



Published in final edited form as:

J Am Coll Cardiol. 2015 March 3; 65(8): 830–845. doi:10.1016/j.jacc.2014.12.033.

Genome-Wide Significant Loci: How Important Are They?:

Systems Genetics to Understand Heritability of Coronary Artery Disease and Other Common Complex Disorders

Johan L.M. Björkegren, MD, PhD^{*,†,‡,§}, Jason C. Kovacic, MD, PhD[†], Joel T. Dudley, PhD^{*}, and Eric E. Schadt, PhD^{*}

^{*}Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York

[†]Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, New York

[‡]The Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

[§]The Department of Pathological Anatomy and Forensic Medicine, University of Tartu, Tartu, Estonia

Abstract

Genome-wide association studies (GWAS) have been extensively used to study common complex diseases such as coronary artery disease (CAD), revealing 153 suggestive CAD loci, of which at least 46 have been validated as having genome-wide significance. However, these loci collectively explain <10% of the genetic variance in CAD. Thus, we must address the key question of what factors constitute the remaining 90% of CAD heritability. We review possible limitations of GWAS, and contextually consider some candidate CAD loci identified by this method. Looking ahead, we propose *systems genetics* as a complementary approach to unlocking the CAD heritability and etiology. Systems genetics builds network models of relevant molecular processes by combining genetic and genomic datasets to ultimately identify key “drivers” of disease. By leveraging systems-based genetic approaches, we can help reveal the full genetic basis of common complex disorders, enabling novel diagnostic and therapeutic opportunities.

Keywords

atherosclerosis; atherosclerotic plaque; genome-wide association study; myocardial infarction; primary prevention; regulatory gene networks

© 2015 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION PUBLISHED BY ELSEVIER INC.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Johan L.M. Björkegren, Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1498, New York, New York 10029-6574. johan.bjorkegren@mssm.edu.

APPENDIX for a supplemental table, please see the online version of this article.

INTRODUCTION

GENETICS OF RARE SINGLE-GENE DISORDERS—THE FOUNDATIONS OF THE SEARCH FOR HERITABILITY OF COMMON DISORDERS

The heritability of traits between generations is principally carried in deoxyribonucleic acid (DNA), mainly as single nucleotide polymorphisms (SNPs), insertions, deletions, or copy number variants. Another type of heritability, believed to be independent of DNA (1,2), is carried by epigenetic mechanisms, which are attributed with inducing major shifts in DNA transcription (3). Epigenetics is not a topic of this paper, but has been reviewed extensively elsewhere (4). In the case of rare disorders, most carriers of 1 or more highly penetrant risk variant develop the disease at some point (Figure 1A). However, it is also common that some persons carrying a potentially causative variant will not develop that disorder; thus, the penetrance is typically <100%. Penetrance is a time-dependent aspect of phenotypic disease expression. For example, both Huntington disease and cystic fibrosis are virtually 100% penetrant: cystic fibrosis usually soon after birth and Huntington disease by about 70 years of age. In contrast, familial breast cancer-associated mutations in the BRCA1 gene have a lifetime penetrance of 60% to 85%. That risk variants do not necessarily lead to manifest disease in all carriers has been highlighted as a possible path to identify genetic and environmental mechanisms that confer resistance to certain rare diseases (5). Mechanisms that buffer against disease could be potential therapeutic targets.

Despite the high penetrance of disease-causing variants, rare diseases are, by definition, rare. As a result of differences in genetic ancestry, the spectrum and relative frequencies of disease-associated alleles vary among different populations. For example, cystic fibrosis is most common in populations originating from northern and western Europe (~1 in 2,000), whereas sickle-cell anemia is more common in African and Afro-Caribbean populations (~1 in 3,000). Through careful characterization of symptoms and disease phenotypes in families carrying rare disorders, the causes of nearly one-half of the estimated 7,000 single-gene disorders have been identified, mainly by using markers found at increased density across the genome and by linkage analysis in pedigrees of disease-carrying families (6). Although these were key techniques in the discovery of highly-heritable single-gene disorders that typically only affect a limited number of biological pathways and tissues, their scope is inadequate for genetic studies of complex disorders such as atherosclerosis and coronary artery disease (CAD), where disease inheritance is blended with environmental risk factors, and causative genes are likely to be operative across several tissues (7).

GENETICS OF COMMON COMPLEX DISORDERS—GENOME-WIDE ASSOCIATION STUDIES

The successful identification of the genetic causes and mechanisms of many rare single-gene disorders inspired scientists to use a similar approach to study common complex disorders. Because such diseases are widespread, linkage analysis within families was not believed to be appropriate. Furthermore, it was expected (and now confirmed) that many genetic signals underlie common complex disorders, each with a relatively weak effect (e.g., odds ratio: <1.5). A study design was instead chosen that analyzed increasingly dense genomic markers

in thousands of mostly unrelated individuals in case-control association studies. From these lines of argument, the genome-wide association study (GWAS) design was born (8).

One of the first GWAS of CAD came from the Wellcome Trust Case Control Consortium, which discovered that the chromosome 9 locus was associated with CAD (9–11). Since then, 153 suggestive DNA variants associated with CAD have been discovered by GWAS, of which 46 were replicated in metaanalyses of genome-wide association (GWA) datasets (12). These CAD-associated loci are strikingly pervasive across the population, but generally have weak effects. As recently reviewed (8), 50% of the CAD-associated variants occur in over one-half of the population, and at least 25% occur in over 75% of the population. However, each variant usually confers a minimal to modest increase in relative risk, averaging only 18% (corresponding to an odds ratio of 1.18). A common theme of recent GWAS reviews is the success of this approach, both for CAD and for other, more or less complex disorders, for which more than 1,000 loci have been identified (13). Certainly, the sheer number of loci discovered, which has also led to the discovery of many previously unknown disease-causing genes, is a major success. Nevertheless, for most common complex disorders, the combined contribution of these loci to disease variation in the population is frequently <10%. Indeed, the 153 known CAD-associated variants explain <10.6% of the likely genetic variation across the population (12). Thus, ~90% of the heritability of CAD, and of most other common complex disorders, remains unexplained by identified GWA loci, despite the inclusion of remarkably large numbers of subjects: a recent meta-analysis of several GWAS for CAD comprised 63,746 cases and 130,681 controls (12). Although a GWAS-based approach will clearly not reveal the full extent of the heritability of CAD and other common complex disorders, more complete sequencing techniques, particularly expanded whole-exome/whole-genome sequencing (WES/WGS), typically applied to the same case-control cohorts previously used in GWAS, promise to reveal additional rare risk variants, perhaps with larger effects on heritability (14). These results and additional refinements of the analysis of existing GWAS (14) will contribute to reducing the large fraction of missing heritability, but to what extent?

When addressing the “missing heritability” it is also reasonable to question the reliability of CAD heritability estimates. Because the overall fraction of genetic variance in CAD is less than we believe, is missing heritability largely (~90%) exaggerated? From traditional analysis of family pedigrees in twins, the range of genetic variance in CAD is between 40% and 60% (15). Assuming 40% heritability, the 153 genome-wide significant SNPs explain 10.6% of CAD variability. Another way of assessing heritability using GWA datasets is to consider all measured SNPs (16). When applying this “polygenic” model to the complex trait of height, 294,831 SNPs were found to explain as much as 45% of height variance (16). The authors concluded that because individual effects are too small to pass the stringent significance tests ($p < 10^{-8}$) traditionally used in GWAS, most height heritability is not “missing,” but has simply not been detected in the GWA data. Thus, the range of 40% to 60% CAD heritability is probably reasonable. However, the notion of completely independent genetic and environmental risk factors needs reconsideration, as most genetic risk factors increasingly appear to be dependent on environmental influences (17).

GENETICS OF COMMON COMPLEX DISORDERS—THE SEARCH FOR MISSING HERITABILITY

Some of the missing heritability of CAD and other common complex disorders is likely carried by rare variants that may be identified in ongoing WES/WGS projects (as opposed to the relatively common variants discovered to date in GWAS [18]). In addition, epigenetic mechanisms will likely carry an as yet unknown fraction of complex disease heritability (4). Although defining the role of epigenetic mechanism requires other techniques, and despite ambiguity regarding how epigenetic modifications remain conserved across generations (1), these sources will undoubtedly reduce the fraction of missing heritability. Will they provide the full picture? Or are there fundamental problems with how we have been trying to understand and therefore seek heritability for complex traits? We believe this may be the case and that it is timely to critically review our understanding of complex disease inheritance by more carefully investigating how genetic risk variants interact with environmental factors to cause disease.

In this review, we question the notion that the heritability of common complex disorders is best revealed by traditional analyses of DNA sequence data in isolation, as thus far performed in GWAS and WES/WGS projects. Sequential analyses of DNA variants in these studies results in an overwhelming multiple-testing problem, and it can be questioned whether simply making the case-control cohorts larger to enable smaller differences to become genome-wide significant is the only reasonable path to follow. We also critically appraise the potential importance of identified risk loci by examining the suggested roles of their putative candidate genes in disease development. Using CAD development as an example, we suggest that genome-wide significant risk loci most likely underlie an early and protracted phase of CAD development, but are unlikely to regulate the rapid or late phases of CAD development that culminate in clinical events (Figure 1B). We instead believe that there may be a more important subpopulation of risk variants, which exert their effects on CAD only in certain environmental contexts. The context-dependence of this risk variant subpopulation is likely to render itself nominally significant ($p < 0.05$) (Figure 1C) in traditional analyses of GWA datasets considering DNA variants in sequence, but it would fail to reach genome-wide significance ($p < 10^{-8}$) (Figure 2) (17). To grasp these assumptions, it is vital to understand how complex diseases develop. In the following section, we evaluate atherosclerosis and plaque development from the perspective of the likely timing and pathogenesis of genetic effects in CAD.

