

RESEARCH ARTICLE

Open Access



# Longitudinal multi-omics study reveals common etiology underlying association between plasma proteome and BMI trajectories in adolescent and young adult twins

Gabin Drouard<sup>1\*</sup> , Fiona A. Hagenbeek<sup>1,2,3</sup>, Alyce M. Whipp<sup>1</sup>, René Pool<sup>2,3</sup>, Jouke Jan Hottenga<sup>2,3</sup>, Rick Jansen<sup>4,5</sup>, Nikki Hubers<sup>2,3,6</sup>, Aleksei Afonin<sup>7</sup>, BIOS Consortium, BBMRI-N. L. Metabolomics Consortium, Gonneke Willemssen<sup>2,3</sup>, Eco J. C. de Geus<sup>2,3</sup>, Samuli Ripatti<sup>1,8,9</sup>, Matti Pirinen<sup>1,8,10</sup>, Katja M. Kanninen<sup>7</sup>, Dorret I. Boomsma<sup>2,3,6</sup>, Jenny van Dongen<sup>2,3,6</sup> and Jaakko Kaprio<sup>1\*</sup>

## Abstract

**Background** The influence of genetics and environment on the association of the plasma proteome with body mass index (BMI) and changes in BMI remains underexplored, and the links to other omics in these associations remain to be investigated. We characterized protein–BMI trajectory associations in adolescents and adults and how these connect to other omics layers.

**Methods** Our study included two cohorts of longitudinally followed twins: FinnTwin12 ( $N=651$ ) and the Netherlands Twin Register (NTR) ( $N=665$ ). Follow-up comprised 4 BMI measurements over approximately 6 (NTR: 23–27 years old) to 10 years (FinnTwin12: 12–22 years old), with omics data collected at the last BMI measurement. BMI changes were calculated in latent growth curve models. Mixed-effects models were used to quantify the associations between the abundance of 439 plasma proteins with BMI at blood sampling and changes in BMI. In FinnTwin12, the sources of genetic and environmental variation underlying the protein abundances were quantified by twin models, as were the associations of proteins with BMI and BMI changes. In NTR, we investigated the association of gene expression of genes encoding proteins identified in FinnTwin12 with BMI and changes in BMI. We linked identified proteins and their coding genes to plasma metabolites and polygenic risk scores (PRS) applying mixed-effects models and correlation networks.

**Results** We identified 66 and 14 proteins associated with BMI at blood sampling and changes in BMI, respectively. The average heritability of these proteins was 35%. Of the 66 BMI-protein associations, 43 and 12 showed genetic and environmental correlations, respectively, including 8 proteins showing both. Similarly, we observed 7 and 3 genetic and environmental correlations between changes in BMI and protein abundance, respectively. *S100A8*

\*Correspondence:

Gabin Drouard  
gabin.drouard@helsinki.fi  
Jaakko Kaprio  
jaakko.kaprio@helsinki.fi

Full list of author information is available at the end of the article



© The Author(s) 2023. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

gene expression was associated with BMI at blood sampling, and the *PRG4* and *CFI* genes were associated with BMI changes. Proteins showed strong connections with metabolites and PRSs, but we observed no multi-omics connections among gene expression and other omics layers.

**Conclusions** Associations between the proteome and BMI trajectories are characterized by shared genetic, environmental, and metabolic etiologies. We observed few gene-protein pairs associated with BMI or changes in BMI at the proteome and transcriptome levels.

**Keywords** Multi-omics, Proteome, Metabolome, Transcriptome, Polygenic risk scores, Body mass index (BMI), Changes in BMI, Longitudinal twin study

## Background

In recent decades, the prevalence of obesity has been increasing [1], and it is predicted that almost one-fourth of the world's population will be affected by obesity in 2035 [2]. The co-morbidities related to obesity include a wide range of high-prevalence diseases, including type 2 diabetes and cardiovascular disease [3–5], making it a major public health concern.

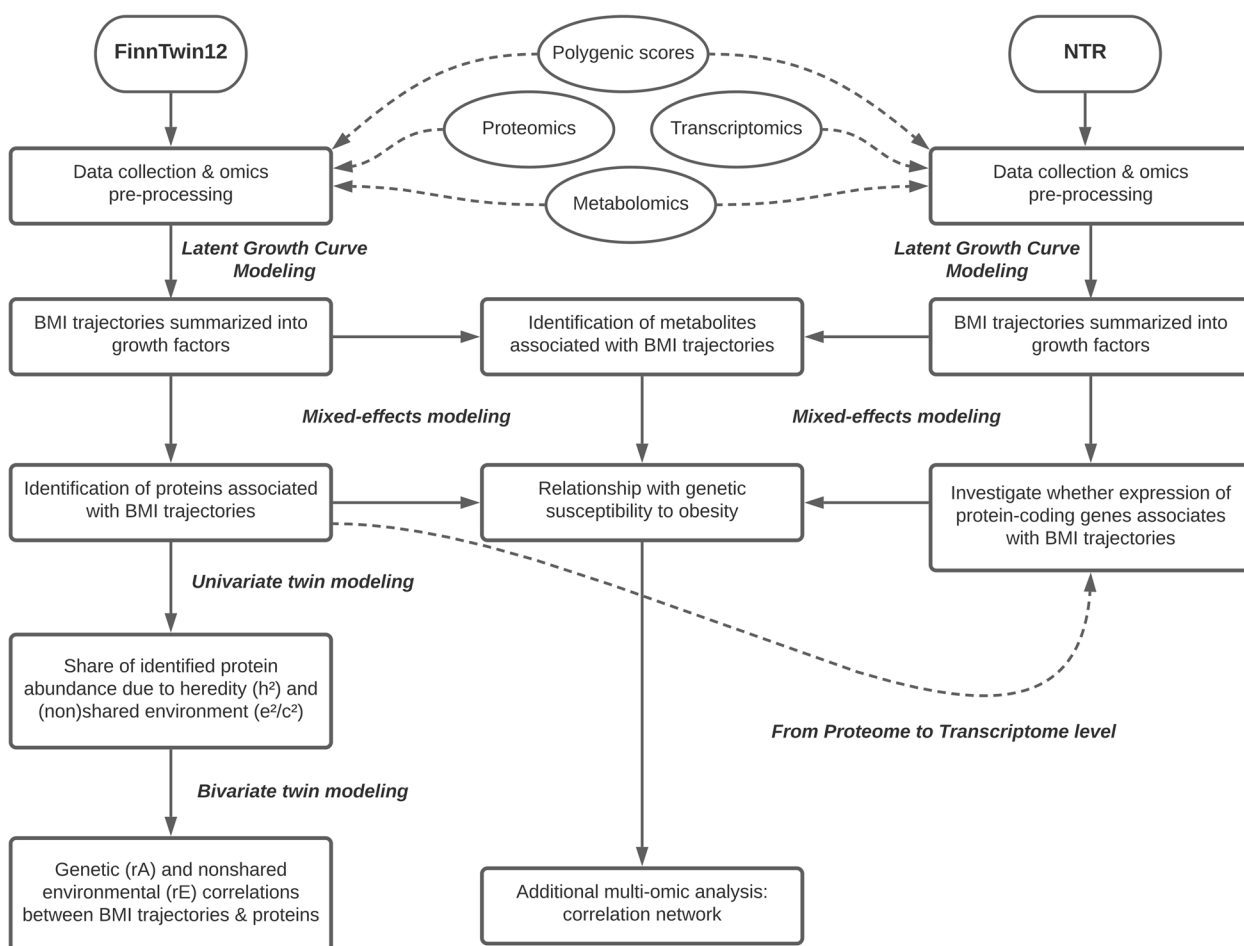
Body mass index (BMI) is commonly used as a marker of obesity and body fat, the former being defined as  $\text{BMI} > 30 \text{ kg.m}^{-2}$ . Both genetics and the environment influence BMI, with twin studies estimating ~75% of its variance attributable to genetic factors in young adults [6]. Multiple proteins and peptides, such as the hormones leptin, ghrelin, and resistin, take part in the complex processes of regulating energy balance, disturbances of which can induce obesity or anorexia [7, 8]. High-throughput technologies have revolutionized obesity research, and omics data are key to an in-depth functional understanding of obesity [5]. An omics technique successful in investigating obesity-related co-morbidities is proteomics, which comprises the large-scale study of proteins [9]. For example, beyond the risk of general obesity in the development of diabetes, plasma protein markers of abdominal fat distribution are associated with an increased risk of diabetes [10]. Mendelian randomization identified unidirectional and bidirectional (i.e., protein-to-BMI and/or BMI-to-protein) causal associations of plasma proteins with BMI [11, 12], signifying a direct imprint of the proteome on obesity, and vice versa. Mendelian randomization leverages genetic variants that influence protein levels and BMI, but little is known about how much of the association between protein levels and BMI is due to genetic and environmental influences.

Multi-omics approaches that combine multiple omics layers in a single analysis better characterize the underlying biology of complex diseases than single-omics approaches [13, 14]. The few multi-omics studies conducted for obesity demonstrated great potential in acquiring a better understanding of obesity, notably when

coupling transcriptomic and metabolomic data [15, 16]. Specifically, one study showed that some metabolite and gene expression associations with BMI were mediated by the epigenome [15], while another study observed associations of weight gain with changes in metabolite levels and blood cell function [16]. Among twin pairs discordant for BMI, weight differences were associated with lipidomics independent of genetic influences [17]. Overall, only a few multi-omics studies quantified biomolecules (i.e., metabolites or proteins) and omics stability in relation to participants' post-intervention weight changes [18–20].

Weight development and obesity in adolescents and children from a multi-omics perspective is largely underexplored. The large inter- and intra-individual variability in the proteome, metabolome, and transcriptome in children [21] holds great promise for identifying obesity biomarkers across omics layers. Better identification of biomarkers for obesity in children and adolescents may allow both prevention of its onset in this age group and in adults, as childhood obesity increases the likelihood of adult obesity [22]. Longitudinal designs can help capture factors associated with weight change, and thus better predict individuals at risk for obesity. A few longitudinal single-omic studies have explored the associations between changes in BMI with proteomic [23–26] or metabolomic data [27]. While the studies involving proteomic data are mainly based on adult populations, one study to investigate the association between weight change and metabolites was conducted in children [28]. No one, to our knowledge, has conducted longitudinal multi-omics studies of weight change and BMI including proteomics data in children or adolescent populations yet.

Plasma proteins are associated with BMI in adults, where some plasma proteins were shown to be causally influencing increases in BMI, while other plasma proteins were causally influenced by higher BMI [11, 12]. As less is known about these associations in adolescence and for changes in BMI, the primary objective of the current study was to identify associations between plasma



**Fig. 1** Study flowchart. The study was divided into two main sequential steps. First, growth factors were calculated and then association studies were performed. In the FinnTwin12 sample, univariate and bivariate twin modeling was performed to quantify the genetic and environmental sources underlying both the protein abundances and the associations between proteins and BMI trajectories. Cross-omics associations were performed to bring a holistic perspective to the analyses. BMI: body mass index. NTR: Netherlands Twin Register

proteins with BMI at blood sampling and changes in BMI during adolescence and in adulthood (Fig. 1). Here, we refer to BMI at blood sampling and changes in BMI together as BMI trajectories. As changes in BMI in adolescents reflect substantial changes in lean body mass [29], observing an overlap of proteins associated with BMI changes during adolescence with those reported in the literature in adults could indicate whether proteins associated with BMI changes are due to changes in body fat or lean mass. We also elucidated the genetic and environmental sources underlying both the protein abundances and their correlations with BMI and changes in BMI. How genetic differences give rise to differences in plasma protein abundance and BMI trajectories is rooted in a complex interplay of multiple omic layers. To characterize the associations of multiple omics layers with BMI

and changes in BMI, we next investigated whether gene expression of relevant protein-coding genes was associated with BMI trajectories in an external cohort. Finally, we linked identified proteins and their coding genes to plasma metabolites and polygenic risk scores (PRS) for obesity and coronary artery disease (CAD) using mixed-effects models and correlation networks.

**Methods**

**Study population and longitudinal measurements**

*FinnTwin12*

FinnTwin12 is a longitudinal cohort based on a population of Finnish twins born between 1983 and 1987 aimed at investigating adolescent behavioral development and health habits [30, 31]. Participants were identified from the Central Finnish Population Register and completed

four questionnaires (response rate range: 85–90%) at approximate ages 12, 14, 17, and 22. A subset of twins, referred to as intensive subset, was studied more intensively, with additional psychiatric interviews and questionnaires starting at the wave corresponding to age 14. Besides health, lifestyle, and psychological indicators, questionnaires collected up to four self-reported weight and height measures for these participants across the four waves, from which BMI values ( $\text{kg}\cdot\text{m}^{-2}$ ) were calculated. At the 22-year assessment wave, 786 twins from the intensive subset participated in in-person assessments and provided venous blood plasma samples after overnight fasting. We assessed the validity of self-reports of weight and height at the 22-year assessment wave versus in-person measurements in a sample of 756 of these participants (Additional file 2). Self-reported and measured values correlated strongly for weight (Pearson correlation:  $r=0.98$ ) and height ( $r=0.99$ ). From the blood samples, extensive omic data were derived, including proteomics, metabolomics, and genotyping to generate the PRS data used in the current study. Of these participants, 651 had complete longitudinal anthropometric measurements over the 10 years of follow-up and constituted the final FinnTwin12 sample (Table 1). The sample included 250 monozygotic twins (106 complete pairs) and 401 dizygotic twins (164 complete pairs). Height and weight across the waves can be found in the supplementary material (Additional file 3: Table S1).

