Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spectral Signatures Reveal a Transient Commitment Bottleneck During Pancreatic Endocrine Differentiation

Hong Liu
Northwestern Polytechnical University
Na Miao
Sichuan University
Yangyang Song
Tongji University
Wanli Yang
Fourth Military Medical University
Liu Hong
Fourth Military Medical University
Wei Liu
Xi'an University of Science and Technology
Zhi-Peng Li
iamzpli@nwpu.edu.cn

Northwestern Polytechnical University

Article

Keywords: Single-cell transcriptomics, Spectral analysis, Lineage commitment, Pancreatic endocrine development, Collective dynamics

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10839985/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spectral Signatures Reveal a Transient Commitment Bottleneck During Pancreatic Endocrine Differentiation

Hong Liu[1&], Na Miao[2&], Yangyang Song[3], Wanli Yang[4], Liu Hong[4*], and Wei Liu[5*], Zhi-Peng Li[1*]

[1]State Key Laboratory of Flexible Electronics (LOFE) & Institute of Flexible Electronics (IFE), Northwestern Polytechnical University, Xi'an, China

[2]West China Hospital of Sichuan University, Sichuan University, Sichuan, China

[3]Putuo People's Hospital, School of Life Sciences and Technology, Frontier Science Center for Stem Cell Research, Tongji University, Xi'an, China.

[4]Department of Digestive Surgery, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an, China

[5]College of Sciences, Xi'an University of Science and Technology, Xi'an, China

[&] These authors contributed equally to this work and should be considered co-first authors.

[*]Correspondence author (s); E-mail (s):

Liu Hong, hongliu180@126.com

Wei Liu, weiliu@xust.edu.cn

Zhi-Peng Li, iamzpli@nwpu.edu.cn

Abstract

Single-cell transcriptomics reconstructs developmental trajectories, but quantitative frameworks for population-level transcriptional reorganization during lineage commitment remain underdeveloped, limiting systems-level insights into cell fate specification. Here we introduce a spectral framework based on singular value decomposition of the cell-by-gene expression matrix, applied to murine pancreatic endocrine development (GSE132188). We derive two collective observables: R_{eff}, the effective number of active modes quantifying the dimensionality of collective transcriptional states, and R_3, a third-order statistic measuring mode-weight distribution asymmetry. Along pseudotime, endocrine differentiation exhibits a transient R_{eff} minimum at the *Ngn3*-high state, marking a low-dimensional population-wide commitment bottleneck. R_3 rises during *Fev*+ and terminal differentiation, reflecting asymmetric stabilization of lineage-specific regulatory programs during fate diversification. A random-order null model reduces these signatures, confirming they reflect coherent developmental progression rather than sampling effects. The leading eigenmodes recover biologically interpretable transcriptional programs underlying progenitor-to-mature endocrine progression and transient endocrine commitment. Together, these results support cell-fate specification as an ordered, collective reorganization in high-dimensional expression-state space, and demonstrate that physics-inspired spectral observables complement standard single-cell trajectory analysis to uncover system-level developmental principles.

Keywords: Single-cell transcriptomics, Spectral analysis, Lineage commitment, Pancreatic endocrine development, Collective dynamics

1. Introduction

Recent advances in single-cell RNA sequencing (scRNA-seq) have enabled the reconstruction of developmental trajectories at cellular resolution[1,2]. In the pancreas, classical studies have delineated a hierarchical model of endocrine differentiation, whereby multipotent epithelial progenitors through a transient *Neurog3*-dependent commitment checkpoint prior to diverging into functionally distinct, hormone-secreting islet cell types[3–5]. Recent single-cell transcriptomic studies have further refined this model, resolving a continuous developmental trajectory that progresses from multipotent progenitors via *Ngn3*-high endocrine precursors and *Fev*-expressing intermediate states to terminal α, β, δ, and ε cell identities[6–8]. While these studies have generated high-resolution molecular atlases of endocrine lineage progression, prevailing analytical frameworks remain centered on cell-type clustering, marker-gene annotation, and pseudotime ordering of individual cells. These approaches excel at defining discrete cellular states but fall short of quantifying the global, collective reorganization of the cell population as it traverses the commitment transition.

From a systems biology perspective, cellular differentiation is not simply a sequential activation of lineage-specific marker genes, but a coordinated collective transition of the cell population within a high-dimensional transcriptional state space. Dynamical systems theory frames lineage commitment as passage through a "regulatory bottleneck" —a critical transition state where cell fate plasticity is rapidly restricted[9–10]. In this framework, multipotent progenitors exhibit high transcriptional entropy, reflecting promiscuous, low-level activation of multiple lineage programs and broad developmental potential[11]. As cells enter the commitment bottleneck, core lineage-specific gene regulatory networks are coherently activated, suppressing alternative fate programs and reducing transcriptional uncertainty. While this conceptual model is widely accepted, most supporting evidence derives from single-cell information-theoretic metrics that assess cell-intrinsic potency, rather than quantifying the global structural organization of the entire cell population across this transition. It remains an open question whether pancreatic endocrine differentiation proceeds through a transiently compressed collective state—a critical point of minimal population-level transcription diversity—prior to re-expanding into the spectrum of terminally differentiated endocrine cell fates.

To address this gap, we build on the eigen-microstate approach developed by Xiaosong Chen and collaborators, which interprets the leading modes and their spectral weights as physically meaningful descriptors of collective organization in complex systems[12–14]. Unlike conventional applications that treat principal component analysis (PCA) solely as a

dimensionality reduction tool, this approach links changes in the singular value spectrum to collective reorganization, finite-size effects, and transition-like behaviour[15–21]. Our recent work extended this framework by introducing third-order spectral signatures for pseudo-transitions in finite systems, enriching the physical interpretation of mode-weight statistics beyond standard variance ranking[22]. Motivated by this perspective, we treat the cell-by-gene expression matrix as a population-state ensemble and decompose it via singular value decomposition (SVD). From the normalized spectral weights wα, we define two observables: R_{eff}, which measures the effective dimensionality of collective states, and R_3, a normalized third-order statistic derived from the second and third cumulants of the spectral-weight distribution.

$$R_{\mathrm{eff}} = \frac{1}{\sum_{\alpha} w_{\alpha}^{2}}, \tag{1}$$

$$R_3 = \frac{K_3}{K_2^3}, \tag{2}$$

Applying this framework to the endocrine lineage of the publicly available murine pancreatic development dataset GSE132188, we identify a transient low-dimensional bottleneck tightly aligned the *Ngn3*-high commitment stage, followed by a marked increase in third-order spectral asymmetry that tracks *Fev*-dependent lineage diversification. Our findings provide quantitative, population-level evidence for a collective commitment transition during pancreatic endocrine development, demonstrating how physics-inspired spectral metrics can uncover systems-level principles of cell differentiation.

