OXFORD

# Data-driven selection of analysis decisions in single-cell RNA-seq trajectory inference

Xiaoru Dong[1], Jack R. Leary[1], Chuanhao Yang[1], Maigan A. Brusko[2,3], Todd M. Brusko[2,3,4], Rhonda Bacher [ID][1,2,*]

[1]Department of Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, FL 32610, United States
[2]Diabetes Institute, University of Florida, Gainesville, FL 32610, United States
[3]Department of Pathology, Immunology, and Laboratory Medicine, College of Medicine, University of Florida, Gainesville, FL 32610, United States
[4]Department of Pediatrics, College of Medicine, University of Florida, Gainesville, FL 32610, United States

*Corresponding author. Rhonda Bacher, Department of Biostatistics, University of Florida, Gainesville, Florida 32610, United States. E-mail: rbacher@ufl.edu

Biographical note: escort assists users in selecting optimal dataset-specific data processing choices in the context of single-cell trajectory inference.

## Abstract

Single-cell RNA sequencing (scRNA-seq) experiments have become instrumental in developmental and differentiation studies, enabling the profiling of cells at a single or multiple time-points to uncover subtle variations in expression profiles reflecting underlying biological processes. Benchmarking studies have compared many of the computational methods used to reconstruct cellular dynamics; however, researchers still encounter challenges in their analysis due to uncertainty with respect to selecting the most appropriate methods and parameters. Even among universal data processing steps used by trajectory inference methods such as feature selection and dimension reduction, trajectory methods' performances are highly dataset-specific. To address these challenges, we developed Escort, a novel framework for evaluating a dataset's suitability for trajectory inference and quantifying trajectory properties influenced by analysis decisions. Escort evaluates the suitability of trajectory analysis and the combined effects of processing choices using trajectory-specific metrics. Escort navigates single-cell trajectory analysis through these data-driven assessments, reducing uncertainty and much of the decision burden inherent to trajectory inference analyses. Escort is implemented in an accessible R package and R/Shiny application, providing researchers with the necessary tools to make informed decisions during trajectory analysis and enabling new insights into dynamic biological processes at single-cell resolution.

**Keywords**: single cell; RNA-seq; trajectory inference; pseudotime inference

## INTRODUCTION

Analyses to computationally order related cell types along an underlying dynamic are referred to as trajectory inference or pseudotime analysis [1]. These analytical methods provide users with the capacity to visualize and quantify complex developmental or maturation states using shared gene expression and/or phenotypic protein markers at single cell resolution [2, 3]. While there are numerous trajectory inference methods developed for single-cell RNA sequencing (scRNA-seq), obtaining an optimal trajectory remains challenging. A comprehensive review of 45 commonly used trajectory methods, of which there are now more than 100 unique methods at the time of this writing, evaluated

each approach's accuracy of inferred topology, cell ordering, and differential feature expression [4]. No method universally outperformed others, though some methods showed distinction in specific metrics. For instance, Monocle [5] and PAGA [6] best captured the underlying trajectory structure and Slingshot [7] most accurately ordered cells. Yet, while this benchmark evaluation helps narrow down the top methods for particular applications, each method requires additional data processing steps and has its own set of hyperparameters.

Specifically, all trajectory methods require, or at least recommend, feature selection and dimension reduction prior to trajectory estimation. Feature selection involves subsetting the total

number of genes to a smaller set of informative genes, typically the most highly variable genes. However, the decisions of how many highly variable genes and what metric should be used to score gene variability are left to the user. Dimension reduction approaches project the high-dimensional dataset into a low-dimensional (typically two-dimensional) space, with most trajectory methods using either principal component analysis (PCA) [1], t-distributed stochastic neighbor embedding (t-SNE) [8], uniform manifold approximation and projection (UMAP) [9], or diffusion maps [10]. Although a number of methods, including popular and well-performing ones such as Slingshot [7], allow users complete flexibility in choosing a dimension reduction technique.

In addition to the decision-making responsibility users encounter during analysis, the concern arises from the possibility that these decisions may result in significant and unanticipated repercussions (Figure 1). Trajectories inferred using different dimension reduction techniques or numbers of highly variable genes may result in visually similar, biologically reasonable orderings of cell-types clusters along a trajectory (Figure 1A). However, even among visually similar trajectories, the distribution of cells along the trajectory can be highly inconsistent (Figure 1B). These discrepancies become most apparent in downstream analysis when identifying genes that are dynamic along pseudotime, where one may obtain contradictory estimates of gene dynamics (Figure 1C). Towards this end, a recently published tool was designed to facilitate selection of an optimal tree-shaped trajectory by examining cell connectivity [11]. However, it was designed for a specific trajectory inference tool, topology type, and only considers one aspect of the trajectory. The Saelens *et al.* (2019) benchmarking study included an interactive guideline tool called dynguidelines to assist users in selecting optimal trajectory inference methods based on several factors, including the data's expected topology, computational constraints (such as time and memory limits), characteristics of the dataset (e.g., known presence of start and end cells) and the usability of methods. However, this tool does not account for the impact of processing steps on the resulting trajectory. As shown in Figure 1, using a single recommended method, in this case, Slingshot, was not alone sufficient to guide users towards an optimal trajectory. Thus, while dynguidelines offers a general overview to aid in the identification of appropriate TI methods, it lacks further consideration of preprocessing and other analysis choices that affect trajectory analysis.

Overall, there is little guidance on how various processing choices or hyperparameter settings might impact trajectory estimation in general, as well as a lack of quantitative information on whether a particular trajectory adequately fits the dataset. As a result, researchers tend to rely on pre-existing knowledge and subjective visual assessment, potentially biasing the analysis and limiting discovery. Evaluation metrics can provide independent assessments of performance when no ground truth exists. For example, the silhouette statistic and other stability measures have been utilized to evaluate clustering of cells in scRNA-seq datasets [12–15]. These metrics assess dataset-specific performance in terms of good or desirable properties, e.g. smaller within and larger between cluster distances. While these types of evaluation metrics allow for more-informed decision making when clustering scRNA-seq data, they are not applicable to the continuous nature of trajectory analysis.

To this end, we introduce Escort, a novel framework for evaluating the impact of choices in trajectory inference, aiming to assist users in navigating trajectory inference analysis in scRNA-seq data. The main contributions of Escort can be summarized as (i)

evaluating a dataset's overall suitability for trajectory inference, (ii) using trajectory-specific evaluation metrics that we developed to quantify desirable trajectory properties, (iii) classifying sets of processing choices as recommended or not-recommended for trajectory estimation and (iv) obtaining more accurate trajectories based on data-driven evaluations. To the best of our knowledge, Escort is the first approach for comprehensively evaluating processing choices in the context of scRNA-seq trajectory inference. We expect Escort to reduce the decision-making burden in trajectory inference analysis, resulting in less biased and more accurate single-cell trajectories. Escort is available as an R package and available for use as an integrated R/Shiny application, which can be downloaded via GitHub (https://github.com/xiaorudong/Escort).

## RESULTS
### Trajectory estimation is susceptible to varied analysis choices

In an effort to create a common framework for cell trajectory analysis, we first demonstrated that processing decisions such as choosing different numbers of highly variable genes and different dimension reduction techniques can affect an inferred trajectory. Although these processing steps are only a subset of specific decisions users face, they were chosen for their near universality within trajectory inference methods. We considered eight different simulation scenarios consisting of differences in simulators for scRNA-seq data and trajectory topologies (Table 1) [4, 16]. We refer to each combination of analysis choices as an 'embedding', given that the vast majority of trajectory estimation methods construct the trajectory in two-dimensional representations. In general, an embedding encompasses all processing choices of interest—including the dimension reduction—and is the dataset upon which any particular method estimates the trajectory. In the simulation, the embeddings under evaluation differed in their proportion of highly variable features selected (20, 40 and 100%) and dimension reduction technique (t-SNE, UMAP, MDS). To avoid over-generalizing, we used both Slingshot [7] and Monocle3 [17] to infer trajectories for each embedding. Accuracy of the estimated trajectories was assessed by two different metrics (Kendall rank correlation and mean squared error), each of which we standardized by ranking embedding performance within datasets independently. Overall, there was no clear advantage observed for any specific dimension reduction algorithm, percent of highly variable genes, or combination thereof in terms of performance across datasets (Figure 2). This demonstrates that it is not feasible to simply pre-select a best dimension reduction algorithm or percent of highly variable genes for every analysis or trajectory method.

