Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Peled JU, Gomes ALC, Devlin SM, et al. Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. N Engl J Med 2020;382:822-34. DOI: 10.1056/NEJMoa1900623

Microbiota as a predictor of mortality in allogeneic HCT
Supplementary Appendix

List of Investigators

Jonathan U. Peled, M.D., Ph.D., Antonio L.C. Gomes, Ph.D., Sean M. Devlin, Ph.D., Eric R. Littmann, B.A., Ying Taur, M.D., Anthony D. Sung, M.D., Daniela Weber, M.D., Daigo Hashimoto, M.D., Ph.D., Ann E. Slingerland, B.S., John B. Slingerland, B.S., Molly Maloy, M.S., Annelie G. Clurman, B.A., Christoph K. Stein-Thoeringer, M.D., Kate A. Markey, M.B.B.S., Ph.D., Melissa D. Docampo, B.S., Marina Burgos da Silva, Ph.D., Niloufer Khan, M.D., Andre Gessner, M.D., Julia A. Messina, M.D., Kristi Romero, B.S., Meagan Lew, B.S., Amy Bush, B.A., Lauren Bohannon, B.S., Daniel G. Brereton, B.A., Emily Fontana, B.A., Luigi A. Amoretti, B.S., Roberta J. Wright, M.S., M.B.S., Gabriel K. Armijo, B.S., Yusuke Shono, M.D., Ph.D., Míriam Sanchez-Escamilla, M.D., Nerea Castillo Flores, M.D., Ph.D., Ana Alarcón Tomas, M.D., Richard J. Lin, M.D.,Ph.D., Lucrecia Yáñez San Segundo, M.D., Ph.D., Gunjan L. Shah, M.D.,M.S., Christina Cho, M.D., Michael Scordo, M.D., Ioannis Politikos, M.D., Kasumi Hayasaka, B.S., Yuta Hasegawa, M.D., Boglarka Gyurkocza, M.D., Doris M. Ponce, M.D., Juliet N. Barker, M.B.B.S., Miguel-Angel Perales, M.D., Sergio A. Giralt, M.D., Robert R. Jenq, M.D., Takanori Teshima, M.D., Ph.D., Nelson J. Chao, M.D., Ernst Holler, M.D., Joao B. Xavier, Ph.D., Eric G. Pamer, M.D., Marcel R.M. van den Brink, M.D., Ph.D.

Supplemental Text

Transplantation Characteristics and Clinical Outcomes

Inclusion criteria for this study were patients with an evaluable stool sample (successfully 16S-amplified and sequenced with >200 reads) that had been collected after day –30 of a first allo-HCT at any of the four centers. Patients who had received an autologous hematopoietic cell transplantation prior to allogeneic HCT were considered eligible for inclusion. Samples from patients who received second allografts were excluded if they were collected after day –10 relative to the second transplant.

Patient exclusions from various stages of analysis are tabulated in **Supplemental Table S1**. Importantly, several subjects at one of the centers (MSK) had participated in a randomized clinical trial of fecal microbiota transplantation (FMT, NCT02269150[1]). Patients who were randomized to the control arm (no FMT) were included in all analyses in this study. For patients randomized to receive FMT, samples collected after the FMT procedure were completely excluded from this study, but the pre-FMT samples from these patients were included in analyses of microbiota composition and dynamics (**Figure 1A** and **Figure 2**). Patients randomized to the FMT arm were excluded from analysis of clinical outcomes (**Figures 1B-F**, **Figure 3C** and **Supplemental Figure S8**).

Similarly, patients who were analyzed in our prior study of diversity and survival[2] were included in analyses of microbiota composition and dynamics (**Figure 1A** and **Figure 2**) but were excluded from analysis of association between clinical outcomes and diversity in the peri-engraftment period (**Figures 1B-F**) so that the MSK cohort described here is independent of the prior study. These patients were included in the analysis of clinical outcomes and diversity in the pre-HCT period. Patients who participated in a trial randomizing the empiric antibiotic regimens for febrile neutropenia (NCT03078010) were included in the analysis because both arms are within standard practice.

Conditioning regimens were categorized by intensity of myeloablation.[3] Clinical data were obtained from institutional clinical research databases and from dedicated chart reviews. In **Table 1**, the "other" disease category includes biphenotypic acute leukemia, natural killer-cell large granular lymphocyte leukemia, plasmacytoid dendritic cell neoplasms, and non-malignant hematologic disorders including familial hemophagocytic lymphohistiocytosis, X-linked lymphoproliferative disease, and paroxysmal nocturnal hemoglobinuria. Among

the 447 recipients of T-cell depleted grafts transplanted at MSK, 437 (97.8%) received grafts that were CD34-selected on CliniMACS CD34 Reagent system (Miltenyi Biotec, Gladbach, Germany). For ten of the T-cell-depleted (2.2%) recipients, grafts were prepared via sheep-erythrocyte-rosetting-based methods. For patients who had transferred their care outside of the four centers within two years of follow-up, outcomes were assessed by telephone interviews with the patients' treating physicians. For the variables reported herein, there were no missing clinical data except in the case of a single patient at MSK who was not evaluable for the hematopoietic cell-transplantation comorbidity index (HCT-CI) because pulmonary function testing was not performed prior to transplantation, and in the case of 38 samples whose time of collection was available only as "pre-transplant" for which a value of day –7 was assigned. Healthy volunteers who provided stool samples provided written informed consent according to a biospecimen collection protocol approved by the MSKCC Institutional Review Board. Raw sequence files from the Human Microbiome Project were downloaded from the human microbiome project website (https://hmpdacc.org)[4] and processed computationally on the same pipeline as samples in this study.

Samples

Samples were collected during different time periods lasting 1.4 – 8.8 years within the years 2009 – 2018 (**Supplemental Table S11**). At Regensburg, Duke, and Hokkaido, and for the majority of the MSK patients, weekly samples were requested during the duration of the patients' inpatient admissions, or for outpatient transplants from the start of conditioning through engraftment. At MSK, there were additional collection efforts that yielded additional samples beyond the weekly samples, including upon readmission, in the outpatient setting at scheduled timepoints, and in a subset of patients, near-daily collections while admitted. For the analysis of microbiota composition dynamics (**Figure 2A-D**) we included any evaluable sample collected on or after day -30 relative to a first allo-HCT. Overall, one quarter of the samples were collected between day -30 and 0, 50% of the samples were collected between day -30 and day 10, 75% of the samples were collected between day -30 and day +25.

In the recent Microbiome Quality Control project[5], biospecimen type (e.g. stool vs. skin swabs), DNA extraction procedures, sample-handling environment, and bioinformatics pipelines were all identified as important sources of variability in microbiome data. In that study, biologic variation among samples was a larger

source of variation than technical or computational differences between laboratories. We aimed to minimize bias by (a) collecting, aliquoting, and freezing samples at each center according to harmonized procedures, and (b) by performing DNA extraction, PCR amplification, sequencing, and computational analysis centrally at MSK.

Stool samples were aliquoted and stored frozen without additives at each center. Aliquots were shipped frozen to a central laboratory (MSK) where bacterial cell walls were disrupted by silica bead-beating, nucleic acids isolated, and the genomic 16S ribosomal-RNA gene V4-V5 variable region was amplified and sequenced on the Illumina MiSeq platform as previously described.[2,6] PCR products were purified either using Qiagen PCR Purification Kit or Agencourt AMPure PCR purification system following the manufacturers' instructions. In cases of poor PCR amplification, the standard PCR buffer was replaced with Ampdirect Plus PCR buffer (Nacalai USA, San Diego, CA). In particular, when amplifying samples from Hokkaido, we observed that 8 of the initial 13 samples (62%) failed to amplify, in comparison with <25% of failed amplifications from the other cohorts. PCR inhibitor removal using Zymo Research OneStep PCR inhibitor removal kit allowed amplification from only one additional sample. Following other microbiota-profiling studies of Japanese populations,[7] we were able to amplify 13 of 13 initial Hokkaido samples using Nacalai AmpDirect Plus buffer. All subsequent Hokkaido samples were amplified according to this protocol.

Identification of Operational Taxonomic Units

The Operational Taxonomic Units[8] (OTUs, referred to in the main text as taxonomic units) were called using a hybrid approach combining *de novo* and closed-reference OTU-calling. Quality-filtered sequences with > 97% identity were grouped into operational taxonomic units (OTUs).[9] For *de novo* calling, we used the vsearch algorithm[9] to dereplicate sequence reads. Reads were filtered to sequences of length between 200-350 nucleotides and abundance size of at least two. The usearch algorithm was used to cluster OTUs (-cluster_otus flag) with parameter –uparse_break 3. The option uchime_ref was further used to filter for chimeras according to a dereplicated version of NCBI 16S Microbial database.[10] OTUs were clustered at 97% identity. For closed-reference OTU calling, the qiime command pick_closed_reference_otus.py was used. A combined set of over 140M reads from approximately 10000 samples were used for *de novo* OTU calling to define the closed-reference

set of OTUs. Reads from subsequent independent sequencing runs were then identified by closed-reference OTU-calling against the reference set.

OTUs were classified to the species level against the Greengenes database,[11] with gaps in taxonomic annotation filled in by classification against the NCBI 16S ribosomal RNA sequence database (release Dec 07, 2016).[10]

Intestinal microbiota diversity

Alpha-diversity is a mathematical value that summarizes an ecological (e.g. microbial) community according to the count of unique species and how evenly their frequencies are distributed. The higher the number of unique species (richness) and the more evenly they are distributed (evenness), the higher the α-diversity. Notably, α-diversity values do not convey any information about the actual species present. Thus, two completely different communities might have identical α-diversity values and share no species in common. Here, we calculated α-diversity using the inverse Simpson index at the level of OTUs. An alternative and commonly used method for α-diversity is the Shannon index. These two metrics are highly correlated with one another, but the Simpson index is slightly less sensitive to the long tail of rare bacteria than the Shannon index.[12]

Taxonomic color scheme

The taxonomic color scheme used in **Figure 2D** was modified from that used in the R package yingtools2 (https://github.com/ying14/yingtools2) and Taur *et al.*[1] These color schemes have been customized to highlight common taxonomic patterns in microbiota community in allo-HCT patients. Each genus is assigned to a distinct color shade derived from a basal color that is assigned to a higher-rank taxonomic group in the dataset. This allows visualization of both genus-level and higher-rank taxonomic information. For example, genera from phylum Actinobacteria are in shades of purple, genera from phylum Bacteroidetes are in shades of teal, and most of phylum Firmicutes is depicted in shades of brown. Certain taxa of biologic interest, including those that we find frequently dominating in these populations, are highlighted separately. For example, genus *Enterococcus* is

in green and family Lachnospiraceae (including genus *Blautia*) is in shades of pink. The reds of phylum Proteobacteria are variegated to allow resolution between genus *Klebsiella* and genus *Escherichia.*

Enterotypes classification

Several large-scale studies of healthy human intestinal communities have discerned recurring patterns in which some configurations of relative microbial abundance are observed more frequently than others. A collaborative report by several workers in the microbiota field recently acknowledged the limitations of the enterotypes approach but recommended it as a standard first-step in the classification of human intestinal microbiota datasets.[13] While each enterotype is complex, it is named according to the dominant taxonomic group that contribute to enterotype clustering (Bacteroides, Prevotella, and Firmicutes). In this analysis, genus-level abundances were classified into Enterotypes using the online tool at http://www.enterotypes.org/.

Statistical Analysis of Clinical Outcomes

Survival analysis was performed using R package *survival.*[14] When peri-engraftment samples (day 7-21) were analyzed, outcomes were considered in landmark analyses of survivors beyond day 21. When pre-HCT samples were analyzed, outcomes were analyzed from HCT day 0. All survivors were censored at two years of follow-up.

