Published in final edited form as:

Pharmacogenomics. 2009 February ; 10(2): 213–223. doi:10.2217/14622416.10.2.213.

Genome-wide association studies of coronary artery disease and heart failure: where are we going?

Gerald W Dorn* and Sharon Cresci†,*

Washington University Center for Pharmacogenomics, 660 S. Euclid Avenue, Campus Box 8086, St Louis, MO 63110, USA

Abstract

The heart diseases that account for a large amount of morbidity and mortality in the developed world (coronary artery disease, myocardial infarction and heart failure) are phenotypically heterogeneous disorders. It has been suspected for many years that genetics may have an important role in these diseases and their poor outcome. However, their complex and likely polygenic pathophysiology has confounded clear understanding of the genetic contribution to their etiology. Despite technological progress and great promise associated with genome-wide association studies, to date the results of their application to coronary artery disease, myocardial infarction and heart failure have yielded limited insights into these common diseases. This review discusses the current status of genome-wide association studies as they have been applied to these cohorts. The potential limitations of these studies, as well as potential future directions for identifying important genes are also discussed.

Keywords

coronary artery disease; genome-wide association study; genomics; heart failure

“If you are asking the wrong question, or the right question in the wrong way, the answer doesn’t really matter.”

– Thomas Pynchon, Novelist

Coronary artery disease and its sequelae (angina, myocardial infarction, cardiac revascularization or transplantation) affect more than 16 million Americans [1] and untold numbers of people world-wide. Approximately 38% of all individuals who suffer a heart attack will die from it [1]. Many who survive will develop chronic heart failure, which afflicts one in five Americans and has a treated mortality rate exceeding that of most malignancies [1]. A genetic contribution to coronary artery disease has been recognized for over 40 years: death rates are strikingly increased in first-degree relatives of young men (age <55 years) and women (age <65 years) with terminal atherosclerotic disease [2]. Large-scale twin pair studies with 26 years of follow-up show a greater degree of concordance for early coronary death risk among monozygotic twins (eightfold increase in risk) than for dizygotic

© 2009 Future Medicine Ltd

†Author for correspondence: Washington University Center for Pharmacogenomics, 660 S. Euclid Avenue, Campus Box 8086, St Louis, MO 63110, USA, Tel.: +1 314 362 8901; Fax: +1 314 362 0186; scresci@dom.wustl.edu.

*The authors contributed to this work equally

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

twins (2.5-fold increase in risk) [3]. Although specific genetic factors that contribute to coronary artery disease have been identified [4-8], and it has been estimated that up to 41% of interindividual variability may be determined by genetic factors [9], the genetic architecture of coronary artery disease remains largely undefined.

The role of genetics in heart failure is even less clear. Genetic mutations have been implicated in some rare inherited cardiomyopathies [10,11], but common nonfamilial heart failure is a complex clinical syndrome. It represents a deteriorating structural and functional state attained when cardiac damage from any cause overwhelms intrinsic compensatory mechanisms. A specific cause for heart failure may be difficult to establish since the possible etiologic factors (e.g., ischemia/atherosclerosis, hypertension, valve dysfunction, viral infection and toxins) are frequently present in combination. Even when a cause for heart failure can be specifically identified and temporally defined (as with a first myocardial infarction) the pathophysiological interactions between arterial wall, circulating platelets, coagulation factors, neurohormones, and myocardium are each subject to common genetic variations that can interact in many different ways to significantly modify individual susceptibility or outcome.

Since human genetic diversity is greater than originally anticipated (0.5% of the total genome differs between individuals [12,13]), and at least 44% of annotated genes exhibit DNA sequence variation [12], it is likely that genetic variation contributes significantly to individual susceptibility to coronary artery disease, as has been found to be the case with asthma and Type II diabetes mellitus [14]. Defining the relevant gene–gene and gene–environment relationships for these and other complex diseases represents a monumental challenge. A decade of study dominated by candidate gene approaches has produced limited success in coronary artery disease and heart failure. Early efforts reflected an underlying assumption of modern reductionist experimentation, that all things must have a cause, and that any series of causes must have a primary cause. Putative associations between cardiac phenotypes and polymorphic variants of candidate causal genes were used to describe mechanistically based individual cause–effect relationships that contrasted with the seeming ambiguity of purely statistical associations in early unbiased genetic studies. Yet, candidate gene studies were often characterized by results conforming to the five Ws, that is, ‘which was what we wanted’, and were often undermined by small population size and the absence of independent validation. As a consequence, genome-wide association studies (GWAS) using advanced, high-resolution platforms have become the current procedure of choice for large-scale comparative genomics research. For some diseases, microarray-based genotyping of hundreds of thousands of tagged SNPs located at set intervals across the complete genome of thousands of individuals have provided robust statistical evidence to implicate particular genetic loci, and have proven to be highly reproducible [15-21]. Here, we review the current status of these types of studies as they have been applied to coronary artery disease (Table 1) and heart failure (Table 2).

Challenges for GWAS in complex heart disease

This review focuses on GWAS investigations using cohorts of individuals with coronary artery disease (typically defined angiographically as having some degree of coronary artery stenosis, although the degree of stenosis differs between studies) or myocardial infarction (typically defined as suggestive patient symptoms and/or electrocardiogram changes [e.g., ST segment elevations] and elevated serum levels of troponin and/or creatine kinase [markers of myocardial injury]). However, it should be mentioned that GWAS analyses have identified important, and physiologically-relevant, associations between SNPs and coronary artery disease risk phenotypes such as Type II diabetes mellitus and dyslipidemia. For example, using GWAS analysis Burkhardt and colleagues identified SNPs in the 3-