### Framework to identify recommended embeddings for trajectory analysis

Given the dataset-specific performance across trajectory analysis decisions, we developed Escort, a data-driven evaluation framework to guide users through trajectory analysis by providing evaluations of sets of analysis choices (Figure 3). In the first step, Escort assesses whether the data support the existence of a trajectory. Non-computational scientists often struggle with the first step of constructing a trajectory, which is deciding whether fitting a trajectory is appropriate for their dataset. There are two scenarios where trajectory fitting is not well-suited: when cells represent biologically distinct cell types or when the data has insufficient cellular heterogeneity (Figure 3A).
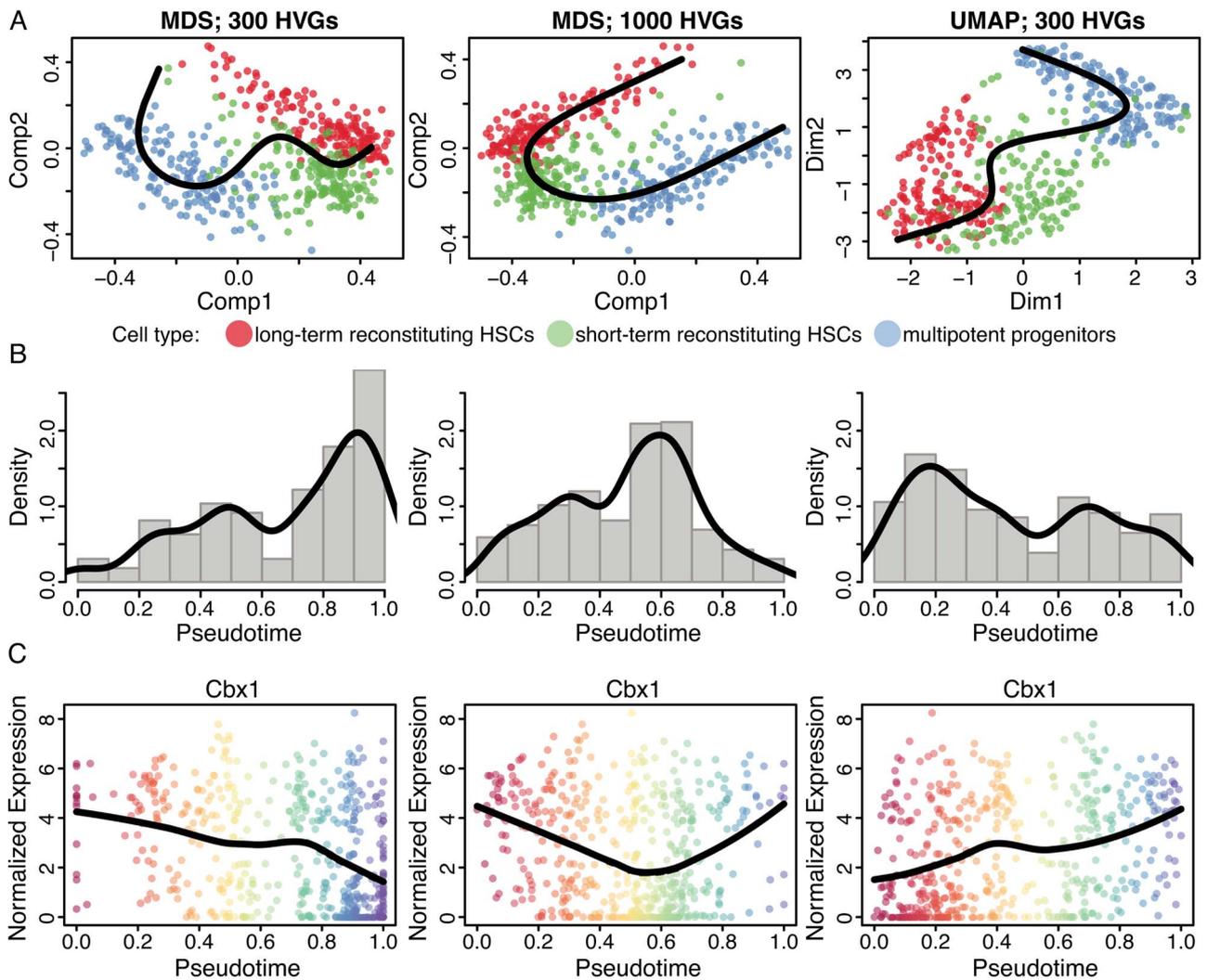
**Figure 1.** Analysis choices significantly impact trajectory estimation in scRNA-seq data. For three different choices of selected genes and dimension reduction methods, trajectory inference and pseudotime estimation were performed on an scRNA-seq dataset of hematopoietic stem cells [53]. (**A**) Dimension-reduced spaces and estimated trajectories with cells colored by cell type. The plot title in each column indicates the dimension reduction used (MDS or UMAP) and the number of highly variable genes selected (300 or 1000). (**B**) Pseudotime distributions for each set of analysis choices. (**C**) Normalized gene expression as a function of pseudotime for *Cbx1*. Cells are colored by pseudotime, i.e., their location along the trajectory. Abbreviations: MDS = multidimensional scaling, UMAP = uniform manifold approximation and projection, HVG = highly variable gene.

**Table 1:** A summary of simulated scRNA-seq data datasets

| Simulation | Method | Trajectory type | Number of cells (post-filtering) |
| --- | --- | --- | --- |
| sim-1 | Scaffold [16] | Linear | 500 |
| sim-2 | dyntoy [54] | Linear | 263 |
| sim-3 | Splatter [55] | Linear | 174 |
| sim-4 | dyntoy [54] | Bifurcation/tree | 748 |
| sim-5 | dyntoy [54] | Multifurcation/tree | 1617 |
| sim-6 | dyntoy [54] | None (disconnected) | 251 |
| sim-7 | dyntoy [54] | None (disconnected) | 674 |
| sim-8 | Scaffold [16] | None (homogenous) | 500 |

Biologically, datasets consisting of distinct cell types indicates that the underlying biological processes are separate or that the experiment did not capture sufficient intermediate-stage cells. Datasets lacking cell heterogeneity may indicate low sensitivity in the experimental assay or excessive technical variability in the dataset. In either case, Escort will flag the dataset in the first step and return a summary with suggestions to the user

specifying how to proceed. If a trajectory signal is detected, in the next step, the analysis choices (embeddings) that could be used for trajectory estimation are evaluated. Users are able to select a number of default embeddings for consideration or input their own embeddings directly. In this step, Escort quantifies how effectively an embedding preserves relationships between cells in the high-dimensional data and assesses the distribution of points
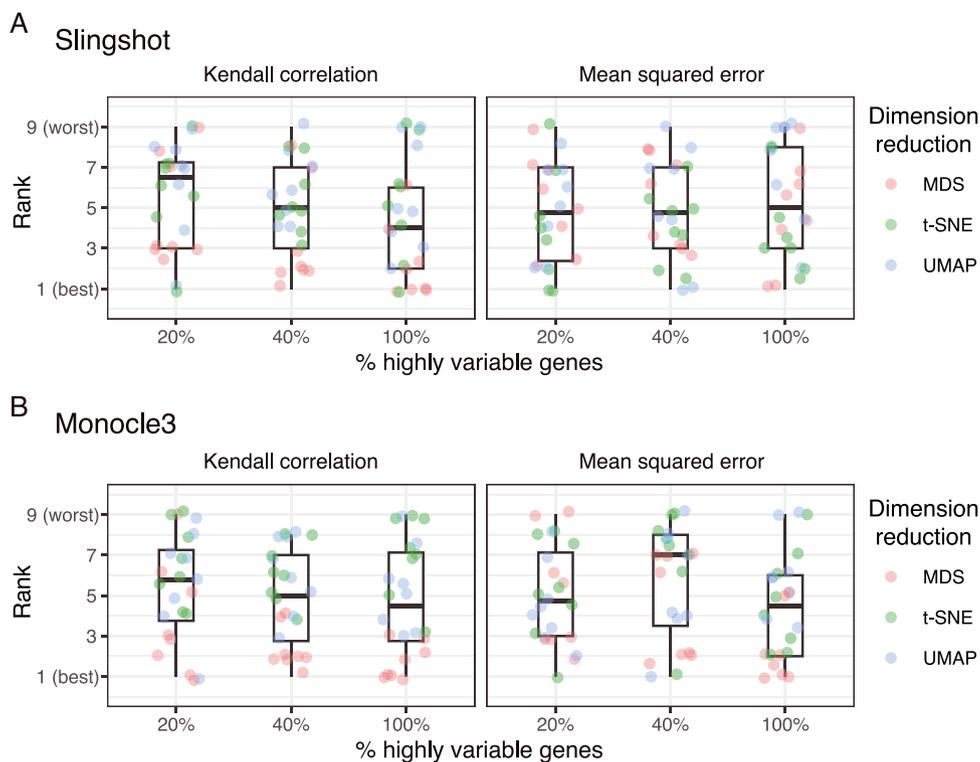
**Figure 2.** Trajectory accuracy is impacted by different dimension reduction algorithms and inclusions of highly variable genes. (**A**) The performance of different embeddings across all eight simulated scenarios is shown. Embeddings were ranked within each dataset separately for the two metrics. The ranks were scaled so that a lower rank indicated better within-dataset performance. (**B**) Similar to A using Monocle3.

