Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies

Shiquan Sun [1,2,4], **Jiaqiang Zhu** [2,4] **and Xiang Zhou** [2,3*]

Identifying genes that display spatial expression patterns in spatially resolved transcriptomic studies is an important first step toward characterizing the spatial transcriptomic landscape of complex tissues. Here we present a statistical method, SPARK, for identifying spatial expression patterns of genes in data generated from various spatially resolved transcriptomic techniques. SPARK directly models spatial count data through generalized linear spatial models. It relies on recently developed statistical formulas for hypothesis testing, providing effective control of type I errors and yielding high statistical power. With a computationally efficient algorithm, which is based on penalized quasi-likelihood, SPARK is also scalable to datasets with tens of thousands of genes measured on tens of thousands of samples. Analyzing four published spatially resolved transcriptomic datasets using SPARK, we show it can be up to ten times more powerful than existing methods and disclose biological discoveries that otherwise cannot be revealed by existing approaches.

The recent emergence of various spatially resolved transcriptomic technologies has enabled gene-expression profiling with spatial localization information on tissues or cell cultures. Such techniques include MERFISH[1] and seqFISH[2], which are based on single-molecule fluorescence in situ hybridization (smFISH)[3] and can measure hundreds of genes with subcellular spatial resolution; TIVA[4], LCM[5], Tomo-Seq[6] and spatial transcriptomics through spatial barcoding[7], which are based on next-generation DNA sequencing and can measure tens of thousands of genes on single cells or on spatial locations consisting of a couple of hundred single cells; and targeted in situ sequencing (ISS)[8] and FISSEQ[9], which are based on in situ RNA sequencing and can measure up to thousands of transcripts with spatial information at a single-cell resolution. These different spatially resolved transcriptomic techniques have made it possible to study the spatial organization of the 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 patterns, which we simply refer to as SE analysis, is an important first step toward 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, SpatialDE[12] and Trendsceek[13], 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 power[16]. Besides direct modeling of count data, identifying SE genes also requires the development of statistical methods that can produce well-calibrated *P* values for control of type I errors. However, some existing methods for SE analysis, such as SpatialDE[12], rely on asymptotic normality and minimal *P*-value-combination rules for constructing hypothesis tests. Consequently, these methods may fail to control for type I errors at the 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 computationally efficient linear mixed models, some other spatial methods, in particular Trendsceek[13], are built without a data-generative model and compute non-parametric test statistics through computationally expensive permutation strategies. Consequently, analyzing even moderately sized spatial transcriptomics datasets with hundreds of genes across hundreds of spatial locations can be a daunting task.

Here we present spatial pattern recognition via kernels (SPARK), a method that addresses the statistical and computational challenges described above. It 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 χ^2 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 the Methods (technical details are provided in the 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 between these methods is provided in the Supplementary

[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi, P. R. China. [2]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [3]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [4]These authors contributed equally: Shiquan Sun, Jiaqiang Zhu. *e-mail: xzhousph@umich.edu

