

Dimension-agnostic and granularity-based spatially variable gene identification

Juexin Wang^{1,2,4,*,#}, Jinpu Li^{3,4,*}, Skyler T Kramer^{3,4}, Li Su^{3,4}, Yuzhou Chang^{5,6}, Chunhui Xu^{3,4}, Qin Ma^{5,6,#}, Dong Xu^{2,3,4,#}

1. Department of BioHealth Informatics, Luddy School of Informatics, Computing, and Engineering, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202, USA
2. Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA
3. MU Institute for Data Science and Informatics, University of Missouri, Columbia, MO 65211, USA
4. Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA
5. Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH 43210, USA
6. Pelotonia Institute for Immuno-Oncology, The James Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA,

* These Authors Contributed Equally

Corresponding Authors

Juexin Wang wangjuex@iu.edu

Qin Ma Qin.Ma@osumc.edu

Dong Xu xudong@missouri.edu

Abstract

Identifying spatially variable genes (SVGs) is critical in linking molecular cell functions with tissue phenotypes. Spatially resolved transcriptomics captures cellular-level gene expression with corresponding spatial coordinates in two or three dimensions and can be used to infer SVGs effectively. However, current computational methods may not achieve reliable results and often cannot handle three-dimensional spatial transcriptomic data. Here we introduce BSP (big-small patch), a spatial granularity-guided and non-parametric model to identify SVGs from two or three-dimensional spatial transcriptomics data in a fast and robust manner. This new method has been extensively tested in simulations, demonstrating superior accuracy, robustness, and high efficiency. BSP is further validated by substantiated biological discoveries in cancer, neural science, rheumatoid arthritis, and kidney studies with various types of spatial transcriptomics technologies.

Keywords

Spatially variable genes, three-dimensional spatial transcriptomics, non-parametric statistical model, granularity

Background

Spatially resolved transcriptomics (SRT) have been rapidly developed and widely used in biological and biomedical research over the past decade¹⁻³. Single-molecule fluorescence *in situ* hybridization (smFISH) (e.g., MERFISH and SeqFISH+) and sequencing-based approaches (e.g., 10X Visium)² are popular SRT technologies on sliced two-dimensional (2D) samples. A shift has recently occurred towards retaining three-dimensional (3D) positional anatomy at cellular resolution. Wang *et al.* developed STARmap, which combined an efficient sequencing approach with hydrogel-tissue chemistry for 3D intact tissue RNA sequencing⁴, with a throughput of up to

1,000 genes or 10,000 genes⁵. Vickovic *et al.* developed protocols on consecutive sections to get 3D spatial profiling of rheumatoid arthritis (RA) synovia⁶. Existing works on 3D imaging data construction and 3D imaging data have significant advantages over 2D data for accurate quantitative interpretation⁷. In contrast to the 2D spatial transcriptomics approach, which depends on sampling strategy (e.g., coronal or sagittal) on sliced samples, the 3D spatial transcriptomics provides a more comprehensive and faithful representation of intact organ structures and functions. It overcomes the inherent 2D bias and enables the visualization of gene expression in relation to the tissue architecture in three dimensions. Such 3D views provide new opportunities in the identification of cell types and states, discovery of new biomarkers, and drug design⁸. However, most of the current analytic methods are developed and validated on 2D SRT data, and cannot be directly applied to diverse types of 3D analyses⁹.

Spatially variable genes (SVGs) are biologically significant as they exhibit variations in expression levels across different regions or cell types within a tissue, indicating their involvement in specific biological processes or functions unique to those regions or cell types. Hence, the inference of SVGs can help researchers gain a deeper understanding of how different cell types and genes contribute to the overall structures and functions of tissues in normal or disease states¹⁰. Additionally, SVGs can be used as molecular markers to track developmental or disease-related changes in the spatial distribution of specific cell types. Identification of SVGs also facilitates the dissection of biological relationships between spatial organization and molecular cell function, providing critical information for biologists and pathologists. For example, in the mouse olfactory bulb, Stahl *et al.* discovered functional regions in the mouse brain by identifying SVGs¹¹, while Maynard *et al.* discovered laminar and nonlaminar genes in the human dorsolateral prefrontal cortex¹². The existing SRT technologies encode the key clues of SVGs, whose expressions rely on the spatial locations of cells^{13,14}, and identifying SVGs from tens of thousands of genes is often a critical first step in analyzing spatial transcriptome data.

Compared to traditional RNA-seq and scRNA-seq analyses that identify differentially expressed (DE) genes, SVGs in SRT data incorporate gene expression information and corresponding spatial context in geometric coordinates. In 2D SRT studies, SpatialDE¹³ and Trendsceek¹⁴ are the first two computational methods for identifying SVGs: SpatialDE utilizes Gaussian process regression to quantify the spatial variance of expression for each gene, while Trendsceek selects SVGs by testing the dependence between the expression and the spatial location for each gene using a permutation process. Afterward, a generalized linear spatial model with Gaussian/periodic kernels, SPARK¹⁵, was proposed to capture the spatial patterns and filter SVGs using the combined p -values from each kernel. A simplified version, SPARK-X¹⁶, was later introduced to reduce computational time and memory usage. In addition to statistical methods, MERINGUE⁹ applied a Voronoi tessellation method to build an adjacency matrix and calculate classical Moran's I score¹⁷ for each gene based on the constructed adjacent matrix to infer SVGs. SpaGCN¹⁸ first identifies the spatial domain with graph neural networks and then employs a statistical test to identify SVGs based on the context of the inferred spatial domains.

Although some preliminary analysis^{4,6} has been conducted on emerging 3D SRT data, significant challenges remain in identifying SVGs in the dimension-agnostic SRT data^{9,16}. The limited spatial information captured by 2D tissue slices may result in incomplete and biased representations of spatial characteristics, potentially leading to inaccurate biological conclusions^{8,19}. Additionally, the existing SVG identification methods require user-defined parameters that can vary across samples and lead to disparate findings that are difficult to justify without prior knowledge of the samples. Hence, a non-parameter method with adequate power is preferred even for 2D data in practical usage. Based on our preliminary analysis, the expression distribution of SVGs tends to exhibit a consistent and specific pattern invariant across different spatial resolutions and views, whereas the expression distribution of non-SVGs has a random pattern with varying

characteristics across different views and resolutions. These distributions can be effectively captured by granularity, a concept relatively underexplored in spatial transcriptomics studies. Granularity refers to the extent or hierarchical level to which a material or system comprises distinguishable pieces²⁰. We propose that the concept of granularity can be leveraged to identify SVGs in a dimension-agnostic geometric manner. With appropriate quantitative measures, granularity-based criteria can distinguish between spatial organizations with biological significance and those with random patterns.

Here we introduce BSP (big-small patch), a spatial granularity-guided and non-parametric model that enables efficient and robust identification of SVGs from two/three-dimensional SRT data. For each spot in the data, BSP selects a set of neighboring spots within a certain distance to capture the regional means with different granularities. The variances of the expression mean across all spots are then calculated under different scales, and genes with high ratios are identified as the SVGs. One of the unique features of BSP is that it does not make any assumption regarding the distribution of the gene expression levels or the spatial pattern of the spots. The model is robust to fluorescence *in situ* hybridization (MERFISH, seqFISH+, and STARMap) and sequencing-based (10X Visium and slide-seq) SRT without requiring pre-defined or well-tuned parameters. Compared with existing methods, BSP outperforms other methods for 3D data, and delivers comparable power and accuracy as current methods for 2D data with a significantly reduced computational cost. In addition, the BSP algorithm is easily implementable, making it versatile and easily integrated into various applications. In our experiments on kidney SRT data and 3D RA synovia study, BSP identified several functional-related SVGs related to the disease mechanisms. In summary, BSP is an accurate, fast, robust, and parameter-free method for identifying SVGs in 2D and 3D SRT data.

2. Results

2.1 The big-small-patch algorithm

The proposed BSP algorithm is a granularity-guided approach for identifying SVGs in dimension-agnostic SRT data (**Figure 1**). BSP defines a patch for each spot in the SRT data, which includes all neighboring spots within a given radius centered on the spot (**Figure 1A**). A pair of patches is then defined, consisting of a small patch with a smaller radius and a large patch with a larger radius (**Figure 1B**). This paired big-small patch captures the ambient local expression characteristics in different granularities, delineating spatial patterns in various contexts. Subsequently, the transcriptomic expression variance of the local means is calculated across all pairs of patches, and the ratio between the variance with a big patch and the variance with a small patch is used as the statistic score for each gene. This statistic score can be used to quantify the conservation of SVGs' spatial patterns in different granularities. The distribution of this statistical score is fitted with a beta distribution, and genes with statistical significance ($p < 0.05$) in the fitted beta distribution are defined as SVGs (**Figure 1C**).

