



Published in final edited form as:

Nat Methods. 2020 February ; 17(2): 193–200. doi:10.1038/s41592-019-0701-7.

Statistical Analysis of Spatial Expression Pattern for Spatially Resolved Transcriptomic Studies

Shiquan Sun^{1,2,*}, Jiaqiang Zhu^{2,*}, Xiang Zhou^{2,3,#}

¹School of Computer Science, Northwestern Polytechnical University, 710072 Xi'an, Shaanxi, P.R. China

²Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

³Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Identifying genes that display spatial expression pattern in spatially resolved transcriptomic studies is an important first step towards characterizing the spatial transcriptomic landscape of complex tissues. Here, we developed a statistical method, SPARK, for identifying such spatially expressed genes in data generated from various spatially resolved transcriptomic techniques. SPARK directly models spatial count data through the generalized linear spatial models. It relies on newly developed statistical formulas for hypothesis testing, providing effective type I error control and yielding high statistical power. With a computationally efficient algorithm based on penalized quasi-likelihood, SPARK is also scalable to data sets with tens of thousands of genes measured on tens of thousands of samples. In four published spatially resolved transcriptomic data sets, we show that SPARK can be up to ten times more powerful than existing methods, revealing new biology in the data that otherwise cannot be revealed by existing approaches.

INTRODUCTION

Recent emergence of various spatially resolved transcriptomic technologies has enabled gene expression profiling with spatial localization information on tissues or cell cultures. Exemplary techniques include MERFISH¹ and seqFISH², which are based on single-molecular fluorescence *in situ* hybridization (smFISH)³ and can measure hundreds of genes with subcellular spatial resolution; TIVA⁴, LCM⁵, Tomo-Seq⁶ and spatial transcriptomics through spatial barcoding⁷, which are based on the next generation DNA sequencing and can measure tens of thousands of genes on single cells or on spatial locations consisting of a couple hundred single cells; targeted *in situ* sequencing (ISS)⁸ and FISSEQ⁹, which are based on *in situ* RNA sequencing and can measure the entire transcriptome with spatial

#: correspondence to XZ (xzhousph@umich.edu).

Contributions

XZ conceived the idea and provided funding support. SS and XZ designed the experiments. SS and JZ developed the method, implemented the software, performed simulations and analyzed real data. SS, JZ and XZ wrote the manuscript.

*: authors contributed equally

Competing interests

The authors declare that they have no competing interests.

information at a single cell resolution. These different spatially resolved transcriptomic techniques altogether have made it possible to study the spatial organization of transcriptomic landscape across tissue sections or within single cells, catalyzing new discoveries in many areas of biology^{10, 11}.

In spatially resolved transcriptomic studies, identifying genes that display spatial expression pattern, which we simply refer to as SE analysis, is an important first step towards characterizing the spatial transcriptomic landscape in tissues. Effective SE analysis faces important statistical and computational challenges. From a statistical perspective, identifying SE genes requires proper modeling of the raw count data generated from both smFISH and sequencing based techniques. Unfortunately, the only two existing approaches for SE analysis, SpatialDE¹² and Trendsceek¹³, transform count data into normalized data before analysis. As is well documented in many other omics sequencing studies^{14, 15}, analyzing normalized data can be suboptimal as this approach fails to account for the mean-variance relationship existed in raw counts, leading to a potential loss of power¹⁶. Besides direct modeling of count data, identifying SE genes also requires the development of statistical methods that can produce well calibrated p -values for type I error control. However, some existing methods for SE analysis, such as SpatialDE¹², rely on asymptotic normality and minimal p -value combination rule for constructing hypothesis tests. Subsequently, these methods may fail to control for type I error at small p -values that are essential for detecting SE genes at the transcriptome-wide significance level, potentially leading to excessive false positives and/or substantial loss of power. From a computational perspective, while some spatial methods such as SpatialDE are based on linear mixed models and are computationally efficient, some other spatial methods, in particular Trendsceek¹³, are built without a data generative model and compute non-parametric test statistics through computationally expensive permutation strategies. Consequently, analyzing even moderate sized spatial transcriptomics data with hundreds of genes across hundreds of spatial locations can be a daunting task for these methods.

Here, we present a new method, which we refer to as Spatial PAttern Recognition via Kernels (SPARK), that addresses the above statistical and computational challenges. SPARK builds upon a generalized linear spatial model (GLSM)^{17, 18} with a variety of spatial kernels to accommodate count data generated from smFISH or sequencing based spatial transcriptomics studies. With a newly developed penalized quasi-likelihood (PQL) algorithm^{19, 20}, SPARK is scalable to analyzing tens of thousands of genes across tens of thousands spatial locations. Importantly, SPARK relies on a mixture of Chi-square distributions to serve as the exact test statistics distribution and takes advantage of a recently developed Cauchy combination rule^{21, 22} to combine information across multiple spatial kernels for calibrated p -value calculation. As a result, SPARK properly controls for type I error at the transcriptome-wide level and is more powerful for identifying SE genes than existing approaches.

RESULTS

Simulations

We provide an overview of SPARK in Online Methods, with technical details provided in Supplementary Notes and a method schematic shown in Fig. 1a. Unlike Trendsceek, SPARK has an underlying data generative model which can be viewed as an extension of SpatialDE. However, unlike SpatialDE, SPARK models count data directly and relies on a proper statistical procedure to obtain calibrated p -values. A more detailed description of the differences among these methods is provided in Supplementary Notes. We performed two sets of simulations to evaluate the performance of SPARK and compared it with two existing approaches, SpatialDE and Trendsceek. Simulation details are provided in Supplementary Notes.

In the first set of simulations, we found that, under the null, SPARK produced well-calibrated p -values at the transcriptome-wide significance levels (Fig. 1b). Some Trendsceek test statistics (e.g. markvario and Vmark) also produced reasonably calibrated p -values while others (e.g. Emark statistics and markcorr statistics) yielded slightly conservative p -values. In contrast, SpatialDE produced overly conservative p -values (Fig. 1b). The failure of SpatialDE for type I error control presumably is due to its use of an asymptotic Chi-square distribution in place of an exact distribution for p -value computation and/or its use of the *ad hoc* minimal p -value combination rule. The p -value calibration results under the null for different methods were consistent across simulation settings and across a range of noise variance levels (Supplementary Fig. 1a). Because some methods failed to control for type I error, we measured power based on false discovery rate (FDR) in the alternative simulations to ensure a fair comparison across methods. Under the alternatives, we found that SPARK was more powerful than the other two methods across a range of FDR cutoffs (Fig. 1c) and across a range of parameter settings (Supplementary Figs. 1b, and c). The power performance of SPARK was followed by SpatialDE, while Trendsceek did not fare well in any of the alternative simulations.

Because of the extremely poor performance of Trendsceek, we performed a second set of simulations fully based on the original Trendsceek paper¹³. The comparison results on the second set of simulations are largely consistent with the results obtained from the first set of simulations. Specifically, under the null, both SPARK and Trendsceek produced well-calibrated p -values, while SpatialDE did not (Fig. 1d). Under the alternative, SPARK was more powerful than the other two methods across a range of FDR cutoffs (Fig. 1e) in almost all parameter settings (Supplementary Fig. 2). The power performance of SPARK was followed by SpatialDE, while Trendsceek did not fare well, even though the power of Trendsceek was largely consistent with the original study¹³. Overall, the two sets of simulations suggest that SPARK produces well-calibrated p -values while being more powerful than the other two methods in detecting SE genes.

Olfactory Bulb Data

We applied SPARK to analyze four published data, including two data obtained through spatial transcriptomics sequencing and two data through smFISH (details in Online

Methods). The first data is a mouse olfactory bulb data⁷, consisting of 11,274 genes measured on 260 spots. Consistent with simulations, both SPARK and Trendsceek produced calibrated p -values under permuted null, while SpatialDE did not (Fig. 2a). SPARK also identified more SE genes compared to SpatialDE and Trendsceek across a range of FDRs (Fig. 2b, and Supplementary Fig. 3a). For example, at an FDR of 5%, SPARK identified 772 SE genes, which is ~10-fold more than that detected by SpatialDE (which identified 67, among which 62 are overlapped with SPARK; Figs. 2b, and f). Trendsceek was unable to detect any SE genes in the data, even though we tried ten different random seeds for the method.

