Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour

John F. Peden and Martin Farrall∗

Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

Received June 24, 2011; Revised and Accepted August 24, 2011

Coronary artery disease (CAD) is the leading cause of death worldwide. Affected individuals cluster in families in patterns that reflect the sharing of numerous susceptibility genes. Genome-wide and large-scale gene-centric genotyping studies that involve tens of thousands of cases and controls have now mapped common disease variants to 34 distinct loci. Some coronary disease common variants show allelic heterogeneity or copy number variation. Some of the loci include candidate genes that imply conventional or emerging risk factor-mediated mechanisms of disease pathogenesis. Quantitative trait loci associations with risk factors have been informative in Mendelian randomization studies as well as fine-mapping of causative variants. But, for most loci, plausible mechanistic links are uncertain or obscure at present but provide potentially novel directions for research into this disease’s pathogenesis. The common variants explain ∼4% of interindividual variation in disease risk and no more than 13% of the total heritability of coronary disease. Although many CAD genes are presently undiscovered, it is likely that larger collaborative genome-wide association studies will map further common/low-penetrance variants and hoped that low-frequency or rare high-penetrance variants will also be identified in medical resequencing experiments.

INTRODUCTION

Coronary artery disease (CAD) is the most frequent cause of death in high-income countries and the second most common cause of death in medium and low-income countries (1). It most commonly presents clinically in cases of angina pectoris and myocardial infarction (heart attack), which are due to atherosclerotic plaques that develop progressivly as we age and occasionally rupture. Genetic epidemiological studies of family history and twin concordance studies are consistent with an underlying multifactorial model of disease susceptibility with a significant polygenic component. This is complemented by genetic analysis of heritable conventional risk factors such as low-density lipoprotein (LDL)-cholesterol and systolic blood pressure, which collectively might explain a minor portion of coronary disease risk (2).

Over the past 4 years, researchers have completed several genome-wide association studies (GWASs) to map underlying common susceptibility variants for coronary disease. In parallel with GWASs of other complex diseases, it was soon apparent that typical effect sizes for individual single-nucleotide polymorphisms (SNPs) were fairly small, so large sample sizes would be required for reliable gene mapping. This has encouraged collaboration between individual research groups and led to the formation of consortia to pool the results of GWASs using meta-analysis techniques. Progress has been facilitated by the availability of phased haplotype training sets (notably, from the HapMap project: http://hapmap.ncbi.nlm.nih.gov/downloads/phasing) and the accompanying genotype imputation software (for example, MACH: www.sph.umich.edu/csg/abecasis/MACH/index.html or IMPUTE: mathgen.stats.ox.ac.uk/impute/impute_v2.html). These population genetic resources and statistical genetic tools provide an efficient solution to the fact that individual GWASs are often carried out on different SNP arrays with variable SNP overlap.

All this effort came to a crescendo this year with the publication of two papers from the CARDIoGRAM (3) and C4D (4) consortia that together scanned nearly 40K coronary disease cases for susceptibility gene signals. Combined with the
highlight an underlying gene. For example, \textit{LIPA} increased expression of (eQTL) SNP, with the CAD risk allele correlated with strongly correlated with expression quantitative trait loci and triglycerides. The lead CAD risk SNP in lipase A, which catalyzes the hydrolysis of cholesteryl esters replication experiments in the Chinese Han population. powered to detect variants that were common to both ancestry and India as well as Europeans; these designs were optimallycovery experiments included South Asians from Pakistan of European descent. The C4D (4) and gene-centric (8) dis-

