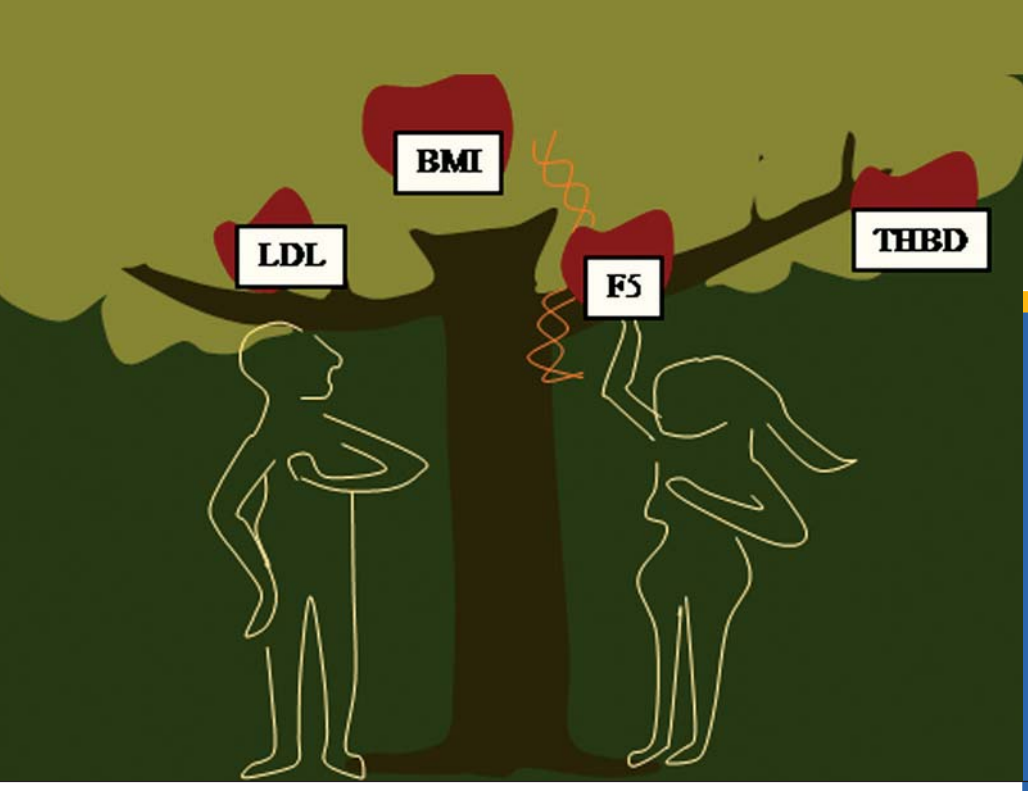


Kirsi Auro

Candidate Gene Studies on Cardiovascular Traits

Publications of the National Public Health Institute A 18/2007

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

Helsinki 2007

Kirsi Auro

CANDIDATE GENE STUDIES ON CARDIOVASCULAR TRAITS

ACADEMIC DISSERTATION

To be presented with the permission of the Medical Faculty, University of Helsinki, for public examination in lecture hall 3, Biomedicum, on December14th, at 12 noon.

National Public Health Institute, Helsinki, Finland

and

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2007

Publications of the National Public Health Institute
KTL A18 / 2007

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 978-951-740-752-6
ISSN 0359-3584
ISBN 978-951-740-753-3 (pdf)
ISSN 1458-6290 (pdf)

Edita Prima Oy
Helsinki 2007

Supervised by

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

Academy Professor Leena Peltonen-Palotie
National Public Health Institute
Department of Molecular Medicine,
University of Helsinki
Department of Medical Genetics
Finland and,
The Sanger Institute
Cambridge, UK

Reviewed by

Professor Tomi-Pekka Tuomainen
School of Public Health and Clinical Nutrition
University of Kuopio
Finland

Docent Jukka Lehtonen
Department of Cardiology
Helsinki University Hospital and
University of Helsinki
Finland

Opponent

Docent Riitta Lassila
Department of Hematology
Helsinki University Hospital and
University of Helsinki
Finland

Kirsi Auro, Candidate gene studies on cardiovascular traits
Publications of the National Public Health Institute, A18/2007, 110 Pages
ISBN 978-951-740-752-6; 978-951-740-753-3 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/4043

ABSTRACT

Cardiovascular diseases (CVD) are a major cause of death and disability in Western countries and a growing health problem in the developing world. The genetic component of both coronary heart disease (CHD) and ischemic stroke events has been established in twin studies, and the traits predisposing to CVD, such as hypertension, dyslipidemias, obesity, diabetes, and smoking behavior, are all partly hereditary.

This thesis aimed to clarify the genetic background of CVD at a population level using large Nordic population cohorts and a candidate gene approach. The first study concentrated on the allelic diversity of the thrombomodulin (*THBD*) gene in two Finnish cohorts, FINRISK-92 and FINRISK-97. The results from this study implied that *THBD* variants do not substantially contribute to CVD risk.

In the second study, three other candidate genes were added to the analyses. The study investigated the epistatic effects of coagulation factor V (*F5*), intercellular adhesion molecuce -1 (*ICAM1*), protein C (*PROC*), and *THBD* in the same FINRISK cohorts. The results were encouraging; we were able to identify several single SNPs and SNP combinations associating with CVD and mortality. Interestingly, *THBD* variants appeared in the associating SNP combinations despite the negative results from Study I, suggesting that *THBD* contributes to CVD through gene-gene interactions.

In the third study, upstream transcription factor -1 (*USF1*) was analyzed in a cohort of Swedish men. *USF1* was associated with metabolic syndrome, characterized by accumulation of different CVD risk factors. A putative protective and a putative risk variant were identified. A direct association with CVD was not observed. The longitudinal nature of the study also clarified the effect of *USF1* variants on CVD risk factors followed in four examinations throughout adulthood.

The three studies provided valuable information on the study of complex traits, highlighting the use of large study samples, the importance of replication, and the full coverage of the major allelic variants of the target genes to assure reliable findings.

Keywords: Cardiovascular diseases, genetic association analysis, candidate gene approach, *F5*, *ICAM1*, *PROC*, *THBD*, *USF1*

Kirsi Auro, Candidate gene studies on cardiovascular traits Kansanterveyslaitoksen julkaisuja, A18/2007, 110 sivua
ISBN 951-740-752-6; 951-740-753-3 (pdf-versio)
ISSN 0359-3584; 1458-6290 (pdf-versio)
http://www.ktl.fi/portal/4043

TIIVISTELMÄ

Sydän- ja verisuonisairaudet ovat länsimaiden johtava kuolinsyy, ja kasvava ongelma kehitysmaissa. Koronaaritaudin ja iskeemisen aivohalvauksen perinnöllisyys on osoitettu kaksostutkimuksissa. Lisäksi useat tautiryhmän riskitekijöistä, kuten kohonnut verenpaine, rasva-aineenvaihdunnan häiriöt, lihavuus, sokeritauti ja nikotiiniriippuvuus ovat osittain perinnöllisiä.

Tämän väitöskirjatyön tavoitteena oli selventää viiden tromboosi- ja lipidiaineenvaihduntaan osallistuvan geenin osuutta sydän- ja verisuonitautien riskitekijöinä. Aihetta lähestyttiin väestötasolla, kookkaita pohjoismaisia väestöotoksia ja ehdokasgeeniasetelmaa hyödyntäen. Kirjan ensimmäinen osatyö keskittyi trombomoduliinigeeniin (*THBD*), jota tutkittiin kahdessa suuressa suomalaisessa väestöotoksessa (FINRISKI-92 ja FINRISKI-97). Yhteyttä *THBD*-geenin ja sydän- ja verisuonitautien välillä ei kuitenkaan havaittu.

Toinen osatyö käsitteli neljän geenin yhteisvaikutuksia sydän- ja verisuonitautien taustalla. Hyytymistekijä V, intersellulaarinen adheesiomolekyyli -1 ja proteiini C genotyypattiin samoissa FINRISKI-otoksissa ja analysoitiin yhdessä trombomoduliinin kanssa. Tulokset olivat rohkaisevia: useat yksittäiset polymorfiat ja polymorfiaparit lisäsivät tautiriskiä, ja tulokset olivat yhteneviä molemmissa väestökohorteissa. Ensimmäisen osatyön negatiivisista tuloksista huolimatta myös *THBD*-variantit lisäsivät tautiriskiä. Siten *THBD* saattaa vaikuttaa sydän- ja verisuonitauteihin geenien välisen interaktion kautta.

Kolmas osatyö keskittyi lipidigeeni upstream transcription factor -1:n (*USF1*) analyyseihin ruotsalaisessa mieskohortissa. Kardiovaskulaaritapahtumien lisäksi tutkittiin metabolista syndroomaa. *USF1*:n geenivariantit vaikuttivat metaboliseen syndroomaan sekä suojaavasti että riskiä lisäten. Suoraa yhteyttä sydän- tai aivotapahtumiin ei havaittu.

Yhdessä nämä kolme osatyötä tuottivat arvokasta tietoa monitekijäisten tautien tutkimuksesta, korostaen suurten väestöotosten merkitystä ja positiivisten löydösten toistamista toisessa aineistossa, sekä kohdegeenien yleisten varianttien kunnollista kattamista luotettavien tutkimustulosten takaamiseksi.

Avainsanat: Sydän- ja verisuonitaudit, geneettinen assosiaatioanalyysi, ehdokasgeeniasetelma, *F5*, *ICAM1*, *PROC*, *THBD*, *USF1*

CONTENTS

ABBREVIATIONS

ABCA1	ATP-binding cassette -1, subtype A gene
AD	Alzheimer's disease
APC	Activated protein C
ApoE	Apolipoprotein E
APOE	Apolipoprotein E gene
ATPIII	National Cholesterol Education Program's Adult treatment panel III
BMI	Body mass index
CHD	Coronary heart disease
chr	Chromosome
Cm	Chylomicrons
CNV	Copy number variation
CVD	Cardiovascular diseases
DNA	Deoxyribonucleid acid
EGIR	European Group for the Study of Metabolic Syndrome
F2	Prothrombin (Coagulation factor II) gene
F5	Coagulation factor V gene
F7	Soluble factor VII
F7	Coagulation factor VII gene
FXIII	Coagulation factor XIII gene
FCHL	Familial combined hyperlipidemia
FDR	False discovery rate
GLM	General linear model
GWAS	Genome-wide association study
HDL	High density lipoprotein cholesterol
HR	Hazard ratio
ICD	International classification for diseases
ICAM1	Intercellular adhesion molecule -1 gene
IDF	International Diabetes Federation
IDL	Intermediate density lipoprotein cholesterol
kb	Kilobase
LD	Linkage disequilibrium
LDL	Low density lipoprotein cholesterol
Lp(a)	Lipoprotein a
LPL	Lipoprotein lipase
MetS	Metabolic syndrome
MI	Myocardial infarction
MTHFR	5,10- methylenetetrahydrofolate reductase gene
n	Number

NO	Nitric oxide
PAI1	Plasminogen activator inhibitor -1
PAI1	Plasminogen activator inhibitor -1 gene
PLAT	Tissue plasminogen activator
PLAT	Tissue plasminogen activator gene
PROC	Protein C gene
sICAM1	Soluble intercellular adhesion molecule -1
SMC	Smooth muscle cell
SNP	Single nucleotide polymorphism
sTM	Soluble thrombomodulin
T2DM	Type 2 diabetes mellitus
TAFI	Thrombin activatable fibrinolysis inhibitor gene
TC	Total cholesterol
TF	Tissue factor
TG	Triglycerides
THBD	Thrombomodulin gene
ULSAM	Uppsala Longitudinal Study of Adult Men
USF1	Upstream transcription factor -1 gene
VLDL	Very low density lipoprotein cholesterol
WHO	World Health Organization

Gene names are italicized 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 **Auro K**, Komulainen K, Alanne M, Silander K, Peltonen L, Perola M, and Salomaa V. Thrombomodulin gene polymorphisms and haplotypes and the risk of cardiovascular events: a prospective follow-up study. *Arterioscler Thromb Vasc Biol*. 2006,26:942-947

II **Auro K**, Alanne M, Kristiansson K, Silander K, Kuulasmaa K, Salomaa V, Peltonen L, and Perola M. Combined effects of thrombosis pathway gene variants predict cardiovascular events. *PloS Genet*. 2007,3:e120

III **Auro K**, Kristiansson K, Zethelius B, Berne C, Taskinen MR, Jauhiainen M, Perola M, Peltonen L, and Syvänen AC. *USF1* gene variants contribute to metabolic traits in males in a longitudinal 32-year follow-up study. *Diabetologia, in press*

INTRODUCTION

Coronary heart disease (CHD) is currently the leading cause of death and disability in Western countries and is likely to become the leading cause of death worldwide by 2020 (Murray and Lopez 1997). The World Health Organization estimated that nearly 8 million people worldwide died of CHD in 2005 (www.who.int). In the United States alone, CHD affects 15 million people (Rosamond et al. 2007), and 300 000 individuals each year suffer sudden cardiac death, 80% of which of atherosclerotic origin (Wang 2005). Further, atherosclerosis, the underlying cause for cardiovascular diseases (CVD), is estimated to account for over 50% of all deaths in the United States (Lusis 2000).

During the last three decades, incidence of both first and recurrent coronary events has, however, declined steeply in Western countries due to major efforts in disease prevention and care. In Finland, mortality from recurrent coronary events dropped approximately 10% per year in 1983-1992 (Salomaa et al. 2003). Especially out-of-hospital mortality decreased during this period (Salomaa et al. 2003). Nevertheless, in 2004 CHD caused over 77 000 hospitalizations and about 12 000 deaths in Finland, and approximately 14 000 first myocardial infarctions (MI) were detected (www.stakes.fi).

Stroke morbidity imposes a substantial health burden, with mortality, lifelong handicap, and loss of "quality years" of life, causing permanent disability for up to 30% of those with non-fatal events (Goldstein et al. 2006). Stroke events also increase the risk for dementia later in life (Ivan et al. 2004). The lifetime risk for stroke was approximately 13% in middle-aged Americans (Seshadri et al. 2006). In Finland, about 14 000 stroke events take place annually (www.stakes.fi). In the past decades especially stroke mortality has decreased significantly (Tuomilehto et al. 1996, Sivenius et al. 2004, Pajunen et al. 2005 a).

This study investigated the role of genetic risk factors in CVD, concentrating on five candidate genes. The focus was at the population level where data on well-established genetic CVD risk variants are scarce. The completion of the Human Genome Project launched projects like HapMap, providing detailed information on haplotype structures across the genome, greatly facilitating the study of complex traits. In addition, technological achievements in the field have provided the means to unravel the genetic basis of common diseases. The following literature review provides an insight to cardiovascular diseases and their genetic background, and explores ways of discovering genes underlying CVD and other complex traits.

REVIEW OF THE LITERATURE

1. ATHEROSCLEROSIS

Atherosclerosis is a progressive disease of the major arteries, giving rise to cardiovascular outcomes, such as coronary heart disease and ischemic stroke. Other atherosclerotic manifestations include atherosclerosis of the lower limbs and ruptures of large vessels such as the aorta. Atherosclerosis is a dynamic disease process involving lipid accumulation, disturbed endothelial function, inflammatory response with recruitment of inflammatory cells and cytokine secretion, smooth muscle cell (SMC) proliferation in the intima, and ulceration and thrombosis of the plaque (Ross 1999, Lusis 2000).

The arterial wall consists of three layers, the intima, media and adventitia. The intima, the inner layer, is lined with endothelium: a single layer of flat epithelial cells and a base membrane of connective tissue and smooth muscle cells. The media is the muscle layer of the artery, outlined with connective tissue of the adventitia. The normally very thin intima is thickened when affected by atherosclerosis. Based on histological changes in the intima, the American Heart Association has suggested a six-stage grading for atherosclerotic lesions (Table 1) (Stary et al. 1994, Stary et al. 1995). Type I lesions contain lipid-laden macrophages, foam cells. Type II lesions are characterized by several layers of foam cells. Type III lesions contain additional extracellular cholesterol, while type IV lesions have a clear lipid core. Type V lesions have a fibrous cap of fibromuscular fibers. Type VI lesions contain ulcerations, hematoma, or thrombosis on the plaque surface. Type I lesions represent very initial atherosclerosis typically seen in children. Fatty streaks belong to type II lesions. Type III lesions, common in young adults, lay between early and advanced atherosclerosis, presenting initial intimal thickening. Only advanced lesions (types IV-VI) cause clinical symptoms.

***Table 1**. Histological grading of atherosclerotic lesions according to the American Heart Association (modified from Stary et al. 1994, Stary et al. 1995).*

Lesion type		Histological characteristics
Early lesions	Type I	Foam cells at lesion site
	Type II	Several separate foam cell layers present
Intermediate lesions	Type III	Extracellular cholesterol droplets
Advanced lesions	Type IV	Distinctive lipid core in the lesion
	Type V	Fibrous cap present, calcification may be seen (Type Vb)
	Type VI	Surface ulceration, hematoma, or thrombosis

1.1. A short overview of lipid metabolism

Lipid particles are transported in plasma in water-soluble lipoproteins (Gotto et al. 1986). The core of lipoproteins consists of triglycerides (TGs) and cholesterol esters, surrounded by phospholipids, apolipoproteins, and free cholesterol. Lipoproteins are divided into five classes by density: chylomicrons (Cms), very low density lipoproteins (VLDLs), intermediate density lipoproteins (IDLs), low density lipoproteins (LDLs), and high density lipoproteins (HDLs) (Gotto et al. 1986). Cms form in the intestine from dietary lipids and bile, and are cleared from the circulation by lipoprotein lipase (LPL). LPL facilitates the breadown of TG of Cm, giving rise to free fatty acids and Cm remnants. If not stored in adipose tissue as free fatty acids, TG can be utilized in the liver to form VLDL. VLDL is secreted from the liver into the circulation, where it is modified by LPL and hepatic lipase to form IDL and further LDL. LDL transports cholesterol to peripheral tissues via LDL receptors.

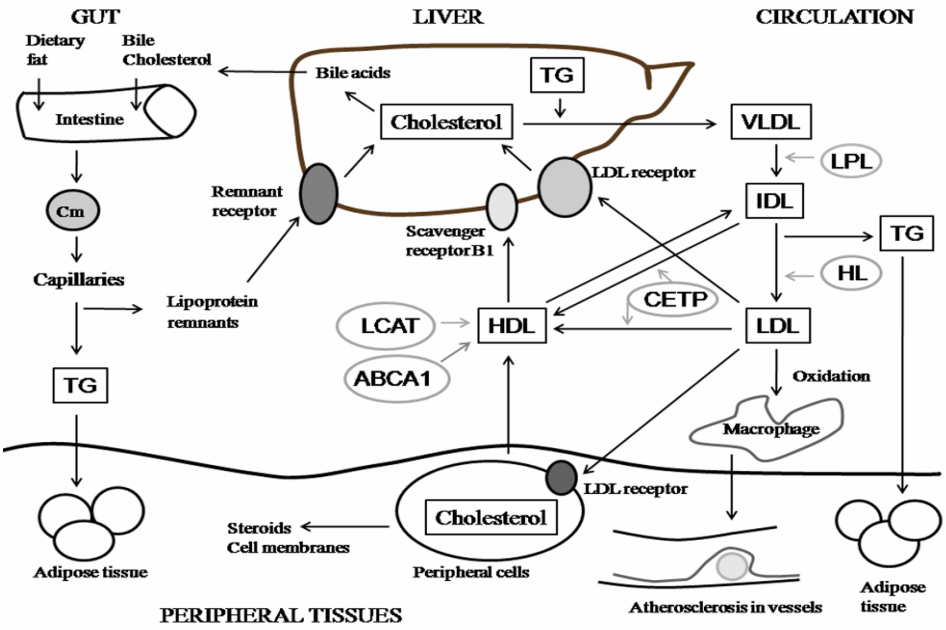
HDL is a key player in reverse cholesterol transport, returning cholesterol from tissues to liver to be exreted into bile. Cholesterol ester transfer protein transfers cholesterol esters from HDL to IDL and LDL. When functioning excessively, this cycle contributes to formation of highly atherogenic, small, dense LDLs. Other HDL-modifying enzymes include lecithin:cholesterol acyl transferase and paraoxonase -1.

Apolipoproteins act as stabilazers in soluble lipoproteins, as enzymatic cofactors, and as ligands for lipoprotein receptors. Apolipoproteins A-1 (ApoA-1) and A2 are the main apolipoproteins in HDL, and apolipoprotein B (ApoB) is the major component of other apolipoproteins. Apolipoprotein E (ApoE) serves as a ligand for LDL receptor and is needed in formation of VLDL and Cms (Eichner 2002). Lipid metabolism is illustrated in Figure 1.

1.2. Early atherosclerotic lesions

Atherosclerotic changes begin with lipid accumulation in the vessel wall. Early lesions are visible in humans already in the first decade of life (Lusis 2000). The process usually begins in the aorta, proceeding to the coronary arteries during the second decade and to the cerebral arteries in the third or fourth decade (Lusis 2000).

***Figure 1**. Overview of lipid metabolism.*



ABCA1: ATP-binding cassette -1, subtype A; CETP: Cholesterol ester transfer protein; Cm: Chylomicrons; HDL: High density lipoprotein cholesterol; HL: Hepatic lipase; IDL: Intermediate density lipoprotein cholesterol; LCAT: Lecithin:cholesterol acyl transferase; LDL: Low density lipoprotein cholesterol; LPL: Lipoprotein lipase; TG: Triglycerides.

Early atherosclerotic lesions, called fatty streaks, contain mainly foam cells; macrophages laden with LDL cholesterol. Lesion formation begins when lipid particles, together with inflammatory monocytes and T lymphocytes, accumulate at the site (Ross and Glomset 1976 a, Ross and Glomset 1976 b, Ross 1999). Monocytes then migrate into the subendothelial space and differentiate into macrophages, which express scavenger receptors necessary for LDL uptake (Mahmoudi et al. 2007). Before LDL can be taken up by the macrophages, it has must be modified by oxidation and other ways including lipolysis, proteolysis, glycation (in diabetes), or different forms of aggregation with other cells or complexes (Ross 1999, Lusis 2000). Accumulation of oxidized LDL into the artery wall triggers inflammatory response with the recruitment of monocytes and lymphocytes at the lesion site (Ross 1999). LDL accumulation can also stimulate these inflammatory cells, being thus the major initiating event in atherogenesis.

1.3. Endothelial dysfunction

The endothelium is a dynamic organ regulating vasodilatation and vasoconstriction, local inflammatory response, and thrombogenic-thrombolytic balance at the vessel wall. The endothelium also controls fibrinolysis and SMC stimulation, inhibition, proliferation, and differentiation (Kinlay et al. 2001, Mahmoudi et al. 2007). Healthy endothelium is a smooth, nonthrombogenic permeability barrier between the circulation and arterial wall (Stary et al. 1992).

Endothelial dysfunction is a highly atherogenic process characterized by an imbalance between vasoconstriction and vasodilatation, and increased vessel wall permeability. Several CVD risk factors impair the normal functions of endothelial cells. Dyslipidemias, vasoconstrictor hormones involved in hypertension, diabetes-associated glycoxidation, proinflammatory cytokines, smoking-derived free radicals, shear stress, infectious agents like herpesviruses and Chlamydia pneumoniae, and genetic variations may injure the endothelial cells (Libby et al. 2005, Mahmoudi et al. 2007).

A key player in vasodilatation, in addition to prostacyclin and bradykinin, is nitric oxide (NO). Lipids, especially LDL, together with local oxidative stress, reduce NO amount and availability, impairing vasodilatation. Small invaginations in the endothelial cell surface, caveolae, contain caveolin –1, which inactivates nitric oxide synthase. LDL stimulates caveolin –1 synthesis and increases nitric oxide synthase inactivation, enhancing vasoconstriction (Kinlay et al. 2001). Lipids also promote inflammation by activation of inflammatory pathways. Inflammatory response leads to accumulation of monocytes and macrophages at the lesion site and their subsequent adhesion and transmigration into the intima. HMG-CoA reductase inhibitors, LDL-lowering statins, in turn, reduce caveolin –1 expression (Kinlay et al. 2001). The endothelium responds to shear stress first by increasing NO and vasodilatation, but in time by vasoconstriction (Gimbrone et al. 2000).

1.4. Inflammation

Recent insights from the molecular level highlight atherosclerosis as an inflammatory process (Ross 1999, Lusis 2000). Inflammation is present at all stages of atherosclerosis, from early fatty streaks to the most complex lesions, contributing importantly to lesion growth. The key inflammatory cells involved in atherosclerosis are monocyte-derived macrophages and T lymphocytes, both which gather at the lesion site due to local changes caused by endothelial dysfunction, enter the vessel wall, and stimulate series of events leading to lesion enlargement and destabilization (Soll et al. 2006). Early atherogenic lesions are, in fact, pure inflammatory lesions,

consisting of inflammatory cells, mainly monocyte-derived macrophages and T lymphocytes.

Vascular cell adhesion molecules have a central role in atherogenesis (Ridker 2001). Selectins E, L, and P are involved in the leukocyte capture and rolling on the endothelial surface. L selectin is expressed in leukocytes, selectin P in platelets, and selectins E and P in endothelial cells when inflammation prevails (Galkina and Ley 2007). Intercellular cell adhesion molecule -1 (ICAM1) and vascular cell adhesion molecule -1 (VCAM1) contribute to leukocyte adhesion. In endothelial dysfunction, endothelial cells upregulate the expression of monocyte- and T-cell-binding cell surface receptors and cell adhesion molecules ICAM1 and VCAM1 (Mahmoudi et al. 2007). ICAM1 and VCAM1 are also expressed in the SMCs in atherosclerotic plaques (Braun et al. 1999). Connexins are involved in the transmigration of inflammatory cells (Galkina and Ley 2007).

After migration from the blood stream to the arterial intima, the inflammatory cells activate SMCs to migrate from the media to the intima, proliferate, and secrete extracellular matrix (Libby et al. 2005). The matrix contains matrix metalloproteinases, which in turn modulate numerous changes in vascular cells (Libby et al. 2005). In the intima, the macrophages form foam cells by scavenging modified LDL particles and maintain chemotaxis by producing cytokines and proteolytic enzymes. The macrophages themselves induce production of tissue factor (TF), cytokines, and matrix metalloproteinases (Soll et al. 2006). In addition, T lymphocytes secrete cytokines when activated by antigens such as oxidized LDL. Thus, macrophages and T lymphocytes maintain the inflammation process at the lesion site.

Animal studies have provided important insights to the role of inflammation in atherosclerosis. Apolipoprotein E -deficient (apoE -/-) mice express substantially high cholesterol levels and develop extensive atherosclerosis (Piedrahita et al. 1992, Plump et al. 1992). Zhou and associates (1996) addressed the inflammatory process in apoE -/- mice and observed that the atherosclerotic lesions in the mice were laden with inflammatory T cells. Further, cross-breeding apoE -/- mice with severe combined immunodeficiency mice lacking both B and T cells significantly reduced fatty streak development (Zhou et al. 2000). Gene therapy targeted against chemoatractive functions of monocytes or macrophages reduced atherosclerosis in apoE -/- mice (Borin et al. 1998, Inoue et al. 2002). Moreover, crossing apoE -/- mice with mice lacking functional macrophages resulted in diminished lesions (Smith et al. 1995). Mice deficient for selectins have less extensive atherosclerosis under fatty diet (Galkina and Ley 2007).

1.5. Thrombosis

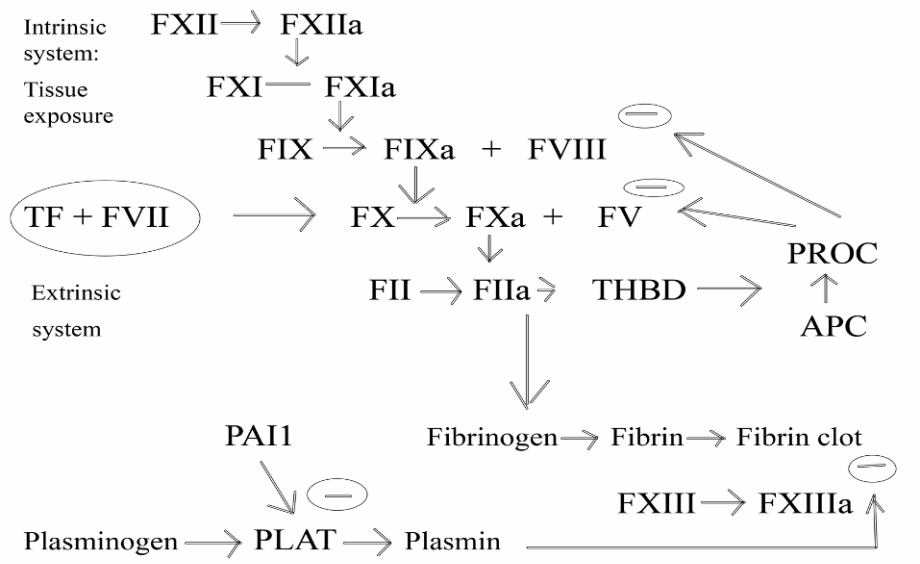
Vascular thrombosis can be divided into two distinct groups differing in the site of occurrence and thrombus composition. The issue has been reviewed by Reitsma (2004): Arterial thrombosis is an end-stage disease of atherosclerotic medium size and large arteries, typically seen in areas with fast blood flow. Arterial thrombi consist primarily of platelets. Venous thrombosis occurs most often in stasis areas of blood flow, mainly in deep, large veins of the lower extremities and in pulmonary arteries. Instead of platelets, the main components in venous thrombi are fibrin and red blood cells.

Thrombosis is a key feature in advanced atherosclerotic lesions (Stary et al. 1995). Plaque ulceration and rupture expose the tissue underlying the endothelium, initiating a series of blood clotting events, leading to thrombus formation on the lesion surface (Spronk et al. 2004). Thrombotic lesions grow in size, restricting the flow of blood and oxygen in vessel lumen. Thrombotic lesions are also prone to rupture, subsequently causing occlusion downstream of the lesion (Libby 2002, Libby et al. 2005).

1.5.1. Blood clotting

The clotting pathway is a cascade of enzymatic and physical reactions, ultimately leading to blood coagulation. Both intrinsic and extrinsic factors can activate the cascade. The intrinsic pathway activates when endothelial damage exposes the underlying tissue; factor (F) XII turns into its activated form FXIIa, activating factor XI. FXIa in turn activates FIX, and FIXa in combination with FVIII activate FX. FXa, together with FV, phospholipid, and calcium, forms a prothrombinase complex, leading to formation of thrombin from prothrombin. Subsequently, fibrin is formed from fibrinogen. Fibrin monomers compose a network where red and white blood cells adhere at the coagulation site. Activation of FXIII into FXIIIa stabilizes the fibrin in the thrombus. The extrinsic pathway activates in the presence of TF. The TF – FVII complex activates FX, joining to the intrinsic pathway at this point. Both pathways are illustrated in Figure 2. (Cecil 2004.)

Figure 2. *The clotting pathway is a cascade of enzymatic and physical reactions, leading to formation of a red thrombus (modified from Cecil 2004).*



APC: Activated protein C; F*x*: Coagulation factor (inactive); F*x*a: Activated coagulation factor; PAI1: Plasminogen activator inhibitor -1; PLAT: Tissue plasminogen activator; PROC: Protein C; TF: Tissue factor; THBD: Thrombomodulin.

Natural anticoagulants dissolve unnecessary thrombi. Plasminogen has both fibrinolytic and anticoagulant functions. Tissue plasminogen activator (PLAT) turns plasminogen into plasmin, which dissolves both fibrinogen and fibrin. The dissolution products further downregulate fibrin formation. Plasmin also has effects on coagulation factors V and VIII. Plasminogen activator inhibitor -1 (PAI1) inhibits the actions of PLAT. Thrombin stimulates anticoagulation by binding to thrombomodulin (THBD), which transforms protein C (PROC) into its activated form (APC). This further inhibits factors FV and FVIII. (Cecil 2004.)

1.5.2. Thrombosis and atherosclerosis

Over time, atherogenesis proceeds from early fatty streaks to more complex lesions. Traditional views see lesion growth as an accumulation of SMCs in plaque. SMCs secrete extracellular matrix, contributing to lesion size. The increasingly large lesion protrudes into the lumen of the artery, finally rupturing as an occlusive thrombus

(Libby 2002). Thrombosis is characteristic for disease conversion from stable into acute phase (Stary et al. 1995).

More recent views on thrombosis challenge traditional ones by introducing a concept of discontinuous lesion growth, where lesion size increases due to sudden microenvironmental changes such as thrombosis (Davies 1996). Three main types of physical disruption can occur at the lesion: superficial erosion, disruption of microvessels, and rupture of the lesion's fibrous cap (Libby 2002, Libby et al. 2005). In superficial erosion, subendothelial collagen is revealed due to endothelial desquamation, leading to platelet activation and adhesion. On the other hand, an atherogenic lesion produces a fragile neovasculature, prone to rupture, leading to thrombosis *in situ*, thrombin generation, and platelet activation. The third type of disruption is the most common; rupture of the lesion's fibrous cap brings coagulation factors in contact with tissue factor, promoting thrombosis. TF, triggering the intrinsic coagulation pathway, is a prominent component of atherosclerotic lesions, present in the macrophages infiltrating plaques (Spronk et al. 2004).

