bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Systematic comparison of sequencing-based spatial transcriptomic methods

Yue You[1†], Yuting Fu[2,3,4,5†], Lanxiang Li[1], Zhongming Zhang[1], Shikai Jia[2,3,4,5], Shihong Lu[1], Wenle Ren[1], Yifang Liu[2,3,4,5], Yang Xu[6], Xiaojing Liu[2,3,4,5], Fuqing Jiang[8,9], Guangdun Peng[8,9], Abhishek Sampath Kumar[10], Matthew E. Ritchie[6,7], Xiaodong Liu[2,3,4,5*], Luyi Tian[1,11*]

[1*]Guangzhou National Laboratory, Guangzhou, China.
[2*]School of Life Sciences, Westlake University, Hangzhou, China.
[3]Research Center for Industries of the Future, Westlake University, Hangzhou, China.
[4]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China.
[5]Westlake Institute for Advanced Study, Hangzhou, China.
[6]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia.
[7]Department of Medical Biology, The University of Melbourne, Parkville, Victoria, Australia.
[8]Center for Cell Lineage and Development, CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, GIBH-HKU Guangdong-Hong Kong Stem Cell and Regenerative Medicine Research Centre, University of Chinese Academy of Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China.
[9]Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, China.
[10]Department of Stem Cell & Regenerative Biology, Harvard University. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, USA.
[11*]GMU-GIBH Joint School of Life Sciences, Guangzhou Medical University, Guangzhou, GuangDong Province, China.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Corresponding author(s). E-mail(s): liuxiaodong@westlake.edu.cn;
tian_luyi@gzlab.ac.cn;
†These authors contributed equally to this work.

Abstract

Recent advancements of sequencing-based spatial transcriptomics (sST) have catalyzed significant advancements by facilitating transcriptome-scale spatial gene expression measurement. Despite this progress, efforts to comprehensively benchmark different platforms are currently lacking. The extant variability across technologies and datasets poses challenges in formulating standardized evaluation metrics. In this study, we established a collection of reference tissues and regions characterized by well-defined histological architectures, and used them to generate data to compare six sST methods. We highlighted molecular diffusion as a variable parameter across different methods and tissues, significantly impacting the effective resolutions. Furthermore, we observed that spatial transcriptomic data demonstrate unique attributes beyond merely adding a spatial axis to single-cell data, including an enhanced ability to capture patterned rare cell states along with specific markers, albeit being influenced by multiple factors including sequencing depth and resolution. Our study assists biologists in sST platform selection, and helps foster a consensus on evaluation standards and establish a framework for future benchmarking efforts that can be used as a gold standard for the development and benchmarking of computational tools for spatial transcriptomic analysis.

1 Main

The advent of high-throughput sequencing technologies has revolutionized transcriptomics, providing unparalleled insights into the complexities of gene expression. Single-cell RNA sequencing (scRNA-seq) has been instrumental in dissecting cellular heterogeneity but falls short in capturing the spatial context essential for understanding tissue architecture, cellular interactions, and functional state [1, 2]. To address this limitation, sequencing-based spatial transcriptomics (sST) has emerged as a pivotal approach, enabling comprehensive transcriptomic profiling while preserving spatial information within tissues [3, 4].

Despite the rapid advancements in sST technologies, the field is still in its very early stages. The imaging-based spatial transcriptomics has a longer history and a collaborative benchmarking effort has been initiated with the SpaceTX consortium [5]. However, a systematic benchmarking study has not been done for sST. Prior studies have established frameworks for comparing single-cell transcriptomic and epigenomic methods, underscoring the necessity for standardized evaluation criteria and reference tissues for technology validation [6–9], since simulated single-cell and spatial data may not be reliable [10]. While sST technologies share common features, such as the use of spatial DNA barcodes analogous to cell barcodes in scRNA-seq, the methods

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

diverge significantly in aspects like spatial resolution and the preparation of spatially barcoded oligo arrays [11]. This variability introduces challenges in method selection and complicates the establishment of universal evaluation standards.

In the present study, we address this critical gap by conducting a systematic comparison of six sST methods. Using a set of reference tissues, including mouse embryonic eyes and hippocampal regions of the mouse brain, we generated cross-platform data for sequencing-based ST benchmarking, referred to as *cadasSTre*. This dataset enables us to evaluate the performance of each technology in terms of spatial resolution, capture efficiency, and molecular diffusion. We updated *scPipe* [12] to enable preprocessing and downsampling of sST data, to further minimize variability and facilitate the incorporation of future technologies. Our analyses reveal that data generated from different sST technologies exhibit varying capabilities in downstream applications, such as clustering, region annotation, and cell-cell communication. Notably, we also highlighted gene detection biases in sST data.

Our study serves multiple purposes: it (i) guides researchers in the selection of appropriate sST methods for their specific biological questions, (ii) establishes a framework for future benchmarking endeavors, and (iii) contributes to the standardization of evaluation criteria in this rapidly evolving field. Furthermore, our work aims to provide a foundation for the assessment of computational tools designed for spatial transcriptomic data analysis.

2 Results

2.1 Benchmarking reference tissues and experimental design

We systematically benchmarked spatial transcriptomics (sST) methods based on distinct spatial indexing strategies, encompassing microarray (10X Genomics Visium [13]), bead-based approaches (HDST [14], BMKMANU S1000, Slide-seq [15]), polony- or nanoball-based technologies (Stereo-seq [16], PIXEL-seq [17]), and microfluidics (DBiT-seq [18]). Details of each sST method are listed in Supplementary Table 1.

We selected the adult mouse brain, E12.5 mouse embryo, and adult mouse olfactory bulb (OB) as reference tissues due to their relatively well-defined morphological characteristics. Adult mouse hippocampus, for instance, exhibits consistent thickness and comprises regions such as Cornu Ammonis (CA)1, CA2, CA3, and Dentate Gyrus (DG), each with distinct expression profiles. E12.5 mouse eyes in embryo exhibit a known structure with a lens surrounded by neuronal retina cells, while mouse olfactory bulbs (OB) feature clear layer separation with various neuron types. These tissues, with their known morphological patterns and heterogeneous expression profile, serve as ideal reference samples for our sST benchmark studies. The use of diverse tissue types allowed us to assess how tissue type influences method performance, and each sample included a technical replicate for variability assessment (Figure 1a). A summary of the datasets in *cadasSTre* is given in Supplementary Table 2. Detailed protocols for obtaining regions of interest have been established and are available in the Methods section, facilitating reproducibility by other researchers. In total, we systematically evaluated 6 sST methods across 22 experiments from 3 tissue types.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

