Gut microbial genes are associated with neurocognition and brain development in healthy children

Kevin S. Bonham [1], Muriel M. K. Bruchhage [2], Sophie Rowland [1], Lexie Volpe [2], Kellyn Dyer [2], RESONANCE Consortium, Viren D'Sa [2,3], Curtis Huttenhower [4], Sean C. L. Deoni [2,3,5,6], Vanja Klepac-Ceraj* [1]

1. Department of Biological Sciences, Wellesley College, Wellesley, MA, 02481, USA

2. Advanced Baby Imaging Lab, Hasbro Children's Hospital, Rhode Island Hospital, Providence, RI, 02903, USA

3. Department of Pediatrics, Warren Alpert Medical School at Brown University, Providence, RI, 02912, USA

4. Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA

5. Department of Radiology, Warren Alpert Medical School at Brown University, Providence, RI, 02912, USA

6. Maternal, Newborn, and Child Health Discovery & Tools, Bill & Melinda Gates Foundation; Seattle WA, USA

ORCID:
Kevin: 0000-0003-3200-7533
Muriel: 0000-0001-9637-0951
Curtis: 0000-0002-1110-0096
Vanja: 0000-0001-5387-5706

* Address correspondence to vklepacc@wellesley.edu

Abstract

Both the brain and microbiome of humans develop rapidly in the first years of life, enabling extensive signaling between the gut and central nervous system (dubbed the "gut-brain axis"[1,2]). While emerging evidence implicates gut microorganisms and microbiota composition in cognitive outcomes and neurodevelopmental disorders (e.g., autism), the role of the gut microbial metabolism on typical neurodevelopment has not been explored in detail.

To examine the relationships between early childhood gut microbiome and neurocognitive development, we collected stool samples from 251 pregnant women and 313 children enrolled in the RESONANCE study of child development - part of the NIH Environmental Influences on Child Health Outcomes (ECHO) Initiative (Table 1). RESONANCE is 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, and rich demographic, socioeconomic, family and medical history information (Supplementary Figure 1).

38

Table 1 - Baseline characteristics of ECHO RESONANCE participants	
All samples (n)	802
Total subjects (n)	561
Pregnant women (n)	251
Kids (n)	310
Kids under 1yo (n)	81
Kids over 2yo (n)	200
Kids with high resolution scan (n)	141
Kids with cognitive function score (n)	263
Kids with both (n)	133
Non-white kids (%)	40.19
Mixed race kids (%)	23.49
Age in years (mean, SD)	4.27, 3.58
BMI (mean, SD)	16.66, 2.68
Maternal SES (mean, SD)	51.68, 10.86

39

40 As an initial characterization step, we used shotgun metagenomic sequencing to
41 generate taxonomic and functional profiles for each of our maternal and child fecal samples.
42 Participant age was the greatest driver of both taxonomic and functional diversity, as expected
43 (Figure 1a, b). Children under one year of age formed a distinct cluster from older children,
44 characterized by high aerobe load and low alpha (within-sample) diversity (Figure 1b,
45 Supplementary Figure 1), and samples from children over two years old were similar to
46 maternal samples (Figure 1a, c, Supplementary Figure 1a).

