nature research

Corresponding author(s): Erik Thiele Orberg, Elisabeth Meedt

Last updated by author(s): Apr 10, 2022

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 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 (*n*) 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 *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 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. *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

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	Microsoft -Office-365 Excel
Data analysis	IBM SPSS 25, phyloseq package version 1.36, pairwiseAdonis (v0.4) package, microbiomeMarker package (v1.0.1.), ggplot2 (v3.3.5.), ggpubr (v0.4), fastp (v0.20.1), dedupe.sh, bbmap suite (v38.76), metaSPAdes (v3.14.0), VirSorter (v1.0.6), CAT (v5.0.4), DeepVirFinder (v1.0), R (v3.2., package vegan), R (v4.0.2, stats package), Prism 9- GraphPad, Origin 2020b, Microsoft Excel, R (v 3.3.3, ggplot2 package), Analyst 1.7 software (Sciex, Darmstadt, Germany), MultiQuant 3.0.3 (Sciex, Darmstadt, Germany), Metaboanalyst 5.0

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 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

The data sets generated and analyzed during the current study are available in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home).

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

Life sciences study design

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

Sample size	Patients analyzed in this study were hospitalized for allogeneic hematopoietic stem cell transplantion (allo-SCT) at the University Hospital rechts der Isar and the University Hospital Regensburg. Inclusion criteria were adult patients (18 years old or older) who underwent allo-SCT and had collection of at least 3 stool samples collected between hospital admission before and up to 5 weeks after allo-SCT. Our objectives were to include at least 70 patients. Between 2018 and 2021, the 78 patients who fulfilled inclusion criteria were prospectivelly included in the study without any selection. For this observational study, no sample size calculation was necessary.
Data exclusions	No data were excluded from the analysis.
Replication	A unique set of patient samples were analyzed without replication.
Randomization	Patients were not randomized in different experimental groups. The clinical course (development of graft-versus-host disease, antibiotic therapy, high and low intestinal bacterial alpha diversity, high and low intestinal metabolite expression) determined the allocation to different groups.
Blinding	Investigators, technicians and analysts where blinded in regards to allocations to the different groups at the time of sample processing, next-generation amplicon and metagenomic sequencing and metabolomics mass-spectrometry.

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
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used	See supplemental table 6.
Validation	See supplemental table 6.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Patient characteristics (provided in Figure 1B) include age, gender, diagnosis and survival were similarly distributed amongst the cohorts at both transplantation centers (Munich and Regensburg) and represent the typical patient cohort undergoing allogeneic stem cell transplantation.
Recruitment	At both centers, Munich and Regensburg, patients undergoing allogeneic stem-cell transplantation were enrolled (Munich: 2019-2021, Regensburg 2018-2021) in this prospective, observational study after informed consent and in accordance with IRB-approved study protocols. Stool samples were collected at pre-detemined time-points (calendar-driven) or in response to clinical occurrences (ecent-driven).

Ethics oversight	Technical University of Munich, IRB 295/18 S; University of Regensburg, IRB 02/220, 14-101-0047, 17-619-101

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	As this trial was not interventional but observational, it has not been registered.
Study protocol	There is no study protocol available. Biosamples and data were collected as described below.
Data collection	We prospectively collected stool samples at calendar-driven and event-driven timepoints for characterization of the intestinal bacterial, fungal and virome structure. In parallel, each sample was assayed by mass-spectrometry for expression of microbiota-derived metabolites including short chain fatty acids, indole derivatives or bile acids. Upon this extensive data set we employed multi-omics factor analysis (MOFA). Biosamples were referenced with extensive clinical metadata, including age, gender, diagnosis, allo-SCT donor type, allo-SCT conditioning regimen, allo-SCT stem cell source, presence or absence of gastrointestinal GvHD, anti-infective medication and laboratory findings. In patients with GVHD, the location, grade, and treatment were recorded. In patients treated with anti-infective medication, antibiotics, antifungals or antivirals were recorded. If relapse or death occurred, the dates and cause of death were also recorded.
Outcomes	Overall survival (OS) and cumulative incidences of transplant-related mortality (TRM) and GvHD were estimated and presented as hazard ratios with 95% confidence interval.