\begin{table}

<table>
<thead>
<tr>
<th>Chr</th>
<th>Position</th>
<th>Locus</th>
<th>SNP</th>
<th>References</th>
<th>Reported effect</th>
<th>SNP-specific heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EAF</td>
<td>OR</td>
</tr>
<tr>
<td>1</td>
<td>55 496 039</td>
<td>PCSK9</td>
<td>rs11206510</td>
<td>MIGen (36)</td>
<td>0.82</td>
<td>1.08</td>
</tr>
<tr>
<td>1</td>
<td>56 962 821</td>
<td>PPAP2B</td>
<td>rs17114036</td>
<td>CARDioGRAM (3)</td>
<td>0.91</td>
<td>1.17</td>
</tr>
<tr>
<td>1</td>
<td>109 822 166</td>
<td>SORT1</td>
<td>rs599839</td>
<td>Samani et al. (58), MIGen (36)</td>
<td>0.78</td>
<td>1.11</td>
</tr>
<tr>
<td>1</td>
<td>222 823 529</td>
<td>MIA3</td>
<td>rs17465637</td>
<td>Samani et al. (58), MIGen (36)</td>
<td>0.74</td>
<td>1.14</td>
</tr>
<tr>
<td>2</td>
<td>44 072 576</td>
<td>ABCG8</td>
<td>rs4299376</td>
<td>HumanCVD (8)</td>
<td>0.29</td>
<td>1.09</td>
</tr>
<tr>
<td>2</td>
<td>203 745 885</td>
<td>WDR12</td>
<td>rs6725887</td>
<td>MIGen (36)</td>
<td>0.15</td>
<td>1.14</td>
</tr>
<tr>
<td>2</td>
<td>138 119 952</td>
<td>MRAS</td>
<td>rs2306374</td>
<td>Erdmann et al. (59)</td>
<td>0.18</td>
<td>1.12</td>
</tr>
<tr>
<td>3</td>
<td>131 867 702</td>
<td>IL5</td>
<td>rs2706399</td>
<td>HumanCVD (8)</td>
<td>0.48</td>
<td>1.02</td>
</tr>
<tr>
<td>6</td>
<td>11 774 583</td>
<td>C6orf105</td>
<td>rs6903956*</td>
<td>Wang et al. (7)</td>
<td>0.07</td>
<td>1.51</td>
</tr>
<tr>
<td>6</td>
<td>12 927 544</td>
<td>PHACTR1</td>
<td>rs12526453</td>
<td>MIGen (36)</td>
<td>0.67</td>
<td>1.10</td>
</tr>
<tr>
<td>6</td>
<td>35 034 800</td>
<td>ANK1S1A</td>
<td>rs17609940</td>
<td>CARDioGRAM (3)</td>
<td>0.75</td>
<td>1.07</td>
</tr>
<tr>
<td>7</td>
<td>134 214 525</td>
<td>TCF21</td>
<td>rs1290287</td>
<td>CARDioGRAM (3)</td>
<td>0.62</td>
<td>1.08</td>
</tr>
<tr>
<td>7</td>
<td>160 961 137</td>
<td>LPA</td>
<td>rs3798220</td>
<td>Clarke et al. (30)</td>
<td>0.02</td>
<td>1.92</td>
</tr>
<tr>
<td>8</td>
<td>161 010 118</td>
<td>LPA</td>
<td>rs10455872</td>
<td>Clarke et al. (30)</td>
<td>0.07</td>
<td>1.70</td>
</tr>
<tr>
<td>10</td>
<td>97 244 545</td>
<td>7q22</td>
<td>rs10953541</td>
<td>CARDioGRAM (3)</td>
<td>0.80</td>
<td>1.08</td>
</tr>
<tr>
<td>10</td>
<td>129 663 496</td>
<td>ZC3H1C</td>
<td>rs11556924</td>
<td>CARDioGRAM (3)</td>
<td>0.62</td>
<td>1.09</td>
</tr>
<tr>
<td>10</td>
<td>126 495 818</td>
<td>TRIB1</td>
<td>rs10808546</td>
<td>C4D 2011 (4), Erdmann (58), MIGen (36)</td>
<td>0.46</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Chr, chromosome; Position, position (in bp) on GRCh37/hg19 (Genome Reference Consortium February 2009); EAF, effect allele frequency; OR, odds ratio; \(h_{\text{SNP}}^2\), SNP-specific heritability estimates are shown for three disease prevalence estimates; \(K_p\), disease prevalence estimate for SNP-specific heritability estimate; \(h_{\text{total}}^2\), total SNP-encoded heritability for each disease prevalence estimate are shown in bold type.

