Pseudotime analysis for time-series single-cell sequencing and imaging data

Gang Li[1,2], Hyeon-Jin Kim[1], Sriram Pendyala[1], Ran Zhang[1], Christine M. Disteche[3,4], Jean-Philippe Vert[5], Xinxian Deng[3], Doug Fowler[1], and William Stafford Noble[1,6]

[1]Department of Genome Sciences, University of Washington, Seattle 98115, USA
[2]eScience Institute, University of Washington, Seattle 98115, USA
[3]Department of Laboratory Medicine and Pathology, University of Washington, Seattle 98115, USA
[4]Department of Medicine, University of Washington, Seattle 98115, USA
[5]Owkin, Paris 75010, France
[6]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle 98115, USA

Abstract

Many single-cell RNA-sequencing studies have collected time-series data to investigate transcriptional changes with respect to various notions of biological time, such as cell differentiation, embryonic development, and response to stimulus. Accordingly, several unsupervised and supervised computational methods have been developed to construct single-cell pseudotime embeddings for extracting the temporal order of transcriptional cell states from these time-series scRNA-seq datasets. However, existing methods, such as psupertime, suffer from low predictive accuracy, and this problem becomes even worse when we try to generalize to other data types such as scATAC-seq or microscopy images. To address this problem, we propose Sceptic, a support vector machine model for supervised pseudotime analysis. Whereas psupertime employs a single joint regression model, Sceptic simultaneously trains multiple classifiers with separate score functions for each time point and also allows for non-linear kernel functions. Sceptic first generates a probability vector for each cell and then aims to predict chronological age via conditional expectation. We demonstrate that Sceptic achieves significantly improved prediction power (accuracy improved by $1.4 - 38.9\%$) for six publicly available scRNA-seq data sets over state-of-the-art methods, and we show that Sceptic also works well for single-nucleus image data. Moreover, we observe that the pseudotimes assigned by Sceptic show stronger correlations with nuclear morphology features than the observed time labels, suggesting that these pseudotimes accurately capture the heterogeneity of nuclei derived from a single time point and thus provide more informative time labels than the observed times. Finally, we show that Sceptic accurately captures sex-specific differentiation timing from both scATAC-seq and scRNA-seq data.

1 Introduction

With the decreasing cost of single-cell RNA-sequencing (scRNA-seq), it is increasingly common to collect time-series scRNA-seq data sets. Such data can be used to investigate transcriptional changes over various notions of biological time, such as cell differentiation [2], embryonic development [21] and response to stimulus [29].

However, single-cell pseudotime analysis comes with significant challenges [24]. First, single-cell measurements are usually sparse and high-dimensional [1, 30]. In practice, only a subset of cells may express any given gene, and the capture rate of the genes also varies, which makes pseudotime inference harder. Second, single-cell measurements frequently suffer from technical noise such as batch effects [1, 14]. Without properly capturing such effects, pseudotime analysis might lead to misleading results. Third, cells in a culture or biological system usually grow asynchronously. Consequently, different cells may be in different cell cycle stages, which confounds the inference of the temporal ordering of cells. Fourth, inferring correct trajectories, and especially identifying branch points, can be challenging, especially when dealing with complex branching trajectories. Last, interpreting the biological relevance of pseudotime values can also be complicated. It is essential to have experiments or other evidence to validate the predicted pseudotimes.

Researchers have developed a variety of unsupervised methods to address all or a subset of the challenges above. PCA [18] is commonly regarded as a straightforward and interpretable baseline method for pseudotime inference [16]. Slingshot identifies cell lineages by treating groups of cells as nodes within a graph and identifying a minimum spanning tree connecting these nodes [27]. Monocle 2 [23] and Monocle 3 [5], widely adopted in various studies [24], employ distinct underlying algorithms. Monocle 2 utilizes a reversed graph embedding strategy to model cell trajectories, effectively constructing a minimum spanning tree among cells [23]. Monocle 3 identifies cell trajectories using a single-rooted directed acyclic graph, which captures the hierarchical organization of cell states [5]. Finally, Palantir is designed to model the trajectories of differentiating cells by treating the determination of cell fate as a probabilistic process, using entropy as a key metric to quantify cell plasticity as cells progress [26].

More recently, Macnair *et al.* proposed psupertime, which adopts a supervised learning approach for pseudotime inference. Psupertime uses a penalized ordinal logistical regression model for time-series scRNAseq data [16]. As in several other cases where supervised methods have been used to improve many unsupervised single-cell analyses [32, 33], psupertime has been shown to perform comparably to or better than other methods like PCA [18], Monocle 2 [23], Slingshot [27] and Tempora [28]. Tempora is another supervised method that primarily focuses on delineating how cell types are interrelated throughout the entire time-series dataset [28]. It relies on the assumption that cells with similar gene expression profiles exhibit closer connections in the underlying cell lineage. However, we hypothesized that psupertime's predictive accuracy suffers from its use of a simple linear combination of gene expression profiles. This problem will likely be exacerbated if we were to apply a similar method to more complex data types, such as single-nucleus microscopy images.

Motivated by this hypothesis, we propose Sceptic, a support vector machine (SVM) framework for supervised pseudotime analysis. Sceptic first generates a probability vector for each cell or nucleus and then predicts chronological age via conditional expectation. By using a multi-class SVM, we aim to achieve higher accuracy when classifying cells from each time point. We validated the performance of Sceptic on multiple public single-cell datasets (Figure 1) and further extended it to different modalities of time-series single-cell data, including our own single-nucleus imaging dataset. First, we demonstrated that Sceptic more accurately classifies cells than psupertime on six scRNA-seq datasets. Second, we showed that Sceptic works well for single-nucleus imaging data. Third, we applied Sceptic to scATAC-seq data and showed that the model captures sex-specific differentiation from both scATAC-seq and scRNA-seq data. Finally, we applied Sceptic to a co-assay data set, where we detected methylation delay that agrees with other independent studies. Sceptic code can be found on github: `https://github.com/Noble-Lab/Sceptic`.

