BACKGROUND: Genomewide association studies have led to an enormous boost in the identification of susceptibility genes for cardiovascular diseases. This review aims to summarize the most important findings of recent years.

CONTENT: We have carefully reviewed the current literature (PubMed search terms: “genome wide association studies,” “genetic polymorphism,” “genetic risk factors,” “association study” in connection with the respective diseases, “risk score,” “transcriptome”).

SUMMARY: Multiple novel genetic loci for such important cardiovascular diseases as myocardial infarction, hypertension, heart failure, stroke, and hyperlipidemia have been identified. Given that many novel genetic risk factors lie within hitherto-unsuspected genes or influence gene expression, these findings have inspired discoveries of biological function. Despite these successes, however, only a fraction of the heritability for most cardiovascular diseases has been explained thus far. Forthcoming techniques such as whole-genome sequencing will be important to close the gap of missing heritability.

© 2011 American Association for Clinical Chemistry

The widespread use of genomewide association studies (GWAS) in the last 5 years has led to an enormous acceleration in discoveries across the entire spectrum of cardiovascular phenotypes. In fact, the multitude of novel gene loci for cardiovascular disease identified in recent years through GWASs has been compared to a “gold rush” in cardiovascular genomics. This impression might have been triggered by the fact that in the 20 years before GWAS became widely available, most genetic associations were based on candidate genes. Although such candidate-gene studies have reported multiple positive findings for cardiovascular phenotypes since the early 1990s, few have actually been replicated and have withstood the test of time (1). Therefore, up until about 2006, the search for causal genes in cardiovascular disease could be compared to a search for the proverbial needle in a haystack.

The current picture in 2011 is vastly different, and many new loci have been identified over the last 5 years for most important cardiovascular diseases and phenotypes, including coronary artery disease (CAD), hypertension, heart failure, hyperlipidemia, and stroke.

Building on a substantive body of work that has been described over the past few years, we discuss the current knowledge of genomics in cardiovascular disease. We provide an overview of the promises and pitfalls of the GWAS approach and systematically review the current knowledge of GWASs in cardiovascular disease. Table 1 provides an overview of important terms used in this review.

GWASs—Technical Aspects

Genetic-association studies gather a large number of individuals affected by a particular trait, such as CAD, myocardial infarction (MI), or related quantitative traits, and search for individual marker alleles (or genotypes) that are more frequent in individuals with disease than in healthy control individuals (2). The results of the Human Genome Project (3) and the International HapMap Project (4) have facilitated the development of a public catalogue of the most common form of human genomic variation, single-nucleotide polymorphisms (SNPs). In addition, this SNP information is being used in the characterization of linkage disequilibrium (LD) structures of the genome (5), thereby providing the foundation for large-scale genotyping and computational technologies, an approach called a GWAS.

GWASs are facilitated by the availability of affordable whole-genome SNP panels and large-scale genotyping technologies (e.g., Affymetrix or Illumina), and powerful association-analysis methods and software.
Association studies, which can be performed on unrelated individuals, rely on the LD between tested loci and disease-predisposing variants at the level of the total population. The rationale of this approach is that if unknown disease-predisposing variants are present somewhere in the genome, they might be detected via their LD with tagging SNPs represented on the genotyping array (5).

By design, GWASs provide an unbiased survey of the effects of common genetic variants. SNPs chosen for GWASs typically have a minor-allele frequency $\geq 0.05$ and are selected to “tag” the most common haplotypes. The power of detection depends directly on the sample size of the study population, the minor-allele frequency, the strength of LD between the SNP and the causal variants, and the effect sizes of the alleles (6).

The effect sizes associated with SNPs identified through GWASs are rather small, with odds ratios usually between 0.7 and 1.3 (7). Therefore, large sample sizes are necessary to uncover the genetic contributions of common variants (8). Another key aspect for a successful GWAS is the quality of phenotypic characterization. Because complex diseases such as cardiovascular disease are characterized by phenotypic heterogeneity (e.g., phenotypes may vary with respect to age or disease onset), detailed and standardized phenotyping is crucial (9, 10). Similar to the effects of genotyping errors, even modest levels of errors in the phenotypic characterization may diminish the power to detect (9).

