

Early-life proteomic and microbiome features signal obesity risk across 26 years of follow-up

Angelica P. Ahrens,¹ Raquel Dias,¹ Tuulia Hyötyläinen,² Pär Anderson White,³ Matej Orešič,^{4,5,6} Eric W. Triplett,¹ Johnny Ludvigsson³

AUTHOR AFFILIATIONS See affiliation list on p. 21.

ABSTRACT Childhood obesity is rising globally. Yet, few studies have examined the microbiome and proteome in early childhood in relation to this outcome, and most are cross-sectional by design. Early-life factors in the ABIS birth cohort ($n = 16,683$) were associated with obesity up to age 26 (mean follow-up 25.3 years, range 23.7–26.5 years): psychosocial stressors, smoking, infections, and diet in the first year. We assessed biomarkers, including cord blood metabolome ($n = 290$) and proteome ($n = 358$), by liquid chromatography, mass spectrometry, and Olink. Gut microbial composition at age one ($n = 1,743$) was assessed using stool samples and 16S rRNA sequencing. In this prospective longitudinal cohort study, significant differences were found in infants with future obesity, including elevated angiopoietin-like 4 (ANGPTL4), follistatin, and hepatocyte growth factor (independently of maternal weight) and reduced isocaproic acid, tryptophan, and oleic acid, with prenatal mediation. *Akkermansia*, asaccharolytic bacteria (*Phascolarctobacterium* and *Senegalimassiliensia*), and equol-producers (*Adlercreutzia* and *Slackia*) were depleted. Machine learning models selecting 40 most predictive features showed long-term prediction from birth proteomics and bacterial taxa at age one (area under the curve [AUC] = $0.83 \pm .05$, $n = 1,877$) and additional metrics, for example, parental and child body mass index in the first 8 years (AUC = $0.89 \pm .02$, $n = 1,877$), suggesting durable biological encoding. Proteomic markers across folds included fibroblast growth factor 19, ANGPTL4, sulfotransferase family 2A member 1, and interleukin 20. These findings suggest clinically relevant biomarkers indicating early-life regulation of bile acid metabolism, lipid storage vs. oxidation, and immune–metabolic signaling and pathways to prospectively prevent childhood- and adult-onset obesity across a 26-year predictive gap.

IMPORTANCE Understanding the origins of obesity is critical for developing preventive strategies, and early life represents a particularly sensitive window. This study leverages a large, general-population cohort with prospectively collected data, including parental body mass index (BMI), cord blood proteomics, and the gut microbiome at age one, linked to obesity outcomes over 26 years. Using integrated machine learning models, we show that in addition to parental BMI, specific proteomic and microbial markers present in infancy can predict long-term obesity risk, highlighting the role of early metabolic programming. Several key markers point to bile acid signaling as a mechanism connecting early microbiome development with fat accumulation and insulin regulation. By identifying these early-life predictors long before obesity manifests, these results provide new insights into intergenerational risk and suggest measurable targets for preventing obesity and related metabolic disorders from the earliest stages of life.

KEYWORDS microbiome, metabolome, environmental toxins, inflammation, pregnancy, metabolic disease, diabetes, machine learning, ANGPTL4, FGF19, SULT2A1, bile acids, carbohydrates

Editor Justin J. van der Hooft, Wageningen University, Wageningen, the Netherlands

Address correspondence to Eric W. Triplett, ewt@ufl.edu.

J.L., E.W.T., and M.O. declare funding from the Horizon Europe Program of the European Union during this study. All other authors declare no competing interests.

See the funding table on p. 21.

Received 3 October 2025

Accepted 20 April 2026

Published 28 May 2026

Copyright © 2026 Ahrens et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Childhood obesity is rising globally (1), driven by biological, social, and environmental factors. Although diet and lifestyle contribute significantly, these alone are not sufficient. Several conceptual models of ecological influences have led to proposed interventions (2). Factors such as reduced breastfeeding, sedentary behavior, home environment, and infections associated with weight gain in children have been studied in recent meta-analyses (3–13). Early-life programming appears to be key to developing evidence-based, personalized treatment approaches (3) to reduce the health burden caused by obesity.

The intrauterine metabolic environment shapes the early microbiome, though its long-term impact on metabolism and obesity risk remains unclear. The link between parental and child obesity is well established (4), particularly during early childhood (5). However, multi-omic, prospective investigations are needed to understand the interactions of factors involved in obesity (6). The gut microbiota are a key mediator of host metabolism, facilitating monosaccharide absorption and hepatic lipogenesis (7). Certain species, like *Akkermansia muciniphila*, are consistently less abundant in individuals with obesity (8), while others like Firmicutes are more abundant. In children with obesity, altered levels of indole—produced by gut bacteria during tryptophan fermentation—suggest dysregulation of tryptophan metabolism (9). While probiotic interventions more broadly have shown variable success as treatments (10, 11), dietary prebiotics have shown moderate success in reducing body weight and improving metabolic profiles (12). Gut dysbiosis is a trigger that can cause inflammation, in part due to increased intestinal permeability (13).

Given that the first 1,000 days of life are critical for microbiome development, investigating early life processes is especially important (14, 15). The host–microbe–environment interactions are especially vulnerable to undernutrition, with intergenerational effects (14). Despite growing interest, few studies have examined the microbiome and proteome in early childhood, especially in the first years of life, in relation to future obesity. Most studies have been cross-sectional (16), with only five of the 42 focused on neonates or infants. A small study of forty 4- to 5-year-old children found increased *Enterobacteriaceae* and decreased *Akkermansia muciniphila*-like bacteria in children with weight issues. However, these differences coincided with obesity rather than preceding it, limiting causal interpretation (17). Another study of 2-year-olds provides preliminary evidence of an association, in which microbial differences preceded later weight problems at age 12, but its small sample size and lack of long-term follow-up into adulthood constrain its implications (18).

Here, in the All Babies in Southeast Sweden (ABIS) prospective longitudinal cohort study, we examined the cord blood proteome and gut microbiome at age one year in relation to obesity outcomes over 26 years of follow-up. This study of a general population enables the prospective identification of early predictors of long-term obesity risk (Fig. 1). Understanding how the intrauterine metabolic environment and early microbial development contribute to the intergenerational transmission of obesity risk could guide preventive strategies, early-life interventions, and development of potential biomarkers.

MATERIALS AND METHODS

Study design

The ABIS study (19–21) is a prospective, general population-based cohort (19) of children born in southeast Sweden from October 1997 to 1999. Parents were recruited from nine obstetric clinics across Östergötland, Småland, Blekinge, and Öland, with 17,055 out of 21,700 (78.6%) families providing informed consent, of whom 16,683 had answered the birth questionnaires. The ABIS study shows a balanced distribution of biological sex, with 48.2% female, and as it consists only of children born in Sweden (primarily to Swedish parents), sociocultural constructs like race and ethnicity were not collected. The ABIS

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.200.55.

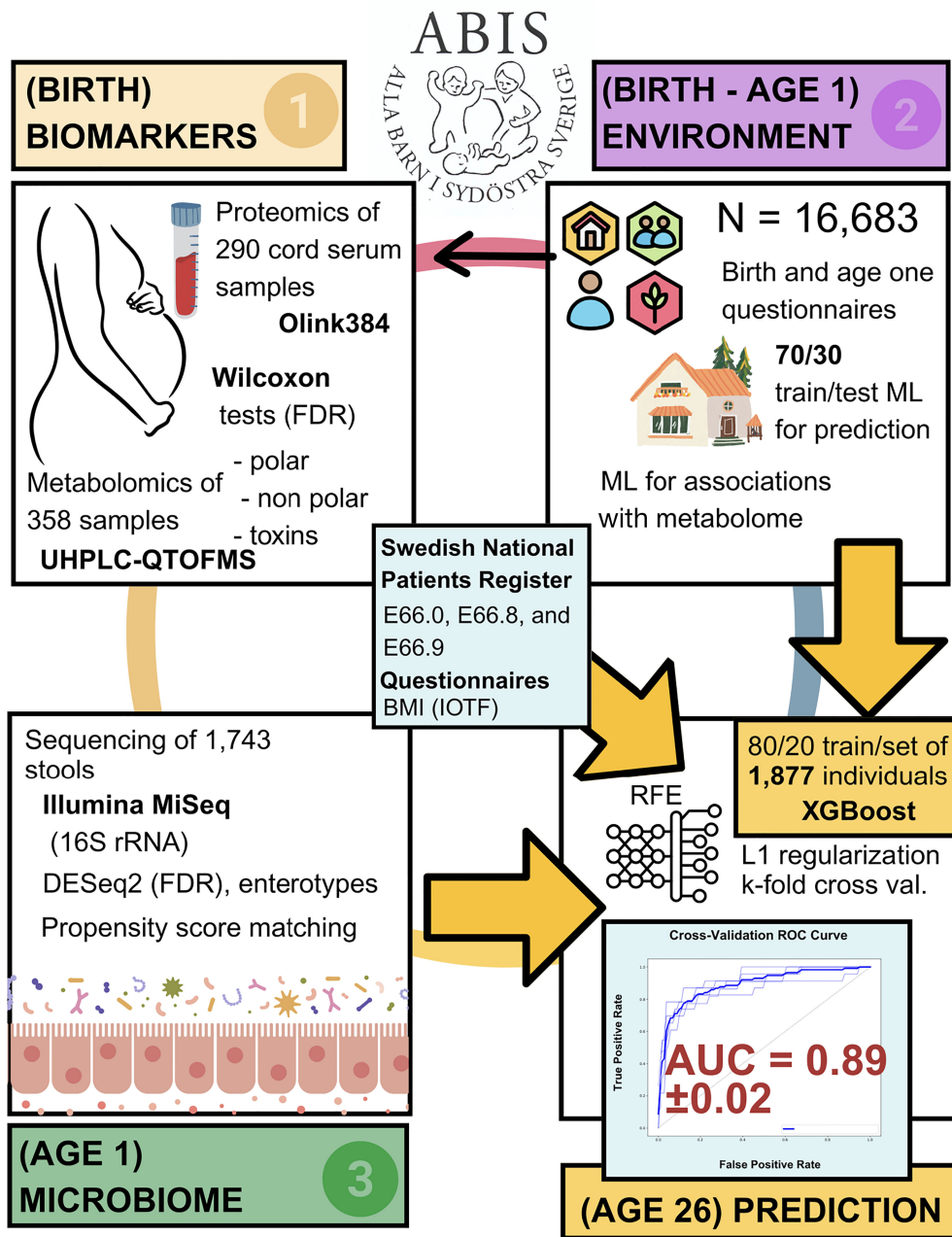


FIG 1 Study design. This cohort study is based on the All Babies in Southeast Sweden (ABIS) general population study ($N = 16,683$). Biomarkers were assessed for association with future obesity outcomes, including body mass index (BMI) classifications based on International Obesity Task Force (IOTF) standards and clinical diagnoses (ICD-10 codes E66.0, E66.8, and E66.9) through age 26. Biomarkers included metabolomics and targeted proteomics from cord blood, as well as microbiome sequencing from stool samples collected at approximately 1 year of age. Wilcoxon tests and differential expression analysis from binomial distribution modeling (via DESeq2) identified top metabolites, proteins, and taxa associated with obesity outcomes. Environmental and lifestyle exposures from pregnancy to the first year were modeled using machine learning (ML) with a 70/30 train-test split. The final integrated data set included 1,877 individuals with complete biological samples, which were used in the XGBoost ML pipeline. Recursive feature elimination (RFE) was employed for dimension reduction, and L1 regularization with k-fold cross-validation was applied to predict obesity diagnosis up to age 26. More details are provided in Materials and Methods. The final model achieved the area under the curve (AUC) of 0.89, using the top 40 features (taxa, proteins, and early-life risk factors) selected by RFE.

register was connected to the Swedish National Patient Register (NPR) (22), giving the confirmed clinical diagnosis of obesity (23).

Definitions of obesity cases and controls and analytical framework

Obesity was defined using two complementary definitions. First, BMI was calculated from parent-reported height and weight from the 3-, 5-, 8-, 10-, and 19-year questionnaires, with weight classifications following IOTF (24) standards. Second, registry-based clinically diagnosed obesity was identified ($n = 407$ of 16,683) using International Classification of Diseases (ICD-10) codes E66.0, E66.8, and E66.9, as recorded in participants' medical records, via the Swedish National Patient Register through December 2023. Controls were defined as normal weight by IOTF standards, based on a parent-completed questionnaire, and with no future autoimmune or neurodevelopmental diagnosis through December 2023.

Questionnaire- and BMI-based overweight and obesity classifications derived from parent-reported height and weight, as well as perinatal and prenatal features, were used to stratify or group participants in the proteomic and microbiome analyses and as predictors in single-layer machine learning analyses linking features to metabolomic data. These measures were not used as outcomes in the integrated multi-omic machine learning models, which focused exclusively on clinically diagnosed obesity obtained from the Swedish National Patient Register (ICD-10 codes), thereby avoiding reliance on self-reported anthropometric measurements for predictive modeling. Given known measurement error in parent-reported anthropometrics, age- and sex-specific IOTF obesity categories were used to improve robustness and reduce sensitivity to modest reporting inaccuracies.

The primary clinical endpoint for etiologic inference and predictive modeling was the registry-confirmed obesity diagnosis obtained from the Swedish NPR (22). While this outcome underestimates population prevalence, it represents a highly specific and physician-recorded diagnosis, providing greater confidence in outcome validity.

Statistics

Associations between questionnaire data (birth and 1-year surveys) and future obesity outcomes

For exploratory associations between early-life questionnaire variables and future obesity, odds ratios (ORs) were calculated from 2×2 contingency tables in SPSS for exposures including gestational diabetes, parental education, parental BMI, infections, and antibiotic use. Odds ratios represent unadjusted associations between exposure and outcome. These analyses were conducted using both obesity definitions: age- and sex-specific IOTF classifications derived from parent-reported questionnaires and registry-confirmed clinical obesity diagnoses.

For this investigation, we primarily focused on the prenatal/birth and 1-year questionnaire environmental data, as these exposures precede the development of obesity and minimize potential reverse causation. Gestational diabetes of the mother was additionally assessed using Swedish National Patient Register data (ICD-10 code O24.4).