### Netherlands Twin Register

The Netherlands Twin Register (NTR) is a population-based cohort which includes twins and multiples from the Netherlands [32]. Adult participants of the NTR have completed a survey on health and behavior every 2–3 years since 1991. In contrast to the FinnTwin12 cohort, where data collection is based on age, the NTR data collection in adult twins is independent of age; adult participants are invited to fill out a survey every ~3 years. The blood samples for the omics experiments described in this study were collected in participants of the NTR biobank project (2004–2008) [33]. Venous blood samples were drawn between 7 and 10 AM after overnight fasting, and fertile women had their blood sample drawn in their pill-free week or on day 2–4 of their menstrual cycle. Information on BMI was obtained during the home visit for blood collection as part of the NTR Biobank project. Additionally, we calculated BMI from self-reported height and weight from NTR Survey 5 (2000), Survey 6 (2002), and Survey 7 (2004).

Array-based transcriptomics data were available for 3369 individuals, of which we selected 965 participants for whom blood sampling was performed before age 30 (i.e., age at blood sampling < 30 years old). Of these participants, we retained those for whom BMI was available at the time of blood sampling as well as at least one other BMI measure at Survey 5, Survey 6, or Survey 7. The final NTR sample comprised 665 participants with 2 to 4 BMI

**Table 1** Descriptive statistics of the FinnTwin12 and NTR cohorts

Timepoint	Cohort	N (% female)	Body mass index (BMI)					
			Male participants		Female participants		Age	
			Mean	SD	Mean	SD	Mean	SD
Wave 12	FinnTwin12	651 (59%)	17.7	2.8	17.6	2.4	11.4	0.3
Wave 14	FinnTwin12	651 (59%)	19.5	3.0	19.5	2.6	14.0	0.1
Wave 17	FinnTwin12	651 (59%)	21.8	2.9	21.2	2.8	17.6	0.2
(*) Wave 22	FinnTwin12	651 (59%)	23.9	3.4	22.6	3.6	22.4	0.7
Survey 5	NTR	412 (71%)	22.0	2.9	21.2	2.7	20.7	2.3
Survey 6	NTR	484 (67%)	22.5	2.8	21.6	2.7	23.2	2.4
Survey 7	NTR	501 (70%)	23.3	3.2	21.9	2.9	25.4	2.5
(*) bb1—biobanking	NTR	665 (69%)	23.5	3.2	22.5	3.2	26.8	2.3
Growth factors								
Intercept	FinnTwin12	651 (59%)	17.6	2.4	17.6	2.0		
Slope (nonlinear)	FinnTwin12	651 (59%)	2.3	0.6	2.0	0.6		
Intercept	NTR	665 (69%)	21.9	2.4	21.2	2.2		
Slope (linear)	NTR	665 (69%)	0.3	0.1	0.2	0.1		

Four longitudinal measurements with ~10 and ~6 years of follow-up were collected in the FinnTwin12 and NTR cohorts, respectively. Time points marked with (\*) included blood sampling from which proteomic, metabolomic, and polygenic risk score data were derived. Intercepts and slopes were obtained by Latent Growth Curve Models, in a logarithmic scheme for FinnTwin12 and linear scheme for NTR. Final samples included 651 participants (female 59%) for FinnTwin12 and 665 (female 69%) for NTR. In FinnTwin12, BMI at baseline and last BMI measurement (wave 22) ranged 11.0–30.1 and 17.2–42.0 respectively. In NTR, BMI at baseline and last BMI measurement (bb1) ranged 15.2–40.3 and 15.8–51.3 respectively. BMI is expressed as  $\text{kg}/\text{m}^2$  and age as years

NTR Netherlands Twin Register, N Number of participants without missing value, sd Standard deviation

measurements over a mean follow-up period of ~6 years (Table 1). The sample included 359 monozygotic twins (141 complete pairs) and 306 dizygotic twins (103 complete pairs).

## Omics processing

### Proteomics

Proteins from the plasma samples of FinnTwin12 participants were subjected to Liquid Chromatography-Electrospray Ionization-Mass Spectrometry (LC-ESI-MS/MS) as described previously [34]. First, proteins were precipitated with acetone and subjected to in-solution digestion according to the standard protocol of the Turku Proteomics Facility (Turku Proteomics Facility, Turku, Finland). After digestion, peptides were desalted using a 96-well Sep-Pak C18 plate (Waters), evaporated to dryness, and stored at  $-20^{\circ}\text{C}$ . A commercial kit (High Select™ Top14 Abundant Protein Depletion Mini Spin Columns, cat. Number: A36370, Thermo Scientific) was used to deplete the 14 most abundant proteins from plasma before the proteomic analysis. Samples were first analyzed by independent data acquisition LC-MS/MS using a Q Exactive HF mass spectrometer. Data were further analyzed using Spectronaut software and included local normalization of the data [35], as described elsewhere [34]. Raw matrix counts were  $\log_2$ -transformed and kit-depleted proteins were removed, including human serum albumin (HSA), albumin, IgG, IgA, IgM, IgD, IgE, kappa and lambda light chains, alpha-1 acid glycoprotein, alpha-1 antitrypsin, alpha-2 macroglobulin, apolipoprotein A1, fibrinogen, haptoglobin, and transferrin. None of the participants had scores greater or less than 5 standard deviations from the mean on the first two principal components (PCs) derived from principal component analysis (PCA), indicating none of the participants were identified as outliers. We excluded proteins with >10% missing values. Following the identification of 4 batches, imputation of missing values was performed (proportion of missing values in dataset: 0.86%) using the lowest observed value per batch. Corrections for batch effects were performed with Combat [36]. The final proteomic dataset comprised 439 proteins and protein abundances were scaled, such that one unit corresponded to one standard deviation (sd) with zero mean.

### Transcriptomics

Details of preprocessing can be found in Jansen et al. [37]. In short, heparinized whole blood was transferred into PAXgene Blood RNA tubes (Qiagen, Valencia, Florida, USA) within 20 min of sampling and stored at  $-30^{\circ}\text{C}$ . Total RNA was extracted using the PAXgene Blood RNA MDx kit protocol in 96-well format with the BioRobot Universal System (Qiagen, Valencia, Florida, USA). RNA

quality and quantity was assessed by Caliper AMS90 with HT DNA5K/RNA LabChips [37–39].

Samples were randomly assigned to plates and co-twins were randomized across plates to avoid bias in family correlation estimates. For cDNA synthesis, 50 ng of RNA was reverse transcribed and amplified in a plate format on a Biomek FX liquid handling robot (Beckman Coulter, Brea, California, USA) using Ovation Pico WTA reagents per the manufacturer's protocol (NuGEN, San Carlos, California, USA). Products purified from single-primer isothermal amplification were then fragmented and labeled with biotin using Encore Biotin Module (NuGEN). Prior to hybridization, the labeled cDNA was analyzed using electrophoresis to verify the appropriate size distribution (Caliper AMS90, HT DNA 5 K/RNA LabChip). Samples were hybridized to Affymetrix U219 array plates, and array hybridization, washing, staining, and scanning were carried out per the manufacturer's protocol (GeneTitan, Affymetrix, Santa Clara, California, USA).

Gene expression data were required to pass standard Affymetrix quality control metrics (Affymetrix expression console). Probes were removed when their location was uncertain or if their location intersected a polymorphic single-nucleotide polymorphism (SNP). Expression values were obtained using robust multi-array average normalization implemented in Affymetrix Power Tools (v 1.12.0). We excluded samples that displayed an average Pearson correlation below 0.8 with the probe set expression values of other samples and samples with incorrect sex chromosome expression. In the analyses, we retained only probes for genes encoding proteins associated with BMI or BMI changes in FinnTwin12. We attempted to match protein-coding gene names; however, we failed to retrieve three protein-coding genes (*AMY2A*, *C4A*, and *CFHR1*) identified in FinnTwin12 from the NTR transcriptomic data. We focused on  $\log_2$ -transformed gene-level expression in our analyses, where we obtained gene-level expression by averaging probe-level expression per gene, for each gene associated with protein-coding genes. The same analyses were also performed for  $\log_2$  probe expression as supplementary analyses.

### Metabolomics

Metabolites were quantified from EDTA plasma samples using high-throughput proton nuclear magnetic resonance spectroscopy ( $^1\text{H-NMR}$ ) on a laboratory setup that combines a Bruker AVANCE III 500 MHz and a Bruker AVANCE III HD 600 MHz spectrometer (Nightingale Health Ltd, Helsinki, Finland) [40, 41]. In short, this method provides simultaneous quantification of routine lipids, lipoprotein subclass profiling with lipid concentrations within 14 subclasses, fatty acid composition,

and various low molecular weight metabolites, including amino acids, ketone bodies, and glycolysis-related metabolites in molar concentration units.

Of the 651 participants from the FinnTwin12 sample, 638 had metabolomic data available of which the pre-processing has been described elsewhere [42]. The data included 114 metabolites, of which 13 had missing values. None of the metabolites had more than 10% missing values. Since the distribution of missing values cannot be assumed to be random, we imputed all missing values with the lowest observed value of each metabolic biomarker. Besides LDL cholesterol as derived from the NMR platform, LDL cholesterol was quantified using the Friedewald equation [43]. Similar to the proteomic data, no outliers were detected on the first two PCs. The metabolite values were  $\log_2$ -transformed and scaled so that one unit corresponded to a change of one sd, with a mean of zero.

In the NTR, apart from six non-overlapping metabolites (Additional file 3: Table S2), we only retained metabolites that were significantly associated with BMI trajectories in FinnTwin12. Description for the NTR metabolomics (NMR platform) has been supplied in detail elsewhere [44]. Missing values represented less than 3% of the total data points, and no metabolite had more than 20% missing values, thus imputation was performed by the lowest observed value for each metabolite. Data were  $\log_2$ -transformed, and a technical variable indicating batch was used as a covariate in all relevant analyses.

#### **Polygenic risk scores**

PRSs were calculated and analyzed to test whether genetic susceptibility could explain the association of protein or metabolite levels with BMI trajectories. In FinnTwin12, genotype data for most participants (645 of 651) were available and genotyping was performed using the Illumina Human CNV370-Duo chip. Processing details are available elsewhere [45]. We calculated three PRSs describing the genetic susceptibility to BMI [46], waist to hip ratio adjusted with BMI (WHR) [47], and CAD [48]. After correcting the PRSs for population stratification [49], by regressing out the top ten genetic PCs, the PRSs were scaled to a mean of zero and unit variance.

Of the 665 NTR participants, 576 were genotyped across multiple platforms (Additional file 4) [46, 49–57]. For genotyped NTR participants, we calculated the PRS of BMI [46], using the infinitesimal prior (LDpred-inf) in Ldpred [57]. Detailed information on genotyping and PRS calculation in NTR participants is included in the supplementary material (Additional file 4). The PRS was first corrected for population stratification [49] and for technical batch (i.e., platform) by regressing out the top ten genetic PCs and platform indicator, respectively, and

then the PRS was scaled to have a mean of zero and unit variance.

#### **Statistical analyses**

The primary analyses comprised two sequential steps: (1) calculating growth factors summarizing BMI trajectories and (2) performing association analyses of omics layers with BMI trajectories using mixed-effects models (Fig. 1). Together, these analyses allowed for the joint investigation of longitudinal BMI measures and omics data, associating proteomic, transcriptomic, metabolomic, and PRS data with BMI trajectories. Genetic levers underlying protein abundances and the associations between BMI trajectories and proteins were further explored by univariate and bivariate twin modeling. A multi-omics correlation network was then constructed to put the omics variables associated with changes in BMI into perspective.

#### **Trajectories computation**

*Latent growth curve modeling* Latent growth curve modeling (LGCM) was used to summarize the anthropometric trajectories into intercepts and slopes, corresponding to baseline measurements and rates of change, respectively. LGCM analyses were run with the *lavaan* package version 0.6–12 [58] in R version 4.1.2. To accommodate the shape of the BMI trajectories, for which non-linearity in time is likely, we considered two modeling schemes, corresponding to linear and logarithmic dependence of time, respectively. When fitting the model, BMI measures were linearly adjusted for age differences within waves, with reference ages corresponding to the observed means per wave in the FinnTwin12 and NTR cohorts (Table 1). We tested for adjustment of the growth factors for sex, and further adjusted the model when the effect of sex was significant ( $p$ -value < 0.05). Missing BMI values were allowed during model fitting by maximum likelihood. We calculated robust standard errors to correct for clustering in the data due to family relatedness.

