Published in final edited form as:

Curr Opin Genet Dev. 2015 August ; 33: 25–31. doi:10.1016/j.gde.2015.06.008.

Insights into blood lipids from rare variant discovery

Ellen M. Schmidt¹ and Cristen J. Willer^{1,2,3}

¹ Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

² Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI 48109, USA

³ Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Large-scale genome wide screens have discovered over 160 common variants associated with plasma lipids, which are risk factors often linked to heart disease. A large fraction of lipid heritability remains unexplained, and it is hypothesized that rare variants of functional consequence may account for some of the missing heritability. Finding lipid-associated variants that occur less frequently in the human population poses a challenge, primarily due to lack of power. Interrogation of the protein-coding regions of the genome using array and sequencing techniques has led to important discoveries of rare variants that affect lipid levels and related disease risk. Here, we summarize the latest methods and findings that contribute to our current understanding of rare variant lipid genetics.

Introduction

Cardiovascular disease is one of the leading causes of death in the U.S. and throughout the world, representing a significant human health burden [1]. Genetic studies of lipid levels, known risk factors for heart disease with heritability ranging from 40% - 60% in humans [2], are logical targets in efforts to prevent and treat heart disease. Modulation of these quantitative lipid traits, which include low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), and total cholesterol (TC), can be effective therapeutically. Lipid traits are not independent, making it challenging to untangle the effects of specific lipids on disease risk. LDL-C and TC generally act in the same direction since the majority of TC is composed of LDL-C. In addition, high TG is associated with high LDL-C and low HDL-C, while LDL-C and HDL-C are positively and inversely associated with heart disease risk, respectively [3].

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

URLs for public data repositories

NHLBI Exome Sequencing Project (ESP) Exome Variant Server: <http://evs.gs.washington.edu/EVS/>; NHGRI ENCyclopedia Of DNA Elements (ENCODE): <http://www.genome.gov/encode/>; database of Genotypes and Phenotypes (dbGAP): <http://www.ncbi.nlm.nih.gov/gap>

We can harness the information from single nucleotide polymorphisms (SNPs) in the human population to gain knowledge about the genetic contribution to plasma lipids. Early genome wide association studies (GWAS) with modest sample sizes ($< 10,000$) uncovered common variants ($MAF > 5\%$) with large effect sizes. The power with which to discover new lipid-associated variants increases with sample size [4-6]. In the largest lipids GWAS to date, researchers jointly meta-analyzed data from a total sample size of $\sim 180,000$, to uncover 157 independent genetic loci containing lipid-associated variants [7]. Most of the associated variants are non-protein-coding however, suggesting a regulatory role [8]. An illustration of the regulatory role of a noncoding LDL-C-associated variant is evident at the SORT1 locus. Researchers demonstrated in human-derived hepatocytes that variant rs12740374 at an LDL-C-associated GWAS locus creates a C/EBP transcription factor binding site, causing altered expression of the nearby SORT1 gene [9]. Still, common variants generally have limited functional consequence. Understanding the underlying biology of noncoding variation and translating these finding into clinical practice remains a universal challenge.

Common variants identified by GWAS explain only a fraction (10-15%) of the trait variance for lipids [6]. In effort to explain some of the ‘missing heritability’, we turn to rare ($MAF < 0.05\%$) and low frequency ($0.5\% \leq MAF < 5\%$) variation. Protein-coding variants with deleterious function are likely to be rarer in the human population due to natural selection acting against them, and have mostly arisen recently in evolutionary history [10]. The Common Disease-Rare Variant hypothesis proposes that the combined effect of a number of low frequency variants with large effect sizes accounts for some of the missing heritability [11]. Indeed, early sequencing studies of candidate genes supported the contribution of multiple rare alleles to plasma HDL-C levels [12]. This rare variation cannot be found with traditional GWAS single-variant approaches due to poor coverage on arrays and difficulties with imputing, thus opening up a niche to investigators dedicated to rare variant discovery.

Mendelian family studies involving large pedigrees are valuable for rare variant genetic studies. In this design, the co-segregation of variants among affected family members can be traced. Investigators recently used whole-exome sequencing in a multi-generation family to uncover a rare variant in a highly conserved codon of SLC25A40 that is associated with TG, giving insight into a previously unknown biological mechanism of Hypertriglyceridemia [13]. In this review, our subsequent focus will be on the rare variant findings from large-scale array and sequencing studies for complex lipid traits.