hydroxy-3-methylglutaryl [HMG]-coenzyme A (CoA) reductase [*HMGCR*] gene that were associated with serum low-density lipoprotein (LDL)-cholesterol levels. They also identified a functional SNP in linkage disequilibrium (LD) with the GWAS-identified SNPs, which resulted in lower expression of an alternatively spliced *HMGCR* mRNA [22]. Willer and colleagues recently performed a meta-analysis of three GWAS including a total of 8816 individuals and identified 11 SNPs associated with increased LDL-cholesterol concentrations in genes that were known to be involved in lipid metabolism [23]. Interestingly, when these investigators examined whether these SNPs were associated with coronary artery disease in the Wellcome Trust Case Control Consortium (WTCCC) cohort (discussed further below), all 11 SNPs were significantly more prevalent in the individuals with coronary artery disease [23]. Recently, two meta-analyses of large (~20,000 individuals in each analysis) cohorts of European ancestry have identified more than 20 loci associated with dyslipidemia [24,25]. Many of these loci were in, or near, genes known to be involved in lipid metabolism or transport or to cause purely genetic forms of dyslipidemia [24,25]

Coronary artery disease is generally considered to have roughly equivalent environmental and genetic components. Determining the specific genetic contribution for an average individual is complicated both by polygenic inheritance and interindividual variations in genetic and epigenetic factors. As an example, genome-wide linkage analyses in small families with extreme phenotypes of premature coronary artery disease or myocardial infarction identified associated genetic loci on chromosomes 1, 2, 3, 12, 13, 14, 16 and X [26]. However, when inclusion criteria were loosened to include less extreme coronary artery disease phenotypes representing a more ‘typical’, but still early, presentation (age at diagnosis: 59 years), a large genome-wide linkage analyses of 4175 individuals in 1933 families identified only one chromosomal locus (2p.12–2q23.3, spanning a region of 80 Mb containing 446 putative genes) having a $\log_{10}$ of odds (LOD) of 1.98 and a p-value ($p = 0.0014$) that did not achieve genome-wide significance [26]. Such are the unavoidable challenges and potential pitfalls encountered when trying to assign pathological significance to one or more common gene variants in a condition that is polygenic, only partially genetic, and phenotypically heterogeneous.

Early studies

In 2002 Ozaki *et al.* performed what may be considered the first large ‘genome-wide’ study, genotyping 92,788 SNPs in 94 unrelated Japanese individuals with myocardial infarction and comparing the allelic frequency results to 658 healthy controls [27]. All SNPs that had a p-value of greater than 0.01 (for either a recessive or dominant model) were excluded from subsequent analysis. The remaining SNPs (1%) were genotyped in a replication cohort of 1133 unrelated Japanese individuals with myocardial infarction and 1006 healthy controls. A total of two SNPs in the lymphotoxin- α (*LTA*) gene, one SNP in the nuclear factor of κ -light polypeptide gene enhancer in B cells, inhibitor-like 1 (*NFKBIL1*) gene, and one SNP in HLA-B associated transcript 1 (*BAT1*) gene achieved statistical significance [27]. One of the *LTA* SNPs was shown by the investigators to affect the transcriptional activity of *LTA*. A SNP located in intron 1 of the gene resulted in enhanced binding of a nuclear factor in *in vitro* experiments, leading the investigators to propose that carriers of this SNP had enhanced inflammation resulting from *LTA*-mediated cytokine release.

In 2004, Helgadottir and colleagues reported the results of a genome-wide linkage analysis in 296 Icelandic families (713 individuals with myocardial infarction compared with 1741 first-degree family members) [28]. The investigators repeated their analyses for the subgroup of individuals with early-onset myocardial infarction, and for males and females separately. None of the analyses achieved genome-wide significance but the region with the highest LOD was investigated further. Fine mapping of this region resulted in identification of a haplotype in the gene encoding 5-lipoxygenase activating protein (*ALOX5AP*) that was

significantly associated with myocardial infarction in 802 unrelated (Icelandic) individuals with myocardial infarction and 837 unrelated controls [28]. The investigators attempted to replicate this association in an additional, independent (British), cohort of 753 individuals with myocardial infarction compared with 730 controls but were unable to. They did, however, identify an unrelated haplotype in the same gene that had a borderline significant association with myocardial infarction [28].

9p21.3, a locus for many studies & multiple diseases

The WTCCC & replication studies—In June 2007 the WTCCC reported one of several landmark studies to apply newer generation GWAS array analyses to coronary disease [29]. Genetic associations were examined in 14,000 individuals with seven common diseases (~2000 individuals with each disease), including coronary artery disease, and 3000 shared controls. Using Affymetrix (CA, USA) 5.0 arrays containing 500,000 SNPs spaced at regular intervals across the genome, an association was revealed between coronary disease and the 9p21 locus (21.93–22.12 [29]). The strongest association ($p = 1.79 \times 10^{-14}$) was with the rs1333049 SNP (C allele; heterozygous odds ratio [OR]: 1.47 [1.27–1.70]; homozygous OR 1.9 [1.61–2.24]), and associations also existed for SNPs located within approximately 100-kb region of the primary locus. This association between rs1333049 and coronary disease was independently replicated in 604 Japanese cases of coronary disease compared with 1151 controls, and 679 Korean cases with 706 controls [30] with similar C allele odds ratios [1.30 (1.13–1.49); $p = 0.00027$ and 1.19 [1.02–1.38]; $p = 0.025$, respectively).

The statistical association between rs1333049 and coronary disease is compelling, but a functional mechanism by which this region of chromosome 9 promotes risk has not yet been established. The authors of the WTCCC study noted that genes encoding the cyclin-dependent kinase inhibitors 2A and 2B (*CDKN2A* and *CDKN2B*, respectively) and the gene encoding 5'-methylthioadenosine phosphorylase (*MTAP*) are in the general area (although relatively distant from rs1333049). *CDKN2B* expression is induced by transforming growth factor (TGF)- β , which is implicated in development and progression of atherosclerosis [31] but, as the authors note, 'further work is required to determine whether the coronary artery disease association at this locus is mediated through *CDKN2A/B*, *MTAP* or some other mechanism' [29]. As diabetes is a potent risk factor for atherosclerosis [1,32], it is interesting that different SNPs within this genomic region were associated with Type II diabetes mellitus in the same study [29].

