nature research

Corresponding author(s): NMED-A96571B-Z

Last updated by author(s): Aug 31, 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 is used.
Data analysis	R/Python standard packages with custom scripts. R packages that are used in the study include stats V3.5.0, ppcor V1.1, Rfit, V0.23.0, Hmisc V.4.1-1. Additional software are also used, including Qiagen Ingenuity Pathway Analysis, SAS 9.4.

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

Data will be available at https://portal.hmpdacc.org

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

Life sciences study design

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

Sample size	Our study is done in two levels of cross sectional and longitudinal analysis: For cross sectional analysis (n = 106) we used median value for all time points of all participants and for longitudinal analysis we used longitudinal data on participants with 5 data points and more. For group-wise comparison (IR vs IS), we used median value of all time points of participants in each group. and used statistical analyses to take the sample size into account. For our analyses focusing on individual characters (n = 43), we examined on those participants who had at least five independent samplings that spanned at least 700 days in our study. For both analyses, we rely on significant test to ensure the sample size is sufficient to draw conclusions.
Data exclusions	As frequently used in microbiome studies, for microbiome analyses, to capture major changes across the whole cohort, we focused on prevalent taxa and genes out of all mapped ones, excluding those that were present in less than half of the cohort.
Replication	To control data generation reproducibility, we included multiple aliquots for selected samples anonymously in different batches of all omic assays, and make sure replicates were similar to each other than to any other samples.
Randomization	Samples were randomly arranged by computing program into various batches of omics assay.
Blinding	Each sample was labeled with an numeric ID whose annotation was kept blinded during data collection and analyses.

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	55 participants in this study were female and 51 male, with ages ranging from 25 to 75 yrs old and BMI 25–40 kg/m2. More information were provided in Methods and Supplemental Table.
Recruitment	The recruitment details were documented in Zhou, et al., Nature, 2019. Briefly, participants were recruited through placement of advertisements in local newspapers and radio stations seeking 'prediabetic volunteers' at risk of T2D for a longitudinal multi-omic study. Screening in the CTRU entailed collection of clinical history, physical examination, anthropometric measurements, and fasting blood tests for exclusions. All different ages of male and female participants were recruited.
Ethics oversight	Stanford University IRB has approved the study.

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