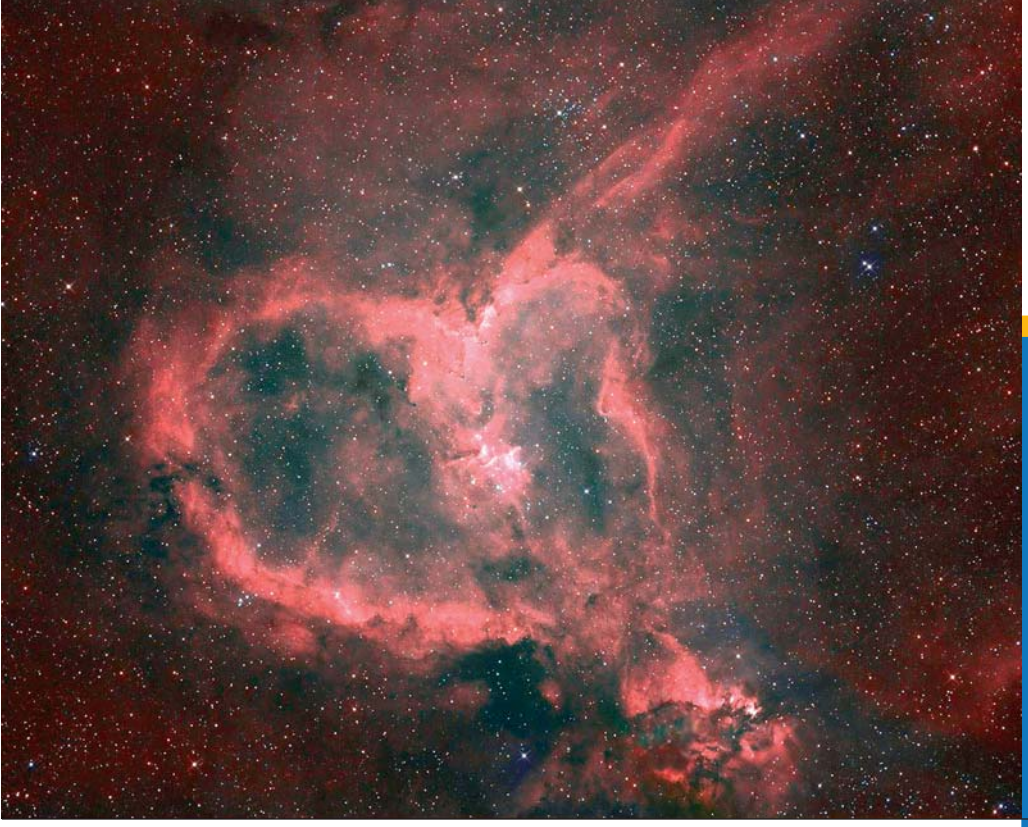

Mervi Alanne

Studies of the Genetic Epidemiology of Cardiovascular Disease:

Focus on Inflammatory Candidate Genes

Publications of the National Public Health Institute A 13/2008

Department of Molecular Medicine, National Public Health Institute,
Helsinki, Finland
and
Department of Medical Genetics, University of Helsinki, Finland

Helsinki, Finland 2008

Mervi Alanne

STUDIES OF THE GENETIC EPIDEMIOLOGY OF CARDIOVASCULAR DISEASE:

FOCUS ON INFLAMMATORY CANDIDATE GENES

ACADEMIC DISSERTATION

To be presented with the permission of the Medical Faculty, University of Helsinki, for public examination in the big lecture hall of the Haartman Institute, Haartmaninkatu 3, Helsinki, on May 16th, at 12 noon.

Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland

and

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2008

Publications of the National Public Health Institute KTL A13 / 2008

Julkaisija-Utgivare-Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 978-951-740-818-9
ISSN 0359-3584
ISBN 978-951-740-819-6 (pdf)
ISSN 1458-6290 (pdf)

Kannen kuva - cover graphic: Heart Nebula, copyright Matthew T. Russell
(edited by Risto Kinnunen)

Yliopistopaino
Helsinki 2008

Supervised by

Adjunct Professor Markus Perola
National Public Health Institute,
Department of Molecular Medicine and
University of Helsinki,
Department of Medical Genetics
Helsinki, Finland

Reviewed by

Adjunct Professor Tarja Laitinen
Helsinki University Central Hospital,
Clinical Research Unit of Pulmonary Diseases
Helsinki, Finland

and

Associate Professor Tomi Pastinen
McGill University and
Genome Quebec Innovation Centre
Montreal, Quebec, Canada

Opponent

Associate Professor Thomas Illig
GSF-National Research
Center for Environment and Health
Institute of Epidemiology
München-Neuherberg, Germany

To my family

Mervi Alanne, Studies of the genetic epidemiology of cardiovascular disease: focus on inflammatory candidate genes
Publications of the National Public Health Institute, A13/2008, 131 Pages
ISBN 978-951-740-818-9; 978-951-740819-6 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/4043

ABSTRACT

Cardiovascular disease (CVD) is a complex disease with multifactorial aetiology. Both genetic and environmental factors contribute to the disease risk. The lifetime risk for CVD differs markedly between men and women, men being at increased risk. Inflammatory reaction contributes to the development of the disease by promoting atherosclerosis in artery walls.

In the first part of this thesis, we identified several inflammatory related CVD risk factors associating with the amount of DNA from whole blood samples, indicating a potential source of bias if a genetic study selects the participants based on the available amount of DNA. In the following studies, this observation was taken into account by applying whole genome amplification to samples otherwise subjected to exclusion due to very low DNA yield.

We continued by investigating the contribution of inflammatory genes to the risk for CVD separately in men and women, and looked for sex-genotype interaction. In the second part, we explored a new candidate gene and its role in the risk for CVD. Selenoprotein S (*SEPS1*) is a membrane protein residing in the endoplasmic reticulum where it participates in retro-translocation of unfolded proteins to cytosolic protein degradation. Previous studies have indicated that SEPS1 protects cells from oxidative stress and that variations in the gene are associated with circulating levels of inflammatory cytokines. In our study, we identified two variants in the *SEPS1* gene, which associated with coronary heart disease and ischemic stroke in women. This is, to our knowledge, the first study suggesting a role of *SEPS1* in the risk for CVD after extensively examining the variation within the gene region.

In the third part of this thesis, we focused on a set of seven genes (angiotensin converting enzyme, angiotensin II receptor type I, C-reactive protein (*CRP*), and fibrinogen alpha-, beta-, and gamma-chains (*FGA*, *FGB*, *FGG*)) related to inflammatory cytokine interleukin 6 (IL6) and their association with the risk for CVD. We identified one variant in the *IL6* gene conferring risk for CVD in men and a variant pair from *IL6* and *FGA* genes associated with decreased risk. Moreover, we identified and confirmed an association between a rare variant in the *CRP* gene and lower CRP levels, and found two variants in the *FGA* and *FGG* genes associating

with fibrinogen levels. The results from this third study suggest a role for the interleukin 6 pathway genes in the pathogenesis of CVD and warrant further studies in other populations. In addition to the IL6 -related genes, we describe in this thesis several sex-specific associations in other genes included in this study. The majority of the findings were evident only in women encouraging other studies of cardiovascular disease to include and analyse women separately from men.

Keywords: Cardiovascular disease, association, proportional hazard models inflammation, single nucleotide polymorphism, DNA

Mervi Alanne, Studies of the genetic epidemiology of cardiovascular disease: focus on inflammatory candidate genes
Kansanterveyslaitoksen julkaisuja, A13/2008, 131 sivua
ISBN 978-951-740-818-9; 978-951-740819-6 (pdf-versio)
ISSN 0359-3584; 1458-6290 (pdf-versio)
http://www.ktl.fi/portal/4043

TIIVISTELMÄ

Sydän- ja verisuonisairaudet kuuluvat geneettisesti monitekijäisiin sairauksiin. Sekä perinteiset riskitekijät, kuten korkea kolesteroli ja miessukupuoli, että perintötekijät altistavat sydän- ja verisuonisairauksille. Tulehdusreaktio liittyy olennaisesti taudin kehitykseen vaikuttaen yhdessä muiden riskitekijöiden kanssa valtimoiden seinämien ateroskleroosin syntyyn.

Väitöskirjan ensimmäisessä osassa tunnistimme useita tulehdusreaktioon liittyviä sydän- ja verisuonisairauksien riskitekijöitä, jotka ovat yhteydessä kokoverestä eristettävän DNA:n saantoon. Tuloksen perusteella voidaan olettaa, että tutkittavien valinta DNA-määrän perusteella geneettisiin epidemiologisiin tutkimuksiin saattaa aiheuttaa harhan tutkimusasetelmassa.

Väitöskirjan seuraavissa osatöissä huomioimme ensimmäisen työn tulokset ja monistimme koko genomin kattavalla monistusmenetelmällä DNA-määrän lisäämiseksi näytteet, jotka vähäisen DNA-määrän johdosta olisivat joutuneet poissuljetuiksi.

Seuraavaksi tutkimme tulehdusreaktioon liittyvien kandidaattigeenien vaikutusta sydän- ja verisuonisairauksien riskiin kahdessa suomalaisessa aineistossa. Toisessa osatyössä syvennyimme uuteen mielenkiintoiseen ehdokasgeeniin, selenoproteiini S:aan (*SEPS1*). SEPS1 on kalvoproteiini, joka sijaitsee solulimakalvostossa. Se osallistuu väärinlaskostuneiden proteiinien takaisinkuljetukseen solulimaan. Aiemmat tutkimukset ovat osoittaneet, että SEPS1 suojaa soluja hapettumisen aiheuttamilta vaurioilta ja että geneettinen vaihtelu geenin alueella on yhteydessä tulehdusreaktiota välittävien liukoisten sytokiinien pitoisuuksiin verenkierrossa.

Tutkimuksemme osoitti, että kaksi polymorfiaa *SEPS1* geenin alueella nostavat koronaaritaudin ja iskemiseen aivohalvauksen riskiä naisilla. Tämä on ensimmäinen kansainvälisellä tasolla julkaistu tutkimus, jossa on tutkittu kattavasti geneettistä variaatiota *SEPS1* geenin alueella ja osoitettu sen vaikuttavan sydän- ja verisuonisairauksien riskiin.

Kolmannessa osatyössä keskityimme seitsemään tulehdusreaktiota säätelevään interleukiini 6 (IL6) sytokiiniin liittyvään geeniin (angiotensiiniä konvertoiva

entsyymi, angiotensin II reseptorityyppi I, C-reaktiivinen proteiini (*CRP*), sekä fibrinogeenin alfa-, beta-, ja gammaketjut (*FGA*, *FGB*, *FGG*)). Tutkimuksen perusteella *IL6* geenin polymorfia on yhteydessä suurentuneeseen sydän- ja verisuonisairauksien riskiin miehillä. Lisäksi toinen polymorfia samassa geenissä yhdessä *FGA* geenissä olevan polymorfian kanssa vaikutti suojaavan sydän- ja verisuonitaudeilta. Edelleen, yhtenevästi aiempien tutkimustulosten kanssa osoitimme *CRP* geenissä sijaitsevan variantin olevan yhteydessä pienentyneisiin CRP pitoisuuksiin sekä tunnistimme kaksi fibrinogeenipitoisuuteen vaikuttavaa varianttia *FGA* ja *FGG* geeneissä.

Tulosten perusteella IL6:een liittyvillä geeneillä on vaikutusta sydän- ja verisuonisairauksien riskiin ja riskitekijöihin, IL6:n ja sen liitännäisgeenien tarkempi rooli sydän- ja verisuonitaudeissa edellyttää kuitenkin lisätutkimuksia muissa populaatioissa. Näiden geenien lisäksi, kuvaamme väitöskirjassa useita kolmanteen osatyöhön liittyvien geenien yhteen sukupuoleen rajoittuneita assosiaatiolöydöksiä. Näistä useimmat esiintyivät naisilla ja korostavat miehillä ja naisilla erikseen tehtävän sydän- ja verisuonitaudin tutkimuksen tärkeyttä.

Avainsanat: Sydän- ja verisuonitauti, assosiaatiotutkimus, suhteellisen vaaran malli, tulehdusreaktio, yhden nukleotidin muutos, DNA

CONTENTS

List of original publications **9**

1 Introduction **10**

2 Review of the literature **11**

2.1 CARDIOVASCULAR DISEASE 11

2.1.1 Atherosclerosis – an inflammatory disease 12

2.1.2 Etiology of cardiovascular disease 17

2.1.3 Inflammatory risk factors 21

2.2 GENETIC EPIDEMIOLOGY OF COMPLEX DISEASES 24

2.2.1 DNA, genes and the human genome 24

2.2.2 Genome-wide statistical methods for identifying genes underlying complex diseases 27

2.2.3 Candidate gene -approach 33

2.2.4 Evaluation of the statistical significance of association studies 34

2.2.5 Study populations, subjects and biological samples 36

2.3 GENETICS OF CARDIOVASCULAR DISEASE 40

2.3.1 Genetic component 40

2.3.2 Important lessons from monogenic disorders 41

2.3.3 Family-based findings on susceptibility loci for common CVD 42

2.3.4 Previous findings in association studies 45

2.3.5 Sex-specific association results 48

2.4 INFLAMMATORY CANDIDATE GENES OF CARDIOVASCULAR DISEASE 49

2.4.1 Selenoprotein S 49

2.4.2 Interleukin 6 -related genes 51

3 Aims of the study **55**

4 Materials and methods **56**

4.1 STUDY SUBJECTS 56

4.1.1 Health 2000 survey 56

4.1.2 FINRISK 92 and 97 surveys 56

4.2 LABORATORY PROCEDURES AND STATISTICAL ANALYSIS 57

4.2.1 Methods in publications 57

4.2.2 Biochemical markers 57

4.2.3 DNA extraction and quality control 60

4.2.4 Candidate genes and variant selection 60

4.2.5 Variant genotyping and quality control 61
4.2.6 Bioinformatic analysis of transcription factor binding 61
4.2.7 Electrophoretic mobility shift assay 64
4.2.8 Statistical analysis 65
4.2.9 Analytical approach for joint analysis of genetic risk factors 67

5 Results and discussion 68

5.1 CARDIOVASCULAR RISK FACTORS AND ASSOCIATION WITH THE VARIATION IN DNA YIELD FROM WHOLE BLOOD SAMPLES (I) 68

5.2 INFLAMMATORY CANDIDATE GENE SELENOPROTEIN S AND THE RISK FOR CARDIOVASCULAR DISEASE (II) 73

5.2.1 Variation in the *SEPS1* gene and the risk for coronary heart disease and ischemic stroke 73

5.2.2 Variation in the *SEPS1* gene and quantitative phenotypes 75

5.2.3 Allele specific binding of nuclear extract according to *SEPS1* variant *rs7178239* 75

5.2.4 Conclusions and discussion of study II 78

5.3 INTERLEUKIN 6 -RELATED CANDIDATE GENES AND SEX-SPECIFIC DIFFERENCES IN THE RISK FOR CARDIOVASCULAR DISEASE (III) 82

5.3.1 Association of gene variants in the interleukin 6 -related pathway with cardiovascular disease and -risk factors 82

5.3.2 Joint analysis of interleukin 6 -related genes 83

5.3.3 Association with CRP, IL6 and fibrinogen levels 85

5.3.4 Sex-specific association with CVD risk factors 88

5.3.5 Sex-specific association of other candidate genes 89

5.3.6 Power analysis of the case-cohort set 89

5.3.7 Conclusions and discussion of Study III 90

6 Concluding remarks 96

7 Acknowledgments 99

8 References 102

ABBREVIATIONS

α	Significance level
ABCA1	ATP-binding cassette transporter A1
ApoB-100	Apolipoprotein B-100
ACE	Angiotensin converting enzyme
ACS	Acute coronary syndrome
Ang I	Angiotensin I
ApoE	Apolipoprotein E
AT1R, *AGTR1*	Angiotensin II receptor type I (protein, gene)
ATP	Adenosine triphosphate
bp	Base pair
BMI	Body-mass index
CAD	Coronary artery disease
CDCV	Common disease, common variant
CETP	Cholesteryl ester transfer protein
CHD	Coronary heart disease
CI	Confidence interval
CNV	Copy-number variant
CVD	Cardiovascular disease
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
ERSE	Endoplasmic reticulum stress response element
FCHL	Familial combined hyperlipidemia
FDR	False discovery rate
FGA, FGB, FGG	Fibrinogen alpha, beta, gamma
FLAP, *ALOX5AP*	5-lipoxygenase activation protein (protein, gene)
GWA	Genome-wide association
HDL	High-density lipoprotein
HR	Hazard ratio
HRT	Hormone replacement therapy
hsCRP	High sensitivity C-reactive protein
htSNP	Haplotype tagging single nucleotide polymorphism
I/D	Insertion/deletion
ICAM1	Intercellular adhesion molecule 1
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th Revision
IL-1β	Interleukin 1 beta
IL6	Interleukin 6
IL10	Interleukin 10
IMT	Intima media thickness
INF-γ	Interferon gamma
ITGB3	β3 integrin

kb	Kilo base
LD	Linkage disequilibrium
LDL	Low density lipoprotein
LDLR	Low density lipoprotein receptor
LGALS2	Galectin-2
LPL	Lipoprotein lipase
LTA	Lymphotoxin alpha
LTA4	Leucotriene A4
MetS	Metabolic syndrome
MI	Myocardial infarction
MPRCA	Multiply-primed rolling circle amplification
mRNA	Messenger ribonucleic acid
MTHFR	Methylenetetrahydrofolate reductase
N	Number
NF-κB	Nuclear factor kappa-B
OR	Odds ratio
PAI1	Plasminogen activator inhibitor-1
PCR	Polymerase chain reaction
RAS	Renin-angiotensin-system
SAA	Serum amyloid A
SD	Standard deviation
SEPS1	Selenoprotein S
SNP	Single nucleotide polymorphism
SNPSpD	Single nucleotide polymorphism spectral decomposition
TE	Tris-EDTA
TF	Transcription factor
TIA	Transient ischemic attack
TNF-α	Tumor necrosis factor alpha
UPR	Unfolded protein response
USF1	Upstream transcription factor 1
UTR	Untranslated region
VCAM1	Vascular cell adhesion molecule 1
VSMC	Vascular smooth muscle cell
WGA	Whole genome amplification
WHO	World Heath Organization
WHR	Waist/hip ratio

Gene names are emphasised with italics in the text.

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles referred to in the text by their Roman numerals:

I. Alanne M, Salomaa V, Saarela J, Peltonen L, Perola M. DNA extraction yield is associated with several phenotypic characteristics: results from two large population surveys. Journal of Thrombosis and Haemostasis 2004; 2: 2069–71

II. Alanne M, Kristiansson K, Auro K, Silander K, Kuulasmaa K, Peltonen L, Salomaa V and Perola M. Variation in the selenoprotein S gene locus is associated with coronary heart disease and ischemic stroke in two independent Finnish cohorts. Human Genetics 2007;122(3-4): 355–65

III. Silander K, Alanne M*, Kristiansson K*, Saarela O, Ripatti S, Auro K, Karvanen J, Kulathinal S, Saarela J, Ellonen P, Kuulasmaa K, Perola M, Salomaa V and Peltonen L. Genetic risk profiles for cardiovascular disease in men and women: a candidate gene study. *Submitted*

* These authors contributed equally to this work

Some previously unpublished data are also presented

1 INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of death among the chronic non-communicable diseases and poses a significant public health burden worldwide. The three major manifestations of atherosclerotic CVD are coronary heart disease (CHD), ischemic stroke and peripheral arterial disease. Atherosclerosis develops as a result of a complex interplay between endothelial dysfunction, inflammation, and coagulation and fibrinolysis. The atherosclerotic plaque can grow either towards the vessel lumen or into the vessel wall. Subsequently, if the plaque does not reduce the blood flow and oxygen supply to the surrounding tissue, CVD can be completely asymptomatic whereas a stenotic plaque often results in angina pectoris. Myocardial infarction or ischemic stroke, are consequences of occluding thrombi diminishing the blood flow. This most often takes place on the surfaces of non-stenotic plaques.

CVD is a heritable trait, heritability estimates for CHD and ischemic stroke death ranging from 0.3 to 0.6. Thus far many genes and chromosomal regions have been suggested as candidates conferring susceptibility to the disease. The common form of CVD is a late-onset multifactorial disease, and as with other multifactorial diseases, the hunt for susceptibility genes has not been very successful. Consistent findings between populations and study samples have been rare until the very recent genome-wide association studies. The underlying genes most likely contribute each with a small effect and exhibit a wide range of allele frequencies.

The purpose of this study was to investigate the contribution of selected inflammatory genes to the risk for CVD. The incidence of CVD is distinctly different between men and women, women being at lower risk, and in this study we aimed to investigate the association separately in both sexes and possible sex-genotype interactions. Since the genes in Study III were selected based on previous knowledge of their biological relationships, we also looked for interactions within a distinct pathway related to cytokine interleukin 6. Furthermore, given that genetic epidemiological studies are vulnerable to many kinds of biases, we wanted to investigate one source of potential bias, the association between inflammatory risk factors of CVD and the amount of DNA, the molecule to which every genetic epidemiological study is based on. The following review will focus on the existing knowledge of the genetic background of cardiovascular disease and on available strategies to identify genes underlying susceptibility to common forms of cardiovascular disease.

2 REVIEW OF THE LITERATURE

2.1 Cardiovascular disease

Coronary heart disease (CHD), ischemic stroke and peripheral arterial disease are the three major manifestations of atherosclerotic cardiovascular disease (CVD). In the year 2005, CVD was the leading cause of death among the chronic non-communicable diseases accounting for 30% of all deaths both in Finland and worldwide (World Health Organization 2006, Statistics Finland 2007). Although the incidence of the disease is declining, ischemic heart disease and cerebrovascular disease are projected to continue to be the leading causes of death also for the next few decades (Mathers and Loncar 2006). The incidence of CVD differs markedly between the sexes: the lifetime risk at the age of 40 for developing CHD is approximately one in two for men and one in three for women. The corresponding figures for stroke are one in six for men and one in five for women (Tunstall-Pedoe et al. 1994, Jousilahti et al. 1999, Lloyd-Jones et al. 1999, Seshadri et al. 2006). Moreover, the age-of-onset of CHD is 7 to 10 years higher for women than for men (Bello and Mosca 2004).

Patients with coronary heart disease may be entirely asymptomatic or present a wide range of symptoms from chest pain to dyspnea and fatigue (Table 1). Differences in symptoms, pathological changes and electrophysiology of the heart divide CHD clinically into a) acute coronary syndrome (ACS) including unstable angina pectoris and both non-ST and ST segment elevation myocardial infarction (MI) and b) stable angina. In clinical practice, it is important to separate the two types of MI and some exclude non-ST elevation MI from ACS. In this thesis, it is included when discussing ACS. The essential cause of the ACS is the abrupt lack of oxygen supply to the heart muscle resulting from a thrombotic occlusion. The thrombotic clot can either originate from a distant site, such as the left atrium or carotid artery, or form *in situ* as a result of the clotting reaction. Systemic factors, such as blood hyperthrombogenicity or rheology (ie. alterations in the blood flow due to, for example, vasoconstriction) can also contribute to the formation of a thrombotic clot. Stable angina pectoris is often associated with exertion and is relieved by rest. The inability of the coronary arteries to respond to an increased demand for oxygen may be due to stenosis in one or more coronary arteries or a vasospasm on a coronary artery lesion. (Grant and Humphries 1999, Goldman and Ausiello 2004)

Stroke is a clinically heterogeneous disease, which is characterized by a neurological deficit lasting for more than 24 hours or leading to death. The underlying cause is either spontaneous hemorrhage or lack of oxygen due to reduced blood flow leading

to ischemic cerebral infarction. Atherosclerosis in the carotid arteries is the most common aetiology of the ischemic strokes. Other causes include cardioembolism and atherosclerosis of the small arteries. If the symptoms of ischemia are only transient, the disease is diagnosed as a transient ischemic attack (TIA). The symptoms related to cerebral haemorrhage are due to either pressure on surrounding tissues or the toxic effects of the blood. (Frizzell 2005)

Table 1. ***Typical coronary heart disease symptoms.***

Crushing or pressing chest pain or discomfort
Radiating pain to the left arm, neck or back
Pain symptoms increasing with exertion
Palpitations, dizziness, syncope
Fatigue, weakness

In genetic epidemiological studies using registries for event identification, the definition of CVD often includes coronary heart disease and ischemic cerebrovascular disease but does not include peripheral artery disease and excludes hemorrhagic stroke and TIA (Laatikainen et al. 2007). Diagnoses based on the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) (World Health Organization 2005) for CHD are: I20-I25, I46, R96, R98 and for stroke I60-I69, G45. In stroke, atherosclerosis contributes mainly to episodes related to ischemia and infarction defined with ICD-10 diagnosis classes I63-I64. Table 2 presents, as an example, diagnostic classification from a study aiming to identify genetic risk factors for CHD and ischemic stroke based on the WHO MONICA project.

2.1.1 Atherosclerosis – an inflammatory disease

Atherosclerotic changes in the arteries of the heart, brain and extremities cause atherosclerotic cardiovascular disease. A complex interplay between dyslipidemia, inflammation, coagulation, fibrinolysis and endothelial dysfunction is involved in the progression of the disease. The intima is the arterial wall layer principally involved in the development of atherosclerosis, although secondary changes are occasionally seen in the media. (Ross and Glomset 1976a, Ross and Glomset 1976b) (Figure 1) The atherosclerotic plaque can grow towards the lumen of the vessel or deeper into the vessel wall (Libby et al. 2002). An acute myocardial infarction or ischemic stroke is a consequence of a thrombus occluding the blood flow. About two-thirds of ACS are caused by occlusive thrombosis on a non-stenotic plaque, and in about one-third, the thrombus occurs on the surface of a stenotic plaque (Falk et

al. 1995). The development of atherothrombosis is a complex process and detailed molecular mechanisms of all the different stages are yet to be fully characterised. However, the key steps in the pathogenesis of the disease are known.

Table 2. ***Diagnostic classification of fatal and non-fatal cardiovascular disease events.*** *An example from a genetic epidemiological study in Finland. The classification is based on the recommendations by the WHO MONICA project.*

Disease classification	ICD-10[a]	Non-fatal	Fatal
Coronary events			
Unstable angina	I20.0	x	
Acute myocardial infarction	I21	x	x
Subsequent myocardial infarction	I22	x	x
Certain current complications following acute myocardial infarction	I23		x
Other acute ischaemic heart diseases	I24		x
Chronic ischaemic heart disease	I25		x
Cardiac arrest	I46		x
Other sudden death, cause unknown	R96		x
Unattended death	R98		x
Other ill-defined and unspecified causes of mortality	R99		x
Stroke events			
Cerebral infarction	I63	x	x

[a] ICD-10: International Statistical Classification of Diseases and Related Health Problems, 10th Revision
Based on Komulainen et al. 2006.

Injury or dysfunction of the endothelium is a key event in the evolution of atherosclerotic plaques. Atherosclerosis develops in the blood vessels at the site of low sheer stress such as curvatures, bifurcations and branch points (Feldman and Stone 2000). At these sites, the shape of the endothelial cells has transformed from an elongated to a more rounded shape with discontinuous tight junctions, irregularly shaped gap junctions and a lack of zonular type of tight junctions, which prevent permeation of serum proteins. These abnormalities may indicate endothelial dysfunction and activation that leads to increased intracellular permeability (Okano and Yoshida 1993, Yoshida et al. 1995). It has also been shown that lipid deposits are more likely to form in low shear stress areas than in areas of elevated shear stress (Zand et al. 1999). Together with other risk factors, such as increased blood pressure, elevated cholesterol levels, toxic substances from tobacco smoking, advanced glycation end products associated with diabetes mellitus, vasoactive amines and immune complexes – shear stress forces induce endothelial dysfunction.

Endothelial dysfunction is linked fundamentally to the initial formation of atherosclerosis via increased permeability of the endothelium, reduced generation of nitric oxide and elevated expression of adhesion molecules; it facilitates vessel wall entry and oxidisation of circulating lipoproteins, as well as infiltration of monocytes and other cells of the immune system (Fuster et al. 2005). In high concentrations, the lipoproteins, particularly low-density lipoprotein (LDL) particles, are trapped in the intima. Self-aggregation of LDL particles and interactions between proteoglycans and extracellular matrix proteins with apolipoprotein B 100 (*APOB*), apolipoprotein E (*APOE*) or lipoprotein lipase (*LPL*) further facilitate their retention (Khalil et al. 2004). Once in the intima, they are modified in many ways: they form complexes with proteglycans and immune cells, become oxidized, hydrolysed and glycated, proteolysis degrades the apoB-100 molecules, and other enzymes further modify LDL particles to form cholesterol-enriched lipid droplets. These modifications, particularly minimally oxidised LDL, stimulate endothelial cells to secrete pro-inflammatory and adhesion molecules, and monocyte chemotactic protein 1 that facilitate the recruitment and transmigration of monocytes and lymphocytes to the arterial wall. Oxidation also makes LDL recognisable to macrophage scavenger receptors leading to their internalisation and macrophage cholesteryl ester accumulation, i.e. foam cell formation. (Tabas 1999)

The size of the LDL particle appears to play a role in defining the atherogenic properties of the lipoprotein: LDL is the principal cholesterol-carrying lipoprotein in the plasma but also larger particles, such as chylomicrons and very-low-density lipoprotein can be atherogenic, especially their remnants (Krauss 1998). It seems, however, that small sized LDL particles are more atherogenic than particles with larger size and confer greater CVD risk (Austin et al. 1988, Veniant et al. 2001). High-density lipoprotein (HDL) balances this atherogenic process by removing cholesterol from the arterial intima macrophages via a process of reverse cholesterol transport and by preventing LDL oxidation via HDL-associated paraoxonase activity (Reddy et al. 2001). The accumulation of modified lipids and inflammatory cells leads to formation of a fatty streak, the early stage of atherosclerotic lesion.

Fatty streaks are already present in early childhood and adolescence and start to develop into more advanced lesions in early adulthood (McGill et al. 2000). A fibrous lesion has characteristically an increasing pool of lipids released from the apoptotic macrophages and it accumulates vascular smooth muscle cells (VSMC) and VSMC-derived extracellular matrix over the core region of lipid droplets and foam cells. Dendritic cells, mast cells and a small number of B cells are also present in the plaque. (Jonasson et al. 1986, Hansson and Libby 2006) Inflammation is crucially involved also in the progression of the fibrous plaques. Inflammatory cells, macrophages and T-lymphocytes continue to infiltrate to the lesion and proliferate at

the site. Macrophages secrete cytokines (i.e. interleukin 1β (IL-1β), interleukin 6 (IL6), interleukin 8, tumor necrosis factor alpha (TNF-α)), chemokines and growth-regulating molecules. These signalling molecules stimulate VSMCs to increase their proliferation rate, to migrate to the site of inflammation and to produce extracellular matrix. (Ross 1999)

Of the lymphocytes, Th1-type T cells, that secrete pro-inflammatory cytokines interferon-γ (INF-γ), TNF-α, and interleukins 12, 15 and 18, is the class most frequently found in the lesion. Other important inflammatory mediators are the cell-surface proteins CD40 and CD40 ligand. In addition to T and B cells, macrophages, endothelial cells, VSMCs and platelets also express CD40 (Mach et al. 1997, Henn et al. 1998). Like pro-inflammatory cytokines, CD40 ligation triggers the secretion of other cytokines, matrix metalloproteinases and macrophage inflammatory protein 1α, and causes macrophages to express procoagulant tissue factor, and endothelial cells to express adhesion molecules E-selectin, vascular cell adhesion molecule 1 (VCAM1) and intercellular adhesion molecule 1 (ICAM1). This further enhances the inflammatory reaction and recruits more lymphocytes and monocytes to the site. Anti-CD40L antibody treatment in LDL receptor deficient mice has led to more stable plaque by reducing lipid content and the number of macrophages and increasing the relative VSMC and collagen content (Schönbeck et al. 2000). Risk factors such as hypertension and elevated homocysteine and hormone levels also contribute to the development of a fibrous lesion: elevated homocysteine causes inhibition of endothelial cell proliferation and stimulates VSMC proliferation and collagen production (Tsai et al. 1994, Majors et al. 1997); Angiotensin II, which is a component of the renin-angiotensin system that regulates blood pressure, stimulates VSMC growth (van Kleef et al. 1996) and the expression of VCAM1 by endothelial cells (Pueyo et al. 2000) and the secretion of matrix metalloproteinases (Browatzki et al. 2005); oestrogen, on the other hand, is an anti-atherogenic hormone that has a beneficial effect on plasma lipoprotein levels, stimulates prostacyclin and nitric oxide production thereby enhancing endothelial function. Oestrogen also reduces VSMC proliferation (Bush et al. 1998, Dubey et al. 2005, Mendelsohn and Karas 2005).

A fibrotic plaque can be stable and asymptomatic for years. Current view on the development of ACS is that the unstable complicated atheroma is the most one likely to rupture and is the main culprit for events leading to MI. Vulnerable plaques exhibit uneven thinning and erosion of the fibrous cap. Thinning occurs most often in the shoulder region of the plaque where inflammatory cells infiltrate to the lesion (Lee and Libby 1997). This region also has more metalloproteinase expression that increases degradation of stabilizing matrix proteins (Shah and Galis 2001). In chronic stable ischemia, excess matrix accumulation is the primary mechanism of

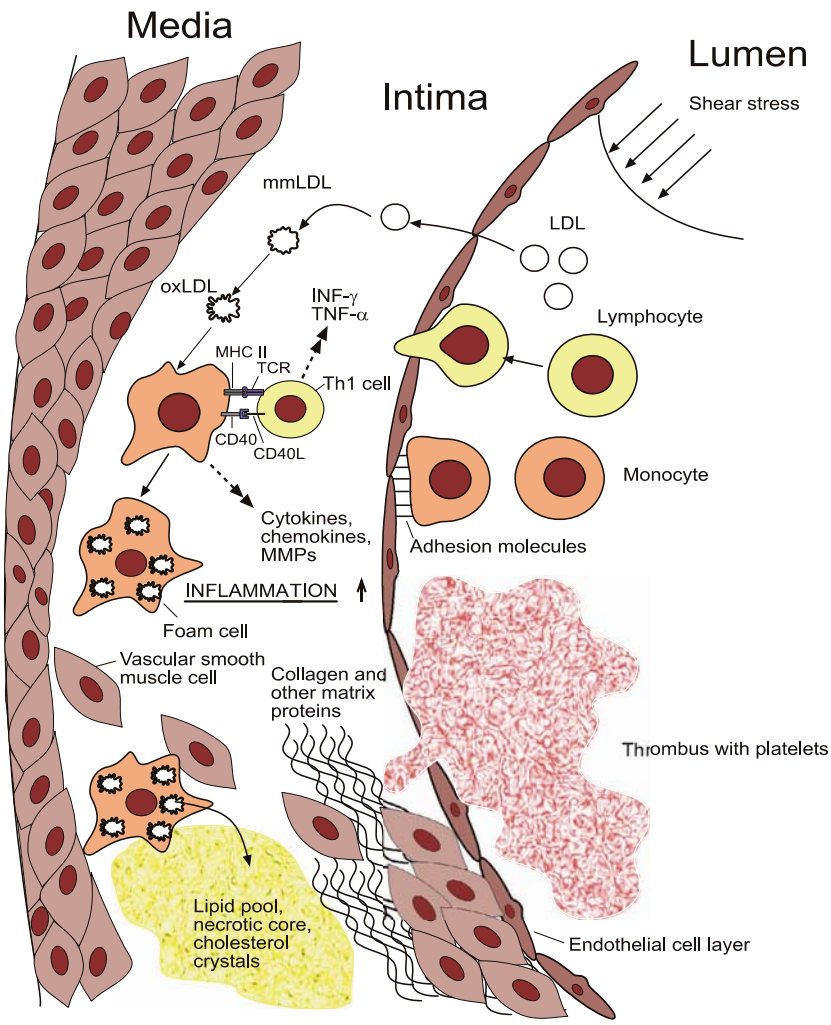

Figure 1. ***A schematic illustration of the development of a healthy vessel wall (above) to a complicated atherosclerotic lesion (below).*** *During the initial phases of atherosclerosis, low-density lipoproteins (LDL) enter the intima and become first minimally modified LDL (mmLDL) and then oxidised LDL (oxLDL). Macrophages internalize oxLDL using scavenger receptors and present it to T cells with major histocompatibility complex (MHC II) receptors. Activated macrophages secrete pro-inflammatory cytokines, proteases, procoagulants and pro-apoptotic factors. Accumulation of oxLDL in macrophages transforms them into cholesterol-loaded foam cells and during their apoptosis the lipid content is relased to the intima. Activated Th1 cells secrete interferon-γ (IFN-γ) and tumor necrosis factor alpha (TNF-α) which increase permeability, the propensity for thrombus formation and expression of adhesion molecules. Vascular smooth muscle cells migrate and profilerate in the lesion and form a fibrous cap over the lipid pool together with extracellular matrix proteins. Rupture of the vulnerable plaque leads to the formation of a thrombus.*

occlusion of the atherosclerotic artery lumen. In contrast, it seems that unstable lesions may fail to synthesize a sufficient mass of matrix to provide strength to the fibrous cap as the number of matrix-producing VSMCs reduces in the fibrotic cap (Davies et al. 1993). Increased calcification, neovascularisation and intra-plaque haemorrhage may also influence the plaque stability (Fuster et al. 2005, Ijäs et al. 2007). In addition to inflammatory activity and vascular remodelling, high shear stress, which occurs at the throat (minimal luminal diameter) of the plaque, may induce endothelial damage and increased platelet deposition, and promote plaque rupture (Feldman and Stone 2000). Either subtle erosion or a more severe rupture of the endothelial layer exposes the underlying tissue to the blood flow. Contact between von Willenbrand factor and collagen and platelet glycoproteins (primary haemostasis) launches the clotting reaction and leads to thrombus formation. Another mechanism involves circulating factor VII(a) and tissue factor that is present on the surfaces of the cells within the plaque. Depending on the severity of the wound, fibrinolysis can degrade the thrombus and heal the wound, while simultaneously making the plaque more complicated. Alternatively, the forming clot is so severe that it causes ACS (Spronk et al. 2004). Figure 1 presents a schematic and simplified view of the atherosclerotic lesion development and the major participating agents.

