MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3194-3197

Graph-Driven Reconstruction of the Hepatic PV–CV Axis: A Topology-Aware Framework for Continuous Zonation and Gradient Gene Discovery

Yi Liu[1*]
[1] *School of Biomedical Engineering, Hainan University, Sanya, 572024, Hainan, China.*
[*]*Corresponding Author: Yi Liu, Email: 3352990378@qq.com*

Abstract: The spatial heterogeneity of liver metabolic function, known as liver zonation, is fundamental to its execution of complex detoxification and synthesis tasks. Although single cell sequencing has revealed cellular diversity, traditional methods often struggle to precisely quantify the continuous metabolic gradient from the portal vein (PV) to the central vein (CV) while preserving intact tissue topology. Existing spatial trajectory inference algorithms also face challenges when handling irregular tissue morphologies and noisy data. This study proposes a robust graph theoretic computational framework to reconstruct the PV–CV axis from spatial transcriptomics data. The method first identifies high confidence PV and CV "seed points" using classic marker genes; subsequently, it constructs a spatial k nearest neighbor graph of spots and employs Dijkstra's algorithm to compute geodesic distance fields from each point to the seed sets. A continuous zonation axis coordinate is then defined by normalizing the ratio of bidirectional distances. Furthermore, we introduce local vector flow field analysis and gradient strength metrics to visualize the directionality and sharpness of metabolic gradients, thereby identifying novel zonation genes that vary monotonically along the axis. Application to mouse liver spatial transcriptomics datasets demonstrates that this method successfully reconstructs a smooth PV–CV axis consistent with physiological structures, effectively overcoming geometric biases caused by tissue section deformation. Local flow field analysis clearly visualizes a virtual metabolic flow directed from the portal tract toward the central vein. PV specific and CV specific genes identified based on this axis show high concordance with established literature, and several potential zonation regulators not previously well documented were discovered. This pipeline provides an interpretable, robust, and training free tool for liver zonation analysis. It not only quantifies spatial metabolic gradients but also offers new quantitative metrics for analyzing zonation disruption in liver disease states.
Keywords: portal vein; central vein; liver zonation; metabolic gradients; graph theoretic computational framework

1. Introduction

As the core metabolic organ, the liver does not function uniformly; instead, it exhibits a precise zonal distribution along the direction of blood flow—from the oxygen-rich periportal region (Portal Vein, PV) to the hypoxic pericentral region (Central Vein, CV)[1]. This spatial functional division enables hepatocytes to efficiently carry out distinct biochemical processes, including gluconeogenesis, the urea cycle, lipid metabolism, and drug detoxification. Although classical histological studies have characterized this macroscopic landscape, approaches to resolve this continuous gradient at the whole-genome molecular level have long been lacking. In recent years, the rapid advancement of spatial transcriptomics (ST) has enabled transcriptome-wide expression profiling while preserving tissue spatial information. However, extracting a continuous physiological axis from discrete spatial spots or cellular data remains a key challenge in computational biology. Existing methods mainly rely on: (1) simple linear regression fitting of spatial coordinates, which ignores the complex hexagonal or polygonal topology of liver lobules; (2) deep learning-based latent variable models, which, despite their high capacity, often suffer from limited interpretability and require large training datasets; (3) threshold-based segmentation using a small set of marker genes, which discards fine-grained information in transition zones. Furthermore, tissue section deformation, irregular vascular organization, and technical noise can all invalidate simple Euclidean distance-based modeling. To overcome these limitations, this study developed a computational pipeline based on graph theory. We hypothesize that changes in hepatocyte metabolic status propagate continuously along the liver sinusoidal network; therefore, geodesic distances derived from a spatial adjacency graph better represent genuine physiological paths than Euclidean distances. Our core strategy is to use well-characterized marker genes to identify high-purity PV and CV "seed points" as graph anchors, compute the shortest path distance from each cell to these two anchor sets, and establish a standardized PV–CV axis via normalized ratio mapping. Building on this framework, we further introduce a physics-inspired flow field concept. By quantifying the coupling between spatial vectors and axial gradients within local neighborhoods, we directly visualize the directional flow of metabolic gradients. Specifically, the objectives of this work are to: (1) construct a robust, topology-aware PV–CV axis that corrects for geometric artifacts in tissue sections; (2) define novel metrics for local metabolic flow and gradient sharpness to characterize functional boundaries; and (3) leverage this continuous coordinate system to discover novel zonation genes and provide an interpretable tool for analyzing spatial metabolic disruptions in liver diseases. This study not only recapitulates the classic liver zonation gene expression profile but also exploits the high resolution of the reconstructed axis to identify a panel of novel candidate zonation genes, providing a general, transparent, and efficient analytical framework for studying spatial reconstruction under both physiological and pathological liver conditions.