in the two-dimensional cell graph. Based on these two evaluations, an embedding may be classified as non-recommended or passed to the next step (Methods). For the third step, embeddings are evaluated in the context of a specific trajectory inference approach in order to allow for consideration of method-specific hyperparameters. Specifically, Escort evaluates the proportion of cells likely to have an ambiguous projection along a trajectory. An overall performance score is calculated based on metrics in the second and third steps, and then the embeddings are classified as recommended or non-recommended. Recommended embeddings are those which are likely to generate trajectories having a better fit to the data and more accurately reflect the underlying dynamic biological processes. Non-recommended embeddings are those that are unlikely to generate accurate trajectories based on the evaluations.

## Escort distinguishes embedding quality in simulations

We demonstrate the effectiveness of Escort in evaluating various embeddings on the eight different simulation scenarios (Table 1). We used Slingshot to construct trajectories for each embedding, although the results were consistent with Monocle3 (Supplementary Figure 1). Accuracy was assessed in terms of cell order and total order error by comparing the estimated trajectory to the ground truth via Kendall rank correlation and mean squared error. Overall, the recommended embeddings tended to produce more accurate trajectories. Specifically, the recommended embeddings had higher correlation and lower error than non-recommended embeddings (Figure 4A). The simulations with no true trajectory were all detected and flagged by Escort in the first step and their embeddings generated significantly less accurate trajectories than even those by the non-recommended

embeddings. The Escort score also correlates well with the individual accuracy measures indicating that a higher Escort score reflects higher trajectory accuracy (Figure 4B).

## Escort guides decision-making for trajectory inference

Additionally, we analyzed five scRNA-seq datasets obtained from publicly available sources and encompassed a range of biological contexts (Table 2). While the scRNA-seq datasets do not have a ground truth in the sense of a known trajectory, we chose datasets that had some degree of biologically relevant time-ordered samples allowing us to evaluate a trajectory's fit to the data. Despite the increased noise and complex data structure, Escort is still able to distinguish embedding quality (Figure 5 and Supplementary Figure 2). The recommended embeddings have a significantly larger accuracy compared to the non-recommended embeddings. The accuracy measures are noisier for these datasets compared to the simulations due to the small number of true times available, however higher accuracy is still correlated with a higher Escort score (Figure 5B).

We also found that the scRNA-seq datasets tended to have more non-recommended embeddings compared to the simulated datasets. Non-recommend embeddings are a result of having a negative Escort score, and, additionally, when the embedding fails to sufficiently preserve complex cell relationships, it is immediately classified as non-recommended. For example, we found that UMAP frequently generated embeddings that exhibited disconnected clusters or grouped small clusters together in unexpected ways [18]. These more complex datasets also have more irregularly spaced sampling times, for example, the sc-3 dataset (Table 2) had single-cells measured at days 100, 130, 175 and then day 450. If too few cells are represented from intermediate
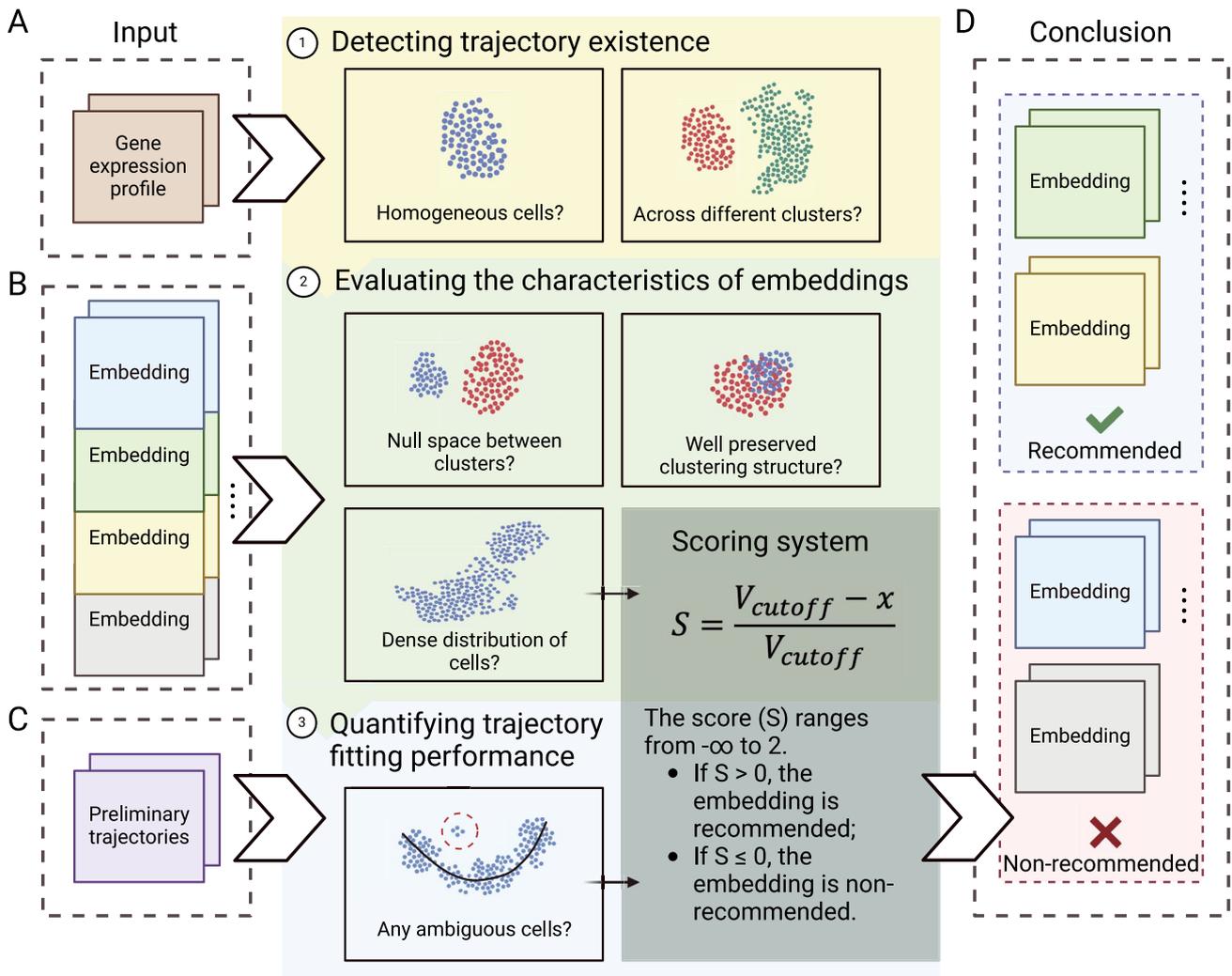
**Figure 3.** Overview of ESCORT. Schematic of the Escort workflow. (**A**) The first step detects the presence of a trajectory signal in the dataset before proceeding to evaluations of embeddings. (**B**) Various metrics are used to evaluate user-defined embeddings regardless of the ultimate trajectory inference method to be used. (**C**) In the final step, the preferred trajectory inference method of the user is used to fit a preliminary trajectory to evaluate method-specific hyperparameters. (**D**) Based on the overall score, embeddings are classified as either recommended or non-recommended.

**Table 2:** A summary of public scRNA-seq datasets. Datasets sc-2 through sc-5 were downloaded from Saelens *et al.* [4]

| Dataset | Organism | Accession | Number of cells after filtering | Number of time points |
|---------|----------|-----------|--------------------------------|-----------------------|
| sc-1 | Human | E-MTAB-3929 [56] | 1469 | 5 |
| sc-2 | Mouse | GSE59114 [53] | 492 | 3 |
| sc-3 | Mouse | GSE99951 [57] | 448 | 4 |
| sc-4 | Mouse | GSE87375 [58] | 321 | 6 |
| sc-5 | Human | GSE86146 [59] | 647 | 12 |

sampling times or cell states, it is unlikely any embedding could overcome experimental design limitations and consistently generate robust connected trajectories.

