Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Maternal-infant multi-omics identifies biological features associated with growth faltering in small vulnerable newborns

Trenton Dailey-Chwalibóg

trenton@dailey-chwalibog.com

Ghent University https://orcid.org/0000-0002-8204-4925

Lishi Deng

Ghent University

Laeticia Toe

Ghent University

Kelsey Fehr

University of Manitoba

Lars Bode

University of California, San Diego https://orcid.org/0000-0002-9914-9571

Daniela Hampel

University of California

Anderson Compaoré

Agence de Formation de Recherche et d'Expertise en Santé pour l'Afrique

Brian DeFelice

Biohub – San Francisco

Bruno De Meulenaer

Ghent University

Joshua Elias

Biohub – San Francisco

Carlos Gonzalez

Biohub – San Francisco

Emma Guiberson

Stanford University School of Medicine

Andrew Mertens

University of California, Berkeley

Annie Moradian

Precision Biomarker Laboratories, Cedars-Sinai Medical Center https://orcid.org/0000-0002-0407-2031

Melissa Manus
University of Manitoba
Patrick McAlpine
University of Oviedo
Matthew Olm
Stanford University School of Medicine
Cheick Ouattara
Agence de Formation de Recherche et d'Expertise en Santé pour l'Afrique
Lionel Ouédraogo
Agence de Formation de Recherche et d'Expertise en Santé pour l'Afrique
Moctar Ouédraogo
Agence de Formation de Recherche et d'Expertise en Santé pour l'Afrique
Anita Reddy
Stanford university
Moulaye Sanou
Unité Nutrition et Maladies Métaboliques, Institut de Recherche en Sciences de la Santé
Sailendharan Sudakaran
Stanford University School of Medicine
Niveda Sundararaman
Cedars-Sinai Medical Center
Eileen Wang
Biohub – San Francisco
Yifan Gao
Stanford University School of Medicine
KIM LAGERBORG
Sapient Bioanalytics
Jennifer Van Eyk
Cedars-Sinai Medical Center
Lindsay Allen
University of California, Davis
Carl Lachat
Ghent University
Justin Sonnenburg
Stanford university
Meghan Azad
University of Manitoba https://orcid.org/0000-0002-5942-4444

Article

Keywords:

Posted Date: July 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10243829/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Maternal-infant multi-omics identifies biological features associated with growth faltering in small vulnerable newborns

Lishi Deng,[1] Laeticia Celine Toe,[1,2] Kelsey Fehr,[3,4] Lars Bode,[5] Daniela Hampel,[6] Anderson Compaoré,[7] Brian C. DeFelice,[8] Bruno De Meulenaer,[1] Joshua E. Elias,[8] Carlos G. Gonzalez,[8] Emma R. Guiberson,[9] Andrew Mertens,[10] Annie Moradian,[11] Melissa B. Manus,[3,4,12] Patrick L. McAlpine,[13] Matthew R. Olm,[9] Cheick Ahmed Ouattara,[2,7] Lionel Oliver Ouédraogo,[7,14] Moctar Ouédraogo,[7] Anita Reddy,[9] Moulaye Sanou,[2] Sailendharan Sudakaran,[9] Niveda Sundararaman,[11] Eileen Wang,[8] Yifan Gao,[9] Kim Lagerborg,[15] Jennifer E. Van Eyk,[16] Lindsay H. Allen,[6] Carl Lachat,[1] Justin L. Sonnenburg,[8,9,17] Meghan B. Azad,[3,4] Trenton Dailey-Chwalibóg,[1,7*]

1 Department of Food Technology, Safety and Health, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium
2 Unité Nutrition et Maladies Métaboliques, Institut de Recherche en Sciences de la Santé (IRSS), Bobo-Dioulasso, Burkina Faso
3 Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Canada
4 Manitoba Interdisciplinary Lactation Centre (MILC), Children's Hospital Research Institute of Manitoba, Winnipeg, Canada
5 Department of Pediatrics and Larsson-Rosenquist Foundation Mother-Milk-Infant Center of Research Excellence (MOMI CORE), and Human Milk Institute (HMI), University of California San Diego, La Jolla, CA, USA
6 Institute for Global Nutrition, Department of Nutrition, University of California, Davis, CA, USA
7 Agence de Formation de Recherche et d'Expertise en Santé pour l'Afrique (AFRICSanté), Bobo-Dioulasso, Burkina Faso
8 Biohub – San Francisco, San Francisco, CA, USA
9 Department of Microbiology & Immunology, Stanford University School of Medicine, Stanford, CA
10 School of Public Health, University of California, Berkeley, CA, USA
11 Proteomics & Metabolomics Core, Precision Biomarker Laboratories, Cedars-Sinai Medical Center, Los Angeles, CA United States
12 Department of Anthropology, University of Texas at San Antonio, San Antonio, TX, USA
13 BIONUC Research Group, Department of Functional Biology, University of Oviedo, Oviedo, Spain
14 Centre Muraz, Bobo-Dioulasso, Burkina Faso
15 Sapient Bioanalytics, LLC, San Diego, CA, United States
16 Smidt Heart Institute, Advanced Clinical Biosystems Research Institute and Barbra Streisand Women's Heart Center, Cedars-Sinai Medical Center, Los Angeles, CA, United States
17 Center for Human Microbiome Studies, Stanford University School of Medicine, Stanford, CA, USA

*Corresponding author: Trenton Dailey-Chwalibóg [Trenton@Dailey-Chwalibog.com]

Abstract

Small vulnerable newborns (SVNs) are at high risk of growth faltering, but the underlying biology remains poorly understood. We performed longitudinal multi-omics profiling in 294 mother-infant dyads from rural Burkina Faso to identify maternal, milk and infant biological features associated with SVN birth and infant growth trajectories. SVN birth was associated with maternal nutritional and household vulnerability together with coordinated prenatal host-microbial, metabolic, and inflammatory perturbations. Postnatally, SVNs exhibited selective remodeling of milk, delayed maturation of infant gut ecosystem, and altered stool metabolite profiles. The strongest biological associations with subsequent growth were concentrated within the first two months of life, when milk-centered signatures were associated with persistent linear growth, transient ponderal catch-up and subsequent recovery, whereas later growth increasingly reflected broader host-microbial biology. Milk glutamine and serine partially mediated the association between SVN status and persistent linear growth deficits, explaining 4-10% of the observed effect.

Introduction

Small vulnerable newborns (SVNs), encompassing infants born preterm, small for gestational age (SGA), or with low birth weight (LBW), account for a substantial share of neonatal morbidity and mortality worldwide[1]. It has been estimated that as many as 80% of all neonatal deaths in the world occur in LBW infants, of whom two-thirds are probably preterm and one-third SGA[2]. Beyond birth, these infants remain at increased risk of postnatal growth faltering, undernutrition, and adverse long-term health outcomes[3–5]. The SVN concept has therefore been proposed as a clinically useful approach for capturing a broader population of vulnerable newborns[1], particularly in low-resource settings where multiple maternal and environmental risk factors co-occur[6].

The pathways leading to a SVN are heterogeneous because the three phenotypes arise through overlapping but distinct biological processes. Preterm birth is often associated with inflammatory, infectious, and maternal-fetal metabolic disturbance[7–9], whereas SGA and LBW more commonly reflect impaired fetal growth in the context of maternal undernutrition, environmental enteric dysfunction[10], or broader maternal and environmental adversity[11–14]. Recent application of the SVN framework has further highlighted the contribution of maternal nutritional status, parity, and household conditions to SVN risk[1,15], underscoring that these newborn phenotypes emerge from interacting maternal, fetal and contextual determinants.

Although the determinants of SVN birth originate before delivery, emerging evidence suggests that biological differences persist across the postnatal maternal-infant interface. Maternal physiology and metabolism continue to shape the postnatal environment through lactation and other maternal-infant interactions, while human milk composition varies by infant birth phenotype and maternal context; in higher-risk births, particularly following preterm birth, milk often contains higher protein content and altered concentrations of bioactive factors during early lactation[16,17]; carbohydrate content may also differ in growth-restricted infants[18]. In parallel, infants born preterm or with LBW exhibit differences in early microbial development compared with lower-risk infants, including altered gut microbiome composition and function during the first months of life[19–22]. Together, these findings suggest that the biological divergence associated with SVN birth persists beyond delivery, spanning maternal physiology, milk composition, and infant microbial development.

Whether these early biological differences contribute to postnatal growth trajectories in SVNs remains incompletely understood. Although SVN phenotypes are consistently associated with elevated risk of postnatal growth faltering, postnatal growth patterns are heterogeneous, with evidence of partial catch-up in some infants but persistent deficits in others[4,23]. Human milk components, including human milk oligosaccharides (HMOs), lipids, micronutrients, and other bioactive factors, have been associated with

infant growth and with broader metabolic and microbial development[24–28], raising the possibility that postnatal biological exposures may influence growth recovery or continued vulnerability. However, most existing studies have examined these factors in isolation, typically within a single biological compartment or at a single timepoint, limiting understanding of how coordinated variation across maternal, milk and infant systems relate to early growth in SVN dyads.

To investigate the biological features associated with SVN birth and subsequent growth trajectories, we performed longitudinal multi-omics profiling across mothers, milk and infants in a cohort of mother-infant dyads in rural Burkina Faso nested within the MISAME-III randomized controlled trial, which evaluated the efficacy of balanced energy-protein (BEP) supplementation during pregnancy and lactation on birth outcomes and infant growth[29–32]. We specifically aimed to identify maternal prenatal signals associated with SVN birth, characterize postnatal biological differences across the maternal-milk-infant interface, and determine how early-life biological variation relates to infant growth faltering and recovery.

Results

Study cohort, sampling scheme, and baseline characteristics

The analytical cohort comprised 294 mother-infant pairs, including 85 SVN cases and 209 non-SVN controls. Longitudinal multi-omics profiling included maternal and infant metabolomics from volumetric absorptive microsampling (VAMS) dried blood samples, maternal plasma metabolomics and proteomics, maternal and infant stool metagenomics, metabolomics, proteomics, and short-chain fatty acid (SCFA), together with comprehensive milk profiling, including macro- and micronutrients, HMOs, microbes, bioactive proteins, proteomics, and targeted and untargeted metabolites, collected across multiple time points from pregnancy through the first 6 months of life (Figure 1A). The cohort comprised four MISAME-III intervention arms (prenatal BEP, postnatal BEP, combined prenatal and postnatal BEP, and control), which were similarly distributed between SVN and non-SVN groups ($p = 0.79$, Table S1). Likewise, prenatal BEP exposure was not associated with SVN incidence ($p = 0.72$, chi-squared test). Accordingly, BEP intervention was not included as a routine covariate, although sensitivity analyses adjusting for intervention yielded similar results unless otherwise specified.

Baseline maternal, infant, and household characteristics are summarized in Table S1 and Figure 1B, C. Compared with controls, mothers of SVN infants had poorer nutritional status, reflected by lower BMI (20.96 ± 2.47 vs. 22.37 ± 3.31 kg/m^2, $p < 0.001$) and mid-upper arm circumference (MUAC, 253.21 ± 23.76 vs. 264.18 ± 28.21 mm, $p = 0.002$), and were more often primigravid (36.5% vs. 20.6%, $p = 0.015$). Households of SVNs also had less frequent access to improved water sources (protected well, borehole, pipe, or bottled water) than controls (50.6% vs. 63.6%, $p = 0.049$), and tended to experience greater food insecurity (68.2% vs. 57.4%, $p = 0.089$). SVN cases were also born at lower gestational age (39.19 ± 2.16 vs. 40.07 ± 1.11 weeks, $p < 0.001$) and exhibited marked lower birth weight (2.49 ± 0.31 vs. 3.22 ± 0.31 kg, $p < 0.001$), length (46.59 ± 1.88 vs. 49.30 ± 1.69 cm, $p < 0.001$), head circumference (32.06 ± 1.39 vs. 33.84 ± 1.24 cm, $p < 0.001$) and MUAC (91.74 ± 7.07 vs. 103.66 ± 6.77 mm, $p < 0.001$). Nearly all infants were exclusively breastfed until 6 months, with no difference between groups.

These baseline differences were followed by persistent postnatal growth deficits (Figure 1E). Across the first year of life, SVNs remained smaller than infants in the control group for weight, length, MUAC, and head circumference. The most sustained separation was observed for length-for-age z-score (LAZ), which remained consistently lower in SVNs than controls across follow-up (birth: -1.56 vs. -0.11; 2 months: -1.61 vs. -0.48; 6 months: -1.25 vs. -0.26; 11 months: -1.25 vs. -0.45; all $p < 0.001$). Weight-for-age z-score (WAZ) showed a similar persistent disadvantage (birth: -1.86 vs. -0.17; 11 months: -1.57 vs. -0.69). In contrast, weight-for-length z-score (WLZ) was lower in SVNs at birth (-1.07

vs. -0.04, $p < 0.001$), showed transient early catch-up around 2 months, and then declined again, with lower WLZ in SVNs reappearing by 6 months (-0.41 vs. -0.04, $p = 0.031$) and persisting thereafter.

Among the 85 SVN cases, 74 were SGA, 40 had LBW, and 12 were preterm, with partial overlap between phenotypes (Figure 1 D). Despite this overlap, the sub-phenotypes showed distinct baseline characteristics and growth trajectories (Table S1). Mothers of LBW and preterm infants tended to have lower baseline anthropometric status compared to mothers of SGA infants, and primigravidity was most common among LBW mothers. At birth, preterm infants showed the most severe anthropometric deficits and the shortest gestation, whereas LBW infants were also substantially smaller than SGA infants, particularly for weight and length. Longitudinally, all three sub-phenotypes showed persistent growth deficits relative to controls across infancy, although preterm infants displayed the most pronounced early deficits and the clearest time-varying changes in weight and length (Figure 1E, Figure S1). Notably, all three sub-phenotypes exhibited an early WLZ improvement followed by later decline, with the transient catch-up pattern appearing most marked among preterm infants.



Figure 1. Study design, baseline characteristics, and longitudinal growth trajectories of SVNs.

(A) Longitudinal sampling framework showing feature numbers by data layer (left) and sample availability across time points (right).

(B-C) Baseline continuous (B) and binary (C) maternal and infant characteristics in SVN cases and controls.

(D) Overlap of SVN sub-phenotypes.

(E-F) Longitudinal growth trajectories during the first year of life in SVN cases and controls (D) and in sub-phenotypes and controls (E). Mean ± SEM shown. P-values indicate group-by-time interactions; asterisks denote between-group differences at individual timepoints.

*** FDR < 0.001, ** FDR < 0.01, * FDR < 0.05, º FDR < 0.1 (Benjamini-Hochberg adjusted).

Prenatal nutritional, inflammatory and microbial signatures of SVN birth

To determine whether prenatal biological differences were associated with subsequent SVN birth, we integrated maternal baseline characteristics with third-trimester multi-omics using DIABLO. Two latent components discriminated pregnancies resulting in SVN infants from controls with good performance (Component 1: AUC = 0.806, p = 0.018; Figure 2; Component 2: AUC = 0.894, p = 0.0018; Figure S2, Supplementary Data 1).

Component 1 was characterized by coordinated microbial metabolic and host inflammatory signatures together with household vulnerability (Figure 2). Among the strongest contributors, fecal isocaproate, two *Prevotella* species (sp002251385 and sp900553155) and the plasma protein vitronectin (P04004), a complement- and coagulation-associated glycoprotein, were higher in mothers who delivered SVNs, whereas nucleophosmin (P06748-2), a multifunctional stress-response nucleolar protein, and access to improved water sources were lower in cases.

Component 2 captured a complementary prenatal axis linking SVN birth with microbial metabolic and proteomic signatures (Figure S2). Household food insecurity emerged as the strongest contributor and was higher in pregnancies resulting in SVNs, together with lower maternal BMI and reduced access to improved water sources. Mothers who delivered SVNs also had higher fecal valerate, the fecal metabolite 1,5-Isoquinolinediol, and plasma proteins PTGDS (P41222, prostaglandin D2 synthase) and complement factor B (P00751), whereas transthyretin (P02766, a thyroid hormone and retinol transport protein), phospholipase B1 (Q6P1J6), and CNDP1 (Q96KN2, a beta-alanyl-L-histidine dipeptidase) were lower.

