Linking the Gut Microbiome to Neurocognitive Development in Bangladesh Malnourished Infants

Theo Portlock[1,*], Talat Sharma[2,*], Shahria Hafiz Kakon[2,*], Berit Hartjen[3,*], Chris Pook[1,*], Brooke Wilson[1,*], Ayisha Bhuttor[3], Daniel Ho[1], Inoli Shennon Wadumesthrige Don[1], Anne-Michelle Engelstad[3], Renata Di Lorenzo[3], Garrett Greaves[3], Caroline Kelsey[3], Peter Gluckman[1], Justin O'Sullivan[1,4,5,6], Terrence Forrester[7], and Charles Nelson[3]

[1]The Liggins Institute, University of Auckland, NZ

[2]Infectious Diseases Division, International Centre for Diarrheal Disease Research, Bangladesh

[3]Department of Pediatrics, Boston Children's Hospital and Harvard Medical School; Harvard Graduate School of Education, Boston, USA

[4]The Maurice Wilkins Centre, The University of Auckland, New Zealand

[5]MRC Lifecourse Epidemiology Unit, University of Southampton, University Road, Southampton, UK

[6]Singapore Institute for Clinical Sciences, Agency for Science Technology and Research, Singapore

[7]Faculty of Medical Sciences, UWI Solutions for Developing Countries, The University of the West Indies (UWI), Jamaica

[*]These authors contributed equally

Key points:

- The gut microbiome of malnourished infants is compositionally distinct from well-nourished infants, characterised by a lower shannon diversity, higher *Prevotella*-to-*Bacteroides* ratio, and lower potential anaerobic pathways involved in the fermentation of pyruvate.
- Depletion of plasma lipids critical for brain development were negatively correlated with gut microbiome pathways, EEG power spectral density, and cognitive outcomes.
- There was a high level of commonality in the shared features between malnutrition and low expressive communication.

Abstract

Malnutrition, affecting approximately 30 million infants annually, has profound immediate and enduring repercussions, with nearly half of child deaths under five linked to malnutrition. Survivors face lasting consequences, including impaired neurocognitive development, leading to cognitive and behavioural deficits, impacting academic performance and socioeconomic outcomes. Despite extensive literature on malnutrition's mechanisms spanning nutrition, infection, metabolism, microbiome, and genomics, knowledge gaps persist. This study employs an interpreted random forest network approach to identify non-overlapping connections between the gut microbiome, plasma lipids, and EEG data, from infants with Moderate Acute Malnutrition (MAM) and well-nourished controls. *Bacteroides fragilis* abundance, linked to fermentation pathways, emerges as a predictive factor for well-nourished infants. In conclusion, network analysis highlights the potential significance of targeted interventions in addressing both the short and long-term impacts of malnutrition.

Key words

Malnutrition, Gut microbiome, Neurocognitive development, Plasma lipidome, Random Forest classification models

Main

Malnutrition is a significant global health issue responsible for an estimated 45% of all child deaths worldwide, making it the leading cause of mortality among children under the age of five [**?**]. Malnutrition is characterised by delayed growth, proportionate reductions in mass of most organs and tissues, and alterations in tissue architecture [**?**]. Children who survive malnutrition are likely to suffer long-term consequences including impaired neurocognitive development, leading to long-term deficits in cognition and behaviour [**?**]. This consequently leads to poor school performance and economic prospects as an adult [**?**]. While much is known about the health, social, and economic ramifications of malnutrition, significant gaps in our knowledge remain. One crucial gap is the contribution of the gut microbiome to the pathology of malnutrition in addition to its impact on brain and cognitive development.

The human gut microbiome is a complex ecosystem comprised of the microorganisms lining the intestinal tract, including bacteria, viruses, fungi, and archaea. Infancy represents a sensitive period in gut microbiome formation as the gut microbiome changes drastically over this time [**?**, **?**]. Importantly, many aspects of malnutrition including host nutritional status, dietary intake, antibiotic administration, and infections impact the diversity, composition, and functionality of the microbiome [**?**, **?**]. To this end, several studies in low- and middle- income countries Low- and Middle- Income Countries (LMIC) have shown differences in gut microbiome profiles between malnourished and well-nourished infants [**?**, **?**]. For example, a study in Bangladesh found that malnourished infants, compared to well-nourished infants, had higher abundances of *Bifidobacterium* and *Escherichia* species [**?**]. Beyond the correlational and descriptive evidence presented, work using mouse models point to a possible causal role of the gut microbiome in growth and weight gain, as mice colonized using fecal microbial transplantation with samples from malnourished children, but not well-nourished controls, showed impairments in weight gain and growth [**?**]. Critically, perturbations of the gut microbiome associated with malnutrition may have downstream

consequences for brain and cognitive development [?, ?].

Malnutrition, like the gut microbiome, is associated with neurocognitive impairments thought to result from structural and functional changes to the brain [?, ?, ?, ?, ?, ?]. More specifically, several studies conducted in healthy infants living in upper-middle-income countries have shown that the gut microbiome is associated with cognitive and brain development; although the directionality remains unclear with both increased and decreased gut microbiota alpha diversity being linked to positive cognitive outcomes and neural development [?, ?, ?, ?]. Previous research has suggested that malnutrition may be associated with alterations in the gut microbiome, including changes in the composition and diversity of the microbial community [?]. Moreover, alterations in the gut microbiome may contribute to negative neurological outcomes observed in malnourished infants, potentially through the disruption of nutrient absorption or the generation of toxic metabolites [?]. Very few studies have examined the link between the gut microbiome and cognition in malnourished children. One notable exception is a randomized control trial of nutrition, stimulation, and hygiene education in a group of rural Ugandan mothers and their infants who were moderately stunted (height-for-age Z-score between -2 and -3 SD). Across a series of studies conducted from 2 years to 3 years of age there were mixed findings with some species such as *Bifidobacterium longum* found to associate with language impairment assessed using the Bayley Scales of Infant and Toddler Development and other developmental assessments but at other time points no associations were found [?, ?, ?]. Therefore, more work is needed to understand how the gut microbiome mediates that association between malnourishment and cognitive development.

Another mechanism by which brain and behavioural development may be impacted by malnutrition is through the circulating plasma lipidome [?, ?]. Several circulating plasma lipids including cholesterol, phosphatidylcholines, phosphatidylethanolamine, and sphingolipids compromise 50% of the dry weight of the brain and have unique roles in neurological structure and function [?]. The brain relies upon nutrients circulating in the blood for its supply of resources. Moreover, the blood brain barrier which plays a crucial role in regulating which circulating lipids enter and exit the brain area is impaired by malnutrition [?]. Circulating plasma lipids represent a means of communication between the gut microbiome and the brain [?] and therefore represent a potential mechanism of influence.

