RESEARCH ARTICLE

A genome-wide association study of childhood adiposity and blood lipids [version 1; peer review: 2 approved with reservations]

Katie O'Nunain ¹, Eleanor Sanderson^{1,2}, Michael V Holmes ¹⁻⁴,
George Davey Smith ^{1,2}, Tom G Richardson ^{1,2,5}

¹Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

²MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK

³Medical Research Council Population Health Research Unit, University of Oxford, Oxford, OX3 7LF, UK

⁴Clinical Trial Service Unit & Epidemiological Studies Unit, University of Oxford, Oxford, OX3 7LF, UK

⁵Novo Nordisk Research Centre, Oxford, OX3 7FZ, UK

v1 First published: 10 Nov 2021, 6:303
<https://doi.org/10.12688/wellcomeopenres.16928.1>
Latest published: 23 Mar 2023, 6:303
<https://doi.org/10.12688/wellcomeopenres.16928.2>

Abstract

Background: The rising prevalence of childhood obesity and dyslipidaemia is a major public health concern due to its association with morbidity and mortality in later life.

Methods: In this study, we have conducted genome-wide association studies (GWAS) for eight measures of adiposity and lipids in a cohort of young individuals (mean age 9.9) from the Avon Longitudinal Study of Parents and Children (ALSPAC). These measures were body mass index (BMI), systolic and diastolic blood pressure, high-density and low-density lipoprotein cholesterol, triglycerides, apolipoprotein A-I and apolipoprotein B. We next undertook functional enrichment, pathway analyses and linkage disequilibrium (LD) score regression to evaluate genetic correlations with later-life cardiometabolic diseases.

Results: Using GWAS we identified 14 unique loci associated with at least one risk factor in this cohort of age 10 individuals ($P < 5 \times 10^{-8}$), with lipoprotein lipid-associated loci being enriched for liver tissue-derived gene expression and lipid synthesis pathways. LD score regression provided evidence of various genetic correlations, such as childhood systolic blood pressure being genetically correlated with later-life coronary artery disease ($r_G = 0.26$, 95% CI = 0.07 to 0.46, $P = 0.009$) and hypertension ($r_G = 0.37$, 95% CI = 0.19 to 0.55, $P = 6.57 \times 10^{-5}$), as well as childhood BMI with type 2 diabetes ($r_G = 0.35$, 95% CI = 0.18 to 0.51, $P = 3.28 \times 10^{-5}$).

Conclusions: Our findings suggest that there are genetic variants inherited at birth which begin to exert their effects on cardiometabolic

Open Peer Review

Approval Status ? ?

	1	2
version 2 (revision) 23 Mar 2023		
version 1 10 Nov 2021	 view	 view
1. Veera Rajagopal , Aarhus University, Aarhus, Denmark		
2. Samuel Jones , University of Helsinki, Helsinki, Finland		

Any reports and responses or comments on the article can be found at the end of the article.

risk factors as early as age 10 in the life course. However, further research is required to assess whether the genetic correlations we have identified are due to direct or indirect effects of childhood adiposity and lipid traits.

Keywords

Early life adiposity, lipoprotein lipids, cardiometabolic disease, genetic correlations, ALSPAC

This article is included in the [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#) gateway.

Corresponding author: Tom G Richardson (Tom.G.Richardson@bristol.ac.uk)

Author roles: **O'Nunain K:** Data Curation, Formal Analysis, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Sanderson E:** Methodology, Supervision, Writing – Review & Editing; **Holmes MV:** Methodology, Writing – Review & Editing; **Davey Smith G:** Methodology, Writing – Review & Editing; **Richardson TG:** Conceptualization, Methodology, Supervision, Writing – Review & Editing

Competing interests: Dr Holmes has collaborated with Boehringer Ingelheim in research, and in adherence to the University of Oxford's Clinical Trial Service Unit & Epidemiological Studies Unit (CSTU) staff policy, did not accept personal honoraria or other payments from pharmaceutical companies. TGR is employed part-time by Novo Nordisk outside of this work. All other authors declare no competing interests.

Grant information: This work was supported by Wellcome (102215); the Integrative Epidemiology Unit which receives funding from the UK Medical Research Council and the University of Bristol (MC_UU_00011/1). MVH works in a unit that receives funding from the UK Medical Research Council, and is supported by a British Heart Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. TGR is a UKRI Innovation Research Fellow (MR/S003886/1).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 O'Nunain K *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: O'Nunain K, Sanderson E, Holmes MV *et al.* **A genome-wide association study of childhood adiposity and blood lipids [version 1; peer review: 2 approved with reservations]** Wellcome Open Research 2021, 6:303 <https://doi.org/10.12688/wellcomeopenres.16928.1>

First published: 10 Nov 2021, 6:303 <https://doi.org/10.12688/wellcomeopenres.16928.1>

Introduction

Childhood obesity is a growing epidemic estimated to affect over 100 million children globally (GBD 2015 Obesity Collaborators *et al.*, 2017). Early intervention for this disease is crucial owing to its detrimental influence on children's psychological and physical health (Vander Wal & Mitchell, 2011). Furthermore, childhood obesity and dyslipidaemia are associated with an increased risk of cardiovascular disease, type 2 diabetes and hypertension in later life (Ayer *et al.*, 2015; Baker *et al.*, 2007; Pulgaron & Delamater, 2014). These chronic disease outcomes have a poor prognosis and place a considerable economic burden on healthcare systems worldwide (Wang *et al.*, 2011). This emphasises the importance of understanding the early life influences of adiposity and lipoprotein lipid traits, even though previous studies have suggested that they ultimately influence cardiometabolic disease outcomes if their levels remain high for many years across the life course (Bjerregaard & Baker, 2018; Newman *et al.*, 1990; Richardson *et al.*, 2020b).

There is strong evidence of a genetic contribution to adiposity, such as previous studies estimating the heritability of body mass index (BMI) at 40% (Hemani *et al.*, 2013; Robinson *et al.*, 2017). Although there have been numerous genome-wide association studies (GWAS) to date of childhood BMI (Bradfield *et al.*, 2019; Felix *et al.*, 2016; Vogelesang *et al.*, 2020), there have been far fewer GWAS of blood pressure (Parmar *et al.*, 2016), and in particular lipoprotein lipid traits, based on measures during childhood.

In this study, we have conducted GWAS of eight measures of adiposity and lipoprotein lipids within a population of young individuals (mean age 9.9) from the Avon Longitudinal Study of Parents and Children (ALSPAC) (Boyd *et al.*, 2013). These were BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, triglycerides, apolipoprotein A-I and apolipoprotein B. We next undertook functional enrichment analyses to highlight the putative underlying tissue types responsible for our GWAS results and to investigate whether they were overrepresented amongst curated biological pathways. In doing so we sought to recapitulate findings from large-scale studies of adult populations, therefore reinforcing that the genome-wide loci identified in our study begin to exert their effects on traits in childhood. Finally, we conducted linkage disequilibrium (LD) score regression to evaluate genetic correlations of childhood adiposity and blood lipid traits with later-life cardiometabolic disease endpoints.