*Most locus assignments are provisional based on proximity (see text).

**Effect allele frequency and odds ratio are given for Chinese Han population.**

results from other recent large-scale studies (5 - 7), 35 common coronary disease variants have been robustly mapped by GWASs (3 - 7) or gene-centric SNP arrays (8) (Table 1). These results are mostly based on cases and controls of European descent. The C4D (4) and gene-centric (8) discovery experiments included South Asians from Pakistan of European descent; these designs were optimally powered to detect variants that were common to both ancestry groups. Wang et al. (7) carried out their GWAS discovery and replication experiments in the Chinese Han population. For some of the SNPs, there is circumstantial evidence to highlight an underlying gene. For example, \textit{LIPA} encodes lipase A, which catalyzes the hydrolysis of cholesteryl esters and triglycerides. The lead CAD risk SNP in \textit{LIPA} (4,8) is strongly correlated with expression quantitative trait loci (eQTL) SNP, with the CAD risk allele correlated with increased expression of \textit{LIPA} mRNA in monocytes (9) and liver (4), suggesting a functional relationship between the disease association signal and this candidate gene. However, for most of the SNPs mapped by GWASs, it is difficult to implicate an underlying gene. Inspection of recombination frequency maps derived from the HapMap project suggest genetic boundaries defined by recombination hotspots, which drives the traditional approach of fine-mapping using haplotype block information (10). However, there is ample evidence that cis-regulatory mechanisms in the soma can operate over tens, hundreds or even millions of base pairs and are presumably unaffected by meiotic recombination (11). It is becoming possible to define functional genomic boundaries based on chromatin architecture-related factors, such as CCCTC-binding factor sites (12) that are found in the vast majority of vertebrate insulator elements. We suspect that, for most coronary disease loci, it will take some time to complete the sequence of functional genomics experiments that will be

\[ h_{\text{total}}^2 = 3.30 \times 4.27 = 5.29 \]
required to confidently lay the blame of disease susceptibility on a specific underlying gene. The loci shown in Table 1 should therefore be interpreted as provisional assignments based mainly on proximity (nearest coding sequence).

It is noteworthy that the lists of confidently assigned genes across the recent crop of studies show little overlap, which at first glance seems unlikely to be due to differences in tagging SNP coverage, phenotypic heterogeneity or ancestry (these studies are heavily weighted towards European ancestry). We suspect that it most likely has its roots in the small effects conferred by the susceptibility genes (each allele increases risk by $\sim5\%$). Such small gene effects would, in a GWAS discovery experiment comprising 20K cases and 20K controls, be expected to have only a 5% probability of passing even a modest stage-1 (tentative discovery) threshold of $P < 0.00001$. Therefore, in absolute terms, this low power to detect an individual gene means that different, and only occasionally overlapping, sets of susceptibility loci are likely to emerge from similarly sized studies (13,14). Of course, even for loci that have surpassed the de facto genome-wide significance threshold of $5 \times 10^{-8}$, there is still an appreciable chance that one or more of the 34 loci will prove to be false-positives (15). We expect based on binomial theory that the maximum number of false-positive associations in Table 1 will be three or fewer. Skol et al. (16) pointed out that a combined analysis of stage 1 discovery data (based on GWAS SNP arrays) and stage 2 validation data (a subset of the most promising SNPs from stage 1) is more efficient than attempting formal replication in stage 2 and such an analysis is now standard practise. However, some researchers, mindful of historical difficulties of interpreting complex genetic data (17), prefer to apply a cautious approach to secure robust (i.e. taking into account stage 2 multiple testing penalties) and independent (of GWAS discovery) replication data (3,4).