Given the importance of the composition and functions of the gut microbiome in maintaining

overall health, there has been increasing interest in understanding how its alterations may contribute to malnutrition and its associated impacts on infant neurocognitive development. The present study examines the impact of malnutrition on the composition of the infant gut microbiome, plasma lipidome, neural activity, and cognitive outcomes in a cross-sectional cohort of well-nourished and malnourished 12-month-old Bangladeshi infants. Random forest models were used to integrate deeply phenotyped multi-modal data and identify correlations that provide putative mechanistic insights into developmental delays the result from malnutrition. Overall, this study provides important information about gut-blood-brain-behaviour links in infants impacted by malnutrition.

Results

Study population characteristics

As a city with the second highest density of population and in a country with childhood malnutrition rate is one of the highest globally, the Mirpur region in Dhaka, Bangladesh was chosen to assess the impact of early-life malnutrition [**?**]. 156 infants with Moderate Acute Malnutrition (MAM) and 74 well-nourished controls at 12 months of age with no history of chronic medical conditions, and no antibiotic use within the past month were recruited from this region (Figure 1a). MAM was defined according to WHO guidelines, using a threshold between two and three standard deviations below the mean z-score for weight-for-length/height (WLZ/WHZ) [**?**]. Confounding variables to measures of MAM (WLZ/WHZ, Mid-upper arm circumference (MUAC), Weight, and Head Circumference (HC)) were measured using Fisher's exact test for categorical variables and Mann-Whitney U test (MWU) for continuous variables (Table 1). Confounding variables included the principal toilet used (Septic-tank/toilet), water treatment method (Boil), toilet facility (shared with other households), length of time lived in current household, mother's income, years of father education, father's education level, monthly total expenditure, and mother's occupation (housewife).

Malnutrition is associated with an elevated gut microbial *Prevotella*-to-*Bacteroides* (P/B) ratio and reduced pyruvate fermentation potential

Nutrition is one of the leading confounding factors that explain the variance of the gut microbiome composition [**?**]. Consequently, the impact that malnutrition imparts on the infant gut microbiome was measured in this cohort. It was hypothesised that malnutrition impacts the diversity and composition of the infant gut microbiome in this cohort. Stool metagenomes were extracted, analysed (shotgun metagenomic sequencing, 40.53 ± 8.5 million reads with no significant difference between MAM and well-nourished ($p = 0.71$)) and profiled according to their species and functional compositions. Across all samples, 3 kingdoms, 17 phyla, 31 classes, 51 orders, 100 families, 226 genera, 749 species, 611 functional pathways, and 2,828,874 gene families were detected.

There was a mean species richness of 50.3 ± 16.4 per sample and mean shannon diversity of 2.96 ± 0.72, commensurate with other infants at that age group [**?**]. Malnutrition was associated with a lower Shannon diversity ($p = 0.025$) and Pielou's evenness ($p = 0.009$) than their well-nourished counterparts (Figure 1, **??**); a result that has been oberved previously in other cohorts of malnourished infants [**?**]. These differences in alpha diveristy were underscored by a significant difference in the Bray-Curtis dissimilarity between the nutritional groups (PERmutational Multivariate ANalysis Of VAriance (PERMANOVA), $R^2 = 2.22$, $p = 0.008$), as a consequence of the differential abundance of 6/350 species (1.7%) (Figure 1b, Table 2). Malnourished infant gut microbiomes had a greater prevalence and abundance of five species including *Prevotella copri* (Log_2(MAM/well-nourished) = 0.64, $p = 0.020$, $q = 0.490$) and *Streptococcus salivarius* (Log_2(MAM/well-nourished) = 2.39, $p = 0.0005$, $q = 0.032$) in microbiomes from MAM infants, compared to well-nourished controls (Figure 1c). Enrichment in these species was associated with the depletion and reduction in the prevalence of the sphingolipid-producing species *Bacteroides fragilis* (Log_2(MAM/well-nourished) = -0.62, $p = 0.021$, $q = 0.49$). This reciprocal relationship was described as an increase to the P/B ratio of the MAM infants (Log_2(MAM/well-nourished) = 2.80, $p = 0.05$) (Figure 1f).

Functional pathway analyses revealed no significant differences in the composition of the overall functionome between MAM and well-nourished controls (PERMANOVA, $R^2 = 8.76$, $p = 0.365$). After false discovery rate adjustment there were no significant differences in the pathway relative abundances (Table 3). However, 28/352 (27 and 1 elevated/depleted in MAM respectively) path-

ways were differentially abundant using MWU without FDR adjustment between the conditions, and a total of 94/352 were approaching significance ($p < 0.1$) (44 and 6 elevated/depleted in MAM respectively)(Figure 1g). Specifically, MAM gut microbiomes had an enrichment of multiple pathways involved in branch chain amino acid biosynthesis (eg. BCAA biosynthesis superpathway Log_2(MAM/well-nourished) = 0.12, p = 7e-4) including L-valine and L-isoleucine (I, III)) and sucrose/glucose degradation (anaglycolysis III (Log_2(MAM/well-nourished) = 0.11, $p = 0.003$)). Conversely, there was a decrease in the relative abundance of threonine metabolism pathways (Log_2(MAM/well-nourished) = -0.27, $p = 0.05$) and pyruvate fermentation pathways to the Short Chain Fatty Acid (SCFA) propionate (Log_2(MAM/well-nourished) = -0.18, $p = 0.08$) within the MAM infant's gut microbiome. Interestingly, fermentation pathways on the whole were increased in the MAM gut microbiome ($p = 0.09$).

