OXFORD

Genetics and Environment

Genetic variation and elevated liver enzymes during childhood, adolescence and early adulthood

Stefan Stender [ID] ,[1]* **George Davey Smith** [ID] [2] **and Tom G Richardson** [ID] [2]

[1]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark and [2]MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, University of Bristol, Bristol, UK

*Corresponding author. Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Blegdamsvej 9, 2100 Copenhagen, Denmark. E-mail: Stefan.stender@regionh.dk

Abstract

Background: Genetic factors influence the risk of fatty liver disease (FLD) in adults. The aim of this study was to test if, and when, genetic risk factors known to affect FLD in adults begin to exert their deleterious effects during childhood, adolescence and early adulthood.

Methods: We included up to 4018 British children and adolescents from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. Three genetic variants known to associate robustly with FLD in adults (*PNPLA3* rs738409, *TM6SF2* rs58542926 and *HSD17B13* rs72613567) were tested for association with plasma levels of alanine transaminase (ALT) and aspartate transaminase (AST) during childhood (mean age: 9.9 years), early adolescence (15.5 years), late adolescence (17.8 years), and early adulthood (24.5 years). We also tested the associations of a 17-variant score and whole-genome polygenic risk scores (PRS) derived from associations in adults with plasma ALT and AST at the same four time points. Associations with elastography-derived liver steatosis and fibrosis were tested in early adulthood.

Results: Genetic risk factors for FLD (individually, combined into a 3-variant score, a 17-variant score and as a genome-wide PRS), were associated with higher liver enzymes, beginning in childhood and throughout adolescence and early adulthood. The ALT-increasing effects of the genetic risk variants became larger with increasing age. The ALT-PRS was associated with liver steatosis in early adulthood. No genetic associations with fibrosis were observed.

Conclusions: Genetic factors that promote FLD in adults associate with elevated liver enzymes already during childhood, and their effects get amplified with increasing age.

Key words: Genetic risk scores, polygenic prediction, liver enzymes, NAFLD, ALSPAC

Key Messages

- Genetic risk factors for fatty liver disease begin to exert their effects already during early childhood.
- The effects get larger with increasing age throughout childhood and adolescence.
- Fatty liver disease in adulthood could partly be due to an 'unmasking' of genetic risk by factors that emerge during childhood and adolescence.

Introduction

Fatty liver disease (FLD) encompasses a spectrum of hepatic pathologies stemming from accumulation of fat inside the liver cells (steatosis). Hepatic steatosis can lead to inflammation of the liver tissue (steatohepatitis) that in turn can progress to fibrosis and cirrhosis of the liver: serious, irreversible diseases that are associated with high morbidity and mortality. Previously an uncommon disease mainly caused by habitual excessive alcohol consumption, FLD due to non-alcoholic causes has become very common in recent decades due to the rising rates of obesity. Approximately 25% to 45% of adults have steatosis, the first stage of the FLD spectrum, and between 3% and 5% have non-alcoholic steatohepatitis.[1–3]

Mirroring the obesity pandemic in adults, the prevalence of paediatric obesity is rapidly growing throughout the world.[4,5] In parallel with this, FLD has become the most common liver disease affecting children. A worrying aspect of paediatric FLD is that the disease is likely to persist into adulthood, conferring a substantial cumulative risk of progressing to chronic liver disease at a young age. Recent data from a study of young British adults (mean age 24 years) found that among those with obesity (body mass index >30 kg/m²), half had severe hepatic steatosis, and among those with severe steatosis, 5% had hepatic fibrosis.[6]

Approximately half of the interindividual variation in propensity to develop FLD in adults is due to genetic factors. So far, genome-wide association studies (GWAS) have identified eight different genetic variants that associate robustly with liver fat content[7–11], 15 associated with chronic liver disease[12–15] and 172 that associate with elevated plasma levels of alanine transaminase (ALT), a biochemical marker of liver cell damage.[16–18] The most widely replicated of these variants are in the genes *PNPLA3*, *TM6SF2* and *HSD17B13*. Homozygous carriers of the risk-increasing variants at each of these three loci have an approximately 2–3-fold higher risk of FLD and/or chronic liver disease as compared with non-carriers.[19]

It remains unclear if the above-mentioned genetic risk variants exert their effects from birth or from some later time point during childhood or adolescence. This is important, because the timing of these effects may shed light on

the underlying biological pathways and provide clues to potentially permissive non-genetic factors (e.g. obesity, sex hormones or alcohol intake). Recent studies of the genetics of obesity during childhood have revealed a striking heterogeneity in onset and duration of the effects of the individual obesogenic variants. Some were associated with body mass index (BMI) at birth, others during infancy and early childhood (but not in adulthood) and some first began to exert their effects at 8 years of age.[20]

The aim of this study was to test if genetic susceptibility factors for FLD display similar age-varying association patterns. The study cohort included up to 4018 British children and adolescents from the Avon Longitudinal Study of Parents and Children (ALSPAC) study. We examined the associations of the three variants known to associate robustly with FLD in adults (in *PNPLA3*, *TM6SF2* and *HSD17B13*) with liver enzymes during childhood (mean age: 9.9 years), early adolescence (15.5 years), late adolescence (17.8 years) and early adulthood (24.5 years). Furthermore, we tested the associations of a 17-variant score and polygenic risk scores (derived from associations in adults) with liver enzymes at the same four time points.