GWAS is a ‘hypothesis-free’ approach to the study of complex diseases, which depends on mutations that unpredictably occurred in previous generations, and purely by chance (genetic drift) or by (balancing) selection is now common. As such the loci that are identified provide a framework for what we know and do not know about pathogenesis from other hypothesis-led (e.g. physiologic, biochemical or cell biological) experiments. In the case of coronary disease genetics, we have examples that illustrate complex genetic architectural features such as allelic heterogeneity, pleiotropy, risk factor QTLs, copy number variation (CNV) or synthetic associations, which are discussed below. There are findings that could lead to tangible clinical benefits relatively soon (LPA and SORT1). But mainly there are leads to point researchers in unanticipated directions, at least some of which we hope will provide novel biological insights into how atherosclerotic plaques develop and rupture.

**HOW MUCH HERITABILITY HAS BEEN MAPPED?**

Manolio et al. (18) have pointed out that despite the success of GWAS in mapping common susceptibility variants for many multifactorial diseases, collectively these variants typically explain a modest fraction of the total heritability of these conditions. The accuracy of such calculations depends on the fidelity of a series of locus-specific heritability estimates as well as the total (i.e. measured plus unmeasured) heritability.

Locus-specific heritability estimates are based on odds ratio estimates that assume that the lead SNP at a locus accurately tags the disease-causing variant. No correction is usually made for potential biases due to the ‘winner’s curse’ (19) or to signal attenuation due to clinically unscreened control data (20). External (to the case–control data) epidemiological information on disease prevalence is required if the multifactorial threshold model is to be used to calculate locus-specific heritabilities. Prevalence estimates can vary substantially with clinical phenotype, sex, and age and are drifting over time with changing environmental risk factor exposures (www.heartstats.org); we suggest that a range of prevalences between 2 and 10% is relevant to coronary disease GWAS case series.

Invaluable coronary disease heritability data are derived from the longitudinal study of over 20K twins in the Swedish Twin Registry (Karolinska Institute, Stockholm, Sweden). The total heritability of coronary disease was estimated for angina in men as 39% (95% CI 29–49%) and in women as 43% (8–51%) (21) and for death from coronary disease in men as 57% (45–69%) and in women as 38% (26–50%) (22). Genetic association studies generally use samples collected from survivors of disease, for ethical and other pragmatic reasons. For coronary disease, case series are clinically heterogeneous if they include different diagnostic subgroups (e.g. chronic stable angina or myocardial infarction) with subtly differing pathologies that might have an impact on susceptibility. So, we propose that a heritability estimate of 40% will encompass the clinical heterogeneity across typical GWAS case series.

Taking all of these issues into account, we estimate that between 8 and 13% of the total heritability of coronary disease can be explained by the 35 common variants (Table 1). So, the vast majority of the heritability is currently unexplained. X-linked single-gene disorders have intrinsic advantages for gene mapping, so it is a pity that the X-chromosome has sometimes been overlooked in the search for coronary disease susceptibility loci (contrast the CARDiOGRAM study which was based on imputed data with C4D which was based solely on genotype data). This can easily be resolved as appropriate analytic means (i.e. phased haplotype training sets and imputation software) are now freely available and have proved productive in the investigation of the X chromosome in other disease areas [e.g. type 2 diabetes (23)].

Although there was little evidence for non-additive genetic effects in the aforementioned twin studies of coronary disease, we note that classic MZ/DZ concordance studies have very limited power to identify non-additive variance components (24). Moreover, dominance and epistatic variance components are inevitably confounded in this design (25). Consequently, there seems no reason why epistatic and high-penetrance/low-frequency alleles should not explain a portion of the missing heritability. Indeed, some of the linkages detected in earlier affected-sib-pair studies (26,27) might be conferred by low-frequency alleles with or without allelic heterogeneity. For example, the locus on chromosome 17p reported by PROCARDIS (27) was associated with a sibling recurrence risk ratio ($\lambda_{sb}$) of 1.29 that could theoretically be conferred by a dominant, low-
frequency (0.5%) allele of intermediate penetrance (17.2%) and with a 1.8% phenocopy rate. Such linkage signals are usually intractable to conventional GWAS based on common SNPs (28) but might be resolved by means of resequencing-based analyses or genotyping arrays with good coverage of low-frequency variants.

