Published in final edited form as:

Curr Opin Lipidol. 2015 April ; 26(2): 114–119. doi:10.1097/MOL.0000000000000156.

Next-generation gene discovery for variants of large impact on lipid traits

Elisabeth Rosenthal^a, Elizabeth Blue^a, and Gail P. Jarvik^{a,b}

^aDepartment of Medicine (Medical Genetics), University of Washington, Seattle, Seattle, Washington, USA

^bDepartment of Genome Sciences, University of Washington, Seattle, Seattle, Washington, USA

Abstract

Purpose of review—Detection of high impact variants on lipid traits is complicated by complex genetic architecture. Although genome-wide association studies (GWAS) successfully identified many novel genes associated with lipid traits, it was less successful in identifying variants with a large impact on the phenotype. This is not unexpected, as the more common variants detectable by GWAS typically have small effects. The availability of large familial datasets and sequence data has changed the paradigm for successful genomic discovery of the novel genes and pathogenic variants underlying lipid disorders.

Recent findings—Novel loci with large effects have been successfully mapped in families, and next-generation sequencing allowed for the identification of the underlying lipid associated variants of large effect size. The success of this strategy relies on the simplification of the underlying genetic variation by focusing on large single families segregating extreme lipid phenotypes.

Summary—Rare, high impact variants are expected to have large effects and be more relevant for medical and pharmaceutical applications. Family data have many advantages over population-based data because they allow for the efficient detection of high-impact variants with an exponentially smaller sample size and increased power for follow-up studies.

Keywords

hyperlipidemia; linkage analysis; next-generation sequencing

Introduction

Hyperlipidemia, defined as elevated LDL cholesterol or high triglycerides (TG), is a major health concern because it is associated with, and may cause, cardiovascular disease (CVD). Current treatments shown to lower the risk of CVD include statins, which reduce the level of circulating LDL, and fibrates, which decrease triglyceride level and increase beneficial HDL

cholesterol levels [1]. These treatments target different biological pathways: statins target cholesterol synthesis in the liver, and fibrates alter the expression of several genes related to the synthesis of triglycerides and HDL [2, 3]. Unfortunately, these treatments are far from perfect: a wide range of side-effects frequently leads to noncompliance [4].

Treatments targeting other pathways or genes may reduce the complications and the health burdens associated with hyperlipidemia [5, 6]. Clinical trials are currently underway, targeting the known underlying genes (e.g., *APOB*, *APOC3*, and *APOE*) and newly discovered genes (i.e., *PCSK9*) [5]. Discovery of additional genes underlying hyperlipidemias may lead to such treatment opportunities.

Previously Discovered Genes

Discovery of genes and causal variants underlying hyperlipidemias proceeded via biochemical assays, linkage analysis, and gene sequencing. The first evidence that lipid levels are heritable came from the segregation studies of familial hyperlipidemias, including familial hypercholesterolemia (FH), familial hypertriglyceridemia (FHTG), and familial combined hyperlipidemia (FCHL) [7]. Subsequent biochemical assays identified low-affinity binding by LDL receptor as a cause of familial hypercholesterolemia [8], implicating mutations in the gene for the LDL receptor, *LDLR*. *In vitro* assays of binding also showed that some LDL particles had lower affinity of binding to their receptors, implicating the apoB-100 particle on the surface of LDL, the major ligand for the LDL receptor [9]. Linkage studies implicated the *APOC3* region with familial lipid disorders [10, 11] as well as triglycerides and the *APOC2/APOE/APOC1/APOC4* gene region on chromosome 19 [12]. Linkage analysis followed by gene sequencing confirmed the segregation of familial hypercholesterolemia with a novel variant in *PCSK9* [13]. Meta-analysis of *APOE* in multiple populations confirmed its association with lipid levels [14]. Comparative sequencing followed by association studies in humans showed that *APOA5* contributes to triglyceride levels [15]. These studies, and others, have clearly shown that hyperlipidemia is a genetically heterogeneous trait (including several genes not listed here) affected by several biological pathways [16].

