OPEN ACCESS

University of BRISTOL

Battram, T., Hoskins, L., Hughes, D., Kettunen, J., Ring, S., Davey Smith, G., ... Ford, D. (2018). Coronary Artery Disease, Genetic Risk and the Metabolome in Young Individuals. *Wellcome Open Research*, *3*, [114]. https://doi.org/10.12688/wellcomeopenres.14788.1

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
10.12688/wellcomeopenres.14788.1

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Check for updates

RESEARCH ARTICLE

Coronary artery disease, genetic risk and the metabolome in young individuals [version 1; referees: 2 approved]

Thomas Battram [1,2], Luke Hoskins [1,2], David A. Hughes[1,2],

Johannes Kettunen[1-4], Susan M. Ring [1,2], George Davey Smith [1,2],

Nicholas J. Timpson [1,2]

[1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[3]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland
[4]Biocenter Oulu, University of Oulu, Oulu, Finland

Abstract

Background: Genome-wide association studies have identified genetic variants associated with coronary artery disease (CAD) in adults – the leading cause of death worldwide. It often occurs later in life, but variants may impact CAD-relevant phenotypes early and throughout the life-course. Cohorts with longitudinal and genetic data on thousands of individuals are letting us explore the antecedents of this adult disease.

Methods: 149 metabolites, with a focus on the lipidome, measured using nuclear magnetic resonance (^{1}H-NMR) spectroscopy, and genotype data were available from 5,905 individuals at ages 7, 15, and 17 years from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. Linear regression was used to assess the association between the metabolites and an adult-derived genetic risk score (GRS) of CAD comprising 146 variants. Individual variant-metabolite associations were also examined.

Results: The CAD-GRS associated with 118 of 149 metabolites (false discovery rate [FDR] < 0.05), the strongest associations being with low-density lipoprotein (LDL) and atherogenic non-LDL subgroups. Nine of 146 variants in the GRS associated with one or more metabolites (FDR < 0.05). Seven of these are within lipid loci: rs11591147 *PCSK9,* rs12149545 *HERPUD1-CETP,* rs17091891 *LPL,* rs515135 *APOB,* rs602633 *CELSR2-PSRC1,* rs651821 *APOA5,* rs7412 *APOE-APOC1.* All associated with metabolites in the LDL or atherogenic non-LDL subgroups or both including aggregate cholesterol measures. The other two variants identified were rs112635299 *SERPINA1* and rs2519093 *ABO.*

Conclusions: Genetic variants that influence CAD risk in adults are associated with large perturbations in metabolite levels in individuals as young as seven. The variants identified are mostly within lipid-related loci and the metabolites they associated with are primarily linked to lipoproteins. This knowledge could allow for preventative measures, such as increased monitoring of at-risk individuals and perhaps treatment earlier in life, to be taken years before any symptoms of the disease arise.

Open Peer Review

Referee Status: ✓✓

	Invited Referees	
	1	**2**
version 1 published 12 Sep 2018	✓ report	✓ report

1 **Timothy M. Frayling** , University of Exeter Medical School, UK

2 **Robert Roberts**, University of Arizona College of Medicine-Phoenix, USA

Discuss this article

Comments (0)

This article is included in the Avon Longitudinal

Study of Parents and Children (ALSPAC) gateway.

Corresponding author: Thomas Battram (tb13101@bristol.ac.uk)

Author roles: Battram T: Data Curation, Formal Analysis, Investigation, Software, Visualization, Writing – Original Draft Preparation; **Hoskins L**: Data Curation, Formal Analysis, Investigation, Writing – Review & Editing; **Hughes DA**: Data Curation, Software, Visualization, Writing – Review & Editing; **Kettunen J**: Methodology, Writing – Review & Editing; **Ring SM**: Data Curation, Resources, Writing – Review & Editing; **Davey Smith G**: Conceptualization, Funding Acquisition, Writing – Review & Editing; **Timpson NJ**: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Introduction
Coronary artery disease (CAD) is the leading cause of adult death worldwide and is a gross contributor to global morbidity[1]. Many of the risk factors have long been established to be modifiable exposures such as low-density lipoprotein (LDL) cholesterol levels, smoking and hypertension[2]. In the developed world, the average age of developing Angina Pectoris, often the first clinical sign of CAD, is typically over 60[3]. However, there is evidence that "fatty streaks", the precursors to athero-sclerosis and thus CAD, form in almost all adolescents from developed countries[4]. Furthermore, there is evidence that the development of atherosclerotic plaques in coronary arteries is prolonged over the life course[5]. Thus, it is unsurprising that risk factors for CAD, including obesity and serum low-density lipo-protein levels, have been associated with an increased rate of plaque formation in children[6]. Also, a recent study suggested that higher BMI early in life was causally associated with adverse cardiovascular health[7]. These observations strongly suggest that at least some CAD risk factors may be contributing to disease development within children and there is potential for early life intervention, even if it involves nothing other than heightened clinical surveillance (not screening) by measured genetic burden.

Genome-wide association studies (GWAS) have been conducted to explore common forms of heritable contributions to this complex disease[8,9]. Over 100 genetic variants have been identi-fied as being reliably associated with an increased risk of CAD in adults[8,9]. These variants, which are likely to be exerting their influence through a diverse collection of mechanisms, are com-mon and exert relatively small effects on disease outcome singularly, but together these variants explain over 10% of CAD heritability[8,9].

It is unclear what effect these variants are having on CAD-relevant phenotypes at an earlier age (i.e. latent disease) or the longitudinal nature of the associations. Elsewhere, work analysing variation near the *FTO* locus and BMI has shown that risk alleles don't always have fixed effects on outcomes throughout life[10]. This may also be the case for other traits, like CAD. This has clinical importance because at risk individuals may gain from treatment or monitoring at various time-points across their life course. There are also implications for applied epidemiology using genetics. Currently, it is often assumed the effect of a genetic variant is fixed across the life course, but whilst the nature of the code itself may be static, the penetrance may be variable, which could be a result of gene-environment interactions.

Proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy offers a cost effective, high throughput technology to analyse multiple metabolic measures from a single sample, providing quantitative information on 149 metabolites[11–13]. The platform focuses largely on lipoproteins and fatty acids and provides the opportunity to examine individual components of lipoproteins in addition to aggregate measures. With such detailed measures of both genotypes and phenotypes, studies have already begun to successfully associate genotypic and metabolic profiles to disease phenotype, such as Type 2 diabetes[14].

Furthermore, single nucleotide polymorphisms (SNPs) have been used as instrumental variables (in a technique called Mendelian randomisation[15]) to begin to appraise the causal relationship between metabolites and CAD in adults[16]. This technique, along with new methods to quantify metabolites are starting to build evidence for the causal associations between metabolites and CAD that are beyond the well-known LDL-C and CAD relationship.

There is a clear need to explore the nature of established adult genetic associations at earlier ages. Thus, this study set out to use a detailed collection of genetic and metabolomic data to assess how genetic risk of CAD is associated with estab-lished and potential risk factors for CAD in young individuals (aged 7, 15, 17).

Methods
Study sample
The study used a single cohort: the Avon Longitudinal Study of Parents and Children (ALSPAC). ALSPAC recruited pregnant women in the Bristol and Avon area, United Kingdom, with an expected delivery date between April 1991 and December 1992. Over 14,000 pregnancies have been followed up (both children and parents) throughout the life-course. Full details of the cohort has been published previously[17]. This study focuses on the children of these pregnancies. EDTA plasma samples were collected for metabolite extraction at ages 7, 15 and 17. Individuals at ages 15 and 17 were fasted prior to sample collec-tion, but individuals at age 7 were not. Samples were aliquoted at 200µl or 500µl and stored below -70°C. Of the 7,176 par-ticipants available, 1,271 were removed due to incomplete data, leaving 5,905 for the analysis. Data at the three ages were combined in order to maximise the power of the study (N = 5,905). This was achieved by taking an individual's metab-olite data at the earliest time point possible. Full details of their characteristics are in Table 1.

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and from the UK National Health Service Local Research Ethics Committees. Full references of committee approval can be found on the ALSPAC website. Written informed consent was obtained from both the parent/ guardian and, after the age of 16, children provided written assent. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary.

Genotyping
Children were genotyped using the Illumina HumanHap550 quad genome-wide SNP genotyping platform (Illumina Inc., San Diego, CA, USA) by the Wellcome Trust Sanger Institute (WTSI; Cambridge, UK) and the Laboratory Corporation of America (LCA, Burlington, NC, USA). Participants were excluded due to having at least one of: incorrectly recorded sex, minimal or excessive heterozygosity, disproportionate levels of individual missingness (>3%), evidence of cryptic relatedness or non-European ancestry. SNPs with a minor allele frequency (MAF) <1%, a genotype missingness >1% and a call rate <95% were removed and only SNPs that passed an exact test of Hardy-Weinberg equilibrium (P<5x10-7) were included.

Table 1. Cohort characteristics.

	f7	tf3	tf4	P
N	4685	856	364	
mean age (sd)	7.54 (0.33)	15.48 (0.36)	17.86 (0.41)	
N female (%)	2265 (48.3)	461 (53.9)	200 (54.9)	
mean CAD score (sd)	0.36 (0.39)	0.37 (0.39)	0.35 (0.40)	0.501

Genotype and metabolite data were available from individuals that attended 3 clinics at different ages. N = sample size and is naturally smaller by age as the largest sample (youngest age) was used as a core collection to which non-overlapping participants from later clinics with ^{1}H-NMR data were added. f7, tf3 and tf4 are all clinics where individuals aged 7, 15 and 17 respectively were invited in to have various measurements taken. CAD (coronary artery disease) score refers to a genetic risk score comprised of 146 coronary artery disease associated genetic variants weighted by their association with the disease. The P value represents a group-wise comparison between the different CAD score values of the clinics.

For imputation, genotypes of ALSPAC mothers and children were combined. Haplotypes were estimated using ShapeIT (v2. r644), which utilises relatedness during phasing. A phased version of the 1000 genomes reference panel (Phase 1, Version 3) was obtained from the Impute2 reference data repository. Imputation was performed using Impute V2.2.2 against the reference panel (all polymorphic SNPs excluding singletons), using all 2186 reference haplotypes (including non-Europeans).

Genetic risk scores

A GWAS meta-analysis conducted using data from UK biobank and CARDIoGRAMplusC4D identified 148 variants associated with CAD at genome-wide significance ($P < 5 \times 10^{-8}$)[9]. 146 of these variants were present in the genotype data after quality control (see above) and were included in the genetic risk score. The effect size of each variant in relation to CAD was used to weight the variants – specifically the natural log of the odds ratio (OR) was used. These weightings were multiplied by the variant dosage and a CAD-GRS was produced for each individual by summing all the weighted variant values. All the loci are outlined in Supplementary Table 1.

Metabolite measures

NMR analyses of the metabolic measures was carried out at the University of Eastern Finland quantifying 149 metabolites from serum samples of the participants. The process has been described elsewhere[12]. Briefly, the samples are prepared automatically with a Gilson Liquid Handler 215, whereby 300μl of sodium phosphate NMR buffer are mixed with 300μl of serum sample. Once prepared the samples are inserted into the SampleJet™ (Bruker BioSpin GmbH, Germany) sample changer. Finally, the data are measured using a Bruker AVANCE III spectrometer. Metabolite data contains known risk factors for CAD, such as LDL-cholesterol, but also many other metabolites, as well as multiple lipoprotein subclasses. All abbreviations of metabolites used can be found in Supplementary Table 2.