Methods

The Avon Longitudinal Study of Parents and Children

ALSPAC is a population-based cohort investigating genetic and environmental factors that affect the health and development of children. The study methods are described in detail elsewhere.[21,22] In brief, 14 541 pregnant women resident in the former region of Avon, UK, with an expected delivery date between 1 April 1991 and 31 December 1992, were eligible to take part in ALSPAC. Of these initial pregnancies there was a total of 14 676 foetuses, resulting in 14 062 live births and 13 988 children who were alive at 1 year of age. Please note that the study website contains details of all the data that are available through a fully searchable data dictionary and variable search tool [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. All research was conducted in accordance with both the Declarations of Helsinki and of Istanbul.

In ALSPAC, there were data available on liver biomarkers at four time points over the life course, measured during clinic visits: 'childhood'—mean age = 9.9 years clinic (range = 8.9 to 11.5 years); 'mid-adolescence'—mean age = 15.5 years clinic (range = 14.3 to 17.7 years); 'late adolescence'—mean age = 17.8 years clinic (range = 16.3 to 20 years); and 'early adulthood'— mean age = 24.5 - years clinic (range = 22.4 to 26.5 years). Measures for ALT and aspartate transaminase (AST) were obtained at each time point. Biomarker levels from the childhood time point at mean age 9.9 years were derived using non-fasting blood samples, whereas all other time points were based on samples after participants had not eaten or drunk anything apart from water for at least 6 h. Lipoprotein lipid levels [high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol and triglycerides] were also available as part of these biomarker panels. Measures of body mass index (BMI) and dual-energy X-ray absorptiometry (DXA)-assessed fat mass were also derived at each time point. BMI was calculated using height measured to the nearest 0.1 cm using a Harpenden stadiometer (Holtain Crosswell), and weight was measured to the nearest 0.1 kg on Tanita electronic scales. Whole-body DXA scans were conducted using a Lunar Prodigy fan-beam densitometer using paediatric software as described previously.[23] Additionally, a subset of the participants examined at mean age = 24.5 years had hepatic fibrosis and steatosis assessed by transient elastography (FibroScan 502 Touch; Echosens, Paris, France).

Genome-wide association studies in the UK Biobank

Genome-wide association studies (GWAS) were undertaken in the UK Biobank (UKB) on ALT (field #30620) and AST (field #30650) to identify weights for genetic and whole-genome polygenic risk scores (PRS) (Supplementary Table S1, available as Supplementary data at *IJE* online). The analysis protocol for these GWAS have been described previously in detail.[24] Briefly, UKB participants of non-European descent based on K-means clustering (K = 4) were excluded from analyses along with individuals with withdrawn consent, mismatch between genetic and reported sex, and putative sex chromosome aneuploidy. GWAS were conducted using the BOLT-LMM software that applied a linear mixed model to account for population structure and cryptic relatedness in UKB.[25] Analyses were undertaken after applying rank inverse normal transformations to liver traits with adjustment for age and sex,

resulting in final sample sizes of $n = 441\,209$ (ALT) and $n = 439\,757$ (AST).

Genotypes and genetic risk scores

Three genetic variants were selected based on previous studies that had elucidated their role in FLD.[8,10,12] These single nucleotide polymorphisms (SNPs) were rs738409 (*PNPLA3*), rs58542926 (*TM6SF2*) and rs13130041 (*HSD17B13*), which were selected because they were previously used to construct a genetic risk score for FLD in an adult population.[19] rs13130041 was identified as a proxy for the original SNP rs72613567 as this was not available in the ALSPAC genotype data ($r^2 = 0.95$). Whereas the variants in *PNPLA3* and *TM6SF2* influence the entire FLD spectrum, the *HSD17B13* variant only affects the later stages of the disorders [non-alcoholic steatohepatitis (NASH), fibrosis, cirrhosis] but not steatosis per se.[12] We also selected 17 variants robustly associated with elevated ALT and with imaging or histologically defined FLD[16] (Supplementary Table S2, available as Supplementary data at *IJE* online).

