{OR 1.58, 95% CI 1.11–2.24, } p = 0.010)\). In this combined population, in a logistic regression analysis with GA, birth weight Z-score and gender as covariates, the SNP remained significant predictor of BPD \((\text{OR 1.80, } p = 0.008)\).

**Table 4. Association of KITLG SNP rs11104948 with BPD in different study populations.**

<table>
<thead>
<tr>
<th>Study population</th>
<th>n (cases)</th>
<th>MAF</th>
<th>OR</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Case / control</td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td>Discovery population</td>
<td>490 (173)</td>
<td>0.115 / 0.070</td>
<td>1.73 (1.09–2.73)</td>
<td>0.018</td>
</tr>
<tr>
<td>Northern Finland</td>
<td>253 (56)</td>
<td>0.136 / 0.061</td>
<td>2.42 (1.22–4.80)</td>
<td>0.009 (^3)</td>
</tr>
<tr>
<td>Southern Finland</td>
<td>111 (59)</td>
<td>0.098 / 0.100</td>
<td>0.98 (0.40–2.42)</td>
<td>0.965</td>
</tr>
<tr>
<td>Canada</td>
<td>126 (58)</td>
<td>0.112 / 0.074</td>
<td>1.59 (0.67–3.78)</td>
<td>0.290</td>
</tr>
<tr>
<td>Finnish replication</td>
<td>282 (41)</td>
<td>0.134 / 0.083</td>
<td>1.71 (0.84–3.49)</td>
<td>0.135</td>
</tr>
<tr>
<td>Hungarian replication</td>
<td>69 (29)</td>
<td>0.155 / 0.138</td>
<td>1.15 (0.44–2.99)</td>
<td>0.771</td>
</tr>
<tr>
<td>Combined Finland</td>
<td>639 (156)</td>
<td>0.122 / 0.076</td>
<td>1.69 (1.11–2.66)</td>
<td>0.013 (^3)</td>
</tr>
<tr>
<td>Combined all populations</td>
<td>834 (239)</td>
<td>0.123 / 0.080</td>
<td>1.58 (1.11–2.24)</td>
<td>0.010 (^3)</td>
</tr>
</tbody>
</table>

\(^1\)Only infants with available data on SNP rs11104948 genotype; \(^2\)Odds ratio (95% confidence interval); \(^3\)significant after multiple-testing. BPD, bronchopulmonary dysplasia; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio
5.2.2 Cord blood levels of Kit ligand predicted BPD

Because of the observed association between KITLG and BPD, the cord blood levels of Kit ligand were compared between infants subsequently developing BPD and those who would not. Kit ligand concentrations were available from 120 infants born in Oulu (n = 35 cases, n = 85 controls). The levels of Kit ligand were higher in infants later developing BPD than in infants surviving without (mean/median ± SD 868/738 ± 383 FU vs. 632/555 ± 266 FU, respectively, \( p < 0.001 \)). Kit ligand concentration associated also with GA. In a ROC analysis, with cut-off value of 738 FU, Kit ligand levels predicted BPD with 49% sensitivity and 81% specificity (area under the curve 0.72, \( p = 0.0002 \)) (present study II, E-Figure 2). In a logistic regression analysis with GA below 28 weeks and IUGR/SGA as covariates, high Kit ligand concentration (>738 FU) associated with BPD (OR 3.46, 95% CI 1.34–8.90, \( p = 0.010 \)). There were no significant differences in Kit ligand concentrations between the genotypes of the studied SNPs, but the study was not sufficiently powered to that analysis.

5.3 CRP – an association with BPD (III)

Fig. 4. Flowchart of the study III
First, in study III, a hypothesis-generating GWAS was performed. Thereafter, the most promising SNPs were genotyped in subsequent replication populations (Figure 4). Because the most promising SNP through this multistage approach was close to the CRP gene, concentrations of plasma CRP were compared between infants developing BPD and those surviving without it.

### 5.3.1 SNP rs11265269 associated suggestively with BPD

The study cohort in GWAS consisted of 174 infants (n = 60 cases i.e. infants with moderate-to-severe BPD, n = 114 controls) (Table 3 and Figure 4). None of the genotyped SNPs revealed the stringent genome-wide significance level ($p < 1.8 \times 10^{-7}$), but a total of 29 SNPs showed suggestive signals with $p$-value for the association with BPD $< 1 \times 10^{-4}$ (present study III, Table 2). Of those, SNP rs11265269 was the most promising (OR 3.22, 95% CI 1.94–5.34, $p = 3.43 \times 10^{-6}$, Figure 5 and Table 5). SNP rs11265269 is located approximately 44 kb upstream of the CRP gene and 23 kb upstream of the DUSP23 gene.

**Fig. 5.** Manhattan plot summarizing the results of GWAS. Each dot represents the $-\log_{10}(p)$ value of a single SNP analysed. The most promising SNP, SNP rs11265269 at chromosome one, is marked. SNPs above the blue line ($-\log_{10}(p) > 3.3$; i.e., $p < 5 \times 10^{-4}$) were selected for further genotyping in the internal replication population.
The first internal replication population consisted of 326 Finnish infants (n = 105 cases, n = 221 controls) (Table 3 and Figure 4). In this population, SNP rs11265269 showed a similar (however, non-significant) association as in the original GWAS (OR 1.42, 95% CI 0.94–2.13, \(p = 0.097\), Table 5). Also in the second internal replication population (n = 31 cases, n = 198 controls) (Table 3 and Figure 4), the difference in the allele frequencies of this SNP between cases and controls was towards the same direction, but the difference was not statistically significant. When the first and second internal replication populations were combined, the association between the SNP rs11265269 and BPD was nominally significant (OR 1.47, 95% CI 1.04–2.06, \(p = 0.029\), Table 5). When all the Finnish populations were combined, the association was clearer (OR 1.84, 95% CI 1.39–2.45, \(p = 2.39 \times 10^{-5}\), Table 5). Finally, in a logistic regression analysis with GA and SGA as covariates, SNP rs11265269 predicted BPD (for all Finnish populations combined; OR 1.82, 95% CI 1.36–2.43, \(p = 5.32 \times 10^{-5}\)) (present study III, Table 6).

Table 5. Association of SNP rs11265269 with BPD in different study populations.