In addition to early-life factors collected at birth and 1 year of age, parental BMI was also analyzed in two distinct contexts. First, to characterize familial patterns of adiposity across early childhood, parental BMI measurements obtained from follow-up questionnaires (up to age eight of the child) were used in correlation analyses to examine associations with child BMI at multiple ages. Second, parental BMI was included as a predictor of clinically diagnosed obesity through adolescence and early adulthood, where it was treated as a stable background characteristic reflecting long-term familial risk rather than a child-level exposure.

Given the large number of interrelated early-life exposures and the descriptive aim of this component of the study, multivariable adjustment was not applied. These analyses

were intended to identify early-life patterns associated with later obesity risk and should not be interpreted as estimates of independent causal effects.

Metabolomic analysis of cord serum and associations with obesity (by IOTF classification)

Cord serum was analyzed in 358 ABIS children, as described previously (19). Briefly, lipidomics profiling was carried out using an ultra-high-performance liquid chromatography quadrupole time-of-flight mass spectrometry (UHPLC-QTOFMS from Agilent Technologies; Santa Clara, CA, USA) and analysis with an ACQUITY UPLC BEH C18 column (2.1 mm × 100 mm, particle size 1.7 μm) by Waters (Milford, MA, USA). Polar and semipolar metabolites were analyzed using an Acquity UPLC system coupled to a triple quadrupole mass spectrometer (Waters Corporation, Milford, MA, USA) with an atmospheric electrospray interface operating in negative-ion mode, with quantification of bile acids (BAs) and PFAS using a 7-point internal calibration. Data were processed using MZmine 2.53. Pooled quality control samples were included, along with a reference and compound standards, and extracted blank samples. Polar metabolites were assessed against BMI IOTF classifications at age 5 years using the Kruskal–Wallis test.

Predicting metabolites and exogenous compounds in cord blood from prenatal/perinatal features

To predict metabolite and exogenous toxin concentrations from prenatal and perinatal factors, and vice versa, we employed ML models including Extreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), Extra Trees, Random Forest (RF), and Logistic Regression (LR), using normalized metabolomic data as a continuous variable. The goal was to assess which early-life factors contain predictive information about the composition of the cord blood. Model performance was assessed using AUC and F1 score. To interpret the direction of the effect of each predictor, explainable AI via SHapley Additive exPlanations (SHAP) was used to learn the contribution of features in the predictions. These associations serve to identify predictive patterns rather than establish causal effects.

Proteomic analysis of cord serum

Cord serum samples were processed by Olink (Uppsala, Sweden), using Olink's Explore 384 Inflammation 1 and 2 panels and the Target Immune Response panel. Cases were intentionally oversampled to overcome class imbalance. Internal and external controls were run for batch effects and technical variations. Normalized protein expression (NPX) values were calculated as a relative protein quantification unit on a log₂ scale, and data intensity was normalized.

Initial differences between case/control were assessed using Wilcoxon test statistics comparing NPX values, based on BMI classifications across time (self-reported questionnaires) or on clinical diagnosis from the Swedish National Patient Register (up to age 26). Median values for each group and the effect size (Cohen's *d*) between the two groups were calculated. Comparisons included the following: 2–10 years global comparison (218 controls vs. 72 a BMI corresponding to obesity at least once across the timepoints) and BMI at 2 (129 controls vs. 45 overweight/obese), 5 (92 controls vs. 51 overweight/obese), and 8 (43 controls vs. 19 overweight/obese) years. For the analysis based on clinical diagnosis, 204 controls and 84 future obesity cases were compared, with E66.0, E66.8, and E66.9 as the qualifying ICD-10 codes.

Sequencing of stool samples at 1 year of age

A stool sample was collected from the participating child's diaper, at a mean age of 11.93 ± 2.94 months, using a sterile spatula and collection tubes provided by the WellBaby Clinic, as previously described (19). Briefly, samples were immediately frozen after collection, either at the infant's home or at the clinic. For home collections, freeze

clamps were used to maintain frozen conditions during transport to the clinic, where samples were subsequently stored at -80°C .

DNA extraction and sequencing of 16S rRNA gene amplicons from 1,748 stool samples were done as previously described (19). Briefly, a total of 1,748 samples were sequenced in 10 pools using Illumina MiSeq 2×300 bp at the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida, Gainesville, Florida, USA, following established protocols. Amplicons for targeted V3-V4 16S rRNA sequencing were produced using Standard Illumina Read 1 sequencing/indexing primers 341F (NN NNCCTACGGGAGGCAGCAG) and 806R (GGGGACTACVSGGGTATCTAAT). Forward primer was as follows: 5'-P5-adapter-linker-SBS3-16S-3', 5'-AATGATACGGCGACCACCGAGCIWTH TAYGGIAARGGIGGGIATHGGIAA-3'. Reverse primer was as follows: 5'-P7 adapter-linker-barcode-SBS12-16S-3', 5'-CAAGCAGAAGACGGCATAACGAGAT-(barcode)-GTGACTGGAGTT CAGACGTGTGCTCTCCGCTCTGGGGACTACVSGGGTATCTAAT-3'. For pooling, barcodes 11 nucleotides in length were used. Each PCR sample was spin-column purified and quantified by Qubit prior to pooling.

Paired ends were joined and demultiplexed in QIIME1, sample inference done using DADA2, and taxonomy assigned using the SILVA database, discarding five samples with low read counts. The resulting pool comprised 107,959,956 reads (with a minimum of 10,059 and a maximum of 776,158 reads per sample), including an average of 61,939.1 reads per sample (57,844 median) and 13,197 unique amplicon sequence variants (ASVs), representing 511 genera and 931 species.

Differential gut microbiome analysis

Differential taxonomic analysis was conducted using a negative binomial generalized linear model approach (DESeq2 [25]) with FDR correction. Significant taxa were defined as those with false discovery rate (FDR)-adjusted $P < 0.05$ and normalized base mean counts of abundance ≥ 10 , representative of the average of the normalized counts (adjusted for sequencing depth across samples) for the taxon across all groups in that data set. This is calculated as the geometric mean of the counts across all samples, regardless of outcome.

A total of 598 controls and 50 future clinical obesity cases, with a qualifying diagnosis of E66.0, E66.8, or E66.9, were included in the analysis. Differential abundance was tested using DESeq2. Independent filtering of features with very low mean counts was applied automatically by DESeq2 to optimize detection power. No manual pre-filtering of taxa was performed.

For the IOTF-based comparisons using the matchIt R package, a propensity score matching of a 3:1 ratio of controls to cases addressed factors related to prenatal, birth, and early diet influences: biological sex, parental education level at birth, maternal smoking during pregnancy, gestational age at birth, mode of delivery, breastfeeding duration, mother's BMI, gestational diabetes, and antibiotic exposure. The matching method used was nearest neighbor. The resulting data sets included 13 children who were obese by age 3 against 39 controls of normal weight (400 taxa represented), and 22 children with obesity by age 5 against 66 controls (464 taxa represented). As before, independent filtering was applied using DESeq2 defaults, and no additional, manual pre-filtering was performed.

Integrated, multi-omic machine learning to predict future obesity diagnosis

ML models were developed to predict future clinically diagnosed obesity using integrated molecular features (microbiome and proteomics) from 1,877 ABIS individuals, where either microbiome or proteomic data were available, including 114 with a clinical obesity diagnosis from the Swedish National Patient Register. For the ML prediction models, XGBoost, a gradient-boosted decision tree algorithm, was implemented in Python, with the outcome variable encoded as a binary classification (0 = no obesity diagnosis; 1 = clinical obesity diagnosis). The XGBoost models were implemented using the XGBClassifier with the following hyperparameters: $n_estimators = 300$, max_depth

= 4, learning_rate = 0.05, alpha = 1 (L1 regularization to induce sparsity), eval_metric = "logloss," random_state = 42, and scale_pos_weight set to the ratio of controls to cases within each training fold to account for class imbalance. The logloss evaluation metric and L1 regularization (alpha = 1) were used to induce sparsity. Default hyperparameters were used based on established performance in similar biomedical prediction tasks with imbalanced data. All ML models were implemented in Python 3.10.12 using: XGBoost (v3.1.3, dmlc implementation), scikit-learn (v1.8.0) for preprocessing and cross-validation, SHAP (v0.50.0) for model interpretation.

Initially, models were trained on the molecular features alone, and then with metadata incorporated, including the first-year questionnaire as well as parental and child BMI up to age 8 to evaluate incremental predictive value, while considering possible interactions between biological and non-biological features. A total of 683 features were brought into testing, including 230 bacterial and 363 proteomic, prior to RFE selection within each fold. Missing data were handled using scikit-learn's SimpleImputer (median strategy) within the testing and training sets separately, and within each fold, to avoid data leakage.

For both approaches, the data set was split into 80% and 20% training and testing sets with a fixed random state for reproducibility. Model performance was evaluated using stratified fivefold cross-validation to preserve case-control proportions across folds. Within each fold, the model was trained on four-fifths of the data and evaluated on the held-out fold. No information from the test fold was used during feature selection or model fitting. To reduce dimensionality, feature selection was performed using RFE with an L1-regularized XGBoost estimator. Importantly, RFE was conducted within each training fold only to prevent information leakage. For each fold, the top 40 features were retained based on model-based feature importance rankings. We evaluated models using both 40 and 50 features selected via RFE. Performance was nearly identical, suggesting that the top 40 features capture the majority of the predictive signal. We selected 40 features for the final model to maximize interpretability while minimizing overfitting risk, given our 114 cases.

The number of times each selected feature was chosen was counted, so that the key features of obesity risk across folds could be identified. Predictive performance was assessed using the area under the receiver operating characteristic curve (AUC), precision, recall, and F1 score. Performance metrics were calculated independently for each fold and summarized as mean \pm standard deviation across cross-validation folds. An average ROC curve was constructed by interpolation across a common false-positive rate grid. Consensus features were defined as those selected by RFE in at least three cross-validation folds, and a final XGBoost model was then trained on the full data set using only the consensus features, with class imbalance addressed via scale_pos_weight and L1 regularization (alpha = 1), to extract the relative importances.

This modeling framework was designed to minimize overfitting, ensure independence between training and testing data, and provide conservative estimates of predictive performance. The combination of within-fold feature selection, L1 regularization, and stratified cross-validation substantially reduced overfitting risk. Importantly, the integrated predictive models relied exclusively on registry-confirmed clinical obesity diagnoses as the primary outcome, thereby avoiding dependence on self-reported anthropometric measurements.

RESULTS

An overview of data availability across questionnaire and omics-based analyses is provided in Fig. 2. BMI-based outcomes (defined according to standard IOTF classifications on the questionnaire-based data across time) were used to characterize growth patterns and stratify participants for proteomic and microbiome analyses, as well as to assess associations with metabolite levels in cord blood. They were not used as outcomes in the integrated multi-omic machine learning models, which relied exclusively on clinically diagnosed obesity from the National Patient Register.

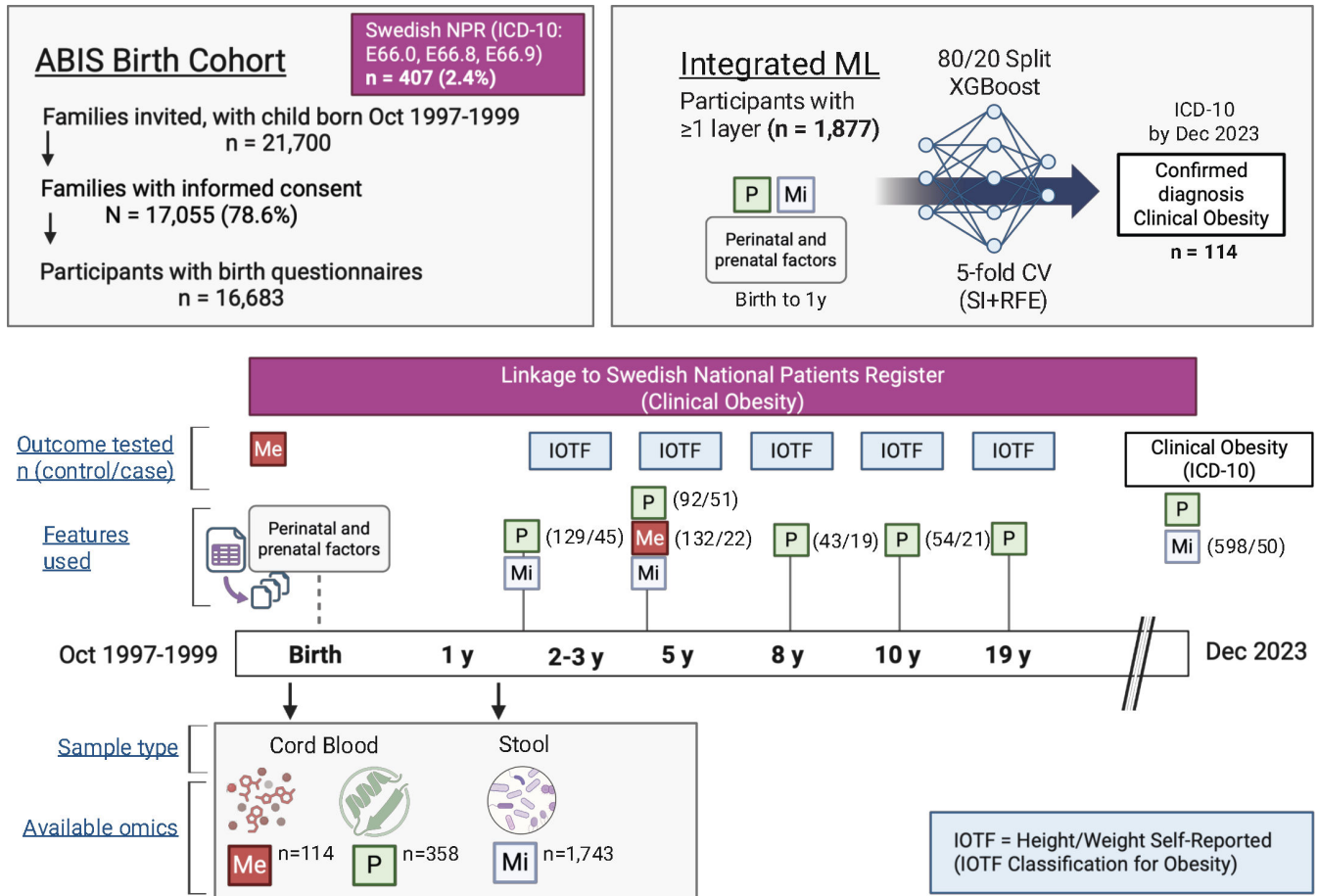


FIG 2 Analytical process. The pipeline for the investigation, demonstrating the metabolomic (Me), proteomic (P), and microbial (Mi) data sets used in each analysis.