2. Results

Figure 1 provides an overview of the spectral analysis framework. Endocrine-lineage single-cell transcriptomes from GSE132188 were assembled into a cell-by-gene expression matrix and decomposed via SVD. Two spectral observables—R_{eff} (effective dimensionality) and R_3 (third-order asymmetry)—were computed from the resulting mode-weight distribution and tracked along diffusion pseudotime in sliding windows to characterize how collective transcriptional organization evolves during endocrine differentiation.

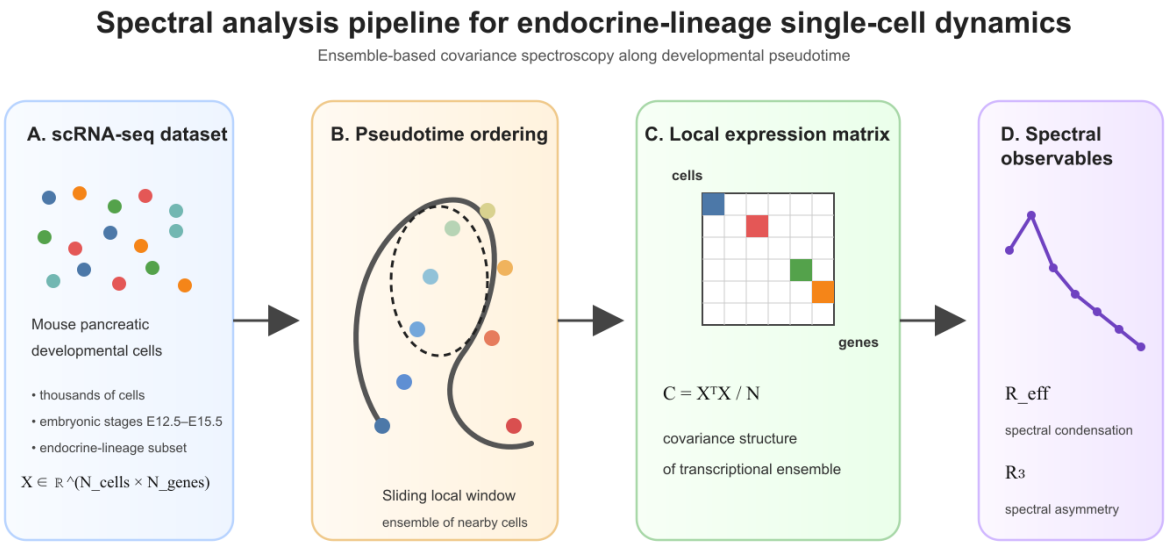


Figure 1. Workflow of spectral analysis of endocrine differentiation: Publicly available single-cell RNA sequencing data from mouse pancreatic development (GSE132188) were first filtered to retain endocrine-lineage cells, including *Ngn3*-low progenitors, *Ngn3*-high endocrine precursors, *Fev*$^{+}$ intermediates, and terminal endocrine populations. After preprocessing and highly variable gene selection, a cell-gene expression matrix was constructed, where rows correspond to cells and columns correspond to genes. SVD was then applied to the normalized matrix to obtain collective transcriptional modes and their spectral weights. Developmental ordering was inferred using diffusion pseudotime, and sliding-window spectral analysis was subsequently performed along pseudotime to quantify the evolution of collective transcriptional organization through endocrine differentiation.

Global endocrine manifold and developmental ordering

We extracted endocrine-lineage cells from GSE132188 using the original annotations, including *Ngn3*-low endocrine progenitors, *Ngn3*-high early and late states, *Fev*$^{+}$ intermediates, and terminal α, β, δ, and ε populations. UMAP visualization revealed that these cells form a continuous, connected transcriptional manifold rather than segregating into discrete, isolated clusters, consistent with a gradual, progressive differentiation process (Fig. 2).

Diffusion pseudotime increased smoothly along this manifold, with early pseudotime enriched in progenitor-like cells and late pseudotime dominated by differentiated endocrine states (Fig. 3). The distribution of annotated cell types across pseudotime bins was temporally ordered in a manner consistent with the inferred progression, with *Ngn3*-low cells concentrated early, *Ngn3*-high states occupying intermediate positions, *Fev*$^{+}$ intermediates appearing later, and terminal endocrine populations enriched toward the end of the trajectory (Fig. 4). Together, these observations support the use of pseudotime as a coarse developmental reaction coordinate (i.e., a one-dimensional ordering variable) for the endocrine lineage.

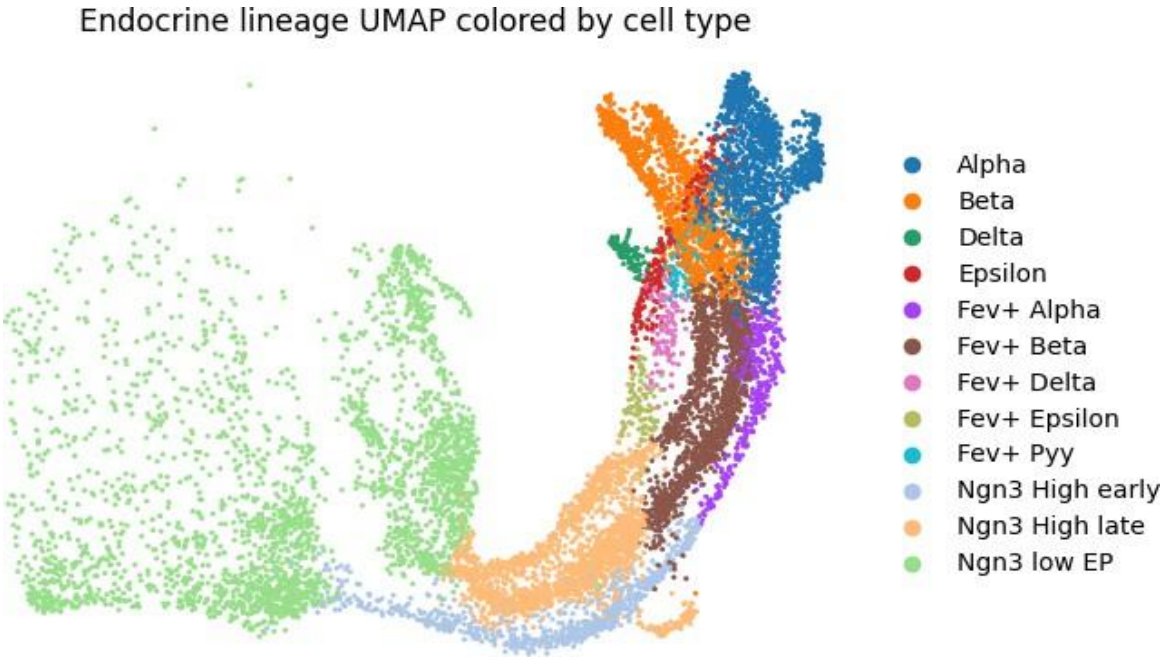


Figure 2: UMAP visualization of endocrine-lineage cells colored by annotated cell type. Cells form a continuous developmental manifold from progenitors to mature endocrine states.