Successful gene discovery in Mendelian disorders has accelerated because of the advent of exome and genome sequencing. The Centers for Mendelian Genomics have identified dozens of causal variants in novel genes since 2012 (www.mendelian.org/Publications; <http://data.mendelian.org/CMG/>). However, discovery of genes with large impact on complex traits, such as lipid levels, has not been as successful because of the difficulty applying traditional analysis methods to these traits.

The complex patterns of inheritance in populations can result from genetic heterogeneity, including multiple variants with large or small effects at one or more genes, incomplete or age-dependent penetrance, environmental effects, and interactions among these. Sampling single families, which have fewer segregating variants, can alleviate some of this complexity. However, inheritance in families can also appear complex because of ascertainment bias and pleiotropy. For example, FCHL, formerly proposed to be Mendelian and characterized by both high LDL and high triglycerides segregating within a family,

often appears to be due to regional multilocus variation, rather than a single-gene disorder [7, 17].

Statistical Genetics Approaches to Complex Traits

Both population-based and family-based methods have been developed to tackle complex traits. Early genome-wide association studies (GWAS) on population-based samples successfully identified many variants of modest effect size underlying lipid traits [18]. However, as expected, fewer novel loci have been detected over time with this strategy. In order to detect novel loci, researchers have used meta-analysis and datasets including massive numbers of individuals and variants ($n > 100K$), so as to overpower these underlying complex factors with sheer numbers. In this way, a weak signal from less-frequent variants of small effect size can be detected [19,20]. This approach often detects single-nucleotide variants (SNVs) near or within the genes already known to influence the trait and provide supportive evidence for candidate genes. Less frequently, some novel and candidate loci with small effect sizes have been identified for lipid traits [18,20,21]. Custom arrays with additional SNVs in genes of interest have also detected both novel-associated SNVs in known and new lipid genes [22].

Advancements in GWAS methods improved the power to detect novel loci and variants within known genes. For example, using novel lipid traits or underrepresented populations and allowing for pleiotropic effects have successfully identified novel loci [18,23,24,25–27]. Often, the most highly associated SNVs do not have a direct effect, but mark a region containing causal variation. These regions generally contain several genes, requiring fine sequencing, additional samples, and animal and cell models to pinpoint the likely causal variants or genes [18,21,28–31]. Methods incorporating genome annotations within GWAS in order to prioritize variants is an active area of research [32].

Linkage analysis is an alternative approach better suited to detecting novel loci with larger effect sizes through the incorporation of family data. Initially, linkage analysis was used for traits with a known Mendelian mode of inheritance. In the case of complex traits, larger sample sizes are needed for adequate statistical power. Linkage analysis using large families ($20 < n < 40$) has successfully localized Mendelian genes that account for some familial, early-onset versions of several complex traits, including breast cancer [33], atherosclerosis [34], and metabolic syndrome [35]. Although larger families are more difficult to ascertain than smaller families, the analysis of lipid traits has an advantage: blood draws and lipid tests are part of routine medical workups, increasing the accessibility of the phenotype and the ability to identify extreme observations in health record data. Furthermore, the phenotype protocol tends to be the same within families at the same care provider, creating uniformity that is often lacking in meta-analysis.

Analysis of larger families ($n > 40$ or larger, depending on trait and family structure) may prove to be a more efficient way to discover new genes underlying complex traits, particularly those with variants of larger effect sizes [17,36,37]. By focusing on one large family, both the potential number of underlying loci and the complexity of the mode of inheritance are dramatically reduced, as long as the trait is correctly adjusted for

ascertainment and environmental variables. In contrast, multiple unrelated families or individuals may exhibit similar phenotypes but have disparate underlying causes, reducing the power to detect any single locus [38]. Furthermore, the information gleaned from each individual within a large family, particularly the phase of the markers or haplotypes, is, on average, greater than those from small families or population-based data, allowing for an overall smaller sample size [39]. Although large samples of smaller families have also proven successful in detecting novel loci with linkage analysis [40], subsequent sequencing is more challenging in these cohorts because of the large numbers of unrelated individuals and the reduced probability of observing the same variant multiple times. Finally, if the family segregates extreme values, then it is likely that the underlying allele(s) are coding changes within the exome.