We assessed the goodness of fit of the linear and logarithmic schemes using the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA). A model with RMSEA less than 0.06, CFI greater than 0.95, and TLI greater than 0.95 was considered a good fit [59, 60]. Whenever the linear modeling did not enable good model fit, the logarithmic scheme was preferred when it did enable a good model fit.

*Sensitivity analyses* We carried out two sensitivity analyses. The first sensitivity analysis was performed in FinnTwin12, which consisted of repeating the analyses of

the association between protein levels and baseline BMI or changes in BMI with another marker of adiposity: the triponderal mass index (TMI). TMI is defined as mass divided by height cubed, in contrast to BMI dividing mass by the square of height [61]. BMI is a good indicator of adiposity for most age groups, but TMI is sometimes preferred as an indicator in children and adolescents [61–63].

We performed the second sensitivity analysis in NTR. In NTR, surveys were collected regardless of the age of the participants; thus, each wave comprised all participants that were 18 years or older at the time of the survey. Because LGCM corrected for within-survey age differences, we used a second approach that did not correct for wave-level age differences, which were larger in NTR than in FinnTwin12 (Table 1). Instead of considering reference ages by which ages are corrected, as in the main analyses, we took each individual's actual age at baseline as the first time point and then modeled changes in BMI from this individual's baseline age. We then quantified the associations between these changes in BMI with gene expression.

#### **Mixed-effects models**

To quantify associations between the plasma proteomics data and BMI trajectories, we used mixed-effects models. Mixed-effects modeling involves using two types of effects, fixed and random, the former being constant across individuals and the latter varying between individuals. We tested associations between proteins and BMI at blood sampling and slope of BMI (i.e., BMI changes) in FinnTwin12. We also tested associations between proteins and intercept of BMI (i.e., baseline BMI) as the intercept of BMI was that of young adolescents, whereas the BMI at blood sampling was that of young adults. BMI at blood sampling, changes in BMI, and intercept of BMI were used as dependent variables, and omic variables (e.g., proteins) were used as independent variables. We included age at blood sampling, age squared at blood sampling, sex, and interactions between sex and age and age squared as covariates. BMI intercept was also included as a covariate to correct for baseline BMI differences when testing associations between proteins and changes in BMI. Use of quadratic age and interactions with sex as covariates aimed to ensure adjustment for potential nonlinear relationships with outcomes and combined age–sex confounding on associations, respectively. The model included two random effects whose main function was to correct for clustering in the data due to family relatedness: a binary indicator for zygosity and the family identifier. We tested the association between the coefficients of the fixed effects and each protein using a *t*-test with Satterthwaite approximation [64]

and Bonferroni correction. We considered an association to be significant if the Bonferroni-corrected *p*-value was less than 0.05. Associations between BMI trajectories and protein levels were therefore significant if the nominal *p*-value was less than  $0.05/439 = 1.1 \times 10^{-4}$ . The same modeling was used for identification of metabolites associated with BMI trajectories (Bonferroni: nominal *p*-value  $< 0.05/114 = 4.4 \times 10^{-4}$ ) as well as to quantify associations between PRSs and proteins (Bonferroni: nominal *p*-value  $< 0.05/439 = 1.1 \times 10^{-4}$ ).

We repeated the same modeling in NTR, replacing proteomic variables, being modeled as independent variables, with probe- or gene-level expression of protein-coding genes identified in FinnTwin12, or with metabolites. We added a technical covariate to correct for batches in the modeling between metabolites and BMI trajectories. Multiple correction was applied to the number of tests performed, with the associations of BMI and changes in BMI with omic variables (e.g., metabolite, gene) studied independently. We did not examine the associations between the omics layers and the BMI intercept, as BMI at baseline and BMI at blood sampling reflected both BMI at adulthood. The association between PRS of BMI and gene expression at the probe and gene level was also quantified.

#### **Classical twin modeling**

We applied classical twin models (CTMs) to quantify the genetic and environmental contributions to protein abundances, and to decompose the associations between protein and BMI trajectories into genetic and environmental correlations. CTMs hinge on the comparison between MZ and DZ twins [65, 66], since MZ twins share approximately 100% of the genomic sequence of their DNA, while DZ twins share, on average, only 50% of their segregating genes. By comparing the magnitude of correlations in DZ twin pairs ( $r_{DZ}$ ) to those in MZ twin pairs ( $r_{MZ}$ ), we can infer the presence of genetic factors. In univariate CTM, the total variance (*V*) of a trait can be decomposed into several components: additive genetic (*A*), non-additive or dominant genetic (*D*), shared or common environmental (*C*), and nonshared or unique environmental (*E*). In classical twin modeling of pairs reared together, simultaneous quantification of the *C* and *D* components is not possible. A ratio  $r_{DZ}/r_{MZ}$  above or below 0.5 is therefore an indication that *C* or *D*, respectively, can be modeled. Since the quantification of *D* requires considerable statistical power, we have only estimated *A* and *E*, with *C* where appropriate. Heritability ( $h^2$ ) is defined as  $A/V$ , and the standardized coefficients  $e^2$  and  $c^2$  denote the part of the variance *V* explained by *E* and *C*, respectively [67, 68].

We first fitted saturated models to obtain estimates of the twin correlations by maximum likelihood (ML), including sex and age at blood sampling as covariates. Analyses were conducted in R software using the *OpenMx* package, version 2.20.7 [69–72]. In univariate settings, proteins for which  $r_{MZ}/r_{DZ} < 0.5$  were fitted by AE models. Proteins for which  $r_{MZ}/r_{DZ} > 0.5$  were fitted with ACE models, which were compared with AE, CE, and E only. Negative variance estimates were allowed [73]. The choice for the best model for these proteins was based on Akaike Information Criterion (AIC). Model comparison by AIC is available in the supplementary material (Additional file 3: Table S3).

In bivariate models, we decomposed phenotypically significant covariances between proteins and BMI trajectories. Since twin correlations indicated an AE model for BMI at blood sampling and changes in BMI in the univariate framework, we decomposed their phenotypic correlations with proteins into genetic ( $r_A$ ) and non-shared environmental ( $r_E$ ) correlations in the bivariate framework. The bivariate twin analysis between changes in BMI and protein abundance included the BMI intercept as a covariate in addition to sex and age. Proteins modeled by CE rather than AE (i.e., having  $h^2 = 0$ ) were excluded from the bivariate modeling. Cross-twin cross-trait correlations can be found in the Supplementary Document (Additional file 3: Table S4, Table S5).

#### Multi-omic network construction

A correlation network was constructed to provide a multi-omics framework of the mechanisms underlying BMI change in FinnTwin12, using the significant findings from the prior analyses. The network included the 14 proteins that were associated with the change in BMI and the 3 PRSs. Metabolites associated with the change in BMI were also included, but the set of 48 highly correlated lipoproteins (out of 53 metabolites) were reduced by PCA into 3 principal components (LIPO.PC1, LIPO.PC2, and LIPO.PC3), explaining 87% of the variance. This allowed for better visual representation of the connections between omics. Each of the 25 final entries was linearly regressed on BMI slope, BMI intercept, and BMI at blood sampling, as well as simple and quadratic age at blood sampling, sex, age by sex interaction, and quadratic age by sex interaction. Correlations between these 25 residual variables were used to construct the multi-omics network. We colored the edges according to the significance of the correlation assessed by Pearson's correlation (i.e., nominal  $p$ -value  $< 0.05$  or Bonferroni-corrected  $p$ -value  $< 0.05$ ) and the direction of the correlation (i.e., positive or negative). In NTR, we used a similar methodology to correlate metabolomics residual variables (3 lipoprotein PCs and 5 low molecular weight molecules) with gene expression ( $N = 14$ ).

## Results

### Longitudinal development of BMI in the two cohorts

In FinnTwin12, only the slope of BMI was associated with sex (intercept:  $p = 0.70$ ; slope:  $p < 0.001$ ) and was therefore further adjusted by sex in the modeling. Logarithmic modeling achieved the best model fits (RMSEA: 0.05 [95% confidence interval (CI) 0.03, 0.06], TLI = 0.97, CTI = 0.98) compared to linear modeling (RMSEA: 0.09 [95% confidence interval (IC) 0.07, 0.11], TLI = 0.93, CTI = 0.94). We therefore kept logarithmic modeling for the calculation of growth factor values. The mean BMI intercept was  $17.6 \text{ kg}\cdot\text{m}^{-2}$ . Male and female participants gained an average of 6.2 and 5.0 BMI units, respectively, during the 10-year follow-up period (Table 1).

In NTR, both the intercept and the slope were sex adjusted (intercept:  $p = 0.01$ ; slope:  $p = 0.02$ ) and linear modeling provided the best model fits (RMSEA: 0.04 [95% CI 0.02, 0.06], TLI = 0.99; CTI = 0.99). Table 1 describes the mean intercepts and slopes. Only twelve NTR participants had a decrease in BMI over the course of follow-up, and overall, the change in BMI in NTR showed substantial weight gain at the population level ( $z$ -value = 3.3,  $p = 0.001$ ). During the 6-year follow-up, male and female NTR participants gained an average of 1.5 and 1.3 BMI units, respectively (Table 1). Growth factor variance was significantly non-zero in both FinnTwin12 and NTR datasets ( $z$ -values  $> 3.0$ ,  $p < 0.01$ ), indicating significant inter-individual differences longitudinally.

### Proteins and BMI trajectories in FinnTwin12

#### Phenotypic associations

We examined associations between plasma protein abundance and BMI trajectories (i.e., BMI at blood sampling and slope of BMI) from mixed-effects models using scaled BMI trajectory markers (Fig. 1). Proteins were defined as independent variables along with covariates in mixed-effects modeling. A total of 66 proteins out of 439 proteins showed a significant association with BMI at blood sampling, of which 42 had a negative and 24 had a positive association with BMI (Table 2). Proteoglycan 4 had the strongest association with BMI (coefficient: 0.46;  $p = 2.1\text{E} - 34$ ); an increase of one sd in the abundance of proteoglycan 4 was associated with an increase of almost a half of sd in BMI (Table 2).

A total of 14 protein abundances were associated with changes in BMI, independent of baseline BMI levels (Table 3). These proteins included complement factors and proteoglycan 4, with proteoglycan 4 having the strongest association with adolescent BMI changes (estimate: 0.27;  $p = 4.3\text{E} - 10$ ). Ten proteins had a positive association with changes in BMI and 4 had negative associations. Thirteen out of 14 proteins were also correlated with BMI at blood sampling.



**Table 2** Mixed-effects model-derived estimates of proteins significantly associated with body mass index at blood sampling in FinnTwin12 participants

