An *et al. BMC Genomics* 2019, **20**(Suppl 2):224
https://doi.org/10.1186/s12864-019-5477-8

BMC Genomics

RESEARCH Open Access

TSEE: an elastic embedding method to visualize the dynamic gene expression patterns of time series single-cell RNA sequencing data

Shaokun An[1,3], Liang Ma[2*] and Lin Wan[1,3*]

From The 17th Asia Pacific Bioinformatics Conference (APBC 2019)
Wuhan, China. 14–16 January 2019

Abstract

Background: Time series single-cell RNA sequencing (scRNA-seq) data are emerging. However, the analysis of time series scRNA-seq data could be compromised by 1) distortion created by assorted sources of data collection and generation across time samples and 2) inheritance of cell-to-cell variations by stochastic dynamic patterns of gene expression. This calls for the development of an algorithm able to visualize time series scRNA-seq data in order to reveal latent structures and uncover dynamic transition processes.

Results: In this study, we propose an algorithm, termed time series elastic embedding (TSEE), by incorporating experimental temporal information into the elastic embedding (EE) method, in order to visualize time series scRNA-seq data. TSEE extends the EE algorithm by penalizing the proximal placement of latent points that correspond to data points otherwise separated by experimental time intervals. TSEE is herein used to visualize time series scRNA-seq datasets of embryonic developmental processed in human and zebrafish. We demonstrate that TSEE outperforms existing methods (e.g. PCA, tSNE and EE) in preserving local and global structures as well as enhancing the temporal resolution of samples. Meanwhile, TSEE reveals the dynamic oscillation patterns of gene expression waves during zebrafish embryogenesis.

Conclusions: TSEE can efficiently visualize time series scRNA-seq data by diluting the distortions of assorted sources of data variation across time stages and achieve the temporal resolution enhancement by preserving temporal order and structure. TSEE uncovers the subtle dynamic structures of gene expression patterns, facilitating further downstream dynamic modeling and analysis of gene expression processes. The computational framework of TSEE is generalizable by allowing the incorporation of other sources of information.

Keywords: Nonlinear dimensionality reduction, Elastic embedding, Visualization, Single-cell RNA sequencing, Time series, Cell fate decisions, Gene expression pattern, Oscillation, In-group proportion

*Correspondence: mal@big.ac.cn; lwan@amss.ac.cn
[1]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190 Beijing, China
[2]Beijing Institute of Genomics, Chinese Academy of Sciences, 100101 Beijing, China
Full list of author information is available at the end of the article

Background

Single-cell RNA sequencing (scRNA-seq) technology provides snapshots of transcriptomes at single-cell resolution, offering a comprehensive approach to study complex biological processes, such as cell fate decisions [1–3]. Given the challenges raised by high-dimensional gene expression profiles of large-scale heterogeneous cell populations at diverse stages during cell state-transitions, many computational methods have been proposed for the visualization, clustering, and reconstruction of scRNA-seq data (see [4], a recent review).

Cells used for a single-cell experiment are snapshot of heterogeneous dynamic populations. However, the dynamic range is usually limited, and only specific cell fate decisions can be analyzed when the samples are collected at a single stage or condition. Therefore, to extend the dynamic range and account for whole processes of cell development, time series scRNA-seq data are emerging. These time series data are generated by sampling single cells collected and sequenced at multiple time stages along the time course of cell developmental processes [5, 6]. Yet, by integrating time series analysis into scRNA-seq, new questions and challenges arise. Importantly, we know that scRNA-seq data across time stages are contaminated by assorted sources of variations during data collection and generation [7]. Moreover , the cell-to-cell gene expression is highly variable across the stochastic dynamics of gene transcription [8]. To address these issues, a handful computational time series scRNA-seq methods have been developed to (1) determine temporal trajectories [9], (2) reconstruct cell development landscapes based on the sophisticated mathematical tool of optimal transport [10], and (3) identify gene-gene interactions, as well as gene networks [8, 11].

However, for such methods to perform model reduction, computation and validation, they must incorporate dimensionality reduction and/or visualization of single-cell high-dimensional gene expression profiles. Accordingly, dimensionality reduction would ideally reveal the intrinsic structure of the data by representing both global structure by preserving the topology and geometry and local structure by preserving the neighborhood relationship. Among the numerous dimensionality reduction and visualization methods, principal component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE) are most widely used in the single-cell community to visualize data structures [4, 12]. However, PCA is designed to preserve linear structure based on eigen-decomposition of a matrix into canonical form, and as such, it is incapable of handling nonlinear structures. While tSNE emphasizes neighborhood information to reveal the local cluster structures of the data, it tends to shatter trajectories and fails to preserve the global structures [12].

To address the limits of dimensionality reduction methods, like PCA and tSNE, we recently adopted the nonlinear dimensionality reduction algorithm, Elastic Embedding (EE) [13], to accurately visualize and reconstruct the embedded intrinsic latent space of cell development trajectory[14]. Both tSNE and EE embed high-dimensional data points into low-dimensional latent space by modeling the data points interactively with two terms. One term is an attractive force that attracts pairs of points towards each other, while the other term activates a repulsive force that simultaneously separates all pairs of points. A similar embedding idea was adopted in force-directed layout embedding [15], which has also been used in the visualization of scRNA-seq [16]. As an extension of tSNE, the EE algorithm penalizes placing latent points far from similar data points, as well as penalizes placing latent points from dissimilar data points close together [13], thereby preserving both local and global intrinsic data structures [17].

However, to the best of our knowledge, no dimensionality reduction and visualization methods can now incorporate temporal information of single-cell experiments into times series scRNA-seq data, i.e., time stages when samples are collected. Instead, scRNA-seq data from different time stages are combined as the input for methods like PCA and tSNE to obtain 2-dimensional visualization of the data structure [5, 6, 10]. TSEE can carefully incorporate experimental temporal information, resulting in significant improvement of temporal resolution of the cells on the 2-dimensional plane, as well as uncovering the subtle structures of dynamic gene expression patterns.

In this study, we propose a time series elastic embedding (TSEE) algorithm for dimensionality reduction and visualization of time series scRNA-seq data by incorporating the temporal information of the experiments. TSEE is an extension of EE by introducing an additional repulsive force term when pairs of data points are collected at distinct time stages (see Eq. 1). TSEE penalizes placing latent points in close proximity to data points otherwise separated by experimental time interval. For developmental processes, the rationale that underlies TSEE holds that data points sampled at the same time stage should be more similar than those sampled from adjacent time stages, while those data points sampled from adjacent time stages should be more similar than data points sampled from the nonadjacent time stages. In this way, TSEE preserves the temporal order and structure of time series data through the input of experimental temporal information. A similar motivation was successfully applied in the time-dependent community detection of time-varying networks [18].

In this paper, we first introduce the TSEE algorithm, and then provide an efficient numerical implementation of TSEE based on the partial-Hessian method developed

by EE [19]. Next, we demonstrate the power of TSEE in the visualization of two datasets of time series scRNA-seq: human preimplantation embryo (hereinafter denoted as HPE) dataset [5] and early zebrafish embryonic development (hereinafter denoted as Zebrafish) dataset [6]. By establishing a new constraint term of temporal repulsive force, TSEE dilutes the distortions of the assorted sources of data variations across time stages and achieves temporal resolution enhancement. Compared to existing methods such as PCA, tSNE and EE, our TSEE shows superior ability to gain time resolution on 2-dimensional space by preserving local, global and temporal structures of time series scRNA-seq data. Furthermore, the visualization represented by TSEE uncovers the subtle patterns of dynamic gene expression by showing continuous patterns with regularity at the interface between samples from adjacent time stages. For example, TSEE reveals the oscillating waves of gene expression for *HER1*, *HER7*, *SOX2* and etc. along the time course, providing a solid foundation for downstream mathematical modeling and analysis. We also demonstrate robustness in the choice of TSEE.

Methods

Time series elastic embedding (TSEE)

Given the time series scRNA-seq dataset, the single-cell samples are collected at n time stages $\{t_1, t_2, \ldots, t_n\}$, and for each time stage $t_i (1 \leq i \leq n)$, a number of m_i single cells are sequenced with the corresponding gene expression vectors $y_1^{(t_i)}, \ldots, y_{m_i}^{(t_i)} \in \mathbb{R}^D$, where D is the number of genes for single cells. Thus, the single cell gene expression profile of a total number of $N = \sum_{i=1}^{n} m_i$ cells is contained in the data matrix

$$Y_{N\times D} = \left(y_1^{(t_1)}, \ldots, y_{m_1}^{(t_1)}, \ldots, \ldots, y_1^{(t_n)} \cdots, y_{m_n}^{(t_n)}\right)^T.$$

TSEE is an extension of EE, a nonlinear dimensionality reduction method [13] by incorporating the information of t_i, embedding the N data points in D-dimensional space into the latent d-dimensional ($d \ll D$) coordinates of

$$X_{N\times d} = \left(x_1^{(t_1)}, \ldots, x_{m_1}^{(t_1)}, \ldots, \ldots, x_1^{(t_n)} \cdots, x_{m_n}^{(t_n)}\right)^T.$$

In a manner that minimizes the pseudo potential energy function as

$$E[X;\lambda,\beta] = \sum_{n,m=1}^{N} w_{nm}^P \|x_n - x_m\|^2 + \lambda \sum_{n,m=1}^{N} \left(w_{nm}^N + \beta t_{nm}\right) \exp\left(-\|x_n - x_m\|^2\right), \quad (1)$$

where the weights $w_{nm}^P, w_{nm}^N, t_{nm}$ are defined as

$$w_{nm}^P = N^+ w_{nm}^+ = N^+ \exp\left(\frac{-\|y_n - y_m\|^2}{2\sigma^2}\right),$$
$$w_{nm}^N = N^- w_{nm}^- = N^- \|y_n - y_m\|,$$
$$t_{nm} = N^- |t(n) - t(m)|.$$

Among these, N^+ and N^- are normalization factors of weights for the two summation terms which, respectively, are $N^+ = \left(\sum_{n,m=1}^{N} w_{nm}^+\right)^{-1}$ and $N^- = \left(\sum_{n,m=1}^{N} w_{nm}^- + \beta|t(n) - t(m)|\right)^{-1}$; $|t(n) - t(m)|$ is the time interval between sample y_n and y_m. TSEE is different from EE by incorporating additional temporal information of t_{nm} in the model.