2 Results

2.1 Overview of Sceptic, Single CEll PseudoTIme Classifier

Sceptic can perform pseudotime analysis on various types of single-cell/single-nucleus data. The model takes as input a collection of single-cell/single-nucleus data and then learns the relationship between the observed

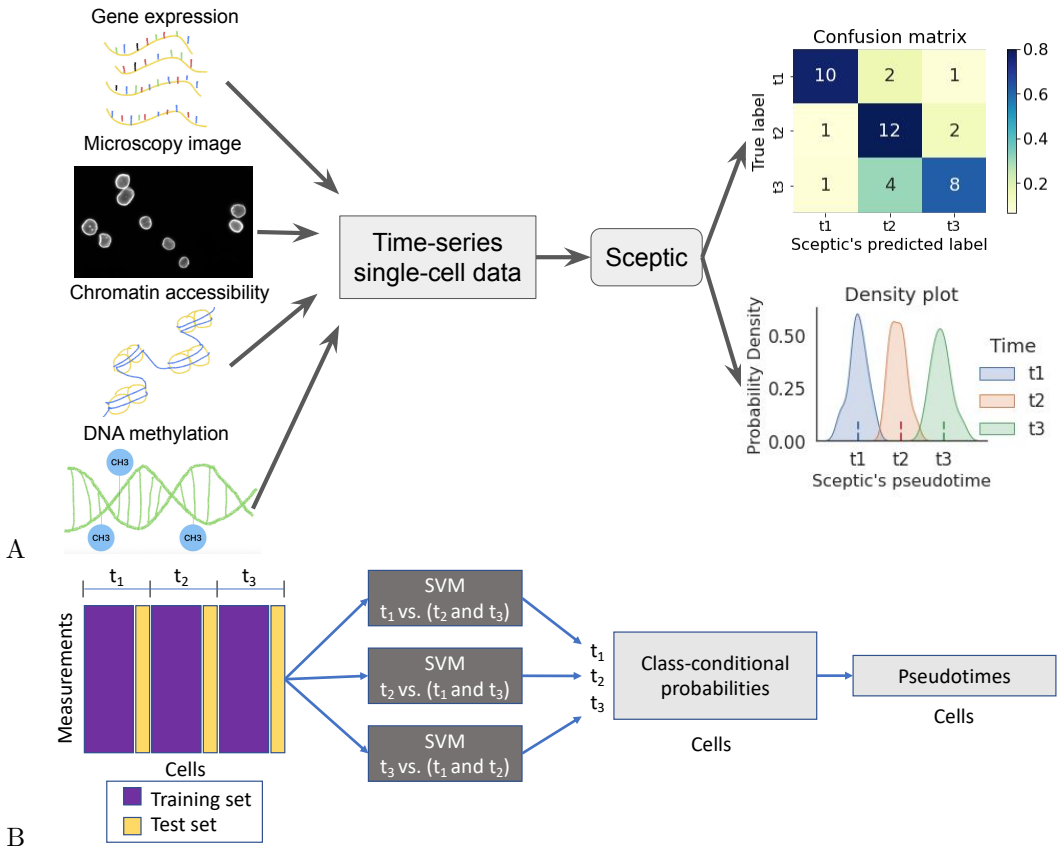

Figure 1: **Sceptic analysis for time-series single-cell sequencing and imaging data. A** Sceptic from the user's perspective. Sceptic can take various types of single-cell/single-nucleus data as input and is trained to predict each cell's time label (shown as a confusion matrix) and to infer the pseudotime for each cell (shown as a density plot across all cells). **B** Overview of Sceptic analysis.

data and the associated time stamps, and finally uses the trained model to assign to each cell a real-valued pseudotime. Ideally, the pseudotimes assigned by Sceptic reflect each cell's progression along a notion of time—developmental, cell cycle, disease progression, aging—that is appropriate to the given data.

Sceptic differs from its predecessor, psupertime [16], in three ways (Figure 1). First, Sceptic uses a nonlinear support vector machine rather than an ordinal logistic regression model. Second, and more importantly, whereas psupertime trains a single regressor with multiple thresholds, Sceptic trains a collection of classifiers, one for each timestamp, using a one-versus-the-rest strategy. The final pseudotime assigned to a given cell is computed as a weighted sum of the different predictions. This setup significantly enhances the classifier's performance. Third, Sceptic employs a cross-validation strategy, in which a collection of models is trained on different subsets of the data and used to make predictions on disjoint test sets. This strategy prevents the model from reporting pseudotime values that may overfit the data.

2.2 Sceptic outperforms psupertime on classification for time-series single-cell RNA-sequencing data

We designed Sceptic to provide a flexible framework for modeling pseudotime trends within a collection of single-cell data. Accordingly, we set out to test the hypothesis that Sceptic could more accurately predict the timestamp associated with a given cell than psupertime. To do this, we used our recently published scRNA-seq data from a mouse embryonic stem cell (mESC) differentiating time-series dataset, collected over five time points: days 0, 3, 7, 11 and neural progenitor cells collected at day 21 [2]. We used five-fold cross-validation, training a model on 80% of the cells and testing on the remaining 20%, and repeating this

