All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Hanna Peterson, 2010

Printed by

www.reproprint.se
Gårdsvägen 4, 169 70 Solna
ABSTRACT

Pre-eclampsia is a multifactorial, pregnancy-specific vascular disorder characterized by hypertension and proteinuria. It affects around 3-5% of pregnancies worldwide. There is a wide range of phenotypes from mild forms developing in the end of pregnancy, to severe forms with extremely high blood pressure that in worst cases could lead to eclampsia, the occurrence of seizures. Pre-eclampsia and eclampsia account for more than 50 000 maternal deaths per year. The etiology and pathophysiology of pre-eclampsia remain poorly understood, but it is generally accepted that defect placentation during the early stage of pregnancy, most likely in combination with maternal and environmental factors could lead to systemic inflammation, endothelial dysfunction and the manifestation of the clinical symptoms. Both large epidemiological and family studies have demonstrated genetic contribution to susceptibility. Although several loci have been mapped by linkage analysis only a few promising positional candidate genes have been identified so far. A number of functional candidate genes encoding for coagulation factors, oxidative stress and vasoactive substances, have been suggested to mediate susceptibility, but attempts to replicate these findings have yielded inconsistent results.

The overall aim of this thesis was to search for genes predisposing for pre-eclampsia using several different approaches. In Paper I we evaluated the role of the first positionally cloned pre-eclampsia candidate gene STOX1 at 10q22 in the Finnish population. We were unable to validate STOX1 as a common pre-eclampsia gene, and our result is in agreement with two other European studies investigating the same gene. An intriguing association in Paper II suggests that pre-eclampsia share a predisposing genetic factor on chromosome 9p21 with coronary artery disease. To the best of our knowledge we were the first to investigate the role of the 9p21 region in pre-eclampsia. The association of this locus has not been confirmed in other populations and further investigations of the genes in this region are warranted. We have previously mapped a candidate susceptibility locus to chromosome 2p25. In Papers III and IV we present our systematic efforts to narrow down the linkage region by fine-mapping followed by association analysis.

In conclusion, our investigations provides an insight into a potential role of a new susceptibility locus for pre-eclampsia at 9p21 in the Finnish population. We were able to narrow down the linkage region at 2p25, but found our sample sets underpowered to evaluate the genes residing within it. Finally, there is no conclusive evidence either for or against STOX1 as a susceptibility gene for pre-eclampsia. To further explore the role of STOX1, much larger sample sets are needed.
LIST OF PUBLICATIONS


II. Hanna Peterson, Katja Kivinen, Leena Hiltunen, Elina Salmela, Tuuli Lappalainen, Vesa Rasi, Ayat Sayed, Lucilla Poston, Matthew P Johnson, Linda Morgan, Eric K Moses, Juha Kere and Hannele Laivuori. Common variants on chromosome 9p21 are associated with pre-eclampsia in the Finnish population. In manuscript


IV. Hanna Peterson, Katja Kivinen, Erja Kerkelä, Hannele Laivuori, Leena Hiltunen, Hong Jiao, Ville- Veiko Mäkelä, Risto Kaaja, Olavi Ylikorkala, Vesa Rasi och Juha Kere. Fine-mapping and characterization of pre-eclampsia susceptibility locus on chromosome 2p25. In manuscript
CONTENTS

1 Background......................................................................................................................... 1
   1.1 Pre-eclampsia........................................................................................................... 1
      1.1.1 Phenotype and symptoms .......................................................................... 1
      1.1.2 Definition.................................................................................................... 1
      1.1.3 Incidence, prevalence, mortality and morbidity........................................ 2
      1.1.4 Long term health effects for mother and child.......................................... 3
      1.1.5 Prevention and treatment ........................................................................... 4
      1.1.6 Risk factors................................................................................................. 5
      1.1.7 Etiology and pathophysiology ................................................................. 6
   1.2 The human genome ............................................................................................... 12
      1.2.1 The human genome project...................................................................... 12
      1.2.2 Sequence variations.................................................................................. 12
      1.2.3 The HapMap project ................................................................................ 14
      1.2.4 The 1000 Genomes Project...................................................................... 14
   1.3 Genetic studies of complex diseases..................................................................... 15
      1.3.1 Genome-wide linkage analysis................................................................ 15
      1.3.2 Association analysis and candidate gene studies .................................... 16
   1.4 Genetics of pre-eclampsia ..................................................................................... 19
      1.4.1 Genetic susceptibility............................................................................... 20
      1.4.2 Susceptibility loci for pre-eclampsia identified by linkage..................... 21
      1.4.3 Association-based mapping approaches.................................................. 24

2 Aims of the thesis............................................................................................................. 28

3 Material and methods....................................................................................................... 29
   3.1 Study subjects ........................................................................................................ 29
      3.1.1 Finnish families (I-IV) ............................................................................. 29
      3.1.2 Finnish case-controls (I-IV) .................................................................... 30
      3.1.3 UK case-controls (II) ............................................................................... 30
      3.1.4 Australian/New Zealand families (II)...................................................... 31
      3.1.5 Finnish placental samples (I-IV)... ......................................................... 31
      3.1.6 Finnish control samples (II)..................................................................... 31
   3.2 Genetic analysis..................................................................................................... 32
      3.2.1 Marker selection (I-IV) ............................................................................ 32
      3.2.2 Genotyping: SNPs and microsatellites (I-IV) ......................................... 32
      3.2.3 Data analysis (I-IV).................................................................................. 34
      3.2.4 DNA re-sequencing (I, III, IV) .............................................................. 35
   3.3 Expression analysis ............................................................................................... 36
      3.3.1 Microarray expression analysis (I-V) ...................................................... 36
      3.3.2 Allele specific expression analysis (III) .................................................. 36

4 Results and discussion...................................................................................................... 38
   4.1 Paper I - Evaluation of STOX1 on 10q22 ............................................................. 38
   4.2 Paper II - Association in 9p21 CAD risk region................................................... 39
   4.3 Paper III, IV - Mapping and characterization of 2p25 linkage region................. 42
   4.4 General aspects ................................................................................................. 44

5 Concluding remarks and future perspectives................................................................... 47

6 Acknowledgements .......................................................................................................... 49

7 References ........................................................................................................................ 51
LIST OF ABBREVIATIONS

ACVR2A  Activin receptor type 2A
ASSHP  Australasian Society for the Study of Hypertension in Pregnancy
Aus/NZ  Australian/New Zealand
BMI  Body mass index
bp  Base pair
CAD  Coronary artery disease
CDCV  Common disease, common variant
CDKN  Cyclin dependent kinase inhibitor
cDNA  Complementary deoxyribonucleic acid
CEPH  Centre d'Etude du Polymorphisme Humain
CI  Confidence interval
cM  Centimorgan
CNV  Copy number variant
DNA  Deoxyribonucleic acid
ddNTP  Dideoxynucleotide
dNTP  Deoxynucleotide
FINNPEC  Finnish Genetics of Pre-eclampsia Consortium
GH  Gestational hypertension
GOPEC  The genetics of pre-eclampsia UK Consortium
GWAS  Genome-wide association study
HGP  Human genome project
HLA  Human leukocyte antigen
HPM  Haplotype pattern mining
H/R  Hypoxia-reoxygenation
HWE  Hardy-Weinberg equilibrium
ICD  International classification of diseases
IBD  Identical by descent
IGF  Insulin growth factor
IL  Interleukin
IVF  In vitro fertilization
ISSHP  International Society for the Study of Hypertension in Pregnancy
kb  Kilo base, 1000 base pairs
KIR  Killer immunoglobulin-like receptor
LD  Linkage disequilibrium
LDL  Low-density lipoprotein
LOG  Logarithm of odds
LPL  Lipoprotein lipase
MAF  Minor allele frequency
MALDI-TOF  Matrix-assisted laser desorption/ionisation time-of-flight
Mb  Mega base, 1 000 000 base pairs
MHC  Major histocompatibility complex
MMP  Matrix metalloproteinase
mRNA  Messenger ribonucleic acid
NHBPEP  National High Blood Pressure Education Program
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>NK</td>
<td>Natural killer</td>
</tr>
<tr>
<td>NPL</td>
<td>Non-parametric linkage</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDT</td>
<td>Pedigree disequilibrium test</td>
</tr>
<tr>
<td>PIGF</td>
<td>Placental growth factor</td>
</tr>
<tr>
<td>RAAS</td>
<td>Renin-angiotensin-aldosterone system</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>ROCK</td>
<td>Rho-associated, coiled-coil containing protein kinase</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple sequence repeat</td>
</tr>
<tr>
<td>STMP</td>
<td>Syncytiotrophoblast microparticle</td>
</tr>
<tr>
<td>STOX</td>
<td>Storkhead box</td>
</tr>
<tr>
<td>T2D</td>
<td>Type 2 diabetes</td>
</tr>
<tr>
<td>TDT</td>
<td>Transmission disequilibrium test</td>
</tr>
<tr>
<td>Th</td>
<td>T helper cell</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>TSC</td>
<td>The SNP consortium</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
</tbody>
</table>
1 BACKGROUND

1.1 PRE-ECLAMPSIA

1.1.1 Phenotype and symptoms

Pre-eclampsia is a pregnancy-specific vascular disorder manifesting in the latter part of the pregnancy (after 20 weeks of gestation), although the pathophysiological process is thought to take place during placental development in early stages of pregnancy. Pre-eclampsia is characterized by elevated blood pressure and protein in the urine (proteinuria). There is a wide range of phenotypes from mild forms of pre-eclampsia developing in the end of pregnancy, generally with few or no symptoms, to severe forms with extremely high blood pressure that may lead to upper abdominal pain, visual disturbances or headache, problems in the liver, kidneys, brain and clotting abnormalities. The fetus is at risk and the most common abnormalities are intrauterine growth restriction as a result of reduced blood supply through placenta and problems of prematurity. A final and severe phase of pre-eclampsia is called eclampsia. It is a rare complication, characterized by the occurrence of seizures, often leading to changes in the circulatory system and kidney failure, and is associated with an increased risk of maternal death. Although the outcome of pre-eclampsia is often good, it can be in some cases life threatening for both mother and child.

1.1.2 Definition

There is no universal classification system of hypertensive disorders of pregnancy and definition of pre-eclampsia, and there has been a wide diversity of terminology and diagnostic criteria over the years for published studies (Harlow and Brown 2001). Currently there are several internationally recognized definitions available (Brown et al. 2001). Reports widely used for classification and definitions are from the Australasian Society for the Study of Hypertension in Pregnancy (ASSHP) and the National High Blood Pressure Education Program (NHBPEP) (ASSHP 1993, NHBPEP 2000). The definitions of hypertension in pregnancy are identical between the two reports; Systolic blood pressure ≥ 140 mmHg and/or a diastolic blood pressure ≥ 90 mmHg. The classifications of hypertensive disorders during pregnancy are similar with four defined categories; pre-eclampsia or eclampsia, gestational hypertension (GH), chronic hypertension and superimposed pre-eclampsia. The categories are summarized in table 1, including generalized and most commonly used definitions based on diagnostic criteria suggested by different reports. Pre-eclampsia is most often defined as a combination of hypertension and proteinuria, and the NHBPEP definition consists of new onset of hypertension (≥ 140 mmHg and/or a diastolic blood pressure ≥ 90 mmHg) after 20 weeks of gestation in combination with proteinuria which is defined as the appearance of ≥ 0.3g/24h of urinary protein or ≥ 1+ reading on a dipstick that correlates to ≥ 0.3g/L in a random urine determination. Attempt to subdivide pre-eclampsia has been done and severe pre-eclampsia is often defined as systolic blood pressure ≥ 160 mmHg and/or a diastolic blood pressure ≥ 110 mmHg in combination with severe proteinuria of ≥3 g/24h.
Table 1. Commonly used diagnostic criteria and classification of hypertensive pregnancy disorders.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Diagnostic criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-eclampsia</td>
<td>Hypertension: Blood pressure of $\geq 140$ mm Hg systolic or $\geq 90$ mm Hg diastolic that occurs after 20 weeks of gestation in a woman with previously normal blood pressure. Proteinuria: Defined as urinary excretion of $\geq 0.3$ g protein in a 24-hour urine specimen.</td>
</tr>
<tr>
<td>Eclampsia</td>
<td>Occurrence, in a woman with pre-eclampsia, of seizures not attributed to other causes.</td>
</tr>
<tr>
<td>Gestational hypertension</td>
<td>Hypertension: Blood pressure of $\geq 140$ mm Hg systolic or $\geq 90$ mm Hg diastolic that occurs after 20 weeks of gestation in a woman with previously normal blood pressure.</td>
</tr>
<tr>
<td>Superimposed pre-eclampsia</td>
<td>Chronic hypertension associated with new-onset of proteinuria during pregnancy.</td>
</tr>
<tr>
<td>Chronic hypertension</td>
<td>Hypertension before 20 weeks of gestation and/or persistent for more than 6 weeks after delivery.</td>
</tr>
</tbody>
</table>

1.1.3 Incidence, prevalence, mortality and morbidity

Approximately 10% of women will have high blood pressure at some point before delivery, and pre-eclampsia complicates around 3-5% of pregnancies worldwide (Hogberg 2005). The incidence is higher in the developing world and specific ethnic groups (Duley 2003, Zhang et al. 2001). Eclampsia, the severe end phase of pre-eclampsia, is associated with mortality and accounts for more than 50,000 maternal deaths per year (Duley 2003). It is rare in Europe with 2-3 cases per 10,000 births, but more common in developing countries with an estimated incidence of 16-69 cases per 10,000 births (Frias and Belfort 2003, Knight 2007, Kullberg et al. 2002). Limited access to maternity services and emergency obstetric care in developing regions is a possible explanation and is reflected by the fact that 99% of all maternal deaths occur in low and middle-income countries (Hogberg 2005). In the countries with low maternal mortality, only a third is associated to eclampsia, compared to countries with high maternal mortality rate where almost all deaths are associated to eclampsia and not pre-eclampsia (Duley 2009). It is noteworthy that the incidence of pre-eclampsia has increased by 40% in the last 15 years (Duley 2009). Plausible causes include worldwide obesity epidemic, a rise in the numbers of older mothers and increase in the frequency of multiple pregnancies. Renal failure, cardiac arrest, stroke, adult respiratory distress syndrome, coagulopathy and liver failure are all severe morbidities associated with eclampsia and pre-eclampsia, and affected women are in need of intensive care (reviewed in (Duley 2009)). There are some studies about the psychological effect of pre-eclampsia on women, since it can be a difficult and
unexpected experience with illness, early deliveries and in worse cases fetal deaths. It may increase the risk of post-traumatic stress disorder (van Pampus et al. 2004).

More than 10% of infants born small for gestational age are from pre-eclamptic pregnancies (Kramer et al. 2000). Perinatal mortality is high after pre-eclampsia and eclampsia, and 25% of all neonatal deaths and stillbirths are associated to the disorders (Ngoc et al. 2006, Roberts et al. 2005). Mortality rates for infants are several times higher in developing world compared to developed countries for both pre-eclampsia (3 times) and eclampsia (4.5 times) (reviewed in (Duley 2009)). The severity of the disorder affects the outcome for both mother and child, with the highest risk for severe pre-eclampsia or eclampsia (Gaugler-Senden et al. 2006). There are also many complications associated with pre-term birth and the infants require intensive care and neonatal facilities, which again are less common in developing countries and will affect the outcome for the infant.

1.1.4 Long term health effects for mother and child

Epidemiological data on long term effects of pre-eclampsia indicate that affected women may also be at increased risk of cardiovascular or cerebrovascular diseases in later life. One of the first studies investigating the long-term effects of eclampsia was published in 1976 (Chesley et al. 1976). Their findings did not show an association to cardiovascular disease or an increase of morbidity and mortality in women who had pre-eclampsia in their first pregnancy. However, women who had experienced eclampsia more than once had a greater incidence of cardiovascular disease as well as higher death rates. More recent data are in conflict with the early findings, and it is now widely accepted that women who have previously experienced pre-eclampsia or eclampsia, even in the first pregnancy, have an increased long-term risk for remote cardiovascular and cerebrovascular disease, such as hypertension, ischemic heart disease, myocardial infarction, and cerebrovascular accidents (reviewed in (Harskamp and Zeeman 2007)). The individual risk is different when looking at specific subgroups of pre-eclampsia and it has been shown that women with recurrent pre-eclampsia, early/severe forms, pre-eclampsia as multiparas, pre-term delivery and women pregnant at older age seem to be at even greater risk (Arnadottir et al. 2005, Irgens et al. 2001, Jonsdottir et al. 1995, Sibai et al. 1986). Moreover, death resulting from cardiovascular causes among pre-eclamptic women has been reported to be 8-12 fold higher than in normotensive women (Irgens et al. 2001).

Many risk factors are shared between pre-eclampsia and coronary artery disease (CAD), including endothelial dysfunction, hypertension, obesity, insulin resistance and dyslipidemia (reviewed in (Garovic and Hayman 2007)). Therefore, it has been suggested that the metabolic syndrome, which refers to a group of conditions including hypertension, dyslipidemia, abdominal fat, and fasting hyperglycemia, may be a possible underlying mechanism common to CAD and pre-eclampsia, and may lead to the different disorders at different time points in a woman’s life (Newstead et al. 2007). Both pre-eclampsia and gestational diabetes (glucose intolerance with onset or first recognition during pregnancy) share features of the metabolic syndrome, and patients with a history of these pregnancy disorders have an increase in the lifelong risk of CAD (Carpenter 2007, Rodie et al. 2004). Supporting the connection between pre-eclampsia
and glucose intolerance, the risk of subsequent type 2 diabetes (T2D) is increased more than 3-fold after severe pre-eclampsia (Lykke et al. 2009). An alternative explanation for the exaggerated risk of CAD and T2D after pre-eclampsia is that pre-eclampsia itself may induce irreversible vascular and metabolic changes that may increase the later risk.

