nature research

Corresponding author(s): Curtis Huttenhower

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	Unless otherwise noted, the evaluations and applications from the paper were based on data produced by HUMAnN2 v0.11.0, which incorporates MetaPhlAn2 v2.6.0, bowtie2 v2.2.1, and DIAMOND v0.8.22. This and other versions of HUMAnN2 are available via http://huttenhower.sph.harvard.edu/humann2.
Data analysis	Data were analyzed and visualized using the Python scientific stack (numpy, scipy, and matplotlib). Many of these analysis scripts are bundled with the HUMAnN2 software. Scripts specific to this publication are available from the authors upon request. HUMAnN2 incorporates a previously generated pangenome database (ChocoPhlAn). The software packages used to produce the ChocoPhlAn database (e.g. PhyloPhlAn and UCLUST) are detailed in PMIDs 24203705 and 26418763. ChocoPhlAn was annotated here against UniRef90 and UniRef50 using DIAMOND v0.8.22.

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

The Human Microbiome Project (HMP) metagenomes analyzed in this work are available via http://hmpdacc.org. The integrative Human Microbiome Project (iHMP) metagenomes and metatranscriptomes analyzed in this work are available via http://ibdmdb.org. The Red Sea metagenomes analyzed in this work are available as NCBI BioProject PRJNA289734. The synthetic metagenomes and metatranscriptomes used in the evaluation of HUMAnN2 are available via the HUMAnN2 website: http://huttenhower.sph.harvard.edu/humann2.

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 calculations were performed. This study is not focused on null hypothesis-based statistical testing, and no new samples were collected toward such a goal. This study re-analyzes all meta'omic samples available from three previously published datasets (modulo samples that were excluded for reasons detailed below).
Data exclusions	In analyses of HMP metagenomes, repeated samples from the same individual and body site were excluded.
Replication	Evaluations of software accuracy and performance were repeated on five distinct synthetic meta'omic samples.
Randomization	This study does not involve any treatment groups that would require randomization of (e.g.) enrolled subjects.
Blinding	This study does not involve any treatment groups that would require blinding during (e.g.) assignment of subjects to groups.

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