SUPPLEMENTARY MATERIAL

Lactose drives *Enterococcus* expansion to promote graft-versus-host disease

Stein-Thoeringer, C.K.[1,3], Nichols, K.B.[1,3], Lazrak, A.[1,3], Docampo, M.D.[1,3], Slingerland, A.E.[1,3], Slingerland, J.B.[1,3], Clurman, A.G.[2], Armijo, G.[1,3], Gomes, A.L.C.[1,3], Shono, Y.[1,3], Staffas, A.[1,3], Burgos da Silva, M.[1,3], Devlin, S.[4], Markey, K.A.[1,2,3], Bajic, D.[5], Pinedo, R.[20], Tsakmaklis, A.[6,19,21], Littmann, E.R.[1,22], Pastore, A.[1], Taur, Y.[7], Monette, S.[8], Arcila, M.E.[9], Pickard, A.J.[10], Maloy, M.[2], Wright, R.J.[1], Amoretti, L.A.[1], Fontana, E.[1], Pham, D.[11], Jamal, M.A.[11], Weber, D.[12], Sung, A.D.[13], Hashimoto, D.[14], Scheid, C.[6], Xavier, J.B.[15], Messina, J.A.[16], Romero, K.[17], Lew, M.[13], Bush, A.[13], Bohannon, L.[13], Hayasaka, K[18], Hasegawa, Y.[14], Vehreschild, M.J.G.T.[6,19,21], Cross, J.R.[10], Ponce, D.M.[2,3], Perales, M.A.[2,3], Giralt, S.A.[2,3], Jenq, R.R.[11], Teshima, T.[14,18], Holler, E.[12], Chao, N.J.[13], Pamer, E.G.[1,3,22], Peled, J.U.[1,2,3]*, van den Brink, M.R.M.[1,2,3]*

Affiliations:
[1]Department of Immunology, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY; [2]Adult Bone Marrow Transplantation Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; [3]Weill Cornell Medical College, NY; [4]Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY; [5]Department of Internal Medicine II, Technical University of Munich, Germany; [6]Internal Medicine I, University Hospital Cologne, Germany; [7]Infectious Disease Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; [8]Laboratory of Comparative Pathology, Memorial Sloan Kettering Cancer Center, The Rockefeller University, Weill Cornell Medicine, NY; [9]Diagnostic Molecular Pathology Laboratory, Memorial Sloan Kettering Cancer Center, NY; [10]Donald B. and Catherine C. Marron Cancer Metabolism Center, Memorial Sloan Kettering Cancer Center, New York, NY; [11]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston; [12]Internal Medicine III, University Clinic Regensburg, Regensburg, Germany; [13]Division of Hematologic Malignancies and Cellular Therapy, Department of Medicine, Duke University Medical Center, Durham, NC; [14]Department of Hematology, Hokkaido University, Faculty of Medicine, Sapporo, Japan; [15]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY; [16]Division of Infectious Diseases, Department of Medicine, Duke University, Durham, NC; [17]Office of Clinical Research, Duke University School of Medicine, Durham, NC; [18]Division of Laboratory and Transfusion Medicine, Hokkaido University Hospital, Sapporo, Japan; [19]Department of Internal Medicine, Infectious Diseases, Goethe University Frankfurt, Germany; [20]Gnotobiotic Facility, Memorial Sloan Kettering Cancer Center, New York, NY; [21]German Center for Infection Research, Partner site Bonn-Cologne, Cologne, Germany; [22]Department of Medicine, Section of Infectious Medicine and Global Health, University of Chicago.
* co-senior authors

MATERIAL and METHODS

Patients and fecal specimens

Stool samples were prospectively collected at four different transplant centers (Memorial Sloan Kettering Cancer Center (MSKCC) and Duke University Medical Center both in the United States, University Medical Center in Regensburg, Germany, and Hokkaido University Hospital in Japan) during different time periods over 1.4–8.8 years within the years 2009 – 2018 (MSKCC: Apr 2009 to Jan 2018; Duke: Jul 2012 to Apr 2018; Regensburg: May 2011 to Jun 2017; Hokkaido: Aug 2016 to Jan 2018). Samples were collected, aliquoted and frozen according to harmonized protocols at each center. DNA extraction, PCR for 16S rRNA amplicon sequencing, and analyses were performed centrally as described below. Written informed consent was received from participants prior to sample collection under the supervision of institutional review boards at each center. Patients with at least one evaluable sample collected after day –30 relative to a first allo-HCT were included. For most patients, samples were requested weekly. Patients underwent transplantation for a range of indications, although acute myeloid leukemia was the most common at all four centers. The patients varied in the intensity of pre-HCT conditioning regimen (**Supplementary Table S1**). The most common graft type was unmodified (i.e., not T-cell depleted) peripheral blood stem cells or bone-marrow (**Supplementary Table S1**). While three of the four centers administered cord-blood grafts, only one center infused grafts that were *ex vivo* T-cell depleted (TCD, allo-HCT patients at MSKCC). The 1,325 patients summarized in **Supplementary Table S1** provided 9,049 stool samples that were analyzed for microbiome composition (MSKCC: 1101 patients, 8472 samples; Duke: 79 patients, 231 samples; Hokkaido: 66 patients, 202 samples; Regensburg: 79 patients, 144 samples).

In the analysis of incidence of domination (**Figure 1a, left**), we computed the cumulative incidence of patients who experienced at least one instance of genus *Enterococcus* domination of the fecal microbiota (domination defined as a relative abundance of the genus ≥0.3) over the course of allo-HCT (day -20 to +24 relative to HCT) using 7-day sliding windows at different transplant centers. If no sample was collected from a patient in a given window, the cumulative incidence did not change but was plotted. We further analyzed the prevalence of domination as the fraction of fecal specimens with enterococcal domination of the gut microbiota (**Figure 1a, right**). Here, we only plotted the data points from time windows in which samples from at least 5 patients were available. The reason for this was that if very few samples were collected in a particular time window as the denominator, then even a low frequency of domination might appear to be a spuriously high frequency.

To determine fold-changes from baseline (pre-transplant), we defined a baseline period as day -30 to day -6 relative to HCT. For the patients with a baseline sample in this pre-transplant window and at least one subsequent sample, we plotted *E. faecium* abundance as a fold-change from the baseline value (**Figure S1c**). Both fold-increases and fold-decreases were observed; considering all samples collected between days 14-21, *E. faecium* abundance increased by a median of 6.4-fold and by a mean of 143.6-fold. In **Figure S1c**, we also present individual patient time courses normalized to baseline, where each thin trendline is a single patient's trend over time as fold-change from *E. faecium* abundance. Considering all samples collected between days 14 to 21 from patients who became dominated at some point, *E. faecium* increased by a median of 120-fold and a mean of 332-fold. Considering all samples collected in the same time period from patients who never became dominated, *E. faecium* increased by a median of 2-fold and a mean of 55.6-fold.

For the analysis of clinical outcomes including GVHD, GVHD-related mortality (GRM), overall survival (OS), recipients of TCD grafts were excluded, and only patients with evaluable samples from day 0 to +21 were considered, yielding a sub-cohort for analysis of clinical outcomes (**Supplementary Table S3**). MSKCC patients (n = 538) served as the discovery cohort while 167 patients from Regensburg, Duke, and Hokkaido were amalgamated into a multicenter validation cohort that included 62 patients from Duke, 38 from Hokkaido, and 67 from Regensburg. GVHD was graded according to the CIBMTR guidelines.

Domination Threshold and Sensitivity Analysis

The concept of microbial domination of the intestinal microbiota has been developed as to better understand microbiota injury and dysbiosis (*33*). We defined domination by a relative-abundance threshold of any single taxonomic unit ≥0.3 which was informed by a previous study from our center by Taur *et al*. (*5*). In this study, domination events by *Enterococcus* were associated with a 9-fold increase in risk of VRE bacteremia, and domination by proteobacteria increased the risk of bacteremia with aerobic gram-negative bacilli 5-fold (*5*). This threshold was recently also used to define intestinal and oral microbiota domination in another clinical cohort of AML patients (*34*). While we selected this threshold *a priori* on the basis of the prior studies, we also conducted a sensitivity analysis to assess the extent to which our observations are robust to various definitions of domination.

As shown in new **Supplemental Fig S2c**, the relative abundance of *E. faecium* in this dataset has a bimodal distribution with a broad peak between 10^{-3} and 10^{-2} and a second peak of very high abundances, including many samples in which *E. faecium* is virtually the only taxon identified.

For this sensitivity analysis, we selected 5 cut-offs in between these two peaks to evaluate: 0.1, 0.2, 0.3, 0.4, and 0.5; the number of samples that are classified as dominated or not dominated are tabulated in **Supplemental Figure S2c**.

The cumulative incidence of genus *Enterococcus* domination was relatively insensitive to the specific threshold at which domination is defined (**Supplemental Figure S2c**). The cumulative incidence of domination dramatically increased around day 0 and then reached a plateau with relatively similar kinetics across all tested threshold definitions, although the height of the plateau varied with the definition such that fewer patients were considered to have had domination events at higher, more stringent thresholds.

Genus *Enterococcus* domination predicted a higher risk of all-cause mortality with high statistical significance at all tested domination thresholds and in both univariate and multivariate Cox models (**Supplemental Figure S2d**). Enterococcus domination predicted GVHD-related mortality at thresholds between 0.2 and 0.5, although in the lowest, least stringent threshold of 0.1 there was no longer a significant association for this outcome. P-values for the univariate and multivariate models are listed in **Table S3**. The multivariate Cox models were adjusted for graft source, age, conditioning intensity, gender, and underlying disease (leukemia vs other).

In the main analysis, we considered patients to have had a domination event if even a single sample collected in the window of day 0 to +21 had *Enterococcus* (genus) abundance ≥0.3. We also observed a similar result when we considered only each patient's last single sample collected in this window: The detection of *Enterococcus* domination (genus level) in the last sample between day 0-21 had an all-cause mortality HR of 1.75 (1.29-2.38), $p<0.001$; and for GVHD-related mortality HR 1.86 (1.07-3.22), $p = 0.03$.

Mice

C57BL/6J, BALB/cJ, 129S1/SvImJ, and LP/J mice were obtained from the Jackson Laboratory and kept under standard housing conditions (cohousing of 3 – 5 mice per cage, chow and water *ad libitum*, 12h-light cycle [6:00 pm: off]). Male mice were used for these experiments at an age of 6 to 9 weeks. For gnotobiotic / germ-free experiments, C57BL/6J mice were bred and housed in the MSKCC gnotobiotic facility with weekly microbiological monitoring of germ-free status (*35*). Experiments with germ-free/gnotobiotic mice were performed in individual gnotobiotic isocages (SentrySPP, Allentown). Mouse allo-HCT experiments were performed as previously described (*10*): conditioning regimens were split-dosed lethal irradiation with 1000 cGy for 129S1 recipients, with 900 cGy for BALB/c recipients, and busulfan/cyclophosphamide i.p. injections (busulfan

20mg/kg for 5 days; cyclophosphamide 100mg/kg for the last 3 days; (*16*)) for C57BL/6J mice (SPF and germ-free/gnotobiotic mice). Mice were transplanted with bone marrow (BM, 5 x 10^6 cells) that was T-cell depleted with anti–Thy-1.2 and Low-Tox-M rabbit complement (CEDARLANE Laboratories). Donor T cells were prepared by harvesting donor splenocytes and enriching T cells by Miltenyi MACS purification using CD5 (routinely >90% purity; T-cell dose indicated for each experiment separately in the main text/figure legends). BM or BM+T were administered by retroorbital injections in transiently isoflurane-anaesthetized mice. Mice were monitored daily for survival and weekly for GVHD clinical scores as described (*10*).

14 days after HCT organs were harvested in a subset of transplanted mice, and parts of the small intestine (ileum; last 2 cm cranial to cecum), proximal large intestine, liver, and skin samples evaluated histologically for evidence of GVHD, whereas spleens, ileum and colon tissues were processed for splenocyte or lamina propria cell preparations for flow cytometry (*10*).

Flow cytometry

Antibodies were obtained from BD Biosciences Pharmingen, eBioscience or Biolegend. For cell analysis of surface markers, cells were stained for 20 min at 4°C in phosphate-buffered saline (PBS) with 0.5% bovine serum albumin (BSA) (PBS/BSA) after Fc block, washing, and resuspending in a viability staining solution (Fixable Dead Cell Stain, Molecular Probes, Life Technologies). Cell-surface staining was followed by intracellular staining with the fixation/permeabilization kit by eBioscience per the manufacturer's instructions. All flow-cytometry experiments were performed on a X50 flow cytometer (BD Biosciences) and analyzed with FlowJo 10.4.1 (Tree Star Software).

Quantitative PCRs for gene expression analyses and SNP genotyping

Duodenal or ileal segments of the small intestine were harvested from transplanted mice vs. mice at steady state. RNA was extracted using a Direct-zol RNA kit (Zymo Research), followed by RT-PCR using a Quantitect RT Kit (Quiagen). For real-time quantitative PCR we used TaqMan probes (Applied Biosystems) for murine Reg3B, Reg3G, lactase and GAPDH and a TaqMan universal master mix according to the manufacturer's instructions. The relative expression of target mRNAs was calculated by the ΔCt method and values were normalized to mRNA expression levels in controls.

SNP genotyping for the functional SNP rs4988235, upstream of the lactase gene, was performed using a TaqMan SNP genotyping protocol by Applied Biosystems (C___2104745_10). All qPCRs were run in 384 well formats on a QuantStudio 6 Flex Real-Time PCR system (Applied

Biosystems). To ensure SNP assessments reflected recipient and not donor genotype, archived buccal DNA samples that had been collected prior to allo-HCT were obtained from the clinical laboratory at MSKCC. All TaqMan-probes used in the present study are listed in **Table S10**.