The cumulative incidences of transplant-related mortality (TRM), relapse (defined here as relapse or progression of disease) and GVHD-related mortality (GRM) were estimated using cumulative incidence functions with *cuminc*. The competing risk for relapse was death without relapse. The competing risk for TRM was relapse. The competing risks for GRM were relapse and death without GVHD. Cox proportional hazards multivariable regression models (*coxph)* were used to assess associations between microbiota and overall survival. For TRM, GRM, and relapse cause-specific Cox regression was used. Hazard ratios are presented in main text with square brackets indicating the 95% confidence interval.

When diversity was considered as a binary (high vs. low) variable, patients were split into above-median and below-median groups using institution-specific median diversity cutoff values. **Supplemental Table S10** also tabulates the clinical characteristics of patients in the high- and low-diversity groups.

As tabulated in **Tables 2**, **S6**, **S7**, and **S8** diversity was also considered as a continuous variable (log_{10}-transformed). In the analysis of peri-engraftment diversity, when patients had more than one sample in the day 7-21 sampling window, the per-patient median value was considered. For pre-HCT analysis, the diversity of the first sample for each patient collected in a sampling window of day –30 to day–6 was used. We considered only the first sample in the pre-HCT analysis because we reasoned that this would be the best estimate of baseline microbiota composition. In multivariate analysis, diversity, age, graft type, conditioning regimen, and the HCT-CI were considered. When samples from more than one institution were analyzed together, the institution of origin was stratified in the *coxph* formula with parameter *strata*(institution).

Identification of bacteria signature associated with patient outcome

We used the R package *glmnet* to perform regularized regression and identify a signature of bacterial abundances during the peri-engraftment period (day 7-21) to predict patient outcomes. We eliminated rare and highly correlated taxonomic groups as follows: Considering taxonomy from the phylum to the OTU levels, our dataset initially contained 8,461 taxonomic groups. We restricted our analysis *a priori* to taxonomic groups that appeared in more than 10% of our samples with relative abundance above 10^{-4}. We used this threshold in a previous study[15] and similar thresholds have been applied by other groups.[16-21] In addition, if the abundance of a parental clade had Pearson correlation greater than 75% to a hierarchically lower clade, the parental clade was removed. We prioritized removal of parental clades in order to favor higher-resolution taxonomic identification. After these filters, 172 taxonomic groups were used as input features in regularized regression (**Supplemental Figure S8** and **Supplemental Table S9**).

The signature of bacterial effect sizes was identified after 10-fold cross-validation using the function *cva.glmnet*, with parameters alpha=c(0,0.1,0.25,0.5,1.0) for the elastic net penalty, family="cox" and maxit = 10000. This function identifies lambda and alpha parameters by minimizing cross-validation loss. The abundance of each taxonomic group was log transformed and a pseudo count of $2 \cdot 10^{-3}$ was given to eliminate the possibility of –Infinity values, i.e. log(abundance + 2e-3). The signature of bacterial effect sizes was trained in the MSK cohort and used to compute a risk score in the combined cohort of Regensburg, Duke, and Hokkaido

(**Supplemental Figure S8, Supplemental Table S9**). This risk score was standardized to assure mean=0 and variance=1.

Mapping high-dimensional data into tSNE projections

We used the *Rtsne* package to perform t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction. The Bray-Curtis dissimilarity index computed at the genus level was used as input to the t-SNE algorithm. The perplexity, theta, and iteration parameters were selected by systematic visual inspection to illustrate patterns in the data such as the large cluster of early/diverse samples and the later/low-diversity dominated clusters.

Compositional domination analysis

In **Figure 2E-G**, domination was defined as any OTU with relative abundance >0.3.[22] The samples were binned into 7-day sliding windows in accordance with the approximately weekly collection schedule. The domination cumulative incidence plot (**Figure 2E**) considers patients with at least one evaluable sample in the pre-HCT period and at least one sample in the post-HCT period. The fraction of patients in whom at least one instance of domination was detected by the given time is plotted in **Figure 2F**.

Estimating geographic and temporal variation in microbiota compositions

We quantified the variation of microbiota composition among the four institutions in comparison to the temporal variation that occurs over time relative to HCT by using a Bray-Curtis beta-diversity matrix (**Supplemental Figure S6**). We defined a reference centroid using pre-HCT samples from MSK patients and computed the distance of each sample in the dataset to the pre-HCT MSK centroid. A mean distance value per patient was computed in cases where patients had multiple samples. The Wilcoxon test was used to compare distances of groups of samples from the centroid.

Antibiotic analysis

We[1,2,22-26] and others[27-29] have previously reported that antibiotics are associated with microbiome disruption and clinical outcomes in allo-HCT patients. This study provided an opportunity to explore the association of antibiotic exposures both with (a) microbiota composition and (b) clinical outcomes. We sought

first to identify key antibiotics most strongly associated with a decline in diversity during allo-HCT, and then to consider exposure to those drugs in a multivariate model of survival.

To identify key drugs associated with a decline in intestinal microbiota diversity, we considered the change in diversity between the pre-HCT period and the peri-engraftment period, using the same definitions for these sampling periods as elsewhere in this study: for pre-HCT the earliest sample available per patient between day -30 and day -6, and for the peri-engraftment period the median per-patient diversity of samples collected between day 7 and 21. We defined an antibiotic exposure window of day -7 and 14 relative to HCT to capture the bulk of antibiotics administered in this population (**Supplemental Figure S5**).[1] Among the 31 unique antibiotic drugs administered to the patients during their transplant courses, we considered only those to which >20% of patients were exposed during this window in at least one institution. We also excluded drugs employed in prophylactic regimens (**Supplemental Table S3**), among which were fluoroquinolones, trimethoprim-sulfamethoxazole, rifaximin,[26] and intravenous vancomycin.[30,31] This yielded a set of five antibiotics for further evaluation: cefepime, doripenem, meropenem, piperacillin-tazobactam and meropenem (**Supplemental Figure S5**).

We modeled variation in microbiota diversity, ΔS, between the peri-engraftment and pre-HCT periods as a function of the time span, Δt, and the effect of antibiotic exposures. We assumed that the impact of antibiotics to microbiota diversity is proportional to burden of exposure. We defined the exposure burden to antibiotic i, a_i, as the number of days of exposure between days -7 and 14 relative to HCT. The time span was computed as day 14 minus the day of the first pre-transplant sample. Formally, the association of drug exposure with diversity was evaluated using linear regression with the following equation in the *stats* package of R using the function *lm()*:

$$\widehat{\Delta S} = \beta_0 \bullet \Delta t + \beta_1 \bullet a_1 + \cdots + \beta_n \bullet a_n$$

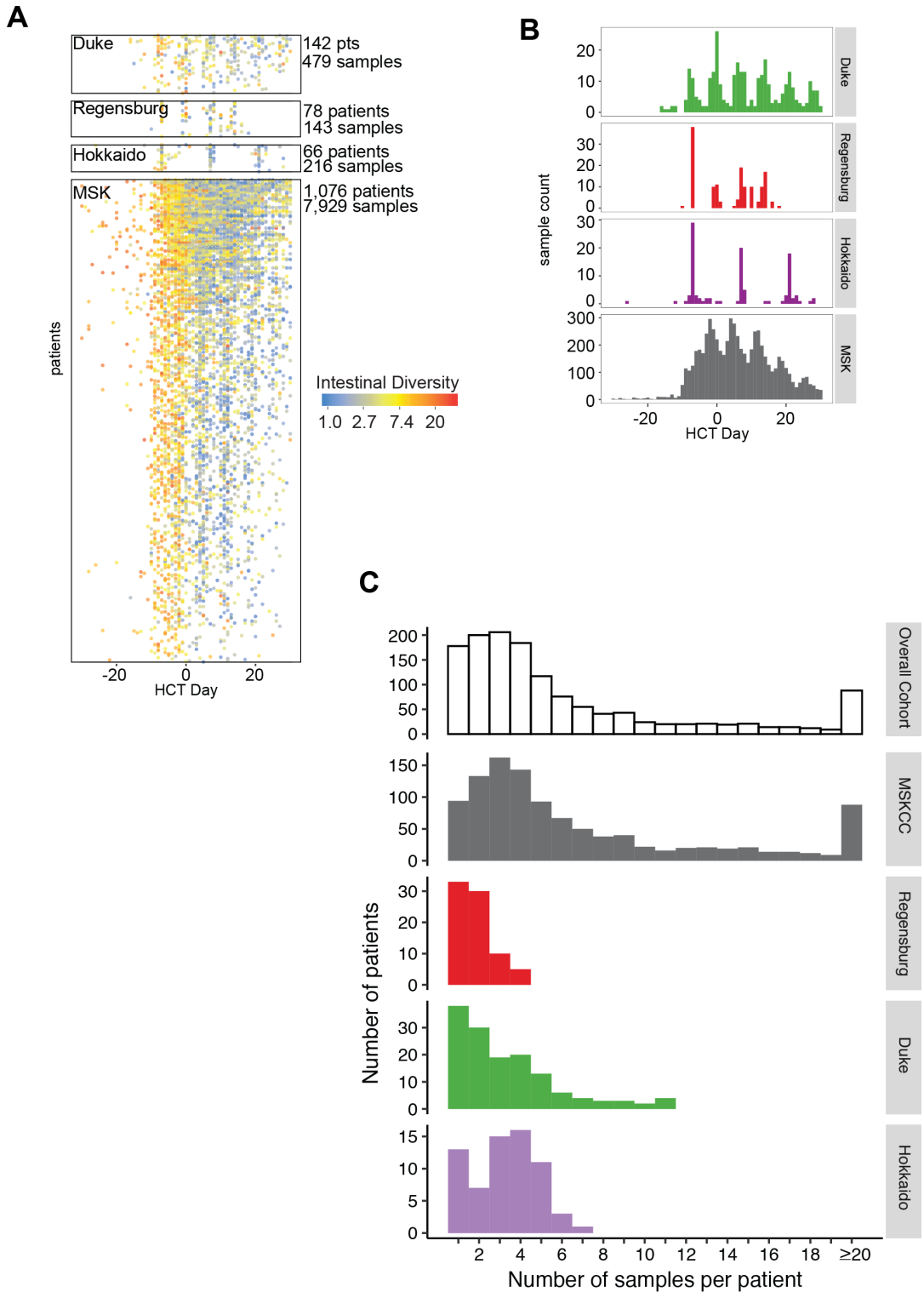
where subscripts *1* to *n* correspond to each drug considered in the model and *ß* represents the regression coefficients. No parameter was used to represent the intercept as it was forced to be 0. Two of the five antibiotics evaluated, piperacillin/tazobactam and meropenem, were significantly associated with variation in microbiota (**Supplemental Table S5**), consistent with our prior observation that exposure to piperacillin/tazobactam and

carbapenems is associated with microbiota disruption.[23] We then modified the multivariable model of microbiota diversity and overall survival that is presented in **Table 2** to include exposure to piperacillin-tazobactam and meropenem (as continuous variables of exposure duration in days). We found that the association between intestinal microbiota diversity and survival remained significant in the MSK and Regensburg cohorts (**Supplemental Table S6**).