A total of 407 ABIS children were diagnosed with obesity (BMI across time shown in Table 1). While only 12.4%–13.8% of children with clinically diagnosed obesity were obese at age 3, 27.8% of females with obesity at age 5 remained clinically obese, as did 53.7% of males (Table 1). Individuals who met IOTF obesity criteria but did not receive a clinical diagnosis were retained in BMI-based analyses but were classified as non-cases for registry-based analyses, reflecting known underdiagnosis of obesity in clinical care rather than exclusion or misclassification. Registry-based ICD-10 obesity diagnoses were used as the primary clinical outcome for proteomic, microbiome, and integrated multi-omic analyses.

Associations with early-life infection, lifestyle, and environment

Parent questionnaires (administered at birth and age one, n = 16,683) were assessed for associations with either clinical diagnosis of obesity or obesity classification based on IOTF standards (at 3, 5, 8, 10, and 19 years). Significant findings are presented in Fig. 3 (see also Fig. S1).

Females were more likely than males to later receive an obesity diagnosis (OR = 1.97 [1.61–2.43, 95% CI], P < 0.0001) or to be obese at ages 3 or 5 (OR = 1.41 [1.06–1.87, 95% CI], P < 0.019 and OR = 1.29 [1.02–1.64], P = 0.036, respectively). Maternal gestational diabetes mellitus (ICD-10 O24.4) was associated with obesity at age 3 (OR = 3.39 [1.35–8.53], P = 0.0096) only. Lower maternal and paternal education levels were also associated with a later clinical diagnosis of obesity in this child (OR = 2.52 [1.92–3.30], P < 0.0001 and OR = 2.61 [1.91–3.567, 95% CI], P < 0.0001, respectively), as well as

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.200.55.

TABLE 1 Obesity in the ABIS cohort^a

Parameter	IOTF	Male future diagnosis of obesity				Female future diagnosis of obesity			
		No		Yes		No		Yes	
		Count	%	Count	%	Count	%	Count	%
BMI at 2 years	Normal or overweight	4,055	98.10	50	86.21	3,690	97.46	113	87.60
	Obese	78	1.89	8	13.79	96	2.54	16	12.40
BMI at 5 years	Normal or overweight	3,392	97.11	25	46.30	2,974	96.12	78	72.22
	Obese	101	2.89	29	53.70	120	3.88	30	27.78
BMI at 8 years	Normal or overweight	1,531	98.27	15	75.00	1,371	97.79	27	75.00
	Obese	27	1.73	5	25.00	31	2.21	9	25.00
BMI at 10 years	Normal or overweight	1,640	97.97	10	55.56	1,616	98.54	28	75.68
	Obese	34	2.03	8	44.44	24	1.46	9	24.32
Obese at 19 years	Normal or overweight	2,316	95.35	32	76.19	2,753	95.23	61	71.76
	Obese	113	4.65	10	23.81	138	4.77	24	28.24

^aPrevalence of obesity based on clinical diagnosis (ICD-10 codes) and IOTF-defined classifications from self-reported data at 3, 5, 8, 10, and 19 years.

with obesity measurements at ages 3, 5, 8, 10, and 19 years (ORs = 1.44–2.28). Children born to overweight or obese mothers had a 1.3- to 6.2-fold increased risk of obesity, depending on the classification and timepoint. Maternal obesity carried the greatest risk, increasing the likelihood of clinical obesity diagnosis in children by 3.1–11.4 times (all P 's < 0.0001). Delivery mode was not significantly associated with obesity risk, though cesarean section showed marginal associations at the 3- and 19-year time points. Higher psychosocial vulnerability index scores (19, 26) were associated with obesity at 3 and 5 years as well as clinical diagnosis of obesity later in life (OR = 1.57 [1.21–2.02, 95% CI], P = 0.0006).

Otitis media increased the odds of clinical obesity diagnosis (OR = 1.45 [1.11–1.90, 95% CI], P = 0.0062), while stomach flu was associated with obesity at age 10 (OR = 1.91 [1.14–3.22, 95% CI], P = 0.0143). Having frequent infections requiring penicillin (three or more times in the first year) increased the odds of a future obesity diagnosis by twofold (OR = 2.08 [1.32–3.22, 95% CI], P = 0.001). Maternal smoking during pregnancy (OR = 1.73–3.23) and exposure to smoking in the home during the first year also elevated risk (OR = 1.58–2.44). Some, but not many, dietary exposures in the first year were identified. Early cessation of breastfeeding (at 1–3 months) was associated with obesity at age 3 and 5, as well as later clinical obesity diagnosis. Early introduction of formula (at 1 month) and carbohydrate-heavy gruel (at 1–3 months) were significantly linked to obesity at all time points—particularly strong for clinical diagnosis (OR = 2.40 [1.68–3.42, 95% CI], P < 0.0001). Early gluten introduction (at 1–3 months) was a significant exposure (P = 0.0001).

Although obesity is multifactorial, an AUC of 0.66–0.67 was achieved using a model with a limited set of early lifestyle factors: early introduction of gruel (a carbohydrate-dense porridge, previously linked to high BMI at age 12 and 18 months (27)), penicillin-treated infections during the first year, otitis, prenatal and household smoking, biological sex, and parental education at birth.

Associations with proteomic markers at birth

Proteomic differences were analyzed across BMI classifications at 3, 5, 8, and 10 years (Table S1) by comparing 72 obese children with 218 controls. Across all ages, 48 proteins were significant, but none after FDR correction (Fig. 4A). In the 2- to 3-year group (45 overweight/obese vs. 129 controls), 52 proteins were significant, 2 of which (follistatin [FST] and hepatocyte growth factor [HGF]) remained significant after FDR (Fig. 4B through D). At 5 years (51 overweight/obese vs. 92 controls), 59 proteins were identified, but none survived FDR (Fig. 4E). FST and HGF were still elevated, albeit with a weaker association. At 8 years (19 overweight/obese vs. 43 controls), 21 proteins were significant, with TNFRSF11B remaining after FDR correction and higher in the overweight/obese group (P = 0.000034; Fig. 4F through H). Elevated ANGPTL4 was also notable (P = 0.0034).

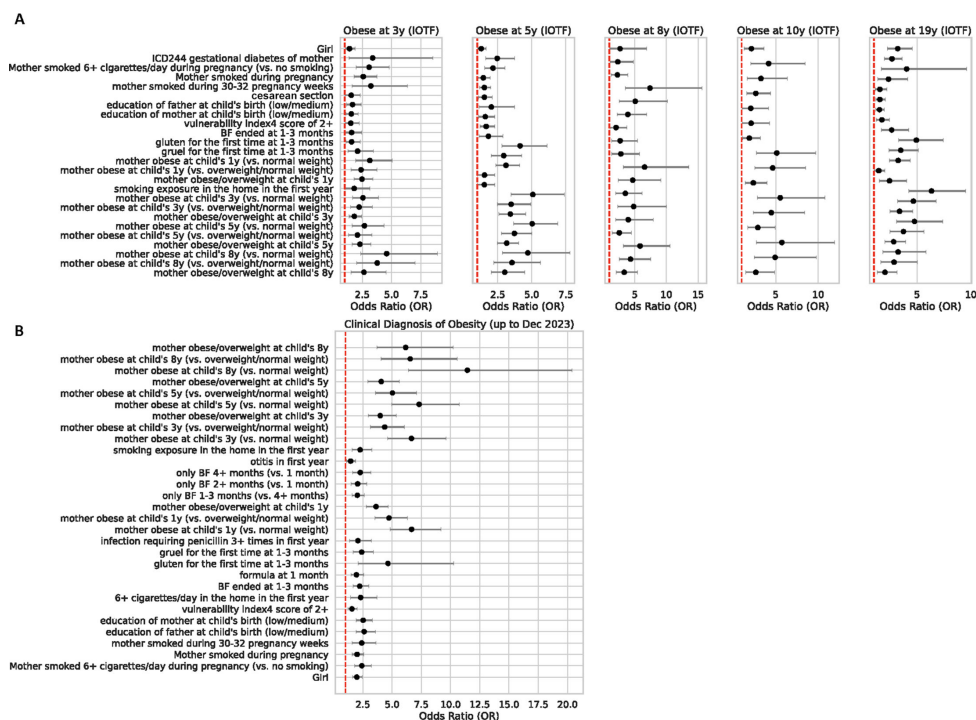


FIG 3 Early-life factors associated with future clinical obesity diagnosis or obesity classification according to IOTF standards (ages 3–19 years). ORs and 95% confidence intervals (CIs) are provided for each reference group and comparison, stratified by IOTF classifications and future obesity diagnosis (up to December 2023). Factors associated with (A) obesity classifications based on BMI at 3, 5, 8, 10, and 19 years of age against IOTF standards, and (B) clinical diagnosis of obesity from the medical record. ORs represent unadjusted associations between exposure and outcome.

At 10 years (21 overweight/obese vs. 54 controls), 39 proteins were identified, but none were significant after FDR.

Among newborns who later received a clinical diagnosis of obesity (up to 26 years later), 84 proteins were differentially expressed, with 19 remaining significant after FDR correction (Fig. 5A and Table S2). All but one of these 19 were more elevated in the obesity group (Fig. 5B through K), while IDS was the lone protein higher in controls at birth. Although mother's weight was associated with ANGPTL4 expression ($P = 1.547e-05$), significance remained after controlling for this factor ($P = 3.45e-06$), with no interaction effect found in ANCOVA ($P = 0.221$).

Using mutual information (MI) scoring, 15 proteins were associated with clinical obesity prediction (Fig. 5L). An XGBoost model for obesity measurement at age 19 years achieved an AUC-ROC of 0.67 (Fig. S2A) and the histogram-based gradient boosting classification tree an AUC of 0.66 for a future diagnosis of obesity (Fig. S2B), with the highest performing model being the XGBoost approach with an AUC of 0.71 (Fig. S2C). Including four additional proteins (ANGPTL4, FST, HGF, and TNFRSF11B) identified in the previous analyses improved the AUC-ROC to 0.71 (Fig. 5M).

Metabolomic differences at birth associated with obesity at age 5 years

Cord serum metabolite concentrations were compared between the normal-weight group ($n = 132$) and a group of overweight or obese children at 5 years of age ($n = 22$). Isocaproic acid ($P = 0.19$; Fig. 5N), tryptophan ($P = 0.019$; Fig. 5O), and oleic acid ($P = 0.035$, Fig. 5Q) were significantly lower in infants who were later overweight/obese.

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.200.55.

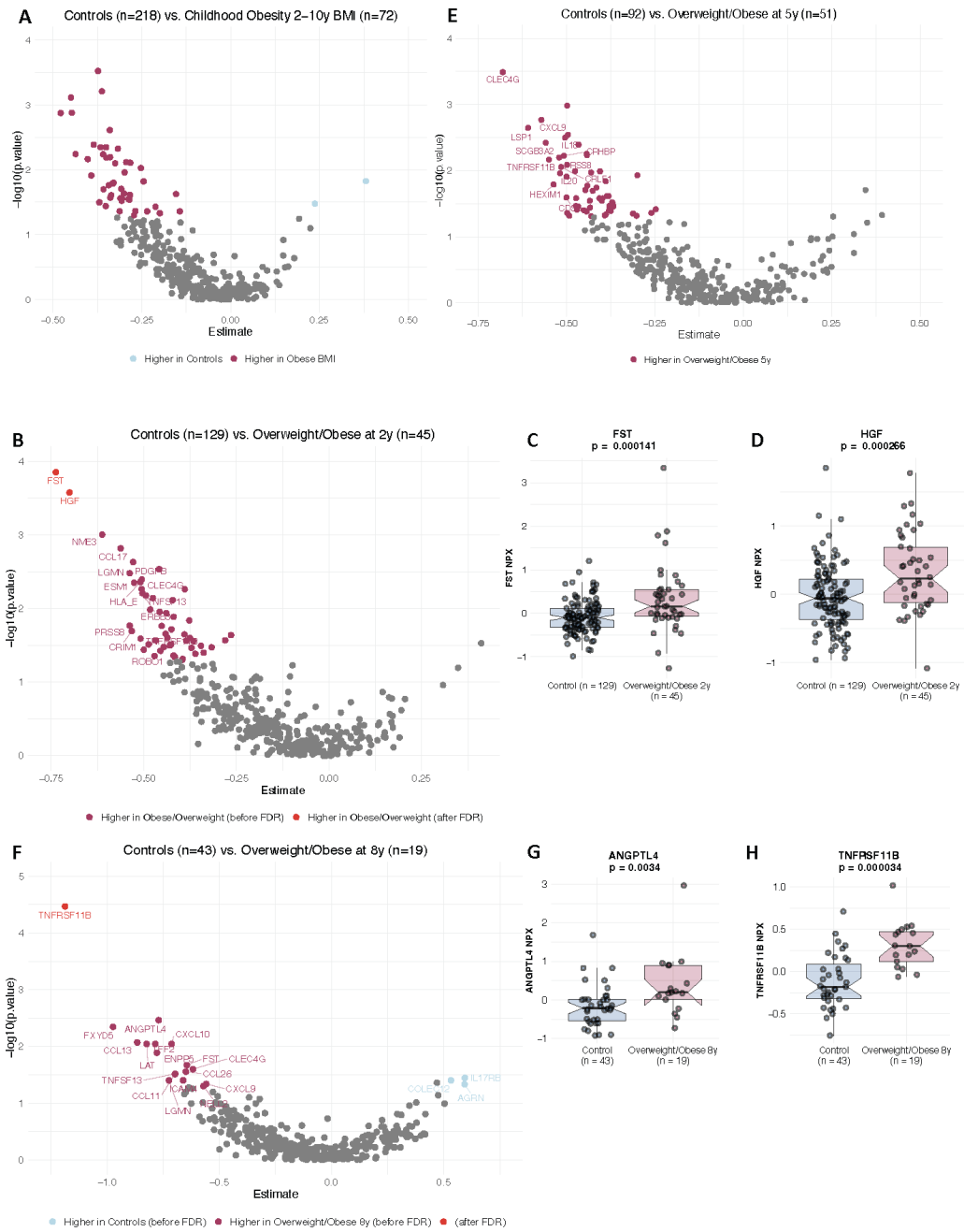


FIG 4 Proteomic differences in cord blood based on childhood BMI classification. Wilcoxon test statistics comparing NPX values of proteins in cord blood, based on BMI classifications across time. Controls are defined as normal-weight by IOTF standards, based on a parent-completed questionnaire, and with no future autoimmune or neurodevelopmental diagnosis from the Swedish National Patient Register. Comparisons include the following: 2–10 years global comparison (218 controls vs. 72 a BMI corresponding to obesity at least once across the timepoints), BMI at 2 years (129 controls vs. 45 overweight/obese), BMI at 5 years (92 controls vs. 51 overweight/obese), and BMI at 8 years (43 controls vs. 19 overweight/obese). (A) Children with BMIs indicative of obesity per IOTF standards at any point between 2 and 3, 5, 8, or 10 years of age were compared to controls using Wilcoxon tests with FDR correction. (B–H) Children classified as overweight or obese per the IOTF definition were compared to controls: (B–D) at age 2 years, (E) at age 5 years, (F–H) at age 8 years. No proteomic findings in panels A or E reached significance after FDR correction. See Table S1 for statistics.

