Federal Aviation Administration

DOT/FAA/AM-15/10
Office of Aerospace Medicine
Washington, DC 20591

Aviation Relevance of Genetic Risk Scores for Cardiovascular Disease

Dennis Burian
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

July 2015

Final Report

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Federal Aviation Administration website.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-15/10		

4. Title and Subtitle	5. Report Date
Aviation Relevance of Genetic Risk Scores for Cardiovascular Disease	July 2015
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Burian D	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.
12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

16. Abstract

Summary

A review of the literature for single nucleotide polymorphisms associated with cardiovascular disease revealed about 30 high-confidence single nucleotide polymorphisms. Individually, none is predictive for cardiovascular disease. Several authors combined some number of these single nucleotide polymorphisms into a genetic risk score modeled on the Framingham Risk Score that is based on traditional risk factors, including lipid profile, hypertension, age, family history, diabetes, and smoking status.

All genetic risk scores were associated with cardiovascular disease and had a similar predictive value as traditional risk factors; however, when genetic risk scores were combined with predictive scores based on traditional risk factors, there was only a minimal increase in predictive value.

Conclusions

Genetic profile discovery to augment the current assessment paradigm is not recommended because traditional risk factors-based scores have similar predictive power, are less expensive, and are more easily interpreted by the medical community.

17. Key Words	18. Distribution Statement
Cardiovascular Disease, Traditional Risk Factors, Genome Wide Association Study, Genetic Risk Score	Document is available to the public through the Internet: www.faa.gov/go/oamtechreports

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	18	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DEFINITIONS AND ABBREVIATIONS

Physiological Factors

AF -- -- - Atrial Fibrillation

BP -- -- - Blood Pressure

CAC -- - Coronary Arterial Calcium

CAD -- - Coronary Artery Disease; clinically considered interchangeable with Atherosclerosis

CHD -- - Coronary Heart Disease; a result of CAD. In this review, not used interchangeably with CAD unless a referenced paper has conflated the two.

CHF -- - Cardiac Heart Failure

CVD -- - Cardiovascular Disease; usually refers to heart disease and stroke combined; however, some investigators appear to conflate CVD and CHD

DBP -- - Diastolic Blood Pressure

HbA1c Glycated Hemoglobin, a marker for plasma glucose level

HDL -- - High-density Lipoprotein

LDL -- - Low-density Lipoprotein

MI -- -- - Myocardial Infarction

SBP - -- - Systolic Blood Pressure

TC -- -- - Total Cholesterol

TRF -- - Traditional Risk Factor; typically, some combination of age, gender, hypertension (SBP, DBP), lipid profile (TC, HDL cholesterol, LDL cholesterol, triglycerides), diabetes, smoking status, and/or family history

Genetic Terminology

eQTL-- - expression Quantitative Trait Locus(Loci); a genetic locus associated with the expression of a gene

FRS - -- - Framingham Risk Score, a risk score developed from the Framingham Study as a 10-year predictor of heart disease

GRS -- - Genetic Risk Score; a single value score developed by combining individual genetic risk factors

kB -- -- - kilobases, 1,000 bases (nucleotides), a unit of measurement for the distance between two loci

LD -- -- - Linkage Disequilibrium; in practice, two loci in strong LD remain associated during meiotic recombination due to relatively close proximity and residence on the same "linkage block"

SNP -- - Single Nucleotide Polymorphism; a single position difference in the nucleotide sequence from the reference genome. The reference genome sequence is a haploid representation built on the majority consensus from multiple sources

Statistical Terms and Measures

AUC -- - Area Under the Curve; the integral of a plot to assess predictive value, 0.5 is random chance

IDI-- -- - Integrated Discrimination Improvement score; a measure of a test's ability to discriminate between two disease states

HR-- -- - Hazard Ratio; a ratiometric to assess the risk of an event associated with a factor(s), 1 is no increased risk, 2 is doubled risk

NRI- -- - Net Reclassification Index; the percentage of people for whom a change in an assessment methodology changes their risk class

OR-- -- - Odds Ratio; a ratiometric to assess the association of two factors, 1 is no association

Other

AHA -- - American Heart Association

AME -- - Aviation Medical Examiner

yoa -- -- - Years of Age

Sources for Referenced Studies

ARIC-- -- -- -- -- -- -- -- -- -- -- -- Atherosclerosis Risk in Communities: https://www2.cscc.unc.edu/aric/desc

ATVB - -- -- -- -- -- -- -- -- -- -- -- Atherosclerosis, Thrombosis, and Vascular Biology Italian Study Group 2003

CARDIoGRAM -- -- -- -- -- -- -- Coronary Artery Disease Genome-wide Replication and Meta-Analysis http://www.cardiogramplusc4d.org/downloads/

FINRISK -- -- -- -- -- -- -- -- -- -- -- http://www.nationalbiobanks.fi/index.php/studies2/7-finrisk

Framingham -- -- -- -- -- -- -- -- -- Framingham Heart Study: https://www.framinghamheartstudy.org/

JNC 7 - -- -- -- -- -- -- -- -- -- -- -- Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; http://www.nhlbi.nih.gov/files/docs/guidelines/express.pdf

Malmo Diet and Cancer Study -- http://snd.gu.se/catalogue/study/610

MORGAM -- -- -- -- -- -- -- -- -- Monica Risk, Genetics, Archiving and Monograph project http://www.thl.fi/morgam/

PROCAM-- -- -- -- -- -- -- -- -- -- PROspective CArdiovascular Munsster study http://en.assmann-stiftung.de/procam/procam-study/

REGICOR- -- -- -- -- -- -- -- -- -- Registre Gironi del Cor (Italian) http://www.regicor.org/

SCORE-- -- -- -- -- -- -- -- -- -- -- Standard Care vs Corticosteroid for Retinal Vein Occlusion study, National Eye Institute: https://web.emmes.com/study/score/

WTCCC -- -- -- -- -- -- -- -- -- -- Wellcome Trust Case Control Consortium: http://www.wtccc.org.uk/

Aviation Relevance of Genetic Risk Scores for Cardiovascular Disease

Introduction 1
What Role Can Genetics Play in Assessing Risk? 1
Non-Modifiable Risk Factors 2
Modifiable Risk Factors 2
Hypertension 2
Lipid Profile 3
Discovery of SNPs Associated With Coronary Heart Disease 3
Development of Genetic Risk Scores for Coronary Heart Disease 4
Association of Single Nucleotide Polymorphisms and Gene Expression in Heart Disease 5
Conclusions 6
Tables 1 and 2 7
References 11

