Title

Gut microbes and their genes are associated with cognitive performance and brain structure in children

Authors

1. Bonham 0000-0003-3200-7533
2. Bottino 0000-0003-1953-1576
3. McCann 0000-0002-9753-7968
4. Beauchemin 0000-0003-2077-4639
5. Elizabeth Weisse
6. Fatoumata Barry
7. Rosa Cano Lorente 0000-0003-1971-7389
8. The RESONANCE Consortium
9. Curtis Huttenhower 0000-0002-1110-0096
10. Muriel Bruchhage 0000-0001-9637-0951
11. D'Sa 0000-0002-3476-8419
12. Deoni 0000-0002-0633-6328
13. Klepac-Ceraj 0000-0001-5387-5706

Abstract

The gastrointestinal tract, its resident microorganisms and the central nervous system are connected by biochemical signaling, also known as "microbiome-gut-brain-axis." Both the human brain and the gut microbiome have critical developmental windows in the first three years of life, raising the possibility that their development is co-occurring and likely co-dependent. Emerging evidence implicates gut microorganisms and microbiota composition in cognitive outcomes and neurodevelopmental disorders (e.g., autism and anxiety), but the influence of gut microbial metabolism on typical neurodevelopment has not been explored in detail. We investigated the relationship of the microbiome with the neuroanatomy and cognitive function of 361 healthy children, demonstrating that differences in gut microbial taxa and gene functions are associated with overall cognitive function as well as with differences in the size of multiple brain regions. Using a combination of multivariate linear models and machine learning (ML) models, we showed that many species, including *Gordonibacter pamelae* and *Blautia wexlerae*, were significantly associated with higher cognitive function, while some species such as *Ruminococcus gnavus* were more commonly found in children with low cognitive scores regardless of age or maternal education. Microbial genes for enzymes involved in the metabolism of neuroactive compounds, particularly short-chain fatty acids such as glutamate and propionate, were also associated with cognitive function. In addition, ML models were able to use microbial taxa to predict the volume of brain regions, with particular taxa often dominating

the feature importance metric for individual brain regions, such as *B. wexlerae* being the most dominant for the parahippocampal region or GABA-producing *Bacteroidetes ovatus* for left accumbens. These findings provide potential biomarkers of neurocognition and may lead to the future development of targets for early detection and early intervention.

Introduction

The gut and the brain are intimately linked. Signals from the brain reach the gut through the autonomic nervous system and the endocrine system, and the gut can communicate with the brain through the vagus nerve and through endocrine and immune (cytokine) signaling molecules (Cerdó et al., 2019; Pronovost and Hsiao, 2019; Sharon et al., 2016; Tognini, 2017). In addition, the products of microbial metabolism generated in the gut can influence the brain, both indirectly by stimulating the enteric nervous and immune systems, as well as directly through molecules that enter circulation and cross the blood-brain barrier. Causal links between the gut microbiome and neural development, particularly atypical development, are increasingly being identified (Spichak et al., 2021). Both human epidemiology and animal models point to the effects of gut microbes on the development of autism spectrum disorder (Laue et al., 2020; Wan et al., 2021) and particular microbial taxa have been associated with depression (Mayneris-Perxachs et al., 2022; Valles-Colomer et al., 2019) and Alzheimer's disease (Fung et al., 2017; Kim et al., 2021). But information about this "microbiome-gut-brain axis" in normal neurocognitive development remains lacking, particularly early in life.

The first years of life are critical developmental windows for both the microbiome and the brain (Laue et al., 2022). Fetal development *in utero* is believed to be mostly sterile, but is rapidly seeded at birth through contact with the birth canal (if birthed vaginally), caregivers, food sources (breastmilk or formula), and other environmental sources such as antibiotics (Bäckhed et al., 2015; Bokulich et al., 2016; Dominguez-Bello et al., 2010; Louwies et al., 2020). The early microbiome is characterized by low microbial diversity, rapid succession and evolution, and domination by Actinobacteria, particularly the genus *Bifidobacterium*, Bacteroidetes, especially *Bacteroides*, and Proteobacteria (Koenig et al., 2011). Many of these microbes have specialized metabolic capabilities for digesting human breast milk, such as *Bifidobacterium infantis* and *Bacteroides fragilis* (Sela et al., 2008; Tso et al., 2021). Upon the introduction of solid foods, the gut microbiome undergoes another categorical transformation, with most taxa of the infant microbiome being replaced by taxa more reminiscent of adult microbiomes (Bäckhed et al., 2015). Many prior studies have focused on either infant microbiomes or adult microbiomes, since performing statistical analyses across this transition poses particular challenges. Nevertheless, since this transition coincides with critical neural developmental windows and neural synaptogenesis, investigation across this solid-food boundary is incredibly important (Tau and Peterson, 2010).

A child's brain undergoes remarkable anatomical, microstructural, organizational, and functional changes in the first years of life. By age 5, a child's brain has reached >85% of its adult size, has achieved near-adult levels of myelination, and the pattern of axonal connections has been

established (Silbereis et al., 2016). Much of this development occurs in discrete windows or "sensitive-" or "critical periods" (CPs) (Knudsen, 2004) when neural plasticity is particularly high, and particular modes of learning and skill development are preferred. The timing of these sensitive periods is driven in part by genetics, but can also be affected by the environment, including the gut microbiome (Cowan et al., 2020). In fact, emerging evidence suggests that the timing and duration of CPs may be driven in part by cues from the developing gut microbiome (Callaghan, 2020). As such, understanding the normal spectrum of healthy microbiome development and how it relates to normal neurocognitive development may provide opportunities for identifying atypical development earlier and offer opportunities for intervention.

To begin to address this need, we investigated the gut microbiome and neurocognitive development of children from infancy through 10 years of age in an accelerated longitudinal study design. Gut microbial communities were assessed using shotgun metagenomic sequencing, enabling profiling at both the taxonomic and gene-functional level, and neurocognitive development was measured using expert-assessment of cognitive function and neuroimaging using magnetic resonance imaging (MRI). Using a combination of classical statistical analysis and machine learning, we showed that the development of the gut microbiome, the brain, and children's cognitive abilities are intimately linked, with both microbial taxa and gene functions able to predict cognitive performance and brain structure.

Results

Cohort overview and summary data

We investigated the co-development of the brain and the microbiome in early childhood in a cohort of typically-developing children in their first years of life using a variety of orthogonal microbial and neurocognitive assessments, including shotgun metagenomic sequencing, cognitive and behavioral assessments, and neuroimaging. In all, 361 children from RESONANCE, an accelerated longitudinal cohort of early development (Forrest, Blackwell, and Camargo 2018) between the ages of 82 days and 10 years old (median age 2.21 years, Supplementary Figure 1) were included in this study (Table 1, Figure 1A). To measure cognitive function, we used age-appropriate assessments that can be normalized to a common, IQ-like scale (Figure 1B), including the Mullen scale of early learning (MSEL) for children from birth to 3 years of age (Mullen and others, 1995), Wechsler preschool and primary scale of intelligence (WPPSI) for 4–5-year-olds (Wechsler, 2012), and the Wechsler intelligence scale for children (WISC) for children 6 years and up (Wechsler, 1949).

As expected, the greatest differences in microbial taxa were driven by age (Figure 1B-C, PCoA axis 1), while older children were primarily stratified into Bacteroidetes-dominant, Firmicutes-dominant, or high-*Prevotella copri* (Supplementary Figure 2). Overall variation in gut microbial genes and also in brain volume profiles was similarly driven largely by subject age (Figure 1C-F).

Table1: Subject demographics

Group	Subgroup	All	Under 6 mo	Over 18 mo
N subjects		361	74	257
Samples	1	262 (72.6%)	74 (100.0%)	203 (79.0%)
	2	72 (19.9%)	0 (0.0%)	46 (17.9%)
	>2	27 (7.5%)	0 (0.0%)	8 (3.1%)
Age (months)	Min	2.8	2.8	18.0
	Max	119.3	6.0	119.3
	Median	25.8	3.7	45.9
Sex	F	163 (45.2%)	37 (50.0%)	114 (44.4%)
	M	198 (54.8%)	37 (50.0%)	143 (55.6%)
Race	White	252 (69.8%)	42 (56.8%)	191 (74.3%)
	Black	40 (11.1%)	16 (21.6%)	21 (8.2%)
	Asian	1 (0.3%)	0 (0.0%)	1 (0.4%)
	Mixed	7 (1.9%)	1 (1.4%)	6 (2.3%)
	Other	48 (13.3%)	12 (16.2%)	29 (11.3%)
	Unknown/Declined	2 (0.6%)	1 (1.4%)	2 (0.8%)
Maternal Ed.	Junior high school	1 (0.3%)	0 (0.0%)	1 (0.4%)
	Some high school	6 (1.7%)	3 (4.1%)	4 (1.6%)
	High school grad	37 (10.2%)	13 (17.6%)	16 (6.2%)
	Some college	95 (26.3%)	25 (33.8%)	57 (22.2%)
	College grad	90 (24.9%)	17 (23.0%)	69 (26.8%)
	Grad/professional school	123 (34.1%)	15 (20.3%)	104 (40.5%)

Several studies have demonstrated links between specific taxa and measures of anxiety and depression (Mayneris-Perxachs et al., 2022; Needham et al., 2022), cognitive flexibility in adults (Magnusson et al., 2015), and atypical neural development (Liu et al., 2019; Wan et al., 2021). We, therefore, set out to identify whether specific taxa or gene functions were linked with normal cognitive development in children. In order to assess whether variations in gut microbial taxa, their genes, or their metabolism are linked with neurocognitive development, we tested whether the beta diversity of microbial taxa and gene functions, as well as variation in brain development as assessed by neuroimaging, were associated with these contemporaneous measures of cognitive function using permutation analysis of variance (PERMANOVA). Due to the large ecological shift in the microbiome that occurs upon the introduction of solid food, and a relatively wide range of ages when infants are transitioned to solid food, we considered the pre-transition (less than 6 months old) and post-transition (over 18 months old) microbiomes separately. Even still, age was a major driver of variation in gut microbiomes in children over 18 months old

(Figure 1G). We also found that overall variation in microbial species in children over 18 months old was associated with a small but significant variation in cognitive function score (Figure 1G, $R^2 = 0.0124$, $q < 0.001$), as was variation in microbial gene functions ($R^2 = 0.0119$, $q < 0.001$). Variation in microbial taxa and genes was not significantly associated with cognitive function in children under 6 months, though this may be due to the low taxonomic diversity, and broad lack of overlap between taxa in infants. As expected, age was significantly associated with microbial beta diversity (taxa $R^2 = 0.0215$, and UniRef90s $R^2 = 0.0245$, $q < 0.001$), and very strongly associated with neuroimaging profiles ($R^2 = 0.256$, $q < 0.001$).

Consistent with prior studies, different microbial measurement types captured overlapping variation, with species profiles and gene function profiles (annotated with clusters of 90% similarity (Suzek et al., 2007)), both generated from DNA sequencing data, being tightly coupled (Figure 1H, $p < 0.001$). Interestingly, other functional groupings (Enzyme commission level4 - ECs (Bairoch, 2000), and KEGG orthologues - KOs (Kanehisa et al., 2004)) overlapped only slightly with taxonomic profiles in both age cohorts, despite being derived from UniRef90 labels. In children over 18 months, some variation (15.9%, $p < 0.01$) in neuroimaging overlapped with microbial measures, though this may be due to the residual variation due to age in both measures.

Microbial species and neuroactive genes are associated with cognitive performance

To assess whether individual microbial species were associated with cognitive function, we fit multivariable linear regression (Mallick et al., 2021) to the relative abundance of each species that had at least 15% prevalence in a given age group (Figure 2A, N = 92 for 0–120 months, N = 46 for 0–6 months, N = 97 for 18–120 months). Only *Blautia wexlerae* was significantly associated (q value = 0.14, $\beta = 0.0015$) with cognitive function in children under 6 months old after adjusting for age and maternal education (Figure 2B). *B. wexlerae* was previously shown to be depleted in children with diabetes (Benítez-Páez et al., 2020), and that oral administration of *B. wexlerae* partially ameliorated weight gain and inflammation from a high-fat diet in a mouse model of T2D (Hosomi et al., 2022; Liu et al., 2021). In children over 18 months of age, several microbial species were significantly enriched (q-value < 0.20) in children with higher cognitive function scores, including *Gordonibacter pamelaeae*, which produces the neuroprotective metabolite urolithin (Gong et al., 2022; Selma et al., 2014), *Asaccharobacter celatus* and *Adelcreutzia equolifaciens*, which produce phytoestrogen-derived equol (Maruo et al., 2008; Thawornkuno et al., 2009), and the SCFA-producing probiotic species such as *Eubacterium eligens* and *Faecalibacterium prausnitzii* (Ghosh et al., 2020; Lopez-Siles et al., 2017; Mukherjee et al., 2020) (Figure 2B).

Given that different microbial species might occupy the same metabolic niche in different individuals, we hypothesized that microbial genes grouped by functional activity would be associated with cognition. To test this, we performed feature set enrichment analysis (FSEA) on groups of genes with neuroactive potential (Valles-Colomer et al., 2019) and concurrent cognitive function score (Table 2, Figure 2C-G) and found that several metabolic pathways were

significantly enriched or depleted in children with higher cognitive function scores. This was true both when considering all age groups together, though the enrichment of most pathways was more pronounced in children under 6 months or over 18 months. For example, genes for degrading the 3-carbon SCFA propionate were significantly depleted in children with higher cognitive function scores across all age groups tested (Table 2; propionate degradation I, under 6 months, enrichment score (E.S.) = -0.542, corrected p-value (q) = 0.020; over 18 months, E.S. = -0.674, q = 0.041). Interestingly, genes for propionate synthesis were also significantly depleted in higher scoring children over 18 months (E.S. -0.676, q = 0.023), as were genes for synthesizing the 2 carbon SCFA acetate in children under 6 months old (acetate synthesis I, E.S. = -0.194, q = 0.153; acetate synthesis II, E.S. = -0.342, q = 0.020; acetate synthesis III, E.S. = -0.31, q = 0.052). SCFAs are produced by anaerobic fermentation of dietary fiber and have been linked with immune system regulation as well as directly with brain function (Dalile et al., 2019).

Synthesis of menaquinone (vitamin K) was also negatively associated with cognitive function score in older children (menaquinone synthesis I, E.S. = -0.170, q = 0.0183). Menaquinone has several isoforms, one of which, MK-4, has been found to be decreased in children diagnosed with ASD (Dong et al., 2021) and is potentially neuroprotective in both rodents and humans (Elkattawy et al., 2022). In children over 18 months old, genes for the synthesis of the amino acids glutamate and tryptophan were significantly enriched in children with higher cognitive function scores (Glutamate synthesis I, E.S. = 0.242, q = 0.047; Tryptophan synthesis, E.S. = 0.119, q = 0.041). Glutamate is a critical neurotransmitter controlling neuronal excitatory/inhibitory signaling along with gamma-aminobutyric acid (GABA), and their balance in the brain controls neural plasticity and learning, particularly in the developing brain (Cohen Kadosh et al., 2015; Palomo-Buitrago et al., 2019). Tryptophan metabolism, including microbial metabolism of tryptophan, has previously been linked with autism in children (Hoshino et al., 1984; Xiao et al., 2021). Taken together, these results suggest that microbial metabolic activity, particularly the metabolism (synthesis and degradation) of neuroactive compounds may have effects on cognitive development.

Table 2 - Feature set enrichment analysis on neuroactive microbial gene sets

	All		under 6 months		over 18 months	
geneset	E.S.	q value	E.S.	q value	E.S.	q value
Acetate syn. I	-0.1223	0.5549	-0.1944	**0.1533**	0.1269	0.6023
Acetate syn. II	-0.2201	**0.0660**	-0.3424	**0.0198**	-0.0946	0.9199
Acetate syn. III	-0.2104	**0.0859**	-0.3098	**0.0518**	-0.0857	0.9199
Glutamate deg. I	-0.4037	**0.1802**	-0.2840	0.8159	-0.2489	0.6023
Glutamate syn. II	0.0953	0.6022	-0.1342	0.9802	0.2421	**0.0466**
Menaquinone syn. I	-0.0853	0.5635	-0.0833	0.8159	-0.1699	**0.1825**
Propionate deg. I	-0.4612	**0.1302**	-0.5421	**0.0198**	-0.6738	**0.0414**
Propionate syn. I	-0.3344	0.5549	-0.3201	0.5713	-0.6760	**0.0227**

Tryptophan syn.	0.0573	0.9742	-0.0725	0.8159	0.1186	**0.0414**

Gut microbial taxonomic and functional profiles predict cognitive function

FSEA relies on understanding functional relationships between individual genes. However, because the relationships between individual taxa are still largely unknown, we turned to Random Forest (RF) models, an unsupervised non-parametric machine learning (ML) method that enables the identification of underlying patterns in large numbers of individual features (here, microbial species). Previous studies have reported successful use of RFs for processing highly-dimensional and sparse data from the domain of genomics (Amaratunga et al., 2008; Brieuc et al., 2018; Chen and Ishwaran, 2012; Franzosa et al., 2019; Stephan et al., 2015), along with other works where it was used to predict cognitive conditions related to Alzheimer's disease in different scenarios (Ardekani et al., 2017; Velazquez et al., 2021). Additionally, given the sequential nature of variable consideration in each tree, RFs are naturally able to work out complex input feature interactions, such as those present in a microbiome-wide study, without the necessity to explicitly compute interaction terms.
Given that gut microbial profiles, as well as neurocognitive development, may partially reflect socioeconomic and demographic factors, we assessed the performance of RF regressors where maternal education (a proxy of socioeconomic status (SES)), sex, and age were included as possible predictors, either alone or in combination with microbial taxonomic profiles (Table 3).