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.200.55.

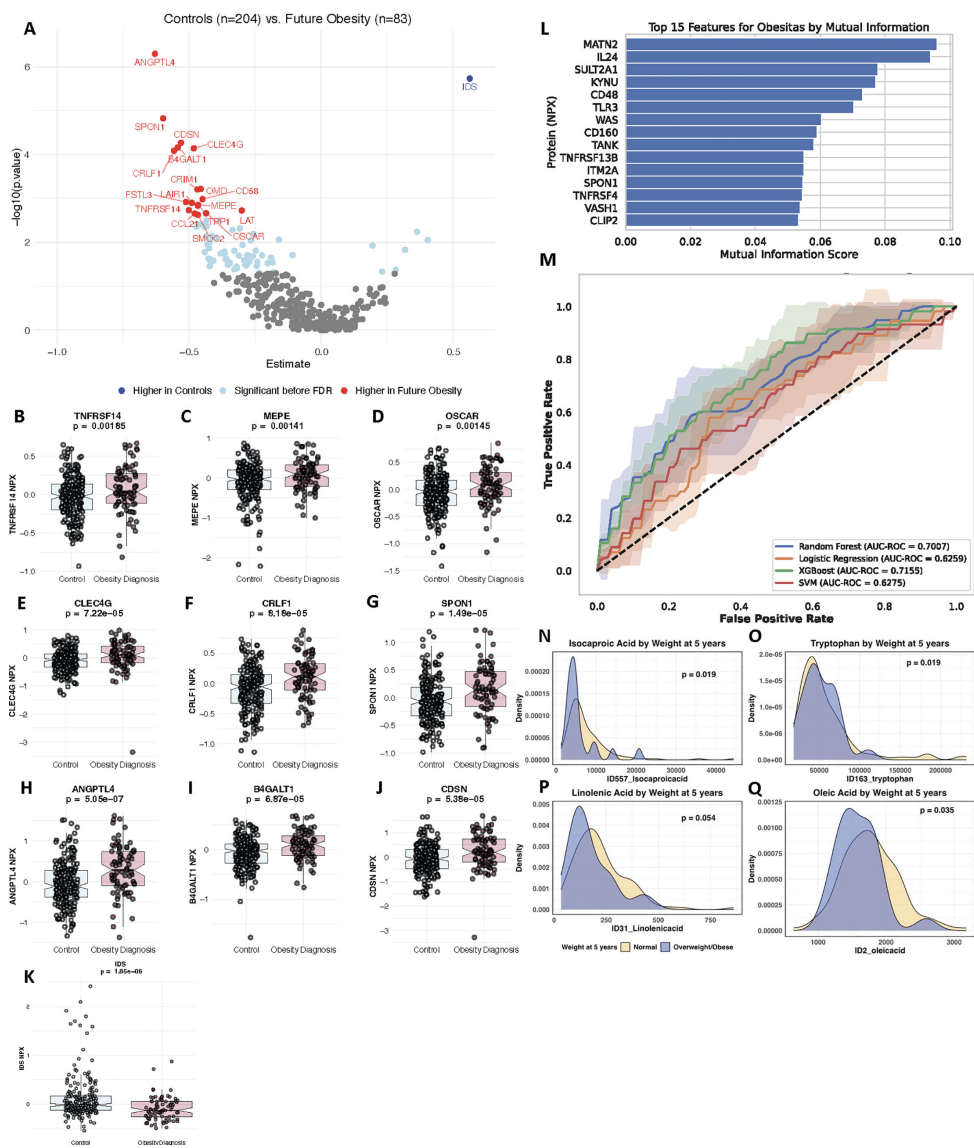


FIG 5 Birth proteome and metabolome associations with future clinically diagnosed obesity. (A) Wilcoxon test results for cord blood protein levels in relation to future ICD-10 diagnosis of clinical obesity (identified in the Swedish National Patient Register), corrected by FDR correction. (B–K) NPX levels for top hits, stratified by future obesity diagnosis case status. See Table S2 for statistics. (L) Top 15 proteins contributing to case–control discrimination of obesity (obesitas), based on mutual information (MI) scores. (M) Predictive performance of ML and logistic regression models predicting future obesity diagnosis using the top 15 proteins from panel K. (N) Improved AUC-ROC (means and standard deviation) after incorporating four additional proteins—ANGPTL4, FST, HGF, and TNFRSF11B—identified in earlier obesity comparisons. (O–P) Concentrations of cord serum metabolites differed between normal-weight and overweight infants, with density curves representing the concentration distributions.

Predicting metabolites and exogenous compounds in cord blood from prenatal/perinatal features

ML models were developed to predict concentrations of fatty acids, bile acids, amino acids, and environmental toxins (Fig. S3A and B) from prenatal and perinatal factors, assessing which early-life factors contain predictive information about the cord blood composition. Table S3 presents the top features predicted by cord serum, selected for their predictive performance, achieving a balance between minor and major class accuracy. SHapley Additive exPlanation values were used to interpret which features

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.200.55.

contribute most to the model's predictions, not to establish causal relationships. Thus, the identified associations represent predictive patterns rather than causal effects. Among the most predictable features with the highest accuracies included delivery mode, week of delivery, and parity, as well as mother's diet during pregnancy, such as coffee consumption, dairy intake, coffee bread consumption, and the type of fat used on sandwiches, as well as maternal smoking during pregnancy three months prior to conception. Children born to mothers who smoked had higher 2-hydroxyacetophenone-sulfate [$n = 42$; Kruskal–Wallis $\chi^2(1) = 45.7$, $P = 1.394 \times 10^{-11}$], with an average peak intensity of 9.35 (SD = 8.8), compared to 2.00 (SD = 3.4) in controls (Fig. S4A). The prediction accuracy for smoking exposure (AUC = 0.80 ± 0.007 ; Fig. S4B) was driven primarily by 2-hydroxyacetophenone-sulfate, a derivative of acetaminophen (paracetamol). This metabolite was among the most accurately predicted by prenatal factors ($R^2 = 0.30$; Fig. S4C), including coffee intake during pregnancy (Fig. S4D and E). Delivery mode associated greatly with the cord blood metabolome (Fig. S4F and G), demonstrating the strongest AUC (0.96 ± 0.02 , Fig. S4F).

Significant microbiome associations with future clinical obesity diagnosis

Sellimonas intestinalis, Family XIII AD3011 group spp., and *Megasphaera micronuciformis* were elevated in children who were later diagnosed with obesity, whereas *Senegalimassilia* spp., *Slackia isoflavoniconvertens*, *Adlercreutzia equolifaciens*, *Phascolarctobacterium faecium*, *Bifidobacterium dentium*, *Ruminococcus gauvreauii* group spp., and *Akkermansia* spp. were markedly depleted (Fig. 6A and B). *Lachnospiraceae* NK4A136 group bacterium, *Veillonella rogosae*, *Roseburia* spp., and *Senegalimassilia* spp. showed reduced prevalence in children later diagnosed with obesity. *Senegalimassilia* spp. was entirely absent in this group (Fig. 6C). Conversely, *Enterobacter* spp. and *Sellimonas intestinalis* were more prevalent (Fig. 6C).

Among the species more abundant in matched controls (those without weight issues at ages 2–3 or 5 years) were *Phascolarctobacterium faecium*, *Akkermansia* spp., *Coprococcus comes*, *Coprococcus eutactus*, *Slackia isoflavoniconvertens*, *Bacteroides thetaiotaomicron*, *Parabacteroides merdae*, *Senegalimassilia* spp., and *Bifidobacterium dentium*. *Senegalimassilia* spp. was detected in only one case at both age 3 and 5 (1.75% prevalence compared to 9.97% in matched controls). Comparisons of the full data sets and the propensity score-matched sample sets are provided in Table 2 to demonstrate covariate balance.

In contrast, *Bacteroides eggerthii* was more abundant in infants who later had weight issues. The bacterium was seen in only 2.1% of infants with normal weight at both 3 and 5 years, but in 7.7% of infants later classified as obese or overweight at 5 years [$\chi^2(1) = 8.5$, $P = 0.0035$; Fig. 6D]. Similarly, ASV-103 *Klebsiella* sp. was more abundant in infants who later became obese at ages 3 or 5 years. This strain was detected in 14.2% and 11.0% of infants with normal weight or overweight at 3 years, respectively, but in 35.5% of cases. At age 5, it was detected in 12.7% of infants with normal weight, compared to 19.7% overweight and 26.2% obese at this age. Although no enterotype-level differences were found based on future diagnosis (Fig. 6E), species-level distinctions suggest more precise differences in the microbiome significant for predicting future weight issues.

To address potential confounding factors influencing microbiome composition, a strict iteration of propensity score matching was applied in a 3:1 ratio using nearest neighbor. This matching included gestational diabetes, the mother's pre-pregnancy BMI, geographic location, gastroenteritis in the first year, total months of breastfeeding, antibiotic exposure during the first year, mode of delivery, biological sex, and psychosocial vulnerability. The psychosocial vulnerability score was calculated from a set of variables including living conditions, parental employment and education, serious stressful life events, disposable income, alcohol, and smoking (21). This adjustment revealed 26 significant taxa when comparing 39 controls against 13 cases at age 2–3, and 30 taxa for cases at age five (22 cases vs. 66 controls), defined by an FDR-adjusted $P < 0.05$ and a normalized base mean of abundance ≥ 10 .

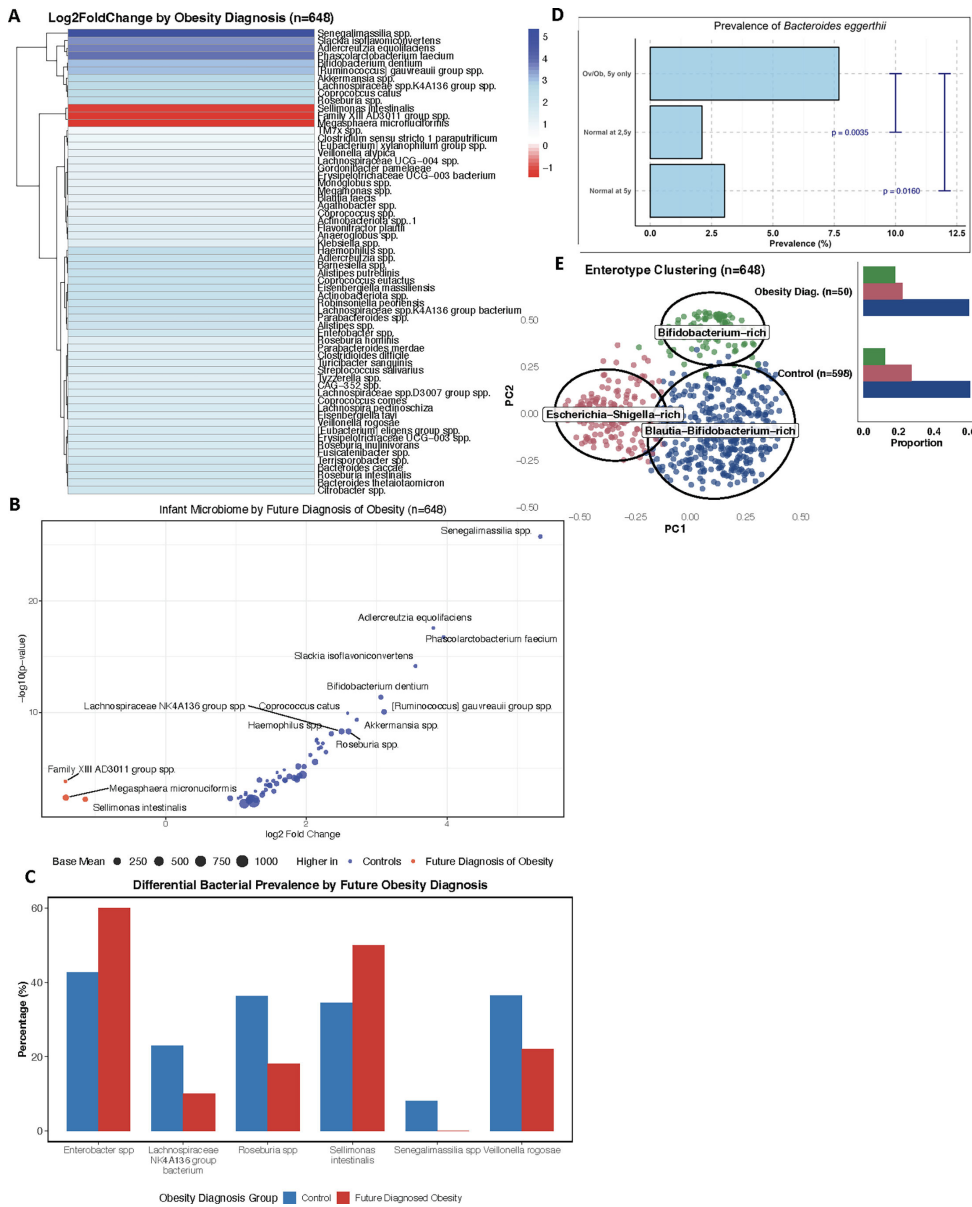


FIG 6 Differentially abundant or prevalent bacterial species by future diagnosis of obesity. Differentially abundant or prevalent bacterial species in future clinical diagnosis of obesity. (A and B) Significant differences in bacterial species abundance, after FDR correction. Controls were defined as those without obesity at the 2–3 years and 5 years timepoints, based on their BMI classifications according to the IOTF standards, as well as no future autoimmune or neurodevelopmental diagnosis. A total of 598 controls were compared to 50 future clinical obesity cases. The 16s rRNA sequencing data were assessed for differentially abundant species between obesity and control groups using DESeq2, with a binomial distribution employed for modeling the counts and using a local fit type to adjust for dispersion differences within the data set. This is important for the heterogeneous variance seen in these data sets. Multiple testing was adjusted for using the FDR method. (C) Significant differences in the prevalence of bacterial species. Chi-square tests assessed differences in bacterial prevalence between control and obesity groups, with Fisher’s exact tests used for taxa with low counts. For each significant taxon, prevalence and percentage presence/absence were calculated for each group. (D) Differences in the prevalence of *Bacteroides egerthii*. (E) Enterotypes representative of obesity diagnosis and control groups, and the proportion of groups belonging to each.

