

Paediatric proteomic signatures of cardiometabolic disease-associated traits predict adult disease outcomes

Received: 27 March 2026

Accepted: 20 July 2026

Published online: 11 September 2026

 Check for updates

A list of authors and their affiliations appears at the end of the paper

While the earliest pathological signs of cardiovascular–kidney–metabolic disease (CKMD) emerge before age 20, current adult-based risk thresholds fail to identify a substantial fraction of children at high risk. With increasing childhood obesity, early, sensitive detection of CKMD risk is critical to timely intervention. We linked 25 CKMD phenotypes spanning liver, adipose, vascular and dysglycaemia traits to the circulating proteome in 273 Hispanic or Latino children and adolescents (13.1 ± 2.7 years; 53% female). Here we show that proteome signatures of CKMD in children are strongly associated with clinical CKMD outcomes in adults. In 685 adults from the same community and 28,256 adults from UK Biobank, we observe high concordance between the paediatric and adult CKMD phenotype–proteome relationships, related to pancreatic beta-cell health and insulin sensitivity, liver homeostasis, inflammation and cholesterol metabolism. Many proteins linked to paediatric CKMD were modifiable with glucagon-like peptide-1 receptor agonist therapy in adults and are linked to genetic liability to CKMD. These findings support an early origin of CKMD disease liability embedded in the human proteome in children, underscoring the value of precision prevention to forestall metabolic disease long before its clinical manifestation.

The earliest pathological signs of cardiovascular–kidney–metabolic disease (CKMD) emerge before age 20 (ref. 1), driven by cumulative exposure to obesity, insulin resistance, hypertension and pro-atherogenic dyslipidaemia². Despite strong epidemiologic evidence highlighting the importance of these risks to premature CKMD^{3–8}, current CKMD prevention efforts remain squarely focused on adults, missing an opportunity to identify high-risk individuals at the earliest stage of childhood. While most children and adolescents harbour intermediate subclinical phenotypes that fall below adult thresholds, even ‘normal’ values in adolescents are linked to CKMD in adulthood^{3,4}. In addition, screening tests commonly used in adults (for example, fasting glucose) miss a substantial fraction (up to 70%)⁹ of dysglycaemia and type 2 diabetes (T2D) in children¹⁰. As a response, molecular biomarkers have been advanced in adults as a ‘precision medicine’ approach to detect disease liability decades before clinical onset^{11–14}, although analogous

studies in paediatric populations are less mature. A key challenge is that identifying paediatric biomarkers to predict clinical risk requires decades-long follow-up data, limiting efforts to identify reversible, actionable, prognostic mechanisms in childhood.

In this Article, we linked 25 CKMD phenotypes in 273 Hispanic or Latino children and adolescents (mean age 13.1 ± 2.7 years; 53% female) to the circulating proteome to identify molecular biomarkers of early CKMD susceptibility in children. CKMD phenotypes spanned clinical and imaging-based measures known to influence risk in adults, including body composition, liver fat and fibrosis, renal function and traditional CKMD risk factors. We identified proteomic patterns of these phenotypes in children at single- and multi-trait levels, generating multi-CKMD phenotype signatures to capture shared molecular risk, high interrelatedness, frequent co-occurrence and shared mechanistic pathways. We subsequently mapped the paediatric multi-protein

✉ e-mail: Kari.E.North@uth.tmc.edu; ravi.shah@vumc.org

composite CKMD signatures to adult CKMD outcomes in two samples: (1) 494 Hispanic or Latino adults from the same community (mean age 51 ± 14.3 years; 68% female) and (2) 28,256 adults from the UK Biobank (median age 58 years; 54% female). Given the application of glucagon-like peptide-1 receptor agonist (GLP-1RA) therapy in adults and children, we evaluated whether proteins implicated in paediatric CKMD were modifiable during GLP-1RA therapy in adults in the Semaglutide Treatment Effect in People with Obesity (STEP) 1 randomized clinical trial¹⁵. In parallel, we constructed genetic instruments¹⁶ of the circulating proteome to conduct proteome-wide association studies (PWAS) to test whether this GLP-1RA-responsive paediatric CKMD proteome would show a potentially causal association with adult CKMD outcomes (such as obesity, T2D, hepatic steatosis, myocardial infarction and kidney function). Finally, we delineated sources of variability in the paediatric CKMD proteome, as observed in recently published resources¹⁷, as understanding such heterogeneity is a prerequisite to targeted intervention. Ultimately, our goal was to map the proteomics of multidimensional CKMD-risk phenotypes in childhood with adult clinical CKMD markers through deep phenotyping and proteomic and genomic analyses, thereby offering a potential precision prevention approach to address clinical CKMD at its earliest manifestation.

Results

Characteristics of paediatric and adult samples for single-trait analyses

Paediatric participants were predominantly adolescents (13.1 ± 2.7 years of age), with approximately 53% self-reporting female sex (Table 1). More than a third of the Border Health Research Cohort (BHRC) paediatric participants had obesity with indications of pro-atherogenic dyslipidaemia (high triglycerides (TG) and low high-density lipoprotein (HDL-C)), hypertension and insulin resistance. Adult participants from the same cohort, distinct from the paediatric participants but including some first-degree relatives, were at similarly high risk for these conditions (51.5 ± 14.3 years of age, 54.1% with obesity, 39.8% with T2D, 68.3% females). We observed expected associations among CKMD phenotypes in youth, with body composition associated with greater insulin resistance, body fat distribution, pro-atherogenic dyslipidaemia and elevated blood pressure (Extended Data Fig. 1). The overall analytical flow for all results is shown in Fig. 1.

We estimated the associations between the circulating proteome and the 25 paediatric CKMD phenotypes across the 6 defined domains (Supplementary Table 1), finding that 1,064 of the 5,420 unique proteins in our platform were associated with at least 1 phenotype. We identified 2,916 protein–phenotype associations (q -value < 0.05 ; the ‘CKMD proteome’; the full results are in Supplementary Table 2). A substantial fraction of the 1,064 associated proteins was statistically significant across multiple paediatric CKMD phenotypes (Fig. 2a), and 65 proteins were consistently associated with at least 4 CKMD domains (Fig. 2b). While a complete mechanistic ontology lies beyond the scope of this work, the paediatric CKMD proteome (see curated entries in Table 2) reveals coherent signals that align with key metabolic processes in CKMD.

Characterizing the paediatric CKMD proteome

Among proteins associated with glycaemic control, we observed markers indexing distinct facets of β -cell function and insulin sensitivity. For example, C-peptide¹⁰ (an indicator of residual β -cell secretory reserve) was associated with many metabolic measures, reinforcing the value of measuring endogenous insulin secretion in early disease. Meanwhile, chromogranin B (CHGB), known to modulate insulin granule trafficking in β -cells by facilitating proinsulin transit from the trans-Golgi to the plasma membrane, was inversely related to dysglycaemia in our paediatric cohort, echoing adult findings of CHGB influence on metabolic traits^{18–20}. In addition, canonical adult markers of insulin sensitivity, such as GUSB²¹, insulin-like growth-factor binding protein 1 (IGFBP1)

and IGFBP2, were detected and have shown consistent directionality in children^{22,23}. For hepatic traits, significant proteins implicate core metabolic machinery in the liver. For example, glycerol-3-phosphate dehydrogenase (GPD1) has a central role in glycerolipid metabolism and gluconeogenesis^{24,25}, while aldehyde dehydrogenase 1 family member A1 (ALDH1A1) is a key node in retinoid and aldehyde detoxification pathways²⁶. Some proteins, such as immunoglobulin superfamily member 9 (IGSF9), have been previously linked in adult cohorts to steatosis or fibrosis. Aminoacylase 1 (ACY1) has been identified in previous metabolic disease or hepatic injury studies^{23,27,28}.

Our results span both protective and risk-conferring associations with the CKMD phenome. For instance, elevated creatine kinase B (CKB), a key enzyme in phosphagen energy shuttling that has been implicated in thermogenesis in highly metabolic tissues, was associated with adiposity-related phenotypes in children. This is consistent with murine knockout studies where CKB influences energy expenditure, inflammation in adipose tissue and obesity outcomes^{29,30}. Similarly, CHGB’s association with lower insulin resistance fits with its secretory role and with adult metabolic directionality¹⁹. We also confirmed previous findings in paediatric populations: IGFBP1 and IGFBP2 (markers of insulin sensitivity) and sex hormone-binding globulin (SHBG), for which rapid early declines predict insulin resistance independently of adiposity, all aligned with our proteomic signals^{31–33}. In addition, proteins long implicated in adult CKMD—such as leptin (LEP), fatty acid-binding protein 4 (FABP4) and retinoic acid receptor responder 2 (RARRES2)—were recapitulated in children, suggesting potential similarities in molecular processes across the lifespan^{34–36}. In addition to well-established proteins, our analysis also revealed several proteins with plausible mechanistic roles but limited previous evidence in CKMD. These include β -site amyloid precursor protein cleaving enzyme 1 (BACE-1), G-protein coupled receptor 158 (GPR158), fibroblast growth-factor binding protein 3 (FGFBP3), ubiquitin specific peptidase 28 (USP28) and cysteine-rich with EGF-like domains 2 (CRELD2), among others^{37–43}. Such proteins operate in signalling, cellular stress or turnover pathways, as further detailed in Table 2, and represent candidates for future functional follow-up. Pathway analysis across the CKMD proteome further revealed associations with pathways of cholesterol metabolism, inflammation and haemostatic mechanisms, and metabolism (Fig. 2c,d). Indeed, mapping to the Human Genetic Evidence (HuGE) framework suggests external genetic support for many of the targets in the paediatric CKMD proteome with CKMD traits studied in adults, including anthropometric, cardiovascular, hepatic and renal phenotypes (Extended Data Fig. 2). Pairwise correlations across proteins indicated some correlation across proteins (average Spearman correlation, 0.258; Supplementary Fig. 1).

Similarities between the paediatric and adult CKMD proteome

Understanding links between the serum proteome and CKMD phenotypes in children—specifically, whether these links recapitulate relationships observed in adults—is key in identifying nascent, potentially modifiable features of CKMD. Across 685 adults in the BHRC and up to 25 phenotypes measured in both children and adults (Table 1), we observed a broad concordance in the magnitude and directionality of phenotype–protein relationships (Fig. 3 and Supplementary Table 3). Of the 2,916 significant paediatric protein–phenotype associations, 1,877 (64%) were significant and had directionally consistent effects. Across domains, we observed the strongest paediatric–adult correlation for the adiposity ($r = 0.61$) and liver domains ($r = 0.69$). While many proteins showed expected associations with phenotypic data (for example, ACY1, GOT1 with hepatic steatosis and fibrosis), we also observed associations with divergent directions in children and adults for some phenotypes, some of which have demonstrated pathophysiologic roles in CKMD (for example, PON3 (ref. 44), APOF⁴⁵ for some dysglycaemia phenotypes) as well as those with no previous evidence in CKMD (for example, MENT, NXPH1).

Table 1 | Demographic summary of BHRC children and adults in proteomics datasets

Domain	Trait (units)	Abbreviation	Multi-trait analysis	Paediatric mean (s.d.)	Paediatric n_{ind}	Adult mean (s.d.)	Adult n_{ind}
Dysglycaemia	Age (years)		No	13.1 (2.7)	273	51.5 (14.3)	494
	Sex, n (% female)		No	145 (53%)	273	356 (68.3%)	494
	Fasting blood glucose (mg dl ⁻¹)	FBG	Yes	88.63 (6.97)	270	92.2 (8.8)	289
	HbA1c (%)	HbA1c	Yes	5.34 (0.29)	271	5.54 (0.32)	245
	HOMA-IR	HOMA-IR	Yes	3.1 (1.9)	260	2.4 (1.6)	259
Blood pressure	Diastolic blood pressure (mm Hg)	DBP	No	62.2 (7)	273	73.8 (9.1)	486
	Systolic blood pressure (mm Hg)	SBP	Yes	102.8 (10)	273	122.9 (19.2)	486
	Pulse pressure (mm Hg)	PP	No	40.7 (6.7)	273	48.8 (14.4)	486
	Mean arterial pressure (mm Hg)	MAP	No	75.7 (7.5)	273	89.7 (11.5)	486
Kidney	eGFR (sex- and age-dependent equation)	eGFR	Yes	108.7 (18.6)	270	98.7 (18.7)	465
Dyslipidaemia	High-density lipoprotein (mg dl ⁻¹)	HDL-C	Yes	48.4 (10.7)	271	50.1 (14.5)	491
	Low-density lipoprotein (mg dl ⁻¹)	LDL-C	Yes	78.2 (22.5)	271	105.2 (32.5)	484
	Total cholesterol (mg dl ⁻¹)	TC	Yes	146.0 (26.1)	271	183.3 (37.4)	491
	Triglycerides (mg dl ⁻¹)	TG	Yes	97.4 (52.5)	271	145.3 (92.4)	491
Liver	Alanine aminotransferase (U l ⁻¹)	ALT	Yes	26.0 (21.0)	271	32.5 (18.3)	473
	Aspartate aminotransferase to platelet ratio index	APRI	Yes	0.24 (0.14)	271	0.32 (0.22)	464
	Aspartate aminotransferase (U l ⁻¹)	AST	Yes	21.4 (10.7)	271	24.4 (13.2)	477
	Controlled attenuation parameter (dB m ⁻¹)	CAP	No	245.8 (57.8)	148	288.1 (63.2)	437
	Fibrosis-4	FIB	Yes	0.23 (0.1)	271	0.97 (0.56)	462
	Liver stiffness measurement (kPa)	kPa	No	4.46 (1.57)	162	5.96 (4.67)	429
Adiposity	BMI z-score	BMI-Z	Yes	0.96 (1.18)	273	NA	NA
	BMI (kg m ⁻²)		No	NA	NA	32.8 (8.9)	494
	Total fat mass (g)	TFM	No	21,563 (10,670)	148	32,161 (11,391)	356
	Total percentage fat	TPF	No	36.0 (9.3)	148	41.21 (8.98)	356
	Total lean mass excluding bone mass (g)	TLM	No	34,712 (10,555)	148	42,942 (10,577)	356
	Total percentage lean excluding bone mass	TPL	No	60.8 (8.9)	148	55.97 (8.66)	356
	Trunk fat (g)	TrF	No	9,475 (5,340)	148	15,964 (5,768)	356
	Visceral fat area (cm ²)	VFA	No	73 (38)	148	199 (80)	356
Clinical classifications	Hypertension (AAP 2017 paediatric hypertension guidelines) (%)		No	7 (2.6%)	273	NA	NA
	Hypertension (SBP ≥ 130 mm Hg, DBP ≥ 80 mm Hg or taking antihypertensive medications) (%)		No	2 (0.70%)	273	232 (45.3%)	494
	Low HDL (<50 mg dl ⁻¹ females, <40 mg dl ⁻¹ males) (%)		No	105 (38.5%)	273	218 (44.4%)	494
	High triglycerides (>150 mg dl ⁻¹ or on lipid-lowering medication) (%)		No	35 (12.8%)	273	248 (50.4%)	494
	Obesity (BMI percentile > 95 th children, BMI > 30 kg m ⁻² adults) (%)		No	95 (34.8%)	273	269 (54.1%)	494
	Diabetes (diagnosed T2D, on diabetes medication, FBG > 126 mg dl ⁻¹ or HbA1c ≥ 6.5%) (%)		No	1 (0.4%)	273	207 (39.8%)	494

Traits analysed and clinical categorizations are listed as either mean (s.d.) or n (%). n_{ind} , number of distinct individuals; NA, not applicable; 95th, 95th percentile.