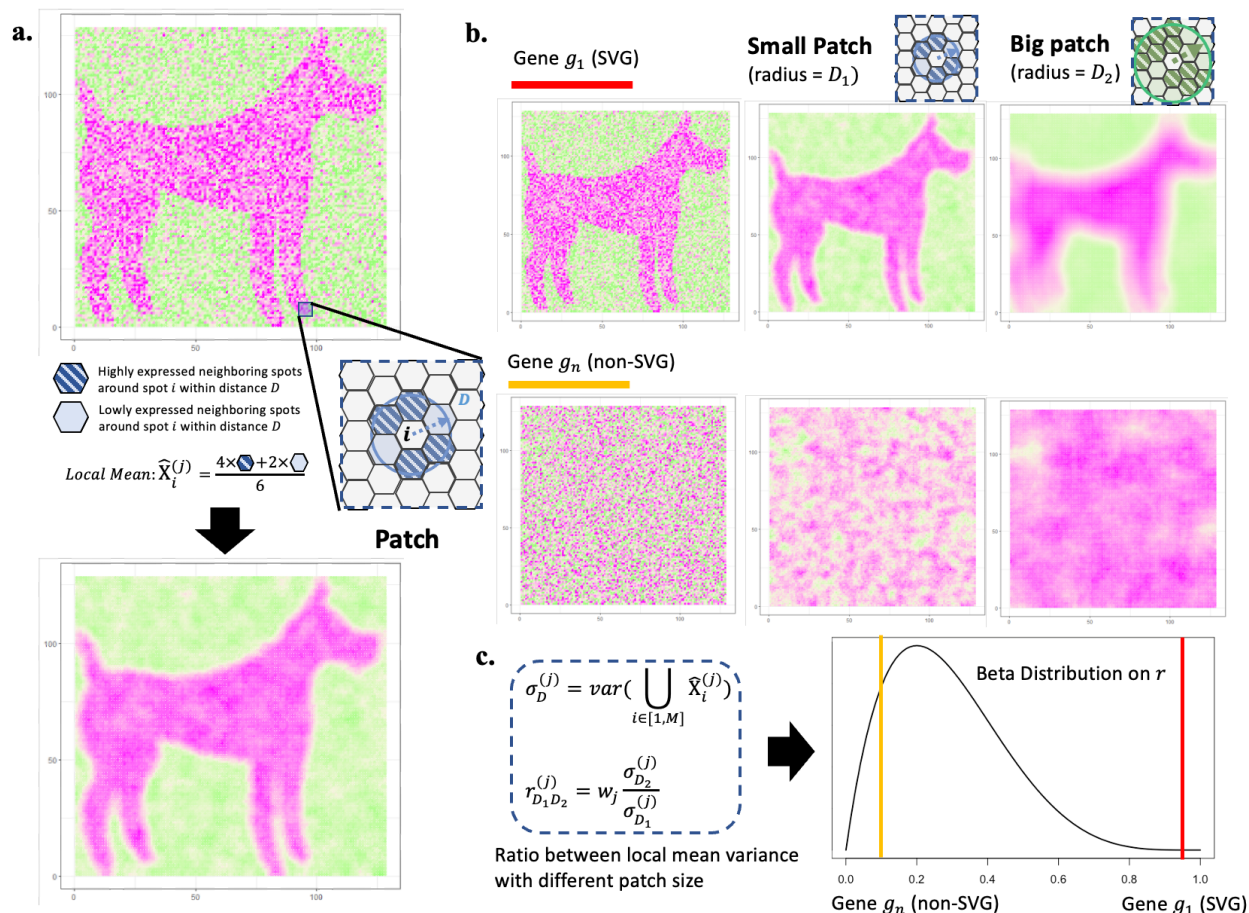


Figure 1. Scheme of BSP. **a)** Definition of the patch in BSP. For spot i , a patch is defined as the set of all its neighboring spots within distance D . Local Mean $\hat{X}_i^{(j)}$ is defined as the average expression of gene j for all spots within the patch. **b)** Identification of SVGs with BSP. Gene g_1 indicates an SVG (red), while g_n is a representative non-SVG gene (orange). Besides the original pattern as the left, we define a small patch of spots with a smaller radius D_1 (blue) in the middle column, and a big batch of spots with a larger radius D_2 (green) in the right column. **c)** The ratio $r_{D_1 D_2}^{(j)}$ of the variance of the local mean between big batch $\sigma_{D_2}^{(j)}$ and the paired small batch $\sigma_{D_1}^{(j)}$ is chosen as the statistical value r . Gene g_1 (red bar) is statistically significant compared with the background gene g_n (orange bar) on the fitted beta distribution of r .

2.2 BSP can accurately and efficiently identify SVGs in 2D simulations

To demonstrate the effectiveness of the BSP model in analyzing 2D transcriptomic data, we generated a set of simulations following the SPARK¹⁵ framework. We then compared the performances of the BSP model with a basic spatial statistic (Moran's I ¹⁷, which is also adopted by MERINGUE⁹) and other established techniques, including SpatialIDE¹³, SPARK¹⁵, and SPARK-X¹⁶. To ensure a fair comparison of the model performances, we measured their statistical power based on the false discovery rate (FDR), considering the differences between the distribution of calibrated p -values from each method. We present three SVG patterns from previous works on the ST mouse olfactory analysis by SpatialIDE and SPARK in **Figure 2A**. Details of the simulations are outlined in the Methods section.

We evaluated the statistical power of different methods using simulations generated with various signal strengths and noise levels. Signal strengths were measured as the fold-change (FC) in cells' expression levels between the pattern and non-pattern areas. To compare the statistical power, we examined different signal-to-noise ratios ($FC = 3,4,5$) with a moderate noise level ($\tau = 0.5$ as defined in SPARK) in **Supplementary Figure 1**. Additionally, we compared the methods' performance under different noise levels ($\tau = 0.2,0.5,0.8$) with a moderate signal-to-noise ratio ($FC = 4$), as shown in **Supplementary Figure 2**. The BSP method consistently showed superior and stable power across a wide range of FDR cutoffs, signal strengths, and noise levels when analyzing the first and third spatial patterns (**Figure 2B**). Furthermore, BSP exhibited comparative power as SPARK and SpatialDE on the second pattern when the signal strength was moderate (3-fold) and the noise level τ equals 0.5. Compared to other existing approaches, BSP was more powerful, particularly on samples with low signal strengths or high noise levels.

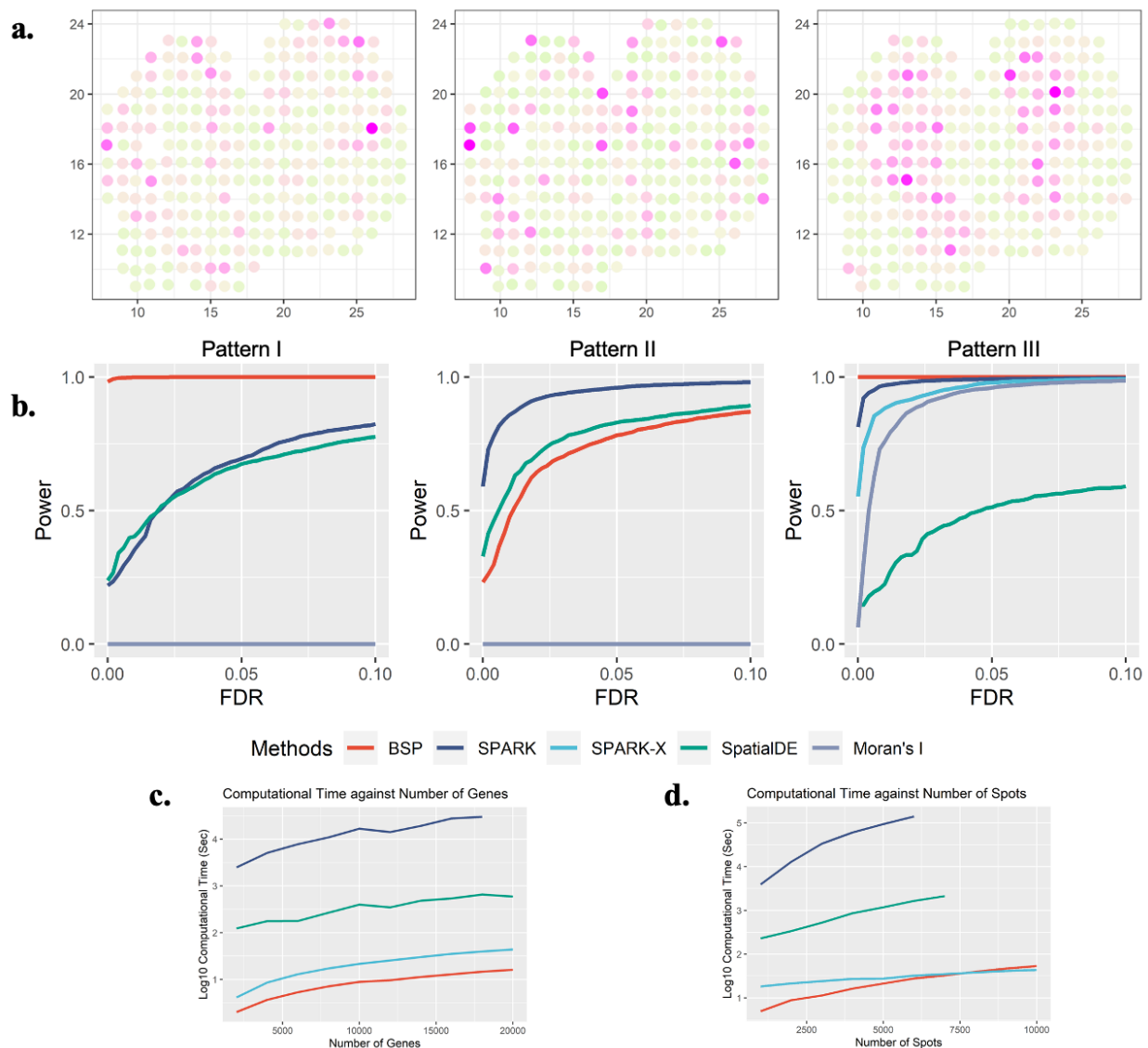


Figure 2: Power analysis for 2D simulation. **a)** Spatial expression patterns I-III (left to right) as defined in SpatialDE and SPARK. **b)** Power comparison among the different methods with moderate noise level ($\tau=0.5$), and moderate signal strengths (3-folds) for the spatial expression patterns I, II, and III (left to right).

