The gut microbiota influences how circulating immune cells in humans change from one day to the next.

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@gmail.com, xavierj@mskcc.org

ABSTRACT

The gut microbiota influences the 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 circulatory immune cells remains limited, particularly in humans where a lack of manipulative experiments makes inference challenging. Here we overcome this challenge by studying hundreds of hospitalized—and closely monitored—bone marrow transplantation patients as they recover from chemotherapy-induced immune ablation. This aggressive treatment causes large shifts in both circulatory immune cell and microbiota populations, allowing the relationships between the two to be studied simultaneously over time with unprecedented resolution. We analyzed daily changes in white blood cell counts from 2,235 patients, and 10,680 longitudinal fecal microbiota samples to identify bacterial genera consistently associated with those changes. Bayesian inference and validation across different patient cohorts revealed consistent associations between intestinal bacteria and peripheral immune cell dynamics in the context of immunomodulatory medications, clinical metadata and homeostatic feedbacks between peripheral immune cells. The quantification of validated microbiota associations enabled us to contrast the potency of fermentatively active, obligate anaerobic bacteria with that of medications with known immunomodulatory mechanism, and this way estimate the microbiota potential to alter peripheral immune cell dynamics directly in patients. Our analysis establishes and quantifies the link between the intestinal microbiota and immune cell dynamics in humans, with implications for microbiota-driven modulation of immunity and immunotherapies that rely on circulatory immune cells.

43 **MAIN TEXT**

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

57 To overcome this limitation, we here quantified the associations between the microbiome and
58 day-by-day changes in peripheral immune cell counts. We investigated the dynamics of immune
59 reconstitution after allogeneic hematopoietic cell therapy (HCT) within the universe of all 2,926
60 patients who underwent HCT at Memorial Sloan Kettering for various hematological malignancies,
61 including leukemia, between 2003 and 2019 (**Figure 1A, Table S1**). The conditioning regimen of
62 radiation and chemotherapy administered to HCT patients is the most severe perturbation to the
63 immune system deliberately performed in humans and thus offers a unique opportunity to investigate
64 dynamic links between gut microbiota and immune system directly in humans. Conditioning depletes
65 circulating white blood cell counts (**Figure 1A**). This can lead to prolonged periods of neutropenia
66 (<500 neutrophils per μl blood) followed by reconstitution after transplanted stem cells have matured
67 sufficiently to release granulocytes from the bone marrow (neutrophil engraftment is defined as 3
68 consecutive days with >500 neutrophils per μl blood, **Figure 1A-C**). The blood of each patient is
69 carefully monitored throughout this recovery, and medications are administered to modulate the

70 immune cell dynamics, including granulocyte-colony stimulating factor (G-CSF) to increase

71 neutrophil counts, and immunosuppressants such as mycophenolate mofetil or tacrolimus to prevent

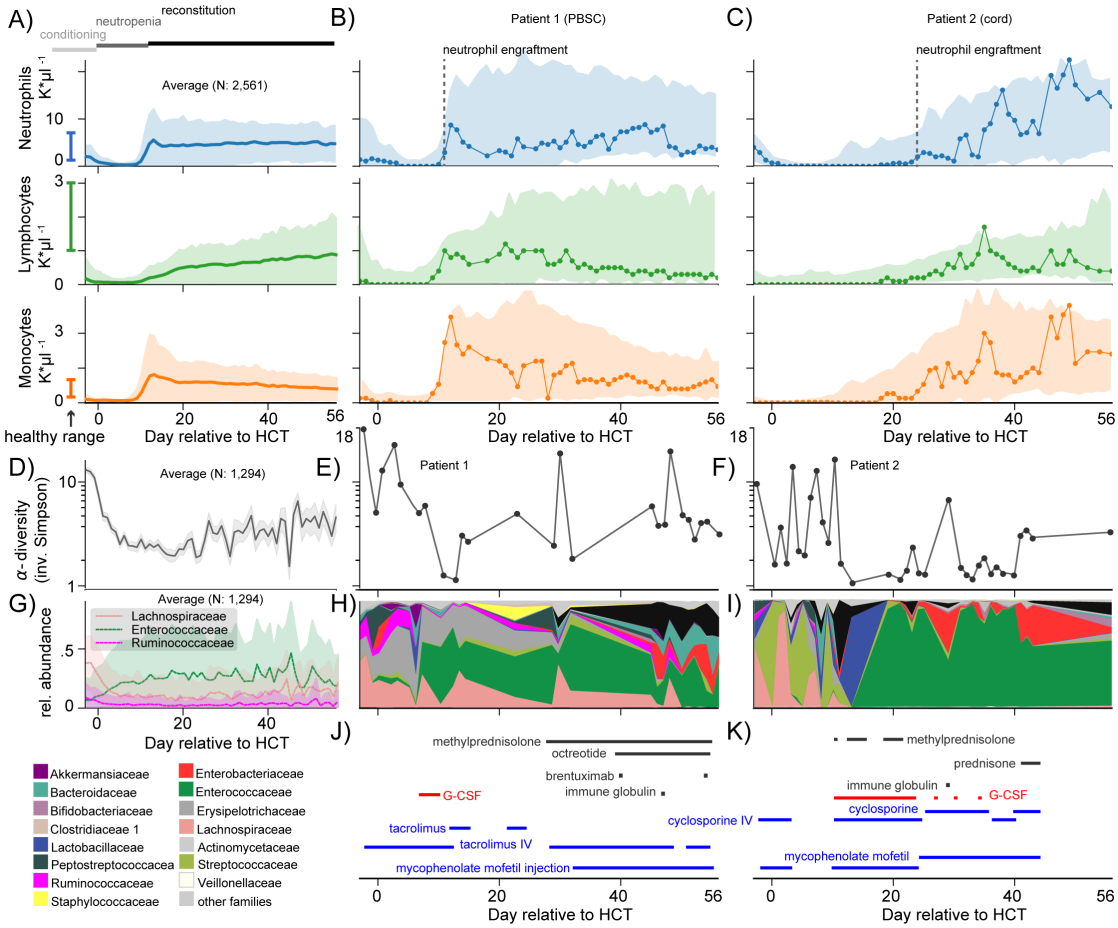


Figure 1: Immune reconstitution and microbiome dynamics after allogeneic hematopoietic cell transplantation (HCT). A) Three major periods of HCT—immunoablation during chemotherapeutic conditioning before HCT, defined as day 0, post-HCT neutropenia, and reconstitution following neutrophil engraftment—lead to recovery trajectories with large variability between patients. Shown are the average 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 (shaded region: mean±1 standard deviation of all PBSC patients), and patient 2 who received a graft of umbilical cord blood (shaded region: mean±1 standard deviation of all cord patients). Fecal samples collected during HCT reveal the loss of microbial diversity reported previously[22,23] (D-F) and commensal families (G-I), often replaced by Enterococcaceae domination. J,K). Administration of immunomodulatory medications for the two example patients.

72 complications such as graft-vs-host disease (**Figure 1 J,K**). To investigate if the composition of the

73 gut microbiota affects the dynamics of circulating white blood cells in our patients, we analyzed

74 detailed blood and clinical metadata of our patients between 3 days before HCT and until 100 days

75 post neutrophil engraftment (excluding pediatric patients, and other exclusion criteria: N=2,235,

76 supplementary methods). During this period patients are monitored carefully, and our analysis

77 included over 140,000 host phenotype measurements in the form of complete blood counts which

78 quantify the most abundant white blood cells—neutrophils, lymphocytes, monocytes, eosinophils—as
79 well as platelet counts (**Figure 1, S1**). We started collecting patients' fecal microbiota data in 2009[21],
80 from which we obtained 10,680 high frequency, longitudinal microbiota compositions from 1,294 of
81 our HCT patients.

82 We have shown previously that HCT patients lose gut microbiota biodiversity and commensal
83 microbial families during their treatment (**Figure 1D-I**)[22,23], and, furthermore, recently associated
84 impaired white blood cell recovery after bone-marrow transplantation in mice with a depleted
85 intestinal flora[24] 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**). This variation is illustrated by the
88 distinct trajectories of patient 1 who received a graft of peripheral blood stem cells (PBSC), retained
89 high microbiota diversity and engrafted earlier (**Figure 1B,E,H**), and patient 2 who received a graft of
90 umbilical cord blood (cord), lost microbiota diversity and engrafted later (**Figure 1C,F,I**). Low
91 microbiota diversity at the time of neutrophil engraftment has been associated with 5-fold increased
92 transplant-related mortality[25], suggesting that that the joint recovery of the microbiota and white blood
93 cells in circulation is critical for clinical outcomes.

94 To detect a causal link between the microbiota and white blood cell recovery dynamics, we
95 first used data from a recent prospective randomized trial of autologous fecal microbiota
96 transplantation (auto-FMT), which represents a deliberate microbiota manipulation in our patients[22]
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 and monocyte counts post-engraftment (i.e. when the transplanted
101 hematopoietic cells begin producing new white blood cells, **Figure 2A**). FMT procedures were
102 conducted at variable time points relative to neutrophil engraftment, but overall, we observed higher
103 counts of each white blood cell type in patients who received an auto-FMT during the first 100 days
104 post neutrophil engraftment ($p<0.01$, **Figure 2B,S3-5**).

