Published in final edited form as:

Curr Atheroscler Rep. 2014 May ; 16(5): 406. doi:10.1007/s11883-014-0406-4.

Dissecting the causal genetic mechanisms of coronary heart disease

Clint L. Miller¹, Themistocles L. Assimes¹, Stephen B. Montgomery², and Thomas Quertermous¹

¹Department of Medicine, Division of Cardiovascular Medicine, and Cardiovascular Institute, Stanford, CA, 94305, USA

²Departments of Pathology and Genetics, Stanford University School of Medicine, Stanford, CA, 94305, USA

Abstract

Large-scale genome wide association studies (GWAS) have identified 46 loci that are associated with coronary heart disease (CHD). Additionally, 104 independent candidate variants (false discovery rate of 5%) have been identified.¹⁻³ A majority of the causal genes in these loci function independently of conventional risk factors, and it is postulated that a number of the CHD associated genes regulate basic processes in the vascular cells involved in atherosclerosis, and that study of the signaling pathways that are modulated in this cell type by causal regulatory variation will provide critical new insights for targeting the initiation and progression of disease. In this review we will discuss the types of experimental approaches and data that are critical for understanding of molecular processes that underlie the disease risk at 9p21.3, *TCF21*, *SORT1*, and other CHD associated loci.

Mechanisms of genome-wide association and disease risk

The majority of non-Mendelian disease-associated variation resides outside structural genes and is presumed to regulate gene expression rather than encoded protein structure.⁴⁻⁷ Thus far, single nucleotide polymorphisms (SNPs) appear to be the type of variation that contribute the majority of known genetic risk, and relatively few studies have identified the causal variants and mechanisms for the expanding number of disease associations. In these cases, the causal SNP has been shown to alter: *i*) transcription factor (TF) binding and causal gene transcription⁸⁻¹⁷, *ii*) post-transcriptional gene expression as mediated by regulatory miRNA^{18, 19}, *iii*) genomic methylation patterns that regulate gene expression^{20, 21}, and *iv*) expression of long non-coding RNAs that likely regulate expression of the causal gene.²²

C. Miller PhD, T. Assimes MD, PhD, T. Quertermous, MD, Stanford University, Cardiovascular Medicine, 300 Pasteur Drive, Stanford, CA 94305-5233, Phone: (650) 723-5013, tomq1@stanford.edu.

S.B. Montgomery PhD, Asst. Professor of Pathology and Genetics, Stanford University, 300 Pasteur Drive, Edwards Building R248A, Stanford, CA 94305, Phone: 650 725 9641, Stephen Montgomery <smontgom@stanford.edu>

Conflict of Interest

Clint L. Miller, Themistocles L. Assimes, Stephen B. Montgomery, and Thomas Quertermous declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent

This article does not contain any studies with human or animal subjects performed by any of the authors.

Importantly, each of the first three mechanisms are expected to result in allele-specific expression of the causal gene, and lncRNAs may also if its effect is in *cis*.²³

Studies to date suggest that disruption of transcription factor and miRNA binding sites are the most common functional mechanisms for causal variation mediated regulation of gene expression.²⁴ To determine the relative contribution of each of these mechanisms to CHD, we estimated the number of GWAS CHD-associated and correlated SNPs in both ENCODE chromatin immunoprecipitation sequencing (ChIP-Seq) TF binding regions (~4,400,000) and miRcode predicted miRNA binding sites (~1,100,000). We identified 3520 total variants, including CHD associated variants and those in high LD ($r^2 > 0.8$). Of these, 676 (19.2%) were found to reside in a TF binding region (of which 62 reside directly in a highly conserved TFBS), and 36 (1%) found to reside in a miRNA binding site (either moderately or highly conserved). While such analyses have limitations, they strongly suggest that disruption of TF binding is much more common than disruption of miRNA binding. Thus, studies concentrating on transcriptional regulatory mechanisms that likely contribute the majority of disease risk would be expected to provide more immediate insights into upstream epigenetic pathways.

Identification of causal variation

Extending human genetic association data

Given whole genome association data there are a number of computational and follow-up genetic analyses that will provide critical insights into which of a number of associated SNPs represent the causal variation. First, it is possible to use available algorithms to increase the number of SNPs for which association data can be generated. In this process termed “imputation” available genotyping data is extended to non-genotyped SNPs through computational analysis from the genotypes at SNPs for which assays were available, and can provide coverage up to, with imputation (filling in missing SNP data) revealing up to several million SNPs in the genome.²⁵ Such studies can point to a causal SNP that was not genotyped on the original array or point to additional associated SNPs that must be considered in the list of possible causal variants. Other purely computational approaches include conditional analyses that allow the association signal at individual associated SNPs to be removed and the relative signal at other SNPs to be quantified.²

Follow-up genetic studies include resequencing and fine mapping with targeted genotyping assays in specific loci in a number of cases to identify common and rare variation that might drive the association. For example, a meta-analysis was conducted by the CARDIoGRAM +C4D consortium with GWAS fine-mapping chip (Metabochip) data that contained dense SNP sets in loci associated with cardiovascular, metabolic and other traits, and these studies identified a number of alleles that were not identified in the initial meta-analysis.^{1, 2} Independent novel alleles were identified in the 9p21.3, *COL4A2*, *CXCL12* and *TCF21* loci with this approach. Also, follow-up association studies in racial-ethnic groups other than those used in the initial analysis may more directly point to causal variation because of local differences in linkage disequilibrium.²⁶ Studies in African American cohorts have provided evidence of a second allele in the 9p21.3 locus²⁷, and studies in Han Chinese have identified a second allele in the *TCF21* locus.²⁸ These findings have significantly added to the

certainty of the association in Caucasians and provided additional avenues of study investigating the mechanisms of association.

Use of genomic data characterizing chromosomal structure

As noted above, there is mounting evidence that altered transcription factor binding is a common mechanism by which causal allelic variation mediates CHD risk. Through ongoing efforts by the ENCODE consortium and the NIH Roadmap Reference Epigenome consortia, genome-wide chromatin organization has been assessed in distinct cultured cells using DNase-Seq and FAIRE-Seq approaches. Additionally, genome-wide post-translational histone modifications that reflect transcriptional activity or repression have been mapped for a number of cultured cell types. A new method to map chromatin accessibility using transposon integration and high-throughput sequencing (ATAC-seq) was recently developed by Drs. Greenleaf and Chang at Stanford, which promises to provide higher resolution with fewer starting cells.²⁹ The value of these data is in highlighting regions of the genome that are likely to mediate transcription, and serve as sites of functional regulatory variation associated with CHD. Several groups have developed statistical and analytical methods to incorporate such data into an analytical paradigm aimed at identifying causal variants.³⁰ Our group has mapped chromatin organization in human coronary artery smooth muscle cells (HCASMC) and used these data to help identify causal variation for two of the disease-associated alleles at *TCF21*.³¹ Finally, it is important to note that scanning genomic sequences for altered miRNA binding sites, often by 3'-UTR variants, may represent another level of gene regulation associated with disease risk, as demonstrated for many cancers.^{32, 33} We have recently identified *TCF21* as the first CHD associated gene to be regulated by a miRNA that is disrupted by disease associated variation.³⁴

Genetics of gene expression

Considerable effort has been focused on defining the links between common polymorphisms in the genome and target gene expression levels, through which causal variation can be identified. By combining whole genome expression profiling by either microarray or RNAseq studies with whole genome genotyping it has been possible to identify allelic variation that correlates with gene expression.^{35, 36} Such expression quantitative trait loci (eQTLs) can be mapped in close proximity (*cis*-eQTL) or at considerable distance (*trans*-eQTL) from the gene of interest. *cis*-eQTL often identify TFBS that mediate local control of gene expression while *trans*-eQTL identify transcription factors that regulate expression of genes that reside in the locus of interest. *cis*-eQTLs have been shown to be highly enriched in those SNPs most closely associated with complex human traits and diseases and often represent causal variation.^{37, 38} eQTL datasets have been generated for a number of human tissues, including brain, circulating leukocytes, adipose and liver tissues, and these datasets are now widely available.³⁹

cis-eQTLs in linkage disequilibrium (LD) with disease-associated SNPs are also good candidates for identifying causal variation, but there are significant limitations with relying on these data. It is important to realize that numerous SNPs in a region of the genome may score as eQTLs due to linkage disequilibrium, with no clear indication of which SNP is driving the effect. Also, it is likely that phenotypic associations reflect eQTLs driven by

specific tissue/cell contexts, and thus examining tissues/cells irrelevant to the disease may miss critical eQTL effects. Ongoing efforts such as the NIH Genotype-Tissue Expression (GTEx) initiative are developing eQTL datasets that may be highly useful for mapping regulatory variation in a broader range of human tissues (<http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi>).