All simulations were generated based on the mouse olfactory bulb data with 260 spots of cells. Each simulation replicate contains 1,000 SVGs and 9,000 non-SVGs. **c)** Computational time costs with an increasing number of genes and fixed 2,000 spots. The y-axis indicates the logarithmic of the running time in seconds. **d)** Computational time costs with an increasing number of spots and fixed 10,000 genes. The y-axis indicates the logarithmic of the running time in seconds. The time cost greater than 48 hours are not shown in the figure.

We assessed the computational time and memory usage required for detecting SVGs on 2D data. Compared to existing approaches, BSP performs the SVG analysis with a feasible computational time and memory consumption on personal computers in most scenarios. Computational resource consumption was recorded on an Ubuntu 16.04.4 LTS workstation with Intel(R) Xeon(R) W-2125 CPU @ 4.00GHz and 32 GB memory in **Supplementary Table 1**. In analyzing a typical spatial transcriptomic sample with 2,000 spots, BSP was much faster than other existing methods, regardless of the number of genes (**Figure 2C**). Similarly, for a spatial transcriptomics sample with 10,000 genes, BSP had the lowest computational time among all methods, despite the number of spots (**Figure 2D**). Although not as low as SPARK-X, the corresponding memory usage was also low, as shown in **Supplementary Figure 3**.

2.3 BSP accurately identifies SVGs in 2D space in biological studies

We applied BSP to four previously published 2D spatial transcriptomic datasets, including mouse olfactory bulb¹¹ and human breast cancer obtained by ST sequencing¹¹, hippocampus by SeqFISH²¹, and mouse hypothalamus preoptic region by MERFISH²². We followed the metric evaluation protocols proposed by SPARK and compared Identified SVGs with the provided marker genes in their original research¹⁵. The results were compared with SpatialDE, SPARK, and Spark-X. All the methods were run with the default parameters.

The mouse olfactory bulb dataset contains 11,274 genes measured on 260 spots using SRT sequencing. BSP detected 8 of 10 marker genes from the original study¹¹, while SpatialDE detected 3, SPARK detected 8, and SPARK-X detected 0. The comparison between different methods is shown in **Figure 3A** and **Supplementary Figure 4**. The two marker genes that BSP missed are *Nmb* and *Sv2b*, with *p*-values of 0.0589 and 0.1806, respectively. We reason these missed marker genes have expression variances confined to many isolated, relatively small regions, which could result in the same variances in both big and small patches (**Supplementary Figure 5**). The human Breast cancer dataset contains 5,262 genes measured on 250 spots by SRT sequencing. BSP detected 12 of 14 marker genes identified as SVGs from the original study, while SpatialDE detected 7, SPARK detected 10, and SPARK-X detected 8. The result comparison is shown in **Figure 3B** and **Supplementary Figure 6**. The two marker genes that BSP missed were *PEG10* and *PIP*, with *p*-values of 0.1728 and 0.4371, respectively. The other two FISH-based datasets include the hippocampus dataset, consisting of 249 genes on 131 cells obtained by SeqFISH, and the mouse hypothalamus preoptic region composed of 160 genes on 257 cells by MERFISH. BSP identified most of the marker genes reported in the original studies. Detailed results for mouse olfactory bulb, human breast cancer, hippocampus, and hypothalamus preoptic regions are provided in **Supplementary Tables 2-5**, respectively.

We extended the application of BSP in studies on Acute Kidney Injury (AKI)²³. We ran BSP on five 10X Visium data on AKI samples collected in the Kidney Tissue Atlas²⁴. BSP identified 285 SVGs (*p*-value<0.05) in one representative AKI sample consisting of 317 spots and 14,988 genes. Annotated by clusterProfiler²⁵, the results were supported by a gene ontology (GO) enrichment analysis in **Figure 3C**, including relevant enrichments in humoral immune response (*q*-value 1.09E-11), and humoral immune response mediated by circulating immunoglobulin (*q*-value 1.63E-10). Reactome enrichment analysis²⁶ identified eukaryotic translation elongation (*q*-value

2.77E-13) and influenza infection (q -value 5.59E-09), as in **supplementary Figure 7**. As innate and adaptive immune responses mediate damage to renal tubular cells and recovery from AKI^{27,28}, these results are consistent with disease enrichment analysis²⁹, which found the most significant terms in urinary system disease (q -value 9.30E-11) and kidney disease (q -value 1.36E-10). **Supplementary Table 6** lists all the SVG results obtained by BSP. **Supplementary Table 7** details the results from GO enrichment analysis, **Supplementary Table 8** details the results from Reactome, and **Supplementary Table 9** details the results from disease ontology enrichment analysis. By demonstrating BSP's utility in kidney research, our study highlights the potential for BSP to advance our understanding of complex diseases in diverse tissue types.

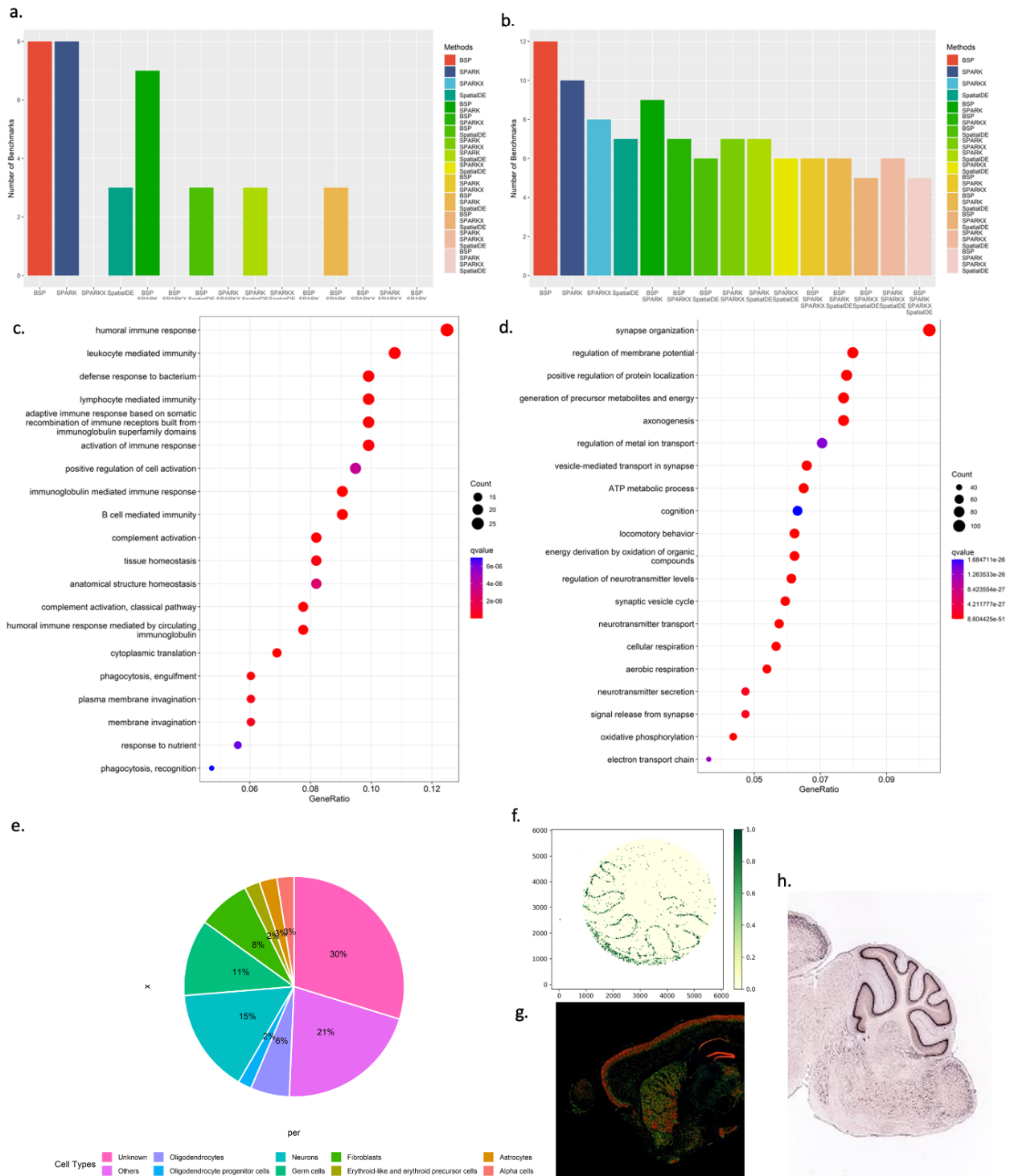


Figure 3. SVGs identified by BSP in biological analysis. **a)** Number of marker genes identified with different methods in mouse olfactory bulb study (the original study identified 10 marker genes). **b)** Number of marker genes identified with different methods in human breast cancer research (the original study identified 14 marker genes). For **a)** and **b)**, intersections between gene identified from different methods were included in analyses. **c)** Gene enrichment analysis on SVGs identified by BSP on mouse cerebellum data using Slide-seq V2. **d)** Gene ontology enrichment analysis on SVGs identified in acute kidney injury studies. **e)** Distribution of cell type marker annotations from Panglaodb on identified SVGs in mouse