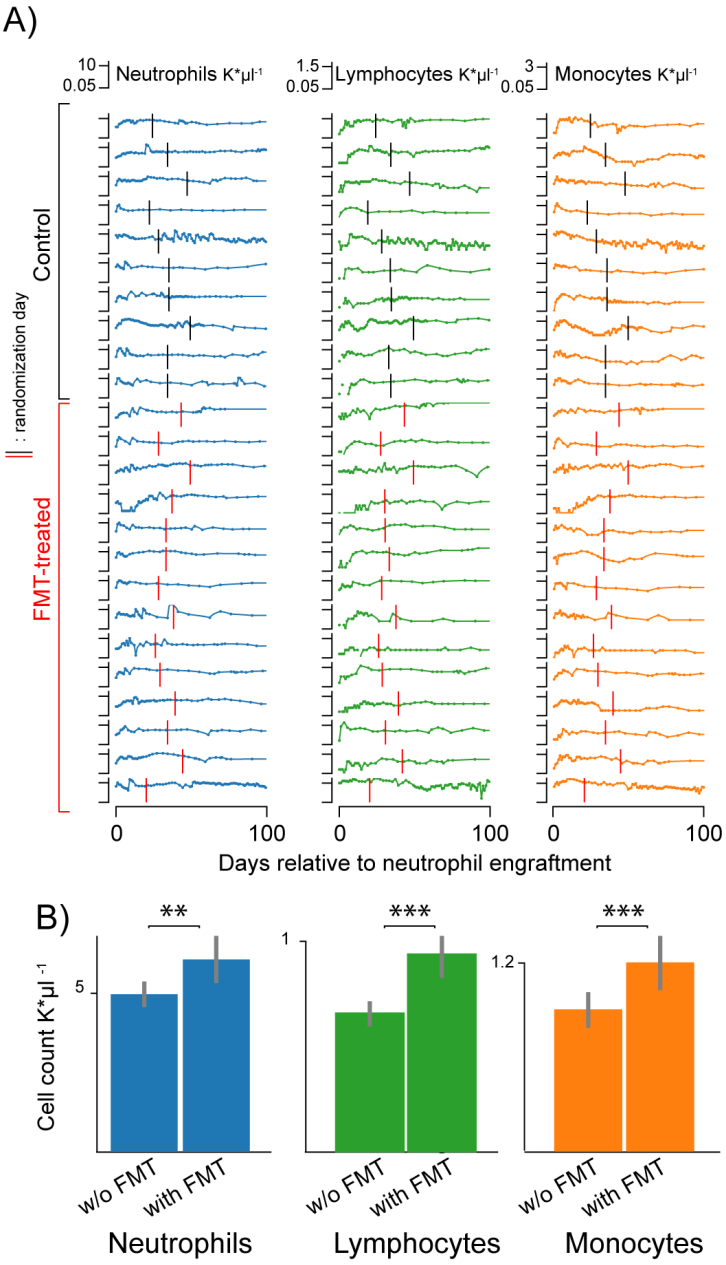


Figure 2: FMT-treated patients showed higher counts of white blood cells. A) Absolute counts of neutrophils (blue), lymphocytes (green) and monocytes (orange) in 24 patients enrolled in a randomized controlled trial to receive an autologous fecal microbiota transplant post-neutrophil engraftment (FMT treatment: red vertical line, control: black vertical line). B) Results from a linear mixed effects model with random effects per patient and per day relative to neutrophil engraftment show elevated neutrophil, lymphocyte and monocyte counts in patients receiving auto-FMT after their treatment as compared to control patients, bars and confidence intervals for the averages of observed white blood cell counts without FMT and post-FMT (***: $p<0.001$, **:$p<0.01$).

The increased white blood cell counts in patients receiving auto-FMT indicated that this deliberate microbiota manipulation affected circulatory immune cell dynamics, supporting a causal role of the microbiota. The increase could be due to the reconstitution of a complex microbiota that

118 we saw in these patients[22] and the associated metabolic capabilities[6,7,24], or it could be a systemic
119 response to a severe therapy which introduced billions of intestinal organisms at once via an enema
120 (no enema was administered to control patients[22]). Moreover, while our analysis used a mixed-effects
121 model which accounted for patient-specific HCT treatments in this randomized patient cohort, chance
122 differences in extrinsic factors such as different immunomodulatory drug exposures may have
123 affected this result due to the small cohort size. Therefore, we next investigated the link between the
124 gut microbiota composition and the dynamics of white blood cell recovery in our large observational
125 cohort of HCT patients.

126 Homeostasis of circulatory white blood cell counts is a complex, dynamic process:
127 neutrophils, lymphocytes and monocytes are formed and released into the blood *de novo* by
128 differentiation of hematopoietic progenitor cells from the bone marrow, and they can be mobilized
129 from thymus and lymph nodes (lymphocytes), spleen, liver and lungs (neutrophils); they can also
130 migrate from the blood to other tissues when needed[26]. These processes are dynamic sources and
131 sinks of circulatory white blood cells, and they can be modulated by drugs administered to patients
132 receiving HCT. To identify factors associated with these dynamic source- and sink-processes—
133 including the microbiota—we developed a two-stage approach analyzing the changes of white blood
134 cell counts between two days (i.e. the rates of cell count increases and decreases). Stage 1 served as a
135 feature selection stage where we used data of 1,096 patients (after filtering for qualifying samples and
136 applying exclusion criteria, see supplementary methods) *without* available microbiome information to
137 identify associations between clinical metadata, including immunomodulatory medications (methods),
138 and changes of white blood cell counts from one day to the next (**Figure 3A**). Stage 2 was performed
139 on data from an independent cohort of 841 different patients at MSK from whom concurrent
140 microbiome samples were available. Stage 2, our main analysis, sought to reveal associations between
141 the abundance of microbial taxa and the daily changes in blood cell counts in context of
142 immunomodulatory medications, additional clinical metadata and the current state of the blood itself.

143 In stage 1 we calculated the changes in neutrophil, lymphocyte and monocyte counts during
144 patients' recovery from >20,000 pairs of post-engraftment blood samples separated by a single day
145 (**Figure 3B, S6-8,** supplementary methods). Using a cross-validated feature selection approach, we

146 detected medications and HCT treatment parameters that were associated with different rates of
147 change in neutrophil, lymphocyte and monocyte counts, including, as expected, G-CSF and the graft
148 stem cell sources (**Figure S6-S8, Table S1**).

149 During stage 2 we sought to identify associations between bacterial taxa of the gut microbiota
150 and the dynamics of immune cells in circulation. For this, we performed Bayesian inferences using
151 data from different sets of patients *with* available microbiome samples. Stage 1 had identified—as
152 expected—that stem cell graft sources are associated with immune reconstitution kinetics (e.g. cord
153 on average slower compared with PBSC[27]), and we therefore stratified our patients by graft source in
154 stage 2. The model of stage 2 now included microbial genera as additional predictors of observed
155 changes in white blood cell counts to the medications selected in stage 1, clinical features
156 (conditioning intensity, age, sex), and the current state of the blood in the form of counts of
157 neutrophils, lymphocytes, monocytes, eosinophils, and platelets. The data comprised 841 patients, but
158 approximately 60% of the stool samples paired with a daily change in white blood cell counts were
159 taken before neutrophil engraftment (**Figure 3B, Table S1,** supplementary methods), i.e. when blood
160 cells counts were zero. In total, the analyzed 2,615 post-engraftment observations of changes in
161 neutrophil counts during immune reconstitution (lymphocytes: 2,006, monocytes: 2,534) with paired
162 stool samples which provided a large sample of observed white blood cell dynamics (**Figure 3B,**
163 **Table S3,S4**). We first focused on the data from the largest (**Figure 3C**) cohort—patients who
164 received a PBSC graft—and withheld the other cohorts (bone marrow, BM; T-cell depleted graft
165 (*ex-vivo*) by CD34+selection, TCD; cord blood, cord) to use as independent validation cohorts. For
166 this validation, we analyzed TCD, BM, and cord patients' data in the same way as PBSC patients'
167 data and compared the resulting posterior coefficient distributions (methods). We assigned
168 coefficients obtained from the PBSC cohort a validation score (v-score) between 0 and 3, representing
169 the number of times that the focal coefficient was validated in the other cohorts; but, conservatively,
170 the score was always set to zero if we observed counter-evidence among any of the other data sets, i.e.
171 evidence that coefficients had the opposite sign. Finally, we analyzed data from another patient cohort
172 consisting of 493 bone marrow transplantation recipients treated at Duke University including 9,603

173 blood samples and a total of 629 microbiota samples from 218 patients, albeit with lower sampling
174 density, and we used the results from this analysis for further validation.

175 Notably, as a verification of our approach, we detected associations between the
176 administration of immunomodulators and increased or decreased rates of immune cell count changes
177 consistent with the known biological mechanism of these medications (**Figure 3C,S9A**). The
178 strongest association we found across all predictors is consistent with the well-known neutrophil-
179 increasing effect of G-CSF[28]; G-CSF administration—used to accelerate recovery from
180 chemotherapy-induced neutropenia[28]—was associated with a 140% increase in the rate of neutrophil
181 changes from one day to the next ([114%, 170%], 95 percent probability density interval [HPDI95]).
182 This finding was observed in all MSK validation data sets (v-score=3, **Figure 3D)**, as well as among
183 Duke University patients (**Figure S10**). We furthermore found a G-CSF-associated increase of 43%
184 ([30%, 58%]HPDI95, v-score=3) in monocyte rates, and, although smaller, in lymphocyte rates (16%,
185 [5%, 27%]HPDI95, v-score=2). Both neutrophil and lymphocyte rates decreased following the
186 exposure to immunosuppressive medications (cetirizine -18%, [-35%,+5%]HPDI95, mycophenolate
187 mofetil -8% [15%,+1%]HPDI95, respectively). Finally, less intensive chemotherapeutic conditioning
188 regimens (reduced intensity and non-ablative conditioning) were associated with higher average
189 lymphocyte count increases during immune reconstitution (**Figure S9C**)

190 Beyond associating medications in agreement with their known biological mechanism, our
191 analysis detected associations between the current count of white blood cells and their rate of change:
192 a negative association among lymphocytes, negative associations between counts of neutrophils and
193 lymphocytes with the rates of monocytes, and a negative association between the counts of platelets
194 and the rates of neutrophils (**Figure 3E**). Conversely, we found positive associations between
195 monocytes and the rates of each of the investigated white blood cell subsets. These associations
196 suggest a feedback network (**Figure 3F**) of circulatory white blood cell counts, reflecting potentially
197 complex homeostatic regulation mechanisms of immune cell compositions in the blood[26]. More
198 importantly, the associations quantified for medications and potential homeostatic feedbacks provided
199 a benchmark against which we could compare associations from gut microbial taxa.