The Ottawa Heart Study

McPherson and colleagues simultaneously reported a GWAS of three sequential case-control study comparisons using a p-value cut off for statistical significance of $p < 0.025$ in *Science* [33]. The primary Ottawa Heart Study recruited severe, premature coronary artery disease, defined as documented onset before the age of 60 years and requiring coronary revascularization. A total of 72,864 SNPs (100,000 SNPs approximately 30-kb intervals apart on a custom array) were examined in 322 Caucasian men or women and 312 controls (men: >65 years old; women: >70 years old with no symptoms or history of coronary artery disease). A total of 9636 of the SNPs significantly deviated from Hardy-Weinberg equilibrium and 17,500 had a minor allele frequency of less than 1% and were excluded from further analysis. Of the approximately 70,000 remaining SNPs, 2586 showed statistically significant associations with coronary disease. These SNPs were then genotyped in a replication cohort consisting of 311 cases and 326 controls from the same Ottawa Heart Study cohort. Of the original 2586 SNPs, 50 achieved statistical significance and were genotyped in a second replication cohort of 11,478 Caucasian individuals enrolled in the Atherosclerosis Risk in Communities (ARIC) study. Two of the 50 SNPs replicated:

rs10757274 and rs2383206. These SNPs are only 20-kB away from each other on chromosome 9p21 and are in strong LD.

The association of rs10757274 and rs2383206 with coronary disease was confirmed in three additional cohorts: another subgroup of the Ottawa Heart Study (547 cases and 847 controls), the Copenhagen City Heart Study (CCHS; a prospective 20-year observational cohort study of 1525 cases with ischemic coronary events compared with 9053 controls), and the Dallas Heart Study which used electron-beam computer tomography to define the presence/absence of coronary artery calcium in 154 cases and 527 controls [33]. The genomic region around these two SNPs was mapped at greater resolution in 500 cases and 500 controls from the second and third Ottawa cohorts by genotyping SNPs averaging 5 kb apart between 175 kb upstream and 175 kb downstream of rs10757274 and rs2383206. A total of eight SNPs were significantly associated with coronary disease, all of which were in significant LD with each other and with rs10757274 and rs2383206. The authors sequenced the coding regions of *CDKN2A* and *CDKN2B* in 96 cases but only found a single SNP in *CDKN2A* in six subjects and concluded that this SNP was ‘unlikely to explain the coronary heart disease risk associated with the locus’ [33].

This study was performed with a very high degree of rigor. Yet, it is striking that each of the levels of SNP genotyping described a genotype–phenotype association with approximately 2% of the parent SNP population: approximately 2500/100,000 in the primary study, 50/2500 in the first replication and 2/50 (which were in LD) in the second replication. This reveals no obvious enrichment through replication, and the final result was a genetic locus in the same chromosomal region as reported by WTCCC and others for Type II diabetes, abdominal aortic aneurysms, intracranial aneurysms, and late-onset Alzheimer’s disease [29,34,35]. The authors noted that no annotated genes localized to this region and did not propose a mechanism for the association. However, they concluded that since African–Americans were adequately represented in only one of the validation cohorts (the Dallas Heart Study), and the associations between rs10757274 and rs2383206 and coronary disease were not observed in this racial subgroup, these particular SNPs are unlikely to be causative; and since the association between rs10757274 and rs2383206 and coronary disease in Caucasians was not affected after adjusting for known coronary risk factors in the ARIC and CCHS studies, the observed genetic effect is independent of conventional clinical risk factors, such as diabetes [33].

Icelandic Myocardial Infarction Study

In the same issue of *Science*, Helgadottir and colleagues reported findings from GWAS of Icelandic subjects with myocardial infarction [36]. Remarkably, despite using a different microarray platform and studying myocardial infarction rather than the broader category of coronary artery disease, the study identified the same chromosomal location as did McPherson *et al.* [33]. A total of 1607 early myocardial infarction cases (myocardial infarction before the age of 70 years in males and 75 years in women) and 6728 control individuals without history of coronary disease were genotyped using Illumina (CA, USA) Hap300 arrays. Although there was no genotype–phenotype association having a p-value adequate for the number of multiple comparisons, ‘more than expected’ were observed with p-values of less than 10^{-5} , and were investigated further. The three SNPs exhibiting strongest associations with myocardial infarction, rs10116277, rs1333040 and rs2383207 ($p = 1.9 \times 10^{-6}$; $p = 6.1 \times 10^{-7}$; $p = 1.4 \times 10^{-6}$, respectively), were all within the same 190-kb region of 9p21 and in significant linkage disequilibrium. A total of 11 additional SNPs within this region had borderline significant p-values (ranging between 0.041 and 9.2×10^{-5}). The rs10116277, rs1333040 and rs2383207 SNPs each replicated in multiple independent cohorts: 665 Icelandic myocardial infarction patients and 3533 controls plus three different case–control cohorts from the USA, PENN CATH (PA, USA), Emory

Genebank (GA, USA), and CATHGEN (NC, USA). When the data from all four case-control cohorts were combined, rs10757278 had the strongest association with myocardial infarction and a strong association with coronary disease when individuals with myocardial infarction were excluded. Further analysis demonstrated that rs10757278 also had a gene-dose effect on age at which the infarction occurred: each copy of the risk allele was associated with an approximately 1 year decrease in age of onset of myocardial infarction; $p = 2.9 \times 10^{-7}$.