Thrombosis is also related to conditions predisposing to atherosclerotic outcomes. Normal insulin actions may reduce thrombus formation (Trovati et al. 1995). Insulin resistance, linked to obesity, may thus increase thrombogeneity through reduced platelet inhibition (Westerbacka et al. 2002). Obesity likely contributes to increased thrombogeneity also through other mechanisms, as reviewed by Darvall et al. (2007). Adipose tissue secretes numerous thrombosis-related substances such as adiponectin, leptin, PAI1, and TF. The secreted cytokines interleukin 6 and tumor necrosis factor -α demonstrate obesity as an inflammatory state. Obesity, further, is related to oxidative stress and endothelial dysfunction, which again increase thrombotic activities.

Endothelial dysfunction, inflammation, and thrombosis tightly intertwine with each other in atherogenesis. Thrombophilias contribute to severe inflammatory conditions such as sepsis (Hofstra et al. 2007). Thrombomodulin - Protein C pathway may promote inflammation independently of their coagulation function (Scaldaferri et al. 2007). CVD risk factors are related to all of these conditions. Obesity, for example, contributes to dyslipidemias, inflammation, and thrombogenesis (Darvall et al. 2007), and smoking increases endothelial dysfunction (Libby et al. 2005, Mahmoudi et al. 2007) and thrombosis (Miller et al. 1998, Fernandez et al. 2002, Pomp et al. 2007).

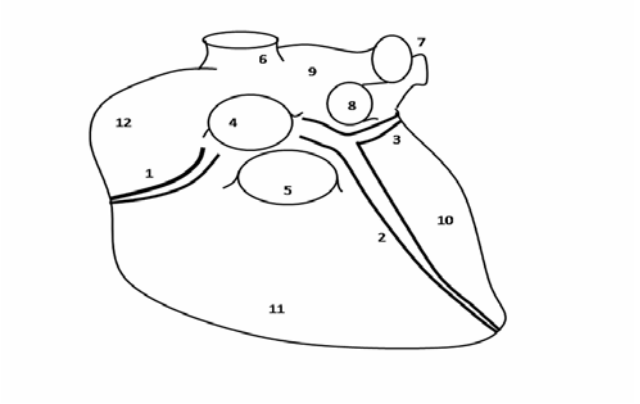
2. CARDIOVASCULAR DISEASES

The major clinical manifestations of atherosclerosis are coronary heart disease (CHD), stroke, and peripheral artery disease. Here we concentrate on the two former, used as endpoints in the genetic studies.

2.1. Coronary heart disease

The heart is typically supplied with three main coronary arteries. The right and left coronary arteries arise from the ascending aorta. After a short main trunk in the left coronary groove between the atrium and the ventricle, the left coronary artery branches into two: The left circumflex branch follows the coronary groove to the posterior surface of the heart, and the left anterior descending (interventricular) branch runs towards the apex of the heart between the ventricles. Typically, the descending branch is larger and supplies both ventricles and the interventricular septum, including the atrioventricular bundle (Moore 1991). The smaller circumflex branch supplies the left atrium and left ventricle from the posterior side (Moore 1991). The right coronary artery, running along the right coronary groove, typically feeds the right atrium, right ventricle, and atrioventricular and sinuatrial nodes (Moore 1991). Figure 3 illustrates the typical location of the coronary arteries.

Figure 3. *The heart and typical locations of the three main coronary arteries (modified from Netter 1989).*



1: Right coronary artery; 2: Left descending (interventricular) coronary artery; 3: Left circumflex coronary artery; 4: Aorta; 5: Left pulmonary artery; 6: Superior vena cava; 7: Left pulmonary veins; 8: Left main bronchus; 9: Left artrium; 10: Left ventricle; 11: Right ventricle; 12: Right atrium.

Coronary heart disease typically manifests as chest pain, although the disease can also be practically symptomless. Angina pectoris refers to a pain symptom experienced in the chest area, possibly radiating to the neck, left shoulder, left arm, jaw, teeth, right arm, or epigastric region. Pain may be experienced as tightness, squeezing, constriction, burning pain, or dull discomfort. Stable angina is diagnosed with recognition of the typical symptoms. Physical effort may trigger angina, and the symptom is relieved by rest. Unstable angina refers to situations where the patient has new-onset or worsening angina, or the symptom occurs in rest. Atherosclerotic narrowing of one or more coronary arteries is the most common cause for angina. Some other ischemic causes for chest pain include coronary spasms, dissections, reduced oxygen supply, or increased oxygen demand. Chest pain may also occur due to infection or of noncardiac origin. (Cecil 2004.)

Myocardial ischemia refers to imbalance between oxygen demand and supply in the myocardium. Acute coronary syndrome points to myocardial ischemia, ranging from unstable angina to myocardial infarction (MI) with no ST-segment elevations in electrocardiogram, characterized by elevation of the markers of myocardial injury (troponins or cardiac isoenzymes). A nonocclusive thrombus in a coronary artery is the most common cause for acute coronay syndrome. In MI with ST-segment elevation, blood flow to the myocardium is completely blocked. Coronary emboli or thrombosis, vasculitis, vasospasm, or trauma in the coronaries are non-atherogenic causes for MI (Cecil 2004.)

Traditionally atherosclerotic narrowing of the vessels was thought to results from protrusion of advanced plaques into the arterial lumen. It is now understood that the vasculature adapts to the proceeding atherosclerosis with remodeling the blood flow (Korshunov et al. 2007), and substantial atherosclerosis can exists even in the absence of protruding plaques. These silent lesions grow outward rather than inward, and do not cause any clinical symptoms – before sudden changes in the lesion, such as thrombosis, give rise to unstable plaques (Libby et al. 2005). Distribution of atherosclerotic lesions in the arteries is not uniform. Flow conditions in the vasculature, namely shear stress, affect lesion development. The lesions are often seen to develop at sites where low shear stress prevails, i.e. at branch points and downstream of stenoses (Gimbrone et al. 2000, Stone et al. 2003). The most common sites of coronary occlusions are the anterior interventricular branch of the left coronary artery, the right coronary artery, and the circumflex branch of the left coronary artery (Cecil 2004).

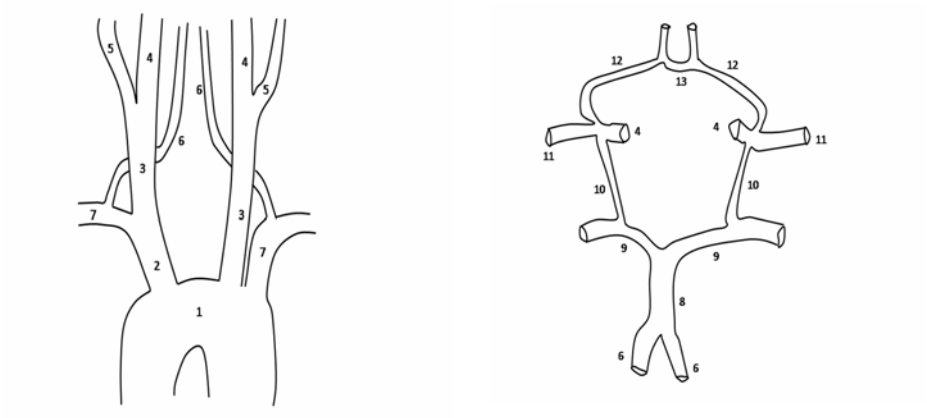
2.2. Ischemic stroke

The brain is supplied by two internal carotid arteries and two vertebral arteries (Figure 4). Together these arteries form an anastomosis ring at the base of the brain, known as the circle of Willis. The circle contains five arteries (Figure 4): posterior cerebral, posterior communicating, internal carotid, anterior cerebral, and anterior communicating arteries (Moore 1991).

Stroke is defined as acute-onset focal loss of cerebral function. Symptoms from a stroke event vary, but persist for 24 hours or more, whereas similar symptoms lasting less than 24 hours refer to transient ischemic attack. Strokes can arise due to ischemic events, intracerebral bleeding, or subarachnoidal hemorrhages. (Cecil 2004.)

Of all stroke events, approximately 80% are of ischemic origin (Humpries and Morgan 2004). Atherosclerosis of the carotid arteries accounts for over 30% of the ischemic strokes (Sandercock et al. 1989). Other causes for ischemic stroke events are cardiogenic embolisms and atherosclerosis of the small arteries. The diagnosis of stroke events is based on persisting clinical symptoms and radiological imaging.

Figure 4. *The carotid arteries and the circle of Willis. (Modified from Netter 1989.)*



1: Aorta; 2: Brachiocephalic trunk; 3: Common carotid artery; 4: Internal carotid artery; 5: External carotid artery; 6: Vertebral artery; 7: Subclavian artery; 8: Basilar artery; 9: Posterior cerebral artery; 10: Posterior communicating artery; 11: Middle cerebral artery; 12: Anterior cerebral artery; 13: Anterior communicating artery.

2.3. Cardiovascular risk factors

CVD events rarely occur in individuals who are unexposed to risk factors. Greenland and coworkers (2003) observed that up to 90% of the affected persons had one or more risk factor. According to Lloyd-Jones et al. (2006 a), individuals free of the risk factor burden at the age of 50 years had a significantly lower lifetime risk for cardiovascular manifestations than those with two or more risk factors. Traditional cardiovascular risk factors comprise advanced age, male gender, smoking, hypertension, elevated total cholesterol (TC) and LDL cholesterol, low HDL cholesterol, obesity, and family history of CVD (Fruchart et al. 2004). Newer risk factors for CVD include disturbances of glucose metabolism, and hemostatic factors and inflammatory markers (Fruchart et al. 2004). Table 2 lists the major risk CVD factors.

Table 2. *Factors increasing the risk for cardiovascular events.*

FACTORS AFFECTING THE RISK OF CARDIOVASCULAR DISEASES
Advanced age
Male gender
Smoking
Hypertension
Low socioeconomic status
Physical inactivity
Dyslipidemias
Obesity
Diabetes
Molecular markers
Family history

Aging is a major risk factor for both CHD and stroke events. The burden of the CVD risk factors tends to increase by age: Advanced age relates to dyslipidemias (Aromaa 2002), obesity (Rahkonen et al. 1998), and hypertension (Aromaa 1981). The proinflammatory state accompanied with aging may also contribute to the development of atherosclerosis (Vasto et al. 2007). Stroke risk increases significantly after 55 years (Goldstein et al. 2006).

Male gender is a significant risk factor for CHD. Over 90 000 CHD events, of which 42 000 fatal, were recorded among men, compared with 36 000 CHD events and 15 000 fatal events among women in the Finnish Hospital Discharge Register and Causes of Death Register during 1991-2001 (Pajunen et al. 2004). Jousilahti et al. (1999) reported a three-fold CHD risk for men compared with women. According to the FINMONICA register, men, compared with women, were also at 1.7 times increased risk for recurrent coronary event (Schreiner et al. 2001). Women, however, tend to fare less well after CHD events worse than men in long-term, possibly due to their older age at disease onset, diabetes, and other factors (Schreiner

et al. 2001). Men also have higher risk for stroke events (Brown et al. 1996, Sacco et al. 1998), except among the young and the old (<44 years or >85 years, Sacco et al. 1998). Stroke events were slightly more common among women than men in Finland during 1991-2001 (Pajunen et al. 2005 a).

Smoking approximately doubles the risk for MI (Prescott et al. 1998). Heavy smokers are reported to hold a double risk for stroke compared to light smokers (Wolf et al. 1988). The stroke risk decreased to level of non-smokers within 5 years from cessation of smoking (Wolf et al. 1988). On the other hand, smoking also decreases life expectancy, and as the risk for cardiovascular events increases with age, smoking may actually decrease the probability of CVD and shorten the duration of the disease (Mamun et al. 2004).

Hypertension is the most important modifiable risk factor for stroke (Goldstein et al. 2006). Hypertension contributes to CHD and stroke risk in a dose-like manner (MacMahon et al. 1990), increasing the risk of cardiovascular outcomes 2- 3 -fold (Kannel 1996). Hypertension commonly clusters with other CVD risk factors (Kannel 1996).

Individuals with lower socioeconomic position are more prone to have a fatal coronary or ischemic stroke event than those with a high socioeconomic status (Salomaa et al. 2001). Socioeconomic factors were observed to explain 30% (women) to 50% (men) of the difference in CHD mortality in Finland (Salomaa et al. 2001). Low socioeconomic status has also been shown to increased the risk for ischemic stroke events (36%) and mortality (56%) (Jakovljevic et al. 2001). Physical activity, occupational or leisure-time, has a positive effect on HDL, blood pressure, and body-mass index (BMI) in a Finnish study, reducing CVD risk in several ways (Barengo et al. 2006).

2.3.1. Dyslipidemias

The lifetime risk for CHD increases substantially with TC levels (Stamler et al. 2000). According to Lloyd-Jones et al. (2003), the lifetime (from 40 to 80 years) CHD risk figures with TC levels of <5.2 mmol/l, 5.2 – 6.2 mmol/l, and <6.2 mmol/l were 31%, 43%, and 57% for men, and 15%, 26%, and 33% for women, respectively.

Familial combined hyperlipidemia (FCHL) patients have elevated TC or TG or both, and are at substantial risk for early coronary event; FCHL is seen in approximately 20% of CHD patients <60 years (Goldstein et al. 1973, Nikkilä and Aro 1973). Besides hypertriglyceridemia and hypercholesterolemia, other dyslipidemias, such as hyperapobetalipoproteniemia (Sniderman et al. 1980, Brunzell et al. 1983),

elevated small, dense LDL particles (Austin et al. 1990, Hokanson et al. 1993), and low HDL cholesterol (de Graaf and Stalenhoef 1998), as well as disturbances in glucose metabolism, such as poor glucose tolerance (Hunt et al. 1989, Aitman et al. 1997) and insulin resistance (Pihlajamäki et al. 2000), are common FCHL-related features. The FCHL phenotype thus overlaps considerably with the phenotype of metabolic syndrome (Ayyobi and Brunzell 2003). The prevalence of FCHL is 1-6% in the western countries (Suviolahti et al. 2006).

Traditionally, TC and LDL levels have been the key indicators of CVD risk, and major guidelines like the National Cholesterol Education Program Adult Treatment Panel III (ATPIII) (NCEP 2001) are based on these markers. Recent findings, however, highlight apolipoprotein B (ApoB) and the ApoB/ ApoA-1 ratio as more accurate predictors of CVD (Barter et al. 2006). All atherogenic lipid particles – VLDL, IDL, LDL, and lipoprotein a (Lp(a)) – contain ApoB-100 and chylomicrons contain ApoB-48. The total number of atherogenic particles is suggested to reflect CVD risk better than LDL alone, highlighting the role of ApoB (Barter et al. 2006). ApoB was observed to predict 10-year CVD risk independently from the traditional factors (Kiechl et al. 2007). Elevated TC, LDL, and ApoB also contribute to the risk of ischemic stroke events (Walldius et al. 2006). Controversial results have been reported on Lp(a) and the risk of CHD (Jauhiainen et al. 1991, Alfthan et al. 1994, Danesh et al. 2000, Murase et al. 2007).

HDL cholesterol has several anti-atherogenic roles. As a key player in reverse cholesterol transport, HDL carries excess cholesterol from tissues to liver, decreasing the amount of circulating cholesterol. HDL also acts as an antioxidant (Navab et al. 2001). Low HDL cholesterol concentration (<0.9 mmol/l) is an independent CVD risk factor (Schaefer et al. 1994).

Dyslipidemias are mainly complex genetic disorders where genes play a role among environmental factors. Life-style factors affecting the cholesterol levels comprise smoking, alcohol consumption, physical activity, gender, and age. The genetics of dyslipidemias is discussed in chapter 3.2.1.

2.3.2. Obesity, metabolic syndrome, and type 2 diabetes

The importance of diabetes and obesity as CVD risk factors is increasing steadily; while cholesterol levels are decreasing due to targeted health programs and pharmaceutical use, obesity and diabetes rates are constantly growing because of poor diet and physical inactivity (Smith 2007). Over 50% of Americans are overweight

(Smith 2007). In Finland, approximately 20% of adults have BMI exceeding 30kg/m^2 (Laatikainen et al. 2003). The risk for type 2 diabetes (T2DM) increases exponentially with BMI; women with BMI >31 kg/m^2 had about 40 times and women with BMI > 35kg/m^2 about 90 times higher risk for T2DM lean lean (BMI <21 kg/m^2) women (Colditz et al. 1995). In men with BMI >35 kg/m^2, a 40-fold T2DM risk compared with lean men has been reported (Chan et al. 1994). Obesity increases the risk for CVD also by predisposing to dyslipidemias (Devroey et al. 2004), hypertension (Huang et al. 1998), and inflammation (Darvall et al. 2007).

Metabolic syndrome (MetS) is characterized as a clustering of cardiovascular risk factors. The diagnostic criteria include different combinations of central obesity, low HDL cholesterol, elevated TG, elevated fasting glucose or insulin resistance, and hypertension. Several different definitions for MetS exist (Table 3). Definitions of the World Health Organization (WHO) (WHO 1999) and the European Group for the Study of Metabolic Syndrome (EGIR) (Balkau et al. 2002) are glucocentric, highlighting insulin resistance as a key feature. The International Diabetes Federation (IDF) (Alberti et al. 2006) describes central obesity as a critical factor, whereas the National Cholesterol Education Program Adult Treatment Panel III (NCEP 2001) includes two lipid criteria instead of one.

Table 3. *Definitions for metabolic syndrome. Male- and ethnic-specific values shown.*

	ATPIII	IDF	WHO [a]	EGIR
	3 or more of the following factors:	Waist grid > 94 cm with two or more of the following:	Glucose intolerance, IGT or insulin resistance with two of the following:	Insulin resistance with two or more of the following:
Central obesity	Waist grid >102cm	-	Waist-hip ratio >0.9 or BMI > 30 kg/m2	Waist grid ≥ 94 cm
Triglycerides	≥1.7 mmol/l [b]	≥1.7 mmol/l [b]	≥1.7 mmol/l [b]	>2.0 mmol/l [b]
HDL-cholesterol	≤1.03 mmol/l [b]	≤1.03 mmol/l [b]	<0.9 mmol/l [b]	<1.0 mmol/l [b]
Blood pressure	≥130/≥ 85 mmHg [c]	≥130/≥85 mmHg [c]	≥140/≥ 90 mmHg [c]	≥140/≥ 90 mmHg [c]
Fasting glucose	>5.6 mmol/l	>5.6 mmol/l	-	>6.1 mmol/l

[a] WHO definition also includes microalbuminuria as a criterion [b] Or lipid lowering medication [c] Or hypertension treatment [d] Or previously diagnosed type II diabetes [e] With non-diabetic individuals

Opinions vary as to which of the various definitions is the best. The ATPIII definition does not require specific measurements for insulin resistance, and is therefore probably the most useful in clinical practice. Lakka and colleagues (2002) observed that the ATPIII definition was less consistent in predicting CVD and all-cause mortality in men. Marchesini and colleagues (2004), in turn, suggested that ATPIII may predict future coronary events in T2DM patients better than the WHO

definition, and Tong et al. (2007) noted that ATPIII predicts future CHD better than the IDF definition. The WHO definition is proposed to have high sensitivity and specificity in predicting future diabetes in MetS patients (Laaksonen et al. 2002).

Metabolic syndrome is known to predispose to T2DM and is associated with a 2- to 4-fold risk for CVD events and mortality (Lakka et al. 2002, Sundström et al. 2006). The mechanisms underlying the contribution of MetS to CVD risk are, however, incompletely understood. Thrombosis is suggested to act as a link between MetS and CVD. Insulin resistance can be viewed as a pro-thrombotic state, characterized by endothelial activation, hypercoagulation, and hypofibrinolysis (Darvall et al. 2007). Obesity, predisposing to insulin resistance and a key feature in MetS, promotes the secretion of several pro-thrombotic substances and directly promotes inflammation and endothelial dysfunction, components of increased thrombosis (Darvall et al. 2007). Platelets express insulin receptors. In nonobese, healthy subjects, insulin prevents platelet aggregation (Trovati et al. 1995), but antiaggregation is suggested to be disturbed in obese (Trovati et al. 1995) and hypertensive (Touyz et al. 1994) subjects. Abnormal insulin actions, linked to obesity, are suggested to increase thrombogenic activities by reducing platelet inhibition and interaction with collagen (Westerbacka et al. 2002). Elevated glucose levels possibly also increase thrombus formation via tissue factor release (Sambola et al. 2003).

People with T2DM have 2-4 times increased risk for coronary events and a 4-fold risk for CHD death (Manson et al. 1991, Koskinen et al. 1992, Haffner and Cassels 2003). Individuals suffering from MI also were at 2- (men) to 4-fold (women) risk of developing diabetes compared with the population free of MI (Pajunen et al. 2005 b). The risk for stroke events is double in the diabetics (Manson et al. 1991, Wannamethee et al. 1999), and diabetes significantly worsens the prognosis of ischemic strokes (Kaarisalo et al. 2005).

2.3.3. Molecular markers

Several molecular markers are associated with the risk of cardiovascular events. Fibrinogen is one of the most significant hemostatic CVD predictors (Voetsch and Loscalzo 2004, Kannel 2005, Rajecki et al. 2005, Smith et al. 2005). People with fibrinogen levels in the highest quartile have approximately double the risk for both MI and ischemic stroke events as those with fibrinogen in the lowest quartile

(Voetsch and Loscalzo 2004). Increased fibrinogen levels are associated with such CVD risk factors as obesity, smoking, advanced age, physical inactivity, hypertension, dyslipidemias, and diabetes (Voetsch and Loscalzo 2004, Kannel 2005). Fibrinogen is also an acute-phase reactant together with C-reactive protein (CRP), and may contribute to CVD via inflammatory response (Kannel 2005). Elevated CRP concentrations are associated with the risk of cardiovascular diseases (Ferranti and Rifai 2007). Homocysteine may not predict the risk of CHD in the general Finnish population (Alfthan et al. 1994, Voutilainen et al. 2000, Voutilainen et al. 2007), but elevated homocysteine concentrations may predict CHD events in individuals with diabetes (Soinio et al. 2004), and are assotiated with the risk of stroke (Virtanen et al. 2005).

Fibrinogen and plasminogen concentrations have showed association with all CVD events, but not with isolated strokes in a Finnish study (Rajecki et al. 2005). Levels of PLAT antigen and PAI1 are associated positively with the degree of carotic stenosis, an important predictor for ischemic stroke events (Soinne et al. 2005). High PAI1 expression is seen in atherosclerotic lesions especially in people with diabetes (Aso 2007). High levels of coagulation factors VIIa and VIII, as well as von Willebrand factor, are associated with atherosclerosis (Meade et al. 1986).

A neuroprotective role has been suggested for activated protein C. The issue was reviewed by Griffin et al. (2006). High APC concentrations are associated with lower incidence of ischemic strokes, whereas the ischemic stroke patients have low APC concentrations. Protein C deficiency has been described in several case reports in patients having arterial thrombotic events (Cakir et al. 2002, Tiong et al. 2003). Activated protein C, in response to thrombin formation, contributed also to reperfusion hemodynamics after cardiac surgery (Raivio et al. 2007).

Several studies have reported an association of high soluble ICAM1 (sICAM1) concentration with CHD (Haugt et al. 1996, Hwang et al. 1997, Malik et al. 2001) or with atherosclerosis (Jenny et al. 2006). High sICAM is also suggested to contribute to CHD in relation to low soluble thrombomodulin (sTM) (Wu et al. 2003). The relationship of soluble VCAM1, E selectin, and P selectin is less well established (Hwang et al. 1997, Malik et al. 2001). High sICAM1 levels have shown association with ischemic stroke events (Simundic et al. 2004, Ehrensperger et al. 2005) and with poor survival after a stroke (Wang et al. 2006, Blum et al. 2006). High sICAM1 also predicted future stroke events in CHD patients (Tanne et al. 2002), and sICAM1 may prove a marker for silent cerebral infarctions (Kawamura et al. 2006). Elevated soluble VCAM1 and soluble E-selectin concentrations may be associated

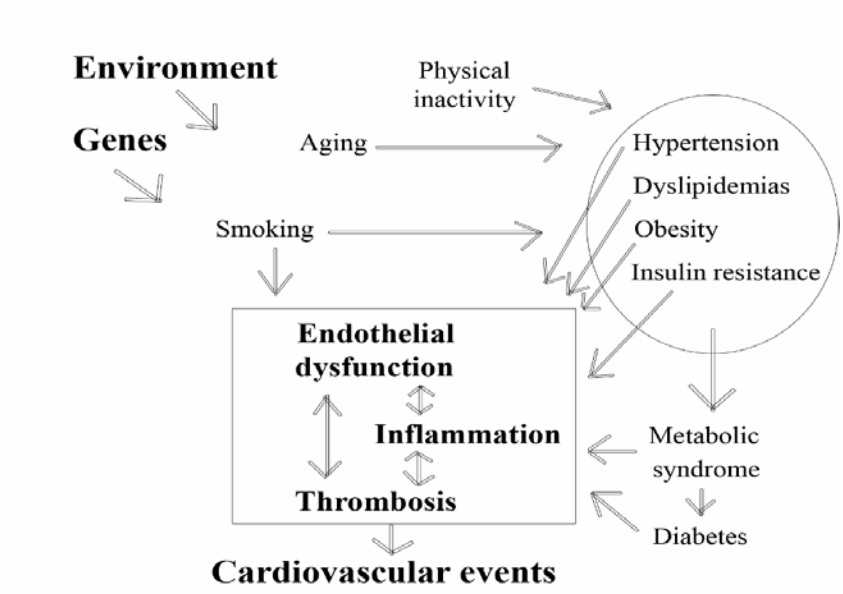
with the risk of stroke (Simundic et al. 2004, Blum et al. 2006), and elevated expression of ICAM1 and VCAM1 (Davies et al. 1993), as well as that of E and P selectins (Dong et al. 1998), has shown association with the risk of atherosclerosis. However, Wang et al. (2006) reported that soluble VCAM1 and E selectin may not be associated with the risk of stroke, and Nuotio et al. (2003) reported a negative association of the expression of adhesion molecules ICAM1, VCAM1, and selectins in symptomatic carotic stenosis plaques.

The relationship of sTM with CVD is complex. Low sTM may contribute to the risk of MI (Salomaa et al. 1999, Wu et al. 2003), but elevated sTM is suggested to relate to the progression of atherosclerosis in CHD patients (Blann et al. 1997). Olivot et al. (2004) suggested that high sTM may be protective for ischemic stroke in the individuals previously free of stroke, but in contrast, in those with prior stroke events, reflect the fatal outcome. Recently Zhang et al. (2007) reported that high sTM may relate to the risk of ischemic stroke, whereas Jauch et al. (2006) stated that sTM may not contribute to ischemic strokes. Soluble thrombomodulin concentrations may reflect natural anticoagulation in the healthy individuals, but in the diseased individuals, sTM may relate to extend of the disease (Olivot et al. 2004, Constans and Conri 2006).

Although several molecular markers show association with the risk of cardiovascular outcomes, the clear predictive value of these factors remains unclear. Adhesion molecules ICAM1, VCAM1, and E and P selectins (Malik et al. 2001), and CRP (Lloyd-Jones et al. 2006 b) may not add substantially to the traditional risk factors in predicting CHD, and according to Matijevic and Wu (2006), the true causal relationship between the markers like fibrinogen, PAI1, von Willebrand factor, FVIII, and sTM with stroke events has not been established.

Cardiovascular diseases belong genetically to common traits, where the interplay of genetic and environmental factors influences the disease outcome. The genetic component of CVD is discussed in detail in the next section. Figure 5 illustrates the complex interplay of the risk factors in atherogenesis.

***Figure 5**. Complex interplay of cardiovascular risk factors in atherogenesis, finally manifesting as cardiovascular events.*



3. CARDIOVASCULAR GENETICS

Genetic factors contribute significantly to the epidemiology of both CHD and ischemic stroke events. Marenberg et al. (1994) estimated in a large twin study that when one twin suffered a fatal MI, the risk for a fatal MI for the other monozygotic twin was 8-fold, and for a dizygotic twin nearly 4-fold when compared with the risk of the general population. Family history of stroke or transient ischemic attack increased the stroke risk 1.5–2.5-fold (Goldstein et al. 2006). Heritability measures the degree the genetic factors explain of the variance in a phenotype. The heritability of death from stroke in Danish families was approximately 30% (Bak et al. 2002). While environment affects the majority of well-established factors contributing to CVD risk (Table 2), these factors themselves belong genetically to complex traits: Heritability estimates lay between 30 and 70 for obesity (Loos and Bouchard 2003), between 40 and 80 for T2DM (Permutt et al. 2005), between 30 and 50 for hypertension (Koskenvuo et al. 1992, Hong et al. 1994), and between 40 and 80 for all LDL, TC, and HDL levels (O'Connell et al. 1988, Heller et al. 1993). Evidence

has also recently accumulated for genetic background of nicotine dependency and smoking behavior (Li et al. 2006, Loukola et al. 2007).

Moreover, there is considerable overlap in the genes predisposing to different traits, further complicating the complexity of CVD genetics. For example, upstream transcription factor -1 (*USF1*) variants are associated with FCHL, different dyslipidemias, diabetes, and CVD, and *ICAM1* variants are related to CVD, type 1 diabetes, asthma, and different cancers. Both genes are discussed in detail below.

3.1. Rare and common cardiovascular and metabolic traits

Viewing genes linked to rare monogenic CVD traits may help to unravel the complicated background of common diseases. Although many of the causative mutations are extremely rare, often even family-specific, these unique disease forms have provided valuable information on the pathophysiology involved in common disease processes (Peltonen et al. 2006).

MODY represents a familial form of diabetes. The disease manifests in early adulthood, accounting for 2-5% of all diabetes cases (Ledermann 1995). Six different genes have been identified as contributing to MODY, all critical for pancreatic cell development (McCarthy 2004). Reflecting the underlying mutation, MODY cases express genetic and clinical heterogeneity (Stride and Hattersley 2002). Studies on MODY have shed light on the pathogenesis of diabetes also in a wider setting.

USF1 gene provides a clear example of how studies ascertained for special rare families can pinpoint candidate genes for common traits. Chromosomal region 1q21-23 was first linked to FCHL in Finnish families (Pajukanta et al. 1998). Closer studies on the region revealed a new FCHL candidate gene, *USF1* (Pajukanta et al. 2004). Subsequent studies provided evidence for functional significance of the associated allele (Naukkarinen et al. 2005) and supported the influence of *USF1* variants on CVD and mortality also at a population level (Komulainen et al. 2006). Interestingly, other studies have reported an association of *USF1* single nucleotide polymorphisms (SNPs) also with MetS and T2DM, suggesting some genetic overlap of these disease groups (Gibson et al. 2005, Ng et al. 2005).

3.2. Some CVD-related loci

Several traits contributing to formation of atherosclerosis, giving rise to CVD exist. The target genes of this thesis contribute to dyslipidemias (*USF1*), thrombosis (*F5*, *PROC*, *THBD*), and endothelial dysfunction (*ICAM1*). The most important genes of these traits in the perspective of cardiovascular outcomes are discussed in the following sections.

3.2.1. Genes related to dyslipidemias

Characteristic of dyslipidemia genes is an association with more than one lipid trait. Many genes have first been linked to rare familial disorders. One of the most famous examples is LDL receptor gene (*LDLR*), whose point mutations predispose to familial hypercholesterolemia (Hegele 2007). More recently, *LDLR* has been linked to Alzheimer's disease (AD) (Kong et al. 2006, Bertram et al. 2007) and HDL and LDL traits (Knoblauch et al. 2004). ATP-binding cassette -1 (*ABCA1*) gene was first identified in rare HDL-deficiency with autosomal recessive inheritance, Tangier's disease (Brooks-Wilson et al. 1999). Subsequently, *ABCA1* has been associated with low HDL concentrations in the general population, as well as with AD (Katzov et al. 2004) and CVD (Clee et al. 2001, Frikke-Schmidt et al. 2005). *APOE* epsilons are among the most studied genetic variants in dyslipidemias and CVD (reviewed by Eichner 2002). The epsilons are associated with TC, LDL, and ApoB concentrations, as well as with AD, CVD, stroke, and peripheral artery disease.

Familial combined hyperlipidemia is currently a hot topic in lipid genetics. The genetic background of FCHL is complex (reviewed by Naukkarinen et al. 2006, Suviolahti et al 2006). In family studies, three main chromosomal regions, 1q21-23, 11p, and 16q22-24, have been linked to FCHL. Replicated evidence for association with FCHL exist for lipoprotein lipase gene (*LPL*) (Naukkarinen et al. 2006, Suviolahti et al 2006), apolipoprotein A1/C3/A4/A5 gene cluster (*APOA1/C3/A4/A5*) (Naukkarinen et al. 2006, Suviolahti et al 2006), and *USF1* (Pajukanta et al. 2004, Coon et al. 2005, Huertas-Vazquez et al. 2005, van der Vleuten et al. 2007).

USF1 has lately been one of the most studied CVD-related genes. As a transcription factor, it influences some 40 genes, many participants of body lipid or glucose metabolism (Naukkarinen et al. 2005). *USF1* was first linked to dyslipidemias in a Finnish sample of FCHL families (Pajukanta et al. 2004), where two *USF1* SNPs were associated with TG and TC levels in males. Subsequently, several family studies have

indicated positive associations of *USF1* SNPs and haplotypes with different dyslipidemias (Coon et al. 2005, Huertas-Vazquez et al. 2005, Komulainen et al. 2006, van der Vleuten et al. 2007). *USF1* variants also showed association with T2DM/MetS in Chinese families (Ng et al. 2005), but results from a French family sample, were by contrast, negative (Gibson et al. 2005). At a population level, *USF1* variants were associated with CVD events and mortality in women in two Finnish population cohorts (Komulainen et al. 2006). In males, however, population studies on *USF1* and T2DM/MetS and CVD have yielded negative results (Gibson et al. 2005, Ng et al. 2005, Komulainen et al. 2006, Zeggini et al. 2006). Functional studies suggest that *USF1* downregulates the *APOE* gene in individuals with dyslipidemias and may in this way contribute to atherogenesis (Naukkarinen et al. 2005).