<table>
<thead>
<tr>
<th>Study population</th>
<th>n (cases)</th>
<th>MAF(^1) case / control</th>
<th>OR(^2) (95% CI)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS</td>
<td>174 (60)</td>
<td>0.392 / 0.167</td>
<td>3.22 (1.94–5.34)</td>
<td>3.4 x 10^{-6}</td>
</tr>
<tr>
<td>Internal replication population 1 (Finnish)</td>
<td>326 (105)</td>
<td>0.278 / 0.230</td>
<td>1.42 (0.94–2.13)</td>
<td>0.097</td>
</tr>
<tr>
<td>Internal replication population 2 (Finnish)</td>
<td>229 (31)</td>
<td>0.274 / 0.204</td>
<td>1.50 (0.75–2.86)</td>
<td>0.216</td>
</tr>
<tr>
<td>Internal replication populations combined</td>
<td>555 (136)</td>
<td>0.277 / 0.218</td>
<td>1.47 (1.04–2.06)</td>
<td>0.029</td>
</tr>
<tr>
<td>All Finnish populations combined</td>
<td>729 (196)</td>
<td>0.313 / 0.207</td>
<td>1.84 (1.39–2.45)</td>
<td>2.4 x 10^{-5}</td>
</tr>
<tr>
<td>External replication population 1 (Caucasian)</td>
<td>312 (98)</td>
<td>0.263 / 0.259</td>
<td>0.99 (0.59–1.37)</td>
<td>0.629</td>
</tr>
<tr>
<td>External replication population 2 (French African)</td>
<td>76 (25)</td>
<td>0.440 / 0.235</td>
<td>2.48 (1.17–5.24)</td>
<td>0.017</td>
</tr>
</tbody>
</table>

\(^1\) Minor allele frequencies of the controls are similar to those of the populations of the 1000genomes project populations (0.239 and 0.223 for the European and African populations, respectively; http://www.1000genomes.org); \(^2\) Odds ratio for minor allele. BPD, bronchopulmonary dysplasia; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio.
Also in the first external replication population consisting of infants with African origin (n = 25 cases, n = 51 controls) (Table 3 and Figure 4), the SNP rs11265269 associated nominally with BPD (OR 2.48, 95% CI 1.17–5.24, \( p = 0.017 \), Table 5). In the second external replication population, the study infants were of Caucasian origin (n = 48 cases, n = 163 controls) (Table 3 and Figure 4). In that population there was no association between the SNP rs11265269 and BPD (Table 5).

**5.3.2 SNP rs3093059 and haplotypes from the CRP region associated with BPD**

Because the most promising SNP, SNP rs11265269, displayed linkage disequilibrium (LD) with some of the SNPs within or nearby the *CRP* gene – rather than *DUSP23* gene – (present study III, Supplemental Figure S4), the genomic region around the *CRP* gene was studied further. Two other SNPs from this region, SNPs rs3093059 and rs12091403 associated with BPD when the population used in GWAS was combined with the first internal replication population \( (p < 0.05) \). In a haplotype analysis of this region, three haplotypes associated with BPD \( (p < 0.05) \) (present study III, Table 4).

**5.3.3 SNP rs3093059 associated with plasma CRP levels**

Because of the above-mentioned evidence of association between SNPs in the *CRP* gene region and BPD, plasma CRP values from the first week of life were assessed in relation to these polymorphisms.

Plasma CRP values from the first week of life were available from the clinical reports for 275 Finnish infants (n = 73 cases, n = 202 controls) (Figure 4). The number of surfactant (Poractant alfa) doses associated significantly with CRP levels \( (p < 0.001) \) and was thus included as covariate in logistic regression analyses. In those analyses, SNP rs3093059 associated with mean plasma CRP concentrations (OR 2.19, 95% CI 1.01–4.72, \( p = 0.046 \)) and showed a trend towards association with maximum plasma CRP concentrations (OR 2.00, 95% CI 0.93–4.27, \( p = 0.074 \)). No association was seen between the best SNP from the GWAS, SNP rs11265269, and CRP levels.
5.3.4 Plasma CRP levels predicted BPD

Since SNPs from the CRP region associated with both BPD and plasma CRP levels, the possible association between plasma CRP levels and BPD was studied. CRP values from the first 12 hours of life were generally low and did not predict BPD. Instead, CRP values from the first week of life were higher in cases i.e. in individuals subsequently developing BPD, compared to controls; for maximum CRP the p-value for the difference between cases and controls was 3.0 x 10^-6, for mean CRP it was 2.0 x 10^-6. The number of surfactant doses associated with CRP values. Instead, no association was seen between CRP values and GA, clinical chorioamnionitis, preterm premature rupture of membranes, cord blood pH or culture proven sepsis (either early- or late-onset during the first week).

In logistic regression analyses with GA and IUGR/SGA as covariates, both the maximum and the mean CRP level (below/above the median) from the first week of life predicted BPD (Table 6).

Table 6. Association of first weeks CRP levels with BPD.

<table>
<thead>
<tr>
<th>Variable</th>
<th>BPD cases/controls with CRP above median, n (%)</th>
<th>OR (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum CRP</td>
<td>53 (72.6) / 82 (40.6)</td>
<td>3.40 (1.76−6.58)</td>
<td>3.0 × 10^-4</td>
</tr>
<tr>
<td>Mean CRP</td>
<td>54 (74.0) / 84 (41.6)</td>
<td>3.57 (1.88−6.77)</td>
<td>9.7 × 10^-5</td>
</tr>
</tbody>
</table>

1 Maximum or mean CRP level (mg/L) during the first week of life; 2 χ² test p values were 3.0 × 10^-6 and 2.0 × 10^-6 for maximum and mean CRP, respectively; 3 Logistic regression with gestational age and IUGR/SGA as covariates. Odds ratio given for maximum and mean plasma CRP (above/below median) during first week of life. BPD, bronchopulmonary dysplasia; CI, confidence interval; CRP, C-reactive protein; OR, odds ratio.