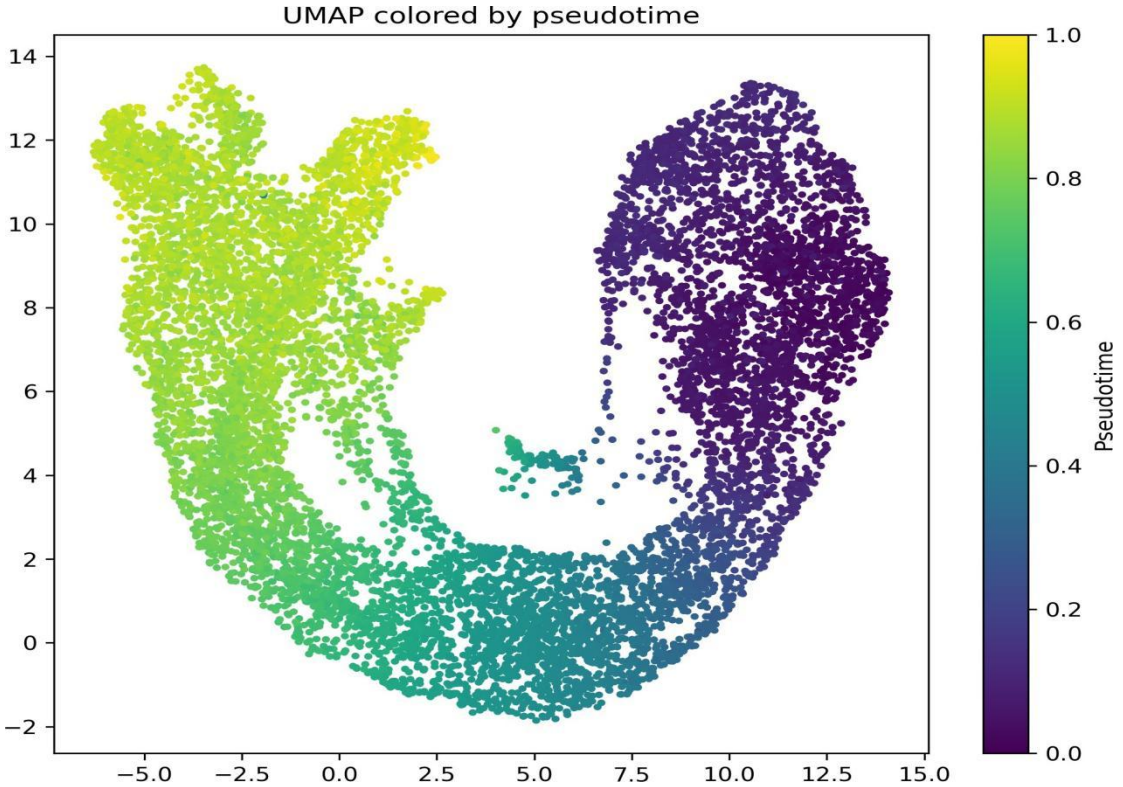


Figure 3: UMAP colored by diffusion pseudotime. Color gradients indicate progressive endocrine differentiation along a continuous trajectory.

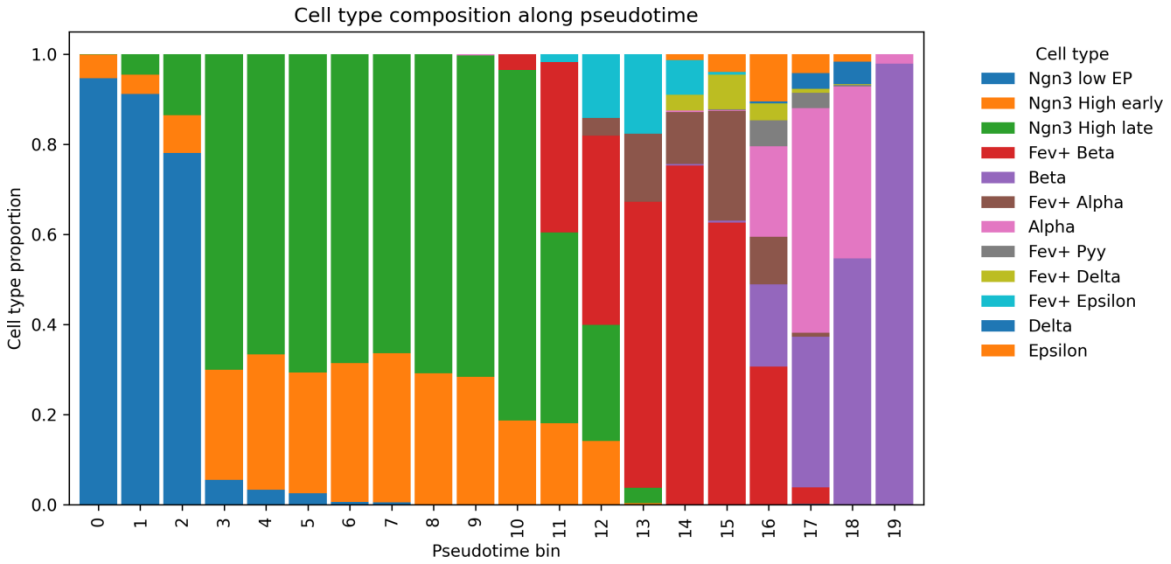


Figure 4: Cell-type composition across pseudotime bins. The ordered replacement of populations supports biologically meaningful lineage progression.

Spectral bottleneck during lineage commitment

Using sliding-window spectral decomposition along pseudotime, we computed the effective mode number R_{eff} and observed a pronounced transient minimum during the transition from *Ngn3*-low progenitors toward *Ngn3*-high commitment states (Fig. 5A). R_{eff} dropped from approximately 220 to 135, representing a reduction in effective dimensionality. Because R_{eff} measures the effective dimensionality of collective transcriptional states, this decrease indicates temporary compression of transcriptional variability into a smaller number of dominant collective modes. The subsequent recovery of R_{eff} at later pseudotime suggests that this compression is not permanent but instead restricted to a discrete interval of lineage progression. Taken together, these results are consistent with a commitment bottleneck in which population-level transcriptional organization becomes transiently constrained.