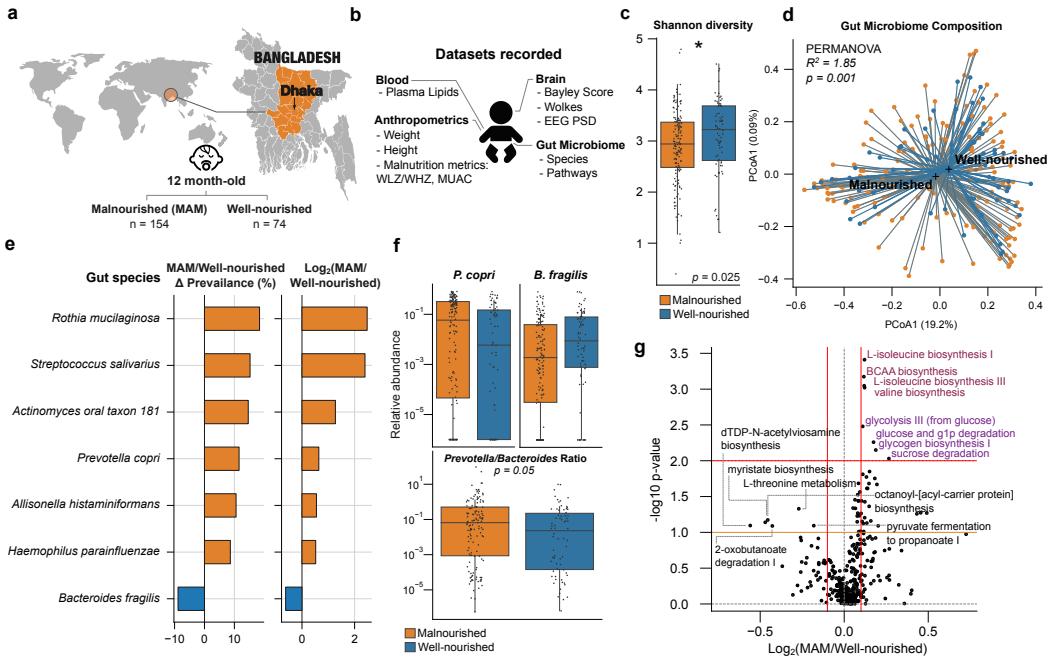

Figure 1: Malnutrition impacts the 12-month-old infant gut microbiome. a) Schematic of study design. b) Summary of data collected. c) Change in diversity of the gut microbiome associated with malnutrition. d) PCoA Scatterplot of Bray-Curtis beta diversities of samples (each marker is a single infant sample). e) Barplot of significant taxonomic differences in relative abundance and prevailance between 12-month-old well-nourished and MAM samples ($p < 0.05$). f) Boxplot of P/B ratio change between study conditions. g) Volcano plot of pathways that associate with malnutrition (red and orange horizontal line signifies $q < 0.05$ and 0.01 respectively. left and right vertical lines represent Log_2(MAM/well-nourished) of -0.1 and 0.1 respectively).

Malnutrition impacts brain activity and expressive communication

Malnourished children often present with long-term impairments in neural and cognitive development [**?**]. The Bayley Scales of Infant and Toddler Development Fourth Edition (BSID-IV; [**?**]) was used to assess development in the cohort. When compared to well-nourished infants, there was a significant reduction in the Expressive Communication, Fine Motor, and Gross Motor scores in the MAM infants (mean difference(MAM - well-nourished) = -2.02, -1.68, -2.69, p = 0.0036, 0.0005, 0.0082) (Figure 2b, Table 4). Expressive communication is a measure of how well a child communicates with others (). There was a reduction in all other Bayley metrics including the receptive language, cognitive, motor abilities, but without significance. To complement this method of assessing development, wolkes scoring was performed for the cohort (Figure 2c, Table 5). As with the Bayley scoring, vocalisation scores were lower in MAM infants (mean difference(MAM - well-nourished) = -1.47, p = 2.05e-10) amongst Activity and Approach scores.

Resting state electroencephalography Electroencephalography (EEG) assessments of participants were performed to enable investigation of the impacts of malnutrition on brain activity. After exploratory comparisons of EEG power spectral density (PSD) between infants with MAM and the well-nourished controls, focus on the high-alpha (9-12 Hz), beta (12-30 Hz) and gamma (30-45 Hz) frequency bands distributed across occipital, temporal and frontal regions of interest will be made. These bands are generally associated with concentration, alertness, and higher mental activity and were observed to have higher amplitudes in the well-nourished infants compared to infants with MAM (Table 6).

Malnutrition is associated with a reduction in circulating odd-chain fatty acids and ceramides

Adequate nutrition in infants is characterised by healthy circulating concentrations of metabolites, including lipids, involved in growth and development [**?**]. Delivery of lipids from the gut to the rest of the body is a crucial process during the developmental window. Therefore, discovery LC-MS/MS was used to assign and quantify the levels of 825 plasma lipids in the infants of the cohort (Figure 3).

Malnutrition was associated with major changes (286/825 - 35%) to the plasma lipidome. Of these

changes, 124 (15%) plasma lipidome compounds increased and 162 (20%) decreased in concentration (Figure 3, Table 7). Enrichment in the abundance of three lipid classes with diverse functions was observed, including two that are known to be specific to neurological development and function (ie. the long chain ceramide Cer 31:5;O2 (Log_2(MAM/well-nourished) = 0.36, q = 2.12e-7) and the lactosylceramide hex2cer 34:1 (Log_2(MAM/well-nourished) = 0.018, q = 0.004)). By contrast, long chain sphingomyelins (SM 44:3;O2, Log_2(MAM/well-nourished) = FIND, q = 5.81e-5)) and others were observed to increase in relative concentration in malnourished infants. Several lysophospholipids from the lysophosphatidylcholine (LPC), and lysophosphatidylethanolamine (LPE) classes were enriched in well-nourished infant plasma.

Multimodal Random Forest classification of malnutrition reveal cross mode influences

Having established the existence of changes associated with malnutrition across the gut microbiome, brain, and plasma lipids, the relative importance of changes in each of these domains for the prediction of malnutrition was measured. Individual and multimodal Random Forest classifiers were trained, using gut microbiome taxonomic and functional neuroimaging (EEG), lipidome and behavioural data (Bayley scale scores), to predict malnutrition in the cohort (Figure 4).

Within the predictors trained on individual feature sets, plasma lipids (AUCROC=1.00, oob=1.00) were the best predictor of malnutrition in 1-year-old infants, followed by brain/behavioural metrics (i.e., EEG, and Bayley AUCROC=0.83, oob=0.64), and the gut microbiome taxonomic and functional profiles (AUCROC=0.59, oob=0.59).

Ensemble models were trained on individually scaled and combined data from the gut microbiome taxonomic and functional profiles, neuroimaging (EEG), plasma lipidome, and Bayley scale and Wolkes scores and evaluated using AUCROC and 10-fold cross validation. The ensemble models had an AUCROC of 0.82.

SHapley Additive exPlanations (SHAP) scoring interpretation was performed to understand the workings of these models and importance of the features without the assumption of linearity of relationship (Figure 4d, Table 8). Those features that changed significantly were more likely to have high importance for the model prediction as evident with a spearman correlation between

SHAP score and MWU -$\log_2(p)$ of 0.74. Comparison with the individual models indicated that inclusion of the other datasets into the ensemble models lead to the identification of non-linear features that contributed to the predictive power of the microbial species within the classification model. For example, these included MAM depleted *Faecalibacerium prausnitzii* (SHAP = 0.0076), and *Odoribacter splanchnicus* (SHAP = 0.0063) or MAM enriched *Bifidobacterium breve* (SHAP = 0.0074), and *Haemophilus parainfluenzae* (SHAP = 0.0065).

A Multimodal Predictive Network Analysis reveals the importance of *Bacteroides fragilis* in infant neurocognitive development

Network analysis is a useful tool to understand complex systems that emerge from interactions between multiple components. To better understand the complexities of the important features and correlations between EEG PSDs, behavior, gut microbial species and functions, and plasma lipids, their architecture was mapped out using co-abundant network analysis (Figure 5). Spearman correlation of the features that were important in predicting malnutrition was calculated, and filtered by significance ($q = <0.05$) (1052/3906 correlations, Table 9). Finally, only edges with an absolute spearman rho of > 0.2 were used to construct the network.

