Early metabolic features of genetic liability to type 2 diabetes: cohort study with repeated metabolomics across early life

Short title: Early features of type 2 diabetes liability

Joshua A. Bell, PhD [1,2 *], Caroline J. Bull, PhD [1,2,3 *], Marc J. Gunter, PhD [4], David Carslake, PhD [1,2], Anubha Mahajan, PhD [5], George Davey Smith, MD, DSc, FRS [1,2], Nicholas J. Timpson, PhD [1,2], Emma E. Vincent, PhD [1,2,3]

* Contributed equally

[1] MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

[2] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[3] School of Cellular and Molecular Medicine, University of Bristol, UK

[4] Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France

[5] Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. Current address: Genentech, 1 DNA Way, South San Francisco, CA 94080, USA

Corresponding author: Joshua A. Bell, Oakfield House, Bristol, UK, BS8 2BN, +44 (0)11 7331 0085, j.bell@bristol.ac.uk, Twitter: @joshuaa_bell

Main text words: 4000 / 4000; **References:** 40 / 40; **Tables & Figures:** 5; **Supplementary files:** 14

Keywords: Type 2 diabetes; Metabolism, Genetics; NMR; Epidemiology; Mendelian randomization; ALSPAC

Abstract

Objective: Type 2 diabetes develops for many years before diagnosis. We aimed to reveal early metabolic features characterising liability to adult disease by examining genetic liability to adult type 2 diabetes in relation to metabolomic traits across early life.

Research Design and Methods: Up to 4,761 offspring from the Avon Longitudinal Study of Parents and Children were studied. Linear models were used to examine effects of a genetic risk score (162 variants) for adult type 2 diabetes on 229 metabolomic traits (lipoprotein-subclass-specific cholesterol and triglycerides, amino acids, glycoprotein acetyls, others) measured at age 8y, 16y, 18y, and 25y. Two-sample Mendelian randomization (MR) was also conducted using genome-wide association study data on metabolomic traits in an independent sample of 24,925 adults.

Results: At age 8y, associations were most evident for type 2 diabetes liability (per SD-higher) with lower lipids in high-density lipoprotein (HDL) subtypes, e.g. -0.03 SD (95% CI=-0.06, -0.003) for total lipids in very-large HDL. At 16y, associations were stronger with pre-glycemic traits including citrate and with glycoprotein acetyls (0.05 SD, 95% CI=0.01, 0.08), and at 18y, associations were stronger with branched chain amino acids. At 25y, associations had strengthened with VLDL lipids and remained consistent with previously altered traits including HDL lipids. Two-sample MR estimates among adults indicated persistent patterns of effect of disease liability.

Conclusions: Our results support perturbed HDL lipid metabolism as one of the earliest features of type 2 diabetes liability, alongside higher branched chain amino acid and inflammatory levels. Several features are apparent in childhood as early as age 8y, decades before the clinical onset of disease.

Background

Type 2 diabetes is a metabolic disease affecting more than 400 million people globally (1). Its incidence is driven largely by increased adiposity (2), a strong causal risk factor (3, 4), yet the difficulty of achieving and maintaining weight loss makes disease management a lifelong and expensive task (5). This is particularly problematic considering that potentially half of those living with type 2 diabetes are undiagnosed and the future burden is expected to be greatest in lower-income countries (1). There is therefore a clear need to minimise the impact of type 2 diabetes and this requires biological understanding of the disease at its very earliest stages.

Type 2 diabetes is typically diagnosed when blood glucose levels exceed 7 mmol/l in the fasting state or 11.1 mmol/l in the post-challenge state, or when glycated haemoglobin levels exceed 6.5% (6), yet glucose spikes relatively late in the disease process. Repeat clinical measures from the Whitehall II cohort study suggest that insulin sensitivity starts declining a decade before glucose changes are detectable (7). Cohort studies with metabolomic measurements also observe associations of numerous subclinical traits with lower insulin sensitivity including higher branched chain amino acid (BCAA) concentrations; higher fatty acid and inflammatory glycoprotein concentrations; and elevated lactate and pyruvate (8-11). Relations with ketone bodies are less clear, with higher levels associated with both higher insulin sensitivity (9) and higher type 2 diabetes risk (12). Hyperglycemia also associates strongly with lipoprotein cholesterol and triglycerides (13). Whether such trait alterations reflect developmental stages of type 2 diabetes is unclear because of inherent confounding by other disease processes.

Another approach to causal inference is to examine genetic liability to type 2 diabetes in relation to metabolic traits, to identify perturbations specific to the development of type 2 diabetes itself – i.e. its early metabolic features (14). The few studies that have investigated this suggest effects of type 2 diabetes liability on cholesterol and triglycerides in high-density lipoprotein (HDL) and very-low density lipoprotein (VLDL) particles (15), BCAAs (16, 17), and ketone bodies (12). Most used a small set of genetic variants from early genome-wide association studies (GWAS) and all relied on one-off measures of metabolic traits among middle-to-older-aged adults which gives little insight as to when in life metabolic alterations first occur. Examining genetic liability to type 2 diabetes in relation to repeated measures of metabolic traits starting in childhood could reveal the existence and timing of subclinical trait perturbations most central to type 2 diabetes development.

We aimed in this study to reveal early metabolic features characterising liability to type 2 diabetes. Using birth cohort study data, we examined genetic liability to adult type 2 diabetes in relation to detailed traits from targeted metabolomics among the same individuals at four key stages of early life – childhood (age 8y), adolescence (16y), young-adulthood (18y), and adulthood (25y). For replication, two-sample Mendelian randomization (MR) (18) was conducted in an independent sample of adults to confirm the persistence of any metabolic features of disease liability observed in early life.

Methods

Study population

Data were from offspring participants of the Avon Longitudinal Study of Parents and Children (ALSPAC), a population-based birth cohort study in which 14,541 pregnant women with an expected delivery date between 1 April 1991 and 31 December 1992 were recruited from the former Avon county of southwest England (19). Since then, 13,988 offspring alive at one year have been followed repeatedly with questionnaire- and clinic-based assessments (20), with an additional 811 children enrolled over the course of the study. Offspring were considered for the current analyses if they had no older siblings in ALSPAC (202 excluded) and were of white-European ethnicity (604 excluded) to reduce confounding of associations by high relatedness and ancestral population structure.

Written informed consent was provided and ethical approval was obtained from the ALSPAC Law and Ethics Committee and the local research ethics committee. The study website contains details of all available data through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/).