2.1.2 Aetiology of cardiovascular disease

Cardiovascular disease is a multifactorial disease with both environmental and genetic risk factors contributing to its development (Figure 2). The most important risk factors are age, male sex, smoking, high blood pressure, a history of MI, type II diabetes and elevated blood cholesterol levels (Goldman and Ausiello 2004). Metabolic syndrome (MetS), defined as a co-occurrence of at least three out of five cardiometabolic risk factors: abdominal obesity, hypertriglyceridemia, low HDL cholesterol, high blood pressure, and high fasting glucose, confers a significant increase in risk for CVD. Currently several diagnostic criteria can be used and a more detailed definition of MetS is available from the World Health Organization 1999, NCEP 2001, Balkau et al. 2002 and Alberti et al. 2006. Population based studies have estimated that individuals with MetS have approximately a two-fold increased risk for MI (Smith 2007). In middle-aged Finnish men, MetS is associated with a 2-3 -fold increase in risk for CVD (CHD and stroke) and approximately 2-fold increased risk for total mortality (Lakka et al. 2002). A similar estimate of a 2-fold increased risk for self-reported CVD has been reported by Ninomiya and colleagues (2004) for both sexes in the United States of America. Furthermore, independent trait components insulin resistance, hypertriglyceridemia, and hypertension are independently and significantly associated with CVD.

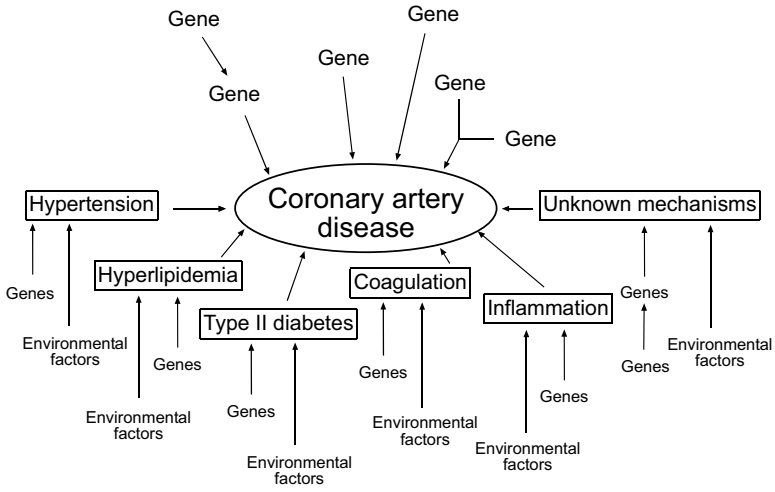

Figure 2. ***Multifactorial genesis of coronary artery disease.*** *Both vascular, environmental and genetic risk factors contribute to the disease.*

Primary and secondary prevention in most western countries has achieved a drastic decrease in the incidence of first CVD events and the occurrence of recurrent events by better treatment (Salomaa et al. 2003) and by being able to reduce cholesterol levels, hypertension and the frequency of smoking (Vartiainen et al. 2000). Recently, trait components of MetS, increased body weight and type II diabetes, have gained importance (Smith 2007). In a recent study of English men, Nanchahal and colleagues (2005) reported that among socioeconomic, anthropometric and life-style risk factors, overweight and obesity now dominate the CHD risk factors in men: excess body weight attributed to 47% of high risk CHD (≥15% 10-year risk) and lack of educational qualifications, low income, smoking, and physical inactivity accounted jointly for 31%. In Finland as well, obesity is independently from other risk factors associated with an increased risk for CHD both in men and in women (Jousilahti et al. 1996b). In addition to genetic susceptibility and traditional risk factors, health behaviour, such as diet and physical activity, and other environmental risk factors, infections and intra-uterine environment modify the risk for CVD (Lusis 2003). It is important to note that many CVD risk factors also have a genetic component (Table 3) and that genes affect also many environmental risk factors. The underlying mechanisms by which different risk factors affect the risk for CVD may be related directly to the development of atherosclerosis or indirectly by

affecting the development other risk factors such as insulin resistance or hypertension.

Although the risk for CVD is clearly associated with ethnicity, population studies suggest that differences in environmental risk factors largely explain the differences in CVD incidence between populations. For example, the incidence is much lower in the Japanese compared to the American population, but has increased significantly in Japanese individuals who have moved to America and adopted a western life-style (Robertson et al. 1977). Based on the World Health Organization (WHO) MONICA study the differences, however, cannot be fully be explained by differences in the known environmental risk factors (Tunstall-Pedoe 2003).

Table 3. ***Heritability estimates for cardiovascular disease risk factors.***

Risk factor	Heritability[a]
Type II diabetes	40–80%
Total cholesterol	40–60%
Systolic blood pressure	50–70%
Diastolic blood pressure	50–65%
HDL-cholesterol	45–75%
Triglycerides	40–80%
Body-mass index	25–60%
Lipoprotein-a levels	90%
Homocysteine levels	45%
Fibrinogen levels	20–50%
C-reactive protein	30–40%

[a] Modified from Lusis (2003) with heritability estimates also from Austin et al. 2004, Pankow et al. 2001 and Vickers et al. 2002.

Sexual dimorphism in the aetiology of CVD

The underlying causes in the different risk for CVD between men and women are numerous. The physiological differences are apparent already in childhood. For example, the constitution of intimal tissue has been shown to be different between male and female children among Ashkenazy children but differences were less apparent among those of Bedouin origin (Vlodaver et al. 1969). Also, already in early adulthood, differences in the extent of fatty streaks and raised lesions between men and women follow the pattern of adults, that is, the abdominal aortas of women have more fatty streaks than abdominal aortas of men but they have an equal extent of raised lesions, and that the coronary arteries of men have a similar extent of fatty streaks as women but they have more raised lesions. (McGill et al. 2000)

CVD risk factors are the same for both sexes, but the frequency and relative impact differs. For example, it appears that in younger women body-mass index (BMI) does not correlate positively with the extent of atherosclerosis as it does in younger men (McGill et al. 2000). In women, increased waist/hip ratio (WHR) seems to confer increased risk for CVD irrespective of BMI while in men WHR was a significant risk factor only among the normal weight individuals (Li et al. 2006). Men have in general less favourable CVD risk factor profile than women (Jousilahti et al. 1999). The relative increase in risk for CHD per change in risk factor level seems to be similar in both sexes, but for women the absolute risk is much lower for any given level of a risk factor (Isles et al. 1992). In the Finnish population, the age-related increase of CHD incidence was associated with increases in blood pressure, total cholesterol, relative body weight and diabetes prevalence to a larger extent in women than in men (Jousilahti et al. 1999). Other studies as well have suggested that diabetes, hypertriglyceridemia and low levels of HDL cholesterol (HDL-C) appear to be stronger risk factors for women (Mosca et al. 1997). On the other hand, in age group stratified analysis, the composite risk factor metabolic syndrome seems to be a stronger risk factor for CHD in males. For women over 55 years of age and men of all ages, MetS confers nearly a 2-fold increased risk for CHD but for women less than 55 years of age there is no increase in risk (Tong et al. 2005).

Sex hormones are the most obvious difference between men and women. Oestrogen has several atheroprotective properties presumably contributing largely to the observed difference in CVD risk. Jensen and colleagues (1990) studied women before and after menopause and observed a significant increase in total cholesterol, LDL cholesterol (LDL-C) and triglycerides and a significant decrease in HDL-C independent from biological age. These changes towards a more atherogenic lipid profile are most likely attributable to the decreased levels of oestrogen. As reviewed by Dubey and co-workers (2005), studies on hormone replacement therapy (HRT) in post-menopausal women report controversial results for the use of HRT and the risk for CVD. It is likely that the timing of the therapy in relation to extent of atherosclerosis is crucial: when administered early HRT has several atheroprotective effects while the effects can be detrimental in established atherosclerotic disease. The use of combined oral contraceptive (including both oestrogen and progestogen) is associated with increased risk for CVD but there is less evidence for progestogen only contraceptives (The ESHRE Capri Workshop Group 2006). In addition to the hormonal factors, the relationship between sociological status and modifiable CVD risk factors differ between the sexes (Helmert et al. 1997, Myint et al. 2006).

2.1.3 Inflammatory risk factors

For long, elevated cholesterol levels and the accumulation of lipid in the artery wall have been considered the main culprit of cardiovascular disease. In the 1970's and 80's a new theory emerged proposing that growth factors and vascular smooth muscle cell proliferation are also involved in the development of atheroma (Ross and Glomset 1973). Later it was noted that cells of the immune system are widely present in the atherosclerotic plaque (Jonasson et al. 1986) and that the disease seems to be driven by inflammatory mechanisms. Moreover, total leukocyte count (Yarnell et al. 1991, Danesh et al. 2000) and levels of systemic inflammatory mediators interleukin 6 (Ridker et al. 2000, Tzoulaki et al. 2007), fibrinogen (Danesh et al. 2005), TNF-α (Tuomisto et al. 2006), serum amyloid A (SAA) and C-reactive protein (CRP) (Danesh et al. 2000) have been shown to predict the onset of CVD.

Meta-analyses of the association between total leukocyte count, and fibrinogen, albumin, SAA and CRP concentrations indicated that in long-term prospective studies, for each of the factors, the difference between the highest third and lowest third was associated with a 1.5-2.0 times increased risk for CHD (Danesh et al. 1998, Danesh et al. 2000). Somewhat unexpectedly, of the specific leukocyte types, neutrophils and granulocytes, not lymphocytes and monocytes which have been shown to contribute to the atherosclerosis development, were associated with increased risk for CHD (Wheeler et al. 2004). The association between granulocytes and CHD was apparent in both sexes separately, but was independent from CRP only in men (Rana et al. 2007).

The inflammatory mediators IL6, fibrinogen, TNF-α and CRP are acute phase proteins that respond to injury. Synthesis of these factors takes place mainly in the liver, but recently adipose tissue has emerged as another source of cytokines (Arner 2007). All of these proteins react differentially in response to infection, injury or trauma. For example, the plasma concentration of CRP and serum amyloid A can rise 10000-fold while fibrinogen and the total leukocyte count rises 3-fold (Danesh et al. 1998, Gruys et al. 2005). While these proteins response rapidly and with a wide range of concentration, the long-term circulating concentrations show low intra-individual variation (Danesh et al. 2004). Studies on the inflammatory markers indicate that, a small increase in the levels of acute phase proteins, also known as low-grade inflammation, is associated with elevated risk for CHD. For example, based on the distribution of plasma high sensitivity CRP (hsCRP) in most adult populations, approximate tertiles <1.0, 1.0-3.0 and >3.0 mg/l have been suggested to correspond to low, intermediate and high risk for CHD. Values >10 mg/l are associated with an acute infection (Pearson et al. 2003).

Infection as a risk factor for CVD

Seroepidemiological studies have suggested that infections due to several different microorganisms are associated with CHD. These studies have identified, among others, herpes viruses, *Clamydia pneumoniae* and *Helicobacter pyroli* bacteria, and dental pathogens within the atherosclerotic lesions. Furthermore, studies on animals have shown that infections can initiate atherosclerosis (Alber et al. 2000, Sunnemark et al. 2000). In humans, acute respiratory infection has been associated with acute myocardial infarction. The hypothesis is that the association between CHD and infections is related to immune defence mechanisms, and acute and chronic inflammation (Meier et al. 1998, Leinonen and Saikku 2002). Although studies have identified individual pathogens within the lesions, antibiotic therapy trials have not been successful in preventing CHD or reversing the development of the disease (Leinonen and Saikku 2002). Another possible mechanism by which pathogens may contribute to the disease is the number of different pathogens (the 'pathogen burden' -hypothesis). An increasing number of pathogens have been associated with an increased risk for MI or death among patients with prevalent coronary artery disease both in cross-sectional and prospective studies (Zhu et al. 2000, Zhu et al. 2001). Spahr and colleagues (2006) have made a similar observation on the periodontal pathogen burden.

Inflammatory mediators: markers or makers of CVD?

Inflammatory markers may be either risk factors (directly or indirectly causal to the disease) or markers of the risk for CVD. In atherosclerotic disease, they might measure other characteristics than the atherosclerotic mass. These include the activity of lymphocyte and macrophage populations within plaque or the degree of plaque destabilization and ongoing ulceration or thrombosis. On the other hand, they may measure the activity of the systemic response to risk factors, such as, diabetes, hyperlipidemia, cigarette smoking or obesity (Pearson et al. 2003). There are alternative ways in which inflammatory mediators may contribute to the development of atherosclerosis. These are: 1) cytokines released from the atherosclerotic lesion promote synthesis of CRP and other mediators in the liver, 2) metabolic changes due to other reasons than atherosclerosis causes increased synthesis of these factors both in the liver and in the plaque and there is no direct causal relationship between the factors and the atherosclerotic disease, 3) metabolic changes induces synthesis of these factors in the liver and they subsequently promote atherosclerosis (Nilsson 2005). In the case of CRP there is some evidence of causality: *in vitro* studies have demonstrated that CRP mediates endothelial apoptosis and secretion of matrix metalloproteinase-9 by human mononuclear cells (Nabata et al. 2007), and that it regulates VCAM1 expression in bovine vascular endothelial cells (Kawanami et al. 2006). On the other hand, a study on the

overexpression of CRP in transgenic mice suggests that CRP does not contribute to the disease and that it might simply be a marker after all (Reifenberg et al. 2005).

Similarly, there is evidence that IL6 is expressed in the human atherosclerotic plaque (Schieffer et al. 2000), and that IL6 mediates atherosclerotic progression by stimulating macrophages to secrete monocyte chemotactic protein 1 (Biswas et al. 1998), and by promoting endothelial cells to express adhesion molecules and to secrete cytokines *in vitro* (Romano et al. 1997). In a study by Schieffer and colleagues (2004), ApoE and IL6 deficient double-knockdown mice showed evidence of increased lesion formation but decreased recruitment of macrophages and lymphocytes. The level of anti-inflammatory cytokine interleukin 10 (IL10) was reduced in these animals, and they had hypercholesterolemic lipid profiles. Supraphysiological concentrations of IL6 have been shown to enhance fatty lesion formation in ApoE deficient mice, but not in atherosclerosis resistant mice (Huber et al. 1999). Together these studies suggest that IL6 contributes directly in many aspects of atherosclerosis progression in atherosclerosis-prone circumstances and that the balance between IL6 and IL10 is crucial for lipid homeostasis and plaque formation.

Mouse models of atherosclerosis - contribution of inflammation

Genetically modified mouse models have increased our knowledge of the contribution of inflammatory mediators in the development of atherosclerotic plaque. For example, the role of T-lymphocytes is exemplified by a study, which showed that severe combined immunodeficient mice on an *Apoe*$^{-/-}$ background had smaller lesions than mice injected with CD4^{+} T-lymphocytes (Zhou et al. 2000). Lymphotoxin alpha (*LTA*), which has been associated with MI in a Japanese study, has been shown to be expressed in early lesions and *LTA*$^{-/-}$ knockout animals showed smaller lesions (Schreyer et al. 2002). Human CRP transgenic mice that were LDL receptor (*LDLR*) null and expressed exclusively ApoB-100 (*Ldlr*$^{-/-}$ *Apob*$^{100/100}$), on the other hand, showed decreased lesion development compared to their control littermates. This finding supports the atheroprotective role of CRP adding another piece to the puzzle of the role of CRP in this disease (Kovacs et al. 2007).

2.2 Genetic epidemiology of complex diseases

2.2.1 DNA, genes and the human genome

Every human nucleated cell contains two genomes: a mitochondrial genome of roughly 16570 base pairs and a nuclear genome of approximately 3 billion base pairs of DNA. The nuclear genome is organized into 22 autosomal chromosomes and sex-determining X and Y -chromosomes. Anderson and colleagues (1981) published the sequence of the mitochondrial genome in the year 1981. Massive academic and private efforts finished the first draft of the human nuclear euchromatic genome in 2001 (Lander et al. 2001, Venter et al. 2001) and three years later the International Human Genome Sequencing Consortium (2004) finished filling the remaining gaps covering approximately 99% of the euchromatic genome.

In the simplest view, genes are the functional units of the genome that encode proteins. Currently the estimate of the number of protein-coding genes in the human genome is ~21800 (www.ensembl.org, database build 47.36i) but the number varies largely between individuals According to Nozawa and co-workers (2007) approximately 14% of the human genes are polymorphic with respect to copy number. The organization of the human genome is, however, much more complex as described by Gerstein and co-workers (2007). In addition to protein-coding genes, the genomic sequence contains information for other functional elements such as non-coding ribonucleic acid (RNA) molecules. Moreover, from the same sequence of a gene there can be alternatively spliced messenger RNAs (mRNA), the reading frame of mRNA can be changed to produce an alternative product, and the RNA molecule can be edited post-transcriptionally. Furthermore, the traditional view of alternating exons and introns of a gene has been revised. The same genomic sequence can code for a different product from both chains, there can be a coding sequence for another gene within an intron of a gene, and regulatory elements of both can be intertwined. All these phenomena regulate how the information in the genome is translated into the function of the cell and acknowledgment of these facts makes the definition of a gene complicated. In addition, interpreting and predicting the function and consequences of variants in the noncoding region is difficult and less well understood than the effect coding variants have on protein function (Glazier et al. 2002). As a result, interpretation of findings from complex disease mapping can be very challenging.

Variation in the genome

It has been estimated that every nucleotide in human genome differs between some individuals in the world if the change is compatible with life. In a recent study Levy

and co-workers (2007) estimated that any given individual differs genetically by 0.1-0.5% from other humans. This fraction of sequence dissimilarity contains several types of variation and polymorphisms, polymorphisms being variations that are not sporadic rare mutations. Different kinds of variations and polymorphisms constitute the basis of human inherited diseases. The majority of this variation is single nucleotide polymorphisms (SNP); the current catalogue includes approximately 6.2 million validated SNPs (www.ncbi.nlm.nih.gov/SNP, Genome build 36.2, dbSNP build 128). Since our knowledge of the human sequence is currently based on only a few sequenced individuals, and hence only from a small amount of diversity, this number is most likely an underestimate. According to the neutral theory of mutation (mutation is selectively neutral), a SNP with >1% frequency occurs in every ~300bp of the 3.2 billion base pair genome adding up to roughly 11 million SNPs in total (Kruglyak and Nickerson 2001). SNPs occur everywhere in the genome, but are more abundant in intragenic non-coding regions and in regions immediately around coding regions than in exons and introns of the genes where they are more likely to have a deleterious effect on the gene product (Salisbury et al. 2003).

Other types of variation include repetitive elements of various sizes: insertions, deletions, inversions, duplications and translocations of DNA sequences. Recently several investigators have identified submicroscopic copy-number differences (also known as copy-number variants, CNVs) of various types ranging from kilobases to megabases in size. These CNV regions covered 12% of the human genome in a study of 270 individuals from four populations and included genes, disease loci, functional elements and segmental duplications. For example, one of the candidate genes of this study (II), selenoprotein S is located in a CNV that is approximately 5% frequent in Caucasians. CNVs have been shown to affect phenotypic variation by influencing gene expression and gene dosage (Redon et al. 2006) and analyzing them parallel to other variants will most probably give more insight into the genetic background of complex human traits.

Patterns of linkage disequilibrium

Humans inherit one copy of the nuclear genome from both parents. During meiosis homologous chromosomes from both gametes align and are independently assorted to the offspring. Two loci in the same chromosome are non-independently inherited unless a recombination event between the loci breaks the link. The result, non-random association of alleles within a chromosome is called linkage disequilibrium (LD) and it arises because the variants share a joint population ancestry (Figure 3). There are several measures for quantifying LD. Of these Lewontin's D, absolute value of D' and r^2 are the most commonly used. The correlation coefficient r^2 depends on the allele frequency and intuitively tells us how much information one locus provides from the other. (Ardlie et al. 2002)

The extent and patterns of LD across the human genome have been a focus of wide interest. In addition to recombination and mutation, different evolutionary events modify the extent and distribution of LD. Genetic drift (the random distribution of alleles to the finite number of offspring), and bottlenecks (the random sampling of gametes during restriction in population size) are manifestations of chance acting on the constitution of the human gene pool. Rapid expansion of population size after a bottleneck reduces genetic drift and decreases LD while admixture and migration can increase LD momentarily due to allele frequency differences in the populations. This spurious LD between unlinked markers decreases rapidly in subsequent generations. Natural selection can modify LD when a larger flanking region of a favourable allele also increases in frequency or when a deleterious allele disappears altogether reducing surrounding variability thus increasing LD. (Ardlie et al. 2002)

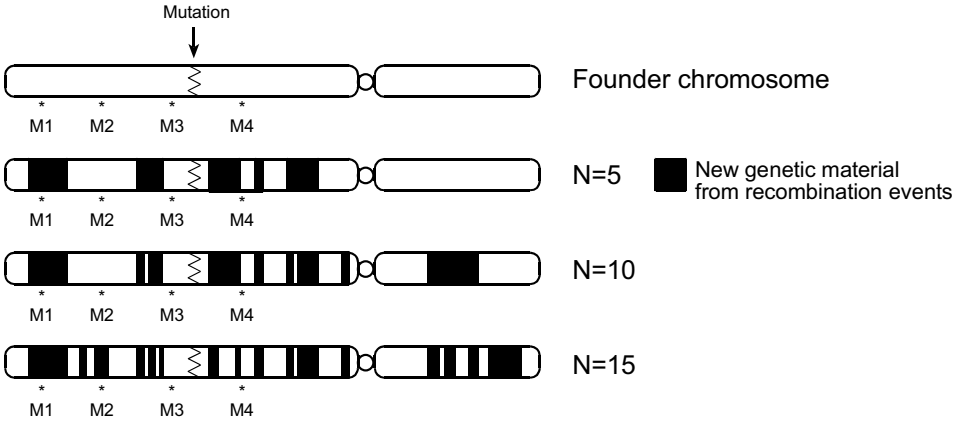

Figure 3. ***The decay of linkage disequilibrium (LD).*** *The shared chromosomal segment around the mutation in the founder chromosome shortens due to recombination events in subsequent generations (N=number of generations). After 15 generations, only a short segment is preserved and only marker (M) 3 remains in LD with the mutation.*

Several studies have investigated the properties of LD in smaller chromosomal regions and in the genome as a whole. Although the studies have varied immensely in regard to size, population, marker type and density as well as measures of LD used, it has become clear that both the quantity and structure of LD varies between populations and between genomic regions (Ardlie et al. 2002). The average length of useful LD in European populations ranges from 6 kb in low LD areas to more than 20 kb in high LD areas and is markedly less in African populations (Reich et al.

2001, Shifman et al. 2003). Studies focusing on long distances have generally reported excess LD whilst studies in shorter segments show less than expected LD across the distance (Pritchard and Przeworski 2001). Demographic features such as a constant size of small population or small sub-isolates seem to lead to extended LD (Kaessmann et al. 2002, Varilo et al. 2003). As a whole, isolated populations exhibit higher overall levels of LD than outbred populations (Service et al. 2006).

Analysis on the LD structure has led to the understanding that in the human genome regions of high LD are punctuated with regions of increased recombination (Daly et al. 2001, Jeffreys et al. 2001). Though the estimates of LD in some of these studies may be biased upwards due to limited number of chromosomes (Terwilliger et al. 2002) and thus leads to an optimistic view of the block structure, regions of high LD are present in the genome in all populations. The block structure, however, differs markedly between populations and even within geographic regions (Gu et al. 2007). Haplotype diversity between populations is great and generally, in a given locus, only a few common haplotypes are frequent in all populations worldwide (Kidd et al. 2004).

2.2.2 Genome-wide statistical methods for identifying genes underlying complex diseases

Genetic markers

In principle, the search for susceptibility genes or loci for common diseases can utilize all types of genetic variation as genetic markers. The feasibility and affordability of the genotyping together with the information content of the markers define the design of a large-scale genetic study. Short tandem repeats, also know as microsatellites, have been the first to be put to use in large-scale genotyping efforts. Polymerase chain reaction (PCR), that can amplify many markers simultaneously, facilitates reduction in the cost and use of DNA material. Microsatellites have multiple alleles, higher mutation frequency than SNPs and thus have high information content (Hearne et al. 1992, Kruglyak 1997). Marker information content affects detectable LD and thus restricts the choice of analysis method and affects the power of the study (Varilo et al. 2003).

SNP markers are most often bi-allelic and occur more frequently in the genome than microsatellites. Recent technological advances have made it possible to analyse thousands of SNPs at a relatively affordable price and with modest consumption of DNA sample (Syvänen 2005). SNP markers have lower information content than microsatellites but are more stable and located more often near and within genes. Although many more SNPs than microsatellites are required to reach a similar

genetic density and information content, with the current genotyping technologies this has become feasible and allows for multiple uses of SNPs in genetic analyses (Kruglyak 1997, Steemers and Gunderson 2007).

Family-based linkage analysis

Genetic mapping aims to identify the chromosomal regions where the disease-associated genes are located. This can be achieved with positional cloning using multi-generational families segregating the disease and genome-wide marker sets. Linkage analysis looks for chromosomal regions that are linked with the disease i.e. segregate together with it in the family. Statistical methods, which calculate the probabilities of the observed recombination between two adjacent markers given the genetic distance, use information of the family or pedigree structure to aid the analysis. Parametric methods rely on *a priori* information of the recombination fraction along the genetic map, mode of inheritance, penetrance, phenocopy rate, marker allele frequencies and disease allele frequency (Terwilliger and Göring 2000). In the case of rare monogenic diseases these parameters have been relatively easy to define correctly and they have aided mapping of several disease genes (Collins 1995, Peltonen et al. 1999). In complex disease mapping, these parameters are difficult, if not impossible, to define reliably beforehand. Non-parametric methods, such as the affected sib pair analysis, investigate in affected individuals the proportion of allele sharing, which is expected to be greater than random Mendelian segregation in a disease-linked locus (Terwilliger and Göring 2000).

Association analysis

Genome-wide genetic association studies (GWA) for complex diseases are likely to be more powerful than linkage studies in identifying several predisposing genes with small effects (Risch and Merikangas 1996). The analysis method can be used for several types of study samples and any number of markers. The most common study sample is population based case-control analysis in affected and unaffected, so called control, individuals. Other types of study designs include a population based case-cohort analysis and family based analysis, where it is possible to analyze both linkage and association separately and together (Terwilliger and Ott 1992). Association study compares whether the allele frequency of the genetic marker is significantly different in the affected and unaffected groups.

Any type of genetic markers can be used for association analysis. Direct association mapping aims to select the disease-causing variant, e.g. a known coding region variant or a variant from a known regulatory region, for the analysis, whereas indirect study design relies on significant LD between the typed marker and the putative functional variant (Figure 4). However, since indirect association analysis can be thought to observe the haplotype background of a variant, which has been

introduced several generations ago into the gene pool in individuals who are currently identical-by-decent for the variant, are SNPs with lower mutation rate better for this analysis than microsatellites (Kruglyak 1997, Terwilliger and Göring 2000).

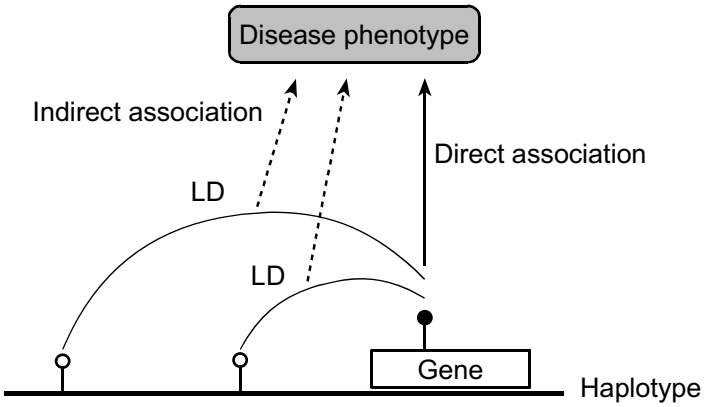

Figure 4. ***Indirect and direct association testing.*** *A test between the causal variant (full circle) and the disease phenotype is a test of direct association whereas a test between a surrogate variant (open circle) and the phenotype is indirect. Indirect association is possible because of linkage disequilibrium (LD) between the variants.*

As described by Manolio and colleagues (2006), association study designs are either retrospective or prospective. The most common has been retrospective design, in which the researchers collect information about the predisposing environmental risk factors after the affected and unaffected study subjects have been identified. Although this design is economically the most feasible, there is a possibility for bias between cases and controls in assessment of past exposure to risk factors. Recall bias may occur if cases and their family members remember, influenced by their current health status, potential harmful exposure more often than the controls. In prospective design, the study collects information about the risk factors at the baseline and identifies, through a follow-up, the individuals who are affected. Prospective cohort study also provides us with unbiased risk estimates, whereas in case-control studies the odds ratio approximates the relative risk only if the probability of getting the disease is small (Zondervan and Cardon 2004).

Case-control studies identify cases often through clinics and have the advantage of collecting cases of rare diseases and mild conditions. For a cohort study to identify as many cases, it would have to be very large and utilize a very powerful follow-up method, something that is often impossible to achieve. The advantage of a case-cohort study is that it can identify several types of diseases and use the same control cohort (a subcohort sampled from the cohort) and exposure information for all end-points thus reducing the cost of genotyping and phenotyping (Prentice 1986, Zondervan and Cardon 2004, Manolio et al. 2006). The principles of both designs are described in Figure 5 and discussed further in 2.2.5.

Allele spectrum of complex diseases

The nature of allelic variants underlying complex diseases has been much debated over the recent years. Multifactorial diseases, by definition, are influenced by incompletely penetrant variants. Furthermore, it is unlikely that a variant would be both common and confer a high risk since it would have to be the principal contributor to the overall prevalence. In addition, it can be assumed that common deleterious variants with large effects would not be the cause of common diseases since they would have been selected against in the course of evolution (Zondervan and Cardon 2004). Linkage studies have generally failed to detect variants with large effects conferring risk for complex diseases. Many researchers have thus suggested that one, or small number of, common variants with small effects in given genes would be the cause of common diseases (Lander 1996, Risch and Merikangas 1996, Collins et al. 1997). In the case of late onset diseases, such as cardiovascular disease, the disease is not likely to affect the reproductive fitness of the individual, and is thought to have developed without selective pressure on the genetic variants. As a result, the variants underlying CVD may have evolved neutrally and exhibit a wide range of frequencies, allelic heterogeneity and epistasis (Weiss and Terwilliger 2000). Alternatively, variants that are now disease-predisposing and common might have been advantageous and not under negative selection in the past (i.e. the 'thrifty-gene' hypothesis) (Neel 1962).

Although there is empirical (Lohmueller et al. 2003) and theoretical (Reich and Lander 2001, Peng and Kimmel 2007) evidence that the common disease, common variant (CDCV) hypothesis holds true, opposing arguments have also been presented. It has been suggested that, since environmental factors are so important in the development of complex disease, individual allelic variants must confer low attributable risk and be rare (Weiss and Terwilliger 2000). Particularly in late-onset diseases, a theoretical modelling proposes that when one assumes neutral selection in constant size population, common variants would contribute little to the genetic variance since they would be close to fixation (Pritchard 2001). Furthermore, empirical evidence from late-onset Mendelian disorders suggests that causal genetic

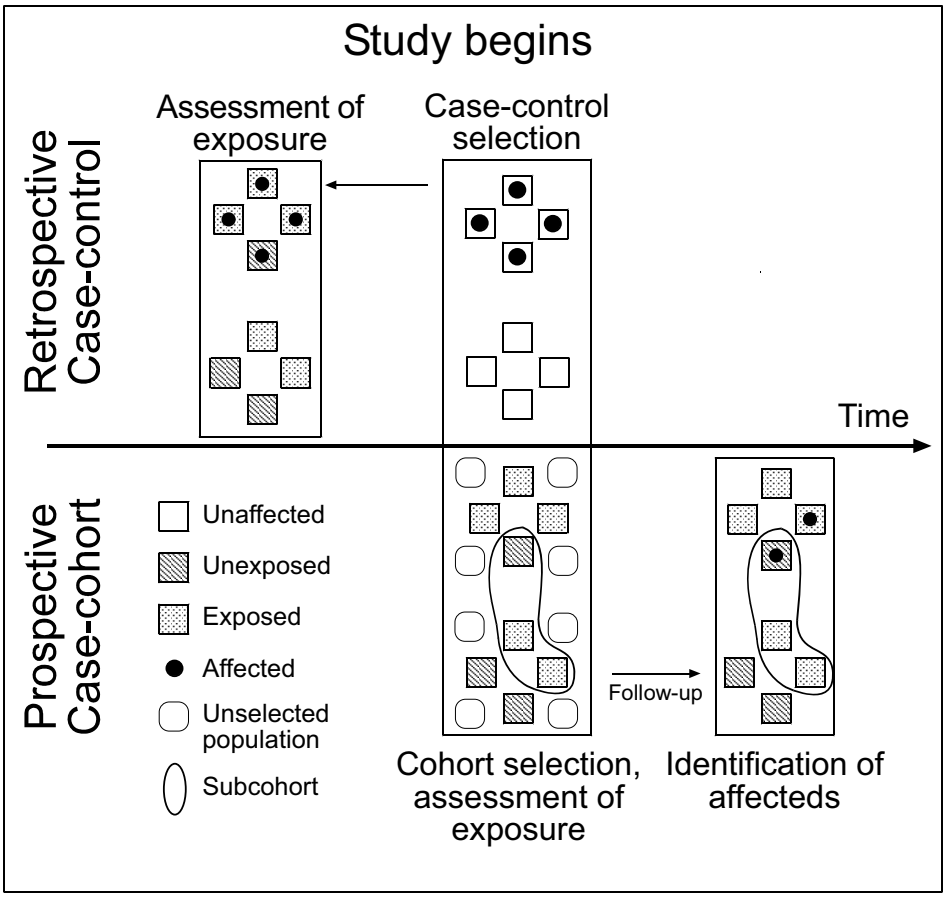

Figure 5. ***Case-control and case-cohort study designs.*** *At the beginning of the study, case-control studies identify affecteds and unaffecteds and then proceed to assess their past exposure to risk factors. In a case-cohort study, the study begins by selecting the representative cohort from the base population, and assessing their exposure. During the follow-up individuals who become affected are identified. A random subcohort drawn from the original cohort represents the controls in the survival analysis.*

variants that have failed to influence fitness have a broad allelic diversity (Wright and Hastie 2001). Currently, we do not have enough empirical evidence to know which model describes the allele spectrum of complex disease most precisely and most likely the spectrum includes both common and rare variants with allelic and locus heterogeneity.

The haplotype map of the human genome

The hypothesis, that each disease-causing locus harbours only one or a few causative variants and that these variants would be common, forms the foundation of a project HapMap that aims to construct a haplotype map of the human genome. The HapMap is a publicly available resource of haplotype information across the genome based on common SNPs. The aim is to enable researchers to study the local structure of LD and select variants from haplotype blocks to represent haplotype variation within their region of interest, or the genome as a whole, and thus reduce the number of variants in the study (The International HapMap Consortium 2003).

Methods for selecting the representing variants are available and have been demonstrated to describe the underlying haplotype diversity. Alternative methods encompass selecting haplotype tagging SNPs (htSNP) to detect common haplotypes, selecting htSNPs that define haplotype blocks, and selecting tagSNPs that describe all common variation within the gene with a specified correlation threshold using r^2 as the metric (Johnson et al. 2001, Stram et al. 2003, Zhang and Jin 2003, Carlson et al. 2004). TagSNPs can generally be fairly well transferred between populations, including population isolates, such as Finland (Willer et al. 2006). However, when Gu and colleagues (2007) selected htSNPs from later populations along the out-of-Africa path: Africa, southwest Asia, Europe and the Americas or east Asia, they did not capture the haplotype diversity in the earlier populations as precisely as did a selection made *vice versa*. This observation agrees with the evolutional hypothesis that older mutations will be more common and have a shorter shared haplotype around them, thus having less correlation with surrounding variants (Kidd et al. 2004).

According to the CDCV and HapMap hypotheses, it is possible to use indirect association mapping to localize disease-predisposing genes for complex diseases (Pritchard and Cox 2002). However, some researchers point out that some of the underlying assumptions of the HapMap project may be questionable. Allelic heterogeneity reduces the power of the study, as do overestimates of the amount of LD and similarity of the structure of LD across populations (Weiss and Terwilliger 2000, Terwilliger et al. 2002, Weiss and Clark 2002). In addition, there is doubt whether it can be statistically possible to detect indirect association even in the case of perfect LD (Terwilliger and Hiekkalinna 2006). Furthermore, the extent of LD around common variants is usually less than around rare variants, since evolution has had more time to increase haplotype diversity (Jorde 1995, Kruglyak 1999). Studies in isolated populations and especially in sub-populations with recent demographic events that have reduced allelic heterogeneity and increased the extent of LD may be better equipped to circumvent some of these problems (Kaessmann et al. 2002, Varilo and Peltonen 2004). The greater extent and homogeneity in LD can

be also achieved by adopting analysis in disease-ascertained pedigrees in isolates (Varilo et al. 2000).

Despite all the doubts listed above, many recent studies have demonstrated that the tagSNP approach can be used for genome-wide association studies. The ~1.3 million SNPs included in the HapMap Phase I have been successfully used in replicating previous associations and identifying new candidate loci and genes for common diseases: type I diabetes (Hakonarson et al. 2007), type II diabetes (Sladek et al. 2007, Steinthorsdottir et al. 2007), and myocardial infarction (Helgadottir et al. 2007) to name a few. The HapMap Phase II launched in 2007 added a further 2.1 million SNPs and has the average spacing of one SNP per kilobase. It now includes more rare variants than the previous release, and with the increased density, allows for more detailed definition of recombination hot spots and haplotype blocks (The International HapMap Consortium 2007). The Wellcome Trust Case Control Consortium GWA studies have identified several new candidate genes and loci for common traits and diseases using an Affymetrix 500K genotyping platform that covers ~70% of the Phase II SNPs (The Wellcome Trust Case Control Consortium 2007). It seems now plausible that GWA studies can identify common variants conferring risk for common traits and diseases, and that the tagSNP approach can be utilized in this type of study design. However, the effect size of identified variants is most often very small and thus their clinical usability remains yet to be determined. Most of the findings are, nevertheless, pathobiologically very interesting.

2.2.3 Candidate gene -approach

While genome-wide studies aim to locate the genomic region predisposing to the disease, candidate gene approach starts with *a priori* hypothesis of the role of the gene in the aetiology of the disease. The gene can be a positional candidate from previous genome-wide studies, or it can be selected based on studies in model organisms, or biological hypothesis and known functions of the gene. In contrast with genome-wide studies, candidate gene studies often select variants with predicted or known functional role affecting protein structure or folding, splicing of the gene, regulatory elements, or gene expression as candidates for direct association testing. The same LD based selection algorithms for indirect association studies are applicable to these studies as for the genome-wide approach (Tabor et al. 2002). Moreover, since the region of interest is generally of manageable size, in-depth re-sequencing can be used to discover a full pattern of variation within the gene, and densely spaced variants can be included in the study. This increases confidence that a full spectrum of variation has been analysed and the null hypothesis of no genetic association can be accepted or rejected with greater certainty (Cordell and Clayton 2005).

