nature research

Corresponding author(s): Julien Gagneur

Last updated by author(s): May 7, 2020

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	We used 3 datasets. The GEUVADIS one (BAM and VCF files) downloaded from https://www.ebi.ac.uk/Tools/geuvadis-das/ and subset using BCFtools and SAMtools. The GTEx one (BAM files) from https://gtexportal.org/home/datasets. And the Kremer one (count matrices) downloaded from https://www.nature.com/articles/ncomms15824
Data analysis	All the code to analyze the data can be found in: https://github.com/gagneurlab/drop

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

The data can be downloaded from https://www.ebi.ac.uk/Tools/geuvadis-das/, https://gtexportal.org/home/datasets and https://www.nature.com/articles/ncomms15824

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	We used the following datasets: 1. GEUVADIS: RNA-seq data and variants from whole genome sequencing from 100 samples. 2. GTEx: 100 samples from whole blood combined with 100 samples of either muscle, skin not sun exposed, liver and brain cerebellum. 3. Combination of 17 samples from Kremer with a known expression outlier with 102 samples from GTEx not sun exposed. 4. Combination of 13 samples from Kremer with a known splicing outlier with 106 samples from GTEx not sun exposed. 5. All 119 samples from the Kremer dataset
Data exclusions	1. The whole GEUVADIS dataset is comprised of more than 1,000 samples, out of which we subset for 100. 2. Each of the selected GTEx tissues has more than 100 samples. 3. We took only 102 samples from GTEx not sun exposed to complete the sample size of 119 of the original Kremer study. 4. Same as point 3. 5. None.
Replication	1. None 2. None. 3. The skin not sun exposed samples were sampled 30 times. 4. Same as point 3. 5. None.
Randomization	1. None. 2. None. 3. The skin not sun exposed samples were selected randomly without replacement. 4. Same as point 3. 5. None.
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

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