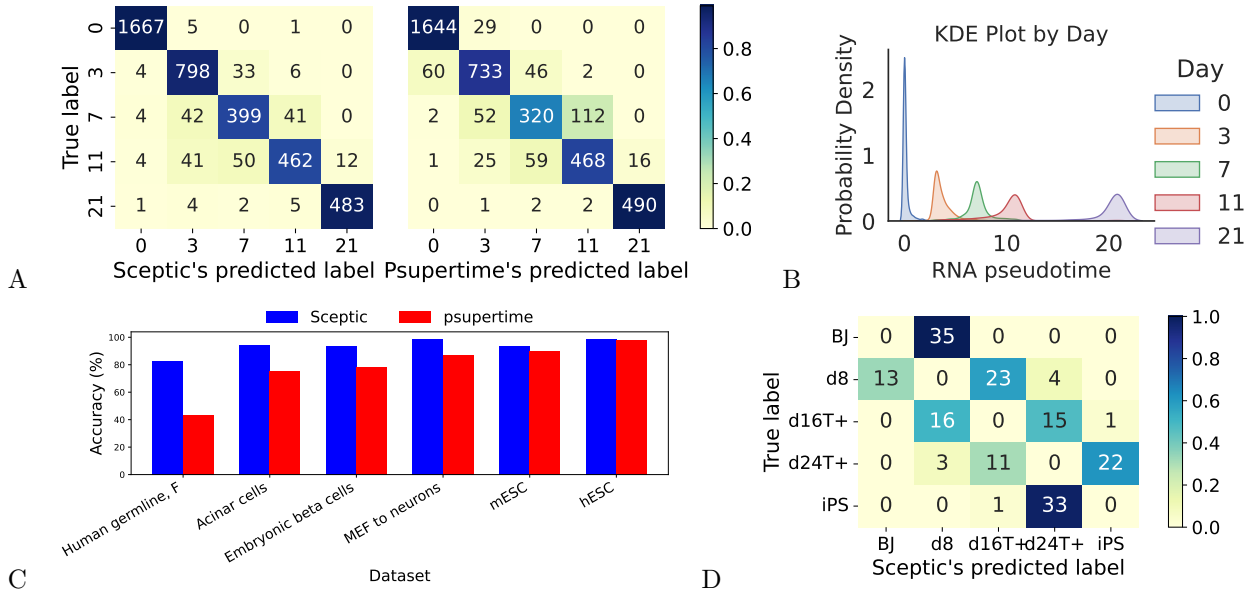

Figure 2: **Sceptic achieves better classification accuracy than psupertime on single-cell RNA sequencing data.** **A** Confusion matrices for Sceptic (left) and psupertime (right) on the mESC data. Sceptic results use cross-validation, whereas psupertime results are from training on the full dataset. **B** Kernel density plot of Sceptic pseudotime, colored by the day of cells, on the mESC data. **C** Accuracy of both methods on six time-series datasets. **D** Confusion matrix from leave-one-time-point-out analysis. A successful model would tend to predict time points adjacent to the diagonal.

procedure five times. For comparison, we also applied psupertime to the same dataset. Because psupertime is not designed to generalize to new cells, we are only able to report psupertime's performance when fitting a single model to the entire dataset. In this sense, the comparison is biased in favor of psupertime.

Despite this bias, the predictions from the two models, summarized in confusion matrices (Figure 2A), show that Sceptic outperforms psupertime. In particular, Sceptic achieves an accuracy of 93.73% (because 3,809 of the 4,064 predictions fall along the diagonal of the confusion matrix), whereas psupertime's accuracy is only 89.94% (3,655 correct predictions). This difference is significant ($p = 4.94e^{-10}$) according to Fisher's exact test. The confusion matrices indicate which timestamps are difficult to differentiate, with most of the off-diagonal predictions occurring between days 3 and 7 and between days 7 and 11. Sceptic achieves a clear separation between cells of day 7 and 11 (Figure 2B), whereas psupertime does not perform very well on these cells (Figure 2A).

To further test our hypothesis, we downloaded the five datasets that were analyzed in the original psupertime paper [16], and we repeated the above analysis. In this case, for psupertime we use the previously reported accuracy values. Again, despite the bias in favor of psupertime in this setup, on all five datasets Sceptic achieves consistently higher classification accuracy than psupertime (Figure 2C), improving by 1.4%-39.0%. The difference is significant ($p = 0.03125$) according to a Wilcoxon signed-rank test. Note that psupertime has previously been shown to perform comparably to or better than other methods like PCA [18], Monocle 2 [23], Slingshot [27] and Tempora [28]. Thus, we only compare our performance with psupertime.

We note that, in any evaluation of a supervised pseudotime inference method, we face the fundamental challenge of distinguishing between batch effects and true pseudotemporal ordering. To understand the problem, consider a theoretical dataset in which each time point is coupled with a very strong, time-specific batch effect. In such a setting, a method that successfully learns to identify this batch effect will perform perfectly in a confusion matrix, despite not having learned anything about the temporal signal in the data. Therefore, additional validation is required to demonstrate that the pseudotemporal ordering is meaningful.

To address this concern, we repeated our analysis but used a cross-validation strategy in which we iteratively left one time point out of the training set on scGEM dataset (see more details of this data in Section 2.5). In this setting, the resulting confusion matrix (Figure 2C) necessarily has zeroes on the diagonal.

Our performance measure is the proportion of values that fall in boxes that are adjacent to the diagonal. The reasoning here is that if the model learned only batch effects, then the errors would be uniformly distributed across all time points. On the other hand, a model that successfully learns temporal patterns will, when forced to predict a time for a novel time point, tend to predict adjacent time points. In our case, we find that 94.9% of the predictions in the leave-one-time-point-out setting fall into adjacent time points. This is substantially greater than the 40.0% that would be expected if the errors were uniformly distributed over time points.

2.3 Sceptic works well for single-nucleus imaging data

Having established that Sceptic's machine learning framework works well for scRNA-seq data, we next investigated whether a similar approach would generalize to single-nucleus microscopy imaging data. Accordingly, we collected F121-6 mESC differentiation 3D images (DNA staining with DAPI and immunostaining of nucleolin) using the same five time points as described in our recent study [2] (Figure 3A). We segmented each 2D-maximum-projection image using a pre-trained Mask R-CNN model [3, 11], and we used a pre-trained ResNet34 model to generate 1024-dimensional representations for each nucleus (Figure 3B). To measure the performance of the classifier, we randomly selected 1,200 single-nucleus image crops from each time point, yielding 6,000 crops in total.