## Computation time benchmark

To assess the computational requirements of Escort, we simulated datasets having varying numbers of cells using the same framework as the simulation study. The full results are shown in Supplemental Table 1. Briefly, with 1000 cells, Escort's first step took an average of 5 min to run, although it increased to 18 min with 3000 cells. Escort's embedding evaluation in Steps 2 and 3

took between 0.4 and 1.5 min for datasets having 1000 or 3000 cells for a single embedding, respectively. Overall, the computational time is reasonable given that trajectory inference involves analyzing a smaller subset of cell types from an experiment and the embedding evaluations are parallelizable.

## Escort guided trajectory analysis of hypertrophic chondrocyte transdifferentiation

Hypertrophic chondrocytes were once thought to be the terminal state of chondrocyte fate prior to apoptosis, yet recent research suggests these cells undergo transdifferentiation towards an
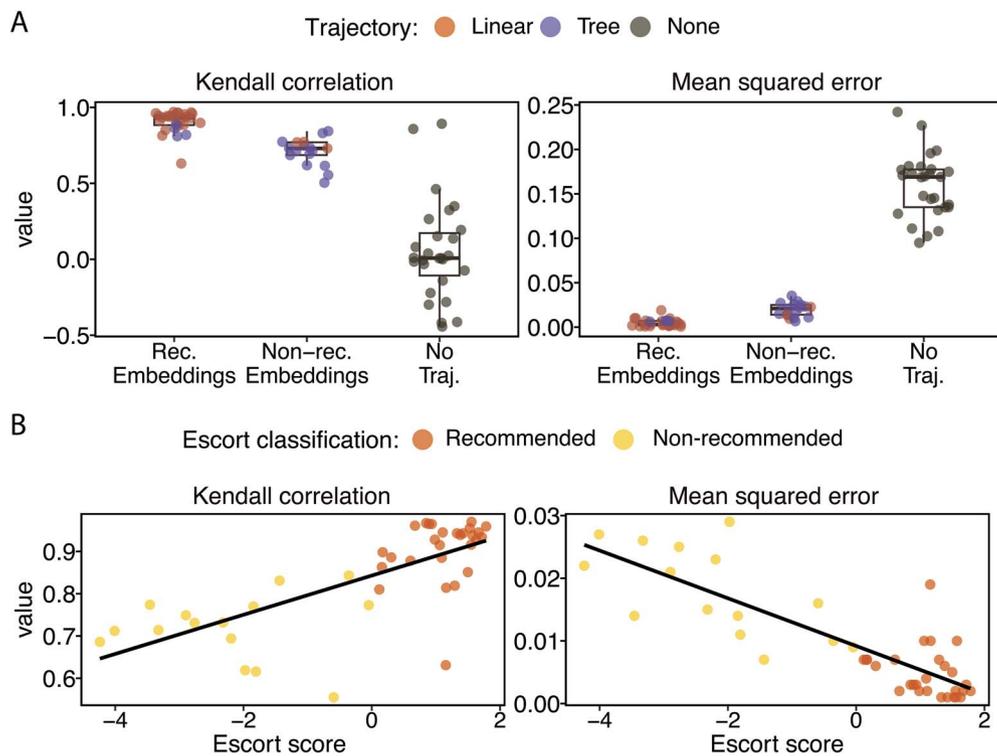
**Figure 4.** Trajectory assessment performance of Escort on simulated datasets. (**A**) The accuracy of trajectories generated on nine different embedding options for each of the eight simulated datasets is shown for different metrics: Kendall rank correlation and mean squared error. Simulated scenarios differ in terms of true trajectory topology (denoted by color) and simulator methods. The y-axis displays the values for the accuracy metric. (**B**) Each embedding's Escort score (x-axis) versus the value for each accuracy metric (y-axis) are shown and colored according to their classification by Escort.
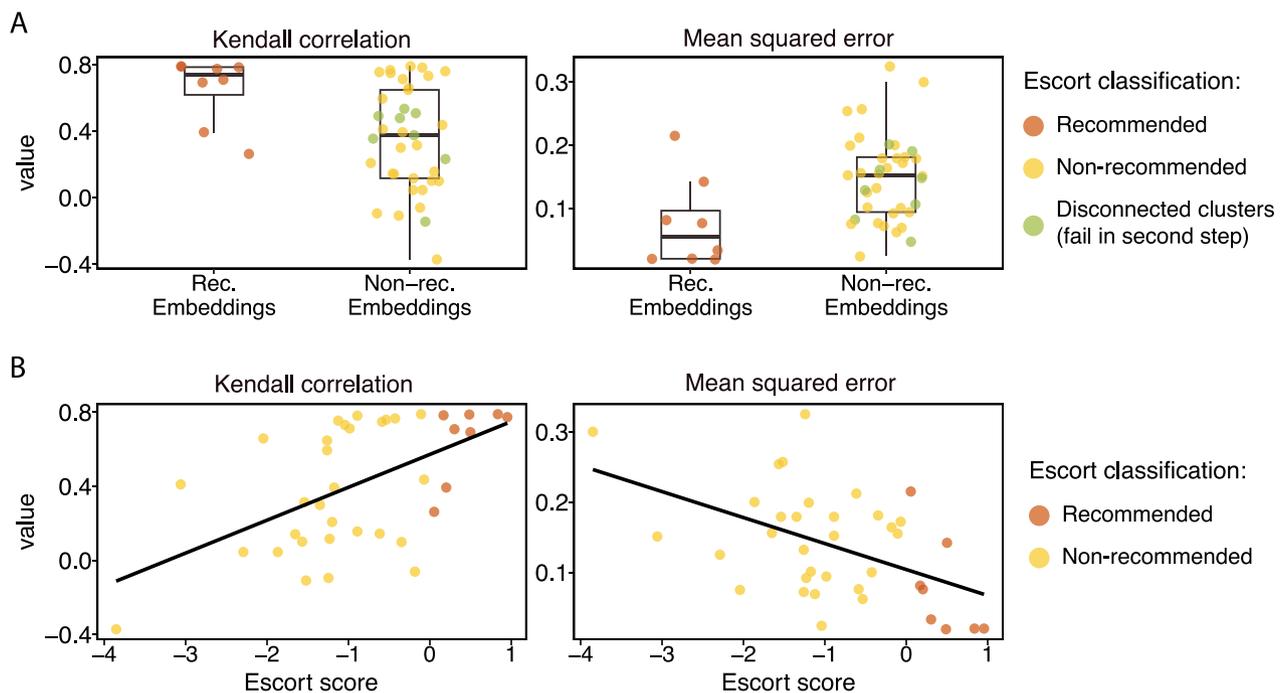


**Figure 5.** Trajectory assessment performance of Escort on public datasets. (**A**) The accuracy of trajectories generated on nine different embedding options is shown for five publicly available datasets assessed using different metrics: Kendall rank correlation and mean squared error. The colors distinguish each embedding classification by Escort, in addition to those embeddings that failed in the second step. The y-axis displays the values for accuracy metrics. The x-axis corresponds to recommendations generated by Escort. (**B**) Similar to (A) with the x-axis showing the Escort score.

osteoblast-like state or other marrow associated cells [19]. To better understand the fates of hypertrophic chondrocytes, Long *et al.* [20] utilized lineage-tracing mouse-reporter models to isolate *Col10a1*-expressing cells and their descendants. The authors

performed trajectory inference on their scRNA-seq data using default settings in Monocle3, resulting in a singular lineage linking hypertrophic chondrocytes to osteoblasts. Using the original paper's publicly available dataset, we sought to examine whether

Escort would recommend a more optimal embedding that might provide additional insight into this differentiation process.

Prior to trajectory estimation, we found that one of the initial cell clusters (cluster 7) appeared to be low-quality cells with very low sequencing depth (Supplemental Figure 3). Removing these cells from the dataset resulted in a highly similar, single-lineage trajectory using the default settings in Monocle3 (Figure 6A). However, for our analysis, given the interest in the multi-fate hypothesis of hypertrophic chondrocytes, we opted to use Slingshot to fit trajectories on both the original embedding and an Escort recommended embedding (Figure 6B and C). Escort evaluated 13 various embedding options. The original embedding scored sub-optimally and was ranked in the bottom quartile among all embeddings (Supplementary Table 2). An Escort-based trajectory was fit using Slingshot with the highest-scoring embedding option.

