

Spatial Transcriptome: Variable Genes Identification Methods

Tianyi Wang *

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai,
200433, China

* Corresponding Author Email: bunny@163.sufe.edu.cn

Abstract. Single-cell RNA (scRNA) sequencing permits the characterization of distinct cellular conditions through the analysis of transcriptomes. Contrasted with bulk sequencing, single-cell methods reveal detailed gene expression, exposing cellular heterogeneity and enabling deep exploration of development and disease. Recently, spatial transcriptomics (ST) emerges and focuses on tissue-level gene expression patterns to capture spatial gene expression profiles. The paper compares and contrasts the approaches currently used to find genes that are differently expressed in ST. This article discusses three noteworthy analysis methods, SpatialDE, BOOST-GP, and CTSV, considers their drawbacks, and suggests new routes for improvement. The purpose of this paper is to provide useful guidelines for performing differential analysis in future ST research by a thorough overview of various techniques. Future research should aim for flexible models, accurate analyses, and richer biological insights. These efforts will enhance comprehension of spatiotemporal gene regulation, illuminate tissue-cell interactions, and advance health and disease studies. This paper is devoted to expanding ST's contribution to life sciences.

Keywords: Spatial Transcriptomics; Differential Gene Identification; Gaussian Process; Zero Inflation.

1. Introduction

Single-cell RNA (scRNA) sequencing technology has played a pivotal role in biological research. Through the sequencing of individual cell genomes or transcriptomes, this technique uncovers distinct cellular characteristics. In comparison to traditional bulk RNA sequencing, scRNA sequencing delves deeper into the molecular states of each cell, thereby shedding light on cellular heterogeneity. scRNA sequencing finds extensive applications in the study of cellular heterogeneity, developmental processes, disease mechanisms, drug development, immunology, neuroscience, and spatial gene expression analysis. This technique is essential for comprehending cellular molecular states, tissue architecture, and disease mechanisms.

As single-cell sequencing continues to advance, there is a growing focus on understanding the correlation between gene expression and the spatial structures within tissues. Over the past few years, the domain of ST experienced a rapid expansion, facilitating the investigation of gene expression distributions within tissues and revealing distinct "spatial patterns" of gene expression. The emergence of various ST techniques drives progress in this domain, including but not limited to the following methods:

1. Microdissection-based: Selecting regions of interest within samples, microdissection isolates these regions, followed by RNA sequencing. Representative techniques include LCM, TIVA, tomo-seq, etc. [1, 2, and 3].

2. In Situ Hybridization (ISH): Probes with distinct fluorescent signals hybridize with transcripts. RNA is not extracted from tissue, and detection and visualization occur directly in the original environment. Representative techniques include seqFISH, seqFISH+, and MERFISH [4, 5, 6].

3. In Situ Sequencing (ISS): Through in-situ-hybridization and sequencing techniques, ISS examines gene expression within tissues and acquires valuable single-cell transcriptome data. Representative techniques include ISS and Barista-seq [7, 8].

4. High-throughput In Situ mRNA Capture: Conducting RNA capture within the tissue environment, followed by subsequent sequencing outside of the tissue context. Techniques encompass ST, Slide-seq, HDST, etc. [9, 10, and 11].

In the realm of biological research, as high-throughput sequencing technology continues to evolve, ST emerges as an effective tool for examining gene expression in cellular tissues. Leveraging the spatial coordinates inherent in ST data, researchers can analyze information concerning gene expression and cell or transcript positions [12], including image recognition, spot-cell type resolution and deconvolution, and spatial distribution analysis. Spatial distribution analysis is particularly vital for addressing challenges in ST. Driven by the coordinates of cell spots and the corresponding gene expression levels, this task is commonly divided into four main stages: detecting genes with spatial variations, clustering spatial patterns, analyzing cell-cell interactions, and performing integrated analyses. [13]. the identification of spatially differential genes represents the initial and fundamental step. Conventional methods for differential genetic analysis such as variance analysis (ANOVA), EdgeR, etc. [14], concentrate on detecting overall differences between various groups, often assuming independence of expression among samples. In ST, the focus of differential gene identification lies in detecting gene expression differences across different locations within tissues or cells.