Lactose measurements

The lactose concentrations in brain-heart-infusion broth, regular cow milk or in a mouse chow suspension (5ml ddH_2O/g chow) were measured by lactose assay kits (Abcam or Cell Biolabs) according to manufacturer's instructions. The mouse chow suspension was pretreated with perchloric acid precipitation (deproteinization protocol by Abcam) to increase assay sensitivity. BHI was pretreated with Lactaid® (3000 IU/ml) for 30min at 37°C.

Microbiome analyses

16S rRNA amplicon sequencing. DNA was extracted from human or mouse fecal samples using a phenol-chloroform bead beating protocol, and the genomic 16S ribosomal-RNA gene V4-V5 variable region was amplified and sequenced on the Illumina MiSeq platform as previously described (*10, 11, 13*). PCR products were purified using the Agencourt AMPure PCR amplicon 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). The Operational Taxonomic Units (OTUs) were called by the vsearch algorithm (*36*) 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 ribosomal RNA sequence database (*37*). OTUs were clustered at 97% identity. OTUs were classified to the species level against the Greengenes database (*38*), with gaps in taxonomic annotation filled in by classification against the NCBI 16S ribosomal RNA sequence database (*37*). We used the Qiime (*39*) function assign_taxonomy.py and the assignment_method mothur (*40*) to map sequences on green genes database. We used blastn on the 16S NCBI database to supplement genus and species annotation in sequences that had no genus/species assigned and a 97% match identity.

Alpha-diversity: We calculated α-diversity using the inverse Simpson index at the level of OTUs. Beta-diversity was computed according to Bray-Curtis distances at the genus level and clustered using a principal coordinates analysis (PCoA) ordination.

Metagenome shotgun sequencing. Extracted and purified DNA was sheared to a target size of 650 bp with a Covaris ultrasonicator and prepared for sequencing with the Illumina TruSeq DNA library preparation kit according to the Illumina protocol. Sequencing was performed on a HiSeq system (Illumina) targeting ~10-20x10^6 reads per sample with 100 bp, paired-end reads. Metagenome sequences were further analyzed using the HUMAnN2 tool developed for the shotgun metagenome analyses of the Human Microbiome Project 2 as described (*41*). Prior to the HUMAnN2 workflow, shotgun sequences were filtered with quality control and removal of host reads using KneadData (http://huttenhower.sph.harvard.edu/kneaddata). Taxonomic profiling was done by MetaPHLAn2, and genes we are annotated to UniRef90 and metabolic pathways to the MetaCyc database. Whole-genome sequences of cultured *Enterococcus* isolates were analyzed by a metagenome web-based analysis tool provided by Pathosystems Resource Integration Center (PATRIC v3.5.27) (https://www.patricbrc.org). In detail, raw Illumina HiSeq reads were quality-filtered by Trimmomatic (version 0.36), and the quality was assessed by FastQC (version 0.11.5). The genome assembly service by PATRIC was used to assemble the reads into contigs. The genome annotation tool was applied to provide annotation of contigs to genomic features using RASTtk. Annotation is done against BLASTN, BLAT, FIGfam collection and SEED databases to compute taxonomy and proteins (*42*). Metabolic pathway modeling was done by ModelSEED v2.4 (modelseed.org).

IgA-BugFACS. Following a previously published protocol (*43*), 100 mg of fecal contents were homogenized in PBS on ice. Clarified supernatants were washed in PBS/1% bovine serum albumin and incubated with blocking buffer (20% normal mouse serum in PBS) for 20 min. Samples were subsequently stained with anti-mouse IgA (clone mA06E1, ebioscience) or isotype (rat IgG1 isotype) and, subsequently, co-stained with SYBR I (Invitrogen). Flow sorting was done in using an FACS Aria device (BD Biosciences) under aseptic conditions of laminar air flow. Sorted fractions were then 16S sequenced as described above. The IgA-coating index (ICI) was calculated using the relative abundance for each individual taxon as ICI = (log(IgA+) – log(IgA-))/(log(IgA+) + log(IgA-)) as described previously (*43, 44*).

Short chain fatty acids (SCFAs) analyses

GCMS for human fecal sample analyses. Samples were weighed into 2 mL microtubes containing 2.8 mm ceramic beads (Omni International). Extraction solvent containing butyric acid-d7 internal standard (Cambridge Isotope Laboratories) in 80% MeOH (Fisher Scientific) was added to the tubes at a ratio of 100 mg sample (wet weight) to 900 µL of extraction solvent. Samples were homogenized using a Bead Ruptor homogenizer (Omni International) at 6 M/s for 30 s for a total

of 6 cycles with 0.01 s dwell time at 4 C°. Samples were centrifuged for 20 minutes at 20,000 x g at 4°C. A 100 µL aliquot of extract supernatant was added to 100 µL of 100 mM borate Buffer (pH 10), 400 µL of 100 mM pentafluorobenzyl bromide (Thermo Scientific) in acetone (Fisher Scientific), and 400 µL of cyclohexane (Acros Organics) in a sealed autosampler vial. Samples were heated to 65°C for 1 hour with shaking. After cooling to room temperature and allowing the layers to separate, 50 µL of the cyclohexane supernatant was added to a 450 µL cyclohexane in an autosampler vial and sealed. 1 µL of the cyclohexane supernatant was analyzed by GCMS (Agilent 7890A GC system, Agilent 5975C MS detector) operating in negative chemical ionization mode, using methane as the reagent gas. Analysis was performed using Mass Hunter GCMS Quantitative Analysis (Version B.09.00, Agilent Technologies) software. Raw peak areas for butyrate were normalized to the butyric acid-d7 internal standard and quantified from a calibration curve ranging from 0.2-100 mM butyrate.

h-NMR for mouse cecal sample analyses. Butyrate concentrations from cecal contents of allo-HCT mice were determined by 1H-NMR spectroscopy using an UNITY INOVA 600 NMR spectrometer (Varian) as previously described (*10*). Spectra were analyzed using the Chenomix NMR software.

GVHD histopathology

Small and large intestines, liver, and skin samples were harvested after allo-HCT. After fixation in 10% formalin, embedding in paraffin and sectioning at 5 µm, H&E staining, as well as Ki67 and CD3 immunohistochemical (IHC) and TUNEL stainings were performed by the Laboratory of Comparative Pathology facility at MSKCC. IHC was performed on a Leica Bond RX automated stainer (Leica Biosystems, Buffalo Grove, IL) following standard manufacturer's instructions except for the following details. Heat induced epitope retrieval was performed in a pH 6.0 (CD3) or 9.0 (ki67) buffer, and the primary antibody, anti-CD3 rabbit monoclonal antibody, clone SP162, ab 135372, or anti-ki67 rabbit monoclonal antibody, clone SP6, ab16667, Abcam, Cambridge, MA, was applied at a concentration of 1:250 and 1:100, respectively, and was followed by application of a polymer detection system (DS9800, Novocastra Bond Polymer Refine Detection, Leica Biosystems). The chromogen was 3,3 diaminobenzidine tetrachloride (DAB), and sections were counterstained with hematoxylin. TUNEL staining was performed as previously described (*45*). Slides were examined by a board-certified veterinary pathologist (S.M.) who was blinded to group treatments during evaluation. Each mouse received a score (from 0 to 4 ["none" to "marked"]) for each of the following parameters. Small intestine: villus blunting, crypt hyperplasia (Ki67), crypt apoptosis (TUNEL), crypt loss, lamina propria fibrosis, CD3+ cells, and mucosal

ulceration. Large intestine: mucosal erosion, crypt hyperplasia (Ki67), crypt apoptosis (TUNEL), crypt loss, lamina propria fibrosis, CD3+ cells, and mucosal ulceration. Liver: portal CD3+ cells, bile duct CD3+ cells, bile duct apoptosis, bile duct sloughing, vascular endothelitis, hepatocyte apoptosis, parenchymal CD3+ cells, parenchymal mitoses, hepatocellular cholestasis, and hepatocellular steatosis. Skin: epithelial and follicular apoptosis, epidermal basal cell vacuolation, epidermal and dermal CD3+ infiltrates. Based on these individual scores, a sum organ-specific GVHD and a compound total score were calculated to evaluate evidence for GVHD (*46*). Representative histopathological microphotographs are presented in **Figure S4b.**

Gnotobiotic experiment with minimal community

Germ-free mice were colonized with a minimal community of six bacterial strains (*Akkermansia muciniphila*, *Lactobacillus johnsonii*, *Blautia producta*, *Bacteroides sartorii*, *Clostridium bolteae,* and *Parabacteroides distasonis*) 21 days before allo-HCT. *Blautia producta*, *Bacteroides sartorii*, *Clostridium bolteae, Parabacteroides distasonis* were previously described by Caballero *et al*. (*19*). Bacteria were individually cultured on Columbia plus 5% sheep blood agar (BD Biosciences) for 3 days at 37°C under anaerobic conditions. *Lactobacillus johnsonii* was isolated previously from our group from feces of GHVD mice (*18*) and grown on MRS agars (BD Biosciences) for 3 days at 37°C under anaerobic conditions. *Akkermansia muciniphila* was obtained from ATCC (strain BAA-835) and anaerobically cultured for 3 days according to ATCC recommendations. Plate cultures were scraped off, mixed in a 1:1 ratio, freshly resuspended in sterile PBS and gavaged to mice (~$1x10^7$CFUs/isolate in 200µl per mouse). In general, these strains represent major phyla of the mouse gut, and have previously been reported to modulate enterococci colonization in mice (*Blautia*), or to characterize GVHD-associated dysbiosis (*Akkermansia, Lactobacillus*) (*10, 18, 19*). One group of colonized mice was spiked with p.o. *E. faecalis* OG1RF ($2x10^7$ CFUs per mouse) two days after 6-strain colonization.

Statistical analyses

Survival analysis was performed using R package *survival* and *cmprsk*. The cumulative incidences of graft-versus-host disease (GVHD), relapse or progression of disease, and GVHD-related mortality (GRM) were estimated using cumulative incidence functions accounting for the corresponding competing risks. The competing risks for GVHD were relapse and death, and for relapse it was death without 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 *Enterococcus* domination or lactase genotypes and outcomes

(survival). Estimates and comparisons across groups were based on a landmark time point among patients without the event by day +21. Patients with T-cell-depleted grafts were excluded from analyses of clinical outcomes. Domination was defined as any OTU with relative abundance ≥0.3, a threshold that we have previously used to define domination (*35*). The samples were binned into 7-day sliding windows in accordance with the approximately weekly collection schedule. The domination cumulative incidence plot considers patients with at least one evaluable sample at pre-HCT and one at post-HCT time. The fraction of patients in whom at least one instance of domination was detected by the given time is plotted. In the analyses of bacterial metagenomes, we used the linear discriminant analysis of effect size (LEfSe), a bioinformatics tool that identifies differentially abundant features between microbial communities (*47*). Correlation statistics were performed using Kendall's Tau rank correlation. For two group comparisons in mouse experiments, we applied independent or dependent T-tests, Wilcoxon rank-sum or signed-rank tests, or the AUC-Vardi test (*48*). Statistical significance was determined based on p-values <0.05.

Data Availability

Sequencing data are deposited into SRA under Bioproject number PRJNA545312.

Table S1

	Overall Cohort
Overall N	1325
Institution (%)	
MSKCC	1101 (83.1)
Regensburg	79 (6.0)
Duke	79 (6.0)
Hokkaido	66 (5.0)
Age at HCT, year (mean (sd))	52.9 (12.8)
Sex = (male, %)	801 (60.5)
Disease (%)	
AML	485 (36.6)
MDS/MPN	244 (18.4)
NHL	223 (16.8)
ALL	124 (9.4)
Myeloma	113 (8.5)
CLL	33 (2.5)
CML	29 (2.2)
Hodgkins	30 (2.3)
AA	9 (0.7)
other	35 (2.6)
Graft type (%)	
BM/PBSC unmodified	660 (49.8)
cord	207 (15.6)
PBSC T-cell Depleted	458 (34.6)
Conditioning intensity (%)	
Ablative	744 (56.2)
Reduced Intensity	466 (35.2)
Nonmyeloablative	115 (8.7)

Supplementary Table S1. Clinical characteristics of the overall cohort used in the analysis of microbiome composition. MSKCC, Memorial Sloan Kettering Cancer Center; The multicenter validation cohort was comprised of Duke, Regensburg, and Hokkaido; HCT, hematopoietic cell transplantation; AML, acute myeloid leukemia; MDS/MPN, myelodysplastic syndromes/myeloproliferative neoplasms; NLH, Non-Hodgkin Lymphomas; ALL, acute lymphoid leukemia; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; AA, aplastic anemia; BM, bone marrow; PBSC, peripheral blood stem cells; sd, standard deviation.

