An association between the gut microbiota and immune cell dynamics in humans.

Jonas Schluter[1], Jonathan U. Peled[2,3], Bradford P. Taylor[1], Melody Smith[2,3], Kate A. Markey[2,3], Ying Taur[4], Rene Niehus[5], Anna Staffas[6], Anqi Dai[2], Emily Fontana[4], Luigi A. Amoretti[4], Roberta J. Wright[4], Sejal Morjaria[4], Maly Fenelus[7], Melissa S. Pessin[7], Nelson J. Chao[8], Meagan Lew[8], Lauren Bohannon[8], Amy Bush[8], Anthony D. Sung[8], Tobias M. Hohl[4], Miguel-Angel Perales[2,3], Marcel R.M. van den Brink[2,3], Joao B. Xavier[1]

1) Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

2) Adult Bone Marrow Transplantation Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

3) Weill Cornell Medical College, New York, NY, USA

4) Infectious Disease Service, Department of Medicine, and Immunology Program, Sloan Kettering Institute, New York, NY, USA.

5) Harvard University, T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115

6) Sahlgrenska Cancer Center, Department of Microbiology and Immunology, Institute of Biomedicine, University of Gothenburg, SE-405 30, Gothenburg, Sweden

7) Department of Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

8) Division of Hematologic Malignancies and Cellular Therapy, Duke University School of Medicine

Correspondence: jonas.schluter+mskcc@gmail.com, xavierj@mskcc.org

ABSTRACT

The gut microbiota influences development and homeostasis of the mammalian immune system[1–3], can alter immune cell compositions in mice[4–7], and is associated with responses to immunotherapy that rely on the activity of peripheral immune cells[8–12]. Still, our understanding of how the microbiota modulates immune cells dynamics remains limited, particularly in humans where a lack of deliberate manipulations makes inference challenging. Here we study hundreds of hospitalized—and closely monitored—patients receiving hematopoietic cell transplantation as they recover from chemotherapy and stem cell engraftment. This aggressive treatment causes large shifts in both circulatory immune cell and microbiota populations, allowing the relationships between the two to be studied simultaneously. We analyzed daily changes in white blood cells from 2,235 patients, and 10,680 longitudinal microbiota samples to identify bacteria associated with those changes. Bayesian inference and validation across patient cohorts revealed consistent associations between gut bacteria and white blood cell dynamics in the context of immunomodulatory medications, clinical metadata and homeostatic feedbacks. We contrasted the potency of fermentatively active, obligate anaerobic bacteria with that of medications with known immunomodulatory mechanism to estimate the potential of the microbiota to influence peripheral immune cell dynamics. Our analysis establishes and quantifies the link between the gut microbiota and the human immune system, with implications for microbiota-driven modulation of immunity.

39 **MAIN TEXT**

40 Experiments in mice provide evidence that the mammalian intestinal microbiome influences the
41 development[1–3] and homeostasis of its host's immune system[4–7,13–15]. In humans, inflammatory bowel
42 diseases correlate with functional dysbiosis in the gut microbiota[16,17]. Children born preterm and at
43 term have different gut microbiome compositions and differ in the development of immune cell
44 populations in their blood[18]. The composition of the gut microbiota may also influence the success of
45 immunotherapies[8–11]. Immune checkpoint inhibitor therapy relies on activation of circulating T-cells
46 and its success has, independently, been associated with abundances of intestinal anaerobic genera
47 such as *Akkermansia*[9] and *Faecalibacterium*[10]. It is therefore an intriguing prospect to augment
48 treatments such as cancer immunotherapy[19], including the burgeoning field of chimeric antigen
49 receptor (CAR) T-cell therapy[20], by leveraging microbiome-driven immune system modulation.

50 Our understanding of how the microbiota influences the dynamics of immune cells in
51 humans, and how this compares to deliberate immunomodulatory interventions nevertheless remains
52 limited. Experiments with animals may not always be sufficient to study mechanisms of
53 microbiome-immune interactions and translate them to human biology as the microbial ecology in the
54 gut of an animal model may be different from humans receiving treatment[21]. On the other hand,
55 studies directly in patients may be criticized when they have small subject numbers, are cross-
56 sectional, lack statistical power, or disregard key confounders such as medications[21].

57 To overcome these limitations, we conducted a large-scale longitudinal study of the gut
58 microbiota and day-by-day changes in circulatory immune cell counts. We investigated immune
59 reconstitution dynamics after allogeneic hematopoietic cell therapy (HCT) within all 2,926 patients
60 who underwent HCT at Memorial Sloan Kettering for various hematological malignancies, including
61 leukemia, between 2003 and 2019 (**Figure 1A, Table S1**). The conditioning regimen of radiation and
62 chemotherapy administered to HCT patients is the most severe perturbation to the immune system
63 deliberately performed in humans and thus offers a unique opportunity to investigate dynamic links
64 between the gut microbiota and the immune system directly in humans. Conditioning depletes white
65 blood cell counts (**Figure 1A**) and can lead to prolonged periods of neutropenia (<500 neutrophils per
66 μl blood). Immune reconstitution begins after transplanted stem cells have matured sufficiently to

release granulocytes from the bone marrow (neutrophil engraftment is defined as 3 consecutive days with >500 neutrophils per µl blood, **Figure 1A-C**). The blood of each patient is carefully monitored

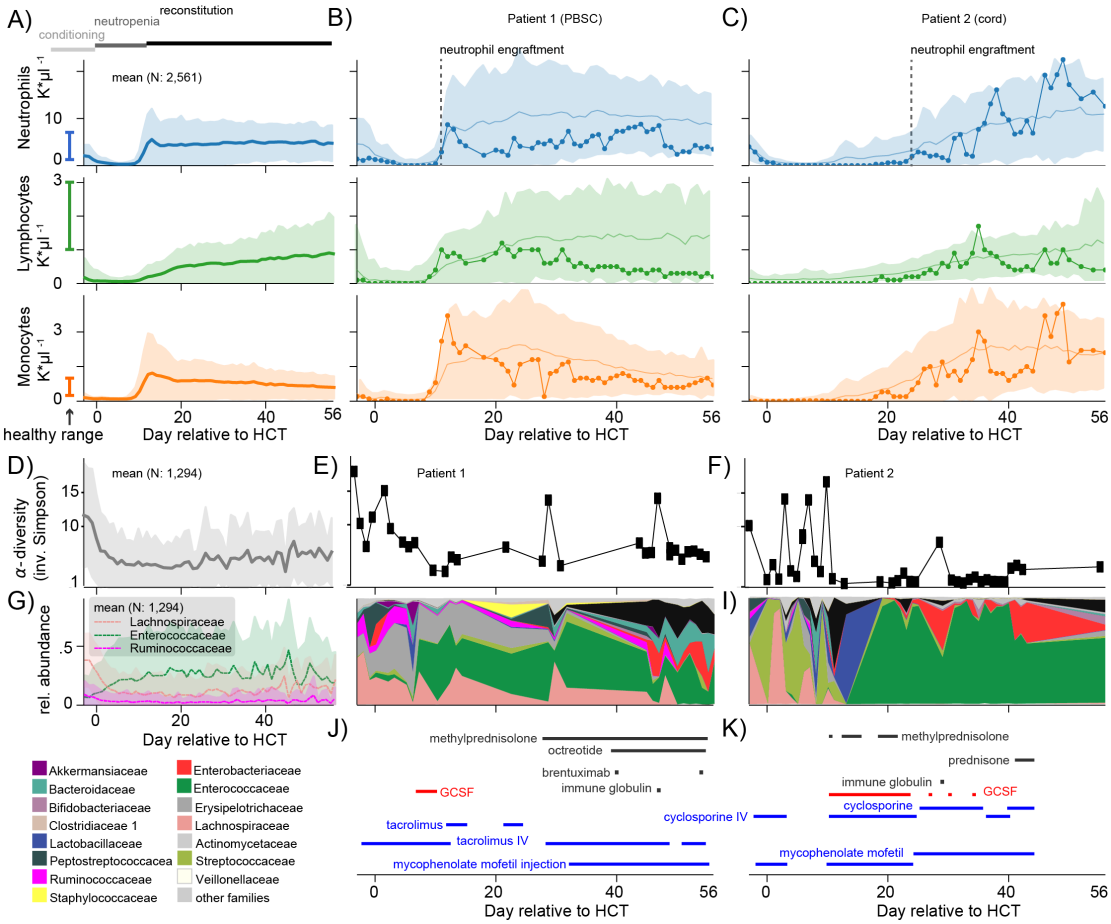


Figure 1: Immune reconstitution and microbiome dynamics after allogeneic hematopoietic cell transplantation (HCT). A) Three major periods of HCT—immunoablation during chemotherapeutic conditioning before HCT, defined as day 0, post-HCT neutropenia, and reconstitution following neutrophil engraftment—lead to recovery trajectories with large variability between patients. Shown are the mean counts (shaded: ±1 standard deviation, σ) of neutrophils, lymphocytes and monocytes per day relative to HCT from patients transplanted between 2003 and 2019 (A), contrasted with two individual patients (B,C) representative of the recovery trajectories for different stem cell graft source; patient 1 who received a peripheral blood stem cell graft, PBSC (line with circles: patient data, solid line and shaded region: mean±1 standard deviation of all PBSC patients), and patient 2 who received a graft of umbilical cord blood (line with circles: patient data, solid line and shaded region: mean±1 standard deviation of all cord patients). Fecal samples collected and analyzed by 16S rRNA gene sequencing reveal the loss of microbial diversity reported previously[23,24] (D, line: mean per day, shaded: ±1 standard deviation); E,F: individual patient measurements) and commensal families (G, line: mean relative abundances of bacterial families, shaded: ±1 standard deviation); H,I: individual patient measurements), often replaced by Enterococcaceae domination. J,K). Administration of immunomodulatory medications for the two example patients.

throughout this recovery, and medications are administered to modulate the immune cell dynamics, including granulocyte-colony stimulating factor (GCSF) to increase neutrophil counts, and immunosuppressants such as mycophenolate mofetil or tacrolimus to prevent complications such as graft-vs-host disease (**Figure 1 J,K**). To investigate if the composition of the gut microbiota is associated with the dynamics of circulating white blood cells, we analyzed detailed blood and clinical

74 metadata of our patients between 3 days before HCT and until 100 days post neutrophil engraftment
75 (excluding pediatric patients, and other exclusion criteria: N=2,235, supplementary methods, **Figure**
76 **S1**). During this period patients are monitored carefully, and our analysis included over 140,000 host
77 phenotype measurements in the form of complete blood counts which quantify the most abundant
78 white blood cells—neutrophils, lymphocytes, monocytes, eosinophils—as well as platelet counts
79 (**Figure 1, S1**). We started collecting patients' fecal microbiota data in 2009[22], and by now obtained
80 10,680 high frequency, longitudinal microbiota compositions.

81 HCT patients lose gut microbiota biodiversity and commensal microbial families during their
82 treatment (**Figure 1D-I**); this figure generated from N=1,294 HCT patients clarifies the preliminary
83 trends observed in previous studies from smaller datasets[23,24]. We have shown recently that mice with
84 a depleted intestinal flora had worse recoveries of white blood cells after bone-marrow
85 transplantation[25]. In our patients, microbial diversity usually recovers slowly during white blood cell
86 reconstitution (**Figure 1D**); however, microbiota recovery as well as immune reconstitution can vary
87 strongly between patients and treatment types (**Figure 1B,C,J,K, S1**). This variation is illustrated by
88 the distinct trajectories of patient 1 who received a graft of peripheral blood stem cells (PBSC),
89 retained high microbiota diversity and engrafted earlier (**Figure 1B,E,H**), and patient 2 who received
90 a graft of umbilical cord blood (cord), lost microbiota diversity and engrafted later (**Figure 1C,F,I**).
91 Low microbiota diversity at the time of neutrophil engraftment has been associated with 5-fold
92 increased transplant-related mortality[26], suggesting that that the joint recovery of the microbiota and
93 white blood cells in circulation is critical for clinical outcomes.

94 To detect a directional and causal link between the microbiota and white blood cells, we first
95 used data from a recent prospective randomized trial of autologous fecal microbiota transplantation
96 (auto-FMT), which is a microbiota manipulation experiment done directly in our patients[23]
97 (supplementary methods). Twenty-four patients (**Figure 2A, Table S2)** underwent randomization,
98 resulting in 10 untreated control and 14 treated patients, including patient 3 in **Figure S2**. To
99 investigate if auto-FMT affected white blood cell reconstitution, we compared the 24 patients'
100 neutrophil, lymphocyte, monocyte (**Figure 2**) and total white blood cell counts (**Figure S3**) post-
101 engraftment (i.e. when the transplanted hematopoietic cells begin producing new white blood cells,

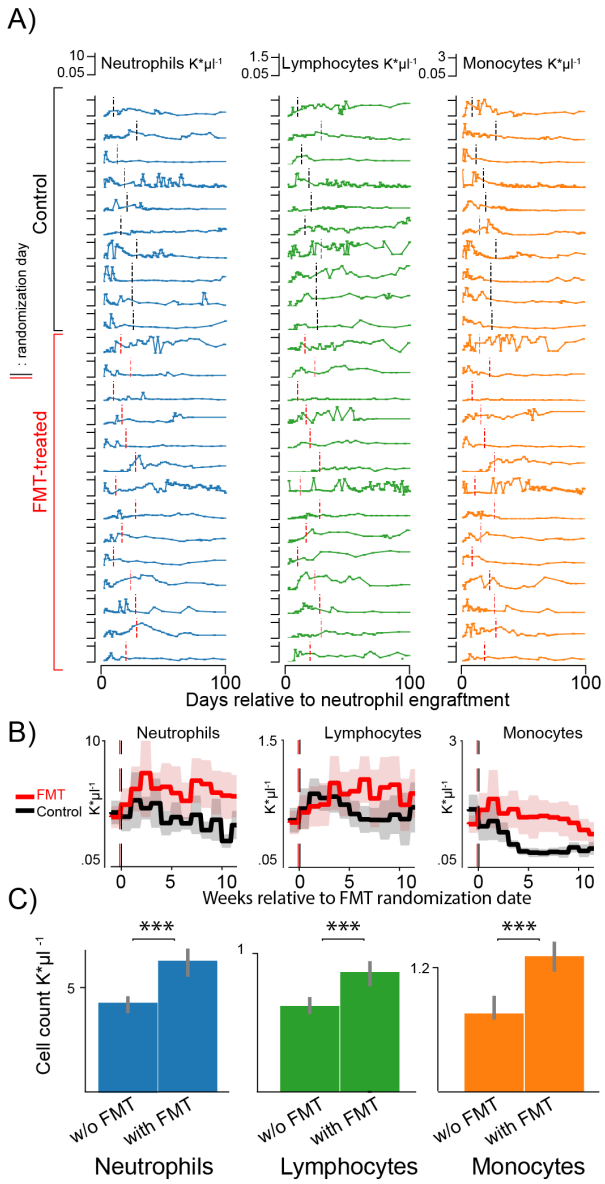
102 **Figure 2A,S3**). FMT procedures were conducted at variable time points relative to neutrophil

103 engraftment, but overall, we observed higher counts of each white blood cell type in patients who

104 received an auto-FMT during the first 100 days post neutrophil engraftment (p<0.001, **Figure 2B,S3-**

105 **S6**).