ATHEROSCLEROSIS DEVELOPMENT CULMINATING IN CLINICAL MANIFESTATIONS—MYOCARDIAL INFARCTION AND STROKE

Generally, diseases are believed to develop according to a sigmoidal (S-shaped) curve (19): commencing slowly with a positive acceleration phase; then increasing rapidly, approaching an exponential growth rate as in a J-shaped curve; and finally, saturating, stabilizing at a near-zero growth rate.

The development of atherosclerotic lesions in the coronary tree generally agrees with this S-shaped model, although the final phase of plaque development and progression may be variable and can include further progression (Figure 1B) (20). The notion that

atherosclerosis develops slowly over a very long period, followed by more rapid progression, is supported by studies in mice (21,22) and humans (23,24). Briefly, early atherosclerotic lesions develop slowly, over 20 to 30 weeks in mice and probably over 30 to 40 years in humans, starting in adolescence (Figure 1B). Atherogenesis involves retention of circulating plasma lipoproteins, mainly low-density lipoprotein (LDL), at sites of turbulent blood flow. Some LDL particles remain and are modified by redox processes in the subendothelial space. Oxidized LDL subtly activates the endothelium, primarily by expressing adhesion molecules, which induce transendothelial migration of leukocytes, predominantly monocytes. Upon entering the subendothelial space and the intima, monocytes differentiate into macrophages, which take up oxidized LDL particles, initiating a key process of atherosclerosis: foam cell formation. As lipid-laden foam cells accumulate in the intima, fatty streaks appear as the first histologically-visible manifestation of atherosclerosis. This early phase of atherosclerosis development ends when foam cells within the fatty streaks start to aggregate, forming small atherosclerotic plaques with well-defined borders (21).

In the second phase, the small plaques expand rapidly, both across the arterial wall and, importantly, into the lumen of the artery, where they can compromise blood flow. The expansion phase is rapid (~10 weeks) in mice, and evidence from ^{14}C dating of human atherosclerotic plaques (23) suggests that it is also rapid (relative to lifespan) in humans (<10 years before clinical symptoms).

In the third and final phase, plaque biology can be quite variable, with approximately 30% of lesions showing rapid progression over 12 months to become fibroatheromas, with a lipid-rich core encapsulated by either a thin (10%) or thick cap (20%). Thin-cap fibroatheromas are considered the most unstable lesions and the most likely to lead to acute myocardial infarction. Over the subsequent 12 months, 75% of thin-cap fibroatheromas stabilize, whereas 5% of thick-cap fibroatheroma develop high-risk features (20,25). Furthermore, depending on its location, a mature plaque can have dramatic or minimal effects on blood flow, possibly leading to angina caused by myocardial ischemia of the heart muscle subtended by the plaque-narrowed artery.

A complex interplay of factors influences whether or not an advanced plaque ruptures, including: the extent and degree of lipid-core necrosis sustained by proliferating macrophages within the plaque; de novo monocyte migration/emigration; the degree of luminal stenosis; plaque burden; positive (outward) vessel remodeling; and the thickness of the fibrous cap (26). Rupture of a coronary plaque leads to intracoronary thrombus formation, which may or may not occlude the vessel. Although a nonocclusive thrombus may increase luminal narrowing and hasten lesion progression, an occlusive thrombus is associated with reduced myocardial perfusion that typically causes an acute coronary syndrome (myocardial infarction or unstable angina) or sudden death.

GWA LOCI FOR COMMON COMPLEX DISEASES—HOW IMPORTANT ARE THEY?

Genome-wide significant loci identified by GWAS conducted for complex diseases may fail to identify central pathological processes and corresponding key DNA variants that contribute to heritability. This notion rests primarily on 3 conventions about the

development and phenotypic expression of complex, relatively rare diseases (Figure 1A and B).

1. The case/control overlap—In comparing cases with population controls in studies of complex diseases (as opposed to rare diseases) (Figures 1A and B), overlap in central disease processes is inevitable, as these processes may be active without having caused clinical symptoms in the control subjects. Consequently, many DNA variants regulating genes active in these processes will not surface at a level of genome-wide significance in GWAS. However, due to a likely over-representation of these processes within cases (albeit not unique), DNA variants regulating these processes will likely instead present in GWAS with nominally significant p values.

2. The context of shifting environments—DNA variants that regulate genes active in central disease processes, and which do so independently of changing environmental contexts, are likely to surface with genome-wide significance in GWAS. In contrast, DNA variants that depend on preexisting environmental contexts to regulate genes in central disease processes are unlikely to surface with genome-wide significance in GWAS. A principal reason is that many of these contexts are variably present (active), probably resulting in context-dependent risk variants presenting with nominally-significant disease associations in GWAS (Figure 1C). Contexts presenting at the macroenvironmental level in CAD are mainly lifestyle factors (smoking, diet, and sedentary lifestyle) or other major disease risk factors such as obesity, diabetes, hypertension, and certain inflammatory diseases. Such factors can be considered in GWAS (27), but most have not yet been accounted for. Other macroenvironmental factors, like highly stressful events (such as death of a spouse or natural disaster) (28,29) also predispose to myocardial infarction, but are harder to define (and thus to consider) in the individual patient. Macroenvironmental factors inevitably result in alteration of a wide array of microenvironments specific to tissues and cell types. The microenvironment in a given cell or tissue is the final determinant of whether a context-dependent DNA variant will be active (affecting gene activity) or not. Take inflammation as an example of a variable microenvironmental context. When early “fatty streak” lesions develop into plaques with intact borders, intralésional foam cells are believed to initiate inflammatory gene activation, which subsequently causes the rapid expansion phase of lesion growth (21,22). Previously silent DNA variants that specifically affect the activation of these inflammatory genes will now be suddenly relevant to disease progression, only to again become silent (or less active) in the late phase of lesion development. There are many reasons why a DNA variant can have changing effects on the genes it regulates—the most obvious being that the genes regulated by a given DNA variant may, during early phases of disease development, be largely silent (not expressed, as exemplified with inflammatory genes). A more complex reason is that a microenvironmental context (e.g., inflammatory stimulation) activates a specific cotranscription factor whose binding and effect depend on a given allele of a DNA variant. In this scenario, the microenvironmental perturbation (that is, inflammation) is needed for the regulatory effect of the DNA variant. Context-dependency of gene expression in this fashion has experimentally been shown to be both common and strong (30–33).

3. The context of time—DNA variants that regulate genes in central disease processes that are active (relevant) over a long period are more likely to surface with genome-wide significance, those that regulate disease processes over a short period (e.g., late in disease development). This is in part related to point 2, because DNA variants that depend on shifting environments for their effect on disease processes also are likely to affect the disease over a shorter time period.

Taking these 3 conventions together, we suspect that DNA variants identified as having genome-wide significance by GWAS are likely to regulate early disease processes in the slow initial growth phase, rather than those in the rapid growth or late phases (Figure 1B). In CAD development, for example, this assumption is partly based on the following developmental characteristics. First, the pathobiology of early CAD development is likely more genetically driven than the later phases and is less affected by environmental exposures linked to later life stages, like diabetes, hypertension, obesity, sedentary lifestyle, and inflammatory states (all less prevalent in adolescence and the early adult years). Also, as the first and longest phase of CAD development, the inclination of the curve of the slow phase will be highly decisive for how case and control subjects are defined in GWAS (Figure 1C). Conversely, the most significant disease-associated DNA variants, defined by comparing case and control subjects (genome-wide significant variants), will likely point toward genes active in the early phase. In contrast, the late and rapid phases are driven more by disease processes that are often shared between case and control subjects (e.g., obesity, hypertension, dyslipidemia, and diabetes) and that are more influenced by environmental factors compared with the early phase. Thus, DNA variants affecting later CAD phases are more likely to be context-dependent. Adding to the complexity of the later phases is that they commonly involve many other parallel disease processes acting across several organs. For CAD, these organs are primarily metabolic (e.g., the liver and pancreas with diabetes, adipose tissue with obesity, or systemic immune activation). Thus, although late processes likely involve additional systemic contributions linked to context-dependent DNA variants that are unlikely to be explained by genome-wide significant loci detected by GWAS, early processes are more likely to be genetically driven and exposed to fewer confounding factors, and therefore, are more likely to contain regulatory DNA variants of genome-wide significance.

Multiple ongoing studies seek to better understand the mechanisms of the 46 genome-wide significant loci thus far identified by GWAS for CAD, and whether they relate to early or late events in CAD development (7). If (as we suspect) most of these CAD loci (and loci for other complex diseases that develop in a similar fashion) are related to early events in disease development, their clinical usefulness may be limited primarily to guiding preventive measures and, possibly, to developing therapies against early disease development (primary prevention). Conversely, the usefulness of these findings for secondary prevention, to prevent the rapid and late phase of CAD development, would be restricted.

To test the assumption that DNA variants with genome-wide significance mainly regulate early CAD development, we examined candidate genes assigned to CAD GWA loci (12). Of the 50 candidate genes proposed for 46 loci confirmed in a meta-analysis of GWA datasets,

10 are involved in regulating plasma lipid levels (7 for LDL, 1 for high-density lipoprotein, and 2 for triglycerides) (8). Plasma lipids are primarily of importance for driving early atherosclerosis development, consistent with the notion that loci identified by GWAS will be more useful for primary prevention and with the experimental finding that atherosclerosis regression in response to LDL lowering is much greater for early lesions than for mature and advanced lesions (22). An additional 6 candidate genes are involved in hypertension, which is clearly important for early endothelial activation; however, its importance for later phases of CAD is less clear. In fact, the guidelines for hypertension treatment in the elderly (>60 years of age) were recently altered; the blood pressure goal has now been eased to <150/90 mm Hg (34) because the risk for stroke and CAD in this group was not increased at the previous lower blood pressure limit (<140/85 mm Hg), as it is in younger people (35,36). These insights again highlight that the variants that drive early atherogenesis are likely to be over-represented among DNA variants with genome-wide significance in GWAS. Interestingly, the only GWA locus (rs579459) involving blood groups (and the gene ABO) is not linked to CAD, but is linked to myocardial infarction (37). Adding weight to our argument that SNPs identified by GWAS are generally associated with early CAD development, this ABO-associated locus is the only SNP thus far identified by GWAS associated with myocardial infarction (obviously a very late event in CAD development) (37).

