Genomics in Atherosclerosis | Towards a clinical application

Article · April 2010

Source: OAI

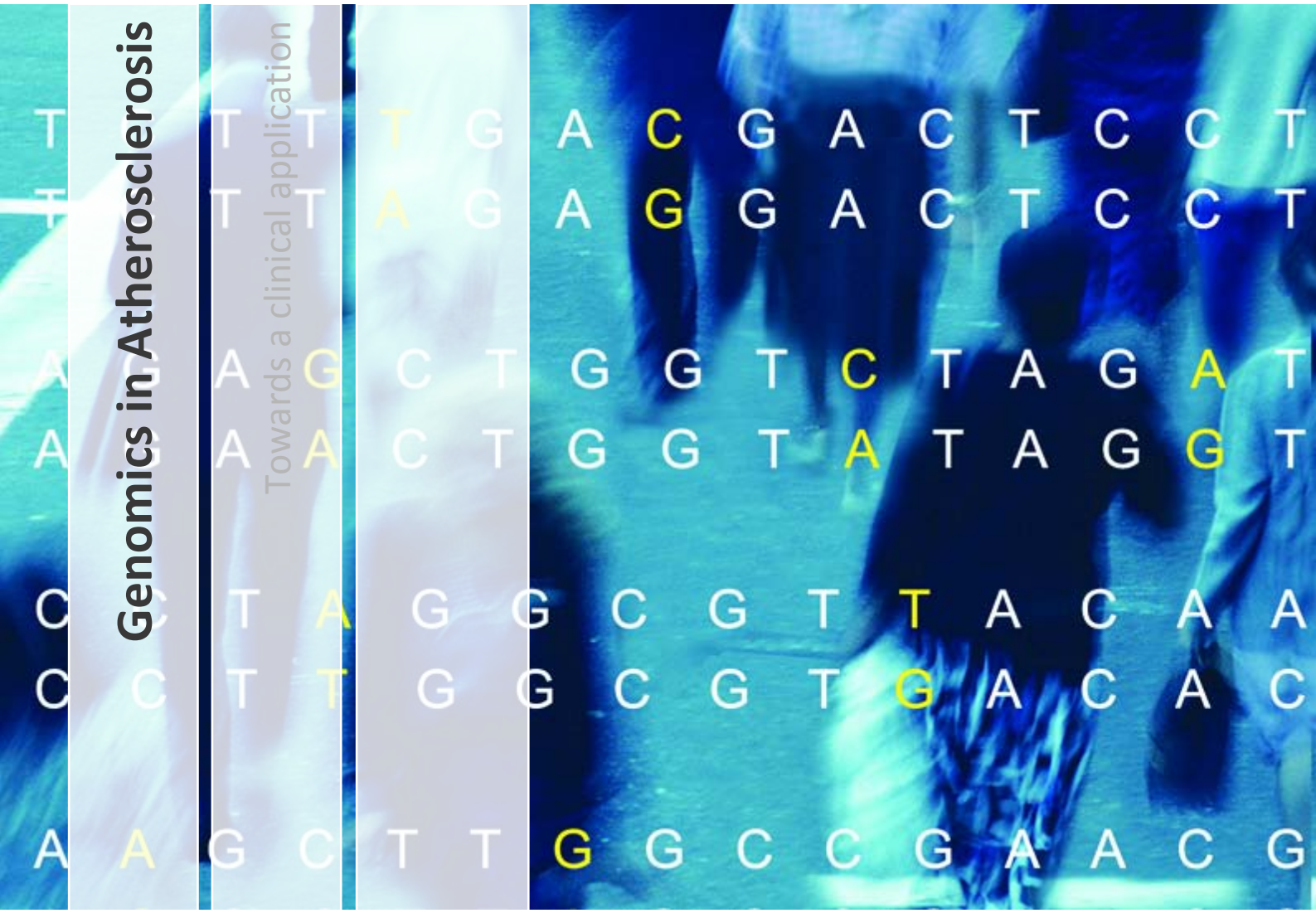
CITATIONS
0

READS
313

1 author:

Sander W. Van der Laan
University Medical Center Utrecht
119 PUBLICATIONS **6,626** CITATIONS

SEE PROFILE



Cardiovascular diseases (CVD) continue to be the major cause of morbidity and mortality in the world. Genome–wide association studies (GWAS) have identified a multitude of commonly occurring genetic loci associated with CVD. Since long atherosclerosis is recognized as the principal cause underlying CVD, and GWAS have led to the discovery of many novel genes and pathways implicated in the aetiology of atherosclerosis (atherogenesis). Traditionally (protein) biomarkers are used to assess the primary risk for adverse clinical events and it is thought genomic discoveries can improve the power to predict. Recent findings in GWAS warrant the assessment of CVD as the phenotypic manifestation of the quantitative trait "atherosclerosis". In addition, genomic investigations have illuminated new possibilities in terms of analysis strategies such as Mendelian randomization and quantitative trait analysis. Assessing atherosclerosis as a quantitative trait could help in a better pinpointing of which genes and pathways are involved in atherogenesis, and not so much its clinical outcomes, *e.g.* stroke or myocardial infarction (MI). Using a polygenic risk score to stratify patients grouped based on an existing or novel risk score algorithm could be useful in secondary prevention. A systems biology approach, combining genomic and proteomic data, could boost biomarker discovery and be a potential short-cut to validation of putative biomarkers.

Utrecht, 27 April 2010

Imprint

Master Thesis

Sander W. van der Laan
Student number 9849009
Master Biology of Disease, Biomedical Sciences
University Medical Centre Utrecht
Experimental Cardiology Laboratory
Heidelberglaan 100
P.O. Box 85500
3508 GA Utrecht
The Netherlands

- (+31) (0)6 429 83 769
- s.w.vanderlaan-2@umcutrecht.nl

Supervisors

Dr. B.P.C. Koeleman
Complex Genetics & Genomics, Medical Genetics
University Medical Centre Utrecht
Universiteitsweg 100
3584 CG Utrecht
The Netherlands

- (+31) (0)88 75 68 427
- b.p.c.koeleman@umcutrecht.nl

Prof. Dr. G. Pasterkamp
Experimental Cardiology Laboratory
University Medical Centre Utrecht
Universiteitsweg 100
3584 CG Utrecht
The Netherlands

- (+31) (0)88 75 57 155
- g.pasterkamp@umcutrecht.nl

Contents

Allele are DNA sequence variations at a physical genomic location (locus). Each gene consist of one paternal and one maternal allele.

Atherosclerosis is an inflammatory disease characterized by endothelial dysfunction. It is a lifelong progressive process of arterial wall changes that can cumulate in cardiovascular events.

In medicine a **biomarker** is a biological substance indicative of a pathophysiological state. It can be used to establish the *risk* for disease (predictive, risk factor), *presence* of disease (diagnostic), the *progression* of disease (prognostic), or as an *intermediate endpoint* as a response to drug treatment (surrogate)[5].

Atherosclerosis is the principal cause of the majority of **cardiovascular diseases (CVD) and traits**. These include: myocardial infarction (MI), cerebral vascular accident (CVA or stroke), transient ischaemic attack (TIA), peripheral artery disease (PAD), claudication, coronary artery disease (CAD), coronary artery calcification (CAC), angina pectoris (AP).

The **common allele** is the most frequently occurring allele at a particular locus in the general population.

Epigenetics refer to changes in the phenotype and expression of genes caused by factors other than changes in the DNA sequence. DNA methylation and acetylation are forms of epigenetic modification of the genome that can regulate gene expression and can be transmitted through generations.

In **epistasis** the effects of one gene are modified by one or more other genes (modified genes).

Linkage disequilibrium (LD) is the non–random association of alleles on two or more loci. These alleles are not necessarily on the same chromosome. Variants in LD are inherited together and not easily separated through recombination[5]. By definition LD is influenced by factors such as non–random mating, the rate of recombination and mutation, genetic linkage, selection, genetic drift, and population structure.

Mendelian traits and diseases are monogenic, rare traits or diseases that follow Mendelian patterns of inheritance.

Introduction

From recent projections by the World Health Organization it is evident, that in terms of national expenditures, medical services, and informal care *cardiovascular disease* continues to be a major burden to society even in regions of low socio–economic class thereby surpassing communicable diseases[10,11]. For decades the assessment of individual risk has been based on risk factors commonly associated with cardiovascular diseases (BOX 1)[12,13]. All exert a certain degree of risk, by contributing to the development and progression of *atherosclerosis*, the single–most important underlying cause of CVD. Regardless of considerable progress in *secondary prevention* and treatment programs, the ability to personalise risk assessment and *primary prevention* has been a major challenge. Recent research indicates that current guidelines such as the Framingham Heart Score (FRS) underestimate the risk in young adults[15]. As a consequence the need for better prognostic, diagnostic, and surrogate *biomarkers* to identify at–risk individuals, stratify patient groups, and devise novel personalised treatment regimes based on the discovery of novel signalling pathways and potential therapeutic targets, is evident and envisioned by many. Serum– and tissue–derived protein biomarkers are disease–related and potentially very strong predictors of adverse cardiovascular events, and consequently could meet the need for biomarkers with greater discriminative power and novel therapeutic targets. However, the extensive inter– and intra–individual differential gene and protein expression patterns at baseline and during disease pose a serious threat to this potential[16]. The proteome is the phenotypic manifestation of the genome which is inherited and inherently unchanging during life. Therefore, many foresee that studying the genomes of the sick and the healthy could provide an opportunity to finding new biomarkers.

BOX 1 | RISK FACTORS FOR CVD

Fixed
- Age
- Gender
- Family history

Modifiable
- Smoking
- BMI
- Hypertension
- Dyslipidaemia
- Diabetes

Indeed, the differences between humans are genetically defined by common and rare variations in the genome. Arguably a better understanding of atherosclerotic genetic variation provides new insights into the aetiology of CVD and offers new avenues of biomarker and therapeutics development. In a relative short time span since the advent of genome–wide association studies (GWAS) 161 papers have reported 480 common variants, termed single–nucleotide polymorphisms (SNP) with an associative p–value < 1.0×10^{-5} for many complex diseases, such as CVD[17,18]. These studies successfully identified variants conveying high risk in terms of populations which could be useful in primary prevention. Recent GWAS undeniably support the notion that atherosclerotic disease is polygenic by nurture and nature, and is heavily influenced by co–morbidity. They reveal

an unambiguous relation between genetic variations, biomarkers and risk factors, and disease (**Figure 1**) which has practical implications.

First, GWAS assess complex diseases as dichotomous (qualitative) traits measuring the difference of individual *risk allele frequencies* among diagnosed cases and controls (case–control study), comparing the tails of the distributions. Given the recent findings of GWAS it is logical to define CVD and atherosclerosis as *quantitative traits*, *i.e.* phenotypic manifestations of normally distributed aggregates of measurable (continuous) variables [19]. Second, the polygenic nature of complex diseases is complicating genetic applications in primary prevention. Some variants are used as biomarkers and indicate a higher risk to develop a disease. The famous variant on chromosome 9p21 (rs133349) increases the risk for MI depending on the *allele* 1.4-fold. In terms of absolute risk the effect of individual variants is small, and traditional risk factors remain more useful in risk prediction. However, the aggregate of genetic variations can be indexed to associate with the biomarkers, risk factors, and proteins known to be involved in the disease, generating a *polygenic risk score* based on quantifiable traits. Third, it is very interesting to focus on secondary clinical outcome in atherosclerotic patients. In aging atherosclerotic patients, traditional risk factors contribute little to the prediction of disease progression or secondary events[20-22]. Compared to individual markers a polygenic risk score could add to the prediction, and be useful in secondary prevention to identify at–risk patients and stratify patient groups. Fourth, the assessment of the predictive value of novel biomarkers is complicated by inter– and intra–individual expression differences. Additionally, genetic variants, proteins, risk factors and biomarkers can be associated with, but are not necessarily causative of the disease[7]. *Mendelian randomization* can be used to investigate the causal relation of associating factors (**Figure 2**). This can be helpful in confirming the validity of the tentative biomarker and in identifying novel therapeutic targets. Furthermore, it effectively indicates a possible patient stratification strategy. An integrative approach would consider atherosclerotic disease as a quantitative trait classifying individuals according to their genotype rather than their diagnosis (phenotype), identifying *quantitative trait loci*, investigating individual risk, facilitating patient stratification, and tapping into the full potential of biomarker discovery and therapeutic target development.

The objective of this thesis is to examine the possibility to apply atherosclerotic genomics for secondary prevention in a diseased population cohort, the Athero–Express Study. The quantitative nature of atherosclerosis could be defined by the stage of atherogenesis, *e.g.* defined by a quantifiable clinical outcome, or assessed by Doppler–technique (stenosis and calcification), immunohistochemical staining of plaque material obtained through vascular surgery, or biomarker expression in

Mendelian randomization studies take advantage of the fact that individuals are randomly assigned at birth to a specific expression level of proteins depending on the alleles they receive from their parents[7]. This randomization is based on Mendel's laws of independent assortment of alleles and in principle undisturbed by any factor, endogenous or exogenous. These analyses determine the effect size of variants on a QT, and the effect size of the QT on the disease. The combination of these results provides a predicted effect size of the variant on the disease. If this agrees with the observed effect, a causal relation can be inferred.