This article aims to explore existing methods for identifying differentially expressed genes in ST, comparing their similarities and differences. The article particularly underscores three representative traditional statistical data analysis methods, examines their limitations, and proposes potential directions for enhancement. Through a comprehensive discussion of these methods, the article seeks to provide valuable guidance for conducting differential analysis in future ST research.

2. Spatial Differential Gene Identification

Within the realm of identifying spatially differential genes, a model is constructed through gene expression profiles and cellular position data. This approach seeks to identify genes that exhibit significant spatial variation, referred to as spatially variable genes (SVGs). Among various methodologies, discussions, and comparisons are prevalent regarding the construction of statistical models based on Gaussian processes and generalized linear models: SpatialDE employs covariance matrices to represent the interrelationships of gene expression values among cells. This method employs likelihood tests comparing various Gaussian kernel linear models against null models to identify genes with spatial variation. Additionally, a general paradigm for establishing ST data using Gaussian processes is provided [15]. Similar to SpatialDE, SPARK considers raw count data directly. It utilizes GLM with a Poisson distribution and identifies SVGs with multiple spatial kernels [16].

In many practical scenarios, ST data often contain a substantial number of zero gene expression counts, exceeding the anticipated amount based on traditional Poisson or negative binomial distributions. This excessive zero-count occurrence can stem from two distinct mechanisms: a completely random zero-count mechanism and a mechanism arising from specific biological processes or technical factors associated with genes. Thus, conventional models of Poisson or negative binomial distribution are challenged to accurately describe the phenomenon of zero inflation in ST data. Consequently, researchers have adapted zero-inflation models from traditional statistics to address the modeling of raw data with excessive zeros. Zero-inflation models consider two potential generation mechanisms: one for zero counts and another for non-zero counts. Frequently utilized models encompass the zero-inflated models including Poisson (ZIP) and negative binomial (ZINB) models, which incorporate supplementary parameters to characterize both the presence of zero counts and non-zero counts, enhancing their ability to accurately capture the abundance of excessive zero values present in count-based data.

For scenarios involving a significant number of zero values in the data, within the framework previously discussed involving Gaussian processes and generalized linear models, GPcounts utilizes Gaussian process regression. It employs the negative binomial distribution and its zero-inflated form to model ST data, achieving better fitting performance for count data compared to SPARK. It offers single-sample and two-sample testing, constructing null hypotheses and calculating gene p-values using the chi-square distribution to discern spatial differential genes [17]. Similarly, BOOST-GP, also

based on Gaussian processes, uses the ZINB distribution and employs a Bayesian hierarchical model [18].

Traditional transcriptomic data allows for the identification of gene expression levels in each cell or tissue sample, but it is not possible to determine the spatial positions that correspond to these expression levels. ST data allow us to associate expression with location. However, due to technical limitations, the resolution of ST data is often relatively low, causing gene expression information from neighboring areas to blend together. This implies that in the data, the measured gene expression levels at each position represent the weighted average expression levels of neighboring cell expressions, rather than the true individual cell expression levels. This confusion makes it challenging for researchers to accurately analyze the specific expression patterns and differences at each cell or tissue location. Additionally, there are further methods currently utilized for differential gene identification in ST data analysis, as shown in Table 1.

Table 1. Summary of variable genes identification methods.