We carefully examined the SE genes and found that most SE genes only detected by SpatialDE tend to have close to zero expression levels (Supplementary Fig. 3b) and appear to locate on either one or two spots (Supplementary Fig. 3d), suggesting potentially false signals. In contrast, the SE genes only detected by SPARK generally have comparable expression levels to the SE genes detected by both methods (Supplementary Fig. 3b). To assess the quality of the SE genes identified by SPARK, we performed clustering on the 772 SE genes and obtained three major spatial expression patterns (dendrogram in Fig. 2e; UMAP visualization in Supplementary Fig. 3c): one representing the mitral cell layer (Pattern I); one representing the glomerular layer (Pattern II); and one representing the granular cell layer (Pattern III); all clearly visualized via three previously known marker genes for the three layers, *Doc2g*, *Kctd12* and *Penk*⁷ (Fig. 2d). We listed 20 random genes only detected by SPARK as representative examples (Supplementary Fig. 4). Almost all these genes showed clear spatial expression patterns that were cross validated by *in situ* hybridization in the Allen Brain Atlas (Fig. 2c), confirming the higher power of SPARK.

We provide three additional lines of evidence to validate the SE genes detected by SPARK. First, we examined the highlighted marker genes in the olfactory system presented in the original study⁷. The list of highlighted marker genes, while is not necessarily the complete list of all marker genes, at least represents the likely set of genes that are both biologically important and detectable in the data. Importantly, SPARK detected 8 of 10 such highlighted genes. SpatialDE only detected 3 and Trendsceek detected none (Supplementary Fig. 5d). Second, we obtained a list of 2,030 cell type specific marker genes identified in a recent single cell RNAseq study in the olfactory bulb²³. Reassuringly, 55% of the unique SE genes identified by SPARK were in the marker list, while only 20% of the unique SE genes identified by SpatialDE were in the same list (Fig. 2g). Third, we obtained a list of 3,262 genes that are related to the olfactory system in the Harmonizome database²⁴. Again, 26% of the unique SE genes identified by SPARK were in the Harmonizome list, while only 20% of the unique SE genes identified by SpatialDE were in the same list (Fig. 2g). These three additional validation analyses provide convergence support for the higher power of SPARK.

Finally, we performed functional enrichment analyses of SE genes identified by SPARK and SpatialDE (details in Online Methods). A total of 1,023 GO terms (Fig. 2h) and 79 KEGG pathways were enriched in the SE genes identified by SPARK at an FDR of 5%, while only 87 GO terms (overlap = 64; Supplementary Fig. 5a) and 2 KEGG pathways (overlap = 2; Supplementary Fig. 5b) were enriched in the SE genes identified by SpatialDE (Supplementary Table 1, Supplementary Fig. 5c). Many enriched GO terms or KEGG

pathways identified only by SPARK are directly related to the synaptic organization and olfactory bulb development. Examples include olfactory lobe development (GO:0021988; SPARK: p -value = 5.81×10^{-3} ; SpatialDE: p -value = 1.21×10^{-1}) and oxytocin signaling pathway (KEGG: mmu04921; SPARK: p -value = 1.59×10^{-9} ; SpatialDE: p -value = 2.15×10^{-1}) for modulating olfactory processing²⁵. An in-depth enrichment analysis using SE genes in Patterns I-III separately provide additional biological insights, revealing the critical role of synaptic organization for the mitral cell layer, importance of cell junction and synaptic connectivity for the nerve layer, as well as function of dendritic morphogenesis and synaptic/dendritic plasticity for the granular layer (details in Supplementary Results; Supplementary Fig. 5 and Supplementary Table 1). Overall, the newly identified GO term and KEGG pathway enrichments highlight the benefits of running SE analysis with SPARK.

Breast Cancer Data

The second data is a human breast cancer biopsy study⁷, which contains 5,262 genes measured on 250 spots. Again, both SPARK and Trendsceek produced calibrated p -values under permuted null, while SpatialDE did not (Fig. 3a). SPARK also identified more SE genes compared to SpatialDE and Trendsceek across a range of FDRs (Fig. 3b, Supplementary Fig. 6a). For example, at an FDR of 5%, SPARK identified 290 SE genes, which is ~3-fold more than that detected by SpatialDE (which identified 115, among which 85 are overlapped with SPARK; Figs. 3b, and d). In contrast, Trendsceek only identified at most 15 SE genes. Again, SE genes only detected by SpatialDE tend to have low expression levels while the SE genes detected only by SPARK generally have comparable expression levels to the SE genes detected by both methods (Supplementary Fig. 6b). We listed 20 random genes only detected by SPARK as representative examples (Supplementary Fig. 7). Most of these genes show clear spatial expression pattern, confirming the higher power of SPARK.

We provide three additional lines of evidence to validate the SE genes detected by SPARK. First, we examined the 14 cancer relevant genes highlighted in the original study⁷. SPARK detected 10 of them. SpatialDE detected 7 and Trendsceek detected two (Supplementary Fig. 6e.) Both SpatialDE and Trendsceek missed three well-known cancer relevant genes: *SCGB2A2*, *KRT17* and *MMP14*. Second, we collected a list of 1,144 genes previously known to be relevant to breast cancer in the CancerMine database²⁶. 14% of SE genes uniquely identified by SPARK were in the list while only 10% by SpatialDE were in the list (Fig. 3c). For example, the well-known proto-oncogene *ERBB2* gene that has tens of thousands of previous literature support can only be identified by SPARK (Fig. 3e). Third, we collected a list of 3,538 genes that are relevant to breast cancer based on the Harmonizome database²⁴. Again, 44% of SE genes uniquely identified by SPARK were in the list while only 37% by SpatialDE were in the list (Fig. 3c). Overall, these three additional lines of evidence provide convergent support for the higher power of SPARK.

Finally, we performed functional enrichment analysis. At an FDR of 5%, SPARK identified 542 GO terms and 20 KEGG pathways (Fig. 3f, Supplementary Table 2) while SpatialDE identified 266 GO terms (overlap = 191) and 3 KEGG pathways (overlap = 3; Supplementary Figs. 6c, and d; (Supplementary Table 2). Many enriched gene sets

discovered only by SPARK are related to extracellular matrix organization and immune responses (representative genes shown in Fig. 3e; detailed analysis and results shown in Supplementary Results, Supplementary Fig. 6f, and Supplementary Table 2), highlighting their importance in cancer development and metastasis.

Hypothalamus Data

The third data is a MERFISH data collected on the preoptic area of the mouse hypothalamus²⁷. The data contains 160 genes measured on 4,975 single cells (Fig. 4c). 155 of these 160 genes were selected in the original study as they are makers of distinct cell populations or are relevant to various neuronal functions of the hypothalamus. Besides these likely true positive genes, another 5 blank control genes were also included in the original study to serve as negative controls. In the analysis, we found that SPARK produced calibrated p -values under permuted null, while SpatialDE did not (Fig. 4a). (Trendsceek was not applied to the permuted null due to its heavy computational burden.) The QQ-plots of p -values from different methods also suggest that both SpatialDE and SPARK are more powerful than Trendsceek (Supplementary Fig. 8a). Because this data contains 5 negative control genes and 155 likely positive genes, we directly compared power of different methods based on the number of SE genes identified given a fixed number of negative control genes identified (Fig. 4b). The power comparison results again support a higher power of SPARK. For example, conditional on only one blank control gene being detected (i.e. one false positive), SPARK identified 145 SE genes, which is 6 more than that detected by SpatialDE (which identified 139, among which 138 are overlapped with SPARK; Fig. 4b, Supplementary Fig. 8b). The performance of SPARK and SpatialDE is followed by Trendsceek, which identified 108 SE genes, among which 103 are overlapped with SPARK. A careful examination suggests that almost all SE genes identified by SPARK show clear spatial expression pattern as one would expect: 9 major cell classes in hypothalamus (Fig. 4d, Supplementary Fig. 8c) along with 9 marker genes²⁷ (Supplementary Fig. 8d) are shown as examples. Importantly, all three SE genes only identified by SPARK (*Avpr1a*, *Chat*, and *Nup62cl*) are closely related to the neuronal functions of the hypothalamus^{28–30} (details of these genes are provided in the Supplementary Results), highlighting the power of SPARK.