Aviation Relevance of Genetic Risk Scores for Cardiovascular Disease

INTRODUCTION

Heart disease is the leading cause of death in the United States, accounting for 24% of the nearly 2.5 million deaths in 2010.[1] Males had a 1.6-fold greater death rate than females, and African Americans had a 1.3-fold higher death rate than Caucasians. Fortunately, the death rate from heart disease has steadily decreased on a log-linear scale from ~600/100,000 in 1958 to ~200/100,000 in 2010. As of 2014,[2] the greatest percentage (40.6%) of heart disease and stroke-related deaths was linked to hypertension, the remainder being linked to smoking (13.7%), poor diet (13.2%), poor exercise habits (11.9%), and abnormal blood glucose levels (8.8%).

Addressing the demographics of most interest to the aviation community, Americans over 19 years of age had a 13.8% prevalence of total cholesterol over 240 mg/dL, 8.3% prevalence of diabetes, and a 33.0% prevalence of hypertension, of which 53% of those who are documented are able to control it. Of the 2,468,435 deaths in 2010, 24.2%, or 597,689, were from heart disease—a rate of 1,638 per day.[3]

CHD results from the accumulation of plaque in the coronary arteries leading to inefficient blood flow and ischemia. The plaque accumulation, itself, is defined as atherosclerosis, or CAD. Angina, or chest pain, can result from low oxygenated blood flow to the heart and usually is a symptom of CHD. Risk factors for CHD include high cholesterol, high blood pressure, diabetes, overweight or obesity, smoking, low physical activity levels, diet, and stress.

The Federal Aviation Administration medical certification process entails an assessment of the heart health of pilots according to commonly accepted TRFs by an Aviation Medical Examiner (AME) (Note: Under FAA guidance, hypertension is defined as SBP/DBP of 155/95, not the AHA standard of 140/90). Certificate issuance involves determining factors that could result in sudden incapacitation of a pilot, then subsetting the factor by qualifying conditions and the license class being issued (Table 1). A decision tree leads to either a decision to issue a medical certificate by the AME, a request for additional testing, or referral of the exam to the Aerospace Medical Certification Division at the Civil Aerospace Medical Institute. To maintain the safest possible airspace, it is crucial that everyone involved in the medical certification process maintain currency by incorporating the latest research and technologies in assessing medical risk. The purpose of this review is to describe the genesis of GRSs associated with heart disease and determine if including GRSs for heart disease would be beneficial in determining risk of an in-flight cardiac event.

WHAT ROLE CAN GENETICS PLAY IN ASSESSING RISK?

Evidence for a genetic underpinning to CardioVascular Disease (CVD) lies in the well documented increased risk within families. Overall, 12.6% of all adults over 19 report a parent or sibling who has suffered from a heart attack or angina before the age of 50. A paternal history of heart attack doubles the risk in men, increases the risk in women by 70%, and an individual of either gender for whom a sibling has had a heart attack is at 50% increased risk.[2] The Odds Ratio (OR) of Myocardial Infarction (MI) range from 1.67 if one parent has had an MI beyond 50 yoa to 6.56 if both parents have an event under 50 yoa.

Genetic marker discovery can be either by linkage analysis or a Genome Wide Association Study (GWAS). Linkage analysis determines the physical distance between two genes, each with a known phenotype, based on the frequency of meiotic cross-over, for example the pioneering work of Morgan using sex and eye color in Drosophila. Linkage studies are performed within related individuals to hone in on an altered genetic locus within affected family members. Meiotic cross-over events between two loci are tested between affected and unaffected individuals. However, the disease-related loci need to be relatively close together on the same chromosome to be detected; the further apart two genes are on a chromosome, the more likely multiple cross-over events are to occur. Typically, variants detected by linkage are rare but have large phenotypic impact.

A GWAS tests the occurrence of a phenotype against a single chromosomal location, usually a Single Nucleotide Polymorphism (SNP) but occasionally a Variable Nucleotide Repeat, in a case-control study design. SNPs are common and easily measured by either microarrays, quantitative polymerase chain reaction (qPCR), or more recently, high-throughput whole genome sequencing. A genetic profile is generated for the phenotype-based cohort groups, controls and patients, for the factor of interest. This method benefits from genotype permanence in individuals allowing samples and data to be collected prospectively. After some period of years allowing time for the phenotype of interest to occur, e.g., CAD or stroke, the analysis can be performed on the cohort groups. Allowing additional time to pass leverages the initial subject recruitment to an individual's lifetime. GWASs usually result in tens to hundreds of SNPs that each account for a small percentage of the incidence of the factor. Due to the imprecision of GWAS SNPs to individually predict disease, additional predictive power is gained by combining SNPs with the greatest phenotypic association to derive a GRS. This strategy is similar to the development of disease Risk Scores from physiology metrics, e.g., the FRS for cardiovascular disease. The significance of

the association between an SNP and disease is complicated by the necessary application of multiple testing correction when assessing half a million to a million individual SNPs. Commonly, a threshold p-value of 5×10^{-8} has been adopted for meta-analysis of GWASs.[4]

Heritability, the proportion of observed variance for a trait that can be attributed to genetics, varies widely for the various CVD TRFs (refer to Go et al., Table 7-3 [2]). For example, from the Framingham Heart Study, lipid profile heritability ranges from 0.48 for triglycerides to 0.59 for LDL cholesterol, which has the highest and most-consistent heritability scores across all lipid profile metrics. However, heritability for fasting glucose is 0.34 and for HbA1c, 0.27. A Danish twin study estimated the heritability of mortality from all heart diseases at 0.55. [5] Taken together, while heart disease in general and the TRFs, specifically, have a genetic component, environment and lifestyle play a significant but difficult to quantify role in the disease process. Several investigators have investigated the genetics of CVD independent of TRFs (see below).