	Publication	Method	Description
SpatialDE	Mar-18	Gaussian Process-based Approach	Uses spatial Gaussian processes and likelihood tests to identify spatially variable genes by comparing Gaussian kernel linear models against null models.
Trendsceek	Mar-18	Point Process-based Method	Utilizes point processes to quantify spatial gene expression, thereby uncovering gradients and hotspots of gene expression within low-dimensional projections.
HMRF	Apr-19	Integrating Sequencing and Imaging Data	Combines imaging data, cell type mapping, and hidden Markov random field model to differentiate between intrinsic gene regulatory networks and spatial influences on gene expression.
SPARK	Feb-20	Generalized Linear Spatial Modeling	Utilizes a GLM to model raw data, controlling Type I error and giving high statistical power. It utilizes efficient computation algorithms based on penalty pseudo-likelihood.
scHOT	Aug-20	Higher-Order Interaction Detection	Identifies changes in high-order interactions between genes, applicable to continuous temporal trajectories and spatial contexts, adapting to diverse higher-order measurements.
singleCellHaystack	Aug-20	Non-Clustering Method for Differential Expression	Predicts differentially expressed genes without explicit cell clustering. Utilizes Kullback-Leibler divergence to identify genes expressed in non-randomly located cell subsets across multi-dimensional space.
Giotto-BinSpect	Mar-21	Spatial Network-based Approach	Enriches spatial network neighbors statistically after binarizing gene expression states, identifying genes showing consistent spatial patterns in ST or proteomic analyses.
sepal	Mar-21	Diffusion Process-based Identification	Extracts spatially patterned genes via simulating gene diffusion, minimizing the impact of gene expression levels compared to existing methods.
MERINGUE	May-21	Spatial Autocorrelation and Cross-Correlation Analysis	In 2D and 3D, spatial autocorrelation and cross-correlation analysis are used to identify genes with expression patterns that

			are spatially heterogeneous. Infuses cell-cell communication and groups cells into space. Utilizes a ZINB model in ST data, leveraging Bayesian inference for parameter estimation, considering zero-inflation and employing a negative binomial model.
BOOST-GP	Jun-21	Bayesian Hierarchical Modeling	
SPARK-X	Jun-21	Non-parametric Method for Large-Scale ST	Designed for large-scale ST studies.
SOMDE	Jun-21	Self-Organizing Maps for Cell Clustering	Utilizes self-organizing maps for cell clustering and Gaussian process fitting for node-level spatial gene expression, suitable for large datasets with over 20,000 sequencing spots.
GPcounts	Jun-21	Gaussian Process Regression for Zero-Inflated Data	Employs Gaussian process regression for negative binomial distribution, using variational Bayesian inference for efficient computation and supporting an additional ZINB model for missing data.
BayesSpace	Nov-21	Enhanced Resolution and Clustering through Bayesian Analysis	Enhances resolution and clustering using spatial neighborhood information and Bayesian analysis, revealing tissue-specific transcriptional features and heterogeneity more accurately.
SpaGCN	Nov-21	Graph Convolutional Networks Integration	Integrates gene expression, position, and histology by graph convolutional networks. This approach detects enriched spatial expression patterns by aggregating gene expression data from neighboring positions.
SpatialDE2	Nov-21	Unified Framework for Spatial Variable Gene Detection	Comprehensive software integrating tissue region mapping and spatial variable gene detection, employs a Bayesian framework considering Poisson count noise.
C-SIDE	Feb-22	Cell Type-Specific Gene Expression Modeling	Establishes gene expression models specific to cell types, considering relationships and localization, enabling statistical inference for multiple samples and repeated measurements.
STAGATE	Apr-22	Graph Attention Autoencoder for 3D Expression Domain Extraction	Expands its application to multiple consecutive tissue sections, mitigates batch effects and efficiently extract expression domains in 3D tissue. This is achieved through the utilization of graph attention autoencoders.
SpaGene	Aug-22	Model for Pattern Discovery without Modeling	A model-free approach rapidly discovering spatial patterns in large-scale omics studies, identifying co-localized interactions.
CTSV	Sep-22	Integration of ZINB Distribution Model	Identifies cell type-specific spatial variable genes by combining ZINB model, cell type ratio, and spatial effect function.
scGCO	Sep-22	Efficient Scalability using Fast Hidden Markov Random Field Optimization	Particularly effective for large datasets, this method's scalability remains linear even with substantial data volume.