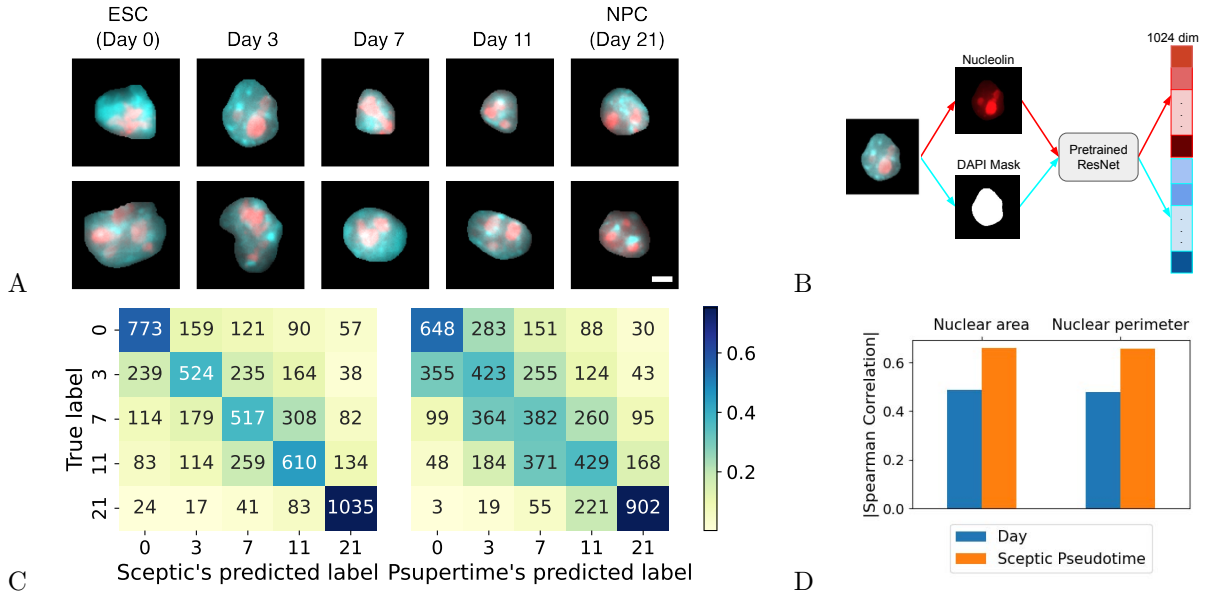

Figure 3: **Sceptic works well for single-nucleus imaging data. A** Examples of single nucleus images. Blue = DAPI; red = nucleolin. Scale bar is 10 μm. ESC: embryonic stem cells. NPC: neural progenitor cells. **B** Image processing pipeline. **C** Sceptic achieves better classification accuracy than psupertime for time-series single-nucleus image data. **D** Pseudotimes inferred by Sceptic are more highly correlated with nuclear morphology features than the observed time label.

Our results show that Sceptic is able to accurately assign time labels to nuclei on the basis of microscopy images, significantly outperforming psupertime on this task (Figure 3C). Overall, Sceptic achieves 57.7% test set accuracy, whereas psupertime achieves a significantly lower 46.4% training set accuracy ($p = 1.08e^{-8}$), Fisher's exact test). For both methods, the overall accuracy is lower in the microscopy setting than in the scRNA-seq setting. As expected, Sceptic achieves the best separation between the ends of the time series—day 0 versus day 21.

Next we investigated whether the time labels assigned by Sceptic show stronger correlations with cell morphology features than the observed time labels. During embryonic stem cell differentiation, neural progenitor cells (day 21) exhibit reduced nucleolus activity, reduced transcription of rRNA genes, and smaller nucleoli [10]. Thus, we expect an accurate time label to be strongly correlated with these morphology

features. We therefore compared the correlations between pseudotime or observed time with respect to two features: nuclear area, and nuclear perimeter. Compared to the observed time points, Sceptic assigns much finer-grained continuous time labels, and thus achieves a higher absolute correlation with nuclear image morphology features compared with the discrete-time labels (Figure 3D). These results suggest that the pseudotime values inferred by Sceptic accurately capture the heterogeneity of nuclei sampled at the same time point and thus provide more informative time labels than the observed ones.

2.4 Sceptic detects sex-specific differentiation from scATAC-seq and scRNA-seq data

Having established that Sceptic can work well for both scRNA-seq and single-nucleus imaging data, we next investigated whether a similar approach could generalize to single-cell transposase-accessible chromatin with sequencing (scATAC-seq) data. This assay identifies, in a cell-specific fashion, punctate regions of open chromatin, which typically correspond to various types of regulatory elements, including promoters, enhancers, and insulators. To test Sceptic, we downloaded F121-6 (female) and F123 (male) mESC differentiation scATAC-seq data from Bonora *et al.* [2]. We randomly downsampled the number of cells from each time point to the smallest cell number across all time points, which left us with 535 female cells (107 cells for each time point) and 560 male cells (140 cells for each time point). Note that the F123 male dataset does not contain neural progenitor cells for day 21. We trained two separate Sceptic models for each dataset, one for each sex.