The majority of children from pre-eclamptic pregnancies survive in countries with good health care, but they may have an increased susceptibility for diseases later in life beyond that mediated by their preterm birth. Higher risk of childhood hypertension have been reported (Seidman et al. 1991, Tenhola et al. 2003, 2006) and it has been found that pre-eclampsia is associated with an increased risk of diabetes in the offspring, although, contradicting results have been published as well (Bache et al. 1999, Dahlquist et al. 1999, Jones et al. 1998, McKinney et al. 1999). In addition, a decreased risk of breast cancer has been reported among female offspring from pre-eclamptic pregnancies, which may be explained by low intrauterine estrogen levels that characterize pre-eclamptic pregnancies (Ekbor et al. 1992, Innes et al. 2000, Sanderson et al. 1998, 2006, Xue and Michels 2007). However, the potential biological mechanisms underlying the association between pre-eclampsia and long-term offspring health remain unknown. One possibility is that genetic factors that predispose the mother for pre-eclampsia and disorders later in life are inherited by the offspring or adaptive responses to the intrauterine environment in pre-eclamptic pregnancy may result in epigenetic changes that affect disease susceptibility later in life (Gluckman et al. 2008, Smith et al. 2001).

1.1.5 Prevention and treatment

Since the etiology and pathogenesis of pre-eclampsia are unclear, the development of strategies for prevention and treatment is difficult. Promising results have been published concerning the use of antiplatelet drugs, primarily low-dose acetylsalicylic acid for prevention and a reduced risk of pre-eclampsia after treatment has been reported (Askie et al. 2007). Treatment with acetylsalicylic acid should be considered for women with a history of severe pre-eclampsia in previous pregnancy. Calcium supplementation is associated with a lower risk of pre-eclampsia as well, however, there is no clear impact on outcome for the infant (Hofmeyr et al. 2007). Treatment is largely directed towards the symptoms, with little evidence that any intervention alters the underlying pathophysiology. Mild pre-eclampsia is often managed with careful observation at maternal care units, at hospital or at home, in combination with activity restriction. Since high blood pressure may lead to direct vascular damage, which in turn could lead to other complications such as renal failure, fetal distress and stroke, antihypertensive drugs are mandatory for women with very high blood pressure (Duley and Henderson-Smart 2000). Magnesium sulfate is known to reduce the risk of eclampsia in pre-eclamptic women and can be injected to prevent eclampsia-related seizures (Duley et al. 2003). To date, the only effective “treatment” is delivery and removal of placenta. It can be challenging to balance between protecting the mother by ending the pregnancy, and maximizing the maturity of the fetus when timing the delivery of women with severe pre-eclampsia before 32 weeks of gestation. Therefore, there is a great need for a safe treatment that would eliminate the need for premature
delivery in the severe cases of pre-eclampsia as well as ensure the well-being of the pregnant women.

1.1.6 Risk factors

Risk factors that predispose women for pre-eclampsia are summarized in table 2. Nulliparity is considered to be a powerful predictor of increased risk and a meta-analysis from 2005 reported that women pregnant for the first time are almost three times more likely to get pre-eclampsia than with a second or later pregnancy (Duckitt and Harrington 2005). Increased risk is also observed for multiparous women when there is a change in paternity (Trupin et al. 1996), prior use of barrier contraceptives and with shorter length of sexual relationship (Klonoff-Cohen et al. 1989, Robillard et al. 1994). A plausible explanation for this is that pre-eclampsia may represent a maternal immunological reaction towards paternal antigens (see 1.1.7.4). However, contradicting results have been reported, suggesting that the effect of these factors may be explained by other confounding factors (Ness et al. 2004). Risk associated with change in paternity could also be explained by the fact that longer intervals between pregnancies seems to predispose regardless of paternity (Skjaerven et al. 2002). Age has also been considered as a risk factor, with women ≥40 years old twice as likely to get pre-eclampsia as woman under 40 (Duckitt and Harrington 2005). This may be explained by other age-related risk factors, such as obesity and chronic hypertension. Increasing body mass index (BMI), even in absence of real obesity, is associated with increased risk of pre-eclampsia (Duckitt and Harrington 2005, Eskenazi et al. 1991). Diastolic blood pressure at ≥110 mm Hg increases the risk of developing superimposed pre-eclampsia five-fold and an increased blood pressure within the normal range is also recognized as a risk factor (reviewed in (Duckitt and Harrington 2005). Other conditions linked to pre-eclampsia are insulin dependent diabetes, renal disease, antiphospholipid syndrome (a blood clotting disorder) and autoimmune diseases (Davies et al. 1970, Duckitt and Harrington 2005, Pattison et al. 1993, Stamilio et al. 2000, Yasuda et al. 1995). Prevalence of pre-eclampsia is higher in African-Americans than in other ethnic groups in United States and ethnicity may be one predisposing factor (Sibai et al. 1997). However, the prevalence of chronic hypertension is also higher in this population (Kurian and Cardarelli 2007). Both personal and familial history of pre-eclampsia are considered as true risk factors, with data showing that women with pre-eclampsia in their first pregnancy have seven times the risk of getting pre-eclampsia in a second pregnancy compared to women with healthy first pregnancies (Duckitt and Harrington 2005). Additionally, a family history of pre-eclampsia triples the risk (Duckitt and Harrington 2005). It has been reported that twin pregnancy triples the risk for pre-eclampsia and triplet pregnancy triples the risk of pre-eclampsia compared with twin pregnancy (Duckitt and Harrington 2005, Skupski et al. 1996). This could have an explanation by the increased placental size associated with multiple pregnancies and the fact that decreased placental perfusion is considered as a central feature of pre-eclampsia (Sibai et al. 2000). Furthermore, in vitro fertilization (IVF) is associated to pre-eclampsia and could be explained by immune maladaptation (Kallen et al. 2005, Shevell et al. 2005). An alternative explanation could be that multiple pregnancies are clearly more common after IVF (reviewed in (El-Toukhy et al. 2006)). However, a meta-analysis from 2004 shows that the risk for pre-eclampsia after IVF is higher even in singleton pregnancies (Jackson et al. 2004).
Table 2. Summary of risk factors for pre-eclampsia

<table>
<thead>
<tr>
<th>Risk factors for pre-eclampsia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternity age</td>
</tr>
<tr>
<td>Multiple pregnancy</td>
</tr>
<tr>
<td>Nulliparity</td>
</tr>
<tr>
<td>Previous pre-eclampsia</td>
</tr>
<tr>
<td>Family history of pre-eclampsia</td>
</tr>
<tr>
<td>Body mass index (BMI)</td>
</tr>
<tr>
<td>Time between pregnancies</td>
</tr>
<tr>
<td>Change of partner</td>
</tr>
<tr>
<td>African-American race</td>
</tr>
<tr>
<td>In vitro fertilization</td>
</tr>
<tr>
<td>Pre-existing medical conditions</td>
</tr>
<tr>
<td>Insulin dependent diabetes</td>
</tr>
<tr>
<td>Insulin resistance</td>
</tr>
<tr>
<td>Chronic hypertension</td>
</tr>
<tr>
<td>Renal disease</td>
</tr>
<tr>
<td>Antiphospholipid syndrome</td>
</tr>
<tr>
<td>Autoimmune disease</td>
</tr>
</tbody>
</table>

1.1.7 Etiology and pathophysiology

Pre-eclampsia has been termed the “disease of theories”, reflecting the confusion surrounding the causes and pathophysiology of the disease. Despite the large number of studies, the pathophysiology of this syndrome is not fully understood. However, placental vasculopathy and endothelial dysfunction appear central to the pathogenesis. Pre-eclampsia is considered as a two-stage disorder; Stage I - Defective placentation due to failed remodeling of maternal vessels that results in a poorly perfused placenta. Stage II - Clinical manifestation of pre-eclampsia (Roberts and Hubel 2009). However, the link between the pathophysiology of abnormal placentation and the physiology of the maternal syndrome remains unclear, but it is widely hypothesized that oxidative stress may be one of the important factors (Gupta et al. 2005). The failed remodeling of maternal vessels observed in pre-eclamptic pregnancies in the first stage of the disorder, resulting in a defective placenta, is probably not sufficient to cause pre-eclampsia. Reduced placental perfusion and pathological evidence of failed placental vascular remodeling is also evident in women who had growth-restricted babies and preterm birth with no maternal signs of pre-eclampsia (Roberts and Post 2008). The occurrence of stage II requires interactions with maternal constitutional factors that may be genetic influences, immune maladaptation, and/or environmental factors (reviewed in (Roberts and Post 2008)). Many of those factors leading to the maternal syndrome are also risk factors for cardiovascular disease in later life (Rodie et al. 2004). Pre-eclampsia is a heterogeneous condition, which is consistent with varying degrees of contribution from mother and infant (reviewed in (Ness and Roberts 1996)). Thus, with profoundly reduced placental perfusion the generation of maternal signs may require very little contribution from the mother. Conversely, the woman with extensive predisposing constitutional sensitivity could develop pre-eclampsia with very little reduced perfusion (Roberts and Hubel 2009).
Figure 1. The two stage model of pre-eclampsia pathophysiology. (Stage I) Abnormal placentation in the first trimester, followed by reduced placental perfusion, which results in release of factors from the placental unit that will influence maternal physiology, lead to endothelial damage, and the maternal syndrome (Stage II). The placental dysfunction is triggered by poorly understood mechanisms, which may include genetic, environmental and immunological factors. The same type of factors could also have a role in later pathophysiological events and initiating the maternal syndrome. Pre-eclampsia is associated to both growth restriction pregnancies and later maternal disorders, such as cardiovascular disease and type 2 diabetes. (Modified from (Parikh and Karumanchi 2008))

1.1.7.1 Stage I - Abnormal placentation

Placenta is a temporally vascular organ essential for the exchange of gases, nutrients and waste between fetal and maternal circulatory systems. A variety of metabolic, hormonal and immunological molecules important for the fetal development are synthesized by the placenta. In normal pregnancy, after implantation, cytotrophoblastic cells of fetal origin attach to the uterine wall via so called anchoring villi. The cytotrophoblasts form a shell, lining the uterine cavity and a small proportion invade the uterine wall and its blood vessels called spiral arteries. A round 12 weeks’ gestation, the process where cytotrophoblastic cells invade the spiral arteries of the myometrium peaks and by 18 to 20 weeks, cytotrophoblast cells have replaced endothelial cells lining the vessels and dismantled the muscle and connective tissue. The previous small, high-resistance vessels are converted into large, low resistance vessels that allow for an increase in placental blood flow needed to sustain the fetus throughout the pregnancy (reviewed in (Pijnenborg et al. 2006)) (Figure 2).
In pre-eclampsia pathophysiology, inadequate invasion of trophoblasts has been implicated and different kinds of pathophysiological hallmarks could be found in the placenta (reviewed in (Cheng and Wang 2009)). One observation in pre-eclamptic placentas is poor cytotrophoblast differentiation, which leads to reduced trophoblast invasion into the myometrial segments of the spiral arteries that remain narrow and undilated. This has been supported by Doppler ultrasound studies of the maternal uterine blood flow (Papageorghiou et al. 2004). The invasive cytotrophoblast cells is known to adopt a cell surface adhesion phenotype characteristic for endothelial cells, but in pre-eclampsia the cells fails to undergo the switching from epithelial to endothelial characteristics (Zhou et al. 1997). It means that the cells fail to express some of the intergrin, cadherin, selectin or immunoglobulin superfamily members important for vascular adhesion phenotype. The end result after defective vascular transformation of the arteries is insufficient placental circulation leading to hypoxia, or at least intermittent perfusion, oxidative stress and release of soluble “toxic” factors from the ischemic placenta that damage the vasculature of the mother, leading to widespread vascular injury and increased permeability.

**Figure 2.** A summary of the suggested pathophysiological events in pre-eclampsia leading to endothelial dysfunction. The upper part of the figure shows the spiral arteries, the normal adaptation at pregnancy and the incomplete remodeling process in pre-eclampsia, followed by the possible pathophysiological events leading to endothelial dysfunction (Modified from (Moffett-King 2002)).
The link between stages I and II

The link between abnormal trophoblast invasion, and later generalized endothelial activation and dysfunction, leading to the maternal syndrome is not clear, but it may be via release of placental factors. Microparticles are cellular derived vesicles that are shed from cell membranes, produced in placenta during the continuous process of growth and apoptosis, and mediate cell-to-cell communication with potential role in processes such as thrombosis, homeostasis, angiogenesis and inflammation (reviewed in (Redman and Sargent 2008)). The levels of these kinds of particles are elevated in conditions associated with enhanced systemic inflammation, such as normal pregnancy and at even higher levels in pre-eclampsia. In pre-eclamptic placentas, an increased apoptosis of particular syncytiotrophoblastic cells lining the outer layer of placenta has been observed (Allaire et al. 2000, Leung et al. 2001) and the release of syncytiotrophoblast microparticles (STMP) into the maternal circulation is elevated in pre-eclampsia (Knight et al. 1998). The STMPs are suggested to stimulate inflammatory responses and directly damage endothelial cells (reviewed in (Germain et al. 2007)), since in vitro studies have shown that STMP could simultaneously disrupt endothelial layers and stimulate production of factors that activate leukocytes (Smarason et al. 1993, von Dadelszen et al. 1999).

Another interesting factor released from the syncytiotrophoblasts is the soluble receptor for vascular endothelial growth factor 1 (sVEGFR-1), also called sFlt-1. It functions as an antagonist of two angiogenic factors called vascular endothelial growth factor (VEGF) and placental growth factor (PIGF) and has been found to be upregulated in pre-eclamptic placentas. When an excess of sFlt-1 is present, it binds and inactivates VEGF and PIGF, which are needed for endothelial survival, and therefore induces endothelial dysfunction (reviewed in (Myatt and Webster 2009)) . Interestingly, it has been demonstrated that administration of sFlt-1 to pregnant rats induces hypertension, proteinuria and glomerular endotheliosis, which are all classical features of pre-eclampsia (Mynard et al. 2003). Altogether, this strongly supports sFlt1 as a disease predisposing factor. However, what is upregulating sFlt-1 in pre-eclampsia is not clear. Moreover, upregulation of sFlt-1 only explains part of the cases as some pre-eclamptic women have normal levels of gene product (Powers et al. 2005).

It has been suggested that reduced blood flow through the spiral arteries leads to chronic hypoxia in placenta. There is also evidence for an alternative mechanism, characterized by hypoxia-reoxygenation (H/R) injury as a result of intermittent placental perfusion, secondary to the abnormal artery remodeling (Hung and Burton 2006). H/R injury could be a possible mechanism causing oxidative stress in placenta. Oxidative stress is excessive in pre-eclampsia, may cause endothelial dysfunction through the action of reactive oxygen species (ROS) and therefore considered to be a key step in the pathogenesis (reviewed in (Poston and Rajmakers 2004)). Oxidative stress has a generally accepted role in the pathogenesis of atherosclerosis. The same dyslipidemia is present in both pre-eclampsia and atherosclerosis, in addition to pathological lesions observed in placenta called atherosclerosis, which have high similarity to atherosclerotic lesions (reviewed in (Belo et al. 2008)). Taken together, this leads to the assumption that oxidative stress could play a significant roll in pre-eclampsia as well. One characteristic of dyslipidemia is increased levels of small dense low-density
lipoproteins (LDL) and the H/R in placenta could lead to peroxidation of the LDL particles, which is known to be increased in pre-eclampsia (Atamer et al. 2005, Wang et al. 1992). Subsequently, the toxic products after lipid peroxidation are transported to distant sites in the body and can cause systemic oxidative stress and cellular damage. Many pro-inflammatory cytokines and modulators are found at increased levels in both circulation and placenta during pre-eclamptic pregnancy. Two of them, tumor necrosis factor α (TNFα) and Interleukin-1 (IL-1) have both been implicated in pre-eclampsia pathophysiology, since they have the ability to stimulate structural and functional alterations in endothelial cells (reviewed in (Conrad and Benyo 1997)). However, the source of these molecules has not yet been identified, but a placental contribution is suggested. Interestingly, infusion of TNFα into rats in their late pregnancy resulted in increased arterial pressure and renal resistance (Alexander et al. 2002, Giardina et al. 2002).