STAMarker	Nov-22	Deep Learning-based Integrated Framework	Integrates graph attention autoencoders, multi-layer perceptron classifiers, and backpropagation gradient to generate significance maps, suitable for large-scale spatial omics studies.
SpatialCorr	Dec-22	Predefined Gene Set Correlation Testing	Examines the correlation of predefined gene sets within and between tissue regions to evaluate the presence of spatially induced variations.
SINFONIA	Feb-23	Integrated Strategy for Spatial Variability Identification	Identifies spatially variable genes using an integrated approach.
SMASH	Mar-23	Non-parametric Method for Spatial Variability Identification	Identifies spatially variable genes using non-parametric methods, balancing statistical power and computational efficiency.
SpaGFT	Mar-23	Hypothesis-Unrelated Model based on Graph Fourier Transform	Utilizes a hypothesis-unrelated model based on graph Fourier transform.
nnSVG	Apr-23	Nearest Neighbor Gaussian Process-based Approach	Uses a scalable method based on nearest neighbor Gaussian processes to continuously identify genes with spatial changes within defined spatial regions, estimating the length scale parameter with gene-specific Gaussian process models.
SPIN-AI	May-23	Deep Learning Model for Predicting Cell Spatial Organization	A deep learning model designed to identify genes that can predict cell spatial organization.
NoVaTeST	Jun-23	Framework for ST Data	A framework for ST data that identifies genes with location-dependent noise variance, capturing spatial variability signal more accurately.
Maxspin	Jun-23	Information Theory Perspective with Jensen-Shannon Divergence	Utilizes an information theory perspective by identifying genes with spatially changing expression patterns through Jensen-Shannon divergence.

In the realm of ST data analysis, various methods are employed to ascertain the proportions of distinct cell types at specific spatial locations and to quantify the gene expression levels of individual cells. SPOTlight uses non-negative least squares initialization to deconvolution ST data and detects locations [19]. SpatialDWLS initially employs enrichment analysis to establish a high likelihood of cell type presence at each location. Subsequently, it employs damped-weighted least squares to accurately ascertain the composition of cell types at specific positions. [20]. RCTD utilizes supervised learning with cell type markers to classify cells in ST, enabling cross-platform cell classification [21]. Methods such as DSTG, Cell2location, and SpatialDecon are employed for spatial deconvolution [22, 23, and 24].

In differential gene identification methods, several approaches based on spatial deconvolution are used to further analyze cell types. SPARK-X, a non-parametric method designed by the same team as SPARK, identifies cell type-specific SVGs. Using the RCTD method, estimates of cell type proportions are obtained, and subsequently, SVGs from spots with the same labeled cell types are detected [25]. CTSV adopts the SPOTlight deconvolution approach, obtaining estimates of cell type proportions at each observation point. By utilizing the ZINB distribution to fit the raw data and weighting by cell type proportions at each position, it explores spatial differential genes specific to particular cell types [26].

SpatialDE, BOOST-GP, and CTSV respectively present general paradigms modeled using statistical methodologies, with considerations for zero inflation and cell type effects. They progressively delve into the gene-cell spatial relationship. In the following sections, specific introductions for these three methods will be provided. Additionally, there are further methods currently utilized for differential gene identification in ST data analysis.

3. Method Analysis

3.1. SpatialDE

SpatialDE builds a model based on Gaussian processes to investigate the connection between cell expression values and their spatial positions. For gene expression $y = (y_1, y_2, \dots, y_N)$ on spatial coordinates $X = (x_1, x_2, \dots, x_N)$, it assumes a multivariate normal distribution:

$$P(y | \mu, \sigma_s^2, \delta, \Sigma) = N(\mu \cdot I, \sigma_s^2 \cdot (\Sigma + \delta \cdot I)). \quad (1)$$