Pathway-level analyses further showed that these prenatal differences converged on three major biological processes (Figure 2D). Maternal fecal metagenomic pathways that were enriched in SVNs were dominated by microbial structural and biosynthetic functions, including flagellar assembly, lipopolysaccharide (LPS) biosynthesis, O-antigen repeat unit biosynthesis, and biosynthesis of nucleotide sugars, amino acids, and cofactors. Maternal metabolomic differences reflected coordinated remodeling of carbohydrate, lipid, and amino acid metabolism, with glycolytic, galactose, glycerophospholipid, and lysine metabolism enriched in SVNs, whereas fatty acid-related pathways, including arachidonic acid metabolism, were relatively depleted. Prenatal proteomic pathways that were higher in SVNs were dominated by complement and coagulation cascades, acute-phase response, complement activation, and plasma lipoprotein remodeling.

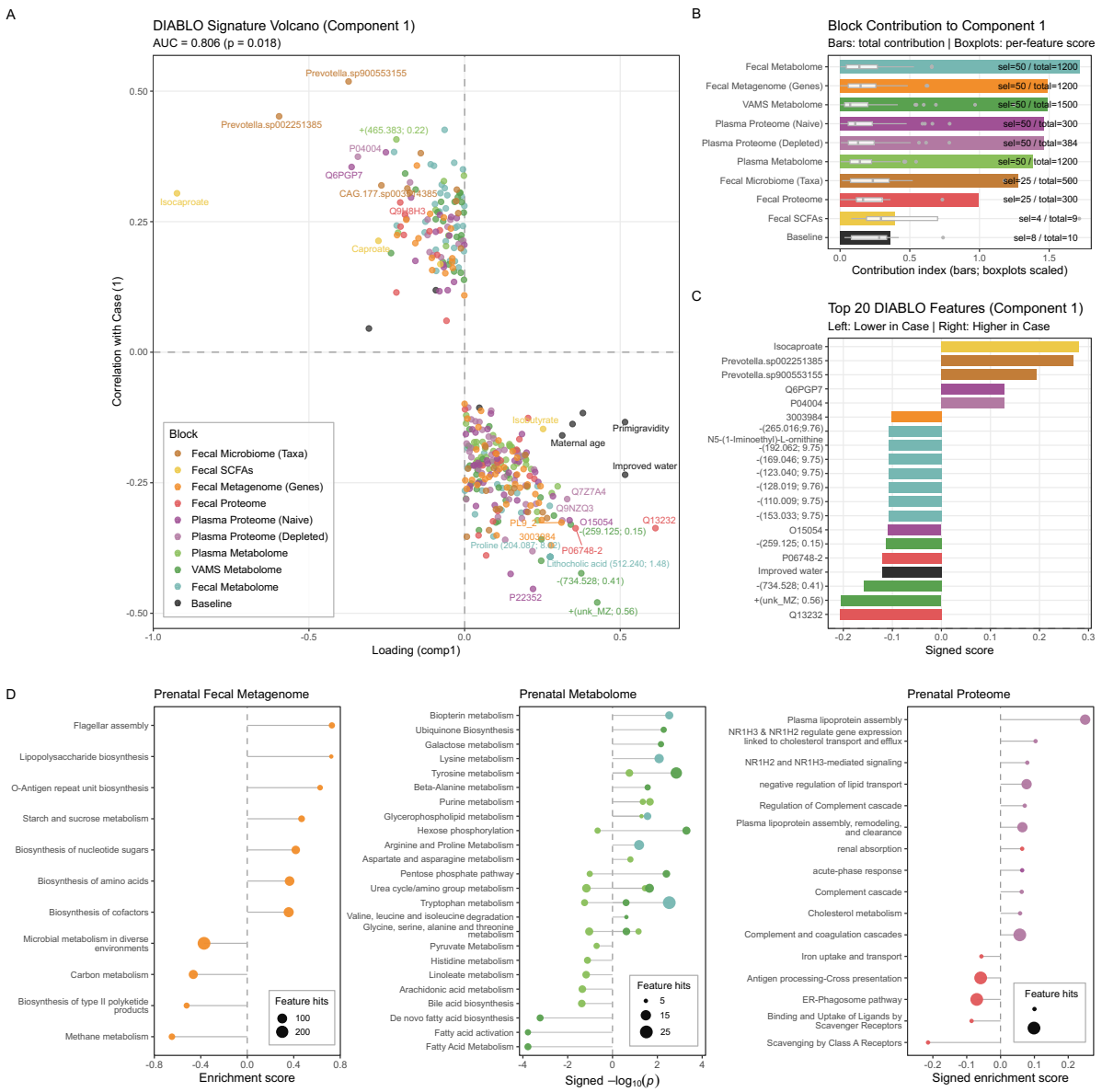


Figure 2. Prenatal maternal DIABLO multi-omics signature associated with SVN at trimester 3.

(A) DIABLO signature volcano plot for Component 1 (n = 58), showing feature loading (x-axis) and correlation with SVN status (y-axis). Positive y-values indicate features higher in SVN cases and negative values indicate features lower in cases.

(B) Block-level contribution to Component 1. Bars indicate total block contribution, and boxplots show the distribution of feature-level scores. Labels indicate selected versus total features included in each block.

(C) Top 20 Component 1 features ranked by signed score. Features to the right were higher in SVN cases, whereas those to the left were lower. Untargeted metabolites are labelled by ion mode (m/z; rt) when no annotation is available. A full feature table is provided in **Supplementary Data 1**.

(D) Prenatal pathway enrichment results for maternal fecal metagenome, maternal metabolome, and maternal proteome at the third trimester comparing SVN cases and controls. For metagenomic pathways, positive enrichment scores indicate pathways enriched in SVN cases, whereas negative scores indicate pathways depleted in cases. For metabolomic and proteomic pathways, positive signed scores indicate that contributing features were relatively higher in SVN cases, whereas negative scores indicate that contributing features were relatively lower in cases. Point size reflects the number of contributing features assigned to each pathway, and colors indicate the contributing compartment.

SVN is associated with postnatal metabolic remodeling and delayed infant gut ecological maturation

Milk differences between SVNs and controls were most evident at 3–4 months and were driven primarily by milk metabolites rather than broad shifts across all measured milk features (Figure 3A). At 14–21 days, SVNs exhibited lower disialyllacto-N-hexaose (DSLNH; FDR < 0.1), together with higher free fatty acids, including arachidonic acid, octadecenoic acid, eicosadienoic acid, eicosatrienoic acid, and γ-aminobutyric acid (GABA), whereas glutamine and serine were lower (Figure 3B-C). At 1–2 months, pyridoxamine was lower, whereas carnitine and GABA were higher in SVNs. By 3–4 months, differences shifted toward lipid metabolism, with 35 triglycerides, 13 phosphatidylcholines, six diglycerides, four ceramides, three sphingomyelins, and two each of dihexosylceramides and hexosylceramides enriched in SVNs, together with higher lysine and α-aminoadipic acid. Sensitivity analysis adjusting for BEP intervention produced similar results, although the associations for serine and GABA were no longer significant.

Differences in infant gut ecology were most evident in microbiome diversity rather than broad taxonomic composition. Shannon diversity and observed richness increased from 1–2 months to 5–6 months ($p = 0.002$ and $p < 0.001$, respectively), whereas no corresponding changes were observed in SVN cases ($p = 0.549$ and $p = 0.496$, respectively; Figure 3D). These differences were therefore driven by divergence in ecological progression, rather than by a marked baseline separation at 1–2 months (Figure S3B). At the microbiome compositional level, genus-level relative abundance profiles were broadly similar between SVN cases and controls within maternal milk, infant gut and maternal gut samples at each time point (Figure S3A). Beta-diversity differences were modest, with no strong case-control separation in most matrices, although maternal stool showed nominal differences at the third trimester ($p = 0.056$) and at 1–2 months ($p = 0.048$, Figure S3C). Differential abundance analysis identified a small number of candidate taxa that differed between groups within maternal milk, infant gut and maternal gut at specific timepoints, but these associations were not retained in sensitivity analyses using ANCOM-BC2 with pseudocount adjustment (Figure S3D).

Selective differences were also evident in infant stool analytes (Figure 3E). Total SCFAs, acetate, and isocaproate showed significant group-by-time interactions (p-int = 0.02, 0.004, and 0.052, respectively), with levels increasing in controls but decreasing in SVNs from 1–2 months to 5–6 months. These SCFAs were also higher in SVNs at 1–2 months (both total SCFAs and acetate $p < 0.05$, isocaproate FDR < 0.05), while acetate was higher in controls at 5–6 months ($p < 0.05$). Calprotectin also showed a significant group-by-time interaction (p-int = 0.039) and was higher in controls at 5–6 months ($p < 0.05$). In maternal stool, caproate and isocaproate were higher in SVN cases ($p = 0.033$

and p = 0.001, respectively, Figure S3E), while no differences were observed for calprotectin at any timepoints (Figure S3F).

Pathway-level analyses echoed the milk and gut differences observed across early and later postnatal timepoints (Figure 4). In metagenomic data, SVN-associated enrichment again centered on microbial structural and biosynthetic pathways, including LPS biosynthesis, together with folate biosynthesis, and related carbohydrate-utilization pathways such as phosphotransferase system (PTS). In postnatal metabolomic data, the clearest differences were seen in milk-associated pathways; features higher in SVN cases mapped predominantly to carbohydrate- and membrane-related pathways, including glycerophospholipid metabolism and glycolysis and gluconeogenesis, whereas features lower in cases were more often represented in amino acid-related pathways including glycine, serine, alanine and threonine metabolism, arginine and proline metabolism, and glutamate metabolism. Some pathways showed time-dependent directional differences, with sialic acid metabolism, fatty acid activation and fructose and mannose metabolism lower in SVN at early postnatal timepoints but higher in SVN at later timepoints. Postnatal proteomic pathways differences were less extensive than prenatal ones but continued to include complement-related signals, including complement activation and complement and coagulation cascades.

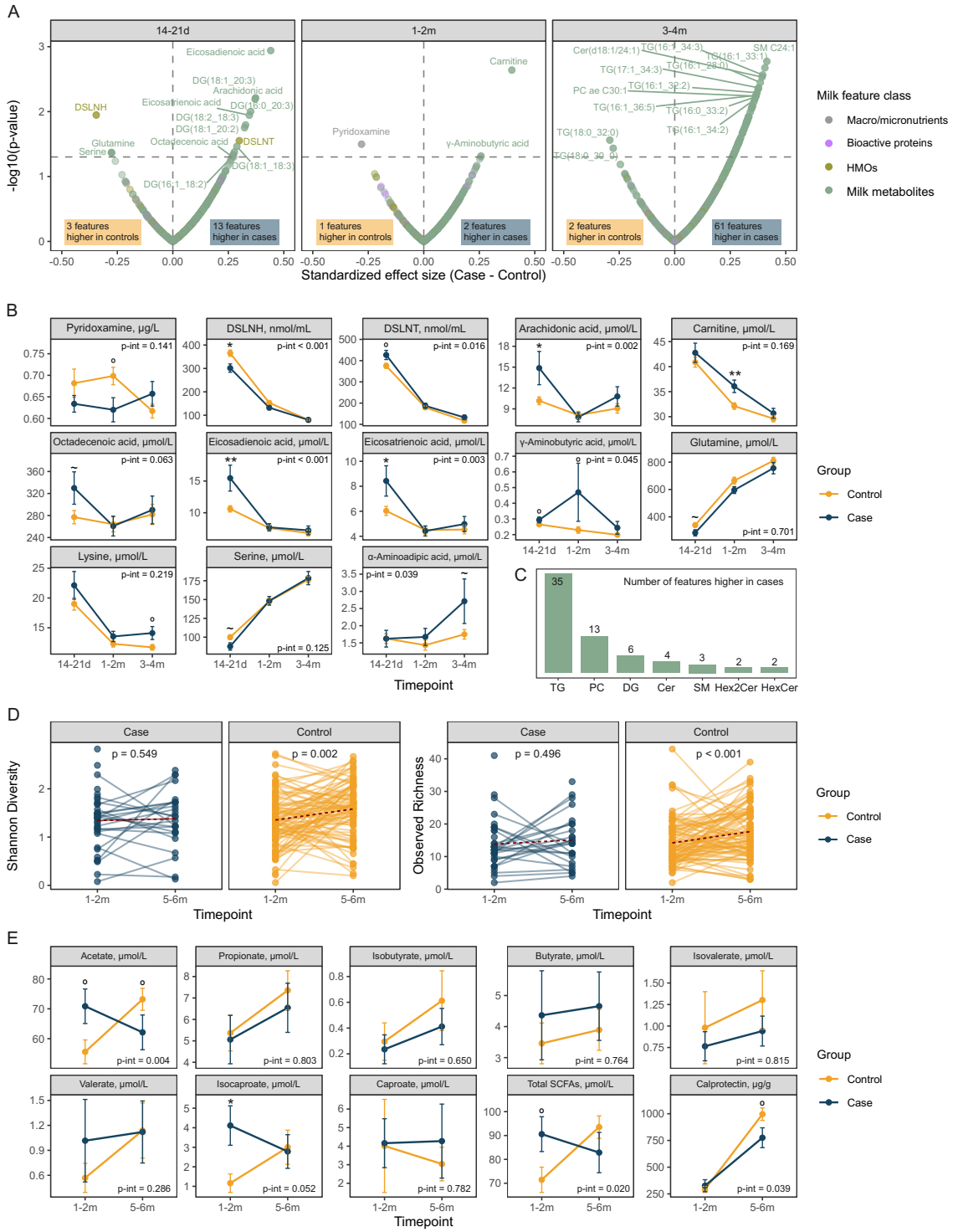


Figure 3. Postnatal milk and infant gut features associated with SVN status.

(A) Volcano plots of combined milk features at each time point comparing SVN cases and controls.

(B) Longitudinal trajectories of milk features identified from panel A, excluding lipids shown in (C).

(C) Distribution of nominally significant milk lipid features higher in SVN cases. Abbreviations: TG, triglycerides; PC, phosphatidylcholines; DG, diglycerides; Cer, dihydroceramides; SM, sphingomyelins; Hex2Cer, dihexosylceramides; HexCer, hexosylceramides.

(D) Infant gut diversity trajectories from 1–2 months to 5–6 months.

(E) Infant stool SCFAs and calprotectin trajectories from 1–2 months to 5–6 months. Mean ± SEM shown. *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.05, º FDR < 0.1.

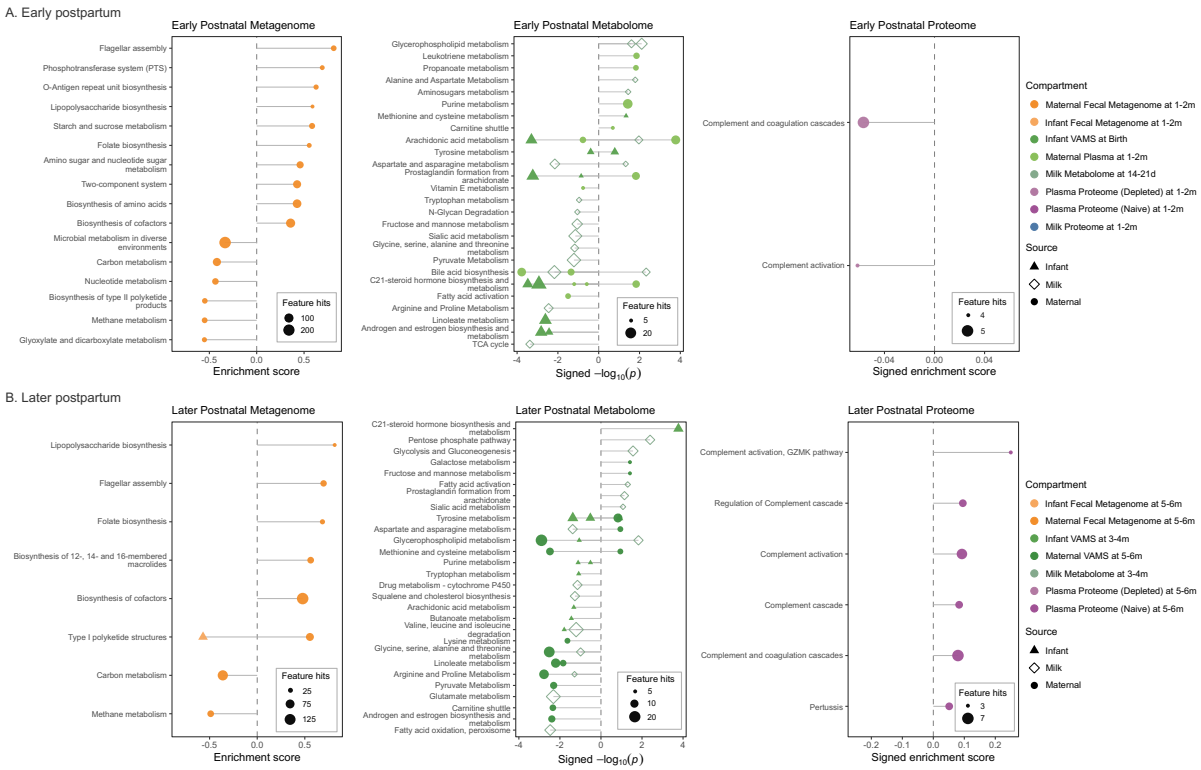


Figure 4. Pathway-level differences between SVN cases and controls for metagenomics, metabolomics, and proteomics across early and later postnatal time windows.