The GWAS paradigm suffers from accuracy-reducing shortcomings. Making the connection between the GWAS-discovered SNP (index SNP) and the mutation (or mutations) that leads to the phenotype must be based on accurate functional understanding of the gene(s) affected by the index SNP and its interaction network. Sequence analysis of mutations in neighboring genes assist in making this determination, especially where the gene mutation results in change in the protein's amino acid sequence. However, distant mutations in promoters or enhancers can alter regulatory protein affinity for the regulatory regions, thereby affecting expression levels of the gene, and be difficult to detect without in-depth experimentation. Less likely, given the current state of knowledge of the human genome, an unknown gene may reside in the region. Also, the mode of action of a mutation may be unknown. A mutation can be dominant, recessive, or additive. Although many GRS calculations assume additive activity, in the absence of other data, there is evidence for this being a relatively safe assumption. Nevertheless, the assumption of additive activity overstates the importance of a recessive and underestimates a dominant mutation.

Gene-by-gene and gene-by-environment interactions further impact phenotype as do epigenetic marks that alter chromatin accessibility to expression-regulating DNA-binding proteins including histones. This effect can be chromosome-specific, such that the presence of a SNP risk allele may be moot if the gene is silenced on that chromosome. As genomic knowledge and methods have expanded, a few recent studies have explored the co-occurrence of disease-associated SNPs with expression data to suggest expression quantitative trait loci (eQTL) for phenotypes of interest, including CVD (see below).

Non-Modifiable Risk Factors

Age, gender, race, and family history are considered as non-modifiable risk factors for heart disease. Several global risk scores for Coronary or Cardiovascular risk have been developed (reviewed in [6]): risk scores from Framingham and SCORE include gender; the PROCAM and Reynolds scores are gender-specific. Age is a risk factor in all but race is not, despite common acceptance of increased association of cardiovascular events with African-American heritage. Family history is a factor in the PROCAM and Reynolds methods.

Modifiable Risk Factors

Modifiable risk factors can be altered or controlled by personal choices of diet, exercise, and lifestyle. In addition to smoking status, metrics or markers for hypertension, lipid profile, and diabetes are incorporated in all five global risk scores (see above). Systolic blood pressure is the chosen metric for hypertension in all risk assessments; Framingham uniquely adds use of hypertensive medication. Total cholesterol and HDL cholesterol are used in the Framingham, PROCAM, and Reynolds, [7,8] whereas SCORE uses the ratio of the two. LDL cholesterol and triglycerides are a factor in only the PROCAM. PROCAM and the female-specific Reynolds score use diabetes, as assessed by fasting blood glucose, but the female Reynolds adds HbA1c as an additional metric for diabetes. Finally, both male and female Reynolds scores utilize high-sensitivity C-reactive protein.

Hypertension

Blood pressure naturally increases with age due to loss of elasticity of the large arteries. There is a consistent, predictive, and etiologically significant relationship with many factors but, because many people have BPs below clinically recommended levels for pharmaceutical intervention, lifestyle changes are recommended to reduce BP, even in non-hypertensive individuals. Meta-analysis of 23 studies showed that decreasing hypertension reduces the risk of stroke by 32% [9]; the method of BP reduction was immaterial and universal in achieving this reduction. Goal BP in non-diabetics, according to the JNC 7 report, is 140/90,[10] and there is evidence that further decreases in BP reduce stroke risk a further 23%.[11]

Twenty-nine SNPs identified from multiple GWASs were analyzed across large multi-ethnic populations to confirm their association with hypertension[12] (Table 2). Non-European ethnicity decreased the number of significant SNPs to nine in East Asians and six in South Asians. However, all 29 SNPs were confirmed at highly significant levels in European ancestry for concordant association with SBP, DBP, and hypertension. A GRS developed from these 29 SNPs was confirmed to be associated with DBP and SBP in a study of 17,688 people enrolled in the Malmo Preventive Project.[13] This GRS was associated with increased SBP and DBP at presentation and with greater than expected increases upon reinvestigation after at least 10 years. However, the AUC for TRFs alone was not significantly increased by inclusion of the GRS. The regions surrounding these SNPs were investigated for genes most likely to regulate blood pressure.[12] Among them were genes encoding two forms of natriuretic peptide, ANP and BNP, and their clearance receptor, NPR-C, where increased levels of the peptides are

known to decrease blood pressure as does the adrenomedullin (ADM) gene also identified in this study. Missense mutations in the metal-ion transporter, HFE, are known to play a role in hemochromatosis; the zinc transporter SLC39A8 was also identified.

Recently, a GRS was developed from SNPs associated with blood pressure and tested for predictive power against cardiovascular events.[14] In addition to nearly 28,000 subjects from four of the FINRISK cohorts, more than 9,000 Finnish subjects from three other cohorts were assessed over an average of 9.8 years. CHD endpoints were non-fatal MI, angina, coronary revascularization, or CHD-related death. From five GWASs, 32 SNPs were chosen to construct separate GRSs for SBP and DBP. Each SNP was weighted for effect size, based on data from the discovery GWAS applied to the copy number of the risk associated allele, 0, 1, or 2. Statistically significant increases in HR were found for both GRSs, with their respective BP (SBP or DBP) and hypertension overall, but there was not a significant increase in AUC or NRI when the GRSs were applied to CHD in conjunction with the Framingham Risk Score based on TRFs. A confounder is that blood pressure, which changes with age, was measured at a single time point for each subject at variable ages.

Interestingly, when all highly significant SNPs for CAD are considered, only two arose from GWASs where hypertension was the endpoint,[15] rs12413409 and rs3184504 (Table 2). The first is located in an intron of the magnesium metabolism gene, CNNM2 on chromosome 10q24.32, the second encodes a missense mutation of the signaling adaptor SH2B3.

Lipid Profile

Of 31 high-confidence SNPs, 23 are associated with CAD independent of traditional risk factors, whereas eight are associated with cholesterol profile by GWAS.[15] Putatively SNP-affected genes include the LDL receptor, a missense mutation in the lipoprotein A coding sequence, and the ABO blood-group locus that is MI-associated and suspected in thrombosis. Potentially affected genes where index SNPs are extragenic include the apolipoprotein gene, APOA5, the receptor degradation mediating PCSK9, the microtubule regulator PSRC1, and the relatively uncharacterized TRIB1. An intronic index SNP is located in the sterol transporter ABCG8, which lies in close proximity to ABCG5 that has a similar function.