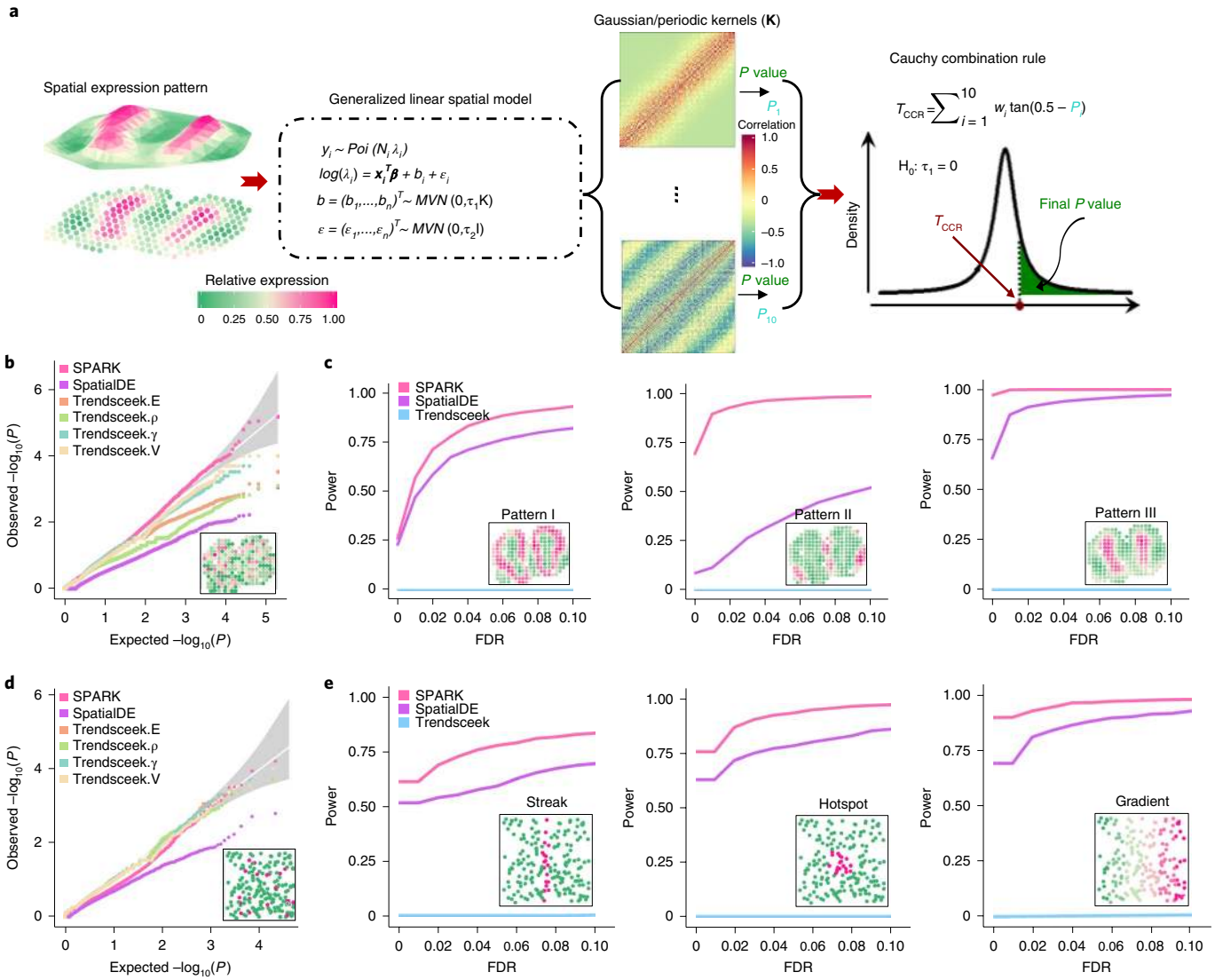

Fig. 1 | Method schematic of SPARK and simulation results. a, Schematic of the SPARK method. w_i is a weight set to be 1/10 here and tan(·) denotes the tangent function. **b**, Quantile–quantile plot of the observed $-\log_{10}(P)$ from different methods against the expected $-\log_{10}(P)$ under the null condition for the first set of null simulations. *P* values were combined across ten simulation replicates. Simulations were performed under moderate noise ($\tau_2 = 0.35$). Compared methods include SPARK (pink), SpatialDE (purple), Trendsceek.E (light salmon; Emark test of Trendsceek), Trendsceek.ρ (yellow-green; the Markcorr test of Trendsceek), Trendsceek.γ (light green; the Markvario test of Trendsceek) and Trendsceek.V (wheat; the Vmark test of Trendsceek). A representative expression pattern for a null gene is embedded inside the panel. **c**, Power plots show the proportion of true positives (*y* axis) detected by different methods at a range of FDRs (*x* axis) for the first set of alternative simulations. Representative genes displaying each of the three spatial expression patterns I–III are embedded inside the panels. The proportion of true positives was averaged across ten simulation replicates. Simulations were 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 $-\log_{10}(P)$ from different methods against the expected $-\log_{10}(P)$ under the null condition for the second set of null simulations. Simulations were performed under moderate sample size ($n = 200$ cells). A representative expression pattern for a null gene is embedded inside the panel. **e**, Power plots show the proportion of true positives (*y* axis) detected by different methods at a range of FDRs (*x* axis) for the second set of alternative simulations. Representative genes displaying each of the three spatial expression patterns are embedded inside the panels. Simulations were performed under a moderate fraction of marked cells (20%) and moderate SE strength (twofold) for the hotspot and streak patterns or under moderate SE strength (40% cells displaying expression gradient) for the linear gradient pattern.

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 the Supplementary Notes.

In the first set of simulations, we found that, under the null condition, SPARK produced well-calibrated *P* values at the transcriptome-wide significance levels (Fig. 1b). Some Trendsceek test statistics (for example, Markvario and Vmark) also produced reasonably calibrated *P* values, while others (for example, Emark statistics and markcorr statistics) yielded slightly conservative *P* values. By contrast, SpatialDE produced overly conservative *P* values (Fig. 1b). The failure of SpatialDE in controlling type I errors is presumably due to its use of an asymptotic χ^2 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 condition 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 errors, we measured power on the basis of 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,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 that were fully based on the original Trendsceek paper[13]. The comparison results on the second set of simulations were largely consistent with the results obtained from the first set of simulations. Specifically, under the null condition, both SPARK and Trendsceek produced well-calibrated *P* values, while SpatialDE did not (Fig. 1d). Under the alternative simulations, 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 study[13]. 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 datasets, including two datasets obtained by spatial transcriptomics sequencing and two datasets obtained by smFISH (Methods). The first dataset was from the mouse olfactory bulb[7], consisting of 11,274 genes measured on 260 spots. Consistent with simulations, both SPARK and Trendsceek produced calibrated *P* values under the permuted null condition, while SpatialDE did not (Fig. 2a). SPARK also identified more SE genes as 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 approximately tenfold more than that detected by SpatialDE, which identified 67 (among which 62 overlapped with SPARK; Fig. 2b,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 tended to have expression levels close to zero (Supplementary Fig. 3b) and appeared to locate on either one or two spots (Supplementary Fig. 3d), suggesting potentially false signals. By 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 (Fig. 2e and 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 patterns were 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 study[7]. The list of highlighted marker genes, which 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 eight of ten such highlighted genes. SpatialDE only detected three 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 RNA sequencing study in the olfactory bulb[23]. 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 database[24]. 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 support for the higher power of SPARK.

Finally, we performed functional enrichment analyses of SE genes identified by SPARK and SpatialDE (Methods). A total of 1,023 Gene Ontology (GO) terms (Fig. 2h) and 79 Kyoto Encyclopedia of Genes and Genomes (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 two KEGG pathways (overlap = 2; Supplementary Fig. 5b) were enriched in the SE genes identified by SpatialDE (Supplementary Table 1 and Supplementary Fig. 5c). Many enriched GO terms or KEGG pathways identified only by SPARK are directly related to synaptic organization and olfactory bulb development. Examples include olfactory lobe development (GO 0021988; SPARK $P = 5.81 \times 10^{-3}$, SpatialDE $P = 1.21 \times 10^{-1}$) and oxytocin signaling pathway (KEGG mmu04921; SPARK $P = 1.59 \times 10^{-9}$, SpatialDE $P = 2.15 \times 10^{-1}$) for modulating olfactory processing[25]. An in-depth enrichment analysis using SE genes in patterns I–III separately provided additional biological insights, revealing the critical role of synaptic organization for the mitral cell layer, the importance of cell junction and synaptic connectivity for the nerve layer, as well as the importance of dendritic morphogenesis and synaptic–dendritic plasticity for the granular layer (Supplementary Results, Supplementary Fig. 5 and Supplementary Table 1). Overall, the newly identified GO term and KEGG pathway enrichment highlight the benefits of running SE analysis with SPARK.