(A) Early postnatal pathway enrichment results for metagenome, metabolome, and proteome compartments comparing SVN cases and controls.

(B) Later postnatal pathway enrichment results for metagenome, metabolome, and proteome compartments comparing SVN cases and controls.

Early postnatal multi-omics prioritizes milk glutamine and serine as mediators of persistent growth deficits

Having identified postnatal biological differences associated with SVN, we next asked which of these features were most closely linked to persistent linear growth during infancy. We therefore analyzed two postnatal windows separately: an early window spanning birth to 1–2 months, and a later window spanning 3–4 to 5–6 months, depending on the dataset. Within each window, DIABLO we first used to identify multi-omics features discriminated SVN cases from controls, after which these candidate features were tested for association with repeated LAZ measurements using mixed-effects models. Early postnatal integrative analysis identified features associated with SVN status, with stronger discrimination in Component 2 (Component 2: AUC = 0.761, p = 0.037; Component 1: AUC = 0.678, p = 0.125; Figure 5; Supplementary Data 2).

Component 2 identified a coherent set of early postnatal features that both distinguished SVNs and were associated with persistent linear growth. Several features that were enriched in SVNs, including milk choline, multiple metabolite features, infant gut taxa at 1–2 months, and two milk microbial taxa, were associated with lower longitudinal LAZ. In contrast, the milk amino acids glutamine and serine, together with thiamine monophosphate (TMP), were lower in SVNs but positively associated with longitudinal LAZ. Component 1 showed a similar but weaker pattern, with several SVN-enriched metabolites, including DFLNT, associated with lower LAZ, whereas complement C1 (P09871) was higher in SVNs but positively associated with LAZ.

Applying the same framework to the later postnatal window yield weaker discrimination of SVN status (Component 2: AUC = 0.733, p = 0.051; Component 1: AUC = 0.644, p = 0.205; Figure S4). Correspondingly, substantially fewer SVN-associated features were also linked to longitudinal LAZ, and these associations were less directionally coherent than those observed during early infancy. The strongest later signal was the maternal plasma proteomic marker MASP1 (P48740, mannan-binding lectin serine protease 1), together with a limited number of maternal plasma and metagenomic features.

To determine whether these candidate growth-associated features partly explained the association between SVN and subsequent growth deficits, we performed mediation analyses. Among the tested features, milk glutamine and serine at 14–21 days showed the most robust and consistent indirect effects across multiple growth outcomes. Glutamine partially mediated the association between SVN and LAZ at 5 and 6 months, WAZ at 6 months, and length at 1 and 6 months, accounting for approximately 4-10% of the total SVN-associated growth deficit (Figure 5E). Serine showed broader mediation across LAZ, WAZ, weight, and MUAC outcomes, explaining approximately 4-9% of the total SVN effect (Figure 5F).

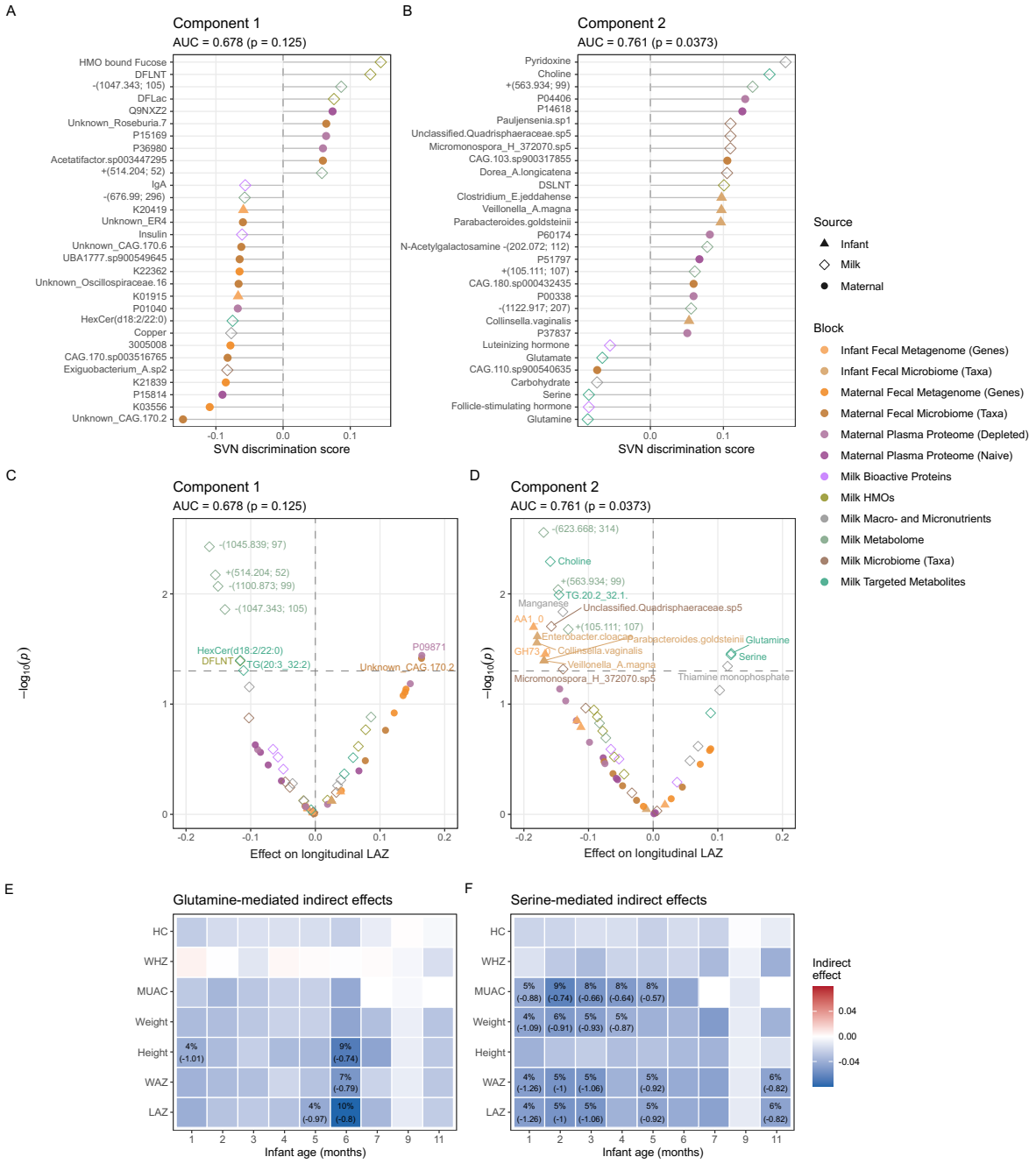


Figure 5. Early postnatal multi-omics features distinguishing SVN and their associations with longitudinal LAZ.
(A,B) Top early postnatal DIABLO features (n = 78) discriminating SVN cases from controls for Component 1 (A) and Component 2 (B). Positive scores indicate features higher in SVNs; negative scores indicate features lower in SVNs.
(C,D) Associations of the same component-specific features with longitudinal LAZ in mixed-effects models for Component 1 (C) and Component 2 (D). The x-axis shows the estimated effect of each feature on longitudinal LAZ, and the y-axis shows -log10(p). Positive coefficients indicate association with higher LAZ, and negative coefficients with lower LAZ.
(E,F) Mediation analyses evaluating whether early milk glutamine (E) and serine (F) mediate associations between SVN status and longitudinal growth outcomes. Heatmap color indicates the estimated indirect (mediated) effect, with red representing positive and blue representing negative mediation effects. Percentages within tiles indicate the proportion of the total SVN effect explained by the mediator and are shown only for significant indirect effects (95% bootstrap confidence interval excluding zero). Values in parentheses indicate the corresponding total effects. Growth outcomes include LAZ, WAZ, height, weight, MUAC, WHZ, and HC measured across infancy. A full feature table is provided in **Supplementary Data 2**.

Early WLZ catch-up in SVNs is associated with milk-centered multi-omics signatures

To identify biological features associated with the transient improvement in WLZ observed during early infancy, we modeled change in WLZ from birth to 2 months using block sparse partial least squares (block-sPLS) integrating maternal, infant, and milk multi-omics datasets collected during the early postnatal period (Figure 6A). Early WLZ catch-up was primarily associated with a milk-centered multi-omics signature, with milk-derived datasets contributing 31 of the top 50 ranked features (Supplementary Data 3). Component 1 showed the strongest overall signal driven by contributions from the milk metabolome, whereas milk bioactive proteins, HMOs and milk macro- and micronutrients showed the highest per-feature importance scores (Figure 6B).

Biological features associated with early WLZ catch-up showed mixed relationships with SVN status. Component 1 identified several SVN-associated features that were positively associated with greater WLZ catch-up, including DSLNT, multiple milk metabolite features, and an infant gut *Prevotella* species (sp900548745) (Figure 6C). Conversely, IGLL1 (P15814, immunoglobulin lambda-like polypeptide 1), which was lower in SVNs, was also positively associated with catch-up. In contrast, several features associated with reduced catch-up were enriched in SVNs, including eicosatrienoic acid and other milk metabolite features, whereas glutamine, DSLNH and the maternal plasma protein KLKB1 (P03952, plasma kallikrein) were lower in SVNs and likewise associated with reduced catch-up.

Additional highly ranked features associated with greater catch-up included multiple HMOs, including lacto-N-tetrose (LNT), sialyl-lacto-N-tetraose b (LSTb), and 6'-sialyllactose (6'SL), together with fibroblast growth factor-21 (FGF-21), nicotinamide adenine dinucleotide (NAD), maternal plasma proteomic markers, and maternal gut microbial taxa (Figure 6D).

Component 2 identified a complementary and broadly consistent signature, again dominant by milk-derived biology, with the milk microbiome providing the largest block contribution and HMOs contributing the highest feature important scores (Figure S5). DSLNT again emerged as a positive correlate early catch-up and was higher in SVNs, together with several infant and milk microbial taxa, including *Prevotella* species (sp900547575), *Sphingoaurantiacus polygranulatus* and *Bacillus kribbensis*, and the maternal plasma proteomic angiotensinogen (P01019). Relatively few features were negatively associated with catch-up, including glutamine and several milk metabolite features that were lower in SVNs. Additional highly ranked features again included multiple HMOs and milk nutrients.

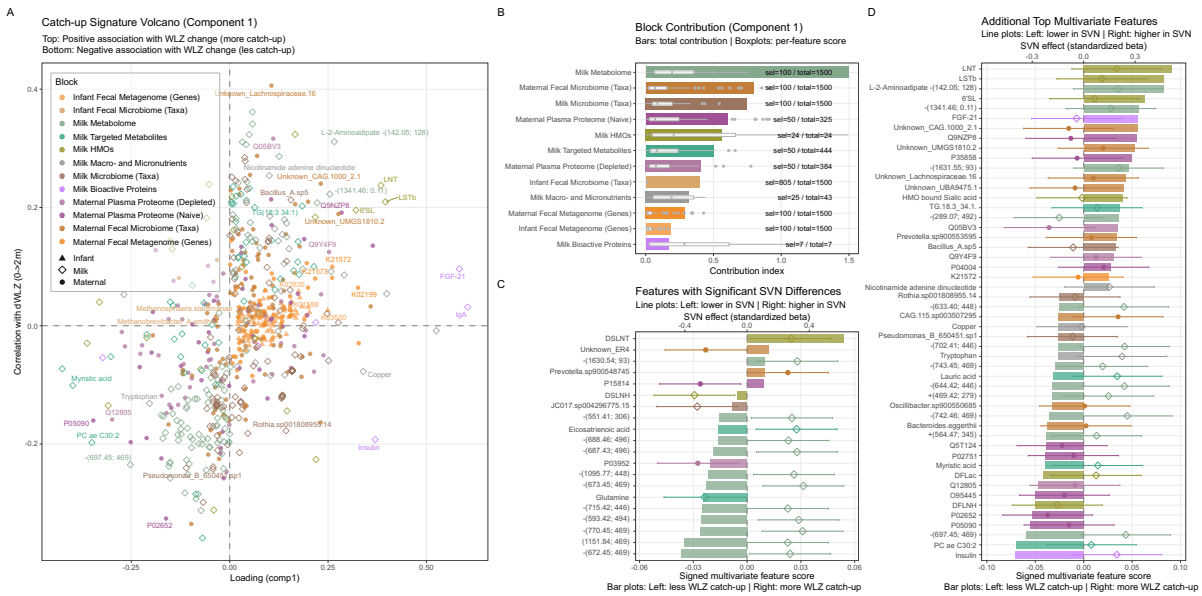


Figure 6. Early-life multi-omics features associated with early WLZ catch-up in SVNs are dominated by milk components

(A) Block sPLS volcano plot for Component 1 (n = 78), showing feature loading (x-axis) and correlation with change in WLZ from birth to 2 months (y-axis). Positive y-values indicate associations with greater early WLZ catch-up.
(B) Block-level contribution to Component 1. Bars indicate block-level contribution and boxplots show the distribution of feature-level scores. Labels indicate selected versus total retained features per block.
(C) Component 1 features with significant differences by SVN status. Bars represent signed multivariate scores (loading : correlation), with positive values indicating association with greater WLZ catch-up. Points and lines show standardized case-control effect estimates and 95% confidence intervals.
(D) Additional top Component 1 features ranked by signed multivariate feature score, displayed as in (C). A full feature table is provided in **Supplementary Data 3**.

Later WLZ faltering in SVNs is associated with multi-matrix multi-omics signatures

In contrast to the milk-centered signatures associated with early catch-up, later WLZ changes from 2 to 6 months was characterized by a shift toward infant gut-centered biology (Figure 7A; Supplementary Data 4). We modeled change in WLZ this period using maternal, infant, and milk multi-omics data collected across the later postnatal period. Component 1 was dominated by the infant fecal microbiome, whereas infant fecal SCFAs, and milk bioactive proteins showed the highest per-feature importance scores (Figure 7B).

Component 1 identified several biological features associated with protection against later WLZ decline. Infant stool acetate, which was lower in SVNs, was positively associated with maintenance of WLZ (Figure 7C). Additional positively associated features included selected infant and maternal metabolites, maternal plasma proteins including IGF2 (P01344, insulin-like growth factor 2) and CD5L (O43866, an immune-associated secreted protein), and the milk species *Acinetobacter towneri*, all of which were higher in SVNs. In contrast, features associated with greater WLZ decline were more numerous and included an infant gut *Prevotella* species (sp900553155), several infant metabolites, and maternal plasma proteins including SRGN (P10124, serglycin), most of which were higher in SVN cases.

Additional highly ranked features associated with maintenance of WLZ included milk bioactive proteins, including calprotectin, leptin, and follicle-stimulating hormone (FSH), together with infant stool butyrate and valerate, several infant gut taxa and milk microbial taxa (Figure 7D). Features associated with greater WLZ decline included several milk minerals, multiple HMOs, and maternal plasma proteomic and metabolomic signals.

Component 2 identified a complementary later postnatal signature that broadly confirmed the findings from component 1, with the milk microbiome providing the dominant block and infant fecal SCFAs and milk bioactive proteins contributing the highest feature important scores (Figure S6). Consistent with Component 1, infant fecal acetate and several metagenomic features were positively associated with WLZ change and were lower in SVNs, whereas relatively few features were negatively associated with WLZ change. Additional highly ranked signals again milk bioactive proteins, HMOs and stool metagenomic features.