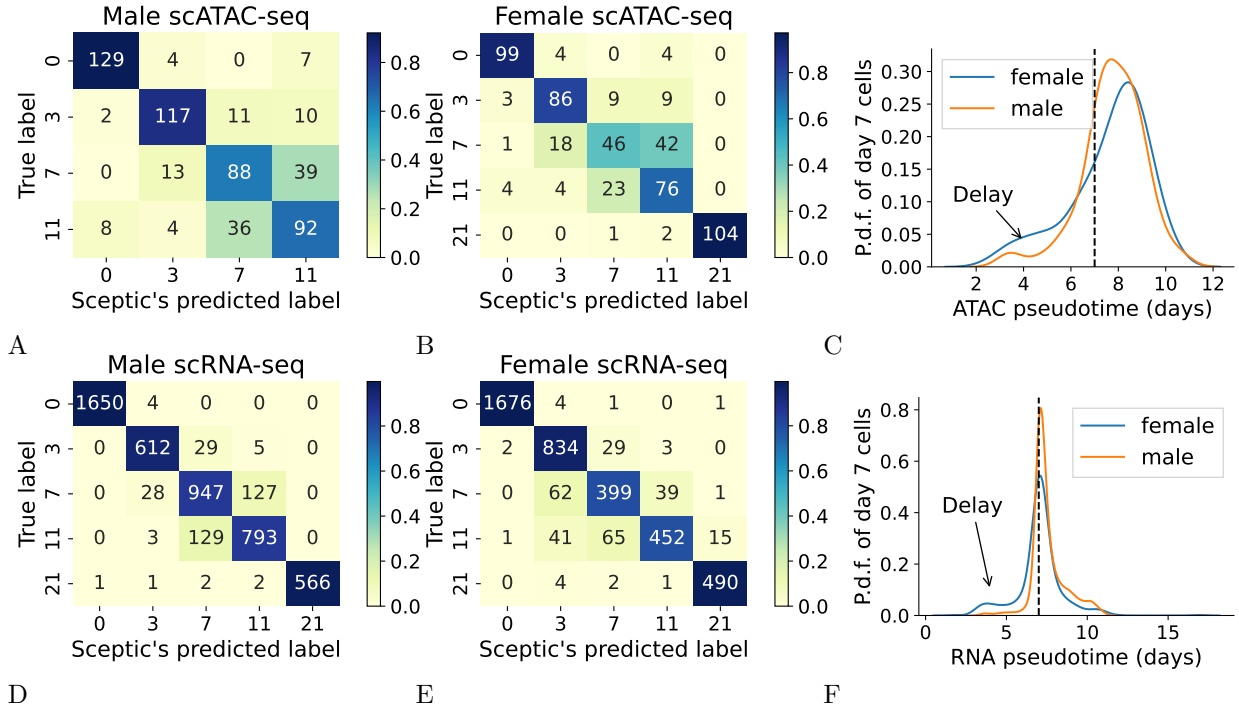

Figure 4: **Sceptic applied to scATAC-seq and scRNA-seq mESC data. A** Sceptic's performance for scATAC-seq data of F123 (male) cells. **B** Sceptic's performance for scATAC-seq data of F121-6 (female) cells. **C** Probability density function of Sceptic's pseudotime for day 7 cells of scATAC-seq data. **D** Sceptic's performance for scRNA-seq data of F123 (male) cells. **E** Sceptic's performance for scRNA-seq data of F121-6 (female) cells. **F** Probability density function of Sceptic's pseudotime for day 7 cells of scRNA-seq data.

In this experiment, Sceptic achieves 76.8% test set accuracy on F121-6 female cells and 76.1% on F123 male cells (Figure 4A–B). We note that Sceptic accurately classifies cells from day 0 and day 3 for both male and female cells, whereas, on both datasets, cells from days 7 and 11 are difficult to distinguish. Cells from day 7 and day 11 share similar chromatin accessibility patterns, and they construct a single cell trajectory

branch (embryonic body branch) in Bonora *et al.* 's analyses. In addition, we observed that for day 7, female cells have a slightly higher proportion of cells classified as earlier time points (day 0 and day 3) than their male counterparts (Figure 4A–C): 17.8% (19 out of 107 cells) for female and 9.3% (13 out of 140 cells) for male. Similar differences in proportion are observed for cells in day 11 (34.3% for female cells and 29.0% for male cells). These differences are potentially indicative of female-delayed onset of differentiation and are consistent with previous observations [25, 9, 2], in which the early establishment of X chromosome inactivation is required for differentiation.

To follow up on this observation, we next use Sceptic to infer pseudotime from scRNA-seq data generated from the same series of mESC samples [2]. For this data, Sceptic achieves an accuracy of 93.2% for male and 93.4% for female across five time points (Figure 4D–E). As in the ATAC-seq setting, we found for day 7, that female cells have a higher proportion of cells classified as earlier time points (12.4%; 62 out of 500 cells) than their male counterparts (2.5%; 28 out of 1102 cells) (Figure 4D–F), and similar proportions difference are observed for cells in day 11 (18.6% for female cells and 14.3% for male cells). This is consistent with our observation on scATAC-seq, but we found the transcript delay pattern is much stronger than the delay of chromatin accessibility.

2.5 Sceptic identifies methylation delay using co-assay data.

We also applied our Sceptic model to a co-assay data, sc-GEM, a protocol that simultaneously measure pre-selected DNA methylation and gene expression markers in single cells via polymerase chain reaction (PCR) [6]. Cheow *et al.* performed sc-GEM on human fibroblasts undergoing induced pluripotent stem (iPS) cell reprogramming. This dataset contains 177 cells from five time points, BJ (day 0; fibroblasts established from skin taken from normal foreskin from a neonatal male), day 8, day 16, day 24 and iPS.

For this data set, Sceptic achieves 92.7% test set accuracy using gene expression data but only 63.8% using DNA methylation data (Figure 5). In particular, using methylation data, Sceptic can accurately distinguish among the later three time points well but struggles to distinguish between day 0 and day 8 cells. This observation suggests that there are enough transcriptomic changes within eight days of the reprogramming process to distinguish untreated cells and day 8 cells, but it takes longer than eight days to distinguish DNA methylation state changes. Sceptic's results are consistent with a previous study, which observed that most of the DNA methylation changes in iPS reprogramming only occur after day 9 [20].

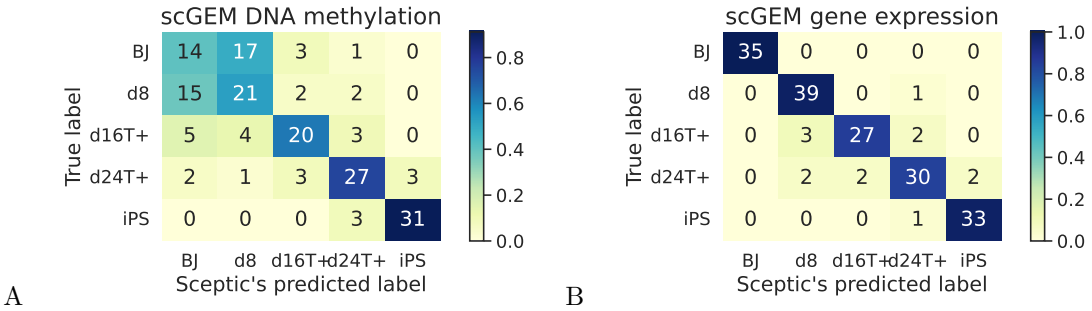

Figure 5: **Sceptic captures methylation delay. A** Sceptic can also separate cells using DNA methylation data, albeit with some ambiguity between day 0 (BJ) and day 8. **B** Sceptic can accurately distinguish cells from 5 times points using gene expression data.

Finally, we explored the concordance between pseudotimes inferred from different modalities. To investigate this, we performed a two-way analysis of variance (ANOVA) to evaluate the influence of RNA-inferred pseudotime and time on methylation pseudotime. Notably, the analysis revealed a significant impact of RNA pseudotime on the variation in methylation pseudotime (P-value = 4.20e-11). This finding implies that the pseudotimes generated by two Sceptic models capture common underlying structures within the same cells, highlighting their shared aspects.

