SPECIAL TOPIC — A celebration of the 90th Anniversary of the Birth of Bolin Hao

Analysis of spatiotemporal dynamic patterns of gene expression during mouse embryonic development based on Moran's I and spatial transcriptomics

Qi-Chao Li(李啟超)[1,2,†], Hai Lin(林海)[2,†], Peng Wang(王鹏)[1,2], Qiutong Dong(董秋彤)[1,2], Kun Wang(王坤)[1,2], Jian-Wei Shuai(帅建伟)[2,‡], and Fang-Fu Ye(叶方富)[1,2,,§]

[1] *Postgraduate Training Base Alliance, Wenzhou Medical University, Wenzhou* 325035, *China*

[2] *Oujiang Laboratory (Zhejiang Laboratory for Regenerative Medicine, Vision and Brain Health), Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou* 325001, *China*

(Received 4 May 2025; revised manuscript received 4 June 2025; accepted manuscript online 13 June 2025)

Spatial transcriptomics technology provides novel insights into the spatial organization of gene expression during embryonic development. In this study, we propose a method that integrates analysis across both temporal and spatial dimensions to investigate spatial transcriptomics data from mouse embryos at different developmental stages. We quantified the spatial expression pattern of each gene at various stages by calculating its Moran's I. Furthermore, by employing time-series clustering to identify dynamic co-expression modules, we identified several developmentally stage-specific regulatory gene modules. A key finding was the presence of distinct, stage-specific gene network modules across different developmental periods: Early modules focused on morphogenesis, mid-stage on organ development, and late-stage on neural and tissue maturation. Functional enrichment analysis further confirmed the core biological functions of each module. The dynamic, spatially-resolved gene expression model constructed in this study not only provides new biological insights into the programmed spatiotemporal reorganization of gene regulatory networks during embryonic development but also presents an effective approach for analyzing complex spatiotemporal omics data. This work provides a new perspective for understanding developmental biology, regenerative medicine, and related fields.

Keywords: Moran's I, spatial transcriptomics, embryonic development, spatiotemporal dynamics, gene regulatory network

PACS: 87.10.Vg, 87.85.mg, 05.45.Tp

DOI: 10.1088/1674-1056/ade427 **CSTR:** 32038.14.CPB.ade427

1. Introduction

A central question in developmental biology is understanding how gene regulatory networks are coordinated in time and space to drive the development of complex organisms from a single cell.[1] While traditional transcriptomics has revealed the expression patterns of many key regulatory factors, the loss of spatial information hinders a comprehensive analysis of the intercellular context within tissues. The recent emergence of spatial transcriptomics technology fills this gap, enabling us to simultaneously capture gene expression and its spatial location on tissue sections.[2,3] In embryonic development research, understanding how cell fate decisions and organ formation occur with precise timing and spatial organization within tissues is crucial.[4,5] Spatial transcriptomics provides a key tool for dissecting these dynamic processes. Meanwhile, advances in single-cell sequencing and high-throughput technologies allow for the collection of transcriptomic data from embryonic tissues at multiple developmental time points, facilitating systematic studies across both spatial and temporal dimensions.[6–9]

To extract biologically meaningful signals from complex spatiotemporal transcriptomics data, this study introduces the spatial statistic Moran's I. Moran's I was initially used in geographic information analysis to measure the spatial autocorrelation of an attribute.[10] Spatial autocorrelation based on Moran's I is widely applied across various research fields: for example, Augustine *et al.* used Moran's I to analyze the relationship between grazing intensity and bare soil spatial heterogeneity, revealing the role of spatial autocorrelation in assessing ecosystem responses;[11] Griffith and Chun used Moran's I for remote sensing information extraction;[12] Youssef *et al.* employed Moran's I to analyze the spatial autocorrelation of landslide occurrences and combined it with spatial logistic regression and GeoDetector methods to improve the accuracy of landslide susceptibility mapping.[13] Similarly, Moran's I has been successfully applied to spatial transcriptomics.[14] For instance, Qiu *et al.* utilized Moran's I to identify differentially expressed genes in spatial transcriptomics data,[15] and a study

[†]These authors contributed equally to this work.

[‡]Corresponding author. E-mail: jianweishuai@xmu.edu.cn

[§]Corresponding author. E-mail: fye@iphy.ac.cn

by Lin *et al*. used Moran's I to measure spatial autocorrelation for identifying tissue architecture.[16] Moran's I measures the correlation of signal intensity between adjacent locations: when the expression levels of a gene at neighboring spatial points are significantly correlated, its Moran's I value is high, indicating that the gene exhibits a clustered spatial expression pattern.[17]

In the present study, we utilize Moran's I to quantitatively assess the spatial expression characteristics of each gene at different embryonic stages, thereby identifying genes associated with spatial pattern development. The analysis pipeline is illustrated in Fig. 1. First, for each stage, a spatial neighborhood graph was constructed based on K-nearest neighbors (KNN).[18] Using this graph, the Moran's I was calculated for every gene within defined tissue regions at each stage. This process generated time-series Moran's I profiles for each gene across the developmental stages within specific tissues. These profiles underwent linear interpolation and smoothing to enhance signal quality, followed by removing genes with excessive missing values. Subsequently, we employed the Tslearn KMeans time-series clustering[19] algorithm to group genes into modules based on similar spatiotemporal dynamic patterns reflected in their Moran's I profiles. Within each identified module, genes were ranked in descending order based on a representative Moran's I value, and the top 50 genes were selected. Finally, functional enrichment and network analyses were performed on these top genes to elucidate their biological significance, particularly concerning their roles in morphogenesis, organogenesis, and tissue maturation.[20] By comparing changes in Moran's I across multiple time points, we can capture the evolution of gene spatial expression patterns over time, laying the groundwork for constructing spatiotemporal dynamic patterns of gene expression.

This study aims to utilize spatiotemporal transcriptomics data, combined with Moran's I analysis, to construct spatiotemporal dynamic patterns during mouse embryonic development and perform functional interpretation. We hypothesized that distinct gene modules govern embryonic development at different periods, which form specific spatial expression distributions and are activated programmatically over time.

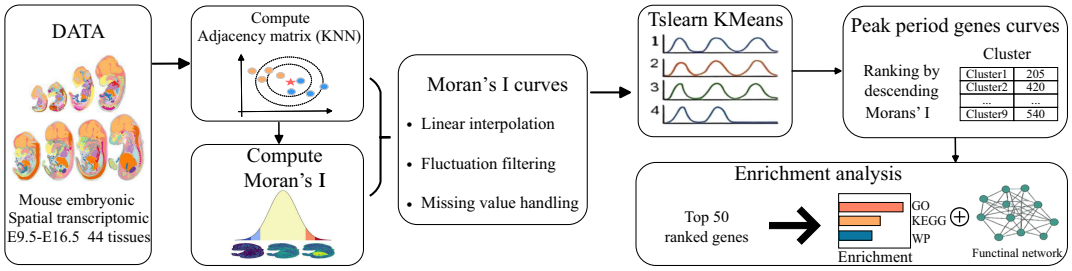

Fig. 1. Workflow for the dynamic identification of gene modules in mouse embryonic development using Moran's I. Gene spatial aggregation dynamics across developmental stages (input: E9.5–E16.5 spatial transcriptomics) were quantified using Moran's I. After processing the temporal Moran's I profiles, time-series clustering (KMeans via Tslearn) grouped genes into modules based on shared dynamic patterns. Top genes from these modules, ranked by descending Moran's I, were subsequently analyzed for functional significance in development through enrichment (GO, KEGG, WP) and network analyses.

2. Materials and methods

2.1. Data source

The data used in this study were sourced from the mouse embryonic development spatiotemporal transcriptomics dataset provided by Zhong *et al*.[21] This dataset contains gene expression information of mouse embryos from developmental stage E9.5 to E16.5 (see Fig. 1), covering spatial expression data for 44 different tissues and the transcript levels of 2000 genes. These 2000 genes were pre-selected by the original data providers (Zhong *et al*.) by first identifying highly variable genes and then further selecting those exhibiting spatial expression patterns based on Moran's I, ensuring their relevance for spatiotemporal dynamic analysis. These data capture, in high resolution, the spatial distribution and temporal evolution of gene expression during embryonic development, providing a rich resource for this research. Using this dataset, we conducted an in-depth analysis of the spatial patterns of gene expression during embryonic development and their dynamic changes.