Methods

The Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is a transgenerational cohort study designed to investigate the influence of genetic and environmental factors on the health of both parents and children. The details of the study are described elsewhere (Boyd *et al.*, 2013; Fraser *et al.*, 2013). In brief, the study recruited 13,761 pregnant women who lived in South West England and were due to deliver between the 1st

April 1991 and 31st December 1992. These women and their children have been followed up at regular intervals over the past 27 years. Detailed phenotypic information, biological samples and genetic data have been collected from the participants which are available through a searchable data dictionary (<http://www.bris.ac.uk/alspac/researchers/our-data/>). Written informed consent was obtained for all study participants. Ethical approval for this study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

Genotyping and imputation. Genome-wide genotyping was undertaken on ALSPAC offspring at a cohort level with quality control, cleaning and imputation, as described previously (Boyd *et al.*, 2013). Genotype data on participants was derived using the Illumina HumanHap550 quad genome-wide single nucleotide polymorphism (SNP) genotyping platform (Illumina Inc, San Diego, USA) by the Wellcome Trust Sanger Institute (WTSI, Cambridge, UK) and the Laboratory Corporation of America (LCA, Burlington, NC, USA). Samples were excluded based on the following criteria: incorrect sex assignment; abnormal heterozygosity (<0.320 or >0.345 for WTSI data; <0.310 or >0.330 for LCA data); high missingness ($>3\%$); cryptic relatedness ($>10\%$ identity by descent) and non-European ancestry (detected by multidimensional scaling analysis). After conducting quality control (QC), the final directly genotyped dataset contained 526,688 SNP loci.

Genotypes with minor allele frequency >0.01 and Hardy-Weinberg equilibrium $P > 5 \times 10^{-7}$ were firstly phased together using ShapeIt (version 2, revision 727) (Delaneau *et al.*, 2013), before undertaking imputation using Impute (v2.2.2) (Howie *et al.*, 2009), with a reference panel from the 1000 Genomes project (phase 1, version 3, phased using ShapeIt version 2, December 2013, using all populations). Subsequently, imputation dosages were converted to best-guess genotypes and filtered to only keep variants with an imputation quality score ≥ 0.8 . The final imputed dataset used for the analyses presented here contained 8,074,398 loci.

Cardiometabolic exposures. We selected eight measures of early-life adiposity and blood lipids from the ALSPAC study to analyse in this research. The measurements were taken from participants who attended the ALSPAC clinic at age 9 (mean age 9.9, range 8.8–11.7) and are detailed as follows. BMI was calculated using the equation $\text{weight}[\text{kg}]/\text{height}[\text{m}]^2$, with weight and height measured to the nearest 0.1kg and 0.1cm, respectively. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured while the participants were at rest using a Dinamap 9301 monitor. Two readings were taken for each, the mean of which was used in our analysis. Plasma lipid concentrations were calculated by taking non-fasting blood samples from the participants. High-density lipoprotein (HDL) cholesterol, total cholesterol and triglycerides were measured by modifying the standard Lipid Research Clinics Protocol with lipid determining reagents (Cooper *et al.*, 1988). LDL cholesterol was determined using the Friedewald equation (Friedewald *et al.*, 1972). Apolipoprotein A-I and apolipoprotein B were calculated using immunoturbidimetric assays (Roche).

Before undertaking analyses, cardiometabolic trait data were cleaned to identify outliers and to check distributions for normality. Outliers were removed from the analysis and were defined as any value four standard deviations (SD) greater or less than the mean. We applied log transformations to ensure normality when distributions were skewed. Individuals with withdrawn consent or those that had an older sibling in the dataset were removed. The mean, SD and sample size for each cleaned trait are listed in Supplementary Table 1 (*Underlying data*, O Nunain *et al.*, 2021a).

Statistical analysis

Genome-wide association study in the ALSPAC cohort. GWAS were conducted for each trait using PLINK v 2.0 software with adjustment for age and sex (Chang *et al.*, 2015). Adjustment for population ancestry is vital as population stratification can introduce confounding and produce spurious associations (Price *et al.*, 2006). Therefore, we repeated analyses for any identified GWAS hits with additional adjustment for the top 10 principal components to verify that our results were not affected by population stratification.

A p-value threshold of 5×10^{-8} was used to assess whether any of the associations reached conventional genome-wide significance corrections. An LD clumping cut-off of $r^2 < 0.001$ was applied to identify independent genetic variants using the 1000 Genomes reference panel. We then sought to evaluate the genetic effects of our lead results on adult measured traits by using findings from previously conducted GWAS in independent adult cohorts (Supplementary Table 2, *Underlying data*, O Nunain *et al.*, 2021a). These were the studies by (Kettunen *et al.*, 2016; Locke *et al.*, 2015; Richardson *et al.*, 2020c; Willer *et al.*, 2013). If the exact SNP was not present in these results, we used a proxy SNP based on $r^2 > 0.8$ using the same reference panel as before.

Gene set and functional analysis using tissue-specific and pathway datasets. We next evaluated whether findings from our GWAS in ALSPAC were enriched for functional tissue types and biological pathways. In doing so, we aimed to recapitulate findings from previous large-scale GWAS, in terms of the responsible tissue types and pathways which play a role in adiposity and lipid synthesis.

This was undertaken by running our results through the Functional Mapping and Annotation (FUMA) of GWAS bioinformatic tool (Watanabe *et al.*, 2017). FUMA was used to assess evidence of enrichment for differentially expressed gene sets using tissue-specific data from the GTEx consortium (v7) (GTEx Consortium *et al.*, 2017), and evaluate overrepresentations of associated genes on established biological pathways using data from the Reactome database (Fabregat *et al.*, 2017). We also used the Multi-marker Analysis of GenoMic Annotation (MAGMA) (de Leeuw *et al.*, 2015) approach to investigate associations between gene sets and each GWAS trait. This was to elucidate potentially overlooked association signals using single SNP analyses in the GWAS.

Genetic correlations with later life cardiometabolic disease. LD score regression was then undertaken to investigate the genetic correlation between our GWAS of early life risk factors and later life cardiometabolic outcomes (Bulik-Sullivan *et al.*, 2015b). These were coronary artery disease (CAD) (Nikpay *et al.*, 2015), type 2 diabetes (T2D) (Mahajan *et al.*, 2018), hypertension and hypercholesterolemia (Elsworth *et al.*, 2020). LD score regression was conducted using LDSC software (Bulik-Sullivan *et al.*, 2015a). The χ^2 values were calculated for each early life trait, and we only undertook LD score regression for exposures with a coefficient of 1.02 or higher. These guidelines are provided by the authors of this method, as they suggest that traits with values lower than this threshold may yield unreliable results (Bulik-Sullivan *et al.*, 2015a).