2.2.4 Evaluation of the statistical significance of association studies

The literature of genetic association studies, including CVD, is full with initial positive findings followed by lack of replication or inconsistencies in the findings (Ioannidis et al. 2001, Ntzani et al. 2007). There are many reasons for this phenomenon (Figure 6). First, the lack of replication may be entirely correct and the initial finding is a false-positive association. The amount of false-positive results in the literature relates mostly to the power of the studies. Power is defined by the probability of not finding an association when a true one exists (i.e. 1-type II error rate). The prior probability of true association, the size of the study, the effect size of the variant, as well as, the required level of statistical significance (i.e. type I error rate, α) has an influence on the power (Colhoun et al. 2003, Livingston and Cassidy 2005). In addition, problems in the power to detect disease association arise when the study uses indirect association and the LD is incomplete, and most importantly, the marker allele frequency differs from the disease allele frequency (Müller-Myhsok and Abel 1997, Zondervan and Cardon 2004). The inability to exclude chance findings at the desired level of statistical significance relates both to the power of the study and to the number of analysed variants, i.e. multiple testing problem. Most studies have investigated too few individuals to face the multiple testing problem especially in the context of small effect sizes of variants predisposing to complex diseases. Estimates for the necessary number of individuals have been presented for various study designs and numbers of markers (Risch and Merikangas 1996, Risch 2000). In principle, the number of individuals increases *n*-fold if the required level of statistical significance reduces *n* powers of 10 (Colhoun et al. 2003).

Several methods to correct statistically for the multiple comparisons are available. Bonferroni correction straightforwardly compensates for *n* independent tests by setting the significance threshold at α/n. It is generally agreed that the Bonferroni correction is too conservative for studies including markers in LD and correlated phenotypes (Balding 2006). A supplementary method called SNPSpD can be used to estimate the effective number of markers and adjust the significance threshold accordingly. The method accounts for the non-independence by using spectral decomposition (Nyholt 2004). An alternative approach is false discovery rate (FDR), which monitors for the proportion of false positive results among all positives (Benjamini and Hochberg 1995). A practical option is to perform permutation testing to obtain a false-positive rate by generating several data sets by breaking the link between genotype data and phenotype data (Churchill and Doerge 1994).

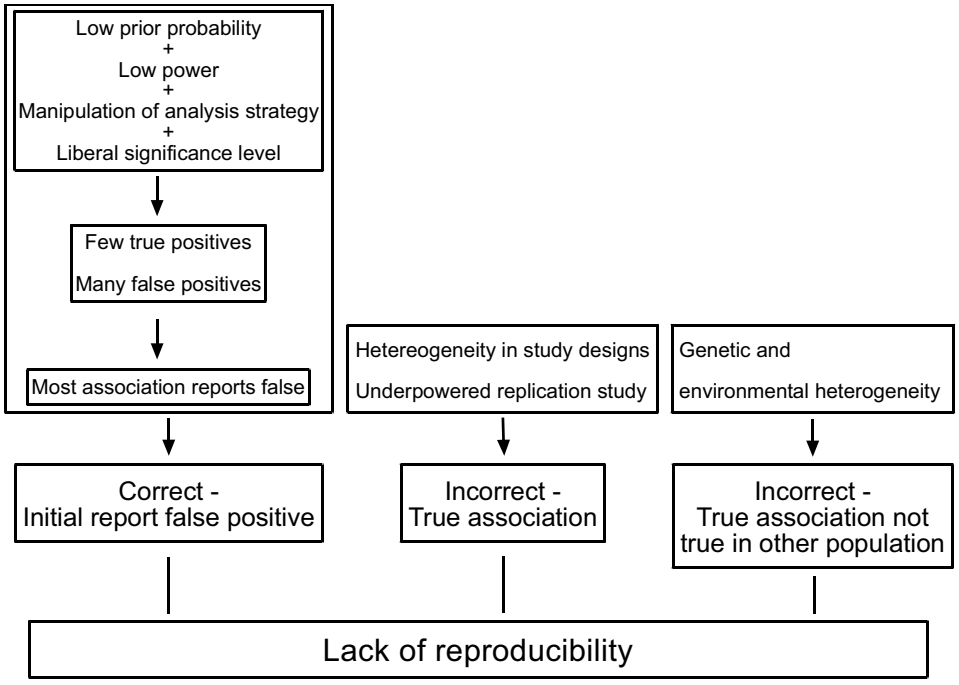

Figure 6. ***The lack of reproducibility in genetic association studies.***

False-positive findings may also arise due to manipulation of analysis strategies. As an example, studies occasionally extend the original analysis plan during the analysis process, and in the end, report findings only from subgroup analysis, such as, sex, age and specific subcategory of phenotypes. Sometimes these exhibit the only signs of association and should be viewed with caution (Colhoun et al. 2003). However, with strong biological hypothesis – e.g. different aetiology of a disease in men and women – subgroup analyses are an important part of the research but should be incorporated in the analysis strategy from the very beginning (Pilote et al. 2007).

Inadequate power leads to the second main form of lack of replication in which a correct report of a true association fails to be replicated in an underpowered follow-up study. In addition, heterogeneity in study designs is one source of inconsistency in association reports. Differences in population stratification, selection and misclassification of cases and analysis of effect modification may lead to substantial differences between studies (Manly 2005). Heterogeneity in genetic or

environmental background may lead to the third major form of unreproducibility where a true association in one population may not necessarily be true in a second.

The power of a study can be increased by increasing the sample size. Since increasing the number of individuals in the study is expensive and sometimes unfeasible, a multi-stage approach has been proposed especially for large-scale genotyping efforts. In this design, interesting findings with less stringent levels of α in the initial step are replicated in following independent study samples while simultaneously increasing the level of statistical significance (van den Oord and Sullivan 2003, Hirschhorn and Daly 2005). Alternatively the power to detect an association can be increased by enriching the genetic component in the study sample. This leads to increased prior probability of observing a genetic effect (Antoniou and Easton 2003, Terwilliger and Weiss 2003).

2.2.5 Study populations, subjects and biological samples

Genetic isolates and perspective from Finland

The use of isolated founder populations, such as, Finland, Iceland, Sardinia, the Bedouines, the Canadian Hutterites, the Old Order Amish, the North American Mennonites, and the Ashkenazi Jews has aided identification of monogenic forms of diseases. In these populations, a small number of founder chromosomes have introduced the disease allele into the population and the extended LD around the rare mutation is still present today. Relative isolation because of cultural or geographical or both factors has limited gene flow from other populations. In addition, bottlenecks, genetic drift, rapid expansion and formation of internal isolates have modified the gene pool leading to reduced genetic diversity and enrichment of some alleles (Peltonen et al. 1999, Peltonen et al. 2000). The success for mapping complex diseases has been more limited, although some examples exist. Linkage analysis followed by association in samples from North-American Mennonites has identified the endothelin B receptor as a candidate for Hirschsprung disease (Puffenberger et al. 1994a, Puffenberger et al. 1994b). Similarly, linkage analysis in one Bedouin extended pedigree identified the chromosomal region of connexin 26 in relation to hereditary nonsyndromic sensorineural deafness (Guilford et al. 1994, Kelsell et al. 1997).

Linkage disequilibrium mapping in isolates has been less successful for complex diseases, although a few recent examples exist of new putative loci for type II diabetes, Parkinson's disease, atrial fibrillation and exfoliation glaucoma from Finland, Iceland and the Netherlands (Bertoli-Avella et al. 2006, Gudbjartsson et al. 2007, Helgadottir et al. 2007, Scott et al. 2007, Thorleifsson et al. 2007). Assuming

CDCV to be the underlying model, the outcome of these studies depends largely on the extent of LD around common variants. Even if the number of founders is relatively small, common alleles have entered the gene pool several times after the origin of the isolate and the shared haplotype around the alleles has become indistinguishable from the ancestral haplotypes (Kruglyak 1999). Nevertheless, isolated populations may aid in identification of complex disease genes in other ways. First, fewer predisposing alleles are likely to contribute to the diseases by virtue of reduced genetic and allelic heterogeneity (Collins 1995). Second, cultural and environmental homogeneity, and standardized and equally available health care reduce phenotypic and etiological variability and diagnostic misclassification, and third, in some of the isolates genealogical records allow for accurate construction of large pedigrees over several generations (Peltonen et al. 2000). Therefore, initial localization of disease-predisposing genes or loci might be somewhat easier in isolated than in outbred populations. Identification of the causative variant or fine mapping could thereafter utilize, for example, African populations with less LD (Reich et al. 2001).

Selection of study subjects in population-based studies

The selection of subjects is crucial for the success and validity of any genetic association study. Genetic epidemiology is as at risk to different kind of bias as is traditional epidemiology. In addition to bias related to collection of exposure information, several type of selection bias can affect the results (Table 4).

In case-control studies, the selected subjects should all represent the same underlying study base. The same study base is inherent to case-cohort studies. If cases are identified as patients of a particular hospital, then the catchment population for controls are the individuals in the same hospital district and include all individuals who are at risk for developing the disease (Miettinen 1985). In case-control studies, the selection of controls is even more important than the selection of cases. In order to reduce false positive findings and to assist detection of true findings, the controls should match the cases by strong confounders (e.g. sex, age, ethnicity) affecting the disease aetiology. However, matching by an intermediate phenotype or a factor in a causal pathway between exposure and disease can lead to overmatching and bias the results (Wacholder et al. 1992). In addition, methods such as selection of hypernormal controls (i.e. old individuals who have not developed the disease) or the use of population controls representative of the population from which the cases originated have been proposed (Hattersley and McCarthy 2005). Sampling from a homogenous population, such as a population isolate, reduces heterogeneity among lifestyle, environmental and genetic factors in both cases and controls and reduces differences in unobserved putative risk factors between these groups (Peltonen et al. 2000).

Table 4. ***Some sources of bias in genetic epidemiology.***

Selection of subjects
Detection signal (unmasking) bias
An exposure may become a suspect if it causes symptoms that are unrelated to the disease and this precipitates a search for the disease.
Diagnosis bias
Information of the current exposure to a know risk factor may influence the thoroughness of the diagnostic procedure or affect the intensity of the outcome.
Drop-out bias
Drop-out bias affects cohort studies where individuals may be lost during the follow-up due to different reasons related or unrelated to the disease in question.
Healthy worker effect
Members of a group, e.g. employed, athletes etc., may have a significantly different degree of health compared to the general population.
Non-response bias
The participation rate of affected and unaffected individuals may be different and this may or may not be related to the exposures and outcomes in the two groups.
Prevalence-incidence or survival bias
The study misses all fatal and silent events and short episodes in cases where evidence of exposure disappears after disease onset.
Referral or admission rate bias
This bias relates to clinical setting, where the inclusion probability might differ between cases and controls depending on the availability of advanced care, burden of symptoms or popularity of disorders.
Surveillance bias
Individuals who are provided with routine medical surveillance and have a mild outcome of the disease are more likely to be included in the study than individuals without routine surveillance.
Information on exposures
Recall bias
Patients and their family members may search their memory more intensely and more accurately for exposures for putative causes of the disease than unaffected individuals.
Family information bias
Information of the exposure or illness in the family may be readily more available to affected individuals or may be directed to these individuals more than to unaffecteds.
Exposure suspicion bias
The putative cause of the disease may be more thoroughly looked for in affected individuals.

From Sackett (1979) and Manolio (2006).

Considerations on the quantity and quality of DNA samples

In order to be successful, a genetic epidemiological study needs to follow several stringent quality control criteria, both in collection and in analysis of biological samples (Altshuler and Daly 2007). Recently the NCI-NHGRI workgroup on replication in association studies (2007) published guidelines for a good initial association study as well as a valid replication study. In their proposition, they outlined that a study should report the exclusion reasons due to DNA sample characteristics, success rate for DNA acquisition and extraction, and comparisons of phenotypic characteristics according to sample availability.

Selection bias (Cardon and Bell 2001) due to sample availability may lead to differences in allele frequencies if the rate of dropout is different among cases and controls or if the underlying cause of dropout is associated with the disease. Apart from differences in willingness to donate a DNA sample and possible difficulties in obtaining the sample, for example, from the elderly or small children, the factors affecting the availability of DNA sample usually relate to the success of the extraction procedure. Several extraction methods are available for different types of source tissues (Vandenplas et al. 1984, Miller et al. 1988, Yoza et al. 2002) and SNP genotyping accuracy is similar in DNA samples extracted from different tissues types (Montgomery et al. 2005). The most commonly used tissue in large-scale genetic epidemiological studies has been whole blood, but non-invasive methods such as buccal swaps and mouthwashes are becoming more common. The amount of available DNA ranges from tens of micrograms per buccal cell sample to hundreds of micrograms from 10 ml of peripheral whole blood sample (Steinberg et al. 2002).

Storage time and temperature of the sample prior to extraction affects the quantity of DNA. Freezing reduces the yield in whole blood samples by approximately 25-50% (Madisen et al. 1987, Feigelson et al. 2001, Nederhand et al. 2003). The studies on the quality of the DNA are more conflicting: in some studies prolonged storage or temperatures other than room temperature did not affect the integrity of DNA measured as visual inspection of degradation on agarose gel (Lahiri and Schnabel 1993, Visvikis et al. 1998), whereas other studies have observed degradation (Farkas et al. 1996) and increased protein contamination (Nederhand et al. 2003). In addition to these technical aspects, the age of the donor correlates negatively with the DNA yield from whole blood samples. This is most likely related to the age-associated decrease in leukocyte count (Erkeller-Yuksel et al. 1992, Richardson et al. 2006).

Whole genome amplification (WGA) methods can resolve the problem of low DNA amounts in some samples (Silander et al. 2005). Methods, such as, degenerate oligonucleotide-primed PCR, primer extension preamplification, and multiply-primed

rolling circle amplification (MPRCA) are available, but they are still quite expensive and laborious to use (Hawkins et al. 2002). In general, MPRCA phi polymerase based methods amplify the genome without great regional or sequence bias and are suitable for samples with very low DNA yield. Pooling two amplifications increases genotyping accuracy even further (Dean et al. 2002). Both Illumina and Affymetrix whole genome genotyping platforms can use WGA samples from various source tissues with acceptable genotyping accuracy but with lower call rates (Tzvetkov et al. 2005, Paynter et al. 2006, Berthier-Schaad et al. 2007).

2.3 Genetics of cardiovascular disease

2.3.1 Genetic component

It has been known for years that cardiovascular disease 'runs in families'. Already in the 60's studies showed that the risk for ischemic heart disease was seven times higher for relatives of CHD cases than controls (Slack and Evans 1966). In North Karelia, a high CVD risk area of Finland, Rissanen (1979a) estimated that the cumulative risk for CHD by the age of 65 was 4.5 higher for brothers and 2.6 higher for sisters of CHD cases when compared with control siblings. In addition, the clinical manifestation of the disease seemed to aggregate in families: fatal MI occurred more often in the family if the proband had died of the disease, and a similar observation was made for angina pectoris. Recently, a German study reported that left main coronary disease is more heritable and confers approximately 2.5-3 times higher risk for unaffected siblings compared to other types of the disease (Fischer et al. 2007). The genetic contribution seems to be even stronger for premature or early onset CHD. In eastern Finland, the risk for having CHD by the age of 55 was 11.4 times higher for brothers of cases who developed the disease by the age of 46 compared to only 1.3 times higher for brothers whose sibling had been diagnosed at the age of 51 to 55 years. Aggregation of hyperlipidemia and hypertension in the early onset families might partially explain the higher risk in these families (Rissanen 1979b). A similar observation of risk factor aggregation has been made in the Framingham Offspring Study. In this study as well, the impact of sibling CVD was stronger in the younger age group (≤ 48 years) than in the older age group (odds ratio 2.22 vs. 1.33) but this difference was not significant (Murabito et al. 2005). Using the same study set Lloyd-Jones and colleagues (2004) reported that premature parental CVD was associated with 2.0 times increased risk for CVD for men and 1.7 times increased risk for women compared to individuals with no parental CVD. Nonpremature parental CVD was not as strong a predictor as premature CVD. Twin studies indicate that heritability of MI ranges from 0.25 to 0.6

and, for example, a Swedish study of 20 966 twins estimates that heritability of coronary death is 0.57 for men and 0.38 for women (Zdravkovic et al. 2002, Lusis 2003). Specifically for stroke, a recent meta-analysis reported that monozygotic twins were more likely to be concordant for stroke than dizygotic twins and positive family history was a significant risk factor in case-control and cohort studies, although severe heterogeneity between studies was observed (Floßmann et al. 2004). Studies using adjustment for traditional risk factors show that also other genetic factors, than those affecting these traits, have a role in the risk for CVD.

2.3.2 Important lessons from monogenic disorders

Studies on rare Mendelian diseases that cause premature coronary artery disease (CAD), and abnormal lipid profiles or blood pressure have revealed a few mutations, which are presumed to explain a fraction of an individuals' inherited risk for common, multifactorial CAD. Although there is still a large gap in our knowledge of the genes that are involved in susceptibility for most patients with multifactorial CAD, these studies have increased our understanding of the pathophysiology of the disease.

Many of the rare Mendelian monogenic diseases share similar pathological components with the multifactorial CAD. For example, familial combined hyperlipidemia (FCHL) causes elevated blood cholesterol and triglyceride levels which make the patient susceptible for premature CAD. In the common form of CAD as well, altered lipid metabolism predisposes to the disease. By studying families with FCHL Pajukanta and colleagues (2004) were able to identify a novel gene, upstream transcription factor 1 (*USF1*), which was associated with abnormal lipid profiles in these families. Variation in the *USF1* gene has later been associated with total and non-HDL-C, waist/hip ratio in both sexes and with CVD and all-cause mortality in females in a population sample (Komulainen et al. 2006). Similarly, studies on Tangier disease have identified a gene ATP-binding cassette transporter A1 (*ABCA1*), which in a truncated form causes low HDL-C levels (Bodzioch et al. 1999). Heterozygous mutations in the *ABCA1* gene are associated with HDL-C levels also in the general population (Frikke-Schmidt et al. 2004).

Apart from direct population associations, studies on monogenic diseases have also taught us about the basic biological mechanisms related to cholesterol metabolism and regulation of blood pressure. Familial hypercholesterolemia, familial ligand-defective apolipoprotein B-100, autosomal recessive hypercholesterolemia, and sitosterolemia are all caused by mutations in the LDL-receptor pathway. Understanding of the mechanisms of LDL excretion and uptake in the liver has facilitated development of the cholesterol lowering drugs statins. Much in the same

way, studies on rare Liddle's, Gitelman's and Bartter's syndromes, which cause abnormal blood pressure, have increased our knowledge of the renal function of epithelial sodium channels, sodium-chloride cotransporter, and genes required for normal salt reabsorption in the thick ascending loop of Henle, and their role in the regulation of blood pressure. (Nabel 2003)

2.3.3 Family-based findings on susceptibility loci for common CVD

To date, (January 2008) there have been 9 published genome-wide linkage mapping studies on CVD. Most of them have been carried out in families with premature CVD. These studies have identified several candidate loci for premature MI, stroke or CHD (Table 5). In addition to linkage scans for CVD, several studies have reported results on linkage scans on CVD risk factors such as hypertension, blood lipid or inflammatory marker levels and type II diabetes. Regions in chromosomes 3q13, 2q and 17p are among the few putative loci where the linkage results for CVD have been replicated if not at a genome-wide significance, but at a suggestive level. However, some of the implied loci harbour interesting candidate genes that have subsequently been shown to associate with the disease. For example, in chromosome 13q12 locus, the *ALOX5AP* gene, which encodes 5-lipoxygenase activation protein (FLAP), was first identified as a candidate for premature MI and ischemic stroke in Icelandic families (Helgadottir et al. 2004). FLAP protein operates in a pathway involved in biosynthesis of leukotriene A4 (LTA4). This pathway has independently been implicated in the pathogenesis of atherosclerosis in humans (Spanbroek et al. 2003) and in mouse models (Mehrabian et al. 2002). Furthermore, a variant in the 5-lipoxygenase (*ALOX5*) gene has been associated with increased carotid artery intima media thickness (IMT) (Dwyer et al. 2004), and a haplotype in the gene encoding LTA4 hydrolase has been associated with modestly increased risk for MI in European Americans (risk ratio = 1.16) and with even greater risk in individuals with African origin (risk ratio = 3.57) (Helgadottir et al. 2006) lending more support to the role of this pathway in the aetiology of CVD. The *ALOX5AP* locus expresses some allelic heterogeneity: in Iceland and in Britain different haplotypes have been associated with MI while the same haplotype was associated with stroke both in Iceland and in Scotland and with suggestive evidence of association in Sweden (Helgadottir et al. 2004, Helgadottir et al. 2005, Kostulas et al. 2007). Although positive association dominate the reports on *ALO5AP*, in a German case-control sample haplotypes in this gene did not associate with MI (Koch et al. 2007).

Table 5. ***Genome-wide linkage studies on coronary artery disease, acute coronary syndrome, and ischemic stroke.*** *See numbering for the references within the table in the first column and in the footnote.*

Reference	Population	Trait	Genome-wide significant locus[a]	Linkage replication for a relevant trait (ref)	Associated gene	Association replication for a relevant trait (ref)
Pajukanta et al. 2000 (1)	Finnish	Early onset CAD	2q21.1-22 Xq23-26			
Francke et al. 2001 (2)	Indo-Mauritian	Early onset MI	16p13-pter[b] 3q27[b] 10q23[b]			
Broeckel et al. 2002 (3)	German	MI	14q32			
Harrap et al. 2002 (4)	Australian	ACS	2q36-q37.3 3q26-q27[b] 20p11-p13[b]			
Gretarsdottir et al. 2002 (5)	Islandic	Stroke	5q12		PDE4D (5)	For review see (14)
Wang et al. 2003 (6)	European-American	MI	15q26		MEF2A	CHD (15)
Hauser et al. 2004 (7)	American	Early onset CAD	3q13	Self-reported CVD/cariotid calcification (16)	KALRN (18), GATA2 (19)	Early onset CAD, KALRN & GATA2
			1q25[b] 7p14[b] 19p13[b]	FCHL (17)	USF1 (20)	CVD (21)
Helgadottir et al. 2004 (8)	Icelandic	MI/ischemic stroke	13q12–13[b]		ALOX5AP	CAD (different haplotype), stroke (22)
Wang et al. 2004 (9)	American, mixed races	Early onset MI	1p34-36		CX37	IMT, CAD see review (23)
			4q32[b]	Self-reported CVD/cariotid calcification (16)		
Samani et al. 2005 (10)	United Kingdom	Early onset CAD/MI	2p12-2q23.3[b]	Early onset MI[b] (9), Early onset CAD (1) (close)		

Reference	Population	Trait	Genome-wide significant locus[a]	Linkage replication for a relevant trait (ref)	Associated gene	Association replication for a relevant trait (ref)
Farrall et al. 2006 (11)	European	MI < 66y	17p11.2-q21	Within this study		
		CAD < 66e	2[b]	Suggestive, distally located, replication (1, 9)		
Nilsson-Ardnor et al. 2007 (12)	Swedish	Stroke	5q13[b] 13q32[b] 18p11[b]	Ischemic stroke (8)		
Engert et al. 2007 (13)	French-Canadian	Early onset CAD	8p22			

[a] As indicated by the authors

[b] Suggestive linkage or subset of sample

References: 1) Pajukanta et al. 2000, 2) Francke et al. 2001, 3) Broeckel et al. 2002, 4) Harrap et al. 2002, 5) Gretarsdottir et al. 2002, 6) Wang et al. 2003, 7) Hauser et al. 2004, 8) Helgadottir et al. 2004, 9) Wang et al. 2004, 10) Samani et al. 2005, 11) Farrall et al. 2006, 12) Nilsson-Ardnor et al. 2007, 13) Engert et al. 2007, 14) Rosand et al. 2006, 15) Larson et al 2007, 16) Bowden et al. 2006, 17) Pajukanta et al. 1998, 18) Wang et al. 2007, 19) Connelley et al. 2006, 20) Pajukanta et al. 2004, 21) Komulainen et al. 2006, 22) Helgadottir et al. 2005, 23) Chanson and Kwak 2007

Genes: PDE4D: phosphodiesterase 4D, MEF2A: myocyte enhancer factor 2A, KALRN: kalirin, GATA2: GATA binding protein 2, USF1: upstream transcription factor 1, ALOX5AP: arachidonate 5-lipoxygenase-activating protein, CX37: gap junction protein, alpha 4

Abbreviations: CAD: coronary artery disease, MI: myocardial infarction, ACS: acute coronary syndrome, CHD: coronary heart disease, FCHL: familial combined hyperlipidemia, IMT: intima media thickness

2.3.4 Previous findings in association studies

Genome-wide studies

Genome-wide studies augment candidate gene studies that rely on *a priori* hypothesis of biological pathways. To date, six published genome-wide association studies have searched for loci and variants conferring risk for coronary heart disease or stroke (Table 6). None of the studies found genome-wide significant association, but reported several interesting suggestive regions and candidate genes. An association in a locus in chromosome 9p21 has been found in three separate studies, in several independent replication study samples, implicating that this region harbours variants with a significant impact on risk for CAD (Helgadottir et al. 2007, Larson et al. 2007, McPherson et al. 2007). Of the linkage studies listed in Table 5, study by Wang and co-workers (2004) report a suggestive linkage to the same region in 9p21. Furthermore, significant association with type 2 diabetes has been found in the neighbouring region (Saxena et al. 2007). The most significant variants in the associating region do not map to any known genes and are in the same 190 kb LD block. There are two genes *CDKN2A* and *CDKN2B*, two exons of an mRNA transcript, a hypothetical methylthioadenosine phosphorylase fusion protein mRNA and several ESTs that are expressed in various tissues within this large block. In-depth-sequencing of the two genes did not reveal any likely candidates for functional variants and further studies are needed to elucidate the role of the region in the development of CAD (Helgadottir et al. 2007, McPherson et al. 2007).

In a Japanese genome-wide association study, Ozaki and colleagues (2002) identified lymphotoxin-α gene as a potential candidate for myocardial infarction. Their study design included roughly 65 700 mostly gene-based SNPs and had consequently less power to detect the variants identified in the more recent GWA studies analysing 100 000– 500 000 SNPs. Later the Japanese group reported that galectin-2 protein interacts with LTA. A variant (C3279T) in intron 1 of galectin-2 encoding gene, *LGALS2*, is even more significantly associated with decreased risk of MI and this variant is associated with decreased secretion of LTA (Ozaki and Tanaka 2005). Replications of the associations with CHD have been more complex. Studies from both Asian and European populations have reported both positive and negative associations for *LTA* variants and a recent meta-analysis concluded that they do not have an effect on CHD risk. A drawback of the meta-analysis is that it did not adjust the analysis for age and sex (Clarke et al. 2006). Only one prospective study has studied both *LTA* and *LGALS2* gene variants and concluded that *LTA* was not associated with CHD while *LGALS2*:C3279T was associated with decreased risk for CHD in American females (Asselbergs et al. 2007).

Table 6. ***Genome-wide case-control association studies on coronary heart disease and stroke.***[a]

Reference	Population	Phenotype	N cases / controls[b]	Genotyping platform, # SNPs analysed	Suggested region(s) (gene(s))
Ozaki and Tanaka 2005	Japan	MI	1133 / 1006	PCR invader assay, 65 671	6p21 (*LTA*)
Helgadottir et al. 2007	Iceland, US	MI	1607 / 6728	Illumina Hap300, 305 953	9p21
McPherson et al 2007	US, Denmark	Early onset CAD	322 / 312	Custom oligonucleotide array, 72 864	9p21
Matarín et al. 2007	US	Ischemic stroke	249 / 268	Illumina Infinium Human-1 and HumanHap300, 408 803	-
Samani et al. 2007[c]	United Kingdom	CAD < 66 years	1926 / 2938	Affymetrix GeneChip Human Mapping 500K Array Set, 377 857	1p13.3 (*PSRC1*), 1q41 (*MIA3*), 2q36.3, 6q25.1 (*MTHFD1L*), 9p21.3, 15q22.33 (*SMAD3*)
	Germany	MI < 60 years	870 & 772 / 1644	Affymetrix *Sty*I & *Nsp*1, 272 602	
Larson et al. 2007	US	CVD CHD Heart failure Atrial fibrillation	142 / 1345 118 / 1345 73 / 1345 151 / 1341	Affymetrix 100K GeneChip, 70 987	Several suggestive loci including 9p21.3

[a] All studies included both sexes and, except for Larson *et al*. 2007, used internal replication of at least one more sample set as validation
[b] In the initial screen
[c] The replication sample from Germany had a different genotyping platform.
Genes: *LTA:* Lymphotoxin-α*; PSRC1:* proline-rich Protein; MIA3: melanoma inhibitory activity 3; *MTHFD1L:* mitochondrial isozyme of C1-tetrahydrofolate synthase.
Abbreviations: MI: myocardial infarction, CAD: coronary artery disease, CVD: cardiovascular disease, CHD: coronary heart disease

The story of LTA exemplifies how genome-wide studies can generate and augment hypotheses of new pathways involved in the pathophysiology of CVD, and at the same time, how replication in large well defined study samples in same and in different populations is essential for the validity of this type of study design.

Candidate gene studies

The field of cardiovascular research and complex diseases in general is famous for a plethora of association studies on candidate genes and lack of replication. Until December, 2005, nearly 5000 candidate gene studies had investigated the association between MI or CAD, and 329 polymorphisms in 152 genes and reported their findings in international journals. Of these, 102 genes and 192 polymorphisms show positive and reproducible results in at least two independent populations. The inconsistency in the results is most likely due to underpowered studies and false positive findings combined with ethnic variation and sex-specific differences in the susceptibility genes (Mayer et al. 2007). Similarly, until January 2003, 120 studies had investigated the association between ischemic stroke and 51 polymorphisms in 32 genes. Only the four most studied polymorphisms (Factor V: Leiden, Methylenetetrahydrofolate reductase (*MTHFR*): C677T, Angiotensin converting enzyme (*ACE*): insertion/deletion (I/D), Prothrombin: G20210A) showed significant modification of risk in the meta-analysis while the others did not (Casas et al. 2004).

Association studies have observed that the effect magnitude of the risk-associated allele is often quite moderate (Altshuler and Daly 2007). Consistent association results for genes like *APOE*, *APOB*, nitric oxide synthase 3 (*NOS3*), *ACE*, plasminogen activator inhibitor-1 (*PAI1* or *SERPINE1*), *MTHFR*, integrin-α-2b (*ITGA2B*), paraoxonase 1 (*PON1*), *LPL*, and *CETP* have inspired researchers to try to establish the predictive value of the combined product of several polymorphisms. In their approach Drenos and co-workers (2007) derived unadjusted summary risk estimates and allele frequencies for these SNPs from published meta-analyses and calculated estimated numbers of individuals with different numbers of risk genotypes. While it was clear based on the analysis that individuals with 6 or more risk alleles had an increased risk for CHD compared to individuals with 3-4 risk alleles, it is still unclear how much these 10 SNPs add to the predictive power of traditional risk factors since adjusted risk estimates were not available.

Previous findings on serum lipid levels in genome-wide association studies

Several large genome-wide association studies including data from thousands of individuals have looked for variants influencing the variability in serum lipid levels (Kathiresan et al. 2007, Saxena et al. 2007, Kathiresan et al. 2008, Sandhu et al. 2008, Wallace et al. 2008). They have confirmed the association of many already

known susceptibility genes but interestingly have identified only very few new candidate loci. For LDL-C the confirmed findings included, for example, *APOB*, *APOE-C1-C4-C2* -cluster and *LDLR*; for HDL-C: *ABCA1*, cholesteryl ester transfer protein (*CETP*) and *LPL*, and for triglycerides: apolipoprotein A5, *APOB*, and *LPL*. Among the most consistent new loci is a region in chromosome 1p13, which has been significantly associated with LDL-C levels in three of the GWA studies (Kathiresan et al. 2008, Sandhu et al. 2008, Wallace et al. 2008). In ~15 200 European samples, two SNPs *rs599839* and *rs646776*, which are located 3' of two genes, cadherin EGF LAG seven-pass G-type receptor 2 and proline/serine-rich coiled-coil 1, explained both around 1% of the variation in the circulating levels of LDL-C (Sandhu et al. 2008). Each variant allele was associated with 15% standard deviation (SD) decrease in LDL-C assuming SD of 1 mmol/l. The small effect this variant has may have prevented the study by Kathiresan and co-workers (2007) from identifying this locus. In their study they had only 48% power to detect an effect of this magnitude. This locus is a very promising new candidate for a susceptibility locus for CHD since, in addition to associating with the levels of a strong CHD risk factor LDL cholesterol, the variant allele of the SNP (*rs599839)* has also been associated with decreased risk for CHD (Samani et al. 2007).

2.3.5 Sex-specific association results

CVD risk factors

Several family and twin studies have demonstrated that many of the cardiovascular disease risk factors have a sexually dimorphic genetic component. Differences in heritabilities or the genetic variance have been presented for the following traits: BMI (Schousboe et al. 2003), LDL-C, HDL-C, systolic blood pressure, fasting insulin, height and lymphocyte count (Weiss et al. 2006). Generally, traits associated with body fat, diabetes and anthropometric measurements have different heritabilities while there are conflicting results from twin versus family studies concerning, for example, blood pressure (Pilia et al. 2006, McCarthy 2007).

There are a few reported sex-specific associations between CVD risk factors and candidate genes. However, most of them did not conduct statistical evaluation for the interaction. In the French-Canadian population endothelin-1 was associated with HDL-C in women while there was no association in men (Paré et al. 2007). Among the Hutterites variations in the β3 integrin (*ITGB3*) gene were associated with lipoprotein(a) levels in women and with serotonin levels in men. In the joint analysis, the gene associated with both traits although there was no correlation between the two phenotypes highlighting the importance of sex-stratified analysis (Weiss et al. 2005).

Sex-specific association results on candidate genes

Previous large-scale association studies on CVD have mostly ignored formal testing of sex-specificity of their findings or have only studied both sexes combined. A study of 71 candidate genes found evidence of sex-specific association in two loci for men (Connexin 37: C1019T and $p22^{phox}$: C242T) and two for women (*PAI-1*: 4G/5G and Stromelysin-1: 5A/6A), but the authors did not report a test of interaction (Yamada et al. 2002). For metabolic syndrome, McCarthy and colleagues (2003) observed a nominally significant sex-gene interaction for nine genes and number of metabolic trait components (LDL receptor-related protein associated protein 1 (*LRPAP1*), Thrombospondin I (*THBS1*), Hepatic lipase (*LIPC*), Acetyl-Coenzyme A acetyltransferase 2 (*ACAT2*), Prolylcarboxypeptidase (*PRCP*), *LDLR*, *ITGB3*, Prothrombin (*F2*), P-selectin (*SELP*)). For CHD, Asselbergs and colleagues (2007) found a female-specific association for variation in the *LGLAS2* gene and demonstrated that the sex-gene interaction was significant at an $\alpha < 0.05$ level.

2.4 Inflammatory candidate genes of cardiovascular disease

2.4.1 Selenoprotein S

Selenoproteins are a group of proteins that contain a selenocysteine residue in place of cysteine. So far, 25 selenoproteins have been identified in humans. Most of them are enzymes that catalyse three types of reactions: glutathione peroxidases catalyse the reduction of hydrogen proxide and organic hydroperoxides thus protecting the cells from oxidative damage; iodothyronine deiodinases are essential in activating and inactivating thyroid hormones; thioredoxin reductases catalyse NADPH-dependent reduction of oxidized thioredoxin and help the cell to regulate redox balance. (Behne and Kyriakopoulos 2001)

Selenoprotein S (*SEPS1,* also called *SELS, SELENOS, VIMP*, listed as *SELS* in human genome build 18) is a protein, which protects cells from oxidative stress and regulates inflammatory responses. It resides in the plasma membrane and in the endoplasmic reticulum (ER) and regulates ubiquitin dependent protein degradation by removing unfolded proteins from the ER. This retro-translocation event requires a close contact between a VCP ATPase and a Derlin-1 protein, and SEPS1 possibly facilitates this interaction in the ER (Kryukov et al. 2003, Ye et al. 2004). ER stress associated calcium depletion, glucose deprivation and cytokines TNF-α and IL-1β regulate *SEPS1* expression (Gao et al. 2004, Gao et al. 2006). SEPS1 expression, on the other hand, reduces secretion of TNF-α and IL-1β from human macrophages (Curran et al. 2005). SEPS1 has been detected in human plasma and is secreted by hepatocytes (Gao

et al. 2007). As part of the unfolded protein response machinery SEPS1 has an important role in protecting the cells and, possibly through this role, protects human macrophages *in vitro* from ER stress induced apoptosis (Kim et al. 2007).

The *SEPS1* gene is located in chromosome 15q26.3 and contains six exons. The protein consists of 189 amino acids and selenocysteine is the second to last one in the C-terminus (Kryukov et al. 2003). *SEPS1* promoter has two nuclear factor kappa-B (NF-κB) transcription factor binding sites and an ER stress response element (ERSE) which overlaps with the other NF-κB site (Gao et al. 2006) (Figure 7). NF-κB participates in hepatic synthesis of acute phase proteins, immune defence, and stress response by regulating transcription of several genes and has been implied in the development of CVD in many different pathways (Pahl 1999, Ahn et al. 2007). Both cytokines TNF-α and IL-1β, and ER stress (inhibition of protein glycosylation) induce *SEPS1* promoter. It is likely that cytokine induction operates via NF-κB and ER stress through ERSE element, but unlike some promoters with multiple transcription factor binding sites, these elements do not act cooperatively in the *SEPS1* promoter (Gao et al. 2006).

