nature portfolio

Corresponding author(s): Peng Gao, Xuebin Zhang, Xiaotao Shen

Last updated by author(s): Dec 15, 2025

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	ProteoWizard: Version. 3.0.19314-fb982f15b
Data analysis	R:4.5.0; Rstudio: 2024.12.1fs_1.6.6 matrixStats_1.5.0 fit.models_0.64 httr_1.4.7 RColorBrewer_1.1-3 doParallel_1.0.17 tools_4.5.0 doRNG_1.8.6.2 backports_1.5.0 R6_2.6.1 DT_0.33 lazyeval_0.2.2 GetoptLong_1.0.5 withr_3.0.2 prettyunits_1.2.0 gridExtra_2.3 preprocessCore_1.70.0 cli_3.6.5 fastDummies_1.7.5 shinyjs_2.1.0 sass_0.4.10 mvtnorm_1.3-3 robustbase_0.99-4-1 readr_2.1.5 randomForest_4.7-1.2 proxy_0.4-27 pbapply_1.7-2 foreign_0.8-90 stringdist_0.9.15 rrcov_1.7-7 MetaboCoreUtils_1.16.1 parallelly_1.44.0 attempt_0.3.1 itertools_0.1-3 limma_3.64.0 readxl_1.4.5 rstudioapi_0.17.1 impute_1.82.0 shape_1.4.6.1 zip_2.3.3 Matrix_1.7-3 MALDIquant_1.22.3 massdbbuildin_1.0.0 abind_1.4-8 bsicons_0.1.2 lifecycle_1.0.4 yaml_2.3.10 SummarizedExperiment_1.38.1 SparseArray_1.8.0 promises_1.3.3 crayon_1.5.3 PSMatch_1.12.0 miniUI_0.1.2 ComplexUpset_1.3.3 KEGGREST_1.48.0 pillar_1.10.2 knitr_1.50 GenomicRanges_1.60.0 rjson_0.2.23 corpcor_1.6.10 codetools_0.2-20 glue_1.8.0 Rdisop_1.68.0 pcaMethods_2.0.0 data.table_1.17.4 remotes_2.5.0 MultiAssayExperiment_1.34.0 vctrs_0.6.5

png_0.1-8	cellranger_1.1.0	gtable_0.3.6	cachem_1.1.0
xfun_0.52	openxlsx_4.2.8	S4Arrays_1.8.0	mime_0.13
tidygraph_1.3.1	pcaPP_2.0-5	ncdf4_1.24	iterators_1.0.14
statmod_1.5.0	robust_0.7-5	progress_1.2.3	GenomeInfoDb_1.44.0
bslib_0.9.0	affyio_1.78.0	rpart_4.1.24	colorspace_2.1-1
DBI_1.2.3	Hmisc_5.2-3	nnet_7.3-20	shinyjqui_0.4.1
tidyselect_1.2.1	curl_6.2.3	compiler_4.5.0	MassSpecWavelet_1.74.0
rvest_1.0.4	htmlTable_2.4.3	xml2_1.3.8	plotly_4.10.4
DelayedArray_0.34.1	colourpicker_1.3.0	checkmate_2.3.2	scales_1.4.0
DEoptimR_1.1-3-1	affy_1.86.0	stringr_1.5.1	digest_0.6.37
rmarkdown_2.29	XVector_0.48.0	htmltools_0.5.8.1	pkgconfig_2.0.3
base64enc_0.1-3	MatrixGenerics_1.20.0	fastmap_1.2.0	rlang_1.1.6
GlobalOptions_0.1.2	htmlwidgets_1.6.4	UCSC.utils_1.4.0	shiny_1.10.0
farver_2.1.2	jquerylib_0.1.4	jsonlite_2.0.0	MsExperiment_1.10.0
mzID_1.46.0	shinyalert_3.1.0	config_0.3.2	massdatabase_1.0.14
Formula_1.2-5	GenomeInfoDbData_1.2.14	patchwork_1.3.0	viridis_0.6.5
MsCoreUtils_1.20.0	vsn_3.76.0	furrr_0.3.1	stringi_1.8.7
ggraph_2.2.1	plyr_1.8.9	shinyFiles_0.9.3	parallel_4.5.0
listenv_0.9.1	ggrepel_0.9.6	Biostrings_2.76.0	MsFeatures_1.16.0
graphlayouts_1.2.2	hms_1.1.3	Spectra_1.18.0	circlize_0.4.16
igraph_2.1.4	QFeatures_1.18.0	rngtools_1.5.2	reshape2_1.4.4
XML_3.99-0.18	evaluate_1.0.3	golem_0.5.1	BiocManager_1.30.25
tzdb_0.5.0	foreach_1.5.2	missForest_1.5	tweenr_2.0.3
httpuv_1.6.16	purrr_1.0.4	polyclip_1.10-7	future_1.49.0
clue_0.3-66	BiocBaseUtils_1.10.0	ggforce_0.4.2	xtable_1.8-4
AnnotationFilter_1.32.0	e1071_1.7-16	RSpectra_0.16-2	ggcorrplot_0.1.4.1
later_1.4.2	viridisLite_0.4.2	class_7.3-23	rARPACK_0.11-0
tibble_3.2.1	memoise_2.0.1	ellipse_0.5.0	IRanges_2.42.0
writexl_1.5.4	cluster_2.1.8.1	shinyWidgets_0.9.0	globals_0.18.0

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The demo data for TidyMassShiny can be accessed on the tutorial website (https://www.tidymass.org/tidymassshiny-tutorial/). The entire analysis pipeline takes approximately 1 hour to complete on a Mac OS system with 16GB of RAM. For the case study, the MS2 data (mgf format) and processed data ("mass_dataset") from the massProcesser package are available on the tidyMass project website (https://www.tidymass.org/databases/). Public databases ("databaseClass") containing metabolite origin information, generated by metid, are also available on the tidyMass project website (https://www.tidymass.org/databases/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

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	In this study, we only re-analyzed data from a previously published studies (Shen X. et al., Longitudinal urine metabolic profiling and gestational age prediction in human pregnancy. Brief Bioinform. Vol 26 (2024); Fu J. et al., Metabolic disorder and intestinal microflora dysbiosis in chronic inflammatory demyelinating polyradiculoneuropathy. Cell Biosci. Vol 13 (2023); Liang L. et al., Metabolic Dynamics and Prediction of Gestational Age and Time to Delivery in Pregnant Women. Cell. Vol 181 (2020)) and no new LC-MS data were produced. A total of 346 urine samples were collected from 36 ethnically diverse women during pregnancy (gestational age 11.8–40.7 weeks) and postpartum. Sample size was determined based on feasibility and the goal of capturing longitudinal variability in urine composition across pregnancy and postpartum in a diverse cohort. No formal power analysis was conducted, as this was an exploratory study aimed at characterizing urinary profiles rather than testing specific hypotheses. The number of participants (n=36) and samples (n=346) is consistent with similar longitudinal metabolomic studies in pregnancy.
Data exclusions	The information for 'duction', 'child sex', and 'child weight' is missing for 10 participants in our cohort due to incomplete data collection
Replication	Sample collection was conducted longitudinally in two batches to account for logistical constraints, with gestational age dating standardized per American Congress of Obstetricians and Gynecologists guidelines.
Randomization	For all datasets, samples were assigned randomly to acquire LC-MS data.
Blinding	The investigators were blinded to group allocation during data collection. At the time of sample acquisition and processing, scientists were completely unaware of the sample group. The data analyses were blinded.

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*