Here, the form of the covariance matrix is pivotal in testing spatial relationships. Intuitively, if gene expression values are related to spatial positions, the off-diagonal elements of the covariance matrix would be zero, implying no correlation between gene expression values at different locations. Otherwise, there would be correlation among expression values. The hypothesis test for the presence of spatial correlation is equivalent to the null hypothesis:

$$P(y | \mu, \sigma^2) = N(\mu \cdot I, \sigma^2 \cdot I). \quad (2)$$

For p-value estimation, a chi-squared distribution is employed to compare the probability of the fitted SpatialDE with that of the model aligned with the null hypothesis. Given the spatial coordinates inherent in ST problems, SpatialDE employs various kernel functions constructed from distances between spatial coordinates to represent the corresponding covariance matrix $\Sigma_{i,j} = k(x_i, x_j)$. Three kernel types are tested in the covariance structure: squared exponential, linear, and periodic. For genes showing significant spatial correlation, the most suitable spatial covariance matrix is chosen based on the Bayesian Information Criterion (BIC) to represent their spatial expression patterns.

SPARK, similar to the SpatialDE framework, is also based on the Gaussian process framework. However, SPARK directly models the raw count data by introducing a Poisson distribution, which is considered more appropriate for modeling the original count data acquired from ST platforms. SPARK provides clearer directions in terms of sample normalization and p-value calibration. It incorporates 10 kernels with different hyperparameters to identify spatial patterns.

SpatialDE provides a paradigm for analyzing ST data using spatial Gaussian processes. In subsequent research, multiple methods based on this framework have delved deeper into analyzing ST data. The following two methods introduced share similar theoretical frameworks with SpatialDE, focusing on the unique characteristics of multi-zero, single-cell multi-cell type ST data.

3.2. BOOST-GP

In contrast to the former method, BOOST-GP takes into account the presence of multiple zero values in ST data by introducing a zero-inflated model. It employs a Bayesian hierarchical pattern for analyzing ST data. To simulate zero inflation and overdispersion in the accounts, a ZINB model is employed. For gene expression, the model is established as follows:

$$y_{ij} | \pi_i, \lambda_{ij}, \phi_j \sim \pi_i I(y_{ij} = 0) + (1 - \pi_i) NB(s_i \lambda_{ij}, \phi_j). \quad (3)$$

Here, y_{ij} represents the gene expression value, $\pi_i \in (0,1)$ is the additional zero value proportion, $I(\cdot)$ is the indicator function, and $NB(\mu, \phi)$ represents the negative binomial (NB) distribution. A

statistical model is formulated to show the standardized expression level of gene j in sample i as follows: $\log \lambda_{ij} = x_i^T \beta_j + \omega_j(t_i)$. The model incorporates covariates x_i and the spatial random process $\omega_j(t_i)$. Similar to the SpatialDE approach, it assumes the spatial random process to follow a stationary Gaussian process with zero-mean and assesses whether the gene in question exhibits spatial variation by testing whether the covariance matrix is diagonal or follows a squared exponential structure.

In terms of parameter estimation, BOOST-GP employs a conjugate Bayesian method. Each coefficient in β_j is given a normal prior, and σ_j^2 is given an inverse gamma prior, i.e., $\beta_j | \sigma_j^2 \sim MVN(0, h\sigma_j^2 I)$ and $\sigma_j^2 \sim IG(a_\sigma, b_\sigma)$.

Compared to the previous methods SpatialDE and SPARK, BOOST-GP utilizes a NB distribution to model count data in ST data analysis. The negative binomial distribution is widely applied in RNA-seq analysis tools (such as DESeq2 and edgeR) to address overdispersion observed in real sequencing data [27]. The accuracy of identifying differentially expressed genes is significantly improved by this method, which significantly improves model fitting.

BOOST-GP employs a Bayesian inference framework for parameter estimation, enabling the quantification of uncertainties and providing more reliable results. Moreover, this method can provide confidence intervals for parameter estimates, aiding in better understanding and interpreting research outcomes.

