**Rebuttal letter**

**Manuscript Number: CELL-SYSTEMS-D-24-00570  
"Geometric Quantification of Cell Phenotype Transition Manifolds with Information Geometry"**  
  
Dear Dr. Wang,  
  
I'm enclosing the comments that reviewers made on your paper, which I hope you will find useful and constructive. As you'll see, they express interest in the study, but they also have a number of criticisms and suggestions. Based on these comments, it seems premature to proceed with the paper in its current form; however, if it's possible to address the concerns raised with additional experiments and/or analysis, we’d be interested in considering a revised version of the manuscript.   
   
As a matter of principle, I usually only invite a revision when I’m reasonably certain that the authors' work will align with the reviewers’ concerns and produce a publishable manuscript.  In the case of this manuscript, the reviewers and I have make-or-break concerns regarding the need for fair comparison to alternative approaches and more rigorous validation with real world data. Additionally, to move forward, the utility of the approach must be demonstrated by delivering biological insights that are not achievable using other techniques.To further guide revision, I’ve highlighted portions of the reviews that strike me as particularly critical.   
   
As you address these concerns, it's important that you and I stay on the same page.  I'm always happy to talk, either over email or by phone, if you’d like feedback about whether your efforts are moving the manuscript in a productive direction. Do note that we generally consider papers through only one major round of revision, so the revised manuscript would be either accepted or rejected based on the next round of comments we receive from the reviewers.  If you have any questions or concerns, please let me know.  More technical information and advice about resubmission can be found below my signature.  Please read it carefully, as it can save substantial time and effort later.   
  
If any changes to the authorship are needed (addition, removal, or a change in order of authors), we require that an [authorship change form](https://track.editorialmanager.com/CL0/https:%2F%2Fwww.cell.com%2Fpb-assets%2Fjournals%2FCell-Press-authorship-change-form.pdf/1/010f0196f9c139e9-5ca35995-0698-42d2-890c-135fa5d272a6-000000/aI-dImldBhTbvDsDL1StfvjfBm-QpuBpmBsbCJ4uRRU=212) be provided with the revised manuscript that indicates the reason for the change and provides written consent for the change from all authors, including any that were removed.

I look forward to seeing your revised manuscript.  
  
All the best,  
  
Ernesto Andrianantoandro, Ph.D.  
Scientific Editor, Cell Systems

**Reviewers' comments:**  
**Reviewer #1:**

This manuscript by Huang et al. presents SCIM (Single Cell Information Manifolds), a novel method that applies information geometry principles to analyze cell phenotype transitions using single-cell RNA sequencing data. The work presents several noteworthy contributions, including the application of Fisher information metrics to characterize cell state transitions and the use of coarse Ricci curvature to identify critical transition points. While the theoretical foundation is sound and the approach shows promise, several aspects require attention before publication.

**A:** Thanks for the concise summary. We have addressed all your concerns in the revised manuscript.

The following major and minor concerns should be addressed to strengthen the manuscript's impact and usefulness to the field:  
  
**Major Issues:**  
**1.** The validation and benchmarking analysis needs substantial enhancement. The current comparison with existing methods (t-SNE, UMAP, diffusion maps, VAE) is primarily limited to toy datasets. While SCIM shows promise, a systematic comparison/benchmarking against other geometric approaches and trajectory inference methods using real biological datasets would significantly strengthen the manuscript. It would be particularly valuable to demonstrate how SCIM captures real-world biological complexities and nonlinearities that existing methods may miss. The addition of quantitative performance metrics would help establish SCIM's advantages over current approaches.

**A:** Thanks. In this method, we focused analyzing the geometric properties of underlying manifold especially the critical point of cell phenotypic transition. SCIM is not designed for trajectory inference.

Different from package Moncle, WaddingtonOT, Plantir and RNA velocity related methods including scVelo and dynamo(1, 2, 3, 4), SCIM provides a geometric understanding of the cell phenotypic transition manifold. Some of the results are presented with the trajectory inference results like pseudo-time. And the calculation of information velocity is based on the inferred RNA velocity. In another words, SCIM performed the analyses that are downstream or parallel with that in trajectory inference method. Besides the geometric perspective, SCIM can be used for branch detection through analyzing the variation of Fisher information matrix reveals increase of stiff direction around branch (Fig.7). Moreover, Fisher information of genes provides a method for quantifying cell state’s sensitivity to genes (Fig. 3).

To the best of our knowledge, there is no such geometric method that do the same analysis as SCIM. Both SCIM and VAE embed single cell vector as Gaussian distributions. The latent distribution in VAE is primarily for generative modeling, whereas in SCIM, it is intended to preserve the local geometric structure. Besides comparing VAE on toy data, we also compared the embeddings of scVI and SCIM on dentate gyrus neurogenesis data. The continuous variation of pseudo-time is preserved by SCIM embeddings but not maintained in the VAE embeddings (Fig. S5). Overall, SCIM is a completely different method compared to VAE.

