nature portfolio

Corresponding author(s): Yi Xing

Last updated by author(s): Aug 12, 2022

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	RNA-seq data for PC3E and GS689 cell lines (Example 1) and RNA-seq data for 1,019 CCLE human cancer cell lines (Example 2) can be downloaded from the SRA archive (https://www.ncbi.nlm.nih.gov/sra) under accession BioProject PRJNA438990 and PRJNA523380, respectively. RNA-seq data were downloaded using the sratoolkit software (v2.9.2), as stated in the "Equipment setup" section of the manuscript. Other required reference files can be downloaded using wget (v1.14), as stated in the "Required data" and "Equipment setup" sections of the manuscript.
Data analysis	The rMATS-turbo software was developed and is publicly available on GitHub (https://github.com/Xinglab/rmats-turbo). The data analysis procedures using rMATS-turbo are documented in detail in the "Procedure" section of the manuscript. Custom scripts for downstream data analysis and visualization are provided in the companion GitHub repository of this tutorial (https://github.com/Xinglab/rmats-turbo-tutorial). More specifically, RNA-seq raw data were processed by STAR (v2.7.1a) and rMATS-turbo (v4.1.1). Sashimi plot visualizations of alternative splicing events were generated by the rmats2sashimiplot software. Epithelial or mesenchymal cell states were inferred by using custom scripts of the two-sample Kolmogorov-Smirnov test based on the expression levels of signature genes. Other data analysis and visualization procedures were performed using Python (v2.7) and R (v3.6.1).

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

rMATS-turbo is publicly available on Github (https://github.com/Xinglab/rmats-turbo) and Bioconda (https://anaconda.org/bioconda/rmats). rmats2sashimiplot is publicly available on GitHub (https://github.com/Xinglab/rmats2sashimiplot). Custom scripts for data analysis and visualization on the Examples 1 and 2 datasets are provided in the companion GitHub repository of this tutorial (https://github.com/Xinglab/rmats-turbo-tutorial). rMATS-turbo output files for both Example 1 and Example 2 datasets are available at https://doi.org/10.5281/zenodo.6647023 and other result files are provided in the companion GitHub repository of this tutorial (https://github.com/Xinglab/rmats-turbo-tutorial).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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

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	No sample size calculation was performed. Publicly available RNA-seq data in SRA archive were used (BioProject PRJNA438990 and PRJNA523380).
Data exclusions	No data were excluded.
Replication	Each cell line in Example 1 (PC3E and GS689 cell lines from BioProject PRJNA438990) includes three biological replicates of RNA-seq experiments.
Randomization	This is not relevant to our study, because our study does not involve the assignment of test subjects or treatments.
Blinding	This is not relevant to our study, because our study does not involve the assignment of test subjects or treatments.

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

Methods

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