Genetics of coronary artery disease

Prins, Bram P.; Lagou, Vasiliki; Asselbergs, Folkert W.; Snieder, Harmen; Fu, Jingyuan

Published in:
ATHEROSCLEROSIS

DOI:
10.1016/j.atherosclerosis.2012.05.015

Publication date:
2012

Atherosclerosis 225 (2012) 1–10

Contents lists available at SciVerse ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis

Review

Genetics of coronary artery disease: Genome-wide association studies and beyond

Bram P. Prins [a,1], Vasiliki Lagou [a,b,1], Folkert W. Asselbergs [c,d,e], Harold Snieder [a,*,1], Jingyuan Fu [a,f,**,1]

[a] Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[b] Oxford Centre for Diabetes, Endocrinology and Metabolism & Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[c] Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands
[d] Department of Medical Genetics, University Medical Center, Utrecht, The Netherlands
[e] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
[f] Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 8 November 2011
Received in revised form
15 May 2012
Accepted 16 May 2012
Available online 23 May 2012

Keywords:
Heritability
Coronary artery disease
Genome-wide association study
Systems genetics

ABSTRACT

Genome-wide association (GWA) studies on coronary artery disease (CAD) have been very successful, identifying a total of 32 susceptibility loci so far. Although these loci have provided valuable insights into the etiology of CAD, their cumulative effect explains surprisingly little of the total CAD heritability. In this review, we first highlight and describe the type of genetic variants potentially underlying the missing heritability of CAD: single nucleotide polymorphisms (SNPs) or structural variants, each of which may either be common or rare. Although finding missing heritability is important, we further argue in this review that it constitutes only a first step towards a fuller understanding of the etiology of CAD development. To close the gap between the genotype and phenotype, we propose a systems genetics approach in the post-GWA study era. This approach that integrates genetic, epigenetic, transcriptomic, proteomic, metabolic and intermediate outcome variables has potential to significantly aid the understanding of CAD etiology.

Contents

1. Interpretation and limitations of genome-wide association findings for CAD 1
2. Factors underlying the missing heritability 4
   2.1. Common SNPs with (very) small effect 4
   2.2. Rare SNPs with large effect 4
   2.3. Structural variants 5
3. Beyond the (narrow sense) heritability 5
4. Integration: role of systems genetics 6
5. Challenges and prospects of systems genetics for CAD 7
   Acknowledgements 8
   References 8

1. Interpretation and limitations of genome-wide association findings for CAD

Coronary artery disease (CAD) is the leading cause of death in Western societies. For example, in the United States the total prevalence of CAD is 7.0% in adults over 20 years of age and it caused about 1 of every 6 deaths in 2007 [1]. It can be viewed to result from a combination of genetic and environmental factors as

* Corresponding author.
** Corresponding author. Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.
E-mail addresses: h.snieder@umcg.nl (H. Snieder), j.fu@umcg.nl (J. Fu).
[1] These authors contributed equally.

0021-9150/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.atherosclerosis.2012.05.015

well as their interactions. Epidemiological studies have identified many traditional risk factors for CAD, including tobacco use, physical inactivity, poor nutrition, obesity, hypertension, high blood cholesterol, and diabetes. In addition to these modifiable risk factors, CAD and its main complication, myocardial infarction (MI), have a strong genetic basis [2]. For example, a family history for CAD was associated with CAD independent of other cardiovascular risk factors [3]. Based on a 36-year follow-up study of more than 20,000 Swedish twins the *heritability* (h^2) of fatal coronary events was estimated at 0.57 for males and 0.38 for females [4].

The CAD gene database (CADgene) [5] includes information on more than 300 candidate genes, but many of the genetic mechanisms that predispose people to CAD remained unknown until the development of a highly dense genotyping array to analyse common variants genome-wide. This provided new opportunities to identify genetic risk factors associated with CAD. The genome-wide association (GWA) study on CAD was pioneered in 2002 by a Japanese group using a genotyping array of >90,000 single nucleotide polymorphisms (SNPs) in 94 MI cases and 658 controls [6,7]. These early studies pointed to two susceptibility loci (*LTA* and *LGALS2*) that are involved in inflammation and induction of cell-adhesion molecules, but later studies failed to replicate their association [8]. However, a third gene in the same pathway (*BRAP*) yielded one of the strongest associations with a single SNP in a CAD GWA study (OR = 1.42; Table 1) [9,10]. The first wave of seven high-throughput array-based association studies on CAD [9,11–16] identified a total of 12 risk loci

Table 1
Summary of 34 independent risk variants at 32 CAD susceptibility loci identified by GWA studies.

Locus No	Genome region	Risk SNP	Risk allele	AF	Odds ratio	Genes	Association with (traditional) risk factors[a]	Ref
1	1p13.3	rs646776	T	0.81	1.19	CELSR2, PSRC1, SORT1	LDL, response to statin, progranulin level, total cholesterol, Lp-PLA2 activity and mass	[25]
2	1p32.2	rs17114036	A	0.91	1.17	PPAP2B		[20]
3	1p32.3	rs11206510	T	0.81	1.15	PCSK9	LDL	[12,20]
4	1q41	rs17465637	C	0.72	1.14	MIA3		[12,14,20]
5	2q33.1	rs6725887	C	0.14	1.17	WDR12		[12,20]
6	3q22.3	rs9818870	T	0.15	1.15	MRAS		[13]
7	6p21.31	rs17609940	G	0.75	1.07	ANKS1A		[14]
8	6p21.33	rs3869109	G	0.56	1.14	HCG27, HLA-C	Triglycerides	[24]
9	6p24.1	rs12526453	C	0.65	1.12	PHACTR1		[12,20,25]
		rs6903956[b]	A	0.03	1.51	c6orf105		[21]
10	6q23.2	rs12190287	C	0.62	1.08	TCF21		[20]
11	6q25.3	rs3798220	C	0.02	1.92	SLC22A3, LPAL2, LPA	Lp(a) level	[20]
		rs10455872[g]	G	0.07	1.70			[19]
12	7q21	rs1859023[c]	A	0.31	0.72[d]	PFTK1		[26]
13	7q22.3	rs10953541	C	0.74	1.08	BCAP29		[25]
14	7q32.2	rs11556924	C	0.62	1.09	ZC3HC1		[20]
15	9p21.3	rs4977574	G	0.56	1.29	CDKN2A, CDKN2[e]	Abdominal aortic aneurysm, intracranial aneurysm	[10,12,14-16, 20,25,27]
16	9p21.3	rs7865618	A	0.59	1.18	MTAP[e]	Type 2 diabetes	[27]
17	9q34.2	rs579459	C	0.21	1.10	ABO	Serum phytosterol level, plasma levels of liver enzymes, venous thromboembolism, E-selectin levels, adhesion levels	[20]
18	10p11.23	rs2505083	C	0.43	1.08	KIAA1462	Non-alcoholic fatty liver disease histology	[20,25]
19	10q11.21	rs1746048	C	0.84	1.17	CXCL12		[12,20]
20	10q23.2	rs1412444	T	0.37	1.1	LIPA	Systolic blood pressure	[25,27]
21	10q24.32	rs12413409	G	0.89	1.12	CYP17A1, CNNM2, NT5C2	Systolic blood pressure, intracranial aneurysm	[20,25]
22	11q22.3	rs974819	T	0.22	1.07	PDGFD		[25]
23	11q23.3	rs964184	G	0.13	1.13	ZNF259, APOA5-A4-C3-A1	HDL, hypertriglyceridemia, triglycerides	[20]
24	12q24.12	rs11066001[b]	C	0.34	1.42	BRAP fs		[10]
	12q24.12	rs671[b]	A	0.23	1.43	ALDH2 fs		[10]
25	13q34	rs4773144	G	0.44	1.07	COL4A1, COL4A2		[20]
26	14q32.2	rs2895811	C	0.43	1.07	HHIPL1		[20]
27	15q25.1	rs3825807	A	0.57	1.08	ADAMTS7,MORF4L1[f]		[20,23]
	15q25.1	rs4380028	C	0.65	1.07	ADAMTS7[f]		[25]
28	17p11.2	rs12936587	G	0.56	1.07	RASD1, SMCR3, PEMT		[20]
28	17p13.3	rs216172	C	0.37	1.07	SMG6, SRR	Aortic root size, type 2 diabetes	[20]
30	17q21.32	rs46522	T	0.53	1.06	UBE2Z, GIP, ATP5G1, SNF8		[20]
31	19p13.2	rs1122608	G	0.75	1.15	LDLR		[12,20]
32	21q22.11	rs9982601	T	0.13	1.20	SLC5A3, MRPS6, KCNE2		[12,20]