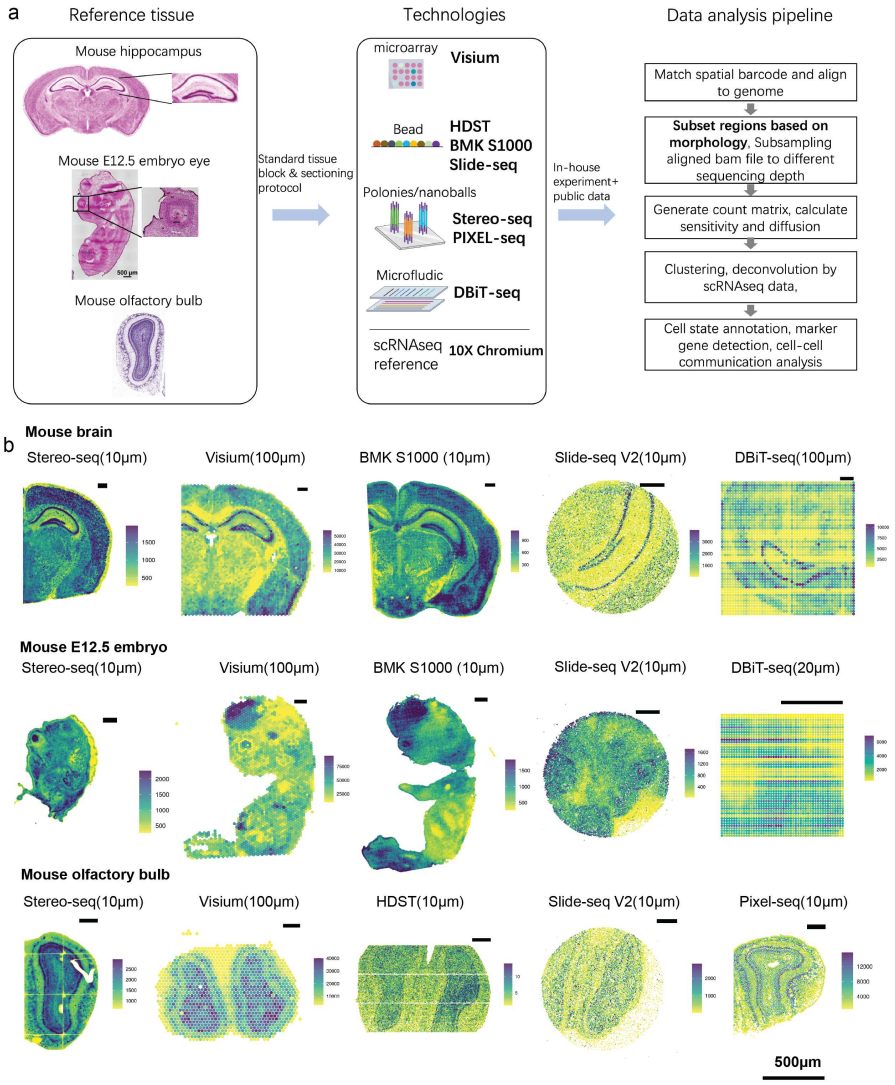


Fig. 1 Overview of experimental design and data processing pipeline. **a)** The experimental design involved the use of reference tissues, namely, adult mouse hippocampus, E12.5 mouse eye, and adult mouse olfactory bulb. We performed sST on these reference tissues using diverse technologies categorized by their distinct spatial indexing strategies. These techniques encompassed microarray-based methods (e.g., 10X Genomics Visium), bead-based approaches (such as HDST, BMKMANU S1000 (abbreviation: BMK S1000), and Slide-seq), polonies or nanoballs techniques (Stereo-seq and PIXEL-Seq), and microfluidic-based methodologies like DBiT-seq. Additionally, the reference tissues were subjected to single-nuclei RNA-sequencing (snRNA-seq) using the 10X platform. The *cadasSTre* datasets underwent a series of processing steps. Initially, spatial barcodes, their corresponding locations, and expression profiles were generated. Subsequently, reads within regions with known morphology were selectively retained, and downsampling was performed to mitigate the impact of sequencing depth variations. Count matrices were then generated for sensitivity and diffusion calculations. This was followed by cell state annotation and a comprehensive analysis of cell-to-cell communication. **b)** The visualization of total counts across the spatial dimension for datasets generated using each platform for reference tissues is shown. The distances from center to center, used in creating the plot, are presented alongside the name of each sST method. The length of the black bar in the visualization corresponds to a distance of 500 microns.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As outlined in the summary pipeline (Figure 1a right-hand panel), we next built a standard benchmarking pipeline to enable homogeneous data processing for sST methods and comparison in a fair way. Initially, spatial barcodes and their corresponding locations, together with expression profiles per spatial location were generated. Figure 1b provides an overview of total counts per spot for each sST method across various tissue types. Clear tissue patterns were observed across the samples. The summary of total counts is presented with varying spot sizes and the distances between spot centers. These differences are clearly depicted in Supplementary Figure 1a. They exhibit clear differences, as depicted in Supplementary Figure 1a. In Figure 1b, we have labeled the distances between spot centers, as we believe this metric better represents the platform's physical resolution, as opposed to using spot sizes. Stereo-seq and BMKMANU S1000 have distances between spot centers smaller than 10μm and spots in them are binned into a 10μm-sized spots for visualization.

We observed that Stereo-seq, Visium, and BMKMANU S1000 managed to capture nearly the entire right brain and the whole E12.5 embryo. In contrast, Slide-seq V2 could capture only a portion of the tissue due to its limited capture size (Supplementary Figure 1b,c). With DBiT-seq, the capture size varies depending on the width of the microfluidic channel, while also posing the risk of contamination across columns and rows in channels. We observed highly consistent tissue morphology among different methods in the H&E image shown in supplementary figure 2-4, which validates that our standard tissue handling and sectioning protocol could generate consistent results in different experiments.

Subsequently, we selectively retained reads within regions with known morphology, including the hippocampus in the mouse brain, and eyes in the E12.5 embryo. We then performed downsampling to address sequencing depth and sequencing cost variations. The purpose of downsampling is to normalize different methods to the same total number of sequencing reads to achieve equivalence in sequencing cost. Count matrices with downsampled data and full data were both then generated for sensitivity and diffusion calculations, followed by cell state annotation, maker gene detection, and analysis of cell-to-cell communication.

2.2 Molecule-capture efficiency

We obtained hippocampus and eye tissues from the adult mouse brain and E12.5 mouse embryo, as illustrated in Figure 2a. This was accomplished by manually delineating boundaries based on tissue patterns indicated by the spatial distribution of total counts and morphological information provided by H&E images. By selecting the same region, we ensure that our comparisons of sST sample performance were not influenced by varying locations within the tissues, as the number of counts from different parts of the tissue may exhibit variations.

Molecule capture efficiency was assessed in two ways. In selected regions, we either 1) used all the reads from that region, or 2) downsampled the data so that different samples had the same number of sequenced reads, which we refer to as *"downsampled data"* in the subsequent results.

Based on the downsampling results (Figure 2b,c), none of the sequencing runs, that ranged from 300 million reads (Visium) to 4 billion reads (Stereo-seq), reached

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

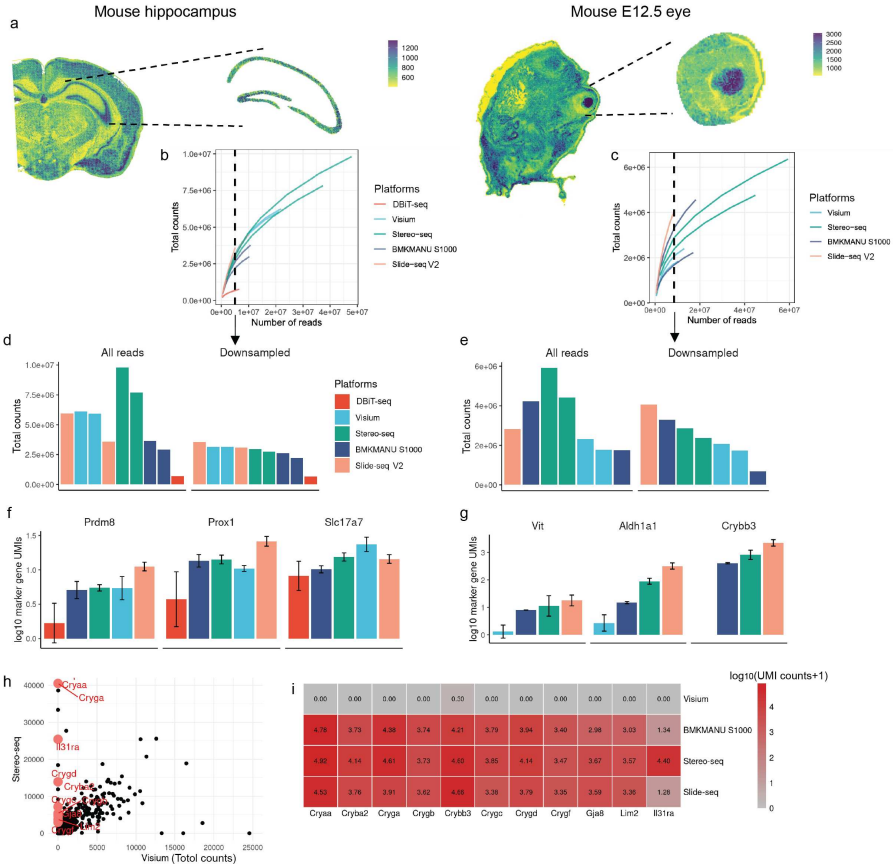


Fig. 2 Comparison of the sensitivity of data generated by different platforms. **a)** Schematic plot illustrating the extraction of regions with known morphology from fully processed samples of the adult mouse hippocampus and E12.5 mouse eye. Total UMI counts are presented as a function of stepwise downsampled sequencing depths for each platform. The data originates from **b)** mouse hippocampus and **c)** E12.5 mouse eye regions. A vertical dashed black line marks the read count used for generating the subsequent downsampled data. **d)** Total unique molecular identifier (UMI) counts were computed for selected regions using all reads and downsampled data for the mouse hippocampus. **e)** Total UMI counts for selected regions using all reads and downsampled data for the E12.5 mouse eye. **f)** The summed UMI counts for marker genes across five individual 100μm×100μm regions in the mouse hippocampus, along with mean and standard deviation. **g)** The summed UMI counts for marker genes across five individual 100μm×100μm regions in the E12.5 mouse eye, along with mean and standard deviation. **h)** Total UMI counts of detected genes are compared between Visium (x-axis) and Stereo-seq (y-axis). Each dot represents a gene, shown in black. Genes that display expression at the 90th percentile with Stereo-seq but are at the 10th percentile in Visium are highlighted in red and labeled with their gene symbols. **i)** A heatmap displays the $\log_{10}$-transformed expression of genes that are specifically not captured by Visium but are captured by Stereo-seq for E12.5 mouse eyes.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

saturation. This observation suggests that sST data requires considerably more reads for optimal performance, with the potential for increased sensitivity.