Generating a paediatric proteomic signature of CKMD

The biological plausibility and concordance of many proteome–CKMD phenome associations across adults and children are consistent with an early nascent phase of CKMD; still, identifying clear links to clinical risk factors or outcomes remains difficult, given the extended time horizon from childhood needed to capture clinical events. Therefore, we next constructed multi-protein-based signatures of CKMD phenotypes in children to test their relevance for clinical CKMD end-points in adults in multiple populations. We used principal components analysis

(PCA) to summarize 16 phenotypes measured in both children and adults, retaining 6 varimax-rotated, orthogonal latent factors (LFs) explaining ≈77% of the overall variation in the measured paediatric phenome (Supplementary Fig. 2). For interpretability, we labelled LFs by clinical interpretation of features with high loadings, recognizing that each LF was loaded heterogeneously on CKMD phenotypes (Fig. 4 and loadings in Supplementary Table 4): (1) pro-inflammatory adiposity (LF1: primarily loaded on pro-atherogenic lipids, homeostatic model of insulin resistance (HOMA-IR) and body mass index (BMI Z score));

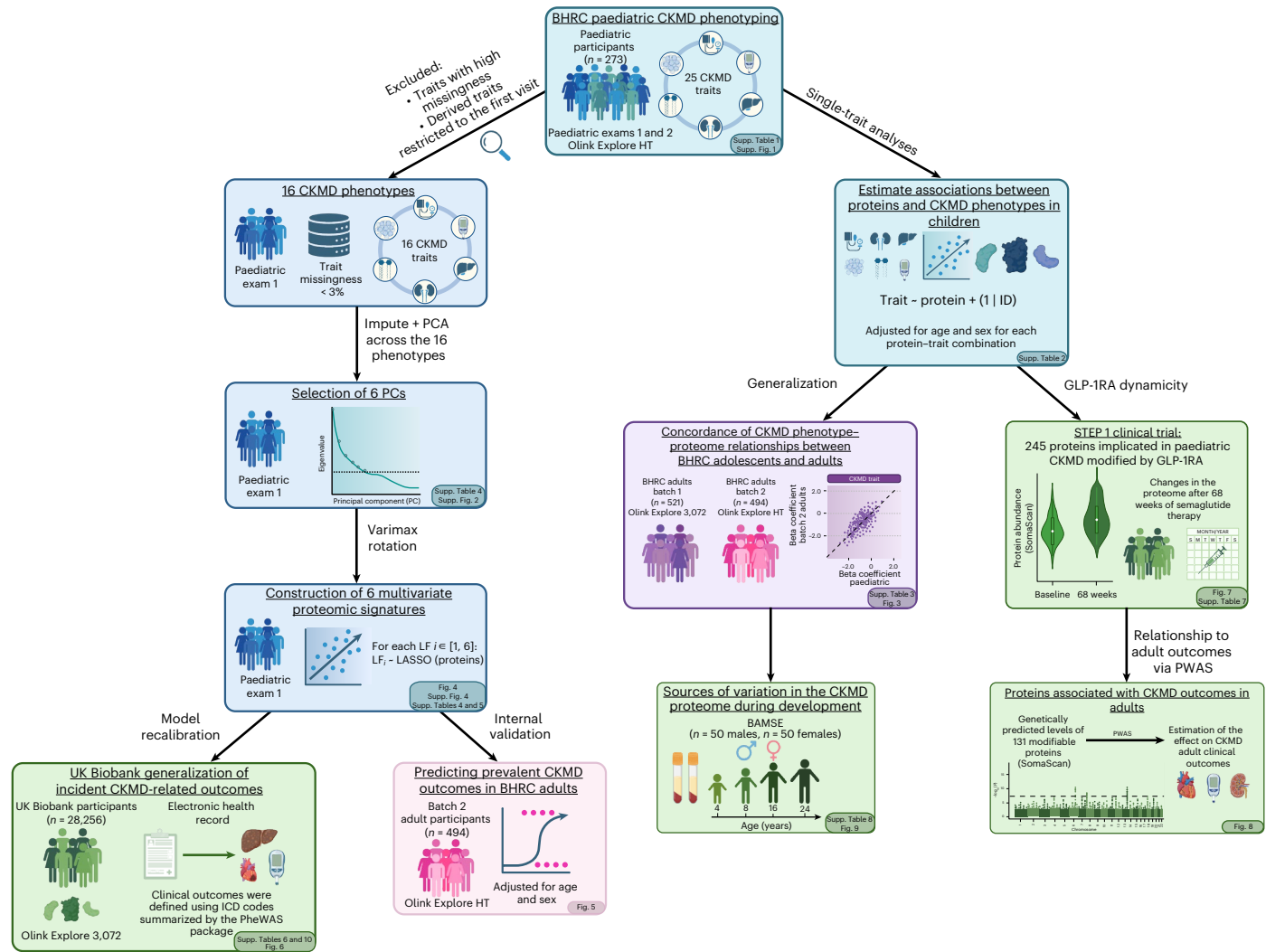


Fig. 1 | Study design. Flow of data and analytical methods in the study. Figure created in BioRender; Frankel, E. <https://biorender.com/aqf2uk2> (2026).

(2) liver fat and fibrosis (LF2: loaded on transaminases and aspartate aminotransferase to platelet ratio index (APRI)); (3) cholesterol (LF3: loaded on low-density lipoprotein (LDL) and total cholesterol (TC)); (4) blood pressure (LF4: loaded on systolic blood pressure (SBP) and diastolic blood pressure (DBP)); (5) kidney function (LF5: loaded on estimated glomerular filtration rate (eGFR) and liver fibrosis); and (6) insulin resistance (LF6: loaded on glycated haemoglobin (HbA1c), HOMA-IR and fasting glucose). The clinical interpretation of 'higher' LF values varied depending on loading directionality, with higher values for all LFs but LF3 representing a more adverse CKMD state. We generated proteomic signatures of each LF in least absolute shrinkage and selection operator (LASSO) regression to serve as molecular scores of the six summary traits. Our model coefficients are in Supplementary Table 5; the mean cross-validated R^2 for each paediatric model and number of proteins per model, k , for each LF are as follows: LF1: $R^2 = 0.485$, $k = 154$; LF2: $R^2 = 0.560$, $k = 89$; LF3: $R^2 = 0.256$, $k = 70$; LF4: $R^2 = 0.072$, $k = 20$; LF5: $R^2 = 0.359$, $k = 52$; and LF6: $R^2 = 0.142$, $k = 63$. Coefficients for the top 50 model coefficients by feature importance (or fewer, for smaller models) for each LF-based model are shown in Supplementary Fig. 3.

Paediatric CKMD proteomic signatures identify adults with susceptibility to clinical CKMD

We next examined whether proteomic signatures of paediatric CKMD would be associated with CKMD in adults. We first conducted a cross-sectional analysis in 494 Hispanic or Latino adults from the

same community, applying the paediatric proteomic LF scores to adult proteomic data and testing associations with CKMD-associated measures. We found that proteomic signatures, derived in children and adolescents, were associated with analogous traits in corresponding LFs in adults, suggesting that paediatric proteomic signatures may be related to comparable CKMD profiles across paediatric and adult populations (Extended Data Fig. 3). Next, we estimated associations between these paediatric CKMD signatures and incident clinical CKMD in adults in the UK Biobank (characteristics in Extended Data Table 1; median follow-up 13.7 years, 25th–75th percentile 13.0–14.5 years for mortality). Pro-inflammatory adiposity (LF1), liver fat and fibrosis (LF2) and insulin resistance (LF6) (for which higher values in childhood signified more advanced paediatric CKMD) were associated with higher risk of all-cause mortality, T2D, multiple cardiovascular diseases, sleep apnoea and fatty liver disease (Fig. 5). Blood pressure (LF4) and kidney function (LF5)—also associated with more advanced paediatric CKMD—were linked to higher T2D risk, with heterogeneous associations with other conditions. By contrast, cholesterol (LF3), for which higher values in childhood were less adverse, had protective associations with cardiovascular and metabolic conditions in adults. Most of our observed associations were robust to comprehensive covariate adjustment in adults (Supplementary Table 6). Collectively, these findings underscore the potential consistency of early CKMD risk, as reflected in the childhood proteome, with the diagnosis of adult cardiovascular and metabolic disease.

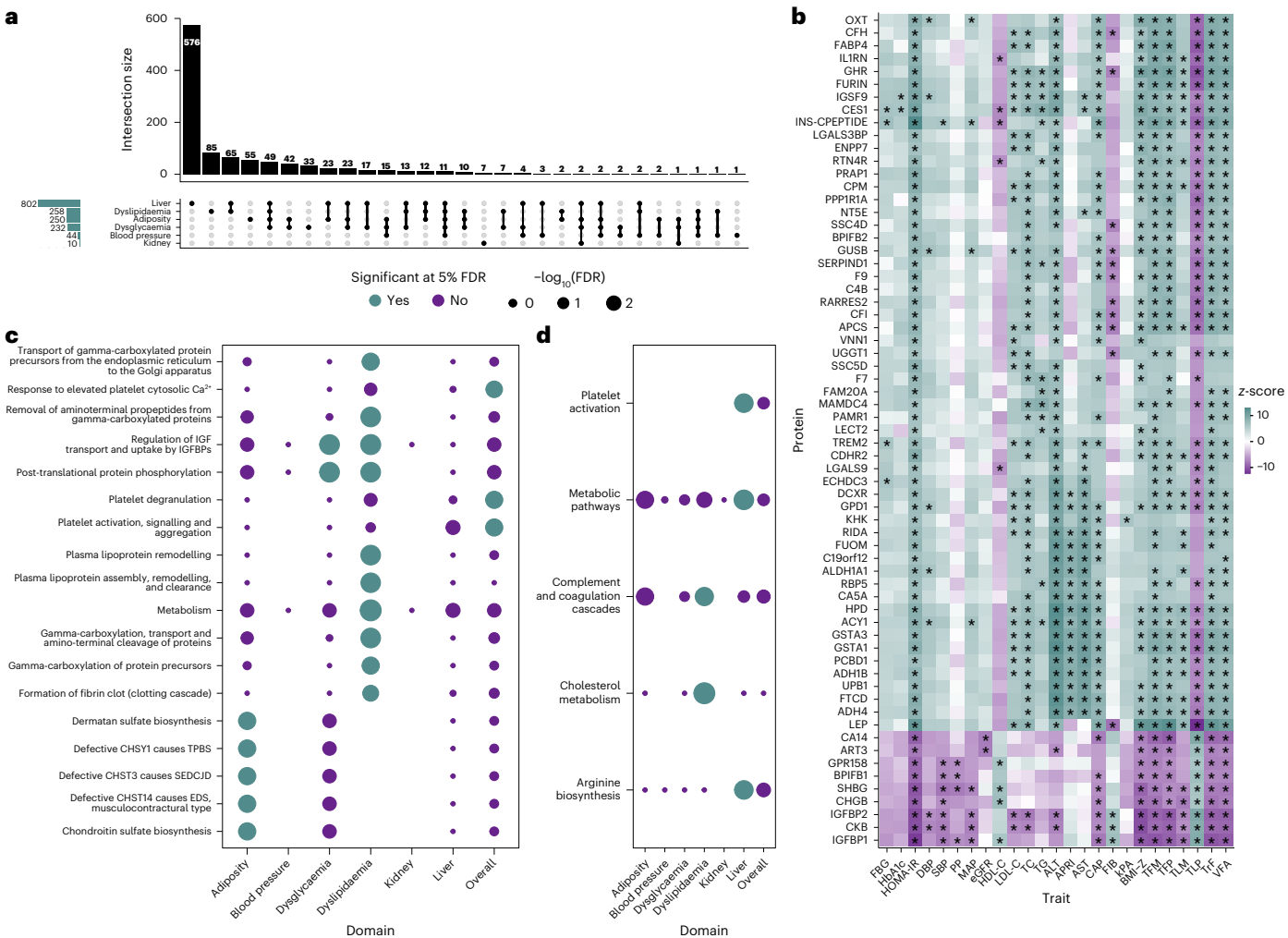


Fig. 2 | Proteomic architecture of CKMD in children. a, Significant proteins across 6 domains (all at the 5% FDR), showing shared association. **b**, Regression z-scores in linear mixed models (adjusted for age and sex) for proteins significant in at least four of the six phenotype domains. The asterisk denotes significance at 5% FDR. **c,d**, Reactome (**c**) and KEGG (**d**) pathway analyses. Full data for single-trait paediatric protein associations (**a** and **b**) are available in

Supplementary Table 2a. Full data for **c** and **d** are available in Supplementary Table 9. CHSY1, chondroitin sulfate synthase 1; TPBS, Temtamy preaxial brachydactyly syndrome; CHST3, carbohydrate sulfotransferase 3; SEDCJD, spondyloepiphyseal dysplasia with congenital joint dislocations; CHST14, carbohydrate sulfotransferase 14; EDS, Ehlers–Danlos syndromes (EDS).

