

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Meal-tray records were captured in the hospital kitchen commercial computer system (Computriton, Bedford, MA), which is linked to the recipe of each item; no version is published for this system. No custom software was used for data collection.
Data analysis	Analyses were performed in R (v4.4.1). Human 16S and shotgun sequencing: DADA2 v1.16.0; BLAST+ v2.11.0 against the NCBI 16S ribosomal RNA reference database (Bacteria and Archaea, downloaded 2020); QIIME 2 v2021.11; vegan v2.5-7; MetaPhlAn v4.0.5 (database mpa_vJan21_CHOCOPhAnSGB_202103), run through the vdblab-shotgun pipeline (https://github.com/vdblab/vdblab-shotgun/) with clumpify and bbduk from the BBTools package. Mouse 16S sequencing: DADA2 and UCLUST as implemented in QIIME v1.9.1 within the ZymoBIOMICS Targeted Sequencing Service; QIIME v1.9.1 for alpha diversity; vegan v2.5-7 for principal coordinates analysis; MaAsLin2 v1.24.1 for differential abundance. Dietary data: food tree built with the Food_Tree pipeline (https://github.com/knights-lab/Food_Tree); Faith's phylogenetic diversity computed in QIIME 2 v2021.11; TaxUMAP (https://github.com/jsevo/taxumap). Enterococcus amplicon specificity: the SYTYCS pipeline (https://github.com/nickp60/sytycs), using EMBOS PrimerSearch, seqkit and vsearch. Statistical modeling: brms v2.16.3 with rstan v2.26.4 for the Bayesian multilevel models; compositions v2.0-6 for the centered-log-ratio transformation; geepack v1.3.9 (dietary and microbiome trend analyses) and v1.3.11 (clinical-outcome analyses) for generalized estimating equations.

RNA-seq: Illumina DRAGEN Bio-IT Platform v3.10; featureCounts v2.1.1; DESeq2 v1.44.0; ashR v2.2-63 for log2 fold-change shrinkage; fgsea v1.30.0 with the MSigDB Hallmark collection v2025.1.Mm via msigdbR v25.1.1; org.Mm.eg.db v3.19.1.

Flow cytometry: FlowJo v10.10.0 (Tree Star).

Visualization: ggplot2 v3.3.5, tidybayes v3.0.2, ggpubr v0.4.0.

Code used for the analysis and figure generation is publicly available at https://github.com/Anqi-Dai/Nutrition_microbiome_reproducibility and is archived at Zenodo (<https://doi.org/10.5281/zenodo.21290618>). R package versions are pinned with renv, and the external tools required to run the pipeline end to end are specified in the repository README. The tables distributed with the repository are the same files archived in the Zenodo data deposit⁷⁸. The code is released under an MIT license.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequencing data analyzed in this study are publicly available through the NCBI Sequence Read Archive (SRA) across several BioProjects, as follows. Accession identifiers for individual samples, together with the sample identifiers used throughout this study, are listed in Data S1. Human fecal 16S rRNA amplicon sequencing, 1,009 samples: PRJNA878528 (476 samples), PRJNA545312 (484 samples²⁷), PRJNA810900 (38 samples), PRJNA972747 (5 samples), PRJNA979931 (5 samples) and PRJNA810891 (1 sample). Human fecal shotgun metagenomic sequencing, 333 samples: PRJNA878528, SRR35981832–SRR35982164. Mouse fecal 16S rRNA amplicon sequencing, 111 samples: PRJNA878528, SRR31624460–SRR31624570. Mouse cecal shallow shotgun metagenomic sequencing, 24 samples: PRJNA878528, SRR37009193–SRR37009216. Mouse intestinal epithelial RNA sequencing, 20 libraries: PRJNA878528, large intestine SRR36928799–SRR36928808 and small intestine SRR39866831–SRR39866840.

Source data are provided for animal experiments (Fig. 4c and Extended Data Figs. 8 and 9) as a spreadsheet workbook file herein and are also archived in a public Zenodo repository (<https://doi.org/10.5281/zenodo.14538105>)⁷⁸ as Data S7, along with de-identified tables of patient data: meal-level dietary intake records (Data S2), stool-sample-level microbiome α -diversity, dietary exposures and clinical covariates (Data S3), medication-exposure records (Data S4), nutrient profiles of the foods consumed from FNDSS 2015–2016 (Data S5), patient-level longitudinal summaries of caloric intake and dietary and fecal microbiota α -diversity (Data S6), clinical metadata, and processed microbiome feature tables (ASV counts and relative abundances, taxonomic annotation, genus-level relative abundances, and shotgun-derived Enterococcus species abundances), along with derived analysis inputs and food-taxonomy files. Analysis code for these panels is included in the public repository.