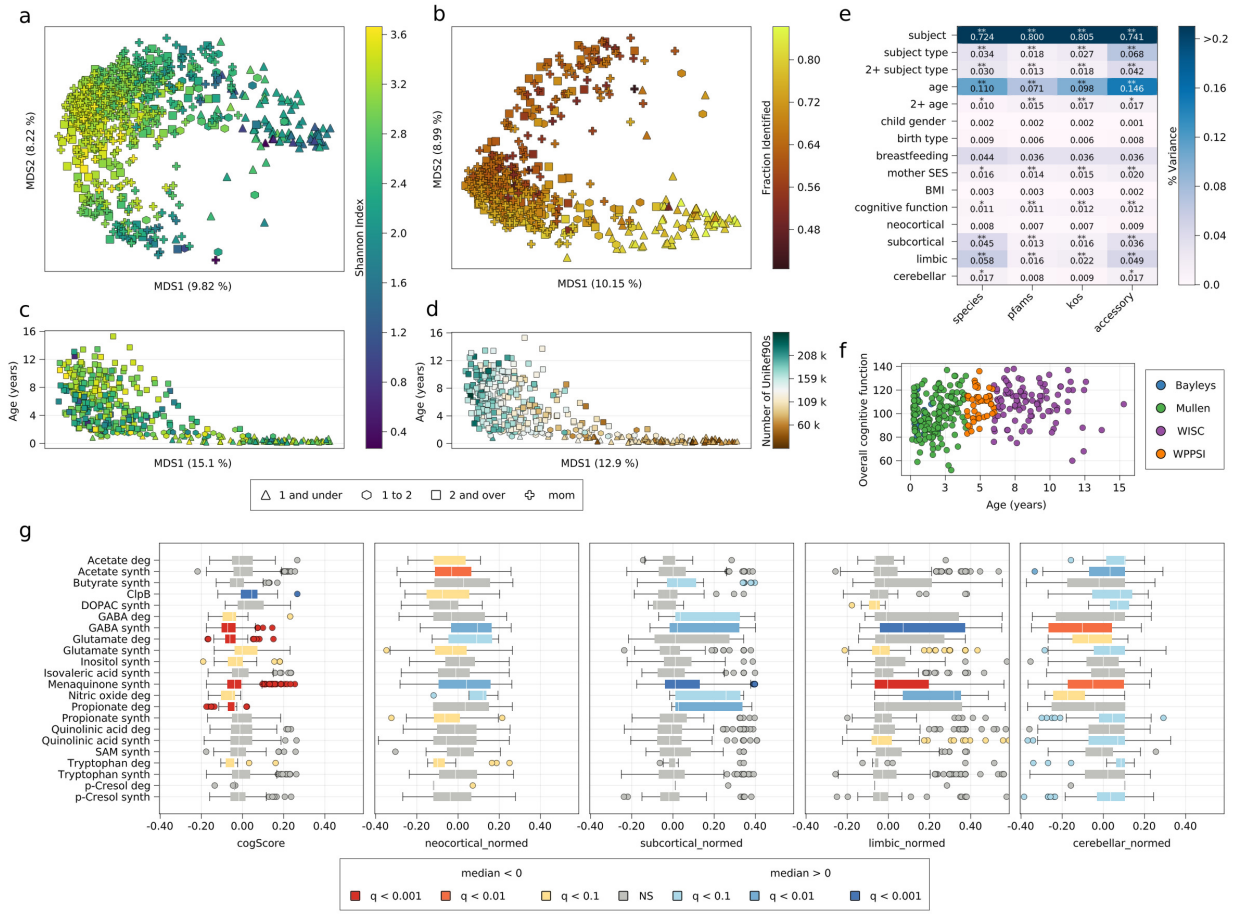

Figure 1: The gut microbiome of children is associated with neurocognition. **a**, Principal coordinates analysis (PCoA) for all subjects based on species-level Bray-Curtis dissimilarity; most variation is driven by age and the presence of *Prevotella*[3,4]. **b**, PCoA based on Bray-Curtis dissimilarity in functional profiles (UniRef90 accessory genes) after removing gene families that were present in >90% of subjects in a given age group; variation is driven by similar effects as for taxonomic profiles. **c**, First principle coordinate axis of species PCoA for child samples vs. age; younger children cluster away from older children, and are lower in diversity. **d**, First principle coordinate axis of accessory gene family PCoA for child samples vs. age; functional diversity increases with age. **e**, PERMANOVA analysis for selected subject metadata vs. pairwiase Bray-Curtis dissimilarity for species-level taxonomic or functional profiles. Functional profiles include Kegg-Orthology (KO), Pfams or accessory UniRef90s; subject and subject type include all subjects, 2+ includes only children over 2 years old, others include all children for which the measure was available; stars indicate significance after Benjamini-Hochberg FDR correction. (* <0.1, ** < 0.01, *** < 0.001) **f**, Child age vs. normalized cognitive function score; age-appropriate assessments achieve similar score distributions for their age groups. **g**, FSEA analysis for gene sets with neuroactive potential (see Methods).

64 As in previous adult and infant cohorts, functional beta diversity was generally lower than
65 taxonomic diversity (Supplementary Figure 1), suggesting that healthy guts select for similar
66 gene functions even when different species contribute those functions. However, this
67 interpretation may be complicated by the fact that as many as 50% of sequencing reads in
68 some samples are not mapped to any of the reference genes used, and are thus unclassified
69 (Figure 1b). Interestingly, though children under 1 tended to have substantially fewer species
70 and, therefore, fewer total genes (Figure 1d, right), those genes tended to be better
71 characterized (Figure 1b, d)[5]. Consistent with previous studies of adult cohorts from
72 industrialized countries[4], another major driver of variation visible from principal coordinates
73 analysis was the presence of Prevotella copri (Supplementary Figure 3). Like samples from very
74 young children, samples with P. copri had reduced diversity compared with samples from older
75 children and pregnant mothers without P. copri (Figure 1a, Supplementary Figure 3). Overall,
76 these results are consistent with prior studies of adult and childhood gut microbiomes.