106



107 **Figure 2: Neutrophil, lymphocyte and monocyte counts increased in FMT-treated patients**
108 **during the weeks following the treatment compared to control patients.** A) Absolute counts of
109 neutrophils (blue), lymphocytes (green) and monocytes (orange) in 24 patients enrolled in a randomized
110 controlled trial to receive an autologous fecal microbiota transplant post-neutrophil engraftment (10 control:
111 black vertical line, 14 FMT treated: red vertical line,). B) Weekly mean cell counts aligned to the date of
112 randomization into FMT treatment arm (red) or control (black). Line: mean per week, shaded region: 95%-CI.
113 C) Results from a linear mixed effects model with random effects per patient and per day relative to neutrophil
114 engraftment confirms that neutrophil, lymphocyte and monocyte counts are higher in patients receiving auto-
115 FMT after their treatment as compared to control patients after the randomization date; bars and confidence
116 intervals for the averages of observed white blood cell counts without FMT and post-FMT (***: p<0.001).
117

118 The increased white blood cells in patients receiving auto-FMT could be due to the
119 reconstitution of a complex microbiota that we saw in these patients[23] and the associated metabolic
120 capabilities[6,7,25], or it could be a systemic response to a severe therapy which introduced billions of
121 intestinal organisms at once via an enema (no enema was administered to control patients[23]).
122 Moreover, while the mixed-effects model accounted for patient-specific HCT treatments in this
123 randomized patient cohort, chance differences in extrinsic factors such as different
124 immunomodulatory drug exposures may have affected this result due to the small cohort size.
125 Nonetheless, observing that auto-FMT recipients had increased white blood cell counts supports the
126 notion that the microbiota can modulate the peripheral immune system. High counts of lymphocytes
127 during immune reconstitution has been associated with improved clinical outcomes[27]. Additionally, in
128 our HCT patients, a higher average level of white blood cells measured across a period of 100 days
129 after neutrophil engraftment (supplementary methods) confirms a positive association with 3-year
130 survival (hazard ratio: 0.91, p:0.04). Determining which taxa modulate immune dynamics could open
131 new ways to improve robust immune reconstitution, which is critical for clinical outcomes[27–29].
132 To address this question, we next investigated the link between the gut microbiota and the
133 dynamics of white blood cell recovery in our large observational cohort of HCT patients. Homeostasis
134 of circulatory white blood cell counts is a complex, dynamic process: neutrophils, lymphocytes and
135 monocytes are formed and released into the blood *de novo* by differentiation of hematopoietic
136 progenitor cells from the bone marrow, and they can be mobilized from thymus and lymph nodes
137 (lymphocytes), spleen, liver and lungs (neutrophils); they can also migrate from the blood to other
138 tissues when needed[30]. These processes are dynamic sources and sinks of circulatory white blood
139 cells, and they can be modulated by drugs administered to patients receiving HCT. To identify factors
140 associated with these dynamic source- and sink-processes—including the microbiota—we developed
141 a two-stage approach analyzing the changes of white blood cell counts between two days (i.e. the
142 rates of cell count increases and decreases). Stage 1 served as a feature selection stage where we used
143 data of 1,096 patients (after filtering for qualifying samples and applying exclusion criteria, see
144 supplementary methods) *without* available microbiome information to identify associations between
145 clinical metadata, including immunomodulatory medications (methods), and changes of white blood

146 cell counts from one day to the next (**Figure 3A**). Stage 2 was performed on data from an independent
147 cohort of 841 different patients at MSK from whom concurrent microbiome samples were available.
148 Stage 2, our main analysis, sought to reveal associations between the abundance of microbial taxa and
149 the daily changes in blood cell counts in context of immunomodulatory medications, additional
150 clinical metadata and the current state of the blood itself.

151 In stage 1 we calculated the changes in neutrophil, lymphocyte and monocyte counts during
152 patients' recovery from >20,000 pairs of post-engraftment blood samples separated by a single day
153 (**Figure 3B, S7-9,** supplementary methods). Using a cross-validated feature selection approach, we
154 detected medications and HCT treatment parameters that were associated with different rates of
155 change in neutrophil, lymphocyte and monocyte counts, including, as expected, GCSF and the graft
156 stem cell sources (**Figure S7-S9, Table S1**).

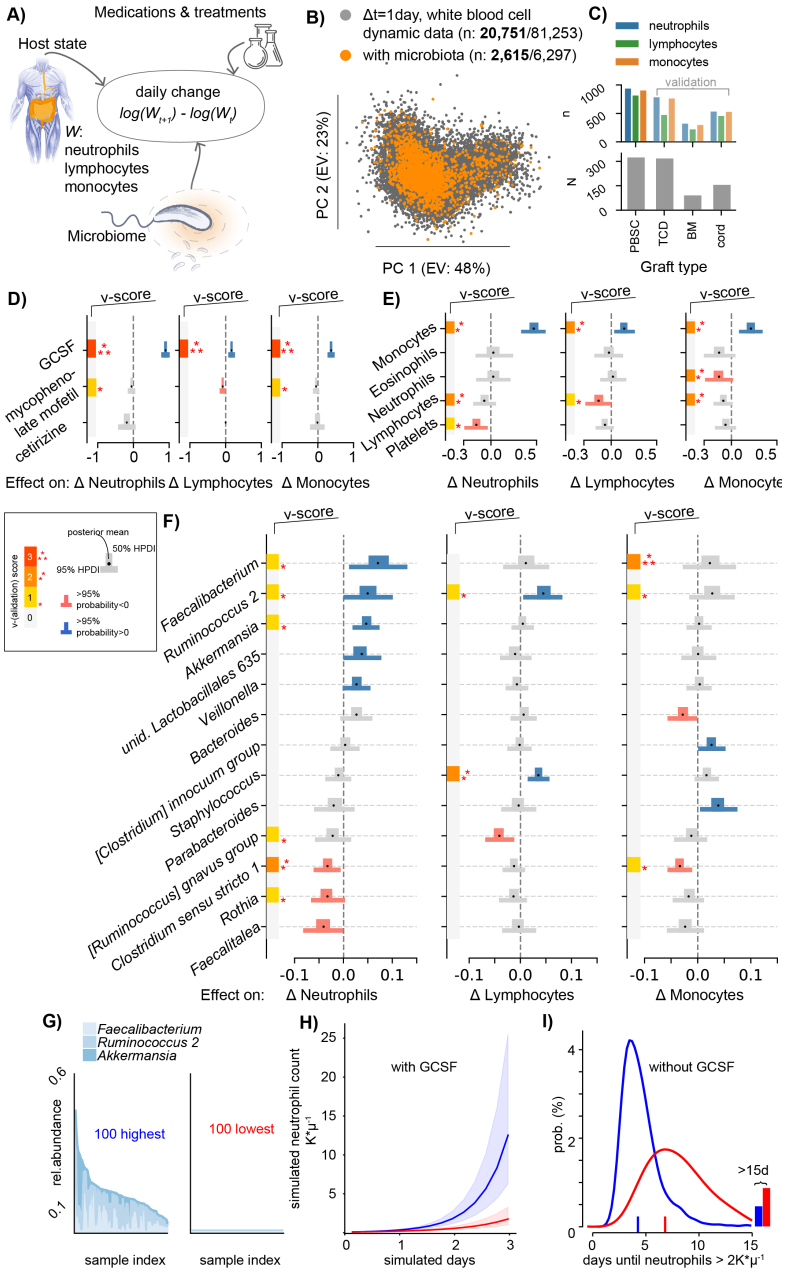
157 During stage 2 we sought to identify associations between bacterial taxa of the gut microbiota
158 and the dynamics of immune cells in circulation. For this, we performed Bayesian inferences using
159 data from different sets of patients *with* available microbiome samples. Stage 1 had identified—as
160 expected—that stem cell graft sources are associated with immune reconstitution kinetics (e.g. cord
161 on average slower compared with PBSC[31]), and we therefore stratified our patients by graft source in
162 stage 2. The model of stage 2 now included microbial genera as predictors of observed changes in
163 white blood cell counts, in addition to the medications selected in stage 1, clinical features
164 (conditioning intensity, age, sex), and the current state of the blood in the form of counts of
165 neutrophils, lymphocytes, monocytes, eosinophils, and platelets. The data comprised 841 patients, but
166 approximately 60% of the stool samples paired with a daily change in white blood cell counts were
167 taken before neutrophil engraftment (**Figure 3B, Table S1,** supplementary methods), i.e. when blood
168 cells counts were zero. In total, we analyzed 2,615 post-engraftment observations of changes in
169 neutrophil counts during immune reconstitution (lymphocytes: 2,006, monocytes: 2,534) with paired
170 stool samples which provided a large sample of observed white blood cell dynamics (**Figure 3B,**
171 **Table S3,S4**). We first focused on the data from the largest (**Figure 3C**) cohort—patients who
172 received a PBSC graft—and withheld the other cohorts (bone marrow, BM; T-cell depleted graft
173 (*ex-vivo*) by CD34+selection, TCD; and cord) to use as independent validation cohorts. For this

174 validation, we analyzed TCD, BM, and cord patients' data in the same way as PBSC patients' data
175 and compared the resulting posterior coefficient distributions (methods). We assigned coefficients
176 obtained from the PBSC cohort a validation score (v-score) between 0 and 3, representing the number
177 of times that the focal coefficient was validated in the other cohorts; but, conservatively, the score was
178 always set to zero if we observed counter-evidence among any of the other data sets, i.e. evidence that
179 coefficients had the opposite sign, ensuring only the most consistent associations were considered as
180 validated. Finally, we analyzed data from another patient cohort consisting of 493 bone marrow
181 transplantation recipients treated at Duke University including 9,603 blood samples and a total of 629
182 microbiota samples from 218 patients, albeit with lower sampling density, and we used the results
183 from this analysis for further validation.

184 Notably, as a verification of our approach, we detected associations between the
185 administration of immunomodulators and increased or decreased rates of immune cell count changes
186 consistent with the known biological mechanism of these medications (**Figure 3C,S10-S13**). The
187 strongest across all predictors is the well-known neutrophil-increasing effect of GCSF[32]; GCSF
188 administration—used to accelerate recovery from chemotherapy-induced neutropenia[32]—was
189 associated with a +140% increase in the rate of neutrophil changes from one day to the next ([+114%,
190 +170%], 95 percent probability density interval [HPDI95]). This finding was observed in all MSK
191 validation data sets (v-score=3, **Figure 3D**), as well as among Duke University patients (**Figure**
192 **S14,S15**). We furthermore found a GCSF-associated increase of +43% ([+30%, +58%]HPDI95, v-
193 score=3) in monocyte rates, and, although smaller, in lymphocyte rates (+16%, [+5%, +27%]HPDI95,
194 v-score=3). Both neutrophil and lymphocyte rates decreased following the exposure to antihistamine
195 or immunosuppressive medications (cetirizine -18%, [-35%, +5%]HPDI95, mycophenolate
196 mofetil -8% [-15%,+1%]HPDI95, respectively). Finally, less intensive chemotherapeutic conditioning
197 regimens (non-ablative conditioning and reduced intensity) were associated with larger lymphocyte
198 and monocyte count growth rates during immune reconstitution (**Figure S10C**)

199 Beyond associating medications in agreement with their known biological mechanism, our
200 analysis detected associations between the current count of white blood cells and their rate of change:
201 a negative association among lymphocytes, negative associations between counts of neutrophils and

202 lymphocytes with the rates of monocytes, and a negative association between the counts of platelets
203 and lymphocytes and the rates of neutrophils (**Figure 3E**). Conversely, we found positive associations
204 between monocytes and the rates of each of the investigated white blood cell subsets. These
205 associations, derived from daily counts of white blood cells, could reflect a complex network
206 underlying the regulation of blood immune cell composition[30]. More importantly, the associations
207 quantified for medications and potential homeostatic feedbacks provided a benchmark against which
208 we could compare associations from gut microbial taxa.



209

210 **Figure 3: Bayesian inference conducted in context of the host state (white blood cell and platelet counts)**
211 **and clinical variables including immunomodulatory medications reveals the microbiota potential to affect**

212 **daily changes in circulatory white blood cell counts.** A) Cartoon of the model: observed changes in white
213 blood cell counts between two consecutive days are associated with the current state of the host in the form of
214 blood cell counts in circulation, the administration of immunomodulatory medications, patient and clinical
215 metadata, and the state of the microbiome. B) Visualization of the dynamic white blood cell data; scatter plot of
216 the principal components (PC) of observed daily changes of neutrophils, lymphocytes and monocytes without
217 (grey, see **Figures S7-9**) and with (orange) concurrent longitudinal microbiome data (bold: post engraftment
218 sample counts). C) PBSC patients (N=312) provided the most blood samples with simultaneous microbiome
219 data (n=995) relative to TCD, BM and cord patients, who were used as validation data sets. D-F) PBSC patient
220 data inference results; bars show the posterior parameter estimate distributions (thin: 95% highest posterior
221 density intervals, HPDI95, thick: HPDI50) for the jointly inferred associations between treatments (D), white
222 blood cells counts (E), and fecal microbiota genus log-relative abundances (F) with the observed daily changes
223 in neutrophils, lymphocytes and monocytes. v-score: number of validation cohort confirming associations,
224 *always* set to zero if invalidated in any of the TCD, BM, or cord cohorts (additional coefficients in **Figure S10)**.
225 G) 100 microbiota samples with highest (left) or lowest (right) relative abundances of *Faecalibacterium*,
226 *Ruminococcus 2* and *Akkermansia*. H) Simulation of the neutrophil population over time in presence of GCSF
227 with microbiota compositions sampled either from those high (blue) or low (red) in *Faecalibacterium*,
228 *Ruminococcus 2* and *Akkermansia* relative abundance as shown in G); line: median of 1,000 simulations, shaded
229 regions: quartile range of simulated neutrophil trajectories. I) In absence of GCSF, equivalent simulations to H)
230 predict that the time to reach neutrophil counts $>2K*\mu l^{-1}$ for the first time after HCT when the microbiota is
231 high (red) in *Faecalibacterium*, *Ruminococcus 2* and *Akkermansia* compared with when these genera are low
232 (blue) will decrease from 6.8 (95%-confidence interval, CI: [6.5, 7]) days to 4.4, (CI: [4.3, 4.5]).
233

234 We identified microbial genera that consistently associated with increases or decreases in
235 white blood cell counts by first using data from the PBSC patients and then validating the associations
236 in the other cohorts (**Figure 3F**). Higher abundances of *Faecalibacterium* (+8%, [+1%,
237 +14%]HPDI95 per $\log_{10}$), *Ruminococcus 2* (+5%, [0%, +10%]HPDI95) and *Akkermansia* (+4%,
238 [+1%, +7%]HPDI95) were associated with greater neutrophil increases, whereas increased *Rothia* (-
239 3%, [-7%, 0%]HPDI95), and *Clostridium sensu stricto 1* (-3%, [-6%, 0%]HPDI95) relative
240 abundances associated with reduced neutrophil rates. These results were validated in univariate
241 analyses conducted in the Duke University cohort (**Figure S14, S15**). We also conducted the
242 inference using total genus abundances as predictors instead of relative abundances; this analysis
243 confirmed *Faecalibacterium* as most strongly associated with neutrophil dynamics (**Figure S16**,
244 supplementary methods). *Staphylococcus* was positively associated with lymphocyte rates (+4%,
245 [+1%, +6%]HPDI95) and, again, *Ruminococcus 2* was also associated with faster lymphocyte
246 increases (+5%, [+1%, +9%]HPDI95). Both *Faecalibacterium* as well as *Ruminococcus 2* also
247 associated with increases in monocytes, and while this association was validated in other cohorts (v-
248 score 3 and 1, respectively), there was higher uncertainty of the association estimate (HPDI50>0).
249 Again, *Clostridium sensu strictu 1* (-3% [-5%, -1%]HPDI95) associated consistently with decreased
250 rates of monocytes. The associations we identified—and validated in other cohorts—between

251 microbial taxa in the gut and daily changes in white blood cell counts support the idea that
252 hematopoiesis and mobilization respond to the composition of the gut microbiome, influencing
253 systemic immunity[33].