Third-order asymmetry marks lineage diversification

In contrast to the transient minimum in R_{eff}, the third-order spectral asymmetry statistic R_3 rose substantially after the bottleneck phase and remained elevated through intermediate-to-late pseudotime (Fig. 5B). This pattern suggests increasingly uneven redistribution of population variance among competing collective transcriptional programs. The timing of the increase is temporally aligned with the emergence of *Fev*$^{+}$ intermediates and subsequent terminal endocrine branches, consistent with progressive diversification of lineage-resolved states. In this sense, R_3 captures higher-order organization in the population structure that is not visible from R_{eff} alone. Together, these observations suggest that endocrine differentiation is accompanied by a shift from transient collective compression to asymmetric diversification.

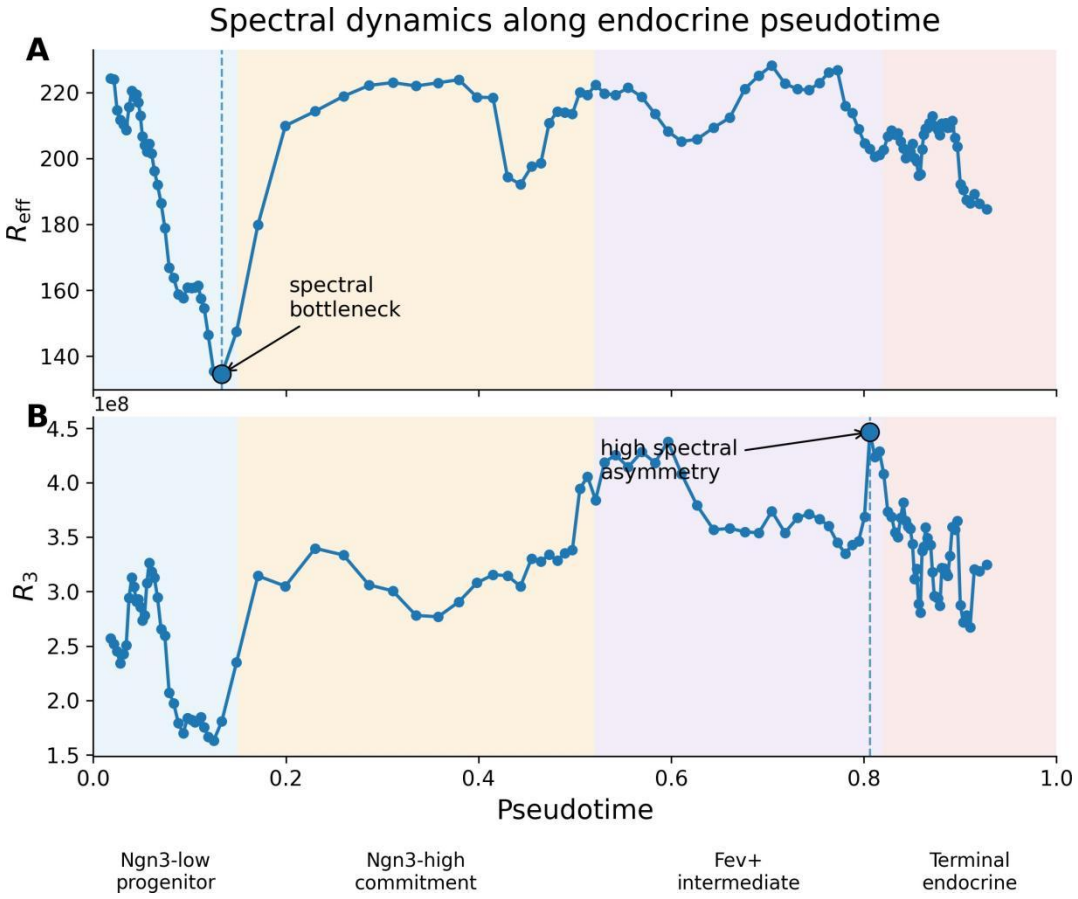


Figure 5. Observed spectral dynamics across endocrine developmental phases. (A) Reff exhibits a transient minimum during the Ngn3-high commitment window, indicating a spectral bottleneck. Colored background regions denote major developmental phases inferred from cell-type composition along pseudotime. (B) R3 increases during the Fev+ intermediate and terminal endocrine phases, consistent with asymmetric lineage diversification.

Although these observed dynamics are temporally aligned with known endocrine developmental stages, statistical validation is required to determine whether they depend on the ordered progression of differentiation rather than on arbitrary sliding-window sampling. We therefore constructed a random-order null model.

Null-model comparison supports trajectory-dependent organization

To test whether the spectral patterns observed in Fig. 5 could arise from random cell ordering or from the marginal expression distribution alone, we used a random-order null model in which the cell expression data were preserved but cell order was randomly permuted before reconstructing sliding windows and recalculating the spectral indicators. As shown in Fig. 6A and B, the gray bands denote the pointwise 95% range of random-order null trajectories, and the dashed gray lines denote the corresponding null medians; these bands are not confidence intervals for the observed trajectories. The observed pseudotime-ordered R_{eff} trajectory deviates strongly from this random-order baseline, including the bottleneck-like decrease identified in Fig. 5, whereas the observed R_3 trajectory shows a later rise beyond the null expectation. These deviations indicate that the spectral bottleneck and third-order spectral

asymmetry depend on the real developmental ordering rather than simply on the expression values present in the dataset. Figure 6 C and D further provide global permutation tests based on the R_{eff} bottleneck-depth statistic (T_{Reff}) and the R_3 asymmetry-rise statistic (T_{R3}), with empirical p values quantifying how often random cell-order permutations produce effects at least as large as those observed.

