

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Ojanen, Xiaowei; Cheng, Runtan; Törmäkangas, Timo; Rappaport, Noa; Wilmanski, Tomasz; Wu, Na; Fung, Erik; Nedelec, Rozenn; Sebert, Sylvain; Vlachopoulos, Dimitris; Yan, Wei; Price, Nathan D.; Cheng, Sulin; Wiklund, Petri

**Title:** Towards early risk biomarkers : serum metabolic signature in childhood predicts cardio-metabolic risk in adulthood

**Year:** 2021

**Version:** Published version

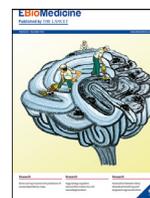
**Copyright:** © 2021 The Authors. Published by Elsevier B.V.

**Rights:** CC BY-NC-ND 4.0

**Rights url:** <https://creativecommons.org/licenses/by-nc-nd/4.0/>

**Please cite the original version:**

Ojanen, X., Cheng, R., Törmäkangas, T., Rappaport, N., Wilmanski, T., Wu, N., Fung, E., Nedelec, R., Sebert, S., Vlachopoulos, D., Yan, W., Price, N. D., Cheng, S., & Wiklund, P. (2021). Towards early risk biomarkers : serum metabolic signature in childhood predicts cardio-metabolic risk in adulthood. *EBioMedicine*, 72, Article 103611. <https://doi.org/10.1016/j.ebiom.2021.103611>



## Research paper

## Towards early risk biomarkers: serum metabolic signature in childhood predicts cardio-metabolic risk in adulthood



Xiaowei Ojanen, Ph.D<sup>1,2,3,\*\*</sup>, Runtan Cheng, MSc<sup>1,2,\*\*</sup>, Timo Törmäkangas, Ph.D<sup>3</sup>, Noa Rappaport, Ph.D<sup>4</sup>, Tomasz Wilmanski, Ph.D<sup>4</sup>, Na Wu, MSc<sup>1</sup>, Erik Fung, M.B.Ch.B., Ph.D<sup>5,6,7</sup>, Rozenn Nedelec, MSc<sup>8</sup>, Sylvain Sebert, PhD<sup>8</sup>, Dimitris Vlachopoulos, Ph.D<sup>9</sup>, Wei Yan, Ph.D<sup>1</sup>, Nathan D. Price, Ph.D<sup>4</sup>, Sulin Cheng, Ph.D<sup>1,2,3,10,\*</sup>, Petri Wiklund, Ph.D<sup>2,3,\*</sup>

<sup>1</sup> Key Laboratory of Systems Biomedicine (Ministry of Education), School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup> Exercise Translational Medicine Center, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China

<sup>3</sup> Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

<sup>4</sup> Institute for Systems Biology, Seattle, WA, USA

<sup>5</sup> CARE Programme, Lui Che Woo Institute of Innovative Medicine, CUHK Hong Kong Hub of Paediatric Excellence, Hong Kong Children's Hospital, Hong Kong SAR, China

<sup>6</sup> Laboratory for Heart Failure + Circulation Research, Li Ka Shing Institute of Health Sciences; Gerald Choa Cardiac Research Centre, Department of Medicine and Therapeutics; Centre for Cardiovascular Genomics and Medicine, Faculty of Medicine, The Chinese University of Hong Kong and Prince of Wales Hospital, Hong Kong SAR, China

<sup>7</sup> Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom

<sup>8</sup> Centre for Life-Course Health Research, Faculty of Medicine, University of Oulu, Finland

<sup>9</sup> Children's Health and Exercise Research Centre, Sport and Health Sciences, University of Exeter, Exeter, United Kingdom

<sup>10</sup> Department of Physical Education, Shanghai Jiao Tong University, Shanghai, China

## ARTICLE INFO

## Article History:

Received 20 February 2021

Revised 13 September 2021

Accepted 17 September 2021

Available online xxx

## Key words:

metabolomics  
cardio-metabolic risk  
children  
longitudinal-study  
ALSPAC

## SUMMARY

**Background:** Cardiovascular diseases may originate in childhood. Biomarkers identifying individuals with increased risk for disease are needed to support early detection and to optimise prevention strategies.

**Methods:** In this prospective study, by applying a machine learning to high throughput NMR-based metabolomics data, we identified circulating childhood metabolic predictors of adult cardiovascular disease risk (MetS score) in a cohort of 396 females, followed from childhood (mean age 11.2 years) to early adulthood (mean age 18.1 years). The results obtained from the discovery cohort were validated in a large longitudinal birth cohort of females and males followed from puberty to adulthood (n = 2664) and in four cross-sectional data sets (n = 6341).

**Findings:** The identified childhood metabolic signature included three circulating biomarkers, glycoprotein acetyls (GlycA), large high-density lipoprotein phospholipids (L-HDL-PL), and the ratio of apolipoprotein B to apolipoprotein A-1 (ApoB/ApoA) that were associated with increased cardio-metabolic risk in early adulthood (AUC = 0.641–0.802, all p < 0.01). These associations were confirmed in all validation cohorts with similar effect estimates both in females (AUC = 0.667–0.905, all p < 0.01) and males (AUC = 0.734–0.889, all p < 0.01) as well as in elderly patients with and without type 2 diabetes (AUC = 0.517–0.700, all p < 0.01). We subsequently applied random intercept cross-lagged panel model analysis, which suggested bidirectional causal relationship between metabolic biomarkers and cardio-metabolic risk score from childhood to early adulthood.

**Interpretation:** These results provide evidence for the utility of a circulating metabolomics panel to identify children and adolescents at risk for future cardiovascular disease, to whom preventive measures and follow-up could be indicated.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

\* Corresponding author and requests for reprints to: Dr. Petri Wiklund and Dr. Sulin Cheng

E-mail addresses: [sulin.cheng@jyu.fi](mailto:sulin.cheng@jyu.fi) (S. Cheng), [petriwiklund@gmail.com](mailto:petriwiklund@gmail.com) (P. Wiklund).

\*\* Xiaowei Ojanen and Runtan Cheng contributed equally to this manuscript.

## Introduction

Cardiovascular diseases are the largest contributors to global mortality and morbidity, [1] and a significant economic burden to the healthcare system. Cardio-metabolic abnormalities, including

<https://doi.org/10.1016/j.ebiom.2021.103611>

2352-3964/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Research in context

### Evidence before this study

Earlier studies have identified a number of circulating biomarkers of cardiovascular disease in adults. However, it is unclear whether in children these biomarkers can predict future cardiovascular disease risk. We performed a literature search in PubMed (cut-off date: 5<sup>th</sup> of November, 2020). Using the search term ["metabolite profiling" AND "cardiovascular risk" AND "children"] with no other limits applied to publication dates. We found 46 publications, and none of these studies examined the associations of childhood metabolite profiles with cardiovascular risk in adulthood.

### Added value of this study

In this prospective study we show that previously identified cardiovascular disease biomarkers (GlycA, L-HDL-PL, ApoB to Apo A-1 ratio) are significant predictors of future cardio-metabolic health already in childhood and adolescence. These findings suggest an opportunity for early intervention and monitoring of cardiovascular disease risk in children.

### Implications of all the available evidence

These results indicate that the pathogenic processes that contribute to cardiovascular diseases originate in childhood. Through the use of high-throughput serum metabolomics, children at increased risk of cardiovascular disease can be identified and the efficacy of interventions monitored effectively.

## Methods

### Study design and participants

This is a prospective study. The study was conducted in the city of Jyväskylä, and the surrounding area in Central Finland. The profile of participants in this study is summarized in **Figure 1**. A total of 396 girls (mean age 11.2 years at baseline) participated in a longitudinal study at different time points for an average of 7.5 years. Detailed meta-data regarding the participants and study design have been reported previously [9,10] and in **Supplementary file 1 and Supplementary Tables 1**.

For validation, we used a large longitudinal cohort and four cross-sectional datasets (**Figure 1**). Detailed information about the validation cohorts are given in **Supplementary Table 2a-d**.