Table 3. Benchmark metrics for the cognitive assessment score prediction models. Confidence intervals are calculated from the distribution of metrics from repeated CV at a confidence level of 95%

Subject Ages (months)	Microbial feature	Demo.	Test set correlation (± C.I.)	Test set Root-mean-square error (± C.I.)
0 to 6	-	+	-0.14 ± 0.01	13.01 ± 0.05
0 to 6	taxa	-	-0.10 ± 0.01	12.66 ± 0.05
0 to 6	taxa	+	-0.13 ± 0.01	12.70 ± 0.05
0 to 6	genes	-	-0.01 ± 0.01	12.56 ± 0.05
0 to 6	genes	+	-0.01 ± 0.01	12.56 ± 0.05
18 to 120	-	+	0.506 ± 0.003	16.27 ± 0.04
18 to 120	taxa	-	0.363 ± 0.003	17.66 ± 0.04
18 to 120	taxa	+	0.429 ± 0.003	17.29 ± 0.04

18 to 120	genes	-	0.303 ± 0.003	18.01 ± 0.04
18 to 120	genes	+	0.333 ± 0.004	17.86 ± 0.04

As with linear models, RF models for children under 6 months old were less generalizable (mean test-set correlation -0.13, mean RMSE 12.70), but RFs were consistently able to learn the relationship between taxa and cognitive function scores in children over 18 months of age (mean test-set correlation 0.429, mean RMSE 17.29). For both age groups, species that were important in RF overlapped with those that were significant when testing the relationship with linear models (Figure 3A-B, Supplementary Table XXA-B), though there were substantial differences. For children under 6 months, only *B. wexlerae* was significantly associated with cognitive function in linear models, but was ranked 20th in importance for RF models. The importance metric employed (MDI) is a measure of how well the variable differentiates (splits) sets of samples by creating subgroups that reduce the intra-group deviations, while also accounting for the amount of samples affected by a split. All taxa significantly associated with cognitive function score in children over 18 months using LMs belong to the top-ranking group responsible for 60% of the total relative importance, except for *Eubacterium ramulus* (Table 3, Figure 3A).

Interestingly, several taxa highly ranked in importance in both age groups, including several that were significant in LMs for children over 18 months old, including *R. gnavus* (0–6 months, rank = 13; 18–120 months, rank = 7) and *G. pamelaeae* (0–6 months, rank = 16; 18–120 months, rank = 20), while others such as *Allistipes finegoldii* were age-group specific. Several taxa important in RF models were not statistically significant when using linear models after multiple hypothesis correction. However, these taxa had small nominal p-values. For example, *Erysipelatoclostridium ramosum* was the most important feature in RF models for children under 6 months old and had an LM p-value of 0.04. Subject age was consistently ranked highly in feature importance, which could indicate that decision branches based on microbial taxa have increased purity when considering the subject's age or that age itself is a useful predictor.

Gut microbial taxonomic profiles predict brain structure differences

If there are causal effects of microbial metabolism on cognitive function, they might be reflected in changes in neuroanatomy. We again employed a Random Forest modeling approach to associate gut taxonomic profiles with individual brain regions identified in MRI scans, normalized to total brain volume. Some brain regions were more readily predicted by RF models trained on microbial taxa (Table 4, Supplementary Table XX), in particular those that were highly correlated with age. These included the L/R lingual gyrus (mean RF correlations, Left = 0.421, Right = 0.434; relative age importances, Left = 7.5%, Right = 8.3%) and the L/R pericalcarine cortex (mean RF correlations, Left = 0.200, Right = 0.273; relative age importances, Left = 3.7%, Right = 6.3%). In many cases, however, age was not an important variable in high-performing models, such as that for the left accumbens area (mean RF correlations = 0.288; relative age importances = 1.2%).

Table 4. Summary statistics for the neuroanatomy prediction benchmarks

Statistic	Mean absolute proportional error (MAPE)	Correlation coefficient (R)
mean	0.072	0.137
Standard deviation	0.028	0.136
maximum	0.206	0.605
75th percentile	0.082	0.195
50th percentile	0.067	0.134
25th percentile	0.056	0.062
minimum	0.035	-0.147

We also observed that many brain regions had high accordance across the left and right hemispheres in terms of both model performance and microbial feature importance. In contrast, other regions had substantial differences between the hemispheres. For example, the left accumbens area, which plays an important role in reward circuits (Ernst et al., 2005; Yau et al., 2012), has one of the highest test-set correlations of our brain region models (R = 0.288), as compared to the right accumbens models which could not adequately generalize, and had a negative mean test-set correlation (R = -0.041). Most healthy individuals have a rightward asymmetry in the nucleus accumbens, and reduced asymmetry has been linked to substance use disorder in young adults (Cao et al., 2021). Feature importance for models of the left accumbens area were dominated by three species of *Bacteroides*, *B. vulgatus* (3.4% relative importance), *B. ovatus* (3.7% relative importance), and *B. uniformis* (3.0% relative importance).. The accumbens area is associated with reward control, and in individuals diagnosed with ADHD, it has been shown to have a divergent neuromorphology (Hoogman et al., 2017). Independently, *B. ovatus*, *B. uniformis* and *B. vulgatus* have been linked to ADHD (Wang et al., 2020). In fact, there are studies showing that alterations on the striatal dopamine transporters can cause effects resembling hyperactivity and attention deficit (Yael et al., 2019), and that *B. uniformis* is gut-microbial modulator of the brain dopamine transporter (Hartstra et al., 2020).

Many of the taxa identified included taxa also identified in LMs and RF models of cognitive function. Interesting to note, while RF models for multiple brain regions had many important microbial taxa, others were dominated by a small number of taxa. In general, for all segments, a consistent number of species between 14 and 22 (median = 20) was responsible for one third of the fitness-weighted cumulative importance (see Methods), regardless of the model benchmarks. We proceeded to select a subset of the most important taxa (averaged over all segments) and the segments whose importance was more heavily loaded to further analyze the multiple relationships unveiled by the RF models (**Figure 4 A and B**)

Our analysis revealed two major patterns of importance distribution from taxa over the brain segments; some species portrayed high contributions to multiple different segments, while others contributed modestly to just one or two brain segments. Notable cases of the first pattern included seven species - *Anaerostipes hadrus*, *Bacteroides vulgatus*, *Fusicatenibacter saccharivorans*, *Ruminococcus torques*, *Eubacterium rectale*, *Coprococcus comes* and *Blautia wexlerae* - that combined, account for approximately one third of the cumulative relative importances, computed after subsetting on the taxa of interest (Figure 4C).

Among these, the most important variable is *Anaerostipes hadrus*, which is a butyrate-producing anaerobe that has been positively associated with cognitive function (Kant et al., 2015; Li et al., 2022). Its importance is, however, heavily loaded on the volume of entire major lobes (frontal, parietal), which cannot be readily correlated to specific cognitive functions. After accounting for the major lobes, it is found to be highly important on the prediction of the cerebellar vermal lobules VIII-X, pars opercularis, cuneus and precuneus, and anterior-cingulate (Figure 4B).

To better understand this relationship, we performed hierarchical cluster analysis, which revealed a well defined cluster of species with high importance loadings on close segments of the lower temporal and close occipital lobe, to which *A. hadrus* belongs. The other species on this cluster are *B. wexlerae*, *R. torques*, *R. intestinalis, R. bicirculans* and *F. saccahrivorans*, who contribute heavily to the entorhinal, fusiform, lingual and parahippocampal segments.

This group of taxa and their related segments drew our attention because they contained *B. wexlerae*, a highlight from the cognitive assessment results. Previous works exploring Parkinson Disease patients found out that issues in the cognitive task of confrontation naming were positively correlated with thinning in the fusiform gyrus and parahippocampal gyrus (Pagonabarraga et al., 2013). Additionally, the left and right parahippocampal, where *B. wexlerae* had the highest importance in both models, are important in visual/spatial processing and memory (Aminoff et al., 2013). *B. wexlerae* can also produce acetylcholine in the gut (Hosomi et al., 2022), and this molecule plays an important role in modulating memory function (Haam and Yakel, 2017).

Another notable importance cluster contains taxa associated with the basal forebrain, especially the cingulate and the accumbens. This cluster contains the previously-discussed *B. ovatus* and *B. uniformis,* but also *Alistipes finegoldii and Streptococcus salivaris.* Reduction in the nucleus accumbens has been associated with depression symptoms (Wacker et al., 2009), and increased levels of the *Alistipes* genus have been observed in patients with depressive disorder (Jiang et al., 2015). While these reports are important, simultaneously probing of the gut metagenome and brain structure relationship is novel.

Both *A. equolofaciens* and *A. celatus,* two closely-related equol-producing species, are examples of the second contribution pattern, and had high importance in predicting the relative volume of the right anterior cingulate (respectively, 3.0% and 2.6% relative importances), which has been linked to social cognition and reward-based decision making (Apps et al., 2016; Boes

et al., 2008; Bush et al., 2002). Equol has a strong estrogenic effect (Setchell et al., 2005), and in anterior cingulate, estrogen, has been shown to regulate pain-related aversion (Xiao et al., 2013). Another example of this pattern is *R. gnavus,* which was the only species with significant negative association with cognitive function. It is heavily associated with the left pars opercularis (2.7% relative importance), a relationship that may be explained by the emerging understanding that this species is increased in individuals with insulin resistance and obesity (Ley et al., 2006), conditions that are known to produce structural abnormalities in the brain (Opel et al., 2021). Finally, *Coprococcus comes* displays an importance distribution that splits almost evenly among the two previously reported clusters. Its loading on the prediction of the left posterior cingulate is the highest for a microbe in RF models (4.0% relative importance) (only age had higher relative importance in any model), while also being one of the most important predictors for neighboring areas of the posterior cingulate such as the pars opercularis (relative importances, Left = 2.3%, Right = 2.8%), along with upper regions like the left precentral (2.3% relative importance) and paracentral lobes (relative importances, Left = 2.6%, Right = 1.8%).

Discussion

The relationship between the gut microbiome and brain function via the gut-microbiome-brain axis has gained increasing acceptance largely as a result of human epidemiological studies investigating atypical neurocognition (eg anxiety and depression, neurodegeneration, attention deficit / hyperactivity disorder, and autism) and mechanistic studies in animal models. The results from these studies point to the possibility that gut microbes and their metabolism may be causally implicated in cognitive development, but this study is the first to our knowledge that directly investigates microbial species and their genes in relation to typical development in young children. Understanding the gut-brain-axis in early life is particularly important, since differences or interventions in early life can have outsized and longer-term consequences than those at later ages. Further, even in the absence of causal impacts of microbial metabolism, identifying risk factors that could point to other early interventions would also have value.

The use of shotgun metagenomic sequencing enabled us to get species-level resolution of microbial taxa. A previous study of cognition in 3 year old subjects used 16S rRNA gene amplicon sequencing, and showed that genera from the Lachnospiraceae family as well as unclassified Clostridiales (now Eubacteriales) were associated with higher scores on the Ages and Stages Questionnaire (Sordillo et al., 2019). However, each of these clades encompass dozens of genera with diverse functions, each of which may have different effects. Indeed, several of the taxa that were positively associated with cognitive function in this study, including *B. wexlerae*, the only species identified by linear models in children under 6 months, *D. longicatena*, *R. faecis*, and *A. finegoldii* are Clostridiales, as is *R. gnavus*, which we found was negatively associated with cognitive function (Figure 2A-B). This kind of species-level resolution is typically not possible with amplicon sequencing.

We identified several species in the family Eggerthelaceae that were associated with cognitive function, including *Gordonibacter pamelaeae, Aldercreutzia equolofaciens, Asaccharobacter*

celatus (formerly regarded a subspecies of *A. equolofaciens* (Takahashi et al., 2021)), and *Eggerthella lenta*. Many members of this family are known in part due to unique metabolic activities. For example, *A. equolofaciens* produces the nonsteroidal estrogen equol from isoflaven found in soybeans (Wang et al., 2005), and *G. pamelaeae* can metabolize the polyphenol ellagic acid (found in pomegranates and some berries) into urolithin, which has been shown in some studies to have a neuroprotective effect (Gong et al., 2022; Selma et al., 2014). *E. lenta* has been extensively studied for its ability to metabolize drug compounds such as the plant-derived heart medication digoxin (Haiser et al., 2013). The metabolic versatility of this clade, and the large number of species that are associated with cognition make these microbes prime targets for further mechanistic studies.

In addition to improved species-level resolution, shotgun metagenomic sequencing also enables gene-functional insight. We showed here that genes for the metabolism of SCFAs, both their degradation and synthesis, are associated with cognitive function scores. However, while the differential abundance of genes for the metabolism of neuroactive compounds like these is suggestive, it is difficult to reason about the relationship between levels of these genes and the gut concentrations of the molecules their product enzymes act on. For example, while it might be intuitive to reason that increased levels of menaquinone synthesis genes is indicative of increased menaquinone, it could be the case that menaquinone deficiency selects for microbes that can synthesize it. For the same reason, increased propionate degradation genes may counterintuitively be indicative of high levels of propionate in the gut lumen, since high propionate would select for microbes that can metabolize it. For this reason, future studies coupling shotgun metagenomics with stool metabolomics could improve our understanding of the relationship between microbial metabolism and cognitive development. Further, strain-level analysis linking specific gene content in species of interest could further refine targeted efforts at identifying specific metabolic signatures of microbe-brain interactions.

The use of multiple age-appropriate cognitive assessments that could be normalized to a common scale enabled us to analyze microbial associations across multiple developmental periods, but carries several drawbacks. In particular, the test-retest reliability, as well as systematic differences between test administrators may introduce substantial noise into these observations, particularly in the youngest children. In addition, our study period overlapped with the beginning of the COVID-19 pandemic, and we and others have observed some reduction in measured scores for children that were assessed after the implementation of lockdowns. In our subject set for this study, these effects are more pronounced in some age groups due to our sampling schedule (Blackwell et al., 2022; Deoni et al., 2021) (Supplementary Figure 3).

This analysis allowed us to establish links between microbial taxa and their functional potential with cognition and brain structure. Although we cannot test causality or the chemistry behind the interactions between gut microbial taxa, gut, and brain, this study provides clear and statistically significant associations between the infant and early child gut microbiota and neurocognition. Future studies should focus on characterizing the early-life microbiome and neurocognitive development across different geographic regions and lifestyles such as covering traditionally understudied low-resource urban, peri-urban and rural communities to obtain the more

comprehensive understanding of the variability within the different gut microbiomes reflects on neurocognition. These studies would also provide us with the wealth of data on different strains from the same species to better understand the effect of genes and their products. Furthermore, culturing and microbial community enrichment studies combined with genetic manipulation and genomic approaches to understand microbial metabolism at the molecular level is the key, as the metabolic functions shape and influence the human host and its health. The discovery of the neuroactive metabolites could provide us with biomarkers for early detection or necessary medicinally useful molecules that can be applied in intervention.

Material and Methods

Study Ethics

All procedures for this study were approved by the local institutional review board at Rhode Island Hospital, and all experiments adhered to the regulation of the review board. Written informed consent was obtained from all parents or legal guardians of enrolled participants.

Participants

Data used in this study were drawn from the ongoing longitudinal RESONANCE study of healthy and neurotypical brain and cognitive development, based at Brown University in Providence, RI, USA. The RESONANCE study is part of the NIH initiative Environmental influences on Child Health Outcomes (ECHO) (Forrest et al., 2018; Gillman and Blaisdell, 2018), a longitudinal observational study of healthy and neurotypical brain development that spans the fetal and infant to adolescent life stages, combining neuroimaging (magnetic resonance imaging, MRI), neurocognitive assessments, bio-specimen analyses, subject genetics, environmental exposures such as lead, and rich demographic, socioeconomic, family and medical history information. From the RESONANCE cohort, 361 typically-developing children between the ages of 2.8 months and 10 years old (median age 2 years, 2 months) were selected for analysis in this study.

General participant demographics are provided in Table 1. Children are representative of the RI population. Children in the RESONANCE cohort were born full-term (< 37 weeks gestation) with height and weight normal for gestational age, and from uncomplicated singleton pregnancies. Children with known major risk factors for developmental abnormalities at enrollment were excluded. In addition to screening at the time of enrollment, on-going screening for worrisome behaviors using validated tools was performed to identify at-risk children and remove them from subsequent analysis.

Exclusion criteria included: *in utero* exposure to alcohol, cigarette or illicit substance exposure; preterm (< 37 weeks gestation) birth; small for gestational age or less than 1500 g; fetal ultrasound abnormalities; preeclampsia, high blood pressure, or gestational diabetes; 5 minute APGAR scores < 8; NICU admission; neurological disorder (e.g., head injury resulting in loss of

consciousness, epilepsy); and psychiatric or learning disorder (including maternal depression) in the infant, parents, or siblings requiring medication in the year prior to pregnancy.

Demographic and other non-biospecimen data such as race and ethnicity, parental education and occupation, feeding behavior (breast- and formula-feeding), child weight and height, were collected through questionnaires or direct examination as appropriate. All data were collected at every assessment visit, if possible.

Cognitive Assessments

Overall cognitive function was assessed using age-appropriate methods. For children from birth to 30 months, we used an Early Learning Composite as assessed via the Mullen Scales of Early Learning (MSEL) (Mullen and others, 1995), a standardized and population-normed tool for assessing fine and gross motor, expressive and receptive language, and visual reception functioning in children from birth through 68 months of age. The Wechsler Intelligence Quotient for Children (WISC) (Wechsler, 2012) is an individually administered standard intelligence test for children aged 6 to 16 years. It derives a full scale intelligence quotient (IQ) score, which we used to assess overall cognitive functioning. The fourth edition of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV) is an individually administered standard intelligence test for children aged 2 years 6 months to 7 years 7 months, trying to meet the increasing need for the assessment of preschoolers. Just as the WISC, it derives a full scale IQ score, which we used to assess overall cognitive functioning.

