1	Psychosocial adversity experienced in utero and early life is associated with
2	variation in gut microbiota: a prospective case-control study
3	
4	Barbara B. Warner <sup>1,&amp;,*</sup> , Bruce A. Rosa <sup>2,&amp;</sup> , I. Malick Ndao <sup>1</sup> , Phillip I. Tarr <sup>1,3</sup> , J. Phillip
5	Miller <sup>4</sup> , Sara K. England <sup>5</sup> Joan L. Luby <sup>6</sup> , Cynthia E. Rogers <sup>6</sup> , Carla Hall-Moore <sup>1</sup> ,
6	Renay E. Bryant <sup>1</sup> , Jacqueline D. Wang <sup>1</sup> , Laura Linneman <sup>1</sup> Tara A. Smyser <sup>7</sup> ,
7	Christopher D. Smyser <sup>7</sup> , Deanna M. Barch <sup>8</sup> , Gregory E. Miller <sup>9</sup> Edith Chen <sup>9</sup> , John
8	Martin <sup>2</sup> and Makedonka Mitreva <sup>10*</sup>
9	
10	<sup>1</sup> Department of Pediatrics, Washington University School of Medicine, St. Louis, MO
11	63110
12	<sup>2</sup> Department of Medicine, Washington University School of Medicine, St. Louis, MO
13	63110
14	<sup>3</sup> Department of Molecular Microbiology, Washington University School of Medicine, St.
15	Louis, MO 63110
16	<sup>4</sup> Division of Biostatistics, Office of Health Information and Data Science, Washington
17	University School of Medicine in St. Louis, St. Louis, MO 63110
18	<sup>5</sup> Department of Obstetrics and Gynecology, Washington University in St. Louis, St.
19	Louis, Missouri, 63110
20	<sup>6</sup> Department of Psychiatry, Washington University School of Medicine in St. Louis, St.
21	Louis, MO 63110
22	<sup>7</sup> Departments of Neurology, Pediatrics and Radiology, Washington University School of
23	Medicine in St. Louis, MO 63110

- <sup>8</sup> Department of Psychological and Brain Sciences, Washington University in St. Louis,
- 25 St. Louis, MO 63130
- <sup>9</sup> Institute for Policy Research & Department of Psychology, Northwestern University,
- 27 Evanston, Illinois 60208
- <sup>10</sup> Departments of Medicine, Genetics and McDonnell Genome Institute, Washington
- 29 University School of Medicine in St. Louis, St. Louis, MO 63110
- 30 \* Correspondence: Warner BB, <u>warnerbb@wustl.edu</u>; Mitreva M, <u>mmitreva@wustl.edu</u>
- 31 & Equal contribution: Warner BB, <u>warnerbb@wustl.edu</u>; Rosa BA, <u>barosa@wustl.edu</u>
- 32
- 33 Email addresses
- 34 Barbara B Warner: warnerbb@wustl.edu
- 35 Bruce A. Rosa: barosa@wustl.edu
- 36 I. Malick Ndao: ndao\_m@wustl.edu
- 37 Phillip I. Tarr: tarr@wustl.edu
- 38 J. Philip Miller: jphilipmiller@wustl.edu
- 39 Sara K. England: englands@wustl.edu
- 40 Joan Luby: lubyj@wustl.edu
- 41 Cynthia Rogers: rogersc@wustl.edu
- 42 Carla Hall-Moore: <u>cmhall@wustl.edu</u>
- 43 Renay E. Bryant: <u>renay@wustl.edu</u>
- 44 Jacqueline D. Wang: jackie.wang@wustl.edu
- 45 Laura Linneman: <u>llinneman@wustl.edu</u>
- 46 Tara A. Smyser: <u>smysert@wustl.edu</u>

- 47 Christopher D. Smyser: <u>smyserc@wustl.edu</u>
- 48 Deanna M. Barch: <u>dbarch@wustl.edu</u>
- 49 Gregory E. Miller: <u>greg.miller@northwestern.edu</u>
- 50 Edith Chen: edith.chen@northwestern.edu
- 51 John Martin: jmartin@wustl.edu
- 52 Makedonka Mitreva: mmitreva@wustl.edu
- 53

54 **Keywords:** infant, neonatal, parental, gut microbiome, social disadvantage, 55 psychological stressors, machine learning, predictive modeling

56

#### 57 Abstract

58 Social disadvantage (SD) and psychological stressors (PS) trap some populations in 59 poverty, resulting in health inequities. How these two factors become biologically 60 embedded and the pathways leading to adverse health outcomes is unclear, especially 61 in infants exposed to psychosocial adversity in utero and during early life. Variation in gut 62 microbiome structure and functions, and systemic elevations in circulating cytokines 63 levels as indices of inflammation, offer two possible causative pathways. Here, we 64 interrogate the gut microbiome of mother-child dyads and maternal inflammatory markers, 65 and compare high-SD/high-PS dyads to pairs with low-SD/Low-PS, and demonstrate that 66 the GM of high-SD and high-PS mothers may already be compromised, resulting in the 67 lowest observed inter-individual similarity in that group. The strong predictors of maternal 68 high-SD and high-PS based on mothers and children microbiomes were phylogenetically 69 very distinct bacteria indicating different GM pathways associated with SD versus PS.

We identified sets of SD- and PS-discriminatory metabolic pathways in the mothers and in the children, however their predictive power was lower compared to the discriminatory bacterial species. Prediction accuracy was consistently greater for IL-6 than for the other inflammatory markers, supporting an association between systemic inflammation and psychosocial adversity. The gut microbiome of the infants can be used to predict the psychosocial adversity of mothers, and are embedded in the gut microbiota of 4month-old infants.

77

### 78 INTRODUCTION

Health inequities experienced by socially disadvantaged populations is an increasing and urgent societal issue [1, 2]. The psychosocial factors contributing to these disparities often begin early in life and include the prenatal environment which has profound effects on fetal and infant outcomes that can last a lifetime [3-5]. How social disadvatage (SD) and psychological stress (PS) becomes biologically embedded, and then leads to disparete health outcomes remains unclear.

The gut microbiome (GM) is one candidate driver of adverse outcomes. Many SDrelated morbidities are associated with systemic chronic inflammation [6], and the GM, by shaping and modulating the immune system [7, 8], is associated with systemic inflammation and autoimmune disease, including many disorders associated with populations in which SD is widespread, and include diabetes [9], obesity [10], cardiovascular disease, and neurologic disorders [11].

91 The GM is itself shaped primarily by environment, with influences including 92 environmental exposures of diet, antibiotics, exercise, stress and sleep deprivation [12].

93 Many of these exposures have distinct characteristics for individuals living with economic 94 and psychosocial hardships. Despite increasing awareness of an intersection between 95 the GM and psychosocial inequities [13, 14], few human studies [15-17] have examined 96 the impact on gut microbial community structure and function, particularly in the perinatal 97 period. Studies in pregnant women have focused on either maternal psychological state 98 [18-22] or socioeconomic status [23-25] but not both, have used 16S RNA analysis, which 99 limits, taxanomic and metabolic pathway identification, and have not linked mother-infants 100 samples to examine maternal transfer within this context.

101 This study aims to fill these knowledge gaps. Using a prospective cohort 102 assembled prenatally, we identify the distinct impact of exposure to SD and PS on GM 103 structure and functions for mothers and their infants. We identify discriminatory taxa 104 driving this association using whole metagenomic shotgun (WMS) sequencing, and 105 attempt to relate systemic inflammation to the GM by examining maternal prenatal serum 106 cytokines. Understanding the interaction of the gut microbiome with social determinants 107 of health holds great promise for future preventions due to its potential alterability and 108 plasticity over time. This is particularly true for the perinatal period, a critical 109 developmental window in which perturbations in GM community structure and functions 110 have long-lasting effects [26-28].

111

# 112 RESULTS AND DISCUSSION

A subset of 121 mother–child dyads were drawn from the larger parent study, Early Life
Adversity Biological Embedding and Risk for Developmental Precursors of Mental
Disorders (eLABE), which enrolled mothers prenatally and followed their infants after

116 These 121 mothers were drawn from across the psychosocial spectrum and brith. 117 representative of the parent cohort with average ages of 29.8 (± 5.1 years) vs. 29.2 (± 5.3 118 years), similar dietary Healthy Eating index profiles 59 (± 10.6%) vs 58 (± 9.9%), and 119 gestational age at delivery 39.0 (± 1.1wks) vs. 38.3 (± 2 weeks). Mothers and their 120 children were classified based on composite scores from two latent constructs, SD and 121 PS [29]. The SD construct included 7 individual domains across income, neighborhood 122 deprivation, insurance, and education. The PS construct included 9 individual domains 123 across depression, stress, short term and lifetime, and racial discrimination (see Methods 124 for details). To define an effect of SD from that of PS on the maternal and children 125 microbiome, we examined the interrelationships among pre- and post-natal adversity and 126 circulating cytokines as biomarkers of inflammation, with the GM. The mothers' GM was 127 profiled using third trimester stool samples before parturition (average  $-18.0 \pm 16.9$ ). Their 128 children's stools at were sampled 130.4 (± 13.1) days after birth. Circulating cytokines 129 were analyzed in maternal blood samples obtained during the 3<sup>rd</sup> trimester of pregnancy.

131

130

# Maternal SD and PS relation to maternal and child pre- and post-natal GM structure and function, based on high level taxonomic profiling

Participant characteristics are provided in **Table 1** and **Supplementary Table S1**.

We performed targeted metagenomic sequencing (see Methods) for 121 mother-child dyads (**Figure 1**). Among the 121 mother–child dyads (**Figure 1A**), the SD scores were positively correlated with PS scores (r = 0.394,  $P < 10^{-5}$ ; **Figure 1B**), corroborating the literature [30] and demonstrating the importance of differentiating effects from these two overlapping variables. Several variables, including race, breast milk feeding frequency,

healthy eating index and income-to-needs ratio, were significantly associated with highvs-low SD and PS scores, also expected because these variables were used as inputs to

141 calculate these scores [29] (Table 1; Supplementary Table S1).

142 Across all samples, 3,072 amplicon sequence variants (ASVs) representing 143 unique 16S nucleotide sequences were detected. Relative abundance values per ASV 144 per sample are provided for all 242 samples in **Supplementary Table S2**, and nucleotide 145 sequences for each identified ASV are provided in **Supplementary Table S3**. In the 146 mothers, SD and PS scores had an overall negative correlation with  $\alpha$ -(within sample) 147 diversity based on the bacterial taxonomic composition, although this correlation was not 148 significant (Figure 2A), consistent with previous reports of decreased a-diversity 149 associated with lower Socioeconomic Status (SES) [15-17]. In contrast, among the 150 children,  $\alpha$ -diversity was positively correlated with SD (r = 0.581, P < 10<sup>-5</sup>) and PS scores 151  $(r = 0.350, P = 8.3 \times 10^{-5})$  (Figure 2B). This observation may be partly explained by the 152 lower frequency of breast feeding in the high-SD mothers (Figure 2B), as it is reported 153 that breast-fed infants had lower GM  $\alpha$ -diversity compared to formula-fed infants at three 154 months of age, but the GM diversity increased by six months [31].

An examination of the similarity between mother and child GM (β-diversity) measured by the Bray-Curtis statistics, showed that mother–child dyads with low-SD scores had significantly less similarity ( $P < 10^{-5}$ , **Figure 2C**). This may be partially due differences in α-diversity, where formula-fed/high-SD infants have GM that are more similar to their mother's GM because of greater overall α-diversity. Bray-Curtis dissimilarity-based clustering identified significant associations between SD scores and PS scores in the microbiome profiles in the children, but not in the mothers (**Figures 2D**-

162 **2G**). We identified four major GM clusters in the mothers, but none significantly differ in 163 SD or PS (**Figures 2D** and **2E**). We also identified four major sample clusters in children, 164 with cluster 4 having significantly greater SD scores than the other three clusters (ANOVA 165  $P < 10^{-5}$ ), and also higher PS scores than clusters 1 and 3 (ANOVA  $P = 3.1 \times 10^{-4}$ ) (**Figures** 166 **2F** and **2G**). These results suggest that high-SD and high-PS scores are significantly 167 associated with a distinct overall GM profile in the children, but not in the mothers.

168 We next compared Bray-Curtis similarity between each sample dyad, within and 169 between high and low SD and PS groups (Figure 3). High-SD mothers had significantly 170 more variable microbiome profiles than those of low-SD mothers, who presumably had 171 more consistent healthy microbiome profiles. This effect was even stronger for the PS 172 comparison, with low-PS mothers having the most similar GM to each other than in any 173 of the other comparisons, and the high-PS mothers having as similar between-sample 174 diversity as the low-SD mothers (Figure 3A). High-SD children had the most similar 175 microbiome profiles (Figure 3B), which may relate to their increased  $\alpha$ -diversity (Figure 176 **2B**). Low-SD children also have some overall similarity in their GM profiles, but the low-177 SD and high-SD children GMs have little similarity (consistent with clustering results in 178 Figure 2F). The same is true for PS in the children, except the within-group similarity for 179 low-PS and high-PS were not significantly different. Overall, these results suggest that 180 high-SD and high-PS mothers have divergent, variable microbiomes compared to low-SD 181 and low-PS mothers, who share some commonality in overall GM profiles. Taken 182 together, these results suggest that high-SD and high-PS mother already have 183 compromised GM resulting in low observed inter-individual similarity compared to the low-184 SD and low-PS group, which has a more similar and healthier mature GM.