For 35 of the 50 candidate genes, their role in atherosclerosis is unknown. Many, like the suggested mechanism for the 9p21 locus, appear to involve the vascular wall, which indicates that they are likely to be primarily involved in early atherosclerosis development. However, perhaps the most compelling evidence for the hypothesis that GWA loci identified for CAD primarily reflect early atherosclerotic development is the lack of hits for inflammatory or immune responses, which are thought to be a central and causal disease mechanism, particularly for late stages of atherosclerosis and CAD (38,39). Further supporting the notion of late activation, microarray studies of atherosclerotic lesions during their progression show activation of inflammatory genes predominantly in the late stages (21). We, therefore, suggest that the conspicuous absence of inflammatory regulators among currently-identified genome-wide significant variants for CAD strongly signals that the GWA approach does not capture the full spectrum of genetically-driven events of coronary atherosclerosis.

In summary, considerable evidence supports the notion that genes so far identified for GWAS loci predominantly regulate early CAD development. We expect that the extensive ongoing studies into the molecular mechanisms of the 46 confirmed GWA-defined CAD loci will shed light on this issue, as these mechanisms will likely be traceable to early versus late events in the pathogenesis of atherosclerosis.

SYSTEMS GENETICS-IDENTIFYING DISEASE-DRIVING NETWORKS AND THEIR GENETIC REGULATION

The proposed conventions synopsise into early CAD development being governed by genetic variants of molecular disease processes that are persistent and less affected by environmental contexts, as opposed to those governing later phases of CAD (Figure 1C).

They are consequently more likely to have surfaced as genome-wide significant in GWAS, at least from how these datasets have been analyzed to date.

A pertinent question then arises: how might one identify context-dependent DNA variants regulating disease processes active over a limited time, as in the rapid expansion and late phases of CAD development, that are not recognized as genome-wide significant variants in a traditional DNA analysis? As these late processes are believed to involve many (as opposed to isolated) causative DNA variants with varying context-dependence (shifting with microenvironments and time), we believe that the key lies in first defining the molecular processes driving these later phases in complex diseases, and then identifying the DNA variants that causally regulate them, thus allowing their contribution to heritability to be weighed. By first addressing the molecular underpinnings of variable complex disease processes, we may be able to unmask a substantial portion of the missing 90% of the heritability of CAD and other complex diseases.

How can this be best achieved? The increasingly frequent answer is *systems genetics* (40–43). To us, the aim of systems genetics can be summarized as using genomic activity measures (e.g., ribonucleic acid [RNA], proteins, metabolites, and DNA modifications) to define disease-driving molecular processes and integrate them with GWA datasets, thereby permitting their contribution to complex disease heritability to be understood. However, the ultimate goal must be to enable diagnosis and treatment of patients on the basis of the status of these complex disease processes and to modulate pathological activity toward a non-pathological state.

It is increasingly understood that individual genetic variants, individual genes, or even linear pathways will never explain the intrinsic complexity of molecular processes underlying common diseases like CAD. Instead, these processes have polygenetic regulation and consist of multiple genes interacting in highly complex, fluid, and dynamic biologic networks reminiscent of intricate wiring diagrams. Fortunately, biological networks are sparse, with most genes (nodes) having only a small number of interaction(s) with other genes (edges), and with only a few highly interconnected nodes acting as hubs with many edges (44). These features can be identified from measures of genome activity (45). Furthermore, biological networks are well-conserved throughout evolution and, because of built-in redundancy, are biologically robust to an individual node's loss (46). In parallel, technological advances in screening genomes and genome activity with evergreater reliability and lower cost, together with increasing capacities for computational analysis of large datasets, have set the stage for more widespread use of systems genetics in biology, medicine, and health care (17).

Presently, causal disease networks are mostly inferred from the combination of genotype (DNA) and gene expression data in *genetics of gene expression studies* (GGES) (Figure 3). Although beyond the scope of this review, this is achieved using network inference algorithms for coexpression (i.e., weighted coexpression networks analysis), Bayesian probabilistic network models (47–50), and direct statistical tests for causality (51,52). To date, most algorithms are designed to infer disease networks from gene expression data generated by microarrays. More recently, modified algorithms that also infer biological

networks from heterogeneous next-generation sequence datasets (e.g., RNA sequence) are emerging (53).

In our CAD research, we have focused on GGES of multiple tissues (Central Illustration), namely the STAGE (Stockholm Atherosclerosis Gene Expression) (7) and STARNET (Stockholm Tartu Atherosclerosis Reverse Network Engineering Task) studies. STAGE was a pilot study for STARNET, with 100 and 900 cases, respectively. Subjects were recruited from patients undergoing open thorax surgery; those having coronary artery bypass grafting served as cases, and those without atherosclerosis or CAD (confirmed by pre-operative angiography) undergoing other forms of open thorax surgery (e.g., isolated mitral valve repair) served as controls.

Parallel sampling of up to 9 CAD-relevant tissues from each patient is a key aspect of the STAGE and STARNET studies (7). RNA samples from case and control subjects were obtained from the arterial wall, liver, visceral abdominal fat, skeletal muscle, subcutaneous fat, primary monocytes, and monocytes that were differentiated in vitro into macrophages and foam cells. The 9 RNA samples were then converted into microarray data (STAGE, custom-made HuRSTA-2a520709 Affymetrix arrays [Affymetrix, Santa Clara, California]) and, more recently, RNA sequence data (STARNET). These RNA expression datasets are now used: 1) to infer causal regulatory disease-driving molecular processes, as reflected in gene networks operating both within and across tissues to cause CAD; and 2) to identify DNA variants that modulate these networks (7). We believe that the STAGE/ (STARNET datasets are unique in allowing us, for the first time to our knowledge, to study the inherent complexity of the molecular process underlying the late, possibly rapid phases of CAD development across the 9 collected tissues.

Even before inferring gene networks, the STAGE/ STARNET dataset can be used to identify expression quantitative trait loci (eQTLs) (54–58) in CAD, especially as related to established GWA loci. An eQTL is a DNA variant (frequently an SNP) that regulates gene expression levels. They are determined by linking alleles of the SNPs found by genotyping the patient's DNA (e.g., in STAGE and STARNET from Affymetrix GenomeWideSNP_6 arrays) with gene expression levels from the various tissues. SNP alleles associated with different levels of gene expression, and therefore acting as eQTLs, can then be identified.

Using the STAGE data, we recently identified 8,156 eQTLs for 6,450 unique genes across 7 CAD-relevant tissues (59). By integrating the analysis with 2 independent GWA datasets for CAD, the Myocardial Infarction Genetics Consortium (60) and the Wellcome Trust Case Control Cohort (61) to assess the enrichment of these eQTLs in inherited risk, we discovered that those eQTLs regulating gene activity across greater numbers of tissues lead to increased CAD risk. Furthermore, eQTLs that were operative across several tissues resided at regulatory genomic “hot spots” (62). In contrast, most of the 22 eQTLs identified in the STAGE study that were established as “CAD GWA hits” affect gene expression in a single tissue, or, at most, in 2 tissues. In our view, the multitissue involvement of the risk-enriched eQTLs suggests that they regulate molecular processes acting across several of these tissues (63) in late CAD development and also may be important contributors to the inherited CAD risk. In contrast, that the 22 STAGE eQTLs were established GWA loci for CAD regulated

genes mostly in 1 or 2 tissues is consistent with their involvement in early CAD development.

The STAGE/STARNET datasets were next used to define groups of genes acting together in modules and networks, primarily on the basis of similar coexpression (17,52,64,65) across the 9 tissues. After identifying these modules, an important next step is to link them to relevant patient phenotypes. For example, we calculated the eigengene value representing the sum of all gene expression values in a given module/network (66), which can subsequently be used to correlate modules with phenotypic characteristics of STAGE and STARNET patients. Gene modules associated with key CAD phenotypes, such as the angiographic SYNTAX score (from preoperative angiograms) and plasma LDL cholesterol levels, can thus be identified. For modules with strong phenotypic associations, Bayesian network algorithms or other statistical causal inference techniques are applied, incorporating information on eQTLs to determine modules that are causally related to CAD, as opposed to those that are reactive or independent. The Bayesian networks can also be employed to infer key driver genes important for regulating the state of the module (50,67,68), where they may serve as diagnostic markers, therapeutic targets, or both (see “Drug Repurposing” section).

Applying a systems genetics approach in clinical medicine (46) is certainly not limited to DNA genotype and RNA expression studies (Central Illustration), and can include more than 1 targeted tissue (Figure 3) (69). The more clinical information obtained from the study subjects the better. Genotype and RNA expression data can be linked to clinical images and histology that indicate stages of severity in cancer development (70,71) and diseases of the central nervous system (72). Genome-wide analyses of proteins, metabolites, and lipids may also be considered. Because of their rapid turnover, tissue protein profiles (as opposed to RNA expression) are more variable, and it is now appreciated that protein, metabolite, and lipid analyses are not as well-suited for systems analyses at the genome-wide level in individual human tissues, given the current state of the technology to measure these different dimensions. However, patterns of protein and metabolite expression in plasma may be an exception (73,74), and integrating RNA expression with protein and metabolite profiles can greatly enhance the predictive power of disease gene networks (75).

Taking all of the previously-mentioned arguments into consideration, we advocate a systems biology-type clinical study design with the basic features outlined for STAGE and STARNET, which we have termed *genome-wide network studies* (GWNS). In addition to DNA and careful clinical phenotypes, this would also include multiple intermediate phenotypes, such as RNA, proteins, metabolites, and screening of molecules that modify the structures of DNA/RNA/proteins (that is, epigenetics) in all tissues relevant for the disease in question. We believe that GWNS will help us to understand the molecular mechanisms underlying genome-wide significant loci identified by GWAS and, eventually, by WES/WGS. However, perhaps more importantly, GWNS will help us to identify the variety and full spectrum of molecular processes driving complex diseases. We think the way forward is to establish network models of these processes, both to uncover the fraction of missing heritability of complex diseases, and also to eventually establish a new paradigm of healthcare on the bases of molecular diagnostics and individually-tailored therapy.