Topologically, the two trajectories were highly similar, although the original embedding generated three lineages while Escort's embedding generated two (Figure 6B–C). Based on the clusters along each branch, we labeled the two shared lineages as Lineage A and Lineage B. Based on the original author's annotations, Lineage A follows hypertrophic chondrocytes to osteoblasts and Lineage B links hypertrophic chondrocytes to terminal hypertrophic chondrocytes. The third lineage, only observed using the original embedding, was enriched for VEGFA signaling and protein transport, which indicated heterogeneity within the dedifferentiating hypertrophic chondrocytes (Supplemental Data 1). For our analysis, we focused on comparing how the two embedding options differed in downstream analysis when characterizing gene dynamics in Lineages A and B. For each lineage, we applied scLANE to identify differentially dynamic genes [21]. scLANE models each gene's expression as a function of pseudotime using a modified multivariate adaptive regression spline approach. The scLANE model identifies locations of significant expression changes along pseudotime (referred to as knots) and estimates the expression slopes along each interval.

For Lineage A, the osteoblastic fate, the two trajectories had similar overall topology and pseudotime distributions, however, the Escort-based trajectory resulted in 12% more genes identified as significantly dynamic compared to the original trajectory (Figure 6D). The distribution of knots, indicating locations of major expression change along the lineage, was multi-modal for the Escort trajectory whereas the original trajectory had a unimodal distribution with a high concentration of knots when transitioning between clusters 6 and 8, indicating disconnectedness. Although the majority (> 85%) of shared differentially dynamic genes had similar trends, the Escort-based dynamics were less noisy and captured more subtle expression changes as seen in chondrogenic differentiation genes *Snorc* [22] and *Id2* [23] (Figure 6F).

The lineage towards more terminal hypertrophic chondroblasts, Lineage B, varied substantially between the two trajectories (Figure 6G). The original trajectory included cells from Cluster 6, thought to be a transitory stem and progenitor-like cell population, which were predominantly limited to Lineage A in the Escort trajectory. The knot distribution for the original trajectory had a large proportion of knots appearing in the initial transition from clusters 1 to 3, likely due to a small gap in the pseudotime, and a large proportion of knots appeared at the transition between clusters 5 and 6 (Figure 6H). These differences in pseudotime distributions lead to 35% of all shared significantly dynamic genes having opposing dynamic trends (upregulation versus downregulation). Genes upregulated in the original trajectory but downregulated in the Escort trajectory were enriched for cartilage and skeletal system development. Genes upregulated in the Escort trajectory and downregulated in the original trajectory were enriched for regulation of RNA splicing and the EGFR1 pathway, which stimulates terminal differentiation and apoptosis in chondrocytes [24] (Supplemental Data 2). Lineage B in the Escort trajectory clearly represented cells heading towards the apoptotic fate, whereas the original trajectory contained blended, less distinct signals among its lineages. The Escort-guided analysis led to a trajectory that more accurately characterized the underlying transdifferentiation of hypertrophic chondrocytes, with two clearly distinct lineages towards osteoblastic fate and terminal apoptosis, whereas the embedding used in the original analysis was only able to characterize the osteoblastic linage.

## DISCUSSION

While all-inclusive or automated analysis pipelines are desirable for ease of use, recent discussions have highlighted issues regarding careful inference in single-cell data [25–27]. In this context, it is crucial to consider the unique characteristics of the data when selecting among trajectory inference methods and performing processing steps. To address this challenge, we developed Escort to evaluate the impact of choices in trajectory inference while guiding users through the estimation of trajectories in scRNA-seq data analysis. Escort effectively selects optimal and more accurate trajectories based on trajectory-specific evaluation metrics as demonstrated by our simulation and case-study results. Escort also assists users in determining whether their dataset is appropriate for trajectory analysis by identifying disconnected cell clusters or homogenous cells lacking a trajectory signal. In either case, Escort will provide relevant data summaries, allowing users to reassess their hypothesis or further refine their dataset. Our framework is implemented in an R package that allows users to evaluate as many embedding options as desired, as well as through an R/Shiny application to guide users more directly through assessing data-specific properties to consider when performing trajectory inference.

Given that trajectory inference methods often differ in their use of dimension reduction techniques, our results observing highly dataset-specific performance are consistent with that observed in the benchmarking study by Saelens *et al*. [4]. In fact, the processing choices for dimension reduction and inclusion of highly variable genes impacted the trajectory accuracy more than the difference between Slingshot and Monocle3 for estimating the trajectory and pseudotime. In addition to standard processing steps, quality control steps may also significantly impact Escort's embedding evaluations. For example, normalization is a key upstream pre-processing step that adjusts for variations among cells due to sequencing depth. Inadequate normalization may compromise accurate trajectory signal detection [28, 29]. Batch effects also have the potential to introduce technical variability between samples [30]. Batch effects in scRNA-seq data can result in incorrect trajectory signal assessment and the emergence of batch-specific clusters or trajectories that do not accurately represent biological characteristics [31–33]. Thus, we highly suggest that these decisions also be defined explicitly in the embedding to allow for evaluation within the Escort framework.

Additionally, we and others have found that the performance of UMAP is heavily dependent on hyperparameters, specifically the minimum distance between points in the lower-dimensional space and the number of approximate nearest neighbors used for constructing the initial high-dimensional graph [34, 35]. The choice of these hyperparameters has a significant impact on
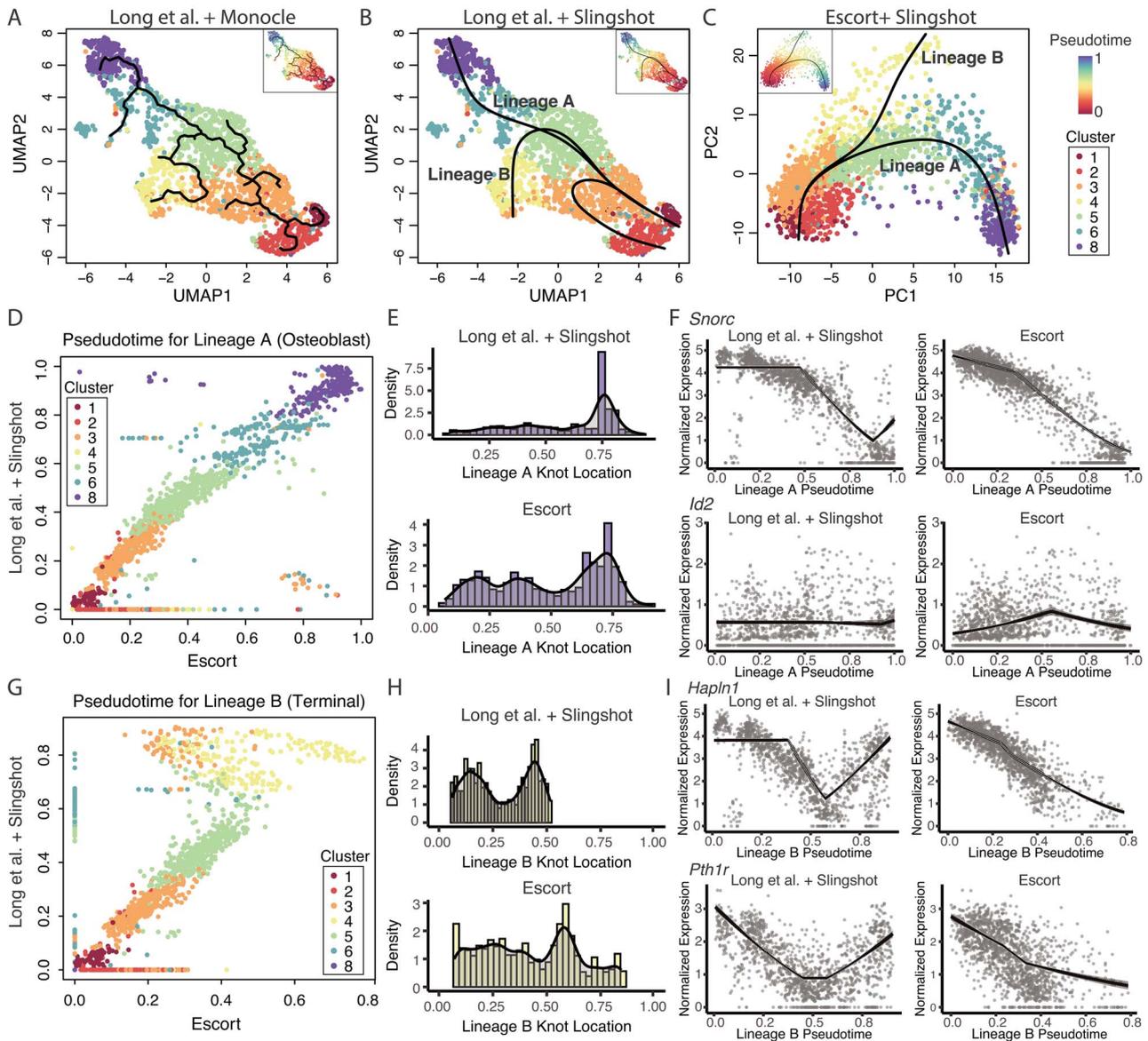
**Figure 6.** Analysis of transdifferentiation of hypertrophic chondroblasts using an Escort-guided trajectory. (**A**) UMAP of the original paper's embedding and the Monocle3 based trajectory. (**B**) UMAP of the original paper's embedding using Slingshot to fit a trajectory. (**C**) Escort recommended embedding using Slingshot to fit a trajectory. (**D**) Correlation of pseudotime between the two Lineage A trajectories. (**E**) Distribution of knots across all significantly dynamic genes for Lineage A. (**F**) Gene expression as a function of pseudotime for *Snorc* and *Id2*. (**G–I**) Similar to (D–F), but for *Hapln1* and *Pth1r* in Lineage B.