5.3.5 Other SNPs with suggestive associations with BPD

In addition to the SNP rs11265269, the GWAS pinpointed 28 other SNPs suggestively associating with BPD with p-value for the association < 1 × 10^-4 (present study III, Table 2). The most promising among those were SNP rs1889268 in COL15A1 and SNP rs5999125 in LARGE (Table 7).
Table 7. Association of SNP rs1889268 (COL15A1) and SNP rs5999125 (LARGE) with BPD in the GWAS, in the combined Finnish population and in the first External replication population (Caucasian)

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>GWAS</th>
<th>Finnish joint</th>
<th>External replication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MAF</td>
<td>OR 1 (95% CI)</td>
<td>MAF</td>
</tr>
<tr>
<td>rs1889268</td>
<td>COL15A1</td>
<td>0.192/0.382</td>
<td>0.38</td>
<td>2.9×10^-4</td>
</tr>
<tr>
<td>rs5999125</td>
<td>LARGE</td>
<td>0.158/0.333</td>
<td>0.38</td>
<td>5.0×10^-4</td>
</tr>
</tbody>
</table>

1 Odds ratio for minor allele in basic association analysis. 2 Odds ratio for minor allele in logistic regression analysis with gestational age as a covariate. GWAS, genome-wide association study; MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism
5.4 The risk of BPD was lower in twins than in singletons (IV)

The occurrence of moderate-to-severe BPD was calculated for the study groups: Singletons (n = 2843, of which 487 cases), all twins (n = 1137, of which cases 121), first-born twins (n = 497, of which cases 66) and second-born twins (n = 489, of which 55 cases). Thereafter, the risk of BPD was compared between singletons and the different groups of twins, separately. These analyses were performed for all study infants, for infants born before 28 weeks of gestation and for infants born between 28 and 32 weeks of gestation, respectively. The analyses were adjusted for GA and intrauterine growth (IU-growth) expressed as a birth-weight Z-score, separately. The composite outcome of BPD or death before 36 weeks PMA (BPD/death) was assessed similarly. In the study of the BPD/death outcome, the number of singletons was 2785 (of which cases 811) and the number of twins 1130 (of which cases 265).

Table 8. Occurrence of BPD according to plurality and gestational age. Occurrences are calculated per 1000 live-born infants of the type studied, i.e., per 1000 live-born singletons for singletons, etc.

<table>
<thead>
<tr>
<th>Study group</th>
<th>BPD (n)</th>
<th>Occurrence</th>
<th>GA adjusted OR¹</th>
<th>95% CI</th>
<th>Birth-weight Z-score adjusted OR¹</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;32 weeks GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singleton</td>
<td>487</td>
<td>194</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Twin</td>
<td>121</td>
<td>123</td>
<td>0.61</td>
<td>0.48–0.77</td>
<td>0.61</td>
<td>0.49–0.76</td>
</tr>
<tr>
<td>First-born</td>
<td>66</td>
<td>132</td>
<td>0.68</td>
<td>0.50–0.92</td>
<td>0.70</td>
<td>0.53–0.92</td>
</tr>
<tr>
<td>Second-born</td>
<td>55</td>
<td>112</td>
<td>0.54</td>
<td>0.39–0.74</td>
<td>0.53</td>
<td>0.39–0.72</td>
</tr>
<tr>
<td>&lt;28 weeks GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singleton</td>
<td>301</td>
<td>414</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Twin</td>
<td>76</td>
<td>327</td>
<td>0.67</td>
<td>0.49–0.93</td>
<td>0.67</td>
<td>0.48–0.92</td>
</tr>
<tr>
<td>First-born</td>
<td>44</td>
<td>377</td>
<td>0.86</td>
<td>0.57–1.31</td>
<td>0.87</td>
<td>0.58–1.32</td>
</tr>
<tr>
<td>Second-born</td>
<td>32</td>
<td>276</td>
<td>0.52</td>
<td>0.33–0.81</td>
<td>0.50</td>
<td>0.32–0.77</td>
</tr>
<tr>
<td>28-31 weeks GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singleton</td>
<td>186</td>
<td>105</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Twin</td>
<td>45</td>
<td>60</td>
<td>0.54</td>
<td>0.38–0.76</td>
<td>0.64</td>
<td>0.45–0.91</td>
</tr>
<tr>
<td>First-born</td>
<td>22</td>
<td>58</td>
<td>0.52</td>
<td>0.33–0.84</td>
<td>0.67</td>
<td>0.42–1.07</td>
</tr>
<tr>
<td>Second-born</td>
<td>23</td>
<td>62</td>
<td>0.56</td>
<td>0.35–0.88</td>
<td>0.62</td>
<td>0.39–0.97</td>
</tr>
</tbody>
</table>

¹Odds ratios (OR) for twins compared with singletons. BPD, bronchopulmonary dysplasia; CI, confidence interval; GA, gestational age; OR, odds ratio
In the whole study population, both GA-and IU-growth adjusted risk of BPD was lower in all groups of twins than in singletons (Table 8 and Figure 6). For infants born before 28 weeks gestation, the similarly adjusted risk of BPD was lower in all groups of twins, but the difference was not statistically significant for first-born infants. For infants born between 28 and 32 weeks gestation, the risk of BPD in all twin groups was still lower than in singletons, but the difference was not statistically significant for first-born infants when adjusting for IU-growth (Table 8). When comparing the risk of BPD in first-and in the second-born twins, the risk was higher in first-born twins, but the difference was not statistically significant (for all twins: GA adjusted OR 1.28, 95% CI 0.84−1.97, \( p = 0.25 \); for twins born < 28 weeks: GA adjusted OR 1.68, 95% CI 0.94−3.01, \( p = 0.081 \)).