cerebellum study using Slide-seq V2. **f)** *Calb1* gene expression in the mouse cerebellum data using Slide-seq V2. The expression values were log-transformed, and those greater than 1.0 were normalized to 1.0. **g)** Expression and **h)** ISH of *Calb1* gene in an adult mouse brain, <http://mouse.brain-map.org/experiment/show/75457491>.

2.4 BSP identifies SVGs on large-scale spatial transcriptomic studies using feasible computational resources

BSP was tested on three large-scale SRT datasets, including Slide-seq data on mouse cerebellum consisting of 18,671 genes on 25,551 beads, Slide-seq V2 data on mouse cerebellum consisting of 23,096 genes on 39,496 beads, and HDST olfactory bulb data consisting of 19,950 genes measured on 181,367 spots. BSP was successful on these large-scale datasets with a reasonable computational time. On the Ubuntu workstation described in Section 2.2, BSP took 7 and 18 minutes to process the Slide-seq mouse cerebellum and Slide-seq V2 mouse cerebellum data, respectively. The memory costs were around 19GB and 32GB, respectively. For the HDST olfactory bulb data, BSP took 4 hours and 90GB of memory on a High-Performance Computer equipped with Intel Xeon(R) CPU E5-2699 v4 @ 2.20GHz. The running details are listed in **Supplementary Table 10**.

BSP detected SVGs with a *p*-value less than or equal to 0.05 (*n* = 842, 1156, and 909 in Slide-seq V1 mouse cerebellum data, Slide-seq V2 mouse cerebellum data, and HDST olfactory bulb data, respectively), and we queried PanglaoDB³⁰ with these detected SVGs. For each of the three implicated datasets, BSP returned numerous neuron-specific and non-neuron-specific genes. Detailed results of detected SVGs from each dataset are listed in **Supplementary Tables 11-13**. Interestingly, in addition to the SVGs corresponding to known cell type composition, many identified genes (30%) were not identified as any cell type markers with PanglaoDB annotations based on the knowledge from previous studies (**Figure 3E**).

On 1,156 identified SVGs in Slide-seq V2 (**Supplementary Table 14**), GO enrichment analysis showed significant enrichments in synapse organization (*q*-value 8.60e-51) (**Figure 3D**, **Supplementary Table 15**). The expression patterns of five representative genes, *Calb1*, *Malat1*, *Nsg1*, *Ttc3*, and *Meg3*, were missed by SPARK-X and annotated as 'unknown' due to the low human brain regional specificity by the Human Protein Atlas³¹. *Calb1* gene (BSP *p*-value 2.73e-14, SPARK-X *p*-value 0.13, **Figure 3F**) is a Ca²⁺ buffering protein found to increase during postnatal development and decrease with aging and neurodegenerative disorders³². *Malat1* (BSP *p*-value 2.73e-14, SPARK-X *p*-value 0.63, **Supplementary Figure 8**) is a highly conserved nuclear-retained lncRNA shown to play a role in regulating genes at both the transcriptional and post-transcriptional levels in a context-dependent manner³³. *Malat1* is shown to be dispensable for normal development and viability in mice³⁴. *Ttc3* gene (BSP *p*-value 2.73e-14, SPARK-X *p*-value 0.30, **Supplementary Figure 9**) is known to play a role in cognitive impairment through protein quality control, which is a common phenotype of Down's syndrome and Alzheimer's disease³⁵. Another representative gene *Nsg1* (BSP *p*-value 2.73e-14, SPARK-X *p*-value 1.00, **Supplementary Figure 10**), is known to be implicated in regulating endosomal recycling and sorting of several important neuronal receptors³⁶. In addition, the *Meg3* gene (BSP *p*-value 2.73e-14, SPARK-X *p*-value 1.00, **Supplementary Figure 11**) modulates AMPA receptor surface expression in primary cortical neurons, and it is in the intricate regulation of the PTEN/PI3K/AKT signaling cascade during synaptic plasticity in neurons³⁷. Besides Slide-seq V2, the spatial patterns of *Calb1*, *Ttc3*, *Nsg1*, and *Meg3* were validated by expression (**Figure 3G**) and ISH (**Figure 3H**) from Allen Brain Atlas³⁸. Overall, the structural and functional compartmentalization in the cerebellum revealed by cell type annotation analysis highlights the utility of BSP.

2.5 BSP accurately identifies SVGs in 3D simulations

We extended the simulation framework in Trendsseek and SPARK further to demonstrate the power of BSP on 3D transcriptomic data. We compared the detection accuracy of SVGs using BSP with that of SPARK-X. The spatial patterns were constructed by a set of center points generated from a random walk with a fixed step length, and any spots within a certain distance from any of the center points were included as the marked cells. We created three 3D patterns, namely, curved stick (Pattern I), thin plate (Pattern II), and irregular lump (Pattern III), controlled by different directions of random walks, as shown in **Figure 4A**.

We performed the power analysis based on the FDR, considering the differences in the distribution of calibrated p -values. Compared to SPARK-X, BSP demonstrated superior and stable power under a wide range of FDR cutoffs with fixed moderate pattern sizes ($r = 2.0$), moderate signal strength ($FC = 2.5$), and moderate noise level ($\sigma = 1$) for the spatial expression patterns I, II, and III (**Figure 4B**). We also varied the pattern sizes, signal strengths, and noise levels while holding the other two parameters constant and found that BSP consistently demonstrated greater power in every scenario tested. **Supplementary Figure 12** shows the power analysis in different pattern sizes using a fixed moderate signal strength (2.5-fold) and low noise level ($\sigma = 0$). Simulations with pattern sizes as small ($r = 1.5$), moderate ($r = 2.0$), and large ($r = 2.5$) are tested on all patterns I, II, and III. **Supplementary Figure 13** demonstrates the results on different signal strengths using a fixed moderate pattern size (radius of 3) and low noise level ($\sigma = 0$). Simulations with signal strengths as low (2-fold), moderate (2.5-fold), and large (3-fold) are tested on patterns I, II, and III. **Supplementary Figure 14** shows the results on various noise levels using a fixed moderate pattern size (radius of 3) and moderate signal strength (3-fold). Simulations with high ($\sigma = 2$), moderate ($\sigma = 1$), and low ($\sigma = 0$) noise levels are tested on patterns I, II, and III.