185

# Species-level gut microbiome profiles and metabolic pathway reconstructions of 89 mother-child dvads

188 We performed WMS sequencing on a subset of stools from 89 of the 121 mother-child 189 dyads, selected based on the distribution of SD and PS scores (Figure 1C) (greater and 190 less than the average value + 0.5 and - 0.5 standard deviations). The profiles of the 178 191 GMs (89 mothers and their 89 children), were generated from an average of 6Gb reads 192 per sample, characterized using the Unified Human Gastrointestinal Genome (UHGG) 193 [32]. Relative abundance values per species per sample were calculated using 194 normalized depths [33] for downstream analysis. A total of 2,219 bacterial genomes were 195 detected across all samples (Figure 4A). Genome taxonomic annotations and relative 196 genome abundance values per sample are provided for all 178 samples (Supplementary 197 Table S2). Twelve genomes were detected frequently across all mothers and all children 198 samples ( $\geq$  40% of both), including five *Bifidobacterium* genus members (*B. infantis*, *B.* 199 breve, B. bifidum, B. catenulatum and B. adolescentis), with Bifidobacterium infantis being 200 the most frequently detected genome across all samples (69.7% of mothers and 89.9% 201 of children). These 12 also included two Bacteroides species (B. dorei and B. 202 xylanisolvens), two Faecalicatena species (F. gnavus and F. unclassified), Flavonifractor 203 plautii and Eggerthella lenta (Figure 4B). Four of these species were also identified as 204 "core mother-infant shared species" in a previous WGS study [34], however the limited 205 overlap may be in part a result of using different genome reference databases (identifying 206 clade-specific marker genes from MetaPhIAn2 [34, 35] vs mapping to the Unified Human 207 Gastrointestinal Genome (UHGG) [32]) in these 2 studies.

208 We next reconstructed the metabolic pathways to compare the functional potential 209 of the GM communities (based on the relative abundance of metabolic pathways [36] in 210 read counts per million, CPM) using HUMAnN3 [37-39]. A total of 468 pathways were 211 detected across all samples, 438 of which were detected in at least 3 mother or 3 children 212 GMs (Figure 4C). Relevant pathway annotations and relative pathway abundance values 213 per sample are provided for all 178 samples (Supplementary Table S2). In contrast to 214 the genomes with relatively sparse identification across samples, almost half (46.3%) of 215 all detected pathways were identified in  $\geq$  90% of samples in both the mothers and the 216 children (top right of the plot, Figure 4D), including 130 pathways (27.8%) detected in all 217 178 samples. Despite the taxonomic differences between the GM of mothers and children 218 (31.71% shared genomes; Figure 4A), the core set of microbial functions is conserved 219 across the groups with 87.2% of the pathways being encoded by both mothers and 220 children's GMs (Figure 4C). This greater similarity between mothers and children in their 221 microbial metabolic pathways compared to their microbial taxonomic profiles has been 222 observed previously, and may be attributed to "core" microbial community functions 223 essential for all species, despite the distinct populations of species adapted to different 224 diets at different stages of life [34].

225

# 226 Bacterial species discriminating between mothers SD and PS

To dissociate the impact of the highly inter-related SD from PS on the maternal GM, and to identify discriminatory bacterial taxa that are strong predictors of mother's SD and PS scores, we analyzed taxonomic and pathway GM profiles using two statistical approaches. First, we used supervised Random Forest (RF; [40]) machine-learning was

231 used with a two-round approach [41] to (i) quantify the ability to predict SD and PS 232 classification based on the mothers' and the childrens' microbiome profiles (Figure 5), 233 and (ii) identify the specific genomes and pathways that most strongly differentiate 234 between the high and low SD and PS scores (Figures 6 and 7; see Methods). Second, 235 we used linear discriminant analysis effect size (LEfSe; [42]) to test differential genome 236 and pathway abundance testing, to assign P values and "effect size" values for 237 significance, as additional validation of RF results. Differential abundance statistics for all 238 genomes and all comparisons are provided in Supplementary Table S4.

239 We identified a set of SD and PS-discriminatory bacteria (Figure 6) whose relative 240 abundances can classify mothers into low- or high-SD 80.5% of the time ( $P = 1.4 \times 10^{-6}$ 241 compared to random assignment; FDR-corrected binomial distribution test; AUC = 0.862, 242  $P = 7.1 \times 10^{-8}$ , Wilcoxon rank sum test), and as low- or high-PS 79.4% of the time (P 243 =1.5x10<sup>-6</sup>; AUC = 0.778,  $P = 8.2x10^{-7}$ ; Figure 5). The best predictors of SD (Figure 6A) 244 and PS (Figure 6B) scores in the mothers were identified according to mean decrease 245 of accuracy (MDA) scores (see Methods) for each genome across the low-SD group and 246 the high-SD groups. The microbiome signature of low-SD and low-PS samples consisted 247 of bacterial species present in high abundance in most of the samples, but which were 248 detected with zero or very low abundance in the high-SD or high-PS samples (Figure 6). 249 Low-SD mothers are identified by increased abundance of many Firmicutes A 250 genomes (Figure 6A). Lawsonibacter asaccharolyticus, a recently-identified butyrate-251 producing species [43, 44], was the strongest predictor of low-SD scores in the mothers (MDA = 10.2%, LEfSe effect size = 2.7,  $P = 3.7 \times 10^{-6}$ ). GM-derived butvrate has a wide 252 253 range of beneficial effects on health including regulating fluid transport, reducing

254 inflammation, reinforcing the epithelial defense barrier, and modulating intestinal motility 255 via mechanisms that include potent regulation of gene expression [45], but L. 256 asaccharolyticus has not been previously and independently associated with these 257 beneficial effects. Streptococcus thermophilus was also among the top 5 predictors (MDA 258 = 5.0%, LEfSe effect size = 3.2,  $P = 1.7 \times 10^{-4}$ ) and is used to produce yogurts and 259 cheeses, has properties beneficial to health including the prevention of chronic gastritis 260 and diarrhea, and immunomodulatory properties with possible benefits in inflammation 261 [46]. S. thermophilus was also among the predictors of low-PS, and this species was 262 detected with zero or low abundance in high-PS and high-SD individuals.

263 The predictors of high-SD and high-PS were phylogenetically guite distinct, 264 enriched for Actinobacteria and Firmicutes C, vs. mainly Firmicutes A, respectively. 265 Genomes from four Bifidobacterium species (B. catenulatum, B. bifidum, B. breve and B. 266 infantis) are among the six most strongly associated with high-SD, and B. catenulatum is 267 associated with high-PS as well. It is recognized that *Bifidobacteria* in the human gut vary 268 with age, and while quantitively some are particularly important in infant GM its presence 269 with aging is stable but abundance changes over time. In general high abundance of 270 Bifidobacteria is related to gut homeostasis and health maintenance and protection, in 271 part by producing a number of potentially health promoting metabolites including short 272 chain fatty acids, conjugated linoleic acid and bacteriocins, and Bifidobacteria is 273 postulated to improve health [47]. However, gualitative and guantitative (increasing 274 abundance) in Bifidobacteria are associated with inflammatory disorders such as 275 diverticulitis, inflammatory bowel disease, and colorectal cancer [48]. Additionally, a 276 recent review of GM variations associated with major depressive disorder (MDD)

identified *Bifidobacterium* as one of three genera most consistently associated with MDD
across studies [49]. While the specific functional role of these high-SD associated *Bifidobacteria* species is unclear, there is a striking increase in overall Bifidobacterial
abundance in the high-SD mothers (**Supplementary Figure S1**).

281 Bacteroides A mediterraneensis was most strongly associated with low-PS in 282 mothers (MDA = 10.9%, effect size = 2.5,  $P = 1.5 \times 10^{-4}$ ; Figure 6B). In mice, stress 283 exposure reduces abundance of *Bacteroides* in the GM [50], and in humans, *Bacteroides* 284 is one of five genera associated with healthy status vs. MDD patients [51], the but this is 285 the first report of Bacteroides A mediterraneensis specifically being associated with PS in 286 a human cohort. The same human MDD study [51] also identified Faecalibacterium and 287 Prevotella as being negatively associated with MDD, and in this study Faecalibacterium sp. and Prevotella sp001275135 were the 5<sup>th</sup> and 7<sup>th</sup> strongest predictors to low-PS 288 289 (respectively).

290 Three species of *Blautia sp.* were among the top four most strong predictors of 291 high-PS in the mothers, and none of these four species were associated with SD, 292 suggesting a specific link with psychological stressors. In human studies, *Blautia* was one 293 of ten genera associated with MDD [51], and Blautia and Eggerthella (represented in the 294 high-PS mothers by *E. lenta*) were significantly correlated with PSS scores [52]. The latter 295 study also identified Blautia and Bifidobacteria (represented in the high-PS mothers by B. 296 catenulatum) as being significantly associated with MDD [52]. However, we wish to note 297 that the overall abundance of *Blautia* in the microbiome remained fairly consistent across 298 the mothers (Supplementary Figure S2), and only the species identified in Figure 6B 299 predict PS.

300

# 301 Bacterial species in the infant GM discriminating prenatal maternal SD and PS

A set of SD and PS-discriminatory bacterial genomes correctly classify children into lowor high-SD in 84.6% of comparisons ( $P = 6.7 \times 10^{-9}$  compared to random assignment; FDRcorrected binomial distribution test; AUC = 0.920,  $P = 7.0 \times 10^{-10}$ , Wilcoxon rank sum test), and as low- or high-PS in 82.1% of comparisons ( $P = 7.7 \times 10^{-8}$ ; AUC = 0.889,  $P = 9.8 \times 10^{-10}$ **%** Figure 5). The top predictors of SD score (Figure 7A) and PS score (Figure 7B) in the children were identified according to MDA scores.

308 Among the predictors that are most important for classification to high-SD in 309 children and with the largest effect size are *Enterobacter nimipressuralis*, nearly absent 310 in low-SD children (detected in only one sample), and Klebsiella pneumoniae, both 311 proinflammatory lipopolysaccharide expressing Proteobacteria [53]. The strongest 312 predictor of children with low-SD was *B. infantis*, a species frequently used as a probiotic 313 to diminish inflammation and associated with breastfeeding [54]. B. infantis represented 314 an average of 28.3% of the GM of the low-SD infants, but only 5.2% of the GM in the 315 high-SD infants (MDA = 10.14%, LEfSe effect size = 5.1, P = 0.0083; Supplementary 316 Figure S1). The top taxa based on the comparison of the SD scores in the mothers were 317 not proinflammatory and belonged to genomes from four Bifidobacterium species 318 previously associated with several putatively beneficial metabolites [47].

There was an overlap of the predictors for low-SD and low-PS, including *Veillonella parvula A*, and several *Collinsella* spp. *Veillonella* is a signature taxa of the 4-month microbiome, and with *Collinsella* have been found in breast-fed microbiome indicating reduced concentration of oxygen, increased production and utilization of lactic acid which

is specific for milk dominated diet [55]. The high-SD and high-PS-discriminating bacteria
were enriched for a broad range of evolutionarily distinct Firmicutes species (Figure 7).
The best high-PS predictors included *F. gnavus* reported as pathobiont associated with
inflammatory bowel disease [56] and *Sutterella* sp. *Sutterella* species are prevalent
commensals in the human GM with mild-proinflammatory properties [57].

328

# 329 Metabolic pathways associated with SD and PS in mothers and children

330 The relative genomic abundance data from the WMS dataset has more features and more 331 taxonomic specificity than metabolic pathway abundance data, because pathways are 332 much more conserved across samples (Figure 4). Genomic data therefore enables better 333 estimation of overall GM associations with SD and PS scores (Figure 5). Nevertheless, 334 the same statistical approaches (see Methods; RF and LEfSe) were used to identify sets 335 of SD- and PS-discriminatory metabolic pathways in the mothers and in the children 336 (defined here using MetaCyc, a curated database of experimentally elucidated metabolic 337 pathways from all domains of life [36]). Differential abundance statistics for all pathways 338 and all comparisons are provided in **Supplementary Table S5**.

Based on metabolic pathway profiles, machine learning correctly classifies mothers as high- and low-SD in 71.8% of comparisons ( $P = 8.5 \times 10^{-4}$ ) and as high- and low-PS in 69.1% of comparisons (P = 0.0023). The top predictive pathways for both comparisons are shown in **Table 2**. The pathways with greatest association with high-SD in mothers include three related to carbohydrate degradation (sucrose degradation IV, glycogen degradation I, starch degradation III), which might relate to the accompanying abundance of *Bifidobacterium* species, which are rich in carbohydrate metabolism

pathways [58]. The "myo-, chiro- and scyllo-inositol degradation" pathway (PWY-7237) was most strongly associated with high-PS in mothers (MDA = 5.95, P =  $2.1 \times 10^{-3}$ ). Myoinositol and chiro-inositol degradation by the gut microbiome contributes to inositol deficiency [59], which includes metabolic disorders involved with insulin function [59] and MDD when myo-inositol is deficient in the prefrontal cortex [60].