**SUSCEPTIBILITY ENCODED BY CNV**

The development of array-based methods to systematically study CNV has allowed researchers to study the role of this rich source of genetic variation in common multifactorial diseases (29). It is ironic that these high-throughput techniques with genome-wide coverage of CNVs have overlooked an exemplar of common disease susceptibility namely that encoded by the apolipoprotein(a) (LPA) gene. This gene includes a highly variable number of kringle IV-2 repeats (range at least 12–44) which result in numerous isoforms that can be typed by protein electrophoresis (30) or genomically quantified by qPCR (30–32). Two SNPs that tag short isoform alleles that are encoded by relatively low copy numbers of kringle IV-2 sequences show strong associations with high lipoprotein(a) levels and with coronary disease risk (OR = ~1.5) (30). These SNPs were not included on commonly used GWAS SNP arrays (but were fortuitously included in the design of the HumanCVD gene-centric SNP array) and have been calcified to genotype imputation due to their frequency or linkage disequilibrium properties. Consequently, they have not been assessed in GWAS meta-analyses of coronary disease susceptibility.

**ALLELIC HETEROGENEITY AND PLEIOTROPY**

Allelic heterogeneity is a regular feature of complex diseases and traits. For example, multiple independent signals were detected at 19 of 180 height QTL (33). Researchers have systematically scanned for secondary association signals of coronary disease (by conditioning on the lead SNP at each locus), an approach that identified multiple independent SNP signals in LPA (30). The coronary disease associations in PCKS9 detected by the non-synonymous R46L SNP rs11591147 (34,35) and a non-coding SNP rs11206510 (36) appear to be independent as the two SNPs show little linkage disequilibrium (r² = 0.04 in PROCARDIS Human CVD data). A possible example of allelic heterogeneity arises for SNPs rs3825807 (3) and rs1994016 (6) that map to alternate flanks of the ADAMTS7 gene on chromosome 15 to rs4380028 (4) and are in moderate linkage disequilibrium (r² = ~0.50).

Coronary disease shows substantial clinical heterogeneity that is reflected in morphological differences in the athero-sclerotic plaques (37) that might in turn reflect differences in inherited susceptibilities. Plaque rupture and subsequent coronary thrombosis causes acute coronary syndromes such as myocardial infarction, thereby motivating searches for genes that might influence plaque stability. Reilly et al. (6) undertook a GWAS of coronary disease patients with angiographic disease contrasting those cases that had suffered myocardial infarction with those who had not. This study mapped a novel association to the ABO blood group system with SNPs that strongly tag the O allele. The CARDIoGRAM consortium (3), which studied a mixture of coronary disease cases of which two-thirds had suffered a myocardial infarction and a mixture of screened and unscreened controls, also mapped a susceptibility signal to the ABO system. Their lead SNP is only in moderate linkage disequilibrium (r² = 0.39) with the Reilly et al. (6) signal so may be due to allelic heterogeneity or pleiotropy. We must await further fine-mapping studies of the ABO and ADAMTS7 loci to fully understand the details of these associations.