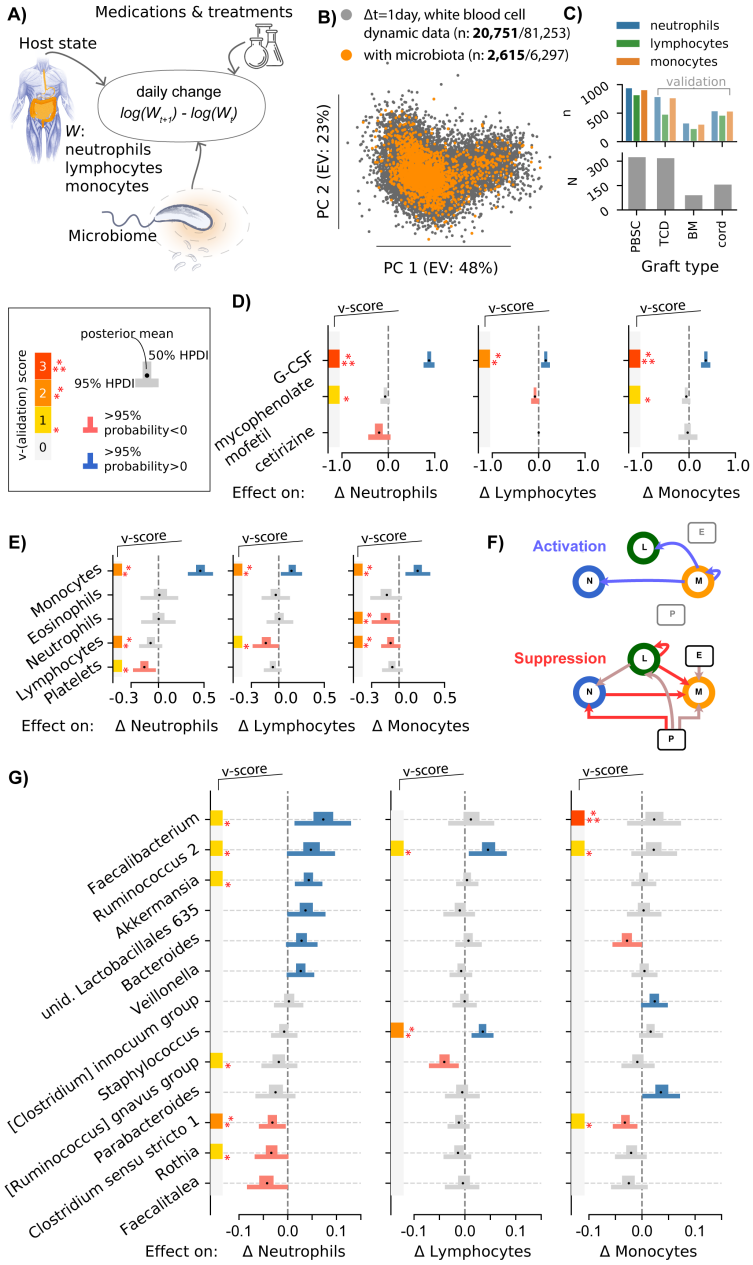


Figure 3: Bayesian inference conducted in context of the host state (white blood cell and platelet counts) and clinical variables including immunomodulatory medications reveals the microbiota potential to affect daily changes in peripheral white blood cell counts. A) Cartoon of the model: observed changes in white blood cell counts between two consecutive days are associated with the current state of the host in the form of blood cell counts in circulation, the administration of immunomodulatory medications, patient and clinical metadata, and the state of the microbiome. B) Visualization of the dynamic white blood cell data; scatter plot of the principal components (PC) of observed daily changes of neutrophils, lymphocytes and monocytes without (grey, see **Figures S6-8**) and with (orange) concurrent longitudinal microbiome data (bold: post engraftment sample counts). C) PBSC patients (N=312) provided the most blood samples with simultaneous microbiome data (n=995) relative to TCD, BM and cord patients, who were used as validation data sets. D-G) PBSC patient data inference results; bars show the posterior parameter estimate distributions (thin: 95% highest posterior density intervals, HPDI95, thick: HPDI50) for the jointly inferred associations between treatments (D), white blood cells counts (E, revealing a potential homeostatic regulation network, F), and log-fecal microbiota genus abundances (G) with the observed daily changes in neutrophils, lymphocytes and monocytes. v-score: number of validation cohort confirming associations, *always* set to zero if invalidated in any of the TCD, BM, or cord cohorts.

218 We identified microbial genera that consistently associated with increases or decreases in
219 white blood cell counts by first using data from the PBSC patients and then validating the associations
220 in the other cohorts (**Figure 3G**). Higher abundances of *Faecalibacterium* (+8%, [+1%,
221 +14%]HPDI95 per $\log_{10}$), *Ruminococcus 2* (+5%, [-0%, +10%]HPDI95) and *Akkermansia* (+4%,
222 [1%, 7%]HPDI95) were associated with greater neutrophil increases, whereas increased *Rothia* (-3%,
223 [-7%, +0%]HPDI95), and *Clostridium sensu stricto 1* (-3%, [-6%, 0%]HPDI95) abundances
224 associated with reduced neutrophil rates. These results were validated in univariate analyses
225 conducted in the Duke University cohort (**Figure S10**). *Staphylococcus* was positively associated with
226 lymphocyte rates (+4%, [+1%, +6%]HPDI95) and, again, *Ruminococcus 2* was also associated with
227 faster lymphocyte increases (+5%, [+1%, +9%]HPDI95). Both *Faecalibacterium* as well as
228 *Ruminococcus 2* also associated with increases in monocytes, and while this association was validated
229 in other cohorts (v-score 3 and 1, respectively), there was higher uncertainty of the association
230 estimate (HPDI50>0). Again, *Clostridium sensu strictu 1* (-3% [-5%,-1%]HPDI95) associated
231 consistently with decreased rates of monocytes. The associations we identified—and validated in
232 other cohorts—between microbial taxa in the gut and daily changes in white blood cell counts support
233 the idea that hematopoiesis and mobilization respond to the composition of the gut microbiome,
234 influencing systemic immunity[29].
235 Most of the taxa that strongly associated with white blood cell dynamics were obligate
236 anaerobes. *Rothia*, was a notable exception: this aerobic genus is typically found in the oral cavity[30]
237 but can become an opportunistic pathogen in immunosuppressed patients and is not known to provide
238 metabolic functions to the host[31]. Obligate anaerobes, on the other hand, produce metabolic
239 byproducts such as short-chain fatty acids[32,33] and bacterial cell-wall molecules[1,34,35] that modulate
240 immune responses and granulopoiesis[6]. Nutritional support from the intestinal microbiota improved
241 hematopoietic reconstitution in a mouse model[24]. To identify a similar association in our patients, we
242 estimated a microbiota potency by multiplying the $\log_{10}$-abundances of microbial genera in a sample
243 with their corresponding posterior coefficients. We analyzed shotgun metagenomics sequences from
244 124 of the samples and observed that samples with positive microbiota potency were associated with
245 enrichment in cholate degradation and vitamin-B1 synthesis related pathways, as well as butanoate

246 formation (**Figure S11**). Our findings are in line with evolutionary theory[36] that essential but broadly
247 available microbial traits such as the production of B vitamins, secondary bile acid metabolism, and
248 fermentation to short-chain fatty acids[37] could be co-opted by the host's immune system as part of the
249 homeostatic interplay between immune system and a complex microbiota[38,39]. For example,
250 *Ruminococcus 2* is a genus that contains *R. bromii*, a keystone species necessary for the efficient
251 release of energy from complex starch in the normal diet[40]. Reassuringly, the genera
252 *Faecalibacterium*, *Ruminococcus 2* and *Akkermansia* that we associated with faster rates of white
253 blood cells were among those best reconstituted by auto-FMT[22], potentially explaining why we found
254 higher counts of neutrophils, monocytes and lymphocytes in patients who received the auto-FMT
255 (**Figure 2B**).
256 The associations we reveal are interpretable as potential effectors on sources and sinks of
257 white blood cell counts in circulation. Intestinal bacteria may affect white blood cell counts in
258 circulation by influencing either their sources in the bone marrow or their cytokine profiles[41] and
259 proliferation rates in the blood, their sinks in different organs, or both. Our analysis cannot distinguish
260 the effects of microbial communities on *de novo* hematopoiesis, for example, in isolation from their
261 impact on other sources and sinks. Unlike the plausible role of obligate anaerobe fermenters in
262 augmenting hematopoiesis via nutritional support[24], the positive association detected between
263 *Staphylococcus* and lymphocyte dynamics could instead result from reduced migration of T cells out
264 of circulation into the gut epithelium[42], especially since high abundances of *Staphylococcus* are
265 associated with low gut microbiota diversity ($p<0.001$**, Figure S12**), which indicates a depleted
266 microbiota.
267 Nevertheless, our approach allows us to leverage the chronology of events and assess
268 "mathematical causality"[43]. Of course, due to the observational nature of these data there are risks of
269 confounding that could explain some of the associations found, but the close temporal
270 correspondence[43] between microbiota and blood cell dynamics, and the validation across cohorts
271 reduces the number of plausible confounders. Our results, therefore, quite naturally suggest candidate
272 microbial taxa to manipulate if we seek to steer complex hematopoietic dynamics and utilize the
273 microbiota as an immunomodulatory component of the human body. *Faecalibacterium*,