351 Children are classified correctly as high- and low-SD in 85.9% of comparisons (P 352 =  $1.3 \times 10^{-9}$ ) and as high- and low-PS in 72.1% of comparisons (P =  $4.1 \times 10^{-4}$ ) based on 353 pathway abundance profiles. The top predictive pathways for both of these comparisons 354 are shown in **Table 3**. Synthesis of L-glutamate and L-glutamine (PWY-5505) was the 355 pathway most strongly associated with high-SD in the children (MDA = 7.9%, P =  $7.9\times10^{-1}$ 356 <sup>7</sup>), and has been previously associated with obesity and visceral fat accumulation [61]. 357 Here, L-glutamine biosynthesis was also associated with high-SD in both mothers and 358 children. In supplementation studies of the gut microbiome, glutamine reduces the ratio 359 of Firmicutes to Bacteroidetes and bacterial overgrowth or bacterial translocation, and 360 increases the density of secretory immunoglobulin A (IgA) and IgA+ cells in the intestinal 361 lumen [62].

362

#### 363 Bacterial species discrimination between SD and PS groups based on maternal

#### 364 circulating cytokines

The relationship between inflammation and the GM is examined through maternal systemic cytokines IL-6, IL-8, IL-10 and TNF alpha, measured across trimesters. In mothers, prediction accuracy of the GM was greatest for third trimester cytokine samples, consistent with timing of maternal stool samples. Prediction accuracy was consistently

369 greater for IL-6 than for the other inflammatory markers (Table 4). Discriminatory bacterial 370 genomes demonstrate the most consistent and greatest predictive accuracy for IL-6, 371 which in preclinical models is centrally important in altering fetal brain development in 372 maternal immune activation models, where placental inflammatory signals are relayed to 373 the fetal brain [63-65]. We identified a set of discriminatory taxa from maternal third 374 trimester stools, whose relative abundance successfully classify mothers into low- or high-375 IL-6 75.6% of comparisons ( $P = 2.2 \times 10^{-4}$ ; Supplementary Figure S3) vs. random 376 assignment (FDR-corrected binomial distribution test), and in 72.1% of comparisons of 377 metabolic pathways ( $P = 2.2 \times 10^{-4}$ ). In this case, high prenatal IL-6 concentrations are 378 associated with the lower abundance of anti-inflammatory Bacteroides species, as 379 evidenced by the RF and LEfSe significance (effect size 2.3, P = 0.017 for Bacteroides A 380 and effect size 2.9, P = 0.02 for *Bacteroides faecis*), as reported in other inflammatory 381 states [47, 66]. These data illustrate the importance of community balance, including 382 presence and absence, on target outcomes, and the importance in the ability to identify 383 these at species/strain levels. Interestingly, we also identified discriminatory taxa in the 384 infant 4-month GM profile (Figure 8). Of all inflammatory marker comparisons, microbial 385 genomes in children had the best overall accuracy for maternal third trimester IL-6 386 concentrations (high/low) based on relative abundance (77.8% accuracy,  $P = 6.0 \times 10^{-5}$ ) 387 and metabolic pathways (75.6%,  $P = 6.9 \times 10^{-4}$ ). However, maternal IL-6 concentration 388 was not significantly correlated with SD or PS in this subsample (Supplementary Figure 389 **S4**), although higher SD is associated with higher maternal IL-6 in the full sample set. 390 Thus, the relationship of the maternal GM to IL-6 is driven by taxa distinct from those 391 identified in composite SD or PS values, but still have proinflammatory characteristics.

392

#### 393 CONCLUSIONS

394 Our prospectively assembled cohort of mother-infant dyads, enabled us to quantify for 395 the first time the impact of exposure to both, SD and PS on GM structure and function for 396 mothers and their infants. Mothers and infants classified as "high' (case) or "low" (control) 397 SD/PS, show distinct discriminatory taxonomic, metabolic and inflammatory features that 398 accurately 'predict' maternal prenatal SD and PS status over 80% of the time, with SD 399 having greater predictive accuracy than PS. Mothers with high-SD/high-PS have highly 400 variable microbiomes compared to low-SD/low PS mothers, reflecting greater 401 permutations from environmental influences. The distinct nature of the taxonomic and 402 functional GM predictors for SD compared to PS, indicate different underlying 403 mechanisms driving the relationship.

404 The human GM modulates inflammatory cytokine production [8, 67] and has been 405 linked to chronic inflammatory disorders [6]. We identified a significant relationship 406 between the maternal prenatal and infant GM, and prenatal circulating cytokine 407 concentrations in mothers. Prediction accuracy is highest and most consistent for IL-6 408 suggesting a contributing role to subtle chronic inflammation, although the discriminating 409 taxa are distinct from SD. Elevated IL-6 concentrations have been linked to specific GM 410 profiles in disease states among adults [68, 69] and recently to neuropathology when 411 elevated during pregnancy in animal models [64, 65].

412 Our findings should be viewed within the strengths and limitations of the study.
413 The finding are associations, and do not necessarily imply causality. Identification of
414 strain level taxa and metabolic pathways through WMG, however, form the basis for

testing mechanistic causation in preclinical models. Use of mother- infant dyads allowed
interrogation of SD and PS between mother and infant GM, and examination of
genomes and metabolic pathways. The population we studied is from a circumscribed
region but contains a broad range of socio-economic backgrounds. We did not examine
the role of race, because of the collinearity of race and SD, with no additional
contribution of race in the model beyond that found with SD alone.

421 While longitudinal studies are needed to determine the stability of these GM 422 signatures from early life to early childhood in children, the results identify unique features 423 of the maternal and infant GMs and host response. Such findings should be pursued to 424 better understand the effect of socioeconomic status and mental health determinants on 425 GM health and stability. More critically, information on causal pathways triggered or 426 sustained by the GM that affect child health and development could lead to new 427 biomarkers and interventions. The potential malleability of the GM leaves room for 428 optimism that unfavorable neurodevelopment outcomes might not be inevitable in 429 children with living with high SD and PS values.

430

#### 431 METHODS

### 432 Study Design and Cohort

The study population Consists of 121 mother-child dyads drawn from a larger parent study of 399 dyads, Early Life Adversity Biological Embedding and Risk for Developmental Precursors of Mental Disorders (eLABE; details provided in [29]). The eLABE study used a prospective observational design to examine the impact of pre- and post-natal psychological and social factors on infant neurodevelopment. Preganant

438 women (N=395) and their offspring were recruited from the March of Dimes Prematurity 439 Research Center at Washington University in St. Louis between 2017-2020, with delivery 440 of singleton births at the Barnes Jewish Hospital in St. Louis [70]. This sub-cohort was 441 chosen from the extremes of maternal social disadvantage and psychosocial stress, 442 among women with relatively healthy offspring, who donated third trimester stools. 443 Exclusions included multiple gestation, congenital malformations and infections, 444 premature birth (< 37 weeks gestational age), maternal alcohol or drug use during 445 pregnancy (excluding tobacco, marijuana, and maternal enteral steroid use). Race and 446 ethnicity were based on maternal self reporting extracted from the medical record. 447 Options included American Indian/Alaskan Native, Asian, Black or African American, 448 Native Hawaiian/Pacific Islander, White, unknown, or other (free text) for race, and the 449 following options for ethnicity: Hispanic/Latina, non-Hispanic/Latina, or unknown/not 450 applicable for ethnicity. All procedures were approved by the Human Research Protection 451 Office, informed consent was obtained from the mother for each dyad. The study was 452 performed in accordance with Strengthening the Reporting of Observational Studies in 453 Epidemiology (STROBE) guidelines [71].

454

## 455 Maternal measures

At each trimester of pregnancy, measures of maternal depression, experiences of stress, as well as demographic information including insurance, education, address, and household composition were obtained from participants or extracted from the medical record by trained staff. Components of the two latent constructs, maternal social

disadvantage and maternal psychosocial stress are previously described [29]. Briefly, thecomponents of each latent variable included:

462

463 Maternal Social Disadvantage: Insurance status obtained was verified at the third 464 trimseter from the medical record and maternal self reporting, Income to Needs ratio in 465 each trimseter based on family income and houshold size (1.0 being the poverty line for 466 the U.S.) was self reported, highest maternal educational level was self reported, the 467 national Area Deprivation Index is a national multidimensional geotracking method based 468 on census block data, providing percentile rankings of neighborhood disadvantage status 469 [72], and maternal nutrition over the past year was categorized using the validated 470 Healthy Eating Index (using National Cancer Institute. The Healthy Eating Index – 471 Population Ratio Method. Updated December 14, 2021; [73]) obtained using the 472 Diet History Questionnaire (DHQII).

473 <u>Maternal Psychosocial Stress</u>: In each trimester mothers completed the Edinburgh 474 Postnatal Depression Scale (EPDS) [74], Perceived Stress scale (PSS) [75] at each 475 trimester, averaged over trimesters a one-time lifetime STRAIN survey [76], a 476 comprehensive measure of lifetime stressful and traumatic life events. Experiences of 477 discrimination based on race were assessed using the Everyday Discrimination Scale 478 [77].

479 <u>Maternal medical risks</u>: were defined by the Maternal Medical Risk score in pregnancy, a
480 validated measure of maternal co-morbidities weighted by severity in pregnancy, and
481 maternal pre-pregnancy Body Mass Index was extracted from the medical record.

482

# 483 Infant measures

484 <u>Gestational age</u> was determined by the best obstetric estimate using last menstrual 485 period or earliest ultrasound dating. <u>Birthweight and route of delivery</u> were extracted from 486 the electronic medical record delivery note. <u>Breastfeeding data</u> was collected by parental 487 reporting at the time of home stool sample collection and based on the Center for Disease 488 Control Infant Feeding Practices II study food frequency checklist data [78, 79].

489

#### 490 Biological specimen collection and processing

491 Maternal blood samples for serum were collected from venous draws during routine clinic 492 visits across each trimester, processed within 12 hours of collection, and stored at -80°C 493 in 1 mL aliquots (details in [70]). Stools from mothers and infants were collected from 494 home and processed as previously described [79]. Briefly maternal samples were 495 collected during the third trimester, and infant stools were retrieved directly from the 496 diaper and scooped into barcoded tubes. Stool samples were placed in insulated 497 packaging (VWR) with U-tex gel packs (Fischer Scientific) and placed in home freezers. 498 A community-based courier system available 24 hours per day was used to retrieve 499 samples within 90 minutes. Weekend or overnight samples were taken to the courier 500 office freezers (-20°C) and then delivered to the laboratory freezer (-80°C) during working 501 hours; weekday stools were couriered directly to the laboratory. Previous testing showed 502 no quality difference between temperatures used in delivery methods [80]. DNA was 503 extracted from stool using the Qiagen (Hilden, Germany) QIAamp Power Fecal Pro DNA 504 Mini Kit (50) and the automated QIAcube platform (Qiagen) as previously described [79, 505 81].

506

# 507 Targeted GM profiling using V4-16S rRNA sequencing, data processing and 508 analyses

509 DNA extracted from stool was sequenced on an Illumina MiSeg, producing 2x250bp 510 paired-end reads spanning the V4 hypervariable region, for 242 samples (121 samples 511 from mothers and 121 from their matched children). Data were imported into QIIME2 512 [82], using standard methods and the developer's docker container (giime2/core:2018.8). 513 V4 region amplicons were assembled and denoised using the Qiime2 method 'DADA2 514 denoise-paired'. Processed V4 amplicons were grouped into amplicon sequence variants 515 (ASVs) with 100% sequence similarity. ASVs were classified using a pre-trained classifier 516 based on SILVA (release 132) [83], a comprehensive database that provides accurate 517 annotations [84]. ASV counts per sample were exported as biom files from a gime2 518 artifact and converted into a human readable tsv file using "biom convert". Read counts 519 per sample were rarefied to 11.929 reads per sample (the lowest count among the 242 520 samples) using the "rrarefy" command in the R package "vegan" (version 2.5-7, 521 https://CRAN.R-project.org/package=vegan), and normalized read counts were 522 calculated per sample by dividing the number of reads associated with each ASV by the 523 total number of reads assigned across ASVs. Taxonomic identifications used are directly 524 provided by SILVA [83]. Raw 16S rRNA can be downloaded from public database 525 (submission in progress).

526

# 527 Whole Metagenome Shotgun (WMS) sequencing and GM profiling

528 Whole Metagenome Shotgun (WMS) datasets for 178 samples (89 samples from mothers 529 and 89 from their respective children) were generated on the Illumina NovaSeg S4. For 530 each sample ~ 6Gbp was generated. The reads for all 178 samples were cleaned of 531 barcodes, adapters and low-quality ends using Trimmomatic [85] (version 0.36). The 532 BMTagger program (installed using conda, July 30th, 2020) was used to identify human 533 contaminant reads using the human reference genome (GRCh38.98 [86]). Reads 534 identified as human were removed to produce final paired-end fastg per each of the 178 535 samples. Raw WMS read data can be downloaded from public database (submission in 536 progress).