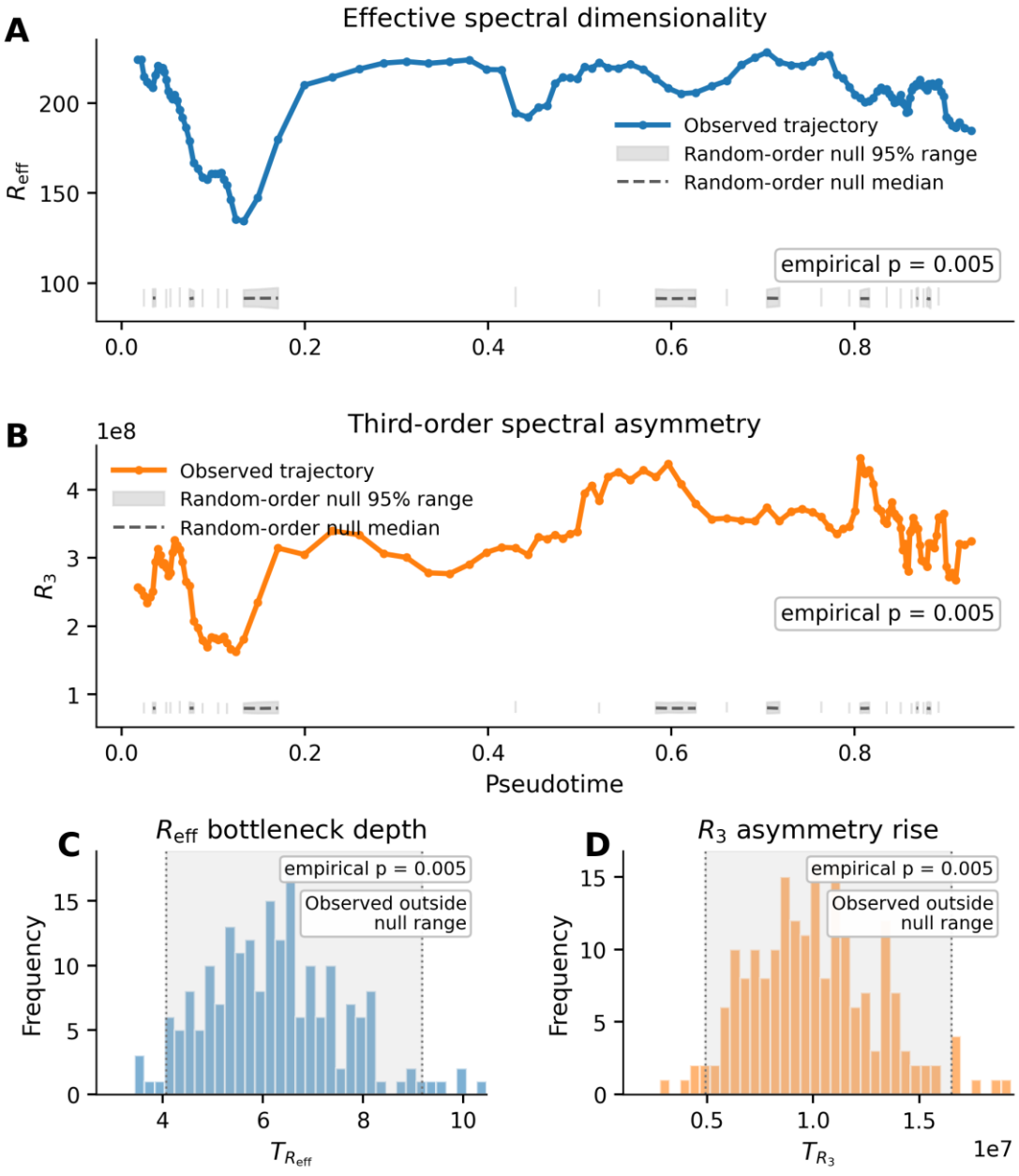


Figure 6. Statistical validation of trajectory-dependent spectral signatures. (A) Observed Reff trajectory along pseudotime compared with the random-order null baseline. The gray band denotes the pointwise 95% range of trajectories obtained by randomly permuting cell order, and the dashed gray line denotes the corresponding null median. The colored curve denotes the observed pseudotime-ordered trajectory. (B) Observed R_3 trajectory compared with the same random-order null baseline. (C) Permutation null distribution of the R_{eff} bottleneck-depth statistic, T_{Reff}. The empirical p value quantifies how often random-order trajectories produce a bottleneck depth at least as large as the observed value. (D) Permutation null distribution of the R_3 asymmetry-rise statistic, T_{R3}. The empirical p value quantifies how often random-order trajectories produce an asymmetry rise at least as large as the observed value. Together, these results show that the observed spectral bottleneck and third-order spectral asymmetry are associated with the ordered progression of endocrine differentiation rather than with random cell ordering or the marginal expression distribution alone.

3. Discussion

Main findings

This study establishes three key systems-level features of pancreatic endocrine differentiation at the population level. First, endocrine commitment is marked by a quantifiable, transient reduction in transcriptional effective dimensionality(R_{eff}), revealing a population-level spectral bottleneck that precedes lineage divergence. Second, the subsequent increase in R_3 indicates an uneven redistribution of transcriptional variance across collective modes, consistent with asymmetric diversification into specialized terminal cell fates. Third, permutation-based null modeling confirms that these spectral features are linked to the ordered terminal progression of differentiation, rather than artifacts of sampling or static expression distributions.

Our identification of a population-level spectral bottleneck expands existing conceptual models of cellular differentiation by providing a global, coordinate-free observable of collective cell state organization. The canonical Waddington epigenetic landscape framework conceptualizes differentiation as cells traversing a topographic surface of branching valleys corresponding to stable cell fates[23]. While this landscape provides an intuitive metaphor for lineage progression, it remains inherently qualitative. Quantitative approaches such as single-cell signaling entropy have offered cell-intrinsic measures of developmental potency, identify a “regulatory bottleneck” via reductions in individual-cell transcription uncertainty[9]. However, entropy-based metrics assess the plasticity of individual cells, rather than the global structural reorganization of the entire cell population.

Our spectral framework addressed this gap by quantifying population-level collective dynamics. The R_{eff} metric, conceptually related to participation ratio analyses used to quantify effective dimensionality in neural system and gene expression correlation

networks[24], captures the global compression of the population's transcriptional state space. By applying these physics-inspired spectral metrics directly to developmental pseudotime trajectories, our work bridges single-cell information theory and macroscopic population topology. Critically, we demonstrate that the commitment "bottleneck" is not solely a property of individual cell states, but a coherent, synchronized contraction of the collective developmental state space across the entire cell population.

Biological interpretation

Biologically, the transient reduction in R_{eff} likely reflects a genome-wide restriction of accessible transcriptional programs as cells synchronously transit through the endocrine commitment checkpoint. The spectral bottleneck aligns precisely with the *Ngn3*-high state, the canonical window of endocrine lineage specification. This is consistent with the established biphasic function of *Neurog3* in pancreatic development: low, sustained *Neurog3* activity maintains proliferating progenitors identity, whereas a sharp, transient upregulation drives cell-cycle exit and activation of the core endocrine differentiation program[25–26]. Accordingly, the R_{eff} minimum provides a population-level molecular readout of the commitment transition, though it does not demonstrate a direct causal for *Neurog3* in driving the spectral compression.

Multiple non-mutually exclusive molecular mechanisms may underlie this population-level spectral bottleneck. First, coordinated chromatin and epigenetic remodeling could globally restrict accessible cis-regulatory elements, narrowing the repertoire of active transcriptional programs and reducing population-level heterogeneity. Second, as cells exit the progenitor state, they may lose responsiveness to a broad range of extracellular signaling inputs, reducing the diversity of signaling-induced transcriptional states. Third, initially heterogeneous progenitor trajectories may coverage onto a shared core endocrine gene regulatory network, aligning individual cell states and reducing collective dimensionality). These hypotheses are consistent with the observed compression of transcriptional degrees of freedom, and distinguishing among them will require targeted perturbation experiments paired with multi-omic profiling.[30]

Following this critical commitment point, the rise in R_3 mirrors the branching and diversification of the endocrine lineage into specialized terminal fates. This spectral asymmetry reflects the uneven stabilization of lineage-specific transcriptional programs as cells differentiate into α-, β-, δ-, and ε subtypes. Mutually inhibitory regulatory modules, such as the *Arx/Pax4* circuit that specifies α- versus β-cell fate, likely contribute to this asymmetric redistribution of transcriptional variance[27], though R_3 does not directly assay this specific molecular pathway.