Linkage analysis of large families is increasingly feasible with advances in the speed and affordability of powerful computers. Parametric linkage analysis of very large families may still be unrealistic. Luckily, there are several model-free linkage methods that can be employed in the analysis for large pedigrees, such as variance components and inheritance-by-descent methods [41–43]. Bayesian oligo-genic linkage analysis methods, although not model free, allow for a large range of models and have been successful [17,44].

Linkage Analysis Narrows the Genomic Space

When a major gene for a trait does exist, linkage analyses may substantially reduce the region of interest from the entire genome to subsets of chromosomes. However, because of the limits of recombination within even large families, linkage detects large regions of interest, containing potentially hundreds of genes. For example, regions on both chromosomes 7 and 17, containing 215 and 78 genes, respectively, were found to be linked to hypertriglyceridemia [17], 143 genes on chromosome 19 were found to be in linkage with phospholipid transfer protein activity [45], and 276 genes on chromosome 4 were found to be in linkage with apoB levels in FCHL families [46]. These regions are a small fraction of the genome, but prohibitively expensive to Sanger sequence.

The advent of affordable and fast exome and whole-genome sequence data allows for the rapid identification of potentially causative variants within these linkage regions. In some cases, using a filtering approach to identify shared rare exon sequence variants can detect causal variants without performing linkage, *a priori* [47–49]. However this strategy can miss relevant variants because of incomplete penetrance and genetic heterogeneity. Furthermore, given that the mode of inheritance is likely unknown, all regions of interest should be queried, and not just those with the highest in-silico significance, as a mismatched trait model can lead investigators astray [17,50].

Variant Prioritization

Even when a linkage region is known, choosing relevant sequence variants remains a challenge. Restricting analysis to one large family segregating an extreme trait allows one to focus on rare variation, which is more likely to have larger effect sizes than common variants [50–52]. Furthermore, single large families allow for multiple observations of a rare

variant underlying the linkage signal, so that the association with the trait can be tested and the effect size estimated.

Most variant filtering methods use a stepwise approach in which allele frequency, coding effect, tissue expression, evolutionary conservation, and predicted damaging effect are considered [53–56]. Newer methods, such as combined annotation-dependent depletion (CADD) [57]. Create a summary score based on a model composed of these correlated factors. Alternatively, if the underlying mode of inheritance is known, an overall test incorporating both a logarithm of the odds ratio score and variant annotation can be applied using pedigree Variant Annotation, Analysis and Search Tool [58■]. The genic intolerance score, which uses sequence variation in large publicly available data-sets to prioritize genes using allele frequencies in a more nuanced way [59], may also be used to determine which variants are most interesting in a specific gene of interest.

As variant databases improve, variant novelty will not be a reliable factor in predicting a pathogenic variant, particularly for adult-onset complex traits. Rather, researchers will need to rely on allele frequency estimates from large populations of varying ancestry and ascertainment, such as the 1000 genomes project, the National Heart, Lung, and Blood Institute Exome Sequencing Project (ESP; evs.gs.washington.edu/EVS/), the UK10K project (www.uk10k.org/), and the Alzheimer's Disease Sequencing Project (www.niagads.org/adsp/content/home). However, the phenotypic data available vary across these studies, making true 'control' data-sets elusive for many traits. Furthermore, researchers should not remove highly conserved synonymous SNVs from consideration, *a priori*, as these SNVs may have a functional role in disease [60]. Similarly, the possibility that a noncoding variant may influence the transcription of nearby genes should be considered [21,61,62].