Similarly to the authors of the above two studies, the Icelandic Myocardial Infarction Study investigators noted that the genes encoding *CDKN2A* and *CDKN2B* were located within the region of interest. Also, like the other studies, resequencing of the exons, exon-intron junctions and regulatory regions of these genes in 93 individuals with early onset myocardial infarction did not identify any SNPs that suggested a mechanistic basis for the association. Notwithstanding, these investigators have also reported an independent association between rs10757278 and abdominal aortic and intracranial aneurysms [34], suggesting a valid biological effect across multiple vascular disease phenotypes.

WTCCC & German Myocardial Infarction Family Study

Samani *et al.* identified SNPs associated with coronary artery disease in 1926 WTCCC cases with a history of early myocardial infarction or coronary revascularization (before age 66 years) and a strong family history of coronary disease, compared with 2938 WTCCC controls [37]. Associated SNPs were validated in 875 similar cases from the German Myocardial Infarction Family Study, in comparison with 1644 random controls. In the WTCCC sample, 396 SNPs were associated with early coronary disease at a p-value of less than 0.001. If the SNP had a false-positive-report probability of less than 0.5 ($n = 30$) or was within 100 kb in either direction of the association SNP and also had a p-value of less than 0.001, the locus was considered as a single result ($n = 9$). A 'lead SNP' (the SNP with the smallest false-positive-report probability) within the locus was identified, and each was tested in the German study. After adjusting for multiple comparisons, three SNPs met the criteria for statistical significance for validation, rs2943634 located on chromosome 2q36.3 (C allele OR: 1.20 [1.06–1.35]; $p = 0.004$), rs6922269 located on chromosome 6q25.1 (A allele OR: 1.24 [1.09–1.41]; $p = 0.001$) and rs1333049 located on chromosome 9p21.3 (C allele OR: 1.33 [1.18–1.51]; $p = 6.8 \times 10^{-6}$) [37]. As previously noted, rs1333049, the SNP with the strongest association, is not located within a known gene on 9p21.3. rs6922269 is located on chromosome 6q25.1 within the gene encoding methylenetetrahydrofolate dehydrogenase (NADP⁺-dependent) 1-like protein (*MTHFD1L*), and rs2943634 is located on chromosome 2q36.3 in a pseudogene of unknown significance.

A secondary analysis combining all SNPs significantly associated ($p < 0.001$) with coronary disease in the WTCCC study, or with myocardial infarction in the German Myocardial Infarction Family Study, identified four additional SNPs with a false-positive-report probability of less than 0.2 and OR between 1.21 and 1.33 in the combined analysis: rs599839, rs17465637, rs501120, rs17228212 [37]. These SNPs tag four genes: pro-line/serine-rich coiled-coil 1 (*PSRC1*), melanoma inhibitory activity family, member 3 (*MIA3*), chemokine (C-X-C) ligand 12 (stromal cell derived factor 1) (*CXCL12*), and SMAD family member 3 (*SMAD3*) involved in cell growth, but that were not previously found to be associated with coronary disease. The genotype-phenotype associations and the known biological functions of the encoded genes suggested a possible effect on atherosclerotic plaque stability [37].

Meta-analyses & SNP-specific confirmatory studies

Schunkert and colleagues performed a meta-analysis of seven case-control studies involving a total of 4645 cases of myocardial infarction or coronary artery disease and 5177 controls, all of which genotyped for rs1333049 SNP [38]. The C variant of rs1333049 was more prevalent in cases than controls in all seven studies, as well as in the combined analysis [38]. The combined OR for the C allele was 1.29 (1.22–1.37; $p = 0.0001$) with an apparent gene-dose effect (homozygous OR: 1.67 [1.46–1.87; $p = 0.001$]) [38]. Pooled analysis of only myocardial infarction or coronary disease excluding myocardial infarction confirmed significantly different genotype frequencies in the cases versus controls (C allele odds ratios of 1.38 [1.24–1.54; $p = 0.0095$] and 1.27 [1.19–1.35; $p = 0.0003$], respectively) [38].

Shen and colleagues confirmed the association between coronary disease and two of the SNPs reported by McPherson, rs10757274, rs2383206 [33], as well as two SNPs reported by Helgadottir, rs2383207, and rs10757278 [34,36] in a South Korean population [39]. In 611 cases of either angiographically-proven severe coronary disease or myocardial infarction, compared with 294 controls, all four SNPs were in significant LD ($>90\%$) and produced ORs between 1.29–1.32, $p = 0.011$. The most common four-SNP haplotype (GGGG) was associated with risk for coronary disease (OR: 1.28 [1.04–1.56]; $p = 0.019$), whereas the second most common haplotype (AAAA) conferred protection from coronary disease (OR: 0.75 [0.60–0.92]; $p = 0.007$) [39]. These investigators further replicated the association between these four SNPs and myocardial infarction in an Italian population (OR: 1.25–1.31; p -values: 0.014–0.037) as well as the haplotype association (GGGG haplotype OR: 1.27 [1.03–1.56]; $p = 0.028$ and AAAA haplotype OR: 0.80 [0.65–1.0]; $p = 0.047$) [40].

Abdullah and colleagues also investigated the association between these four SNPs (rs10757274, rs2383206, rs2383207 and rs10757278) and premature coronary disease in the GeneQuest Caucasian American cohort [41]. A total of 310 cases with strong family histories and premature (men aged 45 years; women aged 50 years) coronary atherosclerosis or myocardial infarction were compared with 560 controls. Significant differences in allele frequencies were found for all four SNPs (OR: 1.67–1.78; p -values: 6.61×10^{-7} to 1.87×10^{-8}) [41]. A sibling trait disequilibrium transmission test in the 310 families showed that the association held for three of the SNPs (rs10757274, rs2383206 and rs10757278; p -values: 0.014, 0.036 and 0.0038, respectively) [41].

van der Net and colleagues genotyped ten SNPs previously reported to be associated with coronary artery disease and myocardial infarction and investigated whether any of the ten was independently associated with coronary artery disease in 2145 individuals with familial hypercholesterolemia [42]. They confirmed the association in four (one of which was the rs10757274 SNP) of the ten SNPs with coronary artery disease in this cohort [42].

