BMJ Open

Metabolomics: population epidemiology and concordance in Australian children aged 11–12 years and their parents

Susan Ellul,[1] Melissa Wake,[1,2,3] Susan A Clifford,[1,3] Katherine Lange,[1,3] Peter Würtz,[4,5] Markus Juonala,[1,6,7] Terence Dwyer,[1,3,8,9] John B Carlin,[1,3] David P Burgner,[1,3,10] Richard Saffery[1,3]

For numbered affiliations see end of article.

Correspondence to
Professor Melissa Wake;
melissa.wake@mcri.edu.au

ABSTRACT

Objectives Nuclear magnetic resonance (NMR) metabolomics is high throughput and cost-effective, with the potential to improve the understanding of disease and risk. We examine the circulating metabolic profile by quantitative NMR metabolomics of a sample of Australian 11–12 year olds children and their parents, describe differences by age and sex, and explore the correlation of metabolites in parent–child dyads.

Design The population-based cross-sectional Child Health CheckPoint study nested within the Longitudinal Study of Australian Children.

Setting Blood samples collected from CheckPoint participants at assessment centres in seven Australian cities and eight regional towns; February 2015–March 2016.

Participants 1180 children and 1325 parents provided a blood sample and had metabolomics data available. This included 1133 parent–child dyads (518 mother–daughter, 469 mother–son, 68 father–daughter and 78 father–son).

Outcome measures 228 metabolic measures were obtained for each participant. We focused on 74 biomarkers including amino acid species, lipoprotein subclass measures, lipids, fatty acids, measures related to fatty acid saturation, and composite markers of inflammation and energy homeostasis.

Results We identified differences in the concentration of specific metabolites between childhood and adulthood and in metabolic profiles in children and adults by sex. In general, metabolite concentrations were higher in adults than children and sex differences were larger in adults than in children. Positive correlations were observed for the majority of metabolites including isoleucine (CC 0.33, 95% CI 0.27 to 0.38), total cholesterol (CC 0.30, 95% CI 0.24 to 0.35) and omega 6 fatty acids (CC 0.28, 95% CI 0.23 to 0.34) in parent–child comparisons.

Conclusions We describe the serum metabolite profiles from mid-childhood and adulthood in a population-based sample, together with a parent–child concordance. Differences in profiles by age and sex were observed. These data will be informative for investigation of the childhood origins of adult non-communicable diseases and for comparative studies in other populations.

Strengths and limitations of this study

► In a large population-based cohort, venous blood was collected for children and their attending parent on the same day using the same methods.

► Rapidly processed, high-quality serum samples with standardised metabolomic data generated as a single batch.

► Cross-sectional design does not enable longitudinal analysis of specific metabolite species over the short term or longer periods of time.

► Assessment of paternal associations with offspring metabolite measures is limited by a relatively small sample size compared with mother–child pairs, reducing the precision of estimates.

► Factors known to influence metabolomic profile (such as body mass index) were not considered as the aim was to describe the distribution of metabolites in children and their parents.

INTRODUCTION

Metabolomics involves the quantitative analysis of a large number of metabolites and lipids involved in a diverse range of biochemical pathways.[1] Genetic/gene expression and environmental exposures are associated with specific metabolic changes across many tissues and body fluids.[2,3] As such, metabolomics is recognised as a powerful top-down approach to understanding genetic and environmental influences on health and disease. Metabolomic profiling also has considerable potential to identify clinically relevant biomarkers for risk stratification and disease monitoring.

Recent advances in nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry have enabled the simultaneous quantitative measurement of hundreds of metabolites. These approaches are sufficiently cost-effective and high throughput to be applicable to large cohort studies. For

example, NMR metabolomics of serum from the Cardio-vascular Risk in Young Finns Study identified many biomarkers from multiple metabolic pathways reflective of fatty liver disease.[4] These were also predictive of risk 10 years prior to diagnosis, indicating that metabolic disruptions precede overt phenotype. Similar population and disease-specific studies have identified metabolomic profiles associated with a range of exposures and health outcomes with the potential to reveal clinically important biomarkers and information on disease mechanisms.[5] In addition, specific serum metabolites can also be considered 'intermediate phenotypes' linking genetic risk with disease outcomes.[6 7]

Previous research indicates that some blood metabolites change with age, particularly from mid to late adulthood.[8 9] However, in adults, sex appears to be a major driver of variation in metabolite profile, potentially interacting with age. For example, the effects of sex appeared to be greater in younger (age 25–35 years) than older Japanese adults.[10] A study of 26 000 Northern European adults identified many sex-specific metabolic species at the population level.[9] In men, several lipid measures begin to rise at early middle age, whereas a similar increase is only observed in women postmenopause. This pattern is consistent for all non-high-density lipoprotein (HDL) cholesterol measures, very low-density lipoprotein, intermediate density lipoprotein and low-density lipoprotein (LDL) subclass particle concentrations, as well as for triglycerides.[9] Physiological states such as pregnancy also have a consistent and measurable influence on serum metabolome.[11] However, it remains unclear how the serum metabolome differs in adults compared with children and by sex, particularly in childhood.

Moreover, factors regulating the metabolic trajectory from early life to adulthood, the role of a metabolomic profile in health at the population level and the extent to which blood metabolomic profiles are concordant for parents and children have not been fully explored. One small study has reported correlations between parents (n=179) and their offspring (n=255) for a range of cardiometabolic risk factors including standard lipid profile measured using conventional methods; this proved stronger for total cholesterol and LDL cholesterol than for HDL cholesterol or triglycerides.[12] Considerable evidence exists that the metabolomic profile is regulated, at least in part, by genetic factors[13 14] and is also influenced by dietary and lifestyle factors. Each of these influences is likely to be shared between parents and their offspring to varying degrees, however, parent–child correlations of metabolites from NMR-based platforms have not been reported previously.

Here, we describe (1) the distribution of NMR-based metabolite measures in a population-based cohort of 11–12-year-old children and their parents, differences in metabolite concentrations; (2) by age (adults compared with children); (3) by sex in children and adults; and (4) report sex-specific parent–child concordance.