[a] The associations were extracted from the GWA Catalog database (www.genome.gov/gwastudies/). The traits are listed here if their associated SNPs are in linkage disequilibrium with CAD SNPs ($r^2 > 0.5$, based on the HapMap II and III CEU panel).
[b] Association detected only in Chinese Han or Japanese populations.
[c] Assocation detected only in African American populations.
[d] Hazard Ratio.
[e] These loci are reported as independent; LD (r^2) < 0.3 in the HapMap II and III CEU panel.
[f] These loci are not reported as independent but map to different genes; $r^2 > 0.7$ for rs11066001 and rs671 in 1000 Genomes Pilot 1 data for CHB + JPT; $r^2 > 0.5$ for rs3825807 and rs4380028 in the HapMap II CEU panel.
[g] This variant has been found through a study employing a gene-centric chip designed to assay SNPs in genes implicated in cardiovascular, metabolic and inflammatory disease [28].

with mostly modest effect sizes of odds ratios (ORs) in the 1.1–1.2 range and collectively explaining only a small part of the estimated CAD heritability. For example, the cumulative effect of nine loci identified in the Myocardial Infarction Genetics (MIGen) Consortium explained just 2.8% of the variance in risk for early-onset MI [12]. This suggests that these GWA studies were probably underpowered due to modest sample sizes. Therefore, the majority of CAD heritability remained missing, which limited the clinical translation of these genetic findings [17,18]. Realizing this, some of the five most recent GWA studies had very large sample sizes [19–24] (e.g., over 22,000 cases and 64,000 controls in the CARDIoGRAM discovery set) and were conducted on different ethnic groups (Europeans, South Asians, Han Chinese and African Americans). These studies have now brought the number of independent risk variants up to a total of 34 at 32 genomic loci (Table 1).

Most of these loci have small effect sizes with ORs in the 1.05–1.20 range. Interestingly, 12 out of the 32 loci are also associated with traditional CAD risk factors and related traits, including blood pressure, low-density lipoprotein (LDL) cholesterol and plasma lipoprotein(a) [Lp(a)] level. This provides genetic evidence for the causal effect of these traditional risk factors on CAD risk.

Translating GWA signals to biological function is seldom straightforward. Therefore, we conducted a *pathway analysis* of all the 31 CAD associated loci discovered up until the end of 2011 in order to provide some initial functional insight. A functional connection network and annotation analysis on 86 potential candidate genes underlying the 31 associated loci highlighted several dominant processes. Some of these were expected such as glycerolipid and steroid metabolism, cell proliferation and the circulatory system (Fig. 1). However, others may not be immediately obvious and may suggest new hypotheses regarding underlying mechanisms for CAD.

Although the recent large-scale GWA studies on CAD have more than doubled the number of risk loci offering more insight into the disease etiology, they did not confirm associations for the majority of the candidate genes from the CADgene database. Moreover, the combined effect of the associated loci still only explains

