Title: Generally-healthy individuals with aberrant bowel movement frequencies show enrichment for microbially-derived blood metabolites associated with impaired kidney function.

Authors: James P. Johnson[1,2], Christian Diener[1], Anne E. Levine[1,3], Tomasz Wilmanski[1], David L. Suskind[3], Alexandra Ralevski[1], Jennifer Hadlock[1], Andrew T. Magis[1], Leroy Hood[1,2,4,5,6], Noa Rappaport[1], Sean M. Gibbons[1,2,7,8,*]

Affiliations: [1] Institute for Systems Biology, Seattle, WA 98109, USA; [2] Department of Bioengineering, University of Washington, Seattle, WA 98195, USA; [3] Seattle Children's Hospital, Seattle, WA 98105, USA; [4] Phenome Health, Seattle, WA 98109; [5] Department of Immunology, University of Washington, Seattle, WA 98195, USA; [6] Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA 98195, USA; [7] Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [8] eScience Institute, University of Washington, Seattle, WA 98195, USA; [*] correspondence can be addressed to sgibbons@isbscience.org

ABSTRACT

Objective: Bowel movement frequency (BMF) variation has been linked to changes in the composition of the human gut microbiome and to many chronic conditions, like metabolic disorders, neurodegenerative diseases, chronic kidney disease (CKD), and other intestinal pathologies like irritable bowel syndrome (IBS) and inflammatory bowel disease (IBD). Slow intestinal transit times (constipation) are thought to lead to compromised intestinal barrier integrity and a switch from saccharolytic to proteolytic fermentation within the microbiota, giving rise to microbially-derived toxins that may make their way into circulation and cause damage to organ systems. However, these phenomena have not been characterized in generally-healthy populations, and the connections between microbial metabolism and the early-stage development and progression of chronic disease remain underexplored.

Design: Here, we examine the phenotypic impact of BMF variation across a cohort of over 2,000 generally-healthy, community dwelling adults with detailed clinical, lifestyle, and multi-omic data.

Results: We show significant differences in key blood plasma metabolites, proteins, chemistries, gut bacterial genera, and lifestyle factors across BMF groups that have been linked, in particular, to inflammation and CKD severity and progression.

Discussion: In addition to dissecting BMF-related heterogeneity in blood metabolites, proteins, and the gut microbiome, we identify self-reported diet, lifestyle, and psychological factors associated with BMF variation, which suggest several potential strategies for mitigating constipation and diarrhea. Overall, this work highlights the potential for managing BMF to prevent disease.

What is already known about this topic: Constipation and diarrhea are linked to several chronic diseases, like IBD, CKD, and neurodegenerative disorders. Chronic constipation, in particular, is associated with the increased production of microbially-derived uremic toxins in the gut due to an ecosystem-wide switch from fiber fermentation to protein fermentation. A build-up of these gut-derived toxins in blood, like p-cresol, has been associated with CKD disease progression and severity.

What this study adds: While prior work has demonstrated associations between microbially-derived uremic toxins, constipation, and CKD severity/progression, here we show similar signatures in a generally-healthy cohort. Overall, we map out the molecular phenotypic effects of aberrant BMFs across individuals without any apparent disease, and show how

these effects precede, and may contribute to, the development of chronic disease. We find that certain lifestyle and dietary patterns, like higher levels of exercise, reduced anxiety levels, a more plant-based diet, and drinking more water, are associated with a more optimal BMF range.

How this study might affect research, policy, or practice: Overall, we suggest that even mild levels of chronic constipation may cause damage to organ systems over time and ultimately give rise to chronic diseases, like CKD or neurodegeneration. These findings pave the way for future research into early interventions for individuals at risk of developing chronic diseases related to BMF abnormalities. Managing BMF abnormalities prior to disease development may be an important disease prevention strategy, but this will require further evidence through longitudinal human intervention trials.

INTRODUCTION

The gut microbiome influences human health in a number of ways, from mediating early life immune system development [1,2], to determining personalized responses to nutritional interventions [3,4] and influencing the central nervous system [5,6]. Stool transit time, defined as the rate at which stool moves through the gastrointestinal tract, is a major determinant of the composition of the human gut microbiota [7]. Transit time is affected by diet, hydration, physical activity, host mucus production, microbe- and host-derived small molecules (e.g., bile acids or neurotransmitters), and peristaltic smooth muscle contractions in the gastrointestinal tract [8,9]. Stool transit time can be inferred or measured using the Bristol Stool Scale [10], edible dyes [7], indigestible food components (e.g., corn) [11], or self-reported bowel movement frequency (BMF) [12]. Aberrant gastrointestinal transit times have been implicated as risk factors in a number of chronic diseases [13–15].

Shorter stool transit times (e.g. diarrhea, defined as more than three watery stools per day), have been associated with lower gut microbiome alpha diversity, increased susceptibility to enteric pathogens, and poorer overall health [12,16–18]. Longer stool transit times (e.g. constipation, defined as fewer than three hard, dry stools per week), have been associated with higher gut microbiome alpha diversity, with an enrichment in microbially-derived urinary metabolites known to be hepatotoxic or nephrotoxic, and with an increased risk for several chronic medical conditions, including neurological disorders and chronic

82 kidney disease (CKD) severity [13,19–21]. Interestingly, the relationship between higher gut
83 alpha-diversity and constipation contrasts with the common belief that increased diversity is
84 a positive marker of gut health, and suggests a more complex relationship between gut
85 commensal diversity and human health [12,13].

86 Constipation is a known risk factor for CKD severity and end-stage renal disease
87 (ESRD) progression [22,23]. In one study, up to 71% of dialysis patients suffered from
88 constipation [24], while the prevalence of constipation in the general population was 14.5% in
89 adults under 60 years old and 33.5% in those over 60 [25]. A nationwide study of veterans
90 found an incrementally higher risk for renal disease progression in those who reported
91 increasingly severe constipation [26]. However, while it is clear that morbidity and mortality
92 risk worsen with constipation in those with active CKD, potential connections between the
93 gut microbiota and the development and early-stage kidney disease are not yet established.
94 Both constipation and CKD associate with declines in gut microbiota-mediated short-chain
95 fatty acid (SCFA) production and a rise in the production of amino acid putrefaction
96 byproducts, including several toxic metabolites, such as p-cresol sulfate (PCS), which has
97 been causally implicated in renal tissue damage [27]. This is consistent with a community-
98 wide transition from saccharolytic to proteolytic fermentation due to the exhaustion of
99 dietary fiber with longer GI transit times [13,28]. Thus, while the relationships between BMF
100 in healthy individuals and future CKD pathogenesis, along with damage to other organ
101 systems like the central nervous system, are not yet understood, the gut metabolic
102 phenotype associated with low BMF in a prodromal cohort suggests an early causal
103 connection.

104 In this study, we focus on categories of self-reported BMF in a large population of
105 generally-healthy individuals with a wide range of molecular phenotypic data, including data
106 on gut microbiome composition, in order to quantify the phenotypic impact of BMF on blood

107 plasma metabolites, blood proteins, clinical chemistries, and gut microbiome composition in
108 a pre-disease context. By exploring the molecular phenotypic consequences of BMF variation
109 in a generally-healthy cohort, we hope to identify early-stage biomarkers for CKD risk and
110 provide further insight into the possible causal connections between BMF and several
111 chronic, non-communicable diseases. Finally, we assess how demographic, dietary, lifestyle,
112 and psychological factors are associated with variation in BMF, in order to identify potential
113 interventions for manipulating BMF and BMF-associated phenotypes.
114

115 **RESULTS**

116 **A cohort of generally-healthy individuals**

117 3,955 Arivale Scientific Wellness program participants with BMF data were analyzed (see
118 Materials and Methods). Arivale, Inc. (USA), was a consumer scientific wellness company
119 that operated from 2015 until 2019. Briefly, participants consented to having their health,
120 diet, and lifestyle surveyed through an extensive questionnaire, along with blood and stool
121 sampling for multi-omic and blood plasma chemistries data generation (**Fig. 1**). Of those
122 participants that self-reported their ethnicity, 80.5% identified as "White", 10.2% identified as
123 "Asian", 2.9% identified as "Black or African-American", 0.3% identified as "American Indian
124 or Alaska Native", 0.8% identified as "Native Hawaiian or other Pacific Islander", and 5.4%
125 identified as "Other". Additionally, Arivale participants responded 92.9% "Non-Hispanic"
126 versus 7.1% "Hispanic". Of the 109 Hispanic cohort participants, 59.6% also self-reported
127 white. Respondents were in the United States, predominantly from the Pacific West. These
128 individuals were generally-healthy, non-hospitalized and aged between 19 and 87 years old.
129 The population was 61% female with a mean ± s.d. body mass index of 27.47 ± 6.15. Self-
130 reported BMF values (responses to typical number of bowel movements per week) were
131 grouped into four categories (**Fig. 1**), which we defined as: "constipation" (≤ 2 bowel

132 movements per week), “low-normal” (3-6 bowel movements per week), “high-normal” (1-3 bowel
133 movements per day), and “diarrhea” (4 or more bowel movements per day). We first looked at
134 potential associations between BMF and relevant covariates: sex, age, BMI, and estimated
135 glomerular filtration rate (eGFR), a measure of renal function (N = 3,682; **Fig. 2; Table S2**).
136 When BMF was coded as an ordinal dependent variable and regressed using ordered
137 proportional odds logistic regression (POLR), only BMI (POLR, FDR-corrected p = 5.09E-6)
138 and sex (POLR, FDR-corrected P = 1.23E-23) showed significant, independent associations
139 with BMF (**Table S2**), with females and individuals with lower BMIs tending to report lower
140 BMFs (**Fig. 2**). All covariates listed above were included in downstream regressions,
141 independent of whether or not they showed a direct association with BMF. The high-normal
142 BMF group was chosen as the reference for all downstream regressions throughout the
143 manuscript where BMF was encoded as a categorical variable.
144
145 **Gut microbiome composition and activity across BMF categories**
146 For a small subset of the Arivale participants (N=38) shotgun metagenomic sequencing
147 data were available in addition to 16S rRNA gene amplicon sequencing data. For this subset,
148 we calculated peak-to-trough ratios (PTR, a proxy for growth/replication rate) for abundant
149 bacterial taxa within each sample. We saw a significant positive pairwise association
150 between community-average PTRs and BMF (**Fig. 2C**, post-hoc t-test low-normal vs. high-
151 normal, P = 0.010), which suggests that we tend to capture a larger number of commensal
152 bacteria in their exponential growth phase when we sample them from individuals with
153 higher BMFs.
154 Next, we looked at a larger cohort of individuals with 16S amplicon sequencing data
155 from stool (N=2,709). Amplicon sequence variant (ASV) richness (linear regression, P =

156 9.02E-4) and Shannon diversity (linear regression, P = 5.89E-3) were both negatively
157 associated with BMF, independent of covariates (BMF encoded as an ordinal variable with a
158 linear coefficient, **Fig. 3**). Pielou's evenness, on the other hand, was positively associated
159 with BMF (linear regression, P = 1.81E-2), independent of covariates (**Fig. 3**). Thus, slow
160 colonic transit times as seen in constipation correspond to a higher community richness and
161 lower community evenness.