Written informed consent was obtained from all participants and their parents. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethic Committee of the Central Hospital of Central Finland and the Finnish National Agency of Medicines (memo 22/8/2008 and 5/2009).

### Background and Assessments

Detailed information regarding background of participants, anthropometry and body composition assessments, circulating biomarker and NMR metabolomics assessments, cardio-metabolic risk assessment and statistical analysis are given in **Supplementary file 1**. Briefly, serum metabolite concentrations were analysed using a high-throughput nuclear magnetic resonance (NMR) metabolomics platform. The experimental protocols and applications of the NMR metabolomics platform have been described in detail elsewhere [11]. To assess cardio-metabolic risk, a standardised continuously distributed variable for clustered metabolic risk (MetS score) was constructed similar to previously published scores [12]. The cardio-metabolic risk score (MetS score) was calculated by standardising and then summing the following continuously distributed metabolic traits: mean arterial pressure ( $[(2 \times \text{diastolic blood pressure}) + \text{systolic blood pressure}]/3$ ); abdominal fat mass; fasting plasma glucose; serum HDL cholesterol  $\times -1$ ; and fasting serum triglyceride z-score. The z-scores for each variable and MetS scores were calculated separately for each time point. A higher score indicates a higher cardio-metabolic risk. In addition, a Youden index was calculated to identify actual cut-off points for these identified metabolites (see **Supplementary file 1**).

### Role of funding source

The funding sources did not have any role in study design, data collection, data analyses, interpretation, or writing of this report.

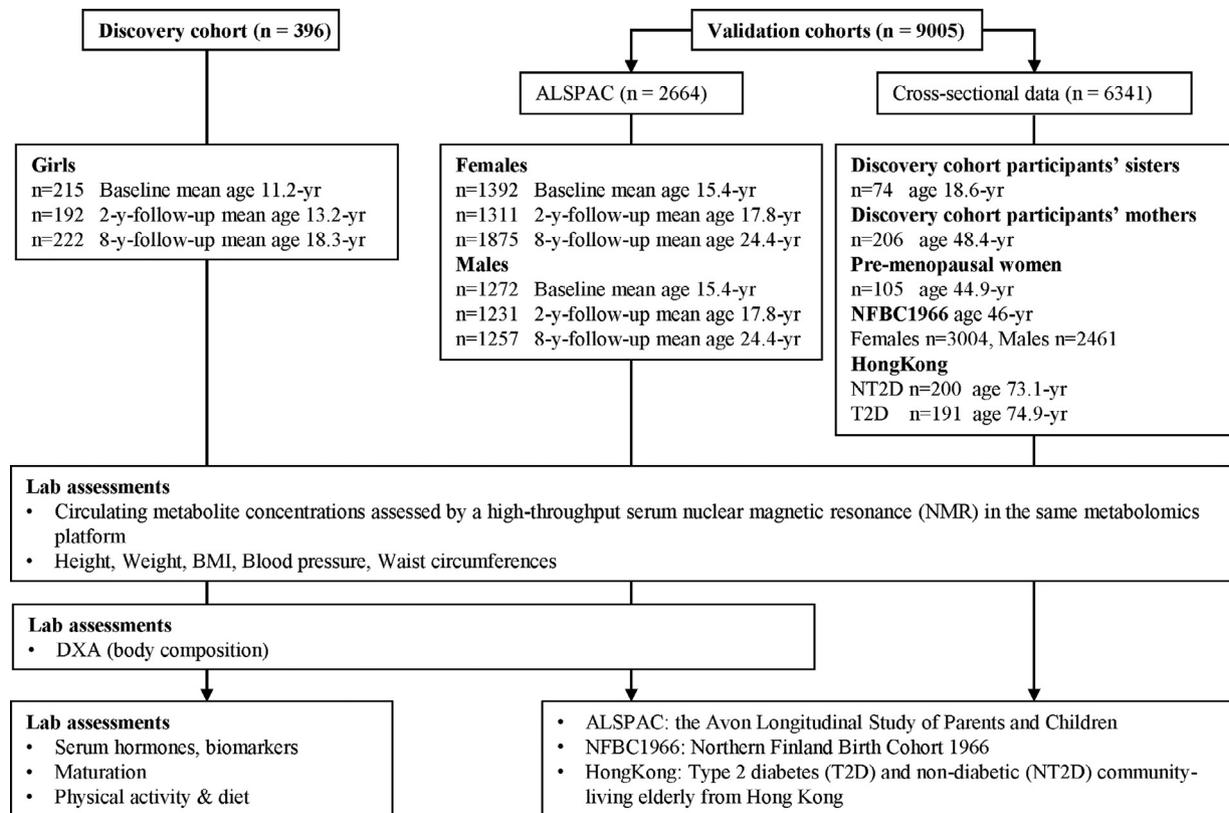
## Results

### Cardio-metabolic risk score

We first examined cardio-metabolic risk factor clustering, i.e. metabolic syndrome, defined as the presence of at least three of the following five risk factors: abdominal obesity (waist circumference  $>88$  cm), elevated blood pressure ( $\geq 130/85$  mmHg), elevated fasting blood glucose ( $\geq 5.6$  mmol/L), elevated serum triglycerides ( $\geq 1.7$  mmol/L) and reduced serum high-density lipoprotein cholesterol (HDL-C) ( $<1.29$  mmol/L) [13] in a total of 396 Finnish adolescent females that participated in a longitudinal study, from pre-puberty (mean age, 11.2 years) to early adulthood (mean age, 18.1 years). At the 7.5 years follow-up, 3.2% of the participants were classified as having metabolic syndrome according to this definition. In addition,

hyperglycaemia, elevated blood pressure, dyslipidemia and abdominal obesity are risk factors for the development of cardiovascular diseases [2]. Although the clinical complications of cardiovascular disease typically manifest in adulthood, autopsy and observational studies have shown that development of atherosclerosis starts in childhood and adolescence, and is associated with the same cardiovascular disease risk factors that are well established in adults [3,4]. Early identification of children who are at risk of developing cardiovascular disease would allow instituting and maintaining optimum health behaviours, at a time when it is likely to be most effective.

The availability of metabolic screens provides an opportunity to identify biomarkers associated with cardio-metabolic risk. For example, using liquid chromatography/mass spectrometry, Cheng et al. [8] demonstrated that obesity, hypertension, insulin resistance and dyslipidaemia were associated with multiple circulating metabolites including branched-chain and aromatic amino acids in adults. These same metabolites have also been consistently associated with future development of Type 2 Diabetes (T2D) [5] and cardiovascular diseases [6]. Metabolomics profiling studies in children and adolescents, however, have reported conflicting results; most of these previous studies are cross-sectional, and the few existing longitudinal studies have short follow-up durations and a small number of participants [7,8]. Therefore, longitudinal studies examining temporal associations between the circulating metabolome and cardio-metabolic risk factors from childhood to adulthood are needed. In this study, we used an NMR-based metabolomics platform to quantify 121 circulating metabolic measures in children followed longitudinally from pre-puberty to early adulthood. The same platform was also used for all validation cohorts. Our results show that a small pre-pubertal metabolic signature predicts cardio-metabolic risk score in adulthood, and provides evidence for a causal role of atherogenic lipoprotein particles and systemic low-grade inflammation in cardiovascular disease pathogenesis.



**Figure 1.** Data profile of discovery and validation cohorts.

10.8% of the participants had two risk factors, 40.1% had one risk factor, and 45.9% had no risk factors.