2.2. Adjacency matrix calculation

In our analysis of mouse embryonic spatiotemporal transcriptomics data, we constructed an adjacency matrix based on the spatial positions of cell spots to quantify the spatial relationships between them. For this purpose, we utilized the spatial_neighbors function from the widely used spatial omics analysis library Squidpy to construct the adjacency matrix. This function computes the adjacency matrix based on spatial coordinates using the KNN algorithm. This method establishes connections by identifying the K nearest neighbors for each spot, ensuring accurate capture of relationships between adjacent spots.[18] The KNN algorithm is a distance-based non-parametric method widely used in classification, regression, and clustering tasks. For each cell spot in the dataset, its Euclidean distance to all other spots is calculated. The algorithm quantifies the spatial distance between any two cell

spots. For each cell spot i, the 6 cell spots with the smallest Euclidean distance to it were selected as its neighbors (i.e., $K = 6$), and the adjacency matrix $\boldsymbol{A}$ was generated. The adjacency matrix $\boldsymbol{A}$ is an $n \times n$ matrix (where n is the number of spots), whose element A_{ij} represents the connection between spot i and spot j. A simple binary method is used to define A_{ij}; the adjacency matrix only records whether a neighbor relationship exists between spots (a value of 1 indicates existence, 0 indicates absence), without considering specific distance weights.

2.3. Moran's I calculation

For each gene, we calculated its Moran's I at different developmental time points (E9.5–E16.5) within each of the 44 annotated tissues, using the expression data from the constituent cell spots. This calculation quantified the spatial autocorrelation of gene expression. The formula for Moran's I is[10]

$$I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\,(x_i - \bar{x})\,(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}, \qquad (1)$$

where, for a specific tissue at a specific time point, N is the number of cell spots, x_i is the expression level of the gene in spot i, $\bar{x}$ is the average expression level of the gene across these N spots, w_{ij} is the spatial weight between spot i and spot j. In this study, w_{ij} is derived from the binary adjacency matrix A_{ij} (based on KNN with $K = 6$, see Subsection 2.2), where $w_{ij} = 1$ if spots i and j are neighbors, and $w_{ij} = 0$ otherwise.

The resulting Moran's I time-series profiles reflect the temporal evolution of the gene's spatial expression pattern within each tissue.

2.4. Processing of Moran's I time series

To ensure the reliability of subsequent analyses and improve data interpretability, we further processed the calculated Moran's I time-series profiles. First, focusing on the Moran's I time series calculated for each gene within each tissue across the E9.5 to E16.5 development period, we applied several processing steps. To handle discontinuities observed in the time series for some tissues (potentially arising from detection sensitivity fluctuations or sample processing variations), we used linear interpolation to fill these gaps (e.g., a "detected–undetected–detected" pattern) to restore temporal continuity, which is crucial for dynamic analysis (illustrated by red bands in Fig. 2(a)). Second, we assessed the data completeness for each tissue across the entire developmental timeline (E9.5–E16.5) and removed tissues where the proportion of missing data exceeded 50%, aiming to reduce interference from potential technical noise in the analysis results (indicated by gray bands in Fig. 2(a)). After completing the tissue-level filtering and interpolation, the temporal data coverage status for each remaining tissue is summarized in Fig. 2(a). While more complex interpolation methods exist, any interpolation approach has its limitations. We adopted linear interpolation based on its simplicity and robustness in filling a few missing points, avoiding excessive assumptions about the local curve shape. Concurrently, our clustering method assesses the overall similarity of two time series across all time points.[22] On a curve comprising multiple time points, the contribution of a single interpolated point to the overall Euclidean distance is limited. Therefore, as long as the main dynamic features of the curve are preserved, this local processing is unlikely to systematically alter the final cluster assignment of a gene.

Finally, based on the volatility of these time series, we selected and retained genes with significant changes in Moran's I while removing those with non-obvious spatiotemporal dynamic changes, specifically by calculating the difference between the maximum and minimum values of the time series. To determine an appropriate fluctuation range threshold (Δ), we first evaluated the impact of different Δ values on downstream clustering analysis. We compared clustering results from gene sets filtered at various Δ thresholds (e.g., 0.05, 0.1, and 0.15) and found that while the number of genes included for analysis varied, the major dynamic clustering patterns were consistently identified across these different thresholds (supplementary Fig. S5). This indicated that, at the level of the primary dynamic patterns of interest, the clustering outcomes exhibited good robustness to variations in Δ within a reasonable range (e.g., 0.05 to 0.15).

Having established a degree of robustness in downstream analysis to the choice of Δ, and to further identify a relatively optimal Δ value based on the data's inherent characteristics, we examined the change in the number of remaining genes after filtering across a broader range of Δ thresholds (from 0 to 0.2, with a step of 0.005). By performing a knee point analysis on the "Number of Remaining Genes vs. Δ Threshold" curve (using the kneed library,[23] with parameters $S = 1.0$, curve='convex', direction='decreasing'), the results showed that the "knee point" of this curve is located near $\Delta \approx 0.1$ (supplementary Fig. S6). In this region, the rate of decrease in the number of filtered genes begins to slow down significantly with increasing threshold, suggesting that $\Delta \approx 0.1$ is a reasonable balance point between effectively filtering out low-dynamic signals and retaining sufficient high-dynamic signals. We therefore filtered the genes, retaining only those with a Moran's I fluctuation range of 0.1 or greater. This "fluctuation filtering" step effectively enriched for genes with strong dynamic spatial signals.

We employed Δ to filter for genes with high volatility, aiming to further pinpoint, among genes already exhibiting variability, those whose spatial aggregation patterns themselves undergo significant dynamic remodeling. Such genes

are more indicative of the key regulatory changes driving transitions between embryonic developmental stages. Concurrently, filtering for highly volatile genes helps reduce static background interference, thereby highlighting core dynamic regulatory modules and rendering time-series clustering analysis, which relies on distinguishable dynamic features, more effective. This "fluctuation filtering" step effectively enriched for genes with strong dynamic spatial signals, as shown in Fig. 2(b): compared to the original data (blue distribution), the range of the Moran's I distribution for the filtered genes (orange distribution) is significantly widened, enhancing the distinction between high and low spatial autocorrelation values, thereby improving the ability to capture key developmental regulatory genes. Notably, due to tissue filtering and the inherent spatiotemporal heterogeneity of gene expression, the number of valid tissues associated with each gene after preprocessing varies. Compared to the theoretical association of each gene with 44 tissues in the original dataset, filtered genes typically retained high-quality spatial autocorrelation information in only 1 to 22 tissues, with the specific distribution shown in Fig. 2(c). After the aforementioned curve processing steps, the dataset's dimensionality was reduced, while data quality and signal-to-noise ratio were significantly improved, facilitating downstream time-series clustering and functional network analysis.