The pseudo potential energy function of TSEE has two terms on the right hand side of Eq. 1. The first term (*attractive* term) attracts pairs of data points towards each other, while the second term (*repulsive* term) separates all pairs of points. The weights $W^P = \left(w_{nm}^P\right)$ of the *attractive* term are defined based on the similarity between data points in high-dimensional space (original space) [20], enabling TSEE to place neighboring points in high-dimensional space still close to each other in low-dimensional embedding space. The weights of the *repulsive* term of TESS are composed of two parts. The weights $W^N = \left(w_{nm}^N\right)$ are the disparities between data points in high-dimensional space (original space), and $T^- = (t_{nm})$ are the time intervals between samples. Thus, in the constructed latent space, TSEE not only preserves the local and global structures of data hidden in original high-dimensional space, as EE does [14, 17], but also further preserves temporal structure of data by penalizing placing close together latent points that correspond to data points separated by experimental time intervals. Intuitively, the experimental temporal information of the experimental time of samples reflects, to some extent, cell development stages. In other words, the later a cell is collected in time, the more likely it will be in later developmental stages, and a longer experimental time interval between two cells typically indicates a longer distance between the cells, along the developmental process.

TSEE has two regularization parameters in Eq. 1. Parameter λ trades off the attractive and repulsive terms, while parameter β trades off the composite weights of data disparities in high-dimensional space and time intervals on the repulsive term. By fixing λ, a smaller β will have smaller effects by the temporal order/structure of samples, leading to less consideration of temporal information. If $\beta = 0$, TSEE degenerates into EE, and no temporal information will be incorporated. If fixing β, a smaller λ makes TSEE focus on local structures, determining $X_{N\times d}$ based on the attractive term, and when $\lambda = 0$, TSEE degenerates to Laplacian Eigenmap methods [13]. Choices

and robustness analysis of both λ and β on embeddings in latent space are discussed in "Results" section. If either λ or β is too large, the result will either be distortion of local structures or loss of global structures in latent space.

The pseudocode of TSEE is in Algorithm 1.

Algorithm 1: TSEE algorithm

Input : single cell expression $Y_{N\times D} = (Y_1, \ldots, Y_N)^T$, time of single cells $T = (t_1, \ldots, t_N)$, λ and β, the default of which are both set to be 10

Output: low-dimensional embedding $X_{N\times d} = (X_1, \ldots, X_N)^T$

Step1: Normalize data

$$Y \leftarrow \frac{Y - \min_{i,j} y_{ij}}{\max_{i,j} y_{ij} - \min_{i,j} y_{ij}},$$

$$T \leftarrow \frac{T - \min_i t_i}{\max_i t_i - \min_i t_i}$$

Step2: Calculate two weight graphs W^P, W^N

$W^P \leftarrow$entropic affinities of $Y_1, \ldots, Y_N$

$W^- \leftarrow \{\|Y_i - Y_j\|\}_{N\times N}$

$T^- \leftarrow \{|t_i - t_j|\}_{N\times N}$

$W^N \leftarrow W^- + \beta T^-$

Normalize graphs:

$W^P \leftarrow \frac{1}{2}\left(W^P + \left(W^P\right)^T\right)$, $diagonal(W_P) \leftarrow 0$, $W^P \leftarrow \frac{W^P}{\sum_{i,j} w^P_{ij}}$

$W^N \leftarrow \frac{1}{2}\left(W^N + \left(W^N\right)^T\right)$, $diagonal(W_N) \leftarrow 0$, $W^N \leftarrow \frac{W^N}{\sum_{i,j} w^N_{ij}}$

Step3: Calculate low-dimensional embedding X

$$X \leftarrow \text{TSEE}\left(W^P, W^N, \lambda\right)$$

The numerical implementation of TSEE

In the numerical implementation of TSEE, we adopt the efficient computational method of partial-Hessian proposed in [19] to find $X_{N\times d}$ that minimizes the function E of TSEE (Eq. 1).

For a given symmetric weighted graph (matrix) $W = (w_{nm})$, the Laplacian matrix graph is defined as $L = D - W$, where $D = \text{diagonal}\left(\sum_{m=1}^{N} w_{nm}\right)$ is defined as degree matrix. L is positive semi-definite if W is non-negative. The gradient of objective function E of TSEE can be represented in terms of Laplacian as

$$\nabla E = 4LX,$$

where L is the Laplacian matrix of $W = (w_{nm})$ with

$$w_{nm} = w^P_{nm} - \lambda\left(w^N_{nm} + \beta t_{nm}\right)\exp\left(-||x_n - x_m||^2\right).$$

This optimization problem can be solved by an iterative approach in the form of

$$x_{k+1} = x_k + \alpha_k p_k,$$

where $\alpha_k > 0$ is the step size in each iteration that satisfies the linear search principles (e.g., Armijo rule), and $p_k \in \mathbb{R}^{ND}$ is the search direction. In each iteration, p_k is determined by

$$B_k p_k = -g_k,$$

where g_k is the gradient at k-th iteration, and B_k is a positive-definite matrix to achieve a descent direction satisfying $p_k^{\mathrm{T}} g_k < 0$. The procedures iterate until certain conditions are satisfied, for example, the iteration terminates when the number of iterations achieves the given constraint, or the distance between x_k and x_{k+1} obtained by two adjacent iterations is less than a given threshold.

The partial-Hessian method relies on the *spectral direction* based on partial Hessian, the attractive Hessian $L^P \otimes I_d$, to strike the best compromise between deep descent and efficient computation, where L^P is the Laplacian matrix of W^P. In order to prevent the problem of singularity, a small $\mu_k I$ is added to the partial Hessian, ensuring positive definiteness of B_k, that is

$$B_k = L^P \otimes I_d + \mu_k I$$

and μ_k is set as $10^{-10}\min_{i,j}\left(L^P\right)_{i,j}$ in this study. Algorithm 2 shows the detailed steps of Partial-Hessian strategies. The authors of EE [19] demonstrated that the spectral direction obtained, as described above, can be rapidly computed, leading to global and fast convergence. Compared with EE, TSEE adds an additional repulsive factor in the objective function, which results in a simple modification of W and L with no effect on computational performance, making the complexity of TSEE comparable to that of EE. In terms of the large-scale Zebrafish dataset with sample size $\sim 40k$, the computational time of EE and TSEE is 38 min and 45 min, respectively.

Datasets

Two publicly available datasets of time series scRNA-seq are used in this study. The Zebrafish dataset [6] was obtained across 12 closely spaced stages of early zebrafish development spanning from high blastula stage (3.3 hours-postfertilization (hpf), just after transcription from the zygotic genome begins) to six-somite stage (12 hpf, shortly after the completion of gastrulation). The number of cells at each time stage ranges from 200 to 7162, comprising a total of 39,505 cells. The UMI count data can be accessed at NCBI GEO (accession no.GSE106587). The human preimplantation embryo (HPE) dataset [5] consists of individually isolated embryonic cells during the human preimplantation process, starting from the 8-cell stage at embryonic day 3 (E3) up to

Algorithm 2: TSEE: Partial-Hessian Strategy

Input : two weighted graphs W^P, W_N, λ
Output: low-dimensional embedding X

Step1: Calculate positive definite matrix based on Partial Hessian

$D^P \leftarrow diagonal\left(\sum_{j=1} w_{ij}^P\right)$, $L^P \leftarrow D^P - W^P$

$B \leftarrow 4 \times \left(L^P + 10^{10} \min\left\{l_{ij}^P : l_{ij}^P > 0\right\} I_{N\times N}\right)$

$R \leftarrow$ upper triangular cholesky factorization of B

Step2: Search spectral directions

initialize:

$opts.maxit \leftarrow 100$, $count \leftarrow 1$

$Xold \leftarrow$random initial coordinates

$Eold \leftarrow$value of objective function at $Xold$

while $count \leq opts.maxit$ **do**

  $W \leftarrow \left\{w_{ij}^P - \lambda w_{ij}^N \exp\left(-\|Xold_i - Xold_j\|^2\right)\right\}_{N\times N}$

  $D \leftarrow diagonal\left(\sum_j w_{ij}\right)$

  $L \leftarrow D - W$

  $G \leftarrow 4LX$ # G is the gradient of objective function

  $P \leftarrow -\left(R^T\right)^{-1} R^{-1} G$ # P is the *spectral direction*

  line search on step size:

  $\alpha \leftarrow 1$, $\rho \leftarrow 0.8$, $c \leftarrow 0.1$

  $E \leftarrow$value of objective function at $Xold + \alpha P$

  $tmp \leftarrow \sum_{i,j} p_{ij} g_{ij}$

  while $E > Eold + \alpha \times c \times tmp$ **do**

    $\alpha \leftarrow \alpha \times \rho$

    $E \leftarrow$ value of objective function at $Xold + \alpha P$

  end

  $X \leftarrow Xold + \alpha P$

  $Xold \leftarrow X$

  $count \leftarrow count + 1$

end

the time point just prior to implantation at E7, consisting of a total of 1529 single cells. The data can be downloaded from https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3929/.