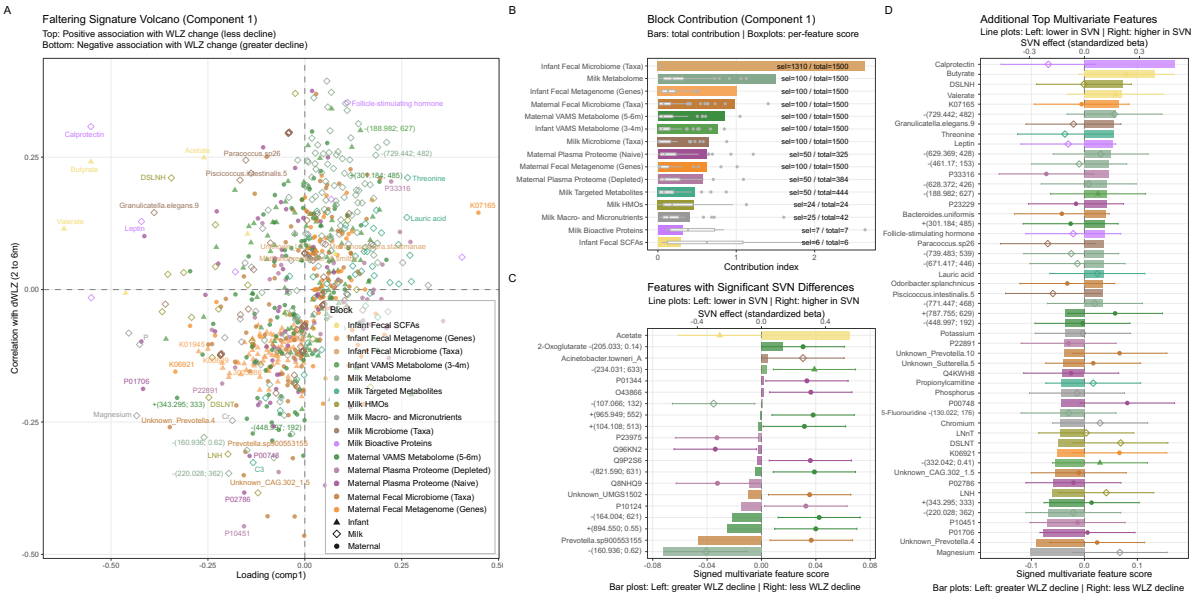


Figure 7. Later postnatal multi-omics features associated with WLZ faltering in SVN infants

(A) Block sPLS volcano plot for Component 1(n = 43), showing feature loading (x-axis) and correlation with change in WLZ from 2 to 6 months (y-axis). Positive y-values indicate association with less WLZ decline and negative values with greater decline.

(B) Block-level contributions to Component 1.

(C) Component 1 features with significant differences by SVN status. Bars represent signed multivariate scores, with positive values indicating association with less WLZ decline and negative values indicating more WLZ decline. Points and lines indicate standardized case–control effect estimates and 95% confidence intervals.

(D) Additional top Component 1 features ranked by signed multivariate feature score, displayed as in (C). A full feature table is provided in **Supplementary Data 4**.

Recovery from SVN is associated primarily with early milk metabolic remodeling

To examine heterogeneity in postnatal growth among SVNs, infants were classified according to recovery across LAZ, WAZ and WLZ trajectories (Figure 8). Recovery was defined separately for each anthropometric domain as attaining a sustain gain of >0.67 z-scores from birth at any follow-up timepoint[33], which was then maintained through the final available measurement. Infants meeting this criterion in at least one domain were classified as having any recovery. Using this definition, 46 SVNs were classified as recovered and 39 as non-recovered (Figure 8B). Recovery overlapped only partially across LAZ, WAZ, and WLZ, with both shared and domain-specific subsets.

Recovered and non-recovered infants diverged earliest for WAZ and WLZ, whereas separation in LAZ emerged later (Figure 8A). By 2 months, recovered infants already exhibited higher WAZ and WLZ than non-recovered infants (0.58 and 0.71 z-score units, respectively; $p = 0.023$ and $p = 0.025$), with these differences increasing through 11 months (WAZ: 0.84 – 1.14, $p < 0.001$; WLZ: 0.78 – 0.90, $p < 0.05$). In contrast, LAZ trajectories remained similar during the first 3 months before gradually diverging from 4–5 months onward, reaching a difference of 0.49 z-score units at 4 months ($p = 0.079$) and 0.63 z-score units at 5 months ($p = 0.027$), and these differences increased further to 1.07 at 11 months ($p < 0.001$).

Recovery status was not explained by the intervention. Distribution across the four intervention arms was similar between recovered and non-recovered infants (Fisher's exact $p = 0.676$) and neither prenatal BEP exposure (45.7% vs. 38.5%, $p = 0.519$) nor postnatal BEP exposure (58.7% vs. 48.7%, $p = 0.389$) differed between groups. Recovery analyses were adjusted for intervention arm to account for potential residual confounding.

Milk differences were concentrated within the same early postnatal period during which WAZ and WLZ trajectories first diverged (Figure 8C). Overall, recovery was characterized by higher abundance of lipid- and carnitine-related milk metabolites, whereas non-recovery was characterized predominantly by vitamin-, amino acid-, and carbohydrate-related metabolites. At 14–21 days, carnitine was higher in recovered infants, whereas EPA was higher in non-recovered infants. At 1–2 months, four triglyceride species, two phosphatidylcholine species, and proline were higher in recovered infants, while pyridoxal, vitamin B6, phenylalanine, and hexose were higher in non-recovered infants. By 3–4 months, butyrylcarnitine was higher in recovered infants, whereas biotin, nicotinamide, vitamin B3, pantothenic acid, lacto-N-fucopentaose III (LNFP.III) and histidine were higher in non-recovered infants.

Infant gut differences were more limited and were driven mainly by stool SCFAs rather than broader ecological changes (Figure 8D). Acetate, propionate, isobutyrate, valerate, and total SCFAs were generally higher in non-recovered infants at one or both postnatal

timepoints, while isocaproate showed differential temporal trajectories between groups (p-int = 0.031). In contrast, infant Shannon diversity did not differ by recovery status, although diversity tended to increase over time among recovered.

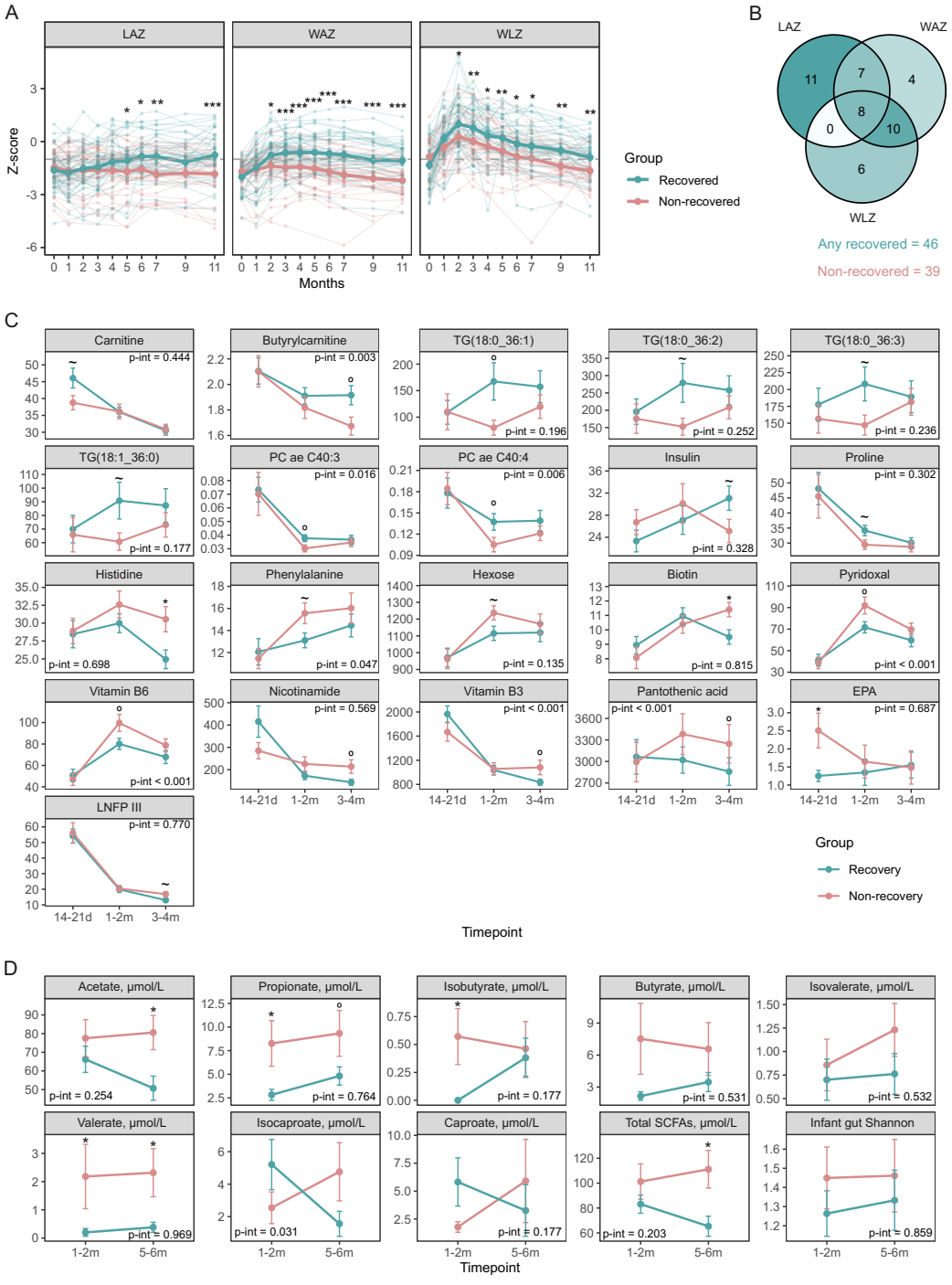


Figure 8. Postnatal milk and infant gut features associated with recovery among SVNs.
(A) Longitudinal LAZ, WAZ, and WLZ trajectories in the recovered and non-recovered groups. Recovery was defined as sustained improvement (>0.67 z-score from birth) in at least one anthropometric domain.
(B) Overlap of recovery across anthropometric domains and final numbers classified as recovered or non-recovered
(C) Longitudinal trajectories of milk features associated with recovery. Abbreviations: EPA, Eicosapentaenoic acid; TG, triglycerides; PC, phosphatidylcholines
(D) Infant stool SCFAs and Shannon diversity trajectories in the recovered and non-recovered groups. *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.05, º FDR < 0.1.

Discussion

In this longitudinal multi-omics study spanning pregnancy through early infancy, we identified coordinated prenatal host-microbial signatures that distinguished pregnancies resulting in SVNs and showed that these biological differences persisted across the maternal-milk-infant axis after birth. The strongest associations with subsequent growth were concentrated during the first two months of life, when early milk metabolites, particularly glutamine and serine, partially mediated persistent linear growth deficits. Distinct milk-centered signatures were associated with early ponderal catch-up and recovery, while broader multi-omics signatures were associated with later faltering. Together, these findings suggest that SVN represents not a single biological entity, but a convergence of vulnerabilities that begins before birth and remains detectable across the maternal-milk-infant interface during early life.

Our prenatal analyses indicate that the biological trajectory leading to SVN is established before birth and reflects the convergence of maternal nutritional, environmental, immune, and microbial factors rather than a single pathogenic pathway. Consistent with previous studies, mothers who delivered SVNs exhibited poorer nutritional status and greater household vulnerability, including lower BMI and MUAC, reduced access to improved water sources, and greater household food insecurity, all of which are well-recognized risk factors for fetal growth restriction and other adverse birth outcomes in low-resource settings[4,6,14,34]. Beyond these contextual determinants, prenatal multi-omics integration revealed coordinated biological alterations spanning maternal immunity, microbial ecology, and metabolism.

Particularly notable was the consistent enrichment of complement-related proteins and pathway across the prenatal analyses. Pregnancy requires tightly regulated complement activation to support implantation, placental development, and immune tolerance, whereas excessive or dysregulated activation has been implicated in placental insufficiency, fetal growth restriction, preeclampsia, and spontaneous preterm birth[35–37]. The accompanying enrichment of microbial structural pathways, including LPS biosynthesis and flagellar assembly, together with alterations in maternal metabolic pathways, further supports the presence of coordinated host-microbial perturbation in pregnancies resulting in SVN.

One of the most striking findings was that multiple independent lines of evidence pointed to the first two months of life as a pivotal developmental window for SVNs. Compared with later postnatal time points, early multi-omics profiles more clearly discriminated SVNs from controls and showed stronger and more coherent associations with persistent longitudinal growth. This same period also coincided with a transient WLZ catch-up and the earliest divergence in recovery trajectories among SVNs, suggesting that biological processes during early infancy may help shape subsequent growth. The first months of life are characterized by rapid physiological

adaptation, as the infant shifts from complete dependence on the intrauterine environment to exclusive reliance on human milk, while the gut microbiome undergoes a period of rapid compositional change, alongside parallel maturation of immune and metabolic systems[38–40]. Together, these findings highlight the first two months of life as a unique window of opportunity during which targeted intervention may have the greatest potential to redirect growth trajectories before biological vulnerabilities become more firmly established.

Within the early postnatal window, human milk showed the strongest and most consistent associations with subsequent growth, although the biological correlates differed according to the aspect of growth being considered. Persistent linear growth deficits were most consistently linked to lower concentrations of free amino acids glutamine and serine. Both metabolites were reduced in the milk of mothers of SVNs, positively associated with longitudinal LAZ, and partially mediated the association between SVN and persistent growth deficits. Glutamine is among the most abundant free amino acids in human milk and plays established roles in intestinal epithelial integrity, immune regulation, and neonatal growth, particularly in preterm and LBW infants[41,42]. These findings align with prior reports linking milk glutamine to infant growth and its lower concentration in the milk of mothers of preterm infants[41,43,44]. Serine supports one-carbon metabolism and phospholipid synthesis, pathways that are essential for tissue growth and development[45,46]. Together with the broader depletion of amino-acid-related pathways, these findings suggest that the milk of mothers of SVNs may provide a less favorable metabolic environment for supporting early linear growth.

In contrast, the biological signals associated with early ponderal growth differed from those underlying persistent linear growth. Early WLZ catch-up was associated predominantly with milk-centered signatures, particularly DSLNT and other HMOs, together with selected milk metabolites and bioactive components. DSLNT has been associated with reduced risk of necrotizing enterocolitis in preterm infants and is thought to support intestinal protection through interactions with the developing mucosal and microbial environment[47–49]. In our cohort, however, higher DSLNT was associated with transient rather than sustained improvement, suggesting that these milk adaptations may provide temporary biological support without fully overcoming the persistent growth vulnerability. Several free fatty acids were likewise elevated in early milk of SVN dyads, including arachidonic acid and other long-chain unsaturated fatty acids involved in energy supply, membrane structure, neurodevelopment, immune function, and intestinal maturation[26,50–52]. Arachidonic acid is preferentially transferred across the placenta during fetal life and continues to be supplied in human milk during the period of rapid postnatal growth[53]. Its elevation in early milk may therefore reflect a compensatory lactational adaptation to infant vulnerability or altered maternal lipid metabolism during the transition from pregnancy to lactation. This interpretation is

supported by the concurrent depletion of prenatal maternal fatty acid metabolic pathways, including arachidonic acid metabolism, in pregnancies resulting in SVNs.

Recovery analyses further suggested that the biology supporting sustained improvement differs from that underlying transient early catch-up. Recovery was characterized by a more carnitine- and lipid-enriched milk profile, whereas non-recovery showed relatively higher concentrations of several B vitamins. Higher carnitine in the milk of recovered SVNs is consistent with its central role in mitochondrial transport and oxidation of long-chain fatty acids, and its importance in neonatal energy metabolism, particularly in preterm and LBW infants with limited endogenous stores or synthesis[54]. Likewise, higher triglyceride and phosphatidylcholine species, major constituents of the milk fat globule membrane[55], may reflect a more lipid-rich nutritional environment that supports rapid ponderal recovery through enhanced energy delivery and lipid utilization[27,56].