A second approach that employs genetics of gene expression data compares relative mRNA levels for the two alleles of the candidate causal gene in informative heterozygous individuals, and thus measures allele-specific expression (ASE). As both alleles are exposed to the same environmental and *trans*-genetic factors as well as subject to the same technical biases when sequencing, relative abundance of allele specific transcripts can reflect *cis*-acting effects. Correlation of the ASE with genotype has been employed to identify multiple types of mechanisms related to a variety of human diseases^{40–44}, and to map causal variation for a number of disease phenotypes.^{11, 40, 45–47} Functional genomic assays and statistical methods have been developed that utilize ASE data to connect trait-associated variants to genes and support mapping of *cis*-acting causal regulatory variation. RNA-Seq is increasingly employed to provide better assessment of gene expression for such studies, but a limitation of these approaches is that ASE information from RNA-Seq can be biased due to differences in sampling depths driven by differences in expression level across genes.⁴⁸ To address this, the Montgomery and Li labs have recently developed a targeted ASE sequencing method using microfluidics-based multiplex PCR and sequencing (mmPCR-Seq).⁴⁹ This system allows one to target and sequence the RNA of as many as 1000 heterozygous positions and with high-fidelity assess allelic imbalance. By further multiplexing samples, this approach provides an efficient and low-cost solution for assessing the impact and extent of ASE in a population sample.

Identification of causal genes

Identification of the causal gene is obviously critical for studies investigating the cellular mechanism of that gene in experimental model systems. In many cases, the causal gene in loci identified through GWAS investigation is the one closest to the genome-wide variation.³⁸ However, this is not always true and in some cases there are multiple proximally spaced protein coding genes in the region of the associated SNPs, requiring additional analysis to discriminate between the candidate genes. Genetics of gene expression data provides the best approach to date for identifying causal genes at GWAS loci. For instance, eQTL information has commonly been incorporated into the CARDIoGRAM and CARDIoGRAM+C4D meta-analyses to better identify the genes that mediate the risk association.^{1, 2} Use of allele-specific gene expression data has been used to identify those genes in the region of association that are in disequilibrium and thus likely to be under the influence of causal variation in the associated loci.

In the context of CHD GWAS loci, challenges associated with causal variant and gene identification at 9p21.3 have significantly impeded progress toward understanding the disease mechanisms associated with this important locus, and highlight the limitations/challenges of gene regulatory based methods in general. CHD associations at 9p21.3 were characterized by the identification of an ~ 50kb block of highly correlated SNPs

(chromosome 9 coronary artery associated region, C9CAR) and these variants were over 50 kb from the nearest protein coding genes, a pair of cyclin dependent kinase inhibitor (CDKI) genes.^{50–52} Interestingly, the CDKI genes were found to be connected by an antisense long non-coding RNA (*CDKN2B-AS1* or ANRIL) transcribed on the opposite strand to the nearest CDKI gene, *CDKN2B*.⁵³ A number of eQTL and ASE studies have been conducted with considerably varied results.⁵⁴ Holdt et al. found increased expression of the large C9CAR lncRNA, ANRIL in peripheral blood mononuclear cells from 1134 CHD patients with the 9p21 risk allele but no change in expression of nearby genes of *CDKN2A*, *CDKN2B*, or *ARF*.⁵⁵ By contrast, Liu et al. found significantly reduced expression of *CDKN2A*, *CDKN2B*, *ARF* and *CDKN2B-AS1* in peripheral blood T-cells from 170 healthy individuals with the risk allele.⁵⁶ Further, Jarinova et al. found in whole blood studies of 124 subjects RNA expression of the short variants of ANRIL was increased by 2.2-fold whereas expression of the long ANRIL variant was decreased by 1.2-fold in healthy subjects homozygous for the risk allele.⁵⁷ Expression levels of the long ANRIL variant (*CDKN2B-AS1*) were positively correlated with that of *CDKN2B*, while short variants were inversely correlated. In general, however, most studies suggested that *CDKN2B-AS1* and *CDKN2B* expression are inversely correlated. The differing results from the various eQTL studies were almost certainly due to inadequate samples size, the use of non-diseased tissues or the use of complex tissues that included multiple non-disease related cell types. Detailed studies by Cunningham et al. with peripheral blood samples showed a low-level positive correlation between *CDKN2A*, *CDKN2B* and *CDKN2B-AS1* mRNA levels, but more importantly showed with elegant ASE studies that ANRIL was the gene most highly correlated with genotype at C9CAR, and that individual SNPs had an inverse effect on *CDKN2B* and *CDKN2B-AS1* expression.⁵⁸ From a mechanistic viewpoint, inverse correlation of these genes was consistent with experiments showing that transcription initiated from the ANRIL promoter suppresses transcription of *CDKN2B*⁵⁹, and studies showing that ANRIL recruits polycomb complex proteins that suppress expression of *CDKN2B*.⁶⁰ As noted previously, studies in African American cohorts and recent large-scale fine mapping studies in Caucasians and South Asians have identified an independent SNP in the *CDKN2B* gene that is not in LD with those at C9CAR and thus implicates this gene as causal. Of course the 9p21 locus has also been the topic of other approaches to identify the causal gene, including those by Harismendy et al. who used chromosome conformation capture and other methods to identify a gamma interferon responsive gene in this locus⁶¹, and CDKI gene deletion studies coupled with mouse genetic models and mechanistic in vitro studies^{62, 63}.

Experimental methods to study transcriptional mechanisms of disease association

Identification of transcription factor binding motif and consensus binding proteins

As described previously, the majority of disease-associated variation likely resides outside of structural genes to regulate gene expression^{4, 5, 24}, and increasingly causal variants are found to be eQTLs.⁶⁴ Thus further understanding of upstream pathways and specific disease mechanisms requires the study of transcriptional pathways that regulate expression of the causal gene, and are disrupted by causal variation. Such studies have been validated for a number of disease-associated variants.^{9–15, 17} It is also possible that regulatory elements

may be created by causal variation that introduces a TFBS and thus aberrant activation of the causal gene, producing a maladaptive phenotype. Scanning the genomic sequence surrounding putative causal variants using *in silico* prediction tools may also reveal cooperative TFBS or the functional identity of TF complexes. Searching relevant gene co-expression datasets might also identify specific TF family members that are predicted to interact with the binding site.

