nature research

Corresponding author(s): Tobias M. Hohl

Last updated by author(s): 11-04-2019

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.

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	No software was used to collect the data in this study.
Data analysis	RStudio Version 1.0.143 and R Version 3.5.3 was used with packages phyloseq, dada2, ggplot, dplyr, reshape2, vegan, parcomp, ggbiplot. Whole genome sequencing data of fungal isolates were analyzed with GATK packages (version 4.0.4.0) HaplotypeCaller, MarkDuplicates, CombineGVCFs, GenotypeGVCFs, VariantsToTable. Additional open source tools used: BWA-mem version 0.7.12., RRHS tool (http://www.cmpg.iee.unibe.ch/content/softwares__services/computer_programs/rrhs/), RAxML (raxmlHPC-PTHREADS) version 8.2.11, BEDTools coverage tool version 2.27.1, RepeatMasker 4.0, TandemRepeatFinder, BLAST v.2.2.31. LEfSe analysis was performed remotely on the Huttenhower lab Galaxy server: http://huttenhower.sph.harvard.edu/galaxy

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. 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

Sequencing data are available from Bioproject with the accession number PRJNA579121, the accession numbers of each entry are listed in Supplementary Table 4.

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	No sample size calculation was performed as this was a retrospective study.
Data exclusions	Patients with candidemia but without at least one banked fecal specimen collected within 10 days preceding BSI were excluded. Samples that failed ITS sequencing were excluded from all mycobiome analyses and further study. Samples that failed 16S sequencing or 16S qPCR reaction were excluded from 16S sequencing and quantitative analyses
Replication	Non-experimental study design in humans. Replication not feasible in this setting.
Randomization	case-control study, no random allocation possible
Blinding	Data and sample collection occured before the conception of this study and was therefore blinded. Sequencing and inference of ASVs by using DADA2 was performed blinded to the group allocation.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	The patients in this study were adult patients (equal to or greater than 18 years of age) undergoing allogeneic hematopoietic stem cell transplantation and receiving care at Memorial Hospital between Jan. 2014 through Dec. 2017.
Recruitment	This was a retrospective study with patient biospecimens. All patients undergoing allogeneic hematopoietic stem cell transplantation at MSKCC are eligible to participate. Patients were informed of the fecal collection study and consent was obtained prior to initiation of hematopoietic stem cell transplantation.
Ethics oversight	MSKCC IRB

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	not a clinical trial
Study protocol	not a clinical trial

Data collection	This is a nested case-control study within a cohort of patients with prospectively collected biospecimens. All patients undergoing allogeneic hematopoietic stem cell transplantation at MSKCC are eligible to participate in this cohort after having provided informed consent. Patients reported in this manuscript have been recruited between January 1, 2014 and December 31, 2017. Clinical data was obtained by chart review.
Outcomes	The outcome (case status) was defined as a positive blood culture with a pathogenic Candida species (Candida albicans, Candida dubliniensis, Candida tropicalis, Candida parapsilosis, Candida orthopsilosis, Candida metapsilosis, Candida famata (Debaryomyces hansenii), Candida lusitaniae (Clavispora lusitaniae), Candida guilliermondii (Meyerozyma guilliermondii), Candida krusei (Pichia kudriavzevii), Candida glabrata, Candida kefyr (Kluyveromyces marxianus), Candida norvegensis (Pichia norvegensis), Candida inconspicua (Pichia cactophila), Candida lipolytica (Pichia lipolytica), Candida fabianii (Cyberlindnera fabianii), and Candida auris.)