In addition, the scRNA-seq dataset of *Drosophila* embryos [21] is utilized here as a demonstration of the generalizability of TSEE for incorporating spatial information. The high-quality data of a total 1297 cells, as defined in [21], were selected. The data are used to reconstruct the spatial positions of cells in the precisely staged embryos. The 84 genes used as predictors of spatial positions of cells in the precisely staged embryos as in [21] are utilized as our spatial genes. The data can be obtained at https://shiny.mdc-berlin.de/DVEX/.

Preprocessing of time series scRNA-seq data

Prior to the implementation of TSEE, the time series scRNA-seq data are preprocessed as follows. First, we select out the most variable genes according to the Z-scores of their variations across all samples. Second, the selected gene expression profile is normalized with all values, subtracting the minimum in the matrix, followed by dividing all values by their maximum of the processed matrix, such that the values of the elements of the normalized profile matrix range from 0 to 1. The time points of samples are also normalized to range from 0 to 1 with the initial time stage set as 0 and the final time stage set as 1. Third, the PCA algorithm is utilized to select the top principal components which preserve most variations, and the dimension of components is determined following the procedure in [14]. Finally, the data of top PCA components are further normalized, as described in the second step.

Performance comparisons

We compare TSEE with PCA, tSNE and EE to evaluate their individual performance on visualizing data, preserving local and global structures, and revealing gene expression patterns.

The scatting plots of cells on the 2-dimensional embedded space are first displayed to compare their visualization results. A good visualization result should reveal the local and global structures, as well as uncover the latent specific gene expression patterns.

To measure the preservation of structures quantitatively, we adopt 1) the in-group proportion (*IGP*) [22] and 2) *IGP2*, a modification of *IGP*, to evaluate local structure preservation. In addition, we adopt the Pearson correlation coefficient (PCC) between experimental time and pseudotime obtained by DensityPath [14], to evaluate global structure preservation.

The *IGP*, defined as the proportion of samples in a group, the nearest neighbors of which are also in the same group [22], can be used to assess the accuracy of distinction of cells on 2-dimensional space. In this study, we use temporal information to define groups. For a given dataset $X = \{x_1, x_2, \ldots, x_N\}$, consisting of N samples which belong to n groups $U = \{1, 2, \ldots, u, \ldots, n\}$, we define $j^N = \arg\min_{k\neq j} \|x_j - x_k\|$ as the index of x_j's nearest sample, and we regard $Class_x\left(j\right)$ as the label for sample x_j. Based on the definition above, *IGP* of group u based on dataset X is defined as

$$IGP(u, X) = \frac{\#\left\{j | Class_x\left(j\right) = Class_x\left(j^N\right) = u\right\}}{\#\{j | Class_x\left(j\right) = u\}}.$$

The value of *IGP* ranges from 0 to 1, and a higher value indicates a better cell distinction between groups.

Because of the heterogeneity of cells, the cells at distinct time stages may still belong to the same cell type, which

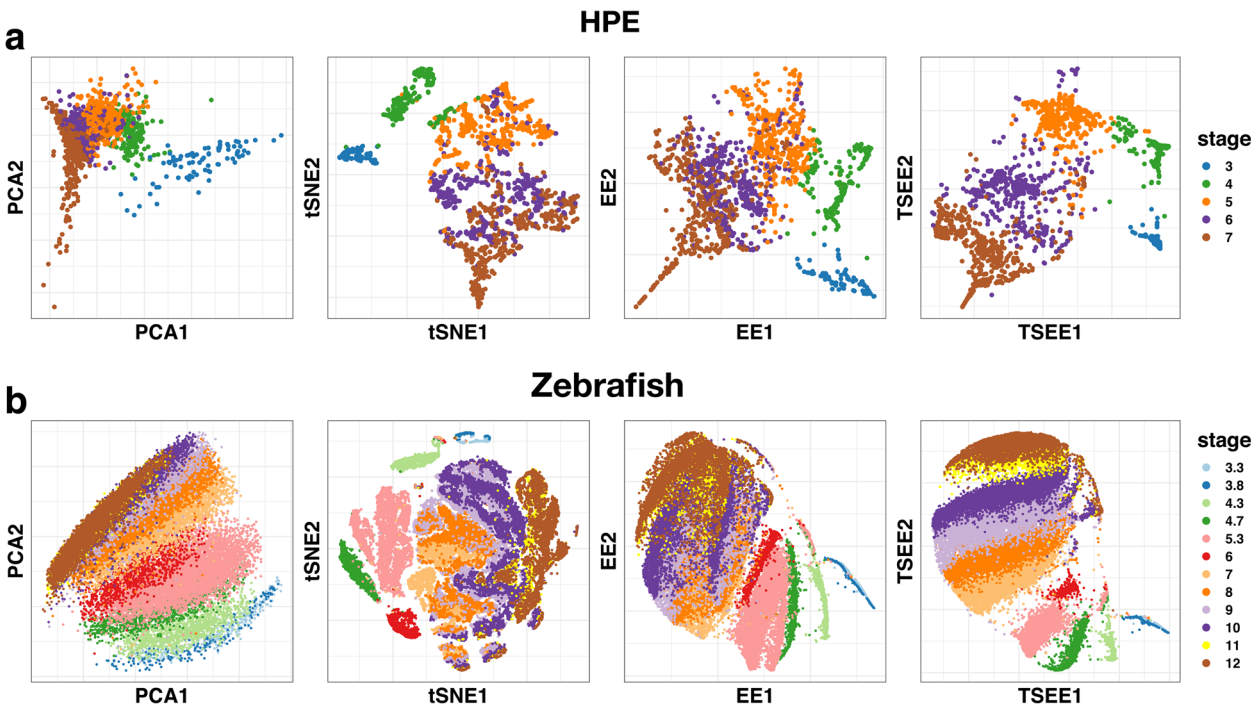

Fig. 1 Visualization results by PCA, tSNE, EE and TSEE. The four dimensionality reduction methods are applied to two datasets of time series scRNA-seq: **a** HPE and **b** Zebrafish. Each cell is colored according to the corresponding experimental time stage

means that *IGP* tends to underestimate the performance of cell-specific distinction. Consequently, we weaken the requirements of *IGP* and propose a new IGP score by defining the proportion of samples in a group, the nearest neighbors of which are also in the same, or adjacent group (adjacent time stages here). The so-called *IGP2* of the adapted index is formulated as

$$IGP2(u,X) = \frac{\#\left\{j||Class_x\left(j\right) - Class_x\left(j^N\right)| \leq 1, Class_x\left(j\right) = u\right\}}{\#\{j|Class_x\left(j\right) = u\}}. \quad (2)$$

In this study, the tSNE algorithm is implemented by the *Rtsne* function in the R package **Rtsne** with default parameters, and the EE algorithm is performed according to the code downloaded from http://faculty.ucmerced.edu/mcarreira-perpinan/software.html.

Results

We analyze two time series datasets: Zebrafish [6] and HPE [5] (see "Methods" section for details). To validate the performance of TSEE, we compare it with three other dimensionality reduction algorithms, PCA, tSNE and EE.

TSEE outperforms other dimensionality reduction methods in visualization and structure preservation

We apply the 4 dimensionality reduction algorithms noted above to HPE and Zebrafish datasets and demonstrate the embedding results on the 2-dimensional embedding space to visualize cell development structures. TSEE not only preserves the local and global data structures, but also has higher temporal resolution of cells with different time stages (Fig. 1).

More specifically, in the HPE dataset, PCA combines cells from E4 to E7 tightly, grouping cells into two groups visually, one mainly containing cells at E3 and the other

Table 1 Evaluation of local structure preservation based on averaged values of *IGP*

	PCA	tSNE	EE	TSEE
HPE	0.7747	0.8900	0.8466	0.9407
Zebrafish	0.4757	0.8799	0.7127	0.8943

Table 2 Evaluation of local structure preservation based on averaged values of *IGP2*

	PCA	tSNE	EE	TSEE
HPE	0.9661	0.9961	0.9936	0.9965
Zebrafish	0.8525	0.9930	0.9196	0.9954

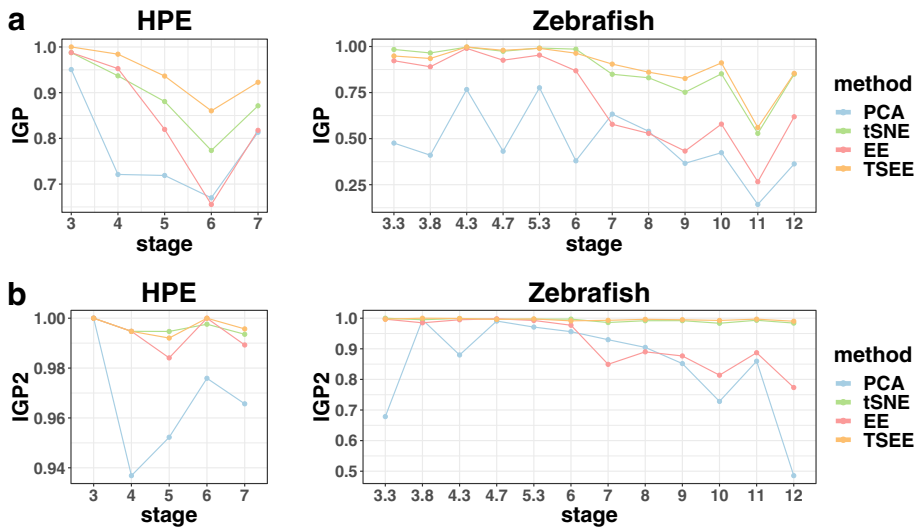

Fig. 2 Comparison of local structure preservation performance by *IGP* and *IGP2*. Performances in local structure preservation of PCA, tSNE, EE and TSEE is individually evaluated based on two metrics, **a** *IGP* and **b** *IGP2*, on the HPE and Zebrafish datasets

filled with cells from E4 to E7, which would affect downstream analysis, such as clustering. The embedded points with various stages obtained by tSNE and EE are both evenly distributed as a whole, but cells from E5 to E7 heavily overlap, with some E7 cells even falling into the area of E5 cells (Fig. 1a). For TSEE, cells at different time stages are well separated so that only cells at adjacent time stages are mixed in agreement with the heterogeneity of cells (Fig. 1a).