There are two highly cited databases that archive and organize knowledge of DNA binding TFs, TFBSs and their regulated genes: TRANSFAC⁶⁵ and JASPAR⁶⁶ databases. TRANSFAC and JASPAR are accessible databases of transcription factor binding models and can be used to identify putative DNA binding sites in a queried DNA sequence. Affinity in the context of TF binding is represented with a position weight matrix (PWM) or a position-specific scoring matrix (PSSM). These models can be used to scan a genome for putative TF binding sites that have a high similarity to the model. Other PWM or PSSM based databases include MatInspector⁶⁷ and UniPROBE⁶⁸, which rely on protein-binding microarray datasets, and databases like ORegAnno⁶⁹ and PAZAR⁷⁰ which have attempted to collect and curate individual experimentally proven transcription factor binding sites from the literature and datasets from CHIP-Seq and other genome-wide experiments. Also, this research direction has been greatly supported by recent high throughput methods, including the use of systematic evolution of ligands by exponential enrichment (SELEX) and next-generation sequencing to define the binding profiles for 800 transcription factors.⁷¹

In the simplest paradigm, with at least one of the database searches the causal variant region will provide a clear contrast between alleles with one allele having a score >90 and the other having a score of <80 (see analyses in Miller et al.).³¹ In most cases, the protective allele is expected to have the higher score but it is also possible that the risk allele will have the higher score. Also, the two alleles may show predicted preferences for different members of the same family, with both alleles showing a high score but to different family members or structurally related TFs. In some cases the search algorithms return family names (e.g., AP1), and in other cases will predict binding of specific factors (e.g., ATF2) in cases where there might be minor differences in binding sequences among the family members.

Identification of candidate transcriptional networks that are associated with causal CHD variants and genes provide valuable information regarding upstream signaling pathways that regulate activation of these TFs and thus mediate disease risk. For instance, studies in this laboratory have identified embryonic and growth factor signaling pathways that may mediate a portion of the disease risk at the *TCF21* locus. Platelet-derived growth factor (PDGF) dependent AP-1 factors (c-JUN, JUND, and ATF3) were shown to preferentially bind and *trans*-activate an enhancer element containing the risk allele at rs12190287.³¹ However, binding of the epicardial developmental factor Wilms' tumor (WT1) to the risk allele led to *trans*-repression of this element.³¹ The complex architecture of transcription factor binding complexes containing both activators and repressors thus requires robust experimental validation to elucidate causal mechanisms of *in silico* predicted TF binding to disease associated variants.

In vitro studies of transcription factor binding to causal regulatory motifs

Many families of transcriptional regulatory factors have now been identified, their binding sequences defined, and their association with basic aspects of cell-specific, hormonal and biophysical gene regulation investigated.^{66, 72} Also, many TFs have been studied in the context of upstream signaling pathways that modulate their function, and identification of the TF often points to fundamental growth factor or cytokine signaling pathways that modulate cellular functions. TF binding sequences have historically been investigated with in vitro DNase I footprinting and gel mobility shift assays, but these techniques have now given way to chromatin immunoprecipitation (ChIP) that better reflects the in vivo state of chromatin in the region under study. For this approach, cultured cells or tissues are fixed in the native state providing crosslinking of TFs to DNA, and when coupled with immunoprecipitation with specific TF antibodies, provide assessment of binding to DNA regions of interest. Antibodies are now available for a large number of the protein factors that mediate transcriptional regulation, and a number of vendors have validated their antibodies to TFs specifically for the ChIP procedure.

If differences in TF binding in the CHD locus and resulting overall differences in transcription level is the mechanism by which causal variants mediate disease risk, it is predicted that TF binding to the risk versus protective allele in vivo is different. This can be assessed by performing ChIP and quantifying the relative amount of DNA representing each allele that is immunoprecipitated from vascular cells that are heterozygous for the causal SNP (Haplo-ChIP).⁷³ Allele-specific quantitation of binding is assessed by employing genotyping primers in a qPCR reaction, or another methodology that quantifies the number of DNA molecules for each allele. We have employed this approach in recent studies of CHD risk association at the 6q23.2 region, using primarily qPCR with genotyping primers and pyrosequencing as a second confirmatory method.³¹

If there are allele-specific differences in TF binding in the causal variant region, it is also expected that this will correlate with differences in local chromatin architecture. To investigate this possibility well-validated histone modification antibodies are employed to verify that the allele with greater TF binding shows a more open configuration, as evidenced by increased amplification of that allele (HistoneMod-ChIP). We have recently used this approach in studies of chromatin configuration at rs12190287 in the CHD associated region at 6q23.2.³¹ Identification of specific TFs will point to upstream signaling pathways and epigenetic mechanisms that could be further investigated in the context of overall gene expression for the CHD-associated genes.³¹

Reporter gene studies

It is essential to follow up the TFBS prediction and ChIP studies described in the previous sections with reporter gene transfection experiments investigating the ability of the causal regulatory variants to alter expression in an allele-specific fashion. These studies address the critical issue of directionality of predicted TF binding factors and whether there is a difference in transcriptional activity between the two alleles when the TFBS is accessible for binding. Also, it is important to verify that transcriptional activity is appropriately modified when cells are grown in the presence or absence of upstream pathway activation. The

direction of effect seen with the reporter assays is important, and when combined with the direction of effect observed in the GWAS studies provides valuable information regarding the upstream signaling pathways. Lastly, by incorporating loss-of-function studies (using siRNAs against upstream TFs) with the reporter assays it is possible to confirm the identification and function of the specific TFs implicated by the ChIP experiments. This is important since some antibodies detect more than one TF family member, and evidence of TF binding does not imply activation.

The recent development of high-throughput methods opens new opportunities for the study of putative TFBS identified through GWAS and follow-up studies, one can investigate all the associated loci and those in high LD simultaneously to look for those putative TFBS identified by disease-related SNPs. Massively parallel reporter assays have been used to simultaneously interrogate over 2000 putative enhancers in vitro, and STARR-Seq and CRE-Seq methods have been developed for the purpose of querying a large number of putative enhancer regions related to GWAS loci.^{74–76}

It is anticipated that luciferase assay results from these reporter gene transfection studies will track with TF binding, with the two alleles showing different transcriptional activity, and contextual dependence of protein binding expected to be mirrored in transcriptional activity and specificity. Identification of specific TFs through changes in transcriptional activity in association with siRNA inclusion in the reporter gene transfections is expected in most cases to be congruent with findings in ChIP experiments. Verification of a transcription regulatory sequence with allelic functional differences, along with identification of the cognate regulatory protein, will point to growth factor, cytokine, or other regulatory pathways that signal to the CHD locus and thus warrant further study.

Evaluation of enhancer function at modified TFBS

Previous in vitro studies will establish that when the TFs have access, i.e., in the presence of open chromatin, their function is modulated by the alteration in sequence introduced by the causal SNP. However, there will remain doubt regarding the ability of the TFBS to function as predicted in the in vivo setting, i.e., whether this site is truly an “enhancer” like region in vivo. We postulate that the majority of the TFBSs and cognate TFs that are associated with CHD causal variation mediate signaling through fundamental enhancer pathways that are involved with embryonic development, SMC lineage determination, and/or vascular wall specific functions such as SMC response to injury and contribution to atherosclerotic disease. This hypothesis builds on numerous examples whereby enhancer region function appears to be altered by complex disease-associated allelic variation, for metabolic disease in the *TCF7L2* locus^{12, 77}, colorectal cancer in the *SMAD7* gene at 18q21¹³ and the *EIF3H* gene at 8p23.3¹⁴, and several forms of cancer linked to a SNP in a *MYC* enhancer region.^{9, 10, 15} For a small number of disease-associated causal regulatory variants, transgene expression has validated the in vivo relevance of an associated enhancer in mouse or *Xenopus* models.^{13, 15} To investigate the gene expression patterns that are specific to transcription factor pathways associated with CHD causal variants, we propose that expression of transgenic constructs containing either the risk or protective allele TFBS, in the context of mouse embryonic development and the response to vascular injury and

atherosclerotic disease. These studies would address: *i*) whether the causal regulatory variant resides in a functional enhancer region, and *ii*) whether the causal regulatory SNP alters the function of this enhancer.