537 The 178 WMS samples were mapped against the Unified Human Gastrointestinal 538 Genome (UHGG) collection, comprising 204,938 nonredundant genomes from 4,644 gut 539 prokaryotes, each theoretically representing an individual bacterial or archaeal species 540 (95% average nucleotide identity [32]) using bowtie2 [39] (v2.3.5.1). The profile module 541 of the inStrain [33] program was then run to generate sequencing breadth and depth of 542 coverage statistics for every genome, in addition to nucleotide diversity measures per 543 genome per sample. The depth of coverage values were normalized within every sample 544 by dividing each genome's depth by the sum of the depths across all genomes.

545 The 178 WMS samples were also used as input for HUMAnN [37] (version 3), 546 which was ran from the biobakery/humann docker container (latest version as of October 547 2020) using the Chocophlan nucleotide database and Uniref90 [87] protein database. 548 HUMAnN3 runs the MetaPhlAn [38] program as an intermediate step to assign organism-549 specific functional profiling, and we used the developer-provided Metaphlan3 [38] bowtie2 550 [39] database for this intermediate step. The HUMAnN3 pipeline was used to generate

551 MetaCyc [36] pathway abundance per sample. The "humann\_renorm\_table" script 552 (included in the HUMAnN3 distribution) was used to convert Reads Per Kilobase (RPK) 553 values in the MetaCyc abundance table to a normalized value, Copies Per Million (CPM), 554 which can be compared across samples.

555

#### 556 Statistical analysis

557 Significant differences between Low and High SD and PS sample sets based on 558 metadata classifications (Table 1, Supplementary Table S1) were calculated using two-559 tailed T-tests with unequal variance for continuous variables and using Fischer exact tests 560 for categorical data. Correction for multiple testing was not performed across tests, 561 because significant differences indicate potential bias in the data. Significant differences 562 in components of SD and PS were expected since samples were chosen from the 563 extremes of phenotype for SD and PS to improve detection of GM differences between 564 groups [29].

565 For the 16S/ASV sample analysis, Shannon index diversity values were calculated 566 for each sample using the normalized read counts across all taxa using the "diversity" 567 function in the "vegan" library in R, and Bray-Curtis distance diversity values were 568 calculated using the "vegdist" function [88]. Correlation R<sup>2</sup> values and Pearson r values 569 were calculated using MS Excel, and the significance of the correlation was tested using 570 the two-tailed t statistic with degrees of freedom N -1. ASV-based sample clustering was 571 performed using the relative abundance profiles of all ASVs across all samples as input 572 for Bray-Curtis dissimilarity-based clustering (complete linkage) using the "hclust" 573 function in R, with additional metadata visualized using MS Excel. Significant differences

in SD and PS between clusters were identified using ANOVA with a Post Hoc Tukey HSDtest.

576 For the WMG sequence analysis, samples were divided into "high" and "low" SD 577 and PS based on the distribution of these values across the sample set. Samples above 578 the average value + 0.5 standard deviations were considered "high" and samples below 579 the average value – 0.5 standard deviations were considered "low" (35 "low-SD", 43 "high-580 SD, 36 "low-Psych" and 32 "high-Psych"; **Figure 1C**). The same approach was used to 581 separate samples into "high" and "low" sample sets based on inflammatory marker data 582 (IL-6, IL-8, IL-10 and TNF $\alpha$ ).

583 To identify bacterial taxa that strongly predict mothers' SD and PS scores, we 584 analyzed taxonomic and pathway GM profiles using two approaches. First, a supervised 585 machine-learning approach (Random Forest [40]) that identifies non-linear relationships 586 from high dimensional and dependent data [40] and two-round approach using only the 587 best 25 predictors (as previously described for gut microbiome associations with mental 588 health measures [41]) was used to (i) quantify the ability to predict metadata classification 589 based on the microbiome profiles, indicative of the overall association between the 590 microbiome and the composite scores, and (ii) for each comparison, identify the specific 591 genomes and pathways that most strongly differentiate between the high and low SD and 592 PS scores. The generalization error of the model was evaluated by out of bag (OOB) 593 error. The association of the metadata with the microbiome was quantified using the RF 594 classification accuracy, and the significance of the accuracy was measured using FDR-595 corrected binomial distribution tests. RF model accuracy was also examined using 596 receiver operating characteristic (ROC) curves, guantified using the area under the curve

597 (AUC). Significance values for the ROC curves were assigned by Mann-Whitney U 598 statistics [89], using the "roc.area" function in the R library "ROCR".

599 Second, linear discriminant analysis effect size (LEfSe [42]), the most frequently 600 used statistical tool to determine significant differences in microbiome member 601 abundance [90], was used for differential genome abundance testing (default settings at 602 an adjusted  $P \le 0.05$  for significance) for the non-parametric factorial Kruskal-Wallis (KW) 603 sum-rank test, and requiring a linear discriminant analysis (LDA) "effect size" of at least 604 2 in order to identify differentially abundant taxa). The same approach was used for the 605 pathway analysis, but the "effect size" test cutoff applied was reduced to a value of 1 606 instead of 2, since the effect size is designed for the more sparse nature of metagenomic 607 abundance data [42]. However, the same MDA and KW cutoffs were applied for pathway 608 analysis.

609

#### 610 Figure legends

Figure 1: Overview of cohorts. (A) Portrayal of the 16S rRNA and WMS sample sets. (B)
Social Disadvantage (SD) scores are positively correlated with Psychological Stressor
(PS) scores (across the 121 sample pairs). (C) The 89 WMS samples are divided into
"low-SD", "high-SD", "low-PS" and "high-PS" groups based on the distribution of values
(samples within 1 standard deviation of the average value for each variable are excluded).

Figure 2: GM sample diversity and composition comparisons with Social Disadvantage
(SD) and Psychological Stressors (PS) scores. (A) SD scores and PD scores do not
significantly correlate with GM α-diversity (Shannon diversity index) in the 121 stool

620 samples from mothers. (B)  $\alpha$ -diversity in the children is positively correlated with SD (P < 621  $10^{-5}$ ) and PS (P = 8.3x10<sup>-5</sup>) scores. Relative proportions of human milk feeding are also 622 shown, to provide additional context for potential sources of differential diversity. (C)  $\beta$ -623 diversity, measured by Bray-Curtis dissimilarity between the GM of each mother-child 624 dyad, is positively correlated with SD ( $P < 10^{-5}$ ) and PS ( $P = 1.9 \times 10^{-3}$ ) scores. (**D**) 625 Hierarchical clustering (based on Bray-Curtis dissimilarity across all ASVs, complete 626 linkage) identifies four major GM profile-based clusters in the mother. (E) No significant 627 differences in SD or PS scores were identified between the four clusters based on 628 maternal GM profiles (one-way ANOVA). (F) Hierarchical clustering identifies four major 629 GM clusters in the children. (G) Among the GM clusters in children, cluster 4 has greater 630 SD (one-way ANOVA  $P < 3.4 \times 10^{-5}$ ) and PS (3.1x10<sup>-4</sup>) scores. Tukey post-hoc tests were used to identify significant differences between each cluster. \* $P \le 0.05$ . \*\*\*\* P  $\le 10^{-4}$ . \*\*\*\*\*. 631 P ≤ 10<sup>-5</sup>. 632

633

**Figure 3:** Comparisons of β-diversity between sample sets based on high-vs-low SD and PS, quantified by Bray-Curtis similarity between sample pairs, and FDR-corrected twotailed T-tests with unequal variance used to test significance (**A**) Comparisons of withinand between-group β-diversity of the GM for mothers with high-SD, low-SD, high-PS and low-PS. (**B**) Comparisons of within- and between-group GM β-diversity in children with high-SD, low-SD, high-PS and low-PS.

640

641 Figure 4: Genome and pathway detection in the 89 mother and 89 child WMS samples.642 (A) The total number of bacterial genomes and genomes with species-level taxonomic

643 annotation detected in at least 3 maternal samples and/or at least 3 child samples. (B)
644 For each of the 2,219 bacterial genomes detected in any sample, the proportion of mother
645 and child samples that were detected in the dataset. (C) The total number of metabolic
646 pathways detected in the GM from at least 3 mother and at least 3 child stools. (D) For
647 each the 468 metabolic pathways detected in any sample, the proportion of mother and
648 child samples with detection in the dataset.

649

650 Figure 5: The Random Forest (RF) classification accuracy (low-SD vs high-SD and low-651 PS vs high-PS, out-of-bag error) based on relative UHGG genome abundance in the 652 mothers and children. (A) Overall classification accuracy, with P values indicating 653 significance based on FDR-corrected binomial distribution tests (compared to random 654 sample assignment). (B) For each of the four classification tests, receiver operating 655 characteristic (ROC) curves are shown based on RF models, with the area under the 656 curve (AUC) scores and associated Wilcoxon rank sum test results for each ROC curve 657 indicated.

658

**Figure 6**: WMS genome differential abundance in the mothers, based on comparisons of high-vs-low SD scores and high-vs-low Psych scores. (**A**) Based on WMS genome profiles, RF can successfully classify mothers as having low and high SD 80.5% of the time ( $P = 1.4 \times 10^{-6}$  compared to random assignment; FDR-corrected binomial distribution test). Taxonomy and relative abundance per sample for the genomes with the highest predictive value in the top-25 RF model are shown (ranked by mean decrease in accuracy of the RF model; MDA). Also displayed are -Log of the Kruskal-Wallis test P values from

LEfSe (no value shown if the effect size was <2) and the overall abundance of the taxa when present. (**B**) Based on WMS genome profiles, Random Forest (RF) can successfully classify mothers into high and low Psych groups 79.4% of the time (P =1.5x10<sup>-6</sup>). \*Associated with low-SD in the children, \*\*associated with high Psych in the children.

671

672 Figure 7: WMS genome differential abundance in the children, based on comparisons of 673 high-vs-low Social Disadvantage (SD) scores and high-vs-low Psychological Stressors 674 (PS) scores. (A) Based on WMS genome profiles, Random Forest (RF) can successfully 675 classify children as low- and high-SD 84.6% of the time ( $P = 6.7 \times 10^{-9}$  compared to random 676 assignment; FDR-corrected binomial distribution test). Taxonomy and relative abundance 677 per sample for the genomes with the highest predictive value in the top-25 RF model are 678 shown (ranked by mean decrease in accuracy of the RF model; MDA). Also displayed 679 are -Log of the Kruskal-Wallis test P values from LEfSe (no value shown if the effect size 680 was <2) and the overall abundance of the taxa when present. (B) Based on WMS genome 681 profiles, RF can successfully classify children into high and low-PS groups 82.1% of the 682 time ( $P = 7.7 \times 10^{-8}$ ). \*Associated with high-SD in the mothers, \*\*associated with low Psych 683 in the mothers, \*\*\*also associated with high-PS in the mothers.

684

**Figure 8**: MGS genome differential abundance in children, based on comparisons of highvs-low IL-6 abundance. Based on MGS genome profiles, Random Forest (RF) can successfully classify children into high and low IL-6 abundance 77.8% of the time (P =688 6.0x10<sup>-5</sup>) compared to random assignment; FDR-corrected binomial distribution test).

Taxonomy and relative abundance per sample for the genomes with the highest predictive value in the top-25 RF model are shown (ranked by mean decrease in accuracy of the RF model; MDA). Also displayed are -Log of the Kruskal-Wallis test P values from LEfSe (no value shown if the effect size was <2) and the overall abundance of the taxa when present.

694

**Table 1**. Patients characteristics at study entry for each of the primary comparisons of interest. "Low" and "High" Disadvantage and Psych scores are separated according to the distribution of the metadata, as shown in **Figure 1**. Complete characteristics of all samples organized per individual sample is available in **Supplementary Table S2**.

