Bivariate GWAS scan identifies six novel loci associated with lipid levels and coronary artery disease

Katherine M. Siewert¹, Benjamin F. Voight²

¹ Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

² Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Corresponding author contact: Benjamin F. Voight, bvoight@pennmedicine.upenn.edu

Abstract

Background- Plasma lipid levels are heritable and genetically associated with risk of coronary artery disease (CAD). However, genome-wide association studies (GWAS) routinely analyze these traits independently of one another. Joint GWAS for two related phenotypes can lead to a higher-powered analysis to detect variants contributing to both traits.

Methods and Results- We performed a bivariate GWAS to discover novel loci associated with heart disease, using a CAD Meta-Analysis (122,733 cases and 424,528 controls), and lipid traits, using data from the Global Lipid Genetics Consortium (188,577 subjects). We identified six previously unreported loci at genome-wide significance ($P < 5 \times 10^{-8}$), three which were associated with Triglycerides and CAD, two which were associated with LDL cholesterol and CAD, and one associated with Total Cholesterol and CAD. At several of our loci, the GWAS signals jointly localize with genetic variants associated with expression level changes for one or more neighboring genes, indicating that these loci may be affecting disease risk through regulatory activity.

Conclusions- We discovered six novel variants individually associated with both lipids and coronary artery disease.

Introduction

Genome-wide association studies (GWAS) for coronary artery disease (CAD) have identified numerous susceptibility loci associated with the disease.^{1–3} These loci account for only a fraction of the total heritability for CAD, indicating that additional loci remain to be discovered.⁴ For many CAD-associated regions, an obvious mechanistic connection to heart disease is not clear. To identify novel mechanistic connections, a recent trend is to utilize *data integration*, whereby associations for a phenotypic endpoint (like CAD) are jointly statistically analyzed with additional data types to better understand the underlying phenotypic association. This raises two questions, namely: (i) which datasets should be used, and (ii) what statistical procedures should be employed for data integration?

The community is poised to make substantial progress in the understanding of CAD through integration of diverse, publicly available genome-wide datasets. This is due in part to a basic understanding of heritable risk factors that contribute risk to CAD.⁵ Specifically, it is known that low-density lipoprotein cholesterol (LDL-C) is a causal risk factor for CAD, and emerging evidence supports elevated triglycerides as an additional causal risk factor.^{6–10} Because serum lipid levels are heritable, large-scale association scans for these traits are available, and thus can be readily integrated with CAD association data.^{11,12} In addition, large-scale studies have now emerged which map genetic variation associated with changes in gene expression (*i.e.*, eQTLs) across major human organs and tissues.¹³ These data provide an opportunity to link a CAD association signal to a putative causal gene and potentially a tissue of action.

To address the second question, several new statistical approaches have emerged. Multivariate studies that combine association data from multiple traits can help interpret the underlying associations at existing loci and have higher power for novel locus discovery than

univariate studies.^{14,15} We and others have demonstrated the utility of such approaches through their use in identifying novel loci affecting CAD and type 2 diabetes,¹⁵ or loci affecting fasting insulin, triglyceride levels and HDL cholesterol levels.¹⁶ A key advantage of this approach is the potential for direct mechanistic insight into how a locus may affect the phenotypic endpoint of interest. For example, a locus associated with both cholesterol and heart disease might suggest a mechanism of action and therapeutic hypothesis to lower disease risk. However, coincidental associations of two traits at a genomic locus may not reflect a unified etiology. Thus, new methods have been developed which quantify the evidence that an overlapping association pattern across two datasets reflects a single underlying association (*i.e.*, statistical colocalization), instead of coincidental overlap in which the associations are due to different causal variants.¹⁷

Here, we report a bivariate GWAS scan to discover novel loci associated with CAD and four lipid-related traits: LDL cholesterol (LDL-C), HDL cholesterol (HDL-C), Total Cholesterol (TC) and Triglycerides (TG) levels. To focus on a set of loci that potentially point to a shared causal variant and cognate gene of interest across traits, we applied statistical colocalization for the lipid and CAD association signals with each other and with gene expression datasets. We report six novel genome-wide significant loci where the GWAS association signals statistically colocalize: three affecting CAD and TG levels, two affecting CAD and LDL-C levels and one affecting CAD and TC. Of these loci, four signals also colocalized with eQTL signals from GTEx,¹³ indicating gene expression may be influencing disease risk at these loci.