274 *Ruminococcus 2*, and *Akkermansia* most strongly associated with increases in white blood cell counts
275 from one day to the next. Intriguingly, members of *Faecalibacterium* and *Ruminococcus* in one
276 study[10], and *Akkermansia* in another[9], were identified as enriched in patients with better responses to
277 anti–PD-1 immunotherapy, which suggests a disagreement between the two studies[44]. Our results
278 consolidate these findings, and, furthermore, allow us to quantitatively relate the potency of
279 manipulating these intestinal commensals to that of immunomodulatory drugs. While these genera are
280 common in the gut microbiota of healthy people[17], the abundance of each genus can drop below
281 detection in our patients during the intestinal damage related to HCT[23]. Therefore, realistic ranges of
282 3-5 orders of magnitude in bacterial abundances (**Figure S13**) can yield effect sizes similar to that of
283 homeostatic feedbacks between white blood cells and several immunomodulatory medications (e.g. a
284 change in *Ruminococcus 2* from below detection to 1% relative abundance associated with a +67%
285 and +63% increases in neutrophil and lymphocyte rates, respectively). The relatively small effect
286 sizes of intestinal microbes, in concert and over time, could therefore have significant impact on
287 systemic immunity even in individuals with less severely injured microbiomes.
288 In sum, our work links the gut microbiota to the dynamics of the human immune system via
289 peripheral white blood cells populations. We were able to infer and quantify this link because we
290 aggregated an unprecedented resource: vast high-frequency longitudinal data of blood cell counts in
291 conjunction with microbiota samples and detailed clinical metadata. The links between microbiota
292 components and immune cell dynamics could be used in the future to improve immunotherapy and
293 help identify microbiota treatments to cure inflammatory diseases[9,10,45–49].
294

295

296

297 **Bibliography**

298

299 1. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory
300 molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,**
301 107–118 (2005).
302 2. Gomez de Agüero, M. *et al.* The maternal microbiota drives early postnatal innate
303 immune development. *Science* **351,** 1296–1302 (2016).
304 3. Sonnenberg, G. F. & Artis, D. Novel connections and precision approaches. *Nat Rev*
305 *Immunol* **19,** 75–76 (2019).
306 4. Geva-Zatorsky, N. *et al.* Mining the human gut microbiota for immunomodulatory
307 organisms. *Cell* **168,** 928–943.e11 (2017).
308 5. Tan, T. G. *et al.* Identifying species of symbiont bacteria from the human gut that, alone,
309 can induce intestinal Th17 cells in mice. *Proc Natl Acad Sci U S A* **113,** E8141–E8150
310 (2016).
311 6. Balmer, M. L. *et al.* Microbiota-derived compounds drive steady-state granulopoiesis via
312 MyD88/TICAM signaling. *J Immunol* **193,** 5273–5283 (2014).
313 7. Deshmukh, H. S. *et al.* The microbiota regulates neutrophil homeostasis and host
314 resistance to Escherichia coli K1 sepsis in neonatal mice. *Nat Med* **20,** 524–530 (2014).
315 8. Tanoue, T. *et al.* A defined commensal consortium elicits CD8 T cells and anti-cancer
316 immunity. *Nature* **565,** 600–605 (2019).
317 9. Routy, B. *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy
318 against epithelial tumors. *Science* **359,** 91–97 (2018).
319 10. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti-PD-1
320 immunotherapy in melanoma patients. *Science* **359,** 97–103 (2018).
321 11. Vétizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut
322 microbiota. *Science* **350,** 1079–1084 (2015).
323 12. Matson, V. *et al.* The commensal microbiome is associated with anti-PD-1 efficacy in
324 metastatic melanoma patients. *Science* **359,** 104–108 (2018).
325 13. Atarashi, K. *et al.* ATP drives lamina propria T(H)17 cell differentiation. *Nature* **455,**
326 808–812 (2008).
327 14. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-
328 helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4,** 337–349 (2008).
329 15. Cahenzli, J., Köller, Y., Wyss, M., Geuking, M. B. & McCoy, K. D. Intestinal microbial
330 diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe*
331 **14,** 559–570 (2013).
332 16. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel
333 diseases. *Nature* **569,** 655–662 (2019).
334 17. Integrative HMP (iHMP) Research Network Consortium. The integrative human
335 microbiome project. *Nature* **569,** 641–648 (2019).
336 18. Olin, A. *et al.* Stereotypic immune system development in newborn children. *Cell* **174,**
337 1277–1292.e14 (2018).
338 19. Brandi, G. & Frega, G. Microbiota: Overview and Implication in Immunotherapy-Based
339 Cancer Treatments. *Int J Mol Sci* **20,** (2019).
340 20. Xin Yu, J., Hubbard-Lucey, V. M. & Tang, J. The global pipeline of cell therapies for
341 cancer. *Nat Rev Drug Discov* **18,** 821–822 (2019).
342 21. Taur, Y. *et al.* Intestinal domination and the risk of bacteremia in patients undergoing

343 allogeneic hematopoietic stem cell transplantation. *Clin Infect Dis* **55,** 905–914 (2012).
344 22. Taur, Y. *et al.* Reconstitution of the gut microbiota of antibiotic-treated patients by
345 autologous fecal microbiota transplant. *Sci Transl Med* **10,** (2018).
346 23. Morjaria, S. *et al.* Antibiotic-Induced Shifts in Fecal Microbiota Density and
347 Composition during Hematopoietic Stem Cell Transplantation. *Infect Immun* **87,** (2019).
348 24. Staffas, A. *et al.* Nutritional Support from the Intestinal Microbiota Improves
349 Hematopoietic Reconstitution after Bone Marrow Transplantation in Mice. *Cell Host*
350 *Microbe* **23,** 447–457.e4 (2018).
351 25. Taur, Y. *et al.* The effects of intestinal tract bacterial diversity on mortality following
352 allogeneic hematopoietic stem cell transplantation. *Blood* **124,** 1174–1182 (2014).
353 26. Scheiermann, C., Frenette, P. S. & Hidalgo, A. Regulation of leucocyte homeostasis in
354 the circulation. *Cardiovasc Res* **107,** 340–351 (2015).
355 27. Thompson, P. A. *et al.* Umbilical cord blood graft engineering: challenges and
356 opportunities. *Bone Marrow Transplant* **50 Suppl 2,** S55-62 (2015).
357 28. Gabrilove, J. L. *et al.* Effect of granulocyte colony-stimulating factor on neutropenia and
358 associated morbidity due to chemotherapy for transitional-cell carcinoma of the
359 urothelium. *N Engl J Med* **318,** 1414–1422 (1988).
360 29. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell*
361 **157,** 121–141 (2014).
362 30. Hall, M. W. *et al.* Inter-personal diversity and temporal dynamics of dental, tongue, and
363 salivary microbiota in the healthy oral cavity. *npj Biofilms and Microbiomes* **3,** 2 (2017).
364 31. Ramanan, P., Barreto, J. N., Osmon, D. R. & Tosh, P. K. Rothia bacteremia: a 10-year
365 experience at Mayo Clinic, Rochester, Minnesota. *J Clin Microbiol* **52,** 3184–3189
366 (2014).
367 32. Ulven, T. Short-chain free fatty acid receptors FFA2/GPR43 and FFA3/GPR41 as new
368 potential therapeutic targets. *Front Endocrinol (Lausanne)* **3,** 111 (2012).
369 33. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic
370 Treg cell homeostasis. *Science* **341,** 569–573 (2013).
371 34. Hergott, C. B. *et al.* Peptidoglycan from the gut microbiota governs the lifespan of
372 circulating phagocytes at homeostasis. *Blood* **127,** 2460–2471 (2016).
373 35. Ladinsky, M. S. *et al.* Endocytosis of commensal antigens by intestinal epithelial cells
374 regulates mucosal T cell homeostasis. *Science* **363,** (2019).
375 36. Foster, K. R., Schluter, J., Coyte, K. Z. & Rakoff-Nahoum, S. The evolution of the host
376 microbiome as an ecosystem on a leash. *Nature* **548,** 43–51 (2017).
377 37. Yoshii, K., Hosomi, K., Sawane, K. & Kunisawa, J. Metabolism of dietary and microbial
378 vitamin B family in the regulation of host immunity. *Front. Nutr.* **6,** 48 (2019).
379 38. Zaza, G. *et al.* Impact of maintenance immunosuppressive therapy on the fecal
380 microbiome of renal transplant recipients: Comparison between an everolimus- and a
381 standard tacrolimus-based regimen. *PLoS ONE* **12,** e0178228 (2017).
382 39. McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L. & Foster, K. R. Host
383 selection of microbiota via differential adhesion. *Cell Host Microbe* **19,** 550–559 (2016).
384 40. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone
385 species for the degradation of resistant starch in the human colon. *ISME J* **6,** 1535–1543
386 (2012).
387 41. Schirmer, M. *et al.* Linking the human gut microbiome to inflammatory cytokine
388 production capacity. *Cell* **167,** 1125–1136.e8 (2016).
389 42. Fu, Y.-Y. *et al.* T Cell Recruitment to the Intestinal Stem Cell Compartment Drives
390 Immune-Mediated Intestinal Damage after Allogeneic Transplantation. *Immunity* **51,**
391 90–103.e3 (2019).
392 43. Gerber, G. K. The dynamic microbiome. *FEBS Lett* **588,** 4131–4139 (2014).