#### Cross-sectional correlation network of circulating metabolites with cardio-metabolic risk score

We next explored the cross-sectional associations of our constructed continuous variable for clustered metabolic risk (MetS score) with serum metabolites at each measurement wave (baseline, 2-year and 7.5-year follow-ups) to assess the stability of associations during the course of pubertal growth. The significant cross-sectional correlations among metabolites (outer cycle) and between metabolites and MetS score (inner cycle) are presented in **Figure 2** and **Supplementary Figure 1a and b**, and **Supplementary Table 3 and Table 4**. Several lipoprotein subclasses, high-density lipoprotein diameter (HDL-D), HDL-2C and VLDL-TG, apolipoprotein B to apolipoprotein A1 ratio (ApoB/ApoA1) and glycoprotein acetyls (GlycA) were consistently associated with MetS score at all three time points after correcting for multiple hypothesis testing. In addition, we found associations of apolipoprotein A1 and omega 6 fatty acids with the MetS score at baseline and 2-year follow-up, very-low-density lipoprotein (VLDL-D) at baseline and 7.5-year follow-up, fatty acid length and triglycerides/phosphoglycerides ratio (TG/PG) at 2-year follow-up, and apolipoprotein B isoleucine, monounsaturated fatty acids (MUFA) and TG/PG at 7.5-year follow-up. Further adjustment for physical activity, dietary intake and other covariates (i.e., sex hormones, SHBG and IGF-1, insulin, adiponectin, leptin, PTH and vitamin D) did not change the results (**Supplementary Table 4**).

#### Serum metabolites predict future MetS score

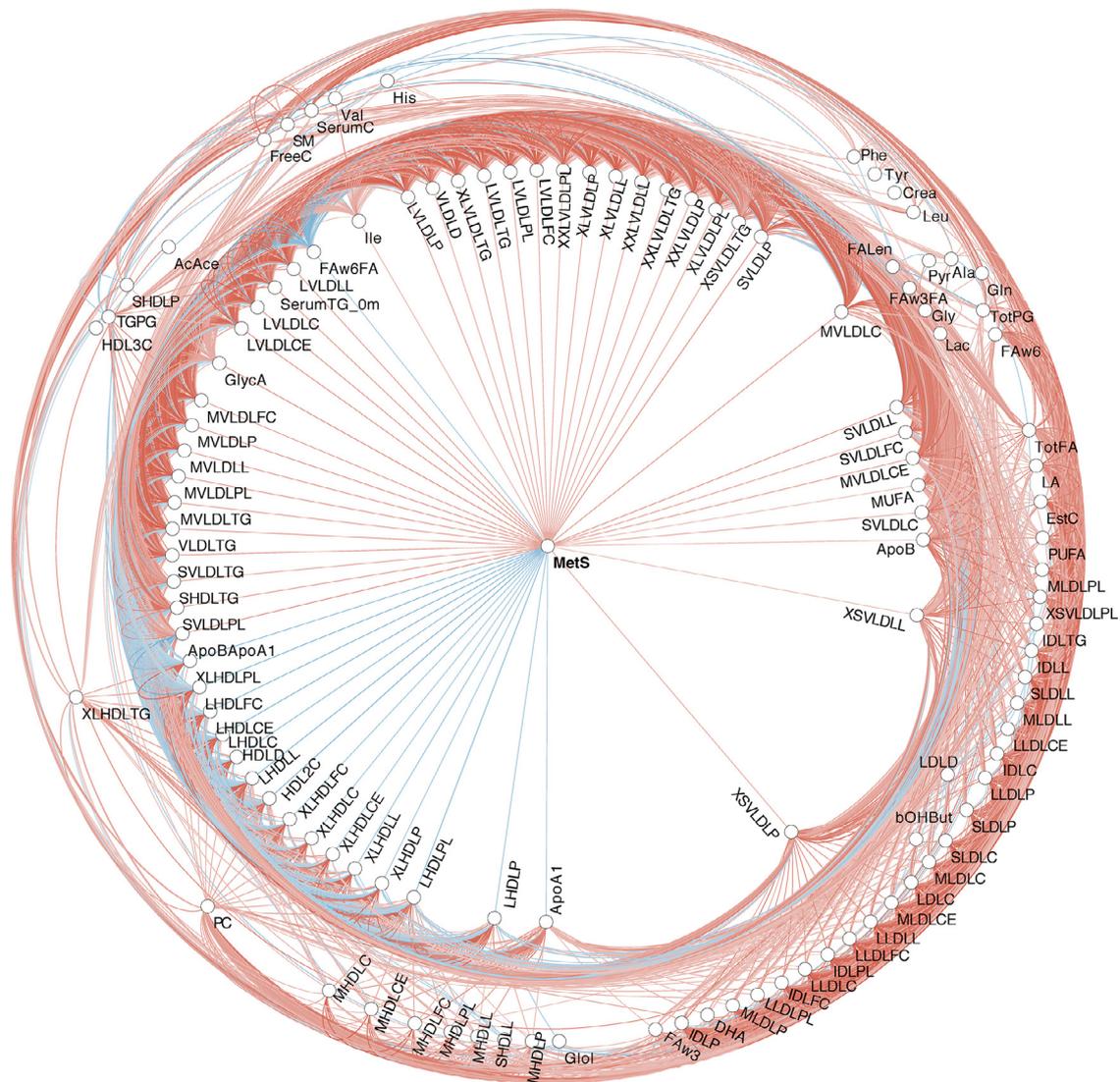
We used Least Absolute Shrinkage and Selection Operator (LASSO) method to identify circulating metabolites that predict the MetS score from pre-puberty to early adulthood using a five-fold cross-validation scheme (**Supplementary File 2**). Of the 121 metabolites measured for each study participant, ten metabolites at baseline and 11

metabolites at 2-year follow-up were retained in the final model. These subsets of metabolites were predictive of MetS score at 7.5 year follow-up, explaining on average 36.3% ( $r^2$  ranged from 0.10 to 0.62) and 32.7% ( $r^2$  ranged from 0.09 to 0.75) of the variance, respectively (**Figure 3a,b,d,e**, **Supplementary File 2** and **Supplementary Table 5**). The identified metabolites included different amino acids, glycolysis and inflammation related metabolites, ketone bodies, fatty acids, apolipoproteins, and lipoprotein subclass particles. Four metabolites (glutamine, L-HDL-PL, ApoB/ApoA1 ratio, GlycA) were retained in the model at both time points, showing greater associations with increasing age (**Supplementary Table 5**).

We next performed a regression analysis with MetS score as the dependent variable and metabolic biomarkers identified by LASSO as independent variables. After Bonferroni correction for multiple testing, only ApoB/ApoA ratio, GlycA and L-HDL-PL remained significant predictors of MetS score ( $p < 0.0001$  for all, **Supplementary Table 6**). These associations were also robust to multi-covariate adjustment, including insulin, leptin, adiponectin, sex steroids, IGF-1, physical activity and energy yield nutrient intakes. We found that baseline ApoB/ApoA ratio and GlycA positively while L-HDL-PL negatively predicted 7.5-year MetS ( $r = 0.471$  and  $p = 2.66 \times 10^{-5}$ ;  $r = 0.400$  and  $p = 0.0005$ ; and  $r = -0.465$  and  $p = 3.47 \times 10^{-5}$ , respectively,  $p$ : adjusted for multiple comparisons by Bonferroni). And 2-year ApoB/ApoA ratio and GlycA positively and L-HDL-PL negatively predicted 7.5-year MetS ( $r = 0.449$  and  $p = 6.75 \times 10^{-5}$ ;  $r = 0.440$  and  $p = 9.93 \times 10^{-5}$ ; and  $r = -0.445$  and  $p = 8.12 \times 10^{-5}$ , respectively,  $p$ : adjusted for multiple comparisons by Bonferroni).