Minor allele is the less frequently occurring allele at a particular locus in the general population. By definition the **minor allele frequency (MAF)** cannot be higher than 50%.

A **misclassification bias** results from the failure to correctly assign individuals to a group in case–control studies[8].

When a single gene influences multiple (different) phenotypic traits, it is said to be **pleiotropic**[14].

Polygenic risk score refers to the set of genetic variants that are associated with and confers risk to develop a disease. Polygenic risk scores have been used to determine individual lifetime risk in a population–wide manor.

The objective of **primary prevention** is to avoid the development of a disease.

A **quantitative trait (QT)** is a phenotypic characteristic that is measurable and attributed by the (inter)actions of multiple genes and environmental factors.

Quantitative trait loci (QTL) are not necessarily genes, but are closely linked to the genes underlying quantitative traits.

The risk allele is the allele conferring risk for a particular trait or disease. It frequently is the minor allele, although certain risk allele can have a higher frequency than 50%.

The **risk allele frequency (RAF)** is the frequency of the risk allele in a particular population. It is analogous to the MAF.

The aim of **secondary prevention** is the early detection of disease, to be able to take appropriate action to prevent disease progression and the onset of symptoms.

circulating cells or plaques. In order to understand the scope of the field a brief description of atherosclerosis, known factors involved, and biomarkers is given. Additionally, an overview of the basics in genomics research is given with atherosclerotic disease to exemplify.

This would be an assessment of which combination of genetic variation, additional knowledge, and genomic methods would be "best" to pursuit, when the goal is to develop a clinically useful application for an already diseased population (secondary prevention).

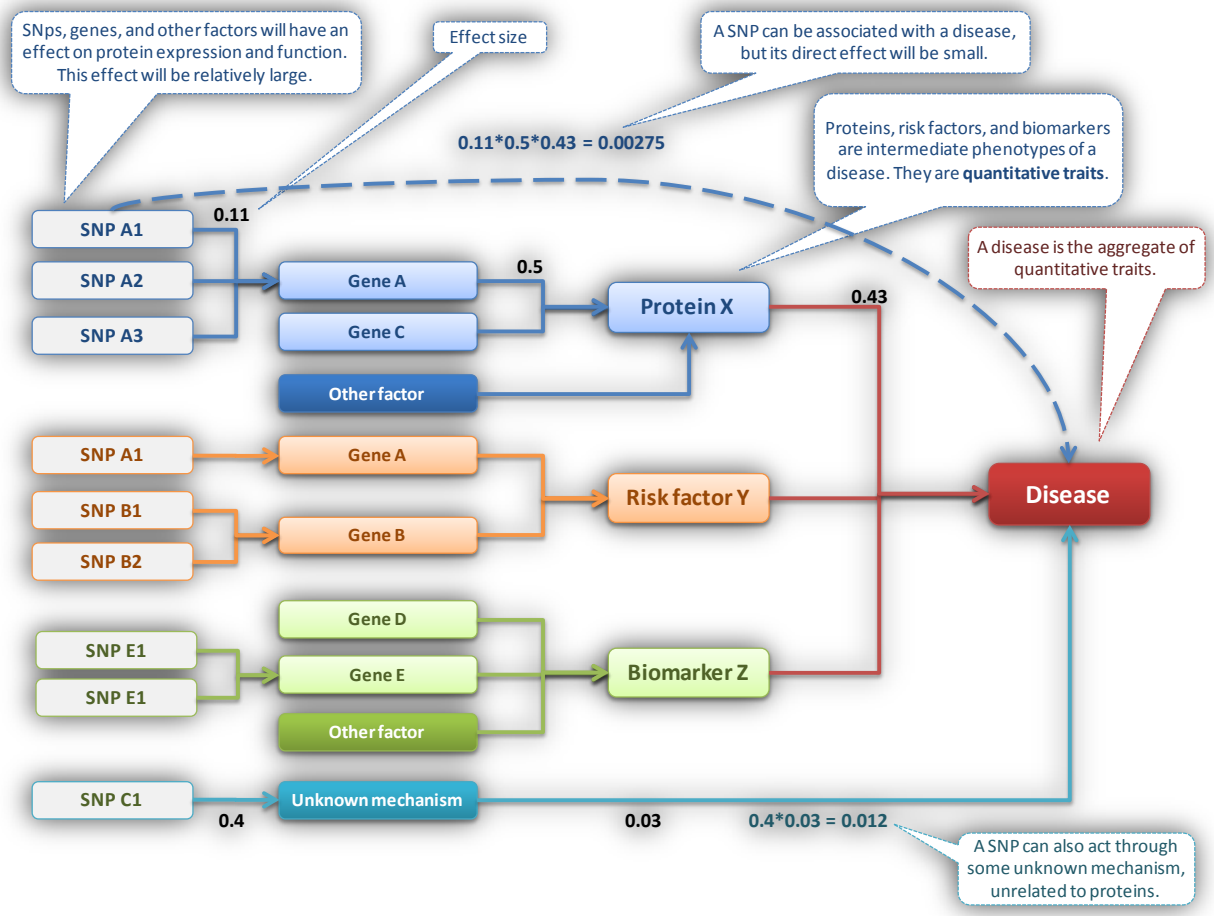


Figure 1: From Genes to Disease

Complex diseases are polygenic by nature. Many genes affect the functions and expression of a variety of biological macromolecules, such as proteins. Defective or low expressed proteins are intermediate traits of the disease (**Protein X**); these or the downstream targets can be measured to assess the risk (**Risk factor Y**); or can be used as a surrogate marker of disease (**Biomarker Z**). For example, *APOB* (Entrez Gene ID: 338) encodes for the *protein* apolipoprotein B that is part of LDL cholesterol particles (LDL). A high LDL has medical significance and is a *risk factor* for CVD; as such APOB and LDL can act as a *biomarker* for disease. Any combination of genes, variants, and other (environmental or behavioural) factors can affect the intermediate traits to some extent.

In this example **SNP A**, **Gene A and C** and some an **other factor** affect the expression or function of the **Protein X** to some degree, the effect size varying from 0.03 to 0.5. **Protein X** itself also has some effect on the **disease**, in this example 0.43. The effect of the individual genes is diluted and conferred through **Protein X**, and therefore the direct effect of a gene on the **disease** is smaller. For instance, **Gene A** affects **Protein X** by 0.05 and **Protein X** affects the **disease** by 0.43, so the overall effect of **Gene A** on the **disease** is only 0.215 (0.03x0.43). A SNP can be coding thereby affecting the protein and indirectly the disease. In this example, **SNP A1** affects **Gene A** by 0.11, **Gene A** affects **Protein X** by 0.5, and **Protein X** affects the disease by 0.43, the overall effect of **SNP A1** on the disease would be 0.00275. A SNP can also be non–coding,

and have an apparent direct effect on the disease through a process that does not necessarily involve known biological pathways. For instance, **SNP C1** affects some **unknown mechanism** by 0.4, and the **unknown mechanism** affects the disease by 0.03. The overall effect of **SNP C1** on the disease would be 0.012. If the effect size of SNP A1 and SNP C1 is sufficiently large to be discovered by a GWAS, the SNP can be associated with the disease.

Variants and genes indirectly influence the disease through intermediate (quantitative) traits which can be proteins, risk factors, or biomarkers. In essence any complex disease is the aggregate of its quantitative traits. Genes and variants can be investigated in the setting of quantitative traits through a method called *Mendelian randomization*.

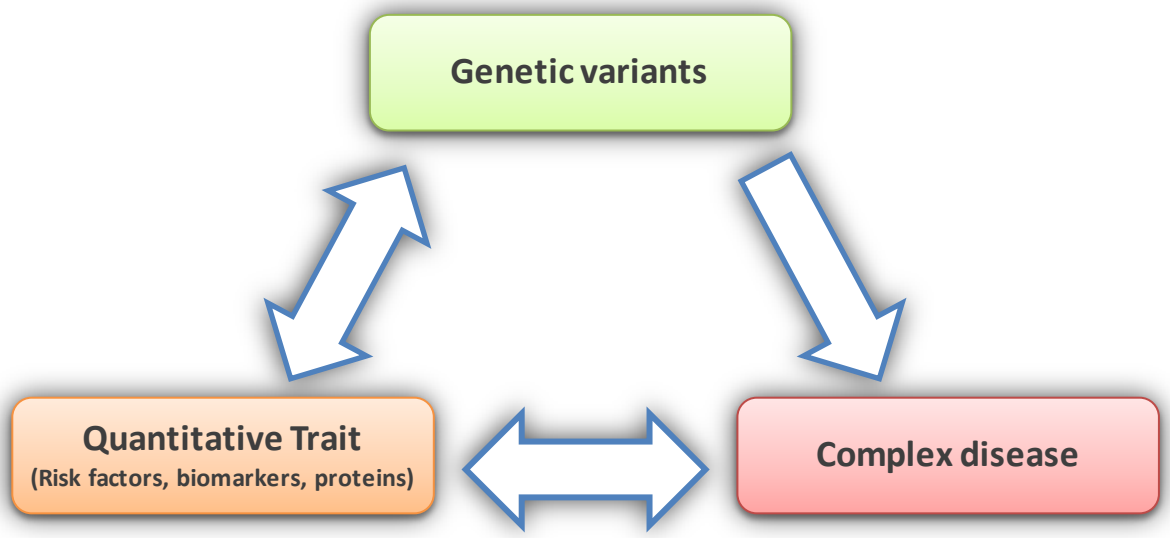


Figure 2: Mendelian randomization can assess biomarker validity through quantitative trait analysis

In *Mendelian randomization* the effect size of genetic variants on quantitative traits (risk factors, biomarkers or proteins), and of quantitative traits on the disease is assessed. From the combinatory results the effect size of the genetic variant on the disease is predicted. If the predicted effect agrees with the observed effect, the quantitative trait and the genetic variant are likely to be causal to the disease. Adapted from Schunkert *et al.*[7].

A Bird's–eye View on Atherosclerosis

Atherosclerosis is a chronic inflammatory disease

Several excellent reviews describe the processes involved in atherosclerosis [23-26]; a summary of literature should suffice. The term "atherosclerosis" was first coined by Jean Lobstein in 1829 and it has been a human disease for over 3500 years as revealed by a pathological report on the aorta of the Egyptian Kin Menphtah[27,28]. Nowadays it is considered to be a chronic inflammatory disease of the arterial wall at susceptible sites in large and medium-sized elastic and muscular arteries. Atherogenesis is initiated by the formation of endothelial lesions, and progressively develops on a lifelong continuum of arterial wall changes[23]. The initial lesions typically begin in the intima and progress into the media and adventitia; the current American Heart Association (AHA) classification categorizes lesions into six types (**Table 1**).

Type I (initial) lesions consist of monocyte-derived macrophages and T lymphocytes, and can be considered as purely inflammatory, most common in infants and young children[29,30]. In hypercholesterolaemic individuals the influx of inflammatory cells is preceded by the deposition, accumulation, oxidation and modification of (circulating) lipids in the vascular wall[29]. In either case the inflammatory response is triggered by endothelial dysfunction, caused by (a combination of) many factors[25]. Endothelial dysfunction leads to increased endothelial permeability and adhesiveness for leukocytes. The endothelium becomes pro–coagulant and the dysfunction triggers the release of vasoactive molecules, cytokines, chemokines, and growth factors. During atherogenesis primarily monocyte-derived macrophages (recruited from the vessel lumen) and subtypes of T lymphocytes (recruited from the adventitia and *vasa vasorum*) mediate the inflammatory response that is able to neutralize these stressors. However, when ineffective the response will go on indefinitely. In time monocyte-derived macrophage will differentiate into foam cells and smooth-muscle cells (SMCs) will migrate and proliferate, leading to Type II lesions (fatty streak). These cells will become intermixed with the area of inflammation leading to an intermediate lesion (Type III), and can thicken the artery wall leading to dilation, thus preventing luminal narrowing ("outward remodelling"). While the inflammation continues, both the influx and the proliferation of macrophages and lymphocytes increase. The activation of the cells triggers the release of more cytokines, chemokines, growth factors, and proteases, leading to lesion progression and the deposit of an extracellular lipid core (Type IV, atheroma). The combination of activated inflammatory cells, migrated and proliferating smooth-muscle cells, and fibrosis induces the restructuring of the lesion to an advanced lesion with a fibrous cap overlying a calcified, lipid, or necrotic core (Type V, fibroatheroma). Type VI (complicated) lesions are characterized by endothelial defects, a thrombus, and intraplaque haemorrhage facilitated by intraplaque neovascularisation. The process of remodelling can no longer compensate and luminal narrowing alters the blood flow, increasing the chance of occlusion and distal ischaemia of tissue and/or organs.