Genome editing applications for CHD loci

Recent advances in site-specific genome editing/targeting technologies have enabled highly efficient targeting of genetic loci in mammalian cells and tissues. It seems likely these tools will be adopted for the functional interrogation of GWAS loci for complex diseases. First generation genome-editing approaches involved zinc-finger nucleases (ZFN) to produce either double-strand breaks (DSBs) or edited alleles using natural repair mechanisms: homologous recombination (HR) or nonhomologous end-joining (NHEJ). More recently, Transcription Activator-Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)/CRISPR-associated (Cas) systems have been developed. These modern genome editing tools have enabled the modification of genomic sites of interest in both mammalian cells and model organisms. Importantly, when coupled to induced Pluripotent Stem Cell (iPSC) derived primary human cells, these approaches hold translational potential to investigate GWAS signals in patient samples from case-controlled studies for either diagnostic or therapeutic purposes.

ZFNs represent a class of engineered nucleases that create DSBs as heterodimers and can be modified to recognize specific genomic sequences. Essentially, site-specific DNA-binding domains are fused to a *FokI* derived nuclease domain to direct DNA cleavage. ZFN based genome-targeting was recently used to introduce disease causing mutations into the safe harbor gene locus (PPP1R2C/AAVS1) in iPSCs and human embryonic stem cells.⁷⁸ Importantly these ZFN edited iPSC-endothelial cell and cardiomyocytes and mouse lines were tracked over time using multi-modality imaging techniques.⁷⁸ ZFNs have been extensively used for targeted gene inactivation in zebrafish models^{79, 80} and more recently in rats to investigate the function of the hypertension candidate gene *Adamts16*.⁸¹

Derived from bacterial plant pathogens, TAL effector proteins have been engineered as the next generation of genome targeting systems. Similar to ZFNs, TALENs are generated by fusing DNA binding domains to the *FokI* nuclease domain and have highly efficient nuclease activity, along with more reliable gene targeting. More recently, bacteria and archaea derived CRISPR/Cas based genome targeting systems have further revolutionized RNA-guided genome engineering with even greater targeting efficiency and scalability in multiple model systems. By designing site-specific guide RNAs with a short protospacer adjacent motif (PAM) sequence at the 3' end, complementary genomic target sites can be targeted by the Cas9 nuclease for cleavage. It is now possible to inactivate, repress, or activate multiple genes in both human cells, and mouse models with approximately 80% efficiency.^{82, 83} The ease of use of this technology is expected to accelerate the functional interrogation of multiple CHD GWAS loci in iPSC-derived vascular cells from patients harboring multiple disease variants, as well as in mouse models of atherosclerosis. Up to five gene regions have been targeted at once,⁸³ suggesting knockouts of disease associated gene networks or pathways in future studies. For instance, once allele-specific TF complexes have been fully characterized in relevant model systems, these could be

inactivated along with the target locus to disrupt CHD-associated regulatory networks. Recent modifications of the Cas9 system to generate nuclease deficient Cas9 (dCas9) fused to functional regulatory domains such as VP16 (activating) and KRAB (repressing) has enabled the modulation of specific genomic regions.⁸⁴ Potential limitations of these approaches include off-targeting effects, which may complicate functional interpretation if mutations are introduced in transcripts within the same disease network. However, recent transcriptome-wide evidence suggests these off-target mutations are rare events using the CRISPR based system⁸³. Taken together, these tools may be directly applied to define the function of multiple noncoding regulatory variants at GWAS loci.

Identification of upstream signaling pathways

A major challenge in investigating the causal mechanisms of GWAS signals is defining the relevant upstream stimuli converging on regulatory variants. As mentioned above, these pathways could be indirectly inferred from identified TFs at *cis*-regulatory elements using various pathway analysis tools and databases, which can be customized with cell or gene specific expression datasets (e.g. Microarray or RNA-seq) or TF binding datasets (e.g. ChIP-seq). Proprietary knowledge-based pathway analysis tools include Ingenuity Pathway Analysis (IPA) (Ingenuity/Qiagen), MetaCore (GeneGO/Thomson Reuters), and Pathway Studio (Ariadne/Elsevier), which are manually curated from experimentally validated interactions. Open-source tools include Cytoscape (<http://www.cytoscape.org>), STRING (string-db.org), and Pathway Commons (www.pathwaycommons.org). Many of these pathway visualization tools utilize gene ontology (GO) based annotations and publicly available pathway databases (e.g. KEGG (www.genome.jp/kegg/), Reactome (<http://www.reactome.org/PathwayBrowser/>), MSigDB (<http://www.broadinstitute.org/gsea/msigdb/>). Alternatively, gene-set enrichment tools can be useful to prioritize functionally relevant pathways specific to biological or disease processes (e.g. DAVID (<http://david.abcc.ncifcrf.gov/>), GREAT (<http://bejerano.stanford.edu/great/>) GSEA (<http://www.broadinstitute.org/gsea/>). For instance, if a regulatory variant is predicted to bind AP-1 factors (e.g., c-Jun) and the target gene is enriched in smooth muscle cells, the candidate upstream pathways may include PDGF-BB-MAPK and TGF- β 1-Smad, which are linked to the regulation of smooth muscle cell proliferation and differentiation, migration and matrix remodeling. While this approach enabled identification of candidate regulatory pathways upstream of the *TCF21* association,³¹ some associations may require the integration of multiple high-throughput approaches.

Genome-wide or proteome-wide techniques have the potential to elucidate the unanticipated binding of TF complexes and inference of relevant pathways at GWAS loci. As noted previously, ChIP-seq is a powerful high-throughput approach to uncover multiple TFBS in the native chromatin context. When coupled to gene-set enrichment tools such as GREAT and PRISM, the upstream pathways can be inferred from experimental evidence and phylogenetic conservation. Recent modifications of ChIP-seq include ChIP-exo, through which exonucleases can be applied to gain single nucleotide resolution of TF binding.⁸⁵ However, remaining limitations of ChIP-based approaches include the lack of validated antibodies to the majority of the TFs in the human proteome, the requirement of large numbers of cells (especially for low abundant TFs), and the inability to define dynamic and

heterogeneous TF complexes. Signaling pathways are driven by specific ligand or mechanical cues to alter defined spatial and temporal cascades. Further, modulation of TFs to turn on/off expression often relies on rapid, site-specific post-translational modifications, such as protein phosphorylation. Complementary approaches to capture this depth of signaling at sites of genetic variation include DNA-affinity pulldown-mass-spectrometry.⁸⁶ These studies have applied allele-specific biotinylated oligo-precipitation to capture TF protein binding complexes to regulatory SNPs, which can be identified from eluate fractions using standard LC-MS mass spectrometry. If applied to cross-linked chromatin samples at distinct cellular states and time points, this approach has the potential to provide a more comprehensive view of functional TF complexes at GWAS loci. Alternatively, RNA-immunoprecipitation (RIP) followed by next generation sequencing⁸⁷ or LC-MS mass-spectrometry⁸⁸ could also reveal RNA dependent interactions at disease variants, involving RNA cofactors as well as miRNA and long non-coding RNA (lncRNA).