Results and Discussion

Overview of Bivariate Scan

Using a bivariate GWAS scan approach, we identified six novel loci associated with lipid levels and heart disease. As input into this scan, we used data from a meta-analysis of coronary artery disease, which combined results from the CARDIOGRAM+C4D consortium (88,192 CAD cases and 162,544 controls)² and the UK BioBank (34,541 cases and 261,984 controls)³ and GWAS results for each of four lipid traits: HDL-C, LDL-C, Total Cholesterol, and Triglycerides, from the Global Lipid Genetics Consortium (188,577 individuals).¹² Our trait covariances and bivariate p-values appeared calibrated (**Supplementary Table 1**, **Supplementary Figure 1**) and had a strong correlation with single-trait P-values, as would be expected (**Supplementary Figures 2 and 3**). In addition, the top hits from our scan included established lipid and CAD associated loci, including *PCSK9* (significant for CAD paired with LDL-C and TC), *APOE* (CAD paired with all four lipids traits), and *LPL* (CAD paired with HDL-C and TG).

We performed several filtering steps to identify variants with non-trivial associations ($P < 5 \times 10^{-8}$) not yet established from individual trait scans. We first removed loci from our set with prior reported associations with either CAD or the lipid trait under consideration.^{1,11,18–20} Next, we narrowed our set of variants to those that were nominally associated with both traits from single-trait association data ($P < 5 \times 10^{-3}$ for the lipid trait and CAD). To focus on sites where we could hypothesize that the same causal variant contributes to both the lipid and CAD signals, we filtered out variants where the patterns of association for both traits did not statistically overlap ($\text{COLOC PP4}/(\text{PP3}+\text{PP4}) < 0.5$, **Methods, Supplementary Table 2**). After these filters, we observed two loci associated with LDL-C and CAD, three with CAD and Triglycerides and one

with CAD and TC (**Table 1**). Sentinel SNPs at these loci were not strongly linked to a nonsynonymous variant (**Methods**), consistent with a gene regulatory mechanistic hypothesis to explain phenotypic variability.²¹ We subsequently focus on associations with eQTLs or relevant prior phenotypic connections. One additional lead SNP, rs7033354, is associated with LDL-C and CAD ($P = 1.16 \times 10^{-8}$, **Table 1, Supplementary Figure 4**).²²

The monogenic diabetes gene glucokinase is associated with TG and CAD

Our first variant, rs2908290, is associated with TG and CAD (Bivariate $P = 1.6 \times 10^{-8}$) and is in an intron of the GCK (glucokinase) gene (**Table 1, Figure 1**). Although this SNP is not an eQTL in any assayed tissue, prior evidence supports GCK as the causal gene at this locus. Glucokinase phosphorylates glucose as the first step of glucose metabolism and is an established gene for maturity-onset diabetes of the young (MODY), a monogenic disorder causing early onset diabetes.²³ It has been robustly associated with metabolic phenotypes in GWAS, including fasting glucose, glycated hemoglobin and type 2 diabetes (T2D).^{15,19,24,25} A mutation in the promoter of *GCK* was previously found to be associated with CAD in a candidate gene analysis,²⁶ however, their variant is in very low linkage with our lead variant (European LD $r^2=0.05$).

The association signals at this locus are complex and suggest the possibility of several causal variants. Bien et al. found evidence of two independent associations at this locus for fasting glucose, with rs2908290 as the lead SNP of the second association.²⁷ In line with this, we found evidence of three independent genome-wide significant causal variants for glycated hemoglobin, a marker of long-term elevated glucose, within 100kb of our lead SNP using approximate conditional analysis (**Methods**).²⁵ However, none of the three lead variants are

strongly linked with rs2908290, indicating the possibility of many different casual variants for various metabolic traits at this locus. To examine if the TG and CAD signal at rs2908290 statistically overlap, we performed a colocalization analysis on a window 100kb to either side of this SNP, to exclude a nominally significant SNP several hundred kb upstream of *GCK*. The TG and CAD GWAS showed strong evidence of statistical overlap (COLOC PP4 = 0.87). Long-term overexpression of *GCK* has been shown to cause both hyperinsulinemia and hypertriglyceridemia in mice²⁸, which is broadly consistent with the pattern of genetic associations observed at this site (**Supplementary Table 3**).