Maximum attention to individual/phenotype characterization by means of standardized, reproducible, and quality-controlled measurements needs to be ensured (11), and patients and controls should undergo the same phenotyping protocol to establish both the presence of disease in patients and the absence of disease in controls (12). Accordingly, studies with smaller sample sizes and precise phenotypic measurement may be as powerful as studies that are even 20 times larger but use less precise measurements (13).

A large number of statistical tests and a large sample size are required in GWASs to detect modestly sized associations. Typically, millions of statistical tests spread across the spectrum of known common SNPs have to be performed; thus, GWASs require extreme levels of statistical significance (14). The primary method used to address the issue of spurious statistical significance arising from multiple comparisons involving up to 1 million SNPs is the Bonferroni correction (15). In such evaluations, the conventional $\alpha$ level of $P (<0.05)$ is divided by the number of tests performed (e.g., $0.05/1000000$, or $5 \times 10^{-8}$).

The imputation of unmeasured genotypes by using information about human LD patterns revealed by the HapMap Project has become an essential tool in conducting GWASs (4). Use of that information makes it possible to supplement directly genotyped SNPs obtained from commercially available arrays with millions of “free” genotypes spanning the human genome at the relatively low costs of computation time and additional statistical tests, leading to an increased power to detect associations (16). Furthermore, imputation tools simplify the combination of association re-

---

### Table 1. Legend: glossary of important terms in cardiovascular genetics.

<table>
<thead>
<tr>
<th>Glossary</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomewide association study (GWAS)</td>
<td>Study investigating statistical association between observable traits or the presence/absence of a disease and a large number of genetic markers. GWASs do not rely on a priori hypotheses.</td>
</tr>
<tr>
<td>Single-nucleotide polymorphism (SNP)</td>
<td>Variation of single nucleotide in a genetic sequence; common form of human variation.</td>
</tr>
<tr>
<td>Minor-allele frequency (MAF)</td>
<td>Proportion of the less common of 2 alleles in a population.</td>
</tr>
<tr>
<td>Linkage disequilibrium (LD)</td>
<td>Nonrandom correlation between 2 alleles located near each other on the same chromosome.</td>
</tr>
<tr>
<td>International HapMap Project</td>
<td>Effort to identify and catalog genetic similarities and differences in humans. Genomewide database of common human variations among multiple ancestral population samples.</td>
</tr>
<tr>
<td>Haplotype</td>
<td>Combination of specific alleles at different loci on the chromosome that tend to be inherited together.</td>
</tr>
<tr>
<td>Expression quantitative trait loci (eQTL)</td>
<td>Genomic loci that show statistically significant correlation to gene expression level. In eQTL studies, genetic variants at specific genomic loci are associated with gene expression levels to pinpoint polymorphic genetic regions affecting gene expression (in cis or trans).</td>
</tr>
<tr>
<td>cis-/trans-acting</td>
<td>Spatial designation of the location of a genetic variant in relation to a locus (gene) affected by this variation, e.g., an eQTL cis-acting indicates a location near the affected locus (e.g. within 1 MB upstream and downstream); trans-acting indicates a location elsewhere in the genome.</td>
</tr>
</tbody>
</table>
sults across studies that use different genotyping platforms and thus have relatively few SNPs in common (16).

Alleles identified in GWASs are often not causative but are likely in LD with the true causative ones (17). Therefore, results obtained in discovery GWASs are only a first step. They require replication in independent study populations and ultimately experimental validation.

Review of GWAS Results for the Most Important Cardiovascular Phenotypes

CORONARY ARTERY DISEASE
CAD is the most common cause of death in industrialized countries, and its prevalence is rapidly increasing in developing countries. CAD has a complex and heterogeneous etiology involving numerous environmental and genetic factors of disease risk (18).

In 2007, three landmark studies used the GWA approach successfully and simultaneously reported the identification of variants on chromosome 9p21.3 associated with risk for MI and CAD (19–21). The SNPs in the region were in strong LD and defined a haplotype associated with a 15%–20% increased risk in heterozygous individuals and a 30%–40% increased risk in homozygous individuals (5). Since then, several studies have confirmed the role of the chromosome 9p21.3 region in the risk of CAD and have postulated, on the basis of a number of significant associations, the identification of variants on chromosome 9p21.3 as associated with risk for MI and CAD (22–24). The first GWASs for blood lipid concentrations were published in 2008 (41–43). Several aspects of GWASs for blood lipids are particularly noteworthy. First, a substantial number of the genes well known to cause rare mendelian hyperlipidemia disorders, such as LDLR3 (low density lipoprotein receptor), APOB [apolipoprotein B (including Ag(x) antigen)], and PCSK9 (proprotein convertase subtilisin/kexin type 9) in familial hypercholesterolemia, have been rediscovered through GWASs. In fact, hyperlipidemia is better than most other phenotypes in that there are substantial number of gene loci for which common genetic variation and rare mendelian mutations overlap. Second, it is intriguing that a large number of GWA loci for hyperlipidemia are also significantly associated with CAD risk, proving the unequivocal link between blood lipid concentration and CAD risk. Finally, several newly dis-