162 Differential abundance analysis of the commensal gut bacterial genera across BMF
163 categories was conducted using beta-binomial regression (CORNCOB) with BMF encoded as
164 a categorical variable. Of the 68 genera that passed our prevalence filter (i.e., detection
165 across ≥ 30% of the individuals), 47 were significantly associated with BMF (see **Table S3** for
166 detailed list of coefficients and p-values), independent of covariates and following an FDR
167 correction for multiple tests on the likelihood ratio test (LRT) P values (LRT, FDR-corrected P
168 < 0.05). Of the 47 significant taxa, we plotted the top ten most abundant (**Fig. 4A-J** and
169 **Table S4**) and the following top 10 most significant taxa (i.e., according to the LRT FDR-
170 corrected P value), including *Akkermansia* (**Fig. 4K-T**). *Bacteroides, Blautia,*
171 *Family_XIII_AD3011_group, Ruminococcaceae_NK4A214_group,* Ruminococcaceae *UBA1819,*
172 *Ruminococcaceae_UCG-005, Anaerotruncus, Butyricicoccus, Lachnospiraceae_UCG-004,*
173 Ruminococcaceae *GCA-900066225,* Ruminococcaceae *Ruminiclostridium_5* were each
174 differentially abundant between constipation and the high-normal (reference) category (LRT,
175 FDR-corrected ratio test P < 0.05). *Agathobacter, Subdoligranulum, Lachnospira,*
176 *Lachnoclostridium, Butyricicoccus*, and *Lachnospiraceae_UCG-004* all showed decreasing
177 abundances with lower BMFs (LRT, FDR-corrected P < 0.05). *Lachnoclostridium* rose in
178 abundance with BMF, and was highest in individuals who reported having diarrhea (LRT,
179 FDR-corrected P < 0.05). In contrast, *Blautia*, *Alistipes*, Ruminococcaceae *UCG-005,*
180 *Ruminococcus_2, UBA1819, Ruminococcaceae_NK4A214_group, Anaerotruncus, GCA-*

181 *900066225,* and *Ruminiclostridium_5* showed the opposite behavior, where decreasing BMF
182 was associated with a increasing abundance of these taxa (LRT, FDR-corrected $P < 0.05$).
183 Some genera appeared to exhibit local minima or maxima (U-shaped vs. peaked relationship
184 with BMF), indicating non-linear trends. These taxa included *Bacteroides*, *Faecalibacterium*,
185 *GCA-900066225, Akkermansia*, and a genus from Family XIII AD3011. However, we had
186 limited power to confidently identify putative non-monotonic trends due to the small number
187 of individuals in the constipation and diarrhea groups.
188
189 **Variation in blood metabolites across BMF categories**
190 Blood metabolite-BMF regression analyses were run using a generalized linear modeling
191 (GLM) framework (LIMMA), with BMF as a categorical variable. Of the metabolites that
192 passed our abundance and prevalence filters (N=1,296, see Materials and Methods), 27
193 unique metabolites were significantly associated with BMF (0 with diarrhea, 24 with low-
194 normal, 4 with constipation, and 1 overlapping metabolite, PCS, associated with both low-
195 normal and constipation), independent of covariates and following an FDR correction for
196 multiple tests (GLM, FDR-corrected $P < 0.05$, **Fig. 5**, **Table S5**). 20 out of 27 metabolites were
197 enriched in the low-normal and/or constipation BMF groups, showing a monotonically
198 decreasing trend with BMF, while the rest showed a monotonically increasing trend (**Fig. 5**).
199 One metabolite, phenylacetylcarnitine, showed a slight, apparent local minimum ("U-
200 shaped" behavior) with lowest levels in the high-normal BMF category (**Fig. 5**). Several
201 unannotated metabolites (e.g. X-12544) showed significant associations with BMF (GLM,
202 FDR-corrected $P < 0.05$), but their identities and physiological roles are unknown (**Table S5**).
203

204 **Blood plasma chemistries across BMF categories**

205 Of the 68 blood plasma chemistries tested, four were significantly different across BMF
206 categories after adjusting for covariates and multiple-testing (N=3,682, GLM, FDR-corrected
207 $P < 0.05$). These included Omega-6/Omega-3 ratio, eicosapentaenoic acid (EPA),
208 homocysteine, and eosinophils levels in the blood (**Fig. 6**). All of these were elevated in the
209 low-normal BMF category compared to the high-normal reference (FDR-corrected $P < 0.05$),
210 except for EPA, which was lower in the low-normal BMF group (**Fig. 6** and **Table S6**).

211

212 **Blood proteins across BMF categories**

213 Of the 274 blood proteins that passed our prevalence filter (see Materials and Methods), 26
214 showed significant associations with BMF after adjusting for covariates and multiple-testing
215 (N=1,999, GLM, FDR-corrected $P < 0.05$). Hepatitis A virus cellular receptor 1 (HAVCR1) was
216 depleted in the low-normal BMF category, relative to the reference group (GLM, FDR-
217 corrected $P < 0.05$). The remaining 25 proteins were significantly depleted in the high BMF
218 (diarrhea) group, relative to the reference group (GLM, FDR-corrected $P < 0.05$). The most
219 significant diarrhea-related protein (GLM, FDR-corrected $P < 0.05$) was TNFRSF11B (tumor
220 necrosis factor receptor superfamily, member 11b; **Fig. 7**, **Table S7**).

221

222 **Self-reported diet, lifestyle, anxiety and depression histories associated with BMF**
223 **categories and demographic covariates**

224 182 survey questions on mental and physical health, diet, and lifestyle were examined from
225 3,002 participants from the Arivale cohort in order to identify covariate-independent
226 associations with BMF. Tests were run using the "polr" package in R (ordinal
227 regression)[29], including the same set of covariates from the prior analyses, and with BMF
228 coded as a categorical variable with high-normal BMF as the reference group (**Fig. 8**).

229 Response categories for each question ascended ordinally in value or intensity (i.e., low to
230 high), so that a positive association represented an increase in that variable. Across the 182
231 questions, the top results with significant odds ratios related to BMF categories were
232 displayed relative to high-normal BMF (**Fig. 8**), colored by the variable category
233 (Diet/Lifestyle or Digestion/Health). BMI, age, and sex were also associated with many of
234 these questionnaire-derived features, independent of BMF (**Fig. 8**). In particular, females took
235 more laxatives, ate more vegetables (including salad and cruciferous vegetables), drank
236 more water, ate breakfast more often, and suffered from greater abdominal pain and
237 bloating. Males, on the other hand, tended to exercise more frequently, drank alcohol more
238 frequently, had an easier time passing bowel movements, and were more likely to have used
239 cholesterol-reducing drugs (**Fig. 8**). Constipation was negatively associated with exercise,
240 alcohol intake, bowel movement completion, diarrhea symptoms, and ease of bowel
241 movement, and positively associated with bloating, cholesterol drug use, reduced appetite,
242 and reported laxative usage, independent of covariates (**Fig. 8**). Membership in the diarrhea
243 BMF category was positively associated with self-reported diarrhea (i.e., a separate question
244 from BMF on the questionnaire), increased bloating, and abdominal pain (**Fig. 8**).

245 A subset of participants self-reported their history of depression and anxiety,
246 including: "self-current", "self-past", and "family" history of depression and anxiety (see
247 Supplement). After logistic regression, one question related to "self-current" history of
248 depression appeared marginally significant (logistic regression, FDR-corrected $P < 0.1$), with
249 a "true" response associated with constipation. Similarly, questions related to a "self-past"
250 (any time) history of anxiety (logistic regression, FDR-corrected $P = 0.01$) and a more recent
251 "self-past" (within the last year) history of anxiety (logistic regression, FDR-corrected $P =$
252 0.048) were significantly associated with constipation.

253

254 **DISCUSSION**

255 In this study, we delve into the multi-omic fingerprint of cross-sectional BMF variation in a
256 large, generally-healthy population. We find that aberrant BMFs are associated with a wide
257 array of phenotypic features, from changes in the ecological composition of the gut
258 microbiota, to variation in plasma metabolites, clinical chemistries, and blood proteins.
259 Overall, we observe an enrichment of microbially-derived uremic toxins resulting from
260 protein fermentation in individuals with lower BMFs. These toxins have been implicated in
261 disease progression and mortality in CKD [23,30] and many of the same metabolites have
262 been associated with other chronic diseases like neurodegeneration [31,32]. We suggest that
263 BMF should be managed throughout the lifespan in order to minimize the build-up of
264 microbially-derived toxins in the blood and to prevent chronic disease. We provide a number
265 of common-sense dietary and lifestyle suggestions for managing BMF, which emerge from
266 our analysis of this generally-healthy cohort.

267

268 **Diet, lifestyle, mood, and demographic factors associated with BMF variation**

269 Of the core set of covariates used in these analyses, only sex and BMI were independently
270 associated with BMF, with females and individuals with lower BMIs showing lower average
271 BMF (**Fig. 2**). Prior work has shown that women are at higher risk of kidney dysfunction [33]
272 and that both BMF and kidney function decline with age [34,35]. In addition to demographic
273 factors associated with BMF, the questionnaire results indicate a number of dietary and
274 lifestyle factors that influence BMF, like exercise frequency, eating fruits and vegetables (i.e.,
275 sources of dietary fiber), sleep, and stress (**Fig. 8**). We also saw evidence that constipation
276 was marginally associated with depression and significantly associated with anxiety, which
277 aligns with prior work showing higher prevalence of anxiety and depression (between 22-
278 33%) on the Hospital Anxiety and Depression Scale (HADS) and the Mini International

279 Neuropsychiatric Interview (MINI) in patients with chronic constipation [36]. The strong
280 positive association between reported cholesterol drug use and constipation suggests that
281 these drugs may influence BMF directly, or perhaps that a "heart healthy" diet/lifestyle that
282 precludes the need for cholesterol medication promotes a healthier BMF range. Diets
283 enriched in complex plant-based carbohydrates, such as starches and fibers, encourage
284 saccharolytic fermentation in the gut microbiome, which likely reduces the level of
285 proteolytic fermentation associated with kidney disease risk and other GI symptoms (**Fig. 8**).
286

287 **The association between BMF and chronic disease may be mediated by the gut**
288 **microbiota**

289 The barrier integrity of the intestinal epithelium, as well as gastrointestinal peristalsis, can
290 be impaired by the enrichment or depletion of certain microbially-derived metabolites
291 [28,37]. BMF-related changes in the composition of the gut microbiota observed in this study
292 reveal a reduction in SCFA-producing genera, like *Bacteroides* and *Faecalibacterium*, in the
293 aberrant BMF groups. Reduced SCFA production is known to weaken smooth muscle
294 contractions that drive peristalsis [38–40], acting as a positive feedback on constipation, and
295 inducing mechanical damage to the epithelium [41–43], which may contribute to subclinical
296 inflammation and disruption of epithelial integrity [30,44,45]. This subclinical inflammation
297 and epithelial damage may give rise to chronic peripheral and systemic inflammation over
298 time and allow for excess luminal metabolites to leak into the blood, which can drive tissue
299 damage throughout the body and exacerbate conditions like CKD [30,46–48].