Methods for rare variant association testing

Because of the rare nature of most variants with functional consequence, studies carefully designed to uncover rare variant associations are critical [14, 15]. Recent advances in exome sequencing and exome array technologies have facilitated larger and more accurate studies for interrogating the protein-coding 1%-2% of the genome. However, single variant association tests commonly used by GWAS carry a heavy multiple testing burden and still lack power when applied to rare variants of high impact. Thus, aggregation-based tests that group multiple variants by a single gene or functional unit have become standard for rare variant association analysis.

Several regression-based approaches have been developed in recent years to optimize rare variant discovery. In a simple burden test, multiple rare variants are collapsed into a genetic score representing the cumulative effect of those variants in a single unit. Then, the score is tested for association with a trait or disease. This idea, which has been implemented by numerous investigators [16-20], assumes that all variants in a single unit are causal and that all alleles affect the phenotype with the same magnitude and direction of effect. More robust modifications of a simple burden test introduce adaptive weights or thresholds [21-26]. We can account for the protective or deleterious impact of alleles on phenotype by considering the magnitude and direction of effect using variance-component tests [27-29]. Finally, we can combine burden and component tests [30-32] or score statistics [33] to achieve more robust power. Choosing the optimal strategy for grouping rare variants is flexible and may depend on the genetic architecture of a particular trait [34].

Meta-analysis is a powerful tool to jointly analyze datasets from multiple studies, especially when individual-level data are not available. Traditional Fisher's method [35] for combining P-values is straight-forward, but does not weight by sample size or consider magnitude or direction of effect. When studies provide score statistics for each variant and a variance-covariance matrix, a fixed-effects meta-analysis will achieve improved power [36-38]. Both genotyping and sequencing-based replication strategies are appropriate for confirmation of rare variant association findings [39].

Rare coding variation implicated in lipids

Interrogating the protein-coding genome through re-sequencing coding regions [40] and whole-exome sequencing [41] is hypothesized to reveal rare mutations with a large effect on phenotype. A splice variant in APOC3 associated with TG was identified using whole-genome sequencing, representing one of the first rare variants of large effect to be found using this sequencing approach at the population scale [42]. Sequencing the exome is more cost-effective than whole-genome sequencing, however, and fewer statistical tests are performed, reducing the multiple testing burden. The potential of exome sequencing has resulted in studies powered for the discovery of novel rare variation implicated in blood lipids. For example, investigators of the NHLBI Exome Sequencing Project (ESP) used exome sequencing to identify the burden of rare variants in four genes (PNPLA5, PCSK9, LDLR and APOB) significantly associated with LDL-C [43] (**Table 1**). In contrast to modest effect sizes observed from individual SNPs identified by GWAS, the burdens of rare variants in these genes have substantially higher effect sizes (**Figure 1**).

Although it is clear that increasingly large sample sizes will improve power in rare variant discovery, this raises the question of how many individuals need to be sequenced to assess the true impact of rare variation on lipid traits and related diseases. LDL-C is a particularly heritable quantitative trait compared to others, and associated variants have relatively high effect sizes. Given the genetic architecture of lipids, more than just a few thousand samples will likely need to be sequenced to comprehend the full impact of rare variation.

Additionally exome chip, a custom genotyping array, allows for large-scale efficient genotyping of low frequency coding variants with large effect sizes. Exome wide association

studies for lipids and related diseases revealed several significant variants at both established and previously unknown lipid loci. Rare variants at ANGPTL4, LIPC and LIPG, for example, were found to be associated with TG and HDL-C [44] (**Table 1**). In addition, a more common variant in the protein-coding gene, TM6SF2, was found to be associated with total cholesterol and myocardial infarction risk [44, 45]. Functional follow-up revealed that modulation of Tm6sf2 in mice alters lipid levels, providing the causal gene at a GWAS locus that was previously intractable for follow-up due to a large number of genes in the associated region. These findings illustrate the importance of interrogating changes in the exome in guiding our knowledge of the functional gene at lipid loci.