Assessment of genetic liability to adult type 2 diabetes

Genotype was assessed using the Illumina HumanHap550 quad chip platform. Quality control measures included exclusion of participants with sex mismatch, minimal or excessive heterozygosity, disproportionately missing data, insufficient sample replication, cryptic relatedness, and non-European ancestry. Imputation was performed using the Haplotype Reference Consortium panel. Since this study aims to address causation, not prediction,

genetic liability to type 2 diabetes was based on genetic variants associated with type 2 diabetes case status at genome-wide significance in the largest GWAS to date which identified up to 403 independent polymorphisms among adults (74,124 cases and 824,006 controls) of white-European ethnicity, explaining 17.4% of variance (21). This set of variants was refined by excluding: 105 variants with P-values $\geq 5.00 \times 10^{-8}$, 12 additional variants identified only when adjusting for body mass index (BMI), and variants that were in linkage disequilibrium based on $R^2 > 0.001$ (retaining those SNPs with the lowest P-values) using the TwoSampleMR package (22). This left 167 variants highly independently associated with adult type 2 diabetes (explaining 7.1% of variance in the UK Biobank study, using phenotyping as prior (21)) (**Supplementary Table-1**), 162 of which were available in imputed ALSPAC genotype data post-quality-control. This set was combined into a genetic risk score (GRS) using PLINK 1.9, specifying the effect (type 2 diabetes raising) allele and coefficient (odds ratio) from the GWAS as external weights. Scoring was done by multiplying the number of effect alleles (or probabilities of effect alleles if imputed) at each SNP (0, 1, or 2) by its weighting, summing these, and dividing by the total number of SNPs used. The score therefore reflects the average per-SNP effect on type 2 diabetes (**Figure-1**).

Assessment of metabolic traits

Participants provided non-fasting blood samples during a clinic visit while aged approximately 8y, and fasting blood samples from clinic visits while aged approximately 16y, 18y, and 25y. Proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy from a targeted metabolomics platform (23) was performed using EDTA-plasma (stored at or below -70 degrees Celsius pre-processing) to quantify 229 traits (149 concentrations plus 80 ratios). These included the concentration and size of lipoprotein subclass particles and their

cholesterol and triglyceride content, apolipoproteins, fatty acids, pre-glycemic factors including lactate and glucose, amino acids, ketone bodies, and inflammatory glycoprotein acetyls.

Assessment of adiposity and type 2 diabetes status

For descriptive purposes, BMI was calculated at each time point as weight (kg) divided by squared-height (m^2) based on clinic measures of weight to the nearest 0.1 kg using a Tanita scale and height measured in light clothing without shoes to the nearest 0.1 cm using a Harpenden stadiometer. Type 2 diabetes/prediabetes status was not assessed at 8y because blood glucose was not quantified based on fasting samples and no data were collected regarding physician diagnosis. Type 2 diabetes was defined at 16y as fasting glucose ≥ 7 mmol/l, and at 18y and 25y as fasting glucose ≥ 7 mmol/l or reported physician diagnosis of type 2 diabetes by that age (6). Prediabetes was defined at 16y, 18y, and 25y as fasting glucose between 5.6 mmol/l and 6.9 mmol/l (6).

Statistical approach

In the first set of analyses, separate linear regression models with robust standard errors were used to estimate coefficients and 95% confidence intervals for associations of the type 2 diabetes GRS (in standard deviation (SD) units) with each metabolic trait measured at 8y (also in SD units) as dependent variables, adjusted for sex and age at the time of metabolic trait assessment. These were repeated for metabolic traits measured at 16y, 18y, and 25y. Estimates are interpreted within a 'reverse MR' framework (14), wherein the direction of causation is from disease liability to metabolic traits, and are taken to reflect 'metabolic features' of type 2 diabetes liability. To compare evidence of linear change in coefficients

across time points, we ran separate linear mixed models utilising repeated measures of each metabolic trait and examined P-values from an interaction term between the type 2 diabetes GRS and mean age at each time point (as an occasion/time variable). All models were additionally run using original (non-SD; mostly mmol/l) units to aid clinical interpretation and external comparisons.

To allow full use of measured data, the aforementioned analyses were first conducted using maximum numbers of participants (N varying across ages and between traits). Participants were eligible for analyses at any age if they had data on genotype, sex, age, and at least 1 metabolic trait. This resulted in 6,216 eligible participants (i.e. who contributed data to any analysis), including up to 4,761, 2,928, 2,612 and 2,560 participants in models at 8y, 16y, 18y and 25y, respectively. Analyses were repeated using a consistent (complete-case) sample of participants with data on genotype, sex, age, and all metabolic traits at each time point (**Figure-2**).

Additional analyses

Effects of type 2 diabetes liability on measured BMI at each time point were also examined. To examine mediation of effects of type 2 diabetes liability on metabolic traits by BMI and insulin (not measured by NMR), we conducted additional analyses wherein we adjusted associations between the type 2 diabetes GRS and metabolic traits for measured BMI at the time of each trait assessment, and separately for measured fasting insulin (measured using conventional clinical assays) at the time of each metabolic trait assessment (except for 8y when insulin was not measured). We repeated these adjustments using a GRS for BMI based on 312 SNPs independently associated with adult BMI (at $R^2<0.001$ and $P<5.00x10^{-8}$) (24) (**Supplementary Table-1**; 308 of which were available in imputed

ALSPAC post-quality-control), and separately for a GRS for fasting insulin based on 14 SNPs independently associated with adult fasting insulin (at $R^2<0.001$ and $P<5.00x10^{-8}$; unadjusted for BMI**; Supplementary Table-1**) (25). ALSPAC analyses were conducted using Stata 15.1 (StataCorp, College Station, Texas, USA).

We conducted two-sample MR analyses to examine metabolic features of type 2 diabetes liability using SNP-outcome (metabolic trait) estimates from a GWAS of 123 traits quantified using the same NMR platform as used in ALSPAC (26) among an independent sample of 13,476 to 24,925 adults of European ancestry (26), Across studies included in this GWAS, mean (SD) age ranged from 23.9y (2.1y) to 61.3y (2.9y) and female sex ranged from 37% to 64%. These estimates were combined with SNP-exposure (type 2 diabetes) estimates based on the 167 SNPs for type 2 diabetes described previously (**Supplementary Table-1**). Three statistical methods were used to generate MR estimates using the TwoSampleMR package (22): random-effects inverse variance weighted (IVW) (22), MR-Egger and weighted-median, which make differing assumptions about directional pleiotropy (27, 28). Estimates are interpreted as SD-unit differences in metabolic trait per 1 log odds of type 2 diabetes.

This study involves describing global patterns of effect estimates; we therefore provide exact P-values and focus on effect size and precision (29, 30).

Results

Sample characteristics

Eligible participants were 49.7% male with higher BMIs at later ages (**Table-1**). Prevalence of type 2 diabetes and prediabetes was very low across time points, e.g. < 5 cases of type 2 diabetes at 16y, 5 cases (0.2%) at 18y, and 7 cases (0.4%) at 25y. Participants who were ineligible for any analysis had slightly higher BMI than those who were eligible; type 2 diabetes prevalence and summary metabolic traits were also comparable (**Supplementary Table-2**).