Stool Sample Collection and Sequencing

Stool samples (n=493) were collected by parents in OMR-200 tubes (OMNIgene GUT, DNA Genotek, Ottawa, Ontario, Canada), stored on ice, and brought within 24 hrs to the lab in RI where they were immediately frozen at -80 ˚C. Stool samples were not collected if the subject had taken antibiotics within the last two weeks. DNA extraction was performed at Wellesley College (Wellesley, MA). Nucleic acids were extracted from stool samples using the RNeasy PowerMicrobiome kit, excluding the DNA degradation steps. Briefly, the samples were lysed by bead beating using the Powerlyzer 24 Homogenizer (Qiagen, Germantown, MD) at 2500 rpm for 45 s and then transferred to the QIAcube (Qiagen, Germantown, MD) to complete the extraction protocol. Extracted DNA was sequenced at the Integrated Microbiome Resource (IMR, Dalhousie University, NS, Canada).

Shotgun metagenomic sequencing was performed on all samples. A pooled library (max 96 samples per run) was prepared using the Illumina Nextera Flex Kit for MiSeq and NextSeq from 1 ng of each sample. Samples were then pooled onto a plate and sequenced on the Illumina NextSeq 550 platform using 150+150 bp paired-end “high output” chemistry, generating 400 million raw reads and 120 Gb of sequence per plate.

Computational analysis of metagenomes

Shotgun metagenomic sequences were analyzed using the bioBakery suite of computational tools (Beghini et al. 2021). First, KneadData (v0.7.7) was used to perform quality control of raw sequence reads, such as read trimming and removal of reads matching a human genome reference. Next, MetaPhlAn (v3.0.7, using database mpa_v30_CHOCOPhlAn_201901) was used to generate taxonomic profiles by aligning reads to a reference database of marker genes. Finally, HUMAnN (v3.0.0a4) was used to functionally profile the metagenomes.

Machine learning for cognitive development

Prediction of cognitive scores was carried out as a set of regression experiments targeting real-valued continuous assessment scores. Different experiment sets were designed to probe how different representations of the gut microbiome (taxonomic profiles or functional profiles encoded as ECs) would behave, with and without the addition of demographics (sex and maternal education as a proxy of socioeconomic status) on participants from different age groups. Age (in months) was provided as a covariate for all models (Table 3).

Table 5. Experimental design and input composition for Random Forest experiments

Input set	Age bracket	Microbiome encoding type	Demographics Provided? (sex, education)
1	0 to 6 months	Not provided	yes
2	0 to 6 months	Taxonomic profile	no
3	0 to 6 months	Taxonomic profile	yes
4	0 to 6 months	Functional Profile (ECs)	no
5	0 to 6 months	Functional Profile (ECs)	yes
6	18 to 120 months	Not provided	yes
7	18 to 120 months	Taxonomic profile	no
8	18 to 120 months	Taxonomic profile	yes
9	18 to 120 months	Functional Profile (ECs)	no
10	18 to 120 months	Functional Profile (ECs)	yes

Random Forests (RFs) (Breiman 1996) were selected as the prediction engine and processed using the DecisionTree.jl (Sadeghi et al., 2022) implementation, inside the MLJ.jl (Blaom et al., 2020) Machine Learning framework. Independent RFs were trained for each experiment, using a set of default regression hyperparameters from Breiman and Cutler (Breiman, 2001), on a

repeated cross-validation approach with different RNG seeds. One hundred repetitions of 3-fold CV with 10 different intra-fold RNG states each were employed, for a total of 3000 experiments per input set.

After the training procedures, the root-mean-square error (RMSE) for cognitive assessment scores and mean absolute proportional error (MAPE) for the brain segmentation data, along with Pearson's correlation coefficient (R) were benchmarked on the validation and train sets. MAPE was chosen as the metric for brain segments due to magnitude differences between median volumes of each segment, which would hinder interpretation of raw error values without additional reference.

To derive biological insight from the models, the covariate variable importances for all the input features, measured by mean decrease in impurity (MDI, or GINI importance), was also analyzed. Leveraging the distribution of results from the extensive repeated cross validation experiments, rather than electing a representative model or picking the highest validation-set correlation, a measure of model fitness (**Equation 1**) was designed to weight the importances from each trained forest. The objective was to give more weight to those with higher benchmarks on the validation sets (or higher generalizability), while penalizing information from highly overfit models, drawing inspiration from the approach used on another work employing repeated CV on Random Forests with high-dimensional, low sample size microbiome datasets (Woodruff et al., 2022). The resulting fitness-weighted importances were used to generate the values in Figure 3.

$$fitness = \sqrt{max(r_{train}, 0) \cdot max(r_{test}, 0)}$$

Equation 1. Mathematical expression of the fitness measure used to weight feature importances based on model benchmarks

MRI / segmentation

MRI data was acquired at a 3T Siemens Trio scanner with the following parameters: TE=5.6msec, TR=1400msec, FA=15 degrees, 1.1x1.1x1.1mm resolution, 160x160 matrix, with an average of 112 slices. FOV was adjusted to infant size. Using a combination of linear and nonlinear image registration, we created representative age-specific templates using the ANTs package (Avants et al., 2014). After age specific templates were created, a single flow from each age to the 12-month template was estimated and a final warp from the 12-month template to standard adult MNI space was performed. For each individual infant/ child brain image, we then calculated the warp from their native T1w image space to their nearest-in-age template. Using the resulting warps (native → nearest age template → 12-month template → MNI template), we could move the standard adult brain atlas to the space of an individual infant in a single step. In this case, we used the Harvard-Oxford brain atlas to provide a coarse-grained parcellation of individual brains into subcortical regions (e.g., thalamus, putamen) and total grey and white matter volumes (included as part of the FSL package (Jenkinson et al., 2012). Total tissue and brainstem volume as well as left and right hemisphere volumes were derived for total white and cortical gray matter, lateral ventricle, thalamus, caudate, putamen, pallidum,

hippocampus, amygdala, and accumbens as well as total brainstem volume (Bruchhage et al., n.d.).

Data and code availability

Taxonomic and functional microbial profiles, as well as subject demographics necessary for statistical analyses and machine learning are available on the Open Science Framework (Bonham et al., 2022). Data processing, generation of summary statistics, and generation of plots was performed using the julia programming language (Bezanson et al., 2017; Bonham et al., 2021; Danisch and Krumbiegel, 2021). All code for data analysis and figure generation, as well as scripts for automated download of input files are available on github .

Acknowledgements

Research funding: NIH UG3 OD023313 and Wellcome LEAP 1kD.

References and Citations

https://doi.org/10.1093/cercor/bhs201
Amaratunga, D., Cabrera, J., Lee, Y.-S., 2008. Enriched random forests. Bioinformatics 24, 2010–2014. https://doi.org/10.1093/bioinformatics/btn356
Aminoff, E.M., Kveraga, K., Bar, M., 2013. The role of the parahippocampal cortex in cognition. Trends Cogn. Sci. 17, 379–390. https://doi.org/10.1016/j.tics.2013.06.009
Apps, M.A.J., Rushworth, M.F.S., Chang, S.W.C., 2016. The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. Neuron 90, 692–707. https://doi.org/10.1016/j.neuron.2016.04.018
Ardekani, B.A., Bermudez, E., Mubeen, A.M., Bachman, A.H., Initiative, for the A.D.N., 2017. Prediction of Incipient Alzheimer's Disease Dementia in Patients with Mild Cognitive Impairment. J. Alzheimers Dis. 55, 269–281. https://doi.org/10.3233/JAD-160594
Avants, B.B., Tustison, N.J., Stauffer, M., Song, G., Wu, B., Gee, J.C., 2014. The insight ToolKit image registration framework. Front. Neuroinform. 8, 44. https://doi.org/10.3389/fninf.2014.00044
Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., Khan, M.T., Zhang, J., Li, J., Xiao, L., Al-Aama, J., Zhang, D., Lee, Y.S., Kotowska, D., Colding, C., Tremaroli, V., Yin, Y., Bergman, S., Xu, X., Madsen, L., Kristiansen, K., Dahlgren, J., Wang, J., 2015. Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe 17, 690–703. https://doi.org/10.1016/j.chom.2015.04.004
Bairoch, A., 2000. The ENZYME database in 2000. Nucleic Acids Res. 28, 304–305. https://doi.org/10.1093/nar/28.1.304
Benítez-Páez, A., Gómez del Pugar, E.M., López-Almela, I., Moya-Pérez, Á., Codoñer-Franch, P., Sanz, Y., 2020. Depletion of *Blautia* Species in the Microbiota of Obese Children Relates to Intestinal Inflammation and Metabolic Phenotype Worsening. mSystems 5, e00857-19. https://doi.org/10.1128/mSystems.00857-19
Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: A fresh approach to numerical

computing. Siam Rev. 59, 65–98. https://doi.org/10.1137/141000671

Blackwell, C.K., Mansolf, M., Sherlock, P., Ganiban, J., Hofheimer, J.A., Barone, C.J., Bekelman, T.A., Blair, C., Cella, D., Collazo, S., Croen, L.A., Deoni, S., Elliott, A.J., Ferrara, A., Fry, R.C., Gershon, R., Herbstman, J.B., Karagas, M.R., LeWinn, K.Z., Margolis, A., Miller, R.L., O'Shea, T.M., Porucznik, C.A., Wright, R.J., 2022. Youth well-being during the COVID-19 pandemic. Pediatrics 149. https://doi.org/10.1542/peds.2021-054754

Blaom, A.D., Kiraly, F., Lienart, T., Simillides, Y., Arenas, D., Vollmer, S.J., 2020. MLJ: A Julia package for composable machine learning. J. Open Source Softw. 5, 2704. https://doi.org/10.21105/joss.02704

Boes, A.D., Tranel, D., Anderson, S.W., Nopoulos, P., 2008. Right anterior cingulate cortex volume is a neuroanatomical correlate of aggression and defiance in boys. Behav. Neurosci. 122, 677–684. https://doi.org/10.1037/0735-7044.122.3.677

Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., D Lieber, A., Wu, F., Perez-Perez, G.I., Chen, Y., Schweizer, W., Zheng, X., Contreras, M., Dominguez-Bello, M.G., Blaser, M.J., 2016. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci. Transl. Med. 8, 343ra82. https://doi.org/10.1126/scitranslmed.aad7121

Bonham, K., Bottino, G.Z.M.F., Klepac-Ceraj, V., 2022. ECHO RESONANCE Microbiome. https://doi.org/10.17605/OSF.IO/YBS32

Bonham, K., Kayisire, A., Luo, A., Klepac-Ceraj, V., 2021. Microbiome.jl and BiobakeryUtils.jl - Julia packages for working with microbial community data. J Open Source Softw 6, 3876. https://doi.org/10.21105/joss.03876

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brieuc, M.S.O., Waters, C.D., Drinan, D.P., Naish, K.A., 2018. A practical introduction to Random Forest for genetic association studies in ecology and evolution. Mol. Ecol. Resour. 18, 755–766. https://doi.org/10.1111/1755-0998.12773

Bruchhage, Muriel, Chen, Y., Cataldo, A.G., Müller, H.-G., Weisee, E., Wilson, S., Pietsch, M., D'Sa, V., Marquand, A., Madhow, S., Bouchard, K., Cole, J.H., Biondo, F., Elison, J., O'Muirchheartaigh, J., Deoni, C, S., n.d. Longitudinal brain and cognitive development of the first 1000 Days:A large Multi-Cohort Multi-Scanner study. Presented at the International Society of Magnetic Resonance Medicine.

Bush, G., Vogt, B.A., Holmes, J., Dale, A.M., Greve, D., Jenike, M.A., Rosen, B.R., 2002. Dorsal anterior cingulate cortex: A role in reward-based decision making. Proc. Natl. Acad. Sci. 99, 523–528. https://doi.org/10.1073/pnas.012470999

Callaghan, B., 2020. Nested sensitive periods: how plasticity across the microbiota-gut-brain axis interacts to affect the development of learning and memory. Curr. Opin. Behav. Sci. 36, 55–62. https://doi.org/10.1016/j.cobeha.2020.07.011

Cao, Z., Ottino-Gonzalez, J., Cupertino, R.B., Schwab, N., Hoke, C., Catherine, O., Cousijn, J., Dagher, A., Foxe, J.J., Goudriaan, A.E., Hester, R., Hutchison, K., Li, C.-S.R., London, E.D., Lorenzetti, V., Luijten, M., Martin-Santos, R., Momenan, R., Paulus, M.P., Schmaal, L., Sinha, R., Sjoerds, Z., Solowij, N., Stein, D.J., Stein, E.A., Uhlmann, A., van Holst, R.J., Veltman, D.J., Wiers, R.W., Yücel, M., Zhang, S., Jahanshad, N., Thompson, P.M., Conrod, P., Mackey, S., Garavan, H., 2021. Mapping cortical and subcortical asymmetries in substance dependence: Findings from the ENIGMA Addiction Working Group. Addict. Biol. 26, e13010. https://doi.org/10.1111/adb.13010

Cerdó, T., Diéguez, E., Campoy, C., 2019. Early nutrition and gut microbiome: interrelationship between bacterial metabolism, immune system, brain structure, and neurodevelopment. Am. J. Physiol. Endocrinol. Metab. 317, E617–E630. https://doi.org/10.1152/ajpendo.00188.2019

Chen, X., Ishwaran, H., 2012. Random forests for genomic data analysis. Genomics 99, 323–329. https://doi.org/10.1016/j.ygeno.2012.04.003

Cohen Kadosh, K., Krause, B., King, A.J., Near, J., Cohen Kadosh, R., 2015. Linking GABA and glutamate levels to cognitive skill acquisition during development. Hum. Brain Mapp. 36, 4334–4345. https://doi.org/10.1002/hbm.22921

Cowan, C.S.M., Dinan, T.G., Cryan, J.F., 2020. Annual Research Review: Critical windows - the microbiota-gut-brain axis in neurocognitive development. J. Child Psychol. Psychiatry 61, 353–371. https://doi.org/10.1111/jcpp.13156

Dalile, B., Van Oudenhove, L., Vervliet, B., Verbeke, K., 2019. The role of short-chain fatty acids in microbiota-gut-brain communication. Nat. Rev. Gastroenterol. Hepatol. 16, 461–478. https://doi.org/10.1038/s41575-019-0157-3

Danisch, S., Krumbiegel, J., 2021. Makie.jl: Flexible high-performance data visualization for Julia. J. Open Source Softw. 6, 3349. https://doi.org/10.21105/joss.03349

Deoni, S.C., Beauchemin, J., Volpe, A., Dâ Sa, V., RESONANCE Consortium, 2021. Impact of the COVID-19 pandemic on early child cognitive development: Initial findings in a longitudinal observational study of child health. MedRxiv Prepr. Serv. Health Sci. https://doi.org/10.1101/2021.08.10.21261846

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., Knight, R., 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. U. S. A. 107, 11971–11975. https://doi.org/10.1073/pnas.1002601107

Dong, H., Wang, B., Feng, J., Yue, X., Jia, F., 2021. Correlation between serum concentrations of menaquinone-4 and developmental quotients in children with autism spectrum disorder. Front. Nutr. 8, 748513. https://doi.org/10.3389/fnut.2021.748513

Ernst, M., Nelson, E.E., Jazbec, S., McClure, E.B., Monk, C.S., Leibenluft, E., Blair, J., Pine, D.S., 2005. Amygdala and nucleus accumbens in responses to receipt and omission of gains in adults and adolescents. NeuroImage 25, 1279–1291. https://doi.org/10.1016/j.neuroimage.2004.12.038

Forrest, C.B., Blackwell, C.K., Camargo, C.A., 2018. Advancing the science of children's positive health in the national institutes of health environmental influences on child health outcomes (ECHO) research program. J. Pediatr. 196, 298–300. https://doi.org/10.1016/j.jpeds.2018.02.004

Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker, S., Vatanen, T., Hall, A.B., Mallick, H., McIver, L.J., Sauk, J.S., Wilson, R.G., Stevens, B.W., Scott, J.M., Pierce, K., Deik, A.A., Bullock, K., Imhann, F., Porter, J.A., Zhernakova, A., Fu, J., Weersma, R.K., Wijmenga, C., Clish, C.B., Vlamakis, H., Huttenhower, C., Xavier, R.J., 2019. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat. Microbiol. 4, 293–305. https://doi.org/10.1038/s41564-018-0306-4

Fung, T.C., Olson, C.A., Hsiao, E.Y., 2017. Interactions between the microbiota, immune and nervous systems in health and disease. Nat. Neurosci. 20, 145–155. https://doi.org/10.1038/nn.4476

Ghosh, T.S., Rampelli, S., Jeffery, I.B., Santoro, A., Neto, M., Capri, M., Giampieri, E., Jennings, A., Candela, M., Turroni, S., Zoetendal, E.G., Hermes, G.D.A., Elodie, C., Meunier, N., Brugere, C.M., Pujos-Guillot, E., Berendsen, A.M., Groot, L.C.P.G.M.D., Feskins, E.J.M., Kaluza, J., Pietruszka, B., Bielak, M.J., Comte, B., Maijo-Ferre, M., Nicoletti, C., Vos, W.M.D., Fairweather-Tait, S., Cassidy, A., Brigidi, P., Franceschi, C., O'Toole, P.W., 2020. Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: the NU-AGE 1-year dietary intervention across five European countries. Gut 69, 1218–1228. https://doi.org/10.1136/gutjnl-2019-319654

Gillman, M.W., Blaisdell, C.J., 2018. Environmental influences on child health outcomes, a research program of the NIH. Curr. Opin. Pediatr. 30, 260.