Table 1: Classification of atherosclerotic lesions

Nomenclature	Histology	Growth	Age period	Clinical outcome
Type I *initial lesion*	Influx of inflammatory cells, macrophage and foam cell formation	Mainly lipid accumulation	From first decade	Asymptomatic, clinically silent
Type II *fatty streak*	Intracellular lipid accumulation, foam cell formation			
Type III *intermediate lesion*	Continuing process of Type II including extracellular lipid accumulation		From third decade	
Type IV *atheroma*	Continuing process of Type II, lipid core			Asymptomatic, or clinically evident (symptomatic)
Type V *fibroatheroma*	A fibrotic cap with either or a combination of a lipid, calcified and/or necrotic core	Smooth muscle cell and collagen increase, while lipid accumulation continues	From fourth decade	
Type VI *complicated*	Endothelial erosion, plaque rupture, (intra)plaque haemorrhage, thrombosis	Possible hematoma and thrombosis		

Stability and Vulnerability

While remodelling continues and luminal narrowing is inevitable, the actual plaque can remain "stable", *i.e.* it is not rupture prone, and therefore is not the cause of acute clinical manifestations such as MI or stroke. Instead, it is thought that upon disruption of the atheroma thrombogenic plaque components are exposed which leads to thrombus formation and accelerates luminal stenosis of the artery leading to a clinical event. In symptomatic atherosclerotic carotid disease the thrombus could also cause distal embolisation and subsequent cerebral ischaemia[31,32].

Histological analyses have identified several plaque characteristics associated with clinical events and the vulnerable (unstable) nature of the plaque. The most common rupture prone vulnerable plaque usually has a large lipid core, a thin fibrous cap depleted of SMCs, and extensive infiltration of inflammatory cells – mostly macrophages. Vulnerable plaques exhibit neovascularisation contributing to plaque progression and intraplaque haemorrhage. The seemingly beneficial process of outward remodelling increases the biomechanical stress on the fibrous cap. Less commonly a calcified nodule might protrude into the lumen, conferring high risk for thrombosis. In addition superficial plaque erosion could expose the thrombogenic connective tissue to the blood leading to thrombus formation[33].

BOX 2 | FACTORS MODULATING ATHEROSCLEROSIS

Age-related metabolism
Gender specific metabolism
Hormonal regulation
Dyslipidaemia
Hyperhomocysteinaemia
Hyperglycemia
Free radical formation
Hypertension
Diabetes mellitus
Chronic kidney disease
Systemic Erythematosus Lupus (SLE)
Infectious microorganisms
Inflammatory status
Substance use
Medicine use

Risk Factors modulating Atherosclerosis

Numerous biological and environmental factors influence the acceleration and extent of atherosclerotic lesion formation (several are mentioned here and in BOX 2). By facilitating development and progression, and modulating atherosclerosis into a very complicated systemic vascular disease, they contribute heavily to the clinical outcome. Risk factors are causally related to the atherosclerosis and used in risk assessment algorithms such as ATPIII or the Framingham Heart Score (BOX 1)[5].

Hypertension

High blood pressure (hypertension) is a major risk factor for endothelial dysfunction and subsequent atherosclerotic disease[34]. Worldwide prevalence is estimated to 1 billion and causative to an estimated 7.1 million deaths annually. The relationship between hypertension and CVD has been consistent and independent from other risk factors. The lifetime risk for hypertension is reported to be approximately 90% for men and women who are non–hypertensive at 55 or 65 years old and survive to age 80 to 85. For every increase of 20 mm Hg systolic blood pressure (BP) mortality rates from stroke and ischaemic heart disease (IHD) increases two-fold. Hypertension is defined as a systolic BP greater than 120 mm Hg, and a diastolic BP greater than 80 mm Hg (Table 2).

Table 2: Classification of hypertension

Classification	Systolic BP [mm Hg]	Diastolic BP [mm Hg]
Normal	< 120	and < 80
Prehypertension	120-139	or 80-89
Stage I Hypertension	140-159	or 90-99
Stage II Hypertension	≥ 160	or ≥ 100

Atherosclerosis can be primary to renovascular hypertension in renal artery stenosis and thereby indirectly influencing cardiac function. In hypertensive patients the risk for MI is increased due to a greater demand of myocardial oxygen which is exacerbated by coronary artery stenosis and calcification of the arterial wall leading to coronary stiffness[34]. As a consequence shear stress on the vessel wall and atherosclerotic plaques is increased, thereby increasing the chance of plaque disruption and subsequent clinical events[34].

Dyslipidaemia

Hypercholesterolaemia, hyperlipidaemia, and hypertriglyceridaemia are closely related disorders of lipid metabolism (dyslipidaemia) and are considered to be principal risk factors for vascular dysfunction and atherosclerosis. Hypercholesterolaemia is characterized by elevated plasma levels of total cholesterol (TC) due to abnormalities in lipoprotein levels, a hallmark of hyperlipidaemia. Lipoproteins are carriers of cholesterol esters, triglycerides (TAG) and apolipoproteins. Hypertriglyceridaemia is characterised by abnormal serum levels of free fatty acids (triglycerides). Disturbances in the expression of lipoproteins are a consequence of genetic defects, diet, comorbidity, and/or endocrinal disorders.

In dyslipidaemia an excess of LDL can enter the vascular tissue and is oxidized (oxLDL). The CD36 multi–ligand scavenger receptor on the resident monocyte–derived macrophages can bind oxLDL and facilitate endocytosis. Endothelial cells (ECs) and macrophages are activated by accumulating levels of subendothelial oxLDL which increases the expression of many proinflammatory chemo– and cytokines in addition to ectopic expression of vascular adhesion molecules. Macrophage gradually transform into foam cells through incorporation of oxLDL. Foam cell formation is a hallmark of atherosclerotic lesion development, and thus LDL facilitates atherosclerosis by initiating a viscous circle of LDL oxidisation, oxLDL accumulation, EC and macrophage activation, proinflammatory signalling, foam cell formation, and increasing endothelial permeability.

Hyperhomocysteinaemia

Homocysteine (Hcy) is derived from metabolic conversion of methionine by methylenetetrahydrofolate reductase (MTHFR), a potentially toxic amino acid. Homocysteine

increases reactive oxygen species (ROS) in the endothelium, thereby decreasing availability of endothelium-derived nitric oxide (NO) and impairing the intracellular redox buffer system facilitating endothelial dysfunction[35]. Homocysteine potentiates platelet aggregation and increases macrophage–derived tissue factor activity, thereby conferring prothrombotic effects[35]. Additionally it increases CD36 expression and oxLDL formation, *i.e.* aggravating atherogenesis. To complicate things even further proinflammatory signalling, such as the nuclear factor kappa–B (NF–κB) pathway, is influenced by Hcy[35]. Furthermore, evidence suggests that homocysteine has corrosive actions on structural proteins such as collagens, elastins, and proteoglycans[36].

Circulating Cells and the Immune System

Circulating cells play a pivotal role in atherogenesis trough the expression of pro– and anti–inflammatory cytokines capable of modulating the innate and adaptive immune response. Macrophages are involved in all stages of lesion development and progression, and in addition subtypes of T lymphocytes, eosinophiles, neutrophiles, mast cells, B cells, and platelets are play a role in both pro– and anti–atherosclerotic processes.

Upon activation through interaction with macrophages and dendritic cells T lymphocytes proliferate and differentiate into T helper cells. Constituting up to 20% of the total amount of nucleated cells, T–lymphocytes produce pro–inflammatory cytokines, such as interferon gamma (*IFN–γ*) and tumour necrosis factor alfa (*TNF–α*)[37-39]. B–lymphocyte transfection in mice suggests an atheroprotective role evident by the reduced lesion development in recipients[40]. In LDL receptor deficient mice immunoglobulin M is required for atheroprotection[41]. Activated mast cells are found in the shoulder region of the plaque, thus playing a role in plaque destabilization which is possibly mediated through secretion of soluble factors by T–cells or macrophages and frequently associated with MI[42-45]. Hence, both the adaptive and innate immune system play a role in atherogenesis.

Biomarkers of Atherosclerotic Disease

Identifying ideal biomarkers

Biomarkers are objectively measured indicators of a biological or pathophysiological stat that can be useful in disease management. Vasan has written an excellent review on the characteristics a (novel) ideal biomarker should have[5]. In brief, several types can be discerned depending on the intended use (diagnostic, prognostic, surrogate). When a potential novel biomarker is identified it should be validated in terms of sensitivity, specificity and reproducibility (method validation). During method validation reference limits, discrimination limits and the thresholds defining risk are determined. Subsequently the biomarker should be qualified in terms of the relation it has to the disease biology or clinical outcome (clinical validation). The value of a novel biomarker should be incremental compared to current standards. The significance and the area under the receiver-operator curve (AUC) should be evaluated as well as the reclassifying value of the novel biomarker[46,47].

Discovery strategies

Several strategies are applied in the discovery of biomarkers[5]. Proteomics approaches assess the differences in protein expression between diseased and healthy tissues. Methodologies include Western blotting, ELISA, and quantitative proteomics (Q–Proteomics). In cancer research gene expression analysis has matured and nowadays it is used to identify biomarkers of CVD. Changes in mRNA expression reflect the disease state and several methods can be applied including Northern

BOX 3 | SELECTION OF NOVEL GENETIC BIOMARKERS FOR CVD[4]

Atherosclerosis
- IL6
- APOC3
- APOB
- APOE
- FGA
- MTHFR
- MMP3

Coronary artery disease
- IL6
- PON1
- PON2
- APOB
- CRP
- HP
- ADRB3
- ADD1
- NOS3
- MMP3

Obesity
- ADRA2
- ADRB2

Myocardial infarction/ischaemia
- ADRA2
- CCR5
- CRP
- FGB
- PECAM1
- SELP
- HSP70A1
- MMP3

Adbominal aortic aneurysm
- HMOX1
- MTHFR

Stroke
- MMP1
- MMP3
- MTHFR
- FGB
- NOS3
- CD14

Adapted from Gibson, *et al.*[4].

blotting, RT–PCR, and microarray technologies which assess thousands of transcripts at once. Microarray and Q–Proteomics are limited by the availability of transcripts on the chips (microarrays), the high false discovery rate, and the heterogeneity of the tissue used[5]. Genomics strategies emphasising the use of SNPs associated with disease have been very successful in identifying numerous novel markers of disease (BOX 3)[4]. These studies take the form of linkage (**Family linkage analysis**) or association studies (**Genome–Wide Association Studies**) and advanced bioinformatics are used to evaluate results.

Examples

The list of established and putative biomarkers seems never–ending (BOX 4). Several risk factors (dyslipidaemia, homocysteinaemia, inflammatory status – see also relevant sections) include biomarkers of disease, such as ApoB, LDL, CETP, homocysteine, IL6, IL8. C–reactive protein (CRP) is an established inflammatory and CVD biomarker, although several studies have been inconclusive regarding its causal role in CVD[48,49]. It has been suggested CRP could be a marker of underlying atherosclerosis[50]. Galectin–3 has been associated with MI and is found in atherosclerotic lesions[51,52]. Mice deficient for ApoE and Galectin–3 develop significantly fewer atherosclerotic lesions and show less inflammatory infiltrates compare to ApoE-deficient mice. Cystatin C is used as a measure of glomerular filtration rate and has been associated with hypertension and diabetes[53]. Several studies have linked Cystatin C to heart failure, CAD, PAD, and stroke. Lipoprotein–associated phospholipase 2 (Lp-PLA2) acts on oxLDL and indirectly promotes vascular inflammation[54-56]. It has been consistently link to risk for CVD and is used in both primary and secondary prevention. Several drugs, including Darapladib and statins, can inhibit or lower Lp-PLA2 expression.

BOX 4 | BIOMARKERS FOR CVD[1-5]

Molecular

Established
- LDL cholesterol
- HDL cholesterol
- HbA1c
- cTn
- CK–MB

Emerging
- CRP
- Lp–PLA2
- NT–proBNP
- NT–proANP

Developing
- sCD40 ligand
- Myeloperoxidase (MPO)
- Cystatin C
- Galectin–3
- Fatty acid binding protein (FABP)

Imaging
- Carotid intima–media thickness
- Coronary intravascular ultrasound
- Brachial artery flow–mediated vasodilation
- Carotid artery MRI

Genomics Basics

Genomic variants

One is the loneliest number

The International HapMap Consortium (HapMap) recognised the need for a freely available database of common variations in the human genome to facilitate the discovery of the variants affecting common diseases and the development of new therapeutic strategies[57]. Single–nucleotide polymorphisms are the best known examples of common, qualitative genetic variations influencing diseases and quantitative traits. By convention common SNPs are defined as occurring in more than 1% of the general population, whereas rare SNPs (point mutations) occur in less than 1% of the general population (Figure 3)[58]. As a rule rare mutations have a high functional impact, though common SNPs with a high impact have been identified. Base pair changes involving SNPs occur with a frequency of ≈1 in every 300 bases of DNA sequence[57]. In terms of genetic diversity among individuals SNPs with a frequency of more than 1% represent ≈0.3% of the genome, about 10 million base pairs, constituting about 90% of the total extent of genetic variations[57,59]. The majority of single–nucleotide changes, *e.g.* the replacement of a guanine (G) in the paternal DNA strand for an adenine (T) in the maternal allele, are *noncoding*, have no functional impact on the proteome and phenotype (silent), though gene expression and splicing may be affected. In contrast *coding* SNPs can be *non–synonymous*, if the replacement of the nucleotide changes the codon and the amino acid it codes for. Non–synonymous SNPs could have functional impact in amino acid structure (*missense*) or premature truncation of the protein (*nonsense*). In contrast synonymous SNPs change the codon, but do not affect the amino acid structure. Generally synonymous SNPs are silent, though they could affect gene splicing patterns.