Results

Genome-wide association studies of childhood adiposity and lipoprotein lipids

Our GWAS analyses identified 14 unique loci associated with at least one measure of early life adiposity based on conventional genome-wide corrections ($P < 5 \times 10^{-8}$, Table 1). Repeating GWAS analyses with further adjustment for the top 10 principal components identified very little differences in the effect estimates for our top hits, with all their corresponding p-values remaining robust to $P < 5 \times 10^{-8}$ (Supplementary Table 3, *Underlying data*, O Nunain *et al.*, 2021a). Manhattan plots illustrating results for a selection of the cardiometabolic exposures analysed (BMI, triglycerides, apolipoprotein B and apolipoprotein A-I) can be found in Figure 1.

Results from this analysis included well established loci known to influence cardiometabolic traits in adulthood, such as *FTO* ($P = 1.25 \times 10^{-9}$) and *MC4R* ($P = 1.80 \times 10^{-9}$) associated with BMI, *CETP* ($P = 1.19 \times 10^{-67}$) associated with HDL cholesterol, *SORT1* ($P = 1.26 \times 10^{-12}$) and *FADS1* ($P = 4.16 \times 10^{-10}$) associated with LDL cholesterol, *APOA1* ($P = 4.02 \times 10^{-15}$) associated with apolipoprotein A-I, *APOB* ($P = 3.48 \times 10^{-14}$) associated with apolipoprotein B, *LPL* ($P = 8.71 \times 10^{-10}$) and *APOC3* ($P = 4.39 \times 10^{-12}$) associated with triglycerides and various other known lipid loci (including *LIPC*, *LIPG* and *APOE*). All the loci have also been identified previously in independent adult cohorts (Supplementary Table 4, *Underlying data*, O Nunain *et al.*, 2021a), suggesting that these loci begin to strongly exert their effects on adiposity and lipids traits in early life.

Functional enrichment and pathway analysis

Leveraging tissue-specific data from the GTEx consortium (v7) (GTEx Consortium *et al.*, 2017) using the FUMA platform (Watanabe *et al.*, 2017) suggested that the genes underlying our GWAS hits for lipoprotein lipids are expressed predominantly in liver tissue (Supplementary Figures 1–4, *Extended data*, O Nunain *et al.*, 2021b), as it was the most enriched tissue type for HDL cholesterol, triglycerides and apolipoprotein B. Liver tissue also provided the strongest evidence of enrichment for apolipoprotein A-I as depicted in Figure 2. Undertaking gene-set enrichment analyses using data from the Reactome

Table 1. Genome-wide association study results for measures of childhood adiposity. A summary of the genetic loci identified in the genome-wide association studies which reached the conventional p-value threshold of 5×10^{-8} . CHR - Chromosome, BP - base position, SE - standard error, P - p-value.

Trait	Lead SNP	CHR	BP	Gene	Effect allele	Other allele	Beta	SE	P
Apolipoprotein A-I	rs613808	11	116710968	<i>APOA1</i>	A	G	0.196064	0.0248607	4.02E-15
Apolipoprotein A-I	rs2070895	15	58723939	<i>LIPC</i>	A	G	0.207788	0.0273232	3.54E-14
Apolipoprotein A-I	rs17231506	16	56994528	<i>CETP</i>	T	C	0.289439	0.0229323	7.96E-36
Apolipoprotein A-I	rs77960347	18	47109955	<i>LIPG</i>	G	A	0.578242	0.101669	1.38E-08
Apolipoprotein B	rs7528419	1	109817192	<i>SORT1</i>	G	A	-0.209157	0.0267949	7.51E-15
Apolipoprotein B	rs580889	2	21290067	<i>APOB</i>	C	T	-0.216929	0.0285166	3.48E-14
Apolipoprotein B	rs174548	11	61571348	<i>FADS1</i>	G	C	-0.131635	0.0237988	3.39E-08
Apolipoprotein B	rs8107974	19	19388500	<i>TM6SF2</i>	T	A	-0.407674	0.0402857	8.79E-24
Apolipoprotein B	rs7412	19	45412079	<i>APOE</i>	T	C	-0.796266	0.039398	1.99E-86
Body mass Index	rs4477562	13	54104968	<i>OLFM4</i>	T	C	0.447695	0.0786661	1.33E-08
Body mass Index	rs55872725	16	53809123	<i>FTO</i>	T	C	0.321356	0.0528212	1.25E-09
Body mass Index	rs6567160	18	57829135	<i>MC4R</i>	C	T	0.376185	0.0624368	1.80E-09
HDL cholesterol	rs7946869	11	116963312	<i>APOA1</i>	T	C	0.162588	0.0289905	2.18E-08
HDL cholesterol	rs1077835	15	58723426	<i>LIPC</i>	G	A	0.192653	0.027557	3.19E-12
HDL cholesterol	rs17231506	16	56994528	<i>CETP</i>	T	C	0.40379	0.0227858	1.19E-67
HDL cholesterol	rs6857	19	45392254	<i>APOE</i>	T	C	-0.171339	0.0308595	3.01E-08
LDL cholesterol	rs599839	1	109822166	<i>SORT1</i>	G	A	-0.192309	0.0269997	1.26E-12
LDL cholesterol	rs580889	2	21290067	<i>APOB</i>	C	T	-0.199764	0.02869	3.89E-12
LDL cholesterol	rs174548	11	61571348	<i>FADS1</i>	G	C	-0.149623	0.0238879	4.16E-10
LDL cholesterol	rs58542926	19	19379549	<i>TM6SF2</i>	T	C	-0.389789	0.0410813	3.94E-21
LDL cholesterol	rs7412	19	45412079	<i>APOE</i>	T	C	-0.718021	0.0400997	6.04E-69
Triglycerides	rs11984636	8	19885726	<i>LPL</i>	C	T	-0.221942	0.0361091	8.71E-10
Triglycerides	rs2072560	11	116661826	<i>APOC3</i>	T	C	0.332775	0.0479112	4.39E-12
Triglycerides	rs584007	19	45416478	<i>APOC1</i>	A	G	-0.159748	0.0236322	1.59E-11

database (Fabregat *et al.*, 2017) suggested that these genes were overrepresented on pathways involved in the metabolism and transport of lipids and lipoproteins (Supplementary Figures 5–9, *Extended data*, O Nunain *et al.*, 2021b). We were unable to run enrichment analyses for the BMI GWAS results due to only one SNP being reliably mapped to a gene by FUMA.

Conducting gene-based tests on our results using MAGMA (Multi-marker Analysis of GenoMic Annotation) (de Leeuw *et al.*, 2015) identified evidence of association for additional genetic loci which did not meet GWAS corrections in the initial analysis (Supplementary Figures 10–15, *Extended data*, O Nunain *et al.*, 2021b). These included *ADCY3* which was associated with BMI, which encodes adenylate cyclase 3 and

has been previously implicated in obesity risk through loss-of-function studies (Grarup *et al.*, 2018). Likewise, we identified evidence that *HMGCR*, the therapeutic target for statin inhibitors, was associated with LDL cholesterol in our dataset based on age 10 individuals from the ALSPAC cohort.