44. Jobin, C. Precision medicine using microbiota. *Science* **359,** 32–34 (2018).
45. Zitvogel, L., Ma, Y., Raoult, D., Kroemer, G. & Gajewski, T. F. The microbiome in cancer immunotherapy: Diagnostic tools and therapeutic strategies. *Science* **359,** 1366–1370 (2018).
46. Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155,** 1451–1463 (2013).
47. Lakritz, J. R. *et al.* Gut bacteria require neutrophils to promote mammary tumorigenesis. *Oncotarget* **6,** 9387–9396 (2015).
48. Lakritz, J. R. *et al.* Beneficial bacteria stimulate host immune cells to counteract dietary and genetic predisposition to mammary cancer in mice. *Int J Cancer* **135,** 529–540 (2014).
49. Chen, H. *et al.* A forward chemical genetic screen reveals gut microbiota metabolites that modulate host physiology. *Cell* **177,** 1217–1231.e18 (2019).
50. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6,** 1621–1624 (2012).
51. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13,** 581–583 (2016).
52. Murali, A., Bhargava, A. & Wright, E. S. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome* **6,** 140 (2018).
53. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41,** D590-6 (2013).
54. Pinheiro, J. C., Bates, D. M., DebRoy, S. S. & Sarkar, D. Nlme: Linear and Nonlinear Mixed Effects Models. (2013).
55. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58,** 267–288 (1996).
56. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* (2011).
57. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* **2,** e55 (2016).
58. Hoffman, M. D. & Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* (2014).
59. Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* **15,** 962–968 (2018).

428 **Supplementary Information**

429 Appended and available online.

430

431 **Acknowledgments and Conflicts of Interest**

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

449

450 **Author Contributions**

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

455 **SUPPLEMENTARY INFORMATION**

456 **Methods**

457 **Complete blood count collection and characterization**

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

465

466 **16S rRNA gene amplification and multiparallel sequencing**

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

479

480 **Sequence analysis**

481 The 16S (V4-V5) paired-end reads were merged and demultiplexed. Amplicon sequence variants
482 (ASVs) were identified using the Divisive Amplicon Denoising Algorithm (DADA2) pipeline
483 including filtering and trimming of the reads[51]. Reads were trimmed to the first 180 bp or the first

484 point with a quality score Q<2. Reads were removed if they contained ambiguous nucleotides (N) or
485 if two or more errors were expected based on the quality of the trimmed read. We assigned taxonomy
486 to ASVs using a 8-mer based classifier trained by IDTaxa[52] using the SILVA database[53].
487
488 **Diversity calculations**
489 Microbiome alpha-diversity was measured by the inverse Simpson (IS) index of a sample. It was
490 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
491 sample i.
492
493 **Linear mixed-effects model of white blood cell counts**
494 To study the effect of auto-FMT on white blood cells, we investigated the white blood cell counts of
495 24 enrolled patients of this trial from the day of neutrophil engraftment until 100 days after. FMT
496 occurred on different days relative to neutrophil engraftment. Thus, we performed an analogous
497 analysis to that conducted in the original publication that demonstrated how FMT re-established a
498 diverse microbiome in the post-FMT period[22]. To answer if white blood cell counts differed post-
499 FMT, we used a linear mixed effects model of white blood cell counts, y, modeled as a function of the
500 FMT treatment as well as patient and timepoint specific random effects. We included random
501 intercept terms for each day i and each patient j, and a fixed effects term for the post-FMT period with
502 associated coefficient "armpost", using the indicator variable "FMT", that is 1 when a patient was
503 from the FMT treated arm of the trial and *day* was greater than or equal to the day of the FMT
504 procedure. We conducted independent analyses for neutrophil, lymphocyte and monocyte counts.
505 This resulted in the following model of a cell count, y, for patient j on day i:
506 $$y_{ij} = \beta_0 + armpost * FMT_{ij} + day_i + patient_j + \varepsilon_{ij}, \qquad i = 0, \dots, D, j = 0, \dots, P$$
507 with prior distributions $day_i \sim \mathcal{N}(0,\ \sigma_{day}^2)$, and $patient_j \sim \mathcal{N}(0,\ \sigma_{patient}^2)$, independent error
508 $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and fixed intercept β_0, for the D days post neutrophils engraftment and P patients,
509 (D=100, P=24). For convenience of those interested in reanalyzing our data, the part of our data
510 concerning the auto-FMT analysis is available in tidy format (data SD1).[54]

511 **Dynamic systems analyses**

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

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

518 $$\frac{\mathrm{d(W)}}{\mathrm{d}t} = W(gr + \sum_{j=1}^{P} \beta_j X_j)$$

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

522 $$\frac{\mathrm{d}\, ln(\mathrm{W})}{\mathrm{d}t} = gr + \sum_{j=1}^{P} \beta_j X_j$$

523 And we parameterized the corresponding discrete difference equation:

524 $$\frac{\Delta \ln(\mathrm{W})}{\Delta t} = gr + \sum_{j=1}^{P} \beta_j X_j$$

525 where Δln(W) is the log-difference between single days of neutrophils, lymphocytes or monocytes
526 counts, and Δt=1 for all intervals. Predictors include the counts of neutrophils, lymphocytes,
527 monocytes, eosinophils and platelets during an interval (homeostatic feedbacks), immunomodulatory
528 medication and clinical observations such as a blood stream infection and the onset of graft versus
529 host disease, HCT parameters such as graft types and conditioning regimens, and, additionally, the
530 microbiota composition in "stage 2" of our analysis (supplementary methods for data exclusion and
531 additional details on interval definitions). Importantly, by parameterizing a dynamic equation and
532 analyzing *rates of change*, our coefficient estimates have an immediate causal interpretation within
533 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"[43].

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:

$$y(x|\beta, gr) = gr + \sum_{j=1}^{p} \beta_j X_j + \varepsilon,$$

with intercept, *gr*, and residual error $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. 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[55] for feature selection using the sklearn package for the Python programming language[56]. 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[57]. Stage 1 identified important difference between transplant types, and we therefore stratified our data

561 into 4 cohorts according to their stem cell graft source. Using data independently from each cohort,
562 we applied "no U-turn" sampling[58] to produce 10,000 posterior samples from 5 independent MCMC
563 chains that parameterized the model:

564 $$y \sim \mathcal{N}(\mu, \sigma^2)$$

565 $$\mu = gr + \sum_{j=1}^{\hat{P}} x_j \beta_j$$

566 with uninformative prior distributions

567 $$gr \sim \mathcal{N}\ (mean = 0,\ standard\ deviation = 100)$$

568 $$\beta_j \sim \mathcal{N}\ (mean = 0,\ standard\ deviation = 100)$$

569 $$\sigma \sim HalfCauchy\ (\ beta = 2)$$

570 where y is the observed daily change of a focal white blood cell type as in stage 1 with normal
571 distributed mean, μ, and σ, the model uncertainty. μ was a function of the baseline growth rate, gr,
572 and predictors, $\hat{P}$: medications with non-zero coefficients in stage 1, the white blood cell counts,
573 patient age and sex, and HCT conditioning intensities; additionally, $\hat{P}$ now included the log-
574 abundances of microbial genera as measured by 16S sequencing from DNA in the stool collected on
575 the second day of a daily interval (see supplementary methods for details). We considered taxa that
576 were among the 100 most abundant, or had reached maximum relative abundances of at least 10%,
577 and selected those who were non-zero in more than 75% of our samples. White blood cell counts and
578 microbiota data present during a daily interval were log-transformed, and zeros were filled with half
579 of the minimum observed non zero counts (i.e. 0.5e3 and 2e-6, respectively). We focused on the
580 largest cohort (PBSC) and used the independent inference results from TCD, BM, and cord cohorts
581 for validation.
582
583 *Validation score*
584 Coefficients learnt from the PBSC patient cohort were assigned a "validation score" based on the
585 results obtained from the other three MSK patient cohorts. Our requirements for validation were
586 conservative; we required evidence from our validation data sets as well as absence of counter

evidence. For regression results from each of the validation graft type cohorts, i.e. TCD, BM, and cord, we checked if a coefficient had more than 75% probability (50%HPDI) to have the same sign as the mean of the PBSC coefficient posterior for a given predictor. If so, this was considered evidence of validation, and we summed the evidence over the three validation sets (i.e. maximum score of 3, 1 from each of TCD, BM, and cord cohorts). Conversely, if we found more than 75% probability among any of the validation data sets that a given predictor had the *opposite* sign as the posterior mean calculated from PBSC data, this was considered counter evidence and the validation score was always set to zero.

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 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.

Shotgun sequencing

Sequencing of 124 post-neutrophil engraftment was conducted on the Illumina HiSeq platform. For details and the processing of the FASTQ files, see supplementary methods. We used the HUMAnN2 pipeline[59] with default settings for functional profiling of our samples, with the UniRef90 data base and ChocoPhlAn for alignment, and we renormalized our samples by library depth to copies per million. We used MetaCyc to obtain stratified and unstratified pathway abundances.