Lipoprotein groupings

To examine the association between GRS and different classes of lipoproteins, lipoproteins were split into six groups based on their size and density. The groups are labelled LDL, atherogenic non-LDL, large very low-density lipoproteins (VLDL), small high-density lipoproteins (HDL), large HDL, and very large HDL (Supplementary Table 3). Groups were split in this way as it is hypothesised that: 1. The roles of lipoproteins of different sizes and densities differ 2. Only certain lipoprotein particles (here LDL and atherogenic non-LDL particles) cross into the intima, or inner most layer of a blood vessel[18,19], which is required for atherosclerosis.

HMGCR variant analysis

We sought to gauge whether a lipid lowering therapy may impact the metabolome similarly in young individuals and adults, and thus potentially reduce risk of CAD in later life. Two additional variants, external to the GRS, within the *HMGCR* locus, rs17238484 and rs12916, were chosen as proxies for statin use, as has been done previously[20]. As these were separate from the GRS, the variants were not weighted by their association with CAD and their impact on metabolite concentrations was assessed separately to all the other variants.

Statistical analyses

Metabolites were rank normalised prior to analyses to approximate normal distributions and to remove the impact of outliers. Linear regression models were used to estimate the association between metabolites in adolescence and genotype. Separately, metabolite concentrations were fitted against the CAD-GRS and each of the individual variants. Age was the only covariate in the models. An FDR-corrected P value < 0.05 was calculated using the Benjamini and Hochberg method[21].

The metabolites measured here do not necessarily represent independent phenotypes, as many are the product of the same biological event or pathway. As such, to estimate the number of independent metabolites or features present in our dataset we performed a hierarchical clustering and tree cutting analysis on the metabolite abundance data, in R[22]. Specifically, distances among metabolites was estimated by 1. subtracting the absolute Pearson's correlation coefficient from one, 2. performing hierarchal clustering on a matrix of those distance with the hclust() function and the method "complete",

3. followed by a tree cutting step at the height of 0.2 with the function cutree(). The functions hclust() and cutree() are both available in the 'stats' package[22].

All analyses were conducted in R[22] (version 3.2.2).

Results

Biological and phenotypic grouping of metabolites

5,905 individuals aged 7, 15 and 17 had NMR-measured metabolite data and genotype data (Table 1). Many metabolites share similar metabolic pathways, thus we attempted to deduce the number of independent features. Using hierarchical clustering we observe 42 independent metabolite clusters (-0.2 < r < 0.2), 23 of which are made up of a single metabolite.

When grouping lipoproteins based on their size and density we found a large overlap between the biological groupings and the clusters, with the lipoproteins within each group mostly mapping to a single cluster. This is with the exception of atherogenic non-LDL particles, where the metabolites overlap largely with clusters containing LDL and VLDL particles. Supplementary Table 3 shows the number of independent metabolite clusters represented by each grouping.

CAD-GRS metabolite associations

A GRS produced from 146 CAD-associated variants associated with 118 of the 149 metabolites tested (FDR < 0.05) (Figure 1). The 118 metabolites were observed in 20 independent metabolite clusters (-0.2 < r < 0.2), seven of which contained single metabolites. The majority of the associated metabolites are either lipoproteins or fatty acids; the only ones not in these categories were citrate, acetate and glycoprotein acetyls. The full table of results can be found in Supplementary Table 4.

When considering our lipoprotein biological groupings, it was observed that the GRS associated most strongly with atherogenic non-LDL particles and LDL (Figure 2). Furthermore, there was good evidence that the median effect size on the groups differed (Kruskal-Wallis test, P = 3.1 × 10^-14) and the median effect size on LDL and atherogenic non-LDL are larger than those observed for the other four groups (post-hoc Dunn's test, FDR < 0.05), but not among each other (P = 0.85).

Individual variant-metabolite analysis

To explore the variants driving the association between the GRS and various metabolites, all of the metabolites were regressed against each variant individually (Figure 3). In total
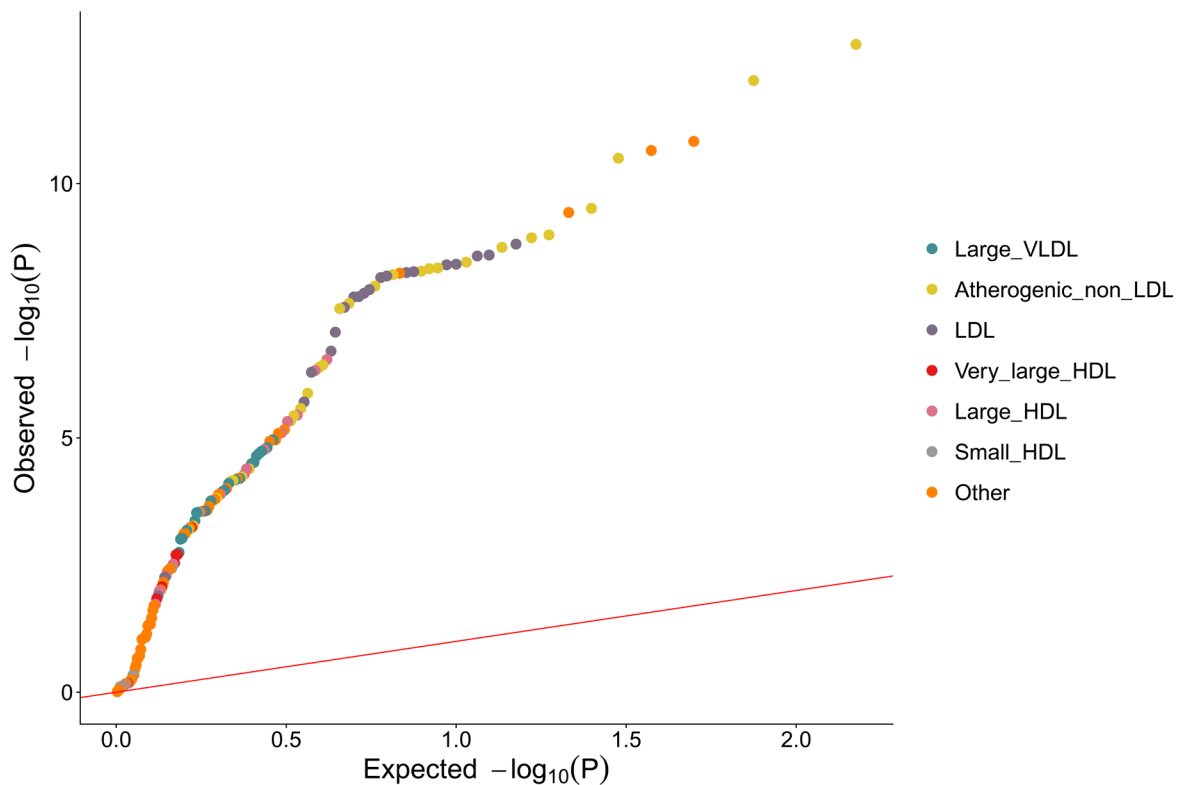
Figure 1. Association between 149 metabolites and a coronary artery disease genetic risk score. QQ-plot where each dot represents an association between one of the 149 metabolites and the genetic risk score comprised of 146 common variants. 98 of the 149 metabolites were lipoproteins and put into six groups based on size and density (Supplementary Table 3). The "other" group contains the rest of the 51 metabolites. LDL = low-density lipoprotein, VLDL = very low-density lipoprotein, HDL = high-density lipoprotein, Other = non-lipoprotein metabolites.

