

Published in final edited form as:

Curr Opin Lipidol. 2012 April ; 23(2): 98–103. doi:10.1097/MOL.0b013e328350fad2.

Finding genes and variants for lipid levels after genome-wide association analysis

Cristen J. Willer¹ and Karen L. Mohlke²

¹Departments of Internal Medicine (Cardiovascular Medicine) and Human Genetics, University of Michigan, Ann Arbor, MI

²Department of Genetics, University of North Carolina, Chapel Hill, NC

Abstract

Purpose of review—We review the main findings from genome-wide association studies (GWAS) for levels of HDL-cholesterol, LDL-cholesterol and triglycerides, including approaches to identify the functional variant(s) or gene(s). We discuss study design and challenges related to whole genome or exome sequencing to identify novel genes and variants.

Recent findings—GWAS have detected ~100 loci associated with one or more lipid trait. Fine-mapping of several loci for LDL-c demonstrated that the trait variance explained may double when the variants responsible for the association signals are identified. Experimental follow-up of three loci identified by GWAS has identified functional genes at *GALNT2*, *TRIB1*, and *SORT1*, and a functional variant at *SORT1*.

Summary—The goal of genetic studies for lipid levels is to improve treatment and ultimately reduce the prevalence of heart disease. Many signals identified by GWAS have modest effect sizes, useful for identifying novel biologically-relevant genes, but less useful for personalized medicine. Whole genome or exome sequencing studies may fill this gap by identifying rare variants of larger effect associated with lipid levels and heart disease.

Keywords

genome-wide association study; lipids; cholesterol; next-generation sequencing

Introduction

High levels of circulating low-density lipoprotein cholesterol (LDL-c) and low levels of high-density lipoprotein cholesterol (HDL-c) are strong risk factors for stroke and heart disease, the leading cause of death in industrialized nations [1]. Blood lipid levels including LDL-c, HDL-c and triglyceride levels (TG) are heritable, and known genetic effects explain ~10–15% of the trait variances [2]. These accurately-measured quantitative traits provide more power to find genes than a dichotomous trait such as myocardial infarction. Identifying genes that influence lipid levels may provide targets for pharmaceuticals to efficiently lower the risk of heart disease and related events.

Genome-wide association studies (GWAS) test hundreds of thousands to millions of genetic markers across the genome for association with a trait. GWAS have successfully identified

Contact information for corresponding author: Cristen J. Willer, Division of Cardiovascular Medicine, Department of Internal Medicine, 1241 E. Catherine St. Ann Arbor, MI 48103, cristen@umich.edu.

Conflicts of interest

There are no conflicts of interest.

novel loci for many disease-related quantitative traits including blood glucose and insulin levels [3–5], systolic and diastolic blood pressure [6], body mass index [7], waist-hip ratio [8], HDL-c, LDL-c, triglyceride levels and total cholesterol [9]. An initial GWAS for lipid levels in 8,816 individuals identified 18 loci at a genome-wide significant threshold ($P < 5 \times 10^{-8}$), 7 of which were novel [10]. Concurrently, the Diabetes Genetics Initiative identified an overlapping set of 6 novel loci [11]. Several subsequent meta-analyses of these or other studies identified additional novel lipid loci [12–17, 18]. In 2010, a meta-analysis of 46 cohorts, including ~100,000 individuals, identified a total of 31 loci associated with HDL-c, 22 loci associated with LDL-c, 16 loci associated with triglyceride levels and 39 loci associated with total cholesterol [9]. These genetic variants together accounted for ~25–30% of the genetic component of these traits, suggesting many lipid-associated genetic variants remain to be found. Although the majority of published GWAS studied European individuals, novel loci also have been identified in non-European populations [19].