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.2.200.55.

TABLE 2 Baseline characteristics before and after propensity score matching for differential abundance testing^a

Parameter	Comparisons (SD)							
	Obesity 2 yrs (all)		Obesity 2 yrs (propensity)		Obesity 5 yrs (all)		Obesity 5 yrs (propensity)	
	0	1	0	1	0	1	0	1
<i>n</i>	648	13	39	13	501	22	66	22
BMI pre-pregnancy (mean)	24.2 (3.9)	26.2 (4.3)	25.9 (4.3)	26.2 (4.3)	24.3 (4.0)**	26.3 (4.0)**	25.9 (4.8)	26.3 (4.0)
Region (%)								
East	146 (23)	2 (15)	9 (23)	2 (15)	106 (21)	2 (9.1)	5 (7.6)	2 (9.1)
North	180 (28)	5 (39)	19 (49)	5 (39)	140 (28)	9 (41)	25 (38)	9 (41)
South	158 (24)	2 (15)	2 (5.1)	2 (15)	130 (26)	6 (27.3)	23 (35)	6 (27)
West	164 (25)	4 (31)	9 (23)	4 (31)	125 (25)	5 (23)	13 (20)	5 (23)
Gastroenteritis in 1st year (yes, %)								
1 = yes	185 (29)	6 (46)	16 (41)	6 (46)	142 (28)	8 (36)	23 (35)	8 (36)
2 = no	447 (69)	7 (54)	22 (56)	7 (54)	345 (69)	14 (64)	42 (64)	14 (64)
3 = unk	16 (2.5)	0 (0.0)	1 (2.6)	0 (0.0)	14 (2.8)	0 (0.0)	1 (1.5)	0 (0.0)
Total breastfeeding, months (%)								
1	24 (3.7)	1 (7.7)	1 (2.6)	1 (7.7)	17 (3.4)	0 (0.0)	5 (7.6)	0 (0.0)
2	25 (3.9)	2 (15)	3 (7.7)	2 (15)	18 (3.6)	1 (4.5)	2 (3.0)	1 (4.5)
3	30 (4.6)	0 (0.0)	2 (5.1)	0 (0.0)	21 (4.2)	1 (4.5)	4 (6.1)	1 (4.5)
4	20 (3.1)	0 (0.0)	2 (5.1)	0 (0.0)	19 (3.8)	0 (0.0)	3 (4.5)	0 (0.0)
5	25 (3.9)	0 (0.0)	2 (5.1)	0 (0.0)	21 (4.2)	2 (9.1)	4 (6.1)	2 (9.1)
6	59 (9.1)	0 (0.0)	3 (7.7)	0 (0.0)	42 (8.4)	4 (18.2)	4 (6.1)	4 (18)
7	70 (11)	2 (15)	2 (5.1)	2 (15)	51 (10)	3 (14)	10 (15)	3 (14)
8	104 (16)	4 (31)	6 (15)	4 (31)	79 (16)	6 (27)	6 (9.1)	6 (27)
9	291 (45)	4 (31)	18 (46)	4 (31)	233 (46)	5 (23)	28 (42)	5 (23)
Infection with antibiotics in first year (%)								
1 = none	435 (67)	2 (15)	12 (31)	2 (15)	342 (68)	14 (64)**	52 (79)**	14 (64)
2 = 1–2×	175 (27)	9 (69)	13 (33)	9 (69)	130 (26)	8 (36)	12 (18)	8 (36)
3 = 3–5×	36 (5.6)	2 (15)	12 (31)	2 (15)	29 (5.8)	0 (0.0)	2 (3.0)	0 (0.0)
4 = more	2 (0.3)	0 (0.0)	2 (5.1)	0 (0.0)				
Gestational age, weeks (%)								
33	2 (0.3)	0 (0.0)			2 (0.4)	0 (0.0)	1 (1.5)	0 (0.0)
34	1 (0.2)	0 (0.0)			2 (0.4)	0 (0.0)		
35	6 (0.9)	0 (0.0)			4 (0.8)	0 (0.0)	1 (1.5)	0 (0.0)
36	14 (2.2)	0 (0.0)	1 (2.6)	0 (0.0)	13 (2.6)	0 (0.0)	1 (1.5)	0 (0.0)
37	36 (5.6)	0 (0.0)	2 (5.1)	0 (0.0)	24 (4.8)	3 (13.6)	1 (1.5)	3 (14)
38	58 (9.0)	1 (7.7)	2 (5.1)	1 (7.7)	44 (8.8)	2 (9.1)	2 (3.0)	2 (9.1)
39	133 (21)	3 (23)	6 (15)	3 (23)	112 (22)	1 (4.5)	18 (27)	1 (4.5)
40	185 (29)	2 (15)	12 (31)	2 (15)	144 (29)	4 (18)	15 (23)	4 (18)
41	133 (21)	4 (31)	8 (21)	4 (31)	110 (22)	8 (36)	16 (24)	8 (36)
42	67 (10)	3 (23)	6 (15)	3 (23)	41 (8.2)	4 (18)	10 (15)	4 (18)
43	13 (2.0)	0 (0.0)	2 (5.1)	0 (0.0)	5 (1.0)	0 (0.0)	1 (1.5)	0 (0.0)
Delivery mode (C-section, %)	74 (11)	3 (23)	10 (26)	3 (23)	59 (12)	4 (18)	11 (17)	4 (18)
Sex (female, %)	316 (49)	7 (54)	26 (67)	7 (54)	231 (46)*	16 (73)*	49 (74)	
Vulnerability index 4 (%)								
0 = low	321 (50)	6 (46)	16 (41)	6 (46)	243 (49)	8 (36)	19 (29)	8 (36)
1	241 (37)	5 (39)	16 (41)	5 (39)	198 (40)	10 (46)	35 (53)	10 (46)
2	69 (11)	2 (15)	6 (15)	2 (15)	51 (10)	2 (9.1)	11 (17)	2 (9.1)
3 = high	17 (2.6)	0 (0.0)	1 (2.6)	0 (0.0)	9 (1.8)	2 (9.1)	1 (1.5)	2 (9.1)

^aComparisons between groups were performed using the CreateTableOne package in R. By default, continuous variables were compared using a t-test (mean ± SD reported for BMI pre-pregnancy), and categorical variables were compared using a chi-square test, with Fisher's exact test used for variables with expected cell counts <5. Only significant *P*-values below 0.05 are shown. ****P* = 0.001, ***P* = 0.021. **P* = 0.026, unk, unknown.

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.200.55.

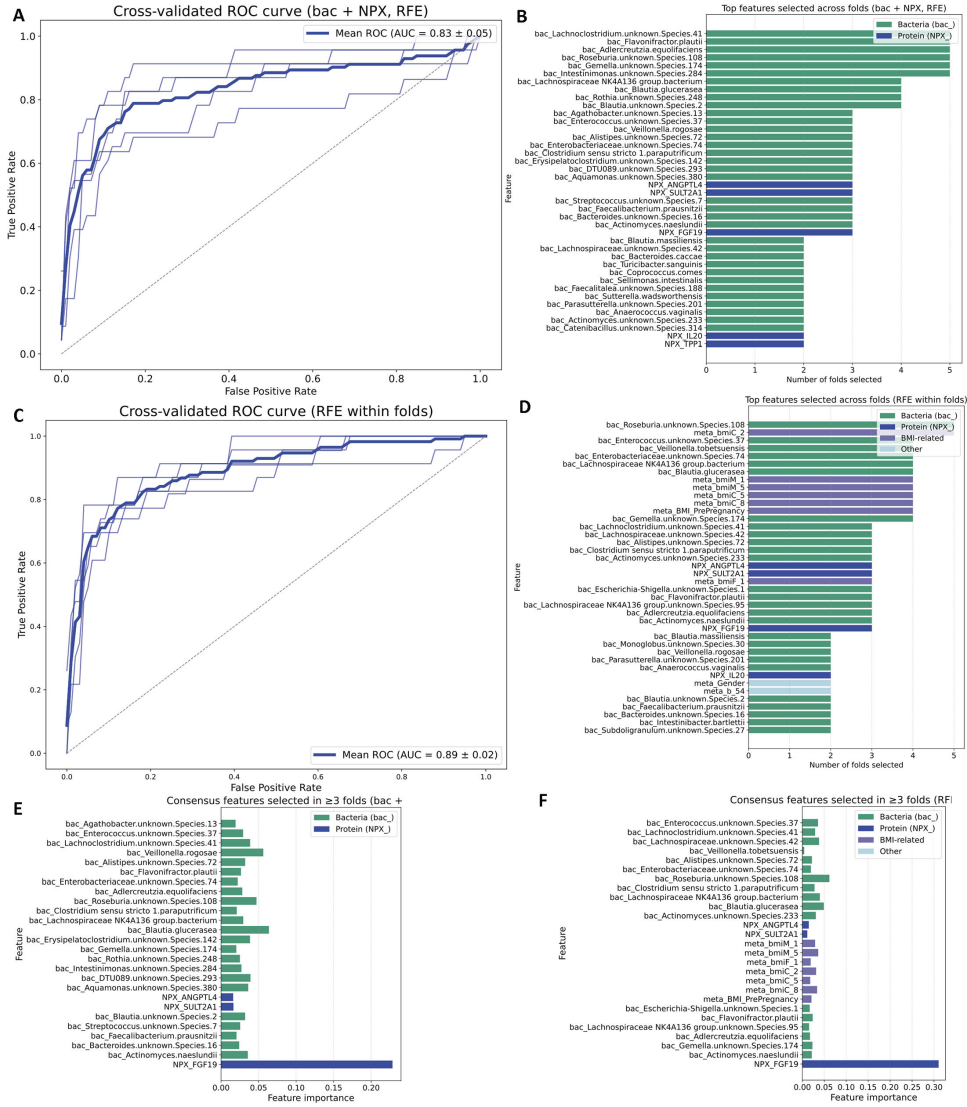


FIG 7 Integrated machine learning models predicting future clinical diagnosis of obesity. Prediction of clinical obesity diagnosis in 1,877 individuals, based on the XGBoost Classifier with an 80/20 train/test split and RFE integrated within fivefold stratified cross-validation. The primary outcome was clinical diagnosis in the Swedish NPR (22). The integrated predictive models relied exclusively on registry-confirmed clinical obesity diagnoses as the primary outcome, thereby avoiding dependence on self-reported anthropometric measurements. Panels A and B show predictions based on proteins at birth and taxa present in one-year stool samples, while panels C and D additionally incorporate metadata from the birth and 1-year questionnaires. For each fold, the training set was used to select the top 40 features via RFE (from a total of 683 features, including 230 bacterial and 363 proteomic). The model was subsequently trained on these features and evaluated on the held-out test set. L1 regularization ($\alpha = 1$) and class weighting (scale_pos_weight) were applied to handle potential overfitting and class imbalance. Missing values were handled using median imputation fit on the training data within each fold and then applied to the corresponding test set, preventing information leakage across folds. Performance metrics—including ROC AUC, F1 score, precision, and recall—were averaged across the folds. The top features selected across folds were summarized by counting the number of times each feature was chosen. The selected features indicate patterns associated with obesity risk rather than independent causal effects. AUC average and standard deviation across folds are shown in panels A and C, with feature importances for the top features in panels B and D. Feature naming conventions are as follows: bacterial features begin with “bac_” while proteomic birth features begin with “NPX_” Environmental and anthropometric variables include meta_b_54, corresponding to dietary vegetable intake frequency at 1 year of age; maternal BMI measures (meta_BMI_PrePregnancy, meta_bmiM_1, meta_bmiM_5); paternal BMI at 1 year (meta_bmiF_1); and child BMI at ages 2–3, 5, and 8 years (meta_bmiC_2, meta_bmiC_5, and meta_bmiC_8). (E) Feature importance of consensus predictors (Continued on next page)

Downloaded from https://journals.asm.org/journal/mSystems on 15 June 2026 by 130.232.200.55.

Fig 7 (Continued)

selected in at least three cross-validation folds. Features retained in every fold (consensus features) are shown, with colors indicating feature type: proteins (NPX_, blue), bacteria (bac_, green), and other metadata features (gray). Feature importance values represent the contribution of each feature to a final XGBoost model predicting future obesity diagnosis.

Prediction of future clinical obesity diagnosis using multi-omics in the first year of life

Among 1,877 ABIS individuals (including 114 cases with a future clinical obesity diagnosis), microbiomic and/or proteomic data from early life were available. These data were used to predict future clinically diagnosed obesity using eXtreme Gradient Boosting. Machine learning models incorporated L1 regularization, class-imbalance weighting, and stratified fivefold cross-validation. Metadata were integrated to account for possible interactions between biological and non-biological features. Clinical obesity diagnosis was defined exclusively using registry-confirmed diagnoses from the Swedish National Patient Register.