254 Most of the taxa that strongly associated with white blood cell dynamics were obligate
255 anaerobes. *Rothia*, was a notable exception: this aerobic genus is typically found in the oral cavity[34]
256 but can become an opportunistic pathogen in immunosuppressed patients and is not known to provide
257 metabolic functions to the host[35]. Some obligate anaerobes, on the other hand, produce short-chain
258 fatty acids[36,37] and bacterial cell-wall molecules[1,38,39] that modulate immune responses and
259 granulopoiesis[6]. Nutritional support from the intestinal microbiota improved hematopoietic
260 reconstitution in a mouse model[25]. To identify a similar association in our patients, we estimated a
261 microbiota potency by multiplying the $\log_{10}$-relative abundances of microbial genera in a sample with
262 their corresponding posterior coefficients. We analyzed shotgun metagenomics sequences from 124 of
263 the samples and observed that samples with positive microbiota potency were associated with
264 enrichment in cholate degradation and vitamin-B1 synthesis related pathways, as well as butanoate
265 formation (**Figure S17**). Our findings are in line with evolutionary theory[40] that essential but broadly
266 available microbial traits such as the production of B vitamins, secondary bile acid metabolism, and
267 fermentation to short-chain fatty acids[41] could be co-opted by the host's immune system as part of the
268 homeostatic interplay between immune system and a complex microbiota[42,43]. For example,
269 *Ruminococcus 2* is a genus that contains *R. bromii*, a keystone species necessary for the efficient
270 release of energy from complex starch in the normal diet[44]. Reassuringly, the genera
271 *Faecalibacterium*, *Ruminococcus 2* and *Akkermansia* that we associated with faster rates of white
272 blood cells (**Figure 3F**) were among those best reconstituted by auto-FMT[23], potentially explaining
273 why we found higher counts of neutrophils, monocytes and lymphocytes in patients who received the
274 auto-FMT (**Figure 2B,C**).

275 The associations we reveal are interpretable as potential effectors on sources and sinks of
276 white blood cell counts in circulation. Intestinal bacteria may affect white blood cell counts in
277 circulation by influencing either their sources in the bone marrow or their cytokine profiles[45] and
278 proliferation rates in the blood, their sinks in different organs, or both. The human immune system in

279 turn can interact with the microbiota and modulate its composition, for example via
280 immunoglobulin A responses targeting specific bacteria as studied in mice[43,46,47]. To investigate a
281 reverse effect of the peripheral immune cell system onto bacterial populations, we employed an
282 analogous approach to the stage 1 analysis of white blood cell dynamics. Dynamics of white blood
283 cells can be estimated from changes in absolute cell counts, and to obtain the necessary measurements
284 in absolute bacterial abundances, we measured total bacterial 16S rRNA gene copies per gram of stool
285 for a subset of our samples (3,995 samples from 481 patients). Using absolute abundances of bacteria
286 as predictors in addition to medications, we jointly inferred the association network of dynamics
287 between the gut bacterial ecosystem and the peripheral immune system. All of our patients receive
288 antibiotics on some days during their treatment[24] and their strong effects on microbiota dynamics
289 were the dominant effects that survived cross-validated regularized elastic net regression (**Figure**
290 **S18**). Relaxing the strength of the regularization (methods), however, revealed several bi-directional
291 relationships between immune cells in circulation and bacterial dynamics in the gut (**Figure S19**). Of
292 note, we detected a negative association of absolute *[Ruminococcus] gnavus* group abundance with
293 lymphocytes dynamics, confirming our main result based on relative bacterial abundances (**Figure 3**).
294 In the reverse direction, we saw a positive association of *[Ruminococcus] gnavus* group dynamics
295 with lymphocyte counts. This result agrees with findings that *Ruminococcus gnavus* thrives in and
296 promotes inflammatory conditions such as Crohn's disease and other inflammatory bowel diseases
297 (IBD)[48]; our analysis suggests it may drive high neutrophil to lymphocyte ratios that are broadly
298 characteristic for poor disease outcomes in IBD[49,50] and beyond[51,52].
299 Overall, our analysis identified that the microbiome is associated with immune cell dynamics
300 in addition to medications. The effects should be interpreted as net effects since they do not
301 distinguish, for example, how the microbiota impacts *de novo* hematopoiesis in isolation from its
302 impact on other sources and sinks. Unlike the plausible role of obligate anaerobe fermenters in
303 augmenting hematopoiesis via nutritional support[25], the positive association detected between
304 *Staphylococcus* and lymphocyte dynamics could instead result from reduced extravasation of T cells
305 from circulation into the gut epithelium[53], especially since high abundances of *Staphylococcus* are

306 associated with low gut microbiota diversity ($p<0.001$, **Figure S20**), which indicates a depleted
307 microbiota.
308 Nevertheless, our approach allows us to leverage the chronology of events and assess
309 "mathematical causality"[54]. Of course, due to the observational nature of these data there are risks of
310 confounding that could explain some of the associations found, but the close temporal
311 correspondence[54] between microbiota and blood cell dynamics, and the validation across cohorts
312 reduces the number of plausible confounders. Our results, therefore, quite naturally suggest candidate
313 microbial taxa to manipulate if we seek to steer complex hematopoietic dynamics and utilize the
314 microbiota as an immunomodulatory component of the human body. Intriguingly, members of
315 *Faecalibacterium* and *Ruminococcus* in one study[10], and *Akkermansia* in another[9], were identified as
316 enriched in patients with better responses to anti–PD-1 immunotherapy, which suggested a
317 disagreement between the two studies[55]. Our results, however, revealed *Faecalibacterium*,
318 *Ruminococcus 2,* and *Akkermansia* as the most strongly associated taxa with increases in white blood
319 cell counts from one day to the next. Therefore, our results agree with the findings of both anti PD-1
320 therapy studies that these taxa are associated with immune modulation in humans. Our results also
321 allow us to compare the potency of manipulating these intestinal commensals to that of
322 immunomodulatory drugs. While these genera are common in the gut microbiota of healthy people[17],
323 the relative abundance of each genus can drop below detection in our patients during the intestinal
324 damage related to HCT[24]. Therefore, realistic ranges of 3-5 orders of magnitude in bacterial log-
325 relative abundances (**Figure 3G, S21**) can yield effect sizes similar to that of homeostatic feedbacks
326 between white blood cells and several immunomodulatory medications (e.g. a change in
327 *Ruminococcus 2* from below detection to 1% relative abundance associated with a +67% and +63%
328 increases in neutrophil and lymphocyte rates, respectively). Therefore, while the effect sizes of
329 intestinal bacteria at first may appear smaller than those of immunomodulatory drugs, the herein
330 estimated homeostatic effects of gut bacteria may not be that small since their coefficients refer to
331 changes in exponential rates of white blood cells and accumulate each day. To better demonstrate how
332 this accumulation of effects would work, we conducted simulations of the inferred dynamic system of
333 white blood cells using our posterior coefficient distributions (methods). We simulated 1,000 time

series for microbiota compositions either chosen from the 100 samples highest or lowest in *Faecalibacterium*, *Ruminococcus 2* and *Akkermansia* (**Figure 3G**), in presence (**Figure 3H**) or absence (**Figure 3I**) of GCSF administration. Simulations predict that a microbiota enriched in these genera could accelerate immune reconstitution, and reduce the time until neutrophils reach >2K*μl^{-1} in absence of GCSF by 2.4 days, from predicted 6.8 (CI: [6.5, 7]) to 4.4 days (CI: [4.3,4.5]) days. Gut bacteria, in concert and over time, could therefore have significant impact on systemic immunity even in individuals with less severely injured microbiomes.

In sum, our work links the gut microbiota to the dynamics of the human immune system via peripheral white blood cell populations. Our analysis uses white blood cells counted directly from patients, which are coarse-grained clinical analyses conducted at large scale but lack details such as lymphocyte and other immune cell subsets. Nonetheless, because it is in humans, this study fills an important gap at a critical time for microbiome research when the clinical relevance of animal models of microbiome-immune interaction has been questioned[21]. By studying a large number of patients over time, we were able to infer and quantify for the first time the association of microbiota components on systemic immune cell dynamics, and our results help to consolidate previous findings[10,9] that seemed in conflict with each other[55]. Our study demonstrates that the composition of the microbiota does indeed modulate systemic immune cell dynamics, a link that could be used in the future to improve immunotherapy and help identify microbiota treatments for inflammatory diseases[9,10,56–60].

Bibliography

1. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–118 (2005).
2. Gomez de Agüero, M. *et al.* The maternal microbiota drives early postnatal innate immune development. *Science* **351,** 1296–1302 (2016).
3. Sonnenberg, G. F. & Artis, D. Novel connections and precision approaches. *Nat. Rev. Immunol.* **19,** 75–76 (2019).
4. Geva-Zatorsky, N. *et al.* Mining the human gut microbiota for immunomodulatory organisms. *Cell* **168,** 928–943.e11 (2017).
5. Tan, T. G. *et al.* Identifying species of symbiont bacteria from the human gut that, alone, can induce intestinal Th17 cells in mice. *Proc. Natl. Acad. Sci. USA* **113,** E8141–E8150 (2016).
6. Balmer, M. L. *et al.* Microbiota-derived compounds drive steady-state granulopoiesis via MyD88/TICAM signaling. *J. Immunol.* **193,** 5273–5283 (2014).
7. Deshmukh, H. S. *et al.* The microbiota regulates neutrophil homeostasis and host resistance to Escherichia coli K1 sepsis in neonatal mice. *Nat. Med.* **20,** 524–530 (2014).
8. Tanoue, T. *et al.* A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. *Nature* **565,** 600–605 (2019).
9. Routy, B. *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359,** 91–97 (2018).
10. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359,** 97–103 (2018).
11. Vétizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350,** 1079–1084 (2015).
12. Matson, V. *et al.* The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* **359,** 104–108 (2018).
13. Atarashi, K. *et al.* ATP drives lamina propria T(H)17 cell differentiation. *Nature* **455,** 808–812 (2008).
14. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4,** 337–349 (2008).
15. Cahenzli, J., Köller, Y., Wyss, M., Geuking, M. B. & McCoy, K. D. Intestinal microbial diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe* **14,** 559–570 (2013).
16. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569,** 655–662 (2019).
17. Integrative HMP (iHMP) Research Network Consortium. The integrative human microbiome project. *Nature* **569,** 641–648 (2019).
18. Olin, A. *et al.* Stereotypic immune system development in newborn children. *Cell* **174,** 1277–1292.e14 (2018).
19. Brandi, G. & Frega, G. Microbiota: Overview and Implication in Immunotherapy-Based Cancer Treatments. *Int. J. Mol. Sci.* **20,** (2019).
20. Xin Yu, J., Hubbard-Lucey, V. M. & Tang, J. The global pipeline of cell therapies for cancer. *Nat. Rev. Drug Discov.* **18,** 821–822 (2019).
21. Walter, J., Armet, A. M., Finlay, B. B. & Shanahan, F. Establishing or Exaggerating

402 Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated
403 Rodents. *Cell* **180,** 221–232 (2020).
404 22. Taur, Y. *et al.* Intestinal domination and the risk of bacteremia in patients undergoing
405 allogeneic hematopoietic stem cell transplantation. *Clin. Infect. Dis.* **55,** 905–914 (2012).
406 23. Taur, Y. *et al.* Reconstitution of the gut microbiota of antibiotic-treated patients by
407 autologous fecal microbiota transplant. *Sci. Transl. Med.* **10,** (2018).
408 24. Morjaria, S. *et al.* Antibiotic-Induced Shifts in Fecal Microbiota Density and
409 Composition during Hematopoietic Stem Cell Transplantation. *Infect. Immun.* **87,**
410 (2019).
411 25. Staffas, A. *et al.* Nutritional Support from the Intestinal Microbiota Improves
412 Hematopoietic Reconstitution after Bone Marrow Transplantation in Mice. *Cell Host*
413 *Microbe* **23,** 447–457.e4 (2018).
414 26. Taur, Y. *et al.* The effects of intestinal tract bacterial diversity on mortality following
415 allogeneic hematopoietic stem cell transplantation. *Blood* **124,** 1174–1182 (2014).
416 27. Savani, B. N. *et al.* Absolute lymphocyte count on day 30 is a surrogate for robust
417 hematopoietic recovery and strongly predicts outcome after T cell-depleted allogeneic
418 stem cell transplantation. *Biol. Blood Marrow Transplant.* **13,** 1216–1223 (2007).
419 28. Mehta, R. S. & Rezvani, K. Immune reconstitution post allogeneic transplant and the
420 impact of immune recovery on the risk of infection. *Virulence* **7,** 901–916 (2016).
421 29. Kim, H. T. *et al.* Absolute lymphocyte count recovery after allogeneic hematopoietic
422 stem cell transplantation predicts clinical outcome. *Biol. Blood Marrow Transplant.* **21,**
423 873–880 (2015).
424 30. Scheiermann, C., Frenette, P. S. & Hidalgo, A. Regulation of leucocyte homeostasis in
425 the circulation. *Cardiovasc. Res.* **107,** 340–351 (2015).
426 31. Thompson, P. A. *et al.* Umbilical cord blood graft engineering: challenges and
427 opportunities. *Bone Marrow Transplant.* **50 Suppl 2,** S55-62 (2015).
428 32. Gabrilove, J. L. *et al.* Effect of granulocyte colony-stimulating factor on neutropenia and
429 associated morbidity due to chemotherapy for transitional-cell carcinoma of the
430 urothelium. *N. Engl. J. Med.* **318,** 1414–1422 (1988).
431 33. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell*
432 **157,** 121–141 (2014).
433 34. Hall, M. W. *et al.* Inter-personal diversity and temporal dynamics of dental, tongue, and
434 salivary microbiota in the healthy oral cavity. *npj Biofilms and Microbiomes* **3,** 2 (2017).
435 35. Ramanan, P., Barreto, J. N., Osmon, D. R. & Tosh, P. K. Rothia bacteremia: a 10-year
436 experience at Mayo Clinic, Rochester, Minnesota. *J. Clin. Microbiol.* **52,** 3184–3189
437 (2014).
438 36. Ulven, T. Short-chain free fatty acid receptors FFA2/GPR43 and FFA3/GPR41 as new
439 potential therapeutic targets. *Front. Endocrinol. (Lausanne)* **3,** 111 (2012).
440 37. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic
441 Treg cell homeostasis. *Science* **341,** 569–573 (2013).
442 38. Hergott, C. B. *et al.* Peptidoglycan from the gut microbiota governs the lifespan of
443 circulating phagocytes at homeostasis. *Blood* **127,** 2460–2471 (2016).
444 39. Ladinsky, M. S. *et al.* Endocytosis of commensal antigens by intestinal epithelial cells
445 regulates mucosal T cell homeostasis. *Science* **363,** (2019).
446 40. Foster, K. R., Schluter, J., Coyte, K. Z. & Rakoff-Nahoum, S. The evolution of the host
447 microbiome as an ecosystem on a leash. *Nature* **548,** 43–51 (2017).
448 41. Yoshii, K., Hosomi, K., Sawane, K. & Kunisawa, J. Metabolism of dietary and microbial
449 vitamin B family in the regulation of host immunity. *Front. Nutr.* **6,** 48 (2019).
450 42. Zaza, G. *et al.* Impact of maintenance immunosuppressive therapy on the fecal
451 microbiome of renal transplant recipients: Comparison between an everolimus- and a