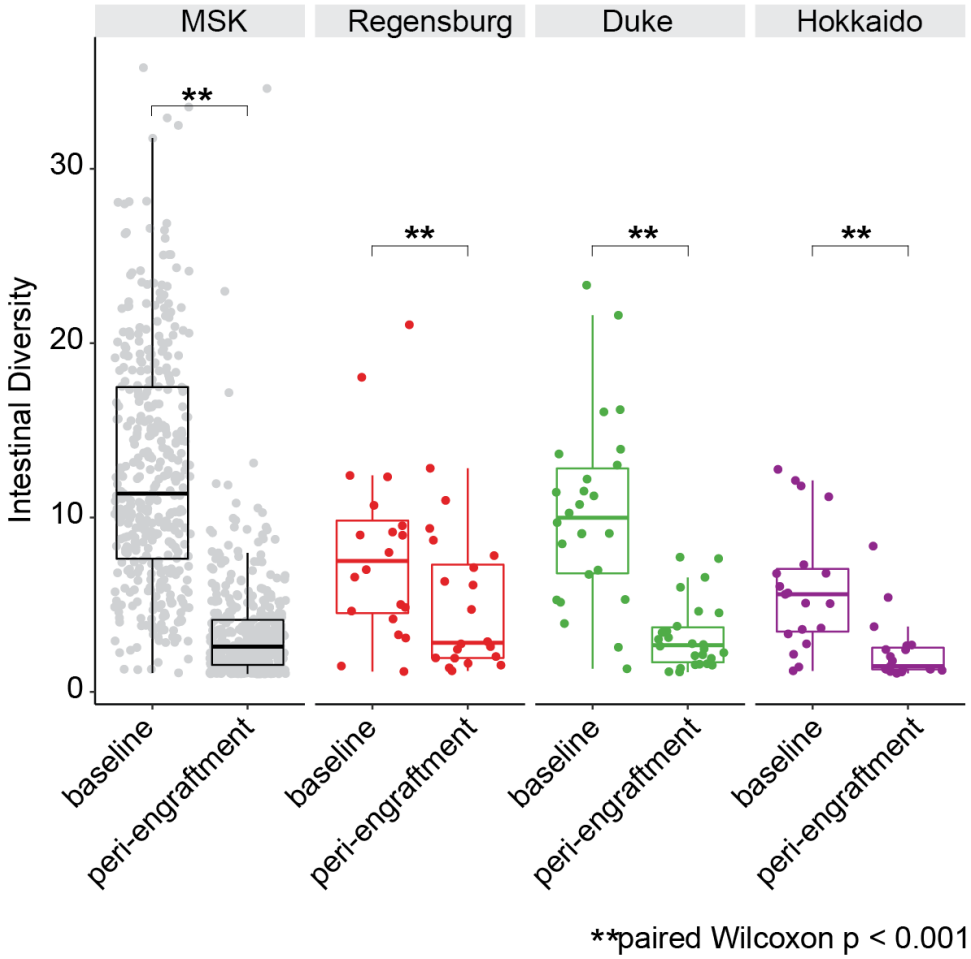
Observational study design considerations

This manuscript describes an observational cohort study and was prepared, where applicable, according to the reporting recommendations for observational studies (STROBE Statement).[32] The stool samples were collected prospectively (since 2009 at MSK, since 2011 at Regensburg, since 2012 at Duke, and since 2016 at Hokkaido, as detailed in **Supplemental Table S11**) with the goal of assembling biospecimen banks that would facilitate many different analyses. In this study, we first aimed to ask whether our previous single-center observation that low diversity after transplantation predicts poorer overall survival[2] would remain true with a much larger sample size and with patients transplanted in multiple centers. Once we observed that the association between peri-engraftment intestinal microbiota diversity and survival was reproducible (**Figure 1B-D**, **Table 2**), we commenced additional exploratory analyses that comprise the balance of the figures and tables herein.

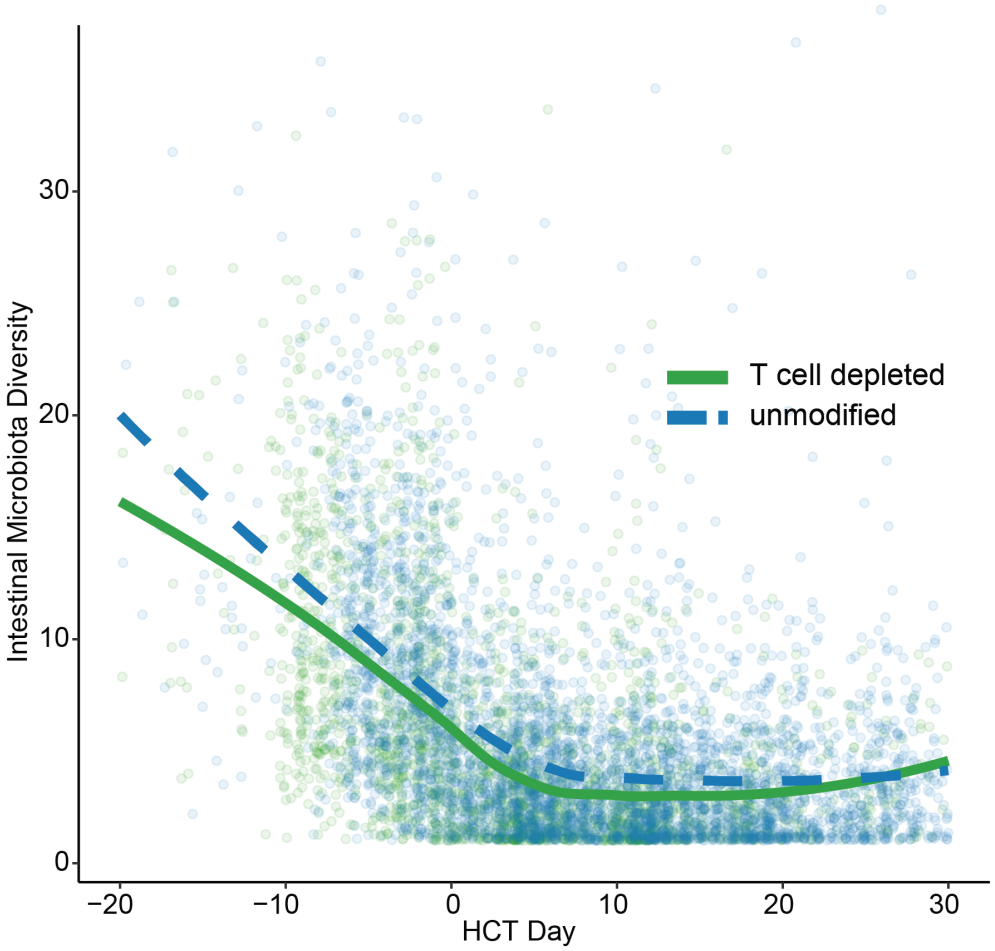
Supplemental Figure S1. Distribution of sample collections. (**A**) Unique patients from each cohort are plotted along the vertical axis against time relative to HCT on the horizontal axis. Each point is a sample, color-coded according to its α-diversity, as measured by the inverse Simpson index. (**B**) Histograms of sample collection frequency across time relative to HCT in each cohort. **(C)** Histogram of the number of samples per patient analyzed. This figure is associated with **Table 1, Supplemental Table S1**, and **Supplemental Table S2.**



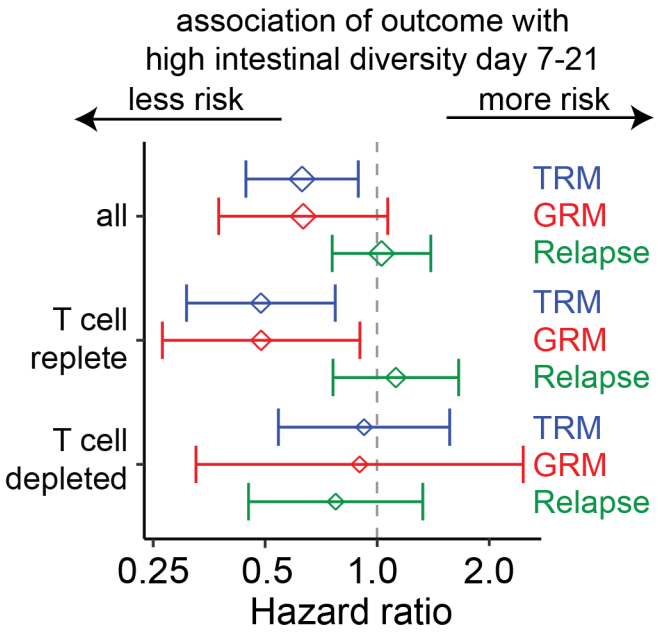
Supplemental Figure S2. A loss of intestinal diversity was observed at all four centers. Plotted are the inverse Simpson index values of fecal samples collected at baseline (earliest sample collected between day -30 and day -6) and the median values of each patient's samples that were collected in the peri-engraftment period (day 7-21). Only patients with at least one sample in each time period are plotted (N = 408, 20, 26, and 19 at MSK, Regensburg, Duke, and Hokkaido, respectively). At MSK, Regensburg, Duke, and Hokkaido diversity decreased 4.3-, 1.7-, 3.3-, and 2.5-fold, respectively. At each center, the reduction in diversity was significant by a paired Wilcoxon test ($p<0.001$).



Supplemental Figure S3. Diversity declines comparably in recipients of T-cell depleted and T-replete grafts. Intestinal microbiota diversity, as measured by 16S sequencing and the inverse Simpson index, declined comparably in 447 recipients of T-cell-depleted grafts as in 629 recipients of unmodified grafts (368 PBSC unmodified, 178 cord-blood, and 83 BM unmodified) at MSK.

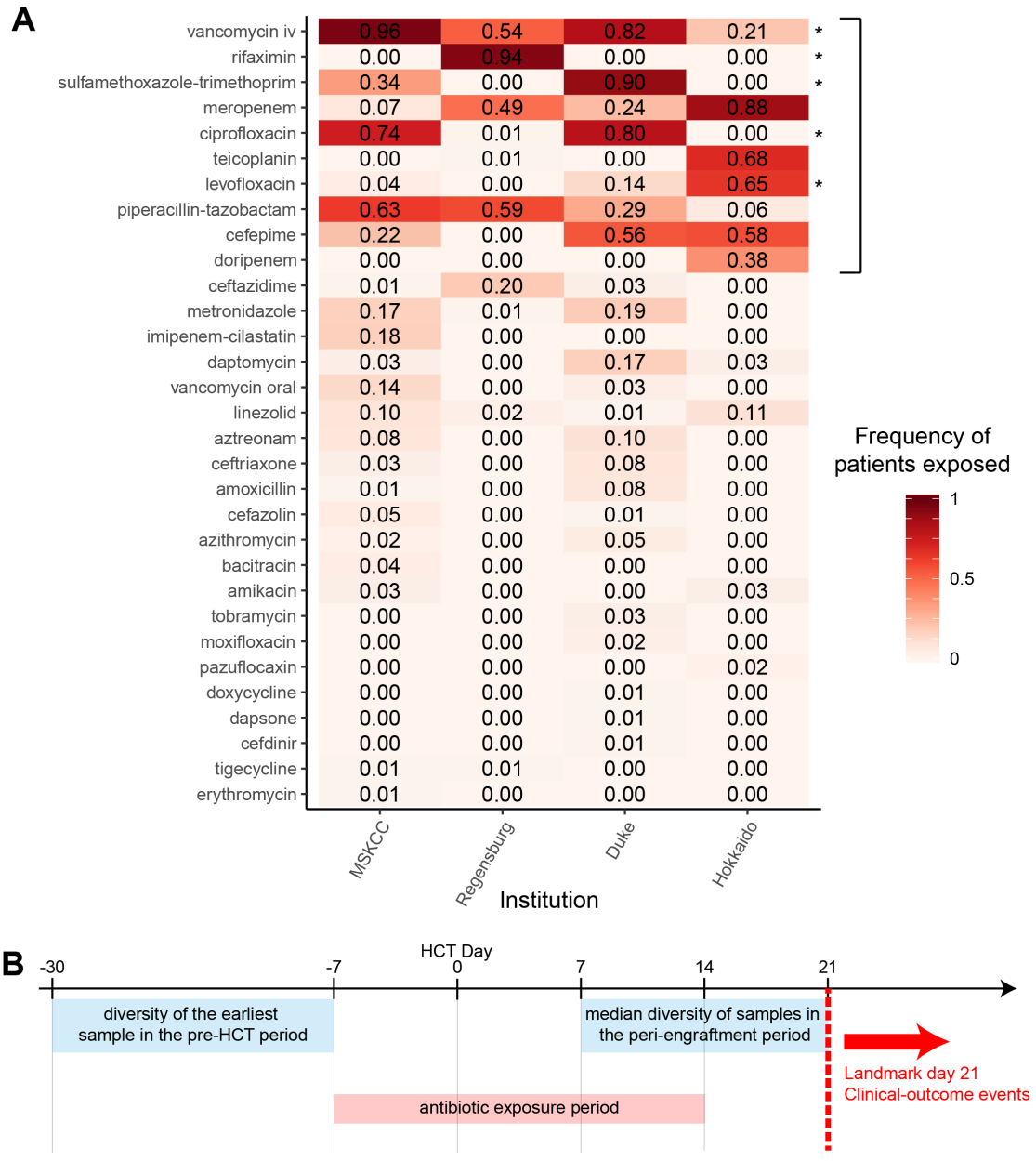


Supplemental Figure S4. Peri-neutrophil engraftment predicts TRM and GRM in recipients of T-replete grafts at MSK. In this subset analysis Forest plot of specific clinical outcomes, hazard ratios for the indicated outcome are plotted. The size of the diamond is proportional to the number of patients in each subset. Whiskers indicate 95% confidence intervals. Cumulative incidence curves of the same data are plotted in **Figure 1E**.