Figure 2. The association between 98 lipoprotein measures split into six subgroups and a coronary artery disease genetic risk score. The lipoproteins were organised into six groups based on size and density (Supplementary Table 3). LDL = low-density lipoprotein, VLDL = very low-density lipoprotein, HDL = high-density lipoprotein.

there was good evidence that nine variants associated with at least one metabolite (FDR < 0.05). Seven of these are within lipid loci: rs11591147 *PCSK9,* rs12149545 *HERPUD1-CETP,* rs17091891 *LPL,* rs515135 *APOB,* rs602633 *CELSR2-PSRC1,* rs651821 *APOA5,* rs7412 *APOE-APOC1.* All associated with metabolites in the LDL or atherogenic non-LDL subgroups or both, including aggregate cholesterol measures LDL-C, VLDL-C and IDL-C. rs2519093 *ABO* associated with three VLDL cholesterol measures and rs112635299 *SERPINA1* associated with glycoprotein acetyl and phenylalanine concentrations. Full tables of results for these nine variants found in Supplementary Table 5–Supplementary Table 13.

Potential for intervention

To assess the potential impact of early life intervention using agents that target lipoproteins, the association between rs17238484 and rs12916 *HMGCR* and the 149 metabolites was investigated. Neither of the SNPs associated with any metabolites at FDR < 0.05. At P < 0.05, rs17238484 and rs12916 associated with 12 and 42 metabolites respectively. These were mostly inverse associations with metabolites in the lipoprotein subclasses LDL and atherogenic non-LDL particles. (Supplementary Table 14, Supplementary Table 15). Supplementary Figure 1 shows the association between these variants and all metabolites alongside the other nine variants associated with one or more metabolites.

Age sensitivity analyses

In these analyses, we combined data at ages 7, 15 and 17. There were 4,685, 856 and 364 individuals from each age group respectively. To understand if grouping the individuals in this way impacted results we conducted sensitivity analyses using only individuals from each age group. We observed no difference between the median metabolite levels at different age groups (Kruskal-Wallis test, P = 0.823), and there was minimal evidence for a difference between the association of the CAD-GRS and metabolites between age groups (Kruskal-Wallis test, P = 0.024). The extent of these differences for each metabolite is displayed in Supplementary Figure 2. The effect estimates for associations between the GRS and lipoprotein groups was largely consistent between those around age 7 and 17 but varied more for those around age 15 (Supplementary Figure 3).

Discussion

In this study a GRS of CAD, made from 146 variants identified in a previous GWAS[9], associated with 118 metabolites in a sample of 5,905 individuals aged 7, 15 and 17. These metabolites were mostly lipoproteins, with stronger associations occurring in LDL and atherogenic non-LDL particles subtypes. Nine of the variants were associated with one or more metabolites. When these variants were removed from a CAD-GRS, the association between the residual GRS and the metabolites