Assessing genome-wide genetic correlations between childhood adiposity and lipids with later life cardiometabolic disease

BMI, SBP, triglycerides and apolipoprotein B provided χ^2 values > 1.02 and were eligible for genetic correlation analyses (Supplementary Table 5, *Underlying data*, O Nunain *et al.*, 2021a). Applying LD score regression suggested that our results for childhood BMI were genetically correlated with later life CAD

Figure 1. Manhattan plots for body mass index, triglycerides, apolipoprotein B and apolipoprotein A-I. Manhattan plots for genome-wide association studies of early life measures of **A)** body mass index, **B)** triglycerides, **C)** apolipoprotein B and **D)** apolipoprotein A-I. The red dashed line indicates the conventional genome-wide correction threshold of $P < 5 \times 10^{-8}$.

Figure 2. Tissue enrichment for apolipoprotein A-I. Evidence of enrichment for genetic variants robustly associated with apolipoprotein A-I levels in childhood and tissue-specific gene expression derived from human samples. The red dotted line represents the multiple testing adjusted P value threshold.

($rG=0.19$, 95% CI=0.03 to 0.35, $P=0.02$), T2D ($rG=0.35$, 95% CI=0.18 to 0.51, $P=3.28 \times 10^{-5}$) and hypertension ($rG=0.20$, 95% CI=0.07 to 0.32, $P=0.002$). Similar results were found for childhood SBP; CAD ($rG=0.26$, 95% CI=0.07 to 0.46, $P=0.009$), T2D ($rG=0.30$, 95% CI=0.15 to 0.45, $P=1.00 \times 10^{-4}$) and hypertension ($rG=0.37$, 95% CI=0.19 to 0.55, $P=6.57 \times 10^{-5}$).

There was weak evidence of a genetic correlation between childhood triglycerides and apolipoprotein B with later-life disease outcomes (Supplementary Table 6, *Underlying data*, [O Nunain et al., 2021a](#)). In particular, the wide confidence intervals for apolipoprotein B is likely attributed to the sample size of our GWAS. As such, there were central correlation estimates, which despite being high (e.g. $rG=0.58$ for hypercholesterolemia), lacked the precision to conclude strong evidence of a genetic correlation. A forest plot of all results from LD score regression analyses can be found in [Figure 3](#).

Discussion

In this study we provide evidence that there are genetic variants associated with adiposity and lipoprotein lipids which begin to exert their effects as early as age 10 in the life course. The variants robustly associated with lipoprotein lipid traits were enriched for genetic loci whose genes are predominantly

expressed in liver tissue and overrepresented on lipid synthesis pathways, supporting their validity as genuine biological effects. Furthermore, we identified strong evidence of genetic correlations between childhood BMI and SBP with later life cardiometabolic disease outcomes.

Our genome-wide association study in a population of young individuals suggested that genetic variation at 14 unique loci has an influence on adiposity and dyslipidaemia even before reaching puberty. Amongst our hits were well-known cardiometabolic loci previously identified in cohorts of adults, such as *FTO* ($P=1.25 \times 10^{-9}$ with BMI), *MC4R* ($P=1.80 \times 10^{-9}$ with BMI), *LPL* ($P=8.71 \times 10^{-10}$ with triglycerides), *CETP* ($P=1.19 \times 10^{-67}$ with HDL cholesterol) and *SORT1* ($P=1.26 \times 10^{-12}$ with LDL cholesterol). Moreover, the association signals at the *APOA1* locus with apolipoprotein A-I ($P=4.02 \times 10^{-15}$) and the *APOB* locus with apolipoprotein B ($P=3.48 \times 10^{-14}$) are very likely real biological effects given that they reside at the coding genes responsible for these lipid-related proteins ([Zannis et al., 2001](#)). The early influence of *APOB* on apolipoprotein B levels is of particular interest from a cardiovascular disease prevention perspective, given that there is increasing evidence highlighting the crucial role it plays in coronary heart disease risk ([Holmes & Ala-Korpela, 2019](#); [Richardson et al., 2020c](#)).

Figure 3. Genetic correlations between early life cardiometabolic risk factors and later life disease outcomes. Forest plots for the linkage disequilibrium (LD) score regression results between early life cardiometabolic risk factors and later life disease outcomes. Genetic correlation coefficients and confidence intervals are shown on the right-hand side. Diastolic blood pressure, high density lipoprotein cholesterol, low density lipoprotein cholesterol and apolipoprotein A-I were not included in this analysis due to having a mean $\chi^2 < 1.02$ suggesting that their correlations may be unreliable.

Additional downstream analysis and evaluation of our GWAS results provided evidence to reinforce that the genetic variants begin to exert their effects very early in the life course. For instance, using tissue-specific gene expression data from the GTEx consortium suggested that the genes underlying our GWAS hits for lipid-related traits are predominantly expressed in liver tissue. This fits with the biology concerning these traits, given that many lipid and lipoproteins are hepatically synthesised (Dietschy *et al.*, 1993). Moreover, we identified evidence of an overrepresentation of these genes on curated biological pathways concerning lipid metabolism and transport.

Gene-based analyses provided evidence of additional genes not identified in the single SNP GWAS. These included *ADCY3*, which was associated with BMI and previously shown through loss-of-function analyses to influence obesity and T2D risk (Grarup *et al.*, 2018). There was also evidence of association between genetic variation at *HMGCR* and LDL cholesterol, which is well-established given that the protein product of this gene (HMG-CoA reductase) is pharmacologically inhibited by statin therapy to lower cholesterol levels (Cholesterol Treatment Trialists' (CTT) Collaboration *et al.*, 2010). These findings highlight how early in the life course these genetic effects begin to influence weight and lipid-related traits, suggesting that a long window of opportunity exists to lower cardiovascular disease risk through lifestyle modifications.

To our knowledge, no previous studies have investigated the genetic correlation between childhood blood pressure and lipoprotein lipids with cardiometabolic disease in adulthood. Despite our GWAS sample sizes being modest, we found evidence for a genetic overlap between childhood SBP with coronary heart disease and hypertension in later life. Furthermore, there was strong evidence of a genetic correlation between childhood BMI and T2D, a result that supports recent findings (Tekola-Ayele *et al.*, 2019; Vogelesang *et al.*, 2020). The genetic correlation between childhood SBP and T2D we identified may be attributed to the vertical pleiotropy which exists between BMI and SBP (i.e. high BMI raising blood pressure levels) (Wade *et al.*, 2018).

A shared genetic basis may partially explain the association between childhood BMI and later life cardiometabolic disease seen in observational studies (Reilly & Kelly, 2011). However, given recent evidence, it is likely that childhood adiposity influences adulthood disease risk due to its persistent effect throughout the life course (Juonala *et al.*, 2011). Although Mendelian randomization studies have been undertaken to support this for childhood adiposity (Richardson *et al.*, 2020b; Richardson *et al.*, 2020a), future research is required to investigate the direct and indirect effects of childhood blood pressure and lipoprotein lipid traits on later life disease risk. Sufficiently powered sample sizes for these traits in the future will likely facilitate such endeavours, allowing a large number of robustly associated genetic variants to be used as instrumental variables.