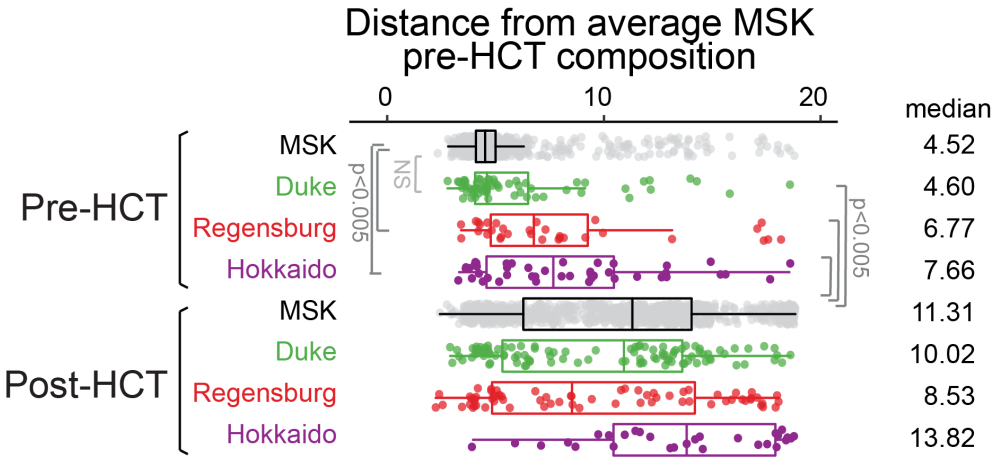
	HR (CI)	No. events	No. at risk
all TRM	0.63 (0.44-0.89)	134	703
all GRM	0.63 (0.37-1.07)	59	703
all Relapse	1.03 (0.76-1.39)	165	703
T cell replete TRM	0.49 (0.31-0.77)	76	428
T cell replete GRM	0.49 (0.26-0.9)	43	428
T cell replete Relapse	1.12 (0.76-1.66)	106	428
T cell depleted TRM	0.92 (0.54-1.57)	58	275
T cell depleted GRM	0.9 (0.33-2.47)	16	275
T cell depleted Relapse	0.77 (0.45-1.33)	59	275

Supplemental Figure S5. Antibiotic exposures. (**A**) Heatmap of drug exposures in the four cohorts. Each row is a drug and each column is one of the four institutions. The values and color-coding indicate the fraction of patients at each center who were exposed to at least one dose in the defined exposure window. The black bracket indicates drugs to which >20% of patients in at least one institution were exposed and were considered for this analysis; asterisks indicate drugs that were employed in prophylactic regimens in this population that were excluded. (**B**) Schematic of the sampling and exposure windows used to identify the drugs associated with a decrease in diversity from the pre-HCT period to the peri-engraftment period (**Supplemental Table S5**). Association of drug exposure with clinical outcomes (**Supplemental Table S6**) was analyzed in a landmark fashion following day 21 as in the rest of the manuscript.



Supplemental Figure S6. Statistical analysis of microbiome composition. (A) The variation in microbiota composition between centers is smaller than the magnitude of change observed during HCT. The variation in microbiota composition between centers is smaller than the magnitude of change observed during HCT. We defined as a reference point an averaged intestinal microbiota composition among samples collected pre-HCT at MSK (day –30 to day –1) Each point represents the distance of a single stool sample from this reference point, as measured by the Bray-Curtis (β-diversity) distance. Among pre-HCT samples, MSK and Duke had comparable distance to the reference (median distance of 4.52 and 4.60, respectively, p>0.05), Regensburg and Hokkaido samples were moderately farther (median distance of 6.77 and 7.66, respectively, p<0.005). In contrast, the median distance of post-HCT (day 0 to day +20) samples from the reference was markedly farther (11.31, 10.02, 8.53 and 13.82 for MSK, Duke, Regensburg and Hokkaido, respectively, p<0.005). Thus, the pre-HCT variation in microbiota composition across geography is smaller than the changes that occur over the course of transplantation. NS, not-significant. **(B)** In a generalized estimating equation with an independence working correlation structure for the binary endpoint of sample dominance, the odds of a observing a dominated sample from Regensburg or Duke was comparable to MSK. The odds of a sample from Hokkaido being dominated was higher than at MSK.

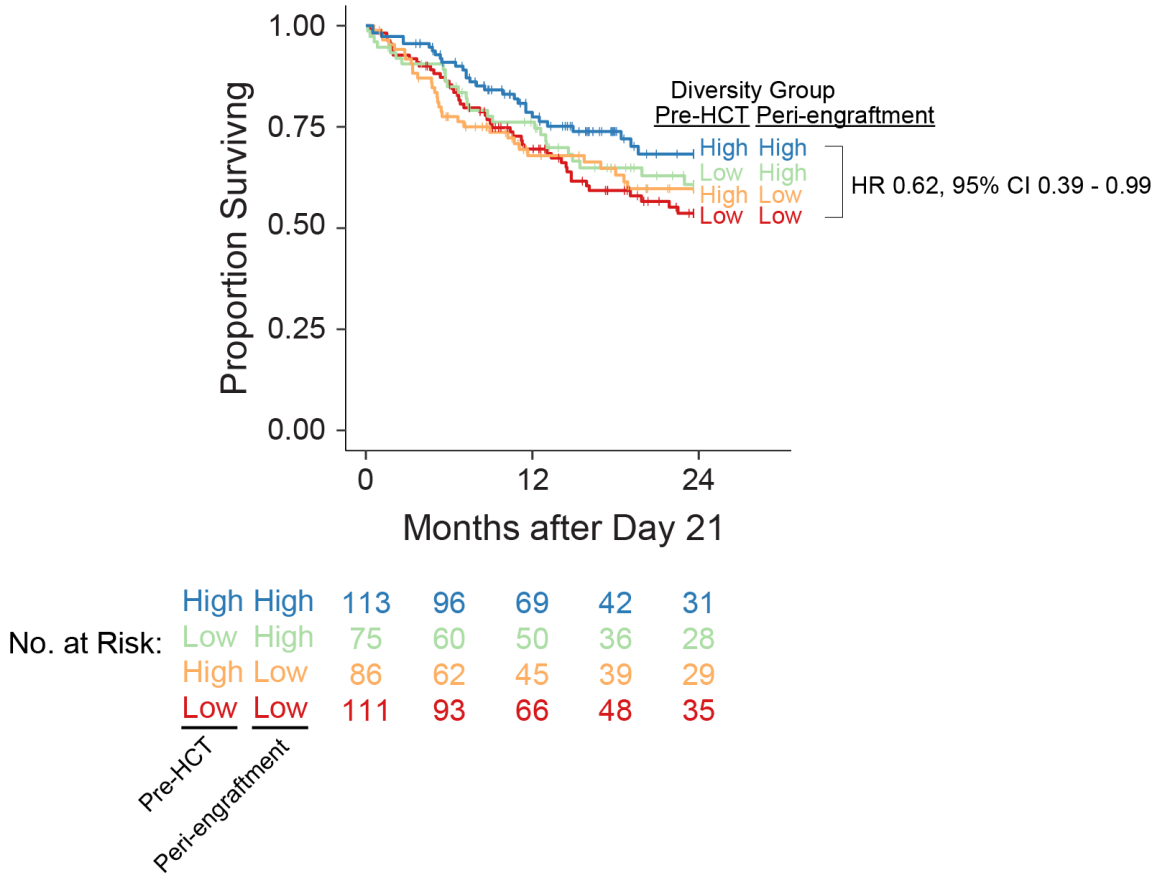
A



B

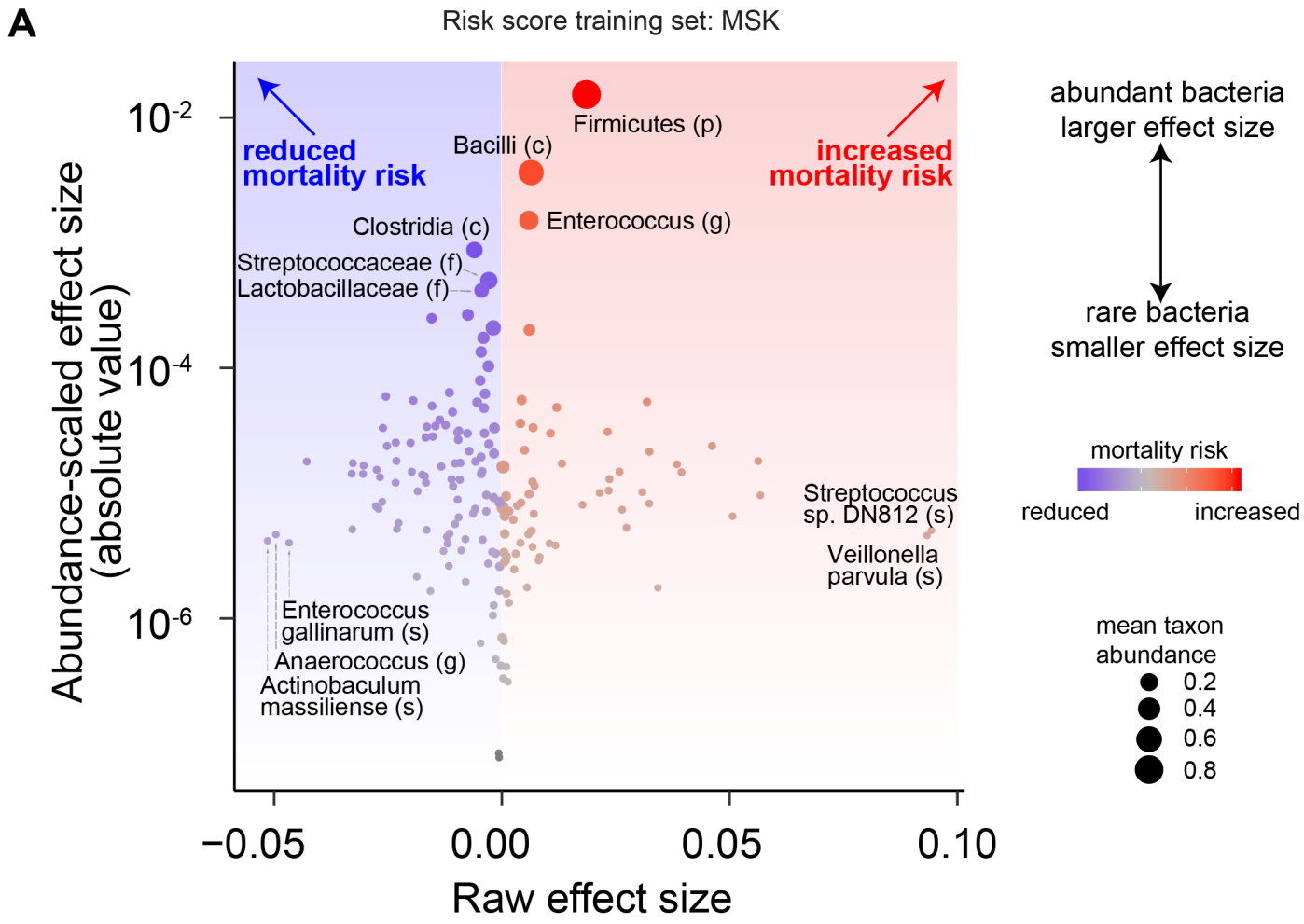
Variable		$N_{Samples}$	$N_{Patients}$	Odds Ratio (95% CI)	P-value
Time from HCT	Linear			1.17 (1.16-1.19)	<0.001
	Quadratic			0.995 (0.995-0.996)	<0.001
MSK		6324	1030	(reference)	
Duke		313	124	1.31 (0.95-1.81)	0.10
Regensburg		141	76	1.01 (0.58-1.75)	0.93
Hokkaido		106	58	2.70 (1.54-4.73)	<0.001

Supplemental Figure S7. Survival analysis by pre-HCT and peri-engraftment diversity. In this analysis within the MSK cohort, these four curves were not statistically different overall (p = 0.2), but the high-high group (blue) had statistically significantly lower risk of mortality than the low-low group (red) (HR 0.62 [0.39-0.99]). High-high group: 29 events in 113 patients. In the low-high group: 26 events in 75 patients. In the high-low group: 31 events in 86 patients. In the low-low group: 45 events in 111 patients. The inverse Simpson diversity cutoff thresholds to define pre-HCT and peri-engraftment groups were 11.2 and 2.66, respectively, as in the rest of the analysis. This analysis is accompanied by a multivariate Cox proportional hazards analysis in **Supplemental Table S8**.