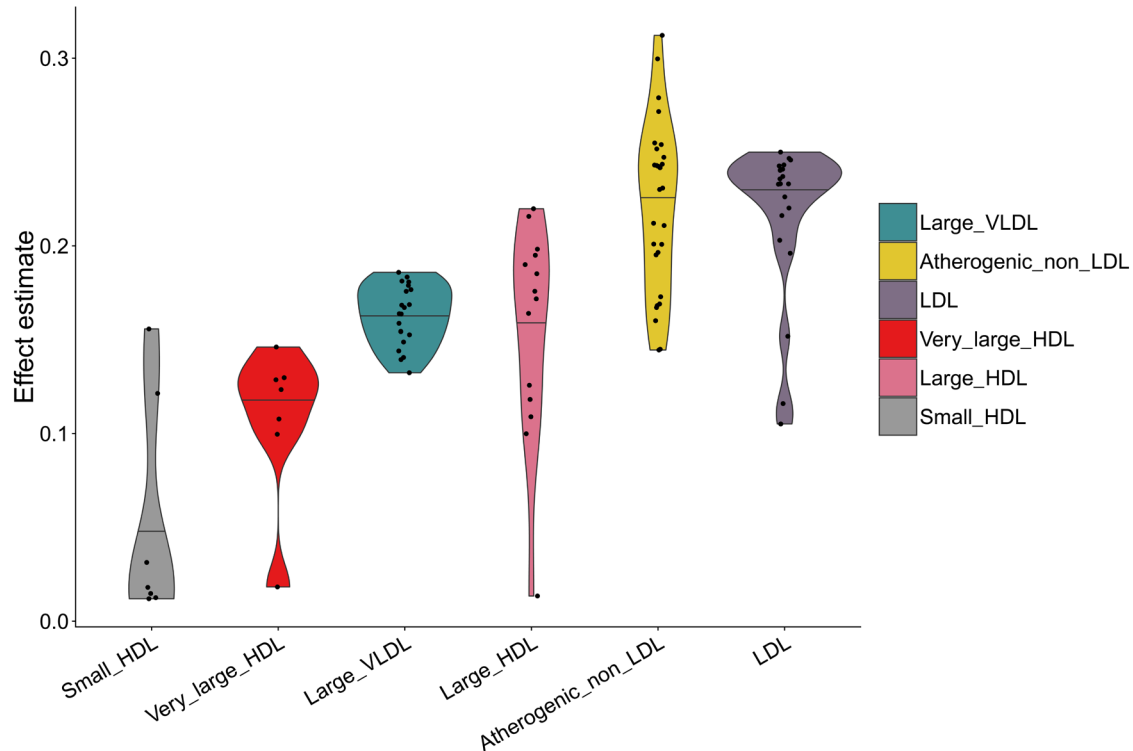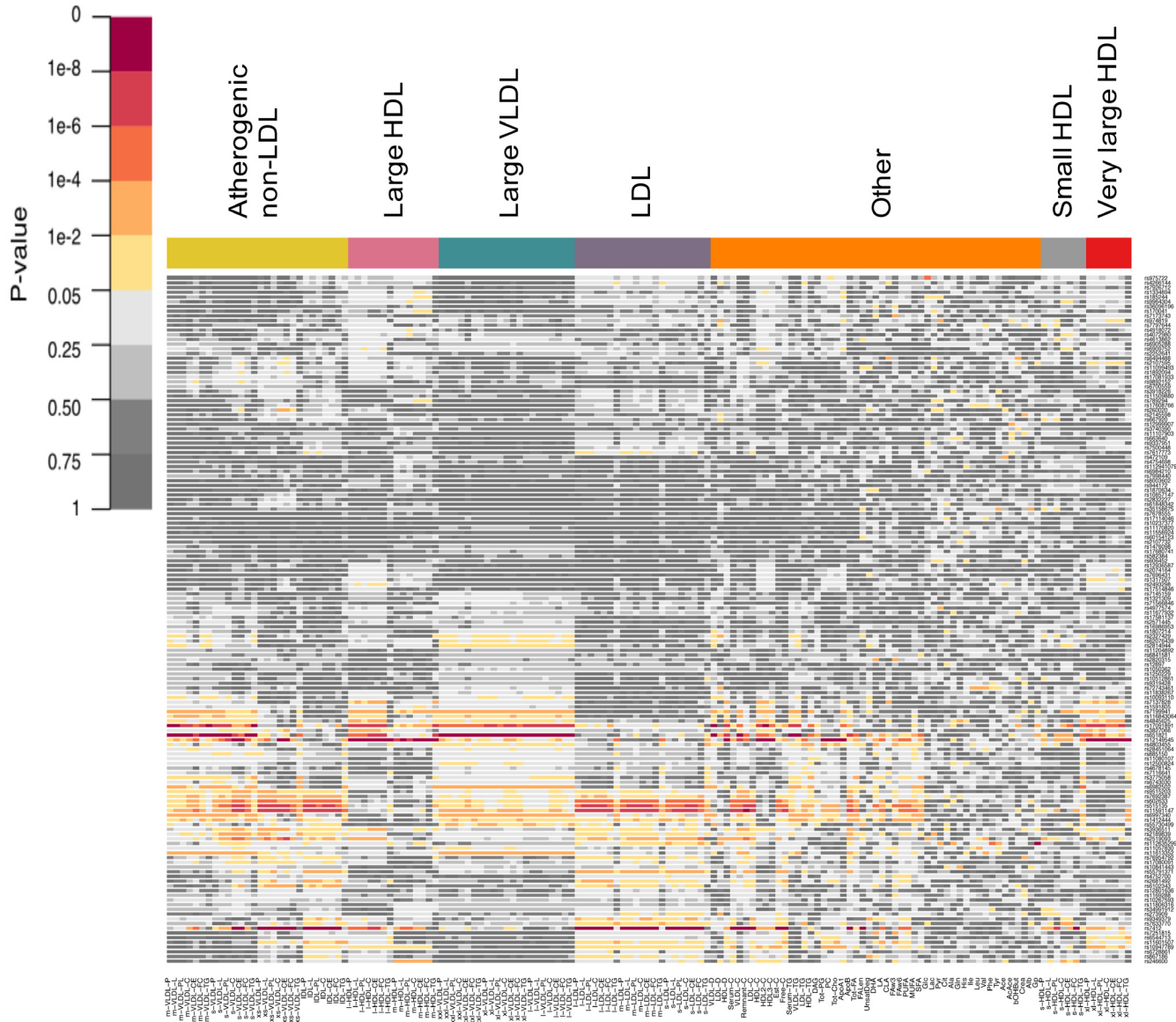**Figure 3. The association between all SNPs and all metabolites.** Each metabolite was regressed against each SNP and the P values from these analyses are presented here. Grey indicates a high P value and red a low one. The lipoproteins were grouped as previously (Supplementary Table 3).

attenuated to the null, strongly suggesting these nine variants were driving CAD-related metabolomic differences in young individuals.