77 To assess the potential role of the microbiome in these neurocognitive processes, child
78 stool samples were collected alongside MRI and age-appropriate neurocognitive evaluations
79 (Figure 1f). Throughout childhood, a child's brain undergoes remarkable anatomical,
80 microstructural, organizational, and functional change. By age 5, a child's brain has reached
81 >85% of its adult size, has achieved near-adult levels of myelination, and the pattern of axonal
82 connections has been established[6]. Measures of overall cognitive ability (e.g., intelligence
83 quotient, IQ)[7,8], MRI measures of cortical volume and morphometry, as well as other potentially
84 relevant clinical metadata, were compared to taxonomic and functional profile dissimilarity by
85 PERMANOVA[9,10]. Consistent with previous studies, inter-individual variation accounted for the
86 majority of variation in microbial taxonomic and functional profiles ([79%, 77%], $q < 0.001$)
87 (Figure 1e, Supplementary Table 1). Subject type (child or mother) accounted for a large
88 amount of variation (4-6% $q < 0.001$), but this effect dropped to 2% when children under 2 years

89 of age were excluded, suggesting that age, rather than subject type, is actually responsible for
90 driving the taxonomic and functional variation. Among children's samples, age accounted for
91 over 20% of variation in both taxonomic and functional profiles, but this effect also largely
92 disappeared when children under 2 were excluded (Figure 1e, Supplementary Table 1),
93 suggesting that the age effect is primarily driven by the enormous changes in the microbiome
94 over the first year.

95 Microbiome taxonomic and functional variation was also associated with moderate but
96 significant differences in several neurocognitive measures including age-appropriate measures
97 of cognitive ability ([2.04%, 2.08%], $q < 0.01$, N=263). We also found significant associations
98 between regional brain volumes and microbial taxonomic and functional variation, including the
99 sizes of the cerebellum ([3.63%, 2.82%], $q < 0.01$), the subcortex ([3.64% 5.42%], $q <$ [0.01,
100 0.001]), neocortex ([1.95% 1.1%], $q <$ [0.01, NS]), and limbic regions ([4.31% 5.9%], $q <$ [0.01,
101 0.001]) (N = 141 for all high resolution scans) even after controlling for the effect of age on brain
102 volume. These results suggest that there is a strong relationship between the gut microbiome
103 and neurocognitive development (Figure 1e). Though the direction of causality cannot be
104 determined, experimental models of brain development neurological disorders have
105 demonstrated that microbes in the intestine may have causal effects on the functioning of the
106 central nervous system[11–14].

107 To investigate potential mechanisms through which the gut microbiome might affect
108 neurostructural and neurocognitive development in infants and young children, we focused on a
109 group of microbial genes with neuroactive potential identified by Valles-Colomer et al.[15] which
110 code for enzymes that metabolize neuroactive compounds. We analyzed the association of
111 each of these gene sets with our neurocognitive measures using feature set enrichment
112 analysis (FSEA)[16,17] (Figure 1g, Supplementary Table 2). Briefly, we calculated the Pearson
113 correlation between the relative abundance of all identified UniRef90 gene families with each

114 neurocognitive measure, then calculated the Mann-Whitney U statistic for each potentially
115 neuroactive subset. Using this analysis, we observed that catabolic and anabolic pathways for
116 several molecules known to be important in the developing brain were significantly associated
117 with overall cognitive function scores and the size of brain subregions.

118 In particular, microbial genes for GABA synthesis were positively associated with
119 neocortical ($q < 0.01$, Figure 1g, Supplementary Table 2), subcortical ($q < 0.01$), and limbic ($q <$
120 0.001) volume, and negatively associated with cerebellar volume ($q < 0.01$) and overall
121 cognitive function ($q < 1e-5$). Interestingly, GABA degradation genes were also positively
122 associated with the size of the subcortex and negatively associated with cognitive function ($q <$
123 0.05). This may be due to higher GABA synthesis selecting for the ability to catabolize this
124 molecule, making it difficult to assess how actual GABA concentrations in the gut are associated
125 with brain development.