Table S2

Species	mean pre HCT	mean post- HCT max	delta of means	p value	FDR adjusted
Enterococcus faecium	4.79E-02	3.19E-01	2.72E-01	1.47E-103	2.93E-102
Enterococcus rivorum	2.49E-04	1.10E-03	8.47E-04	4.45E-83	4.45E-82
Enterococcus mundtii	8.37E-05	1.05E-03	9.65E-04	2.57E-58	1.72E-57
Enterococcus lactis	2.25E-04	7.22E-04	4.97E-04	2.00E-54	1.00E-53
Enterococcus moraviensis	4.02E-05	1.21E-04	8.09E-05	3.22E-28	1.29E-27
Enterococcus gallinarum	1.04E-04	2.12E-04	1.07E-04	1.45E-22	4.82E-22
Enterococcus cecorum	3.62E-06	9.08E-04	9.04E-04	2.39E-21	6.83E-21
Enterococcus durans	2.43E-06	5.07E-05	4.82E-05	1.53E-09	3.83E-09
Enterococcus faecalis	8.87E-07	8.10E-06	7.21E-06	2.65E-09	5.89E-09
Enterococcus ureasiticus	1.28E-05	2.64E-05	1.36E-05	9.16E-08	1.83E-07
Enterococcus silesiacus	2.18E-07	2.02E-06	1.81E-06	1.38E-05	2.50E-05
Enterococcus saccharolyticus	3.49E-07	1.97E-06	1.62E-06	1.37E-04	2.28E-04
Enterococcus sp. CR-303S	6.72E-08	1.09E-06	1.02E-06	5.13E-04	7.89E-04
Enterococcus sp. DME	2.29E-09	1.01E-06	1.01E-06	1.37E-03	1.95E-03
Enterococcus canis	1.81E-07	1.15E-06	9.64E-07	1.08E-01	1.44E-01
Enterococcus sp. DJF SLA47	2.76E-08	2.42E-07	2.14E-07	1.81E-01	2.27E-01
Enterococcus malodoratus	2.45E-07	1.03E-06	7.89E-07	2.81E-01	3.30E-01
Enterococcus sp. 4	6.16E-08	3.90E-07	3.28E-07	7.26E-01	8.07E-01

Supplementary Table S2. FDR-adjusted differences in median *Enterococcus spp.* abundances pre- vs. post-HCT are presented for the overall cohort. For this analysis, the 18 species with non-zero abundance in any pre-HCT sample were included. Pre-HCT abundance was defined as each patient's first sample in a window of day –30 to –6; post-HCT abundance was calculated as each patient's maximum abundance in a window of day 0 to 21. A Wilcoxon rank test with FDR correction was applied. Also tabulated are the mean of the pre-HCT abundances, the mean of the post-HCT-maximum abundances, and the difference (delta) between the means.

Table S3

	Non-dominated	Dominated	p-value	Non-dominated	Dominated	p-value
Cohort, N (overall)	MSKCC, 538 patients			Multicenter validation, 167 patients		
N (non-dominated vs. dominated)	389	149		113	54	
Age at HCT, year (mean (sd))	54.48 (13.30)	53.75 (12.87)	0.56	49.60 (12.94)	49.49 (13.83)	0.96
Sex (male, %)	241 (62.0)	96 (64.4)	0.67	74 (65.5)	33 (61.1)	0.71
Disease (%)			<0.001			0.51
AML	96 (24.7)	84 (56.4)		45 (39.8)	29 (53.7)	
MDS/MPN	65 (16.7)	14 (9.4)		21 (18.6)	7 (13.0)	
NHL	134 (34.4)	21 (14.1)		17 (15.0)	5 (9.3)	
ALL	24 (6.2)	17 (11.4)		14 (12.4)	9 (16.7)	
Myeloma	2 (0.5)	0 (0.0)		8 (7.1)	0 (0.0)	
CLL	22 (5.7)	3 (2.0)		2 (1.8)	1 (1.9)	
Hodgkins	20 (5.1)	2 (1.3)		1 (0.9)	1 (1.9)	
CML	9 (2.3)	3 (2.0)		2 (1.8)	1 (1.9)	
AA	4 (1.0)	0 (0.0)		2 (1.8)	1 (1.9)	
other	13 (3.3)	5 (3.4)		1 (0.9)	0 (0.0)	
Graft type (%)			0.03			0.17
BM/PBSC unmodified	289 (74.3)	96 (64.4)		97 (85.8)	51 (94.4)	
cord	100 (25.7)	53 (35.6)		16 (14.2)	3 (5.6)	
Conditioning intensity (%)			0.10			0.32
Ablative	88 (22.6)	40 (26.8)		68 (60.2)	26 (48.1)	
Reduced Intensity	224 (57.6)	91 (61.1)		44 (38.9)	27 (50.0)	
Nonmyeloablative	77 (19.8)	18 (12.1)		1 (0.9)	1 (1.9)	

Supplementary Table S3. Clinical characteristics of the sub-cohort used in analysis of clinical outcomes. MSKCC, Memorial Sloan Kettering Cancer Center; multicenter validation cohort, transplant centers at Duke, Regensburg, Hokkaido; HCT, hematopoietic cell transplantation; AML, acute myeloid leukemia; MDS/MPN, myelodysplastic syndromes/myeloproliferative neoplasms; NLH, Non-Hodgkin Lymphomas; ALL, acute lymphoid leukemia; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; AA, aplastic anemia; BM, bone marrow; PBSC, peripheral blood stem cells; cord, cord blood; sd, standard deviation.

Table S4

	MSKCC cohort			
Enterococcus	Univariate		Multivariate	
threshold (rel. abundance)	HR (95% CI)	p value	HR (95% CI)	p value
≥0.3				
OS	1.97 (1.45 to 2.66)	<0.0001	2.06 (1.50 to 2.82)	<0.0001
GRM	2.04 (1.18 to 3.52)	<0.05	2.60 (1.46 to 4.62)	<0.01
GVHD (Grade 2-4)	1.44 (1.10 to 1.88)	<0.01	1.32 (1.00 to 1.75)	<0.05

Supplementary Table S4. Univariate and multivariate analysis of *Enterococcus* domination and the outcomes of survival (OS), GVHD-related mortality (GRM) and acute GVHD grade 2-4 in the MSKCC patient cohort. The multivariate Cox models were adjusted for graft source, age, conditioning intensity, gender, and underlying disease (leukemia vs. other).

Table S5

Genus	Domination Frequency Mean	Median	Genus	Domination Frequency Mean	Median
Duke			**MSKCC**		
Enterococcus	0.131	0.135	Enterococcus	0.218	0.29
Streptococcus	0.102	0.095	Streptococcus	0.073	0.094
Lactobacillus	0.099	0.09	Eubacterium	0.04	0.051
Akkermansia	0.097	0.069	Lactobacillus	0.039	0.047
Blautia	0.082	0.076	Blautia	0.034	0.029
Klebsiella	0.035	0.043	Erysipelatoclostridium	0.028	0.027
Megasphaera	0.035	0.035	Staphylococcus	0.021	0.021
Parabacteroides	0.035	0.027	Klebsiella	0.02	0.017
Bacteroides	0.032	0.032	Akkermansia	0.016	0.009
Erysipelatoclostridium	0.031	0.045	Escherichia	0.015	0.014
Pediococcus	0.031	0.03	Clostridium	0.014	0.014
Eubacterium	0.03	0.03	Pediococcus	0.012	0.015
Escherichia	0.028	0.022	Actinomyces	0.01	0.007
Staphylococcus	0.027	0.021	Lactococcus	0.01	0.009
Leuconostoc	0.024	0.024	Bacteroides	0.008	0.011
Scardovia	0.024	0.024	Bifidobacterium	0.008	0.003
Clostridium	0.023	0.022	Parabacteroides	0.007	0.006
Enterobacter	0.021	0.021	Veillonella	0.006	0.005
Hokkaido			**Regensburg**		
Enterococcus	0.497	0.521	Enterococcus	0.453	0.329
Clostridium	0.165	0.059	Eubacterium	0.096	0.076
Akkermansia	0.133	0.14	Akkermansia	0.073	0.073
Eubacterium	0.093	0.086	Streptococcus	0.057	0.067
Streptococcus	0.066	0.033	Staphylococcus	0.054	0.054
Lactobacillus	0.061	0.083	Erysipelatoclostridium	0.046	0.046
Staphylococcus	0.061	0.061	Parabacteroides	0.033	0.033
Bifidobacterium	0.051	0.051	Escherichia	0.032	0.032
Leuconostoc	0.042	0.042	Faecalibacterium	0.028	0.028
Pediococcus	0.042	0.042	Blautia	0.022	0
Klebsiella	0.041	0.041	Clostridium	0.016	0
Erysipelatoclostridium	0.04	0.04	Coprococcus	0.014	0.016
Megamonas	0.033	0.033	Bacteroides	0.012	0.012
Blautia	0.026	0.026	Ruminococcus	0.006	0
Ruminococcus	0.017	0.026	[Clostridium]	0	0
Faecalibacterium	0.005	0	Coprobacillus	0	0
Bacillus	0	0	Klebsiella	0	0
Bacteroides	0	0	Oscillospira	0	0

Supplementary Table S5. The most commonly dominating genus was *Enterococcus*. Tabulated are the frequencies of samples with domination events attributed to each genus that are also plotted in **Figure S3b**. The mean and median frequency of samples dominated by each genus across the time windows (vertical bars of **Figure S3b**) is tabulated. For example, among samples collected from the Duke cohort, the mean frequency of domination by *Enterococcus* over time was 13.1%. The top 18 genera within each cohort are shown.

Table S6

MSKCC cohort

Genera	pre-HCT	post-HCT	p value	FDR adjusted
Streptococcus	0.009584	0.117621	4.77E-38	2.96E-50
Enterococcus	0.002072	0.026001	4.87E-50	3.12E-62
Clostridium	0.114629	0.025575	1	1
Lactobacillus	0.001135	0.025010	1.64E-49	1.03E-61
Eubacterium	0.022226	0.018278	0.186210	1
Blautia	0.177585	0.015736	1	1
Actinomyces	0.001059	0.007879	1.67E-24	9.36E-37
Ruminococcus	0.107578	0.005834	1	1
Veillonella	0.000144	0.003551	2.56E-50	1.66E-62
Bacteroides	0.004375	0.002965	0.509099	1
Erysipelatoclostridium	0.006779	0.002697	0.448567	1
Faecalibacterium	0.024954	0.001367	1	1
Rothia	5.63E-05	0.001317	9.72E-37	5.93E-49
Parascardovia	0.000209	0.0012722	1.73E-18	9.33E-31
Staphylococcus	5.44E-05	0.001075	1.15E-58	7.60E-71
Lactococcus	0.000186	0.000978	3.06E-15	1.59E-27
Bifidobacterium	0.007026	0.000957	1	1
Oscillospira	0.005133	0.000861	1	1
Coprococcus	0.024411	0.000764	1	1
unclassified	0.000393	0.000763	0.000278	0.008885
Dorea	0.014056	0.000648	1	1
Escherichia	0.000255	0.000595	1.02E-08	3.68E-07
Akkermansia	0.000206	0.000485	4.60E-09	1.70E-07
Klebsiella	0.000206	0.000467	4.68E-15	1.87E-13
Granulicatella	4.53E-05	0.000460	3.95E-44	2.29E-42
Scardovia	0	0.000367	6.29E-46	3.71E-44
Atopobium	8.81E-06	0.000363	7.04E-43	4.01E-41
Parabacteroides	0.000288	0.000351	0.005805	0.741493
Roseburia	0.007539	0.000296	1	1
Pediococcus	9.29E-06	0.000277	3.13E-24	1.60E-22
Gemella	5.16E-05	0.000274	5.59E-21	2.68E-19
Peptostreptococcus	3.77E-05	0.000256	4.33E-23	2.12E-21
Coprobacillus	0.000963	0.000225	1	1
Megasphaera	0	0.000173	3.75E-32	1.99E-30
Mogibacterium	3.39E-05	0.000161	3.83E-13	1.49E-11
unclas. Peptostreptococcaceae	0.0002531	0.000157	0.929596	1
Propionibacterium	0	0.000154	6.75E-47	4.05E-45
Corynebacterium	0	0.000143	1.34E-34	7.38E-33
Phascolarctobacterium	3.40E-05	0.000134	6.57E-06	0.000217
Butyrivibrio	2.25E-05	0.000112	5.12E-08	1.75E-06
Clostridium	0.000605	0.000111	1	1
Leuconostoc	0	9.19E-05	7.53E-16	3.16E-14
Prevotella	0	9.02E-05	3.19E-15	1.31E-13
Turicibacter	6.28E-05	8.71E-05	0.250066	1
Weissella	0	8.71E-05	1.39E-19	6.40E-18
Alistipes	4.63E-05	7.10E-05	0.350451	1
Eggerthella	0.000345	5.84E-05	1	1
Fusobacterium	0	4.10E-05	6.62E-20	3.11E-18
Enterobacter	0	3.57E-05	5.01E-08	1.75E-06
Sutterella	0	3.14E-05	0.001565	0.048501
Abiotrophia	0	1.15E-05	7.15E-24	3.57E-22
Adlercreutzia	0	0	0.999698	1
Allobaculum	0	0	4.91E-18	2.16E-16
Anaerococcus	0	0	0.045427	1
Anaerofustis	0	0	0.999978	1
Anaeroplasma	0	0	9.12E-10	3.46E-08
Anaerostipes	3.56E-05	0	0.999998	1
Anaerotruncus	0	0	1	1
Bilophila	0	0	0.594937	1
Collinsella	0	0	0.991462	1
Dehalobacterium	0	0	0.902237	1
Dialister	0	0	0.999772	1
Holdemania	0	0	0.999936	1
Lachnobacterium	0	0	1	1
Mycoplasma	0	0	5.22E-17	2.24E-15
Oribacterium	0	0	2.53E-19	1.14E-17

Multicenter validation cohort:

Genera	pre-HCT	post-HCT	p value	FDR adjusted
Enterococcus	0.014575	0.036330	0.000328	0.021348
Clostridium	0.089745	0.021384	0.994378	1
Streptococcus	0.009601	0.015100	0.036976	1
Eubacterium	0.010862	0.012890	0.125806	1
Lactobacillus	0.009066	0.009271	0.381552	1
Blautia	0.036646	0.007191	0.948537	1
Ruminococcus	0.041457	0.006115	0.983023	1
Faecalibacterium	0.014595	0.004206	0.999133	1
Staphylococcus	0.000182	0.003645	5.50E-12	3.63E-10
Bacteroides	0.005864	0.003385	0.822737	1
Erysipelatoclostridium	0.000285	0.002928	0.018971	1
Coprococcus	0.005156	0.001285	0.973864	1
Akkermansia	0.000571	0.001215	0.228781	1
Dorea	0.001380	0.000945	0.532264	1
Oscillospira	0.005900	0.000788	0.998362	1
unclassified	0.001110	0.000664	0.817973	1
Parabacteroides	0.000449	0.000629	0.674802	1
Lactococcus	0.000000	0.000526	0.000462	0.029596
Roseburia	0.001125	0.000526	0.902407	1
Escherichia	0.000586	0.000455	0.965396	1
Bifidobacterium	0.002037	0.000416	0.999639	1
Klebsiella	0.000391	0.000308	0.945188	1
Actinomyces	0.000294	0.000285	0.570908	1
Coprobacillus	1.84E-05	0.000249	0,.50710	1
Veillonella	0.000681	0.000222	0.999123	1
Clostridium	5.53E-05	0.000165	0.174948	1
Pediococcus	5.92E-05	0.000114	0.281845	1
Propionibacterium	0	6.40E-05	0.001645	0.101963
Leuconostoc	0	3.70E-05	0.147866	1
Abiotrophia	0	0	0.382553	1
Adlercreutzia	0	0	0.713186	1
Alistipes	0	0	0.534649	1
Allobaculum	0	0	0.001036	0.065287
Anaerococcus	0	0	0.983413	1
Anaerofustis	0	0	0.301359	1
Anaeroplasma	0	0	0.065151	1
Anaerostipes	0	0	0.997931	1
Anaerotruncus	0	0	0.997181	1
Atopobium	0	0	0.698718	1
Bilophila	0	0	0.193525	1
Butyrivibrio	8.69E-05	0	0.993740	1
Collinsella	0	0	0.029315	1
Corynebacterium	0	0	0.036327	1
Dehalobacterium	0	0	0.477840	1
Dialister	7.40E-05	0	1	1
Eggerthella	0	0	0.741851	1
Enterobacter	0	0	0.752952	1
Fusobacterium	0	0	0.433932	1
Gemella	0	0	0.213045	1
Granulicatella	0	0	0.569902	1
Holdemania	0	0	0.833104	1
Lachnobacterium	0	0	0.995089	1
Megasphaera	4.86E-05	0	0.996650	1
Mogibacterium	0	0	0.349363	1
Mycoplasma	0	0	0.041193	1
Oribacterium	0	0	0.899635	1
Parascardovia	0	0	0.366910	1
Peptostreptococcus	0	0	0.179367	1
Phascolarctobacterium	0	0	0.490673	1
Prevotella	0	0	0.711281	1
Rothia	0	0	0.08097	1
Scardovia	0	0	0.515801	1
Sutterella	5.19E-05	0	0.962346	1
Turicibacter	0.000191	0	0.99996	1
Weissella	0	0	0.280937	1
unclas. Peptostreptococcaceae	0	0	0.895275	1

Supplementary Table S6. FDR-adjusted differences in median genera abundances pre- vs. post-HCT are presented for the MSKCC and the multicenter validation cohorts. For this analysis, the 66 genera that had a relative abundance over 10^{-4} in at least 10% of the samples were included. Pre-HCT abundance was defined as each patient's first sample

in a window of day –30 to –6; post-HCT abundance was calculated as each patient's maximum abundance in a window of day 0 to 21. A Wilcoxon rank test with FDR correction was applied.

Table S7

Genera	BM	BM+T	p value	FDR
Akkermansia	0.2461	0.0691	0.008	0.049
Ruminococcus	0.0586	0.0177	0.008	0.049
Clostridium	0.3265	0.1303	0.008	0.049
Enterococcus	0.0016	0.4064	0.008	0.049
Oscillospira	0.0150	0.0035	0.016	0.066
Adlercreutzia	0.0020	0.0004	0.016	0.066
Dehalobacterium	0.0011	0.0000	0.025	0.079
Anaerostipes	0.0004	0.0000	0.025	0.079
Coprobacillus	0.0108	0.0711	0.056	0.139
Coprococcus	0.0078	0.0018	0.056	0.139
Lachnobacterium	0.0010	0.0000	0.072	0.150
Turicibacter	0.0000	0.0053	0.072	0.150
Bacteroides	0.0019	0.0001	0.139	0.267
Enterobacter	0.0002	0.1213	0.151	0.269
Staphylococcus	0.0002	0.0022	0.161	0.269
unclas. Peptostreptococcaceae	0.0009	0.0028	0.346	0.540
Anaerotruncus	0.0002	0.0006	0.389	0.552
Butyrivibrio	0.0041	0.0007	0.398	0.552
Roseburia	0.0005	0.0000	0.424	0.558
Dorea	0.0006	0.0001	0.526	0.657
Blautia	0.0014	0.0031	0.600	0.690
Mycoplasma	0.0000	0.0001	0.607	0.690
Anaeroplasma	0.0019	0.0010	0.666	0.724
Lactobacillus	0.0726	0.1190	0.841	0.876
Streptococcus	0.0001	0.0001	0.906	0.906

Supplementary Table S7. Table presenting median relative abundances of individual taxa (day 8 after allo-HCT) in BM vs BM+T mice top-ranked according to relative abundance levels; statistical testing by Kruskal-Wallis rank comparison, adjusted by FDR correction.

Table S8

	Control diet	Lactose-free diet
Vendor	LabDiet	LabDiet
Catalogue #	5053 irradiated	5WEV (modified after 5053, no added dried whey) irradiated
Nutrients (%)	Protein 21.0 Fat (ether extract) 5.0 Fat (acid hydrolysis) 6.3	Protein 21.0 Fat (ether extract) 5.2 Fat (acid hydrolysis) 5.8
	Fiber 4.6 Neutral detergent fiber 16.0 Acid detergent fiber 5.8	Fiber 4.6 Neutral detergent fiber 15.6 Acid detergent fiber 5.8
	Nitrogen-Free extract 53.4 Starch 28.2 Sucrose 3.25	Nitrogen-Free extract 53.5 Starch 30.0 Sucrose 3.37
	Minerals 5.9	Minerals 5.7
Vitamins	Carotene (ppm) 1.5 Vitamin A (IU/g) 15 Vitamin D (IU/g) 2.3 Vitamin E (IU/kg) 99 Vitamin K (ppm) 3.3 Thiamin (ppm) 17 Riboflavin (ppm) 8.0 Niacin (ppm) 85 Panthothenic acid (ppm) 17 Folic acid (ppm) 3.0 Pyridoxine (ppm) 9.6 Biotin (ppm) 0.3 Vitamin B12 (mcg/kg) 51 Choline chloride (ppm) 2,000	Carotene (ppm) 1.6 Vitamin A (IU/g) 15 Vitamin D (IU/g) 2.3 Vitamin E (IU/kg) 101 Vitamin K (ppm) 3.3 Thiamin (ppm) 16 Riboflavin (ppm) 2.4 Niacin (ppm) 88 Panthothenic acid (ppm) 17 Folic acid (ppm) 3.0 Pyridoxine (ppm) 9.6 Biotin (ppm) 0.3 Vitamin B12 (mcg/kg) 51 Choline chloride (ppm) 2,000
Calories (%) provided by	Protein 24.5 Fat 13.1 Carbohyrdates 62.4	Protein 24.4 Fat 13.5 Carbohyrdates 62.1

Supplementary Table S8. Ingredients of the conventional laboratory mouse diet at MSKCC that was used as a control diet vs. lactose-free diet for feeding mice in allo-HCT experiments.

Table S9

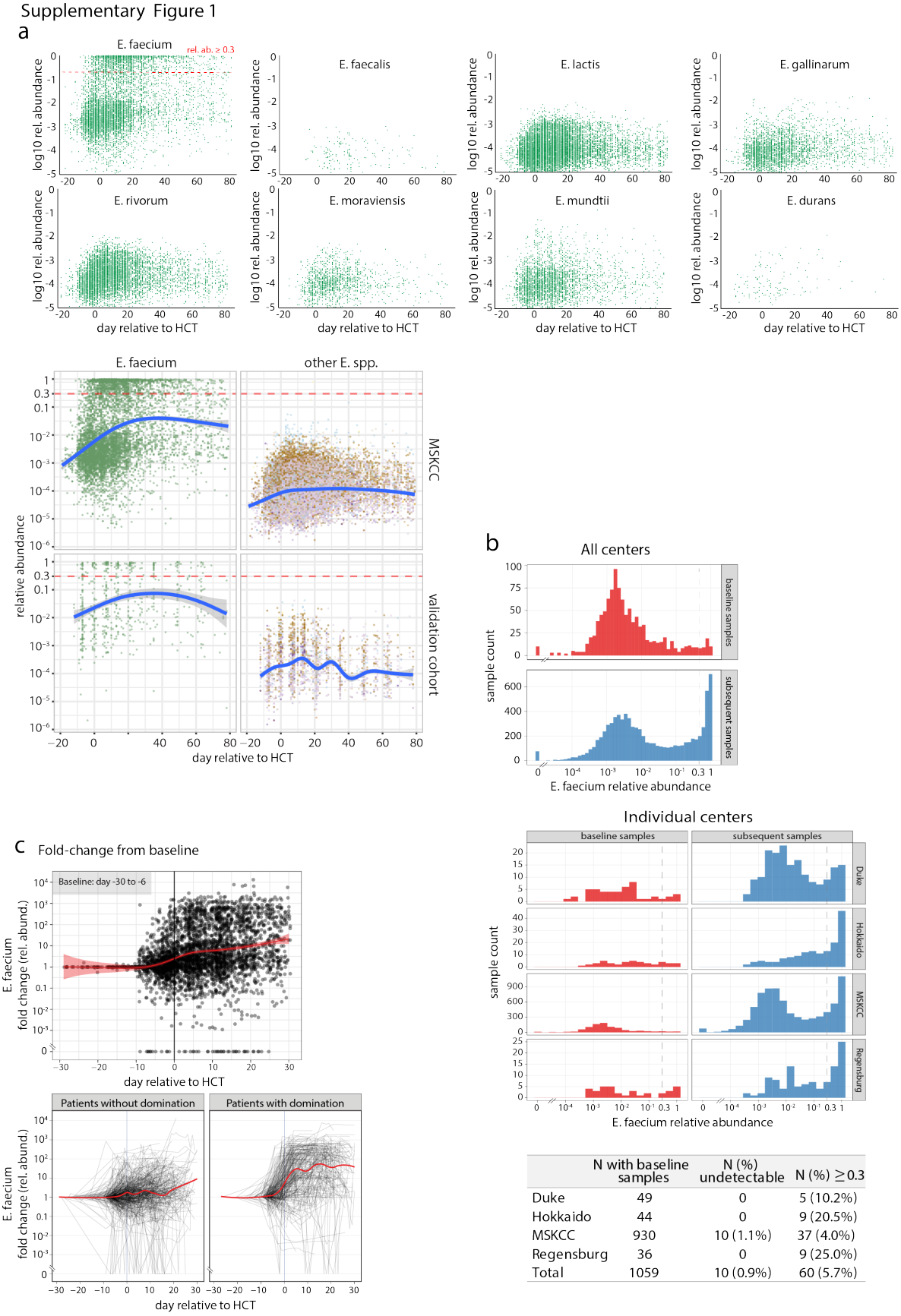
Genera	Ctr	LF	p value	FDR
Clostridium	0.0774	0.2803	0.001	0.049
unclas. Peptostreptococcaceae	0.0002	0.0012	0.003	0.049
Oscillospira	0.0067	0.0338	0.004	0.049
Lactobacillus	0.0249	0.0586	0.010	0.089
Lachnobacterium	0.0003	0.0000	0.013	0.089
rc4.4	0.0418	0.0410	0.017	0.089
Anaerostipes	0.0002	0.0015	0.025	0.104
Leuconostoc	0.0000	0.0000	0.026	0.104
Coprococcus	0.0011	0.0028	0.055	0.198
Sutterella	0.0354	0.0165	0.065	0.214
Olsenella	0.0000	0.0001	0.146	0.386
Prevotella	0.0000	0.0000	0.146	0.386
Butyrivibrio	0.0000	0.0000	0.155	0.386
Bacteroides	0.0329	0.0174	0.182	0.386
Bifidobacterium	0.0008	0.0022	0.191	0.386
Streptococcus	0.0000	0.0000	0.193	0.386
Lactococcus	0.0000	0.0000	0.193	0.386
Adlercreutzia	0.0012	0.0033	0.211	0.400
Ruminococcus	0.0100	0.0130	0.243	0.434
Anaerofustis	0.0000	0.0000	0.277	0.434
Akkermansia	0.3457	0.2141	0.278	0.434
Dehalobacterium	0.0006	0.0009	0.278	0.434
Blautia	0.0004	0.0000	0.303	0.437
Anaeroplasma	0.0037	0.0013	0.315	0.437
Allobaculum	0.0003	0.0005	0.315	0.437
Staphylococcus	0.0001	0.0000	0.477	0.636
Filifactor	0.0001	0.0000	0.517	0.665
Paraprevotella	0.0000	0.0000	0.563	0.699
Anaerotruncus	0.0002	0.0002	0.652	0.758
Candidatus_Arthromitus	0.0000	0.0000	0.656	0.758
Dorea	0.0000	0.0000	0.674	0.758
Propionibacterium	0.0000	0.0001	0.775	0.845
Coprobacillus	0.0005	0.0004	0.902	0.955
Chlamydia	0.0001	0.0002	0.935	0.962
Turicibacter	0.0215	0.0197	0.968	0.968

Supplementary Table S9. Comparison of relative abundances of taxa at genus level in GVHD mice at day 7 after T cell replete transplants that received either control (Ctr) or lactose-free chow (LF) [LP/J → C57BL/6 allo-HCT]; ranking according to magnitude of differences; see Figure 4b for relative abundance plot of the genus *Enterococcus*. Feeding of LF to transplanted mice was associated with increased abundances of *Clostridium*, *Oscillospora* and unclassified peptostreptococci (after FRD adjustment). Statistical testing by Kruskal-Wallis rank comparisons adjusted by FDR correction.