#### Performance of cardio-metabolic risk prediction

To test the ability of the metabolic biomarkers to distinguish individuals with high cardio-metabolic risk from those with low risk, we stratified the study participants into quartiles based on their MetS score values at 7.5-year follow-up and performed Area Under the Receiver Operating Characteristics (AUROC) analysis. The highest

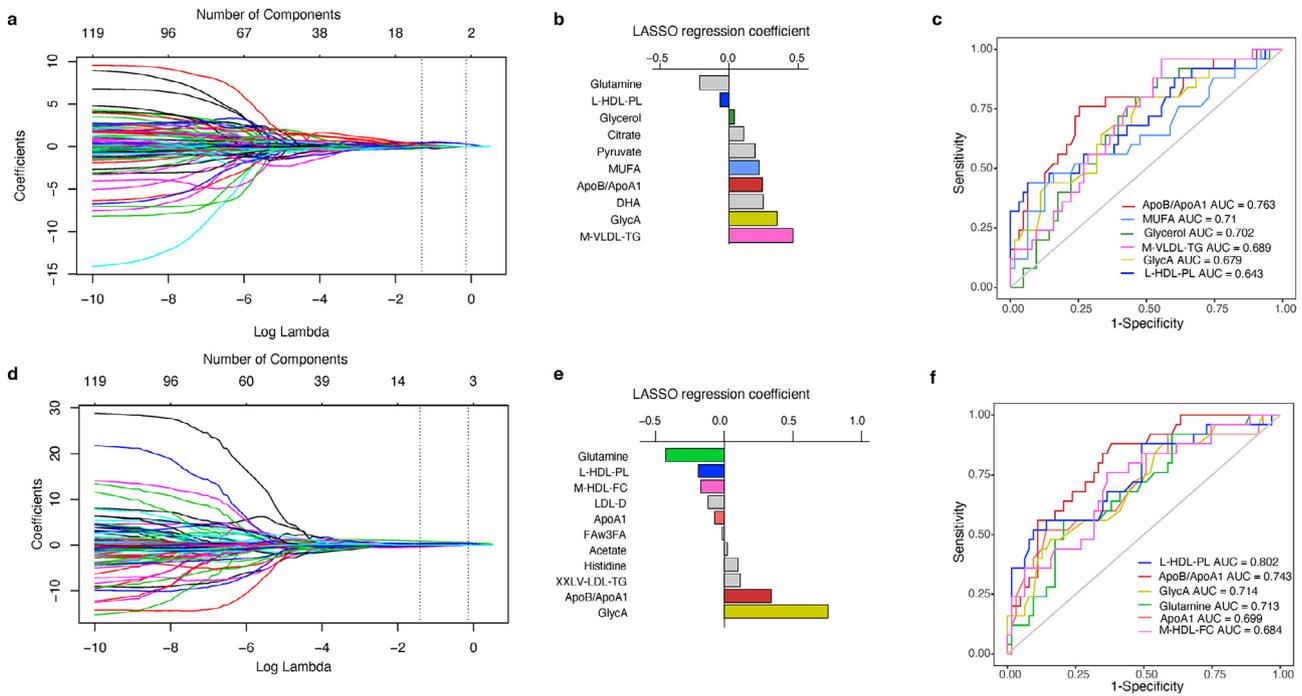


**Figure 2.** Edge bundling by R ‘geom\_conn\_bundle’ for bivariate Spearman correlation network among metabolites and between metabolites and MetS at baseline. Each node represents a variable (such as GlycA or MetS) while the edges linking variable pairs (such as Phe and Val) represent significant correlations between the corresponding nodes. Significant correlations between metabolites and MetS are shown in the inner cycle. There are 118 nodes and 2483 edges. The significant association among metabolites are shown in the outer cycle.

quartile was considered as the high-risk group and the other three quartiles as low-risk. The area under the curve (AUC) showed that baseline levels of ApoB/ApoA1, GlycA and L-HDL-PL were able to classify individuals with high MetS score both at baseline (AUC: 0.643 to 0.763,  $p = 0.001$  to  $p < 0.0001$ , ROC asymptotic significant, **Figure 3c**), and at 2-year follow-up (AUC: 0.684 to 0.802,  $p = 0.03$  to  $p < 0.0001$ , ROC asymptotic significant, **Figure 3f**). For the purpose of comparison, we also tested the ability of body mass index (BMI) to distinguish individuals with high cardio-metabolic risk. At baseline BMI was not able to classify individuals with high MetS score at 7.5-year follow-up (AUC: 0.582,  $p = 0.287$ ), while at 2-year follow-up BMI was a significant predictor of high MetS score in early adulthood (AUC: 0.667,  $p = 0.012$ ). We further calculated Youden index for the metabolic biomarkers and found that the baseline (age 11 years) ApoB/ApoA1, GlycA and L-HDL-PL predicting 7.5-year follow-up (age 18 years) MetS, the Youden index was 0.518, 1.14 and 0.286, respectively. The Youden index of 2-year follow-up (age 13 years) predicting 7.5-year follow-up was 0.496, 1.31 and 0.327, respectively (**Supplementary Table 7**).

#### *Directional influences between the serum metabolites and cardio-metabolic risk*

We used the random intercept cross-lagged panel model (RI-CLPM) to assess the direction of effects between metabolic biomarkers and the MetS score based on the definition of Granger causality. In the model (see **Figure 4a**), the random intercepts ( $iA$  and  $iM$ ) represent stable part of between-subjects variation varying across but not within individuals. We additionally modelled within-subjects variation across time points to see, if there were associations between the two variables over time (cross-lag parameters  $c1$  and  $c3$  for the biomarker, and  $c2$  and  $c4$  for the metabolite, respectively) when the variables were adjusted by their previous measurements (autocorrelation parameters  $d1$  and  $d3$  for the biomarker, and  $d2$  and  $d4$  for the metabolite, respectively). The results suggested a causal effect of baseline ApoB/ApoA1 on MetS score at 2-year follow-up, while MetS score at 2-year follow-up had causal effect on ApoB/ApoA1 at 7.5-year



**Figure 3. LASSO regression identifies metabolite predictors of adult MetS score.** The upper panel represents metabolites at baseline predicted MetS score at 7•5-year follow-up. The lower panel represents metabolites at 2-year predicted MetS score at 7•5-year follow-up. A and d: The vertical dashed lines indicate the LASSO fit for which the mean squared error is smallest. The lower x-axis represents LASSO parameter lambda while the upper x-axis represents the number of variables with non-zero coefficients. The y-axis represents LASSO regression coefficient which is depending on the value of lambda. b and e: LASSO regression coefficients of selected metabolites. c and f: The area under the curve of these significant metabolites (AUC) by Receiver Operating Characteristics (ROC) analysis. ApoB/ApoA1 = apolipoprotein B to apolipoprotein A1 ratio; GlycA = glycoprotein acetyls; HDL2C = high-density lipoprotein two cholesterol; M-VLDL-TG = medium lipoprotein triglycerides; L-HDL-PL = very large high-density lipoprotein phospholipids; M-HDL-FC = medium high-density lipoprotein free cholesterol; XXLV-LDL-TG = extremely large very-low-density lipoprotein triglycerides; His = Histidine.

follow-up (Figure 4b). Similarly, the MetS score at baseline had a causal effect on L-HDL-PL at 2-year follow-up, while L-HDL-PL at 2 year-follow-up had a causal effect on MetS score at 7.5-year follow-up (Figure 4d). The results also suggested causal predominance of MetS score at baseline and at 2 year-follow-up on GlycA both at 2-year and 7.5-year follow-ups, but the associations were not significant (Figure 4c). All RI-CLPM results are shown in **Supplementary Table 8**.