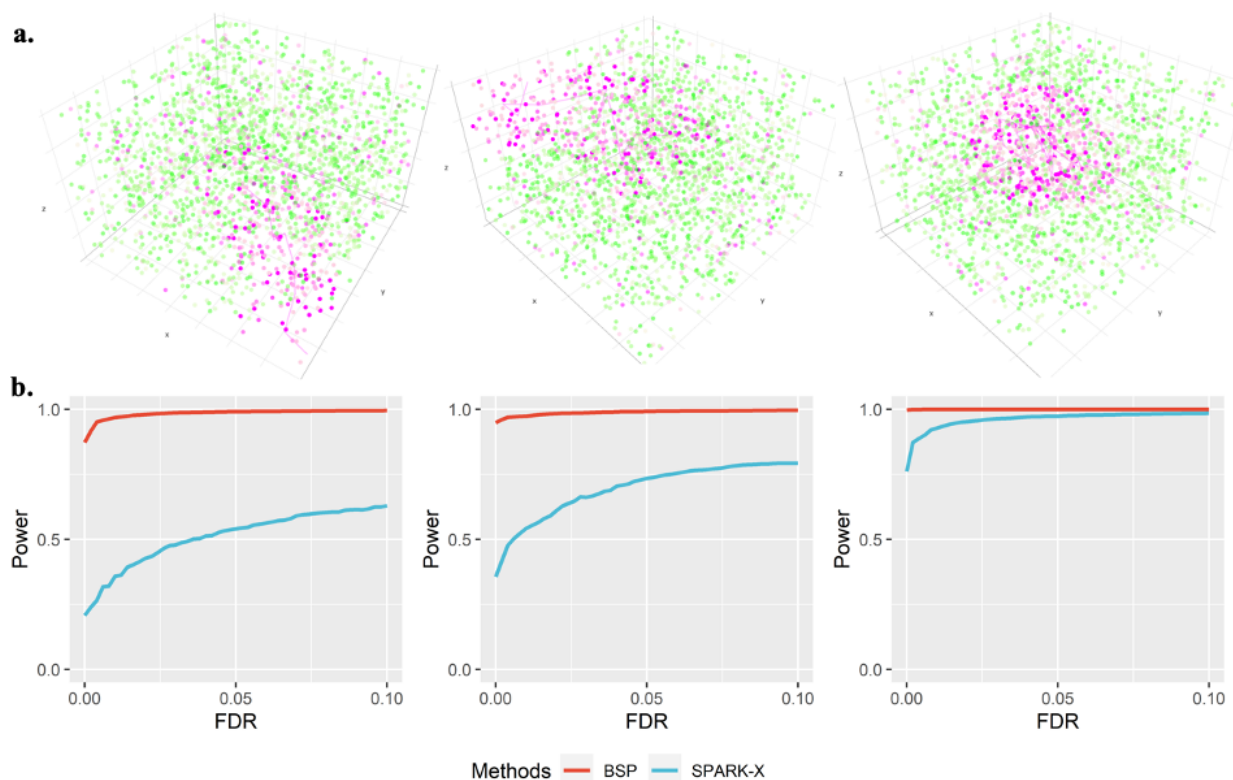


Figure 4: Power analysis for 3d simulation. a) Spatial patterns I-III are controlled by the direction of random walks. **Left:** Pattern I (curved stick): the movements of random walk are monotonic in two directions (x- and z- coordinates, or y- and z-coordinates). **Middle:** Pattern II (thin plate): the movements of the random walk are monotonic in one direction (z-coordinates). **Right:** Pattern III (irregular lump): the movements of the random walk are non-monotonic in any direction. **b)** Power comparison of the different methods under varied pattern sizes. Power charts show the averaged true positive rates (y-axis) across 10 replicates against the false discovery rates (x-axis) for the detected SVGs using each method. Simulations were performed using fixed moderate pattern sizes ($r = 2.0$), moderate signal strength ($FC = 2.5$), moderate noise level ($\sigma = 1$), and three spatial expression patterns I-III (left to right). All simulations were generated based on the seqFISH data with 10 segments (z-coordinate) and 225 spots of cells on each piece (x- and y-coordinates). Each simulation replicate contains 1,000 SVGs and 9,000 non-SVGs.

2.6 BSP identifies more meaningful SVGs in the 3D study than stacking results on the 2D analysis

We utilized BSP on two publicly available 3D transcriptomics datasets, mouse visual cortex through STARmap sequencing⁴ and human RA synovium using stacking SRT⁶. The STARmap dataset contains 28 known SVGs (23 cell-type markers and 5 activity-regulated genes) measured in 33,598 spots. For these low throughput SRT with few genes, BSP adopted the generated null gene approach proposed by SPARK, and identified all these 28 genes as SVGs.

A study on human RA synovium contains 3D spatial transcriptomic sequencing from six RA patients by stacking 2D slices. Each sample consisted of approximately 13,000 genes on three to seven 2D slices with approximately 1,200 spots in each slice. To evaluate the power of 3D transcriptomics, BSP was first applied to each 2D slice, and then to the stacked 3D volume. Using the first sample (patient RA1) as an example, 260 genes were identified as the SVGs by taking the intersection of SVGs from four independent analyses on each 2D slice. However, 1,257 genes

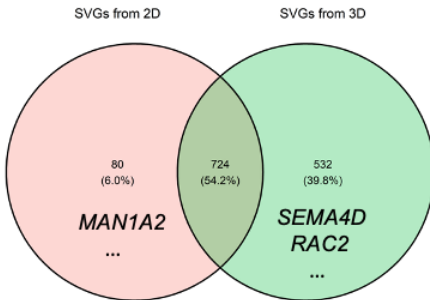
were detected as the SVGs by analyzing the stacked 3D SRT. All 260 genes from the 2D analysis were included in the gene list detected in 3D space, while 997 additional genes were discovered only in 3D space. We further examined these 997 genes neglected by 2D analysis with the DAVID functional annotations³⁹ and found significant enrichments in host-virus interaction (Benjamini: p -value $4.3e-23$), respiratory chain (Benjamini: p -value $3.4e-8$), innate immunity (Benjamini: p -value $2.5e-6$), neutrophil degranulation (Benjamini: p -value $7.1e-31$), and viral process (Benjamini: $7.7e-17$) among biological processes.

We also performed a classical meta-analysis by combining four individual analyses on each 2D slice using Fisher's combined probability test with SciPy packages. The meta-analysis identified 804 genes as statistically significant (p -value < 0.05). Compared to the 1,257 SVGs identified by the 3D analysis, 724 genes were detected as SVGs by both the 2D meta-analysis and 3D settings (**Supplementary Table 16**), 532 genes were only significant in 3D settings (**Supplementary Table 17**), and 80 genes were only significant in the 2D meta-analysis setting (**Supplementary Table 18**). **Figure 5A** shows the Venn diagram of differences between meta-analysis and 3D analysis. **Figure 5B** shows GO enrichment analysis on all SVGs identified in 3D settings (**Supplementary Table 19**). Several immune-related gene ontologies are highlighted in RA studies, including response to interferon-gamma (q -value $2.25e-11$)⁴⁰, myeloid leukocyte migration (q -value $3.39e-09$), leukocyte migration (q -value $3.45e-12$), leukocyte chemotaxis (q -value $2.29e-09$), regulation of leukocyte migration (q -value $5.11e-09$)⁴¹. **Supplementary Figure 15** and **Supplementary Table 20** show GO enrichment results on 724 genes, both identified by 2D meta-analysis and 3D settings. The same GO enrichment analysis is proceeded on 532 genes uniquely identified by 3D settings in **Supplementary Figure 16** and **Supplementary Table 21**. These highlighted GO terms are indicative of key immune responses in immunizations in RA progression⁴².

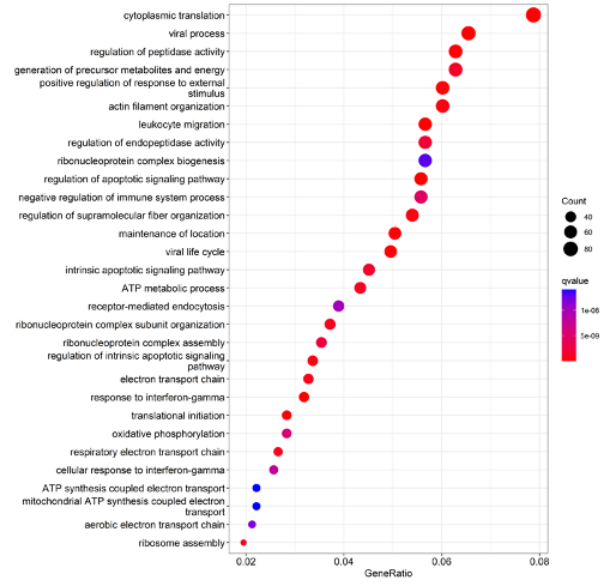
2D meta-analysis may lead to some misleading results. Among these genes, *MAN1A2* (**Figure 5C**) gets Fisher's combined p -value $7.56e-08$ with four individual 2D p -values 0.8755, $6.83e-06$, 0.0129, and $3.6e-04$. However, the p -value of *MAN1A2* in 3D settings is 0.3648, making it unlikely as an SVG in 3D space when considering all the slices as a volume. On the other hand, among SVGs only significant in 3D analysis, *SEMA4D* plays a role in the immune system and induces B-cells to be aggregated and improves their viability (*in vitro*)⁴³. Although the individual 2D p -values of *SEMA4D* are 0.9505, 0.9495, 0.9616, and 0.9735, its Fisher's combined p -value in meta-analysis is 1.0, which has the least possibility of being an SVG in all individuals and meta-analysis on biased 2D analysis. However, the BSP test results of gene *SEMA4D* is 0.0482 on the 3D volume, making it stand out from the genes (**Figure 5D**). Another example that fails in the 2D analysis is *RAC2* (**Figure 5E**), which encodes a member of the Ras superfamily of small guanosine triphosphate (GTP)-metabolizing proteins involved in generating reactive oxygen species. Although Fisher's combined p -value is 0.0942 with four individuals as 0.1173, 0.0846, 0.3286, and 0.3500, it is very significant with a p -value of $4.4929e-15$ in the 3D setting. This analysis demonstrates that the analysis by BSP on intact 3D volume provides new opportunities in identifying SVGs compared to 2D analysis with potential bias.

The same analyses were conducted on each of the six RA patients individually. The enriched GO terms of 3D SVGs identified in each patient were presented in **Figure 5B** and **Supplementary Figures 17-21**. We observed that most GO terms were consistently enriched in all the patients (**Figure 5F**), indicating that BSP robustly identified 3D SVGs across various samples.