Next, we compared the sensitivity of each sST method by summing the total counts within the selected regions. Stereo-seq had many more sequencing reads for the same region compared to other platforms, resulting in higher total counts when all reads are used (Figure 2d and e, left panel). However, when the effect of sequencing depth is controlled, Slide-seq V2 data consistently demonstrated higher sensitivity than other platforms, in both the eye and hippocampus. This observation aligns with the saturation plot results (Figure 2b,c), where the total counts from Slide-seq V2 data exhibited a greater increase with increasing read number. In contrast, DBiT-seq data consistently showed the lowest sensitivity (Figure 2d and e, right panel). Additionally, the impact on the relationship between the number of counts and features per spot is more pronounced in Stereo-seq data when comparing downsampled results to the results obtained using all reads (Supplementary Figure 5).

To provide a more detailed assessment of the differences in sensitivity among selected sST methods, we proceeded to measure the RNA content of marker genes known to be expressed in specific regions using downsampled data. In CA3 of the hippocampus, we compared the sum of counts for *Prdm8*, *Prox1*, and *Slc17a7* within 100μm$\times100\mu$m regions (selected based on the largest physical resolution value among the sST methods applied). Our findings revealed that the expression patterns of these marker genes mirrored the total count results, with Slide-seq exhibiting the highest sensitivity and DBiT-seq displaying the lowest (Figure 2f). In the case of E12.5 mouse eyes, we compared the sum of counts for *Vit*, *Crybb3* (lens), and *Aldh1a1* (neuron retina) within 100μm$\times100\mu$m regions. Similarly, Slide-seq demonstrated the highest sensitivity, while Visium did not generate as many counts for marker genes in regions where their expression was expected (Figure 2g). Furthermore, through pairwise comparisons, we identified genes consistently expressed in the lens across all sST methods, except for data generated by Visium (Figure 2h, Supplementary Figure 6a), including *Crybb3* and *Cryaa* (Figure 2i). Importantly, this inconsistency did not appear to be attributed to the preprocessing pipeline and gene annotations (Supplementary Figure 6b), indicating a systematic gene-specific bias of Visium towards the lens. In an attempt to correlate this bias with various gene attributes, including gene biotypes, length, and GC content percentage, we discovered that these biased genes, which exhibit low expression in Visium, are predominantly protein-coding. Moreover, no significant bias was detected in terms of GC content or gene length (Supplementary Figure 7).

In our investigation of the mouse OB, after annotation, we assessed the sensitivity of selected sST methods, considering layers with varying densities of total counts. Notably, PIXEL-seq exhibited the highest sensitivity, while HDST demonstrated the lowest sensitivity at a 10μm physical resolution (Supplementary Figure 8).

2.3 Molecule-lateral diffusion

In addition to molecule capture sensitivity per unit area, another crucial quality parameter is the spatial accuracy of mRNA detection. To assess such accuracy, we employed two analysis methods to measure molecule lateral diffusion: 1) Plotting the

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

intensity profile of a specific gene across the selected region. 2) Quantifying the distance between the left width at half-maximum (LWHM) of intensity in the chosen region [19], focusing on histological structures where the expression of the selected gene should exhibit a significant difference—showing high expression in one part of the region and minimal to no expression in the rest. These analyses were conducted using count data generated from all reads.

In our evaluation of the OB, we selected *Slc17a7* as the marker gene due to its expected expression specifically in Mitral and Tufted (M/T) cells, which form distinct layers [20] and in glutamatergic neurons located at the base of the glomerular layer (GL) [21]. We confirmed *Slc17a7*'s expression at these locations via in situ hybridization (ISH) from the Allen Brain Atlas [22]. In this analysis, our focus was on *Slc17a7*'s expression in M/T cells. As illustrated by the expression plots of *Slc17a7* in each sST dataset (Figure 3a, left panel, Supplementary Figure 9), we specifically selected regions (N=6) where *Slc17a7* was expressed in the middle. Our observations, based on intensity plots and LWHM measurements, revealed significant lateral diffusion by Stereo-seq V1 of *Slc17a7* in the OB. Notably, Slide-seq V1.5 and PIXEL-seq exhibited relatively better control over this diffusion (Figure 3b-d left panel).

In our analysis of the brain, we selected *Ptgds* as the marker gene, as it has been confirmed by ISH to be specifically expressed in a particular location within vascular cells [23] (Supplementary Figure 10a,b). By examining the expression plots of *Ptgds* and its intensity plots along with LWHM measurements, we noted severe lateral diffusion in the Stereo-seq dataset. In contrast, Slide-seq V2, followed by BMKMANU S1000, exhibited better control over such lateral diffusion issues (Figure 3a-d middle panel, Supplementary Figure 10c). We further validated these observations by conducting a diffusion analysis on downsampled Stereo-seq data, confirming that the challenge of lateral diffusion persisted despite a lower sequencing depth compared to other sST datasets. (Supplementary Figure 10d-f) This suggests that downsampling could not resolve the lateral diffusion issue for Stereo-seq data.

For our examination of eye tissue, we selected *Pmel* as the marker gene due to its specific expression in melanocytes, which encircle the lens and form a circular pattern [24]. Interestingly, in this context, Stereo-seq demonstrated the best control over lateral diffusion, followed by Slide-seq V2. (Figure 3a-d right panel, Supplementary Figure 11) This observation contrasts with our findings in the other two tissue types, indicating that tissue type exerts a considerable influence on the diffusion process. Diffusion is greatly impacted by permeabilization time. We have showed in our permeabilization optimization experiment (Supplementary Figure 2B,3B,4B) that different permeabilization time significant impact the diffusions.

2.4 Clustering and cell annotation across technologies

We next applied selected sST methods to gain insight into biological questions where higher capture sensitivity and well-controlled diffusion are important.

We selected E12.5 mice eyes, known for their distinctive structure featuring the lens, surrounded by the retina, and then melanocytes [25–27].

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

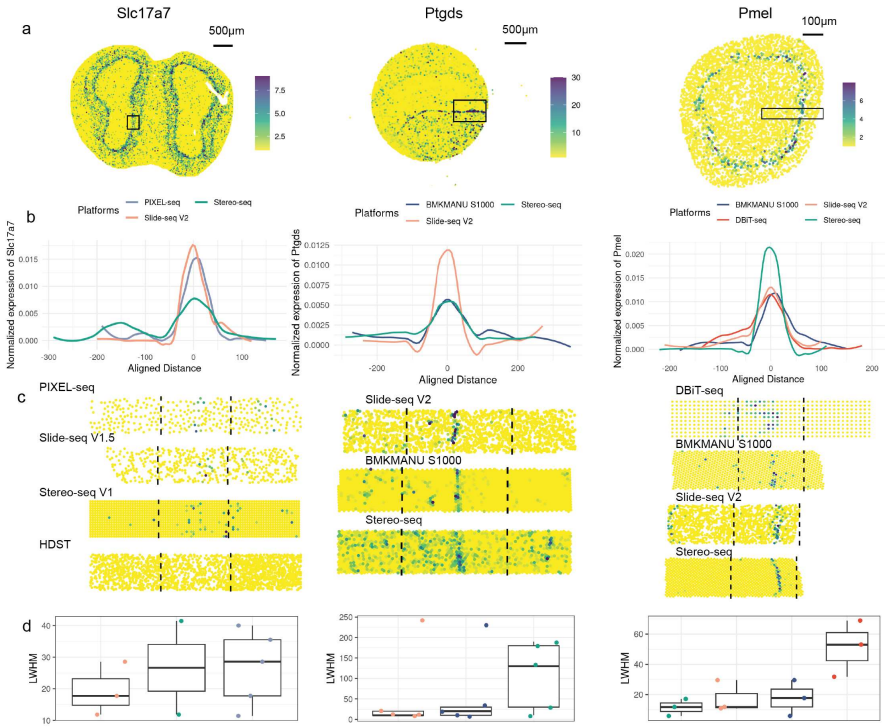


Fig. 3 Comparison of diffusion of data generated by different platforms. **a)** Expression patterns of selected marker genes known to be highly expressed in specific regions. These markers include *Slc17a7* in the mouse olfactory bulb (left panel), *Ptgds* in the mouse brain (middle panel), and *Pmel* in the E12.5 eye (right panel). The plots are based on raw count values. Black boxes indicate the selected regions used for diffusion calculation. **b)** Expression levels of the aforementioned marker genes (from panel a) are aggregated for every 10μm along 50μm in the olfactory bulb, 500μm in the brain, and 300μm in the eyes, as shown in a). UMI counts are averaged across modalities, normalized for each platform, and presented in a density plot with the area under the curve set to 1 (details in Methods). **c)** Expression level of the marker genes as mentioned above (from panel a) within selected modalities are provided, with black dashed lines delineating the boundaries used for diffusion calculations. **d)** The left half-width half maximum (LWHM) of the profile was then calculated for each gene (from panel a) in each modality and displayed in boxplots. Each dot represents the LWHM for a given modality. Modalities for which LWHM could not be calculated were removed.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.4.1 Annotating regions by clustering results

With the basic knowledge of general cell states within the eye area, our next objective was to annotate the spots captured by selected sST platforms using various clustering methods. We aimed to determine whether we could consistently identify more detailed and coherent cell subsets across all samples.

Such resulting annotations of cell subsets not only served as a benchmark for evaluating the methods employed in this study but also provided valuable insights into the intricacies of cell states within the developing eye of E12.5 mice.

