Year: 2015

Title: **Preeclampsia: Specific genetic risk factors and shared predisposition with cardiovascular disease**


Author: **Liv Cecilie Vestrheim Thomsen**

Print: AIT OSLO AS / University of Bergen

# SCIENTIFIC ENVIRONMENT

This PhD project has been performed at the Department of Clinical Science at the University of Bergen, the Department of Obstetrics and Gynecology at Haukeland University Hospital in Bergen, the Department of Cancer Research and Molecular Medicine and the Centre of Molecular Inflammation Research (CEMIR) at the Norwegian University of Science and Technology (NTNU) in Trondheim, and at the Centre for Genetic Origins of Health and Disease at the University of Western Australia in Perth. Professor Line Bjørge has been my main supervisor and Senior Researcher Ann-Charlotte Iversen and Professors Rigmor Austgulen, Ottar Nygård and Eric Moses have been my co-supervisors.

Professor Line Bjørge, Medical Director of the Gynaecologic Cancer Unit at the Department of Obstetrics and Gynaecology at Haukeland University Hospital in Bergen, has experience with management of translational research projects focusing on both targeted therapy and genetic profiling of complex disorders. She is also part the Bergen Gynecologic Cancer Research Group at the University of Bergen, which for several years has created a solid foundation in Bergen for translation research of gynaecological cancer.

The Research Group of Inflammation and Genetics in Pregnancy at the CEMIR is headed by Senior Researcher Ann-Charlotte Iversen and Professor MD Rigmor Austgulen. The group has been working for many years to obtain insight in mechanisms inducing preeclampsia and cardiovascular disease in women. The research is concentrated on the association of clinical manifestations and basic genetic and inflammatory mechanisms. Rigmor Austgulen has performed extensive pregnancy-related genetic analyses based on the HUNT Study for more than a decade and Ann-Charlotte Iversen is partner in the world´s largest genome-wide association study (GWAS) on preeclampsia; the EU FP7 project InterPregGen. The group has established molecular inflammation studies, functional risk gene discovery, and genetic population studies in preeclampsia, and are skilled in molecular methodology, advanced imaging and biobanking.

Professor Ottar Nygård, senior consultant at the Cardiologic Department at Haukeland University Hospital in Bergen and co-head of the Pharmacology/Cardiology research group at the Department of Clinical Science at the University of Bergen, has through the last decade build up a large research portfolio with cardiovascular diseases as the main focus. Ottar Nygård has been principal in the establishment of large clinical cohorts and conduction of clinical trials. His translational research approach, involving epidemiologic methods, molecular and genetic analyses, has revealed biological mechanisms that are being used for intervention.

Professor and Director of the Centre for Genetic Origins of Health and Disease Eric K. Moses is leading a multidisciplinary team working towards identification of the genetic, epigenetic and environmental contributions to risk of preeclampsia and other common human diseases. His research involves contemporary molecular genetics approaches, including genome-wide association mapping in population based cohorts of unrelated individuals and linkage, whole exome and whole genome DNA sequencing-based strategies in families. One of the focus points of his research team is to develop statistical genetics methods for use in disease gene mapping studies.

The project initially started as part of a joint project-cluster between Line Bjørge, Rigmor Austgulen and Ann-Charlotte Iversen, focusing on the pathophysiologic mechanisms of preeclampsia and has developed to involve cardiovascular disease and Ottar Nygård with his expertise within the field. As the project has evolved, Eric Moses, a long term collaborator of Rigmor Austulen, and his team have participated prominently in the genetic studies.

The main funding sources have been the Norwegian Women's Public Health Association, the Norwegian ExtraFoundation for Health and Rehabilitation, the Centres of Excellence funding scheme (CEMIR) at the NTNU, the Research Council of Norway through the programs Functional Genomics (FUGE) and FRIMEDBIO, the Unger-Vetlesen Medical Fund and the Norwegian Heart and Lung Patient Organization (LHL).

# CONTENTS


| | Page |
|---|---|
| Abstract | 12 |
| List of abbreviations | 16 |
| List of publications | 19 |
| 1. Introduction | 21 |
| 1.1 Preeclampsia | 21 |
| 1.1.1 Definitions of preeclampsia | 22 |
| 1.1.2 Risk factors | 27 |
| 1.1.3 Etiology/pathogenesis | 33 |
| 1.1.3.1 The normal pregnancy | 33 |
| 1.1.3.2 The preeclamptic pregnancy | 36 |
| 1.1.4 Disease prediction and prevention | 41 |
| 1.1.5 Disease management | 42 |
| 1.1.6 Future health of women and their offspring after PE | 43 |
| 1.2 Cardiovascular disease | 45 |
| 1.2.1 Epidemiology | 45 |
| 1.2.2 Definitions and phenotypic subgroups | 47 |
| 1.2.3 Risk factors for atherothrombotic cardiovascular diseases (aCVD) | 51 |
| 1.2.4 The etiology of aCVD | 54 |
| 1.2.5 Screening, prevention and treatment for aCVD | 56 |
| 1.3 Genetics | 58 |
| 1.3.1 The basics on human genetics | 59 |

1.3.2 The genetic variation of populations 63

1.3.2.1 Evolutionary sources of genetic variation 63

1.3.2.2 Polymorphisms 64

1.3.3 Genetics in Medicine 67

1.3.4 Epidemiologic observational study designs and their application in genetic studies 68

1.3.5 Genetic approaches to complex diseases 71

1.3.6 Genetic databases and bioinformatics 74

1.3.7 Phenotyping 75

1.3.8 Ethical considerations 76

1.4 Biobanking 77

1.5 The relationship between PE and aCVD

2. Aims of the thesis 81

2.1 Overall objective 82

2.2 Specific objectives 82

3. Materials and Methods 84

3.1 Materials 84

3.1.1 The Medical Birth Registry of Norway 84

3.1.2 The Preeclampsia Family Biobank 85

3.1.3 The HUNT Study 88

3.1.4 The genetic databases utilized 90

3.1.5 Ethical approvals 92

3.2 Methods 93

3.2.1 Diagnostic criteria 93

3.2.2 Phenotypes 94

3.2.3 Heritability 96

3.2.4 Candidate SPN-array 96

3.2.5 Statistical analyses 97

4. Results 101

4.1 Study I 101

4.2 Study II 102

4.3 Study III 103

5. Discussion 105

5.1 Diagnostic validity 105

5.2 Phenotyping 107

5.3 Heritability 109

5.4 Genetic analysis 110

6. Conclusion 115

7. Future perspectives 117

8. References 119

# ABSTRACT

**Introduction.** Preeclampsia (PE) is a serious multifactorial gestational disease affecting 2-8% of pregnancies world-wide, and preeclamptic women demonstrate increased risk of developing cardiovascular disease (CVD) after affected pregnancies. Despite the uncertainty surrounding the pathophysiology of PE, a genetic component of the etiology has been demonstrated. The combination of life-style associated and genetic risk factors with diverging disease presentations complicates research on the disorder. Careful classification of individuals according to disease phenotype is important to identify how risk factors contribute to different presentations of PE, and to determine if certain PE phenotypes lead to specific phenotypic presentations of CVD. Diagnostic validity is important to ensure that the diagnosis registered and used in research identifies only the individuals with the phenotype of interest. Before we know how and why PE develops and preeclamptic women contain a higher risk of CVD, it is challenging to implement effective methods for screening, identification of high-risk individuals for PE, find effective treatment, and to identify women at highest risk for later CVD. Application of existing knowledge on genetic risk factors for CVD could help us understand PE, and genetic approaches will be important for discovery of links between PE and CVD.

**Objectives.** The main objective for this thesis has been to identify pathophysiologic mechanisms for PE and overlap between the etiology of PE and CVD by illuminating disease specific and shared genetic risk factors of the disorders. We first aimed to validate the PE diagnosis in the Medical Birth Registry of Norway (MBRN) and examine how use of different criteria influence the diagnostic validity. Next, we aimed to establish a cohort of preeclamptic families for phenotypic and genetic heritability studies on PE and related conditions, based on valid preeclampsia diagnoses and careful phenotyping. The final objective was to identify genetic risk factors for PE and possibly detect a genetic link between development of PE and CVD in two cohorts with known diagnostic validity of the PE diagnosis; a preeclampsia family cohort and a PE case-control cohort from the HUNT Study (HUNT).

**Material and methods.** In Study I we examined the registration of the diagnosis of PE in the MBRN according to two sets of criteria. Two cohorts were used, both based on women with a registered diagnosis of PE in the MBRN. One cohort included 992 pregnancies of women who had participated in the second survey of the HUNT Study and in addition had the diagnosis PE, and the other cohort, including 1201 pregnancies in 1003 women, was based on pairs of first-degree female relatives where both had given birth in at least one preeclamptic pregnancy at one of five participating hospitals according to registrations in the MBRN. A broad PE criteria was applied to the women selected from the HUNT Study, and a more restricted PE criteria for the related women. In Study II we established a PE Family Biobank based on families of the women from Study I where both individuals had a valid PE diagnosis according to the stricter set of criteria. A thourough phenotyping of the PE presentations and conditions related to PE development, such as CVD, diabetes mellitus and chronic inflammatory diseases, was performed. We further applied a general pedigree variance-component method in the Sequential Oligogenic Linkage Analysis Routines (SOLAR) software to estimate the genetic heritability of each defined phenotype before we cross-examined in pairs the heritable traits for phenotypic correlations. For Study III we performed a candidate single nucleotide polymorphism (SNP) array including SNPs associated with hypertension or inflammation in genome-wide association studies (GWAS). The SNPs were identified through two curated databases, the NHGRI genome-wide association study (GWAS) catalog and the Phenotype-Genotype Integrator. The selected SNPs were examined in samples from the HUNT PE case-control cohort and the PE Family Biobank. Of 122 successfully genotyped SNPs 119 passed quality control and were included in the analyses.

**Results.** The diagnostic validity of PE diagnosis in the MBRN was found to be 88.3% when examined against the criteria given by the MBRN for diagnosis registration (Study I). The positive predictive value (PPV) of the diagnosis decreased to 63.6% when the stricter criteria was applied (Study I). The PE Family Biobank was established and classified according to extensive phenotypic criteria (Study II). Heritable phenotypes in the PE Family Biobank included PE, being born in a

preeclamptic pregnancy, giving birth to neonates small for gestational age (SGA), presence and severity of CVD, chronic hypertension, diabetes mellitus, and pulmonary disease (Study II). Of the heritable phenotypes, several had pairwise phenotypic overlap, such as PE and SGA, and chronic hypertension and diabetes mellitus type 2 (Study II). In the HUNT cohort the SNP rs17367504 in the methylenetetrahydrofolate reductase (MTHFR) gene was identified as a protective genetic factor for PE (Study III). In the PE Family Biobank cohort the SNP rs1327235 near the jagged1 (JAG1) gene demonstrated a borderline significant fetal protective effect on development of PE.

**Conclusions.** When examined against the criteria given by the MBRN for diagnosis registration, the diagnostic validity of PE in the MBRN is high (Study I). The PE Family Biobank has been established and extensively phenotyped (Study II). Several of the examined phenotypes were heritable in the PE Family Biobank, confirming a genetic factor for PE, and for giving birth to SGA neonates, chronic hypertension, CVD, diabetes mellitus and pulmonary disease (Study II). The identified phenotypic correlations could indicate that partially shared genetic mechanisms influence the development of the overlapping conditions (Study II). One novel genetic locus in the MTHFR gene with established protective effect on chronic hypertension has been identified to have a protective effect on development of PE in the HUNT cohort (Study III). One novel genetic locus near the JAG1 gene, not previously associated with PE but as a genetic risk factor for chronic hypertension, is suggestive as having a protective effect for PE, but the finding is inconsistent.

# LIST OF ABBREVIATIONS

| | |
|---|---|
| aCVD | atherothrombotic cardiovascular diseases |
| A | Adenine |
| Biomarker | Biologic marker |
| BMI | body mass index |
| C | Cytosine |
| CI 95% | 95% confidence interval |
| COPD | chronic obstructive pulmonary disease |
| CVD | Cardiovascular disease |
| dbGaP | Database of Genotype and Phenotype |
| DNA | Deoxyribonucleic acid |
| DVT | Deep venous thrombosis |
| G | Guanine |
| GWAS | genome-wide association study |
| H2r | heritability |
| HDL | High density lipoproteins |
| HELLP | organ failure with hemolysis, elevated liver enzymes and low platelets |
| HUNT1-3 | The HUNT Study 1, 2 and 3 |
| HWE | Hardy-Weinberg equilibrium |
| ISSHP | the International Society for the Study of Hypertension in Pregnanacy |
| IUGR | intrauterine growth retardation |
| LD | Linkage disequilibrium |

| | |
|---|---|
| LDL | Low density lipoproteins |
| MBRN | Medical Birth Registry of Norway |
| MTHFR | methylenetetrahydrofolate reductase |
| NHGRI | The National Human Genome Research Institute |
| NHGRI Catalog | NHGRI Catalog of Published Genome- Wide Association Studies |
| NTNU | the Norwegian University of Science and Technology |
| OR | Odds ratio |
| PE | preeclampsia |
| PE Family Biobank | The Norwegian Preeclampsia Family Biobank |
| PheGenI | Phenotype–Genotype Integrator |
| PPV | Positive predictive value |
| QC | quality control |
| RNA | Ribonucleic acid RNA |
| RR | relative risk |
| SGA | small for gestational age |
| SNP | Single nucleotide polymorphism |
| SOLAR | Sequential Oligogenic Linkage Analysis Routines |
| SPSS | Statistical Package for Social Sciences |
| STMP | syncytiotrophoblast microparticles |
| T | Thymine |
| teCVD | Thromboembolic CVD |
| U | Uracil |

# LIST OF PUBLICATIONS

I. Thomsen LCV, Klungsøyr K, Roten LT, Tappert C, Araya E, Bærheim G, Tollaksen K, Fenstad MH, Macsali F, Austgulen R, Bjørge L. Validity of the pre-eclampsia diagnosis in the Medical Birth Registry of Norway. Acta Obstet Gynecol Scand. 2013;92(943-950).

II. Thomsen LCV, Melton PE, Tollaksen K, Lyslo I, Roten LT, Odland LM, Strand KM, Nygård O, Sun C, Iversen A-C, Austgulen R, Moses EK, Bjørge L. Refined phenotyping identifies links between preeclampsia and related diseases in a Norwegian preeclampsia family cohort. Submitted December 2014.

III. Thomsen LCV, McCarthy N, Melton PE, Cadby G, Austgulen R, Moses EK, Bjørge L, Iversen A-C. MTHFR Gene Polymorphism is a Novel Genetic Link between Preeclampsia and the Cardiovascular Risk Factor Chronic Hypertension in the HUNT Study. Manuscript in preparation for submission

# 1. INTRODUCTION

## 1.1 Preeclampsia

Preeclampsia (PE) is a heterogeneous gestational disorder with disease presentations ranging from mild symptoms and barely affected women and / or fetuses to life-threatening maternal multi-organ failure and convulsions (eclampsia), intrauterine fetal growth retardation and death. PE is generally defined by development of hypertension combined with excretion of protein into the urine in the latter part of a pregnancy, and delivery of the child and placenta is still the only certain treatment option for this disease. Globally PE involves 2-8% of all pregnancies and it is the second most common direct cause of maternal deaths [1, 2]. In Norway the prevalence of PE is currently 3.6 % [3].

Women with a history of PE have an up to 8-fold increase in the risk for development of cardiovascular risk factors and dying from cardiovascular causes [4-6], and the overall risk for cardiovascular disease (CVD) following PE is doubled [7]. How PE is linked to later life CVD development is not well understood, but pathophysiologic mechanisms related to endothelial dysfunction and inflammation are important in both disorders. Despite epidemiologic and biologic indicators to the existence of common predisposing factors, it is still uncertain if the shared environmental and biologic risk factors are the sole reasons for a woman to develop CVD after experiencing PE. The pathophysiologic reactions seen in PE might in addition generate lasting biologic changes and influence how and when a woman develops CVD [8, 9].

Multifactorial diseases like PE result from multiple genetic changes (see also paragraph 1.3) combined with life-style and environmental influences. The heterogeneous presentation and complexity of interactions involved in development of

such disorders complicate any discovery of etiologic factors. Although the necessity of placental tissue presence for PE development has been established, as have several causative pathophysiologic mechanisms and risk factors involved, a large part of the etiology of PE is unknown. New discoveries are challenged by discrepancies in disease presentation and in diverging diagnostic criteria used for classification of PE. The complexity of the maternal-fetal interplay further complicates research on PE, as interactions between the mother and fetus influence how the disorder develops and how the disease develops in a particular pregnancy [10].

Before we know why and how women develop PE it will be difficult to discover and implement effective tools for screening, clinical risk stratification, prevention, diagnostics and prognosis, and to develop appropriate follow-up regimes and treatment for the disease(s).

### *1.1.1 Definitions of preeclampsia*

Reflecting the scarce knowledge on how and why this disease develops, PE is commonly known by obstetricians, midwifes and researchers as *the disease of theories*. It might just as well be named *the disease of definitions*.

To define and classify PE seems to have been considered difficult but important throughout the history of medicine, as those aiding the pregnant women always have needed tools for correctly predicting the outcome of a pregnancy and a birth. With the etiology unknown, PE is, and has been, defined by symptoms and signs rather than by markers of relevant pathophysiologic changes. After the discovery of blood pressure in 1773, and in 1881 a non-invasive way of measuring it [11], medical interest widened. Previously the sole focus was eclampsia. In the last decades the attention has shifted to also address the associated disease PE as well as the need for a common international understanding and classification of diseases [12]. Consequently, the different manifestations and expressions associated with a specific disease, its phenotypes (see also paragraph 1.6), are gradually included more extensively in the disease definitions.

As PE is present world-wide in pregnant women, a definition of the condition ought to be globally available for ready clinical use in rich, urban, rural, and poor areas, and encompass all women with increased risk for severe disease threatening the health of mother or child. Despite a substantial overlap between diagnostic criteria for PE, the symptoms and signs and measurable levels of pathologic changes used for diagnosis still differ according to national and international obstetric societies, advisory boards and advocacy groups (Table 1 in Study I).

Although no universal consensus has been established on how to define the different phenotypes of PE, all classifications include reproducible elevated maternal blood pressure developing in the latter half of pregnancy. Previously a generalized requirement for PE was that the gestational hypertension must be combined with proteinuria [13, 14]. Recently several guidelines where non-proteinuric PE is a phenotype have been published from Obstetric and Gynecological Societies, among these the Canadian and Australasian, and the International Society for the Study of Hypertension in Pregnancy (ISSHP). Proteinuria is one of several interchangeable diagnostic criteria, belonging to a list of adverse conditions and complications whereof at least one is required in addition to hypertension to give the diagnosis PE [15-17]. The guidelines developed in the different European countries and for the World Health Organization have maintained the combination of hypertension and proteinuria as a diagnostic necessity for PE [17-19]. Most classifications distinguish between transient gestational and chronic hypertension, using the 21st gestational week as cut-off. If a woman demonstrates hypertension before this time it is considered a chronic condition, while transient gestational increases in blood pressure develop after 20 weeks of pregnancy and return to normal ranges by 12 weeks after delivery. Women with chronic hypertension who acquire proteinuria in the second half of pregnancy are commonly diagnosed with "superimposed preeclampsia" [20, 21]. The Norwegian Association for Obstetricians and Gynecologists defines PE as persisting hypertension (blood pressure $\geq$ 140/90) and proteinuria developing after the twentieth gestational week (Table 1). Proteinuria is classified as 0.3 g/24 hours urine excretion or more, total protein/creatinine ratio above 0.3, or minimum two findings of at least +1 on a urine dip stix [21].

## Table 1 Current Norwegian definitions of preeclampsia (2014)

| **Diagnosis** | **Blood pressure**[a] | | **Protein excretion in urine** | | | **Time of diagnosis** |
|---|---|---|---|---|---|---|
| | Systolic | Diastolic | Dip-stix[a] | Urine excretion grams / 24 hours | Protein / creatinine ratio | |
| PE | ≥ 140 mmHg | ≥ 90 mmHg | ≥ +1 | > 0.3 | > 0.3 | Hypertension and proteinuria after gestational week 20 |
| Severe PE | ≥ 160 mmHg | ≥ 110 mmHg | | ≥ 3 g | | Hypertension and proteinuria after gestational week 20 |
| Early PE | ≥ 140 mmHg | ≥ 90 mmHg | ≥ +1 | > 0.3 | > 0.3 | Hypertension and proteinuria before gestational week 34 |
| Late PE | ≥ 140 mmHg | ≥ 90 mmHg | ≥ +1 | > 0.3 | > 0.3 | Hypertension and proteinuria after gestational week 34 |
| Super-imposed PE[b] | ≥ 140 mmHg | ≥ 90 mmHg | ≥ +1 | > 0.3 | > 0.3 | Hypertension before and addition of proteinuria after gestational week 20 |

Based on the 2014 guidelines of the Norwegian Association for Obstetricians and Gynecologists [21]

[a] Reproducible findings, measured on at least two separate occasions

[b] Preexisting chronic hypertension with gestational proteinuria

*PE* preeclampsia

## Table 2 Severe phenotypes of preeclampsia

| | **Severe preeclampsia**[a] | **HELLP**[b] | **Eclampsia**[b] |
|---|---|---|---|
| **Symptoms** | Generalized convulsions[c] | | Generalized convulsions[c] |
| | Epigastric pain | Epigastric pain | |
| | Dizziness and nausea | Nausea | |
| | Cerebral affection:<br>Severe headaches<br>Irritability<br>Visual disturbances<br>Hyperreflexia | | Cerebral affection:<br>Severe headaches<br>Irritability<br>Visual disturbances<br>Hyperreflexia |
| | Rapid development of subcutaneous edema | Rapid development of subcutaneous edema | |
| | Pulmonary edema | Pulmonary edema | |

Based on the 2014 guidelines of The Norwegian Association for Obstetricians and Gynecologists [21]

[a] Diagnosed when both the criteria for preeclampsia is fulfilled and at least one criteria listed is present.

[b] Considered both as part of a severe preeclampsia presentation and as separate disease entities

[c] Occurring during pregnancy, antepartum or postpartum

*HELLP* hemolysis, elevated lipids and low platelets; *LD* lactate dehydrogenase; *ASAT* aspartate aminotransferase; *ALAT* alanine aminotransferase

.

**Table 2 Severe phenotypes of preeclampsia cont.**

| | **Severe preeclampsia**[a] | **HELLP**[b] | **Eclampsia**[b] |
|---|---|---|---|
| **Signs** | Consistent blood pressure<br><br>≥ 160/110 mmHg | | |
| | Microangiopathic hemolytic anemia:<br><br>Decreasing serum haptoglobin<br><br>Increasing plasma LD levels | Hemolysis:<br><br>Serum haptoglobin < 0.2 g/L<br><br>Increased plasma bilirubin levels<br><br>Increased plasma LD levels | |
| | Quickly decreasing number of platelets | Reproducible low platelets < 100 x $10^9$/L | |
| | Liver affection:<br><br>Increased enzymes excreted to plasma from the liver: ASAT, ALAT and LD | Liver affection:<br><br>Increased enzymes excreted to plasma from the liver: ASAT, ALAT and LD | |
| | Oliguria (< 500 mL urine / 24 hours) | | |

Based on the 2014 guidelines of The Norwegian Association for Obstetricians and Gynecologists [21]

[a] Diagnosed when both the criteria for preeclampsia is fulfilled and at least one criteria listed is present.

[b] Considered both as part of a severe preeclampsia presentation and as separate disease entities

[c] Occurring during pregnancy, antepartum or postpartum

*HELLP* hemolysis, elevated lipids and low platelets; *LD* lactate dehydrogenase; *ASAT* aspartate aminotransferase; *ALAT* alanine aminotransferase

Most PE definitions incorporate severity subtypes, classified according to a range of signs and clinical symptoms (Table 1 and 2). Pregnant women who develop severe disease might contact health personnel acutely due to symptoms of severe disease manifestations like eclampsia, headache caused by hypertension, epigastric pain correlating with development of multi-organ failure, or fetal demise (Table 2). The milder presentations of this disease, on the other hand, such as moderate increases of

blood pressure, peripheral edema increasing rapidly, or decreasing fetal growth are often detected incidentally on regular antenatal controls.

It is important to keep in mind that all given cut-off values for pathologic measurements of blood pressures and proteinuria are arbitrary. As the etiology of the syndrome is obscure, so are the exact levels of blood pressure, proteinuria, thrombocytes and liver enzymes correlated with the important pathophysiologic changes causing disease. The gestational age at onset or delivery is used to differ between early and late development of PE with an arbitrary cut-off at gestational week 34. The differing severity definitions include to a variable extent early PE as well as development of headache, peripheral edema, visual disturbances, upper abdominal pain, oliguria, convulsions, pulmonary edema, eclampsia and the HELLP syndrome.

IUGR is a complicating factor of PE. As the diagnostic confirmation of IUGR depends on repetitive ultrasound measurements, "small for gestational age" (SGA) is often used as proxy for IUGR when measurements are lacking or are unavailable to researchers, as in most epidemiological studies. Neonates, whose weights are at the lower percentiles of the normal spectrum according to gestational week at birth, are said to be SGA. In 18-57% of preeclamptic pregnancies the neonate is SGA/IUGR [22, 23], but IUGR is not pathognomonic for PE, and the condition is not included in most definitions of the disease. IUGR develops as a separate entity as well as combined with other complications of pregnancy like diabetes mellitus, placenta previa, and use of some types of medication [24, 25].

### *1.1.2 Risk factors*

Epidemiologic research has established several risk factors for development of PE (Table 3). Ethnicity, family history of PE development, giving birth to a first child (primiparity), having had a previous pregnancy complicated by PE, and high age of the pregnant woman, all increase the likelihood of an affected pregnancy [26-29]. In pregnancies with multiple fetuses women also have increased chances for developing the syndrome [30, 31]. Development of PE is seen more commonly in pregnancies achieved by assisted reproductive technology [32, 33], and the risk of disease

increases further when the procedure involves donated oocytes [34]. Women with specific pre-existing diseases like chronic hypertension, diabetes mellitus or systemic lupus erythematosus (Table 3) [29, 35]. Other known environmental influences on the risk profile are obesity, diet and smoking (the latter is protective) [36-39].

Many genetic risk factors have been given a role in the pathophysiology of PE. The most genetic changes demonstrated are those associated with mechanisms of and biologic functions known to take part in PE development, such as inflammation, angiogenesis, hemostasis, lipid metabolism and the renin-angiotensin system[40]. Different genetic approaches have been utilized to map disease-associated genetic changes associated with PE (see also paragraph 1.3.5), mainly based on family cohorts and case-control cohorts. In the family cohorts linkage studies have been performed (Table 4), while case-control studies on PE mainly have been investigating candidate genes or SNPs. Few genome-wide association studies (GWAS) have been performed on PE. Only an Australian GWAS on PE our research group participated in, has a positive gene finding near the Inhibin beta B gene, but the finding was not replicable in independent cohorts [41]. In a meta-analysis from 2013 six genes were identified as associated with PE [40]. These identified regions represent replicated loci that have been included in many studies, introducing possible bias in a meta-analysis by excluding variants less commonly examined, independent of results. Table 5 presents a selection of the most commonly studied genes, for which associations with PE has been detected. How these genes and genetic changes affect disease development in PE needs to be studied further, possibly with focus on PE subtype and by a mixed molecular approach.

