nature research

Corresponding author(s): Michael Snyder, Francois Haddad

Last updated by author(s): 2/24/2018

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	R 3.4 (specified in online methods for which analysis which version was used); HugeSeq, TopHat 2.0.11, HTseq 0.6.1, DESEQ2 v3.5; Perseus v. 1.4.2.40; Progenesis QI (Nonlinear Dynamics); ISEC program;
Data analysis	R 3.5; SAS 9.4; IMPaLA v. 11 (Build April 2018); Cytoscape 3.4.0; R packages: corrplot v. 3.3.2; qvalue v. 1.36.0 in R (v. 3.0.1); Hmisc package (v3.15-0) in R (v 3.0.1); igraph (v.0.7.1); R 3.5 function kmeans; SAS procs: Proc Mixed, Proc Univariate Plot, Proc GAM; QIIME2 with DADA2 denoising plugin; MXM v0.9.7; R v 3.4.1 ridge regression; smartpca tool in the PLINK2 suite; Stringdb v. 10.5

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

Raw data included in this study are hosted on the NIH Human Microbiome 2 project site (https://portal.hmpdacc.org/) under the study T2D. Data for participants who have not consented to make their data public was deposited into dbGap under accession phs001719.v1.p1.

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	Sample size for the main HMP project was calculated based on the pilot iPOP study findings of levels of autocorrelation in multiomics data and the number of timepoints needed per participant to reconstruct the time series with < 1% uncertainty. For this study, we used all iPOP participants with available clinical data.
Data exclusions	All available data was used for analyses. There was no data excluded from analyses.
Replication	This was an observational study which did not involve experiments.
Randomization	Participants were not randomized in this study. There was no allocation to groups, since there is no group-based analysis in the study. In some analyses (mainly mixed effects models involving glucose measures and hsCRP) we controlled for age at time of consent and sex since these covariates were considered potential confounders that could obscure the underlying biological relationships between our omics measures and clinical outcome measures.
Blinding	There was no blinding in this study because it is not relevant to the study (There is no allocation to groups or interventions in the study). Omics data was processed without knowledge of participants' clinical status.

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 mean age of iPOP participants at time of enrollment was 53.4 years old (standard deviation 9.2, range 25-75) and 51% were women. The self-reported ethnic distribution of the cohort was 60% European, 11.9% East Asian, 10.1% South Asian, 6.4% Jewish, 5.5% Hispanic, 5.5% African American, 5.5% Mixed/Other. The most common self-reported health conditions at entry were dyslipidemia (34%) and systemic hypertension (27.5%); participants were generally healthy. Over 78% of the cohort was overweight (BMI between 25 and 30, n = 56) or obese (BMI ≥ 30, n = 28). Our cohort which was enriched for DM risk, had a higher than normal family history of DM (55.0%). Family histories of systemic hypertension (54.1%), coronary artery disease (50.5%) and stroke (23.9%) were also common.
Recruitment	Participants were recruited from the Stanford University surrounding community with the goal of enriching the cohort with individuals at risk for diabetes and thus included individuals who expressed interest in other studies related to diabetes. We do not think the biological results presented are affected by the enriched interest in diabetes studies. However our results related to behavioral change may have more limited generalizability due to the unique characteristics of the Stanford University surrounding community and that there may have been a self selection bias of people interested in intensive monitoring of health which may have impacted their response to such monitoring.
Ethics oversight	Stanford University Institutional Review Board (IRB 23602).

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