Important features (ie. mean absolute SHAP score > 0.002) were more likely to be significantly correlated ($q < 0.05$) with one another and had greater measures of Betweenness Centrality (Figure 5, Supplementary Table X) than unimportant features (mean absolute SHAP < 0.002). Plasma lipids that were enriched/depleted in the MAM condition (Supplementary Table X) were positively correlated with the anthropometric measures WLZ/WHZ, MUAC, and weight. Unsurprisingly, cluster analyses revealed that those features which were different between MAM and well-nourished were positively correlated with each other (i.e., change in the same direction, supp table). A subcluster of *B. fragilis*, pyruvate fermentation pathways, plasma ceramides, EEG PSD and Expressive Communication was identified that was highly correlated with the well-nourished state (Figure 5). Those plasma lipids that were depleted ($q < 0.05$, Log2(MAM/well-nourished) < 0) from the MAM infant samples were also positively correlated with EEG PSD amplitudes. Notably, EEG metrics were also correlated with bacterial pyruvate fermentation pathways and *B. fragilis* relative abundance. Conversely, a correlated subcluster of *P. copri*, glycolysis, peptidoglycan biosynthesis, BCAA pathways, and plasma sphingomyelins that were associated with the MAM condition was

identified.

Discussion

A central goal of this study was to obtain a better understanding of how disturbances in host-microbiome interactions impact neurocognitive development in malnutrition. It was observed that malnutrition was characterised by a higher P/B ratio and reduced pyruvate fermentation potential in the gut. *Prevotella* rich microbiomes (previously referred to as the Prevotella enterotype) have been typically understudied due to their underrepresentation in non LMIC [**?**]. This ratio has previously been implicated in diet and lifestyle, particularly in adults [**?**] and *Bacteroides* have been observed previously to be depleted in Bangladesh infants [**?**]. Other studies of malnutrition have shown similar reductions in alpha diversity [**?**]. Accelerated ageing of the gut microbiome, as indicated by the presence of specific markers such as *P. copri* and *Bifidobacterium adolescentis*, is one possible hypothesis for the differential *Prevotella* abundance. This could indicate that the development of the microbiome-gut-brain axis is pathologically accelerated during infant malnutrition. Alternatively, selective microbiome community driven interactions might explain the inverse correlations that were observed between *P. copri* and *Bifidobacterium longum* and *B. breve*. *B. longum* and other anaerobic species have been previously linked to moderate and severe acute malnutrition in Bangladesh [**?**, **?**].

Comparisons of the MAM and control infants identified deficits both in neural activity and expressive communication that were associated with the malnourished condition. When investigating differences in neural activity, disruptions were evident for higher frequency power bands (alpha, beta, and gamma) but not lower frequency bands (delta and theta) in frontal, temporal and occipital areas. Therefore, changes in expressive language are early and readily assessed indicators of long-term developmental consequences associated with MAM.

The plasma lipidomes of malnourished children were substantially different from those of controls, with significant differences in the levels of ceramides and lysolipids (i.e. lipid derivatives in which one or both acyl derivatives have been removed by hydrolyisis). Numerous specific changes stand out as being potentially important for neural development. Firstly, lactosylceramide (hex2cer 34:1) is an essential precursor for synthesis of all complex glycosphingolipids [**?**] that was depleted

by around 50% in malnourished infants. Secondly, lysophosphatidylcholine (LPC) and lysophosphatidylethanolaine (LPE) are essential for brain development and growth as they carry fatty acid across the blood-brain barrier, via the major facilitator superfamily domain-containing protein 2A (Mfsd2a) [**?**].

Phosphatidylcholine (PC) is a precursor to acetylcholine, an essential neurotransmitter for memory and cognitive function. Supplementing neuron differentiation medium with phosphatidylcholine reduces the impact of inflammatory stress and neuronal damage, increasing the numbers of healthy neurons and modulating neuronal plasticity [**?**].

Propranoate is a short chaing fatty acid that has been demonstrated to be neuroprotective and induce neuroregeneration in the peripheral nervous system during inflammation induced neuropathy. It is also an Odd Chain Fatty Acid (OCFA); a class which was observed to be reduced in malnutrition. OCFAs are unable to be synthesised by mammals and instead their presence in the circulation can be attributed to either change in diet or gut microbiome [**?**]. Conversely, higher serum levels of propanoate have been associated with increased odds of cognitive decline in a cohort of French individuals above 65 years of age [**?**, **?**]. Propanoate is a key precursor in lipid biosynthesis and can be metabolised to propionyl-CoA, which can subsequently be incorporated into sphingolipid biosynthesis pathways [**?**]. It remains possible that this is due to extensive metabolic and microbial programming during this period [**?**]. The other class of lipids that was depleted from the MAM infants was ceramides. Sphingomyelinases (SMases) hydrolyse sphingomyelin, releasing ceramide and creating a cascade of bioactive lipids. These ceramides and their conversion to gangliosides have been shown previously to be important for myelin sheath development and so a depletion in this area may impact brain maturation.

Random Forest classification models trained on the gut microbiome, neuroimaging data, and the plasma lipidome accurately predicted the malnutrition condition. Integrating the important features of these models and spearman correlation using network analysis provided a holistic view of the malnutrition mechanism and highlighs the potential importance of *B. fragilis* as a keystone species for infant neurocognitive and microbiome-gut-brain axis development.

Recent studies have emphasised the significant role of the gut microbiome in mediating dietary effects on host physiology, in addition to its influence on the development and function of the nervous system [**?**, **?**, **?**, **?**]. Multiomics analysis examined associations between infant malnutri-

tion, altered brain function, and the infant microbiome and revealed a mechanism that links the fermentation of pyruvate to butanoate and ceramide biosynthesis to brain function and language development. However, in the absence of causal animal studies, it remains unclear if the gut microbiome changes are a result of, or contribute causally to the wider malnutrition phenotype.

Conclusion

Collectively, integrative multi-omic study highlights associations between the gut microbiome, plasma lipids, brain connectivity, and cognitive function. This evidence may inform the development of meaningful, targeted and effective interventions for infants experiencing malnutrition.

Methods

Ethics

The M4EFaD intervention was registered NCT05629624 on clinicaltrials.gov. The study was approved by icddr,b Ethical Review Committee PR-21084 and the Bangladesh Directorate General of Drug Administration. Ethical review for the analytical component was obtained from Auckland Health Research Ethics Committee approval AH23922 (metabolomics, metagenomics, machine learning).

Study Design and Participants

The study was performed on the baseline data from three cohorts of infants who were enrolled (between Jan – December 2022) as part of the M4EFaD intervention within the Mirpur slum, Dhaka, Bangladesh. The cohort consisted of: a control group of 73 well-nourished children at 12 $\pm$ 1 months (WLZ z-score $>$ -1 SD); an intervention group of 156 children with WLZ $<$ -2 and $>$ -3 z-score, and/or MUAC $<$ 12.5 and $>$ 11.5 cm having MAM at 12 $\pm$ 1 months; and an outcome reference group of 73 children with WHZ $<$ -2 and $>$ -3 z-score, and/or MUAC $<$ 12.5 and $>$ 11.5 cm having stable MAM at 3 years $\pm$ 2m. Inclusion criteria included a diagnosis of malnutrition,

no history of chronic medical conditions, and no antibiotic use within the past month. The study protocol has been submitted for publication and is available on MedRxiv [**?**].