Hippocampus Data

The final data is a small seqFISH data with 249 genes measured on 131 single cells in the mouse hippocampus³¹ (Supplementary Fig. 9a). These 249 genes include 214 genes that were selected in the original study³¹ as transcription factors and signaling pathway components and 35 remaining genes that are previously known cell identity markers. In the analysis, we found that both SPARK and Trendsceek produced calibrated p -values under permuted null while SpatialDE yielded conservative p -values (Supplementary Fig. 9b). SPARK again identified more SE genes compared to SpatialDE and Trendsceek across a range of FDRs (Supplementary Figs. 9c, and d). For example, at an FDR of 5%, SPARK identified 17 SE genes. SpatialDE and Trendsceek identified 11 (all overlap with SPARK) and 4 (one overlap with SPARK) SE genes, respectively (Supplementary Fig. 9e). The 11 SE genes identified by both SpatialDE and SPARK show clear spatial expression patterns (Supplementary Fig. 9f), so are the 6 SE genes identified only by SPARK (Supplementary Fig. 9g). The three SE genes only detected by Trendsceek tend to express uniformly highly

in most cells and show less apparent spatial pattern (Supplementary Fig. 9h). The higher number and apparent spatial expression pattern of SE genes identified by SPARK support its higher power. We carefully examined all six SE genes that were only identified by SPARK. Four of them are cell identity markers: *Foxo1* and *Slc17a8* for glutamatergic neurons; *Igtp* for GABAergic neurons; and *Opalin* for oligodendrocytes³². All of them are closely related to neuronal functions in hippocampus. For example, the spatial expression pattern of *Foxo1* detected by SPARK is consistent with the previous observation that it is highly enriched in the ventral CA3 area of the hippocampus as well as in the amygdalohippocampal region^{33, 34}. *Foxo1* is activated in hippocampal progenitor stem cells following cortisol exposure to prenatal stress and mediates the negative effect of stress on neurogenesis³⁵. Besides these four marker genes, the remaining two genes, *Pou4f1* and *Gfil1*, encode neural transcription factors and play important roles in the sensory nervous system development^{36, 37}. These important genes that are missed by other methods again highlight the power of SPARK

DISCUSSION

We have presented a new method, SPARK, for identifying SE genes in spatially resolved transcriptomic studies. Compared with existing approaches, SPARK produces well-calibrated *p*-values and yields more statistical power. SPARK is also easily applicable to three-dimensional data sets such as STARmap³⁸ or even higher dimensional data sets where other coordinates (e.g. time) are recorded.

SPARK incorporates a data generative model and relies on a model-based hypothesis test framework for spatial pattern detection. The data generative model in SPARK distinguishes it from previous spatial data exploratory tools that rely on variogram or semi-variogram for visualizing spatial autocorrelation patterns^{39,40}. The model-based hypothesis test in SPARK also distinguishes it from previous simple spatial test statistics such as Moran's I and Geary's C^{41, 42} for detecting spatial autocorrelation patterns. To illustrate the benefits of SPARK over previous simple spatial test statistics, we have applied Moran's I test to all four real data sets examined here. We found that the *p*-values from Moran's I under permuted null were highly inflated, presumably due to its use of asymptotic normality for *p*-value computation (Supplementary Fig. 10a). We also found that the power of Moran's I was lower than SPARK across all data (Supplementary Figs. 10b, c, and d), likely due to its inability to detect spatial patterns other than simple autocorrelations.

SPARK directly models raw counts to account for the mean-variance dependency observed in the spatial data (Supplementary Fig. 11a), resulting in an appreciable power gain. Such power gain is especially apparent in data with low counts such as the first two spatial transcriptomics data we examined here. However, we acknowledge that the power gain brought by count modeling may be small in data with high counts (e.g. the MERFISH and seqFISH data), since a normal distribution can often approximate high counts as well as an over-dispersed Poisson distribution. Therefore, we develop a Gaussian version of SPARK (details in Supplementary Notes) to ensure more robust modeling and scalable computation for data with high counts. The Gaussian version of SPARK produced well-calibrated *p*-values in all permuted data (Supplementary Fig. 11b), had comparable power with the

Poisson version of SPARK for data with high counts, though was inferior in data with low counts (Supplementary Fig. 11c). Importantly, the Gaussian version of SPARK is much more computationally efficient than its Poisson counterpart. While the Poisson version of SPARK is fast (Supplementary Table 3, Supplementary Fig. 12), the Gaussian version of SPARK may represent an attractive alternative for analyzing large data collected from emerging techniques such as Slide-seq⁴³.

Finally, several potential extensions exist for SPARK. For example, we have aggregated p -values across different kernels to ensure stable performance across a range of possible spatial expression patterns. However, some kernels may work preferentially well for certain data sets (Supplementary Fig. 13), for certain spatial patterns, and/or for certain genes. Subsequently, it could be beneficial to estimate the weights of the ten kernels for each gene separately or to estimate them in an empirical Bayes fashion by borrowing spatial expression information across genes. It could also be beneficial to incorporate prior knowledge on the tissue structure into the kernel functions to facilitate the detection of genes that are specifically expressed in known structures. These future extensions will likely improve the power of SPARK further.

Online Methods

SPARK: Model and Algorithm

We consider modeling gene expression data collected by various high-throughput spatial sequencing techniques such as smFISH and spatial transcriptomics technology. These spatial techniques simultaneously measure gene expression levels of m different genes on n different spatial locations on a tissue of interest (which we simply refer to as samples). The gene expression measurements are often obtained in the form of counts: they are collected either as the number of barcoded mRNA for any given transcript in a single cell through smFISH based techniques or as the number of sequencing reads mapped to any given gene through sequencing based spatial techniques. The number of genes, m , varies across different spatial sequencing techniques and often ranges from a couple hundred (in the case of smFISH) to the whole transcriptome (in the case of spatial transcriptomics technology). The sample composition varies across different spatial sequencing techniques and can consist of either a single cell (in the case of smFISH) or a small set of approximately homogenous single cells residing in a small region of the sampled location known as a spot (in the case of spatial transcriptomics technology). The sampled locations have known spatial coordinates that are recorded during the experiment. These sampled locations can either be considered as random (in the case of smFISH; as expressions are measured on single cells that are randomly scattered across the tissue/culture space) or are pre-determined by researchers before the experiment (in the case of spatial transcriptomics technology). We denote $s_i = (s_{i1}, s_{i2})$ as the spatial coordinates (i.e. location index) for i th sample, with $i \in (1, \dots, n)$. These spatial coordinates vary continuously over a two-dimensional space R^2 , or $s_i \in R^2$. While we only focus on the case where samples are collected on a two-dimensional space of a tissue/culture layout, our model and method are general, capable of handling three-dimensional cases where the depth of the sample location in the tissue can be

recorded or handling even higher dimensional cases where other coordinates (e.g. time) are also recorded.

Our primary goal is to detect genes whose expression level displays spatial pattern with respect to the sample locations. We simply refer to these genes as SE genes (genes with spatial expression pattern), in parallel to DE genes (differentially expressed genes) used in other settings. To identify SE genes, we examine one gene at a time and model its expression level across sampled locations using a generalized linear spatial model (GLSM)^{44, 45}. GLSM, also known as the generalized linear geostatistical model or the spatial generalized linear mixed model, is a generalized linear mixed model that directly models non-Gaussian spatial data and uses random effects to capture the underlying stationary spatial process. GLSM has been commonly used for interpolation and prediction of spatial data, with applications in spatial disease mapping and spatial epidemiologic studies^{46, 47}. However, different from all these previous GLSM development, we instead focus on developing a hypothesis testing framework for GLSM. Here, for the gene of focus, we denote $y_i(s_i)$ as the gene expression measurement in terms of counts for the i th sample. We denote $\mathbf{x}_i(s_i)$ as a k -vector of covariates that include a scalar of one for the intercept and $k-1$ observed explanatory variables for the i th sample. These explanatory variables could contain batch information, cell cycle information, or other information that are important to adjust for during the analysis. We denote $N_i(s_i)$ as the normalization factor for i th sample. Here, we set $N_i(s_i)$ as the summation of the total number of counts across all genes for the sample as our main interest is in analyzing the relative gene expression level. Other choices of $N_i(s_i)$ are possible; for example, $N_i(s_i)$ can be set to one if the main interest is in the absolute gene expression level. We consider modeling the observed expression count data with an over-dispersed Poisson distribution