Recently, Broadbent and colleagues analyzed 11 SNPs in the 9p chromosomal region previously reported to have a significant association with coronary artery disease (seven SNPs) or Type II diabetes mellitus (four SNPs) in 4251 individuals with coronary artery disease before the age of 65 years and 4443 controls (free of coronary artery disease at age 65 and no family history of coronary artery disease by self-report) [43]. Two of the genotyped SNPs with reported association with coronary artery disease deviated significantly from Hardy–Weinberg equilibrium and were excluded from further analysis. The remaining seven coronary artery disease SNPs were all significantly associated with coronary artery disease in this validation cohort [43]. SNP rs2891168 had the strongest association. Interestingly, analyses investigating the association of this SNP with coronary artery disease in distinct risk factor subgroups (e.g., hypertension, cigarette smokers and diabetes) were unable to affect the strength of the association. The investigators identified a

haplotype that had increased prevalence in individuals with coronary artery disease and demonstrated that this haplotype overlapped with exons 13–19 of a recently described gene that encodes a large antisense noncoding RNA (*ANRIL*). *ANRIL* expression was, determined by RT-PCR, was found to be present in coronary smooth muscle cells, human umbilical vein endothelial cells, human microvascular endothelial cells, carotid endarterectomy cells, and abdominal aortic aneurysm cells but not immortalized T-cells [43]. The authors speculate that the high-risk 9p21 haplotype is involved in influencing the level of *ANRIL* expression but acknowledge that how *ANRIL*, or the downstream target(s) of *ANRIL*, is/are involved in coronary artery disease and myocardial infarction remains to be determined [43]. Nevertheless, these data represent an important step towards understanding a potential molecular mechanism that may explain the strong association of this locus with coronary artery disease and myocardial infarction [43].

GWAS of heart failure

“The multitude which does not reduce itself to unity is confusion.”

– Blaise Pascal, French Mathematician

Several studies have reported findings from genome-wide linkage analyses in families with dilated cardiomyopathy that have yielded potential important insights into the mechanisms of purely genetic heart failure. Ellinor and colleagues identified a locus associated with cardio-myopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25–q26.13 [44]. Song and colleagues identified a locus for left ventricular hypertrophy, dilatation, heart failure and sudden cardiac death – without the myocyte disarray characteristic of classical hypertrophic cardiomyopathy – on chromosome 7p21–7q21 [45]. Rampazzo and colleagues identified a locus on chromosome 1q42–43 associated with right ventricular cardiomyopathy, arrhythmia and sudden death [46]. This locus was later fine mapped to the *RYR2* gene encoding the ryanodine receptor, which is essential to making intracellular calcium available for excitation–contraction coupling [47].

To date, only one GWAS has been reported on unrelated individuals with heart failure [48]. Larson and colleagues investigated the association of four cardiovascular disease outcomes, including heart failure, in the Framingham Heart Study (FHS) cohort (a prospective observational cohort studies with 33 years of follow-up) [48]. All subjects free of cardiovascular disease in both cohorts of the FHS (5209 men and women from original study and 5124 offspring in offspring cohort) were analyzed for disease onset during follow-up. Heart failure was defined as having at least two major criteria [48]:

- Paroxysmal nocturnal dyspnea
- Pulmonary rales
- Distended jugular veins
- Enlarging heart size on chest x-ray
- Acute pulmonary edema
- Hepato–jugular reflex
- Third heart sound
- Jugular venous pressure of 16 cm or more
- Weight loss of 4.5 kg or more in response to diuretics
- Pulmonary edema
- Visceral congestion

- Cardiomegaly on autopsy

Or one major and two minor criteria:

- Bilateral ankle edema
- Nocturnal cough

Shortness of breath on ordinary exertion

- Hepatomegaly
- Pleural effusion
- Vital capacity decrease to one third from previous maximum
- Heart rate of 120 beats/min or more

Utilizing the Affymetrix (CA, USA) 100K array as the genotyping platform, 70,987 SNPs (1595 SNPs deviated significantly from Hardy–Weinberg equilibrium, 38,062 SNPs had a minor allele frequency of less than 10% and 2346 SNPs had a call rate of less than 80%) were analyzed, but no association met the p-value criterion of 4.4×10^{-8} appropriate for the multiple comparisons performed in the analyses [48]. Two statistical analyses were employed to determine genetic associations: using generalized estimation equations, six SNPs had association p-values less than 10^{-5} [48]. One of these, rs740363, was associated with heart failure ($p = 8.8 \times 10^{-6}$). No known genes are annotated to this region of chromosome 10 and the authors did not speculate on the possible mechanism of this association [48]. Family-based association testing analyses identified rs10515869 as having the lowest p-value for an association with heart failure ($p = 4.72 \times 10^{-5}$). This SNP is also not located in or near any known gene. Interestingly, a SNP (rs939698) in *RYR2*, a gene that was identified by association with a familial form of right ventricular cardiomyopathy, arrhythmia and sudden death (discussed above) was identified by this GWAS approach to be associated with heart failure ($p = 3.6 \times 10^{-4}$) [48].