how UMAP behaves [36, 37]. Hyperparameters such as these can also be incorporated within the Escort evaluation framework. Alternatively, there are numerous dimension reduction methods available for scRNA-seq data, including those derived from RNA velocity estimations. Including these alternatives directly into the Escort framework is an area of future work, although users are free to implement them when generating embeddings. Presently, Escort only considers trajectory inference approaches that use two-dimensional representations. While this encompasses the vast majority of existing approaches, assessing methods outside of this restriction requires further evaluation.

Finally, Escort is not guaranteed to identify a recommended embedding among the choices being evaluated. We recommend evaluating a wide variety of processing choices, however maintaining connected relationships within an embedding in scRNA-seq data is challenging, as previously reported by others [18, 38,

39]. Insufficient experimental designs may also present situations with small discontinuities. Given that the Escort score was highly correlated with accuracy of the trajectory, if a reasonable number of embedding options have been explored, Escort's score can be utilized in cases where it may be necessary to choose the best possible option, understanding that there is some unavoidable discontinuity present. In any case, Escort provides users with the ability to identify more optimal analysis choices in the context of trajectory inference.

## METHODS

Our three-step framework aims to improve the accuracy of trajectory estimation while guiding users through the decisions involved in fitting a trajectory.

## Step 1: detecting trajectory existence

The first step of trajectory analysis is deciding whether fitting a trajectory is appropriate for a given dataset. There are two scenarios where trajectory fitting is not appropriate: when cells represent diverse cell types or when cells appear homogeneous.

To identify distinct cell types, we use two single-cell specific approaches to identify clusters: scLCA [40] and SC3 [41]. The scLCA approach estimates the optimal number of clusters and cell clusters in the dataset based on cosine-similarity and spectral clustering [14]. SC3 is a consensus cluster approach; a consensus matrix is constructed from clustering with multiple distance metrics and then split into a user-specified number of clusters. In scenarios where cell type information is lacking, Escort depends on the optimal number of clusters and their assignments determined by scLCA. Conversely, if prior information regarding cell types is known, Escort utilizes SC3 to compute the cell clusters. Next, Escort computes distances between the cell clusters in the high-dimensional space to evaluate their connectivity. This assessment assumes that large distances between clusters indicates disconnectedness between the cell types. To evaluate connectivity between clusters, let $N$ be the total number of cells in the dataset. Each cell is denoted by $c_i$, for $i \in \{1, 2, \ldots, N\}$. The Manhattan distance matrix $D_{N \times N}$ is defined as

$$D_{N \times N} = \begin{bmatrix} d_{1,1} & \cdots & d_{1,N} \\ \vdots & \ddots & \vdots \\ d_{N,1} & \cdots & d_{N,N} \end{bmatrix}, \quad (1)$$

where $d_{i,j}$ represents the Manhattan distance between cell $c_i$ and $c_j$.

The $m$ cell clusters are denoted as $T_1, T_2, \ldots, T_m$, such that the cluster $T_k$ consists of a subset of $n_k$ cells. If $c_i \in T_k$, then $D_{within}^{c_i}$ is the set of 'within-cluster distances' for $c_i$ such that $D_{within}^{c_i} = \left\{ d_{i,k_1}, d_{i,k_2}, \ldots d_{i,k_{n_k}} \right\}$ and $D_{between_q}^{c_i}$ represents a set of 'between-cluster distances' for $c_i$ between cells in $T_q$, given by $D_{between_q}^{c_i} = \left\{ d_{i,q_1}, d_{i,q_2}, \ldots d_{i,q_{n_q}} \right\}$, where $q \in \{1, 2, \ldots, m\}$ and $q \notin k$. Jaccard index scores are then computed for $c_i$ based on the distributions of $D_{within}^{c_i}$ and $D_{between_q}^{c_i}$. Cells with high Jaccard indices across multiple clusters indicate low cluster specificity and their potential presence at the boundaries of clusters. The Jaccard index is recalculated for the top 20% of cells with the highest Jaccard index score within each cluster. Two clusters are considered to be connected if, in pairwise comparisons, there are more than 10 cells meeting the 0.3 cutoff criteria for each cluster. The threshold of 0.3 and a minimum of 10 cells was determined based on simulations. Otherwise, clusters are considered disconnected.

To assess the homogeneity of the cells in the dataset, Escort analyzes the correlation between the first principal component of the normalized expression data (PC1) and a random subset of the top 100 highly variable genes. This assumes that if a true trajectory signal exists, then it would be encoded within multiple highly variable genes and also captured by the first principal component. A significant correlation between PC1 and more than 46% of the highly variable genes at a significance level of 0.05 was used as our criterion, determined through a simulation study, for identifying a trajectory signal. Significance is assessed via a permutation test with 20 000 iterations of the Spearman correlations accompanied by a false discovery rate correction.

## Step 2: evaluating the characteristics of embeddings

Once Escort has determined the suitability of trajectory analysis, the second step is designed to evaluate embeddings for performing trajectory inference. Since all methods employ some form of dimension reduction, we refer to the collective set of choices as an 'embedding' of the dataset. However, in practice, an embedding may also consist of methods for normalization, selection of highly variable genes, or other method-specific hyperparameters.

The first evaluation of an embedding is the retention of intercellular relationships that are present in the high-dimensional data. While we do not expect a (typically) lower dimension embedding to perfectly preserve high dimensional relationships, Escort penalizes severe distortions as the accuracy of trajectory prediction is heavily dependent on the extent to which these relationships are preserved [4]. The distance-based method introduced in Step 1 is utilized again to evaluate cell connectivity on the embedding. Since disconnected clusters were resolved in Step 1, a reliable embedding should not exhibit newly disconnected clusters. Consequently, any embeddings found to be disconnected are classified immediately as not recommended for trajectory inference.

For any remaining embeddings, the preservation of similarity relationships in the low-dimensional embedding is evaluated by employing what we define as the 'same group level' method. For each cell in the embedding, the three closest cells based on Euclidean distance in low-dimensional embeddings are identified. A cell has a high 'same group level' when at least two of its closest neighbors belong to its same cluster as defined in Step 1. The clustering structure observed in the higher-dimensional space is considered to be well-preserved in a given embedding when a large number of cells exhibit a high 'same group level'. For example, consider the detection of the three closest neighbors for a cell $c_i$. The set $NS_i = \{c_{i1}, c_{i2}, c_{i3}\}$ contains the three closest neighbors for $c_i$. If $c_i \in T_k$ and $\{c_{i1}, c_{i2}, c_{i3}\} \notin T_k$, then the 'same group level' is considered 0. If $\{c_i, c_{i1}\} \in T_k$ and $\{c_{i2}, c_{i3}\} \notin T_k$, the 'same group level' is considered 1. If the 'same group level' is equal to or greater than 2, we conclude that the cell exhibits a high level of similarity with its neighboring cells. The preservation score for similarity relationships is determined by the percentage of cells exhibiting a high level of similarity.

The next embedding evaluation considers cell density. If cells are more uniformly distributed in the two-dimensional embedding space, the trajectory inference method faces challenges in identifying a robust trajectory. Conversely, if cells are denser and exhibit a distinct topology, the trajectory inference method is more likely to generate a well-defined curve. Since methods frequently fit trajectories on two-dimensional representations, we quantify the cell density by calculating the 'cell coverage area' using the area of the $\alpha$-convex hull [42]. The proportion of this area to a minimum circle enclosing all cells in the two-dimensional embedding space is then calculated.