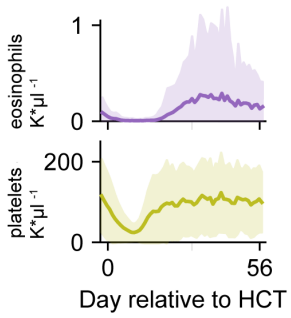
Statistical analysis of shotgun data

We calculated the predicted microbiota potency score for each sample and separately for neutrophils,

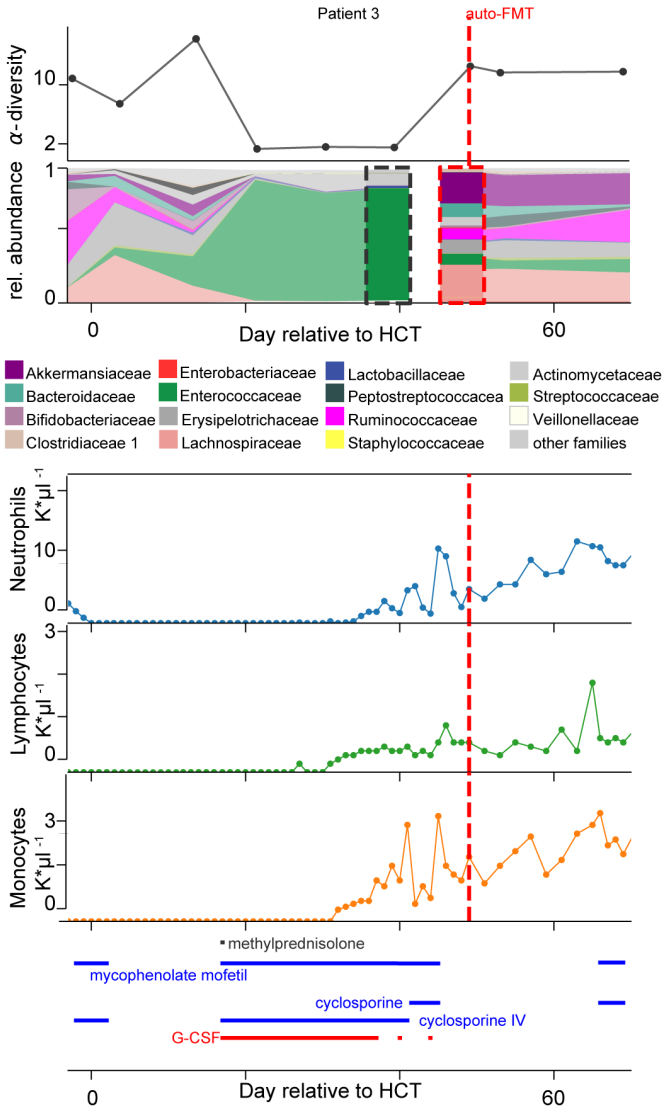
615 lymphocytes and monocytes, by multiplying the abundances of taxa in each of the 124 samples with
616 the corresponding posterior coefficients obtained from the PBSC inference. To distinguish the sets of
617 metabolic functions that separate samples with positive and negative predicted potencies, we
618 converted the pathway abundances into presence and absences profiles. We performed a linear
619 discriminant analysis between positive and negative potency samples with a least squares solver and
620 automatic shrinkage using the Ledoit-Wolf lemma using the sklearn package for the Python
621 programming language[56]. To assess differences in the presence or absence of pathways between
622 samples with positive and negative potency, we used Fisher's exact test for each pathway.
623
624
625
626

627 **Supplementary Figures**

628
629



630
631
632 Figure S1: Eosinophil counts and platelet counts per day relative to HCT; line: mean, shaded: ± standard
633 deviation)
634



635
636
637 Figure S2: HCT patient who received an autologous fecal microbiota transplant (auto-FMT, dashed red line)
638 that restored commensal microbial families and ecological diversity in the gut microbiota, with concurrent cell
639 counts of peripheral neutrophils, lymphocytes and monocytes and immunomodulatory drug administrations.

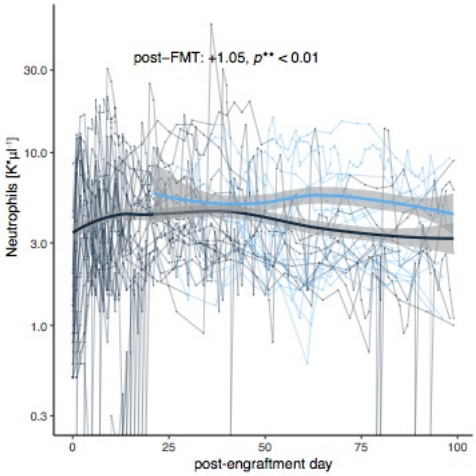


Figure S3: Neutrophil counts in 24. FMT trial patients. Thin lines: raw data (blue: post-FMT); thick black: mean per day, thick blue: mean+post-FMT coefficient. Means and confidence intervals (shaded region) from linear mixed effects model (methods).

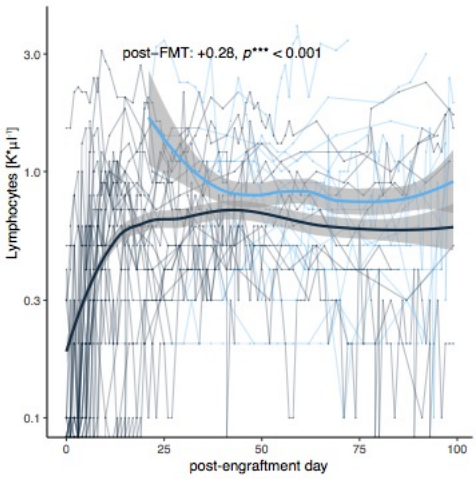


Figure S4: Lymphocyte counts in 24. FMT trial patients. Thin lines: raw data (blue: post-FMT); thick black: mean per day, thick blue: mean+post-FMT coefficient. Means and confidence intervals (shaded region) from linear mixed effects model (methods).

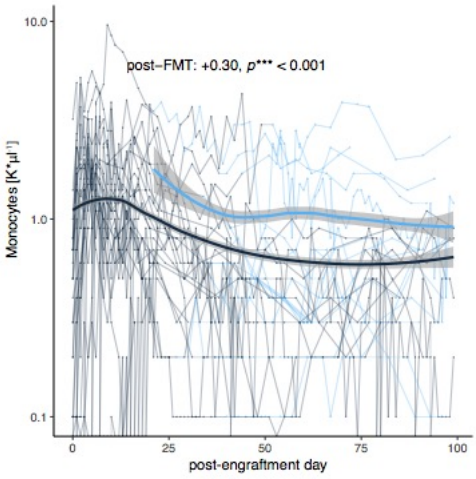


Figure S5: Monocyte counts in 24. FMT trial patients. Thin lines: raw data (blue: post-FMT); thick black: mean

per day, thick blue: mean+post-FMT coefficient. Means and confidence intervals (shaded region) from linear mixed effects model (methods).

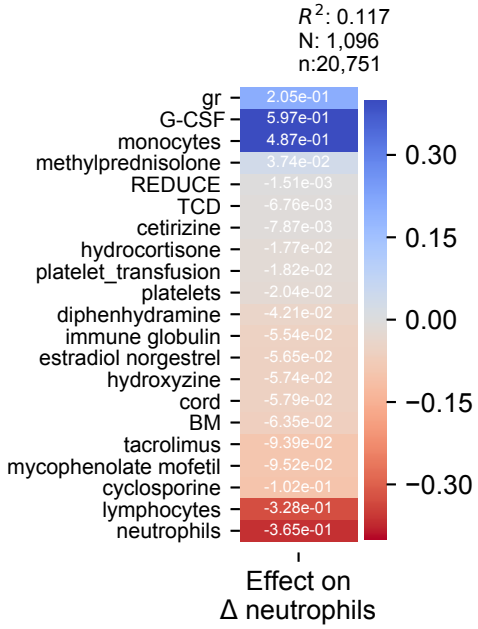


Figure S6: "Stage 1" regression on neutrophil dynamics on patients without microbiome data. Coefficients from 10-fold cross-validated elastic net regression daily changes in neutrophils. gr: intercept; TCD: T-cell depleted graft (ex-vivo) by CD34+selection; PBSC: peripheral blood stem cells; BM: bone marrow; cord: umbilical cord blood; REDUCE: reduced-intensity conditioning regimen; F: female; N: patients, n: samples (daily changes in neutrophils).