METHODS
Study design
Details of the initial Longitudinal Study of Australian Children (LSAC) study design and recruitment are outlined elsewhere.[15 16] The LSAC commenced in 2004 when two cohorts (the 'B' and 'K' cohorts, of which the B cohort only was included in the present study) were recruited who have since been followed biennially. The Child Health CheckPoint comprised a detailed cross-sectional assessment of physical health and biomarkers in a population-based national sample of children (age 11–12 years) and their parents between February 2015 to March 2016. The CheckPoint was nested between waves 6 (2014) and 7 (2016) of the LSAC. Further details regarding the CheckPoint study design and methods are available elsewhere.[17 18]

Participants
Of the 8921 families contacted to be part of the LSAC B cohort, 5107 families (57%) agreed to take part in the first wave of data collection in 2004; 4484 families were retained for Wave 6 in 2014. During the Wave 6 LSAC home visit, B cohort families were introduced to the upcoming Child Health CheckPoint and asked to consent to their contact details being shared with the CheckPoint team. A total of 3513 families provided permission to receive an information pack by mail and an information and recruitment phone call regarding the CheckPoint study (78% of Wave 6 cohort, 69% of the original cohort). Of the families agreeing to receive information about the CheckPoint study, 1874 families took part (53% of eligible participants, 42% of Wave 6 cohort and 37% of the original cohort).

Consent
The attending parent/caregiver provided written informed consent for themselves and their child to participate in the study, and asked to provide optional consent for the collection and use of biological samples.

Procedure
The specialised CheckPoint assessment centre sequentially visited seven Australian cities and eight regional towns between February 2015 and March 2016.[18] Each participating child attended the centre with one parent or caregiver (usually the biological mother) at which both participated in a wide range of measures relevant to non-communicable disease. Those families that could not attend a centre were offered a home visit. Participants were included in the current analyses if metabolomic data from CheckPoint were available (figure 1). Venous blood was not available for home-visit participants but was collected at all city and most regional assessment centres. Participant pairs were excluded from the concordance analyses in this study if the attending parent was not the biological parent.

An experienced phlebotomist collected approximately 28 mL of blood from the brachial vein of the non-dominant
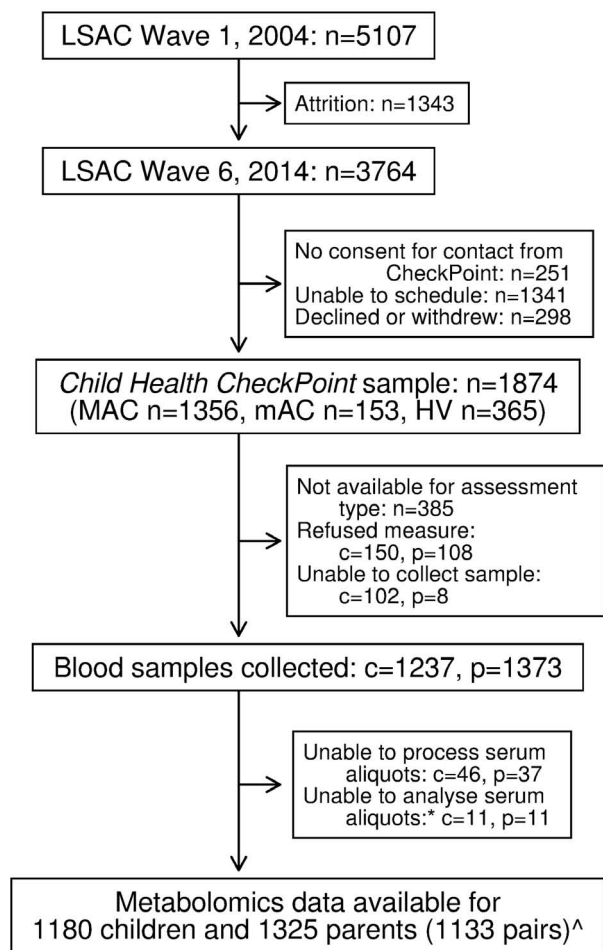
Figure 1 Participant flow chart. *Unable to analyse due to insufficient volume or poor quality sample. Data from six non-biological child–parent pairs excluded from concordance analyses. c, number of children; HV, home visit; LSAC, Longitudinal Study of Australian Children; MAC, main assessment centre; mAC, mini assessment centre; n, number of families; p, number of attending adults.

arm of semi-reclining, semi-fasted participants (at the time of collection, participants reported when they last ate or drank). Blood was collected sequentially into four Becton Dickinson (BD) Vacutainer tubes using a butterfly needle so only a single venepuncture was required. Order of collection was (1) 2.7 mL EDTA, (2) 9 mL EDTA, (3) 9 mL serum and (4) 7.5 mL Lithium Heparin. The latter two tubes were immediately inverted six times to ensure mixing with anticoagulant, and all tubes were transferred to the on-site laboratory. Time of collection was scheduled earlier in the visit for parents than for children.

Collection tube barcodes were linked to the participant and samples were immediately transported to an on-site laboratory where they were processed within 2 hours. Blood clotting was allowed at room temperature for at least 30 min after collection. The sample tubes were spun at 550 g relative centrifugal force for 10 min at room temperature and distributed into 0.5 mL aliquots of plasma, serum, buffy coat (lymphocytes), whole blood and/or an aliquot tube containing a blood clot (1.0 mL

FluidX screwcap tubes, Cheshire, UK) and stored immediately at −80°C (Thermo Fisher Scientific, Waltham, Massachusetts,USA). Each FluidX tube contained a unique 2D barcode linked to the original collection tube and participant. As each assessment centre closed, samples were shipped on dry ice to the Melbourne Children's Bioresource Centre for long term storage at −80°C (serum, whole blood, plasma, blood clot) or vapour phase liquid nitrogen (lymphocytes). At a later date, single 0.5 mL serum aliquot was removed for every CheckPoint participant and the combined aliquots were shipped in a single batch to Nightingale Health (Helsinki, Finland) on dry ice for NMR metabolomics.

Measures
Metabolomic profiling
The Nightingale NMR metabolomics platform (Helsinki, Finland) was used to obtain metabolomics for children and parents using the 2016-version quantification algorithm. Details of this platform and methodology have been extensively described elsewhere,[6 19] and epidemiological applications were recently reviewed.[20] Briefly, metabolites were measured from 0.35 mL of serum using a single high-throughput experimental set up for the simultaneous quantification of routine lipids, lipoprotein subclass distributions, particle size and composition, fatty acids and other low-molecular-weight metabolites such as amino acids and glycolysis-related metabolites. This generated data on 228 serum metabolite measures in absolute concentration units (eg, millimoles per litre) and ratios (summarised in table 1). Although widely used for epidemiological research, the NMR-based quantification has not been certified for clinical diagnostics. Further analytical validation of the quantification protocols for the biomarker subset routinely used in clinical settings (eg, established cholesterol measures and creatinine) is expected to lead to a recalibration of certain metabolite concentrations to better match clinical gold standards.[20]

Many of the 228 metabolomics measures correlate substantially both in children (see the online supplementary figure 1) and adults (see the online supplementary figure 2) and the pattern of correlations were similar for children and adults. For clarity, we therefore focused on a subset of 74 metabolites in analyses. We eliminated the five ratio measures for each of the 14 lipoprotein subclass particles. In addition, the seven other measures within each of the lipoproteins (esterified cholesterol, free cholesterol, total cholesterol, triglycerides, phospholipids, total lipids and particle concentration) are all highly correlated and therefore we only reported total lipids for each of the lipoprotein subclass particles.