Beyond the early postnatal window, however, the biology associated with ponderal growth shifted substantially. Although WLZ initially improved during the first two months, it declined again during later infancy, resulting in renewed separation between SVNs and controls. Correspondingly, early WLZ catch-up was associated predominantly with milk-centered signatures, whereas later WLZ changes was associated with broader host-microbial signatures incorporating maternal, milk, and infant gut features. Together, these findings suggest that early milk-derived factors may contribute to short-term compensatory growth, but may be insufficient to sustain durable recovery, which may depend more on successful maturation of the infant gut ecosystem and broader postnatal metabolic adaptation.

The increasing importance of broader host-microbial signatures during later infancy was accompanied by altered maturation of the infant gut ecosystem in SVNs. While non-SVNs showed age-related increase in microbial diversity together with progressive changes in microbial fermentation products[57–59], these developmental trajectories were comparatively blunted in SVNs. Such delayed gut microbial maturation has previously been linked to undernutrition and growth faltering in other low-resource settings[60–62]. Our findings suggest that this delayed maturation forms part of a broader biological transition in which the determinants of growth shift from predominantly milk-centered processes during early infancy toward increasing contributions from the developing infant gut ecosystem.

Although fecal calprotectin is widely used as a marker of intestinal inflammation in adults, its interpretation during early infancy remains challenging because concentrations are also influenced by breastfeeding[63]. In our cohort, where exclusive breastfeeding was nearly universal until 6 months, the tendency toward higher milk and fecal calprotectin in controls raises the possibility that fecal calprotectin may partly reflect exposure to milk-derived calprotectin rather than intestinal inflammation alone.

This study has several strengths. It integrates longitudinal multi-omics data across maternal, milk, and infant compartments, providing a systems-level view of early-life biology beyond single-domain analyses. Repeated sampling across key developmental windows enabled characterization of temporal dynamics in biological features and growth trajectories. The detailed maternal and household data allowed biological signatures to be interpreted within broader nutritional and environmental contexts. In addition to identifying biological features associated with SVN status, the study also examined heterogeneity in postnatal growth recovery, providing insight into factors associated with resilience among vulnerable infants. Finally, the study was conducted in a population with high breastfeeding prevalence, providing an important context to investigate the biology of SVNs in a low-resource setting.

Several limitations should be considered. Although the cohort was nested within a randomized trial, the omics analyses presented here are observational and cannot establish causality. The multivariate models were most informative as integrative feature-prioritization tools rather than prediction models. Sample sizes did not allow for sufficiently powered analyses of individual SVN sub-phenotypes, which may differ biologically. Many untargeted metabolite features remained partially or not annotated, constraining biological interpretation. Finally, milk volume was not measured, limiting interpretation of whether observed differences in milk composition reflected altered concentrations, differences in milk production, or both.

In summary, our findings support a model in which SVN reflects the convergence of prenatal and postnatal vulnerabilities across the maternal-milk-infant interface. Different biological systems contribute to distinct phases of early growth, with prenatal host-microbial perturbations preceding birth, early milk biology relating most closely to persistent linear growth and transient ponderal catch-up, and later host-microbial interactions accompanying subsequent growth faltering. These findings suggest that improving outcomes in SVNs will likely require interventions that begin before birth and continue through the first months of life, when biological trajectories appear most amenable to change.

Methods

Study design and samples

The present study was conducted within the BioSpé biospecimen sub-study nested in the MISAME-III trial, an individually randomized 2 × 2 factorial controlled trial in rural Burkina Faso evaluating balanced energy–protein (BEP) supplementation during pregnancy and/or lactation on birth outcomes and infant growth. The MISAME-III trial design has been described previously[29]. Briefly, pregnant women aged 15–40 years were enrolled between October 2019 and December 2020 in six rural health centers in the Houndé district (Hauts-Bassins region) and randomized to receive BEP or iron-folic acid (IFA) supplementation during pregnancy and independently during the first six months postpartum, generating four intervention groups.

The BioSpé sub-study was initiated after completion of MISAME-III recruitment, when most women were already in late pregnancy[30]. A total of 309 mother–infant dyads were enrolled and contributed biospecimens collected longitudinally from pregnancy through early infancy. For the present analysis, we included 294 dyads with available birth outcome classification and at least one relevant biological dataset, comprising 85 small vulnerable newborn (SVN) cases and 209 non-SVN controls. SVN was defined as the presence of one or more of the following birth phenotypes: preterm birth, small for gestational age (SGA), or low birth weight (LBW).

Biological sampling included maternal plasma, dried blood spots collected using volumetric absorptive microsampling (VAMS), stool, and human milk, as well as infant VAMS and stool samples. Maternal samples were collected from the second trimester through 6 months postpartum, and infant samples from birth through 6 months of age. Depending on the platform, data were available for metabolomic, proteomic, microbiome, metagenomic, short-chain fatty acid (SCFA), and milk composition analyses.

All participants received routine antenatal care according to national guidelines, including iron–folic acid supplementation and malaria prophylaxis. In the present study, intervention group was retained as a design variable and accounted for where appropriate in downstream analyses. Ethical approval was obtained from the Ethics Committee of Ghent University Hospital (B670201734334) and the Comité d'Éthique Institutionnel de la Recherche en Sciences de la Santé (50-2020/CEIRES). Written informed consent was obtained from all participants. The MISAME-III trial was registered at ClinicalTrials.gov (NCT03533712).

Biospecimen collection and processing

Whole blood collection by VAMS

Capillary whole blood samples were collected from mothers and infants using VAMS (Mitra™, Neoteryx, Torrance, CA, USA), which enables standardized low-volume collection and is particularly suitable for pregnant women and newborns because it simplifies sample handling, storage, and transport. For each collection, a total of 60 µL of whole blood was obtained using two 10 µL and two 20 µL VAMS tips. Blood was collected from mothers by finger prick (middle or ring finger) and from infants by heel prick using age- and weight-appropriate safety lancets. After local disinfection, the first drop of blood was discarded, and the VAMS tips were gently applied to the blood droplet until fully saturated, taking care to avoid milking of the puncture site. After collection, VAMS tips were placed in clamshell holders and air-dried at room temperature for approximately 3 hours before storage in sealed specimen bags. Samples were transported according to manufacturer recommendations and stored at −80 °C until analysis.

Maternal plasma collection and processing

Maternal plasma was collected by capillary blood sampling into plastic BD Microtainer whole-blood tubes spray-coated with K2 ethylenediaminetetraacetic acid (EDTA) (BD, Franklin Lakes, NJ, USA). After finger disinfection, approximately 500 µL of whole blood was obtained by finger prick using a high-flow safety lancet, discarding the first drop and avoiding excessive pressure at the puncture site. Samples were centrifuged on-site at 2,000 × g for 17 minutes using a microcentrifuge (VWR International, Leuven, Gelgium). Plasma was then separated, and 100 µL was aliquoted by single-channel pipette (Thermo Fisher Scientific, Merelbeke, Belgium) into barcoded terile cryotubes (Bioaigma, Cona, VE, Italy). Cryotubes were placed immediately in insulated containers, flash-frozen in liquid nitrogen, and subsequently stored at −80 °C until analysis.

Fecal sample collection and processing

Fecal sample collection has been described previously[64]. Maternal and infant fecal samples (~8 g) were collected using standardized protocols. Maternal samples were self-collected using sterile stool collection containers, whereas infant samples were collected using a sterile 38 × 50 cm protection sheet (Kimberley-Clark, Irving, TX, USA) wrapped around the infant to minimize contamination.

Samples were processed according to downstream analytical requirements. For microbiome analyses requiring stabilization, infant stool was transferred into OMNIgene•GUT OM-200 collection kits (DNA Genotek, Ottawa, Canada) and mixed with stabilizing solution according to manufacturer instructions. For flash-frozen

analyses, maternal stool and additional infant stool aliquots were transferred into sterile cryotubes (Biosigma, Cona, VE, Italy), immediately frozen, and stored at −80 °C until analysis.

Stool consistency was assessed using the Bristol Stool Scale, and liquid samples were homogenized prior to aliquoting to ensure uniformity. Standardized handling procedures were applied to minimize contamination and variability across samples.

Human milk collection and processing

Breastmilk samples collection has been described previously[65]. Briefly, milk samples were collected at health centers by trained staff using a standardized hygienic protocol. Milk was expressed using a hospital-grade electric breast pump (Medela Symphony®, Medela, Baar, Switzerland) with sterile, single-use collection kits. When necessary, both breasts were used to obtain approximately 25 mL of milk. To ensure representative sampling, milk expression continued until the breast was emptied and samples were gently mixed to homogenize fore- and hindmilk fractions.

Milk was aliquoted into four 2 mL amber microtubes and one larger polypropylene tube, protected from light, and kept at 4 °C prior to transfer to liquid nitrogen storage. Samples were subsequently stored at −80 °C until analysis.

Laboratory analysis

Human milk analyses

Human milk samples were analyzed for microbiota composition, human milk oligosaccharides (HMOs), macronutrients, micronutrients, bioactive proteins and metabolites following standardized protocols developed within the International Milk Composition (IMiC) consortium[66].

Milk microbiota profiling was performed at Baylor College of Medicine using 16S rRNA gene sequencing targeting the V4 hypervariable region. Given the low-biomass nature of human milk, sample processing incorporated extensive contamination control measures, including negative controls at all stages from aliquoting to sequencing. DNA was extracted from 1 mL of milk following centrifugation (10,000 × g, 5 min) to remove fat and supernatant, using the DNeasy 96 PowerSoil Pro HT Kit (Qiagen) on a QIACube HT automated platform. Sequence data were processed using the DADA2 (v1.26.0) pipeline implemented in QIIME2 (v2023.2). Taxonomic assignment was performed using the GreenGenes2 (v2022.19) database[67]. Contaminant features were identified and removed using both Decontam[68] and SCruB[69] algorithms, informed by negative controls.

Human milk oligosaccharides were quantified at the Mother–Milk–Infant Center of Research Excellence (MOMI CORE, University of California San Diego). HMOs were isolated by high-throughput solid-phase extraction, fluorescently labeled and analyzed by high-performance liquid chromatography with fluorescence detection (HPLC-FLD) as previously described[70,71]. Identification and quantification were based on retention times and mass spectrometry, using raffinose as an internal standard for absolute quantification.

Macronutrients (fat, protein, carbohydrate/lactose and energy content) were measured by near-infrared spectroscopy using a SpectraStar XT (KPM analytics, Westborough, MA, USA) calibrated specifically for human milk[72,73]. Analyses were conducted at the Manitoba Interdisciplinary Lactation Center biorepository.

Micronutrients, including multiple B vitamins[74–76], fat-soluble vitamins (A and E)[77], and minerals and trace elements (calcium, copper, iron, magnesium, potassium, selenium, sodium and zinc)[78], were quantified at the USDA Agricultural Research Service–Western Human Nutrition Research Center, Davis, CA using established protocols[79]. Free choline was quantified by liquid chromatography-mass spectrometry (LC-MS/MS) (Biocrates Life Science, Austria) as part of the targeted metabolomics workflow described below[80].

Milk bioactive proteins were quantified using high-performance electrochemiluminescence assays (Meso Scale Discovery, USA at MOMI CORE)[81]. Analytes included secretory immunoglobulin A (sIgA), calprotectin and selected metabolic hormones (including insulin, leptin and fibroblast growth factor-21). Milk samples were diluted in assay-specific buffers, and assays were multiplexed where concentration ranges permitted.

Targeted metabolomic profiling of milk was performed at Biocrates Life Sciences using LC-MS/MS and flow injection analysis–tandem mass spectrometry (FIA–MS/MS) with the MxP® Quant 500 kit (Biocrates Life Science AG, Innsbruck, Austria)[80,82]. This platform enabled absolute quantification of more than 600 metabolites, including triglycerides, free fatty acids, amino acids and acylcarnitines.

Stool analyses

Shotgun metagenomic sequencing: DNA extraction, library preparation and sequencing of maternal and infant stool samples were performed at the UC San Diego IGM Genomics Center. DNA was extracted using the MagMAX Microbiome Ultra kit and quantified by PicoGreen. Libraries were prepared using KAPA HyperPlus kits with 25 ng input DNA and nine PCR cycles. Sequencing was conducted on an Illumina NovaSeq 6000 platform (2 × 150 bp), targeting approximately 25 Gbp per maternal sample and 10

Gbp per infant sample. Reads were quality-filtered and deduplicated using FastP, and host reads were removed by alignment to the human reference genome (hg38) using Bowtie2. Microbial reads were mapped to the HPlusDB_BacteriaArchaea reference genome database optimized for non-industrialized infant microbiomes and annotated using GTDB R95[83–85]. Genomes with >50% breadth of coverage were retained. Functional profiling was performed using the inStrain pipeline under default parameters[86].

Fecal metabolomic profiling was performed using LC-MS/MS. Metabolites were extracted from ~10 mg of fecal material using an extraction mixture of methanol plus stable isotope labeled internal standards following established protocols[87]. Metabolites were separated by HILIC high-performance liquid chromatography and analyzed using electrospray ionization in positive ion mode. Data dependent acquisition was used to collect MS/MS spectra. Raw spectra were processed using MS-DIAL[88] to detect and align peaks across samples and generate non-redundant representative features. Metabolite annotations were assigned by matching MS/MS spectra against an MS2 metabolite library composed of spectra from GNPS, MoNA, and MassBank. Peak intensities were normalized to total ion current, followed by blank filtering (features ≥3× higher in QC than blanks) and QC stability filtering (relative standard deviation <30% across QC runs). Technical replicates were averaged prior to downstream analyses.

Fecal proteomic profiling was performed using LC-MS/MS–based shotgun proteomics. Samples were lysed by bead-beating in SDS-containing buffer, followed by S-trap (Protifi, Fairport, NY, USA) based tryptic digestion of proteins and reverse-phase (RPS) peptide desalting using the Agilent AssayMAP Bravo Protein Sample Prep Platform (Santa Clara, CA, USA), as described before[89]. Peptides were separated over a 35-minute method using a Thermo Vanquish Neo ultra-high performance liquid chromatography (UHPLC) system. Electrospray ionization was performed via a Thermo Nanospray Flex source, and peptides were analyzed on a Thermo Fusion Lumos mass spectrometer in data-dependent acquisition mode (DDA). Spectra were searched using FragPipe (v22.0)[90], against a human Swiss-Prot and TrEMBL protein database (canonical and isoform sequences, downloaded from Uniprot on 2024-01-09)[91], with common contaminants and decoy sequences added using Fragpipe. The default workflow was used, with phosphorylation (STY) and deamidation (NQ) specified as additional variable modifications. Protein intensities were normalized and processed using the MSstats R package[92], including median scaling and missing value imputation using an accelerated failure time model, followed by context-aware imputation for remaining missing values prior to downstream analyses.

Short-chain fatty acids (SCFAs) and branched-chain SCFAs were quantified at ProDigest (Belgium) using gas chromatography with flame ionization detection, as described before[93]. Stool samples were partially freeze-dried to determine water content prior to extraction. Analytes included acetate, propionate, butyrate, isobutyrate,

isovalerate, valerate, isocaproate and caproate. Concentrations were expressed as μmol/g wet feces. Values below the limit of quantification were imputed as half the corresponding limit.

Fecal calprotectin concentrations were measured in maternal and infant stool samples using a commercially validated enzyme-linked immunosorbent assay (ELISA), following manufacturer instructions[94]. The assay demonstrated acceptable intra- and inter-assay precision and dilution linearity across the concentration range.

Untargeted milk and blood metabolomics

Untargeted metabolomic profiling of human milk, maternal plasma, maternal VAMS and infant VAMS samples was performed using rapid liquid chromatography-mass spectrometry (rLC-MS) at Sapient Bioanalytics (San Diego, USA), as previously described. Human milk samples were equilibrated on an orbital shaker (550 rpm, 4 °C, 10 min), after which a 20 μL aliquot was extracted with 80 μL of solvent containing internal standards. Samples were mixed (550 rpm, 4 °C, 10 min) and centrifuged (6,000 × g, 4 °C, 10 min). Supernatants were transferred to 384-well polypropylene plates containing methanol–water mixtures optimized for positive and negative ionization modes, respectively. Extraction solvents, dilution conditions and final sample concentrations were optimized using pooled human milk samples. Both annotated and unannotated metabolomic features were retained for downstream analyses.

