nature communications

Review article

Categorization of 34 computational methods to detect spatially variable genes from spatially resolved transcriptomics data

Guanao Yan [1], Shuo Harper Hua[2] & Jingyi Jessica Li [1,3,4,5,6] ✉

In the analysis of spatially resolved transcriptomics data, detecting spatially variable genes (SVGs) is crucial. Numerous computational methods exist, but varying SVG definitions and methodologies lead to incomparable results. We review 34 state-of-the-art methods, classifying SVGs into three categories: overall, cell-type-specific, and spatial-domain-marker SVGs. Our review explains the intuitions underlying these methods, summarizes their applications, and categorizes the hypothesis tests they use in the trade-off between generality and specificity for SVG detection. We discuss challenges in SVG detection and propose future directions for improvement. Our review offers insights for method developers and users, advocating for category-specific benchmarking.

In multicellular organisms, cells work cooperatively in tissues and organs to fulfill biological functions. Measuring cells' spatial locations along with transcriptome profiles can help unravel collaborative cell organizations and molecular mechanisms in tissues and organs[1]. Spatially resolved transcriptomics (SRT) technologies have been rapidly evolving to enable high-throughput profiling of transcriptomes at spatial locations, which may represent single cells or groups of cells, in a tissue slice. SRT data provide unprecedented insights into genes' spatial expression patterns, tissue's cellular organizations, and cell-cell communications[2].

To date, major SRT technologies are either imaging-based or sequencing-based, and the two types of technologies have complementary advantages[1]. Imaging-based SRT technologies use fluorescence in situ hybridization (FISH) to measure the expression levels of selected 200–400 target genes at the single-cell or subcellular spatial resolution[1]. Examples of imaging-based SRT technologies include in situ sequencing (ISS)[3], sequential fluorescence in situ hybridization (seqFISH)[4–6], multiplexed error-robust fluorescence in situ hybridization (MERFISH)[7], STARmap[8], ExSeq[9], and 10x Xenium[10]. The high spatial resolution of these approaches makes them particularly useful for studies requiring precise localization of transcripts within tissue sections. For example, MERFISH is widely used to annotate cell types in

subregions of the mouse brain[11,12] and to investigate the spatial interaction between specific cell types in the mouse somatosensory cortex[13]. However, imaging-based SRT technologies cannot provide transcriptome-wide gene expression levels.

In contrast, sequencing-based SRT technologies, such as Spatial Transcriptomics[5], 10x Visium[14], Slide-seq[15], and GeoMx[16], capture transcriptome-wide gene expression at a lower spatial resolution. Each spot's diameter ranges between 10 μm and 100 μm, containing multiple cells of possibly different types[1]. Recently developed sequencing-based SRT technologies[17,18] can achieve higher, even single-cell, spatial resolution, with spot diameters reduced to below 1 μm, despite the significantly higher cost. Sequencing-based SRT technologies enable transcriptome-wide analyses, such as identifying novel marker genes for tissue layers[19], and sequencing-read analysis, such as inferring RNA velocity from spliced and unspliced sequencing reads[20].

Although imaging-based and sequencing-based SRT technologies have different spatial resolutions, for simplicity, we use the same term "spots" to refer to the measured spatial locations for both technologies in this review. Hence, SRT data encompass two components: (1) an expression count matrix of p genes at n spots and (2) a spatial coordinate matrix containing the two-dimensional (2D) coordinates of n spots.

[1]Department of Statistics and Data Science, University of California, Los Angeles, CA 90095-1554, USA. [2]Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA. [3]Department of Human Genetics, University of California, Los Angeles, CA 90095-7088, USA. [4]Department of Computational Medicine, University of California, Los Angeles, CA 90095-1766, USA. [5]Department of Biostatistics, University of California, Los Angeles, CA 90095-1772, USA. [6]Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138, USA. ✉e-mail: jli@stat.ucla.edu
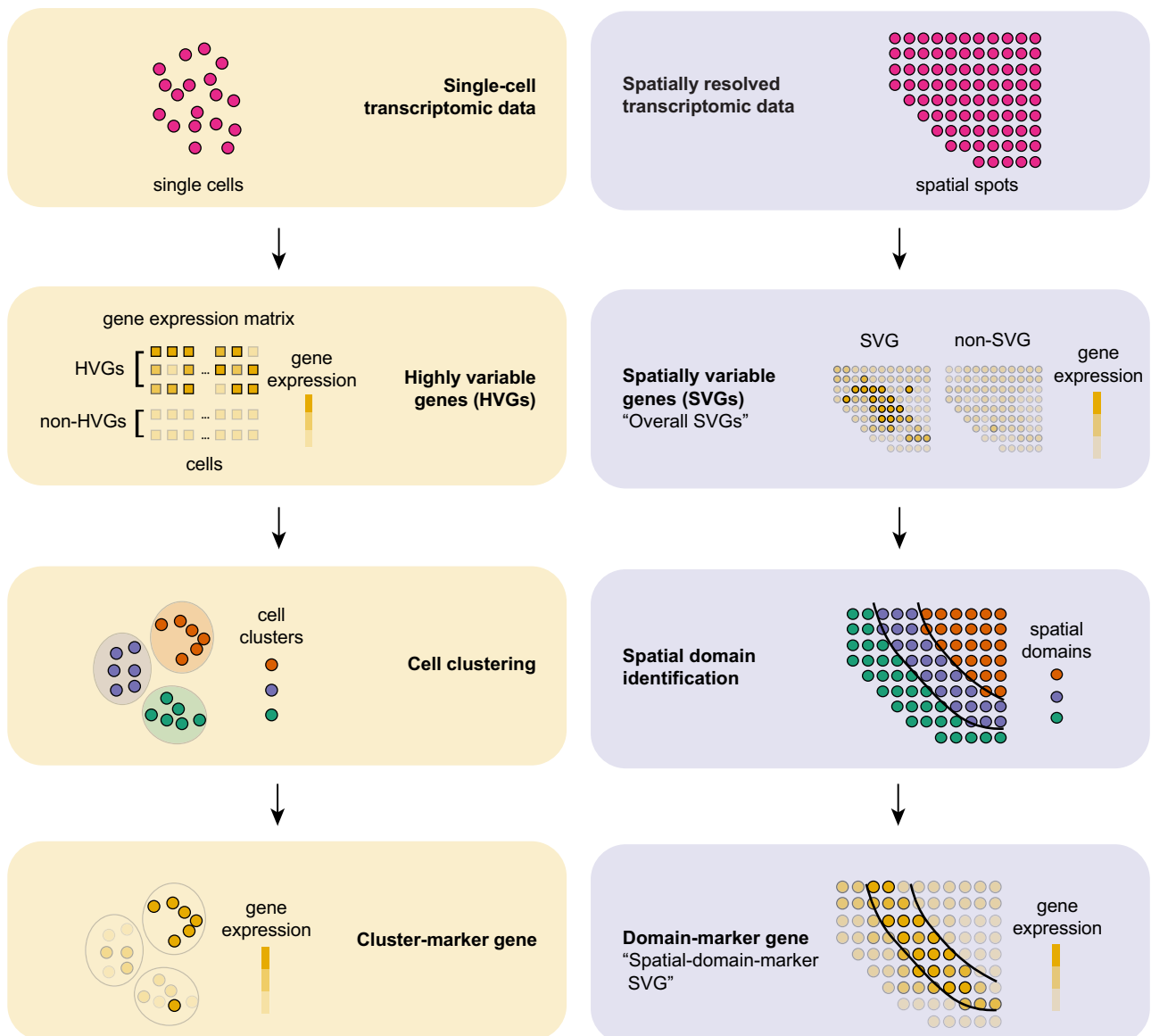
Fig. 1 | General analysis workflows of single-cell transcriptomic data and spatially resolved transcriptomic (SRT) data. The left column shows a general analysis workflow for single-cell transcriptomic data with steps including highly variable gene (HVG) detection, cell clustering, and cluster-marker gene identification. The right column illustrates a workflow for analyzing SRT data with steps including spatially variable gene (SVG) detection, spatial domain identification, and domain-marker gene identification.

As with other types of high-throughput data where hundreds or thousands of genes are measured simultaneously, a crucial early step of SRT data analysis is the identification of informative genes. Prior to SRT data, single-cell RNA-seq data contain gene expression profiles of single cells without spatial information. In typical single-cell RNA-seq data analysis (Fig. 1), highly variable gene (HVG) detection methods are used to screen a proportion of genes (e.g., 10%–20%) with the largest variances (adjusting for cell library sizes in several methods) to reduce the dataset's dimensionality (from p genes to a smaller number of HVGs) and remove the unimportant variations of many genes. The underlying assumption of HVG detection is that genes exhibiting significant expression variations across single cells are more likely to reflect biological variations rather than technical variations caused by sampling effects in sequencing. Analogous to single-cell RNA-seq data analysis, SRT data analysis typically includes an early step to detect *spatially variable genes (SVGs)*, which conceptually generalize HVGs by including the spatial information (Fig. 1). Intuitively, SVGs are genes whose gene expression levels exhibit non-random, informative spatial patterns.

In single-cell RNA-seq data analysis (Fig. 1), common steps after HVG detection include cell clustering and differentially expressed gene (DEG) detection. The goal of cell clustering is to identify potential cell types, and subsequent DEG detection aims to find the genes that are significantly more highly expressed in each cluster. The resulting highly-expressed DEGs of each cell cluster are then used to annotate the cluster as a particular cell type (or subtype), serving as cell-type markers. Similarly, in SRT data analysis, following SVG detection, spots are often partitioned into spatial domains, where each domain contains proximal spots exhibiting similar gene expression profiles. Parallel to DEG detection between clusters in single-cell RNA-seq data analysis, DEGs can be identified between spatial domains, and the resulting highly-expressed DEGs of each spatial domain serve as spatial-domain markers.

Many computational methods have been developed to detect SVGs. However, the SVG definitions in these methods lack consensus, resulting in diverse meanings for the detected SVGs and making understanding difficult. Consequently, the inconsistent definitions of
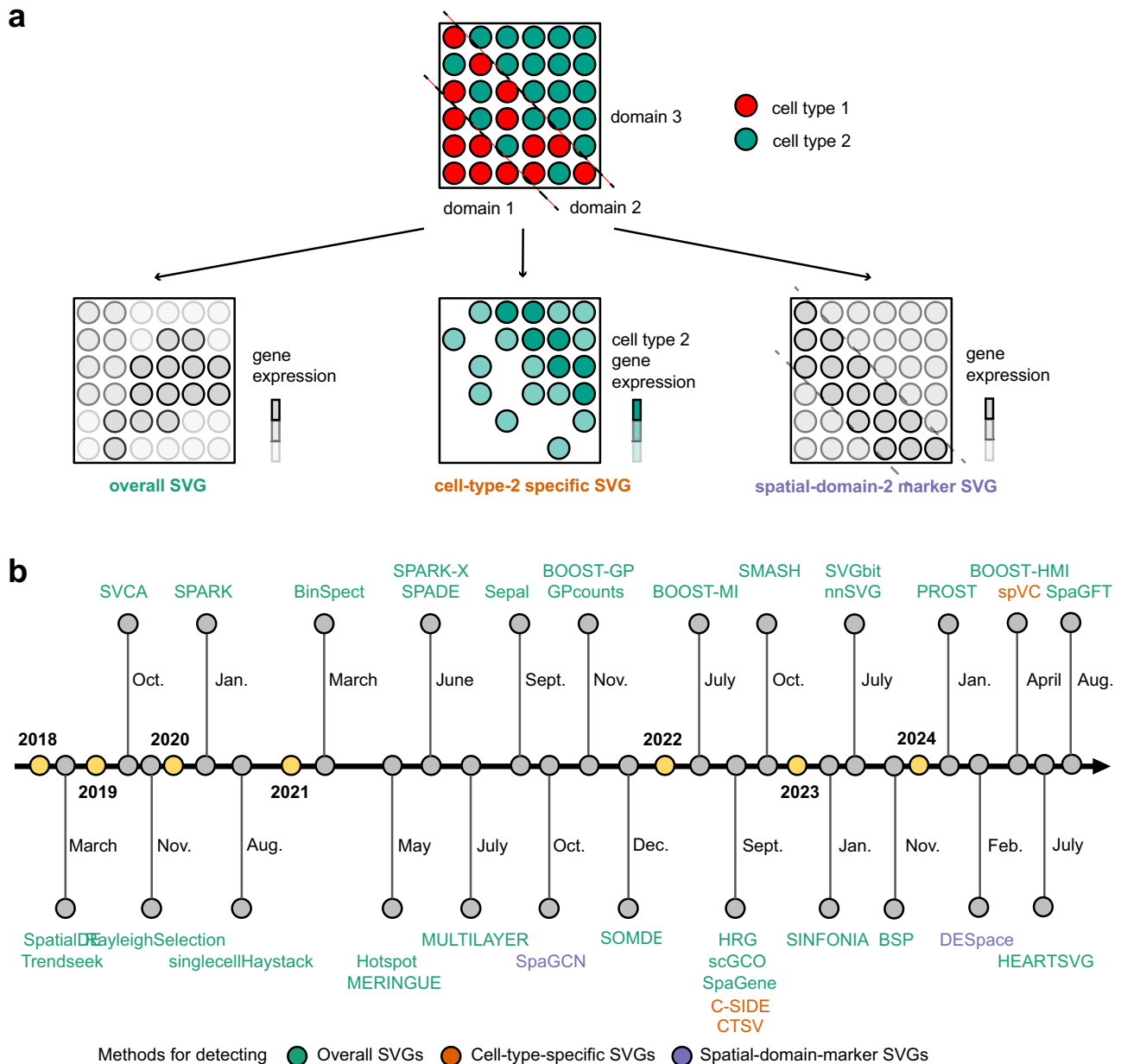
Fig. 2 | Conceptual visualization of three SVG categories and timeline of 34 SVG detection methods. a Conceptual visualization of three SVG categories: overall SVGs, cell-type-specific SVGs, and spatial-domain-marker SVGs. The top row shows a tissue slice with two cell types and three spatial domains. From left to right, exemplar genes with colors representing the expression levels are shown for an overall SVG, a cell-type-specific SVG, and a spatial-domain-marker SVG, respectively. **b** Publication timeline of 34 SVG detection methods. Colors represent three SVG categories: overall SVGs (green), cell-type-specific SVGs (red), and spatial-domain-marker SVGs (purple).

SVGs may lead to ambiguous usage of computational methods. Although 19 methods were reviewed previously[21], the review did not categorize SVG definitions but only focused on summarizing the methodologies, including the input data type (count data vs. normalized data), algorithm type (model-based vs. model-free), statistical paradigm (frequentist vs. Bayesian), the availability of false discovery rate (FDR) control, etc. Nor did the review discuss the biological implications and downstream applications of SVGs.

Motivated by this gap in understanding SVGs, here we review 34 peer-reviewed SVG detection methods (Fig. 2) and define three SVG categories: *overall SVGs, cell-type-specific SVGs*, and *spatial-domain-marker SVGs*. For the first time, our review unveils the biological significance underpinning the three categories of SVGs, summarizes the frequentist hypothesis tests implemented in 23 SVG detection

methods, and discusses the limitations of existing methods and outlines future directions for improvement. Moreover, we construct a hierarchy to summarize the 34 SVG detection methods in terms of methodological differences in Fig. 3 and list the technical details in Table 1 and Supplementary Figs. 1 and 2. Additionally, to help readers better understand our review, we provide Table 2, which lists the definitions of the frequently used terminologies.