Other measures and sample characteristics
Age and sex: for children, LSAC provided date of birth (DOB) and sex, which were originally exported from the Medicare Australia database. In parents, DOB and sex were self-reported in the CheckPoint questionnaire, which was administered on an iPad using the Research

Table 1 Summary of biomarkers and derived variables obtained via high-throughput NMR

Metabolic group	Species and derived measures
Amino acids	Alanine, glutamine, glycine, histidine
	Branched chain: isoleucine, leucine, valine
	Aromatic: phenylalanine, tyrosine
Cholesterol	VLDL, LDL, HDL, HDL2, HDL3, total, free, esterified, remnant
Triglycerides and phospholipids	Triglycerides (VLDL, LDL, HDL, total)
	Phosphoglycerides
	Ratio of triglycerides to phosphoglycerides*
	Phosphatidylcholine
	Sphingomyelins
	Total cholines
Apolipoproteins	ApoA-1
	Apo B
	Ratio of ApoB/Apo A-1*
Fatty acids (FA)	Total, omega-3, omega-6, PUFA, SFA
	MUFA, DHA, LA
	Estimated degree of unsaturation
Fatty acid ratios	Omega-3/total FA*, omega-6/total FA*, PUFA/total FA* (all %)
	SFA/total FA*, MUFA/total FA*, DHA/total FA*, LA/total FA* (all %)
Lipoprotein subclasses*	12 lipid measures in each of 14 subclasses VLDL (XXL, XL, L, M, S, XS), IDL, LDL (L,M,S), HDL (XL, L, M,S): particle concentration, total lipids, esterified cholesterol, total cholesterol, phospholipids, free cholesterol, triglycerides and esterified cholesterol/total lipids (%), free cholesterol/total lipids (%), total cholesterol/total lipids (%), triglycerides/total lipids (%) and phospholipids/total lipids (%)
Lipoprotein size*	Mean diameter of VLDL, LDL and HDL particles
Ketone bodies	Acetate, acetoacetate, 3-hydroxybutyrate
Glycolysis related	Glucose, lactate, pyruvate, citrate, glycerol
Fluid balance	Creatinine, albumin
Inflammation	Glycoprotein acetyls (GlycA)

Information obtained from https://nightingalehealth.com/science/biomarkers.
*Ratio.
ApoA-1, apolipoprotein A-1; Apo B, apolipoprotein B; DHA, docosahexaenoic acid; GlycA, glycoprotein acetyls; HDL, high-density lipoprotein; IDL, intermediate density lipoprotein; L, large; LA, linoleic acid; LDL, low-density lipoprotein; M, medium; MUFA, monounsaturated fatty acid; PUFA, polyunsaturated fatty acid; S, small; SFA, saturated fatty acids; VLDL, very low-density lipoprotein; XL, very large; XXL, chylomicrons and extremely large; XS, very small.

Electronic Data Capture (REDCap) tool.[21] Age in years was calculated as the difference between date of the CheckPoint assessment and DOB divided by 365.

Disadvantage index: LSAC provided contact details of families consenting to be contacted by CheckPoint. The family's residential postcode was confirmed during the CheckPoint recruitment phone call and updated if required. The disadvantage index score of postcode was used to summarise neighbourhood socioeconomic position. Generated by the ABS from the 2011 national Census, the index numerically summarises the social and economic conditions of Australian neighbourhoods; national mean 1000, SD 100; higher scores indicate less disadvantage.[22]

Time of blood collection, processing and fasting time: time of blood collection and the start of laboratory processing were recorded. When missing, collection time was estimated using the midpoint between the time the CheckPoint visit began and time that processing of the sample commenced. Processing lag time was calculated as the minutes between blood collection and the processing commencement. Most samples were processed within 2 hours.

Fasting time was calculated as the hours between last eating/drinking to the time of blood collection. The last time of eating/drinking was cross-checked against when the participant was taking part in other CheckPoint stations (and known not to be eating) as well as sleep and

wake times from accelerometry data (to identify the usual activity, and therefore likely eating patterns) when available. Further details of cleaning processes for the time of last eat/drink can be found elsewhere.[23]

Statistical analysis
Sample characteristics
Continuous descriptive variables and metabolite measures were summarised using means and SD for children and adults separately, by sex and overall. For skewed metabolites, geometric means and relative SD were reported. To provide visual comparisons of distributions of metabolites by age and sex, density plots were used. Population summary statistics were estimated by applying survey weights and survey procedures that corrected for sampling, participation and non-response biases, and took into account clustering in the sampling frame. Standard errors were calculated taking into account the complex design and weights.[24] More detail on the calculation of weights is provided elsewhere.[25]

Differences in metabolite concentration by age (adults compared with children) and by sex (adults and children)
Skewed metabolites (skewness ≥2) were log-transformed. We used two-sided paired and unpaired t-tests (as appropriate) to assess differences in mean metabolite concentrations between adults and children in parent–child dyads, and between males and females for adults and children separately. P values were adjusted using Benjamini-Hochberg (B-H) with a false discovery rate (FDR) of 10% to account for multiple comparisons.

Parent–child concordance
Concordance between parents and children was assessed by (1) Pearson's correlation coefficients (CC) with 95% CIs, and (2) partial correlation coefficients (PCC), adjusting for child and parent age, disadvantage index, fasting time and processing lag time (and for child and parent sex where appropriate). Scatterplots of parent versus child metabolites (log-transformed where needed as above) were examined to check for outliers and to ensure assumptions were met.

The analyses were repeated using weighted multilevel survey analyses and compared with unweighted analyses. As there appeared to be no major effect of response patterns on results, we reported results from unweighted analyses. Analyses were undertaken using Stata V.14.2 and R V.3.3.2.[26]

Patient and public involvement
Because LSAC is a population-based longitudinal study, no patient groups were involved in its design or conduct. To our knowledge, the public was not involved in the study design, recruitment or conduct of the LSAC study or its CheckPoint module. Parents received a summary health report for their child and themselves at or soon after the assessment visit. They consented to take part knowing that they would not otherwise receive individual results about themselves or their child.