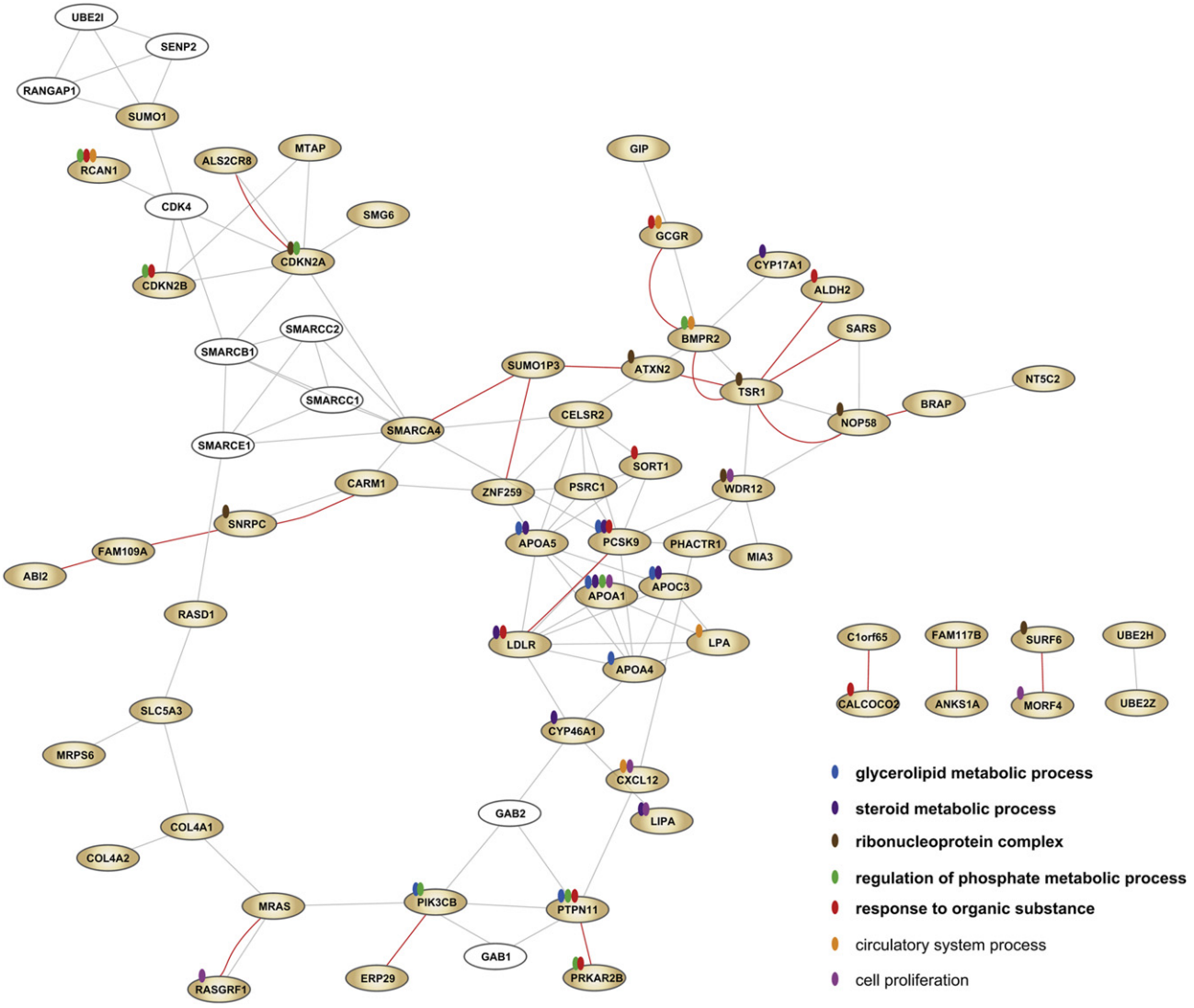

Fig. 1. Pathways underlying CAD associated loci. Each colour coded node represents a gene. The genes in brown are candidate genes at CAD associated loci whereas those in white are genes that show functional connections with the CAD genes. The links between genes indicate their functional connection: those in grey are the combined functional connections with median confidence predicted by STRING, those in red are the direct protein–protein interactions predicted by DAPPLE. Selections of enriched biological processes as annotated by Gene ontology are highlighted for each gene. Only the processes in bold remain significant with FDR<0.05 after taking multiple testing into account. Details of the analysis can be found in the supplementary methods.

approximately 10% of the additive genetic variance of CAD, leaving the majority of its heritability unexplained [20].

In this review, we first highlight and describe the importance of the potential sources of the *missing heritability* in CAD/MI. Then we will argue that pinpointing the specific factors underlying the heritability is essential but far from sufficient to fully understand disease etiology for several reasons. First, it will remain a challenge to translate newly identified genetic signals to biological function. Second, numerous important contributors to disease risk are not covered by the heritability estimate, including gene–gene interaction, *epigenetic* variation and gene–environment interaction. We then propose that a systems biology approach that integrates environmental, genetic and epigenetic factors as well as transcriptomic, proteomic, metabolic and intermediate outcome variables would be the best way forward and would aid significantly to the understanding of CAD etiology.

2. Factors underlying the missing heritability

If we limit ourselves to the *narrow-sense heritability*, which only reflects additive genetic effects (i.e., no dominant or *epistatic effects*) [29], suggested explanations of missing heritability include additional common SNPs with (very) small effect, rare SNPs with larger effects, and structural variants. Optimal designs, technologies and statistical approaches to detect sources of unexplained heritability for common complex traits and diseases have recently been reviewed in considerable depth [30–33] and leading geneticists have offered their opinions on the subject [34]. Therefore, we discuss these issues only briefly here and focus on the sources of missing heritability and their importance for CAD.

2.1. Common SNPs with (very) small effect

Genotyping platforms currently used in GWA studies are designed to tag most known common SNPs (minor allele frequency [MAF] $>$ 0.05), thereby testing the "common disease – common variant" hypothesis. With only a few exceptions, the identified risk alleles of CAD have small to modest effects with ORs between 1.05 and 1.2 and frequencies ranging from 0.13 to 0.91 (Table 1). The as of yet unidentified common risk alleles may have (very) small effects limiting the possibility to detect them individually with the current GWA study sample sizes. This raises the question how many more of these common SNPs with small effect sizes we would need to discover to explain the entire CAD heritability. The total number of such underlying risk variants can be estimated as a function of disease heritability, disease prevalence and some simplifying assumptions that all risk alleles have the same relative risk and frequency [35]. Fig. 2 shows the numbers of risk variants for a range of effect sizes (ORs between 1.05 and 1.20) and different allele frequencies for males and females separately. For example, at a disease prevalence of 7%, 1020 and 1218 risk alleles, each with a relative risk of 1.1 and frequency of 0.1, are needed to explain the heritabilities of 0.38 and 0.57 in females and males, respectively.

However, the number of risk alleles increases at an exponential rate with decreasing relative risk. If the relative risk decreases to 1.05, the total number of risk alleles increases to 4040 for $h^2 = 0.38$ and 4823 for $h^2 = 0.57$. More sophisticated estimates can be obtained by taking into account the full spectrum of expected risk effects and allele frequencies [36]. The detection of risk alleles with small effect requires exponential increases in sample sizes, because required sample sizes scale approximately quadratically with $1/|(OR-1)|$ [32]. Realizing this need, international consortia have emerged, such as the Myocardial Infarction Genetics Consortium (MIGen) [12], CARDIoGRAM [20,37] and the Coronary Artery Disease (C4D) genetics consortium [25]. These consortia have performed meta-analyses combining the association signals from multiple GWA studies, thus maximizing the power to discover risk alleles for CAD.

2.2. Rare SNPs with large effect

SNPs with MAF less than 5% in the general population are considered to be of low frequency (i.e., rare). The occurrence of rare