Breast cancer data. The second dataset was a study of human breast cancer biopsies[7], which contained 5,262 genes measured on 250 spots. Again, both SPARK and Trendsceek produced calibrated *P* values under the permuted null condition, while SpatialDE did not (Fig. 3a). SPARK also identified more SE genes as compared to SpatialDE and Trendsceek across a range of FDRs (Fig. 3b and Supplementary Fig. 6a). For example, at an FDR of 5%, SPARK identified 290 SE genes, which is approximately threefold more than that detected by SpatialDE, which identified 115 (among which 85 overlapped with SPARK; Fig. 3b,d). By 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 patterns, 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-related genes highlighted in the original study[7]. SPARK detected ten of them, SpatialDE detected seven and Trendsceek detected two (Supplementary Fig. 6e). Both SpatialDE and Trendsceek missed three well-known cancer-related 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 database[26]. Fourteen percent of SE genes uniquely identified by SPARK were in the list while only 10% identified by SpatialDE were in the list

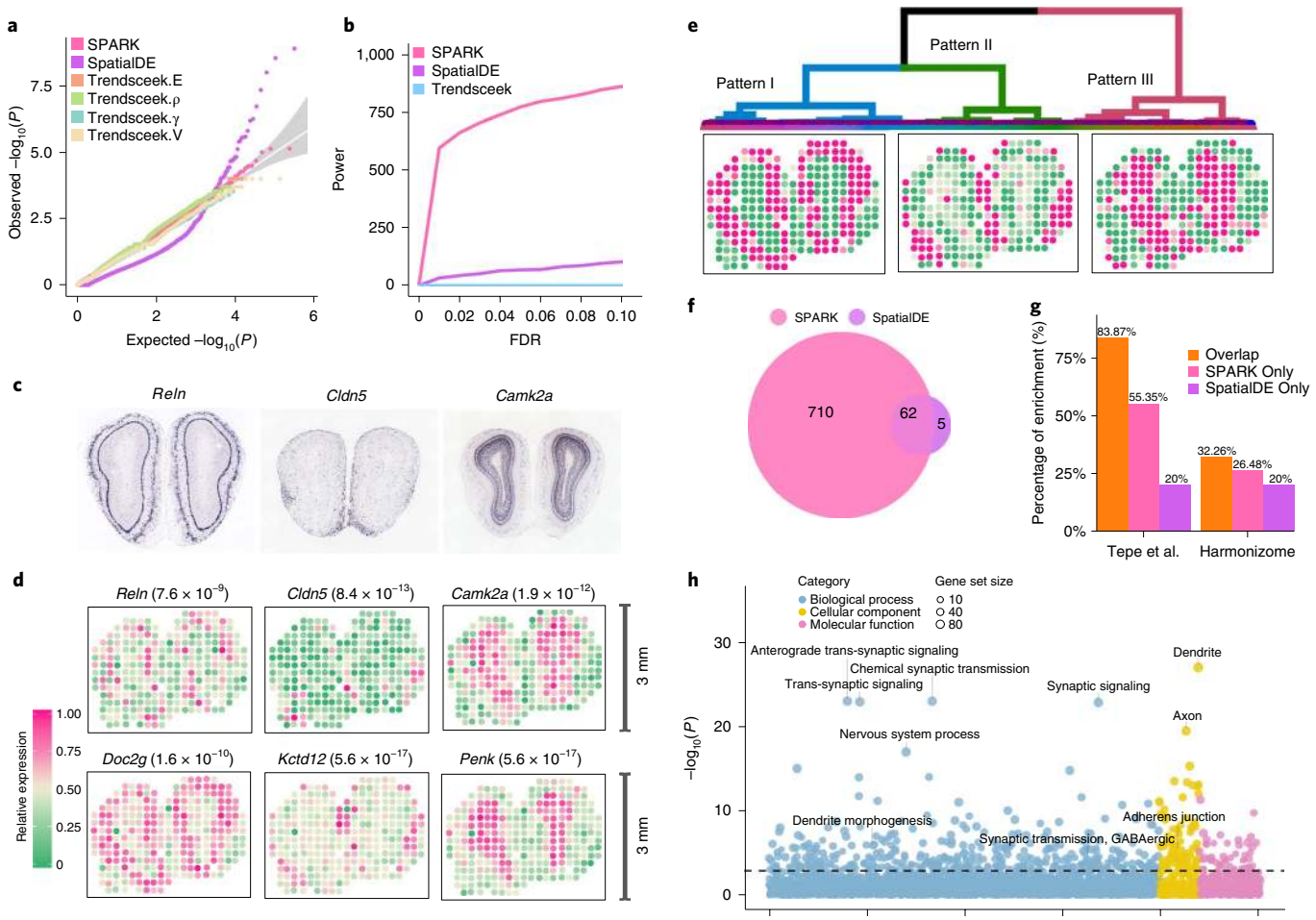

Fig. 2 | Analyzing the mouse olfactory bulb dataset ($n = 260$ spots). a, Quantile–quantile plot of the observed $-\log_{10}(P)$ from different methods against the expected $-\log_{10}(P)$ under the null condition in the permuted data. *P* values were combined across ten permutation replicates. Compared methods include SPARK (pink), SpatialDE (purple), Trendsceek.E (light salmon), Trendsceek.ρ (yellow-green), Trendsceek.γ (light green) and Trendsceek.V (wheat). **b**, Power plot shows the number of genes with spatial expression pattern (*y* axis) identified by different methods at a range of FDRs (*x* axis). Trendsceek (sky-blue) is the combined test of Trendsceek. **c**, In situ hybridization of three representative genes (*Reln*, *Cldn5* and *Camk2a*) obtained from the Allen Brain Atlas. **d**, Spatial expression pattern for representative genes. Top, the same three genes as **c** along with their *P* values from SPARK (in parentheses). Bottom, spatial expression patterns for three additional known marker genes (*Doc2g*, *Kctd12* and *Penk*) for different layers in the mouse olfactory bulb. Color represents relative gene-expression levels (purple, high; green, low). **e**, Three distinct spatial expression patterns summarized on the basis of 772 SE genes identified by SPARK, along with a dendrogram displaying the clustering of the three main patterns. **f**, Venn diagram shows the overlap between SE genes identified by SPARK and SpatialDE on the basis of an FDR cutoff of 0.05. **g**, Bar plot of the percentage of SE genes identified by SPARK (pink) or SpatialDE (purple) or both (orange) that were also validated in two gene lists: one from the literature (left; Tepe et al.[23]) and the other from the Harmonizome database (right). **h**, Bubble plot of $-\log_{10}(P)$ for pathway enrichment analysis on 772 SE genes obtained by SPARK on the basis of an FDR cutoff of 0.05. The dashed line represents 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).