1.1.7.3 Stage II - The maternal syndrome

The end stage of pre-eclampsia is the maternal syndrome defined by cardiovascular and renal features; hypertension and proteinuria. A specific renal endothelial lesion called glomerular endotheliosis is associated to proteinuria and endothelial dysfunction has been implicated in the process leading to hypertension. Injury of the endothelium could lead to a cascade of vasoconstriction, coagulation and redistribution of intravascular fluid and this is the center of the systemic dysfunctions in pre-eclampsia (reviewed in (Hayman et al. 1999)). In normal pregnancy, blood pressure and peripheral vascular resistance are decreased, but in pre-eclampsia the changes are reversed. Vascular constriction is universally present in pre-eclampsia and endothelial dysfunction is believed to be responsible for that. Several markers for endothelial dysfunction and activation are altered and women previously affected by pre-eclampsia are much more responsive to vasopressors (Agatia et al. 2004, Chambers et al. 2001). Additionally, studies in vitro of pre-eclamptic vessels have shown alterations in endothelial function (reviewed in (Roberts 1998)). Prostaglandin I2 (PGI2) is a vasodilator produced by the endothelial and smooth muscle layers of blood vessels and its expression is lower in pre-eclampsia compared to normal pregnancy (Mills et al. 1999). Vasoconstrictors with a suggested role in pre-eclampsia pathophysiology are Thromboxane A2 (TXA2) and Endothelin-1 (Clark et al. 1992, Slowinski et al. 2002). Due to vasoconstriction and endothelial leakage observed in pre-eclampsia, fluids are lost from the vascular compartment and perfusion of organs is reduced. A systemic inflammatory response is observed in normal pregnancy, but exaggerated in pre-eclampsia. The inflammatory response is generated by different networks, mainly involving endothelial cell activation, maternal leukocytes and complement systems (Redman and Sargent 2003). Activation of the coagulation cascade is likely to further reduce organ perfusion by formation of microthrombi (Roberts and Lain 2002).

1.1.7.4 The immune maladaptation theory

The so called “immune maladaptation hypothesis” is one suggested etiology for pre-eclampsia. It is subject to controversy, but several epidemiological studies are supporting its validity by showing that pre-eclampsia risk is much higher in first pregnancy and that multiparity functions as a protective effect that is lost with a change of partner (reviewed in (Saito et al. 2007)). The hypothesis is based on the fact that the
fetoplacental unit contains paternal antigens that are foreign for the mother because of the differences between mother and father in respect to human leukocyte antigens. During early pregnancy, natural killer (NK) cells accumulate around the invading trophoblasts and produce cytokines that are involved in angiogenesis and vascular stability (Parham 2004). The tissues located in the maternal-fetal interface are protected against T-lymphocyte destruction by not expressing major histocompatibility complex (MHC) class Ia and II molecules, except for weak expression of human leukocyte antigen (HLA) C. Instead, invading trophoblasts express an unusual combination of HLA-F, HLA-E and HLA-G (Ishitani et al. 2003, Saftlas et al. 2005). HLA-G is only expressed by the extravillous trophoblasts and may, in part, explain the immune tolerance of the mother to the fetoplacental unit by protecting the cells from lysis by NK cells (O’Brien et al. 2000). A reduction of HLA-G expression on trophoblasts has been observed in pre-eclampsia and that may lead to the impaired trophoblast invasion (Colbern et al. 1994). Additionally, HLA-C is a dominant ligand for NK cells by interaction with killer-cell immunoglobulin-like receptors (KIRs) on the surface of the NK cells. It has been shown that a certain combination of the fetal HLA-C and the maternal KIR genotypes is associated to pre-eclampsia (Hiby et al. 2004). This combination is sending an inhibitory signal from the trophoblast to the NK cell and this inhibition has been suggested to play a role in the defective invasion and transformation of the arteries. However, not all pre-eclamptic pregnancies have this genotype combination and this only indicates that additional factors are most certainly involved in the process leading to defective invasion of cytotrophoblasts and the systemic responses.

Several immune cell types, such as NK cells, monocytes, neutrophils, T and B cells, become hyperactivated in pre-eclampsia as a reaction to trophoblastic debris from damaging processes due to hypoxic conditions, oxidative stress or excessive inflammation (reviewed in (Borzychowski et al. 2006)). This activation enhances the production of cytokines, which may play a role in the maternal-fetal interface or in the whole body. T-helper (Th) cells can be classified into two subgroups known as Th1 and Th2, expressing different kinds of cytokines. Th1 cells enhance the cell-mediated immunity, while Th2 are involved in antibody production and repression of cell-mediated immunity. Th2 type immunity is known to play an important role in successful pregnancy by regulating immune response to the fetus. In pre-eclampsia on the other hand, greater amounts of Th1-type cytokines, such as TNFα, are observed in the circulation and the balance has shifted towards Th1 type immunity, which may be harmful for the invading trophoblasts (Saito and Sakai 2003). In summary, immune maladaptation may cause abnormal trophoblast invasion or, alternatively, cause the release of toxic cytokines, free radicals and enzymes from the decidua, which may cause damage or disturb normal function of the maternal endothelium and the syncytiotrophoblasts.
1.2 THE HUMAN GENOME

1.2.1 The human genome project

The sequence of the human genome is rich in information about human evolution and encodes for genetic instructions of human physiology. The DNA double helix was discovered in 1953 by Watson and Crick and made basis for new research in the field of genetics (Watson and Crick 1953). The human genome project (HGP) was launched in 1990 with the goal of determining the sequence of approximately 3 billion base pairs (bp) that make up the human genome, identifying all human genes, store this information in public databases and develop tools for data analysis. The project was completed in 2003, however, drafts of the human genome sequence, both from the publicly funded project and from the private company Celera, were published already in 2001, comprising roughly 90 % of the total sequence (Lander et al. 2001, Venter et al. 2001). The drafts suggested that human genome contained 30 000-40 000 genes (Lander et al. 2001), much less than previous predictions of 60 000-100 000 (Strachan and Read, 2004). The numbers have been revised further after project completion and are now estimated to 20 000-25 000 (IHGSC 2004). A availability of human genome sequence has made it possible to investigate what types of DNA make up our genome and how we are related to other organisms that have been sequenced already.

1.2.2 Sequence variations

Humans are 99.9 % identical with respect to their DNA sequence, but many genetic variations in the human genome have been observed. Genetic variation explains some of the phenotypic differences among people, such as physical traits and whether a person has a higher or lower risk for certain diseases. However, the vast majority of the variants are believed to be neutral with no effect on phenotypic variation. Variations in the genome can be common (polymorphism), defined as genetic variant with minor allele frequencies (MAF) of at least 1 %, or rare (mutation) with MAF less than 1 % (reviewed in (Frazer et al. 2009)). Depending on the nucleotide composition of the variant, they can be subdivided into different classes and two of the most common ones are; Single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs). Both of them have been widely used as genetic markers and have been very important for genetic research. It was not until recently that copy number variants (CNVs) (i.e. DNA segments that present at variable copy number in comparison to a reference genome), which are structural variations, were found to represent a major source for human genetic variation and genome diversity (reviewed in (Conrad et al. 2009)).

1.2.2.1 Simple sequence repeats

SSRs are blocks of tandem repeats consisting of single repetitive nucleotide or di-, tri-, tetra-, or pentanucleotide repeat (dinucleotides are most frequent) (Figure 3B), and constitute around 3 % of the human genome, one every 2 kb (Lander et al. 2001). According to their size, they can be subdivided into two classes; minisatellites (14-500 bp) and microsatellites (1-13 bp). Microsatellite lengths are known to be highly variable between individuals, probably resulting from slippage during the replication process. Therefore, they have been widely used in disease-gene mapping in pedigrees,
where they can distinguish between maternally and paternally inherited alleles (see 1.3.1).

1.2.2.2 Single nucleotide polymorphisms

The most common genetic variation in the human genome are SNPs, where a single nucleotide (A, C, G, T) is changed (Figure 3A), inserted or deleted. The number of SNPs in a human genome is estimated to be approximately 3.3 million, one every 1000 bp (Levy et al. 2007, Wheeler et al. 2008). SNPs are found in both non-coding and coding regions of the genome, and if a single substitution leads to an amino-acid change (missense), frame shift or termination of translation (nonsense), the variation is called non-synonymous. In reverse, when the variant does not affect the amino-acid sequence, it is referred to as synonymous SNP, however, those could have an effect on mRNA stability or splicing instead. Other regulatory elements such as promoters, enhancers and silencers can be affected by base substitutions located outside coding regions. Information on close to 18 million SNPs have been made publicly via dbSNP database (http://www.ncbi.nlm.nih.gov/snp/). Thus, several of those are spurious findings. SNPs are popular in genome-wide studies because of their high frequency in the genome, providing a dense marker map and availability of powerful methods for large-scale analysis (see 1.3.2.1).

1.2.2.3 Copy number variants

CNV results from genomic rearrangements and represent gain or loss of genomic segment of between a few hundred to several million bp and could be generated by normal mechanisms such as replication, recombination and DNA break repair. Generally, genomic segments with variable copy number could encompass parts of genes, reside entirely outside genes or, in the case of larger variants, include several known genes. The most obvious effect of CNVs is gene-dosage effects that could be observed if regulatory elements or genes are located in duplicated/deleted segments. In the past few years, a large number of CNVs have been identified by various genome-wide technologies (reviewed in Zhang et al. 2009). According to the latest statistics, >29 000 entries of CNVs have been included in the Database of Genomic Variants (http://projects.tcag.ca/variation/).
1.2.3 The HapMap project

The International HapMap Project (www.hapmap.org) was initiated in 2002, with the aim to determine common patterns of SNP variation in human genome and to make the information available to the public domain. The goal was to genotype at least one common SNP every 5 kilo base (kb) across the euchromatic regions of the genome in individuals from four geographically diverse populations: mother-father-child trios from Nigeria; trios of northern and western European ancestry in Utah; and unrelated Chinese and Japanese individuals (IHM C 2003, 2005). In the first phase of the study, around 1.3 million SNPs were genotyped, the project continued and today over 4 million SNPs have been analyzed resulting in a SNP density of at least one every 1 kb (Frazer et al. 2007, www.hapmap.org). The extensive genotyping led to an insight into allele frequency differences among populations, along with characterization of the distribution and extent of linkage disequilibrium (LD) across the entire human genome. Genetic variants that are near each other tend to be inherited together, are in LD with each other, and the combination of those linked variants are known as haplotypes (Daly et al. 2001). SNPs in high LD are located in regions of chromosomes that have not been broken up by recombination, and they are separated by places where recombination has occurred. The HapMap project has been able to build a haplotype map of the genome and identified haplotype tagging SNPs that uniquely identify corresponding haplotypes without genotyping the whole combination of SNPs. Therefore, one of the major applications of the HapMap data has been to guide the design and prioritization of SNP genotyping for disease association studies, and making them much more cost effective, since the likelihood of genotyping a monomorphic SNP decreased dramatically, together with the number of SNPs needed for an association studies. Subsequently, the known knowledge about LD structures and tagging SNPs throughout the genome enabled for whole genome-wide association studies (GWASs) (see 1.3.2.1).

1.2.4 The 1000 Genomes Project

The 1000 Genomes Project (http://www.1000genomes.org/) is an international research effort that aims to provide the most comprehensive map of human genetic variation using next-generation sequencing platforms. Advances in sequencing technology have made it possible to process millions of sequence reads in parallel, allowing for whole genomes to be sequenced in an effective way. The 1000 genomes project involves sequencing the genomes of at least 1000 people from a number of different ethnic groups within the next three years The experience gained from this pilot project will hopefully be helpful in the guidance of future large-scale sequencing projects. A new map of the human genome will be developed, including SNPs, CNVs, and short insertion/deletion polymorphisms that appear in at least 1 %, at a resolution unmatched by current resources. Newly identified low-frequency SNPs and CNVs are essential for development of new generation of genotyping arrays, which will enable integrated analysis of SNPs and CNVs, and have better resolution and coverage of the true sequence variation in the human genome (Hurles et al. 2008, Mccarroll 2008). As with other major human genome reference projects, data from the 1000 Genomes Project will be made available to the worldwide scientific community through freely accessible public databases and that would hopefully improve the sensitivity of disease discovery efforts.
1.3 GENETIC STUDIES OF COMPLEX DISEASES

Many common complex diseases are known to cluster in families and are believed to be influenced by several genetic and environmental factors as well as interactions among those. The common disease, common variant (CDCV) hypothesis suggests that variants with relatively high frequency, but low penetrance, are the major contributors to common complex diseases (Altshuler et al. 2008). A late onset is common among complex diseases, with modest or no impact on reproductive fitness (pregnancy specific disorder such as pre-eclampsia and dystocia are exceptions). Therefore, mildly deleterious alleles can rise to moderate frequencies in the population compared to mutations that cause strongly deleterious phenotypes and are lost by natural selection. Moreover, some alleles that were beneficial or neutral during human evolution may now confer susceptibility to disease because of changes in our environment. However, it has been argued that multiple rare variants contributing to the disease are more consistent with pathobiology than common variants (Schork et al. 2009). The genetic etiology is most likely based on a combination of multiple rare and common susceptibility loci. Efforts to investigate complex diseases initially adopted strategies similar to those employed for the successful mapping of Mendelian disorders. The two most commonly used methods are linkage and association studies.

Positional cloning, which has been extremely successful for mapping Mendelian diseases (Jimenez-Sanchez et al. 2001), has been the most common approach for identifying genes in complex disorders during many years and has been successful in some cases. However, even if susceptibility regions are detected by linkage, extensive candidate gene studies are often needed to narrow down causal genes in the broad linkage regions. Another limitation of the linkage studies is that they lack power to identify common genetic variants with modest genetic risk on disease (Hirschhorn and Daly 2005). Association studies are one strategy for refining the linkage regions and to search for genetic variants of modest effects associated to disease. However, during the time of positional cloning, the association studies were restricted to specific genes or loci, and whole genome analyses were not an option. When GWAS become possible in 2006, it opened new frontiers in the understanding of many complex diseases and it is the most widely used approach for genetic mapping today (McCarthy and Hirschhorn 2008). However, the execution and analysis of those studies require big efforts and so far GWAS have only identified a small fraction of the genetic variance underlying the heritable component of complex diseases (Manolio et al. 2009).

1.3.1 Genome-wide linkage analysis

Two loci are linked because of their physical proximity along a chromosome, which means that they are so close together that their alleles tend to cosegregate within families. Cosegregating loci can be broken up by recombination during meiosis, and the probability of recombination increases the further apart two loci are from each other. Subsequently, the probability, referred to as recombination fraction, is a function of the genetic distance between loci, which is expressed in centimorgans (cM). One cM is defined by 1% recombination chance between two loci and represents approximately 1 mega base (Mb). In linkage analysis, the recombination fraction between individual
markers and the disease locus are estimated. Logarithm of odds (LOD) score compares the likelihood that a locus is linked to the likelihood that the observation is purely by chance and not due to linkage, and is an often used test in linkage analysis. The ratio between the two likelihoods gives the odds of linkage and the linkage is reported as LOD score (Morton 1955). If a positive LOD score is observed the presence of linkage is suggested, whereas negative LOD scores indicate that linkage is less likely. Standard LOD score analyses, also called parametric, require a precise genetic model detailing the mode of inheritance, gene frequencies and penetrance of each genotype and are therefore suitable for Mendelian traits. For complex diseases with no clear inheritance pattern, model-free (non-parametric) methods are preferred. Non-parametric linkage (NPL) analysis ignores unaffected people, and looks for chromosomal segments or alleles that are shared by affected individuals. In families, shared segment analysis can be conducted using identical by descent (IBD) data and NPL LOD score could be calculated with a method based on calculating the extent to which affected relatives share alleles IBD (Kruglyak et al. 1996). Families with increased occurrence of a certain disease are utilized in linkage studies, since affected family individuals are most likely to carry the same genetic predisposition. Genome-wide linkage studies, also called genome-wide scans, have been used over the last two decades to map disease predisposing loci. Markers, preferably microsatellites (see 1.2.2.1) are genotyped across de genome, regularly spaced, and the segregation through families is studied (reviewed in (Borecki and Province 2008)). Linkage analysis identifies large genomic regions that are often tens of Mb and may contain hundreds of genes (Boehnke 1994). Further investigations are needed to be able to map predisposing gene(s) and causal allele(s) and for those purposes, association analysis approaches could be useful (see 1.3.2).

1.3.2 Association analysis and candidate gene studies

Association studies seek a correlation between a specific genetic variation and a trait in a sample of individuals. There are three types of association analysis including the two hypothesis driven approaches focusing on candidate genes (i) or candidate susceptibility regions (ii), often as a result from genome-wide linkage analysis in a complex disease, with linked regions of several Mb in need of several steps of fine-mapping, and the non-hypothesis driven approach GWAS (iii) (see 1.3.2.1). Candidate genes are often chosen for their relevance to pathophysiology of the disease of interest, or they may be picked from previously determined linkage regions. It is generally accepted that susceptibility of complex diseases involves, to different extent, multiple genes with genetic effect size most likely being small to modest. For those variants, association are much more powerful than linkage analysis (Cardon and Bell 2001) and the importance of association analysis has increased during the past year as a consequence of the large number of SNPs mapped and development of the technology, which enable genotyping of millions of SNPs simultaneously. Both association and linkage analysis rely on the coinheritance of adjacent variants, which are separated primarily by recombination, but other factors such as population growth and admixture, natural selection, genetic drift and mutation, could affect LD patterns (Ardlie et al. 2002). Since linkage is focusing on families and rather small recombination events have taken place, the disease loci will often be large. In contrast, association analysis utilizes the recombination history over many hundreds or thousands of generations in
the population and disease loci would be comparatively small (reviewed in (Ardlie et al. 2002, Boehnke 1994)).