For the Zebrafish dataset, all four methods arrange cells along time with relatively large gaps arising in early developmental stages (Fig. 1b). However, PCA and EE results show that cells from 8 to 12 hpf are highly mixed, almost stacked together, even when two or more time intervals exist among them. Although clearly separating points with different time stages, tSNE shows distorted 2-dimensional temporal structures in that cells from 3.3 hpf to 4.3 hpf are distributed along time course, while cells at 4.7 hpf are separated with the cells before 4.7 hpf by cells at 5.3 hpf (Fig. 1b). In contrast, for TSEE, the cells from 3.3 to 6 hpf are obviously separated. Even though cells from 7 hpf to 12 hpf are not separated as distinctly as the cells before 6 hpf, significant gaps still exist between cells at adjacent time stages, thus showing the best performance on preserving structures along time (Fig. 1b).

To quantitatively evaluate the performance of TSEE, we apply *IGP* and *IGP2* scores to measure the distinction of cells, regarding time stages as standard of cell groups. TSEE always has highest averaged values of *IGP* and *IGP2* (see Tables 1 and 2), and its performance is similar to that of tSNE based on the two scores across all time stages in HPE and Zebrafish (Fig. 2), indicating that TSEE preserves the local structures and differentiates cells at various time stages better than the three other methods.

To evaluate the preservation of global and temporal structures by the four methods, we measure performance quantitatively by applying DensityPath [14], which has high accuracy in the reconstruction of cell development

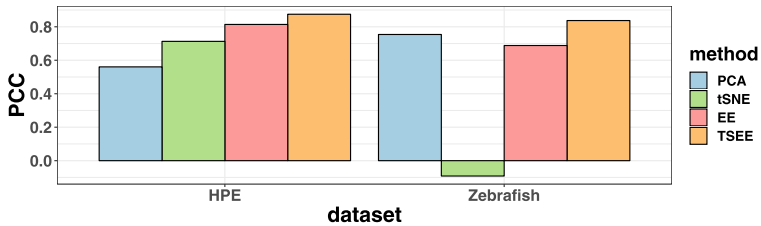

Fig. 3 Quantitative comparison of global structure preservation. Performance in global structure preservation of PCA, tSNE, EE and TSEE is evaluated based on PCC metric of pseudotime calculated by DensityPath and experimental time for the HPE and Zebrafish datasets

trajectory, as well as pseudotime calculation, for both datasets. Since our purpose is to investigate the performance of the four dimensionality reductions, we only apply DensityPath to dimensionality reduced data of 2-dimensional space by the four methods to reconstruct cell state-transition path and calculate pseudotime. After setting the root as 1472-th and 1-st cell in the HPE and Zebrafish datasets, respectively, the accuracy of calculated pseudotime is measured by PCC between the calculated pseudotime and experimental time of the cells. A larger value of PCC indicates more consistency in global structure preservation along time. We find that TSEE preserves temporal structure information best, while the PCC value of tSNE on the Zebrafish dataset is negative, indicating that the embedding of cells by tSNE distorts whole structures of data along time (Fig. 3).

The hierarchical clustering algorithm is also applied to measure the preservation of distance of time stage. The complete linkage method is adopted here based on centroids of samples at each time stages, where the calculated distance of the centroids is based on 2-dimensional Euclidean distances of the four dimensionality reduction methods separately. Although the adjacent time stages are always close to each other in hierarchical clustering trees based on the results of all four methods (Fig. 4), the clustering tree constructed on the basis of TSEE results shows the best consistency between the distance of centroids and the time order. Specifically, the lengths of leaf nodes (3.3, 3.8, 4.3, 4.7, 5.3, and 6) to the root of the clustering tree decrease with increasing time stage from 3.3 to 6 hpf, and then the lengths of leaf nodes (7, 8, 9, 10, 11, and 12) to the root of clustering tree increase with the increasing time stage from 7 to 12 hpf (Fig. 4). As an indication of cell development processes, the clustering tree result by TSEE indicates that the centroids of cell populations propagate along one direction from 3.3 to 6 hpf and then propagate along another direction from 7 to 12 hpf, which can be further supported by

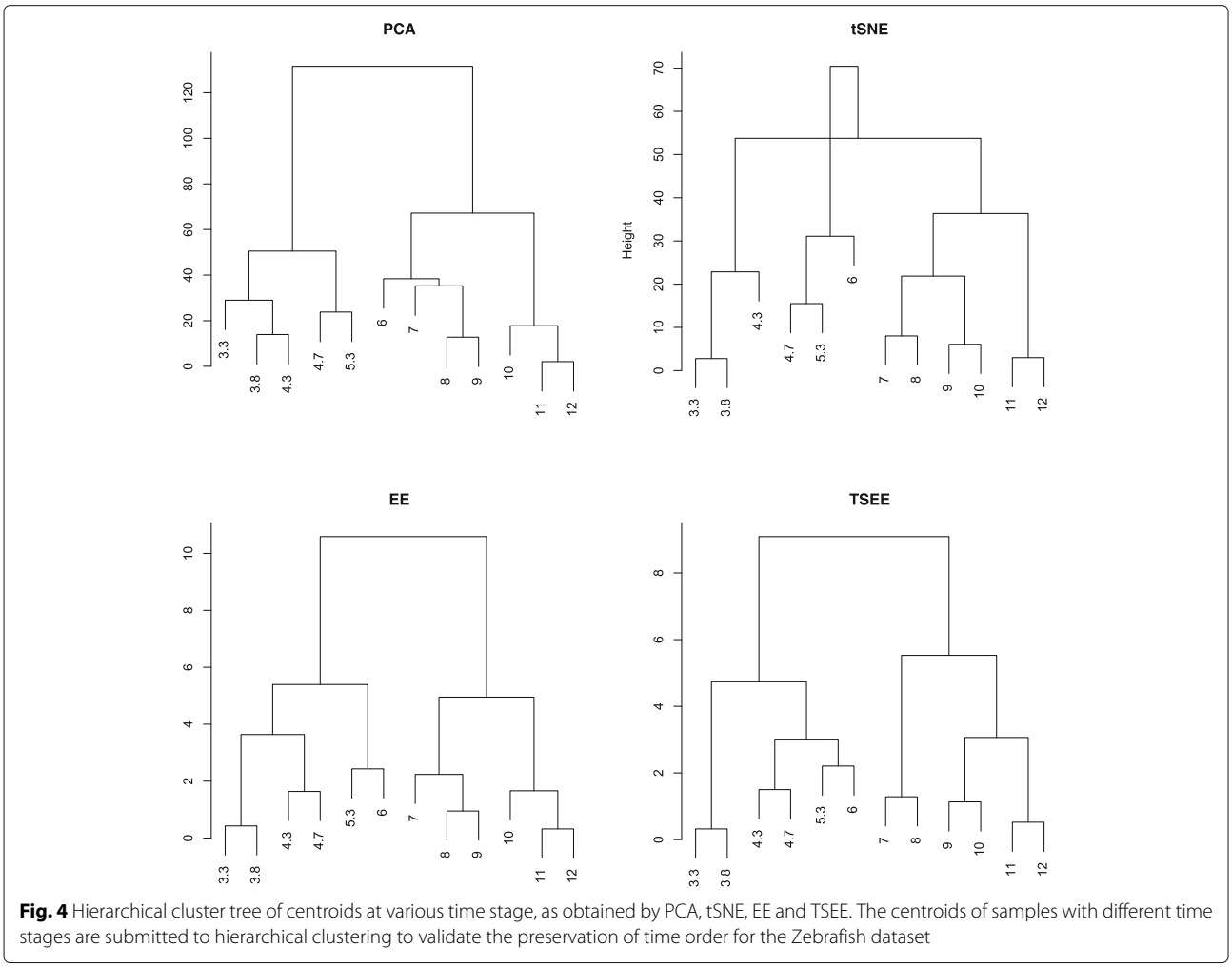

Fig. 4 Hierarchical cluster tree of centroids at various time stage, as obtained by PCA, tSNE, EE and TSEE. The centroids of samples with different time stages are submitted to hierarchical clustering to validate the preservation of time order for the Zebrafish dataset

the large gap between 6 and 7 hpf on the 2-dimensional space by tSNE, EE, and TSEE (Fig. 1b).

TSEE reveals subtle dynamic patterns of gene expression during zebrafish embryogenesis on 2-dimensional space

The large-scale Zebrafish dataset with closely spaced stages allows us to further explore dynamic patterns of gene expression during zebrafish embryogenesis.