(Fig. 3c). For example, the well-known proto-oncogene *ERBB2* gene that has a vast amount of previous literature support was only identified by SPARK (Fig. 3e). Third, we collected a list of 3,538 genes that are relevant to breast cancer from the Harmonizome database[24]. 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 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 and Supplementary Table 2) while SpatialDE identified 266 GO terms (overlap = 191) and three KEGG pathways (overlap = 3; Supplementary Figs. 6c,d and Supplementary Table 2). Many enriched gene sets discovered only by SPARK are related to extracellular matrix organization and immune responses (Fig. 3e, Supplementary Results, Supplementary Fig. 6f and Supplementary Table 2), highlighting their importance in cancer development and metastasis.

Hypothalamus data. The third dataset is a MERFISH dataset collected on the preoptic area of the mouse hypothalamus[27]. The data contains 160 genes measured on 4,975 single cells (Fig. 4c). Of these genes 155 of 160 were selected in the original study as they are markers of distinct cell populations or are relevant to various neuronal functions of the hypothalamus. Besides these likely true-positive genes, another five 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 the permuted null condition, while SpatialDE did not (Fig. 4a; Trendsceek was not applied to the permuted null condition owing to its heavy

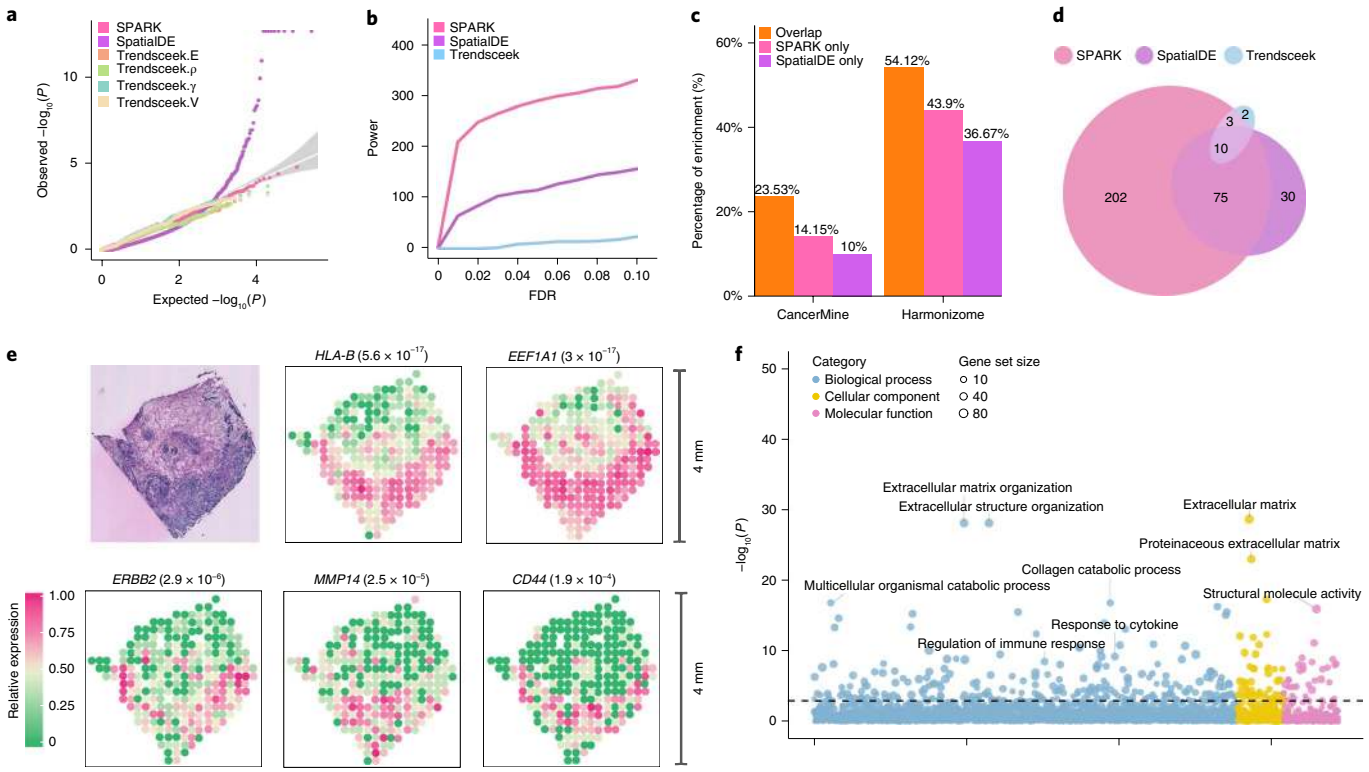

Fig. 3 | Analyzing human breast cancer data ($n = 250$ spots). a, Quantile–quantile plot of the observed $-\log_{10}(P)$ from different methods are plotted against the expected $-\log_{10}(P)$ under the null condition in the permuted data. *P* values were combined across ten permutation replicates. Compared methods include SPARK (pink), SpatialDE (purple), Trendsceek.E (light salmon), Trendsceek.ρ (yellow-green), Trendsceek.γ (light green), Trendsceek.V (wheat). **b**, Power plot shows the number of genes with spatial expression patterns (*y* axis) identified by different methods at a range of FDRs. Trendsceek (sky-blue) is the combined test of Trendsceek. **c**, Bar plot of the percentage of SE genes identified by SPARK (pink), SpatialDE (purple) or both (orange) that are also validated in two gene lists: one from the CancerMine database (left) and the other from the Harmonizome database (right). **d**, Venn diagram of the overlap between SE genes identified by SPARK and SpatialDE on the basis of an FDR cutoff 0.05. **e**, Spatial expression pattern for five known tumor 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 parentheses. Color represents relative gene-expression levels (purple, high; green, low). The hematoxylin and eosin staining is shown in the top left, with dark staining representing the potential tumor region. Reproduced from ref. [7] with permission. **f**, Bubble plot shows $-\log_{10}(P)$ for pathway enrichment analysis on 290 SE genes obtained by SPARK on the basis of an FDR cutoff of 0.05. Dashed line represents 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).

computational burden). The quantile–quantile 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 five negative control genes and 155 likely positive genes, we directly compared the power of different methods on the basis of 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 (that is, one false positive), SPARK identified 145 SE genes, which is six more than the number detected by SpatialDE, which identified 139 (among which 138 overlapped with SPARK; Fig. 4b and Supplementary Fig. 8b). Both SPARK and SpatialDE outperformed Trendsceek, which identified 108 SE genes (among which 103 overlapped with SPARK). A careful examination suggests that almost all SE genes identified by SPARK show clear spatial expression patterns as one would expect: nine major cell classes in hypothalamus (Fig. 4d and Supplementary Fig. 8c) along with nine marker genes[27] (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] (Supplementary Results), highlighting the power of SPARK.

Hippocampus data. The final dataset was a small seqFISH dataset with 249 genes measured on 131 single cells in the mouse hippocampus[31] (Supplementary Fig. 9a). These 249 genes include 214 genes that were selected in the original study[31] as transcription factors and signaling pathway components and 35 remaining genes that were previously known markers of cell identity. In the analysis, we found that both SPARK and Trendsceek produced calibrated *P* values under the permuted null condition while SpatialDE yielded conservative *P* values (Supplementary Fig. 9b). SPARK again identified more SE genes as compared to SpatialDE and Trendsceek across a range of FDRs (Supplementary Figs. 9c,d). For example, at an FDR of 5%, SPARK identified 17 SE genes. SpatialDE and Trendsceek identified 11 (all overlap with SPARK) and four (one overlap with SPARK) SE genes, respectively (Supplementary Fig. 9e). The 11 SE genes identified by both SpatialDE and SPARK showed clear spatial expression patterns (Supplementary Fig. 9f), as did the six SE genes identified only by SPARK (Supplementary Fig. 9g). The three SE genes only detected by Trendsceek tended to express uniformly in most cells and show less apparent spatial patterns (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