Methods for detecting the three SVG categories serve different purposes (Fig. 2a). First, the detection of overall SVGs screens informative genes for downstream analyses, including the identification of spatial domains and functional gene modules. Second, detecting cell-type-specific SVGs aims to reveal spatial variation within a cell type and help identify distinct cell subpopulations or states within cell types. Third, spatial-domain-marker SVG detection is used to find marker
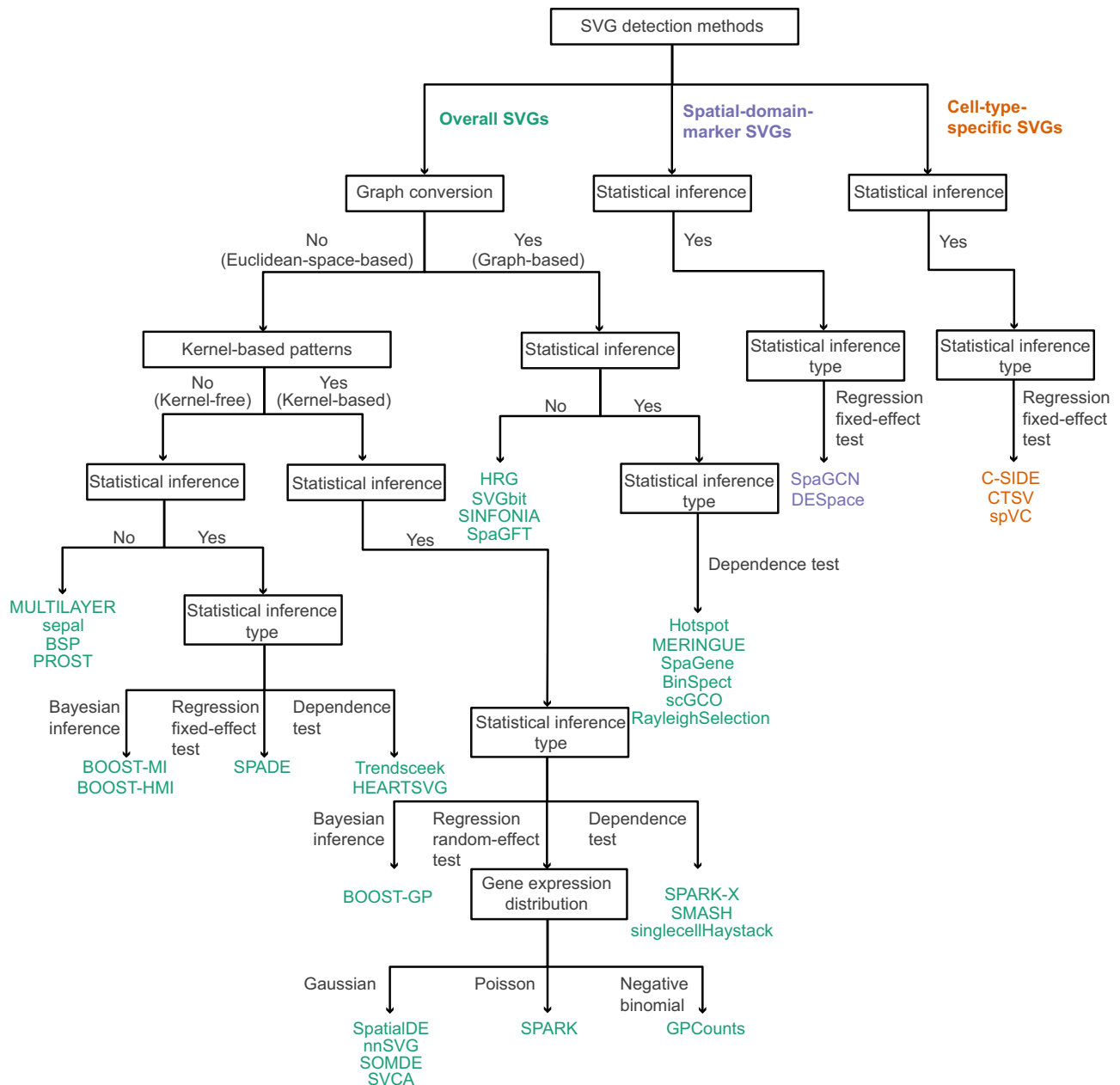
Fig. 3 | A hierarchical summary of 34 SVG detection methods. The hierarchical summary considers the methodological characteristics, including graph conversion, kernel-based patterns, availability of statistical inference, statistical inference types, and gene expression distributions. Colors represent three SVG categories: overall SVGs (green), cell-type-specific SVGs (red), and spatial-domain-marker SVGs (purple).

genes to annotate and interpret spatial domains already detected. These markers help understand the molecular mechanisms underlying spatial domains and assist in annotating tissue layers in other datasets.

The relationship among the three SVG categories depends on the detection methods, particularly the null and alternative hypotheses they employ. If an overall SVG detection method uses the null hypothesis that a non-SVG's expression is independent of spatial location and the alternative hypothesis that any deviation from this independence indicates an SVG, then its SVGs should theoretically include both cell-type-specific SVGs and spatial-domain-marker SVGs. For example, DESpace[22] is a method that detects both overall SVGs and spatial-domain-marker SVGs, and its detected overall SVGs must be marker genes for some spatial domains. This inclusion relationship holds true except in extreme scenarios, such as when a gene exhibits opposite cell-type-specific spatial patterns that effectively cancel each

other out. However, if an overall SVG detection method's alternative hypothesis is defined for a specific spatial expression pattern, then its SVGs may not include some cell-type-specific SVGs or spatial-domain-marker SVGs. For readers with a statistical background seeking deeper insights, we provide a comprehensive discussion on the frequentist hypothesis tests implemented by 23 SVG detection methods in the Section "Theoretical characterization of SVG detection methods that use frequentist hypothesis tests." Other readers can skip this section and still grasp the core content of our review.

Although three benchmark studies were conducted to compare SVG detection methods[23–25], they have three major differences from our review. First, the benchmark studies performed numerical comparisons of SVG detection methods, while our review focuses on categorizing methods conceptually and methodologically. Second, the benchmark studies did not categorize SVGs but focused on detecting

Table 1 | Classification of 34 SVG detection methods along with statistical inference types

Inference types	Categories	Overall SVGs	Cell-type-specific SVGs	Spatial-domain-marker SVGs
Frequentist inference	Dependence tests	SPARK-X		
		SMASH		
		singlecellHaystack		
		Trendsceek		
		HEARTSVG		
		Hotspot		
		MERINGUE		
		SpaGene		
		BinSpect		
		scGCO		
		RayleighSelection		
	Regression fixed-effect tests	SPADE	CTSV C-SIDE spVC	SpaGCN DESpace
	Regression random-effect tests	SpatialDE		
		nnSVG		
		SOMDE		
		SVCA		
		SPARK		
		GPcounts		
Bayesian inference		BOOST-GP		
		BOOST-MI		
		BOOST-HMI		
No statistical inference		MULTILAYER		
		sepal		
		BSP[a]		
		PROST[b]		
		HRG		
		SVGbit		
		SINFONIA		
		SpaGFT[c]		

[a]Although BSP attempts to perform frequentist inference by defining a null distribution for its test statistic, the null distribution is improperly defined as a distribution fitted to the test statistic values of all genes, implying that all genes are non-SVGs. Hence, we do not label BSP as a method with statistical inference in the table.

[b]PROST performs frequentist statistical tests for Moran's I, instead of the PROST index it uses to rank genes as SVGs.

[c]SpaGFT does not perform frequentist statistical tests for the GFT scores it uses to rank genes as SVGs.

the overall SVGs by our definition. For example, spaGCN, a method that detects spatial-domain-marker SVGs, does not perform well in a benchmark study[25], possibly because the study focused on detecting overall SVGs. This observation underscores the need to categorize SVGs. Third, our review has a more comprehensive coverage of SVG detection methods: 34 peer-reviewed methods (all methods to our knowledge) compared to 6 methods in Charitakis et al.[23], 7 methods in Chen et al.[24], and 14 methods (12 peer-reviewed) in Li et al.[25]. In summary, existing benchmark studies complement our review by providing numerical evidence to rank methods in detecting certain types of overall SVGs. Future benchmark studies are needed to compare methods within the three SVG categories and examine more types of null and alternative hypotheses.

To facilitate our discussion, we introduce the following mathematical notations. For each gene with expression levels (in counts or normalized) measured at n spatial spots, we use $y_i \in \mathbb{R}$ to represent its expression level at spot $i = 1, ..., n$. When gene expression levels are considered random variables, we use Y_i to emphasize the random nature. The vector notation is $\mathbf{Y} = (Y_1, ..., Y_n)^\top$. For each spot $i = 1, ..., n$, we denote its 2D spatial location as $\mathbf{s}_i = (s_{i1}, s_{i2})^\top \in \mathbb{R}^2$. The spatial coordinates of all n spots are represented by the matrix $\mathbf{s} = [\mathbf{s}_1, ..., \mathbf{s}_n]^\top \in \mathbb{R}^{n \times 2}$, where each row corresponds to a spot. When spots are annotated with L spatial domain labels, the spatial-domain indicator vector for spot i is $\mathbf{d}_i = (d_{i1}, ..., d_{iL})^\top \in \{0,1\}^L$, with $\sum_{l=1}^{L} d_{il} = 1$; that is, $d_{il} = 1$ means that spot i belongs to spatial domain l. When spots are annotated with K cell type labels, the cell-type proportion vector for spot i is $\mathbf{c}_i = (c_{i1}, ..., c_{iK})^\top \in [0,1]^K$, where c_{ik} indicates the proportion of cell type k at spot i, with $\sum_{k=1}^{K} c_{ik} = 1$. In summary, each spot i has three covariate vectors: spatial location $\mathbf{s}_i$, spatial-domain indicator $\mathbf{d}_i$, and cell-type proportions $\mathbf{c}_i$. We summarize the mathematical notations in Table 3.

Methods for detecting overall SVGs

We define *overall SVGs* as the genes that exhibit non-random spatial expression patterns. This is the most general category of SVGs detected using only SRT data without incorporating external information such as spatial domains or cell types (see Fig. 2a and Table 2). Methods for detecting overall SVGs are generally classified into Euclidean-space-based and graph-based methods (Fig. 3). Euclidean-space-based methods analyze spots in a tissue slice by considering their locations in a 2D Euclidean space. In contrast, graph-based methods first construct a graph of spatial spots and then analyze this graph, focusing on the connections between spots rather than their Euclidean distances. Both approaches, regardless of the spatial representation they use (Euclidean space or graph), study how a gene's expression levels vary across spatial spots, examining the relationship between each gene and its spatial context.

Euclidean-space-based methods

Euclidean-space-based methods are further divided based on whether they use a kernel function to target a specific spatial pattern. Kernel-based methods utilize a kernel function to specify the covariances of spatial spots, enhancing their power to detect the targeted spatial patterns. In contrast, kernel-free methods do not use a kernel function and instead rely on other approaches to capture spatial patterns.

Kernel-based methods. Kernel-based methods use a pre-defined kernel function $K(\mathbf{s}_i, \mathbf{s}_j)$ to specify the covariance between spots i and j based on their spatial locations. This kernel function usually decays as the Euclidean distance between spots i and j increases. Define $\mathbf{K}(\mathbf{s}) = [K(\mathbf{s}_i, \mathbf{s}_j)]$ as the $n \times n$ matrix representing these covariances. The most commonly used kernel functions include the Gaussian kernel for detecting clustered or focal expression patterns, expressed as:

$$K_G(\mathbf{s}_i, \mathbf{s}_j) = \exp\left(-\frac{\left\|\mathbf{s}_i - \mathbf{s}_j\right\|^2}{2\sigma^2}\right), \qquad (2.1)$$

and the cosine kernel for detecting periodic expression patterns, denoted by:

$$K_C(\mathbf{s}_i, \mathbf{s}_j) = \cos\left(2\pi \frac{\left\|\mathbf{s}_i - \mathbf{s}_j\right\|}{\phi}\right). \qquad (2.2)$$

For a given gene, the covariance matrix of its expression vector $\mathbf{Y}$ (may be subject to transformation) can be decomposed into components, one of which depends on $\mathbf{K}(\mathbf{s})$. If the contribution of this term is significantly positive according to a hypothesis test, commonly known as the "variance component test," the gene is detected as an overall SVG.

Table 2 | Definitions of concepts for SVGs and frequentist statistical tests

	Concept	Definition
SVGs	Overall SVGs	Genes that exhibit non-random spatial expression patterns. This is the most general category of SVGs detected using only SRT data without incorporating external information such as spatial domains or cell types.
	Cell-type-specific SVGs	Genes that exhibit non-random spatial expression patterns within a cell type. These genes are detected using both SRT data and external cell-type annotations for the spatial spots.
	Spatial-domain-marker SVGs	Genes that exhibit significantly higher expression in a spatial domain compared to other domains. These genes are detected using SRT data and spatial domains, which are usually detected from the same SRT data.
Frequentist tests	Dependent test	Examines the dependence between a gene's expression level and the spatial location.
	Fixed-effect tests	Examines whether some or all of the fixed-effect covariates $\mathbf{x}_i$ contribute to the mean of the response variable, i.e., a gene's expression.
	Random-effect tests	Examines whether the random-effect covariates $\mathbf{z}_i$ contribute to the variance of the response variable, i.e., a gene's expression.

Table 3 | Mathematical notations and interpretations

Symbol	Interpretation
$n \in \mathbb{Z}^+$	Number of spatial spots
$y_i \in \mathbb{R}$	Expression level of a gene at spot i
$Y_i \in \mathbb{R}$	Random variable representing the expression level of a gene at spot i
$\mathbf{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$	Random vector of a gene's expression levels at all n spatial spots
$\mathbf{s}_i = (s_{i1}, s_{i2})^\top \in \mathbb{R}^2$	2D spatial location of spot i
$\mathbf{s} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]^\top \in \mathbb{R}^{n \times 2}$	Matrix of spatial coordinates for all n spots
$\mathbf{d}_i = (d_{i1}, \ldots, d_{iL})^\top \in \{0, 1\}^L$	Spatial-domain indicator vector for spot i
$\mathbf{c}_i = (c_{i1}, \ldots, c_{iK})^\top \in [0, 1]^K$	Cell-type proportion vector for spot i
$\mathbf{x}_i \in \mathbb{R}^p$	Fixed-effect covariates of spot i
$\mathbf{z}_i \in \mathbb{R}^q$	Random-effect covariates of spot i
$\epsilon_i \in \mathbb{R}$	Random measurement error at spot i
$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top \in \mathbb{R}^n$	Vector of random measurement errors with assumed independence
$\beta_0 \in \mathbb{R}$	(Fixed) intercept in the linear mixed model
$\boldsymbol{\beta} \in \mathbb{R}^p$	Fixed effects in the linear mixed model
$\boldsymbol{\gamma} \in \mathbb{R}^q$	Random effects in the linear mixed model
$\boldsymbol{\Sigma} \in \mathbb{R}^{q \times q}$	Covariance matrix of random effects $\boldsymbol{\gamma}$

Following this general idea, ten kernel-based methods (Fig. 3) make different assumptions about the distribution of $\mathbf{Y}$ and the covariance matrix decomposition. Below, we briefly summarize these methods and highlight their similarities and differences.

SpatialDE[26] assumes that a gene's normalized expression $\mathbf{Y} \in \mathbb{R}^n$ follows an n-dimensional Gaussian distribution, with a covariance matrix that includes a spatial covariance component involving $\mathbf{K}(\mathbf{s})$:

$$\mathbf{Y} \sim \text{MVN}\left(\boldsymbol{\mu}, \sigma_s^2 \cdot \mathbf{K}(\mathbf{s}) + \delta \cdot \mathbf{I}\right),$$

where MVN represents a multivariate Gaussian distribution with a mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and a covariance matrix $\sigma_s^2 \cdot \mathbf{K}(\mathbf{s}) + \delta \cdot \mathbf{I}$. Here, $\mathbf{I}$ is an n-dimensional identity matrix, and σ_s^2 and δ are parameters representing the spatial covariance component and the error variance component, respectively. This model is a realization of a Gaussian process with the kernel function $K(\cdot, \cdot)$. To determine if a gene is an overall SVG, SpatialDE employs a likelihood ratio test with the null hypothesis $H_0 : \sigma_s^2 = 0$, comparing this model with a null model that does not include the spatial covariance component: $\mathbf{Y} \sim \text{MVN}(\boldsymbol{\mu}, \sigma^2 \cdot \mathbf{I})$.

Using the same model as SpatialDE, *nnSVG*[27] improves computational efficiency by replacing the Gaussian process with the nearest-neighbor Gaussian process[28], providing a scalable approximation. This

enhancement allows nnSVG's computational complexity and runtime to scale linearly with n, the number of spatial spots, rather than cubically.

SOMDE[29] also follows the statistical model of SpatialDE but adds a data preprocessing step for scalability. Specifically, SOMDE condenses the original spatial spots into fewer grid points and assigns each grid point a "meta-expression," which aggregates a gene's expression levels at the spots condensed into that grid point. After this preprocessing step, SOMDE uses the same approach as SpatialDE to detect whether the gene is an overall SVG, but it does so using the grid points instead of the original spots.