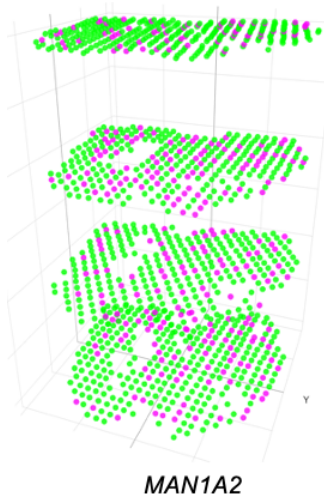
a.



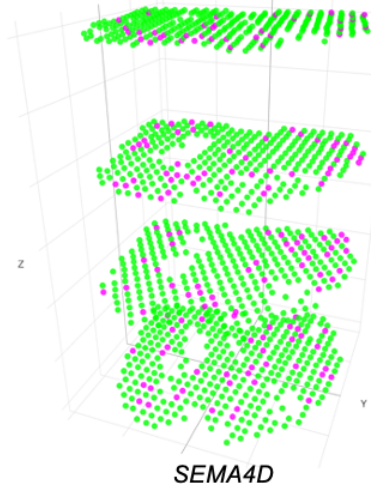
b.



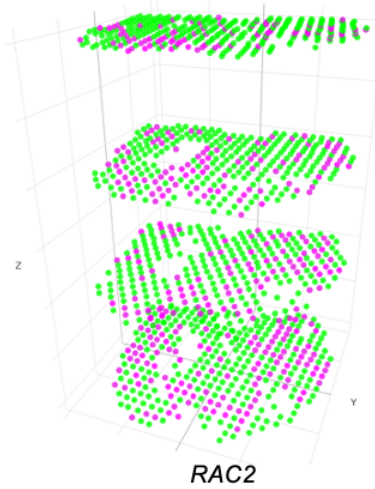
c.



d.



e.



f.

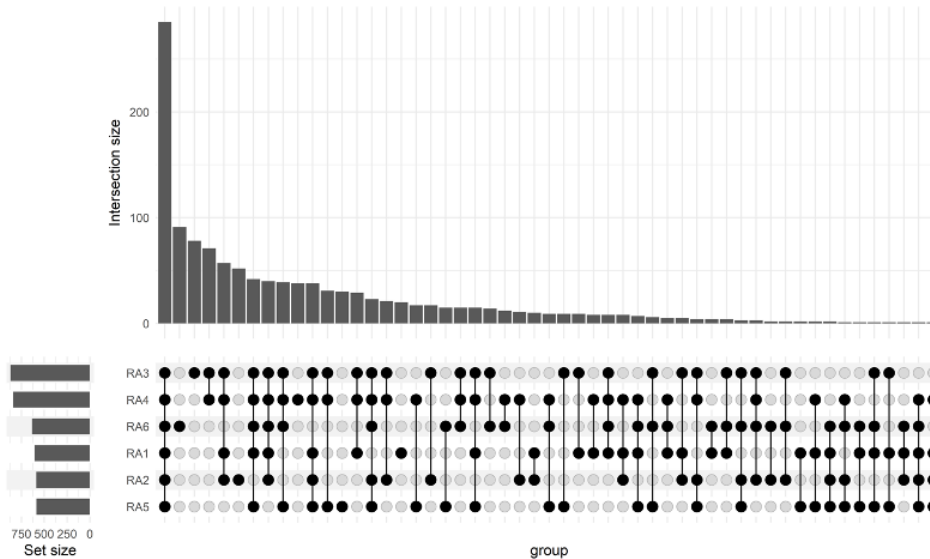


Figure. 5 BSP identifies much more meaningful SVGs in the 3D study than stacking results on 2D analysis. **a)** Venn diagram of SVGs identified by 2D meta-analysis and 3D analysis; **b)** Gene ontology analysis on SVGs identified in patient RA1 in 3D setting; **c)** *MAN1A2* gene, significant in 2D analysis but not an SVG in 3D setting; **d)** *SEMA4D* gene identified as SVGs in 3D transcriptomics but missed by 2D analysis; **e)** *RAC2* gene, a very significant SVGs in 3D transcriptomics but missed by 2D analysis; **f)** Upset plot of enriched gene ontology terms on all six individual RA patients.

3. Discussion

The advancing spatial transcriptomics measures high-throughput multi-cellular- or cellular-level gene expression in the spatial context. This fast-growing 3D technology is critical for understanding the relationship between tissue structure and underlying biological function, posing new challenges in identifying SVGs vital in linking individual genes to spatial expression variance. The proposed BSP provides a dimension-agnostic and utilizes a big-small patch algorithm to identify SVGs at varying levels of granularity. The performance of BSP has been validated in both simulations and real studies using 2D and 3D data. While there is still a debate over the gold standard for SVGs in biological studies, we follow the protocol adopted by SPARK for analysis and annotation. Simulations provide an alternative benchmark for methods development. In the 2D simulation, BSP outperformed existing methods in most scenarios with different signal-to-noise ratios. In the 3D simulations, BSP demonstrated its superiority compared to other well-known criteria, such as Moran's-I. Meanwhile, these 3D simulations can serve as benchmarks for developing new methods. In biological studies using 2D and 3D data, BSP identified more convincing SVGs than existing methods with a good control of false positives. For instance, in a human RA study, BSP revealed that analyzing SVGs as a volume in 3D data outperformed stacking results on individual 2D slices.

The innovation of BSP lies in its dimension-agnostic and granularity-guided approach, which utilizes paired big-small patches. Intuitively, the big patch provides a global view of the spatial pattern with a lower resolution, while the small patch focuses on the local details with a higher resolution. Using ratios between variances of the paired patches, BSP can accurately delineate the spatial patterns in a quantitative manner. Moreover, this approach is applicable to any dimension. These defined patches can effectively capture the characteristics of the expression patterns in both 2D space and 3D volume, making BSP capable of analyzing SRT data in both dimensions.

This granularity-guided approach makes the BSP a data-driven, model-free, and parameter-free model. First, BSP is particularly well-suited for the complexities of biological data, especially in the tumor microenvironment, where fixed spatial patterns cannot be assumed to form locally and globally. BSP's effectiveness in these complex scenarios has been demonstrated in both 2D and 3D simulations, without preconceived assumptions about the underlying distributions. Second, BSP is robust to different levels of signal strengths and tolerates occasional noises, and it robustly discover the same persistent results in different samples, as the spatial patterns are invariant in different scale. Third, the BSP algorithm is highly efficient. In the typical scenario of a 10X Visium scale, BSP remains the fastest method among all the existing methods. Even for large-scale datasets, such as Slide-seq, Slide-seqV2, and HDST, BSP remains feasible with reasonable computational resources. Fourth, BSP's core implementation is just a few dozen lines of code, making it easy to implement and adaptable to different usage scenarios.

Although BSP has shown significant improvements in quantitatively measuring spatial patterns using the beta distribution to fit the distribution of test scores of all the genes, some limitations still need to be addressed. Alternative statistical distributions or non-statistical ranking measurements could be explored to further improve the fitting of the distribution of ratios between variances of

the averaged expression in the paired big-small patch. Furthermore, BSP compromises the performance and computational resources in SRT studies. Although gets superior performances on the benchmarks, BSP consumes more time and more memory than SPARK-X on large-scale datasets.

In conclusion, BSP has demonstrated its efficacy as a robust method for identifying SVGs in both 2D and 3D spatial transcriptomics analysis. As 3D sequencing technologies continue to advance and mature, we anticipate BSP to be increasingly valuable in future applications of 3D spatial transcriptomics. Moreover, as time is often considered as the fourth dimension in development biology^{44,45}, we will also explore the potential for spatiotemporal studies using BSP.

4. Methods

4.1 BSP algorithm

BSP aims to identify spatially variable genes in 2D or 3D SRT data. The algorithm contains several steps, including (1) normalizing expression and spatial coordinates, (2) defining big and small patches for each spot based on neighboring spots with a larger or small radius, (3) calculating local means of gene expression for both big and small patches, (4) computing the ratio between the variances of local means between for big and small patches for each gene, and (5) fitting the ratio of each gene with a beta distribution and adjusting the p-values for each gene. The flowchart of the BSP algorithm is shown in **Supplementary Figure 22**.

4.1.1 Problem setting and data normalization

On an SRT sample with M spots and N genes. The coordinates of spot i are (x_i, y_i) for 2D spatial transcriptomics, (x_i, y_i, z_i) for 3D spatial transcriptomics. The expression level of gene j in spot i is denoted as $X_i^{(j)}$, where $1 \leq i \leq M$, $1 \leq j \leq N$. The goal of BSP is to identify SVGs from all N genes with significant spatial patterns.