Despite the abundance of genes related to different dyslipidemias, direct associations with CVD events are scarce; replicated evidence exists only for *APOE* (Eichner 2002) and *ABCA1* (Clee et al. 2001, Frikke-Schmidt et al. 2005), followed by single findings with *LPL* (Leshinsky-Silver et al. 2006) and *USF1* (Komulainen et al. 2006). With *APOA1/C3/A4/A5*, the findings regarding CVD have been inconsistent (reviewed by Lai et al. 2006).

It is nevertheless worth pointing out that all of the strong FCHL candidates have shown some association with CVD. Other studies have in addition suggested a potential common genetic background for FCHL, metabolic syndrome, and T2DM (Gibson et al. 2005, Ng et al. 2005). These findings make the FCHL-linked genes tempting candidates for common traits beyond familial dyslipidemias. Table 4 lists some important genes linked to dyslipidemias.

3.2.2. CVD-related thrombosis and cell adhesion molecule genes

Genes related to arterial thrombosis include blood coagulation factors, fibrinolytic factors, and platelet-membrane receptors (reviewed by Lane and Grant 2000, Voetsch and Loscalzo 2004). Characteristic of thrombosis genes and CVD is that although plasma concentrations of various hemostatic agents are associated with CVD, the link between soluble protein concentration and gene variants, and further, with gene variants and CVD is often somewhat unclear.

Elevated fibrinogen levels are consistently linked with increased CVD risk. Although fibrinogen gene variants are associated with fibrinogen levels, the relationship of the SNPs and CVD is less well established (Lane and Grant 2000). Several fibrinogen-β (*FGB*) polymorphisms are suggested to be associated with

CVD, physiologically the most significant being the binding-site altering promoter region SNPs -445G/A and -854G/A (Voetsch and Loscalzo 2004). Results from very large studies on *FGB* SNPs and CVD have, however, been negative (Lane and Grant 2000, Voetsch and Loscalzo 2004). Same is true for fibrinogen-α polymorphism Thr312Ala (Voetsch and Loscalzo 2004).

Table 4. *Some genes linked to dyslipidemias.*

GENE	TRAIT	REFERENCE
ABCA1	HDL, CVD, AD, Tangier's disease	1-4
APOA2	HDL	5-6
APOA1/C3/A4/A5	FCHL, HDL, TG	7-9
APOB	HDL, LDL, TC	11
APOE	ApoB, LDL, TC, CVD	9, 12-14
CETP	HDL, LDL, HDL/LDL, CVD	9, 11, 15
FOXC2	BMI, HDL, TG, obesity	16
HNF1A	T2DM	17
HNF4A	T2DM, TC, TG, MetS, glucose markers	18-22
LCAT	FCHL, HDL	23-24
LDLR	FH, AD, HDL, LDL	11, 25-26
LEPR	FCHL, ApoB, HDL	27-29
LIPC (*HL*)	HDL, ApoB, TC	30-31
LPL	HDL, Familial hyperchylomicronemia, CVD, AD	32-34
PCSK9	LDL	35-36
PPARA	FCHL, LDL, MetS	9, 37
SOAT1	HDL, AD	9, 32
TNFRSF1B	FCHL	38
USF1	FCHL, LDL, TC, TG, MetS, BMI, lipolysis, CVD	39-47

References: 1) Brooks-Wilson et al. 1999, 2) Clee et al. 2001, 3) Cohen et al. 2004, 4) Frikke-Schmidt et al. 2005, 5) Lilja et al. 2002, 6) Welch et al. 2004, 7) Gagnon et al. 2003, 8) Pennacchio et al. 2003, 9) Klos et al. 2006, 10) Lai et al. 2006, 11) Knoblauch et al. 2004, 12) Eichner 2002, 13) Martins et al. 2006, 14) Stengard et al. 2006, 15) Leshinsky-Silver et al. 2006, 16) Carlsson et al. 2004, 17) Winckler et al. 2005, 18) Hansen et al. 2005, 19) Damcott et al. 2004, 20) Love-Gregory et al. 2004, 21) Silander et al. 2004, 22) Weissglas-Volkov et al. 2006, 23) Zhang et al. 2004, 24) Aouizerat et al. 1999, 25) Kong et al. 2006, 26) Bertram et al. 2007, 27) Allayee et al. 2002, 28) Norman et al. 1998, 29) van der Vleuten et al. 2007, 30) Allayee et al. 2000, 31) Hoffer et al. 2000, 32) Papassotiropoulos et al. 2005, 33), 34) Leshinsky-Silver et al. 2006, 35) Cohen et al. 2006, 36) Fasano et al. 2006, 37) Eurlings et al. 2003, 38) Geurts et al. 2000, 39) Pajukanta et al. 2004, 40) Coon et al. 2005, 41) Huertas-Vazquez et al. 2005, 42) van der Vleuten et al. 2007, 43) Komulainen et al. 2006, 44) Putt et al. 2004, 45) Gibson et al. 2005, 46) Ng et al. 2005, 47) Zeggini et al. 2006.

ABCA1: ATP-binding cassette -1; AD: Alzheimer's disease; *APOA2*: Apolipoprotein A2; *APOA1/C3/A4/A5*: Apolipoprotein A1/C3/A4/A5 gene cluster; *APOB*: Apolipoprotein B; *APOE*: Apolipoprotein E; *CETP*: Cholesterol ester transfer protein; *FOXC2*: Forkhead box C2; *HNF1A*: Hepatic nuclear factor -1α; *HNF4A:* Hepatic nuclear factor -4α; *LCAT*: Lecithin:cholesterol acyl transferase; *LDLR*: LDL receptor; *LIPC* (*HL*): Hepatic lipase; *LPL*: Lipoprotein lipase; *PCSK9*: Proprotein convertase subtilisin/kexin type -9; *PPARA*: Peroxisome proliferator-activated receptor α; *SOAT1*: Sterol o-acyltransferase -1; *TNFRSF1B*: Tumor necrosis factor receptor subfamily 1 member β; *USF1*: Upstream transcription factor -1.

Coagulation factor VII (*F7*) variants, the most important ones being Arg353Gln, hypervariable region 4 (HVR4), -401G/T, and -402G/A, are estimated to explain about 30% of the F7 plasma level variation. Again, the connection of the *F7* SNPs to the actual disease outcome, CVD events, is somewhat unclear (Lane and Grant

2000). *FXIII* Val34Leu has been repeatedly associated with decreased MI risk (Kohler et al. 1998, Wartiovaara et al. 1999).

The Leiden mutation of coagulation factor V gene (*F5*) was characterized in 1994 (Voorberg et al. 1994, Zöller et al. 1994). The polymorphism G1691A results to the substitution of arginine for glutamine at amino acid site 506, which acts as a cleavage site for APC. The *F5* Leiden polymorphism creates APC resistance and strongly increases the risk for venous thrombosis. Heterozygotes for the *F5* Leiden have approximately a three-fold risk, whereas for homozygotes, the risk is estimated to be over ten-fold (Lane and Grant 2000, Simmonds et al. 2001). Prevalence of the Leiden mutation in Caucasians ranges from 1% to 15% (Rees et al. 1995, Ridker et al. 1997, Franco et al. 1999), and the mutation is suggested to be present in 10-50% of the deep venous thrombosis cases (Franco et al. 2001). Additional mutations, such as a *F5* haplotype H2R and other point mutations affecting the APC cleavage site, have also been described (Bernardi et al. 1997, Chan et al. 1998, Williamson et al. 1998).

Studies on *F5* Leiden and CVD have been controversial. The Leiden mutation and prothrombin (*F2*) G20210A polymorphism associated with CHD in a recent large meta-analysis (Ye et al. 2006). *F5* is a large gene spanning over 70 kilobases (kb), located at chromosome (chr) 1q23. Data on gene variants other than the Leiden mutation are scarce.

Plasminogen activator inhibitor -1 contributes to thrombosis as a potent inhibitor of tissue plasminogen activator. High PAI1 concentrations are associated with CVD risk, especially in persons with diabetes (Lane and Grant 2000). *PAI1* gene promoter variant -675 4G/5G has been widely studied, but although this variant contributes to soluble PAI1 levels, the association with CVD remains unclear (Lane and Grant 2000, Ye et al. 2006). Similarly, *PLAT* SNP -7351C/T has been studied extensively, but the association with CVD is unclear (Lane and Grant 2000). A recent study suggested that two *PLAT* haplotypes increase CVD risk (Kathiresan et al. 2006).

Of thrombomodulin (*THBD*) variants, Ala25Thr has been associated with MI (Doggen et al. 1998), but not with cerebrovascular disease (Warner et al. 2000). An association of Ala455Val with MI was reported by Ireland et al. (1997), with CHD by Wu et al. (2001), and with ischemic stroke by Cole et al. (2004). Konstantoulas et al. (2004) found the combination of -1208–1209TTdelTT and Ala455Val to be associated with MI in men. Chao et al. (2004) stated that -33G/A vatiant may relate to premature MI. Ohlin et al. (2004) concluded that *THBD* may not contribute to acute coronary syndrome. The role of thrombin activatable fibrinolysis inhibitor gene variants in CVD

is unclear (Voetsch and Loscalzo 2004). 5,10-methylenetetrahydrofolate gene (*MTHFR*) polymorphism 677C/T contributes to homocystenemia, but the relation with CVD is unclear (Voetsch and Loscalzo 2004).

Reitsma et al. (1995) have described over 160 polymorphisms at the *PROC* locus on chr 2q13-14. However, little is known of their relation to thrombophilia. Spek et al. (1995) introduced three SNPs on the promoter region of the gene contributing to protein C level. However, Buil et al. (2004) estimated these SNPs to account for only 6% of the variation in plasma protein C levels, and reported a new influencing locus on chr 16.

A recent study suggested an association of *ICAM1* Lys469Glu variant (Podgoreanu et al. 2006) with MI after cardiac surgery. Previously McGlinchey et al. (2004) found no association of this variant with CHD. Recent reports suggest that *VCAM1* (Miyoshi et al. 2007) SNPs may contribute to the development of atherosclerosis. The relation of *SELP* variants with CHD and stroke risk is unclear (Zee et al. 2004, Volcik et al. 2006, Volcik et al. 2007). Results on the role of the glycoprotein (GP) family variants in CVD have been controversial. In a recent large meta-analysis, no association was seen with any of the most studied GP variants *GPIa* C807T, *GPIb* T-5C, *GPIIIa* C1561T (Ye et al. 2006).

Taken together, although concentrations of several hemostatic factors contribute to CVD risk, the same seems to be true only with very few genetic variants; in recent large studies, only the *F5* Leiden mutation and *F2 20210G/A* variant have provided positive results. This might reflect Reitsma's (2004) postulate that inherited hypercoaguability states, such as the factor V Leiden mutation, increase the probability for the first thrombosis event; in the arterial side, this is true in the absence of inflammation and atherosclerosis, i.e. very rarely. When atherosclerosis advances, inherited hypercoaguability loses importance, similar to recurrent venous thrombosis. Table 5 summarizes the most studied hemostatic variants.

Table 5. *Gene variants of thrombosis genes and cell adhesion molecule genes showing association with CVD events*.

Gene	Polymorphism	Reference
F2	20210G/A	Ye et al. 2006
F5	Arg506Gln (The Leiden mutation)	Ye et al. 2006
F7	Arg353Gln, HVR4, -401G/T, -402G/A	Voetsch and Loscalzo 2004
FXIII	Val34Leu	Voetsch and Loscalzo 2004
FGB	-445G/A, -854G/A, *Bcl*1, Thr312Ala	Voetsch and Loscalzo 2004
GPIa	807C/T	Ye et al. 2006
		Voetsch and Loscalzo 2004
GPIbα	-5T/C, Thr145Met	Voetsch and Loscalzo 2004
GPIIIa	Leu33Pro	Ye et al. 2006
		Voetsch and Loscalzo 2004
ICAM1	Lys469Glu	Podgoreanu et al. 2006
MTHFR	677C/T	Voetsch and Loscalzo 2004
PAI1	-675 4G/5G	Lane and Grant 2000
PLAT	-7351C/T	Voetsch and Loscalzo 2004
		Kathiresan et al. 2006
TAFI	Ala147Thr, 1542C/G	Voetsch and Loscalzo 2004
THBD	Ala25Thr	Doggen et al. 1998
	Ala455Val	Wu et al. 2001
	-1208–1209TTdelTT	Ireland et al. 1997
		Konstantoulas et al. 2004

F2=Prothrombin, *F5*=Coagulation factor V, *F7*= Coagulation factor VII, *FXIII*= Coagulation factor XIII, *FGB*=Fibrinogen β, *GPIa*=Integrin α-2, *GPIbα*=Glycoprotein Ib, platelet, α-polypeptide, *GPIIIa*=Integrin β-3, *ICAM1*=Intercellular adhesion molecule -1, *MTHFR*=5,10-methylenetetrahydrofolate reductase, *PAI1*=Plasminogen activator inhibitor -1, *PLAT*=Tissue plasminogen activator, *TAFI*= Thrombin activatable fibrinolysis inhibitor, *THBD*=Thrombomodulin.

3.2.3. Genes related to ischemic stroke events

The putative candidate genes for ischemic stroke events comprise genes related to hypertension, inflammation, thrombosis, and dyslipidemias. Accordin to Humpries and Morgan (2004), the most promising stroke candidates include *APOE*, angiotensin converting enzyme, and *MTHFR*. Evidence replicated in several studies exists for these three genes, although some studies have also shown negative results (Humpries and Morgan 2004). A recent meta-analysis by Ariyaratnam et al. (2007) stated that strong evidence exists for the same three genes also in Asians of non-European origin.

Other putative genes contributing to stroke events include interleukin 6 (*IL6*), cholesterol ester transfer protein (*CETP*), hepatic lipase (*HL*), and paraoxinase (*PON1*) (Humpries and Morgan 2004), followed by *PAI1* (Ariyaratnam et al. 2007), *LPL* (Shimo-Nakanishi et al. 2001), gene encoding 5-lipoxygenase activating protein (*ALOX5AP*) (Helgadottir et al. 2004, Worrall and Mychaleckyj 2006), and phosphodiesterase 4D (*PDE4D*) (Gretarsdottir et al. 2003).

4. STRATEGIES TO IDENTIFY GENES FOR CARDIOVASCULAR TRAITS

4.1. Studying complex diseases

Cardiovascular diseases belong to complex traits, where the disease outcome is affected by environmental as well as genetic factors. The genetic structure of complex traits is still largely unknown, and several competing hypotheses prevail. According to the common variant – common disease hypothesis, the genetic basis of complex traits consists of several common but low-penetrance predisposing loci: The numerous risk variants occur frequently in the population, but each of them likely has only a modest effect on the disease phenotype (Collins et al. 1997). According to the competing heterogeneity model, the structure of complex traits is based on rare alleles with major phenotypic effects, whereas the "neutral" hypothesis suggests that the spectrum of disease alleles in complex traits reflects that of all the genomic variants (Wang et al. 2005).

In addition to the issue of unknown inheritance, several other factors complicate the study of complex traits (Table 6). Such genetic factors include genetic heterogeneity, diversity in allele frequencies across different populations, pleiotrophy (the same genetic variant is responsible for more than one disease phenotype), and epistatic coeffects between several genes. Of clinical features, uncertain and diverse phenotypes, difficulties in disease diagnostics, and late disease onset further complicate the study designs. In addition, technical hinders such as genotyping costs and accuracy, and statistical issues like multiple testing should be taken into account, also bearing publication bias towards positive results in mind. (Lander and Schork 1994, Risch 2000).

4.2. Gene identification strategies

Effective identification of genes underlying complex traits is likely to require the use of several different techniques (Miller et al. 2007). According to Lander and Schork (1994), the main strategies in the search for novel loci in common diseases are linkage analyses, association analyses, and experimental animal models. Accordingly, Glazier et al. (2002) proposed a stepwise strategy for gene identification in complex traits: First, a linkage study of the whole genome is performed. Next, area under the linkage peak is fine-mapped to narrow down the linked region. Candidate genes in the region can be sequenced to define allelic

variants. Alternatively, markers for the fine-mapping can be selected from an allelic variation database according to the haplotype block assumption (discussed below). Finally, functional studies are performed with the associated alleles. Advantages and disadvantages of the different gene identification methods introduced below are presented in Table 7.

Table 6*. Several factors complicate gene identification in common disorders compared with monogenic traits. These factors can be divided into genetic, phenotypic, and other factors related to statistical and technical issues.*

SOME FACTORS COMPLICATING THE STUDY OF COMPLEX TRAITS	
Genetic factors	Unknown inheritance pattern
	Unknown allele frequencies
	Genetic heterogeneity
	Epistasis
Phenotypic factors	Diagnostic difficulties
	Late onset of a disease
	Pleiotrophy
	Phenocopies
	Unknown penetrance
Statistical factors	Limited sample size
	Multiple testing
Technical issues	Genotyping costs
	Genotyping errors
Publication bias	Negative results do not arouse sufficient interest to be published

4.2.1. Linkage studies

Linkage studies are a common way to search for new disease loci (Curtis et al. 1995). A whole-genome linkage scan contains markers across the genome, analyzed in a family sample. Linkage studies look for co-segregation of a genetic marker and a trait. Linkage studies are commonly divided into parametric or nonparametric methods. The parametric approach requires estimation of inheritance model and disease allele frequency and penetrance in the population. This is often difficult in complex traits, which more commonly rely on nonparametric methods, based on allele sharing between relatives sharing the same phenotype. In diseases with late onset, ancestral information is often lacking, and a large number of siblings are needed to gain sufficient power. This may prove problematic if the disease of interest is rare. In addition, the marker coverage of a whole-genome linkage scan is relatively poor for association purposes and requires further focusing with fine-mapping after a positive linkage signal (Lander and Schork 1994, Glazier et al. 2002, Miller et al. 2007).

4.2.2. Linkage disequilibrium and population isolates

Two alleles from separate loci inherited together more often than expected by chance only are in linkage disequilibrium (LD). R^2 is a measure for LD, reflecting correlation between two different variables. Several driving forces for LD exist, including population founder effects, recent mutations, and natural selection favoring certain alleles (Slatkin 1994).

LD patterns and allelic spectra differ between populations. Small, long-isolated, and stable-sized founder populations are observed to hold extended LD distances compared with more heterogeneous populations (Laan and Pääbo 1997, Service et al. 2006). Isolated populations are also likely to possess fewer disease alleles and to express greater extend of allele-sharing around the disease allele (Peltonen et al. 1999, Peltonen 2000, Varilo et al. 2003). LD patterns provide means of discovering linkage with a marker and a disease without genotyping the actual disease-causing locus (indirect association), reducing the number of markers necessary to cover the genome in a study.

Population isolates have indeed been useful in mapping genes for rare monogenic diseases. Population isolates may prove useful also in complex traits (Collins et al. 1997, Peltonen 2000, Amos 2007), although the case is likely to be less straightforward than with monogenic disorders, and the LD patterns around disease-causing variants in founder populations may not differ extensively from mixed populations (Hirschhorn and Daly 2005). Isolated populations, however, hold an important advantage; the similarity of lifestyle factors likely reduces the noise in genetic studies (Peltonen 2000).

4.2.3. The Finnish population and genetic studies

The Finns as a population offer several advantages in the study of genetics. Finns are a typical isolated population. The country was inhabited in two major waves: The early settlement, approximately 2000 years ago, inhabited the southern part of the country, following the coastal line (Nevanlinna 1972, Varilo et al. 2003). The late settlement took place in the 16th century and was characterized by internal migration from a small region in the southeast of the country to northern and western parts, forming small, long-isolated and stable-sized subpopulations. Some of these subisolates present accumulation of certain disease alleles, giving rise to regional differences in the disease prevalence within the country (Pastinen et al. 2001). The founder effect has facilitated the study of many monogenic traits

(Peltonen et al. 1999). In addition to the nearly 2000 years of isolation, bottlenecks such as wars and famines have molded the Finnish population and regional gene pools.

The Finns also have other features that may pave the way in the study of monogenic and complex traits. The opinion of the well-educated population towards medical studies tends to be open-minded, yielding high study participation rates. Finland has high healthcare standards, and the relatively homogeneous education for medical professionals facilitates the disease diagnostics and phenotyping. Availability of medical and genealogical records dates back several hundred years. Table 8 summarizes the main advantages.

Table 8. *The Finnish population is well suited well for genetic studies due to reduced genetic and environmental heterogeneity.*

ADVATAGENS OF THE FINNISH POPULATION IN GENETIC STUDIES	
Reduced number of disease alleles:	Founder effect
	Isolation
	Bottlenecks
Accurate healthcare data and sampling:	Homogeneity in disease diagnostics
	High standard of medical records
	Long availability of genealogical records
Homogeneity in lifestyle factors	
Positive attitude towards medical research	

4.2.4. Haplotypes

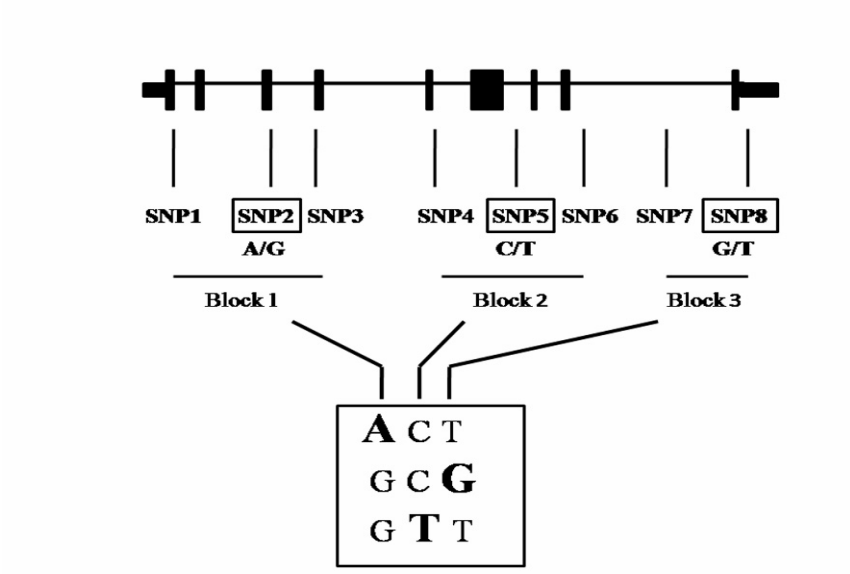
Completion of the Human Genome Project has opened new gateways to genetic research, one of the most significant being the the International HapMap project. HapMap's aim is to produce a physical map of the human genome by providing a limited number of "tag-variants" across the genome (www.hapmap.org). Tag-variants are mainly SNPs, i.e. tag-SNPs, but also include small insertions and deletions. The idea of HapMap is based on uneven distribution of recombinations in the genome, creating distances of genomic sequence, "blocks", where recombinations are only seldom observed (Daly et al. 2001). The SNPs in each block are in substantial LD. Cover most genetic variation at a given locus is therefore possible by genotyping one SNP in each block of the locus.

Each tag-SNP represents a block, tagging one haplotype at the locus. Haplotype refers to the nucleotide structure in a single DNA strand. Within each haplotype block, increased LD and reduced recombination rate are observed. According to the HapMap theory, one or a few polymorphism(s) from each block is therefore

sufficient to capture most of the common variation within a block. This theory is based on several simplifications and has also been criticized (Terwilliger and Hiekkalinna 2006).

The tag-SNP may itself prove a causative variant for a disease (direct association). Alternatively, the tag-SNP can be in LD with the actual causative variant (indirect association). Figure 5 gives an example of tag-SNPs defining haplotypes at a gene locus.

Figure 6. *The imaginary gene locus contains 8 SNPs, which can be divided into three haplotype blocks: SNP 2 is in substantial LD with SNPs 1 and 3, and therefore tags block 1. SNP 5 tags haplotype block 2, representing SNPs 4 and 6. SNP 8 tags the third haplotype block. Thus, tag-SNPs 2, 5, and 8 cover the allelic variants of this gene locus. Together, these tag-SNPs form three haplotypes, ACT, GCG, and GTT. Allele A of SNP 2 is only present in haplotype ACT, whereas allele G of SNP 2 appears in the other two haplotypes. The Allele A therefore tags the haplotype ACT: Allele A of SNP 2 always appears together with allele C of SNP 5 and allele T of SNP 8.*



4.2.5. Candidate gene studies

Linkage studies are utilized to define a chromosomal region linked with the disease studied. The linked region may contain potential candidate genes for the disease. The region of interest can then be further addressed by fine-mapping, forming a tight marker map across the candidate gene locus to capture its allelic variation. Alternatively, interesting candidate genes can be selected based on their known functions. For example, soluble plasma coagulation factors involved in the clotting cascade contribute to thrombotic events, and the genes encoding these proteins are intriguing candidates for venous and arterial thrombosis.

Association analysis is commonly used in candidate gene studies. In association analysis, allele frequencies of a marker are compared between the disease and control groups (case-control analysis). Association analysis can be utilized in both family and population samples. The detected association can then be direct or indirect, LD-based. Moreover, natural selection, where a certain allele of a marker is favored in the population, population stratification, and false-positive findings can cause the allele frequencies to differ between the cases and the controls, and should be taken into account as possible causes for association.

Advantages of candidate gene study are customized selection of markers and lower genotyping costs compared with linkage studies based on genome scans. Candidate gene studies are, however, narrow in perspective, focusing only on the selected genes.

4.2.6. Genome-wide association studies

Due to completion of the Human Genome Project and developing genotyping techniques, larger study samples and wider marker sets are now available for analysis. In genome-wide association studies (GWAS), a panel of SNPs across the whole genome is selected for genotyping. GWAS covering the whole genome can be though of as an alternative to genome scan -based linkage studies. GWAS can, however, utilize population data, whereas linkage studies are tied to family samples. Several commercial SNP panels exist, covering 100 000 – 1 000 000 SNPs. The SNP selection can concentrate on tag-SNPs, coding SNPs, or equal SNP distribution across the genome.

Genome-wide coverage decreases the need for “educated guessing” or the *a priori* hypothesis utilized in candidate gene studies to select the regions genotyped. Interestingly, as Hunter and Kraft (2007) point out, the most robust associations from the GWAS have been with novel regions, instead of those already reported;

even including regions previously thought not to contain any genes at all. GWAS may thus reveal novel candidate genes as well as whole pathways contributing to the disease pathophysiology (Drazen and Phimister 2007, Hunter and Kraft 2007). Although the genome-wide association studies greatly facilitate the genotyping task, they face severe bioinformatic challenges (discussed in section 3.2.9) and require multiple rounds of replication (Evans and Cardon 2006). In addition, gaps are likely to remain in the coverage, and valuable information may thus be missed. Amos (2007) states that key factors to successful GWAS include use of large, homogeneous population samples and careful dissection of the phenotype.

Currently, results from genome-wide association studies related to cardiovascular traits are being published at an accelerating pace, almost on a weekly basis during the preparation of this thesis. Ozaki et al. (2002) observed an association of lymphotoxin -α gene with MI, and Samani et al. (2007) reported several chromosomal regions associated with CHD. Herbert et al. (2006) discovered a common variant associated with obesity, and Hakornarson et al. (2007) noted that *KIAA0350* gene was associated with type 1 diabetes.

4.2.7. Gene expression studies

Gene expression studies address the functional role of a variant of interest, typically first identified with linkage or association studies. Based on messenger RNA levels, expression studies provide first-hand kownledge on the potential pathophysiological changes of a variant at a tissue level. Compared with candidate gene studies targeting one or a few genes, tailored expression arrays facilitate the analysis of entire pathways and address the role of variants studied as a process rather than as separate pieces. Disadvantages of expression studies include difficulty in obtaining the target tissues. Dissecting specific brain regions in animals, for example, is challenging, not to speak of the difficulty of obtaining such tissues from humans. The levels of mRNA may in addition not fully correlate with the actual protein levels.

4.2.8. Interaction studies

Several different genes are likely to contribute to the genetic basis of complex traits. A meaningful way to address this complexity therefore is to simultaneously investigate a number of putative risk variants and their relations, instead of single genes.

Epistasis, interaction between two or more genes, has recently been a hot topic in genetic studies. Epistasis is one of the factors complicating the study of complex traits compared with monogenic diseases; two interacting loci can alter – mask or

enhance – each other's effects, making it difficult to recognize the loci contributing to the traits. This complexity substantially increases when hundreds or thousands of potentially contributing loci are taken into account.

Table 7. *Advantages and disadvantages of some common gene identification methods. Different gene identification strategies address different research goals. Genome scan -based linkage studies on family data have been useful in identifying genes for monogenic traits. The linked region can then be fine-mapped for variants in a candidate gene study. Association studies can utilize both family and population data. Genome-wide association studies cover thousands of SNPs across the genome, facilitating the genotyping task. Expression studies focus on the functionality of the variants (modified from Miller et al. 2007).*

STUDY TYPE	ADVANTAGES	DISADVANTAGES
Linkage study in families	Information genome-wide	Requires family samples: problematic in very rare diseases
	Identifies large-scale genetic effects	Low resolution: fine-mapping needed
		Limited power in complex traits
Association study in population-based samples	More robust in complex traits than linkage studies	Requires large samples to detect small effects
	High resolution	Requires replication samples
	Suitable for both population and family samples	
Candidate gene study	Customized SNP selection	Analysis of the selected gene(s) only
	Lower costs compared with genome-wide studies	
Genome-wide association study	Information genome-wide	Statistical challenges
	Detection of small genetic effects	Questionable overage
	Effective genotyping	Genotyping costs
Expression study	Functional information at tissue level	Difficulty in obtaining target tissues: small sample size
	Analysis of whole pathways: customized arrays	mRNA[a] and protein levels may not correlate
		Bias due to chip selection

[a] mRNA: Messenger RNA

Essentially, epistasis indicates "not independent". Several definitions and interpretations for epistasis do, however, exist. Cordell (2002) reviewed this confusing issue. The term epistasis was first introduced by Bateson in 1909 to describe a situation where a variant at locus A prevents another variant, B, from expressing its effects. The difference between epistasis and traditional models of inheritance is the location of the alleles: The epistatic variants are at different sites, whereas alleles of the same locus determine dominant and recessive inheritance patterns. In 1918, Fisher gave another definition for epistasis as deviation from additivity. According to Cordell, Bateson's definition corresponds closely to biological interaction whereas Fisher's definition is closer to mathematical interaction models.

Currently, three commonly used interaction models exist to define the outcome, the penetrance (p_{ij}) of genotype i at locus α and genotype j at locus β:

an additive model (Risch 1990):

$$p_{ij} = \alpha_i + \beta_j \qquad (I),$$

a multiplicative model (Hodge 1981, Risch 1990):

$$p_{ij} = \alpha_i \beta_j \qquad (II),$$

and a heterogeneity model (Risch 1990, Neuman and Rice 1992):

$$p_{ij} = \alpha_i + \beta_j - \alpha_i \beta \qquad (III).$$

The multiplicative model can be used when the different contributing loci are not considered independent of each other; the additive and heterogeneity models, in turn, assume at least some extend of independence in the loci (Cordell 2002).

The challenge in complex trait studies is to achieve sufficient statistical power to detect minor effects in the numerous loci contributing to the disease. The ability to detect such minor effects can be enhanced by interaction studies and by selecting a suitable model for interaction (Cordell et al. 2001), although one should proceed cautiously when drawing biological conclusions from statistical interaction.

No clear consensus regarding how to address interaction in genetic studies exists since the models of inheritance remain mainly unknown. In recent works, the multiplicative model has been the dominant choice for interaction, although the results have been negative. Weedon et al. (2006) analyzed three common variants previously suggested to predispose to T2DM using deviation from the multiplicative model as a measure for interaction, but found no evidence of epistasis between the loci. Maier et al. (2005) tested several variants of three type 1 diabetes susceptibility genes under the multiplicative model, but found no evidence of interaction either. Cordell et al. (2001) tested all three common interaction models using two pairs of susceptibility loci for type 1 diabetes, and found evidence in favor of the additive model, rejecting the multiplicative one.

Classification trees are one possible approach to study interaction. Classification trees attempt to find a variable best predicting a desired disease outcome from a dataset, also determining a cut-off point for this variable. The first (best-predicting) variable then becomes a root node for a tree-like structure. The data are subsequently divided into smaller groups. The branches of trees grow to define subgroups holding certain features; in the case of complex traits, mixtures of genetic

and environmental risk factors. Classification trees are thus useful in structuring noisy datasets with numerous variables, such as genetic data containing several SNPs. Gruenewald et al. (2006) used classification trees to identify factors contributing to mortality in the elderly and discovered several risk-increasing biomarker combinations. Among complex trait genetic studies, Baessler et al. (2007) analyzed epistatic effects of several genes in CHD using classification tree analysis and found evidence of SNP interactions predisposing to MI.

Besides gene-gene relations, genes can interact with environmental factors such as smoking or nutritional products. Evidence of gene-environment interactions have been observed, for example, with interleukin -1 beta gene and fatty acids in development of metabolic syndrome (Shen et al. 2007), with tumor necrosis factor alpha gene and smoking in asthma (Wu et al. 2007), with thrombomodulin and smoking (Li et al. 2002) in CHD, and with epoxide hydrolase gene and smoking in coronary atherosclerosis (Wei et al. 2007).