						5			
Comparison	Statistic		All 16S	<b>A</b> 11	Disadvantage score		Psy	/ch	
type			samples	All	Low	High	Low	High	
	# Sample pairs	s (Mother and Child)	121	89	35	43	36	32	
		Min	18.8	19.3	25.7	19.3	21.5	19.3	
	Mother delivery	Max	41.3	41.3	41.3	38.7	41.3	41.0	
	age	Average ± Std. dev.	29.8 ± 5.1	30.4 ± 5.2	33.2 ± 4.4	28.5 ± 5.2	32.0 ± 4.8	28.4 ± 5.4	
	(years)	P value (T-test)			3.5×	10 <sup>-5</sup>	4.5×	10 <sup>-3</sup>	
		African American	52.1%	53.9%	2.9%	93.0%	36.1%	75.0%	
	Deee	Caucasian	44.6%	42.7%	91.4%	7.0%	61.1%	21.9%	
	Race	Other	3.3%	3.4%	5.7%	0.0%	2.8%	3.1%	
		P value (Fisher test)			≤ 1	0 <sup>-5</sup>	1.3×	10 <sup>-3</sup>	
		Min	0.4	0.4	3.0	0.4	0.4	0.5	
	Income:Needs	Max	12.0	12.0	12.0	3.0	12.0	11.8	
	ratio	Average ± Std. dev.	3.4 ± 3.2	3.6 ± 3.5	7.5 ± 2.6	$0.9 \pm 0.4$	$5.0 \pm 3.9$	2.0 ± 2.2	
Matariaal		P value (T-test)			≤ 1	0 <sup>-5</sup>	1.8×	10-4	
Maternal	0	Min	-2.2	-2.2	-2.2	0.4	-2.2	-2.2	
	Social	Max	1.3	1.3	-0.8	1.3	1.3	1.3	
	Disadvantage	Average ± Std. dev.	-0.24 ± 1.01	-0.23 ± 1.1	-1.48 ± 0.39	0.78 ± 0.26	-0.71 ± 1.11	0.31 ± 0.81	
	00016	P value (T-test)			≤ 1	≤ 10 <sup>-5</sup>		10 <sup>-5</sup>	
		Min	-1.7	-1.7	-1.7	-1.4	-1.7	0.34	
	Psychological	Max	2.4	2.4	1.3	2.4	-0.73	2.4	
	Stress Score	Average ± Std. dev.	-0.21 ± 0.86	-0.21 ± 0.98	-0.70 ± 0.75	0.12 ± 0.91	-1.13 ± 0.25	0.92 ± 0.51	
		P value (T-test)			3.8x10⁻⁵		≤ 1	0 <sup>-5</sup>	
	Perceived Stress Scale (PSS)	Min	0	0	13.7	0	13.0	8.5	
		Max	29.7	29.7	27.5	28.0	21.0	29.7	
		Average ± Std. dev.	18.9 ± 4.5	18.6 ± 5.0	18.7 ± 2.7	$18.2 \pm 6.3$	17.3 ± 1.9	$22.4 \pm 4.3$	
		P value (T-test)			0.63		≤ 1	0 <sup>-5</sup>	
		Min	2200	2200	2760	2200	2300	2200	
	Birthweight	Max	4665	4627	4370	4627	4370	4270	
	(g)	Average ± Std. dev.	3319 ± 538	$3283 \pm 556$	3538 ± 459	3077 ± 548	3379 ± 555	$3042 \pm 494$	
		P value (T-test)			1.3×	10 <sup>-4</sup>	0.0	)10	
		Min	37	37	37	37	37	37	
	Gestational age	Max	41	41	41	41	41	41	
	(weeks)	Average ± Std. dev.	39.0 ± 1.1	39.0 ± 1.1	$39.5 \pm 0.95$	38.6 ± 1.03	39.1 ± 1.1	38.7 ± 1.1	
		P value (T-test)			2.38	E-04	0.	23	
Children		Female	43.8%	40.4%	42.9%	39.5%	41.7%	46.9%	
Children	Child gender	Male	56.2%	59.6%	57.1%	60.5%	58.3%	53.1%	
		P value (Fisher test)			0.	82	0.	81	
		NSVD	65.3%	64.0%	60.0%	67.4%	69.4%	59.4%	
	Route of	VAVD	6.6%	4.5%	5.7%	4.7%	2.8%	9.4%	
	delivery*	Cesarean section	28.1%	31.5%	34.3%	27.9%	27.8%	31.3%	
		P value (Fisher test)			0.	62	0.79		
	Breast milk	≥50%	47.9%	46.1%	71.4%	23.3%	58.3%	31.3%	
	feeding	<50%	52.1%	53.9%	28.6%	76.7%	41.7%	68.8%	
	frequency	P value (Fisher test)			2.9x	10 <sup>-5</sup>	0.0	0.031	

700 \*NSVD Normal spontaneous vaginal delivery, VAVD vacuum assisted vaginal delivery

701 Table 2. Metabolic pathways differential abundance in the mothers, based on 702 comparisons of high-vs-low SD scores and high-vs-low PS scores. Machine learning 703 algorithm can classify mothers as low and high SD 71.8% of the time ( $P = 8.6 \times 10^{-4}$ 704 compared to random assignment; FDR-corrected binomial distribution test). Average 705 abundance and association (high or low) are shown for the pathways predictive value in 706 the top-25 RF model (ranked by mean decrease in accuracy of the RF model; MDA), and 707 with LEfSe effect size  $\geq$  1. The -log of the Kruskal-Wallis test *P* values from LEfSe is also 708 shown.

Comp-	MetaCvc pathway ID	Avera	age abund (CPM)	dance	Random Forest (top 25) MDA (%)	LEfSe	e results
arison		Low SD	High SD	Dir- ection		Effect size	Kruskal wallis <i>P</i> value
	PWY-7198: pyrimidine deoxyribonucleotides de novo biosynthesis IV	61.9	100.9	↑	8.51	1.30	9.6E-06
	PWY-5384: sucrose degradation IV (sucrose phosphorylase)	31.2	61.8	Ť	7.67	1.19	2.4E-04
	*PWY-6549: L-glutamine biosynthesis III	35.2	60.1	Ť	7.38	1.13	1.6E-05
	PWY-622: starch biosynthesis	26.9	64.7	Ť	3.78	1.28	2.4E-04
Mother	PWY-241: C4 photosynthetic carbon assimilation cycle, NADP-ME type	45.8	84.0	Ť	2.68	1.30	2.9E-04
30	PWY-5913: partial TCA cycle (obligate autotrophs)	42.4	85.9	Ť	2.67	1.35	1.2E-04
	GLYCOCAT-PWY: glycogen degradation I	38.7	72.7	Ť	2.17	1.21	3.6E-03
	PWY-6731: starch degradation III	22.3	42.3	Ť	1.39	1.00	1.3E-03
	UDPNAGSYN-PWY: UDP-N-acetyl-D-glucosamine biosynthesis I	172.4	208.8	Ť	0.91	1.27	0.015
	NAGLIPASYN-PWY: lipid IVA biosynthesis (E. coli)	80.9	58.2	Ļ	1.84	1.10	0.020
	PWY-7237: myo-, chiro- and scyllo-inositol degradation	225.9	285.6	Ť	5.95	1.49	2.1E-03
	GLYCOGENSYNTH-PWY: glycogen biosynthesis I (from ADP-D-Glucose)	297.0	382.7	Ť	5.77	1.60	1.9E-04
	RHAMCAT-PWY: L-rhamnose degradation I	140.9	168.7	Ť	3.15	1.22	6.1E-03
	PWY-7357: thiamine phosphate formation from pyrithiamine and oxythiamine	283.5	333.4	Ť	2.38	1.38	1.2E-03
Mother	COMPLETE-ARO-PWY: superpathway of aromatic amino acid biosynthesis	435.5	470.6	Ť	1.89	1.22	6.4E-03
Psych	*PWY-6731: starch degradation III	24.0	42.9	Ť	1.83	1.01	5.8E-03
	PWY-622: starch biosynthesis	34.1	61.1	Ť	1.81	1.14	6.0E-03
	THISYNARA-PWY: superpathway of thiamine diphosphate biosynthesis III	214.5	242.2	Ť	0.76	1.14	0.014
	ANAGLYCOLYSIS-PWY: glycolysis III (from glucose)	392.8	427.7	Ť	0.51	1.27	4.6E-04
	PWY-7115: C4 photosynthetic carbon assimilation cycle, NAD-ME type	43.4	63.2	Ť	0.12	1.01	0.021

709 \*Also significant in the same comparison in the children; ↑ higher with high disadvantage / psych, ↓ higher with low disadvantage / psych

711 Table 3. Metabolic pathways differential abundance in the children, based on 712 comparisons of high-vs-low SD scores and high-vs-low Machine learning algorithm can 713 correctly classify children as low and high mother SD 71.8% of the time ( $P = 8.6 \times 10^{-4}$ 714 compared to random assignment; FDR-corrected binomial distribution test). Average 715 abundance and association (high or low) are shown for the pathways predictive value in 716 the top-25 RF model (ranked by mean decrease in accuracy of the RF model; MDA), and 717 with LEfSe effect size  $\geq$  1. The -log of the Kruskal-Wallis test *P* values from LEfSe is also 718 shown.

Comp-		Average	abundanc	e (CPM)	Random	LEfSe results	
Comp- arison	MetaCyc pathway ID	Low Disadv.	High Disadv.	Dir- ection	Forest (top 25) MDA (%)	Effect size	Kruskal wallis <i>P</i> value
	PWY-5505: L-glutamate and L-glutamine biosynthesis	9.3	38.7	<b>↑</b>	7.91	1.20	7.9E-07
	PWY-7210: pyrimidine deoxyribonucleotides biosynthesis from CTP	29.2	78.8	Ť	6.35	1.42	4.0E-04
	*PWY-6549: L-glutamine biosynthesis III	33.9	58.5	Ť	6.30	1.12	6.5E-05
	GOLPDLCAT-PWY: superpathway of glycerol degradation to 1,3-propanediol	35.1	58.2	Ť	4.90	1.00	1.4E-03
	TEICHOICACID-PWY: poly(glycerol phosphate) wall teichoic acid biosynthesis	21.8	48.9	Ť	4.51	1.15	2.1E-04
Child –	PWY-6897: thiamine diphosphate salvage II	167.3	223.2	Ť	3.19	1.43	9.9E-05
advantage	PWY-6700: queuosine biosynthesis I (de novo)	207.5	292.3	Ť	2.70	1.61	5.4E-04
Ū	PWY-6470: peptidoglycan biosynthesis V (β-lactam resistance)	26.6	74.8	Ť	2.59	1.34	1.3E-04
	PWY-6124: inosine-5'-phosphate biosynthesis II	275.7	341.1	Ť	2.47	1.54	3.8E-04
	PWY-1042: glycolysis IV	396.4	483.2	Ť	1.68	1.60	1.3E-03
	PWY0-1296: purine ribonucleosides degradation	205.7	255.8	Ť	1.18	1.39	0.032
	PWY0-1479: tRNA processing	187.7	132.5	$\downarrow$	6.01	1.46	1.6E-03
	*PWY-6549: L-glutamine biosynthesis III	29.2	56.9	1	6.72	1.15	7.2E-05
	PWY-4981: L-proline biosynthesis II (from arginine)	68.9	115.6	Ť	4.54	1.42	0.010
	PWY-5505: L-glutamate and L-glutamine biosynthesis	13.5	37.6	Ť	4.00	1.13	1.6E-04
Child -	*PWY-6731: starch degradation III	59.1	86.1	Ť	3.53	1.16	0.034
Psych	ANAEROFRUCAT-PWY: homolactic fermentation	281.5	331.0	Ť	3.00	1.39	8.9E-03
	TRPSYN-PWY: L-tryptophan biosynthesis	266.0	321.2	Ť	2.26	1.43	5.1E-03
	PWY-241: C4 photosynthetic carbon assimilation cycle, NADP-ME type	83.4	119.1	<b>↑</b>	1.72	1.26	8.0E-03
	FUC-RHAMCAT-PWY: superpathway of fucose and rhamnose degradation	76.4	53.8	$\downarrow$	5.18	1.08	0.021

719

720

722	Table 4. Random forest (RF) machine learning predictive accuracy for high-vs-low
723	inflammatory markers based on GM taxonomic and pathway profiles. Prediction accuracy
724	was consistently greater for IL-6 than for the other inflammatory markers. Bolded values
725	correspond to comparisons for which $P < 0.005$ (after FDR correction).

		Prediction accuracy				FDR-adjusted P value for prediction accuracy					
		IL-6 IL-8 IL-10 TNFα					IL-6	IL-8	IL-10	TNFα	
Mothora	Genomes	75.6%	64.0%	70.6%	74.1%		6.9E-04	0.039	2.8E-03	2.8E-03	
Mothers	Pathways	72.1%	62.0%	68.6%	61.8%		6.9E-04	0.067	5.6E-03	0.18	
Children	Genomes	77.8%	76.0%	74.5%	75.0%	· ·	6.9E-04	6.9E-04	6.9E-04	2.8E-03	
Children	Pathways	75.6%	66.0%	66.7%	56.4%		1.8E-03	0.02	0.01	0.45	

# 

728	Table 5. MetaCyc pathways differential abundance in the mothers and children, based
729	on comparisons of high-vs-low IL-6 abundance. Based on MGS HUMAnN3 MetaCyc
730	pathway abundance, RF can successfully classify mothers as low and high IL-6 73.3% of
731	the time ( $P = 6.9 \times 10^{-4}$ compared to random assignment; FDR-corrected binomial
732	distribution test). Average abundance and association (high or low) are shown for the
733	pathways predictive value in the top-25 RF model (ranked by mean decrease in accuracy
734	of the RF model; MDA), and with LEfSe effect size ≥ 1. The -log of the Kruskal-Wallis test
735	<i>P</i> values from LEfSe is also shown.