$$y_i(s_i) \sim \text{Poi}(N_i(s_i)\lambda_i(s_i)), i = 1, 2, \dots, n$$

where $\lambda_i(s_i)$ is an unknown Poisson rate parameter that represents the underlying gene expression level for the i th sample. In the spatial setting, $\lambda_i(s_i)$ can also be viewed as the unobserved spatial random process occurred at location s_i . We model the log scale of the latent variable $\lambda_i(s_i)$ as a linear combination of three terms,

$$\log(\lambda_i(s_i)) = \mathbf{x}_i(s_i)^T \boldsymbol{\beta} + b_i(s_i) + \epsilon_i,$$

where $\boldsymbol{\beta}$ is a k -vector of coefficients that include an intercept representing the mean log-expression of the gene across spatial locations together with $k-1$ coefficients for the corresponding covariates; ϵ_i is the residual error that is independently and identically distributed from $N(0, \tau_2)$ with variance τ_2 ; and $b_i(s_i)$ is a zero-mean, stationary Gaussian process modeling the spatial correlation pattern among spatial locations

$$\mathbf{b}(s) = (b_1(s_1), b_2(s_2), \dots, b_n(s_n))^T \sim \text{MVN}(0, \tau_1 \mathbf{K}(s)),$$

where the covariance $K(s)$ is a kernel function of the spatial locations $s = (s_1, \dots, s_n)^T$, with ij 'th element being $K(s_i, s_j)$; τ_1 is a scaling factor of the covariance kernel; and MVN denotes a multivariate normal distribution. We will discuss the choice of the kernel function in more details below. In the above model, the covariance for the latent variables $\log(\lambda(s))$ is $\Sigma = \tau_1 K(s) + \tau_2 I$, where $\lambda(s) = (\lambda_1(s_1), \lambda_2(s_2), \dots, \lambda_n(s_n))^T$ and I is an n -dimensional identity matrix. In spatial statistics, τ_1 is commonly referred to as the partial sill which effectively measures the expression variance in $\log(\lambda(s_i))$ captured by spatial patterns or spatial location information; τ_2 is commonly referred to as the nugget which effectively measures the expression variance in $\log(\lambda(s_i))$ due to random noise independent of spatial locations.

In the GLSM defined above, testing whether a gene shows spatial expression pattern can be translated into testing the null hypothesis $H_0: \tau_1 = 0$. The statistical power of such hypothesis test will inevitably depend on how the spatial kernel function $K(s)$ matches the true underlying spatial pattern displayed by the gene of interest. For example, a periodic kernel will be particularly useful to detect expression pattern that is periodic across the location space, while a Gaussian kernel will be particularly useful to detect expression pattern that is clustered in focal areas. The true underlying spatial pattern for any gene is unfortunately unknown and may vary across genes. To ensure robust identification of SE genes across various spatial patterns, we consider using a total of ten different spatial kernels, including five periodic kernels with different periodicity parameters and five Gaussian kernels with different smoothness parameters. The detailed construction of these kernels is described in Supplementary Notes. These ten kernels cover a range of possible spatial patterns that are observed in common biological data sets (Supplementary Fig. 13e) and are used as default kernels in our software implementation for all analysis results presented here. However, we note that our method and software implemented can easily handle many other kernel functions or incorporate different number of kernel functions as the users see fit.

We fit the above GLSM and test the null hypothesis using the ten kernels one at a time. Parameter estimation and hypothesis testing in GLSM is notoriously difficult, as the GLSM likelihood consists of an n -dimensional integral that cannot be solved analytically. To overcome the high dimensional integral and enable scalable estimation and inference with GLSM, we develop an approximate inference algorithm based on the penalized quasi-likelihood (PQL) approach^{20, 48}. The algorithmic details are provided in the Supplementary Notes. With parameter estimates from the PQL-based algorithm, we compute a p -value for each of the ten kernels using the Satterthwaite Method⁴⁹ based on score statistics, which follow a mixture of chi-square distributions. Afterwards, we combine these ten p -values through the recently developed Cauchy p -value combination rule²¹. To apply the Cauchy combination rule, we convert each of the ten p -values into a Cauchy statistic, aggregate the ten Cauchy statistics through summation, and convert the summation back to a single p -value based on the standard Cauchy distribution. The Cauchy rule takes advantage of the fact that combination of Cauchy random variables also follows a Cauchy distribution regardless whether these random variables are correlated or not^{21, 22}. Therefore, the Cauchy combination rule allows us to combine multiple potentially correlated p -values into a single p -value without loss of type I error control. After obtaining m p -values across m genes, we

control for false discovery rate (FDR) using the *Benjamini–Yekutieli* (BY) procedure, which is effective under arbitrary dependence across genes⁵⁰.

We refer to the above method as the Poisson version of SPARK (Spatial Pattern Recognition via Kernels) and is the main method used in the present study. Besides the Poisson version, we have also developed a Gaussian version of SPARK for modeling normalized spatial data (Supplementary Notes). Both versions of SPARK are implemented in the same R package with multiple threads computing capability, and with underlying efficient C/C++ code linked through Rcpp. The software SPARK, together with all analysis code used in the present study for reproducing the results presented in the manuscript, are freely available at www.xzlab.org/software.html.

Clustering SE Genes Detected by SPARK

We summarized the spatial expression patterns detected by SPARK by dividing SE genes into different categories. To do so, we first applied variance-stabilizing transformation (VST) to the raw count data¹² and obtained the relative gene expression levels through adjusting for the log-scale total read counts. We then used the hierarchical agglomerative clustering algorithm in the R package *amap* (v0.8–17) to cluster identified SE genes detected by SPARK into five groups. Afterwards, we summarized the gene expression patterns by using the expression level of the five cluster centers (Supplementary Figs. 3e, and f). In the hierarchical clustering, we set the two optional parameters in the R function to be Euclidean distance and Ward's criterion, respectively.

Gene Sets and Functional Enrichment Analysis

For each of the first two real data sets, we obtained lists of genes that can be used to serve as unbiased validation for the SE genes identified by different methods. Specifically, for the olfactory bulb data, we obtained a gene list directly based on the three layers (mitral, glomerular and granule) of the main olfactory bulb listed in the Harmonizome database (<https://amp.pharm.mssm.edu/Harmonizome/>). For the breast cancer data, we obtained from the Harmonizome database a gene list that consists of breast cancer related genes from six different data sets (OMIM Gene-Disease Associations; PhosphoSitePlus Phosphosite-Disease Associations; DISEASES Text-mining Gene-Disease Association Evidence Scores; GAD Gene-Disease Associations; GWAS Catalog SNP-Phenotype Associations). For the breast cancer data, we also obtained from the CancerMine database (<http://bionlp.bcgsc.ca/cancermine/>) another gene list that consists of breast cancer related genes that are either cancer drivers, oncogenes, or tumor suppressors. We used these gene lists to validate the SE genes identified by different methods.

We also performed the functional enrichment analysis of significant SE genes identified by SPARK and SpatialDE with Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. We performed all enrichment analyses using the R package *clusterProfiler*⁵¹ (v3.12.0). In the package, we used the default “BH” method for *p*-value multiple testing correction and set the default number of permutations to be 1,000.

Spatial Transcriptomics Data Sets

We downloaded two spatial transcriptomics data sets from the Spatial Transcriptomics Research (<http://www.spatialtranscriptomicsresearch.org>). These two data sets include a mouse olfactory bulb data and a human breast cancer data. These data consist of gene expression measurements in the form of read counts that are collected on a number of spatial locations known as spots. Following the SpatialDE paper, we used the *MOB Replicate 11* file for mouse olfactory bulb data, which contains 16,218 genes measured on 262 spots; and *Breast Cancer Layer 2* file for the breast cancer data, which contains 14,789 genes measured on 251 spots. We filtered out genes that are expressed in less than 10% of the array spots and selected spots with at least 10 total read counts. With these filtering criteria, we analyzed a final set of 11,274 genes on 260 spots in the mouse olfactory bulb data and 5,262 genes on 250 spots in the breast cancer data. In the analysis, we performed permutations to construct an empirical null distribution of p -values for each method by permuting the spot coordinates 10 times. Afterwards, we examined type I error control of different methods based on the empirical null distribution of p -values.