RESULTS
Sample characteristics
The recruitment and retention of participants in the Child Health CheckPoint are described elsewhere.[18] Of the 1874 families who participated in CheckPoint assessment centres, blood serum samples of analysable quality from 1180 children and 1325 parents (figure 1) were sent for NMR quantification of metabolites. The majority of excluded families undertook home visits or attended a regional centre, where blood samples could not be collected (n=385, 20.5%), while some participants declined a blood sample (children, n=150, 8.0%; adults, n=108, 5.8%). Few data were lost due to insufficient volume or poor quality samples at the assessment centre (figure 1). The sample characteristics of parents and children are outlined in table 2. Summary statistics for our main child and parent metabolite measures are presented in the online supplementary table 1. Online supplementary figures 3–7 show density plots comparing the distributions of metabolites for boys, girls and adults.

Differences in metabolite levels: adults compared with children
Figure 2 shows mean differences in metabolite levels for adults relative to children in SD units. Most concentrations were higher in adults than in children. Values that were similar in adults and children included total lipids in very large HDL lipoprotein subclass particles, acetoacetate, tyrosine and glucose. Levels in children were higher than those of adults for the majority of glycolysis-related measures (lactate, pyruvate, citrate and glycerol), the ketone body 3-hydroxybutrate, the amino acid glutamine, many fatty acid ratios and all lipoprotein particle sizes. Online supplementary table 2 lists the corresponding estimates in absolute concentration units.

Sex differences in metabolite levels in children and adults
Figure 3 shows differences in mean metabolite levels by sex for children and adults separately in SD units, with estimates in absolute concentration units listed in online supplementary tables 3 and 4.

In general, sex differences were more pronounced in adulthood, resulting in distinct overall patterns for children and adults. Children generally showed smaller differences by sex than adults. Of note, sex differences for apolipoproteins and fatty acid measures showed different patterns in children compared with adults.

Girls had lower levels of apolipoprotein-A-1 (ApoA-1) and higher ApoB than boys. In adults, the opposite pattern was observed with women having higher ApoA-1 and lower ApoB than men. In children, some fatty acid concentrations were higher in girls than boys. In contrast, many adult fatty acid measures were higher in men. There was no evidence of a difference in the level of inflammation (GlycA) by sex in children, whereas in adults, GlycA levels tended to be higher in men than women.

Table 2 Sample characteristics; values are weighted mean (SD)

Characteristic	All	Male	Female
Child			
n	1152–1180	558–575	594–605
Age (years)	12.0 (0.4)	12.0 (0.4)	12.0 (0.4)
BMI (kg/m^2)	19.4 (3.5)	19.2 (3.4)	19.6 (3.7)
BMI z-score	0.38 (1.0)	0.40 (1.0)	0.37 (1.0)
Disadvantage Index	1012 (63)	1011 (65)	1014 (61)
Fasting time (hours)	4.2 (1.2)	4.3 (1.3)	4.2 (1.1)
Time of day - blood collection	14.16 (2.0)	14.12 (2.0)	14.20 (2.1)
Processing lag time (hours)	1.16 (0.5)	1.18 (0.5)	1.14 (0.5)
Parent			
n	1272–1325	174–177	1098–1148
Age (years)	43.9 (5.6)	46.9 (6.9)	43.4 (5.2)
BMI (kg/m^2)	28.4 (6.4)	28.9 (4.7)	28.4 (6.6)
Fasting time (hours)	3.3 (1.6)	3.6 (2.0)	3.2 (1.5)
Time of day – blood collection	13.10 (2.0)	13.18 (2.1)	13.09 (2.0)
Processing lag time (hours)	1.26 (0.5)	1.31 (0.5)	1.26 (0.5)

Values are weighted mean (SD).
Disadvantage Index is the Index of Relative Socioeconomic Disadvantage.
n, number of participants in cohort with this measure.

For some metabolites, sex differences in children mirrored (but were smaller in magnitude than) those of adults, particularly for the ketone body acetate and some key amino acids. At both ages, the amino acid glycine was higher in females, but the branched-chain amino acids leucine and valine tended to be higher in males.

Parent–child concordance

Figure 4 shows the correlations between metabolite measures for all children with all parents and for boys and girls with mothers (but not with the 177 fathers, given the small numbers). The corresponding correlation coefficients and partial correlation coefficients are listed in the online supplementary tables 5 and 6.

Correlations for all parents and all children showed similar patterns to that observed for mother and child by sex. Although there was little suggestion of substantial correlation within parent–child dyads for some metabolites (eg, glucose and acetate), a positive correlation was found for many metabolite measures irrespective of child sex. For example, positive correlations were observed for isoleucine (CC 0.33, 95% CI 0.27 to 0.38), total serum cholesterol (CC 0.30, 95% CI 0.24 to 0.35) and omega 6 fatty acids (CC 0.28, 95% CI 0.23 to 0.34) in parent–child comparisons. Additional adjustment for factors that potentially influence metabolite levels (age, socioeconomic status, fasting time and processing lag time) had little effect on the degree of correlation in any comparison (see the online supplementary tables 5 and 6).

DISCUSSION
Principal findings

Here we present age and sex differences, describing the distribution of detailed/NMR-based metabolite measures in Australian 11–12-year-old children and their parents, and demonstrate that many metabolite measures have moderate parent-child concordance and in general there is a high level of agreement in the magnitude of concordance across metabolites. In accord with previous studies, we observed major differences in metabolite levels between childhood and adulthood and also differences by sex in both childhood and adulthood. We also observed variability in the magnitude of differences by sex for several metabolites in childhood compared with adulthood and identified a complex interplay of correlations of specific metabolites between parents and their children according to parent–child sex relationships.