3 Discussion

In this work, we developed a method, Sceptic, to infer pseudotime for single-cell sequencing and imaging data using a supervised machine-learning approach. Our experiments suggest that Sceptic offers improved accuracy relative to the previous supervised method, psupertime. Furthermore, to our knowledge, Sceptic is the first supervised computational approach that infers pseudotime using time-series single-cell ATAC-seq and single nucleus imaging data. Our extensive evaluation demonstrates that Sceptic effectively infers pseudotime for time-series sequencing and imaging data. Application of Scetpic to multiple single-cell RNA-sequencing data has shown uniformly improved performance compared with the state-of-the-art methods. The extension to scATAC-seq and single nucleus imaging is straightforward and effective, and it successfully captures the cells in the same cell trajectory but from different time points and the cell heterogeneity within the same time points, providing a better time label for each cell.

There are several potential avenues for improvements to Sceptic. First, whereas psupertime's linear model is relatively easy to interpret, Sceptic's non-linear kernel functions make such interpretation more challenging. Intuitively, for more complex data structures like imaging, it is natural to use non-linear decision boundaries for high-dimensional classification problems. But it would still be interesting to find which genes are significantly important for Sceptic's classifiers. Permutation tests could be an effective option when the number of genes and the number of cells are not too large.

Second, Sceptic could be made more scalable. In this work, we focused mainly on datasets of intermediate to small size. With the decreasing costs of sequencing, datasets containing millions of cells are becoming more common. Thus, future work could explore the scalability of the Sceptic model for application to such datasets. The current SVM implementation in Sceptic will likely be impractical beyond tens of thousands of cells. For large datasets, LinearSVC or SGD Classifier might be feasible. With enough cells to train, one could also take advantage of recently developed deep learning techniques to potentially further improve the classification performance. Note that a simple neural network did not work well for our mESC data, and we speculated that the main reason was a lack of training data.

Third, our current Sceptic model has been shown to work well for a simple cell differentiation trajectory or reprogramming process. However, one might consider how to compare pseudotimes across multiple concurrent lineages or cell trajectories.

Finally, when the sampling time points are very dense (say, sequencing for every six hours) and the number of time points is large (say, more than 40), then one might also consider formulating the pseudotime analysis problem directly as a regression problem. Along these lines, Calderon *et al.* adopted a neural network as a regression tool to infer pseudotime for *Drosophila* embryonic development [4]. Because their dataset was collected at overlapping time windows and we mainly focused on classification evaluation, we did not evaluate this dataset. It would be great to combine both the classification and regression nature of pseudotime analysis into a single loss function.

4 Methods

4.1 Single CEll PseudoTIme Classifier (Sceptic)

Sceptic is an SVM-based model for pseudotime analysis. It takes as input a single-cell measurement matrix X, with M rows of measurements and N columns corresponding to cells, and a corresponding time label vector $\vec{t} = (t_1, t_2, ..., t_N)$. We assume that the cells are collected from D different observed time points, i.e., $t_i \in \{T_1, T_2, ..., T_D\}$. Sceptic trains one support vector machine classifier for each timestamp, optimizing a coefficient vector $\mathbf{a} = (a_1, a_2, ..., a_N)$ to minimize the following loss:

$$
\begin{aligned}
L(\mathbf{a}) &= \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{n=1}^{N} \sum_{j=1}^{N} a_i a_j t_i t_j K\left(\mathbf{x}_i, \mathbf{x}_j\right) \\
\text{subject to } 0 \leq a_i &\leq C, \quad i = 1, \ldots, N \\
\sum_{i=1}^{N} a_i t_i &= 0.
\end{aligned}
\tag{1}
$$

where $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel function. Sceptic default uses a radial basis function (RBF) kernel $K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right)$ but also allows other kernels such as a linear kernel $K(x_i, x_j) = x_i^T x_j$. Once fitted, Sceptic estimates a class membership probability $\mathbf{P} = (p_1, p_2, ..., p_D)$ where p_d is the probability that the cell is collected from time point T_d based on Wu *et al.* [31]. The pairwise estimated probabilities between any pair of classes are first calibrated using Platt scaling and then used to solve a quadratic optimization where the solution is the class membership probability. Finally, the predicted pesudotime Y for cell x is given by

$$E(Y|X = x) = \sum_{d=1}^{D} Prob(\text{cell collected from } T_d|X = x) \times T_d.$$

Nested cross validation To select hyperparameters while avoiding overfitting, we train Sceptic using a nested cross-validation scheme. Cross-validation involves randomly splitting the data into, say, five equal-sized folds, and then iteratively using four folds to train the model and the remaining fold for testing. Nested cross-validation repeats another cross-validation splitting procedure on each training set. We use the internal 4-fold cross-validation step to select a set of optimal hyperparameters for each training set, and these hyperparameters are then used to train a model on the full training set. The trained model is then used to make predictions on the test set. We repeat this process five times as the external 5-fold cross-validation step to make predictions on all the data.

Hyperparameter search In the inner loop of the nested cross-validation, Sceptic performs a grid search over three hyperparameters. The first hyperparameter R is a Boolean indicating whether the SVM uses a linear or RBF kernel. The second hyperparameter C controls the relative weight assigned to violations of the SVM soft margin. The third hyperparameter γ specifies the inverse of the width of the RBF kernel. We adopt sklearn's default γ for support vector classifier: $\gamma = 1/(M * Var(vector(X))$, which adjusts the number of features M and the variability of the measurement matrix via the variance of the measurement matrix X in vector format. Sceptic employs a six-valued grid with dimensions $R \in \{0, 1\}$, $C \in \{0.1, 1, 10\}$.

4.2 Psupertime

For five datasets that psupertime evaluated, we directly compared Sceptic's results with the reported performance in the psupertime paper. For mESC data, we follow their tutorial using the default setting to select the hyperparameter λ for the weight of the L1 regularization penalty, selecting the $\lambda.1se$ with performance within one standard error of the best performance $\lambda.min$, and then re-training the model using all the cells.