4.2.9. Multiple testing

Multiple testing is a major challenge in current genetic approaches. Genome-wide association studies may comprise over 1 000 000 SNPs, exponentially increasing the number of tests compared with candidate gene studies, which include only a few variants. Achieving sufficient statistical power in such large-scale studies requires extremely large sample sizes; several thousand cases and controls are likely needed to identify relatively small genetic effects in complex traits (Amos 2007). Traditional corrections for multiple testing directly adjust a stringent p-value based on the number of tests performed. With Bonferroni correction, the significance level for GWAS with 1 000 000 SNPs can be set to 10^{-7} (p=0.05/1 000 000). Such conventional methods are stated overly conservative, since LD between the nearby variants analyzed likely exists (Conneely and Boehnke 2007). The contributions of single variants to a common trait are likely to modest, with risk ratios around 1.5 (Hunter and Kraft 2007), and such low-risk loci may not reach the stringent "genome-wide significance" of 10^{-7}. Thus, alternative methods to validate the findings of large-scale SNP studies are warranted.

Replication of the results in several different study samples increases the probability of a true association (Hunter and Kraft 2007). Currently, however, extensive inconsistency in findings of complex trait studies exists; different studies may analyze different variants of the same gene, and even when a certain variant is linked

to a disease, the associating alleles may differ (Clarke et al. 2007). Morgan et al. (2007) addressed this issue by analyzing 85 variants previously associated with acute coronary syndrome, observing a modest association with a single variant. Similarly, Shiffman et al. (2006) were able to replicate only 15 of the 172 previously CHD-associated variants.

False-positive findings are likely to exist in the majority of the published studies. False discovery rate (FDR) method sets a selected, study sample and analysis group-specific threshold for statistical significance. Among the findings passing the selected FDR limit, a certain number may prove to be false-positive (Benjamini and Yekutieli 2005). The FDR limit can, for example, be set to 10%. FDR does not, however, identify which of the positive findings are the false ones. Other methods to correct multiple testing include permutation, where a sample-specific threshold for statistical significance is calculated based on simulations performed (Conneely and Boehnke 2007). Functional studies on associated variants also provide evidence of a true association.

AIMS OF THE STUDY

The purpose of this thesis work was to investigate the role of five biologically relevant candidate genes in cardiovascular diseases at the population level using large study samples from the Nordic countries. Specific aims of the study were as follows:

- To investigate the role of thrombomodulin variants in coronary and ischemic stroke events (I),
- To investigate the epistatic effects of *F5*, *ICAM1*, *PROC*, and *THBD* in coronary and ischemic stroke events (II),
- To clarify the role of the allelic variants of *USF1* in cardiovascular diseases, metabolic syndrome, and its component traits (III).

MATERIALS AND METHODS

1. STUDY SAMPLES

1.1. FINRISK studies

The FINRISK samples are independent and random population samples collected in Finland every five years (www.ktl.fi/finriski). The aim of the FINRISK project is to evaluate cardiovascular risk factors at a population level.

This study utilized two FINRISK samples drawn in 1992 (Vartiainen et al. 1993) and 1997 (Vartiainen et al. 1998). The FINRISK-92 sample comprises 5999 participants from four different geographical regions: Helsinki region, southwestern Finland (Turku-Loimaa), North Karelia, and Kuopio region. At baseline, the participants were 25-64 years of age. FINRISK-97 contains 8141 participants collected from the same four regions and in addition from the province of Oulu. These participants were aged 25-74 years at baseline. The participation rate was 76% in FINRISK-92 and 73% in FINRISK-97. Data on smoking, medication, previous CVD events, and family history was collected at baseline with a self-administrated questionnaire. All participants underwent a physical examination with blood pressure and anthropometric measurements. Biological measurements performed at baseline include serum lipids (at semi-fasting state) and CRP.

The FINRISK-92 cohort was followed for 10 years (1992-2001). The follow-up time for FINRISK-97 was seven years (1997-2003). Fatal and nonfatal cardiovascular events were monitored using the National Causes of Death Register, the National Hospital Discharge Register, and specific registers for coronary (FINAMI, Salomaa et al. 2003) and stroke events (FINSTROKE, Sivenius et al. 2004). International Classification for Diseases (ICD) coding was used to recognize CVD events appropriate for the study. ICD-9 codes for fatal coronary events were 410-414 and 798 and for nonfatal events 410-411, together with ICD-10 codes I20-I25, I46, R96, R98, and R99 (fatal CHD) and I20.0, I21, and I22 (nonfatal CHD). For fatal and non-fatal ischemic stroke events, ICD-9 codes 433 and 434 (excluding 4330X, 4331X, 4339X, and 4349X) and ICD-10 code I63 were used. ICD-9 was utilized until January 1996 in Finland.

1.1.1. Case-cohort study design and genotypic sample

The FINRISK studies utilize a nested case-cohort study design, where a random subsample, a subcohort, is selected from the original study cohort to represent the general study population, i.e. the Finns. The case-cohort design differs from the classic case-control setting in that the selected subcohort also comprises diseased participants – like any normal population.

Altogether 14 140 people participated in the FINRISK-92 and FINRISK-97 studies. The genotypic sample comprised 2222 participants selected with the following criteria:

1) Everyone with first-ever coronary event (n=401) or ischemic stroke (n=149) during follow-up and free of CVD at baseline was selected for genotyping as incident CVD cases.
2) Everyone deceased for any reason during follow-up was selected as all-cause mortality cases (n=610).
3) Individuals with CVD events at baseline (n=662).
4) Subcohorts were selected to represent the original study populations (n=400, FINRISK-92 and n=386, FINRISK-97).

1.2. ULSAM sample

The Uppsala Longitudinal Study of Adult Men (ULSAM) is a population cohort collected in Uppsala County, Sweden (www.pubcare.uu.se/ULSAM/index). All men living in the county, born in 1920-1924, were invited to investigations at four different timepoints: at 50, 60, 70, and 77 years of age. All four surveys were performed at Uppsala University Hospital.

The first survey took place in 1970-1973. The response rate was 81.7%; 2322 out of the 2841 invited 50-year-old men participated. After 10 years, in 1981-1984, all the 2130 eligible men aged around 60 years were re-invited to participate. By this time, 98 had died and 94 were not eligible for other reasons, mainly for having moved out of the county. The second survey gathered 1860 participants (87.3% of those eligible). The third survey was arranged in 1991-1995, re-inviting all participants of the first survey in 1970-1973, including those who did not participate in the second survey. The third survey reached 1221 (73%) of the 1681 eligible 70-year-old men.

During the 20 years of follow-up, 422 men had died and 219 had moved from Uppsala County. In 1997, all 1389 eligible men, now around 77 years of age, were once again re-invited to investigations. Altogether 839 men (60%) participated, whereas 748 were deceased, and 176 were ineligible for other reasons.

Each survey contained a self-administrated questionnaire on cardiovascular risk factors and medication. Also at each time point, all participants underwent a physical examination, including measurements of blood pressure and body anthropometric features. Of biological measurements, samples for serum cholesterols and fasting glucose were drawn in all four surveys. ApoA-1, ApoB-100, and Lp(a) were measured at the first and thrid surveys. Lipid samples drawn in 1970-1973 were stored in liquid nitrogen until analyzed in 1981 or 1988 (all apolipoproteins). The values were adjusted with conversion factors: 1.06 for LDL and TC, 0.9 for TG, and 1.17 for HDL, to enable comparison with the Monarch method used in the last two surveys. LDL was calculated with Friedewald's formula at all ages using 4.0 mmol/l as a cut-off limit for TG. The insulin clamp test was performed at the 70 year survey. Table 9 summarizes participant data and measurements at different timepoints.

Table 9. *Characteristics of the ULSAM study design and measurements relevant to the study III (modified from Auro et al. in press).*

	Survey I 50 years 1970-1973 (n=2322)	**Survey II** 60 years 1981-1984 (n=1860)	**Survey III** 70 years 1991-1995 (n=1221)	**Survey IV** 77 years 1997 (n=839)	**End of follow-up** 2002
CVD events [a]	14	134	373	518	701
MetS ATPIII [b]	452	662	878	969	-
Total mortality	-	98	442	748	1078
Ineligible	-	94	219	176	-
Measurements:	Questionnaire [c] Cholesterols [d] Blood pressure Fasting glucose BMI ApoA-1, ApoB-100, Lp(a) [e] IVGTT [f]	Questionnaire [c] Cholesterols [d] Blood pressure Fasting glucose BMI, waist grid IVGTT [f]	Questionnaire [c] Cholesterols [d] Blood pressure Fasting glucose BMI, waist grid ApoA-1, ApoB-100, Lp(a) [e] Insulin clamp Blood samples for DNA	Questionnaire [c] Cholesterols [d] Blood pressure Fasting glucose Waist grid	

[a] Coronary event or ischemic stroke during follow-up. Nine persons with baseline CVD (50 years of age) had another event during follow-up [b] ATP III criteria (male-specific) used. Individuals with missing values excluded at each timepoint. [c] Questionnaire gathered information on CVD risk factors such as family history, medication, and smoking [d] Cholesterols: Total cholesterol, HDL, TG. LDL calculated from Friedewald's formula [e] ApoA-1: Apolipoprotein A-1, ApoB-100: Apolipoprotein B-100, Lp(a): Lipoprotein (a) [f] IVGTT: Intravenous glucose tolerance test.

The cohort was followed for cardiovascular events using the National Causes of Death Register and the National Hospital Discharge Register. The follow-up lasted 32 years, from 1970 until 2002. ICD-9 codes 410-414 and ICD-10 codes I20-I21, and I25 were used for coronary events, and ICD-9 codes 433-434 and ICD-10 code I63 for ischemic strokes.

2. METHODS

2.1. DNA extraction

In the FINRISK study, whole blood samples for DNA extraction were collected at baseline, stored at -20℃, and extracted with a fenolization method modified from Vandenplas et al. (1984). Altogether 100 FINRISK samples had low genomic DNA yield. These samples were whole-genome amplified with multiply-primed whole circle amplification (Silander et al. 2004) from 10 ng of DNA prior to genotyping.

In the ULSAM study, whole blood samples for DNA extraction were collected at the thrid survey from 1150 participants and extracted with a standard salting-out procedure. In addition, DNA from paraffin block samples collected from about 500 individuals who had died during follow-up was extracted from thin sections of paraffin-embedded tissues using proteinase K digestion and QIAamp DNA minikits (www.qiagen.com). The majority of the paraffin samples derived from non-tumor tissue of cancer patients.

2.2. SNP selection

To capture the common allelic variance of gene loci, all genes were covered with dense SNP maps. SNP selection was based on the SeattleSNPs variation discovery resource (SeattleSNPs 2005, http://pga.mbt.washington.edu/) using data for European descent. One SNP was selected to represent each haplotype block (bin). All blocks were initially covered with tag-SNPs, but when the genotyping of the tag-SNP was challenging, another SNP from the bin was selected. All blocks with a frequency of >10% were covered. With the thrombomodulin gene, additional SNPs in the single exon and its near vicinity were also genotyped. The factor V Leiden mutation is a rare variant not present in the SeattleSNPs database. The mutation has, however, previously shown association with CHD (Ng et al. 2006) and stroke events (Lalouschek et al. 2005), and was therefore included based on the literature.

Initially, 57 SNPs were selected to cover the five gene loci: 24 SNPs in *F5*, 7 in *ICAM1*, 7 in *PROC*, 15 in *THBD*, and 4 in *USF1*. LD patterns of the selected SNPs were analyzed with Haploview3.2 (Barrett et al. 2005). If two SNPs were in tight LD (r^2>0.8) with each other, only one was included in the data analyses. In addition, all SNPs monomorphic in 370 Finns were excluded.

2.3. SNP genotyping

SNPs of *F5*, *ICAM1*, *PROC*, and *THBD* – the SNPs analyzed in the FINRISK cohorts – were genotyped with Sequenom MassARRAY (www.sequenom.com) with 10 ng of DNA, using hME chemistry. Due to technical difficulties, SNP *rs3216183* of *THBD* was genotyped with TaqMan (Roche Molecular Systems, www.roche.com).

Several procedures were performed to assure genotyping quality before analyzing the data. All SNPs were first genotyped in 60 Finnish trio families of mother, father, and child, and the Mendelian inheritance of each SNP was checked. The FINRISK cohorts also contained 2% open duplicates and 5% blinded duplicates, which were unknown for the laboratory during genotyping. The error rate for genotyping was <1/400 in the FINRISK samples, and the success rate 93% or more for each SNP.

USF1 gene was genotyped in the ULSAM sample. The 1150 DNA samples extracted from whole blood were genotyped using single-base primer extension assay with fluorescence polarization detection (Hsu et al. 2001) and PerkinElmer Life Sciences reagents (las.perkinelmer.com). The genotyping call rate was 95-99%, with 100% reproducibility in 300 genotype comparisons (9%) between independent experiments for these samples. Of the 500 paraffin samples, only 158 gave reliable genotyping results. The rest of the paraffin samples were excluded from data analyses.

2.4. Haplotype analyses

THBD haplotypes in the Study I were constructed with PHASE2.1.1 program (Stephens and Donelly 2003). Haplotype analyses were excluded from Study II because the SNP selection for *F5*, *ICAM1*, and *PROC* was based on haplotype block information, and LD between the SNPs within each gene was relatively low. Haplotypes for *USF1* in the ULSAM sample (Study III) were constructed with both PHASE2.1.1 (Stephens and Donnelly 2003) and Haploview3.2 (Barrett et al. 2005). The *USF1* SNPs were in substantial LD with each other, and the minor alleles of the

selected SNPs tagged perfectly the three common haplotypes. Thus, separate haplotype analyses were excluded.

2.5. Sequencing

Despite the selection criterion of allele frequency of >10%, some of the genotyped SNPs were very rare in the Finnish population. If the minor allele frequency was <5%, the genotypes for all the minor allele homozygotes were assured by sequencing. To achieve reliable sequencing results, three samples with the other two genotyped classes were also sequenced. Sequencing was performed with ABI 3730xl DNA analyzer (www.appliedbiosystems.com) and BigDye3.1 sequencing chemistry, using 5 ng DNA and utilizing both strands. In addition to the rare SNPs, all *THBD* variants were sequenced in 30 chromosomes. No discrepancies with the sequenom data were detected.

2.6. Statistical analyses

Allele frequency comparisons were performed with χ^2 test. In Studies I and II containing FINRISK data, the allele frequencies were compared between the cases (CVD or deceased) and the modified subcohort after excluding CVD cases from the subcohort. In the ULSAM sample, CVD and MetS cases were compared with the individuals free of these conditions.

Associations of the SNPs with phenotypic variables, such as cholesterol and glucose levels, BMI, waist grid, CRP, and blood pressure, were analyzed with general linear model (GLM). In Studies I and II, the model was adjusted for age, sex, and cohort. Necessary transformations (logtransformation to achieve normality or rank procedure) were performed prior to using GLM.

2.6.1. Time-to-event analyses

Hazard ratios (HRs) for all separate SNPs were calculated with Cox's proportional hazards model (Studies I-III) (Barlow 1994). HR calculations were contained for haplotypes in Study I, and in Study II for SNP pairs. Studies I and II used four different endpoints: coronary events, ischemic stroke events, the combination of these two (i.e. all CVD events), and total mortality. The model was stratified for east-west origin in the studies I and II. Study III used all CVD as an endpoint. In all

studies, SNPs were analyzed in univariate and multivariate models, adjusted for traditional CVD risk factors.

2.6.2. Classification trees

Classification tree analysis with combined FINRISK-92 and FINRISK-97 cohorts was utilized in Study II. The aim of classification tree analysis was to detect specific subgroups, represented by tree branches, holding a substantial CVD risk as well as to search for epistatic effects between SNPs from the four different genes. Analysis was performed with 20 subsamples, each containing 60% of the whole FINRISK data. Separate samples were drawn for men (n=10) and women (n=10).

Classification tree algorithm AnswerTree3.0 of SPSS (www.spss.com) was used in the analyses. AnswerTree searches for the variable best explaining the desired endpoint, here all CVD (combination of incident coronary and stroke events). The best variable is selected as a root node for the tree structure, followed with the next best splits futher dividing the data into subgroups. Significant splits (n=95 altogether) from the 20 subsamples were determined with χ^2 test.

Next, all significant splits were used as root nodes one at a time. This way "a forest" of a total of 95 trees was achieved. The purpose of the forest was to reduce the importance of the first split since the significance levels of the splits were relatively close to each other, and which of the CVD risk factors is the most important remains unknown. All SNPs present in >10% of the sex-specific trees (n=11) were selected for further analyses with Cox's proportional hazards model.

A separate sensitivity analysis was finally performed to address the SNP selection. All 36 SNPs were analyzed in Cox's proportional hazards model to observe any significant findings missed by the 11 SNPs selected with classification trees.

2.6.3. Epistatic effects

The interaction between gene variants was analyzed in Study II. Classification trees were first used to search for epistatic effects in stage 1 of the study. In stage 2, the interaction was analyzed with Cox's proportional hazards model in all SNP pairs of the 11 SNPs selected for stage 2 analyses. The interaction was determined as a deviation from the multiplicative model (Hodge 1981, Risch 1990). A dominant inheritance model was used in SNP pairs to gain power, unless a SNP showed significant results with a recessive inheritance model when analyzed as a single variant.

2.6.4. Longitudinal analyses

Longitudinal analysis was utilized in Study III. In the ULSAM data, each of the participants had 1-4 longitudinal measurements for the same variables from different timepoints during the follow-up. The longitudinal component of the data was addressed with longitudinal GLM analysis (PROC MIXED procedure of SAS v.8 for Windows, www.sas.com). Longitudinal models analyze change in a given variable in different subgroups (here, genotype classes) as a function of time. The rationale was to detect significant variance between the carriers of different allelic variants. For example, a change in cholesterol values measured during the four surveys may differ between the genotype classes, i.e. carriers of certain allelic variants may be at increased risk for dyslipidemias, and longitudinal measurements may detect these trends better than single baseline measurements.

2.6.5. Multiple testing

Multiple testing was conducted in all three studies. Study I yielded negative results. Multiple testing was addressed with the use of two independent study cohorts, FINRISK-97 as a replication cohort for FINRISK-92. Due to inconsistent findings in the two cohorts, multiple testing was not further addressed in the first study. Study II included 36 SNPs in the first stage and 11 SNPs in the second. The selection of the 11 SNPs based on classification trees substantially reduced the number of tests performed in the second stage. Multiple testing in stage 2 was further corrected with FDR (Benjamini and Yekutieli 2005), using a 10% cut-off limit, predicting that 10% of the results may prove false positive. In addition, all SNPs and SNP pairs were analyzed in two independent study cohorts, requiring consistent results from both cohorts. In Study III, multiple testing was corrected by calculating an appropriate significance level with 1000 permutations.

RESULTS AND DISCUSSION

1. ASSOCIATION OF ALLELIC VARIANTS OF THROMBOMODULIN WITH CVD AT POPULATION LEVEL (Study I)

Soluble thrombomodulin levels have repeatedly been linked with the risk of CHD (Salomaa et al. 1999, Wu et al. 2003). Previous works on the *THBD* gene have concentrated on single SNPs, of which *rs1042579* and *rs3216183* have shown association with MI and ischemic stroke (Ireland et al. 1997, Doggen et al. 1998, Wu et al. 2001, Cole et al. 2004, Konstantoulas et al. 2004). Ohlin et al. (2004), in contrast, concluded that *THBD* may not contribute to acute coronary syndrome. A gene-environment interaction has also been suggested with the *THBD* gene variants and smoking (Li et al. 2002).

Studies capturing the complete allelic diversity of any relevant candidate gene had not been published when initiating this study. We aimed to fully cover all known major variants of the thrombomodulin gene and to address their role in CVD at a population level using two independent prospective population cohorts, FINRISK-92 and FINRISK-97.

We genotyped 15 SNPs at the thrombomodulin gene locus: 3 SNPs in the promoter region, 5 in the single exon of the gene, 4 in the 3' untranslated region, and 3 in the intergenic region downstream from the *THBD* gene (Figure 7, Table 10). Of the 15 SNPs, 6 were excluded as monomorphic in the Finnish population, including 4 of the exonal SNPs. The fifth coding region SNP (*rs1042579*) was in perfect LD with *rs3176123*. The rest, 8 SNPs, were selected for data analyses.

***Figure** 7. Thrombomodulin gene and the SNP studied (modified from Auro et al. 2006).*

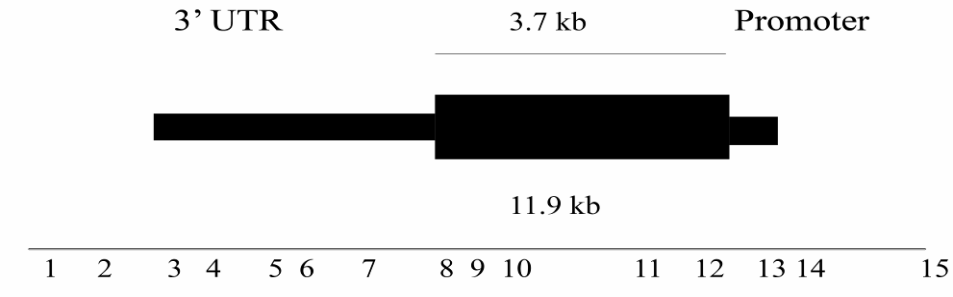


Table 10. *Minor allele frequencies of the SNPs genotyped at THBD locus (modified from Auro et al. 2006).*

SNP	FINRISK-92 Men CVD [a] (n=170)	Subcohort [b] (n=220)	Women CVD [a] (n=86)	Subcohort [b] (n=104)	FINRISK-97 Men CVD [a] (n=204)	Subcohort [b] (n=231)	Women CVD [a] (n=68)	Subcohort [b] (n=71)
1=*Rs6113909*	0.37	0.38	0.41	0.37	0.44	0.45	0.39	0.39
2=*Rs6082986*	0.31	0.28	0.33	0.28	0.36	0.36	0.30	0.31
3=*Rs1962*	0.15	0.17	0.14	0.20	0.16	0.17	0.17	0.14
4=*Rs3176123*	0.32	0.33	0.33	0.32	0.29	0.25	0.30	0.31
5=*Rs1042580*	0.30	0.30	0.32	0.29	0.37	0.36	0.33	0.31
6=*Rs3176122*	0	0	0	0	0	0	0	0
7=*Rs3176121*	0	0	0	0	0	0	0	0
8=*Rs1800579*	0	0	0	0	0	0	0	0
9=*Rs1800578*	0	0	0	0	0	0	0	0
10=*Rs1042579* [c]	0.32	0.33	0.33	0.32	0.29	0.25	0.30	0.31
11=*Rs1800577*	0	0	0	0	0	0	0	0
12=*Rs1800576*	0	0	0	0	0	0	0	0
13=*Rs3176119*	0.06	0.03	0.02	0.03	0.03	0.03	0.04	0.04
14=*Rs3216183*	0.19	0.21	0.18	0.21	0.36	0.13	0.18	0.17
15=*Rs6048519*	0.43	0.43	0.44	0.46	0.43	0.48	0.45	0.37

[a] CVD: Indicent coronary event or incident ischemic stroke [b] CVD cases removed from the subcohort
[c] Excluded due to perfect LD with *rs3176123*

During follow-up, 401 incident CHD events and incident 148 ischemic strokes were recorded, and 610 participants deceased. Of the participants, 21 had both incident coronary and incident ischemic stroke event, and 138 of the 610 deceased also had an incident CVD event. Of the subcohort members, 88 had CVD already at the baseline and 72 had an incident CVD event during follow-up. (Table 11.)

Table 11*. Characteristics of the FINRISK-92 and FINRISK-97 study cohorts: CVD events and total mortality at baseline and during follow-up (modified from Auro et al. 2007).*

Event	FINRISK-92 Men All (n=2833)	Subcohort (n=286)	Women All (n=3166)	Subcohort (n=114)	FINRISK-97 Men All (n=4092)	Subcohort (n=304)	Women All (n=4049)	Subcohort (n=82)	Total
Incident CHD event	143	30	53	4	156	14	49	3	401
Incident stroke	35	8	33	3	59	10	21	2	148
All-cause mortality	188	34	94	9	242	35	86	8	610
Baseline CVD	159	29	63	3	322	50	118	6	662

Smoking, diabetes, and hypertension were more frequent among the CVD cases than in the subcohort. CVD cases also had higher BMI and CRP values. No major differences were seen in the cholesterol values. Table 12 summarizes main phenotypic characteristics of the genotypic study samples.

Table 12. *Phenotypic characteristics (mean [standard deviation] or %) of the FINRISK-92 and FINRISK-97 genotypic samples (modified from Auro et al. 2007).*

	FINRISK-92				FINRISK-97			
	Men		Women		Men		Women	
	CVD (n=170) [a]	Subcohort (n=220) [b]	CVD (n=86) [a]	Subcohort (n=104) [b]	CVD (n=204) [a]	Subcohort (n=231) [b]	CVD (n=68) [a]	Subcohort (n=71) [b]
Hypertension (%)	70.6	54.5	78.8	63.5	75.0	65.4	67.6	74.6
Smoking (%)	39.4	24.5	16.5	9.6	27.0	23.4	20.6	11.3
Diabetes (%)	9.4	4.5	15.3	3.8	13.7	5.6	17.6	7.0
Age (years)	54.2 (7.7)	52.8 (9.1)	54.9 (8.0)	55.9 (6.7)	61.9 (8.7)	59.9 (10.2)	62.8 (7.4)	63.1 (7.6)
BMI (kg/m^2) [c]	28.6 (3.9)	27.4 (3.9)	29.3 (5.1)	26.9 (4.7)	28.6 (3.5)	27.5 (3.9)	29.3 (4.5)	27.9 (4.4)
TC (mmol/l) [d, g]	6.3 (1.1)	5.8 (1.0)	6.1 (1.3)	6.0 (1.0)	5.8 (1.0)	5.7 (1.0)	6.1 (0.9)	6.0 (1.1)
HDL (mmol/l) [e, g]	1.2 (0.3)	1.3 (0.4)	1.4 (0.4)	1.6 (0.4	1.2 (0.3)	1.3 (0.3)	1.4 (0.4)	1.5 (0.4)
CRP (mg/l) [f]	4.3 (5.5)	2.6 (3.7)	3.6 (4.8)	1.0 (3.9)	2.9 (3.2)	1.9 (2.4)	3.3 (3.9)	2.7 (4.2)

[a] CVD: Incident coronary event or ischemic stroke [b] CVD cases removed from the subcohorts [c] BMI: Body-mass index [d] TC: Total cholesterol [e] HDL: high density lipoprotein cholesterol [f] CRP: C-reactive protein
[g] Individuals with lipid medication removed

Haplotype analyses with PHASE2.1.1 yielded 35 haplotypes; 25 in the FINRISK-92 cohort and 30 in FINRISKI-97. The majority of the observed haplotypes were rare, and only 7 had a frequency of >5%. The haplotype AAGGTG(-TT)G was significantly less common among the CVD cases than in the subcohort in the FINRISK-92 sample (14.6% vs. 19.9%, p=0.0054). In the FINRISK-97, the same haplotype was, however, more common among the CVD cases than in the subcohort (15.1% vs. 12.7%). Similar inconsistency was observed with the haplotypes AAGTTG(+TT)T and GGGTCG(+TT)T. AAGTTG(+TT)T seemed a risk haplotype for CVD in the FINRISK-92 cohort (frequency of 13.2% among the CVD cases vs. 10.8% in the subcohort, p=0.039), but not in FINRISK-97 (12.2% vs. 12.3%). The haplotype GGGTCG(+TT)T was more common among the CVD cases than in the subcohort in FINRISK-92 (18.1% vs. 13.1%, p=0.037), but not in FINRISKI-97, where it, instead, seemed protective for total mortality (18.9% vs. 12.7%, p=0.0041). (Table 13).

Cox's proportional hazards model yielded inconsistent results between the FINRISK-92 and FINRISK-97 cohorts. Results from the two separate cohorts were even opposite to each other. Table 14 illustrates results achieved in time-to-event analysis with the haplotype GGGGTCG(+TT)T. SNPs or haplotypes did not associate with any of the phenotypic risk factors, either. Gene-environment interaction between the *THBD* variants and smoking was not observed.

Table 13. *Major (frequency >5%) haplotypes of the 8 SNPs, and the haplotype frequencies in the subcohort and among the CVD patients and the deceased ones (modified from Auro et al. 2006).*

	FINRISK-92			FINRISK-97		
Haplotype [a]	Subcohort [b]	CVD [c]	Deceased [d]	Subcohort [b]	CVD [c]	Deceased [d]
GGGTCG(+TT)G	16.2	16.9	18.8	19.7	16.1	17.5
GGGTCG(+TT)T	13.1	18.1	16.7	18.9	16.7	12.7
AAGTTG(+TT)G	7.3	9.2	8.9	7.5	7.9	7.6
AAGTTG(+TT)T	10.8	13.2	13.3	12.3	12.2	12.7
AAGGTG(+TT)G	13.1	14.2	14.7	10.9	12.4	13.9
AAGGTG(-TT)G	19.9	14.6	13.8	12.7	15.1	14.8
AAAGTT(+TT)T	12.8	10.6	10.3	11.8	10.7	13.3

[a] For SNPs *rs6113909*, *rs6082986*, *rs1962*, *rs3176123*, *rs1042580*, *rs3176119*, *rs3216183*, and *rs6048519;*+ TT indicates an insertion; -TT, deletion in *rs3216183* [b] full subcohorts used [c] CVD at baseline or during follow-up [d] Deceased during follow-up

Table 14. *Association of the haplotype GGGTCG(+TT)T with CVD events and total mortality in time-to-event analysis. The two study cohorts provided inconsistent results (modified from Auro et al. 2006.)*

		FINRISK-92			FINRISK-97		
Endpoint		HR [a]	95% CI [b]	p	HR [a]	95% CI [b]	p
Incident coronary events	Men	1.73	0.97-3.10	0.064	0.86	0.49-1.51	0.603
	Women	1.61	0.40-6.44	0.502	0.62	0.16-2.43	0.495
	All	1.79	1.10-2.91	0.018	0.76	0.47-1.23	0.259
Incident ischemic stroke events	Men	0.41	0.09-1.86	0.244	1.44	0.67-3.10	0.353
	Women	1.35	0.32-5.58	0.683	0.36	0.05-2.39	0.290
	All	0.63	0.27-1.45	0.277	0.95	0.50-1.80	0.878
All incident CVD events [c]	Men	1.45	0.83-2.54	0.197	1.06	0.64-1.73	0.829
	Women	1.56	0.54-4.52	0.414	0.52	0.15-1.73	0.285
	All	1.45	0.91-2.23	0.121	0.85	0.55-1.30	0.846
Total mortality	Men	1.56	0.96-2.55	0.073	0.57	0.31-0.82	0.006
	Women	1.75	0.80-3.81	0.167	0.84	0.29-2.45	0.836
	All	1.57	1.06-2.33	0.024	0.58	0.39-0.87	0.007

[a] HR=Hazard ratio [b] CI=Confidence interval [c] CVD=Coronary events and ischemic strokes combined

This study comprised haplotype analyses, in addition to single SNP analyses. Haplotypes represent the allele structure in a single DNA strand and determine the genetic variance at a locus more precisely than individual SNPs. Haplotype analysis may guide us to discover also more rare variants embedded in the background of common

haplotypes (Clark 2004), subsequently identified with regional deep sequencing. Joint effects of several individual SNPs contributing to haplotypes may also affect important genetic structures, for example protein-binding sites (Clark 2004).

Haplotype estimation in nonrelatives lacking parental information is, however, somewhat problematic. Haplotype estimation programs such as Haploview proceed well in estimating general haplotype frequencies in a study sample, but to exceed this estimation to individual level is uncertain (Clark 2004). In Study I, this was assessed with 40 separate PHASE runs. The haplotypes were then cross-checked between individual runs. The 7 major haplotypes yielded consistent results. *THBD* is a very small gene with only one exon, and the SNPs analyzed were within 12 kb distance. Study I comprised SNPs selected prior to sequence variation resources, such as HapMap or SeattleSNPs, and LD structure between individual *THBD* SNPs ranged from r^2 0.8 – 0.01. These factors facilitated the haplotype construction at the *THBD* locus.

The relationships between soluble hemostatic factors and their genetic determinants and between genetic variants and CVD are often complex (Lane and Grant 2000, Voetsch and Loscalzo 2004). The major weakness of Study I was the lack of soluble thrombomodulin measurements, which could have provided additional information on the relations of TM, *THBD* gene variants, and CVD.

In conclusion, no consistent associations with CHD or stroke events or total mortality were observed with any of the SNPs or haplotypes analyzed. Although some suggestive associations were seen in the separate study cohorts, these findings could not be confirmed in the other study sample. The results from this study suggested that thrombomodulin gene variants are not major independent risk factors for CVD, at least not in our population sample. We could not replicate previously reported associations of the SNPs *rs1042579* and *rs3216183* with CHD, nor did we find evidence of gene-environment interaction between the *THBD* variants and smoking. Association of the *THBD* gene variants with cardiovascular events has not been firmly established in later studies to date.

2. ANALYZING *F5*, *ICAM1*, *PROC*, AND *THBD* VARIANTS AND THEIR CO-EFFECTS AND THE RISK OF CVD (Study II)

This study aimed to investigate the role of four CVD-related genes – *F5*, *ICAM1*, *PROC*, and *THBD* – in cardiovascular events at a population level. *F5*, *PROC*, and

THBD are all participants of the clotting cascade, and defects in the genes or in gene products predispose to increased clotting.