Characteristics of the complete-case sample (N=699) were comparable to those of the full sample (**Supplementary Table-3**), and differences between excluded and included participants appeared small **(Supplementary Table-4**).

Associations of genetic liability to adult type 2 diabetes with metabolic traits at different early life stages in ALSPAC

At 8y, higher type 2 diabetes liability (per SD-higher GRS) was unassociated with lipids in most lipoprotein types including very-low density lipoproteins (VLDL) with effects of inconsistent direction and magnitudes near zero. Associations were more consistent with lower cholesterol, triglycerides, and other lipids in very-large and large high-density lipoprotein (HDL) – e.g. -0.03 SD (95% CI=-0.06, -0.003) for total lipids in very-large HDL per SD-higher GRS (**Figure-3; Supplementary Table-5**).

At 16y, higher type 2 diabetes liability was weakly but more consistently associated with higher lipids in VLDL and lower lipids in LDL. Associations were again strongest with

lower lipids in very-large and large HDL – e.g. -0.08 SD (95% CI=-0.11, -0.04) for total lipids in very-large HDL per SD-higher GRS. Pre-glycemic traits with the strongest evidence of association included citrate (-0.06 SD, 95% CI=-0.09, -0.02) and glucose (0.05 SD, 95% CI=0.02, 0.08); glycoprotein acetyls were associated at 0.05 SD, 95% CI=0.01, 0.08.

At 18y, higher type 2 diabetes liability remained weakly associated with higher lipids in VLDL and LDL, but more strongly associated with lower lipids in HDL, particularly very-large and large HDL. Associations with BCAAs and aromatic amino acids had strengthened (e.g. valine: 0.06 SD; 95% CI=0.02, 0.09 and tyrosine: 0.04 SD; 95% CI=0.001, 0.07). Associations remained stable with glycoprotein acetyls (0.06 SD; 95% CI=0.02, 0.10).

At 25y, associations had strengthened between type 2 diabetes liability and lipids in VLDL subtypes such that effect sizes were comparable to those seen with lipids in HDL subtypes – e.g. 0.05 SD (95% CI=0.01, 0.09) higher total cholesterol in VLDL vs. -0.06 SD (95% CI=-0.09, -0.02) lower total cholesterol in very-large HDL. Increasing effect size for VLDL lipids were supported by relatively low P-values for trend across time points based on linear mixed models (**Supplementary Table-5**) – e.g. P-trend=0.01 for total cholesterol in VLDL. These P-trend values were higher for lipids in HDL (e.g. P-trend=0.15 for total lipids in very-large HDL), indicating more stable effect sizes across time points. Associations were also more evident with several fatty acids including a lower ratio of linoleic-to-total fatty acids (-0.07; 95% CI=-0.10, -0.03) and lower ratios of omega-6-to-total and polyunsaturated-to-total fatty acids. Associations remained relatively strong with BCAAs – e.g. with leucine at 0.06 SD; 95% CI=0.03, 0.10, and with glycoprotein acetyls at 0.06 SD; 95% CI=0.01, 0.10.

Association patterns were comparable when using a complete-case sample (N=699) (**Supplementary Table-6; Supplementary Figure-1**), with estimates s most consistent

across time points for lipids in very-large and large HDL. Results based on non-SD units for metabolic traits are in **Supplementary Table-7-8**. Mean and SD values for metabolic traits are in **Supplementary Table-9**.

Associations of genetic liability to adult type 2 diabetes with BMI at different early life stages in ALSPAC

Type 2 diabetes liability (per SD-higher GRS) was associated with higher measured BMI on each occasion; estimates were similar to those seen for metabolic traits: 0.03 SD (95% CI=0.01, 0.05) higher BMI at 8y; 0.05 SD (95% CI=0.02, 0.08) higher BMI at 16y; 0.04 SD (95% CI=0.01, 0.07) higher BMI at 18y; and 0.04 SD (95% CI=0.003, 0.07) higher BMI at 25y. The type 2 diabetes GRS explained a low amount of variance in measured BMI (0.1% at 8y, 0.3% at 16y, 0.2% at 18y, and 0.2% at 25y).

Associations of genetic liability to adult type 2 diabetes with metabolic traits at different early life stages in ALSPAC, adjusted for BMI and fasting insulin

When adjusting for measured BMI, associations between type 2 diabetes liability and metabolic traits on each occasion were largely consistent in terms of direction and magnitude (**Supplementary Table-11**) – e.g. -0.05 SD (95% CI=-0.08, -0.01) for lipids in very-large HDL at 25y. This was also apparent when adjusting for a GRS for BMI (**Supplementary Table-12**). Associations were also consistent when adjusting for measured fasting insulin on available occasions (**Supplementary Table-13**) – e.g. 0.06 SD (95% CI=0.03, 0.09) for leucine at 25y; likewise when adjusting for a GRS for fasting insulin (**Supplementary Table-14**) – e.g. 0.05 SD (95% CI=0.01, 0.09) for glycoprotein acetyls at 25y.

Associations of genetic liability to adult type 2 diabetes with metabolic traits in adulthood in GWAS summary data

Results of two-sample MR analyses in an independent sample of adults indicated a largely persistent pattern of associations between genetic liability to type 2 diabetes and metabolic traits seen across early life (**Figure-4; Supplementary Table-10**). Higher genetic liability to type 2 diabetes was generally positively associated with VLDL lipid subtypes and inversely associated with HDL lipid subtypes, again for large and very-large HDL specifically – e.g. -0.004 SD (95% CI=-0.007, -0.002) per 1-log-odds of type 2 diabetes for total lipids in large HDL. Type 2 diabetes liability was positively associated with BCAA levels (e.g. 0.004 SD of leucine, isoleucine and valine per 1-log-odds). There was less evidence of association between type 2 diabetes liability and glycoprotein acetyls, at 0.003 SD (95% CI=0.0001, 0.005) per 1-log-odds. The strongest association was seen for glucose, at 0.008 SD (95% CI=0.006, 0.010) per 1-log-odds. Evidence of effect heterogeneity was strong for most metabolic traits – e.g. Cochrane's Q P-value=7.83×10^{-16} for the glucose IVW estimate. Where IVW estimates suggested evidence of effect, weighted median estimators were consistent whereas MR-Egger estimates were imprecise, although there was little evidence to suggest that MR-Egger intercept estimates differed from zero for metabolic traits (all $P > 0.003$).

Discussion

We aimed to reveal early metabolic features of type 2 diabetes liability by integrating genetic liability to adult disease with detailed metabolic traits measured across early life (from 8y to 25y). These metabolic traits were measured long before the expected clinical onset of type 2 diabetes (31), and consequently, their perturbations are expected to reflect early signs of disease that are detectable in circulation. Our findings suggest that one of these earliest features is lower lipid content in HDL particles, particularly in large and very-large subtypes, alongside lower citrate and higher BCAA and inflammatory glycoprotein acetyl levels. Several features are apparent in childhood as early as age 8y, several decades before the clinical onset of disease. Persistent patterns of effect were observed in an independent sample of adults using two-sample MR, supporting their continued relevance with advancing age.