![Fig. 6. The occurrence of BPD in twins and singletons](image)

The results for the composite outcome of BPD/death were towards the same direction; compared to singletons, the risk of BPD/death was lower in all groups of twins, but the adjusted differences were not statistically significant for all subgroups (present study IV, Supplemental Table E4).
5.4.1 The antenatal risk factors of BPD did not differ between twins and singletons

In previous studies, GA, low BW, restricted IU-growth, male gender and in most of the studies, also maternal pre-eclampsia, have shown to associate with BPD. In the present study, in contrast to previous studies, these factors were assessed in separate study groups of singletons, twins and first- and second-born twins. Male gender associated with increased risk of BPD in all study groups, but the association was statistically significant only in singletons (Table 9). Also maternal pre-eclampsia associated with increased risk of BPD, but the association was non-significant in first-born twins ($p = 0.09$). However, because the ORs for both male gender and pre-eclampsia were about the same size for all the study groups, the above-mentioned differences in significance levels of the risk increase associating with male gender and preeclampsia, were determined to be due to limited sample size. Thus, our interpretation of the data is that no significant differences were noticed between twins and singletons among the assessed antenatal risk factors of BPD (Table 9). The results of the risk factor analyses for the composite outcome of BPD/death, were similar (present study IV, Supplemental Table E5).
<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Singletons</th>
<th>Twins</th>
<th>First-born twins</th>
<th>Second-born twins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P</td>
<td>OR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td><strong>A) Univariate analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male gender</td>
<td>1.58 (1.26–1.97)</td>
<td>&lt;.001</td>
<td>1.26 (0.82–1.94)</td>
<td>.292</td>
</tr>
<tr>
<td>Birth weight (100 grams)</td>
<td>0.77 (0.73–0.81)</td>
<td>&lt;.001</td>
<td>0.64 (0.53–0.79)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Birth weight Z-score</td>
<td>0.67 (0.62–0.72)</td>
<td>&lt;.001</td>
<td>0.54 (0.41–0.70)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SGA</td>
<td>3.99 (3.03–5.26)</td>
<td>&lt;.001</td>
<td>8.00 (2.55–25.13)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pre-eclampsia</td>
<td>2.20 (1.70–2.85)</td>
<td>&lt;.001</td>
<td>2.58 (0.87–7.60)</td>
<td>.086</td>
</tr>
<tr>
<td><strong>B) Multivariate analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male gender</td>
<td>2.09 (1.62–2.69)</td>
<td>&lt;.001</td>
<td>1.79 (1.08–2.99)</td>
<td>.025</td>
</tr>
<tr>
<td>Birth weight (100 grams)</td>
<td>0.81 (0.74–0.89)</td>
<td>&lt;.001</td>
<td>0.62 (0.49 – 0.78)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SGA</td>
<td>1.66 (1.07–2.58)</td>
<td>.025</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Adjusted for GA. For singletons and for first- and second-born twins, separately, the ORs were calculated with logistic regression analysis for univariate analyses, and with logistic regression analyses with backward elimination for multivariate analyses. For all twins analyzed as a group, the ORs were calculated with the generalized estimating equation with logit link function and twin pairs clustered for both univariate analyses. In the multivariate analyses, SGA was chosen instead of birth-weight Z-score because of its greater impact in the univariate analyses.

2 Birthweight Z-score describes distribution of birthweight at given length of gestation in SD. BPD, bronchopulmonary dysplasia; OR, odds ratio; CI, confidence interval; SD, standard deviation; SGA, small for gestational age (Z-score ≤ −2SD).
6 Discussion

6.1 Methodological considerations

6.1.1 Study populations

Due to the population history of Finland, with its restricted number of founders, the Finnish population is genetically relatively homogeneous and has therefore been used extensively in genetic studies (Peltonen et al., 1999; Peltonen et al., 1999; Sabatti et al., 2009). Furthermore, the care of very preterm infants in Finland is centralized and the treatment protocols standardized, which is of importance when studying BPD, the incidence of which is affected by neonatal treatment practices.

In the present genetic studies, we performed initial genotyping on discovery populations that consisted of infants of presumed Finnish origin. To test the generalizability of the associations revealed in the discovery populations, genetically more heterogeneous populations were represented among the replication populations (Table 3). However, as a rule the associations observed in the discovery populations of the present studies were replicated only in replication populations of Finnish origin.

In all but GWAS (where the population structure was considered, and the outliers left out), we based our presumption of the ethnicity of the study infants in the present studies on the surnames of their parents; thus, the infants’ ethnicity was conjectured rather than verified. Ideally, ethnicity should be genetically determined since, at least in some populations, self-reported ethnicity correlates poorly with ethnicity based on ancestral markers (Ortega & Kumar, 2015; Wang et al., 2013). However, compared to the populations used in most other genetic studies of BPD, the Finnish study populations in the present studies were likely genetically homogeneous. For example, in Hadchouel et al.’s GWAS, the ethnicity of the study infants was either Caucasian or African, while Wang et al.’s GWAS included four identified groups (Caucasian, African American, Hispanic, and Asian) and Ambalavanan’s study included Caucasians and African Americans.

The use of genetically related study individuals is advantageous, because the greater the ancestral distance among study individuals, the lower the chance that individuals share a common disease susceptibility variant. In other words, it is plausible that all infants who develop BPD have several DNA variants that increase
their susceptibility to BPD. However, those variants are different among individuals, and the degree to which they differ increases with increasing time elapsed since the shared common ancestor. For example, Ambalavanan et al.’s GWAS of BPD found large differences in pathways associated with BPD by race/ethnicity (Ambalavanan et al., 2014). Thus, in genetic studies of complex diseases it may be easier to identify relevant variants in genetically homogeneous populations compared to more heterogeneous populations, in which ethnicity must be taken into account during analyses.

6.1.2 Use of candidate gene and genome-wide association studies to identify genes involved in BPD

The human genome contains >3 billion nucleotides and >21,000 protein-coding genes; thus, it is not surprising that most candidate gene studies of complex diseases have been unsuccessful in revealing replicable associations (Harrow et al., 2012; Marian, 2012). However, there are exceptions; for example, several associations found in candidate gene studies of diabetes type 1 have been replicated in GWASs (Polychronakos & Li, 2011).

There are currently approximately 70 published candidate gene studies of BPD. In addition to the general limitations of the candidate gene study as a method, a limitation shared by almost all candidate gene studies of BPD is the small sample size. Furthermore, only a few genes have been associated with BPD in more than one candidate gene study: genes encoding mannose-binding lectins (MBL) (Cakmak et al., 2012), surfactant proteins (Rova et al., 2004), and angiotensin-converting enzyme (ACE) (Ryckman, Dagle, Kelsey, Momany, & Murray, 2012). Furthermore, none of the SNPs identified in candidate gene studies were significantly associated with BPD in genome-wide studies.