Currently, there are several methods for calculating curvature. In Tram Huynh’s work(5), they also use the notion of coarse Ricci curvature and found that transition cells exhibit low curvature which is consistent with our analysis though the methods of calculating curvature are different. Duluxan Sritharan proposed an alternative method for estimating the curvature of underlying manifold in single cell RNA-seq data (6). However, their definition of curvature is extrinsic that is different from our intrinsic definition. In SCIM, both CRC and information velocity reveals the transition cell. Hence, we compared several methods dedicated to identifying transition cells, including scGeom, scTite, and CellTran(5, 7, 8). These three algorithms were tested on dentate gyrus, endocrinogenesis, epithelial-mesenchymal transition of A549 cells, and CellTran simulated datasets.

scGeom works on cell-cell networks by computing Ollivier-Ricci curvature i.e. coarse Ricci curvature (CRC) and persistent homology to characterize geometric and topological properties of the network. CRC and total persistence are used to distinguish transition cells from stable cells. These features are subsequently employed as inputs for training cell classifiers, demonstrating their ability to capture biologically meaningful information. As reported, CRC decreases at transition cells, while total persistence increases at these regions. However, when applying scGeom to CellTran simulated data, we didn't observe low curvature at transition cells; in fact, CRC values were lower in stable cells than in transition cells (Fig.R1b). In contrast, the CRC computed by SCIM successfully distinguished transition and stable cells in the same simulated data (Fig.R1e). Furthermore, results on real datasets revealed that scGeom's CRC often exhibit low resolution and occasionally misidentify transition cells (Fig.R1b). For instance, CRC results on dentate gyrus show that immature granules are more likely to be transition cells than neural blast cells. Notably, although both follow Ollivier's methodology, calculation of CRC in scGeom is based on transition probability on discrete cell-cell networks, whereas SCIM's CRC operates on the continuous Gaussian embeddding. The results of total persistence can be found in Fig.S11.

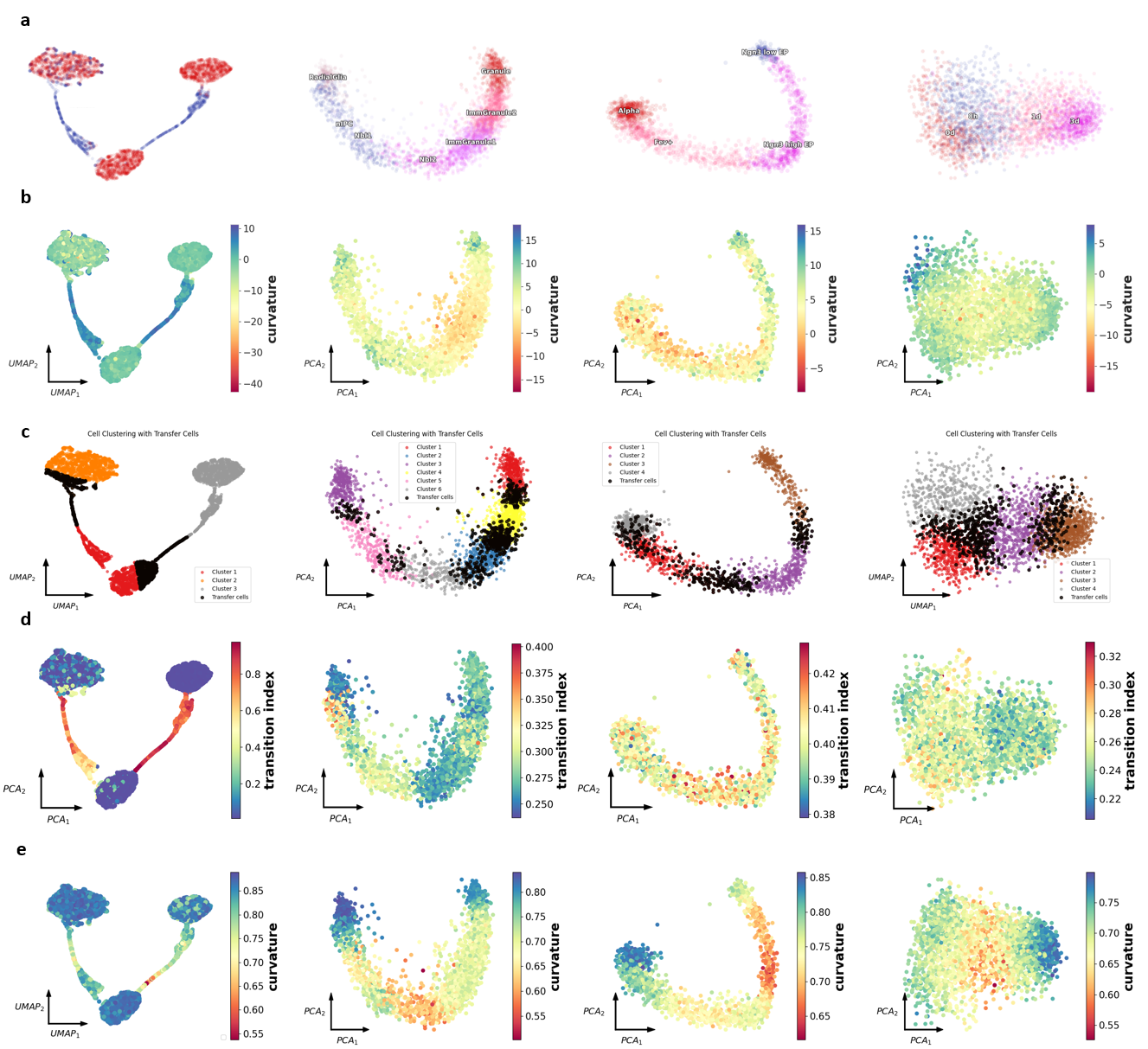


Fig.R1 Comparison of different transition cells identification methods

Dataset: CellTran simulated data, Dentate Gyrus (DG), Endocrinogenesis α-branch (EG), A549 emt (A549)

a. Cell type annotation. For CellTran simulated data, stable cells are colored in red, while transition cells are in blue.

b. ORC calculated via scGeom

c. Results of scTite. Transition cells are colored in black.

d. Transition index calculated via CellTran.

e. CRC calculated via SCIM.

scTite distinguishes transition cells using transition entropy. Specifically, it first applies dimension reduction to scRNA-seq data and apply Gaussian Mixture Model (GMM) soft clustering. Each cell is assigned a discrete probability distribution representing cluster membership, and the entropy of this distribution—transition entropy—quantifies uncertainty in cluster assignment. Cells with higher entropy are identified as transition cells due to their mixed identity (Fig.R1c). However, evaluations on CellTran-simulated and real datasets revealed that scTite is highly sensitive to dimension reduction methods (Fig.S12) and it just identifies cluster-boundary cells as transition cells.