Compared to the first three methods, *SVCA*[30] modifies the covariance matrix decomposition by adding two additional components:

$$\mathbf{Y} \sim \text{MVN}\left(\boldsymbol{\mu}, \mathbf{K}_{\text{int}} + \mathbf{K}_{\text{c-c}} + \sigma_s^2 \cdot \mathbf{K}(\mathbf{s}) + \sigma^2 \cdot \mathbf{I}\right),$$

where $\mathbf{K}_{\text{int}}$ denotes the intrinsic cell-state covariance, and $\mathbf{K}_{\text{c-c}}$ represents the cell-cell interaction covariance. Specifically, $\mathbf{K}_{\text{int}}$ is defined solely using gene expression data without spatial information, while $\mathbf{K}_{\text{c-c}}$ incorporates both gene expression data and Euclidean distances among spots.

Instead of using normalized gene expression data, *SPARK*[31], *GPcounts*[32], and *BOOST-GP*[33] directly model a gene's expression count at spot i, $Y_i \in \mathbb{N}$, using Poisson, negative binomial (NB), and zero-inflated negative binomial (ZINB) distributions, respectively. For instance, *SPARK* assumes that:

$$Y_i \overset{\text{ind}}{\sim} \text{Poisson}(\mu_i(\mathbf{s}_i)), \quad i = 1, 2, \cdots, n.$$

For spot i, the Poisson mean parameter $\mu_i(\mathbf{s}_i)$ is a function of the spatial location $\mathbf{s}_i$, specified by:

$$\log(\mu_i(\mathbf{s}_i)) = \mathbf{x}_i^\top \boldsymbol{\beta} + b_i(\mathbf{s}_i) + \epsilon_i,$$

where $\mathbf{x}_i$ indicates spot i's covariates, and $b_i(\mathbf{s}_i)$ is the random intercept at spot i. The n random intercepts, $b_1(\mathbf{s}_1), \ldots, b_n(\mathbf{s}_n)$, are assumed to follow a Gaussian process with the kernel function $K(\cdot, \cdot)$:

$$(b_1(\mathbf{s}_1), b_2(\mathbf{s}_2), \cdots, b_n(\mathbf{s}_n))^\top \sim \text{MVN}(0, \sigma_s^2 \cdot \mathbf{K}(\mathbf{s})).$$

To decide whether the gene is an overall SVG, SPARK tests the null hypothesis $H_0 : \sigma_s^2 = 0$. Thus, SPARK shares the same variance component test idea with SpatialDE but uses a hierarchical model, incorporating a Gaussian process in the top layer and Poisson distributions in the bottom layer, to account for count data.

Unlike the previous seven methods that use parametric models, three methods—*SPARK-X*[34], *SMASH*[35], and *singlecellHaystack*[36]—employ non-parametric tests to determine whether a gene's expression is independent of its spatial location.

SPARK-X[34] tests whether two $n \times n$ spot similarity matrices are independent to decide if a gene is an overall SVG. One similarity matrix

is based on the gene's expression levels at the n spots, defined as $\mathbf{E} = \mathbf{Y}(\mathbf{Y}^\top \mathbf{Y})^{-1}\mathbf{Y}^\top$. The other is based on the kernel-transformed spatial locations of the n spots, defined as $\mathbf{K} = \mathbf{s}'((\mathbf{s}')^\top \mathbf{s}')^{-1}(\mathbf{s}')^\top$, where $\mathbf{s}'$ is an $n \times 2$ matrix whose ith row $\mathbf{s}'_i = (s'_{i1}, s'_{i2})^\top$ encode the kernel-transformed spatial location of spot $i = 1, \ldots, n$. Specifically, SPARK-X transforms the spatial location $\mathbf{s}_i = (s_{i1}, s_{i2})^\top$ of spot i using two kernel-based functions: a Gaussian transformation $s'_{il} = \exp(\frac{-s^2_{il}}{2\sigma^2_l})$, $l = 1, 2$, to detect clustered or focal patterns, and a cosine transformation $s'_{il} = \cos(\frac{2\pi s_{il}}{\phi_l})$, $l = 1, 2$, to detect periodic patterns, where σ_1, σ_2, ϕ_1, and ϕ_2 are tuning parameters. Eventually, SPARK-X uses the Pearson correlation between the two similarity matrices, trace($\mathbf{EK}$)/n, as the test statistic to decide if a gene is an overall SVG.

SMASH[35] generalizes SPARK-X by adding two alternatives to the spatial location similarity matrix $\mathbf{K}$. The two alternatives include the Gaussian kernel covariance matrix $\mathbf{K}_G = [K_G(\mathbf{s}_i, \mathbf{s}_j)]$ and the cosine kernel covariance matrix $\mathbf{K}_C = [K_C(\mathbf{s}_i, \mathbf{s}_j)]$, where $K_G(\cdot, \cdot)$ and $K_C(\cdot, \cdot)$ are the kernel functions defined in (2.1) and (2.2), respectively. Then, SMASH uses the same test statistic as SPARK-X.

singlecellHaystack[36] is a test of independence between a gene's expression level and its spatial location. It involves two pre-processing steps: first, the gene's expression levels at various spots are binarized into two states: detected and undetected; second, the 2D Euclidean space of a tissue slice is divided into a grid along both axes, with intersection points defined as grid points, serving as coarse spatial coordinates. After pre-processing, the method tests if the gene's two expression states are randomly distributed across the grid points. If they are not, the gene is considered an overall SVG. singlecellHaystack uses a Gaussian kernel to define three distributions of grid points: a reference distribution based on all spatial spots, a conditional distribution based on spatial spots in the detected state, and another conditional distribution based on spatial spots in the undetected state. Technically, for a grid point with a 2D spatial location $\mathbf{s} = (s_1, s_2)^\top$, its reference distribution density is defined as $\sum_{i=1}^{n} \exp(-\frac{\|\mathbf{s}_i - \mathbf{s}\|^2}{2})$, subject to normalization. The two conditional distribution densities are similarly defined, with the sum across all spots replaced by sums across the detected and undetected spots, respectively. The test statistic is defined as the sum of two Kullback-Leibler (KL) divergences, each representing the deviance of one conditional distribution from the reference distribution. Intuitively, the larger the test statistic, the more likely the gene is an overall SVG. Finally, the test statistic is converted to a p value using a permutation test, which shuffles the gene's expression levels across spots.

To summarize, all ten kernel-based methods require pre-specification of the kernel function $K(\cdot, \cdot)$ to target specific spatial patterns. Consequently, the statistical power of these methods depends on how well the kernel function captures the spatial expression patterns of biologically informative genes in the SRT data. Due to the limited choices and subjectivity of kernel functions, these methods may be insufficient for detecting complex spatial expression patterns, such as those found in cancer SRT data.

Kernel-free methods. Nine kernel-free methods employ diverse approaches to detect overall SVGs without relying on pre-specified kernel functions (Fig. 3). Instead, they utilize various statistical and computational techniques to capture spatial patterns. Below, we briefly summarize each method.

Trendsceek[37] tests if a gene's expression level is dependent of the spatial location using a marked point process, which models the joint probability distribution of spatial locations $\mathbf{s}_i$, $i = 1, \ldots, n$, as "points" and a given gene's expression levels y_i, $i = 1, \ldots, n$, as "marks." If deemed dependent, the gene is detected as an overall SVG. For a pair of points $(\mathbf{s}_i, \mathbf{s}_j)$ and their corresponding marks (y_i, y_j), Trendsceek

parameterizes their joint probability density by the distance (called "radius") $r_{ij} = |\mathbf{s}_i - \mathbf{s}_j|$ and the marks y_i, y_j as $f(y_i, y_j, r_{ij})$. Formally, the dependence of the gene's expression on spatial location is formulated as a conditional distribution of a pair of gene expression levels $(y_i = y_1, y_j = y_2)$, given the radius $r_{ij} = r$:

$$M(y_1, y_2 | r) = \frac{f(y_1, y_2, r)}{f(r)}.$$

This conditional distribution can then be summarized into four mark-segregation summary statistics as functions of the radius r, such as Stoyan's mark-correlation function. To test if the gene is an overall SVG, Trendsceek implements a permutation test by sampling marks without replacement and randomly reassigning them to points. For each radius r, summary statistics are calculated and compared to the null distributions derived from the permutations to obtain p values, which are then combined into a single p value across all radius values. Note that Trendsceek has inspired the "markvariogram" approach used in the single-cell and SRT analysis toolkit Seurat[38] for SVG detection.

MULTILAYER[39], similar to singlecellHaystack, discretizes spatial spots for each gene based on the gene's expression levels. The discretization procedure in MULTILAYER consists of three steps. First, each spot receives a $\log_2$ fold change, defined as the ratio of the gene's normalized expression level at the spot to its average normalized expression level across all spots. The $\log_2$ fold changes with absolute values greater than 1 are truncated to 1 with the corresponding sign. Second, MULTILAYER applies agglomerative clustering to the spots based on the $\log_2$ fold change values, using a pre-specified distance threshold. Third, if a cluster contains contiguous spots with all positive $\log_2$ fold change values, MULTILAYER assigns the cluster a label of "1." Finally, MULTILAYER uses the size of the largest cluster labeled as "1" as a summary statistic for the gene and ranks genes from high to low based on this statistic. Thus, if a gene has a large contiguous cluster labeled as "1," it is considered to have a high-expression neighborhood and is regarded as an overall SVG. However, MULTILAYER does not provide statistical significance for its overall SVG ranking.

sepal[40] is a diffusion-process-based method that identifies a gene as an overall SVG if its spatial expression pattern deviates from randomness. To quantify this deviation, sepal simulates a diffusion process from the gene's observed expression pattern until it converges to randomness, using the convergence time as a measure. The longer the convergence time, the more likely the gene is an overall SVG. However, this method ranks genes but does not provide statistical significance for the detected overall SVGs. Specifically, sepal uses the diffusion equation based on Fick's second law[41] to simulate the diffusion of the gene's transcripts in the 2D Euclidean space of a tissue slice. Note that sepal is included in the SRT analysis toolkit squidpy[42] for SVG detection.

BSP[43] is built upon the intuition that an overall SVG should have its spatial expression pattern vary significantly at different spatial resolutions; otherwise, the pattern should remain consistent. To implement this, BSP follows a four-step procedure. First, for each spot, it defines "big" and "small" patches consisting of neighboring spots within pre-specified large and small radii, respectively, so each spot has one big patch and one small patch. Second, for each gene, BSP calculates the average expression levels within the big patches and the variance of these averages, then similarly calculates the variance for the small patches. Third, BSP uses the ratio of the big-patch variance to the small-patch variance as a statistic to summarize the change in the gene's expression pattern with changing spatial resolution. Finally, BSP ranks genes by this statistic, identifying those with large values as overall SVGs. Although BSP attempts frequentist inference by defining a null distribution for its statistic, this null distribution is improperly defined as a distribution fitted to the statistic values of all genes, implying that all genes are non-SVGs. Consequently, a p value

threshold of 0.05 will always lead to 5% of genes being detected as overall SVGs, regardless of how many genes are truly overall SVGs. Therefore, we do not consider BSP to provide valid statistical inference in our review. While the ranking of genes by the BSP statistic could still be informative, the lack of valid statistical inference means that we do not have a threshold on the statistic to reflect a target FDR.

PROST[44] determines if a given gene is an overall SVG based on computer image segmentation. First, treating the gene's spatial expression in a tissue slice as an image, PROST applies image segmentation techniques to detect multiple foreground regions, where the gene is considered expressed, and one background region, where the gene is treated as unexpressed. Second, PROST defines two statistics: (1) a significance factor summarizing the overall elevation of the gene's expression from the background region to the foreground regions, and (2) a separability factor measuring the overall homogeneity of the gene's expression within each foreground region. Finally, PROST combines these two statistics into a single index. Based on this index, PROST ranks genes from high to low and selects the top-ranking genes as overall SVGs. Although PROST performs frequentist statistical tests, they are for Moran's I[45] rather than its index. Therefore, we do not consider PROST to provide valid statistical inference in our review.

HEARTSVG[46] is a time-series-based method designed to detect SVGs by first converting a gene's expression levels in a 2D tissue slice, structured as a matrix with n_{row} rows and n_{col} columns, into four one-dimensional time series. Notably, HEARTSVG assumes that the tissue slice is square, which may impose a limitation. These four time series, termed "marginal expression series," are generated through a process called "semi-pooling." For example, if a tissue slice has $n_{row} = 120$ and $n_{col} = 60$, HEARTSVG defines the following four time series for each gene: (1) a per-row-average time series with 120 entries, each representing the average expression level within the corresponding row; (2) a within-row-window-average time series with $120 \times \frac{60}{\ln(120)} = 1440$ entries, where each entry is derived from averaging expression levels within a non-overlapping sliding window of size $\ln(120) = 5$ within each row; (3) a per-column-average time series with 60 entries, each representing the average expression level within the corresponding column; and (4) a within-column-window-average time series with $60 \times \frac{120}{\ln(60)} = 1800$ entries, where each entry is derived from averaging expression levels within a non-overlapping sliding window of size $\ln(60) = 4$ within each column. HEARTSVG then applies the Portmanteau test to each of these four time series to assess whether there are significant autocorrelations, which would suggest that the time series displays a trend or periodic pattern, potentially indicating an informative spatial pattern in the original 2D square. The four p values, obtained from the Portmanteau tests on the four time series, are then combined into a single p value using Stouffer's method; that is, the gene should be considered an overall SVG if the combined p value falls below a certain threshold. Finally, to determine the appropriate p value threshold and account for multiple testing across all genes, HEARTSVG applies Holm's method to adjust the p values and control the family-wise error rate (FWER). It is important to note that HEARTSVG may have certain conceptual limitations, particularly the requirement for a square tissue slice and the seemingly arbitrary definitions of the four time series, especially regarding the chosen window sizes and the construction of the time series, given the many possible alternatives.

SPADE[47] leverages the hematoxylin-and-eosin (H&E) image accompanying the SRT data (e.g., 10x Visium) to detect overall SVGs. It defines a gene as an overall SVG if the gene's expression levels can be predicted by H&E features, which are extracted by a pre-trained convolutional neural network and expected to contain critical spatial information, in a linear model. Specifically, SPADE first performs principal component analysis (PCA) on the 512 H&E features for dimensionality reduction. Then, it tests if the resulting principal components can predict the gene's expression levels in a linear model. If so, the gene is defined as an overall SVG.

BOOST-MI[48] and *BOOST-HMI*[49] are Bayesian methods that detect overall SVGs based on the Ising model and its extension to the geostatistical mark interaction model, respectively. The Ising model, historically used to model ferromagnetism, represents each magnetic moment (or spin) in a material as a discrete variable that can take one of two values: +1 or −1 (interpreted as spin up or spin down). In ferromagnetic materials, spins tend to align with their neighbors to minimize the system's energy. The Ising model exhibits a phase transition at a critical temperature, below which spins align, resulting in net magnetization corresponding to the ferromagnetic phase. Above this temperature, thermal fluctuations dominate, and spins are randomly oriented, corresponding to the paramagnetic phase. BOOST-MI applies the Ising model to SRT data by treating spatial spots as spins, with the positive and negative values corresponding to high and low expression levels of a gene after dichotomization. BOOST-MI detects a gene as an overall SVG if it finds the gene to be in a "ferromagnetic phase." BOOST-HMI extends this approach for imaging-based SRT data to accommodate the irregular spatial distribution of measured spots.

Graph-based methods

In manifold learning, when data originate from a non-linear manifold within Euclidean space, using a graph to represent the data effectively captures the intrinsic geometry and preserves the local structures of the manifold. Extending this concept to spatial spots on a non-linear manifold within the 2D Euclidean space of a tissue slice, graph-based SVG detection methods first construct a neighborhood graph by connecting nearby spots in Euclidean space. These methods then operate on the graph to detect overall SVGs. The way the graph is constructed is crucial, as it determines how well the graph represents the manifold, significantly impacting the performance of graph-based methods.

In this section, we introduce ten graph-based methods (Fig. 3), focusing on their graph construction approaches and subsequent operations on the graph for detecting overall SVGs.