Modifiability of proteins associated with paediatric CKMD traits during GLP-1RA therapy in adults

Beyond statistical associations with CKMD, we expect the most clinically relevant biomarkers to change in response to disease-modifying therapy. To explore the potential modifiability of protein levels implicated in paediatric CKMD risk in our previous analyses, we mapped 1,032 proteins associated with any CKMD outcome to data from adult participants in the STEP 1 clinical trials of semaglutide (Methods)¹⁵. Protein levels associated with more adverse CKMD in children and adolescents were altered in a favourable direction following semaglutide therapy in the STEP 1 sample (Fisher’s exact test $P < 2.2 \times 10^{-16}$; Fig. 6a and Supplementary Table 7). These findings may be mediated through a variety of systemic changes, including weight loss. Notable examples included LEP, FABP4, CES1 and ACY1, which are all associated with greater paediatric insulin resistance, adiposity and liver fat and fibrosis, and having plausible CKMD-relevant mechanisms in adults (all reduced among adults on GLP-1RA). Conversely, SHBG, WFIKN2 (ref. 27) and IGFBP2 represent examples of proteins with generally opposing effects relative to LEP in the paediatric sample and increased after GLP-1RA treatment in adults. We also identified proteins associated with early CKMD in our paediatric cohort (for example, BCAN, ART3) that have not

previously been implicated in adult T2D or CKMD. A heat map of the top 10 proteins (ranked by effect size) for each phenotype in the paediatric cohort is shown in Fig. 6b. Proteins implicated by the CKMD phenotypes in children that showed evidence of changes in concentration with GLP-1RA therapy in adults appeared distributed across key tissues and cell types, with strong representation of hepatocytes, pancreatic cells, adipose tissue and brain tissue (Supplementary Fig. 4). While we cannot disentangle the relative contributions of GLP-1RA-mediated weight loss versus other metabolic effects on individual CKMD-related proteins, these findings suggest that many proteomic signatures associated with paediatric CKMD risk are favourably modifiable with modern CKMD-targeted therapy in adults.

Genetic liability of the dynamic paediatric CKMD proteome is linked to lifetime CKMD

We next studied whether genetic liability of the CKMD proteome is linked to key CKMD phenotypes in adults, a window into lifelong CKMD susceptibility associated with a proteome identified in childhood. We performed a PWAS using large published genome-wide association studies (GWAS) of heart failure, BMI, T2D, renal function, myocardial infarction and metabolic-associated steatotic liver disease (MASLD)

Table 2 | Selected functional characterization of proteins associated with multiple domains

Protein	Domains associated with protein in paediatric cohort	Direction	Putative function and previous studies
C-peptide	Dysglycaemia , blood pressure, liver, adiposity	+	Marker of beta-cell function and reserve, with higher levels suggesting hyperinsulinaemia; increased in children with impaired glucose tolerance after glycaemic challenge ¹⁰
Immunoglobulin superfamily member 9 (IGSF9)	Dysglycaemia, blood pressure, dyslipidaemia, liver , adiposity	+	IGSF9 involved in cell adhesion; related to hepatic steatosis and fibrosis in adults with HIV ⁹⁹
Glycerol-3-phosphate dehydrogenase 1 (GPD1)	Dysglycaemia, blood pressure, dyslipidaemia, liver , adiposity	+	GPD1, involved in carbohydrate metabolism in the liver; mixed mechanistic findings, with congenital deficiency linked to hypertriglyceridaemia ⁹⁰ and liver disease ⁹¹ , while expression increased in adipose tissue in adults with obesity and is associated with BMI ^{24,25} ; expression responsive to weight loss intervention ⁹²
Aminoacylase 1 (ACY1)	Dysglycaemia, blood pressure, dyslipidaemia, liver , adiposity	+	ACY1; high liver expression, levels in adults related to diabetes and hepatic steatosis ^{23,27,28}
Chromogranin B (CHGB)	Dysglycaemia , blood pressure, dyslipidaemia, liver, adiposity	–	CHGB; involved in secretion of catecholamines (nervous system) and in insulin granule trafficking ¹⁸ (pancreatic beta-cells); genetic studies of <i>CHGB</i> variation and <i>Chgb</i> -null mice suggest lower levels related to higher blood pressure ¹⁹ ; previous association with metabolic traits in adults ²⁰
Aldehyde dehydrogenase 1 family member A1 (ALDH1A1)	Liver	+	Aldehyde dehydrogenase; important in metabolic processing of aldehydes in the liver; expression in paediatric livers lower than adults ⁹³ ; increased expression may protect against hepatic oxidative stress and inflammation during liver injury ²⁶
Creatine kinase B (CKB)	Dysglycaemia, blood pressure, dyslipidaemia, liver, adiposity	–	CKB; involved in high-energy phosphate transfer in highly metabolic tissues; increase in endoplasmic reticular stress during obesity epigenetically reduces <i>CKB</i> expression, increasing inflammatory chemokine expression in fat ²⁹ ; reduction in CKB in mice impairs thermogenesis and increases obesity ³⁰
Beta-glucuronidase (GUSB)	Dysglycaemia , blood pressure, dyslipidaemia, liver, adiposity	+	GUSB; previous studies in adults link circulating levels to diabetes ²¹
Insulin-like growth-factor binding protein (IGFBP1 and IGFBP2)	Dysglycaemia , blood pressure, dyslipidaemia, liver, adiposity	–	Insulin-like growth-factor binding proteins; linked to lower risk of hepatic steatosis, obesity, atherosclerotic heart disease and insulin resistance ^{22,23} ; previous evidence in children suggests IGFBP1 is an early marker of insulin resistance ³¹
Sex hormone-binding globulin (SHBG)	Dysglycaemia , blood pressure, dyslipidaemia, liver , adiposity	–	SHBG; involved in sex hormone metabolism; early reduction in SHBG related to faster decline in insulin sensitivity independently of weight ³² ; levels are lower in children with obesity, reversible with weight loss, with higher levels linked to hepatic steatosis and widespread metabolic dysfunction ³³
β-site amyloid precursor protein cleaving enzyme 1 (BACE-1)	Dysglycaemia , adiposity	–	Overexpression can lead to increased hepatic insulin resistance in murine models, potentially via BACE-1-mediated cleavage of the insulin receptor; pharmacologic targeting of BACE-1 can abrogate these effects ^{37,38}
G-protein coupled receptor 158 (GPR158)	Dysglycaemia , blood pressure, dyslipidaemia, adiposity	–	Involved in CNS-based regulation of energy metabolism and nutrient intake ³⁹ ; variants in GPR158 may impact liability to weight gain ⁴⁰
Guanylate binding protein 1 (GBP1)	Liver	+	Involved in pyroptosis in inflammatory liver injury ⁹⁴
Fibroblast growth-factor binding protein 3 (FGFBP3)	Dyslipidaemia	+	FGFBP3-null mice have lower lipid and TG levels, and supplementation of FGFBP3 in mice with obesity can ameliorate metabolic dysfunction, weight gain and liver steatosis; mechanism via interaction with FGFs ⁴¹
Ubiquitin specific peptidase 28 (USP28)	Liver	+	High expression in human metabolic-associated steatotic liver disease and steatohepatitis; functions via stabilizing PPAR-gamma via blocking ubiquitination; lack of USP28 may lead to reduction in PPAR-gamma signalling and reduced hepatic inflammation ⁴²
Cysteine-rich with EGF-like domains 2 (CRELD2)	Liver	–	May limit ER stress and resilience; lack of CRELD2 may augment unfolded protein response and contribute to MASLD and changes in energy expenditure; may even have sex-specific effects ⁴³
Fibroblast growth-factor receptor 4 (FGFR4)	Dysglycaemia , adiposity	–	Divergent results; FGFR4 modulates insulin sensitivity in murine obesity models ^{95,96}
Cathepsin B (CTSB)	Dysglycaemia, adiposity	–	Lysosomal protease; increased in white adipose tissue in mice; may regulate perilipin-1, which regulates lipolysis in fat ³⁷
Transmembrane protein 108 (TMEM108)	Adiposity	–	Traditionally implicated in psychiatric disorders; TMEM108 mutant mice have poorer insulin sensitivity and shift metabolism towards glucose (away from lipid); metformin can rescue this effect ⁹⁸

This list includes proteins related to five domains in the BHRC paediatric cohort (ACY1, ALDH1A1, CHGB, CKB, GPD1, GUSB, IGFBP1, IGFBP2, IGSF9, C-peptide and SHBG) as well as additional selected proteins related to one or more domains with previous evidence in CKMD. Domains in bold reflect those with previous functional evidence (previous studies provided). +, higher protein abundance is more adverse; –, lower protein abundance is more adverse. CNS, central nervous system; ER, endoplasmic reticular; FGFs, fibroblast growth factors; PPAR, Peroxisome proliferator-activated receptor.

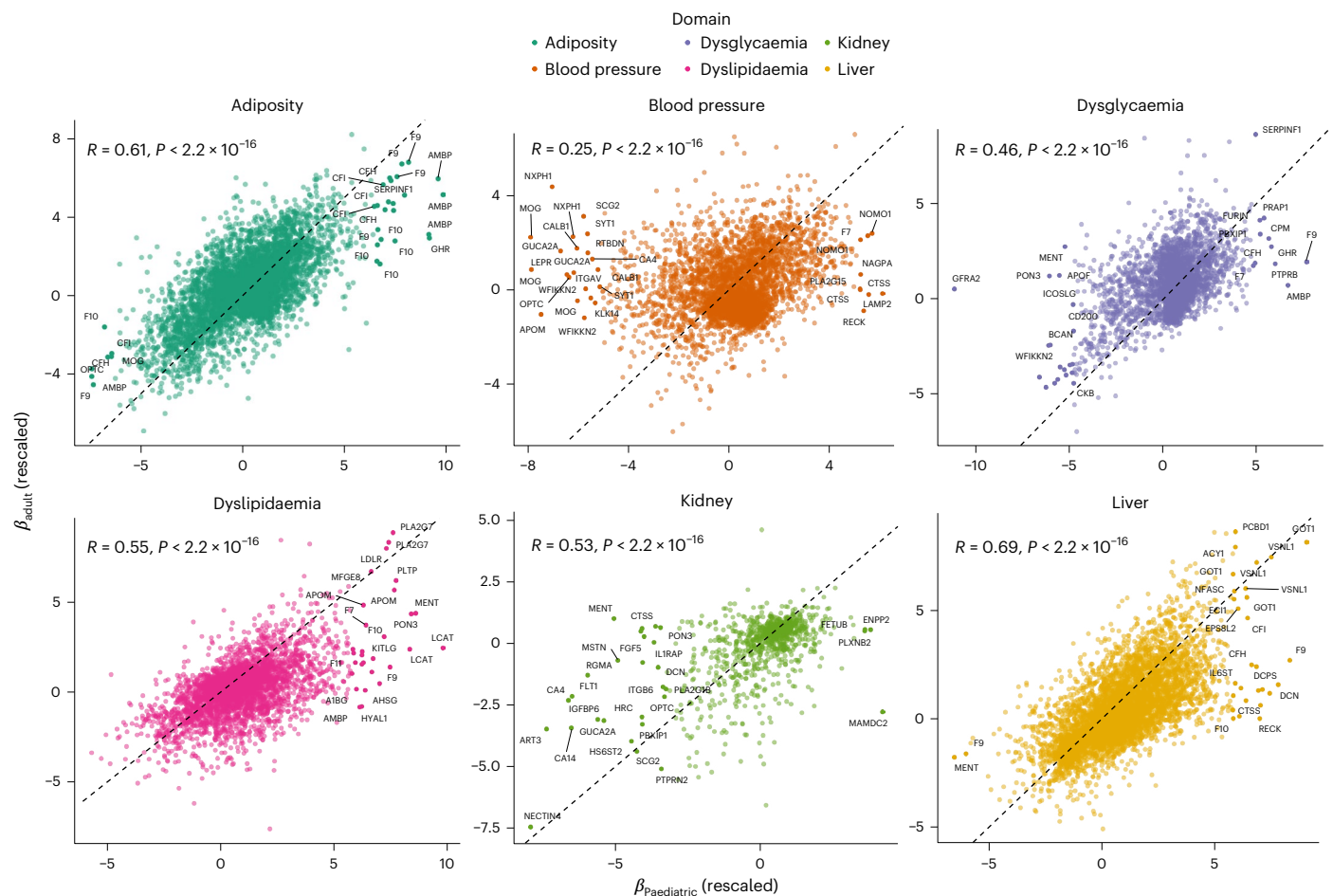


Fig. 3 | Concordance of CKMD phenotype–proteome relationships between adults and youth. This figure depicts the correlation of paediatric and adult regression model effect sizes. The x axis shows rescaled beta coefficients for paediatric results, and the y axis shows rescaled beta coefficients for adult

batch2 results. Protein labelling is arbitrary. Each protein–trait association represents a distinct single-trait model. Pearson correlations were two-sided and not corrected for multiple testing. The full data are available in Supplementary Tables 2a and 3.

spanning up to 2.5 million individuals in a single trait (see Methods for sample sizes). This approach is aimed at using genetic models of circulating proteins linked to CKMD in childhood to quantify the relation with adult CKMD outcomes. To restrict our analyses to proteins potentially dynamic and linked to CKMD, we limited our PWAS analysis to 131 proteins that (1) were associated with paediatric CKMD phenotypes (q -value < 0.05), (2) had evidence of changing after GLP-1RA therapy in adults (q -value < 0.05) and (3) had models of genetically predicted expression.

We observed strong associations with multiple CKMD outcomes (Extended Data Fig. 4), including targets with cross-system and canonical physiologic roles with potential therapeutic relevance. We observed paediatric CKMD targets that reflect BMI-specified mechanisms of energy homeostasis (for example, *SLIT2*, previously implicated in beige fat thermogenesis improvement of glucose sensitivity⁴⁶), appetite (for example, *SCG3*, previously implicated in food intake regulation⁴⁷), and adipose tissue inflammation and adipogenesis (*SERPINE1*/*PAI-1* (ref. 48), *COMP*⁴⁹). Paediatric CKMD-related proteins also implicated pathways across multiple CKMD mechanisms and outcomes, including activin and TGF-beta signalling (*INHBC*^{50,51}), matrix remodelling (*TIMP4*), oxidative homeostasis (*GSTA1*), autophagy (*SPON2* (ref. 52)) and lipoprotein metabolism (*APOH*, implicated in lipoprotein(a) metabolism⁵³). Of note, *NCAN* was strongly related to both MASLD and T2D, a finding supported by previous studies in MASLD in adult populations⁵⁴ but not widely in T2D. In addition to known targets, this

proteogenomic approach also identified targets not previously widely reported (for example, hyaluronan catabolism with potential roles in T2D and CKMD⁵⁵). Collectively, these results provide genetic support for a consistent set of CKMD-related proteins identified by discovery in a paediatric population and prognostic for CKMD outcomes in adults.