Association between a marker and a phenotype could appear under two different circumstances, either by typing the causal allele itself (Figure 4A) or, more likely in complex diseases, by typing neighboring variants in LD (Figure 4B). However, in the latter case the power to detect association depends on the strength of the LD (Borecki and Province 2008). The statistical power in association studies is not only depending on the proximity of the marker and causal allele, but also on the contribution of a specific allele to the phenotype (effect size), allele frequencies and sample size. At weak effect sizes, low allele frequencies or modest LD, there is a need for large sample sets. Closely located genetic markers, showing strong intermarker LD, are often transmitted as haplotype blocks and different blocks are suggested to be separated by recombination hot-spots (Daly et al. 2001). LD patterns across the human genome have been characterized in multiple populations by the HapMap project (see 1.2.3), which has made it possible to detect association with only a few tagging SNPs genotyped when LD is high. However, in regions with low LD, dense SNP maps and higher numbers of genotyped markers are needed to find potential effects. The evolutionary history of LD patterns and haplotypes can vary in different ethnical populations (Gu et al. 2007, Service et al. 2006). It has been suggested that the choice of particular ethnic groups might facilitate association studies as these might arise from limited numbers of founders and would provide less disease heterogeneity and larger regions of LD. Sardinia, Iceland and certain areas of Finland, have all been considered as suitable (Peltonen 1996).

**Figure 4.** (A) Direct association analysis by genotyping the causal SNP. (B) The causal SNP is tested indirectly by genotyping neighboring SNPs in linkage disequilibrium (LD).

Case-control study design has been the most widely applied strategy of association analysis. Briefly, a number of unrelated affected and unaffected from a population has been collected, a set of markers at a locus or loci are genotyped, and subsequently, the genotypes are tested to evaluate their frequencies between cases and controls (McCarthy and Hirschhorn 2008). The selection of controls is crucial because any systematic allele frequency differences between cases and controls can appear as disease association, even though they may reflect differences in for example evolutionary or migratory history. Therefore, controls must be carefully matched to reflect the ethnic and genetic composition of cases. In order to reduce the effect of
population stratification, various family-based association approaches have been developed that use controls selected from the families of affected probands (Cardon and Palmer 2003). At present, one of the most popular approaches is the transmission disequilibrium test (TDT), providing a joint test of linkage and association, focusing only on heterozygous parental genotypes and is applied to single probands and parents (Spielman et al. 1993). TDT compares the frequencies of transmitted vs untransmitted alleles in affected offspring and uses untransmitted parental alleles as controls. Limitations of TDT include not utilizing all genotype data since transmissions from homozygote parents are not used and not being applicable for extended pedigrees. Pedigree disequilibrium test (PDT) is an extension of TDT that evaluates evidence of linkage disequilibrium (LD) in general pedigrees by using data from both related nuclear families and discordant sib pairs (Martin et al. 2000).

To evaluate association results and get conclusive proof of an effect, associated variants need to be replicated in several independent populations. The SNP showing association in a study is often not the causal one and instead only reflects existing LD between variants. It is a difficult task to identify a causal allele in a susceptibility gene or region, thus the ultimate proof that a gene is connected to a disease phenotype is the identification of a variant clustering in affected subjects and with a plausible effect on normal gene function, affecting for example protein function, gene expression or mRNA splicing etc. Variants in coding regions affecting amino acid sequences or at splice sites can generally easily be assessed and mutational screening is often exon-centric. However, in complex diseases it is highly plausible to find causal variants in regulatory regions, such as the promoter, in introns or at more distant regulatory element that could be find several kb away from the gene (Frazer et al. 2009). Other variants with an likely effect on disease could affect localization, stability and translation of mRNA. Functional studies are required to determine the consequences of a causal allele.

1.3.2.1 Genome-wide association - Past and future

GWASs allow a hypothesis free approach for finding novel loci and genes associated with diseases. This has been facilitated by the development of tools and methods required, such as mapping of millions of common variants in the human genome and LD information, techniques to capture all of those variants in thousands of individuals, analytical approaches to handle huge amount of data and to distinguish true associations. Although, several new candidate genes and loci have been suggested, some replicated and accepted, these only explain a small proportion of the observed phenotypic variation (reviewed in (Manolio, 2009)). Additionally, GWAS are still missing for many complex diseases such as pre-eclampsia. For the published GWAS, it is unlikely that common variants of large effect have been missed, however, effect sizes for common variants are typically modest, often between 1.1-1.3 (Wray et al. 2008), and achieving enough power for a common SNP with 20% frequency requires over 8000 samples (Altshuler et al. 2008). Thus, most of the original GWAS were clearly underpowered. The importance of less common SNPs (MAF of 0.5-1%) with modest effects, which are not well covered by current GWAS chips that are restricted to high frequency variants, and variants to rare to prove statistical evidence of association, must be considered. Fine-mapping around GWAS hits, denser SNP panels and gene-centric
approaches could be an alternative that will capture variants not always represented on GWAS platforms and resulting in higher locus risk than estimated for the original GWAS. Data from the 1000 genomes project (see 1.2.4) will aid further identification of less common variants (both SNPs and structural variants) and a more comprehensive catalog of genomic variation. In order to identify disease specific rare variants influencing risk, the ultimate approach would be deep sequencing in thousands of cases and controls, but not the most cost-effective. There are many aspects to consider such as how to select the most appropriate cases, how many to sequence per group and is it possible to lower the coverage and still getting useful information. Hopefully, some knowledge will be gained from the first pilot projects that will be helpful for guidance of future studies. It is most likely that common diseases are affected by gene-gene and gene-environment interactions. However, these are difficult to identify with current methods and data set sizes and further investigations are needed (Altshuler et al. 2008). Large scale deletions and insertions, known as CNVs (see 1.2.2.3), are known to account for a substantial fraction of human genetic variation (Conrad et al. 2009, Hurles et al. 2008), and have been shown to play a role in human evolution and variation in gene expression (Perry et al. 2007, Stranger et al. 2007). Therefore, the involvement of CNVs in common diseases is highly plausible. The SNP microarrays that have been used in GWAS can be used to detect a small portion of CNVs indirectly, however, the mast majority remain invisible. Instead, high-resolution tiling arrays have been used for exploring CNVs in some areas, but they break down for the large fraction containing repetitive elements (Emanuel and Saitta 2007). The new generation of genotyping arrays already contains some CNVs, and the 1000 genomes project will be essential for developing arrays with even better coverage of the structural variants (McCarroll 2008). Epigenetics represent inherited information not carried in the DNA, usually in the form of chemical modification of DNA without changing the sequence and may influence human disease risk (reviewed in (Soejima 2009)), but are not detectable by GWAS, since the technology is based on DNA sequence and epigenetics needs to be tackled with a new type of high-throughput technology. Despite the failure to uncover the majority of genetic risk for common diseases, GWAS has contributed substantially to our understanding of disease mechanisms and we are now approaching the “post-GWAS era”, collectively with new challenges. An explosion of targeted deep sequencing will be seen in near future and finally, complete sequencing of hundreds to thousands of individuals will hopefully help us to pinpoint new disease variants and give us further insight into pathophysiological pathways of complex diseases.

1.4 GENETICS OF PRE-ECLAMPSIA

The identification of genes predisposing for pre-eclampsia has not been very successful over the years, with no universally accepted susceptibility gene. However, both large epidemiological and family studies demonstrate genetic contribution to the pre-eclampsia susceptibility. Genome-wide linkage analysis is one strategy for identifying candidate loci. So far, three loci showing significant linkage to pre-eclampsia have been identified on chromosomes 2p13, 2p25 and 9p13 (Arngrimsson et al. 1999, Laivuori et al. 2003) and evidence of suggestive linkage has been reported for eight additional chromosomal loci (Arngrimsson et al. 1999, Harrison et al. 1997, Hayward et al. 1992,
A summary of all loci can be found in Table 3. Although several loci have been mapped, only a few promising positional candidate genes have been identified so far. Parent-of-origin effect to the storkhead box 1 gene (STOX1) on chromosome 10 was suggested in the Dutch population (see 1.4.2.3) and evidence for association has been reported in two separate populations for activin receptor type II (ACVR2A) at 2q22 and endoplasmic reticulum aminopeptidase 2 (ERAP2) at 5q (see 1.2.2.2). (Fitzpatrick et al. 2009, Roten et al. 2009). Several genes have been suggested to mediate susceptibility based on functional properties and a number of studies have reported association to genes encoding for coagulation factors, oxidative stress and vasoactive substances, but attempts to replicate these findings have yielded inconsistent results.

The lack of consistent reproducibility and convincing results has been observed for many complex disorders and pre-eclampsia is no exception. Possible explanations include genetic heterogeneity, phenotyping, possible involvement of fetal genes and study power (reviewed in (Chappell and Morgan 2006)). Efforts are required to carefully select the study population and to minimize phenotypic differences. Several classifications of pre-eclampsia are used still used and the definition does not always have to reflect the different pathogenic mechanisms. For example, it has been suggested that familial and sporadic representation of pre-eclampsia have different genetic background (Oudejans et al. 2007). Furthermore, the severity of the disorder as well as involvement of various organs is highly variable between women. Most of the genetic studies in pre-eclampsia until today have focused on maternal genes, but it is possible that both maternal and fetal genes play a role and it is not possible to distinguish between these using a case-control design. Although there are few genes and gene variants that look promising, allelic and locus genetic heterogeneity can be a plausible explanation for the lack of replication. Finally, larger studies of statistical power to detect small genetic effects are needed in combination with comprehensive SNP typing in order to pinpoint the causal allele to reliably identify or exclude susceptibility genes in pre-eclampsia.

1.4.1 Genetic susceptibility

The incidence of pre-eclampsia is three times higher among women with a family history of pre-eclampsia than among women without, suggesting a strong familial predisposition (Esplin et al. 2001). The past research has focused primarily on the maternal factors that predispose to pre-eclampsia. However, the observation that a high number of monozygotic twin sets are discordant for the development of pre-eclampsia during their own pregnancies suggests that environmental factors and the fetal genotype are also important in determining susceptibility (Thornton and Macdonald 1999, Thornton and Onwude 1991, Treloar et al. 2001). A study including both monozygotic and dizygotic twin pairs estimated that genetic and environmental effects are of equal importance (Salonen Ros et al. 2000). Additionally, Cnattingius et al analyzed over 200 000 sibling pairs and their pregnancy outcomes and estimated that genetic factors contribute to the incidence of pre-eclampsia for more than 50% of the total phenotypic variance, with 35% of this variance originating from maternal effects and 20% from fetal effects combining genetic factors originating from both parents (Cnattingius et al. 2004). The contribution of the paternal genotype to pre-
Eclampsia is supported by large database studies. One of them shows that mothers who become pregnant by a man who has fathered a pre-eclamptic pregnancy in another woman have nearly a double risk in their own pregnancies (Lie et al. 1998). Additionally, children of a man born from a pre-eclamptic pregnancy have twice the risk to be born from a pre-eclamptic pregnancy themselves than children of men born from an uncomplicated pregnancy (Esplin et al. 2001). Pre-eclampsia in the second-generation has also been studied and men born from a pre-eclamptic pregnancy have 1.5-fold increased risk of fathering a pre-eclamptic pregnancy (Skjaerven et al. 2005). No clear inheritance pattern of pre-eclampsia is known. Family studies designed to test different models of inheritance have reported inconclusive findings, including a single maternal recessive gene (Sutherland et al. 1981), homozygosity for a single recessive gene shared by mother and fetus (Liston and Kilpatrick 1991), and either recessive or dominant gene inheritance (Arngrimsson et al. 1990). Adding to the genetic complexity of pre-eclampsia, possible involvement of mitochondrial inheritance has been reported (Folgero et al. 1996). However, a further study has excluded direct contribution of mitochondrial genes to pre-eclampsia (Lie et al. 1998). Finally, it has also been proposed that the unclear inheritance pattern might be related to maternally or paternally imprinted genes (Graves 1998).

1.4.2 Susceptibility loci for pre-eclampsia identified by linkage

1.4.2.1 Chromosome 2p25 and 9p13

The most recent genome-wide scan in 15 Finnish pre-eclamptic multiplex families found two loci exceeding the threshold for significant linkage on chromosome 2p25 (NPL score 3.77; P=0.000761) and 9p13 (NPL score 3.74; P=0.000821 (Laivuori et al. 2003). Two other linkage studies from Iceland and Australia/New Zealand have reported significant linkage to this chromosome, but the 2p25 loci is clearly different from 2p12 and 2q22-23 (Arngrimsson et al. 1999, Moses et al. 2000). The families in the study were recruited predominantly from the Kainuu province in Finland, a known founder population, to reduce genetic heterogeneity. Significant linkage was only observed when a general diagnostic criteria (pre-eclampsia, eclampsia or GH without proteinuria) were used. However, it is likely that a woman in a pre-eclampsia family with GH shares genetic risk factors with her affected relatives with diagnosed pre-eclampsia.

The chromosome 2p25 region directly overlaps a previously reported risk locus for general hypertension (Angius et al. 2002, Zhu et al. 2001). The gene for Rho-associated coiled-coil protein kinase 2 (ROCK2), previously implicated in general hypertension, is located in the region of overlap. ROCK2 is widely expressed, but appears to be most abundant in smooth muscle cells, suggesting a role in vascular contraction (Riento and Ridley 2003). Interestingly, expression levels of RhoA mRNA have been shown to be significantly higher in pre-eclamptic placentas compared to placentas from normotensive women (Friel et al. 2008). However, the gene did not show any evidence for association to pre-eclampsia in a Finnish population (Paper III) (Peterson et al. 2009). A another plausible candidate gene found in the 2p25 region is lipin 1 (LPIN1). LPIN1 gene encodes for phosphatidate phosphatase (PAP) enzyme required for glycerolipid biosynthesis (Carman and Han 2006). Several recent studies have
confirmed the relationship between adipose tissue lipin-1 levels and insulin sensitivity in humans (Reue 2009). Metabolic changes, such as reduced glucose utilization, hyperinsulinemia and hyperlipidemia are characteristic for pre-eclampsia, making LPIN1 an interesting candidate gene (Kaaja et al. 1999, Kaaja et al. 1995).

The second significant finding at 9p13 has been shown to be a candidate region for T2D in two genome-wide studies from Finland and China (Lindgren et al. 2002, Luo et al. 2001). Additionally, chromosome 9p21, near the pre-eclampsia linkage borders, has shown association for both T2D and CAD in several GWAS during the past 3 years (Helgadottir et al. 2007, McPherson et al. 2007, Samani et al. 2007, Saxena et al. 2007, Scott et al. 2007, WTCCC 2007, Zeggini et al. 2007). There is a clear relationship between pre-eclampsia, CAD, and T2D. For example, pre-eclampsia predisposes women for elevated lifelong risk of CAD (Carpenter 2007, Rodie et al. 2004) and the risk of subsequent T2D is increased more than 3-fold after severe pre-eclampsia (Lykke et al. 2009). Therefore, it is possible that the three different disorders share genetic risk factors.

1.4.2.2 Chromosome 2q22 and 5q - ACVR2A and ERAP2

Moses et al identified a suggestive linkage region at 2q23 in Australian/New Zealand families and they concluded that their findings were consistent with the Arngrimsson et al study and therefore proposed that the linked locus should be designated the “PREG1” (pre-eclampsia, eclampsia gene 1) locus (Fitzpatrick et al. 2004, Moses et al. 2000). However, none of the other linkage studies have pointed to this region. The 2q locus was then strengthened and resolved further to 2q22 by fine-mapping and the application of a genetic linkage method assuming that the underlying liability of pre-eclampsia susceptibility is inherently quantitative (Moses et al. 2006). The same approach was used by Johnson et al for re-analysis of the original genome-wide scan data and led to the identification of two additional susceptibility regions at 5q and 13q (Johnson et al. 2007). Subsequently, a gene prioritization strategy assigned the highest priority to the ACVR2A gene at 2q22, which is a receptor for the cell signaling protein activin A (Moses et al. 2006). Activin A is known to play a role in cell differentiation, implantation and decidualization (Caniggia et al. 1997, Jones et al. 2006). Evidence for association was found for ACVR2A in both an extended cohort of the Aus/NZ families and an independent Norwegian case-control cohort (Fitzpatrick et al. 2009, Roten et al. 2009). At the 5q locus, several genes were investigated and significant association for the endoplasmatic reticulum aminopeptidase 2 (ERAP2) was found in both families and Norwegian case-control data set (Johnson et al. 2009). The ERAP2 gene encode for an enzyme that play a role in blood pressure regulation, in addition to a suggested role in innate immune and inflammatory responses, which are known pathways involved in pre-eclampsia pathophysiology (Johnson et al. 2009, Tanioka et al. 2003).