Table S10

Protocol	Target	TaqMan ID	Species
Gene expression	Reg3B	Mm00440616_g1	mouse
	Reg3G	Mm00441127_m1	mouse
	lactase	Mm01285112_m1	mouse
	GAPDH	Mm99999915_g1	mouse
SNP genotyping	rs4988235, C/T(-1390) SNP		human

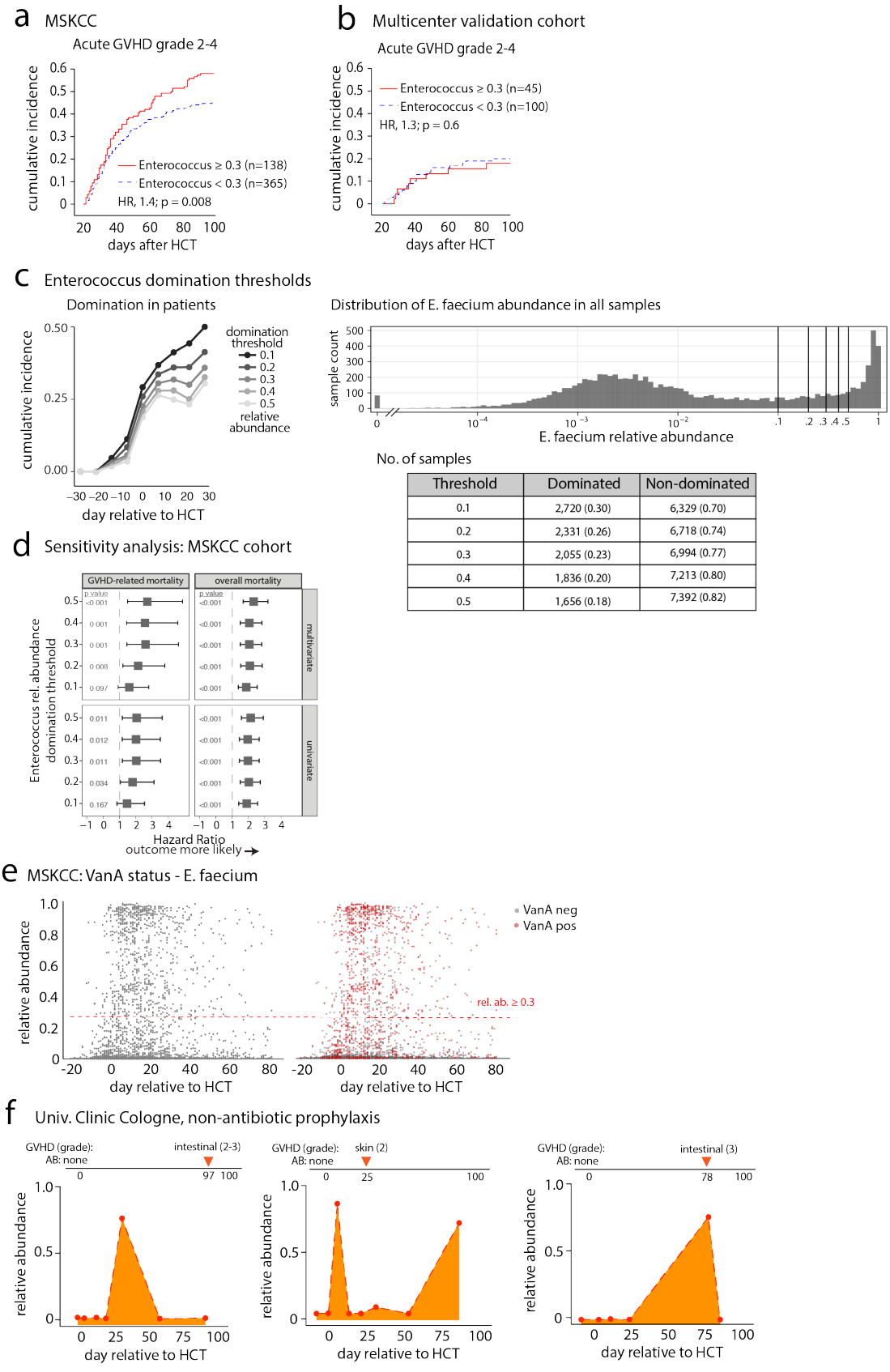
Supplementary Figure 1



	N with baseline samples	N (%) undetectable	N (%) ≥0.3
Duke	49	0	5 (10.2%)
Hokkaido	44	0	9 (20.5%)
MSKCC	930	10 (1.1%)	37 (4.0%)
Regensburg	36	0	9 (25.0%)
Total	1059	10 (0.9%)	60 (5.7%)

Supplementary Figure 1. (a) *Top 8 panels*: Relative abundance (log_{10} transformed) of *Enterococcus spp*. in fecal samples of MSKCC patients collected at different time points relative to allo-HCT (dotted line indicates domination threshold at 0.3 relative abundance). *Bottom 4 panels*: A version of **Figure 1b** plotted on a log_{10} vertical axis that extends to a relative abundance of 10^{-6}. A quantitative analysis of all species within genus *Enterococcus* is tabulated in **Supplemental Table S2**. (b) In baseline samples (first sample per patient collected between days -30 and -6), most samples had a low level of *E. faecium* abundance in a unimodal distribution centered at ~10^{-3}. *E.* faecium was undetectable in 10 (0.9%) samples, and domination with *E. faecium* was observed in 60 (5.7%) of baseline samples. In subsequent samples, *E. faecium* abundance developed into a bimodal distribution with a second peak of samples with relative abundances above the domination threshold. *Top*, histograms of relative abundance of *E. faecium* at baseline and in the subsequent samples using merged data of all centers (MSKCC and multicenter-validation cohort; dotted line indicates domination threshold at 0.3 rel. abundance); *Middle*, distribution analysis for each center individually; *Bottom*, tabulated numbers of patients at each center with either undetectable or already-dominated samples at baseline. (c) Analysis of *E. faecium* expansion expressed as fold-change normalized to each patient's baseline sample. *Top*, *E. faecium* fold-change in 4,039 samples from 563 patients who had a baseline sample collected between day -30 and day -6 and at least one subsequent sample collected up to day +30. Ten patients had undetectable *E. faecium* abundance in baseline samples (i.e, zero rel. abundance); a small constant (1×10^{-4}) was added to all samples to allow fold-change calculations. Both fold-increases and fold-decreases were observed; considering all samples collected between days 14-21, *E. faecium* abundance increased by a median of 6.4-fold and by a mean of 143.6-fold. *Bottom*, the same data are plotted with a trendline for each patient and in separate panels for those patients who became dominated vs. those who did not. Considering all samples collected between days 14-21 from patients who became dominated at some point, *E. faecium* increased by a median of 120-fold and a mean of 332-fold. Considering all samples collected in the same time period from patients who never became dominated, *E. faecium* increased by a median of 2-fold and a mean of 55.6-fold. In all panels, blue and red lines indicate smoothed average trends.

Supplementary Figure 2

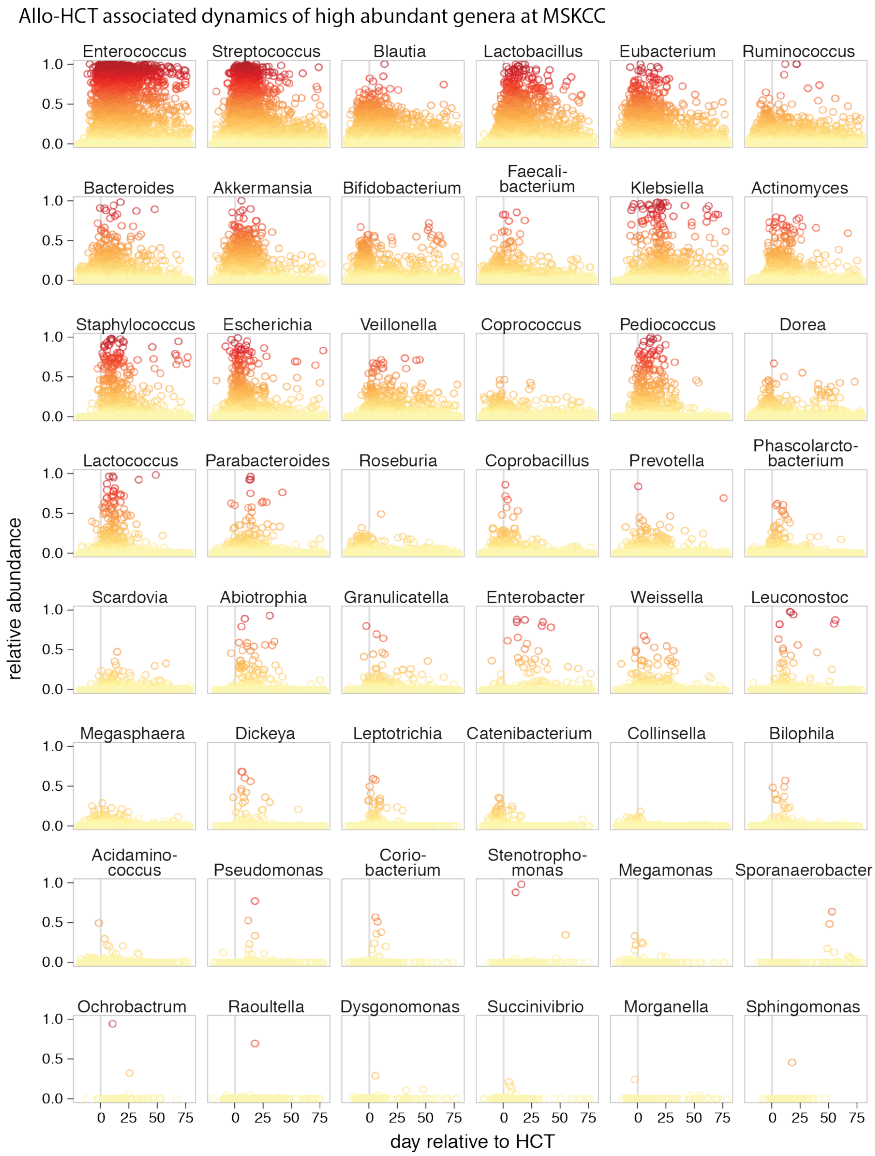


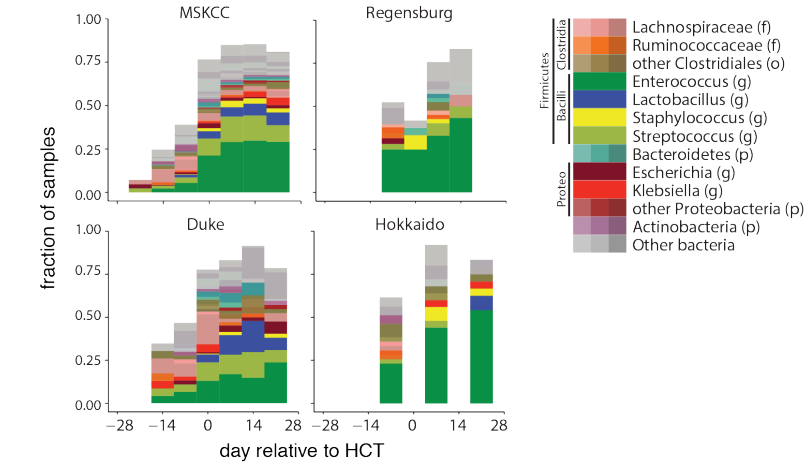
Threshold	Dominated	Non-dominated
0.1	2,720 (0.30)	6,329 (0.70)
0.2	2,331 (0.26)	6,718 (0.74)
0.3	2,055 (0.23)	6,994 (0.77)
0.4	1,836 (0.20)	7,213 (0.80)
0.5	1,656 (0.18)	7,392 (0.82)

1

Supplementary Figure 2. (a) Cumulative incidence of grade 2-4 acute GVHD (landmark analysis of survivors beyond day +21) in MSKCC patients with genus *Enterococcus* domination vs. non-domination. (b) Cumulative incidence grade 2-4 acute GVHD in the multicenter validation cohort. (c) Domination threshold sensitivity analysis with the total dataset for the cumulative incidence of *Enterococcus* domination. *Left*, the cumulative incidence of domination dramatically increases around day 0 and then reaches a plateau with relatively similar kinetics across all tested threshold definitions, although the height of the plateau varies with the definition such that fewer patients are considered to have had domination events at higher, more stringent thresholds. *Right upper*, distribution of *E. faecium* relative abundances in the entire dataset with the cut-offs tested in the sensitivity analysis indicated. *Right lower*, the number of samples in the entire dataset that would be defined as dominated by *E. faecium* at each cut-off tested. (d) Sensitivity analysis for the association of genus *Enterococcus* domination with GVHD-related mortality and overall survival: The domination predicts higher risk of all-cause mortality with high statistical significance at all tested domination thresholds and in both univariate and multivariate Cox models. Genus *Enterococcus* domination also predicts GVHD-related mortality at thresholds between 0.2 and 0.5, although in the lowest, least stringent threshold of 0.1, there was no longer a significant association for this outcome. The multivariate Cox models were adjusted for graft source, age, conditioning intensity, gender, and underlying disease (leukemia vs. other). (e) In a subset of 3,833 samples from 406 patients, presence of the VanA gene in fecal samples was assessed by PCRs. 152 patients were *vanA* positive (i.e., vancomycin resistant), and among them 115 patients had stool samples with a relative abundance of *E. faecium* ≥0.3 (red dotted line). (f) *Enterococcus* expansion and blood-stream infections after allo-HCT have been attributed to antibiotic exposure (*5*, *6*, *49*). The vast majority of patients in our study received antibiotics (prophylactic and/or therapeutic antibiotics). To explore whether enterococcal domination can occur in patients in the absence of antibiotics, we profiled fecal microbiota communities collected at University Clinic of Cologne, where samples were available from patients who did not receive prophylactic antibiotics. Presented are three anecdotal case in which genus *Enterococcus* domination was observed in allo-HCT patients who did not receive antibiotic prophylaxis nor antibiotics for treatment for neutropenic fever or infections. All three patients subsequently developed either skin or gut GVHD.