To prevent information leakage, all preprocessing and feature selection steps were performed independently within each cross-validation fold. Specifically, RFE was applied using only the training data in each fold to select the top 40 predictors, enhancing model interpretability, after which models were trained on these features and evaluated on the corresponding held-out test set. Median imputation was likewise performed separately within testing and training partitions. In total, 683 candidate features were evaluated per fold, including 363 proteomic markers and 230 bacterial taxa, along with questionnaire data from birth and 1-year surveys, and calculated BMI scores up to age 8 (for the child and parents).

Models integrating both microbiomic and proteomic data achieved a mean ROC AUC of 0.83 ± 0.05 (Fig. 7A) with a mean F1 score of 0.53 ± 0.06 . At the selected classification threshold, mean precision (0.51 ± 0.06) indicated that approximately half of the individuals predicted to develop obesity were true positives, while mean recall (0.56 ± 0.10) demonstrated that the model correctly identified over half of all true future obesity cases. These results indicate robust discriminative performance while maintaining a balance between sensitivity and specificity. The most predictive feature was FGF19 in cord blood, followed by ANGPTL4 and SULT2A1 as proteomic markers at birth (Fig. 7B and E) and consensus bacterial markers at age 1 (identified in ≥ 3 folds), including *Blautia glucerasa*, *Veillonella rogosae*, *Roseburia* spp., *Erysipelatoclostridium* spp., *Actinomyces naeslundii*, and *Lachnoclostridium* spp. (Fig. 7B and E).

In models additionally incorporating early-life metadata and early-life BMI metrics, discriminative performance improved slightly further, achieving a mean ROC AUC of 0.89 ± 0.02 , indicating strong overall discriminative performance (Fig. 7C). Feature stability across folds was summarized by counting the number of times the feature was chosen (Fig. 7D). The most frequently selected proteomic predictors included ANGPTL4, SULT2A1, FGF19, and IL20 (Fig. 7D), while predictive bacterial included *Roseburia* spp., *Enterococcus* spp., *Veillonella tobetsuensis*, *Enterobacteriaceae* spp., *Lachnospiraceae* NK4A136 group, *Blautia glucerasea*, and *Gemella* spp. (Fig. 7D). Parental BMI (pre-pregnancy and during early childhood) and child BMI at ages 2–3, 5, and 8 years also emerged as consistently informative features (Fig. 7D). Among consensus predictors (identified in at least three of the five folds), FGF19 exhibited the highest feature importance (score > 0.3), followed by *Roseburia* spp., *Blautia glucerasea*, *Lachnospiraceae* spp., *Enterococcus* spp., *Lachnospiraceae* NK4A136 group, the BMI metrics (from mother, father, and child), and ANGPTL4 (Fig. 7E).

Accuracy metrics for both the Bacterial +Proteomic and the Bacterial +Proteomic + Metadata models are provided in Table S4. Together, these findings indicate that, even across a 26-year predictive gap, cord blood protein profiles, early-life bacterial profiles, and parental anthropometrics provide stable and biologically informative signals associated with long-term obesity risk.

DISCUSSION

Studying early-life biological markers associated with metabolic disorders and obesity can identify targets useful in future studies for the prediction and prevention of these disorders. While other prospective studies (18, 28–32) have explored these links, this is the first to examine microbial differences at age 1 year, alongside metabolomic and proteomic differences at birth, using integrated prediction models powered by interpretable algorithmic machine learning methods. This focus allows identification of critical markers of early-life programming, long before clinical outcomes manifest.

The microbiome is influenced by many well-studied factors, some of which were controlled for here, including breastfeeding practices, antibiotic use (19, 33–35), gastroenteritis, maternal gestational diabetes, pre-pregnancy weight, and psychosocial factors. Parental obesity increases the risk of a child becoming obese as an adult (5). Infants born to obese mothers have reduced *Proteobacteria* and increased *Coriobacteriaceae*, *Erysipelotrichaceae*, *Lachnospiraceae*, and *Ruminococcaceae* (36). Children born to an overweight mother by cesarian section are at the greatest risk of obesity between ages 1 and 3 years, with increased abundance of *Lachnospiraceae* possibly associated with changes in weight (36).

Akkermansia muciniphila is associated with immune system regulation (37), mucin health, anti-inflammatory processes (38–40), and metabolic status (8, 16, 41, 42). Here, *Akkermansia* spp. were less abundant in infants with future obesity. Conversely, *B. eggerthii*, known to promote colitis (43), and possessing a polysaccharide-utilizing locus (44), was higher. Asaccharolytic bacteria, that is, unable to metabolize sugars, *Senegalimassilia* and *Phascolarctobacterium*, were more abundant in controls, while equol-producing bacteria *Slackia* and *Adlercreutzia* were higher in infants with future obesity, which fits the previous observation that equol has been associated with lower BMI, fat mass percentage, and serum triglycerides in a community-based prospective study of adults (45). Equol and equol-producing microbes have also been favorably associated with other obesity markers like liver fat (46). *Klebsiella* ASV-103, more prevalent in infants with future weight problems, is the same strain that we found more abundant in infants with future neurodevelopmental disorders (19), often co-occurring with obesity (47).

Among the most significantly elevated proteins at birth in those with future obesity were FST, HGF, and ANGPTL4. Circulating FST has been shown to be elevated in patients with type 2 diabetes, potentially due to adipose insulin resistance and metabolic dysregulation (48). HGF, which plays a key role in insulin resistance through the HGF/c-Met signaling pathway (49), is elevated in conditions associated with insulin resistance (50) and obesity (51). ANGPTL4, a protein studied with respect to metabolic and cardiovascular diseases (52), is induced by fasting and lipid signals. ANGPTL4 is secreted by many cell types, including macrophages, and plays critical roles in energy homeostasis and lipid metabolism (53). Knockout models exhibit enhanced clearance of triacylglycerol-rich lipoproteins, reduced inflammation, and improved glucose metabolism (54). The association of ANGPTL4 with future obesity outcomes, independently of maternal weight status, highlights the potential of early-life disturbances in metabolism as independent risk factors, separate from maternal influences.

Where stronger associations with BMI at ages 2–3 years occur, such as the FST and HGF proteins, this may be owing to a clinical diagnosis closer to sampling. We hypothesize that attenuation of associations with BMI at later time points (e.g., only one marker reaching FDR significance at age 8) is due to smaller sample sizes and increased variability in BMI reporting. Despite these limitations, our findings suggest that clinically diagnosed obesity provides a more reliable, biologically grounded phenotype with stronger proteomic correlates.

We found that the metabolome strongly associates with delivery mode, preterm birth, and whether the child went to a “newborn ward,” which generally refers in Sweden to a neonatal intensive care unit (NICU), providing specialized care for newborns. Branched-chain amino acids (BCAAs), medium- and long-chain acylcarnitines, nonesterified fatty acids, and triglycerides have been positively associated with birth weight

(55). These are negatively associated with C-peptide, a byproduct of insulin secretion that impacts maternal and fetal health (56, 57). Cholestenone and BCAAs have been previously identified as predictors of rapid growth and overweight/obesity in newborns (58), with lysophosphatidylcholines and glycerophospholipid fatty acids (GPL-FA) related to birth weight (59). Prenatal factors like maternal education, pre-pregnancy BMI, weight gain during pregnancy, tobacco smoke exposure, maternal diet, delivery mode, and birth order influence the composition of cord blood metabolites, consistent with past studies (60–62). In a smaller study, aligned with our findings of the metabolome's strong association with parity, birth order correlated with histidine metabolism and the tricarboxylic acid cycle; sex was associated with BCAA metabolism and vitamin B6 metabolism; whereas mode of delivery was associated with glyoxylate and dicarboxylic acid metabolism (61). Higher intake of margarine or soy oil was associated with E- and Z-octadecanoic acid, hydroxyphenylacetic acid, and uric acid, which are metabolites predictive of allergies (63). One of the most predictable features was whether the mother smoked during pregnancy, which has been linked to alterations in cord blood metabolites from lipid metabolism, such as phosphatidylcholines and sphingomyelins (64). Offspring DNA methylation significantly associates with nicotine and its downstream metabolites in maternal and cord sera (65), with MYO1G, AHRH, and GF11 genes implicated in cardiovascular complications, as are advanced oxidation protein products in the cord blood (62). We previously reported on the adverse consequences of maternal smoking in our study of neurodevelopment (19). These associations offer explanations for the metabolic profiles observed in the cord blood and possible areas for intervention in pregnancy.

Limited studies have explored the predictive value of early, integrated omics for future obesity risk (36). Here, bacterial and proteomic markers measured within the first year of life provided an unprecedented opportunity to model long-term prediction accuracy for clinical obesity diagnoses many years later using a base 40-marker microbiome and proteomic panel. Among the top proteomic predictors were fibroblast growth factor 19 (FGF19), angiopoietin-like 4 (ANGPTL4), sulfotransferase family 2A member 1 (SULT2A1), and interleukin 20 (IL20), proteins that converge on bile acid metabolism, lipid handling, and immune regulation.

FGF19 and SULT2A1 are centrally involved in bile acid metabolism and signaling. FGF19, an enterohepatic hormone primarily produced in the ileum, regulates carbohydrates, glucose homeostasis, and bile acid synthesis (66), and SULT2A1 is significant for the elimination and detoxification of bile acids as well as the sulfation of steroid hormones, altering their bioavailability and activity (67). FGF19 is induced by bile acid signaling. Because bile acid availability and composition are strongly shaped by the gut microbiome (52), altered levels of these proteins at birth may reflect early-life bile acid signatures that contribute to later metabolic risks. ANGPTL4 further supports this axis, as it inhibits intestinal lipase activity (68), suggesting mechanisms of fat accumulation and insulin resistance. Importantly, ANGPTL4 facilitation of bile acid absorption is gut microbiota-dependent (69), which underscores the interaction between ANGPTL4 and the key bacteria identified in our integrated model. ANGPTL4 has been considered for cardiometabolic therapies, studied most recently for its roles in glucose intolerance (70). Bile acid composition and amounts are largely shaped by the gut microbiome (71, 72). Here also, oleic acid levels were depleted at birth in children who were later overweight or obese. Bile acids are inversely associated with abdominal fat, possibly due to downregulation of inflammatory responses (73). The consensus bacterial markers, including *Roseburia* spp., *Enterococcus* spp., *Veillonella tobetsuensis*, *Lachnospiraceae* NK4A136 group, *Blautia glucerasea*, and *Gemella* spp. are notable for their significance to bile acid pool composition, short-chain fatty acid (SCFA) production, and carbohydrate utilization. In addition, B4GALT1, which was significantly higher in future cases, plays key roles in synthesizing glycoproteins and lactose metabolism (74), further supporting altered carbohydrate-related pathways in early life. Meanwhile, IL20, a critical pro-inflammatory cytokine involved in immune modulation (75), mediating immunity

and epithelial tissue health, may reflect early immune tone relevant to metabolic health. It is involved in the dysregulation of macrophages and modulates adipogenesis in *in vitro* models, demonstrating a possible target (76). Consistent with this immune-metabolic interface, the consensus bacterial marker *Enterobacteriaceae* spp. is associated with lipopolysaccharide (LPS) production and low-grade inflammation (77).

Together, these results provide evidence that cord blood proteomic signatures encode biologically meaningful information about future obesity risk. Although precision and recall illustrate the challenges inherent in long-term prediction, the models' AUC and F1 scores present strong signals from birth, demonstrating the reliability of these markers. The association, for instance, of ANGPTL4 being independent of maternal weight, supports the idea that markers reflect intrinsic biological risk rather than simply mirroring maternal adiposity downstream. Consistent with prior literature (5), parental BMI in the child's early life remained an important predictor; however, the integrated model highlights the additional and complementary contribution of early-life biological signatures to long-term obesity risk.

Our findings present the earliest timepoints of microbial dysbiosis to study risk for childhood obesity. The discovery of proteomic and microbiome markers at birth and in infancy, long before the onset of obesity, also underscores the importance of early metabolic programming. While our models show strong discrimination (AUC ~ 0.89), we note that predicted probabilities should not be interpreted as calibrated clinical risk estimates without external validation. These probabilities are not yet calibrated for clinical use. The models are optimized for ranking/classification rather than probability calibration. Future work will assess calibration using methods such as Platt scaling or isotonic regression in external cohorts before clinical deployment. This observational study is intended to generate hypotheses for further testing. Nevertheless, identification of these early-life biomarkers may suggest early interventions to prevent disease.

Limitations

There are several limitations to this investigation. (i) The relatively small numbers of obesity cases in some key omics contrasts should be noted (e.g., 22 overweight/obese vs. 132 normal-weight in the metabolomics comparison; 50 obesity cases vs. 598 controls in microbiome analyses), which may limit statistical power and increase uncertainty in effect estimates. (ii) Obesity diagnosis through early adulthood was observed in 2.4% of subjects ($n = 407$ of 16,683), likely indicating underdiagnosis rather than representing true population prevalence. Since clinical diagnoses of obesity often capture only the most severe cases, these approaches rely on these strict criteria, which may underestimate the broader burden of obesity. However, both registry-based clinical diagnoses and BMI were included to mitigate this potential underestimation. (iii) The omics data were not uniformly available across the cohort at the time of these analyses, such that not all individuals with clinical obesity had metabolomic, proteomic, or microbiome data ready to analyze. Different combinations of omics layers were available for subjects included in this study (e.g., proteomics without metabolomics, microbiome without proteomics), which limited sample sizes for some analyses and may affect the robustness of specific contrasts. As additional data become available, more complete multi-omic coverage will provide greater clarity regarding the stability and reproducibility of these findings. (iv) Furthermore, microbiome findings will require follow-up using shotgun metagenomic approaches to enable functional characterization and identification of microbial pathways and interactions, which will be essential for contextualizing the biological significance of the observed taxonomic differences. (v) Residual confounding is possible in both questionnaire-based and omics analyses, including unmeasured dietary factors, physical activity patterns, and genetic predisposition, which were not accounted for in the present study. In addition, generalizability may be limited beyond the Swedish population, given its specific healthcare, diet, and socioeconomic context, underscoring the importance of replication in other large cohorts. (vi) Finally, the absence of an external validation cohort for the machine learning models represents an important

limitation for clinical translation. While internal cross-validation was applied, external validation will be essential in future studies to confirm robustness and portability.