## Table 3 Risk factors for preeclampsia

| **Maternal biologic factors** | **Pregnancy-related / fetal factors** | **Factors related to life-style** |
|---|---|---|
| Maternal age >= 40 years | Primipara | Diabetes mellitus / Gestational diabetes |
| Renal disease | Multiple pregnancy | Chronic hypertension |
| Inflammatory diseases like systemic lupus erythematosus | Intrauterine growth retardation | BMI >26 |
| Heritable thrombophilias | Low PIGF and high sFIt1 concentrations in maternal blood | Lower socio-economic status |
| Antiphospholipid syndrome | Abnormal uterine artery Doppler velocimetry | Non-smoking |
| Family history of PE | In vitro fertilization<br>Higher if donated sperm or ovum | Excessive weight gain in pregnancy |
| Previous pregnancy complicated by PE or preterm delivery | Inter-pregnancy interval <2 years or >=10 years | |
| Ethnicity: Nordic, Black,<br>South Asian or Pacific<br>Island | New father | |
| Family history of early-onset cardiovascular disease | Molar pregnancy (hydatidiform mole) | Cocaine and methamphetamine use |
| | Infections during pregnancy (periodontitis, urinary tract infections) | |

Based on the 2014 guidelines of the Norwegian Association for Obstetricians and Gynecologists [21] and the 2013 guidelines of the Society of Obstetricians and Gynaecologists of Canada [42]. *BMI* body mass index, *PlGF* Placental Growth Factor, *sFlt1* Soluble fms-like tyrosine kinase-1, *PE* preeclampsia

**Table 4 Genetic associations with preeclampsia identified through linkage studies and genome-wide association studies**

| Gene/region | Study | Publication |
|---|---|---|
| Chromosome 2p25[a] | Family linkage | Laivuori, 2003 [43] |
| Chromosome 2p13[a] | Family linkage | Arngrimsson, 1999 [44] |
| Chromosome 2q | Family linkage | Lachmeijer, 2001 [45] |
| Chromosome 2p25 and 9p13 [a] | Family linkage | Laivuori, 2003 [43] |
| Chromosome 2p12 [a], between markersD2S112 and D2S151 | Family linkage | Moses, 2000 [46] |
| Chromosome 4q[a] | Family linkage | Harrison, 1997 [47] |
| Chromosome 5q[a] | Family linkage | Johnson, 2007 [48] |
| Chromosome 13q[a] | Family linkage | Johnson, 2007 [48] |
| 2q14, near inhibin, beta B gene[a] | GWAS | Johnson, 2012 [49] |
| Chromosome 2 near marker D2S168[b] | Family linkage | Majander, 2013 [50] |

Based on linkage studies and genome-wide association studies identified through search on PubMed

a Association to maternal preeclampsia development

b Association to fetal and maternal preeclampsia development

**Table 5 Genetic associations with preeclampsia identified through meta-analyses**

| Gene/region | SNP | Study | Publication |
|---|---|---|---|
| Factor V Leiden | rs6025 and rs6020 | Meta-analysis | Buurma, 2013 [40] (ref) |
| Factor V Leiden | G1691A | Meta-analysis | Wang, 2014 [51] |
| SERPINE1 | rs1799889 | Meta-analysis | Buurma, 2013 [40] |
| ACE | rs4646994 | Meta-analysis | Buurma, 2013 [40] |
| CTL4 | rs231775 | Meta-analysis | Buurma, 2013 [40] |
| LPL | rs268 | Meta-analysis | Buurma, 2013 [40] |
| eNOS | G894T | Meta-analysis | Qi, 2013 [52] |
| Prothrombin/ coagulation factor II (F2) | rs1799963 | Meta-analysis | Buurma, 2013 [40] |
| Prothrombin/ F2 | G20210A | Meta-analysis | Wang, 2014 [51] |
| VEGF | +936C/T, -634G/C | Meta-analysis | Cheng, 2013 [53] |
| MTHFR | C677T, G894T | Meta-analysis | Wang, 2013 [54] |
| IL-10 | -819C/T, -592C/A | Meta-analysis | Wang, 2014 [55] |

Based on meta-analyses identified through searches on PubMed

*MTHFR* methylenetetrahydrofolate reductase; *SERPINE1* serpin peptidase inhibitor/plasminogen activator inhibitor type 1; *ACE* angiotensin I converting enzyme; *eNOS* nitric oxide synthase; *CTLA4* cytotoxic T-lymphocyte-associated protein 4; *LPL* lipoprotein lipase; *APOE* apolipoprotein E; *VEGF* vascular endothelial crowth factor; *IL-10* interleukin 10

### *1.1.3 Etiology*

#### 1.1.3.1 The normal pregnancy

The normal pregnancy starts when an egg is fertilized by a spermatozoa and becomes a zygote in the distal part of the uterine tube. The cell number of the zygote rapidly increases until it reaches 64, and the cells start dividing into external cells

(trophoblasts) and internal cells that eventually develop into the embryo and its membranes. The trophoblasts then differentiate and transform the cell mass into a blastocyst, which enters the uterus. The blastocyst attaches to the uterine endometrium and is embedded in the uterine wall through the process called implantation.

The trophoblasts multiply and invade deeper while constituting a protective barrier between the embryo and uterine cells and the maternal blood. Columns of junctional trophoblasts help anchor the placenta to the uterus. At the boundary between myometrium and endometrium in the uterus, the radial arteries branch into basal arteries, which in the basalis part of the endometrium transform into highly coiled spiral arteries. During pregnancy the terminal arterial coils dilate [56] and trophoblasts invade the lumen of the vessels, including the myometrial parts, as well as the endometrial tissues forming the decidual layer.

The maternal uterine blood vessels are changed into a blood-filled network of sinuses without an endothelial cell lining in the spaces between a network of trophoblast-covered branches of vessels called chorionic villi. The villi closest to the maternal blood supply will during the progression of the pregnancy form the chorion, while the other villi and the maternal sinuses degenerate. The area of sinuses and chorionic villi constitutes what is called the intervillous space (Figure 1). In the intervillous space oxygen, nutrients and waste products are exchanged between the maternal and fetalcirculation through diffusion and osmosis.

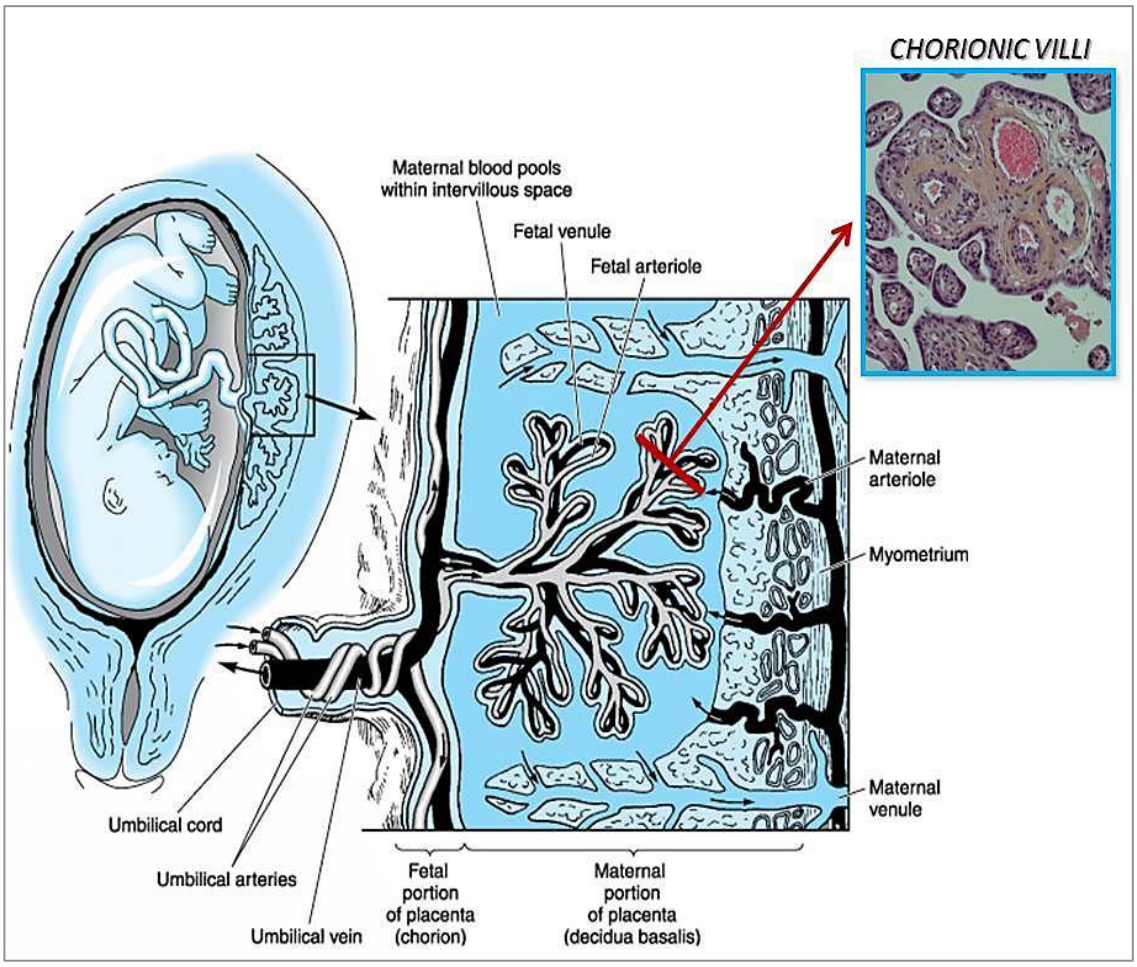


**Figure 1**. *Structure of the human term placenta. The fetal portion, the chorion, is covered by the amnion and is the insertion site of the umbilical cord. Fetal vessels ramify from the umbilical cord into the villi. The villous trees project from the chorion into the intervillous space. From the maternal portion, the decidua basalis, uterine spiral arteries provide blood supply rich in oxygen and nutrients to the placental tissue. The upper right figure illustrates the histological aspect of the chorionic villi, stained with hematoxylin and eosin. Adapted by G. B. Silva (NTNU, Trondheim, Norway) from Morgan et al [57].*

The trophoblasts of the chorionic villi that are directly in contact with the maternal blood in the sinuses, will fuse into a continuous layer of larger multinucleated cells: the syncytiotrophoblasts (Figure 2a). In addition to creating a maternal-fetal blood barrier, the syncytiotrophoblasts are a source of placental hormones, cytokines and enzymes, which are released to the extracellular space or shedded through

immunoactive microparticles (syncytiotrophoblast microparticles (STBM)) with impact on local inflammatory functions and maternal systemic responses [58-60].

As the fetus and placenta consist of maternal and paternal genetic material it can be considered as a semi-allograft to the maternal organism. Inflammation is an immune response to danger signals from infections or damaged tissue, aiming to protect the body by eliminating the microbe and/or repairing the damage. During a pregnancy the maternal biologic system balances enhanced pro-inflammatory mechanisms that protect the woman and her fetus against harmful infections and tissue damage against the immunosuppression necessary to avoid rejection of the semi-foreign fetus. Physiologic immune processes are central for normal implantation and placentation, partly by facilitating trophoblast infiltration and the necessary changes in the spiral arteries. Release of pro-inflammatory cell signaling substances (such as cytokines and chemokines) and increased levels of inflammation both locally at the maternal-fetal interface (decidua and placenta) and generalized maternally, are important for the normal development of a pregnancy. The physiologic low-grade systemic inflammation of gestation results in release of angiogenic factors, oxidative stress and adaptive metabolic changes like hyperlipidemia and increased insulin resistance [61].

#### 1.1.3.2 The preeclamptic pregnancy

Currently the placenta is recognized as the central factor for the development of PE (ref). In PE the changes in spiral arteries, which in normal pregnancies include the myometrial part of the arteries, do not extend longer than to the border between myometrium and decidua (Figure 2b) [62, 63]. This less extensive modulation of the arteries prohibits the physiologic velocity reduction in blood delivered into the intervillous space, leading to destruction of chorionic villi and increased shedding of STBM into the maternal circulation. The result, decreased time for placental uptake of oxygen, induces aggregation of hyperoxic blood in the intervillous space [24]. The ischemia and reperfusion of placental tissue and the increased release of reactive oxygen species [64], STBM and other pro-inflammatory substances augment both local and generalized inflammatory responses in the placenta and the pregnant woman.

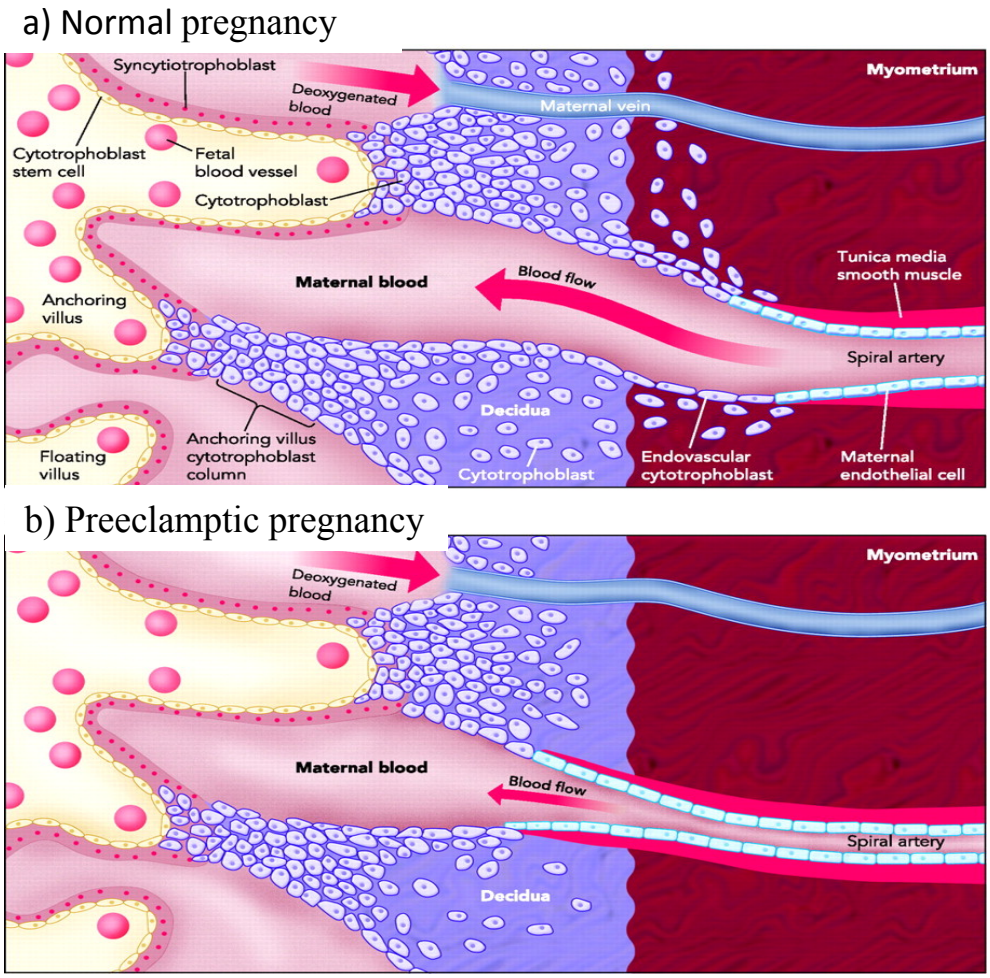


**Figure 2.** *The maternal - fetal interface with spiral arteries and veins during a) normal pregnancy and b) preeclamptic pregnancy. From Wang A et al. [65]*

A model of disease development, containing two main stages, is commonly presented (modified in Figure 3) [66]:

*Stage 1:* The abnormal placentation stage causing endothelial dysfunction and abnormal placental vascularization.

*Stage 2*: The inflammatory stage defined by dysfunctional maternal inflammatory responses.

*Stage 1*. The first stage is often named the fetal stage of preeclampsia as at this stage there are no maternal symptoms. The pathological changes of the placental vascularization, where trophoblast invasion of the spiral arteries is shallow or absent, is often referred to as poor placentation, and is a process linked repeatedly both to early onset of PE and to IUGR [66, 24].

Poor placentation can cause high levels of tissue hypoxia followed by ischemia and tissue reperfusion, thereby inducing placental and generalized oxidative stress and endothelial dysfunction followed by activation of platelets and the coagulation system and production of lipid peroxidases. The increased levels of lipid peroxidases in placental cell membranes give rise to a process called acute atherosis. The acute atherosis develop in the vessel wall of the decidual spiral arteries and is a process involving uptake of lipids to the wessel wall, infiltration of macrophages developing into foam cells, perivascular infiltration of leukocytes and vascular necrosis, and is comparable to the process causing atherosclerosis in cardiovascular disease patients (Figure 4).

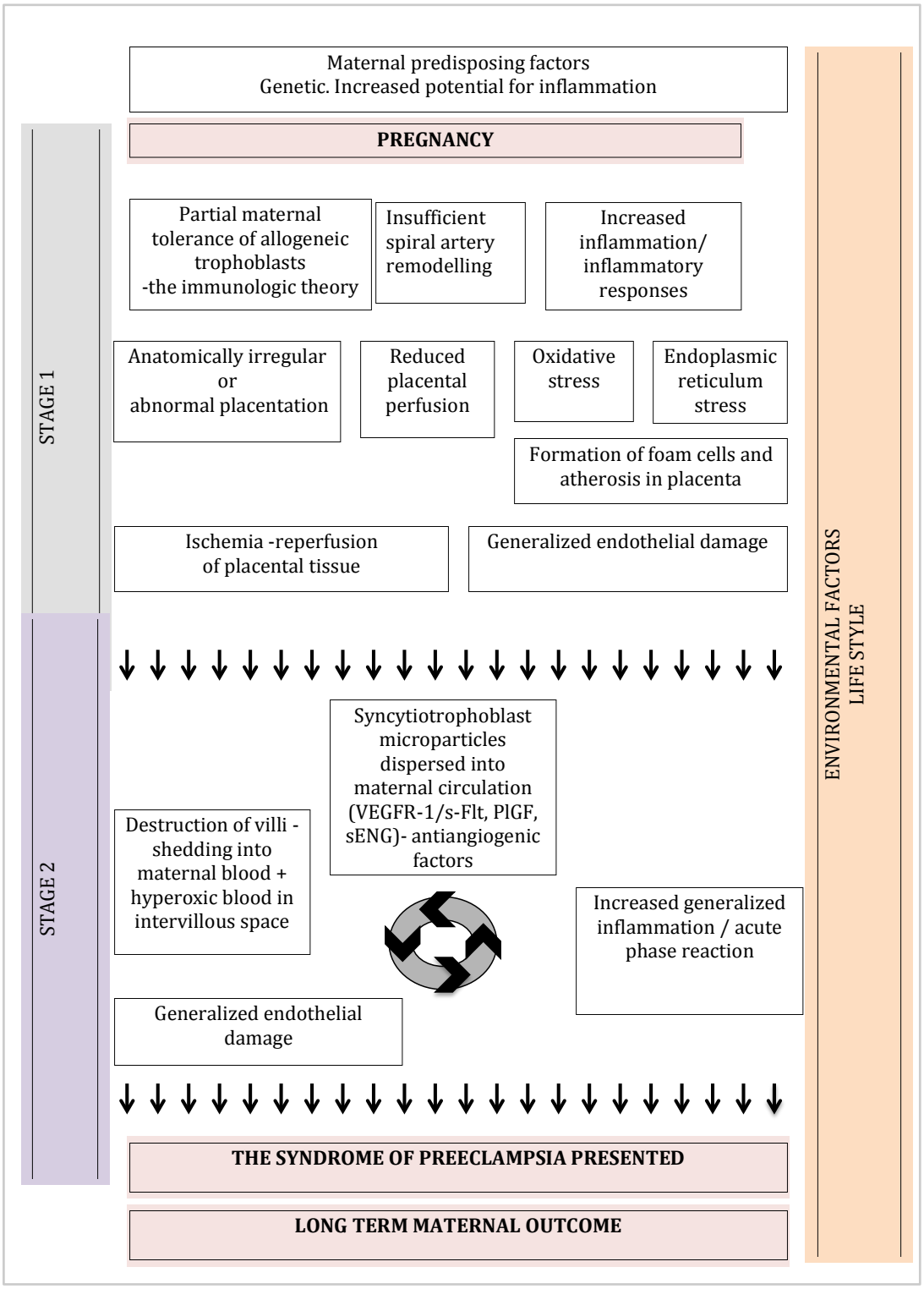


**Figure 3.** *Diagram presenting the stages and interacting factors in preeclampsia development. Modified from Redman&Sargent 2003 [66]*

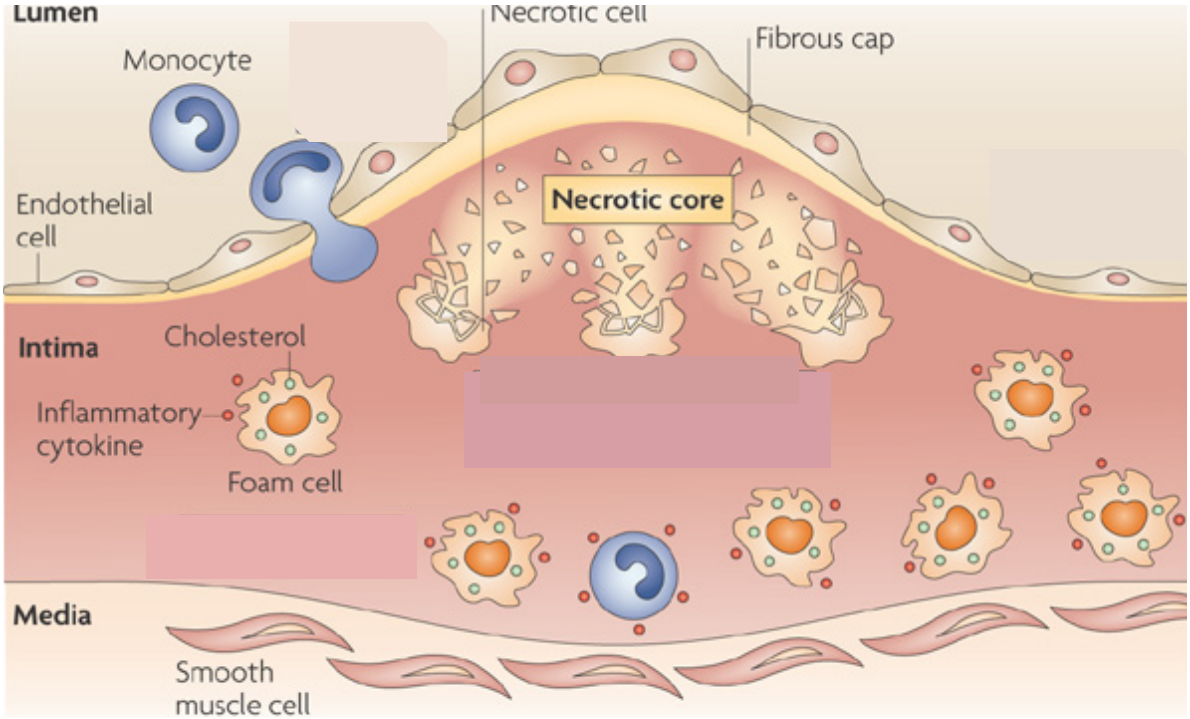


**Figure 4.** *Modified from I.Tabas [67]. Illustration of vessel wall changes in atherosclerosis. Inflammatory cells, including lipid-laden macrophage foam cells, accumulate in the intima. Apoptotic macrophages undergo secondary necrosis. This process contributes to the formation of the necrotic core, which promotes thinning of the fibrous cap and predispose to plaque disruption.*

*Stage 2.* The second stage is sometimes called the maternal stage. The changes resulting from stage 1 disturb the precious physiological balance between the pro-inflammatory and immunosuppressive mechanisms of pregnancy [24, 66]. To establish and continue a normal pregnancy immunologic changes and systemic inflammatory responses of leukocytosis, changes in endothelial cells and endothelial function and enhanced coagulation are required. In women affected by PE all these responses escalate. Further, women who develop PE have increased amount of circulating fetal components (fetal cells, hemoglobin and nucleic acids) in their blood compared to women with healthy pregnancies [68]. In PE an increased shedding of anti-inflammatory and anti-angiogenic particles from syncytiotrophoblasts into the maternal circulation is observed, resulting in a generalized inflammation and endothelial dysfunction in the second part of pregnancy [69].

The abnormal maternal immune responses, including reactions against antigens secreted from fetus and/or placenta, induce a hyperinflammatory state and facilitate development of the maternal presentation of the clinical syndrome. Women who have a massive predisposition for abnormal inflammatory reactions (Table 3, maternal biologic factors) and PE have small changes in the first stage of the developmental model. These individuals develop the disease mainly as a result of the changes initiated in Stage 2 with the expression mainly manifested by the woman, and not by the fetus (e.g. no IUGR) [66, 24].

As knowledge about the separate elements of the pathophysiologic mechanisms behind PE has expanded the two-stage theory has been revised, and the stages and their connections have developed into a more complicated structure of interacting biologic and pathologic processes. Both Stage 1 and Stage 2, and possibly the connections between the stages, involve generalized inflammation and endothelial damage. Before, behind and throughout the disease process a genetic predisposition, environmental influences and the life-style of a woman influence the development of PE (Figure 3).

The importance of placental tissue presence in the disease development indicates a fetal contribution to the disorder. Still, the pathogenesis of PE is most probably based on several mechanisms linking reduced placental perfusion and maternal systemic responses. The relative functional contribution of each phenomenon seems to differ between PE phenotypes, as some women with PE demonstrate severe maternal disease with massive intrauterine changes, while others with a similar maternal presentation have seemly nearly normal placental tissue and no fetal affection [70].

### *1.1.4 Disease prediction and prevention*

Despite arduous research efforts to avail early prediction of PE, aiming to improve and target surveillance and prevention, no effective screening for PE has been discovered and implemented yet. The regular antenatal care visits for all pregnant women screen for risk factors and signs and symptoms of the disease; hypertension, proteinuria and intrauterine growth retardation, and in most industrialized countries additional

ultrasound examinations are undertaken. As a screening method women with an expected high risk of PE development are offered Doppler ultrasound measurements of the placental and fetal blood flow around gestational week 24 to identify the individuals with increased risk of development of placental disease for intensified follow-up throughout gestation.

Within the field of medicine the search for specific biological markers in blood or urine (biomarkers) have surged, especially within cancer research, resulting in clinical usage of several well-documented biomarkers. A multitude of completed and ongoing studies are searching for biomarkers of PE. The aim is to predict with high sensitivity and specificity the subsequent development of PE, as well as the probable pace and severity of the disease. Current biomarker candidates include fetal and placental products, hormones, and angiogenic or anti-angiogenic factors estimated in the blood of the pregnant woman [71-77]. Despite an increasing amount of studies on biomarkers related to PE development, no biomarker has yet been implemented in nation-wide screening programs, and there is an ongoing discussion if any or a range of biomarkers may have a place within an algorithm for PE screening, by adding to the specificity and sensitivity of prediction based on a woman's medical and family history, other risk factors, and Doppler ultrasound examinations [78, 79].

Aspirin is an antiplatelet agent with a potential for inhibiting the platelet aggregation and so hinder initiation of the coagulation induced by increased oxidative stress in a preeclamptic placenta. Prophylactic use of the medication may inhibit the enhanced thrombosis formation of PE in the placenta and the associated placental ischemia-reperfusion injuries might be reduced or avoided. Currently a low dose aspirin is the only preventive measure widely recommended for high-risk women from gestational week 12 [19, 80, 81, 21].

### *1.1.5 Disease management*

Currently the management of PE is largely dependent on the disease presentation. In economically stable, industrialized countries observation of the affected woman is always included in the management scheme, either as an out-patient when she has a

supposedly mild disease phenotype without symptoms, or as a patient in the obstetric ward, admitted for closer surveillance, especially if severe disease is suspected or established.

In Norway current practice is for all affected women to be examined by an obstetrician, and to have blood pressure, urine samples and blood samples examined. To evaluate the well-being of the fetus cardiotocography (CTG) is required, trans-abdominally measuring the fetal heartbeat, movements and muscle activity of the uterus, as is trans-abdominal ultrasound examinations [21]. Both premonitory symptoms of eclampsia and established eclampsia would be treated by MgSO4-infusions, the only efficient treatment of eclampsia established [82]. As only delivery of the child and placenta will permanently reverse the development of PE, this is always the ultimate treatment, either by induction of labor, or a cesarian section is performed. After the birth the women with PE have their blood pressure, urine and liver values in blood monitored until normalization.

Some symptoms are treated during the pregnancy: antihypertensive medication is administered for severe hypertension as a prophylactic measure against cerebral insults and hemorrhage, and diuretics are given if the women develop lung or brain edema. In severe HELLP syndrome coagulation factors might be administered in connection with a cesarean section according to the amount and cause of blood loss during the operation. In cases of very preterm PE a woman would receive corticosteroids for fetal lung maturation before the preterm delivery [83].

There are currently no guidelines or formalized plan in Norway or internationally containing well-founded long-term follow-up strategies of the preeclamptic women or the children of preeclamptic pregnancies in regard to later development of cardiovascular risk factors and aCVD (see also paragraph 1.1.6).

### *1.1.6 Future health of women and their offspring after PE*

There is growing awareness of how a pathologic pregnancy not only influences the health of a woman and her fetus during the nine months of gestation, but that it also affects their long-term health negatively.