All gene expression levels are normalized and scaled to $[0,1]$ using a min-max normalization across all spots, $0 \leq X_i^{(j)} \leq 1$ for all $X_i^{(j)}$. The normalization of spatial coordinates of the spots on SRT is based on the density of spots. The coordinates of spots in each direction are divided by the estimated density, which is calculated as the total number of spots divided by the area (2D) or volume (3D) of the sample. For simplicity, a rectangle is defined as the 2D space, and a cube is defined as the 3D volume. The rescaling functions f for 2D space is defined as Eq. (1), and for 3D space as Eq. (2):

$$f(x_i, y_i) = \sqrt{\frac{N}{\Delta X \Delta Y}} \cdot (x_i, y_i) \quad \text{Eq. (1)}$$

$$f(x_i, y_i, z_i) = \sqrt[3]{\frac{N}{\Delta X \Delta Y \Delta Z}} \cdot (x_i, y_i, z_i) \quad \text{Eq. (2)}$$

where x_i , y_i , and z_i are the coordinates of spot i . ΔX , ΔY , and ΔZ denote the ranges of the sample space. They can be calculated as the differences between the maximum and minimum coordinates in each direction for the cube, as: $\Delta X = \max(x) - \min(x)$, $\Delta Y = \max(y) - \min(y)$, and $\Delta Z = \max(z) - \min(z)$.

This spatial coordinate normalization step ensures an adequate number of spots captured by the pre-defined radii D_1 and D_2 . The goal of this step is to minimize the average spot-to-spot distance to slightly less than one unit. Typically, the default value of D_1 is set as one unit to capture the nearest neighbors, while D_2 is set to three units to include more spots in the patches.

4.1.2 Big-small path

After coordinates normalization, the Euclidean distance between spots i_1 and i_2 is calculated as $dist(i_1, i_2)$. For a given spot i , a **patch** S_i is defined as the set of neighboring spots l within the radius of D

$$S_i = \{l: dist(i, l) < D, l = 1 \leq l \leq M \text{ and } l \neq i\}$$

With a patch S_i , $\widehat{X}_i^{(j)}$ is defined as the **Local Mean**, the average expression level of gene j in this patch.

$$\widehat{X}_i^{(j)} = \frac{\sum_{i \in S_i} X_i^{(j)}}{num(S_i)} \quad \text{Eq. (3)}$$

where $num(S_i)$ is the cardinal number of spots within S_i . Local Mean describes the expression characteristics in the patch. $\sigma_D^{(j)}$ is defined as the variance of Local Means of all patches on gene j .

$$\sigma_D^{(j)} = var(\cup_{i \in [1, M]} \widehat{X}_i^{(j)}) \quad \text{Eq. (4)}$$

For a gene j without any spatial expression pattern, i.e., the distributions of the expression levels being identical across all spots, $\sigma_D^{(j)}$ equals to 0. Otherwise, $\sigma_D^{(j)} > 0$. If distance D is big enough to cover the radius of the sample, then $\sigma_D^{(j)}$ also equals 0 for each patch containing all spots, as the Local Means are the same for each spot.

For each spot i , we define a paired big-small patch, i.e., a small patch is defined as S_i' with a radius D_1 , a big patch is defined as S_i'' with a radius D_2 , where $D_1 < D_2$. We take the $r_{D_1 D_2}^{(j)}$, the ratio between the variances of the paired local averaged expression levels between big patch and small patch, describes the characteristics of the spatial pattern on gene j , defined as:

$$r_{D_1 D_2}^{(j)} = w_j \frac{\sigma_{D_2}^{(j)}}{\sigma_{D_1}^{(j)}} \quad \text{Eq. (5)}$$

w_j is the weight to normalize the intrinsic gene expression variance, i.e., $w_j = \frac{\sigma_j^2}{\max(\sigma_j^2)}$, where σ_j^2 is the variance of expression levels of gene j of all the spots in the sample.

4.1.3 Fitting beta distribution and adjusted p-value

After $r_{D_1 D_2}^{(j)}$ is calculated with all the genes, $j \in [1, M]$, a beta distribution is approximated using the *stat* packages from sklearn⁴⁶. If $r_{D_1 D_2}^{(j)}$ is greater than the $100 * (1 - \alpha)\%$ upper tail of the fitted beta distribution, where α refers to the significance level (usually sets as 0.05), the significance p-value of each gene is addressed. In situations where there are insufficient null genes, a set of randomly permuted genes is generated to estimate the null distribution for practical usage.

4.2 2D simulation on mouse olfactory bulb data

We utilized the mouse olfactory bulb data within the framework of SPARK to construct 2D simulations. The simulation was based on mouse olfactory bulb data consisting of three spatial expression patterns measured on 260 spots (as shown in **Figure 2a**). Each simulation contained 1000 simulated SVGs with identified patterns in SpatialDE and SPARK, as well as 9000 non-

SVGs generated through gene permutation without any spatial expression pattern. The p-values from basic spatial autocorrelation statistics Moran's I, SpatialDE, SPARK, SPARK-X, and BSP were calculated to quantify the corresponding power (true positive rates) given a false discovery rate (FDR). To illustrate the rate of true positives (y-axis) identified by each method at different FDRs (x-axis) in power analysis, we generated ten replicates for each simulation. Specifically, simulation data was generated under different signal-noise ratios ($FC = 3,4,5$) with a medium level of noise ($\tau = 0.5$, as defined in SPARK). Then another set of simulation data was generated under different noise levels ($\tau = 0.2,0.5,0.8$) with a moderate signal-noise ratio ($FC = 4$).

4.3 3D simulation on FISH data

For the simulations in 3D space scenarios, we extended the framework originally introduced by Trendsceek and SPARK. All simulations were generated based on seqFISH data, with 10 segments in the z-coordinate and 225 spots representing cells in each piece in the x- and y-coordinates. We assume the sample was cryosectioned into 10 sections, with each section placed on an individual array without any direct contact between array surfaces. To generate spatial locations for a fixed number of cells ($n = 225$) in each section, we used a random-point-pattern Poisson process. These spatial locations for each section were then stacked together with the index of the section serving as the z-coordinates ($z = 1,2, \dots, 10$).

The 3D spatial patterns were constructed using a set of spheres with center points generated through a random walk with a fixed step length of 2. We included three types of spatial patterns in the simulations by controlling the range of directions (**Figure 4a**). These patterns include Pattern I (curved stick), the movements of a random walk are monotonic in two directions (x- and z- coordinates, or y- and z-coordinates); Pattern II (thin plate), the movements of a random walk are monotonic in one direction (z- coordinates); Pattern III (irregular lump), the movements of a random walk are non-monotonic in any directions. We produced 1000 SVGs with 3D patterns for each simulation, and generated 9000 non-SVGs without any spatial expression pattern by permutating known patterns.

The expression of SVGs was sampled based on whether the cell was inside or outside the pattern, distinguishing between marked and non-marked cells. For marked cells inside the pattern, we randomly selected gene expression values from the upper quantile of the gene expression distribution in the seqFISH data. For non-marked cells and those outside the pattern, we assigned gene expression randomly from the expression measurements in the seqFISH data. Non-SVGs were generated by permutating gene expressions of SVGs. For each SVG, the expression values were permuted and repeated 9 times (i.e., randomly assigning values to all cells without replacement). Finally, random noise was added proportionally to the averaged standard deviation of expressions in all genes.

To systematically explore the influences under different scenarios, we held two parameters constant while manipulating the third to vary the patterns' sizes, signal strengths, and noise levels. We tested three sphere radius values (r) of 1.5, 2.0, and 2.5, which determined the pattern size. Quantile thresholds of 0.66, 0.80, and 0.88 were set, corresponding to expected expression fold changes (FC) of 2, 2.5, and 3 between marked cells and non-marked cells, indicating low, moderate, and strong signal strengths, respectively. We applied random noise following a Gaussian distribution with mean zero and the standard deviation (σ) of 0, 1, and 2 times the averaged standard deviation of the expressions of all simulated genes to represent low, moderate, and high noise levels. In the 3D simulations we varied the pattern sizes ($r = 1.5,2.0,2.5$), expression fold changes ($FC = 2,2.5,3$), and noise levels ($\sigma = 0,1,2$) across spatial patterns I, II, and III. For each combination of pattern size, signal strength, and noise level, we conducted 10 replicates to perform the power analysis.

4.4 Biological data collection and analysis.

For studies on mouse olfactory bulb, human breast cancer obtained by SRT sequencing, hippocampus by SeqFish, and mouse hypothalamus preoptic region by MERFISH, we followed the analysis protocol adopted by SPARK. For studies on Slide-seq data, Slide-seqV2 data, HDST data, we followed the analysis protocol adopted by SPARK-X.

In the case of human RA synovium studies, the spatial locations in 2D slices were normalized with unit one. These 2D slices were stacked together with interval one on the z-axis to construct a volume on 3D transcriptomics. Analysis was performed based on the normalized data provided by the authors.

For kidney analysis, all the data were generated using 10X Visium platforms and processed with CellRange. Expression data is quality controlled and preprocessed by Seurat with scTransform⁴⁷.

4.5 Annotations

The annotations using PanglaoDB were performed by rPanglaoDB (<https://github.com/dosorio/rPanglaoDB>). Go enrichment analysis was performed by topGO⁴⁸ (Version 3.16). Reactome pathway analysis was performed by ReactomePA²⁶ (Version 3.16). Disease Ontology Semantic and Enrichment analysis was performed by DOSE²⁹ (Version 3.16). Meta-analysis is performed by SciPy from the sklearn package⁴⁶ (Version 1.1.2) in python 3.9.12.