HYPERTENSION
Hypertension is the most widespread cardiovascular disease, with a global burden estimated at 25% to 30% of the population. All genetic loci discovered thus far explain only about 2% of the overall variation in blood pressure. The identification of genetic loci for hypertension through GWASs has proved to be exceptionally difficult, however, with the first GWA analysis (35) having yielded no loci with genomewide significance. This problem may be due to the particular difficulty in defining this phenotype: Single blood pressure measurements in population-based cohorts are often unreliable, and many individuals receive medications, such as B-blockers and angiotensin-converting enzyme inhibitors, that influence blood pressure. Several large-scale GWASs and metaanalyses have recently identified broadly replicated risk loci for hypertension with genomewide significance (12, 35–40). The odds ratios associated with risk loci for hypertension identified in GWASs are very modest and do not exceed 1.05. With some authors arguing that the genetic contribution to developing hypertension might be as large as 50%, it is evident that there is still a huge amount of missing heritability for this important disease.

HYPERLIPIDEMIA
The first GWASs for blood lipid concentrations were published in 2008 (41–43). Several aspects of GWASs for blood lipids are particularly noteworthy. First, a substantial number of the genes well known to cause rare mendelian hyperlipidemia disorders, such as LDLR3 (low density lipoprotein receptor), APOB (apolipoprotein B (including Ag(x) antigen)), and PCSK9 (proprotein convertase subtilisin/kexin type 9) in familial hypercholesterolemia, have been rediscovered through GWASs. In fact, hyperlipidemia is better than most other phenotypes in that there are substantial number of gene loci for which common genetic variation and rare mendelian mutations overlap. Second, it is intriguing that a large number of GWA loci for hyperlipidemia are also significantly associated with CAD risk, proving the unequivocal link between blood lipid concentration and CAD risk. Finally, several newly dis-

3 Genes: LDLR, low density lipoprotein receptor; APOB, apolipoprotein B (including Ag(x) antigen); PCSK9, proprotein convertase subtilisin/kexin type 9; SLC10A1, solute carrier organic anion transporter family, member 1B1; ABCB1, ATP-binding cassette, sub-family B (MDR/TAP), member 1; CYP2C9, cytochrome P450, family 2, subfamily C, polypeptide 19; CYP2C8, cytochrome P450, family 2, subfamily C, polypeptide 9; VKORC1, vitamin K epoxide reductase complex, subunit 1; CYP4F2, cytochrome P450, family 4, subfamily F, polypeptide 2; CDKN2A, cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4); CDKN2B, cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4); CDKN2B-AS, CDKN2B antisense RNA (non-protein coding) previously known as ANRIL (antisense noncoding RNA in the INK4a locus); MTA1, methylthioadenosine phosphorylase; Mus musculus gene Cdkn2a, cyclin-dependent kinase inhibitor 2A; Mus musculus gene Cdkn2b, cyclin-dependent kinase inhibitor 2B; Mus musculus gene Mtap, methylthioadenosine phosphorylase; Mus musculus gene Dmrt1, doublesex and mab-3 related transcription factor like family A1; CELSR2, cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila); PSRC1, proline/serine-rich coiled-coil 1; SORT1, sortilin 1; Mus musculus gene Sort1, sortilin 1; Mus musculus gene Ldb, low density lipoprotein receptor; LIPA, lipase A, lysosomal acid, cholesterol esterase.
covered GWAS loci for blood lipids have already been functionally elucidated in animal studies [e.g., (44–46)], providing good evidence for how GWAS findings can inspire functional experiments. A large GWAS metaanalysis for blood lipid concentrations in >100,000 individuals published in 2010 has identified 95 loci showing genomewide significance for association with blood lipids (46).