Hotspot[50] aims to detect overall SVGs as the genes that exhibit high expression levels in local spot neighborhoods (i.e., "hotspots"). It first constructs a K-nearest neighbor (KNN) directed graph by treating each spot as a node and connecting it to its closest K spots in Euclidean distance. For a spot i connected to a spot j, with Euclidean distance denoted by d_{ij}, the directed edge from i to j is assigned a weight defined as $w_{ij} = e^{-d_{ij}^2/\sigma_i^2}$, where σ_i represents the bandwidth of spot i (defined as the distance from spot i to its $[K/3]$-th neighbor). Next, for each gene, denoting its expression count at spot i by Y_i, Hotspot uses an autocorrelation statistic H to quantify the dependence of the gene's expression level on the graph structure:

$$H = \sum_i \sum_{j \neq i} w_{ij} Y_i Y_j,$$

which is related to Moran's I[45], a spatial autocorrelation measure that quantifies the degree to which similar expression levels are clustered together in space. Moran's I is defined as:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2}, \quad (2.3)$$

where n is the number of spatial spots, and $\bar{Y}$ is the gene's mean expression level across spots. Similar to I, a large H indicates that spots where the gene is highly expressed are clustered in local neighborhoods, suggesting the gene is an overall SVG. To assess the statistical significance of H, Hotspot converts H to a z-statistic by subtracting the expectation of H under the null model, where all spots are assumed to be independent, and dividing by the standard deviation of H under the null model. Specifically, Hotspot assumes two null models where Y_i independently follows either an NB distribution or a Bernoulli

distribution. The z-statistic is then assumed to follow the standard Gaussian distribution under the null hypothesis, and a one-sided p value is computed.

HRG[51] differs from Hotspot in three key aspects. First, it replaces the KNN graph with a shared-nearest-neighbors (SNN) graph, where the edge weight w_{ij} between spots i and j is defined as the Jaccard index of their respective K-nearest neighbors. The Jaccard index is calculated as the size of the intersection of the two sets divided by the size of their union. Second, HRG modifies Y_i, the gene count at spot i, to a z-score $Z_i = \frac{Y_i - \bar{Y}}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(Y_j - \bar{Y})^2}}$, where $\bar{Y}$ is the mean gene count and n is the total number of spots. Third, unlike Hotspot, HRG does not provide statistical significance for its detected overall SVGs but instead ranks the genes based on H.

SVGbit[52] is also a spatial-autocorrelation-based method that ranks genes as overall SVGs based on the spatial autocorrelations of local neighborhoods. The method does not provide statistical significance for its rankings. For each gene, SVGbit uses three steps to calculate a summary statistic, called the "aggregation index," which is used to rank genes. First, SVGbit identifies "hotspots" as spatial spots with large and statistically significant local Moran's I values, using a z-test-based p value thresholded at the 5% FDR. Specifically, for spot i, local Moran's I is defined as:

$$I_i = (Y_i - \bar{Y})\sum_j w_{ij}(Y_j - \bar{Y}),$$

where the terms are defined similarly to Moran's I (2.3). A high I_i value indicates that spot i is part of a neighborhood of similar values (either high-high or low-low). Hence, spot i is treated as a hotspot if I_i is deemed significantly large. Second, for each hotspot, SVGbit considers its neighboring hotspots among the 6-nearest-neighbor spots and defines the "local aggregation density" as the sum of these neighboring hotspots' normalized log-transformed p values (i.e., the log-transformed p value of each neighboring hotspot divided by the sum of the log-transformed p values of the 6 neighboring spots). Thus, a hotspot with more neighboring hotspots has a higher local aggregation density. Third, the gene's aggregation index is computed as the average of all hotspots' local aggregation densities. Intuitively, the overall SVGs with large aggregation indices are those whose expression patterns exhibit clustered hotspots.

SINFONIA[53] aims to detect overall SVGs as those exhibiting positive spatial autocorrelations, where similar expression levels are clustered in neighboring spots, either globally or locally. It begins by constructing a KNN directed graph, setting the edge weight from spot i to spot j (if connected) as $w_{ij} = 1 - \frac{d_{ij}}{\max_j d_{ij}}$, where d_{ij} is the Euclidean distance between spots i and j, and $\max_j d_{ij}$ is the maximum distance between spot i and its K nearest neighbors. Given this graph, SINFONIA calculates two measures of spatial autocorrelation for each gene: (1) Moran's I[45], which provides a global measure of spatial autocorrelation, with values ranging from -1 to 1, where a value close to 1 indicates strong positive spatial autocorrelation; (2) Geary's C[54], which is more sensitive to local spatial differences, with values ranging from 0 to 2, where a value close to 0 indicates strong positive spatial autocorrelation. Geary's C is defined as:

$$C = \frac{(n-1)}{2\sum_i\sum_j w_{ij}}\frac{\sum_i\sum_j w_{ij}(Y_i - Y_j)^2}{\sum_i(Y_i - \bar{Y})^2},$$

where the terms are defined similarly to Moran's I (2.3). SINFONIA then ranks genes from high to low based on Moran's I and from low to high based on Geary's C. To identify overall SVGs, SINFONIA takes the union

of the top J genes from both rank lists, where J is a pre-specified positive integer. However, SINFONIA does not provide statistical significance for the detected overall SVGs.

MERINGUE[55] also uses Moran's I to detect overall SVGs but differs from previous methods in its graph construction approach. Instead of using a KNN graph, MERINGUE constructs a graph of spatial spots using Delaunay triangulation, making it suited for spatial spots with non-uniform density, such as cells in MERFISH data. Delaunay triangulation[56] connects points in a 2D space to form triangles such that no point lies inside the circumcircle (the circle that passes through all three vertices) of any triangle. In simpler terms, it maximizes the minimum angle of the triangles, avoiding skinny triangles. Unlike KNN, Delaunay triangulation does not require selecting an arbitrary K parameter and automatically adapts to variations in point density, providing more connections in denser regions and fewer in sparser regions. Given the graph, MERINGUE assigns an edge weight between spots i and j as $w_{ij} = 1$ if the two spots are connected, and $w_{ij} = 0$ otherwise. To determine if a gene is an overall SVG, MERINGUE uses Moran's I, transforms it to a z-statistic, and calculates a one-sided p value based on the standard Gaussian distribution.

SpaGene[57], similar to previous methods that use Moran's I, aims to detect overall SVGs as those highly expressed at neighboring spots. Starting from a KNN graph of spots, it uses an alternative approach to boost computational efficiency. For each gene, it dichotomizes spots, coding them as 0 or 1 corresponding to low or high expression levels of the gene. It then extracts the high-expression spots (i.e., the spots labeled as 1) from the graph and summarizes the distribution of these spots' degrees (the number of edges connected to each spot) in the subgraph. Finally, it compares the degree distribution of these high-expression spots to that of the whole graph, and calculates the earth mover's distance between the two distributions. A small distance indicates that the high-expression spots are densely connected to each other and rarely connected to other spots, so the gene is likely an overall SVG. Finally, SpaGene converts the distance to a p value using a permutation test.

SpaGFT[58] detects overall SVGs by leveraging a graph signal processing approach. Starting from a KNN graph of spots, SpaGFT applies the graph Fourier transform (GFT) to decompose each gene's expression levels on this graph into components of varying spatial frequencies, analogous to decomposing a signal into sine and cosine waves. While the components are common to all genes, each gene has unique coefficients associated with these components. Low-frequency components capture spatially organized patterns, while high-frequency components primarily represent noise. The classification of low- and high-frequency components is based on the eigenvalues derived from the GFT, with thresholds determined at two inflection points of the eigenvalue curve, identified using the Keedle algorithm[59]. For each gene, SpaGFT calculates a GFT score, which is a weighted sum of the gene's coefficients for all components, where lower-frequency components are assigned larger weights. This GFT score reflects the gene's overall concentration in low-frequency components. Genes with high GFT scores are considered to exhibit substantial spatial variation and are potential SVGs. To identify SVGs, SpaGFT employs two thresholds: a GFT score cutoff and a p value cutoff. The GFT score cutoff is determined by ranking genes in descending order of their GFT scores and identifying the inflection point of the GFT score curve across all genes. The p value calculated by SpaGFT for each gene is not derived from the GFT score; rather, it is based on the Wilcoxon rank-sum test comparing the coefficients of low-frequency and high-frequency components. However, since the ranking of SVGs does not depend on the p value and the detection threshold is not solely based on the p value, we do not consider SpaGFT to provide valid statistical inference in our review.

BinSpect, part of the Giotto toolbox[60], detects overall SVGs by examining whether the dichotomized spots (i.e., spots coded as 0 or 1

corresponding to low and high expression levels of a gene, respectively) are randomly distributed in a graph. Specifically, BinSpect constructs a graph of spots either as a KNN graph or by Delaunay triangulation. Then, for each gene, BinSpect follows three steps to determine if the gene is an overall SVG. First, it assigns a binary label to each spot by either thresholding at the top 30% expressed spots or using k-means clustering ($k = 2$). Second, it constructs a 2-by-2 contingency table by counting the edges that connect spots with labels (0,0), (0,1), (1,0), or (1,1). Finally, it applies Fisher's exact test to the contingency table to decide whether spots with the same label tend to be connected, thereby detecting a gene as an overall SVG.

Note that, unlike Giotto, two other popular toolkits with SRT functionalities—Seurat and squidpy—implement different methods for detecting overall SVGs. Seurat[38] uses two approaches: Moran's I and a mark-point-process-based method called "markvariogram" similar to Trendsceek[37]. On the other hand, squidpy[42] employs sepal[40], a diffusion-process-based method.

scGCO[61] is similar to BinSpect at a high level by testing whether a gene's discretized expression levels are dependent on spatial neighborhoods. However, scGCO approaches the problem in a more complex way. After constructing an undirected graph of spots using Delaunay triangulation, scGCO performs three steps. First, it fits a Gaussian mixture model to the gene's log-transformed expression levels to discretize the gene's expression into a number of levels. Second, using a graph cut algorithm, it fits a hidden Markov random field to the gene's log-transformed expression levels on the spot graph to divide the spots into segments. Finally, it tests whether each discretized level is randomly distributed in each segment, using the homogeneous Poisson process as the underlying null model. In summary, a gene is considered an overall SVG if at least one of its discretized levels is non-randomly distributed in any segment.

RayleighSelection[62] is a theoretical approach that generalizes a graph to a simplicial complex[63], incorporating additional higher-dimensional topological information about spatial spots compared to a graph. The method first constructs the Vietoris-Rips complex, a type of simplicial complex, of spatial spots by considering balls of a fixed radius around each spot and forming edges, triangles, and higher-dimensional simplices based on the intersections of these balls. For each gene, it then computes the combinatorial Laplacian score, which measures how well the gene respects the local structure of the simplicial complex. If a gene has high expression levels in highly connected regions, it will have a low combinatorial Laplacian score, indicating a significant spatial pattern and suggesting it to be an overall SVG. Finally, RayleighSelection computes a p value for the combinatorial Laplacian score using a permutation test.

Relationship between Euclidean-space-based methods and graph-based methods

When it is reasonable to assume data lie on a non-linear manifold in Euclidean space, using a graph to represent the data offers significant advantages. A graph can capture the intrinsic geometry of the manifold by representing data points as nodes and their similarity relationships as edges based on some similarity or distance measure. Graph-construction methods like KNN (where each node's neighborhood is defined by a fixed number of neighbors, K) or ϵ-neighborhoods (where each node's neighborhood is defined by a fixed radius, ϵ) ensure that each node is connected to its nearest neighbors, preserving the local relationships among nodes. The shortest-path distance between two nodes in the graph approximates their "geodesic distance" on the manifold, reflecting the data's intrinsic structure more accurately than the Euclidean distance. Additionally, graph-based methods are more robust to the curse of dimensionality: in high-dimensional spaces, where Euclidean distances often lose meaning, a graph can effectively capture relationships among data points.

However, in the context of SVG detection, using Euclidean distance to represent distances between spots might be more appropriate. Biological relationships between spots inherently follow Euclidean distances in a physical 2D Euclidean space in tissue. Moreover, since Euclidean distances are based on 2D spatial locations rather than high-dimensional gene expression data, the curse of dimensionality is not an issue.

It remains an open question how to construct a graph to accurately represent a non-linear manifold of spatial spots. The choice of graph-construction method, such as KNN or Delaunay triangulation, is crucial. KNN connects each spot to its K nearest neighbors, preserving the local structure of the data. In contrast, Delaunay triangulation handles the non-uniform spatial distribution of spots in SRT data, such as MERFISH or Slide-seq data. Delaunay triangulation connects spots to form triangles such that no point is inside the circumcircle of any triangle, resulting in a graph that more naturally respects the non-uniform distribution of spots and preserves local density variations. However, Delaunay triangulation is more computationally expensive and more sensitive to outliers than KNN. Future studies could explore whether the computational cost of using Delaunay triangulation for imaging-based SRT data is justified, and how this decision may vary depending on the tissue structure-whether the tissue is structured (such as in the brain) or unstructured (such as in tumors). Additionally, research could focus on improving the scalability and robustness of Delaunay triangulation in these contexts. Additionally, the definition of edge weights in the graph is another important consideration. Edges can be directional or undirectional, depending on whether the relationship between spots has a direction (e.g., one cell type influencing another cell type's gene expression). The edge weights can take multiple forms. For example, weights can be determined by a Gaussian kernel, where weights decay exponentially with the Euclidean distance between spots, emphasizing closer spots. Alternatively, weights can decrease linearly with distance. These choices affect how the graph captures the manifold structure of data and, consequently, how effectively it can be used to detect spatial patterns of gene expression that are biologically relevant.

Applications of overall SVGs

The detection of overall SVGs mainly serves as a pre-processing step in SRT data analysis. While the definition of overall SVGs may vary across methods, the common objective is to select informative genes for downstream analysis. A common downstream analysis is to identify *spatial domains* (also referred to as *spatial communities*) by partitioning a tissue slice into regions so that spots have similar expression profiles of overall SVGs in each region.

Identifying spatial domains can help uncover tissue layers where morphological architecture is less defined. For example, SPADE[47], an overall SVG detection method, uses its detected overall SVGs to identify spatial domains as substructures in cortical layers and amygdala in a 10x Visium mouse brain immunofluorescence dataset. Another use of spatial domain identification is to reveal gene expression profiles underlying tissue structures. For example, scGCO[61], another overall SVG detection method, identifies spatial domains from a mouse breast cancer biopsy dataset sequenced by Spatial Transcriptomics[64], and the identified domains align with annotated tissue structures including invasive ductal cancer, ductal cancer, and normal tissue. This alignment confirms that these tissue structures have distinct gene expression profiles, so the SRT data can potentially lead to the discovery of new marker genes for these tissue structures, a task tackled by spatial-domain-marker SVG detection.

A common approach to identifying spatial domains is clustering spatial spots using overall SVGs' expression levels and spot locations. For example, graph-based clustering (e.g., Louvain clustering) can be applied after spatial spots are connected into a graph based on their spatial proximity and gene expression levels; then, the identified

clusters contain spatial spots in proximity and exhibiting similar gene expression levels. For example, using Louvain clustering, the overall SVGs detected by SPADE[47] and SINFONIA[53] resulted in spatial domains in well-structured brain tissues in the mouse olfactory bulb Spatial Transcriptomics dataset[64], as well as in the mouse hippocampus seqFISH[5] and Slide-seq[65] datasets.

The spatial domains identified from overall SVGs can be validated, directly or indirectly, in three ways. First, spatial domains can be compared with tissue layers annotated by pathologists from the H&E image accompanying the SRT data (e.g., 10x Visium). For example, in the SpaGCN[66] method paper, three annotated layers of the primary pancreatic cancer were compared to the spatial domains identified from a Spatial Transcriptomics dataset[67]. Second, spatial domains can be annotated with cell-type labels in external transcriptomic data to verify if distinct domains have different cell-type compositions. For example, in a study of the human dorsolateral prefrontal cortex 10x Visium dataset[19], researchers annotated the identified spatial domains as cortical layers based on the cytoarchitecture and marker genes obtained from external large-scale single nucleus RNA-seq datasets. Third, for well-structured tissue types like the brain, spatial domains can be verified by transferring tissue layer labels from external public annotations. For instance, in the SINFONIA[53] method paper, annotations from Allen Brain Atlas[68] and Mouse Brain Gene Expression Atlas (http://mousebrain.org/) were transferred to several SRT datasets, including the mouse brain coronal section 10x Visium dataset, the mouse hippocampus Slide-seqV2 dataset[65], and the mouse olfactory bulb STEREO-seq dataset[17]. Although the detection of overall SVGs may serve as a feature screening step before spatial domain identification to circumvent the curse of dimensionality (i.e., dealing with too many genes as features), it is important to note that this screening step may not be necessary for spatial domain detection methods that employ alternative strategies to reduce feature dimensions. For example, BayesSpace[69] and StLearn[70] implement spatial domain identification using lower-dimensional representations of spatial spots based on selected HVGs.