As it is known that *GCKR* (glucokinase regulator) inhibits the activity of *GCK*, we next investigated the CAD association at the common variant in this gene, Pro446Leu. Previous functional studies have shown that this variant diminishes the capability of *GCKR* to inhibit *GCK*, which results in increased glycolic flux and liver metabolites that may promote *de novo* lipogenesis²⁹, providing a mechanism to explain the observed human genetic associations of decreased fasting glucose and T2D risk but increased triglyceride levels^{12,15,24,29,30}. In addition to these established associations, we observed a modest association of higher CAD for carriers of the triglyceride-increasing allele (OR = 1.018, $P = 1.9 \times 10^{-3}$, **Supplementary Table 3**). This observed effect is consistent with the predicted combined effect of elevated triglycerides and lower fasting glucose levels on CAD risk (Predicted OR = 1.022, **Methods**). These data cumulatively suggest that genetic disruption of *GCK* influences diabetes and cardiovascular disease risk through known, causal intermediates.

Locus associated with TG and HDL-C may regulate genes affecting tyrosine catabolism or platelet traits

The second locus is associated in our bivariate scan with both TG and CAD (Bivariate $P = 2.35 \times 10^{-9}$), with HDL-C and CAD (Bivariate $P = 8.78 \times 10^{-8}$, **Figure 2**), and is also an eQTL for several nearby genes. The lead variant, rs895953, is in an intron of the *SET1B* gene and, to our knowledge, has no prior established GWAS associations for related traits. Our sentinel SNP is an eQTL for three genes in the region in multiple tissues (**Supplementary Table 4**) including *RHOF* (Ras Homolog Family Member F, Filopodia Associated) and *HPD* (4-Hydroxyphenylpyruvate Dioxygenase). In contrast to epidemiological expectation, the major allele of this variant is associated with an increase in Triglycerides, but a reduced risk of CAD.

There is strong evidence of colocalization between *RHOF* and *HPD* expression and the TG signal (**Supplementary Table 4**). The CAD colocalization was weaker due to an unlinked CAD signal nearby (European LD $r^2=0.0$, **Figure 2**, 3rd panel denoted by the arrow) which would violate the assumptions of the approach used to quantify colocalization. To confirm that the CAD signal overlapping the lipids and eQTL signals was not due to this secondary signal, we performed an approximate conditional analysis (**Methods**), conditioning on the top SNP from the nearby peak (rs10849885). The association strength for our sentinel SNP improved slightly ($P_{\text{unconditional}} = 9.1 \times 10^{-4}$ to $P_{\text{conditional}} = 3.6 \times 10^{-4}$) upon conditioning, consistent with the lack of linkage disequilibrium between our sentinel SNP and the secondary CAD signal. Furthermore, the probability of colocalization between the eQTLs for *RHOF*, *HPD*, and each GWAS signal, as well as between the lipid and CAD GWAS signals themselves, also increased when we performed a colocalization analysis using a 200kb window in this region, which excludes the secondary CAD signal (**Supplementary Table 2, Supplementary Table 4**). These results are

consistent with the hypothesis that a single causal variant underlies the GWAS signals and the expression association for *RHOF* and *HPD* at this locus.

Review of the literature suggests that one or both genes are plausible candidate genes for the CAD and lipid associations. *HPD* is involved in the catabolism of tyrosine, and has been associated with 2-Hydroxyisobutyrate levels, though this previous variant is not strongly linked with ours (European LD $r^2=0.03$ between our lead SNP and that from Suhre and Raffler et al.),^{31,32}. 2-Hydroxyisobutyrate concentration is a known biomarker for insulin resistance and adiposity.^{33,34} However, literature may also support the hypothesis that *RHOF* as causal. This gene lies in an GWAS association for platelet count, platelet volume, and reticulocyte fraction of red cells.²² Platelet biology has been implicated in lipid metabolism and atherosclerosis, suggesting rs895953 could be affecting CAD risk through platelet-related pathways.^{35,36}

Locus associated with LDL-C and CAD may affect actin remodeling

A locus tagged by the sentinel SNP rs3741782 was associated with LDL-C and CAD (Bivariate $P = 3.2 \times 10^{-9}$) and fell within an intron of the *COROIC* gene (*i.e.*, Coronin, Actin Binding Protein 1C, **Figure 3**). This SNP is an eQTL for *COROIC* in visceral adipose tissue and in the adrenal gland, and for *SSH1* (*i.e.*, Slingshot Protein Phosphatase 1) in whole blood, with strong evidence of colocalization of these eQTL peaks with the CAD and LDL-C GWAS signals (**Supplementary Table 4**). Both *SSH1* and *COROIC* regulate actin reorganization.^{37,38} It has been demonstrated that uptake and degradation of lipoproteins by macrophages requires an actin cytoskeleton,³⁹ and oxidization of the associated lipids in macrophages can result in foam cells, which contribute to atherosclerosis.⁴⁰ However, to our knowledge, it is unknown whether *COROIC* or *SSH1* participate in this mechanism.