**STRUCTURAL HEART DISEASE AND HEART FAILURE**

Echocardiography constitutes the most reliable tool for diagnosing structural changes of the heart. Such structural changes as diastolic dysfunction, cardiac hypertrophy, dilation of the atria, and valvular disease are detectable in their early stages with echocardiography and may precede the phenotype of congestive heart failure by many years. A metaanalysis of GWASs on echocardiographic data for left ventricular function and dimension, atrial size, and aortic root diameter was recently published (47). This study identified many genes that might provide a functional link between intermediate phenotypes such as hypertension or obesity and resulting structural changes of the heart.

Congestive heart failure constitutes a spectrum of heterogeneous conditions, such as dilated cardiomy-
RHYTHM DISORDERS

Disorders of heart rhythm and electrical conductance of the heart have a strong heritable component. Recently, several metaanalyses of GWASs have been published on electrocardiographic components of cardiac or electrical conductance include those for ion channels or proteins that regulate the intracellular concentrations of electrolytes, such as phospholamban for calcium. Of particular clinical interest are several recent GWASs on atrial fibrillation [e.g., (50, 54–57)] and ventricular fibrillation after MI (58). Genetic risk factors associated with the risk of ventricular fibrillation and/or sudden cardiac death after MI might be clinically useful in helping to identify individuals who would benefit from antiarrhythmic therapy or implantation of an automated defibrillator. Prospective trials will be important for determining whether such genetic variants carry clinical implications. For atrial fibrillation, it is worth noting that the 4q25 and 16q22 GWAS loci are also associated significantly with the risk of embolic stroke (59, 60), results that fit the underlying etiology linking atrial fibrillation with embolic stroke.

Pharmacogenomics

The importance of the pharmacogenomics of drugs in the treatment of cardiovascular disease has been addressed in several GWAS analyses. Several prospective observational studies have suggested an association between the SNP encoding the Trp719Arg substitution in kinesin-like protein 6 (rs20455) and the development of clinical CAD. In addition, results from the PROVE IT–TIMI 22 (Pravastatin or Atorvastatin Evaluation and Infection Therapy – Thrombolysis in Myocardial Infarction 22) trial suggest that carriers of this variant will have a strong benefit of risk reduction when treated with statins (61). Rather prematurely, a commercial genetic test for statin response was then made available; however, recent large-scale meta-analyses of the SNP encoding Trp719Arg could not confirm a role for this variant in CAD risk and statin response (62).

Statin-induced rhabdomyolysis is a rare but potentially devastating side effect of statin therapy. A recent GWAS comparing 85 individuals with statin-induced myopathy with control individuals taking the same medication identified a common risk-factor SNP in the SLC01B1 (solute carrier organic anion transporter family, member 1B1) gene, which encodes a transmembrane transporter already known to regulate intracellular statin concentrations (63). Homozygosity for this SNP is associated with a 17-fold increased risk of myopathy under statin treatment. This association has thus far been demonstrated only for simvastatin, but the same mechanism likely applies to other statin drugs as well. Genetic testing for this risk factor could greatly increase the safety of statin treatment, particularly in situations in which myopathy risk is increased, as in cotreatment with amiodarone, cyclosporine, and other drugs.

Currently, clopidogrel is the most extensively used drug for inhibiting platelet aggregation, particularly after coronary artery stenting and/or treatment for acute coronary syndrome. The clopidogrel prodrug is metabolized and bioactivated in the liver by various cytochrome oxidases. The intracellular concentration is also regulated by ABC transporter proteins, such as P-glycoprotein [encoded by ABCB1, ATP-binding cassette, sub-family B (MDR/TAP), member 1]. Several GWASs have linked genetic variation in CYP2C19 (cytochrome P450, family 2, subfamily C, polypeptide 19) and P-glycoprotein with the bioavailability of clopidogrel, the degree of platelet inhibition, and cardiovascular events after stent implantation (64–66). In contrast to clopidogrel, the metabolism of prasugrel, a newly developed platelet inhibitor, is not affected by these genetic polymorphisms (67). In the near future, genetic testing for polymorphisms associated with clopidogrel metabolism may have an impact on both the choice and dose of drug in high-risk individuals.