1.4.2.3 Chromosome 10q22 - STOX1

Lachmeijer et al studied affected sib-pairs families and found suggestive linkage with pre-eclampsia to chromosomes 10q22 and 22q12 (Lachmeijer et al. 2001). The 10q22 locus was confirmed and maximal allele sharing was shown between pre-eclamptic sisters for the maternal, but not paternal alleles (Oudejans et al. 2004). Van Dijk et al identified five different missense mutations in the gene STOX1 at 10q22, co-
segregating with pre-eclampsia in the families, identical between affected sisters and following matrilineal inheritance (van Dijk et al. 2005). One of the mutations, Y153H, was suggested to be highly mutagenic according to conservation analysis of the protein. They hypothesize that STOX1 could be involved in the transition process of noninvasive to invasive extravillous trophoblasts, known to be central in normal placental development. However, Iglesias-Platas et al showed that expression of STOX1 was very low in placentas and invading trophoblastic cells, and detected biallelic expression in both normal and pre-eclamptic placentas (Iglesias-Platas et al. 2007). Additionally, two other studies investigating STOX1 variants in Dutch and Finnish (Paper I) populations failed to detect any association to pre-eclampsia (Berends et al. 2007, Kivinen et al. 2007). However, Rigourd et al suggested that STOX1 should still be considered as a key factor in pre-eclampsia, since their study strongly suggests that STOX1 plays a major part in normal placental development (Rigourd et al. 2008). They hypothesize that STOX1 is able to act directly on specific gene targets and induce a transcriptional profile approximating that observed in pre-eclampsia. A recent study comparing gene expression of early placentas (10-12 weeks of gestation) of women later developing pre-eclampsia or leading to a normal gestation, showed that STOX1 was induced 2-fold in pre-eclampsia, suggesting that a dysregulation of STOX1 occurs early in pre-eclampsia (Founds et al. 2009).

1.4.2.4 Other reported linkage based studies

The first genome-wide study for pre-eclampsia was designed as an exclusion map, and suggested a role for a gene or genes on chromosomes 1, 3, 9 and 18 (Hayward et al. 1992). The second scan suggested the presence of a candidate region on chromosome 4q34 in an Australian population (Harrison et al. 1997). An Icelandic study, comprising a large number of affected women, revealed a maternal susceptibility locus for pre-eclampsia on chromosome 2p13 and was the first one to meet the criteria for genome-wide significance (Arngrimsson et al. 1999). However, findings were mostly supported by two large pedigrees and exclusion of those lead to identification of a suggestive locus on 2q23, the same region later found linked to pre-eclampsia in Australian/New Zealand families, where ACTRV2 was identified as a susceptibility gene. Additionally, Angrimsson et al has reported linkage to both the Angiotensinogen gene (AGT) at 1q42-43 and Endothelial nitric oxide synthase (eNOS3) at 7q36 in two separate studies before their whole-genome screen was published (Arngrimsson et al. 1993, 1997).

<table>
<thead>
<tr>
<th>Country</th>
<th>Criteria</th>
<th>Markers</th>
<th>Families</th>
<th>Chromosome with nominal or suggestive linkage</th>
<th>Chromosome with significant linkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scotland</td>
<td>S</td>
<td>43</td>
<td>35</td>
<td>3, 9</td>
<td>4q34 (G)</td>
</tr>
<tr>
<td>Australia</td>
<td>S, G</td>
<td>90</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>S, G</td>
<td>440</td>
<td>124</td>
<td>2q23 (G)</td>
<td>2p13 (G)</td>
</tr>
<tr>
<td>Aus,NZ</td>
<td>S, G</td>
<td>400</td>
<td>34</td>
<td>2q23 (S), 11q23 (G)</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>S, HELLP</td>
<td>293</td>
<td>67</td>
<td>10q, 22q (S), 12q (HELP)</td>
<td>2p25 (G), 9p13 (G)</td>
</tr>
<tr>
<td>Finland</td>
<td>S, G</td>
<td>435</td>
<td>15</td>
<td>4q32 (G)</td>
<td></td>
</tr>
</tbody>
</table>

(Aus, NZ) Australia, New Zealand
(HELP) Hemolysis, elevated liver enzymes, and low platelets
1.4.3 Association-based mapping approaches

Candidate gene studies have failed to identify universally accepted susceptibility genes for pre-eclampsia, but studies of more than 50 candidates have been reported. However, around 70% of the published studies have focused on only eight genes (Factor V Leiden (F5), Methylene tetrahydrofolate (MTHFR), Prothrombin (F2), Angiotensin converting enzyme (ACE), Angiotensin II type 1 and type 2 receptor (AGTR1, AGTR2), Angiotensinogen (AGT), Endothelial nitric oxide synthetase 3 (eNOS3) and Tumor necrosis factor α (TNF-α) (reviewed in (Chappell and Morgan 2006, Mutze et al. 2008)). The candidate genes studied belongs to different groups according to their functional properties and plausible role in the pathophysiology. The different subgroups and most important genes are reviewed below and are summarized in table 4.

1.4.3.1 Trombophilia

Since several studies have shown that pre-eclampsia is associated with underlying thrombophilic disorders (Dekker 2005, Kupferminc et al. 1999, van Pampus et al. 1999), placental interactions and villous thrombosis are characteristic for pre-eclampsia, genes in this category have been suggested as candidates and have been investigated in a large number of association studies. F5, MTHFR and F2 are three of the most widely studied genes in pre-eclampsia and thrombophilia mutations have been identified in those genes in pre-eclampsia, but several studies have also provided contradictory results (reviewed in (Mutze et al. 2008). However, a meta analysis reported by Lin et al, including all studies for F5 up to November 2002, showed a 2-fold increase in pre-eclampsia risk for a 1691G>A mutation, but no effect for mutations in MTHFR or F2 (Lin and August 2005). To date, the number of studies not confirming the mutations as pre-eclampsia risk factors for all three genes is much higher than the ones confirming. A 4G/5G deletion/insertion polymorphism of the plasminogen activator factor-1 gene (PAI-1), an inhibitor of fibrinolysis, has been associated to pre-eclampsia, but several replication attempts have failed (Dalmaz et al. 2006, Fabbro et al. 2003, Gerhardt et al. 2005, Morrison et al. 2002). The β integrin protein has a suggested role in the process of cytotrophoblast failing to adopt a vascular phenotype in pre-eclampsia (Zhou et al. 1997). Therefore, the Integrin glycoprotein IIIa gene (GPIIIa), encoding for a subunit of the β integrin, has been investigated, but with conflicting results (O’Shaughnessy et al. 2001, Pegoraro et al. 2003).

1.4.3.2 Hemodynamics, endothelial function and angiogenesis

The renin-angiotensin-aldosterone system (RAAS) is known to play an important role in regulation of renal function and arterial pressure during pregnancy and several studies have implicated the RAAS in pre-eclampsia pathophysiology (reviewed in (Shah et al. 2006). Therefore, genes in this system have been considered as plausible candidates for pre-eclampsia (Gilbert et al. 2008). ACE, AGTR1, AGTR2, and AGT have all been studied extensively in pre-eclampsia with inconsistent results (reviewed in (Mutze et al. 2008)). A gene associated to RAAS is the eNOS3, widely expressed in placental tissue (Myatt et al. 1991), but with reduced activity in pre- eclamptic placentas (Brennecke et al. 1997), which could lead to reduced vasodilation, and vascular remodeling. Association studies in populations with different ethnical background have yielded both positive and negative findings. A meta-analysis including studies up to
2006 investigating a E298D polymorphism, initially associated to pre-eclampsia in Colombian women, did not report an increased risk for that particular polymorphism (Medica et al. 2007, Serrano et al. 2004). Vascular endothelial growth factor (VEGF) is an angiogenic factor and plays an important role in endothelial cell proliferation, migration, survival and regulation of vascular permeability. Soluble fms-like tyrosine kinase-1 (sFlt-1) is the soluble form of the VEGFR-1, an antagonist for VEGF, and up-regulated in pre-eclampsia (Levine et al. 2004, 2006, Maynard et al. 2003, Staff et al. 2005, Thadhani et al. 2004). Interestingly, during placental development, the expression of VEGFR-1 is not only detected in vascular endothelial cells, but also in the developing trophoblasts (Clark et al. 1996) and it has been shown that increased sFlt-1 levels can be detected approximately five to six weeks before the onset of clinical symptoms of pre-eclampsia. The number of studies that has investigated SNPs in the genes involved in the VEGF system is very small. Two polymorphisms of the VEGF, 405G>C and 936C>T were found associated to the severe form of the disease in two small studies (Banyasz et al. 2006, Papazoglou et al. 2004) and are not considered as major risk factors.

1.4.3.3 Oxidative stress and lipid metabolism

Placental oxidative stress, which is the result of abnormal placentation, reduced blood flow and subsequently a hypoxic placenta, is reported to play a central role in the pathogenesis of pre-eclampsia (Gupta et al. 2005). Several genes involved in the generation of ROS could, if they are defective in some way, increase endothelial dysfunction through lipid peroxidation. Despite the strong correlation between oxidative stress and pre-eclampsia, only a few genes have been investigated. Polymorphisms in the genes microsomal epoxide hydrolase (EPHX) that catalyzes the hydrolysis of certain oxides and may produce toxic intermediates that could be involved in pre-eclampsia, and glutathione S-transferase (GST), a protein protecting cells against ROS, have shown association. However, conflicting results have been reported (Canto et al. 2008, Gebhardt et al. 2004, Laasanen et al. 2002, Ohta et al. 2003, Zusterzeel et al. 2000, 2001).

Lipid metabolism is associated with oxidative stress and abnormal lipid profiles have been observed in pre-eclampsia (Sattar et al. 1997). Lipoprotein lipase (LPL) and Apolipoprotein E (ApoE) are the two major regulators of lipid metabolism, abundantly expressed in placenta and have therefore been proposed as possible candidate genes (Descamps et al. 2005, Kim et al. 2001). A missense mutation Asn291Ser in LPL seems to be the most promising variant, since it is associated with lowered LPL activity and increased dyslipidemia in two separate studies. However, others have failed to replicate these findings (Hubel et al. 1999, Kim et al. 2001, Zhang et al. 2006). It has also been reported that the fetal genotype of the two genes could contribute to the metabolism of maternal lipoproteins (Descamps et al. 2005).

1.4.3.4 Immunoregulation

It has been suggested that a maternal-fetal (paternal) immune maladaptation is central in pre-eclampsia pathogenesis and genes encoding for various factors of the immune system, therefore, appear to be good candidates. Several HLA molecules have been
TNFα, a pro-inflammatory cytokine, has been implicated in the etiology. It has been shown that an excessive release of TNFα by activated macrophages and monocytes may contribute to endothelial activation and damage that later result in the maternal syndrome of pre-eclampsia (LaMarca et al. 2007). Interestingly, TNFα induces hypertension in pregnant, but not in non-pregnant rats (Alexander et al. 2002) and plasma levels of TNFα in pre-eclamptic women are significantly higher compared to matched controls (Sharma et al. 2007). TNF is also involved in the generation of ROS and subsequently oxidant-mediated endothelial cell injury (Chen et al. 1996). One of the most commonly studied variants in pre-eclampsia is -308G>A transition in the promoter region that is associated with higher TNF levels and an increased risk for pre-eclampsia linked disorders T2D, CAD and dyslipidemia (Elahi et al. 2009, Saarela et al. 2005). A meta-analysis from 2008 combined sixteen studies investigating the promoter SNP, but did not detect significant association to pre-eclampsia (Bombell Mcguire 2008). Several cytokines are produced at the maternal-fetal interface and have an impact on trophoblast invasion. It has been suggested that deficiency of IL-10 may contribute to an enhanced inflammatory response towards trophoblast cells, resulting in defective transformation of spiral arteries, hypoxia, thrombosis and infarction of the placenta (Renaud et al. 2007). The gene expression of IL-10 is reduced in pre-eclamptic placentas (Makris et al. 2006) and studies have investigated variants in the gene, but with conflicting results (Daher et al. 2006, Goddard et al. 2007, Haggerty et al. 2005, Kamali-Sarvestani et al. 2006). Association has been detected to two additional pro-inflammatory genes, Interleukin 1α (IL-1α) and the interleukin 1 receptor antagonist (IL1Ra), but few studies have addressed the role of polymorphisms in the genes so far (Faisal et al. 2003, Haggerty et al. 2005).

1.4.3.5 Placentation

Defective placentation is known to be central in pre-eclampsia pathophysiology, as a result of defective transformation of spiral arteries, reduced blood flow through the placenta and subsequently ischemia. Therefore, placental genes appear to be good candidates, however, not many genes have been studied and the results are not convincing (reviewed in (Roberts and Cooper 2001)). Expression studies of pre-eclamptic placentas show evidence for matrix metalloproteinases (MMPs) and apoptosis related gene, both with a possible implication in trophoblast invasion (Pang and Xing 2003, 2004, Sitras et al. 2009). However, association studies of those genes are very limited and effects of variants in the genes need to be further investigated. The best candidate that can be implicated in placentation to date is STOX1 that was positionally cloned following a genome-wide linkage scan (see 1.4.2.3). Interestingly, a study by Rigourd et al suggests that STOX1 plays a major part in normal placental development and is able to act directly on specific gene targets, and induce a transcriptional profile approximating that observed in pre-eclampsia (Rigourd et al. 2008). It is plausible that genes working in the same pathway as STOX1 could influence pre-eclampsia risk as well. One interesting candidate is glial cells missing homolog 1 (GCM1), which is a crucial transcription factor for placental development, down-regulated in pre-eclamptic placentas (Chen et al. 2004) and there is evidence suggesting that GCM1 could be under regulation of STOX1 at both a transcriptional level and by proteasome-mediated post-translational degradation (Rigourd et al. 2009).
Table 4. Summary of investigated functional candidate genes in pre-eclampsia confirmed at least in one study

<table>
<thead>
<tr>
<th>Gene name (symbol)</th>
<th>Immunoregulation</th>
<th>Hemodynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trombophilia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor V Leiden (FV)</td>
<td>Tumor necrosis factor α (TNFα)</td>
<td>Angiotensinogen (AGT)</td>
</tr>
<tr>
<td>M ethylenetetrahydrofolate (MTHFR)</td>
<td>Interleukin 1 receptor antagonist (IL1Rα)</td>
<td>Angiotensin converting enzyme (ACE)</td>
</tr>
<tr>
<td>Prothrombin (F2)</td>
<td>Interleukin 1α (IL-1α)</td>
<td>Angiotensin II type 1 receptor (AGTR1)</td>
</tr>
<tr>
<td>Plasminogen activator factor-1 (PAI-1)</td>
<td>Interleukin 10 (IL-10)</td>
<td>Angiotensin II type 2 receptor (AGTR1)</td>
</tr>
<tr>
<td>Integrin glycoprotein IIIa (GPIIIa)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oxidative stress and lipid metabolism</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein E (ApoE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipoprotein lipase (LPL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microsomal epoxide hydrolase (EPHX)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutathione S-transferase (GST)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Endothelial function and angiogenesis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endothelial nitric oxide synthetase 3 (eNOS3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascular endothelial growth factor receptor 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascular endothelial growth factor (VEGF)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\section*{2 AIMS OF THE THESIS}

The overall aim of this study was to search for genes that may predispose to pre-eclampsia.

The specific aims were:

\begin{enumerate}
\item To assess the previously published pre-eclampsia candidate \textit{STOX1} for associations in Finnish population. (Paper I)
\item To evaluate previously published type 2 diabetes (T2D) and coronary artery disease (CAD) susceptibility loci on chromosome 9p21 for increased pre-eclampsia risk. (Paper II)
\item Fine-map the previously implicated pre-eclampsia susceptibility locus on chromosome 2p25 and identify the most likely susceptibility gene within the region. (Papers III and IV)
\end{enumerate}
3 MATERIAL AND METHODS

3.1 STUDY SUBJECTS

Individuals were recruited mainly from Finland, but samples from United Kingdom and Australia/New Zealand were utilized in Paper II. The number of families and independent cases and controls from each population used in each study are presented in Table 5 and the study materials are described in detail below.

Table 5. Sample collections and numbers of individuals used in each of the studies

<table>
<thead>
<tr>
<th>Finland</th>
<th>United Kingdom</th>
<th>Australia/New Zealand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic DNA</td>
<td>Genomic DNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>Placental samples</td>
<td>Case-controls</td>
<td>Case-controls</td>
</tr>
<tr>
<td>Families</td>
<td>Case-controls</td>
<td>Case-controls</td>
</tr>
<tr>
<td>15</td>
<td>340/357</td>
<td>8/6</td>
</tr>
<tr>
<td>Study I-IV</td>
<td>Study I-IV</td>
<td>Study I-IV</td>
</tr>
<tr>
<td></td>
<td>251/198</td>
<td>Study II</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Families</td>
</tr>
<tr>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study II</td>
</tr>
</tbody>
</table>

3.1.1 Finnish families (I-IV)

In total, the family cohort consisted of 15 multiplex families recruited predominantly from the Kainuu province in central eastern Finland and in the Helsinki region for the original genome-wide scan that formed the basis of this thesis. The families (174 individuals) were recruited through pre-eclampsia probands by sending letters to patients who had been hospitalized with the diagnosis of severe pre-eclampsia in their first pregnancies in Kainuu Central Hospital (Kajaani, Finland) or the Department of Obstetrics and Gynecology, Helsinki University Central Hospital (Helsinki, Finland) and by advertisement for the study, between 1988 and 1997. Ten multiplex families were recruited from Kainuu and five from Helsinki. To verify the origin of families, the birthplaces of parents and grandparents were traced by using church records. The reproductive medical records of all women were reviewed to confirm or exclude the diagnosis. Pre-eclampsia was defined as systolic blood pressure >140 mm Hg or diastolic blood pressure >90 mm Hg in a woman who was normotensive before 20 weeks of gestation, and the urinary excretion of ≥0.3 g protein in a 24-h specimen or ≥1+ reading on dipstick in a random urine determination at least twice with no evidence of urinary tract infection. For more detailed description of definitions, see (Laivuori et al. 2003). In families that included at least one patient with pre-eclampsia, women with gestational hypertension (GH) without significant proteinuria were considered to have mild pre-eclampsia under the general diagnostic criteria. In total, fifty women met the criteria of pre-eclampsia or eclampsia and severe pre-eclampsia was diagnosed in 29 of these 50 women (58%). Thirteen women met the criteria of GH in their first pregnancies. Each family was ascertained on the basis of at least one proteinuric pre-eclampsia case. Clinical characteristics for patients are summarized in Table 6.
3.1.2 Finnish case-controls (I-IV)

A nation-wide sample set consisting of 340 cases and 357 controls were recruited by combining the National Hospital Discharge register with diagnosis classified according to International Classification of Diseases (ICD-10 1996) with The National Register of Blood Groups and Blood Group Antibodies of Pregnant, from which data were obtained for 100,000 consecutive pregnancies during 1997-1998. In the cohort of 100,000 pregnancies, 686 were picked as case-candidates, medical records were reviewed by two clinicians and pre-eclampsia was defined according the definitions of the NHBPEP (NHNPEP, 2000). Phenotypes were ascertained by two clinicians and further subdivided into pre-eclampsia (N=245), eclampsia (N=5), superimposed pre-eclampsia (N=46) and GH without proteinuria (N=44). Controls were matched with cases for the province of residence, number of pregnancies, age of mother and number of fetuses. The cases and controls were invited by letters and reminders. All participants gave their written informed consent and the study was approved by the ethics committee of Finnish Red Cross Blood Service, and Ministry of Social Affairs and Health. For more detailed description of the recruitment process and diagnostic criteria see (Hiltunen et al. 2009). Clinical characteristics for cases and controls are summarized in Table 6.