Data Availability

The mouse olfactory bulb and human breast cancer data are available at <http://www.spatialtranscriptomicsresearch.org>, the MERFISH data can be downloaded from (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248>), and the SeqFISH data is available at <https://www.cell.com/cms/10.1016/j.neuron.2016.10.001/attachment/759be4dc-04a6-4a58-b6f6-9b52be2802db/mmc6.xlsx>. Slide-seq data, Slide-seqV2 data, HDST data, and human rheumatoid arthritis synovium data are available at Broad Institute's single-cell repository (https://singlecell.broadinstitute.org/single_cell/) with ID SCP354, SCP948, SCP420, and SCP1414. The STARmap data set is available at <https://www.starmapresources.com/data>. The kidney spatial transcriptomics data can be downloaded from the Kidney Tissue Atlas (<https://atlas.kpmp.org/>). The generated simulation data is available at https://mailmissouri-my.sharepoint.com/:f/g/personal/wangjue_umsystem_edu/EnjH6hbt1ptBjWBzoQhGfPoBhFVVynkeJbrdpszmeqnYpA?e=jDHmEJ , and will be deposited in the Zenodo database.

Code Availability

The source code of BSP is freely available at <https://github.com/juexinwang/BSP/>.

Acknowledgments

This work was supported by grants R35-GM126985 and R01-GM131399 from the National Institutes of Health, as well as the Pelotonia Institute of Immuno-Oncology (PIIO).

References

- 1 Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211-220 (2021).
- 2 Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat Methods* **19**, 534-546, doi:10.1038/s41592-022-01409-2 (2022).
- 3 Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat Rev Genet* **23**, 741-759, doi:10.1038/s41576-022-00515-3 (2022).
- 4 Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, doi:10.1126/science.aat5691 (2018).
- 5 Eng, C. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239, doi:10.1038/s41586-019-1049-y (2019).
- 6 Vickovic, S. *et al.* Three-dimensional spatial transcriptomics uncovers cell type localizations in the human rheumatoid arthritis synovium. *Commun Biol* **5**, 129, doi:10.1038/s42003-022-03050-3 (2022).
- 7 Young, D. M. *et al.* Constructing and optimizing 3D atlases from 2D data with application to the developing mouse brain. *Elife* **10**, doi:10.7554/eLife.61408 (2021).
- 8 Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nat Methods* **19**, 567-575, doi:10.1038/s41592-022-01459-6 (2022).
- 9 Miller, B. F., Bambah-Mukku, D., Dulac, C., Zhuang, X. & Fan, J. Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Res* **31**, 1843-1855, doi:10.1101/gr.271288.120 (2021).
- 10 Chen, S. *et al.* Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle temporal gyrus in Alzheimer's disease. *Acta Neuropathol Commun* **10**, 188, doi:10.1186/s40478-022-01494-6 (2022).
- 11 Stahl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82, doi:10.1126/science.aaf2403 (2016).
- 12 Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**, 425-436, doi:10.1038/s41593-020-00787-0 (2021).
- 13 Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat Methods* **15**, 343-346, doi:10.1038/nmeth.4636 (2018).
- 14 Edsgard, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat Methods* **15**, 339-342, doi:10.1038/nmeth.4634 (2018).
- 15 Sun, S. Q., Zhu, J. Q. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat Methods* **17**, 193-+, doi:10.1038/s41592-019-0701-7 (2020).
- 16 Zhu, J., Sun, S. & Zhou, X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol* **22**, 184, doi:10.1186/s13059-021-02404-0 (2021).
- 17 Moran, P. A. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17-23 (1950).
- 18 Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* **18**, 1342-1351, doi:10.1038/s41592-021-01255-8 (2021).
- 19 Xue, Y. *et al.* A 3D Atlas of Hematopoietic Stem and Progenitor Cell Expansion by Multi-dimensional RNA-Seq Analysis. *Cell Rep* **27**, 1567-1578 e1565, doi:10.1016/j.celrep.2019.04.030 (2019).
- 20 Hobbs, J. R. in *Readings in qualitative reasoning about physical systems* 542-545 (Elsevier, 1990).

- 21 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, doi:10.1016/j.neuron.2016.10.001 (2016).
- 22 Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, doi:10.1126/science.aau5324 (2018).
- 23 Safari, S., Hashemi, B., Forouzanfar, M. M., Shahhoseini, M. & Heidari, M. Epidemiology and Outcome of Patients with Acute Kidney Injury in Emergency Department; a Cross-Sectional Study. *Emerg (Tehran)* **6**, e30 (2018).
- 24 Lake, B. B. *et al.* An atlas of healthy and injured cell states and niches in the human kidney. *bioRxiv* (2021).
- 25 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141, doi:10.1016/j.xinn.2021.100141 (2021).
- 26 Yu, G. & He, Q. Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* **12**, 477-479, doi:10.1039/c5mb00663e (2016).
- 27 Bonavia, A. & Singbartl, K. A review of the role of immune cells in acute kidney injury. *Pediatr Nephrol* **33**, 1629-1639, doi:10.1007/s00467-017-3774-5 (2018).
- 28 Jang, H. R. & Rabb, H. Immune cells in experimental acute kidney injury. *Nat Rev Nephrol* **11**, 88-101, doi:10.1038/nrneph.2014.180 (2015).
- 29 Yu, G., Wang, L. G., Yan, G. R. & He, Q. Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608-609, doi:10.1093/bioinformatics/btu684 (2015).
- 30 Franzen, O., Gan, L. M. & Bjorkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019**, doi:10.1093/database/baz046 (2019).
- 31 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).
- 32 Goffigan-Holmes, J., Sanabria, D., Diaz, J., Flock, D. & Chavez-Valdez, R. Calbindin-1 Expression in the Hippocampus following Neonatal Hypoxia-Ischemia and Therapeutic Hypothermia and Deficits in Spatial Memory. *Dev Neurosci*, 1-15, doi:10.1159/000497056 (2019).
- 33 Xu, W. W., Jin, J., Wu, X. Y., Ren, Q. L. & Farzaneh, M. MALAT1-related signaling pathways in colorectal cancer. *Cancer Cell Int* **22**, 126, doi:10.1186/s12935-022-02540-y (2022).
- 34 Arun, G., Aggarwal, D. & Spector, D. L. MALAT1 Long Non-Coding RNA: Functional Implications. *Noncoding RNA* **6**, doi:10.3390/ncrna6020022 (2020).
- 35 Zhou, X. *et al.* TTC3-Mediated Protein Quality Control, A Potential Mechanism for Cognitive Impairment. *Cell Mol Neurobiol* **42**, 1659-1669, doi:10.1007/s10571-021-01060-z (2022).
- 36 Yap, C. C., Digilio, L., McMahon, L. & Winckler, B. The endosomal neuronal proteins Nsg1/NEEP21 and Nsg2/P19 are itinerant, not resident proteins of dendritic endosomes. *Sci Rep* **7**, 10481, doi:10.1038/s41598-017-07667-x (2017).
- 37 Tan, M. C. *et al.* The Activity-Induced Long Non-Coding RNA Meg3 Modulates AMPA Receptor Surface Expression in Primary Cortical Neurons. *Front Cell Neurosci* **11**, 124, doi:10.3389/fncel.2017.00124 (2017).
- 38 Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-176, doi:10.1038/nature05453 (2007).
- 39 Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50**, W216-221, doi:10.1093/nar/gkac194 (2022).
- 40 Kato, M. New insights into IFN-gamma in rheumatoid arthritis: role in the era of JAK inhibitors. *Immunol Med* **43**, 72-78, doi:10.1080/25785826.2020.1751908 (2020).

- 41 Tarrant, T. K. & Patel, D. D. Chemokines and leukocyte trafficking in rheumatoid arthritis. *Pathophysiology* **13**, 1-14, doi:10.1016/j.pathophys.2005.11.001 (2006).
- 42 Weyand, C. M. & Goronzy, J. J. The immunology of rheumatoid arthritis. *Nat Immunol* **22**, 10-18, doi:10.1038/s41590-020-00816-x (2021).
- 43 Hall, K. T. *et al.* Human CD100, a novel leukocyte semaphorin that promotes B-cell aggregation and differentiation. *Proc Natl Acad Sci U S A* **93**, 11780-11785, doi:10.1073/pnas.93.21.11780 (1996).
- 44 Wang, M. *et al.* High-resolution 3D spatiotemporal transcriptomic maps of developing *Drosophila* embryos and larvae. *Dev Cell* **57**, 1271-1283 e1274, doi:10.1016/j.devcel.2022.04.006 (2022).
- 45 Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature* **571**, 505-509, doi:10.1038/s41586-019-1338-5 (2019).
- 46 Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825-2830 (2011).
- 47 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome biology* **20**, 1-15 (2019).
- 48 Alexa, A. & Rahnenführer, J. Gene set enrichment analysis with topGO. *Bioconductor Improv* **27**, 1-26 (2009).