#### Key metabolite predictors are confirmed in validation cohorts

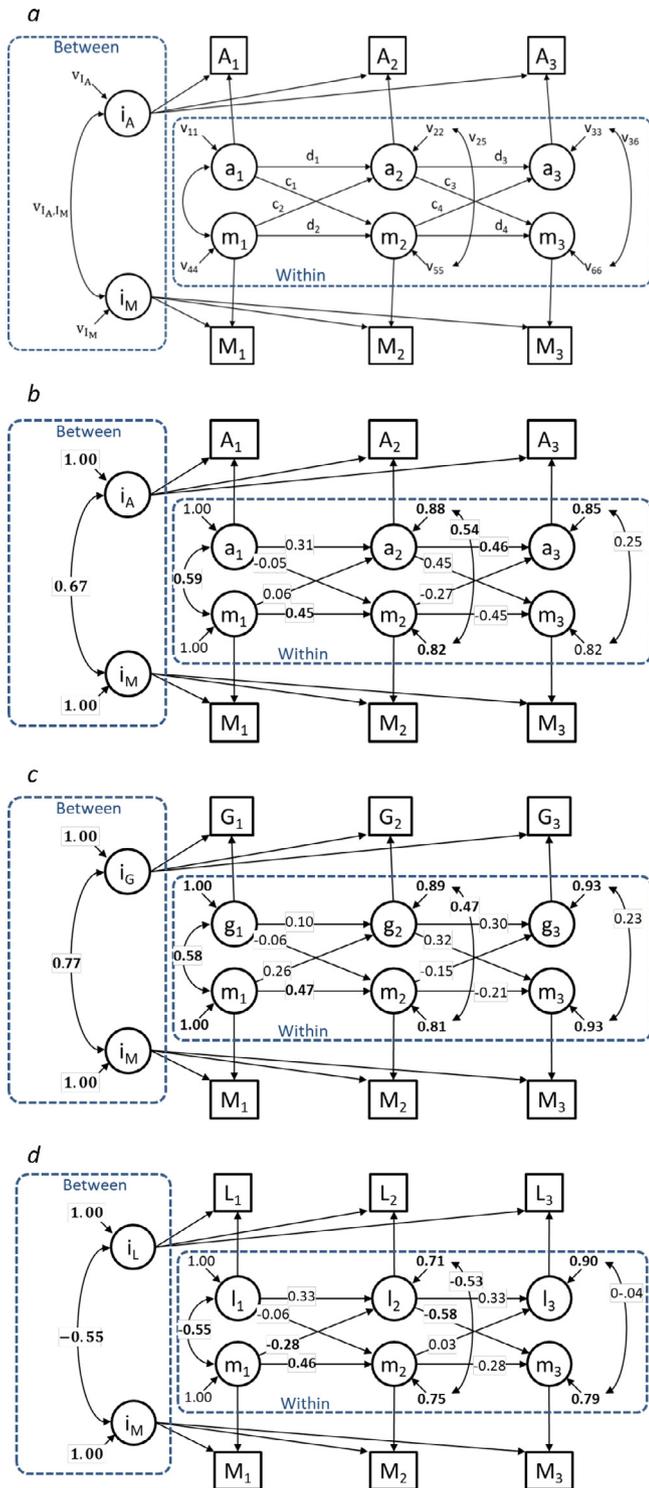
To validate our findings, we first sought to test the main results in a large longitudinal cohort study (Avon Longitudinal Study of Parents and Children (ALSPAC) [14]. The general characteristics of the ALSPAC cohort are shown in **Supplementary Table 2a**. We stratified the study participants into quartiles by their MetS score and by the cutoff for high risk as found in the discovery cohort and compared the metabolic biomarker (ApoB/ApoA1 ratio, GlycA and L-HDL-PL) levels between the groups. All metabolic biomarker levels were significantly different between the highest and lowest quartiles, and the area under the curve (AUCs) of these three metabolic biomarkers were similar to the discovery cohort at all-time points from puberty to early adulthood in both females and males (**Supplementary Figure 2a,b,c**), confirming the ability of the identified metabolite markers to predict future cardiometabolic risk.

We next explored the main results in four cross-sectional datasets. There were two datasets from the same laboratory including the discovery cohort participants' sisters and their biological mothers, and a cohort of middle-aged overweight and obese pre-menopausal women (**Supplementary Table b**). As the validation cohort includes overweight and obese women with more cardio-metabolic risk factors, this allowed us to compare the concentrations of the metabolite biomarkers identified in the discovery cohort between healthy

women and women with metabolic syndrome, to validate that the differences observed in puberty persist into adulthood. Furthermore, in mothers and sisters we compared the high risk and low risk groups by quartiles (the highest quartile represents high risk) (**Figure 5a,d,g and Supplementary Table 9**).

We further used LASSO in the validation datasets (sisters, mothers and menopausal women) to assess whether the performance was similar to the discovery cohort. Comparing to the discovery cohort at baseline ( $r^2 = 0.77$ ) and 7.5-year follow-up ( $r^2 = 0.59$ ), the explained variances of metabolites predicting MetS in the sisters ( $r^2 = 0.47$  and  $r^2 = 0.46$ , respectively), mothers ( $r^2 = 0.48$  for both) and pre-menopausal women ( $r^2 = 0.48$  and  $r^2 = 0.52$ , respectively, **Supplementary file 2**), which were lower than the discovery cohort due to cross-sectional nature of the datasets. However, the AUCs were similar to the discovery cohort (**Figure 5 b, e and h and Supplementary Table 10**). Additionally, a regression analysis on metabolic biomarkers identified by LASSO showed a strong correlation between the same key metabolic biomarkers (ApoB/ApoA1 ratio, GlycA and L-HDL-PL) and MetS score in both validation cohorts (**Figure 5 c, f, i**).

We then replicated the main findings in the large Northern Finland Birth Cohort 1966 study (NFBC1966) at 46 years dataset [15]. Characteristics of the participants are shown in **Supplementary Table 2c**. We found all three biomarker levels were significantly different between the high and low cardio-metabolic risk groups both in women (**Supplementary Figure 3a**,  $p < 0.001 - 0.0001$ , ANOVA adjusted for multiple comparisons by Bonferroni) and in men (**Supplementary Figure 3d**,  $p < 0.05 - 0.0001$ ). The regression analysis showed that metabolites were correlated with MetS significantly in similar ways both in women ( $r$  range from 0.644 to 0.674, **Supplementary Figure 3c**) and men ( $r$  range from 0.613 to 0.698), respectively, (**Supplementary Figure 3f**). The AUCs were also similar in women and men, and showing even stronger performance than in the discovery cohort (**Supplementary Figure 3b**, women: ApoB/



**Figure 4. Directional influences between the common metabolite predictors.** a: Conceptual random intercept cross-lagged panel model (RI-CPLM). The path coefficients of main interest include the between- participants random intercept variances ( $v_{i_A}$ ,  $v_{i_M}$ ) and covariance ( $v_{i_A, i_M}$ ) and within- participants autoregressive coefficients ( $d$ ) and cross-lagged coefficients ( $c$ ). The remaining parameters ( $v$ ) include variances / residual variances and correlations / residual correlations among the variables. Directed arrows with no coefficient attached were constrained to unity. b: The RI-CPLM for ApoB/ApoA1; c: the RI-CPLM for GlycA; d: the RI-CPLM for L-HDL-PL. Likelihood ratio tests indicated a significant between-participants effect in the model when ApoB/ApoA1 ( $\chi^2_{df=3} = 17.0$ ,  $p = 0.001$ , b), GlycA ( $\chi^2_{df=3} = 10.2$ ,  $p = 0.017$ , c), or L-HDL-PL ( $\chi^2_{df=3} = 12.1$ ,  $p = 0.007$ , d), were considered, while the within- participants effects did not have a significant contribution for either ApoB/ApoA1 ( $\chi^2_{df=4} = 3.7$ ,  $p = 0.442$ ) or GlycA ( $\chi^2_{df=4} = 4.6$ ,  $p = 0.333$ ) or L-HDL-PL ( $\chi^2_{df=4} = 8.6$ ,  $p = 0.072$ ). The between components, indicated by A, are the random intercepts:  $i_A$  for metabolic biomarkers and  $i_M$  for MetS score. The within components, indicated by the letters a and m, are the differences between observed measurements and the expected score based on the means and their random intercepts.

ApoA1 ratio = 0.846, GlycA = 0.905 and L-HDL-PL = 0.877; Men: ApoB/ApoA1 ratio = 0.807, GlycA = 0.889 and L-HDL-PL = 0.839 **Supplementary Figure 3e**).