Besides spatial domain identification, another downstream analysis is the unsupervised identification of spatial gene modules among the detected overall SVGs. Each module is a cluster of overall SVGs with similar spatial expression patterns, representing a common spatial pattern shared by these genes (also referred to as a co-expression pattern). Compared to the spatial expression patterns of individual overall SVGs, the patterns of gene modules are less noisy and can provide clearer insights into the molecular architecture underlying tissue structures—specifically, which genes co-determine a tissue structure in a synchronized manner. For example, in downstream applications of MULTILAYER[39], MERINGUE[55], and Binspect in Giotto[60], the detected overall SVGs were grouped into modules using hierarchical clustering based on a constructed gene co-expression matrix or a gene-gene similarity graph.

It is worth noting that the identification of spatial gene modules can precede and facilitate the discovery of non-global spatial domains that are specific to one or more gene modules. For example, among the overall SVGs detected by MULTILAYER[39], the genes *ACTA2* and *ELN* were found to be spatially co-expressed, leading to the identification of five tissue substructures in human heart tissue. Similarly, spatial gene modules derived from the overall SVGs detected by MERINGUE[55] enabled the detection of functional spatial regions in mouse brain tissue. Additionally, spatial gene modules identified from the overall SVGs detected by Binspect in Giotto[60] successfully recovered known anatomical structures in mouse kidney tissue.

Methods for detecting cell-type-specific SVGs
In both sequencing- and imaging-based SRT data, a gene's expression variance across spots can result from the gene's different expression levels in distinct cell types. Given that cell types are typically distributed non-uniformly within a tissue slice, neglecting the cell type information of spatial spots may hinder the discovery of genes whose expression patterns reflect informative spatial variation beyond the variation attributed to cell-type composition[71,72]. Thus, we define *cell-type-specific SVGs* as the genes that exhibit non-random spatial expression patterns within a cell type. These genes are detected using both SRT data and external cell-type annotations for the spatial spots (see Fig. 2a and Table 2). Unlike the methods for detecting overall SVGs that do not consider cell type annotations, the three methods for detecting cell-type-specific SVGs—*CTSV*[71], *C-SIDE*[73], and *spVC*[74]—begin by annotating the cell types of the spatial spots in SRT data and then use a regression framework to identify SVGs within cell types by examining the interaction effects between cell types and spatial locations (see section "Regression tests" for a detailed discussion on the general framework).

CTSV assumes a ZINB distribution for a given gene's expression count Y_i at spot i:

$$Y_i \overset{\text{ind}}{\sim} \text{ZINB}(\mu_i(\mathbf{s}_i), \boldsymbol{\theta}),$$

where $\mu_i(\mathbf{s}_i)$ is the mean expression function of the spatial location $\mathbf{s}_i = (s_{i1}, s_{i2})^\top$, and $\boldsymbol{\theta}$ indicates the other parameters (dispersion and zero-inflated probability) necessary to describe the distribution. To determine if the gene is a cell-type-specific SVG, CTSV includes K cell types' cell-type-level mean functions of $\mathbf{s}_i$: $\eta_k(\mathbf{s}_i)$, $k = 1, ..., K$, and assumes that:

$$\log \mu_i(\mathbf{s}_i) = \log \ell_i + \sum_{k=1}^{K} \eta_k(\mathbf{s}_i) w_{ik},$$
$$\eta_k(\mathbf{s}_i) = \beta_{k0} + \beta_{k1} b_1(s_{i1}) + \beta_{k2} b_2(s_{i2}),$$

where ℓ_i denotes spot i's library size (i.e., total expression count), and the weights w_{ik}, $k = 1, ..., K$, are pre-estimated cell type proportions of spot i obtained by spatial deconvolution methods such as SPOTlight[75] and RCTD[72], and $b_1(s_{i1})$ and $b_2(s_{i2})$ are two non-cell-type-specific, non-parametric, univariate functions of the spatial coordinates in the two dimensions. CTSV then tests if β_{k1} and β_{k2} are both zero. If not, the gene is considered an SVG specific to cell type k.

C-SIDE is similar to CTSV but has three major differences. First, C-SIDE assumes a Poisson distribution for Y_i. Second, C-SIDE assumes a different relationship between $\mu_i(\mathbf{s}_i)$ and $\eta_k(\mathbf{s}_i)$, $k = 1, ..., K$. Third, C-SIDE uses a different non-parametric function form for each $\eta_k(\mathbf{s}_i)$:

$$Y_i \overset{\text{ind}}{\sim} \text{Poisson}(\mu_i(\mathbf{s}_i)),$$
$$\log(\mu_i(\mathbf{s}_i)) = \gamma_0 + \log \ell_i + \log\left(\sum_{k=1}^{K} \eta_k(\mathbf{s}_i) w_{ik}\right) + \epsilon_i,$$
$$\log(\eta_k(\mathbf{s}_i)) = \beta_{k0} + \sum_{\ell=1}^{L} \beta_{k\ell} b_\ell(\mathbf{s}_i),$$

where the common terms are defined similarly to CTSV, ϵ_i is a zero-mean random-effect term, and $\sum_{\ell=1}^{L} \beta_{k\ell} b_\ell(\mathbf{s}_i)$ is a cell-type-specific, bivariate smooth-spline function of $\mathbf{s}_i$ consisting of L basis functions. C-SIDE then tests if $\beta_{k1}, ..., \beta_{kL}$ are all zero. If not, the gene is considered an SVG specific to cell type k.

spVC is similar to C-SIDE but assumes a different relationship between $\mu_i(\mathbf{s}_i)$ and $\eta_k(\mathbf{s}_i)$, $k = 1, ..., K$:

$$\log(\mu_i(\mathbf{s}_i)) = \gamma_0 + \log \ell_i + \sum_{k=1}^{K} w_{ik} \beta_k + \sum_{k=1}^{K} \eta_k(\mathbf{s}_i) w_{ik} + \gamma(\mathbf{s}_i),$$

which considers cell types' non-spatial effects as $\sum_{k=1}^{K} w_{ik} \beta_k$ and a baseline spatial effect $\gamma(\mathbf{s}_i)$. Then, spVC implements a two-step testing procedure. First, it tests whether $\beta_1, ..., \beta_K$ are all zero and whether $\gamma(\cdot) = 0$ in a reduced model without the cell-type-specific spatial effect

term $\sum_{k=1}^{K} \eta_k(\mathbf{s}_i) w_{ik}$. If both null hypotheses are rejected, the second step is to test whether $\eta_k(\cdot) = 0$. If not, the gene is considered a SVG specific to cell type k.

Applications of cell-type-specific SVGs

Although the publications of the three cell-type-specific SVG detection methods[71,73,74] did not specifically outline applications for cell-type-specific SVGs, we envision potential applications that extend the two major applications of overall SVGs—identifying spatial domains and spatial gene modules—to a more granular, cell-type level. First, cell-type-specific SVGs can be used to identify and characterize distinct cell subpopulations or states within particular cell types. Second, cell-type-specific SVGs can enable the identification of gene modules within specific cell types. These applications have the potential to provide new molecular insights into tissues with complex structures and diverse cell types, such as cancer and brain tissue. For example, in cancer research, cell-type-specific SVGs can help identify distinct states of cancer cells and immune cells, leading to a better understanding of tumor heterogeneity and the tumor microenvironment. Additionally, cell-type-specific SVGs can facilitate the identification of cell-type-specific gene modules that could be crucial for developing targeted therapies and personalized treatment strategies. Similarly, in brain tissue, cell-type-specific SVGs can identify diverse neuronal subpopulations that may correspond to functional brain regions. Furthermore, cell-type-specific SVGs can enable the discovery of cell-type-specific gene modules that may regulate neuronal activity and connectivity, potentially advancing our understanding of brain development and neurodegenerative diseases.

Methods for detecting spatial-domain-marker SVGs

We define *spatial-domain-marker SVGs* as the genes that exhibit significantly higher expression in a spatial domain compared to other domains. These genes are detected using SRT data and spatial domains, which are usually detected from the same SRT data (see Fig. 2a and Table 2). Unlike previous SVG detection methods, the two methods for detecting spatial-domain-marker SVGs—*SpaGCN*[66] and *DESpace*[22]—first partition spatial spots into more than one spatial domain. Then, they implement hypothesis tests to assess a gene's mean expression differences between these spatial domains. A gene is defined as a marker SVG of a spatial domain if it shows significantly higher expression in that domain than in other domains.

SpaGCN identifies spatial domains using a pre-trained graph convolutional network applied to SRT data and the paired H&E image. For each gene, it performs the Wilcoxon rank-sum test on normalized expression levels between each domain and the neighboring spots. If the gene is found to have significantly higher expression in a domain, it is considered a marker SVG of that domain.

DESpace first applies existing spatial clustering methods, such as BayesSpace[69] and StLearn[70], to SRT data to identify spatial clusters as spatial domains. DESpace then offers two modes for detecting SVGs: the first mode identifies spatial-domain-marker SVGs, while the second mode detects overall SVGs. In the first mode, DESpace uses an NB generalized linear model to assess if the spatial domains have a significant effect on a gene's expression. If a significant effect is found, the gene is identified as a spatial-domain-marker SVG, associated with the domain where it has significantly higher or lower expression compared to other domains. In the second mode, DESpace tests the null hypothesis that all spatial domains have the same effect on a gene's expression. If this hypothesis is rejected, indicating that the gene exhibits differences among spatial domains, the gene is detected as an overall SVG. Therefore, the spatial-domain-marker SVGs detected by DESpace must be included among the overall SVGs it identifies. Conversely, the overall SVGs detected by DESpace must serve as marker genes for some spatial domains.

Applications of spatial-domain-marker SVGs

Spatial-domain-marker SVGs have two applications for characterizing predefined spatial domains in a tissue slice, particularly when these domains align with tissue structures. First, spatial-domain-marker SVGs can help elucidate the underlying molecular mechanisms of tissue structures. For example, SpaGCN[66] detected spatial-domain-marker SVGs *KRT17* and *MMP11* in a pancreatic cancer region, which align well with pancreatic cancer biology. *KRT17* functions as a tumor promoter and regulates proliferation in pancreatic cancer, while *MMP11* is a prognostic biomarker for the disease. Second, spatial-domain-marker SVGs can assist in annotating tissue structures in other SRT datasets. For instance, the spatial-domain-marker SVGs identified by SpaGCN[66] from a human dorsolateral prefrontal cortex slice (No. 151673 in the LIBD (Lieber Institute for Brain Development) SRT dataset) were used to annotate the tissue layers in a different brain slice (No. 151507).

Theoretical characterization of SVG detection methods that use frequentist hypothesis tests

Among the 34 SVG detection methods, 23 of them implement statistical hypothesis tests to detect SVGs using *frequentist inference* (i.e., defining a test statistic, deriving the test statistic's null distribution, and converting the test statistic value to a p value). For these frequentist hypothesis-testing-based methods, their null and alternative hypotheses direct their SVG detection. In general, methods that use different null hypotheses are not directly comparable. The reason is that a gene may satisfy one null hypothesis (and is considered a true non-SVG) but not the other null hypothesis (and is defined as a true SVG). Hence, we would like to clarify and categorize the null hypotheses used in the 23 methods to deepen our understanding of these methods' conceptual similarities and differences.

Based on the types of null hypotheses, in Table 1 we summarize the hypothesis tests used in the 23 methods into three types: dependence tests, regression fixed-effect tests, and regression random-effect tests (also known as variance component tests). So far, dependence tests and regression random-effect tests have only been used for overall SVG detection, while regression fixed-effect tests have been used to detect all three categories of SVGs. For each method, besides the test type, we also list the test statistic and null distribution in Supplementary Fig. 2. Moreover, we define the three test types in Table 2 and include a conceptual diagram in Fig. 4 that illustrates the relationships between the three test types and the three SVG categories.

Among the three types of hypothesis tests for SVG detection (Fig. 4), dependence tests have the most general null hypothesis: a gene's expression level is independent of the spatial location. In contrast, the two types of regression tests rely on specific assumptions for a regression model, which has a gene's expression level as the response variable and the spatial location as the predictor (also known as the covariate or explanatory variable). The co-existence of dependence tests and model-specific regression tests is common in the statistics literature[76]. In general, dependence tests, thanks to their general independence null hypothesis, can capture SVGs with more diverse patterns but can be less powerful for detecting SVGs of specific patterns, compared to the regression tests. In contrast, relying on specific model assumptions, regression tests are more powerful for discovering the SVGs satisfying the assumptions but, meanwhile, more prone to false discoveries when the model assumptions do not hold[77]. Moreover, an advantage of regression tests is that they can more easily incorporate other covariates for adjustment compared to dependence tests. Additionally, their test statistics typically have theoretical null distributions, making p value calculation straightforward and efficient. It remains an open question to benchmark the three types of tests regarding robustness to model misspecification and the trade-off between robustness and power. Additionally, it is important to identify
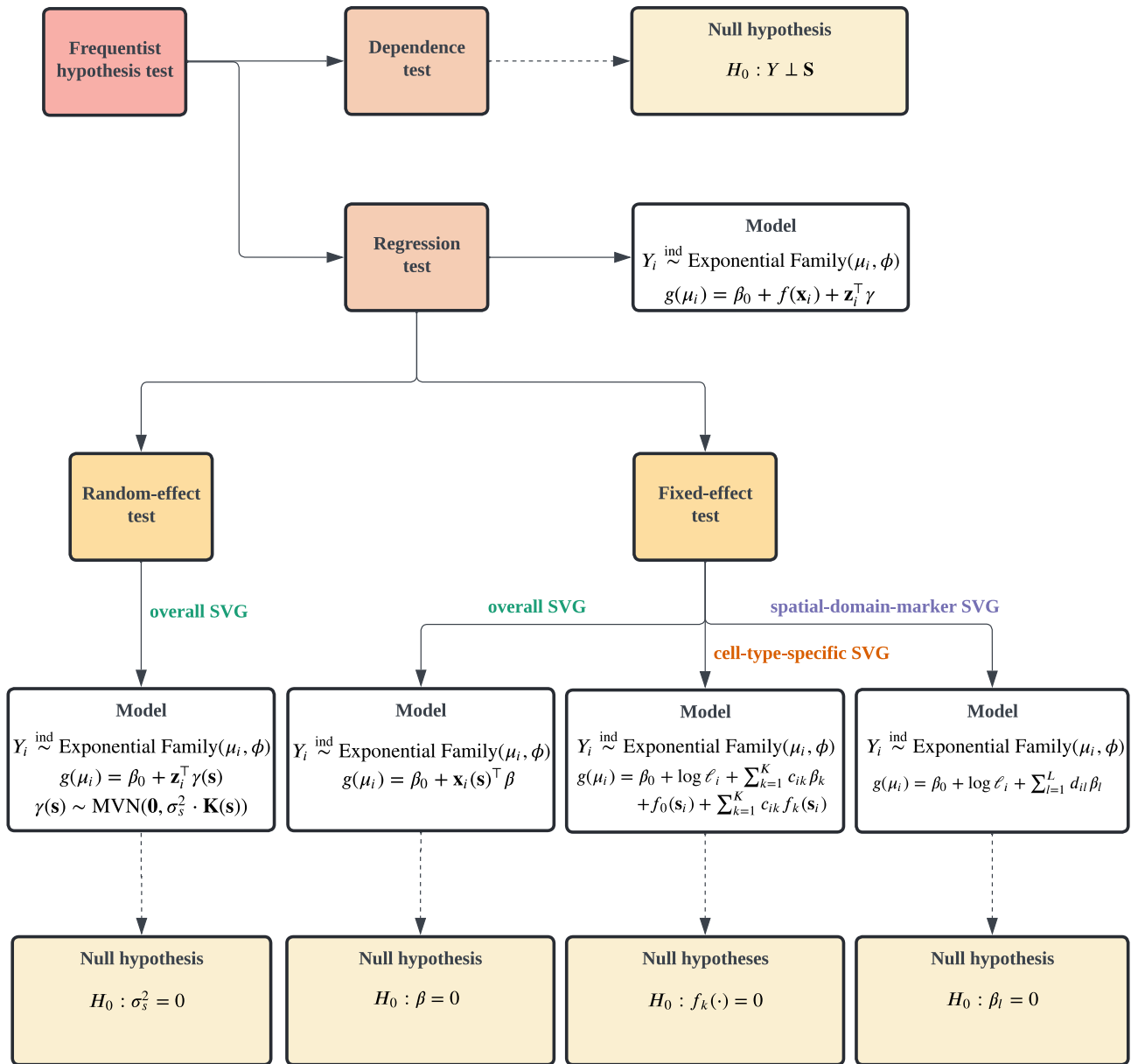
Fig. 4 | Conceptual diagram for section "Theoretical characterization of SVG detection methods that use frequentist hypothesis tests". This diagram illustrates the logical relationships among the three types of statistical tests (dependence test, regression fixed-effect test, and regression random-effect test) used by the 23 SVG detection methods that rely on frequentist hypothesis tests. The diagram also introduces the general form of statistical models upon which regression-based tests are performed and the corresponding null hypotheses for detecting SVGs.

which assumptions are more reasonable for SRT data generated from various tissues and by different technologies.