This study also used the following public resources: the USDA Food and Nutrient Database for Dietary Studies (FNDSS) 2015–2016 (<https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/fndss-download-databases/>); the USDA Food Patterns Equivalents Database (FPED) 2015–2016; the NCBI 16S ribosomal RNA database of Bacteria and Archaea type strains (RefSeq Targeted Loci, BioProjects PRJNA33175 and PRJNA33317; at <https://ftp.ncbi.nlm.nih.gov/blast/db/>), downloaded in 2020 and used to assign taxonomy to the human amplicon sequence variants with blastn (BLAST+ v2.11.0); and the Zymo Research Database to assign taxonomy to the mouse 16S data with Uclust in QIIME v1.9.1; MetaPhlAn v4.0.5 with database mpa_vJan21_CHOCOPhlanSGB_202103; the GENCODE mouse basic gene annotation release vM37 (GRCm39; https://www.gencodegenes.org/mouse/release_M37.html), used for gene-level quantification of the mouse RNA sequencing; and the MSigDB Hallmark gene set collection (v2025.1.Mm, 50 gene sets), accessed through the msigdbR R package v25.1.1.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Supplementary Table 1
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	Supplementary Table 1
Recruitment	In this cohort study conducted at Memorial Sloan Kettering Cancer Center in New York between 2017 and 2022, eligible participants had been admitted for an allogeneic hematopoietic cell transplantation. Participants consented in writing to collection and analysis of biospecimens; specimens were analyzed for this study with the approval of the Memorial Sloan Kettering Cancer Center Institutional Review Board under protocol #16-834, which has been previously published (PMID: 32101664) but is not listed in a clinical trial registry. The study was conducted in accordance with the Declaration of Helsinki. Thrice weekly, patients who had been recently admitted for allo-HCT were invited to participate in dietary-intake collection.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For human studies, eligible patients were invited to participate over a 5-year period (2017-2022) during which staffing and resources were available; the cohort size was the number of patients who agreed to participate during that period; a formal power calculation was not done. For mouse studies, the minimum number of animals for statistical significance were used based on preliminary data; no statistical method was used to formally predetermine sample size.
Data exclusions	Detailed in Extended Fig. E1a
Replication	Experiments were repeated at least twice as successful, independent experiments. Details as to the exact number of replicates are included in the Figure Legends.
Randomization	For antibiotic and diet experiments under conventional (specific-pathogen-free) conditions, experiments were conducted in singly housed mice. Batches of vendor-supplied mice were separated into individual cages and assigned randomly to treatment arms. For bone marrow transplants, mice were segregated into treatment groups by cage, 4-5 mice per cage, randomly, except with care that mice at the extremes of body mass were not allocated to the same treatment cages. For gnotobiotic experiments, germ-free mice were randomized to treatment arm and co-housed by treatment assignment in gnotobiotic isolator cages.
Blinding	Blinding was not done in mouse experiments as the outcomes were objective quantitative measures (e.g., abundance of Enterococcal colony forming units, next-generation sequencing data). Additionally, treatment-arm allocations, treatments, and processing of samples for a given experiment were often handled by the same individual, so blinding was not logistically practical.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	C57BL/6 mice, aged 6–8 weeks. Animals were housed under conditions including 6:00 am to 6:00 pm light/ dark cycle, at a temperature of 72 degrees Fahrenheit with a set point +/- 2 degrees Fahrenheit and relative humidity of 30-70%.
Wild animals	No wild animals were used in the study.
Reporting on sex	Females
Field-collected samples	No field-collected samples were used in the study.

Ethics oversight

Memorial Sloan Kettering Institutional Animal Care and Use Committee (protocols #23-03-006 and #99-07-025).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

NA

Study protocol

In this cohort study of adult allo-HCT recipients at Memorial Sloan Kettering Cancer Center in New York between 2017 and 2022, eligible participants had been admitted for an allogeneic hematopoietic cell transplantation. Participants consented in writing to collection and analysis of biospecimens; specimens were analyzed for this study with the approval of the Memorial Sloan Kettering Cancer Center Institutional Review Board under protocol #16-834, which has been previously published (PMID: 32101664) but is not listed in a clinical trial registry. The study was conducted in accordance with the Declaration of Helsinki.

Data collection

New York, 2017-2022

Outcomes

NA

Plants

Seed stocks

NA

Novel plant genotypes

NA

Authentication

NA