

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <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>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	CellRanger (v9.0.1) was used to preprocess single-cell sequencing data.
Data analysis	CONCORD (v1.0.8) was used for data integration, dimensionality reduction, simulation, and benchmarking, and is archived at https://pypi.org/project/concord-sc/1.0.8/ ; the latest version is available at https://github.com/Gartner-Lab/Concord . All benchmarking codes used to generate results in this manuscript are deposited at https://github.com/Gartner-Lab/Concord_benchmark . Additional dimensionality reduction analyses were performed using scikit-learn (v1.5.1), PHATE (v1.0.11), and ZIFA (https://github.com/epierson9/ZIFA). Comparative data integration analyses were conducted using scVI (v1.2.2.post2), Scanorama (v1.7.4), Harmony-pytorch (v0.1.8), PyLiger (v0.2.4), Seurat (v5.3.0), and Scanpy (v1.10.1). Additional analyses and visualizations were performed using AnnData (v0.10.6), SciPy (v1.15.2), FAISS (v1.8.0), PyTorch (v2.2.1), NumPy (v1.26.4), scib-metrics (v0.5.2), giotto-tda (v0.5.1), UMAP-learn (v0.5.7), pandas (v2.2.3), seaborn (v0.13.2), gseapy (v1.1.4), plottable (v0.1.5), peakRAM (v1.0.2), and matplotlib (v3.10.1).

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](#)

Single-cell RNA-seq data of *C. elegans* early embryos have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE305031. Public datasets analyzed in this study include: the human lung atlas, compiled by Luecken et al.³³ and obtained from the scIB-metrics website (https://scib-metrics.readthedocs.io/en/stable/notebooks/lung_example.html); GTEx v9, HypoMap, immune cell atlas, mouse pancreatic islet, Tabula Sapiens datasets, sourced from the Open Problems in Single-Cell Analysis website (https://openproblems.bio/benchmarks/batch_integration?version=v2.0.0); the *C. elegans* embryogenesis atlas, downloaded from GEO under accession GSE126954; the joint *C. elegans* and *C. briggsae* dataset, available under GEO accession GSE292756; and the mouse intestinal developmental atlas, acquired from GEO under accession GSE233407.

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

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 if 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.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.

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

A total of 12,899 single cells were recovered from wild-type *C. elegans* N2 embryos. This sample size was chosen to achieve near-complete coverage of early embryonic cell states, as established in prior single-cell studies of *C. elegans* embryogenesis (e.g., Packer et al., Science 2019). No formal statistical sample size calculation was performed, as the dataset encompasses the full cellular diversity of the developing embryo.

Data exclusions

No data were excluded from the analysis.

Replication

Single-cell collections were performed once—comprising a single library preparation and sequencing run—from multiple independent embryos. All analyses, including cell state diversity and lineage relationships, were consistent across biological replicates, leveraging the

invariant lineage structure of *C. elegans*.

Randomization Worms were randomly collected for embryo dissociation. No predefined experimental groups or interventions were applied.

Blinding Not applicable. Data collection and computational analyses were not influenced by group assignment, as no experimental groups or treatment conditions were applied. All analyses were performed using standardized and automated pipelines, making investigator blinding unnecessary.

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 |

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

- | | |
|-------------------------|--|
| Laboratory animals | Wild-type Bristol N2 strain of the nematode <i>Caenorhabditis elegans</i> (hermaphrodite; source: <i>Caenorhabditis</i> Genetics Center, University of Minnesota; developmental stage: early embryos). |
| Wild animals | The study did not involve wild animals. |
| Reporting on sex | The study used hermaphroditic <i>C. elegans</i> ; sex was not a factor in study design or analysis, as the N2 strain is predominantly self-fertilizing hermaphrodites. |
| Field-collected samples | The study did not involve field-collected samples. |
| Ethics oversight | No ethical approval or guidance was required, as the study involved invertebrates (<i>Caenorhabditis elegans</i>), which are not subject to animal welfare regulations under institutional or national guidelines. |

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

Plants

- | | |
|-----------------------|--|
| Seed stocks | <i>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.</i> |
| Novel plant genotypes | <i>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.</i> |
| Authentication | <i>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, mosaicism, off-target gene editing) were examined.</i> |