Early developmental variation in a shared paediatric–adult CKMD proteome

The variability of CKMD-linked proteins during the adolescent–adult–hood transition, and whether they reflect developmental age- or sex-related changes, are critical to positioning their use as potential biomarkers of early CKMD. To address this, we leveraged a recently reported longitudinal study of 100 individuals from age 4 to 24 years in the Swedish BAMSE cohort¹⁷ to examine sources of variability in protein abundance. Focusing on proteins with evidence of significance (q -value < 0.05) in children that replicated in BHRC adults, we observed broad variability in the CKMD-associated proteome (Supplementary Fig. 5 for four select domains for visualization purposes; the full results are in Supplementary Table 8 (ref. 17)). Overall, we observed a high degree of variation unexplained by age or sex for a large fraction of proteins associated with CKMD in children and adults (example trajectories for key metabolic proteins *LEP*, *IGFBP2*, *ACY1* and *CES1* in Supplementary Fig. 5). Age or sex explained a relatively low proportion of variability for our traits with the notable exception of BMI-Z, which reflects growth and development. Of note, many

LF loadings

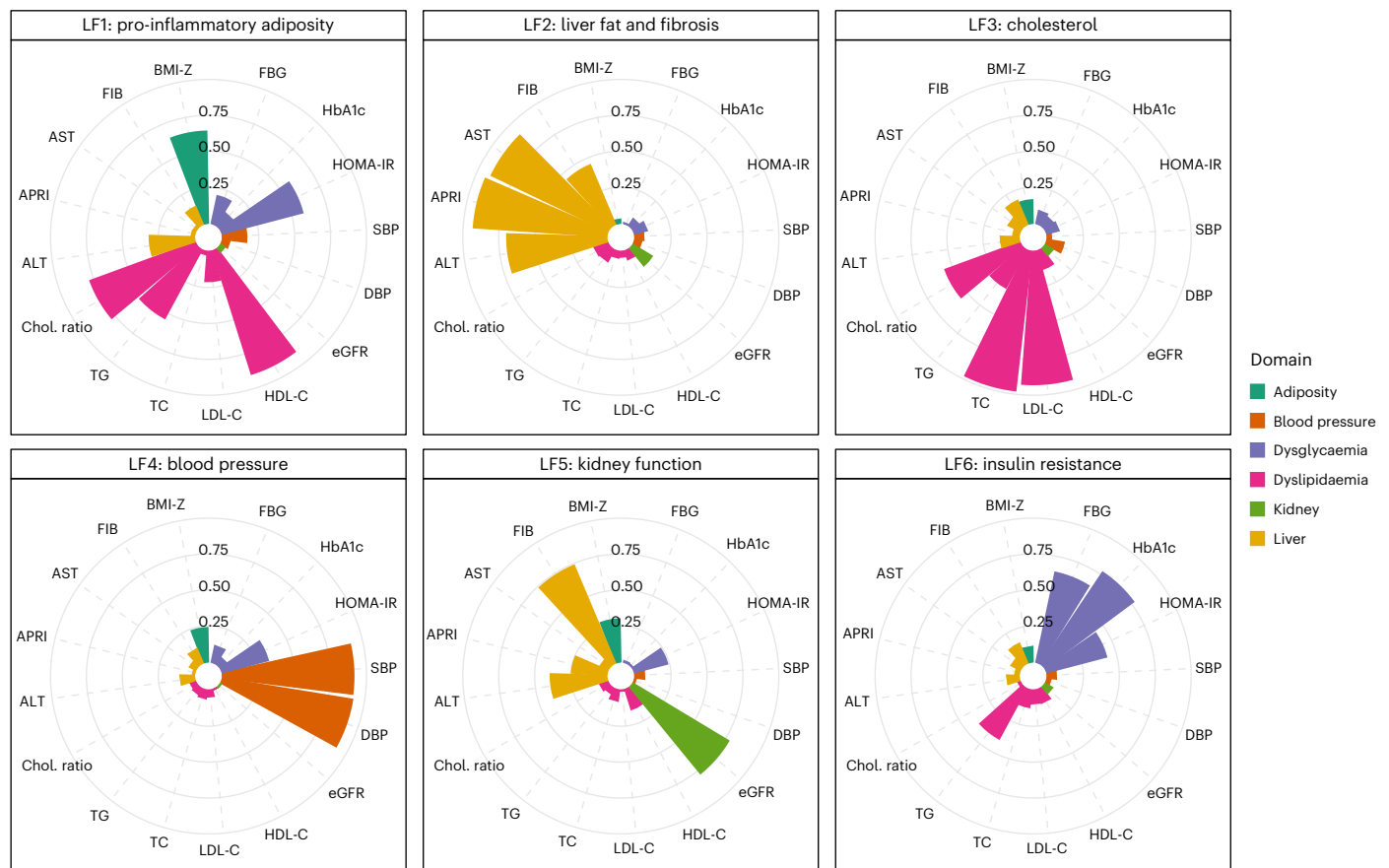


Fig. 4 | Varimax-rotated principal components (LFs) of the paediatric CKMD phenotype. These circular bar plots indicate the magnitude of the loadings per trait for each LF of the paediatric phenotype matrix. The bars are coloured by the domain of each trait. Abbreviations are as in Table 1. Chol. ratio, ratio of TC to HDL-C.

proteins that showed a substantial variability not explained by age or sex changed with GLP1-RA therapy in adults (for example, IGFBP1, ILIRN, TNFSF14). This substantial, potentially modifiable residual variability in the CKMD-relevant proteome suggests potential targets for intervention to improve CKMD outcomes.

Discussion

Climbing rates of CKMD-related morbidity over the past two decades have necessitated increasingly early screening and preventive measures to address population burden. Over the past four decades, evidence from early coronary artery pathology reports¹, studies in Bogalusa⁵⁶ and recent data in ≈36,000 children from the i3C Consortium⁵ have implicated obesity and related CKMD phenotypes (hypertension, pro-atherogenic dyslipidaemia, insulin resistance) as drivers of premature disease. In response, pharmacologic approaches⁵⁷ and lifestyle interventions (for example, nutrition, physical activity) are increasingly emphasized in childhood to mitigate early CKMD risk. CKMD screening in children is triggered by the presence of obesity, which often misses at-risk children⁵⁸. Furthermore, clinical intervention thresholds for blood pressure and lipids in children are less stringent than those used in adult care^{3,5}, and standard screening tests (for example, fasting glucose) often miss a substantial amount of disease-defining risk in children^{9,10}. Given that at-risk children generally do not suffer from clinical cardiovascular events, defining precise biomarkers in childhood that reflect relevant early CKMD traits and have prognostic and genetic links to CKMD outcomes in adults is critical to deploying surveillance and treatment strategies to prevent CKMD at its origin⁵⁷.

Here, using complementary analyses across the paediatric CKMD phenotype, circulating proteome and genome, we derived

proteomic signatures of CKMD-related traits in children that showed notable concordance with an adult CKMD proteome. Composite, multi-protein signatures of CKMD risk in children were strongly associated with cross-sectional and incident CKMD clinical outcomes over a decade in ≈28,000 adults. The childhood CKMD proteome implicated both known and potentially novel biological mechanisms in CKMD. Proteins linked to CKMD traits in children showed (1) altered abundance following GLP1-RA therapy and (2) a potentially causal effect on adult CKMD outcomes (spanning cardiovascular, renal and metabolic diseases). Finally, proteins linked to CKMD in adults and children generally showed substantial variation not explained by age or sex in childhood through early adulthood, highlighting potentially modifiable components of early metabolic risk. These results are consistent with the presence of proteomic signals linked to adult CKMD beginning early in life, highlighting opportunities for use of the childhood proteome in enhanced precision risk assessment starting earlier in life.

Although overt clinical disease in childhood remains rare, classical risk factors for CKMD in adults also appear to promote early CKMD phenotypes in children: for example, reduced physical activity and low physical fitness levels promote insulin resistance⁵⁹ and sub-clinical atherosclerosis in children⁶⁰. Paediatric CKMD is also related to early myocardial tissue changes that presage heart failure⁶¹ as well as pro-atherogenic dyslipidaemia and inflammatory biochemical states ultimately linked to adverse outcomes in adulthood⁵. Given the limitations of standard clinical tests to capture this metabolic risk early (for example, fasting labs), there is growing recognition that molecular biomarkers—‘omics’—can potentially detect risk decades before disease onset and often independent of established clinical risk factors^{14,62–66}.

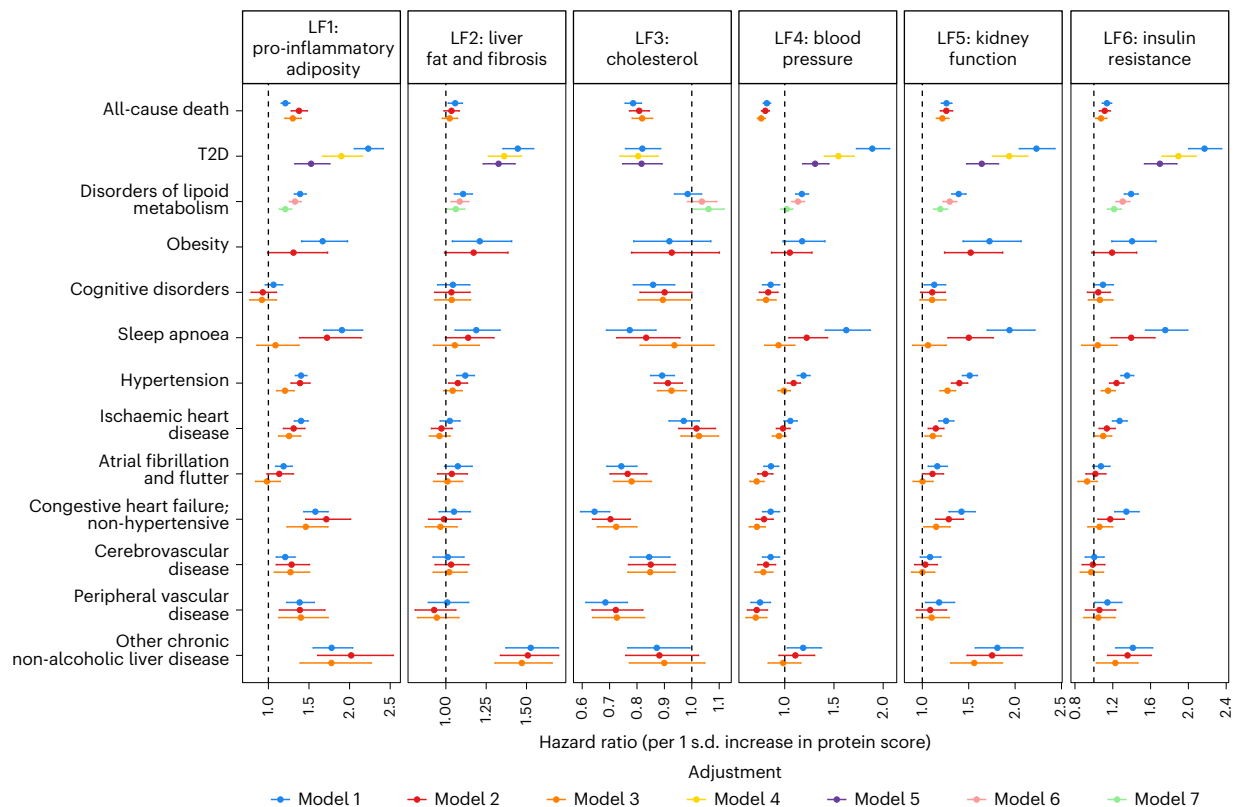


Fig. 5 | Long-term implication of the paediatric CKMD proteome on adult clinical risk. The forest plot depicts the hazard ratio (Cox models) of childhood-derived proteomic scores for CKMD-related end-points in the UK Biobank. The colour of the marks in the plot indicates for which variables each model was adjusted. The error bars represent 95% confidence intervals of the hazard

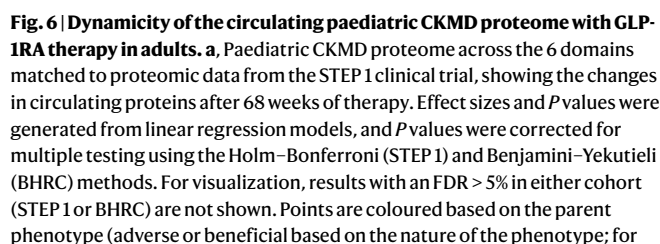
ratio. The models control for covariates as follows: model 1: age, age², sex, ethnicity, smoking, alcohol and Townsend deprivation index; model 2: model 1 + HbA1c + TG + HDL; model 3: model 2 + BMI; model 4: model 1 + TG + HDL; model 5: model 1 + BMI + TG + HDL; model 6: model 1 + HbA1c; and model 7: model 1 + BMI + HbA1c.

Our study contributes to emerging, considerable interest in addressing paediatric CKMD with findings that are broadly consistent with multiple published studies in CKMD (Extended Data Table 2). For example, in one of the largest studies to date in paediatric CKMD, 149 proteins (Olink) were quantified in 4,024 children in a cross-sectional analysis and 200 children during a 1-year obesity intervention⁶⁷ (HOLBAEK study). In addition to identifying similar proteins in cross-sectional association with obesity status as in our study, proteins related to CKMD traits in our study showed changes in over 1 year that were associated with changes in CKMD traits in the HOLBAEK study, suggesting important phenome–proteome dynamics of our findings. Furthermore, seminal work from the same group linking mass spectrometry proteomics in children to human genetics⁶⁸ showed that circulating protein quantitative trait loci in children were linked to adult CKMD (analogous to our approach using adult proteogenomic models for proteins identified in paediatric CKMD).

Our results fundamentally extend previous findings via several key innovations: (1) we surveyed a broader proteomic space than previously quantified (Extended Data Table 2); (2) we focused on a high-risk paediatric population, with a high prevalence of obesity; (3) we included direct, imaging-based measures of MASLD, an early mediator of CKMD risk; and (4) we benchmarked paediatric proteomic signatures against adult outcomes and phenotypes, including adults from the same community as well as from a geographically and ethnically distinct population. Of note, the multisystem CKMD outcomes were most strongly associated with proteomic signatures of hepatic disease, a condition that affects nearly 10% of children⁶⁹ and may serve as a harbinger of adult CKMD⁷⁰. Moreover, we observed strong concordance in phenome–proteome relationships between youth and adults. Several proteins linked to paediatric CKMD mapped to known

adult CKMD-relevant pathways across age groups, including beta-cell function and inflammation, as well as new molecular mechanisms of early CKMD biology. Finally, changes in protein abundance under GLP-1RA therapy, genetic associations with CKMD outcomes across the life-course and proteomic variation at the adolescent–adulthood transition not explained by age and sex provide strong support from external sources that these proteins mark phenotypic states in children that are biologically and therapeutically relevant.