Supplemental Figure S8. Risk score taxa. (**A**) The risk score for post-HCT mortality was computed as a function of the intestinal abundance of 172 bacterial taxa. The score was derived in the training (MSK) cohort using regularized regression. Each point is one of the 172 bacterial taxa, and the diameter of the points is proportional to the mean abundance of the taxon in the intestinal communities. Red indicates an association with increased mortality and blue indicates an association with decreased mortality. The overall risk score was computed as a weighted average of the taxa (where the weights were defined by the regularized Cox model) and was plotted on the horizontal axis. The average magnitude of each taxon's contribution to the overall risk score was plotted on the vertical axis by multiplying the estimated weight by the average abundance in the training cohort. For example, the class Bacilli (large red point near the top of the graph) had a relatively modest small effect size in the direction of predicting increased mortality risk, but due to its high abundance it made a large contribution to the overall risk score. In contrast, *Enterococcus gallinarum* (small blue point in the lower left) had a large effect size in the direction of predicting reduced mortality risk, but due to its low abundance it made a small contribution to overall risk score. The names of taxa with the largest effect sizes are annotated; the full list is tabulated in **Supplemental Table S9**. p, phylum; c, class; f, family; g, genus; s, species. (**B**) The risk score trained in the MSK cohort was tested in a validation cohort combined of patients from Duke + Regensburg + Hokkaido. The risk score association with patient overall survival was evaluated as a continuous variable in a multivariable model adjusted for age, conditioning intensity, graft source, and HCT-CI with stratification by institution. (**C**) Non-parametric estimate of the hazard rate for categories of the standardized risk score (plotted using the R package *bshazard*). Patients at the highest risk of death are in the highest-risk tertile; the difference in the hazard rate across the three groups is most prominent in the first year post-transplant and attenuates over time. 24 events in 59 patients in the highest-risk tertile, 16 events in 59 patients in the middle-risk tertile, 13 events in 60 patients in the lowest-risk tertile. **(D)** Concordance indices (C-index) for various models of mortality in the validation cohort combined of patients from Duke + Regensburg + Hokkaido. All models were stratified by institution.

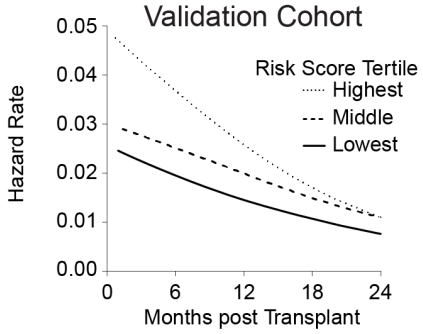
Supplemental Figure S8 (legend on preceding page)



B

Risk score validation cohort:
combined Duke + Regensburg + Hokkaido
HR 1.39 [1.02-1.91]
53 events in 178 patients

C



D

Model	c-statistic	SE
conditioning intensity	0.504	0.025
graft source	0.560	0.029
age	0.564	0.044
Risk score	0.566	0.035
age + graft source + conditioning intensity	0.592	0.049
Risk score + age + graft source + conditioning intensity	0.648	0.044

All models were stratified by institution. SE, Standard error

Patient Flow through the Study	Overall		MSK		Regensburg		Duke		Hokkaido	
	Patients	Samples	Patients	Samples	Patients	Samples	Patients	Samples	Patients	Samples
- evaluable* samples collected ≥ day –30 of a first allo-HCT - exclude samples collected after FMT** - exclude samples collected ≥ day –10 of a subsequent (second) allo-HCT										
Cohort for microbiota composition and dynamics analysis (Figures 1A, 2)	**1362**	8767	**1076**	7929	**78**	143	**142**	479	**66**	216
- exclude patients randomized to FMT** • include only patients with samples collected day 7-21 - exclude patients who died ≤ day 21										
Cohort for regularized regression (Supplemental Figure S8)	**947**	2471	**769**	2228	**56**	73	**84**	121	**38**	49
- exclude patients previously analyzed in Taur *Blood* 2014										
Cohort for landmark analysis of peri-engraftment diversity clinical outcomes (Figure 1B-F)	**882**	2366	**704**	2123	**56**	73	**84**	121	**38**	49
• include only first sample for patients with samples collected between days –30 and –6										
Cohort for pre-HCT diversity analysis (Figure 3)	**606**	606	**501**	501	**28**	28	**40**	40	**37**	37
• include patients with samples collected day between days 7 and 21 or day –30 and –6 - include only first sample for patients with samples collected between days –30 and –6										
Cohort for survival analysis of combined pre-HCT & peri-engraftment diversity (Table S10)	**385**	1614	**385**	1614						

Supplemental Table S1. Patient flow through the study (CONSORT Table).

* successfully amplified and sequenced with ≥ 200 reads/sample

** FMT; fecal microbiota transplantation on study NCT02269150 and Taur *Science Transl Med* 2018

Supplemental Table S2. Numbers of samples per patient. Summary of the number of serial samples collected per patient in the overall cohort and at each center. For example, 2 serial samples were analyzed from each of 200 patients (15% of the overall cohort), of whom 133 were from MSK (12% of the MSK cohort). When the overall cohort is ranked according to the number of samples per patient, the minimum number of samples per patient was 1 sample/patient. The 25th percentile was 2 samples/patient. The median (50th percentile) was 4 samples/patient. The 75th percentile was 8 samples/patient. The mean number of samples per patient was 6.4.

N samples	Overall		MSK		Regensburg		Duke		Hokkaido	
per patient	N patients	Percent	N patients	Percent	N patients	Percent	N patients	Percent	N patients	Percent
1	178	0.13	94	0.09	33	0.42	38	0.27	13	0.20
2	200	0.15	133	0.12	30	0.38	30	0.21	7	0.11
3	206	0.15	162	0.15	10	0.13	19	0.13	15	0.23
4	184	0.14	143	0.13	5	0.06	20	0.14	16	0.24
5	117	0.09	93	0.09			13	0.09	11	0.17
6	76	0.06	67	0.06			6	0.04	3	0.05
7	55	0.04	50	0.05			4	0.03	1	0.02
8	41	0.03	38	0.04			3	0.02		
9	43	0.03	40	0.04			3	0.02		
10	24	0.02	22	0.02			2	0.01		
11	20	0.02	16	0.01			4	0.03		
12	20	0.02	20	0.02						
13	21	0.02	21	0.02						
14	19	0.01	19	0.02						
15	21	0.02	21	0.02						
16	14	0.01	14	0.01						
17	14	0.01	14	0.01						
18	12	0.01	12	0.01						
19	9	0.01	9	0.01						
≥20	88	0.06	88	0.08						

Supplemental Table S3. Institutional antibiotic clinical practices

Scenario	MSK	Duke	Regensburg	Hokkaido
ppx during leukemia induction	cipro	levo	cipro	levo
oral decontamination prior to allo-HCT	none	none	rifaximin from the day of conditioning until d+21 (since 2012)	none
antibacterial ppx peri-HCT	NMA: cipro PO or IV start d–2 RIC/MA: cipro PO or IV + vanco IV start d–2	cipro PO with start of conditioning	rifaximin with start of conditioning (since 2012)	levo PO with start of conditioning
duration of ppx abx	cipro until initiation of empiric abx for F&N or engraftment vanco until d+7	cipro until initiation of empiric abx for F&N. Upon engraftment: NMA: stop, or for alemtuzumab recipients resume until 6-months post-HCT. MA: stop MA cord: resume until 6 months post-HCT	continuation of rifaximin concurrently with empiric abx for F&N through engraftment.	levo until initiation of empiric abx for F&N or until calcineurin inhibitors are converted from IV to PO (d~30)
first-line empiric F&N	pip/tazo	cefepime	pip/tazo	cefepime
alternative F&N drug in penicillin-allergic patients	non-severe allergy: cefepime* severe allergy: vanco/aztreonam*	non-severe allergy: cefepime* severe allergy: aztreonam*	non-severe allergy: ceftazidime* severe allergy: meropenem*	meropenem or pip/tazo*
duration of empiric F&N abx with no source identified	until engraftment	until engraftment or de-escalation to ppx regimen at clinician's discretion	until CRP normal and no fever for ≥3 days; independent of engraftment	until engraftment

Supplemental Table S4. Numbers of Patients at Risk. Number at risk at 3-month intervals for Kaplan-Meier and cumulative-incidence plots.

Figure	Cohort	Sampling Period	Outcome	Diversity Group	Median Cutoff	Months after Day 21								
						0	3	6	9	12	15	18	21	24
1B	MSK	peri-engraftment	all-cause mortality	Low	≤2.64	350	317	281	235	204	184	164	143	129
				High	>2.64	354	320	289	251	220	188	159	135	116
1C	Reg+Duk+Hok	peri-engraftment	all-cause mortality	Low	≤2.64	92	72	57	43	37	30	24	20	15
				High	>2.64	87	74	60	55	44	37	34	29	26
1D	MSK	peri-engraftment	transplant-related mortality	Low	≤2.64	349	297	251	213	175	154	139	121	109
				High	>2.64	354	303	250	215	185	161	138	117	100
			relapse/progression of disease	Low	≤2.64	349	297	251	213	175	154	139	121	109
				High	>2.64	354	303	250	215	185	161	138	117	100
1E	MSK, T-cell replete	peri-engraftment	all-cause mortality	Low	≤2.64	185	161	140	115	103	93	82	69	62
				High	>2.64	244	219	198	172	153	134	108	88	73
			transplant-related mortality	Low	≤2.64	184	146	124	104	90	80	73	60	52
				High	>2.64	244	204	167	145	125	110	90	74	61
			GVHD-related mortality	Low	≤2.64	184	146	124	104	90	80	73	60	52
				High	>2.64	244	204	167	145	125	110	90	74	61
			relapse/progression of disease	Low	≤2.64	184	146	124	104	90	80	73	60	52
				High	>2.64	244	204	167	145	125	110	90	74	61
	MSK, T-cell depleted	peri-engraftment	all-cause mortality	Low	≤2.64	165	156	141	120	101	91	82	74	67
				High	>2.64	110	101	91	79	67	54	51	47	43
			transplant-related mortality	Low	≤2.64	165	151	127	109	85	74	66	61	57
				High	>2.64	110	99	83	70	60	51	48	43	39
			GVHD-related mortality	Low	≤2.64	165	151	127	109	85	74	66	61	57
				High	>2.64	110	99	83	70	60	51	48	43	39
			relapse/progression of disease	Low	≤2.64	165	151	127	109	85	74	66	61	57
				High	>2.64	110	99	83	70	60	51	48	43	39
3C	MSK	pre-HCT	all-cause mortality	Low	≤11.2	251	223	211	178	160	135	120	103	92
				High	>11.2	250	237	208	181	155	139	120	98	91
S8	MSK	pre-HCT & peri-engraftment combined	all-cause mortality											
			Pre-HCT / Peri-engraftment											
			Low / Low			111	102	93	77	66	54	48	40	35
			Low / High			75	66	60	53	50	40	36	31	28
			High / Low			86	78	62	54	45	44	39	32	29
			High / High			113	108	96	81	69	57	42	34	31

Supplemental Table S5. Exposure to piperacillin-tazobactam or meropenem is associated with a decrease in diversity during allo-HCT. Exposure to pip-tazo and meropenem are significantly associated with changes in intestinal microbiota diversity. A linear regression model was constructed to evaluate the association of drug exposures with the change in microbiota diversity during HCT (see Supplemental Text for details). The estimates and the corresponding standard errors (S.E.) represent the decrease in diversity (in Simpson reciprocal units) between the baseline pre-HCT sample to the peri-engraftment period as a function of exposure duration (in number of days) during a window of day -7 to 14. The model also accounts for the time duration, Δt, between each patient's peri-engraftment and pre-HCT samples. A schematic for the design of this analysis is in **Supplemental Figure S5**. The key antibiotic exposures are shaded and used in a survival analysis in **Supplemental Table S6**.