GWASs have also led to considerable progress in the pharmacogenetics of oral anticoagulants. Genetic polymorphisms in CYP2C9 (cytochrome P450, family 2, subfamily C, polypeptide 9), VKORC1 (vitamin K epoxide reductase complex, subunit 1), and CYP4F2 (cytochrome P450, family 4, subfamily F, polypeptide 2) have been identified as major determinants of warfarin response (68, 69). Applying these genetic polymorphisms in a genetic score makes it possible to predict effective warfarin doses more precisely and safely than with a clinical or fixed-dose algorithm (70).
Variants responsible for disease susceptibility. In the next section, we provide an overview of recent major attempts to unravel the basis for CAD susceptibility.

Gene Desert with Regulatory Function? The 9p21 Locus

When the first large-scale GWASs on CAD identified the chromosome 9p21.3 region in 2007 (19–21), it came as a surprise that this particular genomic region is located in a gene desert with no annotated genes. Subsequently, a large number of genetic and experimental analyses have been performed to elucidate the effects of this locus.

The identified region is adjacent to the INK4/ARF locus, which comprises CDKN2A [cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)], CDKN2B [cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)], and CDKN2B-AS [CDKN2B antisense RNA (non-protein coding); previously known as ANRIL (antisense noncoding RNA in the INK4 locus)], a gene encoding a non–protein-coding RNA. This locus also contains the MTAP (methylthioadenosine phosphorylase) gene (Fig. 2). Several groups have investigated the expression of transcripts by the INK4/ARF locus with respect to SNPs conferring CAD risk (79–82) and demonstrated an effect of the 9p21 locus on the expression of INK4/ARF transcripts, including the ANRIL transcript, but not on MTAP gene expression (79). Interestingly, multiple sites in the 9p21 region seem to independently influence the production of INK4/ARF transcripts (82), but CAD-associated SNPs within this region are all highly associated with CDKN2B-AS (i.e., ANRIL) expression, suggesting a prominent role for ANRIL as the prime susceptibility candidate gene of the 9p21 locus (82, 83). That conclusion has been underlined by results demonstrating ANRIL expression in different cells and tissues relevant for atherosclerosis (80, 81, 83). Additionally, ANRIL transcripts of different lengths have been detected, with production of the short transcript increased and the large transcript decreased in carriers of the 9p21 risk haplotype (80, 81). Jininova et al. hypothesized that by reducing production of the long ANRIL transcript or increasing production of the short ANRIL transcripts, the risk allele may modify CDKN2B expression, thereby promoting proliferative phenotype cell types relevant to atherosclerosis (80). In line with these results, Visel et al. (84) showed in a transgenic mouse model that deletion of the orthologous region on mouse chromosome 4 decreased the expression of the neighboring Cdkn2a (cyclin-dependent kinase inhibitor 2A) and Cdkn2b (cyclin–dependent kinase inhibitor 2B) genes [but not Mtap (methylthioadenosine phosphorylase) and Dmrt1 (doublesex and mab-3 related transcription factor like family A1)] in heart tissue, in-
indicating that the CAD risk interval is required for appropriate expression of Cdkn2a and Cdkn2b in the heart and thus might impact pathophysiological processes. Furthermore, an influence on cell proliferation and senescence, as well as an increased mortality under a diet containing high fat and high cholesterol (but without differences in aortic fatty-lesion formation), was observed in chr4\(^{Δ70kb}/Δ70kb\) mice (84). The region on mouse chromosome 4 shows only 50% homology with the human 9p21 region, however, and no ANRIL ortholog exists in the mouse (85), thus raising additional questions about the discrepancies and commonalities of the different 9p21 studies. A recent study (85) investigated the cellular expression pattern of the CDKN2B, CDKN2A, and MTAP genes in human vascular lesions and found abundant production of the encoded proteins in the vessel walls, thus confirming these genes as interesting candidates potentially implicated in atherogenesis (but independent of the 9p21 atherosclerotic susceptibility).