Translation of plasma lipid levels to disease risk

Understanding the relationship between plasma lipid concentrations and disease risk is paramount in human health interest. Several lipid loci contain variants associated with risk for diseases such as coronary artery disease (CAD) [46], highlighting the causal role of blood lipids on CAD. Although the correlation between increased triglyceride levels and increased risk of CAD is well established [47], the causality of this relationship is a separate question that has claimed recent attention. After adjusting for the effects of both LDL-C and HDL-C, investigators found correlation between the effect size of TG-associated SNPs and the magnitude of the effect on CAD risk, suggesting causality [48]. Exome sequencing revealed a set of rare variants, including one missense and 3 loss-of-function driver mutations in APOC3 associated with low plasma TG [49]. Carriers of such mutations showed significantly reduced risk for CAD.

Although the causal relationship between low LDL-C and reduced risk for myocardial infarction (MI) is well established and used in treatment [50], the causality of high HDL-C with a similar outcome has been challenged [51]. A Mendelian randomization study ($n > 25,000$ participants) involving a low frequency variant p.Asn396Ser (2.6%) in LIPG that is associated with high HDL-C showed that carriers did not have a significantly reduced risk for MI [51]. This questions the utility of high plasma HDL-C to either predict or treat heart disease. Researchers sequenced the exons of the NPC1L1 gene and found naturally occurring mutations that can mimic the activity of an LDL-C lowering drug and thus reduce coronary heart disease (CHD) risk [52]. Low frequency variants in PCSK9 are also known to be associated with low LDL-C and reduced CHD risk [53-55]. This prompted the interrogation of rare coding variation to find similar occurrences elsewhere in the genome. Using an exome array, researchers identified four low-frequency variants associated with HDL-C and TG levels, but not CHD [56] (**Table 1**). Undoubtedly, studies of heritable rare variation such as these will be instrumental in guiding physicians toward appropriate risk prevention, diagnosis, and treatment of cardiovascular disease.

Understanding the function of rare noncoding lipid-associated variation

Because we expect that rare variants of large effects are more likely to occur in coding regions, detection of rare noncoding variants with comparable effect sizes will generally require much larger sample sizes. It is thought that pooling for rare variant discovery using both exonic and non-exonic rare variants together is more powerful, and that the increase in

signal that results from including noncoding risk variants outweighs the noise of noncoding neutral variants [57]. Whole-genome sequencing of nearly 1000 individuals revealed more about the genetic architecture of HDL-C [58]. Morrison et al. found that common variation explains a whole 61.8% of the trait heritability for HDL-C, and individuals with extreme HDL-C harbor rare variants with large effect sizes [58]. [42]

Array designs with densely genotyped regions such as the Metabochip [59] allow us to finely interrogate genetic loci of interest. This fine mapping can guide discovery of rare functional variants that were not previously interrogated and help narrow the association signal [60, 61]. Fine mapping on Metabochip in Europeans highlighted the LDL-C associated rare R46L variant (allele frequency 0.03) at PCSK9 that differs from the GWAS signal [7]. In addition, methods to evaluate enrichment of noncoding rare variants in regulatory regions of the genome will give insight into the biological mechanisms involved and help prioritize rare functional variants [62].

Discussion

New methods to detect rare variation associations in combination with evolving sequencing and array technologies have advanced our understanding of the role of rare variation in determining lipid levels in recent years. A central objective for the study of rare variants is to explain some remaining fraction of complex trait heritability and inform about the biology of the trait or disease. As individual variants, it is not likely that there exists a multitude of low-frequency and rare variants that alone have a robust influence on lipid levels and heart disease risk. However, we have seen that when considered in aggregate, multiple rare variants at a single locus can explain a substantial portion of the genetic heritability for lipid levels and subsequent disease risk [63]. Increasing sample sizes will increase our power for detecting rare variant associations to some extent, but it is unknown how many of these variants exist (**Figure 1**). Although studies of rare variation are informative, there still remains a portion of the heritability that is unexplained.

A majority of our understanding of lipid genetics is derived from studies of European participants. Lipid studies of underrepresented groups, with the help of customized arrays and larger reference panels for imputation, will provide guidance toward more patient-specific treatment. In addition, meta-analysis of data from various cohorts will facilitate detection of population-specific rare variant associations. Population-specific rare variants do contribute to phenotype differences (e.g. African-specific LDL-C-associated variants in PCSK9 and HDL-C-associated variants in the APOA/APOC gene cluster) [64]. In addition, an exome chip study in a Greek population isolate revealed a rare variant in APOC3 associated with increased HDL-C and decreased TG [65].