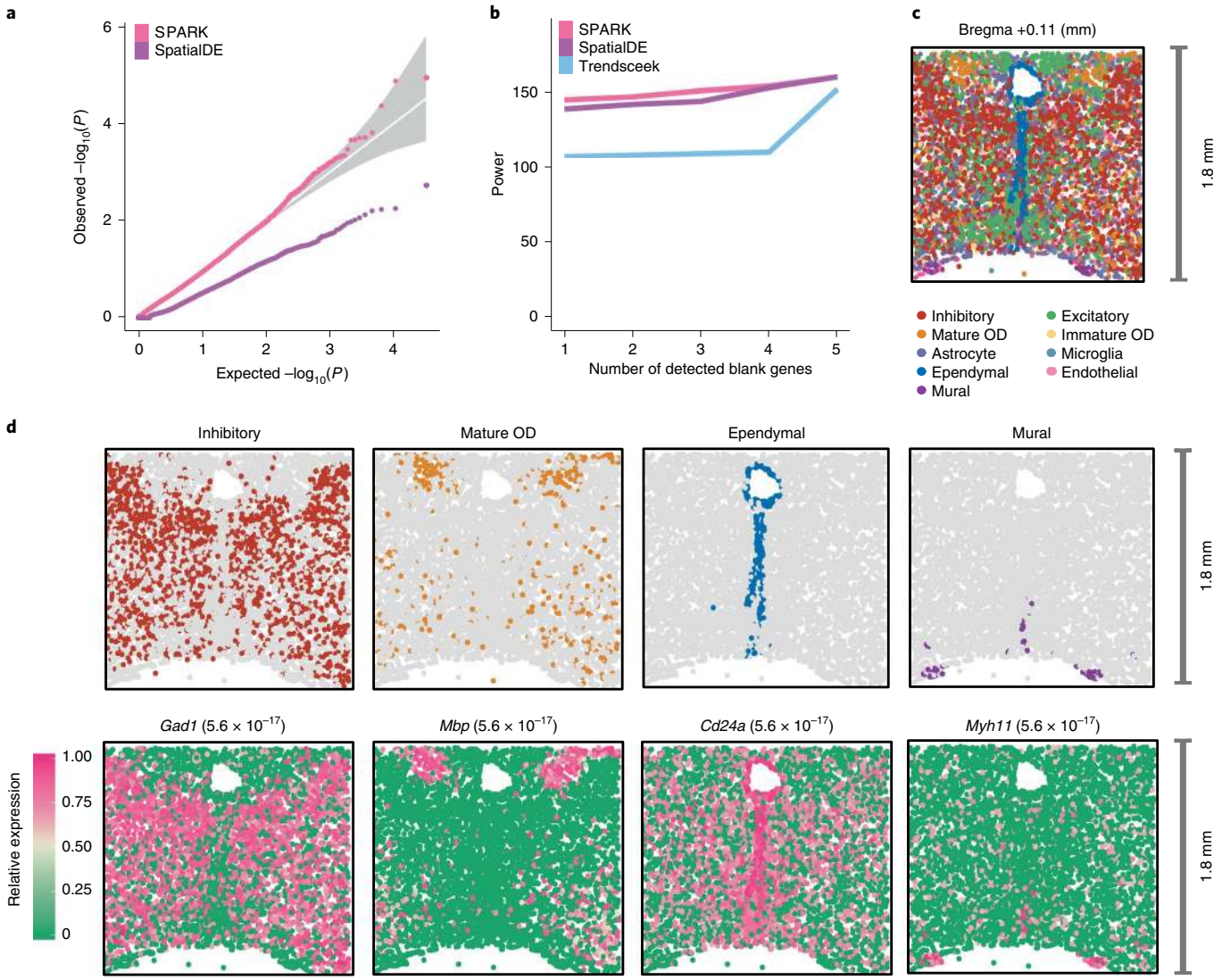

Fig. 4 | Analyzing the mouse hypothalamus data ($n = 4{,}975$ cells). a, Quantile–quantile plot of the observed $-\log_{10}(P)$ from different methods against the expected $-\log_{10}(P)$ under the null condition in the permuted data. The *P* values were combined across 100 permutation replicates. Compared methods include SPARK (pink) and SpatialDE (purple). Results for Trendsceek (sky-blue) are not included here owing to its heavy computational burden. **b**, Power plot shows the number of genes with spatial expression pattern (*y* axis) identified by different methods versus the number of blank control genes identified at the same threshold. **c**, Spatial distribution of all major cell classes on the imaged slice (1.8 mm × 1.8 mm) from a single female mouse. Cells are colored by cell classes shown in the legend, where the cell class information were obtained from the reference. **d**, Top, 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. Bottom, spatial expression pattern for four representative genes (*Gad1*, *Mbp*, *Cd24a* and *Myh11*) that were identified by SPARK are shown, with *P* values from SPARK shown inside parentheses. Color represents relative gene-expression level (purple, high; green, low).

are markers of cell identity: *Foxo1* and *Slc17a8* for glutamatergic neurons; *Igtp* for GABAergic neurons; and *Opalin* for oligodendrocytes[32]. All of them are closely related to neuronal functions in the 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 after cortisol exposure to prenatal stress and mediates the negative effect of stress on neurogenesis[35]. Besides these four marker genes, the remaining two genes, *Pou4f1* and *Gfi1*, encode neural transcription factors and play important roles in the development of the sensory nervous system[36,37]. These important genes, which are missed by other methods, again highlight the power of SPARK.

Discussion

We present a new method, SPARK, for identifying SE genes in spatially resolved transcriptomic studies. In comparison to existing approaches, SPARK produces well-calibrated *P* values and yields more statistical power. SPARK is also easily applicable to three-dimensional datasets such as STARmap[38] or even higher-dimensional datasets where other coordinates (for example, time) are recorded.

SPARK incorporates a data-generative model and relies on a model-based hypothesis-testing framework for detection of spatial patterns. The data-generative model in SPARK distinguishes it from previous spatial data-exploration tools that rely on variograms or semivariograms 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] when 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 datasets examined here. We found that the *P* values from Moran's I under the permuted null condition were highly inflated, presumably owing 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–d), likely owing 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 a power gain is especially apparent in data with low counts, such as the first two spatial transcriptomics datasets we examined here. However, we acknowledge that the power gain brought by count modeling may be small in datasets with high counts (for example, MERFISH and seqFISH data), as a normal distribution can often approximate high counts as well as an overdispersed Poisson distribution. Therefore, we developed a Gaussian version of SPARK (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 to 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 and Supplementary Fig. 12), the Gaussian version of SPARK may represent an attractive alternative for analyzing large datasets collected from emerging techniques such as Slide-seq[43].