Strengths and weaknesses

This is the first major cohort study to report both sex and cross-generational differences in metabolomic concentrations in mid-childhood to adulthood utilising the NMR platform. Further strengths include a large number of parent–child dyads representing a wide range of parent ages, the national population-based sample and the state-of-the-art measurements. Replication studies exploring sex differences in earlier and later stages of childhood and adolescence would strengthen findings.
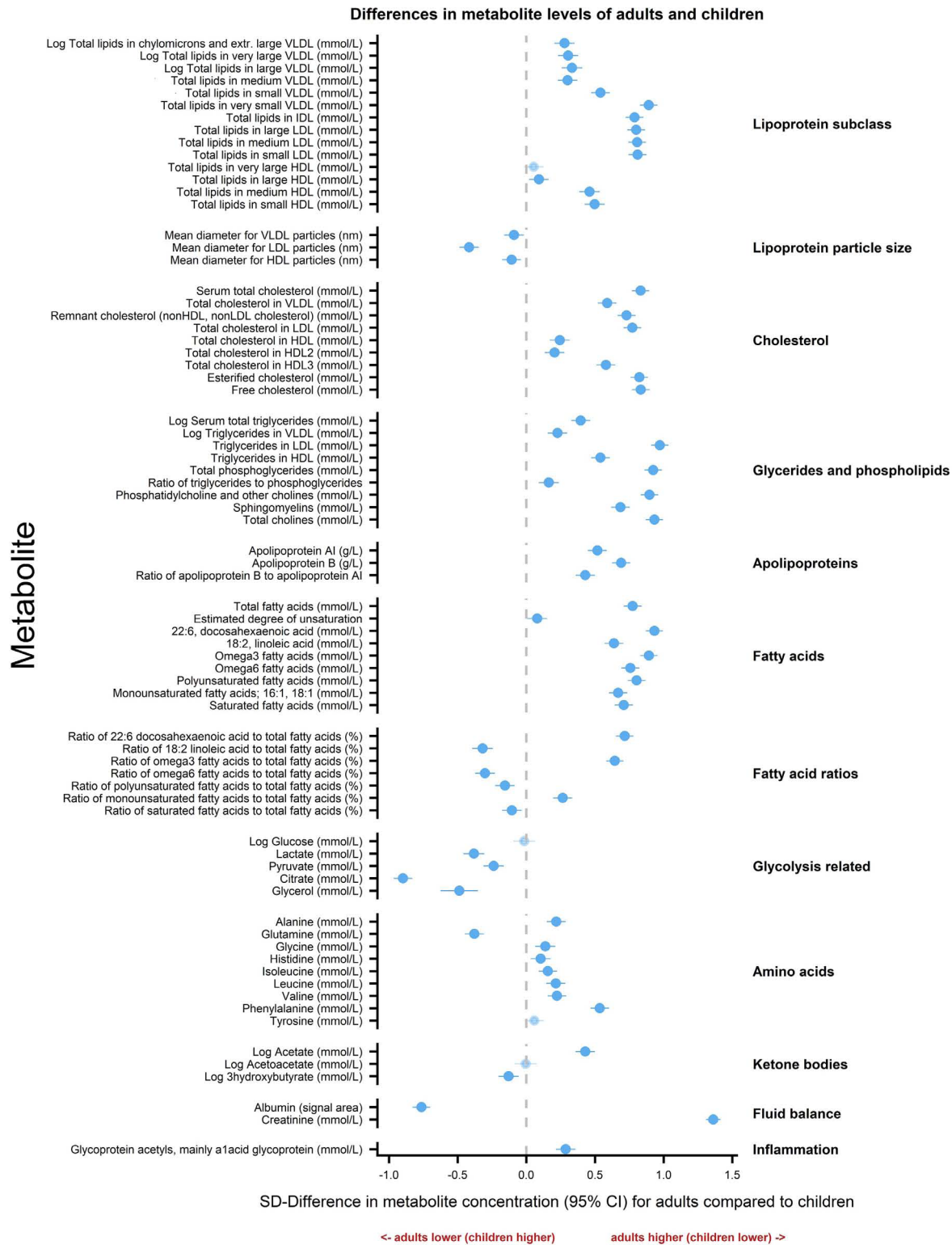
Figure 2 Differences in metabolite levels between children and adults. Association measures are SD difference in metabolite concentration for adults compared with children. Error bars represent 95% CIs. Significant associations after p values adjusted for multiple testing using Benjamini-Hochberg procedure are shown in bold (FDR=0.10). Association measures in absolute concentration units, 95% confidence intervals and associated p-values are listed in the online supplementary table 2. HDL, high-density lipoprotein; IDL, intermediate density lipoprotein; LDL, low-density lipoprotein; VLDL, very low density lipoprotein.

An important limitation is that paternal factors were not fully represented, as most parental samples were from mothers (a well-documented problem in longitudinal cohort studies). This also limited sex-specific parental contribution analysis; further studies including more fathers are warranted. Additional limitations are that, without samples from both parents for each child, we could not estimate heritability, and our results might not apply to
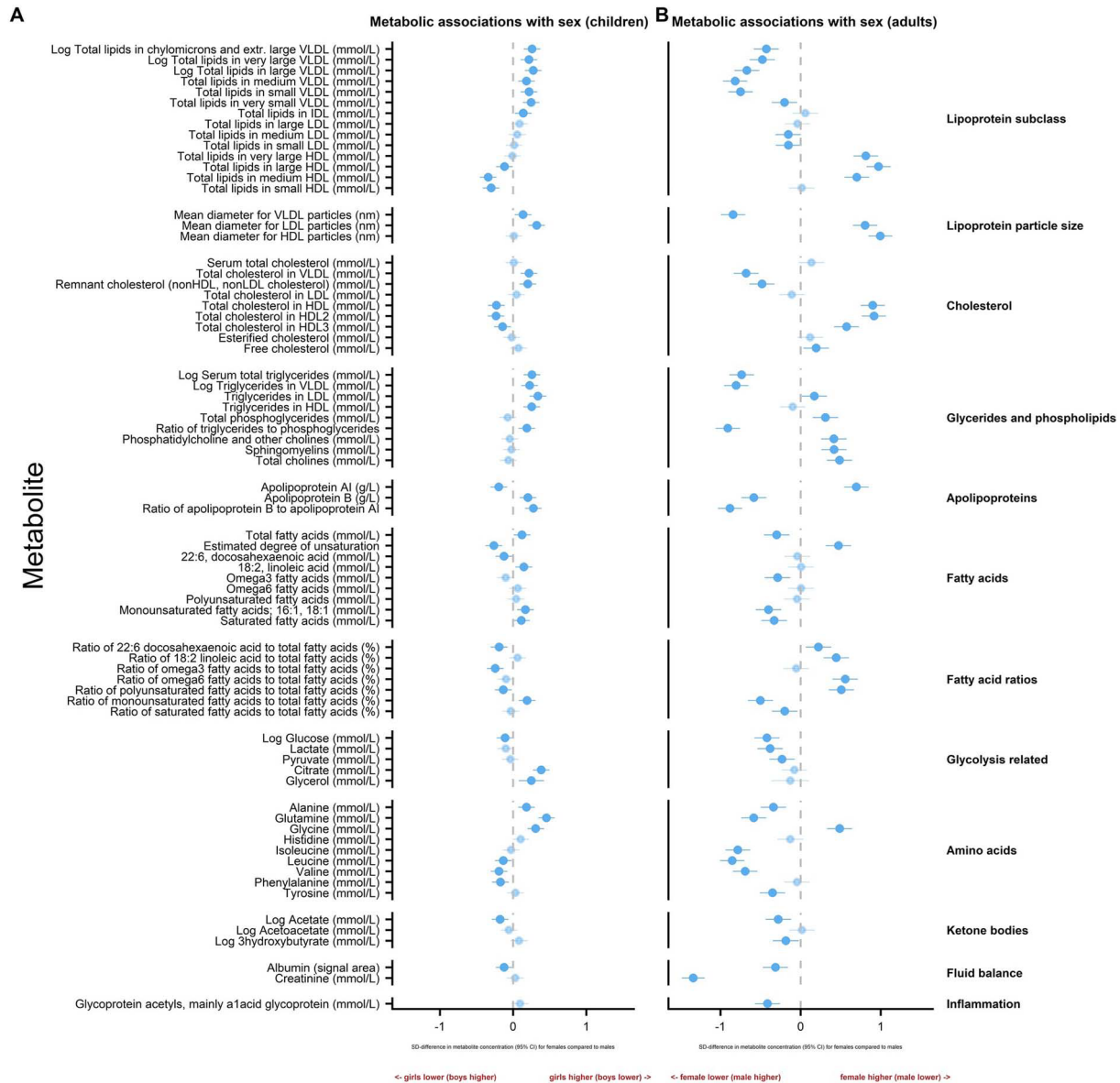
Figure 3 Sex differences in metabolite levels in childhood and adulthood. Association measures are SD difference in metabolite concentration for females compared to males in children (**A**) and adults (**B**). Error bars represent 95% CIs. Significant associations after p values adjusted for multiple testing using Benjamini-Hochberg procedure are shown in bold (FDR=0.10). Association measures in absolute concentration units, 95% CIs and associated p values are listed in the online supplementary tables 3 and 4. HDL, high-density lipoprotein; IDL, intermediate density lipoprotein; LDL, low-density lipoprotein; VLDL, very low-density lipoprotein.