Importantly, in applying GWNS, or any high-dimensional data analysis approach to complex traits, it is important to consider strict statistical thresholds correcting for multiple testing. For example, false discovery rate control (a form of statistical correction for multiple comparisons) should be used when detecting eQTLs (76) and when assessing gene network associations (66). Of note, as opposed to testing individual genes/SNPs in GWAS, GWNS reduce the problem with multiple testing by 1 to 3 orders of magnitude. Nonetheless, the issue of multiple testing remains relevant for GWNS.

For network associations (e.g., eigengenes) with phenotypes, the family-wise false positive rate is rigorously controlled by empirically estimating the null distribution of those associations via permutation testing and then controlling the false discovery rate by setting an appropriate p value threshold on the basis of that distribution. The discovery of eQTLs is similarly assessed by controlling for the false discovery rate. The network itself is a stable structure in that if we randomize the data so that the correlation among molecular features is destroyed, no credible network structure results (the resulting network is not scale-free, no subnetworks/modules are identified, and so on). Regarding pathway enrichment in subnetworks, their significance in the context of multiple testing is again assessed by empirically estimating the distribution of the enrichments in the context of the network topology, shuffling gene names in the network, but maintaining network topology. On this basis, p value thresholds are set to control the false discovery rate.

The relative focus of the STAGE and STARNET studies on the transcriptome (as opposed to the epigenome/proteome/metabolome) is for several reasons. Besides the limited amount of tissue sample and superior technology development for RNA screening, the transcriptome also appears sufficiently stable to capture meaningful variations relating to disease development. However, protein turnover is far more rapid, which introduces additional biological variability such that patient-to-patient comparisons of the proteome are more challenging. Nonetheless, we are strong advocates of integrating all types of “-omics” data in GWN analyses, and have previously shown that integrating proteome data with the analysis of DNA and RNA improves the predictive power of the ensuing networks (67). Therefore, although prone to greater variability, we believe that the assessment of plasma proteins remains particularly important for GWNS of CAD and other complex traits. Systemic integration of plasma proteins with established roles in CVD (Online Table 1) (77) and unbiased mass-spectrometry analysis (78) are currently being performed in plasma from the STARNET study primarily to identify novel markers for risk of clinical CAD events and, potentially, for CAD therapy (79). Similarly, we are working on computational strategies to integrate genome-wide epigenetic measures in GWN analyses (80).

Another pertinent question for GWNS is that of ethnicity. The STAGE and STARNET study participants are predominantly of northern European ancestry, which is similar to European American (EAs). But, what is the relevance of the STARNET GWNS for CAD in African Americans (AAs) and Hispanic Americans (HAs)? CAD risk factors (and thus, with a high degree of certainty, the networks/ pathways in which they operate) are generally believed to be similar across ethnicities. This would suggest that the CAD networks to be inferred from the STARNET datasets should also be relevant to CAD in AAs and HAs. The frequent, successful use of animal models, predominantly mouse, to study CAD/

atherosclerosis suggests that, at least for the early stages of CAD, many disease pathways are similar, even across species. Importantly, however, even if the main CAD risk factors are operative across most or all ethnicities, this is different from stating that every risk factor is equally important across ethnicities. In fact, the relative importance of CAD risk factors in EAs, AAs, and HAs differs. For example, insulin resistance, hypertension, and obesity are much more prevalent causes of CAD in AA than in EA (81). Thus, STARNET should improve our general understanding of causal CAD networks and their key drivers. To decipher the relative roles of these CAD networks in individual ethnicities, 1 strategy is to examine inherited risk profiles of CAD networks for AA, EA, and HA. For this, it is necessary to compute CAD network eQTLs for associations with CAD (risk) using GWA datasets specific to these ethnicities (12,82,83) (see part 2 in “The Role of GWAS in the Era of Systems Genetics”). Nonetheless, additional GWNS are certainly required on the basis of study designs similar to those of STAGE and STARNET, but on non-Caucasians, and preferably across the entire spectrum of complex disorders.

In this review, we suggest that we are on the verge of a new era of discovery of the genetics of CAD and other complex disorders, primarily on the basis of GWNS and GGES. We therefore strongly advocate that additional GWNS should be encouraged by funding bodies, as they hold great promise to decipher complex disease etiologies and represent an alternative route to extract further meaningful information from existing GWA datasets.

THE ROLE OF GWAS IN THE ERA OF SYSTEMS GENETICS

Although GWAS, followed by WES and WGS (“GWA datasets”), will remain fundamentally important in the search for the genetic causes of disease, we believe that integrating these datasets in GWNS provides a parallel approach for clinical studies that should help to define additional genetic regulators of CAD and other complex diseases. We anticipate that GWNS may uncover a significant portion of the missing heritability. We suspect that GWA datasets contain untapped information about the heritability of complex diseases and that, in the era of systems genetics, by integrating the analysis of GWNS with GWA datasets, we can prioritize risk variants that fail to reach genome-wide significance. With this perspective, we foresee that GWA datasets will be reutilized in at least the following 3 ways:

1. Reanalysis of GWA datasets based on common risk factors for CAD

Reanalyzing GWA datasets subcategorized on the basis of common risk factors for CAD is a straightforward, but remarkably seldomused, strategy to identify DNA variants for heritability of complex diseases. For example, a recent study focusing on the chromosome 9 locus for CAD (27) concluded that risk variants identified by GWAS can help explain the risk of CAD in particular subgroups of patients defined by traditional risk factors. To extend such analyses beyond risk variants identified by GWAS, we suggest that GWA datasets be reanalyzed after subjects are sorted into groups with and without given risk factors. The underlying objective is to increase the likelihood of identifying DNA variants of genome-wide significance in CAD cases with a given risk factor (e.g., diabetes, hyperlipidemia, or hypertension). Such a strategy would not merely help to reidentify established CAD risk variants discovered by traditional case-control GWA comparisons, but will also help to

identify risk variants previously found to be suggestive or to have nominally-significant associations with CAD (Figure 2). The results might point toward additional molecular mechanisms important for risk assessment and, possibly, for therapies for CAD patients with certain risk factors.

2. Reusing GWA datasets to define inherited risk-enrichment of genes believed to be involved in complex disease development

Another strategy for reusing GWA data is to consider inherited risk-enrichment analysis of groups of genes suspected to be associated with disease (84,85) (Central Illustration). These genes can either be differentially expressed between disease and control samples (22) or active in disease-related modules (7), networks, or pathways (86–88). Regardless of the gene list origin, the principal concept is that DNA variants regulating the genes of interest (eQTLs) should carry increased association for complex disease (i.e., be risk-enriched) if they are casually related to the disease. In brief, SNPs corresponding to the eQTLs, and highly-correlated SNPs in their immediate vicinity (“the experimental set”), are examined for disease association in a relevant GWA dataset. Typically, nominal significance ($p < 0.05$) is chosen as the threshold for “disease association,” but this may vary between studies. Next, control sets ($n > 5,000$) containing the same number of SNPs as the experimental set are randomly selected (located on the same chromosome and in areas with similar gene density). The fold risk-enrichment is determined by comparing the number of disease-associated SNPs in the experimental set with the average number of disease-associated SNPs in the control sets. We believe this is an especially promising technique to evaluate a set of genes for their relevance to a given disease. In contrast to gene ontology or pathway gene enrichment analysis, the method is data-driven and unbiased. In addition, a gene set regulated by risk-enriched eQTLs, according to the GWA dataset, is not merely involved in disease; the risk-enrichment also indicates that the genes are casually related to disease, as opposed to being reactively related. This strategy’s power has been demonstrated in studies of type 2 diabetes (89), CAD (59,87,88), and early-onset Alzheimer’s disease (72).

3. Reanalysis of GWA datasets by subcategories defined by the status of disease-driving networks

As we suggested earlier for established macroenvironmental risk factors, the status of disease-driving molecular processes at the microenvironmental level represented by gene networks (defined by their gene connectivity and activity) can be used to assign patients to well-defined subgroups. These subgroups can then be used to reanalyze GWA datasets to identify genome-wide significant risk variants associated with the status of these complex disease processes.

THERAPEUTIC TARGETING OF CANDIDATE GENES FOR GWA LOCI AND DRUG REPURPOSING OF KEY DRIVERS IN DISEASE-DRIVING MOLECULAR NETWORKS

Many research programs are working to understand the biological mechanisms underlying the genome-wide significant risk variants identified by GWAS (90), with the goal of

targeting these novel mechanisms therapeutically. Although we believe this approach will primarily target early CAD processes, it may also point toward pathways that can be modulated to affect the disease in its later stages. For instance, targeting PCSK9 to improve plasma cholesterol lowering may indeed help reduce risk in patients with more advanced forms of CAD (91).

Nevertheless, we believe that targeting key drivers of disease networks will be an equally, or even more, successful strategy. For this purpose, we developed an integrated informatics approach to systematically screen genome-wide transcriptional signatures of drug perturbation (treated vs. untreated) from the public domain (e.g., Connectivity Map [92]) against genome-wide transcriptional signatures of disease states (on up to networks) to identify repurposing candidates (93,94). Drug screening approaches typically require knowledge of drug target profiles or mechanisms of action, which is often lacking. In contrast, our approach requires no knowledge of the mechanism of action and can consider the system-wide properties of drug-induced molecular perturbations (e.g., genome-wide transcriptional changes) to enable discovery of novel connections between drugs and disease states through a data-driven approach (94). This approach also allows for rational repurposing of multitarget compounds (poly-pharmacology) that might exhibit therapeutic effects against a complex disease like CAD through modulation of multiple network driver nodes (90).