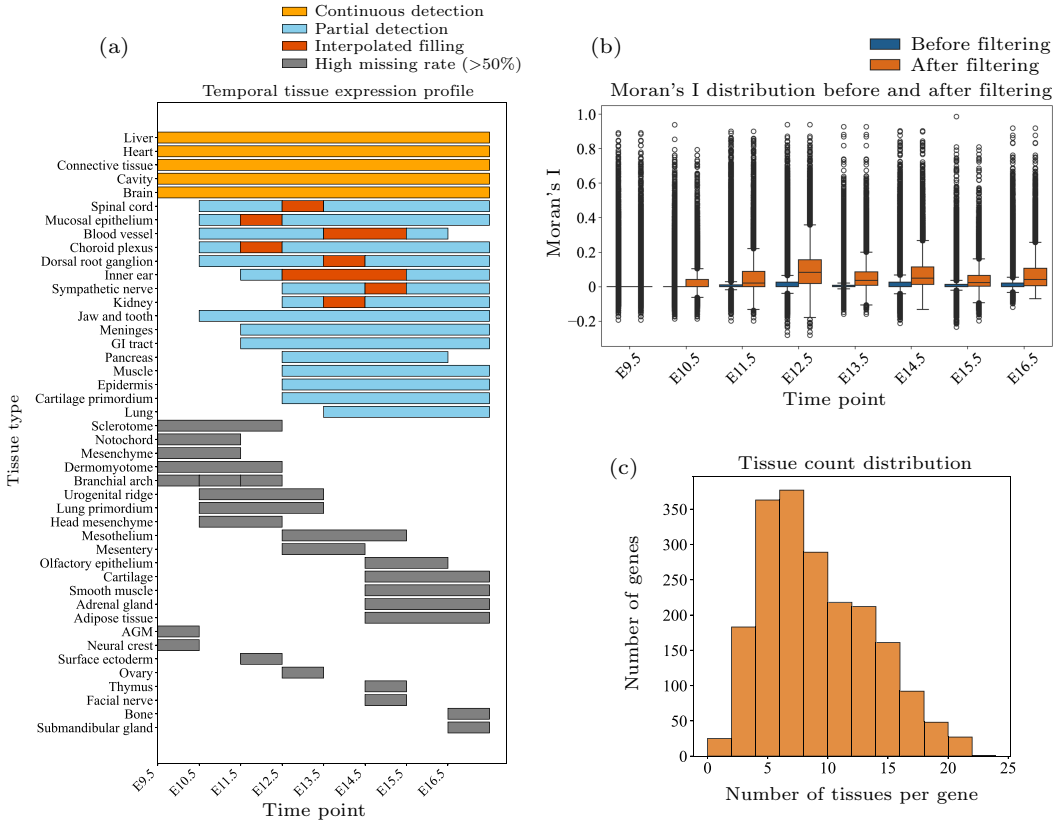

Fig. 2. Processing Moran's I curves enhances the interpretability of spatial autocorrelation signals. (a) Tissue data coverage map (E9.5–E16.5), showing continuous presence (yellow), partial detection (blue), interpolation (red), and removal (>50% missing, gray). (b) Moran's I distributions before (blue) and after (orange) filtering; filtering enriches dynamic signals, widening the distribution range. (c) Post-processing histogram showing the number of tissues covered per gene.

2.5. Clustering

In studying the spatial dynamic patterns of gene expression during mouse embryonic development (stages E9.5 to E16.5), we performed clustering analysis on all qualified Moran's I time series obtained after the curve processing steps. Moran's I is an indicator measuring spatial autocorrelation, used to characterize the changes in the spatial heterogeneity of gene expression during development.[24] To group these genes according to similar spatial expression trends, we employed the Tslearn KMeans algorithm, where each resulting cluster represents a group of genes exhibiting similar spatial dynamic patterns during embryonic development. This analysis revealed the spatiotemporal characteristics of gene expression and laid the foundation for constructing dynamic gene network models.

2.5.1. Tslearn KMeans algorithm principles

Tslearn KMeans is a clustering algorithm specifically designed for time-series data. Compared to traditional K-means, it is particularly suitable for handling data with temporal dependencies and dynamic characteristics.[25] Traditional K-means typically assumes data points are independent static vectors, whereas time-series data (such as the Moran's I of

a gene changing across developmental stages) possess continuity and temporal correlation, thus requiring specialized distance metrics and optimization strategies.

In time-series clustering, the distance metric is a core component of the algorithm. Tslearn KMeans supports various distance calculation methods, such as Euclidean distance and dynamic time warping (DTW).[26] In this study, since the developmental time points are fixed and the focus is on overall trends rather than local time warping, we chose Euclidean distance as the metric. This is because our time-series data are aligned (i.e., the Moran's I for each gene from E9.5 to E16.5 is sampled at fixed time points), eliminating the need to consider non-linear stretching of the time axis. For two time series $\boldsymbol{X} = (x_1, x_2, \ldots, x_T)$ and $\boldsymbol{Y} = (y_1, y_2, \ldots, y_T)$, the Euclidean distance is defined.

$$d(\boldsymbol{X}, \boldsymbol{Y}) = \sqrt{\sum_{t=1}^{T} (x_t - y_t)^2}. \tag{2}$$

Here, T represents the length of the time series, corresponding in this study to the number of consecutive observation points during the E9.5 to E16.5 developmental stages; and x_t and y_t represent the Moran's I values for the two variables at the t-th time point, respectively. This distance metric effectively captures the overall dissimilarity between two time series during their dynamic evolution. Euclidean distance is simple and efficient, suitable for time series that are aligned and of equal length, and capable of reflecting differences in the overall expression trends of genes during development.

The objective of Tslearn KMeans is consistent with traditional K-means: to optimize the clustering result by minimizing the total sum of squared distances from the time series within a cluster to the cluster's center. Mathematically, this objective function can be represented as

$$\min \sum_{i=1}^{k} \sum_{x \in C_i} d(\boldsymbol{X}, \boldsymbol{\mu}_i)^2. \tag{3}$$

where k is the number of clusters. C_i represents the set of time series in the i-th cluster. $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{iT})$ is the center of the i-th cluster (i.e., the mean vector of all time series within the cluster). $d(\boldsymbol{X}, \boldsymbol{\mu}_i)$ is the Euclidean distance between time series $\boldsymbol{X}$ and the cluster center $\boldsymbol{\mu}_i$. By iteratively optimizing this objective function, the algorithm groups genes into clusters with similar spatial expression patterns.

2.5.2. Clustering process

We first constructed a time-series vector for each gene, representing its Moran's I changes during the E9.5 to E16.5 period. For a given gene, its time series vector consists of its Moran's I values across the consecutive time points from E9.5 to E16.5

$$\boldsymbol{X} = (x_{\mathrm{E9.5}}, x_{\mathrm{E10.5}}, x_{\mathrm{E11.5}}, \ldots, x_{\mathrm{E16.5}}). \tag{4}$$

To eliminate differences in scale and improve clustering performance, we standardized the Moran's I time series for all genes. The standardization formula is

$$x'_t = \frac{x_t - \mu_x}{\sigma_x}, \tag{5}$$

where x_t is the original Moran's I value, u_t and σ_x are the mean and standard deviation of the gene's time series, respectively, and x'_t is the standardized value. After standardization, each gene's time series is transformed into a vector with a mean of 0 and a variance of 1, facilitating subsequent distance calculations and clustering. Tslearn KMeans begins its iterations by randomly initializing k cluster centers. The initial cluster centers can be chosen by randomly selecting k time series from the 2000 genes, or by using the KMeans++[27] algorithm for optimized initial selection (the latter improves convergence speed by selecting initial points that are distant from each other). In this study, we used the default random initialization method.

2.5.3. Iterative optimization

The clustering process follows the iterative steps of the classic K-means algorithm:

For each gene's time series $\boldsymbol{X}$, calculate its Euclidean distance to all cluster centers $\boldsymbol{\mu}_i$, and assign it to the cluster with the nearest center:

$$C_i = \left\{ X | d(\boldsymbol{X}, \boldsymbol{\mu}_i) \leq (\boldsymbol{X}, \boldsymbol{\mu}_j),\ \forall j \neq i \right\}. \tag{6}$$

For each cluster C_i, recalculate the cluster center $\boldsymbol{\mu}_i$, i.e., by taking the mean of all time series within the cluster

$$\boldsymbol{\mu}_i = \frac{1}{|C_i|} \sum_{\boldsymbol{X} \in C_i} \boldsymbol{X}, \tag{7}$$

where $|C_i|$ is the number of genes in cluster C_i. Repeat the assignment and update steps until the cluster centers no longer change (or the change is less than a certain threshold), or the maximum number of iterations is reached.

2.6. Functional enrichment and functional network

2.6.1. Functional enrichment

Enrichment analysis, an important step in interpreting genomics data, is used to identify biological functions or pathways that are significantly over-represented in a given list of input genes.[28] Based on the time-series clustering analysis of Moran's I, we identified gene expression modules at different stages of mouse embryonic development and performed functional enrichment analysis on these gene clusters using the Metascape tool.[29] Representative gene lists were uploaded to Metascape's enrichment analysis module. Metascape supports multiple gene identifiers (e.g., Symbol, Ensembl, UniProt), facilitating cross-database comparisons. The analysis covers databases such as gene ontology (GO) biological processes and KEGG pathways. GO provides hierarchical functional annotations,[30] including molecular function, cellular component, and biological process; while KEGG focuses

on metabolic and signaling pathways.[31] Metascape uses the hypergeometric test to assess the enrichment significance for each term. The hypergeometric test[32] is based on the following formula:

$$P(X \geq k) = \sum_{x=k}^{\min(n,K)} \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}. \tag{8}$$

where N represents the total number of background genes, K represents the number of genes in the background annotated with the term, n is the number of genes in the input list, and k is the number of genes in the input list associated with the term. This test calculates the probability of observing at least k genes associated with a specific term, assuming a random distribution under the null hypothesis (no enrichment).