CellTran defines a transition index based on gene expression correlation between cells and their neighbors. While this index successfully identified transition cells in CellTran's own simulated data, it performed inconsistently on real datasets (Fig.R1d). Only in EG data did the transition index yield biologically plausible results. For dentate gyrus data and A549\_emt datasets, high transition index values appear in cells at terminal stages of the developmental process (Fig.R1a, d). This result contradicts biological intuition, as terminal developmental stages typically represent stable cells rather than transitional states.

We further validated SCIM on CellTran-simulated data and scGeom-provided Olsson and iPSC real datasets. Results demonstrated that SCIM’s CRC reliably recapitulates CellTran and scGeom findings while accurately identifying transition cells across all tested datasets. (Fig.R2, R3)

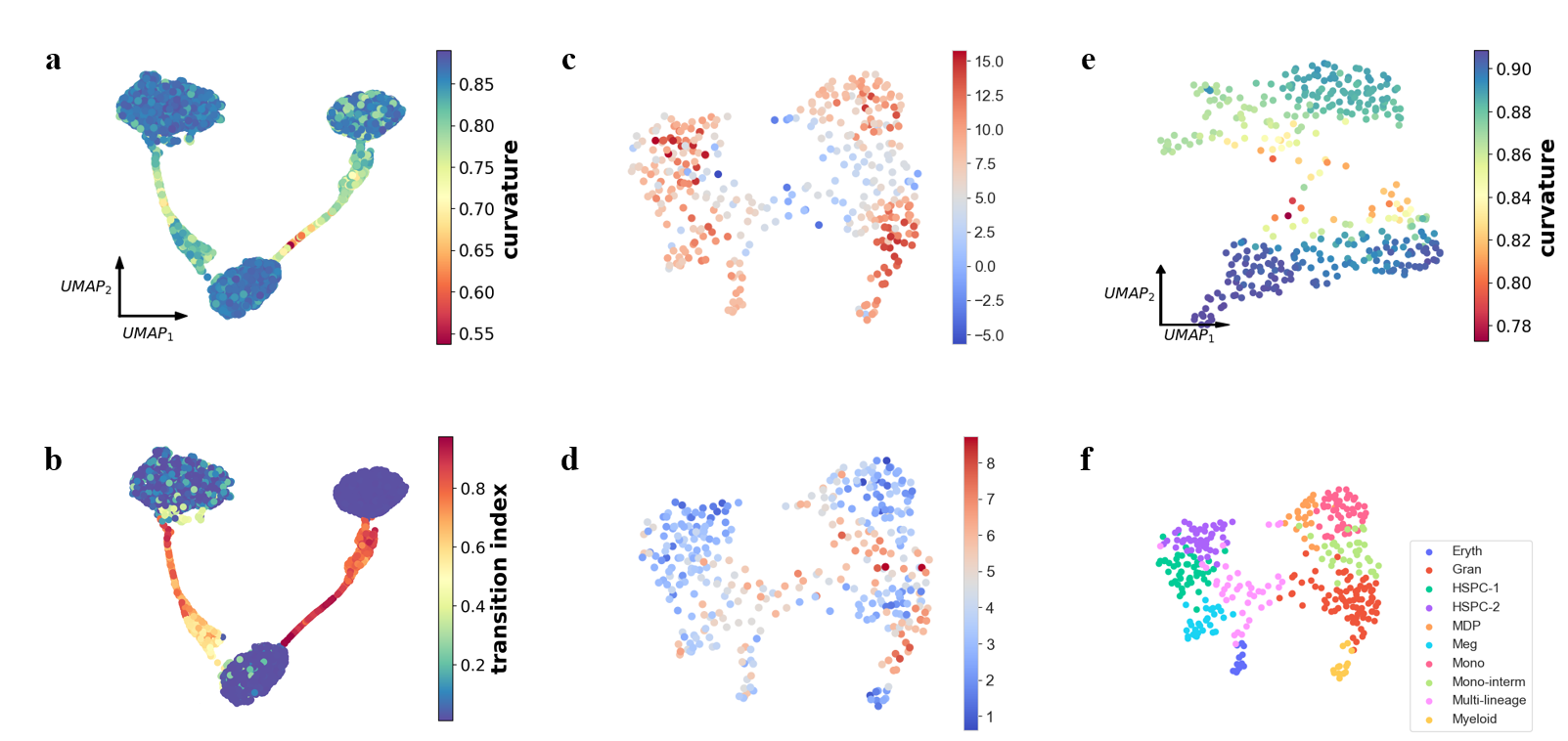


Fig.R2 SCIM on CellTran simulated data and Olsson （used in scGeom） dataset

a. CRC on CellTran simulated data. b. Transition cells and stable cells annotation on CellTran simulated data. c. scGeom ORC on Olsson data. d. scGeom H0 total persistence on Olsson data. e. CRC on Olsson data. f. Cell type annotation of Olsson data. Transition cells are Multi-lineage and Mono-interm

c, d and f are from the scGeom paper.