Instead, R_3 provides a global, unbiased metric of population-level asymmetric diversification, capturing how a transiently unified committed state expands into a spectrum of functionally distinct terminal cell identities.

Figure 7 presents a conceptual model integrating these spectral dynamics with the established biological progression of pancreatic endocrine differentiation. The upper layer depicts the known lineage sequence from *Ngn3*-low progenitors via *Ngn3*-high commitment and Fev^+ intermediates to terminal endocrine fates, whereas the lower layer maps these stages to the corresponding population-level spectral states. In this module, early progenitors from a high-dimensional ensemble with many active collective modes. The Ngn3-high window corresponds to a transient R_{eff} minimum, representing a collective compression of transcriptional degrees of freedom at commitment. bsequent progression through Fev^+ intermediates to terminal fates is marked by elevated R_3, reflecting asymmetric diversification and the emergence of branch-specific transcriptional programs. This framework complements, rather than replaces, marker-based and gene regulatory network analyses, offering a complementary state-space perspective on population reorganization during cell fate transitions.

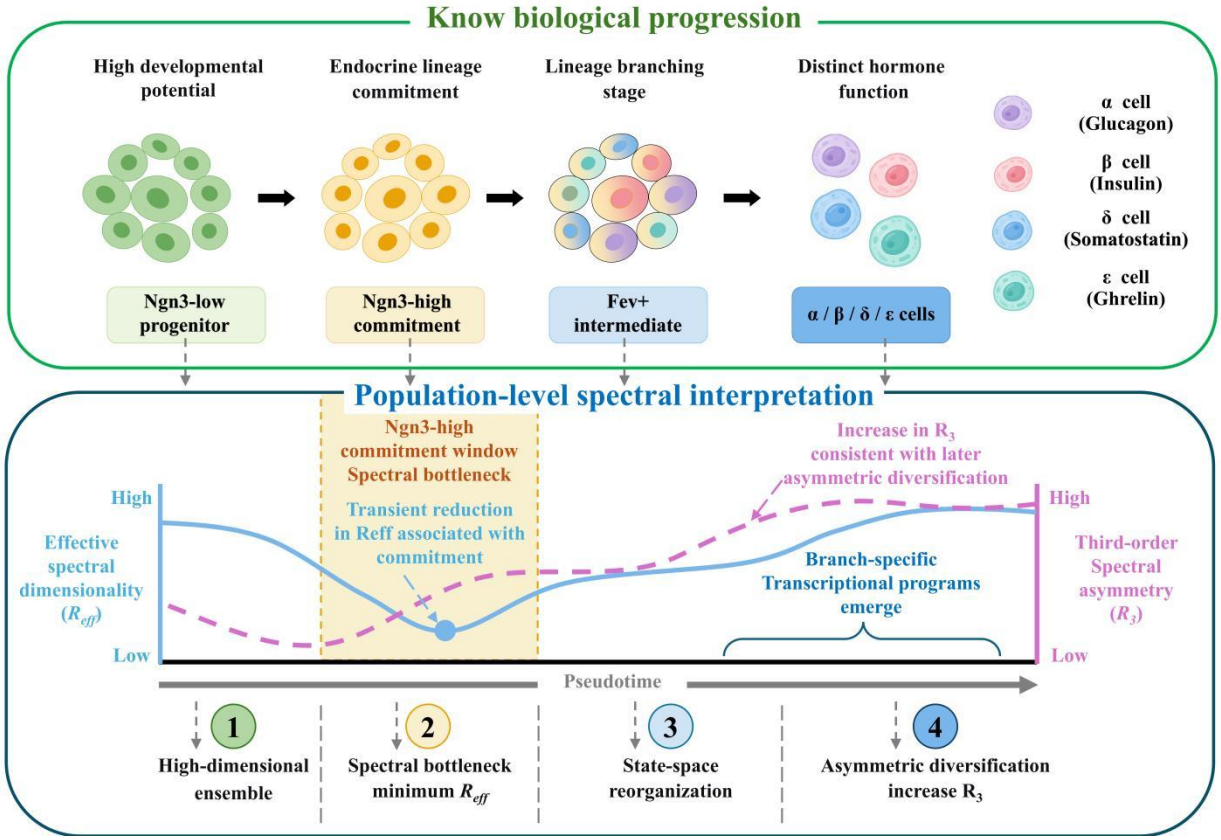


Figure 7. Conceptual model linking pancreatic endocrine lineage progression to population-level spectral reorganization. The upper layer summarizes the known biological progression from *Ngn3*-low progenitors through *Ngn3*-high endocrine commitment and *Fev*$^{+}$ intermediates toward terminal α, β, δ, and ε endocrine fates. The lower layer illustrates the corresponding population-level spectral interpretation proposed in this study. Early progenitor states are represented as a high-dimensional transcriptional ensemble with many active collective modes. During the *Ngn3*-high commitment window, R_{eff} reaches a transient minimum, indicating a spectral bottleneck and compression of collective transcriptional degrees of freedom. Subsequent progression through *Fev*$^{+}$ intermediates and terminal endocrine fates is associated with increased R_3, consistent with asymmetric diversification and the emergence of branch-specific transcriptional programs. This model is intended as a conceptual bridge between known developmental biology and the spectral state-space interpretation, rather than as a direct molecular mechanism.

Our model also generates testable experimental hypotheses. Perturbations that block endocrine commitment, such as *Neurog3* loss or sustained Notch pathway activation, would be expected to attenuate or eliminate the transient R_{eff} minimum by preventing cells from the entering the commitment window. Conversely, perturbations that alter lineage allocation without blocking commitment, including *Pax4* or *Arx* perturbation, would be predicted to modify the magnitude or timing of the late-stage R_3 increase. These experiments would help validate a spectral observable as robust population-level readouts of developmental regulatory perturbation.

Methodological significance

The present framework provides a set of population-level quantitative descriptors that

complement l single-cell analytical approaches, such as marker-gene identification, clustering, and trajectory visualization. Unlike standard methods that focus on individual gene expression or discrete cell states, our spectral metrics capture the global collective organization of the cell population and its evolution across developmental time. This macro-level perspective is particularly valuable for identifying systems-level properties of differentiation that are invisible to cell-centric analyses. [31]

This framework is broadly generalizable to other developmental systems where coordinated state-space reorganization underlines cell fate specification. For example, hematopoiesis has classically been described as a rigid hierarchical tree of discrete lineage choices[28], but recent single-cell studies support a more continuous model of gradual commitment[29]. Applying our spectral framework to hematopoietic differentiation could resolve this debate: R_{eff} dynamics would reveal whether multipotent progenitors transit through a synchronized, low-dimensional priming bottleneck prior to lineage divergence, while R_3 would quantify the asymmetric balance between myeloid and lymphoid lineage outputs. More generally, this approach enables cross-system comparisons of commitment dynamics, providing a common quantitative language to test whether transient collective bottlenecks are a universal principle of multicellular differentiation.