Finally, we compared the metabolic biomarkers between the T2D ( $n = 191$ ) and non-diabetic controls ( $n = 200$ ) (**Supplementary Table 2d**), and found a significant difference in GlycA and L-HDL-PL but not ApoB/ApoA ratio between the groups. After adjusting for age, sex, BMI, and use of statin, antihypertensive and anti-diabetic medications, the significant differences in GlycA and L-HDL-PL between the groups remained. We further performed ROC analysis and found a moderately high performance of GlycA and L-HDL-PL in classifying T2D in the elderly (AUC: 0.656 and 0.668, respectively,  $p < 0.001$  for both, ROC asymptotic significant) (**Supplementary Figure 4**).

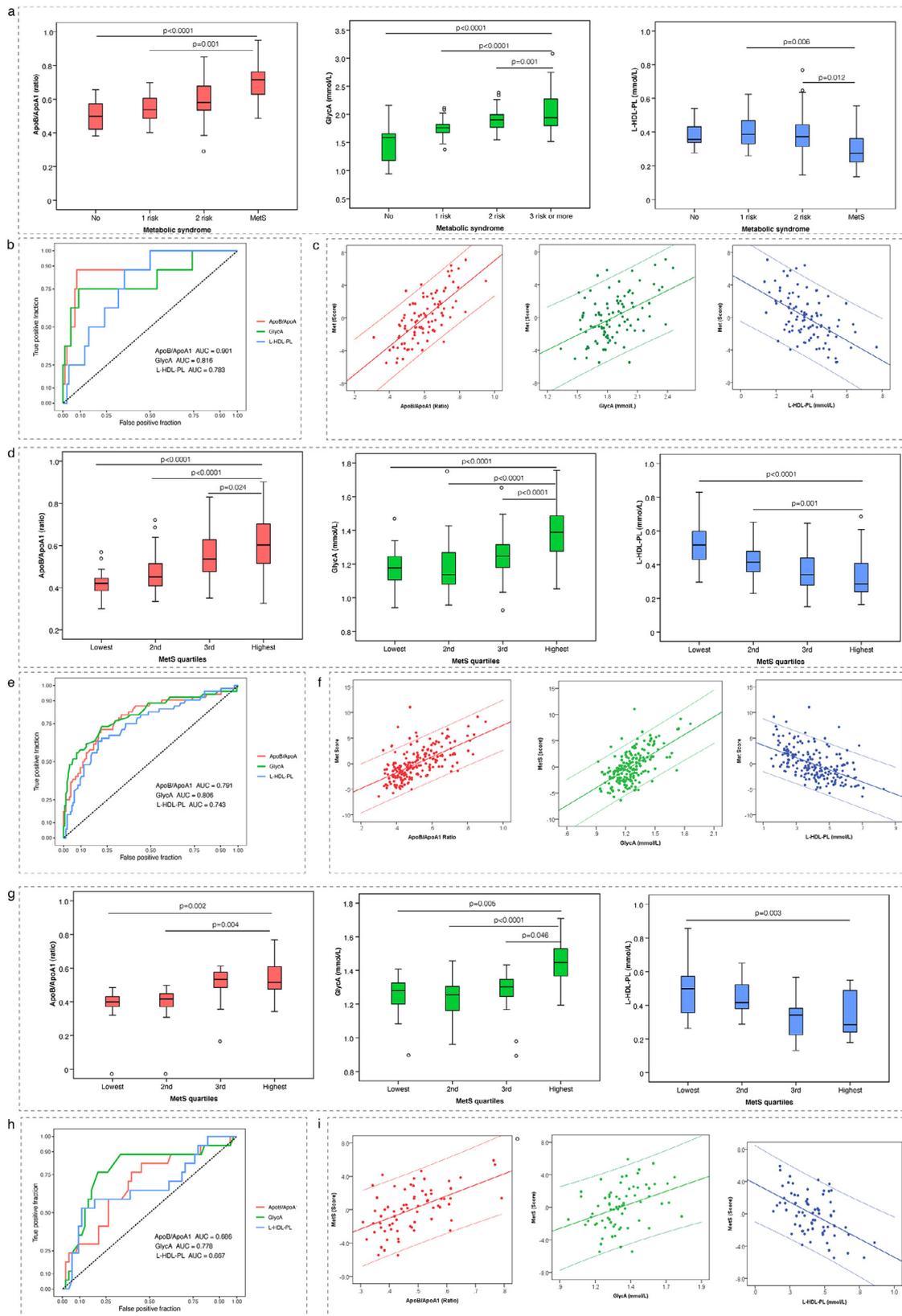
Since the validation cohort of ALSPAC has three-time points measurements in females and males, we also calculated the Youden index for ApoB/ApoA1, GlycA and L-HDL-PL at 14-year or 17-year old to predict the MetS at 24-year old. The Youden indexes (in females: for ApoB/ApoA1 = 0.504 or 0.468, for GlycA = 1.269 or 1.189 and for L-HDL-PL = 0.372 or 0.350, respectively; in males: for ApoB/ApoA1 = 0.455 or 0.504, for GlycA = 1.145 or 1.169 and for L-HDL-PL = 0.310 or 0.291, respectively) were comparable to the discover cohort (**Supplementary Table 7**).

Taken together, these validation analyses confirm that the identified metabolites are elevated in both males and females who are at high risk for cardio-metabolic disease across adolescence, adulthood, and older age.

## Discussion

In this study, we utilised a rich longitudinal data resource, machine learning techniques, and statistical modelling to identify a metabolic signature in childhood that predicts increased cardiovascular disease risk in adulthood. These metabolic measures (ApoB/ApoA1 ratio, GlycA, L-HDL-PL) have atherogenic properties and reflect chronic systemic inflammation, previously associated with future cardiovascular disease and pre-mature mortality in older individuals [16]. Thus, our results provide further evidence that the pathogenic processes that contribute to cardiovascular diseases in later life originate in childhood and adolescence, providing an impetus to earlier intervention strategies to reduce the global burden of cardiovascular disease.

We found that the serum ApoB/ApoA1 ratio in childhood strongly and consistently predicted future cardio-metabolic risk score (MetS) across all time points. ApoB and ApoA1 are the two main lipoproteins involved in lipid transport. ApoB is the main protein in VLDL and LDL (atherogenic) particles, while ApoA1 is the main protein in HDL (anti-atherogenic) particles. This result is supported by an earlier 9-year follow-up study in school girls from childhood to adulthood, demonstrating that ApoB/ApoA1 ratio was associated with metabolic syndrome and its components [17]. In another prospective study, Juonala et al. reported that ApoB and ApoA1 levels and their ratio in adolescence were associated with carotid artery intima-media thickness and brachial artery flow-mediated dilation in adulthood [18]. In that study, ApoB and ApoA1 were stronger predictors of abnormal vascular changes than conventional cholesterol measurements (LDL-C and HDL-C), which suggests that the carriers (apolipoproteins) might play a more central role than the actual lipid content transported in these lipoprotein particles. These findings are further supported by the Pathobiological Determinants of Atherosclerosis in Youth study which found apolipoproteins associated with post-mortem arterial lesions, [19] and the Bogalusa Heart Study, which showed that high ApoB/ApoA1 ratio in children was associated with incidence of parental myocardial infarction, [20] and a recent publication identifying ApoB/ApoA1 ratio as a predictor of cardiovascular risk in adolescent SLE patients [21]. A recent study also showed that



**Figure 5. Validations cohorts.** Comparison of selected metabolites across different cardio-metabolic risk groups. a, b and c: premenopausal women (MS = metabolic syndrome, NMS = without MS); d, e and f: biological mothers; g, h and i: biological sisters. a: comparison of metabolic syndrome groups; d and g comparison of MeTS quartiles; plot boxes represent median with 95% confident interval (CI) and the circles out of the plot boxes are the outliers corresponding to each group. b, e and h: the area under the curve of these significant metabolites (AUC) by Receiver Operating Characteristics (ROC) analysis. c, f and i: regression between MeTS and selected metabolites. ApoB/ApoA1 = apolipoprotein B to apolipoprotein A1 ratio; GlycA = glycoprotein acetyls; L-HDL-PL = very large high-density lipoprotein phospholipids.

increased ApoB/ApoA1 ratio predicts cardiometabolic risk in patients with juvenile onset systemic lupus erythematosus [21].