Before delving into our comparative analyses, Figure 4a showcases our findings about the cell subsets that we expected to observe within an E12.5 mouse eye. In this tissue, the anticipated morphological structure unfolds from the innermost space, housing the lens and lens vesicle, which are enveloped by neuronal retina cells forming distinct subsets in specific locations. The neuronal retina cells are encircled by melanocytes, with the rostral side hosting corneal mesenchyme, while the caudal side is composed of epithelial cells. These annotations provided us with a foundation for our subsequent evaluations and comparative assessments.

2.4.2 Comparison between clustering results

In our comparative analysis of clustering results, we conducted evaluations from two perspectives:

1) Clustering Methods: We systematically employed three distinct clustering methods: *Seurat* [28], which exclusively considers transcriptomic profiles, and *DR.SC* [29] and *PRECAST* [30], which incorporate spatial information alongside gene expression data. Recent benchmark studies have reported that methods leveraging spatial location information demonstrate promising clustering results in specific datasets. However, they do not consistently surpass or exhibit greater robustness compared to methods that solely rely on gene expression data [31]. Our observations align with this conclusion, with *Seurat* consistently demonstrating robust and stable performance compared to the other 2 methods in detecting expected cell subsets as shown in Figure 4a, left panel, and Supplementary Figure 12.

2) sST Methods: In our comparisons between sST platforms, we focused primarily on the results generated by *Seurat*. We annotated spots for each sST method individually (Figure 4a and Supplementary Figure 13). Our analyses unveiled variations in the ability of different methods to consistently identify the expected cell subsets. Notably, Slide-seq V2 and Stereo-seq data delivered a nice separation of spots for comprehensive subset annotations, successfully capturing all anticipated subsets. Conversely, BMKMANU S1000 data faced challenges in cell state detection, particularly in identifying melanocytes. This difficulty may stem from the pronounced lateral diffusion observed in BMKMANU S1000 data (as depicted in Figure 3c,d right panel), making it difficult for clustering methods relying solely on expression profiles to retain this specific cell type. On the other hand, Visium data faced certain limitations in detecting the anticipated cell subsets. These challenges were primarily attributed to the relatively low physical resolution and hence a restricted number of spots available in the eye area (approximately 75 spots in total for each sample). Within each of these

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

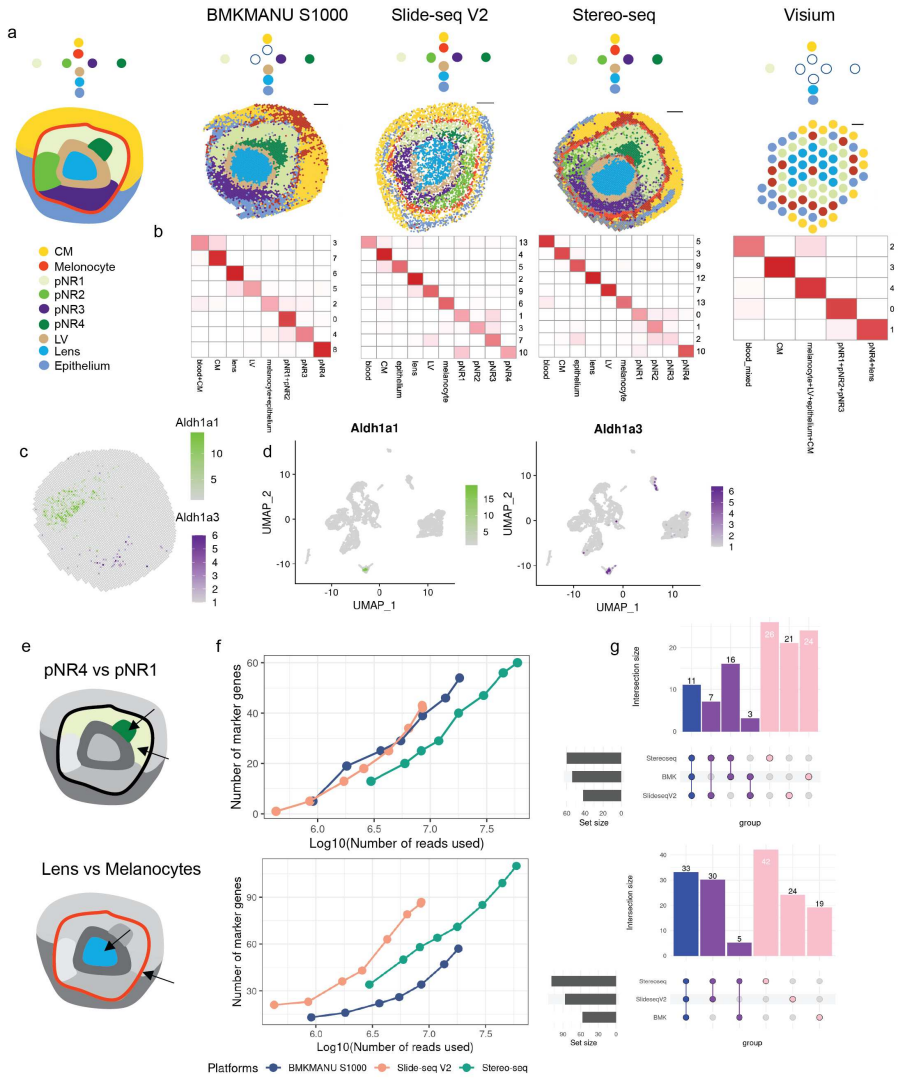


Fig. 4 Comparison on downstream performance. **a)** Expression profiles generated by each platform were processed to obtain clustering results. Known cell types and states are colored in the left-most panel. Additionally, a schematic plot represents the expected cell states, arranged from outer space to inner space and from top to bottom. On the right-hand side, clustering results are presented, with spots color-coded by annotated cell states depicting the identifiable cell states. CM represents corneal mesenchyme; pNR represents presumptive neural retina; LV represents lens vesicle. **b)** Clustering was conducted on downsampled eye data from each platform, with an equal total read count across platforms in the eye area. The correspondence between annotations obtained from clustering based on all reads and clustering based on downsampled data is visualized in a heatmap. The number of spots in this correspondence is presented after $\log_{10}$ transformation without scaling. **c)** With Stereo-seq data as an example, spatial expression profiles of *Aldh1a1* and *Aldh1a3*, which are expressed in pNR2 and pNR3 are shown at 10 μm resolution. The number of spots with the expression of *Aldh1a1* above 0 is 1,329, and of *Aldh1a3* above 0 is 217. **d)** Expression profiles of *Aldh1a1* and *Aldh1a3* in snRNA-seq data are presented, The number of cells having an expression of *Aldh1a1* above 0 is 93, and of *Aldh1a3* above 0 is 161. **e)** An overview of cell states compared in the marker gene detection analysis, with pNR4 and pNR1 highlighted in the top panel, and lens and melanocytes highlighted in the bottom panel. **f)** Number of marker genes detected with different numbers of reads used for each sST method in the comparison between pNR4 and pNR1 (top panel). The same analysis is applied to the lens and melanocytes in the bottom panel. **g)** In the top panel, an Upset plot displays the intersection of marker genes obtained by different sST methods using all reads for the pNR4 and pNR1 comparison. Genes shared among all three platforms are denoted in blue, those shared between two platforms are in purple, and uniquely obtained genes are represented in pink. The bottom panel presents a similar analysis for lens and melanocytes.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

234 100μm$\times100\mu$m spots, cells were mixed, making identifying intricate cell subsets more
235 challenging (Figure 4a, right panel).

236 2.4.3 Influence of downsampling on clustering results

237 We observed that sequencing depth influences the total counts of spatial transcrip-
238 tomic data (Figure 2b-e). In light of this, we set out to investigate how sequencing
239 depth impacts clustering results. Our exploration of clustering results on downsam-
240 pled data involved two key aspects: 1) We assessed the correspondence between the
241 downsampled data and the full data. 2) We calculated entropy measures for cluster
242 purity (ECP) and accuracy (ECA) based on the clustering results obtained with the
243 full data as a reference for downsampled data generated at various proportions as
244 shown in Figure 4b. Remarkably, we discovered that the downsampled data was capa-
245 ble of detecting nearly all of the cell subsets identified by the full data (Figure 4b,
246 Supplementary Figure 14). However, when evaluating ECP and ECA across differ-
247 ent proportion values, we observed relatively high values, signifying a notable degree
248 of inconsistency. This inconsistency could be attributed to the fact that while the
249 majority of cell subsets effectively formed distinct clusters, a portion of cells grouped
250 into different clusters, notably between cells from different subsets of neuronal retina
251 cells. This effect was particularly pronounced in subsets between populations such as
252 lens and lens vesicles; 4 neuron retina subsets (Figure 4b), which are more similar in
253 expression profiles.