Untargeted milk and blood proteomics

Untargeted proteomic profiling of human milk and maternal plasma was conducted at Precision Biomarker Laboratories (Cedars-Sinai Medical Center, California, USA) using data-independent acquisition mass spectrometry (DIA-MS). Maternal plasma was analyzed in both naive and top-14 protein-depleted fractions, following a standardized blood biofluid proteomics workflow described previously[95]. Human milk samples were processed using an S-Trap-based protein digestion workflow optimized for milk matrices, while plasma samples were diluted in trifluoroethanol and digested using Trypsin enzyme. Sample preparation was fully automated on a BioMek i7 workstation (Beckman Coulter). Peptides were analyzed on a U3000-Exploris 480 LC-MS platform using a 30-minute dual-trap DIA method for milk samples and a 20-minute DIA method for plasma samples. Indexed retention time (iRT) standards, pooled reference samples, and other quality control (QC) measures were used to monitor sample processing, system performance, and reproducibility. Proteomic data were processed using DIA-NN to generate peptide- and protein-level quantification for downstream analyses.

Statistical analyses

All analyses were conducted in R (v4.5.2). Unless otherwise specified, analyses were two-sided. For feature-level analyses within omics blocks, false discovery rate (FDR) correction was applied within each blocks using Benjamini-Hochberg method. Multi-omics multivariate models were interpreted primarily as integrative feature-prioritization approaches rather than prediction tools.

Univariate analyses

Baseline maternal, infant, and household characteristics were compared between SVN cases and controls using linear regression for continuous variables and logistic regression for binary variables.

To compare longitudinal growth trajectories between SVN cases and controls over the first year of life, mixed-effects models were fitted for repeated anthropometric outcomes. Models included fixed effects for time, SVN group, and their interaction, with a random intercept for infant. Time was modeled categorically when comparing trajectories across study visits. Estimated marginal means and 95% confidence intervals were derived from fitted models, and cross-sectional group contrasts at individual timepoints were obtained using post hoc comparisons. P values for repeated cross-sectional contrasts within each outcome were adjusted using the Benjamini-Hochberg method.

To characterize postnatal biological differences between SVN cases and controls, feature-wise analyses were conducted separately within each compartment and time window. For cross-sectional comparisons at individual timepoints, linear regression models were used with SVN status as the primary exposure. For repeated-measures analyses across postnatal timepoints, mixed-effects models were fitted with fixed effects for SVN group, time, and group-by-time interaction, with a random intercept for participant. These models were applied to milk macro- and micronutrients, HMOs, bioactive proteins, milk metabolites, stool SCFAs, fecal calprotectin, and diversity metrics. For feature-level analyses within omics blocks, p values were adjusted using the Benjamini-Hochberg method. For the recovery analyses, BEP intervention group was fitted as a covariate to adjusted for confounding.

Microbiome diversity analysis

Alpha and beta diversity analyses were performed for the infant gut, maternal gut, and milk microbiome using the phyloseq (v1.44.0)[96] and vegan (v2.6-4)[97] R packages. Samples from all three microbiome datasets were rarefied to a common sequencing depth defined as the minimum sample depth across the milk, maternal gut, and infant gut phyloseq objects.

Alpha diversity metrics, including Shannon diversity index and observed richness, were calculated based on rarefied count data for each biological compartment and time point. Linear mixed-effects models were fitted using the nlme (v3.1-163)[98] package, with SVN status as the fixed effect and participant ID as a random intercept. Pairwise comparisons between SVN cases and controls were performed using emmeans (estimated marginal means), and p-values were adjusted for multiple testing using the Benjamini-Hochberg FDR correction. Longitudinal changes in alpha diversity were also evaluated within each group by fitting mixed-effects models with time point as the fixed effect and participant ID as a random intercept.

Beta diversity was assessed using weighted UniFrac distance metrices, calculated from phylogenetic trees within phyloseq (v1.44.0). Principal Coordinate Analysis (PCoA) was used for ordination visualization, and 95% confidence ellipses were drawn for each group. Permutational multivariate analysis of variance (PERMANOVA) was conducted using adonis2 from the vegan (v2.6-4) package to test the association of SVN status, time point, and their interaction with microbiome composition. P-values from time point-specific PERMANOVA tests were adjusted using the Benjamini–Hochberg method.

Differential abundance analysis

Differential abundance analysis of microbial taxa was performed at the species level using ANCOM-BC2 (v2.2.1)[99] implemented in R. Analyses were conducted separately for the infant gut samples, milk samples and maternal gut samples at each time point. Raw count data from phyloseq objects were used as input. Low-prevalence taxa were excluded using a prevalence cutoff of 5%, and samples with sequencing depth below 500 reads were removed. Models included SVN status and intervention group as fixed effects, with SVN status specified as the primary group variable. Structural zeros were identified and accounted for, and negative lower-bound correction was applied. P-values were adjusted using the Benjamini–Hochberg method to control the FDR. Differential abundance results were summarized as log fold-changes for SVN cases relative to controls, together with FDR-adjusted p-values.

Metagenomic functional data analysis

Gene-level functional profiling was conducted using hierarchical differential abundance analysis on count data normalized by the trimmed mean of M-values (TMM) method implemented in treeclimR (v0.1.5)[100] and edgeR (v3.42.4)[101]. Genes annotated in the KEGG, HMO, CAZyme, and CARD databases were organized into a hierarchical structure based on their corresponding enzymatic class, functional category, HMO

utilization cluster, or antibiotic resistance mechanism. To identify pathways showing coordinated functional shifts, gene set enrichment analysis (GSEA) was performed using clusterProfiler (v4.18.2)[102] with KEGG pathway mappings. Enrichment scores and adjusted p-values were calculated for each pathway, applying an adjusted $p < 0.05$ and a minimum gene set size of two.

Metabolome pathway analysis

Pathway-level analyses of untargeted metabolomic data were performed separately by biological matrix and time point. Untargeted metabolomic datasets were first preprocessed by imputing missing values, capping extreme outliers, and standardizing feature intensities using z-score scaling. For each dataset, metabolomic features were tested for association with SVN status using linear regression models. Features were retained for downstream pathway analysis if they had sufficient non-zero variance. For each retained feature, the regression coefficient, p-value, FDR-adjusted p-value, and group means were extracted.

To preserve biological directionality, features were classified as higher or lower in SVN cases according to the sign of regression coefficient. Because metabolite annotation depends on ionization polarity, features were additionally separated by positive and negative ionization mode; fecal metabolomic data were analyzed on positive ion mode only. Consequently, pathway enrichment analyses were conducted independently for each combination of ionization model and direction of association.

Pathway enrichment was conducted using mummichog package (v2.7)[103], which infers pathway activity directly from high-resolution mass spectrometry data without requiring full metabolite identification. For each subset, input files containing mass-to-charge ratio (m/z), retention time, p-values, and effect size statistics (beta-coefficients) were provided to mummichog. Analyses were run using a nominal p-value cutoff of 0.05 and 100 permutations. Enriched pathways were extracted from the mummichog output and summarized using pathway p-values, pathway size, overlap size, and overlap features.

This workflow was applied uniformly across all untargeted metabolomic datasets, ensuring that pathway inference was conducted in a time- and matrix-specific manner while maintaining consistent analytical logic across analyses. For visualization and interpretation, pathway enrichment results were into three developmental periods: prenatal, early postnatal, and later postnatal. Prenatal analyses included third-trimester maternal fecal, plasma, and VAMS metabolomes. Early postnatal analyses included infant VAMS metabolome at birth, maternal plasma metabolome at 1–2 months postpartum, and milk metabolome at 14–21 days postpartum. Later postnatal analyses included infant VAMS metabolome at 3–4 months, maternal VAMS metabolome at 5–6 months postpartum, and milk metabolome at 3–4 months

postpartum. Pathway enrichment results were visualized as signed enrichment dot plots, where pathways depleted in SVN were displayed on the left and pathways enriched in SVN on the right. Point color represented the biological compartment, point size reflected the number of overlapping metabolite features contributing to the pathway, and point shape indicated the sample source (maternal, infant, or milk).

Proteomic pathway enrichment analysis

Proteomic pathway enrichment analyses were performed separately by biological matrix and time point. proteomic datasets were first preprocessed by imputing missing values, capping extreme outliners, and standardizing feature intensities using z-score scaling. For each dataset, proteins were tested for association with SVN status using linear regression models. Features were retained for downstream analysis if they had non-zero variance. For each retained protein, the regression coefficient, p-values, FDR-adjusted p-values, and group means were extracted.

Proteins showing nominal evidence of association with SVN status ($p<0.05$) were stratified into proteins higher or lower in SVN cases according to the sign of the regression coefficient. Protein identifiers were harmonized by removing isoform suffixes where appropriate and, for fecal proteomic features, extracting UniProt identifiers from compound feature labels. UniProt identifiers were then mapped to Entrez Gene IDs using the org.Hs.eg.db annotation database. Enrichment analyses were conducted separately for proteins higher and lower in SVN within each dataset.

Pathway enrichment was using over-representation analysis implemented in the clusterProfiler (v4.18.2)[102] and ReactomePA. Gene Ontology (GO) biological processes, KEGG pathways, and Reactome pathways were tested separately. GO biological process terms were simplified using semantic similarity-based reduction. Enrichment pathways were retained if they included at least five proteins, had a RichFactor ≥ 0.1, and met an FDR-adjusted q-value threshold of $q < 0.05$. RichFactor was calculated as the ratio of the number of overlapping proteins to the pathway background size. A signed enrichment score was derived from the RichFactor, with positive values indicating pathways enriched among proteins higher in SVN and negative values indicating pathways enriched among proteins lower in SVN.

This workflow was applied to prenatal, early postnatal, and later postnatal proteomic datasets. Prenatal analyses included third-trimester fecal proteome and maternal plasma proteome measured in naive and top-14 protein-depleted fractions. Early postnatal analyses included milk proteome and maternal plasma proteome measured in naive and top-14 protein-depleted fractions at 1–2 months. Later postnatal analyses included maternal plasma proteome measured in naive and top-14 protein-depleted fractions at 5–6 months. Proteomic pathway results were visualized as signed

enrichment dot plots, with pathways lower in SVN shown on the left and pathways higher in SVN shown on the right. Point color denoted biological compartment, point size indicated the number of overlapping proteins, and point shape denoted pathway source (GO, KEGG, or Reactome).

DIABLO modeling

To identify prenatal maternal biological features associated with SVN birth, third-trimester maternal datasets were integrated using DIABLO (Data Integration Analysis for Biomarker discovery using Latent cOmponents), implemented in mixOmics (v6.34.0)[104]. Omics blocks included maternal stool metagenome, stool microbiome, stool metabolome, stool proteome, stool short-chain fatty acids, plasma metabolome, plasma proteome, VAMS metabolome, and baseline clinical characteristics. Prior to model fitting, feature blocks were restricted to participants with overlapping samples across the included datasets, and variables were centered and scaled within each block. SVN status was used as the supervised outcome. Model discrimination was summarized using component-specific area under the receiver operating characteristic curve (AUC). Features contributing most strongly to each component were ranked according to variable loadings and their correlation structure across blocks, and were interpreted as prenatal biological signatures associated with SVN birth.

To investigate whether postnatal biological features associated with SVN also related to persistent linear growth deficits, DIABLO was applied separately to an early postnatal window (birth, 14-21 days, and 1-2 months) and a later postnatal window (3-4 and 5-6 months), again using SVN status as the supervised outcome. Omics blocks were centered and scaled before model fitting, and only participants with overlapping data across the included blocks were retained. Model discrimination was summarized using component-specific AUC. Features contributing to the main discriminating components were then carried forward into mixed-effects models of repeated LAZ, with infant included as a random effect. This two-stage framework was used to identify postnatal features that both discriminated SVN status and were associated with subsequent longitudinal LAZ.

Block-sPLS modeling

To identify biological features associated with short-term ponderal catch-up, block sparse partial least squares (block-sPLS) regression was performed using mixOmics (v6.34.0)[104], with change in WLZ from birth to 2 months as the continuous outcome. Maternal, milk, and infant omics blocks measured during the early postnatal window were included as predictors. An analogous block-sPLS framework was used to identify

biological features associated with later WLZ faltering, using change in WLZ from 2 to 6 months as the outcome and omics blocks measured during the later postnatal window as predictors.

Prior to model fitting, each feature block was restricted to participants with non-missing outcome data and overlapping samples across blocks. Feature blocks with fewer than 40 samples overlapping the outcome were excluded from the multiblock model. Within each retained block, features were centered and scaled. Block-sPLS models were then fitted using a fixed number of latent components, with sparse feature selection applied within each block. Model interpretation focused on component loadings and ranked feature contributions, which were used to identify biological features most strongly associated with WLZ change.

Model performance was assessed by repeated cross-validation and summarized using root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). For the early postnatal model, 78 participants had overlapping outcome and predictor data (n = 78), and cross-validated performance was modest (RMSE = 1.57, MAE = 1.18, R^2 = -0.069). For the later postnatal model, 43 participants contributed overlapping data (n = 43), and performance was similarly modest (RMSE ≈ 0.94, MAE ≈ 0.70, R^2 ≈ -0.048). Because predictive performance was limited, these models were interpreted primarily as integrative feature-prioritization approaches rather than as predictive models.

To aid biological interpretation, top-ranked features from the block-sPLS models were subsequently examined in complementary univariate analyses. These included feature-wise models testing association with WLZ change, as well as case-control models comparing feature abundance between SVN infants and controls. This two-step strategy was used to identify features that both contributed to the multivariate WLZ-associated signature and showed interpretable directionality in relation to SVN status.

Data availability

The raw metagenomic and 16S rRNA gene sequencing data analyzed in this study have been deposited at the NCBI database under BioProject accession code PRJNA1113165 and PRJNA1292286. The milk composition data can be accessed upon request to meghan.azad@umanitoba.ca, and the de-identified personal data can be access upon request to carl.lachat@ugent.be.

Acknowledgements

This research was funded by grant INV-001734 and OPP1175213 from Bill & Melinda Gates Foundation. L. D. is supported by the China Scholarship Council (Grant No. 202207650056). M.B.A. holds a Tier 2 Canada Research Chair in Early Nutrition and the Developmental Origins of Health and disease and is a Fellow of the CIFAR Humans and the Microbiome Program. The funders had no role in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript. The authors thank all participants from Burkina Faso and the data collection team. Milk and plasma proteomics was carried out by Cedars-Sinai Precision Biomarker Laboratories and the Cedars-Sinai Proteomics and Metabolomics Core.

Author contributions

Conceptualization, formal analysis, visualization, and writing – original draft: L.D.;
Project administration: L.C.T., C.L., T.D.-C.;
Funding acquisition: L.C.T., C.L., M.B.A., T.D.-C.;
Methodology: L.C.T., K.F., J.E.E., C.G.G., D.H., M.R.O., M.B.M., S.S., Y.G., K.L., J.E.V.E., L.H.A., J.L.S., M.B.A.;
Investigation: K.F., L.B., A.C., D.H., A.Mo., M.B.M., L.O.O., M.S., N.S., K.L., J.E.V.E., L.H.A., J.L.S., M.B.A.;
Resources: K.F., L.B., B.C.D., K.L., J.E.V.E., L.H.A., J.L.S., M.B.A., T.D.-C.;
Data curation: K.F., A.C., B.C.D., J.E.E., C.G.G., E.R.G., A.Me., M.R.O., C.A.O., A.R., P.L.M., L.O.O., S.S., E.W., Y.G., T.D.-C.;
Supervision: B.D.M., B.C.D., C.G.G., M.O., C.L., J.L.S., M.B.A., T.D.-C.;
Writing – review & editing: all authors

Declaration of Interests

M.B.A. has received speaking honoraria from non-profit organizations that support breastfeeding (Institute for the Advancement of Breastfeeding & Lactation Education, Thai Breastfeeding Centre, UK Baby Friendly), and companies that produce human milk-

related products (Prolacta Biosciences, Medela). She is a scientific advisor to Human-Milk.com, TinyHealth (an infant microbiome testing company), and PumpKin Baby (a human milk preservative company). She has consulted for All G (a food ingredient company).

Reference

1. Ashorn, P., Ashorn, U., Muthiani, Y., Aboubaker, S., Askari, S., Bahl, R., Black, R.E., Dalmiya, N., Duggan, C.P., Hofmeyr, G.J., et al. (2023). Small vulnerable newborns-big potential for impact. Lancet *401*, 1692–1706. https://doi.org/10.1016/S0140-6736(23)00354-9.