Table 6. Clinical characteristics for the two Finnish sample sets studied in this thesis

<table>
<thead>
<tr>
<th></th>
<th>Finnish families</th>
<th>Finnish case-controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strict General</td>
<td>Strict General Controls</td>
</tr>
<tr>
<td>Cases (N=50) Cases (N=63)</td>
<td>Cases (N=250) Cases (N=294)</td>
<td>Controls (N=357)</td>
</tr>
<tr>
<td>Mean ±SD Mean ±SD</td>
<td>Mean ±SD Mean ±SD</td>
<td>Mean ±SD</td>
</tr>
<tr>
<td>Age (years) 27.2 ± 4.4 26.7 ± 4.4</td>
<td>28.9 ± 5.59 29.1 ± 5.57</td>
<td>28.8 ± 4.99</td>
</tr>
<tr>
<td>Body mass index (kg/m²)* 21.2 ± 2.3 21.4 ± 2.4</td>
<td>24.2 ± 4.6 24.3 ± 4.5</td>
<td>22.8 ± 3.6</td>
</tr>
<tr>
<td>Systolic BP (mmHg) 156.5 ± 27.1 155.8 ± 24.8</td>
<td>169.8 ± 16.4 168.3 ± 16.4</td>
<td>NA</td>
</tr>
<tr>
<td>Diastolic BP (mmHg) 103.6 ± 16.3 103.5 ± 14.8</td>
<td>105.6 ± 9.5 105.3 ± 9.2</td>
<td>NA</td>
</tr>
<tr>
<td>Proteinuria (g/24h) ** NA</td>
<td>4.3 ± 4.2 NA</td>
<td>NA</td>
</tr>
<tr>
<td>Birth weight 2482.3 ± 929.2 2615.5 ± 895.8</td>
<td>2622.3 ± 835.6 2709.3 ± 847.9</td>
<td>3569.2 ± 410.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strict criteria</th>
<th>Women with pre-eclampsia or eclampsia</th>
</tr>
</thead>
<tbody>
<tr>
<td>General criteria</td>
<td>Women with pre-eclampsia, eclampsia or gestational hypertension without proteinuria</td>
</tr>
</tbody>
</table>

** ≥ 0.3 g protein in 24-h specimen or ≥ plus 1 or more reading on dipstick in a random urine determination at least twice with no evidence of urinary tract infection.

* Pre-pregnancy body mass index

3.1.3 UK case-controls (II)

The sample set consisting of 102 women with pre-eclampsia and 198 women with healthy normotensive pregnancies that were recruited during pregnancy in maternity units in Nottingham, UK. All patients were of white Western European origin and were recruited according to a definition from the International Society for the Study of Hypertension in Pregnancy (ISSHP) (Brown et al. 2001). A further 149 cases were white Western European women recruited from 25 UK hospitals, using the same diagnostic criteria, to a study investigating prophylactic oral vitamins C and E in women at high risk of pre-eclampsia. In these women, a history of eclampsia or pre-eclampsia in a previous pregnancy leading to delivery before the 37th week of gestation was used as the sole risk factor for pre-eclampsia. Exclusion criteria applied to both
cases and controls were: previous chronic hypertension, renal or cardiovascular disease, hydatidiform mole in the index or previous pregnancy and multiple pregnancy. Detailed information for the two sample collections can be found here (Morgan et al. 1999, Poston et al. 2006).

3.1.4 Australian/New Zealand families (II)

The cohort consists of 480 individuals from 74 pre-eclampsia families recruited over a period of 15 years through the Royal Women’s Hospital and the Monash Medical Centre in Melbourne, Australia, through print and electronic advertisements in Sydney and through the National Women’s Hospital in Auckland, New Zealand. Thirty-four of the families have been used to identify a susceptibility locus for pre-eclampsia on chromosome 2q22 locus (Moses et al. 2000). In summary, the cohort included 20 women with eclampsia, 120 women with pre-eclampsia, 56 women who had GH without proteinuria, and 90 women who had normal pregnancies. Diagnosis was based on clinical assessment, using criteria of the ASSHP (ASSHP 1993). Women who met these criteria and experienced convulsions or unconsciousness in the prenatal period were classified as having eclampsia. Women with the pattern of hypertension outlined above but with no proteinuria were classified as mild pre-eclampsia. Women with pre-existing hypertension or other medical conditions known to predispose them to pre-eclampsia (e.g. renal disease, diabetes, twin pregnancies or fetal chromosomal abnormalities) were excluded. Ethical approval was obtained from the research and ethics committees of the Royal Women’s Hospital, Melbourne, and written informed consent was obtained from family members.

3.1.5 Finnish placental samples (I-IV)

Placental samples were obtained immediately after delivery from eight primiparous pre-eclamptic women and six primiparous controls at the Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Finland. Samples were included only if there was no indication of diabetes, renal disease, twin pregnancy, smoking, or any other condition known to affect pre-eclampsia risk or placental function. Pre-eclampsia was defined according to the NHBPEP (NHBPEP 2000). The biopsies were taken in a separate room dedicated for handling of tissue samples, snap frozen in liquid nitrogen and stored at -80°C until further analysis.

3.1.6 Finnish control samples (II)

A Finnish population-wide sample set consisting of 260 male blood donors who had given informed consent and were collected through the Finnish Red Cross Blood Service. This sample set was collected for the study of population genetic variation in Finland (Salmela et al. 2008).
3.2 GENETIC ANALYSIS

3.2.1 Marker selection (I-IV)

In Paper IV, microsatellite markers in the 2p25 linkage region were selected primarily from the deCODE genetic map (Kong et al. 2002). The Marshfield (http://research.ashfieldclinic.org/genetics/) genetic map was used to fill the gaps when no deCODE markers were available. The order and physical distances between markers were established using the NCBI (http://www.ncbi.nlm.nih.gov/units/) and Ensembl (http://www.ensembl.org/) public databases. SNPs in Paper IV were chosen from the SNP consortium (TSC) database and dbSNP based on availability and information content. Information regarding variant tagging properties and population frequencies were limited at the time the study was designed, but re-sequencing a genomic region allowed us to identify novel SNPs that were chosen for genotyping. Coding SNPs in STOX1 (I) were selected based on the work by van Dijk and others (van Dijk et al. 2005). An additional SNP was selected for the 3' end of the gene based on its tagging properties to cover the whole gene region. Tagging SNPs were chosen for the plausible candidate genes ROCK2 (III), TRIB2, GREB1, LPIN1 and E2F6 (IV) at the 2p25 locus. The SNP selection was performed in the CEU population genotype data downloaded from HapMap (http://www.hapmap.org) with the Tagger program implemented in Haploview (Barrett et al. 2005, de Bakker et al. 2005). Two SNPs in GREB1 were chosen based on previously reported association to hypertension (Kamide et al. 2005). Additionally, ROCK2 (III), LPIN1 and TRIB2 (IV) were re-sequenced in affected individuals from a subset of families contributing to the 2p25 linkage signal to increase our chances of identifying the most relevant functional variants and other novel variants. SNPs were selected for genotyping based on the sequencing results (non-synonymous coding, splice site, other functional SNPs, insertion/deletion polymorphisms or observed allele frequency differences between cases and controls). Markers at the 9p21 locus (II) were selected based on previous GWA studies showing association for T2D and CAD, and tagging properties to cover majority of common variation within the region (Helgadottir et al. 2007, McPherson et al. 2007, Samani et al. 2007, Saxena et al. 2007, Scott et al. 2007, WTCCC 2007, Zeggini et al. 2007).

3.2.2 Genotyping: SNPs and microsatellites (I-IV)

SNPs and microsatellites were genotyped using Sequenom (I-IV) (Sequenom Inc. San Diego, CA, USA), TaqMan (II) (Applied Biosystems, Foster City, CA, USA), Illumina (II), Affymetrix (II) and MegaBACE (I, III, IV) (Amersham Biosciences, Uppsala, Sweden) platforms.

The Sequenom methodology is based on matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry and it was our method of choice for SNP genotyping in all studies (I-IV). The typing procedure is based on an allele specific primer extension reaction of an oligonucleotide probe over a SNP site in a PCR product, with a mixture of deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs), to produce different size products for each allele of a SNP. The extension products are analyzed by MALDI-TOF mass spectrometry, and the time-of-flight is proportional to mass, permitting precise determination of the size of products.
generated, which can be converted into genotype information. This genotyping assay allows multiplexing of up to 40 SNPs simultaneously, where differentiation of genotypes for each of the SNPs in one assay is a result of unique mass ranges for the extension primers. An overview of the assay technology is shown in Figure 5. PCR assays and extension primers were designed using the SPECTROdesigner software (Sequenom). Extension products were identified by a MassARRAY mass spectrometer (Bruker, Daltonik GmbH, Bremen, Germany) and resulting mass spectra were analyzed for peak identification using the SpectroTYPER RT software (Sequenom Inc). Genotypes were verified independently by two investigators, and Hardy-Weinberg equilibrium (HWE) was assessed for each marker as a quality control procedure. To identify possible genotyping errors in the families, data were checked for Mendelian consistency using Pedcheck (O’Connell and Weeks 1998).

In Paper II, one SNP was genotyped using the TaqMan chemistry, which is based on commercially available on-demand SNP genotyping assays with fluorescent-labelled allele-specific probes provided by Applied Biosystems (Applied Biosystems, Foster City, CA, USA). The PCR reactions and allelic discriminations were performed using ABI PRISM 7900 Sequence Detection System instrument (Applied Biosystems, Foster City, CA, USA) at the Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, USA. In the same study (II) one SNP could not be genotyped with TaqMan, and Illumina’s GoldenGate assay was used instead. It is based on a combination of oligo ligation assay (OLA), allele specific extension reactions and universal PCR.

Figure 5. Schematic overview of the Sequenom methodology based on allele specific primer extension over a SNP site in a PCR product and MALDI-TOF mass spectrometry.
Automated sequencing was utilized to genotype microsatellites (IV) and SNPs (I, III, IV). The technique this method is based on is described in more detailed under the paragraph "DNA re-sequencing" (see 3.2.4). Genotyping of microsatellites was performed using PCR with fluorescently labeled primers for amplifying the microsatellite region, followed by capillary electrophoresis on a MegaBACE 500 sequencer (Amersham Biosciences). A labeled size standard was co-injected and allowed accurate size estimation of the microsatellites. Allele calling was carried out using the Genetic Profiler v 2.0 software. SNP genotyping was performed by direct sequencing of PCR products including the SNP site utilizing the DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE DNA Analysis Systems (Amersham Biosciences) and a MegaBACE 1000 instrument. The sequences were analyzed using the MegaBACE Sequence Analyser software (3.0) (Amersham Biosciences) and the Staden Package computer programs Pregap4 and Gap4 (http://staden.sourceforge.net/).

3.2.3 Data analysis

3.2.3.1 Linkage analysis (IV)

In Paper IV, linkage was assessed by the affecteds-only non-parametric multipoint linkage (NPL) analysis method using Genehunter 2.1 software, which performs reconstruction of haplotypes and complete multipoint analysis of allele sharing IBD among all affected family members at each location in the genome (Kruglyak et al. 1996).

3.2.3.2 Association analysis

Both family (II, IV) and case-control (I-IV) based association analysis have been performed using several different statistical approaches depending on the study design.

Power calculations (I-IV)
We used the Genetic Power Calculator program (http://pngu.mgh.harvard.edu/~purcell/gpc/) and the case-control for discrete traits option to evaluate our ability to detect association to certain markers in the case-control cohorts (Purcell et al. 2003). We used pre-eclampsia prevalence of 8% as reported previously in the Finnish population (Kaaja et al. 2005).

Haplotype pattern mining (IV)
Haplotype pattern mining (HPM) was used in Paper IV for the initial association analysis of SNPs genotyped in the Finnish families at the 2p25 locus, when the markers were sparsely distributed over the studied region. HPM is a data mining-based algorithm that searches for shared patterns among haplotypes, scores each marker location for the number of shared patterns, and evaluates empirically the significance of the scores by a permutation test (Toivonen et al. 2000). The same genotype data was also analyzed using both TDT and PDT.

Transmission and pedigree disequilibrium test (II, IV)
TDT is the simplest and one of the best known family-based association tests using genotype data from trios, which consist of an affected offspring and the two parents. It
was first developed to test for linkage between alleles and phenotypes in qualitative traits (Spielman et al. 1993). TDT compares the frequencies of transmitted vs untransmitted alleles in affected offspring and uses untransmitted parental alleles as controls. In paper IV, we used TDT as implemented in the Haploview software 4.1 to analyze SNP genotype data from our Finnish families. One limitation of TDT is that, although it remains a valid test of linkage, it is not a valid test of association if related nuclear families and/or sibships from extended pedigrees, such as our Finnish families, are used. The PDT on the other hand can use data from extended pedigrees and is valid, just like TDT, even when there is population substructure (Martin et al. 2003). This test uses data from related nuclear families in combination with discordant sibships. Therefore, PDT and PDTPHASE, which is an extension of PDT that can deal with missing data and haplotype analysis, were used to analyze the family data in paper II and IV (Dudbridge 2003). Additionally, LD between markers in paper IV was estimated using PDTPHASE.

Haploview (I-IV)
The Haploview software was used to assess linkage disequilibrium (LD) between markers, calculate allele and haplotype frequencies, and to run single-marker and haplotype association tests in the Finnish (I-IV) and British (II) case-control materials (Barrett et al. 2005). The Finnish material has been subdivided into pre-eclampsia, eclampsia, superimposed pre-eclampsia and GH without proteinuria. Therefore, that sample set was analyzed by using different inclusion/exclusion criteria. We used chi-square as the test statistic and corrected the overall significance for multiple testing using 10,000 permutations. Haplotype blocks were defined according to the confidence interval (CI) of D' as implicated in Haploview (Gabriel et al. 2002).

Meta-analysis (II)
In Paper II, combined Odds Ratios (OR) for marker rs7865618 were calculated by using data from all four pre-eclampsia materials. Mantel-Haenszel meta-analysis Forest plots were constructed using rmeta v2.14 on the R platform (http://www.r-project.org/) and significance was determined by a two-sided t-test.

3.2.4 DNA re-sequencing: Genomic and gene-centric mutational screening (I, III, IV)
The Sanger sequencing method is based on a technique utilizing 2',3'-dideoxy-nucleotide triphosphate (ddNTP) termination of in vitro replication of single stranded DNA (Sanger et al. 1977). A mixture of ddNTP and dNTP is utilized in sequence reactions and by including a DNA polymerase, elongation of a DNA strand will take place and terminates whenever a ddNTP is incorporated into the growing strand. ddNTPs terminate the elongation since they cannot form a phosphodiester bond with the next dNTP, resulting in DNA fragments of varying length that could be separated by size with gel electrophoresis. By fluorescently labeling the four different ddNTPs, automated sequencing was made possible with parallel termination for all ddNTPs followed by capillary electrophoresis. The automated sequencing in this study was performed utilizing the DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE DNA Analysis Systems (Amersham Biosciences). All fragments were
PCR-amplified and sequenced in both directions. Purified sequencing products were resolved using a MegaBace 1000 instrument. The sequences were analyzed using the MegaBACE Sequence Analyser software 3.0 (Amersham Biosciences) and the Staden Package computer Programs; Pregap4 for preparing the sequence trace data before entry into the assembly program and Gap4 for analysing, since it allows careful and efficient alignment of all corresponding sequences (http://staden.sourceforge.net/).

To explore the genetic variation in a 75 kb region at the 2p25 locus (IV), the non-repetitive DNA segments were PCR-amplified and re-sequenced in one pre-eclamptic woman, homozygous for the associated risk haplotype observed after the first round of SNP fine-mapping in Paper IV. The genes ROCK2 (III), LPIN1 and TRIB2 (IV) at the 2p25 locus were screened for polymorphisms by direct sequencing exons, including 100 bp flanking sequence on both sides, untranslated regions (UTRs) and 1 kb upstream of the first exon in genomic DNA from pre-eclampsia cases and controls.