The progression of cardiometabolic risk from childhood to adulthood has important clinical implications. With increasing pharmacological and surgical efforts to combat paediatric obesity (nearly 600% increase in paediatric GLP-1RA prescriptions between 2020 and 2023)^{71,72}, a precision-based framework to prioritize youth at substantial CKMD risk to receive therapies will be critical to balance benefits with therapeutic and financial burden and clinical risk. Our identification of proteins in childhood CKMD that are predictive of adult CKMD and are associated with observed improvement via the GLP-1RA therapy in adults strongly supports the potential for precision proteomic risk assessments in paediatric CKMD. Nevertheless, our study has several limitations. The primary studied population was derived from a single population in South Texas. While consistent findings in the UK Biobank offer external validity, additional research in populations of diverse ancestries and socio-economic and lifestyle exposures is needed to extend generalizability and identify novel exposures and confounders relevant to proteomic prediction. In addition, our paediatric sample did not have serial sampling with power to detect dynamicity in the paediatric proteome over time, although we plan to continue to follow our cohort to accomplish this goal. Our results should also be viewed in the context of differences in proteomic platforms across data sources that may introduce variability, issues



example, lean body mass is designated 'beneficial', whereas BMI is designated 'adverse'). Generally, we observed that proteins with adverse relations to CKMD phenotypes were decreased with GLP-1RA therapy and proteins with beneficial relations were increased. **b**, Visualization of the top 10 proteins (ranked by effect size) related to individual CKMD phenotypes in the paediatric cohort (with several proteins overlapping among phenotypes). Proteins that are positively associated with phenotypes (for example, BMI, HOMA-IR) were generally reduced after GLP-1RA therapy. The full data are available in Supplementary Table 7. Pairwise protein correlations are included in Supplementary Fig. 1.

In summary, we characterized the proteomic architecture of CKMD in children living at the US–Mexico border across a range of cardiometabolic phenotypes, demonstrating similarities between the CKMD phenotype–proteome relation across age, its biological plausibility, relevance to long-term clinical CKMD risk in adults through large cohort observational studies and human genetics, and mutability with GLP-1RA therapy. Our results support the hypothesis that the molecular profile of CKMD appears early in the life-course, with proteomic signals relevant to long-term clinical outcomes appearing even in childhood. In total, these observations underscore the potential for precision proteomic risk assessment in children to identify at-risk groups who may benefit from earlier preventive strategies aimed at improving CKMD-free longevity. Further work in large paediatric populations to understand modifiability is critical to clinical deployment of precision strategies to detect and forestall cardiometabolic complications that originate in childhood.

Paediatric and parental or guardian assent and consent were obtained according to the approved institutional review board protocol (HSC-SPH-14-0141, UT Health Houston School of Public Health, Brownsville), including the use of optional bilingual instruments. Written informed consent was obtained from all participants under the approved institutional review board protocol.

Adults from the same community aged ≥ 18 are enrolled and followed longitudinally every 5 years. Of the 4,956 adult BHRC participants, 887 exams from 686 individuals had proteomics available and were included in analyses validating associations in adults; 494 participants were included for testing the fit of paediatric models to adult phenotypes. No participants included in the paediatric analyses were included in the internal adult replication; however, there were 46 first-degree relatives (parents or siblings).

Descriptive statistics for each trait and outcome are reported in Table 1, with additional details available in Supplementary Table 1. Participants

were examined in the Clinical Research Unit in Brownsville, Texas. Children were accompanied by a parent or guardian. Venipuncture was performed for fasting glucose, fasting insulin, HbA1c, lipid profile (HDL-C, TC and TG), creatinine, and alanine and aspartate aminotransferase (ALT, AST). From these measures, we derived HOMA-IR, calculated LDL-C, TG to HDL ratio, eGFR⁷³, APRI and fibrosis-4 (FIB-4) index. Resting SBP and DBP were measured three times, with the average of the last two measures used, from which mean arterial pressure (MAP) and pulse pressure (PP) were calculated. Non-invasive ultrasound imaging of the liver using FibroScan (Echosens) was used to measure liver stiffness (kPa) and controlled attenuation parameter (CAP). Participants underwent whole-body dual energy X-ray absorptiometry for body composition via standard clinical protocols (Hologic) to derive the total and percentage fat mass, total and percentage lean mass (excluding bone mass), trunk fat and visceral fat area (cm²). Weight and height were measured by study staff for BMI. Age- and sex-specific BMI-Zs for children and adolescents were calculated based on the Centers for Disease Control and Prevention growth charts. Trait-specific quality control strategies are described in Supplementary Table 1.

Proteomics

We measured 5,420 circulating proteins (Olink Explore HT) in 394 frozen serum samples from children and adolescents; three samples were removed owing to poor proteomic sample quality. Normalized protein expression (NPX; protein abundance in log₂ units) was calculated as indicated by the manufacturer (version 6.7.2 Olink proprietary software). Methods for Olink-based proteome quantification have previously been described⁷⁴. There were no missing NPX data; although some values did fall below the limit of detection, they were included in analyses to optimize discovery potential. In adults, we used the same method and panel for 509 exams from 495 participants ('batch 2') and an earlier panel (Olink Explore 3072 with 2,940 proteins) for 378 exams in 191 participants ('batch 1'). Overlapping proteins were mapped across the two batches for *N*-weighted meta-analysis. The subset of 494 adult participants in batch 2 was used for comparisons of effect size.

Association of the circulating proteome with CKMD phenotypes in children and adults

We used linear mixed-effects regression to estimate associations between proteins and CKMD phenotypes in children and adolescents. HOMA-IR and TG levels were log₁₀- and log₂-transformed, respectively, to ensure normality. All other variables were assumed to meet the assumptions of the models. We modelled each protein as a function of a single phenotype and included a participant-level random effect to account for repeated measurement of phenotypes over exams. In our primary analyses, we adjusted for age and self-reported sex. As a sensitivity analysis, we additionally adjusted each model for BMI-Z (or adult BMI), except for those with outcomes relating to body composition (for example, dual energy X-ray absorptiometry measures). Trait-specific modelling strategies are described in Supplementary Table 1. Type 1 error was controlled using the Benjamini–Yekutieli false discovery rate (FDR) across all trait–protein comparisons. In adults, we applied the same modelling strategy (and as genetic data were available for these participants, we additionally controlled for the first three genetic principal components) to measure the association between each phenotype and protein in adults (using standard deviation-scaled model coefficients). Effect sizes were compared with the paediatric results via Pearson correlation. Pathway analyses were performed using proteins associated with phenotypes (*q*-value < 0.05 in children), with all proteins used in the regression analysis as the universe background on Kyoto Encyclopedia of Genes and Genomes (KEGG) and WikiPathways (*q*-value < 0.05 for enrichment).

To assess curated evidence from published GWAS, we classified the phenotypes into six domains: dysglycaemia (fasting glucose, HbA1c, HOMA-IR), body composition (BMI-Z, total fat mass (TFM),

total percentage fat (TPF), total lean mass excluding bone mass (TLM), total percentage lean mass excluding bone mass (TPM), trunk fat mass (TrF) and visceral fat area (VFA)), blood pressure (SBP, DBP, MAP, PP), dyslipidaemia (LDL-C, HDL-C, TC, TG, TG to HDL-C ratio), hepatic (ALT, AST, APRI, FIB-4 index, CAP, kPa) and renal (eGFR). We queried the top 30 proteins by effect size in the paediatric analyses associated with each of the six phenotype domains (*q*-value < 0.05; *n* = 245 unique proteins), along with proteins associated with at least one individual phenotype in children but not in adults (*n* = 109 proteins) in the Common Metabolic Diseases Knowledge Portal (CMDKP)⁷⁵. From the CMDKP, we used the HuGE scores for all phenotypes to summarize extant evidence of the corresponding gene's association with 22 CMDKP catalogued phenotype groups. Evidence was grouped into standard bins of compelling (HuGE score ≥ 350), extreme (≥ 100), very strong (≥ 30), strong (≥ 10), moderate (≥ 3), anecdotal (≥ 1) and no evidence (< 1) for visualization in heat maps⁷⁵.

Identifying a proteomic signature across CKMD phenotypes in childhood using machine learning approaches

We next investigated joint associations between the circulating proteome and 16 CKMD phenotypes across the 6 phenotype domains in 273 children with both proteomic and clinical data (Table 1). Our selected approach required a complete cross-sectional proteomic–phenotype dataset. Traits derived from linear combinations of other traits (for example, MAP and PP) were excluded from the PCA. Consequently, we excluded phenotypes analysed in the single-trait analyses that had over 50% missingness across the union of all participants in those analyses, leaving us with only 11 individuals with any missing data. We imputed missing phenotype data for those 11 individuals via random forest imputation (missForest). Before imputation, TG and HOMA-IR were log-transformed, all hepatic traits were inverse rank normalized, and all proteins and phenotypes were mean-centred and standardized.

We performed PCA across the 16 phenotypes, selected 6 principal components by scree plot evaluation and performed varimax post-rotation to generate LFs. Using each of the six LFs as a dependent variable and our proteomic data as the independent variables, we used L1-regularized regression with five-time repeated fivefold cross-validation (LASSO regression) to construct six parsimonious multivariable proteomic signatures. Model fit (*R*²) was evaluated by performance in cross-validation folds in caret.

Paediatric proteomic signatures of CKMD and long-term adult CKMD outcomes

We next measured the relationship between the LASSO-derived proteomic scores for children (corresponding to each of the six LFs) and adult cross-sectional and longitudinal phenotype outcomes. We estimated the proteomic scores for each multi-phenotype LF in adults from two samples: (1) BHRC adults (*n* = 494) and (2) adults from the UK Biobank (*n* = 28,256)⁷⁶. Given the differences in coverage across Olink platforms (Explore 3072 versus HT) between paediatric and adult samples (BHRC and UK Biobank used HT and Explore 3072, respectively), we 'recalibrated' our LASSO paediatric score for application across datasets, following published methods¹¹. In brief, we refit each of our six models for the subset of proteins available on the Explore 3072 platform using the paediatric proteomic scores for each phenotypic LF as the dependent variable and paediatric proteins common across samples as independent variables. We confirmed adequate performance of the recalibrated LASSO model within children (*R*² across LFs ranging from 0.72 to 0.98). Coefficients from the recalibrated LASSO-based LF proteomic models were applied to mean-centred and standardized protein abundance in the adult sample to generate adult LF proteomic scores for association analyses. LF proteomic scores were defined as the linear combination of the paediatric LASSO coefficient and the standardized protein values across all LASSO-included

proteins. Standardized scores were then used as an independent variable in Cox regression models for all-cause mortality and selected CKMD outcomes.

Our analytic sample in the UK Biobank included 28,256 unique participants that had complete proteomics data for at least 1 of the scores generated in BHRC paediatric participants. From this full sample, we defined 6 different subsets of participants (with participants allowed to be in >1 subset) based on having complete data for each set of proteins in each LF-based proteomic score of paediatric CKMD. Clinical outcomes in the UK Biobank were defined using the International Classification of Diseases (ICD) codes summarized by the PheWAS package⁷⁷. We excluded participants from models if a confounding ICD code was present or if they reported a prevalent diagnosis or related condition (UK Biobank Data Fields 2443, 20002). Time-to-death was defined by registry data (UK Biobank Data Field 40000), and non-deaths were censored on 30 November 2022. Time-to-event data for non-death outcomes were defined as the time to first qualifying ICD code, and non-event participants were censored at the date of death (if applicable) or region-specific dates (UK Biobank Data Field 54): 31 October 2022 for England, 31 July 2021 for Scotland and 28 February 2018 for Wales. Cox regression models included adjustments to capture metabolic risk measures used in clinical practice and more general confounders for mortality: age, sex, race and ethnicity, BMI, HbA1c, TG, HDL-C, Townsend deprivation index, smoking and alcohol use. We included non-linear terms as appropriate, given the potential non-linear dependence of CKMD risk on age. Our methods were adapted with minimal modification from previous work in our group to ensure scientific reproducibility¹¹ (with attribution by this statement).

Paediatric proteomic signatures related to dynamic changes in the circulating proteome with GLP-1 receptor agonist therapy in adults

Proteins associated with paediatric CKMD that are dynamic with GLP-1RA therapy may identify actionable biomarkers of long-term risk and support utilization of GLP-1RA in at-risk children. We mapped 1,032 proteins related to any CKMD phenotype in our paediatric sample (in age- and sex-adjusted models; q -value < 0.05) to a recently reported study of circulating proteins in the STEP 1 randomized clinical trial in adults (semaglutide)¹⁵. We examined changes in the circulating proteome after 68 weeks of treatment with semaglutide in the STEP 1 cohort (SomaScan aptamer-based platform), matching proteins identified in our cohort by UniProt identifier and restricting to those proteins with significant on-therapy change in STEP 1 (q -value ≤ 0.05). For duplicate aptamers for a given protein, the minimal absolute value effect size was chosen. The proteomic data from STEP 1 can be found in the supplemental data from a previous study¹⁵ (sheet 'S2_tx_STEP1', columns 'effect_size' and 'qvalue').

Genetically predicted PWAS

We next investigated whether proteins implicated by paediatric CKMD phenotypes have relevance to disease development in adults, using genetic models for expression of these proteins in adults. To investigate the impact of paediatric CKMD-associated proteins on adult disease outcomes, we leveraged genetic prediction models derived from 7,213 European-ancestry participants with circulating proteomic data in the Atherosclerosis Risk in Communities (Somalogic aptamer-based proteomics)⁷⁸. We conducted PWAS¹⁶ to identify proteins associated with adult CKMD outcomes, prioritizing those previously linked to paediatric CKMD phenotypes and responsive to GLP-1RA treatment (FDR < 5%). We integrated the protein expression models into the Mendelian randomization–joint tissue imputation pipeline to support causal inference⁷⁹. Mendelian randomization–joint-tissue imputation incorporates gene expression prediction models and applies Mendelian randomization to estimate the causal effect of gene expression on complex traits. By leveraging multiple genetic instruments and

cross-tissue information, it improves power and helps distinguish true causal genes from associations driven by linkage disequilibrium or co-expression; in the case of a single tissue, the method is equivalent to conventional PrediXcan. While a lack of large aptamer-based proteogenomic studies in childhood guided us to use adult genetic models of the proteome in this analysis, adult genetic models have been used to support the importance of lifelong exposure to key molecules in cardiovascular disease (for example, LDL and TG⁸⁰).