2. Materials and Methods

2.1 Data Preprocessing and Marker Scoring

Raw spatial transcriptomics data underwent standard quality control, normalization, and log-transformation before being imported into an AnnData object. We selected PV marker genes (e.g., *Glul*, *Oat*, and others widely validated in literature) and CV marker genes (e.g., Sds,*Cyp2f2*). For each spot $_i$, we calculated the average expression of the PV and CV marker gene sets, defined as the marker scores $S_{PV}(i)$and $S_{CV}(i)$:

$$S_{\text{marker}}(i)=\frac{1}{|G|}\sum_{g\in G}E_{i,g} \quad (1)$$

where G is the list of marker genes and $E_{i,g}$ is the gene expression leve.

2.2 PV and CV Seed Detection

To determine the two extreme anchors of the axis, we employed a percentile thresholding method. The sets of spots with SPV and SCV scores above the 90th percentile were selected as the PV seed set IdxPV and CV seed set IdxCV, respectively. These high-confidence regions serve as source points for subsequent graph distance calculations.

2.3 Construction of Spatial Adjacency Graph and Geodesic Distance Calculatio

1) Using the spatial coordinates (x, y) of the spots, we constructed a k-nearest neighbor graph (k=10). In this graph, nodes represent spots, and edge weights are defined as the Euclidean distance between nodes. Based on this sparse graph structure, we used Dijkstra's algorithm[2] to calculate the distance $d_{PV}(i)$ from each node i to the nearest PV seed point and $d_{CV}(i)$ to the nearest CV seed point. This geodesic distance effectively bypasses tissue voids or non-parenchymal regions, reflecting the true topological path of the tissue.

2.4 Definition of PV–CV Axis Coordinates

To eliminate biases caused by absolute distances dependent on tissue size, we defined the normalized zonation axis coordinate .

$$A(i)=\frac{d_{PV}(i)}{d_{PV}(i)+d_{CV}(i)+\epsilon} \quad (2)$$

where $\epsilon= 10^{-8}$ prevents division by zero. This metric ranges from [0, 1], where 0 represents a pure PV phenotype, 1 represents a pure CV phenotype, and intermediate values represent transitional states.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3194-3197

2.5 Local Flow Field and Gradient Strength Analysis

To validate the biological plausibility of the axis, we calculated the local metabolic flow field. For each spot, within its k-nearest neighbors (k=6), we computed the weighted average of spatial displacement vectors and axial changes to generate a flow vector . Simultaneously, the standard deviation of axis coordinates within the neighborhood was calculated as the local gradient strength to identify functional zone boundaries.

2.6 Zonation Gene Identification and Visualization

We calculated the Pearson correlation coefficient between the genome-wide gene expression profiles and the reconstructed axis coordinate A. Genes with the highest positive correlation coefficients were defined as CV-trending genes, while those with the highest negative correlation (i.e., decreasing as the axis increases) were defined as PV-trending genes. Results were visualized multidimensionally using heatmaps, scatter plots, and spatial mapping. All analyses were completed using the Python ecosystem (Scanpy[3], Scipy[4], Scikit-learn[5]).

3 Results

3.1 High-Confidence Identification of PV and CV Seed Regions Based on Marker Genes

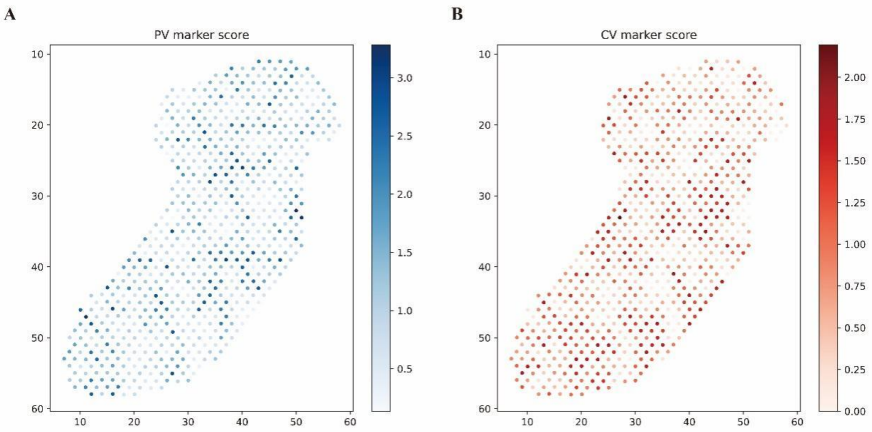

Figure 1. Spatial distribution maps of computed PV and CV marker gene scores on the tissue section.