Since multiple terms are tested simultaneously, it is necessary to control the false positive rate. Metascape uses the Benjamini–Hochberg method to adjust p-values,[33] generating q-values (FDR-corrected p-values). This study sets $q < 0.01$ as the significance threshold to ensure the reliability of the results. To reduce redundancy, Metascape performs enrichment clustering, grouping terms with high gene overlap. Metascape uses the Kappa similarity coefficient (Cohen's Kappa) to quantify the association between functional terms:[34]

$$k = \frac{p_{\mathrm{o}} - p_{\mathrm{e}}}{1 - p_{\mathrm{e}}}, \tag{9}$$

where p_{o} is the proportion of gene overlap actually observed between terms, and p_{e} is the proportion of overlap expected by chance. Based on the resulting Kappa matrix (with a similarity threshold >0.3), similar terms are merged into non-redundant functional clusters via hierarchical clustering, and the most significant term (lowest p-value) within each cluster is chosen as its representative term.

2.6.2. Functional network

In this study, we utilized the Metascape platform to construct functional network diagrams, enabling a systematic analysis of the active biological processes and pathways during key stages of mouse embryonic development. The enriched term network graph relies on identifying significantly enriched functional terms within each gene cluster, obtained via the hypergeometric test with Benjamini–Hochberg correction (FDR <0.01). To enhance interpretability, typically, the top few most significant functional terms (e.g., the top 10 terms from the enrichment results of each input gene cluster) are retained for network construction.

The Kappa similarity coefficient is calculated between any two enriched terms to measure the degree of shared genes between them. If the Kappa coefficient $k > 0.3$, an edge is established between the two corresponding enriched terms in the network. Through hierarchical clustering[35] applied to the Kappa distance matrix between terms, modular clustering is performed to further identify functional clusters, where each cluster represents a group of biologically similar or related functional processes.

These diagrams visualize the interactions between significantly enriched functional terms using: nodes (representing biological processes or pathways, such as GO terms or KEGG pathways), node color (indicating functional clusters), node size (representing statistical significance or degree of gene enrichment), and edges (representing shared genes or functional relatedness based on Kappa similarity). Through this network analysis, we elucidated the dynamic interactions and stage-specific characteristics of gene regulatory networks during embryonic development, providing new insights into the molecular mechanisms of morphogenesis and differentiation.

3. Results and analysis

3.1. Analysis of clustering results

To analyze the dynamic changes in spatial expression patterns of genes during embryonic development, we performed clustering analysis on the preprocessed genes based on their Moran's I time series. Selecting an appropriate number of clusters (K) is crucial for balancing pattern resolution with biological interpretability. Our primary goal in clustering was to distinguish gene modules peaking at different embryonic periods. As we needed to uncover the dynamic changes of genes in each time period, we initially discarded cluster numbers below 8, as these resulted in insufficient separation of dynamic patterns across key developmental stages.

We systematically evaluated the clustering performance when partitioning genes into 8, 9, and 10 clusters (Fig. 3(a)). Our selection of $K = 9$ was principally based on a detailed comparison of the clarity of clustering patterns and their biological significance across these K values.

Specifically, when $K = 8$ or less, some critical period dynamic patterns were not effectively separated (e.g., the peaks for E12.5 and E13.5 were merged within the same cluster, as illustrated by the red curve in the 8-cluster plot in Fig. 3(a)). This obscured the precise signal for specific developmental periods and failed to sufficiently distinguish the dynamic features of this critical transition period.

Conversely, when $K = 10$ or more, although the partitioning was finer, we observed redundancy where multiple clusters peaked simultaneously during the same developmental stage (e.g., E12.5, as shown by red curves in the 10-cluster plot in Fig. 3(a)). This suggested potential pattern redundancy or over-segmentation, increasing the complexity of subsequent functional interpretation; therefore, cluster numbers above 10 were also not considered further.

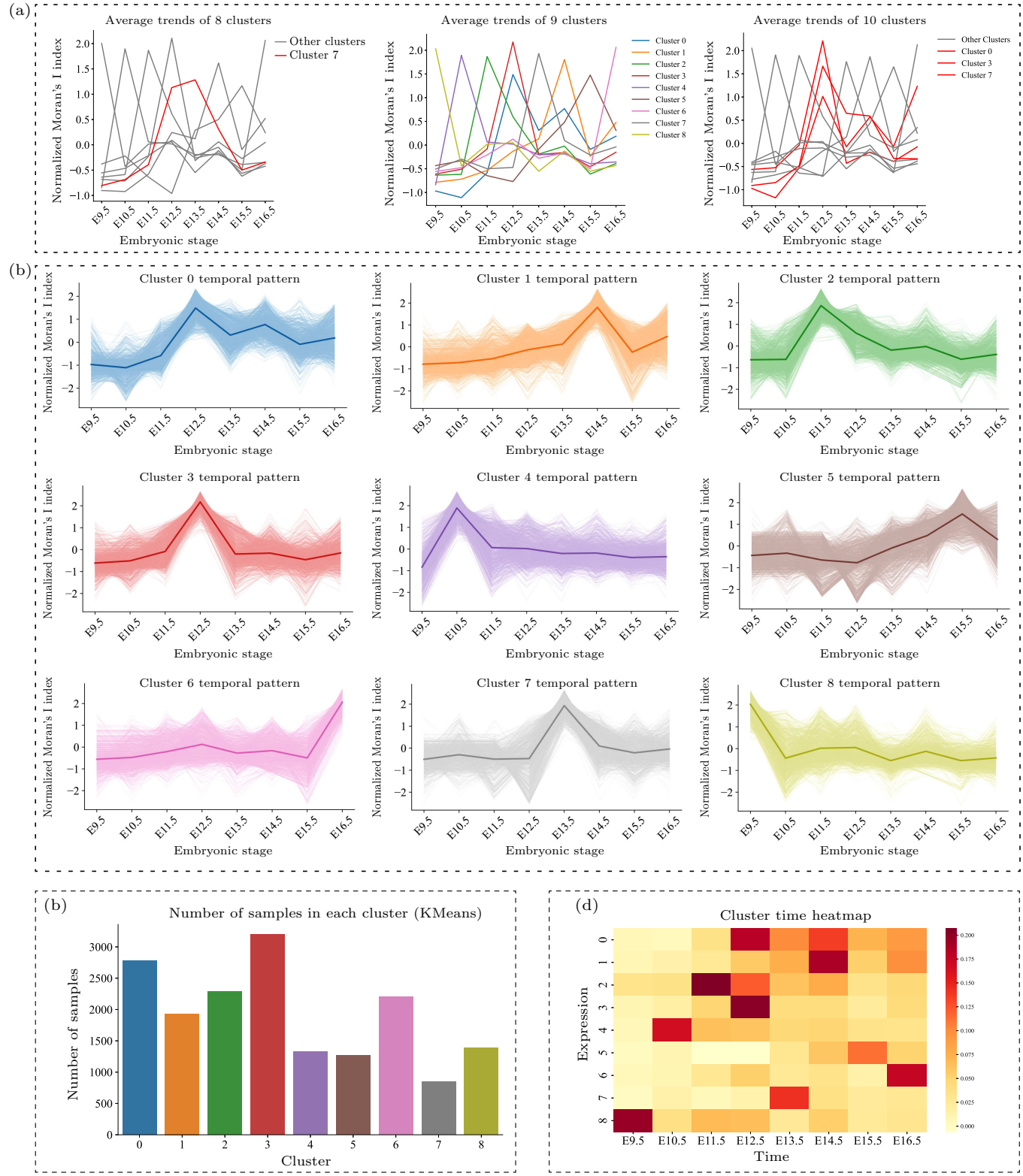

Fig. 3. Time-series clustering analysis of genes during embryonic development stages based on spatial autocorrelation dynamics. (a) Comparison of average Moran's I time curves for each cluster resulting from Tslearn K-means clustering of genes when setting the number of clusters to 8, 9, and 10. (b) Standardized Moran's I time-series curves for each gene cluster in the 9-cluster result. The thick line represents the average time-series trend of all genes within the cluster, while thin lines represent the specific expression trajectories of individual genes. (c) Bar chart showing the number of genes contained in each cluster in the 9-cluster result. (d) Heatmap displaying the average Moran's I value for each gene cluster across different embryonic development stages (E9.5–E16.5), highlighting the temporal specificity and peak activity periods of each gene module.