Recently, Harismendy et al. took a further step toward elucidating the molecular mechanism of the 9p21 locus (86). This group investigated the functions of the underlying variants of the CAD association in detail by using cellular assays and population sequencing. The study predicted the 9p21 interval to be the second densest gene desert for predicted enhancers in the human genome and to contain the most disease-associated variants, indicating an important regulatory function for the 9p21 gene desert. The group identified 33 predicted enhancers within 9p21, and by sequencing the region and computational determination of the variants disrupting consensus transcription factor–binding sites in these predicted enhancers, they found CAD-associated SNPs (rs10811656 and rs10757278) to be located in one of these enhancers. In the presence of the risk haplotype, a binding site for STAT1 (a transcription factor involved in inflammatory responses) was disrupted. This finding was experimentally confirmed by chromatin-immunoprecipitation experiments. Furthermore, this group observed an effect of small interfering RNA–mediated STAT1 knock-down on CDKN2B-AS (ANRIL) and CDKN2B expression, highlighting an important role of STAT1/enhancer interaction in the regulation of the INK4/ARF locus and CAD susceptibility.

Despite this major progress made in the elucidation of the 9p21 susceptibility mechanisms, further in vitro and in vivo studies are required to fully establish the involvement of the INK4/ARF locus and MTAP in the pathologic process and to translate the GWAS signal into a biological meaning.

The 1p13 Sortilin Story

In 2007, another promising locus was identified on chromosome 1p13.3 through a genomewide approach in CAD patients (19). These results have been widely replicated, and this locus has consistently been associated in several studies with LDL cholesterol (LDL-C) concentrations (41–43, 46, 87, 88). Most studies have identified rs599839 as the lead SNP with the strongest correlation with CAD and LDL-C (89). In European populations, the less frequent alleles have been associ-
regulation by the rs599839 genotype genes. In human liver, expression of all 3 genes showed differences in the regulation of the expression of these

1), and

as a lipoprotein receptor that mediates the uptake of cells. These data suggest a potential function of sortilin of radiolabeled LDL and RAP particles into transfected 293 cells and found a significant increase in the uptake of sortilin in transfected human embryonic kidney (44, 94).

enzyme lipoprotein lipase, and apolipoprotein A-V LDL receptor–associated protein (RAP), the lipolytic factor family, acts as a multiligand receptor, and binds the

LDL particles into cells (44). In 2010, Musunuru et al. reported on an elegant series of experimental studies. These investigations identified a SNP at 1p13.3 that created a CCAAT enhancer–binding protein (C/EBP) transcription factor–binding site that had functional consequences. The less frequent allele (that correlates with reduced LDL-C concentrations) introduced the C/EBP binding site and increased the expression of luciferase reporter constructs in human hepatoma cells. In liver-specific experiments in a humanized mouse model, sortilin overproduction decreased total plasma cholesterol and LDL-C, results that were consistent with the genetic findings in human cohorts. Conversely, small interfering RNA–mediated knockdown of Sort1 (sortilin 1) expression in a mouse model led to an increase in total plasma cholesterol. Musunuru et al. concluded that their results identified SORT1 as the causal gene at the 1p13.3 locus for LDL-C and CAD/MI (93).

In contrast to the studies of Linsel-Nitschke et al. and Musunuru et al. that reported a negative correlation between Sort1 expression and LDL-C concentrations, Kjolby et al. reported a positive correlation that was supported by experiments with whole-animal Sort1<sup>−/−</sup>Ldr<sup>−/−</sup> double knock-out mice (92).

Apparent discrepancies exist in the current results for the SORT1 story. It is likely that the use of different in vivo and in vitro models is responsible for these discrepant outcomes; thus, further studies are needed to fully reconcile and explain the SORT1 story. Nevertheless, current data have uncovered regulatory pathway(s) in hepatic lipoprotein export by logical, hypothesis-based experiments and have provided the first functional explanations for cardiovascular risk being