Adiposity is expected to be a key driver of type 2 diabetes and its metabolic intermediates. This is supported by several MR studies suggesting strong effects of BMI on metabolic trait levels (32) and type 2 diabetes in adulthood (3, 4); and by the close resemblance of the effects of adiposity to those seen presently for type 2 diabetes liability – e.g. lower cholesterol in HDL, higher cholesterol in apolipoprotein-B-containing lipoproteins, higher BCAA levels, and higher glycoprotein acetyls. Presently, adult type 2 diabetes liability was found to raise BMI in childhood, with effect sizes appearing consistent at later time points and similar to those seen for metabolic traits. Adjusting metabolic effects of type 2 diabetes liability for BMI (phenotypically or genetically) produced little-to-no attenuation, however. This suggests that higher adiposity is one early feature of type 2 diabetes liability, but that metabolic effects of liability captured by the genetic variants used here act largely

independently of adiposity. This is further supported by lower variance captured in measured BMI by the type 2 diabetes GRS (0.2% at 25y in ALSPAC), compared with the variance explained in type 2 diabetes itself (7.1% among adults in UK Biobank (21)). The type 2 diabetes GRS also explained considerably less variance in BMI than did the BMI GRS (0.2% vs 5% at 25y in ALSPAC, respectively).

The apparent specificity of effects of type 2 diabetes liability on lipids within large and very-large HDL suggests distinct molecular functions of HDL subtypes. Medium-and-smaller HDLs have shown more overlapping immune-related gene expression profiles with triglyceride- and apolipoprotein-B-containing particles than have larger HDLs (33), suggesting that smaller HDLs are more functionally atherogenic whereas larger HDLs are more functionally involved in non-atherogenic reverse cholesterol transport and more representative of conventional/non-subtyped HDL measurements (34).

Bi-directional effects between metabolic traits and type 2 diabetes liability remain plausible. One recent MR study among adults examined effects of type 2 diabetes liability on metabolic traits derived from NMR and mass-spectrometry, and effects in the reverse direction for 20 traits (17). Genetically higher cholesterol in HDL (again in very-large and large subtypes) were most associated with lower fasting glucose, but these effects did not extend to type 2 diabetes. Metabolic traits with the strongest evidence of effect on type 2 diabetes were phospholipids in VLDL and intermediate-density lipoproteins (IDL), and total triglycerides (17). Conversely, type 2 diabetes liability had the greatest effect on alanine, along with several phosphatidylcholine alkyl-acyls, supporting such perturbations as consequences rather than causes of type 2 diabetes liability.

Another MR study that instrumented BCAAs suggested that higher levels raise type 2 diabetes risk (35). Instrumenting metabolic traits is difficult, however, because their genetic

architecture overlaps greatly (36, 37), resulting in the use of genetic variants that are typically not specific to one metabolic trait. Using the same genetic variant for multiple traits outside of a multivariable MR framework can lead to inflated MR estimates via a 'double-counting' of allele effects, as was likely the case previously (35, 37) where the variant rs1440581 was used to instrument both leucine and valine. Other MR evidence suggested that higher genetic liability to insulin resistance (a type 2 diabetes precursor) raises BCAA levels (38). Strong effects of insulin resistance were also found on higher lipids in apolipoprotein-B containing particles; on lower lipids in very-large, large, and medium HDL; and on lower citrate (38) – patterns like those seen presently. Effect sizes were much larger in that previous study, likely because insulin resistance is more biologically distinct than type 2 diabetes (a heterogenous disease), with more precise metabolic effects; and because of much older ages at which traits were measured. These results, together with ours, position such perturbations as consequences of type 2 diabetes development. Our results further indicate that effects of type 2 diabetes liability on metabolic traits are generally not attenuated with adjustment for fasting insulin (phenotypically or genetically), suggesting that effects of genetic liability to type 2 diabetes are potentially independent of insulin resistance. The effects of liability presently examined are based on a generalised disease phenotype, however. Genetic partitioning of type 2 diabetes into distinct sub-types may enable refined mechanistic insights.

Limitations

Limitations of this study include modest sample sizes and thus power/precision for ALSPAC estimates, particularly for complete-case analyses. Descriptive comparisons were made for key measured traits between excluded and included participants and these differences appeared small – e.g. BMI was 25.0 kg/m^2 vs. 24.5 kg/m^2 at age 25y in the full vs

compete-case sample, respectively. Blood samples from the first occasion of metabolic trait assessment were derived while non-fasting, but trait concentrations have previously shown stability over different durations of fasting time (39). Adjustments were made for measured BMI and fasting insulin at the time of metabolic trait measurement to assess mediation via estimate attenuation; this carries potential for collider bias from unmeasured confounding of the mediator-outcome association (40), but close agreement between phenotypic and genetic adjustments suggests this bias is minor.

Our analyses were restricted to white-Europeans; this helps to reduce confounding by ancestral population structure but limits inference to other groups. This requires more comprehensive GWAS studies of non-white-European populations together with metabolomic measurements in cohort studies with higher representation of those groups. Our two-sample MR analysis confirmed the same metabolic features in an independent sample, but differences in exposure units prevent direct comparison of effect sizes. We aimed to reveal early metabolic features of type 2 diabetes, but objectives are only feasible within a framework of 'liability' to type 2 diabetes because the current study population is without clinical disease. This was deliberate as part of an approach for identifying early features of disease activity in early life, with pre-clinical implications. Lastly, although we interpret results as reflecting metabolic effects of type 2 diabetes liability, alternative explanations including bias or pleiotropy remain possible (7 such scenarios are proposed (14)). Methodological flexibility together with an increasingly large scale and scope of genomic and metabolomic data should make interrogating these scenarios increasingly feasible.

Conclusions

Our results based on genetic liability to adult type 2 diabetes in relation to repeated measures of detailed metabolic traits across early life suggest that one of the earliest metabolic features of type 2 diabetes liability is lower lipid content in HDL particles – particularly in very-large and large subtypes – alongside lower citrate and higher BCAA and inflammatory glycoprotein acetyl levels. Several features are apparent in childhood as early as age 8y, several decades before the clinical onset of disease.

Acknowledgements

We are grateful to the families who participated in this study, the midwives for their help in recruiting them, and the entire ALSPAC team which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

Individual-level ALSPAC data are available following an application. This process of managed access is detailed at www.bristol.ac.uk/alspac/researchers/access. Cohort details and data descriptions for ALSPAC are publicly available at the same web address. Summary-level GWAS data used in this study are publicly available without the need for application through the MR-Base platform, which is accessible at http://www.mrbase.org/.