Recruitment and anthropometric data collection

Enrolment was initiated on February 7, 2022, and will continue until February 2024. Study surveillance workers (SWs) conducted a door-to-door census (approximately 100,000 households) in Mirpur DNCC wards ward 2, 3 and 5 between January and December 2022. Verbal consent was obtained to participate in the census. The census identified 5736 children aged between 11 to 13 months and 2,314 children aged between 34 to 38 months. During the census, if the guardian verbally consented to the study procedure, and the babies met the inclusion and exclusion criteria of the study (Table 1), the SWs proceeded to measure the MUAC of the child. Mothers of babies who were within the MUAC range were invited to visit the icddr,b study clinic for further assessment and enrolment.

Final screening for eligibility and study consent occurred at the icddr,b Mirpur study clinic. The consenting process was tailored to each mother's literacy level and involved reviewing the inclusion and exclusion criteria. Comprehension of the study was assessed using scripted points and open-ended questions.

Following consent, the clinical screening team completed a screening form, capturing the date of enrolment, sex, date of birth (DOB), weight (in kg), length/ height (in cm), head circumference (in cm), and Mid-Upper Arm Circumference MUAC measurements of the child. The WLZ/WHZ Z-score for each child was calculated using the WHO anthropometric calculator. The child's age was validated using the EPI vaccination card. Neurological measures, Bailey scores, EEG data were collected upon enrolment to evaluate neurological development.

EEG data collection and analysis

Continuous scalp EEG was recorded using NetStation 4.5.4. and 128-channel Hydrocel Geodesic Sensor Nets modified to remove eye electrodes (Electrical Geodesics, Inc. (EGI), Eugene, OR, USA). Data was sampled at 500 Hz. Impedances were kept under 100 k ω when possible and

measured once at the beginning of the session, and again halfway through. Sessions were conducted in a dimly lit room with the participants sitting on the parent's lap. The participants were separated from the research staff conducting the session by a curtain, but the testing area was not acoustically or electrically shielded. A second research staff member was present in the testing area to help keep the participant engaged. EEG sessions consisted of 6 paradigms, i.e., resting state, visual working memory, flanker, disengagement, visual evoked potential, and auditory stimuli. The subsequent (pre-)processing steps were applied to the resting state data where participants watched a 3-minute video that featured toys.

EEG data were preprocessed offline with MatLab (R2021B) using the Harvard Automated Processing Pipeline for Electroencephalography (HAPPE) Version 3 (Gabard-Durnam et al., 2018). A specified subset of 30 channels was excluded ('E1', 'E8', 'E14', 'E17', 'E21', 'E25', 'E32', 'E38', 'E43', 'E44',' E48', 'E49', 'E56', 'E63', 'E68', 'E73', 'E81', 'E88', 'E94', 'E99', 'E107', 'E113', 'E114', 'E119', 'E120', 'E121', 'E125', 'E126', 'E127', 'E128'). Data were downsampled to 250Hz, bandpass filtered (1-100Hz), and filtered using a 50Hz cleanline filter for line noise removal. Bad channels were then automatically identified and rejected, and wavelet-enhanced Independent Component Analysis (ICA) and the Multiple Artifact Rejection Algorithm (MARA) were performed to detect and impute artifacts. Resting state data were segmented into 2s epochs; epochs with an amplitude ¿±150mV were rejected. Segments were also rejected using segment similarity criteria. Data were then re-referenced to the average of all channels.

EEG outputs from HAPPE were then reformatted and processed using the Batch Electroencephalography Automated Processing Platform (BEAPP) (Levin et al., 2018) to extract power spectra for each participant across the following frequency bands: delta (2-4Hz), theta (4-6Hz), low alpha (6-9Hz), high alpha (9-12Hz), beta (12-30Hz), and gamma (30-45Hz) and the following regions of interest (see Supp Figure 2): occipital ('E70', 'E71', 'E75', 'E76', 'E83'), temporal ('E36','E40', 'E41', 'E45', 'E46', 'E102', 'E103', 'E104', 'E108', 'E109'), parietal ('E52', 'E53', 'E59', 'E60', 'E85', 'E86', 'E91', 'E92'), and frontal ('E5', 'E6', 'E12', 'E13', 'E24', 'E27', 'E28', 'E33', 'E34', 'E112', 'E116', 'E117', 'E122', 'E123', 'E124'). Further, PSD values were normalized by a Log_{10} transform.

Developmental Outcomes (Bayley)

The Bayley Scales of Infant and Toddler Development, Fourth Edition (BSID-IV) cognitive, language, and motor subscales were administered to all participants. Research assistants were trained to research reliability in the administration and scoring of the Bayley-4. Due to cultural differences between the Bangladesh and the United States where the assessment was developed, Bangladeshi researchers modified some assessment stimuli to improve cultural responsiveness and relevancy. For example, pictures for the item naming series and action naming series of the expressive language and receptive language subscales were adapted to include items that Bangladeshi children are more likely to be familiar with and bedtime clothing that would signify the child in the picture was going to sleep instead of the one-piece pajamas worn in the original picture, which the Bangladeshi children would not be familiar with. SECTION ON WOLKES.

Biological sample collection

Stool samples were collected from each infant at their home at the baseline visit. Samples were collected in DNA/RNA Shield Fecal Collection Tubes (Zymo Research, #R1101). Peripheral venous blood samples were collected in EDTA Vacutainers, separated into plasma and RBCs and immediately frozen at -80 C. Batches of blood and stool samples were air-freighted on dry ice from Bangladesh to the Liggins Institute, New Zealand for processing and analysis.

Microbiome DNA extraction and sequencing

DNA was extracted from stool samples using the ZymoBIOMICS MagBead DNA/RNA extraction kit (Zymo Research, #R2136) following the standard protocol. Samples (1 mL) were mechanically lysed in bead bashing tubes using the MiniG tissue homogenizer prior to extraction of DNA. 200 µL of the sample was used post-bead bashing for extraction of DNA following the protocol. A volume of 50 µ L of elute was collected in DNAse/RNAse Free Water. Samples with a DNA concentration < 14.5 ng/µL were re-extracted following the ZymoBIOMICS DNA extraction protocol. Samples were sequenced (Illumina NovaSeq 150PE reads) to an average sequencing depth of 20M read-pairs/sample. Raw sequences were processed using BioBakery3 tools [**?**], specifically read quality filtering and human decontamination with KneadData (Version 1), taxonomic pro-

filing with MetaPhlAn3 (Version 3.1, using the mpa_v31_CHOCOPhlAn_201901 database) and functional profiling using presence/absence and abundance of microbial pathways (MetaCyc) with HUMAnN3 (Version 3.6). A minimum threshold of $> 0.1\%$ relative abundance and $> 5\%$ prevalence for all detected species was applied.