A subsequent meta-analysis of two GWASs for association with LDL cholesterol, HDL cholesterol, triglycerides, and total cholesterol annotated 157 high-confidence loci across nearly 190,000 individuals of multiple ethnicities.[16] A total of 62 new SNPs were identified; however, trait variance values were low: 1.6% for the 24 SNPs associated with HDL, 2.1% for the eight associated with triglycerides, 2.4% for the 15 associated with LDL, and 2.6% for the 15 associated with total cholesterol.

Given these results and the moderate degree of variance captured by the SNPs replicated by this study from previous studies, it is not surprising that when significance levels were loosened in this analysis, strong directional concordance was found with HDL, LDL, triglycerides, and total cholesterol. A literature search using genes within 100 kB of new index SNPs identified candidate lipid metabolism related genes in 52% of these SNPs. Fifteen of the 62 novel SNPs were associated with expression of a nearby eQTL, but remarkably, 52 of the 62 novel index SNPs were associated with chromatin marks for active regulatory DNA regions including enhancers and promoters. The Global Lipids Genetics Study[14] also confirmed that SNPs uniquely associated with increased LDL, triglyceride, or total cholesterol levels, but not decreased HDL, were strongly predictive for increased risk of CAD. These authors further demonstrated that for the 149 SNPs for which there were CAD data, LDL, triglyceride, and total cholesterol were correlated with an effect on CAD. Taken together, these results suggest: 1) there are additional loci involved in lipid metabolism, 2) these additional loci may be associated with the identified high-confidence SNPs or are more weakly associated but capture a significant level of the variance in lipid profiles, and 3) that intergenic regions, including enhancers and promoters, play a significant role in lipid metabolism. Not surprisingly, in addition to CAD, lipid profile SNPs were also strongly associated with type 2 diabetes, body mass index, and systolic and diastolic blood pressure.

Discovery of SNPs Associated With Coronary Heart Disease

The single most replicated genomic region associated with heart disease is 9p21,[17-19] first associated in the wide-ranging WTCCC study.[19] This region encodes the cell-cycle regulatory genes, CDKN2A and CDKN2B, but the most strongly CAD-associated SNPs are in a region encoding a large non-coding, alternatively-spliced RNA variously known as ANRIL[20] or CDKN2B-AS1, whose expression has been demonstrated in disease epithelium[21] and other CAD-associated cell-types[21,22] and is strongly associated with coronary artery disease.[21,22] Expression of CDKN2A, CDKN2B, and CDKN2B-AS1 are positively correlated,[20,23] suggesting coregulation; however, homozygosity of CAD risk alleles increased expression of ANRIL overall, especially the short splice forms, decreased CDKN2A and CDKN2B expression, and was weakly correlated to symptomatic patients without incident MI.[22] Individuals heterozygous for the risk allele had a HR for incident CVD of 1.25, the HR for homozygosity was 1.32,24 but factoring 9p21 status into a model based on TRFs did not improve risk prediction, NRI, or IDI, demonstrating that inclusion of a single genetic locus is insufficient to improve risk assessment for CVD.

In a GWAS designed to discover alterations in blood cell type levels,[25] an SNP was identified on 12q24 that associated with a general increase in blood cell numbers and showed a moderately significant (p=0.002) association with MI. This SNP is located in the adaptor protein gene SH2B3 and had been previously weakly associated with coronary artery disease.[19] Two other weakly CAD-associated genes from the WTCCC study were MTHFD1L in the mitochondrial tetrahydrofolate synthesis pathway, and the putative metallopeptidase, ADAMTS.[17]

In Finland, two non-overlapping case-controlled cohorts with available 10 or seven-year follow-up outcome data were accessed to determine the association of 27 genes with CHD, stroke, CVD, or total mortality and determine gender-specificity.[26] SNPs in angiotensin receptor (AGTR1; recessive model), APOE (additive and dominant models), carboxypeptidase B2 (CPB2; additive and recessive models), and coagulation factor XII (F12; additive model) were associated with CHD risk. F12 was strongly associated in both genders combined and individually, whereas the remaining markers were gender-specific. Increased risk of CHD was associated with SNPs in the hyperlipidemia-associated transcription factor USF1 (multiplicative model), CPB2 (dominant model), and Factor XIII, polypeptide A1 (F13A1; dominant model). A male-specific allele was found in the APOBEC2 gene.

Using the WTCCC, and German MI case control studies, Samani et al.[27] replicated known risk regions of 9p21.3, 6q25.1, and 2q36.3, and identified four additional genomic risk regions: 1p13.3, 1q41, 10q11.21, and 15q22.33 (Table 2). Endpoints for these studies were CAD and MI, respectively. The Myocardial Infarction Genetics Consortium[28] utilized subject populations from Europe and the USA in a case-controlled study using early-onset MI as the endpoint. Study populations were the Italian ATVB, Heart Attack in Puget Sound, REGICOR (Spain), MGH Premature Coronary Artery Disease Study (USA), FINRISK (Finland), and Malmo Diet and Cancer Study (Sweden). They confirmed Samani's risk-associated SNPs on 9p21, 1p13, 1q41, and 10q11, but not the SNPs on 6q25, 15q22, and 2q36. However, they identified risk-associated SNPs at 21q22, 6p24, 2q33, 19p13, and 1p32.

Development of Genetic Risk Scores for Coronary Heart Disease

Numerous GWAS studies of CHD have led to several GRS models. The ARIC study enrolled 15,792 subjects between 1987 and 1989 from four communities: Forsyth County, North Carolina; Minneapolis, Minnesota; Washington County, Maryland; and African Americans only from Jackson, Mississippi. This population has been sampled for various reasons, including GWASs.[29] Morrison et al.[30] combined a GWAS for severe CAD and MI with two MI GWASs. They surveyed the combined dataset against 14,000 SNPs for non-synonymous amino acid substitutions, RNA splice sites, or gene regulatory regions to arrive at 92 well-validated SNPs. They then developed the ARIC Cardiovascular Risk Score (ACRS) based on age, SBP, use of hypertensive medication, total cholesterol, HDL cholesterol, gender, diabetes, and smoking status, and determined the increase in predictive power of the ACRS when a race-specific GRS was added to the ACRS over a 13-year followup endpoint. The GRS for Caucasians included 10 SNPs, whereas the African-American GRS included 11; only a SNP in the KIF6 gene was in common between the GRSs. The GRS HR was 1.10 (95% CI 1.06 to 1.14) for Caucasians and 1.2 (1.11 to 1.29) for African Americans. Bootstrap bias-corrected values were 1.05 and 1.09, respectively. When applied to ROC curves, the GRS improved the AUC of the ACRS to 0.766 from 0.764, whereas the AUC for African Americans improved to 0.769 from 0.758, thus was significant only in the African American population.