Protein description	UniProt ID	Estimate	SE	t-value	P-value	
					Nominal	Bonferroni
Proteoglycan 4	Q92954	0.46	0.03	13.6	4.70E-37	2.10E-34
Complement factor H	P08603	0.33	0.04	9.1	9.00E-19	4.00E-16
Serum amyloid P-component	P02743	0.34	0.04	9.0	1.70E-18	7.60E-16
C-reactive protein	P02741	0.31	0.04	8.5	1.30E-16	5.60E-14
Secreted phosphoprotein 24	Q13103	-0.30	0.04	-8.1	3.40E-15	1.50E-12
Complement component C7	P10643	-0.28	0.04	-7.5	2.30E-13	1.00E-10
Sex hormone-binding globulin	P04278	-0.33	0.05	-7.2	1.50E-12	6.60E-10
Antithrombin-III	P01008	-0.25	0.04	-6.7	4.30E-11	1.90E-08
Complement factor I	P05156	0.23	0.04	6.5	2.20E-10	9.80E-08
Complement C4-A	P0COL4	0.23	0.04	6.3	6.50E-10	2.90E-07
C4b-binding protein alpha chain	P04003	0.23	0.04	6.2	1.20E-09	5.40E-07
Corticosteroid-binding globulin	P08185	-0.24	0.04	-5.9	6.20E-09	2.70E-06
Protein S100-A8	P05109	0.21	0.04	5.8	9.30E-09	4.10E-06
Fumarylacetoacetase	P16930	0.20	0.03	5.8	1.40E-08	6.10E-06
Apolipoprotein B-100	P04114	0.21	0.04	5.7	2.00E-08	8.60E-06
Inter-alpha-trypsin inhibitor heavy chain H2	P19823	-0.20	0.04	-5.6	2.80E-08	1.20E-05
Pancreatic alpha-amylase	P04746	-0.20	0.04	-5.6	3.40E-08	1.50E-05
72 kDa type IV collagenase	P08253	-0.20	0.04	-5.5	6.50E-08	2.80E-05
Contactin-1	Q12860	-0.20	0.04	-5.4	8.50E-08	3.70E-05
Interleukin-1 receptor accessory protein	Q9NPH3	-0.20	0.04	-5.4	1.10E-07	4.70E-05
Apolipoprotein D	P05090	-0.20	0.04	-5.3	1.30E-07	5.80E-05
Properdin	P27918	0.20	0.04	5.3	1.70E-07	7.30E-05
Serum amyloid A-4 protein	P35542	0.19	0.04	5.2	2.20E-07	9.90E-05
Kallistatin	P29622	-0.19	0.04	-5.2	2.70E-07	1.20E-04
ADAMTS-like protein 4	Q6UY14	-0.19	0.04	-5.2	2.70E-07	1.20E-04
Limbic system-associated membrane protein	Q13449	-0.19	0.04	-5.2	3.00E-07	1.30E-04
Coagulation factor XIII A chain	P00488	-0.18	0.04	-5.1	5.00E-07	2.20E-04
Vascular cell adhesion protein 1	P19320	-0.19	0.04	-5.0	7.50E-07	3.30E-04
Cholinesterase	P06276	0.21	0.04	5.0	9.00E-07	3.90E-04
Endoglin	P17813	-0.18	0.04	-5.0	9.40E-07	4.10E-04
Fibulin-1	P23142	-0.18	0.04	-4.9	1.20E-06	5.30E-04
Insulin-like growth factor-binding protein 2	P18065	-0.18	0.04	-4.9	1.50E-06	6.50E-04
Mast/stem cell growth factor receptor kit	P10721	-0.18	0.04	-4.8	1.60E-06	7.10E-04
Serum amyloid A-1 protein	P0DJJ8	0.18	0.04	4.8	1.80E-06	8.00E-04
Inter-alpha-trypsin inhibitor heavy chain H1	P19827	-0.18	0.04	-4.8	2.10E-06	9.20E-04
Transthyretin	P02766	-0.19	0.04	-4.8	2.20E-06	9.50E-04
Serum amyloid A-2 protein	P0DJJ9	0.18	0.04	4.8	2.20E-06	9.70E-04
C4b-binding protein beta chain	P20851	0.18	0.04	4.7	3.90E-06	1.70E-03
Protein S100-A9	P06702	0.16	0.04	4.6	4.60E-06	2.00E-03
Transforming growth factor-beta-induced protein ig-h3	Q15582	-0.17	0.04	-4.6	6.10E-06	2.70E-03
Gelsolin	P06396	-0.18	0.04	-4.5	6.90E-06	3.00E-03
Complement component C8 gamma chain	P07360	0.17	0.04	4.4	1.00E-05	4.60E-03
Cadherin-13	P55290	-0.17	0.04	-4.4	1.10E-05	4.60E-03
Voltage-dependent calcium channel subunit alpha-2/delta-1	P54289	-0.16	0.04	-4.4	1.60E-05	6.80E-03
Apolipoprotein A-IV	P06727	-0.17	0.04	-4.3	1.80E-05	8.00E-03
Serum paraoxonase/lactonase 3	Q15166	-0.16	0.04	-4.3	2.20E-05	9.40E-03
Basement membrane-specific heparan sulfate proteoglycan core protein	P98160	-0.16	0.04	-4.3	2.20E-05	9.80E-03

**Table 2** (continued)

Protein description	UniProt ID	Estimate	SE	t-value	P-value	
					Nominal	Bonferroni
Transketolase	P29401	0.16	0.04	4.2	2.80E-05	1.20E-02
ADAM DEC1	O15204	-0.15	0.03	-4.2	3.00E-05	1.30E-02
Gamma-glutamyl hydrolase	Q92820	0.15	0.04	4.2	3.10E-05	1.40E-02
4F2 cell-surface antigen heavy chain	P08195	-0.16	0.04	-4.2	3.50E-05	1.50E-02
SPARC-like protein 1	Q14515	-0.15	0.04	-4.2	3.60E-05	1.60E-02
Complement factor H-related protein 1	Q03591	0.15	0.04	4.1	4.00E-05	1.70E-02
Bone marrow proteoglycan	P13727	-0.15	0.04	-4.1	4.00E-05	1.80E-02
Apolipoprotein A-II	P02652	-0.15	0.04	-4.1	4.70E-05	2.10E-02
Adiponectin	Q15848	-0.15	0.04	-4.1	4.70E-05	2.10E-02
Neural cell adhesion molecule L1-like protein	O00533	-0.16	0.04	-4.1	4.90E-05	2.20E-02
Complement C3	P01024	0.15	0.04	4.1	5.10E-05	2.30E-02
Complement factor B	P00751	0.15	0.04	4.1	5.50E-05	2.40E-02
Insulin-like growth factor I	P05019	-0.15	0.04	-4.0	6.00E-05	2.60E-02
Prolow-density lipoprotein receptor-related protein 1	Q07954	-0.15	0.04	-4.0	6.60E-05	2.90E-02
Fructose-bisphosphate aldolase B	P05062	0.14	0.04	4.0	7.20E-05	3.20E-02
Vitamin D-binding protein	P02774	-0.15	0.04	-4.0	7.80E-05	3.40E-02
Carboxypeptidase N catalytic chain	P15169	-0.16	0.04	-3.9	8.70E-05	3.80E-02
Neural cell adhesion molecule 1	P13591	-0.15	0.04	-3.9	9.70E-05	4.30E-02
Metalloproteinase inhibitor 2	P16035	-0.14	0.04	-3.9	1.00E-04	4.60E-02

Protein estimates whose abundance were significantly associated with body mass index are shown if Bonferroni-corrected coefficient null-test *p*-values were less than 0.05. Mixed-effects models included as covariates: age at blood sampling, age squared, sex, and interactions of sex with age and age squared. Random effects included the zygosity indicator variable and family identifiers. Body mass index was z-scored; fixed effects estimates indicate effect on 1 sd change  
se Standard error

**Table 3** Mixed-effects model-derived estimates of proteins significantly associated with the changes (slope) of body mass index over 10 years of follow-up in FinnTwin12 participants

Protein description	UniProt ID	Estimate	SE	t-value	P-value	
					Nominal	Bonferroni
Proteoglycan 4	Q92954	0.27	0.04	7.3	9.80E-13	4.30E-10
Complement factor H	P08603	0.25	0.04	6.6	7.70E-11	3.40E-08
Cholinesterase	P06276	0.24	0.04	5.6	3.00E-08	1.30E-05
Complement component C7	P10643	-0.20	0.04	-5.2	2.40E-07	1.10E-04
Sex hormone-binding globulin	P04278	-0.24	0.05	-5.1	4.50E-07	2.00E-04
C4b-binding protein alpha chain	P04003	0.18	0.04	4.8	1.70E-06	7.60E-04
Serum amyloid P-component	P02743	0.18	0.04	4.6	6.40E-06	2.80E-03
Complement factor I	P05156	0.16	0.04	4.5	1.00E-05	4.40E-03
Complement factor B	P00751	0.16	0.04	4.3	1.90E-05	8.10E-03
Phosphatidylcholine-sterol acyltransferase	P04180	0.16	0.04	4.3	2.30E-05	9.90E-03
Secreted phosphoprotein 24	Q13103	-0.17	0.04	-4.3	2.20E-05	9.90E-03
C-reactive protein	P02741	0.16	0.04	4.2	2.70E-05	1.20E-02
Properdin	P27918	0.15	0.04	4.0	7.10E-05	3.10E-02
Inter-alpha-trypsin inhibitor heavy chain H2	P19823	-0.15	0.04	-4.0	8.10E-05	3.60E-02

Protein estimates whose abundance were significantly associated with changes in body mass index (i.e., slope of body mass index) are shown if Bonferroni-corrected coefficient null-test *p*-values were less than 0.05. Mixed-effects models included as covariates: the intercept of body mass index, age at blood sampling, age squared, sex, and interactions of sex with age and age squared. Random effects included the zygosity indicator variable and family identifiers. The slope of body mass index was z-scored; fixed effects estimates indicate effect on 1 sd change  
se Standard error

Similarly, we investigated associations between proteins and BMI intercept (i.e., baseline BMI) (Additional file 3: Table S6), of which 11 significant associations were detected. Of these, 9 were also associated with BMI at blood sampling. These findings suggest that the proteome in young adults contains proteins associated with both childhood and adulthood BMI. Proteoglycan 4, complement factor H, C-reactive protein, and secreted phosphoprotein 24 were associated with BMI at blood sample collection, BMI intercept, and BMI slope.

Associations between proteins with triponderal mass index (TMI) at baseline (~12 years old) and slope of TMI can be found in the supplementary material (Additional file 3: Table S7). In total, we found 10 and 22 proteins associated with baseline TMI and TMI changes, respectively. Seven of the ten proteins associated with baseline TMI were also associated with baseline BMI. All proteins associated with BMI changes were also associated with TMI changes. This shows that BMI and TMI associations with the plasma proteome tend to have similar proteins involved.

#### **Genetic and environmental sources underlying protein abundance**

Univariate twin modeling was applied to quantify the genetic and environmental sources of variance in the 66 proteins identified as associated with either BMI at blood sampling and/or the slope of BMI (Fig. 2A). According to the ratio between  $r_{MZ}$  and  $r_{DZ}$  (Fig. 2B), we obtained 13 CE models, 1 ACE model, and 52 AE models. On average, genetic and nonshared environmental factors explained 35% (range: 0–78) and 59% (range: 22–89) of the variance in protein abundances, respectively (Fig. 2C). Thirteen proteins showed no evidence for genetic effects ( $h^2=0$ ), and all of these had non-zero shared environmental components that accounted for an average of 28% (range: 19–41) of the variance, with nonshared environmental factors accounting for 72% (range: 59–81) of the variance in the abundances of these proteins (Fig. 2A). Heritability of BMI at blood sampling was 72% [95% CI: 62,78], and the nonshared environment accounted for the remaining (28%) of the variance in BMI at blood sampling. Heritability of the slope of BMI was 63% [95% CI: 51,73] while the nonshared environment explained the remaining (37%) variance in BMI changes. All univariate estimates and their 95% CIs are available in the supplementary material (Additional file 3: Table S3, Table S8).

#### **Genetic and environmental correlations between BMI trajectories and proteins**

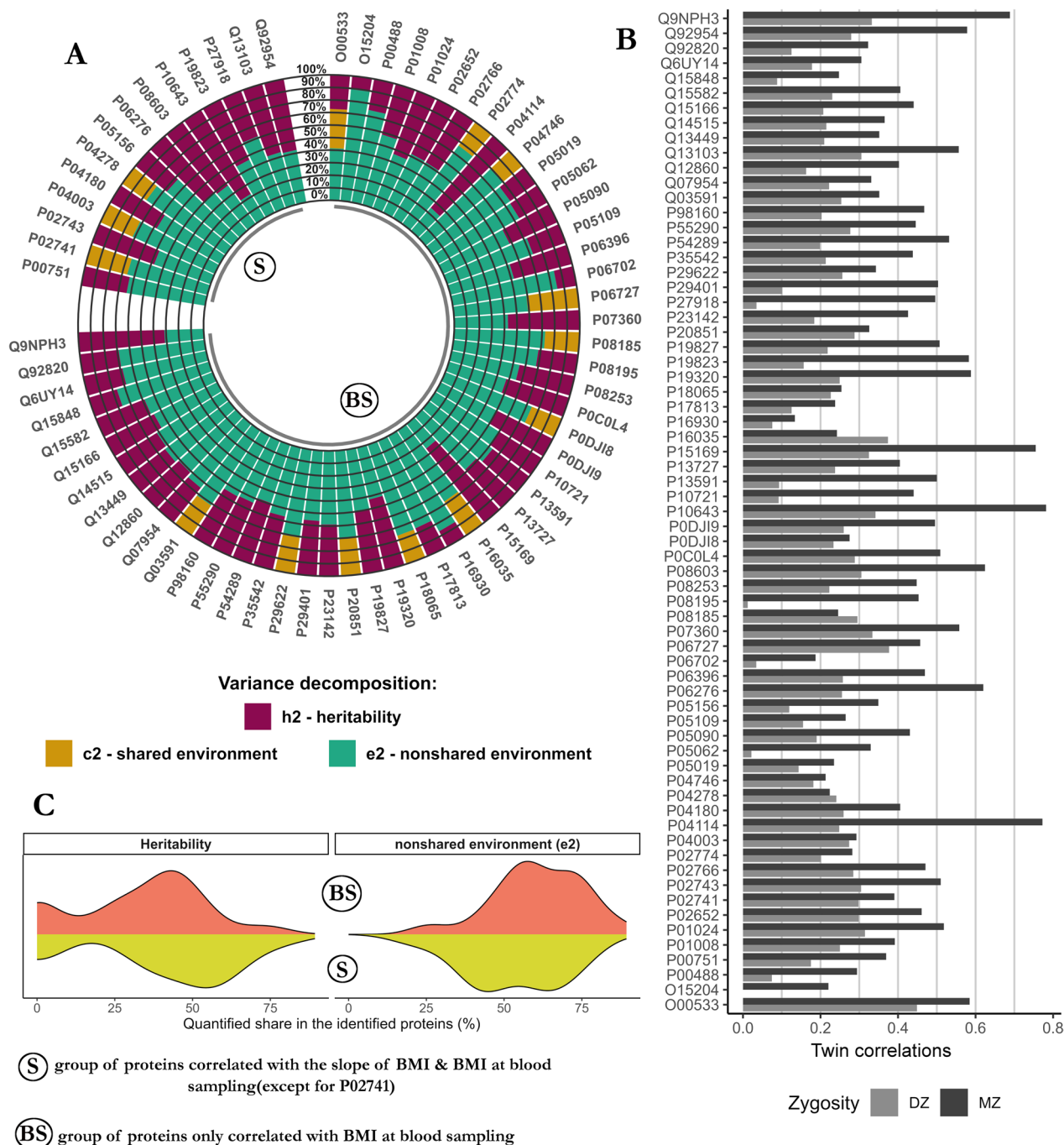
The decomposition of significant phenotypic covariation between plasma protein abundances ( $h^2>0$ ) with