The association between circulating metabolite levels and CAD has been demonstrated many times, especially with lipoproteins[2,23,24]. Therefore, it is potentially unsurprising that Figure 1 suggests that all metabolites measured were associated with CAD variants, especially as the NMR platform contains a greater proportion of lipoproteins and lipids than anything else. However, to see such a perturbation in metabolite profiles in young individuals (aged 7, 15 and 17) suggests that there are long term effects of metabolites on CAD risk and thus early-life

intervention of abnormal metabolite levels could be useful in preventing or delaying onset of this highly heritable disease. Although further analysis on the relevance of many of these metabolites to CAD needs to be done before drawing any strong conclusions.

The accumulation of lipoprotein particles, particularly LDL, within the intima has long been observed in atherosclerotic plaques[23]. *In vivo* experiments suggest that not all lipoprotein particles can cross the intima[18,19]. Interestingly, the CAD-GRS associated most strongly with LDL and atherogenic non-LDL particles, both of which are hypothesised to be small enough to cross the intima. Furthermore, there is good evidence from randomised

controlled trials and Mendelian randomization studies that lowering LDL-C (a conglomerate of all the cholesterol found within all sizes of LDL and some atherogenic non-LDL particles) reduces risk of CAD[16,25]. Thus, our results suggest genetic variants associated with CAD can drive an increase in metabolites that have evidence for causally influencing the disease.

Only 9 of the 146 variants associated with CAD, showed good evidence they associated with NMR measured metabolites in young individuals in this study. A recent GWAS of metabolites that featured 112 of our 149 metabolites was conducted in adults (mean age = 44.6). All nine genetic variants identified in our study had good evidence for association with the same or similar metabolites in the adult GWAS. Interestingly, the GWAS identified five additional variants that were present in our study but had little evidence for association with metabolites. As only five more genetic variants were identified, it suggests many variants associated with CAD are acting through pathways independent of the metabolites measured here. The discrepancy between the studies could be due to 1. chance differences, 2. a lack of power in our study; the GWAS conducted in adults had ~15,000 more individuals, 3. the effect of CAD-associated genetic variants on metabolites may vary temporally. Thus, it could be preferential to target some pathways within critical windows of time.

We assessed whether statin use might have a similar effect in young individuals as in adults in reducing LDL-C levels, to explore whether early-life drug-intervention may be a possibility for some individuals. A previous study by Swerdlow *et al.* showed that variants rs17238484 and rs12916 (*HMGCR*) inversely associate with LDL-C[20]. Here we observed weak evidence that these variants associate with LDL and intermediate-density lipoprotein (IDL) subtypes in young individuals. Along with the association between the *PCSK9* variant and metabolites, it suggests that treatments attempting to target metabolites to reduce risk of CAD or prevent other adverse CAD-related outcomes, may have similar influences within young individuals, even if the effect is reduced. These results agree with the current treatment of familial hypercholesterolemia, whereby statins are administered at young ages[26]. Nevertheless, the consequences, negative and positive, of administering these agents early in life to "seemingly healthy" individuals need to be examined. However, there is the hypothetical potential that administering treatment early in life could delay onset of disease for at risk individuals.

Even though it is unlikely clinicians will prescribe pharmaceutical agents for CAD to very young people, the variants identified in this study could be used to select those who would benefit from a less dangerous lipoprotein lowering treatment. If no treatments became available, the identification of high-risk individuals could still be used to monitor them so that intervention could begin before symptoms start to arise. Furthermore, notification of those at risk could increase caution amongst parents and individuals over environmental exposures such as diet, physical activity and smoking.

Limitations

The study combined metabolite data from young people aged 7, 15 and 17. Even though age was used as a covariate in the main models, sensitivity analysis revealed a potential difference in CAD-GRS associations with metabolites at different ages.

These data also combine metabolite data that was collected after fasting and non-fasting. There is evidence that fasting and non-fasting metabolite data are similar[27], but the study should be replicated using only fasting or only non-fasting data.

Rank-normalisation of the metabolite data removes the influence of outliers on the data but prevents true quantification of association between genotype and metabolite concentrations, i.e. with the addition of one risk allele the level of metabolite X increases by Y.

There is redundancy in the metabolite data, as many of the metabolites are highly correlated. This leads to an increase in false negatives when correcting for multiple testing. To reduce this, the Benjamini and Hochberg (FDR) method[21] was used to correct for multiple tests, rather than a more stringent family-wise error rate correction method such as Bonferroni correction. Furthermore, the study investigated how the GRS of CAD influenced lipoproteins grouped based on previous biological knowledge.

Conclusion

A CAD-GRS associated with differential abundance of 118 metabolites in young individuals. The majority of these metabolites were lipoproteins and fatty acids, and it associated most strongly with lipoproteins that are hypothesised to causally influence CAD development. This suggests that increased monitoring and early life intervention for high-risk individuals identified by their genetic profile, either by pharmaceutical agents or by behavioural changes, could help prevent onset of disease.

Data availability

ALSPAC data access is through a system of managed open access. The steps below highlight how to apply for access to the data included in this data note and all other ALSPAC data. The datasets presented in this article are linked to ALSPAC project number B2714, please quote this project number during your application. The ALSPAC variable codes highlighted in the dataset descriptions can be used to specify required variables.

1. Please read the ALSPAC access policy (PDF, 627kB) which describes the process of accessing the data and samples in detail, and outlines the costs associated with doing so.

2. You may also find it useful to browse our fully searchable research proposals database, which lists all research projects that have been approved since April 2011.

3. Please submit your research proposal for consideration by the ALSPAC Executive Committee. You will receive a

response within 10 working days to advise you whether your proposal has been approved.