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 datasets (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 of 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.

References

1. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
2. 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).
3. Femino, A. M., Fogarty, K., Lifshitz, L. M., Carrington, W. & Singer, R. H. 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).
5. Simone, N. L., Bonner, R. F., Gillespie, J. W., Emmert-Buck, M. R. & Liotta, L. A. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet.* **14**, 272–276 (1998).
6. Junker, J. P. et al. Genome-wide RNA tomography in the zebrafish embryo. *Cell* **159**, 662–675 (2014).
7. Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
8. Ke, R. Q. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
9. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
10. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet* **16**, 57–66 (2015).
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).
12. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).
13. Edsgard, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* **15**, 339–342 (2018).
14. Lea, A. J., Alberts, S. C., Tung, J. & Zhou, X. A flexible, efficient binomial mixed model for identifying differential DNA methylation in bisulfite sequencing data. *PloS Genet.* **11**, e1005650 (2015).
15. Sun, S. Q. et al. Differential expression analysis for RNAseq using Poisson mixed models. *Nucleic Acids Res.* **45**, e106 (2017).
16. Lun, A. Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data. Preprint at *BioRxiv* https://doi.org/10.1101/404962 (2019).
17. Li, Y., Tang, H. C. & Lin, X. H. Spatial linear mixed models with covariate measurement errors. *Stat. Sin.* **19**, 1077–1093 (2009).
18. Ben-Ahmed, K., Bouratbine, A. & El-Aroui, M. A. Generalized linear spatial models in epidemiology: a case study of zoonotic cutaneous leishmaniasis in Tunisia. *J. Appl. Stat.* **37**, 159–170 (2010).
19. Breslow, N. E. & Lin, X. H. Bias correction in generalized linear mixed models with a single-component of dispersion. *Biometrika* **82**, 81–91 (1995).
20. Sun, S. Q. et al. Heritability estimation and differential analysis of count data with generalized linear mixed models in genomic sequencing studies. *Bioinformatics* **35**, 487–496 (2019).
21. Liu, Y. W. 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).
22. Pillai, N. S. & Meng, X. L. 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).
24. Rouillard, A. D. et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, baw100 (2016).
25. Adan, R. A. H. et al. Rat oxytocin receptor in brain, pituitary, mammary-gland, and uterus—partial sequence and immunocytochemical localization. *Endocrinology* **136**, 4022–4028 (1995).
26. Lever, J., Zhao, E. Y., Grewal, J., Jones, M. R. & Jones, S. J. M. CancerMine: a literature-mined resource for drivers, oncogenes and tumor suppressors in cancer. *Nat. Methods* **16**, 505–507 (2019).
27. Moffitt, J. R. 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, S. T. & Kamboh, M. I. Genetic variation in the choline acetyltransferase (CHAT) gene may be associated with the risk of Alzheimer's disease. *Neurobiol. Aging* **27**, 1440–1444 (2006).
30. Kiaris, H., Schally, A. V. & Kalofoutis, A. Extrapituitary effects of the growth hormone-releasing hormone. *Vitam. Horm.* **70**, 1–24 (2005).
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).
32. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
33. Li, X. H., Polter, A. & Yang, S. FoxO transcription factors—regulation in brain and behavioral manifestation. *Biol. Psychiat.* **63**, 150–159 (2008).
34. Hoekman, M. F. M., Jacobs, F. M. J., Smidt, M. P. & Burbach, J. P. H. Spatial and temporal expression of FoxO transcription factors in the developing and adult murine brain. *Gene Expr. Patterns* **6**, 134–140 (2006).