BMI and BMI changes were performed with bivariate genetic twin modeling. For all associations between proteins and BMI at blood sampling or BMI change, we applied bivariate AE models. Of the 53 BMI-protein correlations tested, we observed 43 significant genetic correlations and 12 significant nonshared environmental correlations (Fig. 3A). For 8 protein–BMI associations, we observed both significant genetic and nonshared environmental correlations. Overall, genetic correlations ranged from  $-0.50$  to  $0.50$  (absolute mean:  $r_A=0.29$ ) and environmental correlations from  $-0.33$  to  $0.44$  (absolute mean:  $r_E=0.14$ ). Seven proteins had genetic correlations with BMI greater than 0.4 in absolute value, and proteoglycan 4 had the highest nonshared environmental correlation with BMI among all proteins tested ( $r_E=0.44$  [95% CI 0.28,0.57]). The full set of genetic and nonshared environmental correlations is available in the supplemental material (Additional file 3: Table S9).

Similarly, the significant associations between proteins and changes in BMI were decomposed into genetic and nonshared environmental correlations for 11 of the 14 proteins associated with changes in BMI (Fig. 3B). We observed 7 significant genetic correlations and 3 significant nonshared environmental correlations (Additional file 3: Table S10). Because BMI change during adolescence predicted adult BMI independently of baseline BMI (linear regression:  $R^2=45\%$ ), we examined whether the association between BMI at blood sampling and BMI change might confound the observed correlations between proteins and BMI. After adjusting the slope of BMI by BMI at blood sampling before bivariate twin modeling (Fig. 3B), only apolipoprotein A-IV remained genetically correlated with the slope of BMI ( $r_A=0.20$  [95% CI 0.02, 0.38]), while all other correlations were no longer significant (Additional file 3: Table S11). These results suggest that the genetic and environmental factors influencing the association between changes in BMI during adolescence and protein levels in adulthood may reflect associations with adult BMI.

#### **Gene expression and BMI trajectories in NTR**

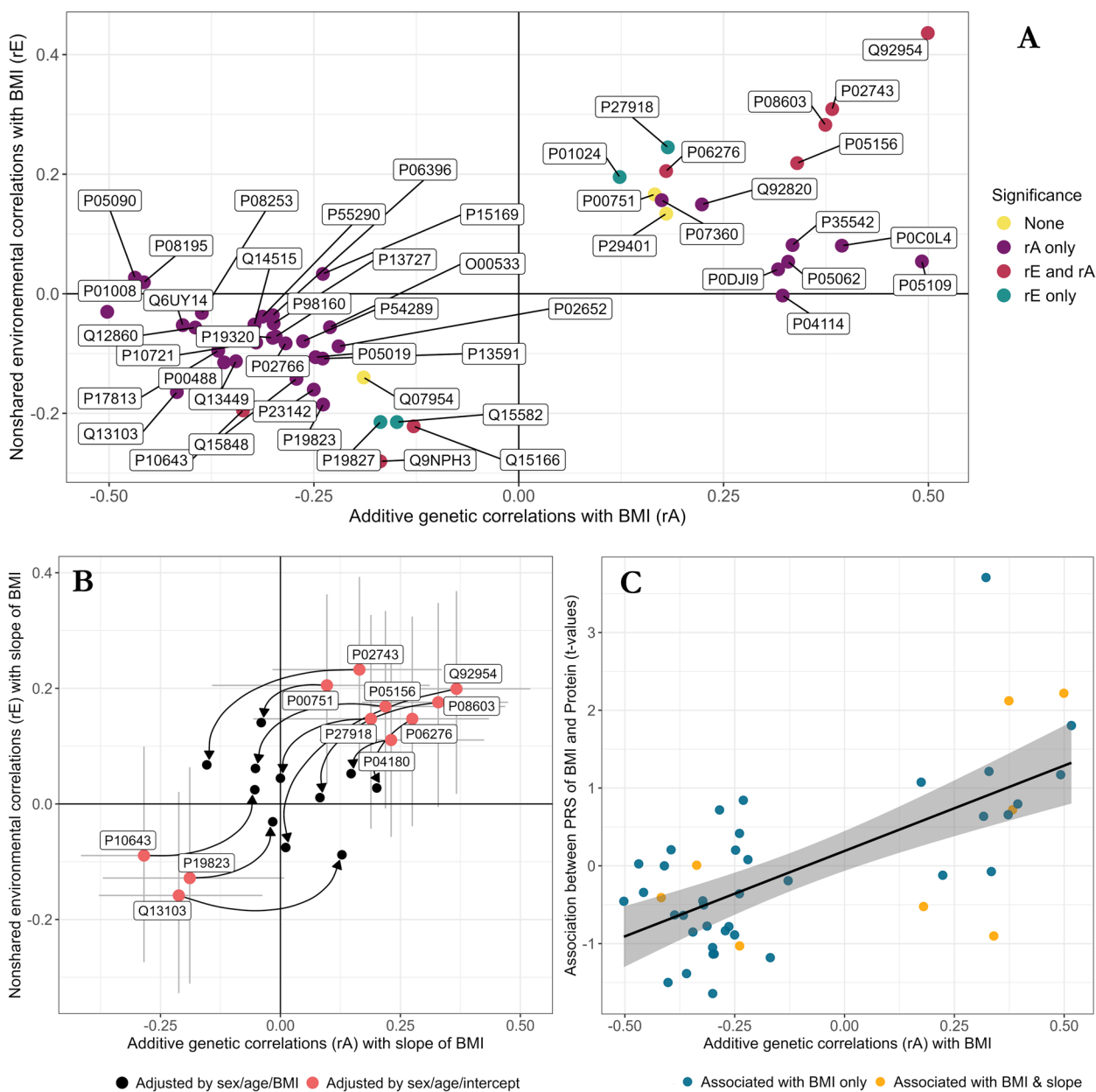
Mixed-effects models quantified the associations of BMI trajectories with the expression of genes encoding the proteins identified in FinnTwin12. *S100A8* (protein encoded: S100 Calcium Binding Protein A8) was significantly associated with BMI at blood sampling (estimate=0.21; sd=0.05;  $p=1.5E-03$ ). The *PRG4* gene (protein encoded: proteoglycan 4) was significantly associated with the slope of BMI (estimate=0.09; sd=0.03;  $p=0.05$ ). While the association of the *CFI* gene (protein encoded: complement factor I) with the slope of BMI did not pass Bonferroni correction ( $p=0.21$ ), one



**Fig. 2** Genetic and environmental sources underlying proteins for which plasma abundance was associated with BMI or BMI change in the FinnTwin12 sample. **A** Univariate twin models were used to decompose the variance of proteins identified as associated with BMI or BMI changes in FinnTwin12. **B** The models were determined by the ratio of the correlations in MZ and the correlations in DZ twins. **C** Fractions attributed to nonshared environmental factors and to heritability. BMI: body mass index

of its probes (ID: 11718481\_s\_at; start: chr4:110661852; end: chr4:110663751) was significantly associated with changes in BMI ( $p=0.02$ ) at the probe level. Summary statistics are available in the supplementary

material (Additional file 3: Table S12, Table S13, Table S14, Table S15), and estimates from the sensitivity analysis are also available in the supplementary material (Additional file 3: Table S16, Table S17).



**Fig. 3** Genetic and nonshared environmental correlations between BMI trajectories and protein levels in the FinnTwin12 sample. **A** Among the 53 heritable proteins for which associations with BMI were significant, bivariate twin models were conducted to dissect the nature of the associations. **B** Genetic and nonshared environmental correlations were also quantified in the association between slope of BMI and protein levels; lines correspond to confidence intervals at the 95% threshold, arrows point to estimates obtained by adjusting protein levels by BMI at blood sampling prior to bivariate twin modeling rather than using BMI intercept as a covariate (pink). **C** The direction of the genetic correlations observed between BMI at blood sampling and protein levels were consistent with the  $t$ -values observed by crossing the PRS of BMI with protein levels. BMI: body mass index. PRS: polygenic risk score

### Metabolites and BMI trajectories

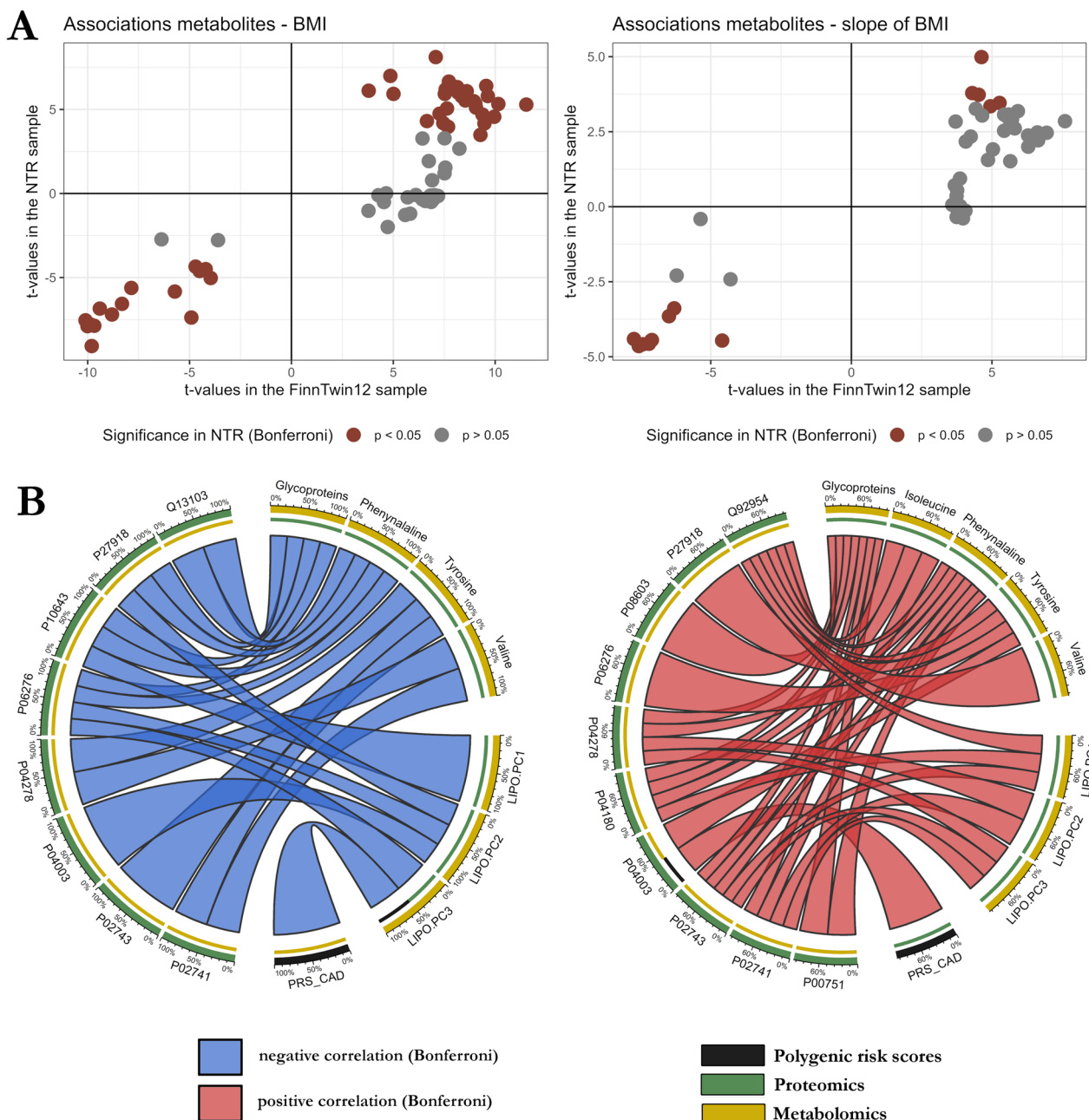
A total of 78 out of 114 metabolites were significantly associated with BMI at blood sampling in FinnTwin12. The mean absolute effect, which is 1.4 times greater than that observed for proteins, was 0.28 sd change in BMI per 1 sd change in metabolite level. Of the 78 metabolites,

68 were lipoproteins. The rest included Friedewald LDL cholesterol [43] and 9 low molecular weight molecules (LMWM): citrate, glycerol, glycoprotein acetyls, isoleucine, leucine, valine, CH<sub>2</sub> groups of mobile lipids, phenylalanine, and tyrosine. In NTR, 44 out of the 78 metabolites were also significantly associated with BMI

at blood sampling (Fig. 4A) (Additional file 3: Table S18, Table S2).