Finally, the authors performed a separate analysis of 408 SNPs in, or near (within 200 kb 5' or 3' of the gene) 46 candidate genes selected by a review of the literature for the phenotype of interest (each of the four cardiovascular disease outcomes) [48]. Using a liberal p-value cut-off of less than 0.05, this analysis provided 'suggestive evidence' for four genes previously identified in the literature to be associated with coronary heart disease: *ALOX5AP* (seven out of 23 SNPs within the gene achieved an association p-value of <0.05), connexin37 (*GJA4*; 6/14), myocyte enhancer factor 2A (*MEF2A*; 2/5) and proprotein convertase subtilisin/kexin type 9 (*PCSK9*; 3/11) and for three genes previously identified to be associated with heart failure: phospholamban (*PLN*; 4/16), β_2 adrenergic receptor (*ADRB2*; 2/18) and α -tropomyosin 1 (*TPM1*; 2/7) [48].

Future perspective

"All extremes are error. The reverse of error is not truth, but error still. Truth lies between these extremes."

– Lord David Cecil, English Scholar

The greatest promise of GWAS is that a purely statistical approach, not biased by investigator preconceptions regarding biological mechanisms, will identify genetic factors which modify disease in previously unsuspected ways. By this criterion, GWAS studies of coronary artery disease have succeeded. There can be little doubt that the rs1333049 SNP on chromosome 9p21.3 and other SNPs in this region are markers for a genetic influence on a variety of diseases that broadly affect the vascular system. The association between rs1333049 and atherosclerotic disease is too powerful, and has been replicated in too many

independent cohorts to really doubt. These strikingly positive results clearly demonstrate that the GWAS paradigm can work with complex disease.

So what have we learned from this association? What new insights into disease mechanism have we derived? What new drug targets have been identified? Why haven't the obvious genetic associations with atherosclerosis, lipid receptors or transporters and inflammatory factors, been detected using the nonbiased approach? At this point there is little to show from the GWAS population studies. None of the candidate genes within the 9p21 region have shown a powerful association with disease, and so the biology remains unclear. It seems that there are two possible outcomes with regard to this locus. Either a mechanistic basis for the genotype–phenotype associations will be discovered, or it will not. Either way, this locus would seem to pose problems for GWAS studies of atherosclerosis. As depicted in Figure 1, the current state of knowledge regarding atherosclerosis is that it is a multifactorial disease which may, or may not (depending upon other factors), lead to myocardial infarction, which may or may not lead to heart failure, which is itself multifactorial. A conservative estimate would be that at least 100 genes have the potential to affect each of the modifying factors (Figure 1, light gray boxes), with each having a genetic contribution of as much as 2% to the phenotype. Thus, there are hundreds of possible modifier genes that have the potential to combinatorially interact in hundreds of thousands of different ways to subtly modify atherosclerosis. The numbers get worse when myocardial infarction is used as a surrogate endpoint or a phenotypic extreme, because the number of potential genetically variant disease modifiers more than doubles (Figure 1). This vast number of genetic influences is analogous to acoustical pink noise. Nevertheless, GWAS has the potential to identify a powerful genetic event if the genotype–phenotype signal is sufficiently greater than the genetic noise. The presence of rs1333049 and its neighbors at 9p21 may make this more difficult, however, since this locus appears to be the 'Tiger Woods' of cardiovascular genes. When Tiger is playing, the rest of the field is largely indistinguishable. Likewise, when the most powerful association with coronary artery disease is on 9p21.3, it is hard to see any others. This may be due to the phenotypic heterogeneity of coronary artery disease and myocardial infarction. As discussed above, SNPs identified in GWAS of coronary artery disease or myocardial infarction often fail to replicate this association when investigated in a cohort with a coronary artery disease–risk phenotype such as hypercholesterolemia [43]. In contrast, SNPs identified by GWAS of intermediate coronary artery disease-risk phenotypes, such as the 11 SNPs identified by Willer and colleagues' GWAS analysis of hypercholesterolemia, are often found to have increased frequency in cohorts of individuals with coronary artery disease [23].

Coronary artery disease, myocardial infarction, and heart failure may, therefore, be diseases for which neither GWAS, nor single candidate gene studies are optimal. Fortunately, there are new approaches that combine the advantages of both. Microarray platforms are becoming available for high resolution SNP analysis of thousands of biologically relevant genes, such as the Illumina IBC chip (55,000 SNPs in 2100 genes related to blood pressure, myocardial infarction, inflammation, heart failure and diabetes) [49]. The next step, which is imminent, is using second generation DNA sequencing technology to resequence multiple candidate genes from pooled case and control DNA with individual bar coding. This will make it possible to generate comprehensive individual sequences within candidate 'secondary' loci in thousands of individuals or entire study cohorts and, perhaps, to identify new rare mutations with large effects. Careful selection of study populations with defined phenotypes, such as affected/unaffected twin studies, or large subgroup meta-analyses of individuals with coronary artery disease grouped into homogenous risk factor groups, may improve the signal:noise ratio. Likewise, comparing extreme phenotypes, such as individuals with extensive coronary atherosclerosis without myocardial infarction versus

individuals with infarction and minimal or no coronary atherosclerosis, is also likely to provide new genetic associations.

Bibliography

Papers of special note have been highlighted as:

- of interest
- of considerable interest

1. Rosamond W, Flegal K, Furie K, et al. Heart disease and stroke statistics – 2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*. 2008; 117(4):E25–E146. [PubMed: 18086926]
2. Slack J, Evans KA. The increased risk of death from ischaemic heart disease in first degree relatives of 121 men and 96 women with ischaemic heart disease. *J Med Genet*. 1966; 3(4):239–257. [PubMed: 16175706]
3. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med*. 1994; 330(15):1041–1046. [PubMed: 8127331]
4. Goldstein JL, Brown MS. Familial hypercholesterolemia: identification of a defect in the regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity associated with overproduction of cholesterol. *Proc Natl Acad Sci USA*. 1973; 70(10):2804–2808. [PubMed: 4355366]
5. Goldstein JL, Dana SE, Brunschede GY, Brown MS. Genetic heterogeneity in familial hypercholesterolemia: evidence for two different mutations affecting functions of low-density lipoprotein receptor. *Proc Natl Acad Sci USA*. 1975; 72(3):1092–1096. [PubMed: 236556]
6. Brown MS, Goldstein JL. Expression of the familial hypercholesterolemia gene in heterozygotes: mechanism for a dominant disorder in man. *Science*. 1974; 185(4145):61–63. [PubMed: 4366052]
7. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N Engl J Med*. 2000; 342(12):836–843. [PubMed: 10733371]
8. Ridker PM, Stampfer MJ, Rifai N. Novel risk factors for systemic atherosclerosis: a comparison of C-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. *JAMA*. 2001; 285(19):2481–2485. [PubMed: 11368701]
9. Peyser PA, Bielak LF, Chu JS, et al. Heritability of coronary artery calcium quantity measured by electron beam computed tomography in asymptomatic adults. *Circulation*. 2002; 106(3):304–308. [PubMed: 12119244]
10. Burkett EL, Hershberger RE. Clinical and genetic issues in familial dilated cardiomyopathy. *J Am Coll Cardiol*. 2005; 45(7):969–981. [PubMed: 15808750]
11. Olson TM, Michels VV, Thibodeau SN, Tai YS, Keating MT. Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. *Science*. 1998; 280(5364):750–752. [PubMed: 9563954]
12. Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. *PLoS Biol*. 2007; 5(10):E254. [PubMed: 17803354]
13. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008; 452(7189):872–876. [PubMed: 18421352]
14. Topol EJ, Frazer KA. The resequencing imperative. *Nat Genet*. 2007; 39(4):439–440. [PubMed: 17392801]
15. Yasuda K, Miyake K, Horikawa Y, et al. Variants in *KCNQ1* are associated with susceptibility to Type 2 diabetes mellitus. *Nat Genet*. 2008; 40:1092–1097. [PubMed: 18711367]
16. Cauchi S, El Achaby Y, Choquet H, et al. *TCF7L2* is reproducibly associated with Type 2 diabetes in various ethnic groups: a global meta-analysis. *J Mol Med*. 2007; 85(7):777–782. [PubMed: 17476472]
17. Cauchi S, Froguel P. *TCF7L2* genetic defect and Type 2 diabetes. *Curr Diab Rep*. 2008; 8(2):149–155. [PubMed: 18445358]

18. Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in *KCNQ1* are associated with susceptibility to Type 2 diabetes in East Asian and European populations. *Nat Genet.* 2008; 40:1098–1102. [PubMed: 18711366]
19. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007; 447(7148):1087–1093. [PubMed: 17529967]
20. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet.* 2007; 39(7):870–874. [PubMed: 17529973]
21. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet.* 2007; 39(5):631–637. [PubMed: 17401366]
22. Burkhardt R, Kenny EE, Lowe JK, et al. Common SNPs in *HMGCR* in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. *Arterioscler Thromb Vasc Biol.* 2008; 28(11):2078–2084. [PubMed: 18802019]
23. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet.* 2008; 40(2):161–169. [PubMed: 18193043]
24. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* 2008; 41(1):56–65. Meta-analyses of genome-wide association studies (GWAS) performed on approximately 20,000 individuals with the coronary artery disease risk phenotype of dyslipidemia. [PubMed: 19060906]
25. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* 2008; 41(1):47–55. Meta-analyses of GWAS performed on approximately 20,000 individuals with the coronary artery disease risk phenotype of dyslipidemia. [PubMed: 19060911]
26. Samani NJ, Burton P, Mangino M, et al. A genomewide linkage study of 1,933 families affected by premature coronary artery disease: The British Heart Foundation (BHF) Family Heart Study. *Am J Hum Genet.* 2005; 77(6):1011–1020. [PubMed: 16380912]
27. Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin- α gene that are associated with susceptibility to myocardial infarction. *Nat Genet.* 2002; 32(4):650–654. What may be considered the first large ‘genome-wide’ study in individuals with myocardial infarction. [PubMed: 12426569]
28. Helgadottir A, Manolescu A, Thorleifsson G, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet.* 2004; 36(3):233–239. [PubMed: 14770184]
29. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007; 447(7145):661–678. One of several landmark GWAS studies of individuals with coronary disease that identified the 9p21.3 locus. [PubMed: 17554300]
30. Hinohara K, Nakajima T, Takahashi M, et al. Replication of the association between a chromosome 9p21 polymorphism and coronary artery disease in Japanese and Korean populations. *J Hum Genet.* 2008; 53(4):357–359. [PubMed: 18264662]
31. Singh NN, Ramji DP. The role of transforming growth factor- β in atherosclerosis. *Cytokine Growth Factor Rev.* 2006; 17(6):487–499. [PubMed: 17056295]
32. Greenland P, Knoll MD, Stamler J, et al. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *JAMA.* 2003; 290(7):891–897. [PubMed: 12928465]
33. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science.* 2007; 316(5830):1488–1491. One of several landmark GWAS studies of individuals with coronary disease that identified the 9p21.3 locus. [PubMed: 17478681]
34. Helgadottir A, Thorleifsson G, Magnusson KP, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet.* 2008; 40(2):217–224. [PubMed: 18176561]
35. Zuchner S, Gilbert JR, Martin ER, et al. Linkage and association study of late-onset Alzheimer disease families linked to 9p21.3. *Ann Hum Genet.* 2008; 72(Pt 6):725–731. [PubMed: 18761660]