Supplementary Figure 3

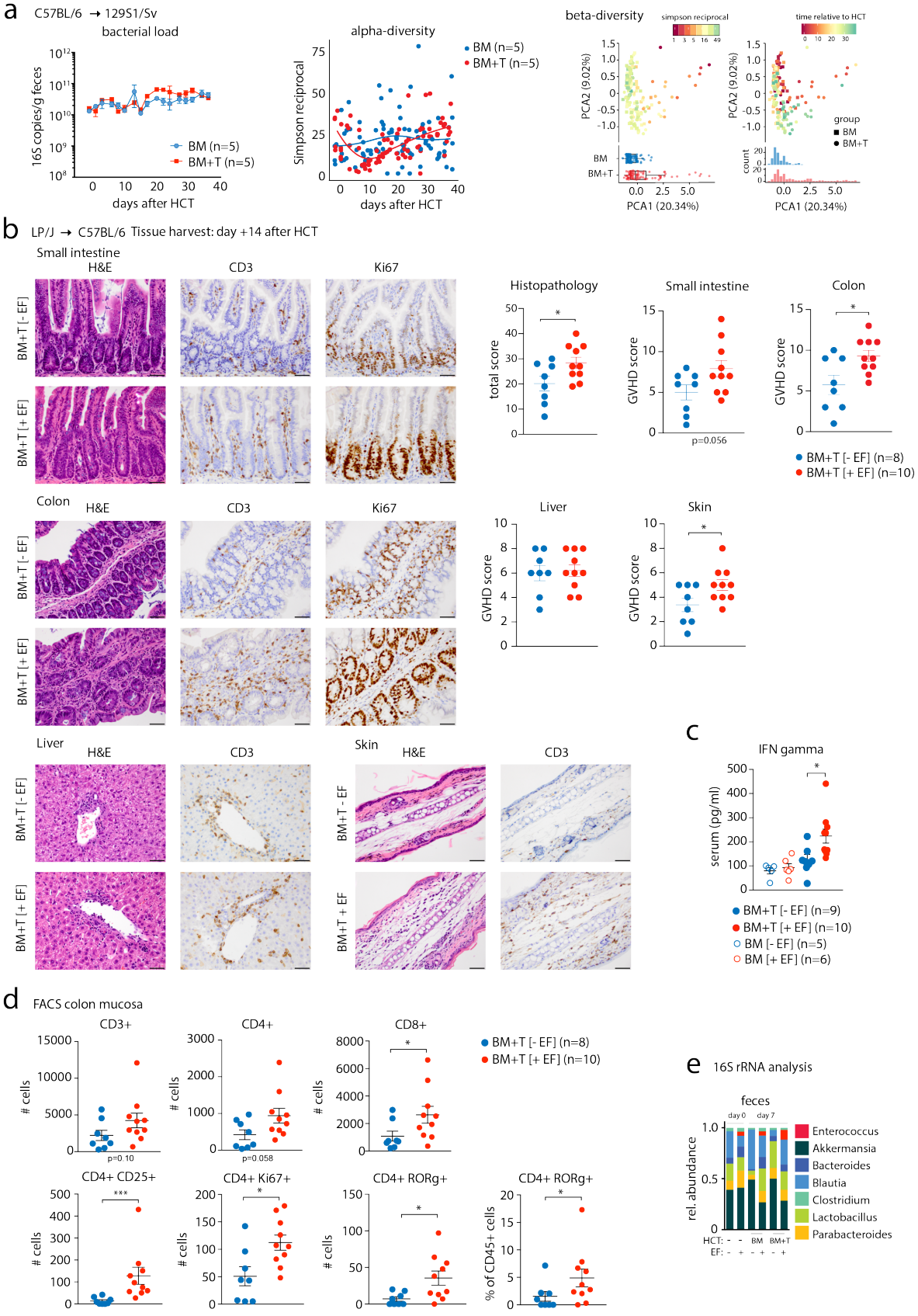




1

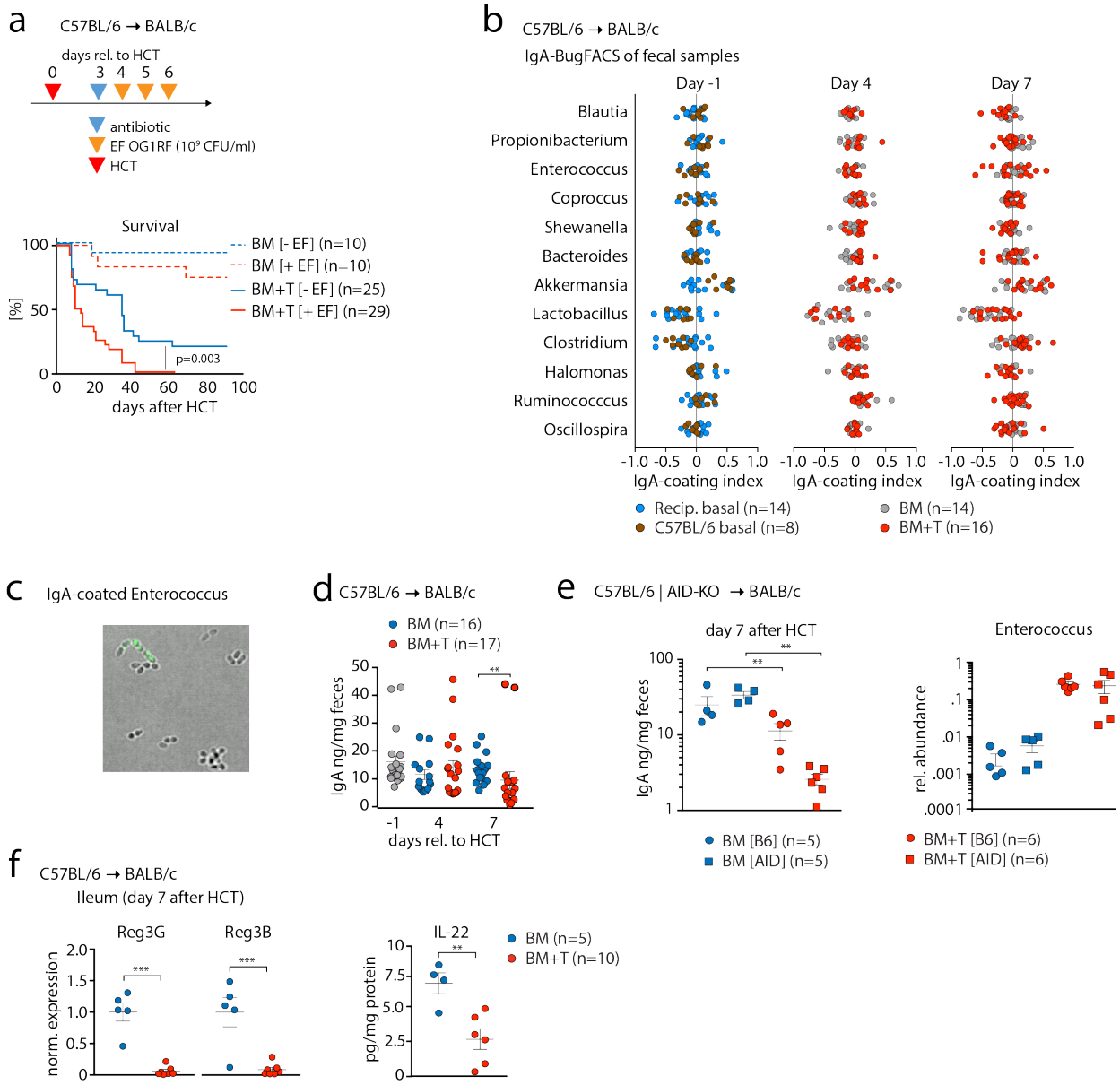
Supplementary Figure 3. (a) Relative abundance dynamics of selected genera in 7,321 fecal samples from 1056 recipients of allo-HCT at MSKCC. *Enterococcus* showed the most pronounced bloom after transplantation followed by *Streptococcus*, *Blautia* and *Lactobacillus*. Plotted are the 48 genera with the highest mean abundances across all samples, in decreasing order of mean abundance. The yellow-to-red color scale recapitulates the y-axis abundance values for emphasis. (b) Taxa contributing to domination events in the MSKCC and the 3 centers in the Multicenter Validation Cohorts. Domination was defined at the level of OTUs; color-coding of higher-rank taxa is defined in the color legend. Samples were binned in 7 day-windows; bars are plotted for windows that contain at least 5 samples. The MSKCC panel includes the recipients of T-replete grafts analyzed in this study. The data plotted in these barplots are tabulated in **Supplemental Table S5.** A similar but distinct analysis of a larger cohort that included recipients of T-cell-depleted grafts is presented in a separate manuscript.

Supplementary Figure 4



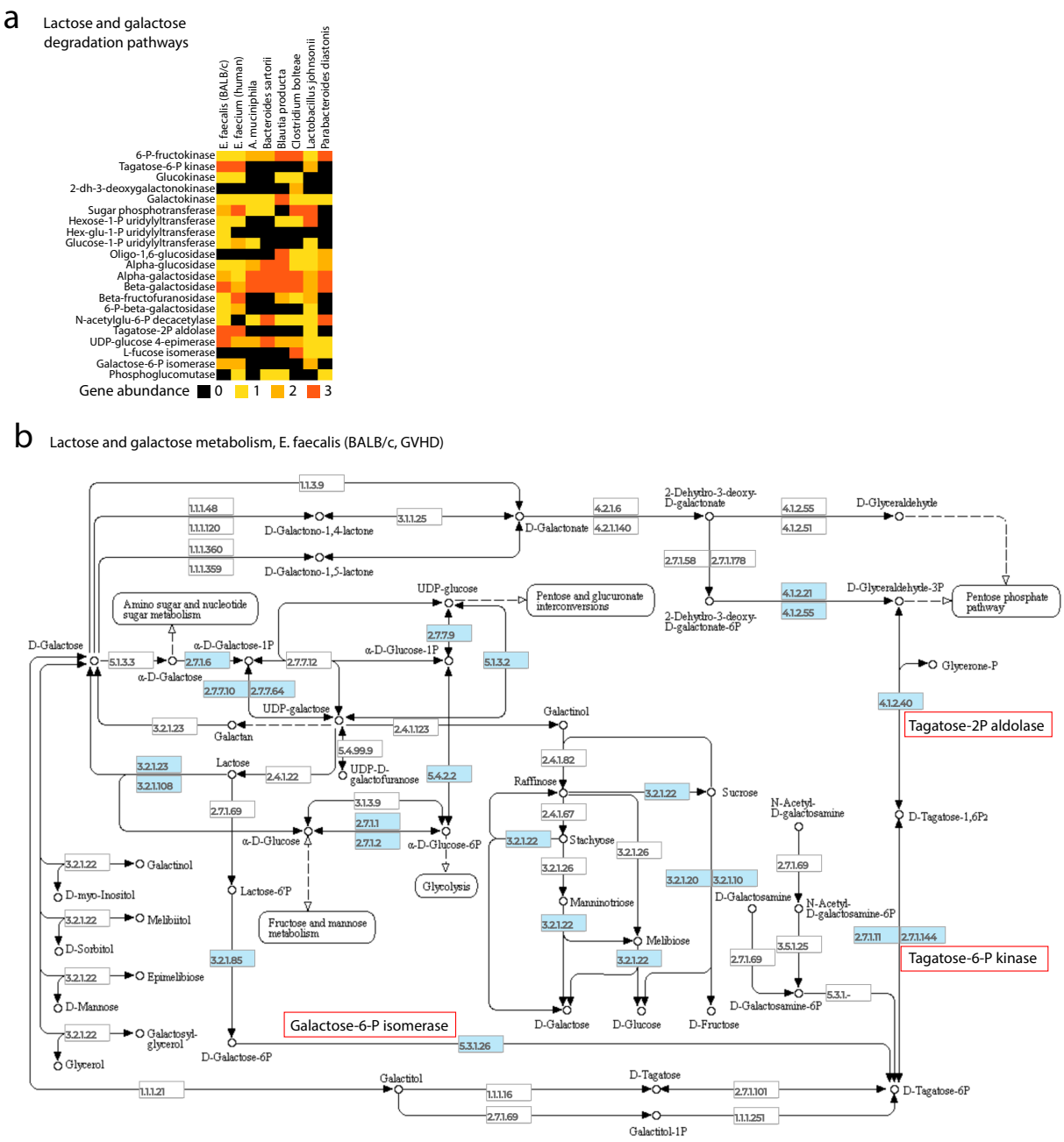
Supplementary Figure 4. (a) *Left*, in transplanted mice that develop GVHD or BM-only controls, the overall bacterial load remains relatively stable, as measured by quantitative PCR using universal 16S primers in serially collected fecal samples from 129S1/Sv recipients of C57BL/6 bone marrow (BM) or T cell-replete BM (BM+T, $4x10^6$ T cells); *Middle*, alpha diversity transiently decreases in BM+T recipients around day 10. *Right*, representation of Bray-Curtis beta-diversity in a PCoA plot. (b) Representative images of histopathological sections of several GVHD target organs in gnotobiotic C57BL/6 mice that were colonized with the bacterial 6-strain mix spiked with *E. faecalis OG1RF* in the +EF group (see **Figure 2**) and tissue was analyzed 14 days after HCT. Compound histopathological scores (including all four organs) vs. GVHD scores for each organ are displayed; the scale bars in the right lower corner represent 50μm. (c) Serum IFNγ concentrations of gnotobiotic mice harvested 14 days after allo-HCT. IFNγ is a pleiotropic cytokine that is produced in large amounts by Th1 and Tc1 cells early after HCT and is the archetypal "Th1" cytokine generated during GVHD. IFNγ is pathogenic in the development of intestinal GVHD (*50*), enhancing the sensitivity of macrophages to LPS, hence increasing their production of proinflammatory cytokines (*51*) in addition to causing crypt hypertrophy and villous atrophy (*52, 53*) via direct signaling of the IFNGR expressed on recipient gut tissue (*50*). (d) Flow cytometric analysis of lamina propria isolates from colon tissues in the recipient gnotobiotic mice on day 14 after HCT. Number of cells per $1x10^6$ cells. Our observed increase in proliferation of CD4+ T cells is consistent with the increased infiltration of antigen-specific donor T cells, which occurs in the gastrointestinal tract of GVHD mice (*16*). (e) Median relative abundances of each member of the 6-strain community +/- *E. faecalis* in fecal samples collected at the day of HCT and 7 days later. Data are combined from two independent experiments in b and c. Values represent mean ± S.E.M. *$p<0.05$.