Burden tests have proven powerful in detecting rare variant associations for lipids. There is still ambiguity in choosing the most effective way to group variants, especially when the trait genetic architecture is unclear. We have found that different genes harbor associated variants that differ by frequency and function. Although challenges still remain in translating genetic findings to medical applications, research advances in recent years have proven successful in furthering our understanding of lipid etiology and heart disease risk.

Acknowledgments

This work has been supported by the National Institutes of Health (grant numbers: HL094535 and HL109946) (CJW) and by the National Science Foundation Open Data Integrative Graduate Education and Research Traineeship (IGERT) grant 0903629 (EMS).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Dai S, Ford ES, Fox CS, Franco S, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. *Circulation*. 2014; 129:e28–e292. [PubMed: 24352519]
2. Weiss LA, Pan L, Abney M, Ober C. The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet*. 2006; 38:218–222. [PubMed: 16429159]
3. Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, Thompson A, Wood AM, Lewington S, Sattar N, Packard CJ, et al. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA*. 2009; 302:1993–2000. [PubMed: 19903920]
4. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*. 2008; 40:161–169. [PubMed: 18193043]
5. Kathiresan, S.; Melander, O.; Guiducci, C.; Surti, A.; Burt, NP.; Rieder, MJ.; Cooper, GM.; Roos, C.; Voight, BF.; Havulinna, AS., et al. *Nat Genet*. Vol. 40. 189-197: 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans..
6. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010; 466:707–713. [PubMed: 20686565]
- 7**. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013; 45:1274–1283. [PubMed: 24097068] [This is a large-scale meta-analysis of previous GWAS cohorts and newly genotyped individuals on the custom MetaboChIP (total n > 180,000). It reports 157 independent lipid-associated loci as well as numerous bioinformatics follow-up analyses to explain the genetic variation underlying biology. This work is the largest lipids association study to-date, representing studies from all over the world as part of the Global Lipids Genetics Consortium.]
8. Hindorf, LA.; MacArthur, J.; Morales, J.; Jenkins, HA.; Hall, PN.; Klemm, AK.; Manolio, TA. A Catalog of Published Genome-Wide Association Studies. 2010. www.genome.gov/gwastudies
9. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*. 2010; 466:714–719. [PubMed: 20686566]
10. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*. 2013; 493:216–220. [PubMed: 23201682]
11. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet*. 2002; 11:2417–2423. [PubMed: 12351577]
12. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, H. H. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*. 2004; 305:869–872. [PubMed: 15297675]

13. Rosenthal EA, Ranchalis J, Crosslin DR, Burt A, Brunzell JD, Motulsky AG, Nickerson DA, NHLBI GO Exome Sequencing Project. Wijsman EM, Jarvik GP. Joint linkage and association analysis with exome sequence data implicates SLC25A40 in hypertriglyceridemia. *Am J Hum Genet.* 2013; 93:1035–1045. [PubMed: 24268658]
14. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES. Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci U S A.* 2014; 111:E455–464. [PubMed: 24443550]
- 15**. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014; 95:5–23. [PubMed: 24995866] [This is a comprehensive review of analytic issues involved in rare variant association studies including study design and association testing methods.]
16. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res.* 2007; 615:28–56. [PubMed: 17101154]
17. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet.* 2008; 83:311–321. [PubMed: 18691683]
18. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 2009; 5:e1000384. [PubMed: 19214210]
19. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol.* 2010; 34:188–193. [PubMed: 19810025]
20. Asimit JL, Day-Williams AG, Morris AP, Zeggini E. ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. *Hum Hered.* 2012; 73:84–94. [PubMed: 22441326]
21. Han F, Pan W. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered.* 2010; 70:42–54. [PubMed: 20413981]
22. Hoffmann TJ, Marini NJ, Witte JS. Comprehensive approach to analyzing rare genetic variants. *PLoS One.* 2010; 5:e13584. [PubMed: 21072163]
23. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet.* 2010; 86:832–838. [PubMed: 20471002]
24. Liu DJ, Leal SM. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet.* 2010; 6:e1001156. [PubMed: 20976247]
25. Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet.* 2011; 89:354–367. [PubMed: 21885029]
26. Ionita-Laza I, Buxbaum JD, Laird NM, Lange C. A new testing strategy to identify rare variants with either risk or protective effect on disease. *PLoS Genet.* 2011; 7:e1001289. [PubMed: 21304886]
27. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* 2011; 89:82–93. [PubMed: 21737059]
28. Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol.* 2009; 33:497–507. [PubMed: 19170135]
29. Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. Testing for an unusual distribution of rare variants. *PLoS Genet.* 2011; 7:e1001322. [PubMed: 21408211]
- 30*. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics.* 2012; 13:762–775. [PubMed: 22699862] [This paper introduces a widely used method for rare variant association testing called the optimal sequence kernel association test (SKAT-O). It is optimal for scenarios in which the underlying biological mechanisms of genomic variation are unknown and vary across the genome. This test outperforms burden tests in a wide range of scenarios, making it especially efficient for modern studies using massively parallel sequencing.]