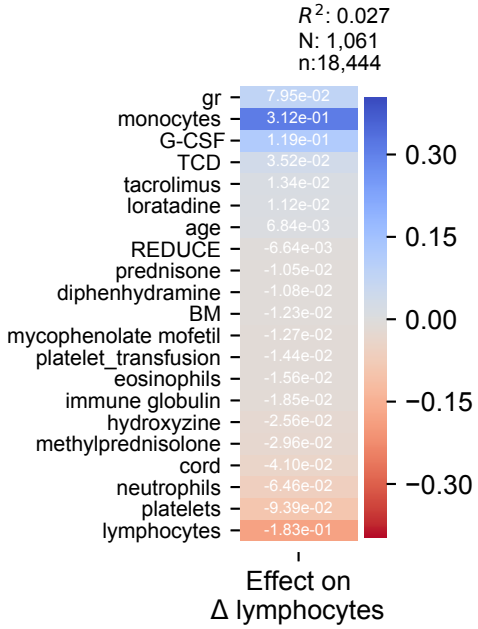


Figure S7: "Stage 1" regression on lymphocytes dynamics on patients without microbiome data. Coefficients from 10-fold cross-validated elastic net regression daily changes in lymphocytes. gr: intercept; TCD: T-cell depleted graft (ex-vivo) by CD34+selection; PBSC: peripheral blood stem cells; BM: bone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

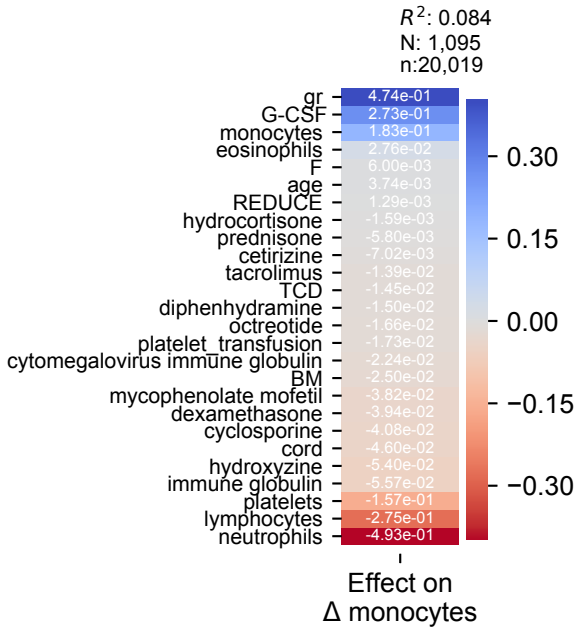


Figure S8: "Stage 1" regression on lymphocytes dynamics on patients without microbiome data. Coefficients from 10-fold cross-validated elastic net regression daily changes in lymphocytes. gr: intercept; TCD: T-cell depleted graft (ex-vivo) by CD34+selection; PBSC: peripheral blood stem cells; BM: bone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

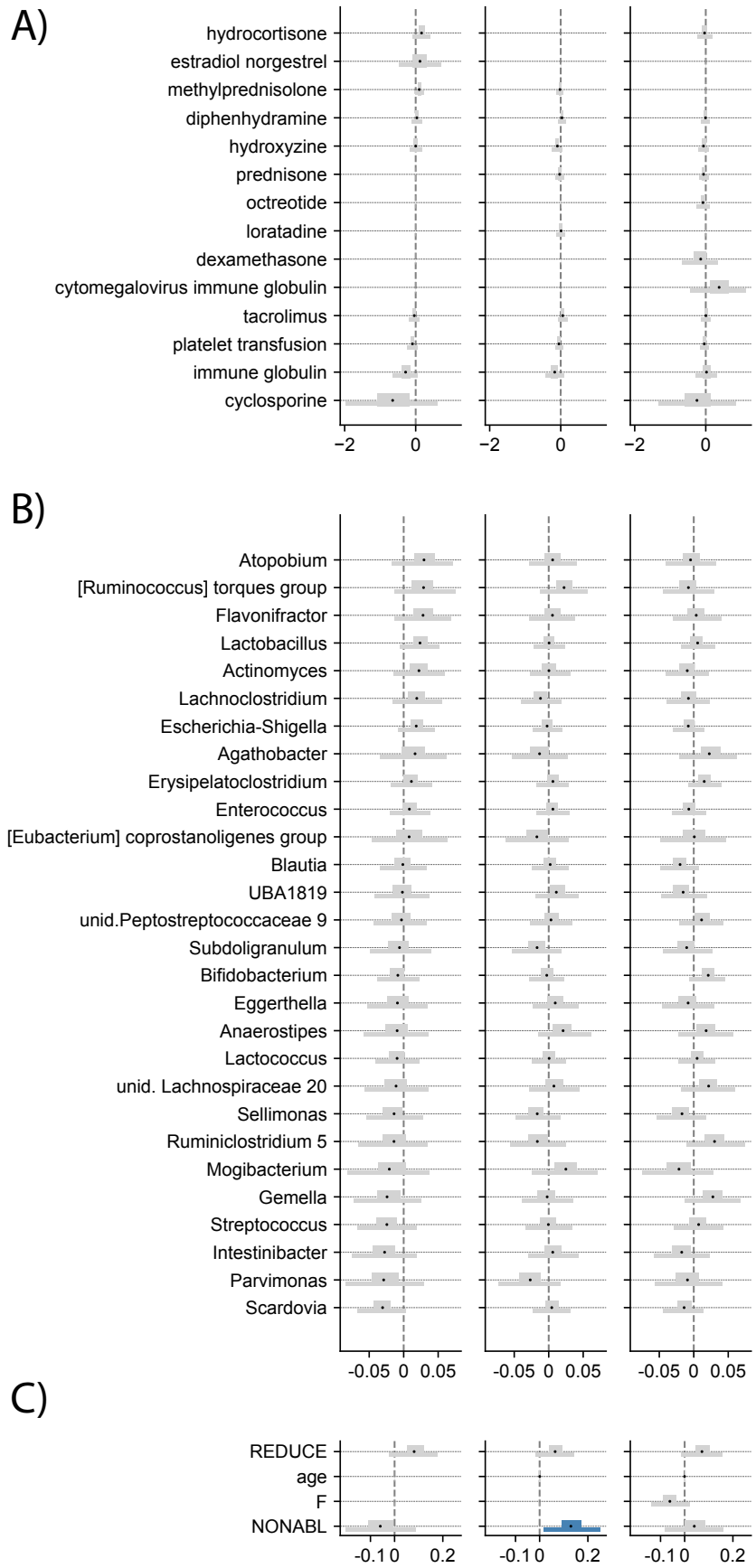


Figure S9: Additional coefficient estimates from the Bayesian regression, see also Figure 3. REDUCE: reduced-intensity conditioning regimen; NONABL: non-myeloablative conditioning regimen. F: female

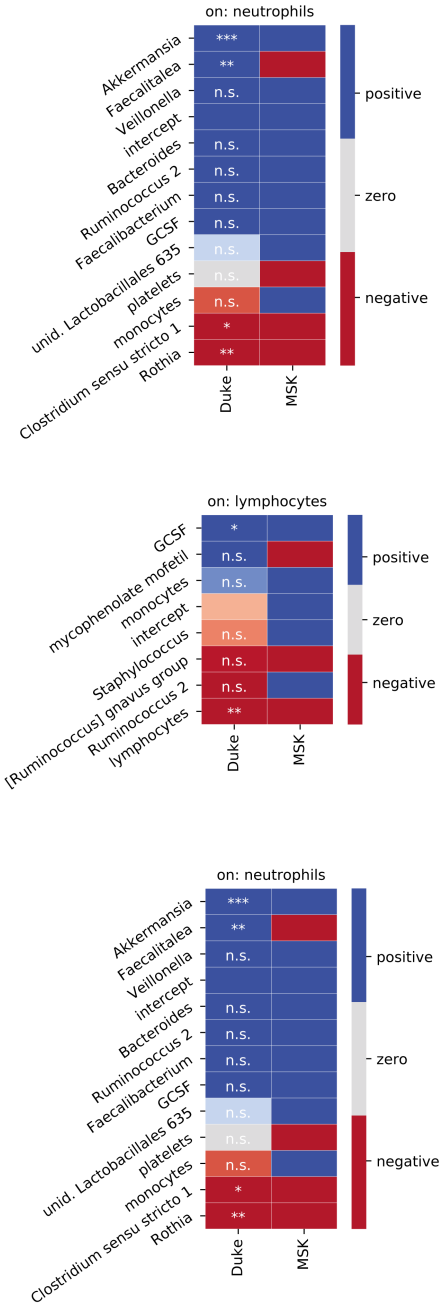


Figure S10: Validation analysis of predictors on white blood cell dynamics using data from patients treated at DukeHealth. Individual univariate regressions of microbiome and clinical predictors identified in stage 2 of our analysis on daily changes in neutrophils, lymphocytes and monocyte. Bonferroni corrected p-values: ***<0.001, **<0.01, *<0.05; p>0.05: n.s. Sign of coefficients from MSK PBSC patients for comparison.