Locus associated with TG and CAD may regulate expression of *FBRSL1*

A fourth locus, tagged by sentinel SNP rs12423664 (Bivariate $P = 1.2 \times 10^{-8}$), is located in an intron of *FBRSL1* (i.e., Fibrosin Like 1). Evidence has shown that the protein encoded by this gene may be a member of the polycomb repressive complex 1 (*PRC1*), which regulates gene expression via epigenetic markers.^{41,42} Our lead SNP is an eQTL for this gene in both pancreas and whole blood. Our sentinel variant is partially linked to a previous GWAS association with red blood cell count (European LD $r^2=0.52$ between rs34390795 and our sentinel SNP).²²

Available evidence suggests that the eQTL association may underlie the lipid and CAD association. Colocalization analysis supports a shared underlying causal variant between the lipid, CAD signal, and the eQTL (**Supplementary Table 1, Supplementary Table 4**). That being said, the underlying lipid GWAS was not as densely genotyped as the CAD and eQTL scans (**Figure 3**). Indeed, the lead GWAS variant for CAD (European LD $r^2=0.08$ between rs4883525 and rs12423664), along with several other top CAD variants were not assayed in the lipid GWAS data. We expect larger-scale imputation studies for the lipids scan should enable more well-powered colocalization analysis at this locus.

The Fibrosin-Like 1 protein has been found to have lower expression in human hearts with dilated cardiomyopathy, relative to control hearts.⁴³ In the eQTL signal colocalizing with our bivariate peak, the alternative allele (A) of our lead SNP is associated with lower expression of this gene, though this association was observed in pancreas, not heart.¹³ This same allele is associated with increased lipid levels and an elevated risk of heart disease. These results indicate that Fibrosin Like 1 may play a role in healthy heart processes, and thus reduced expression of this gene may increase risk of heart disease. However, the diverse biological processes that

PRCI may influence make it difficult to interpret how this gene may be affecting triglycerides and heart disease.

Variant affecting blood pressure, CAD and Total cholesterol

Our final variant, rs1378942, is associated in our bivariate scan with CAD and total cholesterol (Bivariate $P = 2.22 \times 10^{-8}$), and also nominally associates with LDL-C (**Figure 5**). The lipid traits strongly colocalize with each other and moderately colocalize with the CAD locus (**Supplementary Table 4**). In addition, this index SNP has an established association with systolic blood pressure ($P = 6 \times 10^{-23}$),⁴⁴ with moderate to strong evidence of colocalization between both lipid traits, CAD and SBP (**Supplementary Table 5**). The ‘A’ allele of this SNP increases LDL and TC while decreasing SBP and CAD, pointing to SBP as the directionally consistent causal phenotype for the CAD association.

rs1378942 is an eQTL for *CSK* (C-terminal Src Kinase), in both subcutaneous adipose and the left ventricle of the heart, and *CYP11A1* (Cytochrome P450 Family 1 Subfamily A Member 1) in Subcutaneous Adipose and Muscle Skeletal, among several other genes.¹³ The eQTL signals for these two genes have moderate to strong evidence of colocalization with the CAD and TC GWAS peaks, although visual inspection shows that there may be more than one causal eQTL variant for the *CSK* eQTL association (**Supplementary Figure 5, Supplementary Table 4**). siRNA knockdown of *CSK* was found to increase blood pressure, suggesting it is as a possible causal gene for SBP at this locus.⁴⁵ However, *CYP11A1* is another possible causal gene. Although the substrate of the gene is not firmly established, members of the cytochrome p450 family are known to affect cardiovascular disease risk through effects on a diverse set of processes, including lipid metabolism and vasoconstriction.⁴⁶

Conclusion

Using a bivariate approach, we identified six novel variants associated with both CAD and one or more lipid traits, demonstrating the power of bivariate GWAS approaches to discover novel loci influencing phenotypes of interest, with potential candidate genes identified via integration with gene expression data. These loci are of particular interest, because they provide human genetic association evidence of altering CAD risk through lipid levels. However, other scenarios are also possible: namely, a variant may not affect CAD directly through lipid levels, and could instead influence other processes by which lipid and CAD risk levels are altered independently of one another.

