nature research

Corresponding author(s): Prof Matthias Mann

Last updated by author(s): Feb 20, 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	The LC-MS data collection was performed with a Thermo Fisher Easy LC 1200, employing a Pharmafluidics 200 cm µPAC column using a Thermo Fisher Q Exactive HF-X mass spectrometer. Electrospray conditions were monitored by the SprayQC software.
Data analysis	LC-MS data processing was performed using MaxQuant software (version 1.6.1.13), employing the Andromeda search engine. Data analysis was done using Perseus software and custom code written in Python (version 3.6). For the interactive analyses a Neo4j (version 3.5.8 Community edition) was used to develope a graph database which is acces via Python (version3.6) and uses Cypher as query language.

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

LC-MS raw files, FASTA files and MaxQuant output tables are available via the PRIDE partner repository via ProteomeXchange with identifier PXD014877.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	100 Organisms across the tree of life, thereof 19 archaea, 49 bacteria and 32 eukaryotes were analyzed in singlets by MS based proteomics.
Research sample	To collect a diverse set of representative organisms across the tree of life, we considered the availability of assembled genome sequences, the accessibility of cultured or tissue material and included common model organisms for comparison.
Sampling strategy	Cohort size of 100 organisms was chosen by a trade of between biological impact and data acquisition time. As a pioneer study in cross organism proteome comparison this represents a sufficient overview of the proteome of living organisms. Samples were prepared for bottom up proteomics with LC-MS, by denaturationg and lysing (boiling, sonication, grinding), reduction/ alkylation and tryptic digestion of proteins, followed by subsequent purification of peptides by desalting.
Data collection	Data was obtained by bottom up proteomic measurement of tryptic digested proteins with mass spectrometry. The experiments were carried out by the authors.
Timing and spatial scale	Data were collected from September 2018 to January 2019.
Data exclusions	No data were excluded.
Reproducibility	No repetitive measurements were conducted.
Randomization	Sample handling and measuring was done in batch processing. Fractions of the same samples were measured consecutively.
Blinding	Samples were not blinded due to batch processing.

Did the study involve field work? ☐ Yes ☒ No

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	ATCC-CCL-22; ATCC-CCL-34; ATCC-CCL-61; ATCC-CRL-2050; ATCC-CRL-1840; 6C2; ATCC-CCL2; ATCC-CRL-3209; ATCC-CRL-6503; ATCC-CRL-1721; LLC-PK1
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	One cell line was positive for Mycoplasma contamination which was used to generate a proteomics dataset for Mycoplasma.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell line was used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Drosophila melanogaster, strain CantonS with a mean age of 24 h and mixed sex were used. Caenorhabditis elegans N2 wild.type strain were used.
Wild animals	Study did not involve wild animals.
Field-collected samples	Study did not involve field-collected samples.
Ethics oversight	No ethical approvement was required.

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