300 Many of the genera and metabolites that were associated with constipation in this
301 study have been associated with constipation in other disease cohorts and with a variety of
302 risk factors for chronic diseases, like CKD, cardiovascular disease, and metabolic syndrome
303 [8,23,49,50]. *Alistipes* and *Ruminococcus* were enriched in end-stage renal disease (ESRD)

304 patients [51], as well as in our generally-healthy cohort at lower BMF levels (**Fig. 4**). In
305 general, families like Ruminococcaceae and Lachnospiraceae dominate the pool of significant
306 BMF-microbiome hits (**Fig. 4**). In particular, *Roseburia*, a genus in the Lachnospiraceae family
307 observed to be lower in abundance at all stages of CKD and ESRD [52], was found to be
308 lower in abundance in individuals with lower BMFs in our cohort (**Fig. 4**). *Akkermansia,* a
309 mucus-degrading genus generally associated with metabolic health [53], but also enriched in
310 patients with Parkinson's disease (PD) and in constipated individuals [32,54], was enriched
311 at lower BMF in our cohort (**Fig. 4**). *Akkermansia* was positively associated with constipation
312 across several studies [54], likely due to its specialization on breaking down host mucus
313 rather than dietary substrates, but its absence also appears to have a detrimental impact on
314 metabolic health and CKD progression [53,55,56]. Finally, we saw that the average gut
315 bacterial community replication rate was positively associated with BMF (**Fig. 2**) and
316 negatively associated with the production of several protein fermentation byproducts that
317 are known uremic toxins (**Fig. 5**). Findings such as these suggest that constipation may drive
318 pre-clinical risk and progression towards chronic diseases, mediated in part by BMF-induced
319 switch from saccharolytic to proteolytic metabolism in the gut microbiota.
320

321 **BMF-associated blood metabolites are implicated in chronic disease risk and severity**

322 Several blood metabolites found to be enriched at lower BMF were gut microbiome-derived
323 uremic toxins linked to kidney function decline and neurodegenerative diseases. PCS, for
324 example, has been associated with deteriorating kidney function and with damage to
325 nephrons [57,58]. PCS showed the strongest association with BMF (**Fig. 5**), exhibiting a dose-
326 response effect, increasing substantially in both the low-normal and constipation categories
327 (**Fig. 5**). P-cresol glucuronide (PCG) is another uremic toxin, derived from microbe-produced
328 p-cresol, which was significantly enriched at lower BMF (**Fig. 5**). Overall, we see an

329 enrichment in several microbially-derived toxins in the blood of generally-healthy individuals
330 with lower BMFs, like PCS, PCG, phenylacetylglutamine, 6-hydroxyindole sulfate, and
331 phenylacetylcarnitine [58–60], which may drive long-term chronic disease risk.
332
333 **BMF-associated blood plasma chemistries results linked to inflammation and diet**
334 Eicosapentaenoic acid (EPA) levels were lower in the lower-BMF groups (**Fig. 6**). Higher
335 levels of EPA have been associated with lower inflammation [61] and lower cardiovascular
336 disease risk [62]. Conversely, the Omega-6/Omega-3 ratio, homocysteine levels, and
337 eosinophil counts, have all been positively associated with inflammation [63,64], and these
338 features were elevated in the low-normal BMF group (**Fig. 6**). The Omega-6/Omega-3 ratio, in
339 particular, may be related to higher levels of pro-inflammatory Omega-6 lipids and lower
340 levels of anti-inflammatory Omega-3 lipids in the diet [65]. A diet enriched in processed foods
341 and animal products is known to drive increased risk of chronic kidney disease [66,67]. The
342 directionality of these associations point towards lower BMFs being associated with higher
343 systemic inflammation, which may lead to increased chronic disease risk potentially through
344 compromised gut epithelia.
345
346 **BMF-associated proteins connected to inflammation and renal injury**
347 Hepatitis A virus cellular receptor 1 (HAVCR1) was the only protein that was depleted in the
348 low-normal BMF group (**Fig. 7A**). HAVCR1 is, notably, an early biomarker for acute renal
349 injury and a predictor of long-term renal disease, as it is shed into the urine following kidney
350 injury [68]. Tumor necrosis factor receptor superfamily, member 11b (TNFRSF11B) showed
351 the strongest association with BMF and was enriched in individuals with diarrhea (**Fig. 7B**).
352 TNFRSF11B dysregulation has been associated with osteoporosis and with a number of
353 cancers, and TNFSF members are involved in the pathogenesis of irritable bowel syndrome

354 (IBS), a disease often associated with diarrhea [69,70]. The remaining proteins associated
355 with aberrant BMFs were related to inflammation and undesirable immune responses, organ
356 damage, and cancer (**Fig. 7**) [71,72].
357
358 **Current limitations and considerations on designing future research**
359 There are some important limitations to consider when interpreting the results of this study.
360 The generally-healthy cohort studied here was overwhelmingly "White", predominantly
361 female, and from the West Coast of the US, which limits the generalizability of these results.
362 In addition, the diet, lifestyle, and mood data were self-reported and subject to biases and
363 errors, and are not indicative of clinical diagnoses. In designing future follow-up trials, it
364 would be ideal to manage BMF as a preventative measure for chronic disease and to target
365 interventions that are low-risk with fewer side effects than drugs like laxatives. For example,
366 BMF can be managed through exercise, hydration, and diet. However, high-fiber diets can
367 lead to bloating and other issues in those with active disease. CKD patients, usually on
368 multiple medications that may affect gut health and BMF, often need to eat a diet that
369 restricts many plant-based fiber-rich foods because they contain high levels of potassium
370 and phosphorus [73]. However, these low-fiber diets may act as a positive feedback on
371 constipation and inflammation, as they promote protein fermentation in the gut. This
372 highlights the importance of intervening at the prodromal stage, before disease manifests,
373 when a healthy, plant-based diet is well-tolerated by the individual. Alternatively, low-
374 potassium and low-phosphorus, high-fiber diets could be formulated for CKD patients.
375 Ultimately, future work should be done to assess the potential for managing BMF throughout
376 the lifespan to reduce chronic disease risk.

377

378 **Conclusion**

379 Bowel movement abnormalities, such as constipation or diarrhea, have been linked to
380 diseases ranging from enteric infections [18], CKD, and IBD to dementia-related
381 neurodegenerative diseases like Alzheimer's disease (AD) and PD [31,74,75]. Indeed, we see
382 many of the phenotypic markers of these diseases manifested in generally-healthy
383 individuals who report having aberrant BMFs, with constipation in particular associated
384 with a build up of microbially-derived uremic toxins in the blood. Mitigating chronic
385 constipation may be key to reducing uremic, hepatic, and neurological toxin build-up in the
386 blood. Our results underscore common-sense dietary and lifestyle changes, like increasing
387 dietary fiber intake, eating a lower protein diet and exercising more, may help to normalize
388 BMF and reduce BMF-associated phenotypic risk factors for chronic disease, well before the
389 onset of disease.

390

391 **MATERIALS AND METHODS**

392 **Institutional review board approval for the study**

393 The procedures for this study were reviewed and approved by the Western Institutional
394 Review Board, under the institutional review board study number 20170658 for the Institute
395 for Systems Biology and 1178906 for Arivale, Inc.

396

397 **Patient and Public Involvement Statement**

398 There was no patient or public involvement in the conception or implementation of this
399 research study.

400

401 **Generally-healthy cohort**

402 All study participants were subscribers in the Arivale Scientific Wellness program (2015-
403 2019) and provide informed consent for the use of their anonymized, deidentified data for
404 research purposes. Participants were community-dwelling, representative of the populations
405 in Washington State and California (which are slightly leaner and healthier than other parts
406 of the USA), over the age of 18, non-pregnant, but were not screened for the presence or
407 absence of any particular disease. Participants provided questionnaire data, along with
408 blood and stool samples that were used to generate blood plasma metabolomics, proteomics,
409 chemistries, and gut microbiome data (**Fig 1** and **Table S1**).

410 Only baseline time point samples were used for each participant, prior to the
411 beginning of a personalized wellness intervention. A 70% prevalence filter was implemented
412 across the gut microbiome, blood plasma metabolomics, proteomics, chemistries, and ordinal
413 questionnaire data analyses. This meant that each final feature in the data could contain no
414 more than 30% missing data from the final cohort of samples in order to be retained for
415 downstream analysis. For microbiome analyses, a filtered subcohort of 2,709 individuals with
416 ASV-level taxa counts, BMF, sex, age, eGFR, and BMI data were selected. This filtering
417 resulted in a total of 68 genera. For the metabolomics analysis, a cohort of 2,043 participants
418 with BMF, sex, age, eGFR, BMI, and blood metabolomics data were selected. 973
419 metabolites were retained for downstream analyses. 274 blood proteins that met the
420 prevalence filter in the cohort of 1,999 individuals were retained for downstream analyses. A
421 similar prevalence filter was applied to 3,682 samples with blood plasma chemistries data,
422 resulting in 68 features retained for downstream analyses. Similarly, for ordinal regression of
423 the questionnaire data (e.g. diet, lifestyle, and stress/pain/health factors,) using the
424 respective R package, polr [29], we collated all the responses and filtered out questions that
425 contained more than 30% "NAs". We also excluded binary responses, which are incompatible

426 with ordinal regression using polr, which resulted in 277 variables across 2,291 participants,
427 in addition to having paired data on age, sex, eGFR, BMI, and BMF. BMF data was captured
428 from responses to a survey question on how many bowel movements an individual has per
429 week, on average. The available responses to this question were: (1) Twice per week or less;
430 (2) 3-6 times per week; (3) 1-3 times daily; or (4) 4 or more times daily. While the normal
431 range of BMF encompasses both the second and third responses to this question (i.e.,
432 between three times a week and three times a day) [76], we chose to define 1-3 times per
433 day (high-normal) as the reference group for the purposes of regression.
434
435 **Gut Microbiome Data**
436 Fecal samples from Arivale participants were collected (described in Diener et al [12] and
437 detailed here) from proprietary at-home kits developed by two microbiome vendors (DNA
438 Genotek and Second Genome) that stabilize the DNA collected at ambient room temperature.
439 Using the KingFisher Flex instrument, the MoBio PowerMag Soil DNA isolation kit (QIAGEN)
440 enabled the isolation of stool DNA from 250 ml of homogenized human feces, after performing
441 an additional glass bead-beating step. Qubit measurement and spectrophotometry were also
442 performed using an A260/A280 absorbance ratio. Either 250-bp paired-end MiSeq profiling of
443 the 16S V4 region (Second Genome, USA) or the 300-bp paired-end MiSeq profiling of the 16S
444 V3-V4 region (DNA Genotek, USA) was used to obtain the raw amplicon sequencing data
445 (ASVs).
446 16S amplicon sequencing was run on a MiSeq (Illumina, USA) with either paired-end
447 300-bp protocol (DNA Genotek) or paired-end 250-bp protocol (SecondGenome). The FASTQ
448 files were provided by the Illumina Basespace platform after the phiX reads were removed
449 with basecalling. Length cutoffs of 250-bp for the forward reads and 230-bp for the reverse
450 reads as well as manual inspection of the error rate across sequencing cycles were

451 determined from the respective profiles. Any greater than 2 expected errors under the
452 Illumina error model resulted in eliminating that specific read from the data along with reads
453 containing ambiguous ("N" nucleotides) base calls. Over 97% of the reads passed these
454 filters, resulting in approximately 200,000 reads per sample.

455 Shotgun metagenomic sequencing libraries for Arivale samples were prepared by
456 DNA Genotek using the NexteraXT kit, along with QC on a Bioanalyzer and quantification of
457 DNA using qPCR for pooling. Sequencing was run on an Illumina NovaSeq6000 (300-
458 multiplex on S2 flow cell), with a paired-end 150-bp protocol. The target sequencing depth
459 was 3Gb, equivalent to about 20M total reads per sample.

460 Final truncated and filtered reads were then used to infer amplicon sequence variants
461 (ASV) with DADA2. Each sequencing run separately resulted in its own error profiles. The
462 final ASVs and counts were then joined, with chimeras being removed using DADA2's
463 "consensus" strategy. After this step, almost 16% of all reads were removed. Taxonomic
464 assignment of ASVs was then achieved using the naive Bayes classifier in DADA2 with the
465 SILVA database (version 128).