ACKNOWLEDGMENTS

We acknowledge the fundamental contributions of our ABIS participants and their families. We are grateful to Hanna Karlsson and Austeja Rutkauskaite for help with ABIS data and delivery of samples.

This work was made possible by funding from Barndiabetesfonden (Swedish Child Diabetes Foundation); Swedish Council for Working Life and Social Research; Swedish Research Council; Östgöta Brandstodsbolag; Medical Research Council of Southeast Sweden; JDRF-Wallenberg Foundation; ALF County Council, Östergötland; and Joanna Coccozza Foundation. This work was also funded by the “Inflammation in human early life: targeting impacts on life-course health” (INITIALISE) consortium funded by the Horizon Europe Program of the European Union under Grant Agreement 101094099.

ABIS coordinators, under the supervision of J.L., collected the questionnaire data. A.P.A. integrated the microbiome and proteome data, while A.P.A. and P.A.W. integrated the questionnaire and clinical data, and T.H. integrated the metabolome data. P.A.W. was responsible for calculating IOTF designations, and P.A.W. and J.L. confirmed clinical diagnoses from the Swedish National Registry. A.P.A. and R.D. analyzed the data and prepared the results, with input from all authors. A.P.A. led the microbiome and proteome analyses, while R.D. led the metabolomic analyses. A.P.A. wrote the first draft of the manuscript, incorporating feedback from all authors. E.W.T. oversaw the microbiome sequencing, and T.H. and M.O. performed the metabolomic analysis. A.P.A., J.L., and E.W.T. primarily conceived the investigation, with input from all authors. J.L. oversaw the research process, and J.L., A.P.A., and E.W.T. handled the ethical review process for this investigation, with J.L. responsible for the initial and foundational human subjects’ approvals for ABIS. J.L., M.O., and E.W.T. secured funding and provided financial oversight for the project. All authors had access to and verified the data from the ABIS study. The corresponding author had the final responsibility for the decision to submit the manuscript for publication.

AUTHOR AFFILIATIONS

¹Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, College of Agricultural and Life Sciences, University of Florida, Gainesville, Florida, USA

²School of Science and Technology, Örebro University, Örebro, Sweden

³Crown Princess Victoria Children’s Hospital and Division of Pediatrics, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

⁴School of Medical Sciences, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

⁵Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

⁶Department of Life Technologies, University of Turku, Turku, Finland

AUTHOR ORCID_s

Angelica P. Ahrens  <http://orcid.org/0000-0002-1478-7559>

Eric W. Triplett  <http://orcid.org/0000-0002-1845-4866>

FUNDING

Funder	Grant(s)	Author(s)
HORIZON EUROPE European Institute of Innovation and Technology	101094099	Eric W. Triplett
HORIZON EUROPE European Innovation Council	101094099	Tuulia Hyötyläinen

Funder	Grant(s)	Author(s)
HORIZON EUROPE European Institute of Innovation and Technology	101094099	Johnny Ludvigsson
HORIZON EUROPE European Innovation Council	101094099	Matej Orešič

AUTHOR CONTRIBUTIONS

Angelica P. Ahrens, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Raquel Dias, Formal analysis, Methodology, Writing – review and editing | Tuulia Hyötyläinen, Conceptualization, Data curation, Formal analysis, Methodology, Writing – review and editing | Pär Anderson White, Conceptualization, Methodology, Writing – review and editing | Matej Orešič, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Writing – review and editing | Eric W. Triplett, Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review and editing | Johnny Ludvigsson, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

The processed data underlying the proteomic and microbiome analyses in this study are publicly available in the Dryad repository (<https://doi.org/10.5061/dryad.2v6wwq042>). The Jupyter notebook used to implement the machine learning pipelines is publicly available at <https://github.com/aahren/obesity-ML-ABIS>. Researchers who are interested in accessing additional data or materials underlying this work should contact the senior author, Dr. Ludvigsson. For data not available through open access, bona fide researchers may need to develop a formal collaboration with the ABIS study group for specific scientific purposes.

ETHICS APPROVAL

All participating families provided informed consent. The ABIS study and its connection to the Swedish National Patient Register were approved by the Research Ethics Committee, Linköping University, and Lund University, approval numbers Dnr 03-092; LiU 287.96, LU 83-97, LiU 321-99, and Linköping University Dnr 36,287, Dnr 03-513, Dnr 2018/380-32. The analysis of the gut microbiome was conducted with additional approvals by the University of Florida's Institutional Review Board (IRB) IRB201800903 and IRB202301239.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (mSystems01424-25-s0001.docx). Supplemental text, tables, and figures.

Table S1 (mSystems01424-25-s0002.xlsx). Differential protein expression in cord blood by future BMI-derived obesity status during childhood.

Table S2 (mSystems01424-25-s0003.xlsx). Differential protein expression in cord blood by future clinical diagnosis of obesity.

REFERENCES

1. Phelps NH, Singleton RK, Zhou B, Heap RA, Mishra A, Bennett JE, Paciorek CJ, Lhoste VP, Carrillo-Larco RM, Stevens GA, et al. 2024. Worldwide trends in underweight and obesity from 1990 to 2022: a pooled analysis of 3663 population-representative studies with 22 million children, adolescents, and adults. *Lancet* 403:1027–1050. [https://doi.org/10.1016/S0140-6736\(23\)02750-2](https://doi.org/10.1016/S0140-6736(23)02750-2)
2. Smith JD, Fu E, Kobayashi MA. 2020. Prevention and management of childhood obesity and its psychological and health comorbidities. *Annu*

- Rev Clin Psychol 16:351–378. <https://doi.org/10.1146/annurev-clinpsy-100219-060201>
3. Jebeile H, Kelly AS, O'Malley G, Baur LA. 2022. Obesity in children and adolescents: epidemiology, causes, assessment, and management. *Lancet Diabetes Endocrinol* 10:351–365. [https://doi.org/10.1016/S2213-8587\(22\)00047-X](https://doi.org/10.1016/S2213-8587(22)00047-X)
 4. Lee JS, Jin MH, Lee HJ. 2022. Global relationship between parent and child obesity: a systematic review and meta-analysis. *Clin Exp Pediatr* 65:35–46. <https://doi.org/10.3345/cep.2020.01620>
 5. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. 1997. Predicting obesity in young adulthood from childhood and parental obesity. *N Engl J Med* 337:869–873. <https://doi.org/10.1056/NEJM199709253371301>
 6. Oberbach A, Blüher M, Wirth H, Till H, Kovacs P, Kullnick Y, Schlichting N, Tamm JM, Rolle-Kampczyk U, Murugaiyan J, et al. 2011. Combined proteomic and metabolomic profiling of serum reveals association of the complement system with obesity and identifies novel markers of body fat mass changes. *J Proteome Res* 10:4769–4788. <https://doi.org/10.1021/pr2005555>
 7. Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JL. 2004. The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci USA* 101:15718–15723. <https://doi.org/10.1073/pnas.0407076101>
 8. Dao MC, Everard A, Aron-Wisniewsky J, Sokolovska N, Prifti E, Verger EO, Kayser BD, Levenez F, Chilloux J, Hoyles L, et al. 2016. *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut* 65:426–436. <https://doi.org/10.1136/gutjnl-2014-308778>
 9. Zhang S, Dang Y. 2022. Roles of gut microbiota and metabolites in overweight and obesity of children. *Front Endocrinol* 13:13. <https://doi.org/10.3389/fendo.2022.994930>
 10. Sergeev IN, Aljoutai T, Walton G, Huarte E. 2020. Effects of synbiotic supplement on human gut microbiota, body composition and weight loss in obesity. *Nutrients* 12:222. <https://doi.org/10.3390/nu12010222>
 11. Danielsson P, Putri RR, Marcus C, Hagman E. 2023. Evaluating probiotic efficacy on weight loss in adults with overweight through a double-blind, placebo-controlled randomized trial. *Sci Rep* 13:18200. <https://doi.org/10.1038/s41598-023-45395-7>
 12. Nicolucci AC, Hume MP, Martínez I, Mayengbam S, Walter J, Reimer RA. 2017. Prebiotics reduce body fat and alter intestinal microbiota in children who are overweight or with obesity. *Gastroenterology* 153:711–722. <https://doi.org/10.1053/j.gastro.2017.05.055>
 13. Cox AJ, West NP, Cripps AW. 2015. Obesity, inflammation, and the gut microbiota. *Lancet Diabetes Endocrinol* 3:207–215. [https://doi.org/10.1016/S2213-8587\(14\)70134-2](https://doi.org/10.1016/S2213-8587(14)70134-2)
 14. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. 2019. The human microbiome and child growth - first 1000 days and beyond. *Trends Microbiol* 27:131–147. <https://doi.org/10.1016/j.tim.2018.09.008>
 15. Taddei CR, Neu J. 2023. Editorial: microbiome in the first 1000 days: multi-omic interactions, physiological effects, and clinical implications. *Front Cell Infect Microbiol* 13:1242626. <https://doi.org/10.3389/fcimb.2023.1242626>
 16. Depommier C, Everard A, Druart C, Plovier H, Van Hul M, Vieira-Silva S, Falony G, Raes J, Maiter D, Delzenne NM, et al. 2019. Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. *Nat Med* 25:1096–1103. <https://doi.org/10.1038/s41591-019-0495-2>
 17. Karlsson CLJ, Onnerfält J, Xu J, Molin G, Ahrné S, Thorngren-Jerneck K. 2012. The microbiota of the gut in preschool children with normal and excessive body weight. *Obesity (Silver Spring)* 20:2257–2261. <https://doi.org/10.1038/oby.2012.110>
 18. Stanislawska MA, Dabelea D, Wagner BD, Iszatt N, Dahl C, Sontag MK, Knight R, Lozupone CA, Eggesbø M. 2018. Gut microbiota in the first 2 years of life and the association with body mass index at age 12 in a Norwegian birth cohort. *mBio* 9:e01751-18. <https://doi.org/10.1128/mBio.01751-18>
 19. Ahrens Angelica P, Hyötyläinen T, Petrone JR, Igelström K, George CD, Garrett TJ, Orešič M, Triplett EW, Ludvigsson J. 2024. Infant microbes and metabolites point to childhood neurodevelopmental disorders. *Cell* 187:1853–1873. <https://doi.org/10.1016/j.cell.2024.02.035>
 20. Östensson M, Björkqvist O, Guo A, Stördal K, Halfvarson J, Mårdil K, Ludvigsson J. 2023. Epidemiology, validation, and clinical characteristics of inflammatory bowel disease: the ABIS birth cohort study. *BMC Gastroenterol* 23:199. <https://doi.org/10.1186/s12876-023-02840-1>
 21. Ahrens A.P, Dias R, Hyötyläinen T, Orešič M, Triplett EW, Ludvigsson J. 2026. The inflammatory path toward type 1 diabetes begins during pregnancy. *Nat Commun* 17:979. <https://doi.org/10.1038/s41467-025-67712-6>
 22. Ludvigsson Jonas F, Almqvist C, Bonamy A-KE, Ljung R, Michaëlsson K, Neovius M, Stephansson O, Ye W. 2016. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 31:125–136. <https://doi.org/10.1007/s10654-016-0117-y>
 23. Ludvigsson J.F, Otterblad-Olausson P, Pettersson BU, Ekblom A. 2009. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 24:659–667. <https://doi.org/10.1007/s10654-009-9350-y>
 24. Cole TJ, Lobstein T. 2012. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatr Obes* 7:284–294. <https://doi.org/10.1111/j.2047-6310.2012.00064.x>
 25. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
 26. Karlén J, Ludvigsson J, Hedmark M, Faresjö Å, Theodorsson E, Faresjö T. 2015. Early psychosocial exposures, hair cortisol levels, and disease risk. *Pediatrics* 135:e1450–7. <https://doi.org/10.1542/peds.2014-2561>
 27. Almqvist-Tangen G, Dahlgren J, Roswall J, Bergman S, Alm B. 2013. Milk cereal drink increases BMI risk at 12 and 18 months, but formula does not. *Acta Paediatr* 102:1174–1179. <https://doi.org/10.1111/apa.12418>
 28. Kalliomäki M, Collado MC, Salminen S, Isolauri E. 2008. Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr* 87:534–538. <https://doi.org/10.1093/ajcn/87.3.534>
 29. Vael C, Verhulst SL, Nelen V, Goossens H, Desager KN. 2011. Intestinal microflora and body mass index during the first three years of life: an observational study. *Gut Pathog* 3:8. <https://doi.org/10.1186/1757-4749-3-8>
 30. Luoto R, Kalliomäki M, Laitinen K, Isolauri E. 2010. The impact of perinatal probiotic intervention on the development of overweight and obesity: follow-up study from birth to 10 years. *Int J Obes* 34:1531–1537. <https://doi.org/10.1038/ijo.2010.50>
 31. Korpela K, Zijlmans MAC, Kuitunen M, Kukkonen K, Savilahti E, Salonen A, de Weerth C, de Vos WM. 2017. Childhood BMI in relation to microbiota in infancy and lifetime antibiotic use. *Microbiome* 5:26. <https://doi.org/10.1186/s40168-017-0245-y>
 32. Forbes JD, Azad MB, Vehling L, Tun HM, Konya TB, Guttman DS, Field CJ, Lefebvre D, Sears MR, Becker AB, et al. 2018. Association of exposure to formula in the hospital and subsequent infant feeding practices with gut microbiota and risk of overweight in the first year of life. *JAMA Pediatr* 172:e181161. <https://doi.org/10.1001/jamapediatrics.2018.1161>
 33. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, D Lieber A, Wu F, Perez-Perez GI, Chen Y, et al. 2016. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* 8:343ra82. <https://doi.org/10.1126/scitranslmed.aad7121>
 34. Korpela Katri, Salonen A, Virta LJ, Kekkonen RA, Forslund K, Bork P, de Vos WM. 2016. Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat Commun* 7:10410. <https://doi.org/10.1038/ncomms10410>
 35. Korpela Katri, Salonen A, Saxen H, Nikkonen A, Peltola V, Jaakkola T, de Vos W, Kolho K-L. 2020. Antibiotics in early life associate with specific gut microbiota signatures in a prospective longitudinal infant cohort. *Pediatr Res* 88:438–443. <https://doi.org/10.1038/s41390-020-0761-5>
 36. Tun HM, Bridgman SL, Chari R, Field CJ, Guttman DS, Becker AB, Mandhane PJ, Turvey SE, Subbarao P, Sears MR, et al. 2018. Roles of birth mode and infant gut microbiota in intergenerational transmission of overweight and obesity from mother to offspring. *JAMA Pediatr* 172:368–377. <https://doi.org/10.1001/jamapediatrics.2017.5535>
 37. Rodrigues VF, Elias-Oliveira J, Pereira IS, Pereira JA, Barbosa SC, Machado MSG, Carlos D. 2022. *Akkermansia muciniphila* and gut immune system: a good friendship that attenuates inflammatory bowel disease, obesity, and diabetes. *Front Immunol* 13:934695. <https://doi.org/10.3389/fimmu.2022.934695>
 38. Greer RL, Dong X, Moraes ACF, Zielke RA, Fernandes GR, Peremyslova E, Vasquez-Perez S, Schoenborn AA, Gomes EP, Pereira AC, et al. 2016. *Akkermansia muciniphila* mediates negative effects of IFN γ on glucose metabolism. *Nat Commun* 7:13329. <https://doi.org/10.1038/ncomms13329>
 39. Zhai R, Xue X, Zhang L, Yang X, Zhao L, Zhang C. 2019. Strain-specific anti-inflammatory properties of two *Akkermansia muciniphila* strains on