Perspective

The above approaches are mostly restricted to pooled populations of cells, which may not entirely resolve the causal mechanisms and upstream pathways at individual loci. Further refinements may include single-cell analyses of TF binding at sites of regulatory variation using “SMART” RNA-seq or microfluidics.⁸⁹ By sequencing transcriptomes of single immune cells, these studies revealed wide temporal heterogeneity in gene expression and splicing patterns. Also, multiple omics datasets of gene expression, chromatin binding, protein-protein interaction or other functional measures could be integrated to further define causal regulatory mechanisms at disease loci.^{90, 91} Finally, more robust statistical or machine learning tools could be applied to existing methods to accurately infer disease mechanisms.^{92, 93} It is anticipated that a combination of both experimental and computational paradigms would be required to prioritize, predict, and target functional mechanisms of the expanding number of CHD associations.

Conclusions

Progress with human genetic whole genome studies over the past few years have finally provided investigators with loci and genes that are incontrovertibly linked to causal mechanisms in the atherosclerotic vessel wall. While it is early days with respect to the identification of molecular and cellular causal effects that underlie disease initiation and progression, there are now highly sophisticated tools that allow the identification of causal variation, causal genes, and the signaling pathways upstream and downstream of the disease gene that can be targeted for treatment of the primary disease process.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

- **1. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardisino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeier J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoop JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011; 43:333–8. One of two large initial meta-analyses for coronary heart disease, identified 13 new loci and investigated causal gene identity with eQTL analyses. [PubMed: 21378990]
- **2. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, Konig IR, Cazier JB, Johansson A, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikainen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney AS, Mokhtari NE, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Muller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schafer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, Schoot CE, Wagner PJ, Wells GA, Wild PS, Yang TP, Amouyel P, Arveiler D, Basart H, Boehnke M, Boerwinkle E, Brambilla P, Cambien F, Cupples AL, de Faire U, Dehghan A, Diemert P, Epstein SE, Evans A, Ferrario MM, Ferrieres J, Gauguier D, Go AS, Goodall AH, Gudnason V, Hazen SL, Holm H, Iribarren C, Jang Y, Kahonen M, Kee F, Kim HS, Klopp N, Koenig W, Kratzer W, Kuulasmaa K, Laakso M, Laaksonen R, Lee JY, Lind L, Ouwehand WH, Parish S, Park JE, Pedersen NL, Peters A, Quertermous T, Rader DJ, Salomaa V, Schadt E, Shah SH, Sinisalo J, Stark K, Stefansson K, Tregouet DA, Virtamo J, Wallentin L, Wareham N, Zimmermann ME, Nieminen MS, Hengstenberg C, Sandhu MS, Pastinen T, Syvanen AC, Hovingh GK, Dedoussis G, Franks PW, Lehtimäki T, Metspalu A, Zalloua PA, Siegbahn A, Schreiber S, Ripatti S, Blankenberg SS, Perola M, Clarke R, Boehm BO, O'Donnell C, Reilly MP, Marz W, Collins R, Kathiresan S, Hamsten A, Kooner JS, Thorsteinsdottir U, Danesh J, Palmer CN, Roberts R, Watkins H, Schunkert H, Samani NJ. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* 2012; 45:25–33. Large meta-analysis of Metabochip data performing fine-mapping in known coronary disease loci, identifying new alleles in known associated loci, and identified novel loci from list of marginally associated variants. Employed early pathway analysis of GWAS data. [PubMed: 23202125]
- **3. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet.* 2011; 43:339–44. One of two large initial meta-analyses for

coronary heart disease, employed Caucasian and South Asians genome-wide data, identified five new loci. [PubMed: 21378988]

4. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res.* 2012; 22:1748–59. [PubMed: 22955986]
5. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009; 106:9362–7. [PubMed: 19474294]
- *6. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutayvin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012; 337:1190–5. One of a number of superb studies emanating from the ENCODE consortium, this publication described the use of DNase I hypersensitivity assays to link disease causal variation to regulatory genetic regions. [PubMed: 22955828]
7. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 2010; 6:e1000888. [PubMed: 20369019]
8. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Kotliansky V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature.* 2010; 466:714–9. [PubMed: 20686566]
9. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, Yao K, Kehoe SM, Lenz HJ, Haiman CA, Yan C, Henderson BE, Frenkel B, Barretina J, Bass A, Tabernero J, Baselga J, Regan MM, Manak JR, Shivdasani R, Coetzee GA, Freedman ML. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet.* 2009; 41:882–4. [PubMed: 19561607]
10. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei G, Yan J, Niittymäki I, Mecklin JP, Jarvinen H, Ristimäki A, Di-Bernardo M, East P, Carvajal-Carmona L, Houlston RS, Tomlinson I, Palin K, Ukkonen E, Karhu A, Taipale J, Aaltonen LA. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet.* 2009; 41:885–90. [PubMed: 19561604]
- **11. Conde L, Bracci PM, Richardson R, Montgomery SB, Skibola CF. Integrating GWAS and Expression Data for Functional Characterization of Disease-Associated SNPs: An Application to Follicular Lymphoma. *Am J Hum Genet.* 2013; 92:126–30. Work described in this paper used allele-specific expression and innovative analysis methods to map disease loci in follicular lymphoma. [PubMed: 23246294]
12. Savic D, Ye H, Aneas I, Park SY, Bell GI, Nobrega MA. Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism. *Genome Res.* 2011; 21:1417–25. [PubMed: 21673050]
13. Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, van Wezel T, Morreau H, Sullivan K, Fielding S, Twiss P, Vijaykrishnan J, Casares F, Qureshi M, Gomez-Skarmeta JL, Houlston RS. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res.* 2009; 19:987–93. [PubMed: 19395656]
14. Pittman AM, Naranjo S, Jalava SE, Twiss P, Ma Y, Olver B, Lloyd A, Vijaykrishnan J, Qureshi M, Broderick P, van Wezel T, Morreau H, Tuupanen S, Aaltonen LA, Alonso ME, Manzanares M, Gavilan A, Visakorpi T, Gomez-Skarmeta JL, Houlston RS. Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. *PLoS Genet.* 2010; 6.
15. Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res.* 2010; 20:1191–7. [PubMed: 20627891]