35. Cattaneo, A. et al. *FoxO1*, *A2M*, and TGF-β 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, B. R. et al. Sensory neuron diversity in the inner ear is shaped by activity. *Cell* **174**, 1229–1246 (2018).
37. Sun, Y. F. et al. A central role for Islet1 in sensory neuron development linking sensory and spinal gene regulatory programs. *Nat. Neurosci.* **11**, 1283–1293 (2008).
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, R. M., Heuvelink, G. B. M., Bishop, T. F. A., Burgess, T. M. & Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *Eur. J. Soil Sci.* **31**, 315–331 (2019).
41. Li, H. F., Calder, C. A. & Cressie, N. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geogr. Anal.* **39**, 357–375 (2007).
42. Radeloff, V. C., Miller, T. F., He, H. S. & Mladenoff, D. J. Periodicity in spatial data and geostatistical models: autocorrelation between patches. *Ecography* **23**, 81–91 (2000).
43. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

Methods

SPARK: model and algorithm. We aimed to model 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 expression is measured on single cells that are randomly scattered across the tissue or culture space) or are predetermined by researchers before the experiment (in the case of spatial transcriptomics technology). We denote $\boldsymbol{s}_i=(s_{i1},s_{i2})$ as the spatial coordinates that is, location index) for ith sample, with $i\in(1,\ldots,n)$. These spatial coordinates vary continuously over a two-dimensional space R^2, or $\boldsymbol{s}_i\in R^2$. While we only focus on the case where samples are collected on a two-dimensional space of a tissue or culture layout, our model and method are general, and therefore are capable of handling three-dimensional cases where the depth of the sample location in the tissue can be recorded or handling cases with even higher dimensions where other coordinates (for example, time) are also recorded.

Our primary goal is to detect genes whose expression levels display spatial patterns 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, unlike these previous GLSM development, we focused on developing a hypothesis-testing framework for GLSM. Here, for the gene of focus, we denote $y_i(\boldsymbol{s}_i)$ as the gene-expression measurement in terms of counts for the ith sample. We denote $\boldsymbol{x}_i(\boldsymbol{s}_i)$ as a k vector of covariates that include a scalar of one for the intercept and $k-1$ observed explanatory variables for the ith 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 $N_i(\boldsymbol{s}_i)$ as the normalization factor for the ith sample. Here we set $N_i(\boldsymbol{s}_i)$ as the summation of the total number of counts across all genes for the sample as our main interest is analyzing the relative gene-expression level. Other choices of $N_i(\boldsymbol{s}_i)$ are possible, for example, $N_i(\boldsymbol{s}_i)$ 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 overdispersed Poisson distribution

$$y_i(\mathbf{s}_i) \sim \mathrm{Poi}(N_i(\mathbf{s}_i)\lambda_i(\mathbf{s}_i)), i = 1, 2, \cdots, n$$

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

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

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 $b_i(\boldsymbol{s}_i)$ is a zero-mean, stationary Gaussian process modeling the spatial correlation pattern among spatial locations

$$\mathbf{b}(\mathbf{s}) = (b_1(\mathbf{s}_1), b_2(\mathbf{s}_2), \cdots, b_n(\mathbf{s}_n))^T \sim \mathrm{MVN}(0, \tau_1\mathbf{K}(\mathbf{s})),$$

where the covariance $\boldsymbol{K}(\mathbf{s})$ is a kernel function of the spatial locations $\boldsymbol{s}=(\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n)^T$, with the ijth element being $\boldsymbol{K}(\boldsymbol{s}_i,\boldsymbol{s}_j)$; 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 detail below. In the above model, the covariance for the latent variables $\log(\boldsymbol{\lambda}(\boldsymbol{s}))$ is $\boldsymbol{\Sigma} = \tau_1\mathbf{K}(\mathbf{s}) + \tau_2\mathbf{I}$, where $\boldsymbol{\lambda}(\mathbf{s}) = (\lambda_1(\mathbf{s}_1), \lambda_2(\mathbf{s}_2), \cdots, \lambda_n(\mathbf{s}_n))^T$ and $\boldsymbol{I}$ 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 $\log(\lambda_i(\boldsymbol{s}_i))$ captured by spatial patterns or spatial location information; τ_2 is commonly referred to as the nugget, which effectively measures the expression variance in $\log(\lambda_i(\boldsymbol{s}_i))$ owing to random noise independent of spatial locations.

In the GLSM defined above, testing whether a gene shows a spatial expression pattern can be translated into testing the null hypothesis 'H_0: $\tau_1=0$'. The statistical power of such a hypothesis test will inevitably depend on how well the spatial kernel function $\boldsymbol{K}(\boldsymbol{s})$ 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 datasets (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 implementation can easily handle many other kernel functions or incorporate a 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 that is based on the 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 method[49] on the basis of score statistics, which follow a mixture of χ^2 distributions. Afterward, we combine these ten P values through the recently developed Cauchy P value combination rule[21]. 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 on the basis of the standard Cauchy distribution. The Cauchy rule takes advantage of the fact that a combination of Cauchy random variables also follows a Cauchy distribution regardless of 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 losing control over type I errors. After obtaining m P values across m genes, we control for FDR using the Benjamini–Yekutieli procedure, which is effective under arbitrary dependence across genes[50].

We refer to the above method as the Poisson version of SPARK, and it 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 and are capable of computing over multiple threads, with underlying efficient C and 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 to the raw count data[12] 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 (v.0.8–17) to cluster identified SE genes detected by SPARK into five groups. Afterward, we summarized the gene-expression patterns by using the expression level of the five cluster centers (Supplementary Figs. 3e,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 datasets, 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 that is 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 a gene list from the Harmonizome database that consists of breast-cancer-related genes from six different datasets (OMIM gene–disease associations, PhosphoSitePlus phosphosite–disease associations, DISEASES text-mining gene–disease association evidence scores, GAD gene–disease associations and GWAS catalog SNP–phenotype associations). For the breast cancer data, we also obtained another gene list from the CancerMine database (http://bionlp.bcgsc.ca/cancermine/) that consists of genes related to breast cancer 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 GO terms and KEGG pathways. We performed all enrichment analyses using the R package clusterProfiler[51] (v.3.12.0). 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 datasets. We downloaded two spatial transcriptomics datasets from Spatial Transcriptomics Research (http://www.spatialtranscriptomicsresearch.org). These two datasets include mouse olfactory bulb data and 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 the '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 ten 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 ten times. Afterward, we examined control of type I errors by the different methods on the basis of the empirical null distribution of *P* values.

MERFISH dataset. We obtained the MERFISH dataset 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 + 0.11 mm 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 preselected in the original study as either known markers for major cell classes or are relevant to various neuronal functions of the hypothalamus (for example, some are neuropeptides and some are neuromodulator receptors). Most of these 155 genes are expected to have spatial expression pattern in the hypothalamus. The remaining five genes are blank control genes without a spatial expression pattern in the hypothalamus and thus can serve as negative controls. The downloaded data contain 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 by 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 control of type I errors by the different methods.

SeqFISH dataset. We obtained the seqFISH dataset collected on the mouse hippocampus from the supplementary file of the original paper[31] (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 dataset 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 markers[31]. Following Trendsceek[13] and the original study[31], we filtered out cells with *x*- or *y*-axis values falling outside the range of 203–822 pixels 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 control of type I errors by the different methods.

Comparison of 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; v.1.0.0; downloaded on 20 December 2018). We followed the same procedure described in the original Trendsceek paper[13] to filter and normalize count data. Specifically, for the two spatial transcriptomics data, we excluded genes that were expressed in less than three spots and excluded spots that contained less than five read counts. We then performed $\log_{10}$-transformation on raw count data (by adding a pseudocount of one to avoid 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. Afterward, 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 $\log_{10}$-transformation on the count data to obtain normalized expression values. For MERFISH data, we performed $\log_{10}$-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 dataset using ten different seeds and reported results on the basis of the seed that yielded the highest number of discoveries, thus, the power estimates of Trendsceek are likely upwardly biased. One disadvantage of Trendsceek is its slow computation: it takes over 48 h 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 an adjusted *P* value (that is, 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 was below the threshold of 0.05.

The second method we used for comparison was SpatialDE (Python package; v.1.1.0; downloaded on 12 December 2018). For the mouse olfactory data and human breast cancer data, we directly used the analysis code provided by the SpatialDE authors on Github (https://github.com/Teichlab/SpatialDE) to perform analysis. For the mouse hippocampus data, we applied their analysis code to the border-artifacts-adjusted dataset 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 (that is, FDR) from SpatialDE was below the threshold of 0.05.

The last method used for comparison was Moran's *I* test. We used the function moran.test implemented in the R package spdep (v.1.1.2) for analysis. The results on Moran's *I* are presented only in the Discussion.

Data availability

This study made use of four publicly available datasets. 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) and 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 raw data and processed data used for analysis are also available at https://github.com/xzhoulab/SPARK.

Code availability

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

References

44. Diggle, P. J., Tawn, J. A. & Moyeed, R. A. Model-based geostatistics. *J. R. Stat. Soc. Ser. C Appl. Stat.* **47**, 299–326 (1998).
45. Christensen, O. F. & Waagepetersen, R. Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics* **58**, 280–286 (2002).
46. Rousset, F. & Ferdy, J. B. 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. *Stat. Med.* **29**, 1580–1607 (2010).
48. Lin, X. H. & Breslow, N. E. Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Am. Stat. Assoc.* **91**, 1007–1016 (1996).
49. Satterthwaite, F. E. 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, G. C., Wang, L. G., Han, Y. Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
52. Moffitt, J. R. et al. Data from: molecular, spatial and functional single-cell profiling of the hypothalamic preoptic region. *Dryad Digital Repository* https://doi.org/10.5061/dryad.8t8s248 (2018).

Acknowledgements

This study was supported by National Institutes of Health (NIH) grants R01HG009124 and R01GM126553, and National Science Foundation (NSF) grant DMS1712933. S.S. was supported by NIH grant R01HD088558 (PI Tung), the National Natural Science Foundation of China (grant 61902319) and the Natural Science Foundation of Shaanxi Province (grant 2019JQ127). J.Z. was supported by NIH grant U01HL137182 (PI Kang).

Author contributions

X.Z. conceived the idea and provided funding support. S.S. and X.Z. designed the experiments. S.S. and J.Z. developed the method, implemented the software, performed simulations and analyzed real data. S.S., J.Z. and X.Z. wrote the manuscript.

Competing interests

The authors declare that they have no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.Z.