254 2.4.4 Comparison between sST data and snRNA-seq data

255 We consistently observed well-patterned expression of *Pmel*, *Crybb3*, *Atoh7*, *Enfa5*,
256 *Aldh1a1*, and *Aldh1a3* across all sST datasets. These genes were selected as they serve
257 as markers for specific cell types, such as melanocytes, lens, presumptive neural retina
258 (pNR)2, and pNR3 (Figure 4c, Supplementary Figure 15). Although the absolute
259 position for some of the region is not exactly the same but their relative position
260 remains consistent, such as *Aldh1a3* located in rostral ragion of the retina layer while
261 *Aldh1a1* patterned towards caudal region (Supplementary Figure 16).
262 In addition, we obtained snRNA-seq data with eye region sectioned as input. How-
263 ever, a limited number of cells were found to express the aforementioned genes, and
264 these cells were primarily clustered in the lower corner of the UMAP plot (Figure
265 4d, Supplementary Figure 15e). Unfortunately, we were unable to further categorize
266 this small subset into more detailed subgroups, as was the case with the sST data.
267 Interestingly, we also noted that *Crybb3*'s expression values in the snRNA-seq data
268 were relatively lower than expected. This is in line with our earlier observation that
269 *Crybb3* was not found to be expressed in the eye area captured by Visium, indicating
270 a potential capture bias associated with 10X technology.
271 While snRNA-seq data may not capture as many cells in the eye area as sST
272 methods do, it serves as a useful reference dataset for annotating the sST data. As
273 illustrated in Supplementary Figure 17, the integration of snRNA-seq data with sST
274 data using *Seurat* aided the annotation of sST data. For instance, it improved the
275 annotation of epithelium cells in Stereo-seq data, which had been relatively challenging
276 due to an unknown cluster of cells with mixed expression profiles. This cluster was

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

better resolved using the projection of snRNA-seq epithelium cells (Supplementary Figure 17a). Additionally, the projection facilitated the separation of melanocytes and epithelium cells in BMKMANU S1000 data (Supplementary Figure 17b).

Another issue that deserves mention is the susceptibility of sST technologies to blood contamination, which is often introduced during the tissue preparation and sectioning process and is difficult to avoid. In contrast, snRNA-seq can mitigate this effect using microfluidic techniques. We used the *Hba-a1* gene as an example to evaluate the influence of blood contamination in these sST methods. Our findings revealed that Visium, followed by BMKMANU S1000, were significantly impacted by blood contamination, with all Visium spots and 70% of BMKMANU S1000 spots expressing *Hba-a1*. In contrast, Stereo-seq data exhibited a relatively similar level of blood contamination compared to snRNA-seq, and Slide-seq V2 had the lowest amount of blood contamination (Supplementary Figure 18).

2.5 Marker gene detection across technologies

Prior studies have underlined the effectiveness and robustness of using a Wilcoxon rank-sum test when identifying marker genes [32]. We employed this test within *Seurat* to find marker genes between clusters. Analysis of top marker genes reveals technology-specific biases in the selection of these markers. For instance, *Pax6*, a transcription factor known as a master regulator of neural lineages, particularly in the retina [33], exhibited variations in representation among different technologies. Specifically, Stereo-seq data highlighted *Pax6* exclusively in the pNR3 cluster, whereas Slide-seq V2 and BMKMANU S1000 data depicted *Pax6* expression across the entire neural retina (pNR1-4) (Supplementary Table 3), consistent with existing literature. This observation underscores the influence of technology choice on the identification of top markers for specific cell types or clusters. Similarly, disparities were observed in the expression of Hes genes in progenitors of the neural retina and *Sox2* in pNR1 and pNR2 (Supplementary Table 3).

The analysis of clustering results on downsampled data has shown that general cell subsets can still be adequately retained even with fewer sequencing reads. However, it appears that a few subsets, particularly those sharing similar expression profiles are a challenge to be clearly separated. To further investigate the effects of downsampling, we compared the marker genes identified in downsampled data with those in the full dataset. We selected 2 pairs of cell subsets to compare the detection performance for cell subsets with relatively similar expression profiles and those that are more distinct. We employed conducted the marker gene detection in two scenarios: 1) pNR4 and pNR1, which exhibit higher similarity, and 2) lens and melanocytes, which have lower similarity, to identify marker genes (Figure 4e).

Our observations revealed that the number of marker genes increased as the number of reads increased in both pairs of comparisons. Notably, the increase in marker genes was more pronounced with deeper sequencing, as illustrated in Figure 4f and Supplementary Figure 19. The ranking of marker gene detection performance across sST methods aligns with the results depicted in Figure 2d in the comparison between cell subsets with relatively distinct expression profiles. In particular, Slide-seq V2

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

exhibits higher sensitivity (Figure 4f, bottom panel). Furthermore, our analysis identified a set of genes consistently identified as marker genes across different platforms in each of the comparison pairs. However, each platform exhibited a great number of unique marker genes as well (Figure 4g).

Cell-to-cell communication was applied afterward, but no consistent results could be found across the communication methods applied including *CellChat* [34] and *CellPhoneDB v4* [35] and sST methods (Supplementary Figure 20).

3 Discussion

Evaluating spatial transcriptomic methods is more challenging than evaluating scRNA-seq methods. First, it is harder to design a reference tissue for spatial transcriptomics. For scRNA-seq, one could use cell line mixtures/PBMC samples [8, 36, 37], or even purified and diluted mRNA to obtain consistent inputs for different technologies [36]. For spatial transcriptomics, if we use genuine tissues with clear cell type and gene expression patterns, the position and ground truth are then less obvious and limited by our understanding of reference tissues. Second, the measurements are not performed on the same unit. For methods like Visium, the diameter of a spot is larger than 50 microns resembling a mini-bulk RNA-seq. For methods such as Stereo-seq, the spot size is sub-micron, which is much smaller than a single cell.

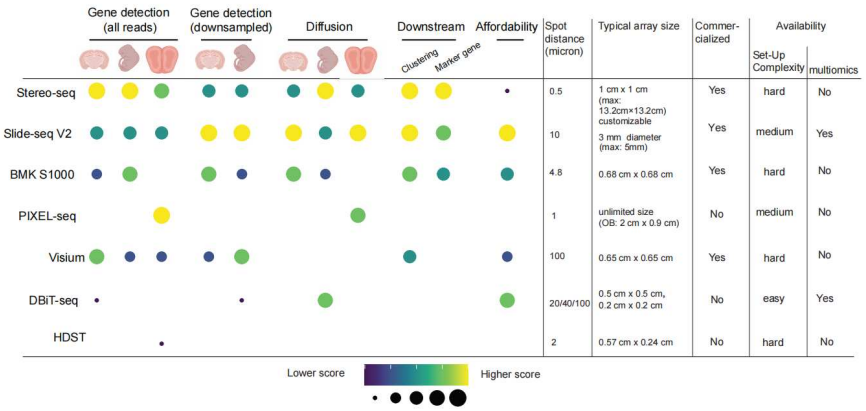


Fig. 5 Summary of results and characteristics of sST methods. The sST methods have been ranked based on their performance in the specified categories, with the highest-performing methods positioned at the top. In the left panel, each ranking is represented by color and spot size. In the right panel, essential characteristics of the sST methods examined are outlined. Set-up complexity represent how difficult it is to build the method from scratch.

We carefully designed our benchmarking study to address these challenges. For the first problem, we selected a set of reference tissues with the following criteria: 1) the tissue should be from the widely used model organism, accessible at most research institutes; 2) the tissue should have stable cell-type patterns and specific marker gene expression; and 3) the reference region should have clear morphology that is easy

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to find in sectioning. Together with the reference tissue, we developed a sectioning protocol to help people reproduce and generate comparable data in the future. For the second challenge, we employed multiple benchmarking metrics and workflows to compare different methods on the same tissue region. We used both all-reads and downsampled data in our comparisons. Downsampling was implemented to mitigate the impact of variations in sequencing depth and cost, however as this may not bring all methods to the same standard as the required number of reads to achieve satisfactory results may differ, we also used all reads in the analysis as complementary results.

In this study, we generated *cadasSTre*, a cross-platform dataset for sequencing-based ST benchmarking that allowed systematic evaluation of 6 sST methods across 22 experiments. We compared various aspects of data from basic metrics to downstream analysis, ranging from sensitivity, and diffusion to clusterability and marker gene detection (Figure 5). Our results suggest spatial transcriptomics requires more sequencing to reach saturation and data generated in this study are well below the saturation level. Technologies such as Stereo-seq require much more sequencing cost to generate high-quality data. Stereo-seq shows the best capture efficiency with raw sequencing depth while Slide-seq v2 gives the best capture efficiency with normalized sequencing depth. Interestingly, we found unexpected gene capturing bias on the Visium platform, with marker genes consistently captured by other technologies not showing up in the Visium data. Considering Visium is the most widely used commercial platform, it is important to further verify its gene-capturing bias on other tissues.