0		Avera	ige abunda (CPM)	ance	Random Forest	LEfSe results	
comp- arison	MetaCyc pathway ID	Low Disadv.	High Disadv.	Dir- ection	(top 25) MDA (%)	Effect size	Kruskal wallis <i>P</i> value
	PWY-5104: L-isoleucine biosynthesis IV	83.3	115.6	↑	5.21	1.36	1.1E-02
Mother -	PWY3O-355: stearate biosynthesis III (fungi)	37.3	55.8	Ť	4.58	1.20	3.5E-02
IL-6	PWY-5505: L-glutamate and L-glutamine biosynthesis	75.3	90.7	Ť	4.54	1.13	3.3E-02
	COBALSYN-PWY: superpathway of adenosylcobalamin salvage from cobinamide I	205.8	231.9	Ť	2.46	1.18	2.5E-02
	PWY-7237: myo-, chiro- and scyllo-inositol degradation	152.9	286.4	Ť	6.23	1.81	3.6E-04
	PWY-6549: L-glutamine biosynthesis III	34.5	71.9	Ť	2.96	1.26	3.3E-03
	PWY0-1586: peptidoglycan maturation (meso-diaminopimelate containing)	369.8	517.5	Ť	2.41	1.84	1.2E-03
Child	COBALSYN-PWY: superpathway of adenosylcobalamin salvage from cobinamide I	112.5	168.0	<b>↑</b>	1.32	1.48	4.5E-03
IL-6	PWY-6897: thiamine diphosphate salvage II	177.0	222.3	<b>↑</b>	0.19	1.38	0.003
	PWY-7328: superpathway of UDP-glucose-derived O-antigen building blocks biosynthesis	208.8	149.3	Ļ	5.71	1.50	0.031
	GLUCOSE1PMETAB-PWY: glucose and glucose-1-phosphate degradation	375.7	237.1	$\downarrow$	5.62	1.87	1.6E-02
	GLYCOCAT-PWY: glycogen degradation I	487.9	327.3	$\downarrow$	4.84	1.91	1.5E-02

 $\uparrow$  higher with high IL-6,  $\downarrow$  higher with low IL-6

# 738 Acknowledgments

- 739 Funding, Washington University in St. Louis' GTAC@MGI for library construction and
- sequencing, NIMH grant RO1 MH113883, (BBW, BAR, IMN, PPT, PJM, SKE, JLL, CER,
- TAS, CDS, DMB, GEM, EC, JM, MM) March of Dimes (SKE), Children's Discovery
- 742 Institute II MD-II-2015-489 (BBW, IMN), Biobank Core P30DK052574 (PIT).
- 743

# 744 Author Contributions

- 745 Conceived and designed the study: BBW, MM JL, DB, CER CDS
- 746 Assembled the cohort, collected specimens: IMN, LL, SKE
- 747 Biological and clinical data database maintenance: IMN, REB, JDW, CH-M
- 748 Sample preparation and extraction of DNA: IMN, REB, JDW, CH-M
- 749 Performed data analysis: BAR, MM, JM, PJM, GEM, EC
- 750 Interpreted the data: BAR, MM, BBW, JM PPT, IMN
- 751 Wrote the paper: BAR, MM, BBW
- All authors approve the manuscript for publication.
- 753
- 754 **Competing Interest**
- 755 The authors have no competing interests to disclose.
- 756

# 757 Additional information

- 758 We thank the families involved in the study, the MARCH of Dime Premature Research
- 759 Center at Washington University, and the GTAC@MGI for the sequence data generation.

# 761 Supplementary Figures

762 Supplementary Figure S1: The relative abundance of *Bifidobacterium* species identified
763 in all mothers and children in the high-SD vs low-SD comparisons.

764

765 Supplementary Figure S2: The relative abundance of *Blautia* species identified in all
766 mothers and children in the high-SD vs low-SD comparisons.

767

**Supplementary Figure S3:** Based on WMS genome profiles, RF can successfully classify mothers as low and high IL-6 abundance 75.6% of the time ( $P = 2.2 \times 10^{-4}$ compared to random assignment; FDR-corrected binomial distribution test). Taxonomy and relative abundance per sample for the genomes with the greatest predictive value in the top-25 RF model are shown (ranked by mean decrease in accuracy of the RF model; MDA). Also displayed are -Log of the Kruskal-Wallis test P values from LEfSe (no value shown if the effect size was <2) and the overall abundance of the taxa when present.

775

Supplementary Figure S4: Comparison of Social Disadvantage scores and
Psychological Stressors scores to maternal third trimester inflammatory marker serum
concentrations. Only IL-8 and PS scores correlated significantly.

779

780 Supplementary Tables

Supplementary Table S1: Complete patient characteristics at study entry for each of the primary comparisons of interest. "Low" and "High" Disadvantage and Psych scores are separated according to the distribution of the metadata, as shown in Figure 1.

# 784 Supplementary Table S2: Database of sample data, including metadata, indications of

- sample groups for each comparison, relative 16S ASV abundance, relative MGS genome
- abundance and relative MGS pathway abundance values.
- 787 **Supplementary Table S3**: Full ASV sequences for each unique ASV identifier from Table
- 788 S2.
- 789 Supplementary Table S4: Differential abundance statistics for each UHGG genome in
- each comparison.
- 791 **Supplementary Table S5**: Differential abundance statistics for each MetaCyc pathway
- in each comparison.
- 793

# 794 References

The L. Taking urgent action on health inequities. Lancet. 2020;395(10225):659. doi:
10.1016/S0140-6736(20)30455-4. PubMed PMID: 32113488.

797 2. Woolf SH, Schoomaker H. Life Expectancy and Mortality Rates in the United States, 1959798 2017. JAMA. 2019;322(20):1996-2016. doi: 10.1001/jama.2019.16932. PubMed PMID:
799 31769830; PMCID: PMC7146991.

Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease
in England and Wales. Lancet. 1986;1(8489):1077-81. doi: 10.1016/s0140-6736(86)91340-1.
PubMed PMID: 2871345.

- 803 4. Nobile S, Di Sipio Morgia C, Vento G. Perinatal Origins of Adult Disease and Opportunities
  804 for Health Promotion: A Narrative Review. J Pers Med. 2022;12(2). Epub 20220125. doi:
  805 10.3390/jpm12020157. PubMed PMID: 35207646; PMCID: PMC8877993.
- Simeoni U, Armengaud JB, Siddeek B, Tolsa JF. Perinatal Origins of Adult Disease.
  Neonatology. 2018;113(4):393-9. Epub 20180531. doi: 10.1159/000487618. PubMed PMID: 29852488.
- 809 6. Furman D, Campisi J, Verdin E, Carrera-Bastos P, Targ S, Franceschi C, Ferrucci L, Gilroy
  810 DW, Fasano A, Miller GW, Miller AH, Mantovani A, Weyand CM, Barzilai N, Goronzy JJ, Rando
  811 TA, Effros RB, Lucia A, Kleinstreuer N, Slavich GM. Chronic inflammation in the etiology of
  812 disease across the life span. Nat Med. 2019;25(12):1822-32. Epub 20191205. doi:
- 813 10.1038/s41591-019-0675-0. PubMed PMID: 31806905; PMCID: PMC7147972.
- 814 7. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. Nat Rev
  815 Microbiol. 2021;19(1):55-71. Epub 20200904. doi: 10.1038/s41579-020-0433-9. PubMed PMID:
  816 32887946.
- 817 8. Olin A, Henckel E, Chen Y, Lakshmikanth T, Pou C, Mikes J, Gustafsson A, Bernhardsson
- AK, Zhang C, Bohlin K, Brodin P. Stereotypic Immune System Development in Newborn Children.
  Cell. 2018;174(5):1277-92 e14. doi: 10.1016/j.cell.2018.06.045. PubMed PMID: 30142345;
- 820 PMCID: PMC6108833.

9. Penders J, Thijs C, van den Brandt PA, Kummeling I, Snijders B, Stelma F, Adams H, van
Ree R, Stobberingh EE. Gut microbiota composition and development of atopic manifestations in
infancy: the KOALA Birth Cohort Study. Gut. 2007;56(5):661-7. Epub 20061017. doi:
10.1136/gut.2006.100164. PubMed PMID: 17047098; PMCID: PMC1942165.

Mbakwa CA, Hermes GDA, Penders J, Savelkoul PHM, Thijs C, Dagnelie PC, Mommers
M, Zoetendal EG, Smidt H, Arts ICW. Gut Microbiota and Body Weight in School-Aged Children:
The KOALA Birth Cohort Study. Obesity (Silver Spring). 2018;26(11):1767-76. Epub 20181008.
doi: 10.1002/oby.22320. PubMed PMID: 30296366; PMCID: PMC6646907.

Hou K, Wu ZX, Chen XY, Wang JQ, Zhang D, Xiao C, Zhu D, Koya JB, Wei L, Li J, Chen
ZS. Microbiota in health and diseases. Signal Transduct Target Ther. 2022;7(1):135. Epub
20220423. doi: 10.1038/s41392-022-00974-4. PubMed PMID: 35461318; PMCID: PMC9034083.
Karl JP, Hatch AM, Arcidiacono SM, Pearce SC, Pantoja-Feliciano IG, Doherty LA, Soares

32 12. Kan JP, Hatch AM, Alcidiacono SM, Pearce SC, Pantoja-Peliciano IG, Donenty LA, Soares
33 JW. Effects of Psychological, Environmental and Physical Stressors on the Gut Microbiota. Front
Microbiol. 2018;9:2013. Epub 20180911. doi: 10.3389/fmicb.2018.02013. PubMed PMID:
30258412; PMCID: PMC6143810.

836 Amato KR, Arrieta MC, Azad MB, Bailey MT, Broussard JL, Bruggeling CE, Claud EC, 13. 837 Costello EK, Davenport ER, Dutilh BE, Swain Ewald HA, Ewald P, Hanlon EC, Julion W, 838 Keshavarzian A, Maurice CF, Miller GE, Preidis GA, Segurel L, Singer B, Subramanian S, Zhao 839 L, Kuzawa CW. The human gut microbiome and health ineguities. Proc Natl Acad Sci U S A. 840 2021;118(25). doi: 10.1073/pnas.2017947118. PubMed PMID: 34161260; PMCID: 841 PMC8237592.

Herd P, Palloni A, Rey F, Dowd JB. Social and population health science approaches to
understand the human microbiome. Nat Hum Behav. 2018;2(11):808-15. Epub 20181022. doi:
10.1038/s41562-018-0452-y. PubMed PMID: 31457107; PMCID: PMC6711373.

845 15. Bowyer RCE, Jackson MA, Le Roy CI, Ni Lochlainn M, Spector TD, Dowd JB, Steves CJ.
846 Socioeconomic Status and the Gut Microbiome: A TwinsUK Cohort Study. Microorganisms.
847 2019;7(1). Epub 20190111. doi: 10.3390/microorganisms7010017. PubMed PMID: 30641975;
848 PMCID: PMC6351927.

849 16. Miller GE, Engen PA, Gillevet PM, Shaikh M, Sikaroodi M, Forsyth CB, Mutlu E,
850 Keshavarzian A. Lower Neighborhood Socioeconomic Status Associated with Reduced Diversity
851 of the Colonic Microbiota in Healthy Adults. PLoS One. 2016;11(2):e0148952. Epub 20160209.
852 doi: 10.1371/journal.pone.0148952. PubMed PMID: 26859894; PMCID: PMC4747579.

He Y, Wu W, Wu S, Zheng HM, Li P, Sheng HF, Chen MX, Chen ZH, Ji GY, Zheng ZD,
Mujagond P, Chen XJ, Rong ZH, Chen P, Lyu LY, Wang X, Xu JB, Wu CB, Yu N, Xu YJ, Yin J,
Raes J, Ma WJ, Zhou HW. Linking gut microbiota, metabolic syndrome and economic status
based on a population-level analysis. Microbiome. 2018;6(1):172. Epub 20180924. doi:
10.1186/s40168-018-0557-6. PubMed PMID: 30249275; PMCID: PMC6154942.

858 18. Zijlmans MA, Korpela K, Riksen-Walraven JM, de Vos WM, de Weerth C. Maternal
859 prenatal stress is associated with the infant intestinal microbiota. Psychoneuroendocrinology.
860 2015;53:233-45. Epub 20150119. doi: 10.1016/j.psyneuen.2015.01.006. PubMed PMID:
861 25638481.

Aatsinki AK, Keskitalo A, Laitinen V, Munukka E, Uusitupa HM, Lahti L, Kortesluoma S,
Mustonen P, Rodrigues AJ, Coimbra B, Huovinen P, Karlsson H, Karlsson L. Maternal prenatal
psychological distress and hair cortisol levels associate with infant fecal microbiota composition
at 2.5 months of age. Psychoneuroendocrinology. 2020;119:104754. Epub 20200605. doi:
10.1016/j.psyneuen.2020.104754. PubMed PMID: 32531627.

Kang LJ, Vu KN, Koleva PT, Field CJ, Chow A, Azad MB, Becker AB, Mandhane PJ,
Moraes TJ, Sears MR, Lefebvre DL, Turvey SE, Subbarao P, Lou WYW, Scott JA, Kozyrskyj AL.
Maternal psychological distress before birth influences gut immunity in mid-infancy. Clin Exp
Allergy. 2020;50(2):178-88. Epub 20200115. doi: 10.1111/cea.13551. PubMed PMID: 31845414.

871 21. Hantsoo L, Jasarevic E, Criniti S, McGeehan B, Tanes C, Sammel MD, Elovitz MA,
872 Compher C, Wu G, Epperson CN. Childhood adversity impact on gut microbiota and inflammatory
873 response to stress during pregnancy. Brain Behav Immun. 2019;75:240-50. Epub 20181103. doi:
874 10.1016/j.bbi.2018.11.005. PubMed PMID: 30399404; PMCID: PMC6349044.

875 22. Hechler C, Borewicz K, Beijers R, Saccenti E, Riksen-Walraven M, Smidt H, de Weerth
876 C. Association between Psychosocial Stress and Fecal Microbiota in Pregnant Women. Sci Rep.
877 2019;9(1):4463. Epub 20190314. doi: 10.1038/s41598-019-40434-8. PubMed PMID: 30872645;
878 PMCID: PMC6418257.