(这里b,e要重新画结果！)

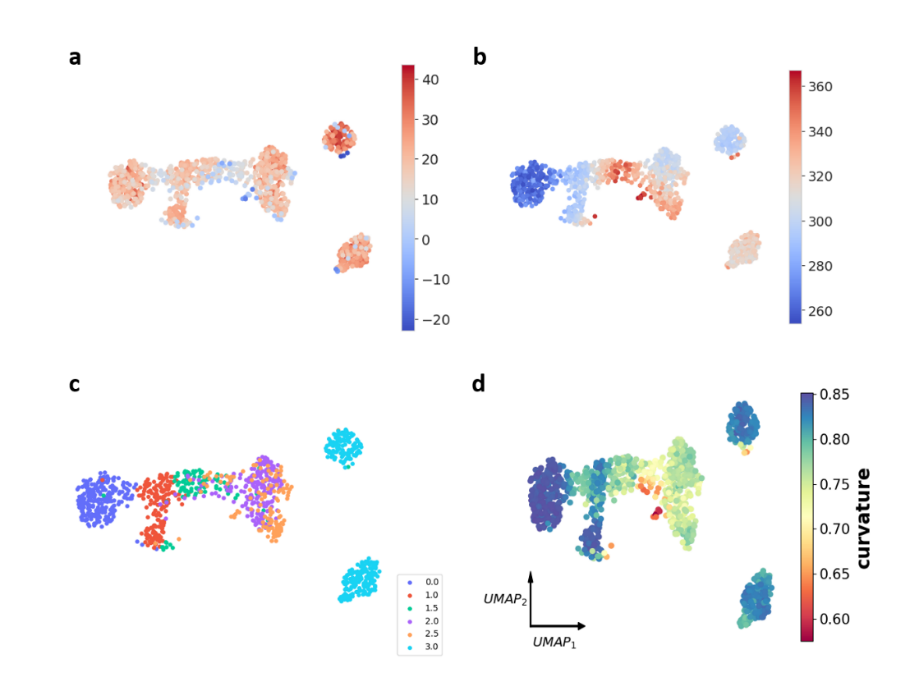
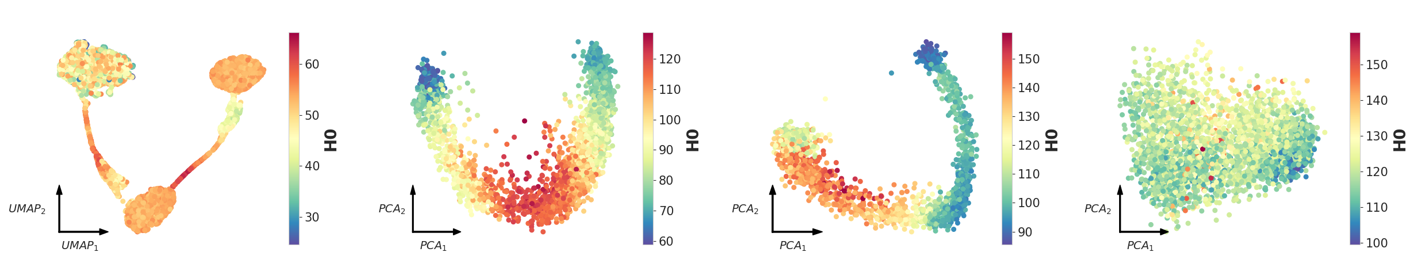


Fig.R3 SCIM on iPSC (used in scGeom) dataset

a. scGeom ORC on iPSC data. b. scGeom H0 total persistence on iPSC data. c. Time annotation of iPSC data. Transition cells are 1.5 and 2.5. d. CRC on iPSC data.

a, b, c are from the scGeom paper.

Fig.S11 scGeom H0 total on CellTran simulated data, DG, EG, A549.

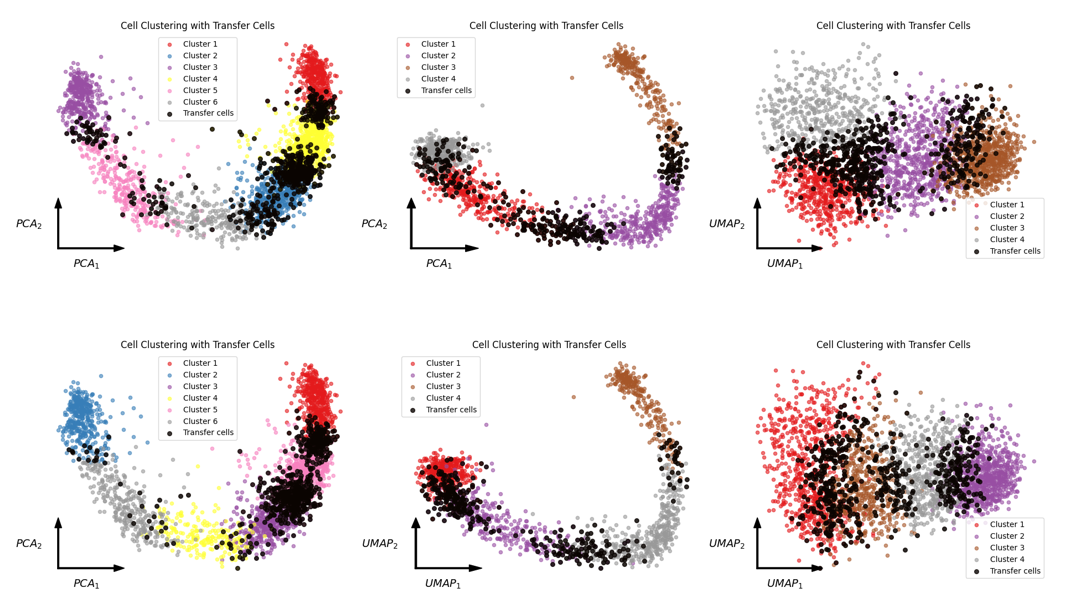


Fig.S12 scTite results using different dimension reduction methods.