Dependence tests

For a given gene, the most general hypothesis test for SVG detection is to decide whether the gene's expression level Y is independent of the spatial location $\mathbf{S}$, i.e., the null hypothesis is:

$$H_0 : Y \perp \mathbf{S}.$$

In this formulation (Fig. 4), we assume that $(y_1, \mathbf{s}_1), \ldots, (y_n, \mathbf{s}_n)$ are independently sampled from the joint distribution of $(Y, \mathbf{S})$, where $(y_i, \mathbf{s}_i)$ indicates spot i's expression level of the given gene and spatial location. When the independent null hypothesis H_0 does not hold, the gene is a *true overall SVG*.

Out of the 23 frequentist hypothesis-testing-based SVG detection methods, 11 methods adopt the dependent test formulation,

constructing a test statistic to summarize the dependence between a gene's expression level and the spatial location. The 11 methods are SPARK-X, SMASH, singlecellHaystack, Trendsceek, HEARTSVG, Hotspot, MERINGUE, SpaGene, BinSpect, scGCO, and RayleighSelection. We categorize the test statistics used in the 11 methods into two types: test statistics with theoretical null distributions (SPARK-X, SMASH, HEARTSVG, Hotspot, MERINGUE, BinSpect, and scGCO) and test statistics without theoretical null distributions (singlecellHaystack, Trendsceek, SpaGene, and RayleighSelection).

Among the seven methods that use test statistics with theoretical null distributions, SPARK-X is a representative method[34]. The SPARK-X test statistic is defined to capture the "agreement" between two spot similarity matrices (referred to as "covariance matrices" by the SPARK-X authors), one defined based on a gene's expression levels at the n spots and the other based on the spatial locations of the n spots (subject to a transformation of the locations before defining the similarity between spots, so that the similarity matrix reflects a specific spatial pattern such

as the Gaussian or cosine kernel). Specifically, the test statistic is defined as $(1/n)$ times the trace of the product of the two similarity matrices, which is essentially the Pearson correlation between the two vectorized matrices. The rationale is that under the independence null hypothesis, the test statistic should be small and follow a theoretical, asymptotic null distribution as a mixture chi-square distribution.

Four methods use test statistics that lack theoretical null distributions and therefore employ a permutation strategy. They generate the null distribution of a test statistic by permuting a gene's expression levels across the n spots, thus removing any dependence between the gene's expression level and the spatial location. The permutation procedure is a general solution for obtaining the null distribution when the theoretical null distribution is difficult to derive, but it is computationally intensive due to the need for repetitive permutations to generate the null distribution. For example, singlecellHaystack[36] defines a test statistic based on the KL divergence of the conditional density of spots' spatial locations, which is conditional on a gene's dichotomized expression level (on or off), from the unconditional density of spots' spatial locations. Specifically, both densities are defined for each spatial grid, which consists of many spots and is predefined by the singlecellHaystack algorithm. Then, the test statistic is a summation of the KL-divergence-like statistics across the spatial grids. As this test statistic is complex and has no theoretical null distribution, the permutation is used to generate the null distribution.

Regression tests

There are two types of regression tests (Fig. 4): fixed-effect tests, where the effect of the spatial location is assumed to be fixed, and random-effect tests, which assume the effect of the spatial location as random. To explain these two types of tests, we start with a *linear mixed model* for a given gene:

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \boldsymbol{\gamma} + \epsilon_i, \tag{5.4}$$

where the response variable Y_i is the gene's expression level at spot i, $\mathbf{x}_i \in \mathbb{R}^p$ indicates the fixed-effect covariates of spot i, $\mathbf{z}_i \in \mathbb{R}^q$ denotes the random-effect covariates of spot i, and ϵ_i is the random measurement error at spot i with $\mathbb{E}[\epsilon_i] = 0$. In the model parameters, β_0 is the (fixed) intercept, $\boldsymbol{\beta} \in \mathbb{R}^p$ indicates the fixed effects, and $\boldsymbol{\gamma} \in \mathbb{R}^q$ denotes the random effects with zero means $\mathbb{E}[\boldsymbol{\gamma}] = 0$ and the covariance matrix $\mathrm{Var}(\boldsymbol{\gamma}) = \boldsymbol{\Sigma} \in \mathbb{R}^{q \times q}$. In this linear mixed model, independence is assumed between $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ and among $\epsilon_1, \ldots, \epsilon_n$.

Fixed-effect tests examine whether some or all of the fixed-effect covariates $\mathbf{x}_i$ contribute to the mean of the response variable $\mathbb{E}[Y_i]$. If all fixed-effect covariates make no contribution, then $\mathbb{E}[Y_i|\mathbf{x}_i] = \mathbb{E}[Y_i]$. The null hypothesis:

$$H_0 : \boldsymbol{\beta} = \mathbf{0},$$

implies $\mathbb{E}[Y_i|\mathbf{x}_i] = \mathbb{E}[Y_i]$, $i = 1, \ldots, n$.

Random-effect tests examine whether the random-effect covariates contribute to the variance of the response variable $\mathrm{Var}(Y_i)$, focusing on the decomposition:

$$\mathrm{Var}(Y_i) = \mathrm{Var}(\mathbb{E}[Y_i|\mathbf{z}_i]) + \mathbb{E}[\mathrm{Var}(Y_i|\mathbf{z}_i)] = \mathbf{z}_i^\top \boldsymbol{\Sigma} \mathbf{z}_i + \mathrm{Var}(\epsilon_i) \tag{5.5}$$

and testing if the contribution of the random-effect covariates $\mathrm{Var}(\mathbb{E}[Y_i|\mathbf{z}_i])$ is zero. The null hypothesis:

$$H_0 : \boldsymbol{\Sigma} = \mathbf{0}$$

implies $\mathrm{Var}(\mathbb{E}[Y_i|\mathbf{z}_i]) = 0$, $i = 1, \ldots, n$.

To generalize the linear mixed model (5.4) for data where the response variable Y_i is not continuous (e.g., binary or count), we can decompose the model into a *random structure*, where each Y_i independently follows a distribution with mean $\mu_i = \mathbb{E}[Y_i]$, and a *systematic structure* that links μ_i to the fixed-effect covariates $\mathbf{x}_i$. Specifically, for the model (5.4), if we assume that $\boldsymbol{\gamma} \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\epsilon_i \sim N(0, \sigma^2)$, then the *random structure* is:

$$Y_i \overset{\text{ind}}{\sim} N(\mu_i, \mathbf{z}_i^\top \boldsymbol{\Sigma} \mathbf{z}_i + \sigma^2),$$

and the *systematic structure* is:

$$\mu_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}.$$

The generalization of the linear mixed model can occur in both the random and systematic structures. For simplicity, we only focus on the fixed-effect covariates $\mathbf{x}_i$ for the generalization and omit the random-effect covariates $\mathbf{z}_i$. If we change the random structure from the Gaussian distribution to another exponential-family distribution, we have:

$$Y_i \overset{\text{ind}}{\sim} \text{Exponential Family}(\mu_i, \phi), \tag{5.6}$$

where ϕ is a nuisance parameter not of primary interest. An example of exponential-family distribution is the Poisson distribution $Y_i \sim \mathrm{Poisson}(\mu_i)$. According to the random structure change, the systematic structure can be written in general as:

$$g(\mu_i) = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{5.7}$$

where the function $g(\cdot)$ is referred to as the *link function* and specified based on the distribution in the random structure. We refer to the model specified by (5.6)–(5.7) as a *generalized linear model*. If we further generalize the systematic structure so that the effects of the p covariates in $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ on μ_i is non-linear and additive:

$$g(\mu_i) = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}), \tag{5.8}$$

we referred to the model specified by (5.6) and (5.8) as a *generalized additive model*.

For regression models with different complexities (roughly speaking, different numbers of parameters to estimate), model selection centers on the bias-variance trade-off. An oversimplified model misses key data characteristics, leading to biased parameter estimates, though with low variance. In contrast, an overly complex model wastes parameters on unimportant details, increasing variance in parameter estimates. As a trade-off, a reasonable model appropriately addresses the data characteristics without overfitting the data noise, thereby being more powerful and robust.

For SVG detection, the effect of the spatial location $\mathbf{s}_i$ on the given gene's expression level Y_i can be formulated as either fixed or random. In the following two subsections, we will discuss the SVG detection methods that adopt each formulation.

Regression fixed-effect tests. Among the 23 frequentist hypothesis-testing-based SVG detection methods, six methods use the regression fixed-effect test formulation: SPADE, C-SIDE, CTSV, spVC, SpaGCN, and DESpace. Notably, unlike dependence tests and regression random-effect tests, only regression fixed-effect tests cover all three SVG categories: overall SVGs, cell-type-specific SVGs, and spatial-domain-marker SVGs. Methods within each category use the same fixed-effect covariates $\mathbf{x}_i$ for spot i, while methods of different categories use different $\mathbf{x}_i'$ s. Below, we introduce the six methods in the three categories.

Among the six methods, SPADE[47] is the only method that detects overall SVGs. For a given gene, it uses a linear model without random-effect covariates:

$$\mu_i = \beta_0 + \mathbf{x}_i(\mathbf{s})^\top \boldsymbol{\beta},$$

where the fixed-effect covariates $\mathbf{x}_i(\mathbf{s})$ are defined as some processed features from the n spots' spatial locations $\mathbf{s}$ and an H&E image. Specifically, the processed features are based on 512 features from a pre-trained convolutional neural network applied to an H&E image. The gene is a *true overall SVG* if $\boldsymbol{\beta} \neq \mathbf{0}$. SPADE uses the R package limma[78] to fit the linear model and tests if each component of $\boldsymbol{\beta}$ is zero using a t-test.

For cell-type-specific SVG detection, the fixed-effect covariates of spot i (i.e., $\mathbf{x}_i$) include the library size ℓ_i (i.e., spot i's total expression count), the cell type $\mathbf{c}_i = (c_{i1}, \ldots, c_{iK})^\top \in [0,1]^K$, and the spatial location $\mathbf{s}_i \in \mathbb{R}^2$ (see section "Introduction" for definitions). The considered fixed effects include the marginal effects of $\mathbf{c}_i$ and $\mathbf{s}_i$ and their interactive effects. Hence, in a generalized additive model formulation for a given gene, the systematic structure is:

$$g(\mu_i) = \beta_0 + \log \ell_i + \sum_{k=1}^{K} c_{ik}\beta_k + f_0(\mathbf{s}_i) + \sum_{k=1}^{K} c_{ik}f_k(\mathbf{s}_i), \qquad (5.9)$$

where β_0 is the overall intercept, $\log \ell_i$ is the intercept effect of spot i's library size, β_k indicates the cell-type-specific spatial-invariant effect for cell type k, $f_0(\mathbf{s}_i)$ is the overall spatial effect at spot i, and $f_k(\mathbf{s}_i)$ is the cell-type-specific spatial effect for cell type k at spot i. For identifiability, constraints are needed for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^\top$ and $f_0(\,\cdot\,)$, $f_1(\,\cdot\,), \ldots, f_K(\,\cdot\,)$. The gene is a *true cell-type-specific SVG* of cell type k if $f_k(\,\cdot\,) \neq 0$.

C-SIDE, CTSV, and spCV are three methods for cell-type-specific SVG detection. As an example method, spVC[74] assumes a gene's expression count Y_i follows a Poisson distribution (i.e., the random structure (5.6)) with the systematic structure (5.9). To decide whether the gene is a cell-type-specific SVG, spVC implements a two-step procedure for sequential hypothesis tests. First, it considers a reduced model without interactive effects between $\mathbf{c}_i$ and $\mathbf{s}_i$, so the systematic structure becomes:

$$g(\mu_i) = \beta_0 + \log \ell_i + \sum_{k=1}^{K} c_{ik}\beta_k + f_0(\mathbf{s}_i).$$

Then it tests two null hypotheses: $H_0 : \boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^\top = \mathbf{0}$ and $H_0: f_0(\,\cdot\,) = 0$ using the likelihood ratio test and the Wald test, respectively. If both null hypotheses are rejected, it proceeds to the second step. Second, it considers the full model with interactive effects, with the systematic structure (5.9). It tests if any of the interactive effects $f_1(\,\cdot\,), \ldots, f_K(\,\cdot\,)$ are 0 using the Wald test. Similar to spVC, CTSV[71] assumes a generalized additive model. However, CTSV has a different random structure and testing procedure. Specifically, CTSV assumes the random structure to be that Y_i follows a ZINB distribution, and the systematic structure is written as:

$$g(\mu_i) = \beta_0 + \log \ell_i + \sum_{k=1}^{K} c_{ik}\beta_k + \sum_{k=1}^{K} c_{ik}f_k(\mathbf{s}_i), \qquad (5.10)$$

where $f_k(\mathbf{s}_i)$ is further assumed to be additive $f_k(\mathbf{s}_i) = f_{k1}(s_{i1}) + f_{k2}(s_{i2})$. Unlike spVC's two-step testing procedure, CTSV directly tests if any of the interactive effects $f_1(\,\cdot\,), \ldots, f_K(\,\cdot\,)$ are 0 using the Wald test. Same as spVC, C-SIDE[73] assumes the random structure to be Poisson. However, it has a different systematic structure from those of spVC and CTSV, allowing for random effects and additional non-linear transformations. C-SIDE uses a two-sided z-test.

For spatial-domain-marker SVG detection, the fixed-effect covariates of spot i (i.e., $\mathbf{x}_i$) include the library size ℓ_i and the spatial domain, indicated by $\mathbf{d}_i = (d_{i1}, \ldots, d_{iL}) \in \{0,1\}^L$, with $\sum_{l=1}^{L} d_{il} = 1$ (that is, $d_{il} = 1$ if spot i belongs to domain l). Then the systematic structure

becomes:

$$g(\mu_i) = \beta_0 + \log \ell_i + \sum_{l=1}^{L} d_{il}\beta_l, \qquad (5.11)$$

where β_l indicates the effect of spatial domain l on $g(\mu_i)$, i.e., the transformed gene i's true expression level. For identifiability, constraints are needed for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)^\top$, e.g., $\beta_1 = 0$. The gene is a *true spatial-domain-marker SVG* (i.e., a DEG across spatial domains) if $\boldsymbol{\beta} \neq \mathbf{0}$.

SpaGCN and DESpace are two methods for spatial-domain-marker SVG detection. Following the systematic structure (5.11), DESpace[22] assumes the gene expression Y_i follows an NB distribution and tests the null hypothesis $H_0: \boldsymbol{\beta} = \mathbf{0}$ using the likelihood ratio test statistic, which follows an asymptotic chi-square distribution under the null hypothesis. For SpaGCN[66], although it does not explicitly adopt a regression framework but instead uses the Wilcoxon rank-sum test on normalized gene expression levels (with the library size effect removed) to decide if a gene is significantly more highly expressed in one spatial domain compared to its neighboring spots, the Wilcoxon rank-sum is a score test statistic of the proportional odds ordinal logistic regression model. Hence, the test in SpaGCN can broadly be considered to follow the systematic structure (5.11).

Regression random-effect tests. Unlike the 11 methods that adopt dependence tests and the six methods that use regression fixed-effect tests, the last six of the 23 frequentist hypothesis-testing-based methods use regression random-effect tests: SpatialDE, nnSVG, SOMDE, SVCA, SPARK, and GPcounts.