Levin AM, Sitarik AR, Havstad SL, Fujimura KE, Wegienka G, Cassidy-Bushrow AE, Kim
H, Zoratti EM, Lukacs NW, Boushey HA, Ownby DR, Lynch SV, Johnson CC. Joint effects of
pregnancy, sociocultural, and environmental factors on early life gut microbiome structure and
diversity. Sci Rep. 2016;6:31775. Epub 20160825. doi: 10.1038/srep31775. PubMed PMID:
27558272; PMCID: PMC4997337.

Search 24. Jahnke JR, Roach J, Azcarate-Peril MA, Thompson AL. Maternal precarity and HPA axis
functioning shape infant gut microbiota and HPA axis development in humans. PLoS One.
2021;16(5):e0251782. Epub 20210520. doi: 10.1371/journal.pone.0251782. PubMed PMID:
34015045; PMCID: PMC8136730.

Lewis CR, Bonham KS, McCann SH, Volpe AR, D'Sa V, Naymik M, De Both MD,
Huentelman MJ, Lemery-Chalfant K, Highlander SK, Deoni SCL, Klepac-Ceraj V. Family SES Is
Associated with the Gut Microbiome in Infants and Children. Microorganisms. 2021;9(8). Epub
20210728. doi: 10.3390/microorganisms9081608. PubMed PMID: 34442687; PMCID:
PMC8398307.

893 26. Jain N. The early life education of the immune system: Moms, microbes and (missed)
894 opportunities. Gut Microbes. 2020;12(1):1824564. Epub 20201012. doi:
895 10.1080/19490976.2020.1824564. PubMed PMID: 33043833; PMCID: PMC7781677.

896 27. Cowan CSM, Dinan TG, Cryan JF. Annual Research Review: Critical windows - the
897 microbiota-gut-brain axis in neurocognitive development. J Child Psychol Psychiatry.
898 2020;61(3):353-71. Epub 20191126. doi: 10.1111/jcpp.13156. PubMed PMID: 31773737.

899 Laursen MF, Bahl MI, Licht TR. Settlers of our inner surface - factors shaping the gut 28. 900 microbiota from birth to toddlerhood. FEMS Microbiol Rev. 2021;45(4). doi: 901 10.1093/femsre/fuab001. PubMed PMID: 33428723; PMCID: PMC8371275.

29. Luby JL, Barch DM, Warner B, Rogers C, Smyser CD, Triplett R, Arora J, Smyser T,
Birds SK, Slavich G, Stout M, Miller P. Modeling Prenatal Adversity/Advantage: Effects on
Birth Weight. MedRxiv. 2021. doi: doi: https://doi.org/10.1101/2021.12.16.21267938.

30. Ridley M, Rao G, Schilbach F, Patel V. Poverty, depression, and anxiety: Causal evidence
and mechanisms. Science. 2020;370(6522). Epub 2020/12/12. doi: 10.1126/science.aay0214.
PubMed PMID: 33303583.

Ma J, Li Z, Zhang W, Zhang C, Zhang Y, Mei H, Zhuo N, Wang H, Wang L, Wu D.
Comparison of gut microbiota in exclusively breast-fed and formula-fed babies: a study of 91 term
infants. Sci Rep. 2020;10(1):15792. Epub 2020/09/27. doi: 10.1038/s41598-020-72635-x.
PubMed PMID: 32978424; PMCID: PMC7519658.

32. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS, Sakharova
E, Parks DH, Hugenholtz P, Segata N, Kyrpides NC, Finn RD. A unified catalog of 204,938
reference genomes from the human gut microbiome. Nat Biotechnol. 2021;39(1):105-14. Epub
2020/07/22. doi: 10.1038/s41587-020-0603-3. PubMed PMID: 32690973; PMCID: PMC7801254.

916 33. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF.
917 inStrain profiles population microdiversity from metagenomic data and sensitively detects shared
918 microbial strains. Nat Biotechnol. 2021;39(6):727-36. Epub 2021/01/20. doi: 10.1038/s41587919 020-00797-0. PubMed PMID: 33462508.

920 34. Wang S, Zeng S, Egan M, Cherry P, Strain C, Morais E, Boyaval P, Ryan CA, E MD, Ross
921 RP, Stanton C. Metagenomic analysis of mother-infant gut microbiome reveals global distinct and

922 shared microbial signatures. Gut Microbes. 2021;13(1):1-24. doi: 923 10.1080/19490976.2021.1911571. PubMed PMID: 33960282; PMCID: PMC8115609.

35. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower
C, Segata N. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods.
2015;12(10):902-3. doi: 10.1038/nmeth.3589. PubMed PMID: 26418763.

927 36. Časpi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK,
928 Paley S, Subhraveti P, Karp PD. The MetaCyc database of metabolic pathways and enzymes - a
929 2019 update. Nucleic Acids Res. 2020;48(D1):D445-D53. Epub 2019/10/06. doi:
930 10.1093/nar/gkz862. PubMed PMID: 31586394; PMCID: PMC6943030.

931 37. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson
932 KS, Knight R, Caporaso JG, Segata N, Huttenhower C. Species-level functional profiling of
933 metagenomes and metatranscriptomes. Nat Methods. 2018;15(11):962-8. Epub 2018/11/01. doi:
934 10.1038/s41592-018-0176-y. PubMed PMID: 30377376; PMCID: PMC6235447.

38. Beghini F, McIver LJ, Blanco-Miguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A,
Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower
C, Franzosa EA, Segata N. Integrating taxonomic, functional, and strain-level profiling of diverse
microbial communities with bioBakery 3. Elife. 2021;10. Epub 2021/05/05. doi:
10.7554/eLife.65088. PubMed PMID: 33944776; PMCID: PMC8096432.

940 39. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
941 2012;9(4):357-9. Epub 2012/03/06. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PMCID:
942 PMC3322381.

943 40. Knights D, Costello EK, Knight R. Supervised classification of human microbiota. FEMS
944 Microbiol Rev. 2011;35(2):343-59. Epub 2010/11/03. doi: 10.1111/j.1574-6976.2010.00251.x.
945 PubMed PMID: 21039646.

946 41. Malan-Muller S, Valles-Colomer M, Foxx CL, Vieira-Silva S, van den Heuvel LL, Raes J, 947 Seedat S, Lowry CA, Hemmings SMJ. Exploring the relationship between the gut microbiome and 948 mental health outcomes in a posttraumatic stress disorder cohort relative to trauma-exposed 949 controls. Eur Neuropsychopharmacol. 2022:56:24-38. Epub 20211216. doi: 950 10.1016/j.euroneuro.2021.11.009. PubMed PMID: 34923209.

951 42. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C.
952 Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):2011-12.

953 43. Sakamoto M, Iino T, Yuki M, Ohkuma M. Lawsonibacter asaccharolyticus gen. nov., sp.
954 nov., a butyrate-producing bacterium isolated from human faeces. Int J Syst Evol Microbiol.
955 2018;68(6):2074-81. Epub 20180510. doi: 10.1099/ijsem.0.002800. PubMed PMID: 29745868.

956 44. Sakamoto M, Ikeyama N, Yuki M, Ohkuma M. Draft Genome Sequence of Lawsonibacter
957 asaccharolyticus JCM 32166(T), a Butyrate-Producing Bacterium, Isolated from Human Feces.
958 Genome Announc. 2018;6(25). Epub 20180621. doi: 10.1128/genomeA.00563-18. PubMed
959 PMID: 29930067; PMCID: PMC6013597.

45. Canani RB, Costanzo MD, Leone L, Pedata M, Meli R, Calignano A. Potential beneficial
effects of butyrate in intestinal and extraintestinal diseases. World J Gastroenterol.
2011;17(12):1519-28. doi: 10.3748/wjg.v17.i12. 1519

963 10.3748/wjg.v17.i12.1519. PubMed PMID: 21472114; PMCID: PMC3070119.

964 46. Uriot O, Denis S, Junjua M, Roussel Y, Dary-Mourot A, Blanquet-Diot S. Streptococcus
965 thermophilus: From yogurt starter to a new promising probiotic candidate? Journal of Functional
966 Foods. 2017;37:74-89. doi: <u>https://doi.org/10.1016/j.jff.2017.07.038</u>.

967 47. Arboleya S, Watkins C, Stanton C, Ross RP. Gut Bifidobacteria Populations in Human
968 Health and Aging. Front Microbiol. 2016;7:1204. Epub 2016/09/07. doi:
969 10.3389/fmicb.2016.01204. PubMed PMID: 27594848; PMCID: PMC4990546.

970 48. Gueimonde M, Ouwehand A, Huhtinen H, Salminen E, Salminen S. Qualitative and
971 quantitative analyses of the bifidobacterial microbiota in the colonic mucosa of patients with
972 colorectal cancer, diverticulitis and inflammatory bowel disease. World J Gastroenterol.

973 2007;13(29):3985-9. Epub 2007/08/01. doi: 10.3748/wjg.v13.i29.3985. PubMed PMID: 974 17663515; PMCID: PMC4171173.

49. Knudsen JK, Bundgaard-Nielsen C, Hjerrild S, Nielsen RE, Leutscher P, Sorensen S. Gut
microbiota variations in patients diagnosed with major depressive disorder-A systematic review.
Brain Behav. 2021;11(7):e02177. Epub 20210528. doi: 10.1002/brb3.2177. PubMed PMID:

978 34047485; PMCID: PMC8323045.

50. Bailey MT, Dowd SE, Galley JD, Hufnagle AR, Allen RG, Lyte M. Exposure to a social
stressor alters the structure of the intestinal microbiota: implications for stressor-induced
immunomodulation. Brain Behav Immun. 2011;25(3):397-407. Epub 20101030. doi:
10.1016/j.bbi.2010.10.023. PubMed PMID: 21040780; PMCID: PMC3039072.

- 51. Jiang H, Ling Z, Zhang Y, Mao H, Ma Z, Yin Y, Wang W, Tang W, Tan Z, Shi J, Li L, Ruan
  B. Altered fecal microbiota composition in patients with major depressive disorder. Brain Behav
  Immun. 2015;48:186-94. Epub 20150413. doi: 10.1016/j.bbi.2015.03.016. PubMed PMID:
  25882912.
- 52. Chung YE, Chen HC, Chou HL, Chen IM, Lee MS, Chuang LC, Liu YW, Lu ML, Chen CH,
  Wu CS, Huang MC, Liao SC, Ni YH, Lai MS, Shih WL, Kuo PH. Exploration of microbiota targets
  for major depressive disorder and mood related traits. J Psychiatr Res. 2019;111:74-82. Epub
  20190119. doi: 10.1016/j.jpsychires.2019.01.016. PubMed PMID: 30685565.
- 991 53. Wang J, Chen WD, Wang YD. The Relationship Between Gut Microbiota and 992 Inflammatory Diseases: The Role of Macrophages. Front Microbiol. 2020;11:1065. Epub 993 2020/06/26. doi: 10.3389/fmicb.2020.01065. PubMed PMID: 32582063; PMCID: PMC7296120.
- 994 54. Groeger D, O'Mahony L, Murphy EF, Bourke JF, Dinan TG, Kiely B, Shanahan F, Quigley
  995 EM. Bifidobacterium infantis 35624 modulates host inflammatory processes beyond the gut. Gut
  996 Microbes. 2013;4(4):325-39. Epub 2013/07/12. doi: 10.4161/gmic.25487. PubMed PMID:
  997 23842110; PMCID: PMC3744517.
- 55. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie
  H, Zhong H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding
  C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J. Dynamics
  and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe.
  2015;17(5):690-703. Epub 2015/05/15. doi: 10.1016/j.chom.2015.04.004. PubMed PMID:
  25974306.
- 1004 56. Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, Lagoudas GK, Vatanen T,
  1005 Fornelos N, Wilson R, Bertha M, Cohen M, Garber J, Khalili H, Gevers D, Ananthakrishnan AN,
  1006 Kugathasan S, Lander ES, Blainey P, Vlamakis H, Xavier RJ, Huttenhower C. A novel
  1007 Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. Genome Med.
  1008 2017;9(1):103. Epub 2017/12/01. doi: 10.1186/s13073-017-0490-5. PubMed PMID: 29183332;
  1009 PMCID: PMC5704459.
- 1010 57. Hiippala K, Kainulainen V, Kalliomaki M, Arkkila P, Satokari R. Mucosal Prevalence and
  1011 Interactions with the Epithelium Indicate Commensalism of Sutterella spp. Front Microbiol.
  1012 2016;7:1706. Epub 2016/11/12. doi: 10.3389/fmicb.2016.01706. PubMed PMID: 27833600;
  1013 PMCID: PMC5080374.
- 1014 58. Pokusaeva K, Fitzgerald GF, van Sinderen D. Carbohydrate metabolism in Bifidobacteria.
  1015 Genes Nutr. 2011;6(3):285-306. Epub 20110216. doi: 10.1007/s12263-010-0206-6. PubMed
  1016 PMID: 21484167; PMCID: PMC3145055.
- 1017 59. Caputo M, Bona E, Leone I, Sama MT, Nuzzo A, Ferrero A, Aimaretti G, Marzullo P,
  1018 Prodam F. Inositols and metabolic disorders: From farm to bedside. J Tradit Complement Med.
  1019 2020;10(3):252-9. Epub 20200324. doi: 10.1016/j.jtcme.2020.03.005. PubMed PMID: 32670820;
  1020 PMCID: PMC7340869.
- 1021 60. Coupland NJ, Ogilvie CJ, Hegadoren KM, Seres P, Hanstock CC, Allen PS. Decreased
  1022 prefrontal Myo-inositol in major depressive disorder. Biol Psychiatry. 2005;57(12):1526-34. doi:
  1023 10.1016/j.biopsych.2005.02.027. PubMed PMID: 15953489.