Gong, Q.-Y., Cai, L., Jing, Y., Wang, W., Yang, D.-X., Chen, S.-W., Tian, H.-L., 2022. Urolithin A alleviates blood-brain barrier disruption and attenuates neuronal apoptosis following traumatic brain injury in mice. Neural Regen. Res. 17, 2007. https://doi.org/10.4103/1673-5374.335163

Haam, J., Yakel, J.L., 2017. Cholinergic modulation of the hippocampal region and memory function. J. Neurochem. 142, 111–121. https://doi.org/10.1111/jnc.14052

Haiser, H.J., Gootenberg, D.B., Chatman, K., Sirasani, G., Balskus, E.P., Turnbaugh, P.J., 2013. Predicting and Manipulating Cardiac Drug Inactivation by the Human Gut Bacterium Eggerthella lenta. Science 341, 295–298. https://doi.org/10.1126/science.1235872

Hartstra, A.V., Schüppel, V., Imangaliyev, S., Schrantee, A., Prodan, A., Collard, D., Levin, E., Dallinga-Thie, G., Ackermans, M.T., Winkelmeijer, M., Havik, S.R., Metwaly, A., Lagkouvardos, I., Nier, A., Bergheim, I., Heikenwalder, M., Dunkel, A., Nederveen, A.J., Liebisch, G., Mancano, G., Claus, S.P., Benítez-Páez, A., la Fleur, S.E., Bergman, J.J., Gerdes, V., Sanz, Y., Booij, J., Kemper, E., Groen, A.K., Serlie, M.J., Haller, D., Nieuwdorp, M., 2020. Infusion of donor feces affects the gut–brain axis in humans with metabolic syndrome. Mol. Metab. 42, 101076. https://doi.org/10.1016/j.molmet.2020.101076

Hoogman, M., Bralten, J., Hibar, D.P., Mennes, M., Zwiers, M.P., Schweren, L., van Hulzen, K.J.E., Medland, S.E., Shumskaya, E., Jahanshad, N., de Zeeuw, P., Szekely, E., Sudre, G., Wolfers, T., Onnink, A.M.H., Dammers, J.T., Mostert, J.C., Vives-Gilabert, Y., Kohls, G., Oberwelland, E., Seitz, J., Schulte-Rüther, M., di Bruttopilo, S.A., Doyle, A.E., Høvik, M.F., Dramsdahl, M., Tamm, L., van Erp, T.G.M., Dale, A., Schork, A., Conzelmann, A., Zierhut, K., Baur, R., McCarthy, H., Yoncheva, Y.N., Cubillo, A., Chantiluke, K., Mehta, M.A., Paloyelis, Y., Hohmann, S., Baumeister, S., Bramati, I., Mattos, P., Tovar-Moll, F., Douglas, P., Banaschewski, T., Brandeis, D., Kuntsi, J., Asherson, P., Rubia, K., Kelly, C., Di Martino, A., Milham, M.P., Castellanos, F.X., Frodl, T., Zentis, M., Lesch, K.-P., Reif, A., Pauli, P., Jernigan, T., Haavik, J., Plessen, K.J., Lundervold, A.J., Hugdahl, K., Seidman, L.J., Biederman, J., Rommelse, N., Heslenfeld, D.J., Hartman, C., Hoekstra, P.J., Oosterlaan, J., von Polier, G., Konrad, K., Vilarroya, O., Ramos-Quiroga, J.-A., Soliva, J.C., Durston, S., Buitelaar, J.K., Faraone, S.V., Shaw, P., Thompson, P., Franke, B., 2017. Subcortical brain volume differences of participants with ADHD across the lifespan: an ENIGMA collaboration. Lancet Psychiatry 4, 310–319. https://doi.org/10.1016/S2215-0366(17)30049-4

Hoshino, Y., Yamamoto, T., Kaneko, M., Tachibana, R., Watanabe, M., Ono, Y., Kumashiro, H., 1984. Blood Serotonin and Free Tryptophan Concentration in Autistic Children. Neuropsychobiology 11, 22–27. https://doi.org/10.1159/000118045

Hosomi, K., Saito, M., Park, J., Murakami, H., Shibata, N., Ando, M., Nagatake, T., Konishi, K., Ohno, H., Tanisawa, K., Mohsen, A., Chen, Y.-A., Kawashima, H., Natsume-Kitatani, Y., Oka, Y., Shimizu, H., Furuta, M., Tojima, Y., Sawane, K., Saika, A., Kondo, S., Yonejima, Y., Takeyama, H., Matsutani, A., Mizuguchi, K., Miyachi, M., Kunisawa, J., 2022. Oral administration of Blautia wexlerae ameliorates obesity and type 2 diabetes via metabolic remodeling of the gut microbiota. Nat. Commun. 13, 4477. https://doi.org/10.1038/s41467-022-32015-7

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jiang, H., Ling, Z., Zhang, Y., Mao, H., Ma, Z., Yin, Y., Wang, W., Tang, W., Tan, Z., Shi, J., Li, L., Ruan, B., 2015. Altered fecal microbiota composition in patients with major depressive disorder. Brain. Behav. Immun. 48, 186–194. https://doi.org/10.1016/j.bbi.2015.03.016

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., Hattori, M., 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. 32, D277–D280.

https://doi.org/10.1093/nar/gkh063
Kant, R., Rasinkangas, P., Satokari, R., Pietilä, T.E., Palva, A., 2015. Genome Sequence of the Butyrate-Producing Anaerobic Bacterium Anaerostipes hadrus PEL 85. Genome Announc. 3, e00224-15. https://doi.org/10.1128/genomeA.00224-15
Kim, C.-S., Cha, L., Sim, M., Jung, S., Chun, W.Y., Baik, H.W., Shin, D.-M., 2021. Probiotic supplementation improves cognitive function and mood with changes in gut microbiota in Community-Dwelling older adults: A randomized, Double-Blind, Placebo-Controlled, multicenter trial. J Gerontol Biol Sci Med Sci 76, 32–40. https://doi.org/10.1093/gerona/glaa090
Knudsen, E.I., 2004. Sensitive periods in the development of the brain and behavior. J. Cogn. Neurosci. 16, 1412–1425. https://doi.org/10.1162/0898929042304796
Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., Ley, R.E., 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. U. S. A. 108 Suppl 1, 4578–4585. https://doi.org/10.1073/pnas.1000081107
Laue, H.E., Coker, M.O., Madan, J.C., 2022. The developing microbiome from birth to 3 years: The Gut-Brain axis and neurodevelopmental outcomes. Front. Pediatr. 10, 815885. https://doi.org/10.3389/fped.2022.815885
Laue, H.E., Korrick, S.A., Baker, E.R., Karagas, M.R., Madan, J.C., 2020. Prospective associations of the infant gut microbiome and microbial function with social behaviors related to autism at age 3 years. Sci. Rep. 10, 15515. https://doi.org/10.1038/s41598-020-72386-9
Ley, R.E., Turnbaugh, P.J., Klein, S., Gordon, J.I., 2006. Human gut microbes associated with obesity. Nature 444, 1022–1023. https://doi.org/10.1038/4441022a
Li, Y.-M., Yu, Q., Huang, L., Fu, J., Feng, R.-J., 2022. [Correlations Between Gut Microbiota Changes and Cognitive Function in Patients with Post-Stroke Cognitive Impairment in the Early Stage]. Sichuan Da Xue Xue Bao Yi Xue Ban 53, 857–865. https://doi.org/10.12182/20220960105
Liu, S., Li, E., Sun, Z., Fu, D., Duan, G., Jiang, M., Yu, Y., Mei, L., Yang, P., Tang, Y., Zheng, P., 2019. Altered gut microbiota and short chain fatty acids in Chinese children with autism spectrum disorder. Sci. Rep. 9, 287. https://doi.org/10.1038/s41598-018-36430-z
Liu, X., Mao, B., Gu, J., Wu, J., Cui, S., Wang, G., Zhao, J., Zhang, H., Chen, W., 2021. *Blautia*—a new functional genus with potential probiotic properties? Gut Microbes 13, 1875796. https://doi.org/10.1080/19490976.2021.1875796
Lopez-Siles, M., Duncan, S.H., Garcia-Gil, L.J., Martinez-Medina, M., 2017. Faecalibacterium prausnitzii: from microbiology to diagnostics and prognostics. ISME J. 11, 841–852. https://doi.org/10.1038/ismej.2016.176
Louwies, T., Johnson, A.C., Orock, A., Yuan, T., Greenwood-Van Meerveld, B., 2020. The microbiota-gut-brain axis: An emerging role for the epigenome. Exp. Biol. Med. 245, 138–145. https://doi.org/10.1177/1535370219891690
Magnusson, K.R., Hauck, L., Jeffrey, B.M., Elias, V., Humphrey, A., Nath, R., Perrone, A., Bermudez, L.E., 2015. Relationships between diet-related changes in the gut microbiome and cognitive flexibility. Neuroscience 300, 128–140. https://doi.org/10.1016/j.neuroscience.2015.05.016
Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren, B., Schwager, E.H., Chatterjee, S., Thompson, K.N., Wilkinson, J.E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J.N., Franzosa, E.A., Bravo, H.C., Huttenhower, C., 2021. Multivariable association discovery in population-scale meta-omics studies. https://doi.org/10.1101/2021.01.20.427420
Maruo, T., Sakamoto, M., Ito, C., Toda, T., Benno, Y., 2008. Adlercreutzia equolifaciens gen. nov., sp. nov., an equol-producing bacterium isolated from human faeces, and emended

description of the genus Eggerthella. Int. J. Syst. Evol. Microbiol. 58, 1221–1227. https://doi.org/10.1099/ijs.0.65404-0

Mayneris-Perxachs, J., Castells-Nobau, A., Arnoriaga-Rodríguez, M., Martin, M., de la Vega-Correa, L., Zapata, C., Burokas, A., Blasco, G., Coll, C., Escrichs, A., Biarnés, C., Moreno-Navarrete, J.M., Puig, J., Garre-Olmo, J., Ramos, R., Pedraza, S., Brugada, R., Vilanova, J.C., Serena, J., Gich, J., Ramió-Torrentà, L., Pérez-Brocal, V., Moya, A., Pamplona, R., Sol, J., Jové, M., Ricart, W., Portero-Otin, M., Deco, G., Maldonado, R., Fernández-Real, J.M., 2022. Microbiota alterations in proline metabolism impact depression. Cell Metab. 34, 681-701.e10. https://doi.org/10.1016/j.cmet.2022.04.001

Mukherjee, A., Lordan, C., Ross, R.P., Cotter, P.D., 2020. Gut microbes from the phylogenetically diverse genus Eubacterium and their various contributions to gut health. Gut Microbes 12, 1802866. https://doi.org/10.1080/19490976.2020.1802866

Mullen, E.M., others, 1995. Mullen scales of early learning. AGS Circle Pines, MN.

Needham, B.D., Funabashi, M., Adame, M.D., Wang, Z., Boktor, J.C., Haney, J., Wu, W.-L., Rabut, C., Ladinsky, M.S., Hwang, S.-J., Guo, Y., Zhu, Q., Griffiths, J.A., Knight, R., Bjorkman, P.J., Shapiro, M.G., Geschwind, D.H., Holschneider, D.P., Fischbach, M.A., Mazmanian, S.K., 2022. A gut-derived metabolite alters brain activity and anxiety behaviour in mice. Nature 602, 647–653. https://doi.org/10.1038/s41586-022-04396-8

Opel, N., Thalamuthu, A., Milaneschi, Y., Grotegerd, D., Flint, C., Leenings, R., Goltermann, J., Richter, M., Hahn, T., Woditsch, G., Berger, K., Hermesdorf, M., McIntosh, A., Whalley, H.C., Harris, M.A., MacMaster, F.P., Walter, H., Veer, I.M., Frodl, T., Carballedo, A., Krug, A., Nenadic, I., Kircher, T., Aleman, A., Groenewold, N.A., Stein, D.J., Soares, J.C., Zunta-Soares, G.B., Mwangi, B., Wu, M.-J., Walter, M., Li, M., Harrison, B.J., Davey, C.G., Cullen, K.R., Klimes-Dougan, B., Mueller, B.A., Sämann, P.G., Penninx, B., Nawijn, L., Veltman, D.J., Aftanas, L., Brak, I.V., Filimonova, E.A., Osipov, E.A., Reneman, L., Schrantee, A., Grabe, H.J., Van der Auwera, S., Wittfeld, K., Hosten, N., Völzke, H., Sim, K., Gotlib, I.H., Sacchet, M.D., Lagopoulos, J., Hatton, S.N., Hickie, I., Pozzi, E., Thompson, P.M., Jahanshad, N., Schmaal, L., Baune, B.T., Dannlowski, U., 2021. Brain structural abnormalities in obesity: relation to age, genetic risk, and common psychiatric disorders. Mol. Psychiatry 26, 4839–4852. https://doi.org/10.1038/s41380-020-0774-9

Pagonabarraga, J., Corcuera-Solano, I., Vives-Gilabert, Y., Llebaria, G., García-Sánchez, C., Pascual-Sedano, B., Delfino, M., Kulisevsky, J., Gómez-Ansón, B., 2013. Pattern of Regional Cortical Thinning Associated with Cognitive Deterioration in Parkinson's Disease. PLoS ONE 8, e54980. https://doi.org/10.1371/journal.pone.0054980

Palomo-Buitrago, M.E., Sabater-Masdeu, M., Moreno-Navarrete, J.M., Caballano-Infantes, E., Arnoriaga-Rodríguez, M., Coll, C., Ramió, L., Palomino-Schätzlein, M., Gutiérrez-Carcedo, P., Pérez-Brocal, V., Simó, R., Moya, A., Ricart, W., Herance, J.R., Fernández-Real, J.M., 2019. Glutamate interactions with obesity, insulin resistance, cognition and gut microbiota composition. Acta Diabetol. 56, 569–579. https://doi.org/10.1007/s00592-019-01313-w

Pronovost, G.N., Hsiao, E.Y., 2019. Perinatal interactions between the microbiome, immunity, and neurodevelopment. Immunity 50, 18–36. https://doi.org/10.1016/j.immuni.2018.11.016

Sadeghi, B., Chiarawongse, P., Squire, K., Jones, D.C., Noack, A., St-Jean, C., Huijzer, R., Schätzle, R., Butterworth, I., Peng, Y.-F., Blaom, A., 2022. DecisionTree.jl - A julia implementation of the CART decision tree and random forest algorithms. https://doi.org/10.5281/zenodo.7359268

Sela, D.A., Chapman, J., Adeuya, A., Kim, J.H., Chen, F., Whitehead, T.R., Lapidus, A., Rokhsar, D.S., Lebrilla, C.B., German, J.B., Price, N.P., Richardson, P.M., Mills, D.A., 2008. The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. Proc. Natl. Acad. Sci. 105,

18964–18969. https://doi.org/10.1073/pnas.0809584105
Selma, M.V., Beltrán, D., García-Villalba, R., Espín, J.C., Tomás-Barberán, F.A., 2014. Description of urolithin production capacity from ellagic acid of two human intestinal Gordonibacter species. Food Funct. 5, 1779–1784. https://doi.org/10.1039/c4fo00092g
Setchell, K.D., Clerici, C., Lephart, E.D., Cole, S.J., Heenan, C., Castellani, D., Wolfe, B.E., Nechemias-Zimmer, L., Brown, N.M., Lund, T.D., Handa, R.J., Heubi, J.E., 2005. S-Equol, a potent ligand for estrogen receptor β, is the exclusive enantiomeric form of the soy isoflavone metabolite produced by human intestinal bacterial flora. Am. J. Clin. Nutr. 81, 1072–1079. https://doi.org/10.1093/ajcn/81.5.1072
Sharon, G., Sampson, T.R., Geschwind, D.H., Mazmanian, S.K., 2016. The central nervous system and the gut microbiome. Cell 167, 915–932. https://doi.org/10.1016/j.cell.2016.10.027
Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., Sestan, N., 2016. The cellular and molecular landscapes of the developing human central nervous system. Neuron 89, 248–268.
Sordillo, J.E., Korrick, S., Laranjo, N., Carey, V., Weinstock, G.M., Gold, D.R., O'Connor, G., Sandel, M., Bacharier, L.B., Beigelman, A., Zeiger, R., Litonjua, A.A., Weiss, S.T., 2019. Association of the infant gut microbiome with early childhood neurodevelopmental outcomes: An ancillary study to the VDAART randomized clinical trial. JAMA Netw. Open 2, e190905. https://doi.org/10.1001/jamanetworkopen.2019.0905
Spichak, S., Bastiaanssen, T.F.S., Berding, K., Vlckova, K., Clarke, G., Dinan, T.G., Cryan, J.F., 2021. Mining microbes for mental health: Determining the role of microbial metabolic pathways in human brain health and disease. Neurosci. Biobehav. Rev. 125, 698–761. https://doi.org/10.1016/j.neubiorev.2021.02.044
Stephan, J., Stegle, O., Beyer, A., 2015. A random forest approach to capture genetic effects in the presence of population structure. Nat. Commun. 6, 7432. https://doi.org/10.1038/ncomms8432
Suzek, B.E., Huang, H., McGarvey, P., Mazumder, R., Wu, C.H., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23, 1282–1288. https://doi.org/10.1093/bioinformatics/btm098
Takahashi, H., Yang, J., Yamamoto, H., Fukuda, S., Arakawa, K., 2021. Complete Genome Sequence of Adlercreutzia equolifaciens subsp. celatus DSM 18785. Microbiol. Resour. Announc. 10, e00354-21. https://doi.org/10.1128/MRA.00354-21
Tau, G.Z., Peterson, B.S., 2010. Normal development of brain circuits. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol. 35, 147–168. https://doi.org/10.1038/npp.2009.115
Thawornkuno, C., Tanaka, M., Sone, T., Asano, K., 2009. Biotransformation of Daidzein to Equol by Crude Enzyme from Asaccharobacter celatus AHU1763 Required an Anaerobic Environment. Biosci. Biotechnol. Biochem. 73, 1435–1438. https://doi.org/10.1271/bbb.80908
Tognini, P., 2017. Gut microbiota: a potential regulator of neurodevelopment. Front. Cell. Neurosci. 11, 25. https://doi.org/10.3389/fncel.2017.00025
Tso, L., Bonham, K.S., Fishbein, A., Rowland, S., Klepac-Ceraj, V., 2021. Targeted High-Resolution taxonomic identification of bifidobacterium longum subsp. infantis using human milk oligosaccharide metabolizing genes. Nutrients 13, 2833. https://doi.org/10.3390/nu13082833
Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., Claes, S., Van Oudenhove, L., Zhernakova, A., Vieira-Silva, S., Raes, J., 2019. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat. Microbiol. 4, 623–632. https://doi.org/10.1038/s41564-018-0337-x
Velazquez, M., Lee, Y., Initiative, for the A.D.N., 2021. Random forest model for feature-based

Alzheimer's disease conversion prediction from early mild cognitive impairment subjects. PLOS ONE 16, e0244773. https://doi.org/10.1371/journal.pone.0244773

Wacker, J., Dillon, D.G., Pizzagalli, D.A., 2009. The role of the nucleus accumbens and rostral anterior cingulate cortex in anhedonia: Integration of resting EEG, fMRI, and volumetric techniques. NeuroImage 46, 327–337. https://doi.org/10.1016/j.neuroimage.2009.01.058

Wan, Y., Zuo, T., Xu, Z., Zhang, F., Zhan, H., Chan, D., Leung, T.-F., Yeoh, Y.K., Chan, F.K.L., Chan, R., Ng, S.C., 2021. Underdevelopment of the gut microbiota and bacteria species as non-invasive markers of prediction in children with autism spectrum disorder. Gut gutjnl-2020-324015. https://doi.org/10.1136/gutjnl-2020-324015

Wang, L.-J., Yang, C.-Y., Chou, W.-J., Lee, M.-J., Chou, M.-C., Kuo, H.-C., Yeh, Y.-M., Lee, S.-Y., Huang, L.-H., Li, S.-C., 2020. Gut microbiota and dietary patterns in children with attention-deficit/hyperactivity disorder. Eur. Child Adolesc. Psychiatry 29, 287–297. https://doi.org/10.1007/s00787-019-01352-2

Wang, X.-L., Hur, H.-G., Lee, J.H., Kim, K.T., Kim, S.-I., 2005. Enantioselective Synthesis of S-Equol from Dihydrodaidzein by a Newly Isolated Anaerobic Human Intestinal Bacterium. Appl. Environ. Microbiol. 71, 214–219. https://doi.org/10.1128/AEM.71.1.214-219.2005

Wechsler, D., 2012. Wechsler preschool and primary scale of intelligence—fourth edition. Psychol. Corp. San Antonio TX.