- **16. Karczewski KJ, Dudley JT, Kukurba KR, Chen R, Butte AJ, Montgomery SB, Snyder M. Systematic functional regulatory assessment of disease-associated variants. *Proc Natl Acad Sci U S A*. 2013; 110:9607–12. These data suggest that variation-related alteration of NFkB binding may be a common mechanism for complex human disease. [PubMed: 23690573]
17. Schodel J, Bardella C, Sciesielski LK, Brown JM, Pugh CW, Buckle V, Tomlinson IP, Ratcliffe PJ, Mole DR. Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat Genet*. 2012; 44:420–5. S1–2. [PubMed: 22406644]
18. Chen K, Song F, Calin GA, Wei Q, Hao X, Zhang W. Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. *Carcinogenesis*. 2008; 29:1306–11. [PubMed: 18477647]
19. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer*. 2010; 10:389–402. [PubMed: 20495573]
20. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013; 31:142–7. [PubMed: 23334450]
21. Leung A, Schones DE, Natarajan R. Using epigenetic mechanisms to understand the impact of common disease causing alleles. *Current opinion in immunology*. 2012; 24:558–63. [PubMed: 22857822]
- **22. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B, Almeida R, Zhernakova A, Reinmaa E, Vosa U, Hofker MH, Fehrmann RS, Fu J, Withoff S, Metspalu A, Franke L, Wijmenga C. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet*. 2013; 9:e1003201. Important observations pointing to lncRNAs as possible intermediaries for linking causal variation to protein coding function. [PubMed: 23341781]
23. Popadin K, Gutierrez-Arcelus M, Dermitzakis ET, Antonarakis SE. Genetic and Epigenetic Regulation of Human lincRNA Gene Expression. *Am J Hum Genet*. 2013; 93:1015–26. [PubMed: 24268656]
24. Gaffney DJ. Global properties and functional complexity of human gene regulatory variation. *PLoS Genet*. 2013; 9:e1003501. [PubMed: 23737752]
25. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011; 1:457–70. [PubMed: 22384356]
26. Zaitlen N, Pasaniuc B, Gur T, Ziv E, Halperin E. Leveraging genetic variability across populations for the identification of causal variants. *Am J Hum Genet*. 2010; 86:23–33. [PubMed: 20085711]
- **27. Kral BG, Mathias RA, Suktitipat B, Ruczinski I, Vaidya D, Yanek LR, Quyyumi AA, Patel RS, Zafari AM, Vaccarino V, Hauser ER, Kraus WE, Becker LC, Becker DM. A common variant in the CDKN2B gene on chromosome 9p21 protects against coronary artery disease in Americans of African ancestry. *J Hum Genet*. 2011 Small association study for CHD in African Americans identified a novel association in the CDKN2B gene, showing the importance of doing association studies in non-Caucasian groups and in African Americans in particular who have shorter regions of linkage disequilibrium.
- *28. Lu X, Wang L, Chen S, He L, Yang X, Shi Y, Cheng J, Zhang L, Gu CC, Huang J, Wu T, Ma Y, Li J, Cao J, Chen J, Ge D, Fan Z, Li Y, Zhao L, Li H, Zhou X, Chen L, Liu D, Chen J, Duan X, Hao Y, Wang L, Lu F, Liu Z, Yao C, Shen C, Pu X, Yu L, Fang X, Xu L, Mu J, Wu X, Zheng R, Wu N, Zhao Q, Li Y, Liu X, Wang M, Yu D, Hu D, Ji X, Guo D, Sun D, Wang Q, Yang Y, Liu F, Mao Q, Liang X, Ji J, Chen P, Mo X, Li D, Chai G, Tang Y, Li X, Du Z, Liu X, Dou C, Yang Z, Meng Q, Wang D, Wang R, Yang J, Schunkert H, Samani NJ, Kathiresan S, Reilly MP, Erdmann J, Peng X, Wu X, Liu D, Yang Y, Chen R, Qiang B, Gu D. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet*. 2012; 44:890–4. Important GWAS study showing replication of a number of Caucasian CHD associated variants and identifying new associated loci in an East Asian population. [PubMed: 22751097]
29. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013

30. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* 2011; 21:447–55. [PubMed: 21106904]
- **31. Miller CL, Anderson DR, Kundu RK, Raiesdana A, Nurnberg ST, Diaz R, Cheng K, Leeper NJ, Chen CH, Chang IS, Schadt EE, Hsiung CA, Assimes TL, Quertermous T. Disease-Related Growth Factor and Embryonic Signaling Pathways Modulate an Enhancer of TCF21 Expression at the 6q23.2 Coronary Heart Disease Locus. *PLoS Genet.* 2013; 9:e1003652. Initial study from the authors investigating mechanisms by which causal variation, and upstream signaling pathways, regulate CHD gene TCF21 expression. [PubMed: 23874238]
32. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009; 136:215–33. [PubMed: 19167326]
33. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, Wojcik SE, Ferdin J, Kunej T, Xiao L, Manoukian S, Secreto G, Ravagnani F, Wang X, Radice P, Croce CM, Davuluri RV, Calin GA. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res.* 2010; 70:2789–98. [PubMed: 20332227]
34. Miller CL, Haas U, Diaz R, Leeper NJ, Kundu RK, Patolla B, Assimes TL, Kaiser FJ, Ljubica P, Hedin U, Maegdefessel L, Schunkert H, Erdmann J, Quertermous T, Sczakiel G. Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation. 2014 in press.
35. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008; 6:e107. [PubMed: 18462017]
- *36. Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, Durbin R, Deloukas P, Dermitzakis ET. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.* 2012; 8:e1002639. This study investigates the genetics of gene expression in multiple populations, providing insights regarding the transferability of regulatory variation across different racial ethnic groups. [PubMed: 22532805]
37. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, Weingarh DT, Zhang B, Greenawalt D, Dobrin R, Hao K, Woo S, Fabre-Suver C, Qian S, Tota MR, Keller MP, Kendziora CM, Yandell BS, Castro V, Attie AD, Kaplan LM, Schadt EE. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* 2010; 6:e1000932. [PubMed: 20463879]
38. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Zhao JH, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, Konig IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N, Haiqing S, Hartikainen AL, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpelainen TO, Koiranen M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Pare G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietilainen KH, Pouta A, Ridderstrale M, Rotter JJ, Sambrook JG, Sanders AR, Schmidt CO, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle

- WL, Nelis M, Peters MJ, Ripatti S, van Meurs JB, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kahonen M, Kaprio J, Kathiresan S, Kiemeny L, Kocher T, Launer LJ, Lehtimäki T, Melander O, Mosley TH Jr, Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tonjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Gronberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Lathrop GM, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Cupples LA, Erdmann J, Fox CS, Gudnason V, Gyllenstein U, Harris TB, Hayes RB, Jarvelin MR, Mooser V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Volzke H, Watkins H, Wilson JF, Groop LC, Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010; 467:832–8. [PubMed: 20881960]
39. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013; 45:580–5. [PubMed: 23715323]
 40. Lefebvre JF, Vello E, Ge B, Montgomery SB, Dermitzakis ET, Pastinen T, Labuda D. Genotype-based test in mapping cis-regulatory variants from allele-specific expression data. *PLoS One*. 2012; 7:e38667. [PubMed: 22685595]
 41. Ria M, Lagercrantz J, Samnegard A, Boquist S, Hamsten A, Eriksson P. A common polymorphism in the promoter region of the TNFSF4 gene is associated with lower allele-specific expression and risk of myocardial infarction. *PLoS One*. 2011; 6:e17652. [PubMed: 21445270]
 42. Milani L, Lundmark A, Nordlund J, Kiialainen A, Flaegstad T, Jonmundsson G, Kanerva J, Schmiegelow K, Gunderson KL, Lonnholm G, Syvanen AC. Allele-specific gene expression patterns in primary leukemic cells reveal regulation of gene expression by CpG site methylation. *Genome Res*. 2009; 19:1–11. [PubMed: 18997001]
 43. Valle L, Serena-Acedo T, Liyanarachchi S, Hampel H, Comeras I, Li Z, Zeng Q, Zhang HT, Pennison MJ, Sadim M, Pasche B, Tanner SM, de la Chapelle A. Germline allele-specific expression of TGFBR1 confers an increased risk of colorectal cancer. *Science*. 2008; 321:1361–5. [PubMed: 18703712]
 44. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, Urban AE, Montgomery SB, Levinson DF, Koller D. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res*. 2014; 24:14–24. [PubMed: 24092820]
 45. Campino S, Forton J, Raj S, Mohr B, Auburn S, Fry A, Mangano VD, Vandiedonck C, Richardson A, Rockett K, Clark TG, Kwiatkowski DP. Validating discovered Cis-acting regulatory genetic variants: application of an allele specific expression approach to HapMap populations. *PLoS One*. 2008; 3:e4105. [PubMed: 19116668]
 46. Li Y, Grupe A, Rowland C, Nowotny P, Kauwe JS, Smemo S, Hinrichs A, Tacey K, Toombs TA, Kwok S, Catanese J, White TJ, Maxwell TJ, Hollingworth P, Abraham R, Rubinsztein DC, Brayne C, Wavrant-De Vrieze F, Hardy J, O'Donovan M, Lovestone S, Morris JC, Thal LJ, Owen M, Williams J, Goate A. DAPK1 variants are associated with Alzheimer's disease and allele-specific expression. *Hum Mol Genet*. 2006; 15:2560–8. [PubMed: 16847012]
 47. Serre D, Gurd S, Ge B, Sladek R, Sinnett D, Harmsen E, Bibikova M, Chudin E, Barker DL, Dickinson T, Fan JB, Hudson TJ. Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet*. 2008; 4:e1000006. [PubMed: 18454203]
 48. Sun W. A statistical framework for eQTL mapping using RNA-seq data. *Biometrics*. 2012; 68:1–11. [PubMed: 21838806]