452 standard tacrolimus-based regimen. *PLoS One* **12,** e0178228 (2017).
453 43. McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L. & Foster, K. R. Host
454 selection of microbiota via differential adhesion. *Cell Host Microbe* **19,** 550–559 (2016).
455 44. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone
456 species for the degradation of resistant starch in the human colon. *ISME J.* **6,** 1535–1543
457 (2012).
458 45. Schirmer, M. *et al.* Linking the human gut microbiome to inflammatory cytokine
459 production capacity. *Cell* **167,** 1125–1136.e8 (2016).
460 46. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota
461 and the immune system. *Science* **336,** 1268–1273 (2012).
462 47. Palm, N. W. *et al.* Immunoglobulin A coating identifies colitogenic bacteria in
463 inflammatory bowel disease. *Cell* **158,** 1000–1010 (2014).
464 48. Henke, M. T. *et al.* Ruminococcus gnavus, a member of the human gut microbiome
465 associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc. Natl.*
466 *Acad. Sci. USA* **116,** 12672–12677 (2019).
467 49. Okba, A. M. *et al.* Neutrophil/lymphocyte ratio and lymphocyte/monocyte ratio in
468 ulcerative colitis as non-invasive biomarkers of disease activity and severity. *Auto*
469 *Immun. Highlights* **10,** 4 (2019).
470 50. Celikbilek, M. *et al.* Neutrophil-lymphocyte ratio as a predictor of disease severity in
471 ulcerative colitis. *J Clin Lab Anal* **27,** 72–76 (2013).
472 51. Gao, Y. *et al.* Neutrophil/lymphocyte ratio is a more sensitive systemic inflammatory
473 response biomarker than platelet/lymphocyte ratio in the prognosis evaluation of
474 unresectable pancreatic cancer. *Oncotarget* **8,** 88835–88844 (2017).
475 52. Choi, S.-J. *et al.* High neutrophil-to-lymphocyte ratio predicts short survival duration in
476 amyotrophic lateral sclerosis. *Sci. Rep.* **10,** 428 (2020).
477 53. Fu, Y.-Y. *et al.* T Cell Recruitment to the Intestinal Stem Cell Compartment Drives
478 Immune-Mediated Intestinal Damage after Allogeneic Transplantation. *Immunity* **51,**
479 90–103.e3 (2019).
480 54. Gerber, G. K. The dynamic microbiome. *FEBS Lett.* **588,** 4131–4139 (2014).
481 55. Jobin, C. Precision medicine using microbiota. *Science* **359,** 32–34 (2018).
482 56. Zitvogel, L., Ma, Y., Raoult, D., Kroemer, G. & Gajewski, T. F. The microbiome in
483 cancer immunotherapy: Diagnostic tools and therapeutic strategies. *Science* **359,** 1366–
484 1370 (2018).
485 57. Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities
486 associated with neurodevelopmental disorders. *Cell* **155,** 1451–1463 (2013).
487 58. Lakritz, J. R. *et al.* Gut bacteria require neutrophils to promote mammary tumorigenesis.
488 *Oncotarget* **6,** 9387–9396 (2015).
489 59. Lakritz, J. R. *et al.* Beneficial bacteria stimulate host immune cells to counteract dietary
490 and genetic predisposition to mammary cancer in mice. *Int. J. Cancer* **135,** 529–540
491 (2014).
492 60. Chen, H. *et al.* A forward chemical genetic screen reveals gut microbiota metabolites
493 that modulate host physiology. *Cell* **177,** 1217–1231.e18 (2019).
494 61. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the
495 Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).
496 62. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon
497 data. *Nat. Methods* **13,** 581–583 (2016).
498 63. Murali, A., Bhargava, A. & Wright, E. S. IDTAXA: a novel approach for accurate
499 taxonomic classification of microbiome sequences. *Microbiome* **6,** 140 (2018).
500 64. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data
501 processing and web-based tools. *Nucleic Acids Res.* **41,** D590-6 (2013).

65. Pinheiro, J. C., Bates, D. M., DebRoy, S. S. & Sarkar, D. Nlme: Linear and Nonlinear Mixed Effects Models. (2013).
66. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58,** 267–288 (1996).
67. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* (2011).
68. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* **2,** e55 (2016).
69. Hoffman, M. D. & Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* (2014).
70. Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15,** 962–968 (2018).
71. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *arXiv* (2019).

519 **Supplementary Information**

520 Appended and available online.

521

522 **Acknowledgments and Conflicts of Interest**

523 We thank Marc Lipsitch, Sandra B. Andersen, Kevin R. Foster, Jonathan Kevin Sia, Eric G. Pamer,
524 Kat Coyte, Sibylle Mitschka and the members of the Xavier lab for helpful discussion and comments
525 on the manuscript. This work was supported by the National Institutes of Health (NIH) grant U01
526 AI124275 to JBX and grant R01 AI137269 to JBX, by the MSKCC Cancer Center Core Grant P30
527 CA008748, the Parker Institute for Cancer Immunotherapy at Memorial Sloan Kettering Cancer
528 Center, the Sawiris Foundation, the Society of Memorial Sloan Kettering Cancer Center, MSKCC
529 Cancer Systems Immunology Pilot Grant and Empire Clinical Research Investigator Program. MS
530 received funding from the Burroughs Wellcome Fund Postdoctoral Enrichment Program, the Damon
531 Runyon Physician-Scientist Award, and the Robert Wood Johnson Foundation. MRMvdB and JUP
532 received financial support from Seres Therapeutics. TMH is investigator in the Pathogenesis of
533 Infectious Diseases from the Burroughs Wellcome Fund, and funded via an award from Geoffrey
534 Beene Foundation, and NIH RO1 AI093808. M-AP has received honoraria from AbbVie, Bellicum,
535 Bristol-Myers Squibb, Incyte, Merck, Novartis, Nektar Therapeutics, and Takeda; has received
536 research support for clinical trials from Incyte, Kite (Gilead) and Miltenyi Biotec; and serves on data
537 and safety monitoring boards for Servier and Medigene and scientific advisory boards for MolMed
538 and NexImmune. The funders had no role in study design, data collection and analysis, decision to
539 publish, or preparation of the manuscript.

540

541 **Author Contributions**

542 J.S. and J.B.X. wrote the manuscript. J.S. and J.B.X. designed the analyses with expert help from
543 R.N.. J.U.P. and Y.T. contributed to the clinical data preparation, B.P.T. provided the 16S data
544 processing pipelines, A.D. provided the shotgun processing pipelines. All authors contributed to the
545 writing and interpretation of the results.

546 **SUPPLEMENTARY INFORMATION**

547 **Methods**

548 **Complete blood count collection and characterization**

549 Absolute white blood cells count data were obtained from routine complete blood counts ordered by
550 clinicians during normal clinical practice, used to obtain informative diagnostic and monitoring
551 information. Blood samples received in the clinical hematology laboratory were analyzed using
552 Sysmex XN automated hematology analyzers (Sysmex, Lincolnshire, IL) and, when needed based on
553 specific flags and parameters as per MSKCC standard operating procedures, were validated manually
554 using the Sysmex DI-60 Slide Processing System or CellaVision DM9600 Automated Digital
555 Morphology System (Sysmex, Lincolnshire, IL).

556

557 **16S rRNA gene amplification and multiparallel sequencing**

558 For each sample, duplicate 50-µl PCRs were performed, each containing 50 ng of purified DNA, 0.2
559 mM deoxynucleotide triphosphates, 1.5 mM MgCl2, 2.5 U Platinum Taq DNA polymerase, 2.5 µl of
560 10× PCR buffer, and 0.5 µM of each primer designed to amplify the V4-V5: 563F (5′-nnnnnnnn-
561 NNNNNNNNNNNN-AYTGGGYDTAAAGNG-3′) and 926R (5′-nnnnnnnn-NNNNNNNNNNNN-
562 CCGTCAATTYHTTTRAGT-3′). A unique 12-base Golay barcode (Ns) precedes the primers for
563 sample identification[61], and one to eight additional nucleotides were placed in front of the barcode to
564 offset the sequencing of the primers. Cycling conditions were 94°C for 3 min, followed by 27 cycles
565 of 94°C for 50 s, 51°C for 30 s, and 72°C for 1 min. For the final elongation step, 72°C for 5 min was
566 used. Replicate PCRs were pooled, and amplicons were purified using the QIAquick PCR Purification
567 Kit (Qiagen). PCR products were quantified and pooled at equimolar amounts before Illumina
568 barcodes and adaptors were ligated, using the Illumina TruSeq Sample Preparation protocol. The
569 completed library was sequenced on an Illumina MiSeq platform following the Illumina
570 recommended procedures with a paired-end 250 × 250 bp kit

571

572 **Sequence analysis**

573 The 16S (V4-V5) paired-end reads were merged and demultiplexed. Amplicon sequence variants
574 (ASVs) were identified using the Divisive Amplicon Denoising Algorithm (DADA2) pipeline
575 including filtering and trimming of the reads[62]. Reads were trimmed to the first 180 bp or the first
576 point with a quality score Q<2. Reads were removed if they contained ambiguous nucleotides (N) or
577 if two or more errors were expected based on the quality of the trimmed read. We assigned taxonomy
578 to ASVs using a 8-mer based classifier trained by IDTaxa[63] using the SILVA database[64]. We
579 determined the copy number of 16S rRNA genes per gram of stool for 4,158 of our samples as
580 reported previously[24], by quantitative PCR on total DNA extracted from fecal samples.
581
582 **Quantification of total microbiota density per gram of stool and estimation of total genus**
583 **abundances.**
584 Quantitative PCR (qPCR) was performed on DNA extracted from the 1g wet weight of a stool sample
585 using DyNAmo SYBR Green qPCR kit (Finnzymes) and 0.2 μM of the universal bacterial primer 8F
586 (5′-AGAGTTTGATCCTGGCTCAG) and the broad-range bacterial primer 338R
587 (5′-TGCTGCCTCCCGTAGGAGT-3′). Standard curves were prepared by serial dilution of the PCR
588 blunt vector (Invitrogen) containing 1 copy of the 16s rRNA gene. Cycling conditions were 95°C for
589 10 minutes followed by 40 cycles of 95°C for 30 seconds, 52°C for 30 seconds, and 72°C for 1
590 minute. We used the measurements of total 16S rRNA gene counts per gram of stool to multiply the
591 relative abundances of taxa obtained from 16S amplicon sequencing to obtain the estimate of their
592 total abundance per gram of stool (supplementary methods). Importantly, this does not account for
593 16S copy number variation between taxa, but the observed dynamic ranges in total abundances of taxa
594 in our data set span up to 9 orders of magnitude, exceeding the potential inaccuracies due to copy
595 number variation.
596
597 **Diversity calculations**

598 Microbiome alpha-diversity was measured by the inverse Simpson (IS) index of a sample. It was

599 calculated by $\mathrm{IS_i} = \frac{1}{\sum_{j=1}^{N} p_{ij}^2}$, where p is the relative abundance of the jth ASV out of N total ASVs in

600 sample i.

601

602 **Linear mixed-effects model of white blood cell counts**

603 To study the effect of auto-FMT on white blood cells, we investigated the white blood cell counts of

604 24 enrolled patients of this trial from the day of neutrophil engraftment until 100 days after. FMT

605 occurred on different days relative to neutrophil engraftment. Thus, we performed an analogous

606 analysis to that conducted in the original publication that demonstrated how FMT re-established a

607 diverse microbiome in the post-FMT period[23]. To answer if white blood cell counts differed post-

608 FMT, we used a linear mixed effects model of white blood cell counts, y, modeled as a function of the

609 FMT treatment as well as patient and timepoint specific random effects. We included random

610 intercept terms for each day i and each patient j, and a fixed effects term for the post-FMT period with

611 associated coefficient "armpost", using the indicator variable "FMT", that is 1 when a patient was

612 from the FMT treated arm of the trial and *day* was greater than or equal to the day of the FMT

613 procedure. We conducted independent analyses for neutrophil, lymphocyte and monocyte counts.

614 This resulted in the following model of a cell count, y, for patient j on day i:

615 $$y_{ij} = \beta_0 + armpost * FMT_{ij} + day_i + patient_j + \varepsilon_{ij}, \qquad i = 0, \dots, D, j = 1, \dots, P$$

616 with prior distributions $day_i \sim \mathcal{N}(0,\ \sigma_{day}^2)$, and $patient_j \sim \mathcal{N}(0,\ \sigma_{patient}^2)$, independent error

617 $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and fixed intercept β_0, for the D days post neutrophils engraftment and P patients,

618 (D=100, P=24). For convenience of those interested in reanalyzing our data, the part of our data

619 concerning the auto-FMT analysis is available in tidy format (supplementary data 11), and the

620 analysis code conducted in the R programming language is available as an exported notebook

621 (fmt_effect_on_wbc.pdf) on github: https://github.com/jsevo/wbcdynamics_microbiome/.[65] We

622 conducted an additional analysis with "day" as a continuous predictor which did not change our

623 conclusions (supplementary methods).

624

625 **Dynamic systems analyses**

626 We analyzed factors associated with the observed changes of absolute counts of neutrophils,
627 lymphocytes and monocytes between two days. In the following we describe how chronology of
628 events and biological samples were encoded, and the models used to infer a role of medications,
629 clinical parameters and the microbiome on dynamics of white blood cells.

630 To reveal factors that associate with day-to-day changes in white blood cell counts, we started
631 from a first-order differential equation of white blood cell (W) dynamics:

632 $$\frac{\mathrm{d(W)}}{\mathrm{d}t} = W(gr + \sum_{j=1}^{P} \beta_j X_j)$$

633 Where g*r* represents the intercept, i.e. the base line rate of change during immune reconstitution, and
634 β_j are the to-be-estimated coefficients of the P predictors $X_j, j \in P$, of the white blood cell dynamics.
635 This equation was then linearized to

636 $$\frac{\mathrm{d}\, ln(\mathrm{W})}{\mathrm{d}t} = gr + \sum_{j=1}^{P} \beta_j X_j$$

637 And we parameterized the corresponding discrete difference equation:

638 $$\frac{\Delta \ln(\mathrm{W})}{\Delta t} = gr + \sum_{j=1}^{P} \beta_j X_j$$

639 where Δln(W) is the log-difference between single days of neutrophils, lymphocytes or monocytes
640 counts, and Δt=1 for all intervals. Predictors include the counts of neutrophils, lymphocytes,
641 monocytes, eosinophils and platelets during an interval (homeostatic feedbacks), immunomodulatory
642 medication and clinical observations such as a blood stream infection and the onset of graft versus
643 host disease, HCT parameters such as graft types and conditioning regimens, and, additionally, the
644 microbiota composition in "stage 2" of our analysis (supplementary methods for data exclusion and
645 additional details on interval definitions). Importantly, by parameterizing a dynamic equation and
646 analyzing *rates of change*, our coefficient estimates have an immediate causal interpretation within
647 our modeling framework (i.e. a β_j>0 implies that higher levels of the corresponding X_j *increases* the

respective white blood cell type, *W*). To differentiate such results from other associations, analyses of this type have been termed "mathematical causality"[54].

Stage 1 analysis: Feature selection. Identifying medications and clinical observations associated with white blood cell dynamics from patients without microbiome data

Stage 1 uses data of patients without any available microbiome samples and the following model of white blood cell changes, y:

$$y = gr + \sum_{j=1}^{p} \beta_j X_j,$$

with intercept, *gr*.The predictors, *X*, include dummy variables for the HCT graft type, patients' age on the date of HCT, sex, 13 most frequently observed positive blood cultures with remaining other blood stream infections grouped into a separate category "other infections", an indicator for the onset of graft versus host disease, administrations of 55 different, most common immunomodulatory medications and platelet transfusion events, and HCT conditioning intensity regimens as well as the log-transformed geometric mean counts of neutrophils, lymphocytes, monocytes, eosinophils and platelets during the respective interval. We used elastic net regression[66] for feature selection using the sklearn package for the Python programming language[67]. For elastic net regression with 50% L1-penalty, predictors were scaled between zero and 1, and we used 10-fold cross validation (i.e. leaving out 10% of patients at each cross-validation step) to choose the regularization strength, λ, solving for

$$argmin_{gr,\beta} \left\{ \frac{1}{2N} \sum_{i=1}^{N} (y_i - gr - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \frac{1}{2}\lambda \sum_{j=1}^{p} |\beta_j| + \frac{1}{2}\lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

Stage 1 yielded a sparse coefficient matrix of predictors used to design the model in stage 2.