Subsequently, investigators at Celera[31] narrowed the eligible SNPs from the ACRS (above) to those associated with CHD in ARIC and at least two other GWASs, and chose a cut-off to assess predictive significance of their GRS as opposed to treating it as a continuous variable. Study population (ARIC), TRFs, clinical endpoints, and the time span of the study were unchanged. SNPs associated with five genes met their screening criteria, MYH15, KIF6, VAMP8, PALLD, and SNX19 (Table 2).

Only homozygotes were counted in this GRS; risk alleles were +1, non-risk alleles were counted as -1 leading to a maximum risk-associated score of +5; +3 was chosen as the cutoff value. After correction for TRFs, the HR for GRS=>+3 was 1.57 (CI 1.2 to 2.0) in Caucasians and estimated to be 1.43 in an external population. However, when the GRS HR was compared to the HR for TRFs, it was no more predictive than any of the individual TRFs. The results were even less promising for the African Americans enrolled in the ARIC study and was not significant; due to possible differences in allele frequency between African Americans and Caucasians or smaller sample size, the GRS in this population did not have as great a range and no documented subjects with CHD had a score of 5.

Utilizing samples gathered from eight case-controlled subject sets with a validation cohort-controlled study of five of the seven subject sets, a GRS for Coronary Heart Disease was developed.[32] Here, 13 (Table 2) well-validated SNPs that originally met a p-value $< 5x10^8$, a standard metric for large GWAS studies, were used. Of the 13 SNPs, seven were associated with at least one endpoint: Coronary Heart Disease, Cardiovascular Disease, or MI. The GRS was significantly associated with CHD, CVD, and MI; however, inclusion of the GRS with a risk model based on TRFs for the three conditions did not significantly improve the predictive power of the traditional risk models alone, although there was significant NRI improvement of intermediate risk with the inclusion of the GRS.

The same 13 SNPs were assayed in a North American sample set, the Framingham Heart Study Offspring and Third-Generation studies. [33] Endpoints were CAC, CVD, here defined as: cardiovascular-related death, MI, cardiac insufficiency, angina, stroke, transient ischemic attack, intermittent claudation or CHF, and hard CHD, defined as CV death or MI. The 13 SNP GRS was associated with hard CHD and slightly improved the predictive power of a risk model based on age, sex, and conventional CVD factors. The 13 SNP GRS also was a slightly better predictor of CAC over the conventional model.

A GRS was developed from published and replicated GWAS SNPs and candidate genes associated with CHD but not with traditional or intermediate risk factors: age, sex, smoking, lipid levels, or blood pressure.[34] The cohorts for the study were non-Hispanic whites enrolled in the ARIC, Rotterdam, and Framingham Offspring Studies. CHD was the endpoint defined

as definite, probable, or silent (by ECG) MI, CHD-related death, or revascularization. Nine GWAS-based SNPs and four from other literature sources were identified; however, their screening criteria resulted in replacement of five SNPs used above and utilization of two SNPs upstream of the CXCL12 gene on chromosome 10 (Table 2). SNPs were weighted based on their level of association between patient FRS, individual SNPs, and both an unweighted and weighted GRS were utilized. Both GRSs were statistically associated with the primary endpoint of incident CHD in all three study populations, and the weighted GRS was an improvement. AUC was increased only in the ARIC study. After addition to the traditional risk factor methods, the weighted GRS improved NRI compared to the unweighted GRS in all three cohort studies.

From eight SNPs associated with CHD independent of TRFs, a GRS was tested against subjects without incident disease in the Spanish REGICOR subject sample and the Framingham Original and Offspring cohorts.[35] The authors note that the REGICOR cohort has a low CHD event rate, the Framingham cohort a high event rate, and a resulting survivor bias against the GRS. In building the GRS, SNPs were weighted for effect size from the CARDIoGRAM study in an additive model. Study endpoints were MI, angina, coronary revascularization, or death; suspected events were assessed by a committee using previously established criteria[36] The FRS was used for disease prediction by TRFs. When individual SNPs were assessed for association with CHD, only the SNP in CDKN2B-AS1 was significant. This GRS increased predictive power in the highest quintile after meta-analysis of the combined cohort groups, and in the Framingham cohort but not the REGICOR. It also had some reclassification and discrimination advantage over the TRF-based score for subjects of intermediate risk but not when applied across all subjects.

With the goal of increasing the classification accuracy for males at high risk of CHD over indices based on TRFs, three GRSs were developed from haplotype data from four SNPs in the LPA locus, and 12 highly significant SNPs replicated in at least two GWASs for CHD or MI.[37] The three GRSs varied in the methods used to weight effects, either directly from that calculated in the GWASs, or applying alternative regression strategies to the study data. The GRSs were applied to a cohort-controlled study sub-population of males without prevalent CHD of the much larger MORGAM project. In the net, all three GRSs performed similarly to others discussed above, i.e., reclassification was improved, suggesting that the primary advantage of adding a GRS to a TRF-based risk score is increasing accurate identification of high-risk individuals. However, after adding family history to the FRS, the GRSs offered no additional improvement to reclassification indices except in 50-59 year old males; here, a regressive weighting strategy outperformed the GWAS-derived effect-weighted GRS, resulting in mostly upward reclassification of 13.8% of this early onset population. Unfortunately, since this GRS was weighted using study data, it may be population-specific, of limited use, and difficult to replicate.

Finally, two gender-specific GRSs were developed from the SNP data in the Women's Health Genome Study.[38] In the first, 101 SNPs associated with any heart disease-related endpoint, including CVD (MI, stroke, CAD, or related death), lipid profile, hypertension, hemoglobin, or CRP were included. The second GRS was developed from the 12 CVD-associated SNPs (Table 2). For neither GRS did inclusion with an age-adjusted TRF covariate model improve prediction of CVD over the TRF model alone, although inclusion of family history showed a significant increase in predictive power, suggesting that the highest confidence SNPs do not capture the entire predictive power genomic factors.