We conducted four analyses. First, we analysed the association between the three variants in *PNPLA3*, *TM6SF2* and *HSD17B13* individually against ALSPAC measures of ALT and AST, stratified by age. Second, to increase power we combined the three variants (each weighted by its effects on ALT and AST in UKB) into weighted genetic risk scores. We tested the association between increasing genetic risk score and ALT and AST in ALSPAC, again stratified by age. Third, we tested the association with ALT of a 17-variant weighted genetic risk score based on the 17 variants robustly associated with FLD mentioned above. The weights of the individual variants in this risk score were derived from beta ALT as reported in the GWAS. Our fourth set of analyses was based on whole-genome PRS for ALT and AST, based on the full set of genome-wide results from our UKB GWAS. PRS were constructed by applying linkage disequilibrium (LD) clumping to our UKB GWAS results to identify genetic variants based on an $r^2 < 0.1$, window distance of 1000 kb and $P < 0.05$. A reference panel of 10 000 random UKB European participants was used to estimate correlation of variants for LD clumping.[26] PRS were then built using ALSPAC genotype data for ALT and AST as separate scores, by summing trait-increasing alleles weighted by their GWAS effect estimates. PLINK v2.0 was used for all LD clumping and PRS construction.[27]

Statistical analysis

Rank inverse normal transformations were applied to all liver biomarker data and measures of fibrosis in ALSPAC

prior to analyses. The association between each of the three genetic variants was then evaluated with ALT and AST in turn in ALSPAC, using linear regression with adjustment for age and sex. The genotypes were coded 0, 1 and 2 for number of ALT-increasing alleles in the single variant analyses. Previous analyses have demonstrated that adjusting for principal components in ALSPAC does not make meaningful differences to results, given that study participants were recruited from the same geographical area (county of Avon in the UK).[28] We therefore did not adjust for principal components in the present study. The three-variant and 17-variant scores and whole-genome PRS were entered as continuous variables in linear regression.

Evaluations of an interaction with time points in ALSPAC were conducted by stacking liver biomarker observations (i.e. so that each individual had multiple measures) and then regressing this against genetic variables with adjustment for age, sex and time point. Analyses were then repeated, along with adjustment for the interaction between genetic variables and time point as undertaken previously,[29] by stacking observations (i.e. so that individuals had multiple ALT/AST measurements) and then regressing the liver enzymes measures on genetic variables with adjustment for age, sex and time point. An inherent feature of quantitative interactions (i.e. interactions that affect the magnitude but not the direction of an association) is that they depend on the scale on which they are analysed. To assess the effects of scale transformations, we repeated all interaction tests using untransformed or logarithmically transformed values of ALT or AST.

Associations of the ALT-PRS with BMI and plasma lipid levels were tested at the four time points, using linear regression adjusted for age and sex. Furthermore, given previous evidence of an interaction between genetic variants which influence risk of FLD and body BMI,[30] we subsequently included an interaction term between liver biomarkers and BMI in models to formally assess this. Analyses were repeated to evaluate interactions with DXA-assessed fat mass substituted for BMI in the model. Last, we evaluated the association between individual genetic variants, three-variant and 17-variant scores and PRS, and the Fibroscan-measures of fibrosis and steatosis taken at the early adulthood time point in ALSPAC, as before using linear regression adjusting for age and sex. Finally, we used the whole-genome PRS to stratify the ALSPAC population into deciles, and for their respective biomarkers assessed linear trends across groups using linear regression.

Figures in this paper were generated using the R package 'ggplot2'.[31] All analyses were conducted using R (version 3.5.3).

Results

Baseline characteristics

The baseline characteristics of up to 4018 participants from the ALSPAC cohort, stratified by age group, are given in Supplementary Table S3 (available as Supplementary data at *IJE* online).

PNPLA3, TM6SF2 and HSD17B13 variants and liver enzymes during childhood through to early adulthood

We first tested the associations of three variants known to associate robustly with FLD in adults. The strongest evidence of association was found between the *PNPLA3* variant rs738409 and ALT measured during early adulthood (i.e. mean age = 24.5 years) (beta = 0.12 per SD change in ALT levels, 95% CI = 0.05 to 0.18, $P = 5.8 \times 10^{-4}$). Furthermore, we found that the magnitude of effect for this variant on both ALT and AST levels increased incrementally across time points in the ALSPAC cohort (Figure 1A). For example, the per-allele effect on inverse-normalized ALT was 0.045 at 9.9 years old, 0.059 at 15.5 years old, 0.094 at 17.8 years old and 0.115 by 24.5 years old (*P* for interaction between time and *PNPLA3* genotype on ALT = 0.048). A similar, albeit less consistent, pattern of increasing effects on inverse-normalized AST with age was seen for the *TM6F2*-variant (*P* for interaction between time and *TM6SF2* genotype on inverse normalized AST = 0.075). The interactions were similar when analysing ALT or AST either as raw, untransformed values or after logarithmical transformation (Supplementary Table S4, available as Supplementary data at *IJE* online). The full results underlying Figure 1A can be found in Supplementary Table S5 (available as Supplementary data at *IJE* online).