Supplementary Figure 5



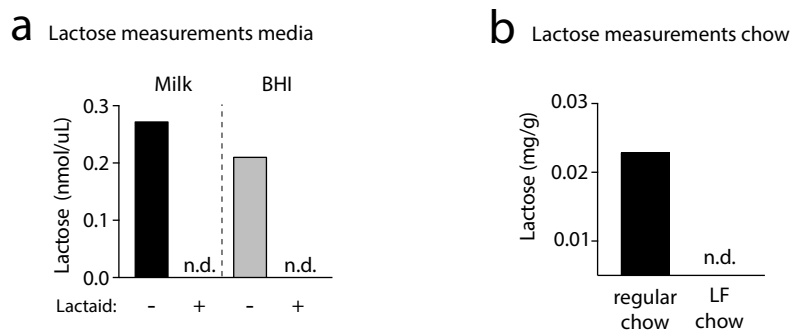
Supplementary Figure 5. (a) Survival data combined from three experiments of BALB/c mice receiving either C57BL/6 BM or BM+T (5x10^5 T cells), and received oral gavages of *E. faecalis OG1RF* for three consecutive days after a one-time vancomycin (25mg/kg, p.o.) gavage to facilitate intestinal strain engraftment. (b) IgA – BugFACS on fecal samples from C57BL/6 donors and BALB/c recipients at different days relative to HCT (BM vs. BM + 1x10^6 T cells). IgA-coating presented as IgA-coating index (ICI) for several genera (ICI 0-1: IgA+ fraction, ICI -1-0: IgA- fraction). (c) Immunofluorescence microphotograph (by confocal laser microscopy) of *E. faecalis* incubated with polyclonal human IgA and stained with anti-human IgA – Alexa 488. (d) IgA in the feces of BALB/c mice at baseline and 4 and 7 days after HCT with BM or BM+T (1x10^6 T cells) of C57BL/6 mice. (e) Left, IgA in the feces and, right, *Enterococcus* relative abundance in the stool of BALB/c mice transplanted with either BM or BM+T (1 Mio T cells) of wildtype C57BL/6 mice or of AID-KO mice lacking IgA. (f) mRNA expression of Reg3B and Reg3G (relative to GADPH and normalized to BM as controls) and concentration of IL-22 (as pg protein per mg total protein) in the ileum of BALB/C mice transplanted with C57BL6 BM or BM+T (1x10^6 T cells). Values represent mean ± S.E.M. *p<0.05, **p<0.01, ***p<0.001 (Student's T test).

Supplementary Figure 6



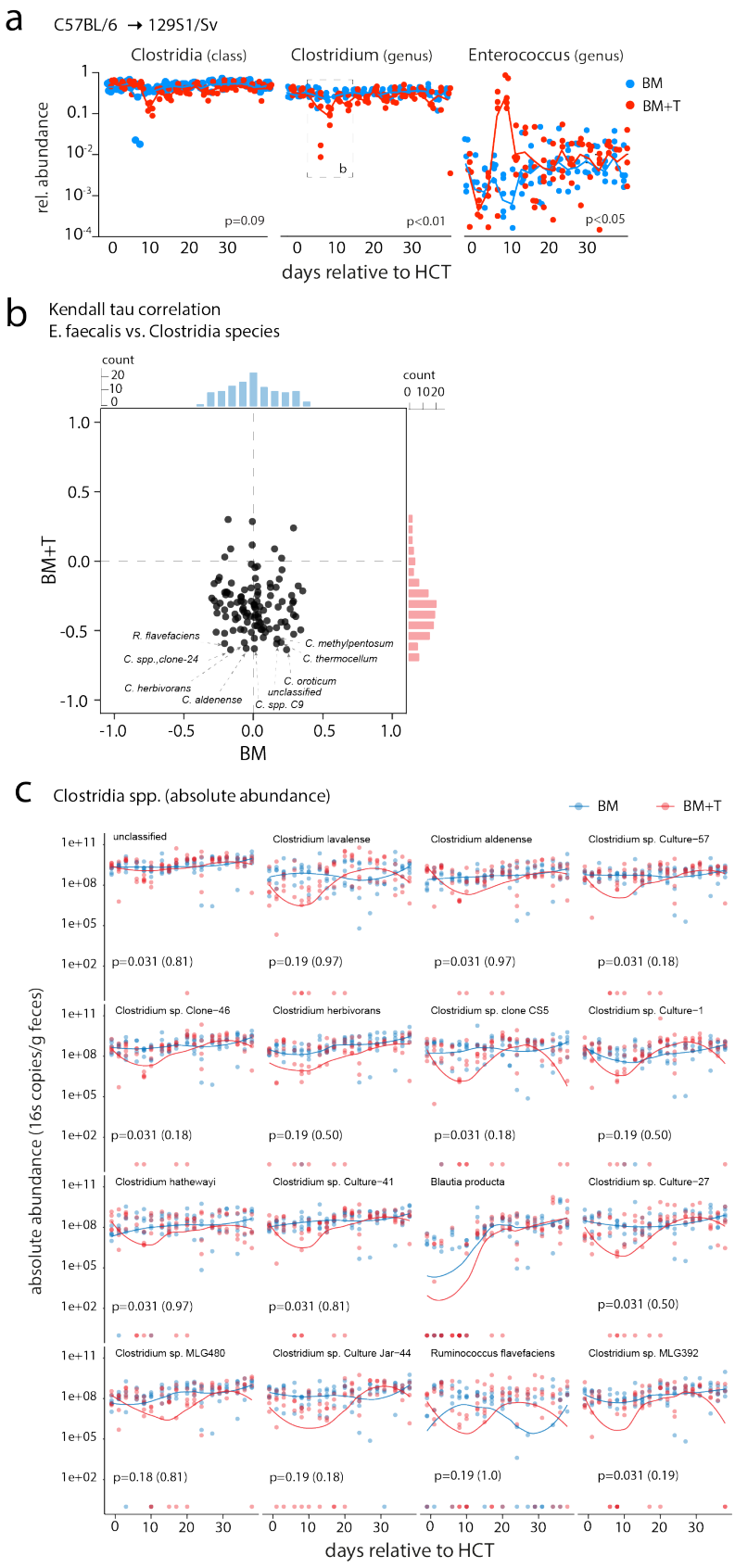
Supplementary Figure 6. (a) Comparing the *Enterococcus* genomes (mouse GVHD and human isolate from Figure 3c) and publicly available genomes (from PATRIC 3.5.27) of other members of the gnotobiotic 6-strain consortium revealed that enterococci are enriched in enzymes for the Leloir pathway (common lactose utilization), and in the tagatose pathway: Heat map of gene abundances of individual genes of the lactose and galactose degradation pathway. (b) Overview of the lactose and galactose degradation pathways in bacteria based on pathway analysis and reconstruction by ModelSEED v2.3. Highlighted in red boxes are enzymes that are predominantly found in *E. faecalis* and *E. faecium*.

Supplementary Figure 7



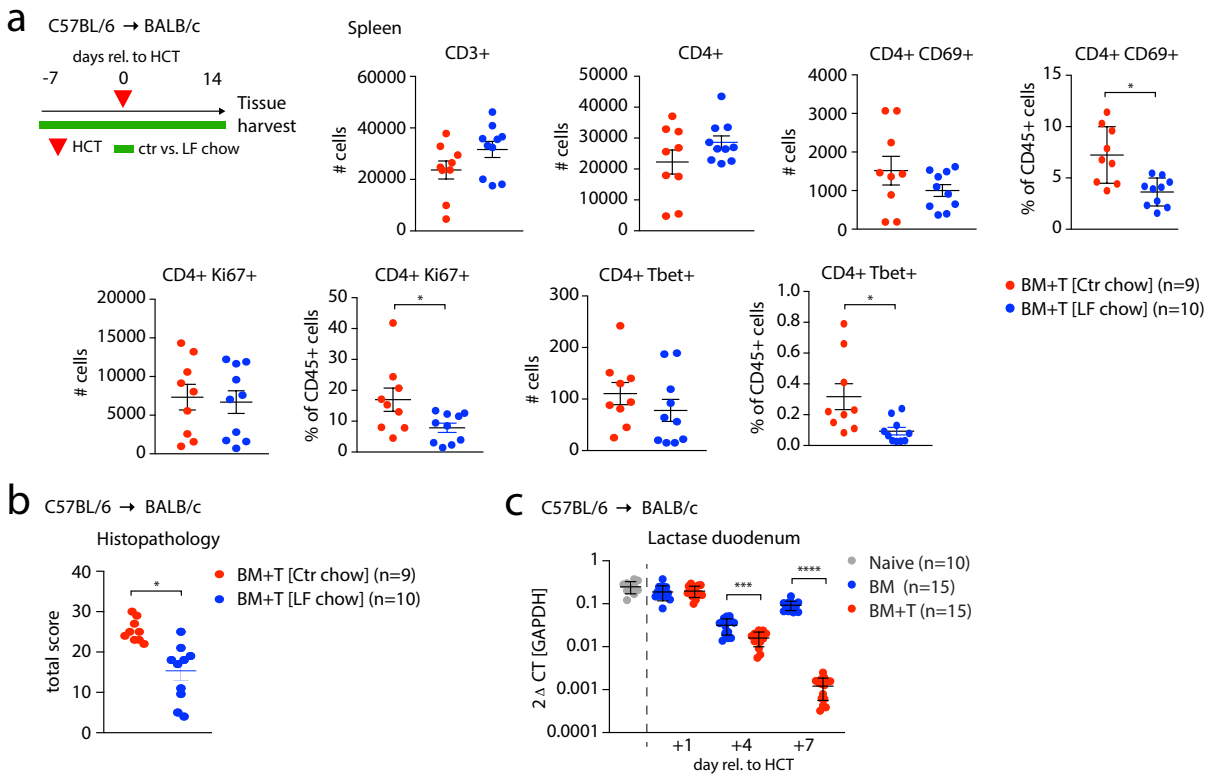
Supplementary Figure 7. (a) Lactase treatment reduced the amount of lactose in regular milk and BHI broth to non-detectable (n.d.) concentrations; the levels of glucose were unaffected by the treatment (3 replicates: regular BHI = 178.3 mg/dl, lactase-pretreated BHI = 181.7 mg/dl). (b) Lactose contents of regular chow or lactose-free chow (5WEV) measured in chow homogenates by a fluorometric lactose assay.