Frequent criticisms of GWAS are that the variants discovered have small effect sizes and that their discovery requires sample sizes of tens or even hundreds of thousands of individuals. These statements are generally true. Most lipid-associated variants discovered by GWAS are both common (minor allele frequency $> .05$) and have small effect sizes that require large sample sizes to detect [9]. Larger effect sizes are observed for a few loci discovered prior to GWAS, such as *CETP* for HDL and *APOE* for LDL [20]; these signals are the low-hanging fruit of complex genetic traits. Few common variants with large effect sizes exist, likely due to natural selection. Irrespective of effect sizes, GWAS analyses detect novel genes and pathways that were not prior disease gene candidates. By leveraging nature's experiment on humans over evolutionary history, we can identify common variants with small influences on lipid levels that can lead us to genes with large influences on lipid levels. For example, LDL-c levels were significantly associated with GWAS SNPs near HMG Co-A reductase (*HMGCR*), the rate-limiting enzyme for cholesterol biosynthesis [9]. Typical of GWAS-identified variants, an LDL-associated genetic variant near *HMGCR* has an allele frequency of 39% and influences LDL cholesterol levels by a modest 2.5 mg/dL. However, use of statins, which inhibit the function of the rate-limiting enzyme of cholesterol synthesis, encoded by *HMGCR*, typically decreases LDL cholesterol levels by 20–40%, or ~14–70 mg/dL [21]. The goal of GWAS is not to catalogue all of the tiny effect size variants into a laundry list of genetic markers but to provide clues about biologically-relevant genes that remain undetected by previous approaches.

From GWAS loci to genes

One challenge of GWAS is to convert an association signal into an underlying functional gene. Infrequently, an associated genetic variant is a nonsense or nonsynonymous substitution in a protein coding region. For example, a genetic variant at 8p21 that shows strong association with HDL-c ($p = 3 \times 10^{-94}$) encodes a premature stop in the lipoprotein lipase gene (*LPL*) at codon 474 of a protein that typically has 475 amino acids (rs328, S474X) [9]. At most loci, the associated variants are found in either intronic or intergenic positions. At the 175 SNPs that showed genome-wide significance in a GWAS of ~100,000 individuals [9], only 2 loci had a nonsense variant with strong evidence for association and 52 had a missense variant with strong evidence for association (defined as $r^2 > 0.5$ with the most significant SNP at the locus to allow identification of untyped nonsynonymous variants).

To help identify the functional gene at regions where there are no associated coding variants, bioinformatics and experimental approaches can help prioritize among positional candidate genes. One such approach is to determine whether variants also influence expression levels of nearby genes in expression quantitative trait locus (eQTL) studies. eQTL results in liver

may be most relevant to lipid levels because the liver is where cholesterol, phospholipids and lipoproteins are synthesized and packaged for delivery to the rest of the body, although expression in macrophages and adipose may also be pertinent. One clear eQTL example is rs10468017, an intergenic HDL-associated SNP ($p=4 \times 10^{-93}$) that is also associated with expression levels in liver of *LIPC*, the hepatic lipase gene ($p=7 \times 10^{-23}$) [9]. We hypothesize that rs10468017 either disrupts a binding site for a transcription factor that may lead to decreased expression of *LIPC*, which in turn alters circulating HDL-c levels, or is correlated with a variant that disrupts a transcription factor binding site. Among 95 loci listed as significantly associated with blood lipid levels [9], 32 have significant eQTL results for at least one nearby gene, defined as a gene within 500 kb with a p -value $< 5 \times 10^{-8}$. However, at many lipid-associated loci, eQTL results are either not significant after correcting for the number of genes examined, or suggest a gene with no previously known role in lipid metabolism. To convincingly demonstrate a functional role of these genes, experimental approaches are required.

Mouse experiments that either knock-down or overexpress candidate genes from associated regions have resulted in the identification of several novel lipid-related genes. SNPs in intron 1 of a GalNac transferase (*GALNT2*) were identified as a novel lipid-associated region from GWAS [10], and subsequent knock-down and overexpression of this gene in mouse liver clearly demonstrated that *GALNT2* can influence HDL-c levels [9]. At an LDL-c locus, liver eQTL studies highlighted three nearby genes, *PSRC1*, *CELSR2* and *SORT1*; hepatic overexpression in mouse of each gene identified a role in LDL metabolism of *Sort1* [9]. Molecular biology experiments revealed that the minor allele of rs12740374 creates a transcription factor binding site for C/EBP alpha, which may influence expression of *SORT1* to affect LDL-c levels [22]. Both confirmatory [23] and conflicting reports [24] in other mouse models suggest that further work is needed to clarify the role of this gene in humans [25].