Transition cells are colored black.

Upper panel: reduce dimension via UMAP; reduce dimension via Gaussian embedding on DG, EG, A549.

Following your advice, we developed a simple method for cell type classification and compare its prediction with the pre-defined cell types. In addition, we also compared it with several other clustering method based on geometric concept.

Through Gaussian embedding, we get the Gaussian distribution and corresponding Fisher information matrix (FIM) of each cell, and then use the diagonal element of the FIM of each cell as a vector to represent the cell. We use the vectors’ cosine distance to perform agglomerative hierarchical clustering to obtain the clustering results.

Agglomerative hierarchical clustering is a bottom-up clustering algorithm. It starts by treating each data point as an independent cluster, and then iteratively merges the two most similar clusters until all data points are finally merged into one large cluster, or a preset stopping condition is reached (here we give the number of clusters).

We compared various clustering methods on three real single-cell datasets, namely Leiden, Louvain, DTNE and multiscale Phate algorithms (Fig. R6)(9, 10, 11, 12). The comparison results with ground truth show that SCIM can get clustering results very well. On these datasets, the various indicators of the Leiden and Louvain algorithms are significantly higher than those of other algorithms, which is related to the annotation method of the ground truth of the original data (The ground annotation results of these datasets were implemented using the Louvain algorithm)(14,15,16).

Rand index (RI) evaluates the match between the clustering results and the true label by statistically analyzing the distribution consistency of sample pairs. Mutual Information (MI) measures the degree of information sharing between the clustering results and the true label based on information entropy. ARI and AMI are the results of their correction based on the expected value of the random distribution cluster label assumption. F1-score is the harmonic mean of recall and precision. Larger is better for these clustering performance metrices.

Here are clustering results and comparisons:

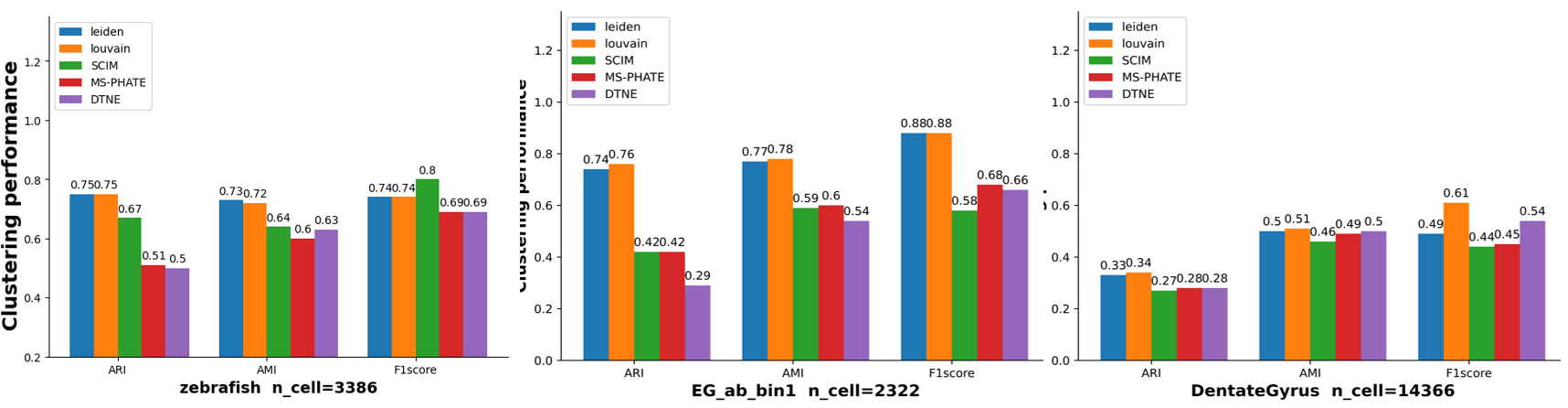


Fig.R6 Clustering results

Clustering performance of 5 different clustering methods on three real single-cell datasets with ARI, AMI and F1-score as performance metrices.

**2.** The manuscript requires a more thorough assessment of how noise and technical variations in scRNA-seq data affect Fisher information calculations. A robust demonstration that the Fisher information-related calculations, particularly the identification of 'stiff directions,' remain stable despite typical noise levels in single-cell data would significantly strengthen the method's credibility.

**A:** Thanks for this advice. In SCIM, the primary purpose is to preserve the relative distance between neighboring cells. Since the Gaussian distributions are derived from training results of neural network, there are some randomness in the obtained embedding. We found that distribution of CRC is barely affected by different Gaussian embeddings (Fig.5c). However, the values of Fisher information of each dimensionality vary in different runs. If we choose a large value of latent dimensionality, there are more choice of embeddings can that preserve the relative distance, which will introduce randomness in analyzing the stiff directions.

Thanks for pointing this out.