<table>
<thead>
<tr>
<th>Table 2. Legend: summary of different theories/models of mode of action of sequence variations on the chromosome 1p13 CAD risk locus based on experimental data from 3 different studies.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models of sortilin function and relationship to 1p13 variants</td>
</tr>
<tr>
<td>Model 1 [Linsel-Nitschke et al. (44), Dubé et al. (94)]</td>
</tr>
<tr>
<td>Identification of colocalization of sortilin with apolipoprotein B-100, which facilitates maturation of pre-VLDL and hepatic secretion of VLDL. Increased SORT1 expression stimulates hepatic release of VLDL and increases plasma LDL-C.</td>
</tr>
<tr>
<td>Increased circulating LDL-C concentrations are contrary to the clinical biochemical consequence (i.e., decreased LDL-C concentrations) expected on the basis of earlier GWAS and eQTL studies.</td>
</tr>
<tr>
<td>Identification of SNP (rs12740374) within binding site for C/EBP. Disruption of C/EBP binding by the risk allele (G) of rs12740374, leading to reduced transcription of SORT1 and increased hepatic LDL precursor VLDL secretion.</td>
</tr>
<tr>
<td>Lowering of circulating LDL-C concentrations consistent with the expected clinical biochemical consequence (i.e., decreased LDL-C concentrations) of increased SORT1 expression.</td>
</tr>
<tr>
<td>G allele of SNP (rs599839) is associated with increased SORT1 expression and increased LDL-C uptake into liver (lowering blood LDL-C concentrations), presumably through increased interactions with the LDL receptor and increased receptor-mediated lowering of circulating LDL-C concentrations consistent with expected clinical biochemical consequences (i.e., decreased LDL-C concentrations) of increased SORT1 expression.</td>
</tr>
<tr>
<td>Identification of association signals and the locus for LDL-C and CAD/MI (93). These investigations identified a SNP at 1p13.3 that tightly linked genes, namely CELSR2 [cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)], PSRC1 (proline/serine-rich coiled-coil 1), and SORT1 (sortilin 1). Different cell types showed differences in the regulation of the expression of these genes. In human liver, expression of all 3 genes showed regulation by the rs599839 genotype (42, 90). In human blood cells, Linsel-Nitschke et al. observed a significant association between rs599839 and levels of SORT1 expression but detected no association with CELSR2 or PSRC1 (44). However, using the largest population-based human monocyte biobank available, Zeller et al. could not identify the cis association between SORT1 mRNA levels and the 1p13.3 locus, whereas they did prove a strong association between PSRC1 mRNA levels and the less frequent rs629301 allele (a tag of rs599839) (91). Despite these discrepancies, recent investigations have focused on the SORT1 gene to provide functional characterization of the association signals (44, 92, 93).</td>
</tr>
<tr>
<td>Sortilin 1 is a member of a VPS10P-domain receptor family, acts as a multiligand receptor, and binds the LDL receptor–associated protein (RAP), the lipolytic enzyme lipoprotein lipase, and apolipoprotein A-V (44, 94). Linsel-Nitschke et al. reported overproduction of sortilin in transfected human embryonic kidney 293 cells and found a significant increase in the uptake of radiolabeled LDL and RAP particles into transfected cells. These data suggest a potential function of sortilin as a lipoprotein receptor that mediates the uptake of LDL particles into cells (44). In 2010, Musunuru et al. reported on an elegant series of experimental studies. These investigations identified a SNP at 1p13.3 that created a CCAAT enhancer–binding protein (C/EBP) transcription factor–binding site that had functional consequences. The less frequent allele (that correlates with reduced LDL-C concentrations) introduced the C/EBP binding site and increased the expression of luciferase reporter constructs in human hepatoma cells. In liver-specific experiments in a humanized mouse model, sortilin overproduction decreased total plasma cholesterol and LDL-C, results that were consistent with the genetic findings in human cohorts. Conversely, small interfering RNA–mediated knockdown of Sort1 (sortilin 1) expression in a mouse model led to an increase in total plasma cholesterol. Musunuru et al. concluded that their results identified SORT1 as the causal gene at the 1p13.3 locus for LDL-C and CAD/MI (93).</td>
</tr>
</tbody>
</table>

In contrast to the studies of Linsel-Nitschke et al. and Musunuru et al. that reported a negative correlation between Sort1 expression and LDL-C concentrations, Kjolby et al. reported a positive correlation that was supported by experiments with whole-animal Sort1<sup>−/−</sup>Ldr<sup>−/−</sup> double knock-out mice (92).
associated with 1p13.3. Table 2 summarizes the models and recent results of functional analyses of the 1p13/SORT1 locus.

**Linking GWAS Results with the Transcriptome**

Variation in gene expression is important in mediating disease susceptibility. The identification of genetic variants that potentially affect regulatory elements and thereby regulate gene transcription is sought as an aid in mapping human disease genes (95). eQTL (expression quantitative trait loci) mapping, a widely used tool for identifying genetic variants that affect gene regulation, provides a means for detecting transcriptional regulatory relationships on a genomewide scale (87, 91, 92, 95–98).