126 Unlike the metabolism of GABA, glutamate synthesis and degradation genes have an
127 inverse relationship (Figure 1g, Supplementary Table 2). The glutamate degradation gene set
128 was negatively associated with cognitive function ($q < 1e-4$) and cerebellar volume ($q < 0.05$)
129 and positively associated with the size of the neocortex ($q < 0.05$), while glutamate synthesis
130 was marginally negatively associated with overall cognitive function and the size of the
131 neocortex, while positively associated with the size of the cerebellum ($q < 0.05$). However, it
132 remains difficult to predict how gut concentrations of glutamate might be related to microbial
133 metabolism; while it might be intuitive to expect that higher glutamate synthesis and lower
134 glutamate degradation would lead to higher gut concentrations of glutamate, it might also be the
135 case that lower glutamate concentrations select for microbes that can synthesize it and against
136 those that break it down. Glutamate is also far more prevalent in the diet and can be rapidly
137 metabolized by gut epithelial cells, making the relationship between gut concentrations and
138 microbial metabolism even more complex[18].

139 Understanding the relationship of intestinal GABA and glutamate metabolism may be
140 particularly important to understanding the role of the gut microbiome in early childhood
141 cognitive development. Together, GABA and glutamate make up the main cerebellar output
142 neurons[19,20] with inhibitory GABAergic cells being widely distributed in the cerebellum projecting
143 out of the cerebellar cortex and excitatory glutamatergic neurons receiving afferent input. The
144 cerebellum, coined 'the little brain', is one of the earliest brain regions to develop[6,21], but
145 continues to grow into adulthood[22], making it especially vulnerable for disorder and disease[23].
146 Specifically, neurodevelopmental disorders, such as autism spectrum disorder, have been
147 associated with an imbalance of the inhibitory/excitatory system regulated by glutamate and
148 GABA, with recent evidence suggesting an impaired conversion of glutamate to GABA in the
149 disorder[24].

150 This is the first look at an ongoing study of child neurocognitive and microbiome
151 development. Using cross-sectional data, we have shown that differences in gut microbial taxa
152 and genes are associated with the structural development of the brain and with cognitive
153 development. In addition, we have shown that particular microbial gene sets with neuroactive
154 potential are associated with neurocognitive development, thus perhaps playing a direct role in
155 the gut luminal exposure of children to neuroactive metabolites. This study is ongoing, and we
156 are collecting additional clinical data such as participant genetic profiles, lead exposure, air
157 quality data and more to understand how the environment, microbiome, and biological
158 development interact to shape neurocognition. Future studies assessing gut metabolite pools
159 combined with MR spectroscopic methods to quantify concentrations of neurotransmitters such
160 as GABA and glutamate in the brain, as well humanized mouse models and longitudinal human
161 data, will provide further insight into the interactions of microbial metabolism and neurocognitive
162 development. As we continue to follow these subjects, we will be able to identify how early life

microbial exposures, including exposures of mothers *in utero*, might affect future neurocognitive outcomes.

Materials and methods

Cohort description

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. From the RESONANCE cohort, 310 typically-developing children and 251 healthy pregnant women were selected for analysis in this study. General participant demographics are provided in **Table 1**, with children being representative of the RI population. As a broad background, children in the BAMBAM cohort were born full-term (>37 weeks gestation) with height and weight average 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 were performed to identify at-risk children and remove them from subsequent analysis.

Additional data collection

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, scheduled on the same day of the MRI scan or at least within 2 weeks of the scan date.

183 ## Approval for human patient research

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

187 ## Stool sample collection and handling

188 Stool samples (n=802) were collected by parents in OMR-200 tubes (OMNIgene GUT, DNA
189 Genotek, Ottawa, Ontario, Canada), stored on ice, and brought within 24 hrs to the lab in RI
190 where they were immediately frozen at -80˚C. Stool samples were not collected if the infant had
191 taken antibiotics within the last two weeks.

192 ## DNA extraction and sequencing of metagenomes

193 All processing of the samples was done at Wellesley College (Wellesley, MA). Nucleic acids were
194 extracted from stool samples using the RNeasy PowerMicrobiome kit automated on the QIAcube
195 (Qiagen, Germantown, MD), excluding the DNA degradation steps. Extracted DNA was
196 sequenced at the Integrated Microbiome Resource (IMR, Dalhousie University, NS, Canada).