3.3. CTSV

Building upon the methods discussed earlier, CTSV takes into account overdispersion, zero-inflation, and cell types by utilizing a ZINB distribution model:

$$Y_{gi} \sim \pi_g \delta_0 + (1 - \pi_g) NB(c_i \lambda_{gi}, \psi_g). \quad (4)$$

Here, π_g represents the probability of gene g being technically zero at location i , and δ_0 denotes the Dirac measure at zero.

Considering that each spatial position in ST data corresponds to multiple cell types, the CTSV method employs a deconvolution technique similar to SPOTlight to infer the cell types and their proportions at each individual spatial position. Building upon the previously discussed negative binomial modeling, the natural logarithm of a particular variable, denoted as λ_{gi} is represented as a sum of weighted contributions from various cell types. This represents the level of expression specific to the cell type of gene g and position i :

$$\log \lambda_{gi} = \sum_{k=1}^K \mu_{gki} w_{ik}. \quad (5)$$

In this context, $w_{ik} = \mu_{gki}$ represents the proportion and the cell type k 's relative average expression level at position i . μ_{gki} is modeled based on the coordinates s_i of position i :

$$\mu_{gki} = c + \beta_{gk1} h_1(s_{i1}) + \beta_{gk2} h_2(s_{i2}) h_2(\cdot). \quad (6)$$

Where η_{gk} the baseline expression is level of cell type k , and functions $h_1(\cdot)$ and $h_2(\cdot)$ describe the spatial effects of the mean η_{gk} . The goal is to test whether β_{gk1} and β_{gk2} are zero through a ZINB regression model, thereby identifying cell-type-specific spatial variation for each gene.

CTSV focuses on addressing cell-type specificity and cell type proportions. It introduces the notion of genes that display spatial variability that is specific to certain cell types. This highlights the significance of such genes and emphasizes their differentiation from more general spatially varying genes. In practical biological research, certain genes might exhibit spatial variation in specific cell

types but not in others. Moreover, CTSV takes into account overdispersion. Similar to the approach of BOOST-GP, CTSV directly focuses on raw data utilizing ZINB distribution. This approach is more reasonable compared to other methods, as the ZINB distribution employed by CTSV provides a better explanation and modeling of the observed overdispersion and excess zero counts.

4. Conclusion

This study encompassed a comprehensive exploration to underscore the significance of ST techniques in analyzing the correlation between gene expression and cells' spatial positions. Intricate gene expression data can be acquired at the individual cell level, thereby unveiling cellular heterogeneity and enabling profound explorations into aspects such as cell development, disease mechanisms, and beyond. However, traditional statistical methods encounter challenges when analyzing ST data due to characteristics such as high zero counts and the influence of cell types.

This article introduced three representative methods – SpatialDE, BOOST-GP, and CTSV – which are grounded in statistical frameworks for analyzing ST data. SpatialDE establishes a relationship between gene expression and spatial position using Gaussian process modeling, validating spatial correlations. BOOST-GP introduces a negative binomial distribution to model excessive zero counts in count data, enhancing the accuracy of differential gene identification. CTSV, on the other hand, comprehensively considers zero inflation, cell type specificity, and cell type proportions to more accurately identify spatially variable genes among cell types.

However, as the field of ST continues to evolve, numerous avenues for future research warrant exploration. Firstly, adopting non-parametric structures and non-stationary Gaussian process models could better describe complex data structures and spatiotemporal variations. Secondly, the potential introduction of spatial lag interference by latent factors such as cell types should be further investigated, along with strategies to effectively address such influences. Moreover, custom analysis methods can be developed for different datasets, encompassing human and animal tissues, and tailoring to the specific features of diverse biological systems. Additionally, methods to enhance data resolution remain essential, as they enable a more precise reflection of single-cell expression patterns. Finally, the integration of multiple data sources should be pursued, as it promises a more comprehensive understanding of biological complexity.