2. Blencowe, H., Krasevec, J., de Onis, M., Black, R.E., An, X., Stevens, G.A., Borghi, E., Hayashi, C., Estevez, D., Cegolon, L., et al. (2019). National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: a systematic analysis. The Lancet Global Health *7*, e849–e860. https://doi.org/10.1016/S2214-109X(18)30565-5.

3. Christian, P., Lee, S.E., Donahue Angel, M., Adair, L.S., Arifeen, S.E., Ashorn, P., Barros, F.C., Fall, C.H., Fawzi, W.W., Hao, W., et al. (2013). Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. International Journal of Epidemiology *42*, 1340–1355. https://doi.org/10.1093/ije/dyt109.

4. Rocha, A.S., Ribeiro-Silva, R. de C., Silva, J.F.M., Pinto, E.J., Silva, N.J., Paixao, E.S., Fiaccone, R.L., Kac, G., Rodrigues, L.C., Anderson, C., et al. (2024). Postnatal growth in small vulnerable newborns: a longitudinal study of 2 million Brazilians using routine register-based linked data. Am J Clin Nutr *119*, 444–455. https://doi.org/10.1016/j.ajcnut.2023.12.009.

5. de Mendonça, E.L.S.S., de Lima Macêna, M., Bueno, N.B., de Oliveira, A.C.M., and Mello, C.S. (2020). Premature birth, low birth weight, small for gestational age and chronic non-communicable diseases in adult life: A systematic review with meta-analysis. Early Hum Dev *149*, 105154. https://doi.org/10.1016/j.earlhumdev.2020.105154.

6. Huang, Y., Zhang, H., Xiong, Z., Lv, Y., Jia, Z., Liu, H., Xia, W., Xu, S., Chen, T., and Li, Y. (2025). Mortality and risk factors of small vulnerable newborns in 32 low- and middle-income countries. BMC Med *23*, 558. https://doi.org/10.1186/s12916-025-04406-2.

7. Romero, R., Espinoza, J., Gonçalves, L.F., Kusanovic, J.P., Friel, L., and Hassan, S. (2007). The role of inflammation and infection in preterm birth. Semin Reprod Med *25*, 21–39. https://doi.org/10.1055/s-2006-956773.

8. Battat, T.L., and Erez, O. (2023). Spontaneous Preterm Birth: a Fetal-Maternal Metabolic Imbalance. Matern Fetal Med *5*, 223–228. https://doi.org/10.1097/FM9.0000000000000205.

9. Lien, Y.-C., Zhang, Z., Barila, G., Green-Brown, A., Elovitz, M.A., and Simmons, R.A. (2020). Intrauterine Inflammation Alters the Transcriptome and Metabolome in Placenta. Front Physiol *11*, 592689. https://doi.org/10.3389/fphys.2020.592689.

10. Lauer, J.M., Duggan, C.P., Ausman, L.M., Griffiths, J.K., Webb, P., Agaba, E., Nshakira, N., Tran, H.Q., Gewirtz, A.T., and Ghosh, S. (2018). Biomarkers of maternal environmental enteric dysfunction are associated with shorter gestation and reduced length in newborn infants in Uganda. The American Journal of Clinical Nutrition *108*, 889–896. https://doi.org/10.1093/ajcn/nqy176.

11. Short Maternal Stature Increases Risk of Small-for-Gestational-Age and Preterm Births in Low- and Middle-Income Countries: Individual Participant Data Meta-Analysis and Population Attributable Fraction (2015). J Nutr *145*, 2542–2550. https://doi.org/10.3945/jn.115.216374.

12. Workicho, A., Belachew, T., Argaw, A., Roba, A., Ghosh, S., Kershaw, M., Lachat, C., and Kolsteren, P. (2020). Maternal nutritional status mediates the association between maternal age and birth outcomes. Matern Child Nutr *16*, e13015. https://doi.org/10.1111/mcn.13015.

13. Lee, A.C., Kozuki, N., Cousens, S., Stevens, G.A., Blencowe, H., Silveira, M.F., Sania, A., Rosen, H.E., Schmiegelow, C., Adair, L.S., et al. (2017). Estimates of burden and consequences of infants born small for gestational age in low and middle income countries with INTERGROWTH-21st standard: analysis of CHERG datasets. BMJ *358*, j3677. https://doi.org/10.1136/bmj.j3677.

14. Kozuki, N., Lee, A.C.C., Silveira, M.F., Sania, A., Vogel, J.P., Adair, L., Barros, F., Caulfield, L.E., Christian, P., Fawzi, W., et al. (2013). The associations of parity and maternal age with small-for-gestational-age, preterm, and neonatal and infant mortality: a meta-analysis. BMC Public Health *13 Suppl 3*, S2. https://doi.org/10.1186/1471-2458-13-S3-S2.

15. Thiruvengadam, R., Ayushi, Murugesan, D.R., Desiraju, B.K., Misra, S., Sharma, D., Subbaian, S.S., Mehta, U., Singh, A., Sharma, S., et al. (2024). Incidence of and risk factors for small vulnerable newborns in north India: a secondary analysis of a prospective pregnancy cohort. The Lancet Global Health *12*, e1261–e1277. https://doi.org/10.1016/S2214-109X(24)00212-2.

16. Gates, A., Marin, T., De Leo, G., and Stansfield, B.K. (2021). Review of Preterm Human-Milk Nutrient Composition. Nutr Clin Pract *36*, 1163–1172. https://doi.org/10.1002/ncp.10570.

17. Szyller, H., Antosz, K., Batko, J., Mytych, A., Dziedziak, M., Wrześniewska, M., Braksator, J., and Pytrus, T. (2024). Bioactive Components of Human Milk and Their Impact on Child's Health and Development, Literature Review. Nutrients *16*, 1487. https://doi.org/10.3390/nu16101487.

18. Phattraprayoon, N., Kraisonsin, N., and Kanjanapattanakul, W. (2018). Comparison of Breast Milk Compositions Among Mothers Delivering Small-for-Gestational Age, Appropriate-for-Gestational Age, and Large-for-Gestational Age Infants. Breastfeed Med *13*, 627–630. https://doi.org/10.1089/bfm.2018.0043.

19. Groer, M.W., Luciano, A.A., Dishaw, L.J., Ashmeade, T.L., Miller, E., and Gilbert, J.A. (2014). Development of the preterm infant gut microbiome: a research priority. Microbiome *2*, 38. https://doi.org/10.1186/2049-2618-2-38.

20. Unger, S., Stintzi, A., Shah, P., Mack, D., and O'Connor, D.L. (2015). Gut microbiota of the very-low-birth-weight infant. Pediatr Res *77*, 205–213. https://doi.org/10.1038/pr.2014.162.

21. Dahl, C., Stanislawski, M., Iszatt, N., Mandal, S., Lozupone, C., Clemente, J.C., Knight, R., Stigum, H., and Eggesbø, M. (2017). Gut microbiome of mothers delivering prematurely shows reduced diversity and lower relative abundance of Bifidobacterium and Streptococcus. PLoS ONE *12*, e0184336. https://doi.org/10.1371/journal.pone.0184336.

22. Chi, C., Xue, Y., Lv, N., Hao, Y., Liu, R., Wang, Y., Ding, X., Zeng, H., Li, G., Shen, Q., et al. (2019). Longitudinal Gut Bacterial Colonization and Its Influencing Factors of Low Birth Weight Infants During the First 3 Months of Life. Front. Microbiol. *10*. https://doi.org/10.3389/fmicb.2019.01105.

23. Sinha, B., Choudhary, T.S., Nitika, N., Kumar, M., Mazumder, S., Taneja, S., and Bhandari, N. (2022). Linear Growth Trajectories, Catch-up Growth, and Its Predictors Among North Indian Small-for-Gestational Age Low Birthweight Infants: A Secondary Data Analysis. Front. Nutr. *9*. https://doi.org/10.3389/fnut.2022.827589.

24. Puccio, G., Alliet, P., Cajozzo, C., Janssens, E., Corsello, G., Sprenger, N., Wernimont, S., Egli, D., Gosoniu, L., and Steenhout, P. (2017). Effects of Infant Formula With Human Milk Oligosaccharides on Growth and Morbidity: A Randomized Multicenter Trial. J Pediatr Gastroenterol Nutr *64*, 624–631. https://doi.org/10.1097/MPG.0000000000001520.

25. Brockway, M.M., Daniel, A.I., Reyes, S.M., Gauglitz, J.M., Granger, M., McDermid, J.M., Chan, D., Refvik, R., Sidhu, K.K., Musse, S., et al. (2023). Human Milk Bioactive Components and Child Growth and Body Composition in the First 2 Years: A Systematic Review. Adv Nutr, 100127. https://doi.org/10.1016/j.advnut.2023.09.015.

26. Innis, S.M. (2014). Impact of maternal diet on human milk composition and neurological development of infants123. The American Journal of Clinical Nutrition *99*, 734S-741S. https://doi.org/10.3945/ajcn.113.072595.

27. Brockway, M. (Merilee), Daniel, A.I., Reyes, S.M., Granger, M., McDermid, J.M., Chan, D., Refvik, R., Sidhu, K.K., Musse, S., Patel, P.P., et al. (2024). Human Milk Macronutrients and Child Growth and Body Composition in the First Two Years: A Systematic Review. Advances in Nutrition *15*, 100149. https://doi.org/10.1016/j.advnut.2023.100149.

28. Reyes, S.M., Brockway, M.M., McDermid, J.M., Chan, D., Granger, M., Refvik, R., Sidhu, K.K., Musse, S., Monnin, C., Lotoski, L., et al. (2024). Human Milk Micronutrients and Child Growth and Body Composition in the First 2 years: A

Systematic Review. Adv Nutr *15*, 100082.
https://doi.org/10.1016/j.advnut.2023.06.005.

29. Vanslambrouck, K., Kok, B. de, Toe, L.C., Cock, N.D., Ouedraogo, M., Dailey-Chwalibóg, T., Hanley-Cook, G., Ganaba, R., Lachat, C., Huybregts, L., et al. (2021). Effect of balanced energy-protein supplementation during pregnancy and lactation on birth outcomes and infant growth in rural Burkina Faso: study protocol for a randomised controlled trial. BMJ Open *11*, e038393.
https://doi.org/10.1136/bmjopen-2020-038393.

30. Bastos-Moreira, Y., Ouédraogo, L., De Boevre, M., Argaw, A., de Kok, B., Hanley-Cook, G.T., Deng, L., Ouédraogo, M., Compaoré, A., Tesfamariam, K., et al. (2023). A Multi-Omics and Human Biomonitoring Approach to Assessing the Effectiveness of Fortified Balanced Energy–Protein Supplementation on Maternal and Newborn Health in Burkina Faso: A Study Protocol. Nutrients *15*, 4056.
https://doi.org/10.3390/nu15184056.

31. De Kok, B., Toe, L.C., Hanley-Cook, G., Argaw, A., Ouédraogo, M., Compaoré, A., Vanslambrouck, K., Dailey-Chwalibóg, T., Ganaba, R., Kolsteren, P., et al. (2022). Prenatal fortified balanced energy-protein supplementation and birth outcomes in rural Burkina Faso: A randomized controlled efficacy trial. PLoS Med *19*, e1004002.
https://doi.org/10.1371/journal.pmed.1004002.

32. Argaw, A., De Kok, B., Toe, L.C., Hanley-Cook, G., Dailey-Chwalibóg, T., Ouédraogo, M., Compaoré, A., Vanslambrouck, K., Ganaba, R., Kolsteren, P., et al. (2023). Fortified balanced energy–protein supplementation during pregnancy and lactation and infant growth in rural Burkina Faso: A 2 × 2 factorial individually randomized controlled trial. PLoS Med *20*, e1004186.
https://doi.org/10.1371/journal.pmed.1004186.

33. Rocha, A.S., Ribeiro-Silva, R. de C., Silva, J.F.M., Pinto, E.J., Silva, N.J., Paixao, E.S., Fiaccone, R.L., Kac, G., Rodrigues, L.C., Anderson, C., et al. (2024). Postnatal growth in small vulnerable newborns: a longitudinal study of 2 million Brazilians using routine register-based linked data. Am J Clin Nutr *119*, 444–455.
https://doi.org/10.1016/j.ajcnut.2023.12.009.

34. Rahman, M.M., Abe, S.K., Kanda, M., Narita, S., Rahman, M.S., Bilano, V., Ota, E., Gilmour, S., and Shibuya, K. (2015). Maternal body mass index and risk of birth and maternal health outcomes in low- and middle-income countries: a systematic review and meta-analysis. Obes Rev *16*, 758–770.
https://doi.org/10.1111/obr.12293.

35. Girardi, G. (2018). Complement activation, a threat to pregnancy. Semin Immunopathol *40*, 103–111. https://doi.org/10.1007/s00281-017-0645-x.

36. Burwick, R.M., Java, A., and Regal, J.F. (2025). The role of complement in normal pregnancy and preeclampsia. Front. Immunol. *16*.
https://doi.org/10.3389/fimmu.2025.1643896.

37. Girardi, G., Lingo, J.J., Fleming, S.D., and Regal, J.F. (2020). Essential Role of Complement in Pregnancy: From Implantation to Parturition and Beyond. Front Immunol *11*, 1681. https://doi.org/10.3389/fimmu.2020.01681.

38. Neville, M.C., Anderson, S.M., McManaman, J.L., Badger, T.M., Bunik, M., Contractor, N., Crume, T., Dabelea, D., Donovan, S.M., Forman, N., et al. (2012). Lactation and Neonatal Nutrition: Defining and Refining the Critical Questions. J Mammary Gland Biol Neoplasia *17*, 167–188. https://doi.org/10.1007/s10911-012-9261-5.

39. Ward Platt, M., and Deshpande, S. (2005). Metabolic adaptation at birth. Seminars in Fetal and Neonatal Medicine *10*, 341–350. https://doi.org/10.1016/j.siny.2005.04.001.

40. Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arboleya Montes, S., Mancabelli, L., et al. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiol Mol Biol Rev *81*, e00036-17. https://doi.org/10.1128/MMBR.00036-17.

41. Manta-Vogli, P.D., Schulpis, K.H., Loukas, Y.L., and Dotsikas, Y. (2020). Birth weight related essential, non-essential and conditionally essential amino acid blood concentrations in 12,000 breastfed full-term infants perinatally. Scandinavian Journal of Clinical and Laboratory Investigation *80*, 571–579. https://doi.org/10.1080/00365513.2020.1818280.

42. Neu, J. (2001). Glutamine in the Fetus and Critically Ill Low Birth Weight Neonate: Metabolism and Mechanism of Action. The Journal of Nutrition *131*, 2585S-2589S. https://doi.org/10.1093/jn/131.9.2585S.

43. Zhang, Z., Adelman, A.S., Rai, D., Boettcher, J., and Lőnnerdal, B. (2013). Amino Acid Profiles in Term and Preterm Human Milk through Lactation: A Systematic Review. Nutrients *5*, 4800–4821. https://doi.org/10.3390/nu5124800.

44. Larnkjær, A., Bruun, S., Pedersen, D., Zachariassen, G., Barkholt, V., Agostoni, C., Mlgaard, C., Husby, S., and Michaelsen, K.F. (2016). Free Amino Acids in Human Milk and Associations With Maternal Anthropometry and Infant Growth. J Pediatr Gastroenterol Nutr *63*, 374–378. https://doi.org/10.1097/MPG.0000000000001195.

45. Holeček, M. (2022). Serine Metabolism in Health and Disease and as a Conditionally Essential Amino Acid. Nutrients *14*, 1987. https://doi.org/10.3390/nu14091987.

46. Korsmo, H.W., and Jiang, X. (2021). One Carbon Metabolism and Early Development: A Diet-Dependent Destiny. Trends Endocrinol Metab *32*, 579–593. https://doi.org/10.1016/j.tem.2021.05.011.

47. Masi, A.C., Embleton, N.D., Lamb, C.A., Young, G., Granger, C.L., Najera, J., Smith, D.P., Hoffman, K.L., Petrosino, J.F., Bode, L., et al. (2021). Human milk oligosaccharide DSLNT and gut microbiome in preterm infants predicts necrotising enterocolitis. https://doi.org/10.1136/gutjnl-2020-322771.

48. Autran, C.A., Kellman, B.P., Kim, J.H., Asztalos, E., Blood, A.B., Spence, E.C.H., Patel, A.L., Hou, J., Lewis, N.E., and Bode, L. (2018). Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants. Gut *67*, 1064–1070. https://doi.org/10.1136/gutjnl-2016-312819.