197 Shotgun metagenomic sequencing was performed on all samples. A pooled library (max 96
198 samples per run) was prepared using the Illumina Nextera Flex Kit for MiSeq and NextSeq from
199 1 ng of each sample. Samples were then pooled onto a plate and sequenced on the Illumina
200 NextSeq 550 platform using 150+150 bp paired-end “high output” chemistry, generating ~400
201 million raw reads and ~120 Gb of sequence.

202 ## Computational methods, statistical analyses and data availability

203 Raw and processed data (excluding PHI) is available through SRA and Zenodo.org[25]. All code
204 used for statistical and other analysis was version controlled and is available on github[26].

205 Software packages included vegan (R package) for PERMANOVAs, MultivariateStats.jl for MDS
206 analysis, HypothesisTests.jl for Mann-Whitney U tests (used in FSEA analysis),
207 MultipleTesting.jl for false discovery rate correction, and Makie.jl for plotting.

208 Metagenomic data were analyzed using the bioBakery[27] family of tools with default parameters.
209 Briefly, KneadData (v0.7.1) was used to trim and filter raw sequence reads and to separate human
210 reads from bacterial sequences. Samples that passed quality control were taxonomically profiled
211 to the species level using MetaPhlAn2 (v2.7.7) and functionally profiled by HUMAnN2 (v0.11.1).

212 MRI Acquisition and data processing

213 Structural T_1-weighted MRI scans were acquired on a 3T Siemens Trio scanner with a 12-
214 channel head RF array, preprocessed using a multistep registration procedure. Cortical
215 reconstruction and volumetric segmentation was performed with the Freesurfer image analysis
216 suite, which is documented and freely available for download online
217 (http://surfer.nmr.mgh.harvard.edu/). Brain regions were divided into neocortex, cerebellum,
218 limbic and subcortical regions (for more details on acquisition and processing, see extended
219 methods).

220 Neurocognitive assessments

221 Overall cognitive function was defined by the Early Learning Composite as assessed via the
222 Mullen Scales of Early Learning (MSEL)[7], a standardized and population normed tool for
223 assessing fine and gross motor, expressive and receptive language, and visual reception
224 functioning in children from birth through 68 months of age.

225 The third edition of the Bayley Scales of Infant and Toddler Development (Bayley's III) is a
226 standard series of measures primarily to assess the development of infants and toddlers,
227 ranging from 1 to 42 months of age[28].

228 The Wechsler Intelligence Quotient for Children (WISC)[29] is an individually administered
229 standard intelligence test for children aged 6 to 16 years. It derives a full scale intelligence
230 quotient (IQ) score, which we used to assess overall cognitive functioning.

231 The fourth edition of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV)[29] is
232 an individually administered standard intelligence test for children aged 2 years 6 months to 7
233 years 7 months, trying to meet the increasing need for the assessment of preschoolers. Just as
234 the WISC, it derives a full scale IQ score, which we used to assess overall cognitive functioning.

235 Acknowledgments

236 The authors would like to thank Christopher Loiselle and Jennifer Beauchemin for assistance
237 with biospecimen and metadata collection, Nisreen Abo-Sido for substantial sample preparation
238 and DNA extraction, and Julius Krumbiegel for extensive help with figure generation code.

239 This work was supported by the National Institutes of Health (UH3 ODD023313 and R01
240 MH087510).

241 References

242 1. Cryan, J. F. & O'Mahony, S. M. The microbiome-gut-brain axis: from bowel to behavior.
243 *Neurogastroenterol. Motil.* **23**, 187–192 (2011).
244 2. Grenham, S., Clarke, G., Cryan, J. F. & Dinan, T. G. Brain–Gut–Microbe Communication in
245 Health and Disease. *Front. Physiol.* **2**, (2011).
246 3. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180
247 (2011).
248 4. Tett, A. *et al.* The Prevotella copri complex comprises four distinct clades that are
249 underrepresented in Westernised populations. (2019) doi:10.1101/600593.