SNPs approximately 40 kb from *TRIB1* are associated with levels of triglycerides, HDL-c and LDL-c [15]. A similar model of hepatic knock-down and overexpression in mouse liver was applied to the candidate gene *Trib1*. These experiments in mouse confirmed a role of *Trib1* in plasma cholesterol, triglyceride levels, and very low density lipoprotein production [26]. These experiments are more extensively discussed in a review [27]. Identifying a functional gene at these associated loci may improve our understanding of the biological mechanisms of cholesterol metabolism and synthesis, and identifying functional genetic variants can reveal the mechanisms by which the variants influence genes and traits.

From GWAS loci to underlying functional variants

One challenge in using GWAS to pinpoint functional variants, or even to create a list of potential functional variants, is the sparseness of markers. Typically, less than 10% of common genomic variation is directly assessed with GWAS panels, and the remainder is indirectly represented by correlated markers. These correlations are often sufficient to discover a region of association. However, a more complete picture of candidate functional variants can be obtained by identifying co-inherited markers in large reference panels of fully-sequenced individuals, such as the publicly available 1000 Genomes Project resource [28]. Further, imputing variants from dense reference panels into samples with GWAS genotypes may lead to the discovery of novel loci not tagged by GWAS panels.

Identifying the functional genetic variants at a locus can increase the estimated proportion of trait variance explained by that locus. This is because imperfect proxies of functional variants detected by relatively sparse GWAS panels likely underestimate the variance explained [29]. Sanna and colleagues demonstrated this difference by performing targeted

exon sequencing of 7 genes in 5 LDL-c-associated regions: *APOE*, *APOC1*, *APOC2*, *SORT1*, *LDLR*, *APOB*, and *PCSK9* [30]. The most strongly associated variants identified from sequencing 256 Sardinians were then genotyped in an additional 5,524 Sardinians and ~10,000 Norwegians and Finns. At *PCSK9*, the GWAS-identified SNP was common (24% allele frequency), located ~10 kb upstream of the transcription start site, and exhibited a modest effect size of 3.7 mg/dL. However, the association signal was determined to be driven by a less common (3.7% allele frequency) coding variant, R46L, that exhibited a larger effect size of 12.9 mg/dL. More complete assessment at each locus, including additional rare and common variants and fine-mapping the functional variant, increased the trait variance explained at these seven genes from 3.1% to 6.5%. Another targeted sequencing study in individuals with hypertriglyceridemia revealed an excess of rare variants in several genes when considered in aggregate: *APOA5*, *GCKR*, *LPL* and *APOB* [31].

Progress in prioritizing functional variants has been aided by the ENCODE project (Encyclopedia of DNA Elements), which attempts to decipher important regulatory regions and transcription factor binding domains in the vast sea of non-coding DNA [32]. Comparisons of the positions of the most significant lipid-associated SNPs to regulatory domains defined by patterns of histone methylation marks revealed an excess of lipid-associated variants located within these domains [33]. Analysis of SNPs located in regulatory elements can focus experimental study of candidate causal variants from amongst large numbers of variants that show similar strength of association at a locus [34]. For example, at the intergenic associated region near *LIPG*, rs7239867 falls in a strong enhancer specific to HepG2 liver carcinoma cell line and appears to overlap with binding sites of transcription factors SP1, HNF4 α , HNF4 γ , and CEBP/ β . Interestingly, rare variants in the proximal promoter or 5' UTR of *LIPG* identified by sequencing individuals with high or low HDL cholesterol appear to impact expression levels using promoter assays [35].

Another approach to distinguish functional variants from nearby variants in linkage disequilibrium is to test for SNP-trait association in non-European individuals. Genotypes may be obtained for a dense set of variants spanning the European association signal. If linkage disequilibrium extends over shorter distances, as expected especially in individuals with African ancestry, then the association signal would be restricted to a subset of the variants [36–38]. One challenge of this approach is that detecting many of the GWAS loci required tens of thousands of individuals, hence many non-European samples may be required to detect sufficient evidence of association to narrow the signals [39]. Failure to replicate an association signal across populations using only the most-significant genetic variant may be due to lack of power or due to differences in linkage disequilibrium.

Sequencing to identify genes and variants

An even more comprehensive approach to pinpoint functional variants is to perform whole genome sequencing in thousands of individuals and then carry out direct tests of association with the identified variants. This approach will allow for nearly complete discovery of variants in associated regions to create a complete catalogue of potential functional variants before expensive functional studies are initiated. Sequencing will allow more low frequency trait-relevant variants (< 5%) to be discovered than is possible from GWAS panels.