mid-life adults who are not parents (although we see no good reason why these would differ greatly). The original uptake of just over 50% and subsequent attrition within LSAC and then the CheckPoint have led to a relatively advantaged sample, but nonetheless, participants varied widely on key potential confounders (eg, disadvantage and age) and this was at least partly offset by application or consideration of survey weights. Given a large number of metabolites and modest sample size, considerable uncertainty remains in any ranking of the various effects across metabolites. In addition, given the descriptive aims of the paper, additional factors and potential confounders not considered could explain some of the results observed.

Meaning and implications for clinicians and policy-makers

Overall, we found a difference in metabolite profile between children and their parents. This was apparent for specific metabolite measures (such as some amino acids) as well as the distribution of metabolites (such as lipid composition of lipoproteins of different density). Some measures were higher in adults, some similar, while a minority were lower. Previous studies, largely in adults, have identified a range of specific metabolite changes with age, particularly from mid to late adulthood.[27] This includes a general decrease in several amino acid species, which contrasts with our findings from childhood to mid-adulthood.[8] Only the amino acid glutamine showed this pattern in our dataset.
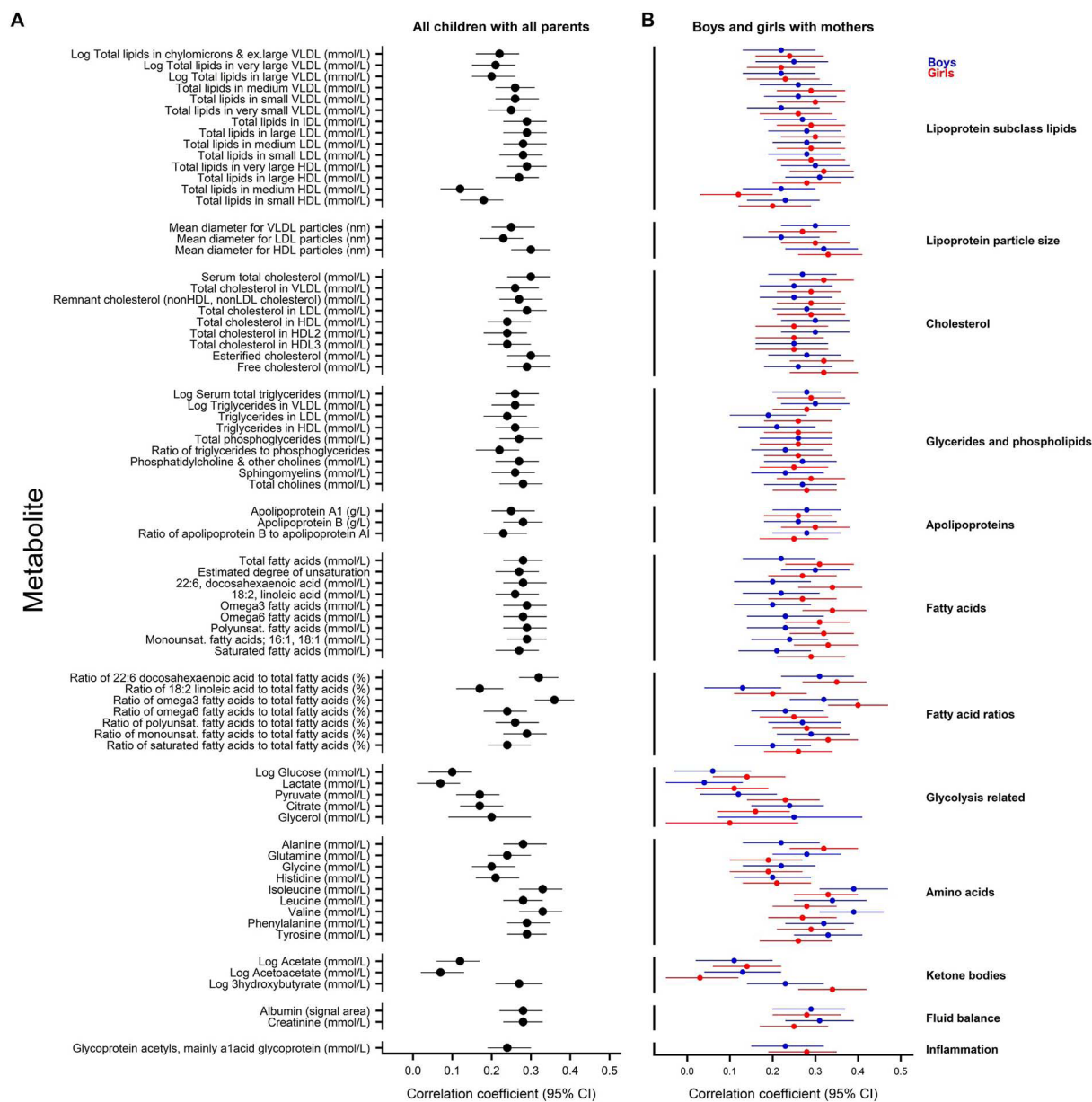
Figure 4 Parent–child correlation for metabolite measures. Pearson's correlation coefficients for all children with all parents (A); and for boys (blue) with mothers and for girls (red) with mothers (B). Error bars represent 95% CIs. Correlation coefficients with associated 95% CIs are listed in the online supplementary tables 5 and 6. HDL, high-density lipoprotein; IDL, intermediate density lipoprotein; LDL, low-density lipoprotein; VLDL, very low-density lipoprotein.