The broad biological relevance of this approach is supported by experimental validation of several novel drug indications identified by a computational drug-repurposing pipeline. This system was recently used to transcriptionally profile intestinal samples obtained from inflammatory bowel disease (IBD) patients to experimentally validate a novel IBD indication predicted for the anticonvulsant agent, topiramate (94). Topiramate has no history of efficacious use for IBD or other inflammatory diseases. There are no established therapeutic targets for IBD or other inflammatory diseases among the canonical targets of topiramate, which enhances GABA-A receptor activity, antagonizes AMPA/kinate glutamate receptor subtypes, and weakly inhibits carbonic anhydrase isozymes II and IV (93). In a rodent model of IBD induced with 2,4,6-trinitrobenzenesulfonic acid, topiramate significantly reduced the severity of IBD, as judged by gross pathophysiological and histopathological measures (94).

In a separate study (95), this computational drug-repurposing approach was applied to transcriptional profiles of tumor versus adjacent normal tissue to identify novel drug-repurposing candidates for small-cell lung cancer (SCLC). Several pharmacologically-diverse compounds were identified as novel drug-repurposing candidates for SCLC, including: the tricyclic antidepressant, imipramine; the calcium-channel blocker, bepridil; and the phenothiazine antihistamine, promethazine. Anti-SCLC or other antineoplastic effects were not previously established for these drugs or for other drugs of the same pharmacological class. The anti-SCLC activity of these compounds was validated in experiments in human and animal model systems in vitro and in vivo (95).

In summary, given the shortage of new drugs reaching the market for CAD and many other complex diseases, drug repurposing, using a systems genetics approach to define new agents

and novel indications for existing therapies, will be an essential path toward personalized and preventive drug therapies.

SUMMARY AND FUTURE DIRECTIONS

To answer our own question posed in the title of this paper, there is no doubt that GWAS are important in providing datasets to reveal risk variants that explain the heritability of complex diseases. In particular, genome-wide significant loci point to potentially important genes and molecular mechanisms responsible for the pathobiology of these diseases. However, we believe that traditional analyses of GWA datasets overlook the potential role of context-dependent risk variants that exert risk only when certain environmental influences are operative. These influences typically arise over a shorter period and, increasingly, at later stages of disease development. We believe these context-dependent risk variants can have a large effect on key disease processes active during limited windows of complex disease development. As such, they are unlikely to emerge in traditional GWA dataset analyses as of genome-wide significance, but are likely to be detected as suggestive or nominally-significant risk variants. If this is the case, much information on the heritability of complex disease remains hidden in GWA datasets. We believe that, by applying GWNS to identify the disease-driving molecular processes reflected in molecular networks and their genetic regulators, this information can be revealed. We propose that these disease-driving gene networks can distinguish true-from false-positive risk variants with nominal disease associations in GWAS. As more network biology underlying complex disorders is revealed, the relevant activity and type of disease network can be used as a diagnostic marker to subcategorize patients into groups requiring different therapeutic measures. To reach this goal, we must enter a post-GWAS era, in which priority is given to clinical studies that include intermediate phenotypes (with the general design described earlier for GWNS, but also considering other genome-wide measures besides RNA) and where the screening of patients ranges from gross disease phenotypes (i.e., CAD), to histological and image-based patient characteristics, and ultimately to clinical outcomes. Such studies will be essential for solving the puzzle of interconnected molecules (e.g., genes) in disease-driving networks, perhaps allowing fulfillment of the long-sought goal of the “genomic revolution”: the preventive and individual care of patients in molecularly-defined subcategories.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Drs. Björkegren, Kovacic, and Schadt are supported by the American Heart Association (14SFRN20490315; 14SFRN20840000) and are members of the Consortium “CAD Genomics” (Drs. Björkegren and Schadt) and the Consortium “Cellular and Molecular Targets to Promote Therapeutic Cardiac Regeneration” (Dr. Kovacic), both of which are funded by the Leducq Foundation (Transatlantic Network of Excellence Awards). Dr. Björkegren is also supported by the Swedish Heart-Lung Foundation, the Swedish Research Council, the University of Tartu (SP1GVARENG), the Estonian Research Council, and by a grant from Astra-Zeneca Translational Science Centre-Karolinska Institutet (joint research program in translational science); and is the founder, a main shareholder, and chairman of the board of Clinical Gene Networks AB (CGN), which has invested interests in the STARNET and STAGE cohorts. Dr. Kovacic is also supported by the National Institutes of Health (K08HL111330) and by a research grant from AstraZeneca. Dr. Dudley is supported in part by funding from the National Institutes of Health

(R01 DK098242 and U54 CA189201) and by the PhRMA Foundation. Dr. Schadt is a CGN board member and shareholder. Robert Roberts, MD, served as Guest Editor for this paper.

ABBR EVIATIONS AND ACRONYMS

CAD	coronary artery disease
eQTL	expression quantitative trait locus
GGES	genetics of gene expression studies
GWA	genome-wide association
GWAS	genome-wide association studies
GWNS	genome-wide network studies
LDL	low-density lipoprotein
SNP	single-nucleotide polymorphism
WES/WGS	whole-exome/whole-genome sequencing

REFERENCES

1. Hughes V. Epigenetics: the sins of the father. *Nature*. 2014; 507:22–24. [PubMed: 24598623]
2. Cortijo S, Wardenaar R, Colome-Tatche M, et al. Mapping the epigenetic basis of complex traits. *Science*. 2014; 343:1145–1148. [PubMed: 24505129]
3. Holliday R. DNA methylation and epigenetic mechanisms. *Cell Biophys*. 1989; 15:15–20. [PubMed: 2476223]
4. Felsenfeld G. A brief history of epigenetics. *Cold Spring Harb Perspect Biol*. 2014; 6 a018200.
5. Friend SH, Schadt EE. Translational genomics. Clues from the resilient. *Science*. 2014; 344:970–972. [PubMed: 24876479]
6. Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet*. 2011; 129:351–370. [PubMed: 21331778]
7. Hägg S, Skogsberg J, Lundström J, et al. Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) study. *PLoS Genet*. 2009; 5:e1000754. [PubMed: 19997623]
8. Roberts R. Genetics of coronary artery disease. *Circ Res*. 2014; 114:1890–1903. [PubMed: 24902973]
9. Samani NJ, Erdmann J, Hall AS, et al. WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *New Engl J Med*. 2007; 357:443–453. [PubMed: 17634449]
10. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007; 316:1491–1493. [PubMed: 17478679]
11. McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007; 316:1488–1491. [PubMed: 17478681]
12. Deloukas P, Kanoni S, Willenborg C, et al. CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013; 45:25–33. [PubMed: 23202125]
13. Hindorf, LA.; MacArthur, J.; Morales, J., et al. [Accessed December 21, 2014] A catalog of published genome-wide association studies. Available at: <http://www.genome.gov/gwastudies>
14. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009; 461:747–753. [PubMed: 19812666]