	MSKCC Estimate ± S.E.	Regensburg Estimate ± S.E.	Duke Estimate ± S.E.	Hokkaido Estimate ± S.E.
Δt	**-0.33 ± .02**	-0.10 ± 0.17	**-0.15 ± 0.06**	-0.01 ± 0.06
cefepime	0.07 ± .07	-	-0.13 ± 0.22	0.18 ± 0.19
doripenem	-	-	-	0.11 ± 0.23
meropenem	0.08 ± 0.21	-0.02 ± 0.24	-0.51 ± 0.33	**-0.36 ± 0.15**
piperacillin-tazobactam	**-0.13 ± 0.05**	-0.24 ± se 0.29	-0.35 ± 0.28	-
teicoplanin	-	-	-	-0.15 ± 0.12

Supplemental Table S6. Diversity and survival remain significantly associated in multivariable models adjusted for exposure to key antibiotics. Multivariable Cox proportional hazards analyses of the association of peri-engraftment intestinal diversity (median of samples collected day +7 to +21) with overall survival at each institution. The multivariate models were stratified by institution and adjusted for age, conditioning intensity, graft source, the hematopoietic cell transplantation comorbidity index (HCT-CI), and duration of exposure to the two key antibiotics identified in **Supplemental Figure S5** and **Supplemental Table S5**. Intestinal diversity was measured by the inverse Simpson (S) index and is considered here separately as either a $\log_{10}$-transformed continuous variable or a median-stratified binary variable. See **Table 2** for univariate results.

	Multivariate I log transformed HR(95% CI)	Multivariate II two groups HR(95% CI)
MSK		
log(S); 240 events in 704 patients	0.59 (0.36 to 0.95)	
S: two-groups		
S≤2.64; 136 events in 350 patients		
S>2.64; 104 events in 354 patients		0.76 (0.58 to 0.99)
Regensburg + Duke + Hokkaido		
log(S); 53 events in 179 patients	0.45 (0.16 to 1.23)	
S: two-groups		
S≤2.64; 35 events in 92 patients		
S>2.64; 18 events in 87 patients		0.55 (0.30 to 1.01)

Supplemental Table S7. Multivariate analysis of pre-HCT diversity at MSK. Multivariate Cox proportional hazards analysis of the association of pre-HCT (first sample collected day -30 to day -6) intestinal diversity with overall survival at MSK. Intestinal diversity was measured by the inverse Simpson (S) index and is considered here separately as either a log-transformed continuous variable or a median-stratified binary variable. The model was adjusted for age, conditioning intensity, graft source, and the hematopoietic cell transplantation comorbidity index (HCT-CI).

	Univariate	Multivariate I log transformed	Multivariate II two groups
Parameters	HR(95% CI)	HR(95% CI)	HR(95% CI)
log(S); 173 events in 501 patients	0.38 (0.22 to 0.64)	0.41 (0.24 to 0.71)	
S: two-groups			
S≤11.18; 101 events in 251 patients	(reference)		
S>11.18; 72 events in 250 patients	0.71 (0.52 to 0.96)		0.74 (0.54 to 1.01)

Parameters	Univariate	Multivariate Analyses			
pre-HCT:		continuous		2-groups	
peri-engraftment:		continuous	2-groups	continuous	2-groups
	HR(95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
peri-engraftment S: continuous	0.53 (0.35 to 0.80)	0.42 (0.21 to 0.83)		0.41 (0.20 to 0.81)	
peri-engraftment S: 2-groups					
S≤2.66	(reference)				
S>2.66	0.71 (0.56 to 0.91)		0.80 (0.56 to 1.15)		0.80 (0.55 to 1.16)
pre-HCT S: continuous	0.53 (0.29 to 0.97)	0.59 (0.30 to 1.15)	0.55 (0.28 to 1.05)		
pre-HCT S: 2-groups					
S≤11.18	(reference)				
S>11.18	0.81 (0.57 to 1.14)			0.90 (0.63 to 1.28)	0.86 (0.60 to 1.23)

Supplemental Table S8. Multivariate analysis of pre-HCT and peri-engraftment diversity at MSK. Multivariate Cox proportional hazards analysis of the association of overall survival at MSK according to both pre-HCT and peri-engraftment intestinal diversity. Intestinal diversity was measured by the inverse Simpson (S) index and is considered in each sampling window both as a log-transformed continuous variable and separately as a median-stratified two-group variable. Pre-HCT samples were the first sample collected day -30 to day -6. Peri-engraftment values are the median of samples collected day +7 to +21. The models were also adjusted for age, conditioning intensity, graft source and HCT-CI. For the continuous-variable analysis there were 131 events in 385 patients. For the binary analysis: in the high-diversity peri-engraftment group there were 55 events in 188 patients and in low-diversity peri-engraftment group there were 76 events in 197 patients. This table is accompanied by the survival curves in **Supplemental Figure S7**.

Supplemental Table S9. Taxa in the Risk Score. Effect size is the coefficient of each term in the model. Positive values indicate increased mortality risk; negative values decreased mortality risk. The abundance columns tabulate the abundance of each taxon at MSK day 7-21. Score contributions are the effect sizes scaled according to taxon abundance. k, kingdom; p, phylum; c, class; o, order; f, family; s, species. An electronic version of this table is available upon request.

tax_group	effect size	abudance (mean)	abudance (median)	score contribution (mean)	score contribution (median)
k__Bacteria|p__Firmicutes	0.01858226	0.824897124	0.921632412	0.015328453	0.017126013
k__Bacteria|p__Firmicutes|c__Bacilli	0.006419545	0.567933871	0.628757696	0.003645877	0.004036338
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Enterococcaceae|g__Enterococcus	0.00593896	0.254442086	0.019330415	0.001511121	0.000114803
k__Bacteria|p__Firmicutes|c__Clostridia	-0.006039346	0.144839721	0.052605143	-0.000874737	-0.000317701
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae	-0.002878403	0.173676107	0.048320566	-0.00049991	-0.000139086
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae	-0.004464276	0.093222578	0.011639543	-0.000416171	-5.20E-05
k__Bacteria|p__Actinobacteria	-0.00744442	0.035648491	0.009017625	-0.000265382	-6.71E-05
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Pediococcus|s__Pediococcus_acidilactici	-0.015405481	0.016209037	0.000117082	-0.000249708	-1.80E-06
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales	-0.001862565	0.112109806	0.007344043	-0.000208812	-1.37E-05
k__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales	0.006044613	0.033259648	0.000757006	0.000201042	4.58E-06
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Coprobacillaceae	-0.004016455	0.043266577	0.001121915	-0.000173778	-4.51E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae	-0.004524725	0.02969127	0.004174668	-0.000134345	-1.89E-05
k__Bacteria|p__Verrucomicrobia|c__Verrucomicrobiae|o__Verrucomicrobiales|f__Verrucomicrobiaceae|g__Akkermansia|s__Akkermansia_muciniphila	-0.002929923	0.035142525	0.000223964	-0.000102965	-6.56E-07
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacteriales|f__Enterobacteriaceae|g__Escherichia|s__Escherichia_coli	-0.004774801	0.016593926	0.000231102	-7.92E-05	-1.10E-06
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus_mitis	-0.011526578	0.005513562	0.000499685	-6.36E-05	-5.76E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Veillonellaceae	-0.003676564	0.016938698	0.001986583	-6.23E-05	-7.30E-06
k__Bacteria|p__Proteobacteria|c__Deltaproteobacteria|o__Desulfovibrionales|f__Desulfovibrionaceae|g__Bilophila|s__NA	-0.025467042	0.002326579	0	-5.93E-05	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacteriales|f__Enterobacteriaceae|g__Klebsiella|s__ambiguous_Klebsiella	0.004342555	0.012784667	7.62E-05	5.55E-05	3.31E-07
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Actinomycetales|f__Actinomycetaceae|g__Parascardovia	-0.019446787	0.00282671	0.000301341	-5.50E-05	-5.86E-06
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_delbrueckii_subsp._bulgaricus	0.031868701	0.001682892	0	5.36E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus_salivarius	-0.005419843	0.009825872	0.000174617	-5.33E-05	-9.46E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Coriobacteriales|f__Coriobacteriaceae	-0.015337341	0.003241494	0.00032	-4.97E-05	-4.91E-06
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Coprobacillaceae|g__Coprobacillus	0.012009739	0.004030127	5.53E-05	4.84E-05	6.64E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae	-0.003923031	0.012230883	0.000626953	-4.80E-05	-2.46E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Veillonellaceae|g__Phascolarctobacterium|s__Phascolarctobacterium_faecium	-0.010806242	0.004116472	1.26E-05	-4.45E-05	-1.36E-07
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_plantarum	-0.013635147	0.002831401	0	-3.86E-05	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Porphyromonadaceae|g__Parabacteroides	0.004060634	0.00892424	8.98E-05	3.62E-05	3.64E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_tertium	-0.012353841	0.002834322	8.07E-05	-3.50E-05	-9.97E-07
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Erysipelotrichaceae|g__Eubacterium|s__Eubacterium_biforme	-0.014566811	0.002362044	0	-3.44E-05	0
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria	-0.01646401	0.002061079	1.32E-05	-3.39E-05	-2.17E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae|g__Ruminococcus	0.00685218	0.004881552	0.000363719	3.34E-05	2.49E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae	-0.001632194	0.020347078	0.001425545	-3.32E-05	-2.33E-06
k__Bacteria|p__Firmicutes|c__Bacilli|o__Gemellales|f__Gemellaceae|g__Gemella|s__Gemella_haemolysans	-0.026128407	0.001269416	4.01E-05	-3.32E-05	-1.05E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Dorea	-0.009453025	0.003329931	0.000202194	-3.15E-05	-1.91E-06
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus_mutans	0.023236318	0.001330146	3.06E-05	3.09E-05	7.11E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Coprococcus	-0.00970865	0.003144756	0.00022416	-3.05E-05	-2.18E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae|g__Ruminococcus|s__NA	0.010658758	0.002827952	0.000161935	3.01E-05	1.73E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae|g__Clostridium	-0.003787824	0.007941007	0.000241955	-3.01E-05	-9.16E-07
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Actinomycetales|f__Micrococcaceae|g__Rothia	-0.00755143	0.003982853	0.000264046	-3.01E-05	-1.99E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Veillonellaceae|g__Megasphaera	-0.015190475	0.001869936	2.81E-05	-2.84E-05	-4.27E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Coriobacteriales|f__Coriobacteriaceae|g__Atopobium	-0.016713979	0.001666346	6.26E-05	-2.79E-05	-1.05E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae|g__Clostridium|s__Clostridium_glycolicum	-0.009577513	0.00279757	3.55E-05	-2.68E-05	-3.40E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Dorea|s__Coprococcus_comes	-0.023292119	0.001088898	6.71E-05	-2.54E-05	-1.56E-06
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Coprobacillaceae|g__[Clostridium]|s__[Clostridium]_spiroforme	-0.020047616	0.001258357	1.37E-05	-2.52E-05	-2.75E-07
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_fermentum	-0.00280125	0.008803936	0.000159026	-2.47E-05	-4.45E-07
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Enterococcaceae|g__Enterococcus|s__Enterococcus_rivorum	0.046156253	0.000517621	0.00016474	2.39E-05	7.60E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._826	-0.025198899	0.000944941	0	-2.38E-05	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacteriales|f__Enterobacteriaceae|g__Enterobacter|s__Enterobacter_ludwigii	0.004995862	0.004426714	0	2.21E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_animalis	-0.007165477	0.0030297	0	-2.17E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_leptum	0.032407265	0.000659947	0	2.14E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Ruminococcus	-0.001709083	0.012144775	0.00070007	-2.08E-05	-1.20E-06
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus_parasanguinis	-0.004544749	0.004294704	0.000267415	-1.95E-05	-1.22E-06
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._ID5	-0.023123975	0.00078342	0	-1.81E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Enterococcaceae|g__Enterococcus|s__Enterococcus_lactis	0.056278911	0.00032088	7.39E-05	1.81E-05	4.16E-06
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Carnobacteriaceae|g__Granulicatella|s__Granulicatella_adiacens	-0.005756792	0.003125136	9.35E-05	-1.80E-05	-5.38E-07
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Actinomycetales|f__Actinomycetaceae|g__Parascardovia|s__Actinomyces_graevenitzii	-0.042821836	0.00041826	0	-1.79E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae|g__Clostridium|s__Clostridium_bartlettii_DSM_16795	-0.009155058	0.001909506	7.62E-05	-1.75E-05	-6.97E-07
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__S24-7|g__NA|s__NA	-0.032682203	0.000531859	0	-1.74E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_algidixylanolyticum	-0.010264597	0.001688164	0	-1.73E-05	0
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Bifidobacteriales|f__Bifidobacteriaceae|g__Bifidobacterium|s__Bifidobacterium_dentium	0.013205528	0.001306544	0	1.73E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Eubacteriaceae|g__Eubacterium|s__Eubacterium_limosum	0.038413129	0.000443405	0	1.70E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_homohiochii	-0.030332547	0.000548572	0	-1.66E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Dorea|s__Tyzzerella_nexilis	-0.013301833	0.001229924	1.60E-05	-1.64E-05	-2.13E-07
k__Bacteria|p__Proteobacteria	0.000277319	0.058590422	0.002054986	1.62E-05	5.70E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Mogibacterium|s__Mogibacterium_neglectum	-0.027487838	0.000559892	0	-1.54E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_reuteri	-0.004344434	0.003443281	5.58E-05	-1.50E-05	-2.42E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_saccharogumia	-0.019748414	0.000754613	1.18E-05	-1.49E-05	-2.33E-07
k__Bacteria|p__Tenericutes|c__Mollicutes|o__Mycoplasmatales|f__Mycoplasmataceae|g__Mycoplasma	0.025871615	0.000576	0	1.49E-05	0
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Erysipelotrichaceae|g__Holdemania|s__Holdemania_filiformis	0.039453045	0.000373683	0	1.47E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Coprococcus|s__[Eubacterium]_hallii	-0.032948136	0.000435495	1.88E-05	-1.43E-05	-6.20E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Ruminococcus|s__Drancourtella_massiliensis	-0.004534883	0.003151982	7.56E-05	-1.43E-05	-3.43E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_perfringens	-0.030475093	0.000467051	0	-1.42E-05	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_stercoris_ATCC_43183	-0.017494164	0.000803343	0	-1.41E-05	0