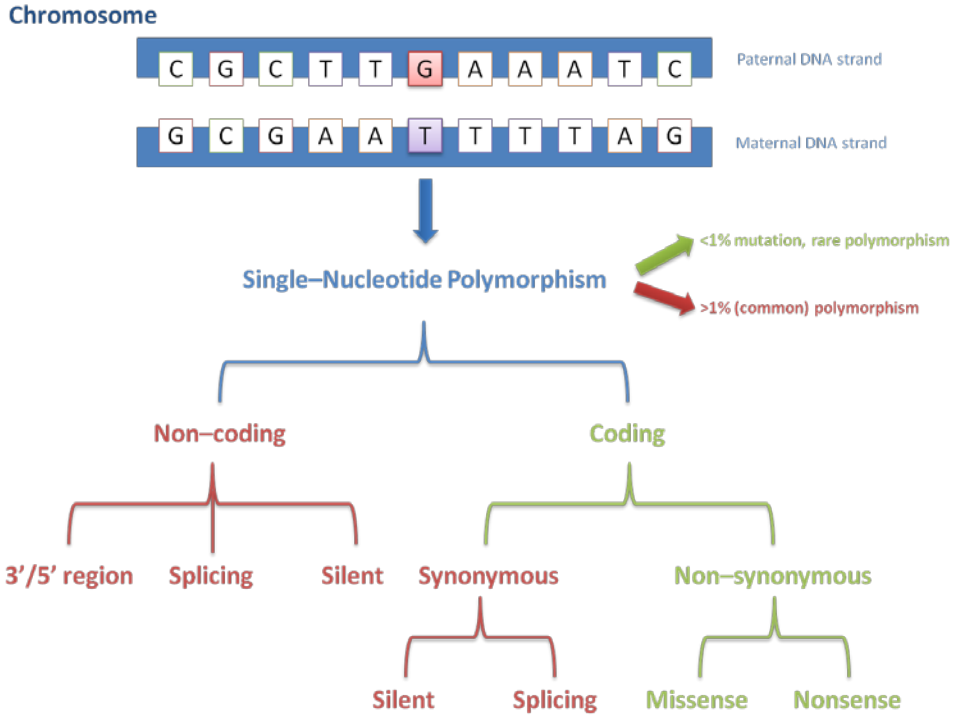


Figure 3: Single–nucleotide polymorphisms

A genetic locus consists of two alleles. An individual receives one allele from each parent. One parent might have a SNP on one allele. A SNP can be *non–coding* or *coding*. A non–coding SNP could have no effect (*silent*) on the amino acid sequence. However, it could result in alternate *splicing* or affect the 3'- or 5'-UTR of a transcript. A coding SNP could be *synonymous*, or *non–synonymous*. Synonymous SNPs have no change in amino acid sequence, and can be silent or affect splicing. Non–synonymous SNPs entail a change in amino acid sequence, thereby altering protein structure (*missense*) or leading to a premature translational stop (*nonsense*). Adapted from Pollex *et al.*[9].

Haplotype

A set of adjacent SNPs that are in *linkage disequilibrium* with each other can be compiled into a *haplotype*. Alleles of a haplotype are transmitted together and are statistically associated with each other; therefore they can be informative of the polymorphic nature of the genomic region they cover. Within haplotypes several SNPs can be "tagging" (*tag SNP*), they uniquely identify the haplotype (**Figure 4**). The advantage of inferring haplotypes is that not all SNPs need to be genotyped. It estimated that roughly 300,000-600,000 tag SNPs are informative of the ≈10 million common SNPs in human populations.

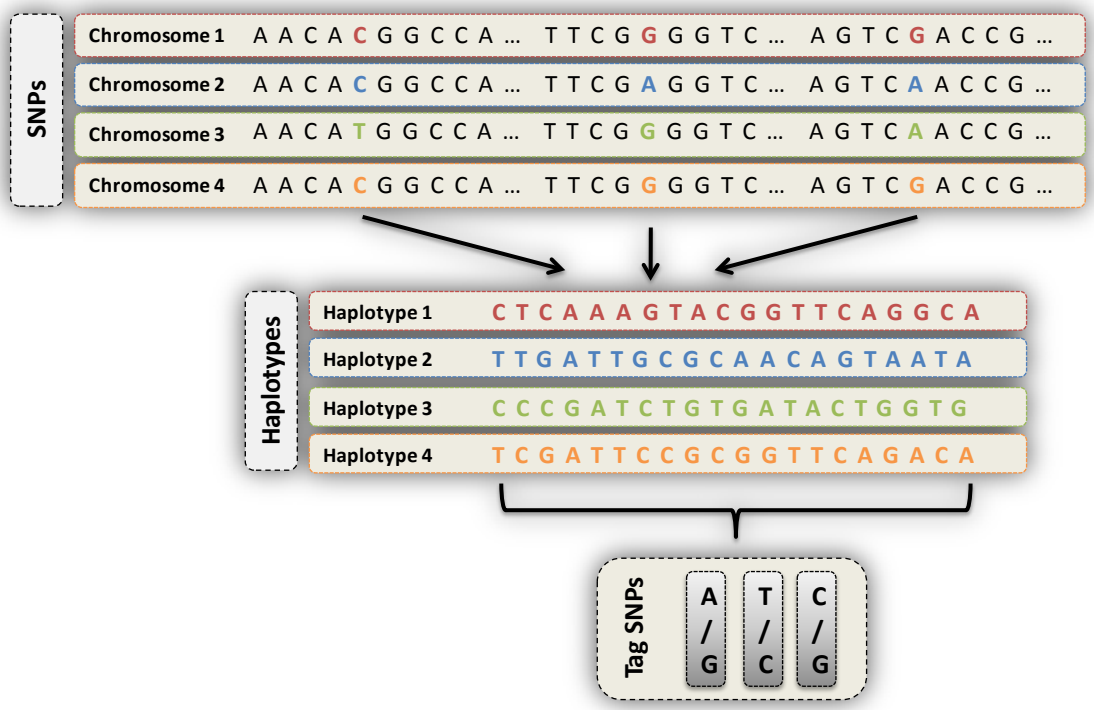


Figure 4: From SNPs through Haplotypes to Tag SNPs

On each chromosome several SNPs are identified that are in LD with each other. Consequently the haplotype of each chromosomal region can be inferred. Within each haplotype tag SNPs can be identified that uniquely identify the haplotype. Adapted from The International HapMap[57].

Structural variants

The recent discovery of copy number variations (CNV) has shifted the paradigm. Contrary to SNPs, CNVs are structural variants quantitative in nature which affect gene dosage and copy number of a particular genomic region[9,58,59]. Copy number variations are intermediate size structural variants spanning a region between 1,000 to $5x10^{-6}$ nucleotides. Insertions, deletions and inversions are types of the cytogenic rearrangements comprising CNVs. Other structural variants include short tandem repeats (STR, microsatellites), transposons, and the less common but highly informative variable number of tandem repeats (VNTR).

Family & heritability

It runs in the family

A positive history of cardiovascular disease and associated risk factors tend to aggregate in families. Several studies have shown that a (self–reported) positive family history for cardiovascular risk

factors, such as hypertension, diabetes, and hypercholesterolaemia, is strongly and independently associated with early–onset CVD[60]. In addition, the assessment of family history increases the ability to discriminate and reclassify patients[60]. The Framhingham Heart Score is frequently used to classify patients into risk groups, and family history is a valuable, independent addition to this algorithm[61,62]. Even individuals with a low Framingham score but positive family history have a higher risk for developing CVD[63]. Common behaviour, environmental factors, and genetic susceptibility are reflected in family history as a factor of complex diseases.

Heritability

Heritability is mathematically described as "narrow–sense heritability", h^2, and defined as "the proportion of phenotypic variance in a population attributable to additive genetic factors in individuals"[64,65]. In a formula:

$$h^2 = \frac{Var\,(A)}{Var(P)} \qquad \textit{Equation (1)}$$

Where h^2 (the heritability) equals the ratio of the total variance (Var) of additive genetic factors (A) and the total variance of the phenotype (P). The overall estimates of heritability for any disease are based on a proper classification of the disease and extensive family studies. The misclassification bias is particularly important in case–control studies and could negatively influence the outcome[8]. The accuracy of the estimated heritability is further shaped by family studies and can be inflated by genetic dominance, *epistasis*, shared familial environments, and correlations between genes and environmental factors[64].

Study designs

Family linkage analysis

Genome–wide family linkage analysis has been used as an instrument to address the familial genetic component driving CVD. In a family based linkage analysis clinical data is collected from parents and siblings, and a genome–wide scan using microsatellites distributed evenly across the genome every 10 cM (about 1 million base pairs) is performed[66]. The families included have at least one parent with a history of disease, and preferably sibling pairs of disease. The genome is systematically screened for linkage with disease: affected siblings and family members extensively share alleles (linkage peak) compared to unaffected individuals (families)[66]. For each microsatellite marker a logarithm of the odds (LOD) score is calculated, and a LOD > 3.5, or probability of $<1x10^{-6}$, indicates a significant association of a locus near or in the microsatellite region with the disease[66]. Family linkage studies survey the entire genome and identify genes with a large effect on the phenotype. These studies are best–suited to identify genes involved in monogenic, *Mendelian traits and diseases* which are rare and segregate in an autosomal, dominant or recessive, X-linked, or mitochondrial mode. Therefore family–studies are likely to identify rare genes that have little influence in a general population.

Genome–Wide Association Studies

Marker sets

Genome–wide association studies take advantage of the fact that the majority of SNPs are in LD with each other. Commercially available DNA genotyping chips can achieve high accuracy and coverage (≈68-92% of all known SNPs) in non–African samples (from North-American descent, CEU)[67]. The

producers of such chips apply different strategies when selecting SNP represented on their chips[67]. Illumina, Inc. (www.illumina.com, San Diego, CA, U.S.A.) carefully selected 300,000 to > 1,000,000 tag SNPs with a high LD for its HumanHap and HumanOmni series of genotyping chips. By comparison Affymetrix, Inc. (www.affymetrix.com, Santa Clara, CA, U.S.A.) randomly chose SNPs evenly spread over the genome for its products such as the Genome–Wide Human SNP Array 5.0 (SNP5) with over 500,000 SNPs and 420,000 CNVs. Both parties also offer chips (Affymetrix' MegAllele and Illumina's Human-1 BeadChip) that contain a panel of functional SNPs (nsSNPs).

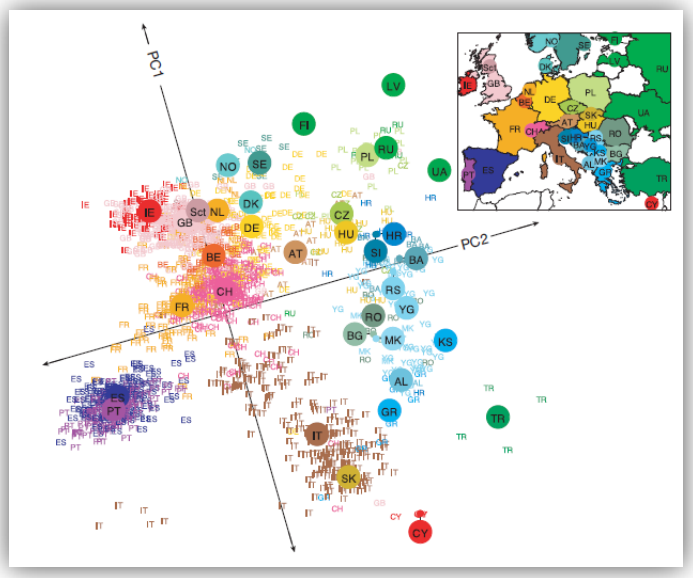


Figure 5: A visualisation of principle component analysis of genetic variation mimics the European geographical map

Principle component analysis can be used to infer population structure based on thousands of genetic markers. This multi-dimensional scaling technique can then be used to control for substructure and increase power to detect true associations. Taken from Novembre *et al.*[6].

Population structure

The HapMap is based on 270 samples from different populations making it possible to address questions regarding population. The Yoruba people of Ibadan, Nigeria provided 30 trio-sets (two parents and one adult child). Additionally, 45 unrelated individuals from the Tokyo area, Japan, and 45 unrelated individuals from Beijing, China, provided samples. Lastly 30 trios resident of Utah, U.S.A. with northern and western European ancestry provided samples.