All the six methods generalize the linear mixed model (5.4) for each gene, by encoding the spatial location $\mathbf{s}_i$ of spot i as the random-effect covariate $\mathbf{z}_i$. With n spots, $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})^\top \in \{0,1\}^n$ is a binary indicator vector with only one 1 indicating spot i. Without loss of generality, we assume that $z_{ii} = 1$. Then, the corresponding random-effect vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^\top \in \mathbb{R}^n$ has γ_i indicating the random effect of $\mathbf{s}_i$. The covariance matrix of $\gamma_1, \ldots, \gamma_n$, $\mathrm{Var}(\boldsymbol{\gamma})$, is assumed to depend on the spatial proximity of $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Accordingly, $\boldsymbol{\gamma}$ can be explicitly written as $\boldsymbol{\gamma}(\mathbf{s})$. When the null hypothesis $H_0: \mathrm{Var}(\boldsymbol{\gamma}(\mathbf{s})) = \mathbf{0}$ does not hold, the gene is a *true overall SVG*.

As an early and exemplar method, for each gene SpatialDE[26] assumes a linear random-effect model that only contains $\mathbf{z}_i$ as the random-effect covariates:

$$Y_i = \beta_0 + \mathbf{z}_i^\top \boldsymbol{\gamma}(\mathbf{s}) + \epsilon_i, \qquad (5.12)$$

where the random errors $\epsilon_1, \ldots, \epsilon_n$ independently follow a Gaussian distribution $N(0, \delta)$, and the random effects $\boldsymbol{\gamma}(\mathbf{s})$ jointly follow a multivariate Gaussian distribution $\mathrm{MVN}(\mathbf{0}, \sigma_s^2 \cdot \mathbf{K}(\mathbf{s}))$, with $\mathbf{K}(\mathbf{s}) = [K(\mathbf{s}_i, \mathbf{s}_j)]_{n \times n}$ specified by a kernel function $K(\,\cdot\,, \,\cdot\,)$ applied to the n spatial locations. This model is essentially a Gaussian process. The null hypothesis then becomes $H_0: \sigma_s^2 = 0$.

Two methods, nnSVG[27] and SOMDE[29], also use the Gaussian process model as in SpatialDE but adopt more efficient computational algorithms. The other three methods extend the Gaussian process model. Specifically, SVCA[30] considers two additional variance terms to accommodate other covariates, such as cell-cell interactions. Instead of assuming that Y_i follows a Gaussian distribution, SPARK[31] and GPcounts[32] assume Poisson and NB distributions, respectively, and generalize the linear mixed effect model (5.12) to generalized linear mixed effect models.

All six methods use the likelihood ratio test with an asymptotic chi-square distribution under the null hypothesis.

Discussion and future directions
Below, we discuss the comparative advantages and trade-offs of existing SVG detection methods in terms of detection power,

specificity, and scalability. Further, we outline two future directions for improving SVG detection methods: (1) accommodating differences among various SRT technologies and tissue types, as well as supporting multi-sample SRT data and (2) enhancing statistical rigor and method validation. Last, we summarize the limitations of existing benchmark studies and provide guidance for future benchmark studies.

Power, specificity, and scalability of SVG detection methods

Among the methods for detecting overall SVGs, the Euclidean-space-based, kernel-based methods define more specific spatial patterns for overall SVGs than the Euclidean-space-based, kernel-free methods and the graph-based methods. The statistical power of these kernel-based methods depends on the alignment between the pre-defined kernels' spatial patterns and the biologically relevant spatial patterns in SRT data. This alignment is often questionable because biologically relevant patterns are rarely as regular as the pre-defined kernel patterns. As a result, kernel-based methods may lose statistical power when this alignment is poor. To address this, some kernel-based methods, such as SPARK[31], incorporate multiple pre-defined kernels to capture a broader diversity of spatial patterns. Despite the potential loss of overall power, kernel-based methods are more specific and powerful for discovering overall SVGs whose spatial expression patterns align with the kernels compared to methods that do not use kernels. In short, there is a trade-off between overall power and specificity: methods targeting specific spatial patterns are less powerful at discovering other patterns, leading to a loss of overall power, but more powerful at discovering the targeted patterns with higher specificity.

Four types of genes that exhibit interesting but non-global expression patterns can be easily missed by methods that detect overall SVGs. First, some genes of interest may only be highly expressed in small regions of interest (ROIs) and can be overlooked by methods that do not distinguish small ROIs. For such genes, methods that detect spatial-domain-marker SVGs, such as SpaGCN[66], might be more appropriate because these methods can identify small ROIs as spatial domains in the first step before identifying marker genes in the second step. Second, there are genes that exhibit spatial expression patterns within spatial domains. These genes, referred to as *spatial-domain-specific SVGs*, may be a subset of spatial-domain-marker SVGs because, although marker SVGs are highly expressed within a spatial domain, they may not exhibit informative spatial variation within that domain. While no existing methods are specifically designed to detect these spatial-domain-specific SVGs, methods for detecting overall SVGs can likely be applied to specific spatial domains to discover these genes. Identifying spatial-domain-specific SVGs may help uncover spatial subdomains, capturing the internal variation within spatial domains. An alternative approach is to perform clustering within a spatial domain and then identify the resulting subdomain marker genes. However, these genes are not conceptually equivalent to those identified by an overall SVG detection method within the spatial domain, reflecting the same conceptual difference as that between overall SVGs and spatial-domain-marker SVGs. Third, cell-type-specific SVGs might also be easily missed by methods that detect overall SVGs if the cell types of interest have small proportions. To address this, methods for detecting cell-type-specific SVGs have been developed, including CTSV[71], C-SIDE[73], and spVC[74]. These methods rely on regression models, whose goodness-of-fit to SRT data remains to be explored. Ensuring a good fit is essential to avoid spurious discoveries. Fourth, some genes may exhibit sharp expression changes at tissue layer boundaries, which are too local to be detected by methods looking for overall SVGs. Belayer[79] aims to detect such genes by examining gene expression change rates (gradients) in the 2D space. As a future direction, adding the accompanying H&E image can help refine tissue boundaries, ensuring that the identified genes are interpretable. Moreover, if users have prior knowledge on some interesting

genes that should be detected as SVGs, how to incorporate this knowledge into SVG detection to improve the specificity remains an open question.

The scalability of an SVG detection method is determined by the computational time of two steps: calculating a summary statistic for each gene and converting the summary statistic to a p value based on a null distribution. The second step is only included in methods that provide statistical significance for the detected SVGs in a frequentist manner. In the first step of summary statistic calculation, computational time is measured in terms of n, the number of spatial spots. For example, fitting a Gaussian process takes $O(n^3)$ time in SpatialDE[26] and SPARK[31], but this time is reduced to $O(n)$ by using the nearest-neighbor Gaussian process approximation in nnSVG[27]. By changing the modeling framework from a Gaussian process to a Pearson correlation between two similarity matrices, SPARK-X[34] also achieves a fast computational time of $O(n)$. In the second step of p value calculation, computational time is fast if the summary statistic is a test statistic with a closed-form theoretical null distribution, as is the case for most methods that use regression tests. However, methods that use test statistics without theoretical null distributions require hundreds to thousands of rounds of permutation to calculate the null distribution, making them computationally intensive. To improve the scalability of a method, considerations can be put into expediting both steps. The first step can be accelerated using approximation algorithms. The number of permutations in the second step can be reduced through approaches such as p value-free FDR control[80], parametric smoothing of the permutation-based null distribution[81], and adaptive strategies that use a large number of permutations only for potentially small p values[82].

Future direction 1: accommodating differences in SRT technology and tissue types, and supporting the analysis of multi-sample SRT data

There are two key differences among SRT technologies, yet most current SVG detection methods do not account for these differences in the pre-processing and modeling of SRT data. Below, we introduce these differences and discuss their potential impact on SVG detection:

(1) Spatial resolution: imaging-based SRT measures per transcript's spatial localization at a single-cell or subcellular resolution, whereas sequencing-based SRT captures data at a multicellular level with relatively coarse resolution. In imaging-based SRT data, detected SVGs might result from the irregular distribution of cell types, such as cancer cells spreading across tumor tissue or mixed cell types at the boundaries of tissue layers. In contrast, the coarse resolution of sequencing-based SRT likely results in smaller variance of gene expression levels across spatial spots compared to imaging-based SRT data. Consequently, the spatial resolution differences between the two types of SRT technologies necessitate different interpretations for the detected SVGs. However, most existing SVG detection methods do not distinguish between the SRT technologies but use the same approach to identify SVGs, indicating a need for improvement.

(2) Positional randomness of spatial spots: different SRT technologies use various strategies to record the positions of spatial spots, resulting in differences in positional randomness. For instance, technologies like Spatial Transcriptomics and 10x Visium capture transcripts in predefined rectilinear or hexagonal grids on micro-slides, leading to structured spatial spots that lack positional randomness. In contrast, technologies like Slide-seq, MERFISH, and SeqFISH capture transcripts wherever they are located, without using predefined grids, leading to unstructured spatial spots with positional randomness. This difference in spatial positional randomness necessitates different modeling strategies for spatial variance. For example, sepal[40] designs distinct spatial modeling frameworks—rectilinear, hexagonal, and unstructured spots—for Spatial Transcriptomics,

10x Visium, and Slide-seq, respectively. MERINGUE[55] and several other graph-based methods use Delaunay triangulation to account for the nonuniform density of unstructured spots. BOOST-HMI[49] uses the mark interaction model, an extension of the Ising model used in BOOST-MI, to address the irregular distribution of spatial spots for imaging-based SRT data. However, the majority of SVG detection methods have not considered the positional randomness differences among SRT technologies, potentially leading to biased detection of SVGs.

Moreover, the type of tissue plays a crucial role in SVG detection. Whether a tissue is morphologically structured or not should be considered when defining SVGs to ensure they are biologically meaningful. In well-structured tissues, such as the main olfactory bulb or cerebellum, meaningful SVGs are expected to exhibit regular expression patterns that align with the tissue's morphological structures. Therefore, kernel-based SVG detection methods, which identify SVGs based on regular kernel patterns, may be particularly relevant in these cases. In contrast, for tissues without clear morphological structures, such as tumors or diseased tissues, meaningful SVGs may display irregular patterns. In other words, the expression patterns of SVGs can differ significantly between well-structured and unstructured tissues. Consequently, SVG detection should account for the morphological characteristics of the tissue. Among existing methods, SpaGCN[66] follows this rationale by defining spatial domains through the joint analysis of SRT data and H&E images, encouraging the identified spatial domains to reflect the underlying tissue structures. The SVGs identified by SpaGCN then serve as spatial-domain markers that explain these spatial domains. However, it remains an open question whether incorporating H&E images into overall SVG detection is necessary, especially if overall SVGs are primarily used as pre-screened features for further analysis and computational efficiency is a primary concern. For cell-type-specific SVG detection, H&E images have not yet been incorporated into existing methods. This raises an interesting future research question: how do cell types interact with morphological structures?

Another open question and future direction is the detection of SVGs from multi-sample SRT data, such as consecutive slices of a single tissue or tissue slices from different patients. Existing methods typically detect SVGs from a single tissue slice; however, intuitively, SVGs should be meaningful features that can be compared across multiple tissue slices. The challenge, therefore, is how to effectively identify SVGs from multi-sample SRT data to enable downstream multi-sample comparisons (e.g., normal vs. diseased tissue). In multi-sample SRT data of well-structured tissues, a common strategy is to align the tissue slices into a common 2D coordinate system[83–86]. This alignment facilitates the detection of SVGs. However, for unstructured tissues (e.g., tumors), alignment may not be feasible or reasonable, necessitating new strategies for SVG detection that can summarize the common, meaningful spatial variations in multi-sample SRT data of unstructured tissues.

Future direction 2: enhancing statistical rigor and method validation

Similar to single-cell RNA-seq data analysis[87–89], SRT data analysis faces the "double-dipping" challenge: the same data are analyzed more than once, and the final statistical tests rely on variables inferred from the same data in previous steps, leading to a "confirmation bias". This issue is prominent in spatial-domain-marker SVG detection, where the genes involved in identifying spatial domains would inherently be identified as spatial-domain markers in a subsequent step, even if they do not exhibit significant expression changes between spatial domains. Therefore, strategies are needed to remove false-positive discoveries resulting from double-dipping, such as ClusterDE[89].

Moreover, as many SVG detection methods use complex algorithms with implicit assumptions, interpretable ways for sanity checks on the detected SVGs are essential. For example, using in silico negative control SRT data can help users test SVG detection methods and identify spurious discoveries. Additionally, fast visualization tools will assist users in interpreting the top detected SVGs, providing a clearer understanding of the results and enhancing the reliability of the findings.

No method can be optimal in every aspect, making method choice for users' goals crucial. Users should select methods that capture the spatial patterns of interest, ensuring the method's strengths align with their research needs. Consequently, the diversity of methods is indispensable to cater to various user needs, and method validation and benchmarking are necessary for users to choose the appropriate methods. However, benchmarking is challenging due to the lack of well-annotated datasets with SVG ground truths. As a result, methods often justify their effectiveness indirectly, which may or may not reflect the biological questions users seek answers for. Early methods like SpatialDE and Trendsceek validate their detected SVGs by visual inspection. Another prevalent strategy is to use synthetic datasets with artificial spatial patterns, such as hotspots, streaks, and rings (Fig. 5a), to evaluate methods' detection power. However, these artificial patterns are oversimplified and might not be biologically relevant (Fig. 5b). Hence, benchmarking needs well-annotated SRT datasets from diverse tissues (e.g., structured tissues like the brain and unstructured tissues like tumors) and realistic SRT data simulators[90,91], not only for comprehensive method validation but also for the potential development of supervised methods. We curated the SRT datasets used by the existing SVG detection methods in Supplementary Data 1. Future research is needed to enhance the realism, comprehensiveness, and scalability of SRT data simulators.

Guidance for future benchmark studies

To date, three benchmark studies have been conducted to evaluate the performance of selected SVG detection methods[23–25]. These studies highlight the complexity of the SVG detection task and the variability in the performance of existing methods. In Supplementary Table 1, we summarize the three benchmark studies, detailing their choices of methods, benchmark datasets, evaluation metrics, and conclusions.

A limitation of the existing benchmark studies is their lack of consideration for the conceptual differences among SVG detection methods (in our categorization: overall SVGs, cell-type-specific SVGs, and spatial-domain-marker SVGs). This limitation can lead to unnecessary comparisons that are essentially like comparing apples to oranges. For instance, two benchmark studies[23,25] compared overall SVG detection methods (such as SpatialDE and SPARK) with spatial-domain-marker SVG detection methods (such as SpaGCN), using the number of SVGs detected as the evaluation metric. However, these two types of methods have distinct goals: overall SVG detection methods are intended to serve as a preprocessing step for feature selection and are inherently more likely to detect a greater number of genes as SVGs, while spatial-domain-marker SVG detection methods are specific to pre-defined spatial domains and tend to identify fewer genes as spatial domain markers. Conceptually, the number of detected genes is not a fair metric for comparing these two types of methods. Indeed, this anticipated result was confirmed in both benchmark studies. Had this conceptual categorization been recognized before the benchmark studies were conducted, computational resources could have been conserved by avoiding such comparisons.