In FinnTwin12, 53 metabolites were associated with the slope of BMI, and all of these were also associated with BMI at the time of blood sampling. These metabolites comprised 48 lipoproteins and five LMWMs: glycoprotein acetyls, isoleucine, valine, phenylalanine, and tyrosine. In

NTR, 14 of the 53 metabolites were also significantly associated with the BMI slope (Fig. 4A), of which 13 were lipoproteins besides glycoprotein acetyls. In FinnTwin12, 19 metabolites were associated with the intercept of BMI (i.e., baseline BMI; baseline age: ~12 years). Summary statistics are available in the supplementary material (Additional file 3: Table S19, Table S20, Table S21).



**Fig. 4** Extended cross-omic analyses. **A** The  $t$ -values of the associations between metabolites with BMI and BMI change have proportional directions between the two samples NTR and FinnTwin12. **B** Multi-omic correlations between PRSs with proteins and metabolites associated with BMI change in the FinnTwin12 sample, independent of BMI trajectory markers (BMI blood sampling, BMI slope and BMI intercept). PRS: polygenic risk score. BMI: body mass index. NTR: Netherlands Twin Register

### Polygenic risk scores and BMI-associated omics traits

We first examined the association of PRSs with proteins identified as associated with BMI and changes in BMI in FinnTwin12. Of the 66 proteins associated with BMI at blood sampling, none were associated with the PRS of BMI after Bonferroni correction ( $p > 0.05$ ) (Additional file 3: Table S22). However, we observed consistent directions of effect between the genetic correlations of proteins and BMI and the associations between the PRS of BMI and proteins (Fig. 3C). In NTR, we quantified the association of BMI protein-encoding gene expression with the PRS of BMI, but found no significant associations. We then compared the  $t$ -values of associations between gene expression and the PRS of BMI with the  $t$ -values of associations between gene expression and BMI, and found weak evidence for consistency in the direction of the compared  $t$ -values (linear regression coefficient nullity-test:  $p = 0.015$ ).

In FinnTwin12, we also investigated the associations between proteins and the PRSs of CAD and WHR (Additional file 3: Table S22). None of the associations between proteins and the PRS of WHR were significant, with P06702 (Bonferroni-corrected  $p$ -value = 0.06) and Q15848 ( $p = 0.10$ ) having the lowest Bonferroni-corrected  $p$ -values. We found two significant associations between proteins and the PRS of CAD: C4b-binding protein alpha chain (protein ID: P04003; nominal  $p$ -value =  $6.0E-04$ ; Bonferroni  $p$ -value =  $4.0E-02$ ) and Apolipoprotein B-100 (protein ID: P04114; nominal  $p$ -value =  $5.2E-04$ ; Bonferroni  $p$ -value =  $3.4E-02$ ).

We also conducted additional analyses between BMI protein-encoding gene expression and PRS of BMI (Additional file 3: Table S23, Table S24), between PRSs and proteins associated with the slope of BMI in the FinnTwin12 cohort (Additional file 3: Table S25), between the PRS of BMI and genes encoding the proteins identified associated with the slope of BMI in NTR (Additional file 3: Table S26, Table S27), and between PRSs and metabolites in both cohorts (Additional file 3: Table S28, Table S29).

### Multi-omics correlation network

To examine the association of the proteome, metabolome, and PRSs with BMI trajectories, we constructed a multi-omics correlation network in FinnTwin12 (Fig. 4B) with previously identified proteins and metabolites adjusted for covariates and BMI at blood sampling, slope of BMI, and BMI intercept. We observed 49 correlations between omic layers, of which 30 were positive and 19 negative (Fig. 4B). Apart from associations between the PRS of CAD with P04003 and LIPO.PC3, the remainder were metabolome-proteome associations. Twelve of the fourteen proteins associated with changes in BMI shared at least one significant correlation with a residual metabolomic variable, the average number of

connections being 3.9 out of 8 possible connections for these proteins. These correlations indicate multi-omic relationships between omics variables associated with a 10-year change in BMI, independent of BMI trajectories markers. We also examined the associations between the transcriptome and metabolome in NTR, but found no significant multi-omics relationships (Additional file 3: Table S30, Table S31).

### Discussion

Overall, a substantial number of plasma proteins were associated with adult BMI and BMI change during adolescence. The estimates for the heritability of plasma abundance of the identified proteins varied. Some proteins were influenced by shared (i.e., familial) environmental factors rather than genetics. The etiology of the observed protein–BMI trajectory associations revealed genetic and/or nonshared environmental correlations between BMI and protein levels. Few associations between blood expression levels of protein-encoding genes and BMI trajectories were observed in NTR adults. This suggests a potential for observing gene-protein pairs associated with BMI trajectories at two different omic levels, among both adolescents and adults from two different countries. Finally, the use of metabolomic data and PRSs allowed both the observation of cross-omic associations and a better understanding of the connections that link BMI trajectories to different omic layers.

The twin design allowed the quantification of the genetic and environmental sources of variation underlying the associations between BMI trajectories and the proteome from an epidemiologic perspective. Including two cohorts of different ages and populations also empowered the identification of biomolecules likely to be stable across generations and populations. This is in line with two longitudinal multi-omics studies that identified stable biomolecules in response to interventions and investigated proteome resistance to weight change [18, 19]. Although BMI measurements at blood sampling were similar between the two cohorts (Table 1), both the length of follow-up and the intensity of BMI changes between the two cohorts differed. This may have limited our ability to compare biomolecule associations with changes in BMI between cohorts. Moreover, the relatively small sample sizes of the cohorts in the current study might have limited the statistical power to detect associations with PRSs or transcriptomic data. Using less conservative multiple testing corrections and cohorts with larger sample sizes may allow more associations to be detected. However, availability of larger multi-omics twin data is limited. An increase in the size and number of omics data in twin cohorts is therefore necessary to realize the full potential of twin designs.

BMI changes in FinnTwin12 adolescents were substantial from early adolescence to young adulthood in both male and female participants. These changes reflect multiple processes such as increase in lean body mass, and increase in body fat as growth continued from pre-puberty onwards. Longitudinal omics data would allow for a more in-depth investigation of the relationship that links changes in BMI to the proteome or metabolome. Investigations among adolescents need to evaluate changes in body composition, while among adults BMI weight gain is primarily change in fat depots. Beyond describing body composition, it is still needed to characterize the causal relationships between changes in BMI and proteins or metabolites, for which our design is not suited. Therefore, we reported associations between BMI changes and omics variables, but could not infer their direction (i.e., whether changes in BMI affected proteins and metabolites, or vice versa). Twin designs combining MR and direction of causation (DOC) twin models, e.g., MR-DOC models, could be used for this purpose [74]. However, there is a lack of sufficiently large genome-wide association studies of BMI changes, which precludes the use of sufficiently robust instrumental variables.

Complement factors associated with BMI at blood sampling in FinnTwin12, including complements I, B, and H, have been associated with BMI previously [11, 12, 75]. Other proteins with strong associations with BMI at blood sampling ( $|\text{estimate}| > 0.3$ ; Table 2) in FinnTwin12 replicate findings from the literature, such as C-reactive protein [11, 12, 18, 75], sex hormone-binding globulin [11, 12, 75], and proteoglycan 4 [76]. Several proteins associated with BMI at blood sampling have also been reported in the literature to be causally associated with BMI. These include those encoded by the *CRP*, *NCAM1*, *IGFBP2*, or *SHBG* genes [11, 12], the last four of which are causally influenced by BMI (i.e., BMI-to-protein associations) [12].

We identified several of the proteins associated with changes in BMI (i.e., gains in BMI) during adolescence that were also identified in adult populations that experienced weight loss. For example, proteoglycan 4, which was most strongly associated with BMI changes in our study, was previously associated with weight loss in adults [25] with the same direction of effect, i.e., positive covariation between change in BMI and proteoglycan 4 level. C-reactive protein, although previously shown to rapidly respond to weight loss [25], was also associated with changes in BMI over a 10-year follow-up in our study. These observations suggest the proteins identified in adult populations are, at least in part, relevant to the study of adolescents. Moreover, we observed that the sign of the effect between protein

levels and BMI change echoed those in the literature, i.e., positive associations with BMI gain were negative with weight loss and vice versa. This indicates a stability of certain biomolecules in the proteome to characterize changes in BMI across ages, both in the study of weight gain and weight loss. Weight loss involves changes in lean mass as well as in body fat [77], which would be analogously the reverse regarding growth in adolescence.

Translating findings from observational studies of weight change in adolescents into clinical practice requires additional efforts. Whether proteins causally influence or are influenced by weight change, and how quickly the proteome responds to changes in diet or increased physical activity is to be considered. Furthermore, we reported proteins associated with gains in BMI among relatively healthy participants over a long follow-up period. Whether these proteins would remain associated with short-term body fat loss among adolescents with obesity is not assured. Additional studies in younger populations, including proteomic weight loss interventions, would be highly valuable designs for the identification of optimal biomarkers for clinical use of proteomics data.

Overall, we found few proteins for which the expression of their coding genes were associated with BMI or changes in BMI. Post-transcriptional modifications affecting proteins, e.g., the post-transcription addition of chemical groups or polypeptides [78, 79], can explain this lack of association, as they increase the complexity of the transcriptome–proteome connection. However, we observed that the gene expression of *S100A8* (protein: S100A8), *PRG4* (protein: proteoglycan 4), and *CFI* (protein: complement factor 1) were associated with BMI at blood sampling, BMI changes at the gene level, and BMI changes at the probe level, respectively. The *S100A8* gene encodes the eponymous protein belonging to the S100 family corresponding to  $\text{Ca}^{2+}$ -binding proteins widely known to be associated with BMI [80] in both adults [81, 82] and children [83]. *S100A8* expression levels have previously been associated with BMI [84], echoing the dual association with BMI at both the transcriptome and proteome levels in our study. In contrast, no previous study reported associations between the gene expression of *CFI* and *PRG4* genes with changes in BMI to our knowledge.

Heritability estimates of BMI trajectories in FinnTwin12 adolescents ( $h^2=63\%$ ) were slightly larger than those observed in adult Finnish twins ( $h^2=52\text{--}57\%$ ) [85], which is consistent with the higher heritability of BMI reported in younger populations [6] and further motivates the study of factors genetically associated with BMI changes during adolescence. Of body mass, fat accounts for a



smaller proportion in children and adolescents in general, while gaining fat is more typical in adult populations. The previous estimates for the heritability of plasma protein abundance associated with BMI trajectories appear to be rare in the literature. Liu et al. reported that the heritability of 342 plasma proteins in a relatively small sample of older twins [86] varied widely, which is in line with our observations. Other studies have quantified the heritability of disease-associated plasma proteins [87], but none seem to focus on obesity. The heritability estimates of the proteins identified in our study therefore constitute a valuable resource for the further study of the proteome–obesity relationship.

C4b-binding protein alpha chain (Bonferroni  $p$ -value =  $4.0E-02$ ) and apolipoprotein B-100 ( $p = 3.4E-02$ ) were associated with the PRS of CAD, with apolipoprotein B-100 being genetically correlated to BMI ( $r_A = 0.32$ ). While the links between C4b-binding protein alpha chain and CAD are new, apolipoprotein B-100 is genetically associated with cardiovascular disease [88, 89], as well as chronic kidney disease, blood pressure, and various lipids [90, 91]. A better understanding of the mechanisms induced or inferred by this protein could provide insights into both obesity and its co-morbidities.