Wechsler, D., 1949. Wechsler intelligence scale for children.

Woodruff, M.C., Bonham, K.S., Anam, F.A., Walker, T.A., Ishii, Y., Kaminski, C.Y., Ruunstrom, M.C., Truong, A.D., Dixit, A.N., Han, J.E., Ramonell, R.P., Haddad, N.S., Rudolph, M.E., Jenks, S.A., Khosroshahi, A., Lee, F.E.-H., Sanz, I., 2022. Inflammation and autoreactivity define a discrete subset of patients with post-acute sequelae of COVID-19, or long-COVID. https://doi.org/10.1101/2021.09.21.21263845

Xiao, L., Yan, J., Yang, T., Zhu, J., Li, T., Wei, H., Chen, J., 2021. Fecal Microbiome Transplantation from Children with Autism Spectrum Disorder Modulates Tryptophan and Serotonergic Synapse Metabolism and Induces Altered Behaviors in Germ-Free Mice. mSystems 6, e01343-20. https://doi.org/10.1128/mSystems.01343-20

Xiao, X., Yang, Y., Zhang, Y., Zhang, X.-M., Zhao, Z.-Q., Zhang, Y.-Q., 2013. Estrogen in the Anterior Cingulate Cortex Contributes to Pain-Related Aversion. Cereb. Cortex 23, 2190–2203. https://doi.org/10.1093/cercor/bhs201

Yael, D., Tahary, O., Gurovich, B., Belelovsky, K., Bar-Gad, I., 2019. Disinhibition of the Nucleus Accumbens Leads to Macro-Scale Hyperactivity Consisting of Micro-Scale Behavioral Segments Encoded by Striatal Activity. J. Neurosci. 39, 5897–5909. https://doi.org/10.1523/JNEUROSCI.3120-18.2019

Yau, W.-Y.W., Zubieta, J.-K., Weiland, B.J., Samudra, P.G., Zucker, R.A., Heitzeg, M.M., 2012. Nucleus Accumbens Response to Incentive Stimuli Anticipation in Children of Alcoholics: Relationships with Precursive Behavioral Risk and Lifetime Alcohol Use. J. Neurosci. 32, 2544–2551. https://doi.org/10.1523/JNEUROSCI.1390-11.2012

Figures

Figure 1: Cohort overview and summary data.

(A) Study design and overview. Stool samples, cognitive assessments and neuroimaging were collected from participants in an accelerated longitudinal design. (B) Cognitive function scores are assessed by different instruments, but may be normalized to a common IQ-like scale. (C-D) Principal coordinates analysis using Bray-Curtis dissimilarity on taxonomic profiles demonstrates high beta diversity, with much of the first axis of variation explained by increasing age and alpha diversity. Differences in gene function profiles (E) and neuroimaging (PCA based on the Euclidean distance of brain region volumes) (F) are likewise dominated by changes as children age. (G) Permutation analysis of variance (PERMANOVA) of taxonomic profiles, functional profiles (annotated as UniRef90s) and neuroimaging, against the metadata of interest. Variations in taxonomic and functional profiles explain a modest but significant percent of the variation in cognitive development in children over 18 months of age. (H) Mantel testing of different microbial feature matrices, shows overlapping but distinct patterns of variation. Dotted lines in (B) and (C) show 6 and 18 months, which are used as cut-offs in some following models.

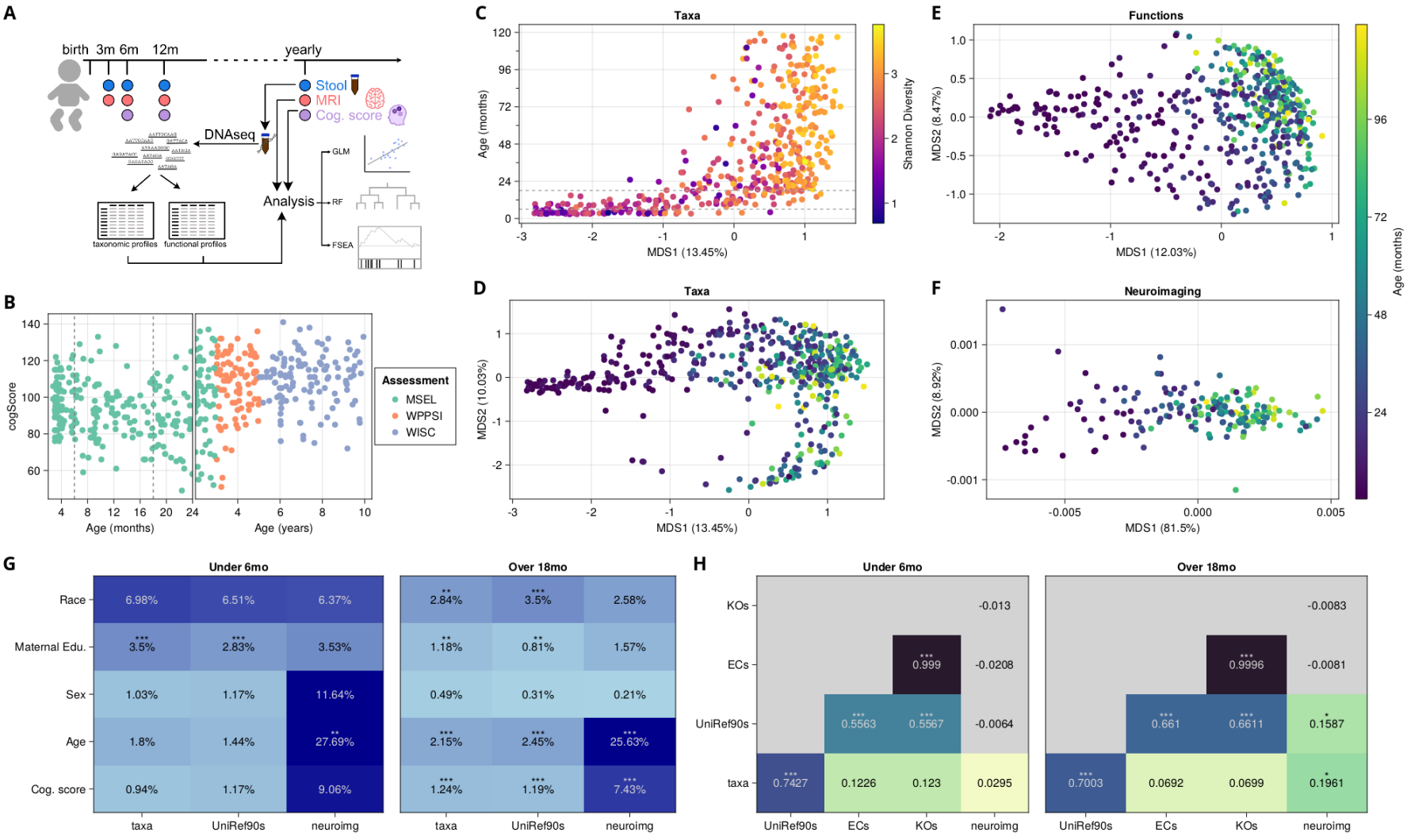

A
birth 3m 6m 12m
yearly
Stool
MRI
Cog. score
DNAseq
Analysis
GLM
RF
FSEA
taxonomic profiles
functional profiles
B
cogScore
Age (months)
Age (years)
Assessment
MSEL
WPPSI
WISC
C
Taxa
Age (months)
MDS1 (13.45%)
Shannon Diversity
D
Taxa
MDS2 (10.03%)
MDS1 (13.45%)
E
Functions
MDS2 (8.47%)
MDS1 (12.03%)
Age (months)
F
Neuroimaging
MDS2 (8.92%)
MDS1 (81.5%)
G
Under 6mo
Over 18mo
Race
Maternal Edu.
Sex
Age
Cog. score
6.98%
6.51%
6.37%
3.5%
2.83%
3.53%
1.03%
1.17%
11.64%
1.8%
1.44%
27.69%
0.94%
1.17%
9.06%
2.84%
3.5%
2.58%
1.18%
0.81%
1.57%
0.49%
0.31%
0.21%
2.15%
2.45%
25.63%
1.24%
1.19%
7.43%
taxa
UniRef90s
neuroimg
H
Under 6mo
Over 18mo
KOs
ECs
UniRef90s
taxa
0.999
0.5563
0.5567
0.7427
0.1226
0.123
0.0295
-0.013
-0.0208
-0.0064
0.9996
0.661
0.6611
0.7003
0.0692
0.0699
0.1961
-0.0083
-0.0081
0.1587
neuroimg

Figure 2 - Taxa and gene functional groups are associated with cognitive function.

(A) Volcano plot of multivariable linear model results showing the relationship between individual taxa and cognitive function score in children over 18 months of age, controlling for age, maternal education, and sequencing depth. Taxa that were significant after FDR correction ($q < 0.2$) are colored red. (B) For taxa that were significantly associated with cognitive function, heatmaps of prevalence, mean relative abundance, and correlation with cognitive function in different age groups. (C-F) Enrichment plots for selected neuroactive gene sets used in feature set enrichment analysis (FSEA). (G) summary of FSEA results across all samples (left) as well as in the under 6 months (middle) and over 18 months (right) subsets, colored based on the significance of the association. Markers indicate the individual correlation of genes within a gene set, and vertical bars represent the median correlation of that gene set.

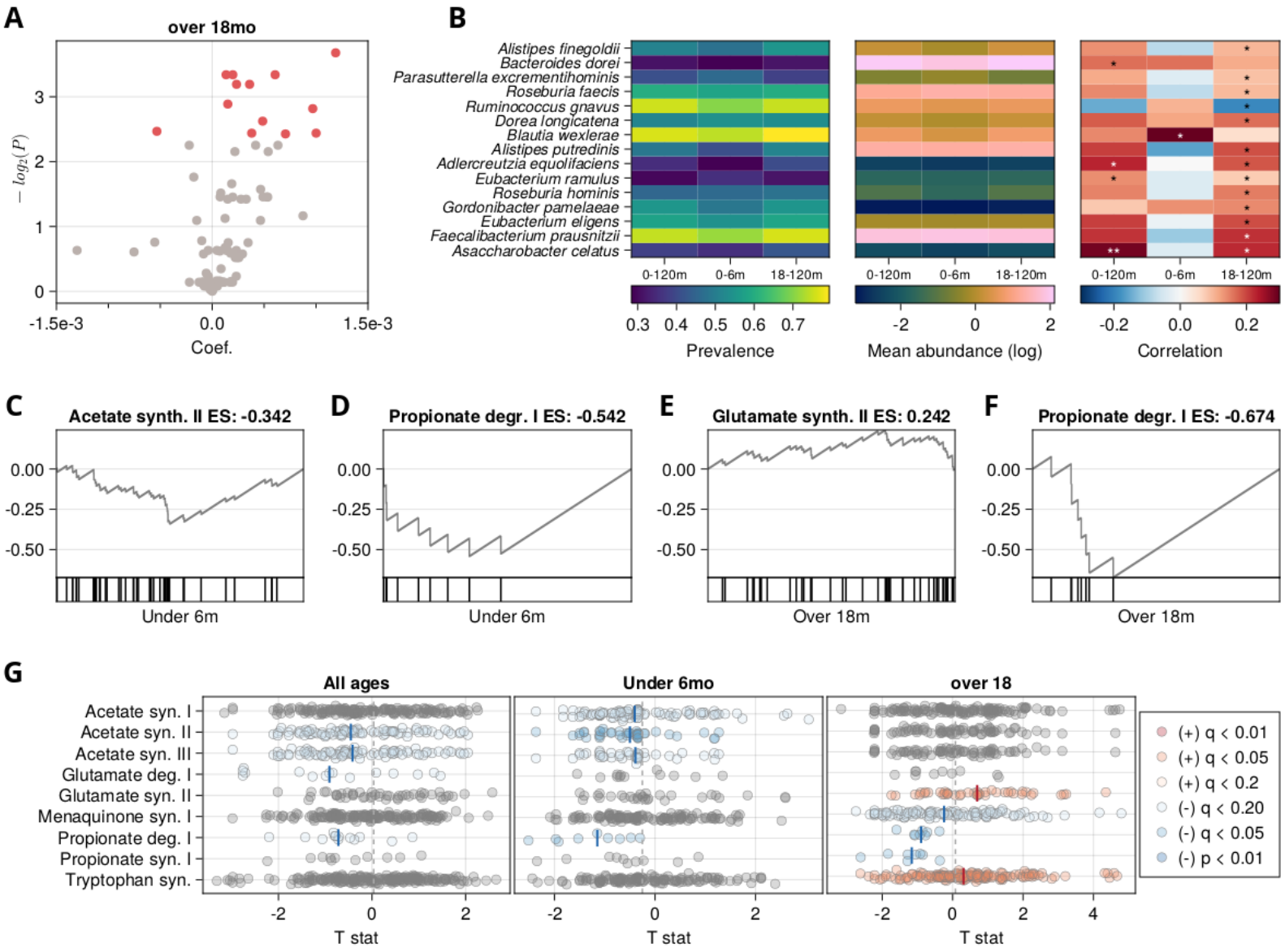

A
over 18mo
$-log_2(P)$
Coef.
-1.5e-3
0.0
1.5e-3
B
Alistipes finegoldii
Bacteroides dorei
Parasutterella excrementihominis
Roseburia faecis
Ruminococcus gnavus
Dorea longicatena
Blautia wexlerae
Alistipes putredinis
Adlercreutzia equolifaciens
Eubacterium ramulus
Roseburia hominis
Gordonibacter pamelaeae
Eubacterium eligens
Faecalibacterium prausnitzii
Asaccharobacter celatus
0-120m
0-6m
18-120m
Prevalence
Mean abundance (log)
Correlation
C
Acetate synth. II ES: -0.342
Under 6m
D
Propionate degr. I ES: -0.542
Under 6m
E
Glutamate synth. II ES: 0.242
Over 18m
F
Propionate degr. I ES: -0.674
Over 18m
G
All ages
Under 6mo
over 18
Acetate syn. I
Acetate syn. II
Acetate syn. III
Glutamate deg. I
Glutamate syn. II
Menaquinone syn. I
Propionate deg. I
Propionate syn. I
Tryptophan syn.
T stat
(+) q < 0.01
(+) q < 0.05
(+) q < 0.2
(-) q < 0.20
(-) q < 0.05
(-) p < 0.01

Figure 3 - Random forest models predict concurrently measured cognitive function.