In terms of study limitations, the relatively modest sample size of our childhood GWAS (in comparison to modern standards)

limited the statistical power of our study, and hence our ability to detect associations. It is likely this is the reason we didn't observe any SNP associations for SBP after adjusting for conventional multiple-testing corrections applied in GWAS (i.e. $P < 5 \times 10^{-8}$), although a recent high-powered GWAS, of which ALSPAC was a participating study, identified two SNPs associated with SBP in childhood (Parmar *et al.*, 2016). Furthermore, the modest sample size of the GWAS also limited the power of our downstream analyses, particularly the LD score regression which is indicated by the low χ^2 values of several traits.

In conclusion, our findings suggest that future GWAS endeavours should focus on traits during childhood to elucidate variants which have lifelong effects. These will also pave the way for Mendelian randomization analyses to disentangle the contribution of early life exposures to disease risk, independent of the same exposures measured in adulthood. Doing so can help discern whether genetic correlations between childhood traits and disease outcomes, such as those identified in our study, are due to either a direct or indirect effect of early-life risk factors.

Data availability

Underlying data

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 article, and all other ALSPAC data:

- Please read the [ALSPAC access policy](#) which describes the process of accessing the data and samples in detail, and outlines the costs associated with doing so.
- 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.
- 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.

Figshare: Supplementary tables for a genome-wide association study of childhood adiposity and blood lipids, <https://doi.org/10.6084/m9.figshare.15134409.v3> (O Nunain *et al.*, 2021a)

This project contains the following underlying data:

- Supplementary Table 1: Trait characteristics from the ALSPAC cohort at mean age 9.9
- Supplementary Table 2: Dataset of adult populations used in this study to evaluate genetic effects identified in ALSPAC
- Supplementary Table 3: Genome-wide association study results for measures of childhood adiposity adjusted for population stratification
- Supplementary Table 4: Evaluation of genome-wide association study hits in adult populations
- Supplementary Table 5: χ^2 coefficients for each childhood exposure to assess eligibility for genetic correlation analyses

- Supplementary Table 6: Linkage disequilibrium score regression results

Extended data

Figshare: Extended data for a genome-wide association study of childhood adiposity and blood lipids, <https://doi.org/10.6084/m9.figshare.15172824.v3> (O Nunnain *et al.*, 2021b)

This project contains the following:

- Supplementary figures for a genome-wide association study of childhood adiposity and blood lipids

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer

and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data were generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

This research was conducted at the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. This publication is the work of the authors and TGR will serve as guarantor for the contents of this paper.

References

- Ayer J, Charakida M, Deanfield JE, *et al.*: **Lifetime risk: childhood obesity and cardiovascular risk.** *Eur Heart J.* 2015; **36**(22): 1371–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Baker JL, Olsen LW, Sorensen TI: **Childhood body-mass index and the risk of coronary heart disease in adulthood.** *N Engl J Med.* 2007; **357**(23): 2329–37.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bjerregaard LG, Baker JL: **Change in Overweight from Childhood to Early Adulthood and Risk of Type 2 Diabetes.** *N Engl J Med.* 2018; **378**(26): 2537–2538.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Boyd A, Golding J, Macleod J, Lawlor DA, *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): 111–27.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bradfield JP, Vogelzang S, Felix JF, *et al.*: **A Trans-ancestral Meta-Analysis of Genome-Wide Association Studies Reveals Loci Associated with Childhood Obesity.** *Hum Mol Genet.* 2019; **28**(19): 3327–3338.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bulik-Sullivan B, Finucane HK, Anttila V, *et al.*: **An atlas of genetic correlations across human diseases and traits.** *Nat Genet.* 2015a; **47**(11): 1236–41.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bulik-Sullivan B, Finucane HK, Anttila V, *et al.*: **LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.** *Nat Genet.* 2015b; **47**(3): 291–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chang CC, Chow CC, Tellier LC, *et al.*: **Second-generation PLINK: rising to the challenge of larger and richer datasets.** *Gigascience.* 2015; **4**: 7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cholesterol Treatment Trialists' (CTT) Collaboration; Baigent C, Blackwell L, *et al.*: **Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials.** *Lancet.* 2010; **376**(9753): 1670–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- GBD 2015 Obesity Collaborators; Afshin A, Forouzanfar MH, *et al.*: **Health Effects of Overweight and Obesity in 195 Countries over 25 Years.** *N Engl J Med.* 2017; **377**(1): 13–27.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; *et al.*: **Genetic effects on gene expression across human tissues.** *Nature.* 2017; **550**(7675): 204–213.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cooper GR, Myers GL, Smith SJ, *et al.*: **Standardization of lipid, lipoprotein, and apolipoprotein measurements.** *Clin Chem.* 1988; **34**(8B): B95–105.
[PubMed Abstract](#)
- de Leeuw CA, Mooij JM, Heskes T, *et al.*: **MAGMA: generalized gene-set analysis of GWAS data.** *PLoS Comput Biol.* 2015; **11**(4): e1004219.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Delaneau O, Howie B, Cox AJ, *et al.*: **Haplotype estimation using sequencing reads.** *Am J Hum Genet.* 2013; **93**(4): 687–96.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dietschy JM, Turley SD, Spady DK: **Role of liver in the maintenance of cholesterol and low density lipoprotein homeostasis in different animal species, including humans.** *J Lipid Res.* 1993; **34**(10): 1637–59.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Elsworth B, Lyon M, Alexander T, *et al.*: **The MRC IEU OpenGWAS data infrastructure.** *bioRxiv.* 2020.
[Publisher Full Text](#)
- Fabregat A, Sidiropoulos K, Viteri G, *et al.*: **Reactome pathway analysis: a high-performance in-memory approach.** *BMC Bioinformatics.* 2017; **18**(1): 142.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Felix JF, Bradfield JP, Monnereau C, *et al.*: **Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index.** *Hum Mol Genet.* 2016; **25**(2): 389–403.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fraser A, Macdonald-Wallis C, Tilling K, *et al.*: **Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.** *Int J Epidemiol.* 2013; **42**(1): 97–110.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Friedewald WT, Levy RI, Fredrickson DS: **Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge.** *Clin Chem.* 1972; **18**(6): 499–502.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Grarup N, Moltke I, Andersen MK, *et al.*: **Loss-of-function variants in *ADCY3* increase risk of obesity and type 2 diabetes.** *Nat Genet.* 2018; **50**(2): 172–174.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hemani G, Yang J, Vinkhuyzen A, *et al.*: **Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs.** *Am J Hum Genet.* 2013; **93**(5): 865–75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Holmes MV, Ala-Korpela M: **What is 'LDL cholesterol'?** *Nat Rev Cardiol.* 2019; **16**(4): 197–198.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Howie BN, Donnelly P, Marchini J: **A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.** *PLoS Genet.* 2009; **5**(6): e1000529.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Juonala M, Magnussen CG, Berenson GS, *et al.*: **Childhood adiposity, adult adiposity, and cardiovascular risk factors.** *N Engl J Med.* 2011; **365**(20): 1876–85.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kettunen J, Demirkan A, Wurtz P, *et al.*: **Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA.** *Nat*