Supplemental Table S9, continued

tax_group	effect size	abudance (mean)	abudance (median)	score contribution (mean)	score contribution (median)
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Ruminococcus|s__Ruminococcus_faecis	-0.016907813	0.000809283	2.55E-06	-1.37E-05	-4.32E-08
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._PPf35E6	-0.026744177	0.00050289	0	-1.34E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Lactococcus|s__Lactococcus_piscium	0.02364148	0.000550045	0	1.30E-05	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_uniformis	-0.01104081	0.001171065	0	-1.29E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Leuconostocaceae|g__Weissella|s__Weissella_confusa	-0.009235372	0.001391195	7.56E-06	-1.28E-05	-6.98E-08
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Blautia|s__NA	0.006753751	0.001824627	6.83E-05	1.23E-05	4.61E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Veillonellaceae|g__Megasphaera|s__Megasphaera_micronuciformis	-0.016669157	0.00073024	0	-1.22E-05	0
k__Bacteria|p__Actinobacteria|c__Actinomycetales|o__Propionibacterineae|f__Propionibacteriaceae|g__Propionibacterium	-0.023297247	0.000521282	2.92E-05	-1.21E-05	-6.81E-07
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_thetaiotaomicron	0.007196619	0.001594384	0	1.15E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae|g__Peptostreptococcus|s__Peptostreptococcus_sp._MDA2346-2	-0.010705236	0.001065977	2.86E-06	-1.14E-05	-3.06E-08
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Coprobacillaceae|g__Coprobacillus|s__Massiliomicrobiota_timonensis	0.023451125	0.000446903	0	1.05E-05	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_fragilis	-0.01848356	0.000560736	0	-1.04E-05	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Blautia|s__Blautia_hydrogenotrophica	0.030831158	0.00033113	0	1.02E-05	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Enterococcaceae|g__Enterococcus|s__Enterococcus_mundtii	0.02149635	0.00046799	4.58E-06	1.01E-05	9.85E-08
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Prevotellaceae|g__Prevotella|s__Prevotella_melaninogenica	0.005897007	0.001674424	0	9.87E-06	0
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacteriales|f__Enterobacteriaceae|g__Dickeya|s__Erwinia_chrysanthemi	0.00609968	0.00161463	0	9.85E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae|g__Oscillospira	-0.003390357	0.002853402	0.00014242	-9.67E-06	-4.83E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_nexile_DSM_1787	0.056765607	0.000169357	0	9.61E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Veillonellaceae|g__Veillonella	0.000953008	0.009945513	0.000699475	9.48E-06	6.67E-07
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_gasseri	-0.001819181	0.005098767	0.000174871	-9.28E-06	-3.18E-07
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Actinomycetales|f__Micrococcaceae|g__Rothia|s__Rothia_dentocariosa	-0.009695668	0.000916299	3.09E-05	-8.88E-06	-3.00E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sporosphaeroides	-0.026266655	0.000325531	0	-8.55E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Lactococcus	-0.000491388	0.017314986	0.000385607	-8.51E-06	-1.89E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae|g__Clostridium|s__Clostridium_difficile	0.00421071	0.001998454	0	8.41E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Blautia|s__[Ruminococcus]_obeum	0.032359558	0.000256023	0	8.28E-06	0
k__Bacteria|p__Cyanobacteria|c__Chloroplast|o__Streptophyta|f__NA|g__NA|s__NA	0.017714261	0.000456981	0	8.10E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Ruminococcus|s__NA	0.003414661	0.002326911	4.04E-05	7.95E-06	1.38E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_hathewayi	-0.027665194	0.000284837	0	-7.88E-06	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales	0.000183614	0.041594069	0.001761639	7.64E-06	3.23E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_methylpentosum_DSM_5476	-0.027051702	0.000277366	0	-7.50E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._cTPY-17	-0.005849082	0.001279527	0	-7.48E-06	0
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Actinomycetales|f__Actinomycetaceae|g__Parascardovia|s__NA	0.026407746	0.000279444	0	7.38E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_clostridioforme	0.001570447	0.004570813	0.000204983	7.18E-06	3.22E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._enrichment_culture_clone_NHT38	-0.00340411	0.002103315	0.000236834	-7.16E-06	-8.06E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Ruminococcus|s__Ruminococcus_torques	-0.006314759	0.001094621	3.24E-05	-6.91E-06	-2.05E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Blautia|s__Blautia_obeum	0.007005156	0.000978505	2.62E-05	6.85E-06	1.84E-07
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Enterobacteriales|f__Enterobacteriaceae|g__Klebsiella|s__Klebsiella_oxytoca	0.000617388	0.010656567	2.37E-05	6.58E-06	1.46E-08
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Erysipelotrichaceae|g__Bulleidia|s__Bulleidia_moorei	0.050648001	0.000129261	0	6.55E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._cf3-PUG	-0.009510236	0.000676862	0	-6.44E-06	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Porphyromonadaceae|g__Parabacteroides|s__Parabacteroides_merdae	0.002530365	0.002428855	0	6.15E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_buchneri	-0.022646512	0.000255484	0	-5.79E-06	0
k__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rickettsiales|f__mitochondria|g__NA|s__NA	-0.010271271	0.000551264	0	-5.66E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Butyrivibrio|s__Shuttleworthia_satelles	0.027346726	0.000194476	0	5.32E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Coprococcus|s__NA	-0.023044804	0.000224485	0	-5.17E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Bacillaceae	-0.032812453	0.000156982	0	-5.15E-06	0
k__Bacteria|p__Fusobacteria|c__Fusobacteriia|o__Fusobacteriales|f__Fusobacteriaceae|g__Fusobacterium|s__Fusobacterium_nucleatum	-0.016488666	0.000308318	0	-5.08E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus_sp._DN812	0.094292902	5.37E-05	0	5.06E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_cellulosi	0.006604899	0.000758653	0	5.01E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_acidophilus	0.006360097	0.000783952	0	4.99E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Coriobacteriales|f__Coriobacteriaceae|g__Eggerthella|s__Eggerthella_lenta	-0.011469671	0.0004157	0	-4.77E-06	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_ovatus	0.000615364	0.007683304	6.37E-05	4.73E-06	3.92E-08
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_hylemonae	0.005819405	0.000808336	0	4.70E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Anaerococcus	-0.04956495	9.43E-05	0	-4.68E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Veillonellaceae|g__Veillonella|s__Veillonella_parvula	0.0933585	4.93E-05	0	4.60E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._Culture-54	-0.012060995	0.000371128	0	-4.48E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Coprococcus|s__Anaerostipes_caccae	-0.007980946	0.000537459	0	-4.29E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Blautia|s__Blautia_faecis	-0.004243403	0.00100395	3.30E-05	-4.26E-06	-1.40E-07
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Actinomycetales|f__Actinomycetaceae|g__Actinobaculum|s__Actinobaculum_massiliense	-0.051428745	8.11E-05	0	-4.17E-06	0
k__Bacteria|p__Firmicutes|c__Erysipelotrichi|o__Erysipelotrichales|f__Coprobacillaceae|g__Coprobacillus|s__Longibaculum_muris	0.004070436	0.000986397	0	4.02E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Enterococcaceae|g__Enterococcus|s__Enterococcus_gallinarum	-0.04670738	8.58E-05	0	-4.01E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Butyrivibrio	0.010451853	0.000379729	0	3.97E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_sp._MSTE9	-0.011890152	0.000332811	0	-3.96E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae|g__unclassified_Peptostreptococcaceae|s__Peptostreptococcaceae_ba	0.011790518	0.000324712	0	3.83E-06	0
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Bifidobacteriales|f__Bifidobacteriaceae|g__Alloscardovia|s__Alloscardovia_omnicolens	0.006735229	0.000553539	0	3.73E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Leuconostocaceae|g__Leuconostoc|s__ambiguous_Leuconostoc	-0.008900832	0.00039279	0	-3.50E-06	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Paraprevotellaceae	-0.012787111	0.000270429	0	-3.46E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Roseburia	-0.002096878	0.001625235	0.000106496	-3.41E-06	-2.23E-07
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Lactobacillaceae|g__Lactobacillus|s__Lactobacillus_salivarius	0.000480257	0.007001854	0.00014241	3.36E-06	6.84E-08
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Peptostreptococcaceae|g__unclassified_Peptostreptococcaceae	-0.001371481	0.002406348	8.11E-06	-3.30E-06	-1.11E-08
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_scindens	0.003032448	0.001085495	1.20E-05	3.29E-06	3.63E-08
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Rikenellaceae|g__Alistipes	0.001009088	0.003105564	0	3.13E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_lavalense	0.008291395	0.000373356	0	3.10E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Anaerostipes|s__NA	0.008094988	0.000358584	0	2.90E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Aerococcaceae|g__Abiotrophia|s__Abiotrophia_defectiva	0.000810256	0.003569111	0	2.89E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Leuconostocaceae	0.000593109	0.004794776	8.30E-05	2.84E-06	4.92E-08
k__Bacteria|p__Fusobacteria|c__Fusobacteriia|o__Fusobacteriales|f__Fusobacteriaceae|g__Fusobacterium	-0.00300684	0.000911299	0	-2.74E-06	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Coriobacteriales|f__Coriobacteriaceae|g__Atopobium|s__Atopobium_rimae	-0.011642434	0.000225393	0	-2.62E-06	0