Association of Single Nucleotide Polymorphisms and Gene Expression in Heart Disease

In a study of 166 SNPs associated by GWAS with the heart disease-related factors, hypertension, cholesterol profile, waist circumference, aortic aneurysm, left ventricular mass, or MI, gene expression profiles of genes within 200 kB of the risk SNPs were determined by microarray.[39] RNA was purified from liver, the medial and adventitial layers of mammary artery, and dilated and nondilated ascending aorta. These biopsies were obtained from aortic valve surgery patients. RNA from atherosclerotic tissue was obtained from carotid endarterectomy surgery patient samples. Forty-seven risk-associated SNPs were correlated with the expression of a "nearby gene." Gene function was loosely associated with tissue; SNPs discovered in GWASs for lipid phenotype were more likely differentially expressed in liver or plaque; SNPs associated with MI were differentially expressed in the aorta or mammary artery. Regulation of single genes associated with SNPs within 35kB is evident, although not always with the SNP originally investigated.

There are two additional classes of regulation exhibited, both highlighting the long-range regulation of gene expression by factors such as chromatin structure. In the first, there is evidence of regulatory association in mammary artery between a single SNP with a block of three genes, all of which are greater than 100 kB away from the risk SNP. In the second case, no risk SNPs are found in strong linkage disequilibrium with the differentially expressed gene. Another interesting finding is for the risk SNP in 2q33 associated with MI. This SNP is in the WDR12 gene that has been used in many other GRSs (Table 2); however, the researchers found that WDR12 expression is unaffected, but a neighboring gene, NBEAL1, is differentially expressed in all blood vessel sample types tested.

A second CAD-focused study used associative data between GWAS SNPs and gene expression data to discover disease-associated pathways and regulatory molecules.[40] This pathway-discovery approach potentially highlights genes that contribute to disease but do not rise to significance in a GWAS. SNPs were gathered from 16 GWASs where endpoints were CHD, arterial calcification, or MI. These were filtered for quality and significance, pooled into two independent SNP sets for confirmatory power, then combined for a meta-analysis of a single list.

To determine the genes affected by the SNPs, expression data correlated to the SNPs were empirically determined in adipose tissue, liver, human aortic epithelium, blood, and a pool of multiple tissue and cell types. The resulting tissue-specific lists of SNP-associated differentially expressed genes were mapped to known functional and interaction networks in the openly available molecular interaction databases, Reactome, the Kyoto Encyclopedia of Genes and Genomes (KEGG), and BioCarta. Then, these investigators utilized the wide public availability of gene expression data from other investigators to include genes strongly co-expressed with the genes in their gene list. Overlapping networks (>20%) were combined to "supersets" and further filtered to arrive at truly robust CAD-associated gene sets. Using a Bayesian approach, tissue-specific regulatory molecules were identified in each superset. In addition, a dataset of co-expressed genes was developed from CAD-related tissues in human and mouse. These data from multiple cohorts were experimentally mapped back to SNPs shown to influence the expression of the gene—an eSNP—and were filtered to remove eSNPs in strong LD with each other.

Of 833 canonical pathways tested, 79 were significant including those related to the TRFs—dyslipidemia, inflammation, and vascular dysfunction. Novel pathways revealed in this analysis were neuroprotection, cell-cycle, and proliferation, DNA or RNA metabolism/regulation, and protein turnover. Further, from the 22 supersets, lipid profile and immune system involvement were top annotations. The genes within eight of the supersets were not over-represented within any known canonical pathway; nonetheless, central regulators shared across multiple tissues and with networks of known function were identified within these interaction networks of unknown function. Subnetworks of known function are annotated within these supersets, suggesting that either the trimming step eliminated true positives affecting sensitivity of the over-representation analysis, and/or our knowledge of the sum total of biological activity is lacking. Within the discovery of regulatory molecules, the authors note that CAD-associated SNPs do not affect the regulatory molecules, themselves, but rather effect the downstream GWAS molecules.

CONCLUSIONS

Heritability of the top SNPs is estimated to be 10%,[5] whereas heritability from family-histories may be as high as 50% [41] although, especially for heart disease and stroke, family histories are difficult to gather and assess,[42] unless a health-care provider is directly involved. In addition, this discrepancy is likely a combination of biological factors: loci that fall under the false discovery rate of GWASs, gene-gene and gene-environment interactions, low-frequency high-effect loci, and epigenetics and their contribution to gene-gene interactions. GRS development relies on numerous assumptions and methods of data manipulation. As demonstrated above, investigators have used very different methods and made highly variable assumptions in their choice of SNPs, the methods of modeling the data, determining significance, and calculating a GRS. Much experimentation will be necessary to tease out how each of these factors, biological and statistical, individually and combined, affect heart disease and determine which tests yield the greatest information in a cost-effective manner.

The medical community is likely to see a vast increase in understanding heart disease. An untapped reservoir of information lies in understanding gene expression patterns and how they are linked to SNPs. Recent papers from the Encyclopedia of DNA Elements project (encodeproject.org) have added to our understanding of chromatin organization, histone marks and their effects on gene expression.[43,44] It will require designing a study *de novo* to gather these data as past studies focused on genotype. Teasing out additional important SNPs will likely require new statistical tools. Until such time, the data speak for themselves: GRSs currently contribute little to risk prediction for heart disease over the information already easily gathered in a doctor's office. As a result, the American Heart Association does not recommend adding a GRS to heart disease risk assessment, and this recommendation applies equally to the airman medical examination.