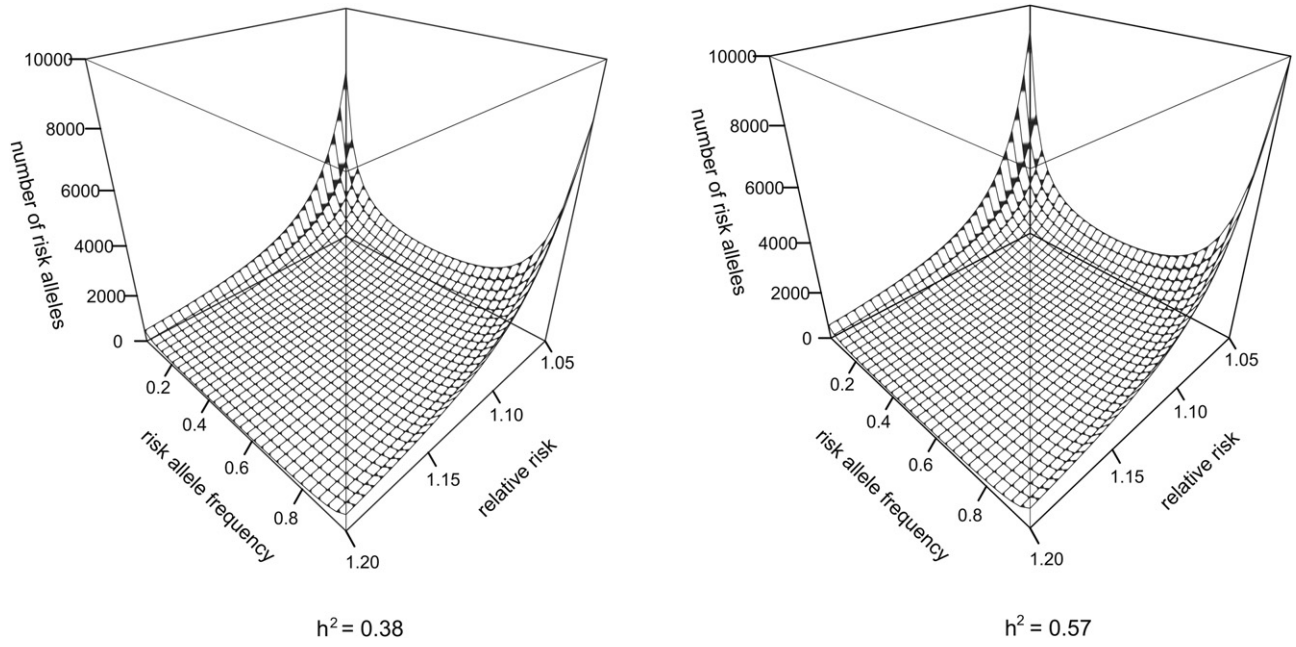

Fig. 2. Estimate of the number of common variants that contribute risk to CAD. The number of risk variants underlying heritability can be modelled as $n = ln[h^2 + (1 - h^2)K] - ln(K)/2\{ln[1 + p(\lambda^2 - 1)] - ln[1 + p(\lambda - 1)]^2\}$, where n refers to the number of risk variants; h^2 is the heritability of the disease; K is the disease prevalence in the population; λ is the relative risk of a risk allele and p is the frequency of the risk allele, assuming all risk alleles have the same relative risk and frequency [35]. At the disease prevalence K of 0.07, the plots show the number of common variants needed to explain heritabilities of 0.38 for females and 0.57 for males based on a 36-year follow-up study of >20,000 Swedish twins [4] depending on the relative risk and allele frequency.

variants in the population can be due to selection pressure, random genetic drift or introduction of recent mutations [46,47]. These alleles, although individually rare, are collectively frequent and might contribute substantially to genetic susceptibility underlying complex traits and diseases [31,48]. However, GWA arrays predominantly include common SNPs and in general these do not tag the rare variants well. Therefore our knowledge of the impact of rare variants on CAD remains limited. To address this issue a number of novel experimental strategies and statistical models for the detection of rare variants and their association with complex traits and diseases have emerged [30,49]: 1) as rare variants are often of recent origin they can typically be tagged by the *haplotype* on which they arose, because recombination has had insufficient time to break down the *linkage disequilibrium* (LD) surrounding the variant [30,49]; 2) recent advances in high-throughput sequencing technology have accelerated the discovery of rare risk alleles; 3) custom-made arrays have specifically included rare variants in target regions thereby allowing genotyping of such variants in large sample sizes. These advances have uncovered the association between CAD and several rare variants. For example, the association of rs3798220 (MAF: 0.02) at the *SLC2A-LPAL2-LPA* locus was first detected by haplotype association [11] and subsequently identified by a custom-made array [19] (Box 2). Application of haplotype association analysis to the Wellcome Trust Case Control Consortium (WTCCC) GWA data identified rare variants at one known locus (*CDKN2B*) and three novel loci for CAD: *EIF4H*, *HFE2*, and *ZBTB43* [50]. These rare variants often have larger effects than common variants. For example, the OR of rs3798220 is 1.92, much larger than the effects of common variants with an OR between 1.05 and 1.2.

2.3. Structural variants

Besides variation at a single nucleotide position, a segment of DNA can be deleted, duplicated or rearranged. This type of DNA variation is known as structural variation. One common type of structural variation is the copy number variant (CNV) that refers to DNA deletion or duplication >1000 base pairs in size [51], which might contribute substantially to risk for common diseases as shown recently for obesity [52]. Previous studies have identified the association of CAD risk with low kringle IV type 2 copy number at the *SLC2A-LPAL2-LPA* locus [44,53] and high number of CA repeats at the *NOS3* locus [54]. Another well-known example in relation to CAD involves Heterozygous Familial Hypercholesterolemia (HeFH), which is an autosomal dominant disorder that affects 1 in 500 people. The genetics underlying the disease in the majority of HeFH patients include SNPs as well as CNVs or small deletions within the LDL receptor gene *LDLR*, making it impossible for the liver to catabolize LDL cholesterol. The resulting rise in plasma LDL cholesterol leads to atherosclerosis and up to a 100 times greater risk of CAD [55].

So far, GWA studies have been unsuccessful in detecting effects of CNVs on CAD, perhaps because they only capture the common CNVs. Despite good coverage of CNVs no significant associations were detected in the MIGen [12] and WTCCC studies [56]. They concluded that common CNVs that can be typed on existing platforms are unlikely to contribute greatly to the genetic basis of common human diseases.

3. Beyond the (narrow sense) heritability

Above we discussed the potential genetic sources of missing (narrow-sense) heritability of CAD. Although important, just finding the missing heritability is only a first step towards a fuller understanding of the disease etiology because the narrow-sense heritability does not capture at least three factors that are believed to be of vital importance for disease development: gene–gene interactions, gene–environment interactions and epigenetic effects.

Box 1. Glossary of terms.

Epigenetic effects – Heritable changes in gene expression that are not caused by changes in DNA sequence, such as DNA methylation (addition of methyl groups to a DNA base) and histone modification (histones are proteins that enable dense packing of DNA in cell nuclei).

Epistasis – Interaction between genes that may result in a phenotype different from the expected phenotype in the case that these genes would not interact.

Heritability – The proportion of individual differences (i.e. variation) in a certain trait (or phenotype) that can be attributed to genetic variation. If the genetic variation includes the total genetic variation, consisting of additive genetic effects, dominance genetic effects (representing interactions between alleles at the same locus), and epistatic genetic effects (representing interactions between alleles at different loci), this is called the *broad-sense heritability*. If this genetic variation is limited to the additive genetic variation only, this is called the *narrow-sense heritability*.

Missing heritability – For all of the diseases and traits that have been studied by means of GWA studies, the identified variants explain only a small proportion of the total heritability. The proportion of heritability that remains unaccounted for is generally referred to as the missing heritability.

Haplotype – A set of alleles or variants that is inherited together as a unit.

Linkage disequilibrium (LD) – The extent to which two alleles are non-randomly associated, which is determined by the degree of recombination.

Omics– A suffix that is added to a wide variety of analyses to indicate they occur on a large or genome-wide scale. Transcriptomics for example refers to analysis of genome-wide expression level of messenger RNAs – the transcriptome.

Pathway-based analysis – An approach in which genome-wide results for a trait or disease are analysed and interpreted in the context of predefined pathways, which are collections of genes or proteins with know interaction, instead of investigating the individual effects of genes. This type of analysis is frequently applied as part of *post GWAS analyses*, to identify potentially important molecular mechanisms underlying the disease or trait of interest.

Post GWAS analyses – A recently coined term that refers to a collection of methods and approaches that aim to reveal the functional consequences of loci identified in GWA studies.

Quantitative trait loci (QTLs) – These are specific regions in the genome that influence a quantitative trait. Examples of quantitative traits include RNA levels (genome-wide referred to as the transcriptome), and levels of metabolites (genome-wide referred to as the metabolome).

Systems genetics – A recently coined term that refers to an integration of genetic analysis approaches aiming to understand the complexity of genotype and phenotype relationships in complex traits and diseases, in analogue fashion to systems biology.

Box 2. Genetic architecture and function of the *LPA* locus.