An increasing number of studies have identified preeclamptic women with an enhanced risk throughout life of developing cardiovascular risk factors and established aCVD [84, 7]. A metaanalysis have demonstrated increased risks of chronic hypertension (relative risk (RR) 3.7), ischemic heart disease (RR 2.16), stroke (RR 1.81) and venous thromboembolism (RR 1.79) [84] while other epidemiologic studies find higher risk for diabetes mellitus type 2 (hazard ratio 1.42-2.08) [85, 86], and renal disease (RR 3.2-15.5) [87] in women after a preeclamptic pregnancy. With early-onset PE and recurrent disease the risks are even higher [87]. The shared pathophysiology of PE paired with aCVD, diabetes mellitus type 2 or renal disease is intensively studied using molecular methods. Other studies try to identify shared etiologic mechanisms of PE and the other conditions using a genetic approach (see also paragraph 1.3.4, 1.3.5).

The stress-test hypothesis postulates why these women appear to have increased risk [88] (see also paragraph 1.1.1). Some women have an underlying constitutional predisposition for abnormal systemic inflammatory responses. Any pregnancy will cause an increase in the cardiovascular stress and low-grade systemic inflammation. Some women demonstrate an abnormal bodily reaction to a pregnancy, as in PE. For these individuals the inflammatory responses of gestation are disproportional to what is expected for pregnant women, and inducing cardiovascular changes and hypertension. These individuals could continue to display similar immune responses when exposed to risk factors later in life. A persistent low-grade triggering of the immune system through years by risk factors or environmental changes such as smoking, weight gain, or diabetes mellitus could result in development of atherosclerosis and subsequent aCVD [88]. Another theory on the link between PE and aCVD indicates that the development of aCVD after a preeclamptic pregnancy is not only due to risk factors, biologic and environmental, shared by the conditions as described in the stress-test hypothesis. The theory postulates that in addition to shared

risk factors of the diseases, the pathophysiologic mechanisms of PE modulates and enhances the risk the woman had pre-pregnancy [8, 9, 89].

The concept of how adaptations and environmental influences in utero impact the fetus' development of multifactorial disorders as an adult is often named the Barker Hypothesis or Fetal Origins Hypothesis [90]. The future health of children born in preeclamptic pregnancies is not widely examined. Influencing and confounding the meager results are the long-term health effects related to prematurity and IUGR as well as environmental factors and life-style, potentially shared with the mother. Most studies on health effects of intrauterine environmental influences / pathologic pregnancies have been performed on fetal growth restriction, demonstrating an increased risk for coronary heart disease, diabetes mellitus, obesity and insulin resistance [90]. Although IUGR is often associated with PE and the conditions share some pathophysiologic mechanisms, IUGR is not part of the syndrome of PE (see also paragraph 1.3.1). When the effects of IUGR and prematurity have been corrected for it is still uncertain how being born in a PE pregnancy effects the health of the child. Some studies demonstrate that these children do have an increased risk for development of risk factors of or established aCVD, but the results are discrepant [91-94]. Currently the only established long-term effects of being born in a preeclamptic pregnancy are higher BMI (0.62 kg/m2), increased systolic (2.26 mmHg) and diastolic (1.48 mmHg) blood pressure (after correcting for SGA in neonates born to term) and a 1.9-fold increased stroke mortality [95] [96, 97].

## 1.2 Cardiovascular disease

### *1.2.1 Epidemiology*

Cardiovascular disease (CVD) is a collective description of disorders affecting the heart and blood vessels. Despite declines in disease-related death rates the last four decades the condition is the leading cause of death worldwide [98, 99], as well as a main cause of diminished quality of life and a direct and indirect financial burden on health care services [100].

The CVD mortality rates are disproportionally higher in low and middle income countries (where over 80% of all CVD deaths take place) compared to developed countries, illustrated by the span in age-standardized mortality rates for CVD from below 200/100,000 of the population in industrialized countries to over 500/100,000 of inhabitants in other countries exemplified by Russia and Egypt (Figure 5) [100].

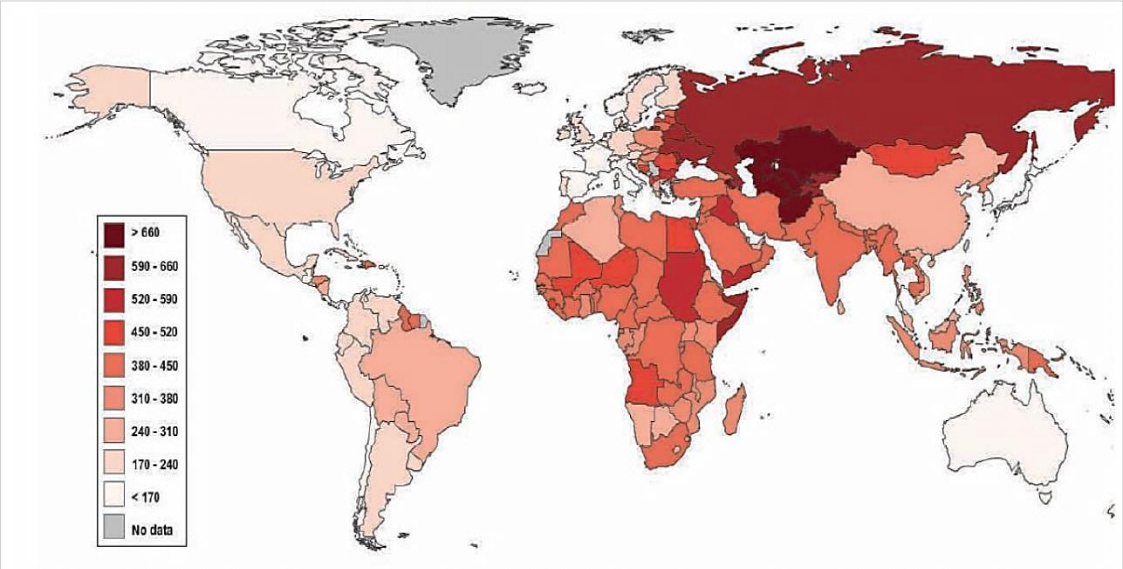


**Figure 5.** *Age-standardized deaths due to cardiovascular disease, 2004. Rates per 100,000, adjusted according to the WHO world standard population. Institute of Medicine (US) Committee on Preventing the Global Epidemic of Cardiovascular Disease [100].*

General estimates of prevalence and incidence for CVD and the conditions included in the diagnosis are difficult to give. The lack of consistent and long- term health statistics is partly due to discrepancies within countries and across borders, both in disease definitions and in how diagnoses are registered. Recently the disease term atherothrombotic CVD (aCVD) (see also paragraph 1.2.2) has been introduced, incorporating the main part of cardiovascular conditions in the definition of CVD, making a separation between inborn cardiovascular malformations and diseases that develop with aging. The definition is recent enough not to have been included in most registers yet. For persons affected by aCVD, the main causes of death are acute myocardial infarction and stroke [101]. The existing differences in incidence and

prevalence of the diseases between age-specific and socioeconomic groups, combined with gender differences in presentation and age of occurrence, further challenge a homogenous, reliable recording of the condition [102].

Despite the difficulties and inconsistencies in data on incidence and prevalence of CVD, receding CVD death rates are observed in industrialized countries, albeit at a faster rate in men compared to in women. The decline can in part be explained by the preventive focus aiming to reduce known risk factors such as smoking and a sedate life-style. As a consequence of these measures, including the usage of medicines to decrease levels of circulating cholesterol, CVD events take place less frequently. Further, the decreasing numbers in CVD mortality partly reflect how improved medical treatment modalities and increasingly successful surgical interventions extend survival in persons affected by CVD in industrialized countries [98].

### *1.2.2 Definitions and phenotypic subgroups*

The term *cardiovascular disease (CVD)* encompasses any condition of the heart and/or blood vessels. In this thesis, the main focus will be on the disorders included in the CVD phenotypes atherothrombotic CVD (aCVD) and thromboembolic CVD (teCVD) (Table 4). Further, in this thesis the conditions of teCVD have been merged into the term aCVD and the term used consistently, from paragraph 1.2.3 onwards, for both disorders is aCVD.

*Atherosclerosis* is the process of hardening the wall of blood vessels and thereby reducing tissue elasticity. Through deposition of fat substances, fibrin and cellular waste products a plaque is formed in the inner part of the muscular wall causing a roughened luminal surface and narrowed lumen susceptible to blockage (Figure 4). Pathophysiologic details of atherosclerosis and atherothrombotic cardiovascular diseases are given in paragraph 1.2.4. Atherosclerosis is the main cause of aCVD but the disorders of this condition can also result from anomalies and systemic pathologic processes including degenerative diseases such as Marfan and Ehlers-Danlos syndromes, dysplastic disorders like fibromuscular dysplasia, and vascular inflammation as in giant cell arteritis or rheumatoid arthritis.

**Table 4. Phenotypes and definitions of cardiovascular disease relevant for this thesis**

| Phenotype / Subgroup | Affected blood vessels | Causative mechanisms | Clinical conditions |
|---|---|---|---|
| Coronary heart disease[a] | Arteries of the heart muscle | Atherosclerosis and thrombosis | Angina pectoris<br>Cardiac arrhytmia<br>Acute myocardial infarction<br>Heart failure<br>Sudden cardiac death |
| Cerebrovascular disease[a] | Arteries supplying the brain | Atherosclerosis, thrombosis, embolism and hemorrhage<br>Atrial fibrillation | Transient ischemic attack Ischemic stroke<br>Hemorrhagic stroke |
| Peripheral arterial disease[a] | Aorta and arteries to the lower limbs | Atherosclerosis<br>Degenerative disease<br>Dysplastic disease<br>Inflammation | Aorta aneurysm (and rupture)<br>Claudicatio intermittens<br>Acute ischemia due to embolus |
| Venous thrombosis[b] | Deep veins of the lower limbs<br>Pulmonary arteries | Blood stasis<br>Trauma<br>Heritable hypercoagulability states<br>Disseminated cancer<br>Medication<br>Atrial fibrillation Cardiac failure | Deep venous thrombosis<br>Pulmonary embolism |

Adapted and modified from the World Health Organization definition of cardiovascular diseases 2014: http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cardiovascular-diseases/cardiovascular-diseases2/definition-of-cardiovascular-diseases. Checked 260914. And Yusuf et al 2004 [103]

a Conditions included in the term atherothrombotic cardiovascular disease

b Conditions included in the term thromboembolic cardiovascular disease

*Atherothrombosis* can be defined as development of atherosclerosis, lesion disruptions or fissures, and associated thrombotic complications [104].

*Atherothrombotic cardiovascular disease (aCVD)* covers diseases in which atherosclerosis development is recognized as the main underlying mechanism: coronary heart diseases, cerebrovascular diseases, and peripheral vascular diseases [105].

The term *thromboembolic CVD (teCVD)* describes the formation of thrombosis, usually in low-pressure systems such as the veins, the pulmonary artery or the atria. Major causes of teCVD are blood stasis as caused by immobilization, trauma and endothelial injury, heritable hypercoagulability states, disseminated cancer, medication, atrial fibrillation, and cardiac failure [105].

An *embolus* is a part of a blood cloth or a complete thrombus that ruptures from the site of development, subsequently plugging either the lumen of the same artery more distant relative to the heart or blocking the lumen of another, more distally located artery. A thrombosis of the venous system may be mobilized and create an embolus to the pulmonary arteries. In patients with impaired left ventricular function of the heart, an embolus may originate from a thrombosis formed within the left ventricle. A frequent cause of embolization is thrombosis formation in the left atrium of patients with atrial fibrillation. These emboli originate from a low-pressure system and usually end in the cerebrovascular bed. The origin of emboli in the venous system is less studied but a route of transfer from veins to the arterial systems by an open foramen ovale or atrial septum defect has been established.

In the lower extremities veins are classified as superficial or deep according to their physical placing relative to the muscular fascia. The deep veins drain the muscles of the lower limbs, returning the blood to the right atrium of the heart and to the lungs. The term *deep venous thrombosis* (DVT) usually refers to thrombosis formation in the deep vessels of the lower extremities.

The term *ischemia* embodies the metabolic changes that occur in tissues due to an inadequate oxygen supply (hypoxia), and may be induced impaired arterial blood supply or an increased oxygen demand. Acute or chronic ischemia may present different clinically according the organ affected.

*Coronary heart disease* encompasses disorders of the arteries supplying the heart muscle (ref). If the structure of the artery is changed, most commonly due to depositions of cholesterol and other lipids within the inner wall as in atherosclerosis, a subsequent narrowing of the lumen may evolve and cause hypoxia of the cardiac muscle. The associated ischemia may result in angina pectoris, atrial or ventricular arrhythmias or heart failure [106].

*Cerebrovascular diseases* are caused by acute or chronic changes in the arteries supplying the brain tissues. Partial or total occlusion of a vessel by atherosclerotic changes, a thrombus, or an embolus, may temporarily cause cerebral ischemia, leading to a transient ischemic attack [104] (defined as acute appearance of symptoms of focal loss of cerebral functions or sight, lasting less than 24 hours) or an ischemic stroke. Occlusion of cerebral vessels most commonly evolves from an embolus originating in the carotid arteries, the aorta or the heart. Thus, transient ischemic attack and ischemic stroke can be cause by either aCVD or teCVD. The condition cerebrovascular disease also includes cerebral hemorrhage [107, 108].

*Periferal arterial diseases include* disorders of the aorta and the arteries to the lower limbs, conditions that may lead to transient or permanent ischemia (Table 4).

### *1.2.3 Risk factors for aCVD*

Multiple risk factors are acknowledged to influence development of aCVD, including genetic, environmental and behavioral factors such as smoking and inactivity, and presence of other disorders like diabetes mellitus type 2 (Table 5) [103, 109]. Individuals commonly present a combination of risk factors as the risk conditions often influence and interact with each other, and the underlying disorders often share etiologic functions with aCVD [110]. The different risk factors are commonly separated into modifiable and non-modifiable influences (Table 5) [103, 111].

Main risk factors for atherosclerosis and consequent aCVD development that can be modified by changes in life-style or medication are cigarette smoking, an unhealthy diet, alcohol consumption, physical inactivity, overweight/obesity, hypertension, high

cholesterol levels, and diabetes mellitus type 2. Overweight (body mass index > 25 kg2/m) and obesity (body mass index > 30 kg2/m) are variants of an established risk factor for aCVD. The risk has been demonstrated repeatedly [98, 112-114].

In one meta-analysis individuals with increased BMI were found to have a hazard ratio of 1.26 (overweight) and 1.69 (obese) of coronary heart disease, while the correlating numbers for stroke are 1.13 and 1.47 when compared to persons within the normal range of BMI [98, 112-114]. Hypercholesterolemia is closely related to coronary heart disease and cerebrovascular disease, as is dysregulation of subsets of lipoprotein molecules with increased low-density lipoprotein (LDL) cholesterol and decreased high-density lipoprotein (HDL) cholesterol [115]. Presence of diabetes mellitus type 2 increases an individual's risk of coronary heart disease 2- to 4-fold, the risk of ischemic stroke 3-fold and of peripheral arterial disease 3.5-fold (men) and 8.6-fold (women). Several publications demonstrate either how a low intake of fruit and vegetables impact negatively on cardiovascular health or that high-energy diets increase the risk of developing obesity, a modifiable risk factor for aCVD [116, 117].

Non-modifiable risk factors for aCVD are ethnicity, family history and the gestational history of women, genetic predisposition and sex. Heritability of several aCVD phenotypes has been demonstrated, such as for coronary heart disease and hypercholesterolemia. Individuals known to have first-degree relatives who have developed premature aCVD (occuring in men <55 years of age and in women <65 years of age) are defined with a more severe risk profile than the general population. If a person has experienced stroke previously she/he has a 39% risk of a new incident within a ten-year period [118, 119].

Hypertensive diseases in pregnancy and gestational diabetes are now recognized as independent risk factors for aCVD development by the American Heart Association Guidelines [109]. In the 2012 version of the European Guidelines the gestational diseases are not considered in risk estimation of aCVD [120]. In many guidelines male sex is listed as a risk factor for aCVD as men develop the disorders at a younger age.

**Table 5. Non-modifiable and modifiable risk factors for atherothrombotic cardiovascular diseases**

| Risk factor | Non-modifiable | Modifiable |
|---|---|---|
| Unhealthy diet | | x |
| Physical inactivity | | x |
| Overweight and obesity (especially central adiposity). | | x |
| Hypertension | | x |
| Diabetes mellitus type 2 | | x |
| Cigarette smoking | | x |
| Hypercholesterolemia (total cholesterol ≥200 mg/dL) | | x |
| High LDL-C | | x |
| Low HDL-C (< 50 mg/dL) | | x |
| Age (male>=45 years, female>=55 years) | x | |
| Sex | x | |
| Ethnicity | x | |
| Genetic predisposition | x | |
| Family history of premature CVD occurring in first-degree relatives in men <55 years or women <65 years | x | |
| Low socio-economic status | | x |
| Modified from The 2011 American Heart Association Guidelines[109], Japan Atherosclerosis Society (JAS) guidelines 2012 and Systematic review by BS Ferket et al. 2010 [110]. | | |

The discrepancy between male and female aCVD development is ten years on average. The last decade there has been an increased focus on how aCVD might develop differently according to sex, demonstrating some discrepancy in risk factors and presentations, and also the influence of menopausal state in women. Possibly some divergence in the pathophysiology of the disease exists between the sexes, as

inflammation appears to have a more fundamental role in the pathophysiology of female aCVD [121].

The increased occurrence in families of diseases encompassed by the aCVD definition (see also paragraph 1.2.2) indicates genetic contributions to the pathophysiology of these conditions. Studies based on such families have identified several genetic regions with a high functional impact on the generation of aCVD phenotypes. Genetic changes in *LDLR*, *APOB*, *PCSK9*, and *LDLRAP* are associated with familiar hypercholesterolemia, and defined genetic variants in the Factor V Leyden gene (*F5*) heightens the risk of deep venous thrombosis [122, 123].

By use of other genetic approaches (see also paragraph 1.3.5) several genetic areas and genetic variants are found to be associated with development of aCVD. The more established associations are found between defined aCVD phenotypes and the genes in the pathways involved in expression of nitric oxide synthase (like NOS-3 and endothelial dysfunction [124], adrenergic receptors, endothelin (ET-1 and hypertension [125] and natriuretic peptide, as well as in genes potentially regulating the renin-angiotensin-aldosterone system and those involved in the biological mechanisms of oxidative stress [126].

Using hundred thousands of individuals in genome-wide studies on the multifactorial condition of chronic hypertension has, on the other hand, as yet only resulted in identifying genetic changes counting for about 2.0-4.6 mmHg of blood pressure differences [127-129]. Despite the absence of large direct genetic influences on blood pressure regulation, the biologic mechanisms influenced by the identified genetic changes could represent potent antihypertensive drug targets.

### *1.2.4 The etiology of atherothrombotic cardiovascular diseases*

The main pathophysiologic mechanism of aCVD development is the atherosclerotic process. Development is slow, but atherosclerosis formation has high potential of rapid progress. In 1977 the term and theory "response-to-injury" was introduced, describing how atherosclerosis results from an initiating injurious event such as chronic

hyperlipidemia, infections, or mechanical stressors of the vessel endothelium (Ross, 1977). This theory forms a basis for the present knowledge of atherosclerosis formation in vessels.

Cholesterol is a molecule necessary for functional cell membranes and for vitamin D, hormone and bile acid production in the body. Cholesterol is generated intrinsically by the liver and is additionally delivered to the body through food high in saturated fats. In blood HDL transport cholesterol to the liver, while LDL transport the cholesterol molecules from the liver to other cells in the body.

Vessel walls are constituted by three main layers; the tunica intima, tunica media and tunica adventitia. At highly susceptible sites of the vessel wall, especially at vessel branching points, mechanical forces of disturbed blood flow can cause endothelial injury and thereby initiate in the vessel an inflammatory process, which may become chronic. Through the injured endothelium LDL enter the tunica intima where the lipids undergo oxidative modification and induce endothelial expression of adhesion molecules and production of cytokines and reactive oxygen species.

Monocytes, a subtype of leukocytes, are attracted to the place of injury and enter the tunica intima, where the cells transform into macrophages. As the mature macrophages phagocytose the LDL-particles containing cholesterol they become foam cells, which then accumulate in the vessel wall and form fatty streaks. A fibrous capsule gradually surrounds the fatty streaks, forming an atherosclerotic plaque containing fatty substances, small cholesterol crystals, cellular waste products and fibrin (Figure 4) [129, 189].

Early plaque development does not impact the lumen as the pathologic changes first grow outward toward an expanding internal elastic lamina. Later in the process the internal elastic layer is disrupted, the tunica media is involved, and smooth muscle cells start to multiply and move to the surface of the plaque, contributing to the formation of a fibromuscular cap. The luminar surface of the cap may rupture and induce formation of a hematoma and thrombus. Through subsequent healing the thrombus can be incorporated into the plaque and combine with calcium deposits to

form a new fibromuscular capsule, resulting in a plaque with a lipid-rich, necrotic core, hardening of the vessel and loss of vessel elasticity. Gradually such processes induce a plaque that intrudes into and partially or totally blocks the vascular lumen. Microvessels are formed within the plaque, possibly as a result of inflammation-related signals released from the cells in or surrounding the atherosclerotic plaque (Figure 9) [189]. In addition to transporting blood to the plaque, the neovascularization does provide an increased surface area for monocyte recruitment and adhesion, resulting in a sustained inflammatory response [189].

Increases in the pressure of blood squeezing through the smaller lumen damage the fibromuscular capsule, particularly at the thinnest and most vulnerable point, and thereby rendering the plaque unstable. Secreted products from macrophages and foam cells further induce weakening of the fibrous cap and promote thrombosis formation. As the capsule of the plaque ruptures, the necrotic core is exposed to the blood and the thrombogenic, lipid-rich content triggers the thrombotic process. The resulting thrombus is either incorporated into the plaque by formation of a new fibrous capsule or breaks into circulation becoming an embolus. In relocating to smaller vessels the embolus may obstruct the blood flow, thereby causing a hypoxic state in tissues. Hypoxia induces ischemia and, if sustained, necrosis of the tissues supplied by the blocked vessel.

### *1.2.5 aCVD screening and prevention*

There are few generalized guidelines for how to and when to screen for aCVD, or for regulation of follow-up regimes when risk factors have been established. The existing recommendations for prevention and treatment of the conditions of aCVD differ in details between guidelines but the principal components are compatible (ref, ref). The treatment options and available modalities of therapy vary with disease phenotype and severity at presentation. Currently no nation-wide or regional systematic routine screening programmes for atherosclerosis-related diseases exist globally, nor an international agreement on the potential age at which screening should commence or on the intervals between examinations.

Recommendations are that individuals of identified risk groups (see also paragraph 1.2.3) are evaluated for high risk of development of aCVD according to their own medical history, presence of unfavourable dietary and life-style habits, and their family history for detection of diseases which are partly heritable, such as premature aCVD. Except for persons with a family history of certain diseases like familiar hypercholesterolemia, where the affected genes are known and preventive treatment has a demonstrated effect on disease development, no genetic screening is recommended. As mentioned in the 2012 version of the European Guidelines the gestational diseases are not considered in risk estimation of aCVD [120]. The initial screening for persons with increased risk of aCVD who could benefit from preventive treatment or who have established disease should also include an examination of blood pressure, BMI and waist-circumference as well as laboratory measures of relevant risk factors in blood.

Further mapping of risk factors should be offered persons with very high risk for development of aCVD and those with suspected established disease. Subtypes of the disorder and some precursors can potentially be identified by several non-invasive methods like electrocardiogram, echocardiography, measures of ankle-brachial blood pressure ratios, as well as ultrasound estimations of the vessel wall thickness of carotid arteries and aorta, or computer tomography scanning to identify cerebral ischemia or hemorrhage, cerebral or aorta aneurisms. Angiography is a favored invasive method of mapping atherosclerosis development in coronary arteries. Currently there are no biological markers (biomarkers) for aCVD established for regular use, even though several promising biomarkers in blood are described: lipid-associated markers for primary aCVD, fibrinogen for primary stroke, and ischemic-related biomarkers such as C-reactive protein as possible part of the follow-up regime after a first disease event [130, 120].

Intervention would differ according to the risk factors exposed and the disease subtype experienced. Improval of unfavorable eating habits, increases in levels of physical activity, reduction of weight (if overweight), and cessation of smoking are regarded as appropriate actions both for primary and secondary prevention [131] as is the adequate

usage of anti-hypertensiva, lipid lowering drugs, and strict management of diabetes mellitus by diet and/or medication.

## 1.3 Genetics

Virtually all biological characteristics including diseases, contain a genetic component, and consequently genetic changes shape the patterns of evolution and biodiversity of an organism. Genetics, the branch of science concerned with the consequences of biologic heredity and variations, has emerged as an area of major importance in almost all fields of medicine. The research area investigating the processes of how environmental factors influence gene expression of the corresponding protein production is called epigenetics [132]. Epigenetics will not be discussed further in this thesis. The increasing focus on medical genetic research may be due to the discovery that common health conditions and diseases (cancer, diabetes, obesity, psychiatric disease) are complex genetic diseases and that revealing the causality could influence the health of large parts of the population, contrasting with initial genetic approaches mainly covering rare diseases caused by one or a few deleterious genetic changes. Concomitant development of laboratory methods and advanced software for both research and diagnostics, at decreasing costs, has made large, complex data sets accessible and analyses feasible. When applied to common diseases it could diminish unnecessary physical discomfort and anxiety in patients, while parallelly more effective and less expensive treatment regimes could be developed. Another incentive to combine genetic research and findings with clinical medicine is the growing awareness and demand for individually adapted therapies for those patient groups where it would be efficient and health-promoting. By understanding the interplay between genes, environment and diseases we can further improve our understanding of physiological processes and the mechanisms that are altered when disease develops.

### *1.3.1 The basics on human genetics*

The central dogma in molecular biology states that DNA (comprising a gene) transcribes into RNA, which again translates into proteins (Figure 6).

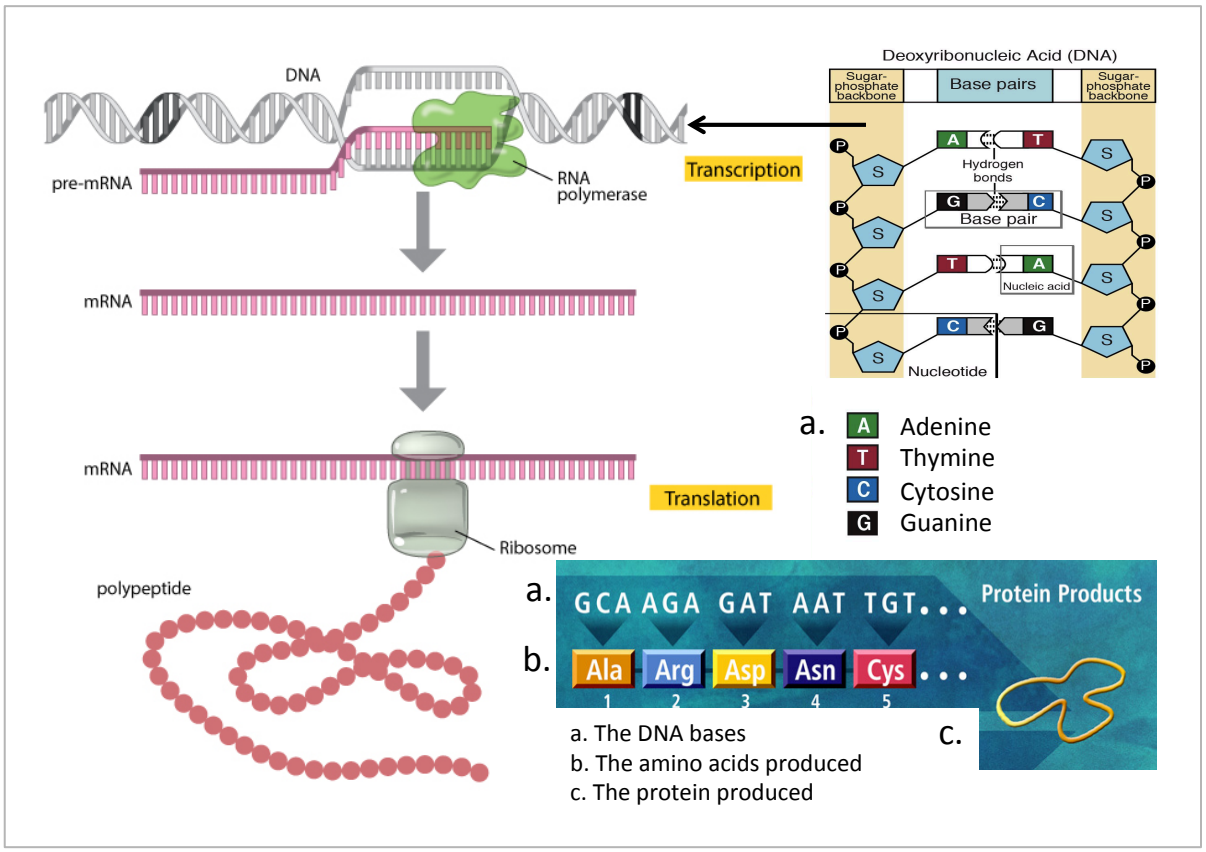


**Figure 6 .** *Illustration of the transcription and translation processes in the cell. From DNA to production of polypeptides and proteins. Modified from Clancy & Brown 2008 [133].*

*Deoxyribonucleic acid (DNA)* contains all information required to sustain life of an organism and for transmission of this data from one generation to the next. The genetic information is conveyed to the somatic cells, thereby regulating the synthesis of proteins and cellular functions. The four different nitrogenous bases which make up the genetic code are adenine (A), cytosine (C), thymine (T) and guanine (G) [134]. The combination of a sugar molecule bound by a phosphate group with a nitrogenous base is called a *nucleotide*, and the link of connected nucleotides creates a single DNA strand. The combining of two strands, running in opposite directions, by weak hydrogen bonds, constitutes the DNA helix. The double stranded DNA loops around histone octamers to form nucleosomes, which are coiled to form chromatin fibers, located to the nucleus of all cells (Figure 7).