The lack of replication of associations noted in candidate gene studies does not automatically mean that the initial association was a false positive. Small sample sizes predispose studies to beta errors, and the multiethnic populations included in most of the studies complicate analyses. For example, the exonic SNP rs2536512 in SOD3 was associated marginally with a reduced risk of BPD with GA and BW adjusted OR of 0.49 (95% CI 0.21–1.1) and \( p = 0.09 \) in a candidate gene study with an Italian population (Giusti et al., 2012). The same SNP showed a trend towards association in an analysis of exons in a mostly Mexican–Hispanic population in the Wang et al. GWAS (\( p = 0.03 \)). In our study III, the same SNP once again showed a trend towards association (\( p = 0.07 \)) in the Finnish GWAS population. Thus, this
SNP has not reached an association with BPD at a formal significance level in any of the published studies, but it has demonstrated a consistent trend towards association.

Since the mid-2000s, GWASs have discovered >50,000 unique SNP-trait associations, which have been published in >3000 publications (MacArthur et al., 2017). Most of the conducted GWASs investigated SNPs (i.e., common variants), and thus the great majority of the associated variants, even in meta-analyses of GWASs, have had modest effects, with ORs of <1.2. Moreover, the underlying causal variant has been identified only rarely and most of the associations have been in noncoding areas (Cirulli & Goldstein, 2010; Manolio, 2013). On the other hand, as revealed by the ENCODE project, 80% of the genome contains elements linked to biochemical functions, and many of the associations found in GWASs are located in enhancer and other regulatory elements (ENCODE Project Consortium, 2012; Ward & Kellis, 2012). Thus, even if GWASs have not fulfilled initial expectations, they have revealed much about the biological pathways and processes of complex diseases; for some diseases, such as Crohn disease, their impact has been remarkable (Parkes, 2012; Robinson, Wray, & Visscher, 2014; Strachan et al., 2015). Other achievements of GWASs include disease classification, drug development, drug selection, and even risk prediction (e.g., type I diabetes) (Polychronakos & Li, 2011) (Manolio, 2013).

Before the present study III, three GWASs of BPD have been published. One of those found a significant association between a gene encoding SPOCK2 and BPD (Hadchouel et al., 2011). This finding did not replicate in the other GWASs, and the associations noted in them remained at the level of suggestion (Ambalavanan et al., 2014; Wang et al., 2013). A likely reason for this “missing heritability” is the genetic heterogeneity of study populations, discussed in detail in the preceding section. Another plausible reason for not finding significant associations is that the GWASs were designed to investigate only common variants, thus excluding rare variants. However, because of the human population explosion of the past 1000–2000 years, most of our genetic variation is recent and thus the vast majority of single nucleotide variants (SNVs) in humans are rare. In an exome-sequencing study of 1351 individuals of European and 1088 individuals of African ancestry, 86% of SNVs had a MAF of <0.05 (i.e., a standard GWAS would not capture them), 82% were previously unknown, and 82% were population-specific. Furthermore, 96% of the SNVs predicted to be functionally important (i.e., predicted to affect protein function) were estimated to be these rare variants (Tennessen et al., 2012). The results were similar in another large exome-
sequencing study (Fu et al., 2013). Thus, rare variants are likely extremely important to phenotypic variation and disease susceptibility, and the common disease–rare variant hypothesis explains at least part of the current “missing heritability” of complex diseases (i.e., why methods based on common variation have managed to explain only a modest fraction of the heritability in most complex traits). This probably also applies to BPD.

Sample size is another issue in candidate gene and GWA studies of BPD. It is less straightforward to calculate the sample size for genetic studies compared to clinical trials, for example, because the information required (e.g., interactions among genetic variants and among genetic variants and environmental factors) is not available. The sample size required depends, among other things, upon the type of variation studied (common/rare), the effect size of the SNPs chosen, the LD structure of the study population, and the density of the genotyped SNPs (Spencer, Su, Donnelly, & Marchini, 2009). In general, the more samples the better (McCarthy et al., 2008). However, increasing the sample size should not lead to studying individuals with less risk of the disease; for example, by including late-preterm infants in BPD studies. However, the most important factor, also within the context of sample size, is the phenotype studied; the more heterogeneous the phenotype, the higher the number of study individuals required. Once again, we are hindered by the great variation in the pathogenesis of BPD and the imprecise definition of BPD. In general, GWASs that have been successful in finding associations for other complex diseases than BPD have had sample sizes that were much larger than those of the BPD GWASs (MacArthur et al., 2017).

Yet another limitation of both candidate gene and GWA studies of BPD is that they have been poorly or not at all designed to assess gene–gene and gene–environment interactions. However, dynamic gene–gene interactions (i.e., epistasis) probably play a very important role in the genome (Sadee et al., 2014).

Finally, there is the question of the heritability of BPD. As discussed earlier in the review of the literature, evidence for the high heritability of BPD is based on two relatively small twin studies that compared the concordance of BPD in mono- and dizygotic twins. However, zygosity in these studies was not confirmed by DNA analysis. Furthermore, in complex diseases, there is evidence of a fundamental gap between the heritability estimated from sibling studies and the heritability calculated based on known genetic variants (Zuk, Hechter, Sunyaev, & Lander, 2012). The aforementioned gene–gene interactions may contribute to this gap. Thus, the “true” heritability of BPD may be far less than the heritability estimated from the twin studies.
6.2 Main Findings

6.2.1 VEGF and BPD (I)

We assessed the association between BPD and the genes encoding VEGF and its main receptor, VEGFR-2. Thirty-one tSNPs were genotyped in a discovery population of 160 infants of Finnish origin born in Northern Finland (n = 44 BPD cases). One of the genotyped SNPs of \textit{VEGFR2} was associated with BPD in the discovery population, but in the more heterogeneous replication population the allele frequencies of that SNP were the opposite. Thus, the association was studied at the level of common variation and reached significance only in a small and genetically relatively homogeneous study population.