Lipoprotein(a) [Lp(a)] levels have long been known to be a risk factor for CAD [38] with very high heritability (~ 90%) [39,40]. It has also been known that this heritability could almost entirely be explained by variation at the apolipoprotein(a) gene on 6q25 as shown in linkage studies [41–43]. The identification of this poster child locus (*SLC2A-LPAL2-LPA*) as the strongest for CAD to date and elucidation of (part of) its genetic architecture is particularly intriguing. CAD was initially observed to be associated with two haplotypes of four SNPs [11] (rs2048327, rs3127599, rs7767084, and rs10755578) that turned out to tag two rare variants rs3798220 and rs10455872 (see Table 1), and a CNV of kringle IV type 2 (KIV-2) repeats. Furthermore, this locus was observed to be associated with expression of the *LPA* gene in the liver (rs9355814, $P = 2.24\text{x}10^{-28}$) [19]. Interestingly, the CAD associated SNPs rs3798220 and rs10455872 were also highly associated with Lp(a) levels in the serum, together explaining 36% of the total Lp(a) variation. After the adjustment for Lp(a) level, their associations with CAD were abolished, which indicates that Lp(a) level is indeed a causal intermediary factor [19]. Further research showed that the CAD risk variants rs3798220 and rs10455872 together with the KIV-2 repeat explained a larger proportion of variation in Lp(a) concentration than the SNPs by themselves [44]. This suggests that both SNPs and CNVs contribute to CAD risk through their influence on Lp(a) concentrations. A recent GWA study by Kivimäki et al. [45] detected a common SNP (rs783147 with a MAF of 0.47) with a very strong effect on Lp(a) ($P = 3.1 \times 10^{-58}$) that explained 11.7% of its variance. In conclusion, these results show a fairly complicated genetic architecture of the *LPA* locus with multiple independent variants contributing to Lp(a) levels including rare SNPs, common SNPs and a CNV.

First, genes do not function in isolation. There is increasing awareness that gene–gene interaction or epistasis plays a role in susceptibility to complex diseases. We observe that disease-associated genes identified by GWA studies often converge on pathways, co-expression networks and protein–protein interaction networks [49] as illustrated by the functional connection network of CAD loci shown in Fig. 1. Some gene products even show direct interaction as observed between PCSK9 and LDLR. The proteinase PCSK9 can bind to the LDL receptor and mediate the degradation of LDLR [57]. However, detecting epistatic effects statistically remains challenging. GWA studies have typically used single-locus strategies and a risk variant may thus be missed if its marginal effect is not strong enough to pass the genome-wide significance level. Cordell and others have provided a critical survey of the methods and software to detect interactions in the context of GWA studies and showed that epistasis analysis is statistically feasible [58]. In the near future, pathway-based association analyses are expected to provide a new paradigm for the second-wave of GWA studies [59].

Second, the expression of some genes may be dependent on environmental factors. Sabatti and co-workers performed a GWA analysis of gene–environment interaction for nine metabolic traits in the Northern Finland Birth Cohort [60], including some traditional CAD risk factors such as triglycerides, HDL, LDL, body mass index and blood pressure. Although the gene–environment interactions detected in this study need further replication, it shows that prospectively investigating such interactions for CAD risk may be fruitful. Lanktree and Hegele specifically discussed gene–gene and gene–environment interactions in CAD development and concluded that accounting for gene–gene and gene–environment interactions is important for future strategies of diagnosis, prognosis and management of CAD [44].

Third, one possible mechanism through which environmental factors contribute risk to complex disease such as CAD is through mediation of the epigenome [61]. Epigenetic effects refer to all meiotically and mitotically heritable changes in gene expression that are *not* coded in the DNA sequence such as those caused by methylation and histone modification. Epigenetic mechanisms collectively enable the cells to respond quickly to environmental changes. Several studies have argued that epigenetic variation is a driving force of development, evolutionary adaption, and complex diseases [62,63]. Recent studies have shown differential global DNA methylation levels in peripheral blood leukocytes in CAD patients compared to controls, but the direction of the effect is inconsistent [61,64–66] probably due to the limited resolution of the global methylation measures.

4. Integration: role of systems genetics

Whether part of the heritability or not, integration of abovementioned disparate determinants of disease etiology in a common framework is badly needed. We therefore argue that in the *post-GWA* era, a *systems genetics* approach may help us move from finding heritability towards understanding the complex biological networks that underlie complex diseases such as CAD. Systems genetics, by definition, is the approach that studies genetic effects within the larger scope of systems biology, which integrates environmental, genetic and epigenetic factors as well as transcriptomic, proteomic, metabolomic and intermediate (e.g., physiological) outcome variables, ideally within the same population [67].

Variation in methylation states and the abundance of molecular levels (transcripts, proteins and metabolites) in cellular systems can be treated as quantitative traits. Their associated genomic loci are therefore called *quantitative trait loci* (QTLs) for methylation (methQTL), expression (eQTL), protein (pQTL), metabolites (mQTL), and physiological traits (phQTL), respectively. Studies in humans have investigated the genomic architecture of methylation [68], gene expression [69], lipids [70], and other proteins and metabolites of clinical importance [71]. The resulting comprehensive maps of different QTLs are valuable resources for prioritizing causal variants and designing functional experiments. Integrating such data from multiple molecular levels into explanatory models (i.e., systems genetics) provides a powerful holistic approach to study complex traits and holds several promises.

In the context of evolutionary adaptation, systems genetics may provide insight into the robustness of biological systems and buffering of the propagation of genomic variation to the phenotype level. The HapMap and 1000 Genome projects have catalogued many millions of genetic variants in the human genome. However, robustness at the phenotypic level is essential to keep processes and traits in any living organism within narrow limits, even in the face of all this genetic variation. We were one of the first to provide system-wide molecular evidence for phenotypic buffering using a systems genetics approach in a model system [72]. That is, the largest fraction of genomic and transcriptomic variants is silent at the phenotypic level with only a few influential QTL hot spot regions causing major phenotypic variation across a wide range of environments. These results are in agreement with recent findings that many human diseases share their genetic origin with other diseases to some extent [73]. Fragilities at crucial nodes in the molecular networks may underlie this phenomenon.

One important promise of systems genetics relevant for complex diseases, is its potential to improve understanding of the way genetic information is integrated, coordinated and ultimately

transmitted through molecular, cellular and physiological networks to enable higher-order functions and alterations of phenotypes. Causal inference through the construction of causal networks can provide insight into the route from genotype to phenotype. For example, eQTL maps have provided an important reference source for categorizing the effect of disease-associated SNPs on the expression of genes [74]. SNPs that affect expression of genes at larger distances or on different chromosomes (trans-eQTLs) allow us to identify affected genes downstream, with the potential to reveal novel pathways underlying disease etiology [75]. Similarly, QTLs for proteins and metabolites may coincide with disease-associated SNPs as illustrated in Fig. 3 and exemplified by the recent identification of *SORT1* as the causal gene responsible for the CAD GWA signal at the 1p13 locus (Table 1). Munsunuru et al. [76] integrated eQTL and pQTL (for LDL) information at that locus and uncovered that a novel pathway for lipoprotein metabolism regulated by altered expression of the *SORT1* gene underlies the MI etiology.