Expanded analysis on patients with microbiome data – stage 2

To identify associations between microbiota and white blood cell dynamics, we conducted an analogous, Bayesian regression using the package PyMC3 for the Python programming language[68]. Stage 1 identified important difference between transplant types, and we therefore stratified our data into 4 cohorts according to their stem cell graft source. Using data independently from each cohort,

675 we applied "no U-turn" sampling[69] to produce 10,000 posterior samples from 5 independent MCMC
676 chains that parameterized the model:

677 $$y \sim \mathcal{N}(\mu, \sigma^2)$$

678 $$\mu = gr + \sum_{j=1}^{\hat{P}} x_j \beta_j$$

679 with uninformative prior distributions

680 $$gr \sim \mathcal{N}\ (mean = 0,\ standard\ deviation = 100)$$

681 $$\beta_j \sim \mathcal{N}\ (mean = 0,\ standard\ deviation = 100)$$

682 $$\sigma \sim HalfCauchy\ (\ beta = 2)$$

683 where y is the observed daily change of a focal white blood cell type as in stage 1 with normal
684 distributed mean, μ, and σ, the model uncertainty with a thick-tailed half Cauchy prior (importantly,
685 our posterior estimates do not depend on this choice as we obtain the same results with an inverse
686 Gamma prior, figure S19). μ was a function of the baseline growth rate, gr, and predictors, $\hat{P}$:
687 medications with non-zero coefficients in stage 1, the white blood cell counts, patient age and sex,
688 and HCT conditioning intensities; additionally, $\hat{P}$ now included the log-abundances of microbial
689 genera as measured by 16S sequencing from DNA in the stool collected on the second day of a daily
690 interval (see supplementary methods for details). We considered taxa that were among the 100 most
691 abundant, or had reached maximum relative abundances of at least 10%, and selected those who were
692 non-zero in more than 75% of our samples. White blood cell counts and microbiota data present
693 during a daily interval were log-transformed, and zeros were filled with half of the minimum observed
694 non zero counts (i.e. 0.5e3 and 2e-6, respectively). We focused on the largest cohort (PBSC) and
695 used the independent inference results from TCD, BM, and cord cohorts for validation.
696

697 *Validation score*

698 Coefficients learnt from the PBSC patient cohort were assigned a "validation score" based on the
699 results obtained from the other three MSK patient cohorts. Our requirements for validation were
700 conservative; we required evidence from our validation data sets as well as absence of counter

701 evidence. For regression results from each of the validation graft type cohorts, i.e. TCD, BM, and
702 cord, we checked if a coefficient had more than 75% probability (50%HPDI) to have the same sign as
703 the mean of the PBSC coefficient posterior for a given predictor. If so, this was considered evidence
704 of validation, and we summed the evidence over the three validation sets (i.e. maximum score of 3, 1
705 from each of TCD, BM, and cord cohorts). Conversely, if we found more than 75% probability
706 among any of the validation data sets that a given predictor had the *opposite* sign as the posterior
707 mean calculated from PBSC data, this was considered counter evidence and the validation score was
708 always set to zero.
709
710 *Analysis of white blood cell dynamics with absolute bacterial abundances as predictors instead of*
711 *relative abundances*
712 We conducted an ordinary least squares regression using the statsmodels package in the Python
713 programming language of the same model as in the main Bayesian analysis using total bacterial
714 abundances as predictors. This was only possible on a subset of 389 neutrophil, 331 lymphocyte and
715 376 monocyte rate observations from PBSC patients.
716
717 *Forwards simulation of predicted immune system reconstitution kinetics*
718 To assess the impact of the estimated microbiota coefficients on immune system dynamics, we
719 conducted 1,000 simulations of the system of 3 differential equations describing the dynamics of
720 neutrophils, lymphocytes and monocytes. We ran 1,000 simulations four times: in presence and
721 absence of GCSF, each with microbiota compositions enriched or depleted in *Faecalibacterium*,
722 *Ruminococcus 2* and *Akkermansia*. To identify these compositions, we ranked the observed
723 microbiota compositions by these taxa, and chose randomly either from the top or bottom 100. The
724 coefficients for white blood cell interactions, interactions with the microbiota and the effect of GCSF
725 were sampled from our posterior coefficient distributions. Using these coefficients sampled at the start
726 of the simulation, and using 50 cells*μl^{-1} of neutrophils, lymphocytes and monocytes as initial values,
727 we simulated these differential equations forwards in time using the odeint function of the scipy
728 package for the Python programming language.

Validation on data from Duke University

We analyzed 9,603 blood samples with 25,581 associated administrations of immunomodulatory medications, and 741 microbiota samples from Duke as an orthogonal data set to validate our findings. The temporal resolution of this data was much lower, and after filtering for samples from the relevant post neutrophil engraftment period, and by requiring daily intervals, 83 valid, complete data points were available. Using these data, we correlated daily blood cell changes individually in univariate, or jointly in a partial least squares regression, with those predictors that achieved more than 95% probability density in the positive or negative domain in the PBSC data regression. For each of these predictors, we present the sign of slopes and Bonferroni corrected *p*-values from individual linear regressions.

Joint analysis of the effect of antibiotics and white blood cell counts on the microbiota and the microbiota and immunomodulatory medications on white blood cell counts

Analogous to *stage 1*, we performed cross-validated, regularized linear regressions (ElasticNet) using the scikit-learn package for the Python programming language to jointly estimate the association network between microbiota and circulatory white blood cells. For this, we constructed a block matrix **X** of predictor matrices $\boldsymbol{X_i}$ that include the absolute bacterial abundances, drug data (antibiotics for bacterial dynamics and immune modulators for white blood cell dynamics), as well as the counts of white blood cells and a separate intercept term per block. Each block $\boldsymbol{X^l_{n_l,p_l}}$, with n_l observations and p_l predictors (l=0...k), on the diagonal of **X** corresponds to the indices of the observed daily log-changes of one of the 41 bacterial genera considered in our main analysis or the log changes in neutrophil, lymphocyte and monocyte counts from PBSC patients contained in **Y** (in total we calculated 15,833 rates from 256 patients). Our regression problem can thus be written as:

$$argmin_{\beta}\ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \text{ where } \mathbf{X} = \begin{matrix} \boldsymbol{X^0_{n_0,p_0}} & \cdots & \mathbf{0}_{n_0,p_k} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{n_k,p_0} & \cdots & \boldsymbol{X^k_{n_k,p_k}} \end{matrix}$$

with k=44, i.e. 41 bacterial genera and 3 white blood cell types, the to-be estimated coefficient vector **β** and **0** the zero matrix. This system is underdetermined and we therefore chose the same approach as

756 in stage 1, elastic net regression, for feature selection. Predictors were scaled between zero and 1, and
757 we used 3-fold cross validation, leaving out 1/3rd of the patients at each iteration to identify a global
758 regularization strength, λ, solving for

759 $$argmin_{\beta}\left\{\frac{1}{2\eta}\sum_{i=1}^{\eta}(y_i-\sum_{j=1}^{\rho}x_{ij}\beta_j)^2+\frac{1}{2}\lambda\sum_{j=1}^{\rho}|\beta_j|+\frac{1}{2}\lambda\sum_{j=1}^{\rho}\beta_j^2\right\}$$

760

761 where η is the total number of observed daily log changes in genera and white blood cells, and ρ the
762 total number of predictors. This yielded a strongly regularizing λ_s, and thus few predictors. To
763 characterize potential bidirectional relationships between white blood cell counts and the gut
764 microbiota, we iteratively reduced the regularization strength until the strongest interaction between
765 microbiota and white blood cell dynamics, i.e. *Faecalibacterium* with neutrophil dynamics, was
766 detected. We than re-ran the regression with this reduced regularization strength, λ_r.

767

768 *Shotgun sequencing*

769 Sequencing of 124 post-neutrophil engraftment was conducted on the Illumina HiSeq platform. For
770 details and the processing of the FASTQ files, see supplementary methods. We used the HUMAnN2
771 pipeline[70] with default settings for functional profiling of our samples, with the UniRef90 data base
772 and ChocoPhlAn for alignment, and we renormalized our samples by library depth to copies per
773 million. We used MetaCyc to obtain stratified and unstratified pathway abundances.

774

775 *Statistical analysis of shotgun data*

776 We calculated the predicted microbiota potency score for each sample and separately for neutrophils,
777 lymphocytes and monocytes, by multiplying the abundances of taxa in each of the 124 samples with
778 the corresponding posterior coefficients obtained from the PBSC inference. To distinguish the sets of
779 metabolic functions that separate samples with positive and negative predicted potencies, we
780 converted the pathway abundances into presence and absences profiles. We performed a linear
781 discriminant analysis between positive and negative potency samples with a least squares solver and
782 automatic shrinkage using the Ledoit-Wolf lemma using the sklearn package for the Python
783 programming language[67]. To assess differences in the presence or absence of pathways between
784 samples with positive and negative potency, we used Fisher's exact test for each pathway.

Supplementary Figures

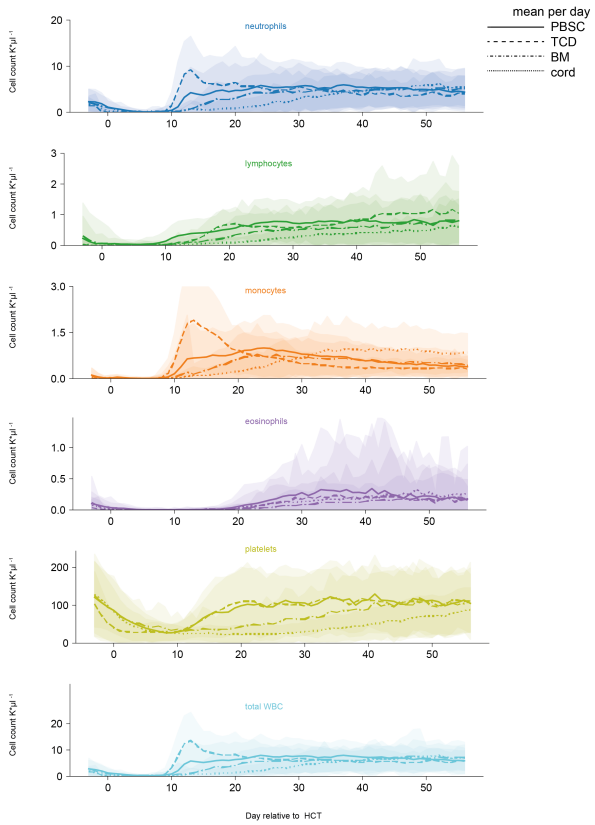


Figure S1: White blood cell counts and platelet counts per graft source over the first 100 days post HCT per day relative to HCT; lines: mean, shaded: ± standard deviations).

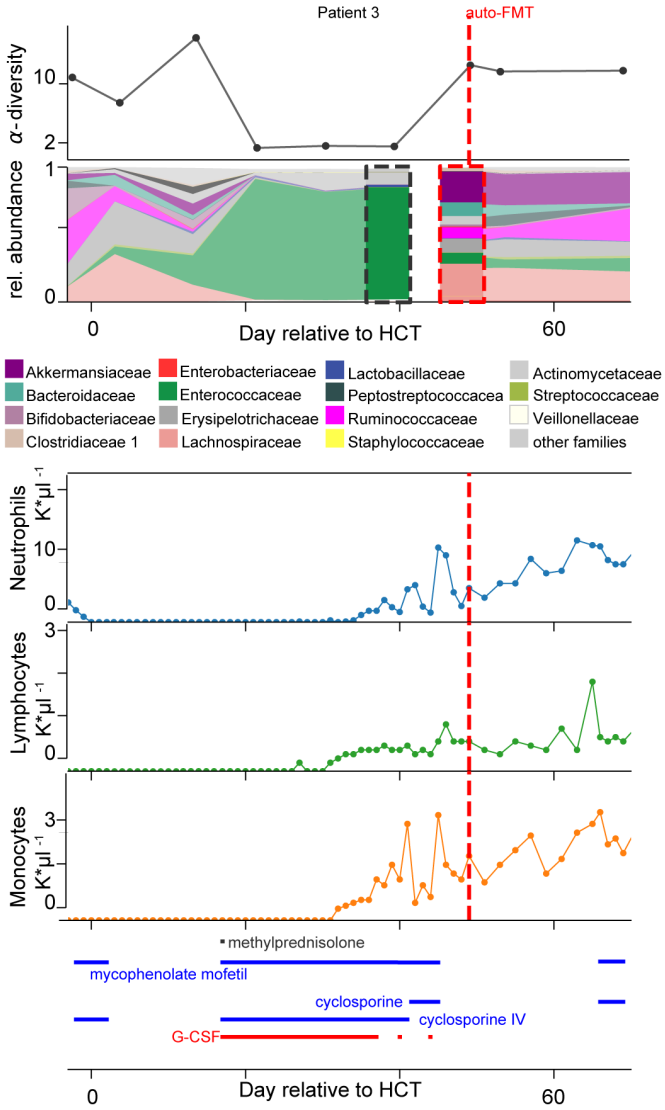


Figure S2: HCT patient who received an autologous fecal microbiota transplant (auto-FMT, dashed red line) that restored commensal microbial families and ecological diversity in the gut microbiota, with concurrent cell counts of peripheral neutrophils, lymphocytes and monocytes and immunomodulatory drug administrations.

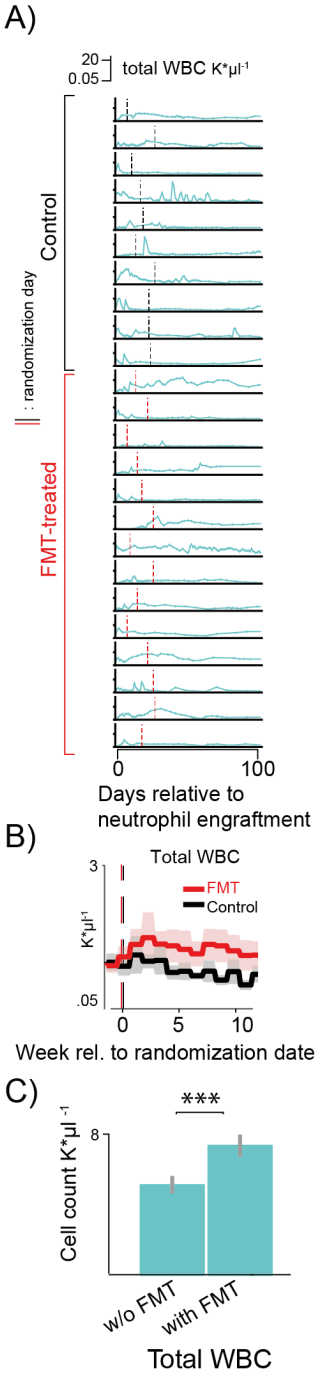


Figure S3: Total counts of white blood cells increased in FMT-treated patients in the weeks following the treatment compared to control patients. A) Total white blood cell counts in 24 patients enrolled in a randomized controlled trial to receive an autologous fecal microbiota transplant post-neutrophil engraftment (10 control: black vertical line, 14 FMT treated: red vertical line,). B) Weekly mean cell counts aligned to the date of randomization into FMT treatment arm (red) or control (black). Line: weekly mean, shaded region: 95%-CI C) Results from a linear mixed effects model with random effects per patient and per day relative to neutrophil engraftment confirms that the total white blood cell counts is higher in patients receiving auto-FMT after their treatment as compared to control patients after the randomization date, bars and confidence intervals for the averages of observed white blood cell counts without FMT and post-FMT (***: $p<0.001$).