Results from various studies indicate important difference among ethnicities in terms of coverage by marker sets[6,67-70], and population structure[6,68,70]. Novembre *et al.* vividly showed how a geographical map of Europe arises as a two-dimensional digest of genetic variation in Europeans (**Figure 5**). A principle component analysis (PCA) is applied to studies thousands of markers and capture population structure in a three-dimensional space[6,68-71]. These types of studies have important consequences for genotype-phenotype association studies, because spurious associations might arise when comparing cases and controls[6,68-71]. PCA can be efficiently used to correct for population structure and increase power to detect true associations[6,68-71]. The accuracy of genotyping and inferring an individual's ancestry is surprisingly high (within a few hundred kilometers) and could be used to verify self-reported ancestry[6].

CDCV hypothesis

The rationale for GWAS is the "common disease, common variant" (CDCV) hypothesis which dictates that common variants explain the majority of genetic variance in complex disease, and rare variants affect Mendelian traits and diseases[72]. In theory rare variants have an intermediate to high *penetrance* (the proportion of carriers that also express the associated trait), and therefore have a higher *effect size* measured in odds ratio (OR), hazard ratio (HR) or relative risk (RR)[64]. In contrast common variants confer relatively small increments of risk, and disease loci have relatively simple spectra (a handful of alleles account for a large proportion of the disease class)[64,73]. Population genetics suggests that any high penetrance variant results in decreased reproductive fitness, thus facilitating selection bias and a reduction in risk allele frequency. It appears that common variants

will rarely be solely responsible for complex diseases and related traits, whereas rare variants will seldom be associated with common diseases but frequently with Mendelian diseases and traits (Figure 6). However, Schork *et al.* and Manolio *et al.* note that both rare and common variants can be associated with a common disease such as hypertension[64,72].

High Numbers equal more Power

For a random sampling of controls in a case–control GWAS in order to achieve a power of 80% and a p-value (α) of 10^{-6} for a trait with a prevalence of 5% and an minor allele frequency of 1%, the sample size (*N*) would have to be at least 2,500[8]. The sample size would increase even more with increasing allele frequency (MAF > 5%) and prevalence of the trait (> 20%). GWAS are a very powerful tool to investigate the genetics of complex disease, yet the common variants discovered explain only a small proportion of the heritability, and over 80% of the variants fall outside the coding regions of the genome[64]. The "missing heritability" is partly explained by the array designs which focus on common variants with a frequency of >1-5% and poorly capture structural or rarer variants[64]. In addition GWAS have low power to detect gene–gene and gene–environment interactions[64].

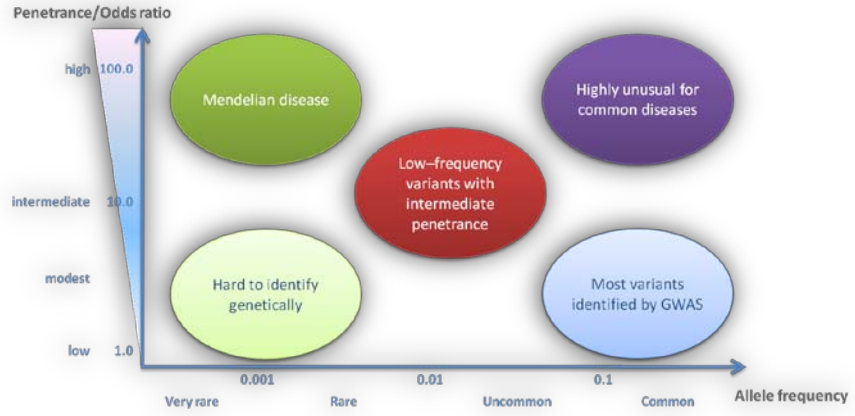


Figure 6: Penetrance, Odds ratio, Allele frequency, and *N*

Low allele frequency implies high penetrance and a high odds ratio. Mendelian diseases often show this type of relation. GWAS applied to study common diseases usually identify common alleles with a high frequency (> 0.01-0.10) and subsequently low penetrance and odds ratio. Adapted from McCarthy *et al.*[8].

The majority of GWAS identified variants with an OR of 1.1-1.5. From estimates of power the numbers needed to achieve a power >0.20 and an OR of 1.1-1.5 would range from 1000-3000[74]. This issue can be circumvented by testing multiple variants (multimarkers test) for association with disease, substantially increasing power[75], or by focusing on less common variants with frequencies from 1%-5%[76]. Additionally searches for any allelic combination (haplotype) can be run to find a putative variant associated with disease (exhaustive multimarker test)[76].

Advantages

Compared to linkage analysis studies, GWAS have several advantages. First, a GWAS can achieve greater statistical power for complex disease, compared to linkage analyses, and uncover variants with modest effect size (OR 1.1-2.0), unlikely to be discovered by linkage studies. Second, GWAS take an agnostic approach towards identifying genetic factors involved in the disease under investigation. Third, although a GWAS can be family based, unrelated individuals can be used to investigate a disease. Fourth, GWAS have a *finer resolution* due to the larger number of variants typed by the technique which makes the identification of variants with a relative small impact size possible.

Imputation

Imputation analysis can overcome the disadvantages one marker set might have over the other, facilitate meta-analysis[77], or the large number of samples needed to find genome–wide significance.

Moreover missing genotype data can be imputed from typed markers based on LD. Several (open-source) imputation methods are available (MACH, IMPUTE, fastPHASE, PLINK and Beagle) and recently the performance of each were examined. When imputing SNP5 data to Affymetrix Genome–Wide Human Array 6.0 or Illumina HumanHap 550 Bead array data Beagle, MACH, and IMPUTE consistently outperformed fastPHASE and PLINK in terms of accuracy and efficiency[78,79]. Nowadays a combinatory approach, imputing data and performing multimarker tests, can decrease spurious associations, increase power[80] and the ability to detect true associations of (less) common risk alleles closer to the actual causal variant[74,81]. Additionally, the coverage of SNP5 data after imputation of missing genotypes increases from 65 to 73 for $r^2 > 0.8$ or 43 to 54 for $r^2 = 1$[80].

Candidate Genes

The discovery of monogenetic disorders and the identification of genes involved in CVD through familial linkage analyses have been the basis for more in–depth investigations to the nature of atherosclerotic genetics. Candidate gene association studies are performed under the assumption that the genes examined are biologically relevant or (causally) associated with the disease of interest.

Potentially powerful, in essence candidate gene studies are biased against the identification of novel genes for they are based on prior knowledge. For example, several studies have identified a significant influence of CRP levels on cardiovascular risk. Yet, the genomic investigation of SNPs in the CRP gene has yielded little results and the overall variance on CRP levels explained by *CRP* genotypes has been ≈2%[7,50,82,83]. When results from family based studies are the rationale for pursuing a candidate gene study, the results are likely only applicable to a specific, small subgroup of the population. As in the case of CRP, the statistical strength of association was leading in many studies, but is by no means definitive proof of any causality[7,50,83]. Many traditional risk factors independently associate with CRP levels, and atherosclerosis itself could trigger an increase in CRP levels (reverse causality)[7].

Regulating Gene Expression

The central dogma of molecular genetics is the storage of the gist of life in DNA, which is transcribed to RNA, and subsequently translated into amino acid sequences, culminating in a vast array of proteins. Indeed, many genes are differentially expressed in atherosclerotic disease, including *ABCA1*, *CCR2* (MCP-1 receptor), *CCR5* (Rantes-receptor), *SPP1*, *TIMP1*, and *VCAM1*. This central dogma is misleading since recently miRNA have been identified as important modifiers of gene transcription and translation, influencing development, differentiation, proliferation, and human disease through changes in protein synthesis[84-86]. One can add even more layers of complexity when taking into consideration the vast array of genetic variants that can modulate gene expression – especially those variants that are not silent, but result in alternative splicing, affect 3'– and 5'–regions, or are non–synonymous, and can be imprinted or epigenetically modified. Moreover *cis*– and *trans*–acting regulatory elements, DNA sequences that are not transcribed or translated, impact on gene expression[87,88].

The exact mechanisms by which variants discovered through GWAS influence the phenotype often remain intangible. In contrast, many expressed biomarkers and risk factors have been identified that are not only associated but also causally or consequently related with the disease. Several studies have identified expression quantitative trait loci (eQTL)[89] and protein QTL[90], showing how genetic

variants can have *cis*– and *trans*–acting regulatory effects on mRNA levels and influence protein expression. A polymorphism in *VKORC1* (-1639G>A) generates a suppressor E-box binding site, thereby regulating gene expression and directly influencing warfarin dosage accross ethnicities[91]. These studies underpin the notion that most diseases are affected by variations in regulatory rather than functional genomic regions.

Genomics of risk factors

Hypertensive genomics

The Wellcome Trust Case Control Consortium (WTCCC) investigated the association of common variants in clinically hypertensive individuals compared to two population control cohorts, the British 1958 Birth Cohort Controls (58BC) and the UK Blood Services Controls (UKBS)[92]. Previous studies identified several variants, specifically those in *WNK1* (WNK lysine deficient protein kinase 1)[92]. The WTCCC study did not identify any variants with a signal of $p > 5.0x10^{-8}$, which is the minimum significance signal needed for a "true" genome–wide association[92]. The strongest, albeit moderate in terms of genome–wide significance, signal was variant rs2820037 (genetic test, $p = 7.7x10^{-7}$). The genes in the vicinity of the signal do not correspond to clinically or genetically previously identified candidates. The nearest genes are *RYR2* (ryanodine receptor 2) which is associated with stress–induced tachycardia and arrhythmia; *CHRM3* (cholinergic receptor muscarinic 3), a G–protein–coupled receptor; and *ZP4* (zona pellucida glycoprotein 4).

The authors offer three possible explanations for the lack of any significant genome–wide association. First, as a phenotype hypertension may have fewer common variants with strong effect size. Second, the lack signal could be the consequence of poor tagging by the used chip, Affymetrix' GeneChip® Human Mapping 500K Array Set (the previously identified candidate *WNK1* is poorly tagged by this chip). Third, compared to other phenotypes hypertension could be particularly susceptible to the *misclassification bias*, because of the presence of (latent) hypertensive individuals in the control cohorts.

In the WTCCC study, the inclusion of hypertensive individuals is based on three main criteria[92]. First, the individual had a reported history of hypertension prior to the age of 60. Second, the hypertension was based on one confirmed blood pressure recording at seated level of > 150/100 mmHg, or the mean of three separate recordings of > 145/95 mmHg. Third, the individuals included had body mass indices (BMI) of < 30 kg m^{-2}. These criteria were not based on the previously described classification of hypertension used by the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (**Table 2**)[34]. In addition, blood pressure is influenced by lifestyle factors (physical activity, diet, alcohol consumption), but can also be secondary to renal disease or obesity. The WTCCC criteria for hypertensive phenotype definition and the problem of misclassification bias underline the fact that hypertension might be a particularly difficult phenotype to define as a dichotomous trait.

Fatty Genomics

In familial hypercholesterolaemia and combined hyperlipidaemia the majority of the cases (frequency 1:200-1.000) is polygenic and due to defects in a variety of genes involved in lipid metabolism genes. Heterozygous familial hypercholesterolaemia (FH) is a relatively common genetic disorder, occurring 1:500, whereas homozygous FH is rare, occurring 1 in a million births. The majority of the cases are caused by mutations in the gene encoding for LDL receptor (*LDLR*), but other monogenic disorders with a clinical phenotype resembling FH have been reported for *ApoB*, *PCSK9*, *LDLRAP1*, *ABCG5*, and *ABCG8*[93-97]. Mutations in these genes result in a lack of synthesis, impaired transportation, impaired ligand–binding function, impaired clustering in clathrin–coated pits, or impaired recycling of the gene products. The recent confirmation of variants in *ABCG5* and *ABCG8* modulating plasma cholesterol levels in FH which could be influenced by smoking, indicates

the importance of gene–environmental interactions[93]. Familial hypercholesterolaemia is a classical monogenic disorder and characterized by autosomal dominant inheritance with high cholesterol levels, tendon xanthomas, and premature atherosclerotic disease (before the age of 40)[98]. On the other side of the lipid metabolic spectrum stands the identification of a splicing defect in the cholesteryl–ester transfer protein (*CETP*) gene causing increased HDL levels conferring an antiatherogenic effect[99].

In a recent GWAS 22 loci were found to be associated with serum lipid levels (TC, HDL, LDL, TAG, $p < 5x10^{-8}$) in 16 population–based cohorts[100]. Six loci were previously unknown, they include variants in *ABCG5* (total cholesterol, LDL), *TMEM57* (TC), *CTCF-PRMT8* region (HDL), *DNAH11* (LDL), *FADS3-FADS2* (total cholesterol, LDL), and *MADD-FOLH1* region (HDL). The authors performed a GWA network analysis and showed how the SNPs associated with TC levels are involved in major pathways linking cholesterol and sterol metabolism, lipid transporters, and even nutrient response. Additionally the authors analysed risk profiles containing the associated SNPs and determined that the TC-specific SNP risk profile was significantly associated ($p < 0.001$) with clinically defined hypercholesterolaemia (serum cholesterol > 6.5 mmol/L). Area under the receiver–operating–characteristic curve analysis revealed that this risk profile improved prediction beyond age, gender, and BMI from 63% to 66%. Also the TC genetic risk score was significantly associated with carotid intima–media thickness (IMT) and incident coronary heart disease (CHD). However, after adjusting for circulating levels of TC the association with CHD did not remain significant. Furthermore, the TC genetic risk profile improved CHD risk classification when added to existing risk algorithms (such as the Framingham score). This study shows how analysing the genetics of quantitative traits involved in atherosclerotic disease can be useful in a clinical setting, improving risk prediction and classification.