Comparison of RF predictor importance versus linear models for children between birth and 6 months old (A), and for those older than 18 months (B). Colors represent whether the species belong to the group of top-important features that account for 60% of the cumulative importance on the RF model, if that species was significant ($q < 0.2$) in linear models, both, or neither. Ranked feature importance for taxa in RF models for children between birth and 6 months old (C), and for those older than 18 months (D). Taxa that are important for RF models both for children under 6 months and for those over 18 months (E), only for children under 6 months (F), and only for children over 18 months (G)

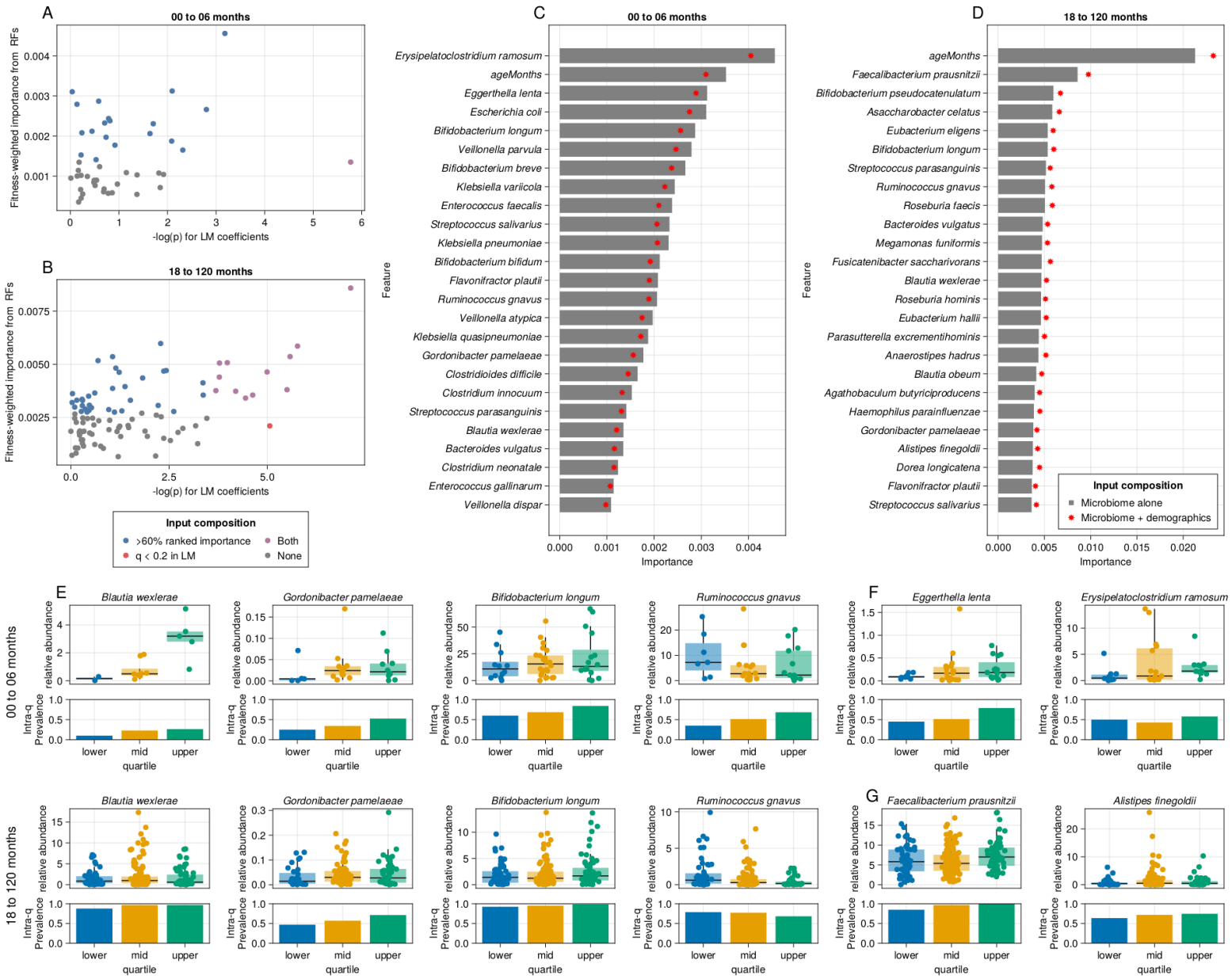

A
00 to 06 months
Fitness-weighted importance from RFs
-log(p) for LM coefficients
B
18 to 120 months
Fitness-weighted importance from RFs
-log(p) for LM coefficients
Input composition
>60% ranked importance
q < 0.2 in LM
Both
None
C
00 to 06 months
Erysipelatoclostridium ramosum
ageMonths
Eggerthella lenta
Escherichia coli
Bifidobacterium longum
Veillonella parvula
Bifidobacterium breve
Klebsiella variicola
Enterococcus faecalis
Streptococcus salivarius
Klebsiella pneumoniae
Bifidobacterium bifidum
Flavonifractor plautii
Ruminococcus gnavus
Veillonella atypica
Klebsiella quasipneumoniae
Gordonibacter pamelaeae
Clostridioides difficile
Clostridium innocuum
Streptococcus parasanguinis
Blautia wexlerae
Bacteroides vulgatus
Clostridium neonatale
Enterococcus gallinarum
Veillonella dispar
Feature
Importance
D
18 to 120 months
ageMonths
Faecalibacterium prausnitzii
Bifidobacterium pseudocatenulatum
Asaccharobacter celatus
Eubacterium eligens
Bifidobacterium longum
Streptococcus parasanguinis
Ruminococcus gnavus
Roseburia faecis
Bacteroides vulgatus
Megamonas funiformis
Fusicatenibacter saccharivorans
Blautia wexlerae
Roseburia hominis
Eubacterium hallii
Parasutterella excrementihominis
Anaerostipes hadrus
Blautia obeum
Agathobaculum butyriciproducens
Haemophilus parainfluenzae
Gordonibacter pamelaeae
Alistipes finegoldii
Dorea longicatena
Flavonifractor plautii
Streptococcus salivarius
Feature
Importance
Input composition
Microbiome alone
Microbiome + demographics
E
F
G
Blautia wexlerae
Gordonibacter pamelaeae
Bifidobacterium longum
Ruminococcus gnavus
Eggerthella lenta
Erysipelatoclostridium ramosum
Faecalibacterium prausnitzii
Alistipes finegoldii
00 to 06 months
18 to 120 months
relative abundance
Intra-q Prevalence
lower
mid
upper
quartile

Figure 4: Microbial feature importance in predicting brain volumes in children over 18 months of age.

(A) Average test-set correlations for prediction of MRI segmentation data from microbiome and demographics on select segments after repeated cross validation. (B) Heatmap of average individual relative taxa importances on each brain segment. Importances are reported as proportions relative to the sum of importances for each model - since every model is trained with 132 features, and even distribution of importance would be 0.75% for each feature. Segments and taxa ordered by HCA on a select list of species with high ranks on average importances, and their respective highest-load segments. (C)

Supplementary Figures

Figure S1 - Samples collected at different ages

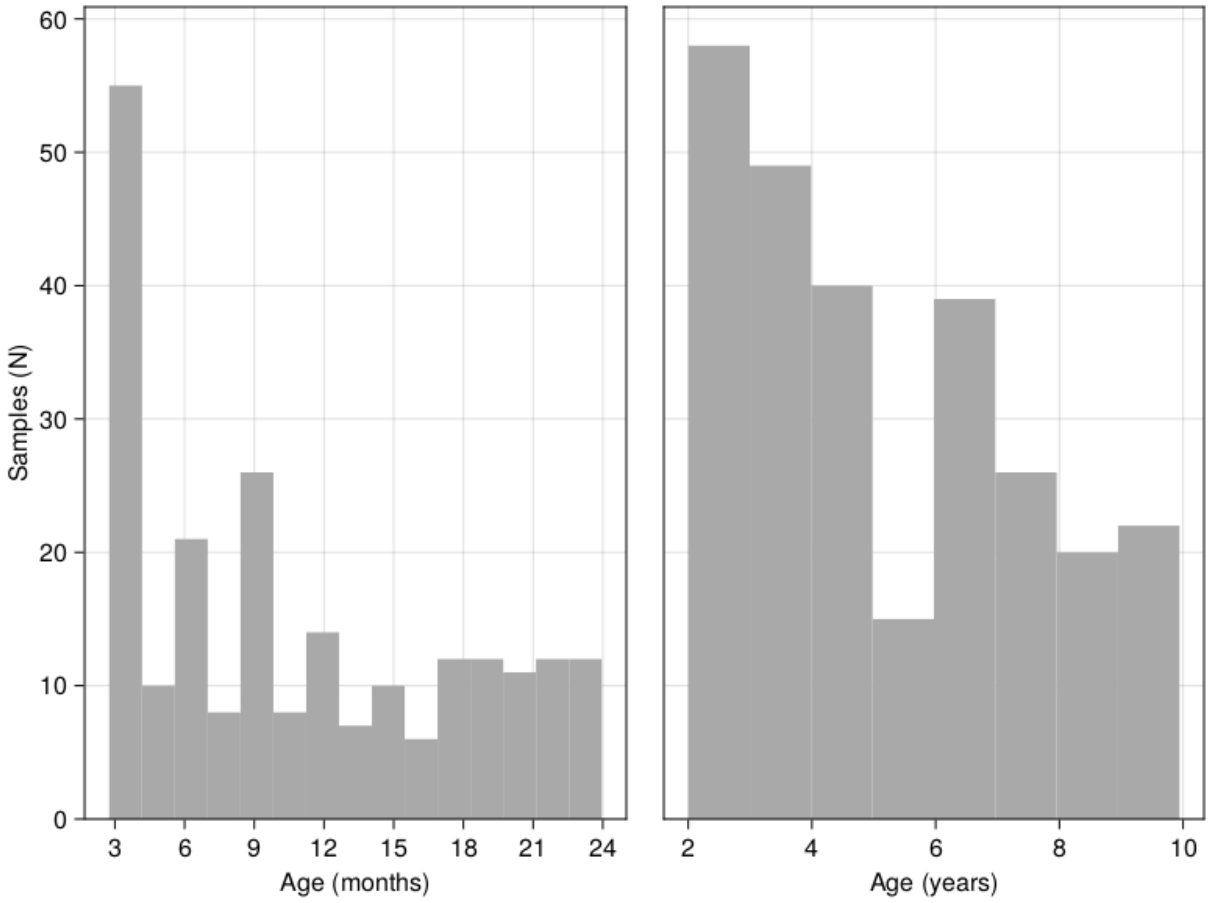


Figure S2 - Relative abundance of select phyla and genera

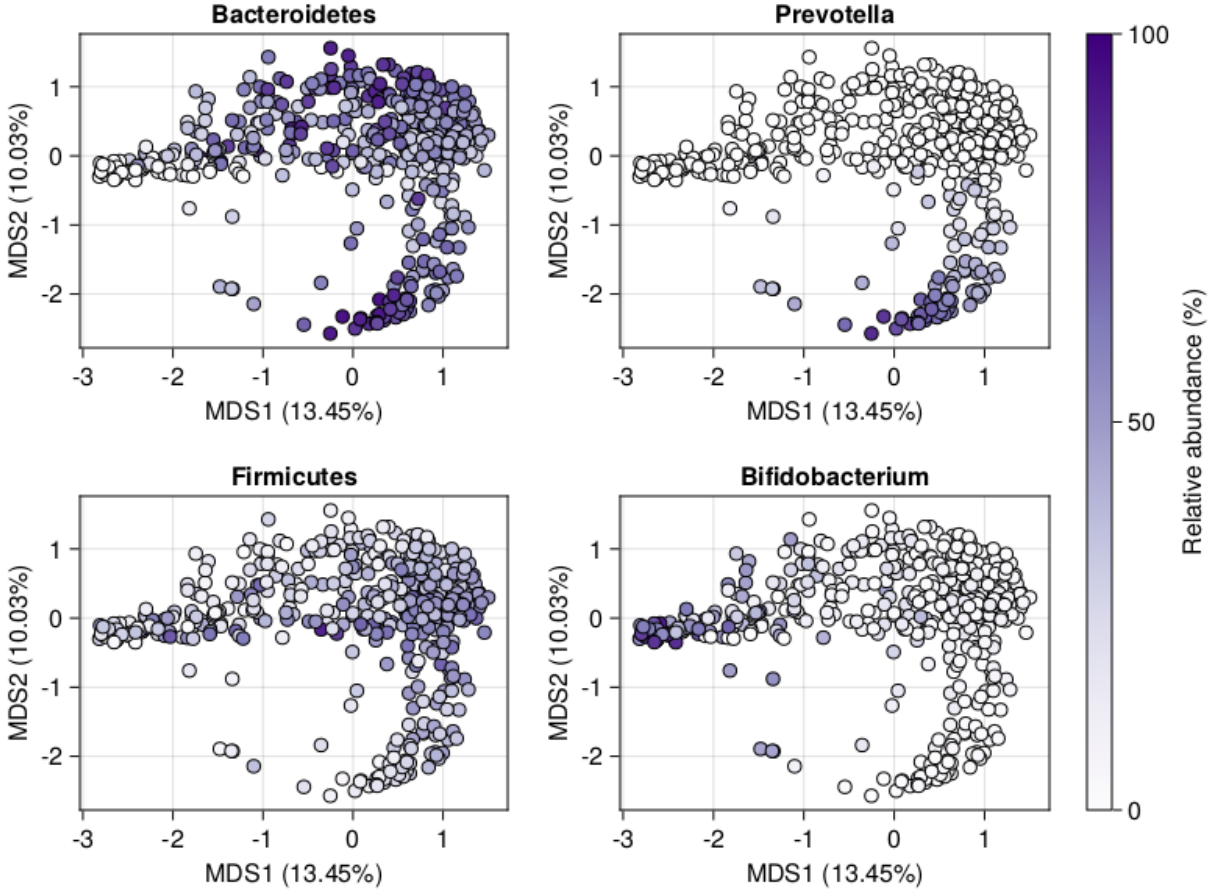
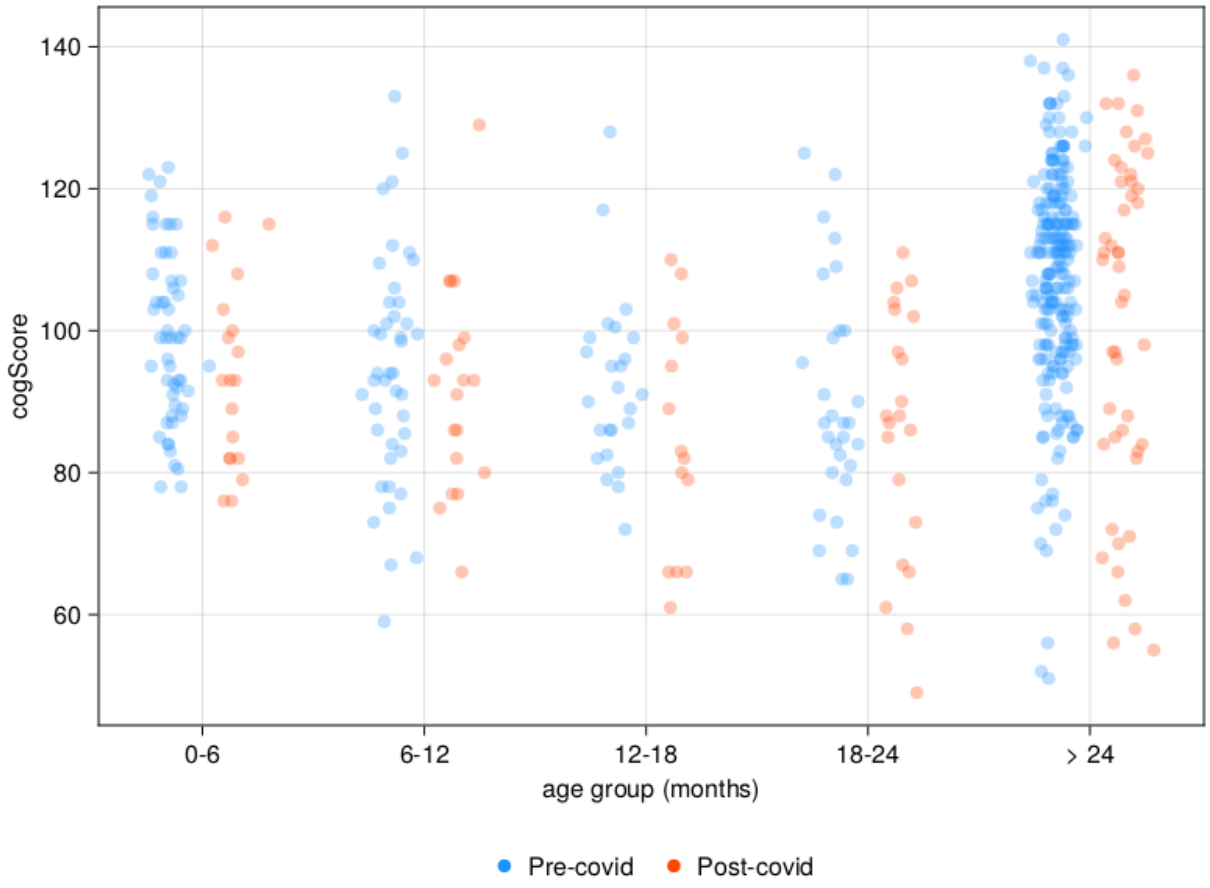


Figure S3 - Cognitive function scores in different age groups collected before or after March 2020.