Limitations and outlook

Several limitations should be acknowledged. First, the analysis is based on a single public dataset, and validation in independent pancreatic development datasets, as well as across species, will be critical to establish the generality of our findings. Second, the temporal positioning of the spectral signatures depends on pseudotime inference, and alternative trajectory reconstruction methods may shift the precise location of the R_{eff} minimum and R_3 increase. Third, the spectral metrics are descriptive and correlative, and do not on their own identify the molecular mechanisms driving observed population dynamics. Fourth, the potential sensitivity of results to sliding-window parameters (window size, step size) requires systematic characterization in future work.

Looking forward, it will be important to test whether transient spectral bottlenecks represent a general feature of cell differentiation across systems, including hematopoiesis, neurogenesis, and tissue regeneration. Integrating this spectral framework with lineage tracing, genetic perturbation, and single-cell multi-omic data will help establish causal links between underlying gene regulatory networks and population-level spectral dynamics. Such studies will advance our understanding of the systems-level principles that

govern cell fate specification across developmental biology.

4. Conclusion

In summary, we demonstrate that pancreatic endocrine differentiation proceeds through a quantifiable population-level spectral bottleneck followed by asymmetric diversification, revealing systems-level structure of cell-fate specification that complements marker-gene and clustering analyses. The transient R_{eff} minimum aligns with the *Ngn3*-high endocrine commitment window, while the subsequent rise in R_3 reflects asymmetric branching into terminal endocrine cell fates. Physics-inspired spectral observables thus provide a quantitative framework for characterizing collective differentiation dynamics, enabling cross-system comparisons of cell commitment mechanisms. Future validation across independent datasets and perturbation studies will establish the broader generality of these findings and elucidate how population-level spectral signatures relate to underlying molecular regulatory programs.

5. Materials and Methods

Dataset and endocrine-lineage selection

We analyzed the publicly available single-cell RNA sequencing dataset GSE132188, which profiles mouse pancreatic development across multiple embryonic stages. Processed expression matrices and accompanying cell-type annotations were obtained from the original study. The analysis was based on a population-level single-cell expression matrix rather than temporal trajectories of individual cells. After preprocessing and endocrine-lineage selection, the dataset consisted of thousands of individual cells sampled from multiple developmental stages. To focus on endocrine differentiation, we extracted the endocrine-lineage subset, including *Ngn3*-low progenitors, *Ngn3*-high early and late endocrine precursors, *Fev*$^{+}$ intermediates, and terminal α, β, δ, and ε endocrine cells.

Gene filtering and preprocessing

Each row therefore represents the transcriptional state of a single cell in high-dimensional gene-expression space. Genes expressed in fewer than 10 cells were removed, and highly variable genes were selected. Unless otherwise noted, the analysis was performed on the top 2000 highly variable genes.

The resulting matrix therefore represents a statistical ensemble of single-cell transcriptional states across endocrine differentiation, rather than a longitudinal time series of individual cells. The expression matrix was written as$X_{i,j}$, where i labels cells and j labels genes. SVD was then applied to this collective expression matrix to identify dominant transcriptional modes and their spectral organization.

Let the resulting expression matrix be:

$$X \in \mathbb{R}^{N \times M}, \tag{3}$$

where N is number of cells and M is the number of selected genes. For each gene, expression values were centered across cells and scaled to unit variance,

$$X_{ij} \leftarrow X_{ij} - \frac{1}{N}\sum_{k=1}^{N} X_{kj}, \tag{4}$$

$$X_{ij} \leftarrow \frac{X_{ij}}{\sigma_j}, \tag{5}$$

where σ_j denotes the standard deviation of gene j. This preprocessing reduces dominance by highly expressed genes and emphasizes covariance structure across the cell population.

Singular value decomposition and spectral weights

To characterize population-level transcriptional structure, we performed SVD of the normalized matrix,

$$X = U\Sigma V^{\mathrm{T}}, \tag{6}$$

where U contains cell scores, V contains gene loadings, and Σ = diag (σ_1, σ_2, . . . , σ_r) contains the singular values. We defined the normalized spectral weight of mode α as

$$w_\alpha = \frac{\sigma_\alpha^2}{\sum_{\beta=1}^{r} \sigma_\beta^2}, \qquad \sum_{\alpha=1}^{r} w_\alpha = 1, \tag{7}$$

The set w_α therefore forms a probability-like distribution over collective transcriptional modes.

Definition of R_{eff} and R_3

We first defined the effective number of active modes as

$$R_{\mathrm{eff}} = \frac{1}{\sum_\alpha w_\alpha^2}, \tag{8}$$

This quantity measures the effective dimensionality of collective transcriptional states. Larger values indicate that variance is distributed across many modes, whereas smaller values indicate concentration into fewer dominant modes.

To quantify higher-order asymmetry in the spectral-weight distribution, we computed the mean spectral weight:

$$\bar{w} = \frac{1}{r}\sum_{\alpha=1}^{r} w_\alpha, \tag{9}$$

and the second and third central moments:

$$K_2 = \frac{1}{r}\sum_{\alpha=1}^{r} (w_\alpha - \bar{w})^2, \tag{10}$$

$$K_3 = \frac{1}{r}\sum_{\alpha=1}^{r} (w_\alpha - \bar{w})^3, \tag{11}$$

The third-order spectral asymmetry statistic was then defined as

$$R_3 = \frac{K_3}{K_2^3}, \tag{12}$$

This normalization is not conventional skewness. It suppresses dominant second-order concentration effects and enhances third-order spectral redistribution.

Pseudotime inference

Developmental ordering was inferred using diffusion pseudotime (DPT) as implemented in Scanpy. A nearest-neighbor graph was constructed from principal-component coordinates, followed by diffusion map embedding. *Ngn3*-low progenitors were used as the root population, and each cell was assigned a continuous pseudotime coordinate τ .