- chronic colitis in mice. *Front Cell Infect Microbiol* 9:239. <https://doi.org/10.3389/fcimb.2019.00239>
40. Zhang L, Qin Q, Liu M, Zhang X, He F, Wang G. 2018. *Akkermansia muciniphila* can reduce the damage of gluco/lipotoxicity, oxidative stress and inflammation, and normalize intestine microbiota in streptozotocin-induced diabetic rats. *Pathog Dis* 76:76. <https://doi.org/10.1093/fempd/fty028>
 41. Xu Y, Wang N, Tan H-Y, Li S, Zhang C, Feng Y. 2020. Function of *Akkermansia muciniphila* in obesity: interactions with lipid metabolism, immune response and gut systems. *Front Microbiol* 11:11. <https://doi.org/10.3389/fmicb.2020.00219>
 42. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, Guiot Y, Derrien M, Muccioli GG, Delzenne NM, et al. 2013. Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci USA* 110:9066–9071. <https://doi.org/10.1073/pnas.1219451110>
 43. Dziarski R, Park SY, Kashyap DR, Dowd SE, Gupta D. 2016. Pglyrp-regulated gut microflora *Prevotella falsenii*, *Parabacteroides distasonis* and *Bacteroides eggerthii* enhance and *alisters finegoldii* attenuates colitis in mice. *PLoS One* 11:e0146162. <https://doi.org/10.1371/journal.pone.0146162>
 44. Kmezik C, Krska D, Mazurkewich S, Larsbrink J. 2021. Characterization of a novel multidomain CE15-GH8 enzyme encoded by a polysaccharide utilization locus in the human gut bacterium *Bacteroides eggerthii*. *Sci Rep* 11:17662. <https://doi.org/10.1038/s41598-021-96659-z>
 45. Wu Y-Y, Gou W, Yan Y, Liu C-Y, Yang Y, Chen D, Xie K, Jiang Z, Fu Y, Zhu H-L, et al. 2022. Gut microbiota and acylcarnitine metabolites connect the beneficial association between equol and adiposity in adults: a prospective cohort study. *Am J Clin Nutr* 116:1831–1841. <https://doi.org/10.1093/ajcn/nqac252>
 46. Zhang H, Liang S, Yin K, Mo Y, Li Y, Lv Y, Zhan H, Zhang Z, Shan Z, Guo Z, et al. 2024. Urinary equol and equol-predicting microbial genera are favorably associated with body fat measures among chinese adults. *J Nutr* 154:2843–2851. <https://doi.org/10.1016/j.tjnut.2024.07.024>
 47. Garcia-Argibay M, Lundström S, Cortese S, Larsson H. 2024. Trends in body mass index among individuals with neurodevelopmental disorders. *JAMA Netw Open* 7:e2431543. <https://doi.org/10.1001/jamanetworkopen.2024.31543>
 48. Wu C, Borné Y, Gao R, López Rodríguez M, Roell WC, Wilson JM, Regmi A, Luan C, Aly DM, Peter A, et al. 2021. Elevated circulating follistatin associates with an increased risk of type 2 diabetes. *Nat Commun* 12:6486. <https://doi.org/10.1038/s41467-021-26536-w>
 49. Oliveira AG, Araújo TG, Carvalho B de M, Rocha GZ, Santos A, Saad MJA. 2018. The role of hepatocyte growth factor (HGF) in insulin resistance and diabetes. *Front Endocrinol* 9:503. <https://doi.org/10.3389/fendo.2018.00503>
 50. Tsukagawa E, Adachi H, Hirai Y, Enomoto M, Fukami A, Ogata K, Kasahara A, Yokoi K, Imaizumi T. 2013. Independent association of elevated serum hepatocyte growth factor levels with development of insulin resistance in a 10-year prospective study. *Clin Endocrinol (Oxf)* 79:43–48. <https://doi.org/10.1111/j.1365-2265.2012.04496.x>
 51. Rehman J, Considine RV, Bovenkerk JE, Li J, Slavens CA, Jones RM, March KL. 2003. Obesity is associated with increased levels of circulating hepatocyte growth factor. *J Am Coll Cardiol* 41:1408–1413. [https://doi.org/10.1016/s0735-1097\(03\)00231-6](https://doi.org/10.1016/s0735-1097(03)00231-6)
 52. Collins SL, Stine JG, Bisanz JE, Okafor CD, Patterson AD. 2023. Bile acids and the gut microbiota: metabolic interactions and impacts on disease. *Nat Rev Microbiol* 21:236–247. <https://doi.org/10.1038/s41579-022-00805-x>
 53. Kersten S. 2021. Role and mechanism of the action of angiopoietin-like protein ANGPTL4 in plasma lipid metabolism. *J Lipid Res* 62:100150. <https://doi.org/10.1016/j.jlr.2021.100150>
 54. Aryal B, Singh AK, Zhang X, Varela L, Rotllan N, Goedeke L, Chaube B, Camporez J-P, Vatner DF, Horvath TL, et al. 2018. Absence of ANGPTL4 in adipose tissue improves glucose tolerance and attenuates atherogenesis. *JCI Insight* 3:3. <https://doi.org/10.1172/jci.insight.97918>
 55. Kadakia R, Talbot O, Kuang A, Bain JR, Muehlbauer MJ, Stevens RD, Ilkayeva OR, Lowe LP, Metzger BE, Newgard CB, et al. 2019. Cord blood metabolomics: association with newborn anthropometrics and C-peptide across ancestries. *J Clin Endocrinol Metab* 104:4459–4472. <https://doi.org/10.1210/jc.2019-00238>
 56. Meek CL, Oram RA, McDonald TJ, Feig DS, Hattersley AT, Murphy HR. 2021. Reappearance of C-peptide during the third trimester of pregnancy in type 1 diabetes: pancreatic regeneration or fetal hyperinsulinism? *Diabetes Care* 44:1826–1834. <https://doi.org/10.2337/dci-21-0028>
 57. Niknam A, Ramezani Tehrani F, Behboudi-Gandevani S, Rahmati M, Hedayati M, Abedini M, Firouzi F, Torkestani F, Zokaee M, Azizi F. 2022. Umbilical cord blood concentration of connecting peptide (C-peptide) and pregnancy outcomes. *BMC Pregnancy Childbirth* 22:764. <https://doi.org/10.1186/s12884-022-05081-4>
 58. Handakas E, Keski-Rahkonen P, Chatzi L, Alfano R, Roumeliotaki T, Plusquin M, Maitre L, Richiardi L, Brescianini S, Scalbert A, et al. 2021. Cord blood metabolic signatures predictive of childhood overweight and rapid growth. *Int J Obes* 45:2252–2260. <https://doi.org/10.1038/s41366-021-00888-1>
 59. Hellmuth C, Uhl O, Standl M, Demmelair H, Heinrich J, Koletzko B, Thiering E. 2017. Cord blood metabolome is highly associated with birth weight, but less predictive for later weight development. *Obes Facts* 10:85–100. <https://doi.org/10.1159/000453001>
 60. Alfano R, Plusquin M, Robinson O, Brescianini S, Chatzi L, Keski-Rahkonen P, Handakas E, Maitre L, Nawrot T, Robinot N, et al. 2022. Cord blood metabolites and rapid postnatal growth as multiple mediators in the prenatal propensity to childhood overweight. *Int J Obes* 46:1384–1393. <https://doi.org/10.1038/s41366-022-01108-0>
 61. Ross AB, Barman M, Hartvigsson O, Lundell A-C, Savolainen O, Hesselmar B, Wold AE, Sandberg A-S. 2021. Umbilical cord blood metabolome differs in relation to delivery mode, birth order and sex, maternal diet and possibly future allergy development in rural children. *PLoS One* 16:e0242978. <https://doi.org/10.1371/journal.pone.0242978>
 62. Rua E de AO, Porto ML, Ramos JPL, Nogueira BV, Meyrelles SS, Vasquez EC, Pereira TC. 2014. Effects of tobacco smoking during pregnancy on oxidative stress in the umbilical cord and mononuclear blood cells of neonates. *J Biomed Sci* 21:105. <https://doi.org/10.1186/s12929-014-0105-z>
 63. Lundell A-C, Hesselmar B, Nordström I, Adlerberth I, Wold AE, Rudin A. 2015. Higher B-cell activating factor levels at birth are positively associated with maternal dairy farm exposure and negatively related to allergy development. *J Allergy Clin Immunol* 136:1074–1082. <https://doi.org/10.1016/j.jaci.2015.03.022>
 64. Rolle-Kampczyk UE, Krumsiek J, Otto W, Röder SW, Kohajda T, Borte M, Theis F, Lehmann I, von Bergen M. 2016. Metabolomics reveals effects of maternal smoking on endogenous metabolites from lipid metabolism in cord blood of newborns. *Metabolomics (Los Angeles)* 12:76. <https://doi.org/10.1007/s11306-016-0983-z>
 65. Kheirkhah Rahimabad P, Anthony TM, Jones AD, Eslamimehr S, Mukherjee N, Ewart S, Holloway JW, Arshad H, Commodore S, Karmaus W. 2020. Nicotine and its downstream metabolites in maternal and cord sera: biomarkers of prenatal smoking exposure associated with offspring DNA methylation. *Int J Environ Res Public Health* 17:9552. <https://doi.org/10.3390/ijerph17249552>
 66. Wu A-L, Coulter S, Liddle C, Wong A, Eastham-Anderson J, French DM, Peterson AS, Sonoda J. 2011. FGFR4 regulates cell proliferation, glucose and bile acid metabolism via FGFR4-dependent and independent pathways. *PLoS One* 6:e17868. <https://doi.org/10.1371/journal.pone.0017868>
 67. Alnouti Y. 2009. Bile acid sulfation: a pathway of bile acid elimination and detoxification. *Toxicol Sci* 108:225–246. <https://doi.org/10.1093/toxic/ci/kfn268>
 68. Mattijssen F, Alex S, Swarts HJ, Groen AK, van Schothorst EM, Kersten S. 2014. Angptl4 serves as an endogenous inhibitor of intestinal lipid digestion. *Mol Metab* 3:135–144. <https://doi.org/10.1016/j.molmet.2013.11.004>
 69. Janssen AWF, Dijk W, Boekhorst J, Kuipers F, Groen AK, Lukovac S, Hooiveld GJEJ, Kersten S. 2017. ANGPTL4 promotes bile acid absorption during taurocholic acid supplementation via a mechanism dependent on the gut microbiota. *Biochim Biophys Acta Mol Cell Biol Lipids* 1862:1056–1067. <https://doi.org/10.1016/j.bbalip.2017.07.005>
 70. Morris A. 2018. ANGPTL4 — the link binding obesity and glucose intolerance. *Nat Rev Endocrinol* 14:251–251. <https://doi.org/10.1038/nrendo.2018.35>
 71. Ridlon JM, Kang DJ, Hylemon PB, Bajaj JS. 2014. Bile acids and the gut microbiome. *Curr Opin Gastroenterol* 30:332–338. <https://doi.org/10.1097/MOG.0000000000000057>
 72. Wahlström A, Sayin SI, Marschall H-U, Bäckhed F. 2016. Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metab* 24:41–50. <https://doi.org/10.1016/j.cmet.2016.05.005>

73. Tutunchi H, Ostadrahimi A, Saghafi-Asl M. 2020. The effects of diets enriched in monounsaturated oleic acid on the management and prevention of obesity: a systematic review of human intervention studies. *Adv Nutr* 11:864–877. <https://doi.org/10.1093/advances/nmaa013>
74. Montasser ME, Van Hout CV, Miloscio L, Howard AD, Rosenberg A, Callaway M, Shen B, Li N, Locke AE, Verweij N, et al. 2021. Genetic and functional evidence links a missense variant in *B4GALT1* to lower LDL and fibrinogen. *Science* 374:1221–1227. <https://doi.org/10.1126/science.abe0348>
75. Rutz S, Wang X, Ouyang W. 2014. The IL-20 subfamily of cytokines—from host defence to tissue homeostasis. *Nat Rev Immunol* 14:783–795. <https://doi.org/10.1038/nri3766>
76. Hsu Y-H, Wu C-H, Chiu C-J, Chen W-T, Chang Y-C, Wabitsch M, Chang M-S. 2021. IL-20 is involved in obesity by modulation of adipogenesis and macrophage dysregulation. *Immunology* 164:817–833. <https://doi.org/10.1111/imm.13403>
77. Boopathi S, Priya PS, Haridevamuthu B, Nayak SPRR, Chandrasekar M, Arockiaraj J, Jia A-Q. 2023. Expanding germ-organ theory: understanding non-communicable diseases through enterobacterial translocation. *Pharmacol Res* 194:106856. <https://doi.org/10.1016/j.phrs.2023.106856>