15. Marenberg ME, Risch N, Berkman LF, et al. Genetic susceptibility to death from coronary heart disease in a study of twins. *New Engl J Med*. 1994; 330:1041–1046. [PubMed: 8127331]
16. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010; 42:565–569. [PubMed: 20562875]
17. Schadt EE, Björkegren JL. NEW: network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med*. 2012; 4 115rv1.
18. Joyner MJ, Prendergast FG. Chasing Mendel: five questions for personalized medicine. *J Physiol*. 2014; 592:2381–2388. [PubMed: 24882820]
19. Allaby, M. *A Dictionary of Zoology*. 2nd edition. New York, NY: Oxford University Press; 1999. p. 597vi
20. Narula J, Kovacic JC. Putting TCFA in clinical perspective. *J Am Coll Cardiol*. 2014; 64:681–683. [PubMed: 25125299]
21. Skogsberg J, Lundstrom J, Kovacs A, et al. Transcriptional profiling uncovers a network of cholesterol-responsive atherosclerosis target genes. *PLoS Genet*. 2008; 4:e1000036. [PubMed: 18369455]
22. Björkegren JL, Hagg S, Talukdar HA, et al. Plasma cholesterol-induced lesion networks activated before regression of early, mature, and advanced atherosclerosis. *PLoS Genet*. 2014; 10:e1004201. [PubMed: 24586211]
23. Hagg S, Salehpour M, Noori P, et al. Carotid plaque age is a feature of plaque stability inversely related to levels of plasma insulin. *PloS One*. 2011; 6:e18248. [PubMed: 21490968]
24. Lusis AJ. Atherosclerosis. *Nature*. 2000; 407:233–241. [PubMed: 11001066]
25. Kubo T, Maehara A, Mintz GS, et al. The dynamic nature of coronary artery lesion morphology assessed by serial virtual histology intravascular ultrasound tissue characterization. *J Am Coll Cardiol*. 2010; 55:1590–1597. [PubMed: 20378076]
26. Bentzon JF, Otsuka F, Virmani R, et al. Mechanisms of plaque formation and rupture. *Circ Res*. 2014; 114:1852–1866. [PubMed: 24902970]
27. Lusk CM, Dyson G, Clark AG, et al. Validated context-dependent associations of coronary heart disease risk with genotype variation in the chromosome 9p21 region: the Atherosclerosis Risk in Communities study. *Hum Genet*. 2014; 133:1105–1116. [PubMed: 24889828]
28. Peters MN, Moscona JC, Katz MJ, et al. Natural disasters and myocardial infarction: the six years after Hurricane Katrina. *Mayo Clin Proc*. 2014; 89:472–477. [PubMed: 24656058]
29. Chan C, Elliott J, Troughton R, et al. Acute myocardial infarction and stress cardiomyopathy following the Christchurch earthquakes. *PloS One*. 2013; 8:e68504. [PubMed: 23844213]
30. Smith EN, Kruglyak L. Gene-environment interaction in yeast gene expression. *PLoS Biol*. 2008; 6:e83. [PubMed: 18416601]
31. Smirnov DA, Morley M, Shin E, et al. Genetic analysis of radiation-induced changes in human gene expression. *Nature*. 2009; 459:587–591. [PubMed: 19349959]
32. Romanoski CE, Lee S, Kim MJ, et al. Systems genetics analysis of gene-by-environment interactions in human cells. *Am J Hum Genet*. 2010; 86:399–410. [PubMed: 20170901]
33. Orozco LD, Bennett BJ, Farber CR, et al. Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell*. 2012; 151:658–670. [PubMed: 23101632]
34. James PA, Oparil S, Carter BL, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014; 311:507–520. [PubMed: 24352797]
35. Ogihara T, Saruta T, Rakugi H, et al. Valsartan in Elderly Isolated Systolic Hypertension Study Group Target blood pressure for treatment of isolated systolic hypertension in the elderly: valsartan in elderly isolated systolic hypertension study. *Hypertension*. 2010; 56:196–202. [PubMed: 20530299]
36. JATOS Study Group. Principal results of the Japanese trial to assess optimal systolic blood pressure in elderly hypertensive patients (JATOS). *Hypertension Res*. 2008; 31:2115–2127.
37. 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; 377:383–392. [PubMed: 21239051]
38. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *New Engl J Med*. 2005; 352:1685–1695. [PubMed: 15843671]
 39. Libby P, Ridker PM, Hansson GK. Leducq Transatlantic Network on Atherothrombosis. Inflammation in atherosclerosis: from pathophysiology to practice. *J Am Coll Cardiol*. 2009; 54:2129–2138. [PubMed: 19942084]
 40. Kidd BA, Peters LA, Schadt EE, et al. Unifying immunology with informatics and multiscale biology. *Nat Immunol*. 2014; 15:118–127. [PubMed: 24448569]
 41. Lusis AJ, Weiss JN. Cardiovascular networks: systems-based approaches to cardiovascular disease. *Circulation*. 2010; 121:157–170. [PubMed: 20048233]
 42. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*. 2014; 15:34–48. [PubMed: 24296534]
 43. Weiss JN, Karma A, MacLellan WR, et al. “Good enough solutions” and the genetics of complex diseases. *Circ Res*. 2012; 111:493–504. [PubMed: 22859671]
 44. Barabasi AL, Oltvai ZN. Network biology: understanding the cell’s functional organization. *Nat Rev Genet*. 2004; 5:101–113. [PubMed: 14735121]
 45. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature*. 2009; 461:218–223. [PubMed: 19741703]
 46. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011; 12:56–68. [PubMed: 21164525]
 47. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005; 4 Article 17.
 48. Chen Y, Zhu J, Lum PY, et al. Variations in DNA elucidate molecular networks that cause disease. *Nature*. 2008; 452:429–435. [PubMed: 18344982]
 49. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. *Nature*. 2008; 452:423–428. [PubMed: 18344981]
 50. Wang IM, Zhang B, Yang X, et al. Systems analysis of eleven rodent disease models reveals an inflammatoe signature and key drivers. *Mol Syst Biol*. 2012; 8:594. [PubMed: 22806142]
 51. Millstein J, Zhang B, Zhu J, et al. Disentangling molecular relationships with a causal inference test. *BMC Genet*. 2009; 10:23. [PubMed: 19473544]
 52. Schadt EE, Lamb J, Yang X, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet*. 2005; 37:710–717. [PubMed: 15965475]
 53. Liu Y, Qiao N, Zhu S, et al. A novel Bayesian network inference algorithm for integrative analysis of heterogeneous deep sequencing data. *Cell Res*. 2013; 23:440–443. [PubMed: 23318583]
 54. Hubner N, Wallace CA, Zimdahl H, et al. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*. 2005; 37:243–253. [PubMed: 15711544]
 55. Cookson W, Liang L, Abecasis G, et al. Mapping complex disease traits with global gene expression. *Nat Rev Genet*. 2009; 10:184–194. [PubMed: 19223927]
 56. Nicolae DL, Gamazon E, Zhang W, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010; 6:e1000888. [PubMed: 20369019]
 57. Kang HP, Morgan AA, Chen R, et al. Coanalysis of GWAS with eQTLs reveals disease-tissue associations. *AMIA Jt Summits Transl Sci Proc*. 2012:35–41. [PubMed: 22779046]
 58. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013; 45:1238–1243. [PubMed: 24013639]
 59. Foroughi Asl H, Talukdar AH, Kindt AS, et al. Expression quantitative trait loci acting across multiple tissues are enriched in inherited risk for coronary artery disease. *Circ Cardiovasc Genet*. 2014 In press.
 60. Kathiresan S, Voight BF, Purcell S, et al. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009; 41:334–341. [PubMed: 19198609]

61. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
62. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
63. Dobrin R, Zhu J, Molony C, et al. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol*. 2009; 10:R55. [PubMed: 19463160]
64. Zhu J, Lum PY, Lamb J, et al. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res*. 2004; 105:363–374. [PubMed: 15237224]
65. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol*. 2008; 6:e107. [PubMed: 18462017]
66. Langfelder P, Horvath S. Eigengene networks for studying the relationships between coexpression modules. *BMC Syst Biol*. 2007; 1:54. [PubMed: 18031580]
67. Zhu J, Zhang B, Smith EN, et al. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet*. 2008; 40:854–861. [PubMed: 18552845]
68. Yang X, Zhang B, Molony C, et al. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res*. 2010; 20:1020–1036. [PubMed: 20538623]
69. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013; 45:580–585. [PubMed: 23715323]
70. Lin B, White JT, Lu W, et al. Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. *Cancer Res*. 2005; 65:3081–3091. [PubMed: 15833837]
71. Nibbe RK, Koyuturk M, Chance MR. An integrative -omics approach to identify functional sub-networks in human colorectal cancer. *PLoS Comput Biol*. 2010; 6:e1000639. [PubMed: 20090827]
72. Zhang B, Gaiteri C, Bodea LG, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013; 153:707–720. [PubMed: 23622250]
73. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014; 46:543–550. [PubMed: 24816252]
74. Kim J, Ghasemzadeh N, Eapen DJ, et al. Gene expression profiles associated with acute myocardial infarction and risk of cardiovascular death. *Genome Medicine*. 2014; 6:40. [PubMed: 24971157]
75. Zhu J, Sova P, Xu Q, et al. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol*. 2012; 10:e1001301. [PubMed: 22509135]
76. Qi J, Asl HF, Björkegren J, et al. kruX: matrix-based non-parametric eQTL discovery. *BMC Bioinformatics*. 2014; 15:11.
77. Huynh JL, Garg P, Thin TH, et al. Epigenome-wide differences in pathology-free regions of multiple sclerosis-affected brains. *Nat Neurosci*. 2014; 17:121–130. [PubMed: 24270187]
78. Gu GJ, Friedman M, Jost C, et al. Protein tag-mediated conjugation of oligonucleotides to recombinant affinity binders for proximity ligation. *New Biotechnol*. 2013; 30:144–152.
79. Pan S, Aebersold R, Chen R, et al. Mass spectrometry based targeted protein quantification: methods and applications. *J Proteome Res*. 2009; 8:787–797. [PubMed: 19105742]
80. Chan JC, Piper DE, Cao Q, et al. A proprotein convertase subtilisin/kexin type 9 neutralizing antibody reduces serum cholesterol in mice and nonhuman primates. *Proc Natl Acad Sci U S A*. 2009; 106:9820–9825. [PubMed: 19443683]
81. Clark LT, El-Atat F. Metabolic syndrome in African Americans: implications for preventing coronary heart disease. *Clin Cardiol*. 2007; 30:161–164. [PubMed: 17443653]
82. Lettre G, Palmer CD, Young T, et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARE Project. *PLoS Genet*. 2011; 7:e1001300. [PubMed: 21347282]

83. Qi L, Campos H. Genetic predictors for cardiovascular disease in Hispanics. *Trends Cardiovasc Med*. 2011; 21:15–20.
84. Zhong H, Beaulaurier J, Lum PY, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet*. 2010; 6:e1000932. [PubMed: 20463879]
85. Zhong H, Yang X, Kaplan LM, et al. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet*. 2010; 86:581–591. [PubMed: 20346437]
86. Hsu YH, Zillikens MC, Wilson SG, et al. An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility loci for osteoporosis-related traits. *PLoS Genet*. 2010; 6:e1000977. [PubMed: 20548944]
87. Makinen VP, Civelek M, Meng Q, et al. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet*. 2014; 10:e1004502. [PubMed: 25033284]
88. Shang MM, Talukdar HA, Hofmann JJ, et al. Lim domain binding 2: a key driver of trans-endothelial migration of leukocytes and atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2014; 34:2068–2077. [PubMed: 24925974]
89. Yang X, Peterson L, Thieringer R, et al. Identification and validation of genes affecting aortic lesions in mice. *J Clinical Invest*. 120:2414–2422. [PubMed: 20577049]
90. Keiser MJ, Setola V, Irwin JJ, et al. Predicting new molecular targets for known drugs. *Nature*. 2009; 462:175–181. [PubMed: 19881490]
91. Rashid S, Curtis DE, Garuti R, et al. Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9. *Proc Natl Acad Sci U S A*. 2005; 102:5374–5379. [PubMed: 15805190]
92. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006; 313:1929–1935. [PubMed: 17008526]
93. Knox C, Law V, Jewison T, et al. DrugBank 3.0: a comprehensive resource for ‘omics’ research on drugs. *Nucleic Acids Res*. 2011; 39:D1035–D1041. [PubMed: 21059682]
94. Dudley JT, Sirota M, Shenoy M, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci Transl Med*. 2011; 3:96ra76.
95. Jahchan NS, Dudley JT, Mazur PK, et al. A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer Disc*. 2013; 3:1364–1377.