466 Wherever possible, the 16S gene in SILVA was used to perform by using an exact
467 match of the inferred ASV to the gene. Nearly 90% of the reads were able to be classified
468 down to the genus level, which was the taxonomic level chosen for this analysis. 3,694
469 samples across 609 taxa were available from these methods, which were then filtered down
470 to 68 taxa after using a 70% prevalence filter (no more than 30% of data was permitted to be
471 missing per filtered taxa). The diversity of the gut microbiomes of the cohort was
472 characterized and rarefied to an even depth across ASVs where count parity is preserved
473 across samples. Observed ASVs, a measure of species diversity, were used to obtain
474 Shannon diversity and Pielou's evenness. After BMI, sex, age, and eGFR data were merged
475 to the taxa dataset, 2,709 samples remained across the 68 taxa.

476 The diversity of the gut microbiomes of the cohort was characterized and rarefied to
477 an even depth (using the "rarefy_even_depth( )" function in the phyloseq R package [77]; rng
478 seed = 111) with observed amplicon sequence variants (ASV), a measure of species
479 diversity, to obtain Shannon diversity and Pielou's evenness.
480
481 **Olink Proteomics**
482 Blood plasma proteomic data were generated by Olink Biosciences using the ProSeek
483 Cardiovascular II, Cardiovascular III, and Inflammation arrays. The proteins were filtered
484 down to 274 proteins and 1,999 samples and included based on whether or not they had 30%
485 or less missingness across samples as well as BMI, sex, age, and BMF data. NA data values
486 were assumed to be below detection and imputed to be the median across samples for that
487 particular protein. The values used for the proteomics analysis were from protein readings
488 previously batch-corrected and normalized based on the overlapping reference samples
489 within the batch plates. The corrected values were also scale-shifted to the reference sample
490 and the original delivered data (using the seventh run as a baseline). The method is
491 described further in the study by Zubair et al [78]. All data were merged with BMI, sex, age,
492 and eGFR data for the cohort.
493
494 **Metabolon Metabolomics**
495 Metabolon obtained metabolomics data on the previously mentioned plasma samples using
496 preparation, quality control, and collection methods described in previous studies [50]. 2,043
497 total metabolites across 1,297 samples were filtered down using the same prevalence filter
498 as for proteins. In this analysis, missing values were imputed to be the median of the non-
499 missing samples for the metabolite, and final downstream metabolites were log-transformed
500 and merged with available BMI, sex, age, and eGFR data.

Blood Plasma Chemistries

LabCorp and Quest phlebotomists collected blood from Arivale participants within 21 days of their gut microbiome samples being taken, during the same blood draw as the metabolomics and using methods described previously by Wilmanski et al and others [12]. Individuals were asked to abstain from alcohol, vigorous exercise, monosodium glutamate and aspartame at least 24 hours prior to drawing of the blood, as well as fasting at least 12 hours beforehand. Blood samples were collected for blood plasma chemistries, metabolomics and proteomics at the same time, and within 21 days of stool sampling. BMI was calculated from weight and height using the following formula $BMI = \frac{weight\ (kg)}{(height\ (m))^2}$. 4,881 samples and 127 laboratory values were filtered down using the same prevalence filtering as with metabolomics and proteomics. The final 68 features were log-transformed, with missing samples imputed to be the median value of the non-missing samples. These features were merged with other data and covariates. eGFR was calculated based on the CKD Epidemiology Collaboration (CKD-EPI) creatinine Equation (2021), as recommended by the current guidelines of the National Kidney Foundation [cite PMID: 34563581]: $eGFR_{cr} = 142 \times \min(Scr/\kappa, 1)^{\alpha} \times \max(Scr/\kappa, 1)^{-1.200} \times 0.9938^{Age} \times 1.012$ [if female], where Scr = standardized serum creatinine in mg/dL, κ = 0.7 (female) or 0.9 (male), and α = -0.241 (female) or -0.302 (male).

Questionnaire Data

4,402 self-reported results to questionnaire data with 3,002 samples were retrieved from Arivale participants at the beginning of the study. After filtration, 283 downstream features remained, which were subsequently filtered down again to 182 question features by removing factored features with less than 10 responses per level and at least 2 nonmissing

526 levels to the factor. Category responses were organized and numbered to be ordinally
527 ascending in magnitude or intensity with relatively even-spaced differences in magnitude
528 between categories wherever possible (i.e. for a factored feature with levels from 1,…,n, the
529 level labeled "1" represented responses such as "Strongly Disagree", "Never", "None", or
530 the lowest frequency/intensity, and the level labeled "n" represented responses such as
531 "Strongly Agree", "Always", or the greatest frequency/intensity). These features were
532 merged with BMI, sex, age, and eGFR data available for this subcohort.
533
534 **Depression and Anxiety Health History Data**
535 We used logistic regression to scrutinize associations between 11 (anxiety) and 10
536 (depression) independent binary ("true" or "false") self-reported questions based on asking
537 self-reported "self-current", "self-past", and "family" histories of depression or anxiety and
538 BMF, with depression or anxiety encoded as a binary dependent variable, and BMF encoded
539 as a categorical independent variable, and with the standard set of covariates (sex, age, BMI,
540 and eGFR).
541
542 **Statistical Analyses**
543 For the blood proteomics, plasma chemistries, and metabolite associations, generalized
544 linear regression models were run using the LIMMA package in R [79]. BMF was encoded as
545 a categorical variable (or in the case of analyzing alpha-diversity, it was also computed as an
546 ordinal variable with a linear model coefficient) with categories: 1 = constipation (1-2 bowel
547 movements per week), 2 = low-normal (3-6 bowel movements per week), 3 = high-normal (1-
548 3 bowel movements per day), and 4 = diarrhea (4 or more bowel movements per day). In
549 each regression covariates BMI, sex, age, and eGFR were included, in addition to BMF, to the
550 response variable. The response variables were either: centered log ratio-transformed taxa

551 data, log-transformed plasma metabolomics data, corrected plasma proteomics data, log-
552 transformed plasma chemistries data, or ordinal response variables from questionnaire data,
553 depending on the analysis. For gut microbiome data, genus-level counts were modeled with
554 a beta-binomial distribution using the CORNCOB package in R [80]. Finally, for the
555 questionnaire data (ordinal response categories across diet, exercise, stress, pain, and other
556 lifestyle factors), the depression questions data, and the anxiety questions data, polr in R
557 was used for the ordinal regression analysis.
558
559 **Community Replication Rate (PTR) of Gut Microbiome**
560 FASTQ files from the metagenomic shotgun sequencing were first filtered and trimmed using
561 FASTP. Here the first 5-bp of the 5' end of the read were trimmed to remove partial adapter
562 sequences and the 3' end was trimmed using a sliding window that would trim the read as
563 soon as the window average fell below a quality score of 20. Reads shorter than 50-bp after
564 trimming or with more than 1 ambiguous base call were removed from further analysis.
565 Filtered and trimmed reads were then passed to COPTR to estimate PTRs [81,82]. In brief,
566 preprocessed reads were aligned to a database of 2,935 species representative genomes
567 from the human gut contained in the IGG database version 1.0 using BOWTIE2. Coverage
568 profiles were extracted from the generated alignments and log2-transformed PTRs were
569 calculated by COPTR for each reference genome with at least 5,000 mapped reads. For each
570 sample an overall measure of bacterial replication was estimated as the mean of all log2 PTR
571 estimates in the sample. The mean log2 PTR was then used in a regression model as the
572 dependent variable and regressed against BMF categories correcting for sex, age, and BMI.
573 Significant associations with overall BMF were obtained from an F-test comparing the full
574 model with a nested model containing only the confounding variables.
575

576 **Data Availability**

577 Code used to analyze 16S rRNA gene amplicon sequencing data can be found at
578 https://github.com/gibbons-lab/mbtools while code used to run the statistical analysis
579 described in this paper is available at https://github.com/jajohnso29/Generally-Healthy-
580 Cohort-BMF .

581 Pipelines for the processing of the metagenomic shotgun sequencing data and
582 estimation of PTRs can be found at https://github.com/gibbons-lab/pipelines.

583 Qualified researchers can access the full Arivale deidentified dataset, including all
584 raw data, supporting the findings in this study for research purposes through signing a Data
585 Use Agreement (DUA). Inquiries to access the data can be made at data-
586 access@isbscience.org and will be responded to within 7 business days.

587

588 **ACKNOWLEDGMENTS**

589 This research was funded by Washington Research Foundation Distinguished Investigator
590 Award and by startup funds from the Institute for Systems Biology (to SMG). Research
591 reported in this publication was supported by the National Institute of Diabetes and
592 Digestive and Kidney Diseases of the National Institutes of Health (NIH) under award no.
593 R01DK133468 (to SMG). The content is solely the responsibility of the authors and does not
594 necessarily represent the official views of the NIH. The funders had no role in designing,
595 carrying out or interpreting the work presented in the manuscript.

596

597 **Patient and Public Involvement**

598 There was no patient or public involvement in the conception or implementation of this
599 research study.

600

601 **Ethics Declaration**

602 L.H. is a former shareholder of Arivale. A.T.M. was a former employee of Arivale. Arivale is no
603 longer a commercially operating company as of April 2019. The remaining authors report no
604 competing interests.

605

606 **REFERENCES**

607 1 Magalhães-Guedes KT. Psychobiotic Therapy: Method to Reinforce the Immune System.
608 *Clin Psychopharmacol Neurosci* 2022;**20**:17–25. doi:10.9758/cpn.2022.20.1.17

609 2 Kim J, Lee HK. Potential Role of the Gut Microbiome In Colorectal Cancer Progression.
610 *Front Immunol* 2021;**12**:807648. doi:10.3389/fimmu.2021.807648

611 3 Hughes RL. A Review of the Role of the Gut Microbiome in Personalized Sports Nutrition.
612 Frontiers in Nutrition. 2020;**6**. doi:10.3389/fnut.2019.00191

613 4 Ceballos D, Hernández-Camba A, Ramos L. Diet and microbiome in the beginning of the
614 sequence of gut inflammation. *World J Clin Cases* 2021;**9**:11122–47.
615 doi:10.12998/wjcc.v9.i36.11122

616 5 Cryan JF, O'Riordan KJ, Cowan CSM, *et al.* The Microbiota-Gut-Brain Axis. *Physiol Rev*
617 2019;**99**:1877–2013. doi:10.1152/physrev.00018.2018

618 6 Kim DS, Zhang T, Park S. Protective effects of and water extract against memory deficits
619 through the gut-microbiome-brain axis in an Alzheimer's disease model. *Pharm Biol*
620 2022;**60**:212–24. doi:10.1080/13880209.2022.2025860

621 7 Asnicar F, Leeming ER, Dimidi E, *et al.* Blue poo: impact of gut transit time on the gut
622 microbiome using a novel marker. *Gut* 2021;**70**:1665–74. doi:10.1136/gutjnl-2020-323877

623 8 Müller M, Canfora EE, Blaak EE. Gastrointestinal Transit Time, Glucose Homeostasis
624 and Metabolic Health: Modulation by Dietary Fibers. *Nutrients* 2018;**10**.
625 doi:10.3390/nu10030275

626 9 Vanuytsel T, Tack J, Farre R. The Role of Intestinal Permeability in Gastrointestinal
627 Disorders and Current Methods of Evaluation. *Front Nutr* 2021;**8**:717925.
628 doi:10.3389/fnut.2021.717925

629 10 Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand*
630 *J Gastroenterol* 1997;**32**:920–4. doi:10.3109/00365529709011203

631 11 Keendjele TPT, Eelu HH, Nashihanga TE, *et al.* Corn? When did I eat corn?
632 Gastrointestinal transit time in health science students. *Adv Physiol Educ* 2021;**45**:103–8.
633 doi:10.1152/advan.00192.2020

634 12 Wilmanski T, Rappaport N, Earls JC, *et al.* Blood metabolome predicts gut microbiome -