MERFISH Data Set

We obtained the MERFISH data set collected on the mouse preoptic region of the hypothalamus from Dryad^{27,52} (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248>). We used the slice at Bregma +0.11mm from animal 18 for analysis, as it contains all 160 genes measured on the largest number of single cells (5,665) across all nine cell classes. Among the 160 genes, 155 of them were pre-selected in the original study as either known markers for major cell classes or are relevant to various neuronal functions of the hypothalamus (e.g. some are neuropeptides and some are neuro-modulator receptors). Most of these 155 genes are expected to have spatial expression pattern in the hypothalamus. The remaining 5 genes are blank control genes without spatial expression pattern in the hypothalamus and thus can serve as negative controls. The downloaded data contains normalized gene expression values, which were previously computed as read counts divided by either the cell volume (combinatorial smFISH) or arbitrary fluorescence units per μm^3 (non-combinatorial, sequential FISH) and further scaled by 1,000. To obtain the raw count data, we rescaled the expression values by first multiplying 1,000, adjusted for cell volume, and then converted the rescaled value into integers by taking the ceiling over the rescaled data. After removing the ambiguous cells that were identified as putative doublets in the original data, we analyzed a final set of 160 genes on 4,975 cells. In the analysis, we permuted the location coordinates 100 times to construct an empirical null distribution, with which we examined type I error control of different methods.

SeqFISH Data Set

We obtained the seqFISH data set collected on the mouse hippocampus from the supplementary file of the original paper³¹ (<https://www.cell.com/cms/10.1016/j.neuron.2016.10.001/attachment/759be4dc-04a6-4a58-b6f6-9b52be2802db/mmc6.xlsx>). Following the SpatialDE paper, we extracted the field 43 data set for analysis. The data are in the form of raw count data for 249 genes measured in 257 cells with known spatial location information. Among 249 measured genes, 214 were selected from a list of

transcription factors and signaling pathway components, and the remaining 35 were selected from cell identity markers³¹. Following Trendsceek¹³ and the original study³¹, we filtered out cells with x- or y-axis values falling outside the range of 203 – 822 pixels in order to address border artifacts. After filtering, we analyzed a final set of 249 genes measured on 131 cells. In the analysis, we permuted the location coordinates 100 times to construct an empirical null distribution, with which we examined type I error control of different methods.

Compared Methods

We compared SPARK with three existing methods for detecting genes with spatial expression patterns. All these methods are designed for normalized data. The first method is Trendsceek (R package *trendsceek*; v1.0.0; download date: 12/20/2018). We followed the same procedure described in the original Trendsceek paper¹³ to filter and normalize count data. Specifically, for the two spatial transcriptomics data, we excluded genes that are expressed in less than 3 spots and excluded spots that contain less than 5 read counts. We then performed log10-transformation on raw count data (by adding a pseudo-count of one to void log transformation of zero values). For the real data analysis, we focused on analyzing the top 500 most variable genes to ensure sufficient power as well as computational feasibility as described in the Trendsceek paper. For the permuted data, we analyzed all the genes to construct an empirical null distribution. For seqFISH data, we first removed boundary cells as described in the previous section. Afterwards, following the Trendsceek recommendation, for each gene in turn, we performed a one-sided winsorization procedure to remove outlier effects by setting the first four largest values to be the fifth largest value. We then applied log10-transformation on the count data to obtain normalized expression values. For MERFISH data, we performed log10-transformation on raw count data and included all genes for analysis. Besides filtering and normalization, Trendsceek relies on permutation to compute *p*-values. Here, we set the number of permutations to be the default of 10,000. In addition, because the results of Trendsceek depend on the seeds used in the software, we analyzed each data using ten different seeds and reported results based on the seed that yields the highest number of discoveries; thus the power estimates of Trendsceek are likely upward biased. One disadvantage of Trendsceek is its slow computation: it takes over 48 hours to analyze one single gene in the mouse hypothalamus data. Therefore, in that data, we only applied the Trendsceek to the real data but not to the permuted data. Following the Trendsceek paper, we used the *Benjamini-Hochberg* procedure implemented in Trendsceek software to obtain adjusted *p*-value (i.e. FDR). With the adjusted *p*-value, we declared an SE gene significant if at least one of the four adjusted *p*-value outputs from (the four tests of) Trendsceek is below the threshold of 0.05.

The second method we compared with is SpatialDE (python package; v1.1.0; download date: 12/12/2018). For the mouse olfactory data and human breast cancer data, we directly used the analysis code provided by the SpatialDE authors on the Github (<https://github.com/Teichlab/SpatialDE>) to perform analysis. For the mouse hippocampus data, we applied their analysis code to the border artifacts adjusted data set described above to avoid detection of border artifacts and ensure fair comparison across methods. For the mouse hypothalamus data, we also directly applied the MERFISH analysis code described in the SpatialDE paper.

Following the SpatialDE paper, we declared an SE gene as significant if the output q -value (i.e. FDR) from SpatialDE is below the threshold of 0.05.

The last method is Moran's I test. We used the function *moran.test* implemented in the R package *spdep* (v1.1.2) for analysis. The results on Moran's I are presented only in the Discussion.

Data availability

This study made use of four publicly available data sets. These include the mouse olfactory bulb and human breast cancer data <http://www.spatialtranscriptomicsresearch.org>), the MERFISH data (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248>), the SeqFISH data (<https://www.cell.com/cms/10.1016/j.neuron.2016.10.001/attachment/759be4dc-04a6-4a58-b6f6-9b52be2802db/mmc6.xlsx>). In addition, all these raw data and processed data used for analysis are also made available at <https://github.com/xzhoulab/SPARK>.

Code availability

All source code used in our experiments have been deposited at www.xzlab.org/software.html.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported by the National Institutes of Health (NIH) Grants R01HG009124 and R01GM126553, and the National Science Foundation (NSF) Grant DMS1712933. S.S. was supported by NIH Grant R01HD088558 (PI Tung), the National Natural Science Foundation of China (Grant No. 61902319), and the Natural Science Foundation of Shaanxi Province (Grant No. 2019JQ127). J.Z. was supported by NIH Grant U01HL137182 (PI Kang).

REFERENCES

1. Chen KH, Boettiger AN, Moffitt JR, Wang S & Zhuang X RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348, aaa6090 (2015).
2. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M & Cai L Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* 11, 360–361 (2014). [PubMed: 24681720]
3. Femino AM, Fogarty K, Lifshitz LM, Carrington W & Singer RH Visualization of single molecules of mRNA in situ. *Method Enzymol* 361, 245–304 (2003).
4. Lovatt D et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nat Methods* 11, 190–196 (2014). [PubMed: 24412976]
5. Simone NL, Bonner RF, Gillespie JW, Emmert-Buck MR & Liotta LA Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends In Genetics* 14, 272–276 (1998). [PubMed: 9676529]
6. Junker JP et al. Genome-wide RNA Tomography in the Zebrafish Embryo. *Cell* 159, 662–675 (2014). [PubMed: 25417113]
7. Stahl PL et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82 (2016). [PubMed: 27365449]