Thus we explore a more robust way for identifying branch point. The principle is still increase of stiff directions at branch points. First, the parameters at branch points should contribute to variation of distance between neighboring cell. Second, when stiff direction varies, some parameter’s Fisher information start to increase and some parameters’ Fisher information start to decrease, which reveals negative correlation between those parameters. Hence we use both criteria to identify the branch points. We found that the latent dimensions that exhibit large Wasserstein distances and negative correlation strength with other dimensions peak at branch points.

For each single cell, we calculated the contribution of each latent dimension in Wasserstein distance between each cell and its neighboring cells. And the negative correlation strength of each latent dimension. The negative correlation strength is defined as the total value of negative correlation with other dimensions. For each dimensionality, if the peak of negative correlation strength and its contribution on Wasserstein distance both peaks at the region of low CRC, it can be account as variation of stiff parameters. With this criteria, we identify the branch points

In Fig. S6, we varied the number of latent dimensions in the Gaussian embedding neural network, the number of k-nearest neighbors in building the cell graph, and *khop*in sampling triplets from the graph. The correspondence between CRC and information velocity remained unaffected. Additionally, we employed a different gene selection method called DUBstepR (Fig. S8) (13).

We compared the influence of different distance metrics for constructing the cell graph, and different random seeds on the results (Fig. S7 top and middle). Additionally, we performed subsampling on the dataset (Fig. S7 bottom). By randomly selecting different proportions of cells from the original dataset, we calculated the CRC and information velocity. The results were essentially the same as those obtained from the full dataset.

Following your advice, we test adding noise and subsampling on identifying stiff directions (Fig. Sx).

**Minor Issues:  
1.** The notation in the Fisher information matrix definitions (page 32 and Figure 1) switches between and parameters. While mathematically trained readers may understand what does it mean, an explicit statement that encompasses would improve clarity. Besides this, some other technical terms need better introduction for the broader biological audience.

**A:** Thanks for the question. θ is the general symbol representing parameters. And represents in Gaussian embedding. We have added more explanation on the technical terms in Page12 and caption of Fig.1 following your suggestion.

**2.** About the Gaussian embedding, the parameters selected: (L=10, k\_nei=10, k\_hop=2) needs better justification. It would be helpful if there could be a general guideline for parameter selection in different scenarios are needed.

**A:** Thanks. We have updated a [Readme.md](https://github.com/wwklab/SCIM/tree/main) file on Github for parameter selection.

The following is a brief explanation:

K\_hop and k\_nei are closely related to the total number of cells n:

K\_hop will affect the approximation of the neural network loss function, so they should not be too large. Usually, 2, 3, or 4 are selected. After determining k\_hop, k\_nei is selected using the following formula:

Finally, according to the original literature, the selection of L is determined by the number of dimensions where the sigma remains stable after overfitting. According to our prior tests, 6 or 10 are generally selected.

**3.** Several key citations are missing for the comparative methods. For example, the UMAP citation (McInnes et al. 2018) should be included when discussing this method.

**A:** Thanks. We have added this references and other references.

**4.** The Gene Ontology (GO) enrichment analysis in Figure 5 requires more detailed methodology, including specification of statistical parameters such as p-value cutoffs and multiple testing correction procedures.

**A:** Thanks. We added the details of GO enrichment analysis to the revised manuscript.

**5.** Last but not least, while the mathematical and physical foundations of the method are solid, the biological significance of the geometric features (curvature, information velocity) needs clearer exposition. It would be helpful to have more discussion of how the method relates to existing biological knowledge. Additional evidence (to explore other biological references) demonstrating that these features capture meaningful biological phenomena rather than technical artifacts would enhance the manuscript's impact.

**A:** Thanks for your insightful comments and questions. Here, we provide some clarifications:

**1. Biological Association Between Low Curvature and Transitory Cell States**

**Low CRC regions correspond to critical transition points for cell fate determination:** During dentate gyrus neurogenesis, low curvature regions (Fig. 5b box ii) connect two stable states of radial glia cells and granule cells, which highly aligns with the concept of "transitory states" in developmental biology. In fact, from a physical perspective, we consider that stable states of a system correspond to high-curvature, low-energy regions/attractors, while unstable transitory states correspond to low-curvature saddle points. If we view CPT as a system controlled by gene regulatory networks, its stable states correspond to stable cell phenotypes, and unstable transitory states correspond to transitory state cells, namely low CRC regions. Therefore, the biological "transitory states" and the low curvature regions emphasized in this paper are essentially two sides of the same concept—the former describes intermediate state cells connecting two stable cell states in CPT from a dynamical perspective, while the latter approaches it from a geometric angle. Identifying transitory states is critical for control cell fate and identify early warning signal of disease. Several studies like scGeom, scTite, and CellTran also explore the transitiory states from different perspective.

**2. Biological Interpretation of Information Velocity and Cell State Variation Rate**

**Single cell is a high-dimensional dynamical system. Through embedding single cell state as probability distribution, the information velocity provides a measure for quantifying the rate of cell state variation.**

**Quantifying the rate of cell state changes**: Information velocity reflects the actual rate at which cell states undergo transitions. In gene expression space, we have RNA velocity and other dynamical quantities that well describe the rate of changes in cellular gene expression patterns. However, describing the rate of gene expression pattern changes is not equivalent to describing the rate of cell state changes, because CPT is constrained on a low-dimensional manifold rather than in Euclidean space. Directly calculating the magnitude of RNA velocity or observing the rate of change in key gene expression levels cannot fully quantify the rate of cell state changes. Information velocity, through the Fisher metric on SCIM, recalculates the magnitude of RNA velocity. By considering the metric on the manifold where CPT resides, rather than simply treating it as Euclidean space, information velocity can more accurately describe the rate of cell state changes. As shown in Fig. 6, high information velocity regions precisely correspond to critical periods of cell type transitions, while the magnitude of RNA velocity cannot achieve this. And information velocity has been already used thermodynamics and non-linear dynamics.