Three-variant risk score and liver enzymes during childhood through to early adulthood

Evidence of an association between the three-variant genetic risk score and liver enzymes was strongest when analysing data from the late adolescent (i.e. mean age = 17.8 years) and early adulthood (i.e. mean age = 24.5 years) time points (Figure 1B). For example, the AST risk score was most strongly associated with AST levels during late adolescence (beta = 1.18 per SD change in AST levels, 95% CI = 0.60 to 1.76, $P = 6.7 \times 10^{-5}$). Evidence of an interaction with time for the three variant risk scores for inverse normalized ALT and AST was $P = 0.062$ and $P = 0.021$, respectively (Supplementary Table S4). The evidence for interaction was stronger when
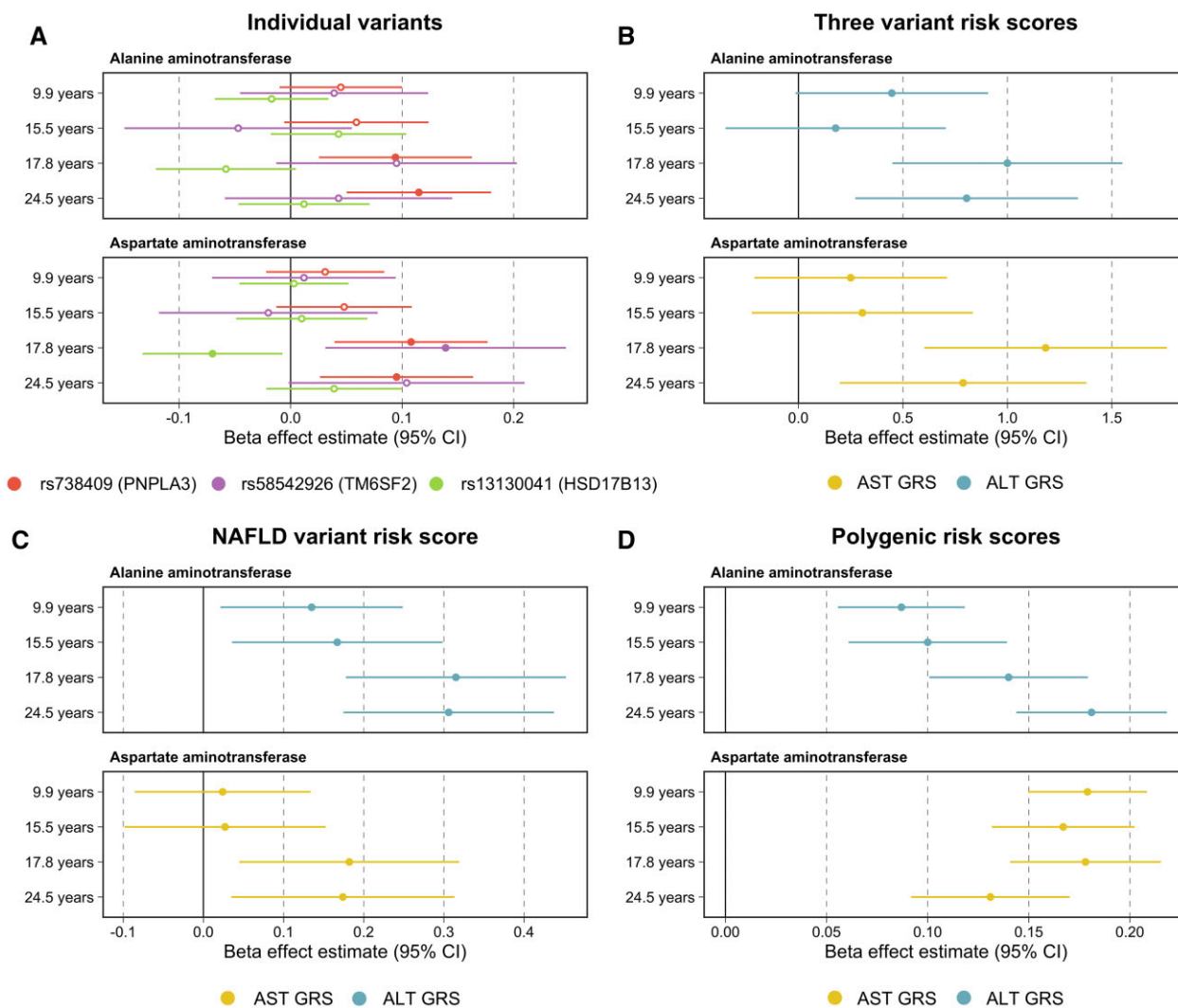
Figure 1 Forest plots illustrating the effect estimates per one standard deviation change in liver enzymes, and corresponding 95% confidence intervals for A) single variant analyses encoded 0, 1 and 2 per ALT-increasing allele. Filled circles: $P<0.05$, unfilled circles: $P\geq0.05$. B) Genetic risk scores consisting of all three variants encoded additively per ALT-increasing allele. C) Genetic risk score consisting of 17 variants associated with ALT as well as histologically or imaging-verified fatty liver disease in adults. D) Polygenic risk scores weighted by effect estimates derived from UK Biobank genome-wide association studies for corresponding traits after rank inverse normal transformation. Analyses were conducted at four time points across childhood, adolescence and early adulthood using liver enzyme data from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. ALT, alanine transaminase; GRS, genetic risk score; NAFLD, non-alcoholic fatty liver disease

analysing ALT or AST on the untransformed scale or after logarithmical transformation (all *P*-values <0.02; Supplementary Table S4). The full results underlying Figure 1B can be found in Supplementary Table S6 (available as Supplementary data at *IJE* online).