4.3 Imaging data

Sample Collection F121-6 mouse embryonic stem cells (ESCs) were grown and differentiated into embryoid bodies (EBs) as described in Bonora *et al.* [2]. Neural progenitor cells (NPCs) were derived from EBs at day 11 and collected at day 21. ESCs and EBs at days 3, 7, and 11, as well as NPCs were collected and frozen in 10% DMSO in 2–3 million cell aliquots.

Immunostaining Frozen cell aliquots were thawed at 37 C and resuspended in 1x DPBS. Approximately 250,000 cells were plated on a 24-well glass bottom, black-walled plates (CellVis, Mountain View, CA, P24-1.5H-N) via centrifugation at 300g for 5 mins. The cells were then fixed with 4% paraformaldehyde (Electron Microscopy Sciences Cat. No. 50-980-492) for 15 m at room temperature. After fixation, they were washed three times with 1x PBS and incubated in blocking buffer (3% BSA and 0.25% Tx100 in 1x PBS) for 1 hour at room temperature. They were then incubated with primary antibodies (Polyclonal nucleolin antibodies, 1:500, Invitrogen Cat. No.PA5-85972) diluted in blocking buffer for an hour at room temperature. The cells were washed three times with 1x PBS and incubated with secondary antibodies (Donkey anti rabbit Alexa 647, Invitrogen, Cat. No. A-31573) and DAPI diluted in blocking buffer for an hour at room temperature. They were then washed three times with 1x PBS and stored in 600 μL of 1x PBS.

Image acquisition Image acquisition was performed with an Automated Leica DMi8 inverted microscope with Adaptive Focus technology configured with a 40X 0.95 NA objective. The microscope was illuminated using a six-line Lumencor Spectra X Light Engine LED with Semrock multi-band dichroic filters (Spectra Services, Ontario, NY). The images were captured with a Zyla 4.2 sCMOS camera (Andor, Windsor, CT, USA) and the Piezo-driven stage (Okolab) was used to acquire z-stacks.

Image segmentation Nuclear segmentation was performed with a pre-trained Mask R-CNN model [3, 11] with a nuclear size cutoff of 900. Each nucleus was padded and cropped to a 128×128 pixel image.

Image embeddings For each single-nucleus image crop (128×128) of mESC dataset that we collected, we use pre-trained (on ImageNet [13]) Resnet34 [12](pytorch/vision:v0.10.0) from Pytorch [17] to generate a less-noisy vector-representation of the image. We generate 512-dim embeddings from the nucleolin-tagged image and 512-dim embeddings from the DAPI mask image. By concatenating them together, we generated a 1024-dim representation for each image crop.

4.4 Processing scATAC-seq data

We follow the epiScanpy pipeline to process the mESC scATAC-seq data [7]. Briefly, we first binarized the cell by peak matrix and identified the most variable peaks. The standard library size normalization and log normalization are performed. Finally, we use the top 50 principal components from normalized data as the input for our Sceptic model.

4.5 scRNA-seq data

Datasets	Type	Time Label	# time points	# cells	Accession
mESC [2]	sciRNA	Embryonic day	5	4064	GSE184554
hESC [19]	scRNA	Embryonic day	5	1529	E-MTAB-3929
Human germline [15]	scRNA	Age (weeks)	12	992	GSE86146
Embryonic beta cells [22]	scRNA	Developmental stage	7	575	GSE87375
Acinar cells [8]	scRNA	Donor age	8	411	GSE81547
MEF to neurons [29]	scRNA	Days since induction	5	315	GSE67310

Table 1: **Time-series scRNA-seq datasets.**

We used six publicly available scRNA-seq data sets (Table 1). We downloaded the five scRNA-seq datasets that were previously analyzed by Manair *et al.* from the R package "psupplementary" (`https://github.com/wmacnair/psupplementary`). We also downloaded the mESC scRNA-seq and scATAC-seq datasets analyzed by Bonora *et al.* [2] from bitbucket repository (`https://bitbucket.org/noblelab/mouse-sci-omics/src/master/`). We downloaded the processed scGEM dataset from UnionCom's GitHub page `https://github.com/caokai1073/UnionCom/tree/master/scGEM`. The image data are available upon request and in the process of submission to 4D Nucleome Consortium data portal.

Software availability Sceptic code is available with an Apache license at `https://github.com/Noble-Lab/Sceptic`.

Acknowledgments The authors would like to thank Yang Lu, Anupama Jha, and Laetitia Meng-Papaxanthos (Google Research, Switzerland) for their suggestions and comments. This work was supported by NIH award UM1 HG011531. G.L.'s research is supported by the eScience Institute at the University of Washington.

Author contributions G.L.: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Visualization, Writing—Original draft preparation, and Software. H.K., S.P.: Formal analysis, Image collection, Writing—Original draft preparation. R.Z.: Methodology. J.-P.V.: Supervision, Methodology, and Writing—Review and Editing. D.F., C.D.: Supervision, and Writing—Review and Editing. X.D.: ESC Differentiation, Cell Collection, Supervision, and Writing—Review and Editing. W.S.N.: Conceptualization, Supervision, Methodology, Writing—Review and Editing, and Funding Acquisition.

References

[1] Ricard Argelaguet, Anna SE Cuomo, Oliver Stegle, and John C Marioni. Computational principles and challenges in single-cell data integration. *Nature biotechnology*, 39(10):1202–1215, 2021.

[2] Giancarlo Bonora, Vijay Ramani, Ritambhara Singh, He Fang, Dana L Jackson, Sanjay Srivatsan, Ruolan Qiu, Choli Lee, Cole Trapnell, Jay Shendure, et al. Single-cell landscape of nuclear configuration and gene expression during stem cell differentiation and x inactivation. *Genome Biology*, 22(1):1–36, 2021.

[3] Juan C Caicedo, Allen Goodman, Kyle W Karhohs, Beth A Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, et al. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature methods*, 16(12):1247–1253, 2019.

[4] Diego Calderon, Ronnie Blecher-Gonen, Xingfan Huang, Stefano Secchia, James Kentro, Riza M Daza, Beth Martin, Alessandro Dulja, Christoph Schaub, Cole Trapnell, et al. The continuum of drosophila embryonic development at single-cell resolution. *Science*, 377(6606):eabn5800, 2022.