Differences in children by sex (±0.2 SD) were generally much smaller than in adults (±0.8 SD). Large metabolomic studies using alternative platforms have previously reported reproducible, sex-specific signatures in circulating metabolite profile in adults.[28 29] This includes differences in amino acid and lipid serum concentrations, potentially influenced by sex-specific effects of genetic polymorphisms on metabolite levels.[29 30] As in our study, most amino acids have usually been reported to be higher in men than women.[29 31] For example, in a recent study of 507 metabolic markers in 1756 individuals (903 women and 853 men aged ~60 years), one third of metabolites showed significant sexual dimorphism. These were predominantly related to pathways of steroid metabolism,

fatty acids, other lipids and a large proportion of amino acids.[31] Of particular note, branched chain amino acids (BCAAs) and their related metabolic products were among the most differentially represented, with much higher isoleucine, leucine and valine in men. A similar finding of higher leucine and valine was also noted in the Cooperative Health Research in the Region of Augsburg (KORA) follow-ups 3 (F3) and 4 (F4) analysis of >3000 adults,[29] consistent with our observations in adulthood.

In children, we found that sex differences for leucine and valine were smaller but in the same direction as adults. Several lines of evidence implicate BCAA metabolism with metabolic risk in humans. For example, three candidate genes for obesity and/or type 2 diabetes

mellitus (T2DM) are involved in the BCAA metabolic pathway.[32] In a recent large meta-analysis of metabolomics in diabetes, a >30% higher risk of T2D was found per SD increase in isoleucine, leucine, valine or tyrosine, whereas glycine and glutamine were inversely associated with risk.[32] Several clinical studies have also reported that BCAAs positively correlate with insulin resistance, homeostatic model assessment index and levels of haemoglobin A1c, while longitudinal studies have reported that increased blood BCAAs are predictive of future insulin resistance and T2D.[33] It is intriguing to speculate that the higher BCAA in males from early life could contribute to the well-described increasing prevalence of T2D in men. Levels of BCAA are elevated in women with polycystic ovary syndrome, potentially contributing to the associated insulin resistance.[34] However, it remains unclear whether BCAAs are on the causal pathway to T2D or result from adverse metabolic health. Our demonstration that the sex differences in BCAA possibly arise early in life offers potential to track their association with sex-specific measures of metabolic health from an early age to help clarify where they lie on the causal pathway.

In accord with previous adult studies,[29] we found higher levels of glycine in mothers than fathers, and (less markedly) in girls than boys. Interestingly, recent metabolomics and genetic analyses of ~10 000 adults with cardiovascular disease (CVD), with replication in >53 000 subjects, identified a genetic variant in carbamoyl-phosphate synthetase 1 (CPS1) (linked to plasma glycine levels) to be strongly associated with a reduced risk of CVD in women ($p=6.3 \times 10^{-5}$) but not men ($p=0.95$), suggesting a direct link between glycine levels and CVD risk, although whether this is a causal association remains unclear.[35] It will be interesting in the future to explore the link between variants in CPS1 and circulating glycine levels from early life to adulthood in relation to markers of cardiovascular health in females.

The small sex differences of HDL cholesterol and ApoA-1 in children compared with adults is consistent with modest differences in children, whereas substantial differences in adulthood have previously been reported.[36] ApoA-1 was more abundant in boys, whereas ApoB was higher in girls, leading to a higher ApoB/ApoA-1 ratio in girls. The opposite pattern was found in our limited sample of fathers relative to mothers. These data are surprising and differ from a similarly sized study of slightly older European adolescent children (mean age 15 years) that found higher ApoA-1 and ApoB in girls relative to boys.[37] Interestingly, a higher ApoB/ApoA-1 ratio has been strongly linked to increased coronary risk in adults,[38–40] suggesting that sex differences may alter with increasing age, in keeping with the increased CVD risk in adult men. ApoA-1 is the main protein component of HDL cholesterol[41] thus the differences in trajectories in lipids and HDL cholesterol for boys and girls across childhood that have been reported[42 43] could partially explain this observation.

These are the first data on the mother–child or parent–child correlations of NMR metabolites. Smaller studies have reported positive correlations between parents and children for a limited range of cardiometabolic risk factors including total cholesterol, LDL cholesterol, HDL cholesterol and triglycerides measured using conventional methods. We found positive correlations between parents and children for the same lipid measures (although measured using NMR) consistent with previously reported findings. One study reported a positive association between the serum lipid levels of 4-year-old children (n=127) and their parents (122 mothers and 118 fathers)[44] whereas another study of children aged 6–18 years (n=255) and their parents (n=179) found that the age of the child influenced the degree of correlation of several lipid measures, with older (10–18 years) children more similar to their parents in terms of triglyceride levels than younger individuals (6–9 years).[12]

Unanswered questions and future research
The temporal and sex-specific dynamism of the metabolomics data we describe here offer considerable opportunities for identification of biomarkers of risk for a range of non-communicable diseases early in life to inform targeted interventions and monitor their efficacy. Combining metabolomics with other 'omics data (such as genetics), as is increasingly reported from large adult studies, offers considerable promise in understanding the causal pathways that link early life exposures, genetics and intermediate phenotypes with later onset chronic disease and in identifying clinically relevant biomarkers.

In conclusion, we describe the metabolite profile from mid-childhood and adulthood in a population-based sample, together with parent–child concordance and differences by sex in children and adults. In this descriptive paper, distinct differences in profiles were observed by age and sex, as well as considerable evidence of a correlation between parent and child measures. These data will be informative for investigation of the childhood origins of adult non-communicable diseases and for comparative studies across populations.