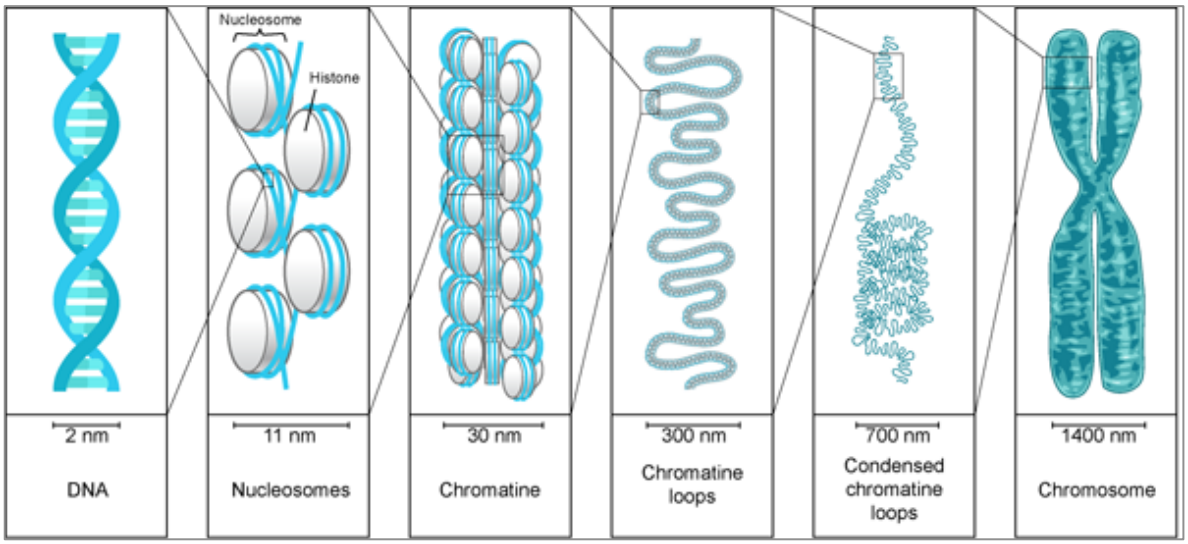


**Figure 7.** *The structure of DNA (from left to right) with a double-stranded helix, DNA packed around histones into nucleosomes, nucleosomes forming a chromatine, chromatine forming loops and condensing, finally attaching to another chromatine at the centromere (the three boxes to the right). From homepage: http://www.shmoop.com/dna/dna-packaging.html.*

*Ribonucleic acid (RNA)* is a macromolecule similar to DNA in composition, but with ribose sugar backbone instead of deoxyriboses as found in DNA, and uracil (U) nucleobase instead of thymine. A single stranded RNA is formed on a chromosome in the process called transcription as the DNA strands separate and the genes are made accessible. While the genetic code on DNA is copied into messenger RNA (mRNA: for transport of genetic code to the ribosomes), transfer RNA (tRNA; for amino acid transport to ribosomes) and ribosomal RNA (rRNA; for peptide synthesis) are also produced in the nucleus. Single RNA strands peel away from the DNA template and leave from the nucleus of the cell and form large complexes with the ribosomes of the cytoplasm. The ribosomes read the templates where amino acids are added to form peptides during *translation*. The nucleotide sequence of mRNA transcripts is read in triplets, each tri-nucleotide combination constitutes a specific amino acid upon translation and formation of protein molecule.

*Chromosomes* are formed as part of the cell division processes, where the genetic material is replicated and two double-stranded DNA molecules are formed for the daughter cells, in principle identical to the original DNA helix. The two identical molecules, chromatids, are attached through their centromere part as the cell divides into two daughter cells to form a chromosome. Chromosomes are inherited one from each parent and are found in the nucleus of cells. Human somatic cells contain 23 chromosomes, 22 pairs of autosomes and one pair of sex-determining chromosomes (XX for females, XY for males). Each of the chromosomes has been assigned a number and is identified microscopically according to size, where the centromere is positioned, and the chromosomal banding pattern.

A *gene* is a stretch of DNA from chromosomes which is transcribed into RNA. The genes are the known functional units of heredity and each gene occupies a specific place, a *locus*, on a chromosome. Genes are capable of reproducing themselves exactly during each cell division, and they are the part of DNA directing the formation of proteins. Human beings have approximately 21.000 genes, resulting in hundreds of thousand accompanying protein products. A gene contains components called *exons,* which are the coding parts of the gene which are expressed as peptides or proteins, and intragenetic intervening non-coding segments named *introns* with potentially regulatory function. About 2 to 3 percent of total DNA is exonic. The *promoter* region of a gene is the upstream flaking region located on the same DNA string, near the site where transcription of that gene starts. The distance between genes on a DNA string is named *intergenic regions or intergenetic deserts*.

*Alleles* are the existing variant forms of a genetic locus [135]. All different variants located at the same physical position on a specific chromosome are defined as alleles. Humans have two alleles at each genetic locus, one allele inherited from each parent. Each pair of alleles represents the *genotype* of that specific locus. Which protein being produced by a gene is dependent on the combination of alleles present. *Genotype* can refer to the combination of alleles in a single gene and to the total set of genes in the genome. Locus *homozygosity* is when two identical alleles exist at a particular locus on the paired chromatins, as opposite to locus *heterozygosity* when the two alleles at a

particular locus differ. As the name indicates, a dominant allele will dominate the expression of its genetic locus, either an individual is homozygous or heterozygous, having one dominant and one recessive allele at the specific locus. The trait for which a *recessive allele* codes will only be expressed if the locus is homozygous for that allele.

A *haplotype* refers to a group of alleles at a genetic locus on a single DNA strand that tend to be inherited together as a unit [136]. The basis of Mendel's principle of independent assortment alleles is how all genes will segregate independently during meiosis, the division of germ cells forming gametes, and that during each meiosis a random exchange (crossing over) of genetic material will take place. These principles have in the years since been demonstrated to be inaccurate for human beings. *Linkage disequilibrium (LD)* describes how specific alleles at two or more loci are inherited together with a frequency greater than expected by chance [137]. The degree of linkage is increased with increasing proximity of the alleles. Complete linkage between two genes in a parent will result in all offspring containing that allelic combination/haplotype, while *genetic independence* between two alleles in a parent will appear with the same allelic combination of those loci in 50% of the offspring and as recombined alleles in 50%.

In a model population, the relative frequencies of alleles in a genotype in that population will remain constant from generation to generation, and the genotype is said to be in *Hardy-Weinberg equilibrium (HWE)*[138]. Despite evolutionary sources (see also paragraph 1.3.2) influencing the proportion of genotype and disturbing HWE, significant deviations from HWE are rarely seen for the larger part of genotypes in most populations. Estimation of HWE is used for quality control of genetic studies. If deviations are detected for any genotype, this usually indicates methodological genotyping errors, but the divergence may also be caused by bias in the selected cohort [138].

### *1.3.2 The genetic variation of populations*

The genomes of human beings are largely identical and adapted for the basic requirements of the biologic functions and mechanisms the body depends on to survive. Still, to sustain genetic variation and diversity between individuals and within populations, favorable genes must be permitted to be transmitted to offspring. Heterogeneity reduces the proportion of disadvantageous variants in a population, such as alleles predisposing to disease phenotypes and reduced survival. The required genetic variation in populations is added through several evolutionary sources: gene flow, genetic drift, selection and mutations (genetic variants spontaneously induced into the genome during a cell division).

#### 1.3.2.1 Evolutinary sources of genetic variation

The term gene flow incorporates the exchange of genetic variants between populations, most commonly achieved through geographical migration of individuals. Gene flow is important to evolution as spread of favorable genetic variants from one population to another to generate increased survival of individuals and thereby stability in a population. Disadvantageous, often rare and recessive, mutations commonly accumulate through inbreeding, leading to homozygocity and disproportionately high prevalence of rare diseases. The transfer of genes can decrease allele clustering within one population.

Genetic drift influences the development of heterogeneity within populations through random changes in allele frequencies from one generation to the next. These intergenerational alterations commonly involve mutations and alleles with low frequency in the population. Genetic drift results in decreased variation within a population but this source of evolution cause increased variation between populations. In effect it may lead to complete disappearance of uncommon genetic variants from a population. Another effect of this evolutionary process can be seen when a new population is established from a larger population and based on very small numbers of individuals genetic variants which are rare originally may aggregate due to restricted selection when it comes to procreation (the founder effect) [139].

In populations the evolutionary process of natural selection is based on the existence of biologic variation combined with imbalances between reproduction and availability of food supplies for each generation. The individuals possessing favorable genetic variants and phenotypes have a comparably higher chance of survival and production of offspring. Consequently, those genetic variants and phenotypes associated with survival within a contemporary environment are biologically selected for and inherited to the next generation.

Mutations add new genetic information, and as such, they are the creative factor in evolution. A mutation is a random insertion of base sequence, which may lead to altered alleles into a population. The mutational change can occur in germ line cells or in somatic cells, and on average each person have ten genetic mutations throughout the genome. If a mutation occurs in the gametes and is compatible with life, and also is passed from one generation to the next, the mutated allele may disseminate in a population and develop into a polymorphism.

#### 1.3.2.2 Polymorphisms

The genetic variation introduced into populations through gene flow, genetic drift, selection and mutation are defined as polymorphisms, variations in the DNA sequence. A locus on the DNA string, gene or chromosome is polymorphic if two or more relatively common alleles co-exist in the population. The most prevalent genetic polymorphism is the single nucleotide polymorphism (SNP) but also polymorphisms like insertion/deletion variants, microsatellites/short tandem repeats, and copy number variations occur frequently in humans.

A SNP is a variation among individuals within a single genetic position, at the locus of a single nitrogenous base (Figure 8). SNPs are commonly biallelic, which means that at the given genetic locus one of two possible allelic variants can be found.

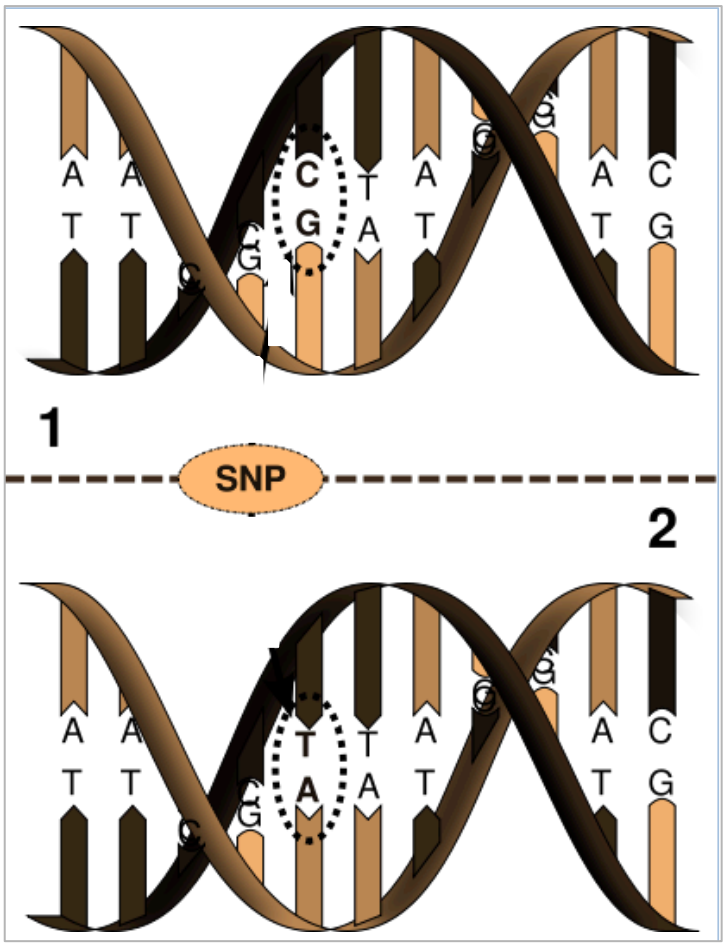


**Figure 8.** *A single nucleotide polymorphism exists at the same locus in the two alleles of a gene, with CG basepair in allele 1 (1) and TA basepair in allele 2 (2).*

An abundant number of SNPs exists in the genome, dispersed throughout coding regions, introns, promoter regions and intergenic loci. While a multitude of SNPs and genetic changes in regulatory sequences of or near genes have been examined for possible influence on phenotypic presentations, SNPs located in non-coding regions have been indicated as having little or no phenotypic effects. As few studies have focused on variation in non-regulatory and intergenetic deserts compared to the interest given SNPs and loci involving genes, little is currently known about the importance of these regions. The abundance, stability and dispersion of SNPs in the genome, and that they are readily detectable in most tissues with available and affordable methods make these polymorphisms attractive targets for genetic research.

The insertion/deletion polymorphism (indel) is a form for polymorphism induced by adding (inserting) or removing (deleting) one or a few bases in the DNA sequence, resulting in alleles of different lenghts [140]. Thus, an indel within a coding region of a gene could cause a frameshift, which usually induce a premature stop codon in mRNA during translation. The resulting preterminated protein could cause a so-called loss-of-function phenotype. The majority of indels occur in regions of DNA containing microsatellites.

Microsatellites, also named short tandem repeats or simple sequence repeats, are loci within DNA where short sequences of 2-6 base pairs are repeated one right after the other, in tandem arrays. These polymorphisms are highly polymorphic, remarkably heritable, and they demonstrate a pattern of co-dominance where both alleles contribute to phenotypic expression. Several thousand microsatellites have been characterized and genetically mapped, and the polymorphisms are commonly used as molecular markers in population research as it is an efficient tool for mapping genetic relationship, as well as in studies investigating effects of duplication or deletion of genes [141].

Copy number variations (CNVs) are chromosomal insertions, deletions and duplications of DNA segments of at least thousand base pairs. Most of these polymorphisms occur spontaneously and the affected region can encompass entire genes. When whole genes are involved, changes in the quantity of produced RNA are observed due to altered gene expression, and pathologic changes may be induced as a consequence. Around 12% of the genome is copy number variable and genomic loci around genes involved in neurologic and immunologic development and activity appear to be enriched for CNVs. Even if CNVs generally result in copy numbers differing from those seen in a standard genome, many of these polymorphisms do not cause disease.

### *1.3.3 Genetics in Medicine*

The aim of genetic studies performed today within the field of medicine is to investigate genetic alterations against physiologic changes and pathogenic

mechanisms resulting in disease. The field of genetic medicine incorporates both inherited diseases and genetic disease caused by mutations, both in germ cells and in somatic cells. Medical conditions with partly or completely genetic origins can be classified as monogenetic disorders, chromosome abnormalities, polygenic disease, multifactorial/complex genetic disease or acquired somatic genetic disorder.

Monogenetic disorders are caused solely by alterations in a single predefined gene with extensive impact on phenotype, without environmental influences. The different monogenetic conditions classify either as dominant disease (requiring only one affected allele to develop) such as the neuromuscular condition Huntington's disease, recessive disease (where both alleles have to be mutated for the presentation of phenotype) like cystic fibrosis, or X-linked (presentation is related to the sex of an individual) as in the bleeding disorders named hemophilias. Monogenetic diseases are considered rare disorders, and these conditions are commonly hereditary and thus accumulate in families. The diseases usually have devastating effects on health and survival.

Multifactorial genetic diseases result from combinations of several genetic changes and environmental influences. To reveal the complex mechanisms of these interactions and how they may cause disease is a challenging task. The genetic basis of most complex diseases is still unknown, but two hypotheses have been frequently advocated: The common disease–common variant hypothesis proposes that one disease will be caused by a few common polymorphisms in the majority of cases, and the multiple rare-variant hypothesis conceptualize the theory that the same disease phenotype in different individuals would be caused by disparate rare genetic variants [138].

The common disease–common variant hypothesis has been supported by identification of a causative gene in familial forms of common diseases such as hypercholesterolemia and diabetes, but until recently few variants with a large impact on phenotype have been described for non-familial forms of these disorders [142]. SNPs that occur in less than 1% of the population are classified as rare variants, and

may have an extensive effect on phenotypic expression as they may identify rare disease-associated variants. Rare SNP mutations can demonstrate co-inheritance with disease in some families. Although individually rare, these SNPs may be sufficiently clustered within a gene or small genomic region to be identified statistically as an important locus of mutational variation.

### *1.3.4 Epidemiologic observational study designs and their application in genetic studies*

Epidemiology started as a study of epidemics, of how contagious diseases were dispersed in a population, but has developed to also investigate the distribution and incidence of noncommunicable diseases, chronic disorders, and health related conditions in populations. The aim of epidemiologic studies is to produce knowledge about the occurrence and causes of conditions in the population to identify risk factors and possible preventive measures, as well as guide management of disease on a population-based level [143]. Within the field of epidemiology the main study approaches can be divided into experimental and non-experimental or observational. In experimental studies investigators manipulate the exposure to a condition within a population to investigate if the outcome changes according to presence or absence of exposure. For observational studies the researchers do not influence conditions, but observe the outcomes in populations according to reported exposure.

Genetic epidemiology is a branch of epidemiology that has developed during the last four decades and which investigates the potential influences of genetic factors on disease development at a population-level [144]. This field of medical research has evolved in parallel with advances in molecular techniques, statistical methods and software development.

The main types of observational epidemiologic studies are the case-control, cohort, cross-sectional, and ecological studies. For cross-sectional studies data relevant to exposure and outcome of interest are assembled from the general population at a set point in time and participants are included irrespective to disease-state/outcome and exposure. Ecological studies investigate incidence and prevalence of conditions within a set population, comparing different time periods or geographic areas. The analysis is

not performed on individual levels, but based on group levels, and thereby not relevant as a genetic approach.

In this chapter the focus will be on the study designs of case-control and cohort studies. As a considerable part of the projects included in the thesis are based on the PE Family Biobank, the family study design, one of the main study designs of genetic epidemiology, will be explained in some detail.

A cohort is a group of people who share a common characteristic, phenotype, condition or experience within a defined period, such as being born the same year, or having been exposed to a vaccine within the same 10-year period [143, 144]. This type of study follows a group of people who do not have the disease and uses correlations to determine the absolute risk of development of the disease in question. In cohort studies a comparison group may be the general population from which the cohort is drawn, or one can compare subgroups within the cohort. A challenge for this type of investigations can be logistic, such as how loss to follow-up, usually for unknown reasons, may influence the composition of the cohort. Although end-points like survival can be given accurately it can be demanding to define and determine the time of event precisely for many outcomes in longitudinal cohort studies. As time from exposure to end-point can be long for many conditions, the data collection in cohort studies must at least continue the estimated time span from exposure to outcome. For cohort studies to generate valuable scientific results the defined outcome needs to be prevalent in the examined population.

A case–control study is an observational epidemiological study of persons with a disease and a suitable control group of persons without the disease [143, 144]. The potential relationship of a suspected risk factor to the disease is examined by comparing the diseased and non-diseased subjects with regard to how frequently the factor is present in each of the groups. If the exposure examined has been much more frequent in the cases than in the controls this suggests a connection between exposure and outcome. This is a useful and also cost-efficient study design for investigation of diseases with low incidence, as individuals can be identified after they have

established the disease and then be included as cases. Matching of cases and controls is an integral part of these studies, as matching is the process where individuals without the outcome of interest, but otherwise similar to cases regarding selected basic characteristics like age, sex and ethnicity are chosen as controls. Using too many variables during the matching process may cause difficulties in identifying enough individuals for an appropriate control group. Further, the variables used for matching cannot be included in the main analyses for possible disease associations.

In all studies that examine a population according to past exposures there will be the bias of how the individual remembers the information that is collected retrospectively [143, 144]. An individual with the disease is more likely to remember specific exposures potentially related to development of the condition than a healthy person (ref). This recall bias might lead to identification of a false positive association between that exposure and the selected outcome.

Nested case-control designs are often selected to avoid recall bias. This type of study designates a group of participants from a prospective cohort study in which descriptive data and information related to exposure and outcome of interest is available. The exposure data used in the nested study have been collected prospectively, and bias recall bias would be minimized as well as potential bias caused by the data collector knowing if the participant has been affected or not.

Family studies include families with increased prevalence of the disease of interest to illuminate transmission of genetic variants with phenotypes. Common family study designs are twin studies, examining concordance between disease presentation in monozygotic versus dizygotic twins, and studies on trios based on an affected child and both parents irrelevant of disease affection. Another type of family study design, used in this thesis, is where all available family members are included, and several generations if possible. When the term "family study" is applied in the remaining part of the thesis it will consistently allude to this last type of family study design.

Through family studies it is possible to investigate how specific genetic variants are transmitted within the families and if they are more frequently inherited by the individuals expressing the examined phenotype compared to unaffected participants. If

a SNP increases risk of disease, that SNP transmit from parent to offspring in high-risk families more often than in the general population. Genetic studies use families to enrich the study population with affected individuals who are more likely to have a genetic susceptibility to the condition of interest. Consequently, family-specific genetic loci associated with the condition may be identified. The changes in that specific locus might be specific for a familial form of the disease and not associated with the condition on a population level. The functional mechanisms dictated or influenced by that loci may still be affected in unrelated individuals but driven by other genes. If a variant identified in families with increased predisposition to a disease also could be identified in a case–control study based on the same disease, or in a cohort study, it would strengthen the possibility that the finding is a true association and that the genomic loci is related to disease development. In family studies healthy family members can be used as controls for the affected participants and by use of the pedigree information the risk for including genotyping errors is reduced. As the participants share both genetic background and environmental exposures some of the biases commonly seen in case-control studies can be minimized. It is still challenging to perform such studies as enough families and family members must participate to achieve the power needed to discover genetic changes truly associated with the disease in question. In this respect, most family studies are underpowered.

### *1.3.5 Genetic approaches to complex diseases*

Two main study approaches have been utilized to identify the molecular genetic basis of multifactorial diseases; the hypothesis-free positional cloning studies where no prior assumptions about the variants' involvement in disease should exist, and the hypothesis-driven genetic candidate studies (ref). In the first approach, researchers do not have a specific locus of interest, but perform a generalized study examining loci dispersed throughout the genome against one or more defined phenotypes to identify significant correlations, and possibly novel loci for further examination (ref). Examples of this type of research is the genomewide-association studies (GWAS) and WES. In contrast, a hypothesis driven genetic candidate approach would focus on

genetic regions based on existing knowledge of biological function, and a priory assumption about their involvement in the disease or phenotype of interest. One example of a hypothesis driven approach would be candidate SNP-based assays [148], as has been applied in Study III of this thesis project. Linkage studies usually include large pedigrees with several generations affected by the condition examined, and can be performed using either study approach, but the hypothesis-free approach is currently the preferred method of linkage studies on complex diseases [138].

Replication of associations found between genotype and phenotype in studies using other cohorts helps ensure that the results represent a credible association and not a chance or false positive finding. Further, if the genotype-phenotype connection can be replicated in different but related disease entities, it would increase the reliability of the link between that specific genetic locus and a specific biologic mechanism. It has proven to be a difficult task to replicate the results of genetic studies on complex genetic diseases.

In GWAS the position within the genome of variants significantly associated with disease is determined. Multiple GWAS have been performed for complex traits and over 14,000 common SNPs associated with disease are registered in online databases. Several of the SNP associations have been replicated in other studies, such as between hypertension and SNP rs 1378942 at the c-src tyrosine kinase gene [127, 194-196], implicating that the association is real. Still, how the identified loci influence disease development is currently uncertain for most findings. The statistical significant SNP-variants detected through this method are not necessarily the variants responsible for causing disease, rather, they may be in linkage equilibrium with the causative variant (Figure 9), but they do pinpoint regions of interest. Replication of significant findings is more commonly performed successfully by use of other genetic approaches, like candidate SNP arrays, based on few genes or SNPs. In one recent study 4 loci associated with blood pressure regulation, identified through GWAS, have been associated to the same phenotype in a different population [145].

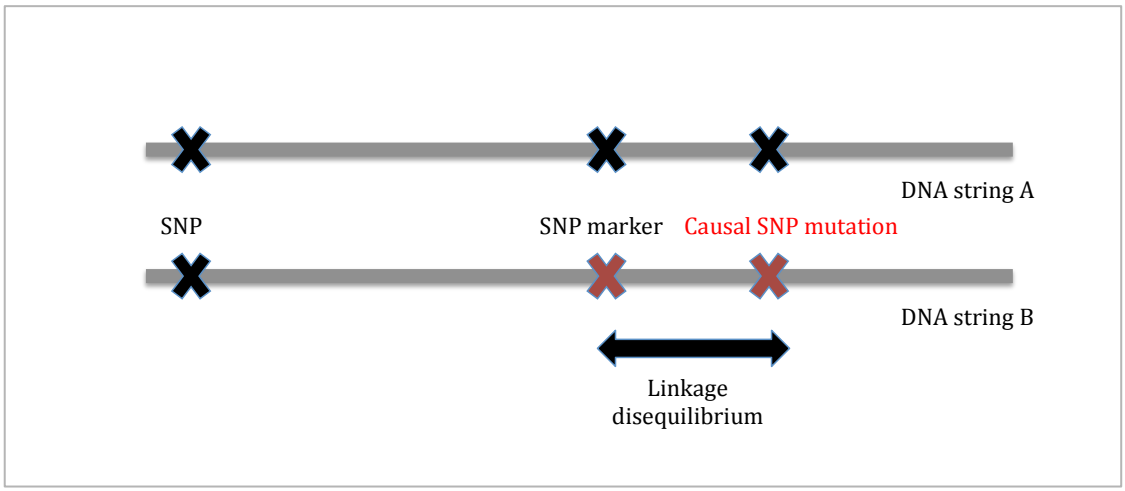


**Figure 9.** *Demonstrating how a SNP marker not neccessarily is the causative allele, but that it might identify a locus containing the causal variant by being in linkage disequilibrium with this SNP.*

Genes, genetic regions or SNPs become candidates when genetic or biological data establish a link between the variant and the trait of interest. Candidate studies can be performed in multiple versions, as exemplified here by the catechol-O-methyltransferase (COMT) gene, a gene with regulatory effects on steroid hormone metabolism and PE. Candidate studies can be based on one SNP in one genetic locus, such as the SNP rs4680 in COMT [146], a few SNPs allocated to the same gene, as in the study on the polymorphisms rs4680 and rs6269 that our group has previously performed on COMT gene activity in PE [147], or the candidates to study can be SNPs located to separate genes believed to interact in the condition examined (demonstrate epistasis), as has been investigated for PE between a haplotype on COMT consisting of four SNPs (rs6269, rs4633, rs4680, and rs4818) and one SNP, rs1801133, on the methylenetetrahydrofolate reductase (MTHFR) gene.

In complex diseases genetic candidate studies aim to detect small-to-moderate relative risks, as well as identifying associations between genetic loci with acknowledged functional impact and disease development. To increase probability of detecting true associations the selection of SNPs needs to be firmly based on plausible, consistent and established functionality of the locus, carefully selected phenotype and study design. Case-control populations are most commonly used in candidate studies [148].