635 diversity in humans. Nature Biotechnology. 2019;**37**:1217–28. doi:10.1038/s41587-019-
636 0233-9

637 13 Roager HM, Hansen LBS, Bahl MI, *et al.* Colonic transit time is related to bacterial
638 metabolism and mucosal turnover in the gut. *Nature Microbiology* 2016;**1**:1–9.
639 doi:10.1038/nmicrobiol.2016.93

640 14 Wiskur B, Meerveld BG-V. The Aging Colon: The Role of Enteric Neurodegeneration in
641 Constipation. *Current Gastroenterology Reports* 2010;**12**:507–12. doi:10.1007/s11894-010-
642 0139-7

643 15 Stirpe P, Hoffman M, Badiali D, *et al.* Constipation: an emerging risk factor for
644 Parkinson's disease? *European Journal of Neurology* 2016;**23**:1606–13.
645 doi:10.1111/ene.13082

646 16 Duvallet C, Gibbons SM, Gurry T, *et al.* Meta-analysis of gut microbiome studies
647 identifies disease-specific and shared responses. *Nat Commun* 2017;**8**:1784.
648 doi:10.1038/s41467-017-01973-8

649 17 McDonald D, Hyde E, Debelius JW, *et al.* American Gut: an Open Platform for Citizen
650 Science Microbiome Research. *mSystems* 2018;**3**. doi:10.1128/mSystems.00031-18

651 18 Tomkovich S, Taylor A, King J, *et al.* An Osmotic Laxative Renders Mice Susceptible to
652 Prolonged Clostridioides difficile Colonization and Hinders Clearance. *mSphere* 2021;**6**.
653 doi:10.1128/mSphere.00629-21

654 19 Falony G, Joossens M, Vieira-Silva S, *et al.* Population-level analysis of gut microbiome
655 variation. *Science* 2016;**352**:560–4. doi:10.1126/science.aad3503

656 20 Vandeputte D, Falony G, Vieira-Silva S, *et al.* Stool consistency is strongly associated
657 with gut microbiota richness and composition, enterotypes and bacterial growth rates.
658 Gut. 2016;**65**:57–62. doi:10.1136/gutjnl-2015-309618

659 21 Adams-Carr KL, Bestwick JP, Shribman S, *et al.* Constipation preceding Parkinson's
660 disease: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry*
661 2016;**87**:710–6. doi:10.1136/jnnp-2015-311680

662 22 Ramos CI, Nerbass FB, Cuppari L. Constipation in Chronic Kidney Disease: It Is Time to
663 Bridge the Gap. Kidney and Dialysis. 2022;**2**:221–33. doi:10.3390/kidneydial2020023

664 23 Sumida K, Yamagata K, Kovesdy CP. Constipation in CKD. *Kidney Int Rep* 2020;**5**:121–34.
665 doi:10.1016/j.ekir.2019.11.002

666 24 Ikee R, Sasaki N, Yasuda T, *et al.* Chronic Kidney Disease, Gut Dysbiosis, and
667 Constipation: A Burdensome Triplet. Microorganisms. 2020;**8**:1862.
668 doi:10.3390/microorganisms8121862

669 25 Gr BAPJ. American Gastroenterological Association technical review on constipation.
670 *Gastroenterology* 2013;**144**:218–38. doi:10.1053/j.gastro.2012.10.028.

671 26 Sumida K, Molnar MZ, Potukuchi PK, *et al.* Constipation and Incident CKD. Journal of the
672 American Society of Nephrology. 2017;**28**:1248–58. doi:10.1681/asn.2016060656

673 27 Brocca A, Virzì GM, de Cal M, *et al.* Cytotoxic effects of p-cresol in renal epithelial
674 tubular cells. *Blood Purif* 2013;**36**:219–25. doi:10.1159/000356370

675 28 Ikee R, Yano K, Tsuru T. Constipation in chronic kidney disease: it is time to reconsider.
676 Renal Replacement Therapy. 2019;**5**. doi:10.1186/s41100-019-0246-3

677 29 polr function - RDocumentation.
678 https://www.rdocumentation.org/packages/MASS/versions/7.3-58.1/topics/polr
679 (accessed 8 Feb 2023).

680 30 Wan M, King L, Baugh N, *et al.* Gutted: constipation in children with chronic kidney
681 disease and on dialysis. *Pediatr Nephrol* Published Online First: 9 January 2023.
682 doi:10.1007/s00467-022-05849-y

683 31 Kim JE, Park JJ, Lee MR, *et al.* Constipation in Tg2576 mice model for Alzheimer's
684 disease associated with dysregulation of mechanism involving the mAChR signaling
685 pathway and ER stress response. *PLoS One* 2019;**14**:e0215205.
686 doi:10.1371/journal.pone.0215205

687 32 Baldini F, Hertel J, Sandt E, *et al.* Parkinson's disease-associated alterations of the gut
688 microbiome predict disease-relevant changes in metabolic functions. *BMC Biol*
689 2020;**18**:62. doi:10.1186/s12915-020-00775-7

690 33 Neugarten J, Golestaneh L. Gender and the Prevalence and Progression of Renal
691 Disease. Advances in Chronic Kidney Disease. 2013;**20**:390–5.
692 doi:10.1053/j.ackd.2013.05.004

693 34 Li G, Chen Y, Hu H, *et al.* Association between age-related decline of kidney function
694 and plasma malondialdehyde. *Rejuvenation Res* 2012;**15**:257–64.
695 doi:10.1089/rej.2011.1259

696 35 Coresh J, Selvin E, Stevens LA, *et al.* Prevalence of Chronic Kidney Disease in the United
697 States. *JAMA* 2007;**298**:2038–47. doi:10.1001/jama.298.17.2038

698 36 Hosseinzadeh ST, Poorsaadati S, Radkani B, *et al.* Psychological disorders in patients
699 with chronic constipation. *Gastroenterol Hepatol Bed Bench* 2011;**4**:159–
700 63.https://www.ncbi.nlm.nih.gov/pubmed/24834176

701 37 Glorieux G, Gryp T, Perna A. Gut-Derived Metabolites and Their Role in Immune
702 Dysfunction in Chronic Kidney Disease. *Toxins* 2020;**12**. doi:10.3390/toxins12040245

703 38 Martin-Gallausiaux C, Marinelli L, Blottière HM, *et al.* SCFA: mechanisms and functional
704 importance in the gut. *Proc Nutr Soc* 2021;**80**:37–49. doi:10.1017/S0029665120006916

705 39 Zhao Y, Liu Q, Hou Y, *et al.* Alleviating effects of gut micro-ecologically regulatory
706 treatments on mice with constipation. *Front Microbiol* 2022;**13**:956438.
707 doi:10.3389/fmicb.2022.956438

708 40 Vicentini FA, Keenan CM, Wallace LE, *et al.* Intestinal microbiota shapes gut physiology
709 and regulates enteric neurons and glia. *Microbiome* 2021;**9**:210. doi:10.1186/s40168-021-
710 01165-z

41 Bodnar HB, Bodnar BM. [Peculiarities of a large bowel microstructure in children, suffering chronic constipation, caused by its inborn anomaly]. *Klin Khir* 2014;:30–3.https://www.ncbi.nlm.nih.gov/pubmed/25097973

42 Xiao Y, Kashyap PC. Microbially derived polyunsaturated fatty acid as a modulator of gastrointestinal motility. J. Clin. Invest. 2022;**132**. doi:10.1172/JCI161572

43 Negussie AB, Dell AC, Davis BA, *et al.* Colonic Fluid and Electrolyte Transport 2022: An Update. *Cells* 2022;**11**. doi:10.3390/cells11101712

44 Clayburgh DR, Shen L, Turner JR. A porous defense: the leaky epithelial barrier in intestinal disease. *Lab Invest* 2004;**84**:282–91. doi:10.1038/labinvest.3700050

45 Mitsuoka T. Intestinal flora and human health. *Asia Pac J Clin Nutr* 1996;**5**:2–9.https://www.ncbi.nlm.nih.gov/pubmed/24394457

46 Usuda H, Okamoto T, Wada K. Leaky Gut: Effect of Dietary Fiber and Fats on Microbiome and Intestinal Barrier. *Int J Mol Sci* 2021;**22**. doi:10.3390/ijms22147613

47 Zhang L, Xie F, Tang H, *et al.* Gut microbial metabolite TMAO increases peritoneal inflammation and peritonitis risk in peritoneal dialysis patients. *Transl Res* 2022;**240**:50–63. doi:10.1016/j.trsl.2021.10.001

48 Yu W, Venkatraman A, Menden HL, *et al.* Short-chain fatty acids ameliorate necrotizing enterocolitis-like intestinal injury through enhancing Notch1-mediated single immunoglobulin interleukin-1-related receptor, toll-interacting protein, and A20 induction. *Am J Physiol Gastrointest Liver Physiol* 2023;**324**:G24–37. doi:10.1152/ajpgi.00057.2022

49 Rysz J, Franczyk B, Ławi□ski J, *et al.* The Impact of CKD on Uremic Toxins and Gut Microbiota. *Toxins* 2021;**13**. doi:10.3390/toxins13040252

50 Manor O, Zubair N, Conomos MP, *et al.* A Multi-omic Association Study of Trimethylamine N-Oxide. *Cell Rep* 2018;**24**:935–46. doi:10.1016/j.celrep.2018.06.096

51 Lohia S, Vlahou A, Zoidakis J. Microbiome in Chronic Kidney Disease (CKD): An Omics Perspective. *Toxins* 2022;**14**. doi:10.3390/toxins14030176

52 Chung S, Barnes JL, Astroth KS. Gastrointestinal Microbiota in Patients with Chronic Kidney Disease: A Systematic Review. *Adv Nutr* 2019;**10**:888. doi:10.1093/advances/nmz028

53 Perraudeau F, McMurdie P, Bullard J, *et al.* Improvements to postprandial glucose control in subjects with type 2 diabetes: a multicenter, double blind, randomized placebo-controlled trial of a novel probiotic formulation. *BMJ Open Diabetes Res Care* 2020;**8**. doi:10.1136/bmjdrc-2020-001319

54 Yarullina DR, Shafigullin MU, Sakulin KA, *et al.* Characterization of gut contractility and microbiota in patients with severe chronic constipation. *PLoS One* 2020;**15**:e0235985. doi:10.1371/journal.pone.0235985

55 Dao MC, Everard A, Aron-Wisnewsky J, *et al.* Akkermansia muciniphila and improved

749 metabolic health during a dietary intervention in obesity: relationship with gut
750 microbiome richness and ecology. *Gut* 2016;**65**:426–36. doi:10.1136/gutjnl-2014-308778

751 56 Li F, Wang M, Wang J, *et al.* Alterations to the Gut Microbiota and Their Correlation
752 With Inflammatory Factors in Chronic Kidney Disease. *Front Cell Infect Microbiol* 2019;**9**.
753 doi:10.3389/fcimb.2019.00206

754 57 Levi I, Gurevich M, Perlman G, *et al.* Potential role of indolelactate and butyrate in
755 multiple sclerosis revealed by integrated microbiome-metabolome analysis. *Cell Rep*
756 *Med* 2021;**2**:100246. doi:10.1016/j.xcrm.2021.100246

757 58 Sun C-Y, Hsu H-H, Wu M-S. p-Cresol sulfate and indoxyl sulfate induce similar cellular
758 inflammatory gene expressions in cultured proximal renal tubular cells. *Nephrol Dial*
759 *Transplant* 2012;**28**:70–8. doi:10.1093/ndt/gfs133

760 59 Yu F, Feng X, Li X, *et al.* Gut-Derived Metabolite Phenylacetylglutamine and White
761 Matter Hyperintensities in Patients With Acute Ischemic Stroke. *Front Aging Neurosci*
762 2021;**13**. doi:10.3389/fnagi.2021.675158

763 60 Miyamoto Y, Miyazaki T, Honda A, *et al.* Retention of acetylcarnitine in chronic kidney
764 disease causes insulin resistance in skeletal muscle. *J Clin Biochem Nutr* 2016;**59**:199.
765 doi:10.3164/jcbn.15-146

766 61 Calder PC. Omega-3 Fatty Acids and Inflammatory Processes. *Nutrients* 2010;**2**:355.
767 doi:10.3390/nu2030355

768 62 Eicosapentaenoic acid (EPA). Mount Sinai Health System.
769 https://www.mountsinai.org/health-library/supplement/eicosapentaenoic-acid-epa
770 (accessed 6 Feb 2023).