Scores were calculated based on the average expression of classic PV- and CV-specific genes. These regions were selected as "seed points" for subsequent graph-theoretic analysis. The color bar indicates the intensity of the normalized marker gene score.

To initiate the reconstruction of the liver zonation axis, we first quantified the spatial enrichment of classic marker genes. We calculated the PV marker score (S_{PV}) and CV marker score (S_{CV}) for each spatial spot based on the average expression of established periportal (e.g., *Glul*, *Oat*) and pericentral (e.g., *Sds*, *Cyp2f2*) gene sets[6], respectively. As shown in **Fig. 1B** and **Fig. 1A**, these scores revealed distinct spatial domains: high PV scores localized around the portal triads, while high CV scores clustered around the central veins. We defined spots with scores above the 90th percentile as high-confidence "seed points" Idx_{PV} and Idx_{CV}, which serve as anchors for our subsequent graph-theoretic modeling.

3.2 Reconstruction of a Continuous PV–CV Axis Based on Graph Theory

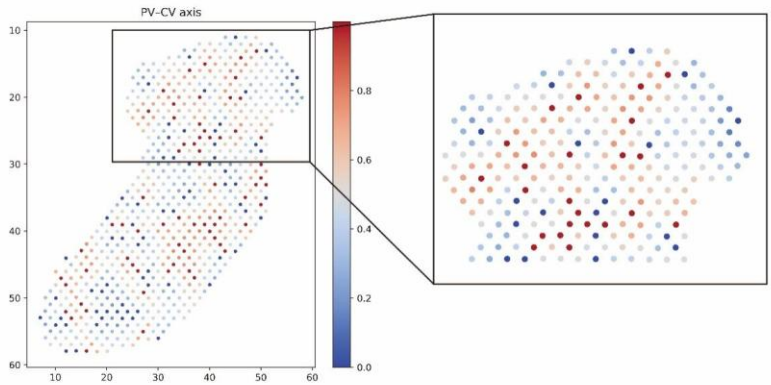

Figure 2. Spatial visualization of the reconstructed PV–CV axis. The axis values range from 0 (pure PV phenotype, blue) to 1 (pure CV phenotype, red).

Considering that liver micro-architecture is often irregular and Euclidean distance may not reflect true physiological paths, we constructed a spatial K-nearest neighbor graph[7] to simulate tissue topology. Using Dijkstra's algorithm, we calculated the geodesic distance from each spot to the nearest PV seed point (d_{PV}) and to the nearest CV seed point (d_{CV}). Subsequently, we defined the normalized PV-CV axis coordinate A as the ratio $d_{PV} / d_{PV} + d_{CV}$. The generated spatial map (**Fig. 2A**) displays a continuous gradient smoothly transitioning from 0 (portal vein end) to 1 (central vein end), effectively connecting discrete vascular structures. This topology-aware axis successfully captures the hexagonal organization of liver lobules and remains robust even in regions where tissue morphology is distorted, providing a reliable coordinate system for downstream analysis.