If you have any questions about accessing data, please email alspac-data@bristol.ac.uk.

The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access.

All code for the analysis is freely available on GitHub: https://github.com/thomasbattram/CAD_analysis

Archived code at time of publication: http://doi.org/10.5281/zenodo.1410263[28]

Licence: MIT

Consent
Written informed consent was obtained from both the parent/guardian and, after the age of 16, children provided written assent. Children were invited to give assent where appropriate. Study members have the right to withdraw their consent for elements of the study or from the study entirely at any time. Full details of the ALSPAC consent procedures are available of the study website.

Supplementary material
Supplementary File 1: File containing the following supplementary figures –

Click here to access the data.

Supplementary Figure 1. SNPs associated with one or more metabolites at false discovery rate (FDR) < 0.05, along with the 2 SNPs within the *HMGCR* region (rs12916, rs17238484).

Supplementary Figure 2. A forest plot comparing the effect estimates for the association between the coronary artery disease (CAD) genetic risk score and 149 metabolites within each age.

Supplementary Figure 3. Comparison of effect estimates (transformed so all estimates are positive) for the association between the coronary artery disease-genetic risk score (CAD-GRS) and the lipoprotein sub-groups stratified by age.

Supplementary File 2: File containing the following supplementary tables –

Click here to access the data.

Supplementary Table 1. Coronary artery disease (CAD)-associated loci

Supplementary Table 2. Nuclear magnetic resonance (NMR) measured metabolites

Supplementary Table 3. Biological grouping of lipoprotein subclasses

Supplementary Table 4. The effect estimates and P values for the associations between the coronary artery genetic risk score and 149 metabolites.

Supplementary Table 5. All associations between rs112645299 *SERPINA1* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 6. All associations between rs11591147 *PCSK9* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 7. All associations between rs12149545 *HERPUD1-CETP* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 8. All associations between rs17091891 *LPL* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 9. All associations between rs515135 *APOB* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 10. All associations between rs2519093 *ABO* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 11. All associations between rs602633 *CELSR2-PSRC1* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 12. All associations between rs651821 *APOA5* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 13. All associations between rs7412 *APOE-APOC1* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 14. All associations between rs12916 *HMGCR* and metabolites (false discovery rate (FDR) < 0.05)

Supplementary Table 15. All associations between rs17238484 *HMGCR* and metabolites (false discovery rate (FDR) < 0.05)

References

1. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, *et al.*: **Large-scale association analysis identifies new risk loci for coronary artery disease.** *Nat Genet.* 2013; **45**(1): 25–33.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Jensen MK, Bertoia ML, Cahill LE, *et al.*: **Novel metabolic biomarkers of cardiovascular disease.** *Nat Rev Endocrinol.* 2014; **10**(11): 659–72.
   PubMed Abstract | **Publisher Full Text**

3. Moran AE, Forouzanfar MH, Roth GA, *et al.*: **The global burden of ischemic heart disease in 1990 and 2010: the Global Burden of Disease 2010 study.** *Circulation.* 2014; **129**(14): 1493–1501.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. McGill HC Jr, McMahan CA, Herderick EE, *et al.*: **Origin of atherosclerosis in childhood and adolescence.** *Am J Clin Nutr.* 2000; **72**(5 Suppl): 1307S–1315S.
   PubMed Abstract | **Publisher Full Text**

5. Insull W Jr: **The pathology of atherosclerosis: plaque development and plaque responses to medical treatment.** *Am J Med.* 2009; **122**(1 Suppl): S3–S14.
   PubMed Abstract | **Publisher Full Text**

6. Berenson GS, Srinivasan SR, Bao W, *et al.*: **Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study.** *N Engl J Med.* 1998; **338**(23): 1650–1656.
   PubMed Abstract | **Publisher Full Text**

7. Wade KH, Chiesa ST, Hughes AD, *et al.*: **Assessing the causal role of body mass index on cardiovascular health in young adults.** *Circulation.* 2018.
   Reference Source

8. Nikpay M, Goel A, Won HH, *et al.*: **A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease.** *Nat Genet.* 2015; **47**(10): 1121–30.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. van der Harst P, Verweij N: **Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease.** *Circ Res.* 2018; **122**(3): 433–443.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Sovio U, Mook-Kanamori DO, Warrington NM, *et al.*: **Association between common variation at the *FTO* locus and changes in body mass index from infancy to late childhood: the complex nature of genetic association through growth and development.** *PLoS Genet.* 2011; **7**(2): e1001307.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Soininen P, Kangas AJ, Würtz P, *et al.*: **Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics.** *Circ Cardiovasc Genet.* 2015; **8**(1): 192–206.
   PubMed Abstract | **Publisher Full Text**

12. Soininen P, Kangas AJ, Würtz P, *et al.*: **High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism.** *Analyst.* 2009; **134**(9): 1781–1785.
   PubMed Abstract | **Publisher Full Text**

13. Dunn WB, Broadhurst DI, Atherton HJ, *et al.*: **Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy.** *Chem Soc Rev.* 2011; **40**(1): 387–426.
   PubMed Abstract | **Publisher Full Text**

14. Stančáková A, Paananen J, Soininen P, *et al.*: **Effects of 34 risk loci for type 2 diabetes or hyperglycemia on lipoprotein subclasses and their composition in 6,580 nondiabetic Finnish men.** *Diabetes.* 2011; **60**(5): 1608–1616.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Davey Smith G, Hemani G: **Mendelian randomization: genetic anchors for causal inference in epidemiological studies.** *Hum Mol Genet.* 2014; **23**(R1): R89–98.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Holmes MV, Asselbergs FW, Palmer TM, *et al.*: **Mendelian randomization of blood lipids for coronary heart disease.** *Eur Heart J.* 2015; **36**(9): 539–550.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Boyd A, Golding J, Macleod J, *et al.*: **Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children.** *Int J Epidemiol.* 2013; **42**(1): 111–127.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Shaikh M, Wootton R, Nordestgaard BG, *et al.*: **Quantitative studies of transfer *in vivo* of low density, Sf 12-60, and Sf 60-400 lipoproteins between plasma and arterial intima in humans.** *Arterioscler Thromb.* 1991; **11**(3): 569–577.
   PubMed Abstract | **Publisher Full Text**