## Step 3: quantifying trajectory fitting performance

So far Escort has provided independent evaluations of embeddings for trajectory inference based on general properties, specific methods may impose additional specific graph structure. Thus, the final step for Escort accounts for any method-specific variations in the performance of embeddings.

For each embedding, a rough trajectory is fit using an assumed method. We use Slingshot as the default, however, other methods are easily incorporated at this stage and described in the user

vignette. With the embedding-specific trajectory, Escort estimates the proportion of cells positioned along the trajectory such that their projection is ambiguous. For example, trajectories in a U-shape tend to be less accurate due to the presence of cells that map with similar probability to the beginning or end of the trajectory. Each cell's pseudotime is computed by each trajectory inference method by projecting cells onto the trajectory. Escort calculates each cell's pseudotime standard deviation based on the closest 10% of its nearest neighbor. Ambiguous cells are then defined as those with an extremely large standard deviation exceeding the upper fence identified by a skew-adjusted approach [43]. If multiple lineages are present, then this step is performed per lineage, followed by the calculation of the total number of uniquely ambiguous cells.

## Escort scoring

A comprehensive score incorporating the assessments above is used to evaluate the overall performance of each embedding. The score consists of three components: proportion of cells having a high same-group level (Step 2), cell coverage area (Step 2) and the proportion of ambiguous cells (Step 3). Embeddings deemed non-recommended based on the disconnectedness evaluation in the second step are not included in the score comparison.

To establish a realistic and standardized benchmark scoring system, we conducted simulations using Scaffold based on a human pancreas single-cell RNA-seq dataset [16, 44]. We simulated a total of 50 scRNA-seq datasets, each comprising 500 cells, 15 894 genes and including 20% dynamic genes. For each of the simulated datasets, we produced nine different embeddings by employing MDS, UMAP and t-SNE based on a subset of 2000, 4000 or 10 000 highly variable genes. Subsequently, we assessed all generated embeddings by computing both mean squared error (MSE) and Kendall rank correlation coefficients. We identified good embeddings as those with correlation coefficients exceeding 0.85 and MSE less than 0.01. We then fit a beta distribution to the cell coverage area values generated by the good embeddings, as well as a gamma distribution to the proportion of ambiguous cells within these embeddings. We established suitable cutoffs as the 99th percentile of the standard fitted distributions. This led to a cutoff for cell coverage area of 0.53 and for the proportion of ambiguous cells as 0.029. In practice, these cutoffs have proven reasonable.

To classify embeddings as recommended or not recommended based on the score, Escort scales the scores separately for the density and ambiguous components as follows:

$$S = \frac{V_{cutoff} - x}{V_{cutoff}}, \tag{2}$$

where $x$ is the value of cell density or proportion of ambiguous cells. For cell density, we set the $V_{cutoff}$ to 0.53, while for the proportion of ambiguous cells, the $V_{cutoff}$ is 0.029. The percentage of cells with a high 'same group level' is included in the score as a weight. The total score for each embedding is calculated as

$$Total\ score = (S_{density} + S_{ambiguous}) \times W_{same\ group\ level}, \tag{3}$$

where $S_{density}$ is the scaled score for the coverage area components, $S_{ambiguous}$ denotes the scaled score for ambiguous cells component and $W_{same\ group\ level}$ represents the percentage of cells with a high 'same group level.' The score ranges from negative infinity to two, with higher scores indicating better performance. Escort

reports embeddings with a score greater than zero as recommended, while those with a score less than or equal to zero as non-recommended. Through this scoring system, users can readily identify more optimal embedding choices to construct a trajectory.

## Application of Escort to simulation and single-cell RNA-seq datasets

We generated two simulated datasets using the Scaffold v0.2.0 R package [16], one consisting of homogenous cells and the other featuring a linear trajectory. The reference data for simulation was obtained from human pancreas single-cell RNA-seq [44]. The estimateScaffoldParameters() function was utilized to configure all simulation parameters. In the dataset with homogenous cells, no dynamically expressed genes (i.e., genes influencing a trajectory), were specified. For the linear structure dataset, we specified that 20% of the genes should be dynamic. The simulateScaffold() function was applied to run the simulations based on the parameters from the estimateScaffoldParameters() function. Additionally, we utilized six simulated datasets that are publicly available [45].

For each simulated dataset, we initially eliminated cells with duplicated ground truth time. Then, we preprocessed each dataset by removing low-quality cells based on extreme outliers (low or high) counts of unique genes, total molecule counts per cell, and a high percentage of mitochondrial genome reads using Seurat v5.0.1 [46]. Genes having less than three total counts (or less than 10 for the sc-1 dataset) were also filtered out. The normalization method 'LogNormalize' was then applied using the Normalize-Data() function in Seurat. Highly variable genes were identified using scran v1.28.2 [47]. The modelGeneVar() function identified genes demonstrating higher variability than expected given their mean expression. We selected features based on biological variability by utilizing the 'bio' argument in the getTopHVGs() function, picking the top 20, 40 and 100% proportion of genes. Next, t-SNE, UMAP and MDS were applied as dimension reduction techniques to generate embeddings. t-SNE was performed by the Rtsne v0.16 package, involving an initial PCA on the normalized data with the default setting retaining the top 50 dimensions. UMAP was implemented using the R package UMAP v0.2.10.0 directly on normalized data with default settings. SCORPIUS v1.0.9 [48] was used to implement MDS with the Spearman distance metric. A trajectory was inferred for each embedding using both Slingshot v2.8.0 [7] and Monocle3 v1.3.4 [17]. The mclust package v6.0.1 [49] was used to generate clusters based on hierarchical clustering, and these clusters were subsequently employed as input for Slingshot.

## Computation time evaluation

To assess the computational time required for Escort, we utilized datasets varying in cell count—specifically, 250, 500, 1000, 1500 and 3000 cells, each containing 20 000 genes. This range was chosen to represent a variety of dataset sizes encountered in analyses. The data was simulated using the Scaffold R package v0.2.0 [16] as described for the simulation study. Our tests were performed on a laptop equipped with an Apple M1 Pro chip and 16 GB of RAM. We structured our computational analysis to separately measure the runtime for Step 1 (testing the suitability of trajectory inference) which does not involve a specific embedding and Steps 2 and 3 combined. For Steps 2 and 3, we report the time of evaluating a single embedding. We repeated each scenario five times and calculated both the mean runtime and the standard deviation in min.

## Analysis of hypertrophic chondrocytes

The preprocessed scRNA-seq data from Long *et al.* (2022) was downloaded from GSE190616 [20]. Escort was applied to the normalized expression matrix for 13 different embedding comparisons (Supplemental Table 1), the embedding option using PCA with 10% of the most highly variable genes performed best was selected for further trajectory analysis. Slingshot v2.8.0 [7] was used to fit the trajectory under the Escort embedding option, as well as the embedding used in the original paper's analysis. The R package scLANE v0.7.8 [21] was used to test each gene's expression for trajectory differential expression using the default settings. Genes were considered significantly dynamic if they have an overall false discovery rate adjusted p-value <0.01. Genes were further filtered for enrichment if they had a last segment slope > 5 or < −5. Enrichment was carried out using both Enrichr [50], specifically using the BioPlanet 2019 collection and GSEA Molecular Signatures Database for mouse [51, 52], specifically the GO biological processes collection.

---

**Key Points**

- The performance of trajectory inference methods for scRNA-seq data is dataset-specific and highly vulnerable to data processing choices made during the analysis.
- Users are unaware of the effect that, say, choosing a different dimension reduction approach or number of features may have on the estimated trajectory, both of which are data processing steps for all trajectory inference methods.
- We developed Escort (R package and R/Shiny app) to guide users through trajectory analysis by quantitatively evaluating sets of data processing choices based on metrics assessing desirable trajectory properties.
- Escort also first evaluates whether a given dataset is appropriate for trajectory analysis. This occurs, for example, when datasets contain distinct cell populations, or when the cells are too homogenous.
- Overall, Escort reduces the decision-burden on users during trajectory analysis, leading to more optimal trajectories and ultimately, a better understanding of dynamic biological processes.

---

## SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/bib.

## FUNDING

## DATA AVAILABILITY

The processed simulated, scRNA-seq and hypertrophic chondrocyte datasets are available on GitHub, along with all analysis codes for reproducibility (https://github.com/xiaorudong/Escort-paper).

## CODE AVAILABILITY

Code used to analyze the hypertrophic chondrocyte data along with analysis results is provided on GitHub (https://github.com/xiaorudong/Escort-paper). The Escort package and tutorials are available on our GitHub repository (https://github.com/xiaorudong/Escort).