Our results speak to the complexity of interpreting functional mechanisms of associated variants. Three of our loci suggest the possibility of several different regulatory targets, indicating these loci may not affect CAD and lipids through a single regulatory mechanism. Instead, these loci may harbor regulatory variants for several genes, which independently affect lipid levels and therefore CAD. Alternatively, as shown in the case of rs3741782, these regulatory targets may be involved in the same biological pathway. Furthermore, conditional analyses reveal that several of the GWAS and eQTL signals may harbor more than one causal variant, as has been previously demonstrated for several metabolic-related loci.^{47,48}

The combination of lipid and heart disease associations and their colocalization with eQTLs for genes involved with cardiometabolic process supports the potential therapeutic potential of the loci found in our bivariate scan. Further functional experimental work will be required to verify and elucidate these candidates' association with lipid and CAD pathways.

Methods

GWAS Data Collection

Our CAD GWAS data was from a meta-analysis that was performed using the UK BioBank and several prior CAD GWAS studies.³ Lipid GWAS data was obtained from the Global Lipids Genetics consortium.¹² The joint analysis of Metabochip and GWAS data was used for all four lipid traits.

To generate z-scores for the lipid traits with genomic control correction, we used the following expression:

$$Z = \Phi^{-1}\left(\frac{p}{2}\right) \frac{\beta}{|\beta|}$$

where Φ^{-1} is the inverse-cumulative distribution function of the normal distribution, p is the genomically-controlled p-value and β is the effect size of the single-trait association, as was done in Zhao et al..¹⁵

Bivariate Scan

To perform the bivariate scan, alleles were first harmonized using the `harmonise_data` function in MRbase (<https://github.com/MRCIEU/TwoSampleMR>). A set of LD-independent variants was obtained using `plink` with the command `--indep-pairwise 1000 5 0.2`,⁴⁹ using 1000 Genomes phase 3 data from Europeans for linkage information. This set of independent variants was used to estimate the means and covariance matrix of the bivariate normal distribution of the z-scores. These parameters were then used to calculate a p-value for each SNP contained in both GWAS datasets using a chi-squared test with two degrees of freedom (code available at https://github.com/WWinstonZ/bivariate_scan).

Clusters of significantly associated variants were produced using the `--clump-r2 0.2` command in R. To focus on novel loci, we filtered out variants that had a $P < 5 \times 10^{-8}$ for either

trait in our GWAS datasets, as these represented known loci. We additionally filtered out variants with a $P > 5 \times 10^{-3}$ in one of the two traits to focus on variants with convincing evidence of association with both traits. We further removed GWAS variants that were within 1MB and $r^2 > 0.2$ of a significant loci associated with CAD or the relevant lipid trait in the GWAS catalog (downloaded 2/20/18),¹⁹ Nelson et al.,¹ Klarin et al.,²⁰ Lu et al.¹⁸ or Liu et al.¹¹ We used the annovar package to search for functional mutations in $r^2 \geq 0.8$ of each lead SNP.⁵⁰

GTEEx Data Collection

eQTL data from the GTEx consortium version 7 was downloaded for the following tissues: Adipose Subcutaneous, Adipose Visceral Omentum, Artery Aorta, Artery Coronary, Artery Tibial, Heart Atrial Appendage, Heart Left Ventricle, Liver, Muscle Skeletal, Pancreas, Thyroid and Whole Blood.¹³ We focused on this subset of tissues, as (i) the most plausible for our collection of phenotypes and disease endpoints, and (ii) eQTL discovery power, based on the sample size.

Locus Characterization

We performed colocalization analysis of the GWAS using the `coloc.abf()` function from the `coloc` package (<https://github.com/chr1swallace/coloc>).¹⁷ A threshold of $PP4/(PP3+PP4)$ greater than 0.5 was used for colocalization, with default region size of 500kb on either side of the SNP of interest.

We also conducted a colocalization analysis between eQTL signals and our top bivariate GWAS hits using `coloc`. We performed `coloc` analysis between both the lipid GWAS and the eQTL signal, and the CAD GWAS and the eQTL signal. A threshold of $PP4/(PP3+PP4) \geq 0.8$ was used for “strong” evidence of colocalization and ≥ 0.5 was used for “moderate” evidence.