Funding: The UK Medical Research Council, Wellcome (102215/2/13/2), and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. This publication is the work of the authors who are guarantors for its contents. JAB is supported by Cancer Research UK (C18281/A19169) and the Elizabeth Blackwell Institute for Health Research, University of Bristol and the Wellcome Trust Institutional Strategic Support Fund (204813/Z/16/Z). EEV and CB are supported by Diabetes UK (17/0005587). NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/2/13/2), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-20011), the MRC Integrative Epidemiology Unit (MC_UU_12013/3) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). DC and GDS work in a unit funded by the UK Medical Research Council

(MC_UU_00011/1,6) and the University of Bristol. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: All authors contributed to the planning of the study. JAB and CJB had access to data and conducted analyses. JAB, CJB, and EEV wrote the first draft. All authors critically reviewed the intellectual content of manuscript drafts and approved the final version for submission. JAB is the guarantor.

Competing interests: As of January 2020, AM is an employee of Genentech, and a holder of Roche stock. No others to declare.

References

1. International Diabetes Federation. IDF Diabetes Atlas: Eighth edition. 2017.
2. Abarca-Gómez L, Abdeen ZA, Hamid ZA, Abu-Rmeileh NM, Acosta-Cazares B, Acuin C, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. Lancet. 2017;390(10113):2627-42.
3. Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Davey Smith G, et al. Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation. Diabetes. 2016:db160418.
4. Dale CE, Fatemifar G, Palmer TM, White J, Prieto-Merino D, Zabaneh D, et al. Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus. Circulation. 2017;135(24):2373-88.
5. Dombrowski SU, Knittle K, Avenell A, Araujo-Soares V, Sniehotta FF. Long term maintenance of weight loss with non-surgical interventions in obese adults: systematic review and meta-analyses of randomised controlled trials. BMJ. 2014;348:g2646.
6. Association AD. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2020. Diabetes Care. 2020;43(Supplement 1):S14-S31.
7. Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. The Lancet. 2009;373(9682):2215-21.
8. Würtz P, Soininen P, Kangas AJ, Rönnemaa T, Lehtimäki T, Kähönen M, et al. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. Diab Care. 2013;36(3):648-55.
9. Würtz P, Mäkinen V-P, Soininen P, Kangas AJ, Tukiainen T, Kettunen J, et al. Metabolic signatures of insulin resistance in 7,098 young adults. Diabetes. 2012:DB_111355.
10. Mahendran Y, Cederberg H, Vangipurapu J, Kangas AJ, Soininen P, Kuusisto J, et al. Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. Diab Care. 2013:DC_130800.
11. Würtz P, Tiainen M, Mäkinen V-P, Kangas AJ, Soininen P, Saltevo J, et al. Circulating metabolite predictors of glycemia in middle-aged men and women. Diab Care. 2012:DC_111838.
12. Mahendran Y, Vangipurapu J, Cederberg H, Stančáková A, Pihlajamäki J, Soininen P, et al. Association of ketone body levels with hyperglycemia and type 2 diabetes in 9,398 Finnish men. Diabetes. 2013:DB_121363.
13. Wang J, Stančáková A, Soininen P, Kangas A, Paananen J, Kuusisto J, et al. Lipoprotein subclass profiles in individuals with varying degrees of glucose tolerance: a population-based study of 9399 Finnish men. J Intern Med. 2012;272(6):562-72.
14. Holmes MV, Davey Smith G. Can Mendelian Randomization Shift into Reverse Gear? Clin Chem. 2019;65(3).
15. Stančáková A, Paananen J, Soininen P, Kangas AJ, Bonnycastle LL, Morken MA, et al. Effects of 34 risk loci for type 2 diabetes or hyperglycemia on lipoprotein subclasses and their composition in 6,580 nondiabetic Finnish men. Diabetes. 2011:DB_101655.
16. Stančáková A, Civelek M, Saleem NK, Soininen P, Kangas AJ, Cederberg H, et al. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. Diabetes. 2012:DB_111378.
17. Liu J, van Klinken JB, Semiz S, van Dijk KW, Verhoeven A, Hankemeier T, et al. A mendelian randomization study of metabolite profiles, fasting glucose and type 2 diabetes. Diabetes. 2017:db170199.

18. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Gen. 2014;23(R1):R89-R98.
19. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42(1):97-110.
20. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson 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. 2012(1):111-27.
21. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nature Genet. 2018;50(11):1505.
22. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.
23. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum NMR Metabolomics in Large-Scale Epidemiology: A Primer on-Omic Technology. Am J Epidemiol. 2017:kwx016.
24. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Gen. 2019;28(1):166-74.
25. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan Ja, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44(9):991-1005.
26. Kettunen J, Demirkan A, Würtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun. 2016;7:11122.
27. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. Epidemiology. 2017;28(1):30.
28. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40(4):304-14.
29. Sterne JA, Davey Smith G. Sifting the evidence—what's wrong with significance tests? BMJ. 2001;322(7280):226-31.
30. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. Am Statistician. 2016;70(2):129-33.
31. Sharma M, Nazareth I, Petersen I. Trends in incidence, prevalence and prescribing in type 2 diabetes mellitus between 2000 and 2013 in primary care: a retrospective cohort study. BMJ Open. 2016;6(1):e010210.
32. Würtz P, Wang Q, Kangas AJ, Richmond RC, Skarp J, Tiainen M, et al. Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. PLoS Med. 2014;11(12):e1001765.
33. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol Systems Biol. 2010;6(1).
34. Tall A. An overview of reverse cholesterol transport. Eur Heart J. 1998;19:A31-5.
35. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan Ja, Tillin T, et al. Genetic predisposition to an impaired metabolism of the branched-chain amino acids and risk of type 2 diabetes: a Mendelian randomisation analysis. PLoS Med. 2016;13(11):e1002179.
36. Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46(6):543-50.
37. Holmes MV, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol. 2017;14:577-90.

38. Wang Q, Holmes MV, Davey Smith G, Ala-Korpela M. Genetic support for a causal role of insulin resistance on circulating branched-chain amino acids and inflammation. Diab Care. 2017:dc171642.
39. Sidhu D, Naugler C. Fasting time and lipid levels in a community-based population: a cross-sectional study. Arch Intern Med. 2012;172(22):1707-10.
40. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013;42(5):1511-9.