3.3 EXPRESSION ANALYSIS

3.3.1 Microarray expression analysis (I-V)

Microarrays were used for genome-wide gene expression profiling in pre-eclamptic cases and controls. The expression data for the genes studied in paper I-IV are collected from the same study using placental samples from eight pre-eclamptic and six normotensive placentas analyzed on an Affymetrix GeneChip® Human Genome U133 Plus 2.0 arrays (Affymetrix, Santa Clara, CA, USA) for the analysis of over 47 000 transcripts simultaneously. Affymetric GeneChip® technology is based on oligonucleotide probes and complementary hybridization.

Placental RNA was isolated using TRIzol kit (Invitrogen, Carlsbad, CA, USA) and purified with RNAeasy total RNA isolation kit (Qiagen, Hilden, Germany), 6 µg of total RNA was reverse transcribed, amplified, labeled according to the alternative protocol for one-cycle cDNA synthesis and hybridized on the chip. Arrays were stained and scanned according to Affymetrix protocols. Transcript levels were determined from data images with Affymetrix GeneChip® DNA analysis software (GDAS). Arrays were background corrected and normalised using GC Robust Multi-array Average (GCRMA) adjustment, and differential expression was assessed using linear models from Limma. Both software packages are available from the open-source Bioconductor project, and were used under the R environment for statistical computing (Gentleman et al. 2004).

3.3.2 Allele specific expression analysis (III)

Allele specific mRNA expression of ROCK2 in Paper III was assessed in three pre-eclamptic women heterozygous for marker rs17366517. The assay was based on the comparison of allelic signal intensities in cDNA and genomic sequence from each individual. RNA isolated from our placental samples was reverse-transcribed and both cDNA and placental genomic DNA from each individual were sequenced in six independent reactions, originating from three separate PCR amplifications. Signal
intensities in each channel were compared and an allelic ratio was calculated for each sequence. To estimate the degree of attenuation of one allele, the average cDNA allelic ratio was divided by the average genomic DNA allelic ratio. Standard deviation of the measurements was calculated on replicated experiments.
RESULTS AND DISCUSSION

The aims of this thesis were to further investigate a previously identified linkage region for pre-eclampsia, evaluate the pre-eclampsia candidate STOX1, and a locus identified by GWAS associated to CAD and T2D. We found association to the CAD locus at 9p21 (II) and narrowed down the linkage region at 2p25 (IV). We did not find any evidence for STOX1 as a susceptibility gene in the Finnish population (I) and failed to confirm or reject any of the genes within 2p25 linkage region as risk factors for pre-eclampsia (III-IV). Our studies highlight the challenges of gene mapping in complex diseases and more specifically in pre-eclampsia. Implications of our results, limitations of the separate studies and challenges in genetic studies of pre-eclampsia will be discussed on the following pages.

4.1 PAPER I - EVALUATION OF STOX1 ON 10q22

In Paper I, we used three methods to assess STOX1 as a candidate gene for pre-eclampsia. STOX1 was one of the first positionally cloned susceptibility genes for pre-eclampsia, with seven coding variations co-segregating with the pre-eclamptic phenotype in the Dutch population (van Dijk et al. 2005). The suggestion of STOX1 as a susceptibility gene for pre-eclampsia was followed by attempts to replicate results in different populations. At the time our study begun, no other studies of this gene in pre-eclampsia had been reported. Today, additional reports have rejected STOX1 as a common European pre-eclampsia susceptibility factor, based on both association studies and by investigating the suggested maternal inheritance (Berends et al. 2007, Iglesias-Platas et al. 2007).

No evidence for STOX1 in the Finnish population
We aimed to genotype all seven markers co-segregating with disease in the Dutch study and one additional SNP for covering the 3’ end of the gene. Only five markers were successfully genotyped, but covered all common haplotype variants even in the absence of the three failed markers, and were therefore likely to represent a comprehensive set to be tested for association. The five markers had similar allele frequencies in Finland and in the CEPH reference material obtained from the HapMap project, and formed a single LD block with identical haplotypes in both data sets. One marker was excluded from the haplotype analysis due to its low minor allele frequency (4%). In summary, none of the single markers or haplotypes tested showed significant association in our case-control data set. It is evident that additional population samples are needed to increase the power further for detecting genetic effects that may be weaker than those that we could assess with good power. It is also possible that the specific mutations observed in the Dutch families are causal in only a subset of families, and therefore not easily detected by large population-based analyses. In that case, linkage approaches should be more powerful to implicate this locus. Therefore, we re-examined our linkage analysis results in the extended Finnish families, that had been published previously (Laivuori et al. 2003), but there was no evidence for linkage (linkage scores were negative or just above zero throughout chromosome 10).
The expression pattern of STOX1 were investigated in placental biopsies taken from pre-eclamptic and uncomplicated deliveries using Affymetrix Human Genome arrays, but neither of the two STOX1 specific probe sets showed statistically significant difference in expression levels between cases and controls.

We searched for genes that had sequence similarity to STOX1 within the linkage peaks in our genome-wide scan, since it has been suggested that paralogous genes located on different chromosomes may be responsible for the same disease, although they are likely to segregate in different populations due to different founders (Leegwater et al. 2001). We did not find obvious paralogs to STOX1 within our linkage peak regions; the closest candidate was STOX2 (28% identical to STOX1) on chromosome 4q35, but it was located >20 Mb downstream from the 3’ end of our candidate region. If other paralogs exist, their sequences may have diverged too far from STOX1 to be recognized by standard similarity search methods.

In summary, none of our three tests showed any evidence in support of STOX1 as a pre-eclampsia susceptibility gene, and we were therefore unable to confirm the association of STOX1 with pre-eclampsia in the Finnish population. This is in agreement with two other studies reporting bi-allelic expression of STOX1 in the placenta and no association to pre-eclampsia (Berends et al. 2007, Iglesias-Platas et al. 2007). However, a study by Rigourd et al suggested that STOX1 should still be considered as a key factor in pre-eclampsia, since their study strongly suggest that STOX1 plays a major part in normal placental development (Rigourd et al. 2008). Furthermore, evidence for an increased STOX1 expression in early pre-eclamptic placentas has been shown (Founds et al. 2009). Altogether, there is no conclusive evidence either for or against STOX1 as a susceptibility gene, but much larger sample sets than previously used in any study are needed to understand its role in pre-eclampsia.

4.2 PAPER II - ASSOCIATION IN 9p21 CAD RISK REGION

The number of published GWAS has been growing rapidly since 2007 leading to the discovery and replication of many new disease loci. Interestingly, several GWAS have identified a region on chromosome 9p21 that is associated with both CAD and T2D.

Since there is a clear link between pre-eclampsia, CAD and T2D, and our linkage peak borders the GWA signals, we decided to genotype previously reported associated and tagging SNPs in our 15 families and the nation-wide case-control cohort. To our knowledge, we are the first to examine an association between this chromosomal region and pre-eclampsia.

Association in Finnish, but not in UK and Australian/New Zealand populations

Twenty-three markers were successfully genotyped and single marker association test revealed four markers with nominal association in the case-controls when GH women were excluded from the analysis (Table 7). We observed a trend towards association for the same four markers in the extended families. However, this association was very weak due to the small number of families available for the analysis. We included
genotypes for three (rs7044859, rs496892 and rs564398) out of the four SNPs showing nominal association from 260 additional Finnish controls to improve our power and this resulted in significant association for two SNPs, rs7044859 and rs496892 (corrected P = 0.02 and 0.04). The significant association was observed when women without any proteinuria were excluded from the analysis, even though the sample size was reduced by almost 15%. Although pre-eclampsia and GH are often seen as a continuum of the same disease, it is not clear whether they are entirely different diseases sharing a measurable part of their phenotype (hypertension) or whether some GH patients present an early or mild stage of pre-eclampsia. Some evidence suggests that GH and pre-eclampsia are different diseases with different pathophysiological mechanisms. Altogether, the evidence suggests that CAD and pre-eclampsia share a potential genetic risk factor on 9p21 that manifests as pre-eclampsia under the metabolic changes caused by pregnancy and is revealed as CAD morbidity in later life.

We tried to replicate our findings in 74 Aus/NZ families and UK case-control sample set. Twenty-two markers were successfully genotyped in the UK samples. The two markers showing best association in our case-controls were genotyped in the Aus/NZ families, but none of the genotyped markers showed signs of association in either of the two sample sets (Table 7). Additionally, a meta-analysis of data collected from all four sample sets for marker rs7865618 resulted in decreased effect size with pooled OR from 1.34 (1.03-1.73) in the Finnish sample sets to 1.10 (0.91-1.32), when including the UK and Aus/NZ cohorts. The UK cohort consisted of 230 cases and 200 controls. Given the effect size of the associated marker obtained from studying the Finnish cohorts, power calculations indicated that the sample size of the UK cohort only had 35% power to detect association of this particular marker with pre-eclampsia. Even in the situation with a marker with a higher relative risk (e.g. 1.6), we would be underpowered to detect association in the UK cohort alone. Only two markers were genotyped in the Aus/NZ families and no association was observed. However, we cannot rule out that SNPs located in other LD blocks in this region could be associated to pre-eclampsia in different populations. Therefore, two markers at this locus are not enough for rejecting it as a susceptibility region for pre-eclampsia in the Australian population.

Table 7. Genetic association analysis of chromosome 9p21 SNPs rs7044859, rs496892, rs564398 and rs7865618 in the data sets from Finland, UK and Australia/New Zealand.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Allele</th>
<th>x²</th>
<th>P-val (sim)</th>
<th>x²</th>
<th>P-val (sim)</th>
<th>x²</th>
<th>P-val (sim)</th>
<th>P-val</th>
<th>P-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7044859</td>
<td>A</td>
<td>7.33</td>
<td>0.007 (0.092)</td>
<td>7.91</td>
<td>0.005 (0.016)</td>
<td>0.10</td>
<td>0.749 (1)</td>
<td>0.052</td>
<td>0.377</td>
</tr>
<tr>
<td>rs496892</td>
<td>C</td>
<td>5.07</td>
<td>0.024 (0.290)</td>
<td>5.90</td>
<td>0.015 (0.044)</td>
<td>0.01</td>
<td>0.945 (1)</td>
<td>0.277</td>
<td>0.192</td>
</tr>
<tr>
<td>rs564398</td>
<td>T</td>
<td>6.72</td>
<td>0.010 (0.130)</td>
<td>4.72</td>
<td>0.030 (0.081)</td>
<td>1.08</td>
<td>0.298 (1)</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>rs7865618</td>
<td>A</td>
<td>8.10</td>
<td>0.004 (0.061)</td>
<td>0.02</td>
<td>0.888 (1)</td>
<td>0.061</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 260 additional Finnish controls
(sim) P-values after 10 000 permutations
(GH) Gestational hypertension
Genes in the associated region

The two associated SNPs in the Finnish cohort are both located in the same LD region previously known to harbour CAD associated SNPs (Samani et al. 2007, Schunkert et al. 2008). The region contains the cyclin-dependent kinase inhibitors 2A/2B genes (CDKN2A/2B) and a antisense noncoding RNA in the INK 4 locus gene (ANRIL), but none of the genes has previously been implicated in the pathogenesis of pre-eclampsia. Interestingly, the protein p16\(^{\text{INK4a}}\) encoded by CDKN2A has been shown to inhibit smooth muscle cell proliferation and is therefore an interesting functional candidate since inadequate transformation and replacement of the smooth muscle cells of spiral arteries in the placental bed has been considered to be a major pathophysiological event in pre-eclampsia. It has recently been shown that ANRIL is expressed in vascular endothelial cells, monocyte derived macrophages, and coronary smooth muscle cells, all of which could play a role in the pathophysiology (Broadbent et al. 2008). Additionally, an investigation of the functional significance of conserved sequences within the 9p21 risk locus for CAD demonstrated an effect of genetic variation on ANRIL expression (Jarinova et al. 2009). It is plausible that ANRIL could regulate the two genes CDKN2A/2B through RNA interference, gene co-suppression, gene silencing, chromatin structure maintenance, imprinting, or DNA demethylation. Our two associated SNPs are not in strong LD with markers showing an effect on ANRIL expression and the corresponding conserved sequence with enhancer activity. However, given the structural similarity to other long ncRNA genes, ANRIL is likely to rely on multiple transcriptional and splicing regulators (Ng et al. 2007, Storz et al. 2005) and other sequences within the 9p21 locus may modulate activity of other ANRIL splicing enhancer(s).

Another plausible explanation for the association within this region is that a functional variant could have a potential role in transcriptional control of CDKN2B, without any involvement of ANRIL. It has been reported that our best associated SNP, located in the putative promoter region of CDNK2B, could alter the binding of the transcription factor hepatic nuclear factor (HNF) 1\(\alpha\), known to regulate the expression of a number of liver genes and may play a key role in linking metabolic and inflammatory pathways underlying the pathogenesis of CAD (Armendariz and Krauss 2009, Ruiz-Llorente et al. 2007). This could be the variant responsible for the observed association, but without further analysis we cannot say that a putative causal variants at this locus has been identified. This will require follow-up studies involving deep re-sequencing in preferably extreme phenotypes (e.g. severe cases) in the Finnish population to define the full allele spectrum and uncover functional variants. Finally, our gene expression analysis did not detect any differences between cases and controls at the 9p21 locus using the microarray data.
4.3 PAPER III, IV - MAPPING AND CHARACTERIZATION OF 2p25 LINKAGE REGION

Our previously identified susceptibility locus on chromosome 2p25 directly overlaps with a risk locus for general hypertension (Angius et al. 2002, Zhu et al. 2001, Laivuori et al. 2003). The linkage region contains several plausible susceptibility genes based on their functional properties. In Paper III, we first focused on ROCK2, which has been implicated in general hypertension and is located in the region of overlap between pre-eclampsia and hypertension linkage peak (see Figure 6B). In Paper IV, we present our efforts to narrow down the linkage region at the 2p25 locus and evaluate genes within it by association test, re-sequencing and gene expression analysis.

Significant linkage, but no association in families
Genotyping of eight new microsatellites between markers D2S398 and D2S305 (28cM-42cM) was done at approximately 1 cM intervals (Paper IV) (Figure 6A), and the region was narrowed down to 1.4 M b with the highest peak showing significant linkage (NPL score 4.09, P= 0.00036) using general diagnostic criteria (women with GH but without proteinuria were considered as cases). We attempted to refine the 1.4 M b region further by genotyping more SNPs in our 15 families, but ended up with several clearly separated potential hits showing weak association signals. However, we may have missed associated variants by chance since our first stages of fine-mapping were inefficient due to absence of knowledge of tagging SNP and allele frequency information. Even after it became possible to select reasonably common tagging SNPs to cover variation within the region of interest there were gaps in our genotyping. Preferably, the whole region of 1.4 M b and corresponding LD blocks should have been covered more densely with tagging SNP to increase the power of the study.

No significant evidence for the genes investigated
We decided to focus on the genes in the 1.4 M b region; E2F6, GREB1, LPIN1 and TRIB2 (Figure 6B). The four genes, and ROCK2 investigated in Paper III, were tested for association by genotyping tagging SNPs in our case-control sample set. Single marker and haplotype analysis did not reach statistical significance. However, markers in TRIB2 and GREB1 were nominally associated (P<0.04) and a common haplotype in TRIB2 (P=0.01). SNPs with MAF <5% were not considered due to lack of power to detect association. However, rare SNPs are more likely to be functionally deleterious than common SNPs and it would be important to cover them (Gorlov et al. 2008). Our association studies were gene-centric, however, intronic or intergenic variants could modulate disease risk by affecting gene expression pattern (tissue specificity, timing, expression level) and these are missed by our approach that focuses on the genes only. The weak association could be a reflection of a causal allele that is located somewhere else in the region and is not in complete LD with the associated SNPs. The nominally associated genes need to be further investigated in well powered sample sets and with comprehensive SNP typing over the gene and plausible regulatory regions.

Re-sequencing the genes could have helped us to identify rare variants that confer susceptibility to pre-eclampsia. Based on both functional properties and association results, the genes TRIB2, LPIN1 and ROCK2 were re-sequenced in pre-eclamptic
women from families contributing to the 2p25 linkage signal to improve our chances of identifying the most relevant functional variants and healthy controls. ROCK2 is a widely expressed gene, but appears to be most abundant in smooth muscle cells, suggesting a role in vascular contraction (Riento and Ridley 2003). Interestingly, expression levels of RhoA mRNA have been shown to be significantly higher in pre-eclamptic placentas compared to placentas from normotensive women (Friel et al. 2008). However, we were not able to detect any expression differences for any of the specific probes included in the microarray expression analysis between the case-control groups. This is in agreement with a previous study that evaluated ROCK2 expression levels in placental tissues using RT-PCR and did not detect alterations between pre-eclamptic women and controls (Friel et al. 2008). However, the lack of alteration in mRNA expression levels does not rule out the possibility of alterations in protein levels that have been reported using immunohistochemistry and Western blot for ROCK2 (Ark et al. 2005).