The role of the *F5* Leiden mutation in venous thrombosis is well-established (Lane et al. 2000, Simmonds et al. 2001), but its role in arterial thrombosis is controversial. Ye et al. (2006) recently suggested in a large meta-analysis that the variant predisposes to coronary events. Besides the Leiden mutation, data on *F5* variants and arterial thrombosis are scarce. Protein C deficiency in patients with arterial thrombosis has been characterized in single case reports (Cakir et al. 2002, Tiong et al. 2003), and protein C is suggested to have a neuroprotective function in ischemic strokes (reviewed by Griffin et al. 2006). The relation of genetic *PROC* variants to cardiovascular diseases is largely unknown. ICAM1 is involved in leukocyte adhesion and migration at the site of atherosclerotic lesions, a crucial step in atherosclerosis development (Ross 1999). High soluble ICAM1 concentrations are linked to the risk of CHD (Haugt et al. 1996, Hwang et al. 1997, Malik et al. 2001, Jenny et al. 2006). Interestingly, Wu et al. (2003) have suggested that the interaction of high sICAM1 and low sTM predict CHD risk. High sICAM1 also relates to ischemic stroke events (Simundic et al. 2004, Ehrensperger et al. 2005, Blum et al. 2006, Wang et al. 2006). Elevated ICAM1 expression has been associated with atherosclerosis risk (Davies et al. 1993). However, Nuotio et al. (2003) reported a negative association of ICAM1 expression in symptomatic carotic stenosis plaques.

Our aim was to analyze these four genes simultaneously in order to address potential epistatic effects between the gene variants. We hypothesized that simultaneous analysis would reveal co-effects missed if examining the genes separately. The study covered all known common (>10%) variants of the target genes. We started with 53 SNPs: 24 in *F5*, 7 in *ICAM1*, 7 in *PROC,* and 15 in *THBD* (discussed in previous chapter). Four *F5* SNPs (*rs3766103*, *rs2227245*, *rs6670678*, and *rs6029*), one *ICAM1* SNP (*rs5030352*), one *PROC* SNP (*rs2069921*), and one *THBD* SNP (*rs1042579*) were excluded based on LD, whereas 11 SNPs (*rs9332566*, *rs9332625*, *rs9332587*, *rs1046712* of *F5*, *rs5030388* of *ICAM1*, and *rs1800576*, *rs1800577*, *rs1800578*, *rs1800579*, *rs3176122*, *rs3176121* of *THBD*) were excluded as monomorphic in 370 Finns. This yielded 36 SNPs for stage 1 analyses (Tables 15-17).

The data analyses were performed in three stages, illustrated in Table 15. In stage 1, we built a forest of 95 classification trees, using phenotypic risk factors (diabetes status, BMI, blood pressure, TC, HDL, TG, WHR, and CRP) and 36 SNPs, and selected 12 SNPs present in >10% of the trees. These included 6 *F5* SNPs, and 2 SNPs of *ICAM1*, *PROC*, and *THBD* each. *F5* SNPs *rs2420369* and *rs970741*

belonged to the same haplotype block in Haploview-analyses; thus the latter was excluded, yielding 11 SNPs to be analyzed at stage 2 (Table 17).

Table 15. *Analysis strategy used in Study II.*

1.	53 SNPs of *F5*, *ICAM1*, *PROC*, and *THBD*
2.	Removal of monomorphic SNPs and SNPs in thight LD: 36 SNPs analyzed in stage 1
3.	Stage 1: 11 SNPs selected for stage 2 based on classification trees
4.	Stage 2: Single SNPs and SNP pairs analyzed in Cox's proportional hazards model
5.	Stage 3: Sensitivity analysis for stage 1: All 36 SNPs analyzed in Cox's proportional hazards model

Table 16. *Minor allele frequencies of SNPs excluded from stage 2 analyses (modified from Auro et al. 2007).*

SNP	GENE	FINRISK-92 Men CVD [a] (n=170)	Subcohort [b] (n=220)	Women CVD [a] (n=86)	Subcohort [b] (n=104)	FINRISK-97 Men CVD [a] (n=204)	Subcohort [b] (n=231)	Women CVD [a] (n=68)	Subcohort [b] (n=71)
Rs970741	*F5*	0.19	0.19	0.13	0.23	0.20	0.20	0.20	0.21
Rs6013	*F5*	0.08	0.08	0.09	0.09	0.07	0.09	0.08	0.09
Rs9332640	*F5*	0.43	0.46	0.45	0.47	0.47	0.43	0.45	0.49
Rs6030	*F5*	0.28	0.27	0.23	0.32	0.28	0.29	0.28	0.31
Rs9332618	*F5*	0.17	0.16	0.18	0.19	0.16	0.16	0.22	0.18
Rs9332695	*F5*	0.05	0.06	0.05	0.02	0.05	0.03	0.04	0.02
Rs9332590	*F5*	0.34	0.34	0.33	0.29	0.34	0.34	0.32	0.26
Rs6035	*F5*	0.03	0.03	0.04	0.07	0.06	0.05	0.04	0.05
Rs9332575	*F5*	0.14	0.10	0.11	0.08	0.10	0.12	0.15	0.11
Rs6019	*F5*	0.02	0.02	0.02	0.04	0.03	0.03	0.04	0.02
Rs3753305	*F5*	0.37	0.38	0.33	0.36	0.36	0.37	0.32	0.32
Rs5030390	*ICAM1*	0.02	0.04	0.05	0.05	0.04	0.04	0.01	0.04
Rs281432	*ICAM1*	0.45	0.42	0.48	0.48	0.44	0.46	0.42	0.49
Rs3093030	*ICAM1*	0.41	0.41	0.39	0.40	0.42	0.43	0.43	0.36
Rs1799810	*PROC*	0.33	0.35	0.39	0.33	0.35	0.33	0.25	0.35
Rs2069920	*PROC*	0.43	0.42	0.43	0.46	0.42	0.42	0.43	0.39
Rs2069923	*PROC*	0.03	0.03	0.04	0.03	0.02	0.03	0.01	0.04
Rs2069928	*PROC*	0.24	0.21	0.18	0.21	0.24	0.24	0.29	0.25
Rs6113909	*THBD*	0.37	0.38	0.41	0.37	0.44	0.45	0.39	0.39
Rs6082986	*THBD*	0.31	0.28	0.33	0.28	0.36	0.36	0.30	0.31
Rs1962	*THBD*	0.15	0.17	0.14	0.20	0.16	0.17	0.17	0.14
Rs3176123	*THBD*	0.32	0.33	0.33	0.32	0.29	0.25	0.30	0.31
Rs3176119	*THBD*	0.06	0.03	0.02	0.03	0.03	0.03	0.04	0.04
Rs3216183	*THBD*	0.19	0.21	0.18	0.21	0.36	0.13	0.18	0.17

[a] CVD: Indicent coronary event or incident ischemic stroke [b] CVD cases removed from the subcohort

Table 17*. Minor allele frequencies of SNPs selected for stage 2 analyses (modified from Auro et al. 2007).*

		FINRISK-92				FINRISK-97				
		Men		Women		Men		Women		
SNP	GENE	CVD [a] (n=170)	Subcohort [b] (n=220)	CVD [a] (n=86)	Subcohort [b] (n=104)	CVD a (n=204)	Subcohort [b] (n=231)	CVD [a] (n=68)	Subcohort [b] (n=71)	Location
Rs2420369	*F5*	0.37	0.38	0.31	0.42	0.41	0.36	0.36	0.42	Intron
Rs9332591	*F5*	0.12	0.13	0.11	0.13	0.10	0.12	0.13	0.17	Intron
Rs6025	*F5*	0.03	0.03	0.03	0.01	0.03	0.01	0.03	0.03	Exon [c]
Rs7542281	*F5*	0.37	0.34	0.36	0.26	0.32	0.35	0.43	0.33	Intron
Rs2269648	*F5*	0.24	0.28	0.23	0.30	0.27	0.25	0.24	0.29	Promoter
Rs5030347	*ICAM1*	0.23	0.21	0.23	0.16	0.23	0.18	0.24	0.17	Intron
Rs5030341	*ICAM1*	0.33	0.33	0.37	0.38	0.30	0.34	0.30	0.37	Intron
Rs5937	*PROC*	0.27	0.31	0.31	0.34	0.37	0.35	0.38	0.39	Exon [d]
Rs1401296	*PROC*	0.31	0.28	0.22	0.28	0.36	0.36	0.30	0.31	3’
Rs1042580	*THBD*	0.30	0.30	0.32	0.29	0.37	0.36	0.33	0.31	3’
Rs6048519	*THBD*	0.43	0.43	0.44	0.46	0.43	0.48	0.45	0.37	Promoter

[a] CVD: Indicent coronary event or incident ischemic stroke [b] CVD cases removed from the subcohort

The study revealed several novel CVD-associated gene variants and their pairwise combinations (Table 18). The minor allele of *F5 rs7542281* was associated with increased risk of both coronary and ischemic stroke events in females. The risk of incident CHD events was 2.63 (CI 95% 1.38-5.00, p=0.0033), and of all CVD events 2.65 (CI 95% 1.50-4.39, p=0.00061). *F5 rs2420369* was associated with all CVD events in females, the major allele being the risk-increasing allele. Of SNP combinations, *F5 rs7542281* and *THBD rs14012580* jointly contributed to the risk of CHD in females (HR 4.45, CI 95% 1.94-10.19, p=0.00042), the minor alleles being the risk increasing alleles.

The minor allele of *ICAM1* variant *rs5030347* was associated with increased mortality risk in men (HR 2.28, CI 95% 1.55-3.36, p=0.000029). In addition, *PROC rs1401296* was associated with total mortality in men jointly with *ICAM1 rs5030341* (HR 1.63, CI 95% 1.18-2.24, p=0.0028) and with *F5 rs2269648* (HR 1.80, CI 95% 1.19-2.86, p=0.0069).

The minor allele of *PROC rs1401296* showed association with increased risk of stroke events in men (HR 2.84, CI 95% 1.32-6.12, p=0.0077). The *F5* Leiden mutation (*rs6025*) was also associated with ischemic stroke in men (HR 4.47, CI 95% 1.63-12.29, p=0.0037), but the variant is very rare (minor allele frequency <5%, Table 17), and could only be analyzed in the combined FINRISK sample.

With the incident stroke events in females, the high HRs and wide confidence intervals of both *rs7542281* and *rs2269648* likely reflect the insufficient number of cases.

Table 18. *SNPs and SNP pairs associated with CVD in Study II with Cox's proportional hazards model adjusted for smoking, BMI, hypertension, total cholesterol/HDL ratio, diabetes, and CRP. Only results for the combined sample of FINRISK-92 and FINRISK-97 and significant after multiple testing correction with FDR (10% limit) shown (modified from Auro et al. 2007).*

Variant	Gene	Inheritance model	Endpoint (sex)	HR [c]	CI 95% [d]	p
Rs7542281	*F5*	Additive [a]	CHD (F)	2.63	1.38-5.00	0.0033
Rs7542281	*F5*	Additive [a]	Stroke (F)	13.51	2.82-64.77	0.0011
Rs7542281	*F5*	Additive [a]	CVD (F)	2.65	1.50-4.39	0.00061
Rs2420369	*F5*	Additive [b]	CVD (F)	1.81	1.21-2.72	0.0040
Rs2269648	*F5*	Additive [b]	Stroke (F)	11.51	2.25-58.83	0.0034
Rs6025	*F5*	Dominant [a]	Stroke (M)	4.47	1.63-12.29	0.0037
Rs1401296	*PROC*	Recessive [a]	Stroke (M)	2.84	1.32-6.12	0.0077
Rs5030347	*ICAM1*	Recessive [a]	Mortality (M)	2.28	1.55-3.36	0.000029
Rs7542281 x Rs1042580	*F5/THBD*	Recessive [a]/Dominant [a]	CHD (F)	4.45	1.94-10.19	0.00042
Rs5030341 x Rs1401296	*ICAM1/PROC*	Dominant [a]/Dominant [a]	Mortality (M)	1.63	1.18-2.24	0.0028
Rs2269648 x Rs1401296	*F5/PROC*	Dominant [b]/Dominant [a]	Mortality (M)	1.80	1.18-2.76	0.0069

[a] Minor allele is the risk allele [b] Major allele is the risk allele [c] HR: Hazard ratio [d] CI: Confidence interval

Reports on the role of inherited thrombophilias in arterial thrombosis are controversial. Linnemann et al. (2007), together with Boekholdt and Kramer (2007), concluded that arterial thrombotic events in patients with venous thrombosis are rare, suggesting that inherited thrombophilias do not majorly contribute to arterial thrombosis. De Moerloose and Boehlen (2007) stated that inherited thrombophilias may play a role in arterial thrombosis in specific subgroups, for example in young individuals, or after revascularization procedures. Several reports, on the other hand, speak for genetic involvement (Lalouschek et al. 2005, de Paula Sabino et al. 2006, Eterovic et al. 2007). These reports have, however, mainly concentrated on the Leiden mutation and *F2 20210GA* variant. Our results with *F5*, *PROC*, and *THBD* variants speak for the involvement of thrombosis genes in CHD and ischemic stroke in the Finnish population.

PROC variant *rs1401296* was associated with total mortality jointly with *ICAM1 rs5030341* or *F5 rs2269648* variant. *ICAM1* variants have previously been linked to type 1 diabetes (Ma et al. 2006), bronchial asthma (Nejentsev et al. 2003, Li et al. 2005), celiac disease (Abel et al. 2006), inflammatory bowel diseases (reviewed by Papa et al. 2004), and some cancers (Kammerer et al. 2004, Chen et al. 2006, Vinceti et al. 2006), and activated protein C is suggested to reduce body organ damage by reducing leukocyte and monocyte activity in inflammatory processes independent of its thrombotic actitivities (Grey et al. 1994, Mizutani et al. 2000). Contribution of *F5*, *ICAM1*, and *PROC* variants to all-cause mortality may also here be distinct from atherosclerotic causes.

Classification trees may help to structure datasets comprising numerous variables. The approach may reveal relations between individual variables or groups of variables. Gruenewald et al. (2006), for example, used classification trees to determine pathways of biomarkers contributing to mortality in the elderly. In genetic studies, classification trees can be used in determination of gene-gene and gene-environment interactions. Besides cancer genetics (Garcia-Closas et al. 2006, Wu et al. 2006, Briollais et al. 2007), classification trees are currently not widely utilized in complex diseases, although Baessler et al. (2007) recently identified epistatic interactions in two genes affecting the susceptibility of MI and CHD using the approach.

In Study II, classification trees were utilized to filter out the most significant variants from among many analyzed. The approach failed to identify any direct epistatic effects between the SNPs studied, but did reveal a distinct feature; the most significant splits in the trees were for the traditional CVD risk factors, such as cholesterols and BMI, whereas the SNPs appeared only in the lower tree branches. This suggests that the role of genetic variants is most evident in specific subgroups determined by the traditional risk factors. The *F5* Leiden mutation, for example, has previously been associated with CVD risk among young females with unhealthy lifestyles (Rosendaal et al. 1997). Classification tree analyses may facilitate the recognition of such specific subgroups of individuals holding a substantial disease risk.

To explore the reliability of the classification tree approach, we performed a sensitivity analysis for stage 1 to observe if any significant SNPs were left out from stage 2 analyses. All 36 SNPs were analyzed in Cox's proportional hazards model in the separate study cohorts. The results indicated that classification trees performed well in detecting the significant SNPs. No consistent associations were seen with

any of the other SNPs besides the 12 selected ones. Another way to address the reliability of this kind of study would be the use of a separate study sample in the classification tree selection step (stage 1), performing the stage 2 analyses using other study samples. This, however, was beyond our resources.

Multiple testing is a major issue in large studies comprising numerous gene variants. In complex diseases, the risk of each predisposing variant to the trait is likely modest (Hunter and Kraft 2007), and a significance threshold required in strict multiple correction methods, such as Bonferroni correction, may prove extremely challenging. Inheritance model of the variants in complex traits is, in addition, often unknown, which also here increased the number of tests, performed using all additive, dominant, and recessive models. We corrected multiple testing with a SNP selection procedure performed with classification trees (stage 1), and with FDR (stage 2). We chose to use a 10% FDR limit, stating that up to 10% of the results presented in Table 18 may prove false positives.

Replicating the results in other independent study cohorts and populations decreases the chance for false-positive findings. Replication, however, may refer to associations seen with the same gene, haplotype, variant, or allele. We chose to use the gold standard of replication (Clarke et al. 2007); to state any finding "true", the same allele was required to be associated with the traits in each of the two independent study cohorts.

Sample size proved a major limitation of Study II. Reliable interaction studies require very large study samples. Here, the total number of incident CVD cases (>500) turned out to be insufficient. Especially ischemic stroke events in the sex-specific analyses were few. To address rare variants, such as the Leiden mutation, with recessive inheritance models would require a substantial numbert of cases (Ye et al. 2006), beyond the capacity of our study. Further studies on larger cohorts and different populations are therefore necessary to confirm our observations.

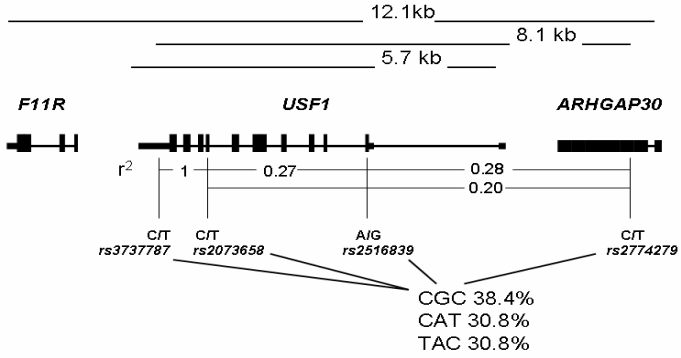
To conclude, the results from Study II suggest that the *F5* gene variants may contribute to CHD and ischemic stroke independently or jointly with *THBD* variants. Variants of *F5*, *ICAM1*, and *PROC* may contribute to all-cause mortality, either as single variants or through gene-gene interactions.

3. ADDRESSING THE ROLE OF *USF1* IN CVD AND METABOLIC TRAITS (Study III)

Extensive studies on *USF1* have demonstrated the gene's impact on traditional cardiovascular risk factors, such as serum lipid levels, in FCHL families. Pajukanta et al. (2004) first reported an association of the gene with FCHL and dyslipidemias. Several other studies have subsequently provided supporting results, confirming the role of *USF1* in dyslipidemias (Coon et al. 2005, Huertas-Vazquez et al. 2005, Komulainen et al. 2006, van der Vleuten et al. 2007). However, the majority of these works have used family data ascertained for specific endpoints such as FCHL, and concentrated on only few SNPs, mainly *rs3737787* and *rs207365*. Thus, the impact of the gene on a population level remains largely unknown, and reports covering the full allelic variance are scarce. Komulainen et al. (2006) reported an association of *USF1* variants with CVD and mortality in women. In men, no association with CVD at the population level has been published. Reports on *USF1* and MetS/T2DM at population level have yielded negative results (Gibson et al. 2005, Ng et al. 2005, Zeggini et al. 2006).

Study III aimed to address the role of *USF1* in a large population cohort of Swedish men, focusing on incident CHD and ischemic stroke events and CVD risk factors, including metabolic syndrome. We analyzed four SNPs, covering the allelic variance of the gene. *Rs3737787* was in perfect LD with *rs2073658* (Figure 8), and the former was therefore excluded. The minor alleles of the SNPs were observed to precisely tag three major haplotypes: *Rs2516839* tagged the haplotype CGC, *rs2073658* the haplotype TAC, and *rs2774279* the haplotype CAT, respectively (Figure 8).

Figure 8. *USF1 gene, SNPs genotyped, and the major haplotypes.*



The cohort was followed up for 32 years for CVD events, and the components of metabolic syndrome (TC, TG, HDL, BMI, waist, fasting glucose, and blood pressure) were measured at four different surveys over 30 years. Individuals free of MetS and CVD in all the four surveys served as healthy controls. Table 19 states the study design.

Table 19. *Statistical analyses in the study III were performed in four parts.*

1.	Allele frequencies of CVD and MetS groups were compared with those of healthy controls free of these conditions (χ^2 test). The allele frequencies of the cohort were also compared with a group of healthy Swedish blood donors (n=98) because the genotypic sample was drawn in the middle of the follow-up when about half of the original participants were deceased.
2.	Hazard ratios were calculated with Cox's proportional hazards model using the combination of incident CHD and ischemic stroke as an endpoint.
3.	Longitudinal GLM analysis was performed to observe the influence of the *USF1* variants with the phenotypic variables. This analysis was done with the whole cohort and then in three subgroups: CVD, MetS, and healthy controls.
4.	Cross-sectional analysis at the third survey utilized the four different MetS definitions.

Table 20 summarizes main phenotypic characteristic of the study cohort. Mean blood pressure and fasting glucose values increased by age, whereas TC and LDL values declined. This decline was also observed among those participating to all the four surveys, suggesting that it was due to environmental factors, rather than selective mortality.

SNP *rs2774279* was associated with metabolic syndrome. The frequency of MetS was significantly lower in the minor allele carriers of this SNP than in the control group, and the trend was present regardless of the MetS definition used, with the exception of the EGIR definition. The association was strongest when using the ATPIII definition at 70-years only ($p<0.001$), followed by the IDF definition at 70 years, and both ATPIII and IDF when all the four surveys were taken into account ($p<0.01$ for all). With the WHO definition at 70 years, a borderline association was seen ($p<0.05$ but failing multiple testing limit). MetS was more common among the minor allele carriers of *rs2073658* than in the control group with the IDF definition in the third survey ($p<0.05$), suggesting that this variant increases the risk for MetS.

Rs2516839 was not associated with MetS. None of the SNPs were associated with CVD when comparing the allele frequencies or with Cox's hazard model. (Table 21.)

Table 20. *Phenotypic characteristics (mean [standard deviation]) of the ULSAM cohort (modified from Auro et al. in press).*

Variable	**Survey I** 50 years 1970-1973 (n=2 322)	**Survey II** 60 years 1981-1984 (n=1 860)	**Survey III** 70 years 1991-1995 (n=1 221)	**Survey IV** 77 years 1997 (n=839)
TC (mmol/l) [a]	6.87 (1.31)	6.31 (1.10)	5.82 (1.00)	5.48 (0.96)
LDL (mmol/l) [a]	5.26 (1.24)	4.28 (0.96)	3.90 (0.89)	3.56 (0.84)
HDL (mmol/l) [a]	1.36 (0.40)	1.19 (0.34)	1.28 (0.35)	1.32 (0.33)
TG (mmol/l) [a]	1.93 (1.22)	1.85 (1.01)	1.45 (0.78)	1.36 (0.68)
Fasting glucose (mmol/l) [b]	4.96 (0.61)	4.81 (0.58)	5.63 (0.58)	5.48 (0.58)
Blood pressure (mmHg) [c]	132/83 (17/11)	140/86 (19/9)	144/82 (17/9)	149/80 (21/10)
Lp(a) (U/l) [a]	245.87 (305.16)	-	305.35 (360.37)	-
ApoA-1 (g/l) [a]	1.43 (0.25)	-	1.28 (0.23)	-
ApoB-100 (g/l) [a]	1.24 (0.28)	-	1.03 (0.23)	-

[a] Individuals with lipid lowering medication removed (n=166) [b] Individuals using glucose lowering medication removed (n=196) [c] Individuals using antihypertensive medication removed (n=466)

Table 21*. Allele frequencies among individuals with metabolic syndrome and among healthy controls in the ULSAM study cohort (modified from Auro et al. in press).*

Polymorphism	*Rs2073658*	*Rs2774279*	*Rs2516839*
MetS ATPIII at 70 years (n=384)	0.33	0.27 [d]	0.39
MetS IDF at 70 years (n=435)	0.35 [b]	0.28 [c]	*0.38*
MetS EGIR at 70 years (n=269)	0.34	*0.30*	*0.36*
MetS WHO at 70 years (n=413)	0.33	0.29 [b]	*0.39*
MetS ATPIII life-long (n=969) [a]	*0.32*	0.29 [c]	*0.39*
MetS IDF life-long (n=905) [a]	*0.32*	0.29 [c]	*0.39*
Healthy controls (n=998)	*0.30*	*0.33*	*0.38*

[a] Life long: Metabolic syndrome diagnosed at one or more surveys

[b] $p < 0.05$ for difference in the minor allele frequencies between the MetS group and the healthy controls

[c] $p < 0.01$

[d] $p < 0.001$

To further dissect the putative protective role of *rs2774279* and the putative risk-increasing role of *rs2073658*, we analyzed the variants with longitudinal GLM in the MetS component traits. Minor allele carriers of *rs2774279* had lower BMI in the whole cohort ($p<0.05$) and lower fasting glucose concentrations in the CVD group, whereas minor allele carriers of *rs2073658* had higher ApoB-100 and TC (illustrated in Figure 9) in the MetS group ($p<0.05$). Borderline findings suggest that the minor allele carriers of *rs2774279* also had higher HDL cholesterol levels (whole cohort,

CVD, and MetS groups), lower fasting glucose values (whole cohort), lower BMI (CVD group), and smaller waist grid (whole cohort), but these results did not reach statistical significance after multiple testing was corrected with permutations. Similarly, the minor allele carriers of *rs2073658* had higher LDL (Mets group) and higher Lp(a) concentrations (CVD and control groups), although the results were statistically insignificant after multiple testing corrections. (Table 22.)

***Table 22**. Association (p-values) of the USF1 SNPs with longitudinally measured cardiovascular risk factors in longitudinal linear analysis (PROC MIXED of SAS v. 8). SNPs analyzed modeling dominant inheritance (modified from Auro et al. in press).*

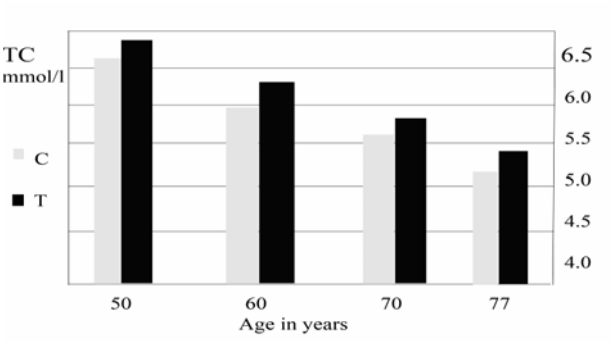
SNP	BMI	Waist	TC [d]	LDL [d]	HDL [d]	ApoA-1 [d]	ApoB-100 [d]	Lp(a) [d]	Fasting glucose	HOMA β-cell [e]
Rs2073658										
All										
MetS [a]			0.0142 [f]	0.0242			0.0136 [f]			
CVD [b]								0.0232		
Healthy [c]								0.0277		
Rs2774279										
All	0.0188 [f]	0.0685			0.0620				0.0397	
MetS [a]					0.0242					
CVD [b]	0.0484				0.0611				0.0116 [f]	0.0532
Healthy [c]										
Rs2516839										
All										
MetS [a]										
CVD [b]	0.0460									
Healthy [c]										

[a] Metabolic syndrome by the ATP III criteria [b] CVD: Individuals with cardiovascular event [c] Individuals with no CVD events or metabolic syndrome

[d] Individuals with lipid-lowering medication removed from the analyses. [e] Fasting insulin / (fasting glucose - 3.5) [f] Significant finding after multiple testing correction as 5th percentile of 1000 permutations

In the cross-sectional GLM analysis of the third survey at 70 years, minor allele carriers of *rs2774279* had higher HDL and lower Lp(a) values compared with major allele homozygotes with all the four MetS definitions. Further, the minor allele carriers of *rs2774279* had lower BMI, higher HDL, lower fasting glucose, and higher mean glucose disposal in the clamp test ($p<0.05$ for the whole cohort).

***Figure 9**. USF1 influences metabolic syndrome component traits throughout adulthood. Minor allele (T) carriers of SNP rs2073658 had significantly higher total cholesterol (TC) values (mmol/l) than major allele homozygotes (C) among individuals with metabolic syndrome in the ULSAM study cohort.*



Which one of the various MetS definitions is the most useful one is not well established. ATPIII is suggested to perform better than IDF in predicting future CHD overall (Tong et al. 2007) and in patients with diabetes (Marchesini et al. 2004), whereas the WHO definition may be sensitive in predicting future diabetes in patients with MetS (Laaksonen et al. 2002). We determined MetS using four different definitions; ATPIII (NCEP 2001) and IDF (Alberti et al. 2006) in all the four surveys, and in addition EGIR (Balkau et al. 2002) and WHO (WHO 1999) definitions at 70 years. In this study, the overall prevalence of MetS was 41.7% with the ATPIII definition, and 39.0% with IDF (Table 23). Of the participants having MetS at baseline (determined by ATPIII), 43.1% had a CVD event in the next three decades (Table 23). MetS at baseline was thus a major risk factor for future CVD. At 70 years, the ATPIII definition had the best overlap with CVD (33.9%, Table 23).

***Table 23.** Metabolic syndrome and cardiovascular events in the study cohort (modified from Auro et al. in press).*

Definition	MetS (n)	CVD and MetS (n)	CVD and MetS (%)
ATPIII (50 years) [a]	452	195	43.1
IDF (50 years) [b]	394	159	40.4
ATPIII (70 years) [a]	384	130	33.9
IDF (70 years) [d]	435	128	29.4
EGIR (70 years) [c]	296	80	29.7
WHO (70 years) [d]	413	123	29.8
ATPIII (life-long) [a]	969	358	36.9
IDF (life-long) [b]	905	30	33.1

[a] ATPIII: National Cholesterol Education Program's Adult treatment panel III [b] IDF: International Diabetes Federation [c] EGIR: European Group for the Study of Metabolic Syndrome [d] WHO: World Health Organization

With longitudinal data, this study provided valuable information of the influence of the *USF1* variants on CVD risk factors in adulthood. The study sample has, however, some downsides. The cohort comprised only men. The drop-out during the 30 years of follow-up was large, cutting the number of participants from 2322 to 839. In addition, measurements carried out varied between the individual surveys (Table 9), although performed at same hospital. The major weakness was the collection of the DNA samples in the third survey instead of the baseline. To assess if this late collection caused bias to the genotype frequencies, additional DNA was extracted from paraffin blocks from the deceased ones, although these samples are also likely biased as collected mainly from cancer patients. No significant differences in the allele frequencies were seen between the whole-blood samples, paraffin samples, and healthy Swedish blood donors, suggesting that late DNA sampling did not cause major bias. The relation of the *USF1* gene with mortality could not be assessed in this study.

As a transcription factor, *USF1* regulates the function of numerous genes involved in body energy and lipid metabolism (Naukkarinen et al. 2005). Although associations of the *USF1* variants with FCHL and dyslipidemias have been replicated in several works, direction of association is less well established. For example with *rs2073658,* the major allele was the risk allele in the most of the family studies (Pajukanta et al. 2004, Huertas-Vazquez et al. 2005, van der Vleuten et al. 2007), whereas two population studies have reported the minor allele to be the risk allele (Komulainen et al. 2006, Zeggini et al. 2006). Our results in this matter support the previous population studies. The inconsistency in the results suggests that underlying gene-gene or gene-environment interactions or both modulate the effects of *USF1* in CVD and dyslipidemias. A thorough understanding of the specific role of the *USF1* variants in the human pathophysiology may be achieved with future functional studies.

To conclude, the results from this study on Swedish men suggested that the *USF1* variants are associated with metabolic syndrome at a population level. The minor allele of SNP *rs2774279* may prove protective for MetS, whereas the minor allele of *rs2073658* may increase MetS risk. Analyses with individual MetS component traits support these findings. No association was observed between *USF1* variants and CHD or ischemic stroke events.

4. GENERAL DISCUSSION

Cardiovascular diseases are and continue to be a major cause of death and disability worldwide (Murray and Lopez 1997). Despite advances in disease prevention, diagnostics, and treatment, coronary heart disease causes annually about 77 000 hospitalizations in Finland (www.stakes.fi), accompanied by approximately 14 000 yearly stroke events (www.stakes.fi). Better understanding of the epidemiology of cardiovascular diseases is therefore a challenge. Genetic studies on cardiovascular traits may reveal novel insights to recognition of individuals in a substantial disease risk. Furthermore, CVD disease mechanisms and research methods used in CVD genetics may be extended to other common traits.

Genetic basis of complex diseases is proposed to consist of several common loci (Collins et al. 1997). However, each of the predisposing variants likely has only a modest effect on the trait (Collins et al. 1997, Hunter and Kraft 2007). An alternative view speaks for rare predisposing loci with strong phenotypic effects, often seen in Mendelian traits (Wang et al. 2005). The study of cardiovascular traits, as well as that of other common traits, has turned out to be a complex story. Although several important candidate genes for CVD have been suggested, the wider genetic structures of the traits remain unknown. Focusing on few, previously reported gene variants, such as *APOE* epsilons (Eichner 2002) or the *PAI1* -675 4G/5G mutation (Ariyaratnam et al. 2007), has been characteristic for the majority of genetic studies on CVD in the past decades. This holds also for the target genes of this work, with the *F5* Leiden mutation, Lys469Glu of *ICAM1*, Ala455Val and -1208–1209TTins/del of *THBD*, and *rs3737787* and *rs2073658* of *USF1*.

We are, however, beginning to understand how polygenic diseases should be addressed. Targeting on single polymorphisms fails to explore the full-scale influence of any candidate gene or locus for a given trait, and studies comprising several variants from different genes are needed to structure the complexity of common diseases. Some of the key factors facilitating the study of complex traits are listed in Table 24.

