

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 No software was used for data collection.

Data analysis Metagenomic sequence data were analyzed using Metaphlan2 (v2.7), Humann2 (v2.8.1). RNA-Seq data were analyzed using CIBERSORT. were Luminex assay results were read and quantified using a BioPlex multiplex plate reader and Microplate Manager software (v5). Statistical analysis was conducted using R v4.0.3 and R packages vegan (v2.5), DESeq (v1.30), lme4 (v1.1-27). Custom code to run analyses and generate figures is available at github.com/cworby/UMB-study

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](#)

Metagenomic sequence data are available from the Sequence Read Archive under Bioproject PRJNA400628. Questionnaire data, outout files from Metaphlan2, Humann2 and StrainGE are available from github.com/cworby/UMB-study. PBMC RNASeq data are PBMC RNASeq data are available from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002728.v1.p1. Human genome hg19 used in PBMC RNAseq analysis available under Bioproject PRJNA31257.

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	Sample size calculation for longitudinal microbiome analyses is not straightforward, we note that each high-risk patient represents their own controlled experiment. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous longitudinal microbiome studies which were able to detect significant effects, e.g. Dethlefsen & Relman PNAS 2011. 108 Suppl 1: p.4554-61; Turnbaugh et al., Nature, 2006. 444(7122): p. 1027-31.
Data exclusions	No data were excluded from our study.
Replication	Our study was an observational cohort study and no replication was performed, although we have described the recruitment process and sampling strategy sufficiently such that the study may be replicated.
Randomization	Our study was an observational study with no intervention and cohorts based on pre-determined criteria; as such, no randomization was required. Control participants were age-matched to rUTI participants, and few dietary differences existed between the cohorts based on survey responses. We adjusted for race in cohort comparisons of microbiome structure.
Blinding	Our study was an observational study with no intervention and cohorts based on pre-determined criteria; as such, blinding was not relevant.

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

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Women from the St. Louis, MO area reporting three or more UTIs in the past 12 months were recruited into the rUTI study arm, while women with no history of UTI (at most one UTI ever) were recruited into the control arm via the Department of Urological Surgery at Barnes-Jewish Hospital in St. Louis, MO. We excluded women who: i) had inflammatory bowel disease (IBD) or urological developmental defects (e.g., ureteral reflux, kidney agenesis, etc.), ii) were pregnant, iii) take antibiotics as prophylaxis for rUTI, and iv) were younger than 18 years or older than 45 at the time of enrollment.
Recruitment	rUTI women were recruited based on clinical history via the Department of Urological Surgery, along with age-matched control women with no history of rUTI. Flyers were posted around Wash U Medical School, Wash U in St. Louis campus, and the Barnes-Jewish Hospital. Participants were remunerated midway through the study, and at the end of the study upon completion, via gift cards. Self selection biases may therefore exist; in particular we did not collect socio-economic data on participants. However, given age matching and the similarity in self-reported dietary habits between cohorts, we do not anticipate any such bias to have a significant impact on the composition of the gut microbiome.
Ethics oversight	This study was conducted with the approval and under the supervision of the Institutional Review Board of Washington University School of Medicine in St. Louis, MO

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