[5] J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, and J. Shendure. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745):496–502, 2019.

[6] L. Cheow, E. T. Courtois, Y. Tan, R. Viswanathan, Q. Xing, R. Tan, D. S. Tan, R. Robson, Y. Loh, S. R. Quake, et al. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nature methods*, 13(10):833–836, 2016.

[7] Anna Danese, Maria L Richter, Kridsadakorn Chaichoompu, David S Fischer, Fabian J Theis, and Maria Colomé-Tatché. Episcanpy: integrated single-cell epigenomic analysis. *Nature Communications*, 12(1):5228, 2021.

[8] Martin Enge, H Efsun Arda, Marco Mignardi, John Beausang, Rita Bottino, Seung K Kim, and Stephen R Quake. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell*, 171(2):321–330, 2017.

[9] Oriana Genolet, Anna A Monaco, Ilona Dunkel, Michael Boettcher, and Edda G Schulz. Identification of x-chromosomal genes that drive sex differences in embryonic stem cells through a hierarchical crispr screening approach. *Genome Biology*, 22(1):1–41, 2021.

[10] Shivani Gupta and Raffaella Santoro. Regulation and roles of the nucleolus in embryonic stem cells: from ribosome biogenesis to genome organization. *Stem Cell Reports*, 15(6):1206–1219, 2020.

[11] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[14] Gang Li, Yuchen Yang, Eric Van Buren, and Yun Li. Dropout imputation and batch effect correction for single-cell rna sequencing data. *Journal of Bio-X Research*, 2(04):169–177, 2019.

[15] Li Li, Ji Dong, Liying Yan, Jun Yong, Xixi Liu, Yuqiong Hu, Xiaoying Fan, Xinglong Wu, Hongshan Guo, Xiaoye Wang, et al. Single-cell rna-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell stem cell*, 20(6):858–873, 2017.

[16] Will Macnair, Revant Gupta, and Manfred Claassen. psupertime: supervised pseudotime analysis for time-series single-cell rna-seq data. *Bioinformatics*, 38(Supplement_1):i290–i298, 2022.

[17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., Vancouver, Canada, 2019.

[18] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

[19] Sophie Petropoulos, Daniel Edsgärd, Björn Reinius, Qiaolin Deng, Sarita Pauliina Panula, Simone Codeluppi, Alvaro Plaza Reyes, Sten Linnarsson, Rickard Sandberg, and Fredrik Lanner. Single-cell rna-seq reveals lineage and x chromosome dynamics in human preimplantation embryos. *Cell*, 165(4):1012–1026, 2016.

[20] Jose M Polo, Endre Anderssen, Ryan M Walsh, Benjamin A Schwarz, Christian M Nefzger, Sue Mei Lim, Marti Borkent, Effie Apostolou, Sara Alaei, Jennifer Cloutier, et al. A molecular roadmap of reprogramming somatic cells into ips cells. *Cell*, 151(7):1617–1632, 2012.

[21] Chengxiang Qiu, Junyue Cao, Beth K Martin, Tony Li, Ian C Welsh, Sanjay Srivatsan, Xingfan Huang, Diego Calderon, William Stafford Noble, Christine M Disteche, et al. Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nature genetics*, 54(3):328–341, 2022.

[22] Wei-Lin Qiu, Yu-Wei Zhang, Ye Feng, Lin-Chen Li, Liu Yang, and Cheng-Ran Xu. Deciphering pancreatic islet β cell and α cell maturation pathways and characteristic features at the single-cell level. *Cell metabolism*, 25(5):1194–1205, 2017.

[23] Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A Pliner, and Cole Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nature methods*, 14(10):979–982, 2017.

[24] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A comparison of single-cell trajectory inference methods. *Nature biotechnology*, 37(5):547–554, 2019.

[25] Edda G Schulz, Johannes Meisig, Tomonori Nakamura, Ikuhiro Okamoto, Anja Sieber, Christel Picard, Maud Borensztein, Mitinori Saitou, Nils Blüthgen, and Edith Heard. The two active x chromosomes in female escs block exit from the pluripotent state by modulating the esc signaling network. *Cell stem cell*, 14(2):203–216, 2014.

[26] Manu Setty, Vaidotas Kiseliovas, Jacob Levine, Adam Gayoso, Linas Mazutis, and Dana Pe'Er. Characterization of cell fate probabilities in single-cell data with palantir. *Nature biotechnology*, 37(4):451–460, 2019.

[27] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19:1–16, 2018.

[28] Thinh N Tran and Gary D Bader. Tempora: cell trajectory inference using time-series single-cell rna sequencing data. *PLoS computational biology*, 16(9):e1008205, 2020.

[29] Barbara Treutlein, Qian Yi Lee, J Gray Camp, Moritz Mall, Winston Koh, Seyed Ali Mohammad Shariati, Sopheak Sim, Norma F Neff, Jan M Skotheim, Marius Wernig, et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell rna-seq. *Nature*, 534(7607):391–395, 2016.

[30] Di Wu, Deepti S Karhade, Malvika Pillai, Min-Zhi Jiang, Le Huang, Gang Li, Hunyong Cho, Jeff Roach, Yun Li, and Kimon Divaris. Machine learning and deep learning in genetics and genomics. *Machine Learning in Dentistry*, pages 163–181, 2021.

[31] T. Wu, C. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. In S. Thrun, L. K. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16, pages 529–532, Cambridge, MA, 2003. MIT Press.

[32] Yuchen Yang, Gang Li, Huijun Qian, Kirk C Wilhelmsen, Yin Shen, and Yun Li. Smnn: batch effect correction for single-cell rna-seq data via supervised mutual nearest neighbor detection. *Briefings in bioinformatics*, 22(3):bbaa097, 2021.

[33] Yuchen Yang, Gang Li, Yifang Xie, Li Wang, Taylor M Lagler, Yingxi Yang, Jiandong Liu, Li Qian, and Yun Li. ismnn: batch effect correction for single-cell rna-seq data via iterative supervised mutual nearest neighbor refinement. *Briefings in Bioinformatics*, 22(5):bbab122, 2021.