771 63 Liu R, Peng Y, Ye H, *et al.* Peripheral Eosinophil Count Associated with Disease Activity
772 and Clinical Outcomes in Hospitalized Patients with Lupus Nephritis. *Nephron* 2023;:1–9.
773 doi:10.1159/000528486

774 64 Eosinophilia. Mayo Clinic.
775 2021.https://www.mayoclinic.org/symptoms/eosinophilia/basics/definition/sym-
776 20050752 (accessed 3 Feb 2023).

777 65 Western Diet and the Immune System: An Inflammatory Connection. *Immunity*
778 2019;**51**:794–811. doi:10.1016/j.immuni.2019.09.020

779 66 Sharma M, Singh V, Sharma R, *et al.* Glomerular Biomechanical Stress and Lipid
780 Mediators during Cellular Changes Leading to Chronic Kidney Disease. *Biomedicines*
781 2022;**10**. doi:10.3390/biomedicines10020407

782 67 Kim OY, Lee SM, An WS. Impact of Blood or Erythrocyte Membrane Fatty Acids for
783 Disease Risk Prediction: Focusing on Cardiovascular Disease and Chronic Kidney
784 Disease. *Nutrients* 2018;**10**. doi:10.3390/nu10101454

785 68 Song J, Yu J, Prayogo GW, *et al.* Understanding kidney injury molecule 1: a novel
786 immune factor in kidney pathophysiology. *Am J Transl Res*
787 2019;**11**:1219.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6456506/ (accessed 3 Feb

788 2023).

789 69 TNFRSF11B gene. https://medlineplus.gov/genetics/gene/tnfrsf11b/ (accessed 3 Feb
790 2023).

791 70 lebioda TJ, Kmieć Z. Tumour necrosis factor superfamily members in the pathogenesis
792 of inflammatory bowel disease. *Mediators Inflamm* 2014;**2014**:325129.
793 doi:10.1155/2014/325129

794 71 Sundbøll J, Thygesen SK, Veres K, *et al.* Risk of cancer in patients with constipation. *Clin*
795 *Epidemiol* 2019;**11**:299. doi:10.2147/CLEP.S205957

796 72 Peng Y, Liu F, Qiao Y, *et al.* Association of abnormal bowel health with major chronic
797 diseases and risk of mortality. *Ann Epidemiol* 2022;**75**:39–46.
798 doi:10.1016/j.annepidem.2022.09.002

799 73 Eating right for chronic kidney disease. National Institute of Diabetes and Digestive and
800 Kidney Diseases. 2022.https://www.niddk.nih.gov/health-information/kidney-
801 disease/chronic-kidney-disease-ckd/eating-nutrition (accessed 24 Feb 2023).

802 74 Blanc F, Bouteloup V, Paquet C, *et al.* Prodromal characteristics of dementia with Lewy
803 bodies: baseline results of the MEMENTO memory clinics nationwide cohort. *Alzheimers*
804 *Res Ther* 2022;**14**:96. doi:10.1186/s13195-022-01037-0

805 75 Nedelec T, Couvy-Duchesne B, Monnet F, *et al.* Identifying health conditions associated
806 with Alzheimer's disease up to 15 years before diagnosis: an agnostic study of French
807 and British health records. *Lancet Digit Health* 2022;**4**:e169–78. doi:10.1016/S2589-
808 7500(21)00275-2

809 76 Frequent Bowel Movements. Cleveland Clinic.
810 https://my.clevelandclinic.org/health/diseases/17791-frequent-bowel-movements
811 (accessed 3 Feb 2023).

812 77 McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis
813 and graphics of microbiome census data. *PLoS One* 2013;**8**:e61217.
814 doi:10.1371/journal.pone.0061217

815 78 Zubair N, Conomos MP, Hood L, *et al.* Genetic Predisposition Impacts Clinical Changes
816 in a Lifestyle Coaching Program. *Sci Rep* 2019;**9**. doi:10.1038/s41598-019-43058-0

817 79 Ritchie ME, Phipson B, Wu D, *et al.* limma powers differential expression analyses for
818 RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47–e47.
819 doi:10.1093/nar/gkv007

820 80 Martin BD, Witten D, Willis AD. MODELING MICROBIAL ABUNDANCES AND
821 DYSBIOSIS WITH BETA-BINOMIAL REGRESSION. *Ann Appl Stat* 2020;**14**:94–115.
822 doi:10.1214/19-aoas1283

823 81 Korem T, Zeevi D, Suez J, *et al.* Growth dynamics of gut microbiota in health and disease
824 inferred from single metagenomic samples. *Science* 2015;**349**:1101–6.
825 doi:10.1126/science.aac4812

82 Joseph TA, Chlenski P, Litman A, *et al.* Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates. *Genome Res* 2022;**32**:558–68. doi:10.1101/gr.275533.121

Figures and Figure Captions

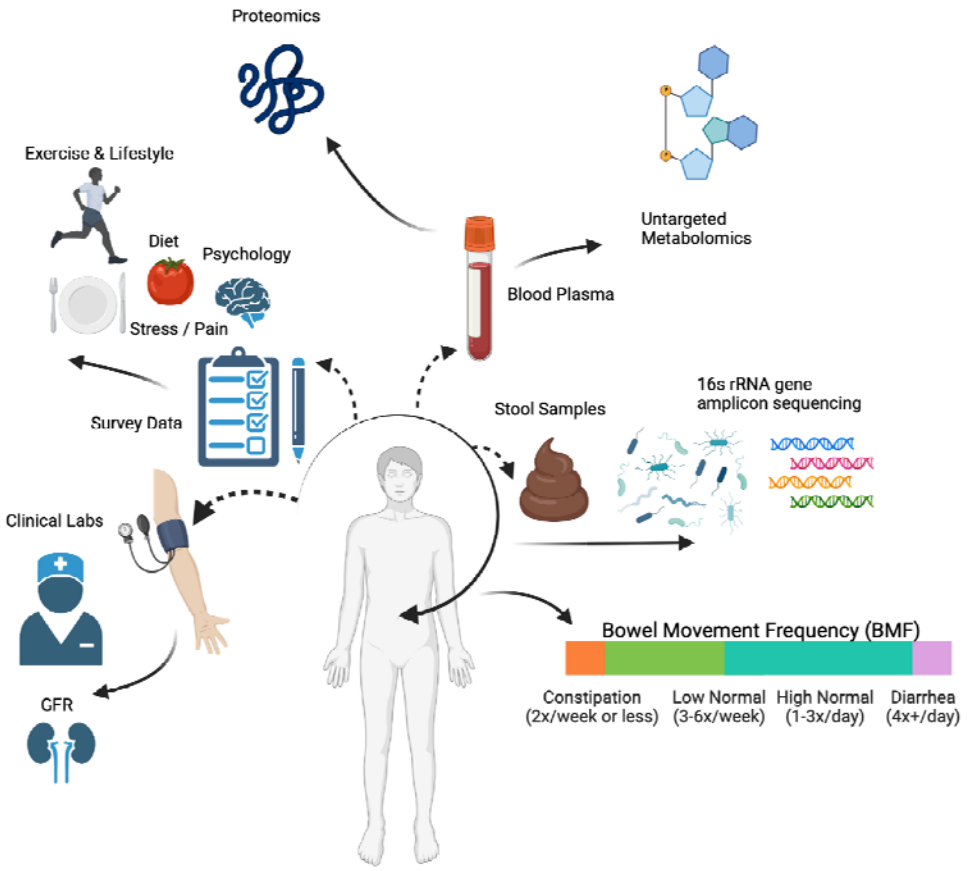

Figure 1. Data collection strategy. Arivale participants had their multi-omics, survey, and clinical data collected through various methods: interviewing, blood plasma collection, and stool samples. Interview data consisted of several questions with categorical responses, either ordinal or binary (True/False) answers (which were excluded in this analysis), which were then used in ordinal POLR to determine likelihoods of different response categories across BMF and its covariates. Clinical labs, untargeted metabolomics, and proteomics data were obtained from collected blood plasma samples (the earliest sample available per participant in the cohort). Gut microbiome ASV data were collected from stool samples provided using an at-home kit. BMF data were determined as categorical ranges of reported bowel movements per week or day depending on the response to lifestyle questionnaire data from the interviews.

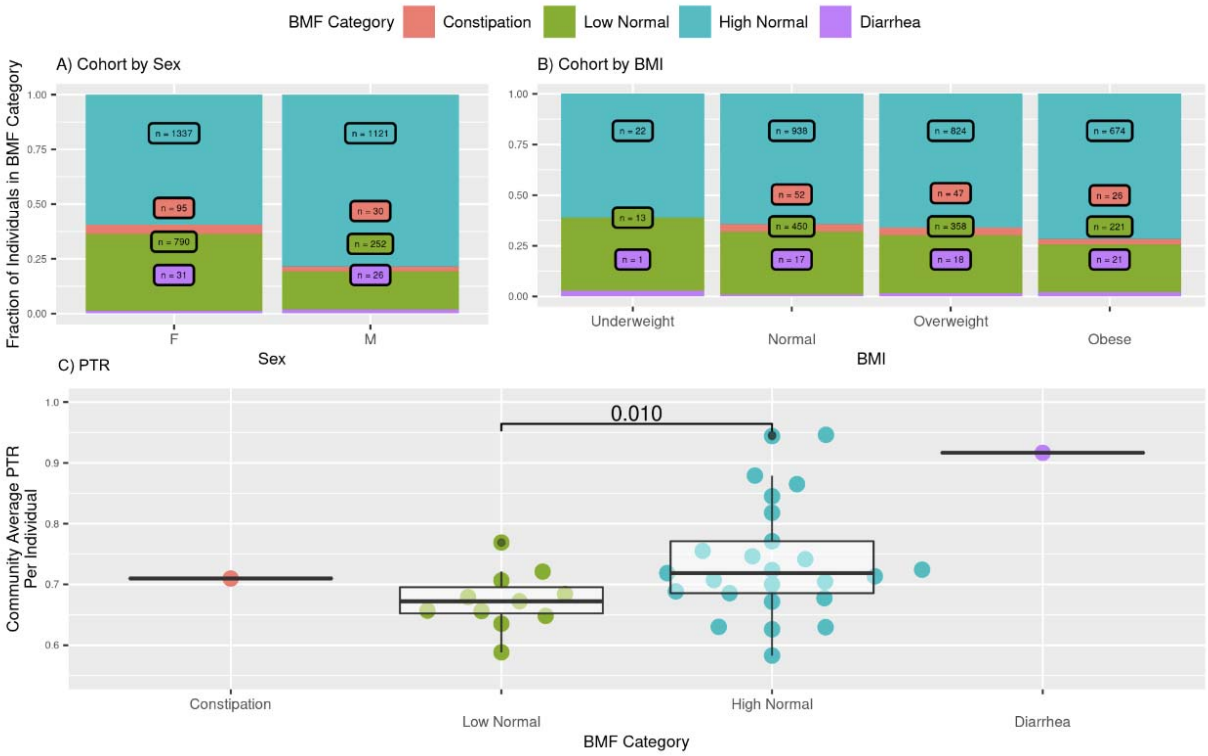

Figure 2. Distribution of BMF categories across sex and BMI groups and the relationship between average community growth rate and BMF. (A-B): Significant unevenness in the distribution of BMF across sex and BMI are highlighted here. POLR was used to regress BMF against the covariates sex, age, BMI, and eGFR. The result was that only sex (P = 1.23E-23) and BMI (P = 5.09E-6) were significantly associated with variations in BMF. (C): Community Average PTR Per Individual (the mean growth rate across all growth rates of all taxa for a given individual). There is a significant difference (linear regression, P value = 1.56E-2; post-hoc t-test P value = 1.0E-2) between the higher and lower BMF "normal" categories, showing a general directional trend of increasing community average PTR with rising BMF level, indicative of higher BMF representing a higher "flow rate" of material through the gut which is associated with higher growth community growth rates.