Of the 245 proteins that were significantly associated with paediatric CKMD phenotypes and showed evidence of change after GLP-1RA based on STEP 1 trial findings, 131 proteins were available among the 1,318 elastic net genetic models of the proteome. We used these genetic models in concert with large-scale GWAS to test whether genetically predicted protein levels of these 131 proteins would be associated with CKMD clinical outcomes in adults, prioritizing key CKMD outcomes: (1) heart failure⁸¹ ($n = 1,946,349$ individuals from 42 studies, 153,174 cases), (2) BMI ($n = 419,163$ in the UK Biobank⁸²), (3) T2D⁸³ ($n = 2,535,601$ individuals, 428,452 cases), (4) renal function⁸⁴ (creatinine; $n = 1,004,040$), (5) myocardial infarction⁸⁵ (3,961 African American or Afro-Caribbean cases, 51,544 African American or Afro-Caribbean controls, 2,384 Hispanic or Latin American cases, 26,930 Hispanic or Latin American controls, 225 East Asian ancestry cases, 3,322 East Asian ancestry controls, 39,300 European ancestry cases, 276,368 European ancestry controls) and (6) MASLD⁸⁶ (imaging $n = 66,814$ and diagnostic code (3,584 cases versus 621,081 controls)). We estimated the effect (z -score) of the genetically predicted protein levels on each clinical outcome as follows^{87,88}:

$$Z_g \approx \sum_{l \in \text{Model}_g} \omega_{lg} \frac{\hat{\sigma}_l}{\hat{\sigma}_g} \frac{\hat{\beta}_l}{\text{s.e.}(\hat{\beta}_l)}.$$

Here ω_{lg} is the weight assigned to the variant l by the trained protein model; $\hat{\beta}_l$ and $\text{s.e.}(\hat{\beta}_l)$ are the GWAS effect size and corresponding standard error summary statistics, respectively; and $\hat{\sigma}_l$ and $\hat{\sigma}_g$ are the estimated variance for the variant l and protein g , respectively.

Sources of variability in early development

To investigate sources of variability in early development, we used recent published data from a population-based birth cohort from Sweden (BAMSE)¹⁷. We extracted reported variance explained by age or sex and subject-specific random effect from linear mixed models analysing Olink Explore-derived NPX protein levels in 50 males and 50 females measured longitudinally at ages 4, 8, 16 and 24 (ref. 17). We plotted variance explained across variance components for proteins associated with both paediatric and adult CKMD in our BHRC sample. We highlighted a selection of participant-level proteins across time in BAMSE from the Human Protein Atlas (images taken directly from <https://www.proteinatlas.org>). Images for each protein were taken directly from the Human Protein Atlas (with credit by this statement for images included in Supplementary Fig. 5). Figures taken from the Atlas can be found at <https://www.proteinatlas.org>. An example URL is <https://v25.proteinatlas.org/ENSG00000054356-PTPRN/blood#bamse>.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

UK Biobank data used for this analysis is publicly available (<https://www.ukbiobank.ac.uk>) and was accessed using approved proposal 57492. Individual-level cohort data from the BHRC cannot be made publicly available owing to ethical constraints and restrictions imposed by the terms of participant informed consent. Raw proteomics data may be shared with qualified investigators for research purposes

following review by the BHRC Project Review Committee (comprising internal and external advisors) and execution of a formal Data Use Agreement (DUA) with UTHealth Houston; these data are not available to the general public. Applications for data access can be initiated by contacting BHRCResearch@uth.tmc.edu. Source data are provided with this paper.

References

- McGill, H. C. Jr. et al. Associations of coronary heart disease risk factors with the intermediate lesion of atherosclerosis in youth. The Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Research Group. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1998–2004 (2000).
- Carr, J. J. et al. Association of coronary artery calcium in adults aged 32 to 46 years with incident coronary heart disease and death. *JAMA Cardiol.* **2**, 391–399 (2017).
- Hartiala, O. et al. Adolescence risk factors are predictive of coronary artery calcification at middle age: the cardiovascular risk in young Finns study. *J. Am. Coll. Cardiol.* **60**, 1364–1370 (2012).
- Jacobs, D. R. Jr., Sinaiko, A. R. & Woo, J. G. Childhood risk factors and adult cardiovascular events. *N. Engl. J. Med.* **387**, 473–474 (2022).
- Jacobs, D. R. Jr. et al. Childhood cardiovascular risk factors and adult cardiovascular events. *N. Engl. J. Med.* **386**, 1877–1888 (2022).
- Sun, X. Life-course associations between blood pressure-related polygenic risk scores and hypertension in the Bogalusa Heart Study. *Genes* **13**, 1473 (2022).
- Shi, M. et al. Association of genome-wide polygenic risk score for body mass index with cardiometabolic health from childhood through midlife. *Circ. Genom. Precis. Med.* **15**, e003375 (2022).
- Li, S. et al. Variability and rapid increase in body mass index during childhood are associated with adult obesity. *Int. J. Epidemiol.* **44**, 1943–1950 (2015).
- Libman, I. M., Barinas-Mitchell, E., Bartucci, A., Robertson, R. & Arslanian, S. Reproducibility of the oral glucose tolerance test in overweight children. *J. Clin. Endocrinol. Metab.* **93**, 4231–4237 (2008).
- Sinha, R. et al. Prevalence of impaired glucose tolerance among children and adolescents with marked obesity. *N. Engl. J. Med.* **346**, 802–810 (2002).
- Perry, A. S. et al. Proteomic analysis of cardiorespiratory fitness for prediction of mortality and multisystem disease risks. *Nat. Med.* **30**, 1711–1721 (2024).
- Gerszten, R. E. & Wang, T. J. The search for new cardiovascular biomarkers. *Nature* **451**, 949–952 (2008).
- Khan, S. S. et al. Development and validation of the American Heart Association's PREVENT equations. *Circulation* **149**, 430–449 (2024).
- Wang, T. J. et al. Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17**, 448–453 (2011).
- Marett, L. et al. Proteomic changes upon treatment with semaglutide in individuals with obesity. *Nat. Med.* **31**, 267–277 (2025).
- Zhang, J. et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* **54**, 593–602 (2022).
- Álvarez, M. B. et al. A human pan-disease blood atlas of the circulating proteome. *Science* **390**, ead2678 (2025).
- Bearrows, S. C. et al. Chromogranin B regulates early-stage insulin granule trafficking from the Golgi in pancreatic islet beta-cells. *J. Cell Sci.* **132**, jcs231373 (2019).
- Zhang, K. et al. Autonomic function in hypertension; role of genetic variation at the catecholamine storage vesicle protein chromogranin B. *Circ. Cardiovasc. Genet.* **2**, 46–56 (2009).
- Anwar, M. Y. et al. Multi-omics of human obesity and related multi-system diseases. *J. Clin. Endocrinol. Metab.* **111**, 506–523 (2025).
- Cronje, H. T. et al. Plasma proteomic risk markers of incident type 2 diabetes reflect physiologically distinct components of glucose-insulin homeostasis. *Diabetes* **72**, 666–673 (2023).
- Rajwani, A. et al. Increasing circulating IGFBP1 levels improves insulin sensitivity, promotes nitric oxide production, lowers blood pressure, and protects against atherosclerosis. *Diabetes* **61**, 915–924 (2012).
- Perry, A. S. et al. A prognostic molecular signature of hepatic steatosis is spatially heterogeneous and dynamic in human liver. *Cell Rep. Med.* **5**, 101871 (2024).
- Sledzinski, T. et al. Association between cytosolic glycerol 3-phosphate dehydrogenase gene expression in human subcutaneous adipose tissue and BMI. *Cell. Physiol. Biochem.* **32**, 300–309 (2013).
- Swierczynski, J. et al. Enhanced glycerol 3-phosphate dehydrogenase activity in adipose tissue of obese humans. *Mol. Cell. Biochem.* **254**, 55–59 (2003).
- Zhou, J. et al. Liver regeneration and ethanol detoxification: a new link in YAP regulation of ALDH1A1 during alcohol-related hepatocyte damage. *FASEB J.* **36**, e22224 (2022).
- Ngo, D. et al. Proteomic profiling reveals biomarkers and pathways in type 2 diabetes risk. *JCI Insight* **6**, e144392 (2021).
- Sveinbjornsson, G. et al. Multiomics study of nonalcoholic fatty liver disease. *Nat. Genet.* **54**, 1652–1663 (2022).
- Renzi, G. et al. Epigenetic suppression of creatine kinase B in adipocytes links endoplasmic reticulum stress to obesity-associated inflammation. *Mol. Metab.* **92**, 102082 (2025).
- Rahbani, J. F. et al. Creatine kinase B controls futile creatine cycling in thermogenic fat. *Nature* **590**, 480–485 (2021).
- Bhangoo, A. et al. Fasting serum IGFBP-1 as a marker of insulin resistance in diverse school age groups. *Front. Endocrinol.* **13**, 840361 (2022).
- Le, S. et al. Does sex hormone-binding globulin cause insulin resistance during pubertal growth? *Endocr. Connect.* **8**, 510–517 (2019).
- Aydin, B. & Winters, S. J. Sex hormone-binding globulin in children and adolescents. *J. Clin. Res. Pediatr. Endocrinol.* **8**, 1–12 (2016).
- Prentice, K. J., Saksi, J. & Hotamisligil, G. S. Adipokine FABP4 integrates energy stores and counterregulatory metabolic responses. *J. Lipid Res.* **60**, 734–740 (2019).
- Nakamura, R. et al. Serum fatty acid-binding protein 4 (FABP4) concentration is associated with insulin resistance in peripheral tissues, a clinical study. *PLoS ONE* **12**, e0179737 (2017).
- Liu, R. et al. Adipocyte-derived chemerin rescues lipid overload-induced cardiac dysfunction. *iScience* **26**, 106495 (2023).
- Lin, T., Liang, T., Shen, Y. & Gao, F. BACE1 inhibition protects against type 2 diabetes mellitus by restoring insulin receptor in mice. *Int. J. Mol. Sci.* **26**, 5100 (2025).
- Meakin, P. J. et al. The beta secretase BACE1 regulates the expression of insulin receptor in the liver. *Nat. Commun.* **9**, 1306 (2018).
- Lin, J., Li, Q., Lei, X. & Zhao, H. The emerging roles of GPR158 in the regulation of the endocrine system. *Front. Cell Dev. Biol.* **10**, 1034348 (2022).
- Piaggi, P. et al. A genome-wide association study using a custom genotyping array identifies variants in GPR158 associated with reduced energy expenditure in American Indians. *Diabetes* **66**, 2284–2295 (2017).
- Tassi, E. et al. Fibroblast growth factor binding protein 3 (FGFBP3) impacts carbohydrate and lipid metabolism. *Sci. Rep.* **8**, 15973 (2018).

42. Cai, C. et al. The deubiquitinase USP28 maintains the expression of PPAR γ and its inactivation protects mice from diet-induced MASH and hepatocarcinoma. *Mol. Ther.* **33**, 1825–1841 (2025).
43. Kern, P. et al. Creld2 function during unfolded protein response is essential for liver metabolism homeostasis. *FASEB J.* **35**, e21939 (2021).
44. She, Z. G., Chen, H. Z., Yan, Y., Li, H. & Liu, D. P. The human paraoxonase gene cluster as a target in the treatment of atherosclerosis. *Antioxid. Redox Signal.* **16**, 597–632 (2012).
45. Shen, X. et al. Unveiling the impact of ApoF deficiency on liver and lipid metabolism: insights from transcriptome-wide m6A methylome analysis in mice. *Genes* **15**, 347 (2024).
46. Svensson, K. J. et al. A secreted Slit2 fragment regulates adipose tissue thermogenesis and metabolic function. *Cell Metab.* **23**, 454–466 (2016).
47. Hotta, K., Hosaka, M., Tanabe, A. & Takeuchi, T. Secretogranin II binds to secretogranin III and forms secretory granules with orexin, neuropeptide Y, and POMC. *J. Endocrinol.* **202**, 111–121 (2009).
48. Wang, L. et al. PAI-1 exacerbates white adipose tissue dysfunction and metabolic dysregulation in high fat diet-induced obesity. *Front. Pharmacol.* **9**, 1087 (2018).
49. Denton, N., Pinnick, K. E. & Karpe, F. Cartilage oligomeric matrix protein is differentially expressed in human subcutaneous adipose tissue and regulates adipogenesis. *Mol. Metab.* **16**, 172–179 (2018).
50. Loh, N. Y. et al. Bidirectional Mendelian randomization highlights causal relationships between circulating INHBC and multiple cardiometabolic diseases and traits. *Diabetes* **73**, 2084–2094 (2024).
51. Windelinckx, A. et al. Comprehensive fine mapping of chr12q12-14 and follow-up replication identify activin receptor 1B (ACVR1B) as a muscle strength gene. *Eur. J. Hum. Genet.* **19**, 208–215 (2011).
52. Guan, Y. et al. SPON2 suppresses autophagy by regulating the PI3K/AKT/mTOR signaling pathway in renal cell carcinoma. *Curr. Cancer Drug Targets* <https://doi.org/10.2174/0115680096352140250101041725> (2025).
53. Hoekstra, M. et al. Genome-wide association study highlights APOH as a novel locus for lipoprotein(a) levels—brief report. *Arterioscler. Thromb. Vasc. Biol.* **41**, 458–464 (2021).
54. Zhong, Q. et al. A proteome-wide association study reveals novel plasma proteins as potential therapeutic targets for metabolic dysfunction-associated steatotic liver disease. *Front. Endocrinol.* **16**, 1664691 (2025).
55. Ding, H. Y. et al. Roles of hyaluronan in cardiovascular and nervous system disorders. *J. Zhejiang Univ. Sci. B* **20**, 428–436 (2019).
56. Webber, L. S., Srinivasan, S. R., Wattigney, W. A. & Berenson, G. S. Tracking of serum lipids and lipoproteins from childhood to adulthood. The Bogalusa Heart Study. *Am. J. Epidemiol.* **133**, 884–899 (1991).
57. Fox, C. K. et al. Liraglutide for children 6 to <12 years of age with obesity—a randomized trial. *N. Engl. J. Med.* **392**, 555–565 (2025).
58. Palmer, A. B. et al. Prevalence of obesity with confirmed excess adiposity in US children and adolescents. *JAMA Pediatr.* **179**, 1227–1229 (2025).
59. Dwyer, T. et al. Decline in physical fitness from childhood to adulthood associated with increased obesity and insulin resistance in adults. *Diabetes Care* **32**, 683–687 (2009).
60. Agbaje, A. O., Barker, A. R., Lewandowski, A. J., Leeson, P. & Tuomainen, T. P. Accelerometer-based sedentary time, light physical activity, and moderate-to-vigorous physical activity from childhood with arterial stiffness and carotid IMT progression: a 13-year longitudinal study of 1339 children. *Acta Physiol.* **240**, e14132 (2024).
61. Shah, R. V. et al. Myocardial tissue remodeling in adolescent obesity. *J. Am. Heart Assoc.* **2**, e000279 (2013).
62. Ahola-Olli, A. V. et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* **62**, 2298–2309 (2019).
63. Geyer, P. E. et al. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* **12**, 901 (2016).
64. Nowak, C. et al. Multiplex proteomics for prediction of major cardiovascular events in type 2 diabetes. *Diabetologia* **61**, 1748–1757 (2018).
65. Luo, Y. et al. SOMAscan proteomics identifies serum biomarkers associated with liver fibrosis in patients with NASH. *Hepatol. Commun.* **5**, 760–773 (2021).
66. Shah, R. V. et al. Targeted proteomics reveals functional targets for early diabetes susceptibility in young adults. *Circ. Genom. Precis. Med.* **17**, e004192 (2024).
67. Stinson, S. E. et al. Identification of modifiable plasma protein markers of cardiometabolic risk in children and adolescents with obesity. *Nat. Commun.* **17**, 1718 (2026).
68. Niu, L. et al. Plasma proteome variation and its genetic determinants in children and adolescents. *Nat. Genet.* **57**, 635–646 (2025).
69. Schwimmer, J. B., Biddinger, S. B. & Ibrahim, S. H. MASLD in children: integrating epidemiological trends with mechanistic and translational advances. *J. Clin. Invest.* **135**, e186422 (2025).
70. Duell, P. B. et al. Nonalcoholic fatty liver disease and cardiovascular risk: a scientific statement from the American Heart Association. *Arterioscler. Thromb. Vasc. Biol.* **42**, e168–e185 (2022).
71. Miller, M. G., Terebuh, P., Kaelber, D. C., Xu, R. & Davis, P. B. Characterizing GLP-1 receptor agonist use in preadolescent and adolescent populations. *JAMA Netw. Open* **7**, e2439887 (2024).
72. Lee, J. M., Sharifi, M., Oshman, L., Griaude, D. H. & Chua, K. P. Dispensing of glucagon-like peptide-1 receptor agonists to adolescents and young adults, 2020–2023. *JAMA* **331**, 2041–2043 (2024).
73. Pierce, C. B. et al. Age- and sex-dependent clinical equations to estimate glomerular filtration rates in children and young adults with chronic kidney disease. *Kidney Int.* **99**, 948–956 (2021).
74. Lindman, B. R. et al. Integrated multiomics of pressure overload in the human heart prioritizes targets relevant to heart failure. *Nat. Commun.* **16**, 6889 (2025).
75. Dornbos, P. et al. Evaluating human genetic support for hypothesized metabolic disease genes. *Cell Metab.* **34**, 661–666 (2022).
76. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
77. Wu, P. et al. Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Med. Inform.* **7**, e14325 (2019).
78. The ARIC Investigators The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. *Am. J. Epidemiol.* **129**, 687–702 (1989).
79. Zhou, D. et al. A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat. Genet.* **52**, 1239–1246 (2020).
80. Ference, B. A. et al. Association of triglyceride-lowering LPL variants and LDL-C-lowering LDLR variants with risk of coronary heart disease. *JAMA* **321**, 364–373 (2019).
81. Henry, A. et al. Genome-wide association study meta-analysis provides insights into the etiology of heart failure and its subtypes. *Nat. Genet.* **57**, 815–828 (2025).
82. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