250 5. Vatanen, T. *et al.* The human gut microbiome in early-onset type 1 diabetes from the
251 TEDDY study. *Nature* **562**, 589–594 (2018).
252 6. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular
253 landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016).
254 7. Mullen, E. M. & others. *Mullen scales of early learning*. (AGS Circle Pines, MN, 1995).
255 8. Wechsler, D. Wechsler intelligence scale for children. (1949).
256 9. McArdle, B. H. & Anderson, M. J. Fitting Multivariate Models to Community Data: A
257 Comment on Distance-Based Redundancy Analysis. *Ecology* **82**, 290–297 (2001).
258 10. Anderson, M. J. Permutational Multivariate Analysis of Variance (PERMANOVA). in *Wiley*
259 *StatsRef: Statistics Reference Online* 1–15 (American Cancer Society, 2017).
260 doi:10.1002/9781118445112.stat07841.
261 11. Clarke, G. *et al.* The microbiome-gut-brain axis during early life regulates the hippocampal
262 serotonergic system in a sex-dependent manner. *Mol. Psychiatry* **18**, 666–673 (2013).
263 12. Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities
264 associated with neurodevelopmental disorders. *Cell* **155**, 1451–1463 (2013).
265 13. Blacher, E. *et al.* Potential roles of gut microbiome and metabolites in modulating ALS in
266 mice. *Nature* **572**, 474–480 (2019).
267 14. Gao, W. *et al.* Gut microbiome and brain functional connectivity in infants-a preliminary
268 study focusing on the amygdala. *Psychopharmacology (Berl.)* (2019) doi:10.1007/s00213-
269 018-5161-8.
270 15. Valles-Colomer, M. *et al.* The neuroactive potential of the human gut microbiota in quality of
271 life and depression. *Nat. Microbiol.* **4**, 623–632 (2019).
272 16. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of
273 gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).
274 17. Metwally, A. A., Aldirawi, H. & Yang, J. A review on probabilistic models used in microbiome
275 studies. (2018).

18. Reeds, P. J., Burrin, D. G., Stoll, B. & Jahoor, F. Intestinal Glutamate Metabolism. *J. Nutr.* **130**, 978S-982S (2000).
19. Hoshino, M. Molecular mechanisms underlying glutamatergic vs. GABAergic neuronal subtype specification in the cerebellum. *Seikagaku* **78**, 130 (2006).
20. Carletti, B. & Rossi, F. Neurogenesis in the cerebellum. *The Neuroscientist* **14**, 91–100 (2008).
21. Leto, K. *et al.* Consensus paper: cerebellar development. *The Cerebellum* **15**, 789–828 (2016).
22. Tiemeier, H. *et al.* Cerebellum development during childhood and adolescence: a longitudinal morphometric MRI study. *Neuroimage* **49**, 63–70 (2010).
23. Wang, S. SH, Kloth, AD, & Badura, A.(2014). *Cerebellum Sensitive Periods Autism Neuron* **83**, 518–532.
24. Fatemi, S. H. *et al.* Consensus paper: pathological role of the cerebellum in autism. *The Cerebellum* **11**, 777–807 (2012).
25. Bonham, K. S. *et al.* Data for 'Gut microbial genes are associated with neurocognition and brain development in healthy children'. (2020) doi:10.5281/zenodo.3633793.
26. Kevin Bonham. *Klepac-Ceraj-Lab/resonance_paper1: Zenodo Release*. (Zenodo, 2020). doi:10.5281/zenodo.3648677.
27. McIver, L. J. *et al.* bioBakery: a meta'omic analysis environment. *Bioinforma. Oxf. Engl.* **34**, 1235–1237 (2018).
28. Bayley, N. *Bayley scales of infant and toddler development*. (PsychCorp, Pearson, 2006).
29. Wechsler, D. Wechsler preschool and primary scale of intelligence—fourth edition. *Psychol. Corp. San Antonio TX* (2012).