Supplementary Tables

Supplementary Table XXA. List of most important taxa for Random Forest models on the 0 to 6 months age bracket, microbiome + demographics, ordered by relative weighted importance rank

Rank	variable	Average fitness-weighted importance	Relative fitness-weighted importance	Rank-cumulative relative importance
1	*Erysipelatoclostridium ramosum*	4.56E-03	5.57%	5.57%
2	ageMonths	3.52E-03	4.31%	9.88%
3	*Eggerthella lenta*	3.12E-03	3.82%	13.70%
4	*Escherichia coli*	3.10E-03	3.80%	17.50%
5	*Bifidobacterium longum*	2.87E-03	3.51%	21.01%
6	*Veillonella parvula*	2.79E-03	3.42%	24.43%
7	*Bifidobacterium breve*	2.66E-03	3.26%	27.68%
8	*Klebsiella variicola*	2.44E-03	2.98%	30.66%
9	*Enterococcus faecalis*	2.38E-03	2.91%	33.57%
10	*Streptococcus salivarius*	2.33E-03	2.84%	36.42%
11	*Klebsiella pneumoniae*	2.31E-03	2.82%	39.24%
12	*Bifidobacterium bifidum*	2.12E-03	2.59%	41.83%
13	*Flavonifractor plautii*	2.08E-03	2.55%	44.38%
14	*Ruminococcus gnavus*	2.06E-03	2.52%	46.90%
15	*Veillonella atypica*	1.97E-03	2.41%	49.31%
16	*Klebsiella quasipneumoniae*	1.87E-03	2.29%	51.60%
17	*Gordonibacter pamelaeae*	1.77E-03	2.17%	53.77%
18	*Clostridioides difficile*	1.65E-03	2.02%	55.79%
19	*Clostridium innocuum*	1.53E-03	1.87%	57.66%

20	*Streptococcus parasanguinis*	1.41E-03	1.73%	59.38%
21	*Blautia wexlerae*	1.35E-03	1.65%	61.03%
22	*Bacteroides vulgatus*	1.35E-03	1.65%	62.68%
23	*Clostridium neonatale*	1.24E-03	1.51%	64.19%
24	*Enterococcus gallinarum*	1.14E-03	1.40%	65.59%
25	*Veillonella dispar*	1.09E-03	1.33%	66.92%
26	*Collinsella aerofaciens*	1.08E-03	1.32%	68.24%
27	*Clostridium paraputrificum*	1.04E-03	1.27%	69.51%
28	*Parabacteroides distasonis*	1.03E-03	1.26%	70.78%
29	*Collinsella stercoris*	1.02E-03	1.25%	72.02%
30	*Intestinibacter bartlettii*	1.00E-03	1.22%	73.25%
31	*Bacteroides uniformis*	9.93E-04	1.21%	74.46%
32	*Streptococcus mitis*	9.52E-04	1.16%	75.63%
33	*Bacteroides ovatus*	9.00E-04	1.10%	76.73%
34	*Klebsiella oxytoca*	8.97E-04	1.10%	77.82%
35	*Hungatella hathewayi*	8.43E-04	1.03%	78.86%
36	*Veillonella infantium*	8.01E-04	0.98%	79.84%
37	*Clostridium sp 7 2 43FAA*	7.74E-04	0.95%	80.78%
38	*Haemophilus parainfluenzae*	7.67E-04	0.94%	81.72%
39	*Parabacteroides merdae*	7.16E-04	0.88%	82.60%
40	*Clostridium bolteae*	6.81E-04	0.83%	83.43%
41	*Clostridium perfringens*	6.76E-04	0.83%	84.26%
42	*Bifidobacterium adolescentis*	6.67E-04	0.82%	85.07%
43	*Bifidobacterium pseudocatenulatum*	6.09E-04	0.75%	85.82%

44	*Enterococcus avium*	5.86E-04	0.72%	86.53%
45	*Bacteroides thetaiotaomicron*	5.72E-04	0.70%	87.23%
46	*Akkermansia muciniphila*	5.55E-04	0.68%	87.91%
47	*Bacteroides caccae*	5.46E-04	0.67%	88.58%
48	*Citrobacter youngae*	5.18E-04	0.63%	89.21%
49	*Sellimonas intestinalis*	5.10E-04	0.62%	89.84%
50	*Enterobacter cloacae complex*	4.94E-04	0.60%	90.44%
51	*Streptococcus thermophilus*	4.90E-04	0.60%	91.04%
52	*Lactobacillus rhamnosus*	4.75E-04	0.58%	91.62%
53	*Bacteroides stercoris*	4.56E-04	0.56%	92.18%
54	*Klebsiella michiganensis*	4.51E-04	0.55%	92.73%
55	*Enterococcus faecium*	3.64E-04	0.45%	93.18%
56	*Veillonella sp T11011 6*	3.55E-04	0.43%	93.61%
57	*Streptococcus vestibularis*	3.48E-04	0.43%	94.04%
58	*Ruthenibacterium lactatiformans*	3.41E-04	0.42%	94.45%
59	*Bacteroides fragilis*	3.31E-04	0.41%	94.86%
60	*Anaerostipes caccae*	3.24E-04	0.40%	95.25%
61	*Enterococcus casseliflavus*	3.22E-04	0.39%	95.65%
62	*Citrobacter freundii*	3.16E-04	0.39%	96.03%
63	*Proteus mirabilis*	3.16E-04	0.39%	96.42%
64	*Collinsella intestinalis*	3.06E-04	0.37%	96.80%
65	*Varibaculum cambriense*	3.00E-04	0.37%	97.16%
66	*Bilophila wadsworthia*	2.75E-04	0.34%	97.50%
67	*Bifidobacterium scardovii*	2.48E-04	0.30%	97.80%

68	*Aeriscardovia aeriphila*	2.39E-04	0.29%	98.10%
69	*Morganella morganii*	2.16E-04	0.26%	98.36%
70	*Bacteroides massiliensis*	2.08E-04	0.25%	98.61%
71	*Clostridium symbiosum*	1.82E-04	0.22%	98.84%
72	*Lactococcus lactis*	1.79E-04	0.22%	99.05%
73	*Bacteroides dorei*	1.66E-04	0.20%	99.26%
74	*Actinomyces sp HPA0247*	1.51E-04	0.19%	99.44%
75	*Bacteroides xylanisolvens*	1.10E-04	0.14%	99.58%
76	*Phascolarctobacterium faecium*	9.68E-05	0.12%	99.70%
77	*Streptococcus pasteurianus*	8.79E-05	0.11%	99.80%
78	*Fusicatenibacter saccharivorans*	8.52E-05	0.10%	99.91%
79	*Ruminococcus torques*	7.58E-05	0.09%	100.00%

Supplementary Table XXB. List of most important taxa for Random Forest models on the 18 to 120 months age bracket, microbiome + demographics, ordered by relative weighted importance rank

Rank	variable	Average fitness-weighted importance	Relative fitness-weighted importance	Rank-cumulative relative importance
1	*ageMonths*	2.13E-02	6.48%	6.48%
2	*Faecalibacterium prausnitzii*	8.59E-03	2.61%	9.09%
3	*Bifidobacterium pseudocatenulatum*	5.98E-03	1.82%	10.91%
4	*Asaccharobacter celatus*	5.86E-03	1.78%	12.70%
5	*Eubacterium eligens*	5.36E-03	1.63%	14.33%
6	*Bifidobacterium longum*	5.36E-03	1.63%	15.96%
7	*Streptococcus parasanguinis*	5.17E-03	1.57%	17.53%
8	*Ruminococcus gnavus*	5.07E-03	1.54%	19.08%
9	*Roseburia faecis*	5.06E-03	1.54%	20.62%
10	*Bacteroides vulgatus*	4.82E-03	1.47%	22.08%
11	*Megamonas funiformis*	4.75E-03	1.45%	23.53%
12	*Fusicatenibacter saccharivorans*	4.71E-03	1.43%	24.96%
13	*Blautia wexlerae*	4.68E-03	1.42%	26.38%
14	*Roseburia hominis*	4.64E-03	1.41%	27.79%
15	*Eubacterium hallii*	4.62E-03	1.41%	29.20%
16	*Parasutterella excrementihominis*	4.40E-03	1.34%	30.54%
17	*Anaerostipes hadrus*	4.36E-03	1.33%	31.87%
18	*Blautia obeum*	4.13E-03	1.26%	33.12%
19	*Agathobaculum butyriciproducens*	3.95E-03	1.20%	34.32%

20	*Haemophilus parainfluenzae*	3.86E-03	1.18%	35.50%
21	*Gordonibacter pamelaeae*	3.80E-03	1.16%	36.66%
22	*Alistipes finegoldii*	3.76E-03	1.14%	37.80%
23	*Dorea longicatena*	3.73E-03	1.14%	38.94%
24	*Flavonifractor plautii*	3.65E-03	1.11%	40.05%
25	*Streptococcus salivarius*	3.62E-03	1.10%	41.15%
26	*Adlercreutzia equolifaciens*	3.55E-03	1.08%	42.23%
27	*Veillonella parvula*	3.55E-03	1.08%	43.31%
28	*Intestinibacter bartlettii*	3.50E-03	1.07%	44.38%
29	*Alistipes putredinis*	3.41E-03	1.04%	45.41%
30	*Ruthenibacterium lactatiformans*	3.35E-03	1.02%	46.43%
31	*Bacteroides ovatus*	3.31E-03	1.01%	47.44%
32	*Bacteroides fragilis*	3.31E-03	1.01%	48.45%
33	*Bacteroides uniformis*	3.30E-03	1.00%	49.45%
34	*Eubacterium rectale*	3.20E-03	0.97%	50.43%
35	*Roseburia inulinivorans*	3.07E-03	0.94%	51.36%
36	*Streptococcus thermophilus*	3.06E-03	0.93%	52.29%
37	*Roseburia intestinalis*	3.05E-03	0.93%	53.22%
38	*Bacteroides caccae*	2.98E-03	0.91%	54.13%
39	*Ruminococcus bicirculans*	2.96E-03	0.90%	55.03%
40	*Parabacteroides distasonis*	2.89E-03	0.88%	55.91%
41	*Ruminococcus torques*	2.87E-03	0.87%	56.78%
42	*Clostridium symbiosum*	2.83E-03	0.86%	57.65%
43	*Collinsella stercoris*	2.81E-03	0.85%	58.50%
44	*Bifidobacterium bifidum*	2.78E-03	0.85%	59.35%

45	*Collinsella aerofaciens*	2.78E-03	0.85%	60.19%
46	*Ruminococcus bromii*	2.75E-03	0.84%	61.03%
47	*Eubacterium sp CAG 38*	2.75E-03	0.84%	61.87%
48	*Erysipelatoclostridium ramosum*	2.70E-03	0.82%	62.69%
49	*Gemmiger formicilis*	2.66E-03	0.81%	63.50%
50	*Eggerthella lenta*	2.61E-03	0.79%	64.29%
51	*Dorea formicigenerans*	2.53E-03	0.77%	65.06%
52	*Parabacteroides merdae*	2.47E-03	0.75%	65.81%
53	*Odoribacter splanchnicus*	2.46E-03	0.75%	66.56%
54	*Bacteroides thetaiotaomicron*	2.46E-03	0.75%	67.31%
55	*Coprococcus eutactus*	2.46E-03	0.75%	68.06%
56	*Coprococcus comes*	2.45E-03	0.74%	68.80%
57	*Tyzzerella nexilis*	2.41E-03	0.73%	69.54%
58	*Megamonas hypermegale*	2.36E-03	0.72%	70.25%
59	*Ruminococcus lactaris*	2.36E-03	0.72%	70.97%
60	*Eubacterium siraeum*	2.34E-03	0.71%	71.69%
61	*Clostridium sp CAG 58*	2.27E-03	0.69%	72.38%
62	*Clostridium spiroforme*	2.25E-03	0.69%	73.06%
63	*Veillonella atypica*	2.19E-03	0.67%	73.73%
64	*Bifidobacterium adolescentis*	2.14E-03	0.65%	74.38%
65	*Monoglobus pectinilyticus*	2.12E-03	0.64%	75.03%
66	*Alistipes shahii*	2.10E-03	0.64%	75.67%
67	*Eubacterium ramulus*	2.10E-03	0.64%	76.31%
68	*Prevotella copri*	2.09E-03	0.64%	76.94%
69	*Bacteroides xylanisolvens*	2.02E-03	0.62%	77.56%

70	*Oscillibacter sp 57 20*	2.01E-03	0.61%	78.17%
71	*Blautia sp CAG 257*	1.98E-03	0.60%	78.77%
72	*Sutterella parvirubra*	1.98E-03	0.60%	79.38%
73	*Akkermansia muciniphila*	1.92E-03	0.58%	79.96%
74	*Clostridium bolteae*	1.86E-03	0.57%	80.53%
75	*Clostridium leptum*	1.85E-03	0.56%	81.09%
76	*Lachnospira pectinoschiza*	1.85E-03	0.56%	81.66%
77	*Bacteroides stercoris*	1.85E-03	0.56%	82.22%
78	*Clostridium innocuum*	1.84E-03	0.56%	82.78%
79	*Escherichia coli*	1.83E-03	0.56%	83.33%
80	*Eubacterium sp CAG 180*	1.79E-03	0.55%	83.88%
81	*Intestinimonas butyriciproducens*	1.79E-03	0.55%	84.42%
82	*Dialister invisus*	1.77E-03	0.54%	84.96%
83	*Anaerotruncus colihominis*	1.71E-03	0.52%	85.48%
84	*Bifidobacterium breve*	1.68E-03	0.51%	85.99%
85	*Sellimonas intestinalis*	1.66E-03	0.51%	86.50%
86	*Veillonella infantium*	1.62E-03	0.49%	86.99%
87	*Bacteroides dorei*	1.61E-03	0.49%	87.48%
88	*Proteobacteria bacterium CAG 139*	1.60E-03	0.49%	87.97%
89	*Megamonas funiformis CAG 377*	1.52E-03	0.46%	88.43%
90	*Bilophila wadsworthia*	1.52E-03	0.46%	88.89%
91	*Hungatella hathewayi*	1.45E-03	0.44%	89.33%
92	*Veillonella dispar*	1.44E-03	0.44%	89.77%
93	*Turicimonas muris*	1.41E-03	0.43%	90.20%
94	*Roseburia sp CAG 471*	1.41E-03	0.43%	90.63%

95	*Eisenbergiella massiliensis*	1.37E-03	0.42%	91.05%
96	*Firmicutes bacterium CAG 83*	1.27E-03	0.39%	91.44%
97	*Barnesiella intestinihominis*	1.26E-03	0.38%	91.82%
98	*Turicibacter sanguinis*	1.22E-03	0.37%	92.19%
99	*Bifidobacterium catenulatum*	1.17E-03	0.36%	92.55%
100	*Bacteroides finegoldii*	1.15E-03	0.35%	92.90%
101	*Phascolarctobacterium faecium*	1.15E-03	0.35%	93.25%
102	*Clostridium disporicum*	1.15E-03	0.35%	93.60%
103	*Eubacterium ventriosum*	1.06E-03	0.32%	93.92%
104	*Haemophilus sp HMSC71H05*	1.06E-03	0.32%	94.24%
105	*Coprococcus catus*	1.04E-03	0.32%	94.56%
106	*Firmicutes bacterium CAG 145*	1.02E-03	0.31%	94.87%
107	*Collinsella intestinalis*	9.58E-04	0.29%	95.16%
108	*Clostridium clostridioforme*	8.79E-04	0.27%	95.43%
109	*Oscillibacter sp CAG 241*	8.16E-04	0.25%	95.68%
110	*Clostridium lavalense*	8.11E-04	0.25%	95.93%
111	*Eisenbergiella tayi*	7.55E-04	0.23%	96.16%
112	*Butyricimonas virosa*	7.45E-04	0.23%	96.38%
113	*Bacteroides cellulosilyticus*	7.29E-04	0.22%	96.60%
114	*Ruminococcus callidus*	7.17E-04	0.22%	96.82%
115	*Bacteroides massiliensis*	6.84E-04	0.21%	97.03%
116	*Clostridium citroniae*	6.73E-04	0.20%	97.24%
117	*Dielma fastidiosa*	6.66E-04	0.20%	97.44%

118	*Lawsonibacter asaccharolyticus*	6.54E-04	0.20%	97.64%
119	*Clostridium bolteae CAG 59*	6.10E-04	0.19%	97.82%
120	*Holdemania filiformis*	5.40E-04	0.16%	97.99%
121	*Coprobacillus cateniformis*	5.23E-04	0.16%	98.15%
122	*Romboutsia ilealis*	5.10E-04	0.16%	98.30%
123	*Alistipes indistinctus*	4.73E-04	0.14%	98.45%
124	*Coprobacter sp*	4.65E-04	0.14%	98.59%
125	*Clostridium sp CAG 299*	4.63E-04	0.14%	98.73%
126	*Roseburia sp CAG 309*	4.17E-04	0.13%	98.86%
127	*Lactobacillus rogosae*	4.17E-04	0.13%	98.98%
128	*Bacteroides faecis*	4.16E-04	0.13%	99.11%
129	*Butyricimonas synergistica*	4.00E-04	0.12%	99.23%
130	*Holdemanella biformis*	3.42E-04	0.10%	99.33%
131	*Clostridium sp CAG 253*	3.21E-04	0.10%	99.43%
132	*Firmicutes bacterium CAG 110*	3.21E-04	0.10%	99.53%
133	*Eubacterium sp CAG 251*	3.20E-04	0.10%	99.63%
134	*Slackia isoflavoniconvertens*	3.08E-04	0.09%	99.72%
135	*Bacteroides intestinalis*	2.97E-04	0.09%	99.81%
136	*Clostridium scindens*	2.47E-04	0.08%	99.89%
137	*Clostridium sp CAG 167*	2.17E-04	0.07%	99.95%
138	*Paraprevotella xylaniphila*	1.55E-04	0.05%	100.00%