Table 1. Current FAA Guidance for Medical Certificate Issuance for Cardiovascular Conditions

Factor	Qualifer	License Class	Documentation	On-Line AME Guide Result
Atrial Fibrillation	None for >5 years	All	Previous workup for CAD or heart disease	Issue if no ischemia, emboli or structural heart disease; else FAA decision
	Chronic	All	CVE (cardiovascular evaluation) w ECHO, EST and 24-hour Holter	FAA decision
	Paroxysmal/Lone	All	Same as Chronic	Initial Special Issuance- Req FAA decision Follow up Special Issuance
Coronary Heart Disease; Myocardial Infarction	Open or left coronary artery stent; >6 months	All (see reference for 1st or unlimited 2nd)	Current status/CVE, hosp adm sum, post-op report, Graded Exertion Test (GXT),	FAA decision
	Percutaneous ex. Left coronary artery; >3 months	All (see reference for 1st or unlimited 2nd)	Same	FAA decision
	Uncomplicated MI, no open or percutaneous intervention; >3 months	All (see reference for 1st or unlimited 2nd)	Same	FAA decision
	Non-CAD MI; >3 months	All	Current status report; copies of all medical records	FAA decision
Hypertension	>155/95 for 3 days	All	Current status report, Hypertension worksheet (ref)	Issue or FAA decision or Special Issuance

Relevant Web Sources

Atrial Fibrillation: http://www.faa.gov/abou/office_org/headquarters_offices/avs/offices/aam/ame/guide/app_process/exam_tech/item36/amd/afib/

Coronary Heart Disease/Myocardial Infarction: https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/ame/guide/dec_cons/disease_prot/coronary/

http://www.faa.gov/licenses_certificates/medical_certification/specialissuance/coronary/

Hypertension: https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/ame/guide/app_process/exam_tech/item36/amd/hypertension/

https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/ame/guide/media/C-CACIHypertension.pdf

Table 2. SNPs from GWA Studies for Coronary Heart Disease Utilized in Gene Risk Score Development

SNP Identity*	Chromosomal Location	Genes Genomically Nearby	Link to CHD [15]	Original SNP Discovery Reference	GRS Usage by Reference
rs2505093	10p11.23	KIAA1462		46	
rs1746048	10q11	near CXCL12		28	32, 33, 34, 35, 38
rs501120	10q11	near CXCL12			34, 37
rs10822891	10q21	CTNNA3			30 (AA) #
rs1412444	10q23.31	LIPA		52, 53	
rs2296436	10q24	HPS1			30 (C) #
rs12413409	10q24.32	CYP17A1, CNNM2, NT5C2	Hypertension	47	
rs11016076	10q26	MKI67			30 (C)
rs1799963	11p11	F2			30 (AA)
rs974819	11q22.3	PDGF		52	
rs964184	11q23.3	ZNF259, APOA5-A4-C3-A1	LDL cholesterol	47	
rs2298566	11q24	SNX19 – non-synonomous, L>R			30 (C), 31, 34
rs89962	12q13	KRT5			30 (AA)
rs3184504	12q24	SH2B3	Hypertension	25	32, 33, 37
rs2259816	12q24.31	Near HNFA1, c12orf43		45	32, 33, 34, 37
rs4773144	13q34	COL4A1, COL4A2		47	
rs2895811	14q32.2	HHIPL1		47	
rs121022003	15q21	DMX2			30 (AA)
rs3825807	15q25.1	ADAMTS7		47, 52	
rs12936587	17p11.2	RASD1, SMCR3, PEMT		47	
rs216172	17p13.3	SMG6, SRR		47	
rs4796603	17q21	HAP1			30 (AA)
rs546522	17q21.32	UBE2Z, GIP, ATP5G1, SNF8		47	
rs1122608	19p13	SMARCA4, near LDLR	LDL cholesterol	28	32, 33, 37, 38
rs1122955	19q13	ZNF132			30 (AA)
rs1800437	19q13	GIPR			30 (AA)
rs646776	1p13	Near CELSR2, PSRC1, SORT1	LDL cholesterol	28, 50, 51	32, 33, 37, 38
rs11206510	1p32	PCSK9	LDL cholesterol	48	32, 33, 37, 38
rs17114036	1p32.2	PPAP2B		27	

Table 2. SNPs from GWA Studies for Coronary Heart Disease Utilized in Gene Risk Score Development (Continued)

SNP Identity*	Chromosomal Location	Genes Genomically Nearby	Link to CHD [15]	Original SNP Discovery Reference	GRS Usage by Reference
rs2213948	1q23	LOC646377			30 (C)
rs2505093	10p11.23	KIAA1462		46	
rs17465637	1q41	MIA3		28	32, 33, 35
rs3008621	1q41	MIA3			37
rs4002007	21q21	ADAMTS1			30 (C)
rs9982601	21q22	Near SLC5A3, MRPS6, KCNE1		28	32, 33, 34, 35, 37, 38
rs1010	2p11	VAMP8 – 3'UTR			30 (C), 31, 34
rs4299376	2q21	ABCG8	LDL cholesterol	53	
rs6725887	2q33	WDR12		28	32, 33, 34, 35, 37, 38
rs3900940	3q13	MYH15 – non-synonomous, T>A			30 (C), 31, 34
rs9818870	3q22	MRAS		28	32, 33, 34, F41 37, 38
rs2306174	3q22.3	MRAS		45	
rs7439293	4q32	PALLD3 – intron			30 (C), 31, 34
rs3749817	5q31	FSTL4			30 (AA)
rs2706399	5q31.1	IL5		53	
rs20455	6p21	KIF6			30 (C, AA), 31
rs17609940	6p21.31	ANKS1A		47	
rs12526453	6p24	PHACTR1		28	32, 33, 34, 35, 37, 38
rs6903956	6p24.1	C6orf105		55	
rs12190287	6q23.2	TCF21		47	
rs6922269	6q25	MTHFD1L – intron			34, 38
rs2048327	6q25.3	SLC22A3			37, haplotype SNP in 6q25
rs3127599	6q25.3	LPA2			37, haplotype SNP in 6q25
rs7767084	6q25.3	LPA			37, haplotype SNP in 6q25
rs10755578	6q25.3	LPA			37, haplotype SNP in 6q25
rs3798220	6q26-27	LPA	coronary artery disease	49	32, 33
rs10455872	6q26-27	LPA	independent	54	
rs8089	6q27	THBS2			30 (AA)
rs10953541	7q22.3	BCAP29		52	

Table 2. SNPs from GWA Studies for Coronary Heart Disease Utilized in Gene Risk Score Development (Continued)

SNP Identity*	Chromosomal Location	Genes Genomically Nearby	Link to CHD [15]	Original SNP Discovery Reference	GRS Usage by Reference
rs2961135	7q35	OR2A25			30 (AA)
rs4994	8p12	ADRB3			30 (C)
rs2505093	10p11.23	KIAA1462		46	
rs17321515	8q24.13	TRIB1	LDL cholesterol	53	
rs4977574	9p21	CDKN2B-AS1/ANRIL		28	32, 33, 34, 35, 37, 38
rs579459	9q34.2	ABO	LDL cholesterol	47	
rs2200733	4q25		Stroke		37
rs12425791	12p13.3	NINJ2	Stroke		37

*SNP Identity chosen from one study and may vary but listed SNPs are in linkage disequilibrium with the listed genes # 28 (AA) and 28 (C) – GRSs in this study (Morrison) were specific for African Americans or Caucasians

REFERENCES

1. Murphy, S.L., Xu, J. & Kochanek, K.D. Deaths: final data for 2010. *Natl Vital Stat Rep* **61**, 1-117 (2013).

2. Go, A.S. *et al.* Heart disease and stroke statistics--2014 update: a report from the American Heart Association. *Circulation* **129**, e28-e292 (2014).