We chose to study VEGF and VEGFR2 because of strong evidence of the involvement of the VEGF pathway in the pathogenesis of BPD. Most of this evidence arose from animal studies. Expression of \textit{Vegf} and \textit{Vegfr1} and levels of VEGF protein were decreased in a baboon model of BPD (Maniscalco et al., 2002). In a rat model of BPD, levels of VEGFR2 were decreased (Wagenaar et al., 2004). Furthermore, inhibition of VEGFR-2 impaired vascular development and alveolarization and caused pulmonary hypertension in rats (Jakkula et al., 2000; Le Cras et al., 2002). Also in rats, recombinant human VEGF therapy and \textit{VEGF} gene therapy restored vascular growth and lung structure (Kunig et al., 2005; Thebaud et al., 2005). In a recent study, treatment with anti-sFlt-1 monoclonal antibody, which inhibits the action of the decoy receptor for VEGF, preserved lung structure and function in a mouse model of BPD and prevented right ventricular hypertrophy (Wallace et al., 2017). In lung biopsies from infants dying of BPD, mRNA levels of \textit{VEGF} and \textit{VEGFR1} (Bhatt et al., 2001) and VEGFR-2 protein levels (Janer et al., 2008) were decreased. In addition, cord blood VEGF levels (Mestan et al., 2017) and VEGF levels in bronchoalveolar fluid and urine were low during the first days of life in infants who later developed BPD (Been et al., 2010; Hasan, Beharry, Valencia, Strauss, & Modanlou, 2009; Levesque et al., 2013).

In humans, common polymorphisms of both \textit{VEGF} and \textit{VEGFR2} have been shown to influence VEGF level and function (An et al., 2009). Furthermore, common polymorphisms of \textit{VEGF} are associated with lung function in both children and adults (Simpson et al., 2012). Within the context of BPD, in addition to the present study I, common variations of these genes have been investigated in five candidate gene studies. None of those studies included a replication population. However, two of them reported an association between \textit{VEGF} and BPD. First, the
VEGF SNP rs833061 was associated with BPD in a Polish candidate gene study that included 118 cases and 63 controls (MAF 48.7% in cases vs 35.7% in controls, OR 1.71, 95% CI 1.1–2.7) (Kwinta et al., 2008). Second, the VEGF SNP rs2010963 was associated with BPD in a small Japanese study of 55 infants with BPD and 42 controls (MAF 66.4% in cases vs. 50% in controls, \( p = 0.02 \)) (Fujioka et al., 2014). The other candidate gene studies did not observe an association (Banyasz et al., 2006; Floros et al., 2012; Mailaparambil et al., 2010), nor did the GWASs.

As discussed, there is strong evidence for the involvement of the VEGF pathway in BPD. The lack of a replicative association in the genetic studies is evident, but in light of our current understanding, not surprising: While the study population in the discovery set of study I was of high quality (i.e., genetically relatively homogeneous), the study shares the limitations of other candidate gene studies of BPD in that only common variations were assessed, the number of study individuals was low, and the study individuals in the replication population were of varying ethnicities.

6.2.2 Kit ligand and BPD (II)

We also assessed the association between BPD and common variations of the gene encoding Kit ligand. Six tSNPs were genotyped in a discovery population consisting of three populations, one of which was a genetically more homogeneous Northern Finland population (n = 253, of whom 56 were BPD cases). One of the genotyped SNPs, SNP rs11104948, was associated with BPD in that population. The allele differences of that SNP trended towards the same direction in the other study populations, including replication populations, and when all populations were combined (n = 834, including 239 BPD cases), the association remained significant. The finding was supported by the levels of Kit ligand in cord blood, which were higher in infants who later developed BPD. However, no association between the genotyped SNPs and the level of cord blood Kit ligand was noticed. Thus, in parallel to our findings in study I, the genetic association observed in study II was at the level of common variation and the initial association was observed in a genetically more homogeneous population.

The rationale behind choosing to study Kit ligand as a candidate gene for BPD first arose from a prospective cohort that analysed levels of Kit ligand and 106 other immunoproteins in the umbilical cord blood as potential predictors of BPD (Kaukola et al., 2009). In addition, Kit ligand is an important growth factor of mast cells, and mast cells have been found in excess in human infants dying of BPD.
More recently, genome-wide transcriptional profiling demonstrated that three of the five most highly induced genes in human BPD lungs were specific for mast cells. In addition, a substantial accumulation of connective tissue mast cells was observed in the lungs of infants dying of BPD, and also in a mouse model of BPD (Bhattacharya et al., 2012).

The present study II is the first to report an association between KITLG, the gene encoding Kit ligand, and BPD. Several mutations of KIT, the gene encoding the main receptor of Kit ligand, have been associated with excessive accumulation, growth, and activation of mast cells (El-Agamy, 2012). Neither of these genes showed association with BPD in our or in others GWASs (Ambalavanan et al., 2014; Bhattacharya et al., 2012; Hadchouel et al., 2011; Wang et al., 2013).

The mechanisms by which the associating KITLG SNP might increase susceptibility to BPD remain speculative. Because of the observed difference in cord blood Kit ligand levels between infants who developed BPD and those who did not, the observed polymorphism could contribute to altered expression of Kit ligand. Kit ligand protein exists in two isoforms: a membrane-bound and a soluble form (Kassel, da Silva, & Frossard, 2001). One of the hypothetical effects of the observed polymorphism could be an altered ratio between these two. Another possible effect of the associating SNP could be an impact on mast cell subtypes, since mast cells exist in several forms. Bhattacharya et al.’s gene expression study found that the chymase-producing type of mast cells in particular accumulated in the lungs of BPD infants (Bhattacharya et al., 2012).

Thus, analogously to study I, study II found a significant association only in the genetically more homogeneous discovery population. However, the allele frequency differences in all study populations trended towards the same direction, and when populations were combined in joint analysis the association remained significant. This type of joint analysis does not respect the strict requirements of replication of the association finding, but the overall strength of the association is, however, considered to be the most important (McCarthy et al., 2008). Of course, since no “true” association was reached in the replication populations, it is possible that the association observed in the study is a spurious association due to chance.

6.2.3 CRP and BPD (III)

The initial GWAS with 174 Finnish infants (of which 60 were BPD cases) revealed several candidate loci for further analyses in replication populations. The most promising among these was SNP rs11265269, which is located approximately 44
kb upstream of \textit{CRP} and 23 kb upstream of \textit{DUSP23}. This SNP was associated with BPD in the Finnish replication populations and also in the replication population, which consisted of infants of African origin. A significant association was also noted between CRP levels in the first week after birth and the risk of BPD.