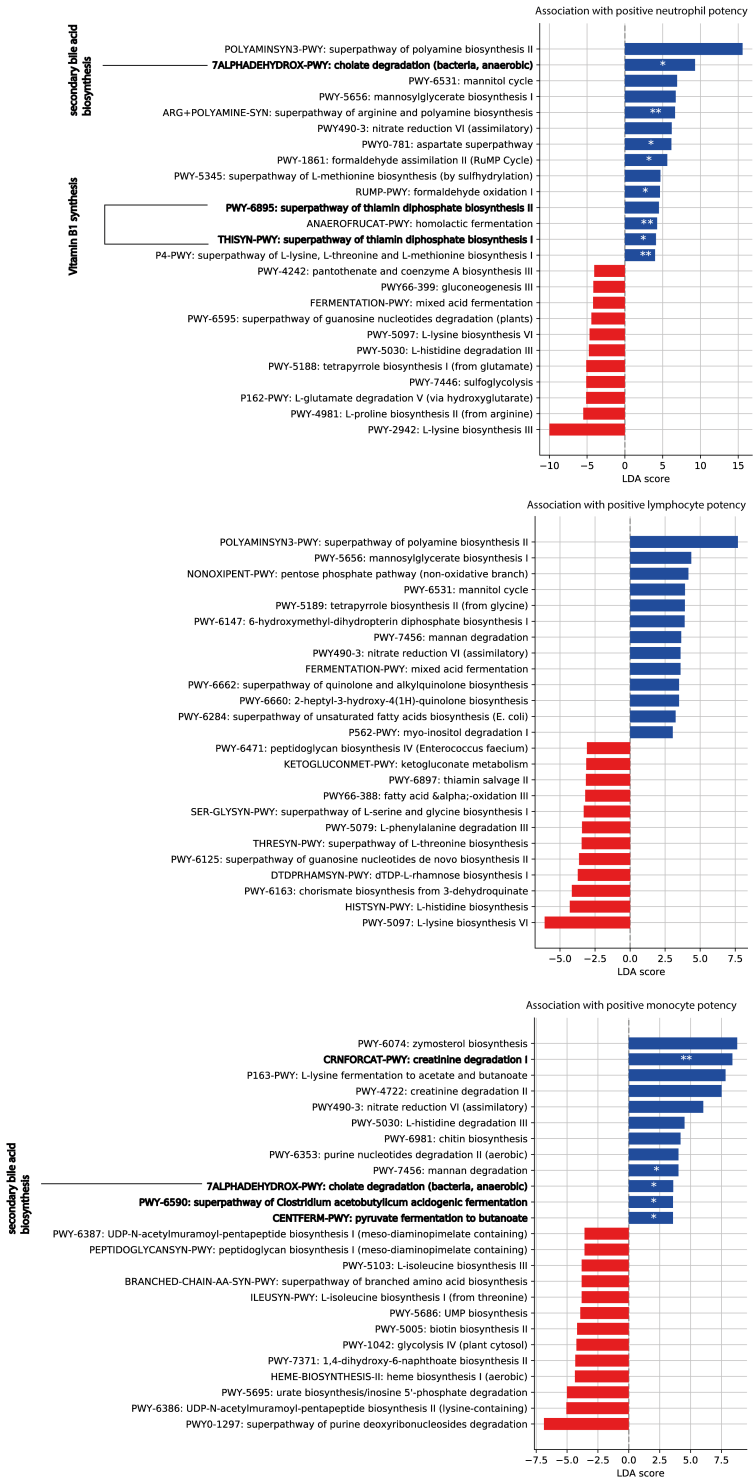


Figure S11: Functional analysis of microbiota samples. To distinguish samples predicted to increase rates of white blood cells, a microbiota potency score was calculated from posterior coefficients (Figure 3, methods) and the relative abundance of taxa in samples. Bars show linear discriminant analysis (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:*.

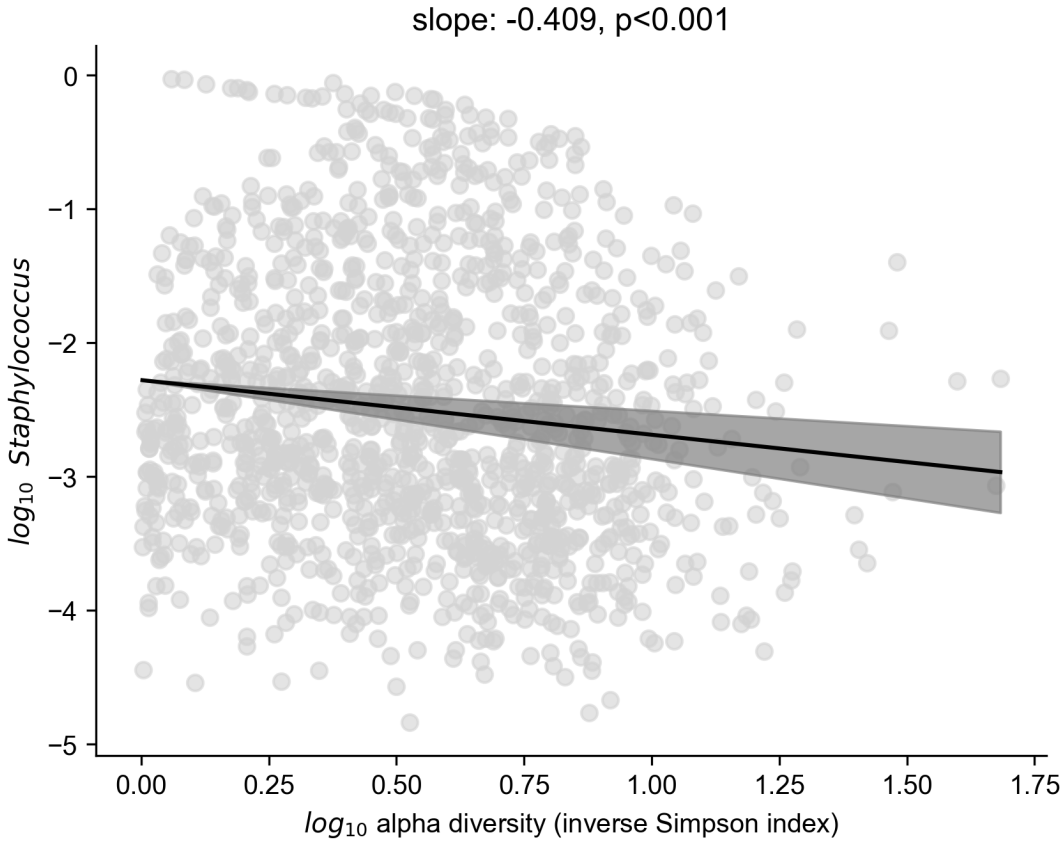


Figure S12: The relative non-zero abundance of Staphylococcus is inversely related to microbiome alpha diversity, shaded: 95% confidence intervals.

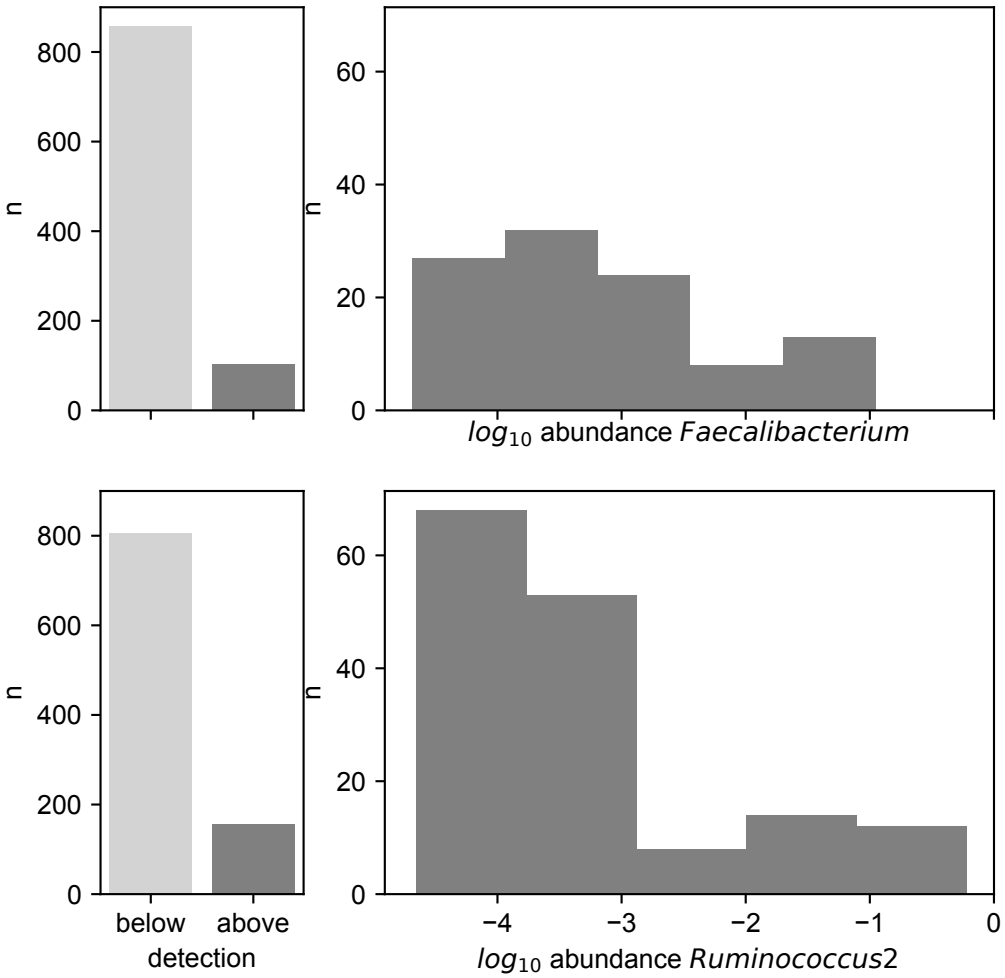


Figure S13: Abundance profiles of the two genera, *Faecalibacterium* and *Ruminococcus 2*, most strongly associated with white blood cell increase; number of times detected (left) and log_{10} abundance distribution when above detection (right).

735 Table S1: Data set summary and patient characteristics. HCT-graft types: TCD: T-cell depleted graft (*ex-vivo*)
736 by CD34+selection; PBSC: peripheral blood stem cells; BM: bone marrow; cord: umbilical cord blood;
737 Conditioning intensity: Bacigalupo classification, graded categories from most to least intense (ABLATIVE,
738 REDUCE, NONABL).
739

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

740 *) some patient received several HCTs
741
742
743 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

744
745
746
747
748
749
750
751

752 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

753
754
755 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

756
757
758 Table S5: Patient and HCT characteristics of the Duke University patient cohort.
759

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

Supplementary Data

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. *cIVS.csv*: diversity (inverse Simpson index) 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. FMT_analysis.csv: convenience table for Figure 2