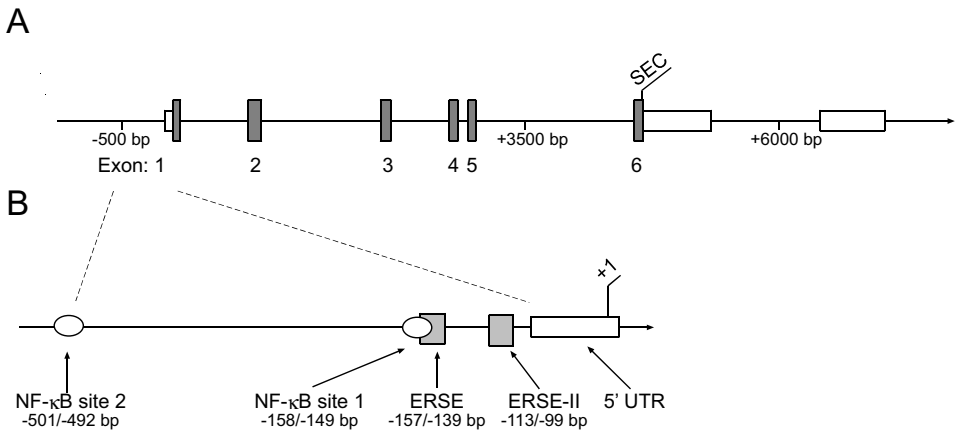

Figure 7. ***Selenoprotein S (SEPS1) gene region****. A) Location of exons and a selenocysteine residue in the SEPS1 gene (to scale). The open boxes indicate two major alternative splice variants in the 3' end according to the AceView genes (http://www.ncbi.nlm.nih.gov/ IEB/Research/Acembly/) B) The promoter area indicating nuclear factor kappa B (NFκB) binding sites and endoplasmic reticulum stress elements (ERSE).*

Three SNPs (*rs28665122*, *rs4965814*, *rs4965373*) in the *SEPS1* gene region associate with plasma concentrations of IL-1β, TNF-α and IL6. Variant *rs28665122* (G>A) is located near the ERSE element and in the middle of an ERSE-II-like element. During ER stress caused by inhibition of protein glycosylation, the presence of the A-allele reduces *SEPS1* promoter activity in HepG2 hepatocytes and the more frequent G-allele is required for the normal ER stress response. The minor allele A is associated with increased cytokine concentrations (Curran et al. 2005).

2.4.2 Interleukin 6 -related genes

Interleukin 6

Interleukin 6 is a pleiotropic cytokine that is a potent mediator of inflammatory processes and especially the acute-phase response (Heinrich et al. 1990). Several mechanisms including infection, cytokines IL1 and TNF-α, INF-γ, bacterial endotoxins and catecholamines induce its synthesis (van Deventer et al. 1990, Ng et al. 1994, Sanceau et al. 1995). Elevated levels of IL6 are associated with increased risk for myocardial infarction in healthy men independent of the levels of CRP (Ridker et al. 2000). Inflammatory markers CRP and IL6 have been associated with carotid atherosclerosis both prospectively and retrospectively (Elias-Smale et al. 2007, Lee et al. 2007, Thakore et al. 2007).

The interleukin 6 gene is located in chromosome 7p15.3 (Table 7). Variation in the gene has been associated with IL6 levels and haplotypes within the gene region have been associated with transcriptional regulation of the gene (Terry et al. 2000, Walston et al. 2007). Some studies have found an association between variation in the IL6 gene and CVD outcome (Georges et al. 2001, Humphries et al. 2001), while others have not (Walston et al. 2007). In particular, a SNP in the promoter region (-174G>C, *rs1800795*) has been widely studied. Recently a meta-analysis concluded that *IL6*: -174G>C SNP is not associated with CHD (Sie et al. 2006). It is possible, however that smoking is a significant modifier of the effects of IL6. Thakore and colleagues (2007) observed a significant interaction between smoking status and IL6 levels and internal carotid artery IMT, and both Georges (2001) and Rosner and colleagues (2005) observed a significant interaction with the risk for MI. The meta-analysis was not stratified according to smoking, but in the same report, the authors investigated this relationship in a prospective cohort of 463 incident cases and approximately 5 500 event-free individuals and found no association in smokers nor in non-smokers.

Table 7. ***Interleukin 6 -related CVD candidate genes in this study.***

Gene	Symbol	Location	Exons	Genomic coverage (kb)[a,b]	Polypeptide length (amino acids)[b]
Angiotensin converting enzyme	ACE	17q23.3	25	20.5	1306
Antiotensin II receptor, type I	AGTR1	3q24	2	45.1	359
Interleukin 6	IL6	7p15.3	5	4.8	212
C-reactive protein	CRP	1q23.2	2	2.3	224
Fibrinogen alpha	FGA	4q32.1	6	7.6	866
Fibrinogen beta	FGB	4q32.1	8	8.1	491
Fibrinogen gamma	FGG	4q32.1	9	8.2	453

[a] According to AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html) all genes exhibit alternative transcript lengths or alternative splicing
[b] From USCS Gene Sorter (http://genome.ucsc.edu)

The renin-angiotensin-system

The Renin-angiotensin-system (RAS) is an important regulator of cardiovascular hemodynamics and is also involved in the development of cardiovascular disease. In this system, renin cleaves angiotensinogen to form angiotensin I (Ang I) and ACE then converts Ang I into vasoactive angiotensin II (Ang II). ACE also inactivates a vasodilator agent bradykinin efficiently regulating the RAS system. Angiotensin II receptor type I (AT1R, gene: *AGTR1*) is the main mediator of the physiological effects of Ang II (Davis and Roberts 1997) and also mediates the Ang II stimulated release of IL6 from human adipocytes via activation of NF-κB (Skurk et al. 2004). The effect of the RAS system on atherosclerosis include the stimulatory effect of Ang II via AT1R1 on expression of VCAM1 in endothelial cells (Pueyo et al. 2000), and secretion of matrix metalloproteinase by human VSMCs (Browatzki et al. 2005). Furthermore, oxidized LDL has been shown to increase the expression of *AGTR1* in human coronary artery endothelial cells and VSMCs suggesting a link between the progression of atherosclerosis and AT1R (Li et al. 2000).

The genes encoding components of the RAS system are located in chromosomes, 17q23.3 (*ACE*), 3q24 (*AGTR1*) (Table 7), and 1q42-q43 (angiotensinogen). Genetic polymorphisms in RAS system and their relation to the risk for hypertension and CVD have been studied extensively. In particular, *ACE* insertion/deletion polymorphism in the 16th intron of the gene and *AGTR1*: A1166C (*rs5186*) in 3'-untranslated region (UTR) of the gene have been a focus of wide interest. The *ACE*: I/D polymorphism is associated with levels of ACE enzyme (Rigat et al. 1990). In meta-analysis reports these two variants were associated with increased risk for MI: *ACE* I/D homozygosity for the D-allele with pooled odds ratio (OR) of 1.21 (95% confidence interval (CI): 1.11-1.32) and C-allele of *AGTR1*: A1166C with an OR 1.13 per allele (CI: 1.01-1.23) (Agerholm-Larsen et al. 2000, Ntzani et al. 2007).

Some studies have also found epistatic interactions between genes in the RAS system (Ye et al. 2003, Tsai et al. 2007).

C-reactive protein and fibrinogen

Interleukin 6 is the main inducer of the hepatic production of CRP and fibrinogen (Heinrich et al. 1990). CRP and fibrinogen are both acute phase proteins and increased levels of both of them have been associated with increased risk for CVD and MI (Danesh et al. 2004, Mannila et al. 2004). In addition to having inflammatory effects, fibrinogen is an essential component of the coagulation cascade and may contribute to the risk for CVD via increased risk for coagulation (hypercoagulability hypothesis), lesion progression and enhanced endothelial dysfunction (Spronk et al. 2004).

The *CRP* gene is located in chromosome 1q23.2 (Table 7). Several variants in the gene have been associated with CRP levels in plasma. Of these variants a tri-allelic SNP in the promoter region has been confirmed to be functional (Carlson et al. 2005, Miller et al. 2005, Szalai et al. 2005, Kathiresan et al. 2006a). SNP and haplotype analysis in the *CRP* locus have found both supporting and opposed evidence for the effect of *CRP* gene variants and the association with CAD (Miller et al. 2005, Kardys et al. 2006, Lange et al. 2006).

Fibrinogen is a dimeric protein with each unit comprising of three polypeptides Aα, Bβ and γ. Genes encoding the polypeptides, *FGA*, *FGB* and *FGG*, are located in chromosome 4q28 in a tight cluster (Table 7). Transcription of the genes is coordinated and synthesis of the β-chain is the rate-limiting step of fibrinogen synthesis (Roy et al. 1990). Genetic polymorphisms in the *FGB* gene but not *FGA* or *FGG* have been associated with the levels of fibrinogen (Kathiresan et al. 2006b). Several studies have investigated the relationship between genetic variation in the fibrinogen gene locus and the risk for CVD. Most studies have focused on the promoter region of *FGB* gene and to a SNP -148C>T (*rs1800787*). A recent meta-analysis concluded that this variant is not associated with CHD (Smith et al. 2005). Studies using haplotypes across the fibrinogen gene cluster have identified a risk associating haplotype from *FGG* and *FGA* genes (Mannila et al. 2005) and contradicted the role of haplotypes in the *FGG* gene (Uitte de Willige et al. 2006) and in the whole fibrinogen gene cluster (Kathiresan et al. 2006b).

Interactions between interleukin 6 -related genes

IL6 participates in many reactions relevant to atherosclerosis and risk for CVD. The relationship between IL6 and the renin-angiotensin system is complex and apparently works in two directions. The Ang II induced hypertension is partially controlled by IL6 (Coles et al. 2007); Ang II induces secretion of IL6, facilitated by

AT1R, from leukocytes *ex vivo* during oxidative stress (Willemsen et al. 2007) and from human VSMCs *in vitro* (Kranzhöfer et al. 1999); and, on the other hand, IL6 induces expression of *AGTR1* in rat VSMCs both *in vitro* and *in vivo* (Wassmann et al. 2004).

Polymorphisms in the *IL6* gene have been associated with circulating levels of CRP and fibrinogen (Vickers et al. 2002, Ferrari et al. 2003, Sie et al. 2006, Walston et al. 2007, Wong et al. 2007). Furthermore, a Korean study demonstrated that variants *IL6*: −572C>G (*rs1800796*) and *CRP*: −717G>A (*rs2794521*), +1444C>T (*rs1130864*), and +2147A>G (*rs2808631*) show additive effects on serum hsCRP and insulin concentration, and measurement of insulin resistance (homeostasis model assessment, HOMA-IR) (Paik et al. 2007).

3 AIMS OF THE STUDY

The aim of the present study was to investigate the inflammatory mechanism and its relation with cardiovascular disease in large-scale genetic studies. The following specific aims were addressed:

1. To identify underlying causes of variation in the DNA extraction yield in large population surveys (I).
2. To examine the relationship between the risk for CVD and common variants in the selenoprotein S gene (II), and to analyse the function of the downstream polymorphism (unpublished data).
3. To study the risk for CVD and 46 different CVD candidate genes related to inflammation, thrombosis and lipid and energy metabolism focusing on a subset of these genes all related to interleukin 6 by using two prospectively followed Finnish case-cohort sets (III).

4 MATERIALS AND METHODS

4.1 Study subjects

4.1.1 Health 2000 survey

The Health 2000 survey aims to assess the health and functional capacity of Finns. The cohort is based on a nationally representative clustered random sampling drawn from the Finnish population register and included 8028 individuals aged 30 years and above. Of these 6968 (87%) were interviewed at home and after the interview 6770 (84%) either participated in a clinical examination or were examined at home. Fasting blood samples were drawn only from those participating in clinical examination. The study was approved by the local ethics committees, the participants signed an informed consent and the principles of the Helsinki Declaration were followed (Aromaa and Koskinen 2002). A whole blood sample for DNA extraction was available from 6451 (80%) individuals.

4.1.2 FINRISK 92 and 97 surveys

The FINRISK surveys are designed to assess the risk factors levels of chronic, noncommunicable diseases including cardiovascular disease in defined geographical areas. The studies cover south-western, southern, north-eastern (1992 & 1997) and northern (1997) parts of Finland. The survey samples were drawn from the National Population Register and the selection was stratified by age, sex and geographical area. Participants of these two studies were 25 to 64 years of age at the baseline examination (25-75 in the FINRISK 97 study) and the participation rates were 76% (n = 5999) and 73% (n = 8141), respectively. Both cohorts were followed up for fatal or non-fatal coronary event or ischemic stroke or venous thromboembolic events through the myocardial infarction and stroke registers (Salomaa et al. 2003, Sivenius et al. 2004), the National Hospital Discharge Register and the National Causes of Death Register. All participants gave their informed consent, both studies were approved by the local ethics committees, and they followed the principles of the Helsinki Declaration. In the FINRISK 97 cohort, a whole blood sample for DNA extraction was available from 7982 individuals (69% of the original selection).

Case-cohort selection

The two sample sets for studies II and III were derived from the two independent FINRISK cohorts recruited and sampled in 1992 and 1997 and followed until 2001 and 2003, respectively (Vartiainen et al. 2000). From these two cohorts, two separate sample sets were selected for genotyping in a case-cohort design (n = 999 and 1223, respectively). The both FINRISK cohorts participated in a multi-national project MORGAM examining genetic predisposition to CHD and the stroke and the case-cohort selection in this study was made according to principles agreed to in the project (Evans et al. 2005, Kulathinal et al. 2007). The selection included individuals with history of myocardial infarction or ischemic stroke defined as prevalent cases and individuals who suffered from a cardiovascular disease (myocardial infarction, unstable angina or ischemic stroke) event during the follow-up period defined as incident cases. Individuals who had suffered from a venous thromboembolic event or had died during the follow-up were also genotyped. A sex and geographic region matched random subcohort was also drawn from both of the original cohorts with unequal sampling probabilities so that their age distributions were similar to the cases. These subcohorts represented the general population of the areas included in the study. Figure 8 presents the design of studies II and III with number of individuals in each study end-point group from study III.

4.2 Laboratory procedures and statistical analysis

4.2.1 Methods in publications

Table 8 lists the methods used in this study.

4.2.2 Biochemical markers

Serum lipid levels were measured from fresh samples using routine enzymatic methods (CHOD-PAP, Monotest, Boehringer, Mannheim, Germany). Inflammation markers hsCRP, IL6 and TNF-α were measured in the FINRISK 92 cohort with the solid-phase cheminoluminescent immunometric assay (Immulite®, Diagnostic Products Corporation, Los Angeles, USA) and hsCRP with an immunoassay (Sentinel Diagnostics, Milan, Italy) using the Architect c8000 instrument (Abbott Laboratories, Irving, USA) from the FINRISK 97 samples. CRP measurement was not available from 167 individuals (16.7%) in the FINRISK 92 and from 67 individuals (5.5%) in the FINRISK 97 cohort. Careful examination of FINRISK 92 cohort revealed no significant differences in clinical or laboratory characteristics between individuals with and without CRP measurement.

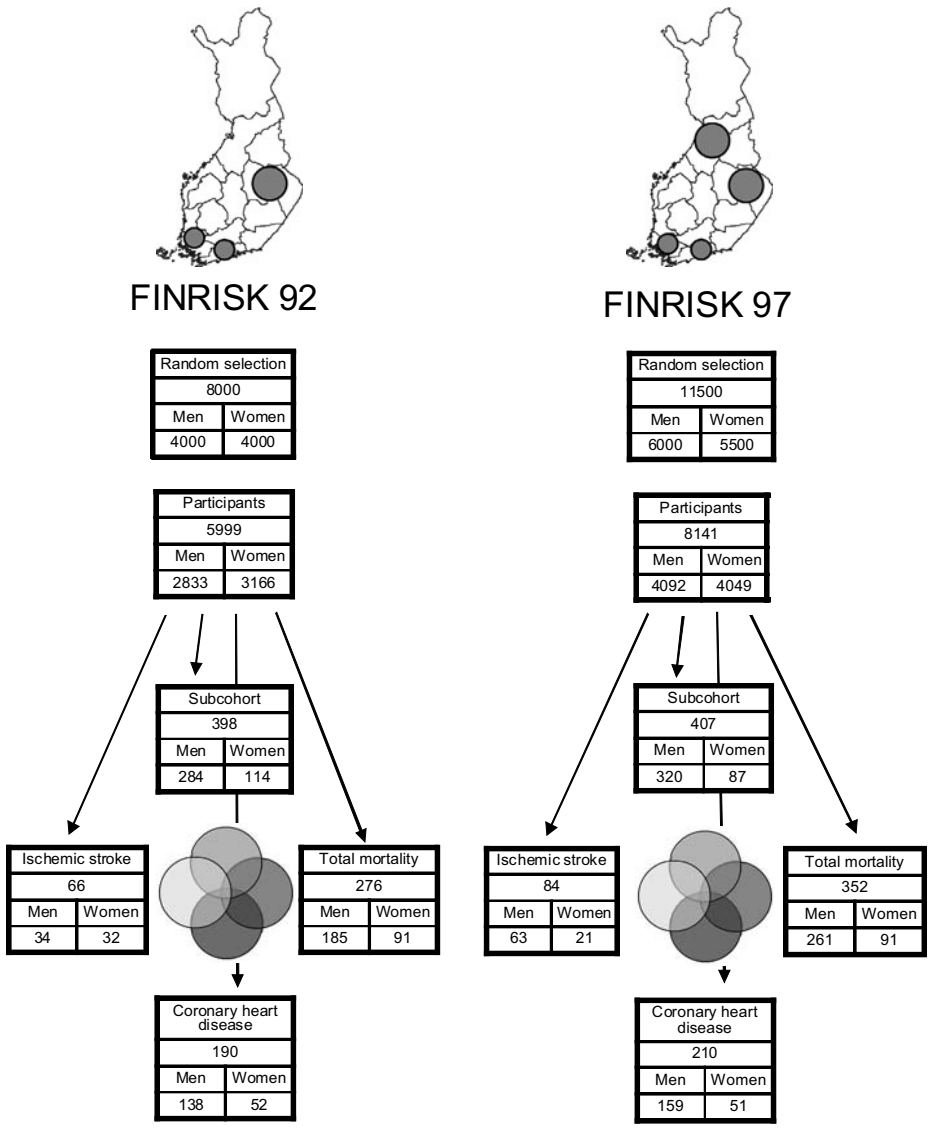

Figure 8. ***FINRISK study regions and number of selected subjects in study III.*** *The analysis set included an age-weighted, sex-and geographic region stratified, randomly selected control subcohort representing the study population, fatal and non-fatal incident ischemic stroke and coronary heart disease cases, and individuals who died during the follow-up. The different end-point groups are not mutually exclusive.*

Table 8. *Methods used in the present study.*

Method	Original publication
Study samples	
Health 2000	I
FINRISK 97	I
FINRISK 92 & 97 Case-cohort sample set	II, III
Laboratory procedures and measurements	
DNA extraction	I, II
DNA sample quality control	
Spectrophotometry	I, II, III
PicoGreen measurement	II, III
Polymerase chain reaction (PCR)	II, III
Whole genome amplification	II, III
Agarose gel electrophoresis	III
Allele-specific primer extension -based micro array	III
Sequenom MassArray	II, III
TaqMan SNP genotyping	II, III
Biochemical measurements	II, III
Analysis programs	
SNPsnapper	III
Primer3	III
Sequenom AssayDesigner (V3.0.1.1 beta)	II, III
Sequenom MassaArray Typer (v. 3.4)	II, III
ABI 7900HT SDS 2.2	II, III
Pedcheck 1.1	II, III
Alibaba2	II
P-match	II
Statistical methods and programs	
t-test	I
Linear regression	I, III
Cox's proportional hazard model, time-to-event analysis	II, III
Analysis of covariance	II
Wilcoxon–Mann–Whitney rank test	II
Phase 2.1	II, III
haplo.glm	III
SNPSpD	II
False discovery rate	III
Haploview 3.32	II, III
SPSS 11.0	I
SAS 9.1 for Windows	II, III
R-statistical environment	III

Interleukin 6 measurements were available only from 309 subcohort individuals free of baseline CVD from the FINRISK 92 cohort. Serum fibrinogen levels were also available only from a selected subset of individuals: in the FINRISK 92 cohort, from 203 women and 377 men participating in the FINRISK 92 Hemostasis study (Rajecki et al. 2005), and in the FINRISK 97 cohort only from 528 men of the PAIS sub-sample (Jousilahti et al. 2001). Among men, the baseline characteristics of these groups differed significantly: FINRISK 97 participants were older ($p < 0.0001$), were more often daily smokers ($p < 0.0001$), had a history of diabetes ($p = 0.046$) and CVD ($p < 0.0001$), had hypertension ($p = 0.017$), lower levels of triglycerides ($p = 0.002$), HDL-C to total cholesterol ratio ($p < 0.0001$), and hsCRP ($p = 0.0004$) and higher levels of fibrinogen ($p < 0.0001$). Fibrinogen was measured from snap-frozen plasma with the Clauss method (Clauss 1957) in the FINRISK 97 cohort and in the FINRISK 92 cohort with a prothrombin time assay (IL Test PT-Fibrinogen, Instrumentation Laboratories, Milan, Italy) measuring the scattered light from the clot.

4.2.3 DNA extraction and quality control

EDTA anticoagulated whole blood samples were stored in -20ºC in both the Health 2000 and FINRISK surveys. DNA from the FINRISK samples was extracted with a phenol-chloroform protocol modified from Vandenplas and colleagues (1984) and DNA from the Health 2000 samples with Gentra Puregene reagents (QIAGEN, GmbH, Hilden, Germany). We quantified the accurate concentration of DNA aliquots prior to normalization with the PicoGreen dsDNA Quantitation Kit (Invitrogen (Molecular Probes), Carlsbad, USA) and amplified samples with low yield of DNA with GenomePhi whole genome amplification kit (GE Healthcare (Amersham Biotech), Chalfont St. Giles, UK). We determined the sex from the DNA sample with sex chromosomal markers and excluded all samples with discrepant results. Furthermore, we genotyped additional autosomal microsatellite markers in order to identify samples contaminated with another DNA.

4.2.4 Candidate genes and variant selection

Study II focused on the *SEPS1* gene and study III on IL6 -related genes: *ACE*, *AGTR1*, *CRP*, *IL6*, *FGA*, *FGB*, and *FGG*. All the genes investigated in study III are listed in Table 9. We selected common (>5%) variants for study II from dbSNP (www.ncbi.nlm.nih.gov/SNP/) and HapMap (The International HapMap Consortium 2003) databases based on available valid frequency data, haplotype information aiming at an even distribution along the gene. In addition, variants from previous association reports were included. In study III we searched the SeattleSNP variation discovery database (SeattleSNPs) and selected all haplotype bin tagging SNPs

(tagSNP) with >5% minor allele frequency in the European decent sample to fully capture the variation in the genes. In addition, we selected all SNPs with prior evidence of association with CVD, metabolic syndrome or protein product levels. For genes, such as, *AGTR1* where tagSNP data was not available in the SeattleSNP database, we selected the SNPs based on frequency, validation data and location from the dbSNP database. Once HapMap phase I data became available, we also used the additional information it provided. From the *ACE* gene we also genotyped the Alu insertion/deletion polymorphism in the FINRISK 92 sample. Table 10 presents variants selected for IL6 -related genes.

4.2.5 Variant genotyping and quality control

This study utilized three different genotyping methods. Primarily we used the Sequenom MassArray system (Sequenom, San Diego, USA), either with the homogeneous Mass Extension (hME) reaction or the iPLEX reaction. For the hME genotyping we used 5-7.5 ng of genomic and 10 ng of WGA DNA. The second genotyping method was allele-specific primer extension on microarray (Pastinen et al. 2000) with 10 ng of genomic or WGA DNA. A few variants were genotyped using the TagMan SNP genotyping system (Applied Biosystems, Foster City, USA). The *ACE*: I/D polymorphism was genotyped using PCR amplification and separation on an agarose gel. The genotyping sample included 5% of blinded duplicates and each 96-well plate included two open duplicate samples, two CEPH (Centre d'Etude du Polymorphisme Humain) controls and two negative controls. In addition, WGA samples were genotyped in duplicate.

4.2.6 Bioinformatic analysis of transcription factor binding

We used two transcription factor (TF) binding site prediction programs Alibaba2 and P-match to investigate putative TFs interacting with 36 bp surrounding sequence of variant *rs7178239*. Alibaba2.1, by Niels Grabe, utilizes the TRANSFAC 4.0 library of known TF-binding site alignments to predict binding the TFs with the target sequence. (http://www.gene-regulation.com/pub/programs.html#alibaba2) We used three different prediction conditions: 1) loose criteria with pair-wise similarity score 36, matrix width 10 bp, minimum number of sites = 4, low matrix conservation, 1% similarity between the sequence and matrix, factor class level = Family; 2) medium tight criteria: pair-wise similarity score 50, matrix width 10 bp, minimum number of sites = 4, medium matrix conservation, 1% similarity between the sequence and matrix, factor class level = Subfamily; 3) tight criteria: pair-wise similarity score 50, matrix width 10 bp, minimum number of sites = 4, medium

Table 9. ***Selected candidate genes in study III.***

Pathway and genes	Gene symbol	Location
Inflammation & thrombosis		
carboxypeptidase B2 (plasma)	CPB2	13q14.12
CD14 molecule	CD14	5q31.3
coagulation factor II (prothrombin)	F2	11p11.2
coagulation factor II (thrombin) receptor	F2R	5q13.3
coagulation factor V (proaccelerin, labile factor)	F5	1q24.2
coagulation factor VII (serum prothrombin conversion accelerator)	F7	13q34
coagulation factor X	F10	13q34
coagulation factor XII	F12	5q35.3
coagulation factor XIII, A1 polypeptide	F13A1	6p25.1
C-reactive protein, pentraxin-related	CRP	1q23.2
fibrinogen alpha chain	FGA	4q32.1
fibrinogen beta chain	FGB	4q32.1
fibrinogen gamma chain	FGG	4q32.1
integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	ITGA2	5q11.2
integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	ITGB3	17q21.32
intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	ICAM1	19p13.2
interleukin 1, alpha	IL1A	2q13
interleukin 1, beta	IL1B	2q13
interleukin 10	IL10	1q32.1
interleukin 6 (interferon, beta 2)	IL6	7p15.3
lectin, mannose-binding, 1	LMAN1	18q21.32
lymphotoxin alpha (TNF superfamily, member 1)	LTA	6p21.33
plasminogen activator, tissue	PLAT	8p11.21
protein C (inactivator of coagulation factors Va and VIIIa)	PROC	2q14.3
selectin E (endothelial adhesion molecule 1)	SELE	1q24.2
selectin L (lymphocyte adhesion molecule 1)	SELL	1q24.2
selectin P (granule membrane protein 140kDa, antigen CD62)	SELP	1q24.2
selenoprotein S	SEPS1	15q26.3
serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	SERPINE1	7q22.1
thrombomodulin	THBD	20p11.21
tumor necrosis factor (TNF superfamily, member 2)	TNF	6p21.33
vascular cell adhesion molecule 1	VCAM1	1p21.2
Lipids & energy		
apoliprotein A-V	APOA5	11q23.3
apoliprotein E	APOE	19q13.32
forkhead box C2 (MFH-1, mesenchyme forkhead 1)	FOXC2	16q24.1
lactase	LCT	2q21.3
lipin 1	LPIN1	2p25.1
neuropeptide Y	NPY	7p15.3
thioredoxin interacting protein	TXNIP	1q21.1
upstream transcription factor 1	USF1	1q23.3
Others		
5,10-methylenetetrahydrofolate reductase (NADPH)	MTHFR	1p36.22
angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	ACE	17q23.3
angiotensin II receptor, type 1	AGTR1	3q24
apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 2	APOBEC2	6p21.1
fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group)	FUT3	19p13.3
klotho	KL	13q13.1

Table 10. ***Selected variants in the interleukin 6 -related genes.***

Gene	Variant	Alleles	Alias	Location	Amino acid change	Minor allele frequency[a]
ACE	rs4293	G>A		Intron		0.49
	rs4298	C>T		Coding exon	N/N	0.01
	rs4320	G>A		Intron (boundary)		0.43
	indel	Del>Ins	ACE I/D	Intron		0.41[b]
AGTR1	rs275652	T>G	M-713C	Promoter		0.14
	rs1492078	C>T	C-521T	Promoter		0.32
	rs2276735	G>A		Intron		-[c]
	rs1492101	G>A		Intron		0.18
	rs3772616	C>T		Intron		0.2
	rs388915	A>G		Intron		0.21
	rs5182	T>C	M573C	Coding exon	L/L	0.41
	rs5186	A>C	A1166C	3' UTR		0.18
	rs380400	A>G		3' UTR		-[c]
CRP	rs2794521	T>C	-717A/G	Promoter		0.2
	rs3091244	G>A>T	-286C/T/A, -390C/T/A	Promoter		0.36 / 0.10
	rs1800947	C>G	+1059G/C	Coding exon	L/L	0.06
	rs1130864	G>A	+1444C/T	3' UTR		0.34
	rs1205	C>T	+1846G/A	3' UTR		0.38
	rs3093075	G>T		3'		0.07
FGA	rs2070006	C>T		Promoter		0.47
	rs2070016	T>C		Intron		0.14
	rs2070018	A>G		Intron (boundary)		0.15
	rs6050	T>C		Coding exon	A/T	0.31
	rs2070022	G>A		3' UTR		0.22
FGB	rs1800788	C>T		Promoter		0.25
	rs1800787	C>T	-148C/T	Promoter		0.19
	rs2227412	A>G		Intron		0.19
	rs2227439	C>T		3'		0.15
	rs1044291	C>T		3'		0.23
FGG	rs1800792	T>C		Promoter		0.31
	rs2066860	C>T		Intron		0.02
	rs2066861	C>T		Intron		0.3
	rs1049636	A>G		Intron (boundary)		0.39
IL6	rs2069825	+CT>-CT		Promoter		0.49
	rs2069827	G>T		Promoter		0.26
	rs1800796	G>C	-572G/C	Promoter		0.03
	rs1800795	C>G	-174G/C	Promoter		0.46
	rs2069840	C>G		Intron		0.23
	rs1554606	T>G		Intron		0.43

[a] Estimated in the combined subcohort of both sample sets.
[b] Estimated only in the FINRISK 92 subcohort.
[c] Success rate < 95%.

matrix conservation, 100% similarity between the sequence and matrix, factor class level = Subfamily.

P-match, by Dmitry Chekmenev, Carla Haid and Alexander Kel, predicts TF binding using TRANSFAC 6.0 library and combines both pattern matching and nucleotide weight matrix approaches. (http://www.gene-regulation.com/pub/programs.html#pmatch) We used TF-binding prediction with two levels of stringency: cut-off to minimize false positive matches and a less stringent criterion with a cut-off to minimize false positive matches.

4.2.7 Electrophoretic mobility shift assay

Twenty micrograms of single stranded allele specific sense and antisense oligonucleotides for the *rs7178239* (5'-GCTACTTCAAAGAGAAG[C/G]GGGCTT AACCTGTATCAA-3'), non-specific competitor oligo containing a NF-κB binding site (5'-CCCACGTGTGATGGAAAGTCCAAAATTCTACAGGAGTCT-3') and an oligo where the NF-κB binding site was mutated (3'- CCCACGTGTGATTTAAAG TCCAAAATTCTACAGGAGTCT-5'), were annealed in buffer containing 10 mM Tris-HCl, pH 8.0 and 200 mM NaCl. The double stranded oligos were first precipitated using 3M sodium acetate (pH 5.2) and absolute ethanol, diluted in 1xTris-EDTA (TE) buffer (pH 8.0), and then purified by electrophoresis through a 15% polyacrylamide gel, eluted from the gel, filtered with Costar Spin-X columns (Corning, Massachusetts, USA), and precipitated as previously. A total of 100 ng of the oligos were labelled with 48μCi of [γ-^{32}P]ATP (Perkin Elmer, Massachusetts, USA) using 8U of T4 polynucleotide kinase (New England Biolabs, Massachusetts, USA). End-labelled double-stranded oligos were purified from free ATP with QIAquick Spin columns (QIAGEN GmbH, Hilden, Germany) according to the manufacturer's instructions and eluted in 1xTE buffer.

Electrophoretic mobility shift assay (EMSA) reactions were set up in Gel Shift Binding Buffer (Promega, Wisconsin, USA) with 3 μg of HeLaScribe Nuclear Extract (Promega). For competition assays 0.5-, to 10-fold molar excess of cold competitor oligo, or non-specific oligo containing a NF-κB binding site, was added. The mix was incubated on ice for 10 minutes after which radioactively labelled oligos were added, incubated at room temperature for 1h and separated on 4% PAGE in 0.25x Tris-Borate-EDTA buffer at 200V for 1.5 h. The dried gel was exposed to X-ray film at -70°C overnight.

4.2.8 Statistical analysis

In studies II and III we assessed deviations from the Hardy-Weinberg equilibrium in the subcohort by using χ^2-statistic with 1 degree of freedom for bi-allelic SNPs and with 3 degrees of freedom for the tri-allelic SNP ($p < 0.05$ (II), $p < 0.01$ (III)). All variants except *SEPS1*: *rs4965814* (in II) followed the equilibrium. The extent of linkage disequilibrium at each locus was estimated with r^2 and D' statistics using program Haploview (v3.2). (Barrett et al. 2005)

We studied the relationship of the variant allele of each SNP and CHD, CVD (CHD and ischemic stroke combined) and all-cause mortality (II only) with time-to-event analysis using the weighted Cox's proportional hazard model. The model included subcohort sampling probabilities as weights (Kim and De Gruttola 1999) and variance correction based on the Barlow method (Barlow 1994) (in study II). We adjusted the analysis with sex and cohort when applicable, together with traditional CVD risk factors: history of diabetes, presence of absence of high blood pressure (individuals with medication for high blood pressure or systolic blood pressure > 140 mmHg or diastolic blood pressure > 90 mmHg), HDL-C to total cholesterol ratio (in II), HDL-C (in III), non-HDL-C (in III), daily smoking and body-mass index. In study II, the analyses were stratified according to known east-west difference in the incidence of CVD and distribution of the CVD risk factors (Salomaa et al. 1996) while in study III the western, eastern and northern regions of Finland were included as covariates. Prevalent CVD cases were excluded from time-to-event analyses. Study II used a dominant model of variant allele effect in all its genetic analyses and study III explored the genetic effect with recessive, dominant and multiplicative models. The analysis strategies in studies II and III were different: in study II we examined Cox's models first in the FINRISK 92 cohort separately in men and women and then replicated the findings in the FINRISK 97 cohort and estimated the hazard ratio in a joint analysis of both cohorts. In study III both men and women were first analysed together and in the second phase, a formal test of genotype-sex interaction was carried out. We defined interaction as significant ($p \leq 0.01$) departure from the recessive, dominant and multiplicative models. Furthermore, we tested cohort-genotype interaction with a similar method. The SAS statistical software (version 9.1. for Windows) (SAS Institute Inc., Cary, USA) procedure PHREG was used in the time-to-event analysis of study II, whereas the R statistical environment and the coxph function of the package survival and its robust variance estimator was used in study III (R Development Core Team 2006).

Study II analysed the relationship between three inflammatory markers hsCRP, IL6, TNF-α and obesity related measures BMI and WHR, and genetic variation in the *SEPS1* locus by using analysis of covariance. The analysis was adjusted for age and

geographic region, and when applicable, for sex and cohort. It was carried out separately in incident CVD cases and subcohort members free of disease. In study III we applied linear regression models adjusted for cohort, age, geographic region, and sex in subcohort members free of baseline CVD. The quantitative CVD risk factors analysed were serum total cholesterol, HDL-C, triglycerides, LDL-C, mean blood pressure (average of systolic and diastolic blood pressure, each value based on two subsequent measurements), hsCRP, BMI, and WHR for all genes in study III. Individuals with lipid lowering medication at the baseline were excluded from lipid analyses and individuals with medication for hypertension from the analysis of blood pressure. The genetic models studied were recessive, dominant and additive and departure from the models was considered as an indication of genotype-sex interaction. In addition, we analysed in subcohort individuals whether variants in the fibrinogen genes associated with fibrinogen levels and whether *IL6* gene variants associated with fibrinogen levels or IL6 levels in the FINRISK 92 cohort. SAS statistical software was used for the analysis of fibrinogen and IL6 levels and R statistical environment for the rest. All analyses on quantitative variables used logarithmic transformation where applicable.

Both studies used Phase 2.1.1 (Stephens et al. 2001) to estimate best-guess haplotypes for all individuals. The estimation was carried out separately in cases and subcohort individuals and only consistent results from 40 independent iterations were accepted. In addition, study III utilized haplo.glm function in the haplo.stats package of the R environment when analyzing the association between haplotypes and quantitative variables.

Multiple testing correction

This study utilized two different methods for multiple testing correction. In study II, we estimated the effective number of variants using the SNPSpD method (Nyholt 2004) in the combined subcohort of both FINRISK cohorts, and adjusted the significance threshold according to estimated number of independent SNPs. We applied the threshold $p < 0.0125$ within sex and end-point specific groups. Furthermore, study II used two independent cohorts for replicating the results.

Study III applied the false discovery rate method implemented in the R package "fdrtool" (Strimmer 2007) by interpreting the tail area of the FDR as the expected proportion of null results in a given group of tests. The presented time-to-event and quantitative risk factor association results (apart from fibrinogen and IL6 levels) for joint analysis of men and women are expected to have 47% true positive findings (corresponding to FDR value 53%). The FDR estimation included results from all 172 SNPs in the study III with p-value ≤ 0.01 in time-to-event analyses or with $p < 0.01$ in quantitative analysis. Study III also investigated Q-Q plots for individual

test statistics. Sex-genotype interaction analysis also included 172 variants and the reported findings with $p \leq 0.01$ are expected to include 30% of true positive findings. However, we applied an additional criterion of nominal association at the $p \leq 0.01$ level in time-to-event analyses or at the $p < 0.01$ level in quantitative analysis in either sex in joint analysis of both cohorts. This is likely to increase the proportion of true positive findings.

4.2.9 Analytical approach for joint analysis of genetic risk factors

Increasingly more candidate gene and GWA studies will be analysing the joint effects of genetic variants in the future (Daly and Altshuler 2005). Several analytical approaches are available for case-control studies but few can be implemented in case-cohort design. We wanted to investigate the joint effects of all SNPs in the IL6 -related pathway and did this by using Cox's proportional hazard model with dummy variable coding. The model asked whether the difference between having a variant allele from both SNPs and having the variant allele only from either of the SNPs was statistically significant while accounting for the more traditional risk factors.