We performed the coloc analysis on any significant eQTL signal in GTEx at our bivariate SNP of interest in any of the aforementioned tissues.

Linkage disequilibrium for our analyses other than those using plink was calculated with LDlink (<https://analysistools.nci.nih.gov/LDlink/>) using EUR populations from the 1000 Genomes Project.⁵¹ Locus plots were generated using LocusZoom also using EUR populations.⁵²

We used the COJO method,⁵³ implemented in GCTA,⁵⁴ to perform our conditional analyses. In each case, we conditioned the lead GWAS SNP in our bivariate scan on the lead SNP of the nearby GWAS signal from which we were testing independence. To estimate the number of independent effects for the GCKR locus, the stepwise model selection method in COJO with default parameters was used.

Predicted effect of GCKR on CAD risk

We calculated the expected change in CAD risk due to the T2D and TG associations of rs1260326 at the GCKR locus. The odds ratio of CAD associated with 1 standard deviation change in triglycerides has been estimated as 1.28,⁵⁵ corresponding to a β of $\ln(1.28)=0.247$. The corresponding change in CAD risk accompanying the TG β of this SNP of -0.11 (**Supplementary Table 3**) is then $-0.11 * 0.247 = -0.027$. The effect of a 1 standard deviation change in T2D on CAD risk is estimated to be 1.11,⁵⁶ corresponding to a β of $\ln(1.11) = 0.104$. The 0.049 standard deviation change in T2D risk associated with rs1260326 is then expected to cause a CAD β of $0.104*0.049=0.0051$. Assuming an additive model, the predicted effect of rs1260326 on CAD is then $-0.027 + 0.0051 = -0.022$, which is close to the observed effect of -0.018, corresponding to an odds ratio of 0.98.

Funding Sources: This work was supported by a grant from the National Institutes of Health NIDDK R01DK101478 to B.F.V.

Table 1: Significant loci by bivariate scan with colocalization of the CAD and lipid trait GWAS signals

Variant	chr	Position (Mb)	Lipid Trait	EA/OA	EAF	Bivariate P	Lipid Beta	Lipid s.e.	Lipid P	CAD OR	CAD 95% CI	CAD P
rs895953	12	122.2	TG [†]	T/G	0.81	2.4x10 ⁻⁹	0.022	0.0043	8.6x10 ⁻⁸	0.98	0.97-0.99	9.1x10 ⁻⁴
rs3741782	12	109.0	LDL	A/G	0.69	3.2x10 ⁻⁹	-0.020	0.0039	6.3x10 ⁻⁷	0.97	0.96-0.99	3.3x10 ⁻⁶
rs7033354	9	16.9	LDL	T/C	0.63	1.2x10 ⁻⁸	-0.019	0.0038	1.4x10 ⁻⁶	0.98	0.97-0.99	7.4x10 ⁻⁶
rs12423664	12	133.0	TG	A/G	0.13	1.2x10 ⁻⁸	0.026	0.005	8.2x10 ⁻⁸	1.03	1.01-1.05	5.2x10 ⁻⁴
rs2908290	7	44.2	TG	A/G	0.38	1.6x10 ⁻⁸	0.017	0.0034	1.5x10 ⁻⁶	1.02	1.01-1.03	1.4x10 ⁻⁵
rs1378942	15	75.0	TC	A/C	0.62	2.2x10 ⁻⁸	0.012	0.0037	9.1x10 ⁻⁶	0.98	0.97-0.99	1.8x10 ⁻⁵

*EA: Effect Allele, OA: Other Allele, OR: Odds Ratio, CI: Confidence Interval, EAF: Effect Allele. Position is aligned to hg19. Frequency (EUR in 1000 Genomes).

[†] In addition to the TG association, this SNP is nominally associated with HDL (Bivariate P = 8.78 x 10⁻⁸)

Figure 1: GWAS signals for Triglycerides (TG), CAD and T2D at the GSK locus.^{15,25}

Figure 2 GWAS and eQTL associations at the *SET1B* locus. Triglyceride, HDL-C and CAD signals are the GWAS signal corresponding to the respective traits. eQTL associations are for expression of the gene in Adipose Subcutaneous.

Figure 3: GWAS and eQTL Associations at the CORO1C locus. LDL-C and CAD signals are the GWAS signal corresponding to the respective traits. CORO1C eQTL signal is for expression in adipose, and SSH1 is for expression in whole blood.