Supplementary Figure 8



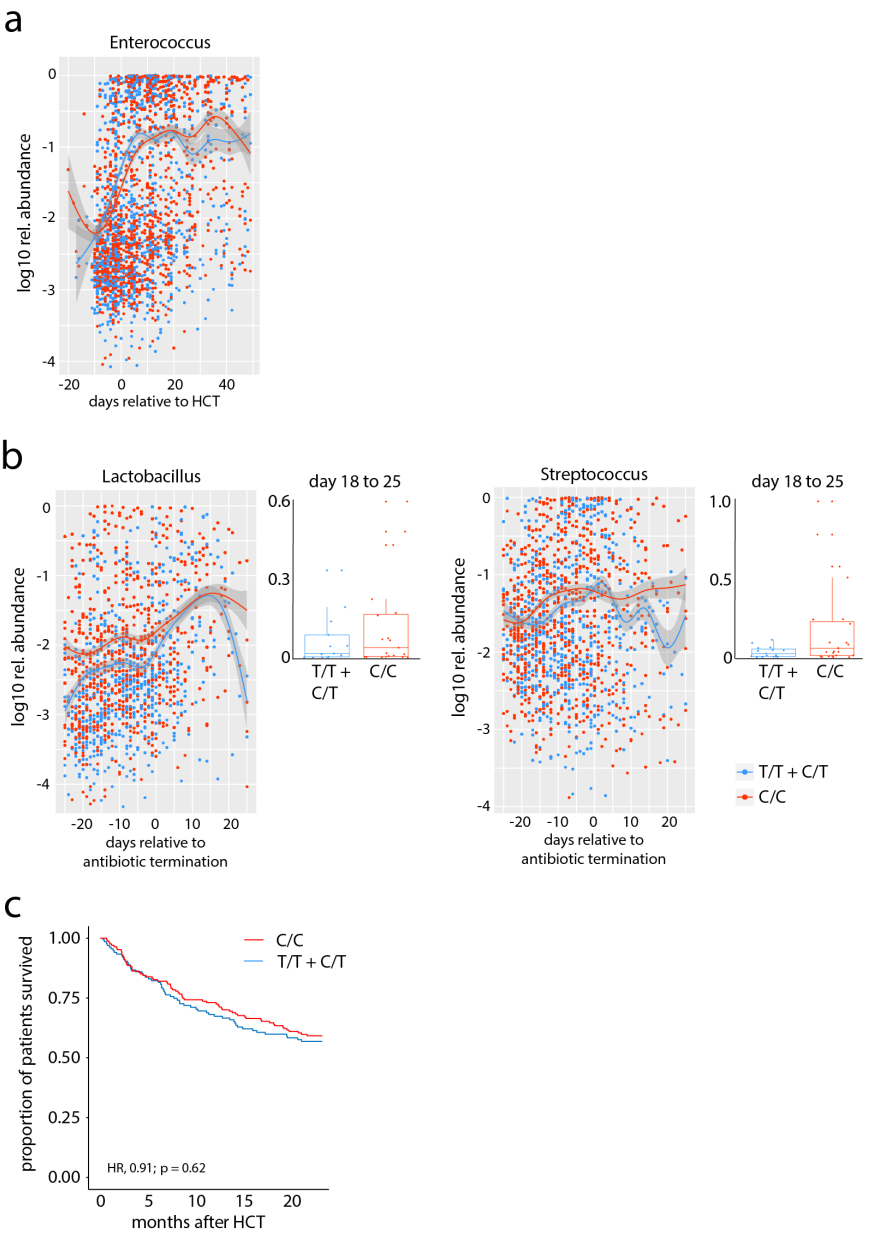
Supplementary Figure 8. (a) Relative abundances of class Clostridia, genus *Clostridium*, and genus *Enterococcus* are plotted over time relative to allo-HCT for mice that develop GVHD (BM+T) vs. transplanted, non-GVHD controls (BM; T cell depleted). These data come from the C57BL/6 into 129S1/Sv HCT shown in Figure 2A. Statistical comparison of areas under the curves was by the AUC-Vardi test. The dotted rectangle indicates samples within the time period day +3 to +13 that were used for correlation statistics in b. (b) Scatter plot of rank correlation between 146 different *Clostridia spp*. and *E. faecalis* (relative abundances; within selected time period after HCT). Within the resolution of V4-V5 16S rRNA amplicon sequencing, species were annotated against the Greengenes and the NCBI 16S rRNA taxonomy databases. These data show that clostridial abundances are negatively correlated with *E. faecalis* relative abundance in the BM+T group ($p<1e-20$, binomial test) consistent with an expansion of *E. faecalis* at the cost of clostridia, but not in BM only mice ($p=0.72$, binomial test). The 10 most negatively correlated *Clostridia spp.* are labeled in the figure. (c) The 16 most abundant *Clostridia spp.* found in the microbiota of transplanted mice are shown. We analyzed contraction of each *Clostridia spp.* in the BM+T vs. BM only groups by measuring the difference between day -1 and day +8 absolute abundances and computed paired binomial tests for each species and transplant group separately. Within each graph, we present p-values for these 16 *Clostridia spp.* (except for *B. producta* where values were missing for paired statistics). Here, we observed that contraction of clostridia occurs primarily in GVHD cases, but not in BM only mice. The contraction over all species is significant only in the BM+T group ($p = 2.061e-42$), but not in BM mice. N = 5 mice / group.

Supplementary Figure 9



Supplementary Figure 9. (a) *Left*, schematic of experimental workflow of HCT in BALB/c recipients of C57BL/6 T cell-replete bone marrow (BM+T, $5x10^5$ T cells) maintained on control (Ctr) or lactose-free (LF) chow. Organs were harvested on day 14 after HCT and used for flow-cytometric analysis for T cell subsets within spleens; two experiments combined. (b) Compound histopathology GVHD score combining small and large intestine, liver and skin pathology of mice treated with control or LF chow; harvest at day 14 after allo-HCT. (c) mRNA levels of lactase in the duodenum of BALB/c mice at steady state, 1, 4 or 7 days after HCT of C57BL/6 BM or BM+T ($1x10^6$ T cells). Values represent mean ± S.E.M. *p<0.05; ***p<0.001, ****p<0.0001 (Student's T test).

Supplementary Figure 10



Supplementary Figure 10. (a) Comparable relative abundances (log10) of *Enterococcus* (genus level) at different time points before and after allo-HCT (day -20 to 50) in MSKCC patients (n=602) dichotomized in rs4988235 lactase SNP genotypes (-13910*T SNP: C/C = 315 patients; T/T = 110 patients; C/T = 177 patients [T/T + C/T = 287 patients]). We reasoned that this might be because antibiotic exposures are likely a major driver of *Enterococcus* expansion in patients (*5, 6*) and that this might overcome any differences in lactose delivery to the lower intestine. We therefore hypothesized that nutritional factors like lactose utilization may affect microbiota compositions in this patient cohort only after antibiotic cessation. (b) We focused on recovery from antibiotic-induced *Enterococcus* expansion by limiting our analysis to patients who had been exposed to one of the three major broad-spectrum antibiotics used in this cohort for empiric treatment of neutropenic fevers (piperacillin-tazobactam i.v., imipenem-cilastatin i.v., or meropenem i.v.) and by synchronizing patient time courses relative to the day of antibiotic cessation. Relative abundances of *Lactobacillus* and *Streptococcus* (two other major genera within the *Lactobacillales* order); data are presented as days relative to the day of last administration of antibiotics; box plot-inserts display the median relative abundances of *Lactobacillus* (left) or *Streptococcus* (right) of time binned (in days) between day 18 to 25 after last day of antibiotic administration. (c) Overall survival of MSKCC patients dichotomized into either C/C (n = 175) or C/T+T/T genotypes (n = 140); allo-HCT recipients of bone marrow, and peripheral blood stem cells (recipients of T cell-depleted grafts were excluded in this analysis).

REFERENCES

1. Y. Litvak, M. X. Byndloss, A. J. Baumler, Colonocyte metabolism shapes the gut microbiota. *Science* **362**, (2018).
2. M. S. Gilmore, F. Lebreton, W. van Schaik, Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Curr Opin Microbiol* **16**, 10-16 (2013).
3. S. Schloissnig *et al.*, Genomic variation landscape of the human gut microbiome. *Nature* **493**, 45-50 (2013).
4. F. Lebreton *et al.*, Tracing the Enterococci from Paleozoic Origins to the Hospital. *Cell* **169**, 849-861.e813 (2017).
5. Y. Taur *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* **55**, 905-914 (2012).
6. C. Ubeda *et al.*, Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J Clin Invest* **120**, 4332-4341 (2010).
7. E. Holler *et al.*, Metagenomic analysis of the stool microbiome in patients receiving allogeneic stem cell transplantation: loss of diversity is associated with use of systemic antibiotics and more pronounced in gastrointestinal graft-versus-host disease. *Biol Blood Marrow Transplant* **20**, 640-645 (2014).
8. C. D. Ford *et al.*, Vancomycin-Resistant Enterococcus Colonization and Bacteremia and Hematopoietic Stem Cell Transplantation Outcomes. *Biol Blood Marrow Transplant* **23**, 340-346 (2017).
9. H. J. Khoury *et al.*, Improved survival after acute graft-*versus*-host disease diagnosis in the modern era. *Haematologica* **102**, 958-966 (2017).
10. Y. Shono *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* **8**, 339ra371 (2016).
11. Y. Taur *et al.*, The effects of intestinal tract bacterial diversity on mortality following allogeneic hematopoietic stem cell transplantation. *Blood* **124**, 1174-1182 (2014).
12. R. R. Jenq *et al.*, Intestinal Blautia Is Associated with Reduced Death from Graft-versus-Host Disease. *Biol Blood Marrow Transplant* **21**, 1373-1383 (2015).
13. Y. Taur *et al.*, Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. *Sci Transl Med* **10**, (2018).
14. N. Steck *et al.*, Enterococcus faecalis metalloprotease compromises epithelial barrier and contributes to intestinal inflammation. *Gastroenterology* **141**, 959-971 (2011).
15. N. Geva-Zatorsky *et al.*, Mining the Human Gut Microbiota for Immunomodulatory Organisms. *Cell* **168**, 928-943.e911 (2017).
16. K. Riesner, M. Kalupa, Y. Shi, S. Elezkurtaj, O. Penack, A preclinical acute GVHD mouse model based on chemotherapy conditioning and MHC-matched transplantation. *Bone Marrow Transplant* **51**, 410-417 (2016).
17. K. Dubin, E. G. Pamer, Enterococci and Their Interactions with the Intestinal Microbiome. *Microbiology spectrum* **5**, (2014).
18. R. R. Jenq *et al.*, Regulation of intestinal inflammation by microbiota following allogeneic bone marrow transplantation. *J Exp Med* **209**, 903-911 (2012).
19. S. Caballero *et al.*, Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. *Cell Host Microbe* **21**, 592-602.e594 (2017).

20. N. W. Palm *et al.*, Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000-1010 (2014).
21. K. Brandl *et al.*, Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. *Nature* **455**, 804-807 (2008).
22. D. Zhao *et al.*, Survival signal REG3α prevents crypt apoptosis to control acute gastrointestinal graft-versus-host disease. *J Clin Invest* **128**, 4970-4979 (2018).
23. C. A. Lindemans *et al.*, Interleukin-22 promotes intestinal-stem-cell-mediated epithelial regeneration. *Nature* **528**, 560-564 (2015).
24. K. H. Schleifer, A. Hartinger, F. Götz, Occurence of D-tagatose-6-phosphate pathway of D-galactose metabolism among staphylococci. *FEMS Microbiology Letters* **3**, 9-11 (1978).
25. D. Weber *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* **23**, 845-852 (2017).
28. P. A. Gill, M. C. van Zelm, J. G. Muir, P. R. Gibson, Review article: short chain fatty acids as potential therapeutic agents in human gastrointestinal and inflammatory disorders. *Aliment Pharmacol Ther* **48**, 15-34 (2018).
27. N. D. Mathewson *et al.*, Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. *Nat Immunol* **17**, 505-513 (2016).
28. B. W. Haak *et al.*, Impact of gut colonization with butyrate-producing microbiota on respiratory viral infection following allo-HCT. *Blood* **131**, 2978-2986 (2018).
29. C. J. Ingram, C. A. Mulcare, Y. Itan, M. G. Thomas, D. M. Swallow, Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet* **124**, 579-591 (2009).
30. J. Vydra *et al.*, Enterococcal bacteremia is associated with increased risk of mortality in recipients of allogeneic hematopoietic stem cell transplantation. *Clin Infect Dis* **55**, 764-770 (2012).
31. S. Manfredo Vieira *et al.*, Translocation of a gut pathobiont drives autoimmunity in mice and humans. *Science* **359**, 1156-1161 (2018).
32. S. G. Kim *et al.*, Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. *Nature* **572**, 665-669 (2019).
33. M. Levy, A. A. Kolodziejczyk, C. A. Thaiss, E. Elinav, Dysbiosis and the immune system. *Nat Rev Immunol* **17**, 219-232 (2017).
34. J. R. Galloway-Pena *et al.*, The role of the gastrointestinal microbiome in infectious complications during induction chemotherapy for acute myeloid leukemia. *Cancer* **122**, 2186-2196 (2016).
35. A. Staffas *et al.*, Nutritional Support from the Intestinal Microbiota Improves Hematopoietic Reconstitution after Bone Marrow Transplantation in Mice. *Cell Host Microbe* **23**, 447-457 e444 (2018).
36. T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
37. T. Tatusova *et al.*, in *The NCBI Handbook [Internet]*. (National Center for Biotechnology Information (US), 2014).
38. T. Z. DeSantis *et al.*, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069-5072 (2006).
39. J. G. Caporaso *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336 (2010).
40. P. D. Schloss *et al.*, Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**, 7537-7541 (2009).

41. J. Lloyd-Price *et al.*, Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61-66 (2017).
42. T. Brettin *et al.*, RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* **5**, 8365 (2015).
43. M. Viladomiu *et al.*, IgA-coated E. coli enriched in Crohn's disease spondylarthritis promote Th17-dependent inflammation. *Sci Transl Med* **9**, eaaf9655 (2017).
44. A. L. Kau *et al.*, Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Sci Transl Med* **7**, 276ra224 (2015).
45. Y. Gavrieli, Y. Sherman, S. A. Ben-Sasson, Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J Cell Biol* **119**, 493-501 (1992).
46. S. Naserian *et al.*, Simple, Reproducible, and Efficient Clinical Grading System for Murine Models of Acute Graft-versus-Host Disease. *Front Immunol* **9**, 10 (2018).
47. N. Segata *et al.*, Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60 (2011).
48. Y. Vardi, Z. Ying, C.-H. Zhang, Two-Sample Tests for Growth Curves under Dependent Right Censoring. *Biometrika* **88**, 949-960 (2001).
49. B. J. Webb *et al.*, Prediction of Bloodstream Infection Due to Vancomycin-Resistant Enterococcus in Patients Undergoing Leukemia Induction or Hematopoietic Stem-Cell Transplantation. *Clin Infect Dis* **64**, 1753-1759 (2017).
50. A. C. Burman *et al.*, IFNgamma differentially controls the development of idiopathic pneumonia syndrome and GVHD of the gastrointestinal tract. *Blood* **110**, 1064-1072 (2007).
51. F. P. Nestel, K. S. Price, T. A. Seemayer, W. S. Lapp, Macrophage priming and lipopolysaccharide-triggered release of tumor necrosis factor alpha during graft-versus-host disease. *J Exp Med* **175**, 405-413 (1992).
52. P. Garside, S. Reid, M. Steel, A. M. Mowat, Differential cytokine production associated with distinct phases of murine graft-versus-host reaction. *Immunology* **82**, 211-214 (1994).
53. A. M. Mowat, Antibodies to IFN-gamma prevent immunologically mediated intestinal damage in murine graft-versus-host reaction. *Immunology* **68**, 18-23 (1989).