Within the last decade methods to classify genetic variations throughout the whole human genome or parts of it has been developed. The term exome incorporates all exons of a genome, and exome sequencing/whole exome sequencing is performed to map all variants of the protein coding regions of the genome. The method is useful for discovery of highly penetrant variants and can be applied in cohorts based on unrelated individuals as well as in family studies. Exome sequencing has been successfully used for detection of causative variants in rare diseases, and it is increasingly used for interrogating variants associated with complex diseases as it encompass most variants in the examined regions and are relatively cost-effective compared to whole genome sequencing [149].

Whole genome sequencing provides genome-wide molecular detail revealing both small and large variation and allowing for determination of the exact sequence of bases of the DNA of an individual [150]. This method maps not only the exome, but also identifies normal variants and mutations in the non-coding regions of the genome. Thus, by this approach, regulatory regions of promoters can be analyzed, as these are acknowledged to direct gene expression and possibly phenotype expression. Further, data on intergenic regions (deserts) that might influence the expression of genes indirectly can be generated.

### *1.3.6 Genetic databases and bioinformatics*

The field of genomic research develops rapidly, and accordingly, increasingly large and complex genetic datasets are produced, comprising each level from DNA and RNA to protein synthesis and function. Huge efforts like the Human Genome Project, the International HapMap Project, and the 1000 Genomes Project have resulted in a deeper understanding of the human genome. In pace with the data generation and technology development, we see an increasing need for sophisticated computoral tools and pipelines. To make the escalating amount of data available and give researchers the possibility of mining study results for relevant findings information is organized in publicly available online databases like the NHGRI Catalog and the PheGenI, used in the third study of this thesis.

Genomic research is dependent on the growing area of bioinformatics which utilize computer science to develop statistical knowledge, databases and software tools for biological and medical research. Tools such as SOLAR, R and PLINK, used in this thesis (Study II and Study III), are required for researchers to be able to assemble, analyze, store and obtain information as the scientific methods become more complex and the sample numbers included in studies increase.

### *1.3.7 Phenotyping*

For rare genetic diseases a thorough phenotyping of the affected individuals and families have been the mainstay for directing genetic examination and diagnosing of the conditions (ref). Research on complex diseases focuses increasingly on the phenotypes being used, as the underlying etiology of conditions may differ with distinct subsets of one disease (see part 7.2.1 and 7.3.1). The lacking replication of genetic results is likely caused partly by the mix of different phenotypes included in the case-population examined and partly by differences in ethnicity, as well as underpowered studies [151]. Another reason for missing replication of results can be that studies are affected by population stratification, a confounding effect originate when alleles at a locus are distributed differently between cases and controls in a population even though the genotype is independent on disease state, resulting in spurious association findings [152]. In most genetic diseases and conditions, a range of presentations co-exist. The phenotypes of one disorder differ regarding the first presentation and age of onset, the severity as well as rapidity of development and treatment response. If the main trait is not studied separately according to similar presentations it will be very difficult to detect genetic variants associated with any presentation of the condition. One potential approach for the studies could be the utilization of phenotypic extremes to enhance genotypic differences in regard to disease state.

Exposures related to environmental influences that are to be included into analyses need to be selected with care, especially as the interplay between genes and environment is intricate and the environmental exposure is difficult to measure precisely (see also paragraph 1.3.4). Further, it will only be possible to know of, and

include, a minority of variables connected with environmental factors.

As the genetic influences on phenotype are complex as well as interacting with environmental influences, so are the biological and pathophysiological mechanisms. The partly overlapping genetic and biologic mechanisms of different disorders and the co-existence of interrelated disorders further complicate interrogation of possible genotype-phenotype association.

Since so many of the biological factors of disease development are extremely complex, and they by nature are complicating illumination of disorders, it is important that at least the phenotype is strictly defined. The benefits of including true cases (persons who really have the phenotype to be examined) and avoiding inclusion of false negative controls (individuals with undetected or subtle phenotype, or with overlapping conditions) encompass increased power to detect genetic associations. Additionally, a detailed knowledge of the phenotypes included as cases and controls, and the factors actively excluded will establish an enhanced setting for interpretation of the results. Strict phenotyping may not be the solution for all multifactorial conditions but the procedure might improve the comparison and replication of studies, and possibly aid in revelation of genetic loci or biologic functions with more general impact on the main disease phenotype**.**

### *1.3.8 Ethical considerations*

As all research involve ethical considerations, this also applies to genetic research. Projects related to genetic research comprise a collection of phenotypic information and biological samples, examination of parts of or the complete genome, data analysis, and dissemination of results. Looking back at how science has been developing the last decades, it is likely that researchers in the future will continue to want to apply new methods on previously collected samples, as well as test hypotheses differing from the original aims on the assembled material. Some important ethical considerations applying to genetic research are listed below:

- When researchers perform analyses on samples from individuals, the family

members of that person will automatically, but commonly unknowingly, be implicated in the study through shared genetic material. This point is of special importance when working with germ line mutations.

- A practical obstruction to a participant's right to withdraw from the study is constructed when study results have been published.
- Currently our understanding of the implications of genetic information is limited, and consequently the possibility of incidental findings that a researcher might have a moral or legal duty to disclose is genuinely low, but this may change rapidly when the clinical significance of genetic changes starts to be revealed.
- When the whole, or part, of the genome is identified, and the data additionally is made publicly available together with some phenotypic variables, proper anonymisation of data and participants poses a challenge.
- The concerns listed further constitute challenges to the requirements for informed consent of study participants.

The society of Genetics is consecutively implementing and revising their guidelines for continuous generation of knowledge and improved health in a sensible and responsible way [153]. As new analyses are developed, new questions will be posed, and as before, each person performing research has the responsibility to be aware of potential challenges and to implement the necessary framework to handle in the best possible way the issues as they arrive.

### 1.4 Biobanking

It has become increasingly customary to collect, contain and use human biologic material within a biobank and research setting. A biobank is a strictly regulated collection of biologic material, such as blood, tissue, or urine. In addition the biobank can contain the results of analysis performed on the collected material. A corresponding data source contains anamnestic and clinical information on the individuals that have contributed the material. The biologic materials are contained under regulated conditions and stabilized according to available facilities and type of

tissue collected to render samples secure, of high quality, and durable for use. Previously the tissues were mostly paraffin-fixated, but as availability of methods, software and information technologies have escalated with new possibilities of discovering disease patterns at molecular levels, tissues are preserved and stored differently, often utilizing freezing techniques.

There are many reasons why and how biobanks are established, not just that biologic material in biorepositories can be contained for use in analysis in the future. Utilization of larger study populations, more samples and data, increase the possibilities for detecting pathophysiological mechanisms of conditions, identify biologic markers (biomarkers) for disease, prognosis or response to therapy. Study opportunities are even more advantageous when the biorepository contains material collected from several stages of a disease in the same individual, several tissues per participant, or biologic material assembled before and after exposure or development of a condition in population-based cohorts. Learning about the heterogeneity of a group of individuals needing treatment for a disease, could help identify potential molecular targets for medication, relevant to specific phenotypes. Consequently better, more phenotypically focused clinical trials could be constructed. To assemble large, high-quality biobanks it is beneficial to have a stable population without too much migration, community-based health care systems and reliable infrastructure.

The classification of biobanks differs. One common division is the separation into population-based and disease-based biobanks [154]. *The population-based biobanks* are further categorized into longitudinal biobanks, biobanks based on families and twins, and isolated population biobanks. Longitudinal biobanks contain material from a general population in large cohorts, with samples and information from repeated examinations, which opens for opportunities to identify risk factors for disorders. Genetic and other molecular examinations of samples from individuals who contract a disease between collection periods could help differentiate between genetic alterations within the specter of normal and disease-related changes. In this type of biobank biological material is collected prospectively. Biobanks based on families or twin studies are used for comparison between genetic and environmental risk factors due to shared exposures and partly homogenicity of genetic material. Twins are commonly examined

as monozygotic versus heterozygotic twins. Biobanks based on isolated populations are collected from cohorts living separately for generations, either due to the geographical setting, or because of diverging from surrounding populations ethnically or by religion. Isolated populations are found in Iceland and within American groups like the Amish and Ashkenazi Jews. In biobanks of this classification the pedigrees can be tracked through generations. Genetic studies performed in isolated populations are powerful for identification of genetic risk profiles as much larger similarities can be found then in a generalized population, both in genes and environmental influences.

*Disease-based biobanks* are biobanks containing tissues/biological material collected as part of clinical intervention, such as diagnostic testing for which the collection of blood, urine or cerebrospinal fluid sampling are necessities or during an operation for cancer. The material contained in these biobanks will be from individuals with the similar disease phenotypes. As disease-based biobanks have been established both for common and rare diseases, researchers get the opportunity to identify etiologic factors of disease and different phenotypic presentations.

To get enough samples from a phenotype to perform useful studies, and obtain the statistical power to detect true associations between disease and molecular changes, it has become usual to run multicenter-studies, collecting data and biobank material from patients from several medical centers.

As biobanking has become more common a compulsory, national register for Norwegian biobanks was established in 2004. The aims of the Biobank Registry at the Norwegian Institute of Public Health were to identify and regulate existing biobanks, and to secure reliable and consistent use according to Norwegian law. In addition the register was established to increase opportunities for utilization of collected material in health related projects and help coordinate medical and scientific work. In 2011 Biobank Norway was established to develop a national biobank platform and improve infrastructure for biobanks and health registers. This consortium is further a part of the Nordic Biobank Network and Biobanking and of the Biomolecular Resources Research Infrastructure, both international endeavors encouraging and helping the establishment of biobank collaboration among the Nordic countries and in Europe,

respectively. It is difficult to identify all existing PE-related biobanks but according to rapported research the collected samples include maternal and fetal blood, maternal urine, placenta, decidua, umbilical cord, adipose tissue and amniotic fluid, and the biobanks described encompass all the abovementioned study designs.

### 1.5 The relationship between PE and aCVD

After a pregnancy the women who experience PE have a doubled risk of developing aCVD compared to individuals with unaffected pregnancies [7, 84]. Both PE and aCVD are genetic complex disorders where several genetic changes and environmental exposures interact to cause disease. Some environmental and life-style associated factors increase the risk for development of both disorders, such as obesity and diet [29, 35] (see also paragraph 1.1.2, Table 3 and paragraph 1.2.3, and Table 5), and underlying health conditions like chronic hypertension and diabetes mellitus type 2 increase the risk for development of either disease [29, 35] (see also paragraph 7.1.2, Table 3 and paragraph 1.2.3, Table 5).

Endothelial dysfunction and inflammation have been linked repeatedly to the etiology of both disorders [69, 129, 189] despite inconsistencies between studies (see also paragraph 1.1.3 and 1.2.4). Some factors contributing to these discrepancies are the lack of replication of findings in other cohorts or by other molecular methods, and difficulties in comparing studies due to application of different inclusion and exclusion criteria.

Each condition has demonstrated familiar aggregation and thereby established that genetic causative factors are involved [122]. Some genetic risk factors have been identified associated with each of the diseases [40, 125, 126, 127, 128, 129] (see also paragraph 1.1.2, Table 4 and Table 5, and 1.2.3.), and some genetic regions have separately been indicated as causative, both for PE and for risk factors for aCVD or developed aCVD, including variants in the vascular endothelial growth factor (VEGF), nitric oxide synthase (NOS), and methylenetetrahydrofolate (MTHFR).

The molecular studies performed are increasingly based on use of biobanks containing biologic specimens and associated information regarding health, environment and lifestyle, and a range of epidemiologic and molecular study designs, and combinations of these, has been applied to examine PE (Table 4 and Table 5, paragraph 7.1.2) and the co-morbidity of PE and aCVD [4, 5, 24, 40, 43 44,47,46, 69, 71,89].

Despite considerable research efforts knowledge on the pathophysiologic background of PE is inadequate and the overlap between PE and development of aCVD largely unexplained.

# 2. AIMS OF THE PROJECT

PE is a gestational disease affecting 2-8% of pregnancies world-wide and the second most common cause of maternal morbidity. The phenotypes of the disorder include a range of presentations, and may result in maternal multiorgan failure, preterm births, stillbirth and maternal death. Women with PE demonstrate co-morbidity with aCVD, a disorder of which one in three women dies globally. In both PE and aCVD complex interactions between genes as well as an interplay beteween genes and environmental factors lead to disease, classifying both conditions as complex genetic diseases.

## 2.1 Overall objective

The overall aim of this thesis is to increase the understanding of both the pathophysiology of PE and of the overlapping etiologic mechanisms between PE and aCVD. By illuminating disease specific and shared genetic risk factors of the disorders we wish to improve identification of women at highest risk for development of PE and preeclamptic women with increased risk for aCVD later in life, and hopefully help generate a basis for new screening and treatment options.

## 2.2 Specific objectives

**2.2.1** To examine the validity of PE registration in the MBRN in two cohorts: women with a familial predisposition to PE, and a case-control cohort based on women participating in the second HUNT Study (Study I). To identify how use of different criteria in different time periods influence the diagnostic validity we wanted to evaluate the diagnosis according to two disease definitions.

**2.2.2** To establish a genetic resource for studies of PE based on families with aggregation of this disorder, based on first-degree female relatives with valid registration of PE in the MBRN, the PE Family Biobank (Study II).

**2.2.3** To identify in the PE Family Biobank conditions related to PE development by phenotyping all individuals according to risk factors of aCVD like chronic hypertension and hypercholesterolemia, as well as established aCVD, and conditions related to PE development, such as underlying health conditions like inflammatory diseases and diabetes mellitus, and related conditions of pregnancy like gestational diabetes and IUGR/SGA (Study II).

**2.2.4** To investigate in the PE Family Biobank the heritability of PE phenotypes, phenotypes of aCVD risk factors and established aCVD, and other phenotypes related to development of PE, and to explore the shared heritability of phenotypes found to be heritable (Study II).

**2.2.5** To evaluate the contribution to PE development of identified candidate risk factors for a) chronic hypertension and b) inflammation by a SNP array approach in the PE Family Biobank, and in the HUNT cohort, an unrelated PE case-control population sample from the HUNT Study (Study III).

**2.2.6** To replicate the associations between chronic hypertension and SNPs identified previously in genome-wide association studies, and to identify novel associations between genetic regions involved in inflammation and chronic hypertension. The associations will be examined using a candidate SNP array in PE two cohorts, the PE Family Biobank and the HUNT cohort (Study III).

# 3. MATERIALS AND METHODS

## 3.1 Materials

### 3.1.1 The Medical Birth Registry of Norway (MBRN)

The MBRN was established in 1967 as the first population based medical birth registry in the world. Since the establishment all deliveries from 16 weeks of gestation have been registered in the MBRN (over 2.7 million) based on compulsory notification. In 2001 information on miscarriages and abortions after the 12th week of gestation were included additionally. After delivery, data from antenatal forms and information from the obstetric records and neonatal ward (where appropriate) are transferred to the MBRN [155]. The MBRN forms have been electronically transmitted since 2006. A generalized printed antenatal form is carried by women throughout their pregnancy, and a copy of this document is archived in the hospital where the delivery took place.

The registered information includes identification of the parents by their unique 11-digit national identification number as well as demographic information on the women, paternal age, data on maternal smoking (voluntary), and if the pregnancy was induced through in vitro fertilization. Selected maternal health conditions that possibly influence pregnancy outcome, such as heart disease, epilepsy, diabetes, asthma, rheumatoid arthritis and use of medication and dietary supplements are registered. Medical complications arising throughout the pregnancy, as documented in the women's records, are routinely transferred to the medical birth notification form by health personnel working in the maternity wards. The conditions registered are infections (rubella, venereal disease), vaginal bleeding, anemia, Rhesus-immunization, thrombosis, gestational diabetes and gestational hypertensive diseases. From the establishment of the MBRN until 1998 the birth notification form was unaltered, requiring information on complications during a pregnancy to be noted as free text. In the registry women are designated the diagnosis PE if the free text-documentation include either "pre-eclampsia" or a combination of "hypertension" and "proteinuria". As the birth notification form was changed in 1998 to include check boxes, the MBRN assigned the diagnosis PE to women where health personnel had checked either of the

following alternatives: “Preeclampsia, mild”; “Preeclampsia, severe” and “Preeclampsia, before 34 weeks”. The form submitted to the MBRN also includes circumstances and conditions directly associated with the birth, such as induction of birth, fetal position (cephalic vs breech), delivery by cesarian section/ventouse/forceps, type of anesthesia given, neonatal outcomes, and data on the placenta and amniotic fluid.

The MBRN defines PE as an increase in blood pressure to at least 140 systolic or 90 mm Hg diastolic (or an increase in blood pressure ≥15 mm Hg from the level measured before 20th gestational week), combined with proteinuria (protein excretion of ≥ 0.3 g per 24 hours or ≥1 + on dip-stick) (Table 1, Study I).

### 3.1.2 The Preeclampsia Family Biobank (PE Family Biobank)

The PE Family Biobank has been founded to address the genetic basis of PE and related conditions by use of phenotypic and genetic information from Norwegian families with an increased susceptibility to PE [156] (Study II). The establishment of the Family PE Biobank was initiated in 2002 and completed in 2012, and is based on collaboration between the Norwegian University of Science and Technology (NTNU) and Haukeland University Hospital in Norway.

Families with aggregation of the disease PE were defined as families containing at least two first-degree female relatives (indexes) where each woman had experienced at least one preeclamptic pregnancy. Due to study design the women had to have given birth in one or more of the five participating hospitals in Mid- and West-Norway: Haukeland University Hospital in Bergen, Stavanger University Hospital in Stavanger, St. Olavs Hospital Trondheim University Hospital in Trondheim, Levanger Hospital in Levanger and Namsos Hospital in Namsos (Supplementary Figure 1, Study II). The pairs of index cases were identified by the MBRN using registered pregnancy/delivery information and diagnoses from the period 1967 - 2005, and the individual identification numbers to determine relationship between individuals and select pairs of either two sisters with the same mother or a mother and her daughter where both individuals have had PE.

The PE diagnosis of the identified 1003 women with a total of 1201 pregnancies was validated against information from hospital records by physicians working in the departments of obstetrics and gynecology in the participating hospitals. Diagnostic criteria for PE were a minimum of two blood pressure measurements ≥ 140/90 mmHg combined with proteinuria (≥0.3 g/24 hrs or a dipstick reading of +1 or more) occurring after 20 weeks of pregnancy.

After receiving the results of the validation the staff at the MBRN performed a quality control where all women with non-confirmed diagnoses, deceased individuals and persons lacking contact information were excluded from the list. Only those index-pairs where both women had their diagnosis confirmed and were approachable were included in the next step of the study.

The information on the 207 pairs totaling 426 women who still fulfilled the study criteria, received written invitations to the study from the hospital where they had given birth. Information on the main purpose of the study, on how the invitees had been identified, and the official approvals of the study was given. Women who wished to participate in the study met with the contact person of her delivery department. Each participating index woman was asked to invite other family members above age 17 to participate, including family members of both sexes, and female relatives with histories of normal or abnormal pregnancies and births. All potential participants were informed about why the family study design was chosen, why partners and family members with healthy pregnancies were invited as well as the diagnosed women, and what the participation entailed in form of information and samples requested. The process of informed consent also included informing the participants that they would not receive results except as data published in a de-identified form for the whole or parts of the cohort. The information on estimated length of the study, the later establishment of a biobank based on the data and samples collected, the possible uses of the collected data and samples in other research projects, as well as in studies linking separate registries and biobanks and with collaborators abroad were included in both the written information, the consent form and in the process of informed

consent. The participants were informed about the possibility of withdrawing consent. All invited individuals were given the opportunity to discuss any concerns they should have about the study and participation. The study language (oral and written) was Norwegian, so only individuals able to communicate in this language were included. Only the medical personnel involved with the study at the respective delivery departments were allowed to contact the index women and carry out the consultations needed. During the period 2009-2011 medical doctors and medical students performed a standardized interview of all included participants and peripheral blood samples were collected. Those individuals who signed the consent form were interviewed concerning their own and their family members' pregnancies as well as current health and family maps were drawn for each participating family. The data collected from the participants of the study can be found in SDC5 Table 3 (Study II).

All data were de-identified and quality controlled at Haukeland University Hospital and a summary pedigree was constructed for each family in Cyrillic 2.1.3 (CyrillicSoftware, Oxfordshire, UK). All clinical data are stored at an internal secure site at the Western Norway Regional Health Authority database. Blood samples were collected from 98% (486/496) of the included family members in serum separator tubes (SST), ethylenediaminetetraacetic acid (EDTA) tubes, and Tempus blood RNA stabilizing tubes (Applied Biosystems, Foster City, CA, USA). Serum, plasma and white blood cells were separated by centrifugation. The biologic materials are labeled with a unique barcode and stored at $-80^0$C at the facilities of the HUNT Research Centre and Biobank, NTNU, Levanger, Norway. Figure 2 (from Paper II) illustrates the step-wise process of establishing in the PE Family Biobank.

When the collection study was finished in January 2012, the PE Family Biobank was formally established, originally containing 138 families with a total of 496 individuals of North-European ancestry. Numbers of participants used in the studies included in this thesis differ due to withdrawals and discrepancy in data for one family in the years after the establishment, as well as availability of DNA. Descriptive characteristics for the cohort are given in Table xyx (from Study II). In Study II of this thesis all

individuals in the PE Family Biobank with phenotypic information (n=493) were included. The cohort selected for the Study III in this thesis contains all those individuals in the PE Family Biobank who both had DNA and reliable pheotypic information available. Descriptive information on the cohort used in the third study can be found in Table 1 and Table 2 (Study III).

### 3.1.3 The HUNT Study

The HUNT Study consists of three main separate multipurpose health surveys (HUNT1-3) and represents the total population above age 19 of the rural county of Nord-Trøndelag in Norway [157].

The first part, HUNT1, took place during the years 1984-1986. The second part of the study, HUNT2, was performed in 1995-1997, and the third survey, HUNT3, in 2006-2008. All county citizens aged 20 years or older (n=92,566) living in Nord-Trøndelag County were invited. In HUNT1 data were gathered through structured questionnaires and physical examinations, while in the other two main parts of the HUNT study, biological materials were sampled additionally [157, 158]. As the information was collected in a similar fashion throughout the HUNT study, longitudinal data are available on persons who took part in two or more of the surveys. Of the included individuals, about 27,000 took part in all HUNT surveys, whereas 46,000 persons participated in both HUNT1 and HUNT2, and 37,000 in both HUNT2 and HUNT3.

The assembled information is contained in the HUNT Databank at the HUNT Research Centre and Biobank, where it is linked to the unique national 11-digit identification numbers, and available in de-identified form to researchers upon applications following strictly defined requirements. The biologic materials collected in HUNT1-3 are labeled with unique barcodes and stored at tissue-specific recommended temperatures (- 20 °C, - 80 °C or - 196 °C) at the facilities of the HUNT Research Centre and Biobank, NTNU, Levanger, Norway.

For the HUNT cohort included in this thesis (hereafter called HUNT) we have selected a subset of women participating in the HUNT Study. By means of the national

identifiers, data from the HUNT Study and the MBRN has been linked to recognize all women who participated in the second and/or third survey of the HUNT Study and who gave birth during the period 1967-2008. Selected as cases in the cohort were all women identified in the MBRN with at least one singleton preeclamptic pregnancy (n=1,179), and a compatible number of women without any pregnancy affected by PE were included as controls (n=1,179). All relevant information available from all of the surveys the identified women participated in was requested as study variables, including data from HUNT1 where applicable. Table 1 and 2 (Study III) contain descriptive characteristics of the case-control cohort as well as those available on all female participants in HUNT1-3.

For the first study in this thesis a subset of the HUNT case-control cohort was used to examine the validity of the PE diagnosis for the included women (Paper I). The 992 women in this subset had participated in HUNT2 and delivered at one of the two regional hospitals in North-Trøndelag (Levanger Hospital and Namsos Hospital). All were registered with the diagnosis PE in the MBRN.

For the third study in this thesis we included all women in the HUNT case - control cohort (n=2358). The ensuing selective exclusion of cases and controls due to clinical information is demonstrated in Supplementary Table 3 and 4 (Study III, manuscript). Some individuals participated in both the HUNT cohort and the PE Family Biobank (n=13), and these were removed from the HUNT cohort. The cases with invalid diagnosis according to our results in the first study of this thesis (n=113) were identified and removed from the cohort. Women identified as controls with no preeclamptic pregnancies by MBRN but who in any of their pregnancies had registered conditions in the MBRN related to PE, such as gestational or chronic hypertension or kidney disease, were identified and removed from the cohort (n=94). The exclusion criteria comprised gestational hypertension, chronic kidney disease, maternal heart failure, and if any pregnancy had ended in a stillbirth or placental abruption. After further exclusion of 149 samples who failed all assays in the genotyping and those individuals with genotyping call rates below the required 98%, the HUNT case-control cohort in the genotype-phenotype analysis of Study III

consisted of 1822 women (1006 cases and 816 controls, sample call rates > 95%). In this project we used well-defined clinical and laboratory information from the HUNT databank and data on gestational health and outcomes from the MBRN to identify phenotypes and covariates related to PE. Various birth-related outcomes were included per pregnancy for each woman to be able to identify associations between phenotype and genotype while adjusting for covariates such as pregnancy-specific birth weight, birth year and gender. The information regarding total obstetric history and outcomes (PE ever, SGA ever) were treated as dichotomous variables. Longitudinal measures available from the HUNT Databank, such as age at participation and measurements of weight, blood pressure and cholesterol levels were averaged and analyzed as continuous variables. Participants were classified as smokers if they in the HUNT Study questionnaires had expressed to be or have been smoking. Otherwise they were treated as non-smokers in our analyses.

### 3.1.4 The genetic databases utilized

For selection of SNPs to be included in the third study of this thesis two online genetic databases were used. The information contained in the databases are partly overlapping, with data from The National Human Genome Research Institute (NHGRI) Catalog of Published Genome-Wide Association Studies (NHGRI Catalog) included in the Phenotype–Genotype Integrator (PheGenI).

The NHGRI Catalog is an online resource published by NHGRI, a division of the National Institutes of Health, USA. This is an online database of SNP-phenotype associations from published genome-wide association studies (GWAS) identified and regularly updated through weekly searches in the PubMed database and daily reports from NIH, as well as intermittent comparisons with the content in other, similar databases. The NHGRI Catalog is curated and publicly available at no cost to the users (http://www.genome.gov/gwastudies) [159].

PheGenI is a publicly available online repository that collects, integrates and presents information extracted from several genetic databases at the National Institutes of Health, USA, according to SNP rs-numbers, genes and / or chromosomal location

[160]. The assembled information is based on published and unpublished genotype-phenotype associations identified through GWAS recorded in the NHGRI Catalog and the database of Genotype and Phenotype (dbGaP) [161]. PheGenI is curated and regularly updated, and part of the information can be freely obtained directly from the resource while part of it (e.g. de-identified phenotypic data from the separate GWAS) can be made use of by researchers on application and on certain conditions. According to the latest publication on PheGenI the resource contained in March 2013 data on 66,063 phenotype-genotype associations in total (54,282 from dbGaP, 11,781 from the NHGRI Catalog [160].

### 3.1.5 Ethical approvals

The studies performed on material from the Preeclampsia Family Study and the PE Family Biobank have obtained approvals from the Regional Committees for Medical Research Ethics (Regional komité for medisinsk og helsefaglig forskningsetikk Midt-Norge ref. 041/03, 157/03, 2012/408, 2012/1876) and the National Data Inspectorate (03/01372-22/CGN).

For the studies involving the North-Trøndelag Health Study cohort the required approvals from the Regional Committees for Medical Research Ethics (Regional komité for medisinsk og helsefaglig forskningsetikk Midt-Norge ref. 041/03, 157/03 and 2012/1907) and from the HUNT Research Centre (13/4894) have been obtained.

Informed consent was obtained from all participants when enrolled in the PE Family Biobank, and for all participants in the HUNT Study when included in the separate surveys.

## 3.2 Methods

### 3.2.1 Diagnostic criteria

For our first study we evaluated the diagnostic validity of PE. For this study we have used the term *diagnostic validity*, which is usually applied with regard to categorical variables and identifies the proportion of true cases (how many of the individuals

assigned a diagnosis that have the disease) and true controls (how many of the persons without the diagnosis do not have the disorder) [162], in an examined population.