Supplemental Table S9, continued

tax_group	effect size	abudance (mean)	abudance (median)	score contribution (mean)	score contribution (median)
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Blautia|s__Blautia_luti	-0.000483695	0.005367175	0.000465611	-2.60E-06	-2.25E-07
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Dorea|s__NA	0.002722814	0.000905451	1.13E-05	2.47E-06	3.09E-08
k__Bacteria|p__Actinobacteria|c__Actinomycetales|o__Propionibacterineae|f__Propionibacteriaceae|g__Propionibacterium|s__Propionibacterium_propioni	-0.018678621	0.000115174	0	-2.15E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus_lutetiensis	-0.007931214	0.000247448	0	-1.96E-06	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Rikenellaceae|g__Alistipes|s__Alistipes_putredinis	0.005513495	0.000321177	0	1.77E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Bacillales|f__Staphylococcaceae|g__Salinicoccus|s__Salinicoccus_qingdaonensis	0.034271968	5.12E-05	0	1.76E-06	0
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Bifidobacteriales|f__Bifidobacteriaceae|g__Scardovia|s__Scardovia_inopinata	-0.000580411	0.002876826	9.64E-05	-1.67E-06	-5.60E-08
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Dorea|s__Dorea_formicigenerans	-0.015688313	0.000105658	0	-1.66E-06	0
k__Bacteria|p__Proteobacteria|c__Betaproteobacteria|o__Burkholderiales|f__Alcaligenaceae|g__Sutterella|s__Parasutterella_excrementihominis	0.000997514	0.001579469	0	1.58E-06	0
k__Bacteria|p__Bacteroidetes|c__Bacteroidia|o__Bacteroidales|f__Bacteroidaceae|g__Bacteroides|s__Bacteroides_caccae	0.001524869	0.000872827	0	1.33E-06	0
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Actinomycetales|f__Corynebacteriaceae|g__Corynebacterium|s__Corynebacterium_pseudogenitalium	-0.001777214	0.000714239	0	-1.27E-06	0
k__Bacteria|p__Firmicutes|c__Bacilli|o__Turicibacterales|f__Turicibacteraceae|g__Turicibacter|s__Turicibacter_sanguinis	-0.00197181	0.000537559	1.01E-06	-1.06E-06	-1.99E-09
k__Bacteria|p__Actinobacteria|c__Actinobacteria|o__Bifidobacteriales|f__Bifidobacteriaceae	7.43E-05	0.009416336	0.00085743	6.99E-07	6.37E-08
k__Bacteria|p__Firmicutes|c__Bacilli|o__Lactobacillales|f__Streptococcaceae|g__Streptococcus|s__Streptococcus_anginosus	0.000477923	0.001388004	2.48E-05	6.63E-07	1.19E-08
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Lachnospiraceae|g__Roseburia|s__Roseburia_faecis	-0.004667884	0.000135785	0	-6.34E-07	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_paraputrificum	-0.001346627	0.000351336	0	-4.73E-07	0
k__Bacteria|p__Fusobacteria	-0.000171216	0.002445486	2.14E-06	-4.19E-07	-3.67E-10
k__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pasteurellales|f__Pasteurellaceae|g__Haemophilus|s__Haemophilus_parainfluenzae	0.001010249	0.000406836	0	4.11E-07	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Ruminococcaceae|g__Oscillospira|s__NA	0.000298961	0.001114031	2.93E-05	3.33E-07	8.75E-09
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_aldenense	0.001311758	0.000238148	0	3.12E-07	0
k__Bacteria|p__Firmicutes|c__Clostridia|o__Clostridiales|f__Clostridiaceae|g__Clostridium|s__Clostridium_symbiosum	-0.000672498	0.000124891	0	-8.40E-08	0
k__Bacteria|p__Actinobacteria|c__Actinomycetales|o__Propionibacterineae|f__Propionibacteriaceae|g__Propionibacterium|s__Propionibacterium_freudenr	-0.000575057	0.000135021	0	-7.76E-08	0

Supplemental Table S10. Clinical characteristics of patients by high- and low-diversity groups.

N = 882	High Diversity	Low Diversity	Median Diversity*
Institution, N (%)	443	439	
MSKCC	354 (79.9)	350 (79.7)	2.64
Regensburg	28 (6.3)	28 (6.4)	2.87
Duke	42 (9.5)	42 (9.6)	2.98
Hokkaido	19 (4.3)	19 (4.3)	1.55
Age at HCT, year (mean (sd))	54.67 (13.13)	52.01 (13.42)	
Sex = M (%)	289 (65.2)	270 (61.5)	
Disease (%)			
AML	144 (32.5)	174 (39.6)	
MDS/MPN	101 (22.8)	78 (17.8)	
NHL	80 (18.1)	65 (14.8)	
ALL	42 (9.5)	50 (11.4)	
Myeloma	27 (6.1)	34 (7.7)	
other	13 (2.9)	11 (2.5)	
CLL	15 (3.4)	6 (1.4)	
Hodgkins	9 (2.0)	10 (2.3)	
CML	9 (2.0)	8 (1.8)	
AA	3 (0.7)	3 (0.7)	
Graft Type (%)			
BM unmodified	48 (10.8)	36 (8.2)	
cord	66 (14.9)	74 (16.9)	
PBSC T-cell Depleted	110 (24.8)	165 (37.6)	
PBSC unmodified	219 (49.4)	164 (37.4)	
Conditioning Intensity (%)			
Ablative	218 (49.2)	261 (59.5)	
Reduced Intensity	176 (39.7)	152 (34.6)	
Nonmyeloablative	49 (11.1)	26 (5.9)	

Supplemental Table S10. Clinical characteristics of patients in the survival analysis according to high- and low-diversity groups day 7-21. For survival analysis, patients were grouped into high- and low-diversity groups according to the institution-specific median diversity.

* median Simpson reciprocal diversity index value per institution.

Supplemental Table S11. Sample-collection periods

Institution	Start	End	Collection Duration (years)
MSKCC	Apr 2009	Jan 2018	8.8
Regensburg	May 2011	Jun 2017	6.1
Duke	Jul 2012	Apr 2018	5.8
Hokkaido	Aug 2016	Jan 2018	1.4

References to the Supplemental Appendix

1. Taur Y, Coyte K, Schluter J, et al. Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Sci Transl Med 2018;10.
2. Taur Y, Jenq RR, Perales MA, et al. The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. Blood 2014;124:1174-82.
3. Bacigalupo A, Ballen K, Rizzo D, et al. Defining the intensity of conditioning regimens: working definitions. Biol Blood Marrow Transplant 2009;15:1628-33.
4. Consortium HMP. Structure, function and diversity of the healthy human microbiome. Nature 2012;486:207-14.
5. Sinha R, Abu-Ali G, Vogtmann E, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nature biotechnology 2017;35:1077-86.
6. Jenq RR, Taur Y, Devlin SM, et al. Intestinal Blautia Is Associated with Reduced Death from Graft-versus-Host Disease. Biol Blood Marrow Transplant 2015;21:1373-83.
7. Nagpal R, Ogata K, Tsuji H, et al. Sensitive quantification of Clostridium perfringens in human feces by quantitative real-time PCR targeting alpha-toxin and enterotoxin genes. BMC microbiology 2015;15:219.
8. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature methods 2013;10:996-8.
9. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. PeerJ 2016;4:e2584.
10. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I, Zaslavsky L. About Prokaryotic Genome Processing and Tools. The NCBI Handbook [Internet]. 2nd ed: National Center for Biotechnology Information (US); 2014.
11. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;72:5069-72.
12. Morris EK, Caruso T, Buscot F, et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. Ecology and evolution 2014;4:3514-24.
13. Costea PI, Hildebrand F, Arumugam M, et al. Enterotypes in the landscape of gut microbial community composition. Nature microbiology 2018;3:8-16.
14. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Australia: R Foundation for Statistical Computing; 2015.
15. Peled JU, Devlin SM, Staffas A, et al. Intestinal Microbiota and Relapse After Hematopoietic-Cell Transplantation. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2017;35:JCO2016703348.
16. Zhernakova A, Kurilshikov A, Bonder MJ, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science (New York, NY) 2016;352:565-9.
17. Forslund K, Hildebrand F, Nielsen T, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 2015;528:262-6.
18. Kugathasan S, Denson LA, Walters TD, et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. Lancet (London, England) 2017;389:1710-8.
19. Morgan XC, Kabakchiev B, Waldron L, et al. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. Genome biology 2015;16:67.
20. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome biology 2012;13:R79.

21. Schirmer M, Denson L, Vlamakis H, et al. Compositional and Temporal Changes in the Gut Microbiome of Pediatric Ulcerative Colitis Patients Are Linked to Disease Course. Cell Host Microbe 2018;24:600-10.e4.
22. Taur Y, Xavier JB, Lipuma L, et al. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2012;55:905-14.
23. Shono Y, Docampo MD, Peled JU, et al. Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic hematopoietic stem cell transplantation in human patients and mice. Sci Transl Med 2016;8:339ra71.
24. Weber D, Jenq RR, Peled JU, et al. Microbiota Disruption Induced by Early Use of Broad Spectrum Antibiotics is an Independent Risk Factor of Outcome after Allogeneic Stem Cell Transplantation. Biol Blood Marrow Transplant 2017;23:845-52.
25. Weber D, Hiergeist A, Weber M, et al. Detrimental Effect of Broad-spectrum Antibiotics on Intestinal Microbiome Diversity in Patients After Allogeneic Stem Cell Transplantation: Lack of Commensal Sparing Antibiotics. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2019;68:1303-10.
26. Weber D, Oefner PJ, Dettmer K, et al. Rifaximin preserves intestinal microbiota balance in patients undergoing allogeneic stem cell transplantation. Bone marrow transplantation 2016;51:1087-92.
27. Golob JL, Pergam SA, Srinivasan S, et al. Stool Microbiota at Neutrophil Recovery Is Predictive for Severe Acute Graft vs Host Disease After Hematopoietic Cell Transplantation. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2017;65:1984-91.
28. Routy B, Letendre C, Enot D, et al. The influence of gut-decontamination prophylactic antibiotics on acute graft-versus-host disease and survival following allogeneic hematopoietic stem cell transplantation. Oncoimmunology 2017;6:e1258506.
29. Simms-Waldrip TR, Sunkersett G, Coughlin LA, et al. Antibiotic-Induced Depletion of Anti-Inflammatory Clostridia is Associated with the Development of GVHD in Pediatric Stem Cell Transplant Patients. Biol Blood Marrow Transplant 2017.
30. Jaffe D, Jakubowski A, Sepkowitz K, et al. Prevention of peritransplantation viridans streptococcal bacteremia with early vancomycin administration: a single-center observational cohort study. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2004;39:1625-32.
31. Seo SK, Xiao K, Huang YT, et al. Impact of peri-transplant vancomycin and fluoroquinolone administration on rates of bacteremia in allogeneic hematopoietic stem cell transplant (HSCT) recipients: a 12-year single institution study. The Journal of infection 2014;69:341-51.
32. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS medicine 2007;4:e296.