The power of TSEE to allow visualization of structures that show dynamic patterns of gene expression can first be illustrated through the expression of marker genes *HER1* and *HER7* on the 2-dimensional space by PCA, tSNE, EE and TSEE in Zebrafish data (Fig. 5). *HER1* and *HER7*, for which the existence of oscillation expression patterns by negative feedback has been reported in presomitic mesoderm of zebrafish [23], clearly show oscillatory expression patterns along time stages on the 2-dimensional space of TSEE, indicating that TSEE successfully reveals the underlying dynamic patterns of gene expression. On the other hand, PCA only shows two wide strips of highly expressed genes on the whole plot, and tSNE gathers cells with highly expressed genes together without any oscillation pattern whatsoever. Meanwhile EE just shows a narrow region with oscillation expression pattern in early developmental stages, and the cells with high-level expression are stacked in the latter developmental stages, resulting in low temporal resolution. For TSEE, however, points on the 2-dimensional plane clearly demonstrate the oscillation patterns of gene expression with high resolution, validating the effectiveness of TSEE in preserving the intrinsic structures of data with high resolution, while, at the same time, allowing visualization of the underlying dynamic structures of genes.

We also explore the expression patterns of genes related to embryogenesis from the list provided by [6] on the 2-dimensional spaces by the four methods to examine the developmental patterns. Here, visualization by TSEE can also show clearer oscillation waves of gene expressions along time compared to the other methods (see Fig. 6 for four examples). PCA tends to distribute the cells with high-level gene expression uniformly, or accumulate them in the region of late developmental stages. The tSNE methods tends to gather together the cells with high gene expression, while small gaps occur among them, but this only shows meaningful clusters of cells without any evidence of patterns of gene expression. EE displays the products resulting from mixtures of cells at various time stages as the region of high-level gene expression, where genes are continuously expressed. The three methods place together the regions where the specific genes are highly expressed, failing to uncover the dynamic structures even if some special gene expression patterns exist among all cells. In TSEE, different patterns of gene expression can be revealed. As examples, *WNT8A*, an essential transcript for zebrafish axis development [24], displays an oscillatory gene expression pattern throughout

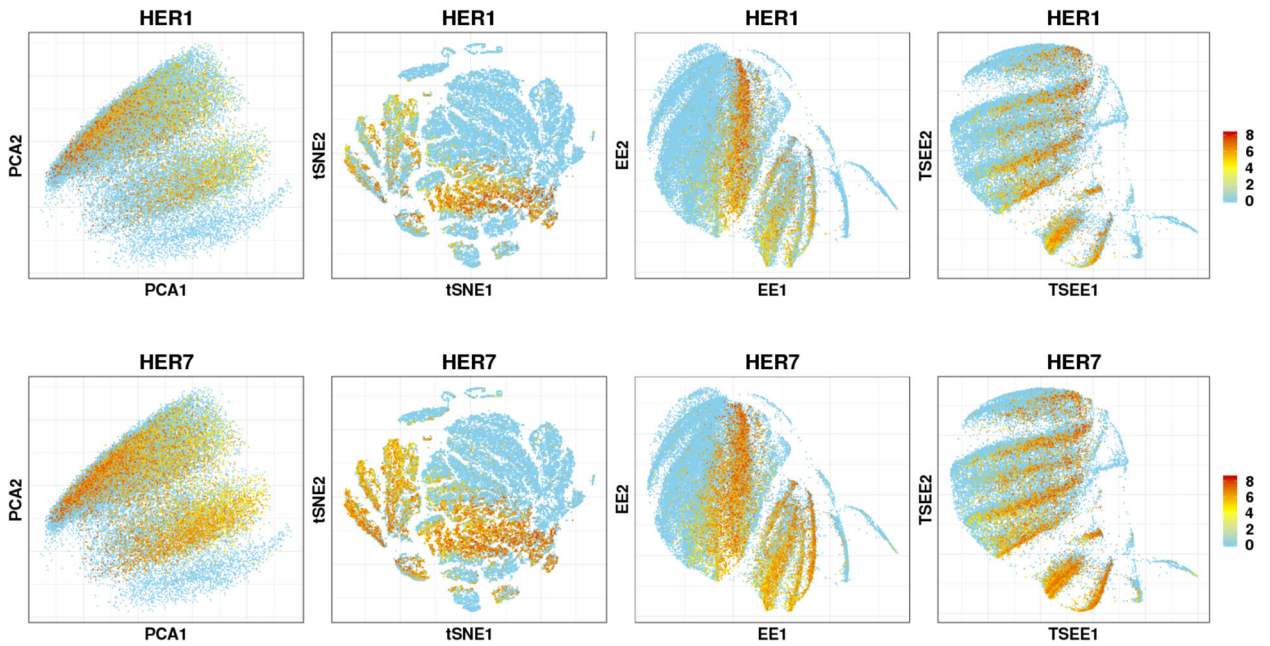

Fig. 5 Distribution of gene expression of *HER1* and *HER7* in 2-dimensional space. The distribution of *HER1* and *HER7* gene expression in 2-dimensional embedding obtained by PCA, tSNE, EE and TSEE, revealing the oscillatory expression pattern along the development of time course during zebrafish embryogenesis

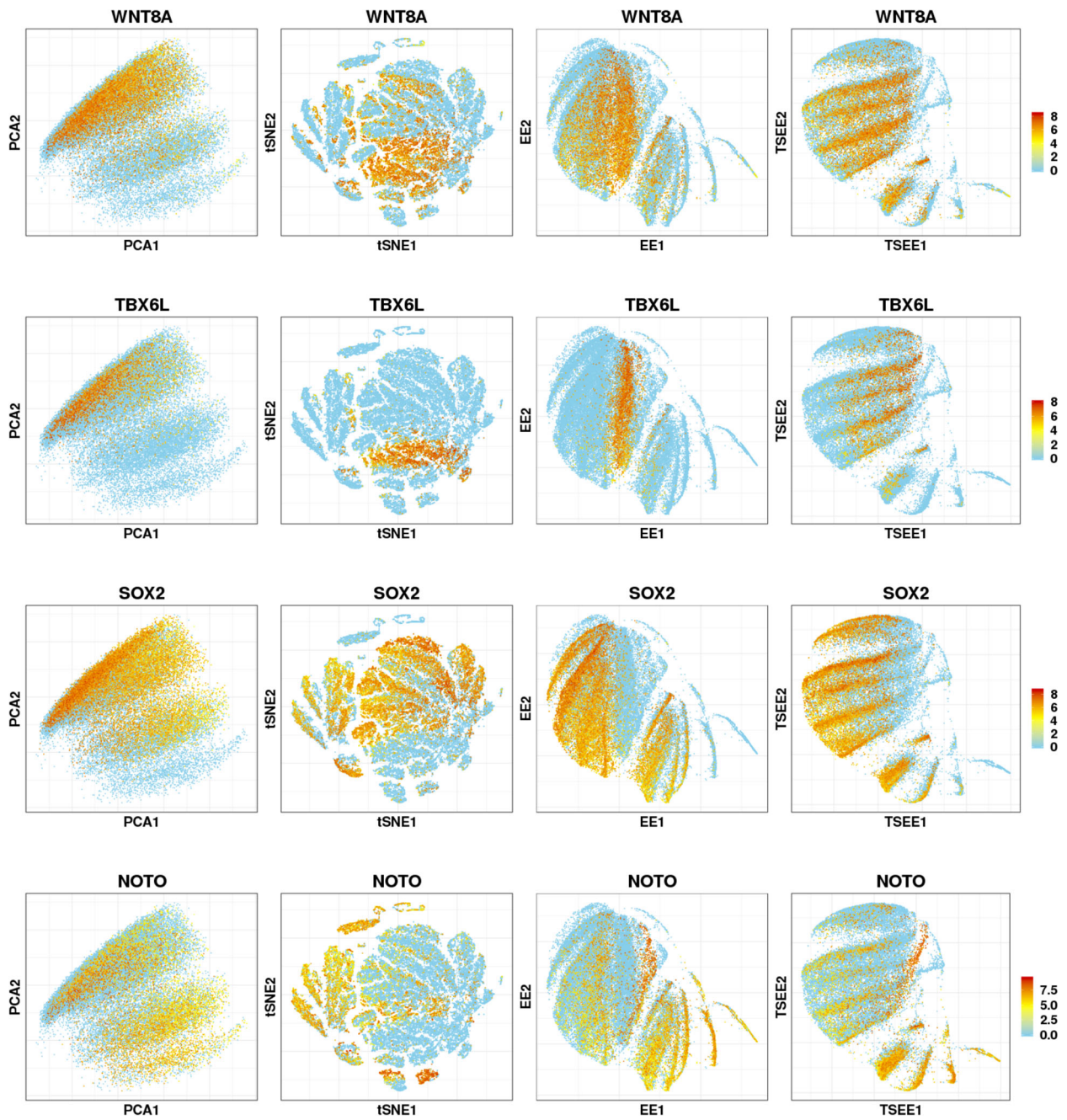

Fig. 6 Distribution of expression of four genes in 2-dimensional space. Distribution of expression of four genes in 2-dimensional embedding obtained by PCA, tSNE, EE and TSEE. Dynamic gene expression patterns are revealed by TSEE

all samples. *TBX6L*, which contributes to posterior paraxial mesoderm formation during zebrafish embryogenesis [25], reflects a periodically expressed pattern in the right region. *SOX2*, identified as an essential transcription factor to maintain self-renewal or pluripotency [26], reflects oscillation, as well, in the left region in the whole stages. Finally, *NOTO* displays a periodic gene expression pattern entirely, but fades away gradually along time. More genes with an oscillating pattern if expression can be found at https://github.com/ShaokunAn/TSEE.