CRP is a prominent clinical marker of infection and inflammation. CRP activates phagocytic cells, stimulates production of inflammatory cytokines, and regulates the classical complement pathway. Phosphocholine is one of the main ligands of CRP. Phosphocholine is a constituent of many bacterial and fungal polysaccharides, but it is also present in dead or dying cells. Interleukins 1 and 6 have a major synergistic effect on the induction of CRP expression (Volanakis, 2001; Weismann & Binder, 2012).

CRP was linked to BPD in two studies that preceded our study III: levels of plasma CRP at day 28 were shown to associate with BPD (Ambalavanan, Ross, & Carlo, 2005), and D’Angio \textit{et al.} found a qualitative association between CRP levels from days 0 to 21 and BPD (D’Angio \textit{et al.}, 2016). Our finding that CRP levels during the first week of life discriminated between infants who later developed BPD and those who did not is novel and interesting: Compared to other identified biomarkers of BPD, plasma CRP is a feasible candidate because it is easy and inexpensive to measure (C. V. Lal & Ambalavanan, 2016). Moreover, plasma CRP may identify high-risk individuals as early as during the first week of life. In our study, CRP levels from the first week of life did not correlate with either early- or late-onset sepsis or with other factors previously associated with perinatal CRP levels (e.g., low GA, preterm premature rupture of membranes, chorioamnionitis, and low cord blood pH). However, neonatal blood cultures tend to remain negative during the first days of life because of the frequent use of antibiotics to the mothers giving birth preterm. It is thus possible, that the association noted between CRP and BPD represents an association between infection and BPD. Hence, of course, CRP values, as all the other laboratory values, need to be interpreted together with other findings and clinical signs. Independent studies are also needed to confirm the association between the first weeks CRP levels and BPD.

Given the central role of inflammation in the pathogenesis of BPD, an association between levels of plasma CRP and BPD is biologically plausible. Genetically predisposed individuals could be prone to a more intense inflammatory reaction, clinically seen as higher CRP levels during the first week of life. This intense reaction could contribute towards a development ultimately leading to a diagnosis of BPD several weeks later. Alternatively, the genetically predisposed individuals could fail to resolve the inflammation caused by clinical risk factors of
BPD frequently faced by all preterm infants at neonatal intensive care units (Figure 1) (Kolls, 2017).

The study used a multistage approach in which only the most promising SNPs were genotyped in subsequent replication populations. This design was used to power the study; the study populations throughout the study were limited in size relative to the number of interesting SNPs (Pahl, Schafer, & Muller, 2009). At the same time, the design excluded many SNPs of interest and thus may have excluded SNPs with true associations. Another issue is the location of the most promising SNP, SNP rs11265269. This SNP is located between the genes encoding CRP and DUSP23. DUSP23 is a phosphatase involved in cell–cell adhesion regulation (Gallegos et al., 2016). There are no studies of DUSP23 and BPD. Because of the LD between the most promising SNP and the SNPs of CRP, we determined that rs11265269 is probably linked to CRP rather than DUSP23. This conclusion is not without weaknesses and thus a potential link between DUSP23 and BPD remains to be tested.

In addition to rs11265269, study III pinpointed several other suggestive associations, such as SNP rs1889268 in COL15A1 (encoding the alpha chain of type XV collagen). The case–control allele frequency differences for this SNP were consistent in all Caucasian populations, although the difference was not significant in the internal replication population. SNP rs2536512 in SOD3 was another SNP of interest. As discussed earlier, this particular SNP was associated nominally with BPD in two previous studies (one candidate gene study and one GWAS) (Giusti et al., 2012; Wang et al., 2013). SOD3 encodes a member of the superoxide dismutase family, which protects the lungs and other tissues from oxidative stress. Furthermore, SOD3 is regulated by methylation during alveolar septation (Cuna et al., 2015). Another SNP of interest with a suggestive association in our study was SNP rs5999125 in LARGE, which encodes a glycosyltransferase. In addition, SNP rs1822471 in RASGRF1 (encoding Ras protein specific guanine nucleotide releasing factor 1) may have represented a real association, since other SNPs located near this gene showed association signals in a previous GWAS (Wang et al., 2013). Variants near SGCD (encoding sarcoglycan delta) are another example of SNPs of interest, because they are known to be associated with airway responsiveness and as such are of plausible biological relevance to BPD.
6.2.4 Twins and BPD (IV)

In this population-based register study of 1139 twins and 2845 singletons (121 and 488 cases, respectively) the risk of BPD was higher in singletons compared to different groups of twins (all twins, first-born twins, and second-born twins considered separately). Male gender increased the risk of BPD in all study groups, but this was not statistically significant in any of the above-mentioned groups of twins. The same applied to the combined outcome of BPD or death before 36 weeks PMA.

As discussed earlier in the review of the literature, studies of the occurrence of BPD exclusively in twins are few and the results are controversial. There are no prior published studies on the risk factors in twins only. Yet twins comprise about 30% of very preterm infants, and in light of new therapies with unknown long-term effects, identification of the individuals at highest risk is becoming increasingly important.

Since the pathogenesis of BPD plausibly starts in utero, most new therapies to prevent BPD will probably be administered soon after birth. Consequently, high-risk individuals need to be identified shortly after birth. Several cellular and humoral biomarkers for BPD have been identified. However, their prospective predictive value is low, and most of them cannot be measured in clinics (C. V. Lal & Ambalavanan, 2016). The same applies to reported genetic associations: as discussed previously, most of them are preliminary and represent common variations with only modest size effects. Thus, at present, clinical risk factors remain the best means to estimate the risk of developing BPD in newborns.

After identifying a lower risk of BPD among twins, we compared the risk of BPD in first- and second-born twins. The risk was lower in second-born twins, in line with the results of two other recent large studies (Hunter et al., 2017; Mei-Dan et al., 2017), but the difference was not statistically significant. Unfortunately, we did not have zygosity data for the twins. Thus, we were unable to calculate the heritability of BPD in this study.