31. Derkach A, Lawless JF, Sun L. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet Epidemiol.* 2013; 37:110–121. [PubMed: 23032573]
32. Sun J, Zheng Y, Hsu L. A unified mixed-effects model for rare-variant association in sequencing studies. *Genet Epidemiol.* 2013; 37:334–344. [PubMed: 23483651]
33. Chen LS, Hsu L, Gamazon ER, Cox NJ, Nicolae DL. An exponential combination procedure for set-based association tests in sequencing studies. *Am J Hum Genet.* 2012; 91:977–986. [PubMed: 23159251]
34. Ladouceur M, Dastani Z, Aulchenko YS, Greenwood CM, Richards JB. The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals. *PLoS Genet.* 2012; 8:e1002496. [PubMed: 22319458]
35. Fisher, RA.; Genetiker, S.; Fisher, RA.; Genetician, S.; Britain, G.; Fisher, RA.; Généticien, S. Statistical methods for research workers. Oliver and Boyd; Edinburgh: 1970.
- 36*. Liu DJ, Peloso GM, Zhan X, Holmen OL, Zawistowski M, Feng S, Nikpay M, Auer PL, Goel A, Zhang H, et al. Meta-analysis of gene-level tests for rare variant association. *Nat Genet.* 2014; 46:200–204. [PubMed: 24336170] [The authors introduce a novel approach for meta-analysis of rare variant associations tests. The method is demonstrated using simulation and empirically in a meta-analysis of lipids in > 18,000 individuals genotyped on an exome array. This meta-analysis approach is as powerful as more cumbersome analyses requiring individual-level data.]
37. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet.* 2013; 93:42–53. [PubMed: 23768515]
38. Hu YJ, Berndt SI, Gustafsson S, Ganna A, Genetic Investigation of ATC. Hirschhorn J, North KE, Ingelsson E, Lin DY. Meta-analysis of gene-level associations for rare variants based on single-variant statistics. *Am J Hum Genet.* 2013; 93:236–248. [PubMed: 23891470]
39. Liu DJ, Leal SM. Replication strategies for rare variant complex trait association studies via next-generation sequencing. *Am J Hum Genet.* 2010; 87:790–801. [PubMed: 21129725]
40. Kryukov GV, Shpunt A, Stamatoyannopoulos JA, Sunyaev SR. Power of deep, all-exon resequencing for discovery of human trait genes. *Proc Natl Acad Sci.* 2009; 106:3871–3876. [PubMed: 19202052]
41. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum Mol Genet.* 2012; 21:R1–9. [PubMed: 22983955]
42. Timpson NJ, Walter K, Min JL, Tachmazidou I, Malerba G, Shin SY, Chen L, Futema M, Southam L, Iotchkova V, et al. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nat Commun.* 2014; 5:4871. [PubMed: 25225788]
- 43*. Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, Bizon C, Lange EM, Smith JD, Turner EH, et al. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet.* 2014; 94:233–245. [PubMed: 24507775] [This is a large whole-exome sequencing study (n = 2,005) that discovered the burden of rare variants in a novel gene *PNPLA5* to be associated with LDL-C, as well as novel associated variants in 3 other lipid-related genes (*PCSK9*, *LDLR*, and *APOB*). The results highlight the range of genetic architecture underlying these associations in terms of allele frequency and predictive function, necessitating a variety of burden approaches.]
44. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, Guo Y, Zhang J, Langhammer A, Lochen ML, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet.* 2014; 46:345–351. [PubMed: 24633158]
45. Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjaerg-Hansen A, Vogt TF, Hobbs HH, Cohen JC. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet.* 2014; 46:352–356. [PubMed: 24531328]
46. CARDIoGRAMplusC4D Consortium. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* 2013; 45:25–33. [PubMed: 23202125]