To avoid analyzing redundant data we selected only one variant from the pair of variants with > 80% (r^2) correlation within gene or across fibrinogen gene cluster (excluded variants: *ACE*: I/D, *FGA: rs2070022*, *FGG: rs2066861*, *IL6: rs1554606*). In addition, we excluded two rare SNPs (<5% minor allele frequency) (*FGG: rs2066860* and *IL6: rs1800796*) and two SNPs (*AGTR1: rs2276735* and *rs380400*) due to low genotyping success rate (<95%). In the joint effect analyses variants *rs1130864* and *rs3093075* were used as proxies for the alleles A and T, ($r^2 = 0.997$ and 0.984), respectively, of three allelic SNP *rs3091244* in the *CRP* gene. After these quality control procedures, 30 SNPs entered final analyses. We constructed models for all SNP by SNP combinations (30 variants, 435 tests) in a) joint analysis of both sexes from one cohort, b) joint analysis of both cohorts of one sex, and c) joint analysis of both sexes and both cohorts and excluded cohort and sex-specific analyses because we expected that some SNP pairs would have a very small number of individuals carrying variant alleles from both SNPs.

In the joint analysis of IL6 -related genes, we controlled for the multiple testing using FDR implemented in the SAS procedure Multtest and with replication in two separate cohorts. An FDR limit <10% was applied in end-point, cohort and sex-specific groups of time-to-event analysis including results from individual SNPs in the interleukin 6 -related genes and joint effects analyses. The validity of the joint effect findings was also evaluated by inspecting the consistency between the two cohorts and the two end-points.

5 RESULTS AND DISCUSSION

5.1 Cardiovascular risk factors and association with the variation in DNA yield from whole blood samples (I)

In the first part of this study, we observed that the distribution of DNA yield from 10 ml whole blood sample approximates normal distribution (Figure 9) and wanted to investigate which factors contribute in this variation. We utilized two independent population-based surveys, Health 2000 (n = 6451) and a sub-sample from the FINRISK 97 survey (n = 1175), in which we knew that homogenous DNA extraction method had been used throughout the sample set. Table 11 describes the characteristics of the study participants.

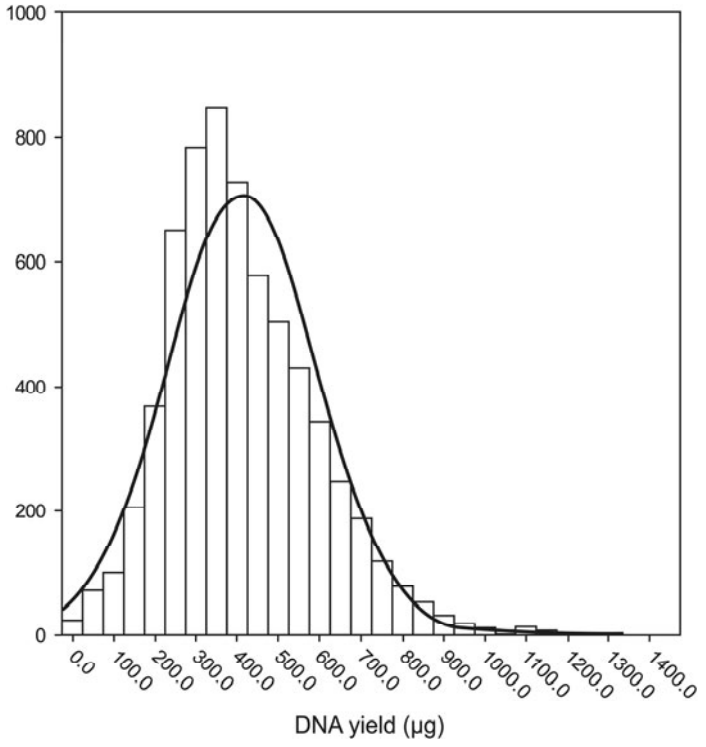

Figure 9. ***Distribution of DNA yields in the Health 2000 survey, (n = 6451).***

Table 11. ***Characteristics of participants in Health 2000 survey and a sub-sample of the FINRISK 97 survey.***

	Health 2000 survey			Sub-sample of FINRISK 97 survey		
Characteristic	N	Mean (SD) or %	Range	N	Mean (SD) or %	Range
Questionnaire						
Age (years)	6451	53.5 (15.4)	30-99			
Males	2903	51.9 (14.2)	30-97	1175	60.5 (8.6)	30-74
Females	3548	54.7 (16.2)	30-99			
Daily smoking (%)	1497	22.1				
Males	847	28.1		277	23.6	
Females	650	17.3				
Diabetes (%)	374	5.8				
Males	163	5.7		89	7.6	
Females	211	6				
History of myocardial infarction (%)	274	4.3				
Males	171	5.9		122	11	
Females	103	2.9				
Physical measurements						
Body mass index (kg/m^2)	5923	26.9 (4.65)	12.30-53.70	1106	26.07 (7.57)	17.45-47.85
Systolic blood pressure (mmHg)	6138	135.6 (21.37)	90-245	1174	147 (21.4)	78-225
Diastolic blood pressure (mmHg)	6031	82.6 (10.79)	60-124	1174	86.2 (11.62)	50-123
Laboratory measurements						
Total blood cholesterol (mmol/l)	6391	5.93 (1.13)	1.00-11.70	1175	5.72 (1.02)	2.40-13.6
HDL-cholesterol (mmol/l)	6448	1.33 (0.38)	0.23-3.41	1175	1.24 (0.34)	0.41-3.00
LDL-cholesterol (mmol/l)	5649	3.87 (1.01)	0.90-8.80	965	3.68 (0.88)	1.15-6.78
Triglycerides (mmol/l)	6445	1.62 (1.05)	0.4-16.6	1175	1.82 (1.11)	0.41-8.00
C-reactive protein (mg/l)				1085		0.0-98.7
Fibrinogen (g/l)				1045		1.84-7.7
Serum amyloid A (µg/l)				1111		0.1-701
White blood cell count (x10^9/l)				352		1.9-15.5
DNA yield (µg/ml of blood, geometric)	6451	36.82 (18.16)	1.0-143.4	1174	17.15 (13.06)	1.2-64.9
Time-to-extraction (month)	6451	1.56 (1.4)	0-15	1177	9.13 (1.06)	7-12

In the joint analysis of both sexes in the Health 2000 survey, the mean DNA yield was greater in individuals who had diabetes, hypertension, or were daily smokers or obese. When the analysis was stratified by sex, the result was not significant for obesity in females. In the FINRISK 97 survey, the only significant association observed was for smoking. In addition, we found a trend for increased DNA yield in myocardial infarction cases in men of both surveys (Table 12).

Table 12. ***The mean yield of DNA (µg/ml of blood) by presence or absence of a disease or a risk factor.***

		Health 2000			FINRISK 97
Disease or risk factor		**Both sexes**	**Females**	**Males**	**Males**
Diabetes	N (No/Yes)	6035/374	3314/211	2721/163	1087/89
	No, mean (SD)	41.1 (18.1)	41.0 (18.0)	41.2 (18.1)	21.4 (13.1)
	Yes, mean (SD)	45.1 (18.8)	44.4 (20.0)	46.1 (17.1)	20.8 (12.4)
	P-value[a]	<0.0001	0.014	0.001	0.69
History of myocardial infarction	N (No/Yes)	6145/274	3429/103	2716/171	979/122
	No, mean (SD)	41.2 (18.2)	41.1 (18.2)	41.4 (18.2)	21.4 (13.0)
	Yes, mean (SD)	42.5 (16.8)	41.5 (16.1)	43.1 (17.2)	23.7 (14.1)
	P-value[a]	0.256	0.833	0.222	0.074
Daily smoking	N (No/Yes)	5007/1444	2924/624	2083/820	898/276
	No, mean (SD)	40.0 (17.8)	40.3 (18.0)	39.7 (17.6)	20.8 (13.0)
	Yes, mean (SD)	45.6 (18.6)	45.2 (18.8)	45.8 (18.5)	22.9 (13.2)
	P-value[a]	<0.0001	<0.0001	<0.0001	0.02
Obesity[b]	N (No/Yes)	4259/1339	1967/569	2292/770	750/258
	No, mean (SD)	40.9 (17.8)	41.1 (17.7)	40.6 (18.0)	21.7 (13.1)
	Yes, mean (SD)	42.9 (19.0)	42.4 (18.9)	43.2 (19.2)	21.8 (12.6)
	P-value[a]	0.001	0.142	0.001	0.916
Hypertension[c]	N (No/Yes)	3288/2850	1379/1409	1909/1441	389/784
	No, mean (SD)	41.0 (18.1)	41.0 (18.0)	41.0 (18.1)	21.0 (12.5)
	Yes, mean (SD)	42.5 (17.9)	42.4 (18.1)	42.6 (17.7)	21.5 (13.3)
	P-value[a]	0.001	0.013	0.034	0.554

[a] t-test
[b] BMI ≥30 vs. <30 kg/m^2
[c] Systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg

In the Health 2000 survey, multivariate linear regression modelling adjusted for age, sex, technician and time of storage, examining each risk factor at the time, revealed significant association between DNA yield and BMI, daily smoking, history of diabetes, systolic blood pressure, triglycerides, HDL-C and history of myocardial infarction (data not shown). Of these, only smoking was significantly associated with DNA yield in the FINRISK 97 survey and the association with MI showed an insignificant trend (p = 0.082). Since many of the risk factors are correlated, we

wanted to evaluate their contribution to the DNA yield by a multivariable regression with a step-wise approach in the Health 2000 survey. This analysis including both sexes, indicated that daily smoking was associated with ~50µg increase ($p < 0.0001$) in DNA yield compared to non-smokers, history of diabetes with 33.1 µg increase ($p < 0.001$) compared to unaffected individuals, one SD increase in systolic blood pressure with 10.7 µg increase ($p < 0.0001$), one SD increase in triglycerides with 10.2 µg increase ($p < 0.0001$), and one SD increase in HDL-C with 10.8 µg decrease ($p < 0.0001$) in the DNA yield. The results were similar for men and women. When we tested the same model in the FINRISK 97 data, we observed that smoking, HDL-C, and triglycerides were significantly associated with the DNA yield and the trend was similar for systolic blood pressure. However, contrary to the results from the Health 2000 survey, the effect of one SD increase in triglyceride levels was associated with 9.9 µg decrease in the DNA yield ($p = 0.012$).

In the FINRISK 97 sample, we were able to analyse the association between DNA yield and inflammatory markers WBC, hsCRP, SAA and fibrinogen. In a multivariate model adjusted for age, sex, laboratory technician responsible for the extraction and time of storage, WBC was significantly and positively associated with the DNA yield and explained 15.2% of the variation ($p < 0.0001$, $n = 352$). The other inflammatory markers were also associated with the DNA yield when analysed separately. However, when WBC was included in the multivariate model, none of the other markers or trait phenotypes remained significant.

In conclusion, CVD trait phenotypes and inflammatory markers, especially WBC, were significantly associated with DNA yield from whole blood samples. We did not observe any significant factors associating with very low DNA yield. Most of the findings of this study are consistent in both cohorts. There are, however, a few discrepant results: the association of prevalent diabetes was not significant in the FINRISK 97 sub-sample, while it was a clearly associated in the Health 2000 survey, and the serum triglyceride concentration was significantly associated with higher DNA yield in the Health 2000 survey and lower DNA yield in the FINRISK 97 sub-sample. The most likely explanation for the first observation is the much smaller number of individuals with diabetes (374 vs. 89) in the FINRISK 97 sub-sample. In addition, there was a marked difference in the mean DNA yield between the two cohorts. Factors, such as different extraction method and longer storage time before extraction, resulting in much lower DNA yield in the FINRISK 97 sub-sample could potentially mask any borderline findings. The discrepant finding in triglyceride concentration could be due to differences in the length of fasting that were not accounted for in the analyses. It should be also noted that the FINRISK 97 sub-sample participants are all men and the upper age limit is higher at 74 years. It is possible that these elderly men have additional inflammatory conditions, such as

dental problems or chronic bronchitis, potentially confounding the present analyses. Finally, we cannot exclude pure chance and multiple testing as possible explanations for these discrepant findings.

Variation in the DNA yield can result in quite large unexpected differences between a group selected based on DNA availability and a group without selection bias. For example, if a study searching for genes associating with MI decides to select their control samples from the Health 2000 survey using only individuals without MI in the top quartile of the DNA yield (n = 1536), their control sample becomes enriched for smokers. The frequency of current smokers is 8% higher (30.5%) in this group than in the whole cohort (22.5%) ($p < 0.001$). Similarly, these results suggest that other CVD risk factors would be more common in this group and most likely the individuals would have higher level of systemic inflammation. Adjusting for the known risk factors associating with variation in DNA yield will aid in balancing the selection bias, but unidentified confounding factors will remain a source for potential bias.

Although we did not identify any factors associating with very low DNA yield, it is likely that they exist but are masked by relatively strong reducing effect of time of storage. To overcome the need to select the samples based on the amount of DNA (usually to avoid using samples with low DNA yield), the study should apply whole genome amplification to the low DNA yield samples so that they can be included in the analyses.

To our knowledge, this study is first to analyse which factors, apart from storage time and the age of the blood sample donor, affect the DNA yield from whole blood samples. Unfortunately, we were able to study only men in the FINRISK 97 sub-sample and data on inflammatory markers was only available in this subset of individuals. Apart from these drawbacks, our study, which analysed DNA samples extracted in a highly experienced central laboratory, utilized two independent cohorts and was able to demonstrate the same trend for the results in both cohorts, although two different extraction methods were used.

5.2 Inflammatory candidate gene selenoprotein S and the risk for cardiovascular disease (II)

5.2.1 Variation in the *SEPS1* gene and the risk for coronary heart disease and ischemic stroke

Selenoprotein S is a membrane protein, which protects cells from oxidative stress and apoptosis and participates in the regulation of inflammation. SNP variants *rs28665122*, *rs4965814*, and *rs4965373* have been shown to associate with circulating levels of IL-1β, TNF-α and IL6 and variant allele of *rs28665122* reduces *SEPS1* promoter activity (Ye et al. 2004, Curran et al. 2005, Kim et al. 2007).

In the second part of this study, we aimed to investigate whether the common variation in this new interesting inflammatory candidate gene associated with the risk for CHD, ischemic stroke, a composite end-point CVD and total mortality in two independent case-cohort sample sets. The extent of LD in the gene region is low and of the selected six variants, four tagged three haplotype bins excluding only one bin with a single SNP in a repeat region based on the publicly available HapMap and Perlegen genotypes. The other two variants have not been genotyped in either project (http://gvs.gs.washington.edu/GVS/index.jsp). Figure 10 presents the selected variants and the haplotype structure within the *SEPS1* locus.

We found two significant associations in the time-to-event analysis. After controlling for the traditional risk factors, a promoter variant *rs8025174* (haplotype H2) was associated with increased risk for CHD in women of the FINRISK 92 cohort (hazard ratio (HR) 6.27 (95% confidence interval: 1.36–28.85, dominant model)), FINRISK 97 cohort (HR: 3.56 (1.06–11.99)), and in combined analysis of both cohorts (HR: 2.95 (1.37–6.39)). SNPSpD estimated four independent loci among the genotyped variants which corresponds to a significance threshold of $p = 0.0125$. The result from the combined analysis ($p = 0.006$) reached this level of significance. There was no association among men or for any other end-point in either sex for this variant.

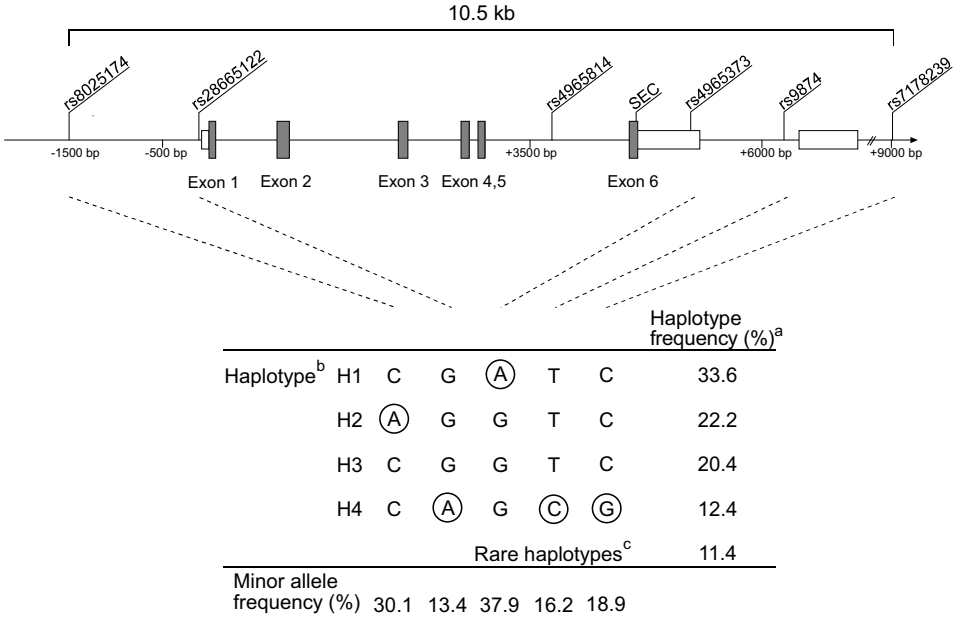

Figure 10. ***Genotyped variants in the SEPS1 gene.*** *a) Haplotype and minor allele frequencies are estimated in the combined subcohort of both cohorts, b) The variant allele is encircled, c) Haplotypes with frequency <5%. Modified from Alanne M.* et al. *"Variation in the selenoprotein S gene locus is associated with coronary heart disease and ischemic stroke in two independent Finnish cohorts" Human Genetics 2007:122;355-65 © Springer-Verlag 2007. Printed with kind permission of Springer Science and Business Media.*

Variant *rs7178239* (in haplotype H4) was associated with increased risk for ischemic stroke in women. Since the number of incident stroke cases in each cohort was low (FINRISK 92: 33, FINRISK 97: 21), we only considered the analysis in the combined female sample to be meaningful. In this group, the variant allele of SNP *rs7178239* was significantly associated with 3.35 times higher risk for ischemic stroke (CI: 1.66–6.76, dominant model) when adjusting for traditional risk factors and applying multiple testing significance threshold. A similar and significant result was observed also for the unadjusted model and for joint analysis of both sexes. The variant was not associated with ischemic stroke in men or with CHD, CVD or total mortality in either sex. Haplotype analysis supported findings from the individual SNP analyses for both variants.

5.2.2 Variation in the *SEPS1* gene and quantitative phenotypes

Variation in the *SEPS1* gene region has previously been associated with inflammatory cytokines. (Curran et al. 2005) In our study, we investigated the association between the selected gene variants and inflammatory markers IL6 and TNF-α together with CVD trait phenotypes BMI and WHR. Adipose tissue is one of the tissues which secrete inflammatory cytokines and obesity can be considered to be a low-grade inflammatory condition. (Ferrante 2007) The only association we observed for the inflammatory markers was for variant *rs7178239* (H4) and levels of IL6. Subcohort women free of disease, who carried the variant allele, had higher levels of IL6 than non-carriers (1.65 vs. 1.24 ng/l, p = 0.008). The result was not significant in men and among CVD cases and we could not observe any consistent trend.

Several of the variants associated with BMI and WHR in the FINRISK 92 cohort but not in the FINRISK 97 cohort. The association with increased BMI was mostly evident among incident CVD cases for the variant allele of *rs8025174* (Figure 11). We observed the same effect of the variant allele for WHR. Furthermore, variants in haplotype H4, including *rs7178239*, associated with increased WHR among women in the subcohort (Figure 12). Time-to-event analysis including BMI or WHR resulted in only a small attenuation in the hazard ratio estimates.

5.2.3 Allele specific binding of nuclear extract according to *SEPS1* variant *rs7178239*

Bioinformatic analysis of the *rs7178239* locus

The allele-specific binding of transcription factors was investigated using two TF-binding site prediction programs, Alibaba2 and P-match. Variant *rs8025174* resides next to a repeat region and we focused on variant *rs7178239C>G*. Since Curran and co-workers (2005) had in their study, which included re-sequencing of the gene, concluded that the association signal was mostly due to the effect of *rs28665122,* other variants than *rs7178239* in the H4 haplotype were not included in the analysis. Variant *rs7178239* is located outside the sequencing region. The results from both programs differed markedly, and did not predict the same TF-binding sites. At the lowest level of stringency, Alibaba2.1 predicted binding of TF Sp-1 (simian-virus-40-protein-1) when either of the SNP alleles was present. However, with increasing stringency, only interaction with the G-allele remained. P-match predicted binding of TF C-rel with both alleles when using a cut-off to minimize false negative findings, but the core matrix similarity was greater with the G-allele (0.85 vs. 0.95).

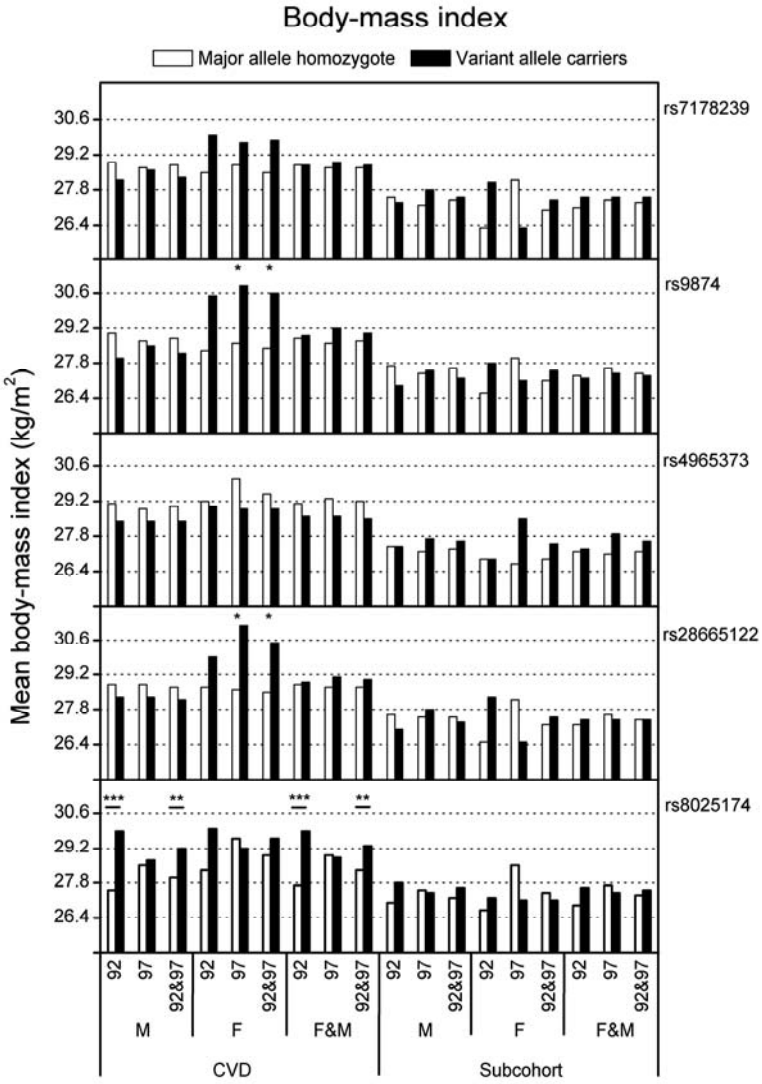

Figure 11. ***Association of allelic variants in SEPS1 gene and body-mass index separately in both sexes and in cardiovascular cases and subcohort members free of disease.*** *Analysis was adjusted for age, sex and cohort when relevant. 92, 97 and 92&97 indicate each FINRISK cohort and their combination. Significance level: P<0.05*, P<0.01**, P<0.001***. Results significant at the 5% level after SNPSpD correction are underlined.*

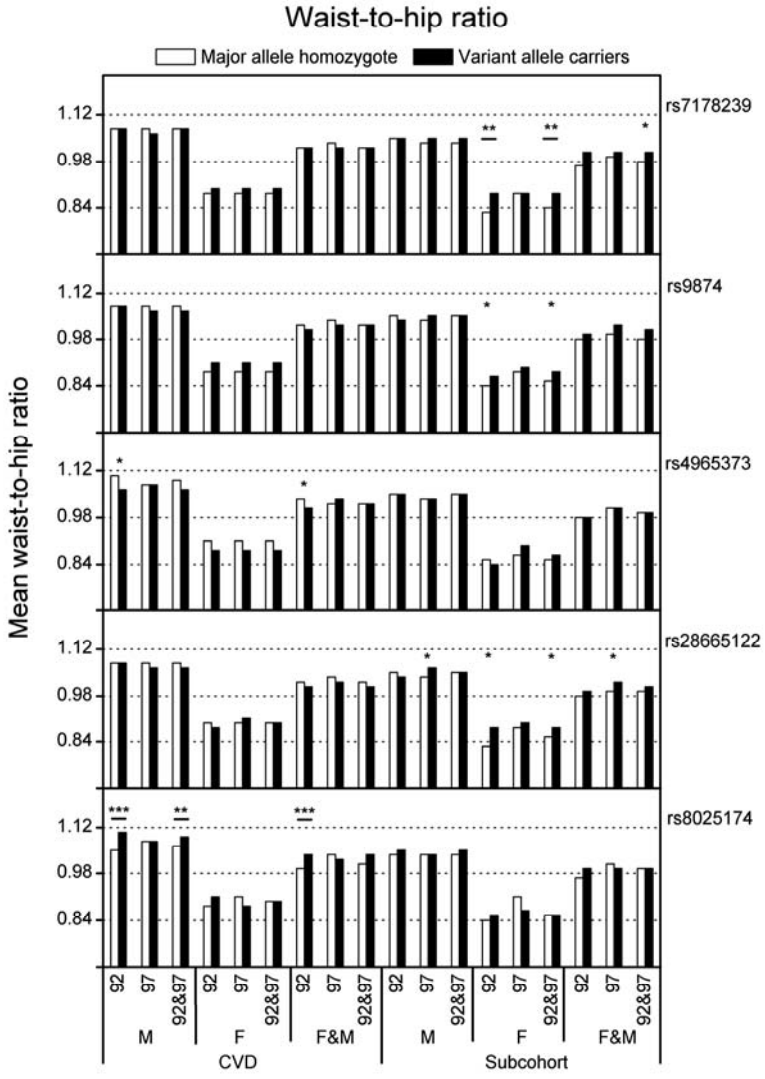

Figure 12. ***Association of allelic variants in SEPS1 gene and waist/hip ratio separately in both sexes and in cardiovascular cases and subcohort members free of disease.*** *Analysis was adjusted for age, sex and cohort when relevant. 92, 97 and 92&97 indicate each FINRISK cohort and their combination. Significance level: P<0.05*, P<0.01**, P<0.001***. Results significant at the 5% level after SNPSpD correction are underlined.*

Electrophoretic mobility shift assay

To further elucidate the function of the associated variant, we investigated whether the sequences exhibit allele-specific DNA-nuclear protein interactions from hepatocyte nuclear extract using EMSA. The results indicated specific and non-specific interactions between the nuclear extract and the 36mer oligo. With the C-allele we observed a DNA-protein band that was absent with the G-allele probe. (Figure 13, panel 1, band B) Furthermore, the binding of shared bands A, C and D appeared to be stronger with the G-allele. Unlabelled competitor analysis with the respective allele indicated that band B results from a specific DNA-protein interaction since the C-probe competed band B already in a small molar excess. Competition of band A was weaker than in bands C and D in 0.5 – 5-fold molar excess of the G-probe. (Figure 13, panel 2) Surprisingly, although the G-allele did not produce band B, competition of the C-probe with the G-allele removed the band. (Figure 13, panel 3) Competition experiment made *vice versa* did not affect binding of band A, but reduced the binding of bands C and D. Non-specific competition with the probe containing a NF-κB binding site and the probe mutated for this binding, indicated that the C-allele band B was specific to the sequence and did not result from binding of NF-κB, a common transcription factor, which has been shown to bind to the *SEPS1* promoter (Gao et al. 2006) (Figure 13, panel 4).

5.2.4 Conclusions and discussion of Study II

Taken together, the results from the case-cohort analysis indicate that two genetic variants in the *SEPS1* gene locus confer risk for CHD (*rs8025174)* and ischemic stroke (*rs7178239)* in women. In men, there was no indication of association for either of the variants suggesting female-specific genetic risk. To our knowledge, neither of these variants have previously been included in association studies for CVD. Interestingly, *SEPS1*: *rs7178239* is located in the Finnish population in the same haplotype (H4) as the variant *rs28665122*, which has previously been associated with increased levels of inflammatory cytokines and transcriptional regulation of the gene. In this study, the variant had a borderline association with ischemic stroke (HR: 2.21 (1.04-4.72)). This promoter variant was not associated with ischemic stroke in a case-control study conducted in Germany and Italy (Hyrenbach et al. 2007). In the German sample, the variant allele was a little more common among cases ($p = 0.17$) whereas in Italy there was no difference. In addition to population and case ascertainment differences in our, German and Italian samples, the Italian controls were selected from the staff of the neurological hospital which may have caused some unobserved bias.

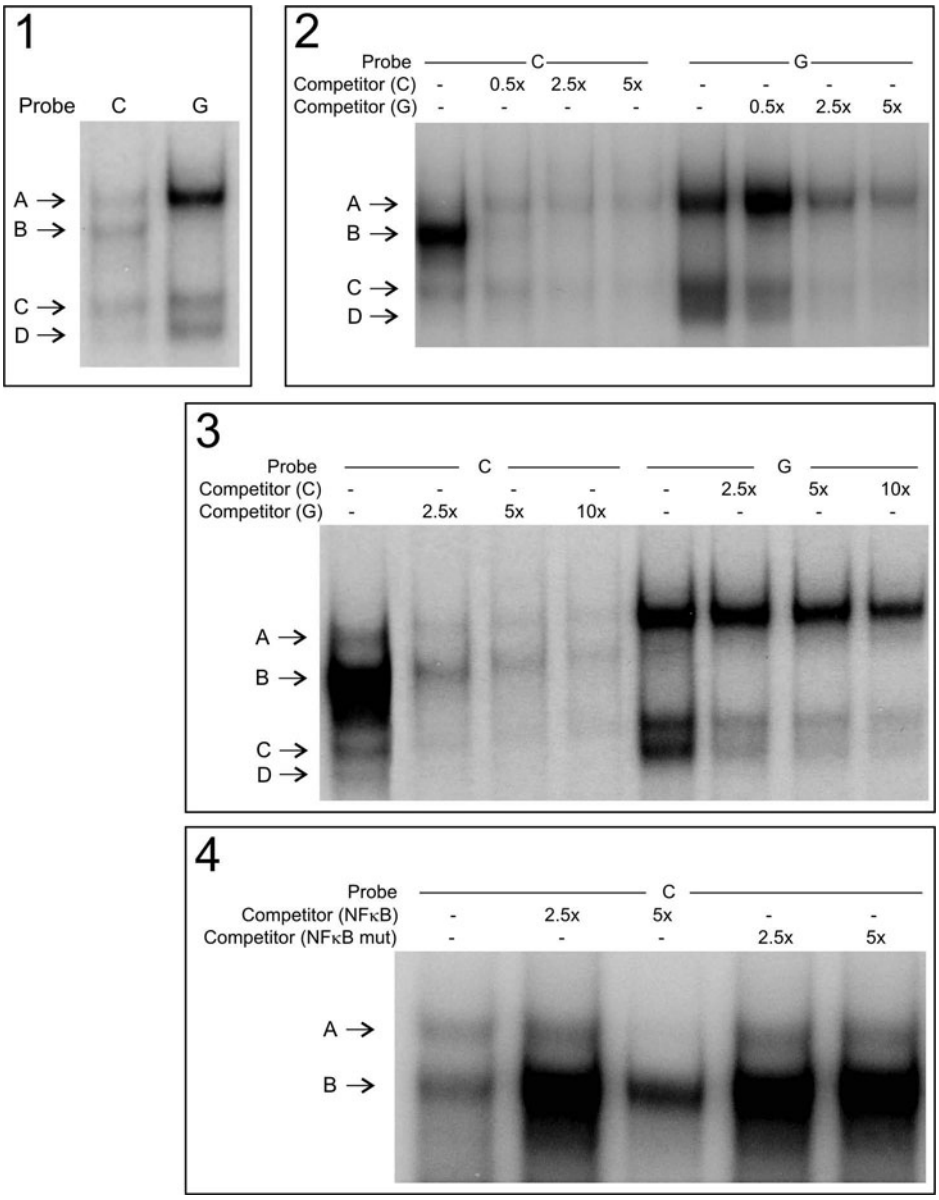

Figure 13. ***EMSA analysis of SEPS1: rs7178239 C>G.*** *Arrows A, B, C and D indicate different sized nuclear protein-probe complexes from hepatocyte nuclear extract. 1) Allele C and G -specific binding patterns, 2) Competitor experiment with unlabelled respective probe of indicated molar excess, 3) Cross competitor analysis with allele-specific probes, 4) Competitor analysis for allele C with unspecific competitor NFκB and mutated NFκB (NFκB mut).*

The female-specificity of our results may, in part, be linked with selenium metabolism. Recent studies have demonstrated that in mice, expression and activity of other selenoproteins is affected by selenium concentration and tissue-dependent post-transcriptional modification and that both of these mechanisms are sexually dimorphic (Riese et al. 2006), and that the uptake and utilization of selenium may be sexually dimorphic also in humans (Méplan et al. 2007).

Genetic variation in the *SEPS1* gene locus, *rs28665122* in particular, has previously been associated with circulating levels of cytokines in humans. In our study, *rs28665122* and other variants in the gene did not associate with circulating levels of inflammatory markers IL6 and TNF-α. A major difference between this study and the previous study reporting the association, is that the study by Curran and colleagues (2005) was ascertained through an obese proband. Although the association was independent of BMI, this difference may have caused dissimilarity in the results. Furthermore, in our study, variation in the *SEPS1* gene locus was associated with BMI and WHR in the FINRISK 92 cohort, but the association with the risk for CHD or stroke was not mediated through them.

The EMSA analysis of variant *rs7178239* supported the hypothesis that this gene region binds regulatory proteins from nuclear extract and that the binding is allele-specific. The results suggests that the major allele C of *rs7178239* binds a different protein from the nuclear extract than the minor allele G, but the interaction may involve the same protein complex, since G-allele could remove this interaction. In addition, band A may result from stronger sequence specific binding of the G-allele than C-allele, but this result needs further clarification. Bioinformatic analysis did not reveal any clear candidates as transcription factors with allele-specific binding.

The CVD associated variants in this study are not located in the coding region or the UTR regions of the gene. It is thus likely that the mechanism by which they confer higher risk for CVD is related to the transcriptional regulation of the *SEPS1* gene or that they are in high LD with such variants. Particularly, the variant *rs7178239* is intriguing. It shows evidence of allele-specific binding of transcription factors or other interacting proteins and it may be located in an enhancer element. On the other hand, it resides in same haplotype as a promoter variant *rs2866512* and the association result may represent the effect of the promoter variant which has previous functional evidence (Curran et al. 2005). The EMSA analysis is susceptible to false positive findings, especially when there is no prior knowledge of the interacting transcription factor and supershift assays are not be performed. These results from the variant *rs7178239* thus need to be interpreted with caution.

Functional studies on the selenoprotein S suggest that it may protect cells from apoptosis. (Kim et al. 2007) Macrophage, smooth muscle cell and endothelial cell apoptosis are important contributors to atherosclerosis development and

understanding of regulation of apoptotic process may lead to greater understanding of the pathogenesis of atherosclerosis (Choy et al. 2001, Akishima et al. 2004). Another way in which SEPS1 could contribute to the development of CVD is regulation of inflammatory processes in endothelial cells as part of the unfolded protein response (UPR). A recent study has demonstrated that ER stress induced UPR is involved in regulation of the inflammatory response in endothelial cells suggesting a key role in regulation of vascular inflammation and endothelial dysfunction (Gargalovic et al. 2006).

In conclusion, SEPS1 is an interesting new candidate for CVD susceptibility. The role of genetic variation in the *SEPS1* gene locus requires further experimental validation and replication in other population-based cohorts.

5.3 Interleukin 6 -related candidate genes and sex-specific differences in the risk for cardiovascular disease (III)

5.3.1 Association of gene variants in the interleukin 6 -related pathway with cardiovascular disease and -risk factors

In study III, we investigated the association between 239 variants in 46 candidate genes and the risk for CHD, ischemic stroke and a composite end-point cardiovascular disease. This thesis focuses on the results from the interleukin 6 -related genes *ACE*, *AGTR1*, *IL6*, *CRP*, *FGA*, *FGB* and *FGG*.

Interleukin 6 is a pleiotropic cytokine that regulates inflammatory reactions. It induces synthesis of CRP, fibrinogens and *AGTR1*, and is involved in a complex regulatory relationship with Ang II. (Heinrich et al. 1990, Wassmann et al. 2004, Coles et al. 2007) Previous genetic association studies have focused mainly only on a few polymorphisms in each gene and we wanted to investigate what is the role of this pathway in the risk for CVD in a population based sample by comprehensively studying the common variation in these genes and their joint effects.

In single SNP time-to-event analysis, we observed several nominally significant associations at the $p < 0.05$ level (uncorrected) and two variants in *AGTR1* and *FGA* genes with p-value < 0.01 (Table 13). In addition, the results indicated that the variant allele carriers of SNP *IL6*: *rs2069840* had an increased risk for CVD. The association was only evident in males (HR: 1.54 (1.12-2.13), $p = 0.006$ for combined analysis of both cohorts) and we identified a significant genotype sex interaction for CVD ($p = 0.004$, uncorrected). None of the other variants in IL6 -related genes associated consistently at the $p < 0.05$ level with the CVD risk in both sexes or showed evidence of genotype-sex interaction.

Based on the FDR analysis, we expect 47% of the results with uncorrected $p \leq 0.01$ in the combined analysis of both cohorts and both sexes including single SNP time-to-event and quantitative trait analysis of all SNPs included in study III (n=172) except fibrinogen and IL6 levels, to present true positive findings.