3.3 Validation via Local Flow Field and Gradient Analysis

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3194-3197

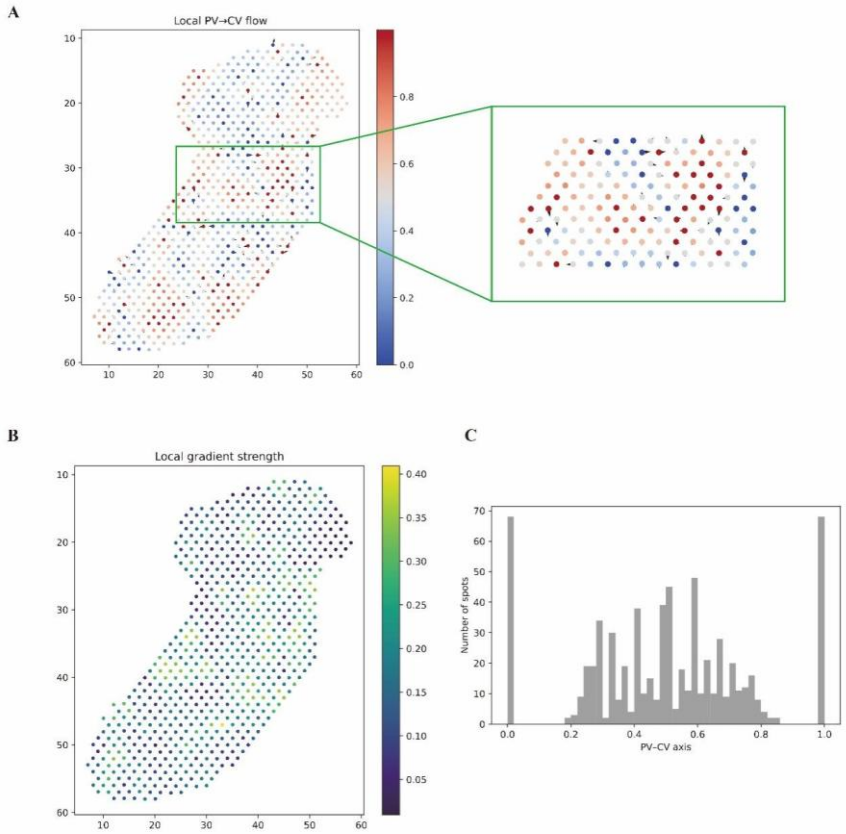

Figure 3. Local gradient variation and flow direction maps. (A) Vector field map superimposed on the PV–CV axis map. Arrows represent local flow vectors obtained by calculating the coupling between spatial displacement and axis gradient among neighboring spots. The arrow direction indicates the predicted propagation direction of the metabolic gradient (from PV to CV). (B) Spatial distribution map of local gradient strength, defined as the standard deviation of axis values within the local neighborhood of each spot. High gradient strength (yellow/green) highlights sharp boundaries between functional zones, contrasting with homogeneous regions in PV or CV cores. (C) Frequency distribution histogram of all tissue spots along the normalized PV–CV axis (0 to 1).

To validate the biological plausibility of the reconstructed axis, we calculated the local metabolic flow field by analyzing the vector coupling between spatial displacement and axis gradient within local neighborhoods. The resulting vector plot (**Fig. 3A**) reveals a coherent directional flow from the portal tract toward the central vein, mirroring the direction of blood perfusion and metabolic signal propagation. Additionally, mapping the local gradient strength[8] (**Fig. 3B**) allows us to identify sharp transition zones where metabolic functions switch rapidly, contrasting with the homogeneous regions in PV or CV cores. The distribution histogram of spot counts along the axis (**Fig. 3C**) confirms that the reconstructed PV–CV axis broadly covers various zonation states, indicating an absence of sampling bias.