47. Sarwar N, Danesh J, Eiriksdottir G, Sigurdsson G, Wareham N, Bingham S, Boekholdt SM, Khaw KT, Gudnason V. Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation*. 2007; 115:450–458. [PubMed: 17190864]
48. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*. 2013; 45:1345–1352. [PubMed: 24097064]
- 49*. TG and HDL Working Group of the Exome Sequencing Project, National Heart Lung and Blood Institute. Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitzel NO, Lange LA, Lu Y, Tang ZZ, et al. Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. *N Engl J Med*. 2014; 371:22–31. [PubMed: 24941081] [Using exome sequencing in 3,734 individuals, this study identified rare mutations in *APOC3* associated with lower triglyceride levels. Mutations include three loss-of-function and one missense mutation. Mutation carriers showed reduced risk for coronary artery disease, which supports the known positive correlation between triglycerides and risk of heart disease.]
50. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, Kirby A, Sourjina T, Peto R, Collins R, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *The Lancet*. 2005; 366:1267–1278.
- 51*. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, Hindy G, Hólm H, Ding EL, Johnson T, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012; 380:572–580. [PubMed: 22607825] [Here the authors use mendelian randomization to test the causal relationship between HDL-C and reduced risk for myocardial infarction (MI). After testing both a loss-of-function coding SNP in *LIPG* associated exclusively with HDL-C as well as a genetic score from 14 common SNPs associated exclusively with HDL-C, they found that genetically raised HDL-C is not associated with lower MI risk. These findings are important because they refute the idea that therapeutics involving raising HDL-C reduces MI risk.]
- 52*. Myocardial Infarction Genetics Consortium Investigators. Stitzel NO, Won HH, Morrison AC, Peloso GM, Do R, Lange LA, Fontanillas P, Gupta N, Duga S, et al. Inactivating mutations in *NPC1L1* and protection from coronary heart disease. *N Engl J Med*. 2014; 371:2072–2082. [PubMed: 25390462] [Exons of *NPC1L1*, for which the encoded protein is a known drug target for lowering LDL-C, were sequenced in coronary artery disease (CAD) cases and controls to test for association between inactivating carriers and LDL-C levels and CAD risk. Carriers of naturally occurring mutations were associated with reduced LDL-C and relative reduction in CAD risk, suggesting the utility of therapies focused on inhibition of *NPC1L1* function.]
53. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence Variations in *PCSK9*, Low LDL, and Protection against Coronary Heart Disease. *N Engl J Med*. 2006; 354:1264–1272. [PubMed: 16554528]
54. Cohen JC, Hobbs HH. Simple Genetics for a Complex Disease. *Science*. 2013; 340:689–690. [PubMed: 23661745]
55. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nat Genet*. 2005; 37:161–165. [PubMed: 15654334]
- 56*. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitzel NO, Brody JA, Khetarpal SA, Crosby JR, Fornage M, et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet*. 2014; 94:223–232. [PubMed: 24507774] [This is a large exome array effort that identified four low-frequency variants with large effects on HDL-C and/or triglycerides. Notably, none of these variants were associated with risk for coronary heart disease.]
57. Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. *Am J Hum Genet*. 2010; 87:604–617. [PubMed: 21070896]