***Table 24**. Factors to be considered in the study of common diseases to assure reliable results.*

FACTORS FACILITATING GENE IDENTIFICATION IN COMPLEX TRAITS	
Adequate sampling	- Large sample size: sufficient statistical power in order to identify small genetic effects - Reliable diagnostics of the disease outcomes - Use of population isolates: genetic and environmental similarity
Replication of findings	- Use of several separate study samples - Functional studies
Simultaneous analysis of several genes and entire pathways	- Interaction studies - Genome-wide studies - Expression studies
Use of multiple study methods	
Addressing multiple testing	

In cardiovascular genetics, association of only few candidate genes with CVD has, in fact, been firmly established. Of the thrombosis-related variants, the *F5* Leiden mutation and *F2 20210G/A* variant have provided consistent results in several studies (Ye et al. 2006). Of lipid genes, *APOE* (Eichner 2002) and *ABCA1* (Clee et al. 2001, Frikke-Schmidt et al. 2005) seem the most intriguing candidates. Limited sample size likely contributes to the inconsistency in the CVD-related findings. Large studies comprising several thousand participants or large meta-analyses have failed to replicate the majority of previous findings, reported in smaller studies (Shiffman et al. 2006, Ye et al. 2006, Morgan et al. 2007). We could not replicate associations previously reported with the *THBD* variants, nor did we find evidence of direct involvement of the *USF1* gene on CHD and ischemic stroke. Inadequate sample size may also here partly explain the lack of replication. Especially the number of ischemic strokes was limited.

Studies on CVD also suffer from phenotypic variability. “CVD” may, in fact, refer to a variety of different endpoints, such as general atherosclerosis, coronary heart disease, MI, family history of CHD, ischemic stroke, or all stroke events, or to any combination of the former. With thrombomodulin, for example, Wu et al. (2001) used general CHD status as an endpoint, whereas Konstatoulas et al. (2004) used coronary events, Cole et al. (2004) ischemic stroke events, Ohlin et al. (2004) acute coronary syndrome, and Chao et al. (2004) premature MI, to mention few.

Heterogeneity of the cardiovascular phenotypes might also here partly explain the negative results. All the three studies utilized, in addition to all-cause mortality, incident coronary and ischemic stroke events and the combination of these two as

endpoints. Although CHD and ischemic stroke have several common risk factors, the precise risk profiles of the two groups differ. This is likely true also for the genetic risk profiles, although some genes, for example *APOE* (Eichner 2002, Humpries and Morgam 2004, Ariyaratnam et al. 2007), have been firmly linked with both CHD and ischemic stroke. In addition, selection of the CVD cases in Studies I-III was based on the National Hospital Discharge Register and the National Causes of Death Register. Precise data on the degree of atherosclerosis, based for example on autopsy data or coronary angiography, could provide more accurate phenotypes and facilitate the future gene hunt in cardiovascular traits.

Aging contributes to cardiovascular diseases in several ways. Atherosclerosis proceeds by age (Lusis 2000), and CVD risk factor burden increases by age (Aromaa 1981, Rahkonen et al. 1998, Aromaa 2002, Goldstein et al. 2006, Vasto et al. 2007). The genetic risk factors may, however, play a more important role in young (<50 years) patients. Age distribution of the FINRISK participants ranged from 25 to 75 years at baseline, and may significantly contribute to results presented here. Genetic susceptibility for CVD events may also prove partly age-dependent. With *USF1*, a possible age-genotype interaction was recently reported (Reiner et al. 2007). In Study III, influence of the *USF1* variants on CVD risk factors was not age-dependent. Age-genotype interaction could not be reliably assessed in this study due to late DNA sampling. Studies I and II did not address age-genotype interactions, either.

In addition to inadequate sample sizes and phenotypic variability, population stratification may increase the risk for negative or false positive findings. Several population bottlenecks in the history of Finland have created regional subisolates (Pastinen et al. 2001, Service et al. 2006), and stratification may be present between such subisolates. For example, differences in the prevalence of coronary heart disease exist between East and West Finland (Ylä-Herttuala et al. 1987). The FINRISK samples, however, are collected from four to five different geographic areas. We also stratified the time-to-event analyses for the geographic area. The ULSAM sample comprised men living in the Uppsala County, Sweden, born in 1920-1924. At the time, the Swedes were likely a very homogeneous population. Population stratification is not likely a major issue in these Nordic study cohorts.

Sex-specific differences appear in the prevalence of CHD and ischemic stroke, men being at greater risk (Brown et al. 1996, Sacco et al. 1998, Jousilahti et al. 1999, Schreiner et al. 2001). The incidence of CVD events increases by age in both genders, but based on autopsy data, women are approximately 15 years “behind” men in coronary calcification (Hoff et al. 2001), and young women have less

atherosclerosis than men at same the age group (McGill et al. 2000). In a Finnish study, men were at 3-fold risk for CHD compared with women (Jousilahti et al. 1999). Differences in risk factors, markedly in HDL cholesterol and smoking, explained up to 50% of the sex-specific risk difference (Jousilahti et al. 1999). Although partly explained by lifestyle, sex-specific genetic factors are likely to contribute. A recent report stated that the heritability of stroke is greater in women than in men (Touze and Rothwell 2007). Pan and coworkers (2007) reported that the heritability estimates for traits contributing to CVD, such as blood pressure, LDL cholesterol, Lp(a), and TG, have sex-specific differences, which may be X-chromosome-related, whereas Weiss et al. (2006) stated that significant sex-specific differences exist in the genetic distribution of blood pressure, cholesterol levels, BMI, and insulin.

In Study II, several variants were associated with CHD and ischemic stroke in women, whereas in men, the findings concentrated on total mortality. The involvement of *F5* SNPs other than the Leiden mutation in CVD, or that of *F5*, *ICAM1*, or *PROC* variants in mortality, has not been previously reported. Here, the Leiden mutation predisposed to ischemic stroke events in men. Lalouschek et al. (2005) and Eterovic et al. (2007) have observed this variant to be more common in female subjects suffering from ischemic stroke events than in control subjects. The number of female stroke cases in our study cohort was limited, and therefore the Leiden mutation as a risk factor for stroke in women could not be analyzed reliably. With *USF1*, Pajukanta et al. (2004) showed that the gene was associated with FCHL especially in males with high TG levels. Subsequently, Komulainen et al. (2006) reported that the *USF1* variants were associated with CVD and mortality in women but not in men. A *USF1* variant was recently observed to influence TG levels and BMI in sex-dependent manner, different alleles of the same variant being the risk-increasing alleles in women and men (Lee et al. 2007). In Study III, the *USF1* variants were associated with metabolic syndrome and lipid profiles in men (the study III). The ULSAM study cohort did not include women.

Unraveling the genetic structures underlying complex traits will likely require use of several different methods. Genome scan -based family studies may reveal new candidate genes and loci, and investigating monogenic, rare disease forms may also provide valuable information on the pathophysiology underlying common disease forms. Family studies, however, provide information only on the specific families, ascertained for a specific trait, and population-based samples are needed to assess the role of the variants at a population level. Linkage studies are also relatively

insensitive, and other methods are needed to address the putative small-sized effects of the numerous contributing loci likely present in polygenic diseases. In addition to large-scale association studies, functional data are required to fully evaluate the role of associated variants in practice.

SNPs occur in average once in every 1000 basepairs (Sachidanandam et al. 2001). Non-synonymous variants directly influence the amino acid sequence. The majority of all SNPs, however, do not have such direct effects. Synonymous and noncoding variants may nevertheless have important effects on gene functions (Knight 2003, Knight 2005). Polymorphisms in promoter regions may influence gene expression by altering the binding of transcription factors. Regulatory polymorphisms can be divided into *cis*-acting, located in the near vicinity of the target gene, and *trans*-acting, which act as regulatory elements for other genes located elsewhere (Knight 2005). The majority of the regulatory variants are likely *trans*-acting (Brem et al. 2002, Schadt et al. 2003). Current research of complex traits is, however, concentrated on *cis*-regulatory variants (Knight 2005); recognition of regulatory regions located in far distance, possible within other genes, is challenging.

Of the associating SNPs in Study II, the Leiden mutation is the only non-synonymous variant. *THBD* SNP *rs1042580*, jointly contributing to CHD events in women with *rs7542281,* and *rs2269648* of *F5*, associated with all CVD in women, are located in the promoter regions. These promoter variants may act as *cis*-regulatory elements, affecting the near-by genes, or as *trans*-regulatory elements, affecting some other genes. *PROC rs1401296*, associated with stroke as a single variant and contributing jointly to all-cause mortality with *ICAM1 rs5030341* of *F5 rs2269648,* locates to 3' untranslated region of the gene. The other SNPs associated with CVD or mortality in Study II are intronic. Of Study III SNPs, *rs2774279* is located within *ARHGAP30*, a novel gene neighboring *USF1*, and *rs2073658* is located in an intron.

Effects of synonymous variants, appearing in the coding sequence, and promoter region variants, putatively altering gene expression, cannot be understood by inspecting the sequence. Expression studies providing information of allele-specific difference in the expression in a target tissue are needed to address the significance of these variants (Wray et al. 2003). Studies on allele-specific expression patterns provide evidence of regulatory polymorphisms (Knight 2005). With *USF1*, allele-specific difference in the expression of *USF1*-regulated genes was seen in fat biopsy (Naukkarinen et al. 2005). In addition to altered gene expression, noncoding variants

may have effects on splicing, and detected associations may prove indirect, linked to underlying, unknown variants. Further functional studies are thus needed to gain a thorough understanding on the impact of our findings in the pathophysiology of cardiovascular traits.

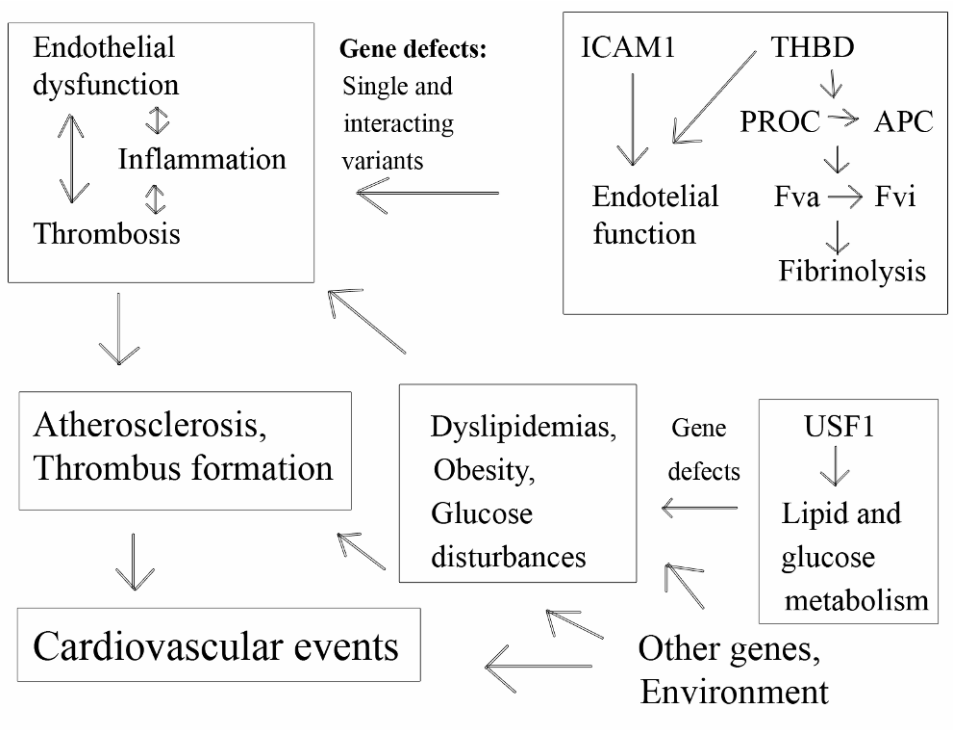
Our results suggest that gene-gene interactions are likely to modify the effects of individual genes and single variants. With *USF1*, evidence of modulatory gene-gene interactions accumulates. Putt et al. (2004) observed that the influence of the *USF1* variants on LDL was BMI-dependent, and Komulainen et al. (2006) reported that the gene variants influenced CVD risk factors among those with CHD or ischemic stroke events, but not in the healthy subcohort. In Study III, the *USF1* variants were associated with TC and ApoB-100 in individuals diagnosed with MetS, with fasting glucose among CVD patients, and with BMI in the whole study cohort. In addition, different alleles of the same *USF1* variant may predict the risk in FCHL families (Pajukanta et al. 2004, Huertas-Vazquez et al. 2005, van der Vleuten et al. 2007) than in the general population (Komulainen et al. 2006, Zeggini et al. 2006). Together these observations suggest that other genes modify the effects of *USF1*.

Enviromental factors may modulate the effects of genetic variants. Here, smoking was the only environmental component measured. Smoking promotes thrombosis (Miller et al. 1998, Fernandez et al. 2002, Pomp et al. 2007), and may enhance the effect of inherited thrombophilias (de Moerloose and Boehlen 2007). Substantial risk for CHD events has been observed among young female smokers carrying the Leiden mutation (Rosendaal et al. 1997). Gene-environment interaction with *THBD* variants and smoking has been proposed previously (Li et al. 2002), but was not observed here. Interaction with smoking was not assessed in Studies II or III.

The genes analyzed in this work are linked to central steps of atherosclerosis development. ICAM1 is involved in leukocyte adhesion and migration into atherosclerotic lesions, a crucial step in atherosclerosis (Ross 1999). Dyslipidemias are central CVD risk factors and key targets of disease prevention and treatment. Lipid accumulation is one of the early signs of atherosclerosis (Stary et al. 1992). Dyslipidemias, among other factors, initiate endothelial dysfunction, which subsequently leads to progression of atherogenesis (Ross 1999, Lusis 2000). As a transcription factor, *USF1* regulates the function of numerous genes involved in body energy and lipid metabolism (Naukkarinen et al. 2005). Association of the *USF1* variants with FCHL and dyslipidemias has been replicated in several works. Thrombosis is a key feature of advanced atherosclerotic lesions giving rise to

clinical complications (Stary et al. 1995). Thrombotic lesions may cause partial or total occlusions in the vasculature, leading to ischemia. As participants of the coagulation cascade, *F5*, *PROC*, and *THBD* genes are tempting candidate genes for arterial thrombosis. Defects in their function result to increased blood clotting. Figure 10 provides a schematic view of how the five target genes may contribute to CVD.

***Figure 10**. Role of the five target genes in CVD.*



The worldwide health burden of cardiovascular diseases is substantial. Better understanding of the pathophysiology of CVD-related traits could help to target disease prevention and clinical treatment to individuals at an especially high disease risk and provide novel pharmaceutical interventions. Although the genetic basis of coronary heart disease and ischemic stroke remains unknown, even single genetic findings may facilitate the recognition of high-risk subgroups. Further genetic and functional studies are, however, needed to determine the specific meaning of our findings in the human pathophysiology.

Results of Studies I-III can be summarized as follows:

1) Thrombomodulin gene variants may not be independent cardiovascular risk factors, but may contribute to coronary heart disease through gene-gene interactions.

2) Variants of *F5*, *ICAM1*, and *PROC* may influence the risk of CHD, ischemic stroke, and total mortality in the Finnish population.

3) Simultaneous study of numerous variants from different genes to reveal epistatic effects may be one step forward in understanding the molecular background of complex traits.

4) *USF1* gene variants may have a role in metabolic syndrome and contribute to various component traits of MetS in the Swedish population.

CONCLUSIONS

In the course of this study, important advances have taken place in the gene hunt for complex traits. Completion of the Human Genome Project handed us the sequence of the human genome. Subsequently, public databases like HapMap and SeattleSNPs have provided detailed information of variation in the human genome, facilitating the genotyping task. Improving techniques have made it possible to efficiently address vast number of genetic variants. Yet, we lack deeper understanding of the genetic basis of most common traits.

This study was designed to address the impact of relevant candidate genes on CVD risk at the population level. With the wave of genome-wide association studies, the information of critical genes for this trait will rapidly increase compared with candidate genes studies; GWAS are not influenced by our previous understanding of the biochemical background of the trait. Especially studies analyzing epistatic effects of more than one gene will be warranted to structure the complexity of common diseases. Genome-wide association studies will at least partly meet this need by providing data simultaneously from all the genes in the genome. Analysis of these very large datasets is somewhat problematic, tackling with issues of replication, multiple testing, and inadequate sampling. In addition, meaningful selection of SNPs genotyped is an issue under debate. Amos (2007) states that key factors for successful genome-wide association studies include use of large homogeneous population samples and careful dissection of the phenotype. The population isolates – such as the Finns – are thus likely to remain applealing beyond the monogenic era and candidate gene studies.

According to Pollex and Hegele (2007), SNP studies may not, however, sufficiently cover the variation in the human genome. Copy number variations (CNVs) are intermediate-sized structural variants – insertions and deletions, duplications, and inversions – ranging from 1 kb to 5 Mb (Pollex and Hegele 2007). Since up to 5% of the human genome is estimated to be structurally variable, CNVs obviously play an important role in complex traits: Stranger et al. (2007) compared the expression levels of some 14 000 genes in 210 unrelated individuals and observed that of the associated genes, 84% were associated with SNPs, 18% with structural variants, and 1.3% with both. Developing techniques, such as affordable whole-genome sequencing, will facilitate the study of CNVs in the future.

The goal of medical genetic studies is to lighten the disease burden by improving the disease prevention and treatment. On the way to complete understanding of the genetic epidemiology of cardiovascular diseases, identifying single risk variants helps to target the efforts on those at substantial disease risk.

ACKNOWLEDGMENTS

This study was carried out in the Department of Molecular Medicine at the National Public Health Institute, Helsinki, Finland during 2000-2007. I wish to aknowledge the former and present directors, Jussi Huttunen and Pekka Puska, for providing outstanding research facilities.

Professor Tomi-Pekka Tuomainen and Docent Jukka Lehtonen are warmly thanked for reviewing this thesis. Their valuable criticism led to some essential improvements.

The study was financially supported by Aarne Koskelo Foundation, Biomedicum Helsinki Foundation, the Center of Excellence on Disease Genetics by the Academy of Finland, Ensio Hyvärinen Foundation, Ernfors Family Foundation, Finnish Cultural Foundation, the Finnish Foundation for Cardiovascular Research, the GenomEUtwin-project under the European Commission "Quality of Life and Management of the Living Resources" of 5th Framework Programme (QLG2-CT-2002–01254), Jenny and Antti Wihuri Foundation, Knut and Alice Wallenberg Foundation, the National Insitutes of Health/National Heart, Lung, and Blood Institute (1R01HL70150-01A1), the Nordic Center of Excellence in Disease Genetics, the Research Foundation of Orion Corporation, Sehlander Foundation, Sigrid Juselius Foundation, Swedish Diabetes Association, which all are gratefully acknowledged. I also wish to thank all the participants of the FINRISK and ULSAM studies.

I wish to express my deepest gratitude to my supervisors, Docent Markus Perola and Professor Leena Palotie. With Markus and Leena I have had a unique chance to grow as a scientist in an inspiring and supporting atmosphere. In the heart of top science I've learned some valuable lessons, the most important one being to never give up. I also want to thank for giving me a lot of responsibility and a chance to work independently, as well as for the privilidge to experience life outside research.

I'm also greatly indebted for Dr. Kaisa Silander for her help and advice in genotyping, and for Professor Veikko Salomaa for his efforts with the first article and, together with Dr. Kari Kuulasmaa, in all the FINRISK matters. Professor Ann-Christine Syvänen is thanked for her guidance with the third article, together with Dr. Matti Jauhiainen and Professor Marja-Riitta Taskinen, and our collaborators from Uppsala, Professor Björn Zethelius, Professor Christian Berne, and Professor Lars Lannfelt. Dr. Janna Saarela is thanked for her guidance especially in the beginning of this study, and Dr. Marjo Kestilä for her efforst duringthe last few months. Special thanks also go to Pekka, Anne, Minna, Siv, and Päivi T for their

extremely valuable help in the lab, to Heli and Susanna for their efforts in sequencing, and to Minttu for all the help with sampling. The MORGAM data center is acknowledged for excellent sample management, and Olli, Juha, and Tero for their statistical assistance.

I want to thank all the Queens, Kati, Mervi, Outi, and Johannes, of our office for their friendship and making the times in KTL and all the conference trips truly unforgettable. Kati and Mervi are thanked for, besides their excellent company, guidance in genetical and statistical issues. Mervi is also thanked for her company in the Marigold times, as well as for always sorting out all the sample questions. Kati, as I've said many times before, you should always be standing next to me in order to keep thing running. I want to thank Outi for all the long conversations we've had over the years and Johannes for his always positive attitude and cheery company. Sampo and Tero are also thanked for their excellent company in the office and conference trips.

My thanks also go to the special people in KTL for making coffee and work enjoyable. I also want to thank the med school girls for their company over the years, and my old friends for traveling to tyttömatkat and reminding me in other ways of life beyond medicine.

I want to thank my parents, sister, and grandfather for always believing in me.
Finally, I thank Sampo and Ansa, without whom nothing would matter.

Helsinki, December 14th, 2007

Kirsi Auro

REFERENCES

Abel M, Cellier C, Kumar N, Cerf-Bensussan N, Schmitz J, Caillat-Zucman S. Adulthood-onset celiac disease is associated with intercellular adhesion molecule-1 (ICAM-1) gene polymorphism. *Hum Immunol*. 2006,67:612-617.

Aitman TJ, Godsland IF, Farren B, Crook D, Wong HJ, Scott J. Defects of insulin action on fatty acid and carbohydrate metabolism in familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol*. 1997,17:748-754.

Alberti KG, Zimmet P, Shaw J: Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med* 2006,23:469-480.

Alfthan G, Pekkanen J, Jauhiainen M, Pitkäniemi J, Karvonen M, Tuomilehto J, Salonen JT, Ehnholm C. Relation of serum homocysteine and lipoprotein(a) concentrations to atherosclerotic disease in a prospective Finnish population based study. *Atherosclerosis*. 1994,106:9-19.

Allayee H, Dominguez KM, Aouizerat BE, Krauss RM, Rotter JI, Lu J, Cantor RM, de Bruin TW, Lusis AJ: Contribution of the hepatic lipase gene to the atherogenic lipoprotein phenotype in familial combined hyperlipidemia. *J Lipid Res* 2000, 41:245-252.

Allayee H, Krass KL, Pajukanta P, Cantor RM, van der Kallen CJ, Mar R, Rotter JI, de Bruin TW, Peltonen L, Lusis AJ. Locus for elevated apolipoprotein B levels on chromosome 1p31 in families with familial combined hyperlipidemia. *Circ Res*. 2002,90:926-931.

Amos CI. Successful Design and Conduct of Genome-Wide Association Studies. *Hum Mol Genet*. 2007 Jun 27

Aouizerat BE, Allayee H, Cantor RM, Dallinga-Thie GM, Lanning CD, de Bruin TW, Lusis AJ, Rotter JI: Linkage of a candidate gene locus to familial combined hyperlipidemia: lecithin:cholesterol acyltransferase on 16q. *Arterioscler Thromb Vasc Biol* 1999, 19:2730-2736.

Ariyaratnam R, Casas JP, Whittaker J, Smeeth L, Hingorani AD, Sharma P. Genetics of ischaemic stroke among persons of non-European descent: a meta-analysis of eight genes involving approximately 32,500 individuals. *PLoS Med.* 2007 Apr;4(4):e131.

Aromaa A. Kohonnut verenpaine ja sen kansanterveydellinen merkitys Suomessa. Helsinki: Kansaneläkelaitoksen julkaisuja AL:17, 1981

Aromaa A, Koskinen S (toim.) Terveys ja toimintakyky Suomessa. Terveys 2000 -tutkimuksen perustulokset. Kansanterveyslaitoksen julkaisuja B3/2002

Aso Y. Plasminogen activator inhibitor (PAI)-1 in vascular inflammation and thrombosis. *Front Biosci*. 2007,1:2957-2966. Review.

Auro K, Komulainen K, Alanne M, Silander K, Peltonen L, Perola M, Salomaa V. Thrombomodulin gene polymorphisms and haplotypes and the risk of cardiovascular events: a prospective follow-up study. *Arterioscler Thromb Vasc Biol*. 2006,26:942-947.

Auro K, Alanne M, Kristiansson K, Silander K, Kuulasmaa K, Salomaa V, Peltonen L, Perola M. Combined effects of thrombosis pathway gene variants predict cardiovascular events. *PLoS Genet*. 2007,3:e120.

Austin MA, Brunzell JD, Fitch WL, Krauss RM. Inheritance of low density lipoprotein subclass patterns in familial combined hyperlipidemia. *Arteriosclerosis*. 1990,10:520-350.

Ayyobi AF, Brunzell JD. Lipoprotein distribution in the metabolic syndrome, type 2 diabetes mellitus, and familial combined hyperlipidemia. *Am J Cardiol*. 2003,18(4A):27J-33J. Review.

Baessler A, Fischer M, Mayer B, Koehler M, Wiedmann S, Stark K, Doering A, Erdmann J, Riegger G, Schunkert H, Kwitek AE, Hengstenberg C. Epistatic interaction between haplotypes of the ghrelin ligand and receptor genes influence susceptibility to myocardial infarction and coronary artery disease. *Hum Mol Genet*. 2007,16:887-899.

Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic liability in stroke: a long-term follow-up study of Danish twins. *Stroke*. 2002,33:769-774.

Balkau B, Charles MA, Drivsholm T, Borch-Johnsen K, Wareham N, Yudkin JS, Morris R, Zavaroni I, van Dam R, Feskins E, Gabriel R, Diet M, Nilsson P, Hedblad B: Frequency of the WHO metabolic syndrome in European cohorts, and an alternative definition of an insulin resistance syndrome. *Diabetes Metab*. 2002,28:364-376.

Barengo NC, Kastarinen M, Lakka T, Nissinen A, Tuomilehto J. Different forms of physical activity and cardiovascular risk factors among 24-64-year-old men and women in Finland. *Eur J Cardiovasc Prev Rehabil*. 2006,13:51-59.

Barlow WE. Robust variance estimation for the case-cohort design. *Biometrics*. 1994,50:1064-1072.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005,21:263-265.

Barter PJ, Ballantyne CM, Carmena R, Castro Cabezas M, Chapman MJ, Couture P, de Graaf J, Durrington PN, Faergeman O, Frohlich J, Furberg CD, Gagne C, Haffner SM, Humphries SE, Jungner I, Krauss RM, Kwiterovich P, Marcovina S, Packard CJ, Pearson TA, Reddy KS, Rosenson R, Sarrafzadegan N, Sniderman AD, Stalenhoef AF, Stein E, Talmud PJ, Tonkin AM, Walldius G, Williams KM. Apo B versus cholesterol in estimating cardiovascular risk and in guiding therapy: report of the thirty-person/ten-country panel. *J Intern Med*. 2006,259:247-258.

Benjamini Y, Yekutieli D. Quantitative trait Loci analysis using the false discovery rate. *Genetics*. 2005,171:783-790.

Bernardi F, Faioni EM, Castoldi E, Lunghi B, Castaman G, Sacchi E, Mannucci PM. A factor V genetic component differing from factor V R506Q contributes to the activated protein C resistance phenotype. *Blood.* 1997, 15:1552-1557.

Bertram L, Hsiao M, McQueen MB, Parkinson M, Mullin K, Blacker D, Tanzi RE: The LDLR locus in Alzheimer's disease: a family-based study and meta-analysis of case-control data. *Neurobiol Aging.* 2007,28:18 e11-14.

Blann AD, Amiral J, McCollum CN. Prognostic value of increased soluble thrombomodulin and increased soluble E-selectin in ischaemic heart disease. *Eur J Haematol*. 1997,59:115-120.

Blum A, Khazim K, Merei M, Peleg A, Blum N, Vaispapir V. The stroke trial - can we predict clinical outcome of patients with ischemic stroke by measuring soluble cell adhesion molecules (CAM)? *Eur Cytokine Netw*. 2006,17:295-298.

Boekholdt SM, Kramer MH. Arterial thrombosis and the role of thrombophilia. *Semin Thromb Hemost*. 2007,33:588-596.

Boring L, Gosling J, Cleary M, Charo IF. Decreased lesion formation in CCR2-/- mice revelas a role for chemokines in the initiation of atherosclerosis. *Nature*. 1998,394:894-897.

Braun M, Pietsch P, Schror K, Baumann G, Felix SB. Cellular adhesion molecules on vascular smooth muscle cells. Cardiovasc Res. 1999,41:395-401. Review.

Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. *Science*. 2002,296:752-755.

Briollais L, Wang Y, Rajendram I, Onay V, Shi E, Knight J, Ozcelik H. Methodological issues in detecting gene-gene interactions in breast cancer susceptibility: a population-based study in Ontario. *BMC Med*. 2007,5:22.

Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, van Dam M, Yu L, Brewer C, Collins JA, Molhuizen HO, et al.: Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet* 1999,22:336-345.

Brown RD, Whisnant JP, Sicks JD, O'Fallon WM, Wiebers DO. Stroke incidence, prevalence, and survival: secular trends in Rochester, Minnesota, through 1989. *Stroke*. 1996,27:373-380.

Brunzell JD, Albers JJ, Chait A, Grundy SM, Groszek E, McDonald GB. Plasma lipoproteins in familial combined hyperlipidemia and monogenic familial hypertriglyceridemia. *J Lipid Res*. 1983,24:147-155.

Buil A, Soria JM, Souto JC, Almasy L, Lathrop M, Blangero J, Fontcuberta J. Protein C levels are regulated by a quantitative trait locus on chromosome 16: results from the Genetic Analysis of Idiopathic Thrombophilia (GAIT) Project. *Arterioscler Thromb Vasc Biol.* 2004,24:1321-1325.

Cakir O, Ayyildiz O, Oruc A, Eren N. A young adult with coronary artery and jugular vein thrombosis: A case report of combined protein S and protein C deficiency. *Heart Vessels.* 2002,17: 74–76.

Carlsson E, Almgren P, Hoffstedt J, Groop L, Ridderstrale M: The FOXC2 C-512T polymorphism is associated with obesity and dyslipidemia. *Obes Res.* 2004,12:1738-1743.

Cecil Textbook of Medicine. Edited by Goldman and Ausiello. 22nd edition, Elsewier. 2004, p. 975, 400-424, 2281-2298.

Chan JM, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care*. 1994,17:961-969.

Chan WP, Lee CK, Kwong YL, Lam CK, Liang R. A novel mutation of Arg306 of factor V gene in Hong Kong Chinese. *Blood.* 1998, 15:1135-1139.

Chao TH, Li YH, Chen JH, Wu HL, Shi GY, Tsai WC, Chen PS, Liu PY. Relation of thrombomodulin gene polymorphisms to acute myocardial infarction in patients <or =50 years of age. *Am J Cardiol*. 2004,93:204-207.

Chen H, Hernandez W, Shriver MD, Ahaghotu CA, Kittles RA. ICAM gene cluster SNPs and prostate cancer risk in African Americans. *Hum Genet*. 2006,120:69-76.

Clark AG. The role of haplotypes in candidate gene studies. *Genet Epidemiol*. 2004,27:321-333. Review.

Clarke GM, Carter KW, Palmer LJ, Morris AP, Cardon LR. Fine Mapping versus Replication in Whole-Genome Association Studies. *Am J Hum Genet*. 2007,81:995-1005.

Clee SM, Zwinderman AH, Engert JC, Zwarts KY, Molhuizen HO, Roomp K, Jukema JW, van Wijland M, van Dam M, Hudson TJ, et al.: Common genetic variation in ABCA1 is associated with altered lipoprotein levels and a modified risk for coronary artery disease. *Circulation* .2001,103:1198-1205.

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science.* 2004,305:869-872.

Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH: Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med.* 2006,354:1264-1272.

Colditz GA, Willett WC, Rotnitzky A, Manson JE. Weight gain as a risk factor for clinical diabetes mellitus in women. *Ann Intern Med*. 1995,122:481-486.

Cole JW, Roberts SC, Gallagher M, Giles WH, Mitchell BD, Steinberg KK, Wozniak MA, Macko RF, Reinhart LJ, Kittner SJ. Stroke Prevention in Young Women Study. Thrombomodulin Ala455Val Polymorphism and the risk of cerebral infarction in a biracial population: the Stroke Prevention in Young Women Study. *BMC Neurol*. 2004,4:21.

Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science.* 1997,278:1580-1581.

Conneely KN, Boehnke M. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. *Am J Hum Genet*. 2007,81.

Constans J, Conri C. Circulating markers of endothelial function in cardiovascular disease. *Clin Chim Acta*. 2006,368:33-47. Review.

Coon H, Xin Y, Hopkins PN, Cawthon RM, Hasstedt SJ, Hunt SC: Upstream stimulatory factor 1 associated with familial combined hyperlipidemia, LDL cholesterol, and triglycerides. *Hum Genet.* 2005,117:444-451.

Cordell HJ, Todd JA, Hill NJ, Lord CJ, Lyons PA, Peterson LB, Wicker LS, Clayton DG. Statistical modeling of interlocus interactions in a complex disease: rejection of the multiplicative model of epistasis in type 1 diabetes. *Genetics*. 2001,158:357-367.

Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet*. 2002,11:2463-2468. Review.

Curtis D, Sham PC, Vallada HP. Genetic analysis of complex disease. *Nat Genet*. 1995,9:13.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet*. 2001,29:229-232.

Damcott CM, Hoppman N, Ott SH, Reinhart LJ, Wang J, Pollin TI, O'Connell JR, Mitchell BD, Shuldiner AR: Polymorphisms in both promoters of hepatocyte nuclear factor 4-alpha are associated with type 2 diabetes in the Amish. *Diabetes* 2004,53:3337-3341.

Danesh J, Collins R, Peto R. Lipoprotein(a) and coronary heart disease. Meta-analysis of prospective studies. *Circulation*. 2000,102:1082-1085.

Darvall KA, Sam RC, Silverman SH, Bradbury AW, Adam DJ. Obesity and thrombosis. *Eur J Vasc Endovasc Surg*. 2007,33:223-233. Review.