The 17-variant risk score and plasma ALT during childhood through to early adulthood

Similarly, we found that the 17-variant risk score provided stronger evidence of association with plasma ALT at the late adolescent (beta = 0.32 per SD change in ALT levels, 95% CI = 0.18 to 0.45, $P=6.1\times10^{-6}$) and early adulthood (beta = 0.31, 95% CI = 0.17 to 0.44, $P=6.0\times10^{-6}$)

time points. Comparatively, the association of this score with AST levels was weaker than with ALT, which was to be expected given that these 17 variants were identified based on elevated ALT levels (Figure 1C). Full results underlying this figure can be found in Supplementary Table S7 (available as Supplementary data at *IJE* online).

Genome-wide PRS and liver enzymes during childhood and early adulthood

There was strong evidence of an association for both the ALT and AST whole-genome PRS and their corresponding traits measured at each of the four time points in the ALSPAC cohort (Supplementary Table S8, available as

Supplementary data at *IJE* online). The ALT-PRS and AST-PRS explained 0.2% to 1.5% of and 0.2% to 2.1% of the total variation in plasma ALT and AST, respectively, across the four time points (Supplementary Table S9, available as Supplementary data at *IJE* online). The association between the genome-wide ALT-PRS and ALT plasma levels increased with age and was strongest among the young adults (Figure 1C, P for interaction between time and PRS on inverse normalized $ALT = 1.5 \times 10^{-4}$). The evidence for interaction was similar when analysing ALT on the untransformed scale (*P* for interaction $= 2.6 \times 10^{-4}$) but was stronger when analysing ALT on the logarithmical scale (*P* for interaction $= 2.1 \times 10^{-7}$, Supplementary Table S4). In contrast, the association between the AST-PRS and AST was similar during childhood and adolescence, and slightly weaker during young adulthood (*P* for interaction between time and AST-PRS on inverse normalized $AST = 0.136$).

Using these genome-wide PRS to stratify the ALSPAC population resulted in a clear incremental trend observed across risk deciles for each ,as illustrated using data from the childhood (i.e. mean age = 9.9 years) in Figure 2. Plots generated using findings from the other three time points can be found in Supplementary Figures S1–S3 (available as Supplementary data at *IJE* online). Tests for linearity in effects on ALT and AST across deciles are shown in Supplementary Table S10 (available as Supplementary data at *IJE* online). The percentage of individuals with

liver enzyme levels greater than the 95th percentile across each PRS decile can be found in Supplementary Table S11 (available as Supplementary data at *IJE* online).

Associations between genetic scores with BMI and plasma lipid levels during childhood and early adulthood

We found that the association between the genome-wide PRS for ALT and BMI increased in magnitude across time points in ALSPAC from mean age 9.9 years (beta = 0.28, 95% CI = 0.06 to 0.49, P = 0.01) to mean age 24.5 years (beta = 0.75, 95% CI = 0.44 to 1.06, $P = 1.7 \times 10^{-6}$). The ALT PRS was likewise associated with triglycerides most strongly at the early adulthood time point (beta = 0.57, 95% CI = 0.23 to 0.91, $P = 9.6 \times 10^{-4}$). In terms of individual genetic variants, the *TM6SF2* variant (rs58542926) was associated consistently with LDL cholesterol across time points in ALSPAC (Supplementary Table S12, available as Supplementary data at *IJE* online).

Interaction between genetics and adiposity on liver enzymes during childhood through to early adulthood