The spot size has become an important metric as a surrogate of the resolution for each method. However, in this study, we highlighted diffusion as a key factor that affects the actual resolution. We found different technologies show distinct diffusion profiles on different tissues. For example, Stereo-seq gives excellent diffusion control on mouse embryo tissue but has much stronger diffusion-induced artifacts in mouse brains. Permeabilization time has a great impact on the molecule diffusion of samples and the tissue-to-tissue variations in diffusion could be a result of it. Although some technologies have sub-micron spot sizes, their real resolution would never reach the same level due to limited sensitivity and high diffusion. Further development of sST would benefit from increased diffusion control and improved assay to determine the permeabilization condition and time, which is a key factors in sST technology development.

Overall, our study generated the first systematic benchmarking scheme of sST methods. Although we strive to make the most of this study, there remain several areas that could be further improved. Also, our understanding of mouse eye development is still limited, making it hard to construct a ground truth in mouse embryo data. The benchmarking dataset generated in this study could be used to further compare computational tools, but it is important to use diverse tissue and data to develop more generalized spatial tools. Although the goal of this study is not to comprehensively benchmark computational tools, we found that clustering tools designed for spatial data may not give better performance than clustering methods for single cells, which agrees with a comparison study [31]. We also found that cell annotations derived from single-cell references may not yield detailed cell states and that clustering derived from

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spatial data could give complementary results that were sometimes better at resolving rare cell states with spatial patterns. It is important to consider both analyses with and without single-cell references in annotating spatial data.

The sST field is rapidly evolving and the performance of each technology is likely to change with time as they are further optimized. Continuing evaluation is required to keep pace with this fast-moving field. Spatial multi-omics methods are still in their early stages of development [38–41], and new technologies need to be established. Therefore, we believe a community-driven spatial benchmarking league would be beneficial to the future of spatial multi-omics. Our benchmarking efforts highlight the benefits and current issues in the sST field, set up standards for comparing sST methods, and take the first step towards benchmarking spatial multi-omics technology.

4 Methods

4.1 Sample preparation

4.1.1 Reference sample

All relevant procedures involving animal experiments presented in this study are compliant with ethical regulations regarding animal research and were conducted under the approval of the Animal Care and Use Committee of Westlake University (license number AP#23-111-LXD). Animals were group housed with a 12-hour light-dark schedule and allowed to acclimate to their housing environment for two weeks post arrival. Mouse embryos were collected from pregnant C57BL/6J female mice at embryonic day 12.5 (E12.5). Mouse brain was dissected from 8-week-old C57BL/6J male mice.

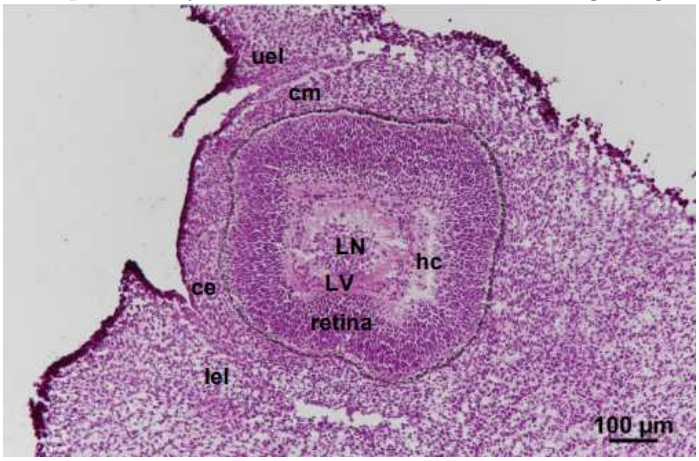
4.1.2 Sample preparation, embedding, sectioning, and histological testing

Mouse embryo

1) E12.5 pregnant female mice were anesthetized with carbon dioxide, and the whole uterus was collected and washed 3 times in ice-cold DPBS; 2) The uterus was separated under a stereo microscope, and each embryo was numbered and photographed with a Motorized Fluorescence Stereo Zoom microscope (ZEISS, Axio Zoom V16); 3) A yolk sac was collected to extract DNA for genotyping (identification of sex); 4) Using dust-free paper to gently wipe the liquid on the surface of the embryo, the embryo was rinsed with ice-cold Tissue-Tek OCT (Sakura, 4583), and then moved to the encapsulation box with ice-cold OCT; 5) Air bubbles were carefully removed with the syringe, and the embryo was placed in sagittal position with tweezers; 6) The location of the embryonic eye was circled and marked the orientation of the embryo, then tissues were transferred to a -80℃ freezer, snap-frozen, and stored; 7) Embryos of average size and normal phenotype were selected for subsequent cryosectioning and sequencing (note: the embryos used in our benchmarking analysis came from a litter of mice); 8) Before sectioning, the tissue block was removed from the -80℃ freezer and placed in a cryostat (Leica, CM1950) to balance for at least 30 min; 9) The tissue block was smoothly glued to the sample head so that the embryo was sectioned in a

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sagittal position. If necessary, the angle can be fine-tuned so that the blade section is strictly parallel to the cross-section of the tissue block; 10) Cryosections were cut at a thickness of 10 μm, both the left eye and right eye can be collected; 11) The structure of the sequenced cryosections is shown in the following image:



12) H&E staining procedure: cryosections were balanced at room temperature for 30 min, and then fixed with 4% PFA for 3 min. Then, the sections were washed with ddH_2O for 2 min, stained with hematoxylin for 6 min, washed with ddH_2O, stained with eosin for 2 min, washed with ddH_2O. After that, sections were gradient dehydrated (75% ethyl alcohol for 1 s, 85% ethyl alcohol for 1 s, 95% ethyl alcohol for 1 s, 100% ethyl alcohol for 1 s, 100% ethyl alcohol for 1 min), cleared (xylene for twice), and sealed with Permount TM Mounting Medium after airing. Finally, the figure was scanned using a Motorized Fluorescence Microscope (Nikon, Ni-E).

Mouse brain

1) 8-week-old male mice were anesthetized with carbon dioxide and decapitated; 2) The whole brain was rapidly dissected, numbered, and photographed with a Motorized Fluorescence Stereo Zoom microscope; 3) Using dust-free paper to gently wipe the liquid on the surface of the brain, the brain was rinsed with ice-cold Tissue-Tek OCT (Sakura, 4583), and then moved to an encapsulation box with ice-cold OCT; 4) Air bubbles were carefully removed with the syringe, and the brain was placed properly with tweezers; 5) The location of the hippocampus was circled and marked the orientation of the brain, then tissues were transferred to a -80℃ freezer for snap-frozen and storage; 6) Brains of average size and normal phenotype were selected for subsequent cryosection and sequencing; 7) Before sectioning, the tissue block was taken out from -80℃ freezer and placed in a cryostat (Leica, CM1950) to balance for at least 1 h; 8) The tissue block was smoothly glued to the sample head, and the cerebellum was oriented towards the experimenter so that the brain was sectioned in a coronal position. If necessary, the angle can be fine-tuned so that the blade section is strictly parallel to the cross-section of the tissue block; 9) Cryosections were cut at a thickness of 10 μm; 10) The structure of the sequenced cryosections is shown in the following image:

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11) H&E staining procedure: cryosections were balanced at room temperature for 30 min, and then fixed with 4% PFA for 3 min. Then, the sections were washed with ddH_2O for 2 min, stained with hematoxylin for 6 min, washed with ddH_2O, stained with eosin for 1 min, washed with ddH_2O. After that, sections were gradient dehydrated (75% ethyl alcohol for 1 s, 85% ethyl alcohol for 1s, 95% ethyl alcohol for 1 s, 100% ethyl alcohol for 1 s, 100% ethyl alcohol for 1 min), cleared (xylene for twice), and sealed after airing. Finally, the figure was scanned using a Motorized Fluorescence Microscope (Nikon, Ni-E).

4.1.3 In-situ imaging with padlock probes

To validate the expression of marker genes, we performed in-situ hybridization and imaging following a simplified version of targeted ExSeq [42]. More specifically. we used 4 fixed barcode regions for distinct fluorescent probes (FAM6, CY3, TXRED, CY5) so we could detect at most 4 genes at the same time without performing muli-round imaging for in-situ sequencing. The tissue was sectioned on Leica CM1950 Cryostats, with 10-micron sections placed on the CITOTEST adhesion microscope slides. The section was then fixed with 4% formalin for 15 minutes at room temperature and washed two times with PBS. Permeabilization of tissue was done with ice-cold 70% EtOH overnight at -20℃. RNase inhibitor (Lucigen) was added at 0.4 U/μl throughout the incubation until the rolling cycle amplification (RCA) was done. The padlock probe was diluted at a final concentration of 5nM per probe, in wash buffers with 2XSSC and 20% formamide. Hybridization was done overnight at 37℃, then washed with the same wash buffer (2XSSC and 20% formamide) 3 times for 15 minutes each, followed by washing with PBS for 15 minutes at 37℃. SplintR ligase (NEB, M0375) was used for probe ligation at 37℃ for 2.5 hours. RCA was performed at 30℃ overnight using Phi29 enzyme mix (NEB, M0269L). fluorescent probes hybridization was done at 2X SCC and 10% formamide buffer mix, diluting the probes at 100 μM and incubate at 37℃ for 1 hour. Imaging was performed with a NIKON A1 confocal microscope with 10X objectives and 2×2 image stitching.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.1.4 Protocols used in different ST methods

See Table1.