## REFERENCES

1. Ji Z, Ji H. TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res* 2016;**44**:e117–7.
2. Farrell JA, Wang Y, Riesenfeld SJ, *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 2018;**360**:eaar3131.
3. Kim M-C, Borcherding N, Ahmed KK, *et al.* CD177 modulates the function and homeostasis of tumor-infiltrating regulatory T cells. *Nat Commun* 2021;**12**:5764.
4. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* 2019;**37**: 547–54.
5. Trapnell C, Cacchiarelli D, Grimsby J, *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;**32**:381–6.
6. Wolf FA, Hamey FK, Plass M, *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* 2019;**20**:59.
7. Street K, Risso D, Fletcher RB, *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 2018;**19**:477.
8. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579–605.
9. McInnes L, Healy J, Melville J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv 2018;1802:03426.
10. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* 2015;**31**:2989–98.
11. Smolander J, Junttila S, Elo LL. Cell-connectivity-guided trajectory inference from single-cell data. *Bioinformatics* 2023;**39**:btad515.
12. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience* 2018;**7**:giy083.
13. Duò A, Robinson MD, Soneson C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res* 2020;**7**:1141.
14. Yu L, Cao Y, Yang JYH, Yang P. Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data. *Genome Biol* 2022;**23**:49.
15. Leary JR, Xu Y, Morrison AB, *et al.* Sub-Cluster Identification through Semi-Supervised Optimization of rare-cell silhouettes (SCISSORS) in single-cell RNA-sequencing. *Bioinformatics* 2023;**39**:btad449.
16. Bacher R, Chu L-F, Argus C, *et al.* Enhancing biological signals and detection rates in single-cell RNA-seq experiments with cDNA library equalization. *Nucleic Acids Res* 2022;**50**:e12.
17. Cao J, Spielmann M, Qiu X, *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 2019;**566**: 496–502.
18. Charrout M, Reinders MJT, Mahfouz A. Untangling biological factors influencing trajectory inference from single cell data. *NAR Genom Bioinform* 2020;**2**:lqaa053.
19. Hallett SA, Ono W, Ono N. The hypertrophic chondrocyte: to be or not to be. *Histol Histopathol* 2021;**36**:1021–36.

20. Long JT, Leinroth A, Liao Y, *et al.* Hypertrophic chondrocytes serve as a reservoir for marrow-associated skeletal stem and progenitor cells, osteoblasts, and adipocytes during skeletal development. *Elife* 2022;**11**:e76932.

21. Leary, J., Bacher, R. Interpretable trajectory inference with single-cell Linear Adaptive Negative-binomial Expression (scLANE) testing. *BioRxiv* 2023;12.

22. Heinonen J, Taipaleenmäki H, Roering P, *et al.* Snorc is a novel cartilage specific small membrane proteoglycan expressed in differentiating and articular chondrocytes. *Osteoarthr Cartil* 2011;**19**:1026–35.

23. Sakata-Goto T, Takahashi K, Kiso H, *et al.* Id2 controls chondrogenesis acting downstream of BMP signaling during maxillary morphogenesis. *Bone* 2012;**50**:69–78.

24. Zhang X, Zhu J, Li Y, *et al.* Epidermal growth factor receptor (EGFR) signaling regulates epiphyseal cartilage development through β-catenin-dependent and -independent pathways. *J Biol Chem* 2013;**288**:32229–40.

25. Johnson EM, Kath W, Mani M. EMBEDR: distinguishing signal from noise in single-cell omics data. *Patterns* 2022;**3**:100443.

26. Neufeld A, Gao LL, Popp J, *et al.* Inference after latent variable estimation for single-cell RNA sequencing data. *Biostatistics* 2022;**25**:270–87.

27. Chari T, Pachter L. The specious art of single-cell genomics. *PLOS Comp Bio* 2023;**19**:e1011288.

28. Lun A. Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data. *BioRxiv* 2018;404962.

29. Tian L, Dong X, Freytag S, *et al.* Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nat Methods* 2019;**16**:479–87.

30. Leek JT, Scharpf RB, Bravo HC, *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010;**11**:733–9.

31. Büttner M, Miao Z, Wolf FA, *et al.* A test metric for assessing single-cell RNA-seq batch correction. *Nat Methods* 2019;**16**:43–9.

32. Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol* 2019;**37**:685–91.

33. Lakkis J, Wang D, Zhang Y, *et al.* A joint deep learning model enables simultaneous batch effect correction, denoising, and clustering in single-cell transcriptomics. *Genome Res* 2021;**31**:1753–66.

34. Wang Y, Huang H, Rudin C, *et al.* Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization. *J Mach Learn Res* 2021;**22**:1–73.

35. Xia L, Lee C, Li JJ. scDEED: a statistical method for detecting dubious 2D single-cell embeddings and optimizing t-SNE and UMAP hyperparameters. *Nature Comm* 2023;**15**:1753.

36. Liu Z. Visualizing single-cell RNA-seq data with semisupervised principal component analysis. *Int J Mol Sci* 2020;**21**:5797.

37. Ehiro T. Feature importance-based interpretation of UMAP-visualized polymer space. *Mol Inform* 2023;**42**:e2300061.

38. Fischer J, Burkholz R, Vreeken J. *Preserving Local Densities in Low-Dimensional Embeddings*. *ArXiv* 2023;**2301**:13732.

39. Xu Y, Zang Z, Xia J, *et al.* Structure-preserving visualization for single-cell RNA-Seq profiles using deep manifold transformation with batch-correction. *Commun Biol* 2023;**6**:1–15.

40. Cheng C, Easton J, Rosencrance C, *et al.* Latent cellular analysis robustly reveals subtle diversity in large-scale single-cell RNA-seq data. *Nucleic Acids Res* 2019;**47**:e143.

41. Kiselev VY, Kirschner K, Schaub MT, *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* 2017;**14**:483–6.

42. Pateiro-López B, Rodríguez-Casal A. Generalizing the convex hull of a sample: the R package alphahull. *J Stat Soft* 2010;**34**:1–28.

43. Hubert M, Vandervieren E. An adjusted boxplot for skewed distributions. *Comput Stat Data Anal* 2008;**52**:5186–201.

44. Baron M, Veres A, Wolock SL, *et al.* A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst* 2016;**3**:346–360.e4.

45. Cannoodt R, Saelens W, Todorov H, *et al. Single-Cell-Omics Datasets Containing a Trajectory*. (Version 2.0.0) [Data set]. Zenodo 2018;1443566. https://doi.org/10.5281/zenodo.1443566.

46. Hao Y, Stuart T, Kowalski MH, *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2024;**42**:293–304.

47. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* 2016;**5**:2122.

48. Cannoodt R, Saelens W, Sichien D, *et al.* SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. *Biorxiv* 2016;079509.

49. Scrucca L, Fraley C, Murphy TB, *et al. Model-Based Clustering, Classification, and Density Estimation Using mclust in R. Chapman and Hall/CRC*, 2023.

50. Xie Z, Bailey A, Kuleshov MV, *et al.* Gene set knowledge discovery with Enrichr. *Curr Protoc* 2021;**1**:e90.

51. Liberzon A, Subramanian A, Pinchback R, *et al.* Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* 2011;**27**:1739–40.

52. Castanza AS, Recla JM, Eby D, *et al.* Extending support for mouse data in the Molecular Signatures Database (MSigDB). *Nat Methods* 2023;**20**:1619–20.

53. Kowalczyk MS, Tirosh I, Heckl D, *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* 2015;**25**:1860–72.

54. Cannoodt R, Saelens W, Deconinck L, Saeys Y. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. *Nat Commun* 2021;**12**:3942.

55. Zappia L, Phipson B, Oshlack A. Splatter: simulation of single-cell RNA sequencing data. *Genome Biol* 2017;**18**:174.

56. Petropoulos S, Sandberg R, Edsgard D, *et al.* Single-cell RNA-seq reveal lineage formation and X-chromosome dosage compensation in human preimplantation embryos. *Cell* 2016;**165**:1012–26.

57. Sloan SA, Darmanis S, Huber N, *et al.* Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron* 2017;**95**:779–790.e6.

58. Qiu W-L, Zhang Y-W, Feng Y, *et al.* Deciphering pancreatic islet β cell and α cell maturation pathways and characteristic features at the single-cell level. *Cell Metab* 2017;**25**:1194–1205.e4.

59. Li L, Dong J, Yan L, *et al.* Single-cell RNA-Seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* 2017;**20**:858–873.e4.