nature research

Corresponding author(s): Ami S. Bhatt

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ An indication of 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ 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 statistics 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. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ 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 *d*, Pearson's *r*), indicating how they were calculated
☒	☐ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect the data in this study.
Data analysis	Open source software tools bcl2fastq v2.20.0.422, SuperDeduper v1.4, TrimGalore v0.4.4, CutAdapt v1.16, SPAdes/metaSPAdes v3.11.0, GAEMR v1.0.1, CheckM v1.0.11, Burrows-Wheeler Aligner v0.7.10, SAMtools v1.7, BamTools v2.4.0, bedtools v2.26.0, MUSCLE v3.8.31, FastTree v2.1.7, BLAST v2.2.31, SRST2 v0.2.0, Prodigal v2.6.3, hmmer v1.2, and iRep v1.10 were used in this analysis. The One Codex web-based tool was used with the v2017 database (https://onecodex.com). R packages ape v5.1, phangorn v2.4.0, ggtree v1.10.5, ggplot2 v2.2.1, reshape2 v1.4.3, and dplyr v0.7.4 were used. Custom R and Python code for this manuscript is available at https://github.com/bhattlab/strainsifter.

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All whole-genome sequencing data have been deposited in the National Center for Biotechnology Information Sequence Read Archive under BioProject PRJNA477326.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	No sample size calculation was performed as this was a retrospective study.
Data exclusions	None.
Replication	The study was a descriptive retrospective study of a cohort from a single institution. As very few institutions have collected longitudinal stool samples on subjects and paired, stored bloodstream isolates, it was not feasible to replicate the results of this study without a prospective collection.
Randomization	Not applicable, this was not an interventional study.
Blinding	Not applicable, this was not an interventional study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	This study used bloodstream isolates (bacteria) and stool samples collected from human subjects under an Institutional Review Board-approved protocol at Stanford University Hospital. Limited stool samples are available and bloodstream isolates may be available through the clinical microbiology laboratory at Stanford. Investigators interested in working with these unique biological materials may contact the study investigators and arrange for a material transfer agreement, pending a successful amendment of the Institutional Review Board approved protocol.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Patients in this cohort were adults equal to or greater than 18 years of age with a history of a hematologic disorder who had undergone autologous or allogeneic hematopoietic stem cell transplantation for which they were hospitalized at Stanford University Hospital between October 2015 through June 2017.
Recruitment	This was a retrospective study of patient samples and did not require active patient recruitment. Patients had previously consented to participate in an Institutional Review Board-approved tissue and biospecimen banking protocol at Stanford University. All patients undergoing hematopoietic cell transplantation are invited to participate in this protocol upon admission to care at Stanford University Hospital.