3.4 Global Transcriptional Gradients and Discovery of Novel Zonation Genes

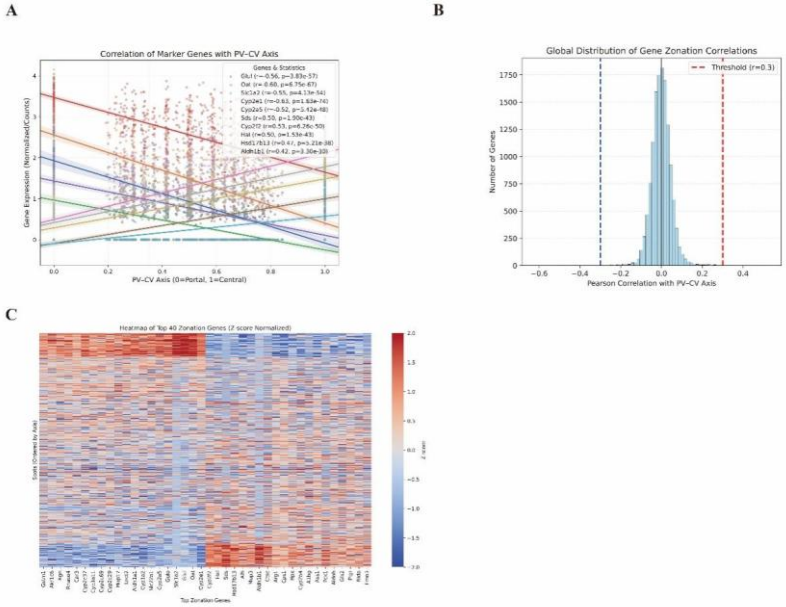

Figure 4. (A) Scatter plots showing the relationship between the expression levels of key PV marker genes (Glul, Oat, Slc1a2, Cyp2e1, Cyp2a5) and CV marker genes (Sds, Cyp2f2, Hal, Hsd17b13, Aldh1b1) and the reconstructed axis values. Each point represents a spot. The strong negative correlation of PV markers and strong positive correlation of CV markers indicate that the computed axis accurately reflects known liver zonation biological gradients. (B) Ranking plot of Pearson correlation coefficients between all detected genes and the PV–CV axis, highlighting genes with the highest positive and negative correlations. (C) Heatmap displaying the expression patterns of the top 20 PV-specific and top 20 CV-specific genes. Rows (spots) are strictly ordered by axis value from 0 (left/top) to 1 (right/bottom). The clear diagonal banding pattern confirms the presence of a highly ordered continuous transcriptional gradient along the reconstructed axis.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 3194-3197

After establishing the continuous axis, we investigated the global transcriptional landscape. Plotting the expression of classic marker genes against the axis values revealed strong monotonic correlations: PV marker genes decreased linearly as the axis value approached 1, while CV marker genes increased linearly (**Fig. 4A**). This confirms that our geometric reconstruction aligns accurately with molecular phenotypes. To visualize the expression patterns of zonation genes, we selected the top 20 PV-specific and top 20 CV-specific genes with the highest correlation to the reconstructed axis. Since baseline expression levels vary significantly among different genes, directly comparing raw counts would cause highly expressed genes to dominate the visual signal. In addition to known markers, this method effectively identifies differential genes between PV and CV regions. First, genes were ranked and filtered based on the Pearson correlation between the reconstructed spatial axis and genome-wide gene expression to select the top 20 genes. Before plotting the heatmap, we performed row-wise Z-score normalization on the expression matrix of the selected genes. Specifically, for each gene, we converted its expression across all spatial spots into Z-scores, ensuring a mean of 0 and a standard deviation of 1 for each gene's expression distribution. This allows the heatmap to clearly display the relative upregulation or downregulation trend of each gene along the PV–CV axis, independent of absolute expression magnitude. Observations revealed that genes such as *Glul* and *Slc22a1*, as well as *Mup3* and *Ctsc*, exhibit significant zoned expression along the PV–CV axis. This provides a new set of marker genes for liver zonation research and reveals potential new features of liver functional zonation.

4 Discussion

This study presents a novel method for reconstructing the liver PV–CV axis based on graph-theoretic geodesic distance. Compared to traditional methods based on straight-line distance or Principal Component Analysis (PCA[9]), our core advantage lies in **topology awareness**. The liver microenvironment is not a uniform Euclidean space; the winding path of liver sinusoids and physical deformation of tissue sections often render straight-line distance inaccurate for reflecting true paths of substance exchange and signal transduction. By introducing Dijkstra's algorithm to calculate geodesic distance, our method adaptively follows tissue structure, reconstructing a smooth and continuous metabolic gradient axis even in the presence of missing vasculature or section artifacts. Our results show that the reconstructed axis coordinates exhibit extremely high monotonic correlation with known marker gene expression, validating the method's effectiveness. In particular, the visualization of the Local Flow Field not only intuitively confirms the unidirectional gradient trend from the portal tract to the central vein but also provides a potential quantitative tool for identifying abnormal blood flow patterns or metabolic retrograde flow (as seen in certain liver disease models). The gradient strength map further reveals the sharpness of zonation boundaries, which may correspond to regions of specific signaling pathway activation thresholds. In terms of gene discovery, in addition to reproducing classic genes such as *Cyp2e1* (PV) and *Cyp2f2* (CV), we identified a batch of novel candidate genes highly correlated with the axis coordinates. These genes may participate in zonation-specific metabolic regulation that has not been fully recognized. Notably, this method does not rely on complex black-box models; every calculation step has clear geometric or statistical significance, making the results highly interpretable and easy for biologists to understand and verify. Of course, this study has certain limitations. First, the selection of seed points relies on prior knowledge of marker genes; if these markers are globally downregulated under specific pathological conditions, it may affect the initialization accuracy of the axis. Future work could explore iterative optimization[10] of seed points or combine unsupervised clustering to automatically identify anchors. Second, the current analysis is primarily based on 2D sections[11] and does not fully capture the complexity of the liver's 3D structure[12]. With the development of 3D spatial transcriptomics technologies, extending this algorithm to 3D graph structures[13] will be an important research direction. In summary, this pipeline provides a powerful tool for the deep mining of spatial transcriptomics data. It is not only applicable to the liver but, with appropriate adjustments (such as modifying seed point definitions), can be extended to other organs with clear spatial polarity, such as the kidney cortex-medulla axis or the intestinal crypt-villus axis[14], aiding in the mapping of finer human organ spatial atlases.