4.2 Data processing

4.2.1 Preprocessing

We preprocessed fastq files from multiple platforms using their respective preprocessing pipeline (where provided) and updated *scPipe* to allow sample processing with unified functions for data from different sST technologies starting from fastq files.

Mouse GRCm39 was used as a reference for alignment in each of the pipelines for locally generated data.

Visium data were processed with spaceranger (v2.1.0), and aligned with STAR 2.7.10b.

BMKMANU S1000 is a technology developed by BMKGENE (https://www.bmkgene.com/). Similar to HDST, it uses barcoded beads deposited on patterned array. Data were processed with BSTMatrix (v2.3.j), and aligned with STAR 2.7.10b.

Slide-seqV2 generated bam files of pucks of mouse eyes (Puck_190926_03) and hippocampus (Puck_191204_01 and Puck_200115_08) were downloaded [15].

Stereo-seq data were processed with SAW (v6.1).

DBiT-seq data underwent initial filtering using a predefined barcode list, and subsequently, fastq file 1 was restructured to adopt the format of spatial barcodes followed by UMIs. The processed data were further analyzed using *scPipe* (v2.0.0) to generate spot-by-gene count matrices.

4.2.2 Selection of region of interest and downsampling

After acquiring count matrices and associated location data for the datasets generated by the aforementioned sST platforms, we aimed to mitigate the impact of variable sequencing depths and costs. To achieve this, we extracted spots located within consensus regions in reference tissues, specifically the hippocampus in the brain and the eye in mouse embryos, for comparative analysis.

Spot selection was guided by histological images (H&E images), feature plots of total counts, and marker genes (*Pmel* for the eye and *Slc17a7* for the brain). These boundaries were meticulously delineated manually.

Subsequently, spots falling within the predefined boundaries for each sample were isolated and used for downsampling. An equal number of reads were chosen within selected spots for both eye and hippocampus data, based on the readID of the selected reads. These selected reads were then isolated from the BAM files generated in the aforementioned pipelines.

The BAM files were further processed to demultiplex based on spatial barcodes and quantified into matrices using UMIs and aligned gene information, using new functions introduced in *scPipe*. In addition to generating count matrices with an equivalent number of reads across platforms, we also processed reads from each platform in specific proportions using *scPipe*.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	Stereo-seq	BMKMANU S1000	10X Genomics Visium	DBiT-seq
Tissue Optimization (Pre-Experiment)	1 Tissue Fixation None None 2 Permeabilization Time Course 3 TRITC cDNA Synthesis 4 Tissue Removal 5 Chip TRITC Imaging	1 Tissue Fixation 2 Tissue HE Staining 3 Tissue brightfield imaging 4 Permeabilization Time Course 5 Cy3 cDNA Synthesis 6 Tissue Removal 7 Slide Cy3 Imaging	1 Tissue Fixation 2 Tissue HE Staining 3 Tissue brightfield imaging 4 Permeabilization Time Course 5 TRITC cDNA Synthesis 6 Tissue Removal 7 Slide TRITC Imaging	**None**
Permeabilization Time	**Brain: 12 min** **Embryo: 18 min**	**Brain: 15 min** **Embryo: 6 min**	**Brain: 12 min** **Embryo: 6 min**	
Spatial Expression(Formal–Experiment)	**1 Cryosection on the Stereo Chip** **2 Fixation: Methanol for 30 min** 3 Tissue ssDNA Staining 4 Tissue FITC Imaging 5 Tissue Permeabilization 6 Reverse Transcription 7 cDNA Release 8 cDNA Amplification 9 cDNA Cleanup and cDNA QC **HE: Adjacent Tissue Section**	**1 Cryosection on the S1000 Gene Expression Slide** **2 Fixation: Methanol for 30 min** 3 Tissue HE Staining 4 Tissue brightfield Imaging 5 Tissue Permeabilization 6 Reverse Transcription 7 Second Strand Synthesis 8 cDNA Amplification 9 cDNA Cleanup and cDNA QC **HE: The Same Tissue Section**	**1 Cryosection on the Visium Gene Expression Slide** **2 Fixation: Methanol for 30 min** 3 Tissue HE Staining 4 Tissue brightfield Imaging 5 Tissue Permeabilization 6 Reverse Transcription 7 Second Strand Synthesis 8 cDNA Amplification 9 cDNA Cleanup and cDNA QC **HE: The Same Tissue Section**	**1 Cryosection on poly-L-lysine coated glass slide** **2 Fixation: formaldehyde 20 min** 3 Incubations with ADTs 4 RT with Barcode A Oligo-dT 5 Ligation with Barcode B 6 Tissue Lysis 7 Template Switch 8 cDNA Amplification 9 cDNA Cleanup and cDNA QC **HE: Adjacent Tissue Section**
Library Construction	cDNA was fragmented and amplified. Then the products were filtered twice (0.6X+0.2X)	To construct the sequencing library, cDNA was fragmented, end-repaired, and A-tailed. Then the adaptor was ligated so that dual index PCR could amplify and distinguish these samples by different index sequences. The distribution of the main peak was between 200 to 600bp and the libraries were sequenced in the same sequencing run.		

Table 1 Protocols used in different ST methods.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.2.3 Sensitivity and diffusion of marker genes

For each sample, we calculated the sum of total counts within selected regions using both the full set of reads and downsampled reads. To assess marker gene sensitivity, we considered specific genes known to be expressed in the dorsal anterior (DA) region of the hippocampus in adult mice (*Prdm8*, *Prox1*, and *Slc17a7*), as well as genes known to be expressed in the lens of the eyes (*Vit* and *Crybb3*) and a subset of neural retina cells in the eyes of E12.5 mice (*Aldh1a1*). In each sample, we selected five regions measuring 50μm by 50μm in the eyes and five regions measuring 100μm by 100μm in the hippocampus, where these genes were known to be expressed. We individually summed the total number of UMIs in these selected regions within downsampled count matrices to ensure the number of reads was consistent across platforms.

We then performed pairwise comparisons of UMI counts for detected genes across platforms. For eye samples, genes expressed in any one of the platforms with total counts above the 99th percentile but below the 10th percentile in any other platforms were selected for heatmap plotting using a $\log_1 0$ scale.

To investigate the gene bias observed in Visium, as demonstrated in the aforementioned pairwise comparisons between platforms, we focused on genes meeting a specific criterion: those expressed across all other platforms with total counts exceeding the 90th percentile and 80th percentile, yet exhibiting number of counts below 1 with Visium. Our exploration of these genes encompassed an analysis of their attributes, including GC content percentage and gene length, using ANOVA analysis. Additionally, we examined the biotypes of these biased genes.

To assess the spatial distribution of marker genes known to be expressed in specific regions of the reference tissues, we used *Pmel* for eye data, *Ptgds* for brain data, and *Slc17a7* for OB data. This analysis used count matrices generated from all reads. We selected regions with expression of these genes roughly in the middle of the chosen regions (6 modalities of 300μm by 50μm in the OB, 6 modalities of 500μm by 50μm in the brain, and 3 modalities of 300μm by 50μm in the eyes). We summed the expression of the aforementioned marker genes for every 10μm along 50μm within 300μm in OB, 500μm in the brain, and 300μm in the eyes. The UMI counts of these marker genes were then aligned based on the location of peak expression and averaged. Modalities with insufficient counts for the selected marker genes were filtered out before plotting.

After the computation of averaged summed values across modalities, these values were subsequently normalized for each platform and depicted in a density plot with the area under the curve standardized to 1. Subsequently, the left half-width half maximum (LWHM) of the profile was computed within each modality, using non-normalized expression values across platforms. It is worth noting that we employed LWHM as an evaluation metric, drawing inspiration from the full-width half maximum (FWHM) method [19]. In the chosen region, only the left half-width half maximum was used, as there could be an expression of selected genes, such as *Slc17a7* in OB, on the right side of the section that is biologically expected but not caused by lateral diffusion. Modalities that could not be computed with LWHM were excluded from the plotting process. For diffusion analysis, it is important to note that the DBiT-seq data pertained to E10 embryonic eyes, whereas the other datasets were associated with E12.5 embryonic eyes.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To address the significant diffusion in data generated by Stereo-seq in the mouse brain, we conducted diffusion analysis on its downsampled data, consisting of 14% of all the reads.

4.2.4 Cell type annotation

Low-quality spots with total counts below 30% of the first quantile of total counts are filtered out before normalization, which was carried out using the median number of total counts from each platform as the scaling factor. Subsequently, the top 2,000 highly variable genes were identified using the `FindVariableFeatures` function and used to scale the data through the `ScaleData` function. A total of 20 principal components (PCs) were then calculated using `RunPCA`. To categorize spots in each eye sample, we employed 3 distinct methods, including *Seurat* (v4.3.0), *DR.SC* (v3.3), and *PRECAST* (v1.6.2).