Two different diagnostic criteria were utilized, a broader and a stricter definition. Both definitions comprise hypertension (blood pressure measured to at least 140 systolic or 90 mm Hg diastolic) combined with proteinuria (protein excretion of ≥ 0.3 g per 24 hours or ≥1 + on dip-stick) developing after gestational week 20. In the broader criteria used for evaluation of PE in a subset of the HUNT cohort one measurement of hypertension with proteinuria was required, while the restricted criteria incorporated reproducible (measured at least twice minimum 6 hours apart) hypertension with proteinuria in pairs of female first-degree relatives who were selected as potential participants for a family study.

The broader criteria were the contemporary criteria set by the MBRN, and until 2006 the definition was concordant with the clinically applied diagnostic criteria in Norway [163]. The other, stricter set of criteria corresponds with current national and international guidelines at the time when the validation of the closely related women took place [164].

The gold standard diagnosis of PE in the study was based on information from antenatal forms and medical records. The diagnosis was confirmed or rejected by a medical doctor working in the department where the birth had taken place (the paired first-degree relatives), or two study nurses, trained by one of the study investigators (the subset of women from HUNT).

In the two other studies incorporated in the thesis we have continued to use the stricter criteria for the PE Family Biobank, where the hospital records were available to us during the study. For the diagnostic criteria applied on the HUNT cohort accessible information on the diagnosis of PE consisted of the data registered in the MBRN. When selecting and using the HUNT case-control cohort in the third study we therefore employed the diagnosis as set by the MBRN.

### 3.2.2 Phenotypes

For the PE Family Biobank participants several phenotypes of PE were examined in the analyses included in the second study of this thesis (Table 1, Study II). The presence / absence of PE, moderate / severe PE, and a high-severity phenotype, which incorporated the conditions eclampsia (seizures) and imminent eclampsia, HELLP (multiorgan failure) and pre-defined symptoms (severe headache, upper right quadrant pain, visual disturbances), were treated as dichotomous variables. The different presentations were defined as severe PE and PE (Table 1, Study II). We did subset the disease according to gestational age, firstly as early or late disorder with cut-off at gestational week 34, but later according to four subsets of gestational age (<30 weeks, 30 to 33 weeks+ 6 days, 34 - 37 weeks, and 37 weeks and longer gestations, data not shown).

For the HUNT cohort in the third study the PE phenotype was defined as a dichotomous variable. Women were defined as cases if they had the PE diagnosis in the MBRN, or controls if none of their pregnancies were registered in the MBRN with hypertensive gestational disorders.

In this thesis we have used the condition small-for-gestational age (SGA) as proxy for intrauterine growth retardation (IUGR). It was not possible with the information available to us to identify the fetuses and neonates with inappropriately low fetal growth, as this diagnosis requires data from serial ultrasound measurements during the pregnancy. The newborn children were defined as SGA in this thesis if their birth weights were below the 5th centile according to Norwegian unisex reference ranges for gestational age [165]. To estimate gestational age at birth we used the due date and date of birth as given in the MBRN for the HUNT cohort and in the information collected in the PE Family Biobank. If available, estimates of due date according to an ultrasound scan performed before 20 weeks of gestation was used in the study, otherwise we have used the date as calculated from time of the last menstrual period.

The other conditions and phenotypes that were examined in the second study of this thesis is listed in Table 1, Study II. The participants were only classified with a disease phenotype if they stated that they had received relevant treatment for the disorder.

To estimate how severe expression of PE a parous woman had experienced, a severity score was constructed based on definitions of severity in the Norwegian and international guidelines [164]. Each phenotypic presentation of PE was assigned a value between zero (absence of disease) and five (high-severity phenotype as defined above). No phenotype was allotted the value of three. The disease severity score for a woman is based on the overall severity summed and multiplied with her number of affected births divided by the total number of her births (Supplementary SDC 4 Table 2, Study II).

The severity of aCVD was estimated for each individual. The aCVD phenotypes were designated a stepwise increasing value according to absence of disease, presence of aCVD risk factors, and information on established aCVD (Supplementary SDC 5 Table 2, Paper II). A severity score of aCVD was calculated for a person by addition of the values for each aCVD trait observed for that person. Waist circumference was measured by circling the participant's trunk with a tape measure held parallel with the floor at the point halfway between the lower ribs and the hipbone.

The phenotype of hypertension was examined in the HUNT cohort for the third study. The phenotype was annotated women who fulfilled at least one of the following criteria: 1) averaged blood pressure measurements of at least 140 systolic or 90 mm Hg diastolic, and 2) antihypertensive medication used previously or at participation.

### 3.2.3 Heritability (H2r)

H2r estimates the relative contribution of genetic factors to the defined phenotype expressed in the cohort. Any observed differences in a phenotype in a population could be due to genetic and/or environmental influences. Heritability estimates indicate to which degree the genetic material of the parents influences the phenotypic presentation in the children, but as the estimates are made for a specific population at a

defined time period, and incorporating an environmental factor in the calculations, this tool cannot be used to identify heritable traits on an individual level or define that a phenotype is genetic or not. The heritability (H2r) of a trait is dependent on the population used, varying according to the frequencies of given genotypes and the environment surrounding and impacting on the population examined.

The examination of phenotypic correlation (ρP) estimates the overlap between two phenotypes and the additive genetic component, and thereby incorporates how much of the overlap is caused by a genotype alone (the genetic correlation) and to which degree environmental factors (environmental correlation) influence the overlap in phenotype.

### 3.2.4 Candidate SPN-array

The third study of this thesis is a candidate SNP study. The decision of which SNPs to include in the study was based on information in the NHGRI Catalog and the PheGenI resource. The SNPs were selected according to one of two categories: either they were found to be associated with blood pressure regulation in a previously performed GWAS, or they were located to one of our selected genes and identified in a GWAS to be associated with a physical phenotype. All SNPs associated with blood pressure regulation at p-values $\leq 10^{-6}$ in published studies in the NHGRI Catalog by February 25th 2013 were included in our study. The selection of other candidate genes was based on focus areas of our research group, and include the following: TLR1, TLR4, TLR5, NOD1, NOD2, AGER, NLRP3, NLRP12, VEGF-A, RGS5. For five additional genes of interest no GWAS data could be found in PheGenI (TLR2, TLR3, TLR9, NLRC4, VEGF-B) and these were therefore not selected to the study. These genes have demonstrated a potential role in the biologic processes of inflammation and/or angiogenesis, but except for VEGF-A associations, few connections have been described between the genes and PE development. By use of PheGenI we retrieved information on all GWAS (published and unpublished) that had identified associations at p-values $\leq 10^{-6}$ for any SNP located to these genes.

A total of 250 SNPs were identified in total, whereof 61 were in duplicate. By use of the software tool SNAP for SNP Annotation and Proxy Search [166] we could identify 74 of the 189 unique SNPs to be in linkage disequilibrium (LD) with another SNP in the dataset ($r^2$=< 0.8, HapMap release 21, CEU population). Based on the output from SNAP the second SNP listed in each LD pair was excluded, leaving 115 SNPs for genotyping (Supplementary Table 1, Paper III) by using Sequenom MassARRAY system (Sequenom iPLEXassay, Australian Genome Research Facility Ltd., St Lucia, Queensland).

Of the 115 SNPs 13 had poor designability. Where SNPs failed designability they were replaced with at least one surrogate SNP (LD = 1, Hapmap release 21). One SNP on the original list failed and was not replaced (rs319690). Due to multiple surrogate SNPs for some on the original list the end result was 122 successfully genotyped SNPs.

### **3.2.5** Statistical analyses

In the study of the validity of the PE diagnosis the statistical analyses were performed with the Statistical Package for Social Sciences (SPSS) for Windows, release 17 (SPSSInc, Chicago, IL) and STATA Statistical Software, Release 10.0 (StataCorp LP, College Station, TX). To establish the positive predictive value (PPV) of PE we calculated how many of the pregnancies in the MBRN that were registered with the diagnosis which also fulfilled the set diagnostic criteria according to information in the medical records. As the study did not include a group of pregnancies registered without PE in the MBRN that could be used as controls we were unable to calculate specificity, sensitivity, or negative predictive value of the registered diagnosis. Time trends in PPV were measured by relative risk models, categorizing year of birth into groups, each including four consecutive years. Time was modeled as a categorical parameter using as reference the first time period (1967-70).

In the second and third study of this thesis the SOLAR (Sequential Oligogenic Linkage Analysis Routines) computer package version 7.2.0 was utilized for analyses. SOLAR is developed for genetic analysis of common disease and incorporates general

pedigree variance components and identity-by-descent probability estimation methods in order to identify genetic loci. Variance component linkage methods including linkage, SNP association, and covariate screening, can be analyzed in pedigrees of arbitrary size and complexity. Additional possibilities for evaluation of multiple quantitative traits and / or discrete traits which involve multiple loci (oligogenic analysis), and models for investigation of dominance effects, household effects and environmental interactions are incorporated in the software package. The analyses for the second and third studies of this thesis were performed within iVEC, a high-performance national (Australian) computing facility for researchers. The supercomputing center is supported by the Western Australian Government and located in Perth.

To perform analyses in the second study family data mapped in the Cyrillic 2.1.3 pedigree software (CyrillicSoftware, Oxfordshire, UK) and Microsoft Office 2011 Excel v14.0 (Microsoft Corporation, 2011) from the PE Family Biobank was reformatted for use in SOLAR. By analyzing these files with software incorporated in SOLAR we could estimate phenotypic heritability of both discrete and quantitative traits and the statistical significance of results. Only those phenotypes that were heritable at $p<0.05$ were explored further.

The covariates participant age, age$^2$, participant sex, age-sex interactions were included in all heritability (H2r) analyses, as well as a weighting factor for PE. All families and participants were specifically selected due to their association with increased occurrence of PE and not invited randomly from a general population. In selection of a cohort on such a basis a potential bias is formed, and consequently a weighting factor was calculated for each participant and included as covariate in the analyses to counteract the selection bias. In the analyses of phenotypic correlation, the covariates significantly associated with the heritability of each examined trait were included. To investigate phenotypic overlap between the included conditions a likelihood-ratio test incorporating non-independence among related individuals was applied to the significantly heritable phenotypes, which were cross-examined in pairs [167, 168].

In a cohort consisting of heterogenous families a phenotype may be irregularly distributed and demonstrate kurtosis [169], thereby influencing results from variance components models. To counteract effects of kurtosis in the study on phenotypic H2r we used inverse Gaussian transformation to normalize the traits with high residual kurtosis [170]. The phenotypes that despite these calibrations retained high residual kurtosis were excluded from further analysis. Dichotomous phenotypic variables were analyzed by employment of a liability threshold model using probit-regression for the mean effect component. For the residual additive genetic component of variance a standard random effects variance component model was adopted [167, 168].

For the candidate SNP study analyses were performed with software packages SOLAR version 7.2.0 (ref), PLINK version 1.07 (14) and R version 2.15.2 (http://www.r-project.org/). Quality controls for the two cohorts were first performed jointly in R software and then separately for the PE Family Biobank in SOLAR and the HUNT cohort in R. For statistical significance we decided to use strict Bonferroni-corrected p-values ($p<0.00042$ when adjusted for 119 SNPs) as cut-off. Statistical tests were two-sided.

Associations between genotype and phenotype in the HUNT cohort were analyzed under an additive genetic model to estimate effect of phenotype by number of minor alleles. Analyses were performed using a multiple logistic regression model incorporated in the PLINK software. Potential covariates were tested against phenotypes in a generalized linear model in R and only variables significantly associated with a phenotype were included as covariates for that trait.

In the third study the family data, which were reformatted and used for the heritability analyses in the second study, were applied in a generalized linear model. Variance component methods in SOLAR were employed to identify significant measured genotype associations between SNPs and the phenotypes PE and hypertension in the PE Family Biobank. Only the covariates associated with the phenotype at $p<0.05$ in the estimations of H2r in the second study were included in the genotype-phenotype analysis. In the estimation of genotype-phenotype associations a null model was used which reveals no disparity in phenotype by SNP genotype. The association analysis

was performed utilizing a likelihood ratio test that examines the additive genetic model against the null model.

# 4. SUMMARY OF RESULTS

## 4.1 Study I

We validated the diagnosis of PE in MBRN in two study groups, each group according to one of two pre-determined sets of diagnostic criteria. In the cohort based on unrelated HUNT2 participants with PE, we could evaluate 97.4% of pregnancies identified through the MBRN against identified hospital records. When applying the broader diagnostic criteria for PE on this cohort a total of 88.3% could be confirmed. The causes of the invalid diagnoses could not be investigated in depth for this part of the study.

For the other cohort, consisting of pairs of first-degree relatives with registered PE, the medical records could be located and examined for 98% of the births notified as affected by PE in the MBRN. According to the restricted diagnostic criteria used, 63.6% of PE diagnoses registered in the MBRN for this cohort of related women could be confirmed. In the 414 pregnancies where the diagnosis in the MBRN did not fulfill the criteria the reasons were absent documentation of proteinuria, missing data on hypertension, and no indication in the records of the woman having experienced PE in any pregnancy (72.2%, 5.8%, and 22.0% respectively).

Time trends of the diagnostic validity were examined for both cohorts using as reference the positive predictive value (PPV) identified in the first time period (1967-70). The cohort based on HUNT2 had high PPVs for all time periods examined, but the proportion of pregnancies with valid diagnosis decreased significantly in the two time periods between 1975 and 1982. For the time categories extending from 1983 to 2002 the results were comparable to the reference value for the cohort based on HUNT2. In the cohort consisting of paired first-degree female relatives, the estimated PPV for the period 1983-86 demonstrates a significant decrease compared to the reference. In this cohort of paired women the proportion of valid diagnoses increased significantly from 1986, and in the two last periods examined the PPV was 81% (1999-2002) and 82% (2003-2005).

## 4.2 Study II

We established the PE Family Biobank, a biobank based on families with increased occurrence of PE, and estimated the heritability (H2r) and phenotypic correlation of several PE-related phenotypes.

The PE Family Biobank included 496 individuals belonging to 138 pedigrees. Average age in the total cohort was 46.8 years, ranging from 18 to 87. Of the invited first degree female relatives 30 mother-daughter and 72 sister-sister pairs (102 pairs in total) participated. In 35 of the other identified families only one of the invited women with PE accepted and participated, and often bringing relatives into the study, but not the affected mother/sister/daughter. Several of the participating female relatives of the invitees reported during interviews to have had PE. Among the female participants who had given birth 75.7% (263/338) had PE. About 36% of the affected women had PE with severe hypertension and 26% had early-onset PE. The disease affected the first pregnancy in 81% of the preeclamptic women, and for 50% of these women it was their only preeclamptic pregnancy. The prevalence of giving birth to children small for gestational age (SGA) was 35.7% in women with PE and 22.4% in the women without a PE diagnosis.

Of the 30 phenotypes examined for the cohort 11 were heritable. Heritability (H2r) estimated for PE as a dichotomous trait (presence versus absence of disease) was 60%. When the offspring contribution to PE was examined in the phenotype "having been born in a preeclamptic pregnancy", H2r was 25%. To deliver children small for gestational age (SGA) were heritable in the cohort, demonstrating H2r 40%. The genetic effects on the variance of pulmonary disease in the cohort was estimated as H2r 91%, and for chronic hypertension and diabetes mellitus type 2 the estimated H2r were 57% and 57%, respectively.

For traits with high residual kurtosis in the primary analysis trait normalization was performed before H2r estimations. The following phenotypes demonstrated significant H2r and normal residuals after normalization: disease severity of PE (H2r 15.1%),

body mass index (BMI) (H2r 60%), severity of aCVD (H2r 31%) and waist circumference (H2r 58%).

In the examination of overlap between phenotypes two traits were found to be phenotypically correlated with PE status: how many times a woman gave birth (19.5%), and giving birth to a SGA child (26.2%). The trait pulmonary disease was found to be positively correlated with BMI (17.2%) and the severity of aCVD presentation (18.8%), and negatively correlated to delivery of children SGA (25.4%).

## 4.3 Study III

We performed an investigation of genetic associations between PE and either genes regulating blood pressure or selected inflammatory/angiogenetic genes in a candidate SNP array. The top ten association results for the phenotypes PE and chronic hypertension Supplementary Tables 3 and 4 (Study III).

### 4.3.1 The HUNT cohort

*PE phenotype*

Of the 119 SNPs passing the quality control (QC) after genotyping, one SNP, rs17367504, was found associated with PE in the HUNT cohort (p=3.52e-05, OR 0.65). The SNP was not associated with chronic hypertension in the cohort (p=0.71). The SNP is located in an intronic region of the methylenetetrahydrofolate reductase (MTHFR) gene and was included in our study due to previous associations with blood pressure regulation.

*Chronic hypertension phenotype*

For the phenotype chronic hypertension we could not identify any significant associations in the HUNT cohort. The highest ranked SNP associated with chronic hypertension in the cohort was rs1572299 (p=0.0063), located in an intergenic region and mapped to a gene involved in inflammatory mechanisms, TLR4. When examined against the PE phenotype this SNP had a p-value of 0.30.

### 4.3.2 The PE Family Biobank

*PE phenotypes*

For the maternal phenotype "PE presence" in the PE Family Biobank no SNP association reached the Bonferroni-adjusted significance treshold. The highest ranked SNP was rs633185 (p=0.0046), located to an intron of the Rho GTPase activating protein 42 (ARHGAP42) gene, selected as it has been associated with blood pressure regulation in GWAS cohorts. The association with chronic hypertension was non-significant in the PE Family Biobank (p=0.40).

In the PE Family Biobank cohort we identified one intergenic SNP, rs1327235, near gene JAG1, as marginally associated with both a protective effect on being born in a preeclamptic pregnancy (p=4.7x10-4, OR 0.81) and with an increased risk for development of chronic hypertension (p=3.4x10-3, OR 1.34). The SNP was not associated with maternal PE (p=0.88).

*Chronic hypertension phenotype*

No SNPs were significantly associated to the phenotype chronic hypertension in the PE Family Biobank. The highest ranked SNP associated with chronic hypertension was rs1327235, the SNP marginally associated with the offspring's contribution to PE, described above.

# 5. METHODOLOGICAL CONSIDERATIONS AND DISCUSSION OF RESULTS

## 5.1 Diagnostic validity

New knowledge on PE and other multifactorial diseases is increasingly generated by use of register data and biobanks collecting data and biologic material over long time periods. If the disorders registered have an unknown diagnostic validity the bias introduced by including false positive cases and false negative controls in studies can severely skew the results [171]. This may either lead to dilution of biologically true findings ending with no significant research results, or to spurious identification of associations, indicating inappropriate targets for potential clinical intervention. As a consequence interventions may neither identify the intended target population for the specific treatment, nor work as efficiently as expected [171].

In the first study of this thesis we examined the registered diagnosis of PE in the MBRN for the time period 1967-2005 against information in medical records. The broader criteria was applied on the hospital records of a group of women from the HUNT cohort, resulting in 88.3% of the diagnoses confirmed as valid (Study I).

Even though several studies have been based on registered preeclampsia diagnoses in the MBRN, the diagnostic validity of PE in the registry had not been validated previously. The validity of the PE diagnosis had been estimated in other national birth registers and for hospital registers in different countries with PPVs ranging from 54% to 93%, and with application of different diagnostic criteria and gold standards in the various studies [172-175].

Since the publication of Study I another study has examined the validity of the PE diagnosis in the MBRN for the recent ten-year period (1999-2010) in a country-wide cohort, resulting in a positive predictive value (PPV) of 83.9 [176], confirming our results for the last two time periods according to both diagnostic criteria applied in Study I. That study further included a control group to estimate sensistivity, specificity and negative predictive value of the diagnosis, thereby complementing our study, as

we had not been able to examine these parameters of the diagnosis in our cohorts, lacking a study group of women with apparently unaffected pregnancies. Together these studies demonstrate that MBRN is a reliable source for further studies on PE.

When we examined the PE diagnosis according to the stricter criteria that was in clinical use at the time Study I was performed, the PPV decreased to 63.6. The strength of a registry is based on the data assembled, and can not be more accurate than the facts committed to it. This phenomenon is illustrated by our results. By validating PE according to two diagnostic criteria we demonstrated how selection of diagnostic criteria impact the PPV. Our results show how the number of false positive cases may increase if a study uses stricter criteria to define a disease than the original data source have applied in registration of the disorder. The effects of changing diagnostic criteria on registered diagnoses, as indicated by the time-trends in our study, would be even better illustrated if we had applied each of the two criteria on both the examined study groups.

Diagnosing of complex syndromes is intricate, and the process is further complicated as the diagnostic criteria changes with time when more about the disease and pathophysiologic mechanisms is revealed, as has happened with PE. For clinical purposes all women at risk of PE need to be encompassed by the disease definition to ensure that those who potentially will suffer from severe disease are given adequate follow-up and treatment. On the other hand, large variation within the phenotypes covered by such inclusive criteria will complicate research on the disorder. By use of stricter disease classifications in research, the greater discrepancies between studies might be avoided. The complexity of PE is reflected in widely varying research results and the lack of valid replication of reported findings, and until the underlying causes of PE are illuminated properly it might be necessary to use different disease classifications for purpose of clinical diagnostics and research [177, 178].

Many studies based on longitudinal and registry data describe the applied diagnosis according to current clinical criteria in use internationally at the time of the study. Usually the registered diagnosis has not been validated according to the study's given

criteria. Diagnoses are given by contemporary criteria, and specific details that are now regarded as valuable information, such as reproducible proteinuria, may not have been registered in the medical records, and not considered significant for disease development at the time of registration. Retrospective application of criteria without evaluating their validity may lead to studies where we neither know if the selected cases represent a comparable phenotype, nor can we be certain that the controls would be registered as healthy by today's standards. This bias ought to be more thoroughly considered by researchers when including case and control groups for studies, and in interpretation of study outcomes and lack of replication of results.

Based on our results in Study I we suggest that for complex disorders with variable diagnostic criteria, it would be valuable to register clinical measurements such as blood pressures and levels of glucose and protein in the urine in addition to the diagnoses. Inclusion of measurements from the first and last antenatal control, at birth and at departure from hospital, would help identify appropriate and comparable case and control groups for retrospective studies, and additionally be a quality control both for registers and the hospital wards.

## 5.2 Phenotyping

Research on multifactorial diseases like PE necessitates large population-based studies, biobanks, and health registries. How disorders and conditions are defined in data repositories and studies will influence the disease phenotype selected for studies on a disorder, as we have demonstrated in Study I. To be able to compare studies and results it is important that the phenotypes used in the studies are similar in their expression, due to the likelihood of similar phenotypes resulting from the same biological or pathophysiological mechanisms, and that they could derive from either the same genetic variants, different changes in the same genes, or from impacts on the same genetic pathway. Using strict phenotypic criteria for a disorder could help identify the subgroups within that disease, and the shared causative factors [178]. Previously deep phenotyping has mainly been reserved for identification and classification of rare genetic diseases like rasopathies [179]. Complex interactions

between genes, and between genes and environmental influences lead to the expression of multifactorial disease phenotypes. PE is often investigated as one phenotype including all subtypes and severities, and not uncommonly including gestational hypertension without proteinuria, which has not been fulfilling the PE diagnostic criteria used currently in Europe, and globally until recently [19, 14, 21].

In Study II we performed a thorough phenotyping of individuals and families in a family biobank based on PE. We selected phenotypic criteria according to risk factors for PE (see 1.1.2 and Table 1, Study II), pregnancy conditions known to co-exist with PE, such as IUGR and gestational diabetes, and long term health effects related to preeclamptic pregnancies such as aCVD, kidney disease and diabetes mellitus type 2. All individuals in the PE Family Biobank (n=496) were classified according to all phenotypes. The prevalence in severe PE phenotypes, and a proporition of SGA higher than expected for women with pregnacies unaffected by PE, indicate that the PE Family Biobank is a unique source for examination of shared pathophysiology of these conditions. Further, although the prevalence in the cohort of the diseases chronic hypertension, aCVD and diabetes mellitus type 2 matched general prevalence estimates for developed countries [180-182] but we were still able to identify a heritable component for these diseases in the cohort. We speculate that the relative young average age in the cohort might camouflage a higher actual prevalence of the conditions in these families.

In Study III we phenotyped the cases and controls in the HUNT Study according to information received on the women both from the MBRN and from the HUNT Databank. We tried, as far as possible, to identify the same phenotypes in the HUNT cohort as in the PE Family Biobank. As the HUNT Databank contain longitudinal data and often several variables registered regarding each conditon, especially concerning physical measurements and conditions unrelated to preganancy, we were able to perform a deep phenotyping of the HUNT cohort as well. The data for the HUNT cohort in the MBRN were not as complete in regard to variables associated with pregnancies as information in the PE Family Biobank, but enough to identify individuals falsely included as controls in our study.

In PE 70% of women are affected by the disease in their first pregnancy only. It might be that these women present a phenotype with lower risk for development of aCVD and renal failure as indicated by several studies [9, 87]. Studies using this phenotype to identify common genetic associations between PE and aCVD would consequently be futile or needing very high participant numbers. Studies including other subsets of women with the disorder is required but so is thorough phenotyping of study subjects, as this would result in corresponding case and control groups across studies.

Not all individuals belonging to a family with inherited predisposition to a complex disease are affected by the disease, and the inheritance patterns are inconsistent. Further, the affected family members do not necessarily present the same phenotype of the disease. Consequently, well-characterized families as well as large enough study samples are a necessity for the family study design. We have established a unique family biobank that contains more affected families, and which we believe is better characterized compared to other family cohorts of this type [183, 44, 46, 45, 184]. For part of the information collected and used for the phenotyping in the PE Family Biobank, and in the HUNT cohort, data were based on interviews and self-reported data on disease, a research setting associated with decreased diagnostic validity and recall bias. To get as exact data as possible, additional information on the use of relevant medication and duration of the disorder was required in the phenotyping processes in Study II and Study III to register a person as afflicted by a disorder. Study II would have benefitted from further standardization of the interview setting, data entry, and coordination of the data and sample collection across the including hospitals. Further improvement to the process of establishment of the PE Family Biobank, and the selection of phenotypic questions, would have been to involve medical doctors and researchers with expertise on the other disorders phenotyped in the cohort, and to include statisticians and bioinformaticians in the project planning phase.

## **5.3** Heritability

Investigation of the heritability of distinct phenotypes in the families with PE in Study II helped map the cohort and confirm that the PE Family Biobank is a cohort with increased risk for PE development. Additionally we found that phenotypes like chronic hypertension and SGA are heritable, manifesting a genetic component of these phenotypic presentations in the cohort. This is the first study estimating heritability of PE in a multigenerational family cohort. Heritability of PE has previously been examined in trios based on the Swedish Medical Birth Registry [175]. In the Swedish study the maternal phenotype of PE was estimated to have H2r = 0.35, which was lower than our results in Study II (60%), and the fetal genetic effects to PE was found to be H2r=0.20, which correlates to the 25% we have identified for the phenotype in the PE Family Biobank. For some of the examined phenotypes there were very few affected individuals in the cohort, as for eclampsia and the HELLP syndrome, which are more severe and rare PE phenotypes. The low numbers of persons having developed acute angina, acute myocardial infarction or cerebral stroke can be attributed to the mean age of 46.8 years in the cohort.

Possible overlap in risk factors (including both genetic and environmental) influencing the presentations of the heritable phenotypes were examined to identify shared risk factors. We found phenotypic correlation between several phenotypes when they were examined in pairs (Supplementary Table 6, S8, Paper II). The identified correlations between BMI and waist circumference and several other phenotypes are consistent with the fact that BMI is an acknowledged risk factor for PE, aCVD and asthma [185] The overlap in phenotype between pulmonary disease and both aCVD severity and SGA was unexpected. A few studies have found women with asthma to have an increased risk for adverse pregnancy outcomes, including PE [186, 187], but we have not been able to identify studies who examine the genetic background of this. Our results indicate the PE Family Biobank as a useful cohort for further examination in regard to novel shared genetic pathways of the overlapping phenotypes. As we did not request information on general smoking habits, only on pregnancy-related smoking, we could not correct for smoking in our analysis, and this might be an important

confounder in the study of heritability and phenotypic overlap for certain conditions like aCVD and pulmonary disease. Another weakness in regard to the analysis of phenotypic correlations is that we have not considered the individual variables of the phenotypic correlation: the genotypic and the environmental correlations. Despite this, we speculate that due to the significant heritability of the overlapping phenotypes the results do indicate shared genetic mechanisms of trait development for several of the examined phenotypes.