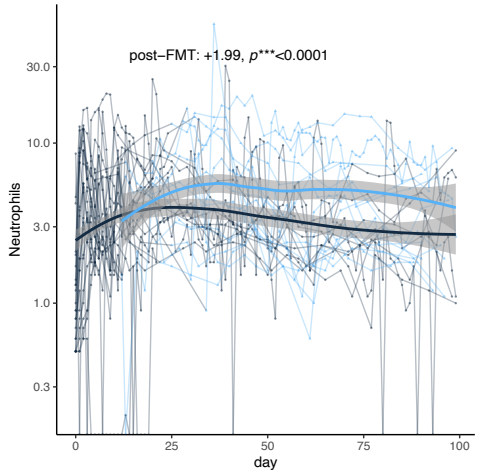


Figure S4: Neutrophil counts in 24. FMT trial patients. Thin lines: raw data (blue: post-FMT); thick black: mean per day, thick blue: mean+post-FMT coefficient. Means and confidence intervals (shaded region) from linear mixed effects model (methods).

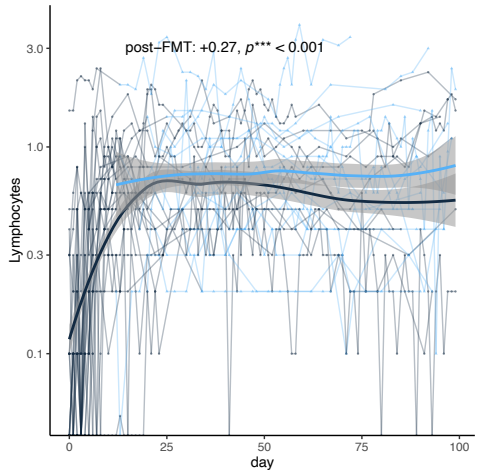


Figure S5: Lymphocyte counts in 24. FMT trial patients. Thin lines: raw data (blue: post-FMT); thick black: mean per day, thick blue: mean+post-FMT coefficient. Means and confidence intervals (shaded region) from linear mixed effects model (methods).

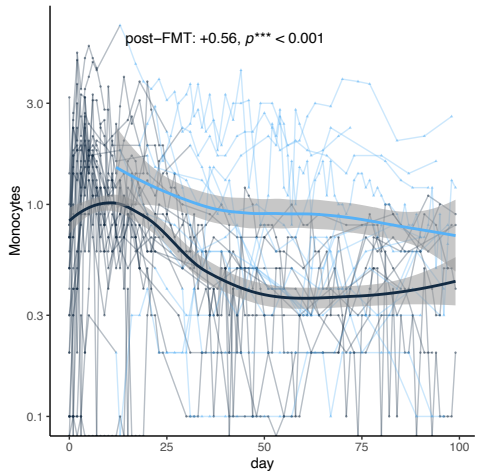


Figure S6: Monocyte counts in 24. FMT trial patients. Thin lines: raw data (blue: post-FMT); thick black: mean per day, thick blue: mean+post-FMT coefficient. Means and confidence intervals (shaded region) from linear mixed effects model (methods).

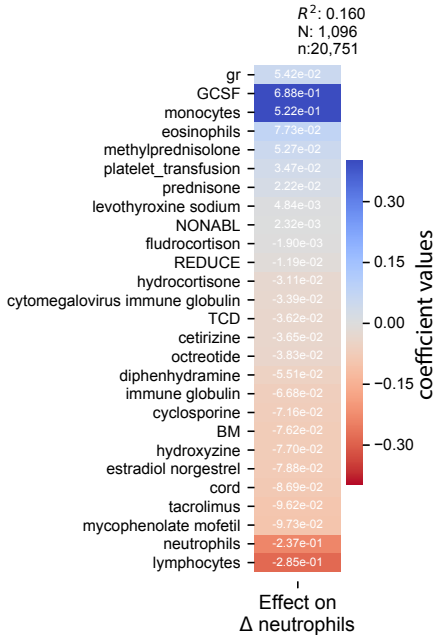


Figure S7: "Stage 1" regression on neutrophil dynamics on patients without microbiome data. Coefficients from 10-fold cross-validated elastic net regression daily changes in neutrophils. gr: intercept; TCD: T-cell depleted graft (ex-vivo) by CD34+selection; PBSC: peripheral blood stem cells; BM: bone marrow; cord: umbilical cord blood; NONABL: Nonmyeloablative; REDUCE: reduced-intensity conditioning regimen; F: female; N: patients, n: samples (daily changes in neutrophils).

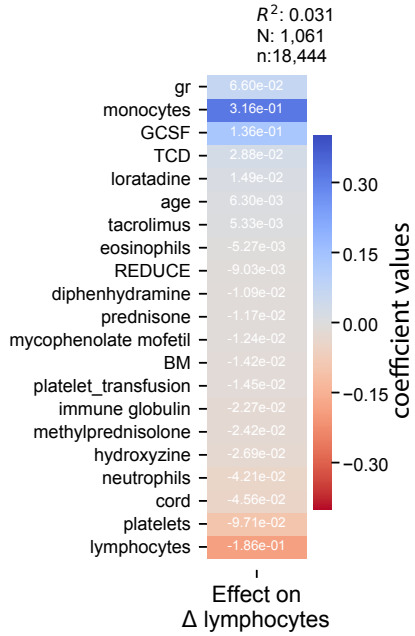


Figure S8: "Stage 1" regression on lymphocytes dynamics on patients without microbiome data. Coefficients from 10-fold cross-validated elastic net regression daily changes in lymphocytes. gr: intercept; TCD: T-cell depleted graft (ex-vivo) by CD34+selection; PBSC: peripheral blood stem cells; BM: bone marrow; cord: umbilical cord blood REDUCE: reduced-intensity conditioning regimen; F: female. N: patients, n: samples (daily changes in lymphocytes).

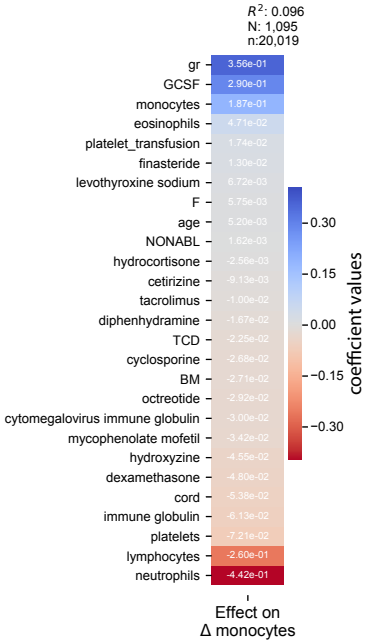


Figure S9: "Stage 1" regression on lymphocytes dynamics on patients without microbiome data. Coefficients from 10-fold cross-validated elastic net regression daily changes in lymphocytes. gr: intercept; TCD: T-cell depleted graft (ex-vivo) by CD34+selection; PBSC: peripheral blood stem cells; BM: bone marrow; cord: umbilical cord blood REDUCE: reduced-intensity conditioning regimen; F: female. N: patients, n: samples (daily changes in lymphocytes).

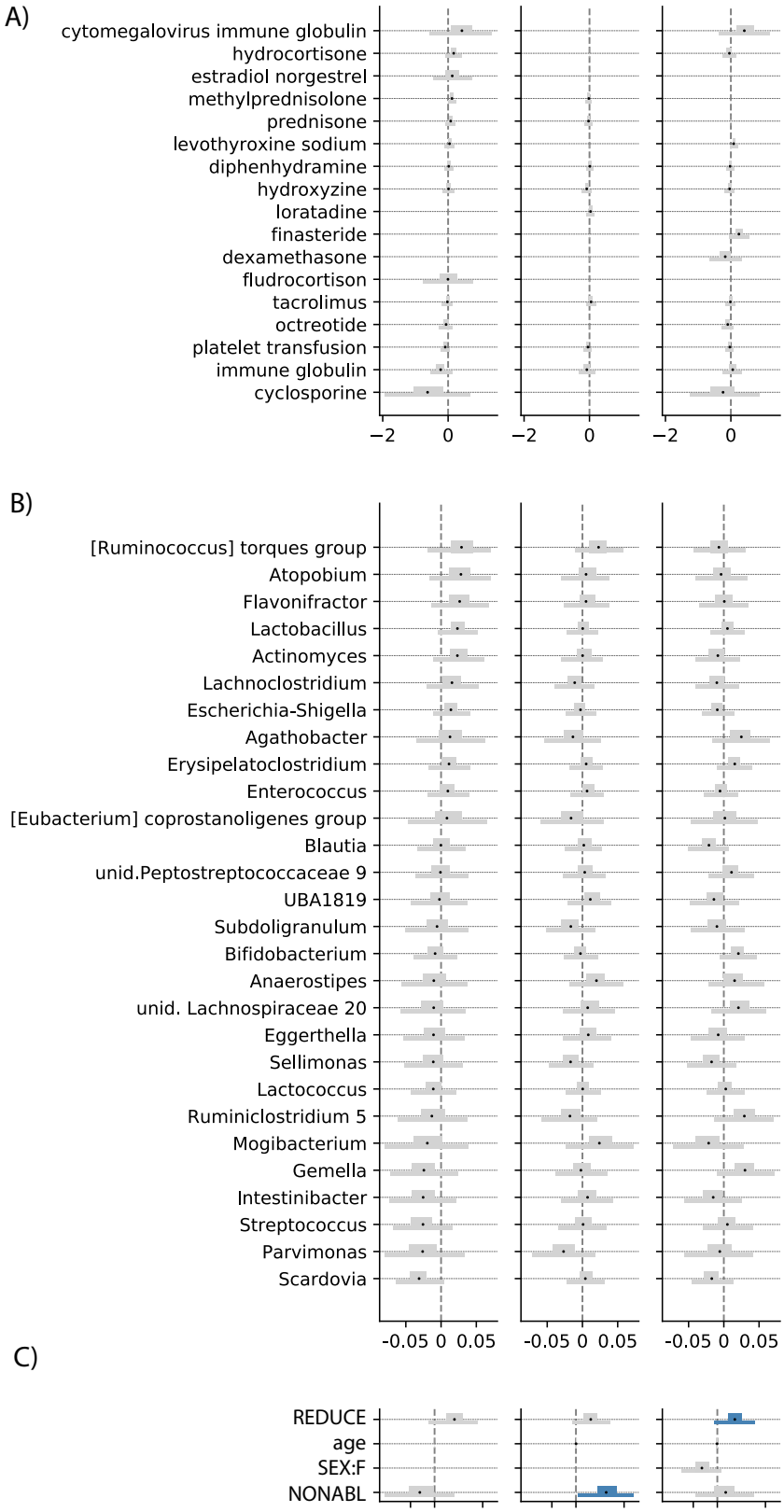


Figure S10: Additional coefficient estimates of medications (A), additional genera (B) and metadata (C) from the Bayesian regression, see also Figure 3. REDUCE: reduced-intensity conditioning regimen; NONABL: non-myeloablative conditioning regimen. F: female

Figure S11: Posterior sampling converged. Histograms of the ranked posterior draws from the model of neutrophil dynamics in PBSC patients (ranked over all chains), plotted separately for each chain (see supplementary methods), show no substantial differences between chains[71].

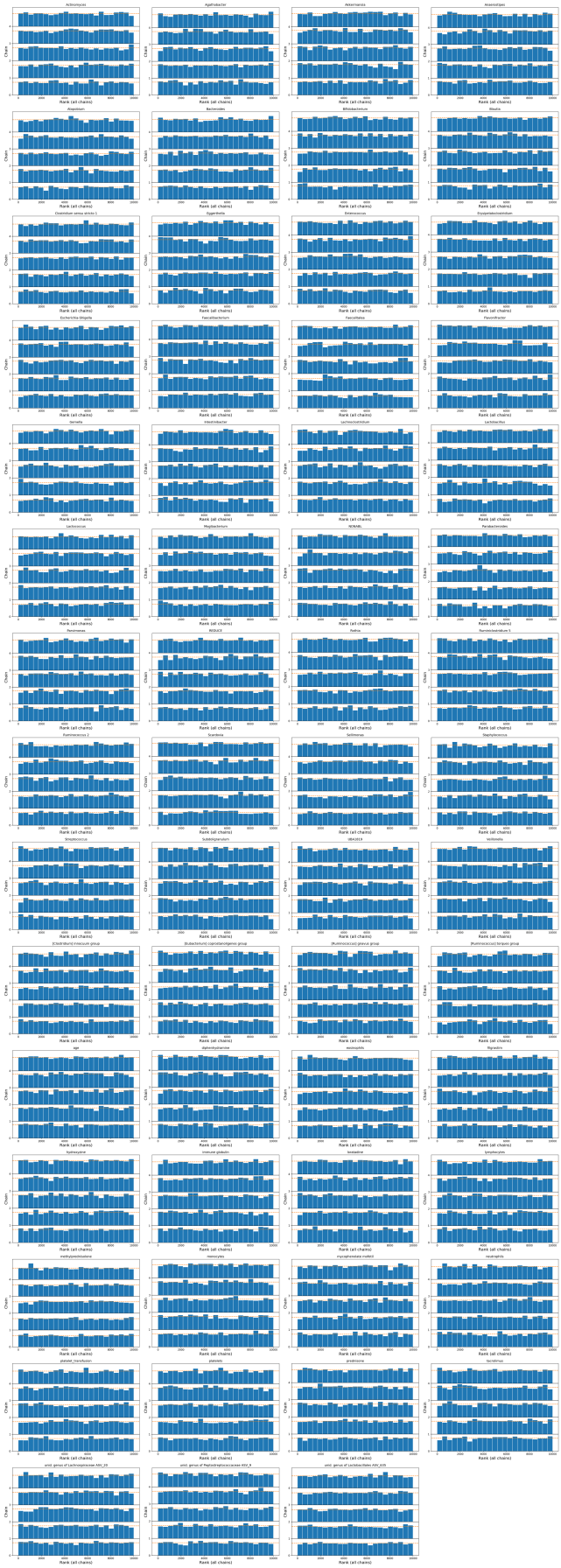


Figure S12: Posterior sampling converged. Histograms of the ranked posterior draws from the model of lymphocyte dynamics in PBSC patients (ranked over all chains), plotted separately for each chain (see supplementary methods), show no substantial differences between chains[71].