Table 13. ***Association of variants in the interleukin 6 -related genes (p < 0.05, uncorrected) with coronary heart disease and cardiovascular disease in joint analysis of both sexes and cohorts.***

Phenotype	Gene	Variant rs#	Model[a]	Hazard ratio (95% CI)	p-value
Coronary heart disease	ACE	rs4293 G>A	GG+AG vs. AA	1.42 (1.02-1.97)	0.0367
	AGTR1	rs388915 A>G	AA vs. AG+GG	0.45 (0.25-0.81)	0.0084
	FGA	rs2070018 A>G	AA vs. AG+GG	1.55 (1.11-2.17)	0.0101
	FGA	rs2070018 A>G	AA>AG>GG	1.48 (1.09-2.03)	0.0133
	IL6	rs1554606 T>G	GG vs. GT+TT	1.43 (1.02-2.00)	0.0369
	IL6	rs1800795 C>G	CC+CG vs. GG	0.71 (0.52-0.99)	0.0424
Cardiovascular disease	FGA	rs2070018 A>G	AA vs. AG+GG	1.49 (1.11-2.02)	0.0091
	FGA	rs2070018 A>G	AA>AG>GG	1.45 (1.09-1.91)	0.0097
	FGA	rs6050 T>C	CC vs. CT+TT	1.60 (1.05-2.42)	0.0279
	FGB	rs1800788 C>T	CC>CT>TT	0.81 (0.66-0.99)	0.0376
	FGB	rs2227439 C>T	CC vs. CT+TT	1.35 (1.01-1.80)	0.046
	FGG	rs2066861 C>T	CC+CT vs. TT	0.62 (0.40-0.94)	0.0247
	IL6	rs2069827 G>T	GG vs. GT+TT	1.30 (1.00-1.68)	0.0495

[a] Time-to-event analysis, adjusted with geographic region, cohort, HDL-C, non-HDL-C, body mass index, mean blood pressure, smoking status, history of diabetes

5.3.2 Joint analysis of interleukin 6 -related genes

Since the IL6 -related genes form a biologically meaningful pathway, we analysed the joint effects of all 30 SNPs in these genes and the risk for incident CHD and CVD using Cox's proportional hazard model. We did not include analysis of ischemic stroke due to the small number of individuals. The analysis revealed one SNP pair *FGA*: *rs2070018* and *IL6*: *rs2069827* in which the carriers of variant alleles from both SNPs had significantly lower risk for CHD and CVD than individuals who carried variant allele from only either of the SNPs. The analysis included 30 CHD and 43 CVD cases with variant alleles from both SNPs, and 182 CHD and 242 CVD cases with variant allele from either of the SNPs and 174 CHD and 219 CVD cases with no variant alleles.

In the combined analysis of both cohorts and sexes the risk for CHD was 61.5% lower (uncorrected p-value 0.0002) and 58.7% lower (p = 0.0001) for CVD for individuals with variant alleles from both SNPs. The difference was similar in both cohorts in combined analysis of both sexes (Figure 14). The risk for individuals carrying variant alleles from both SNPs compared to those who did not carry variant alleles from either of the SNPs was 0.37 (0.22-0.61) for CHD and 0.41 (0.26-0.64) for CVD when both sexes and cohorts were analyzed together. The risk was not different between those who carried the variant allele from either of the SNPs and

those who carried none. The results of joint analysis of both sexes and cohorts for CHD and CVD, and combined analysis of both male cohorts for CVD had a FDR value less than 10%. A clear trend of association in the two variants *FGA*: *rs2070018* and *IL6*: *rs2069827* in sex- and cohort stratified analysis gives further support for the joint effect of both variants. The variant alleles of both SNPs showed evidence of a protective effect on the risk for CHD, ischemic stroke and CVD (Figure 15).

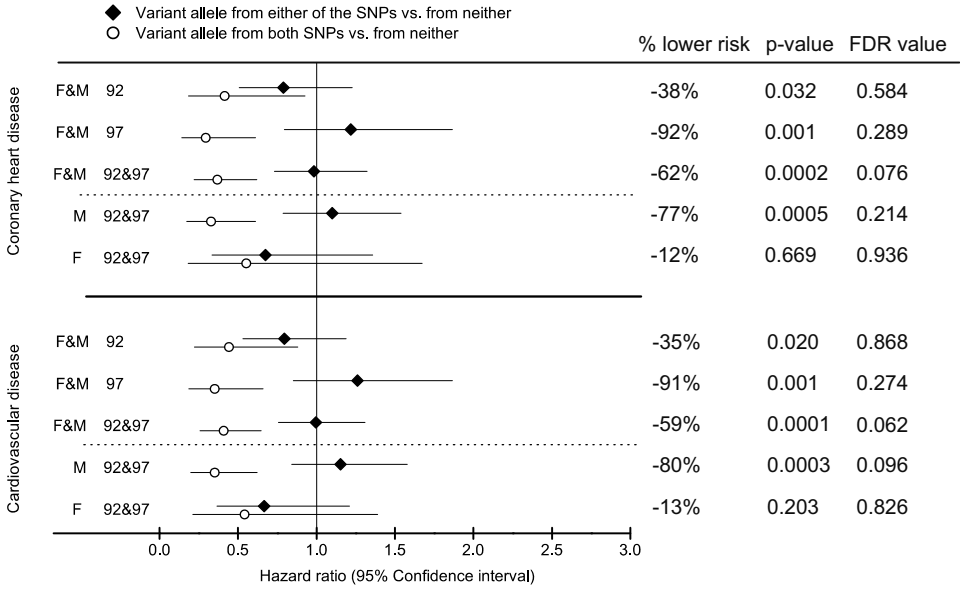

Figure 14. ***Association of joint effects of SNPs FGA: rs2070018 and IL6: rs2069827 with coronary heart disease and cardiovascular disease.*** *Results with FDR value ≤ 10% are expected to include 90% true positive findings.*

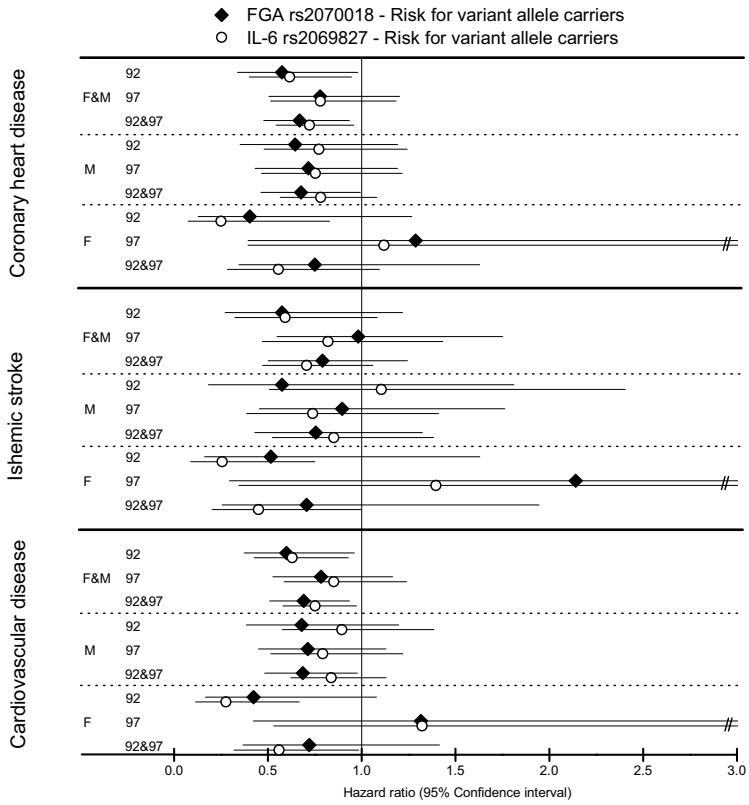

Figure 15. ***Variant allele associated risk of SNPs FGA: rs2070018 and IL6: rs2069827 for coronary heart disease, ischemic stroke and composite end-point cardiovascular disease separately in the two cohorts and sexes.***

5.3.3 Association with CRP, IL6 and fibrinogen levels

In addition to the risk for CVD, we investigated whether genetic variation in genes *CRP*, *FGA*, *FGB*, *FGG* and *IL6* associated with the levels of the respective gene products in subcohort individuals free of CVD at the baseline. Furthermore, we investigated the association between *IL6* gene variants and the levels of fibrinogen and IL6, and between all genes in study III and CRP levels. A rare variant allele of synonymous SNP *rs1800947* (frequency = 0.06) in the second exon of the *CRP* gene associated strongly with decreased CRP concentration in combined analysis males

from both cohorts (p = 0.0001). The trend of the association was similar in women (Figure 16).

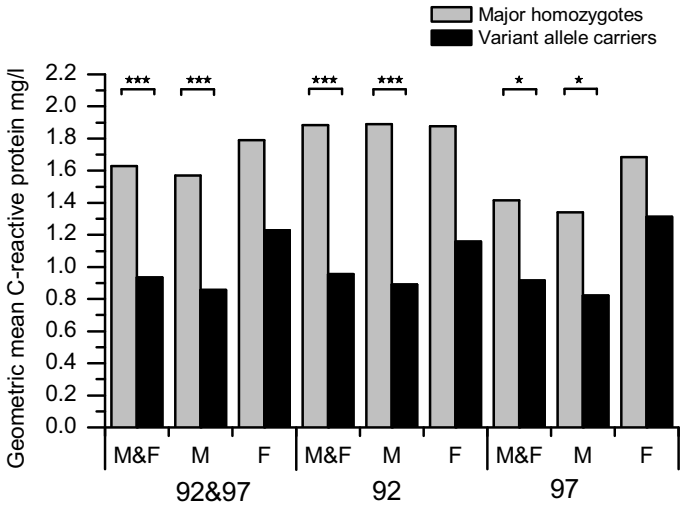

Figure 16. ***Association of CRP gene variant rs1800947 with high sensitivity CRP levels in subcohort individuals free of baseline CVD.*** *Significance level: p<0.05*, p<0.01**, p<0.001*** (uncorrected), for the difference between the two groups.*

The results from linear regression models for the fibrinogen levels, revealed a variant *rs2070018* in the *FGA* gene, which showed evidence of association, the variant allele carriers having lower levels. The association was strongest in the joint analysis of both sexes in the FINRISK 92 cohort (p = 0.004, uncorrected) and the trend was similar in other subgroups analysed (Figure 18). The fibrinogen concentration was not available from women in the FINRISK 97 cohort. We found a trend of an association also for additive effect of variant *rs1800792* in the *FGG* gene in the FINRISK 92 cohort. Each variant allele increased the fibrinogen concentration in the joint analysis of men and women by 0.2 mg/l (p = 0.009, uncorrected). The trend was similar in both men and women but evident only in the FINRISK 92 cohort (Figure 17). The LD between these two variants in the fibrinogen locus is low ($r^2 = 0.08$) suggesting that these associations did not stem from high correlation between the variants. Interleukin 6 gene variants did not associate with CRP, fibrinogen or IL6 levels at p < 0.01.

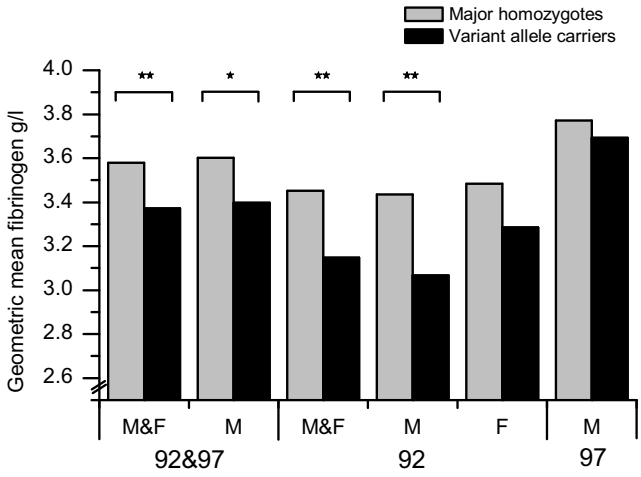

Figure 18. ***Association of variant FGA: rs2070018 with the fibrinogen levels in subcohort individuals free of baseline CVD.*** *Significance level: p<0.05*, p<0.01**, p<0.001*** (uncorrected), for the difference between the two groups.*

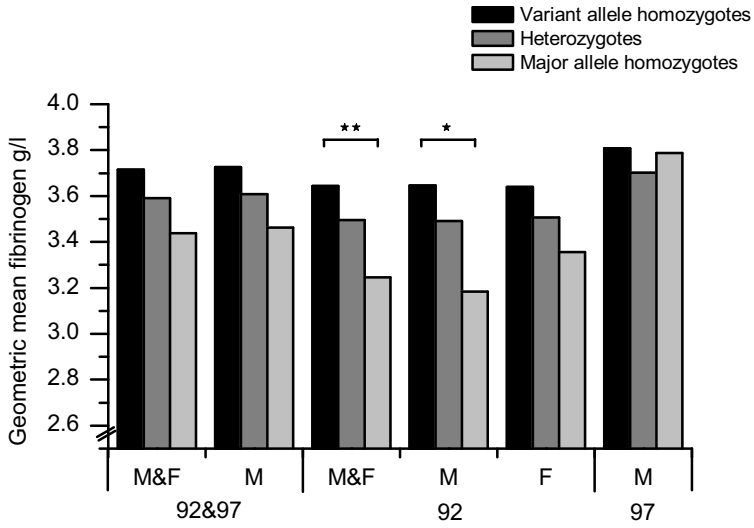

Figure 17. ***Association of variant FGG: rs1800792 with the fibrinogen levels in subcohort individuals free of baseline CVD.*** *Significance level: p<0.05*, p<0.01**, p<0.001*** (uncorrected), for trend.*

5.3.4 Sex-specific association with CVD risk factors

Several of the IL6 pathway genes were associated with CVD risk factors differentially in men and women. Particularly, a variant allele of SNP *rs1800792* in the promoter region of the *FGG* gene associated only in women in an additive fashion with decreased levels of HDL-C and higher BMI (uncorrected p-value for interaction ≤ 0.01) (Table 14). Another variant in the fibrinogen gene cluster *FGA*: *rs2070006* associated also with HDL-C only in women with an unadjusted interaction p-value of 0.0001. Table 14 presents suggestive sex-genotype interaction results only for IL6 -related genes. The interaction test was performed for all 172 SNPs in study III in three different genetic models for CHD and CVD end-points and quantitative traits: mean blood pressure, non-HDL-C, HDL-C, triglycerides, total cholesterol, CRP, BMI and WHR. For sex-genotype interaction results with $p \leq 0.01$ the estimated proportion of true positive findings is 30% but with the additional criterion of nominal association at the $p \leq 0.01$ level in time-to-event analyses or at the $p < 0.01$ level in quantitative analysis in either sex in joint analysis of both cohorts, the expected proportion of true positive findings is likely to be greater.

Table 14. ***Genotype-sex interaction in interleukin 6 -related genes with quantitative CVD risk factors.*** *Results from linear regression analysis, adjusted for geographic region, cohort and age showing evidence of association in either sex (uncorrected p ≤ 0.01) and genotype-sex interaction (uncorrected p ≤ 0.01).*

		Mean value for minor homozygotes / heterozygotes / major homozygotes				
Gene: variant	Model	Women	p-value[b]	Men	p-value[b]	p-value interaction[a/b]
Body mass index (kg/m^2)						
ACE: rs4320 G>A	additive	29.0 / 27.8 / 26.5	0.003	27.5 / 27.6 / 28.1	0.2	0.0009
FGG: rs1800792 T>C	additive	30.8 / 28.2 / 26.7	0.0005	27.9 / 27.6 / 27.7	0.9	0.002
HDL cholesterol (mmol/l)						
CRP: rs1205 C>T	recessive	1.77 / 1.46 / 1.52	0.0008	1.28 / 1.23 / 1.24	0.3	0.02
FGA: rs2070006 C>T	additive	1.73 / 1.53 / 1.38	0.0002	1.24 / 1.23 / 1.29	0.2	0.0001
FGG: rs1800792 T>C	additive	1.38 / 1.46 / 1.62	0.004	1.20 / 1.28 / 1.24	1.0	0.01

[a] Interaction p-value tests the null hypothesis that the genotype effects in regression analysis in men and women do not differ from each other.
[b] Uncorrected p-values.

5.3.5 Sex-specific association of other candidate genes

In study III, we observed several sex-specific interactions between genetic variants and the risk for CHD, ischemic stroke or CVD. First, we confirmed a previously reported sex-specific association with increased risk for CHD and CVD only in women for variant *USF1*: *rs2774279* (Komulainen et al. 2006). Furthermore, three variants in the *SEPS1* gene haplotype H4 showed evidence of association with increased risk for ischemic stroke only in women as described in study II, and we observed evidence for sex-genotype interaction for them (uncorrected p-values for interaction 0.02-0.007). Other genes with evidence of sex-specific effect at $p \leq 0.01$ and evidence of association at $p \leq 0.01$ (uncorrected) were for coronary heart disease: *CPB2* and *F13A1* in women, and for cardiovascular disease: *CPB2*, and *LPIN1* in women. Furthermore, we observed several associations for quantitative risk factors only in either sex with evidence of association at $p \leq 0.01$ and genotype-sex interaction at $p \leq 0.01$ (uncorrected). In male subcohort individuals without CVD at the baseline, a *LPIN1* gene variant showed evidence for association with total cholesterol, a *APOBEC2* gene variant with LDL-C, a *F5* gene variant with triglycerides and a *SERPINE1* gene variant with mean blood pressure. In women we observed evidence of association for *F5*, *LPIN*, *THBD* and *USF1* gene variants with CRP, for *F13A1* and *USF1* gene variants with BMI, and for three *ICAM1* gene variants with WHR.

5.3.6 Power analysis of the case-cohort set

Understanding the power of the study is critical to the correct interpretation of the analysis results from an association study. We estimated the power of the case-cohort set separately for men and women by combining the two FINRISK cohorts but not adjusting for number of variants. Together the simulation model for incident CVD included 881 men (409 cases) and 372 women (177 cases). Simulation of covariates (the same as used in regression modelling) in 5000 independent data sets with six different combinations of allele frequencies and effect sizes revealed that the power of the study is quite low for each sex separately while in joint analysis we had approximately 70% power to detect a hazard ratio of 1.3 (Figure 19) (Kulathinal et al. 2007). With this power, we estimated posterior odds for a true finding amongst the 27 independent genes in study III of 6:1, if we expect to observe two true associations (The Wellcome Trust Case Control Consortium 2007).

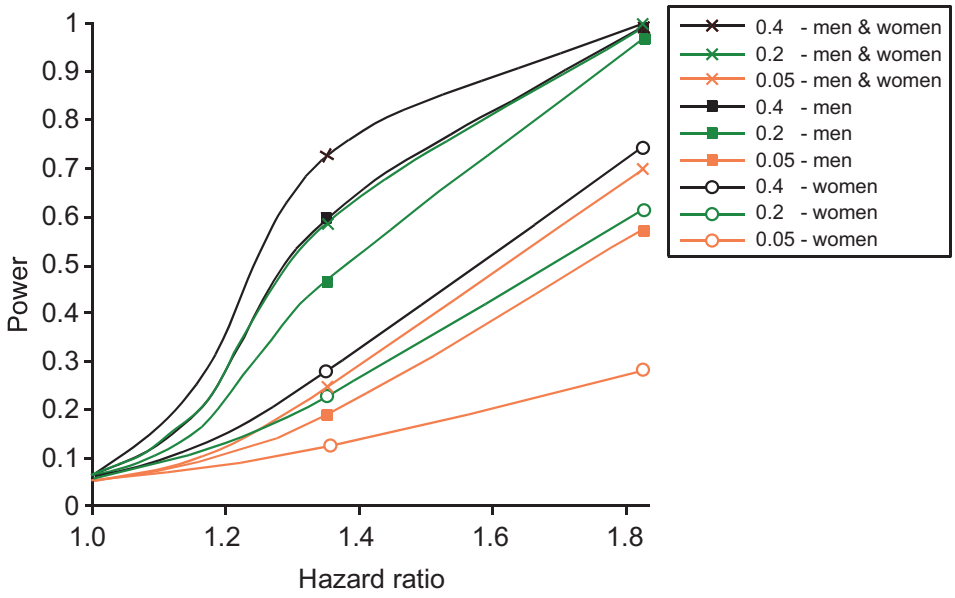

Figure 19. ***Estimated power of the FINRISK case-cohort set.*** *Power was estimated for men, women and joint analysis at three different allele frequencies 0.05, 0.2 and 0.4, and four different hazard ratios 1, 1.16, 1.35 and 1.82 in 5000 simulations.*

5.3.7 Conclusions and discussion of Study III

In a summary, we found evidence of an association between genetic variation in the IL6 pathway and the risk for CVD and CVD risk factors. In the *IL6* gene, variant *rs2069840* associated with an increased risk for CVD in men while another (*rs2069827*) showed a trend of an association for a decreased risk. These two variants are not in tight LD with each other ($r^2 = 0.08$). Furthermore, the second variant was jointly associated with decreased risk of CHD and CVD with a variant *rs2070018* in the *FGA* gene downstream in the IL6 pathway. Unfortunately, the number of individuals homozygote for both variants was too small to allow more detailed analysis of a possible interaction model behind this association. Haplotypes in the IL6 gene promoter area have previously been associated with the transcriptional regulation of the gene but have not included the *IL6*: *rs2069827* variant, which is located ~1.3kb upstream of the gene. The variant is the sole representative of the haplotype bin it resides in according to the SeattleSNPs database. It is thus likely that the association with CVD may involve transcriptional regulation of *IL6* and that this variant adds more to the previously studied haplotype diversity in the regulatory region of the gene.

The *FGA*: *rs2070018* variant is located 37 bp downstream of the fourth exon of the *FGA* gene and can putatively alter the splicing of the fourth intron. Alternatively, it may present signal from a variant in a repeat region in the promoter area of the *FGA* gene (*rs2070008*) which tags the same haplotype bin or signal from variant *rs2227439* in the *FGB* gene, which is also located in a repeat region (r2=0.78). The risk-lowering effect of *FGA*: *rs2070018* may be mediated through its association with lower levels of fibrinogen. The joint effect of these two variants on the risk for CVD may be mediated through altered availability and constitution of IL6 and fibrinogen alpha transcripts, a new interesting pair of molecules that have not together been previously associated with CVD.

Genetic variants in the interleukin 6 gene

Polymorphisms in the *IL6* gene have been inconsistently associated with MI, CHD and the acute coronary syndrome (Rauramaa et al. 2000, Georges et al. 2001, Humphries et al. 2001, Jenny et al. 2002, Sie et al. 2006, Barbaux et al. 2007). Recently, Mälarstig and co-workers (2007) analysed some of the same variants as we did in our study (-572G>C (*rs1800796*), *rs2069827* and two proxies for -174G>C (*rs1800795*)) also selecting SNPs representing haplotype bins in the SeattleSNPs database. In their study of 3027 non-ST elevated ACS patients, who received dalteparin treatment or placebo, and 447 controls they concluded that variation in the *IL6* gene does not associate with ACS.

The *IL6* promoter region haplotype carrying the variant allele of SNP -174G>C has previously been associated with decreased *ex vivo* production of IL6 and decreased affinity for nuclear proteins (Rivera-Chavez et al. 2003). The findings have been inconsistent in regard to -174G>C and circulating levels of IL6 (Fishman et al. 1998, Sie et al. 2006). In this study, we were able to analyze the effect of genetic variants on serum IL6 concentration in 309 population-based controls free of baseline CVD from the FINRISK 92 cohort. None of the five variants in our study associated consistently and significantly with serum IL6 levels, but carriers of the C allele of SNP -174G>C tended to have lower levels. Although the participants were all studied in the semi-fasting state, the diurnal and intra-individual variation in the levels of IL6 and differences in the time before the last meal may have confounded these analyses (Dugue and Leppänen 1998, Undar et al. 1999).

Genetic variants in the fibrinogen genes

Studies on SNPs and haplotypes in the fibrinogen gene cluster and fibrinogen concentration have been conflicting and both negative and positive findings have been presented (Thomas et al. 1995, Mannila et al. 2004, Mannila et al. 2005, Kathiresan et al. 2006b, Keavney et al. 2006, Mannila et al. 2006). Similarly, there has been several positive (Boekholdt et al. 2001, Mannila et al. 2005, Mannila et al. 2006, Mannila et

al. 2007) and negative (Fox et al. 2004, Mannila et al. 2004, Kathiresan et al. 2006b, Keavney et al. 2006, Uitte de Willige et al. 2006, Zee et al. 2006) association reports on genetic variants in the cluster and the risk of MI, CHD or increased carotid intima media thickness. In this study, we observed an association between fibrinogen concentration and two variants *FGA: rs2070018* and *FGG: rs1800792*. Interestingly, two variants from the fibrinogen gene cluster *FGA: rs2070006* and *FGG: rs1800792* associated in women with levels of HDL-C, and the *FGG* variant also with BMI. Furthermore, the same variant associated at the $p < 0.05$ level for interaction with triglycerides and WHR. There is no apparent rationale of how the fibrinogen gene polymorphisms would regulate lipid metabolism and the result may represent statistical fluctuation. It is known, however that hypercoaguability and hypofibrinolysis coincide with lipid disorders and that obesity promotes prothrombotic and hypofibrinolytic state and dyslipidemia (Darvall et al. 2007). Apart from the variant *FGA: rs2070018*, variants in the *FGA*, *FGB* and *FGG* genes did not associate with the risk of CHD, ischemic stroke or CVD.

Genetic variants in the CRP gene

Previous studies on the *CRP* gene locus and the association with CVD have been conflicting. (Miller et al. 2005, Kardys et al. 2006, Lange et al. 2006) In this study (III), we did not observe any evidence of association for any of the six *CRP* variants studied.

Several variants in the *CRP* gene have been associated with the baseline or stimulated serum CRP levels (Miller et al. 2005, Kathiresan et al. 2006a). The association reports are most consistent for a tri-allelic SNP -286C>T>A (*rs3091244*) in the promoter region of the *CRP* gene. Both variant alleles T and A are associated to higher CRP levels in Caucasian populations. The SNP disrupts an E-box binding site for basic helix-loop-helix transcription factors such as USF1, USF2, Myc and Max and the T-allele has been shown to affect binding of nuclear proteins (Szalai et al. 2005). In this study, we investigated the association of the tri-allelic SNP by using two variants *rs1130864* and *rs3093075* as surrogates for the alleles A and T, respectively. We did not find a significant association at $p \leq 0.01$, although the variant allele carriers of *rs3093075* (corresponding to the A-allele of -286C>T>A*)* tended to have higher levels of CRP. In concordance with earlier studies (Carlson et al. 2005, Lange et al. 2006), we observed an association with variant allele of SNP +1059G>C (*rs1800947*) and decreased CRP levels. In our study including 172 variants, this association was one of two results with a FDR value <10%.

Genetic variants in the RAS system

Several studies have reported genetic analyses of the renin-angiotensin-system focusing mainly on three polymorphisms *ACE:* I/D, angiotensinogen M235T and *AGTR1*: A1166C and their interaction. The Beijing atherosclerosis study

investigated more extensively the sequence variation in the *AGTR1* gene and found three SNPs (*rs275650*, *rs2276736* and *rs5182*), but not A1166C, all of which associated with myocardial infarction (Su et al. 2004). Recently the variant allele of A1166C has been shown to produce less mRNA than the more common allele and that the 3' region of the gene, where this variant is located, associates with metabolic syndrome traits (Abdollahi et al. 2007). In this study (III), the *ACE*: I/D polymorphism did not associate with CHD, ischemic stroke or CVD, neither did it have synergistic effects with variants in the *AGTR1* gene. In the *AGTR1* gene, we observed an association for increased risk for CHD with an intronic SNP *rs388915* in a joint analysis of men and women. Although the results were not significant in sex-stratified analyses, the trend of the effect was similar in both sexes. *AGTR1:* A1166C and other variants in the *AGTR1* gene were not associated with these end-points and the variant *rs388915* was not in significant correlation with any of the other variants.

Sex-specific association results

This study also observed evidence of genotype-sex interaction both for the risk for CVD and CVD risk factors. Generally, we identified more associations in women than in men. We confirmed that these results do not stem from just either cohort by conducting a cohort-genotype interaction test. There are a few possible reasons for the female-specificity of our findings. First, the aetiology of the disease is somewhat different for females and males. The biological basis for this difference comes from: 1) differences in the genetic make-up as represented by the XX and XY sex chromosomes, 2) differences in distribution of CVD risk factors, 3) the effects of sex hormones, 4) anatomical and physiological differences in fat distribution and insulin resistance (Mittendorfer 2005a), lipid metabolism (Mittendorfer 2005b), cardiovascular function (Leinwand 2003) and gene expression (Rinn and Snyder 2005).

The differences in distribution of CVD risk factors and the sexual dimorphism in the risk for CHD has been studied before in the Finnish population. In a study by Jousilahti and co-workers (1999), the differences in CHD risk factors, particularly HDL-C to total cholesterol ratio, explained approximately half of the observed difference in the risk for CHD. For some risk factors, the difference diminished with increasing age, suggesting the difference in the size of the effect should also diminish. It is, however, possible that these risk factors have a larger effect in men because of cumulative effects during longer periods of exposure. Although the prevalence of positive family history of CHD does not differ between sexes in the Finnish population (Jousilahti et al. 1996a), it is plausible that, in addition to unaccounted differences in nutrition, alcohol intake and physical activity, genetic factors contribute to the remaining half of the observed difference in the risk for CHD.

It has been shown for many human diseases, that the gender-dependent differences in the progression and extent of disease are partially explained by sex hormones. The differential effects in gene expression in somatic tissues may lead to gender specific susceptibility to disease (Rinn and Snyder 2005). In women, oestrogen has several protective effects against CHD. It induces favourable effects of long-term regulation of arterial blood pressure in the kidney and inhibits many age-related vascular remodelling processes such as VSMC proliferation and endothelial dysfunction. In addition, oestrogen lowers cholesterol and improves vascular tone (Dubey et al. 2005). It is thus possible, that a complex interplay between a hormonal milieu and agents that induce inflammation will have an impact on the expression of the genes associated with CVD.

Strengths and limitations

This study is one of the few, which have investigated extensively common genetic variation and joint effects of a set of candidate genes of CVD that form a biologically meaningful pathway. In the era of GWA, this type of study is likely to become more common when we start to dissect how the new candidates and candidate pathways emerging from the GWAs contribute to the risk for common diseases.

None of the presented findings in study III represent strong evidence of association. There are a few possible reasons for this. The power was somewhat limited first, due to relatively small number of cases especially when analysing the two sexes separately and second, due to very large number of tests. Although we utilized two large prospective cohorts each followed for several years, we were able to include in our analysis only 154 female incident CVD cases and 373 male cases. This represents the down-side of case-cohort studies: even when using large population based cohorts and focusing on a common disease is it hard to be able to include thousands of incident cases in the analyses. Collaborative studies, such as the Morgam project (Evans et al. 2005) merging information from several case-cohort studies would provide a solution for this problem provided that between studies heterogeneity is not very large. Also, replication in several independent studies would provide support for validity of the results. Although cohort studies may not be very efficient in collecting large numbers of cases they have the advantage of being able to model more precisely without biases encountered in retrospective studies the prospective risk effect of strong risk factors that may be affected by the disease or lifestyle changes or have a long latent period (Manolio et al. 2006).

Our study aimed to investigate genes with a biological hypothesis of an effect on CVD and included parts of pathways with close relationships between the genes. The association of most of these genes was studied using common variants describing the common haplotype structure of the gene. Recent large scale genetic

association studies and meta-analyses have shown that the risk effect of the common associating variants is quite small, in the range of 1.1-2 (Lohmueller et al. 2003, Altshuler and Daly 2007). Our study, although describing well the common variation within these genes, may have not been powerful enough to detect such small effects even though the selection of biological candidates increased our *a priori* probability to detect association. Also, the relative heterogeneity in CVD phenotype encompassing ACS and ranging from unstable angina to ST-elevation MI may have influenced the power of this study. More homogenous phenotype restricting the study only to, for example, definite myocardial infarctions might have improved the odds.

Both the number of genotyped variants and analysed phenotypes was quite large in study III and we used a two-staged approach, FDR and additional nominal association limits to control for the number of tests. In addition, we compared results from the two cohorts in joint analysis of IL6 -related genes. From all of the tests, only two single SNP association results had a FDR value <10%. In this type of study with non-independent genetic markers, non-independent genes and a hierarchical nature of the data, FDR is possibly the best method to account for the multiple tests, although it is far from optimal. Replication in additional cohorts would have increased the probability of true positive association.

This study should be considered as being exploratory in nature and its findings should be replicated in other population-based studies. Based on the results of this study, we can not definitely exclude any of these genes from being a candidate gene for CVD. However, this study utilized two separate population cohorts and prospective design reducing the potential of selection and information bias in the study. The results consistent with earlier findings and clearly present in both cohorts, such as the association between the CRP variant *rs1800947* and CRP levels, lend further support to the validity of the study design and its findings. Taken together, these results indicate that the interleukin 6 pathway may be involved in the pathogenesis of CVD at many levels and argue for further studies, especially in women.

6 CONCLUDING REMARKS

We showed in this thesis that certain inflammatory genes contribute to the risk for CHD, ischemic stroke or CVD and interestingly most of the identified associations were among women. Furthermore, we showed that inflammatory factors and traits are associated with the DNA yield from whole blood samples.

During the course of this study, the available tools for genetic epidemiology have developed enormously. The publicly available full sequence of the human genome and many other organisms including primates has enabled scientists to compare sequence homology between species to identify conserved regions and polymorphisms and putative regulatory motifs. Together with the expanding catalogue of human polymorphisms and more detailed haplotype construction, this information has made the attempt to identify sequence variation contributing to the disease susceptibility more advanced.

Although the tools are becoming more easy to use and available to many research groups, genetic epidemiology faces huge challenges in terms of multiple testing with non-independent markers, genes and overlapping phenotype categories. New tools are definitely required to account for the number of tests statistically since increasing the number of samples indefinitely will not be a sustainable option. The accuracy of the results from genetic epidemiological studies will increase with increasing homogeneity in phenotype measurement and definition, and although collaborative studies with a number of participating centres will aid in increasing the number of individuals in the study, unless carefully accounted for via rigorous study design, they may result in greater noise in the genetic signal than aiding in tackling multiple testing problems due to increase in heterogeneity in measured phenotypes. During the follow-up period of this study, the implementation of troponin measurements in Finland between 1997-2000 improved the diagnosis of myocardial infarction (Salomaa et al. 2005). Although this has increased the precision of the diagnosis, it has mostly done it in individuals who would have been diagnosed for ACS also with older diagnostic tools. It is thus unlikely, that this would have substantially affected the heterogeneity in the CHD phenotype of this study.

One drawback of both case-control and -cohort studies is the susceptibility to unobserved differences in allele frequencies in subgroups of the base population, i.e. population stratification. Population based cohort studies are less vulnerable to this than carelessly selected case-control studies. This phenomenon causes overestimates in the effect sizes of the associating allele and false positive results if the causative variant is overrepresented, or if the penetrance of this variant is higher in a subgroup

(Setakis et al. 2006). An Icelandic study observed that in their population, which had been thought to be very homogenous, there was evidence of significant population stratification (Helgason et al. 2005). Differences in allele frequencies are also evident in the Finnish population (Pastinen et al. 2001). Without extensive genotype information from the population, it is impossible to correct for population stratification with sampling strategies in population-based studies. The use of family members as controls and novel statistical methods, such as, genomic control offer some help (Terwilliger and Ott 1992, Devlin et al. 2001). Simulation studies indicate, however, that statistical association methods using the χ^2 statistic are quite robust against population stratification and significant excess of false positive findings occur only when there are two very different subgroups in the analysis (Setakis et al. 2006). In our study, we were able to use the knowledge of geographical east-west difference in allele frequencies present in Finland and adjusted or stratified the analysis accordingly thus reducing the likelihood of false positive findings due to population stratification.

Other challenges are common to traditional epidemiology. It has been shown that some risk factors change over time (Zavaroni et al. 2002), and that although cross-sectional baseline measurements do provide good estimates of the risk, updated measurements offer better predictive power (Karp et al. 2004). It may thus also be justified to invest in repeated measurements and longitudinal study design in addition to large-scale genotyping in traditional case-control or case-cohort designs. More precise models, which would account for the traditional risk factors, might help to tease out the small genetic effects accounting for the rest of the phenotypic variability and enable investigation of so-called variability genes in addition to level genes.

Novel approaches including expression quantitative trait mapping, and the use of new intermediate phenotypes like IMT and non-invasive measures of endothelial function or arterial stiffness, are likely to advance the field of cardiovascular research in the future. Moreover, detailed analysis of the different plaque rupture-like features currently available in murine models will enhance our understanding of the processes leading to acute myocardial infarction events (Falk et al. 2007).

So far, the variants identified in various studies searching for susceptibility genes for common traits have been shown to have a rather small effect at the population level. However, although at this point the variants do not add substantially to the predictive power of traditional risk factors, they may prove to be clinically useful at the individual level. Also, with accumulating knowledge of several variants, the combined information of them may one day add to the predictive power.

In this study we identified several suggestive associations in the selected CVD candidate genes with relatively small effects. Common to these findings was marked

sexual polarization of the results most of the findings being evident in women only. In study II variation in the *SEPS1* gene locus associated with the increased risk for CHD and ischemic stroke in women and study III confirmed significant genotype-sex interaction for haplotype H4 associated variants. Study III identified many more associations with evidence of sex-genotype interaction. Unfortunately, there are very few large scale association studies on cardiovascular disease in women and even fewer have formally tested sex-genotype interaction making it hard to compare our results with others'. In addition to the hormonal difference, it is possible that subtle genetic effects can be more readily identified in women due to a smaller risk factor load than in men. We hope that study III will promote collection and analysis of samples with more female CVD cases.

Based on this study, we can conclude that inflammation and genetic variation in genes related to inflammation play a minor but a consistent role in the risk for CVD, especially in women. This study suggests that new genetic epidemiological studies, in which participants are not ascertained based on the availability or amount of the DNA sample, are warranted in this field.

7 ACKNOWLEDGMENTS

This study was carried out in the Department of Molecular Medicine at the National Public Health Institute, Helsinki, Finland. I wish to acknowledge the former and present directors of the institute, Jussi Huttunen and Pekka Puska, and the head of the Department of Molecular Medicine, Adjunct Professor Anu Jalanko, for providing excellent research facilities which have made this work possible.

This work was financially supported by the Finnish Foundation for Cardiovascular Research, the Center of Excellence on Disease Genetics by the Academy of Finland, the Research Foundation of Orion Corporation, Aarne Koskelo, Maud Kuistila Memorial, Aarne and Auli Turunen, Veritas and Paulo Foundations, Helsinki University and the GenomEUtwin-project under the European Commission "Quality of Life and Management of the Living Resources" of the 5th Framework Programme (QLG2-CT-2002-01254). I also wish to thank all the participants of the FINRISK and Health 2000 studies.

Adjunct Professors Tarja Laitinen and Tomi Pastinen are warmly thanked for their valuable comments and criticism while reviewing this thesis. My thesis committee members, Adjunct Professor Tarja Laitinen and Research Professor Pekka Jousilahti, are thanked for their support in my thesis work during these years.