The most important use of whole genome sequencing may be the discovery of rare variants that may have large impacts on lipid levels and heart disease risk, however, whole genome sequencing studies are not without challenges of their own. The biggest challenge is clearly the expense, as sequencing a single individual costs \$2,000–\$4,000, whereas genotyping a GWAS chip is ~\$300–\$800. Another issue is the computer hardware needed to store

sequences, map billions of reads, call variants, perform extensive quality control, and finally test tens of millions of genetic variants for evidence of trait or disease association. Error rates in genotype calls from sequence data can be approximately 1–2%, which can be more than 10x higher than typically observed from large-scale genotyping approaches. Lastly, phasing of variants into haplotypes can be more challenging with sequencing data due to higher error rates, complex genotype probabilities and the presence of many rare variants, which are difficult to phase accurately.

The statistical analysis of genotypes derived from sequencing studies is much more complex than that used for GWAS approaches. After 5 years of intense activity in the GWAS arena, the processes of genotype calling, imputation, quality control, statistical association testing and meta-analysis became a well-oiled machine. Well-tested and feature-rich programs were developed for each stage of analysis that became extremely popular and easy to use (e.g. BIRDSUITE [40], IMPUTE [41], MACH [42], PLINK [43] METAL [44], LocusZoom [45]). However, the statistical approaches that worked so well for GWAS studies, such as single marker logistic or linear regression models, may not provide the highest power for identifying rare variants associated with disease. Rare variants may have little power to show association individually, motivating the use of models that allow for the aggregation of potentially functional variants across some pre-specified unit, typically a gene. Many different rare variant burden tests have been proposed [46], but relatively few examples of known genes are available to evaluate these tools. A likely outcome of sequencing studies is that different genes will have different genetic architecture; some gene signals driven by singleton variants, some driven by a collection of intermediate frequency variants or intermediate frequency and rare variants, and some by a single intermediate frequency variant. Some loci may harbor a collection of common variants of very small effect and rare or intermediate frequency variants of much larger effect. Possibly no single burden test will work well for every scenario and we expect that several tests that assume different underlying genetic models will be needed to find all discoverable genetic associations.

A priori prediction of ‘potentially functional’ variants in exons also remains challenging. Several tools have been created that predict a variant’s functionality based on conservation and/or the severity of the amino acid change [47]. An average of scores from several tools has been shown to be slightly more informative than any single score considered alone and allows for a variant to be evaluated in the presence of incomplete prediction data for some of the methods [48].

An alternative to sequencing that allows direct testing of coding variants at low cost is genotyping an array of exome variants. One such product has been developed by Illumina in collaboration with many different sequencing studies that have a total of >12,000 whole genome or exome-sequenced individuals. This array will allow 247,870 coding variants to be genotyped for association analysis with lipid levels and cardiovascular disease in much larger sample sizes than can be sequenced at current costs. The array is expected to cover > 99% of variants present in Europeans at > 0.1% frequency. Based on results from genotyping this chip in an anticipated ~1 million individuals, we anticipate a flurry of novel association results that implicate functional genes by identifying association with coding variants.

Toward personalized treatment

The goal of genetic studies is to translate these scientific gains into the clinic, by providing novel genes as drug targets, allowing personalized treatment based on genetics, or perhaps predicting disease risk to target preventive medicine approaches. Currently, genetic efforts are focused primarily on the first goal, which requires determining the underlying genes

using functional approaches, often in model organisms. Personalized medicine based on genetic factors will require studies of genes that impact drug metabolism and response, an active area of research that will likely yield many variants of large effect. Based on the current sets of variants known to be associated with lipid levels, the effect sizes are generally too small, even when considered in aggregate, to enable accurate prediction of at-risk individuals. The use of known risk factors such as age, sex, body mass index, and family history of cardiovascular disease provide much more information than genetic factors alone.

Conclusion

GWAS have identified many loci that will harbor genes relevant to the biology of lipid levels. Experiments in mouse have identified functional genes (*GALNT2*, *TRIB1*, *SORT1*), creating potential new therapeutic targets for decreasing the prevalence of heart disease. At the *SORT1* locus, a functional mechanism has been described by which a variant creates a transcription factor binding site that in turn influences gene expression. Until recently, most genome-wide efforts have used genotyping arrays and imputation to assay most of the common variation across the genome. Recent technological advances have enabled whole genome sequencing approaches, which hold the promise of discovery of novel rare variants with large effects on lipid levels and heart disease risk.