In contrast, setting $K = 9$ provided an ideal balance. It effectively distinguished different dynamic patterns and, importantly, demonstrated strong biological relevance (Fig. 3(a)). The $K = 9$ partitioning clearly resolved distinct gene modules with different peak activity times corresponding to various developmental stages, including the critical period of organogenesis at E12.5 (Fig. 3(b)). Specifically, the 9-cluster partitioning clearly identified two major clusters (cluster 0 and cluster 3)

whose average spatial autocorrelation reached a significant peak during the E12.5 period (Figs. 3(b) and 3(d)). This time point corresponds precisely to the critical period of organogenesis in mouse embryos, during which significant reorganization of spatial gene arrangement is expected. This concordance between the clustering pattern at $K = 9$ and known developmental milestones strongly supports the choice of $K = 9$ as the optimal number of clusters for our analysis.

Finally, based on the $K = 9$ Tslearn K-means clustering, we partitioned the genes into 9 modules with distinct spatiotemporal dynamic characteristics (Fig. 3(b)). The correspondence between clusters and their approximate average peak times in the clustering results is as follows: (0, E12.5), (1, E14.5), (2, E11.5), (3, E12.5), (4, E10.5), (5, E15.5), (6, E16.5), (7, E13.5), (8, E9.5). Each cluster represents a set of gene instances exhibiting similar temporal dynamics of spatial aggregation. The distribution of the number of genes included in each module is shown in Fig. 3(c). Each module's unique average Moran's I time curve (Fig. 3(b)) and its heatmap pattern across different developmental stages (Fig. 3(d)) reveal their respective temporal preferences, laying the foundation for subsequent in-depth investigation into the functional roles of each module at specific stages of embryonic development.

3.2. Spatial enrichment and functional network at key developmental stages

The results reveal the spatiotemporal regulation of gene expression during the E9.5–E16.5 embryonic development process and the dynamic changes in the biological processes and molecular pathways driven by it. To highlight the significance of spatial autocorrelation, for the cluster(s) peaking at each specific time point, we ranked the genes within them in descending order based on their Moran's I values at that respective time point. For example, genes from cluster 8 (peaking at E9.5) were ranked by their Moran's I at E9.5. Furthermore, for the dual peak observed at E12.5 involving clusters 0 and 3, we combined the genes from both clusters and performed a unified descending sort based on their respective Moran's I values at E12.5.

Starting from the top of this sorted list, we initially selected the top 50 gene instances. Since a single gene might appear multiple times in this initial list due to its high spatial aggregation (high Moran's I values) in different tissues, we then deduplicated this top-50 list to obtain a unique set of genes. Finally, the top 50 genes from this unique list were selected as the gene set representing the dynamic features of that specific time point (e.g., E9.5 and E12.5) and inputted into Metascape for functional enrichment analysis. In this manner, the gene sets we obtained represent groups of genes that exhibit peak spatial aggregation dynamics in at least one tissue during a specific developmental stage.

This section will focus on analyzing the functional enrichment results for four key stages: E9.5, E12.5, E15.5, and E16.5. The corresponding results for the remaining analyzed periods (E10.5, E11.5, E13.5, E14.5) are presented in Figs. S1–S4.

3.2.1. E9.5: Establishment of embryonic polarity and formation of basic tissue architecture

At the E9.5 stage of mouse embryonic development, functional enrichment analysis based on the gene set with high spatial aggregation at this stage revealed that a series of biological processes related to early embryonic patterning, tissue morphogenesis, and basic metabolic regulation were significantly activated. Specifically, pathways such as "Retinol metabolism" and "response to retinoic acid" were significantly enriched (Fig. 4(a)). The retinol signaling pathway, particularly involving its active form retinoic acid (RA), plays a central regulatory role in establishing the embryonic anterior-posterior axis pattern and inducing the development of key structures like the neural tube.[36] The enrichment of these pathways is consistent with the known mechanism whereby RA determines embryonic somite segmentation identity by regulating Hox gene expression.[37] The enrichment of the "response to retinoic acid" process further highlights the fundamental role of RA signaling gradients in coordinating morphogenetic and cell differentiation programs.

In addition to the RA signaling pathway, Fig. 4(a) further reveals other key developmental processes. For instance, terms related to early nervous system development, such as "forebrain regionalization", "cell proliferation in forebrain", and "regulation of neurogenesis", were significantly enriched, indicating that the initial construction and regional specialization of the nervous system are important events at this stage. Concurrently, the enrichment of terms like "tissue morphogenesis", "pattern specification process", and "extracellular matrix organization" collectively reflects that the embryo is actively establishing basic tissue structures and spatial layout during this stage. Furthermore, the significant enrichment of "blood circulation" and pathways related to basic metabolism, such as "post-translational protein phosphorylation", also underscores the importance of maintaining cellular functions and establishing early organ support systems.

The functional network diagram for E9.5 (Fig. 4(a), right panel) further visualizes the interconnections between these processes. This highly interconnected network is dominated by key processes of early embryonic patterning and tissue specialization, exhibiting distinct functional clusters. These include red nodes (post-translational protein phosphorylation), dark blue nodes (blood circulation), pink nodes (intracellular chemical homeostasis), and light orange nodes (retinol metabolism). Additionally, clusters such as yellow nodes (regulation of neurogenesis) and purple nodes (pattern specification process) are densely interconnected, indicating coordinated regulation of multiple developmental processes. The rich connections between nodes reveal the interdependence

of protein modification, morphogenesis, and metabolic pathways. Pathways like pattern specification process and regulation of neurogenesis act as central hubs, reflecting their key roles in signal transduction and gene expression regulation. Regulation of neurogenesis is closely linked to brain development, highlighting the importance of neural processes in establishing brain development patterns.

The E9.5 network illustrates the fundamental processes of embryonic development, establishing embryonic polarity and basic tissue architecture through the synergistic action of morphogenetic gradients and tissue specialization genes. This includes the initiation of the brain and skeletal systems, as well as circulatory mechanisms supporting nutrient transport. The significant enrichment of retinol metabolism and tissue morphogenesis pathways underscores the early embryo's reliance on precise spatiotemporal signal regulation.