Systems genetics further offers a means to investigate gene–environment interactions to enhance insight into the pathophysiology of complex diseases. Such interactions involve plasticity of genetic regulation responding to both internal environments (tissues and cell types) [77–80] and external environments [81]. Our recent comparison of gene expression between blood samples and four primary tissues (liver, subcutaneous and visceral adipose tissue and skeletal muscle) characterized four different tissue-dependent manners of genetic regulation of nearby genes (i.e., in *cis*): specific *cis*-regulation, alternative *cis*-regulation, different effect sizes and opposite allelic effects. We further showed that SNPs associated with complex diseases more often exert a tissue-dependent effect on gene expression. As shown for the *SORT1* gene, the MI risk variant alters the expression of *SORT1* in the liver, but not in blood, adipose tissues or muscle [76]. These findings highlight the importance of investigating genetic effects in disease-relevant tissues and environments, in order to correctly characterize the functional effects of disease-associated variants.

5. Challenges and prospects of systems genetics for CAD

Systems genetics is a powerful method, but applying the approach to the study of complex diseases such as CAD in humans is still a challenge and requires the development of more sophisticated experimental strategies and statistical models. First, the most ideal experiment is to perform system-wide profiling on genome, epigenome, transcriptome, proteome, metabolome and phenome on the same subjects. Integrating "omics" data from different experiments on different subjects can only provide indirect support for etiological hypotheses. Second, we are largely unable to control the effect of environmental factors in human genetics. Environmental factors can have different effects on

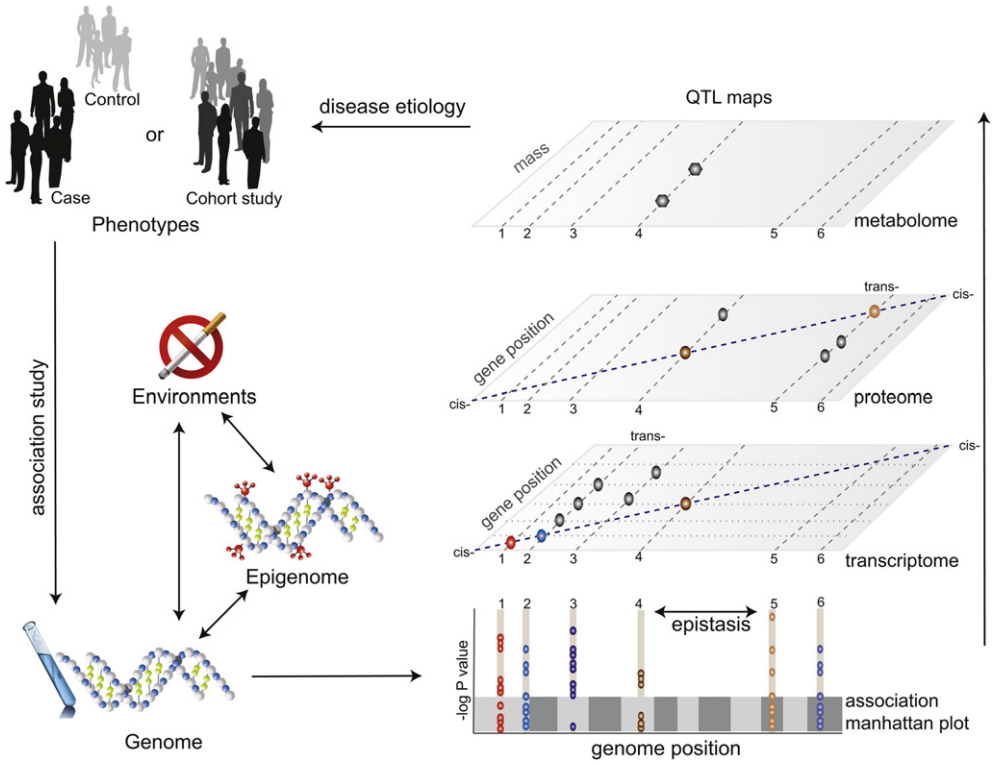

Fig. 3. Systems genetics: from finding sources of missing heritability towards understanding the complex biological networks that underlie complex diseases. Systems genetics aims at constructing a holistic view of biological processes by integrating data from multiple molecular levels into explanatory models of complex diseases. Comprehensive "omics" data from transcripts, proteins and metabolites are used in order to explain how these affect the final disease outcome. GWA studies using case–control or cohort designs may discover underlying risk loci as illustrated by the six significant loci in the association ("Manhattan") plot. Above the Manhattan plot, maps of quantitative trait loci (QTLs) at the level of transcriptome (eQTL), proteome (pQTL) and metabolome (mQTL) are shown. The dots on the QTL maps represent the QTLs at the six risk loci. The *x*-axis of the QTL maps indicates the genome position of the QTLs corresponding to the six risk loci. The *y*-axis for the eQTL and pQTL maps represents the position of genes affected by the risk variants. If the affected genes physically locate at the risk loci (the dots at the blue dashed diagonal line) these are called *cis*-QTLs. If the affected genes reside at different genomic regions (the dots at the grey dashed vertical line) these are called *trans*-QTLs. The *y*-axis of the mQTL map refers to the mass of the metabolites. Risk variants can have different effects on the different molecular levels. For example, there is only a *cis*-eQTL at locus1, both *cis*- and *trans*-eQTLs at locus 2; a *cis*-eQTL, a *cis*- and *trans*-pQTL and mQTLs at locus 4; *cis*- and *trans*-pQTLs at locus 5. Through integration of the genetic effects on multiple levels as well as interactions with environments, the epigenome and amongst the genetic effects (i.e., epistasis), systems genetics endeavours to model the causal network that underlies disease etiology.

"omics" levels than on disease endpoints. For example, smoking and diet can have an acute effect on "omics" levels but a chronic effect on disease outcomes. Compared to case–control studies, the prospective cohort study, in which a group of individuals is followed over time and potential disease outcomes predicted on the basis of factors such as genetics, molecular biomarkers, physiological traits and environmental exposures, will become more valuable in human genetic research [82]. The advantages of this cohort design include better definition of environmental exposures and better characterization of disease and risk phenotypes over time. For example, the LifeLines cohort in the Northern three provinces of The Netherlands, will eventually include 165,000 participants that will be followed for 30 years [83]. Approximately 1000 individuals with MIs will be expected in this cohort after five years of follow up. Through integration of systems genetics into the prospective cohort design this study offers great promise for improving our understanding of the causes and prognosis of the burden of CAD. However, considerable investments in bioinformatics and statistical genetics are necessary in order to deliver on this promise, because the complexity of the statistical analysis and required sample size to correctly infer causality constitute a third challenge. Omics data is most valuable when the different layers of data on genome, epigenome, transcriptome, proteome, metabolome and phenome (Fig. 3) are mathematically integrated into predictions of the underlying causal networks. However, the robustness of biological systems as mentioned earlier may lead to non-linear relationships between these layers [72]. Even for linear relationships, fairly large sample sizes are required to reliably discriminate between different directions of effects (causal, modifying or independent relationships) between two traits associated with the same locus. A simulation study for this simple scenario showed that a GWA study population size >10,000 is needed to provide 50% sensitivity and 90% positive predictive value for causal inference and realistic QTL effect sizes [84]. On the upside, structural and functional data (gene sequences, gene ontology, metabolic pathways, and protein–protein interactions) as well as independent experimental data gleaned from secondary sources (e.g., gene expression databases) can be used post-hoc to verify the defined gene and causal interferences.