**3. Fisher Metric Off-diagonal Elements and Gene Module Switching**

As shown in Fig. 5e, eigengenes form a relay pattern, with the increase of EG3 closely following the decrease of EG1 and EG2, reflecting the switching of gene modules during development. As shown in Fig. 6d, the switching of EGs precisely corresponds to low curvature regions and high information velocity regions, indicating that geometric feature changes in SCIM correspond to the activation/inhibition of key regulatory gene modules.

**Reviewer #2:**

The paper proposes a new geometry-based method that reconstructs the underlying manifold for cell phenotype transition. My major concern is: what does the new method bring to the field? Is there any downstream analysis that would not have been possible without the new method? For indicating high velocity regions, do other velocities (like RNA velocity) fail to do so?

**A:** Thanks for your insightful comments and questions. Here, we provide some clarifications:

In this work, we present a method for quantifying the underlying manifold of cell phenotypic transition where Fisher information matrix is defined as Riemann metric. The original definition of Fisher information reflects the sensitivity of probability density function to the change of parameter. By calculating the Fisher information of eigengenes and genes, the sensitivity of cell state to the change of eigengene and genes can be quantified. As far as we know, there is no such method for quantifying this kind of sensitivity. Fisher information reflects the probable influence of genes on the cell state. Even though it doesn’t indicate causal relation, it still can be used to filtering candidate genes that can change cell state.

Information velocity quantifies the rate of probability distribution variation. As we use Gaussian distribution to represent cell state, the information velocity reflects how fast the cell state changes. From its definition, the magnitude of information velocity is affected by both Fisher metric and parameter velocity. The parameter velocity is calculated with RNA velocity. Information velocity is related with the curvature of the manifold. But we didn’t observe the association between curvature and norm of RNA velocity (Fig.Sx). Norm of RNA velocity can measure h**ow fast a cell's transcriptome is changing. In Riemann geometry, calculation of norm of a vector need the Riemannian metric. Norm of RNA velocity simply relies the hidden assumption that the space is Euclidean. Hence, in our understanding, the transcriptome doesn’t equal with cell state.**

In addition, the paper is not written carefully. Two examples:  
  
**1.** Riemann metric -> Riemannian metric? ...

**A:** Thanks. we have revised Riemann metric/geometry/manifold to Riemannian metric/geometry/manifold.

**2.** A few formulas seem to be inserted as low-quality images, for example the blurry Gaussian distribution in the first paragraph of the Result section.

**A:** Thanks. We have re-edit the formulars.  
  
And how does it connect to other similar works in the field? A careful literature review is definitely necessary. For instance:

**1.** Gaussian embedding: differ from, for instance, scEGOT (<https://www.biorxiv.org/content/10.1101/2023.09.11.557102v1>)?

**A:** Thanks. In SCIM, each cell is embedded as a multi-variate Gaussian distribution. While in scEGOT, the mixture of Gaussian distribution is used to represent a population of cells.

scEGOT processes time series data. When processing single time point data, pseudotime is required to divide the data. However, pseudotime can be omitted in the data processed by SCIM.

At each time point, scEGOT uses Gaussian mixture distribution to group data, and each Gaussian distribution represents a cell group. The purpose of SCIM is to find the low-dimensional manifold on which the gene expression data of all cells depends. According to the theory of Fisher and Rao, the parameterized probability distribution with FIM as the metric matrix constitutes a Riemann manifold (called a statistical manifold). Therefore, SCIM maps each cell into a Gaussian distribution through Gaussian embedding. Note that each Gaussian distribution here represents a cell, which is also a point on the statistical manifold. The meaning of Gaussian distribution is different from that in scEGOT.

Between adjacent time points, since the data at each time point is represented by a Gaussian mixture distribution, scEGOT obtains the transfer path of two Gaussian mixture distributions by solving a continuous time optimal transmission problem and thus obtains the transfer probability between cell groups and the cell speed of each cell. The information velocity in SCIM refers to the speed of change of the cell. It is a scalar whose calculation depends on the above-mentioned metrics and RNA velocity.

**2.** How does information velocity differ from RNA velocity (even just visually)? And when splicing data is unavailable, is it possible to replace RNA velocity by optimal-transport velocity?

**A:** Thanks. The information velocity is the rate of change of probability distribution. More specifically, information velocity is a scalar, while RNA velocity is a multi-dimensional vector in transcriptome space. And as mentioned above, magnitude of RNA velocity is the information velocity when the manifold is Euclidean.

The calculation of information velocity requires the time derivative of parameters i.e. parameter velocity. Parameter velocity can be expressed via chain rule where is the RNA velocity. For instance, For the eigen-gene or PCA coordinates (), the parameter velocity is defined as , where  is the weight of velocity gene in the corresponding coordinates.

The calculation of cell velocity in scEGOT doesn’t need splicing data, so it can be an alternative to RNA velocity. But it can’t calculate cell velocity of the last cell type. Again, because scEGOT needs pseudotime when there is no real timestamp, it has to use a method other than RNA velocity to get pseudotime first.