4 Conclusion

In summary, this study establishes a robust, topology-aware computational framework for reconstructing the hepatic PV–CV axis from spatial transcriptomics data. By shifting from Euclidean distance to graph-theoretic geodesic distance, our method effectively overcomes the geometric distortions and irregular tissue morphologies that have long hindered accurate zonation analysis. The reconstructed axis not only recapitulates known metabolic gradients with high fidelity but also introduces novel quantitative metrics—specifically the local metabolic flow field and gradient strength—that provide a dynamic, physics-inspired perspective on liver micro-architecture. Our application to mouse liver data successfully identified a comprehensive panel of canonical and novel zonation genes, demonstrating the pipeline's power in uncovering hidden spatial patterns without relying on black-box models. Beyond basic liver biology, this interpretable and training-free tool holds significant promise for clinical applications, offering a new paradigm to quantify zonation disruption in fibrosis, steatosis, and hepatocellular carcinoma. Ultimately, by bridging the gap between discrete spatial data and continuous physiological reality, this work provides a foundational resource for mapping functional heterogeneity in the liver and paves the way for extending topological analysis to other complex, polarized organ systems.

Data Sharing Agreement:

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

[1] MCENERNEY L, DUNCAN K, BANG B R, et al. Dual modulation of human hepatic zonation via canonical and non-canonical Wnt pathways [J]. Experimental & molecular medicine, 2017, 49(12): e413.
[2] JASON, SIEVER M, VALENTINO A, et al. Dijkstra's algorithm to find the nearest vaccine location [J]. Procedia computer science, 2023, 216: 5-12.
[3] WOLF F A, ANGERER P, THEIS F J. SCANPY: large-scale single-cell gene expression data analysis [J]. Genome biology, 2018, 19(1): 15.
[4] VIRTANEN P, GOMMERS R, OLIPHANT T E, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python [J]. Nature methods, 2020, 17(3): 261-272.
[5] TANAKA T. [[Fundamentals] 5. Python+scikit-learn for Machine Learning in Medical Imaging] [J]. Nihon Hoshasen Gijutsu Gakkai zasshi, 2023, 79(10): 1189-1193.
[6] HILDEBRANDT F, ANDERSSON A, SAARENPää S, et al. Spatial Transcriptomics to define transcriptional patterns of zonation and structural components in the mouse liver [J]. Nature communications, 2021, 12(1): 7046.
[7] PEI S, CHEN H, NIE F, et al. Centerless Clustering [J]. IEEE transactions on pattern analysis and machine intelligence, 2023, 45(1): 167-181.
[8] LIANG Q, SOLIS SOTO L, HAYMAKER C, et al. LSGI: interpretable spatial gradient analysis for spatial transcriptomics data [J]. Genome biology, 2025, 26(1): 238.
[9] YOUNG M D, BEHJATI S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data [J]. GigaScience, 2020, 9(12).
[10] HAN H, WU S, SONG Z, et al. 3D reconstruction of the specular surface using an iterative stereoscopic deflectometry method [J]. Optics express, 2021, 29(9): 12867-12879.
[11] BEN-MOSHE S, SHAPIRA Y, MOOR A E, et al. Spatial sorting enables comprehensive characterization of liver zonation [J]. Nature metabolism, 2019, 1(9): 899-911.
[12] ZHU Y, BALAJI A, HAN M, et al. High-resolution dynamic imaging of chromatin DNA communication using Oligo-LiveFISH [J]. Cell, 2025, 188(12): 3310-3328.e3327.
[13] LIN S, WANG Z, CUI Y, et al. Bridging the dimensional gap from planar spatial transcriptomics to 3D cell atlases [J]. Nature methods, 2026, 23(2): 360-372.
[14] TOUHARA K K, ROSSEN N D, DENG F, et al. Topological segregation of stress sensors along the gut crypt-villus axis [J]. Nature, 2025, 640(8059): 732-742.