49. Bode, L. (2012). Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology *22*. https://doi.org/10.1093/glycob/cws074.

50. Carlson, S.E. (2001). Docosahexaenoic acid and arachidonic acid in infant development. Seminars in Neonatology *6*, 437–449. https://doi.org/10.1053/siny.2001.0093.

51. Hadley, K.B., Ryan, A.S., Forsyth, S., Gautier, S., and Salem, N. (2016). The Essentiality of Arachidonic Acid in Infant Development. Nutrients *8*, 216. https://doi.org/10.3390/nu8040216.

52. Demmelmair, H., and Koletzko, B. (2018). Lipids in human milk. Best Practice & Research Clinical Endocrinology & Metabolism *32*, 57–68. https://doi.org/10.1016/j.beem.2017.11.002.

53. Innis, S.M. (2003). Perinatal biochemistry and physiology of long-chain polyunsaturated fatty acids. The Journal of Pediatrics *143*, 1–8. https://doi.org/10.1067/S0022-3476(03)00396-2.

54. Arenas, J., Rubio, J.C., Martín, M.A., and Campos, Y. (1998). Biological roles of L-carnitine in perinatal metabolism. Early Human Development *53*, S43–S50. https://doi.org/10.1016/S0378-3782(98)00064-4.

55. Nilsson, Å., Duan, R.-D., and Ohlsson, L. (2021). Digestion and Absorption of Milk Phospholipids in Newborns and Adults. Front. Nutr. *8*. https://doi.org/10.3389/fnut.2021.724006.

56. Much, D., Brunner, S., Vollhardt, C., Schmid, D., Sedlmeier, E.-M., Brüderl, M., Heimberg, E., Bartke, N., Boehm, G., Bader, B.L., et al. (2013). Breast milk fatty acid profile in relation to infant growth and body composition: results from the INFAT study. Pediatr Res *74*, 230–237. https://doi.org/10.1038/pr.2013.82.

57. Chu, D.M., Ma, J., Prince, A.L., Antony, K.M., Seferovic, M.D., and Aagaard, K.M. (2017). Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. Nat Med *23*, 314–326. https://doi.org/10.1038/nm.4272.

58. Łoniewska, B., Fraszczyk-Tousty, M., Tousty, P., Skonieczna-Żydecka, K., Maciejewska-Markiewicz, D., and Łoniewski, I. (2023). Analysis of Fecal Short-Chain Fatty Acids (SCFAs) in Healthy Children during the First Two Years of Life: An Observational Prospective Cohort Study. Nutrients *15*, 367. https://doi.org/10.3390/nu15020367.

59. de Muinck, E.J., and Trosvik, P. (2018). Individuality and convergence of the infant gut microbiota during the first year of life. Nat Commun *9*, 2233. https://doi.org/10.1038/s41467-018-04641-7.

60. Blanton, L.V., Charbonneau, M.R., Salih, T., Barratt, M.J., Venkatesh, S., Ilkaveya, O., Subramanian, S., Manary, M.J., Trehan, I., Jorgensen, J.M., et al. (2016). Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. Science *351*, aad3311. https://doi.org/10.1126/science.aad3311.

61. Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M.A., Benezra, A., DeStefano, J., Meier, M.F., Muegge, B.D., et al. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature *510*, 417–421. https://doi.org/10.1038/nature13421.

62. Robertson, R.C., Manges, A.R., Finlay, B.B., and Prendergast, A.J. (2019). The Human Microbiome and Child Growth – First 1000 Days and Beyond. Trends in Microbiology *27*, 131–147. https://doi.org/10.1016/j.tim.2018.09.008.

63. Savino, F., Castagno, E., Calabrese, R., Viola, S., Oggero, R., and Miniero, R. (2009). High Faecal Calprotectin Levels in Healthy, Exclusively Breast-Fed Infants. Neonatology *97*, 299–304. https://doi.org/10.1159/000255161.

64. Deng, L., Taelman, S., Olm, M.R., Toe, L.C., Balini, E., Ouédraogo, L.O., Bastos-Moreira, Y., Argaw, A., Tesfamariam, K., Sonnenburg, E.D., et al. (2025). Maternal balanced energy-protein supplementation reshapes the maternal gut microbiome and enhances carbohydrate metabolism in infants: a randomized controlled trial. Nat Commun *16*, 2683. https://doi.org/10.1038/s41467-025-57838-y.

65. Deng, L., Fehr, K., Toe, L.C., Allen, L.H., Bode, L., Hampel, D., Manus, M.B., Mertens, A., Robertson, B., Yonemitsu, C., et al. (2026). Time-specific bidirectional links between the maternal microbiome, milk composition, and infant gut microbiota. Cell Host & Microbe *34*, 149-166.e5. https://doi.org/10.1016/j.chom.2025.11.014.

66. Fehr, K., Mertens, A., Shu, C.-H., Dailey-Chwalibóg, T., Shenhav, L., Allen, L.H., Beggs, M.R., Bode, L., Chooniedass, R., DeBoer, M.D., et al. (2025). Protocol: the International Milk Composition (IMiC) Consortium - a harmonized secondary analysis of human milk from four studies. Front. Nutr. *12*. https://doi.org/10.3389/fnut.2025.1548739.

67. McDonald, D., Jiang, Y., Balaban, M., Cantrell, K., Zhu, Q., Gonzalez, A., Morton, J.T., Nicolaou, G., Parks, D.H., Karst, S.M., et al. (2024). Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol *42*, 715–718. https://doi.org/10.1038/s41587-023-01845-1.

68. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 226. https://doi.org/10.1186/s40168-018-0605-2.

69. Austin, G.I., Park, H., Meydan, Y., Seeram, D., Sezin, T., Lou, Y.C., Firek, B.A., Morowitz, M.J., Banfield, J.F., Christiano, A.M., et al. (2023). Contamination source modeling with SCRuB improves cancer phenotype prediction from microbiome data. Nat Biotechnol *41*, 1820–1828. https://doi.org/10.1038/s41587-023-01696-w.

70. Azad, M.B., Robertson, B., Atakora, F., Becker, A.B., Subbarao, P., Moraes, T.J., Mandhane, P.J., Turvey, S.E., Lefebvre, D.L., Sears, M.R., et al. (2018). 'Human Milk Oligosaccharide Concentrations Are Associated with Multiple Fixed and Modifiable Maternal Characteristics, Environmental Factors, and Feeding Practices. The Journal of Nutrition *148*, 1733–1742. https://doi.org/10.1093/jn/nxy175.

71. Bode, L., Kuhn, L., Kim, H.-Y., Hsiao, L., Nissan, C., Sinkala, M., Kankasa, C., Mwiya, M., Thea, D.M., and Aldrovandi, G.M. (2012). Human milk oligosaccharide concentration and risk of postnatal transmission of HIV through breastfeeding123. The American Journal of Clinical Nutrition *96*, 831–839. https://doi.org/10.3945/ajcn.112.039503.

72. Sauer, C.W., and Kim, J.H. (2011). Human milk macronutrient analysis using point-of-care near-infrared spectrophotometry. J Perinatol *31*, 339–343. https://doi.org/10.1038/jp.2010.123.

73. Workman, J., Schumann, Bob, Eilert, Arnold, Persson, J.-A., and Gajewski, Rob (2016). A Guide to Evaluating Instrument Calibration and Performance.

74. Hampel, D., Shahab-Ferdows, S., Adair, L.S., Bentley, M.E., Flax, V.L., Jamieson, D.J., Ellington, S.R., Tegha, G., Chasela, C.S., Kamwendo, D., et al. (2016). Thiamin and Riboflavin in Human Milk: Effects of Lipid-Based Nutrient Supplementation and Stage of Lactation on Vitamer Secretion and Contributions to Total Vitamin Content. PLoS One *11*, e0149479. https://doi.org/10.1371/journal.pone.0149479.

75. Hampel, D., Shahab-Ferdows, S., Domek, J.M., Siddiqua, T., Raqib, R., and Allen, L.H. (2014). Competitive chemiluminescent enzyme immunoassay for vitamin B12 analysis in human milk. Food Chemistry *153*, 60–65. https://doi.org/10.1016/j.foodchem.2013.12.033.

76. Hampel, D., York, E.R., and Allen, L.H. (2012). Ultra-performance liquid chromatography tandem mass-spectrometry (UPLC–MS/MS) for the rapid, simultaneous analysis of thiamin, riboflavin, flavin adenine dinucleotide, nicotinamide and pyridoxal in human milk. Journal of Chromatography B *903*, 7–13. https://doi.org/10.1016/j.jchromb.2012.06.024.

77. Turner, T., and Burri, B.J. (2012). Rapid Isocratic HPLC Method and Sample Extraction Procedures for Measuring Carotenoid, Retinoid, and Tocopherol Concentrations in Human Blood and Breast Milk for Intervention Studies. Chromatographia *75*, 241–252. https://doi.org/10.1007/s10337-012-2193-9.

78. Hampel, D., Islam, M.M., Shahab-Ferdows, S., and Allen, L.H. (2024). Human milk macro- and trace-elements: Simultaneous analysis in sub-milliliter amounts by ICP-

MS and application to assessing acute supplementation effects. Heliyon *10*. https://doi.org/10.1016/j.heliyon.2024.e34709.

79. Allen, L.H., Hampel, D., Shahab-Ferdows, S., Andersson, M., Barros, E., Doel, A.M., Eriksen, K.G., Christensen, S.H., Islam, M., Kac, G., et al. (2021). The Mothers, Infants, and Lactation Quality (MILQ) Study: A Multi-Center Collaboration. Curr Dev Nutr *5*, nzab116. https://doi.org/10.1093/cdn/nzab116.

80. Hampel, D., Shahab-Ferdows, S., Kac, G., and Allen, L.H. (2025). Evaluating Metabolic Profiling of Human Milk Using Biocrates MxP® QUANT 500 Assay. Metabolites *15*, 14. https://doi.org/10.3390/metabo15010014.

81. Li, L., Chen, Y., and Zhu, J.-J. (2017). Recent Advances in Electrochemiluminescence Analysis. Anal. Chem. *89*, 358–371. https://doi.org/10.1021/acs.analchem.6b04675.

82. Hampel, D., Shahab-Ferdows, S., Hossain, M., Islam, M.M., Ahmed, T., and Allen, L.H. (2019). Validation and Application of Biocrates AbsoluteIDQ® p180 Targeted Metabolomics Kit Using Human Milk. Nutrients *11*, 1733. https://doi.org/10.3390/nu11081733.

83. Olm, M.R., Dahan, D., Carter, M.M., Merrill, B.D., Yu, F.B., Jain, S., Meng, X., Tripathi, S., Wastyk, H., Neff, N., et al. (2022). Robust variation in infant gut microbiome assembly across a spectrum of lifestyles. Science *376*, 1220–1223. https://doi.org/10.1126/science.abj2972.

84. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics *36*, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

85. Olm, M., Carter, M., Merrill, B., and Sonnenburg, J. (2023). Data supporting publication: Ultra-deep Sequencing of Hadza Hunter-Gatherers Recovers Vanishing Gut Microbes. (Zenodo). https://doi.org/10.5281/zenodo.8072245 https://doi.org/10.5281/zenodo.8072245.

86. Olm, M.R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B.A., Morowitz, M.J., and Banfield, J.F. (2021). inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat Biotechnol *39*, 727–736. https://doi.org/10.1038/s41587-020-00797-0.

87. Han, S., Guiberson, E.R., Li, Y., and Sonnenburg, J.L. (2024). High-throughput identification of gut microbiome-dependent metabolites. Nat Protoc *19*, 2180–2205. https://doi.org/10.1038/s41596-024-00980-6.

88. Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., and Arita, M. (2015). MS-DIAL: Data Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. Nat Methods *12*, 523–526. https://doi.org/10.1038/nmeth.3393.

89. Gonzalez, C.G., Wastyk, H.C., Topf, M., Gardner, C.D., Sonnenburg, J.L., and Elias, J.E. (2020). High-Throughput Stool Metaproteomics: Method and Application to Human Specimens. mSystems *5*, e00200-20. https://doi.org/10.1128/mSystems.00200-20.

90. Hsiao, Y., Zhang, H., Li, G.X., Deng, Y., Yu, F., Valipour Kahrood, H., Steele, J.R., Schittenhelm, R.B., and Nesvizhskii, A.I. (2024). Analysis and Visualization of Quantitative Proteomics Data Using FragPipe-Analyst. J. Proteome Res. *23*, 4303–4315. https://doi.org/10.1021/acs.jproteome.4c00294.

91. Breuza, L., Poux, S., Estreicher, A., Famiglietti, M.L., Magrane, M., Tognolli, M., Bridge, A., Baratin, D., and Redaschi, N. (2016). The UniProtKB guide to the human proteome. Database (Oxford) *2016*, bav120. https://doi.org/10.1093/database/bav120.

92. Kohler, D., Staniak, M., Tsai, T.-H., Huang, T., Shulman, N., Bernhardt, O.M., MacLean, B.X., Nesvizhskii, A.I., Reiter, L., Sabido, E., et al. (2023). MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-Based Proteomic Experiments with Chromatography-Based Quantification at Scale. J. Proteome Res. *22*, 1466–1482. https://doi.org/10.1021/acs.jproteome.2c00834.

93. Toe, L.C., Kerckhof, F.-M., De Bodt, J., Morel, F.B., Ouedraogo, J.-B., Kolsteren, P., and Van de Wiele, T. (2020). A prebiotic-enhanced lipid-based nutrient supplement (LNSp) increases Bifidobacterium relative abundance and enhances short-chain fatty acid production in simulated colonic microbiota from undernourished infants. FEMS Microbiol Ecol *96*, fiaa105. https://doi.org/10.1093/femsec/fiaa105.

94. Whitehead, S.J., French, J., Brookes, M.J., Ford, C., and Gama, R. (2013). Between-assay variability of faecal calprotectin enzyme-linked immunosorbent assay kits. Ann Clin Biochem *50*, 53–61. https://doi.org/10.1258/acb.2012.011272.

95. Ardle, A.M., Binek, A., Moradian, A., Orgel, B.C., Rivas, A., Washington, K.E., Phebus, C., Manalo, D.-M., Go, J., Venkatraman, V., et al. (2022). Standardized Workflow for Precise Mid- and High-Throughput Proteomics of Blood Biofluids. Clin Chem *68*, 450–460. https://doi.org/10.1093/clinchem/hvab202.

96. McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE *8*, e61217. https://doi.org/10.1371/journal.pone.0061217.

97. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., et al. (2001). vegan: Community Ecology Package. https://doi.org/10.32614/CRAN.package.vegan https://doi.org/10.32614/CRAN.package.vegan.

98. Pinheiro, J., R Core Team, and Bates, D. (2025). nlme: Linear and Nonlinear Mixed Effects Models. (The R Foundation).

99. Lin, H., and Peddada, S.D. (2020). Analysis of compositions of microbiomes with bias correction. Nat Commun *11*, 3514. https://doi.org/10.1038/s41467-020-17041-7.

100. Huang, R., Soneson, C., Germain, P.-L., Schmidt, T.S.B., Mering, C.V., and Robinson, M.D. (2021). treeclimbR pinpoints the data-dependent resolution of hierarchical hypotheses. Genome Biology *22*, 157. https://doi.org/10.1186/s13059-021-02368-1.

101. Chen, Y., Chen, L., Lun, A.T.L., Baldoni, P.L., and Smyth, G.K. (2025). edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Nucleic Acids Research *53*, gkaf018. https://doi.org/10.1093/nar/gkaf018.

102. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

103. Li, S., Park, Y., Duraisingham, S., Strobel, F.H., Khan, N., Soltow, Q.A., Jones, D.P., and Pulendran, B. (2013). Predicting Network Activity from High Throughput Metabolomics. PLoS Comput Biol *9*, e1003123. https://doi.org/10.1371/journal.pcbi.1003123.

104. Rohart, F., Gautier, B., Singh, A., and Cao, K.-A.L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Computational Biology *13*, e1005752. https://doi.org/10.1371/journal.pcbi.1005752.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplmentaryData1.xlsx
- SupplmentaryData2.xlsx
- Deng.et.alSupplementaryInformationNatureMetabolism.pdf
- SupplmentaryData3.xlsx
- SupplmentaryData4.xlsx