Variants in *MTHFR*

Patients with defects in genes necessary for Hcy metabolism develop severe atherosclerosis and many have their first MI before the age of 21[25]. In patients with no genetic defects in Hcy metabolism, elevated homocysteine levels are associated with an increased risk for symptomatic atherosclerosis[25]. Several variants in *MTHFR* on chromosome 1p36.3 have been the subject of extensive investigation, revealing association with thrombosis, hypertension, CVD, and other diseases. One variant, rs1801133, encodes for a nucleotide change (cytosine to thymine) on position 677 (C677T)[101]. The *common allele* has a cytosine which results in the amino acid alanine and a properly functioning enzyme. The *risk allele* has a thymine which results in a valine and a thermolabile enzyme with reduced activity. A meta–analysis in Caucasian and Asian populations revealed a significant association of the rs1801133 with hypertension[102]. In another meta–analysis Casas *et al.* revealed a significant association of rs1801133 (OR 1.24) with stroke[103]. This was supported by the results from a meta–study that associated ischaemic stroke with mild to moderate hyperhomocysteinaemia and the C677T genotype[104]. Klerk *et al.* performed a meta–analysis of the same variant and found a significant association with coronary heart disease (OR 1.16) particularly in relation with low folate[105]. The authors concluded that this supports the hypothesis that an impaired folate metabolism, resulting in hyperhomocysteinaemia, is causally related to an increased risk of CVD. Carriers of the TT genotype have higher total plasma Hcy levels, 13.3 μmol/L compared to 10.4

μmol/L in subjects with the CC genotype. Some suggest vitamin B or betaine* nutritional complements might be beneficial[106,107]. Interestingly, the association between C677T variants and disease seems to be consistent across traditional ethnic boundaries[108]. As a biomarker homocysteine also seems to be causally involved in the development and progression of atherosclerosis and could serve as a quantitative trait of disease.

Immunological genomics

Polymorphisms in the macrophage scavenger receptor 1 (*SCARB1*) gene have been associated plasma levels of total, LDL cholesterol and BMI, and in addition confer a risk for peripheral artery disease[109,110]. In the Multi–Ethnic Study of Atherosclerosis several subclinical phenotypes (CAC, IMT) were examined in relation to SCARB1 polymorphisms. Across ethnicities and in pooled analyses the C allele of SNP rs10846744 was significantly associated and independent from lipid levels and other cardiovascular risk factors[111]. Additionally polymorphisms in several cytokines and chemokines, such as CD40L, LTA, IL6, Rantes, IL8, IL10, have been associated with cardiovascular risk factors, CAD, MI, and in-stent restenosis after percutaneous coronary intervention[112-115]. Also SNPs in coagulation-fibrinolysis genes have been associated with atherosclerotic disease, including SERPINE1, and F7.

Familial Genomics

Cardiovascular diseases & traits

The heritability of familial aggregation of CVD is estimated at >90% in families with CAD onset before the age 46, whereas the genetic contribution drops to <30% in families with late onset CAD[116]. The heritability of CVD in general is estimated at 0.56[22], and the overall heritability is estimated at 0.14–0.62 for cardiovascular related diseases, risk factors, and biological processes relevant for clinical outcome such as estimated glomerular filtration rate and platelet aggregation[61]. The heritability of complex traits such as weight, BMI, and systolic BP associated with CVD was demonstrated to be age–stratified, with a lower heritability in 70–year–old individuals compared to 40–year–old[20].

The Reykjavik Cohort Study prospectively assessed the cardiovascular risk of 9328 males and 10,062 females aged 33–81 years[117]. Depending on gender, hazard ratios of coronary heart disease for individuals with a positive family history (one or more first–degree relatives with MI) were between 1.75 and 1.83 (95% CI 1.59–2.11) attributing at least 15% of all MIs to a positive family history. Angiography studies showed family history as an independent risk factor for angiographically evident CAD[118]. Additionally a positive family history is associated with pre–clinical atherosclerosis as measured by IMT of which the heritability is estimated at 0.35[119,120]. The majority of studies investigating familial aggregation of CVD found that the earlier the age of onset, the higher the risk for relatives[121]. The high concordance rate in monozygotic twins underscores the clear genetic component of atherosclerotic disease[21]. The concordance rate is higher in monozygotic twins compared to dizygotic twins and the genetic effect is lower at higher age[21,22].

Atherosclerosis

Assessing the heritability of atherosclerosis is depending of the type of measurement, the population under investigation and the study design. In one population–based twin study heritability

* Trimethylglycine, a organic chemical that is used to treat hyperhomocysteinaemia. (http://en.wikipedia.org/wiki/Trimethylglycine)

was assessed using carotid IMT, a surrogate marker of early atherosclerosis[122]. Results indicated that carotid IMT was under influence of familial, rather than genetic factors. Heritability was estimates at 0.31, but non–significant (p = 0.15). In one study of 565 individuals from 154 families including an affected parent with CAD, three quantitative measures of atherosclerosis were assessed for heritability: IMT, plaque score, and maximal stenosis[123]. After correction for covariates the plaques score, defined as local thickening compared to the adjacent arterial wall (>50%), no significant heritability estimate could be found. Compared to maximal stenosis, IMT had the highest heritability, h^2 = 0.61 (p = 0.001) versus h^2 = 0.47 (p = 0.006). In another study heritability of atherosclerosis as assessed by carotid IMT was examined in 58 monozygotic (MZ) and 40 dizygotic (DZ) male twin pairs free overt CVD[124]. After correcting for covariates carotid IMT was estimated at 0.59. Coronary artery calcification (CAC) is a marker of coronary artery atherosclerosis and CAC progression is influenced by genetic factors, heritability estimate was 0.40 (p < 0.001)[125]. All in all heritability of atherosclerosis is estimated between 0.21 and 0.64, and was dependent on the method of measurement, and increased by the presence of cardiovascular risk factors and age[126].

Atherosclerosis is a quantitative trait

The WTCCC study is one of the most famous examples of a very successful case–control GWAS, despite its failure to identify variants associated with hypertensive cases. Genomic studies draw attention to the polygenic nature of many common, complex diseases, such as hypertension, T2DM, type 1 diabetes mellitus, and CAD. There is an apparent disconnection between the qualitative diagnosed disease (dichotomous trait, "the case") and their quantitatively distributed polygenic susceptibilities[19]. In psychiatric genomics the understanding that psychiatric disorders are qualitatively diagnosed diseases with a quantitative nature came early, because of the complexity to properly define cases in a measurable manner. Gottesman and Shields first described *endophenotypes* as "internal phenotypes measurable through biochemical tests or microscopic examinations"[127]. Endophenotypes are considered intermediate phenotypes, assisting in the identification "downstream" clinical traits and "upstream" genes involved in the polygenic nature of the psychiatric disorder[127]. They are the psychiatric equivalent of biomarkers and risk factors, and analogous to quantitative traits. Additionally some quantitative traits are surrogate biomarkers for disease, such as plasma cholesterol levels or carotid IMT.

The aggregate of atherosclerotic risk factors (BOX 1) and biomarkers (BOX 4) is informative of the development and progression of atherosclerosis, and the overall risk for the individual. In addition, many factors involved in atherogenesis specifically (*e.g.* thrombogenesis, calcium and lipid metabolism) and organismal function in general (*e.g.* cellular plasticity and longevity, hormonal metabolism, *epigenetics*, and behaviour), shape the eventual clinical outcome. Furthermore, recent evidence suggest a considerable *pleiotropic* overlap between genes involved in different aspects of human diseases and their associated traits and biomarkers[19]. For example several genes involved in autoimmune disorders such as rheumatoid arthritis (RA) or SLE are also involved in atherosclerosis[128-130].

Traditionally the atherosclerotic disease "pie" is cut along the lines of medical diagnostic thresholds. The extremes of the distribution are investigated, neglecting the intermediate phenotypes and focusing on identifying the abnormality in cases[19]. Considering the progressive nature of atherosclerotic disease, it appears to be appropriate to cut that "pie" along the lines of genotypes, quantitative traits and factors known to be involved in the diagnosed diseases. Atherosclerotic disease can be defined as the aggregate of atherosclerotic quantitative traits and the genomic variants indexed to associate with them, generating a *polygenic risk score* that can facilitate an individual assessment of risk for clinical outcome[19].

Atherosclerosis is a progressive disease that can be asymptomatic. In an ever ageing society it becomes more common, and so a quantitative approach becomes more powerful as the difference between diagnosed cases and controls becomes less clear[19]. Studying atherosclerotic disease in a quantitative manner in population cohorts would increase the knowledge of the aetiology, enabling biomarker and therapeutic target discovery. A fine example is the discovery of a robust association of a common variant (rs9939609) in *FTO* conferring a risk for diabetes through an effect on BMI[131]. Individuals homozygous for the minor allele had higher risk (OR 1.67) to develop obesity. The *FTO* variant not only associates with qualitative traits (obesity, diabetes), but also with quantitative traits (BMI, weight).

A different road

The Athero–Express Biobank

The AE started in 2002 and is an ongoing longitudinal study that includes patients undergoing CEA among others[33,132]. The study design was previously describe, but in short: atherosclerotic plaques obtained during CEA and blood taken prior to surgery are collected and processed. After surgery patients undergo follow-up, and all provided written informed consent. Additionally, an extensive questionnaire covering medical history, risk factors, and medication use is filled in, and several (standard) clinical laboratory and patient parameters are collected. The medical ethics boards of the participating centres, the St. Antonius Hospital Nieuwegein and the University Medical Centre Utrecht, have approved the study. The Athero–Express biobank studies (AE) were conceived on the idea that proteins in plaques are informative of systemic disease[33,132]. A recent paper reported the discovery of one such protein, osteopontin (OPN, Entrez Gene Symbol: SPP1, Entrez Gene ID: 6696), with a prognostic value for cardiovascular events supporting this hypothesis[166]. Additionally, high MRP14 expression correlates with rupture-prone lesions, and IL6 and IL8 expression strongly decreases after stroke[133,134].

Novel biomarker discoveries

GWAS of Stenosis

In the quest for novel biomarker GWAS have provided several possible new targets that previously were not known to be involved in atherosclerotic disease (BOX 3)[4]. Clearly GWAS have been successful, and there is no reason to assume that new GWAS will not uncover more genes involved atherosclerotic disease. Despite advances in primary prevention and drug treatment, stenosis of the carotid artery remains one of the chief indications to perform a carotid endarterectomy (CEA) in cardiovascular patients and the primary inclusion criterion of the AE. In terms of novel biomarker discoveries, two possible scenarios would be interesting. First, a GWAS associating stenosis prior to surgery could yield novel targets for drug development and preventive strategies. Second, a GWAS associating stenosis prior to surgery in relation to stenosis and adverse events after CEA could yield novel targets for secondary preventive strategies.

Replication is an issue because of the unique character of the AE. However, not only patients indicated for CEA are included, also patients undergoing restenosis treatment of the iliac and femoral arteries are included. Atherosclerotic disease is systemic and femoral artery plaques hold predictive value for systemic outcome. Therefore, “femoral” patients would be an ideal replication cohort. The advantage of performing a GWAS of stenosis is that it is a (semi-)quantitative trait measured as percentage (%) stenosis of the artery using duplex ultrasound technology following internationally accepted guidelines[135,136]. This renders it a QT study and negates the need for a control cohort such as in case-control study designs. One marker and multimarker regression tests accompanied with false-discovery rate correction can reduce the number of false-positives and the detection of spurious associations. If the GWAS is performed using the SNP5 with approximately 350,000 markers that passed QC and the p-value = 0.05 (α), than the associated p-value (β) would have to be $0.05/350{,}000 \approx 1.43\text{x}10^{-6}$ using Bonferroni correction. Although Bonferroni correction is overly conservative in case of a well defined study, the β is a factor 100 larger than genome–wide significance as defined by the WTCCC. When the study size would be approximately 750, the power to find common or rarer variants associated with stenosis with an OR of 2.0 would be 0.6-0.8[74].

This type of GWAS is would be unprecedented in its field, as the AE is the world's largest biobank containing samples from atherosclerotic patients. Such a GWAS could be useful in light of the continuing biomarker discovery efforts of the AE which are based on proteomics studies. Combining the results of these studies with such GWAS could not only give new insights into aetiology, but would certainly be of value in terms of prioritising targets for validation.