Davies MJ, Gordon JL, Gearing AJ, Pigott R, Woolf N, Katz D, Kyriakopoulos A. The expression of the adhesion molecules ICAM-1, VCAM-1, PECAM, and E-selectin in human atherosclerosis. *J Pathol*. 1993,171:223-229.

Davies MJ. Stability and instability: the two faces of coronary atherosclerosis. *Circulation*. 1996,94:2013-2020.

de Graaf J, Stalenhoef AF. Defects of lipoprotein metabolism in familial combined hyperlipidaemia. *Curr Opin Lipidol.* 1998,9:189-196. Review.

de Ferranti SD, Rifai N. C-reactive protein: a nontraditional serum marker of cardiovascular risk. *Cardiovasc Pathol*. 2007,16:14-21. Review.

de Moerloose P, Boehlen F. Inherited thrombophilia in arterial disease: a selective review. *Semin Hematol*. 2007,44:106-113. Review.

de Paula Sabino A, Ribeiro DD, Carvalho MG, Cardoso J, Dusse LM, Fernandes AP. Factor V Leiden and increased risk for arterial thrombotic disease in young Brazilian patients. *Blood Coagul Fibrinolysis*. 2006,17:271-275.

Devroey D, De Swaef N, Coigniez P, Vandevoorde J, Kartounian J, Betz W. Correlations between lipid levels and age, gender, glycemia, obesity, diabetes, and smoking. *Endocr Res.* 2004,30:83-93.

Doggen CJM, Kunz G, Rosendaal FR, et al. A mutation in the thrombomodulin gene, 127G to A coding for Ala25Thr, and the risk of myocardial infarction. *Thromb Haemost*. 1998;80:743

Dong ZM, Chapman SM, Brown AA, Frenette PS, Hynes RO, Wagner DD. The combined role of P- and E-selectins in atherosclerosis. *J Clin Invest*. 1998,102:145-152.

Drazen JM, Phimister EG. Publishing Genomewide Association Studies. *N Engl J Med*. 2007 Jul 18;

Ehrensperger E, Minuk J, Durcan L, Mackey A, Wolfson C, Fontaine AM, Cote R. Predictive value of soluble intercellular adhesion molecule-1 for risk of ischemic events in individuals with cerebrovascular disease. *Cerebrovasc Dis*. 2005,20:456-462.

Eichner JE, Dunn ST, Perveen G, Thompson DM, Stewart KE, Stroehla BC: Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. *Am J Epidemiol.* 2002,155:487-495.

Eterovic D, Titlic M, Culic V, Zadro R, Primorac D. Lower contribution of factor V Leiden or G202104 mutations to ischemic stroke in patients with clinical risk factors: pair-matched case-control study. *Clin Appl Thromb Hemost*. 2007,13:188-193.

Eurlings PM, van der Kallen CJ, Vermeulen VM, de Bruin TW: Variants in the PPARgamma gene affect fatty acid and glycerol metabolism in familial combined hyperlipidemia. *Mol Genet Metab.* 2003,80:296-301.

Evans DM, Cardon LR. Genome-wide association: a promising start to a long race. *Trends Genet*. 2006,22:350-354

Fasano T, Cefalu AB, Di Leo E, Noto D, Pollaccia D, Bocchi L, Valenti V, Bonardi R, Guardamagna O, Averna M, et al.: A Novel Loss of Function Mutation of PCSK9 Gene in White Subjects With Low-Plasma Low-Density Lipoprotein Cholesterol. *Arterioscler Thromb Vasc Biol.* 2007,27:677-681.

Fernandez JA, Gruber A, Heeb MJ, Griffin JH. Protein C pathway impairment in nonsymptomatic cigarette smokers. *Blood Cells Mol Dis*. 2002,29:73-82.

Franco RF, Elion J, Santos SEB, Araújo AG, Tavella MH, Zago MA. Heterogenous ethnic distribution of factor V Leiden mutation. *Genet Molec Biol*. 1999,22:143-145.

Franco RF, Reitsma PH. Genetic risk factors for venous thrombosis. *Hum Genet*. 2001,109:369-384.

Frikke-Schmidt R, Nordestgaard BG, Schnohr P, Steffensen R, Tybjaerg-Hansen A: Mutation in ABCA1 predicted risk of ischemic heart disease in the Copenhagen City Heart Study Population. *J Am Coll Cardiol.* 2005,46:1516-1520.

Fruchart JC, Nierman MC, Stroes ES, Kastelein JJ, Duriez P. New risk factors for atherosclerosis and patient risk assessment. *Circulation*. 2004,109(Suppl 1):III15-9. Review.

Gagnon F, Jarvik GP, Motulsky AG, Deeb SS, Brunzell JD, Wijsman EM: Evidence of linkage of HDL level variation to APOC3 in two samples with different ascertainment. *Hum Genet* 2003,113:522-533.

Galkina E, Ley K. Vascular adhesion molecules in atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2007,27:2292-2301.

Garcia-Closas M, Malats N, Real FX, Welch R, Kogevinas M, Chatterjee N, Pfeiffer R, Silverman D, Dosemeci M, Tardon A, Serra C, Carrato A, Garcia-Closas R, Castano-Vinyals G, Chanock S, Yeager M, Rothman N. Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2006,15:536-542.

Geurts JM, Janssen RG, van Greevenbroek MM, van der Kallen CJ, Cantor RM, Bu X, Aouizerat BE, Allayee H, Rotter JI, de Bruin TW: Identification of TNFRSF1B as a novel modifier gene in familial combined hyperlipidemia. *Hum Mol Genet* 2000,9:2067-2074.

Gibson F, Hercberg S, Froguel P. Common polymorphisms in the USF1 gene are not associated with type 2 diabetes in French Caucasians. *Diabetes.* 2005,54:3040-3042.

Gimbrone MA, Topper JN, Nagel T, Andersson KR, Garcia-Cardena G. Endothelial dysfunction, hemodynamic forces, and atherogenesis. *Ann N Y Acad Sci*. 2000,902:230-239. Review.

Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. Science. 2002, 298:2345-2349. Review.

Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR, Garside DB, Wilson PW. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *JAMA*. 2003,290:891-897.

Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet.* 2003,35:131-138.

Goldstein JL, Schrott HG, Hazzard WR, Bierman EL, Motulsky AG. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J Clin Invest*. 1973,52:1544-1568.

Goldstein LB, Adams R, Alberts MJ, Appel LJ, Brass LM, Bushnell CD, Culebras A, Degraba TJ, Gorelick PB, Guyton JR, Hart RG, Howard G, Kelly-Hayes M, Nixon JV, Sacco RL; American Heart Association/American Stroke Association Stroke Council; Atherosclerotic Peripheral Vascular Disease Interdisciplinary Working Group; Cardiovascular Nursing Council; Clinical Cardiology Council; Nutrition, Physical Activity, and Metabolism Council; Quality of Care and Outcomes Research Interdisciplinary Working Group; American Academy of Neurology. Primary prevention of ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council: cosponsored by the Atherosclerotic Peripheral Vascular Disease Interdisciplinary Working Group; Cardiovascular Nursing Council; Clinical Cardiology Council; Nutrition, Physical Activity, and Metabolism Council; and the Quality of Care and Outcomes Research Interdisciplinary Working Group: the American Academy of Neurology affirms the value of this guideline. *Stroke*. 2006,37:1583-1633. Erratum in: *Stroke*. 2007,38:207.

Gotto AM Jr, Pownall HJ, Havel RJ. Introduction to the plasma lipoproteins. *Methods Enzymol*. 1986,128:3-41. Review.

Grey ST, Tsuchida A, Hau H, Orthner CL, Salem HH, Hancock WW. Selective inhibitory effects of the anticoagulant activated protein C on the responses of human mononuclear phagocytes to LPS, IFN-gamma, or phorbol ester. *J Immunol*. 1994,153:3664-3672.

Griffin JH, Fernandez JA, Mosnier LO, Liu D, Cheng T, Guo H, Zlokovic BV. The promise of protein C. *Blood Cells Mol Dis*. 2006,36:211-216. Review.

Gruenewald TL, Seeman TE, Ryff CD, Karlamangla AS, Singer BH. Combinations of biomarkers predictive of later life mortality. *Proc Natl Acad Sci U S A*. 2006,103:14158-14163.

Haffner SJ, Cassells H. Hyperglycemia as a cardiovascular risk factor. *Am J Med*. 2003,115 Suppl 8A:6S-11S. Review.

Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu HQ, Polychronakos C. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature*. 2007 Jul 15.

Hansen SK, Rose CS, Glumer C, Drivsholm T, Borch-Johnsen K, Jorgensen T, Pedersen O, Hansen T: Variation near the hepatocyte nuclear factor (HNF)-4alpha gene associates with type 2 diabetes in the Danish population. *Diabetologia.* 2005,48:452-458.

Haugt WH, Mansour M, Rothlein R, Kishimoto TK, Mainolfi EA, Hendricks JB, Hendricks C, Mehta JL. Alterations in circulating intercellular adhesion molecule-1 and L-selectin: further evidence for chronic inflammation in ischemic heart disease. *Am Heart J*. 1996,132(1 Pt 1):1-8.

Hegele RA. Familial hypercholesterolemia. *N Engl J Med*. 2007,356:1779.

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, Gudmundsson G, Grant SF, Thorgeirsson G, Sveinbjornsdottir S, Valdimarsson EM, Matthiasson SE, Johannsson H, Gudmundsdottir O, Gurney ME, Sainz J, Thorhallsdottir M, Andresdottir M, Frigge ML, Topol EJ, Kong A, Gudnason V, Hakonarson H, Gulcher JR, Stefansson K. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet.* 2004,36:233-239.

Heller DA, de Faire U, Pedersen NL, Dahlen G, McClearn GE. Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med.* 1993,328:1150-1156.

Herbert A, Gerry NP, McQueen MB, et al. A common genetic variant is associated with adult and childhood obesity. *Science*. 2006;312:279 – 283.

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

Hodge SE. Some epistatic two-locus models of disease. I. Relative risks and identity-by-descent distributions in affected sib pairs. *Am J Hum Genet*. 1981,33:381-395.

Hoff JA, Chomka EV, Krainik AJ, Daviglus M, Rich S, Kondos GT. Age and gender distributions of coronary artery calcium detected by electron beam tomography in 35,246 adults. *Am J Cardiol*. 2001,87:1335-1339.

Hoffer MJ, Snieder H, Bredie SJ, Demacker PN, Kastelein JJ, Frants RR, Stalenhoef AF: The V73M mutation in the hepatic lipase gene is associated with elevated cholesterol levels in four Dutch pedigrees with familial combined hyperlipidemia. *Atherosclerosis* 2000,151:443-450.

Hofstra JJ, Schouten M, Levi M. Thrombophilia and outcome in severe infection and sepsis. *Semin Thromb Hemost*. 2007,33:604-609.

Hokanson JE, Austin MA, Zambon A, Brunzell JD. Plasma triglyceride and LDL heterogeneity in familial combined hyperlipidemia. *Arterioscler Thromb*. 1993,13:427-434.

Hong Y, de Faire U, Heller DA, McClearn GE, Pedersen N. Genetic and environmental influences on blood pressure in elderly twins. *Hypertension*. 1994,24:663-670.

Hsu TM, Chen X, Duan S, Miller RD, Kwok PY. Universal SNP genotyping assay with fluorescence polarization detection. *Biotechniques*. 2001,31:560, 562, 564-568, passim.

Huang Z, Willett WC, Manson JE, Rosner B, Stampfer MJ, Speizer FE, Colditz GA. Body weight, weight change, and risk for hypertension in women. *Ann Intern Med.* 1998,128:81-88.

Huertas-Vazquez A, Aguilar-Salinas C, Lusis AJ, Cantor RM, Canizales-Quinteros S, Lee JC, Mariana-Nunez L, Riba-Ramirez RM, Jokiaho A, Tusie-Luna T, Pajukanta P: Familial combined hyperlipidemia in Mexicans: association with upstream transcription factor 1 and linkage on chromosome 16q24.1. *Arterioscler Thromb Vasc Biol.* 2005,25:1985-1991.

Humpries SE, Morgan L. Genetic risk factors for stroke and carotid atherosclerosis: insights into pathophysiology from candidate gene approaches. *Lancet Neurol*. 2004,3:227-235. Review.

Hunt SC, Wu LL, Hopkins PN, Stults BM, Kuida H, Ramirez ME, Lalouel JM, Williams RR. Apolipoprotein, low density lipoprotein subfraction, and insulin associations with familial combined hyperlipidemia. Study of Utah patients with familial dyslipidemic hypertension. *Arteriosclerosis*. 1989,9:335-344.

Hunter DJ, Kraft P. Drinking from the Fire Hose -- Statistical Issues in Genomewide Association Studies. *N Engl J Med*. 2007 Jul 18;

Hwang SJ, Ballantyne CM, Sharret AR, Smith LC, Davis CE, Gotto AM jr, Boerwinkle E. Circulating adhesion molecules VCAM-1, ICAM-1, and E-selectin in carotid atherosclerosis and incident coronary heart disease cases: the Atherosclerosis Risk In Communities (ARIC) study. *Circulation.* 1997,96:4219–4225.

Inoue S, Egashira K, Ni W, Kitamoto S, Usui M, Otani K, Ishibashi M, Hiasa K, Nishida K, Takeshita A. Anti-monocyte chemoattractant protein-1 gene therapy limits progression and destabilization of established atherosclerosis in apolipoprotein E.knock-out mice. *Circulation*. 2002,106:2700-2706.

Ireland H, Kunz G, Kyriakoulis K, Stubbs PJ, Lane DA. Thrombomodulin gene mutations associated with myocardial infarction. *Circulation*. 1997,96:15-18.

Ivan CS, Seshadri S, Beiser A, Au R, Kase CS, Kelly-Hayes M, Wolf PA. Dementia after stroke: the Framingham Study. *Stroke.* 2004,35:1264-1268.

Jakovljevic D, Sarti C, Sivenius J, Torppa J, Mähönen M, Immonen-Räihä P, Kaarsalo E, Alhainen K, Kuulasmaa K, Tuomilehto J, Puska P, Salomaa V. Socioeconomic status and ischemic stroke: The FINMONICA Stroke Register. *Stroke.* 2001,32:1492-1498.

Jauch EC, Lindsell C, Broderick J, Fagan SC, Tilley BC, Levine SR; NINDS rt-PA Stroke Study Group. Association of serial biochemical markers with acute ischemic stroke: the National Institute of Neurological Disorders and Stroke recombinant tissue plasminogen activator Stroke Study. *Stroke*. 2006,37:2508-2513.

Jauhiainen M, Koskinen P, Ehnholm C, Frick MH, Mänttäri M, Manninen V, Huttunen JK. Lipoprotein (a) and coronary heart disease risk: a nested case-control study of the Helsinki Heart Study participants. *Atherosclerosis*. 1991,89:59-67.

Jenny NS, Arnold AM, Kuller LH, Sharret AR, Fried LP, Psaty BM, Tracy RP. Soluble intracellular adhesion molecule -1 is associated with cardiovascular disease risk and mortality in older adults. *J Thromb Haemost*. 2006,4:107-113.

Jousilahti P, Vartiainen E, Tuomilehto J, 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*. 1999,99:1165-1172.

Kaarisalo MM, Räihä I, Sivenius J, Immonen-Räihä P, Lehtonen A, Sarti C, Mähönen M, Torppa J, Tuomilehto J, Salomaa V. Diabetes worsens the outcome of acute ischemic stroke. *Diabetes Res Clin Pract*. 2005,69:293-298.

Kammerer S, Roth RB, Reneland R, Marnellos G, Hoyal CR, Markward NJ, Ebner F, Kiechle M, Schwarz-Boeger U, Griffiths LR, Ulbrich C, Chrobok K, Forster G, Praetorius GM, Meyer P, Rehbock J, Cantor CR, Nelson MR, Braun A. Large-scale association study identifies ICAM gene region as breast and prostate cancer susceptibility locus. *Cancer Res*. 2004,64:8906-8910.

Kannel WB. Blood pressure as a cardiovascular risk factor: prevention and treatment. *JAMA.* 1996,275:1571-1576.

Kannel WB. Overview of hemostatic factors involved in atherosclerotic cardiovascular disease. *Lipids.* 2005,40:1215-1220. Review.

Kathiresan S, Yang Q, Larson MG, Camargo AL, Tofler GH, Hirschhorn JN, Gabriel SB, 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.* 2006,26:1405-1412.

Katzov H, Chalmers K, Palmgren J, Andreasen N, Johansson B, Cairns NJ, Gatz M, Wilcock GK, Love S, Pedersen NL, Brookes AJ, Blennow K, Kehoe PG, Prince JA. Genetic variants of ABCA1 modify Alzheimer disease risk and quantitative traits related to beta-amyloid metabolism. *Hum Mutat*. 2004,23:358-367.

Kawamura T, Umemura T, Kanai A, Nagashima M, Nakamura N, Uno T, Nakayama M, Sano T, Hamada Y, Nakamura J, Hotta N. Soluble adhesion molecules and C-reactive protein in the progression of silent cerebral infarction in patients with type 2 diabetes mellitus. *Metabolism*. 2006,55:461-466.

Kiechl S, Willeit J, Mayr M, Viehweider B, Oberhollenzer M, Kronenberg F, Wiedermann CJ, Oberthaler S, Xu Q, Witztum JL, Tsimikas S. Oxidized phospholipids, lipoprotein(a), lipoprotein-associated phospholipase A2 activity, and 10-year cardiovascular outcomes: prospective results from the Bruneck study. *Arterioscler Thromb Vasc Biol.* 2007,27:1788-1795.

Kinlay S, Libby P, Ganz P. Endothelium function and coronary artery disease. *Curr Opin Lipidol*. 2001,12:383-9. Review.

Klos KL, Sing CF, Boerwinkle E, Hamon SC, Rea TJ, Clark A, Fornage M, Hixson JE: Consistent effects of genes involved in reverse cholesterol transport on plasma lipid and apolipoprotein levels in CARDIA participants. *Arterioscler Thromb Vasc Biol.* 2006,26:1828-1836.

Knight JC. Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin Sci (Lond)*. 2003,104:493-501. Review.

Knight JC. Regulatory polymorphisms underlying complex disease traits. *J Mol Med*. 2005,83:97-109. Review.

Knoblauch H, Bauerfeind A, Toliat MR, Becker C, Luganskaja T, Gunther UP, Rohde K, Schuster H, Junghans C, Luft FC, Nurnberg P, Reich JG. Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet*. 2004,13:993-1004.

Kohler HP, Stickland MH, Ossei-Gerning N, Carter A, Mikkola H, Grant PJ. Association of a common polymorphism in the factor XIII gene with myocardial infarction. *Thromb Haemost*. 1998;79:8

Komulainen K, Alanne M, Auro K, Kilpikari R, Pajukanta P, Saarela J, Ellonen P, Salminen K, Kulathinal S, Kuulasmaa K, Silander K, Salomaa V, Perola M, Peltonen L. Risk alleles of USF1 gene predict cardiovascular disease of women in two prospective studies. *PLoS Genet*. 2006, 2:e69.

Kong WJ, Liu J, Jiang JD: Human low-density lipoprotein receptor gene and its regulation. *J Mol Med* 2006, 84:29-36.

Konstantoulas CJ, Cooper J, Warnock G, Miller GJ, Humphries SE, Ireland H. A combination of two common thrombomodulin gene variants (-1208-1209TTdelTT and A455V) influence risk of coronary heart disease: a prospective study in men. *Atherosclerosis*. 2004,177:97-104.

Konstantoulas CJ, Cooper JA, Ohlin AK, Humphries SE, Goodall AH, Toh CH, Mather H, Ireland H. Low soluble thrombomodulin activity and antigen is associated with a family history of heart disease while a high level is associated with a personal history of heart disease in type 2 diabetes. *Thromb Haemost*. 2007,97:161-164.

Korshunov VA, Schwartz SM, Berk BC. Vascular remodeling: hemodynamic and biochemical mechanisms underlying Glagov's phenomenon. *Arterioscler Thromb Vasc Biol*. 2007,27:1722-1728. Review.

Koskenvuo M, Kaprio J, Romanov K. Twin studies in metabolic diseases. *Ann Med.* 1992 24:379-381.

Koskinen P, Mänttäri M, Manninen V, Huttunen JK, Heinonen OP, Frick MH. Coronary heart disease incidence in NIDDM patients in the Helsinki Heart Study. *Diabetes Care.* 1992,15:820-825.

Laaksonen DE, Lakka HM, Niskanen LK, Kaplan GA, Salonen JT, Lakka TA. Metabolic syndrome and development of diabetes mellitus: application and validation of recently suggested definitions of the metabolic syndrome in a prospective cohort study. *Am J Epidemiol*. 2002,156:1070-1077.

Laan M, Pääbo S. Demographic history and linkage disequilibrium in human populations. *Nat Genet*. 1997,17:435-438.

Laatikainen T, Tapanainen H, Alfthan G, Salminen I, Sundvall J, Leiviskä J, Harald K, Jousilahti P, Salomaa V, Vartiainen E. FINRISKI 2002: Tutkimuksen toteutus ja tulokset 1. Kansanterveyslaitos 2003.

Lai CQ, Parnell LD, Ordovas JM. The APOA1/C3/A4/A5 gene cluster, lipid metabolism and cardiovascular disease risk. *Curr Opin Lipidol*. 2005,16:153-166. Review.

Lalouschek W, Schillinger M, Hsieh K, Endler G, Tentschert S, Lang W, Cheng S, Mannhalter C. Matched case-control study on factor V Leiden and the prothrombin G20210A mutation in patients with ischemic stroke/transient ischemic attack up to the age of 60 years. *Stroke*. 2005,361405-1409.

Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994, 265:2037-2048. Review. Erratum in: Science 1994,266:353.

Lane DA, Grant PJ. Role of hemostatic gene polymorphisms in venous and arterial thrombotic disease. *Blood.* 2000 ,95:1517-1532. Review.

Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, Salonen JT: The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA.* 2002,288:2709-2716.

Ledermann HM. Is maturity onset diabetes at young age (MODY) more common in Europe than previously assumed? *Lancet.* 1995,345:648.

Lee JC, Weissglas-Volkov D, Kyttala M, Sinsheimer JS, Jokiaho A, de Bruin TW, Lusis AJ, Brennan ML, van Greevenbroek MM, van der Kallen CJ, Hazen SL, Pajukanta P. USF1 contributes to high serum lipid levels in Dutch FCHL families and U.S. whites with coronary artery disease. *Arterioscler Thromb Vasc Biol*. 2007,27:2222-2227.

Leshinsky-Silver E, Cheng S, Grow MA, Shoshana S, Scharf L, Lev D, Boaz M, Brunner D, Zimlichman R. Candidate gene polymorphism in cardiovascular disease: the BIP cohort. *Isr Med Assoc J*. 2006,8:103-105.

Li MD, Payne TJ, Ma JZ, Lou XY, Zhang D, Dupont RT, Crews KM, Somes G, Williams NJ, Elston RC. A genomewide search finds major susceptibility loci for nicotine dependence on chromosome 10 in African Americans. *Am J Hum Genet*. 2006,79:745-751.

Li YF, Tsao YH, Gauderman WJ, Conti DV, Avol E, Dubeau L, Gilliland FD. Intercellular adhesion molecule-1 and childhood asthma. *Hum Genet*. 2005,117:476-484.

Li YH, Chen JH, Tsai WC, Chao TH, Guo HR, Tsai LM, Wu HL, Shi GY. Synergistic effect of thrombomodulin promoter -33G/A polymorphism and smoking on the onset of acute myocardial infarction. *Thromb Haemost*. 2002,87:86-91.

Libby P. Inflammation in atherosclerosis. *Nature*. 2002,420:868-874. Review.

Libby P, Theroux P. Pathophysiology of coronary artery disease. *Circulation*. 2005,111:3481-3488.

Lilja HE, Soro A, Ylitalo K, Nuotio I, Viikari JS, Salomaa V, Vartiainen E, Taskinen MR, Peltonen L, Pajukanta P: A candidate gene study in low HDL-cholesterol families provides evidence for the involvement of the APOA2 gene and the APOA1C3A4 gene cluster. *Atherosclerosis.* 2002,164:103-111.

Linnemann B, Schindewolf M, Zgouras D, Erbe M, Jarosch-Preusche M, Lindhoff-Last E. Are patients with thrombophilia and previous venous thromboembolism at higher risk to arterial thrombosis? *Thromb Res*. 2007.

Lloyd-Jones DM, Wilson PWF, Larson MG, Leip EP, Beiser A, D'Agostino RB, Cleeman JI, Levy D. Lifetime risk for coronary heart disease by cholesterol levels at selected ages. *Arch Intern Med.* 2003,163:1966–1972.

Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PW, Wolf PA, Levy D. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation.* 2006,113:791-798. (a)

Lloyd-Jones DM, Liu K, Tian L, Greenland P. Narrative review: Assessment of C-reactive protein in risk prediction for cardiovascular disease. *Ann Intern Med*. 2006,145:35-42. Review. (b)

Loos RJ, Bouchard C. Obesity--is it a genetic disorder? *J Intern Med*. 2003,254:401-425. Review.

Loukola A, Broms U, Maunu H, Widen E, Heikkilä K, Siivola M, Salo A, Pergadia ML, Nyman E, Sammalisto S, Perola M, Agrawal A, Heath AC, Martin NG, Madden PA, Peltonen L, Kaprio J. Linkage of nicotine dependence and smoking behavior on 10q, 7q and 11p in twins with homogeneous genetic background. *Pharmacogenomics J*. 2007 Jun.

Love-Gregory LD, Wasson J, Ma J, Jin CH, Glaser B, Suarez BK, Permutt MA: A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4 alpha gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an ashkenazi jewish population. *Diabetes.* 2004,53:1134-1140.

Lusis AJ. Atherosclerosis. *Nature*. 2000,407:233- 241. Review.

Ma J, Mollsten A, Prazny M, Falhammar H, Brismar K, Dahlquist G, Efendic S, Gu HF. Genetic influences of the intercellular adhesion molecule 1 (ICAM-1) gene polymorphisms in development of Type 1 diabetes and diabetic nephropathy. *Diabet Med*. 2006,23:1093-1099.

MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, Abbott R, Godwin J, Dyer A, Stamler J. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet.* 1990,335:765-774.

Mahmoudi M, Curzen N, Gallagher PJ. Atherogenesis: the role of inlammation and infection. *Histopathology*. 2007,50:535-546. Review.

Maier LM, Chapman J, Howson JM, Clayton DG, Pask R, Strachan DP, McArdle WL, Twells RC, Todd JA. No evidence of association or interaction between the IL4RA, IL4, and IL13 genes in type 1 diabetes. *Am J Hum Genet*. 2005,76:517-521.

Malik I, Danesh J, Whincup P, Bhatia V, Papacosta O, Walker M, Lennon L, Thomson A, Haskard D. Soluble adhesion molecules and prediction of coronary heart disease: a prospective study and meta-analysis. *Lancet.* 2001,358:971-976.

Mamun AA, Peeters A, Barendregt J, Willekens F, Nusselder W, Bonneux L. Smoking decreases the duration of life lived with and without cardiovascular disease: a life course analysis of the Framingham Heart Study. *Eur Heart J.* 2004,25: 409–415.

Manson JE, Colditz GA, Stampfer MJ, Willett WC, Krolewski AS, Rosner B, Arky RA, Speizer FE, Hennekens CH. A prospective study of maturity-onset diabetes mellitus and risk of coronary heart disease and stroke in women. *Arch Intern Med.* 1991,151:1141-1147.

Marchesini G, Forlani G, Cerrelli F, Manini R, Natale S, Baraldi L, Ermini G, Savorani G, Zocchi D, Melchionda N. WHO and ATPIII proposals for the definition of the metabolic syndrome in patients with Type 2 diabetes. *Diabet Med*. 2004,21:383-387.

Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med*. 1994,330:1041-1046.

Martins IJ, Hone E, Foster JK, Sunram-Lea SI, Gnjec A, Fuller SJ, Nolan D, Gandy SE, Martins RN: Apolipoprotein E, cholesterol metabolism, diabetes, and the convergence of risk factors for Alzheimer's disease and cardiovascular disease. *Mol Psychiatry.* 2006,11:721-736.

Matijevic N, Wu KK. Hypercoagulable states and strokes. *Curr Atheroscler Rep*. 2006,8:324-329. Review.

McCarthy MI. Progress in defining the molecular basis of type 2 diabetes mellitus through susceptibility-gene identification. *Hum Mol Genet*. 2004,13 Spec No 1:R33-41. Review.

McGill HC Jr, McMahan CA, Zieske AW, Sloop GD, Walcott JV, Troxclair DA, Malcom GT, Tracy RE, Oalmann MC, Strong JP. Associations of coronary heart disease risk factors with the intermediate lesion of atherosclerosis in youth. The Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Research Group. *Arterioscler Thromb Vasc Biol*. 2000,20:1998-2004.

McGlinchey PG, Spence MS, Patterson CC, Allen AR, Murphy G, Belton C, McKeown PP. The intercellular adhesion molecule-1 (ICAM-1) gene K469E polymorphism is not associated with ischaemic heart disease: an investigation using family-based tests of association. *Eur J Immunogenet*. 2004,31:201-206.

Meade TW, Mellows S, Brozovic M, Miller GJ, Chakrabarti RR, North WR, Haines AP, Stirling Y, Imeson JD, Thompson SG. Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. *Lancet*. 1986,2:533-537.

Miller GJ, Bauer KA, Cooper JA, Rosenberg RD. Activation of the coagulant pathway in cigarette smokers. *Thromb Haemost*. 1998,79:549-553.

Miller DT, Ridker PM, Libby P, Kwiatkowski DJ. Atherosclerosis: the path from genomics to therapeutics. *J Am Coll Cardiol*. 2007,49:1589-1599. Review.

Miyoshi T, Yuan Z, Shi W. Association of a Vcam1 mutation with atherosclerosis susceptibility in diet-induced models of atherosclerosis. *Atherosclerosis*. 2007 Jun 15;

Mizutani A, Okajima K, Uchiba M, Noguchi T. Activated protein C reduces ischemia/reperfusion-induced renal injury in rats by inhibiting leukocyte activation. *Blood.* 2000,95:3781-3787.

Moore KL. Clinically oriented anatomy. Williams & Wilkins, 3rd edition. 1991, pp. 98-99, 698.

Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *JAMA*. 2007,297:1551-1561. Review.

Murase T, Okubo M, Amemiya-Kudo M, Hiraga T, Oka J, Shimada M, Igarashi T. Impact of markedly elevated serum lipoprotein(a) levels (> or = 100 mg/dL) on the risk of coronary heart disease. *Metabolism*. 2007,56:1187-1191.

Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet.* 1997,349:1498-1504.

Naukkarinen J, Gentile M, Soro-Paavonen A, Saarela J, Koistinen HA, Pajukanta P, Taskinen MR, Peltonen L. USF1 and dyslipidemias: converging evidence for a functional intronic variant. *Hum Mol Genet*. 2005,14:2595-2605.

Naukkarinen J, Ehnholm C, Peltonen L. Genetics of familial combined hyperlipidemia. *Curr Opin Lipidol.* 2006,17:285-290. Review.

Navab M, Van Lenten BJ, Reddy ST, Fogelman AM. High-density lipoprotein and the dynamics of atherosclerotic lesions. *Circulation*. 2001,104:2386-2387.

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.* 2001,285:2486-2497.

Nejentsev S, Guja C, McCormack R, Cooper J, Howson JM, Nutland S, Rance H, Walker N, Undlien D, Ronningen KS, Tuomilehto-Wolf E, Tuomilehto J, Ionescu-Tirgoviste C, Gale EA, Bingley PJ, Gillespie KM, Savage DA, Carson DJ, Patterson CC, Maxwell AP, Todd JA. Association of intercellular adhesion molecule-1 gene with type 1 diabetes. *Lancet*. 2003,362:1723-1724.

Netter FH. Atlas of human anatomy. Ciba-Ceigy Limited. 1989, pp. 133, 205.

Neuman RJ, Rice JP. Two-locus models of disease. *Genet Epidemiol*. 1992,9:347-365.

Nevanlinna HR. The Finnish population structure. A genetic and genealogical study. *Hereditas.* 1972,71:195-236.

Ng MC, Miyake K, So WY, Poon EW, Lam VK, Li JK, Cox NJ, Bell GI, Chan JC. The linkage and association of the gene encoding upstream stimulatory factor 1 with type 2 diabetes and metabolic syndrome in the Chinese population. *Diabetologia.* 2005,48:2018-2024.

Nikkilä EA, Aro A. Family study of serum lipids and lipoproteins in coronary heart-disease. *Lancet*. 1973,1:954-959.

Norman RA, Tataranni PA, Pratley R, Thompson DB, Hanson RL, Prochazka M, Baier L, Ehm MG, Sakul H, Foroud T, Garvey WT, Burns D, Knowler WC, Bennett PH, Bogardus C, Ravussin E. Autosomal genomic scan for loci linked to obesity and energy metabolism in Pima Indians. *Am J Hum Genet*. 1998,62:659-668.

Nuotio K, Lindsberg PJ, Carpen O, Soinne L, Lehtonen-Smeds EM, Saimanen E, Lassila R, Sairanen T, Sarna S, Salonen O, Kovanen PT, Kaste M. Adhesion molecule expression in symptomatic and asymptomatic carotid stenosis. *Neurology*. 2003,60:1890-1899.

O'Connell DL, Heller RF, Roberts DC, Allen JR, Knapp JC, Steele PL, Silove D. Twin study of genetic and environmental effects on lipid levels. *Genet Epidemiol*. 1988,5:323-341.