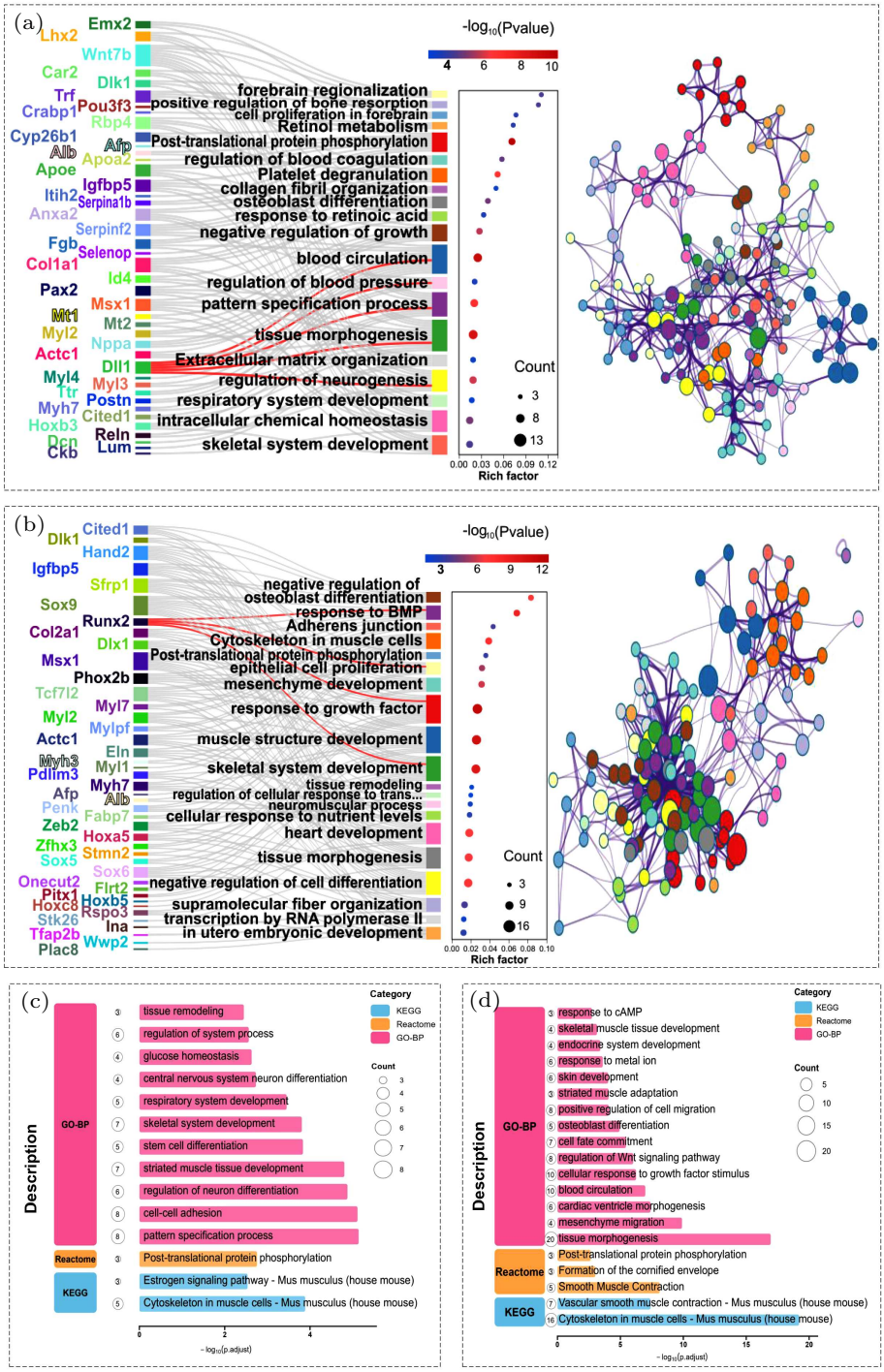

Fig. 4. Functional analysis of gene sets showing peak spatial aggregation at key developmental stages. Analysis results for E9.5 (a) and E12.5 (b). Left: Sankey diagrams link genes to enriched terms, highlighting examples like Dll1 (a) and Runx2 (b). Middle: Bubble plots display enrichment results (color indicates significance $-\log_{10}(P)$, size indicates gene count, x-axis is rich factor). Right: Functional networks visualize term associations (nodes represent terms, sized by significance and colored by functional cluster corresponding to Sankey terms; edge density reflects association strength). Bar charts summarize enrichment results for E15.5 (c) and E16.5 (d), showing significantly enriched terms colored by database (bar length represents significance $-\log_{10}(P)$; numbers indicate gene count).

3.2.2. E12.5: Organ system differentiation and morphological refinement

Entering day 12.5 of embryonic development (E12.5), the developmental focus significantly shifts towards the differentiation, growth, and morphological refinement of specific organ systems. Functional enrichment analysis of the gene set with strong spatial aggregation at this stage (Fig. 4(b)) reveals this transition. This timing closely coincides with key events in cardiac morphogenesis, including heart tube looping, initial partitioning of atria and ventricles, and valve formation. As shown in the comparative study of mouse and human heart development by Krishnan *et al.*, thickening of the ventricular walls and the primordia of the atrioventricular valves are clearly visible by E12.5, laying an important structural foundation for the subsequent functionalization of the circulatory system.[38]

Beyond heart development, the enrichment results in Fig. 4(b) further highlight that the coordinated development of the muscular and skeletal systems is a central theme of the E12.5 stage. Muscle-related pathways, such as "muscle structure development", "cytoskeleton in muscle cellsMus musculus", and "supramolecular fiber organization", all show high enrichment. Concurrently, "skeletal system development" and its regulatory processes, such as "negative regulation of osteoblast differentiation", are also significantly enriched. This indicates that myocyte differentiation, muscle fiber assembly, and the formation and initial shaping of skeletal elements are actively underway at this stage. Coordination of these complex morphogenetic processes relies on precise regulation by signaling pathways. Figure 4(b) shows that "response to BMP" and "response to growth factor" are among the most significantly enriched signaling pathways at this stage, suggesting that BMPs (bone morphogenetic proteins) and various growth factors (like FGFs) play crucial roles in regulating cell proliferation, differentiation, and tissue interactions, especially in coordinating the synchronous development of muscle and bone.[39,40]

These enrichment results are further visually corroborated in the functional association network (Fig. 4(b), right panel). The network structure shows that the blue node group representing "muscle structure development" and the green node group representing "skeletal system development" form the core of the network and are closely interconnected. Concurrently, the purple nodes representing "response to BMP" and the red node cluster representing "response to growth factor" also exhibit dense connections with the node clusters related to muscle and skeletal development. The high degree of connectivity (edge density) both within and between clusters suggests that the differentiation and morphogenesis of the muscular and skeletal systems are under strict, coordinated regulation by these signaling pathways.

In summary, the gene regulatory network at E12.5 focuses on the intensive differentiation and functional specialization of organ systems. Organ formation processes, exemplified by heart and musculoskeletal system development, under the precise control of key signaling pathways like BMP and growth factors, are tightly coupled with tissue morphogenesis and remodeling processes. The enrichment patterns revealed in Fig. 4(b) clearly reflect the key developmental characteristics of the mouse embryo during mid-gestation, transitioning from early broad patterning to the formation of specific organ structures and gradual functional refinement.

3.2.3. E15.5–E16.5: Terminal differentiation and preparation for birth

As embryonic development enters the late stages of E15.5 to E16.5, the main theme of development significantly shifts from early organ patterning and mid-stage system differentiation towards terminal differentiation and functional maturation of various organ systems. This transition is a critical step necessary for the embryo to adapt to the extrauterine environment and prepare for independent survival. The molecular activity underlying this shift is clearly demonstrated through the functional enrichment analysis of spatially aggregated gene sets from this period (Figs. 4(c) and 4(d)).

During the E15.5 stage (Fig. 4(c)), the development of tissues and organs enters a final phase of refinement. Maturation of the nervous system is particularly prominent, with functional enrichment analysis highly concentrated on "central nervous system neuron differentiation" and its regulation ("regulation of neuron differentiation"), suggesting that neurons are undergoing large-scale specialization and establishment of functional connections. Concurrently, other systems crucial for postnatal survival are also undergoing key developments. For instance, the significant enrichment of "respiratory system development" foreshadows the progressive refinement of the structures for future gas exchange. The musculoskeletal system also continues its maturation trajectory, with significant enrichment in "striated muscle tissue development" and "skeletal system development". Moreover, maturation at the cellular structure level, such as the "cytoskeleton in muscle cells" (KEGG), begins to become prominent.

Subsequently, the E16.5 stage (Fig. 4(d)) clearly shifts the developmental focus towards the acquisition of function. Building upon the structural and differentiation foundation established by E15.5, multiple systems exhibit pronounced features of functional maturation. The readiness of the motor system is particularly evident. Enriched pathways not only include ongoing "skeletal muscle tissue development" but also extend to functional adaptation aspects, such as "striated muscle adaptation", as well as pathways directly reflecting function, like "smooth muscle contraction" (Reactome) and "vascular smooth muscle contraction" (KEGG). The skin, as an

important protective barrier, also reaches a peak in its terminal differentiation and functional formation during this stage, evidenced by the significant enrichment of "skin development" and "formation of the cornified envelope" (Reactome), which marks the maturation of the physical barrier. Furthermore, the enrichment of "endocrine system development", along with continued bone tissue differentiation (e.g., "osteoblast differentiation"), points towards the final establishment of specific organ functions. These complex processes of terminal differentiation and maturation are regulated by precise signaling networks. For example, the Wnt signaling pathway ("regulation of Wnt signaling pathway"), cAMP response ("response to cAMP"), and response to growth factors ("cellular response to growth factor stimulus") remain highly active during this stage.

In summary, E15.5 to E16.5 together constitute a critical developmental continuum transitioning from structural refinement to functional completion. Through advanced cell differentiation, tissue functional specialization, and the gradual refinement of physiological regulatory systems, the major organ systems of the embryo (including respiratory, motor, skin, endocrine, etc.) actively undergo final preparations before birth. This series of coordinated terminal differentiation events ensures that the newborn possesses the basic physiological functions necessary for survival in the extrauterine environment.