My passion for science really sparked into flames while I was working at the Finnish Forest Research Institute in the group of Professor Jarkko Hantula. To him, I would like to express my deepest gratitude for guiding my first steps in molecular genetics and for all our discussions. Without this inspiration I would have not started my thesis work.

I wish to thank Academy Professor Leena Peltonen-Palotie for setting the ground for excellent research in complex disease genetics and providing us with ample flexibility and resources in our various research aims. Her never ending enthusiasm and eagerness to explore new research ideas have been an inspiration to me.

I am very grateful to my supervisor, Adjunct Professor Markus Perola, first in trusting and training me as his successor as the DNA-extraction unit coordinator and then as a young scientist. His and Leena's faith in two new graduate students, myself and Kirsi, enabled us to independently learn how to organize a big study, an experience which I value very much. Markus has always been available for me to consult with any (how weird they might have been) research dilemmas and to discuss science and life in general. I also wish to thank him for guiding my steps in the world of big collaborative studies, his cheerful company in many trips abroad and tutorials in wine tasting.

I would like to thank the secretaries, Sari M, Sari K, Tuija, Sanna, and Mika for helping with various practicalities with remarkable efficacy and a positive attitude. Sisko is thanked for smoothly handling the orders and Jari, Juri, Teemu and Tero for their assistance with computer-related problems. I would also like to thank Juha K. for assisting us with our genotype database.

I am thankful for all the help in genotyping and support in my research I have got over the years from the senior scientists Janna Saarela and Kaisa Silander. I want to thank Ismo Ulmanen for initially hiring me to the department and for his always supportive attitude and Matti Jauhiainen for sharing his knowledge on atherosclerosis and for reviewing that chapter of my thesis. I am greatly indebted to Marjo Kestilä for all the help she has given me during the thesis writing process.

I wish to thank Professor Veikko Salomaa and Kari Kuulasmaa for sharing their expertise in epidemiology and Veikko for his involvement and guidance in all stages of my thesis project. I wish also to acknowledge our collaborators Professor Erkki Vartiainen and Joanne Curran, and from the Morgam project: Professor Alun Evans and people from the Morgam Data Center, Zygimantas Cepaitis, Ari Haukijärvi, Bijoy Joseph and Matti Nieminen, whose work has been a corner stone of the project. I would like to thank Sangita Kulathinal, Juha Karvanen and Olli Saarela and especially Samuli Ripatti for patiently explaining statistical details to us less experienced, and Olli for assisting with the power calculations for this thesis. Minna, Siv and Arja are thanked for their skillful assistance in genotyping. Tanja and Anne I would like to thank for introducing me to the world of EMSA, and especially Anne for always good humoured atmosphere she created in the lab and Ismo Ulmanen for helping me interpret the results. I am grateful for the help and advice I have obtained from Pekka in all issues related to genotyping and particularly, in the early days of my project, with the wonders PCR. Minttu and Elina, I wish to thank you for your huge work in aliquotting Morgam DNA samples and Minttu for excellent company in workshops abroad. I am indebted to Päivi for taking over my responsibilities in co-ordinating Morgam sample logistics.

Outi, I am so glad you accepted the job of Sample-Queen! It has been a great pleasure to see how much better you are at that work than I ever was and know that the DNA extraction unit is in very capable hands. My thesis would have not been possible without you. I also want to thank you for coming late to work so I was not the only one, for all our discussions and the positive atmosphere you created in our office.

Kati and Kirsi, your company during these years of genotyping and struggling over analytical issues has always been cheerful and full of joy and has meant a lot to me. I feel that I would not be half the scientist I now am without our discussions. Kirsi, with whom I share a history of team work already from Marigold times, could always be trusted to work towards our common aim and I thank her for excellent company over the years and careful review of my thesis. I wish to thank Kati, in addition to teaching me SAS coding, for all the fun times outside the office, starting

salsa dancing with me, all the wonderful trips together and for your support over the years. I want to thank Outi, Kati, Kirsi, Johannes, Elina, Katja, Henna, Kaisu and Annina for their friendship and making our office a place where it was always nice to come to work. I also wish to thank the guys Johannes, Tero and Sampo for excellent company during lunch breaks and conference trips.

Working in the department was much more fun than I could have ever anticipated mainly thanks to our coffee group. Thank you all for making this time so enjoyable!

I would specially like to thank Pekka and Camilla for sharing my passion for sailing. Our sailing trips, however short during weekdays, have provided me with much needed break from the work and have given me new strength to carry on with my thesis. My friends Anne, Päivi Pöyry, Katriina and Tuomo, Tiina and Mikko, my friends from Marigold, Pikkis, Päivi Putkonen, Minna, Mari and Johanna, I wish to thank for giving me a perspective of what life can be outside the office and the former team members for many memorable bring-a-dish parties.

Finally, I would like to thank my parents Hannu and Inkeri, siblings Erja and Tero, grandmothers Raili and Aino and late grandfathers Pentti and Olavi for their love and support during my studies. It is too seldom that I express my gratitude for all you have done for me. I feel that, for you, it does not matter whether I succeed academically or not – it will always be enough for me to just be me – and this has given me a solid foundation to stand on.

Helsinki, April 16th, 2008

Mervi Alanne

8 REFERENCES

Abdollahi MR, Lewis RM, Gaunt TR, Cumming DV, Rodriguez S, Rose-Zerilli M, Collins AR, Syddall HE, Howell WM, Cooper C, et al. Quantitated transcript haplotypes (QTH) of AGTR1, reduced abundance of mRNA haplotypes containing 1166C (rs5186:A>C), and relevance to metabolic syndrome traits. Hum Mutat 2007;28:365-73.

Agerholm-Larsen B, Nordestgaard BG and Tybjærg-Hansen A. ACE gene polymorphism in cardiovascular disease: meta-analyses of small and large studies in whites. Arterioscler Thromb Vasc Biol 2000;20:484-92.

Ahn KS, Sethi G and Aggarwal BB. Nuclear factor-kappa B: from clone to clinic. Curr Mol Med 2007;7:619-37.

Akishima Y, Akasaka Y, Ishikawa Y, Lijun Z, Kiguchi H, Ito K, Itabe H and Ishii T. Role of macrophage and smooth muscle cell apoptosis in association with oxidized low-density lipoprotein in the atherosclerotic development. Mod Pathol 2004;18:365-73.

Alber DG, Powell KL, Vallance P, Goodwin DA and Grahame-Clarke C. Herpesvirus infection accelerates atherosclerosis in the apolipoprotein E-deficient mouse. Circulation 2000;102:779-85.

Alberti KG and Zimmet P. The metabolic syndrome: time to reflect. Curr Diab Rep 2006;6:259-61.

Altshuler D and Daly M. Guilt beyond a reasonable doubt. Nat Genet 2007;39:813-15.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290:457-65.

Antoniou AC and Easton DF. Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol 2003;25:190-202.

Ardlie KG, Kruglyak L and Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 2002;3:299-309.

Arner P. Introduction: the inflammation orchestra in adipose tissue. J Intern Med 2007;262:404-7.

Aromaa A and Koskinen S, editors. Health and functional capacity in Finland: baseline results of the Health 2000 health examination survey. Helsinki, Finland: National Public Health Institute; 2002.

Asselbergs FW, Pai JK, Rexrode KM, Hunter DJ and Rimm EB. Effects of lymphotoxin-alpha gene and galectin-2 gene polymorphisms on inflammatory biomarkers, cellular adhesion molecules and risk of coronary heart disease. Clin Sci (Lond) 2007;112:291-98.

Austin MA, Breslow JL, Hennekens CH, Buring JE, Willett WC and Krauss RM. Low-density lipoprotein subclass patterns and risk of myocardial infarction. JAMA 1988;260:1917-21.

Austin MA, Zhang C, Humphries SE, Chandler WL, Talmud PJ, Edwards KL, Leonetti DL, McNeely MJ and Fujimoto WY. Heritability of C-reactive protein and association with apolipoprotein E genotypes in Japanese Americans. Ann Hum Genet 2004;68:179-88.

Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet 2006;7:781-91.

Balkau B, Charles MA, Drivsholm T, Borch-Johnsen K, Wareham N, Yudkin JS, Morris R, Zavaroni I, van Dam R, Feskins E, et al. Frequency of the WHO metabolic syndrome in

European cohorts, and an alternative definition of an insulin resistance syndrome. Diabetes Metab 2002;28:364-76.

Barbaux S, Tregouet DA, Nicaud V, Poirier O, Perret C, Godefroy T, Francomme C, Combadiere C, Arveiler D, Luc G, et al. Polymorphisms in 33 inflammatory genes and risk of myocardial infarction-a system genetics approach. J Mol Med 2007;85:1271-80.

Barlow WE. Robust variance estimation for the case-cohort design. Biometrics 1994;50:1064-72.

Barrett JC, Fry B, Maller J and Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-65.

Behne D and Kyriakopoulos A. Mammalian selenium-containing proteins. Annu Rev Nutr 2001;21:453-73.

Bello N and Mosca L. Epidemiology of coronary heart disease in women. Prog Cardiovasc Dis 2004;46:287-95.

Benjamini Y and Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 1995;57:289-300.

Berthier-Schaad Y, Kao WH, Coresh J, Zhang L, Ingersoll RG, Stephens R and Smith MW. Reliability of high-throughput genotyping of whole genome amplified DNA in SNP genotyping studies. Electrophoresis 2007;28:2812-17.

Bertoli-Avella AM, Dekker MC, Aulchenko YS, Houwing-Duistermaat JJ, Simons E, Testers L, Pardo LM, Rademaker TA, Snijders PJ, van Swieten JC, et al. Evidence for novel loci for late-onset Parkinson's disease in a genetic isolate from the Netherlands. Hum Genet 2006;119:51-60.

Biswas P, Delfanti F, Bernasconi S, Mengozzi M, Cota M, Polentarutti N, Mantovani A, Lazzarin A, Sozzani S and Poli G. Interleukin-6 induces monocyte chemotactic protein-1 in peripheral blood mononuclear cells and in the U937 cell line. Blood 1998;91:258-65.

Bodzioch M, Orso E, Klucken J, Langmann T, Bottcher A, Diederich W, Drobnik W, Barlage S, Buchler C, Porsch-Ozcurumez M, et al. The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. Nat Genet 1999;22:347-51.

Boekholdt SM, Bijsterveld NR, Moons AH, Levi M, Büller HR and Peters RJ. Genetic variation in coagulation and fibrinolytic proteins and their relation with acute myocardial infarction: a systematic review. Circulation 2001;104:3063-68.

Bowden DW, Rudock M, Ziegler J, Lehtinen AB, Xu J, Wagenknecht LE, Herrington D, Rich SS, Freedman BI, Carr JJ, et al. Coincident linkage of type 2 diabetes, metabolic syndrome, and measures of cardiovascular disease in a genome scan of the diabetes heart study. Diabetes 2006;55:1985-94.

Broeckel U, Hengstenberg C, Mayer B, Holmer S, Martin LJ, Comuzzie AG, Blangero J, Nurnberg P, Reis A, Riegger GA, et al. A comprehensive linkage analysis for myocardial infarction and its related risk factors. Nat Genet 2002;30:210-14.

Browatzki M, Larsen D, Pfeiffer CA, Gehrke SG, Schmidt J, Kranzhöfer A, Katus HA and Kranzhöfer R. Angiotensin II stimulates matrix metalloproteinase secretion in human vascular smooth muscle cells via nuclear factor-kappaB and activator protein 1 in a redox-sensitive manner. J Vasc Res 2005;42:415-23.

Bush DE, Jones CE, Bass KM, Walters GK, Bruza JM and Ouyang P. Estrogen replacement reverses endothelial dysfunction in postmenopausal women. Am J Med 1998;104:552-58.

Cardon LR and Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001;2:91-99.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L and Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004;74:106-20.

Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, Liu K, Williams OD, Iribarren C, Lewis EC, et al. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. Am J Hum Genet 2005;77:64-77.

Casas JP, Hingorani AD, Bautista LE and Sharma P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. Arch Neurol 2004;61:1652-61.

Chanson M and Kwak BR. Connexin37: a potential modifier gene of inflammatory disease. J Mol Med 2007;85:787-95.

Choy JC, Granville DJ, Hunt DW and McManus BM. Endothelial cell apoptosis: biochemical characteristics and potential implications for atherosclerosis. J Mol Cell Cardiol 2001;33:1673-90.

Churchill GA and Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics 1994;138:963-71.

Clarke R, Xu P, Bennett D, Lewington S, Zondervan K, Parish S, Palmer A, Clark S, Cardon L, Peto R, et al. Lymphotoxin-alpha gene and risk of myocardial infarction in 6,928 cases and 2,712 controls in the ISIS case-control study. PLoS Genet 2006;2:e107.

Clauss A. Rapid physiological coagulation method in determination of fibrinogen. Acta Haematol 1957;17:237-46.

Coles B, Fielding CA, Rose-John S, Scheller J, Jones SA and O'Donnell VB. Classic interleukin-6 receptor signaling and interleukin-6 trans-signaling differentially control angiotensin II-dependent hypertension, cardiac signal transducer and activator of transcription-3 activation, and vascular hypertrophy in vivo. Am J Pathol 2007;171:315-25.

Colhoun HM, McKeigue PM and Davey Smith G. Problems of reporting genetic associations with complex outcomes. Lancet 2003;361:865-72.

Collins FS. Positional cloning moves from perditional to traditional. Nat Genet 1995;9:347-50.

Collins FS, Guyer MS and Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. Science 1997;278:1580-81.

Connelly JJ, Wang T, Cox JE, Haynes C, Wang L, Shah SH, Crosslin DR, Hale AB, Nelson S, Crossman DC, et al. GATA2 is associated with familial early-onset coronary artery disease. PLoS Genet 2006;2:e139.

Cordell HJ and Clayton DG. Genetic association studies. Lancet 2005;366:1121-31.

Curran JE, Jowett JB, Elliott KS, Gao Y, Gluschenko K, Wang J, Abel Azim DM, Cai G, Mahaney MC, Comuzzie AG, et al. Genetic variation in selenoprotein S influences inflammatory response. Nat Genet 2005;37:1234-41.

Daly MJ, Rioux JD, Schaffner SE, Hudson TJ and Lander ES. High-resolution haplotype structure in the human genome. Nat Genet 2001;29:229-32.

Daly MJ and Altshuler D. Partners in crime. Nat Genet 2005;37:337-38.

Danesh J, Collins R, Appleby P and Peto R. Association of Fibrinogen, C-reactive Protein, Albumin, or Leukocyte Count With Coronary Heart Disease: Meta-analyses of Prospective Studies. JAMA 1998;279:1477-82.

Danesh J, Whincup P, Walker M, Lennon L, Thomson A, Appleby P, Gallimore JR and Pepys MB. Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. BMJ 2000;321:199-204.

Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB and Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N Engl J Med 2004;350:1387-97.

Danesh J, Lewington S, Thompson SG, Lowe GD, Collins R, Kostis JB, Wilson AC, Folsom AR, Wu K, Benderly M, et al. Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. JAMA 2005;294:1799-809.

Darvall KA, Sam RC, Silverman SH, Bradbury AW and Adam DJ. Obesity and thrombosis. Eur J Vasc Endovasc Surg 2007;33:223-33.

Davies MJ, Richardson PD, Woolf N, Katz DR and Mann J. Risk of thrombosis in human atherosclerotic plaques: role of extracellular lipid, macrophage, and smooth muscle cell content. Br Heart J 1993;69:377-81.

Davis GK and Roberts DH. Molecular genetics of the renin-angiotensin system: implications for angiotensin II receptor blockade. Pharmacol Ther 1997;75:43-50.

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 2002;99:5261-66.

Devlin B, Roeder K and Bacanu SA. Unbiased methods for population-based association studies. Genet Epidemiol 2001;21:273-84.

Drenos F, Whittaker JC and Humphries SE. The use of meta-analysis risk estimates for candidate genes in combination to predict coronary heart disease risk. Ann Hum Genet 2007;71:611-19.

Dubey RK, Imthurn B, Barton M and Jackson EK. Vascular consequences of menopause and hormone therapy: importance of timing of treatment and type of estrogen. Cardiovasc Res 2005;66:295-306.

Dugue B and Leppänen E. Short-term variability in the concentration of serum interleukin-6 and its soluble receptor in subjectively healthy persons. Clin Chem Lab Med 1998;36:323-25.

Dwyer JH, Allayee H, Dwyer KM, Fan J, Wu H, Mar R, Lusis AJ and Mehrabian M. Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. N Engl J Med 2004;350:29-37.

Elias-Smale SE, Kardys I, Oudkerk M, Hofman A and Witteman JC. C-reactive protein is related to extent and progression of coronary and extra-coronary atherosclerosis; results from the Rotterdam study. Atherosclerosis 2007;195:e195-202.

Engert JC, Lemire M, Faith J, Brisson D, Fujiwara TM, Roslin NM, Brewer CG, Montpetit A, Darmond-Zwaig C, Renaud Y, et al. Identification of a chromosome 8p locus for early-onset coronary heart disease in a French Canadian population. Eur J Hum Genet 2008;16:105-14.

Erkeller-Yuksel FM, Deneys V, Yuksel B, Hannet I, Hulstaert F, Hamilton C, Mackinnon H, Stokes LT, Munhyeshuli V, Vanlangendonck F, et al. Age-related changes in human blood lymphocyte subpopulations. J Pediatr 1992;120:216-22.

Evans A, Salomaa V, Kulathinal S, Asplund K, Cambien F, Ferrario M, Perola M, Peltonen L, Shields D, Tunstall-Pedoe H, et al. MORGAM (an international pooling of cardiovascular cohorts). Int J Epidemiol 2005;34:21-27.

Falk E, Shah PK and Fuster V. Coronary plaque disruption. Circulation 1995;92:657-71.

Falk E, Schwartz SM, Galis ZS and Rosenfeld ME. Putative murine models of plaque rupture. Arterioscler Thromb Vasc Biol 2007;27:969-72.

Farkas DH, Drevon AM, Kiechle FL, DiCarlo RG, Heath EM and Crisan D. Specimen stability for DNA-based diagnostic testing. Diagn Mol Pathol 1996;5:227-35.

Farrall M, Green FR, Peden JF, Olsson PG, Clarke R, Hellenius ML, Rust S, Lagercrantz J, Franzosi MG, Schulte H, et al. Genome-wide mapping of susceptibility to coronary artery disease identifies a novel replicated locus on chromosome 17. PLoS Genet 2006;2:e72.

Feigelson HS, Rodriguez C, Robertson AS, Jacobs EJ, Calle EE, Reid YA and Thun MJ. Determinants of DNA yield and quality from buccal cell samples collected with mouthwash. Cancer Epidemiol Biomarkers Prev 2001;10:1005-8.

Feldman CL and Stone PH. Intravascular hemodynamic factors responsible for progression of coronary atherosclerosis and development of vulnerable plaque. Curr Opin Cardiol 2000;15:430-40.

Ferrante AW, Jr. Obesity-induced inflammation: a metabolic dialogue in the language of inflammation. J Intern Med 2007;262:408-14.

Ferrari SL, Ahn-Luong L, Garnero P, Humphries SE and Greenspan SL. Two promoter polymorphisms regulating interleukin-6 gene expression are associated with circulating levels of C-reactive protein and markers of bone resorption in postmenopausal women. J Clin Endocrinol Metab 2003;88:255-59.

Fischer M, Mayer B, Baessler A, Riegger G, Erdmann J, Hengstenberg C and Schunkert H. Familial aggregation of left main coronary artery disease and future risk of coronary events in asymptomatic siblings of affected patients. Eur Heart J 2007;28:2432-37.

Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S and Woo P. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 1998;102:1369-76.

Floßmann E, Schulz UG and Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. Stroke 2004;35:212-27.

Fox CS, Larson MG, Corey D, Feng D, Lindpaintner K, Polak JF, Wolf PA, D'Agostino RB, Tofler GH and O'Donnell CJ. Absence of association between polymorphisms in the hemostatic factor pathway genes and carotid intimal medial thickness: the Framingham Heart Study. Stroke 2004;35:e65-67.

Francke S, Manraj M, Lacquemant C, Lecoeur C, Lepretre F, Passa P, Hebe A, Corset L, Yan SL, Lahmidi S, et al. A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. Hum Mol Genet 2001;10:2751-65.

Frikke-Schmidt R, Nordestgaard BG, Jensen GB and Tybjaerg-Hansen A. Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. J Clin Invest 2004;114:1343-53.

Frizzell JP. Acute stroke: pathophysiology, diagnosis, and treatment. AACN Clin Issues 2005;16:421-40.

Fuster V, Moreno PR, Fayad ZA, Corti R and Badimon JJ. Atherothrombosis and high-risk plaque: part I: evolving concepts. J Am Coll Cardiol 2005;46:937-54.

Gao Y, Feng HC, Walder K, Bolton K, Sunderland T, Bishara N, Quick M, Kantham L and Collier GR. Regulation of the selenoprotein SelS by glucose deprivation and endoplasmic reticulum stress - SelS is a novel glucose-regulated protein. FEBS Lett 2004;563:185-90.

Gao Y, Hannan NRF, Wanyonyi S, Konstantopolous N, Pagnon J, Feng HC, Jowett JBM, Kim K-H, Walder K and Collier GR. Activation of the selenoprotein SEPS1 gene expression by pro-inflammatory cytokines in HepG2 cells. Cytokine 2006;33:246-51.

Gao Y, Pagnon J, Feng HC, Konstantopolous N, Jowett JB, Walder K and Collier GR. Secretion of the glucose-regulated selenoprotein SEPS1 from hepatoma cells. Biochem Biophys Res Commun 2007;356:636-41.

Gargalovic PS, Gharavi NM, Clark MJ, Pagnon J, Yang WP, He A, Truong A, Baruch-Oren T, Berliner JA, Kirchgessner TG, et al. The unfolded protein response is an important regulator of inflammatory genes in endothelial cells. Arterioscler Thromb Vasc Biol 2006;26:2490-96.

Georges JL, Loukaci V, Poirier O, Evans A, Luc G, Arveiler D, Ruidavets JB, Cambien F and Tiret L. Interleukin-6 gene polymorphisms and susceptibility to myocardial infarction: the ECTIM study. Etude Cas-Temoin de l'Infarctus du Myocarde. J Mol Med 2001;79:300-5.

Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S and Snyder M. What is a gene, post-ENCODE? History and updated definition. Genome Res 2007;17:669-81.

Glazier AM, Nadeau JH and Aitman TJ. Finding genes that underlie complex traits. Science 2002;298:2345-49.

Goldman L and Ausiello D. Cardiovascular disease. In: Cecil Textbook of Medicine. 22nd ed. Philadelfia: Saunders; 2004. p. 242-492.

Grant PJ and Humphries SE. Genetic determinants of arterial thrombosis. Baillieres Best Pract Res Clin Haematol 1999;12:505-32.

Gretarsdottir S, Sveinbjörnsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E, Agnarsson U, Shkolny D, Einarsson G, Gudjonsdottir HM, Valdimarsson EM, et al. Localization of a susceptibility gene for common forms of stroke to 5q12. Am J Hum Genet 2002;70:593-603.

Gruys E, Toussaint MJ, Niewold TA and Koopmans SJ. Acute phase reaction and acute phase proteins. J Zhejiang Univ Sci B 2005;6:1045-56.

Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR and Kidd KK. Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. Eur J Hum Genet 2007;15:302-12.

Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature 2007;448:353-57.

Guilford P, Ben Arab S, Blanchard S, Levilliers J, Weissenbach J, Belkahia A and Petit C. A non-syndromic form of neurosensory, recessive deafness maps to the pericentromeric region of chromosome 13q. Nat Genet 1994;6:24-28.

Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 2007;448:591-94.

Hansson GK and Libby P. The immune response in atherosclerosis: a double-edged sword. Nat Rev Immunol 2006;6:508-19.

Harrap SB, Zammit KS, Wong ZY, Williams FM, Bahlo M, Tonkin AM and Anderson ST. Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. Arterioscler Thromb Vasc Biol 2002;22:874-78.

Hattersley AT and McCarthy MI. What makes a good genetic association study? Lancet 2005;366:1315-23.

Hauser ER, Crossman DC, Granger CB, Haines JL, Jones CJ, Mooser V, McAdam B, Winkelmann BR, Wiseman AH, Muhlestein JB, et al. A genomewide scan for early-onset coronary artery disease in 438 families: the GENECARD Study. Am J Hum Genet 2004;75:436-47.

Hawkins TL, Detter JC and Richardson PM. Whole genome amplification--applications and advances. Curr Opin Biotechnol 2002;13:65-67.

Hearne CM, Ghosh S and Todd JA. Microsatellites for linkage analysis of genetic traits. Trends Genet 1992;8:288-94.

Heinrich PC, Castell JV and Andus T. Interleukin-6 and the acute phase response. Biochem J 1990;265:621-36.

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, Gudmundsson G, Grant SF, Thorgeirsson G, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 2004;36:233-39.

Helgadottir A, Gretarsdottir S, St Clair D, Manolescu A, Cheung J, Thorleifsson G, Pasdar A, Grant SF, Whalley LJ, Hakonarson H, et al. Association between the gene encoding 5-lipoxygenase-activating protein and stroke replicated in a Scottish population. Am J Hum Genet 2005;76:505-9.

Helgadottir A, Manolescu A, Helgason A, Thorleifsson G, Thorsteinsdottir U, Gudbjartsson DF, Gretarsdottir S, Magnusson KP, Gudmundsson G, Hicks A, et al. A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. Nat Genet 2006;38:68-74.

Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 2007;316:1491-93.

Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J and Stefansson K. An Icelandic example of the impact of population structure on association studies. Nat Genet 2005;37:90-95.

Helmert U, Shea S and Bammann K. The impact of occupation on self-reported cardiovascular morbidity in western Germany: gender differences. Rev Environ Health 1997;12:25-42.

Henn V, Slupsky JR, Grafe M, Anagnostopoulos I, Forster R, Muller-Berghaus G and Kroczek RA. CD40 ligand on activated platelets triggers an inflammatory reaction of endothelial cells. Nature 1998;391:591-94.

Hirschhorn JN and Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95-108.

Huber SA, Sakkinen P, Conze D, Hardin N and Tracy R. Interleukin-6 exacerbates early atherosclerosis in mice. Arterioscler Thromb Vasc Biol 1999;19:2364-67.

Humphries SE, Luong LA, Ogg MS, Hawe E and Miller GJ. The interleukin-6 -174 G/C promoter polymorphism is associated with risk of coronary heart disease and systolic blood pressure in healthy men. Eur Heart J 2001;22:2243-52.

Hyrenbach S, Pezzini A, del Zotto E, Giossi A, Lichy C, Kloss M, Werner I, Padovani A, Brandt T and Grond-Ginsbach C. No association of the -105 promoter polymorphism of the selenoprotein S encoding gene SEPS1 with cerebrovascular disease. Eur J Neurol 2007;14:1173-75.

Ijäs P, Nuotio K, Saksi J, Soinne L, Saimanen E, Karjalainen-Lindsberg ML, Salonen O, Sarna S, Tuimala J, Kovanen PT, et al. Microarray analysis reveals overexpression of CD163 and HO-1 in symptomatic carotid plaques. Arterioscler Thromb Vasc Biol 2007;27:154-60.

Ioannidis JP, Ntzani EE, Trikalinos TA and Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001;29:306-9.

Isles CG, Hole DJ, Hawthorne VM and Lever AF. Relation between coronary risk and coronary mortality in women of the Renfrew and Paisley survey: comparison with men. Lancet 1992;339:702-6.

Jeffreys AJ, Kauppi L and Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 2001;29:217-22.

Jenny NS, Tracy RP, Ogg MS, Luong le A, Kuller LH, Arnold AM, Sharrett AR and Humphries SE. In the elderly, interleukin-6 plasma levels and the -174G>C polymorphism are associated with the development of cardiovascular disease. Arterioscler Thromb Vasc Biol 2002;22:2066-71.

Jensen J, Nilas L and Christiansen C. Influence of menopause on serum lipids and lipoproteins. Maturitas 1990;12:321-31.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, et al. Haplotype tagging for the identification of common disease genes. Nat Genet 2001;29:233-37.

Jonasson L, Holm J, Skalli O, Bondjers G and Hansson GK. Regional accumulations of T cells, macrophages, and smooth muscle cells in the human atherosclerotic plaque. Arteriosclerosis 1986;6:131-38.

Jorde LB. Linkage disequilibrium as a gene-mapping tool. Am J Hum Genet 1995;56:11-14.

Jousilahti P, Puska P, Vartiainen E, Pekkanen J and Tuomilehto J. Parental history of premature coronary heart disease: an independent risk factor of myocardial infarction. J Clin Epidemiol 1996a;49:497-503.

Jousilahti P, Tuomilehto J, Vartiainen E, Pekkanen J and Puska P. Body weight, cardiovascular risk factors, and coronary mortality. 15-year follow-up of middle-aged men and women in eastern Finland. Circulation 1996b;93:1372-79.

Jousilahti P, Vartiainen E, Tuomilehto J and Puska P. Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. Circulation 1999;99:1165-72.

Jousilahti P, Salomaa V, Rasi V, Vahtera E and Palosuo T. The association of c-reactive protein, serum amyloid a and fibrinogen with prevalent coronary heart disease -- baseline findings of the PAIS project. Atherosclerosis 2001;156:451-56.

Kaessmann H, Zöllner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, Uhlén M and Pääbo S. Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet 2002;70:673-85.

Kardys I, de Maat MP, Uitterlinden AG, Hofman A and Witteman JC. C-reactive protein gene haplotypes and risk of coronary heart disease: the Rotterdam Study. Eur Heart J 2006;27:1331-37.

Karp I, Abrahamowicz M, Bartlett G and Pilote L. Updated risk factor values and the ability of the multivariable risk score to predict coronary heart disease. Am J Epidemiol 2004;160:707-16.

Kathiresan S, Larson MG, Vasan RS, Guo CY, Gona P, Keaney JF, Jr., Wilson PW, Newton-Cheh C, Musone SL, Camargo AL, et al. Contribution of clinical correlates and 13 C-reactive protein gene polymorphisms to interindividual variability in serum C-reactive protein level. Circulation 2006a;113:1415-23.

Kathiresan S, Yang Q, Larson MG, Camargo AL, Tofler GH, Hirschhorn JN, Gabriel SB and O'Donnell CJ. Common genetic variation in five thrombosis genes and relations to plasma hemostatic protein level and cardiovascular disease risk. Arterioscler Thromb Vasc Biol 2006b;26:1405-12.

Kathiresan S, Manning AK, Demissie S, D'Agostino RB, Surti A, Guiducci C, Gianniny L, Burtt NP, Melander O, Orho-Melander M, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC Med Genet 2007;8 Suppl 1:S17.

Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 2008;40:189-97.

Kawanami D, Maemura K, Takeda N, Harada T, Nojiri T, Saito T, Manabe I, Imai Y and Nagai R. C-reactive protein induces VCAM-1 gene expression through NF-kappaB activation in vascular endothelial cells. Atherosclerosis 2006;185:39-46.

Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, Delepine M, Lathrop M, Peto R and Collins R. Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. Int J Epidemiol 2006;35:935-43.

Kelsell DP, Dunlop J, Stevens HP, Lench NJ, Liang JN, Parry G, Mueller RF and Leigh IM. Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. Nature 1997;387:80-83.

Khalil MF, Wagner WD and Goldberg IJ. Molecular interactions leading to lipoprotein retention and the initiation of atherosclerosis. Arterioscler Thromb Vasc Biol 2004;24:2211-8.

Kidd KK, Pakstis AJ, Speed WC and Kidd JR. Understanding human DNA sequence variation. J Hered 2004;95:406-20.

Kim KH, Gao Y, Walder K, Collier GR, Skelton J and Kissebah AH. SEPS1 protects RAW264.7 cells from pharmacological ER stress agent-induced apoptosis. Biochem Biophys Res Commun 2007;354:127-32.

Kim S and De Gruttola V. Strategies for cohort sampling under the Cox proportional hazards model, application to an AIDS clinical trial. Lifetime Data Anal 1999;5:149-72.

Koch W, Hoppmann P, Mueller JC, Schömig A and Kastrati A. No association of polymorphisms in the gene encoding 5-lipoxygenase-activating protein and myocardial infarction in a large central European population. Genet Med 2007;9:123-29.

Komulainen K, Alanne M, Auro K, Kilpikari R, Pajukanta P, Saarela J, Ellonen P, Salminen K, Kulathinal S, Kuulasmaa K, et al. Risk alleles of USF1 gene predict cardiovascular disease of women in two prospective studies. PLoS Genet 2006;2:e69.

Kostulas K, Gretarsdottir S, Kostulas V, Manolescu A, Helgadottir A, Thorleifsson G, Gudmundsson LJ, Thorsteinsdottir U, Gulcher JR, Stefansson K, et al. PDE4D and ALOX5AP genetic variants and risk for Ischemic Cerebrovascular Disease in Sweden. J Neurol Sci 2007;263:113-17.

Kovacs A, Tornvall P, Nilsson R, Tegnér J, Hamsten A and Björkegren J. Human C-reactive protein slows atherosclerosis development in a mouse model with human-like hypercholesterolemia. Proc Natl Acad Sci U S A 2007;104:13768-73.

Kranzhöfer R, Schmidt J, Pfeiffer CA, Hagl S, Libby P and Kübler W. Angiotensin induces inflammatory activation of human vascular smooth muscle cells. Arterioscler Thromb Vasc Biol 1999;19:1623-29.

Krauss RM. Atherogenicity of triglyceride-rich lipoproteins. Am J Cardiol 1998;81:13B-17B.

Kruglyak L. The use of a genetic map of biallelic markers in linkage studies. Nat Genet 1997;17:21-24.

Kruglyak L. Genetic isolates: separate but equal? Proc Natl Acad Sci U S A 1999;96:1170-72.

Kruglyak L and Nickerson DA. Variation is the spice of life. Nat Genet 2001;27:234-36.

Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigo R and Gladyshev VN. Characterization of mammalian selenoproteomes. Science 2003;300:1439-43.

Kulathinal S, Karvanen J, Saarela O, Kuulasmaa K and Project FT. Case-cohort design in practice -- experiences from the MORGAM Project. Epidemiol Perspect Innov 2007;4:15.

Laatikainen T, Pajunen P, Pääkkönen R, Keskimäki I, Hämäläinen H, Rintanen H, Niemi M, Moltchanov V and Salomaa V. [Online database] National Cardiovascular Disease Register. Available from: http://www.ktl.fi/cvdr/. URN:NBN:fi-fe20041365 [cited 16.10.2007]

Lahiri DK and Schnabel B. DNA isolation by a rapid method from human blood samples: effects of MgCl2, EDTA, storage time, and temperature on DNA yield and quality. Biochem Genet 1993;31:321-28.

Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J and Salonen JT. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. JAMA 2002;288:2709-16.

Lander ES. The new genomics: global views of biology. Science 1996;274:536-39.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.

Lange LA, Carlson CS, Hindorff LA, Lange EM, Walston J, Durda JP, Cushman M, Bis JC, Zeng D, Lin D, et al. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. JAMA 2006;296:2703-11.

Larson MG, Atwood LD, Benjamin EJ, Cupples LA, D'Agostino RB, Sr., Fox CS, Govindaraju DR, Guo CY, Heard-Costa NL, Hwang SJ, et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. BMC Med Genet 2007;8 Suppl 1:S5.

Lee RT and Libby P. The unstable atheroma. Arterioscler Thromb Vasc Biol 1997;17:1859-67.

Lee WY, Allison MA, Kim DJ, Song CH and Barrett-Connor E. Association of interleukin-6 and C-reactive protein with subclinical carotid atherosclerosis (the Rancho Bernardo Study). Am J Cardiol 2007;99:99-102.

Leinonen M and Saikku P. Evidence for infectious agents in cardiovascular disease and atherosclerosis. The Lancet Infectious Diseases 2002;2:11-17.

Leinwand LA. Sex is a potent modifier of the cardiovascular system. J Clin Invest 2003;112:302-7.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, et al. The diploid genome sequence of an individual human. PLoS Biol 2007;5:e254.

Li C, Engstrom G, Hedblad B, Calling S, Berglund G and Janzon L. Sex differences in the relationships between BMI, WHR and incidence of cardiovascular disease: a population-based cohort study. Int J Obes (Lond) 2006;30:1775-81.

Li D, Saldeen T, Romeo F and Mehta JL. Oxidized LDL upregulates angiotensin II type 1 receptor expression in cultured human coronary artery endothelial cells: the potential role of transcription factor NF-kappaB. Circulation 2000;102:1970-76.

Libby P, Ridker PM and Maseri A. Inflammation and Atherosclerosis. Circulation 2002;105:1135-43.

Livingston EH and Cassidy L. Statistical power and estimation of the number of required subjects for a study based on the t-test: a surgeon's primer. J Surg Res 2005;126:149-59.

Lloyd-Jones DM, Larson MG, Beiser A and Levy D. Lifetime risk of developing coronary heart disease. Lancet 1999;353:89-92.

Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr., Levy D, Murabito JM, Wang TJ, Wilson PW and O'Donnell CJ. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA 2004;291:2204-11.

Lohmueller KE, Pearce CL, Pike M, Lander ES and Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003;33:177-82.

Lusis AJ. Genetic factors in cardiovascular disease. 10 questions. Trends Cardiovasc Med 2003;13:309-16.

Mach F, Schonbeck U, Sukhova GK, Bourcier T, Bonnefoy JY, Pober JS and Libby P. Functional CD40 ligand is expressed on human vascular endothelial cells, smooth muscle cells, and macrophages: implications for CD40-CD40 ligand signaling in atherosclerosis. Proc Natl Acad Sci U S A 1997;94:1931-36.

Madisen L, Hoar DI, Holroyd CD, Crisp M and Hodes ME. DNA banking: the effects of storage of blood and isolated DNA on the integrity of DNA. Am J Med Genet 1987;27:379-90.

Majors A, Ehrhart LA and Pezacka EH. Homocysteine as a risk factor for vascular disease. Enhanced collagen production and accumulation by smooth muscle cells. Arterioscler Thromb Vasc Biol 1997;17:2074-81.

Manly KF. Reliability of statistical associations between genes and disease. Immunogenetics 2005;57:549-58.

Mannila MN, Silveira A, Hawe E, Eriksson P, Aillaud MF, Juhan-Vague I, Yudkin J, Margaglione M, di Minno G, Mussoni L, et al. Plasma fibrinogen concentration predicts the risk of myocardial infarction differently in various parts of Europe: effects of beta-fibrinogen genotype and environmental factors. The HIFMECH Study. Thromb Haemost 2004;92:1240-49.