Ohlin AK, Holm J, Hillarp A. Genetic variation in the human thrombomodulin promoter locus and prognosis after acute coronary syndrome. *Thromb Res*. 2004,113:319-326.

Olivot JM, Labreuche J, Aiach M, Amarenco P; GENIC Investigators. Soluble thrombomodulin and brain infarction: case-control and prospective study. *Stroke.* 2004,35:1946-1951.

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet*. 2002,32:650-654. Erratum in: *Nat Genet*. 2003,33:107.

Pajukanta P, Nuotio I, Terwilliger JD, Porkka KV, Ylitalo K, Pihlajamäki J, Suomalainen AJ, Syvänen AC, Lehtimäki T, Viikari JS, Laakso M, Taskinen MR, Ehnholm C, Peltonen L. Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. *Nat Genet*. 1998,18:369-373.

Pajukanta P, Lilja HE, Sinsheimer JS, Cantor RM, Lusis AJ, Gentile M, Duan XJ, Soro-Paavonen A, Naukkarinen J, Saarela J, Laakso M, Ehnholm C, Taskinen MR, Peltonen L. Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nat Genet*. 2004,36:371-376.

Pajunen P, Pääkkönen R, Juolevi A, Hämäläinen H, Keskimäki I, Laatikainen T, Moltchanov V, Niemi M, Rintanen H, Salomaa V. Trends in fatal and non-fatal coronary heart disease events in Finland during 1991-2001. *Scand Cardiovasc J*. 2004,38:340-344.

Pajunen P, Pääkkönen R, Hämäläinen H, Keskimäki I, Laatikainen T, Niemi M, Rintanen H, Salomaa V. Trends in fatal and nonfatal strokes among persons aged 35 to > or =85 years during 1991-2002 in Finland. *Stroke*. 2005,36:244-248. (a)

Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Räihä P, Kärjä-Koskenkari P, Mähönen M, Niemelä M, Kuulasmaa K, Palomäki P, Mustonen J, Lehtonen A, Arstila M, Vuorenmaa T, Lehto S, Miettinen H, Juolevi A, Torppa J, Tuomilehto J, Kesäniemi YA, Pyörälä K, Salomaa V. Five-year risk of developing clinical diabetes after first myocardial infarction; the FINAMI study. *Diabet Med.* 2005,22:1334-1337. (b)

Pan L, Ober C, Abney M. Heritability estimation of sex-specific effects on human quantitative traits. *Genet Epidemiol*. 2007,31:338-347.

Papa A, Danese S, Urgesi R, Grillo A, Guglielmo S, Roberto I, Semeraro S, Scaldaferri F, Pola R, Flex A, Fedeli G, Gasbarrini G, Pola P, Gasbarrini A. Intercellular adhesion molecule 1 gene polymorphisms in inflammatory bowel disease. *Eur Rev Med Pharmacol Sci*. 2004,8:187-191. Review.

Papassotiropoulos A, Wollmer MA, Tsolaki M, Brunner F, Molyva D, Lutjohann D, Nitsch RM, Hock C. A cluster of cholesterol-related genes confers susceptibility for Alzheimer's disease. *J Clin Psychiatry*. 2005,66:940-947.

Pastinen T, Perola M, Ignatius J, Sabatti C, Tainola P, Levander M, Syvanen AC, Peltonen L. Dissecting a population genome for targeted screening of disease mutations. *Hum Mol Genet*. 2001,10:2961-2972.

Peltonen L, Jalanko A, Varilo T. Molecular genetics of the Finnish disease heritage. *Hum Mol Genet.* 1999,8:1913-1923. Review.

Peltonen L. Positional cloning of disease genes: advantages of genetic isolates. *Hum Hered*. 2000,50:66-75.

Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. *Nat Rev Genet*. 2006,1:182-190. Review.

Pennacchio LA, Rubin EM: Apolipoprotein A5, a newly identified gene that affects plasma triglyceride levels in humans and mice. *Arterioscler Thromb Vasc Biol.* 2003, 23:529-534.

Permutt MA, Wasson J, Cox N. Genetic epidemiology of diabetes. *J Clin Invest*. 2005,115:1431-1439. Review.

Piedrahita JA, Zhang SH, Hagaman JR, Oliver PM, Maeda N. Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. *PNAS*. 1992,89:4471–4475.

Pihlajamäki J, Karjalainen L, Karhapää P, Vauhkonen I, Laakso M. Impaired free fatty acid suppression during hyperinsulinemia is a characteristic finding in familial combined hyperlipidemia, but insulin resistance is observed only in hypertriglyceridemic patients. *Arterioscler Thromb Vasc Biol*. 2000,20:164-170.

Plump AS, Smith JD, Hayek T, Aalto-Setälä K, Walsh A, Verstuyft JG, Rubin EM, Breslow JL. Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell*. 1992,71:343-353.

Podgoreanu MV, White WD, Morris RW, Mathew JP, Stafford-Smith M, Welsby IJ, Grocott HP, Milano CA, Newman MF, Schwinn DA. Inflammatory gene polymorphisms and risk of postoperative myocardial infarction after cardiac surgery. *Circulation*. 2006,4(1 Suppl):I275-281.

Pollex RL, Hegele RA. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation.* 2007,115:3130-3138. Review.

Pomp ER, Rosendaal FR, Doggen CJ. Smoking increases the risk of venous thrombosis and acts synergistically with oral contraceptive use. *Am J Hematol*. 2007,Aug 28.

Prescott E, Hippe M, Schnohr P, Hein HO, Vestbo J. Smoking and risk of myocardial infarction in women and men: longitudinal population study. *BMJ*. 1998,316:1043-1047.

Putt W, Palmen J, Nicaud V, Tregouet DA, Tahri-Daizadeh N, Flavell DM, Humphries SE, Talmud PJ; EARSII group. Variation in USF1 shows haplotype effects, gene : gene and gene : environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. *Hum Mol Genet.* 2004,13:1587-1597.

Rahkonen O, Lundberg O, Lahelma E, Huuhka M. Body mass and social class: a comparison of Finland and Sweden in the 1990s. *J Public Health Policy*. 1998,19:88-105.

Raivio P, Fernandez JA, Kuitunen A, Griffin JH, Lassila R, Petäjä J. Activation of protein C and hemodynamic recovery after coronary artery bypass surgery. *J Thorac Cardiovasc Surg*. 2007,133:44-51.

Rajecki M, Pajunen P, Jousilahti P, Rasi V, Vahtera E, Salomaa V. Hemostatic factors as predictors of stroke and cardiovascular diseases: the FINRISK '92 Hemostasis Study. *Blood Coagul Fibrinolysis*. 2005,16:119-124.

Rees DC, Cox M, Clegg JB. World distribution of factor V Leiden. *Lancet*. 1995,346:1133-1134.

Reiner AP, Carlson CS, Jenny NS, Durda JP, Siscovick DS, Nickerson DA, Tracy RP. USF1 Gene Variants, Cardiovascular Risk, and Mortality in European-Americans. Analysis of Two U.S. Cohort Studies. *Arterioscler Thromb Vasc Biol*. 2007 Sep 20

Reitsma PH, Bernardi F, Doig RG, Gandrille S, Greengard JS, Ireland H, Krawczak M, Lind B, Long GL, Poort SR, Satio H, Sala N, Witt I, Cooper D. *Thromb Haemost*. 1995,73:876-889. Review.

Reitsma PH. Is hypercoagulability an issue in arterial thrombosis? No. *J Thromb Haemost*. 2004,2:692-694.

Ridker PM, Miletich JP, Hennekens CH, Buring JE. Ethnic distribution of factor V Leiden in 4047 men and women. Implications for venous thromboembolism screening. *JAMA*. 1997,277:1305-1307.

Ridker PM. Role of inflammatory biomarkers in prediction of coronary heart disease. *Lancet*. 2001,358:946-948. Review.

Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet*. 1990,46:222-228.

Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000,405:847-856. Review.

Rosamond W, Flegal K, Friday G, Furie K, Go A, Greenlund K, Haase N, Ho M, Howard V, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell CJ, Roger V, Rumsfeld J, Sorlie P, Steinberger J, Thom T, Wasserthiel-Smoller S, Hong Y; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*. 2007,115:e69-171.

Rosendaal FR, Siscovick DS, Schwartz SM, Beverly RK, Psaty BM, Longstreth WT Jr, Raghunathan TE, Koepsell TD, Reitsma PH. Factor V Leiden (resistance to activated protein C) increases the risk of myocardial infarction in young women. *Blood.* 1997,89:2817-2821.

Ross R, Glomset JA. The pathogenesis of atherosclerosis (first of two parts). *N Engl J Med.* 1976,295:369-377. Review.

Ross R, Glomset JA. The pathogenesis of atherosclerosis (second of two parts). *N Engl J Med.* 1976,295:420-425. Review.

Ross R. Atherosclerosis – An inflammatory disease. *N Engl J Med.* 1999,340:115-125. Review.

Sacco RL, Boden-Albala B, Gan R, Chen X, Kargman DE, Shea S, Paik MC, Hauser WA. Stroke incidence among white, black, and Hispanic residents of an urban community: the Northern Manhattan Stroke Study. *Am J Epidemiol*. 1998,147:259-268.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D; International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature.* 2001,409:928-933.

Salomaa V, Matei C, Aleksic N, Sansores-garcia L, Folsom A, Juneja H, Chambless LE, Wu KK. Soluble thrombomodulin as a predictor of incident coronary heart disease and symptomless carotid artery atherosclerosis in the Atherosclerosis Risk in Communities (ARIC) Study. *Lancet.* 1999,353:1729–1734.

Salomaa V, Miettinen H, Niemelä M, Ketonen M, Mähönen M, Immonen-Raihä P, Lehto S, Vuorenmaa T, Koskinen S, Palomäki P, Mustaniemi H, Kaarsalo E, Arstila M, Torppa J, Kuulasmaa K, Puska P, Pyorälä K, Tuomilehto J. Relation of socioeconomic position to the case fatality, prognosis and treatment of myocardial infarction events; the FINMONICA MI Register Study. *J Epidemiol Community Health.* 2001,55:475-482.

Salomaa V, Ketonen M, Koukkunen H, Immonen-Raihä P, Jerkkola T, Kärjä-Koskenkari P, Mähönen M, Niemelä M, Kuulasmaa K, Palomäki P, Arstila M, Vuorenmaa T, Lehtonen A, Lehto S, Miettinen H, Torppa J, Tuomilehto J, Kesäniemi YA, Pyörälä K. Trends in coronary events in Finland during 1983-1997. The FINAMI study. *Eur Heart J*. 2003 Feb;24(4):311-9.

Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H; the WTCCC and the Cardiogenics Consortium. Genomewide Association Analysis of Coronary Artery Disease. *N Engl J Med*. 2007 Jul 18;

Sambola A, Osende J, Hathcock J, Degen M, Nemerson Y, Fuster V, Crandall J, Badimon JJ. Role of risk factors in the modulation of tissue factor activity and blood thrombogenicity. *Circulation.* 2003,107:973-977.

Sandercock PA, Warlow CP, Jones LN, Starkey IR. Predisposing factors for cerebral infarction: the Oxfordshire community stroke project. *BMJ.* 1989,298:75-80.

Scaldaferri F, Sans M, Vetrano S, Graziani C, De Cristofaro R, Gerlitz B, Repici A, Arena V, Malesci A, Panes J, Grinnell BW, Danese S. Crucial role of the protein C pathway in governing microvascular inflammation in inflammatory bowel disease. *J Clin Invest*. 2007,117:1951-1960.

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 2003,422:297-302.

Schaefer EJ, Genest JJ Jr, Ordovas JM, Salem DN, Wilson PW. Familial lipoprotein disorders and premature coronary artery disease. *Atherosclerosis*. 1994,108 Suppl:S41-54. Review.

Schreiner PJ, Niemelä M, Miettinen H, Mähönen M, Ketonen M, Immonen-Räihä P, Lehto S, Vuorenmaa T, Palomäki P, Mustaniemi H, Kaarsalo E, Arstila M, Torppa J, Puska P, Tuomilehto J, Pyörälä K, Salomaa V. Gender differences in recurrent coronary events; the FINMONICA MI register. *Eur Heart J.* 2001,22:762-768.

SeattleSNPs. NHLBI Program for Genomic Applications, SeattleSNPs, Seattle, Wash. 2005, Oct. (http://pga.gs.washington.edu)

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin MR, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet*. 2006,38: 556–560.

Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Au R, Kannel WB, Wolf PA. The lifetime risk of stroke: estimates from the Framingham Study. *Stroke.* 2006,37:345-350.

Shen J, Arnett DK, Peacock JM, Parnell LD, Kraja A, Hixson JE, Tsai MY, Lai CQ, Kabagambe EK, Straka RJ, Ordovas JM. Interleukin1{beta} Genetic Polymorphisms Interact with Polyunsaturated Fatty Acids to Modulate Risk of the Metabolic Syndrome. *J Nutr*. 2007,137:1846-1851.

Shiffman D, Rowland CM, Sninsky JJ, Devlin JJ. Polymorphisms associated with coronary heart disease: better by the score. *Curr Opin Mol Ther*. 2006 Dec;8(6):493-9. Review.

Shimo-Nakanishi Y, Urabe T, Hattori N, Watanabe Y, Nagao T, Yokochi M, Hamamoto M, Mizuno Y. Polymorphism of the lipoprotein lipase gene and risk of atherothrombotic cerebral infarction in the Japanese. *Stroke*. 2001,32:1481-1486.

Silander K, Mohlke KL, Scott LJ, Peck EC, Hollstein P, Skol AD, Jackson AU, Deloukas P, Hunt S, Stavrides G, Chines PS, Erdos MR, Narisu N, Conneely KN, Li C, Fingerlin TE, Dhanjal SK, Valle TT, Bergman RN, Tuomilehto J, Watanabe RM, Boehnke M, Collins FS. Genetic variation near the

hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. *Diabetes.* 2004,53:1141-1149.

Silander K, Komulainen K, Ellonen P, Jussila M, Alanne M, Levander M, Tainola P, Kuulasmaa K, Salomaa V, Perola M, Peltonen L, Saarela J. Evaluating whole genome amplification via multiply-primed rolling circle amplification for SNP genotyping of samples with low DNA yield. *Twin Res Hum Genet*. 2005,8:368-375.

Simmonds RE, Hermida J, Rezende SM, Lane DA. Haemostatic genetic risk factors in arterial thrombosis. *Thromb Haemost*. 2001,86:374-385. Review.

Simundic AM, Basic V, Topic E, Demarin V, Vrkic N, Kunovic B, Stefanovic M, Begonja A. Soluble adhesion molecules in acute ischemic stroke. *Clin Invest Med*. 2004 Apr;27(2):86-92.

Sivenius J, Tuomilehto J, Immonen-Raihä P, Kaarisalo M, Sarti C, Torppa J, Kuulasmaa K, Mähönen M, Lehtonen A, Salomaa V; FINSTROKE study. Continuous 15-year decrease in incidence and mortality of stroke in Finland: the FINSTROKE study. *Stroke*. 2004,35:420-425.

Slatkin M. Linkage disequilibrium in growing and stable populations. *Genetics*. 1994,137:331-336.

Smith A, Patterson C, Yarnell J, Rumley A, Ben-Shlomo Y, Lowe G. Which hemostatic markers add to the predictive value of conventional risk factors for coronary heart disease and ischemic stroke? The Caerphilly Study. *Circulation.* 2005; 112: 3080–3087.

Smith JD, Trogan E, Ginsberg M, Grigaux C, Tian J, Miyata M. Decreased atherosclerosis in mice deficient in both macrophage colony-stimulating factor (op) and apolipoprotein E. *PNAS*. 1995, 92:8264-8268.

Smith SC. Multiple risk factors for cardiovascular disease and diabetes mellitus. *Am J Med*. 2007,120(3 Suppl 1):S3-S11. Review.

Sniderman A, Shapiro S, Marpole D, Skinner B, Teng B, Kwiterovich PO Jr. Association of coronary atherosclerosis with hyperapobetalipoproteinemia [increased protein but normal cholesterol levels in human plasma low density (beta) lipoproteins]. *PNAS*. 1980,77:604-608.

Soinio M, Marniemi J, Laakso M, Lehto S, Rönnemaa T. Elevated plasma homocysteine level is an independent predictor of coronary heart disease events in patients with type 2 diabetes mellitus. *Ann Intern Med*. 2004,140:94-100.

Soinne L, Saimanen E, Malmberg-Ceder K, Kovanen P, Lindsberg PJ, Kaste M, Lassila R. Association of the fibrinolytic system and hemorheology with symptoms in patients with carotid occlusive disease. *Cerebrovasc Dis*. 2005,20:172-179.

Soll G, Bendszus M. Inflammation and atherosclerosis. Novel insights into plaque formation and destabilization. *Stroke*. 2006,37:1923-1932. Review.

Spek CA, Koster T, Rosendaal FR, Bertina RM, Reitsma PH. Genotypic variation in the promoter region of the protein C gene is associated with plasma protein C levels and thrombotic risk. *Arterioscler Thromb Vasc Biol*. 1995,15:214-218.

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

Stamler J, Daviglus ML, Garside DB, Dyer AR, Greenland P, Neaton JD. Relationship of baseline serum cholesterol levels in 3 large cohorts of younger men to long-term coronary, cardiovascular, and all-cause mortality and to longevity. *JAMA.* 2000,284: 311–318.

Stary HC, Blankenhorn DH, Chandler AB, Glagov S, Insull W Jr, Richardson M, Rosenfeld ME, Schaffer SA, Schwartz CJ, Wagner WD, et al. A definition of the intima of human arteries and of its

atherosclerosis-prone regions. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Circulation*. 1992,85:391-405. Review.

Stary HC, Chandler AB, Glagov S, Guyton JR, Insull W Jr, Rosenfeld ME, Schaffer SA, Schwartz CJ, Wagner WD, Wissler RW. A definition of initial, fatty streak, and intermediate lesions of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Circulation*. 1994,89:2462-2478. Review.

Stary HC, Chandler AB, Dinsmore RE, Fuster V, Glagov S, Insull W Jr, Rosenfeld ME, Schwartz CJ, Wagner WD, Wissler RW. A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Circulation*. 1995,92:1355-1374. Review.

Stengard JH, Kardia SL, Hamon SC, Frikke-Schmidt R, Tybjaerg-Hansen A, Salomaa V, Boerwinkle E, Sing CF: Contribution of regulatory and structural variations in APOE to predicting dyslipidemia. *J Lipid Res.* 2006,47:318-328.

Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet. 2003,73:1162-1169.

Stone PH, Coskun AU, Yeghiazarians Y, Kinlay S, Popma JJ, Kuntz RE, Feldman CL. Prediction of sites of coronary atherosclerosis progression: In vivo profiling of endothelial shear stress, lumen, and outer vessel wall characteristics to predict vascular behavior. *Curr Opin Cardiol*. 2003,18:458-470. Review.

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007,315:848-853.

Stride A, Hattersley AT. Different genes, different diabetes: lessons from maturity-onset diabetes of the young. *Ann Med*. 2002,34:207-316. Review.

Sundström J, Riserus U, Byberg L, Zethelius B, Lithell H, Lind L: Clinical value of the metabolic syndrome for long term prediction of total and cardiovascular mortality: prospective, population based cohort study. *BMJ.* 2006,332:878-882.

Suviolahti E, Lilja HE, Pajukanta P. Unraveling the complex genetics of familial combined hyperlipidemia. *Ann Med*. 2006,38:337-351. Review.

Tanne D, Haim M, Boyko V, Goldbourt U, Reshef T, Matetzky S, Adler Y, Mekori YA, Behar S. Soluble intercellular adhesion molecule-1 and risk of future ischemic stroke: a nested case-control study from the Bezafibrate Infarction Prevention (BIP) study cohort. *Stroke*. 2002,33:2182-2186.

Terwilliger JD, Hiekkalinna T. An utter refutation of the "Fundamental Theorem of the HapMap". *Eur J Hum Genet*. 2006,14:426-437.

Tiong IY, Alkotob ML, Ghaffari S. Protein C deficiency manifesting as an acute myocardial infarction and ischaemic stroke. *Heart.* 2003,89: E7.

Tong PC, Kong AP, So WY, Yang X, Ho CS, Ma RC, Ozaki R, Chow CC, Lam CW, Chan JC, Cockram CS. The usefulness of the International Diabetes Federation and the National Cholesterol Education Program's Adult Treatment Panel III definitions of the metabolic syndrome in predicting coronary heart disease in subjects with type 2 diabetes. *Diabetes Care*. 2007,30:1206-1211.

Touyz RM, Schiffrin EL. Blunted inhibition by insulin of agonist-stimulated calcium, pH and aggregatory responses in platelets from hypertensive patients. *J Hypertens.* 1994,12:1255–1263.

Touze E, Rothwell PM. Heritability of ischaemic stroke in women compared with men: a genetic epidemiological study. *Lancet Neurol*. 2007,6:125-133.

Trovati M, Mularoni E, Burzacca S, Ponziani MC, Massucco P, Mattiello L, Piretto V, Cavalot F, Anfossi G. Impaired insulin-induced platelet anti-aggregating effect in obesity and in obese non-insulin-dependent diabetes mellitus. *Diabetes*. 1995,44:1318–1322.

Tuomilehto J, Rastenyte D, Sivenius J, Sarti C, Immonen-Raihä P, Kaarsalo E, Kuulasmaa K, Narva EV, Salomaa V, Salmi K, Torppa J. Ten-year trends in stroke incidence and mortality in the FINMONICA Stroke Study. *Stroke.* 1996,27:825-832.

Van der Vleuten GM, Isaacs A, Hijmans A, van Duijn CM, Stalenhoef AF, de Graaf J. The involvement of upstream stimulatory factor 1 in Dutch patients with familial combined hyperlipidemia. *J Lipid Res*. 2007,48:193-200.

Van Hinsberg VWM, Bertina RM, van Wijngaarden A, van Tilburg NH, Emeis JJ, Haverkate F. Activated protein C decreases plasminogen activator inhibitor activity in endothelial cell conditioned media. *Blood.* 1985,65:1914-1920.

Vandenplas S, Wiid I, Grobler-Rabie A, Brebner K, Ricketts M, Wallis G, Bester A, Boyd C, Mathew C. Blot hybridisation analysis of genomic DNA. *J Med Genet*. 1984,21:164-172. Review.

Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD, 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.

Vartiainen E, Jousilahti P, Tamminen M, Korhonen HJ, Tuomilehto J, Sundvall J, Jauhiainen M, Puska P. FINRISKI '92: Tutkimus kansanterveydellisistä riskitekijöistä, niihin liittyvistä elintavoista, oireista ja terveyspalvelujen käytöstä. Tutkimuksen toteutus ja perustaulukot. Kansanterveyslaitoksen julkaisuja B9/1993.

Vartiainen E, Jousilahti P, Juolevi A, Sundvall J, Alfthan G, Salminen I, Puska P. FINRISKI 1997. Tutkimus kroonisten kansantautien riskitekijöistä, niihin liittyvistä elintavoista, oireista ja terveyspalvelujen käytöstä. Tutkimuksen toteutus ja perustaulukot. Kansanterveyslaitoksen julkaisuja B1/1998.

Vasto S, Candore G, Balistreri CR, Caruso M, Colonna-Romano G, Grimaldi MP, Listi F, Nuzzo D, Lio D, Caruso C. Inflammatory networks in ageing, age-related diseases and longevity. *Mech Ageing Dev*. 2007,128:83-91. Review.

Vinceti M, Pellacani G, Casali B, Malagoli C, Nicoli D, Farnetti E, Bassissi S, Bergomi M, Seidenari S. High risk of cutaneous melanoma amongst carriers of the intercellular adhesion molecule-1 R241 allele. *Melanoma Res*. 2006,16:93-96.

Virtanen JK, Voutilainen S, Happonen P, Alfthan G, Kaikkonen J, Mursu J, Rissanen TH, Kaplan GA, Korhonen MJ, Sivenius J, Salonen JT. Serum homocysteine, folate and risk of stroke: Kuopio Ischaemic Heart Disease Risk Factor (KIHD) Study. *Eur J Cardiovasc Prev Rehabil*. 2005,12:369-375.

Voetsch B, Loscalzo J. Genetic determinants of arterial thrombosis. *Arterioscler Thromb Vasc Biol*. 2004,24:216-229. Review.

Volcik KA, Ballantyne CM, Coresh J, Folsom AR, Wu KK, Boerwinkle E. P-selectin Thr715Pro polymorphism predicts P-selectin levels but not risk of incident coronary heart disease or ischemic stroke in a cohort of 14595 participants: the Atherosclerosis Risk in Communities Study. *Atherosclerosis*. 2006,186:74-79.

Volcik KA, Ballantyne CM, Coresh J, Folsom AR, Boerwinkle E. Specific P-selectin and P-selectin glycoprotein ligand-1 genotypes/haplotypes are associated with risk of incident CHD and ischemic stroke: the Atherosclerosis Risk in Communities (ARIC) study. *Atherosclerosis*. 2007,195:e76-82.

Voorberg J, Roelse J, Koopman R, Buller H, Berends F, ten Cate JW, Merents K, van Mourik JA. Association of idiopathic venous thromboembolism with single point-mutation at Arg506 of factor V. *Lancet.* 1994,343:1535-1536.

Voutilainen S, Lakka TA, Hamelahti P, Lehtimäki T, Poulsen HE, Salonen JT. Plasma total homocysteine concentration and the risk of acute coronary events: the Kuopio Ischaemic Heart Disease Risk Factor Study. *J Intern Med*. 2000,248:217-222.

Voutilainen S, Tuomainen TP, Korhonen M, Mursu J, Virtanen JK, Happonen P, Alfthan G, Erlund I, North KE, Mosher MJ, Kauhanen J, Tiihonen J, Kaplan GA, Salonen JT. Functional COMT Val158Met polymorphism, risk of acute coronary events and serum homocysteine: the kuopio ischaemic heart disease risk factor study. *PLoS ONE*. 2007,2:e181.

Walldius G, Aastveit AH, Jungner I. Stroke mortality and the apoB/apoA-I ratio: results of the AMORIS prospective study. *J Intern Med*. 2006,259:259-266.

Wang Q. Molecular genetics of coronary artery disease. *Curr Opin Cardiol*. 2005,20:182-188. Review.

Wang JY, Zhou DH, Li J, Zhang M, Deng J, Gao C, Li J, Lian Y, Chen M. Association of soluble intercellular adhesion molecule 1 with neurological deterioration of ischemic stroke: The Chongqing Stroke Study. *Cerebrovasc Dis*. 2006,21:67-73.

Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet*. 2005,6:109-118. Review.

Wannamethee SG, Perry IJ, Shaper AG. Nonfasting serum glucose and insulin concentrations and the risk of stroke. *Stroke*. 1999,30:1780-1786.

Warner D, Catto A, Kunz G, Ireland H, Grant PJ, Lane DA. The thrombomodulin gene mutation G(127)-->A (Ala25Thr) and cerebrovascular disease. *Cerebrovasc Dis*. 2000,10:359-363.

Wartiovaara U, Perola M, Mikkola H, et al. Association of FXIII Val34Leu with decreased risk of myocardial infarction in Finnish males. *Atherosclerosis*. 1999,142:295.

WHO: Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications. Report of a WHO Consultation. *WHO/NCD/NCS/99.2*, 1999

Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, Rayner NW, Shields B, Owen KR, Hattersley AT, Frayling TM. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med*. 2006,3:e374.

Wei Q, Doris PA, Pollizotto MV, Boerwinkle E, Jacobs DR Jr, Siscovick DS, Fornage M. Sequence variation in the soluble epoxide hydrolase gene and subclinical coronary atherosclerosis: interaction with cigarette smoking. *Atherosclerosis*. 2007,190:26-34.

Weiss LA, Pan L, Abney M, Ober C. The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet*. 2006,38:218-222.

Weissglas-Volkov D, Huertas-Vazquez A, Suviolahti E, Lee J, Plaisier C, Canizales-Quinteros S, Tusie-Luna T, Aguilar-Salinas C, Taskinen MR, Pajukanta P: Common hepatic nuclear factor-4alpha variants are associated with high serum lipid levels and the metabolic syndrome. *Diabetes.* 2006,55:1970-1977.

Welch CL, Bretschger S, Wen PZ, Mehrabian M, Latib N, Fruchart-Najib J, Fruchart JC, Myrick C, Lusis AJ: Novel QTLs for HDL levels identified in mice by controlling for Apoa2 allelic effects: confirmation of a chromosome 6 locus in a congenic strain. *Physiol Genomics.* 2004,17:48-59.

Westerbacka J, Yki-Järvinen H, Turpeinen A, Rissanen A, Vehkavaara S, Syrjälä M, Lassila R. Inhibition of platelet-collagen interaction: an in vivo action of insulin abolished by insulin resistance in obesity. *Arterioscler Thromb Vasc Biol.* 2002,22:167-172.

Williamson D, Brown K, Luddington R, Baglin C, Baglin T. Factor V Cambridge: a new mutation (Arg306-->Thr) associated with resistance to activated protein C. *Blood.* 1998,91:11480-1144.

Winckler W, Burtt NP, Holmkvist J, Cervin C, de Bakker PI, Sun M, Almgren P, Tuomi T, Gaudet D, Hudson TJ, Ardlie KG, Daly MJ, Hirschhorn JN, Altshuler D, Groop L. Association of common variation in the HNF1alpha gene region with risk of type 2 diabetes. *Diabetes*. 2005,54:2336-2342.

Wolf PA, D'Agostino RB, Kannel WB, Bonita R, Belanger AJ. Cigarette smoking as a risk factor for stroke. The Framingham Study. *JAMA*. 1988,259:1025-1029.

Worrall BB, Mychaleckyj JC. PDE4D and stroke: a real advance or a case of the Emperor's new clothes? *Stroke*. 2006,37:1955-1957. Review.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*. 2003,20:1377-1419. Review.

Wu H, Romieu I, Sienra-Monge JJ, del Rio-Navarro BE, Anderson DM, Dunn EW, Steiner LL, Lara-Sanchez Idel C, London SJ. Parental smoking modifies the relation between genetic variation in tumor necrosis factor-alpha (TNF) and childhood asthma. *Environ Health Perspect*. 2007,115:616-622.

Wu KK, Aleksic N, Ahn C, Boerwinkle E, Folsom AR, Juneja H. Thrombomodulin Ala455Val polymorphism and risk of coronary heart disease. *Circulation.* 2001; 103: 1386–1389.

Wu KK, Aleksic N, Ballantyne CM, Ahn C, Juneja H, Boerwinkle E. Interaction between soluble thrombomodulin and intercellular adhesion molecule-1 in predicting risk of coronary heart disease. *Circulation.* 2003,107:1729–1732.

Wu X, Gu J, Grossman HB, Amos CI, Etzel C, Huang M, Zhang Q, Millikan RE, Lerner S, Dinney CP, Spitz MR. Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. *Am J Hum Genet*. 2006,78:464-479.

Ye Z, Liu EH, Higgins JP, Keavney BD, Lowe GD, Collins R, Danesh J. Seven haemostatic gene polymorphisms in coronary disease: Meta-analysis of 66,155 cases and 91,307 controls. *Lancet*. 2006,367:651–658.

Ylä-Herttuala S, Sumuvuori H, Karkola K, Möttönen M, Nikkari T. Atherosclerosis and biochemical composition of coronary arteries in Finnish men. Comparison of two populations with different incidences of coronary heart disease. *Atherosclerosis.* 1987,65:109-115.

Zee RY, Cook NR, Cheng S, Reynolds R, Erlich HA, Lindpaintner K, Ridker PM. Polymorphism in the P-selectin and interleukin-4 genes as determinants of stroke: a population-based, prospective genetic analysis. *Hum Mol Genet*. 2004,13:389-396.

Zeggini E, Damcott CM, Hanson RL, Karim MA, Rayner NW, Groves CJ, Baier LJ, Hale TC, Hattersley AT, Hitman GA, Hunt SE, Knowler WC, Mitchell BD, Ng MC, O'Connell JR, Pollin TI, Vaxillaire M, Walker M, Wang X, Whittaker P, Kunsun X, Jia W, Chan JC, Froguel P, Deloukas P, Shuldiner AR, Elbein SC, McCarthy MI. Variation within the gene encoding the upstream stimulatory factor 1 does

not influence susceptibility to type 2 diabetes in samples from populations with replicated evidence of linkage to chromosome 1q. *Diabetes.* 2006,55:2541-2548.

Zhang K, Zhang S, Zheng K, Hou Y, Liao L, He Y, Zhang L, Nebert DW, Shi J, Su Z, et al.: Novel P143L polymorphism of the LCAT gene is associated with dyslipidemia in Chinese patients who have coronary atherosclerotic heart disease. *Biochem Biophys Res Commun.* 2004,318:4-10.

Zhang X, Hu Y, Hong M, Guo T, Wei W, Song S. Plasma thrombomodulin, fibrinogen, and activity of tissue factor as risk factors for acute cerebral infarction. *Am J Clin Pathol*. 2007,128:287-292.

Zhou X, Stemme S, Hansson GK. Evidence for a local immune response in atherosclerosis: CD4+ T cells infiltrate lesions of apolipoprotein-E-deficient mice. *Am J Pathol*. 1996,149:359-366.

Zhou X, Nicoletti A, Elhage R, Hansson GK. Transfer of CD4+ T cells aggravates atherosclerosis in immunodeficient apolipoprotein E knockout mice. *Circulation*. 2000,102:2919-2922.

Zöller B, Dahlbäck B. Linkage between inherited resistance to activated protein C and factor V gene mutation in venous thrombosis. *Lancet*. 1994,343:1536-1538.