nature portfolio

Corresponding author(s): Erik Thiele Orberg, MD, PhD
Hendrik Poeck, MD

Last updated by author(s): 08/24/2023

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 (*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	Microbial sequencing data, metabolite expression profiles were collected from analysis platforms using Microsoft -Office-365 Excel (16.7.5) Clinical metadata was collected from patient electronic health records using Microsoft -Office-365 Excel (16.7.5)
Data analysis	Analyst 1.7 software (Sciex), bbmap suite (v38.76), bowtie2 (2.5.0), CAT (v5.0.4), CheckV, corrplot (0.91), CoverM (v0.6.1), cutadapt (3.4), dada2 (1.22), DECIPHER (2.26), dedupe.sh, DeepVirFinder (v1.0), fastp (v0.20.1), FastTree (2.1.1), ggplot2 (3.4.0), ggpubr (0.4.0), GraphPad Prism (9), HUMAnN (3.6), IBM SPSS 25, Maaslin2 (1.7.3), MAFFT (v7.475), Metaboanalyst (5.0), MetaPhlAn (4.0.3), meta SPAdes (v3.14.0), microbiomeMarker package (v1.0.1.), Microsoft Excel (16.7.5), minimap2, MMseqs2, MOFA2 (1.4.0), mofapy2 (0.6.4), MultiQuant 3.0.3 (Sciex, Darmstadt, Germany), pairwiseAdonis (v0.4), PHAMB, PHROGs, phyloseq (1.36), Prodigal (v2.6.3), Psych R (v2.1.9), pyCircus, Python (3.6.13), R (4.1.1), R (gggenomes package), R (ggtree package), R (vegan package), rstatix (0.7.0), tBLASTx, trimAl (v1.4.rev15), Trimmomatic (0.39), vConTACT2, ViroProfiler (v0.1), VirSorter2, vsearch (2.22.1)

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

Microbial sequencing (bacterial, and fungal amplicon data and viral metagenomic sequencing data) that support the findings of this study have been deposited at the European Nucleotide Archive under accession number PRJEB53547 (https://www.ebi.ac.uk/ena/browser/view/PRJEB53547). Mass spectrometry data have been deposited at Zenodo under accession number 6603017 (https://zenodo.org/record/6603017). Both repositories are annotated with pseudonymized clinical metadata.
Code, scripts and packages used in the present study can be found at Github (https://github.com/guardianre/MOFA-in-allo-SCT.git).

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	Findings reported in our study apply to both sexes. The study cohort included female and male patients and sex was assigned based on electronic patient records. Detailed numbers of patients in each sex is reported in Supplementary Table 1, male and female patients were equally distributed across both centers. We did not consider a priori sex- or gender-based analysis as they did not affect microbiome-predicted outcomes in published cohorts of allogeneic stem-cell transplantation patients (Taur et al., Blood 2014; Peled et al., NEJM 2020). We included sex as a covariate when comparing IMM-RI low- vs high-risk patients. However, we did not detect statistically significant differences between groups.
Reporting on race, ethnicity, or other socially relevant groupings	The study cohort included patients enrolled at the University Hospital Rechts der Isar in Munich and the University Hospital in Regensburg. Both centers are located in Bavaria, Germany and treat a populace that is largely of Western and Central European descent. In Munich, epidemiological data regarding place of birth of the patients and their parents was recorded, but race and ethnicity was not recorded a priori. In Regensburg, all patients identified as "Non-hispanic/white" based on self-reporting.
Population characteristics	Patient characteristics including age, sex, diagnosis, donor type, conditioning regimen, stem cell source, incidence of acute GI-GvHD, incidence of TRM, incidence of relapse, and 2-year overall survival are reported in Supplementary Table 1. If relapse or death occurred, the dates and cause of death were recorded. Patient characteristics according to IMM-RI are reported in the Supporting Information. Diet in Munich patients is reported in the Supporting Information. In patients with GI-GVHD, the onset, grade, treatment and outcome were recorded. In patients treated with anti-infective medication, antibiotics, antifungals or antivirals were recorded. We recorded type of antibiotics administered as per study-specific standards and the initiation of antibiotic treatment, this is reported in the Supporting Information.
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 cohort after obtaining informed consent and in accordance with IRB-approved study protocols (see "Ethics oversight"). All patients who were admitted for allo-SCT were screened for study enrollment. Stool samples were collected at predetermined time-points (calendar-driven, weekly from Day -7 until Day +35 or discharge, whichever occurred later) or in response to clinical occurrences (event-driven, e.g., onset of acute GI-GvHD). Patient were included in the analysis if a minimum of 3 longitudinal samples were available. We are not aware of a selection bias, since all patients were routinely screened for study participation, and the large majority of patients consented to study participation. There is a potential bias for inclusion of patients with a longer hospital stay in the analysis, since they are more likely to have met the minimum number of samples required.
Ethics oversight	Technical University of Munich, IRB 295/18 S; University of Regensburg, IRB 14-47_1-101, 14-101-0047, 21-2521-01.

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

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 allo-SCT at the University Hospital rechts der Isar in Munich and the University Hospital Regensburg. Inclusion criteria were adult patients (18 years old or older) who received allo-SCT and had a collection of at least 3 longitudinal

	stool samples collected between hospital admission 1 weeks before and up to 5 weeks after allo-SCT. Our prospective objective was to include at least 70 patients based on the size of other published cohorts. Between 2018 and 2021, the 78 patients who fulfilled inclusion criteria were prospectively included in the study without any selection. A total of 356 stool samples were collected (see Figure 1). For this observational cohort, no a priori sample size calculation was performed. However, the size of our cohort (n=78) is similar to those reported in previous studies (Taur et al., Blood 2014).
Data exclusions	Exposure to broad-spectrum antibiotics was an a priori data exclusion criteria. Initiation of antibiotic therapy was recorded for all patient samples. Once a patient was started on antibiotics the first and all subsequent samples were excluded from the "No ABX" subset and classified as “ABX” for all subsequent analysis. Samples obtained after Day +100 after allo-SCT (n=2) were excluded from the analysis, since the microbiome at this late time-point is no longer comparable to that of the early post-transplant period.
Replication	We performed a cross-cohort analysis by comparing microbiome, microbiota-derived metabolites and clinical outcome within two German centers. The study population was well-balanced and there were no statistically significant differences in regard to age, gender, underlying disease, donor type, intensity of conditioning and incidence of GI-GvHD across both centers. We observed a difference between both centers in regard to survival (Munich 52%, Regensburg 72%), which are in the range of allo-SCT reported outcomes and may be explained in part by stem cell source. There were no differences in regard to TRM or relapse between centers. The association of bacterial microbiome diversity and clinical outcome confirmed those of published cohorts (Taur et al., Blood 2014; Peled et al., NEJM 2020).
Randomization	Patients were not randomized in different experimental groups, nor was there an active intervention. 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 subgroups.
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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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