- 36■. Helgadóttir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007; 316(5830):1491–1493. One of several landmark GWAS studies of individuals with coronary disease that identified the 9p21.3 locus. [PubMed: 17478679]
37. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007; 357(5):443–453. [PubMed: 17634449]
38. Schunkert H, Gotz A, Braund P, et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation*. 2008; 117(13): 1675–1684. [PubMed: 18362232]
39. Shen GQ, Li L, Rao S, et al. Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler Thromb Vasc Biol*. 2008; 28(2):360–365. [PubMed: 18048766]
40. Shen GQ, Rao S, Martinelli N, et al. Association between four SNPs on chromosome 9p21 and myocardial infarction is replicated in an Italian population. *J Hum Genet*. 2008; 53(2):144–150. [PubMed: 18066490]
41. Abdullah KG, Li L, Shen GQ, et al. Four SNPs on chromosome 9p21 confer risk to premature, familial CAD and MI in an American Caucasian population (GeneQuest). *Ann Hum Genet*. 2008; 72(Pt 5):654–657. [PubMed: 18505420]
42. van der Net JB, Oosterveer DM, Versmissen J, et al. Replication study of 10 genetic polymorphisms associated with coronary heart disease in a specific high-risk population with familial hypercholesterolemia. *Eur Heart J*. 2008; 29(18):2195–2201. [PubMed: 18599554]
- 43■. Broadbent HM, Peden JF, Lorkowski S, 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*. 2008; 17(6):806–814. Authors speculate on a mechanism for the high risk 9p21 haplotype. [PubMed: 18048406]
44. Ellinor PT, Sasse-Klaassen S, Probst S, et al. A novel locus for dilated cardiomyopathy, diffuse myocardial fibrosis, and sudden death on chromosome 10q25–26. *J Am Coll Cardiol*. 2006; 48(1): 106–111. [PubMed: 16814656]
45. Song L, DePalma SR, Kharlap M, et al. Novel locus for an inherited cardiomyopathy maps to chromosome 7. *Circulation*. 2006; 113(18):2186–2192. [PubMed: 16651466]
46. Rampazzo A, Nava A, Erne P, et al. A new locus for arrhythmogenic right ventricular cardiomyopathy (ARVD2) maps to chromosome 1q42-q43. *Hum Mol Genet*. 1995; 4(11):2151–2154. [PubMed: 8589694]
47. Tiso N, Stephan DA, Nava A, et al. Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy Type 2 (ARVD2). *Hum Mol Genet*. 2001; 10(3):189–194. [PubMed: 11159936]
- 48■. Larson MG, Atwood LD, Benjamin EJ, et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med Genet*. 2007; 8(Suppl. 1):S5. To date, the only GWAS of unrelated individuals with heart failure. [PubMed: 17903304]
49. Keating BJ, Tischfield S, Murray SS, et al. Concept, design and implementation of a cardiovascular gene-centric 50K SNP array for large-scale genomic association studies. *PLoS ONE*. 2008; 3(10):E3583. [PubMed: 18974833]

Executive summary

- Coronary artery disease and heart failure are polygenic, only partially genetic and phenotypically heterogeneous.
- The 9p21.3 locus has been identified and validated in several large genome-wide association study (GWAS) analyses of coronary artery disease.
- To date, the molecular/biological mechanism of this association remains unclear.
- To date only one GWAS has been reported on unrelated individuals with heart failure but no association achieved statistical significance appropriate for the multiple comparisons performed in the analyses.
- Coronary artery disease, myocardial infarction, and heart failure may be diseases for which neither GWAS, nor single candidate gene studies are optimal.

Figure 1. Multiplicative modifying factors (light gray boxes) affecting the pathophysiologic cascade of coronary artery disease, myocardial infarction and heart failure.

Table 1

Genome-wide association studies in coronary artery disease.

Study	Genome-wide association study platform	SNPs identified	Chromosomal location	Annotated gene in region (if any)	Population	Ref.
Wellcome Trust	Affymetrix 5.0 array	rs1333049	9p21	<i>CDKN2A, CDKN2B, MTAP</i>	Caucasian	[29]
Hinohara <i>et al.</i>		rs1333049	9p21	<i>CDKN2A, CDKN2B, MTAP</i>	Japanese, Korean	[30]
McPherson <i>et al.</i>	Custom oligonucleotide arrays	rs10757274, rs2383206	9p21	<i>CDKN2A, CDKN2B, MTAP</i>	Caucasian	[33]
Helgadottir <i>et al.</i>	Illumina Hap300 array	rs1333040, rs2383207, rs10116277, rs10757278	9p21	<i>CDKN2A, CDKN2B, MTAP</i>	Caucasian	[36]
Samani <i>et al.</i>	GeneChip® Human Mapping 500K array set	rs2943634	2q36.3	Pseudogene	Caucasian	[37]
Samani <i>et al.</i>	GeneChip Human Mapping 500K array set	rs6922269	6q25.1	<i>MTHFD1L</i>	Caucasian	[37]
Samani <i>et al.</i>	GeneChip Human Mapping 500K array set	rs1333049	9p21.3	<i>CDKN2A, CDKN2B, MTAP</i>	Caucasian	[37]
Samani <i>et al.</i>	GeneChip Human Mapping 500K array set	rs599839, rs17465637, rs501120, rs17228212		<i>PSRC1, MIA3, CXCL12, SMAD3</i>	Caucasian	[37]
Shen <i>et al.</i>		rs10757274, rs2383206, rs2383207, rs10757278	9p21	<i>CDKN2A, CDKN2B, MTAP</i>	South Korean	[39]
Abdullah <i>et al.</i>		rs10757274, rs2383206, rs2383207, rs10757278	9p21	<i>CDKN2A, CDKN2B, MTAP</i>	Caucasian	[41]

Table 2

Genome-wide association studies in heart failure.

Study	Genome-wide association study platform	SNPs identified	Chromosomal location	Annotated gene in region (if any)	Population	Ref.
Larson <i>et al.</i>	Affymetrix 100K array	rs740363	10		Caucasian	[48]
Larson <i>et al.</i>		rs10515869	5		Caucasian	[48]
Larson <i>et al.</i>		rs939698	1	<i>R YR2</i>	Caucasian	[48]