**QTL MAPPING AND CONVENTIONAL RISK FACTORS**

Conventional risk factors for coronary disease such as circulating lipid levels and blood pressure are heritable (quantitative) traits. There has been much effort in mapping common QTL for these traits using GWAS or other large-scale SNP arrays in population-based as well as case–control samples (38–40). Technical difficulties such as uncontrolled fasting status or on-treatment measurements have been largely overcome to produce a rich crop of risk factor QTL. Notable examples of overlap with coronary disease loci (Table 2) include TRIB1 (Drosophila tribbles homologue, a gene that interacts with the mitogen-activated protein kinase cascade), which has pleiotropic effects on circulating triglyceride, LDL- and HDL-cholesterol levels (40) and CYP17A1 (17α-hydroxylase gene involved in steroid hormone metabolism) which is a systemic blood pressure QTL (38,39). However, most (22 of 34) of the coronary disease loci do not show convincing risk factor QTL effects (Table 2). It may be that novel heritable intermediate phenotypes will eventually be identified that will explain some of the disease associations; this should lead to informative insights into pathological mechanisms. Indeed, expectations that coronary genes would be involved in innate immunity or thrombosis have not emerged from large-scale genetic association studies to date.

**MOVING FROM ASSOCIATED LOCUS TO CAUSATIVE GENE**

Robust assignments of common susceptibility variants are to be applauded, but it may take some time to resolve the underlying molecular genetic mechanisms. For instance, the first coronary disease locus to confidently emerge from GWAS mapped to chromosome 9p21 to a region that was initially believed to be a gene desert. However, it was quickly recognized in fine-mapping studies (10) that the associated region, which was of prior interest to cancer genetics researchers (41,42), was potentially linked to neighbouring cyclin-dependent kinase inhibitor genes CDKN2A (which has multiple synonyms including p16, see www.genecards.org/cgi-bin/carddisp.pl?gene=CDKN2A for more details) and CDKN2B (p15, see www.genecards.org/cgi-bin/carddisp.pl?gene=CDKN2B). Subsequent studies of murine models have highlighted a Cdkn2a/b-mediated mechanism involving smooth muscle cell proliferation (43). Studies of human transcription enhancer elements propose that the large non-coding antisense RNA molecule CDKN2ABAS, which is also known
by the monikers ANRIL and CDKN2B-AS1, is involved in the long-range transcriptional regulation of several genes, including CDKN2A and CDKN2B, in vascular endothelial cell lines (44).

Another notable success story followed the overlap of LDL-cholesterol QTL and coronary disease association signals on chromosome 1p. Here eQTL were particularly informative to resolve the role of SORT1 which encodes sortilin from its neighbours (45). Parallel functional studies have implicated sortilin as a novel regulator of lipoprotein production in the liver (46), thereby providing a mechanistic link to the coronary disease susceptibility although some details of the mechanism need to be reconciled.

DISCREPANCIES BETWEEN MEASURED AND PREDICTED GENETIC RISK

Using results from proxy SNPs that are in almost complete linkage disequilibrium with each other, the CARDIoGRAM and C4D studies provide a joint SORT1 per-allele risk estimate equal to 1.12 (1.09–1.15). The measured per-allele QTL effect on LDL-cholesterol is equal to 0.145 mmol/l (0.135–0.155) (40). Substituting the latter QTL effect into the Framingham coronary heart disease risk equation (47) predicts a per-allele relative risk equal to 1.042 (1.039–1.045). So, the coronary disease risk estimate derived from GWAS is substantially higher than that predicted from the effect on LDL-cholesterol levels derived from long-term prospective studies.

A similar discrepancy was noted in a genotype risk score analysis of the Malmö Diet and Cancer study (48). Given the underlying sample sizes (>100K for the lipid QTL GWAS experiment) and the consequent precision of the risk factor effect sizes, it seems unlikely that these have been systematically underestimated. It is possible that the disease risk estimates have been overestimated [e.g. winner’s curse (19)] and/or the disease associations are only partially mediated through the accompanying risk factor QTL effect (i.e. pleiotropy). Moreover, the Framingham risk equations, which were based on US population data from 1970s onwards, were based on single (baseline)