Previous studies in adults have found that the effects of individual risk variants and polygenic risk scores for liver disease are robustly amplified by underlying obesity or
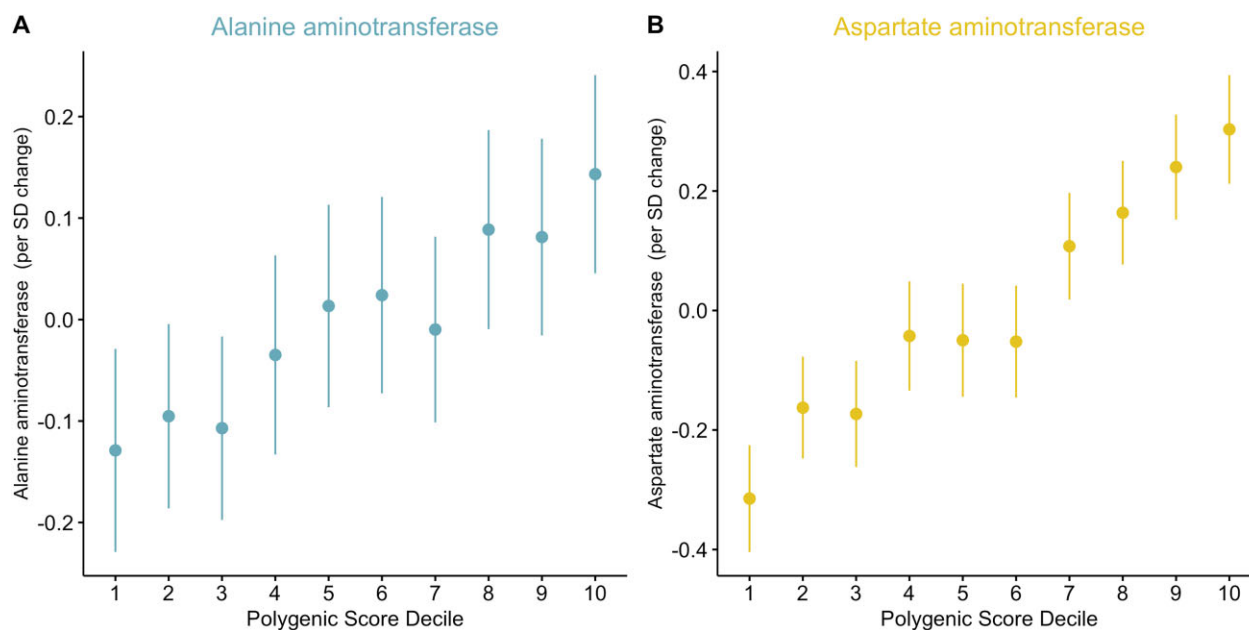

Figure 2 Error plots presenting the mean measurement and 95% confidence intervals for A) alanine aminotransferase and B) aspartate aminotransferase measured in a cohort of young individuals (mean age = 9.9 years, range = 8.8 to 11.7 years) from the Avon Longitudinal Study of Parents and Children (ALSPAC) study, across deciles determined using whole-genome polygenic risk scores (PRS) for their corresponding traits. Weights for polygenic risk scores were derived using findings from genome-wide association studies undertaken in populations of adults (mean age 56.9 years, range = 39 to 73 years) enrolled in the UK Biobank study. Similar plots generated using findings from the other three timepoints can be found in Supplementary Figures S1–S3 (available as Supplementary data at *IJE* online).

insulin resistance (i.e. interaction). We attempted to identify similar interaction in the ALSPAC cohort. Fitting interaction terms between genetic variants (and scores) with measured BMI and DXA-assessed fat mass at each time point in the ALSPAC cohort provided little evidence that BMI may modulate the association of these genetic variables with the early life measures of liver biomarkers (Supplementary Table S13, available as Supplementary data at *IJE* online).

Genetic effects on liver fat and fibrosis at 24.5 years of age

Fibroscan-quantified hepatic steatosis and fibrosis were available in approximately $n = 2400$ from the young adulthood time point in ALSPAC (mean age: 24.5, range: 22.4 to 26.5). Evaluating the relationship between individual genetic variants and polygenic scores with measures of steatosis or fibrosis provided the strongest evidence of association when analysing the ALT PRS with the measure of steatosis based on the controlled attenuation parameter (CAP) (beta = 0.08 per SD change in CAP measure, 95% CI = 0.04 to 0.12, $P = 2.5 \times 10^{-5}$). However, no other findings from this analysis were robust to multiple testing comparisons (Supplementary Table S14, available as Supplementary data at *IJE* online).

Discussion

The main finding of this study is that genetic variants known to confer a higher risk of FLD in adults also associate with elevated liver enzymes during early childhood, adolescence and early adulthood. The liver enzyme-increasing effects of these variants were evident already among the 10-year-old children in our cohort and they seemed to become stronger with increasing age.

Unravelling the genetic underpinning of FLD has progressed rapidly during the past decade, mainly owing to the burgeoning of large-scale GWAS conducted in cohorts of adults. In the latest and largest of these, 172 different genetic loci were found to associate with elevated plasma ALT levels.[17] Comparatively little is known about the genetic basis of FLD in children. The most robustly validated risk variants identified in adults have been found to confer risk of FLD in children too. For example, Goyal and colleagues recently reported that the *PNPLA3* and *TM6SF2* variants were associated with hepatic steatosis among 822 children with histologically verified non-alcoholic FLD.[32] Earlier studies have reported similar results.[33,34] The results reported in our study align with and extend the findings from these previous studies. Taken together, these data support the notion that genetic risk factors that promote the development of FLD exert their effects from childhood and throughout adolescence and early life.

An unexpected finding in our study was that the liver enzyme-increasing effects of the genetic variants increased with higher age. This was true for the individual variants (in *PNPLA3* and *TM6SF2*), for a three-variant score comprising the risk alleles in *PNPLA3*, *TM6SF2* and *HSD17B13*, a 17-variant score based on alleles with robust effects on FLD in adults, and most clearly for a genome-wide ALT-PRS. The mechanisms underlying these time-gene interactions are not obvious. We speculate that they might reflect the effects of certain permissive factors that gradually emerge during childhood and adolescence, effectively 'unmasking' the genetic risk factors.