### 5.4 Genetic analysis

Despite multiple molecular studies knowledge on the genetic causes of PE is still lacking. There is an increasing awareness about how these women not only run the risk of pregnancy-related complications of PE, but also have an increased risk for aCVD after the pregnancy [7, 188]. As already mentioned, the two conditions PE and aCVD share risk factors associated with environment and life-style like obesity, diabetes mellitus, and chronic hypertension [29, 35]. Molecular research has identified shared etiologic factors between the two disorders, such as endothelial dysfunction, inflammation and oxidative stress (see also paragraph 1.1.3 and 1.2.4) [69, 189, 129], but few studies have identified genetic loci regulating the development of both disorders.

In Study III we aimed to identify a genetic connection between PE and aCVD by investigating an association between PE and one or more GWAS findings associated with blood pressure regulation or with genes involved in inflammation. We performed the genetic analysis in two cohorts based on PE, the PE Family Biobank and the HUNT cohort. As already pointed out the cohorts contain women with valid diagnosis and well-classified phenotypes for all participants (see also paragraph X.Y). In the HUNT cohort of unrelated individuals we had a relatively large case-control cohort for this type of study, and in the families we had the additional knowledge that both PE and chronic hypertension were heritable in the collected families (Study II).

In the HUNT cohort one SNP, rs17367504, previously linked to a protective effect on development of chronic hypertension, was discovered to be associated with a protective effect on PE (Study III). The SNP is located in the intron of the MTHFR gene. Several studies have demonstrated that the MTHFR gene take part in the pathophysiology of PE and/or aCVD [190], but our study is the first to identify this SNP in PE, adding a new genetic link between the two conditions PE and aCVD.

We chose to performe a candidate SNP array, which is a hypothesis-driven method containing inherent weaknesses. The method would not lead to detection of genetic regions with small effect on phenotype, and such studies will necessarily be biased as they are performed based on defined possible or plausible biologic mechanism involved in the disease. The candidate SNP array facilitates investigation of specific regions/loci in the genome, and replication of findings from other studies. A GWAS maps larger parts of the genome and may detect novel areas with possible impact on the disease phenotype. The locus with a biological effect has to be relatively common both in the total population examined and in the cases to be detected as a significant genetic variant in this type of studies. The candidate SNP array method was selected for Study III as we did have a defined hypothesis for our study. To our knowledge, this way of applying GWAS results for replication in a different, but ethiologically linked phenotype, is a rather novel approach, utilized by few other studies on PE.

Complex diseases pose their own challenges in genetic studies, as several biases are implemented by the combination of genetic and environmental influences, producing different phenotypic presentations. Incorrect results may be due to population stratification, epistasis or epigenetic modifications. Population stratification describes situations when allele frequencies at a defined locus differ between cases and control irrespective of disease state but rather due to fundamental differences in the population examined [191]. In the process of epistasis one or more genetic loci can modify the protein regulation of a different genetic locus and thereby contribute to which phenotype is expressed [192], while during epigenetic modification environmental factors influence the gene regulation of protein production [132].

All results, also in the field of genetics, require reproducibility. The likelihood that findings indicate a true disease mechanism increases when similar results are found by comparable studies in different cohorts. Unfortunately, most PE-related genetic results to date have not been reproducible in other cohorts [49]. The lacking consistency of genetic associations with a phenotype might in part be due to the selection of diagnostic criteria used in studies. Often wide definitions have been applied and phenotypes are diverging, both within a study and between studies. Further rarely the criteria used are validated in registry-based studies. Consequently the cohorts may include a large proportion of false positive cases and false negative controls, as well as a mix of similar phenotypes with different etiology. The three studies included in this thesis illustrates the need for clean, specific phenotypes and knowledge of the validity of the chosen diagnostic study criteria in the cohort.

The SNPs included in Study III were selected from GWAS for blood pressure regulation and genes with inflammatory functions. Still, we could not reproduce the associations between genotype and chronic hypertension in our PE-based cohorts. One reason for this could be that the cohorts are not collected on the basis of aCVD or chronic hypertension. While the diagnosis of chronic hypertension in the HUNT cohort was based on repeated measurements of blood pressure, the registered diagnosis in the PE Family Biobank of chronic hypertension has not been validated against another source. For the PE Family Biobank the missing validation could be another cause for the lack of replication of results, as could the number of participants and average age in the cohort. Prior to study start we estimated the statistical power of the study in the HUNT cohort in regard to detection of significant associations between the selected SNPs and PE. As fewer individuals in the HUNT cohort has the phenotype chronic hypertension, the statistical power to identify significant associations with chronic hypertension decreases compared to our estimates for PE, and the lack of power may be another reason why the associations between chronic hypertension and SNPs could not be replicated in the HUNT cohort.

We could not replicate in the PE Family Biobank the significant association between rs17367504 and the maternal PE phenotype detected in the HUNT cohort. Nor did thia

SNP associate to the phenotype chronic hypertension in the PE Family Biobank. Rather, in the PE Family Biobank we found a suggestive association between PE and an intronic SNP, rs633185, in the Rho GTPase-Activating protein 42 (ARHGAP42). The SNP have been discovered to associate with chronic hypertension in GWAS, but this association did not replicate with chronic hypertension in the PE Family Biobank, and with neither of the two phenotypes in the HUNT cohort. ARHGAP42 has been assigned a function in the regulation of blood pressure by influencing smooth muscle cell contractility in blood vessels.

After cross-examining the results from the two cohorts we are not yet convinced that we have identified a true association between a genetic locus in ARHGAP42 and PE. The effect of the top SNPs for the families go partly in the opposite directions to the effects indicated for the same SNPs in the HUNT cohort. It may be that the suggested genotype-phenotype associations are spurious, or that the PE Family Biobank contains a population distinct from the HUNT population. The geographic areas of collection overlap, but the diagnostic criteria used for the two cohort differ, with stricter criteria defining the PE diagnosis in the PE Family Biobank.

The PE Family Biobank is solidly founded on families containing at least two close female relatives who have experienced PE according to the study criteria, confirmed by information in their hospital records. That the participants have been strictly classified according to several pregnancy- and disease-related phenotypes in the cohort could strengthen the probability of identifying genotype-phenotype associations. Further, we have identified several heritable phenotypes, and partly overlap in phenotypes, in the families contained in the PE Family Biobank (Study II). These results could indicate that these familes are different from a general preeclamptic population. Still, it is unlikely that a SNP has a protective effect on development of PE in the families while it is a risk factor of the disorder in a general, unrelated population. Before we can report the finding as a true link between the two diseases, or as a negative finding, we need to examine more in depth the genetic region, possibility of reproduction of the finding in other cohorts or by other genetic methods, as well as through functional studies.

# 6. CONCLUSIONS

Much knowledge on the pathophysiologic mechanisms contributing to PE development is still missing. As women with preeclamptic pregnancies have increased risk of development of aCVD there is additional need for identification of the biologic functions affected in both diseases. Through the studies described in this thesis we have added new information to the field on the validity of registered diagnoses of PE, on biological risk factors for PE and about shared genetic mechanisms between PE and aCVD. The following conclusions can be drawn: When the diagnostic criteria, as required by the registry is applied, the validity of PE diagnoses entered in the Medical Birth Registry of Norway (MBRN) is high **(Study I)**.

**6.1** A thorough phenotyping has been performed where all individuals included in the PE Family Biobank have been classified according to several phenotypes of preeclampsia and related phenotypes of pregnancy, as well as phenotypes regarding PE risk factors and health conditions **(Study II).**

**6.2** In the PE Family Biobank several phenotypes of PE and PE-related conditions, including risk factors for aCVD, are heritable **(Study II)**.

**6.3** Several of the heritable phenotypes in the PE Family Biobank, demonstrate phenotypic overlap with at least one of the other heritable traits, indicating shared pathophysiologic mechanisms these phenotypes **(Study II)**.

**6.4** A shared genetic risk factor for PE and the aCVD risk factor chronic hypertension has been identified by a candidate SNP array approach in a nested PE case-control cohort based on the HUNT surveys (**Study III)**.

**6.5** Associations between chronic hypertension and SNPs identified in genome-wide association studies could not be replicated, neither in a family-based PE cohort, the PE Family Biobank, or in a nested PE case-control cohort, the HUNT cohort **(Study III).**

**6.6** After correction for multiple testing no significant associations could be identified between chronic hypertension and the defined genes with inflammatory and angiogenetic impact in the two study cohorts, the PE Family Biobank and the HUNT case-control cohort based on PE **(Study III).**

# 7. FUTURE PERSPECTIVES

The combination of advanced genetic analysis and a realization of the importance of phenotypic classification and standardization will enable us to better understand the relationship between PE and aCVD [193, 12]. Combined international efforts are required to gain knowledge about the molecular and genetic biological networks responsible for disease development. This will be critical in the identification of individuals at risk and for selection of targets for intervention. The information gained in the presented thesis identifies several points to be addressed in future work:

- Further investigation of the genetic region around the locus identified in the MTHFR gene and of the locus mapped to TLR4 in the HUNT cohort is necessary. This could be achieved by genetic replication in other PE cohorts, by gene-centric array, and through functional studies in biological tissues.

- We will investigate further if the SNP findings in the PE Family Biobank demonstrate true associations, by using genetic replication in other PE cohorts and functional studies like immunohistochemistry in maternal blood and placental and decidual tissues.

- It is of interest to examine more deeply the other heritable phenotypes in the PE Family Biobank, both through genetic and functional studies, and by collaboration with researchers working with other family-based cohorts with PE.

- Any significant findings for chronic hypertension and aCVD in women ought to be investigated further to identify possible gender specific mechanisms in aCVD cohorts where both sexes are represented such as the Cohort of Norway.

# 8. REFERENCES


1. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol. 2009;33(3):130-7.
2. Say L, Chou D, Gemmill A, Tuncalp O, Moller AB, Daniels J et al. Global causes of maternal death: a WHO systematic analysis. Lancet Glob Health. 2014;2(6):e323-33. doi:10.1016/S2214-109X(14)70227-X.
3. Klungsøyr K, Morken N, Irgens L, Vollset S, Skjærven R. Secular trends in the epidemiology of pre-eclampsia throughout 40 years in Norway: prevalence, risk factors and perinatal survival. Paediatr Perinat Epidemiol. 2012;26(3):190-8.
4. Magnussen EB, Vatten LJ, Smith GD, Romundstad PR. Hypertensive disorders in pregnancy and subsequently measured cardiovascular risk factors. Obstet Gynecol. 2009;114(5):961-70. doi:10.1097/AOG.0b013e3181bb0dfc.
5. Irgens HU, Reisaeter L, Irgens LM, Lie RT. Long term mortality of mothers and fathers after pre-eclampsia: population based cohort study. BMJ. 2001;323(7323):1213-7.
6. Lykke J, Langhoff-Roos J, Sibai B, Funai E, Triche E, Paidas M. Hypertensive Pregnancy Disorders and Subsequent Cardiovascular Morbidity and Type 2 Diabetes Mellitus in the Mother Hypertension. 2009;53:944-51.
7. Brown MC, Best KE, Pearce MS, Waugh J, Robson SC, Bell R. Cardiovascular disease risk in women with pre-eclampsia: systematic review and meta-analysis. Eur J Epidemiol. 2013;28(1):1-19. doi:10.1007/s10654-013-9762-6.
8. Osol G, Bernstein I. Preeclampsia and maternal cardiovascular disease: consequence or predisposition? J Vasc Res. 2014;51(4):290-304. doi:10.1159/000367627.
9. Rich-Edwards JW, Fraser A, Lawlor DA, Catov JM. Pregnancy characteristics and women's future cardiovascular health: an underused opportunity to improve women's health? Epidemiol Rev. 2014;36(1):57-70. doi:10.1093/epirev/mxt006.

10. Bernardes TP, Mol BW, Ravelli AC, van den Berg PP, Stolk RP, Groen H. Early and late onset preeclampsia versus small for gestational age risks in subsequent pregancies. XIX World Congress for the International Society for the Study of Hypertension in Pregnancy; 26-29.09.14; New Orleans, Louisiana, United States of America2014. p. 46-OR.
11. Smith IB. The impact of Stephen Hales on medicine. J R Soc Med. 1993;86(6):349-52.
12. Myatt L, Redman CW, Staff AC, Hansson S, Wilson ML, Laivuori H et al. Strategy for Standardization of Preeclampsia Research Study Design. Hypertension. 2014;63(6):1293-301. doi:10.1161/HYPERTENSIONAHA.113.02664.
13. Thomsen LC, Klungsoyr K, Roten LT, Tappert C, Araya E, Baerheim G et al. Validity of the diagnosis of pre-eclampsia in the Medical Birth Registry of Norway. Acta Obstet Gynecol Scand. 2013;92(8):943-50. doi:10.1111/aogs.12159.
14. Steegers EA, von Dadelszen P, Duvekot JJ, Pijnenborg R. Pre-eclampsia. Lancet. 2010;376(9741):631-44.
15. Magee LA, Pelsb A, Helewac M, Reyd E, von Dadelszen P, Group ObotCHDoPHW. Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy. Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health. 2014;4(2):105-45.
16. Gynaecologists TRAaNZCoOa. Antenatal care - module 1 Clinical practice guidelines. Department of Health and Ageing (DoHA); 2013. p. 59.
17. Tranquilli AL, Dekker G, Magee L, Roberts J, Sibai BM, Steyn W et al. The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP. Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health. 2014;4:97-104.
18. . WHO Recommendations for Prevention and Treatment of Pre-Eclampsia and Eclampsia. WHO Guidelines Approved by the Guidelines Review Committee. Geneva2011.

19. Visintin C, Mugglestone MA, Almerie MQ, Nherera LM, James D, Walkinshaw S et al. Management of hypertensive disorders during pregnancy: summary of NICE guidance. BMJ. 2010;341:c2207. doi:10.1136/bmj.c2207.
20. Walker RL, Hemmelgarn B, Quan H. Incidence of gestational hypertension in the Calgary Health Region from 1995 to 2004. Can J Cardiol. 2009;25(8):e284-7.
21. Staff A, Andersgaard A, Henriksen T, Langesæter E, Magnussen E, Michelsen T et al. Norwegian [Veileder i fødselshjelp 2014. Hypertensive svangerskapskomplikasjoner]. Norway: The Norwegian Medical Association2014.
22. Rasmussen S, Irgens LM. Fetal growth and body proportion in preeclampsia. Obstet Gynecol. 2003;101(3):575-83.
23. Groom KM, North RA, Poppe KK, Sadler L, McCowan LM. The association between customised small for gestational age infants and pre-eclampsia or gestational hypertension varies with gestation at delivery. BJOG. 2007;114(4):478-84. doi:10.1111/j.1471-0528.2007.01277.x.
24. Roberts JM. Pathophysiology of ischemic placental disease. Semin Perinatol. 2014;38(3):139-45. doi:10.1053/j.semperi.2014.03.005.
25. Salam RA, Das JK, Bhutta ZA. Impact of intrauterine growth restriction on long-term health. Curr Opin Clin Nutr Metab Care. 2014;17(3):249-54. doi:10.1097/MCO.0000000000000051.
26. Nilsson E, Salonen Ros H, Cnattingius S, Lichtenstein P. The importance of genetic and environmental effects for pre-eclampsia and gestational hypertension: a family study. BJOG. 2004;111(3):200-6.
27. Ness RB, Roberts JM. Heterogeneous causes constituting the single syndrome of preeclampsia: a hypothesis and its implications. Am J Obstet Gynecol. 1996;175(5):1365-70.
28. Chesley LC, Cooper DW. Genetics of hypertension in pregnancy: possible single gene control of pre-eclampsia and eclampsia in the descendants of eclamptic women. Br J Obstet Gynaecol. 1986;93(9):898-908.

29. Broughton Pipkin F. Risk factors for preeclampsia. N Engl J Med. 2001;344(12):925-6. doi:10.1056/NEJM200103223441209.
30. Trogstad L, Skrondal A, Stoltenberg C, Magnus P, Nesheim BI, Eskild A. Recurrence risk of preeclampsia in twin and singleton pregnancies. Am J Med Genet A. 2004;126A(1):41-5. doi:10.1002/ajmg.a.20512.
31. Coonrod DV, Hickok DE, Zhu K, Easterling TR, Daling JR. Risk factors for preeclampsia in twin pregnancies: a population-based cohort study. Obstet Gynecol. 1995;85(5 Pt 1):645-50.
32. Tandberg A, Klungsoyr K, Romundstad L, Skjaerven R. Pre-eclampsia and assisted reproductive technologies: consequences of advanced maternal age, interbirth intervals, new partner and smoking habits. BJOG. 2014. doi:10.1111/1471-0528.13051.
33. Chen XK, Wen SW, Bottomley J, Smith GN, Leader A, Walker MC. In vitro fertilization is associated with an increased risk for preeclampsia. Hypertens Pregnancy. 2009;28(1):1-12. doi:10.1080/10641950802001859.
34. Malchau SS, Loft A, Larsen EC, Aaris Henningsen AK, Rasmussen S, Andersen AN et al. Perinatal outcomes in 375 children born after oocyte donation: a Danish national cohort study. Fertil Steril. 2013;99(6):1637-43. doi:10.1016/j.fertnstert.2013.01.128.
35. Sibai BM, Lindheimer M, Hauth J, Caritis S, VanDorsten P, Klebanoff M et al. Risk factors for preeclampsia, abruptio placentae, and adverse neonatal outcomes among women with chronic hypertension. National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. N Engl J Med. 198;339(10):667-71.
36. Jeyabalan A. Epidemiology of preeclampsia: impact of obesity. Nutr Rev. 2013;71 Suppl 1:S18-25. doi:10.1111/nure.12055.
37. O'Brien TE, Ray JG, Chan WS. Maternal body mass index and the risk of preeclampsia: a systematic overview. Epidemiology. 2003;14(3):368-74.

38. Schoenaker D, Soedamah-Muthu SS, Mishra GD. The association between dietary factors and gestational hypertension and pre-eclampsia: a systematic review and meta-analysis of observational studies. BMC Med. 2014;12(1):157. doi:10.1186/s12916-014-0157-7.

39. England L, Zhang J. Smoking and risk of preeclampsia: a systematic review. Front Biosci. 2007;12:2471-83.

40. Buurma AJ, Turner RJ, Driessen JH, Mooyaart AL, Schoones JW, Bruijn JA et al. Genetic variants in pre-eclampsia: a meta-analysis. Hum Reprod Update. 2013;19(3):289-303. doi:10.1093/humupd/dms060.

41. Johnson M, Brennecke S, East C, Göring H, Kent Jr. J, Dyer T et al. Genome-Wide Association Scan Identifies a Risk Locus for Preeclampsia on 2q14, Near the Inhibin, Beta B Gene. PLoS ONE. 2012;7(3):e33666.

42. Magee LA, Pels A, Helewa M, Rey E, von Dadelszen P, Committee SHG. Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy: executive summary. J Obstet Gynaecol Can. 2014;36(7):575-6.

43. Laivuori H, Lahermo P, Ollikainen V, Widen E, Haiva-Mallinen L, Sundstrom H et al. Susceptibility loci for pre-eclampsia on chromosomes 2p25 and 9p13 in Finnish families. Am J Hum Genet. 2003;72(1):168-77.

44. Arngrimsson R, Sigurdardottir S, Frigge ML, Bjarnadottir RI, Jonsson T, Stefansson H et al. A genome-wide scan reveals a maternal susceptibility locus for pre-eclampsia on chromosome 2p13. Hum Mol Genet. 1999;8(9):1799-805.

45. Lachmeijer AM, Arngrimsson R, Bastiaans EJ, Frigge ML, Pals G, Sigurdardottir S et al. A genome-wide scan for preeclampsia in the Netherlands. Eur J Hum Genet. 2001;9(10):758-64. doi:10.1038/sj.ejhg.5200706.

46. Moses EK, Lade JA, Guo G Wilton AN, Grehan M, Freed K, Borg A et al. A genome scan in families from Australia & New Zealand confirms the presence of a maternal susceptibility locus for pre-eclampsia, on chromosome 2. Am J Hum Genet. 2000;67(6):1581-5.

47. Harrison GA. A Genomewide Linkage Study of Preeclampsia/Eclampsia Reveals Evidence for a Candidate Region on 4q. Am J Hum Genet. 1997;60:1158-67.

48. Johnson MP, Fitzpatrick E, Dyer TD, Jowett JB, Brennecke SP, Blangero J et al. Identification of two novel quantitative trait loci for pre-eclampsia susceptibility on chromosomes 5q and 13q using a variance components-based linkage approach. Mol Hum Reprod. 2007;13(1):61-7.

49. Johnson MP, Brennecke SP, East CE, Goring HH, Kent JW, Jr., Dyer TD et al. Genome-wide association scan identifies a risk locus for preeclampsia on 2q14, near the inhibin, beta B gene. PLoS ONE. 2012;7(3):e33666. doi:10.1371/journal.pone.0033666.

50. Majander KK, Villa PM, Kivinen K, Kere J, Laivuori H. A follow-up linkage study of Finnish pre-eclampsia families identifies a new fetal susceptibility locus on chromosome 18. Eur J Hum Genet. 2013;21(9):1024-6. doi:10.1038/ejhg.2013.6.

51. Wang X, Bai T, Liu S, Pan H, Wang B. Association between thrombophilia gene polymorphisms and preeclampsia: a meta-analysis. PLoS ONE. 2014;9(6):e100789. doi:10.1371/journal.pone.0100789.

52. Qi HP, Fraser WD, Luo ZC, Julien P, Audibert F, Wei SQ. Endothelial nitric oxide synthase gene polymorphisms and risk of preeclampsia. Am J Perinatol. 2013;30(10):795-804. doi:10.1055/s-0032-1333406.

53. Cheng D, Hao Y, Zhou W, Ma Y. Vascular endothelial growth factor +936C/T, -634G/C, -2578C/A, and -1154G/A polymorphisms with risk of preeclampsia: a meta-analysis. PLoS ONE. 2013;8(11):e78173. doi:10.1371/journal.pone.0078173.

54. Wang XM, Wu HY, Qiu XJ. Methylenetetrahydrofolate reductase (MTHFR) gene C677T polymorphism and risk of preeclampsia: an updated meta-analysis based on 51 studies. Arch Med Res. 2013;44(3):159-68. doi:10.1016/j.arcmed.2013.01.011.

55. Yang W, Zhu Z, Wang J, Ye W, Ding Y. Evaluation of association of maternal IL-10 polymorphisms with risk of preeclampsia by A meta-analysis. J Cell Mol Med. 2014;18(12):2466-77. doi:10.1111/jcmm.12434.

56. Ramsey EM, Houston ML, Harris JW. Interactions of the trophoblast and maternal tissues in three closely related primate species. Am J Obstet Gynecol. 1976;124(6):647-52.

57. Morgan GE, Mikhail MS, Murray MJ. Clinical Anesthesiologi. 4 ed. McGraw-Hill Professional Publishing; 2005.

58. Germain SJ, Sacks GP, Sooranna SR, Sargent IL, Redman CW. Systemic inflammatory priming in normal pregnancy and preeclampsia: the role of circulating syncytiotrophoblast microparticles. J Immunol. 2007;178(9):5949-56.

59. Smarason AK, Sargent IL, Starkey PM, Redman CW. The effect of placental syncytiotrophoblast microvillous membranes from normal and pre-eclamptic women on the growth of endothelial cells in vitro. Br J Obstet Gynaecol. 1993;100:943-9.

60. Knight M, Redman CW, Linton EA, Sargent IL. Shedding of syncytiotrophoblast microvilli into the maternal circulation in pre-eclamptic pregnancies. Br J Obstet Gynaecol. 1998;105(6):632-40.

61. Sargent IL, Borzychowski AM, Redman CW. Immunoregulation in normal pregnancy and pre-eclampsia: an overview. Reprod Biomed Online. 2006;13(5):680-6.

62. Meekins JW, Pijnenborg R, Hanssens M, McFadyen IR, van Asshe A. A study of placental bed spiral arteries and trophoblast invasion in normal and severe pre-eclamptic pregnancies. Br J Obstet Gynaecol. 1994;101(8):669-74.

63. Knofler M, Pollheimer J. IFPA Award in Placentology lecture: molecular regulation of human trophoblast invasion. Placenta. 2012;33 Suppl:S55-62. doi:10.1016/j.placenta.2011.09.019.

64. Myatt L, Kossenjans W, Sahay R, Eis A, Brockman D. Oxidative stress causes vascular dysfunction in the placenta. J Matern Fetal Med. 2000;9(1):79-82. doi:10.1002/(SICI)1520-6661(200001/02)9:1<79::AID-MFM16>3.0.CO;2-O.

65. Wang A, Rana S, Karumanchi SA. Preeclampsia: the role of angiogenic factors in its pathogenesis. Physiology (Bethesda). 2009;24:147-58. doi:10.1152/physiol.00043.2008.

66. Redman CW, Sargent IL. Pre-eclampsia, the placenta and the maternal systemic inflammatory response--a review. Placenta. 2003;24 Suppl A:S21-7.

67. Tabas I. Macrophage death and defective inflammation resolution in atherosclerosis. Nat Rev Immunol. 2010;10(1):36-46. doi:10.1038/nri2675.

68. Olsson MG, Centlow M, Rutardottir S, Stenfors I, Larsson J, Hosseini-Maaf B et al. Increased levels of cell-free hemoglobin, oxidation markers, and the antioxidative heme scavenger alpha(1)-microglobulin in preeclampsia. Free Radic Biol Med. 2010;48(2):284-91. doi:10.1016/j.freeradbiomed.2009.10.052.

69. Redman CW, Sargent IL. Circulating microparticles in normal pregnancy and pre-eclampsia. Placenta. 2008;29 Suppl A:S73-7. doi:10.1016/j.placenta.2007.11.016.

70. Roberts JM, Hubel CA. The two stage model of preeclampsia: variations on the theme. Placenta. 2009;30 Suppl A:S32-7. doi:10.1016/j.placenta.2008.11.009.

71. Hansson SR, Gram M, Akerstrom B. Fetal hemoglobin in preeclampsia: a new causative factor, a tool for prediction/diagnosis and a potential target for therapy. Curr Opin Obstet Gynecol. 2013;25(6):448-55. doi:10.1097/GCO.0000000000000022.

72. Kenny LC, Black MA, Poston L, Taylor R, Myers JE, Baker PN et al. Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: the Screening for Pregnancy Endpoints (SCOPE) international cohort study. Hypertension. 2014;64(3):644-52. doi:10.1161/HYPERTENSIONAHA.114.03578.

73. Kuc S, Koster MP, Franx A, Schielen PC, Visser GH. Maternal characteristics, mean arterial pressure and serum markers in early prediction of preeclampsia. PLoS ONE. 2013;8(5):e63546. doi:10.1371/journal.pone.0063546.

74. Than NG, Balogh A, Romero R, Karpati E, Erez O, Szilagyi A et al. Placental Protein 13 (PP13) - A Placental Immunoregulatory Galectin Protecting Pregnancy. Front Immunol. 2014;5:348. doi:10.3389/fimmu.2014.00348.

75. Akolekar R, de Cruz J, Foidart JM, Munaut C, Nicolaides KH. Maternal plasma soluble fms-like tyrosine kinase-1 and free vascular endothelial growth factor at 11 to 13 weeks of gestation in preeclampsia. Prenat Diagn. 2010;30(3):191-7. doi:10.1002/pd.2433.

76. Levine RJ, Lam C, Qian C, Yu KF, Maynard SE, Sachs BP et al. Soluble endoglin and other circulating antiangiogenic factors in preeclampsia. N Engl J Med. 2006;355(10):992-1005. doi:10.1056/NEJMoa055352.

77. Venkatesha S, Toporsian M, Lam C, Hanai J, Mammoto T, Kim YM et al. Soluble endoglin contributes to the pathogenesis of preeclampsia. Nat Med. 2006;12(6):642-9. doi:10.1038/nm1429.

78. Poon LC, Kametas NA, Chelemen T, Leal A, Nicolaides KH. Maternal risk factors for hypertensive disorders in pregnancy: a multivariate approach. J Hum Hypertens. 2010;24(2):104-10. doi:10.1038/jhh.2009.45.

79. Skrastad RB, Hov GG, Blaas HG, Romundstad PR, Salvesen KA. A prospective study of screening for hypertensive disorders of pregnancy at 11-13 weeks in a Scandinavian population. Acta Obstet Gynecol Scand. 2014;93(12):1238-47. doi:10.1111/aogs.12479.