Table-1 Summary metabolic traits at different early life stages among 6,216 ALSPAC offspring eligible for inclusion in ≥ 1 analysis

	Childhood	***Adolescence***	***Young adulthood***	***Adulthood***
Age – mean (SD)	7.5 (0.3)	15.5 (0.3)	17.8 (0.4)	24.5 (0.8)
Male – % (N)	49.7 (3,087)	49.7 (3,087)	49.7 (3,087)	49.7 (3,087)
Body mass index (kg/m^2) – mean (SD)	16.2 (2.0)	21.4 (3.5)	22.7 (4.02)	24.8 (4.9)
Has type 2 diabetes – % (N)	NA	NA (< 5)	0.2 (5)	0.4 (7)
Has prediabetes – % (N)	NA	0.4 (11)	0.5 (12)	0.4 (10)
Lipid traits				
Total cholesterol (mmol/l) – mean (SD)	3.9 (0.6)	3.5 (0.6)	3.5 (0.7)	3.6 (0.8)
Cholesterol in VLDL (mmol/l) – mean (SD)	0.6 (0.2)	0.5 (0.1)	0.6 (0.2)	0.4 (0.2)
Cholesterol LDL (mmol/l) – mean (SD)	1.2 (0.3)	1.04 (0.3)	1.03 (0.4)	1.2 (0.4)
Cholesterol HDL (mmol/l) – mean (SD)	1.5 (0.2)	1.4 (0.2)	1.4 (0.2)	1.4 (0.3)
Total triglycerides (mmol/l) – mean (SD)	1.1 (0.4)	0.9 (0.3)	0.9 (0.3)	0.9 (0.4)
Triglycerides in VLDL (mmol/l) – mean (SD)	0.7 (0.3)	0.6 (0.3)	0.6 (0.3)	0.6 (0.4)
Triglycerides in LDL (mmol/l) – mean (SD)	0.1 (0.1)	0.1 (0.04)	0.1 (0.1)	0.1 (0.04)
Triglycerides in HDL (mmol/l) – mean (SD)	0.1 (0.02)	0.1 (0.02)	0.1 (0.02)	0.1 (0.03)
Pre-glycemic traits				
Lactate (mmol/l) – mean (SD)	1.4 (0.5)	1.3 (0.6)	1.0 (0.5)	0.9 (0.5)
Citrate (mmol/l) – mean (SD)	0.1 (0.03)	0.1 (0.02)	0.1 (0.02)	0.2 (0.02)
Isoleucine (mmol/l) – mean (SD)	0.1 (0.02)	0.04 (0.01)	0.04 (0.01)	0.1 (0.01)
Leucine (mmol/l) – mean (SD)	0.1 (0.01)	0.1 (0.01)	0.1 (0.01)	0.1 (0.01)
Valine (mmol/l) – mean (SD)	0.1 (0.03)	0.1 (0.03)	0.1 (0.03)	0.1 (0.03)
Glucose (mmol/l) – mean (SD)	4.2 (0.5)	4.3 (0.3)	4.1 (0.5)	3.9 (0.4)
Inflammatory traits				
Glycoprotein acetyls (mmol/l) – mean (SD)	1.2 (0.1)	1.2 (0.1)	1.2 (0.1)	1.2 (0.2)

Type 2 diabetes is defined in adolescence as a clinic fasting glucose ≥ 7 mmol/l, and in young adulthood and adulthood as a clinic fasting glucose ≥ 7 mmol/l or reported physician diagnosis. Prediabetes is defined as fasting glucose between 5.6 mmol/l and 6.9 mmol/l.

Figure-1 Distribution of the genetic risk score (GRS) for adult type 2 diabetes among ALSPAC offspring