3.3. Gene-specific functional analysis

By combining clustering analysis with functional enrichment analysis, we can systematically analyze the core pathways and their biological significance at different stages of mouse embryonic development, revealing the dynamic evolution of gene regulatory networks across spatiotemporal dimensions. These analyses not only illuminate the clear progression of embryonic development from early patterning to late tissue maturation but also highlight the temporal orchestration of stage-specific gene networks and their crucial role in driving the transformation of the embryo from a simple cell mass into a complex functional organism.

During the E9.5 stage of polarity establishment and basic tissue architecture formation, functional enrichment analysis showed that protein phosphorylation and the retinoic acid signaling pathway are synergistically activated, driving cell fate decisions and the establishment of the embryonic body plan. Among these, the key gene Dll1,[41] related to somite formation, not only exhibited significant spatially aggregated expression (Fig. 5(a)) but functional enrichment analysis also indicated its involvement in key pathways significantly activated during this period, such as tissue morphogenesis, pattern specification, and regulation of neurogenesis (highlighted connections shown in Fig. 4(a), left panel). This intuitively validates its critical role in regulating these core developmental processes, reflecting the guiding role of early morphogens in axial patterning.

By E12.5, the embryo enters the stage of organ system differentiation and initiation of ossification. Pathway enrichment reveals significant activity in skeletal system development, extracellular matrix remodeling, and muscle structure development. At this stage, the spatial expression of the key skeletal development gene Runx2[42] is significantly enhanced, particularly showing high enrichment in future bone-forming regions such as the craniofacial and axial areas (Fig. 5(b)). Correspondingly, functional enrichment analysis confirms Runx2's involvement in the highly enriched skeletal system development pathways and related regulatory processes active at this time, such as response to BMP and regulation of osteoblast differentiation (highlighted connections shown in Fig. 4(b), left panel). Its central role in osteoblast differentiation and the transformation of cartilage to bone tissue reflects the transition of mid-stage embryonic development towards functional specialization.

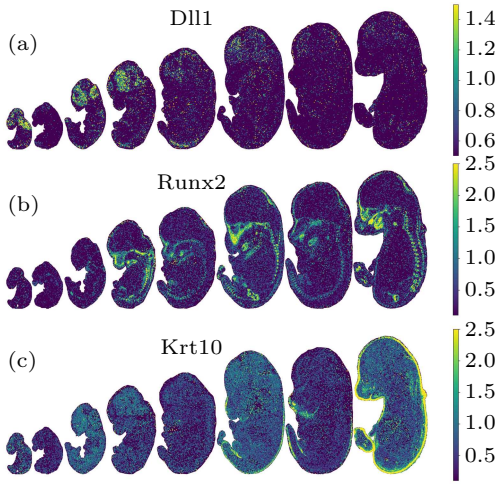

Fig. 5. Spatiotemporal expression patterns of representative stage-specific genes during mouse embryonic development. Spatial transcriptomics data were used to visualize the expression distribution of three key marker genes at different developmental time points. (a) Dll1, involved in somite formation and neurogenesis. Its expression is highly concentrated in the neural and adjacent regions on the brain side of the embryo during early stages (e.g., E9.5). (b) Runx2 is a key regulator of skeletal development, with its expression significantly increased at mid-embryonic stages (e.g., E12.5), mainly localized in craniofacial bones, the spine, and limbs. (c) Krt10, a marker for differentiated epidermal cells. Its expression during late stages (e.g., E16.5) is clearly confined to the outermost skin tissue layer of the embryo. Color intensity in the plots represents the standardized relative expression level of each gene.

By E16.5, the embryo enters the terminal differentiation stage. Functional pathways such as muscle cell cytoskeleton organization, lung development, and hair follicle morphogenesis are significantly activated. At this time, the skin differentiation marker gene Krt10[43] shows clear tissue-specific expression aggregation, primarily localized to the epidermal layer (Fig. 5(c)). This indicates the maturation of the skin and its appendages, laying the structural and molecular foundation for autonomous defense and functional independence after birth.

3.4. Identification of dynamic gene modules

Using time-series clustering methods based on the dynamics of gene spatial expression (Moran's I time series), this study successfully identified gene clusters with similar spatiotemporal expression patterns during mouse embryonic development. Each cluster is considered a co-regulated functional module. These modules exhibit specific biological functions at different developmental stages, collectively driving the continuous process of embryonic development from early morphogenesis to late tissue maturation. Three representative dynamic modules were primarily identified.

First, the early somitogenesis and morphogenesis module (E9.5–E11.5). Gene sets within this module show peak aggregation activity in early embryonic development and are significantly enriched in key signaling pathways regulating embryonic pattern formation, such as Retinoic acid, Wnt, and BMP signaling. The coordinated expression of these genes is crucial for establishing the anterior-posterior and dorsal-ventral axis patterns, guiding somite formation, neural tube closure, and the positioning of major germ layers and tissue primordia. The ordered activation of this module lays down the basic structural blueprint for subsequent complex organogenesis.

Next, the mid-stage organ functional specialization module (E12.5–E14.5). Genes in this module have their aggregation peaks concentrated in mid-embryogenesis. Their functional enrichment characteristics clearly point towards the intensive differentiation and functional refinement of specific organ systems. This module includes numerous key genes involved in the development of the nervous, muscular, and skeletal systems, as well as the establishment of important metabolic pathways. Pathway analysis shows that pathways for skeletal system development, cartilage formation, muscle structure development, and extracellular matrix organization are particularly active during this stage. Concurrently, signaling pathways such as Insulin and IGF signaling also demonstrate important roles during this stage, supporting organ growth and functional construction. Overall, the activation of this module reflects the key transition in embryonic development from establishing basic patterns towards the systematic and functional construction of organ systems.

Finally, the late-stage organ terminal maturation module (E15.5–E16.5). In late embryonic development, the aggregation of genes in this module reaches its peak. Their functions are primarily associated with the terminal differentiation and functional maturation of various organ systems. Enrichment analysis clearly points towards the final refinement of the respiratory system, motor system (final maturation and adaptation of musculoskeletal components), skin system (including the formation of its barrier function), and endocrine system. Genes within this module are crucial for establishing the physiological functions of organs, such as acquiring muscle contraction capability, forming the skin's protective barrier, and establishing endocrine regulatory networks. Activation of this module ensures that the newborn can adapt to the extrauterine environment, achieving the system integration and functional independence necessary for survival.

4. Conclusion

By innovatively integrating the quantification of gene expression spatial aggregation dynamics using Moran's I with time-series clustering analysis, this study systematically modeled and functionally analyzed spatiotemporal transcriptomics data from mouse embryonic development between E9.5 and E16.5. It successfully revealed that embryonic development is driven not by static gene sets but by a "relay-style" activation of a series of gene network modules that possess stage-specific functions and dynamically reorganize in the spatiotemporal dimension. We found that the dynamics of spatial gene expression aggregation themselves contain significant biological information. The identified modules for early morphogenesis, mid-stage organ specialization, and late-stage functional maturation align closely with known developmental milestones. Among these, the early modules sketch the embryonic blueprint centered around morphogen signaling, the mid-stage modules shift towards refining organ structures and functional primordia, and the late-stage modules fine-tune system integration in preparation for birth.

The creation of this "dynamic map" not only deepens the understanding of developmental temporal logic but also highlights the critical role of spatial organization dynamics in coordinating cell behavior and tissue formation. It provides new perspectives for identifying key regulatory nodes, elucidating the roots of developmental diseases, and guiding precise interventions (such as stem cell differentiation control). Methodologically, this study overcomes the limitations of traditional static or non-spatial analyses by capturing spatial aggregation dynamics with Moran's I and combining it with time-series analysis. It provides a quantitative framework for understanding spatiotemporal coordination mechanisms. Furthermore, this dual-dimension integration strategy could potentially be generalized and applied to analyze spatiotemporal data from other species or dynamic biological systems.