Oscillating patterns of gene expression revealed by TSEE may be linked to the genome-scale oscillations in DNA methylation during exit from pluripotency [27]. Therefore, our TSEE results provide new insights and perspectives for subsequent analysis of dynamic transitions and regulation mechanisms of key genes.

Parameter choices and robustness analysis of TSEE

Two tuning parameters, λ and β, are found in the pseudo potential energy function E of TSEE (Eq. 1). The parameter choices of λ and β are critical to the embedding results by TSEE. In the implementation of TSEE, we set the default values of both λ and β equal to 10.

To test the robustness of the parameter choices, we first choose the values of each parameter separately in a wide range from 1 to 1000, while keeping the other parameter fixed at default, and calculate weighted mean of *IGP* scores for both HPE and Zebrafish datasets (Fig. 7). The weighted mean of *IGP* is the sum of *IGPs* at each time stage, weighting by the proportion of samples (cells) at each time stage.

When fixing $\beta = 10$, TSEE achieves the highest weighted mean of *IGP* at $\lambda = 100$ in both datasets (Fig. 7a,b). However, the large $\lambda(\geq 100)$ tends to separate samples into clusters on the 2-dimensional space of TSEE, breaking the structure into discontinuities (Fig. 8a,c), which is inappropriate for continuous embryonic development. Therefore, $\lambda = 10$ best balances the preservation of local and global structures.

When fixing $\lambda = 10$, the weighted mean of *IGP* increases sharply when increasing β from 1 to 10 and tends to be saturated after $\beta \geq 100$ (Fig. 7c,d). Although the global structures of cells on the 2-dimensional space of TSEE are stable by varying β from 1 to 1000, TSEE may sacrifice some local structures with large $\beta(\geq 100)$ since the repulsive terms are determined by the difference between time stages, while the dissimilarities based on gene expression are seldom considered. For example, the two corner sections at stage E7 and the one at stage E4 become less distinguishable as β grows in HPE data. Therefore, $\beta = 10$ is sufficient for the incorporation of temporal information data and trades off the weights between gene expression and time stage.

To determine the parameters more precisely, as well as analyze the robustness of TSEE, we further study the performance of TSEE when the parameters are tested in detail in the region from 1 to 100. The three metrics *IGP*, *IGP2* and PCC are employed here. Based on *IGP* and *IGP2*, TSEE has similar performance for various β and λ from 1 to 50 in the HPE and Zebrafish datasets, indicating that TSEE is quite robust to the change of parameters on preserving local structures. We calculate PCC for HPE and Zebrafish data to analyze the robustness to parameter choice, as well. Figure 9 demonstrates that TSEE is quite robust for the choices of β, as well as choices λ, when varying them separately in the range from 1 to 50.

Based on all the results above, we set the optimal default values of both β and λ to be equal to 10.

Discussion

In summary, we propose a novel visualization method, TSEE, for time series scRNA-seq data. To the best of our knowledge, in the single-cell community, TSEE is the first visualization method that considers additional temporal information.

TSEE enhances resolution of the map by correcting for unknown noise variation. The repulsive force based on time intervals of samples enables TSEE to align cells along time, preserving temporal structures, while the intrinsic structures remain well preserved owing to the incorporation of the attractive term and the repulsive term based on their distance in gene expression space. Figure 10 displays the correction of cells by TSEE. The samples colored in red are those for which the nearest neighboring points are at least two time intervals apart, indicating with high probability that they are misplaced in the 2-dimensional plane. The misplaced points by EE are distributed uniformly (Fig. 10a), but when displayed on the 2-dimensional space obtained by TSEE (Fig. 10b), they are aligned into their respective time stages, indicating that TSEE corrects the misplacement of samples by EE. Moreover, the misplaced cells by TSEE are generally located in the interface of two time stages (Fig. 10c), which makes sense since cells are typically heterogeneous, especially at late developmental stages. Because of the

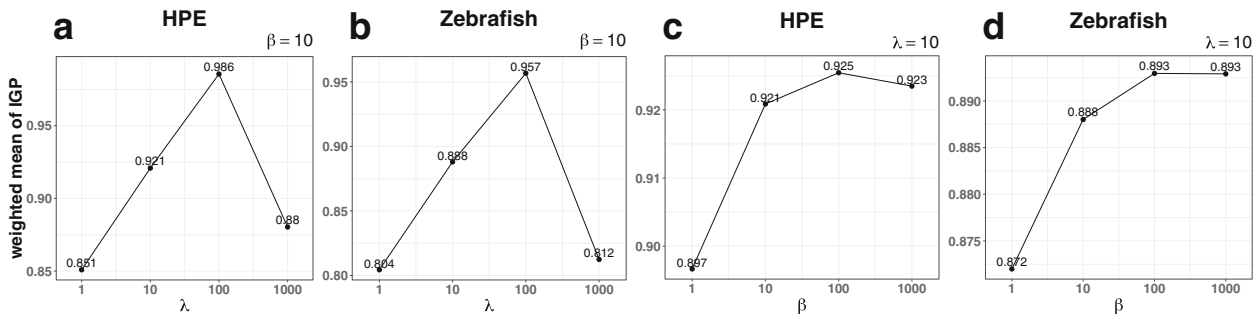

Fig. 7 Weighted mean of *IGP* under different parameter values. The y-axis is the weighted mean of *IGP* where the weights are the proportions of the number of cells at each time stage. The x-axis represents parameter choices (in logarithmic scale). (**a**) results of different λ for HPE dataset, (**b**) results of different λ for Zebrafish dataset, (**c**) results of different β for HPE dataset, (**d**) results of different β for Zebrafish dataset

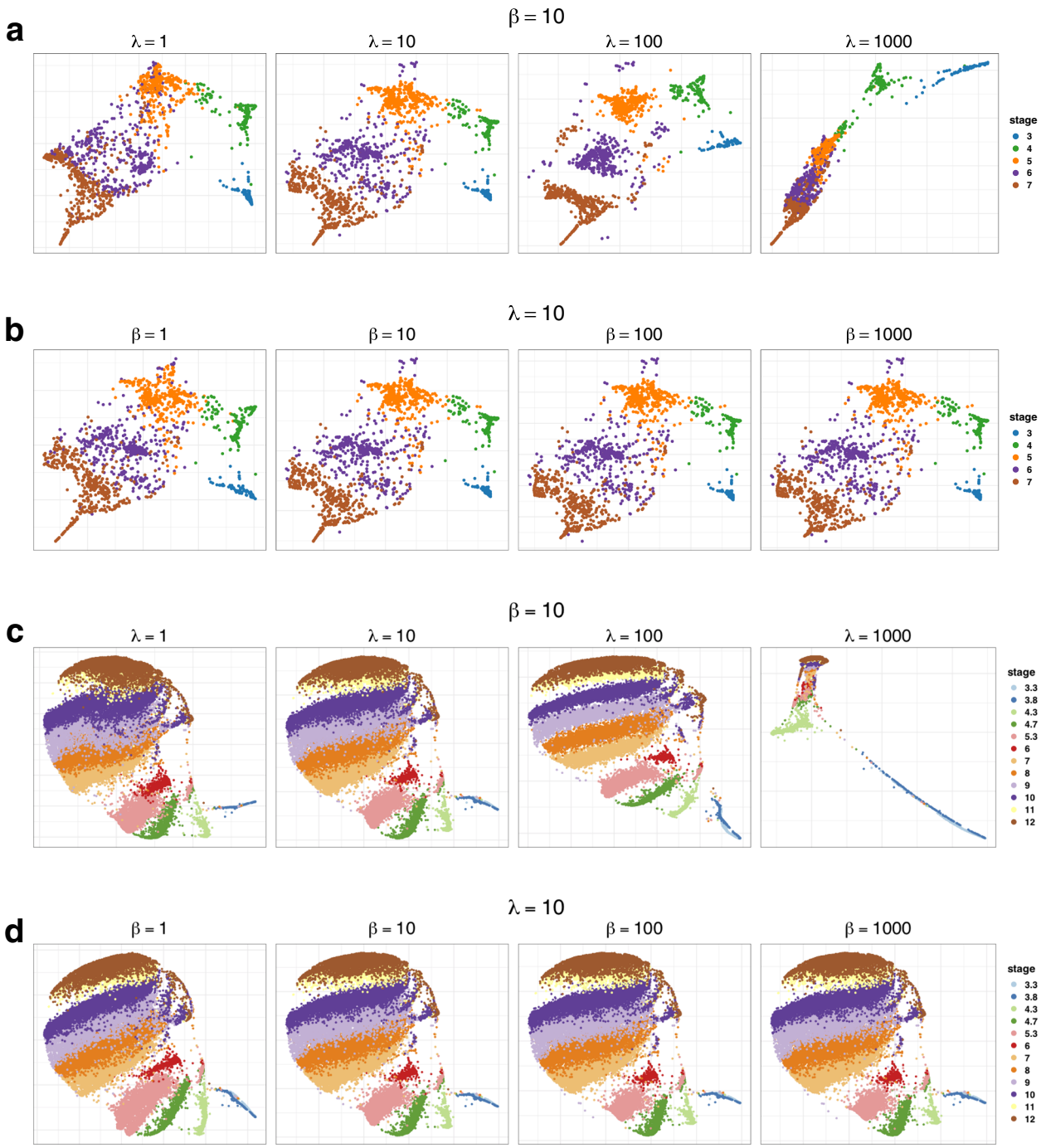

Fig. 8 The 2-dimensional embeddings of TSEE under different parameter settings. The visualization results of TSEE by varying one parameter, while the other is fixed at the default value of 10, are displayed on **a**, **b** for the HPE dataset and on **c**, **d** for the Zebrafish dataset

correction of cells along time, gene expression patterns can be revealed, such as the oscillation process of *HER1* and *HER7* in Zebrafish data, as demonstrated in the "Results" section. The increasing number of genes with oscillatory expression pattern discovered is further supported by the genome-scale oscillations in DNA methylation[27], and the uncovered subtle dynamic structures of time series scRNA-seq data, as, for example, through the use of TSEE, can be utilized for further analysis of gene expression.