49. Zhang R, Li X, Ramaswami G, Smith KS, Turecki G, Montgomery GW, Li JB. Quantifying RNA allelic ratios by microfluidics-based multiplex PCR and sequencing. *Nature Methods*. 2013 in press.
- **50. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007; 316:1488–91. One of two early studies identifying association of 9p21.3 with CHD. [PubMed: 17478681]
- **51. Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, Jonasdóttir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthíasdóttir S, Jónsdóttir T, Palsson S, Einarsson H, Gunnarsdóttir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorsteirsson G, Thorsteinsdóttir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007; 316:1491–3. One of two early studies identifying association of 9p21.3 with CHD. [PubMed: 17478679]
52. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–78. [PubMed: 17554300]
53. Pasmant E, Laurendeau I, Heron D, Vidaud M, Vidaud D, Bieche I. Characterization of a germline deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res*. 2007; 67:3963–9. [PubMed: 17440112]
- *54. Cunnington MS, Keavney B. Genetic mechanisms mediating atherosclerosis susceptibility at the chromosome 9p21 locus. *Curr Atheroscler Rep*. 2011; 13:193–201. [PubMed: 21487702] Studies showing that ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler Thromb Vasc Biol*. 2010; 30:620–7. [PubMed: 20056914]
55. Holdt LM, Beutner F, Scholz M, Gielen S, Gabel G, Bergert H, Schuler G, Thiery J, Teupser D. ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler Thromb Vasc Biol*. 2010; 30:620–7. [PubMed: 20056914]
56. Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Mohlke KL, Ibrahim JG, Thomas NE, Sharpless NE. INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS ONE*. 2009; 4:e5027. [PubMed: 19343170]
57. Jarinova O, Stewart AF, Roberts R, Wells G, Lau P, Naing T, Buerki C, McLean BW, Cook RC, Parker JS, McPherson R. Functional Analysis of the Chromosome 9p21. 3 Coronary Artery Disease Risk Locus. *Arterioscler. Thromb Vasc Biol*. 2009; 29:1671–7.
58. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS Genet*. 2010; 6:e1000899. [PubMed: 20386740]
59. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*. 2008; 451:202–6. [PubMed: 18185590]
60. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene*. 2011; 30:1956–62. [PubMed: 21151178]
61. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature*. 2011; 470:264–8. [PubMed: 21307941]
62. Leeper NJ, Raiesdana A, Kojima Y, Kundu RK, Cheng H, Maegdefessel L, Toh R, Ahn GO, Ali ZA, Anderson DR, Miller CL, Roberts SC, Spin JM, de Almeida PE, Wu JC, Xu B, Cheng K, Quertermous M, Kundu S, Kortekaas KE, Berzin E, Downing KP, Dalman RL, Tsao PS, Schadt EE, Owens GK, Quertermous T. Loss of CDKN2B promotes p53-dependent smooth muscle cell apoptosis and aneurysm formation. *Arterioscler Thromb Vasc Biol*. 2013; 33:e1–e10. [PubMed: 23162013]

63. Visel A, Zhu Y, May D, Afzal V, Gong E, Attanasio C, Blow MJ, Cohen JC, Rubin EM, Pennacchio LA. Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature*. 2010; 464:409–12. [PubMed: 20173736]
64. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, Weingarh DT, Zhang B, Greenawalt D, Dobrin R, Hao K, Woo S, Fabre-Suver C, Qian S, Tota MR, Keller MP, Kendziorzski CM, Yandell BS, Castro V, Attie AD, Kaplan LM, Schadt EE. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet*. 2010; 6:e1000932. [PubMed: 20463879]
65. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCmpel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. 2006; 34:D108–10. [PubMed: 16381825]
66. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2010; 38:D105–10. [PubMed: 19906716]
67. Quandt K, Frech K, Karas H, Wingender E, Werner T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res*. 1995; 23:4878–84. [PubMed: 8532532]
68. Newburger DE, Bulyk ML. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res*. 2009; 37:D77–82. [PubMed: 18842628]
69. Montgomery SB, Griffith OL, Sleumer MC, Bergman CM, Bilenky M, Pleasance ED, Prychyna Y, Zhang X, Jones SJ. ORegAnno: an open access database and curation system for literature-derived promoters, transcription factor binding sites and regulatory variation. *Bioinformatics*. 2006; 22:637–40. [PubMed: 16397004]
70. Portales-Casamar E, Arenillas D, Lim J, Swanson MI, Jiang S, McCallum A, Kirov S, Wasserman WW. The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res*. 2009; 37:D54–60. [PubMed: 18971253]
71. Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J. DNA-binding specificities of human transcription factors. *Cell*. 2013; 152:327–39. [PubMed: 23332764]
72. Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res*. 2003; 31:3576–9. [PubMed: 12824369]
- **73. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet*. 2003; 33:469–75. Elegant study looking at the transcriptional mechanism for allele-specific expression, this study introduced the concept of allele-specific binding assays. [PubMed: 12627232]
74. Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res*. 2013; 23:800–11. [PubMed: 23512712]
75. Kwasniewski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci U S A*. 2012; 109:19498–503. [PubMed: 23129659]
76. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013; 339:1074–7. [PubMed: 23328393]
77. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, Berney T, Montanya E, Mohlke KL, Lieb JD, Ferrer J. A map of open chromatin in human pancreatic islets. *Nat Genet*. 2010; 42:255–9. [PubMed: 20118932]
78. Wang Y, Zhang WY, Hu S, Lan F, Lee AS, Huber B, Lisowski L, Liang P, Huang M, de Almeida PE, Won JH, Sun N, Robbins RC, Kay MA, Urmov FD, Wu JC. Genome editing of human embryonic stem cells and induced pluripotent stem cells with zinc finger nucleases for cellular imaging. *Circ Res*. 2012; 111:1494–503. [PubMed: 22967807]

79. Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol.* 2008; 26:695–701. [PubMed: 18500337]
80. Zhu C, Smith T, McNulty J, Rayla AL, Lakshmanan A, Siekmann AF, Buffardi M, Meng X, Shin J, Padmanabhan A, Cifuentes D, Giraldez AJ, Look AT, Epstein JA, Lawson ND, Wolfe SA. Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development.* 2011; 138:4555–64. [PubMed: 21937602]
81. Gopalakrishnan K, Kumarasamy S, Abdul-Majeed S, Kalinoski AL, Morgan EE, Gohara AF, Nauli SM, Filipiak WE, Saunders TL, Joe B. Targeted disruption of *Adamts16* gene in a rat genetic model of hypertension. *Proc Natl Acad Sci U S A.* 2012; 109:20555–9. [PubMed: 23185005]
82. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* 2013; 339:823–6. [PubMed: 23287722]
83. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell.* 2013; 153:910–8. Cas9 mRNA and single-guide RNAs were introduced into embryonic stem cells or induced pluripotent stem cells to show rapid genome editing in vivo. [PubMed: 23643243]
- *84. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 2013; 154:442–51. Work described in this publication revealed the ability of dCas9 protein to serve as a general platform for RNA-guided targeting, laying groundwork for the use of this approach to the identification of causal variation in cultured human cells, and the rapid manipulation of genomic sequences in vivo. [PubMed: 23849981]
85. Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell.* 2012; 149:1461–73. [PubMed: 22726434]
86. Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, Mann M. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet.* 2012; 8:e1002982. [PubMed: 23028375]
87. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell.* 2010; 40:939–53. [PubMed: 21172659]
88. Scheibe M, Butter F, Hafner M, Tuschl T, Mann M. Quantitative mass spectrometry and PAR-CLIP to identify RNA-protein interactions. *Nucleic Acids Res.* 2012; 40:9897–902. [PubMed: 22885304]
89. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublotte JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JJ, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H, Regev A. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature.* 2013; 498:236–40. [PubMed: 23685454]
90. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* 2010; 6:e1000895. [PubMed: 20369022]
91. Laurila PP, Surakka I, Sarin AP, Yetukuri L, Hyotylainen T, Soderlund S, Naukkarinen J, Tang J, Kettunen J, Mirel DB, Soronen J, Lehtimäki T, Ruokonen A, Ehnholm C, Eriksson JG, Salomaa V, Jula A, Raitakari OT, Jarvelin MR, Palotie A, Peltonen L, Oresic M, Jauhiainen M, Taskinen MR, Ripatti S. Genomic, transcriptomic, and lipidomic profiling highlights the role of inflammation in individuals with low high-density lipoprotein cholesterol. *Arterioscler Thromb Vasc Biol.* 2013; 33:847–57. [PubMed: 23413431]
92. Chung RH, Chen YE. A two-stage random forest-based pathway analysis method. *PLoS One.* 2012; 7:e36662. [PubMed: 22586488]
93. Califano A, Butte AJ, Friend S, Ideker T, Schadt E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet.* 2012; 44:841–7. [PubMed: 22836096]

94. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; 22:1790–7. [PubMed: 22955989]
95. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012; 40:D930–4. [PubMed: 22064851]
96. Griffith OL, Montgomery SB, Bernier B, Chu B, Kasaian K, Aerts S, Mahony S, Sleumer MC, Bilenky M, Haeussler M, Griffith M, Gallo SM, Giardine B, Hooghe B, Van Loo P, Blanco E, Ticoll A, Lithwick S, Portales-Casamar E, Donaldson IJ, Robertson G, Wadelius C, De Bleser P, Vlieghe D, Halfon MS, Wasserman W, Hardison R, Bergman CM, Jones SJ. ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res.* 2008; 36:D107–13. [PubMed: 18006570]
97. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489:57–74. [PubMed: 22955616]
98. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol.* 2010; 28:1045–8. [PubMed: 20944595]
99. Savic D, Bell GI, Nobrega MA. An in vivo cis-regulatory screen at the type 2 diabetes associated TCF7L2 locus identifies multiple tissue-specific enhancers. *PLoS One.* 2012; 7:e36501. [PubMed: 22590553]
100. Jao LE, Wenthe SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A.* 2013; 110:13904–9. [PubMed: 23918387]
101. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet.* 2010; 86:581–91. [PubMed: 20346437]

Table 1

Beyond-GWAS approaches to investigate causal variation, genes, and mechanisms at disease loci

Type of Analysis	Purpose	Strengths and/or Limitations	Examples at CAD or MI loci
Fine-mapping	Identify additional common or rare variants associated at GWAS loci	<ul style="list-style-type: none"> + May reveal variants not detected on initial GWAS SNP chip + Could utilize sequencing or densely covered fine-mapping arrays (e.g. Cardio-Metabochip) selected for specific cardiovascular traits 	9p21.3 ²⁷ , <i>COL4A1</i> ² , <i>CXCL12</i> ² , <i>TCF21</i> ²⁸
eQTL	Identify causal genes at candidate loci by correlating genotype to gene expression	<ul style="list-style-type: none"> + Compares expression across genotypes, detects <i>cis</i> and <i>trans</i> effects + Can be applied to microarray or RNA-seq expression data – Often requires a large number of samples from relevant cell/tissue sources, thus current data limited to accessible types – May not identify causal variants due to linkage disequilibrium 	Multiple CAD loci including: 1p13.3 ^{1,8} , 3q22.3 ¹ , 6q23.2 ¹ , 9p21.3 ¹ , 17p11.2 ¹ , 17q.21.32 ¹ , 19p13.2 ¹ , 21q22.11 ¹
ASE	Identify causal variants at candidate loci by measuring allele-specific expression	<ul style="list-style-type: none"> + Measures allelic ratio of transcripts in cells from individuals heterozygous for candidate eSNPs, thus primarily detects <i>cis</i> effects + Variants eliciting allelic imbalance are likely to be causal + Requires fewer samples, reducing statistical power needed + Can be applied to large-scale analysis using mmqPCR⁴⁹ – Potential biases with qPCR probes or RNA-seq read depth 	<i>CDKN2A</i> , <i>CDKN2B</i> , <i>CDKN2B-AS1</i> ⁵⁸ , <i>TCF21</i> ³¹
In silico prediction	Predict functional variants and regulatory mechanisms <i>in silico</i>	<ul style="list-style-type: none"> + Use accessible databases to predict variant function (e.g. RegulomeDB⁹⁴, HaploReg⁹⁵, ORegAnno⁹⁶) + Map chromatin accessibility and TF binding at candidate loci using genome-wide datasets (e.g. ENCODE⁹⁷, Roadmap⁹⁸ consortia) + Predict allele-specific TFBS using PWM databases (e.g. TRANSFAC) – Requires functional validation in disease-relevant cell types 	9p21.3 ⁴ , <i>TCF21</i> ³¹
Chromatin assays	Identify causal variants and potential regulatory mechanisms for candidate loci in native chromatin context	<ul style="list-style-type: none"> + Measures enrichment of TF binding to candidate region (ChIP) or allelic abundance at candidate SNP (haplo-ChIP) in native chromatin + Chromatin accessibility can be monitored through antibodies against histone modifications (histoneMod-ChIP) + Genome-wide approaches (ChIP-seq⁴, DNaseI-seq³⁰, ATAC-seq²⁹) – Requires carefully validated antibodies and technical expertise – Often lacks directionality and disease relevance unless samples were treated to modulate a particular cellular phenotype 	<i>TCF21</i> ³¹ , <i>TNFSF4</i> ⁴¹ , <i>SORT1</i> ⁸
Reporter assays	Identify causal variants and potential regulatory mechanisms for candidate loci <i>in vitro</i> or <i>in vivo</i>	<ul style="list-style-type: none"> + Useful to determine direction of regulatory variation using allele-specific reporters, and specific treatments in various cells + Luciferase reporters often used in cultured cells, but LacZ reporters useful to measure enhancer activity in animal models⁹⁹ + Applicable to large-scale analysis (massively parallel reporter assay) when coupled to RNA-seq⁷⁴ – Restricted to primary genomic sequence, lacks chromatin organization and structural information 	<i>CDKN2B-AS1</i> ⁵⁷ , <i>TCF21</i> ³¹ , <i>TNFSF4</i> ⁴¹ , <i>SORT1</i> ⁸
Genome-editing	Modulate identified candidate loci using <i>in vitro</i> and <i>in vivo</i> models	<ul style="list-style-type: none"> + CRISPR based genome-editing system is highly efficient and requires less technical experience for vector construction^{83, 100} + Capable of multiplexing to target up to 5 loci^{83, 100} 	

Type of Analysis	Purpose	Strengths and/or Limitations	Examples at CAD or MI loci
		<ul style="list-style-type: none"> – Target region may be restricted by required PAM sequence – Possible off-target cleavage effects may confound results 	
Pathway analysis	Identify upstream and downstream signaling pathways implicated	<ul style="list-style-type: none"> + Provides robust knowledge-based enrichment of genes and pathways from large-scale genomic or proteomic approaches¹⁰¹ – Potential biases due to curation of select pathways and phenotypes 	