Validation

Given a statistically significant variant that explains all or some of the linkage signal in the family and is, thus, the putative causative variant, validation is still necessary. However, we have begun with the premise that the family carries an extremely rare or novel variant, so that we would not expect to find the same variant with any frequency in other data-sets. One possible approach is to identify other families segregating the same or, more likely, another variant in the same gene. Alternatively, the aforementioned datasets of unrelated individuals with sequence data can be used to validate an association of the trait and multiple rare, high-impact (e.g., coding changes with high conservation) variants in the gene [17,63–65]. As the number of tests is reduced to variants at one or a few genes, rather than the whole genome, there are fewer contrasts and thus more statistical power. In addition, the power to detect association with common SNVs of lower effect size is still reasonable in a relatively small sample ($n < 3000$). As with any candidate gene or variant, experimental evidence using appropriate model organisms is imperative to definitively assign causality and to understand the role of the gene in relevant biological pathways [66■].

Future Investment in Methods

Resources must be made available for data analysis to sustain gene discovery and functional assessment. Rich phenotype, genotype, and sequence databases on large families currently

exist [67–69], including large families with hyperlipidemia under study at several institutions [7,70–74]. Within this flood of data, there are many linkage and association signals to pursue and opportunities to improve public health using the currently available methods [50]. Increasingly, other types of data are being measured, such as methylation, RNA expression, isoform abundance within different tissues, and protein interaction networks [75–80,81], which can help prioritize genes and variants. As technology becomes cheaper and faster, analysis becomes the bottleneck. Consideration of what data are most likely to yield in high-impact results is of utmost importance as we consider rich new datasets using established or novel methods. Clearly, multidisciplinary teams of statisticians, biologists, chemists, and clinicians will be needed to identify the genes for hyperlipidemia.

Conclusion

Detection of novel genes that influence disease and related traits is integral to developing better models and treatments for disease. The advent of fast and cheap sequencing has opened up vast possibilities. Although large population-based studies have the appeal of relatively quick and easy analysis methods, compared with family-based studies, they may not be as efficient for detecting novel genes. Family-based studies can be more fruitful with much smaller sample sizes (100's vs. 100K). In addition, the analysis of family data to discover health-related genetic variation allows the researchers to maintain statistical power when transitioning to validation in the population-based setting.

Acknowledgments

Financial support and sponsorship: This study was funded by the NIH grants P01 HL030086 and T32 GM007454.

References and Recommended Reading

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

■ of special interest

■ of outstanding interest

1. 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; 366:1267–1278. [PubMed: 16214597]
2. Endo A, Kuroda M, Tsujita Y. ML-236A, ML-236B, and ML-236C, new inhibitors of cholesterologenesis produced by *Penicillium citrinum*. *J Antibiot (Tokyo)*. 1976; 29:1346–1348. [PubMed: 1010803]
3. Staels B, Van Tol A, Chan L, et al. Alterations in thyroid status modulate apolipoprotein, hepatic triglyceride lipase, and low density lipoprotein receptor in rats. *Endocrinology*. 1990; 127:1144–1152. [PubMed: 2387252]
4. Wierzbicki AS, Hardman TC, Viljoen A. New lipid-lowering drugs: an update. *Int J Clin Pract*. 2012; 66:270–280. [PubMed: 22340447]
5. Ridker PM. LDL cholesterol: controversies and future therapeutic directions. *Lancet*. 2014; 384:607–617. [PubMed: 25131980]
6. Gryn SE, Hegele RA. Pharmacogenomics, lipid disorders, and treatment options. *Clin Pharmacol Ther*. 2014; 96:36–47. [PubMed: 24722394]