Plasma lipidomics

Plasma samples for lipidomics were thawed on ice and extracted according to a method modified from liu2016plasma. Briefly, 10 µL volume was placed in an amber glass autosampler vial and 300 µL of a mixture of Type 1 water, butanol, methanol, chloroform and SPLASH Lipidomix in a ratio of 4:15:15:20:1 was added. The mixture was vortexed and sonicated at room temperature before the protein precipitate was removed by centrifugation and an aliquot of supernatant transferred to an amber glass autosampler for negative ionisation LC-MS/MS. A second aliquot of supernatant was diluted 5 times with 75% IPA for positive ionisation LC-MS/MS. A 5 µL volume of each sample was injected onto a Phenomenex Kinetex F5 column (100 mm × 2.1 mm × 2.6 µm) and lipids were separated using a ternary gradient of Type 1 water, methanol and isopropanol containing ammonium acetate. Lipids were quantified and identified with a Q-Exactive mass spectrometer (Thermo Fisher Scientific, Germany) equipped with a heated electrospray ionisation HESI source. Data was processed using MS-DIAL v4.92 92 [**?**]. For full methodological details see the supplementary information.

Statistical Analyses

Python version 3.9.2 was used to perform all analysis [**?**]. Due to the unequal sample sizes and non-normally distributed data; non-parametric statistical approaches were used for differential abundance analysis. Relative abundances were adjusted by Centred Log Ratio to account for the compositional nature of the dataset [**?**]. Log adjusted fold change significance was measured using (MWU) test using the 'mannwhitneyu' function from 'scipy.stats' and adjusted for multiple testing using the 'fdrcorrection' function from statsmodels.stats.multitest. Principal Coordinates Analysis (PCoA) ordinations (plotted using 'skbio.stats.ordination.pcoa' module) were used to visualise the clustering of the Bray-Curtis dissimilarities (calculated using skbio.distance.pdist) between samples from their species and functional composition. To quantify the variance of

the gut microbiome explained covariates, PERMANOVA p-values were calculated from those Bray-Curtis Dissimilarities using the 'permanova' function from the 'skbio.stats.distance' module. Bray-Curtis were also used to capture the temporal dynamics of the microbiome from baseline. Numerical Associations between species and metadata were measured with Spearman correlation (calculated using 'spearmanr' function from 'scipy.stats' module), where significance was defined as False Discovery Rate (FDR) adjusted p-values of < 0.05 as per 2020SciPyNMeth. Associations between categorical data were measured with Fisher's Exact test (calculated using 'fisher_exact' from 'scipy.stats' module), where significance was defined as p-values of < 0.05.

Machine learning

Machine learning models were used to classify malnourished from well-nourished infants. Extra-trees Random Forest models were trained on functional and microbial taxa relative abundances. Model hyperparameters including the number of trees in the forest, maximum tree depth, and minimum sample numbers needed to split internal nodes were tuned using grid searching. A 5-fold cross-validation was used to measure the performance of each hyperparameter combination and to identify overfitting. Model performance was measured with AUCROC and out-of-bag error analysis (oob). SHAP Value (SHapley Additive exPlanations) interpretation was used to interpret the contributions each feature had on the model's performance using the 'shap' python package [**?**].

Network analysis

Absolute spearman rho of above 0.3 were used as edges and gut bacterial species and functional profiles, EEG, and plasma lipids were used as nodes coloured by their mean SHAP scores for classifier models that distinguish MAM from well-nourished conditions.

Code availability

All analysis code is available on the GitHub repository. The codebase is organised into scripts, providing a comprehensive framework for replicating the experiments. Detailed documentation

and instructions on how to use the code are provided in the repository's README file.

Ethics approval and consent to participate

Ethical approvals were obtained from the Research Review Committee (RRC; August 21, 2021) and Ethical Review Committee (ERC) of icddr,b (protocol no: PR-21084; September 21, 2021), Institutional Review Board of Boston Children's Hospital, USA (for analyses of neuropsychological assessments), University of Auckland, New Zealand (approval AH23922; for analyses of collected biological samples) and University of West Indies (CREC-MN.51, 21/22).

Data availability

Metagenome data is available at PRJNAXXX on the SRA. EEG and metadata are available from the authors, upon reasonable request that meets the ethics of the study.

Competing interests

The authors declare that they have no competing interests.

Funding

Work on this clinical trial is supported by Wellcome Leap (9942 Culver Blvd Unit 1277 Culver City, CA 90232-4167, United States; www.wellcomeleap.org) to PDG, JMO, TF and CAN as part of the 1kD Program. We acknowledge our core donors, Governments of Bangladesh, Canada for providing unrestricted support and commitment to icddr,b's research effort.

Author Contributions

TP, KG and JOS drafted and co-wrote the manuscript. TS, SHK, BCW, BH, CP, AB, DH, IS, AME, RD, GG, CK, PDG, RH, TF, CAN commented on the manuscript. JMO, RH, TF, PDG, CAN designed the study and analyses. TS, SHK performed assessments and obtained samples in Dhaka. RH oversaw the Dhaka group. TP performed multiomic analyses, BCW and IS performed metagenomics, CP performed metabolomics, JOS oversaw the Auckland group. BH performed EEG analyses, CAN oversaw the Boston group.

Acknowledgements

The authors would like to acknowledge the participants in Mirpur, Dhaka, Bangladesh for their contributions to this study. The authors would also like to thank the study team within the Infectious Diseases Division, International Centre for Diarrheal Disease Research, Bangladesh for their work in participant recruitment, sample collection and assessments.

Supplementary material

Table 1: Baseline infant characteristics. Plus minus values are means ± SD from continous variables and their pvalues are calculated using MWU. All other variables are categorical (True vs False) with their pvalues calculated using Fishers Exact test.

Table 2: Changes to gut microbial taxa associated with malnutrition.

Table 3: Changes to gut microbial functional pathways associated with malnutrition.

Table 4: Changes to Bayleys score associated with malnutrition.

Table 5: Changes to Wolkes score associated with malnutrition.

Table 6: Changes to EEG PSD associated with malnutrition.

Table 7: Changes to Wolkes score associated with malnutrition.

Table 8: SHAP score measure of important features by Random Forest classification.

Table 9: Spearman correlation between features.