We identified many metabolites associated with BMI and BMI changes in FinnTwin12, of which we replicated a substantial proportion in NTR. These metabolites included mostly lipoproteins and branched-chain amino acids with known associations with BMI changes [27]. Despite the strong connections observed between the metabolome and proteome in the study of BMI trajectories, we found no significant connections between the metabolome and the transcriptome. Despite the lack of results, which could be explained by statistical power issues, it is still unclear how the different omics layers are connected biologically. Having omics from multiple tissues would be one approach, but perhaps feasible only in model organisms. The identification of mediation mechanisms between omics layers may help in acquiring a better understanding of the multi-omics mechanisms underlying obesity [92, 93].

## Conclusions

In conclusion, the proteome is a promising resource for cross-sectional and longitudinal studies of obesity. Both environmental and genetic factors influence the proteome–obesity association, and integrating multiple omics may pave the way for a better understanding of the biological mechanisms underlying obesity. We observed proteins associated with weight gain in adolescents that are also associated with weight loss in adults in the literature, suggesting that these may have more to do with increases and decreases of lean mass, respectively, than

changes in body fat. Future studies with different designs (i.e., different populations, ages, or follow-up lengths) could provide a holistic view of how the proteome–obesity relationship is expressed.

## Abbreviations

<sup>1</sup> H-NMR	High-throughput proton nuclear magnetic resonance spectroscopy
A	Additive genetic component
AIC	Akaike Information Criterion
BMI	Body mass index
C	Shared or common environmental component
CAD	Coronary artery disease
CFI	Comparative Fit Index
CI	Confidence interval
CTMs	Classical twin models
D	Non-additive or dominant genetic component
DOC	Direction of causation
E	Nonshared or unique environmental component
$h^2$	Heritability
HSA	Human serum albumin
LGCM	Latent Growth Curve Modeling
LMWM	Low molecular weight molecules
ML	Maximum likelihood
NTR	Netherlands Twin Register
PCA	Principal component analysis
PCs	Principal components
PRSs	Polygenic risk scores
$r_A$	Genetic correlations
$r_{DZ}$	Correlations in DZ twins
$r_E$	Nonshared environmental correlations
RMSEA	Root Mean Square Error of Approximation
$r_{MZ}$	Correlations in MZ twins
sd	Standard deviation
SNP	Single-nucleotide polymorphism
TLI	Tucker-Lewis Index
TMI	Triponderal mass index
V	Variance

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-023-03198-7>.

**Additional file 1.** BIOS consortium and BBMRI-NL Consortium Banner.

**Additional file 2: Supplementary document.** Validity of self-reported versus measured anthropometric measurements.

**Additional file 3: Table S1.** Height and weight of FT12 participants across the four questionnaire waves. **Table S2.** Associations between metabolites and body mass index at blood sampling in NTR. **Table S3.** Estimates of hereditary and environmental components underlying abundance of proteins identified as associated with BMI at blood sampling in FinnTwin12 using univariate twin modeling. **Table S4.** Cross-twin cross-trait correlations in bivariate modeling between BMI and protein levels in monozygotic twins. **Table S5.** Cross-twin cross-trait correlations in bivariate modeling between BMI and protein levels in dizygotic twins. **Table S6.** Mixed model-derived estimates of proteins significantly associated with baseline body mass index (mean age: ~12 years old) in FinnTwin12 participants. **Table S7.** Mixed model-derived estimates of proteins significantly associated with baseline and changes in triponderal mass index in FinnTwin12 participants. **Table S8.** Estimates of hereditary and environmental components underlying abundance of proteins identified as associated with BMI changes in FinnTwin12 using univariate twin modeling. **Table S9.** Genetic and nonshared environmental correlations between proteins and BMI. **Table S10.** Genetic and nonshared environmental correlations between proteins and changes in BMI. **Table S11.** Addition of BMI at blood sampling as a covariate in the estimation of the

genetic and nonshared environmental correlations between proteins and changes in BMI. **Table S12.** Associations between gene expression and BMI at blood sampling in NTR at the gene level. **Table S13.** Associations between gene expression and BMI at blood sampling in NTR at the probe level. **Table S14.** Associations between gene expression and changes in BMI in NTR at the gene level. **Table S15.** Associations between gene expression and changes in BMI in NTR at the probe level. **Table S16.** Associations between gene expression and changes in BMI from the sensitivity analysis, at gene level. **Table S17.** Associations between gene expression and changes in BMI from the sensitivity analysis, at probe level. **Table S18.** Associations between metabolites and body mass index at blood sampling in FinnTwin12. **Table S19.** Associations between metabolites and changes in body mass index in FinnTwin12. **Table S20.** Associations between metabolites and changes in body mass index in NTR. **Table S21.** Associations between metabolites and baseline body mass index (mean age ~12 years old) in FinnTwin12. **Table S22.** Associations between PRSs and proteins associated with BMI at blood sampling in FinnTwin12. **Table S23.** Associations between BMI protein-encoding gene expression and PRS of BMI in NTR, gene level. **Table S24.** Associations between BMI protein-encoding gene expression and PRS of BMI in NTR, probe level. **Table S25.** Associations between PRSs and proteins associated with changes in BMI in FinnTwin12. **Table S26.** Associations between changes in BMI protein-encoding gene expression and PRS of BMI in NTR, gene level. **Table S27.** Associations between changes in BMI protein-encoding gene expression and PRS of BMI in NTR, probe level. **Table S28.** Associations between metabolites associated with BMI at blood sampling with PRSs in FinnTwin12. **Table S29.** Associations between metabolites associated with changes in BMI with PRSs in FinnTwin12. **Table S30.** Correlation matrix across omics layers in FinnTwin12. **Table S31.** Correlation matrix across omics layers in NTR.

**Additional file 4: Supplementary document.** Genotyping and polygenic score calculation in NTR.

### Acknowledgements

We gratefully acknowledge the contribution of the Turku Proteomics Facility team supported by Biocenter Finland for mass spectrometry. We warmly thank the participants involved in this study. Biobank-based Integrative Omics Study Consortium and BBMRI-NL Metabolomics Consortium lists of authors and their affiliations appear in the supplementary material (Additional file 1).

### Authors' contributions

The study design was developed and discussed by GD, FAH, AMW, NH, SR, MP, DIB, JvD, and JK. GW, EJCdG, KMK, DIB, JvD, and JK provided funding to collect data and/or participated in data collection. The BIOS Consortium and the BBMRI-NL Metabolomics Consortium were involved in funding and managing samples for generating the NTR data. RP and AA facilitated access to data for analysis. Processing of FinnTwin12 data was performed by GD and AMW. NTR data were processed by GD, FAH, RP, JJH, and RJ. The bioinformatic and statistical analyses were performed by GD. GD and FAH wrote the original manuscript. All authors actively participated in the improvement of the manuscript by critically revising it. All of the authors read and approved the final version of the manuscript.

### Authors' Twitter handles

@FionaHagenbeek (Fiona Hagenbeek); @jenny\_vandongen (Jenny van Dongen).

### Funding

Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). The Biobank-based Integrative Omics Study (BIOS) Consortium and the BBMRI Metabolomics Consortium are funded by BBMRI-NL, a Research Infrastructure financed by NWO, project nos. 184.021.007 and 184033111. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Phenotype and genotype data collection in FinnTwin12 cohort has been supported by the Wellcome Trust Sanger Institute, the Broad Institute, ENGAGE – European Network for Genetic and Genomic Epidemiology, FP7-HEALTH-F4-2007, grant agreement number 201413, National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145,

and AA-09203 to R J Rose; AA15416 and K02AA018755 to D M Dick; R01AA015416 to Jessica Salvatore) and the Academy of Finland (grants 100,499, 205,585, 118,555, 141,054, 264,146, 308,248 to JK, and the Centre of Excellence in Complex Disease Genetics (grants 312,073, 336,823, and 352,792 to JKaprio). This research was partly funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 874724 (Equal-Life). Equal-Life is part of the European Human Exposome Network.

For the Netherlands Twin Register, funding was obtained from the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organisation for Health Research and Development (ZonMW) grants 904-61-090, 985-10-002, 912-10-020, 904-61-193,480-04-004, 463-06-001, 451-04-034, 400-05-717, Addiction-31160008, 016-115-035, 481-08-011, 400-07-080, 056-32-010, Middelgroot-911-09-032, OCW\_NWO Gravity program -024.001.003, NWO-Groot 480-15-001/674, Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI -NL, 184.021.007 and 184.033.111), X-Omics 184-034-019; Spinozapremie (NWO- 56-464-14,192), KNAW Academy Professor Award (PAH/6635) and University Research Fellow grant (URF) to DIB; Amsterdam Public Health research institute (former EMGO+), Neuroscience Amsterdam research institute (former NCA); the European Community's Fifth and Seventh Framework Program (FP5- LIFE QUALITY-CT-2002-2006, FP7- HEALTH-F4-2007-2013, grant 01254: GenomEUtwin, grant 01413: ENGAGE and grant 602,768: ACTION); the European Research Council (ERC Starting 284,167, ERC Consolidator 771,057, ERC Advanced 230,374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the National Institutes of Health (NIH, R01D0042157-01A1, R01MH58799-03, MH081802, DA018673, R01 DK092127-04, Grand Opportunity grants 1RC2 MH089951, and 1RC2 MH089995); the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health as well as the US National Institute of Mental Health (RC2 MH089951). Computing was supported by NWO through grant 2018/EW/00408559, BiG Grid, the Dutch e-Science Grid and SURFSARA.

GD has received funding for his doctoral studies from the Doctoral Programme in Population Health (DOCPOP), University of Helsinki, Finland. JK acknowledges support by the Academy of Finland (grants 265,240, 263,278) and the Sigrid Juselius Foundation.

### Availability of data and materials

FinnTwin12 data analyzed in this study is not publicly available due to the restrictions of informed consent. Requests to access these datasets should be directed to the Institute for Molecular Medicine Finland (FIMM) Data Access Committee (DAC) (fimmdac@helsinki.fi) for authorized researchers who have IRB/ethics approval and an institutionally approved study plan. To ensure the protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data.

The data of the Netherlands Twin Register (NTR) may be accessed, upon approval of the data access committee, through the NTR ([https://tweelingenregister.vu.nl/information\\_for\\_researchers/working-with-ntr-data](https://tweelingenregister.vu.nl/information_for_researchers/working-with-ntr-data)).

### Declarations

#### Ethics approval and consent to participate

In FinnTwin12, ethical approval for all data collection waves was obtained from the ethical committee of the Helsinki and Uusimaa University Hospital District and the Institutional Review Board of Indiana University. All data collection and sampling protocols were performed in compliance with the ethical guidelines. Parents provided consent for the twins aged 12 and 14 years old, while twins aged 17 and 22 years old provided written consent themselves for sample collection.

In NTR, informed consent was obtained from all participants. Projects were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance- FWA00017598; IRB/institute code, NTR 03-180).

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. <sup>2</sup>Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>3</sup>Amsterdam Public Health Research Institute, Amsterdam, The Netherlands. <sup>4</sup>Department of Psychiatry, Amsterdam UMC Location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. <sup>5</sup>Amsterdam Neuroscience, Mood, Anxiety, Psychosis, Sleep & Stress Program, Amsterdam, The Netherlands. <sup>6</sup>Amsterdam Reproduction & Development (AR&D) Research Institute, Amsterdam, The Netherlands. <sup>7</sup>A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland. <sup>8</sup>Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland. <sup>9</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>10</sup>Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