7. Goldstein JL, Schrott HG, Hazzard WR, et al. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J Clin Invest.* 1973; 52:1544–1568. [PubMed: 4718953]
8. Goldstein JL, Brown MS. Binding and degradation of low density lipoproteins by cultured human fibroblasts. Comparison of cells from a normal subject and from a patient with homozygous familial hypercholesterolemia. *J Biol Chem.* 1974; 249:5153–5162. [PubMed: 4368448]
9. Innerarity TL, Weisgraber KH, Arnold KS, et al. Familial defective apolipo-protein B-100: low density lipoproteins with abnormal receptor binding. *Proc Natl Acad Sci USA.* 1987; 84:6919–6923. [PubMed: 3477815]
10. Ott J, Schrott HG, Goldstein JL, et al. Linkage studies in a large kindred with familial hypercholesterolemia. *Am J Hum Genet.* 1974; 26:598–603. [PubMed: 4213615]
11. Elston RC, Namboodiri KK, Go RC, et al. Probable linkage between essential familial hypercholesterolemia and third complement component (C3). *Cyto-genet Cell Genet.* 1976; 16:294–297.
12. Elbein SC, Hasstedt SJ. Quantitative trait linkage analysis of lipid-related traits in familial type 2 diabetes: evidence for linkage of triglyceride levels to chromosome 19q. *Diabetes.* 2002; 51:528–535. [PubMed: 11812765]
13. Abifadel M, Varret M, Rabes JP, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet.* 2003; 34:154–156. [PubMed: 12730697]
14. Dallongeville J, Lussier-Cacan S, Davignon J. Modulation of plasma triglyceride levels by apoE phenotype: a meta-analysis. *J Lipid Res.* 1992; 33:447–454. [PubMed: 1388198]
15. Pennacchio LA, Olivier M, Hubacek JA, et al. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science.* 2001; 294:169–173. [PubMed: 11588264]
16. Brouwers MC, van Greevenbroek MM, Stehouwer CD, et al. The genetics of familial combined hyperlipidaemia. *Nat Rev Endocrinol.* 2012; 8:352–362. [PubMed: 22330738]
17. Rosenthal EA, Ranchalis J, Crosslin DR, et al. NHLBI GO Exome Sequencing Project. Joint linkage and association analysis with exome sequence data implicates SLC25A40 in hypertriglyceridemia. *Am J Hum Genet.* 2013; 93:1035–1045. [PubMed: 24268658]
18. Willer CJ, Mohlke KL. Finding genes and variants for lipid levels after genome-wide association analysis. *Curr Opin Lipidol.* 2012; 23:98–103. [PubMed: 22418572]
19. Lange LA, Hu Y, Zhang H, et al. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet.* 2014; 94:233–245. This study illustrates the difficulty in finding novel, high-impact hyperlipidemia genes using GWAS-based methods, because of the low frequency of causal variants within population-based studies. [PubMed: 24507775]
20. Auer PL, Teumer A, Schick U, et al. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nat Genet.* 2014; 46:629–634. The methodology used by Auer *et al.* is reversed from that proposed here. Upon finding a candidate gene, using tens of thousands of individuals, the researchers luckily happened upon a single family which harbored a linkage variant in the same gene. That family allowed for the exploration of cosegregation. [PubMed: 24777453]
21. Willer CJ, Schmidt EM, Sengupta S, et al. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013; 45:1274–1283. [PubMed: 24097068]
22. Asselbergs FW, Guo Y, van Iperen EP, et al. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet.* 2012; 91:823–838. [PubMed: 23063622]
23. An P, Straka RJ, Pollin TI, et al. Genome-wide association studies identified novel loci for nonhigh-density lipoprotein cholesterol and its postprandial lipemic response. *Hum Genet.* 2014; 133:919–930. [PubMed: 24604477]
24. Ko A, Cantor RM, Weissglas-Volkov D, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nat Commun.* 2014; 5:3983. Ko *et al.* used a clever screening approach to limit the number of considered variants and increase power to detect

novel lipid loci in Latinos, illustrating the need for population-based research in non-European ancestry cohorts. [PubMed: 24886709]