Proteomics & Genomics of Atherosclerosis

From recent GWAS a triangular relation between SNPs, biomarker, and disease is apparent. Schunkert *et al.* discussed how the significant association of CRP with atherosclerotic disease does not imply a causal relation[7]. Zacho *et al.* applied Mendelian randomization to determine a causal relation of CRP with CVD and no causality could be inferred[83]. Despite these results CRP remains on the forefront of inflammatory factors associated with CVD with a possible use as biomarker[3]. Extending the proteomic discoveries in the AE with genomic data could help in identifying valid targets for further research. Several biomarkers have been discovered by the AE, such as MRP14, MIF and OPN with varying predictive power in terms of cumulative survival rates, and estimations of specificity and sensitivity[134,137]. As Schunkert *et al.* have pointed out; a significant association of a particular novel biomarker does not mean causality[7]. It could be simply a case of reverse causality or confounding of unknown origins[7]. Yet, the predictive power of OPN is robust when compared to traditional factors used in risk algorithms, remains strong when measured in serum, and is capable of reclassifying patients into higher (or lower) risk groups[166]. A quantitative trait analysis based on Mendelian randomization of OPN expression in plaques can uncover variants associated with expression. Variants that subsequently associate with clinical outcome are of particular interest, as they indicate a causal relation. Additionally, these variants are useful in patient stratification based on genotypes (polygenic score) and OPN expression for secondary preventive measures.

Established Variants & Biomarker prediction

Contrary to those who apply polygenic scores on top of traditional risk factors[138], a Mendelian randomization approach seems more appropriate when the objective is to apply genomics in biomarker discovery and development for clinical purposes. This is not to say, that adding a polygenic score to traditional models is not useful in primary prevention and reclassification of high risk patients[139-142]. The variant on chromosome 9p21 might not be of influence in terms of absolute risk for the general population; it does reclassify patients with cardiovascular disease to a high risk group which could be beneficial when devising a treatment regime[140,141]. Additionally genomic information is inherently unchanging in life, and therefore a diamante in the rough when compared to expression based risk assessment factors[143].

To assess the value of novel serum– or tissue–derived biomarkers and genetic variants, combining both could be beneficial. Many genetic variants have been consistently replicated in various population cohorts, as have many putative biomarkers. Few have combined all identified variants and biomarkers for a particular trait in a Mendelian randomization experiment. The AE is a perfect setting for such an experiment. It has uncovered many possible biomarkers through proteomics, and holds biomaterial and clinical information of over 1500 well-annotated patients. Initial results demonstrate the power of the AE for previously discovered variants. Measuring putative biomarkers and associating known variants wit expression and outcome could underscore the usefulness of the biomarkers. Additionally, a quantitative trait analysis could uncover new variants associated with expression and secondary outcome.

Traditional case-control based GWAS have uncovered many variants associated with a variety of disease. None of these variants have made it to clinical setting. Additionally, many putative biomarkers have been identified, yet they up to now it have been difficult to prove their usefulness in the clinic. The GWAS have emphasized the polygenic nature of CVD and therefore it seems fitting to qualify atherosclerotic disease as a quantitative trait. The AE has uncovered many putative biomarkers, and is a valid study to validate a combined genotype/biomarker-based stratification strategy for secondary prevention. Such an integrative, systems biology approach could boost biomarker discovery, and be a potential short-cut to validating putative biomarkers for disease.

Acknowledgements

Kim E. Jie, MD. Rob Hurks, MD.

References

1. Duivenvoorden, R., de Groot, E., Stroes, E.S. & Kastelein, J.J. Surrogate markers in clinical trials--challenges and opportunities. *Atherosclerosis* **206**, 8-16 (2009).
2. Melander, O. et al. Novel and conventional biomarkers for prediction of incident cardiovascular events in the community. *Jama* **302**, 49-57 (2009).
3. Jaffe, A.S., Babuin, L. & Apple, F.S. Biomarkers in acute cardiac disease: the present and the future. *J Am Coll Cardiol* **48**, 1-11 (2006).
4. Gibbons, G.H. et al. Genetic markers: progress and potential for cardiovascular disease. *Circulation* **109**, IV47-58 (2004).
5. Vasan, R.S. Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation* **113**, 2335-62 (2006).
6. Novembre, J. et al. Genes mirror geography within Europe. *Nature* **456**, 98-101 (2008).
7. Schunkert, H. & Samani, N.J. Elevated C-reactive protein in atherosclerosis--chicken or egg? *N Engl J Med* **359**, 1953-5 (2008).
8. McCarthy, M.I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**, 356-69 (2008).
9. Pollex, R.L. & Hegele, R.A. Copy number variation in the human genome and its implications for cardiovascular disease. *Circulation* **115**, 3130-8 (2007).
10. Atella, V. et al. Bridging science and health policy in cardiovascular disease: focus on lipid management: A Report from a Session held during the 7th International Symposium on Multiple Risk Factors in Cardiovascular Diseases: Prevention and Intervention--Health Policy, in Venice, Italy, on 25 October, 2008. *Atheroscler Suppl* **10**, 3-21 (2009).
11. Mathers, C.D. & Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* **3**, e442 (2006).
12. Marma, A.K. & Lloyd-Jones, D.M. Systematic examination of the updated Framingham heart study general cardiovascular risk profile. *Circulation* **120**, 384-90 (2009).
13. Vasan, R.S. & Kannel, W.B. Strategies for cardiovascular risk assessment and prevention over the life course: progress amid imperfections. *Circulation* **120**, 360-3 (2009).
14. Frazer, K.A., Murray, S.S., Schork, N.J. & Topol, E.J. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* **10**, 241-51 (2009).
15. Akosah, K.O., Schaper, A., Cogbill, C. & Schoenfeld, P. Preventing myocardial infarction in the young adult in the first place: how do the National Cholesterol Education Panel III guidelines perform? *J Am Coll Cardiol* **41**, 1475-9 (2003).
16. Cheung, V.G. et al. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* **33**, 422-5 (2003).
17. Hindorff, L.A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
18. Hindorff LA, J.H., Mehta JP, and Manolio TA. A Catalog of Published Genome-Wide Association Studies. Vol. 2010 Hindorff LA, Junkins HA, Mehta JP, and Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed [date of access]. (2009).
19. Plomin, R., Haworth, C.M. & Davis, O.S. Common disorders are quantitative traits. *Nat Rev Genet* **10**, 872-8 (2009).
20. Brown, W.M. et al. Age-stratified heritability estimation in the Framingham Heart Study families. *BMC Genet* **4 Suppl 1**, S32 (2003).
21. Marenberg, M.E., Risch, N., Berkman, L.F., Floderus, B. & de Faire, U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med* **330**, 1041-6 (1994).
22. Nora, J.J., Lortscher, R.H., Spangler, R.D., Nora, A.H. & Kimberling, W.J. Genetic--epidemiologic study of early-onset ischemic heart disease. *Circulation* **61**, 503-8 (1980).
23. Insull, W., Jr. The pathology of atherosclerosis: plaque development and plaque responses to medical treatment. *Am J Med* **122**, S3-S14 (2009).
24. Rocha, V.Z. & Libby, P. Obesity, inflammation, and atherosclerosis. *Nat Rev Cardiol* **6**, 399-409 (2009).
25. Ross, R. Atherosclerosis--an inflammatory disease. *N Engl J Med* **340**, 115-26 (1999).
26. Virmani, R., Kolodgie, F.D., Burke, A.P., Farb, A. & Schwartz, S.M. Lessons from sudden coronary death: a comprehensive morphological classification scheme for atherosclerotic lesions. *Arterioscler Thromb Vasc Biol* **20**, 1262-75 (2000).
27. Mayerl, C. et al. Atherosclerosis research from past to present--on the track of two pathologists with opposing views, Carl von Rokitansky and Rudolf Virchow. *Virchows Arch* **449**, 96-103 (2006).
28. Shattock, S.G. A Report upon the Pathological Condition of the Aorta of King Menephtah, traditionally regarded as the Pharaoh of the Exodus. *Proc R Soc Med* **2**, 122-127 (1909).
29. Napoli, C. et al. Fatty streak formation occurs in human fetal aortas and is greatly enhanced by maternal hypercholesterolemia. Intimal accumulation of low density lipoprotein and its oxidation precede monocyte recruitment into early atherosclerotic lesions. *J Clin Invest* **100**, 2680-90 (1997).
30. Stary, H.C. et al. 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* **89**, 2462-78 (1994).
31. Lee, R.T. & Libby, P. The unstable atheroma. *Arterioscler Thromb Vasc Biol* **17**, 1859-67 (1997).
32. Spagnoli, L.G. et al. Extracranial thrombotically active carotid plaque as a risk factor for ischemic stroke. *Jama* **292**, 1845-52 (2004).
33. Hellings, W.E., Peeters, W., Moll, F.L. & Pasterkamp, G. From vulnerable plaque to vulnerable patient: the search for biomarkers of plaque destabilization. *Trends Cardiovasc Med* **17**, 162-71 (2007).
34. Chobanian, A.V. et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* **42**, 1206-52 (2003).
35. Papatheodorou, L. & Weiss, N. Vascular oxidant stress and inflammation in hyperhomocysteinemia. *Antioxid Redox Signal* **9**, 1941-58 (2007).

36. Toohey, J.I. Homocysteine toxicity in connective tissue: theories, old and new. *Connect Tissue Res* **49**, 57-61 (2008).
37. Hansson, G.K. Immune mechanisms in atherosclerosis. *Arterioscler Thromb Vasc Biol* **21**, 1876-90 (2001).
38. Hansson, G.K. et al. Localization of T lymphocytes and macrophages in fibrous and complicated human atherosclerotic plaques. *Atherosclerosis* **72**, 135-41 (1988).
39. Hansson, G.K., Libby, P., Schonbeck, U. & Yan, Z.Q. Innate and adaptive immunity in the pathogenesis of atherosclerosis. *Circ Res* **91**, 281-91 (2002).
40. Caligiuri, G., Nicoletti, A., Poirier, B. & Hansson, G.K. Protective immunity against atherosclerosis carried by B cells of hypercholesterolemic mice. *J Clin Invest* **109**, 745-53 (2002).
41. Lewis, M.J. et al. Immunoglobulin M is required for protection against atherosclerosis in low-density lipoprotein receptor-deficient mice. *Circulation* **120**, 417-26 (2009).
42. Liu, M.C. et al. Human lung macrophage-derived histamine-releasing activity is due to IgE-dependent factors. *J Immunol* **136**, 2588-95 (1986).
43. Sedgwick, J.D., Holt, P.G. & Turner, K.J. Production of a histamine-releasing lymphokine by antigen- or mitogen-stimulated human peripheral T cells. *Clin Exp Immunol* **45**, 409-18 (1981).
44. Kaartinen, M., Penttila, A. & Kovanen, P.T. Accumulation of activated mast cells in the shoulder region of human coronary atheroma, the predilection site of atheromatous rupture. *Circulation* **90**, 1669-78 (1994).
45. Kovanen, P.T., Kaartinen, M. & Paavonen, T. Infiltrates of activated mast cells at the site of coronary atheromatous erosion or rupture in myocardial infarction. *Circulation* **92**, 1084-8 (1995).
46. Hlatky, M.A. et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* **119**, 2408-16 (2009).
47. Cook, N.R. & Ridker, P.M. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med* **150**, 795-802 (2009).
48. Scirica, B.M. & Morrow, D.A. Is C-reactive protein an innocent bystander or proatherogenic culprit? The verdict is still out. *Circulation* **113**, 2128-34; discussion 2151 (2006).
49. Shah, S.H. & de Lemos, J.A. Biomarkers and cardiovascular disease: determining causality and quantifying contribution to risk assessment. *Jama* **302**, 92-3 (2009).
50. Elliott, P. et al. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *Jama* **302**, 37-48 (2009).
51. Liu, F.T. & Rabinovich, G.A. Galectins: regulators of acute and chronic inflammation. *Ann N Y Acad Sci* **1183**, 158-82 (2010).
52. Al-Ansari, S., Zeebregts, C.J., Slart, R.H., Peppelenbosch, M. & Tio, R.A. Galectins in atherosclerotic disease. *Trends Cardiovasc Med* **19**, 164-9 (2009).
53. Madero, M. & Sarnak, M.J. Association of cystatin C with adverse outcomes. *Curr Opin Nephrol Hypertens* **18**, 258-63 (2009).
54. Khakpour, H. & Frishman, W.H. Lipoprotein-associated phospholipase A2: an independent predictor of cardiovascular risk and a novel target for immunomodulation therapy. *Cardiol Rev* **17**, 222-9 (2009).
55. Koenig, W. & Khuseyinova, N. Lipoprotein-associated and secretory phospholipase A2 in cardiovascular disease: the epidemiological evidence. *Cardiovasc Drugs Ther* **23**, 85-92 (2009).
56. Lerman, A. & McConnell, J.P. Lipoprotein-associated phospholipase A2: a risk marker or a risk factor? *Am J Cardiol* **101**, 11F-22F (2008).
57. The International HapMap Project. *Nature* **426**, 789-96 (2003).
58. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nat Rev Genet* **7**, 85-97 (2006).
59. Sharp, A.J., Cheng, Z. & Eichler, E.E. Structural variation of the human genome. *Annu Rev Genomics Hum Genet* **7**, 407-42 (2006).
60. Wandeler, G. et al. Strength of Family History in Predicting Levels of Blood Pressure, Plasma Glucose and Cholesterol. *Public Health Genomics* **13**, 143-154 (2010).
61. Govindaraju, D.R. et al. Genetics of the Framingham Heart Study population. *Adv Genet* **62**, 33-65 (2008).
62. Nasir, K. et al. Family history of premature coronary heart disease and coronary artery calcification: Multi-Ethnic Study of Atherosclerosis (MESA). *Circulation* **116**, 619-26 (2007).
63. Michos, E.D. et al. Women with a low Framingham risk score and a family history of premature coronary heart disease have a high prevalence of subclinical coronary atherosclerosis. *Am Heart J* **150**, 1276-81 (2005).
64. Manolio, T.A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).
65. Visscher, P.M., Hill, W.G. & Wray, N.R. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* **9**, 255-66 (2008).
66. Damani, S.B. & Topol, E.J. Future use of genomics in coronary artery disease. *J Am Coll Cardiol* **50**, 1933-40 (2007).
67. Barrett, J.C. & Cardon, L.R. Evaluating coverage of genome-wide association studies. *Nat Genet* **38**, 659-62 (2006).
68. Freedman, M.L. et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet* **36**, 388-93 (2004).
69. Novembre, J. & Stephens, M. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* **40**, 646-9 (2008).
70. Tian, C., Gregersen, P.K. & Seldin, M.F. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* **17**, R143-50 (2008).
71. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
72. Schork, N.J., Murray, S.S., Frazer, K.A. & Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* **19**, 212-9 (2009).
73. Reich, D.E. & Lander, E.S. On the allelic spectrum of human disease. *Trends Genet* **17**, 502-10 (2001).
74. Spencer, C.C., Su, Z., Donnelly, P. & Marchini, J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet* **5**, e1000477 (2009).
75. Pe'er, I. et al. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet* **38**, 663-7 (2006).
76. de Bakker, P.I. et al. Efficiency and power in genetic association studies. *Nat Genet* **37**, 1217-23 (2005).
77. de Bakker, P.I. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* **17**, R122-8 (2008).
78. Ellinghaus, D., Schreiber, S., Franke, A. & Nothnagel, M. Current software for genotype imputation. *Hum Genomics* **3**, 371-80 (2009).
79. Nothnagel, M., Ellinghaus, D., Schreiber, S., Krawczak, M. & Franke, A. A comprehensive evaluation of SNP genotype imputation. *Hum Genet* **125**, 163-71 (2009).