Mannila MN, Eriksson P, Lundman P, Samnegård A, Boquist S, Ericsson CG, Tornvall P, Hamsten A and Silveira A. Contribution of haplotypes across the fibrinogen gene cluster to variation in risk of myocardial infarction. Thromb Haemost 2005;93:570-77.

Mannila MN, Eriksson P, Ericsson CG, Hamsten A and Silveira A. Epistatic and pleiotropic effects of polymorphisms in the fibrinogen and coagulation factor XIII genes on plasma fibrinogen concentration, fibrin gel structure and risk of myocardial infarction. Thromb Haemost 2006;95:420-27.

Mannila MN, Lovely RS, Kazmierczak SC, Eriksson P, Samnegard A, Farrell DH, Hamsten A and Silveira A. Elevated plasma fibrinogen gamma' concentration is associated with myocardial infarction: effects of variation in fibrinogen genes and environmental factors. J Thromb Haemost 2007;5:766-73.

Manolio TA, Bailey-Wilson JE and Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet 2006;7:812-20.

Mathers CD and Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006;3:e442.

Mayer B, Erdmann J and Schunkert H. Genetics and heritability of coronary artery disease and myocardial infarction. Clin Res Cardiol 2007;96:1-7.

McCarthy JJ, Meyer J, Moliterno DJ, Newby LK, Rogers WJ and Topol EJ. Evidence for substantial effect modification by gender in a large-scale genetic association study of the metabolic syndrome among coronary heart disease patients. Hum Genet 2003;114:87-98.

McCarthy JJ. Gene by sex interaction in the etiology of coronary heart disease and the preceding metabolic syndrome. Nutr Metab Cardiovasc Dis 2007;17:153-61.

McGill HC, Jr., McMahan CA, Herderick EE, Malcom GT, Tracy RE and Strong JP. Origin of atherosclerosis in childhood and adolescence. Am J Clin Nutr 2000;72:1307S-15S.

McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, et al. A common allele on chromosome 9 associated with coronary heart disease. Science 2007;316:1488-91.

Mehrabian M, Allayee H, Wong J, Shi W, Wang XP, Shaposhnik Z, Funk CD and Lusis AJ. Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. Circ Res 2002;91:120-26.

Meier CR, Jick SS, Derby LE, Vasilakis C and Jick H. Acute respiratory-tract infections and risk of first-time acute myocardial infarction. Lancet 1998;351:1467-71.

Mendelsohn ME and Karas RH. Molecular and cellular basis of cardiovascular gender differences. Science 2005;308:1583-87.

Méplan C, Crosley LK, Nicol F, Beckett GJ, Howie AF, Hill KE, Horgan G, Mathers JC, Arthur JR and Hesketh JE. Genetic polymorphisms in the human selenoprotein P gene determine the response of selenoprotein markers to selenium supplementation in a gender-specific manner (the SELGEN study). FASEB J 2007;21:3063-74.

Miettinen OS. The "case-control" study: valid selection of subjects. J Chronic Dis 1985;38:543-48.

Miller DT, Zee RY, Suk Danik J, Kozlowski P, Chasman DI, Lazarus R, Cook NR, Ridker PM and Kwiatkowski DJ. Association of common CRP gene variants with CRP levels and cardiovascular events. Ann Hum Genet 2005;69:623-38.

Miller SA, Dykes DD and Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.

Mittendorfer B. Insulin resistance: sex matters. Curr Opin Clin Nutr Metab Care 2005a;8:367-72.

Mittendorfer B. Sexual dimorphism in human lipid metabolism. J Nutr 2005b;135:681-86.

Montgomery GW, Campbell MJ, Dickson P, Herbert S, Siemering K, Ewen-White KR, Visscher PM and Martin NG. Estimation of the rate of SNP genotyping errors from DNA extracted from different tissues. Twin Res Hum Genet 2005;8:346-52.

Mosca L, Manson JE, Sutherland SE, Langer RD, Manolio T and Barrett-Connor E. Cardiovascular disease in women: a statement for healthcare professionals from the American Heart Association. Writing Group. Circulation 1997;96:2468-82.

Murabito JM, Pencina MJ, Nam BH, D'Agostino RB, Sr., Wang TJ, Lloyd-Jones D, Wilson PW and O'Donnell CJ. Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. JAMA 2005;294:3117-23.

Myint PK, Luben RN, Welch AA, Bingham SA, Wareham NJ and Khaw KT. Effect of age on the relationship of occupational social class with prevalence of modifiable cardiovascular risk factors and cardiovascular diseases. A population-based cross-sectional study from European Prospective Investigation into Cancer - Norfolk (EPIC-Norfolk). Gerontology 2006;52:51-58.

Müller-Myhsok B and Abel L. Genetic analysis of complex diseases. Science 1997;275:1328-29.

Mälarstig A, Wallentin L and Siegbahn A. Genetic variation in the interleukin-6 gene in relation to risk and outcomes in acute coronary syndrome. Thromb Res 2007;119:467-73.

Nabata A, Kuroki M, Ueba H, Hashimoto S, Umemoto T, Wada H, Yasu T, Saito M, Momomura SI and Kawakami M. C-reactive protein induces endothelial cell apoptosis and matrix metalloproteinase-9 production in human mononuclear cells: Implications for the destabilization of atherosclerotic plaque. Atherosclerosis 2007;196:129-35.

Nabel EG. Cardiovascular Disease. N Engl J Med 2003;349:60-72.

Nanchahal K, Morris JN, Sullivan LM and Wilson PW. Coronary heart disease risk in men and the epidemic of overweight and obesity. Int J Obes (Lond) 2005;29:317-23.

NCEP. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA 2001;285:2486-97.

NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. Nature 2007;447:655-60.

Nederhand RJ, Droog S, Kluft C, Simoons ML and de Maat MP. Logistics and quality control for DNA sampling in large multicenter studies. J Thromb Haemost 2003;1:987-91.

Neel JV. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? Am J Hum Genet 1962;14:353-62.

Ng SB, Tan YH and Guy GR. Differential induction of the interleukin-6 gene by tumor necrosis factor and interleukin-1. J Biol Chem 1994;269:19021-27.

Nilsson-Ardnor S, Janunger T, Wiklund PG, Lackovic K, Nilsson AK, Lindgren P, Escher SA, Stegmayr B, Asplund K and Holmberg D. Genome-wide linkage scan of common stroke in families from northern Sweden. Stroke 2007;38:34-40.

Nilsson J. CRP--marker or maker of cardiovascular disease? Arterioscler Thromb Vasc Biol 2005;25:1527-28.

Ninomiya JK, L'Italien G, Criqui MH, Whyte JL, Gamst A and Chen RS. Association of the metabolic syndrome with history of myocardial infarction and stroke in the Third National Health and Nutrition Examination Survey. Circulation 2004;109:42-46.

Nozawa M, Kawahara Y and Nei M. Genomic drift and copy number variation of sensory receptor genes in humans. 2007;104:20412-26.

Ntzani EE, Rizos EC and Ioannidis JP. Genetic effects versus bias for candidate polymorphisms in myocardial infarction: case study and overview of large-scale evidence. Am J Epidemiol 2007;165:973-84.

Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 2004;74:765-69.

Okano M and Yoshida Y. Influence of shear stress on endothelial cell shapes and junction complexes at flow dividers of aortic bifurcations in cholesterol-fed rabbits. Front Med Biol Eng 1993;5:95-120.

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 2002;32:650-54.

Ozaki K and Tanaka T. Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses. Cell Mol Life Sci 2005;62:1804-13.

Pahl HL. Activators and target genes of Rel/NF-kappaB transcription factors. Oncogene 1999;18:6853-66.

Paik JK, Kim OY, Koh SJ, Jang Y, Chae JS, Kim JY, Kim HJ, Hyun YJ, Cho JR and Lee JH. Additive effect of interleukin-6 and C-reactive protein (CRP) single nucleotide polymorphism on serum CRP concentration and other cardiovascular risk factors. Clin Chim Acta 2007;380:68-74.

Pajukanta P, Nuotio I, Terwilliger JD, Porkka KV, Ylitalo K, Pihlajamäki J, Suomalainen AJ, Syvänen AC, Lehtimäki T, Viikari JS, et al. Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. Nat Genet 1998;18:369-73.

Pajukanta P, Cargill M, Viitanen L, Nuotio I, Kareinen A, Perola M, Terwilliger JD, Kempas E, Daly M, Lilja H, et al. Two loci on chromosomes 2 and X for premature coronary heart disease identified in early- and late-settlement populations of Finland. Am J Hum Genet 2000;67:1481-93.

Pajukanta P, Lilja HE, Sinsheimer JS, Cantor RM, Lusis AJ, Gentile M, Duan XJ, Soro-Paavonen A, Naukkarinen J, Saarela J, et al. Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). Nat Genet 2004;36:371-76.

Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH and Tracy RP. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis 2001;154:681-89.

Paré G, Serre D, Brisson D, Anand SS, Montpetit A, Tremblay G, Engert JC, Hudson TJ and Gaudet D. Genetic analysis of 103 candidate genes for coronary artery disease and associated phenotypes in a founder population reveals a new association between endothelin-1 and high-density lipoprotein cholesterol. Am J Hum Genet 2007;80:673-82.

Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L and Syvanen A-C. A System for Specific, High-throughput Genotyping by Allele-specific Primer Extension on Microarrays. Genome Res 2000;10:1031-42.

Pastinen T, Perola M, Ignatius J, Sabatti C, Tainola P, Levander M, Syvanen AC and Peltonen L. Dissecting a population genome for targeted screening of disease mutations. Hum Mol Genet 2001;10:2961-72.

Paynter RA, Skibola DR, Skibola CF, Buffler PA, Wiemels JL and Smith MT. Accuracy of multiplexed Illumina platform-based single-nucleotide polymorphism genotyping compared between genomic and whole genome amplified DNA collected from multiple sources. Cancer Epidemiol Biomarkers Prev 2006;15:2533-36.

Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, 3rd, Criqui M, Fadl YY, Fortmann SP, Hong Y, Myers GL, et al. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation 2003;107:499-511.

Peltonen L, Jalanko A and Varilo T. Molecular genetics of the Finnish disease heritage. Hum Mol Genet 1999;8:1913-23.

Peltonen L, Palotie A and Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet 2000;1:182-90.

Peng B and Kimmel M. Simulations provide support for the common disease-common variant hypothesis. Genetics 2007;175:763-76.

Pilia G, Chen WM, Scuteri A, Orrú M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet 2006;2:e132.

Pilote L, Dasgupta K, Guru V, Humphries KH, McGrath J, Norris C, Rabi D, Tremblay J, Alamian A, Barnett T, et al. A comprehensive view of sex-specific issues related to cardiovascular disease. CMAJ 2007;176:S1-44.

Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986;73:1-11.

Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69:124-37.

Pritchard JK and Przeworski M. Linkage disequilibrium in humans: models and data. Am J Hum Genet 2001;69:1-14.

Pritchard JK and Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet 2002;11:2417-23.

Pueyo ME, Gonzalez W, Nicoletti A, Savoie F, Arnal JF and Michel JB. Angiotensin II stimulates endothelial vascular cell adhesion molecule-1 via nuclear factor-kappaB activation induced by intracellular oxidative stress. Arterioscler Thromb Vasc Biol 2000;20:645-51.

Puffenberger EG, Hosoda K, Washington SS, Nakao K, deWit D, Yanagisawa M and Chakravart A. A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. Cell 1994a;79:1257-66.

Puffenberger EG, Kauffman ER, Bolk S, Matise TC, Washington SS, Angrist M, Weissenbach J, Garver KL, Mascari M, Ladda R, et al. Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. Hum Mol Genet 1994b;3:1217-25.

R Development Core Team. R: A language and environment for statistical computing. Available from: http://www.et.bs.ehu.es/soft/fullrefman.pdf. [cited 18.12.2007]

Rajecki M, Pajunen P, Jousilahti P, Rasi V, Vahtera E and Salomaa V. Hemostatic factors as predictors of stroke and cardiovascular diseases: the FINRISK '92 Hemostasis Study. Blood Coagul Fibrinolysis 2005;16:119-24.

Rana JS, Boekholdt SM, Ridker PM, Jukema JW, Luben R, Bingham SA, Day NE, Wareham NJ, Kastelein JJ and Khaw KT. Differential leucocyte count and the risk of future coronary artery disease in healthy men and women: the EPIC-Norfolk Prospective Population Study. J Intern Med 2007;262:678-89.

Rauramaa R, Vaisanen SB, Luong LA, Schmidt-Trucksass A, Penttila IM, Bouchard C, Toyry J and Humphries SE. Stromelysin-1 and interleukin-6 gene promoter polymorphisms are determinants of asymptomatic carotid artery atherosclerosis. Arterioscler Thromb Vasc Biol 2000;20:2657-62.

Reddy ST, Wadleigh DJ, Grijalva V, Ng C, Hama S, Gangopadhyay A, Shih DM, Lusis AJ, Navab M and Fogelman AM. Human paraoxonase-3 is an HDL-associated enzyme with biological activity similar to paraoxonase-1 protein but is not regulated by oxidized lipids. Arterioscler Thromb Vasc Biol 2001;21:542-47.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al. Global variation in copy number in the human genome. Nature 2006;444:444-54.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, et al. Linkage disequilibrium in the human genome. Nature 2001;411:199-204.

Reich DE and Lander ES. On the allelic spectrum of human disease. Trends Genet 2001;17:502-10.

Reifenberg K, Lehr HA, Baskal D, Wiese E, Schaefer SC, Black S, Samols D, Torzewski M, Lackner KJ, Husmann M, et al. Role of C-reactive protein in atherogenesis: can the apolipoprotein E knockout mouse provide the answer? Arterioscler Thromb Vasc Biol 2005;25:1641-46.

Richardson AJ, Narendran N, Guymer RH, Vu H and Baird PN. Blood storage at 4 degrees C-factors involved in DNA yield and quality. J Lab Clin Med 2006;147:290-94.

Ridker PM, Rifai N, Stampfer MJ and Hennekens CH. Plasma Concentration of Interleukin-6 and the Risk of Future Myocardial Infarction Among Apparently Healthy Men. Circulation 2000;101:1767-72.

Riese C, Michaelis M, Mentrup B, Gotz F, Kohrle J, Schweizer U and Schomburg L. Selenium-dependent pre- and posttranscriptional mechanisms are responsible for sexual dimorphic expression of selenoproteins in murine tissues. Endocrinology 2006:en.2006-0689.

Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P and Soubrier F. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. J Clin Invest 1990;86:1343-46.

Rinn JL and Snyder M. Sexual dimorphism in mammalian gene expression. Trends Genet 2005;21:298-305.

Risch N and Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516-17.

Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847-56.

Rissanen AM. Familial aggregation of coronary heart disease in a high incidence area (North Karelia, Finland). Br Heart J 1979a;42:294-303.

Rissanen AM. Familial occurrence of coronary heart disease: effect of age at diagnosis. Am J Cardiol 1979b;44:60-66.

Rivera-Chavez FA, Peters-Hybki DL, Barber RC and O'Keefe GE. Interleukin-6 promoter haplotypes and interleukin-6 cytokine responses. Shock 2003;20:218-23.

Robertson TL, Kato H, Rhoads GG, Kagan A, Marmot M, Syme SL, Gordon T, Worth RM, Belsky JL, Dock DS, et al. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California. Incidence of myocardial infarction and death from coronary heart disease. Am J Cardiol 1977;39:239-43.

Romano M, Sironi M, Toniatti C, Polentarutti N, Fruscella P, Ghezzi P, Faggioni R, Luini W, van Hinsbergh V, Sozzani S, et al. Role of IL-6 and its soluble receptor in induction of chemokines and leukocyte recruitment. Immunity 1997;6:315-25.

Rosand J, Bayley N, Rost N and de Bakker PI. Many hypotheses but no replication for the association between PDE4D and stroke. Nat Genet 2006;38:1091-92.

Rosner SA, Ridker PM, Zee RY and Cook NR. Interaction between inflammation-related gene polymorphisms and cigarette smoking on the risk of myocardial infarction in the Physician's Health Study. Hum Genet 2005;118:287-94.

Ross R and Glomset JA. Atherosclerosis and the arterial smooth muscle cell: Proliferation of smooth muscle is a key event in the genesis of the lesions of atherosclerosis. Science 1973;180:1332-39.

Ross R and Glomset JA. The pathogenesis of atherosclerosis (first of two parts). N Engl J Med 1976a;295:369-77.

Ross R and Glomset JA. The pathogenesis of atherosclerosis (second of two parts). N Engl J Med 1976b;295:420-25.

Ross R. Atherosclerosis -- An Inflammatory Disease. N Engl J Med 1999;340:115-26.

Roy SN, Mukhopadhyay G and Redman CM. Regulation of fibrinogen assembly. Transfection of Hep G2 cells with B beta cDNA specifically enhances synthesis of the three component chains of fibrinogen. J Biol Chem 1990;265:6389-93.

Sackett DL. Bias in analytic research. J Chronic Dis 1979;32:51-63.

Salisbury BA, Pungliya M, Choi JY, Jiang R, Sun XJ and Stephens JC. SNP and haplotype variation in the human genome. Mutat Res 2003;526:53-61.

Salomaa V, Miettinen H, Kuulasmaa K, Niemelä M, Ketonen M, Vuorenmaa T, Lehto S, Palomäki P, Mähönen M, Immonen-Räihä P, et al. Decline of Coronary Heart Disease Mortality in Finland During 1983 to 1992: Roles of Incidence, Recurrence, and Case-Fatality: The FINMONICA MI Register Study. Circulation 1996;94:3130-37.

Salomaa V, Ketonen M, Koukkunen H, Immonen-Räihä P, Jerkkola T, Kärjä-Koskenkari P, Mähönen M, Niemelä M, Kuulasmaa K, Palomäki P, et al. Trends in coronary events in Finland during 1983-1997. The FINAMI study. Eur Heart J 2003;24:311-19.

Salomaa V, Koukkunen H, Ketonen M, Immonen-Räihä P, Kärjä-Koskenkari P, Mustonen J, Lehto S, Torppa J, Lehtonen A, Tuomilehto J, et al. A new definition for myocardial infarction: what difference does it make? Eur Heart J 2005;26:1719-25.

Samani NJ, Burton P, Mangino M, Ball SG, Balmforth AJ, Barrett J, Bishop T and Hall A. A genomewide linkage study of 1,933 families affected by premature coronary artery disease: The British Heart Foundation (BHF) Family Heart Study. Am J Hum Genet 2005;77:1011-20.

Sanceau J, Kaisho T, Hirano T and Wietzerbin J. Triggering of the human interleukin-6 gene by interferon-gamma and tumor necrosis factor-alpha in monocytic cells involves cooperation between interferon regulatory factor-1, NF kappa B, and Sp1 transcription factors. J Biol Chem 1995;270:27920-31.

Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH, Song K, Yuan X, Johnson T, Ashford S, et al. LDL-cholesterol concentrations: a genome-wide association study. Lancet 2008;371:483-91.

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316:1331-36.

Schieffer B, Schieffer E, Hilfiker-Kleiner D, Hilfiker A, Kovanen PT, Kaartinen M, Nussberger J, Harringer W and Drexler H. Expression of angiotensin II and interleukin 6 in human coronary atherosclerotic plaques: potential implications for inflammation and plaque instability. Circulation 2000;101:1372-78.

Schieffer B, Selle T, Hilfiker A, Hilfiker-Kleiner D, Grote K, Tietge UJ, Trautwein C, Luchtefeld M, Schmittkamp C, Heeneman S, et al. Impact of interleukin-6 on plaque development and morphology in experimental atherosclerosis. Circulation 2004;110:3493-500.

Schousboe K, Willemsen G, Kyvik KO, Mortensen J, Boomsma DI, Cornes BK, Davis CJ, Fagnani C, Hjelmborg J, Kaprio J, et al. Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. Twin Res 2003;6:409-21.

Schreyer SA, Vick CM and LeBoeuf RC. Loss of lymphotoxin-alpha but not tumor necrosis factor-alpha reduces atherosclerosis in mice. J Biol Chem 2002;277:12364-68.

Schönbeck U, Sukhova GK, Shimizu K, Mach F and Libby P. Inhibition of CD40 signaling limits evolution of established atherosclerosis in mice. Proc Natl Acad Sci U S A 2000;97:7458-63.

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 2007;316:1341-45.

SeattleSNPs. [Online database] NHLBI Program for Genomic Applications, SeattleSNPs, Seattle, WA. Available from: http://pga.gs.washington.edu. [cited 13.12.2007]

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat Genet 2006;38:556-60.

Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Au R, Kannel WB and Wolf PA. The lifetime risk of stroke: estimates from the Framingham Study. Stroke 2006;37:345-50.

Setakis E, Stirnadel H and Balding DJ. Logistic regression protects against population structure in genetic association studies. Genome Res 2006;16:290-96.

Shah PK and Galis ZS. Matrix metalloproteinase hypothesis of plaque rupture: players keep piling up but questions remain. Circulation 2001;104:1878-80.

Shifman S, Kuypers J, Kokoris M, Yakir B and Darvasi A. Linkage disequilibrium patterns of the human genome across populations. Hum Mol Genet 2003;12:771-76.

Sie MP, Sayed-Tabatabaei FA, Oei HH, Uitterlinden AG, Pols HA, Hofman A, van Duijn CM and Witteman JC. Interleukin 6 -174 g/c promoter polymorphism and risk of coronary heart

disease: results from the rotterdam study and a meta-analysis. Arterioscler Thromb Vasc Biol 2006;26:212-17.

Silander K, Komulainen K, Ellonen P, Jussila M, Alanne M, Levander M, Tainola P, Kuulasmaa K, Salomaa V, Perola M, et al. Evaluating whole genome amplification via multiply-primed rolling circle amplification for SNP genotyping of samples with low DNA yield. Twin Res Hum Genet 2005;8:368-75.

Sivenius J, Tuomilehto J, Immonen-Raiha P, Kaarisalo M, Sarti C, Torppa J, Kuulasmaa K, Mahonen M, Lehtonen A and Salomaa V. Continuous 15-year decrease in incidence and mortality of stroke in Finland: the FINSTROKE study. Stroke 2004;35:420-25.

Skurk T, van Harmelen V and Hauner H. Angiotensin II stimulates the release of interleukin-6 and interleukin-8 from cultured human adipocytes by activation of NF-kappaB. Arterioscler Thromb Vasc Biol 2004;24:1199-203.

Slack J and Evans KA. The increased risk of death from ischaemic heart disease in first degree relatives of 121 men and 96 women with ischaemic heart disease. J Med Genet 1966;3:239-57.

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881-85.

Smith GD, Harbord R, Milton J, Ebrahim S and Sterne JAC. Does Elevated Plasma Fibrinogen Increase the Risk of Coronary Heart Disease?: Evidence from a Meta-Analysis of Genetic Association Studies. Arterioscler Thromb Vasc Biol 2005;25:2228-33.

Smith SC, Jr. Multiple risk factors for cardiovascular disease and diabetes mellitus. Am J Med 2007;120:S3-S11.

Spahr A, Klein E, Khuseyinova N, Boeckh C, Muche R, Kunze M, Rothenbacher D, Pezeshki G, Hoffmeister A and Koenig W. Periodontal infections and coronary heart disease: role of periodontal bacteria and importance of total pathogen burden in the Coronary Event and Periodontal Disease (CORODONT) study. Arch Intern Med 2006;166:554-59.

Spanbroek R, Grabner R, Lötzer K, Hildner M, Urbach A, Rühling K, Moos MP, Kaiser B, Cohnert TU, Wahlers T, et al. Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. Proc Natl Acad Sci U S A 2003;100:1238-43.

Spronk HM, van der Voort D and Ten Cate H. Blood coagulation and the risk of atherothrombosis: a complex relationship. Thromb J 2004;2:12.

Statistics Finland. [Online database] Kuolemansyytilasto. Available from: http://pxweb2.stat.fi/Database/StatFin/ter/ksyyt/ksyyt_fi.asp. [cited 10.10.2007]

Steemers FJ and Gunderson KL. Whole genome genotyping technologies on the BeadArray platform. Biotechnol J 2007;2:41-49.

Steinberg K, Beck J, Nickerson D, Garcia-Closas M, Gallagher M, Caggana M, Reid Y, Cosentino M, Ji J, Johnson D, et al. DNA banking for epidemiologic studies: a review of current practices. Epidemiology 2002;13:246-54.

Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007;39:770-75.

Stephens M, Smith NJ and Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978-89.

Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE and Pike MC. Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. Hum Hered 2003;55:27-36.

Strimmer K. The fdrtool Package version 1.1.4. 2007

Su S, Chen J, Zhao J, Huang J, Wang X, Chen R and Gu D. Angiotensin II type I receptor gene and myocardial infarction: tagging SNPs and haplotype based association study. The Beijing atherosclerosis study. Pharmacogenetics 2004;14:673-81.

Sunnemark D, Harris RA, Frostegard J and Orn A. Induction of early atherosclerosis in CBA/J mice by combination of Trypanosoma cruzi infection and a high cholesterol diet. Atherosclerosis 2000;153:273-82.

Syvänen AC. Toward genome-wide SNP genotyping. Nat Genet 2005;37 Suppl:S5-10.

Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A, Zakharkin SO, George V, Allison DB, Cooper GS, et al. Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. J Mol Med 2005;83:440-47.

Tabas I. Nonoxidative modifications of lipoproteins in atherogenesis. Annu Rev Nutr 1999;19:123-39.

Tabor HK, Risch NJ and Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002;3:391-97.

Terry CF, Loukaci V and Green FR. Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. J Biol Chem 2000;275:18138-44.

Terwilliger JD and Ott J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. Hum Hered 1992;42:337-46.

Terwilliger JD and Göring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 2000;72:63-132.

Terwilliger JD, Haghighi F, Hiekkalinna TS and Göring HH. A bias-ed assessment of the use of SNPs in human complex traits. Curr Opin Genet Dev 2002;12:726-34.

Terwilliger JD and Weiss KM. Confounding, ascertainment bias, and the blind quest for a genetic 'fountain of youth'. Ann Med 2003;35:532-44.

Terwilliger JD and Hiekkalinna T. An utter refutation of the "Fundamental Theorem of the HapMap". Eur J Hum Genet 2006;14:426-37.

Thakore AH, Guo C-Y, Larson MG, Corey D, Wang TJ, Vasan RS, D'Agostino Sr RB, Lipinska I, Keaney Jr JF, Benjamin EJ, et al. Association of Multiple Inflammatory Markers with Carotid Intimal Medial Thickness and Stenosis (from the Framingham Heart Study). The American Journal of Cardiology 2007;99:1598-602.

The ESHRE Capri Workshop Group. Hormones and cardiovascular health in women. Hum Reprod Update 2006;12:483-97.

The International HapMap Consortium. The International HapMap Project. Nature 2003;426:789-96.

The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851-61.

The International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931-45.

The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.

Thomas A, Green F, Lamlum H and Humphries S. The association of combined alpha and beta fibrinogen genotype on plasma fibrinogen levels in smokers and non-smokers. J Med Genet 1995;32:585-89.

Thorleifsson G, Magnusson KP, Sulem P, Walters GB, Gudbjartsson DF, Stefansson H, Jonsson T, Jonasdottir A, Jonasdottir A, Stefansdottir G, et al. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. Science 2007;317:1397-400.

Tong W, Lai H, Yang C, Ren S, Dai S and Lai S. Age, gender and metabolic syndrome-related coronary heart disease in U.S. adults. Int J Cardiol 2005;104:288-91.

Tsai CT, Hwang JJ, Ritchie MD, Moore JH, Chiang FT, Lai LP, Hsu KL, Tseng CD, Lin JL and Tseng YZ. Renin-angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort: detection of high order gene-gene interaction. Atherosclerosis 2007;195:172-80.

Tsai JC, Perrella MA, Yoshizumi M, Hsieh CM, Haber E, Schlegel R and Lee ME. Promotion of vascular smooth muscle cell growth by homocysteine: a link to atherosclerosis. Proc Natl Acad Sci U S A 1994;91:6369-73.

Tunstall-Pedoe H, editor. MONICA monograph and Multimedia Sourcebook. Geneva: World Health Organization; 2003.

Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM and Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. Circulation 1994;90:583-612.

Tuomisto K, Jousilahti P, Sundvall J, Pajunen P and Salomaa V. C-reactive protein, interleukin-6 and tumor necrosis factor alpha as predictors of incident coronary and cardiovascular events and total mortality. A population-based, prospective study. Thromb Haemost 2006;95:511-18.

Tzoulaki I, Murray GD, Lee AJ, Rumley A, Lowe GD and Fowkes FG. Relative value of inflammatory, hemostatic, and rheological factors for incident myocardial infarction and stroke: the Edinburgh Artery Study. Circulation 2007;115:2119-27.

Tzvetkov MV, Becker C, Kulle B, Nürnberg P, Brockmöller J and Wojnowski L. Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification. Electrophoresis 2005;26:710-15.

Uitte de Willige S, Doggen CJM, de Visser MCH, Bertina RM and Rosendaal FR. Haplotypes of the fibrinogen gamma gene do not affect the risk of myocardial infarction. J Thromb Haemost 2006;4:474-76.

Undar L, Ertugrul C, Altunbas H and Akca S. Circadian variations in natural coagulation inhibitors protein C, protein S and antithrombin in healthy men: a possible association with interleukin-6. Thromb Haemost 1999;81:571-75.

van den Oord EJ and Sullivan PF. False discoveries and models for gene discovery. Trends Genet 2003;19:537-42.

van Deventer SJ, Buller HR, ten Cate JW, Aarden LA, Hack CE and Sturk A. Experimental endotoxemia in humans: analysis of cytokine release and coagulation, fibrinolytic, and complement pathways. Blood 1990;76:2520-26.

van Kleef EM, Fingerle J and Daemen MJ. Angiotensin II-induced progression of neointimal thickening in the balloon-injured rat carotid artery is AT1 receptor mediated. Arterioscler Thromb Vasc Biol 1996;16:857-63.

Vandenplas S, Wiid I, Grobler-Rabie A, Brebner K, Ricketts M, Wallis G, Bester A, Boyd C and Mathew C. Blot hybridisation analysis of genomic DNA. J Med Genet 1984;21:164-72.

Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD and Peltonen L. Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. Eur J Hum Genet 2000;8:604-12.

Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD and Peltonen L. The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. Hum Mol Genet 2003;12:51-59.

Varilo T and Peltonen L. Isolates and their potential use in complex gene mapping efforts. Curr Opin Genet Dev 2004;14:316-23.

Vartiainen E, Jousilahti P, Alfthan G, Sundvall J, Pietinen P and Puska P. Cardiovascular risk factor changes in Finland 1972-1997. Int J Epidemiol 2000;29:49-56.

Véniant MM, Withycombe S and Young SG. Lipoprotein size and atherosclerosis susceptibility in Apoe(-/-) and Ldlr(-/-) mice. Arterioscler Thromb Vasc Biol 2001;21:1567-70.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. Science 2001;291:1304-51.

Vickers MA, Green FR, Terry C, Mayosi BM, Julier C, Lathrop M, Ratcliffe PJ, Watkins HC and Keavney B. Genotype at a promoter polymorphism of the interleukin-6 gene is associated with baseline levels of plasma C-reactive protein. Cardiovasc Res 2002;53:1029-34.

Visvikis S, Schlenck A and Maurice M. DNA extraction and stability for epidemiological studies. Clin Chem Lab Med 1998;36:551-55.

Vlodaver Z, Kahn HA and Neufeld HN. The coronary arteries in early life in three different ethnic groups. Circulation 1969;39:541-50.

Wacholder S, Silverman DT, McLaughlin JK and Mandel JS. Selection of controls in case-control studies. III. Design options. Am J Epidemiol 1992;135:1042-50.

Wallace C, Newhouse SJ, Braund P, Zhang F, Tobin M, Falchi M, Ahmadi K, Dobson RJ, Marcano AC, Hajat C, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. Am J Hum Genet 2008;82:139-49.

Walston JD, Fallin MD, Cushman M, Lange L, Psaty B, Jenny N, Browner W, Tracy R, Durda P and Reiner A. IL-6 gene variation is associated with IL-6 and C-reactive protein levels but not cardiovascular outcomes in the Cardiovascular Health Study. Hum Genet 2007;122:485-94.

Wang L, Fan C, Topol SE, Topol EJ and Wang Q. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. Science 2003;302:1578-81.

Wang L, Hauser ER, Shah SH, Pericak-Vance MA, Haynes C, Crosslin D, Harris M, Nelson S, Hale AB, Granger CB, et al. Peakwide mapping on chromosome 3q13 identifies the kalirin gene as a novel candidate gene for coronary artery disease. Am J Hum Genet 2007;80:650-63.

Wang Q, Rao S, Shen GQ, Li L, Moliterno DJ, Newby LK, Rogers WJ, Cannata R, Zirzow E, Elston RC, et al. Premature myocardial infarction novel susceptibility locus on chromosome 1P34-36 identified by genomewide linkage analysis. Am J Hum Genet 2004;74:262-71.

Wassmann S, Stumpf M, Strehlow K, Schmid A, Schieffer B, Böhm M and Nickenig G. Interleukin-6 induces oxidative stress and endothelial dysfunction by overexpression of the angiotensin II type 1 receptor. Circ Res 2004;94:534-41.

Weiss KM and Terwilliger JD. How many diseases does it take to map a gene with SNPs? Nat Genet 2000;26:151-57.

Weiss KM and Clark AG. Linkage disequilibrium and the mapping of complex human traits. Trends Genet 2002;18:19-24.

Weiss LA, Abney M, Parry R, Scanu AM, Cook EH, Jr. and Ober C. Variation in ITGB3 has sex-specific associations with plasma lipoprotein(a) and whole blood serotonin levels in a population-based sample. Hum Genet 2005;117:81-87.

Weiss LA, Pan L, Abney M and Ober C. The sex-specific genetic architecture of quantitative traits in humans. Nat Genet 2006;38:218-22.

Wheeler JG, Mussolino ME, Gillum RF and Danesh J. Associations between differential leucocyte count and incident coronary heart disease: 1764 incident cases from seven prospective studies of 30,374 individuals. Eur Heart J 2004;25:1287-92.

Willemsen JM, Westerink JW, Dallinga-Thie GM, van Zonneveld AJ, Gaillard CA, Rabelink TJ and de Koning EJ. Angiotensin II type 1 receptor blockade improves hyperglycemia-induced endothelial dysfunction and reduces proinflammatory cytokine release from leukocytes. J Cardiovasc Pharmacol 2007;49:6-12.

Willer CJ, Scott LJ, Bonnycastle LL, Jackson AU, Chines P, Pruim R, Bark CW, Tsai YY, Pugh EW, Doheny KF, et al. Tag SNP selection for Finnish individuals based on the CEPH Utah HapMap database. Genet Epidemiol 2006;30:180-90.

Wong LY, Leung RY, Ong KL and Cheung BM. Plasma levels of fibrinogen and C-reactive protein are related to interleukin-6 gene -572C>G polymorphism in subjects with and without hypertension. J Hum Hypertens 2007;21:875-82.

World Health Organization. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Report of a WHO Consultation. WHO/NCD/NCS/99.2. Geneva: World Health Organization; 1999.

World Health Organization. International Statistical Classification of Diseases and Health Related Problems. 2nd ed. Geneva: World Health Organization; 2005.

World Health Organization. World health statistics 2006. Available from: http://www.who.int/whosis/whostat2006.pdf. [cited 8.10.2007]

Wright AF and Hastie ND. Complex genetic diseases: controversy over the Croesus code. Genome Biol 2001;2:COMMENT2007.1-07.8.

Yamada Y, Izawa H, Ichihara S, Takatsu F, Ishihara H, Hirayama H, Sone T, Tanaka M and Yokota M. Prediction of the risk of myocardial infarction from polymorphisms in candidate genes. N Engl J Med 2002;347:1916-23.

Yarnell JW, Baker IA, Sweetnam PM, Bainton D, O'Brien JR, Whitehead PJ and Elwood PC. Fibrinogen, viscosity, and white blood cell count are major risk factors for ischemic heart disease. The Caerphilly and Speedwell collaborative heart disease studies. Circulation 1991;83:836-44.

Ye S, Dhillon S, Seear R, Dunleavey L, Day LB, Bannister W, Day IN and Simpson I. Epistatic interaction between variations in the angiotensin I converting enzyme and angiotensin II type 1 receptor genes in relation to extent of coronary atherosclerosis. Heart 2003;89:1195-99.

Ye Y, Shibata Y, Yun C, Ron D and Rapoport TA. A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. Nature 2004;429:841-47.

Yoshida Y, Okano M, Wang S, Kobayashi M, Kawasumi M, Hagiwara H and Mitsumata M. Hemodynamic-force-induced difference of interendothelial junctional complexes. Ann N Y Acad Sci 1995;748:104-20; discussion 20-1.

Yoza B, Matsumoto M and Matsunaga T. DNA extraction using modified bacterial magnetic particles in the presence of amino silane compound. J Biotechnol 2002;94:217-24.

Zand T, Hoffman AH, Savilonis BJ, Underwood JM, Nunnari JJ, Majno G and Joris I. Lipid deposition in rat aortas with intraluminal hemispherical plug stenosis. A morphological and biophysical study. Am J Pathol 1999;155:85-92.

Zavaroni I, Ardigo D, Massironi P, Gasparini P, Barilli AL, Vetrugno E, Sartori E, Baroni MC, Delsignore R and Reaven G. Do coronary heart disease risk factors change over time? Metabolism 2002;51:1022-26.

Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI and De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J Intern Med 2002;252:247-54.

Zee RY, Cook NR, Cheng S, Erlich HA, Lindpaintner K and Ridker PM. Multi-locus candidate gene polymorphisms and risk of myocardial infarction: a population-based, prospective genetic analysis. J Thromb Haemost 2006;4:341-48.

Zhang K and Jin L. HaploBlockFinder: haplotype block analyses. Bioinformatics 2003;19:1300-1.

Zhou X, Nicoletti A, Elhage R and Hansson GK. Transfer of CD4(+) T cells aggravates atherosclerosis in immunodeficient apolipoprotein E knockout mice. Circulation 2000;102:2919-22.

Zhu J, Quyyumi AA, Norman JE, Csako G, Waclawiw MA, Shearer GM and Epstein SE. Effects of total pathogen burden on coronary artery disease risk and C-reactive protein levels. Am J Cardiol 2000;85:140-46.

Zhu J, Nieto FJ, Horne BD, Anderson JL, Muhlestein JB and Epstein SE. Prospective study of pathogen burden and risk of myocardial infarction or death. Circulation 2001;103:45-51.

Zondervan KT and Cardon LR. The complex interplay among factors that influence allelic association. Nat Rev Genet 2004;5:89-100.