In a study conducted to assess the overall variability of the monocytic transcriptome for nearly 1500 healthy individuals, Zeller et al. (91) detected 2745 eQTLs (of 12 808 expressed genes), most (90%) of which were cis modulated. The authors further investigated whether monocyte gene expression might mediate the effects of loci recently identified by GWASs of cardiovascular risk factors. Loci identified in previous GWASs of lipids, blood pressure, and body mass index were chosen, and associations between gene expression and the lead or tag SNPs, as well as the respective risk factor, were analyzed. For the phenotypic traits, several SNPs showed strong correlation to gene expression, mainly in cis; however, when the association of the expression trait with the risk factor under consideration was assessed, only a few significant findings were observed. One was the above-mentioned association between the 1p13.3 locus and LDL-C (91).

Within the frame of the “CADomics” study, the same group performed a GWAS via a case–control approach with more than 20 000 CAD cases and 38 000 controls. They subsequently analyzed the effect on gene expression for each CAD-associated SNP (33). This approach revealed a novel CAD-susceptibility locus on chromosome 10q23.31, with the lead SNPs, rs1412444 and rs2246833, located in intronic regions of the LIPA (lipase A, lysosomal acid, cholesterol esterase) gene, which encodes lysosomal acid lipase A. Using the above-mentioned data set of global monocytic gene expression, the authors assessed the influence of both SNPs on gene expression. A strong cis effect on LIPA transcript levels was observed, with the risk alleles being associated with increased LIPA expression. This find provides the first evidence for the functionality of this locus. Furthermore, LIPA transcript levels were significantly associated with prevalent cardiovascular risk factors and phenotypes of subclinical disease. In line with the results from the CADomics study, the Coronary Artery Disease (C4D) Genetics Consortium confirmed LIPA as a novel CAD-associated locus in a metaanalysis of European and South Asian populations and confirmed the association with increased LIPA expression in a subsequent eQTL analysis (99).

These examples demonstrate the benefit of using eQTL analyses to further evaluate the potential functional relevance of genetic variants identified in large-scale association studies and present a first step in the translation of genetic-association signals into biological function. Combining eQTL and DNA sequence–variation data sets in an analysis of coexpressed gene networks, e.g., as shown by Heining et al. (100), will provide even more relevant information for unraveling the genetic underpinnings of cardiovascular disease.

**Beyond GWAS: What Comes Next?**

The broad availability of GWASs has led to an enormous boost in the discovery of risk genes for cardiovascular disease. Genes identified by GWASs have inspired biology and enabled the discovery of novel pathways or mechanisms, as exemplified by the findings for chromosome 9p21/ANRIL, chromosome 1p13/SORT1, and—just recently—the chromosome 10q23 LIPA locus. However, only a very small fraction of the heritability for the most important cardiovascular diseases, such as CAD or hypertension, can be explained by the common genetic risk factors identified by GWASs thus far. A huge gap of missing heritability remains. The field of individual risk prediction for cardiovascular disease using common genetic factors is still in its infancy. Most genetic risk factors identified thus far will show only small odds ratios, rendering the positive predictive values of such tests minimal.

For the most important cardiovascular diseases, such as CAD, hypertension, and hyperlipidemia, large-scale metaanalyses have already been conducted. Because the density of coverage and the number of participants involved in these studies have been optimized, it seems unlikely at present that substantially more discoveries will arise from GWASs.

To prioritize GWAS findings and to translate them into biologically meaningful functions require post-GWAS analyses including bioinformatics-based approaches, epigenetic approaches, and follow-up experimental studies (101, 102). With the rapid advances in next-generation technologies, whole-exome and -genome sequencing is feasible, and data from the 1000 Genomes Project (103) will provide a comprehensive map of novel, rare variants.

For the identification of variants contributing to disease susceptibility and possibly mediating gene expression caused by environmental influences, assessments of epigenetic mechanisms (i.e., histone modification, methyl-
ation, and microRNAs) are promising approaches (102). For example, a recent study described by Richardson et al. (104) used bioinformatics to identify variants that disturb or create microRNA-recognition elements in a genome-wide fashion, thus providing functional hypotheses for observed GWAS associations.

The task remaining for the future is to establish causality and functionality and ultimately to move from gene discovery to translation (101). This road will require in-depth post-GWAS analyses across multiple disciplines involving genetics, molecular biology, and bioinformatics.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design of data, or analysis and interpretation of acquisition; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

References