8. Ke RQ et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* 10, 857–860 (2013). [PubMed: 23852452]
9. Lee JH et al. Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* 343, 1360–1363 (2014). [PubMed: 24578530]
10. Crosetto N, Bienko M & van Oudenaarden A Spatially resolved transcriptomics and beyond. *Nat Rev Genet* 16, 57–66 (2015). [PubMed: 25446315]
11. Fan X et al. Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res* 28, 730–745 (2018). [PubMed: 29867213]
12. Svensson V, Teichmann SA & Stegle O SpatialDE: identification of spatially variable genes. *Nat Methods* 15, 343–346 (2018). [PubMed: 29553579]
13. Edsgard D, Johnsson P & Sandberg R Identification of spatial expression trends in single-cell gene expression data. *Nat Methods* 15, 339–342 (2018). [PubMed: 29553578]
14. Lea AJ, Alberts SC, Tung J & Zhou X A flexible, efficient binomial mixed model for identifying differential DNA methylation in bisulfite sequencing data. *Plos Genet* 11, e1005650 (2015). [PubMed: 26599596]
15. Sun SQ et al. Differential expression analysis for RNAseq using Poisson mixed models. *Nucleic Acids Res* 45, e106 (2017). [PubMed: 28369632]
16. Lun A Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data. *BioRxiv*, 404962 (2019).
17. Li Y, Tang HC & Lin XH Spatial Linear Mixed Models with Covariate Measurement Errors. *Stat Sinica* 19, 1077–1093 (2009).
18. Ben-Ahmed K, Bouratbine A & El-Aroui MA Generalized linear spatial models in epidemiology: A case study of zoonotic cutaneous leishmaniasis in Tunisia. *J Appl Stat* 37, 159–170 (2010).
19. Breslow NE & Lin XH Bias Correction In Generalized Linear Mixed Models with a Single-Component Of Dispersion. *Biometrika* 82, 81–91 (1995).
20. Sun SQ et al. Heritability estimation and differential analysis of count data with generalized linear mixed models in genomic sequencing studies. *Bioinformatics* 35, 487–496 (2019). [PubMed: 30020412]
21. Liu YW et al. ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet* 104, 410–421 (2019). [PubMed: 30849328]
22. Pillai NS & Meng XL An Unexpected Encounter with Cauchy And Levy. *Ann Stat* 44, 2089–2097 (2016).
23. Tepe B et al. Single-Cell RNA-Seq of Mouse Olfactory Bulb Reveals Cellular Heterogeneity and Activity-Dependent Molecular Census of Adult-Born Neurons. *Cell Rep* 25, 2689–2703 (2018). [PubMed: 30517858]
24. Rouillard AD et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database-Oxford* 2016, baw100 (2016). [PubMed: 27374120]
25. Adan RAH et al. Rat Oxytocin Receptor In Brain, Pituitary, Mammary-Gland, And Uterus - Partial Sequence And Immunocytochemical Localization. *Endocrinology* 136, 4022–4028 (1995). [PubMed: 7649111]
26. Lever J, Zhao EY, Grewal J, Jones MR & Jones SJM CancerMine: a literature-mined resource for drivers, oncogenes and tumor suppressors in cancer. *Nat Methods* 16, 505–507 (2019). [PubMed: 31110280]
27. Moffitt JR et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362, eaau5324 (2018).
28. Fabio K et al. Synthesis and evaluation of potent and selective human V1a receptor antagonists as potential ligands for PET or SPECT imaging. *Bioorgan Med Chem* 20, 1337–1345 (2012).
29. Ozturk A, DeKosky ST & Kamboh MI Genetic variation in the choline acetyltransferase (CHAT) gene may be associated with the risk of Alzheimer's disease. *Neurobiol Aging* 27, 1440–1444 (2006). [PubMed: 16223550]
30. Kiaris H, Schally AV & Kalofoutis A Extrapituitary effects of the growth hormone-releasing hormone. *Vitam Horm* 70, 1–24 (2005). [PubMed: 15727800]

31. 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). [PubMed: 27764670]
32. Tasic B et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19, 335–346 (2016). [PubMed: 26727548]
33. Li XH, Polter A & Yang S FoxO transcription factors - Regulation in brain and behavioral manifestation. *Biol Psychiat* 63, 150–159 (2008).
34. Hoekman MFM, Jacobs FMJ, Smidt MP & Burbach JPH Spatial and temporal expression of FoxO transcription factors in the developing and adult murine brain. *Gene Expr Patterns* 6, 134–140 (2006). [PubMed: 16326148]
35. Cattaneo A et al. FoxO1, A2M, and TGF-beta 1: three novel genes predicting depression in gene X environment interactions are identified using cross-species and cross-tissues transcriptomic and miRNomic analyses. *Mol Psychiatr* 23, 2192–2208 (2018).
36. Shrestha BR et al. Sensory Neuron Diversity in the Inner Ear Is Shaped by Activity. *Cell* 174, 1229–1246 (2018). [PubMed: 30078709]
37. Sun YF et al. A central role for Islet1 in sensory neuron development linking sensory and spinal gene regulatory programs. *Nat Neurosci* 11, 1283–1293 (2008). [PubMed: 18849985]
38. Wang X et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361, eaat5691 (2018).
39. Voss S, Zimmermann B & Zimmermann A Detecting spatial structures in throughfall data: The effect of extent, sample size, sampling design, and variogram estimation method. *J Hydrol* 540, 527–537 (2016).
40. Lark RM, Heuvelink GBM & Bishop TFA Burgess TM & Webster R. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *Journal of Soil Science*, 31, 315–331. *Eur J Soil Sci* 70, 7–10 (2019).
41. Li HF, Calder CA & Cressie N Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Geogr Anal* 39, 357–375 (2007).
42. Radeloff VC, Miller TF, He HS & Mladenoff DJ Periodicity in spatial data and geostatistical models: autocorrelation between patches. *Ecography* 23, 81–91 (2000).
43. Rodriques SG et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363, 1463–1467 (2019). [PubMed: 30923225]
44. Diggle PJ, Tawn JA & Moyeed RA Model-based geostatistics. *J R Stat Soc C-Appl* 47, 299–326 (1998).
45. Christensen OF & Waagepetersen R Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics* 58, 280–286 (2002). [PubMed: 12071400]
46. Rousset F & Ferdy JB Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography* 37, 781–790 (2014).
47. Vanhatalo J, Pietilainen V & Vehtari A Approximate inference for disease mapping with sparse Gaussian processes. *Statistics in medicine* 29, 1580–1607 (2010). [PubMed: 20552572]
48. Lin XH & Breslow NE Bias correction in generalized linear mixed models with multiple components of dispersion. *J Am Stat Assoc* 91, 1007–1016 (1996).
49. Satterthwaite FE An Approximate Distribution Of Estimates Of Variance Components. *Biometrics Bull* 2, 110–114 (1946).
50. Benjamini Y & Yekutieli D The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29, 1165–1188 (2001).
51. Yu GC, Wang LG, Han YY & He QY clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics* 16, 284–287 (2012). [PubMed: 22455463]
52. Moffitt JR et al. in *Dryad Digital Repository*, Vol. 362 (ed. Repository DD) eaau5324 (2018).

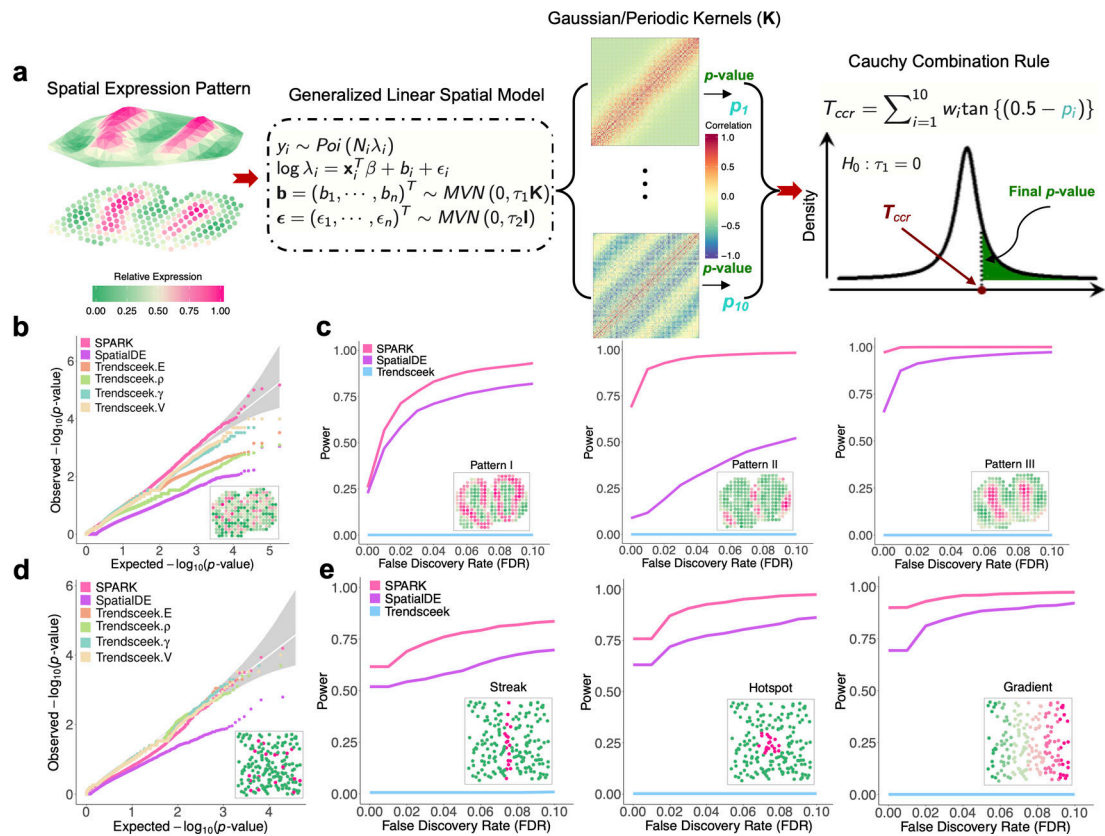


Figure 1: Method schematic of SPARK and simulation results.