In our study, to serve the need of identification of risk individuals at birth, only antenatal risk factors of BPD were analysed. The postnatal risk factors were thus not considered, which may have had an impact on the results. Another issue in our study is that there were less SGA infants among twins than among singletons. However, in regression analyses considering the effect of IU-growth and GA, being a singleton remained a statistically significant risk factor for BPD. Nevertheless, twin pregnancies were also less frequently complicated by pre-eclampsia. Hence,
pregnancies with a single fetus resulting in very preterm delivery may be more profoundly complicated in terms of placental dysfunction –potentially increasing the risk of BPD– than twin pregnancies, where the demands for placenta may be exceeded by the simple fact of the needs of two fetuses instead of one, or the reason for preterm delivery has not to do with the function of placenta. The above-mentioned is of course only hypothetical, and the real reason for the lower occurrence of BPD among twins compared to singletons remains to be discovered.

Our interpretation of the risk factor analyses is that there were no significant differences in the risk factors assessed. However, male gender increased the risk of BPD statistically significant only in singletons. Furthermore, in rodents, the difference in gender of the neighbouring fetus has been shown to influence a phenotype (Nielsen, 1985). However, in our data, the risk of BPD among the twin pairs of the same gender was not higher than among the different-gender twin pairs.

6.2.5 Future studies

For future studies of BPD, both genetic and epidemiological, redefining the phenotype is a priority. In addition to subclassifying infants diagnosed according to the current definition of BPD, the whole disease should be redefined so that the diagnosis of BPD would predict long-term outcome. Other critical aspects when designing future studies are the sample size and heterogeneity of the study populations. The question of sample size is challenging, since, at least for genetic studies, including infants of different ethnic backgrounds tends to yield data that are less readily analysed, while on the other hand, few relatively homogeneous populations are large enough to achieve the required sample sizes within a reasonable amount of time. Biobanks under development in countries including Finland (e.g., the ongoing FinnGen project) could offer a solution: Via imputation and other technical procedures, information gained through direct genotyping can be amplified, thus limiting the actual need for study individuals. To address the other main issue, the genetic heterogeneity among the study individuals, the study populations for at least genetic studies of BPD, should be chosen carefully. The population of Finland offers several advantages in the field of population genetics because of its unique population history combined with comprehensive data registers on health care–related factors. Hence, sequencing and other in-depth analyses of the few infants both of Finnish origin and with the most severe form of BPD could reveal important clues about the pathogenesis of BPD. In-depth analyses would preferably include detailed “omic”-data (e.g., proteomic,
metabolomic, and microbiomic), as well as placental analyses. Finally, continuing to investigate the most promising variants identified in previous studies, such as those of \textit{SOD3}, could be of great value.
7 Conclusions

1. Common polymorphisms of the genes encoding VEGF and its main receptor VEGFR-2 did not associate constantly with BPD.
2. A common variant in the gene encoding Kit ligand associated with BPD. While the association was not confirmed in the replication population, cord blood kit ligand levels associated with BPD.
3. Several genomic regions were found to potentially associate with BPD. The most promising of those was a SNP close to the gene encoding CRP. Furthermore, plasma CRP levels during the first week of life predicted BPD.
4. The risk of BPD was lower in twins compared to singletons. The actual mechanisms protecting twins from BPD remain to be discovered. There was no difference in the robust antenatal risk factors of BPD between twins and singletons.

The results of this study contribute to our understanding of the genetic susceptibility and the pathogenesis of BPD. The study also provides new data on the risk of BPD, especially in twins.
References


Hirschhorn, J. N. (2005). Genetic approaches to studying common diseases and complex traits. Pediatric Research, 57(5 Pt 2), 74R-77R. doi:01.PDR.0000159574.98964.87


102


*Seminars in Neonatology* : SN, 8(1), 39-49. doi:S108427560200194X [pii]

disability to age 5 years after neonatal caffeine therapy for apnea of prematurity. *Jama,* 

motor function, and behavior 11 years after neonatal caffeine citrate therapy for apnea 

Schmidt, B., Roberts, R. S., Davis, P., Doyle, L. W., Barrington, K. J., Ohlsson, A., . . . 
Caffeine for Apnea of Prematurity Trial Group. (2006). Caffeine therapy for apnea of 
20/2112 [pii]

Schmidt, B., Roberts, R. S., Davis, P., Doyle, L. W., Barrington, K. J., Ohlsson, A., . . . 
1893-1902. doi:357/19/1893 [pii]

Canadian Oxygen Trial (COT) Group. (2013). Effects of targeting higher vs lower 
arterial oxygen saturations on death or disability in extremely preterm infants: A 

Non-invasive versus invasive respiratory support in preterm infants at birth: Systematic 
review and meta-analysis. *BMJ (Clinical Research Ed.),* 347, f5980. doi:10.1136/ 
bmj.f5980 [doi]

Seedorf, G., Metoxen, A. J., Rock, R., Markham, N., Ryan, S., Vu, T., & Abman, S. H. 
(2016). Hepatocyte growth factor as a downstream mediator of vascular endothelial 
doi:10.1152/ajplung.00423.2015 [doi]

corticosteroids for preventing bronchopulmonary dysplasia in ventilated very low birth 
weight preterm neonates. *The Cochrane Database of Systematic Reviews,* 10, 
CD002058. doi:10.1002/14651858.CD002058.pub3 [doi]

corticosteroids for the treatment of bronchopulmonary dysplasia in ventilated very low 
birth weight preterm infants. *The Cochrane Database of Systematic Reviews,* 10, 
CD002057. doi:10.1002/14651858.CD002057.pub4 [doi]


110


113
List of original publications

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:


Reprinted with permission from S. Karger AG, Basel (I), Wiley (II) and Springer Nature (III).

Original publications are not included in the electronic version of the dissertation.
1461. Pasanen, Anu (2018) Genetic susceptibility to childhood bronchiolitis
1462. Käkelä, Juha (2018) Family history of mental disorders and long-term outcome in schizophrenia
1463. Xu, Qi (2018) Role of Wnt11 in kidney ontogenesis and development of renal organoid based models to identify candidate oncogenes
1465. Isojarvi, Henri (2018) Association of glucose metabolism, physical activity and fitness with peripheral nervous system function in overweight people
1467. Tikkanmäki, Marjaana (2018) Preterm birth and parental and pregnancy related factors in association with physical activity and fitness in adolescence and young adulthood

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

GENETIC BACKGROUND
AND ANTENATAL RISK FACTORS OF
BRONCHOPULMONARY DYSPLASIA