1024 61. Nie X, Chen J, Ma X, Ni Y, Shen Y, Yu H, Panagiotou G, Bao Y. A metagenome-wide
1025 association study of gut microbiome and visceral fat accumulation. Comput Struct Biotechnol J.
1026 2020;18:2596-609. Epub 20200920. doi: 10.1016/j.csbj.2020.09.026. PubMed PMID: 33033580;
1027 PMCID: PMC7528071.

1028 62. Perna S, Alalwan TA, Alaali Z, Alnashaba T, Gasparri C, Infantino V, Hammad L, Riva A,
1029 Petrangolini G, Allegrini P, Rondanelli M. The Role of Glutamine in the Complex Interaction
1030 between Gut Microbiota and Health: A Narrative Review. Int J Mol Sci. 2019;20(20). Epub
1031 2019/10/28. doi: 10.3390/ijms20205232. PubMed PMID: 31652531; PMCID: PMC6834172.

Smith SE, Li J, Garbett K, Mirnics K, Patterson PH. Maternal immune activation alters fetal
brain development through interleukin-6. J Neurosci. 2007;27(40):10695-702. Epub 2007/10/05.
doi: 10.1523/JNEUROSCI.2178-07.2007. PubMed PMID: 17913903; PMCID: PMC2387067.

1035 64. Choi GB, Yim YS, Wong H, Kim S, Kim H, Kim SV, Hoeffer CA, Littman DR, Huh JR. The
1036 maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. Science.
1037 2016;351(6276):933-9. Epub 20160128. doi: 10.1126/science.aad0314. PubMed PMID:
1038 26822608; PMCID: PMC4782964.

1039 65. Wu WL, Hsiao EY, Yan Z, Mazmanian SK, Patterson PH. The placental interleukin-6
1040 signaling controls fetal brain development and behavior. Brain Behav Immun. 2017;62:11-23.
1041 Epub 20161109. doi: 10.1016/j.bbi.2016.11.007. PubMed PMID: 27838335; PMCID:
1042 PMC5373986.

1043 66. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular1044 phylogenetic characterization of microbial community imbalances in human inflammatory bowel
1045 diseases. Proceedings of the National Academy of Sciences of the United States of America.
1046 2007;104(34):13780-5. Epub 20070815. doi: 10.1073/pnas.0706625104. PubMed PMID:
1047 17699621; PMCID: PMC1959459.

1048 67. Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, Ter Horst
1049 R, Jansen T, Jacobs L, Bonder MJ, Kurilshikov A, Fu J, Joosten LAB, Zhernakova A, Huttenhower
1050 C, Wijmenga C, Netea MG, Xavier RJ. Linking the Human Gut Microbiome to Inflammatory
1051 Cytokine Production Capacity. Cell. 2016;167(4):1125-36 e8. doi: 10.1016/j.cell.2016.10.020.
1052 PubMed PMID: 27814509; PMCID: PMC5131922.

1053 68. Shahini A, Shahini A. Role of interleukin-6-mediated inflammation in the pathogenesis of
1054 inflammatory bowel disease: focus on the available therapeutic approaches and gut microbiome.
1055 J Cell Commun Signal. 2022. Epub 20220916. doi: 10.1007/s12079-022-00695-x. PubMed PMID:
1056 36112307.

Kanazawa A, Aida M, Yoshida Y, Kaga H, Katahira T, Suzuki L, Tamaki S, Sato J, Goto
H, Azuma K, Shimizu T, Takahashi T, Yamashiro Y, Watada H. Effects of Synbiotic
Supplementation on Chronic Inflammation and the Gut Microbiota in Obese Patients with Type 2
Diabetes Mellitus: A Randomized Controlled Study. Nutrients. 2021;13(2). Epub 20210208. doi:
10.3390/nu13020558. PubMed PMID: 33567701; PMCID: PMC7914668.

1062 Stout MJ, Chubiz J, Raghuraman N, Zhao P, Tuuli MG, Wang LV, Cahill AG, Cuculich PS, 70. 1063 Wang Y, Jungheim ES, Herzog ED, Fay J, Schwartz AL, Macones GA, England SK. A 1064 multidisciplinary Prematurity Research Cohort Study. PLoS One. 2022;17(8):e0272155. Epub 1065 doi: 20220825. 10.1371/journal.pone.0272155. PubMed PMID: 36006907; PMCID: 1066 PMC9409532.

 1067
 71.
 Cuschieri S. The STROBE guidelines. Saudi J Anaesth. 2019;13(Suppl 1):S31-S4. doi:

 1068
 10.4103/sja.SJA\_543\_18. PubMed PMID: 30930717; PMCID: PMC6398292.

1069 72. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible 1070 The Neighborhood Atlas. N Engl J Med. 2018;378(26):2456-8. Epub 2018/06/28. doi:
1071 10.1056/NEJMp1802313. PubMed PMID: 29949490; PMCID: PMC6051533.

1072 73. Institute NC. The Healthy Eating Index – Population Ratio Method Updated December 14,
 1073 2021. Available from: <u>https://epi.grants.cancer.gov/hei/population-ratio-method.html</u>.

1074 74. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the
1075 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry. 1987;150:782-6. doi:
1076 10.1192/bjp.150.6.782. PubMed PMID: 3651732.

1077 75. Cohen S, Kessler RC, Gordon LU. Perceived stress scale: Measuring stress: A guide for
1078 health and social scientists. Cohen S, Kessler RC, Gordon LU, editors. New York, NY, US: Oxford
1079 University Press; 1997. xii, 236-xii, p.

1080 76. Slavich GM, Shields GS. Assessing Lifetime Stress Exposure Using the Stress and
1081 Adversity Inventory for Adults (Adult STRAIN): An Overview and Initial Validation. Psychosom
1082 Med. 2018;80(1):17-27. doi: 10.1097/PSY.00000000000534. PubMed PMID: 29016550;
1083 PMCID: PMC5757659.

1084 77. Lewis TT, Yang FM, Jacobs EA, Fitchett G. Racial/ethnic differences in responses to the
1085 everyday discrimination scale: a differential item functioning analysis. Am J Epidemiol.
1086 2012;175(5):391-401. Epub 20120203. doi: 10.1093/aje/kwr287. PubMed PMID: 22306556;
1087 PMCID: PMC3282874.

1088 78. Fein SB, Labiner-Wolfe J, Shealy KR, Li R, Chen J, Grummer-Strawn LM. Infant Feeding
1089 Practices Study II: study methods. Pediatrics. 2008;122 Suppl 2:S28-35. doi: 10.1542/peds.20081090 1315c. PubMed PMID: 18829828.

1091 79. Planer JD, Peng Y, Kau AL, Blanton LV, Ndao IM, Tarr PI, Warner BB, Gordon JI.
1092 Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice.
1093 Nature. 2016;534(7606):263-6. Epub 20160525. doi: 10.1038/nature17940. PubMed PMID:
1094 27279225; PMCID: PMC4902178.

1095 80. Bennett WE, Jr., Gonzalez-Rivera R, Shaikh N, Magrini V, Boykin M, Warner BB, Hamvas
1096 A, Tarr PI. A method for isolating and analyzing human mRNA from newborn stool. J Immunol
1097 Methods. 2009;349(1-2):56-60. Epub 20090803. doi: 10.1016/j.jim.2009.07.013. PubMed PMID:
1098 19660464; PMCID: PMC2850193.

1099 81. Warner BB, Deych E, Zhou Y, Hall-Moore C, Weinstock GM, Sodergren E, Shaikh N,
1100 Hoffmann JA, Linneman LA, Hamvas A, Khanna G, Rouggly-Nickless LC, Ndao IM, Shands BA,
1101 Escobedo M, Sullivan JE, Radmacher PG, Shannon WD, Tarr PI. Gut bacteria dysbiosis and
1102 necrotising enterocolitis in very low birthweight infants: a prospective case-control study. Lancet.
1103 2016;387(10031):1928-36. Epub 20160309. doi: 10.1016/S0140-6736(16)00081-7. PubMed
1104 PMID: 26969089; PMCID: PMC5553277.

1105 Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, 82. 1106 Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown 1107 CT, Callahan BJ, Caraballo-Rodriguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein 1108 PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz 1109 JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, 1110 Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, 1111 Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu 1112 YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, 1113 Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, 1114 Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson 1115 MS, 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford 1116 AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, 1117 Vargas F, Vazguez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren 1118 J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, 1119 Caporaso JG. Reproducible, interactive, scalable and extensible microbiome data science using 1120 QIIME 2. Nat Biotechnol. 2019;37(8):852-7. Epub 2019/07/26. doi: 10.1038/s41587-019-0209-9. 1121 PubMed PMID: 31341288; PMCID: PMC7015180.

1122 83. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.
1123 The SILVA ribosomal RNA gene database project: improved data processing and web-based
1124 tools. Nucleic Acids Res. 2013;41. doi: 10.1093/nar/gks1219.

1125 84. Balvociute M, Huson DH. SILVA, RDP, Greengenes, NCBI and OTT - how do these
1126 taxonomies compare? BMC Genomics. 2017;18(Suppl 2):114. doi: 10.1186/s12864-017-3501-4.
1127 PubMed PMID: 28361695; PMCID: PMC5374703.

1128 85. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
1129 data. Bioinformatics. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. PubMed PMID:
1130 24695404; PMCID: PMC4103590.

1131 Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, Becker A, 86. 1132 Bennett R, Berry A, Bhai J, Bhurji SK, Bignell A, Boddu S, Branco Lins PR, Brooks L, Ramaraju 1133 SB, Charkhchi M, Cockburn A, Da Rin Fiorretto L, Davidson C, Dodiya K, Donaldson S, El 1134 Houdaigui B, El Naboulsi T, Fatima R, Giron CG, Genez T, Ghattaoraya GS, Martinez JG, Guijarro 1135 C, Hardy M, Hollis Z, Hourlier T, Hunt T, Kay M, Kaykala V, Le T, Lemos D, Margues-Coelho D, 1136 Marugan JC, Merino GA, Mirabueno LP, Mushtaq A, Hossain SN, Ogeh DN, Sakthivel MP, Parker 1137 A, Perry M, Pilizota I, Prosovetskaia I, Perez-Silva JG, Salam AIA, Saraiva-Agostinho N, 1138 Schullenburg H, Sheppard D, Sinha S, Sipos B, Stark W, Steed E, Sukumaran R, Sumathipala 1139 D, Suner MM, Surapaneni L, Sutinen K, Szpak M, Tricomi FF, Urbina-Gomez D, Veidenberg A, 1140 Walsh TA, Walts B, Wass E, Willhoft N, Allen J, Alvarez-Jarreta J, Chakiachvili M, Flint B, Giorgetti 1141 S, Haggerty L, Ilsley GR, Loveland JE, Moore B, Mudge JM, Tate J, Thybert D, Trevanion SJ, 1142 Winterbottom A, Frankish A, Hunt SE, Ruffier M, Cunningham F, Dyer S, Finn RD, Howe KL, 1143 Harrison PW, Yates AD, Flicek P. Ensembl 2023. Nucleic Acids Res. 2022. Epub 20221101. doi: 1144 10.1093/nar/gkac958. PubMed PMID: 36318249.

1145 87. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt C. UniRef clusters: a
1146 comprehensive and scalable alternative for improving sequence similarity searches.
1147 Bioinformatics. 2015;31(6):926-32. Epub 2014/11/16. doi: 10.1093/bioinformatics/btu739.
1148 PubMed PMID: 25398609; PMCID: PMC4375400.

88. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL,
Solymos P, Stevens M, Wagner H. vegan: Community Ecology Package. R package version 2.010. 2013. There is no corresponding record for this reference. 2015.

1152 89. Mason SJ, Graham NE. Areas beneath the relative operating characteristics (ROC) and 1153 relative operating levels (ROL) curves: Statistical significance and interpretation. Quarterly 1154 Journal of the Roval Meteorological Society. 2002;128(584):2145-66. doi: 1155 https://doi.org/10.1256/003590002320603584.

1156 90. Kleine Bardenhorst S, Berger T, Klawonn F, Vital M, Karch A, Rubsamen N. Data Analysis
1157 Strategies for Microbiome Studies in Human Populations-a Systematic Review of Current
1158 Practice. mSystems. 2021;6(1). Epub 20210223. doi: 10.1128/mSystems.01154-20. PubMed
1159 PMID: 33622856; PMCID: PMC8573962.

1160

# 1162 Figures

# 1163 Figure 1.



# 1165 Figure 2:



# 1169 Figure 3



1171

# 1172 Figure 4





# 1175 Figure 5.



1176

# 1177 Figure 6.



# 1180 Figure 7.



# 1183 Figure 8.



# 1186 Supplementary Figure S1



# 1191 Supplementary Figure S2



# 1202 Supplementary Figure S3