80. Anderson, C.A. et al. Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. *Am J Hum Genet* **83**, 112-9 (2008).
81. Hao, K., Chudin, E., McElwee, J. & Schadt, E.E. Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. *BMC Genet* **10**, 27 (2009).
82. Miller, D.T., Ridker, P.M., Libby, P. & Kwiatkowski, D.J. Atherosclerosis: the path from genomics to therapeutics. *J Am Coll Cardiol* **49**, 1589-99 (2007).
83. Zacho, J. et al. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med* **359**, 1897-908 (2008).
84. Meola, N., Gennarino, V.A. & Banfi, S. microRNAs and genetic diseases. *Pathogenetics* **2**, 7 (2009).
85. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).
86. Selbach, M. et al. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008).
87. Sadee, W. Measuring cis-acting regulatory variants genome-wide: new insights into expression genetics and disease susceptibility. *Genome Med* **1**, 116 (2009).
88. Wray, G.A. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**, 206-16 (2007).
89. Goring, H.H. et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* **39**, 1208-16 (2007).
90. Melzer, D. et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* **4**, e1000072 (2008).
91. Wang, D. et al. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement. *Blood* **112**, 1013-21 (2008).
92. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
93. Garcia-Rios, A. et al. Genetic variations at ABCG5/G8 genes modulate plasma lipids concentrations in patients with familial hypercholesterolemia. *Atherosclerosis* (2010).
94. Abifadel, M. et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* **34**, 154-6 (2003).
95. Ono, S. et al. A promoter SNP (-1323T>C) in G-substrate gene (GSBS) correlates with hypercholesterolemia. *J Hum Genet* **48**, 447-50 (2003).
96. Fujita, Y. et al. Hypercholesterolemia associated with splice-junction variation of inter-alpha-trypsin inhibitor heavy chain 4 (ITIH4) gene. *J Hum Genet* **49**, 24-8 (2004).
97. Bhatnagar, D. Diagnosis and screening for familial hypercholesterolaemia: finding the patients, finding the genes. *Ann Clin Biochem* **43**, 441-56 (2006).
98. Lamberts, S.W. & Uitterlinden, A.G. Genetic testing in clinical practice. *Annu Rev Med* **60**, 431-42 (2009).
99. Inazu, A. et al. Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation. *N Engl J Med* **323**, 1234-8 (1990).
100. Aulchenko, Y.S. et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* **41**, 47-55 (2009).
101. Sibani, S. et al. Characterization of mutations in severe methylenetetrahydrofolate reductase deficiency reveals an FAD-responsive mutation. *Hum Mutat* **21**, 509-20 (2003).
102. Qian, X., Lu, Z., Tan, M., Liu, H. & Lu, D. A meta-analysis of association between C677T polymorphism in the methylenetetrahydrofolate reductase gene and hypertension. *Eur J Hum Genet* **15**, 1239-45 (2007).
103. Casas, J.P., Hingorani, A.D., Bautista, L.E. & Sharma, P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch Neurol* **61**, 1652-61 (2004).
104. Kelly, P.J. et al. Homocysteine, MTHFR 677C-->T polymorphism, and risk of ischemic stroke: results of a meta-analysis. *Neurology* **59**, 529-36 (2002).
105. Klerk, M. et al. MTHFR 677C-->T polymorphism and risk of coronary heart disease: a meta-analysis. *Jama* **288**, 2023-31 (2002).
106. Hustad, S. et al. The methylenetetrahydrofolate reductase 677C-->T polymorphism as a modulator of a B vitamin network with major effects on homocysteine metabolism. *Am J Hum Genet* **80**, 846-55 (2007).
107. Holm, P.I. et al. Modulation of the homocysteine-betaine relationship by methylenetetrahydrofolate reductase 677 C->t genotypes and B-vitamin status in a large-scale epidemiological study. *J Clin Endocrinol Metab* **92**, 1535-41 (2007).
108. Ioannidis, J.P., Ntzani, E.E. & Trikalinos, T.A. 'Racial' differences in genetic effects for complex diseases. *Nat Genet* **36**, 1312-8 (2004).
109. Ritsch, A. et al. Scavenger receptor class B type I polymorphisms and peripheral arterial disease. *Metabolism* **56**, 1135-41 (2007).
110. Acton, S. et al. Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population. *Arterioscler Thromb Vasc Biol* **19**, 1734-43 (1999).
111. Naj, A.C. et al. Association of scavenger receptor class B type I polymorphisms with subclinical atherosclerosis: the Multi-Ethnic Study of Atherosclerosis. *Circ Cardiovasc Genet* **3**, 47-52 (2010).
112. Vogiatzi, K. et al. Interleukin 8 gene polymorphisms and susceptibility to restenosis after percutaneous coronary intervention. *J Thromb Thrombolysis* **29**, 134-40 (2010).
113. Heiskanen, M. et al. Polymorphism in the IL10 promoter region and early markers of atherosclerosis: the Cardiovascular Risk in Young Finns Study. *Atherosclerosis* **208**, 190-6 (2010).
114. Guzman-Guzman, I.P. et al. Interleukin-6 gene promoter polymorphisms and cardiovascular risk factors. A family study. *Dis Markers* **28**, 29-36 (2010).
115. Vogiatzi, K. et al. Genetic diversity of RANTES gene promoter and susceptibility to coronary artery disease and restenosis after percutaneous coronary intervention. *Thromb Res* **124**, 84-9 (2009).
116. Rissanen, A.M. Familial occurrence of coronary heart disease: effect of age at diagnosis. *Am J Cardiol* **44**, 60-6 (1979).
117. Andresdottir, M.B., Sigurdsson, G., Sigvaldason, H. & Gudnason, V. Fifteen percent of myocardial infarctions and coronary revascularizations explained by family history unrelated to conventional risk factors. The Reykjavik Cohort Study. *Eur Heart J* **23**, 1655-63 (2002).
118. Sharp, S.D., Williams, R.R., Hunt, S.C. & Schumacher, M.C. Coronary risk factors and the severity of angiographic coronary artery disease in members of high-risk pedigrees. *Am Heart J* **123**, 279-85 (1992).
119. Fox, C.S. et al. Genetic and environmental contributions to atherosclerosis phenotypes in men and women: heritability of carotid intima-media thickness in the Framingham Heart Study. *Stroke* **34**, 397-401 (2003).
120. Bensen, J.T., Li, R., Hutchinson, R.G., Province, M.A. & Tyroler, H.A. Family history of coronary heart disease and pre-clinical carotid artery atherosclerosis in African-Americans and whites: the ARIC study: Atherosclerosis Risk in Communities. *Genet Epidemiol* **16**, 165-78 (1999).
121. Scheuner, M.T. Genetic predisposition to coronary artery disease. *Curr Opin Cardiol* **16**, 251-60 (2001).

122. Swan, L., Birnie, D.H., Inglis, G., Connell, J.M. & Hillis, W.S. The determination of carotid intima medial thickness in adults--a population-based twin study. *Atherosclerosis* **166**, 137-41 (2003).
123. Moskau, S. et al. Heritability of carotid artery atherosclerotic lesions: an ultrasound study in 154 families. *Stroke* **36**, 5-8 (2005).
124. Zhao, J. et al. Heritability of carotid intima-media thickness: a twin study. *Atherosclerosis* **197**, 814-20 (2008).
125. Cassidy-Bushrow, A.E. et al. Coronary artery calcification progression is heritable. *Circulation* **116**, 25-31 (2007).
126. Gaeta, G. Heritability of atherosclerosis. *Am J Hypertens* **20**, 255 (2007).
127. Gottesman, II & Gould, T.D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* **160**, 636-45 (2003).
128. Shi, G.P. Immunomodulation of vascular diseases: atherosclerosis and autoimmunity. *Eur J Vasc Endovasc Surg* **39**, 485-94 (2010).
129. van Zonneveld, A.J., de Boer, H.C., van der Veer, E.P. & Rabelink, T.J. Inflammation, vascular injury and repair in rheumatoid arthritis. *Ann Rheum Dis* **69 Suppl 1**, i57-60 (2010).
130. Sherer, Y., Zinger, H. & Shoenfeld, Y. Atherosclerosis in systemic lupus erythematosus. *Autoimmunity* **43**, 98-102 (2010).
131. Frayling, T.M. et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-94 (2007).
132. Verhoeven, B.A. et al. Athero-express: differential atherosclerotic plaque expression of mRNA and protein in relation to cardiovascular events and patient characteristics. Rationale and design. *Eur J Epidemiol* **19**, 1127-33 (2004).
133. Peeters, W. et al. Carotid atherosclerotic plaques stabilize after stroke: insights into the natural process of atherosclerotic plaque stabilization. *Arterioscler Thromb Vasc Biol* **29**, 128-33 (2009).
134. Ionita, M.G. et al. High levels of myeloid-related protein 14 in human atherosclerotic plaques correlate with the characteristics of rupture-prone lesions. *Arterioscler Thromb Vasc Biol* **29**, 1220-7 (2009).
135. Grant, E.G. et al. Carotid artery stenosis: grayscale and Doppler ultrasound diagnosis--Society of Radiologists in Ultrasound consensus conference. *Ultrasound Q* **19**, 190-8 (2003).
136. Grant, E.G. et al. Carotid artery stenosis: gray-scale and Doppler US diagnosis--Society of Radiologists in Ultrasound Consensus Conference. *Radiology* **229**, 340-6 (2003).
137. de Kleijn, D.P. et al. Local Atherosclerotic Plaques Are a Source of Prognostic Biomarkers for Adverse Cardiovascular Events. *Arterioscler Thromb Vasc Biol* (2009).
138. Janssens, A.C. & van Duijn, C.M. Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet* **17**, R166-73 (2008).
139. deCODE Genetics, I. Announcing the New deCODE. Vol. 2010 (deCODE Genetics, Inc., Reykjavik, Iceland, 2010).
140. Gulcher, J. & Stefansson, K. Questions about genetic variation in 9p21 as a predictor of cardiovascular risk. *Ann Intern Med* **150**, 736; author reply 737 (2009).
141. Gulcher, J. & Stefansson, K. Genetic risk information for common diseases may indeed be already useful for prevention and early detection. *Eur J Clin Invest* **40**, 56-63 (2010).
142. Kaiser, J. Biotechnology. Bankruptcy won't stop deCODE, says its founder, Stefansson. *Science* **326**, 1172 (2009).
143. Wray, N.R., Goddard, M.E. & Visscher, P.M. Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev* **18**, 257-63 (2008).