855

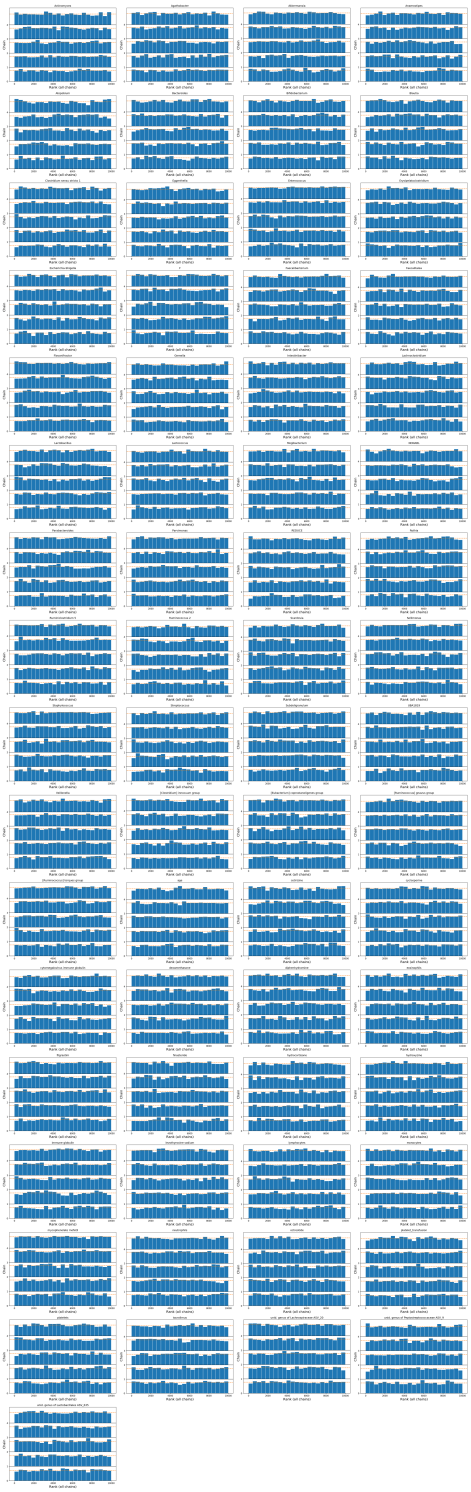


Figure S13: Posterior sampling converged. Histograms of the ranked posterior draws from the model of monocyte dynamics in PBSC patients (ranked over all chains), plotted separately for each chain (see supplementary methods), show no substantial differences between chains[71].

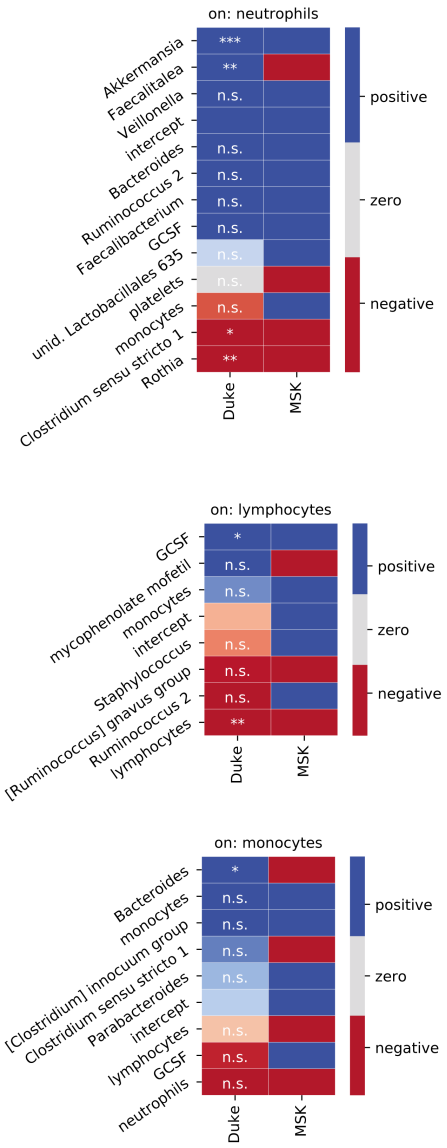


Figure S14: Validation analysis of predictors on white blood cell dynamics using data from patients treated at DukeHealth. Individual univariate regressions of microbiome and clinical predictors identified in stage 2 of our analysis on daily changes in neutrophils, lymphocytes and monocyte. Bonferroni corrected p-values: ***<0.001, **<0.01, *<0.05; p>0.05: n.s. Sign of coefficients from MSK PBSC patients for comparison.

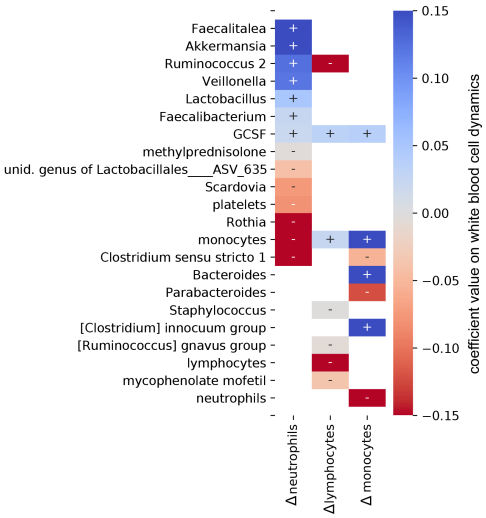


Figure S15: Validation analysis of predictors on white blood cell dynamics using data from patients treated at DukeHealth. Partial least squares regression of microbiome and clinical predictors identified in stage 2 of our analysis on daily changes in neutrophils, lymphocytes and monocyte.

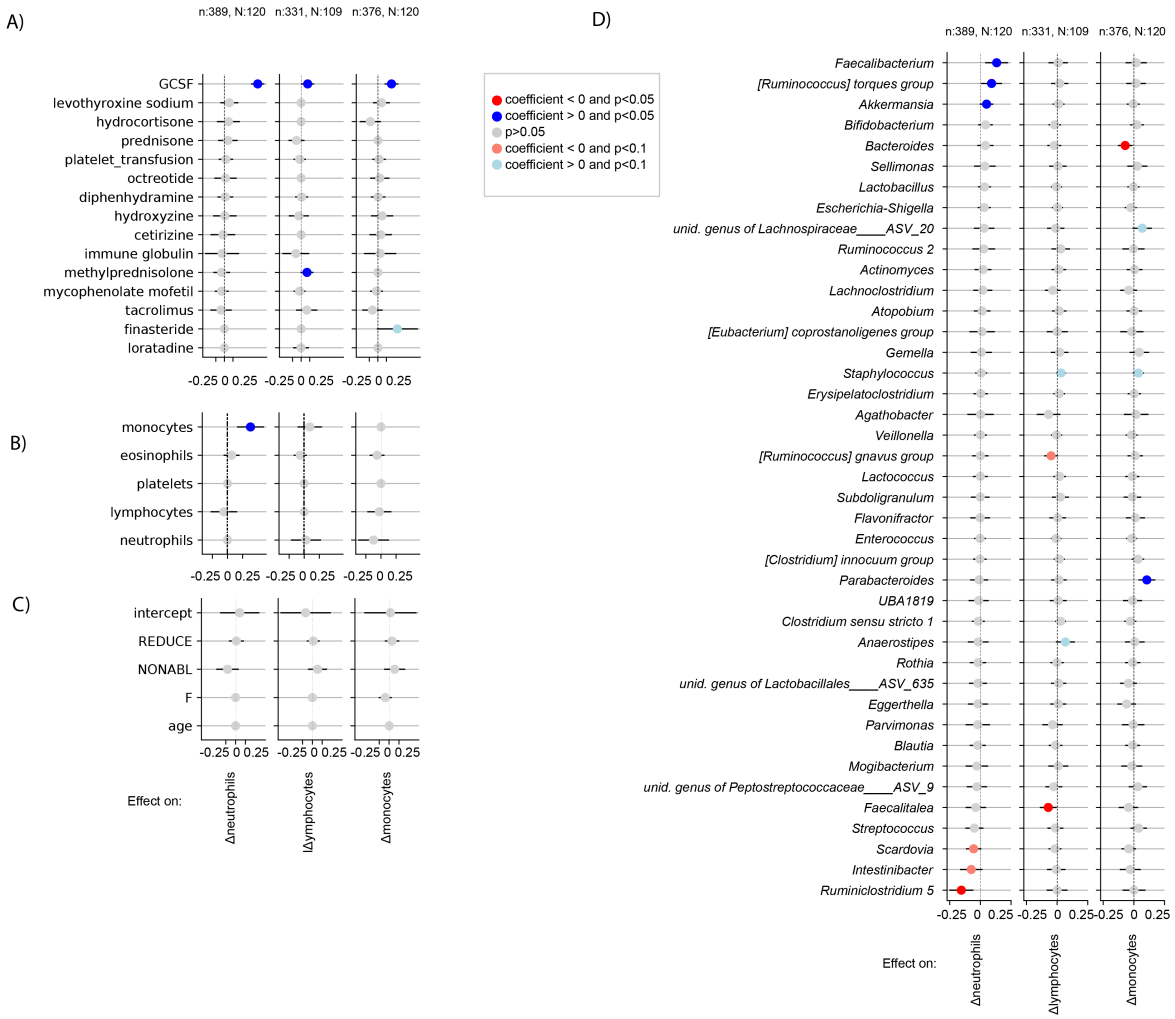


Figure S16: Validation analysis of the main model using absolute bacterial abundances as predictors instead of relative abundances in Figure 3. Results show coefficients from a least squares regression for medications (A), white blood cell feedbacks (B) metadata (C) and total genus abundances (D) of neutrophil, lymphocyte and monocyte daily log-changes. This was only possible for only a subset of the data for which we obtained absolute bacterial abundance estimates (methods), n: samples, N: patients.

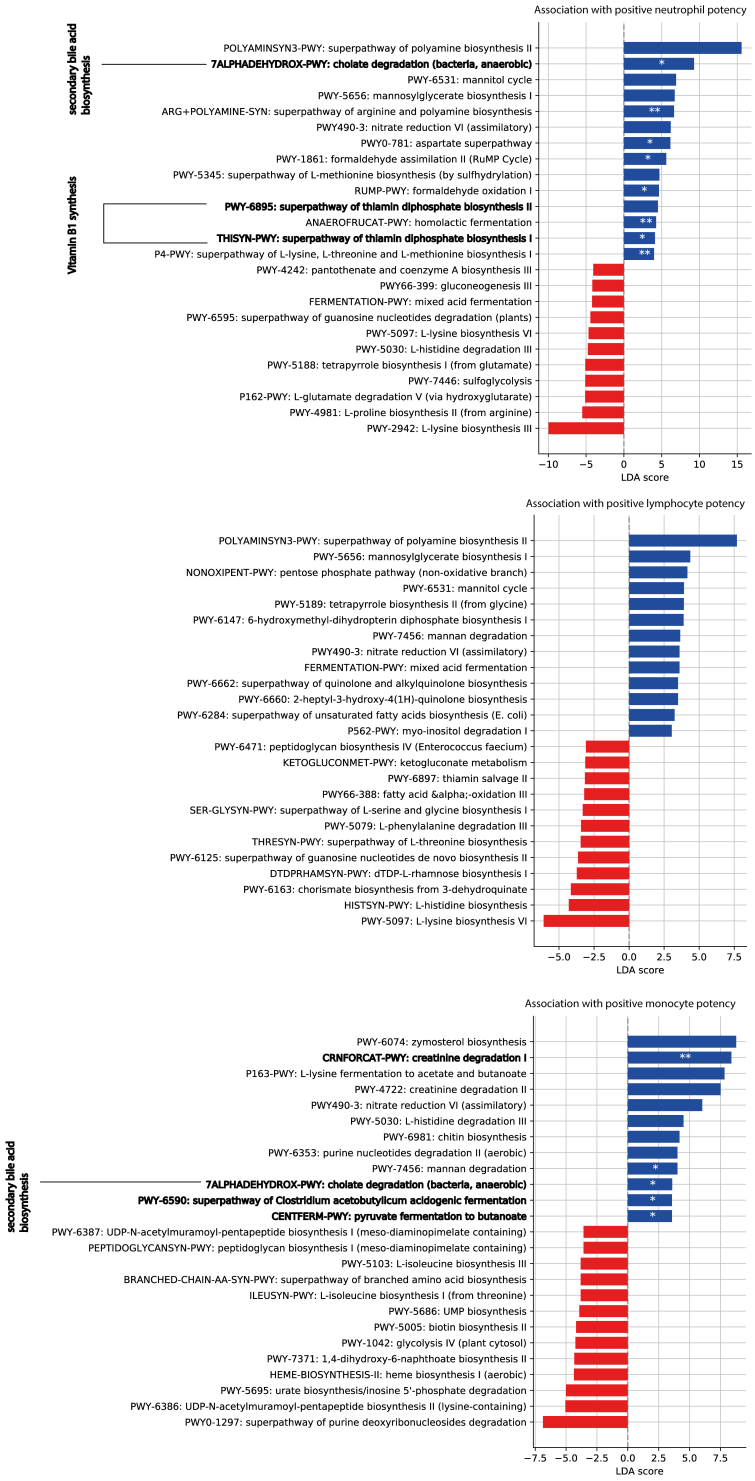


Figure S17: Functional analysis of microbiota samples. To distinguish samples predicted to increase rates of white blood cells, a microbiota potency score was calculated from posterior coefficients (Figure 3, methods) and the relative abundance of taxa in samples. Bars show linear discriminant analysis (LDA) scores of MetaCyc

pathway profiles from 124 shotgun sequenced samples that distinguished positive and negative potency samples the most (LDA-score magnitude in the 95th percentile). Highlighted pathways are discussed in the main text. For each pathway, we tested differences between positive and negative potency samples using Fisher's exact test; p-value <0.001: ***, <0.01:**, <0.05:*.

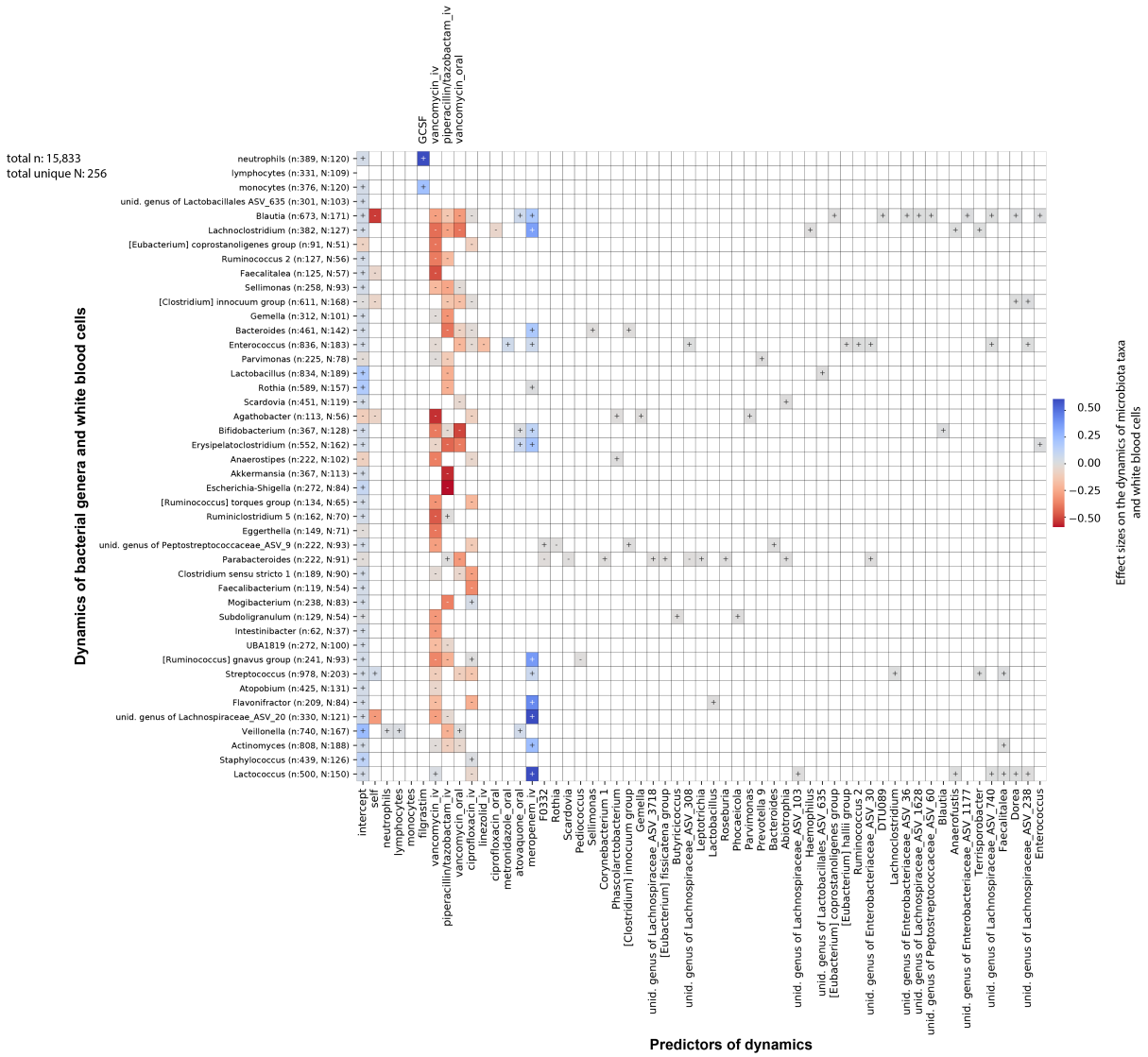


Figure S18: Jointly inferred association network between white blood cell and bacterial genus dynamics (methods). Strong regularization yields few non-zero coefficients and antibiotics dominate the dynamics.