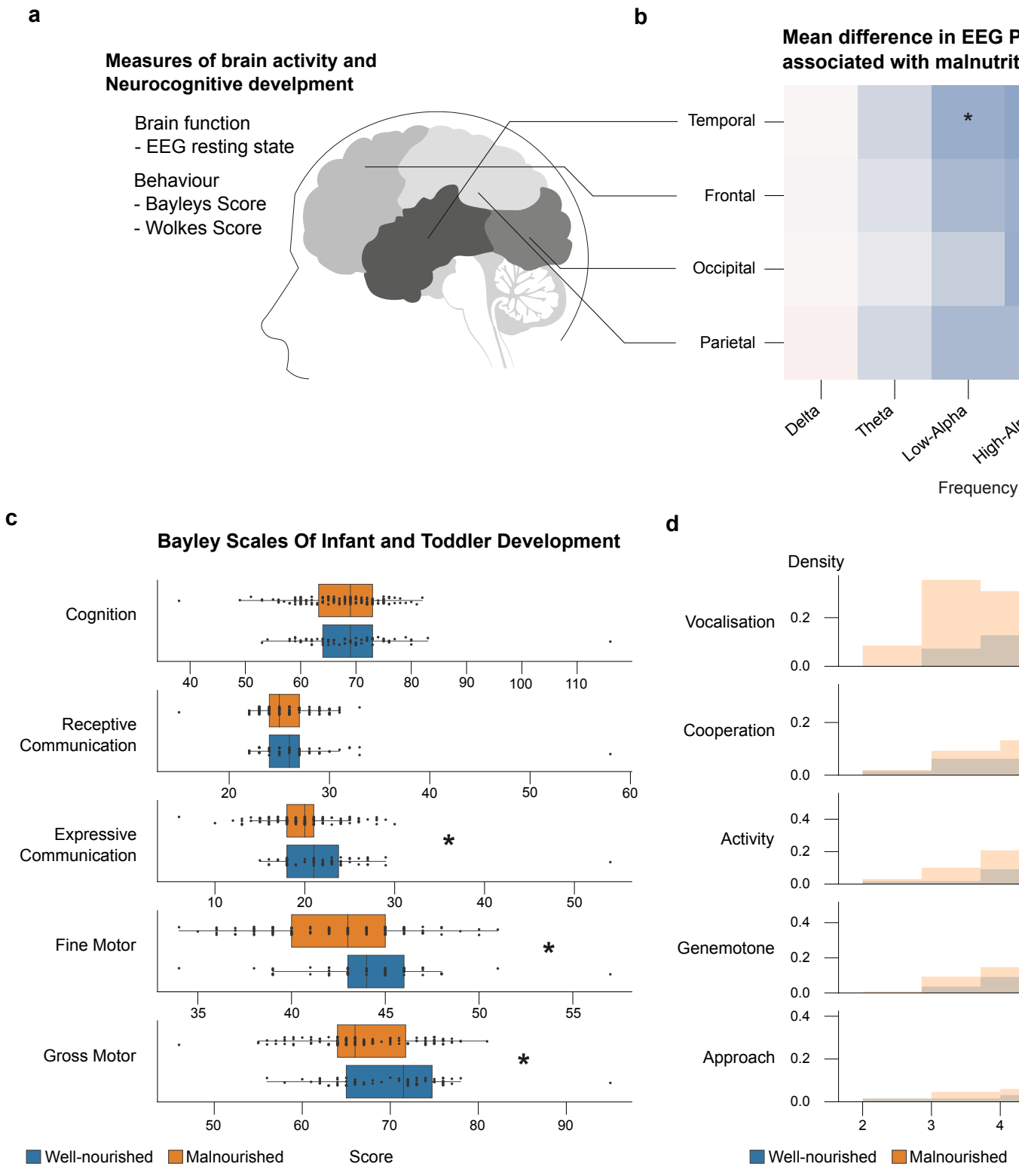

Figure 2: Differences in cognitive development of 12-month-old infants associate with malnutrition. a) Schematic of approach to study neurocognitive function. b) Boxplot of significant difference in Expressive Communication Score of the children with malnutrition compared to well-nourished controls. c) Heatmap of lobe and frequency specific changes in EEG resting state power spectral density (PSD) in MAM versus well-nourished infants. * = $q < 0.05$.

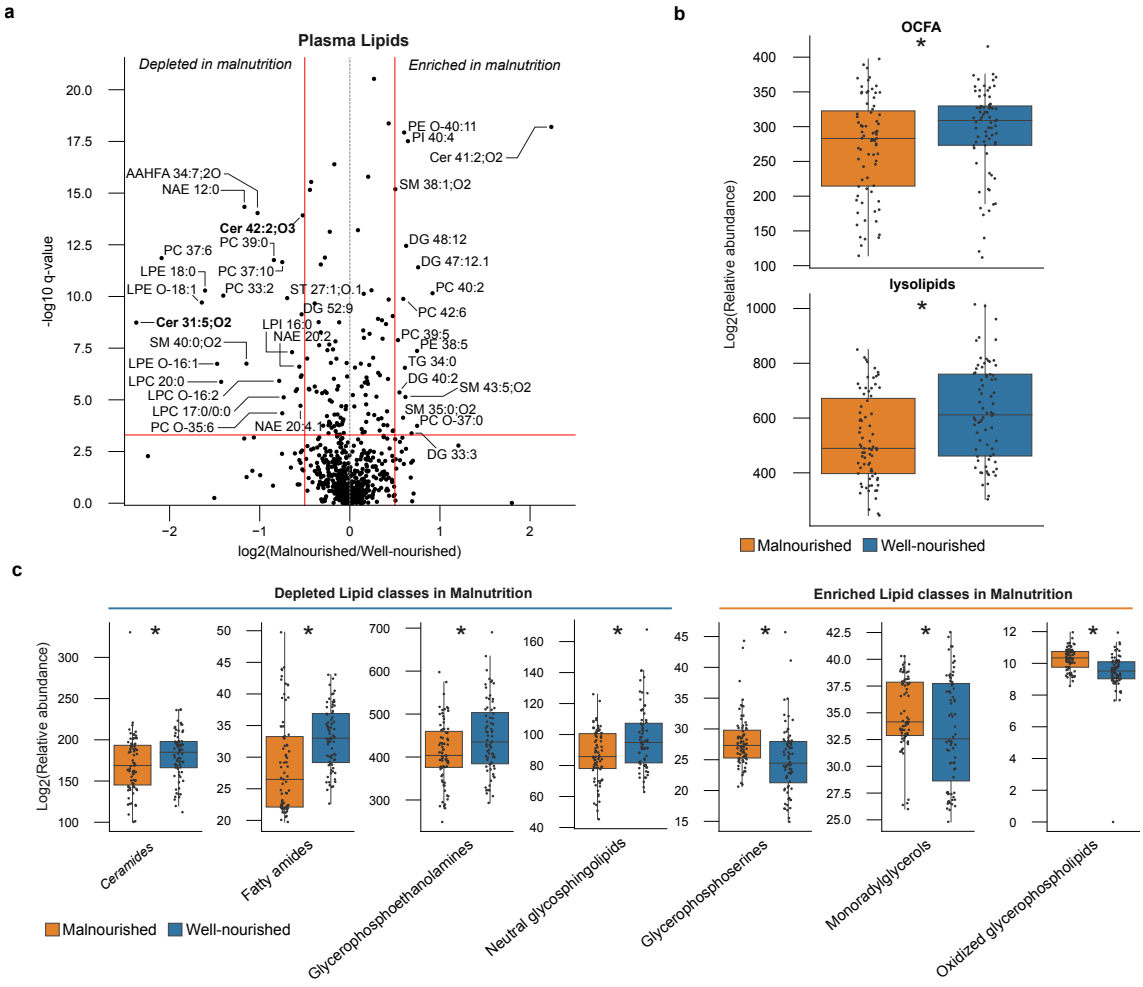

Figure 3: Malnutrition results in major, compositional differences in plasma lipids in 12-month-olds. Volcano plot changes to plasma lipids between well-nourished and MAM 12-month-olds. (Upper left and upper right quadrants signify significant changes where the red horizontal line signifies $q < 0.05$ and vertical lines represent Log_2(MAM/well-nourished) of -0.1 and 0.1 respectively). Boxplot of differences in Lysolipid (b) and Ceramide (c) concentrations associated with malnutrition.