A second limitation of the existing benchmark studies is that, although they compared SVG detection methods using diverse methodologies, their conclusions did not adequately summarize or emphasize how different core methodologies (e.g., graph-based or kernel-based approaches) affect the methods' performance within the benchmarks. This limitation means that the impact of fundamental methodological choices on the effectiveness of SVG detection remains
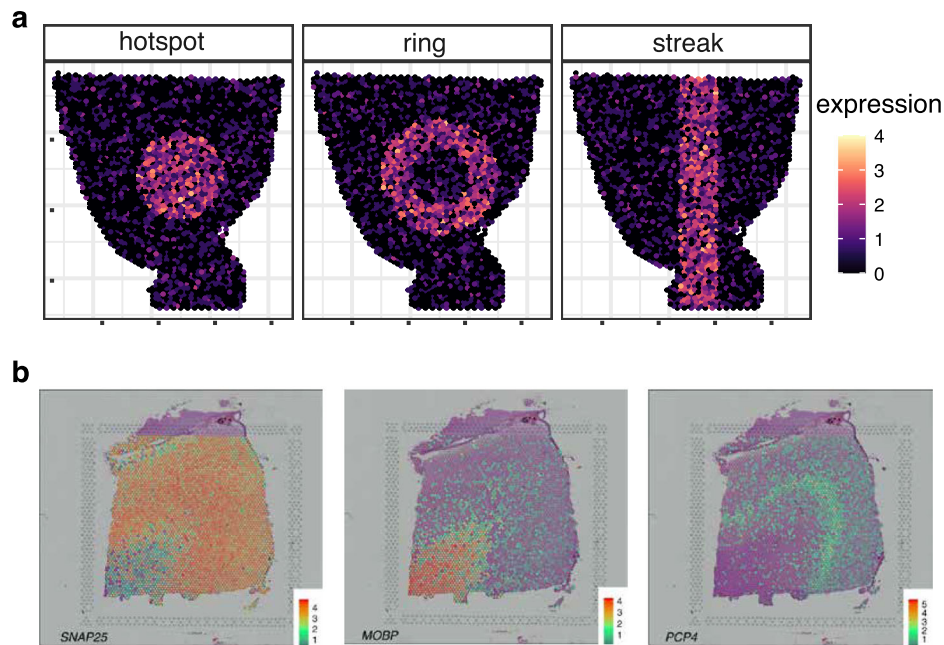
Fig. 5 | Synthetic spatial patterns oversimplify the spatial patterns observed in real SRT data. a Representative spatial patterns used in synthetic SRT data for evaluating SVG detection methods such as Trendsceek[37], SpatialDE[26], and others. **b** Spatial patterns shown in the 10x Visium dataset[19] profiling a dorsal lateral prefrontal cortex sample. The color indicates the log transformed expression for genes *SNAP25*, *MOBP*, and *PCP4*. With permission from Springer Nature.

underexplored, potentially limiting the insights that could guide future method development.

A third limitation of the existing benchmark studies is their varied selection of SRT datasets and benchmarking scenarios, leading to evaluation results and conclusions that are not fully comparable. Using limited benchmarking scenarios may favor methods that perform well under specific conditions while overlooking their robustness across a broader range of spatial expression patterns. Consequently, without a consensus on benchmark design, the conclusions drawn from these studies may not provide a comprehensive assessment of the strengths and weaknesses of the methods being compared.

In light of the three identified limitations in existing benchmark studies, we offer three recommendations for future benchmark studies, guided by our conceptual categorization of SVG detection methods. Recognizing that the definitions of SVGs differ among the three conceptual categories (overall SVGs, cell-type-specific SVGs, and spatial-domain-marker SVGs), it is both unfair and uninformative to compare methods across these categories. Therefore, our first recommendation is that future benchmark studies should evaluate methods within each conceptual category, providing more accurate and insightful assessments of SVG detection methods, and better guiding researchers in selecting the most appropriate methods for their specific needs.

Our second recommendation is to benchmark distinct methodologies aimed at achieving the same detection goal. This approach is valuable for method developers, as it provides a clearer understanding of the sensitivity and specificity of different methodologies, helping to avoid reinventing the wheel or encountering unknown pitfalls in future method development. For example, in the context of overall SVG detection, it would be insightful to compare the performance of dependence tests (with a general independence null hypothesis) against regression tests (with a model-based null hypothesis). Another question worth investigating is whether graph-based methods using Delaunay triangulation, such as MERINGUE and Binspect, can better capture cellular adjacency compared to KNN-based methods like SpaGene and Hotspot in tissues where cells are non-uniformly arranged.

Our third recommendation is to include SRT datasets from diverse tissue types and technologies, along with comprehensive and carefully designed simulations, to ensure broad coverage of informative spatial expression patterns. While this practice is challenging, it is essential to avoid benchmark conclusions that may biasedly favor certain types of SVG detection methods. For instance, if SRT data is simulated based on a specific kernel pattern, benchmark results will inherently favor SVG methods tailored to that kernel. Similarly, if the simulation design includes only a limited range of spatial expression patterns, methods using dependence tests—which operate under a general independence null hypothesis and can detect SVGs with diverse patterns—may be unfairly disfavored. We curated the SRT datasets used by the existing SVG detection methods in Supplementary Data 1. Looking forward, it is essential for the community to collaboratively curate a benchmark database to ensure fair benchmarking and foster future method development.

References

1. Longo, S. K., Guo, M. G., Ji, A. L. & Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* **22**, 627–644 (2021).
2. Marx, V. Method of the year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–14 (2021). **This paper introduces spatially resolved transcriptomics as a transformative methodology for studying gene expression within the spatial context of tissues, highlighting its potential to revolutionize our understanding of cellular organization and interactions.**
3. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–60 (2013).
4. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
5. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
6. Linus Eng, C.-H. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).

7. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

8. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).

9. Alon, S. et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656 (2021).

10. Janesick, A. et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).

11. Zhang, M. et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143 (2021).

12. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).

13. Stogsdill, J. A. et al. Pyramidal neuron subtype diversity governs microglia states in the neocortex. *Nature* **608**, 750–756 (2022).

14. 10xGenomics. Inside visium spatial capture technology. https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_BR060_Inside_Visium_Spatial_Technology.pdf (2019).

15. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

16. Zollinger, D. R., Lingle, S. E., Sorg, K., Beechem, J. M. & Merritt, C. R. Geomx™ RNA assay: high multiplex, digital, spatial analysis of RNA in FFPE tissue. *In Situ Hybridization Protocols, edited by Boye Schnack Nielsen (Molecular Histology, Bioneer A/S, Horsholm, Denmark) and Julia Jones (Cancer Research UK, University of Cambridge, Cambridge, UK),* 331–345 (Springer, 2020).

17. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).

18. Cho, C.-S. et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559–3572 (2021).

19. Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).

20. Abdelaal, T. et al. SIRV: spatial inference of RNA velocity at the single-cell resolution. *NAR genomics and bioinformatics* **6**, 3 (2024).

21. Das Adhikari, S., Yang, J., Wang, J. & Cui, Y. Recent advances in spatially variable gene detection in spatial transcriptomics. *Comput. Struct. Biotechnol. J.* **23**, 883–891 (2024). **This review highlights recent advances in methods for detecting spatially variable genes in spatial transcriptomics data, focusing on the methodological characteristics of methods**.

22. Cai, P., Robinson, M. D. & Tiberi, S. DESpace: spatially variable gene detection via differential expression testing of spatial clusters. *Bioinformatics* **40**, btae027 (2024).

23. Charitakis, N. et al. Disparities in spatially variable gene calling highlight the need for benchmarking spatial transcriptomics methods. *Genome Biol.* **24**, 209 (2023).

24. Chen, C., Kim, H. J. & Yang, P. Evaluating spatially variable gene detection methods for spatial transcriptomics data. *Genome Biol.* **25**, 18 (2024).

25. Li, Z. et al. Benchmarking computational methods to identify spatially variable genes and peaks. *bioRxiv* 2023.12.02.569717 (2023).

26. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018). **This study presents SpatialDE, the first method for detecting spatially variable genes; its kernel-based, Euclidean-space-based approach has inspired many later methods**.

27. Weber, L. M., Saha, A., Datta, A., Hansen, K. D. & Hicks, S. C. nnSVG for the scalable identification of spatially variable genes using nearest-neighbor gaussian processes. *Nat. Commun.* **14**, 4059 (2023).

28. Datta, A., Banerjee, S., Finley, A. O. & Gelfand, A. E. Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *J. Am. Stat. Assoc.* **111**, 800–812 (2016).

29. Hao, M., Hua, K. & Zhang, X. SOMDE: a scalable method for identifying spatially variable genes with self-organizing map. *Bioinformatics* **37**, 4392–4398 (2021).

30. Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J. & Stegle, O. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell Rep.* **29**, 202–211 (2019).

31. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* **17**, 193–200 (2020).

32. BinTayyash, N. et al. Non-parametric modelling of temporal and spatial counts data from RNA-seq experiments. *Bioinformatics* **37**, 3788–3795 (2021).

33. Li, Q., Zhang, M., Xie, Y. & Xiao, G. Bayesian modeling of spatial molecular profiling data via Gaussian process. *Bioinformatics* **37**, 4129–4136 (2021).

34. Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol.* **22**, 1–25 (2021).

35. Seal, S., Bitler, B. G. & Ghosh, D. SMASH: Scalable method for analyzing spatial heterogeneity of genes in spatial transcriptomics data. *PLoS Genet.* **19**, e1010983 (2023).

36. Vandenbon, A. & Diez, D. A clustering-independent method for finding differentially expressed genes in single-cell transcriptome data. *Nat. Commun.* **11**, 1–10 (2020).

37. Edsgärd, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* **15**, 339–342 (2018). **This paper introduces Trendsceek, a representative kernel-free, Euclidean-space-based method for detecting spatially variable genes**.

38. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell.* https://doi.org/10.1016/j.cell.2021.04.048 (2021).

39. Moehlin, J., Mollet, B., Colombo, B. M. & Mendoza-Parra, M. A. Inferring biologically relevant molecular tissue substructures by agglomerative clustering of digitized spatial transcriptomes with multilayer. *Cell Syst.* **12**, 694–705 (2021).

40. Andersson, A. & Lundeberg, J. sepal: identifying transcript profiles with spatial patterns by diffusion-based modeling. *Bioinformatics* **37**, 2644–2650 (2021).

41. Kantorovich, L. V. *Approximate Methods of Higher Analysis* (Interscience, 1958).

42. Palla, G. et al. Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).

43. Wang, J. et al. Dimension-agnostic and granularity-based spatially variable gene identification using BSP. *Nat. Commun.* **14**, 7367 (2023).

44. Liang, Y. et al. PROST: quantitative identification of spatially variable genes and domain detection in spatial transcriptomics. *Nat. Commun.* **15**, 600 (2024).

45. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).

46. Yuan, X. et al. HEARTSVG: a fast and accurate method for identifying spatially variable genes in large-scale spatial transcriptomics. *Nat. Commun.* **15**, 5700 (2024).

47. Bae, S., Choi, H. & Lee, D. S. Discovery of molecular features underlying the morphological landscape by integrating spatial transcriptomic data with deep features of tissue images. *Nucleic Acids Res.* **49**, e55–e55 (2021).

48. Jiang, X., Xiao, G. & Li, Q. A bayesian modified ising model for identifying spatially variable genes from spatial transcriptomics data. *Stat. Med.* **41**, 4647–4665 (2022).

49. Yang, J., Jiang, X., Wang Jin, K., Shin, S. & Li, Q. Bayesian hidden mark interaction model for detecting spatially variable genes in imaging-based spatially resolved transcriptomics data. *Front. Genet.* **15**, 1356709 (2024).

50. DeTomaso, D. & Yosef, N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446–456 (2021). ***This study presents Hotspot, a representative graph-based method for identifying spatially variable genes***.

51. Wu, Y. et al. Highly regional genes: graph-based gene selection for single-cell rna-seq data. *J. Genet. Genomics* **49**, 891–899 (2022).

52. Hong, Y. et al. The spatiotemporal dynamics of spatially variable genes in developing mouse brain revealed by a novel computational scheme. *Cell Death Discov.* **9**, 264 (2023).

53. Jiang, R., Li, Z., Jia, Y., Li, S. & Chen, S. SINFONIA: scalable identification of spatially variable genes for deciphering spatial domains. *Cells* **12**, 604 (2023).

54. Geary, R. C. The contiguity ratio and statistical mapping. *Inco. Stat.* **5**, 115–146 (1954).

55. Miller, B. F., Bambah-Mukku, D., Dulac, C., Zhuang, X. & Fan, J. Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Res.* **31**, 1843–1855 (2021).

56. Preparata, F. P. & Shamos, M. I. *Computational Geometry: An Introduction* (Springer Science & Business Media, 2012).

57. Liu, Q., Hsu, C.-Y. & Shyr, Y. Scalable and model-free detection of spatial patterns and colocalization. *Genome Res.* **32**, 1736–1745 (2022).

58. Chang, Y. et al. Graph fourier transform for spatial omics representation and analyses of complex organs. *Nat. Commun.* **15**, 7467 (2024).

59. Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a "knee-dle" in a haystack: detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, 166–171 (IEEE, 2011).

60. Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 1–31 (2021).

61. Zhang, K., Feng, W. & Wang, P. Identification of spatially variable genes with graph cuts. *Nat. Commun.* **13**, 1–15 (2022).

62. Govek, K. W., Yamajala, V. S. & Camara, P. G. Clustering-independent analysis of genomic data using spectral simplicial theory. *PLoS Comput. Biol.* **15**, e1007509 (2019).

63. Horak, D. & Jost, J. Spectra of combinatorial laplace operators on simplicial complexes. *Adv. Math.* **244**, 303–336 (2013).

64. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

65. Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).

66. Hu, J. et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).

67. Guo, W. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-sequencing reveals tissue architecture in esophageal squamous cell carcinoma. *EBioMedicine* **84**, 104281 (2022).

68. Sunkin, S. M. et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* **41**, D996–D1008 (2012).

69. Zhao, E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).

70. Pham, D. et al. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nat. Commun.* **14**, 7739 (2023).

71. Yu, J. & Luo, X. Identification of cell-type-specific spatially variable genes accounting for excess zeros. *Bioinformatics* **38**, 4135–4144 (2022).

72. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).

73. Cable, D. M. et al. Cell type-specific inference of differential expression in spatial transcriptomics. *Nat. Methods* **19**, 1076–1087 (2022).

74. Yu, H. & Li, W. V. spVC for the detection and interpretation of spatial gene expression variation. *Genome Biol.* **25**, 103 (2024).

75. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded nmf regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50–e50 (2021).

76. Li, J. J., Zhou, H. J., Bickel, P. J. & Tong, X. Dissecting gene expression heterogeneity: generalized pearson correlation squares and the k-lines clustering algorithm. *J. Am. Stat. Assoc.* **119**, 2450–2463 (2024).

77. Li, Y., Ge, X., Peng, F., Li, W. & Li, J. J. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biol.* **23**, 79 (2022).

78. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

79. Ma, C., Chitra, U., Zhang, S. & Raphael, B. J. Belayer: Modeling discrete and continuous spatial variation in gene expression from spatially resolved transcriptomics. *Cell Syst.* **13**, 786–797 (2022).

80. Ge, X. et al. Clipper: *p*-value-free FDR control on high-throughput data from two conditions. *Genome Biol.* **22**, 1–29 (2021).

81. Song, D. & Li, J. J. PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated *p*-values from single-cell RNA sequencing data. *Genome Biol.* **22**, 124 (2021).

82. Tiberi, S., Crowell, H. L., Samartsidis, P., Weber, L. M. & Robinson, M. D. distinct: a novel approach to differential distribution analyses. *Ann. Appl. Stat.* **17**, 1681–1700 (2023).

83. Clifton, K. et al. STalign: alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nat. Commun.* **14**, 8123 (2023).

84. Jones, A., Townes, F. W., Li, D. & Engelhardt, B. E. Alignment of spatial genomics data using deep Gaussian processes. *Nat. Methods* **20**, 1379–1387 (2023).

85. Zhang, M. et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).

86. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3d microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).

87. Zhang, J. M., Kamath, G. M. & David, N. T. Valid post-clustering differential analysis for single-cell RNA-Seq. *Cell Syst.* **9**, 383–392 (2019).

88. Neufeld, A., Gao, L. L., Popp, J., Battle, A. & Witten, D. Inference after latent variable estimation for single-cell RNA sequencing data. *Biostatistics* **25**, 270–287 (2024).

89. Song, D., Li, K., Ge, X. & Li, J. J. ClusterDE: a post-clustering differential expression (DE) method robust to false-positive inflation caused by double dipping. *bioRxiv* 2023.07.21.550107 (2023).

90. Zhu, J., Shang, L. & Zhou, X. SRTsim: spatial pattern preserving simulations for spatially resolved transcriptomics. *Genome Biol.* **24**, 39 (2023).

91. Song, D. et al. scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics. *Nat. Biotechnol.* **42**, 247–252 (2024). ***This paper introduces scDesign3, a cutting-edge tool for generating realistic in silico datasets for multimodal single-cell and spatial omics, facilitating benchmarking and methodological development in the field***.

Author contributions
G.Y., S.H.H., and J.J.L. wrote the manuscript. G.Y. prepared the figures. G.Y., S.H.H., and J.J.L. prepared the tables. G.Y. and J.J.L. revised the manuscript. All authors approved the final version of the manuscript for publication.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-56080-w.

Correspondence and requests for materials should be addressed to Jingyi Jessica Li.