Commun. 2016; 7: 11122.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Locke AE, Kahali B, Berndt SI, *et al.*: **Genetic studies of body mass index yield new insights for obesity biology.** *Nature.* 2015; **518**(7538): 197–206.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Mahajan A, Taliun D, Thurner M, *et al.*: **Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps.** *Nat Genet.* 2018; **50**(11): 1505–1513.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Newman TB, Browner WS, Hulley SB: **The case against childhood cholesterol screening.** *JAMA.* 1990; **264**(23): 3039–43.

[PubMed Abstract](#) | [Publisher Full Text](#)

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): 1121–1130.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

O Nunain K, Sanderson E, Holmes M, *et al.*: **Supplementary tables for a genome-wide association study of childhood adiposity and blood lipids.** *figshare.* 2021a; Dataset.

<http://www.doi.org/10.6084/m9.figshare.15134409.v3>

O Nunain K, Sanderson E, Holmes M, *et al.*: **Extended data for a genome-wide association study of childhood adiposity and blood lipids.** *figshare.* 2021b.

<http://www.doi.org/10.6084/m9.figshare.15172824.v3>

Parmar PG, Taal HR, Timpson NJ, *et al.*: **International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents.** *Circ Cardiovasc Genet.* 2016; **9**(3): 266–278.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Price AL, Patterson NJ, Plenge RM, *et al.*: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet.* 2006; **38**(8): 904–9.

[PubMed Abstract](#) | [Publisher Full Text](#)

Pulgaron ER, Delamater AM: **Obesity and type 2 diabetes in children: epidemiology and treatment.** *Curr Diab Rep.* 2014; **14**(8): 508.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Reilly JJ, Kelly J: **Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review.** *Int J Obes (Lond).* 2011; **35**(7): 891–8.

[PubMed Abstract](#) | [Publisher Full Text](#)

Richardson TG, Mykkanen J, Pahkala K, *et al.*: **Evaluating the direct effects of childhood adiposity on adult systemic metabolism: A multivariable Mendelian randomization analysis.** *medRxiv.* 2020a.

[Publisher Full Text](#)

Richardson TG, Sanderson E, Elsworth B, *et al.*: **Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study.** *BMJ.* 2020b; **369**: m1203.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Richardson TG, Sanderson E, Palmer T M, *et al.*: **Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis.** *PLoS Med.* 2020c; **17**(3): e1003062.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Robinson MR, English G, Moser G, *et al.*: **Genotype-covariate interaction effects and the heritability of adult body mass index.** *Nat Genet.* 2017; **49**(8): 1174–1181.

[PubMed Abstract](#) | [Publisher Full Text](#)

Tekola-Ayele F, Lee A, Workalemahu T, *et al.*: **Shared genetic underpinnings of childhood obesity and adult cardiometabolic diseases.** *Hum Genomics.* 2019; **13**(1): 17.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Vander Wal JS, Mitchell ER: **Psychological complications of pediatric obesity.** *Pediatr Clin North Am.* 2011; **58**(6): 1393–401.

[PubMed Abstract](#) | [Publisher Full Text](#)

Vogelezang S, Bradfield JP, Ahluwalia TS, *et al.*: **Novel loci for childhood body mass index and shared heritability with adult cardiometabolic traits.** *PLoS Genet.* 2020; **16**(10): e1008718.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wade KH, Chiesa ST, Hughes AD, *et al.*: **Assessing the causal role of body mass index on cardiovascular health in young adults: Mendelian randomization and recall-by-genotype analyses.** *Circulation.* 2018; **138**(20): 2187–2201.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wang YC, McPherson K, Marsh T, *et al.*: **Health and economic burden of the projected obesity trends in the USA and the UK.** *Lancet.* 2011; **378**(9793): 815–25.

[PubMed Abstract](#) | [Publisher Full Text](#)

Watanabe K, Taskesen E, van Bochoven A, *et al.*: **Functional mapping and annotation of genetic associations with FUMA.** *Nat Commun.* 2017; **8**(1): 1826.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Willer CJ, Schmidt EM, Sengupta S, *et al.*: **Discovery and refinement of loci associated with lipid levels.** *Nat Genet.* 2013; **45**(11): 1274–1283.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zannis VI, Kan HY, Kritis A, *et al.*: **Transcriptional regulatory mechanisms of the human apolipoprotein genes *in vitro* and *in vivo*.** *Curr Opin Lipidol.* 2001; **12**(2): 181–207.

[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 24 January 2022

<https://doi.org/10.21956/wellcomeopenres.18679.r47034>

© 2022 Jones S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Samuel Jones

Institute for Molecular Medicine (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland

The authors report on a series of eight GWAS of adipose- and lipid-related traits (BMI, triglycerides, LDL- and HDL-cholesterol, Apolipoproteins A-I and B, and systolic and diastolic blood pressure) in a cohort of approximately 5,000 participants of European ancestry. Despite the limited sample size, the authors report the identification of 24 genome-wide significant genetic associations in 14 unique loci across their eight phenotypes. Amongst the identified loci were those previously associated in later-life GWAS for the equivalent traits, such as FTO and MC4R (BMI), APOA1, and APOB (apolipoproteins A-I and B, respectively) and others. The authors follow up their findings by firstly interrogating the associations through genetic correlation with later-life cardiometabolic phenotypes, finding moderate but significant genetic overlap between early-life systolic blood pressure and later-life CAD and hypertension and early-life BMI with later-life type 2 diabetes, though report low heritability estimates for the early-life phenotypes. This was followed by gene-set enrichment analysis in an attempt to understand the biological mechanisms in which the identified genetic variants were implicated, with genes involved in metabolism and lipid transport pathways and those expressed in liver tissue showing evidence of being enriched. The authors conclude that their results demonstrate the ability to detect the early effect of genetic factors on adipose and lipid traits and that further work should be undertaken to understand the effects of (more acute) early-life exposure versus the cumulative (chronic) effect of life-long exposure to these genetic factors.

I feel this manuscript is an important addition to the literature on early-life traits and am pleased to see that focus is not just on trying to replicate findings from later-life GWAS for the equivalent phenotypes. That being said, more could be done to help the reader understand whether the genetics of these early-life phenotypes really are distinct from the genetics identified in later-life GWAS. I have a few suggestions and questions that I feel need to be addressed before the manuscript should be accepted.