80. Villa PM, Kajantie E, Raikkonen K, Pesonen AK, Hamalainen E, Vainio M et al. Aspirin in the prevention of pre-eclampsia in high-risk women: a randomised placebo-controlled PREDO Trial and a meta-analysis of randomised trials. BJOG. 2013;120(1):64-74. doi:10.1111/j.1471-0528.2012.03493.x.

81. Askie LM, Duley L, Henderson-Smart DJ, Stewart LA, Group PC. Antiplatelet agents for prevention of pre-eclampsia: a meta-analysis of individual patient data. Lancet. 2007;369(9575):1791-8. doi:10.1016/S0140-6736(07)60712-0.

82. Altman D, Carroli G, Duley L, Farrell B, Moodley J, Neilson J et al. Do women with pre-eclampsia, and their babies, benefit from magnesium sulphate? The Magpie Trial: a randomised placebo-controlled trial. Lancet. 2002;359(9321):1877-90.

83. Roberts D, Dalziel S. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. Cochrane Database Syst Rev. 2006(3):CD004454. doi:10.1002/14651858.CD004454.pub2.

84. Bellamy L, Casas JP, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. BMJ. 2007;335(7627):974. doi:10.1136/bmj.39335.385301.BE.

85. Feig DS, Shah BR, Lipscombe LL, Wu CF, Ray JG, Lowe J et al. Preeclampsia as a risk factor for diabetes: a population-based cohort study. PLoS Med. 2013;10(4):e1001425. doi:10.1371/journal.pmed.1001425.

86. Mannisto T, Mendola P, Vaarasmaki M, Jarvelin MR, Hartikainen AL, Pouta A et al. Elevated blood pressure in pregnancy and subsequent chronic disease risk. Circulation. 2013;127(6):681-90. doi:10.1161/CIRCULATIONAHA.112.128751.

87. Vikse BE, Irgens LM, Leivestad T, Skjaerven R, Iversen BM. Preeclampsia and the risk of end-stage renal disease. N Engl J Med. 2008;359(8):800-9. doi:10.1056/NEJMoa0706790.

88. Sattar N, Greer IA. Pregnancy complications and maternal cardiovascular risk: opportunities for intervention and screening? BMJ. 2002;325(7356):157-60.

89. Magnussen EB, Vatten LJ, Lund-Nilsen TI, Salvesen KA, Davey Smith G, Romundstad PR. Prepregnancy cardiovascular risk factors as predictors of pre-eclampsia: population based cohort study. BMJ. 2007;335(7627):978. doi:10.1136/bmj.39366.416817.BE.

90. Barker DJ. In utero programming of cardiovascular disease. Theriogenology. 2000;53(2):555-74.

91. Geelhoed JJ, Fraser A, Tilling K, Benfield L, Davey Smith G, Sattar N et al. Preeclampsia and gestational hypertension are associated with childhood blood pressure independently of family adiposity measures: the Avon Longitudinal Study of Parents and Children. Circulation. 2010;122(12):1192-9. doi:10.1161/CIRCULATIONAHA.110.936674.

92. Palmsten K, Buka SL, Michels KB. Maternal pregnancy-related hypertension and risk for hypertension in offspring later in life. Obstet Gynecol. 2010;116(4):858-64. doi:10.1097/AOG.0b013e3181f3a1f9.

93. Wu CS, Nohr EA, Bech BH, Vestergaard M, Catov JM, Olsen J. Health of children born to mothers who had preeclampsia: a population-based cohort study. Am J Obstet Gynecol. 2009;201(3):269 e1- e10. doi:10.1016/j.ajog.2009.06.060.

94. Wu CS, Nohr EA, Bech BH, Vestergaard M, Catov JM, Olsen J. Diseases in children born to mothers with preeclampsia: a population-based sibling cohort study. Am J Obstet Gynecol. 2011;204(2):157 e1-5. doi:10.1016/j.ajog.2010.08.046.

95. Herrera-Garcia G, Contag S. Maternal preeclampsia and risk for cardiovascular disease in offspring. Curr Hypertens Rep. 2014;16(9):475. doi:10.1007/s11906-014-0475-3.

96. Davis EF, Lazdam M, Lewandowski AJ, Worton SA, Kelly B, Kenworthy Y et al. Cardiovascular risk factors in children and young adults born to preeclamptic pregnancies: a systematic review. Pediatrics. 2012;129(6):e1552-61. doi:10.1542/peds.2011-3093.

97. Kajantie E, Eriksson JG, Osmond C, Thornburg K, Barker DJ. Pre-eclampsia is associated with increased risk of stroke in the adult offspring: the Helsinki birth cohort study. Stroke. 2009;40(4):1176-80. doi:10.1161/STROKEAHA.108.538025.

98. Tunstall-Pedoe H, Kuulasmaa K, Mahonen M, Tolonen H, Ruokokoski E, Amouyel P. Contribution of trends in survival and coronary-event rates to changes in coronary heart disease mortality: 10-year results from 37 WHO MONICA project populations. Monitoring trends and determinants in cardiovascular disease. Lancet. 1999;353(9164):1547-57.

99. Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G et al. Executive summary: heart disease and stroke statistics--2010 update: a report from the American Heart Association. Circulation. 2010;121(7):948-54. doi:10.1161/CIRCULATIONAHA.109.192666.

100. . In: Fuster V, Kelly BB, editors. Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. The National Academies Collection: Reports funded by National Institutes of Health. Washington (DC)2010.

101. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2095-128. doi:10.1016/S0140-6736(12)61728-0.

102. Shaw LJ, Bairey Merz CN, Pepine CJ, Reis SE, Bittner V, Kelsey SF et al. Insights from the NHLBI-Sponsored Women's Ischemia Syndrome Evaluation (WISE) Study: Part I: gender differences in traditional and novel risk factors, symptom evaluation, and gender-optimized diagnostic strategies. J Am Coll Cardiol. 2006;47(3 Suppl):S4-S20. doi:10.1016/j.jacc.2005.01.072.

103. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937-52. doi:10.1016/S0140-6736(04)17018-9.

104. Viles-Gonzalez JF, Fuster V, Badimon JJ. Atherothrombosis: a widespread disease with unpredictable and life-threatening consequences. Eur Heart J. 2004;25(14):1197-207. doi:10.1016/j.ehj.2004.03.011.

105. Aursulesei V, Roca IC. Venous pulmonary thromboembolism and atherothrombosis: need for a new approach. Rev Med Chir Soc Med Nat Iasi. 2014;118(3):724-32.

106. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD et al. Third universal definition of myocardial infarction. Circulation. 2012;126(16):2020-35. doi:10.1161/CIR.0b013e31826e1058.

107. Arboix A, Alio J. Cardioembolic stroke: clinical features, specific cardiac disorders and prognosis. Curr Cardiol Rev. 2010;6(3):150-61. doi:10.2174/157340310791658730.

108. Katsanos AH, Giannopoulos S, Kosmidou M, Voumvourakis K, Parissis JT, Kyritsis AP et al. Complex atheromatous plaques in the descending aorta and the risk of stroke: a systematic review and meta-analysis. Stroke. 2014;45(6):1764-70. doi:10.1161/STROKEAHA.114.005190.

109. Anderson JL, Adams CD, Antman EM, Bridges CR, Califf RM, Casey DE, Jr. et al. 2011 ACCF/AHA Focused Update Incorporated Into the ACC/AHA 2007 Guidelines for the Management of Patients With Unstable Angina/Non-ST-Elevation Myocardial Infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation. 2011;123(18):e426-579. doi:10.1161/CIR.0b013e318212bb8b.

110. Ferket BS, Colkesen EB, Visser JJ, Spronk S, Kraaijenhagen RA, Steyerberg EW et al. Systematic review of guidelines on cardiovascular risk assessment: Which recommendations should clinicians follow for a cardiovascular health check? Arch Intern Med. 2010;170(1):27-40. doi:10.1001/archinternmed.2009.434.

111. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med. 2007;356(23):2388-98. doi:10.1056/NEJMsa053935.

112. Global Burden of Metabolic Risk Factors for Chronic Diseases C, Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB et al. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. Lancet. 2014;383(9921):970-83. doi:10.1016/S0140-6736(13)61836-X.

113. Briasoulis A, Agarwal V, Tousoulis D, Stefanadis C. Effects of antihypertensive treatment in patients over 65 years of age: a meta-analysis of randomised controlled studies. Heart. 2014;100(4):317-23. doi:10.1136/heartjnl-2013-304111.

114. Wilson PW. Established risk factors and coronary artery disease: the Framingham Study. Am J Hypertens. 1994;7(7 Pt 2):7S-12S.

115. Perk J, De Backer G, Gohlke H, Graham I, Reiner Z, Verschuren WM et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012) : the fifth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). Int J Behav Med. 2012;19(4):403-88. doi:10.1007/s12529-012-9242-5.

116. World Health Organization. Global health risks: Mortality and burden of disease attributable to selected major risks. Geneva: World Health Organization; 2009.

117. WHO. World Health Organization. Diet, nutrition and the prevention of chronic diseases. Report of a joint WHO/FAO expert consultation. Geneiva: World Health Organization; 2003.

118. Bazzano LA, Serdula MK, Liu S. Dietary intake of fruits and vegetables and risk of cardiovascular disease. Curr Atheroscler Rep. 2003;5(6):492-9.

119. Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA et al. Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. Lancet. 2014;383(9913):245-54.

120. Fifth Joint Task Force of the European Society of C, European Association of E, European Association of Percutaneous Cardiovascular I, European Heart Rhythm A, Heart Failure A, European Association for Cardiovascular P et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012): the Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur J Prev Cardiol. 2012;19(4):585-667. doi:10.1177/2047487312450228.

121. Blum A, Blum N. Coronary artery disease: Are men and women created equal? Gend Med. 2009;6:410.

122. Youngblom E, Knowles JW. Familial Hypercholesterolemia. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT et al., editors. GeneReviews(R). Seattle (WA)1993.
123. Cohen W, Castelli C, Suchon P, Bouvet S, Aillaud MF, Brunet D et al. Risk assessment of venous thrombosis in families with known hereditary thrombophilia: the MARseilles-NImes prediction model. J Thromb Haemost. 2014;12(2):138-46.
124. Casas JP, Cavalleri GL, Bautista LE, Smeeth L, Humphries SE, Hingorani AD. Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. Am J Epidemiol. 2006;164(10):921-35. doi:10.1093/aje/kwj302.
125. Pollock DM. 2013 Dahl Lecture: American Heart Association council for high blood pressure research clarifying the physiology of endothelin. Hypertension. 2014;63(5):e110-7. doi:10.1161/HYPERTENSIONAHA.114.02441.
126. Bian K, Doursout MF, Murad F. Vascular system: role of nitric oxide in cardiovascular diseases. J Clin Hypertens (Greenwich). 2008;10(4):304-10.
127. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9. doi:10.1038/nature10405.
128. Tragante V, Barnes MR, Ganesh SK, Lanktree MB, Guo W, Franceschini N et al. Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. Am J Hum Genet. 2014;94(3):349-60. doi:10.1016/j.ajhg.2013.12.016.
129. Weber C, Noels H. Atherosclerosis: current pathogenesis and therapeutic options. Nat Med. 2011;17(11):1410-22. doi:10.1038/nm.2538.
130. van Holten TC, Waanders LF, de Groot PG, Vissers J, Hoefer IE, Pasterkamp G et al. Circulating biomarkers for predicting cardiovascular disease risk; a systematic review and comprehensive overview of meta-analyses. PLoS ONE. 2013;8(4):e62080. doi:10.1371/journal.pone.0062080.

131. Akesson A, Larsson SC, Discacciati A, Wolk A. Low-risk diet and lifestyle habits in the primary prevention of myocardial infarction in men: a population-based prospective cohort study. J Am Coll Cardiol. 2014;64(13):1299-306. doi:10.1016/j.jacc.2014.06.1190.
132. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000;403(6765):41-5. doi:10.1038/47412.
133. Clancy S, Brown W. Translation: DNA to mRNA to protein. Nature Education. 2008;1(1):101.
134. Strachan T, Read AP. Human Molecular Genetics. 2nd ed. New York1999.
135. Scriver CR. Allelic and Locus Heterogeneity. eLS. . John Wiley & Sons, Ltd.; 2006.
136. Merriam-Webster's Medical Dictionary. Merriam-Webster Inc; 2014.
137. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477-85. doi:10.1038/nrg2361.
138. Al-Chalabi A. Genome-wide association studies. Cold Spring Harb Protoc. 2009;2009(12):pdb top66. doi:10.1101/pdb.top66.
139. Provine WB. Ernst Mayr: Genetics and speciation. Genetics. 2004;167(3):1041-6.
140. Vali U, Brandstrom M, Johansson M, Ellegren H. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. BMC Genet. 2008;9:8. doi:10.1186/1471-2156-9-8.
141. Ellis JR, Burke JM. EST-SSRs as a resource for population genetic analyses. Heredity. 2007;99(2):125-32.
142. Consortium STD, Estrada K, Aukrust I, Bjorkhaug L, Burtt NP, Mercader JM et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA. 2014;311(22):2305-14. doi:10.1001/jama.2014.6511.
143. Juul S. Epidemiologi og evidens. 7th ed. Danmark: København:Munksgaard; 2009.
144. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology

. 3rd ed. Philadelphia USA.

: Lippincott, Williams &Wilkins; 2008.

145. Fujimaki T, Oguri M, Horibe H, Kato K, Matsuoka R, Abe S et al. Association of a transcription factor 21 gene polymorphism with hypertension. Biomed Rep. 2015;3(1):118-22. doi:10.3892/br.2014.371.

146. Liang S, Liu X, Fan P, Liu R, Zhang J, He G et al. Association between Val158Met functional polymorphism in the COMT gene and risk of preeclampsia in a Chinese population. Arch Med Res. 2012;43(2):154-8. doi:10.1016/j.arcmed.2012.03.002.

147. Roten L, Fenstad M, Forsmo S, Johnson M, Moses E, Austgulen R et al. A low COMT activity haplotype is associated with recurrent preeclampsia in a Norwegian population cohort (HUNT2). Mol Hum Reprod. 2011;17(7):439-46.

148. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 2002;3(5):391-7. doi:10.1038/nrg796.

149. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-6. doi:10.1038/nature08250.

150. Ng PC, Kirkness EF. Whole genome sequencing. Methods Mol Biol. 2010;628:215-26. doi:10.1007/978-1-60327-367-1_12.

151. Ehrenreich H, Nave KA. Phenotype-Based Genetic Association Studies (PGAS)-Towards Understanding the Contribution of Common Genetic Variants to Schizophrenia Subphenotypes. Genes (Basel). 2014;5(1):97-105. doi:10.3390/genes5010097.

152. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev. 2002;11(6):505-12.

153. van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson SV et al. Whole-genome sequencing in health care. Recommendations of the European Society of Human Genetics. Eur J Hum Genet. 2013;21 Suppl 1:S1-5.

154. Asslaber M, Zatloukal K. Biobanks: transnational, European and global networks. Brief Funct Genomic Proteomic. 2007;6(3):193-201. doi:10.1093/bfgp/elm023.

155. Irgens LM. The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. Acta Obstet Gynecol Scand. 2000;79(6):435-9.

156. Roten LT, Thomsen LCV, Gundersen AS, Fenstad MH, Odland ML, Strand KM et al. Cohort profile: The Norwegian Preeclampsia Family Cohort Study. To be submitted. 2015.

157. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR et al. Cohort Profile: the HUNT Study, Norway. Int J Epidemiol. 2013;42(4):968-77. doi:10.1093/ije/dys095.

158. Holmen J, Midthjell K, Krüger Ø, Langhammer A, Holmen TL, Bratberg GH et al. The Nord-Trøndelag Health Study 1995-97 (HUNT 2): Objectives, contents, methods and participation. Norsk Epidemiologi. 2003;13(1):19-32.

159. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106(23):9362-7. doi:10.1073/pnas.0903103106.

160. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST et al. Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur J Hum Genet. 2014;22(1):144-7. doi:10.1038/ejhg.2013.96.

161. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007;39(10):1181-6. doi:10.1038/ng1007-1181.

162. Salkind NJ, Rasmussen K. Encyclopedia of Measurement and Statistics. SAGE Research methods2007.

163. Øian P, Henriksen T, Sviggum O. Hypertensive svangerskapskomplikasjoner. Veileder i Fødselshjelp. DnLF: The Norwegian Medical Association; 1998. p. 94-7.

164. Lorentzen B, Øian P, Staff A, Mortensen J. Preeklampsi. Veileder i Fødselshjelp. DnLF: The Norwegian Medical Association; 2006. p. 125-31.
165. Johnsen SL, Rasmussen S, Wilsgaard T, Sollien R, Kiserud T. Longitudinal reference ranges for estimated fetal weight. Acta Obstet Gynecol Scand. 2006;85(3):286-97.
166. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24(24):2938-9. doi:10.1093/bioinformatics/btn564.
167. Duggirala R, Williams JT, Williams-Blangero S, Blangero J. A variance component approach to dichotomous trait linkage analysis using a threshold model. Genet Epidemiol. 1997;14(6):987-92. doi:10.1002/(SICI)1098-2272(1997)14:6<987::AID-GEPI71>3.0.CO;2-G.
168. Almasy L, Blangero J. Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees. Am J Hum Genet. 1998;62(1198-1211).
169. DeCarlo LT. On the meaning and use of kurtosis. Psychol Methods. 1997;2(3):292-307.
170. Beasley TM, Erickson S, Allison DB. Rank-based inverse normal transformations are increasingly used, but are they merited? Behav Genet. 2009;39(5):580-95. doi:10.1007/s10519-009-9281-0.
171. Malay S, Chung KC. The choice of controls for providing validity and evidence in clinical research. Plast Reconstr Surg. 2012;130(4):959-65. doi:10.1097/PRS.0b013e318262f4c8.
172. Geller SE. International Classification of Diseasese9th revision coding for preeclampsia: How accurate is it? Am J Obstet Gynecol. 2004;190:1629-34.
173. Klemmensen ÅK. Diagnostic criteria and reporting procedures for pre-eclampsia: A national survey among obstetrical departments in Denmark. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2005;123:41-5.
174. Kristensen J, Langhoff-Roos J, Skovgaard LT, Kristensen F. Validation of the Danish Birth Registration.] Clin Epidemiol. 1996;49(8):893-7.

175. Salonen Ros H. Genetic effects on the liability of developing pre-eclampsia and gestational hypertension. Am J Med Genet. 2000;91:256-60.
176. Klungsoyr K, Harmon QE, Skard LB, Simonsen I, Austvoll ET, Alsaker ER et al. Validity of pre-eclampsia registration in the medical birth registry of norway for women participating in the norwegian mother and child cohort study, 1999-2010. Paediatr Perinat Epidemiol. 2014;28(5):362-71. doi:10.1111/ppe.12138.
177. Harlow FH, Brown MA. The diversity of diagnoses of preeclampsia. Hypertens Pregnancy. 2001;20(1):57-67. doi:10.1081/PRG-100104172.
178. Myatt L, Redman CW, Staff AC, Hansson S, Wilson ML, Laivuori H et al. Strategy for standardization of preeclampsia research study design. Hypertension. 2014;63(6):1293-301. doi:10.1161/HYPERTENSIONAHA.113.02664.
179. Cizmarova M, Kostalova L, Pribilincova Z, Lasabova Z, Hlavata A, Kovacs L et al. Rasopathies - dysmorphic syndromes with short stature and risk of malignancy. Endocr Regul. 2013;47(4):217-22.
180. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe: epidemiological update. Eur Heart J. 2013;34(39):3028-34. doi:10.1093/eurheartj/eht356.
181. Kearney PM, Whelton M, Reynolds K, Whelton PK, He J. Worldwide prevalence of hypertension: a systematic review. J Hypertens. 2004;22(1):11-9.
182. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2224-60. doi:10.1016/S0140-6736(12)61766-8.
183. Hayward C, Livingstone J, Holloway S, Liston WA, Brock DJ. An exclusion map for pre-eclampsia: assuming autosomal recessive inheritance. Am J Hum Genet. 1992;50(4):749-57.
184. Laivuori H, Lahermo P, Ollikainen V, Widen E, Haiva-Mallinen L, Sundstrom H et al. Susceptibility loci for preeclampsia on chromosomes 2p25 and 9p13 in Finnish families. Am J Hum Genet. 2003;72(1):168-77. doi:10.1086/345311.

185. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. BMC Public Health. 2009;9:88. doi:10.1186/1471-2458-9-88.

186. Murphy VE, Namazy JA, Powell H, Schatz M, Chambers C, Attia J et al. A meta-analysis of adverse perinatal outcomes in women with asthma. BJOG. 2011;118(11):1314-23. doi:10.1111/j.1471-0528.2011.03055.x.

187. Kallen B, Rydhstroem H, Aberg A. Asthma during pregnancy--a population based study. Eur J Epidemiol. 2000;16(2):167-71.

188. Wenger NK. What do the 2011 American Heart Association guidelines tell us about prevention of cardiovascular disease in women? Clin Cardiol. 2011;34(9):520-3. doi:10.1002/clc.20940.

189. Libby P, DiCarli M, Weissleder R. The vascular biology of atherosclerosis and imaging targets. J Nucl Med. 2010;51 Suppl 1:33S-7S. doi:10.2967/jnumed.109.069633.

190. Yang B, Fan S, Zhi X, Li Y, Liu Y, Wang D et al. Associations of MTHFR gene polymorphisms with hypertension and hypertension in pregnancy: a meta-analysis from 114 studies with 15411 cases and 21970 controls. PLoS ONE. 2014;9(2):e87497. doi:10.1371/journal.pone.0087497.

191. Gallacher J. Genetics of Complex Human Diseases: A Laboratory Manual. A Laboratory Manual. United States of America: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; 2009.

192. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet. 2002;11(20):2463-8.

193. Marian AJ. Molecular genetic studies of complex phenotypes. Transl Res. 2012;159(2):64-79. doi:10.1016/j.trsl.2011.08.001.

194. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009;41(6):666-76. doi:10.1038/ng.361.

195. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet. 2011;43(10):1005-11. doi:10.1038/ng.922.

196. International Consortium for Blood Pressure Genome-Wide Association S, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9. doi:10.1038/nature10405.

**Supplementary Table S1** SNP selection of genotyped SNPs

| **Genotyped snps** | **Replaced LD=1 rs#** | **Hypothesis** |
|---|---|---|
| rs1004467 | | Candidate Gene |
| rs10188442 | | Candidate Gene |
| rs10491334 | | Blood pressure regulation |
| rs10493339 | rs10493340 | Candidate Gene |
| rs10493340 | | Blood pressure regulation |
| rs10496288 | | Candidate Gene |
| rs10513308 | | Candidate Gene |
| rs10513311 | | Candidate Gene |
| rs10521209 | | Candidate Gene |
| rs10785581 | | Candidate Gene |
| rs10984096 | | Candidate Gene |
| rs10984338 | | Candidate Gene |
| rs11014166 | | Candidate Gene |
| rs11024074 | | Candidate Gene |
| rs11191548 | | Blood pressure regulation |
| rs11222084 | | Blood pressure regulation |
| rs11578482 | rs10493340 | Candidate Gene |
| rs11583678 | rs10493340 | Candidate Gene |

| | | |
|---|---|---|
| rs11646213 | | Candidate Gene |
| rs11671984 | | Candidate Gene |
| rs1173771 | | Blood pressure regulation |
| rs11823543 | | Blood pressure regulation |
| rs11953630 | | Blood pressure regulation |
| rs12046278 | | Candidate Gene |
| rs12210386 | rs3798440 | Candidate Gene |
| rs12211735 | rs3798440 | Candidate Gene |
| rs12239046 | | Candidate Gene |
| rs12258967 | | Blood pressure regulation |
| rs12522034 | | Candidate Gene |
| rs12940887 | | Candidate Gene |
| rs12946454 | | Blood pressure regulation |
| rs13002573 | | Blood pressure regulation |
| rs13082711 | | Blood pressure regulation |
| rs13107325 | | Blood pressure regulation and Candidate gene |
| rs13139571 | | Blood pressure regulation |
| rs1327235 | | Blood pressure regulation |
| rs13333226 | | Candidate Gene |
| rs1372662 | | Candidate Gene |
| rs1378942 | | Blood pressure regulation |

| | | |
|---|---|---|
| rs1399431 | | Candidate Gene |
| rs1446468 | | Blood pressure regulation |
| rs1458038 | | Blood pressure regulation |
| rs15285 | | Blood pressure regulation |
| rs1530440 | | Blood pressure regulation |
| rs1539019 | | Candidate Gene |
| rs1571093 | | Candidate Gene |
| rs1572299 | | Candidate Gene |
| rs16852030 | | Candidate Gene |
| rs16933812 | | Candidate Gene |
| rs16948048 | | Blood pressure regulation |
| rs16982520 | | Candidate Gene |
| rs16998073 | | Blood pressure regulation |
| | | |
| rs17221417 | | Candidate Gene |
| rs17249754 | | Blood pressure regulation |
| rs17367504 | | Blood pressure regulation |
| rs17477177 | | Blood pressure regulation |
| rs17608766 | | Blood pressure regulation |
| rs17773430 | | Blood pressure regulation |
| rs1799945 | | Blood pressure regulation |

| | | |
|---|---|---|
| rs1918974 | | Blood pressure regulation |
| rs1963982 | | Blood pressure regulation |
| rs200752 | | Candidate Gene |
| rs2070600 | | Candidate Gene |
| rs2071518 | | Blood pressure regulation |
| rs2121070 | | Blood pressure regulation |
| rs2266788 | | Blood pressure regulation |
| rs2384550 | | Blood pressure regulation |
| rs2469997 | | Candidate Gene |
| rs2509458 | | Blood pressure regulation |
| rs2521501 | | Blood pressure regulation |
| rs2782980 | | Blood pressure regulation |
| rs2806117 | | Candidate Gene |
| rs2820037 | | Candidate Gene |
| rs2841993 | | Candidate Gene |
| rs2906766 | | Candidate Gene |
| rs2932538 | | Blood pressure regulation |
| rs2954033 | | Blood pressure regulation |
| rs3096277 | | Blood pressure regulation |
| rs3184504 | | Blood pressure regulation and Candidate gene |
| rs3774372 | | Blood pressure regulation |

| | | |
|---|---|---|
| rs3798440 | | Candidate Gene |
| rs3798441 | rs3798440 | Candidate Gene |
| rs4132476 | | Candidate Gene |
| rs4273687 | rs943072 | Candidate Gene |
| rs4370013 | | Blood pressure regulation |
| rs4513773 | | Candidate Gene |
| rs4590817 | | Blood pressure regulation |
| rs5743289 | | Candidate Gene |
| rs5743594 | | Candidate Gene |
| rs628926 | | Candidate Gene |
| rs633185 | | Blood pressure regulation |
| rs6452524 | | Candidate Gene |
| rs6495122 | | Candidate Gene |
| rs6711736 | | Candidate Gene |
| rs6749447 | | Blood pressure regulation |
| rs6902871 | rs943072 | Candidate Gene |
| rs6905288 | | Candidate Gene |
| rs6921438 | rs4513773 | Candidate Gene |
| rs6940798 | rs943072 | Candidate Gene |
| rs6946944 | | Candidate Gene |
| rs7129220 | | Blood pressure regulation |

| | | |
|---|---|---|
| rs721640 | | Candidate Gene |
| rs7258592 | rs11671984 | Candidate Gene |
| rs7259148 | | Candidate Gene |
| rs729761 | | Candidate Gene |
| rs7513108 | | Candidate Gene |
| rs7591163 | | Blood pressure regulation |
| rs7638110 | | Blood pressure regulation |
| rs7696175 | | Candidate Gene |
| rs7767396 | rs4513773 | Candidate Gene |
| rs780093 | | Blood pressure regulation |
| rs805303 | | Blood pressure regulation |
| rs871606 | | Blood pressure regulation |
| rs910609 | | Candidate Gene |
| rs932272 | | Candidate Gene |
| rs932764 | | Blood pressure regulation |
| rs9350602 | rs3798440 | Candidate Gene |
| rs935334 | | Blood pressure regulation |
| rs943072 | | Candidate Gene |
| rs9472138 | | Candidate Gene |
| rs9472155 | | Candidate Gene |
| rs9472159 | rs4513773 | Candidate Gene |

| | | |
|---|---|---|
| rs991316 | | Candidate Gene |
| rs998584 | | Candidate Gene |
| Genotyped snps | Replaced LD. rs | Hypothesis |
| rs1004467 | | Candidate Gene |
| rs10188442 | | Candidate Gene |
| rs10491334 | | Blood pressure regulation |
| rs10493339 | rs10493340 | Candidate Gene |