83. Suzuki, K. et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).
84. Stanzick, K. J. et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat. Commun.* **12**, 4350 (2021).
85. Verma, A. et al. Diversity and scale: genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).
86. Chen, Y. et al. Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. *Nat. Genet.* **55**, 1640–1650 (2023).
87. Gamazon, E. R., Zwinderman, A. H., Cox, N. J., Denys, D. & Derks, E. M. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat. Genet.* **51**, 933–940 (2019).
88. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
89. van Eekeren, L. E. et al. Plasma proteomic signatures of liver steatosis and fibrosis in people living with HIV: a cross-sectional study. *EBioMedicine* **109**, 105407 (2024).
90. Turk, N. E. et al. GPD1 deficiency—a rare, overlooked cause of liver disease. *J. Hum. Genet.* **70**, 375–379 (2025).
91. Joshi, M. et al. A compound heterozygous mutation in GPD1 causes hepatomegaly, steatohepatitis, and hypertriglyceridemia. *Eur. J. Hum. Genet.* **22**, 1229–1232 (2014).
92. Park, J. J., Berggren, J. R., Hulver, M. W., Houmard, J. A. & Hoffman, E. P. GRB14, GPD1, and GDF8 as potential network collaborators in weight loss-induced improvements in insulin action in human skeletal muscle. *Physiol. Genomics* **27**, 114–121 (2006).
93. Bhatt, D. K., Gaedigk, A., Pearce, R. E., Leeder, J. S. & Prasad, B. Age-dependent protein abundance of cytosolic alcohol and aldehyde dehydrogenases in human liver. *Drug Metab. Dispos.* **45**, 1044–1048 (2017).
94. Duan, H. et al. GBP1 promotes acute rejection after liver transplantation by inducing Kupffer cells pyroptosis. *Biochim. Biophys. Acta Mol. Basis Dis.* **1871**, 167644 (2025).
95. Ge, H. et al. Fibroblast growth factor receptor 4 (FGFR4) deficiency improves insulin resistance and glucose metabolism under diet-induced obesity conditions. *J. Biol. Chem.* **289**, 30470–30480 (2014).
96. Huang, X. et al. FGFR4 prevents hyperlipidemia and insulin resistance but underlies high-fat diet induced fatty liver. *Diabetes* **56**, 2501–2510 (2007).
97. Mizunoe, Y. et al. Cathepsin B overexpression induces degradation of perilipin 1 to cause lipid metabolism dysfunction in adipocytes. *Sci. Rep.* **10**, 634 (2020).
98. Yu, J. et al. The candidate schizophrenia risk gene *Tmem108* regulates glucose metabolism homeostasis. *Front. Endocrinol.* **12**, 770145 (2021).
99. Goulding, N. et al. Inflammation proteomics datasets in the ALSPAC cohort. *Wellcome Open Res.* **7**, 277 (2022).
100. Krakowczyk, D. et al. Proteomic insights into childhood obesity: a systematic review of protein biomarkers and advances. *Int. J. Mol. Sci.* **26**, 8522 (2025).
101. Stratakis, N. et al. Multi-omics architecture of childhood obesity and metabolic dysfunction uncovers biological pathways and prenatal determinants. *Nat. Commun.* **16**, 654 (2025).
102. Zhao, C. et al. Integrative multi-omics analysis reveals molecular signatures of central obesity in children. *Pediatr. Res.* **98**, 1374–1385 (2025).
103. Paine-Cabrera, D. et al. Proteomics used in identifying novel correlates of disease in pediatric metabolic dysfunction-associated steatotic liver disease. *Clin. Transl. Sci.* **18**, e70271 (2025).

Acknowledgements

We would like to thank the BHRC cohort team, particularly R. Uribe, who recruited and interviewed the participants; M. Morris and her team for laboratory support; Valley Baptist Medical Center in Brownsville, Texas, for providing space where our Center for Clinical and Translational Science Clinical Research Unit is located; and the community of Brownsville and the participants who so willingly contributed to this study in their city.

Author contributions

J.M.L., H.M.H., A. S. Perry, Q.S., A.L., A.B.P., S.Z., W.Z., X.Z., V.L.B., E.G.F., R.R., A.S., E.H.F.-E., M.Y.A., A.B., C.A.B., E.R.G., P.G.-L., J.E.B., K.E.N. and R.V.S. contributed to data analysis. J.M.L., H.M.H., J.B.M., S.P.F.-H., P.G.-L., J.E.B., K.E.N. and R.V.S. contributed to study conception and design. J.M.L., H.M.H., A. S. Perry, Q.S., E.R.G., P.G.-L., J.E.B., K.E.N. and R.V.S. contributed to writing the paper. J.M.L., H.M.H., A. S. Perry, A.G.H., A.L., A.B.P., S.Z., W.Z., X.Z., V.L.B., E.G.F., R.R., A.S., J.K.S., T.-C.W., J.T., A.G., L.E.P., L.F.-R., H.-H.C., M.K., M.G., K.A.M., M.L., K.L.Y., Q.W., E.R.G., J.B.M., S.P.F.-H., P.G.-L., J.E.B., K.E.N. and R.V.S. edited the paper. All authors contributed to the critical review of the paper. E.H.F.-E., M.N., A. S. Petty, B.L., J.E.F., E.R.G., J.B.M., S.P.F.-H., P.G.-L., J.E.B., K.E.N. and R.V.S. contributed to data acquisition. J.E.B., P.G.-L., K.E.N. and R.V.S. are co-senior authors. P.G.-L., J.E.B., K.E.N. and R.V.S. contributed equally to this work.

Funding

This study was funded in part by the Center for Clinical and Translational Sciences and a National Institutes of Health (NIH) Clinical and Translational Award (grant UL1TR000371) from the National Center for Advancing Translational Sciences. Funding for the BHRC is supported by the following NIH grants: U54AG089326, R01DK139598, R01HL142302, R01HL163262, R01DK127084 and U01CA288325. J.M.L. was supported by NIH grant T32HG008341. J.M.L., H.M.H., J.B.M., S.P.F.-H., J.E.B. and K.E.N. were supported by NIH grant U01CA288325. J.M.L. and M.L. were supported by R01AG078452. A. S. Perry was supported by NIH grant K23HL179316 and is a co-inventor on patents of molecular signatures of fitness, liver and lung disease. H.M.H., A.G.H., P.G.-L., J.E.B. and K.E.N. were supported by NIH grant R01DK139598. A.B.P., J.K.S., J.B.M., S.P.F.-H., J.E.B. and K.E.N. were supported by NIH grants R01HL142302 and R01DK122503. A.B.P. was supported by NIH grant T32GM145734. J.M.L., A.G., J.B.M., S.P.F.-H., J.E.B. and K.E.N. were supported by NIH grant R01DK127084. M.N. is supported by National Institutes of Health (R01HL156975, R01HL131029 and R21AG086679). M.K. was supported by NIH grant U24DK132715. J.B.M., S.P.F.-H., J.E.B. and K.E.N. were supported by NIH grant U54AG089326. E.R.G. is supported by National Institutes of Health (grants U01DK140952 and R01HG011138). These studies used the Functional Multi-Omics of Diabetes and Metabolism Core of the Vanderbilt Diabetes Research and Training Center (P60DK020593 to E.R.G. and J.E.B.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper.

Competing interests

E.R.G. has performed consulting for Thryv Therapeutics and has pending patents or disclosures on molecular biomarkers of fitness, cardiovascular diseases and phenotypes, and metabolic health; use of RNAs as therapeutics and diagnostic biomarkers in disease; and methods in metabolomics. R.V.S., E.R.G., E.H.F.-E. and Q.W. are supported by Bayer. R.V.S. is supported by grants from the National Institutes of Health and institutional endowment funds. R.V.S. has equity ownership in and is a consultant for Thryv Therapeutics and is a co-inventor on pending patents or disclosures on molecular biomarkers of fitness, lung disease, cardiovascular diseases and phenotypes, and metabolic health; use of RNAs (including spatial) as

therapeutics and diagnostic biomarkers in disease; and methods in metabolomics. K.N. received grant funding from AMGEN and Takeda. J.T. receives research funding from Eli Lilly. A. S. Perry has patents issued or pending for proteomic signatures of fitness, liver disease and lung disease. All other authors have no disclosures to report.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s42255-026-01589-7>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42255-026-01589-7>.

Correspondence and requests for materials should be addressed to Kari E. North or Ravi V. Shah.

Peer review information *Nature Metabolism* thanks Anna di Sessa, Eiichiro Satake, Sara Stinson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Yanina-Yasmin Pesch, in collaboration with the *Nature Metabolism* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

Joshua M. Landman^{1,2,3,21}, Heather M. Highland^{4,5,21}, Andrew S. Perry^{6,21}, Annie G. Howard⁷, Quanhu Sheng⁸, Anna Lorenz¹, Alexandra B. Palmer⁵, Shilin Zhao⁸, Wanying Zhu¹, Xinruo Zhang⁵, Victoria L. Buchanan⁴, Elizabeth G. Frankel⁹, Rashedeh Roshani⁹, Alyssa Scartozzi⁹, Eric H. Farber-Eger¹⁰, Mohammad Y. Anwar^{4,5}, Jessica K. Sprinkles¹¹, Ash Breidenbach⁹, Ting-Chen Wang⁹, Christine A. Ballard⁵, Matthew Naylor¹², Jaclyn Tamaroff¹³, Absalon Gutierrez¹⁴, Lauren E. Petty¹, Alexander S. Petty¹, Benjamin Lippi¹, Lindsay Fernandez-Rhodes¹⁵, Hung-Hsin Chen¹⁶, Mohanraj Krishnan^{5,15}, Mariaelisa Graff¹⁵, Katie A. Meyer^{11,17}, Miryoung Lee⁴, Kristin L. Young⁴, Quinn Wells^{2,18}, Kahraman Tanriverdi¹⁹, Jane E. Freedman¹⁹, Eric R. Gamazon², Joseph B. McCormick⁴, Susan P. Fisher-Hoch⁴, Penny Gordon-Larsen^{11,22}, Jennifer E. Below^{1,2,22}, Kari E. North^{1,2,22} ✉ & Ravi V. Shah^{1,20,22} ✉

¹Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. ²Division of Genetic Medicine and Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ³Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. ⁴Department of Epidemiology, Border Health Research Center, School of Public Health, Brownsville Regional Campus, The University of Texas Health Science Center at Houston, Brownsville, TX, USA. ⁵Department of Epidemiology, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁶Vanderbilt Translational and Clinical Cardiovascular Research Center, Vanderbilt University Medical Center, Nashville, TN, USA. ⁷Department of Biostatistics, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁸Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA. ⁹Division of Genetic Medicine and Clinical Pharmacology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA. ¹⁰Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University, Nashville, TN, USA. ¹¹Department of Nutrition, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ¹²Sections of Cardiovascular Medicine, Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA. ¹³Division of Pediatric Endocrinology and Diabetes, Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA. ¹⁴Division of Endocrinology, Diabetes and Metabolism, Department of Internal Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA. ¹⁵Department of Biobehavioral Health, College of Human and Health Development, Pennsylvania State University, University Park, PA, USA. ¹⁶Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. ¹⁷Nutrition Research Institute, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ¹⁸Division of Cardiovascular Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ¹⁹Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. ²⁰Vanderbilt Diabetes Center, Vanderbilt University Medical Center, Nashville, TN, USA. ²¹These authors contributed equally: Joshua M. Landman, Heather M. Highland, Andrew S. Perry.