The visualization tool of the forced-directed layout [15] utilizes the concept of attractive force and repulsive force, as well, to visualize network structures in low-dimensional space. The data tend to be collapsed along

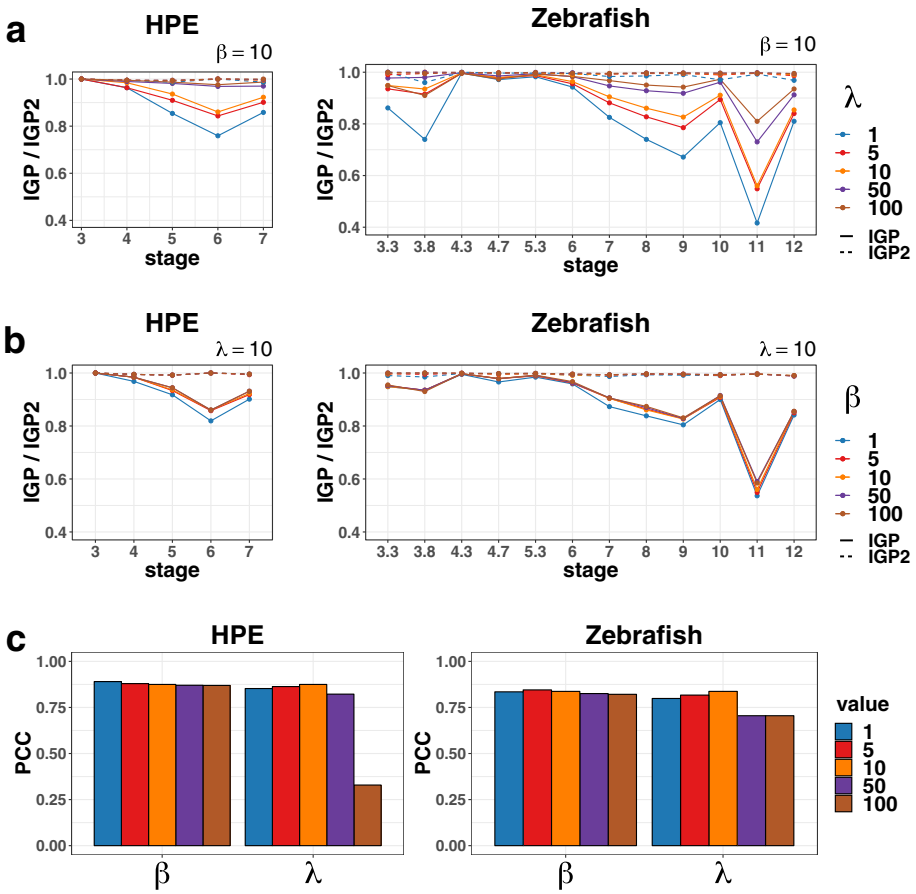

Fig. 9 Robustness analysis of parameters on two datasets. Robustness analysis is performed for λ and β on the two datasets. To evaluate preservation of local structure under each set of parameters, *IGP* and *IGP2* are used. **a** When β is set as default 10, *IGP* and *IGP2* display the effect by the change of λ. **b** When λ is set as default 10, *IGP* and *IGP2* display the effect by the change of β. **c** PCC is used to assess the preservation of global and temporal structures

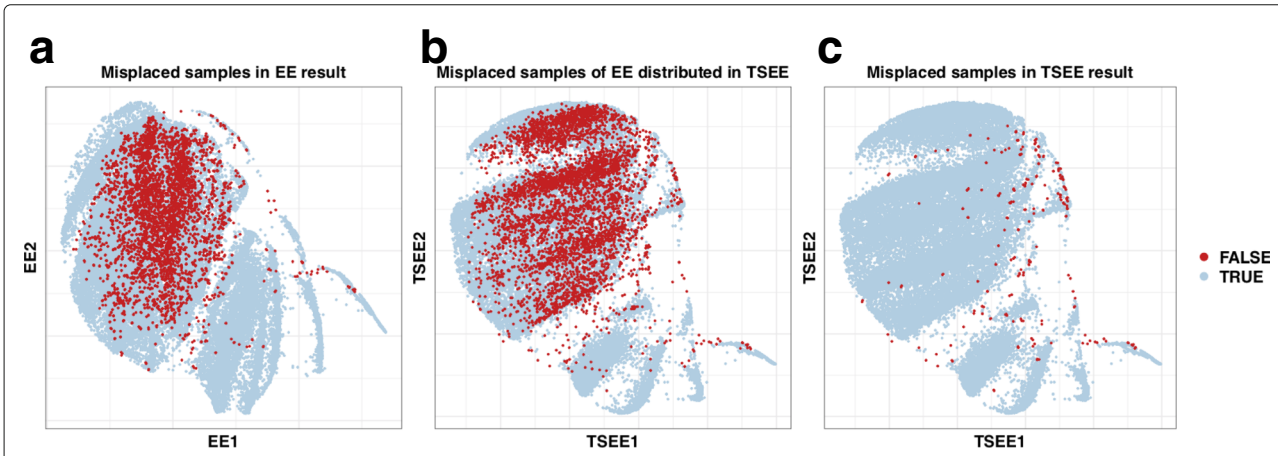

Fig. 10 Distribution of misplaced samples based on 2-dimensional embeddings. When one cell's nearest neighboring point is at least two time intervals away from it, the cell is considered to be misplaced with high probability and is colored in red in above figures. **a** shows the distribution of misplaced cells by EE on latent space of EE, **b** displays the distribution of misplaced cells by EE on latent space of TSEE, and **c** displays the distribution of misplaced cells by TSEE on latent space of TSEE

branches, covering cell-to-cell heterogeneity, as well as latent gene expression pattern, and the results in low-dimensional space generally need to be adjusted manually. For example, the marker *WNT8A* demonstrates an oscillating expression pattern in the TSEE visualization result, but the development tree result in the Supplementary File of [6], which was obtained by a hand-tuned force-direct layout merely shows a growing tendency along time, and the heterogeneity of cells is hidden because cells collapse along branches.

The computational framework of TSEE can also be potentially extended to incorporate other sources of information (e.g., spatial information) of scRNA-seq data. We utilize the scRNA-seq data of embryo cells from [21] to demonstrate the extended application based on spatial information of cells. A set of 84 marker genes are considered to uniquely classify almost every position within the embryo [21]. Therefore, we employ the information from the 84 marker genes to quantify the spatial distance of single cells, defining as $1 - \mathrm{PCC}\left(x_i, x_j\right)$ between samples i and j, where x_i and x_j are the expression vectors of the 84 marker genes. Figure 11 displays the expression of three markers of dorsal ectoderm in 2-dimensional space, as obtained by tSNE, EE, and TSEE, respectively. Compared to the results of EE and tSNE, the highly expressed genes tend to locate along the boundary of the 2-dimensional space by the modified spatial TSEE, showing better consistency with the spatial patterns of the genes in the embryo. The accurate reconstruction of spatial location of single cells will be challenging, and this will be an interesting topic for our future study.

Conclusions

In this study, we propose an efficient algorithm, TSEE, for the visualization of time series scRNA-seq data. TSEE is an extension of EE by introducing an additional repulsive force term when pairs of data points are collected at distinct time stages, thereby balancing the effects from

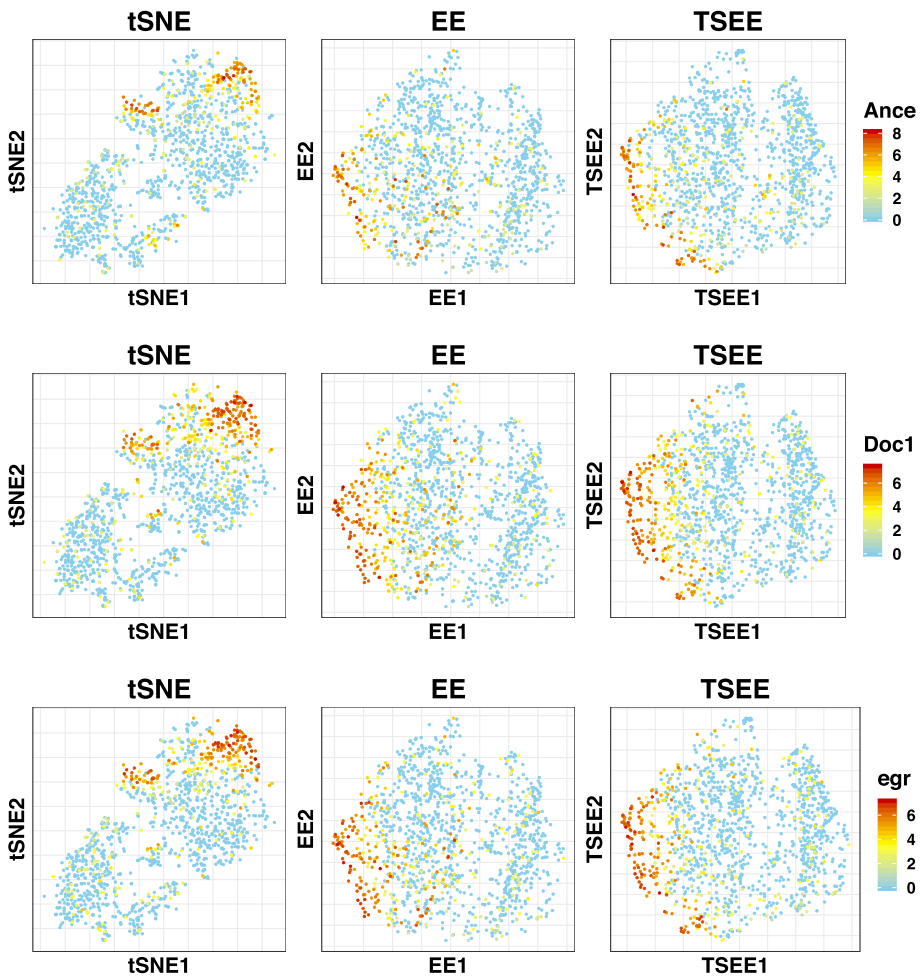

Fig. 11 Distribution of genes expressed in *Drosophila* embryo cells. The expression of *Ance*, *Doc1* and *egr* are displayed in 2-dimensional space, as obtained by tSNE, EE and modified spatial TSEE

disparities between samples in high-dimensional gene expression space.