58. Morrison AC, Voorman A, Johnson AD, Liu X, Yu J, Li A, Muzny D, Yu F, Rice K, Zhu C, et al. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat Genet.* 2013; 45:899–901. [PubMed: 23770607]
59. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, Burt NP, Fuchsberger C, Li Y, Erdmann J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* 2012; 8:e1002793. [PubMed: 22876189]
- 60*. Wu Y, Waite LL, Jackson AU, Sheu WH, Buyske S, Absher D, Arnett DK, Boerwinkle E, Bonnycastle LL, Carty CL, et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet.* 2013; 9:e1003379. [PubMed: 23555291] [The authors fine-mapped 58 lipid-associated loci on Metabochip in individuals from African American, East Asian, and European ancestries. Many loci exhibited two more more signals, accounting for an increase in the phenotypic variance explained. The analysis narrowed the association signal at some loci, limiting the number of candidate variants, and allowed for identification of population-specific variants.]
61. Sanna S, Li B, Mulas A, Sidore C, Kang HM, Jackson AU, Piras MG, Usala G, Maninchedda G, Sassu 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. [PubMed: 21829380]
62. Lo KS, Vadlamudi S, Fogarty MP, Mohlke KL, Lettre G. Strategies to fine-map genetic associations with lipid levels by combining epigenomic annotations and liver-specific transcription profiles. *Genomics.* 2014; 104:105–112. [PubMed: 24997396]
63. Tada H, Won HH, Melander O, Yang J, Peloso GM, S K. Multiple Associated Variants Increase the Heritability Explained for Plasma Lipids and Coronary Artery Disease. *Circ Cardiovasc Genet.* 2014; 7:583–587. [PubMed: 25170055]
64. Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA, Ochs-Balcom HM, Donlon TA, Martin LW, Eaton CB, et al. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. *Am J Hum Genet.* 2013; 92:904–916. [PubMed: 23726366]
65. Tachmazidou I, Dedoussis G, Southam L, Farmaki AE, Ritchie GR, Xifara DK, Matchan A, Hatzikotoulas K, Rayner NW, Chen Y, et al. A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. *Nat Commun.* 2013; 4:2872. [PubMed: 24343240]

Figure 1. Power to detect lipid-associated loci for different study designs

Dotted lines represent 80% power curves and points represent lipid-associated loci from exome sequencing [43] (red), exome chip [44] (blue), and GWAS [7] (black) study designs. Colored loci represent standardized effect sizes and burden frequencies estimated from European American samples [43] (red) and Norwegian samples [44] (blue). Abbreviations are as follows: LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TG, triglycerides; TC, total cholesterol; SD, standard deviation units.

Table 1

Contribution of rare and low frequency variation to population variation in lipid levels