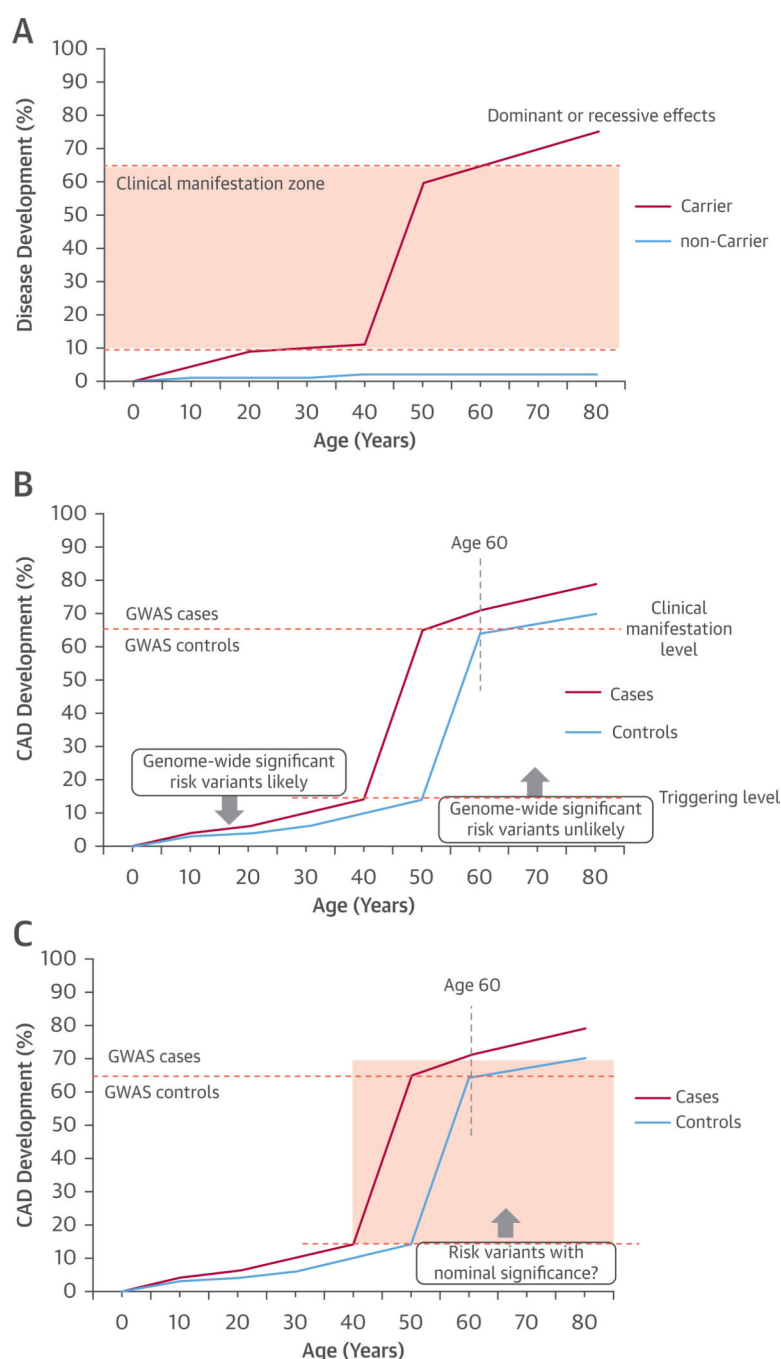


FIGURE 1. Development of Rare Versus Complex Disease, Using CAD as the Example
(A) Rare disease does not develop in noncarriers of deoxyribonucleic acid (DNA) risk variants. The level of clinical manifestation varies with different rare diseases. Of note, the indicated sigmoidal curve for rare disease development merely reflects that most biological events follow this growth pattern; it does not indicate that this is true for all rare diseases. **(B)** The development of complex diseases, such as coronary artery disease (CAD), generally follows a sigmoidal curve, characterized by a slow initial growth phase (30 to 50 years), reaching a triggering level to enter a subsequent rapid expansion phase (10 years), and

finally, a slow-down phase in which clinical symptoms manifest. Subjects in genome-wide association studies (GWAS) are more likely to be defined as cases or controls predominantly on the basis of early disease processes, rather than due to later, largely overlapping processes. As a result, the likelihood of finding risk variants with genome-wide significance is higher for disease processes active in early CAD development. (C) Given the large overlap between cases and controls, increasing dependence of environmental contexts, and the shorter duration of late disease processes, DNA variants regulating these processes are likely associated with disease at nominally-significant p values in traditional analyses of genome-wide association datasets.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

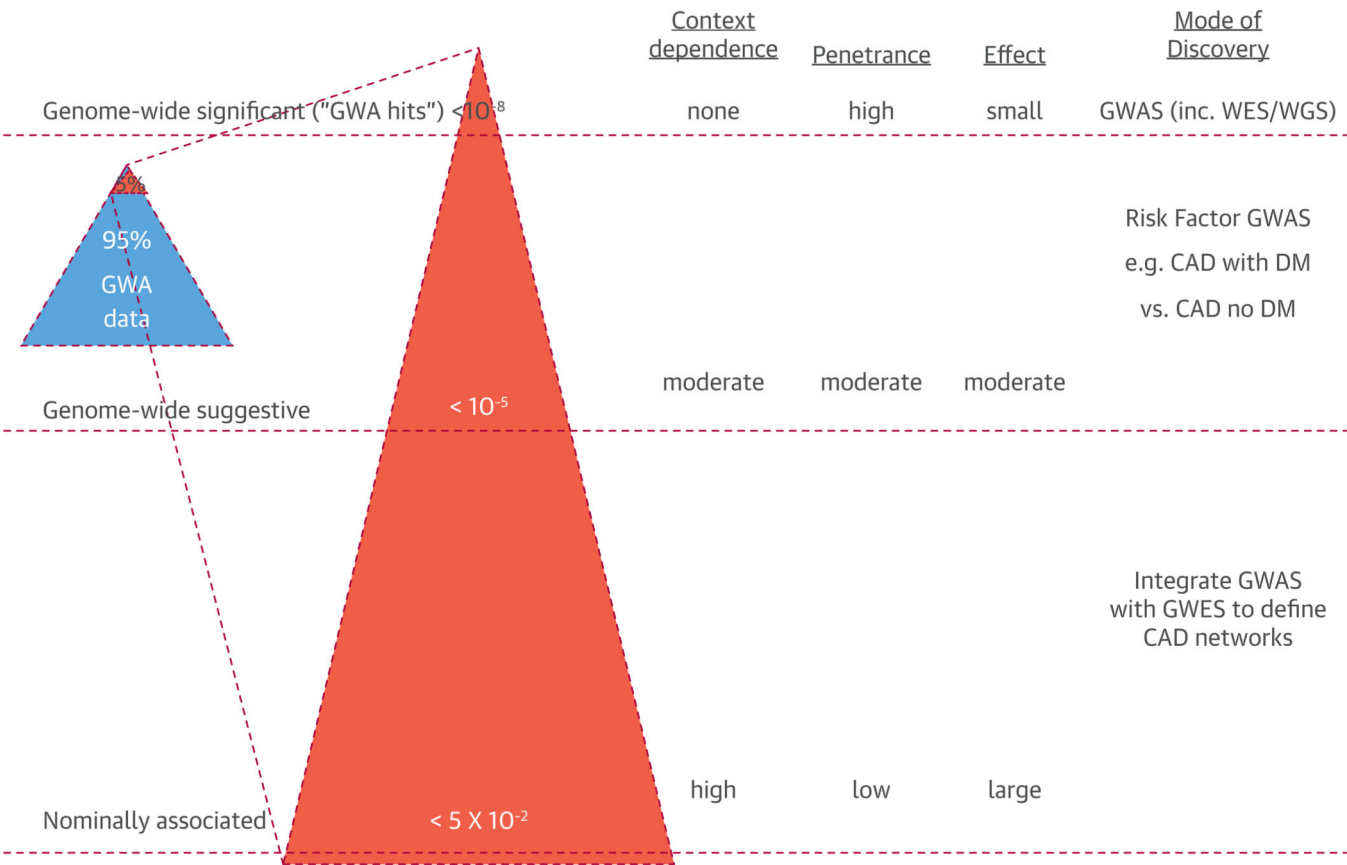


FIGURE 2. Risk Variants Identified in GWAS
(A) The **gray triangle** represents all DNA variants tested for CAD association in GWAS. The 5% (**red triangle**) have nominal significant association ($p < 0.05$). The very **top of the central red triangle** consists of risk variants associated with CAD at a level of genome-wide significance ($p < 10^{-8}$). The **middle** section consists of risk variants associated with CAD at a level suggestive of genome-wide significance ($p < 10^{-5}$). The **base of the triangle** consists of risk variants associated with CAD at a level of nominal significance ($p < 0.05$). On the **right**, the likely influences of different significance levels on context dependency, penetrance, and disease are indicated. On the **far right** is the likely mode of discovery for testing genome-wide significance ($p < 10^{-8}$). DM = diabetes mellitus; GWA = genome-wide association; GWES = genome-wide network studies; other abbreviations as in Figure 1.