(a) Method schematic of SPARK. SPARK examines one gene at a time and models the gene expression measurements on spatial locations using the generalized linear spatial model. To detect whether the gene shows spatial expression pattern, SPARK relies on a series of spatial kernels for pattern recognition and outputs a p -value for each spatial kernel using the Satterthwaite method that enables exact p -value computation. All these p -values from different spatial kernels are subsequently combined into a final SPARK p -value through the Cauchy combination rule. Here, w_i is a weight for the combination (set to be 1/10 throughout the text) and $\tan(\cdot)$ denotes the tangent function. **(b)** Quantile-quantile plot of the observed $-\log_{10} p$ -values from different methods against the expected $-\log_{10} p$ -values under the null for the first set of null simulations based on the mouse olfactory bulb data ($n = 260$ array spots). The p -values are combined across ten simulation replicates, with each replicate containing 10,000 simulated genes. Simulations are performed under moderate noise ($\tau_2 = 0.35$). Compared methods include SPARK (pink), SpatialIDE (purple), Trendsceek.E (light salmon) which is the Emark test of Trendsceek, Trendsceek. ρ (yellow-green) which is the Markcorr test of Trendsceek, Trendsceek. γ (light green) which is the Markvario test of Trendsceek, and Trendsceek.V (wheat) which is the Vmark test of Trendsceek. Representative expression pattern for a null gene that does not show a spatial expression pattern is embedded inside the panel. **(c)** Power plots show the proportion of true positives (y-axis) detected by different methods at a range of false discovery rates (FDR; x-axis) for the first set of alternative simulations based on the mouse olfactory bulb data ($n = 260$ array spots). Representative genes displaying each of the three spatial expression patterns I-III are

embedded inside the panels. The proportion of true positives is averaged across ten simulation replicates, with each replicate containing 1,000 SE genes and 9,000 non-SE genes. Simulations are performed under moderate noise ($\tau_2 = 0.35$) and moderate SE strength (threefold). Trendsceek (sky-blue) is the combined test of Trendsceek. **(d)** Quantile-quantile plot of the observed $-\log_{10} p$ -values from different methods against the expected $-\log_{10} p$ -values under the null for the second set of null simulations based on the seqFISH data. p -values are combined across ten simulation replicates, with each replicate containing 1,000 simulated genes. Simulations are performed under moderate sample size ($n = 200$ cells). Representative expression pattern for a null gene that does not show a spatial expression pattern is embedded inside the panel. **(e)** Power plots show the proportion of true positives (y-axis) detected by different methods at a range of false discovery rates (FDR; x-axis) for the second set of alternative simulations based on the seqFISH data with moderate sample size ($n = 200$ cells). Representative genes displaying each of the three spatial expression patterns are embedded inside the panels. The proportion of true positives is averaged across ten simulation replicates, with each replicate containing 100 SE genes and 900 non-SE genes. Simulations were performed under moderate fraction of marked cells (20%) and moderate SE strength (2-fold) for the hotspot and streak patterns, or under moderate SE strength (40% cells displaying expression gradient) for the linear gradient pattern.

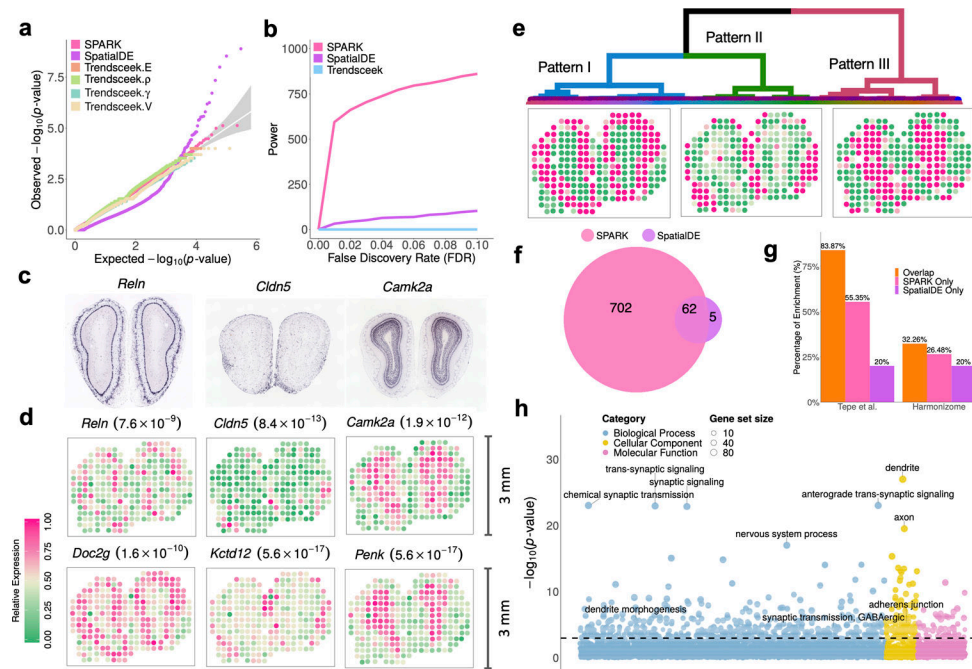


Figure 2: Analyzing the mouse olfactory bulb data ($n = 260$ spots).

(a) Quantile-quantile plot of the observed $-\log_{10} p$ -values from different methods are plotted against the expected $-\log_{10} p$ -values under the null in the permuted data. p -values are combined across ten permutation replicates. Compared methods include SPARK (pink), SpatialDE (purple), Trendsceek.E (light salmon) which is the Emark test of Trendsceek, Trendsceek. ρ (yellow-green) which is the Markcorr test of Trendsceek, Trendsceek. γ (light green) which is the Markvario test of Trendsceek, and Trendsceek.V (wheat) which is the Vmark test of Trendsceek. (b) Power plot shows the number of genes with spatial expression pattern (y-axis) identified by different methods at a range of false discovery rates (x-axis). Trendsceek (sky-blue) which is the combined test of Trendsceek, detected almost none. (c) *In situ* hybridization of three representative genes (*Reln*, *Cldn5*, and *Camk2a*) obtained from the database of the Allen Brain Atlas. *Reln* is spatially expressed in the mitral layer and glomeruli layer. *Cldn5* is spatially expressed in the nerve layer. *Camk2a* is spatially expressed in the granular layer. (d) Spatial expression pattern for representative genes in the spatial transcriptomics data. Top row shows the same three genes as shown in (c) along with their p -values from SPARK (inside parenthesis). These genes are only identified by SPARK, but not by the other two methods. Bottom row shows spatial expression patterns for three additional known marker genes (*Doc2g*, *Kctd12*, and *Penk*) for different layers in the mouse olfactory bulb: *Doc2g* for mitral layer; *Kctd12* for nerve layer; and *Penk* for granular layer. Color represents relative gene expression level (purple: high; green: low). (e) Three distinct spatial expression patterns summarized based on the 772 SE genes identified by SPARK, along with dendrogram displaying the clustering of these three main patterns (pattern I: 119 genes; pattern II: 270 genes; pattern III: 321 genes). (f) Venn diagram shows the overlap between SE genes identified by SPARK and SpatialDE based on an FDR cutoff 0.05. (g) Bar plot shows the percentage of SE genes identified by SPARK (orange/pink) or SpatialDE (orange/purple) that are also validated in two gene lists, one from a literature (left) and the

other from the Harmonizome database (right). The orange bar represents the percentage of SE genes identified by both SPARK and SpatialDE that are in either of the gene lists; the pink bar represents the percentage of unique SE genes identified by SPARK that are in either of the gene lists; the purple bar represents the percentage of unique SE genes identified by SpatialDE that are in either of the gene lists. **(h)** Bubble plot shows $-\log_{10} p$ -values for pathway enrichment analysis on 772 SE genes obtained by SPARK based on an FDR cutoff of 0.05. The p -values are from the *clusterProfiler* analysis and the dash line indicates a p -value cutoff of 0.05. Gene sets are colored by three categories: GO biological process (blue), GO molecular function (purple), and GO cellular component (yellow).