Nevertheless, the study's limitations must be acknowledged: Moran's I has limited sensitivity to complex spatial patterns; clustering results are influenced by the algorithm and parameters used; the study is based on a specific dataset (which may have limitations in gene coverage, resolution, and time-point density); it focuses on the transcriptomic level; potential noise during data processing and biases in gene annotations could affect analysis accuracy; and it does not directly address deeper regulatory layers such as proteomics or epigenetics. Additionally, it should be noted that the functional

enrichment analysis in this study was performed based on the overall spatiotemporal dynamic patterns of genes. Although this approach successfully identified macroscopic functional modules associated with major developmental stages, it did not resolve potential tissue specificity within these modules. Therefore, the enriched pathways likely represent a composite reflection of biological processes concurrently active across multiple tissues within that developmental window, rather than the precise functions driven by spatial aggregation within a specific tissue. Identifying and resolving this tissue-level heterogeneity will be an important direction for future research.

Future studies could more deeply explore the fine-tuned regulation of gene spatial organization within specific anatomical structures or cell types and its functional significance by calculating and analyzing tissue-specific Moran's I dynamics, or by incorporating spatial deconvolution methods. Alternatively, applying higher-resolution spatiotemporal multi-omics technologies (e.g., integrating proteomics, metabolomics, chromatin accessibility analysis), developing more optimized spatial statistics, network analysis approaches (such as quantifying network structural entropy[44] or ranking genes within single-cell networks[45]), time-series analysis algorithms (such as deep learning models), and using machine learning models for validation and prediction would greatly expand the depth and breadth of the research.

In conclusion, this study dissected the dynamic reorganization of gene networks in mouse embryonic development from a spatiotemporally integrated perspective, offering new systemic insights into developmental biology. It builds a bridge connecting fundamental scientific discoveries with potential applications in regenerative medicine, disease modeling, and other fields. We anticipate that future technological integration will further refine this dynamic blueprint of life, aiding in the unraveling of the deeper mysteries of life's construction.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant Nos. 12090052, U24A2014, and 12325405). We thank the CNSknowall platform (https://cnsknowall.com) for providing data analysis services.

References

[1] Liberali P and Schier A F 2024 *Cell* **187** 3461
[2] Tian L, Chen F and Macosko E Z 2023 *Nature Biotechnology* **41** 773
[3] Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro J F, Magnusson J, Giacomello S, Asp M, Westholm J O, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder J, Bergmann O, Lundeberg J and Frisén J 2016 *Science* **353** 78
[4] Rao A, Barkley D, França G S and Yanai I 2021 *Nature* **596** 211
[5] Xu F, Li X, Wu R, Qi H, Jin J, Liu Z, Wu Y, Lin H, Shen C and Shuai J 2024 *Fundam. Res.* in press
[6] Delile J, Rayon T, Melchionda M, Edwards A, Briscoe J, Sagner A, Klein A and Treutlein B 2019 *Development* **146** dev173807
[7] Xu Y, Zhang T, Zhou Q, Hu M, Qi Y, Xue Y, Nie Y, Wang L, Bao Z and Shi W 2023 *Nature Cell Biology* **25** 604
[8] Lin H, Hu H, Feng Z, Xu F, Lyu J, Li X, Liu L, Yang G and Shuai J 2024 *Nucleic Acids Res.* **52** 6114
[9] He Q, Li X, Zhong J, Yang G, Han J and Shuai J 2024 *Smart. Med.* **3** e20240014
[10] Moran P A P 1950 *Biometrika* **37** 17
[11] Augustine D J, Te Booth Dr, Cox S E and Derner J D 2012 *Rangeland Ecology & Management* **65** 39
[12] Griffith D A and Chun Y 2016 *Remote Sensing* **8** 535
[13] Yang J, Song C, Yang Y, Xu C, Guo F and Xie L 2019 *Geomorphology* **324** 62
[14] Zhang J, Liu Z, Zhao W, Li C, Liu F and Wang J 2025 *Smart. Med.* **4** e70005
[15] Qiu Z, Li S, Luo M, Zhu S, Wang Z and Jiang Y 2022 *Front. Neurosci.* **16** 1086168
[16] Lin Y, Wang Y, Liang Y, Yu Y, Li J, Ma Q, He F and Xu D 2022 *Front. Genet.* **13** 912813
[17] Li H, Calder C and Cressie N 2007 *Geographical Analysis* **39** 357
[18] Cover T and Hart P 1967 *IEEE Trans. Inf. Theory* **13** 21
[19] Tavenard R, Faouzi J, Vandewiele G, Divo F, Androz G, Holtz C, Payne M, Yurchak R, Rußwurm M, Kolar K and Woods E 2020 *Journal of Machine Learning Research* **21** 118
[20] Temur N, Eryilmaz-Eren E, Celik I, Yıldız I E, Nisari M, Unal I S, Celik C, Ildiz N and Ocsoy I 2025 *Smart. Med.* **4** e70007
[21] Zhong C, Ang K S and Chen J 2024 *Nature Methods* **21** 2072
[22] Liu J, Zhi X, Fang X, Li W, Zhao W, Liu M, Lai E, Fang W, Wang J, Zheng Y, Zou J, Fu Q, Cui W and Zhang K 2025 *Smart. Med.* **4** e70010
[23] Satopaa V, Albrecht J, Irwin D and Raghavan B 2011 2011 31*st International Conference on Distributed Computing Systems Workshops*, June 20–24, 2011 Minneapolis, MN, USA p. 166
[24] Ding H, Xue L, Zhu S, Wang Z, Gui H, Zhang G, Tang N, Ren H and Chen D 2025 *Smart. Med.* **4** e70009
[25] Zhu F, Niu Q, Li X, Zhao Q, Su H and Shuai J 2024 *Research* **7** 0361
[26] Petitjean F, Ketterlin A and Gançarski P 2011 *Pattern Recognition* **44** 678
[27] Arthur D and Vassilvitskii S 2007 *Proceedings of the* 18*th Annual ACM-SIAM Symposium on Discrete Algorithms*, January 7–9, 2007 New Orleans, Louisiana p. 1027
[28] Li X, Zhang P, Yin Z, Xu F, Yang Z H, Jin J, Qu J, Liu Z, Qi H, Yao C and Shuai J 2022 *Research* **2022** 9838341
[29] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi A H, Tanaseichuk O, Benner C and Chanda S K 2019 *Nat. Commun.* **10** 1523
[30] Consortium G O 2004 *Nucleic Acids Research* **32** D258
[31] Kanehisa M, Furumichi M, Tanabe M, Sato Y and Morishima K 2016 *Nucleic Acids Res.* **45** D353
[32] Rice J A 2007 *Mathematical Statistics and Data Analysis* 3rd Ed. (Belmont, CA: Thomson/Brooks/Cole) p. 42
[33] Benjamini Y and Hochberg Y 1995 *Journal of the Royal Statistical Society: Series B (Methodological)* **57** 289
[34] McHugh M L 2012 *Biochemia Medica* **22** 276
[35] Murtagh F and Contreras P 2012 *WIREs* **2** 86
[36] Lara-Ramírez R, Zieger E and Schubert M 2013 *Int. J. Biochem. Cell Biol.* **45** 1302
[37] Niederreither K, Subbarayan V, Dollé P and Chambon P 1999 *Nat. Genet.* **21** 444
[38] Krishnan A, Samtani R, Dhanantwari P, Lee E, Yamada S, Shiota K, Donofrio M T, Leatherbury L and Lo C W 2014 *Pediatric Research* **76** 500
[39] Row R H, Pegg A, Kinney B A, Farr G H III, Maves L, Lowell S, Wilson V and Martin B L 2018 *eLife* **7** e31018
[40] Wu M, Wu S, Chen W and Li YP 2024 *Cell Res.* **34** 101
[41] Hrabě de Angelis M, McIntyre J and Gossler A 1997 *Nature* **386** 717
[42] Komori T 2022 *Int. J. Mol. Sci.* **23** 5776
[43] Wang S, Drummond M L, Guerrero-Juarez C F, Tarapore E, MacLean A L, Stabell A R, Wu S C, Gutierrez G, That B T, Benavente C A, Nie Q and Atwood S X 2020 *Nat. Commun.* **11** 4239
[44] Liu Z, Lin H, Li X, Xue H, Lu Y, Xu F and Shuai J 2025 *Brief. Bioinform.* **26** bbae698
[45] Xu F, Hu H, Lin H, Lu J, Cheng F, Zhang J, Li X and Shuai J 2024 *Brief. Bioinform.* 25 bbae091