²²These authors jointly supervised this work: Penny Gordon-Larsen, Jennifer E. Below, Kari E. North, Ravi V. Shah. ✉ e-mail: Kari.E.North@uth.tmc.edu; ravi.shah@umc.org

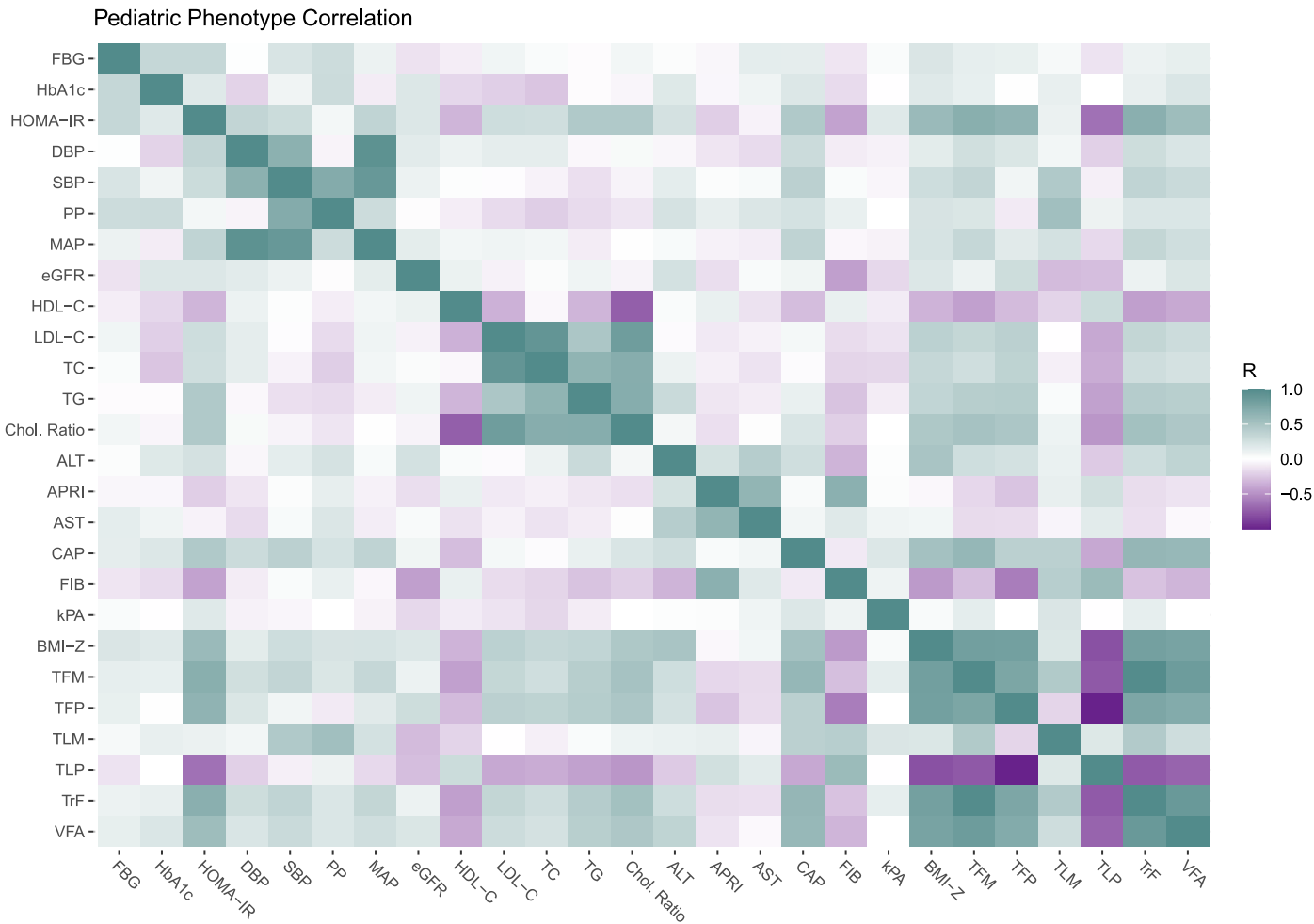
Extended Data Table 1 | Demographic summary for adult participants analyzed from the UK Biobank

Characteristic	N	Value
Age (years)	28,256	58 (50, 64)
Female	28,256	15,273 (54%)
Ethnic Background	28,256	
Asian or Asian British		596 (2.1%)
Black or Black British		690 (2.4%)
Mixed		189 (0.7%)
Unknown Other		495 (1.8%)
White		26,286 (93%)
Body mass index (kg/m ²)	28,108	26.7 (24.1, 29.9)
Hemoglobin A1c (mmol/mol)	26,769	35.3 (32.9, 38.1)
Triglycerides (mmol/L)	26,839	1.47 (1.04, 2.14)
High-density lipoprotein (mmol/L)	24,599	1.40 (1.16, 1.68)
Smoking status	28,223	
Current		3,067 (11%)
Never No answer		15,388 (55%)
Previous		9,768 (35%)
Alcohol use	28,223	
Current		25,753 (91%)
Never No answer		1,371 (4.9%)
Previous		1,099 (3.9%)
Townsend Deprivation Index	28,218	-2.0 (-3.6, 0.8)
Whole body fat mass (kg)	2,425	3.0 (2.1, 5.2)
Whole body fat free mass (kg)	27,601	23 (18, 30)

Models for each protein score included a subset of participants with complete data required to generate the respective protein score. The number of participants for each model is as follows: LF1=22,389; LF2=22,607; LF3=24,504; LF4=25,018; LF5=22,814; LF6=21,955. Fat mass and fat-free mass were measured by bioimpedance. Continuous variables are reported as mean (25th percentile, 75th percentile), and categorical variables are reported as N (%).

Extended Data Table 2 | Contextualization of findings to several recent studies in pediatric CKMD proteomics

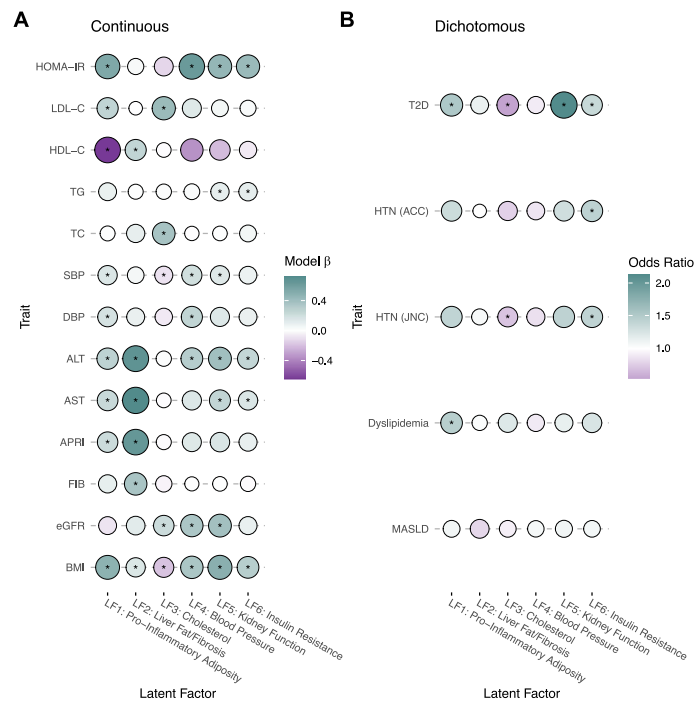
Study	Design & proteomics platform	Contextualization to current findings
Goulding et al. (ALSPAC cohort) ⁹⁹	3005 children (age ≈9 years), Olink (96 proteins, inflammation)	• Analogous findings with inflammation (IL-6) to BMI and CKMD traits in children
Stinson et al. (HOLBAEK study) ⁶⁷	Cross-sectional: 4024 children (2377 with obesity; age 4-20)	• BMI-related proteins in HOLBAEK were similar to those found in association with CKMD traits in children (for example, LEP, CDCP1, HGF)
	Longitudinal: 200 children (age 7-17) with 1-yr obesity therapy	• Protein-BMI association in pediatric population was correlated with association in adults (UK Biobank), similar to findings in BHRC adults/children
	<u>Olink</u> (149 proteins)	• Changes in proteins related to CKMD (from our BHRC study) were associated with changes in cardiometabolic traits with weight loss in the Stinson et al. study (for example, AXIN1, CD40, CDCP1, CX3CL1, FGF21, HGF, IL10RB, IL18R1, IL7, LIFR, TNFSF14)
Krakowczyk et al. (meta-analysis) ¹⁰⁰	Meta-analyses of 20 studies in children for association with obesity (age 0-18; multiple assays used)	• Proteins implicated by CKMD in BHRC children have been shown in other cohorts in meta-analyses (for example, APOA1, ALDH members, IGFBP1, PTX3, UMOD, GSTA1, RBP4)
Stratakis et al. (HELIX project) ¹⁰¹	863 children (mean age ≈7–8 years); multi-omics performed (Luminex for proteins)	• “High-risk” cluster for metabolic disease included proteins associated with CKMD in BHRC (for example, HGF, IL6)
Zhao et al. ¹⁰²	169 children (with and without obesity; mean age 11.3 years; mass spectrometry proteomics)	• Key proteins related to obesity also found associated with CKMD traits in BHRC (for example, SHBG, LEP, ADH1B, ADH4, IGFBP2, CHGA)
Paine-Cabrera et al. ¹⁰³	128 children (with and without MASLD; mean age ≈12–13 years; mass spectrometry proteomics)	• Key proteins related to MASLD also found associated with CKMD traits in BHRC (for example, SHBG, ADH4)



Extended Data Fig. 1 | Phenotype correlation heatmap. This heatmap shows the correlations between each pair of measured phenotypes in the pediatric study sample. Full data is available in Supplementary Table 11.

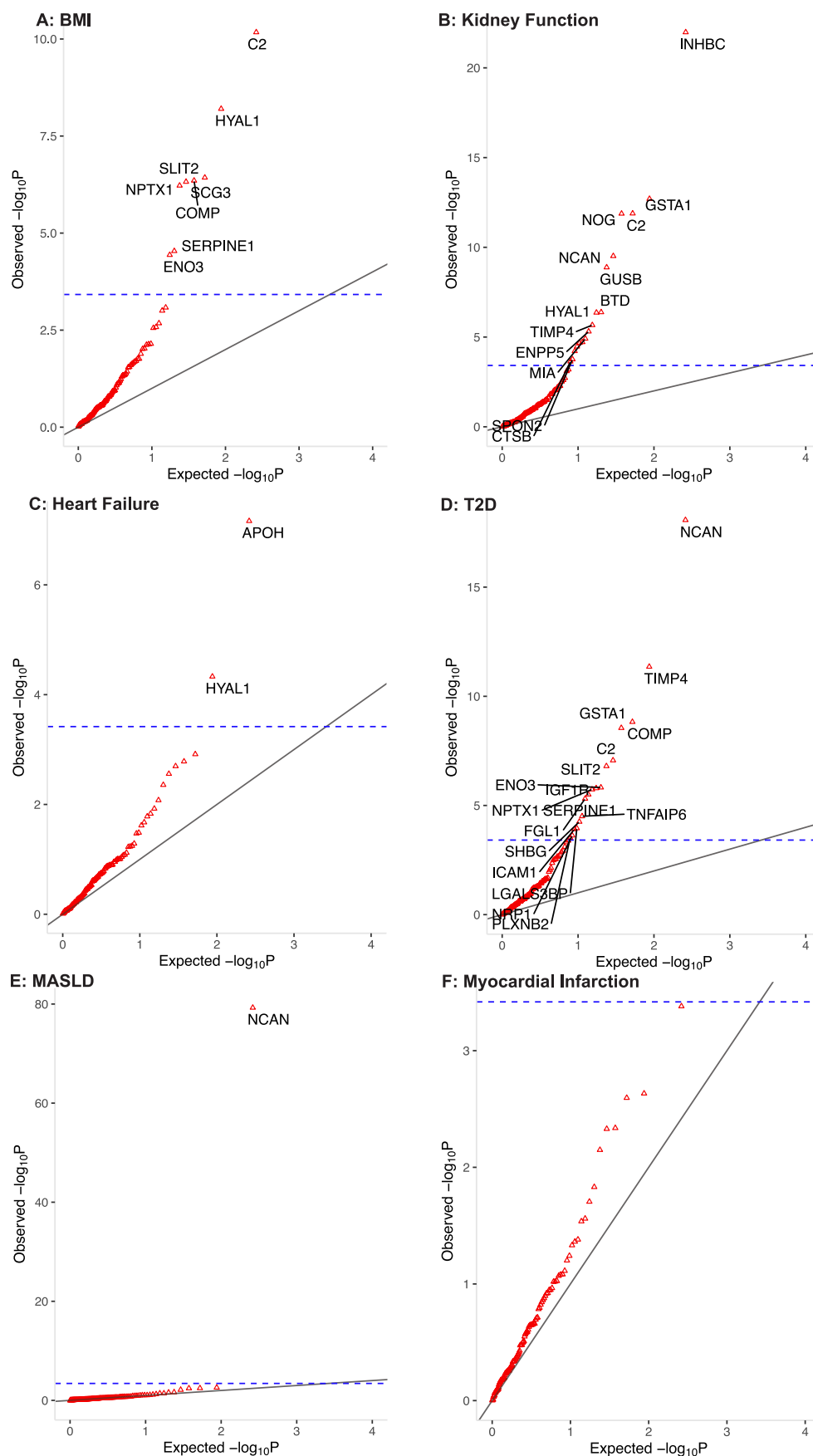


Extended Data Fig. 2 | HuGE genetic evidence heat map. Supplementary Fig. 2 Genetic evidence (HuGE) for select proteins from pediatric studies (as described in Methods).



Extended Data Fig. 3 | Mapping proteomic signatures of the pediatric CKMD proteome to adults. Age- and sex-adjusted association of each proteomic score (built on latent factors of the pediatric CKMD phenome) in up to 521 Hispanic/Latino adults from the same population. Size and color intensity represent the

magnitude of the effect size. The * indicates significance at FDR-adjusted P 0.05. Correlations are Pearson. Panel (a) displays estimates for outcomes that are continuous and panel (b) displays estimates for outcomes that are dichotomous. Full data in Supplementary Table 10.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Proteome-wide association studies (PWAS) for CKMD outcomes in adults. Each panel represents the results of PWAS for proteins that were (1) associated with CKMD phenotypes in BHRC children, (2) responsive to GLP-1RA therapy in adults, and (3) had models of genetically predicted expression. The diagonal line represents complete agreement with the null hypothesis; an upward/leftward shift from the diagonal line represents an

enrichment of significant associations compared to what was expected under the null. The blue dotted line represents a Bonferroni p-value threshold of 0.05 across the proteins tested. Proteins above this line (labeled) exhibit genetic evidence for association with the target outcome. Panels **a–f** detail associations for each outcome: BMI, kidney function, heart failure, T2D, MASLD, and myocardial infarction.