3. Heron, M. Deaths: leading causes for 2010. *Natl Vital Stat Rep* **62**, 1-96 (2013).

4. Roberts, R., Marian, A.J., Dandona, S. & Stewart, A.F. Genomics in cardiovascular disease. *J Am Coll Cardiol* **61**, 2029-37 (2013).

5. Wienke, A., Holm, N.V., Skytthe, A. & Yashin, A.I. The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. *Twin Res* **4**, 266-74 (2001).

6. Greenland, P. *et al.* 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* **122**, e584-636 (2010).

7. Ridker, P.M., Buring, J.E., Rifai, N. & Cook, N.R. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA* **297**, 611-9 (2007).

8. Ridker, P.M., Paynter, N.P., Rifai, N., Gaziano, J.M. & Cook, N.R. C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men. *Circulation* **118**, 2243-51, 4p following 2251 (2008).

9. Psaty, B.M. *et al.* Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. *JAMA* **289**, 2534-44 (2003).

10. Chobanian, A.V. *et al.* The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* **289**, 2560-72 (2003).

11. Turnbull, F. & Blood Pressure Lowering Treatment Trialists, C. Effects of different blood-pressure-lowering regimens on major cardiovascular events: results of prospectively-designed overviews of randomised trials. *Lancet* **362**, 1527-35 (2003).

12. International Consortium for Blood Pressure Genome-Wide Association, S. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).

13. Fava, C. *et al.* Prediction of blood pressure changes over time and incidence of hypertension by a genetic risk score in Swedes. *Hypertension* **61**, 319-26 (2013).

14. Havulinna, A.S. *et al.* A blood pressure genetic risk score is a significant predictor of incident cardiovascular events in 32,669 individuals. *Hypertension* **61**, 987-94 (2013).

15. Roberts, R. & Stewart, A.F. Genes and coronary artery disease: where are we? *J Am Coll Cardiol* **60**, 1715-21 (2012).

16. Global Lipids Genetics, C. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* **45**, 1274-83 (2013).

17. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491-3 (2007).

18. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488-91 (2007).

19. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).

20. Pasmant, E. *et al.* Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* **67**, 3963-9 (2007).

21. Broadbent, H.M. *et al.* Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum Mol Genet* **17**, 806-14 (2008).

22. Jarinova, O. *et al.* Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol* **29**, 1671-7 (2009).

23. Folkersen, L. *et al.* Relationship between CAD risk genotype in the chromosome 9p21 locus and gene expression. Identification of eight new ANRIL splice variants. *PLoS One* **4**, e7677 (2009).

24. Paynter, N.P. *et al.* Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. *Ann Intern Med* **150**, 65-72 (2009).

25. Gudbjartsson, D.F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* **41**, 342-7 (2009).

26.Silander, K. *et al.* Gender differences in genetic risk profiles for cardiovascular disease. *PLoS One* **3**, e3615 (2008).

27.Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* **357**, 443-53 (2007).

28.Myocardial Infarction Genetics, C. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* **41**, 334-41 (2009).

29.The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).

30.Morrison, A.C. *et al.* Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* **166**, 28-35 (2007).

31.Bare, L.A. *et al.* Five common gene variants identify elevated genetic risk for coronary heart disease. *Genet Med* **9**, 682-9 (2007).

32.Ripatti, S. *et al.* A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* **376**, 1393-400 (2010).

33.Thanassoulis, G. *et al.* A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circ Cardiovasc Genet* **5**, 113-21 (2012).

34.Brautbar, A. *et al.* A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis* **223**, 421-6 (2012).

35.Lluis-Ganella, C. *et al.* Assessment of the value of a genetic risk score in improving the estimation of coronary risk. *Atherosclerosis* **222**, 456-63 (2012).

36.Grau, M. *et al.* Why should population attributable fractions be periodically recalculated? An example from cardiovascular risk estimation in southern Europe. *Prev Med* **51**, 78-84 (2010).

37.Hughes, M.F. *et al.* Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. *PLoS One* **7**, e40922 (2012).

38.Paynter, N.P. *et al.* Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* **303**, 631-7 (2010).

39.Folkersen, L. *et al.* Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. *Circ Cardiovasc Genet* **3**, 365-73 (2010).

40.Makinen, V.P. *et al.* Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet* **10**, e1004502 (2014).

41.Zdravkovic, S. *et al.* Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* **252**, 247-54 (2002).

42.Facio, F.M. *et al.* Validation of My Family Health Portrait for six common heritable conditions. *Genet Med* **12**, 370-5 (2010).

43.Gerstein, M.B. *et al.* Comparative analysis of the transcriptome across distant species. *Nature* **512**, 445-8 (2014).

44.Ho, J.W. *et al.* Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449-52 (2014).

45.Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* **41**, 280-2 (2009).

46.Erdmann, J. *et al.* Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. *Eur Heart J* **32**, 158-68 (2011).

47.Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* **43**, 333-8 (2011).

48.Kathiresan, S. & Myocardial Infarction Genetics, C. A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. *N Engl J Med* **358**, 2299-300 (2008).

49.Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med* **361**, 2518-28 (2009).

50.Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**, 189-97 (2008).

51.Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714-9 (2010).

52.Coronary Artery Disease Genetics, C. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* **43**, 339-44 (2011).

53. Consortium, I.K.C. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet* **7**, e1002260 (2011).

54. Shiffman, D. *et al.* Single variants can explain the association between coronary heart disease and haplotypes in the apolipoprotein(a) locus. *Atherosclerosis* **212**, 193-6 (2010).

55. Wang, F. *et al.* Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat Genet* **43**, 345-9 (2011).