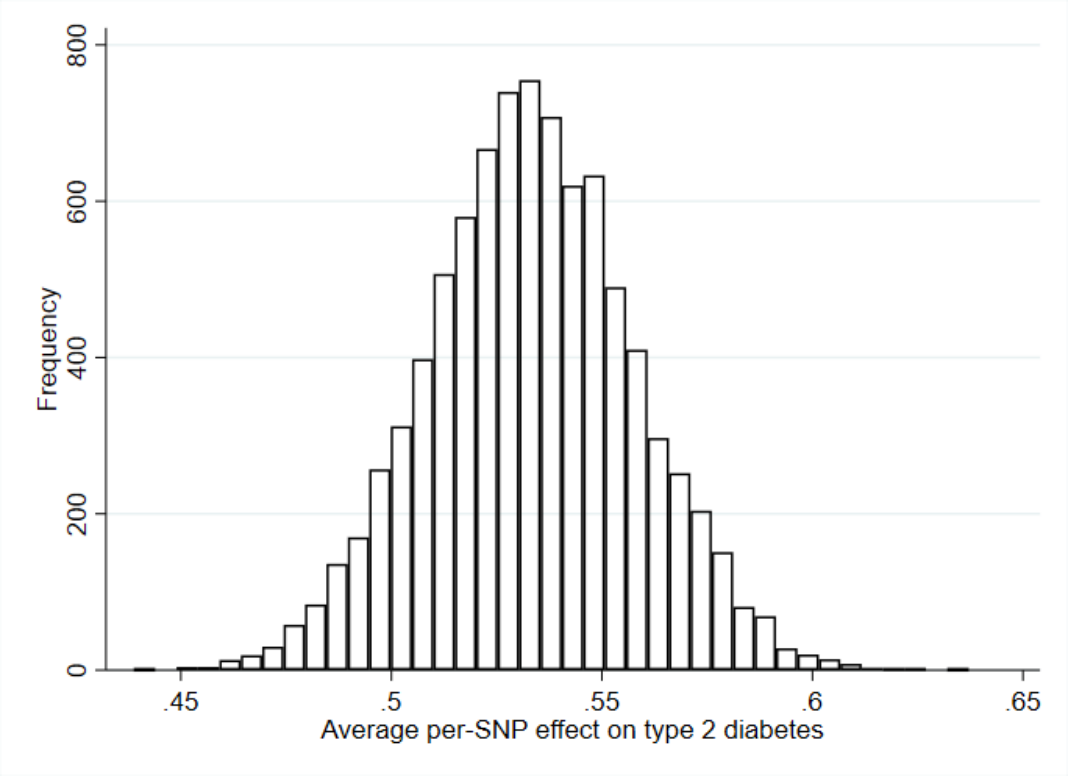


Figure-2 Selection of ALSPAC participants into analyses

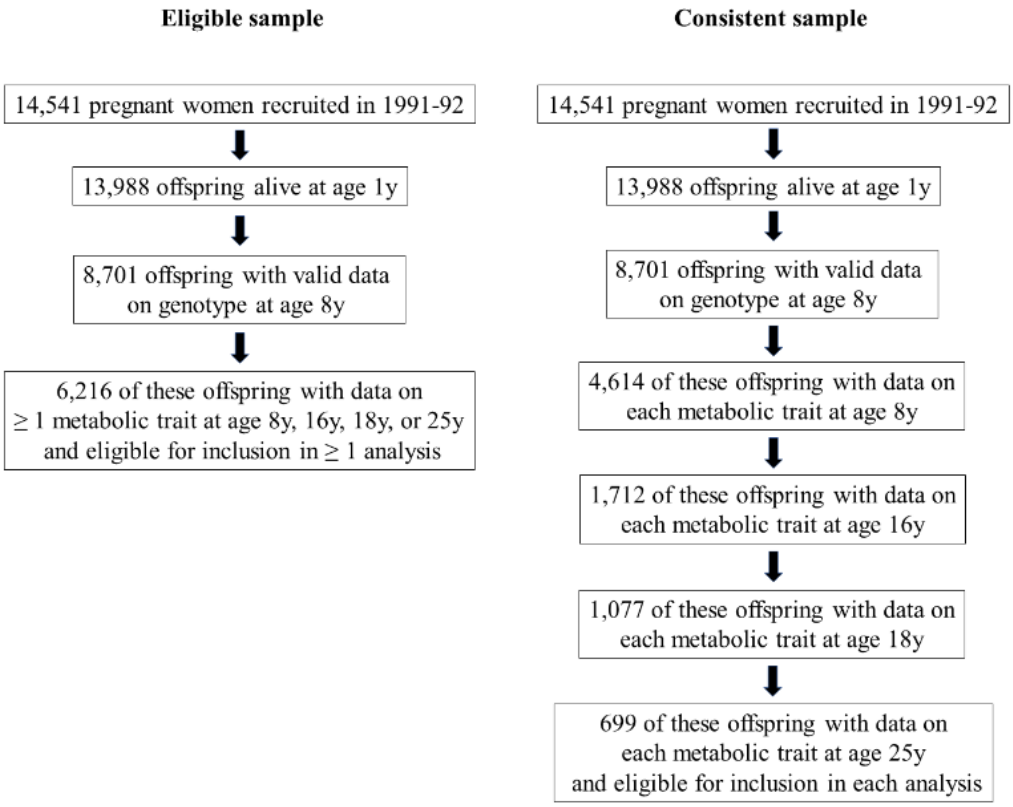


Figure-3 Associations of genetic liability to adult type 2 diabetes with metabolic traits at different early life stages among ALSPAC offspring

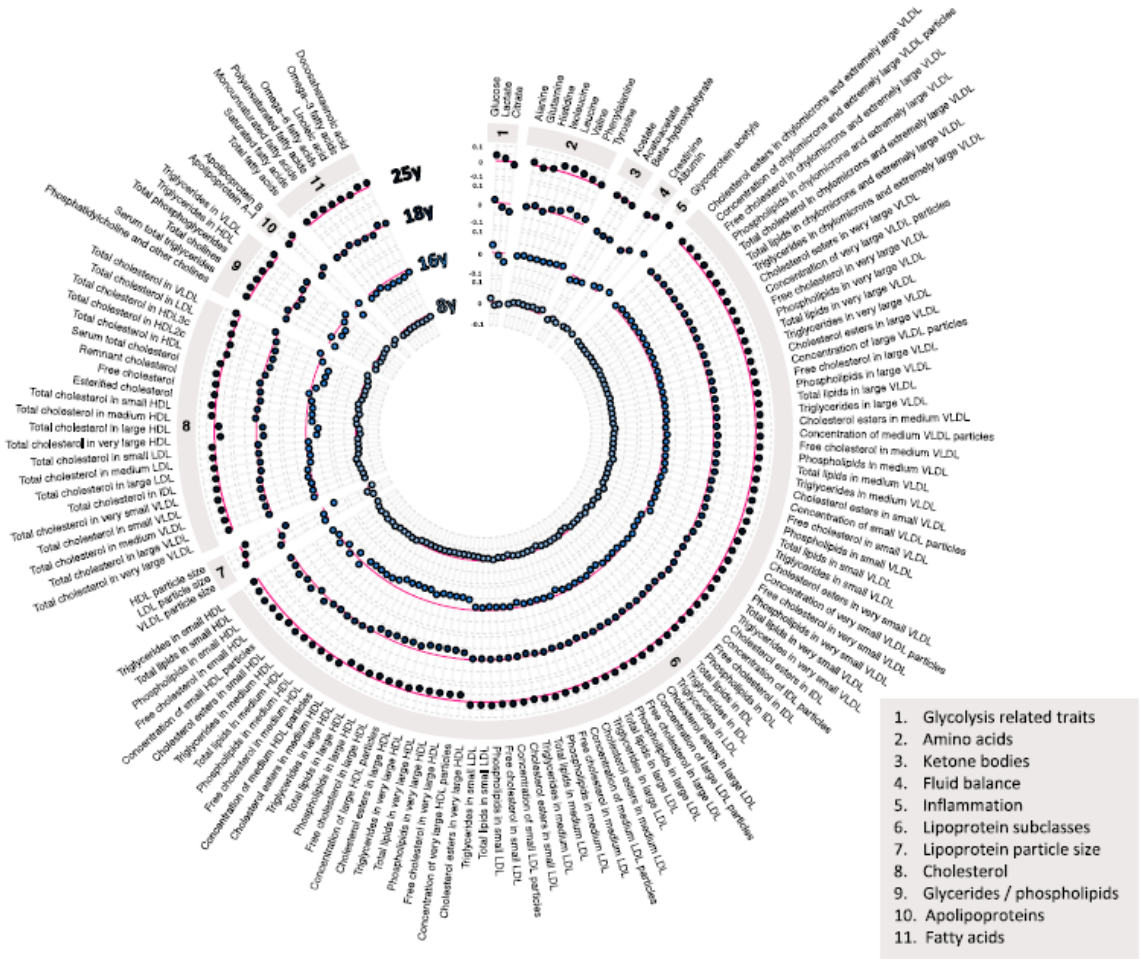


Figure-3 Legend Estimates shown are beta coefficients representing SD difference in metabolic trait per SD higher GRS for type 2 diabetes, ordered concentrically (inner circle to outer circle) by increasing age at measurement. Six metabolic traits were not measured at the 25y time point: diacylglycerol, ratio of diacylglycerol to triglycerides, fatty acid chain length, degree of unsaturation, conjugated linoleic acid, and ratio of conjugated linoleic acid to total fatty acids.

