REVIEW

OPEN

Progress in genetic association studies of plasma lipids

Folkert W. Asselbergs[a,b,c], *Ruth C. Lovering*[d], *and Fotios Drenos*[d]

Purpose of review
This review summarizes recently published large-scale efforts elucidating the genetic architecture of lipid levels. A supplemental file with all genetic loci is provided for research purposes and we performed bioinformatic analyses of the genetic variants to give an oversight of involved pathways.

Recent findings
In total, 52 genes for HDL cholesterol, 42 genes for LDL cholesterol, 59 genes for total cholesterol, and 39 genes for triglycerides have been identified. Genetic overlap is present between the different traits and similar pathways are involved. Most of the SNPs that were detected in the European studies could be replicated in other ethnicities and these SNPs show the same direction of effect suggesting that the underlying genetic architecture of blood lipids is similar between ethnicities.

Summary
Genetic studies have identified many loci associated with plasma lipids and have provided insight into the underlying mechanisms of lipid homeostasis. Future research is needed to determine whether these loci may be novel targets for lipid-lowering therapy and for reducing cardiovascular disease risk. In addition, the proportion of genetic variance explained by these lipid loci is still limited and new large-scale genetic studies are ongoing to identify additional common and rare variants associated with lipid levels.

Keywords
array, cholesterol, functional analysis, genetics, genome-wide, lipids, SNP

INTRODUCTION

Plasma lipid levels such as HDL cholesterol (HDLc), LDL cholesterol (LDLc), total cholesterol (TC), and triglycerides are heritable risk factors for cardiovascular disease. The heritability estimates range from 0.28 to 0.78 in twin and family studies for the different lipid traits [1] suggesting that lipid levels are, at least partially, genetically determined and that genetic information can be used for early prediction of deviations from mean levels. Currently, the field of complex genetics is moving fast and it is hard to keep up-to-date with the literature reporting novel genetic associations. An overview of all variants found in genome-wide association studies (GWASs) is given, and updated regularly, on the website www.genome.gov/gwasstudies, but this resource is only reporting on findings from GWASs. Newer targeted genotyping platforms, such as the candidate-gene ITMAT-Broad-CARe (IBC) array (Illumina, San Diego, California, USA), Metabochip array, and exome chip array, developed to further refine the GWAS signals, are not covered by this catalogue. This review will summarize large-scale genetic studies using different types of arrays published in the last year to provide an overview of loci associated with the different lipid traits. A supplemental file with all loci is provided for research purposes, as investigators may want to perform look-ups in other cohorts to search for pleiotropic effects of the different genes. Finally, we performed bioinformatic analyses of the lipid loci to identify pathways likely to be relevant to these traits.

[a]Division of Heart and Lungs, Department of Cardiology, [b]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, [c]Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands and [d]Centre for Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Sciences, University College London, London, UK

Correspondence to Fotios Drenos, Centre for Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Sciences University College London, Rayne Building, 5 University St, London WC1E 6JF, UK. Tel: +44 2 07 679 6964; e-mail: f.drenos@ucl.ac.uk

Curr Opin Lipidol 2013, 24:123–128

DOI:10.1097/MOL.0b013e32835df2d6

KEY POINTS

- Genetic studies have successfully identified a large number of lipid loci including multiple potential novel targets for drug development.
- The underlying genetic architecture of blood lipids is, for a large part, similar between ethnicities, but population-specific lipid loci have been identified.
- Efforts are ongoing to identify rare variants and additional common variants to explain a larger proportion of the genetic variance in diverse populations.
- Functional follow-up studies of novel lipid loci and studies investigating the relation of novel loci to cardiovascular outcome are crucial to further elucidate the underlying biology and drug target validation.

GENOME-WIDE ASSOCIATION STUDIES

As reviewed previously in this Journal [2], a large GWAS meta-analysis from the Global Lipids Consortium (GLGC) with more than 100 000 samples has identified 95 known and novel genetic loci associated with lipid levels [3]. These loci are all listed in Supplemental Table 1, http://links.lww.com/COL/A4. This study clearly shows the clinical relevance of performing GWASs. For example, a common genetic variant in the *HMGCR* locus was found to be significantly related to LDLc. More importantly, this locus was also related to coronary artery disease (CAD) in line with the well established effect of statins, which inhibit HMGCR (3-hydroxy-3-methylglutaryl CoA reductase). Interestingly, this GWAS identified genes that harbor pathogenic mutations causing familial hypercholesterolemia, such as the *LDLR* and *PCSK9* genes, both of which were also found to be associated with CAD [3].

ITMAT-BROAD-CARE CARDIOCHIP

Recently, the International IBC Lipid Genetics Consortium [4▪▪], using the gene-centric IBC chip covering approximately 50 000 DNA markers across 2000 genes previously implicated with cardiovascular disease, performed a meta-analysis involving genetic data from over 65 000 individuals and replicated the results in an independent set of studies with over 25 000 individuals of European ancestry and the GLGC results [3]. This study identified 21 novel genes associated with levels of LDLc, HDLc, TC, and triglycerides. Some of the new loci reported were found in interesting and unexpected loci, such as the BMI locus *FTO* [5], suggesting a causal relationship between adiposity and HDLc, as this relation became nonsignificant when introducing BMI into the model. Interestingly, the well known breast cancer susceptibility locus *BRCA2* [6] was associated with LDLc, but the underlying mechanism responsible for this association is not yet understood. Future studies are needed to fully comprehend the involved pathways and whether these loci may be potential targets for drug treatment. Other detected loci may become clinically relevant in the short-term such as the *INSR* gene, which was previously associated with triglyceride levels in animal models [7]. Berberine, an isoquinoline alkaloid found in the root, rhizome, and stem bark of many plant species and used as a traditional Chinese medicine, is believed to upregulate the expression of *INSR* through the protein kinase C-dependent pathway [8] and has been associated with lowering of fasting blood glucose, insulin, and triglyceride levels in a clinical study of type 2 diabetes patients, confirming the relationship between the insulin receptor and triglyceride levels [9]. Another identified drug target is the HCAR2 gene, which was related to HDLc. HCAR2 is also known as niacin receptor 1, a well known target of niacin. Niacin has been shown to lower LDLc levels and increase HDLc [10]. Niacin treatment led to regression of carotid intima–media thickness in patients with an LDLc less than 2.6 mmol/l included in the Arterial Biology for the Investigation of the Treatment Effects of Reducing Cholesterol 6-HDL and LDL Treatment Strategies in Atherosclerosis (ARBITER-6-HALTS trial) [11]. Despite their larger effects, the Intervention in Metabolic Syndrome with Low HDL/High Triglycerides: Impact on Global Health Outcomes (AIM-HIGH) study, however, did not demonstrate any significant effect on outcome in patients with LDLc less than 1.8 mmol/l during a mean follow-up of 3 years, despite improvements in HDLc [12]. Niacin lowers LDLc in part through inhibition of DGAT2, another significant locus in the IBC meta-analysis. DGAT2 is also the presumed target of Omacor (Lovaza), a drug on the market for treatment of hypertriglyceridemia [13]. Future studies are necessary to investigate the pleiotropic effects of the detected novel lipid loci from the IBC Cardiochip and assess their relationship with relevant cardiovascular outcomes.

The International IBC Lipid Genetics Consortium was also able to verify 49 of the 136 GLGC reported associations, which due to lack of a large enough independent sample were not previously replicated. Another 38 loci in the GLGC were not represented in the IBC chip showing some of the shortcomings of gene-centric approaches. The researchers also found that some of the strongest signals appeared to have sex-specific effects but in none of the loci was the effect solely present in a

single sex. Based on the denser coverage provided by the IBC chip compared to GWAS arrays, it was revealed that a number of well known loci have a much more complex genetic architecture than previously thought. Almost one in five of all the genotyped SNPs in the study were of frequency lower than 1%, but half of all the statistically significant signals belonged in this category. Of note was the very strong association seen between the familial hypercholesterolemia causing *APOB* R3527Q (rs5742904) and LDLc ($P = 1.039 \times 10^{-46}$). This, along with the great majority of the associations identified with rare variants, was not pursued further due to the restriction imposed on the results in terms of the meta-analysis heterogeneity measure of I^2, which for rs5742904 was 96.6%.

Interestingly, the IBC study further supported that the heritability of lipids, based on the additive effects of SNPs, differs among sexes with women having higher heritability for HDLc but potentially lower heritability for triglycerides compared to men, whereas no difference could be found for LDLc and TC [4▪▪,14]. However, so far we can still only explain less than a third of the expected genetic heritability of lipid levels, around 10–15% [4▪▪]. It is currently believed that a large proportion of this heritability is explained by common variants, as illustrated by Vattikuti *et al.* [15▪] who showed that 58% of the genetic variance of height could be explained by considering all common SNPs. We expect that the explained heritability of lipids will increase as more common functional variants are identified by ongoing initiatives such as the 1000 genomes imputed meta-analysis and specific targeted arrays such as the Metabochip. It has also been speculated that the remaining proportion of the heritability may be explained by rare variants [minor allele frequency (MAF) <5%] [16]. However, despite their larger effects, uncommon variants, such as those found in the *PCSK9* [17] and *LDLR* [18] genes, may not contribute significantly to the problem of 'missing' heritability. A recent study by Park *et al.* [19▪] showed that common SNPs explain a larger fraction of the genetic variance than genetic variants with a lower allele frequency. However, this study did not investigate the contribution of variants with an allele frequency less than 1%.

BIOINFORMATICS ANALYSES

We conducted a bioinformatic analysis of the gene loci identified by the GLGC and IBC consortia. We included the genes listed by the authors, as shown in Supplemental Table S1, http://links.lww.com/COL/A4, as well as several of the genes around the top SNPs, when in an area of dense gene clusters. The Mouse Genome Informatics functional enrichment tool VLAD (http://proto.informatics.jax.org/prototypes/vlad-1.0.3/) was used to look for overrepresentation of Gene Ontology biological processes represented by each gene list dataset relative to the human dataset as a whole. The goa_human annotation set was selected; the query dataset (as UniProt IDs) was pasted into the 'Query Set' field. The 'Universe Set' field was left blank (to specify all human genes in annotation file). The 'Display Settings' options selected were 'Pruning threshold': 3 and 'Collapsing threshold': 6. A graphical summary of the functional analyses is shown in Fig. 1, and the more detailed outputs of these analyses can be found in Supplementary Tables S2–S5, http://links.lww.com/COL/A4. Tables S6, http://links.lww.com/COL/A5 and S7, summarize the functional analyses results obtained. Not surprisingly, some overlap exists between the genetic architecture of the different lipid traits as illustrated in Fig. 1. Most pathways are shared by all four traits and only a few, such as 'digestion', are present in a selection. Although all traits are associated with the 'circulatory system development' category, LDLc has the largest number of genes in this group. From Fig. 1, it is also evident that the four main roles of lipids in the body are represented in almost similar percentages, though 'metabolism' is the most common category and 'coagulation' is only represented in HDLc. The overlap of the processes is not surprising when Fig. 2 and Supplemental Table S1, http://links.lww.com/COL/A4 are considered. Six genes are associated with all fours traits, *CETP*, *APOB*, *FADS1-2-3*, *APOE*, *APOA1*, and *TRIB1*, although not always with the same SNP. HDLc is the trait with the largest number of unique associations, 28, whereas LDLc and TC seem to have the largest overlap with 25 genes in common. Finally, although TC is the lipid trait with the largest number of associated genes, 59, there are only five unique signals verifying its role as a grouping measure of the lipid profile.

ONGOING EFFORTS: METABOCHIP, EXOME CHIP, AND SEQUENCING

Recently, a number of efforts to further investigate the identified lipid signals have been made public. The Metabochip array of nearly 200 000 SNP markers aims to follow up the most significant associated variants from the GWAS meta-analysis on type 2 diabetes, CAD and myocardial infarction, and quantitative traits related to these diseases, in order to find additional variants and refine previous associations through fine mapping [20]. For lipids,

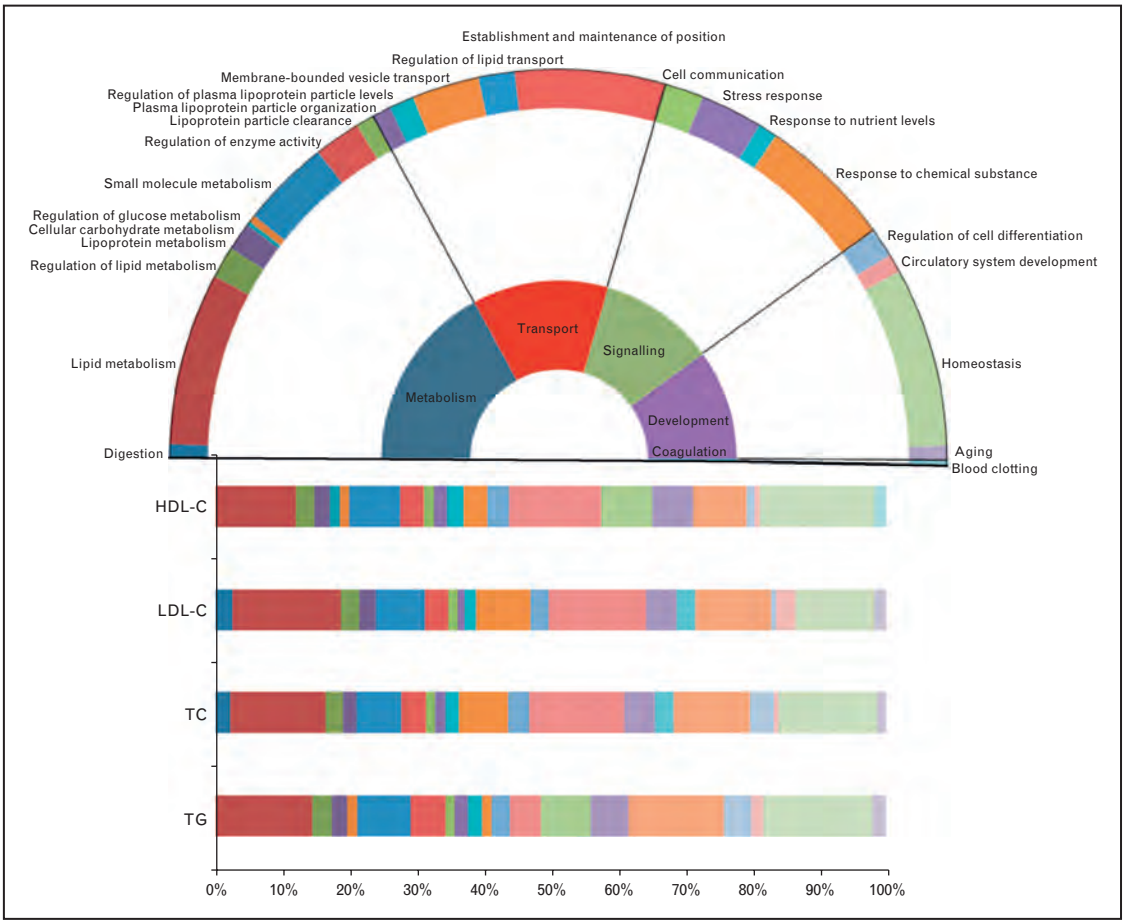

FIGURE 1. Diagrammatic representation of the enrichment of Gene Ontology terms within the genes associated with lipid trait SNPs. A selection of the enriched Gene Ontology terms are listed on the outer arc (full list of enriched Gene Ontology terms in Supplemental Table S7, http://links.lww.com/COL/A4). The inner arc provides biological process grouping terms for the Gene Ontology terms on the outer arc. The arcs represent the percentage of genes in all four lipid trait datasets, which are annotated to these Gene Ontology terms.

65 345 SNPs with preliminary evidence for association were meta-analyzed with the previous GWAS results in a total sample size of up to 188 578 individuals. Lipid levels were associated with 167 genomic regions, of which 63 were novel. These novel signals were in pathways with other, previously known, lipid-related genes. Loci associated with HDLc, LDLc, TC, and triglycerides were also associated with other cardiovascular and metabolic traits such as T2D, CAD, BMI, and blood pressure [21]. Other ongoing efforts include the work by the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium [22], which uses 1000 Genomes imputed GWAS data and preliminary results suggest that another 19 novel associations can be identified by this approach. The increasing affordability of whole genome sequencing has made the sequencing of patients with high lipid levels and their comparison with individuals of normal or low lipid levels possible. This approach can help to not only identify novel loci, but also to fine map already established loci. It is likely that these efforts will be able to extend our ability to detect new rare variants with high effect sizes, although those with more moderate effects will require increasingly bigger sample sizes. As an example of this approach, a recent study by Sanna *et al.* [23] sequenced the *PCSK9* and *LDLR* genes and detected several rare variants that were missed by conventional GWASs. The costs of performing whole exome and whole genome sequencing in individuals has dropped dramatically, but still it remains unaffordable to sequence complete genomes in large populations. As an intermediate solution, both Affymetrix and Illumina have developed an exome array containing variants in coding regions that have been seen more than once in existing whole genome or whole exome sequence datasets. Detailed description of the design of the exome chip can be

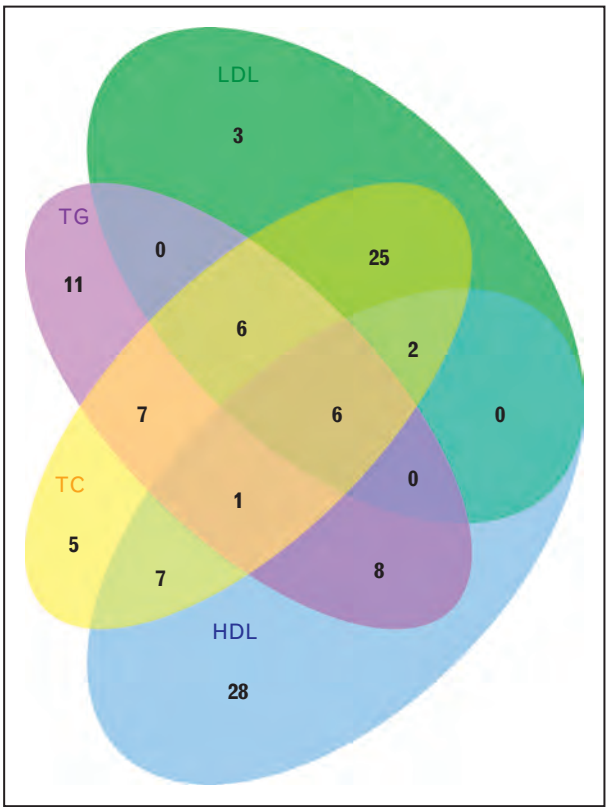

FIGURE 2. Venn diagram showing the overlap between the genes associated with the four lipids traits from large-scale studies up to now.

found on http://genome.sph.umich.edu/wiki/Exome_Chip_Design. The first results of the exome chip on lipid traits are expected soon.

MULTIETHNIC ANALYSES

Genetic studies have predominantly been performed in individuals with European ancestry. However, most of the SNPs that have been detected in European studies have also been replicated in other ethnicities. The Global Lipid Consortium has validated their identified SNPs in several non-European populations including African–Americans, East Asians, and South Asians [3]. In addition, most of these SNPs show the same direction of effect, suggesting that the underlying genetic architecture of blood lipids is similar between ethnicities. However, some of the associated loci seem to be unique for a specific ethnic background. For example, a recent meta-analysis in approximately 10 000 individuals from different ethnicities (e.g., Hispanic, African–American, East Asian) using the previously mentioned IBC Cardiochip found a significant association between HDLc and a nonsense mutation within the CD36 locus in the African–American population [24▪]. This variant, rs3211938-G, has been shown in previous studies to be associated with CD36 deficiency and with susceptibility to malaria [25] and is nearly absent in Europeans, with a MAF of 0.0005, suggesting a population-specific variation.

Additional studies are ongoing to further investigate the contribution of rare variants in other ethnicities. Recently, whole exome data for 3581 individuals from different ethnicities suggested an association between titin (*TTN*) and HDLc, and between thymocyte nuclear protein 1 (*THYN1*) and cholesterol [26]. However, these results are still pending replication in independent studies.

IDENTIFICATION OF CAUSAL VARIANTS

It is important to realize that the identified genetic variant is not *per se* responsible for the association with the lipid trait, as most often they are tagging the functional change which might be located outside the gene or in a different gene altogether. Attention is now focusing on the identification of these causal variants using statistical and laboratory approaches. The recently published Encyclopedia Of DNA Elements (ENCODE) data may play an important role in identifying functional regulatory regions of the genome involved in lipid metabolism, with a number of relevant cell lines included [27▪▪]. A recent study examined the effect of variation upon open chromatin and was able to identify a causal regulatory variant for HDLc levels within the gene encoding LXR-α (rs7120118) [28]. This study highlighted the problems of GWASs, as linkage disequilibrium at this region spans more than 29 genes, and the lead SNP was previously assigned to *F2*, a gene involved in clotting [3]. Novel use of statistical methods can be applied to fine-mapping studies, and a recent Bayesian approach has been applied to identify causal variants for coronary heart disease (CHD) and T2D [29], a method that could equally be applied to lipid traits. A study of functional variants at the *LPL* locus combined both statistical and laboratory methodologies to identify two independent regulatory variants (rs327 and rs3289) [30] associated with triglyceride levels, in addition to the established rs328 (S447X) variant [31].

CONCLUSION

A large number of lipid loci have been identified by meta-analyses of genetic association studies. The identification of these loci provides novel information about the underlying mechanism(s) of lipid biology, which may eventually lead to novel drug targets to reduce lipid levels and thereby the risk of cardiovascular disease. The proportion of the

genetic variance explained by these lipid loci is limited and new studies are ongoing to identify additional loci. These studies predominantly focus on the discovery of additional variants by fine mapping already established loci and aim to detect rare variants with large effect sizes. Adequately powered studies in different ethnicities are necessary to establish whether these variants play a universal role in lipid biology.

Acknowledgements

F.W.A. is supported by a clinical fellowship from the Netherlands Organisation for Health Research and Development (ZonMw grant 90700342). The British Heart Foundation supports F.D. (PG2005/014) and R.C.L (SP/07/007/23671).

Conflicts of interest

There are no conflicts of interest.

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

Additional references related to this topic can also be found in the Current World Literature section in this issue (p. 179).

1. Heller DA, de Faire U, Pedersen NL, *et al.* Genetic and environmental influences on serum lipid levels in twins. N Engl J Med 1993; 328:1150–1156.
2. Willer CJ, Mohlke KL. Finding genes and variants for lipid levels after genome-wide association analysis. Curr Opin Lipidol 2012; 23:98–103.
3. Teslovich TM, Musunuru K, Smith AV, *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010; 466:707–713.
4. ■■ 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.

This large-scale effort identified 21 novel loci and included more than 65 000 individuals. The largest effort since GLGC.

5. Yang J, Loos RJ, Powell JE, *et al.* FTO genotype is associated with phenotypic variability of body mass index. Nature 2012; 490:267–272.
6. Lancaster JM, Wooster R, Mangion J, *et al.* BRCA2 mutations in primary breast and ovarian cancers. Nat Genet 1996; 13:238–240.
7. Galkina EV, Butcher M, Keller SR, *et al.* Accelerated atherosclerosis in Apoe-/- mice heterozygous for the insulin receptor and the insulin receptor substrate-1. Arterioscler Thromb Vasc Biol 2012; 32:247–256.
8. Kong WJ, Zhang H, Song DQ, *et al.* Berberine reduces insulin resistance through protein kinase C-dependent up-regulation of insulin receptor expression. Metabolism 2009; 58:109–119.
9. Yin J, Xing H, Ye J. Efficacy of berberine in patients with type 2 diabetes mellitus. Metabolism 2008; 57:712–717.
10. Carlson LA. Nicotinic acid: the broad-spectrum lipid drug. A 50th anniversary review. J Intern Med 2005; 258:94–114.
11. Taylor AJ, Villines TC, Stanek EJ, *et al.* Extended-release niacin or ezetimibe and carotid intima-media thickness. N Engl J Med 2009; 361:2113–2122.
12. Boden WE, Probstfield JL, Anderson T, *et al.* Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. N Engl J Med 2011; 365:2255–2267.
13. Bays H. Clinical overview of Omacor: a concentrated formulation of omega-3 polyunsaturated fatty acids. Am J Cardiol 2006; 98:71i–76i.
14. Weiss LA, Pan L, Abney M, Ober C. The sex-specific genetic architecture of quantitative traits in humans. Nat Genet 2006; 38:218–222.
15. ■ Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. PLoS Genet 2012; 8:e1002637.

This paper showed that the majority of 'missing' heritability can be explained by common genetic variants detected by GWASs.

16. Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. Nature 2009; 461:747–753.
17. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 2006; 354:1264–1272.
18. Oosterveer DM, Versmissen J, Defesche JC, *et al.* Low-density lipoprotein receptor mutations generate synthetic genome-wide associations. Eur J Hum Genet 2012. doi: 10.1038/ejhg.2012.207.
19. ■ Park JH, Gail MH, Weinberg CR, *et al.* Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc Natl Acad Sci U S A 2011; 108:18026–18031.

Similar to Ref. [15■], this study showed that common SNPs explain a larger fraction of the genetic variance than genetic variants with a lower allele frequency.

20. Voight BF, Kang HM, Ding J, *et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet 2012; 8:e1002793.
21. Willer CJ, Abecasis G, Boehnke M, *et al.* Discovery of 63 novel loci and refinement of known loci associated with lipid levels [abstract 117]. American Society of Human Genetics 62nd Annual Meeting; 6–10 November 2012; San Francisco, California 2012.
22. Surakka I, Sarin A, Magi R, *et al.* Genome-wide screen with 1000 Genomes imputed data identifies 19 new lipid loci and new variants with stronger effects in previously known loci [abstract 119]. American Society of Human Genetics 62nd Annual Meeting; 6–10 November 2012; San Francisco, California 2012.
23. 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.
24. ■ Elbers CC, Guo CY, Tragante V, *et al.* Gene-centric meta-analysis of lipid traits in African, East Asian and Hispanic populations. PLoS One 2012; 7:e50198.

This multiethnic meta-analysis demonstrated that the majority of SNPs that were detected in the European studies could be replicated in other ethnicities. Several population-specific variants were observed.

25. Aitman TJ, Cooper LD, Norsworthy PJ, *et al.* Malaria susceptibility and CD36 mutation. Nature 2000; 405:1015–1016.
26. Highland HM, Sim X, Manning A, *et al.* The impact of genetic variation on diabetes-related quantitative traits from whole exome sequences: the T2D-GENES Consortium [abstract 269]. American Society of Human Genetics 62nd Annual Meeting; 6–10 November 2012; San Francisco, California 2012.
27. ■■ Dunham I, Kundaje A, Aldred SF, *et al.* An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489:57–74.

ENCODE provides integrated functional annotation on the human genome and is freely available for download and analysis.

28. Smith AJ, Howard P, Shah S, *et al.* Use of allele-specific FAIRE to determine functional regulatory polymorphism using large-scale genotyping arrays. PLoS Genet 2012; 8:e1002908.
29. Maller JB, McVean G, Byrnes J, *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet 2012; 44:1294–1301.
30. Smith AJ, Palmen J, Putt W, *et al.* Application of statistical and functional methodologies for the investigation of genetic determinants of coronary heart disease biomarkers: lipoprotein lipase genotype and plasma triglycerides as an exemplar. Hum Mol Genet 2010; 19:3936–3947.
31. Sagoo GS, Tatt I, Salanti G, *et al.* Seven lipoprotein lipase gene polymorphisms, lipid fractions, and coronary disease: a HuGE association review and meta-analysis. Am J Epidemiol 2008; 168:1233–1246.