<table>
<thead>
<tr>
<th>Locus a</th>
<th>QTL</th>
<th>Lead QTL SNP</th>
<th>CAD Risk SNP</th>
<th>r^2</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCSK9</td>
<td>LDL, TC</td>
<td>rs2479409</td>
<td>rs11206510</td>
<td>&lt;0.30</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>PPAP2B</td>
<td>LDL, TC</td>
<td>rs646776</td>
<td>rs99839</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>SORT1</td>
<td>LDL, TC</td>
<td>rs4299376</td>
<td>rs4299376</td>
<td>1.00</td>
<td>Teslovich et al. (40)</td>
</tr>
<tr>
<td>ANRIL</td>
<td>LDL, TC</td>
<td>rs4299376</td>
<td>rs86725887</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>CDKN2B-AS1</td>
<td>LDL, TC</td>
<td>rs6725887</td>
<td>rs2306374</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>MRAS</td>
<td>LDL, TC</td>
<td>rs760399</td>
<td>rs9033956*</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>IL5</td>
<td>LDL, TC</td>
<td>rs12526453</td>
<td>rs17609940</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>Csef105</td>
<td>LDL, TC</td>
<td>rs12190287</td>
<td>rs12190287</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>PHACTR1</td>
<td>LDL, TC</td>
<td>rs11556924</td>
<td>rs11556924</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>ANK31A</td>
<td>LDL, TC</td>
<td>rs10455872</td>
<td>rs10455872</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>TCF21</td>
<td>LDL, TC</td>
<td>rs1084651</td>
<td>rs1084651</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>LPA</td>
<td>LDL, TC</td>
<td>rs398220</td>
<td>rs398220</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>ABCG8</td>
<td>LDL, TC</td>
<td>rs164348</td>
<td>rs164348</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>WDR12</td>
<td>LDL, TC</td>
<td>rs10455872</td>
<td>rs10455872</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>MRAS</td>
<td>LDL, TC</td>
<td>rs10953541</td>
<td>rs10953541</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
<tr>
<td>CDKN2B-AS1</td>
<td>LDL, TC</td>
<td>rs11556924</td>
<td>rs11556924</td>
<td>0.91</td>
<td>Kathiresan et al. (64), Teslovich et al. (40)</td>
</tr>
</tbody>
</table>

r^2, measure of linkage disequilibrium between the lead QTL SNP and the lead risk SNP; LDL, LDL-cholesterol; HDL, HDL-cholesterol; TC, total cholesterol; TG, triglycerides.
aMost locus assignments are provisional based on proximity (see text).
bSNP rs1084651 had a >5% genotyping failure rate in the HapMap database Rel23.
cholesterol measurements. Cholesterol measurements are subject to short-term (e.g. variable fasting) as well as long-term variation (e.g. changes in diet). Such within-individual variation can result in the systematic underestimation of the strength of a risk factor association with disease, an epidemiological effect known as regression dilution bias (49). Whatever the explanation, these findings emphasize that genetic epidemiological inferences from cross-sectional data need cautious interpretation and that information from prospective studies (e.g. UK BIOBANK, www.ukbiobank.ac.uk) will be very informative.

GENETIC INSIGHTS FOR EMERGING RISK FACTORS

Common variants that show quantitative genetic variation for disease risk factors or intermediate phenotypes can probe the putative causal relationship between risk factor and disease. This Mendelian randomization (MR) information (50) is particularly useful when there are no drugs that specifically modulate the exposure. For instance, Lp(a) is an LDL particle that appears to be proatherogenic in cross-sectional and prospective epidemiological studies. But drugs such as niacin that reduce Lp(a) concentrations also beneficially increase the HDL levels. So, it can be difficult in randomized clinical trials (RCT) to unambiguously attribute any clinical benefit to specific mechanisms. Following a large-scale candidate gene study, two SNPs in the apolipoprotein(a) gene were shown to tag short isoform alleles that were strikingly associated with raised Lp(a) concentrations and coronary disease risk (30) (also discussed in CNV and allelic heterogeneity sections above). Simultaneous modelling of disease risk and quantitative genetic variation were consistent with a direct causal link, thus predicting that pharmacological lowering of Lp(a) levels will be beneficial to patients. Loci that carry triglyceride QTL such as TRIB1 and APOA1-C3-A4-A5 also show QTL effects for HDL and LDL. Consequently, these pleiotropic loci will not be useful for MR probing of the role of triglyceride, a well-studied lipid that is presently not routinely included in cardiovascular risk calculations.