We also found a strong inverse association between L-HDL phospholipids in childhood and cardio-metabolic risk in early adulthood. A previous study by Piperi et al. found that HDL-phospholipids were more closely related to coronary artery disease than HDL-C or other lipoproteins studied [22]. Meikle et al. also recently reported that HDL phospholipids, but not HDL cholesterol, distinguished acute coronary syndrome from stable coronary artery disease [23]. Low serum HDL-phospholipid concentrations have also been associated with high coronary artery calcification scores in asymptomatic patients with atherosclerosis, [24] and with increased risk of metabolic syndrome and coronary heart disease, particularly in women [25]. Our key finding of the strong inverse association between L-HDL phospholipids in childhood and cardio-metabolic risk in early adulthood was confirmed in both the validation longitudinal and cross-sectional cohorts, demonstrating that adult women and men with metabolic syndrome had lower serum L-HDL-PL concentration than their healthy counterparts. Our results also demonstrated that L-HDL-PL level was significantly lower in the T2D than the non-T2D of elderly people and was independent of medication for diabetes, hypertension and hyperlipidemia. It has been previously shown that HDL phospholipids play an important role in the cholesterol efflux process, [26] and that metabolic syndrome is associated with progressive reduction in cholesterol efflux capacity, which contributes to development of atherosclerosis [27]. Taken together, our results suggest that exposure to an atherogenic apolipoprotein profile and low HDL phospholipids in childhood may cause reduced cholesterol efflux capacity and changes in the arteries that contribute to the development of atherosclerosis and coronary heart disease in adulthood.

It is increasingly recognised that the atherosclerotic process involves not only lipid and lipoprotein metabolism but it also requires a pro-inflammatory response that includes both the innate and adaptive immune systems. Another NMR marker predicting MetS score in our study was GlycA, which reflects systemic inflammation originating from glycan groups of acute-phase glycoproteins, mainly  $\alpha$ 1-acid glycoprotein, but also other acute-phase reactants such as haptoglobin,  $\alpha$ 1-antitrypsin,  $\alpha$ 1-antichymotrypsin, and transferrin [28]. GlycA has been found to correlate with adiposity, insulin resistance and other markers of metabolic syndrome in adults, suggesting that in addition to being elevated in acute inflammation, GlycA might also be a biomarker of subclinical vascular inflammation [29]. Accordingly, it was recently shown that plasma GlycA is independently associated with the incidence of cardiovascular disease in a large cohort study of initially healthy women [30]. Elevated circulating GlycA levels have also been shown to predict risk of T2D, nonalcoholic fatty liver disease, [31] and all-cause mortality [29] in adults, but so far, no comparable data are available in children and adolescents. In our study, GlycA levels in childhood and adolescence strongly and consistently predicted MetS score in early adulthood, and this finding was confirmed in all validation cohorts. Thus, the findings of our study confirm and extend the results of earlier reports by demonstrating that GlycA is a viable biomarker of systemic subclinical inflammation associated with increased cardio-metabolic risk not only in adults but also in children and adolescents. However, further studies are needed before GlycA could be routinely used in clinical tests for the purposes of risk assessment, management, or follow-up in paediatric populations.

We further investigated the direction of associations between metabolite biomarkers and cardio-metabolic risk using a random intercept cross-lagged path model [32]. This novel analytical approach revealed a bidirectional association pattern in which increased ApoB/ApoA1 ratio in childhood was associated with an increased MetS score in adolescence, and elevated MetS score in adolescence was associated with increased ApoB/ApoA1 ratio in early adulthood. Similar bidirectional associations were observed between

the MetS score and the other two key metabolites i.e., L-HDL-PL and GlycA from childhood to early adulthood. Importantly, these associations were also robust to multi-covariate adjustment, including insulin, leptin, adiponectin, sex steroids, IGF-1, physical activity and energy yield nutrient intakes, suggesting that metabolites and cardiovascular risk factors aggravate each other, leading inexorably to a worsening cardio-metabolic health.

Our results support the theory that cardiovascular diseases originate from childhood. The novelty of our study lies in the application of a machine learning approach to identify the early risk biomarkers and the sophisticated statistical models to reveal the bidirectional relationship between the metabolic biomarkers and MetS score. The AUCs for the biomarkers showed moderate to excellent performance in the discovery and validation cohorts except for the very old T2D patients. In terms of the AUC values, the metabolic biomarkers were superior discriminators of high MetS score than BMI in our study as well as compared to an earlier meta-analysis of cross-sectional studies [33]. Furthermore, we were also able to identify an appropriate cut-off value for the cardio-metabolic risk, which may be of value as potential clinical application for identifying high cardiometabolic risk individuals at an early age. A further major strength of our study is the longitudinal data on children followed from pre-puberty to early adulthood, and validation of the results in several cohorts, including the large longitudinal birth cohort of females and males followed from puberty to adulthood.

There are several limitations that warrant consideration. The prevalence of metabolic syndrome was low among the children and adolescents studied. However, while binary definition of metabolic syndrome might be a useful tool for clinical practice to assess cardiovascular disease risk in adults, [13] continuous MetS score is more appropriate for epidemiological studies for the following reasons: first, dichotomizing continuous outcome variables reduces statistical power; second, the risk of cardiovascular disease is an aggregative progressive function of several risk factors, and third, cardiovascular risk increases progressively with increasing numbers of risk factors. Therefore, numerous studies with paediatric and adult populations have used continuous metabolic risk scores, integrating components of the metabolic syndrome definition to represent clustering of metabolic risk factors. Another limitation to consider is that the discovery cohort included only females. However, the results were well validated in different cohorts including males and females of different age and demographics. Finally, although most of the metabolic biomarkers presented in this study are not routinely used in clinical settings today, the use of high-throughput and cost-effective metabolomics platforms in the clinic are emerging. For example, different biobanks and companies have offered their services for clinical and epidemiological studies, hence the applicability in clinical settings is right around the corner. The findings of our study may open the door for future screening programmes for children and families at high risk for atherosclerotic cardiovascular diseases. Further studies are needed to explore the generalizability and the potential clinical utility of these biomarkers to lower the global burden of cardio-metabolic disease through shaping early-life health strategies.

In conclusion, previous observational studies have found an association of cardio-metabolic risk with atherogenic lipoproteins, cholesterol efflux capacity and subclinical systemic inflammation. Our results show that these metabolite-cardio-metabolic risk associations are present already in early childhood, thus providing evidence for the utility of circulating metabolomics panel to identify children and adolescents at risk for cardiovascular disease in future.

## Funding

This study was financially supported by the Academy of Finland, Ministry of Education of Finland and University of Jyväskylä, the National Nature Science Foundation of China (Grant 31571219), the

111 Project (B17029), the Shanghai Jiao Tong University Zhiyuan Foundation (Grant CP2014013), China Postdoc Scholarship Council (201806230001), the Food and Health Bureau of Hong Kong SAR's Health and Medical Research Fund (HMRF grants 15162161 and 07181036) and the CUHK Direct Grants for Research (2016-033 and 2018-034), and a postdoctoral fellowship from K. Carole Ellison (to T. W.). The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. NFBC1966 received financial support from University of Oulu Grant no. 24000692, Oulu University Hospital Grant no. 24301140, ERDF European Regional Development Fund Grant no. 539/2010 A31592. This work was supported by European Union's Horizon 2020 research and innovation programme LongITools 874739.