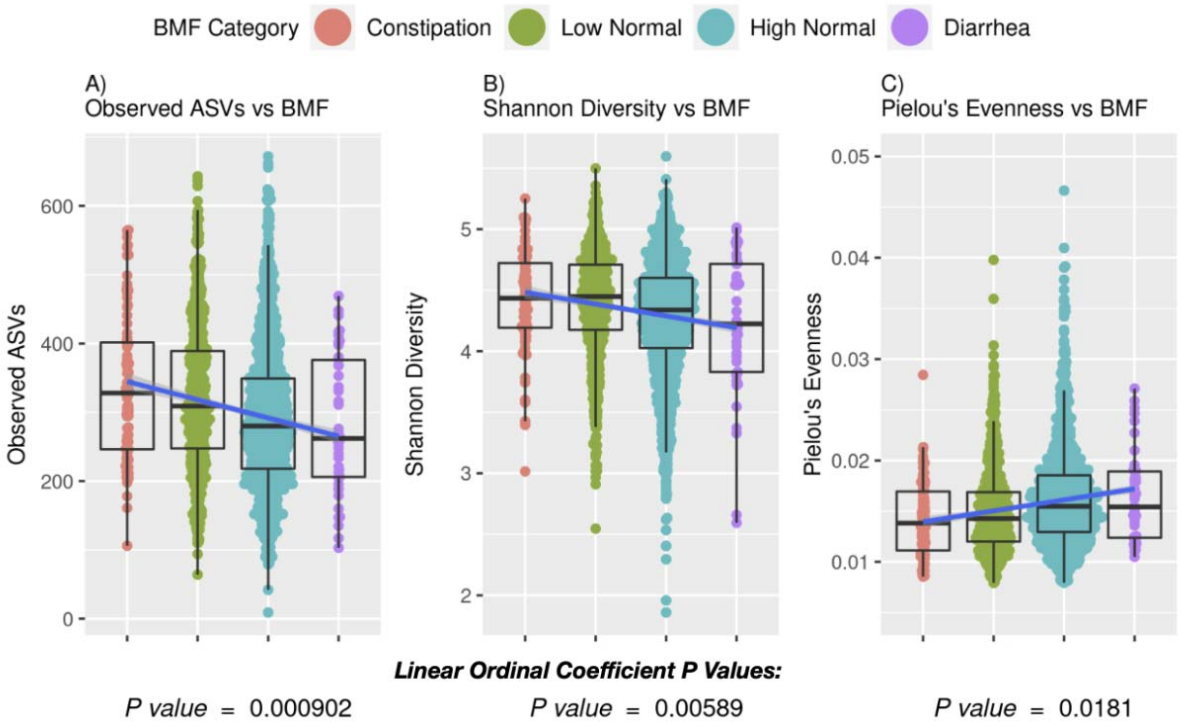

Figure 3. Associations between gut microbiome alpha-diversity measures and BMF. (A) The numbers of taxonomic groups per BMF category, representing the richness of the BMF cohort (ordinal BMF variable - ANOVA P value: 9.02E-4). (B) The number of different taxonomic groups (variety) per BMF category, representing the alpha diversity of the BMF cohort (ordinal BMF variable - ANOVA P value: 5.89E-3). (C) The distribution of abundances of the taxonomic groups, determined proportionally by dividing the diversity by the richness of the cohort (evenness, ordinal BMF variable - ANOVA P value: 1.81E-2). The evenness decreases with BMF, suggesting slow colonic transit times (constipation) correspond to having a higher ratio of richness to diversity and lower evenness.

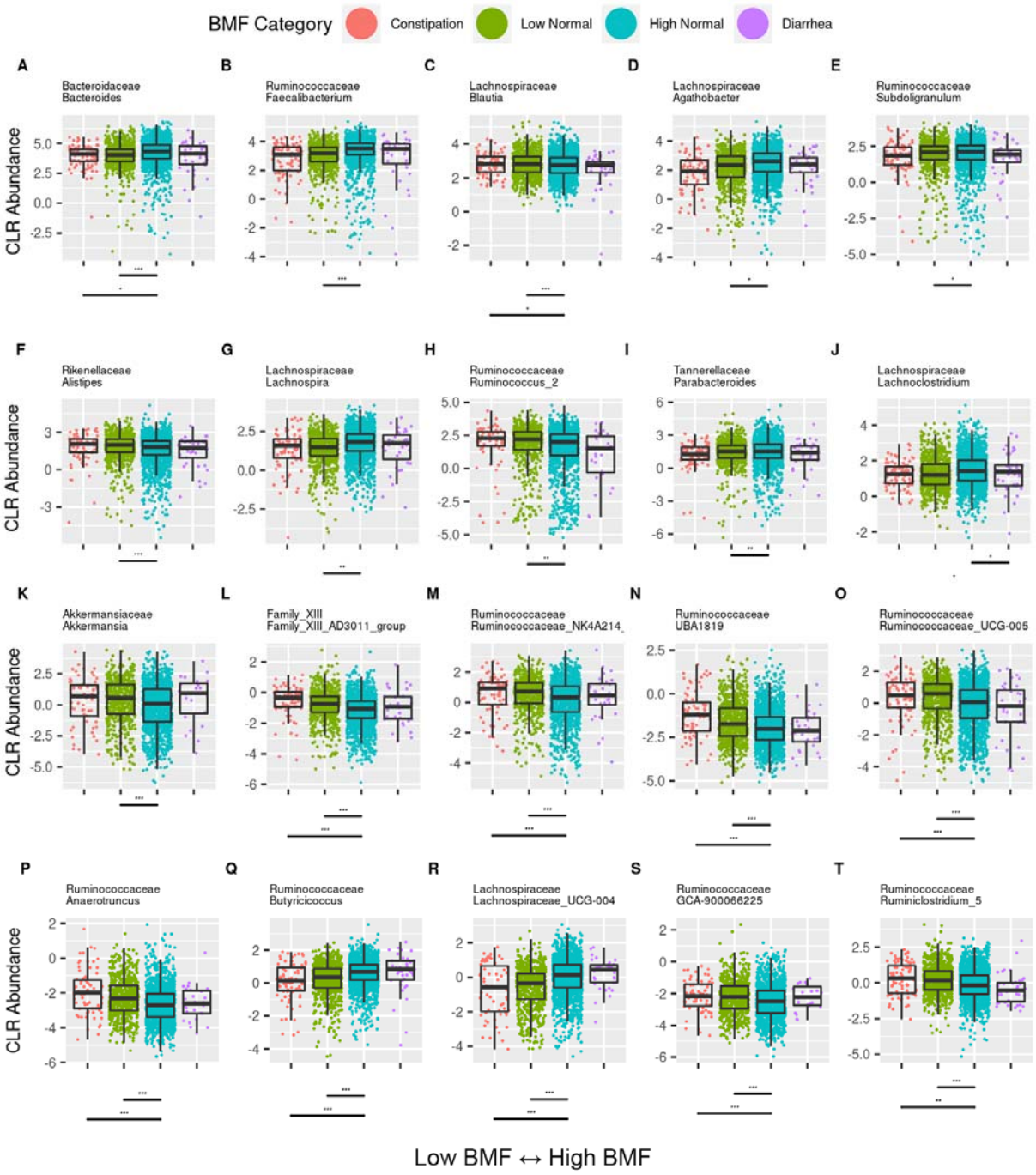

Figure 4. Significant BMF associations for the 10 most abundant genera, *Akkermansia*, and another 9 genera with the lowest remaining P values. The top 10 most abundant significant taxa from the fecal samples and ASV CORNCOB analysis (A-J), *Akkermansia* (K), and the top 9 most significant taxa not already included in the most abundant list (L-T). Lines beneath each plot denote significant differences from the reference category, and asterisks denote FDR-corrected significance threshold. (***): $P < 0.0001$, (**): $0.0001 < P < 0.01$, (*): $0.01 < P < 0.05$.

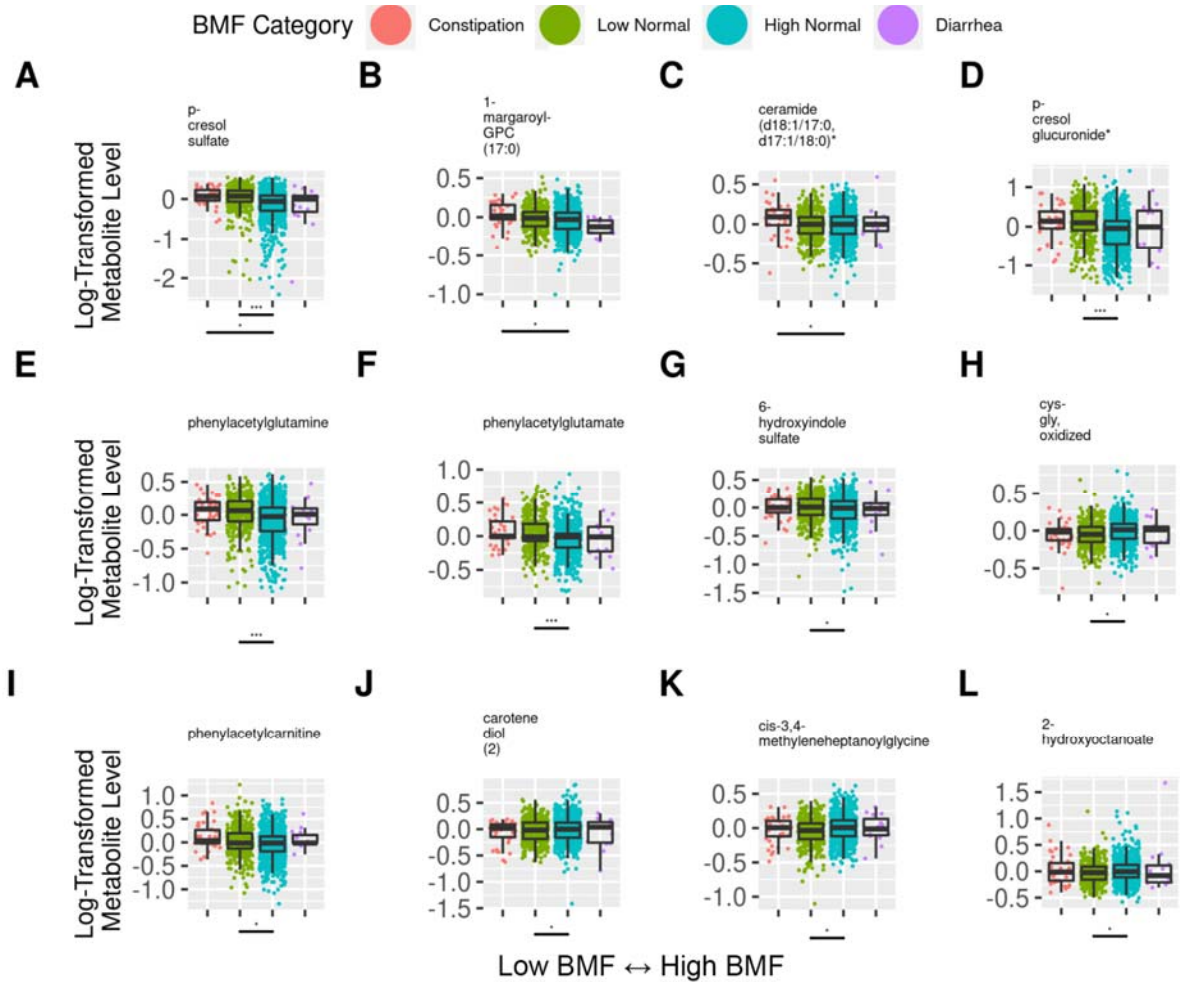

Figure 5. Top 12 BMF-associated blood plasma metabolites with annotations. (A-L) The 12 significant blood plasma metabolites from the LIMMA metabolomics analysis with available annotations. Lines beneath each plot denote significant differences from the reference category, and asterisks denote FDR-corrected significance threshold. (***): $P < 0.0001$, (**): $0.0001 < P < 0.01$, (*): $0.01 < P < 0.05$.