CANDIDATE AND POSITIONALLY CLONED GENES

Before the GWAS epoch, there was much effort expended in scanning candidate genes, those genes with known or predicted functions that might be involved in coronary disease pathogenesis, for susceptibility variants. This research was supplemented by positional cloning experiments that were unbiased in terms of gene candidature (51). In comparison with the levels of statistical support required for GWAS, the evidence for most of the candidate genes was modest despite meta-analyses involving up to 36K subjects (52). For instance, genetic variation in the apolipoprotein E gene (APOE) has well-known effects on LDL-cholesterol and is a highly plausible coronary disease candidate gene. However, a meta-analysis of 17 studies with at least 500 cases that included 21 331 cases and 47 467 controls estimated the risk of carrying the protective ε2 allele (versus ε3/ε3 homozgyotes) as 0.80 (95% CI 0.70–0.90) (53); the significance of this association is approximately \( P = 0.0005 \), four orders of magnitude below genome-wide significance. GWAS, even when enhanced by genotype imputation, may not accurately tag specific candidate gene variants [e.g. LPA and rs10455872 (30)]. So, the absence of a GWAS signal cannot be assumed to negate prior candidate or positional cloned genes (8). Indeed, the design of the HumanCVD gene-centric SNP array has revealed several novel loci LIPA, IL5, TRIB1 and ABCG5/ABCG8 as well as finally robustly confirming the candidature of APOE (8).

FUTURE DIRECTIONS AND PROSPECTS

Models of the genetic architecture of complex traits (33) predict that large numbers of small effect susceptibility loci remain to be discovered, some of which should be tractable to well-powered GWAS. The momentum amongst researchers to meta-analyse GWAS data will be sustained as larger consortia (e.g. the recently merged CARDioGRAMplusC4D consortium) are formed. They can take full advantage of initiatives such as the Metabochip project, a custom array containing 196 725 SNPs that builds on the CARDioGRAM stage-1 discovery results. So, it is reasonable to expect that the number of common coronary disease variants will increase, although as each novel variant will be associated with increasingly tiny effects, it seems that the missing heritability gap will never be filled by common variants alone.

The compilation of an exhaustive list of human genetic variation through the dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP) and HapMap (http://hapmap.ncbi.nlm.nih.gov) projects is being systematically expanded by the 1000 Genomes project (http://www.1000genomes.org). This is particularly useful for low-frequency variants (minimum allele frequency <5%), which were largely absent from the early genome-wide SNPs arrays, which were designed to type common variants in GWAS. Synthetic associations due to tagging of rare variants by common SNPs or haplotypes composed of common SNPs can occasionally be detected (54); this is particularly useful if the rare variant might disrupt gene function (e.g. non-synonymous SNP). Consequently, analyses of imputed genotypes derived from the 1000 genomes project are an immediate research priority. This is encouraged by the detection of a haplotype association with coronary disease (55) that subsequently was partially explained as a synthetic association to an LPA SNP (30). Imputation-based analysis will be complemented by whole-genome or exome resequencing experiments aimed at identifying low-frequency variants and unique high-penetrance mutations; together, these approaches have the potential to reveal novel disease mechanisms.

Finally, as the list of mapped disease variants expands, and fine-mapping and functional genomic studies refine loci to resolve underlying genes, pathway and network analysis should prove useful to provide systems level insights into coronary disease pathogenesis (56,57).

Conflict of Interest statement. None declared.
FUNDING
This work was supported by the European Community Sixth Framework Program (LSHM-CT-2007-037273), the British Heart Foundation, the Oxford BHF Centre of Research Excellence and the Wellcome Trust (090532/Z/09/Z). Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

REFERENCES