25. Musunuru K, Romaine SP, Lettre G, et al. Multiethnic analysis of lipid-associated loci: the NHLBI CARE project. *PLoS One*. 2012; 7:e36473. [PubMed: 22629316]
26. Park SH, Lee JY, Kim S. A methodology for multivariate phenotype-based genome-wide association studies to mine pleiotropic genes. *BMC Syst Biol*. 2011; 5(Suppl. 2):S13.10.1186/1752-0509-5-S2-S13 [PubMed: 22784570]
27. Aguilar-Salinas CA, Tusie-Luna T, Pajukanta P. Genetic and environmental determinants of the susceptibility of Amerindian derived populations for having hypertriglyceridemia. *Metabolism*. 2014; 63:887–894. [PubMed: 24768220]
28. Crosby J, Peloso GM, Auer PL, et al. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014; 371:22–31. [PubMed: 24941081]
29. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010; 466:707–713. [PubMed: 20686565]
30. Blattmann P, Schuberth C, Pepperkok R, Runz H. RNAi-based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol-regulatory function. *PLoS Genet*. 2013; 9:e1003338. [PubMed: 23468663]
31. Slimani A, Harira Y, Trabelsi I, et al. Effect of E670G polymorphism in *PCSK9* gene on the risk and severity of coronary heart disease and ischemic stroke in a Tunisian cohort. *J Mol Neurosci*. 2014; 53:150–157. [PubMed: 24599757]
32. ■ Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet*. 2014; 94:559–573. Pickrell develops a method to combine functional knowledge with GWAS outcomes, in order to create statistical tests of association. Results from this study also inform about the possible mechanisms of disease and which classes of variants (nonsynonymous, methylation hot spot, transcription, etc.) may play a larger role in specific diseases. [PubMed: 24702953]
33. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*. 1990; 250:1684–1689. [PubMed: 2270482]
34. ■ Maiwald S, Sivapalaratnam S, Motazacker MM, et al. Mutation in KERA identified by linkage analysis and targeted resequencing in a pedigree with premature atherosclerosis. *PLoS One*. 2014; 9:e98289. Maiwald *et al.*, illustrate the power of the method proposed here: Parametric linkage analysis followed by sequencing identified a causal non-synonymous variant for atherosclerosis, and elevated KERA as a novel gene. Subsequent experimental analysis supported this hypothesis. [PubMed: 24879339]
35. ■ Keramati AR, Fathzadeh M, Go GW, et al. A form of the metabolic syndrome associated with mutations in *DYRK1B*. *N Engl J Med*. 2014; 370:1909–1919. Keramati *et al.* illustrate the utility of the design described here. In addition, the authors narrowly define their phenotype of interest, allowing for the use of three pedigrees which ultimately had a common ancestral haplotype. [PubMed: 24827035]
36. Zhang M, Chen J, Si D, et al. Whole exome sequencing identifies a novel EMD mutation in a Chinese family with dilated cardiomyopathy. *BMC Med Genet*. 2014; 15:77.10.1186/1471-2350-15-77 [PubMed: 24997722]
37. Santos-Cortez RL, Lee K, Giese AP, et al. Adenylate cyclase 1 (*ADCY1*) mutations cause recessive hearing impairment in humans and defects in hair cell function and hearing in zebrafish. *Hum Mol Genet*. 2014; 23:3289–3298. [PubMed: 24482543]
38. ■ Futema M, Plagnol V, Li K, et al. Simon Broome Consortium. Whole exome sequencing of familial hypercholesterolaemia patients negative for *LDLR*/*APOB*/*PCSK9* mutations. *J Med Genet*. 2014; 51:537–544. This study illustrates the difficulty in detecting novel variants and genes when using unrelated cases. In addition, the sample sizes used in this study are small for the association analysis, but would have been adequate in the case of a linkage study. [PubMed: 24987033]
39. Chapman NH, Wijsman EM. Introduction: linkage analyses in the Hutterites. *Genet Epidemiol*. 2001; 21(Suppl. 1):S222–S223. [PubMed: 11858132]

40. Simino J, Kume R, Kraja AT, et al. Linkage analysis incorporating gene–age interactions identifies seven novel lipid loci: the Family Blood Pressure Program. *Atherosclerosis*. 2014; 235:84–93. [PubMed: 24819747]
41. Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. *Am J Hum Genet*. 2011; 88:173–182. [PubMed: 21310274]
42. Thompson, EA. Statistical inference from genetic data on pedigrees. Vol. 6. Beachwood, Ohio: NSF-CBMS Regional Conference Series in Probability and Statistics; 2000.
43. Amos CI, Elston RC. Robust methods for the detection of genetic linkage for quantitative data from pedigrees. *Genet Epidemiol*. 1989; 6:349–360. [PubMed: 2721929]
44. Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet*. 1997; 61:748–760. [PubMed: 9326339]
45. Rosenthal EA, Ronald J, Rothstein J, et al. Linkage and association of phospholipid transfer protein activity to LASS4. *J Lipid Res*. 2011; 52:1837–1846. [PubMed: 21757428]
46. Wijsman EM, Rothstein JH, Igo RP, et al. Linkage and association analyses identify a candidate region for apoB level on chromosome 4q32.3 in FCHL families. *Hum Genet*. 2010; 127:705–719. [PubMed: 20383777]
47. Reddy MV, Iatan I, Weissglas-Volkov D, et al. Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family. *Circ Cardiovasc Genet*. 2012; 5:538–546. [PubMed: 22923419]
48. Wells QS, Becker JR, Su YR, et al. Whole exome sequencing identifies a causal RBM20 mutation in a large pedigree with familial dilated cardiomyopathy. *Circ Cardiovasc Genet*. 2013; 6:317–326. [PubMed: 23861363]
49. Musunuru K, Pirruccello JP, Do R, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med*. 2010; 363:2220–2227. [PubMed: 20942659]
50. Wijsman EM. The role of large pedigrees in an era of high-throughput sequencing. *Hum Genet*. 2012; 131:1555–1563. [PubMed: 22714655]
51. Gorlov IP, Gorlova OY, Sunyaev SR, et al. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet*. 2008; 82:100–112. [PubMed: 18179889]
52. Ionita-Laza I, Ottman R. Study designs for identification of rare disease variants in complex diseases: the utility of family-based designs. *Genetics*. 2011; 189:1061–1068. [PubMed: 21840850]
53. Davydov EV, Goode DL, Sirota M, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP⁺⁺. *PLoS Comput Biol*. 2010; 6:e1001025. [PubMed: 21152010]
54. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7:248–249. [PubMed: 20354512]
55. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2001; Chapter 7(Unit 7.20)
56. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003; 31:3812–3814. [PubMed: 12824425]
57. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014; 46:310–315. [PubMed: 24487276]
58. Hu H, Roach JC, Coon H, et al. A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. *Nat Biotechnol*. 2014; 32:663–669. Hu *et al.* present the first, to our knowledge, unified statistical approach to incorporating functional prediction and exome sequence data in parametric joint linkage and association analysis. [PubMed: 24837662]
59. Petrovski S, Wang Q, Heinzen EL, et al. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*. 2013; 9:e1003709. [PubMed: 23990802]
60. Korvatska O, Strand NS, Berndt JD, et al. Altered splicing of ATP6AP2 causes X-linked Parkinsonism with spasticity (XPDS). *Hum Mol Genet*. 2013; 22:3259–3268. [PubMed: 23595882]
61. Lagarrigue S, Martin L, Hormozdiari F, et al. Analysis of allele-specific expression in mouse liver by RNA-Seq: a comparison with Cis-eQTL identified using genetic linkage. *Genetics*. 2013; 195:1157–1166. [PubMed: 24026101]

62. De Castro-Oros I, Perez-Lopez J, Mateo-Gallego R, et al. A genetic variant in the LDLR promoter is responsible for part of the LDL-cholesterol variability in primary hypercholesterolemia. *BMC Med Genomics*. 2014; 7:17.10.1186/1755-8794-7-17 [PubMed: 24708769]
63. Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med*. 2014; 371:32–41. [PubMed: 24941082]
64. Wu MC, Lee S, Cai T, et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011; 89:82–93. [PubMed: 21737059]
65. Li BS, 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]
66. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014; 508:469–476. MacArthur *et al.*, emphasize the use of appropriate statistical models and follow-up testing in appropriate model organisms to verify causality of proposed functional rare variants. [PubMed: 24759409]
67. Jaquish CE. The Framingham Heart Study, on its way to becoming the gold standard for Cardiovascular Genetic Epidemiology? *BMC Med Genet*. 2007; 8:63. [PubMed: 17916250]
68. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. 1990; 132:1141–1155. [PubMed: 2260546]
69. Skolnick, MH. The Utah genealogical database: a resource for genetic epidemiology. In: Cairns, J.; Lyon, JL.; Skolnick, M., editors. *Banbury Report 4: cancer incidence in defined populations*. New York, NY: Cold Spring Harbor Laboratories; 1980.
70. Jarvik GP, Brunzell JD, Austin MA, et al. Genetic predictors of FCHL in four large pedigrees. Influence of ApoB level major locus predicted genotype and LDL subclass phenotype. *Arterioscler Thromb*. 1994; 14:1687–1694. [PubMed: 7947591]
71. Dallinga-Thie GM, van Linde-Sibenius Trip M, Rotter JJ, et al. Complex genetic contribution of the Apo AI–CIII–AIV gene cluster to familial combined hyperlipidemia. Identification of different susceptibility haplotypes. *J Clin Invest*. 1997; 99:953–961. [PubMed: 9062353]
72. Pajukanta P, Nuotio I, Terwilliger JD, et al. Linkage of familial combined hyperlipidaemia to chromosome 1q21–q23. *Nat Genet*. 1998; 18:369–373. [PubMed: 9537421]
73. Pajukanta P, Terwilliger JD, Perola M, et al. Genomewide scan for familial combined hyperlipidemia genes in Finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol, and apolipoprotein B levels. *Am J Hum Genet*. 1999; 64:1453–1463. [PubMed: 10205279]
74. Brunzell JD, Schrott HG, Motulsky AG, Bierman EL. Myocardial infarction in the familial forms of hypertriglyceridemia. *Metabolism*. 1976; 25:313–320. [PubMed: 1250165]
75. Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5:621–628. [PubMed: 18516045]
76. Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature*. 2005; 437:1173–1178. [PubMed: 16189514]
77. Sharma A, Gulbahce N, Pevzner SJ, et al. Network-based analysis of genome wide association data provides novel candidate genes for lipid and lipoprotein traits. *Mol Cell Proteomics*. 2013; 12:3398–3408. [PubMed: 23882023]
78. Guay SP, Brisson D, Lamarche B, et al. Epipolymorphisms within lipoprotein genes contribute independently to plasma lipid levels in familial hypercholesterolemia. *Epigenetics*. 2014; 9:718–729. [PubMed: 24504152]
79. DeMille D, Bikman BT, Mathis AD, et al. A comprehensive protein–protein interactome for yeast PAS kinase 1 reveals direct inhibition of respiration through the phosphorylation of Cbf1. *Mol Biol Cell*. 2014; 25:2199–2215. [PubMed: 24850888]
80. To KK, Hu M, Tomlinson B. Expression and activity of ABCG2, but not ABCB1 or OATP1B1, are associated with cholesterol levels: evidence from in vitro and in vivo experiments. *Pharmacogenomics*. 2014; 15:1091–1104. [PubMed: 25084202]

81. Chen H, Wang L, Jiang J. Transcriptome and miRNA network analysis of familial hypercholesterolemia. *Int J Mol Med*. 2014; 33:670–676. Chen *et al.* illustrate how functional annotation and network analysis can be used to detect novel genes of interest underlying hyperlipidemia. [PubMed: 24378567]

Key Points

- Linkage analysis in large families, including next-generation sequencing data, is an efficient method to detect high-impact variants in complex traits such as hyperlipidemia.
- Performing population-based studies after detecting high-impact variants in family-based studies is a statistically powerful strategy to identify and validate novel genes associated with complex traits, such as hyperlipidemia.
- Unified linkage and association methods which incorporate variant annotation and expression data are critical to the analysis of next-generation sequence data for both population and family based studies.