In support of this notion is the fact that previous studies have consistently found that the deleterious effect of genetic risk factors for FLD in adults are amplified (i.e. 'un-masked') by the co-occurrence of environmental risk factors such as obesity, heavy alcohol intake or insulin resistance.[19,30] These synergistic effects are among the most robust examples of gene-environment interaction documented in humans.[35] Some moderately-sized studies have found that adiposity augments the steatogenic effect of the *PNPLA3* variant in children.[36,37] It is, however, unclear if these interactions are generalizable to all genetic risk variants during early life. This is an important question. If obesity 'unmasks' the genetic risk of FLD during early life, the rapidly rising prevalence of obesity in children could herald a future of widespread paediatric and adolescent FLD, especially among ethnic groups genetically predisposed to develop the disorder (e.g. Hispanics).[10,38]

There are limitations to our study that should be considered. First, with $N = 4018$ in the full cohort, power was limited for stratified sub-analyses. For example, although we were able to detect gene-time interactions affecting liver enzyme levels, we did not detect any evidence of BMI-gene or DXA-assessed fat mass-gene interactions. This may reflect lack of power, given that such interactions are well described in adults, and have been hinted at in previous studies of children.[36,37] Additionally, we note that the estimates for our genetic scores derived from the UK Biobank are likely to suffer from 'winner's curse' and that this source of bias may variably affect findings across time points in ALSPAC. Full comparisons of estimates derived in the UK Biobank with those derived in ALSPAC can be found in Supplementary Tables S15–S22 (available as Supplementary data at *IJE* online). Another limitation is that plasma levels of liver enzymes are neither specific nor sensitive markers of FLD. Whereas the variants in *PNPLA3*, *TM6SF2* and *HSD17B13* and those included in the 17-variant score are known to be robustly associated with ALT and FLD, we cannot rule out that some of the

variants used in the whole-genome PRS (derived from associations with ALT or AST) may affect liver enzyme levels via other pathways than FLD. The exact fraction of 'true FLD-variants' in the PRS is unknown. That said, recent GWASs in adults have reported high correlations between genetic effects on plasma ALT and their effects on FLD.[16] Finally, ALSPAC primarily (>95%) comprises children and adolescents of White European ethnicity recruited from the vicinity of Bristol in the UK. This potentially limits the generalizability of our findings to other ethnic groups and geographies.

Conclusion

In conclusion, the genetic risk factors that promote FLD in adults begin to exert their deleterious effects already during childhood, and these effects appear to grow stronger with increasing age. These data highlight the potential of FLD to become a major health problem among children, adolescents and young adults, potentially due to an 'unmasking' of heritable risk by permissive factors such as obesity.

Ethics approval

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. All research was conducted in accordance with the Declarations of Helsinki and of Istanbul.

Data availability

All individual-level data analysed in this study can be accessed via an approved application to ALSPAC [http://www.bristol.ac.uk/alspac/researchers/access/]and the UK Biobank study [https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access].

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

S.S.: study concept and design; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; obtained funding. G.D.S.: study concept and design; acquisition of data; analysis and interpretation of data; critical revision of the manuscript for important intellectual content; obtained funding; administrative, technical or material support. T.G.R.: study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; obtained funding. All authors approved the final version of the draft to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

Conflict of interest

T.G.R. is an employee of GlaxoSmithKline outside this research. All other authors declare no conflict of interest.

References

1. Povsic M, Wong OY, Perry R, Bottomley J. A structured literature review of the epidemiology and disease burden of Non-Alcoholic Steatohepatitis (NASH). *Adv Ther* 2019;**36**:1574–94.
2. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;**64**:73–84.
3. Browning JD, Szczepaniak LS, Dobbins R *et al.* Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* 2004;**40**:1387–95.
4. Younossi ZM, Henry L. Fatty liver through the ages: nonalcoholic steatohepatitis. *Endocr Pract* 2022;**28**:204–13.
5. Li J, Ha A, Rui F *et al.* Meta-analysis: global prevalence, trend and forecasting of non-alcoholic fatty liver disease in children and adolescents, 2000-2021. *Aliment Pharmacol Ther* 2022;**56**:396–406.
6. Anderson EL, Howe LD, Fraser A *et al.* Weight trajectories through infancy and childhood and risk of non-alcoholic fatty liver disease in adolescence: the ALSPAC study. *J Hepatol* 2014;**61**:626–32.

7. Haas ME, Pirruccello JP, Friedman SN *et al.* Machine learning enables new insights into genetic contributions to liver fat accumulation. *Cell Genom* 2021;**1**:100066.

8. Kozlitina J, Smagris E, Stender S *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2014;**46**:352–56.

9. Liu Y, Basty N, Whitcher B *et al.* Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. *Elife* 2021;**10**:e65554.

10. Romeo S, Kozlitina J, Xing C *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2008;**40**:1461–65.

11. Speliotes EK, Yerges-Armstrong LM, Wu J *et al.*; GOLD Consortium. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet* 2011;**7**: e1001324.