Locus	Variant Annotation	Variant ID	Variant Type	Single Variant		Trait	Ethnicity	Reference
				MAF (%)	Effect Size			
<i>ANGPTL4</i>	p.Glu40Lys (E40K)	rs116843064	missense	2.9	-26.9	TG	Norwegian	Holmen <i>et al.</i> 2014 [44]
<i>ANGPTL8</i>	c.361C>T (p.Gln121 *)	rs145464906	nonsense	0.01, 0.1	10 mg/dL	HDL-C	AA, EA	Peloso <i>et al.</i> 2014 [56]
<i>APOC3</i>	IVS2+1G→A	rs138326449	splicing	0.25	-1.43	TG	British	Timpson <i>et al.</i> 2014 [42]
<i>APOC3</i>	p.Arg19* (R19X)	rs76353203	nonsense	1.9	1.471, -0.513	HDL-C, TG	Greek	Tachmazidou <i>et al.</i> 2013 [65]
<i>C3orf14</i>	p.Leu33Met (L33M)	chr3:62307648	missense	0.06, 0.2	0.648	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>COL14A1</i>	p.Alal197Thr (A1197T)	chr8:121292281	missense	0.06, 0.1	0.702	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>COL18A1</i>	c.331G>A (p.Gly111Arg)	rs114139997	missense	1.9, 0.003	-16%	TG	AA, EA	Peloso <i>et al.</i> 2014 [56]
<i>DSEL</i>	p.Alal24Thr (A124T)	chr18:65181506	missense	0.03, 0.2	-0.698	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>FAM175A</i>	p.Arg252Gln (R252Q)	chr4:84384688	missense	0.2	-0.724	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>FAM179A</i>	p.Val852Ala (V852A)	chr2:29259543	missense	0.5, 2	-0.183	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>GIN1</i>	p.Asn515Asp (N515D)	chr5:102423628	missense	1, 0.02	-0.371	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>GLI1</i>	p.Arg382Trp (R382W)	chr12:57863433	missense	0.03, 0.4	-0.476	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>HIVEP3</i>	p.Arg2001Gln (R2001Q)	chr1:41978890	missense	1, 0	0.288	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>LIPC</i>	p.Thr405Met (T405M)	rs113298164	missense	0.75	9.6	HDL-C	Norwegian	Holmen <i>et al.</i> 2014 [44]
<i>LIPG</i>	p.Asn396Ser (N396S)	rs77960347	missense	1.4	5.8	HDL-C	Norwegian	Holmen <i>et al.</i> 2014 [44]
<i>MYCT1</i>	p.Thr54Ala (T54A)	chr6:153019197	missense	0.5, 0	0.586	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>NAV2</i>	p.Thr447Met (T447M)	chr11:19955322	missense	0.3, 1	-0.254	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>PAFAH1B2</i>	c.482C>T (p.Ser161Leu)	rs186808413	missense	0.2, 1.1	3 mg/dL	HDL-C	AA, EA	Peloso <i>et al.</i> 2014 [56]
<i>PCSK7</i>	c.1511G>A (p.Arg504His)	rs142953140	missense	0.2, 0	17 mg/dL	HDL-C	AA, EA	Peloso <i>et al.</i> 2014 [56]
<i>PCSK9</i>	c.137G>T (p.Arg46Leu)	rs11591147	missense	3.2, 0.6	21% decrease	LDL-C	White, Black	Cohen <i>et al.</i> 2006 [53]
<i>PCSK9</i>	c.2037C>A (p.Cys679 *)	rs28362286	nonsense	1.4, <0.1, <0.2	40% decrease	LDL-C	AA, EA, Hisp	Cohen <i>et al.</i> 2005 [55]
<i>PCSK9</i>	c.426C>G (p.Tyr142 *)	rs67608943	nonsense	0.4	40% decrease	LDL-C	AA	Cohen <i>et al.</i> 2005 [55]
<i>RAE1</i>	p.Pro129Ser (P129S)	chr20:55941872	missense	0.1, 0.3	0.563	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]
<i>ZBP2</i>	p.Lys262Glu (K262E)	chr17:38031648	missense	0.03, 0.5	-0.457	TG	AA, EA	Crosby <i>et al.</i> 2014 [49]

Burden							
Locus	Burden Test	Variant ID	Variant Type	Burden Frequency (%)	Burden Effect Size	Trait	Ethnicity Reference
<i>APOB</i>	CMC	NA	LoF < 5%	0.4, 0.1	-1.9, -1.5	LDL-C	EA, AA Lange <i>et al.</i> 2014 [43]
<i>APOC3</i>	Gene-based	NA	missense, nonsense, or splice-site < 1%	0.326, 0.341	-0.55, -0.38	TG	EA, AA Crosby <i>et al.</i> 2014 [49]
<i>LDLR</i>	CMC	NA	NS < 0.1%	2.4, 2.8	0.9, 0.6	LDL-C	EA, AA Lange <i>et al.</i> 2014 [43]
<i>NPC1L1</i>		NA	nonsense, splice-site, frameshift		-13, -12	TC, LDL-C	European Stitzel <i>et al.</i> 2014 [52]
<i>PCSK9</i>	CMC	NA	LoF < 5%	0, 2.1	NA, -1.2	LDL-C	EA, AA Lange <i>et al.</i> 2014 [43]
<i>PCSK9</i>	CMC	NA	NS < 5%	3.2	-0.6	LDL-C	EA Lange <i>et al.</i> 2014 [43]
<i>PNPLA5</i>	CMC	NA	NS < 0.1%	1, 1.5	0.5, 1.1	LDL-C	EA, AA Lange <i>et al.</i> 2014 [43]
<i>SLC25A40</i>		NA	missense		0.42	TG	EA + AA Rosenthal <i>et al.</i> 2013 [13]

* Abbreviations are as follows: MAF, minor allele frequency; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol; AA, African American; EA, European American; Hisp, Hispanic; CMC, combined multivariate and collapsing method; LoF, loss-of-function; NS, nonsynonymous and splice variants