Figure-4 Associations of genetic liability to adult type 2 diabetes with metabolic traits in an independent sample of adults based on two-sample MR

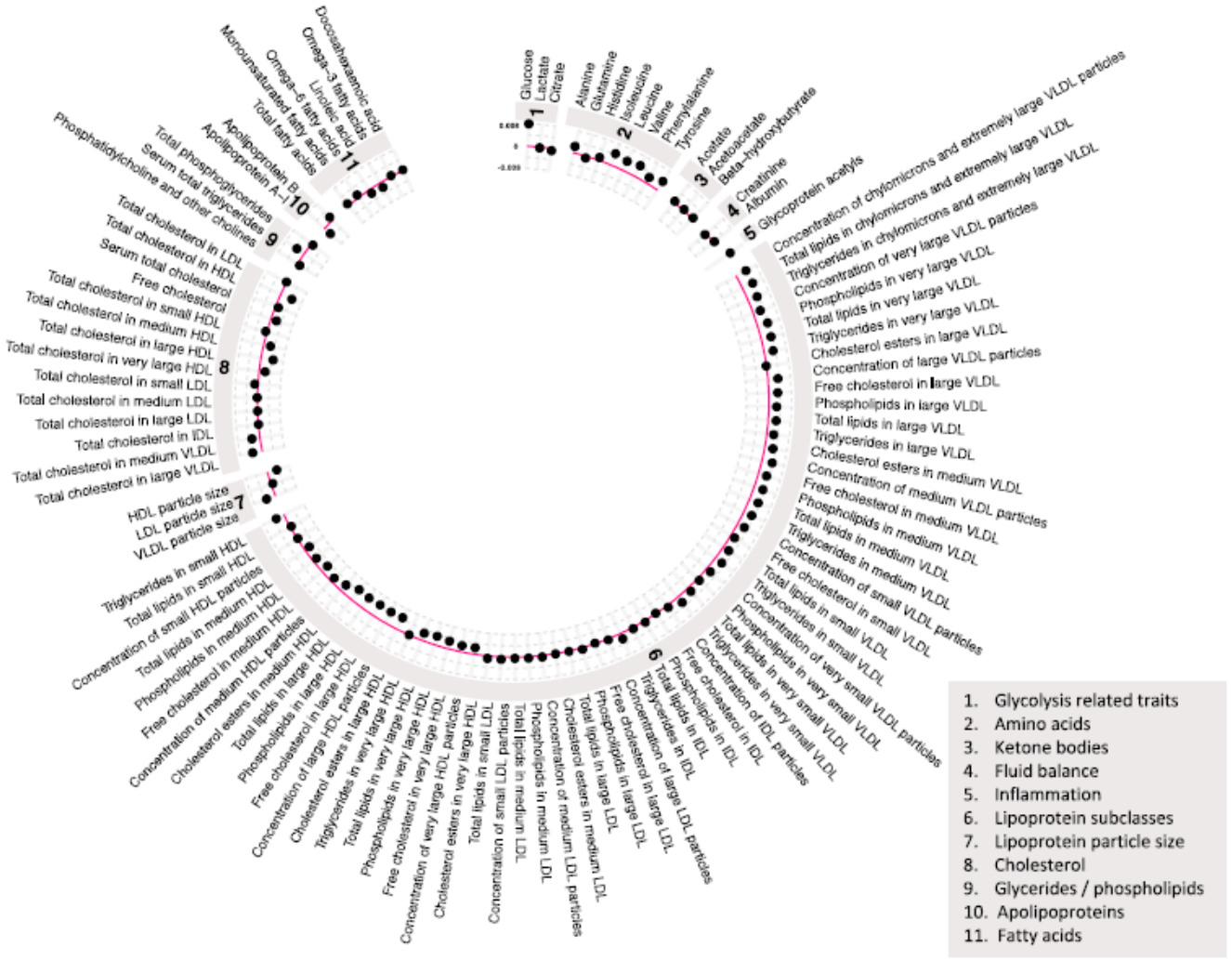
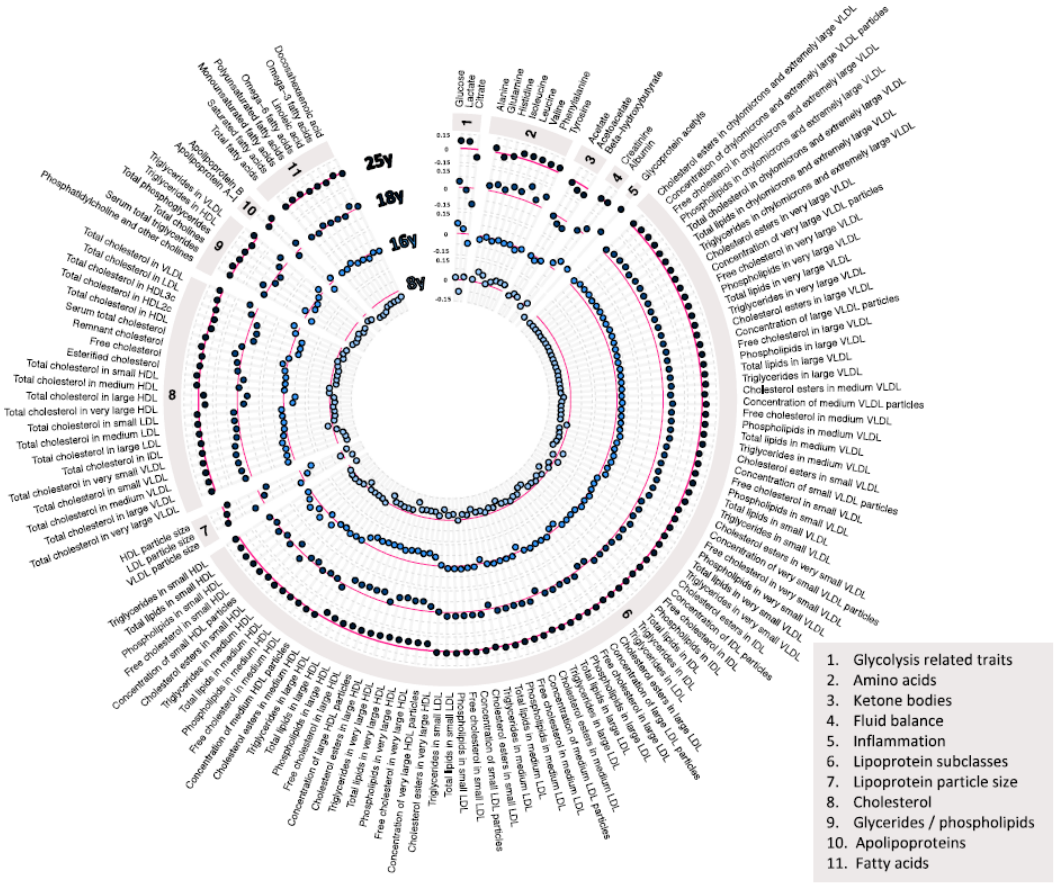


Figure-4 Legend Estimates shown are beta coefficients representing the SD-unit difference in metabolic trait per 1 log odds of type 2 diabetes based on the IVW method.

Supplementary Figure 1 Associations of genetic liability to adult type 2 diabetes with metabolic traits at different early life stages among a complete-case sample (N=699)



Supplementary Figure-1 Legend Estimates shown are beta coefficients representing SD difference in metabolic trait per SD higher GRS for type 2 diabetes, ordered concentrically (inner circle to outer circle) by increasing age at measurement. Six metabolic traits were not measured at the 25y time point: diacylglycerol, ratio of diacylglycerol to triglycerides, fatty acid chain length, degree of unsaturation, conjugated linoleic acid, and ratio of conjugated linoleic acid to total fatty acids.