LPIN1 encodes for a phosphatidate phosphatase (PAP) enzyme required for glycerolipid biosynthesis (Carman and Han 2006). Several recent studies have confirmed the relationship between adipose tissue lipin-1 levels and insulin sensitivity in humans (Reue 2009). Interestingly, LPIN1 variants have previously shown association to pre-eclampsia risk factors such as BMI (Fawcett et al. 2008, Wiedmann et al. 2008), high insulin levels at resting metabolic rate (Loos et al. 2007) and the metabolic syndrome (Wiedmann et al. 2008). LPIN1 and ROCK2 were obvious candidates based on their involvement in lipid metabolism and general hypertension. TRIB2 on the other hand encodes a signaling protein that may regulate a number of physiological and developmental processes. However, the cellular and molecular basis for its action is still poorly characterized and it is therefore impossible to speculate about any plausible involvement in pre-eclampsia pathophysiology.

We identified several novel variants in ROCK2 by re-sequencing and selected five for genotyping in our case-control data set based on their potential functional role (non-synonymous coding, splice site, other functional SNPs and insertion/deletion polymorphisms), but none showed association. Re-sequencing TRIB2 and LPIN1 resulted in identification of several variants, but none co-segregated with pre-eclampsia in the families when they were further investigated.

Figure 6. Mapping stages, re-sequenced genes and peak-markers within the 2p25 linkage region (A) Non-parametric linkage (NPL) analysis in 15 pre-eclampsia families. The NPL results were obtained by increasing the marker density in the previously mapped candidate susceptibility loci for pre-eclampsia on chromosomes 2p25 to approximately 1 cm intervals by genotyping 8 new microsatellites between markers D2S398 and D2S305. (28cm-42cm) (gray horizontal box). Vertical dotted lines highlights the emplacement of peak observed for marker D2S262 (NPL=4.09) and the borders of the region subjected for SNP fine mapping (black horizontal box). (B) Enlarged ~ 1.8 Mb interval between D2S2278 and D2S2199. ROCK2, E2F6, GREB1, NTSR2, LPIN1 and TRIB2 are the protein coding genes located in the region. Rectangles highlights the peak marker D2S262 and the borders of the region subjected for SNP fine mapping (~1.4 Mb). 11 genotyped SNPs are marked on the x-axis. A thickened portion of the x-axis denotes the associated 3-marker haplotype region of ~290 kb. The arrow points out the peak marker rs1453489. Genes that were re-sequenced are colored in gray and genes tested for association in cases and controls are marked with asterisk. Connected arrows point out a susceptibility locus for hypertension (Angius et al. 2002). Marker D2S168, peak marker in genome-wide linkage studies of pre-eclampsia and hypertension (Laivuori et al. 2003, Zhu et al. 2001).
4.4 GENERAL ASPECTS

Replication analyses in pre-eclampsia have often yielded inconsistent results, however, in order for association to be valid, the results need to be replicated in independent sample sets to distinguish between false and true association. There is substantial phenotypic and most likely genetic heterogeneity within and across populations for pre-eclampsia. Thus, it is not surprising that different research groups have identified unique loci and associations, especially since there must exist a strong natural selection against predisposing variants for pre-eclampsia until very recently due to the documented increased mortality for both mother and infant in the absent of effective obstetric care. Utilizing a founder population, such as our Finnish families, is likely to reduce genetic heterogeneity and aid genetic studies. In paper II, we demonstrated some evidence for association for markers at 9p21, but were not able to replicate in the two other populations, which could be due to locus or allelic heterogeneity, especially in the Aus/NZ families where only two markers were genotyped.

Although multiple linkage regions have been identified in pre-eclampsia, the mapping of genes has been unsuccessful, with only a few promising candidates showing association in separate populations. Linkage is known to be more robust against heterogeneity than association and we see a convincing linkage signal at 2p25 in our extended families. However, when we attempt to fine-map the region and test single markers for association in paper III and IV, the signal becomes diluted and disappears. When investigating linkage region or more refined areas, gene-centric approaches are
very commonly used, but intergenic variants that could have an effect on the phenotype will be missed. To complicate things further, there is evidence that even within a causal locus, not all markers show association (Kotowski et al. 2006). Therefore, it is important to perform as exhaustive and comprehensive mapping of variants as possible when investigating candidate genes or loci. Weak associations may be a result of rare variants or causal alleles not in complete LD with the associated variant. Rare SNPs have generally been ignored in association studies because sample size needed to detect association to a rare variant has been too expensive to study. Also, common SNPs are more likely to be shared between populations enabling design of commercial genotyping arrays. However, rare SNPs are more likely to be functionally deleterious than common SNPs and it would be important to cover them (Gorlov et al. 2008).

Definition of the phenotype is crucial in genetic studies, and an accurate phenotype definition is necessary to increase the chance of finding risk genes and to minimize the genetic heterogeneity in the collection of samples. Misclassification can reduce the power to detect association and increase the probability of spurious findings as a result of biased classification or confounding of the phenotype (Ellsworth and Manolio 1999). In general, defining phenotypes for complex diseases is difficult due to issues of heterogeneity, gene-gene and gene-environment interactions. As a likely example of multiple biological pathways leading to disease in pre-eclampsia, defective vascular remodeling is not observed in all affected women (Egbor et al. 2006). It is also possible that familial and sporadic representation of pre-eclampsia have different genetic background. Oudejans et al suggest that pre-eclampsia comes in an early form with a clear familial component, associated with growth restricted babies and abnormal placentation, and a much more common, late form, that arises from interaction between a normal placenta and maternal predisposing factors (Oudejans et al. 2007). The complexity of pre-eclampsia with a heterogeneous genetic background and different manifestations may explain the failures to detect association in linkage regions. Unfortunately, there is no universal definition of pre-eclampsia and inclusion criteria are in some cases inconsistent. For our Finnish sample sets, we have used an internationally accepted definition and all medical records were reviewed by experienced clinicians. Additionally, the risk for population stratification should be small as the controls have been carefully matched according living area and pregnancy number.

Sample size is a clear determinant of a successful association study and it is apparent that majority of the published candidate gene studies and replication attempts in pre-eclampsia have been underpowered, at least for rare variants and variants with low genetic effects. Our case-control data set consists of around 350 cases and controls and gives us 80% power to detect association to common variants (MAF>0.3) with genetic effects of 1.6. If we subdivided the cases into pre-eclampsia/eclampsia, superimposed pre-eclampsia, GH and twin pregnancy to reduce the possibility of heterogeneity, we will end up with approximately 250 pre-eclampsia/eclampsia cases, which reduces the power even more. There is no evidence that pre-eclampsia would be driven by common markers with such strong effects. In fact, genome-wide association studies of many multifactorial diseases have highlighted that most associations are to markers with high MAF and low-medium OR (1.1-1.3) (Wray et al. 2008).
An alternative to association studies for investigating complex diseases is to examine gene expression profiles in diseased and healthy tissues. Pre-eclampsia only occurs in the presence of placenta and abnormal placentation is a known characteristic for the disease. Therefore, placental gene expression profiles may help to identify genes and pathways likely to be involved in the development of the disease. Previous expression study findings on genes and pathways are not consistent, possibly due to small sample size, different maternal ethnicity, different types of arrays and matching variability (Founds et al. 2008). We did not detect significant expression differences when examined our term placentas. However, we found that delivery method (vaginal, acute section, elective section) had a larger impact to differential expression than sample case-control status. True variation in expression levels may also have been missed due to small numbers of samples in each group studied. Moreover, the factors that trigger pre-eclampsia are likely to be expressed in early pregnancy and by examined gene expression later in gestation, as we have done, may only trace the mechanisms that have already occurred. Optimally, we would follow gene expression levels throughout gestation by taking placental biopsies longitudinally during different timepoints, but this is of course ethically unacceptable.
Despite the tremendous technological development in the field of genetics during the past few years, the numbers of identified genetic predisposing factors for pre-eclampsia have not increased as dramatically as for many other complex diseases, with only a few promising susceptibility genes identified so far. This is most likely explained by the fact that there is no GWAS performed in pre-eclampsia so far compared to many other complex diseases. Therefore, the genetics of pre-eclampsia is still an extremely important field to explore and a GWAS in a well characterized population is eagerly awaited, especially since the condition is both life threatening and resource demanding responsible for most of the maternal deaths worldwide. Right now the geneticists in the field are still investigating linkage regions identified almost a decade ago and most of the susceptibility loci are likely to be population-specific and will not allow identification of common predisposing factors with an effect in several populations.

In this thesis, one of our efforts was to investigate the first positionally cloned pre-eclampsia candidate gene STOX1. We were unable to find any evidence for STOX1 association to pre-eclampsia in the Finnish population. Two independent studies have also rejected the impact of STOX1 in pre-eclampsia and its suggested maternal inheritance. Since the link between pre-eclampsia and cardiovascular diseases is well established, it was intriguing to observe the association to the previously reported CAD locus at 9p21, which further strengthen the theory that the different disorders share risk factors that could manifest differently during a woman’s life. However, genetic evidence for this locus in other populations is still lacking and further investigations of genes CDKN2B, ANRIL and plausible regulatory elements are needed. We were able to refine the susceptibility region at 2p25 identified by a genome-wide linkage analysis, but found ourselves underpowered to implicate individual genes residing within this region. Taken together, all our studies highlight the challenges of gene mapping in complex diseases and pre-eclampsia is characterized by a substantial phenotypic and, most likely, genetic heterogeneity within and across populations.

To further elaborate the genetic contribution to pre-eclampsia in the future, carefully chosen definitions and focus on different pre-eclampsia phenotypes are most likely needed. Most important, the suggestion that pre-eclampsia could come in two forms, probably with different genetic background and underlying pathophysiology must be considered. Sample size clearly needs to be increased since we are currently underpowered to detect association to rare variants and variants with low genetic effects that could be expected for a complex disease such as pre-eclampsia. An ongoing project by the Finnish Genetics of Pre-eclampsia Consortium (FINNPEC) is recruiting pre-eclamptic women and healthy pregnant controls, as both trios and separate case-controls. The trio design will make it possible to study the paternal vs maternal contribution in the fetus and plausible effects, which have been implicated in pre-eclampsia. This project is aiming for over 3000 cases and will have sufficient power to detect association to common (MAF >2%) variants with medium effect size (OR > 1.3). A large nationwide clinical and genetic-epidemiological database will be
established, which will be used for studying molecular mechanisms and later consequences of pre-eclampsia and related traits. FINNPEC collaborates with UK Genetics of Pre-eclampsia Consortium (GOPEC), which includes families of over 1000 women whose pregnancies were affected by pre-eclampsia and collaborations as this will enable well-powered future GWAS and replication attempts in pre-eclampsia. GWAS represent an alternative approach to the previous hypothesis free genome-wide linkage studies, with the ability to detect common markers with moderate effect associated to disease and could lead to an increased understanding of diverse molecular pathways underlying pre-eclampsia. Thus, the causal variant are rarely revealed and the statistical power to detect small gene-gene and gene-environment interactions are limited. Furthermore, variants identified in one population are not always transferable to other and up until today, they have not been able to capture information about rare variants or a large fraction of the existing CNVs.

The new sequencing technologies with a dramatic increase in throughput will hopefully overcome many of the challenges of traditional mapping and cloning. Massive cost reduction in sequencing has the potential to accelerate the identification of predisposing genetic variants and genes in pre-eclampsia. One potential approach would be to sequence complete linkage peaks identified in pre-eclampsia, including our susceptibility loci 2p25 and the less studied 9p13, in extended families to discover causal variants in affected individuals. Susceptibility genes found this way may help discover signaling and metabolic pathways involved in pre-eclampsia risk and guide further studies to prevent or treat this disorder. Rare markers, that have been hard to identify with past technologies and limited sample sets, could be discovered by sequencing extreme phenotypes (e.g. individuals suffering from severe pre-eclampsia), are likely to have a stronger effect to susceptibility than common markers tested by genotyping, and may prove more successful in diagnostic screening. Hopefully, international collaborations in the future, in combination with new technologies will shed further light on genetic susceptibility factors and increase the understanding of the underlying pathophysiology in pre-eclampsia.
ACKNOWLEDGEMENTS

The work presented in this thesis has been carried out at the Department of Biosciences and Nutrition, Karolinska Institutet. I wish to express my sincere gratitude to all colleagues, collaborators, friends and family who in one way or another have contributed to this thesis.


Katja for being a fantastic co-supervisor! Thank you for all your help and advise through these years, and finding the time to work with this project although you were working fulltime elsewhere during the last years. I have very much appreciated your encouragement and support in everything. I also want to thank my second co-supervisor Hannele, for introducing me into this field of genetics. You have been a great support to me, and I really appreciated your company at the ISSHP meetings.

I am grateful to everyone that have contributed in the PE-project. Thanks to Erja for guidance through my exam-work, introducing me to this project and for helping me out in the lab during my first time in Novum. Always with a smile on your face 😉. Many thanks to Hong for the initial statistical help in the project and Marco for introducing med to statistical analysis. I also want to acknowledge Hanna N for genotyping assistance.

I would like to express my thanks to all co-authors for excellent collaborations and contributions to the papers in this thesis. Vesa and Leena at the Finnish Red Cross for making our research possible by providing us with samples. I also want to thank our collaborators Eric and Matt in the US and Linda M in UK for letting us investigate our findings in your samples and for always being very helpful. Tack till Michael och Magnus för ett lärorikt och intressant Dystoci-samarbete.

Jag vill passa på att tacka alla på MAF för genotypning och goda råd genom åren: Kristina, Linda, Astrid, Ville, Päivi, Aida, Malin, Lotta, Tiina, Cissi och Eva.

Many thanks to the past and present members of the JKE-group: Juha, Katja, Hannele, Myriam, Isabel, Ulf, Kristiina, Sara, Virpi, Cilla, Per, Marie, Ingegerd, Heidi, Gustaf, Hans, Christina, Liselotte, Hong, Marco, Tiina S, Mauro, Lovisa, Linda, Astrid, Rezin, Ville, Malin, Aida, Eva, Kristina, Päivi, Erik, Tiina, Lotta, Francesca A, Cissi, Rasko, Francesca B, Erja, Thomas, Jingwen, Jun and Pu. TMK-tjejerna: Heidi, Sara, Christina, Kristiina, Linda, Astrid och Anna. Utsökt mat, vin och ”lite” skvaller... kan inte bli så mycket bättre 😊 Anna, tack för allt stöd under dessa år vi följts åt som doktorander och inte minst alla trevliga fikastunder.
Biomed-vännen, ett stort tack till er alla för att ni funnits där genom åren. Mycket roligt har vi upplevt tillsammans. Allt tog sin början i Scheele-labb... Johanna, Susanne, Ying, Maria L och Andreas för alla härliga resor, fester, trevliga fikastunder och mycket mer 😊 Så tacksam att jag har er som vänner!


Tack till alla mina släktingar i Södertälje, Huddinge och närliggande trakter för tervliga sammankomster genom åren.

Slutligen vill jag tacka mina fantastiska föräldrar för all er omtanke, samt ert stöd och uppmuntran under alla år. Ni ska veta hur mycket det har betytt för mig!!!
7 REFERENCES


Biochim Biophys Acta 1792, 163-172.

Ellsworth DL and M anolio TA (1999) The emerging importance of genetics in epidemiologic
research. II. Issues in study design and gene mapping. Ann Epidemiol 9, 75-90.

Gynecol 194, 322-331.

Emanuel BS and Saitta SC (2007) From microscopes to microarrays: dissecting recurrent


Esplin MS, Fausett MB, Fraser A, Kerber R et al. (2001) Paternal and maternal components of

activator inhibitor 1 gene polymorphisms and preeclampsia. Gynecol Obstet Invest 56,
17-22.

Faisal F, Rompapanen EL, Hiltunen M, Helisalmi S et al. (2003) Polymorphism in the
interleukin 1 receptor antagonist gene in women with preeclampsia. J Reprod Immunol
60, 61-70.

Fawcett KA, Grimes Y, Loos RJ, Wheeler E et al. (2008) Evaluating the role of LPIN1
variation in insulin resistance, body weight, and human lipodystrophy in U.K.
Populations. Diabetes 57, 2527-2533.

Fitzpatrick E, Goring HH, Liu H, Borg A et al. (2004) Fine mapping and SNP analysis of
positional candidates at the preeclampsia susceptibility locus (PREG1) on chromosome
2. Hum Biol 76, 849-862.

Fitzpatrick E, Johnson MP, Dyer TD, Forrest S et al. (2009) Genetic association of the activin
A receptor gene (ACVR2A) and pre-eclampsia. Mol Hum Reprod 15, 195-204.


Founds SA, Dorman JS and Conley YP (2008) Microarray technology applied to the complex

expression in first trimester placentas of women destined to develop preeclampsia.

Frazer KA, Ballinger DG, Cox DR, Hind DA et al. (2007) A second generation human

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

Frias A E, Jr. and Belfort MA (2003) Post M appie: how should we be managing severe

Friel AM, Hyne PG, Sexton DJ, Smith TJ et al. (2008) Expression levels of mRNA for Rho
A/Rho kinase and its role in isoprostane-induced vasoconstriction of human placental
and maternal vessels. Reprod Sci 15, 179-188.

in the human genome. Science 296, 2225-2229.


outcome of preeclampsia with an onset before 24 weeks' gestation. A udit in a tertiary


Iglesias-Platas I, Monk D, Jebbink J, Buimer M et al. (2007) STOX1 is not imprinted and is not likely to be involved in preeclampsia. Nat Genet 39, 279-280; author reply 280-271.


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


Wellcome Trust Case Control Consortium (WTCCC) (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-678.