Major Comments

- In the results and discussion sections, it is mentioned that well-known loci were seen, but did the variants identified represent the same signal as in the later-life GWAS? If the variants

aren't the same, what is the LD between your variant and the previously reported one? I'm not sure if you can qualify your discussion of overlapping signals unless we know whether the lead variants are in LD.

- An obvious question is: "How genetically correlated are early-life phenotypes with later-life phenotypes?". I understand that the early-life heritability estimates are low, given the small sample sizes, but it would help contextualise the genetic correlations with later-life cardiometabolic phenotypes that you do report.
- The discussion mentions that the signals at the APOA1 and APOB loci are "very likely real biological effects given that they reside at the coding genes...", but what are the functions of the lead variants at these loci? Are they coding variants within the genes or are they within known eQTLs for these genes? If so, this is definitely worth including in the results/discussion. If not, I don't know if you can claim that they are "likely real biological effects", unless there is other evidence to link these variants to the specific genes.
- Is there a reason that 1) related samples were removed and 2) genotypes were converted to best-guess for GWA analysis? In an ideal situation, I would recommend rerunning the GWAS software that handles related samples and imputed data – is this a possibility?
- It is good practice to make GWAS summary statistics available for use by the wider scientific community, but in your Data Availability section, I don't see any mention of accessing these. Are you planning to make these available? If so, please make it clear how to access them. If not, what are the justifications for not making these available?

Minor Comments

- Would it be possible to add the N of the largest GWAS to the methods section of the abstract, to give the reader a better idea of the cohort size without having to delve into the manuscript?
- If there is space, perhaps add a sentence in the background section of the abstract on why elucidating the genetics is important for these traits?
- In Table 1 and Supp Tables 3 and 4, can you make it clear which genome build the positions are in? This is incredibly useful when other researchers come to use your published results.
- Could you clarify how the genes were identified for each locus? Were they the nearest genes? Or are these the genes mapped using FUMA GWAS?
- Would it be possible to highlight whether the lead variants identified are intergenic, intronic, exonic, etc.?
- In the limitations, where the study that identified two SNPs associated with childhood SBP is mentioned, can you add the sample size of that study to provide some context to their findings in relation to yours? Did you see even nominal associations for these reported variants in your results? Please report negative findings too!

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?

Partly

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?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistical genetics, genetic epidemiology, population genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 11 Mar 2023

Tom Richardson

Reviewer #2 The authors report on a series of eight GWAS of adipose- and lipid-related traits (BMI, triglycerides, LDL- and HDL-cholesterol, Apolipoproteins A-I and B, and systolic and diastolic blood pressure) in a cohort of approximately 5,000 participants of European ancestry. Despite the limited sample size, the authors report the identification of 24 genome-wide significant genetic associations in 14 unique loci across their eight phenotypes. Amongst the identified loci were those previously associated in later-life GWAS for the equivalent traits, such as FTO and MC4R (BMI), APOA1, and APOB (apolipoproteins A-I and B, respectively) and others. The authors follow up their findings by firstly interrogating the associations through genetic correlation with later-life cardiometabolic phenotypes, finding moderate but significant genetic overlap between early-life systolic blood pressure and later-life CAD and hypertension and early-life BMI with later-life type 2 diabetes, though report low heritability estimates for the early-life phenotypes. This was followed by gene-set enrichment analysis in an attempt to understand the biological mechanisms in which the identified genetic variants were implicated, with genes involved in metabolism and lipid transport pathways and those expressed in liver tissue showing evidence of being enriched. The authors conclude that their results demonstrate the ability to detect the early effect of genetic factors on adipose and lipid traits and that further work should be undertaken to understand the effects of (more acute) early-life exposure versus the cumulative (chronic)

effect of life-long exposure to these genetic factors. I feel this manuscript is an important addition to the literature on early-life traits and am pleased to see that focus is not just on trying to replicate findings from later-life GWAS for the equivalent phenotypes. That being said, more could be done to help the reader understand whether the genetics of these early-life phenotypes really are distinct from the genetics identified in later-life GWAS. I have a few suggestions and questions that I feel need to be addressed before the manuscript should be accepted. **Major Comments**

- In the results and discussion sections, it is mentioned that well-known loci were seen, but did the variants identified represent the same signal as in the later-life GWAS? If the variants aren't the same, what is the LD between your variant and the previously reported one? I'm not sure if you can qualify your discussion of overlapping signals unless we know whether the lead variants are in LD.

To clarify, evaluations of the effect estimates in later-life GWAS reported in Table S4 are the exact same variant as the ones found to be lead markers in ALSPAC analyses. Therefore, LD is not an issue for these look-ups. This has been clarified on page 8. "All the loci have also been identified previously in independent adult cohorts (effect estimates for lead variants found Supplementary Table 4, Underlying data, O Nunain et al., 2021a), suggesting that these loci begin to strongly exert their effects on adiposity and lipids traits in early life."

- An obvious question is: "How genetically correlated are early-life phenotypes with later-life phenotypes?". I understand that the early-life heritability estimates are low, given the small sample sizes, but it would help contextualise the genetic correlations with later-life cardiometabolic phenotypes that you do report.

As suggested by reviewer #1, we have conducted a polygenic risk score analysis to evaluate genetic correlations between childhood and adulthood phenotypes. This is now reported on page 8: "Additionally, generating whole genome polygenic risk scores in the UKB using estimates derived from ALSPAC analyses found strong evidence of association for all 8 traits (Supplementary Table 5, Underlying data O Nunain et al., 2021a), suggesting a high level of genetic correlation between their measured obtained during childhood and adulthood."

- The discussion mentions that the signals at the APOA1 and APOB loci are "very likely real biological effects given that they reside at the coding genes...", but what are the functions of the lead variants at these loci? Are they coding variants within the genes or are they within known eQTLs for these genes? If so, this is definitely worth including in the results/discussion. If not, I don't know if you can claim that they are "likely real biological effects", unless there is other evidence to link these variants to the specific genes.

We have now added VEP annotations to Table 1 as also recommended by reviewer #1.

- Is there a reason that 1) related samples were removed and 2) genotypes were converted to best-guess for GWA analysis? In an ideal situation, I would recommend rerunning the GWAS software that handles related samples and imputed data – is this a possibility?

GWAS data in ALSPAC has been prepared internally by the cohort and provided to researchers in the current format to ensure the reproducibility of results. The changes suggested would therefore require an updated application to ALSPAC.

- It is good practice to make GWAS summary statistics available for use by the wider scientific community, but in your Data Availability section, I don't see any mention of

accessing these. Are you planning to make these available? If so, please make it clear how to access them. If not, what are the justifications for not making these available?

Many thanks for this suggestion. We have now uploaded our full summary statistics to the GWAS catalog (accession numbers GCST90104677 to GCST90104684) as mentioned on page 13 of the manuscript: "The full set of summary statistics for the 8 GWAS conducted in this study can be found on the GWAS catalog (accession numbers GCST90104677 to GCST90104684)" Minor Comments

- Would it be possible to add the N of the largest GWAS to the methods section of the abstract, to give the reader a better idea of the cohort size without having to delve into the manuscript?