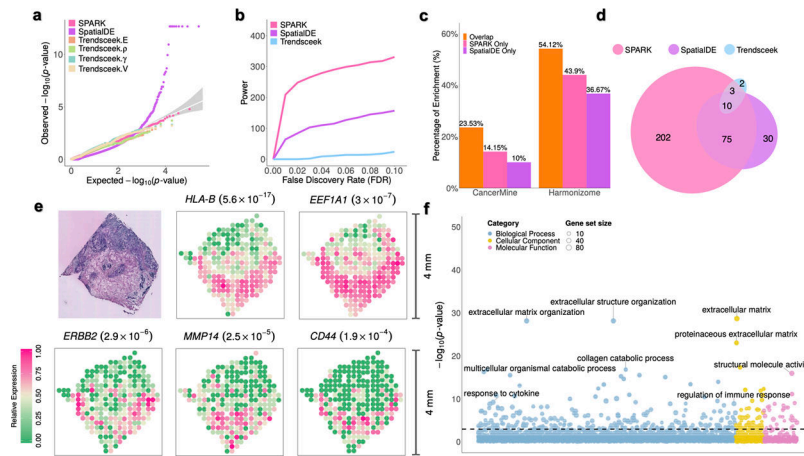


Figure 3: Analyzing the human breast cancer data ($n = 250$ spots)

(a) Quantile-quantile plot of the observed $-\log_{10} p$ -values from different methods are plotted against the expected $-\log_{10} p$ -values under the null in the permuted data. p -values are combined across ten permutation replicates. Compared methods include SPARK (pink), SpatialDE (purple), Trendsceek.E (light salmon) which is the Emark test of Trendsceek, Trendsceek. ρ (yellow-green) which is the Markcorr test of Trendsceek, Trendsceek. γ (light green) which is the Markvario test of Trendsceek, and Trendsceek.V (wheat) which is the Vmark test of Trendsceek. (b) Power plot shows the number of genes with spatial expression pattern (y-axis) identified by different methods at a range of false discovery rates. Trendsceek (sky-blue) which is the combined test of Trendsceek, detected only a few. (c) Bar plot shows the percentage of SE genes identified by SPARK (orange/pink) or SpatialDE (orange/purple) that are also validated in two gene lists, one from the CancerMine database (left) and the other from the Harmonizome database (right). The orange bar represents the percentage of SE genes identified by both SPARK and SpatialDE that are in either of the gene lists; the pink bar represents the percentage of unique SE genes identified by SPARK that are in either of the gene lists; the purple bar represents the percentage of unique SE genes identified by SpatialDE that are in either of the gene lists. (d) Venn diagram shows the overlap between SE genes identified by SPARK and SpatialDE based on an FDR cutoff 0.05. (e) Spatial expression pattern for five genes (*HLA-B*, *EEF1A1*, *ERBB2*, *MMP14*, and *CD44*) that are only identified by SPARK but not by the other two methods. The p -values for the five genes from SPARK are shown inside parenthesis. Color represents relative gene expression level (purple: high; green: low). For reference, the hematoxylin and eosin (H&E) staining of a breast cancer biopsy from ref.7 is shown in the top left panel. The dark staining in the H&E panel represents potential tumors. These five genes are previously known molecular markers associated with tumor induced immune response (*HLA-B*), growth factor (*ERBB2*), or metastasis (*EEF1A1*, *MMP14* and *CD44*). Panel e is reproduced with permission from ref.7. (f) Bubble plot shows $-\log_{10} p$ -values for pathway enrichment analysis on 290 SE genes obtained by SPARK based on an FDR cutoff of 0.05. The p -values are from the *clusterProfiler* analysis and the dash line indicates a p -value cutoff of 0.05. The dash line indicates a p -value cutoff of 0.05. Gene sets are colored by categories: GO biological process (blue), GO molecular function (purple), and GO cellular component (yellow).

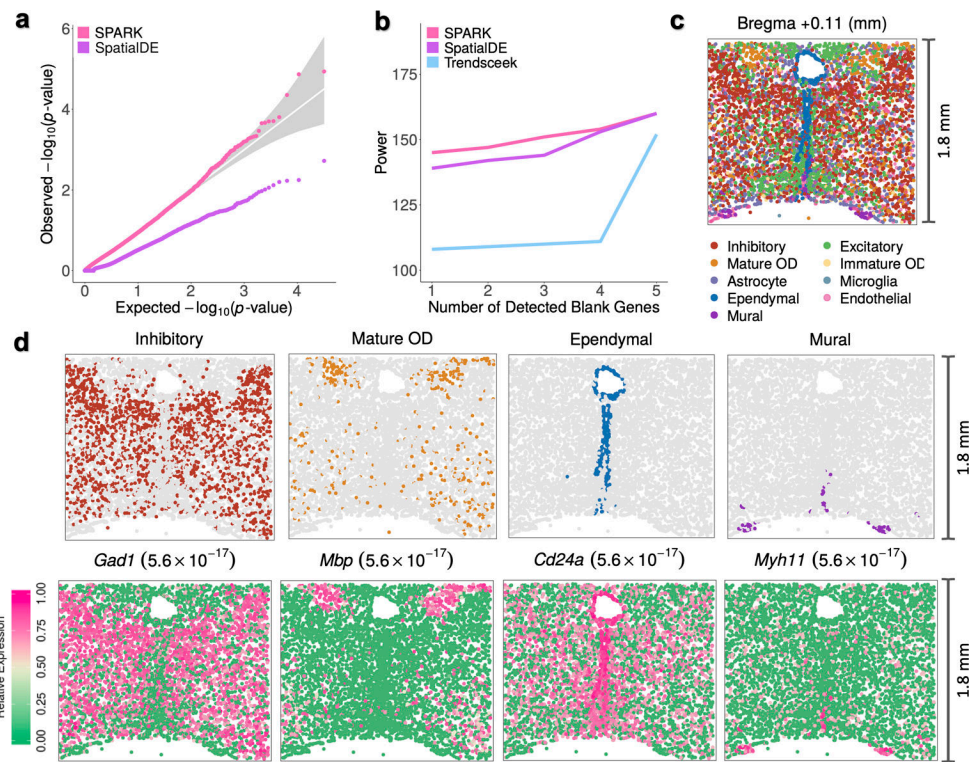


Figure 4: Analyzing the mouse hypothalamus data ($n = 4,975$ cells).

(a) Quantile-quantile plot of the observed $-\log_{10} p$ -values from different methods are plotted against the expected $-\log_{10} p$ -values under the null in the permuted data. The p -values are combined across one hundred permutation replicates. Compared methods include SPARK (pink) and SpatialDE (purple). Results for Trendsceek (sky-blue) are not included here due to its heavy computational burden. (b) Power plot shows the number of genes with spatial expression pattern (y-axis) identified by different methods vs the number of blank control genes identified at the same threshold. (c) Spatial distribution of all major cell classes on the 1.8-mm by 1.8-mm imaged slice from a single female mouse. Cells are colored by cell classes shown in the legend, where the cell class information are obtained from the reference. (d) Spatial distribution of four main cell classes. The spatial distribution of the remaining five cell classes are shown in Supplementary Fig. 8. The cell classes are represented by colored dots while all other background cells are shown as gray dots. Spatial expression pattern for four representative genes (*Gad1*, *Mbp*, *Cd24a*, and *Myh11*) that are identified by SPARK are shown. The p -values for the four genes from SPARK are shown inside parenthesis. Color represents relative gene expression level (purple: high; green: low).