Acknowledgments

CJW and KLM are supported by the National Institutes of Health (HL94535, HL109946, DK72193, DK78150, and DK93757).

References

1. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics--2011 update: a report from the American Heart Association. *Circulation*. 2011; 123:e18–e209. [PubMed: 21160056]
2. Pilia G, Chen WM, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet*. 2006; 2:e132. [PubMed: 16934002]
3. Saxena R, Hivert MF, Langenberg C, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet*. 2010; 42:142–8. [PubMed: 20081857]
4. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010; 42:105–16. [PubMed: 20081858]
5. Strawbridge RJ, Dupuis J, Prokopenko I, et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes*. 2011; 60:2624–34. [PubMed: 21873549]
6. Ehret GB, Munroe PB, et al. International Consortium for Blood Pressure Genome-Wide Association S. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011; 478:103–9. [PubMed: 21909115]
7. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010; 42:937–48. [PubMed: 20935630]
8. Heid IM, Jackson AU, Randall JC, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet*. 2010; 42:949–60. [PubMed: 20935629]
- 9**. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010; 466:707–13. This is the largest published GWAS of lipid traits and includes follow-up of several functional genes. [PubMed: 20686565]
10. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*. 2008; 40:161–9. [PubMed: 18193043]

11. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet.* 2008; 40:189–97. [PubMed: 18193044]
12. Sandhu MS, Waterworth DM, Debenham SL, et al. LDL-cholesterol concentrations: a genome-wide association study. *Lancet.* 2008; 371:483–91. [PubMed: 18262040]
13. Chasman DI, Pare G, Zee RY, et al. Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and Apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ Cardiovasc Genet.* 2008; 1:21–30. [PubMed: 19802338]
14. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* 2009; 41:47–55. [PubMed: 19060911]
15. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* 2009; 41:56–65. [PubMed: 19060906]
16. Hicks AA, Pramstaller PP, Johansson A, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* 2009; 5:e1000672. [PubMed: 19798445]
17. Chasman DI, Pare G, Mora S, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet.* 2009; 5:e1000730. [PubMed: 19936222]
- 18*. Lemaitre RN, Tanaka T, Tang W, et al. Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet.* 2011; 7:e1002193. This is a GWAS for polyunsaturated fatty acids. [PubMed: 21829377]
- 19*. Kim YJ, Go MJ, Hu C, et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet.* 2011; 43:990–5. This large GWAS for lipid levels and other traits in Asians identified several novel loci not discovered in studies of Europeans. [PubMed: 21909109]
20. Hegele RA. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet.* 2009; 10:109–21. [PubMed: 19139765]
21. Baigent C, Keech A, Kearney PM, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet.* 2005; 366:1267–78. [PubMed: 16214597]
- 22**. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature.* 2010; 466:714–9. This paper is the first to report a functional variant underlying an LDL-c GWAS signal. [PubMed: 20686566]
- 23*. Linsel-Nitschke P, Heeren J, Aherrahrou Z, et al. Genetic variation at chromosome 1p13.3 affects sortilin mRNA expression, cellular LDL-uptake and serum LDL levels which translates to the risk of coronary artery disease. *Atherosclerosis.* 2010; 208:183–9. The authors identify a functional gene at the *SORT1* locus by overexpression in transfected cells. [PubMed: 19660754]
- 24*. Kjolby M, Andersen OM, Breiderhoff T, et al. Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metab.* 2010; 12:213–23. The authors created a mouse knock-out of *Sort1* and identified a biological mechanism of gene function. [PubMed: 20816088]
- 25*. Dube JB, Johansen CT, Hegele RA. Sortilin: an unusual suspect in cholesterol metabolism: from GWAS identification to in vivo biochemical analyses, sortilin has been identified as a novel mediator of human lipoprotein metabolism. *Bioessays.* 2011; 33:430–7. This review summarizes the differences between experiments to understand the role of *SORT1* in lipid biology and potential reasons for the conflicting results observed. [PubMed: 21462369]
- 26**. Burkhardt R, Toh SA, Lagor WR, et al. Trib1 is a lipid- and myocardial infarction-associated gene that regulates hepatic lipogenesis and VLDL production in mice. *J Clin Invest.* 2010; 120:4410–4. The authors identified a functional gene and biological mechanism at the *TRIB1* GWAS locus. [PubMed: 21084752]
- 27*. Bauer RC, Stylianou IM, Rader DJ. Functional validation of new pathways in lipoprotein metabolism identified by human genetics. *Curr Opin Lipidol.* 2011; 22:123–8. This review summarizes the functional follow-up of several lipid GWAS loci. [PubMed: 21311327]