Received: 3 July 2023 Accepted: 27 November 2023

Published online: 21 December 2023

**References**

- Blüher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol.* 2019;15(5):288–98.
- World Obesity Federation, World Obesity Atlas 2023. <https://data.worldobesity.org/publications/?cat=19> (2023). Accessed February 2023.
- Kivimäki M, Strandberg T, Pentti J, et al. Body-mass index and risk of obesity-related complex multimorbidity: an observational multicohort study. *Lancet Diabetes Endocrinol.* 2022;10(4):253–63.
- Huang J, Huffman JE, Huang Y, et al. Genomics and phenomics of body mass index reveals a complex disease network. *Nat Commun.* 2022;13(1):7973.
- Paczkowska-Abdulsalam M, Kretowski A. Obesity, metabolic health and omics: current status and future directions. *World J Diabetes.* 2021;12(4):420–36.
- Silventoinen K, Jelenkovic A, Sund R, et al. Differences in genetic and environmental variation in adult BMI by sex, age, time period, and region: an individual-based pooled analysis of 40 twin cohorts. *Am J Clin Nutr.* 2017;106(2):457–66.
- Klok MD, Jakobsdottir S, Drent ML. The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obes Rev.* 2007;8(1):21–34.
- Atalayer D, Gibson C, Konopacka A, Geliebter A. Ghrelin and eating disorders. *Prog Neuropsychopharmacol Biol Psychiatry.* 2013;40:70–82.
- Masood A, Benabdelkamel H, Alfadda AA. Obesity Proteomics: An Update on the Strategies and Tools Employed in the Study of Human Obesity. *High Throughput.* 2018;7(3):27.
- Bao X, Xu B, Yin S, et al. Proteomic Profiles of Body Mass Index and Waist-to-Hip Ratio and Their Role in Incidence of Diabetes. *J Clin Endocrinol Metab.* 2022;107(7):e2982–90.
- Goudswaard LJ, Bell JA, Hughes DA, et al. Effects of adiposity on the human plasma proteome: observational and Mendelian randomisation estimates. *Int J Obes (Lond).* 2021;45(10):2221–9.
- Zaghlool SB, Sharma S, Molnar M, et al. Revealing the role of the human blood plasma proteome in obesity using genetic drivers. *Nat Commun.* 2021;12(1):1279.
- Kreitmaier P, Katsoula G, Zeggini E. Insights from multi-omics integration in complex disease primary tissues. *Trends Genet.* 2023;39(1):46–58.
- Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol.* 2017;18(1):83.
- Zhang Q, Meng XH, Qiu C, et al. Integrative analysis of multi-omics data to detect the underlying molecular mechanisms for obesity in vivo in humans. *Hum Genomics.* 2022;16(1):15.
- Wahl S, Vogt S, Stücker F, et al. Multi-omic signature of body weight change: results from a population-based cohort study. *BMC Med.* 2015;13:48.
- Pietiläinen KH, Sysi-Aho M, Rissanen A, et al. Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects—a monozygotic twin study. *PLoS ONE.* 2007;2(2): e218.
- Watanabe K, Wilmanski T, Diener C, et al. Multiomic signatures of body mass index identify heterogeneous health phenotypes and responses to a lifestyle intervention. *Nat Med.* 2023. <https://doi.org/10.1038/s41591-023-02248-0>.
- Piening BD, Zhou W, Contrepois K, et al. Integrative Personal Omics Profiles during Periods of Weight Gain and Loss. *Cell Syst.* 2018;6(2):157–170.e8.
- Lapatto HAK, Kuusela M, Heikkinen A, et al. Nicotinamide riboside improves muscle mitochondrial biogenesis, satellite cell differentiation, and gut microbiota in a twin study. *Sci Adv.* 2023;9(2):eadd5163.
- Gallego-Paúls M, Hernández-Ferrer C, Bustamante M, et al. Variability of multi-omics profiles in a population-based child cohort. *BMC Med.* 2021;19(1):166.
- Simmonds M, Llewellyn A, Owen CG, Woolcott N. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obes Rev.* 2016;17(2):95–107.
- Figarska SM, Rigdon J, Ganna A, et al. Proteomic profiles before and during weight loss: Results from randomized trial of dietary intervention. *Sci Rep.* 2020;10(1):7913.
- Lind L, Figarska S, Sundström J, Fall T, Ärnlöv J, Ingelsson E. Changes in Proteomic Profiles are Related to Changes in BMI and Fat Distribution During 10 Years of Aging. *Obesity (Silver Spring).* 2020;28(1):178–86.
- Geyer PE, Wewer Albrechtsen NJ, Tyanova S, et al. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol Syst Biol.* 2016;12(12):901.
- Carayol J, Chabert C, Di Cara A, et al. Protein quantitative trait locus study in obesity during weight-loss identifies a leptin regulator. *Nat Commun.* 2017;8(1):2084.
- Vijay A, Valdes AM. The Metabolomic Signatures of Weight Change. *Metabolites.* 2019;9(4):67.
- Wahl S, Yu Z, Kleber M, et al. Childhood obesity is associated with changes in the serum metabolite profile. *Obes Facts.* 2012;5(5):660–70.
- Freedman DS, Wang J, Maynard LM, et al. Relation of BMI to fat and fat-free mass among children and adolescents. *Int J Obes (Lond).* 2005;29(1):1–8. <https://doi.org/10.1038/sj.ijo.0802735>.
- Kaprio J. Twin studies in Finland 2006. *Twin Res Hum Genet.* 2006;9(6):772–7.
- Rose RJ, Salvatore JE, Aaltonen S, Barr PB, Bogl LH, Byers HA, et al. FinnTwin12 Cohort: An Updated Review. *Twin Res Hum Genet.* 2019;22(5):302–11.
- Ligthart L, van Beijsterveldt CEM, Kevenaar ST, et al. The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Res Hum Genet.* 2019;22(6):623–36.
- Willemsen G, de Geus EJ, Bartels M, et al. The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet.* 2010;13(3):231–45.
- Aleksei Afonin, Aino-Kaisa Piironen, Izaque de Sousa Maciel, et al. Proteomic insights into mental health status: plasma markers in young adults. *bioRxiv* 2023.06.07.544039. <https://doi.org/10.1101/2023.06.07.544039>
- Callister SJ, Barry RC, Adkins JN, et al. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res.* 2006;5(2):277–86.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28(6):882–3.
- Jansen R, Batista S, Brooks AI, et al. Sex differences in the human peripheral blood transcriptome. *BMC Genomics.* 2014;15:33.
- Wright FA, Sullivan PF, Brooks AI, et al. Heritability and genomics of gene expression in peripheral blood. *Nat Genet.* 2014;46(5):430–7.
- Vink JM, Jansen R, Brooks A, et al. Differential gene expression patterns between smokers and non-smokers: cause or consequence? *Addict Biol.* 2017;22(2):550–60.
- Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol.* 2017;186(9):1084–96.
- Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet.* 2015;8(1):192–206.
- Whipp AM, Heinonen-Guzejev M, Pietiläinen KH, van Kamp I, Kaprio J. Branched-chain amino acids linked to depression in young adults. *Front Neurosci.* 2022;16:935858.

43. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18(6):499–502.
44. Hagenbeek FA, Pool R, van Dongen J, et al. Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat Commun*. 2020;11(1):39.
45. Kujala UM, Palviainen T, Pesonen P, et al. Polygenic Risk Scores and Physical Activity. *Med Sci Sports Exerc*. 2020;52(7):1518–24.
46. Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641–9.
47. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019;28(1):166–74.
48. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121–30.
49. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9.
50. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
51. Auton A, Brooks LD, 1000 Genomes Project Consortium, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
52. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867–73.
53. Loh PR, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet*. 2016;48(7):811–6.
54. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9(2):179–81.
55. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284–7.
56. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/ROH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. 2016;32(11):1749–51.
57. Vilhjálmsdóttir BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015;97(4):576–92.
58. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw*. 2012;48(2):1–36.
59. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1–55.
60. Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behav Res Methods*. 2019;51(1):409–28.
61. Peterson CM, Su H, Thomas DM, et al. Tri-Ponderal Mass Index vs Body Mass Index in Estimating Body Fat During Adolescence. *JAMA Pediatr*. 2017;171(7):629–36.
62. Gul Siraz U, Hatipoglu N, Mazicioglu MM, Ozturk A, Cicek B, Kurtoglu S. Triponderal mass index is as strong as body mass index in the determination of obesity and adiposity. *Nutrition*. 2023;105: 111846.
63. Bramante CT, Palzer EF, Rudser KD, et al. BMI metrics and their association with adiposity, cardiometabolic risk factors, and biomarkers in children and adolescents. *Int J Obes (Lond)*. 2022;46(2):359–65.
64. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Softw*. 2017;82(13):1–26.
65. Posthuma D, Beem AL, de Geus EJ, et al. The theory and practice in quantitative genetics. *Twin Res*. 2003;6(5):361–76.
66. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nat Rev Genet*. 2002;3(11):872–82.
67. Rijsdijk FV, Sham PC. Analytic approaches to twin data using structural equation models. *Brief Bioinform*. 2002;3(2):119–33. <https://doi.org/10.1093/bib/3.2.119>.
68. Hagenbeek FA, van Dongen J, Pool R, et al. Heritability of Urinary Amines, Organic Acids, and Steroid Hormones in Children. *Metabolites*. 2022;12(6):474.
69. Neale MC, Hunter MD, Pritikin JN, et al. OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*. 2016;81(2):535–49.
70. Pritikin JN, Hunter MD, Boker SM. Modular open-source software for Item Factor Analysis. *Educ Psychol Measur*. 2015;75(3):458–74.
71. Hunter MD. State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment. *Struct Equ Model*. 2018;25(2):307–24.
72. Boker SM, Neale MC, Maes HH, et al. OpenMx 2.21.1 User Guide. 2023.
73. Verhulst B, Prom-Wormley E, Keller M, Medland S, Neale MC. Type I Error Rates and Parameter Bias in Multivariate Behavioral Genetic Models. *Behav Genet*. 2019;49(1):99–111.
74. Castro-de-Araujo LFS, Singh M, Zhou Y, et al. MR-DoC2: Bidirectional Causal Modeling with Instrumental Variables and Data from Relatives. *Behav Genet*. 2023;53(1):63–73.
75. Cominetti O, Núñez Galindo A, Corthésy J, et al. Obesity shows preserved plasma proteome in large independent clinical cohorts. *Sci Rep*. 2018;8(1):16981.
76. Oller Moreno S, Cominetti O, Núñez Galindo A, et al. The differential plasma proteome of obese and overweight individuals undergoing a nutritional weight loss and maintenance intervention. *Proteomics Clin Appl*. 2018;12(1). <https://doi.org/10.1002/prca.201600150>.
77. Mesinovic J, Jansons P, Zengin A, et al. Exercise attenuates bone mineral density loss during diet-induced weight loss in adults with overweight and obesity: A systematic review and meta-analysis. *J Sport Health Sci*. 2021;10(5):550–9.
78. Spoel SH. Orchestrating the proteome with post-translational modifications. *J Exp Bot*. 2018;69(19):4499–503.
79. Wu L, Candille SI, Choi Y, et al. Variation and genetic control of protein abundance in humans. *Nature*. 2013;499(7456):79–82.
80. Riuzzi F, Chiappalupi S, Arcuri C, Giambanco I, Sorci G, Donato R. S100 proteins in obesity: liaisons dangereuses. *Cell Mol Life Sci*. 2020;77(1):129–47.
81. Catalán V, Gómez-Ambrosi J, Rodríguez A, et al. Increased levels of calprotectin in obesity are related to macrophage content: impact on inflammation and effect of weight loss. *Mol Med*. 2011;17(11–12):1157–67.
82. Sekimoto R, Kishida K, Nakatsujii H, Nakagawa T, Funahashi T, Shimomura I. High circulating levels of S100A8/A9 complex (calprotectin) in male Japanese with abdominal adiposity and dysregulated expression of S100A8 and S100A9 in adipose tissues of obese mice. *Biochem Biophys Res Commun*. 2012;419(4):782–9.
83. Grand A, Rochette E, Dutheil F, et al. Body Mass Index and Calprotectin Blood Level Correlation in Healthy Children: An Individual Patient Data Meta-Analysis. *J Clin Med*. 2020;9(3):857.
84. Chen N, Miao L, Lin W, et al. Integrated DNA Methylation and Gene Expression Analysis Identified S100A8 and S100A9 in the Pathogenesis of Obesity. *Front Cardiovasc Med*. 2021;8:631650.
85. Drouard G, Silventoinen K, Latvala A, Kaprio J. Genetic and environmental factors underlying parallel changes in body mass index and alcohol consumption: a 36-year longitudinal study of adult twins. *Obes Facts*. 2023. <https://doi.org/10.1159/000529835>.
86. Liu Y, Buil A, Collins BC, et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol*. 2015;11(1):786.
87. Haj AK, Hasan H, Raife TJ. Heritability of Protein and Metabolite Biomarkers Associated with COVID-19 Severity: A Metabolomics and Proteomics Analysis. *Biomolecules*. 2022;13(1):46.
88. Behbodikhah J, Ahmed S, Elyasi A, et al. Apolipoprotein B and Cardiovascular Disease: Biomarker and Potential Therapeutic Target. *Metabolites*. 2021;11(10):690.
89. Sniderman AD, Thanassoulis G, Glavinovic T, et al. Apolipoprotein B Particles and Cardiovascular Disease: A Narrative Review. *JAMA Cardiol*. 2019;4(12):1287–95.
90. Gottesman O, Drill E, Lotay V, Bottinger E, Peter I. Can genetic pleiotropy replicate common clinical constellations of cardiovascular disease and risk? *PLoS ONE*. 2012;7(9):e46419.
91. Rankinen T, Sarzynski MA, Ghosh S, Bouchard C. Are there genetic paths common to obesity, cardiovascular disease outcomes, and cardiovascular risk factors? *Circ Res*. 2015;116(5):909–22.

92. Auwerx C, Sadler MC, Woh T, Reymond A, Kotalik Z, Porcu E. Exploiting the mediating role of the metabolome to unravel transcript-to-phenotype associations. *Elife*. 2023;12: e81097.
93. Sadler MC, Auwerx C, Lepik K, Porcu E, Kotalik Z. Quantifying the role of transcript levels in mediating DNA methylation effects on complex traits and diseases. *Nat Commun*. 2022;13(1):7559.

### **Publisher's Note**

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