Figure 4: GWAS and eQTL Associations at the FBRSL1 locus. Triglyceride (TG) and CAD signals are the GWAS signal corresponding to the respective traits. FBRSL1 eQTL is signal is for expression of the gene in pancreas.

Figure 5: GWAS Associations at the CSK locus. Total Cholesterol (TC), LDL-C and SBP (systolic blood pressure) and CAD signals are the GWAS signal corresponding to the respective traits.

References

1. Nelson CP, Goel A, Butterworth AS, Kanoni S, Webb TR, Marouli E, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics*. 2017;49(9):1385–1391.
2. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*. 2015;47(10):1121–1130.
3. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circulation research*. 2018;122(3):433–443.
4. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. *Circulation research*. 2016;118(4):564–78.
5. Emerging Risk Factors Collaboration, Di Angelantonio E, Gao P, Pennells L, Kaptoge S, Caslake M, et al. Lipid-Related Markers and Cardiovascular Disease Prediction. *JAMA*. 2012;307(23):2499–506.
6. Voight BF, Peloso GM, Orho-Melanders M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *The Lancet*. 2012;380(9841):572–580.
7. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics*. 2015;47(11):1236–1241.
8. Ruiz-Narváez EA, Yang Y, Nakanishi Y, Kirchdorfer J, Campos H. APOC3/A5 haplotypes, lipid levels, and risk of myocardial infarction in the Central Valley of Costa Rica. *Journal of Lipid Research*. 2005;46(12):2605–2613.
9. Steinberg D. Thematic review series: the pathogenesis of atherosclerosis. An interpretive history of the cholesterol controversy, part V: the discovery of the statins and the end of the controversy. *Journal of lipid research*. 2006;47(7):1339–51.
10. Holmes M V., Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *European Heart Journal*. 2015;36(9):539–550.
11. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nature Genetics*. 2017;49(12):1758–1766.
12. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nature Genetics*. 2013;45(11):1274–1283.
13. Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–213.

14. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics*. 2018;50(2):229–237.
15. Zhao W, Rasheed A, Tikkanen E, Lee J-J, Butterworth AS, Howson JMM, et al. Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nature Genetics*. 2017;49(10):1450–1457.
16. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nature Genetics*. 2016;49(1):17–26.
17. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. Williams SM, ed. *PLoS Genetics*. 2014;10(5):e1004383.
18. Lu X, Peloso GM, Liu DJ, Wu Y, Zhang H, Zhou W, et al. Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nature Genetics*. 2017;49(12):1722–1730.
19. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*. 2017;45(D1):D896–D901.
20. Klarin D, Zhu QM, Emdin CA, Chaffin M, Horner S, McMillan BJ, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nature Genetics*. 2017;49(9):1392–1397.
21. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.)*. 2012;337(6099):1190–5.
22. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*. 2016;167(5):1415–1429.e19.
23. Froguel P, Zouali H, Vionnet N, Velho G, Vaxillaire M, Sun F, et al. Familial Hyperglycemia Due to Mutations in Glucokinase -- Definition of a Subtype of Diabetes Mellitus. *New England Journal of Medicine*. 1993;328(10):697–702.
24. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics*. 2012;44(6):659–669.
25. Wheeler E, Leong A, Liu C-T, Hivert M-F, Strawbridge RJ, Podmore C, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. Gregg E, ed. *PLOS Medicine*. 2017;14(9):e1002383.
26. Marz W, Nauck M, Hoffmann MM, Nagel D, Boehm BO, Koenig W, et al. G(-30)A Polymorphism in the Pancreatic Promoter of the Glucokinase Gene Associated With