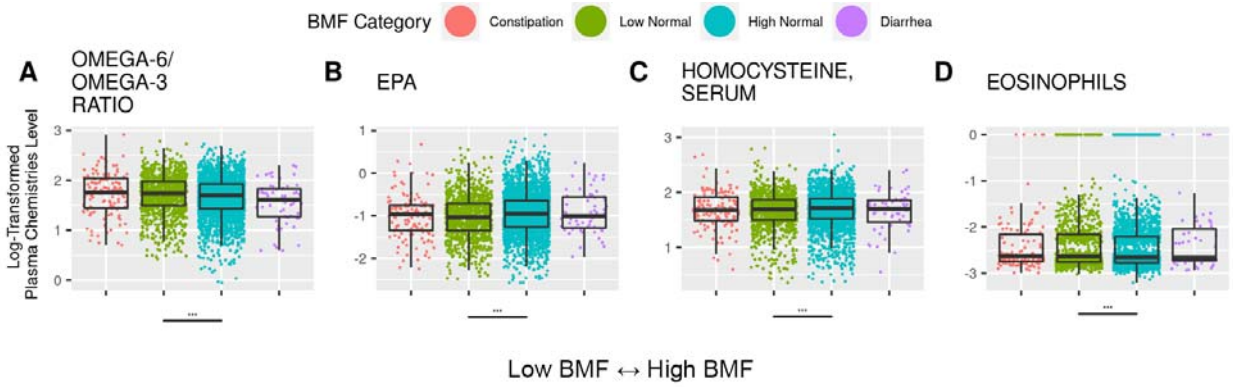

Figure 6. Blood plasma chemistries significantly associated with BMF. The 4 blood plasma chemistries features that showed significant associations with BMF. Lines beneath each plot denote significant differences from the reference category, and asterisks denote FDR-corrected significance threshold. (***): P < 0.0001, (**): 0.0001 < P < 0.01, (*): 0.01 < P < 0.05.

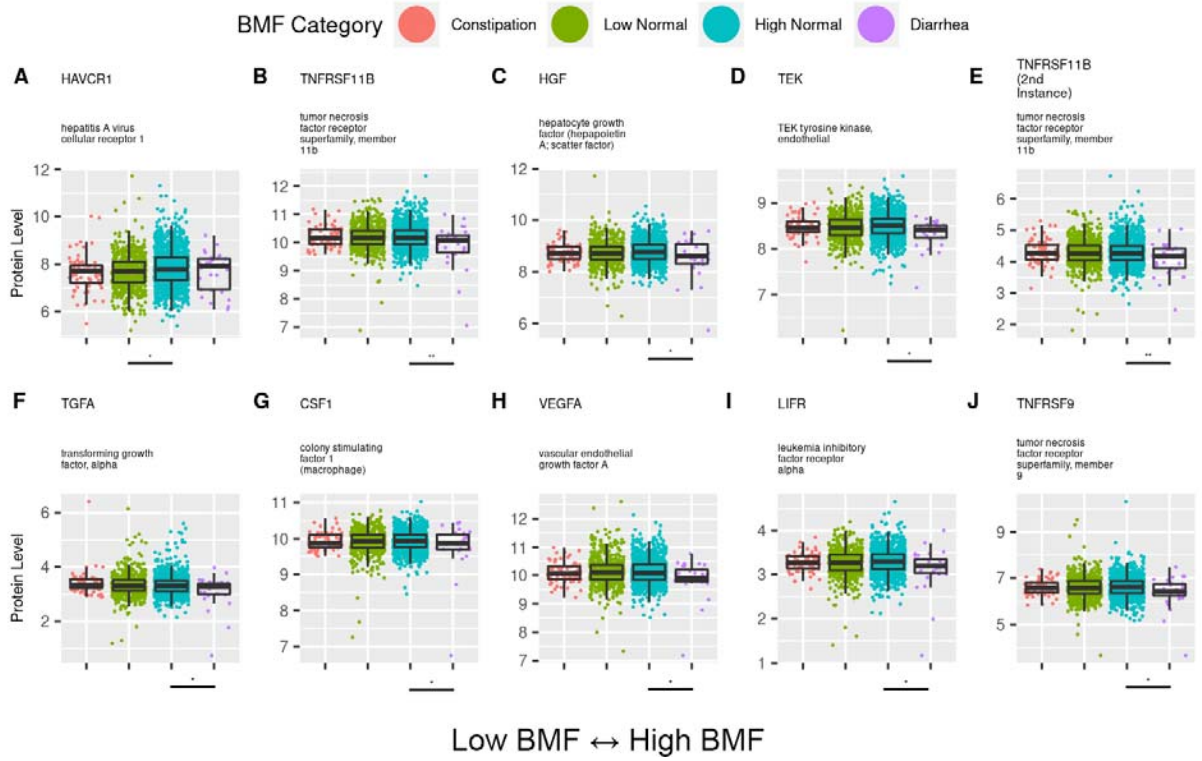

Figure 7. Selected blood plasma proteins significantly associated with BMF. Top 10 most significant blood plasma protein results with associated genes and annotated descriptions from the LIMMA proteomics analysis (A-J). Lines beneath each plot denote significant differences from the reference category, and asterisks denote FDR-corrected significance threshold. (***): P < 0.0001, (**): 0.0001 < P < 0.01, (*): 0.01 < P < 0.05.

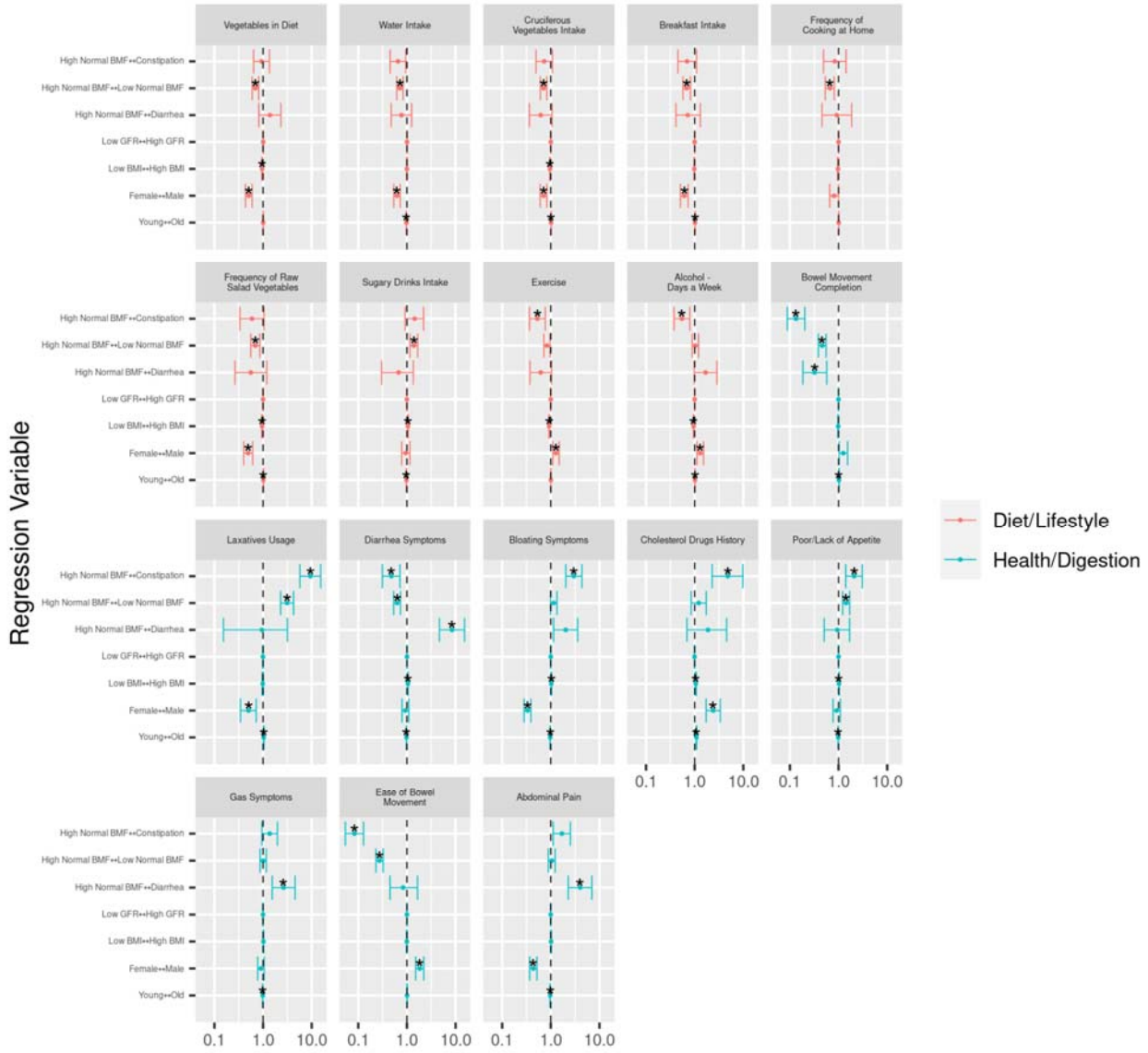

Figure 8. Ordinal regression odds ratio for health, diet, and lifestyle survey data vs BMF and covariates. Response variables are colored by category: questions related to diet, exercise, and lifestyle (Diet/Lifestyle), questions related to current digestive symptoms/function, health/medication history, and appetite (Health/Digestion), and questions related to the Big 5 Personality Test, mood/behavior or pain (Psychological). The BMF reference category was "high-normal" BMF (7-21 bowel movements per week). Each tick on the vertical axes represents a directional association in likelihood across the horizontal axis. The center line over the plots at x = 1.0 represents an equal likelihood of reporting an increase in number, intensity, frequency, or agreement (depending on the response variable) between the left side of the arrow on the vertical axis tick and the right side of the arrow on the vertical axis tick. A confidence interval that does not span the center line is significantly associated with the independent variable on the vertical axis tick. (*): FDR-corrected P-value < 0.05.