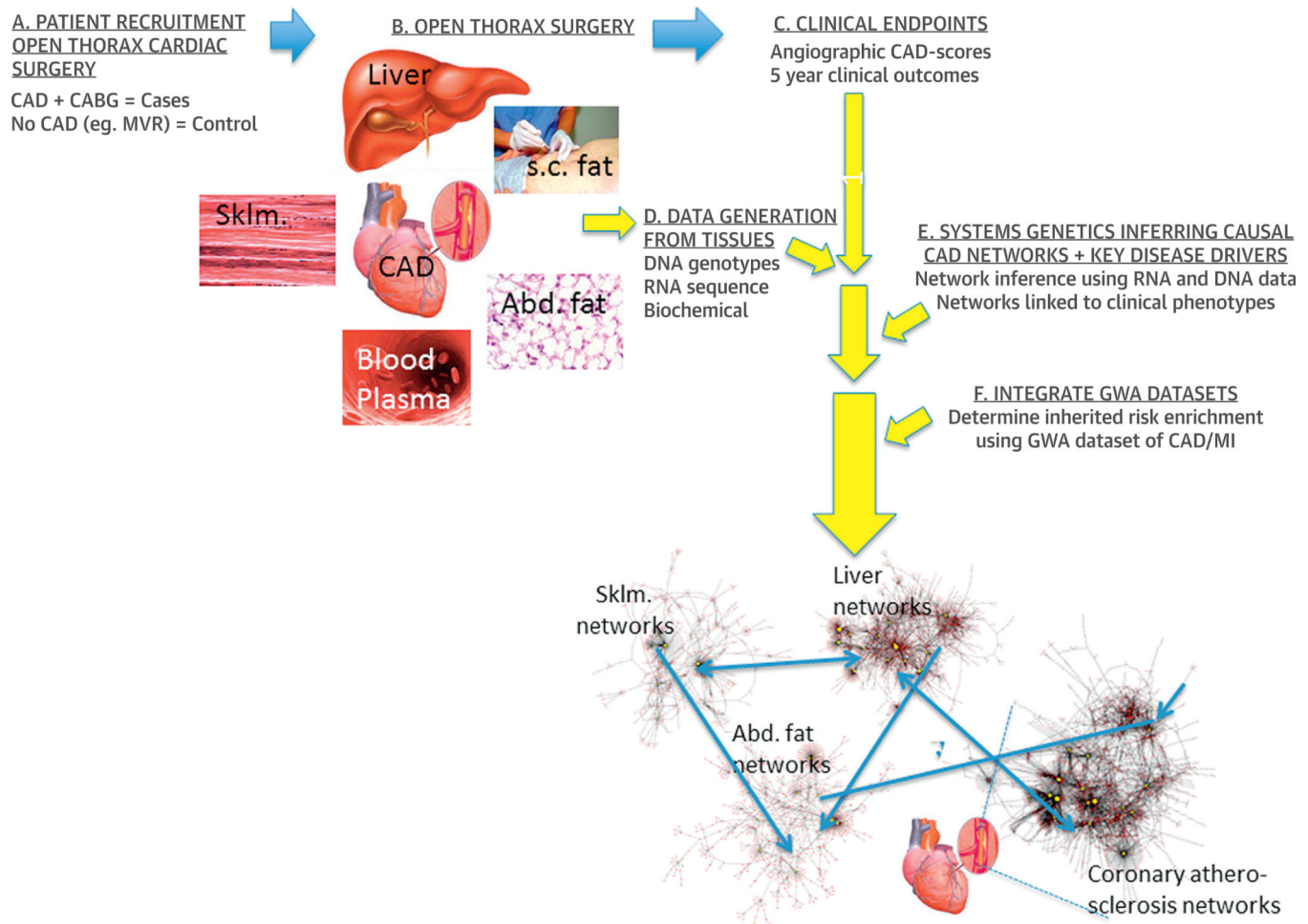
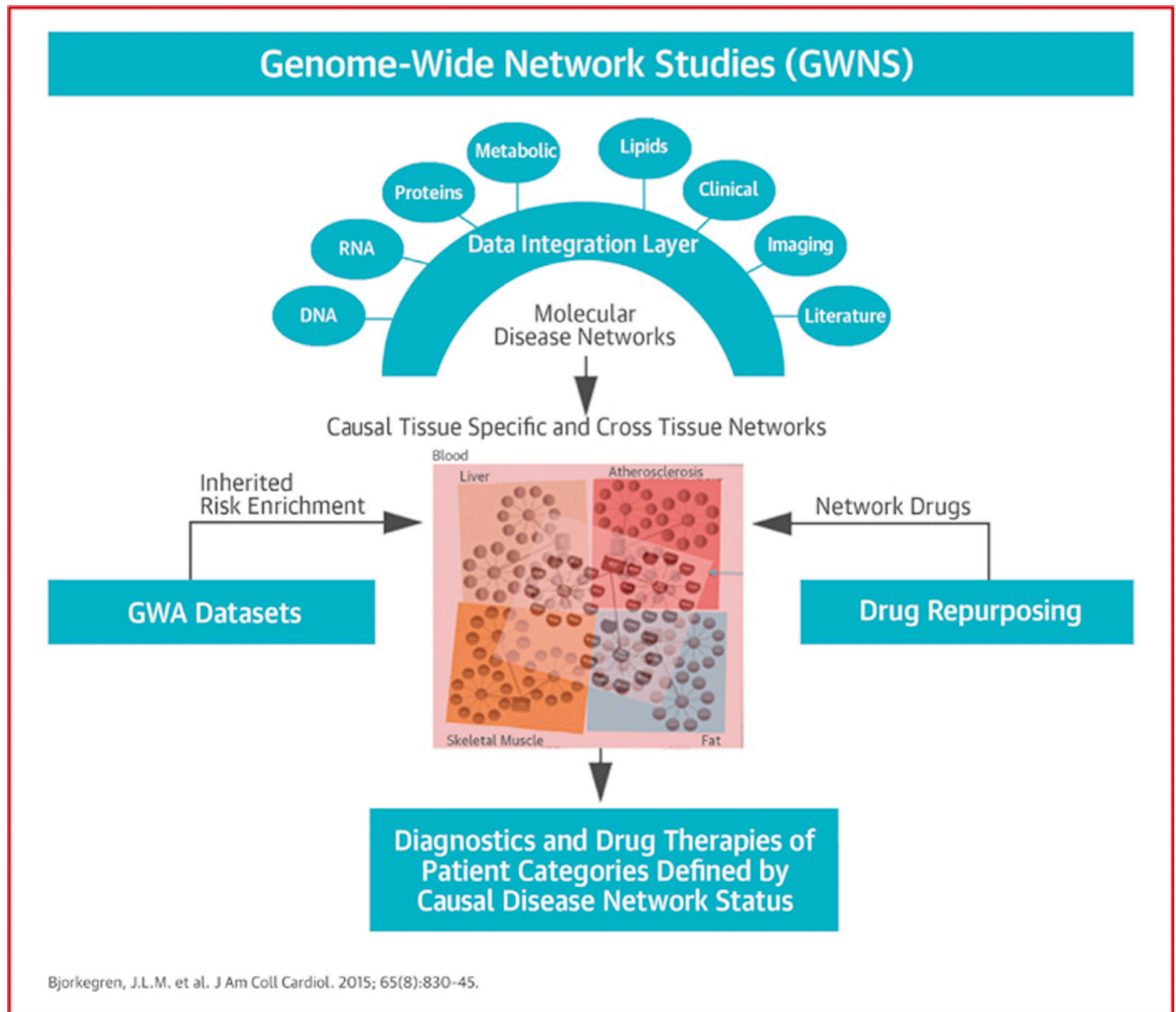


FIGURE 3. Genetics of CAD Gene Expression Studies

The STAGE (Stockholm Atherosclerosis Gene Expression) and STARNET (Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task) genetics of gene expression studies (GGES). **(A)** Patient recruitment. Patients were included if eligible for coronary artery bypass grafting (CABG) or other non-CAD related indications for open thorax surgery and had no other severe systemic diseases (e.g., widespread cancer or active systemic inflammatory disease). The ethical committees of the Karolinska and Tartu University Hospitals approved the studies, and patients gave written consent (Dnr 004-02 and 2007/1521-32). **(B)** Open thorax surgery. Between 7 and 9 vascular and metabolic tissues of 124 (STAGE) and 600 (STARNET) CAD patients and 100 non-CAD control subjects (STARNET), all clinically well-characterized, were sampled during open thorax surgery as described (7). Ribonucleic acid (RNA) samples were isolated from the atherosclerotic arterial wall, internal mammary artery, liver, skeletal muscle (Sklm.), subcutaneous (s.c.) and visceral fat, whole blood, and primary blood monocytes differentiated into macrophages and foam cells in vitro. **(C)** Clinical endpoints. In each patient, a total of 114 clinical characteristics were screened, including the SYNTAX score (from pre-operative angiograms), which shows the clinical degree of coronary atherosclerosis. Currently, the 5-year follow-up of mortality (from CAD and other reasons)

and CAD-related morbidity is being conducted. **(D)** Data generation. GenomeWideSNP_6 arrays (Affymetrix) were used for genotyping DNA, and Custom-made HuRSTA-2a520709 arrays (Affymetrix) were used for gene expression profiling (STAGE) and RNA sequencing (Illumina2500, STARNET). **(E)** Systems genetic analysis. Gene expression and RNA sequence data are used to define groups of genes acting together in modules and networks based on coexpression similarities (17,52,64,65). An eigengene value of each module is calculated and correlated with the phenotypic characteristics of the patients. For modules with strong phenotypic associations, Bayesian network algorithms are applied to infer key drivers genes (50,67,68) (**yellow nodes**) believed to serve as diagnostic markers and therapeutic targets. **(F)** Integrate GWA datasets. GWA datasets for CAD are reanalyzed to assess risk enrichment of identified regulatory gene networks. Abd. = abdominal; MI = myocardial infarction; MVR = mitral valve repair; other abbreviations as in Figures 1 and 2.



CENTRAL ILLUSTRATION. Systems Genetics to Understand Coronary Artery Disease: Genetics of Genome-Wide Expression Studies

A wide range of genomic (deoxyribonucleic acid [DNA], ribonucleic acid [RNA], protein, and metabolite), clinical, imaging, and prior knowledge data (**circles above the half-circle**) are integrated to infer causal gene networks reflecting central disease processes in complex diseases. **(Left)** Genome-wide (GW) association datasets are integrated in the analysis of these networks to determine their load of inherited risk, which also indicates their level of causality. **(Right)** Key drivers in disease-causing networks are analyzed against transcriptomic data from studies of numerous approved and experimental drugs to identify drugs that either increase or decrease the activity of genes in the network (i.e., drug repurposing; heat map indicates gene activity in a network exposed to drugs). Identified drugs can be tested in models of complex disease (e.g., mouse models). The final goal is to enable diagnosis and treatment of patients on the basis of their molecular disease status,

reflected by the activity status of disease networks. The **center parallelogram** shows disease-causing gene networks acting within and across tissues (**red** indicates atherosclerosis; **brown** indicates liver; **orange** indicates skeletal muscle; and **blue** indicates fat) that are integrated through communication (e.g., via blood). The central network indicates the presence of cross-tissue networks.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript