Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Insights from population-based analyses of plasma lipids across the allele frequency spectrum

Gina M Peloso[1] and Pradeep Natarajan[2,3,4]

Plasma lipid levels are heritable quantitative risk factors and therapeutic targets for cardiovascular disease. Plasma lipids have been a model for translating genetic observations across the allele frequency spectrum to unique biological and therapeutic insights. Most large studies to date predominately comprised of individuals of European ancestry. This review focuses on contemporary evidence from 2016 to 2017 looking at the effect of genetic variants on plasma lipid levels across the allele frequency spectrum with incrementally larger sample sizes and the contribution of non-European ancestry studies to the genetic etiology of plasma lipid levels. To date, over 250 loci have been associated with plasma lipid levels and several of these loci have additional evidence of association with rare coding variants providing evidence for causal genes at the locus.

Addresses
[1] Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA
[2] Center for Genomic Medicine and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA
[3] Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[4] Program in Medical and Population Genetics, Broad Institute of Harvard & MIT, Cambridge, MA 02142, USA

Corresponding author: Peloso, Gina M (gpeloso@bu.edu)

Current Opinion in Genetics & Development 2018, **50**:1–6

This review comes from a themed issue on **Molecular and genetic basis of disease**

Edited by **Ines Barroso** and **Jose Florez**

https://doi.org/10.1016/j.gde.2018.01.003

Introduction

Plasma lipid levels including low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol (TC), and triglycerides (TG) are quantitative risk factors for cardiovascular disease, the leading cause of death in the United States and now worldwide [1,2]. Plasma lipids are strongly heritable — estimated to be approximately 60% [3]. Furthermore, lipid measurements are used in routine clinical practice for risk prediction and therapeutic titration [4,5], and are regularly collected within epidemiological cohorts [6–8] making them desirable phenotypes for genetic analyses with the goals of gaining biological insights and determining clinical risk for plasma lipids and cardiovascular disease as well as serving as a model for complex trait genetics since both polygenic and monogenic causes influence plasma lipid levels, with robust evidence of association.

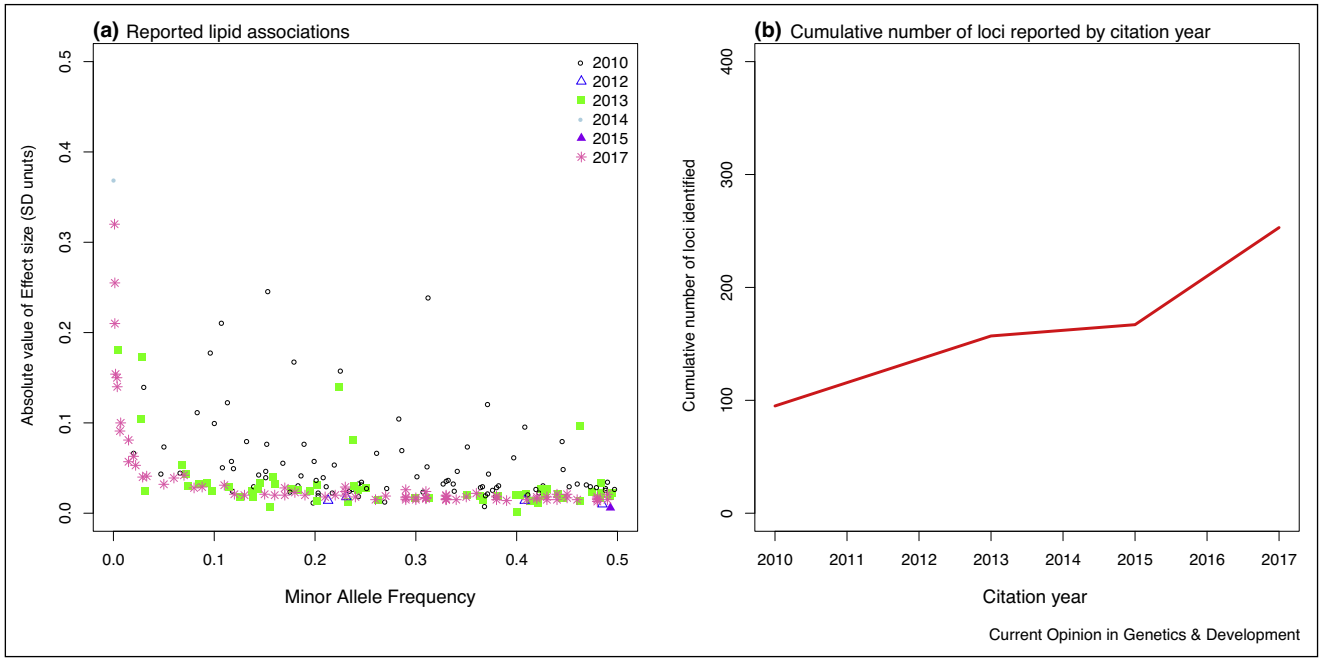
Numerous studies of the effect of genetic variants on plasma lipid levels have identified monogenic (single gene of large effect) and polygenic (multiple genes with small effects) contributors. Initial studies focused on familial hypercholesterolemia (FH), an autosomal dominant disorder characterized by elevated LDL-C and markedly increased coronary heart disease risk (CHD). In the 1970s, Brown and Goldstein described FH as due to dysfunction of the LDL receptor (*LDLR*) and concomitant over-activity of HMG-CoA reductase [9,10]. Such observations were critical to establishing the causal link between LDL-C and atherosclerosis as well as establishing HMG-CoA reductase as a therapeutic target (now the target of statins).

Over the last 50 years, additional disruptive mutations with large effects on lipid levels have been identified [11,12] as well as over 250 single nucleotide polymorphisms (SNPs) have been shown to associate with plasma lipids in the population [13••] using genome-wide association studies (GWAS), mostly in individuals of European ancestry. This review will focus on the incremental contemporary evidence from increasingly larger sample sizes, inclusion of non-European ancestry participants, and population-based analyses of aggregates of rare, coding variation.

Genetic variation detected through single variant testing

Multiple studies of common SNPs (allele frequency >∼1%) have focused on analyzing individual SNPs across the genome for association with plasma lipids [7,8,13••,14–19]. With successively larger GWAS sample sizes, we have been able to study assayed SNPs with lower minor allele frequencies (Figure 1a). Most recently, over 300 000 individuals have contributed to a meta-analysis using the 'exome array' platform of over 242 000 variants, with roughly 87% of the variants being protein-altering coding variants [13••]. With this latest GWAS, 250 loci associated with plasma lipids have been identified (Figure 1b).

Figure 1



Reported lipid associations through genome-wide scans. **(a)** Effect size by minor allele frequency of 229 reported lipid associations using results from the GLGC exome chip (http://csg.sph.umich.edu/abecasis/public/lipids2017/). The reported lead trait was used for each of the reported SNPs with effect sizes in standard deviation units. Exome chip association results were not available for 21 reported SNPs; SNP proxies were used for 16 reported SNPs. **(b)** Cumulative number of loci detected (both novel and known) for plasma lipid levels by citation year.

An application of this data type include using genetic variants as instruments to determine the causality of epidemiological observations (i.e. Mendelian randomization). This entails estimating the relationship between genetic variants that are associated with the risk factor of interest as well as the relationship between the same genetic variants and the outcome. For example, while observational epidemiological studies demonstrated a strong inverse correlation between HDL-C and CHD risk, Mendelian randomization demonstrated that this was not a causal relationship [20]. Mendelian randomization also showed that, similar to LDL-C, triglyceride rich lipoproteins (TRLs) are causally associated with CHD [21]. TRLs transport both cholesterol and triglycerides in the blood. In 2016, Helgadottir *et al.* found that a genetic risk score created for the non-HDL-C phenotype (total cholesterol − HDL-C) confers risk beyond that of LDL-C, suggesting that the cholesterol content of the TRLs may increase cardiovascular risk [22•]. A key challenge is that, by definition, non-HDL-C is correlated with triglycerides (5 * (total cholesterol − HDL-C − LDL-C)). Variants that influence serum triglycerides have been shown to influence cardiometabolic disease risk but with individually variable biological consequences. Liu *et al.* observed that TG-lowering alleles involved in hepatic production of TG-rich lipoproteins (*TM6SF2* and *PNPLA3*) lead to increased liver fat, increased risk for type 2 diabetes (T2D), and reduced risk for CHD, while TG-lowering alleles involved in peripheral lipolysis (*LPL* and *ANGPTL4*) have no effect on liver fat but reduced risks for both T2D and CHD. These findings suggest that therapeutics targeting peripheral lipolysis may decrease risk for both T2D and CHD without steatosis, and moreover, that certain TRL pathways may lead to increased risk of T2D. TRL cholesterol content, or remnant cholesterol, has been shown to associate with incident CHD [23] but the model did not take into account the triglyceride content of TRLs. The relative contributions of both remnant cholesterol and remnant triglycerides on CHD is unknown. Discovering genetic markers influencing only remnant cholesterol or only remnant triglycerides may help disentangle these two components.

Contribution of non-European individuals

A key limitation of GWAS, including for plasma lipids, is that most have primarily consisted of individuals of European ancestry [24]. In 2012, Musurunu *et al.*, found in a study of 25 000 individuals of European ancestry and 9000 individuals of African ancestry evidence of allelic heterogeneity — while the loci tended to be the same, the SNPs identified were often distinct [25]. Recently, others studying body mass index (BMI) [26] and glycemic traits [27] have drawn similar conclusions. Notably, these differences have implications for polygenic risk score accuracy across diverse individuals. In 2017, Wang *et al.*