19. Nordestgaard BG, Wootton R, and Lewis B: **Selective retention of VLDL, IDL, and LDL in the arterial intima of genetically hyperlipidemic rabbits *in vivo*. Molecular size as a determinant of fractional loss from the intima-inner media.** *Arterioscler Thromb Vasc Biol.* 1995; **15**(4): 534–542.
   PubMed Abstract | **Publisher Full Text**

20. Swerdlow DI, Preiss D, Kuchenbaecker KB, *et al.*: **HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials.** *Lancet.* 2015; **385**(9965): 351–361.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc Ser B Stat Methodol.* 1995; **57**(1): 289–300.
   Reference Source

22. R Core Team: **R: A language and environment for statistical computing**. Vienna, 2017.
   Reference Source

23. Linton MF, Yancey PG, Davies SS, *et al.*: **The Role of Lipids and Lipoproteins in Atherosclerosis.** 2000; **111**: 2877.

24. Würtz P, Havulinna AS, Soininen P, *et al.*: **Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts.** *Circulation.* 2015; **131**(9): 774–785.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

25. Taylor F, Huffman MD, Macedo AF, *et al.*: **Statins for the primary prevention of cardiovascular disease.** *Cochrane Database Syst Rev.* 2013; **1**(1): CD004816.
   PubMed Abstract | **Publisher Full Text**

26. Wiegman A, Gidding SS, Watts GF, *et al.*: **Familial hypercholesterolaemia in children and adolescents: gaining decades of life by optimizing detection and treatment.** *Eur Heart J.* 2015; **36**(36): 2425–37.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Sidhu D, Naugler C: **Fasting time and lipid levels in a community-based population: a cross-sectional study.** *Arch Intern Med.* 2012; **172**(22): 1707–1710.
   PubMed Abstract | **Publisher Full Text**

28. Battram T: **thomasbattram/CAD_analysis: Second release of CAD-metabolome analysis code (Version v1.0.1).** *Zenodo.* 2018.
   http://www.doi.org/10.5281/zenodo.1410263

Open Peer Review

Current Referee Status: ✔ ✔

Version 1

Referee Report 02 November 2018

✔ **Robert Roberts**

University of Arizona College of Medicine-Phoenix, Phoenix, AZ, USA

The investigators have explored the genetic risk score versus metabolites mainly lipoproteins. This is done in a group of children and adolescents of which there is a scarcity in the literature. The results are worthy of publication and will serve as part of the increasing number of studies, we hope will be performed in this age group in the future. The paper is well done but the conclusions are perhaps overstated. The conclusion that early life intervention for high risk individuals identified by their genetic profile could help prevent onset of the disease is an overstatement for this study. Why? The investigators correlate the genetic risk score with the lipoproteins which is a good segregate for atherosclerosis but is not atherosclerosis per say. The phenotype of atherosclerosis is not assessed. Secondly, there is no intervention in this study to indicate lowering of lipoprotein or the disease. The reviewer would also like to see the raw genetic risk score and how they vary across this population. This should be included in the manuscript not in the supplemental. Secondly, if the data is available, the plasma lipoprotein level should be included. This is of interest for a variety of reasons and is too important to be included in a supplemental. The reviewer also hopes that in the future, at least, the investigators will test the genetic risk score based on non-lipid related genetic risk variants.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Cardiovascular genetics

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 23 October 2018

https://doi.org/10.21956/wellcomeopenres.16115.r33887

✔

Timothy M. Frayling iD

Genetics of Complex Traits, Royal Devon and Exeter Hospital, University of Exeter Medical School, Exeter, UK

This is a well performed report and I only have minor comments that may help the authors.

Minor points:
1. I found the penultimate sentence of the results hard to parse.
2. The final sentence of the conclusions gets a little too speculative in my opinion – suggest take out or tone down – e.g. "further research needed to find out if useful, given marker in young children.
3. Introduction – penetrance of alleles can vary across the lifecourse but not sure why the need to suggest gene x environment interactions one source – main source is simple biology – or I guess at best an interaction with age – many monogenic diseases – e.g. Huntington's don't manifest until later age. From age 7, after the adiposity peaks and troughs during early growth. FTO does have a relatively stable effect on BMI – see the largest study on longitudinal data from the HUNT study.
4. I am sure you've thought of it but if you take the LDL C SNPs, do they have bigger effects in kids compared to old adults? A separate paper and project I imagine, but could be indicative of survival bias in older cohorts, as well as genuine differences in penetrance.
5. Results. Is it surprising that the most strongly associated metabolites in figure 1 - the top 9, are non LDL based? Worth more of a mention in results or discussion (looks like 7 in figure 2)?
6. In the HMGCoR section – is the FDR calculation correct ? 42 / 149 at p<0.05 sounds like an enrichment! Also clarify directions when talking about "inverse associations" – especially in the context of the SNPs.
7. In the discussion you say you found little evidence of association for five SNPs found in adults, and that that could indicate non lipid pathways but could it also be relative lack of power? What were the sample sizes in the GWAS discovery? Presumably much bigger?
8. In the discussion about giving kids statins could mention the specific dangers – myopathy and T2D.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: I know the Bristol group well , including the senior authors and have recently socialised at a meeting with the first author. Some readers may feel this a compromise, but it is very common in science to for authors and reviewers to know each other as we work in small areas of expertise.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.