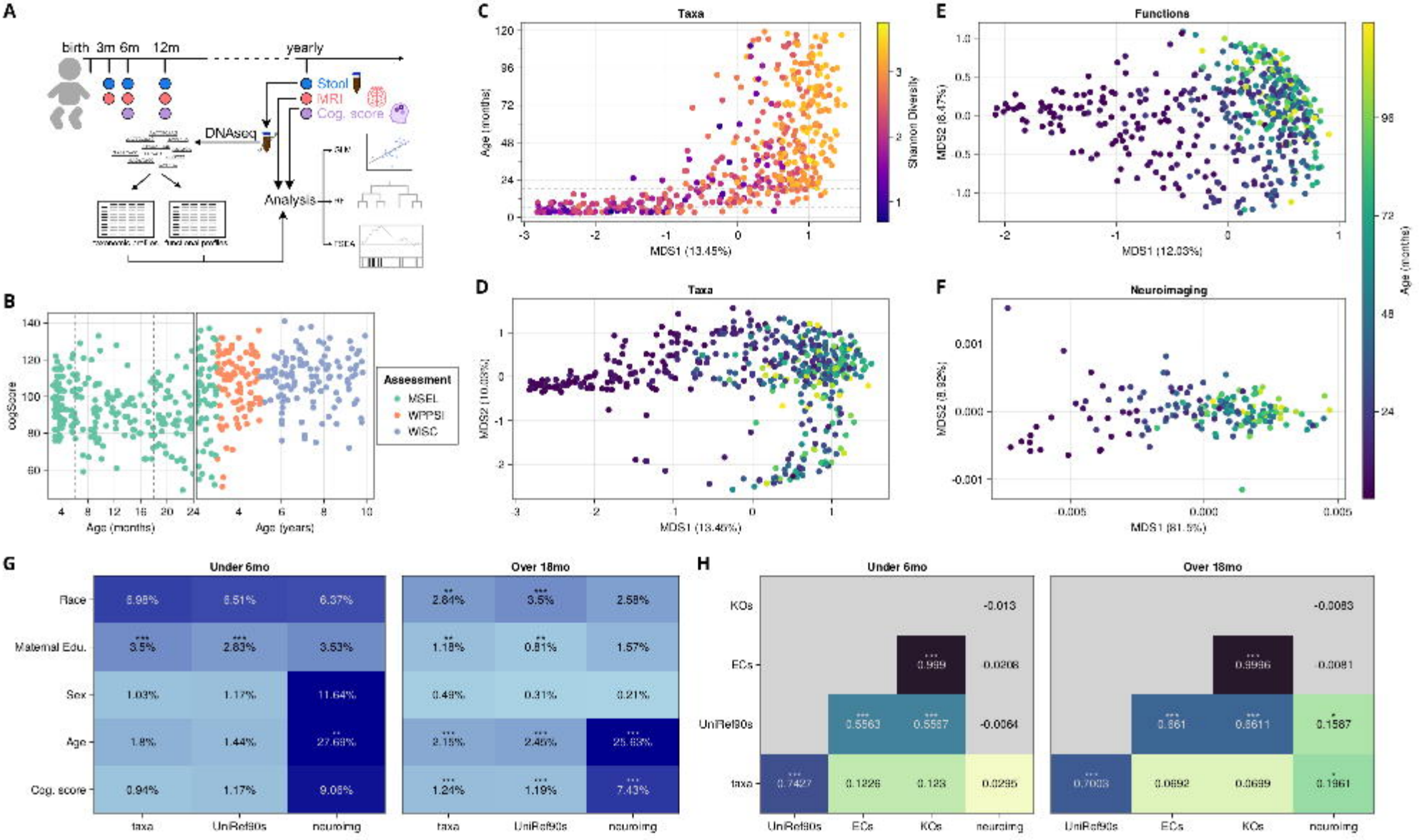
Supplementary Table XX. Benchmark metrics for each brain segment prediction experiment. Confidence intervals are calculated from the distribution of metrics from repeated CV at a confidence level of 95%

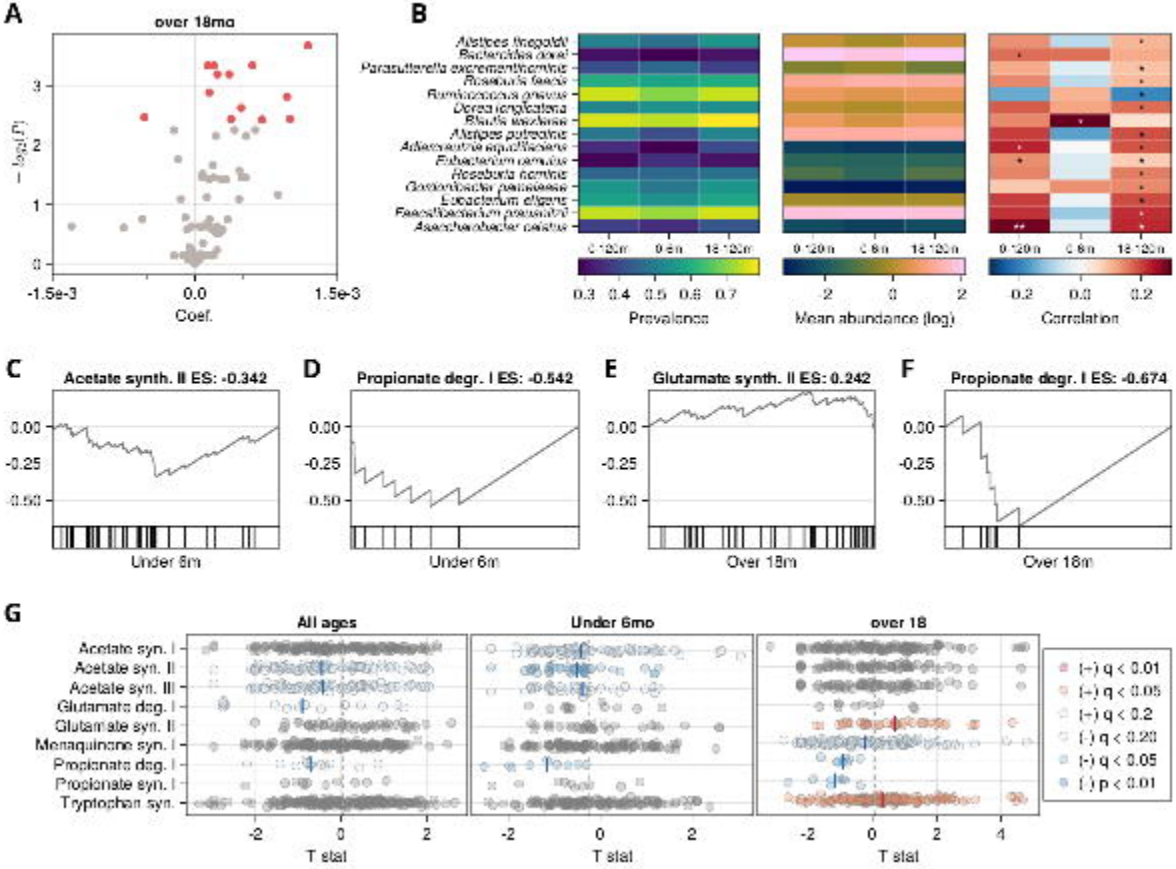
Brain segment	Root-mean-square Error (RMSE)	Mean Absolute Proportional Error (MAPE)	Correlation coefficient (R)
Brain-stem	0.0876	0.077	0.605
cerebellar-vermal-lobules-I-V	0.0321	0.085	0.156
cerebellar-vermal-lobules-VI-VII	0.0132	0.081	0.006
cerebellar-vermal-lobules-VIII-X	0.0217	0.083	0.078
CSF	0.0193	0.162	0.071
left-accumbens-area	0.0020	0.094	0.288
right-accumbens-area	0.0020	0.102	-0.041
left-amygdala	0.0037	0.063	0.288
right-amygdala	0.0035	0.070	0.179
left-anterior-cingulate	0.0375	0.069	0.092
right-anterior-cingulate	0.0264	0.074	0.245
left-basal-forebrain	0.0008	0.206	0.154
right-basal-forebrain	0.0010	0.117	0.123
left-caudate	0.0257	0.099	0.029
right-caudate	0.0221	0.083	-0.010
left-cerebellum-white-matter	0.0689	0.067	0.413
right-cerebellum-white-matter	0.0694	0.070	0.434
left-cuneus	0.0274	0.082	0.234
right-cuneus	0.0264	0.075	0.065
left-entorhinal	0.0107	0.080	0.190
right-entorhinal	0.0105	0.082	0.099

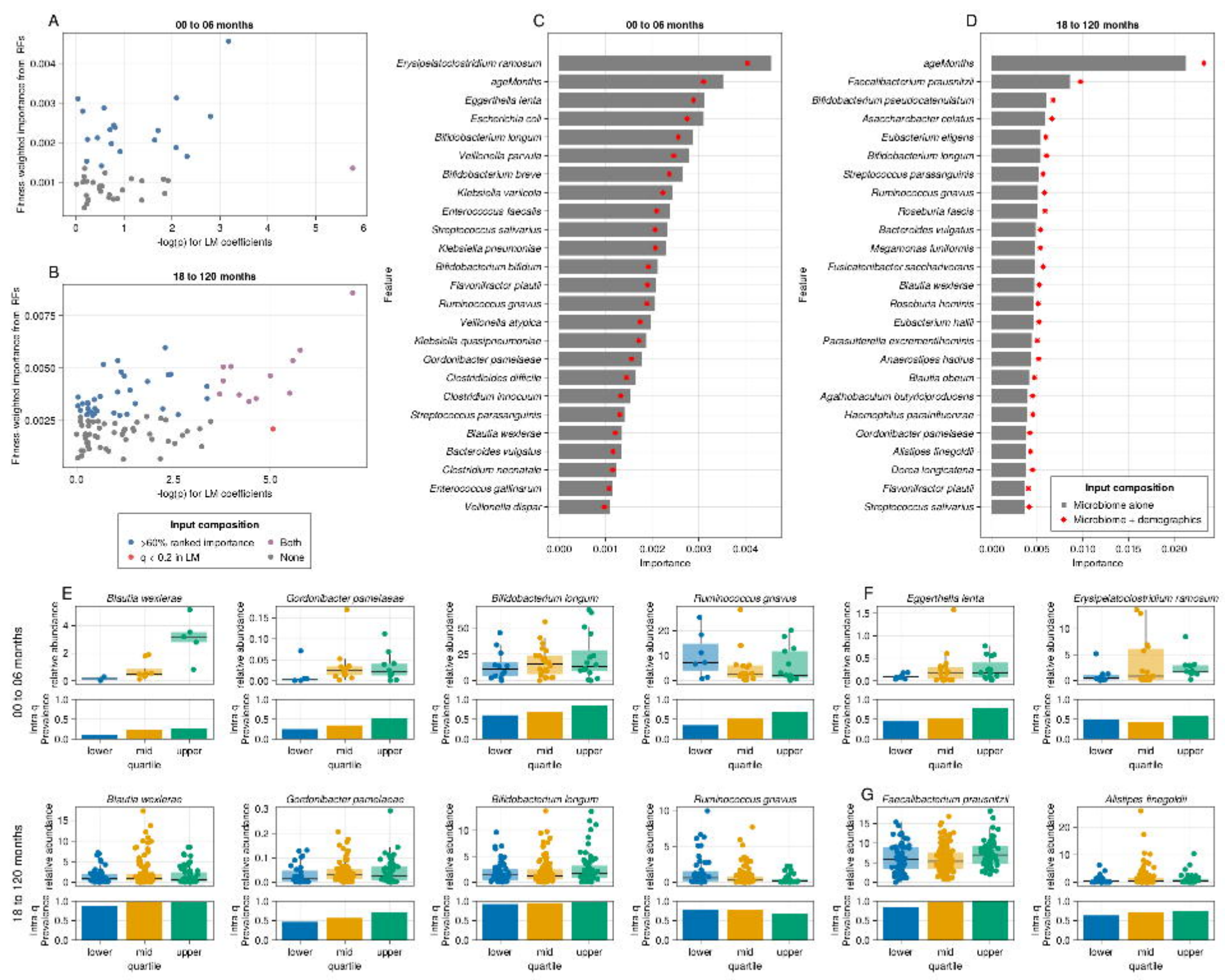
left-fusiform	0.0510	0.057	0.333
right-fusiform	0.0334	0.053	0.248
left-hippocampus	0.0128	0.054	0.075
right-hippocampus	0.0150	0.059	-0.094
left-insula	0.0294	0.057	-0.119
right-insula	0.0303	0.056	0.077
left-isthmus-cingulate	0.0161	0.077	-0.041
right-isthmus-cingulate	0.0166	0.085	0.074
left-lateral-occipital	0.0661	0.062	0.224
right-lateral-occipital	0.0716	0.054	0.221
left-lingual	0.0410	0.069	0.421
right-lingual	0.0440	0.072	0.434
left-middle-frontal	0.0875	0.052	0.000
right-middle-frontal	0.0636	0.039	0.083
left-orbitofrontal	0.0715	0.055	0.186
right-orbitofrontal	0.0814	0.061	0.133
left-pallidum	0.0075	0.064	0.297
right-pallidum	0.0081	0.061	0.264
left-paracentral	0.0335	0.078	0.131
right-paracentral	0.0364	0.084	0.023
left-parahippocampal	0.0098	0.058	0.154
right-parahippocampal	0.0092	0.060	0.134
left-parietal	0.0755	0.035	0.177
right-parietal	0.1099	0.045	-0.147

left-pars-opercularis	0.0197	0.057	0.146
right-pars-opercularis	0.0261	0.060	0.030
left-pars-orbitalis	0.0157	0.082	0.003
right-pars-orbitalis	0.0233	0.112	0.024
left-pars-triangularis	0.0226	0.063	0.174
right-pars-triangularis	0.0322	0.076	0.043
left-pericalcarine	0.0178	0.111	0.200
right-pericalcarine	0.0185	0.089	0.273
left-postcentral	0.0495	0.052	0.168
right-postcentral	0.0552	0.059	0.135
left-posterior-cingulate	0.0460	0.134	0.137
right-posterior-cingulate	0.0480	0.139	0.063
left-precentral	0.0830	0.065	0.248
right-precentral	0.0594	0.046	0.066
left-precuneus	0.0489	0.056	0.144
right-precuneus	0.0622	0.068	0.150
left-putamen	0.0243	0.059	0.154
right-putamen	0.0242	0.066	0.011
left-superior-frontal	0.0867	0.040	0.101
right-superior-frontal	0.1206	0.044	-0.096
left-supramarginal	0.0767	0.067	0.061
right-supramarginal	0.0734	0.068	-0.001
left-temporal	0.1530	0.040	0.128
right-temporal	0.1503	0.039	0.068

left-thalamus-proper	0.0347	0.053	0.170
right-thalamus-proper	0.0304	0.050	0.162


A
birth 3m 6m 12m
yearly
Stool
MRI
Cog. score
DNAseq
Analysis
B
cogScore
Assessment
MSEL
WPPSI
WISC
Age (months)
Age (years)
C
Taxa
Age (months)
MDS1 (13.45%)
Shannon Diversity
D
Taxa
MDS2 (10.03%)
MDS1 (13.46%)
E
Functions
MDS2 (8.47%)
MDS1 (12.03%)
Age (months)
F
Neuroimaging
MDS2 (8.82%)
MDS1 (81.5%)
G
Under 6mo
Over 18mo
Race
Maternal Edu.
Sex
Age
Cog. score
6.98%
6.51%
6.37%
2.84%
3.5%
2.58%
3.5%
2.83%
3.53%
1.18%
0.81%
1.57%
1.03%
1.17%
11.64%
0.48%
0.31%
0.21%
1.8%
1.44%
27.68%
2.15%
2.46%
25.63%
0.94%
1.17%
9.08%
1.24%
1.19%
7.43%
taxa
UniRef90s
neuroimg
H
Under 6mo
Over 18mo
KOs
ECs
UniRef90s
taxa
-0.013
0.0208
-0.0064
0.0295
0.999
0.5563
0.5557
0.7427
0.1226
0.123
-0.0083
0.0081
0.1587
0.1961
0.9996
0.861
0.8611
0.7003
0.0692
0.0699
neuroimg


A
over 18mo
− log(FDR)
Coef.
B
Alistipes finegoldii
Bacteroides dorei
Parasutterella excrementihominis
Roseburia faecis
Ruminococcus gnavus
Dorea longicatena
Blautia wexlerae
Alistipes putredinis
Adlercreutzia equolifaciens
Eubacterium rectale
Roseburia hominis
Gordonibacter pamelaeae
Lachnospira eligens
Faecalibacterium prausnitzii
Anaerobutyricum soehngenii
0-12m
0-6m
18-120m
Prevalence
Mean abundance (log)
Correlation
C
Acetate synth. II ES: -0.342
Under 6m
D
Propionate degr. I ES: -0.542
Under 6m
E
Glutamate synth. II ES: 0.242
Over 18m
F
Propionate degr. I ES: -0.674
Over 18m
G
All ages
Under 6mo
over 18
Acetate syn. I
Acetate syn. II
Acetate syn. III
Glutamate deg. I
Glutamate syn. II
Menaquinone syn. I
Propionate deg. I
Propionate syn. I
Tryptophan syn.
T stat
(−) q < 0.01
(−) q < 0.05
(−) q < 0.2
(-) q < 0.20
(-) q < 0.05
(-) p < 0.01


A
00 to 06 months
Fitness weighted importance from RFs
-log(p) for LM coefficients
B
18 to 120 months
Fitness weighted importance from RFs
-log(p) for LM coefficients
Input composition
>60% ranked importance
q < 0.2 in LM
Both
None
C
00 to 06 months
Feature
Importance
Erysipelatoclostridium ramosum
ageMonths
Eggerthella lenta
Escherichia coli
Bifidobacterium longum
Veillonella parvula
Bifidobacterium breve
Klebsiella varicola
Enterococcus faecalis
Streptococcus salivarius
Klebsiella pneumoniae
Bifidobacterium bifidum
Flavonifractor plautii
Ruminococcus gnavus
Veillonella atypica
Klebsiella quasipneumoniae
Gordonibacter pamelaeae
Clostridioides difficile
Clostridium innocuum
Streptococcus parasanguinis
Blautia wexlerae
Bacteroides vulgatus
Clostridium neonatale
Enterococcus gallinarum
Veillonella dispar
D
18 to 120 months
Feature
Importance
ageMonths
Faecalibacterium prausnitzii
Bifidobacterium pseudocatenulatum
Saccharibacter colatus
Eubacterium eligens
Bifidobacterium longum
Streptococcus parasanguinis
Ruminococcus gnavus
Roseburia faecis
Bacteroides vulgatus
Megamonas funiformis
Fusicatenibacter saccharivorans
Blautia wexlerae
Roseburia hominis
Eubacterium hallii
Parasutterella excrementihominis
Anaerostipes hadrus
Blautia obeum
Agathobaculum butyriciproducens
Haemophilus parainfluenzae
Gordonibacter pamelaeae
Alistipes finegoldii
Dorea longicatena
Flavonifractor plautii
Streptococcus salivarius
Input composition
Microbiome alone
Microbiome + demographics
E
F
G
00 to 06 months
18 to 120 months
Blautia wexlerae
Gordonibacter pamelaeae
Bifidobacterium longum
Ruminococcus gnavus
Eggerthella lenta
Erysipelatoclostridium ramosum
Faecalibacterium prausnitzii
Alistipes finegoldii
relative abundance
Intra q Prevalence
lower
mid
upper
quartile