In conclusion, there is no doubt that the GWA approach has been successful in identifying and elucidating previously unexpected genetic candidates for CAD/MI. Including the recent GWA studies with larger sample sizes a total of 32 CAD loci have been identified (Table 1). Despite the numerous successes, the GWA approach has not delivered on some of its promises. A large proportion of heritability remains unexplained and where to find the missing heritability (e.g., largely due to rare variants or to common variants with very small effect) has been hotly debated [32,34]. On the one hand, some studies focused on common variants and estimated that a substantial proportion of variation for a range of common complex traits and diseases can be explained by considering all common SNPs across the genome simultaneously. Examples include Crohn's disease (~24%), bipolar disorder (~41%), type 1 diabetes (~32%), height (~45%), BMI (~17%), von Willebrand factor (~25%) and QT interval (~21%) [85–87]. On the other hand, it was suggested that the GWA associations of common SNPs may result from multiple unobserved rare variants that are in LD with the common SNP; so-called synthetic associations [88]. However, others have argued that the empirical data does not support this hypothesis [89,90]; where both rare and common alleles are uncovered at the same locus, it is much more likely they constitute independent signals [91]. Finally, some studies argued that current estimates of total heritability may be significantly inflated [92], although assumption free methods to estimate heritability do not confirm this [93]. Arguments on the mystery of missing heritability are likely to continue to rage in human genetics and discussions may benefit from complementary information on model organisms such as mouse, rat and *Drosophila melanogaster* [94]. Here, we further argue that finding (part of) the missing heritability by itself constitutes only a first step towards a fuller understanding of the mechanisms underlying complex diseases. First, complex diseases are the product of the complex interplay between genetic, epigenetic and environmental factors. These interactions are not captured by the (narrow-sense) heritability estimate. Second, even if association can be detected between genotype and phenotype, drawing causal conclusions remains a major challenge. We propose a systems genetics approach within a prospective epidemiological cohort design that integrates molecular traits, including transcripts, metabolites and proteins and a range of (physiological) endophenotypes for CAD. The success of identifying the causal variant and underlying mechanism of the *SORT1* locus illustrates the added value of the systems genetics approach. Although the system-wide application of this approach in humans will require major investments in terms of sample collection, time, and computing-power, in combination with the prospective cohort design it offers great promise in elucidating underlying mechanisms of CAD development. If successful, its findings will not only have implications for disease therapy, but through improvement of risk prediction will also allow prevention efforts to be targeted to those most at risk for CAD [18,95].

Acknowledgements

The authors are indebted to Behrooz Z. Alizadeh and Ilja M. Nolte for their critical review and comments on the manuscript. J. Fu was funded by the Netherlands Organization for Scientific Research (NWO) VENI grant 863.09.007.

Appendix A. Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atherosclerosis.2012.05.015.

References

[1] Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics–2011 update: a report from the American Heart Association. Circulation 2011;123:e18–209.
[2] Marenberg ME, Risch N, Berkman LF, et al. Genetic susceptibility to death from coronary heart disease in a study of twins. N Engl J Med 1994;330:1041–6.
[3] Schildkraut JM, Myers RH, Cupples LA, et al. Coronary risk associated with age and sex of parental heart disease in the Framingham Study. Am J Cardiol 1989;64:555–9.
[4] Zdravkovic S, Wienke A, Pedersen NL, et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J Intern Med 2002;252:247–54.
[5] Liu H, Liu W, Liao Y, et al. CADgene: a comprehensive database for coronary artery disease genes. Nucleic Acids Res 2011;39:D991–6.
[6] Ozaki K, Inoue K, Sato H, et al. Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. Nature 2004;429:72–5.
[7] Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 2002;32:650–4.
[8] Kimura A, Takahashi M, Choi BY, et al. Lack of association between LTA and LGALS2 polymorphisms and myocardial infarction in Japanese and Korean populations. Tissue Antigens 2007;69:265–9.
[9] Ozaki K, Sato H, Inoue K, et al. SNPs in BRAP associated with risk of myocardial infarction in Asian populations. Nat Genet 2009;41:329–33.
[10] Takeuchi F, Yokota M, Yamamoto K, et al. Genome-wide association study of coronary artery disease in the Japanese. Eur J Hum Genet 2011;20:333–40.
[11] Tregouet DA, Konig IR, Erdmann J, et al. Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. Nat Genet 2009;41:283–5.