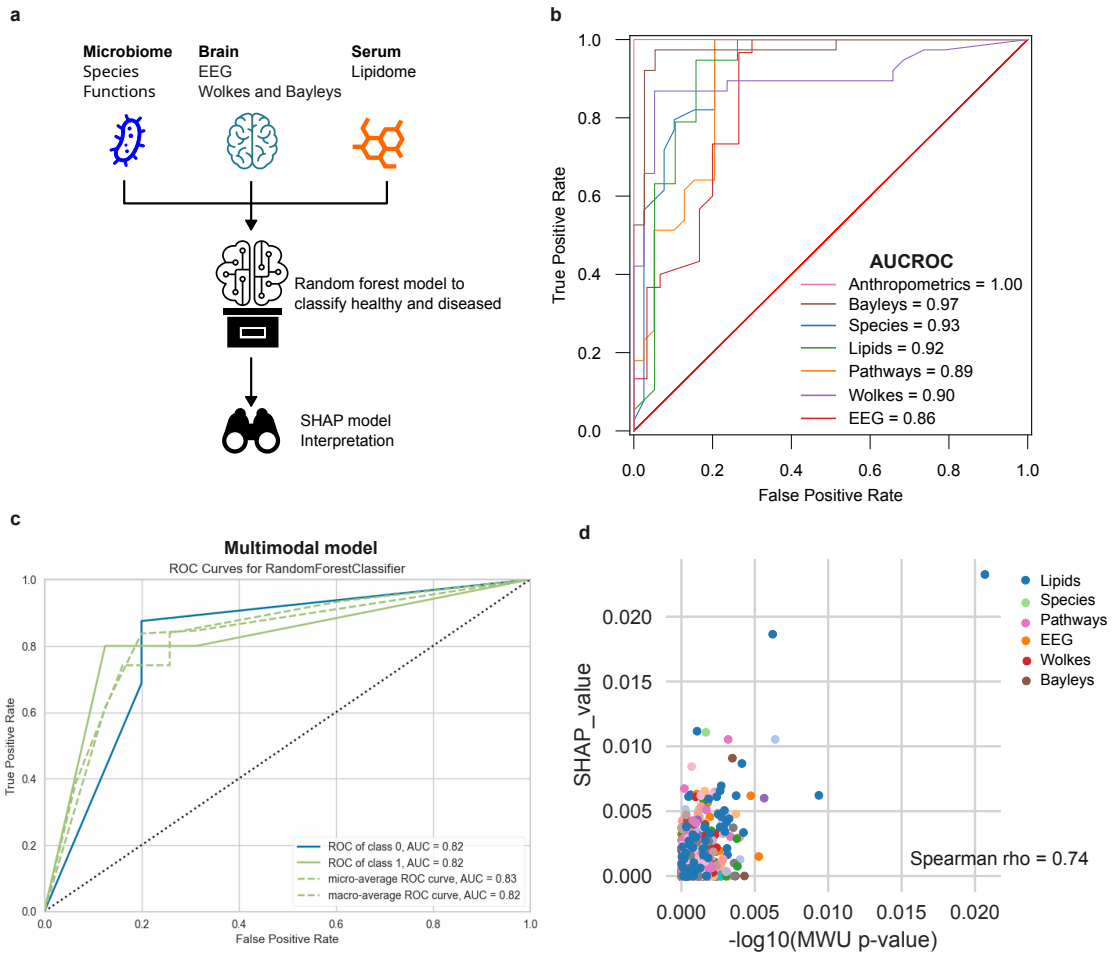

Figure 4: Integration of multimodal datasets boosts the predictive power and affects the relative feature importance of random forest models predicting nutritional status. a) Schematic describing interpreted multimodal approach to predict malnutrition. b) AUCROC curves showing relative predictive power of each modal dataset on predicting nutritional status. c) Multimodal model predicts malnutrition accurately. d) The multimodal model captures non-linear interactions between the features as demonstrated by the SHAP score distribution.

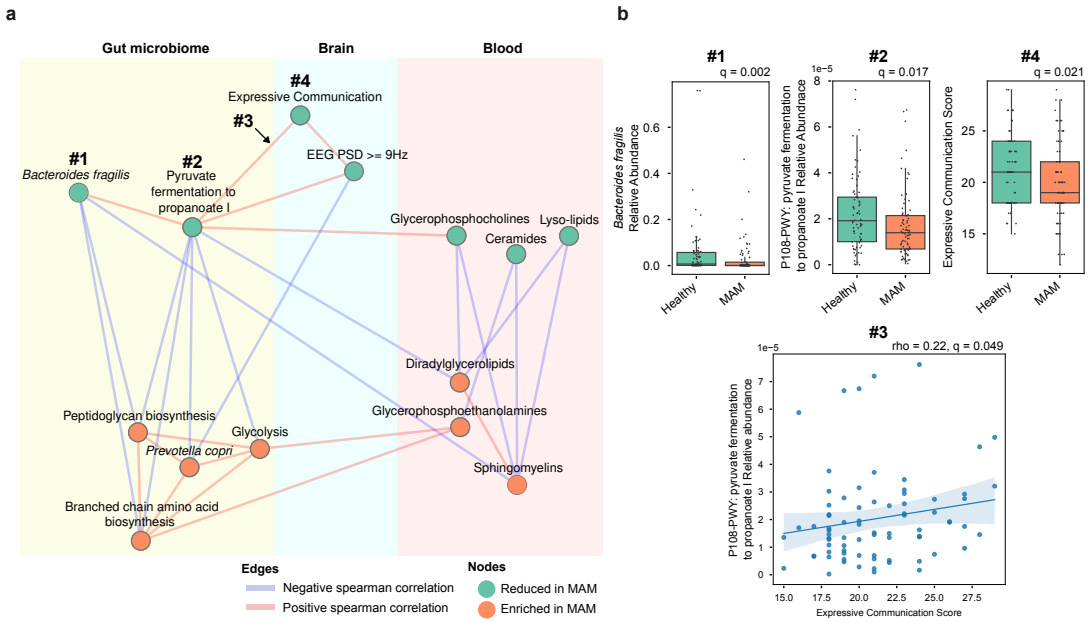

Figure 5: *Bacteroides fragilis* forms a network with propanoate synthesis, EEG and expressive communication that is anti-correlated with a *Prevotela copri* focused cluster of features in healthy and malnourished individuals. a) Network illustrating inter-relationships of feature associations that predict malnutrition. Inclusion in the network requires both a SHAP score for the node ($>$ 0.6) and a significant Spearman rho score for the correlation of $q < 0.05$. Nodes are features coloured by their enrichment in MAM (orange and green are enriched and depleted in MAM respectively). Edges are spearman correlations coloured red and blue being positively and negatively correlated respectively. b) Evidence for relative abundance of *B. fragilis* (#1), pyruvate fermentation to propanoate I pathway relative abundance (#2), correlation between pyruvate fermentation to propanoate I pathway and Expressive communication (#3), Expressive communication score distributions (#4).