nature portfolio

https://doi.org/10.1038/s41586-025-09312-4

Supplementary information

The Biodiversity Cell Atlas: mapping the tree of life at cellular resolution

In the format provided by the authors and unedited

Table S1. BCA single-cell dataset reporting checklist and recommended quality controls.

Specimen information	
Species	Scientific name
TaxID	Using NCBI taxonomy
Collection details	Geolocation and date
Sample type	Details on the tissues and life stages collected
Collection method and storage	Details on collection procedure and sample storage conditions (buffers employed, temperature, total time until single-cell sampling, etc.).
Sample processing information	
Cells/nuclei	Indicate if cells or nuclei were isolated.
Fixation/preservation	Indicate if the sample was preserved pre and/or post cell extraction and provide details on the method employed (fixative used, fixation time, cryopreservation conditions, etc.)
Dissociation method	Describe in detail the cell dissociation procedure, e.g. enzymes used (time and concentration) and physical dissociation methods (e.g. Dounce homogenizer, sonication or automatic dissociators); or the nuclei extraction protocol (e.g. hypotonic lysis or grinding of frozen tissues).
Purification/enrichment	Describe procedures employed to clean extracted cells/nuclei, possibly including filtering steps, FACS gating strategies, centrifugations or density gradients.
Cell viability	If working with live cells, describe how viability was estimated and whether and how dead cells were removed from the sample.
Sample preparation quality controls	
QC – RNA quality	Provide RNA size distribution profiles (e.g. using Tapestation or Bioanalyzer). Samples with highly degraded RNA should be discarded.
QC – RNAses	Measure and report the presence of RNAses in the sample, which can degrade RNA at subsequent steps in the scRNA-seq procedure.
QC – Cell/nuclei imaging	Provide images of the dissociated cells or extracted nuclei, possibly using basic nuclear/cytosolic staining.
QC – Ambient RNA	Centrifuge extracted cells/nuclei and measure the RNA concentration in the supernatant. Samples with abundant ambient RNA should be discarded or further purified (e.g. using FACS-sorting).
scRNA-seq experiment information	
Single-cell technology	For commercial methods, provide details on kit version and instrument used. For custom methods, provide direct access the protocol and/or indicate any variations in the protocol.
Bias	Indicate if the method method used is 5’ biased, 3’ biased, or full-length.
Target number of cells	Indicate the number of cells loaded in the experiment and the final target.
Sequencing method	Indicate sequencing platform and sequencing layout employed.
Number of reads	Report number of reads obtained.
scRNA-seq data processing information	
Reference genome	Indicate genome version and include stable link to genome data (or provide genome FASTA file).

Genome annotation	Indicate version and include stable link. Also indicate if any modification to the original annotation has been applied (e.g. 3' UTR extension) and provide the file used in the analysis.
Quantified features	Indicate whether full gene intervals or only exons were used and also if any particular features were filtered (e.g. non-coding genes).
Demultiplexing	Indicate software used to map and demultiplex reads (i.e. assign them to cells) and to quantify expression, including details on parameters employed.
Cell hashing strategy	If cell hashing or SNP demultiplexing was used, provide details to reproduce the demultiplexing procedure (list of barcodes, algorithm used, etc.).
Clustering	Indicate clustering algorithm and parameters used. Indicate the procedure if selective subclustering is performed.
Batch correction and integration	If performed, indicate the method and parameters used.
scRNA-seq data quality controls	
QC – mapping and saturation	Report read mapping percentage and sequencing saturation estimates, including how saturation was calculated.
QC – Cell/non-cell filtering	Specify method applied to call cells from non-cells and show global transcripts per barcode distributions, highlighting cells.
QC – Ambient RNA	Report % UMIs in background cell barcodes, and algorithm (if any) used to remove ambient RNA.
QC – Theoretical doublet rate	Indicate estimated doublet rates according to the scRNA-seq method used and final number of captured single-cells.
scRNA-seq data release checklist	
Source code	Release any newly developed code/software that has been used in the processing and downstream analysis of the dataset under open source licenses.
Library structure	Specify the single-cell library structure using a *seqspec*[118] to enable reproducible data re-mapping and demultiplexing. This includes (a) the position in reads of cellular barcodes and UMIs and (b) stable lists of used barcodes and barcoding strategy (e.g. combinatorial barcoding or cell hashing barcodes).
Raw gene expression matrix	Provide gene x cell unspliced, spliced, and ambiguous transcript (UMI) count matrices.
Processed gene expression matrix	Provide normalized gene expression matrix after QC and filtering. If batch correction or data integration is performed, provide batch corrected counts or integrated embeddings.
Gene identifiers	Gene or transcript names should be listed as in Ensembl, with gene short names in metadata. If not on Ensembl, provide other standardized ID and a protein FASTA file. This will facilitate cross-species analyses.
Functional gene annotations	Provide a file with any functional gene annotation generated/used (gene names, gene ontologies, structural domains, etc.).
Protein models	Provide FASTA file with the predicted proteins associated to genes in the UMI count matrix and strictly matching gene IDs.
Cell metadata	Table mapping cell IDs to cluster/cell type/broad cell type annotations.