In summary, the trajectory of future ST research should move towards flexible models, accurate data analysis, and profound biological interpretation. These research directions will aid in comprehending the spatiotemporal regulatory mechanisms underlying intracellular gene expression, revealing interactions between tissues and cells, and exerting a greater impact on health and disease investigations. Through persistent effort and innovation, the application domains of ST can be expanded, offering substantial contributions to the advancement of life sciences.

References

- [1] Emmert-Buck M R, Bonner R F, Smith P D, et al. Laser capture microdissection. *Science*, 1996, 274(5289): 998-1001
- [2] Lovatt D, Ruble B K, Lee J, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nature methods*, 2014, 11(2): 190-196.
- [3] Kruse F, Junker J P, Van Oudenaarden A, et al. Tomo-seq: A method to obtain genome-wide expression data with spatial resolution//*Methods in cell biology*. Academic Press, 2016, 135: 299-307.
- [4] Lubeck E, Coskun A F, Zhiyentayev T, et al. Single-cell in situ RNA profiling by sequential hybridization. *Nature methods*, 2014, 11(4): 360-361.
- [5] Shah S, Lubeck E, Zhou W, et al. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron*, 2016, 92(2): 342-357.
- [6] Moffitt J R, Zhuang X. RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH)//*Methods in enzymology*. Academic Press, 2016, 572: 1-49.

- [7] Ke R, Mignardi M, Pacureanu A, et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nature methods*, 2013, 10(9): 857-860.
- [8] Wilbrey-Clark A, Roberts K, Teichmann S A. Cell Atlas technologies and insights into tissue architecture. *Biochemical Journal*, 2020, 477(8): 1427-1442.
- [9] Ståhl P L, Salmén F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 2016, 353(6294): 78-82.
- [10] Rodriques S G, Stickels R R, Goeva A, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 2019, 363(6434): 1463-1467.
- [11] Dries R, Chen J, Del Rossi N, et al. Advances in spatial transcriptomic data analysis. *Genome research*, 2021, 31(10): 1706-1718.
- [12] Vickovic S, Eraslan G, Salmén F, et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nature methods*, 2019, 16(10): 987-990.
- [13] Zhang M, Sheffield T, Zhan X, et al. spatial molecular profiling: platforms, applications and analysis tools. *Briefings in bioinformatics*, 2021, 22(3): bbaa145.
- [14] Robinson M D, McCarthy D J, Smyth G K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010, 26(1): 139-140.
- [15] Svensson V, Teichmann S A, Stegle O. SpatialDE: identification of spatially variable genes. *Nature methods*, 2018, 15(5): 343-346.
- [16] Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nature methods*, 2020, 17(2): 193-200.
- [17] BinTayyash N, Georgaka S, John S T, et al. Non-parametric modelling of temporal and spatial counts data from RNA-seq experiments. *Bioinformatics*, 2021, 37(21): 3788-3795.
- [18] Li Q, Zhang M, Xie Y, et al. Bayesian modeling of spatial molecular profiling data via Gaussian process. *Bioinformatics*, 2021, 37(22): 4129-4136.
- [19] Elosua-Bayes M, Nieto P, Mereu E, et al. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic acids research*, 2021, 49(9): e50-e50.
- [20] Dong R, Yuan G C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome biology*, 2021, 22(1): 145.
- [21] Cable D M, Murray E, Zou L S, et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature biotechnology*, 2022, 40(4): 517-526.
- [22] Song Q, Su J. DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Briefings in bioinformatics*, 2021, 22(5): bbaa414.
- [23] Kleshchevnikov V, Shmatko A, Dann E, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature biotechnology*, 2022, 40(5): 661-671.
- [24] Danaher P, Kim Y, Nelson B, et al. Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. *Nature communications*, 2022, 13(1): 385.
- [25] Zhu J, Sun S, and Zhou X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome biology*, 2021, 22(1): 1-25.
- [26] Yu J, Luo X. Identification of cell-type-specific spatially variable genes accounting for excess zeros. *Bioinformatics*, 2022, 38(17): 4135-4144.
- [27] Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 2014, 15(12): 1-21.