Sliding-window spectral dynamics

To quantify changes in spectral organization along development, cells were sorted by pseudotime and analyzed in overlapping sliding windows. Unless otherwise noted, the window size was 400 cells and the step size was 100 cells. This window size balances spectral estimation stability against temporal resolution, corresponding to approximately 10% of the lineage population per window. Within each window, SVD was recomputed and the observables were evaluated from the corresponding spectral weights. This procedure yielded pseudotime- resolved trajectories of effective dimensionality and third-order spectral asymmetry.

$$R_{\mathrm{eff}}(\tau), R_3(\tau). \tag{13}$$

Null-model control

As a control, cell order was randomly permuted while preserving the expression matrix. The same sliding-window analysis was then applied to each randomized ordering. This random-order null model preserves the underlying expression values but removes coherent developmental ordering, allowing assessment of whether observed spectral structure depends on pseudotime organization rather than generic sampling effects. Pointwise 95% null ranges and null medians were computed from the random-order trajectories, and empirical p values were estimated from permutation distributions of the R_{eff} bottleneck-depth statistic and the R_3 asymmetry-rise statistic.

Software and reproducibility

All analyses were performed in Python using Scanpy, NumPy, SciPy, and Matplotlib. Singular value decompositions were computed using standard numerical linear algebra routines. Unless otherwise noted, the preprocessing and analysis parameters reported above were used throughout.

Data availability

The single-cell RNA-sequencing data analyzed in this study are publicly available from the Gene Expression Omnibus (GEO) under accession number GSE132188. The dataset can be accessed and downloaded from the NCBI GEO database at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132188. No new experimental data were generated in this study.

Author Contribution

H.L. and N.M. are co-first authors and contributed equally to this article.

Conceptualization: W.L., Z-P.L., L.H.; Methodology: N.M., Y.Y.S.; Data Curation: H.L., N.M.; Writing-Original Draft: H.L., W.L., Z-P.L.; Writing-Review & Editing: N.M., Y.Y.S., W.L.Y., Z-P.L.; Resources: L.H., W.L.; Visualization: H.L., W.L. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the Advanced Materials-National Science and Technology Major Project (No. 2026ZD0623600) and the National Natural Science Foundation of China (No. 12575033). Additional support was provided by grants from the National Natural Science Foundation of China (82303427 and 82372693), the Shaanxi Provincial Outstanding Youth Fund (2024JC-JCQN-77), the Shaanxi Provincial Health Research and Innovation Capability

Improvement Program (2026PT-07), and the Program of the National Clinical Research Center for Digestive Diseases (2026NCRCDDZZ07 and 2026NCRCDDZZ13).

References

1. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386 (2014).

2. Zhang, Z., Wang, Z., Sun, Y., Shen, J., Peng, Q., Li, T. & Zhou, P. Deciphering cell-fate trajectories using spatiotemporal single-cell transcriptomic data. npj Syst. Biol. Appl. 12, 5 (2025).

3. Gradwohl, G., Dierich, A., LeMeur, M. & Guillemot, F. Neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc. Natl Acad. Sci. USA 97, 1607–1611 (2000).

4. Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development 129, 2447–2457 (2002).

5. Apelqvist, A. et al. Notch signalling controls pancreatic cell differentiation. Nature 400, 877–881 (1999).

6. Byrnes, L. E. et al. Lineage dynamics of murine pancreatic development at single-cell resolution. Nat. Commun. 9, 3922 (2018).

7. Krentz, N. A. J. et al. Single-cell transcriptome profiling of mouse and hESC-derived pancreatic progenitors. Stem Cell Rep. 11, 1551–1564 (2018).

8. Wang, J. et al. Regulation of endocrine cell alternative splicing revealed by single-cell RNA sequencing in type 2 diabetes pathogenesis. Commun. Biol. 7, 778; 10.1038/s42003-024-06475-0 (2024).

9. Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. Nat. Commun. 8, 15599 (2017).

10. Paczkó, M. et al. A neural network-based model framework for cell-fate decisions and development. Commun. Biol. 7, 323; 10.1038/s42003-024-05985-1 (2024).

11. Wolff, R. et al. Monoallelic gene expression in developing cells increases genetic noise and Shannon entropy. Commun. Biol. 8, 857; 10.1038/s42003-025-08128-2 (2025).

12. Sun, Y. et al. Eigen microstates and their evolutions in complex systems. Commun. Theor. Phys. 73, 065603 (2021).

13. Liu, T. et al. Renormalization group theory of eigen microstates. Chin. Phys. Lett. 39, 080503 (2022).

14. Li, X. et al. Discontinuous and continuous transitions of collective behaviors in living systems. Chin. Phys. B 30, 128703 (2021).

15. Pearson, K. On lines and planes of closest fit to systems of points in space. Philos. Mag. 2, 559–572 (1901).

16. Hotelling, H. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417–441 (1933).

17. Alter, O., Brown, P. O. & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. Proc. Natl Acad. Sci. USA 97, 10101–10106 (2000).

18. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A 374, 20150202 (2016).

19. Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. Nat. Phys. 13, 431–434 (2017).

20. Wetzel, S. J. Unsupervised learning of phase transitions: from principal component analysis to variational autoencoders. Phys. Rev. E 96, 022140 (2017).

21. Muzzi, C. et al. Principal component analysis of absorbing state phase transitions. Phys. Rev. E 110, 064121 (2024).

22. Liu, W. et al. Spectral signatures of third-order pseudo-transitions in finite systems: an eigen-microstate approach. arXiv 10.48550/arXiv.2604.19244 (2026).

23. Moris, N., Pina, C. & Arias, A. M. Transition states and cell fate decisions in epigenetic landscapes. Nat. Rev. Genet. 17, 693–703 (2016).

24. Recanatesi, S., Bradde, S., Balasubramanian, V., Steinmetz, N. A. & Shea-Brown, E. A scale-dependent measure of system dimensionality. Patterns 3, 100555 (2022).

25. Yang, L., Yu, X.-X., Wang, X., Jin, C.-T. & Xu, C.-R. The expression order determines the pioneer functions of NGN3 and NEUROD1 in pancreatic endocrine differentiation. Sci. Adv. 11, eadt4770 (2025).

26. Peng, L. et al. Deciphering ligand-receptor-mediated intercellular communication based on ensemble deep learning and the joint scoring strategy from single-cell transcriptomic data. Comput. Biol. Med. 171, 108123 (2024).

27. Collombat, P. et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes Dev. 17, 2591–2603 (2003).

28. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat. Cell Biol. 19, 271–281 (2017).

29. Mincarelli, L. et al. Single-cell gene and isoform expression analysis reveals signatures of ageing in haematopoietic stem and progenitor cells. Commun. Biol. 6, 558; 10.1038/s42003-023-04936-6 (2023).

30. Hu, Q., Lu, X., Xue, Z. & Wang, R. Gene regulatory network inference during cell fate decisions by perturbation strategies. npj Syst. Biol. Appl. 11, 8 (2025).

31. Herrera, J., Bensussen, A., García-Gómez, M. L., Garay-Arroyo, A. & Álvarez-Buylla, E. R. A system-level model reveals that transcriptional stochasticity is

required for hematopoietic stem cell differentiation. npj Syst. Biol. Appl. 10, 15 (2024).