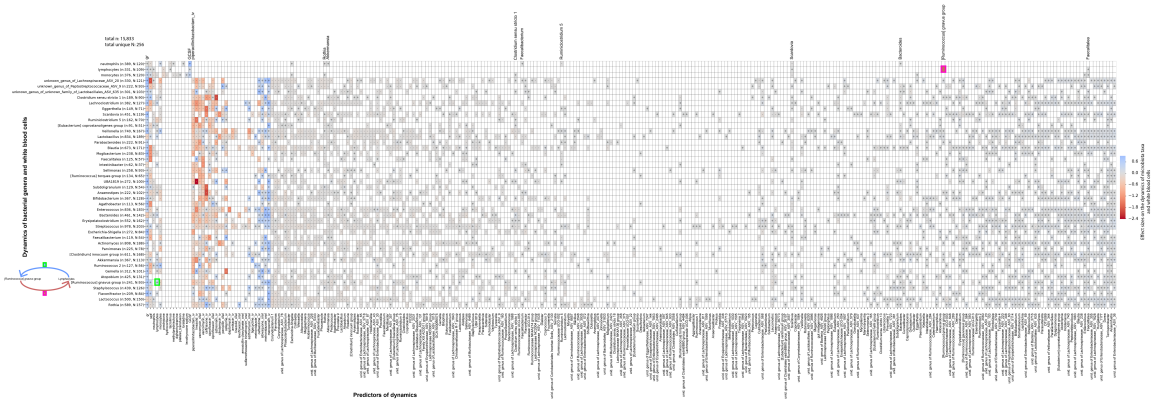


Figure S19: Jointly inferred association network between white blood cell and bacterial genus dynamics with reduced regularization (methods) indicates potential bidirectional feedbacks, e.g. between lymphocytes and *[Ruminococcus] gnavus group* (highlighter green boxes, and cartoon).

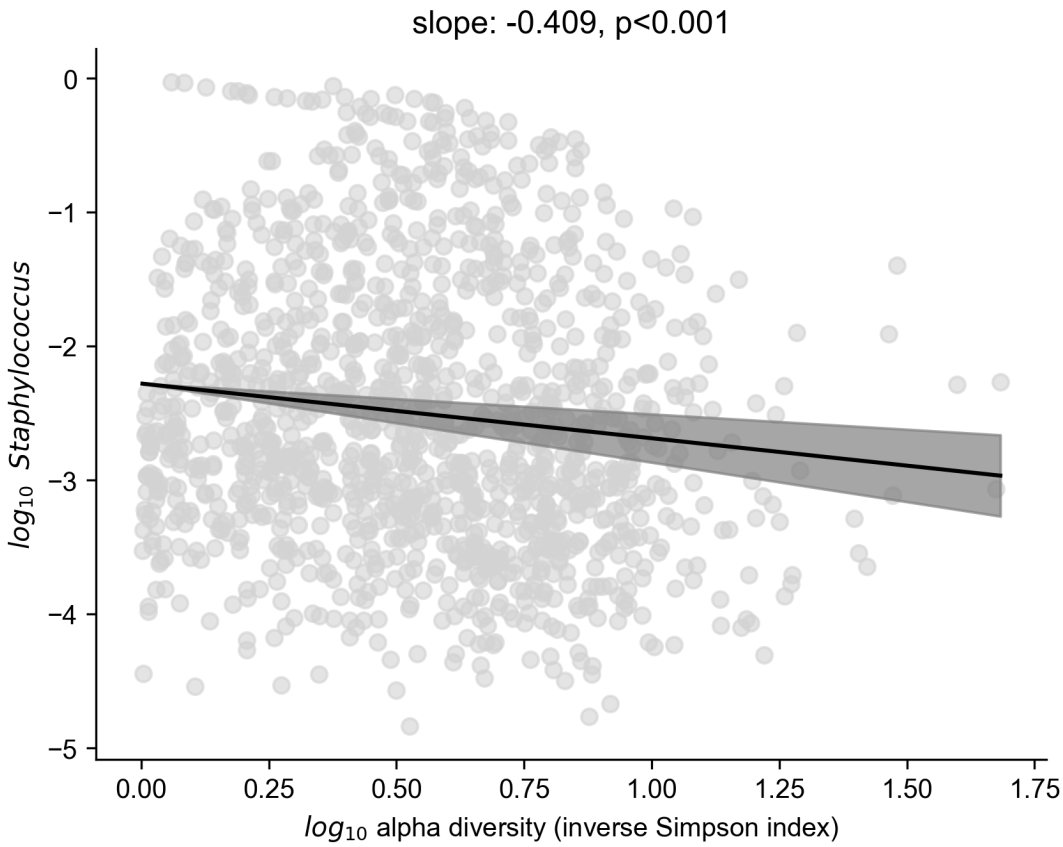


Figure S20: The relative non-zero abundance of Staphylococcus is inversely related to microbiome alpha diversity, shaded: 95% confidence intervals.

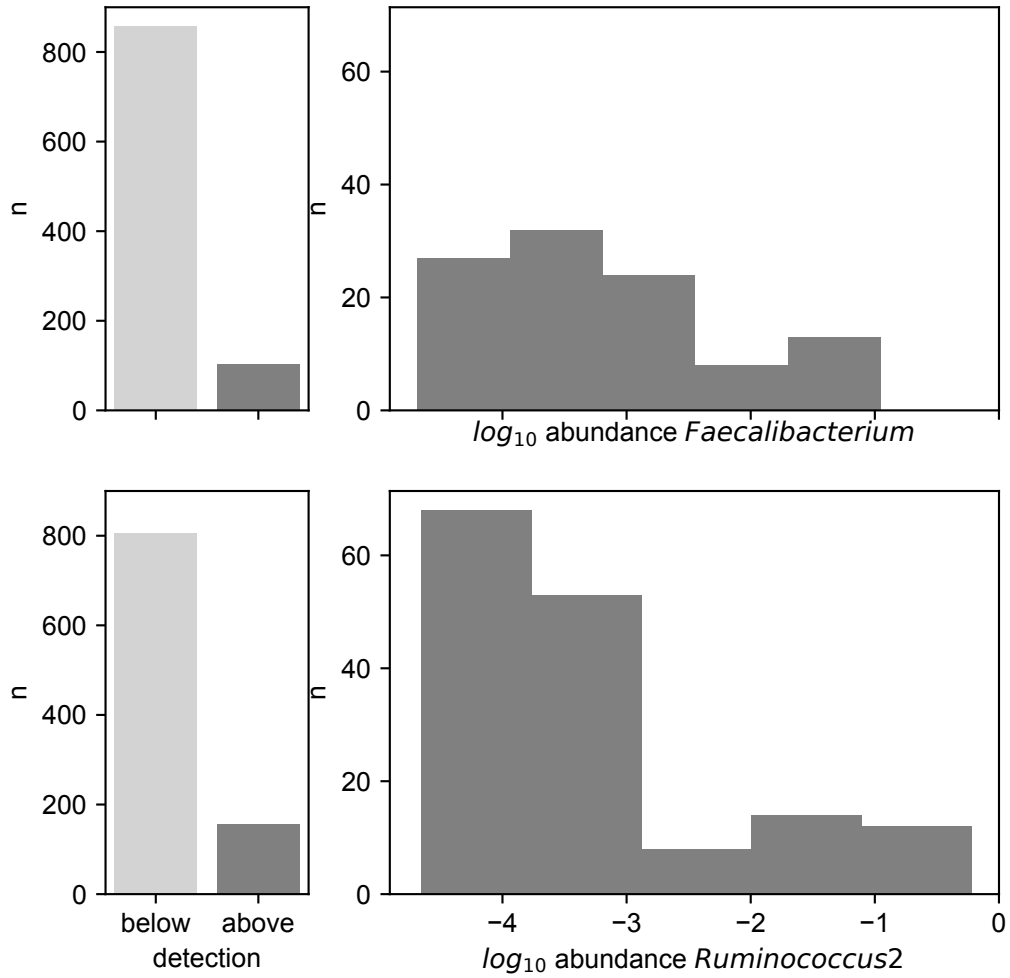


Figure S21: Abundance profiles of the two genera, *Faecalibacterium* and *Ruminococcus 2*, most strongly associated with white blood cell increase; number of times detected (left) and log_{10} abundance distribution when above detection (right).

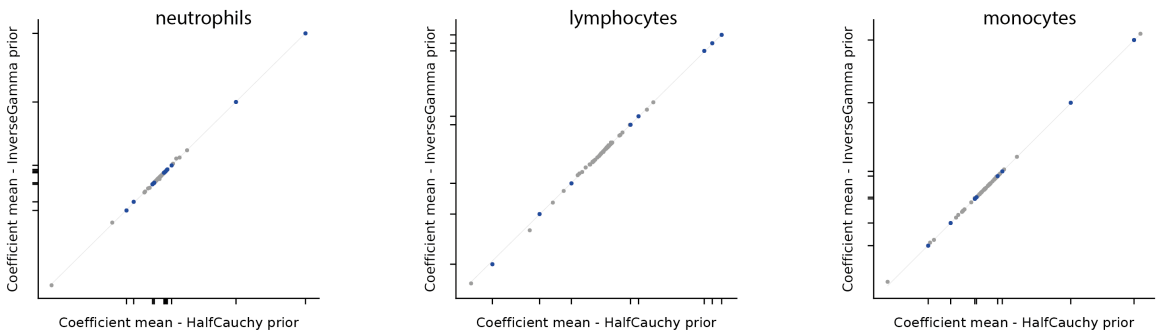


Figure S22: Posterior association coefficients do not depend on the choice of prior for σ in the main Bayesian model. Plotted are the posterior means from our main analysis against the equivalent inference with an inverse Gamma prior (alpha=1, beta=1).

942 Table S1: Data set summary and patient characteristics. HCT-graft types: TCD: T-cell depleted graft (*ex-vivo*)
943 by CD34+selection; PBSC: peripheral blood stem cells; BM: bone marrow; cord: umbilical cord blood;
944 Conditioning intensity: Bacigalupo classification, graded categories from most to least intense (ABLATIVE,
945 REDUCE, NONABL).
946

patients		2,926
HCT therapies*		3,060
blood samples	total	450,635
	between HCT-day -21 and HCT-day 183	193,396
Disease	Leukemia	1,635
	Non-Hodgkin's Lymphoma	415
	Multiple Myeloma	170
	Hodgkin's disease	88
	other	752
HCT graft type	TCD	1,106
	PBSC unmodified	959
	BM unmodified	617
	cord	378
Conditioning intensity	ABLATIVE	65%
	REDUCE	21%
	NONABL	13%
Gender	M	58%
	F	42%
Age of adults (years)	25%-tile	39
	mean	50
	75%-tile	62
Microbiome samples	total	12,633
	- from patients with blood data	10,680
	- of those, post engraftment	4,179
	- of those with daily change in WBC	2,615
	patients with microbiome sample	1,290

947 *) some patient received several HCTs
948
949
950 Table S2: Patient and HCT characteristics of 24 patients enrolled in the randomized controlled FMT trial.

	control	FMT treated
N patients	10	14
ABLATIVE	6	7
REDUCE	4	7
BM unmodified	1	3
PBSC unmodified	3	4
TCD	5	3
cord	1	4

951
952
953
954
955
956
957
958

959 Table S3: Patient and HCT characteristics of the subset of patients who donated microbiota samples.

patients		1,294
Disease distribution	Leukemia	51%
	Non-Hodgkin's Lymphoma	15%
	Multiple Myeloma	8%
	Hodgkin's disease	3%
	other	23%
HCT graft type	TCD	37%
	PBSC unmodified	38%
	BM unmodified	9%
	cord	16%
Conditioning intensity	ABLATIVE	55%
	REDUCE	34%
	NONABL	11%
Gender	M	59%
	F	41%
Age (years)	25%-tile	46
	mean	54
	75%-tile	65

960
961
962 Table S4: Patient and HCT characteristics of the subset of patients who did not donate microbiota samples.

patients		1,010
Disease distribution	Leukemia	53%
	Non-Hodgkin's Lymphoma	17%
	Multiple Myeloma	5%
	Hodgkin's disease	5%
	other	20%
HCT graft type	TCD	40%%
	PBSC unmodified	31%
	BM unmodified	17%
	cord	12%
Conditioning intensity	ABLATIVE	65%
	REDUCE	14%
	NONABL	21%
Gender	M	58%
	F	42%
Age (years)	25%-tile	36
	mean	47
	75%-tile	59

963
964
965 Table S5: Patient and HCT characteristics of the Duke University patient cohort.
966

patients		493
Disease distribution	Lymphoma	11%
	Leukemia	50%
	Non-Hodgkin's Lymphoma	4%
	Multiple Myeloma	8%

	Hodgkin's disease	4%
	other	24%
HCT graft type	TCD	0%
	PBSC unmodified	72%
	BM unmodified	11%
	cord	16%
Conditioning intensity	ABLATIVE	92%
	NONABL	7%
Gender	M	65%
	F	35%
Age (years)	25%-tile	41
	mean	49
	75%-tile	57

Data availability

The data used in our study is organized in supplementary tables (data-tables.zip), with corresponding filenames (italic):

1. *cGENUS.csv*: relative taxon abundances in fecal microbiota samples from 12,633 stool samples
2. *cHCTMETA.csv*: HCT characteristics
3. *cINFECTIONS.csv*: positive blood culture results
4. *cMISAMPLES.csv*: NCBI SRA accession number, diversity (inverse Simpson index), total 16S (where available), stool consistency for each fecal microbiota sample
5. *cMED.csv*: medication data
6. *cPIDMETA.csv*: anonymized patient demographics
7. *cWBC.csv*: absolute counts of neutrophils, lymphocytes, monocytes, eosinophils, and platelets with indication if included in analyses
8. *cDUKE__GENUS.csv*: relative taxon abundances in fecal microbiota samples from 12,633 stool samples
9. *cDUKE__WBC.csv*: absolute counts of neutrophils, lymphocytes, monocytes, eosinophils, and platelets with indication if included in analyses
10. *cDUKE__MED.csv*: medication data
11. *cFMT_analysis*.csv: convenience table for Figure 2

Code availability

The relevant scripts for stage 1, stage 2, and the model assessing the effect of FMT on white blood cell counts are on github:https://github.com/jsevo/wbcdynamics_microbiome.