## Contributors

XO, PW and SC participated in data collection, data analysis and drafted the manuscript. XO, RC, NW, WY, TT performed the data analysis. RC, NW, NR, TW, WY, TT, DV, MA and NDP edited the manuscript. EF provided and analysed the UFO and HARE validation study cohorts and edited the manuscript. RN and SS provided and analysed the NFBC1966 validation study cohorts and edited the manuscript. PW, SC, NR, TW and NDP designed the study. All authors approved the submitted version. PW and SC have full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

## Data sharing

Availability of the data will follow the data management principles for research at the University of Jyväskylä <https://www.jyu.fi/tutkimus/tutkimusaineistot/rdmenpdf>. Established researchers wishing to collaborate will be given access to the de-identified data following approval of a signed research proposal.

NFBC data is available from the University of Oulu, Infrastructure for Population Studies. Permission to use the data can be applied for research purposes via electronic material request portal. In the use of data, we follow the EU general data protection regulation (679/2016) and Finnish Data Protection Act. The use of personal data is based on cohort participant's written informed consent at his/her latest follow-up study, which may cause limitations to its use. Please, contact NFBC project centre (NFBCprojectcenter@oulu.fi) and visit the cohort website ([www.oulu.fi/nfbc](http://www.oulu.fi/nfbc)) for more information.

## Supplementary materials

Supplementary file 1. Methods

## Declaration of Competing Interest

The authors declared no conflict of interest.

## Acknowledgements

We are extremely grateful to all the participants who took part in this study, the study assistants, nurses, midwives for their help in recruiting and other activities, interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, and receptionists.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.ebiom.2021.103611](https://doi.org/10.1016/j.ebiom.2021.103611).

## References

- [1] Duong M, Islam S, Rangarajan S, et al. Mortality and cardiovascular and respiratory morbidity in individuals with impaired FEV1 (PURE): an international, community-based cohort study. *Lancet Glob Health* 2019;7(5):e613–23.
- [2] Ballantyne CM, Hoogeveen RC, McNeill AM, et al. Metabolic syndrome risk for cardiovascular disease and diabetes in the ARIC study. *Int J Obes (Lond)* 2008;32(Suppl 2):S214.
- [3] Daniels SR, Pratt CA, Hayman LL. Reduction of risk for cardiovascular disease in children and adolescents. *Circulation* 2011;124(15):1673–86.
- [4] Berenson GS, Foster TA, Frank GC, et al. Cardiovascular disease risk factor variables at the preschool age. The Bogalusa heart study. *Circulation* 1978;57(3):603–12.
- [5] Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17(4):448–53.
- [6] Tobias DK, Lawler PR, Harada PH, et al. Circulating Branched-Chain Amino Acids and Incident Cardiovascular Disease in a Prospective Cohort of US Women. *Circ Genom Precis Med* 2018;11(4):e002157.
- [7] Christensen JJ, Ulven SM, Retterstol K, et al. Comprehensive lipid and metabolite profiling of children with and without familial hypercholesterolemia: A cross-sectional study. *Atherosclerosis* 2017;266:48–57.
- [8] McCormack SE, Shaham O, McCarthy MA, et al. Circulating branched-chain amino acid concentrations are associated with obesity and future insulin resistance in children and adolescents. *Pediatr Obes* 2013;8(1):52–61.
- [9] Cheng S, Volgyi E, Tylavsky FA, et al. Trait-specific tracking and determinants of body composition: a 7-year follow-up study of pubertal growth in girls. *BMC Med* 2009;7:5.
- [10] Wiklund P, Zhang X, Tan X, Keinänen-Kiukkaanniemi S, Alen M, Cheng S. Serum Amino Acid Profiles in Childhood Predict Triglyceride Level in Adulthood: A 7-Year Longitudinal Study in Girls. *J Clin Endocrinol Metab* 2016;101(5):2047–55.
- [11] Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015;8(1):192–206.
- [12] Cheng S, Wiklund P. The effects of muscle mass and muscle quality on cardiometabolic risk in peripubertal girls: a longitudinal study from childhood to early adulthood. *Int J Obes (Lond)* 2018;42(4):648–54.
- [13] Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;120(16):1640–5.
- [14] Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;42(1):111–27.
- [15] Sebert S, Lowry E, Aumuller N, et al. Cohort Profile: The DynaHEALTH consortium – a European consortium for a life-course bio-psychosocial model of healthy ageing of glucose homeostasis. *Int J Epidemiol* 2019;48(4):1051.
- [16] Wurtz P, Havulinna AS, Soininen P, et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015;131(9):774–85.
- [17] Morrison JA, Glueck CJ, Daniels SR, Horn PS, Wang P. Determinants of ApoB, ApoA1, and the ApoB/ApoA1 ratio in healthy schoolgirls, prospectively studied from mean ages 10 to 19 years: the Cincinnati National Growth and Health Study. *Metabolism* 2012;61(10):1377–87.
- [18] Juonala M, Viikari JS, Kahonen M, et al. Childhood levels of serum apolipoproteins B and A-I predict carotid intima-media thickness and brachial endothelial function in adulthood: the cardiovascular risk in young Finns study. *J Am Coll Cardiol* 2008;52(4):293.
- [19] Rainwater DL, McMahan CA, Malcom GT, et al. Lipid and apolipoprotein predictors of atherosclerosis in youth: apolipoprotein concentrations do not materially improve prediction of arterial lesions in PDAY subjects. The PDAY Research Group. *Arterioscler Thromb Vasc Biol* 1999;19(3):753–61.
- [20] Srinivasan SR, Berenson GS. Serum apolipoproteins A-I and B as markers of coronary artery disease risk in early life: the Bogalusa Heart Study. *Clin Chem* 1995;41(1):159–64.
- [21] Robinson GA, Waddington KE, Coelewij L, et al. Increased apolipoprotein-B:A1 ratio predicts cardiometabolic risk in patients with juvenile onset SLE. *EBioMedicine* 2021;65:103243.
- [22] Piperi C, Kalofoutis C, Papaevaggeliou D, Papapanagiotou A, Lekakis J, Kalofoutis A. The significance of serum HDL phospholipid levels in angiographically defined coronary artery disease. *Clin Biochem* 2004;37(5):377–81.
- [23] Meikle PJ, Formosa MF, Mellett NA, et al. HDL Phospholipids, but Not Cholesterol Distinguish Acute Coronary Syndrome From Stable Coronary Artery Disease. *J Am Heart Assoc* 2019;8(11):e011792.
- [24] Garcia-Sanchez C, Posadas-Romero C, Posadas-Sanchez R, et al. Low concentrations of phospholipids and plasma HDL cholesterol subclasses in asymptomatic subjects with high coronary calcium scores. *Atherosclerosis* 2015;238(2):250.
- [25] Hergenc G, Onat A, Sari I, Yazici M, Eryonuuc B, Can G. Serum total and high-density lipoprotein phospholipid levels in a population-based study and relationship to risk of metabolic syndrome and coronary disease. *Angiology* 2008;59(1):26–35.
- [26] Rohatgi A, Khera A, Berry JD, et al. HDL cholesterol efflux capacity and incident cardiovascular events. *N Engl J Med* 2014;371(25):2383–93.

- [27] Gall J, Frisdal E, Bittar R, et al. Association of Cholesterol Efflux Capacity With Clinical Features of Metabolic Syndrome: Relevance to Atherosclerosis. *J Am Heart Assoc* 2016;5(12).
- [28] Otvos JD, Shalaurova I, Wolak-Dinsmore J, et al. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clin Chem* 2015;61(5):714–23.
- [29] Fischer K, Kettunen J, Wurtz P, et al. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* 2014;11(2):e1001606.
- [30] Connelly MA, Otvos JD, Shalaurova I, Playford MP, Mehta NN. GlycA, a novel biomarker of systemic inflammation and cardiovascular disease risk. *J Transl Med* 2017;15(1):219.
- [31] Kaikkonen JE, Wurtz 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(2):491–500.
- [32] Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 1969;37:424–38.
- [33] Lo K, Wong M, Khalechelvam P, Tam W. Waist-to-height ratio, body mass index and waist circumference for screening paediatric cardio-metabolic risk factors: a meta-analysis. *Obes Rev* 2016;17(12):1258–75.