- 28*. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–73. This paper describes an extremely important contribution to the human genetics community provided by sequencing diverse samples to characterize genetic variation across the genome. [PubMed: 20981092]
29. Nica AC, Parts L, Glass D, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet*. 2011; 7:e1002003. [PubMed: 21304890]
- 30*. Sanna S, Li B, Mulas A, et al. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet*. 2011; 7:e1002198. This study describes fine-mapping GWAS loci using sequencing of seven genes. [PubMed: 21829380]
- 31*. Johansen CT, Wang J, Lanktree MB, et al. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet*. 2010; 42:684–7. Resequencing of individuals with extreme hypertriglyceridemia revealed an excess of rare variants in genes identified by GWAS for TG. [PubMed: 20657596]
- 32*. Myers RM, Stamatoiyannopoulos J, et al. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol*. 2011; 9:e1001046. This paper provides an introduction to the diverse set of ENCODE resources to annotate the non-coding part of the genome. [PubMed: 21526222]
- 33*. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011; 473:43–9. This paper describes patterns of histone methylation marks that characterize functional domains. [PubMed: 21441907]
- 34 *. 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*. 2011 Epub Nov 7, 2011 HaploReg is a useful tool to evaluate SNPs for overlap with histone methylation marks, conservation and regulatory motifs.
- 35*. Khetarpal SA, Edmondson AC, Raghavan A, et al. Mining the LIPG Allelic Spectrum Reveals the Contribution of Rare and Common Regulatory Variants to HDL Cholesterol. *PLoS Genet*. 2011; 7:e1002393. This study ascertains rare variants in individuals with extreme HDL-c levels and characterizes their impact on expression. [PubMed: 22174694]
36. McKenzie CA, Abecasis GR, Keavney B, et al. Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I-converting enzyme (ACE). *Hum Mol Genet*. 2001; 10:1077–84. [PubMed: 11331618]
37. Frere C, Tregouet DA, Morange PE, et al. Fine mapping of quantitative trait nucleotides underlying thrombin-activatable fibrinolysis inhibitor antigen levels by a transethnic study. *Blood*. 2006; 108:1562–8. [PubMed: 16705091]
38. N'Diaye A, Chen GK, Palmer CD, et al. Identification, replication, and fine-mapping of Loci associated with adult height in individuals of african ancestry. *PLoS Genet*. 2011; 7:e1002298. [PubMed: 21998595]
- 39*. Keebler ME, Deo RC, Surti A, et al. Fine-mapping in African Americans of 8 recently discovered genetic loci for plasma lipids: the Jackson Heart Study. *Circ Cardiovasc Genet*. 2010; 3:358–64. Genotyping-based fine-mapping in African-Americans at eight lipid loci refined the association signal at three loci. [PubMed: 20570916]
40. Korn JM, Kuruvilla FG, McCarroll SA, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet*. 2008; 40:1253–60. [PubMed: 18776909]
41. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009; 5:e1000529. [PubMed: 19543373]
42. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet*. 2009; 10:387–406. [PubMed: 19715440]
43. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–75. [PubMed: 17701901]
44. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26:2190–1. [PubMed: 20616382]

45. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26:2336–7. [PubMed: 20634204]
46. Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet*. 2010; 11:773–85. [PubMed: 20940738]
- 47*. Cooper GM, Shendure J. Needles in stacks of needles: finding disease causal variants in a wealth of genomic data. *Nat Rev Gen*. 2011; 12:628–40. This review describes recent advances that enable annotation of coding and non-coding variants.
- 48*. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet*. 2011; 88:440–9. This paper introduces the idea of an average of functional prediction scores, an improvement over any single method. [PubMed: 21457909]

Key points

Genome-wide association scans have successfully identified novel loci associated with blood lipid traits.

Functional follow-up of novel loci for blood lipid levels have identified functional genes and in some cases, functional variants.

The next steps include discovering more loci by targeting rare variants with whole genome sequencing, coding variants with exome genotyping arrays, and common variants with more diverse samples.