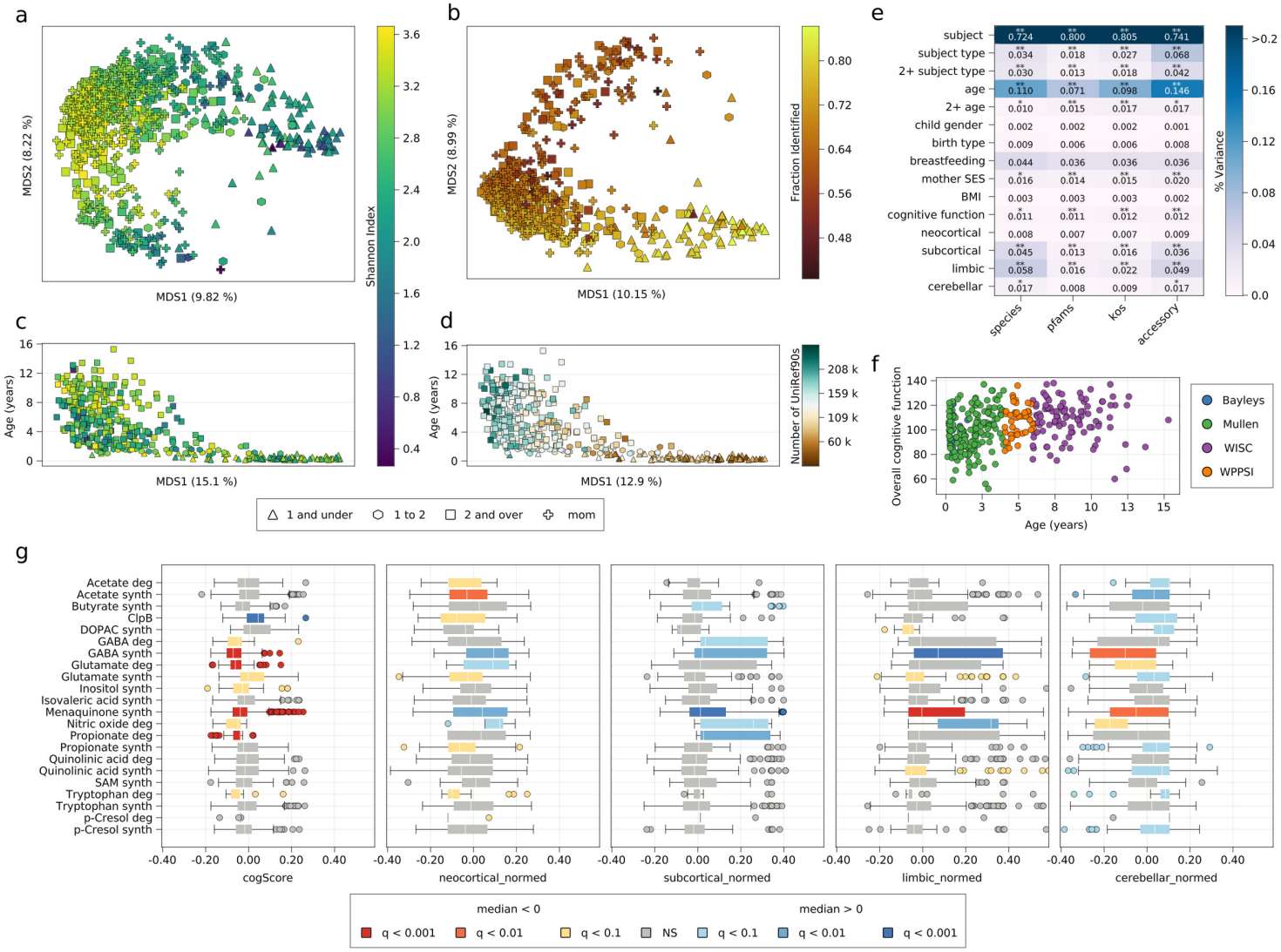

Figure 1: The gut microbiome of children is associated with neurocognition.
a, Principal coordinates analysis (PCoA) for all subjects based on species-level Bray-Curtis dissimilarity; most variation is driven by age and the presence of Prevotella[3,4]. **b**, PCoA based on Bray-Curtis dissimilarity in functional profiles (UniRef90 accessory genes) after removing gene families that were present in >90% of subjects in a given age group; variation is driven by similar effects as for taxonomic profiles. **c**, First principle coordinate axis of species PCoA for child samples vs. age; younger children cluster away from older children, and are lower in diversity. **d**, First principle coordinate axis of accessory gene family PCoA for child samples vs. age; functional diversity increases with age. **e**, PERMANOVA analysis for selected subject metadata vs. pairwiase Bray-Curtis dissimilarity for species-level taxonomic or functional profiles. Functional profiles include Kegg-Orthology (KO), Pfams or accessory UniRef90s; subject and subject type include all subjects, 2+ includes only children over 2 years old, others include all children for which the measure was available; stars indicate significance after Benjamini-Hochberg FDR correction. (* <0.1, ** < 0.01, *** < 0.001) **f**, Child age vs. normalized cognitive function score; age-appropriate assessments achieve similar score distributions for their age groups. **g**, FSEA analysis for gene sets with neuroactive potential (see Methods).