To incorporate the temporal information of data, TSEE adds additional terms of the time intervals between samples into the repulsive terms to balance the effects from the disparities between samples in high-dimensional gene expression space. In this way, TSEE dilutes the distortions of the assorted sources of variations of the data across time stages and achieves temporal resolution enhancement by preserving temporal order and structure. In addition, TSEE uncovers the subtle dynamic structures of gene expression patterns, as exemplified by oscillating waves in our results, facilitating further downstream dynamic modeling and analysis of gene expression processes. The computational framework of TSEE is generalizable for the incorporation of other sources of information.

Abbreviations

EE: Elastic embedding; HPE: Human preimplantation embryo; IGP: In-group proportion; IGP2: A modification of in-group proportion; PCA: Principal component analysis; PCC: Pearson correlation coefficient; scRNA-seq: Single-cell RNA sequencing; TSEE: Time series elastic embedding; tSNE: t-distributed stochastic neighbor embedding

Acknowledgements

We are grateful to Xiangqi Bai and Ziwei Chen for their helpful discussions.

Funding

Publication of this article was sponsored by NSFC grants (Nos.11571349, 91630314, 81673833), the Strategic Priority Research Program of Chinese Academy of Sciences (CAS) (XDB13050000), the NCMIS of CAS, the LSC of CAS, and the Youth Innovation Promotion Association of CAS. LW would like to thank the Mathematical Biosciences Institute (MBI) at Ohio State University for partially supporting this research. MBI receives its funding through NSF grant DMS 1440386.

Availability of data and materials

The data used in the current study are all publicly available. HPE data can be obtained at https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3929/, while Zebrafish data in UMI count format can be assessed at NCBI GEO with accession no.GSE106587, and the scRNA-seq dataset of *Drosophila* embryos is available at https://shiny.mdc-berlin.de/DVEX/.
The TSEE code is available at https://github.com/ShaokunAn/TSEE, and the oscillation genes discovered by TSEE among the genes related to embryogenesis can be found at https://github.com/ShaokunAn/TSEE/tree/Additional-files.

About this supplement

This article has been published as part of BMC Genomics Volume 20 Supplement 2, 2019: Selected articles from the 17th Asia Pacific Bioinformatics Conference (APBC 2019): genomics. The full contents of the supplement are available online at https://bmcgenomics.biomedcentral.com/articles/supplements/volume-20-supplement-2.

Author's contributions

SA participated in method design and data analysis, wrote the program, and drafted the manuscript. LM conceived the project and participated in method design and writing. LW conceived and designed the study and participated in method design and writing. All authors read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190 Beijing, China. [2]Beijing Institute of Genomics, Chinese Academy of Sciences, 100101 Beijing, China. [3]University of Chinese Academy of Sciences, 100049 Beijing, China.

Published: 10 April 2019

References

1. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381–6. https://doi.org/10.1038/nbt.285.
2. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14(10):979–82. https://doi.org/10.1038/NMETH.440.
3. Setty M, Tadmor MD, Reich-Zeliger S, Ange O, Salame TM, Kathail P, Choi K, Bendall S, Friedman N, Pe'er D. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat Biotechnol. 2016;34(6):637–45. https://doi.org/10.1038/nbt.356.
4. Chen J, Renia L, Ginhoux F. Constructing cell lineages from single-cell transcriptomes. Mol Asp Med. 2018;59:95–113. https://doi.org/10.1016/j.mam.2017.10.00.
5. Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, Reyes AP, Linnarsson S, Sandberg R, Lanner F. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell. 2016;165(4):1012–26.
6. Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science. 2018;360(6392):979. https://doi.org/10.1126/science.aar313.
7. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411. https://doi.org/10.1038/nbt.409.
8. Gao NP, Ud-Dean SMM, Gandrillon O, Gunawan R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. Bioinformatics. 2018;34(2):258–66. https://doi.org/10.1093/bioinformatics/btx57.
9. Rashid S, Kotton DN, Bar-Joseph Z. TASIC: determining branching models from time series single cell data. Bioinformatics. 2017;33(16):2504–12. https://doi.org/10.1093/bioinformatics/btx17.
10. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, Liu S, Lin S, Berube P, Lee L, Chen J, Brumbaugh J, Rigollet P, Hochedlinger K, Jaenisch R, Regev A, Lander E. Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming. bioRxiv. 2017. https://doi.org/10.1101/191056. https://www.biorxiv.org/content/early/2017/09/27/191056.full.pdf.
11. Ding J, Aronow BJ, Kaminski N, Kitzmiller J, Whitsett JA, Bar-Joseph Z. Reconstructing differentiation networks and their regulation from time series single-cell expression data. Genome Res. 2018;28(3):383–95. https://doi.org/10.1101/gr.225979.11.
12. Moon KR, van Dijk D, Wang Z, Burkhardt D, Chen W, van den Elzen A, Hirn MJ, Coifman RR, Ivanova NB, Wolf G, Krishnaswamy S. Visualizing transitions and structure for high dimensional data exploration. bioRxiv. 2017. https://doi.org/10.1101/120378.
13. Carreira-Perpiñán MÁ. The elastic embedding algorithm for dimensionality reduction. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel; 2010. p. 167–74. http://www.icml2010.org/papers/123.pdf.
14. Chen Z, An S, Bai X, Gong F, Ma L, Wan L. Densitypath: a level-set algorithm to visualize and reconstruct cell developmental trajectories for large-scale single-cell rnaseq data. bioRxiv. 2018. https://doi.org/10.1101/276311. https://www.biorxiv.org/content/early/2018/03/05/276311.full.pdf.
15. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PLoS ONE. 2014;9(6):. https://doi.org/10.1371/journal.pone.009867.

16. Weinreb C, Wolock S, Klein AM. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics. 2018;34(7): 1246–8. https://doi.org/10.1093/bioinformatics/btx79.
17. Wasserman L. Topological data analysis. Annu Rev Stat Appl. 2018;5(1): 501–32. https://doi.org/10.1146/annurev-statistics-031017-100045.
18. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. Science. 2010;328(5980):876–8. https://doi.org/10.1126/science.118481.
19. Vladymyrov M, Carreira-Perpiñán MÁ. Partial-hessian strategies for fast learning of nonlinear embeddings. In: Proceedings of the 29th International Conference on Machine Learning (ICML-12). 2012. p. 345–352.
20. Vladymyrov M, Carreira-Perpinan M. Entropic affinities: Properties and efficient numerical computation. In: Dasgupta S, McAllester D, editors. Proceedings of the 30th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 28. Atlanta, Georgia, USA: PMLR; 2013. p. 477–85. http://proceedings.mlr.press/v28/vladymyrov13.html.
21. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP. The Drosophila embryo at single-cell transcriptome resolution. Science. 2017;358(6360):194–9. https://doi.org/10.1126/science.aan323.
22. Kapp AV, Tibshirani R. Are clusters found in one dataset present in another dataset?. Biostatistics. 2007;8(1):9–31. https://doi.org/10.1093/biostatistics/kxj02.
23. Kageyama R, Niwa Y, Isomura A, Gonzalez A, Harima Y. Oscillatory gene expression and somitogenesis. Wiley Interdiscip Rev Dev Biol. 2012;1(5): 629–41. https://doi.org/10.1002/wdev.4.
24. Wylie AD, Fleming J-AGW, Whitener AE, Lekven AC. Post-transcriptional regulation of wnt8a is essential to zebrafish axis development. Dev Biol. 2014;386(1):53–63. https://doi.org/10.1016/j.ydbio.2013.12.00.
25. Morrow ZT, Maxwell AM, Hoshijima K, Talbot JC, Grunwald DJ, Amacher SL. tbx6l and tbx16 are Redundantly Required for Posterior Paraxial Mesoderm Formation During Zebrafish Embryogenesis. Dev Dyn. 2017;246(10):759–69. https://doi.org/10.1002/DVDY.2454.
26. Hagey DW, Muhr J. Sox2 Acts in a Dose-Dependent Fashion to Regulate Proliferation of Cortical Progenitors. Cell Rep. 2014;9(5):1908–20. https://doi.org/10.1016/j.celrep.2014.11.01.
27. Rulands S, Lee HJ, Clark SJ, Angermueller C, Smallwood SA, Krueger F, Mohammed H, Dean W, Nichols J, Rugg-Gunn P, Kelsey G, Stegle O, Simons BD, Reik W. Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. Cell Syst. 2018;7(1):63. https://doi.org/10.1016/j.cels.2018.06.01.