

Corresponding author(s): Mingyao Li, Nancy R. Zhang

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 <u>Authors & Referees</u> and the <u>Editorial Policy Checklist</u>.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Cor	nfirmed
		The $\underline{\text{exact sample size}}$ (n) for each experimental group/condition, given as a discrete number and unit of measurement
\boxtimes		An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
	\boxtimes	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.
	X	A description of all covariates tested
	\boxtimes	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	\boxtimes	A full description of the statistics including <u>central tendency</u> (e.g. means) or other basic estimates (e.g. regression coefficient) AND <u>variation</u> (e.g. standard deviation) or associated <u>estimates of uncertainty</u> (e.g. confidence intervals)
	\boxtimes	For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
\boxtimes		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
	\boxtimes	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
	\boxtimes	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
\boxtimes		Clearly defined error bars State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on <u>statistics for biologists</u> may be useful.

Software and code

Data analysis

Policy information about availability of computer code

CIBERSORT[https://cibersort.stanford.edu/]: an analytical tool to provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data;

BSEQ-sc[https://github.com/shenorrLab/bseqsc]: Bioinformatics analysis pipeline that leverages single-cell sequencing data to estimate cell type proportion.

MuSiC[https://github.com/xuranw/MuSiC]: deconvolution method that utilizes cross-subject scRNA-seq to estimate cell type proportions in bulk RNA-seq data.

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 upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

	Dala
--	------

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

his study was a re-analysis of existing data, which is openly available at locations cited in the Reference section. The detailed data informations are

- 1. E-MTAB-5061[https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5061/]
- 2. E-MTAB-5060[https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-5060/]
- 3. GSE81608[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81608]
- 4. GSE81433[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81433]
- 5. GSE50244[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50244]
- 6. GSE107585[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107585]
- 7. GSE81492[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81492]
- 8. GSE56743[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56743]
- 9. GSE65267[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65267]
- 10. GSE79443[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79443]

The source data underlying Figures 2a-d, 3b-e and Supplementary Figure 1c, Supplementary Figures 2-8 are provided as a Source Data file. All relevant data is available upon request.

Field-specific reporting

1131313 3 3 3 3 1 1 1 1 3								
Please select 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 t	he document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf							
Life scier	nces study design							
All studies must dis	close on these points even when the disclosure is negative.							
Sample size	NA							
Data exclusions	NA							
Replication	NA							
Randomization	NA							
Blinding	NA							

Reporting for specific materials, systems and methods

Ma	terials & experimental systems	Me	Methods	
n/a	Involved in the study	n/a	Involved in the study	
\times	Unique biological materials	\boxtimes	ChIP-seq	
\boxtimes	Antibodies	\boxtimes	Flow cytometry	
\boxtimes	Eukaryotic cell lines	\boxtimes	MRI-based neuroimaging	
\boxtimes	Palaeontology			
\times	Animals and other organisms			
\boxtimes	Human research participants			