12. Abul-Husn NS, Cheng X, Li AH *et al.* A protein-truncating HSD17B13 variant and protection from chronic liver disease. *N Engl J Med* 2018;**378**:1096–106.

13. Buch S, Stickel F, Trepo E *et al.* A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet* 2015;**47**:1443–48.

14. Emdin CA, Haas M, Ajmera V *et al.* Association of genetic variation with cirrhosis: a multi-trait genome-wide association and gene-environment interaction study. *Gastroenterology* 2021; **160**:1620–33.e13.

15. Innes H, Buch S, Hutchinson S *et al.* Genome-wide association study for alcohol-related cirrhosis identifies risk loci in MARC1 and HNRNPUL1. *Gastroenterology* 2020;**159**:1276–89.e7.

16. Vujkovic M, Ramdas S, Lorenz KM *et al.*; VA Million Veteran Program. A multiancestry genome-wide association study of unexplained chronic ALT elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. *Nat Genet* 2022;**54**:761–71.

17. Chen VL, Du X, Chen Y *et al.* Genome-wide association study of serum liver enzymes implicates diverse metabolic and liver pathology. *Nat Commun* 2021;**12**:816.

18. Chambers JC, Zhang W, Sehmi J *et al.*; Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC). Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 2011;**43**: 1131–38.

19. Gellert-Kristensen H, Richardson TG, Davey Smith G, Nordestgaard BG, Tybjaerg-Hansen A, Stender S. Combined effect of PNPLA3, TM6SF2, and HSD17B13 variants on risk of cirrhosis and hepatocellular carcinoma in the general population. *Hepatology* 2020;**72**:845–56.

20. Helgeland O, Vaudel M, Sole-Navais P *et al.* Characterization of the genetic architecture of infant and early childhood body mass index. *Nat Metab* 2022;**4**:344–58.

21. 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**:111–27.

22. 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**:97–110.

23. Ness AR, Leary SD, Mattocks C *et al.* Objectively measured physical activity and fat mass in a large cohort of children. *PLoS Med* 2007;**4**:e97.

24. Richardson TG, Sanderson E, Palmer TM *et al.* Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. *PLoS Med* 2020;**17**: e1003062.

25. Loh PR, Tucker G, Bulik-Sullivan BK *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;**47**:284–90.

26. Kibinge NK, Relton CL, Gaunt TR, Richardson TG. Characterizing the causal pathway for genetic variants associated with neurological phenotypes using human brain-derived proteome data. *Am J Hum Genet* 2020;**106**:885–92.

27. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**:7.

28. Richardson TG, O'Nunain K, Relton CL, Davey Smith G. Harnessing whole genome polygenic risk scores to stratify individuals based on cardiometabolic risk factors and biomarkers at age 10 in the lifecourse-brief report. *Arterioscler Thromb Vasc Biol* 2022;**42**:362–65.

29. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: Mendelian randomisation study. *BMJ* 2020;**369**:m1203.

30. Stender S, Kozlitina J, Nordestgaard BG, Tybjærg-Hansen A, Hobbs HH, Cohen JC. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. *Nat Genet* 2017; **49**:842–47.

31. Ginestet C. ggplot2: elegant graphics for data analysis. *J R Stat Soc Stat* 2011;**174**:245–46.

32. Goyal NP, Rosenthal SB, Nasamran C *et al.*; NASH Clinical Research Network. Nonalcoholic fatty liver disease risk and histologic severity are associated with genetic polymorphisms in children. *Hepatology* 2023;**77**:197–212.

33. Riccio S, Melone R, Vitulano C *et al.* Advances in pediatric nonalcoholic fatty liver disease: from genetics to lipidomics. *World J Clin Pediatr* 2022;**11**:221–38.

34. Nobili V, Alisi A, Valenti L, Miele L, Feldstein AE, Alkhouri N. NAFLD in children: new genes, new diagnostic modalities and new drugs. *Nat Rev Gastroenterol Hepatol* 2019;**16**:517–30.

35. Lyon MS, Millard LAC, Davey Smith G, Gaunt TG, Tilling K. Hypothesis-free detection of gene-interaction effects on biomarker concentration in UK Biobank using variance prioritisation. *medRxiv*; doi:10.1101/2022.01.05.21268406, 5 January 2022 Preprint: not peer reviewed.

36. Giudice EM, Grandone A, Cirillo G *et al.* The association of PNPLA3 variants with liver enzymes in childhood obesity is driven by the interaction with abdominal fat. *PLoS One* 2011;**6**: e27933.

37. Romeo S, Sentinelli F, Cambuli VM *et al.* The 148M allele of the PNPLA3 gene is associated with indices of liver damage early in life. *J Hepatol* 2010;**53**:335–38.

38. Kubiliun MJ, Cohen JC, Hobbs HH, Kozlitina J. Contribution of a genetic risk score to ethnic differences in fatty liver disease. *Liver Int* 2022;**42**:2227–36.