**3.** There is no thorough comparison between the embeddings presented in this work and many other methods.

**A:** Thanks. In this method, we employed a probability embedding for single cell RNA sequencing data. And we have focused on geometric properties of the underlying manifold especially characterizing critical points in cell phenotypic transition. In previous version, we already compared the embedding of SCIM and other methods like UMAP and diffusion map.

As far as we know, there is rare method on analyzing geometric properties. Here, we compared SCIM with methods for studying transitory cells including scGeom and cellTrans.

We further developed a clustering method based on SCIM and compared it with other clustering methods based on probability embedding like DTNE and ms-PHATE. Here are clustering results and comparisons:

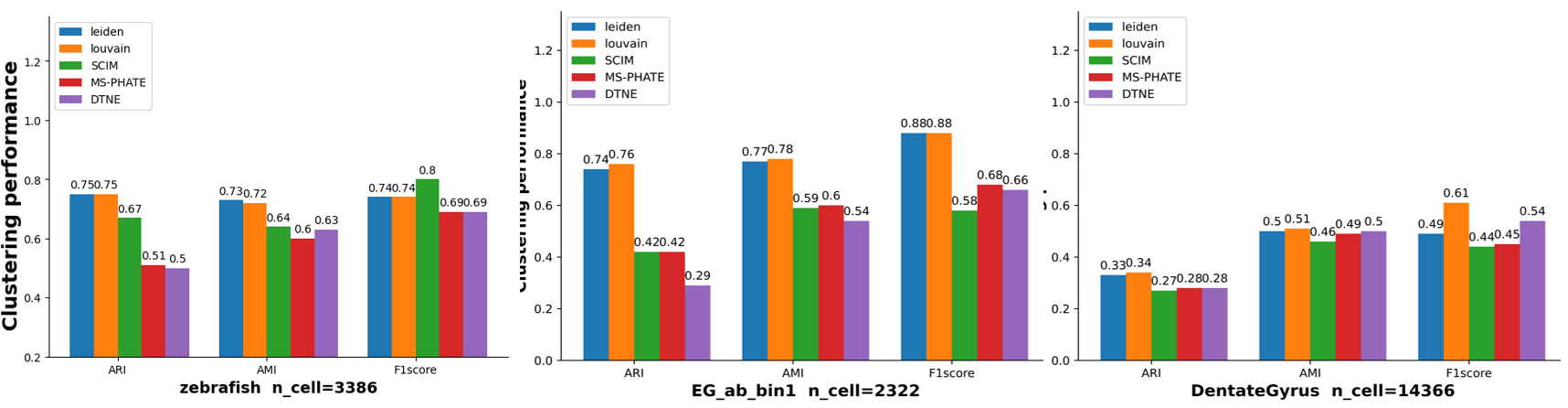


Fig.R6 Clustering results

Clustering performance of 5 different clustering methods on three real single-cell datasets with ARI, AMI and F1-score as performance metrices.

1. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381-6.

2. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, et al. Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. Cell. 2019;176(4):928-43.e22.

3. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe’er D. Characterization of cell fate probabilities in single-cell data with Palantir. Nature Biotechnology. 2019;37(4):451-60.

4. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.

5. Huynh T, Cang Z. Topological and geometric analysis of cell states in single-cell transcriptomic data. Briefings in Bioinformatics. 2024;25(3).

6. Sritharan D, Wang S, Hormoz S. Computing the Riemannian curvature of image patch and single-cell RNA sequencing data manifolds using extrinsic differential geometry. Proc Natl Acad Sci U S A. 2021;118(29).

7. Gan Y, Guo C, Guo W, Xu G, Zou G. Entropy-based inference of transition states and cellular trajectory for single-cell transcriptomics. Briefings in Bioinformatics. 2022;23(4).

8. Wang Y, Dede M, Mohanty V, Dou J, Li Z, Chen K. A statistical approach for systematic identification of transition cells from scRNA-seq data. Cell Reports Methods. 2024;4(12).

9. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports. 2019;9(1):5233.

10. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008;2008(10):P10008.

11. Wei J, Zhang B, Wang Q, Zhou T, Tian T, Chen L. Diffusive topology preserving manifold distances for single-cell data analysis. Proceedings of the National Academy of Sciences. 2025;122(4):e2404860121.

12. Kuchroo M, Huang J, Wong P, Grenier J-C, Shung D, Tong A, et al. Multiscale PHATE identifies multimodal signatures of COVID-19. Nature Biotechnology. 2022;40(5):681-91.

13. Ranjan B, Sun W, Park J, Mishra K, Schmidt F, Xie R, et al. DUBStepR is a scalable correlation-based feature selection method for accurately clustering single-cell data. Nature Communications. 2021;12(1):5849.

14. Bastidas-Ponce A, Tritschler S, Dony L, Scheibner K, Tarquis-Medina M, Salinno C, Schirge S, Burtscher I, Böttcher A, Theis FJ, Lickert H, Bakhti M. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. Development. 2019 Jun 17;146(12):dev173849. doi: 10.1242/dev.173849. PMID: 31160421.

15 La Manno, G., Soldatov, R., Zeisel, A. et al. RNA velocity of single cells. Nature 560, 494–498 (2018). https://doi.org/10.1038/s41586-018-0414-6

16 Lauren M SaundersAbhishek K MishraAndrew J AmanVictor M LewisMatthew B ToomeyJonathan S PackerXiaojie QiuJose L McFaline-FigueroaJoseph C CorboCole TrapnellDavid M Parichy (2019) Thyroid hormone regulates distinct paths to maturation in pigment cell lineages eLife 8:e45181..