We have added sample sizes to the abstract as requested.

- If there is space, perhaps add a sentence in the background section of the abstract on why elucidating the genetics is important for these traits?

We have now added the following sentence to the abstract: "Previous studies have found that genetic variants inherited at birth can begin to exert their effects on cardiometabolic traits during the early stages of the lifecourse."

- In Table 1 and Supp Tables 3 and 4, can you make it clear which genome build the positions are in? This is incredibly useful when other researchers come to use your published results.

We have now clarified that results are reported on the hg19 build of the human genome as recommended in these tables.

- Could you clarify how the genes were identified for each locus? Were they the nearest genes? Or are these the genes mapped using FUMA GWAS?

Genes were mapped at each locus based on previous GWAS published in the literature and functional follow-up studies of these loci.

- Would it be possible to highlight whether the lead variants identified are intergenic, intronic, exonic, etc.?

VEP annotations have now been added to Table 1 to address this point.

- In the limitations, where the study that identified two SNPs associated with childhood SBP is mentioned, can you add the sample size of that study to provide some context to their findings in relation to yours? Did you see even nominal associations for these reported variants in your results? Please report negative findings too!

We have added the sample size of this study to the discussion (n=8,423) as well as providing a look up for this SNP in our own study (page 12). Previously we reported that two variants surpassed genome-wide corrections, although only one of these was based on blood pressure measured at puberty (as in our study). "A previous GWAS (N = 8,423), of which ALSPAC was a participating study, identified one SNPs associated with SBP at puberty (rs872256, $P=8.7 \times 10^{-9}$) ([Parmar et al., 2016](#)), which did not reach genome-wide corrections in ALSPAC alone ($P=6.4 \times 10^{-5}$ in this study)."

Competing Interests: No competing interests were disclosed.

Reviewer Report 07 December 2021

<https://doi.org/10.21956/wellcomeopenres.18679.r47032>

© 2021 Rajagopal V. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Veera Rajagopal

Department of Biomedicine, Aarhus University, Aarhus, Denmark

O'Nunai *et al.* has performed genome-wide association studies (GWASs) for eight cardio-metabolic traits—body mass index (BMI), systolic and diastolic blood pressure (SBP and DBP), high-density and low-density lipoprotein cholesterol (HDL and LDL), triglycerides (TGL), apolipoprotein A-1 and B (apo-A1, apo-B)—in ~5000 children from ALSPAC cohort. Although the sample size is smaller by many orders of magnitude compared to other existing GWASs, the study is interesting as it evaluates the genetic influences of these cardio-metabolic traits in children as opposed to most other studies that studied mainly adults (with the exception of studies of childhood BMI). The authors report the results following a conventional style. As expected many of the known suspects (e.g. FTO, MC4R, APOE, etc.) show up beautifully in the GWASs, highlighting their strong genetic effects. Gene set enrichment analysis implicates disease relevant tissues and pathways and genetic correlation analyses suggest genetic variants influencing cardio-metabolic quantitative traits in children are the same that influence the risk for cardio-metabolic diseases in adults.

Given the major—and perhaps the only—strength of this study is that these phenotypes are measured in children, I'd report the results slightly differently. The main questions, as the authors discuss in the paper, to ask in such a sample are

1. Do the genetic variants that influence cardio-metabolic traits and diseases in adulthood also influence in childhood? (The answer to this question is often yes unless there is a strong biological argument to suggest otherwise)
2. Do the effect sizes of these risk variants differ between childhood and adulthood?

I am not sure if the current version of the paper answers these questions clearly. I recommend the following revisions to improve the manuscript so that it answers the key questions mentioned above.

Variant level associations:

In the current version, the authors report only loci significant above conventional genome-wide significant threshold ($5e-8$). However, I'd not consider the current analysis as discovery in nature, given that the sample size is too small and there exist GWASs for these traits in very large sample sizes. Reporting genome-wide hits is okay. But a better way to report variant associations is to first take all the variants that are reported as genome-wide significant in the most recent GWASs of each of the eight traits and evaluate their significance in the current sample. The P value threshold can be set based on the number of variants being evaluated. We'd expect only those variants with higher statistical power will replicate in the current study. That is, those variants with large effect size and rare MAF or with moderate effect size and common MAF. This can be visualised using an allele frequency vs effect size plot. For an idea, please refer to figure 3 from the recent preprint from global biobank meta-analysis initiative (Zhou *et al.*, MedRxiv, 2021).¹ Reporting such a plot will be very informative and educational for the readers. When replicated and non-replicated variants were differentiated by shape (color differentiating the traits), we would see all the

replicated variants falling within the centre zone within a U shape. This kind of visual inspection is important because—firstly, by reproducing the expected pattern it ensures that the analyses were performed properly and secondly, it helps identify outliers that deviate from the expected pattern (e.g. if a variant with sufficient power does not show a significant association) and study them further. Such outliers are the ones that likely have different effects in childhood vs adulthood.

Additionally, I recommend to compare the effect sizes (standardised betas) of those variants that replicate between childhood and adulthood. Perhaps a scatter plot with effect sizes reported in adult sample GWASs in X axis and effect sizes observed in the current sample in Y axis. Any outliers in this scatter plot might be interesting candidates to study further as they will correspond to variants with differential effects between childhood and adulthood.

MAGMA gene based analysis and tissue specific enrichment:

Gene based analyses and tissue specific enrichment analysis using FUMA do not add anything new and also in such a small sample size I wouldn't do these analyses. Removing these altogether or reporting them in supplementary will help the readers to focus only on the main findings.

Genetic correlation analysis:

LD score regression based genetic correlation analysis between two traits, say A and B, requires adequate sample sizes for both A and B GWASs. Hence, not an ideal analysis for the GWASs reported in the current paper. An alternative would be to perform a polygenic score analysis and report the betas and P values as we have GWAS for these traits in UK Biobank in huge sample sizes that will serve as training samples and will offer better power to detect genetic associations. It would be more informative if the authors could perform a similar analysis also in a set of adult samples (perhaps a small chunk of UK Biobank sample kept out of the training) and compare the effect sizes between childhood and adulthood. If the polygenic score analysis could not be performed for some reason, I recommend that at least the authors report the LDSC r_g for both child and adult GWASs. Otherwise, the genetic correlation analysis results will offer no insight to the readers.

Minor comments

1. Please provide the sample size in abstract, methods, results and in the main tables. When you report genome-wide significant variants as a table, it is essential that it also has an N column. It is not fair to expect the readers to go to supplementary tables to learn this crucial piece of information.
2. I recommend the authors to make the full summary statistics publicly available for the readers.

References

1. Zhou W, Kanai M, Wu K, Humaira R, et al.: Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. *medRxiv*. 2021. [Publisher Full Text](#)

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?

Partly

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

No

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: GWAS, statistical genetics and psychiatric genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.