[12] Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334–41.
[13] Erdmann J, Grosshennig A, Braund PS, et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet 2009;41:280–2.
[14] Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357:443–53.
[15] McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. Science 2007;316:1488–91.
[16] Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science 2007;316:1491–3.
[17] Ioannidis JP. Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. Circ Cardiovasc Genet 2009;2:7–15.
[18] Humphries SE, Drenos F, Ken-Dror G, et al. Coronary heart disease risk prediction in the era of genome-wide association studies: current status and what the future holds. Circulation 2010;121:2235–48.
[19] Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med 2009;361:2518–28.
[20] Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 2011;43:333–8.
[21] Wang F, Xu CQ, He Q, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. Nat Genet 2011;43:345–9.
[22] Erdmann J, Willenborg C, Nahrstaedt J, et al. Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23. Eur Heart J 2011;32:158–68.
[23] Reilly MP, Li M, He J, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. Lancet 2011;377:383–92.
[24] Davies RW, Wells GA, Stewart AF, et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. Circ Cardiovasc Genet 2012;5:217–25.
[25] The Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet 2011;43:339–44.
[26] Barbalic M, Reiner AP, Wu C, et al. Genome-wide association analysis of incident coronary heart disease (CHD) in African Americans: a short report. PLoS Genet 2011;7. e1002199.
[27] Wild PS, Zeller T, Schillert A, et al. A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. Circ Cardiovasc Genet 2011;4:403–12.
[28] Keating BJ, Tischfield S, Murray SS, et al. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. PLoS One 2008;3. e3583.
[29] Visscher PM, Hill WG, Wray NR. Heritability in the genomics era–concepts and misconceptions. Nat Rev Genet 2008;9:255–66.
[30] Bansal V, Libiger O, Torkamani A, et al. Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet 2010;11:773–85.
[31] Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 2010;11:415–25.
[32] Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–53.
[33] McCarroll SA. Extending genome-wide association studies to copy-number variation. Hum Mol Genet 2008;17:R135–42.
[34] Eichler EE, Flint J, Gibson G, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 2010;11:446–50.
[35] Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 2007;17:1520–8.
[36] Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 2010;42:570–5.
[37] Preuss M, Konig IR, Thompson JR, et al. Design of the coronary ARtery DIsease genome-wide replication and meta-analysis (CARDIoGRAM) study: a genome-wide association meta-analysis involving more than 22 000 cases and 60 000 controls. Circ Cardiovasc Genet 2010;3:475–83.
[38] Rhoads GG, Dahlen G, Berg K, et al. Lp(a) lipoprotein as a risk factor for myocardial infarction. JAMA 1986;256:2540–4.
[39] Snieder H, van Doornen LJ, Boomsma DI. The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. Am J Hum Genet 1997;60:638–50.
[40] Snieder H, van Doornen LJ, Boomsma DI. Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: lessons from twin studies. Arterioscler Thromb Vasc Biol 1999;19:2826–34.
[41] Beekman M, Heijmans BT, Martin NG, et al. Two-locus linkage analysis applied to putative quantitative trait loci for lipoprotein(a) levels. Twin Res 2003;6:322–4.
[42] Boerwinkle E, Leffert CC, Lin J, et al. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. J Clin Invest 1992;90:52–60.
[43] Falchi M, Andrew T, Snieder H, et al. Identification of QTLs for serum lipid levels in a female sib-pair cohort: a novel application to improve the power of two-locus linkage analysis. Hum Mol Genet 2005;14:2971–9.
[44] Lanktree MB, Hegele RA. Gene-gene and gene-environment interactions: new insights into the prevention, detection and management of coronary artery disease. Genome Med 2009;1:28.
[45] Kivimaki M, Magnussen CG, Juonala M, et al. Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis: the Young Finns Study. Int J Epidemiol 2011;40:470–8.
[46] Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or "common gene, common disease"? Methods Mol Biol 2007;376:71–84.
[47] Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69:124–37.
[48] Gorlov IP, Gorlova OY, Sunyaev SR, et al. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. Am J Hum Genet 2008;82:100–12.
[49] Lage K, Karlberg EO, Storling ZM, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 2007;25:309–16.
[50] Zhu X, Feng T, Li Y, et al. Detecting rare variants for complex traits using family and unrelated data. Genet Epidemiol 2010;34:171–87.
[51] Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. Science 2004;305:525–8.
[52] Bochukova EG, Huang N, Keogh J, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature 2010;463:666–70.
[53] Kraft HG, Lingenhel A, Kochl S, et al. Apolipoprotein(a) kringle IV repeat number predicts risk for coronary heart disease. Arterioscler Thromb Vasc Biol 1996;16:713–9.
[54] Stangl K, Cascorbi I, Laule M, et al. High CA repeat numbers in intron 13 of the endothelial nitric oxide synthase gene and increased risk of coronary artery disease. Pharmacogenetics 2000;10:133–40.
[55] Yuan G, Wang J, Hegele RA. Heterozygous familial hypercholesterolemia: an underrecognized cause of early cardiovascular disease. CMAJ 2006;174:1124–9.
[56] Craddock N, Hurles ME, Cardin N, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010;464:713–20.
[57] Kwon HJ, Lagace TA, McNutt MC, et al. Molecular basis for LDL receptor recognition by PCSK9. Proc Natl Acad Sci U S A 2008;105:1820–5.
[58] Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 2009;10:392–404.
[59] Peng G, Luo L, Siu H, et al. Gene and pathway-based second-wave analysis of genome-wide association studies. Eur J Hum Genet 2010;18:111–7.
[60] Sabatti C, Service SK, Hartikainen AL, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 2009;41:35–46.
[61] Baccarelli A, Tarantini L, Wright RO, et al. Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study. Epigenetics 2010;5.
[62] Feinberg AP, Irizarry RA. Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A 2010;107-(Suppl. 1):1757–64.
[63] Kong A, Steinthorsdottir V, Masson G, et al. Parental origin of sequence variants associated with complex diseases. Nature 2009;462:868–74.
[64] Castro R, Rivera I, Struys EA, et al. Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease. Clin Chem 2003;49:1292–6.
[65] Kim M, Long TI, Arakawa K, et al. DNA methylation as a biomarker for cardiovascular disease risk. PLoS One 2010;5. e9692.
[66] Sharma P, Kumar J, Garg G, et al. Detection of altered global DNA methylation in coronary artery disease patients. DNA Cell Biol 2008;27:357–65.
[67] Nadeau JH, Dudley AM, Genetics. Systems genetics. Science 2011;331:1015–6.
[68] Morris MR, Ricketts CJ, Gentle D, et al. Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. Oncogene 2011;30:1390–401.
[69] Staal FJ, de Ridder D, Szczepanski T, et al. Genome-wide expression analysis of paired diagnosis-relapse samples in ALL indicates involvement of pathways related to DNA replication, cell cycle and DNA repair, independent of immune phenotype. Leukemia 2010;24:491–9.
[70] Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707–13.
[71] Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;42:137–41.
[72] Fu J, Keurentjes JJ, Bouwmeester H, et al. System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat Genet 2009;41:166–7.
[73] Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet 2009;10:43–55.
[74] Cookson W, Liang L, Abecasis G, et al. Mapping complex disease traits with global gene expression. Nat Rev Genet 2009;10:184–94.
[75] Fehrmann RS, Jansen RC, Veldink JH, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on

intermediate genes, with a major role for the HLA. PLoS Genet 2011;7. e1002197.
[76] Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature 2010;466:714–9.
[77] Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 2009;325: 1246–50.
[78] Gerrits A, Li Y, Tesson BM, et al. Expression quantitative trait loci are highly sensitive to cellular differentiation state. PLoS Genet 2009;5. e1000692.
[79] Nica AC, Parts L, Glass D, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 2011;7. e1002003.
[80] Fu J, Wolfs MG, Deelen P, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet 2012;8. e1002431.
[81] Li Y, Alvarez OA, Gutteling EW, et al. Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. PLoS Genet 2006;2. e222.
[82] Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet 2006;7:812–20.
[83] Stolk RP, Rosmalen JG, Postma DS, et al. Universal risk factors for multifactorial diseases: lifelines: a three-generation population-based study. Eur J Epidemiol 2008;23:67–74.
[84] Li Y, Tesson BM, Churchill GA, et al. Critical reasoning on causal inference in genome-wide linkage and association studies. Trends Genet 2010;26: 493–8.
[85] Lee SH, Wray NR, Goddard ME, et al. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet 2011;88:294–305.
[86] Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010;42:565–9.
[87] Yang J, Manolio TA, Pasquale LR, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 2011;43:519–25.
[88] Dickson SP, Wang K, Krantz I, et al. Rare variants create synthetic genome-wide associations. PLoS Biol 2010;8. e1000294.
[89] Anderson CA, Soranzo N, Zeggini E, et al. Synthetic associations are unlikely to account for many common disease genome-wide association signals. PLoS Biol 2011;9. e1000580.
[90] Wray NR, Purcell SM, Visscher PM. Synthetic associations created by rare variants do not explain most GWAS results. PLoS Biol 2011;9. e1000579.
[91] Visscher PM, Brown MA, McCarthy MI, et al. Five years of GWAS discovery. Am J Hum Genet 2012;90:7–24.
[92] Zuk O, Hechter E, Sunyaev SR, et al. The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 2012;109:1193–8.
[93] Visscher PM, Medland SE, Ferreira MA, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2006;2. e41.
[94] Aitman TJ, Boone C, Churchill GA, et al. The future of model organisms in human disease research. Nat Rev Genet 2011;12:575–82.
[95] Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. Lancet 2010;375:1525–35.