- Angiographic Coronary Artery Disease and Type 2 Diabetes Mellitus. *Circulation*. 2004;109(23):2844–2849.
27. Bien SA, Pankow JS, Haessler J, Lu YN, Pankratz N, Rohde RR, et al. Transethnic insight into the genetics of glycaemic traits: fine-mapping results from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. *Diabetologia*. 2017;60(12):2384–2398.
28. Ferre T, Riu E, Franckhauser S, Agudo J, Bosch F. Long-term overexpression of glucokinase in the liver of transgenic mice leads to insulin resistance. *Diabetologia*. 2003;46(12):1662–1668.
29. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PRV, Orho-Melander M, et al. The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Human Molecular Genetics*. 2009;18(21):4081–4088.
30. Orho-Melander M, Gloyn AL. GCKR: How Genetic Variation Across the Allelic Spectrum Influences Protein Function and Metabolic Traits in Humans. In: *The Genetics of Type 2 Diabetes and Related Traits*. Cham: Springer International Publishing; 2016:317–336.
31. Raffler J, Friedrich N, Arnold M, Kacprowski T, Rueedi R, Altmaier E, et al. Genome-Wide Association Study with Targeted and Non-targeted NMR Metabolomics Identifies 15 Novel Loci of Urinary Human Metabolic Individuality. *PLoS genetics*. 2015;11(9):e1005487.
32. Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, Michael K, et al. A genome-wide association study of metabolic traits in human urine. *Nature genetics*. 2011;43(6):565–9.
33. Gall WE, Beebe K, Lawton KA, Adam K-P, Mitchell MW, Nakhle PJ, et al. α -Hydroxybutyrate Is an Early Biomarker of Insulin Resistance and Glucose Intolerance in a Nondiabetic Population. Federici M, ed. *PLoS ONE*. 2010;5(5):e10883.
34. Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam K-P, et al. Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. *Diabetes*. 2013;62(5):1730–7.
35. Wang N, Tall AR. Cholesterol in platelet biogenesis and activation. *Blood*. 2016;127(16):1949–53.
36. Ahmadsei M, Lievens D, Weber C, von Hundelshausen P, Gerdes N. Immune-mediated and lipid-mediated platelet function in atherosclerosis. *Current Opinion in Lipidology*. 2015;26(5):438–448.
37. Mizuno K. Signaling mechanisms and functional roles of cofilin phosphorylation and dephosphorylation. *Cellular Signalling*. 2013;25(2):457–469.
38. Xavier C-P, Rastetter RH, Blömacher M, Stumpf M, Himmel M, Morgan RO, et al. Phosphorylation of CRN2 by CK2 regulates F-actin and Arp2/3 interaction and inhibits cell migration. *Scientific reports*. 2012;2:241.

39. Hölttä-Vuori M, Vainio S, Kauppi M, Van Eck M, Jokitalo E, Ikonen E. Endosomal actin remodeling by coronin-1A controls lipoprotein uptake and degradation in macrophages. *Circulation research*. 2012;110(3):450–5.
40. Moore KJ, Sheedy FJ, Fisher EA. Macrophages in atherosclerosis: a dynamic balance. *Nature reviews. Immunology*. 2013;13(10):709–21.
41. Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular cell*. 2012;45(3):344–56.
42. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*. 2006;441(7091):349–353.
43. Creemers EE, Bawazeer A, Ugalde AP, van Deutekom HWM, van der Made I, de Groot NE, et al. Genome-Wide Polyadenylation Maps Reveal Dynamic mRNA 3'-End Formation in the Failing Human Heart. *Circulation research*. 2016;118(3):433–8.
44. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478(7367):103–9.
45. Lee H-J, Kang J-O, Kim S-M, Ji S-M, Park S-Y, Kim ME, et al. Gene Silencing and Haploinsufficiency of Csk Increase Blood Pressure. Jia Z, ed. *PLOS ONE*. 2016;11(1):e0146841.
46. Elbekai RH, El-Kadi AOS. Cytochrome P450 enzymes: Central players in cardiovascular health and disease. *Pharmacology & Therapeutics*. 2006;112(2):564–587.
47. Roman TS, Marvelle AF, Fogarty MP, Vadlamudi S, Gonzalez AJ, Buchkovich ML, et al. Multiple Hepatic Regulatory Variants at the GALNT2 GWAS Locus Associated with High-Density Lipoprotein Cholesterol. *The American Journal of Human Genetics*. 2015;97(6):801–815.
48. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507(7492):371–375.
49. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007;81(3):559–575.
50. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010;38(16):e164–e164.
51. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics (Oxford, England)*. 2015;31(21):3555–7.
52. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom:

- regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)*. 2010;26(18):2336–7.
53. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*. 2012;44(4):369–375.
 54. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*. 2011;88(1):76–82.
 55. White J, Swerdlow DI, Preiss D, Fairhurst-Hunter Z, Keating BJ, Asselbergs FW, et al. Association of Lipid Fractions With Risks for Coronary Artery Disease and Diabetes. *JAMA Cardiology*. 2016;1(6):692.
 56. Ahmad OS, Morris JA, Mujammami M, Forgetta V, Leong A, Li R, et al. A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease. *Nature Communications*. 2015;6(1):7060.