Author affiliations
[1]Murdoch Children's Research Institute, Parkville, Victoria, Australia
[2]Department of Paediatrics and The Liggins Institute, University of Auckland, Auckland, New Zealand
[3]Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia
[4]Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland
[5]Nightingale Health Ltd., Helsinki, Finland
[6]Department of Medicine, University of Turku, Turku, Finland
[7]Division of Medicine, Turku University Hospital, Turku, Finland
[8]The George Institute for Global Health, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, United Kingdom
[9]Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia
[10]Department of Paediatrics, Monash University, Clayton, Victoria, Australia

REFERENCES

1. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999;29:1181–9.
2. Nath AP, Ritchie SC, Byars SG, *et al*. An interaction map of circulating metabolites, immune gene networks, and their genetic regulation. *Genome Biol* 2017;18:146.
3. Shah SH, Newgard CB. Integrated metabolomics and genomics: systems approaches to biomarkers and mechanisms of cardiovascular disease. *Circ Cardiovasc Genet* 2015;8:410–9.
4. Kaikkonen JE, Würtz P, Suomela E, *et al*. Metabolic profiling of fatty liver in young and middle-aged adults: Cross-sectional and prospective analyses of the Young Finns Study. *Hepatology* 2017;65:491–500.
5. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 2016;17:451–9.
6. Kettunen J, Tukiainen T, Sarin AP, *et al*. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet* 2012;44:269–76.
7. Suhre K, Gieger C. Genetic variation in metabolic phenotypes: study designs and applications. *Nat Rev Genet* 2012;13:759–69.
8. Yu Z, Zhai G, Singmann P, *et al*. Human serum metabolic profiles are age dependent. *Aging Cell* 2012;11:960–7.
9. Auro K, Joensuu A, Fischer K, *et al*. A metabolic view on menopause and ageing. *Nat Commun* 2014;5:4708.
10. Saito K, Maekawa K, Kinchen JM, *et al*. Gender- and Age-Associated Differences in Serum Metabolite Profiles among Japanese Populations. *Biol Pharm Bull* 2016;39:1179–86.
11. Wang Q, Würtz P, Auro K, *et al*. Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. *BMC Med* 2016;14:205.
12. Halvorsen T, Moran A, Jacobs DR, *et al*. Relation of Cardiometabolic Risk Factors between Parents and Children. *J Pediatr* 2015;167:1049–56.
13. Rueedi R, Ledda M, Nicholls AW, *et al*. Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. *PLoS Genet* 2014;10:e1004132.
14. Kettunen J, Demirkan A, Würtz P, *et al*. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016;7:11122.
15. Edwards B. Growing Up in Australia: The Longitudinal Study of Australian Children: Entering adolescence and becoming a young adult. *Family Matters* 2014;95:5–14.
16. Sanson A, Johnstone R. The LSAC Research Consortium & FaCS LSAC Project Team. Growing Up in Australia takes its first steps. *Family Matters* 2004;67:46–53.
17. Wake M, Clifford SA, York E, *et al*. Introducing Growing Up in Australia's Child Health CheckPoint. *Family Matters* 2014;95:15–23.
18. Clifford SA, Davies S, Wake M, *et al*. Child Health CheckPoint: cohort summary and methodology of a physical health and biospecimen module for the Longitudinal Study of Australian Children. *BMJ Open* 2019;9(suppl 3):3–22.
19. Soininen P, Kangas AJ, Würtz P, *et al*. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* 2009;134:1781–5.
20. Würtz P, Kangas AJ, Soininen P, *et al*. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol* 2017;186:1084–96.
21. Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.
22. Australian Bureau of Statistics. Census of population and housing: Socio-Economic Indexes for Areas (SEIFA) 2011. 2011. Cat. no. 2033.0.55.001.
23. Davies S, Clifford SA, Gillespie AN, *et al*. *Longitudinal Study of Australian Children's Child Health CheckPoint Data Issues Paper -December 2018*. Melbourne: Murdoch Children's Research Institute, 2018.
24. Heeringa SG, West BT, Berglund PA. *Applied survey data analysis*. Boca Raton: CRC press, 2010.
25. Ellul S, Hiscock R, Mensah FK, *et al*. *Longitudinal Study of Australian Children's Child Health CheckPoint Technical Paper 1: Weighting and non-response*. Melbourne: Murdoch Children's Research Institute, 2018.
26. *R: A language and environment for statistical computing [program]*. Vienna, Austria: R Foundation for Statistical Computing, 2018.
27. Menni C, Kastenmüller G, Petersen AK, *et al*. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* 2013;42:1111–9.
28. Dunn WB, Lin W, Broadhurst D, *et al*. Molecular phenotyping of a UK population: defining the human serum metabolome. *Metabolomics* 2015;11:9–26.
29. Mittelstrass K, Ried JS, Yu Z, *et al*. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet* 2011;7:e1002215.
30. Kolz M, Johnson T, Sanna S, *et al*. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet* 2009;5:e1000504.
31. Krumsiek J, Mittelstrass K, Do KT, *et al*. Gender-specific pathway differences in the human serum metabolome. *Metabolomics* 2015;11:1815–33.
32. Guasch-Ferré M, Hruby A, Toledo E, *et al*. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care* 2016;39:833–46.

33. Lynch CJ, Adams SH. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat Rev Endocrinol* 2014;10:723–36.

34. Chang AY, Lalia AZ, Jenkins GD, *et al*. Combining a nontargeted and targeted metabolomics approach to identify metabolic pathways significantly altered in polycystic ovary syndrome. *Metabolism* 2017;71(Supplement C):52–63.

35. Hartiala JA, Tang WH, Wang Z, *et al*. Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat Commun* 2016;7:10558.

36. Davis CE, Williams DH, Oganov RG, *et al*. Sex difference in high density lipoprotein cholesterol in six countries. *Am J Epidemiol* 1996;143:1100–6.

37. Spinneker A, Egert S, González-Gross M, *et al*. Lipid, lipoprotein and apolipoprotein profiles in European adolescents and its associations with gender, biological maturity and body fat–the HELENA Study. *Eur J Clin Nutr* 2012;66:727–35.

38. Walldius G, Jungner I, Aastveit AH, *et al*. The apoB/apoA-I ratio is better than the cholesterol ratios to estimate the balance between plasma proatherogenic and antiatherogenic lipoproteins and to predict coronary risk. *Clin Chem Lab Med* 2004;42:1355–63.

39. Walldius G, Jungner I. The apoB/apoA-I ratio: a strong, new risk factor for cardiovascular disease and a target for lipid-lowering therapy–a review of the evidence. *J Intern Med* 2006;259:493–519.

40. Mozaffarian D, Benjamin EJ, Go AS, *et al*. Heart disease and stroke statistics–2015 update: a report from the American Heart Association. *Circulation* 2015;131:e29–322.

41. Upadhyay RK. Emerging risk biomarkers in cardiovascular diseases and disorders. *J Lipids* 2015;2015:1–50.

42. Hardy R, Lawlor DA, Kuh D. A life course approach to cardiovascular aging. *Future Cardiol* 2015;11:101–13.

43. Jolliffe CJ, Janssen I. Distribution of lipoproteins by age and gender in adolescents. *Circulation* 2006;114:1056–62.

44. Öhlund I, Hernell O, Hörnell A, *et al*. Serum lipid and apolipoprotein levels in 4-year-old children are associated with parental levels and track over time. *Eur J Clin Nutr* 2011;65:463–9.