Seurat initially identified neighbors based on 20 PCs, with a k-value of 5 chosen for the k-nearest neighbor algorithm in `FindNeighbors`. `FindClusters` was subsequently applied with various physical resolutions to group known cell-type spots.

DR.SC was applied by setting K (the number of clusters) as 10.

PRECAST was applied with the number of clusters specified as 10 and using the `SelectModel` function to reorganize the fitting results within `PRECASTObj`.

4.2.5 Integration of sST and scRNA-seq data

We followed the instructions in *Seurat* with parameters `reduction ='cca'`, `k.filter = NA`, and `normalization.method = 'SCT'` in `FindTransferAnchors`. Dims were set as 30 with PCA used as `weight.reduction` in `TransferData`.

4.2.6 Evaluation of clustering on downsampled data

The downsampled data were subjected to the same pipeline as described above, leveraging *Seurat* to generate clustering results. These obtained clustering results were subsequently compared to clustering outcomes obtained through the processing of count matrices generated from the entire set of reads. To visualize this comparison, we created a heatmap with a logarithmic scale, illustrating the corresponding number of spots in each of the downsampled clustering and the overall clustering results. The entropy of accuracy and purity were then calculated. ECA and ECP are defined as follows:

$$ECA = -\frac{\sum_{i=1}^{M}\sum_{j=1}^{N_i} p(x_j)\log(p(x_j))}{M} \tag{1}$$

where M denotes the number of clusters generated from a method (the clustering solution to be evaluated), N_i denotes the number of elements in the ith cluster based on the ground truth (here the provided labels) and

$$ECP = -\frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} p(x_j)\log(p(x_j))}{N}. \tag{2}$$

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.2.7 Marker genes detection

`FindMarkers` in *Seurat* was applied to find marker genes between two pairs of cell subsets: 1) lens and melanocytes; 2) pNR4 and pNR1. Genes that exhibited higher expression in the lens and pNR4, with expression levels exceeding 5% of the specific spots, the log-fold-change greater than 0.25, and an adjusted p-value provided by *Seurat* less than 0.01, were considered as marker genes.

4.2.8 Cell communication analysis

Cell communication analysis was then performed on spots with distinct annotated cell types, using methods including *Cellchat* (v1.6.1), *CellPhoneDB* (v4).

Cellchat used the CellChatDB database of the mouse, creating cellchat objects based on annotation information, and employed the default 'Trimean' statistical method.

CellPhoneDB was applied by transforming mouse genes into their human homologs using the `biomaRt` package. Using the CellPhoneDB database for cellphone analysis, we conducted using 1,000 random permutations in the analysis following the tutorial https://github.com/ventolab/CellphoneDB/blob/master/notebooks/T01_Method2.ipynb. The minimum cell percentage threshold required to consider a gene as expressed in the analysis was set to 0.1, and significance was determined with a p-value threshold of less than 0.05.

Data availability

Raw count matrices are available at the National Genome Data Center (https://www.cncb.ac.cn/) under BioProject accession code PRJCA020621. A summary of individual accession numbers is given in Supplementary Table 2. The *c*adasSTre data collections are continually updated on our website genographix.com. The standard sectioning protocol is deposited in protocols.io: dx.doi.org/10.17504/protocols.io.5qpvo379dv4o/v1.

Code availability

Scripts used to process the data are available at https://github.com/YOU-k/cadasSTre.

References

[1] Crosetto, N., Bienko, M., Van Oudenaarden, A.: Spatially resolved transcriptomics and beyond. Nature Reviews Genetics **16**(1), 57–66 (2015)

[2] Asp, M., Bergenstråhle, J., Lundeberg, J.: Spatially resolved transcriptomes—next generation tools for tissue exploration. BioEssays **42**(10), 1900221 (2020)

[3] Marx, V.: Method of the year: spatially resolved transcriptomics. Nature methods **18**(1), 9–14 (2021)

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[4] Tian, L., Chen, F., Macosko, E.Z.: The expanding vistas of spatial transcriptomics. Nature Biotechnology **41**(6), 773–782 (2023)

[5] Long, B., Miller, J., Consortium, T.S.: Spacetx: A roadmap for benchmarking spatial transcriptomics exploration of the brain. arXiv preprint arXiv:2301.08436 (2023)

[6] Soneson, C., Robinson, M.D.: Bias, robustness and scalability in single-cell differential expression analysis. Nature methods **15**(4), 255–261 (2018)

[7] Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., *et al.*: Systematic comparison of single-cell and single-nucleus rna-sequencing methods. Nature biotechnology **38**(6), 737–746 (2020)

[8] Mereu, E., Lafzi, A., Moutinho, C., Ziegenhain, C., McCarthy, D.J., Alvarez-Varela, A., Batlle, E., Sagar, Gruen, D., Lau, J.K., *et al.*: Benchmarking single-cell rna-sequencing protocols for cell atlas projects. Nature biotechnology **38**(6), 747–755 (2020)

[9] Dance, A.: Which single-cell analysis tool is best? scientists offer advice. Nature **612**(7940), 577–579 (2022)

[10] Crowell, H.L., Morillo Leonardo, S.X., Soneson, C., Robinson, M.D.: The shaky foundations of simulating single-cell rna sequencing data. Genome Biology **24**(1), 1–19 (2023)

[11] Lewis, S.M., Asselin-Labat, M.-L., Nguyen, Q., Berthelet, J., Tan, X., Wimmer, V.C., Merino, D., Rogers, K.L., Naik, S.H.: Spatial omics and multiplexed imaging to explore cancer biology. Nature methods **18**(9), 997–1012 (2021)

[12] Tian, L., Su, S., Dong, X., Amann-Zalcenstein, D., Biben, C., Seidi, A., Hilton, D.J., Naik, S.H., Ritchie, M.E.: scpipe: A flexible r/bioconductor preprocessing pipeline for single-cell rna-sequencing data. PLoS computational biology **14**(8), 1006361 (2018)

[13] Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., *et al.*: Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science **353**(6294), 78–82 (2016)

[14] Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J.F., *et al.*: High-definition spatial transcriptomics for in situ tissue profiling. Nature methods **16**(10), 987–990 (2019)

[15] Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J.,

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Arlotta, P., Macosko, E.Z., Chen, F.: Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqv2. Nature biotechnology **39**(3), 313–319 (2021)

[16] Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., *et al.*: Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. Cell **185**(10), 1777–1792 (2022)

[17] Fu, X., Sun, L., Dong, R., Chen, J.Y., Silakit, R., Condon, L.F., Lin, Y., Lin, S., Palmiter, R.D., Gu, L.: Polony gels enable amplifiable dna stamping and spatial transcriptomics of chronic pain. Cell **185**(24), 4621–4633 (2022)

[18] Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., *et al.*: High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell **183**(6), 1665–1681 (2020)

[19] Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., Macosko, E.Z.: Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science **363**(6434), 1463–1467 (2019)

[20] Nakamura, K., Hioki, H., Fujiyama, F., Kaneko, T.: Postnatal changes of vesicular glutamate transporter (vglut) 1 and vglut2 immunoreactivities and their colocalization in the mouse forebrain. Journal of Comparative Neurology **492**(3), 263–288 (2005)

[21] Takamori, S., Rhee, J.S., Rosenmund, C., Jahn, R.: Identification of a vesicular glutamate transporter that defines a glutamatergic phenotype in neurons. Nature **407**(6801), 189–194 (2000)

[22] Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., *et al.*: Genome-wide atlas of gene expression in the adult mouse brain. Nature **445**(7124), 168–176 (2007)

[23] Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., Van Der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., *et al.*: Molecular architecture of the mouse nervous system. Cell **174**(4), 999–1014 (2018)

[24] Watt, B., Niel, G., Raposo, G., Marks, M.S.: Pmel: a pigment cell-specific model for functional amyloid formation. Pigment cell & melanoma research **26**(3), 300–315 (2013)

[25] Larina, I.V., Syed, S.H., Sudheendran, N., Overbeek, P.A., Dickinson, M.E., Larin, K.V.: Optical coherence tomography for live phenotypic analysis of embryonic ocular structures in mouse models. Journal of biomedical optics **17**(8), 081410–081410 (2012)

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[26] Heavner, W., Pevny, L.: Eye development and retinogenesis. Cold Spring Harbor perspectives in biology **4**(12), 008391 (2012)

[27] Zhang, R., Huang, H., Cao, P., Wang, Z., Chen, Y., Pan, Y.: Sma-and mad-related protein 7 (smad7) is required for embryonic eye development in the mouse. Journal of Biological Chemistry **288**(15), 10275–10285 (2013)

[28] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R.: Comprehensive integration of single-cell data. Cell **177**(7), 1888–1902 (2019)

[29] Liu, W., Liao, X., Yang, Y., Lin, H., Yeong, J., Zhou, X., Shi, X., Liu, J.: Joint dimension reduction and clustering analysis of single-cell rna-seq and spatial transcriptomics data. Nucleic acids research **50**(12), 72–72 (2022)

[30] Liu, W., Liao, X., Luo, Z., Yang, Y., Lau, M.C., Jiao, Y., Shi, X., Zhai, W., Ji, H., Yeong, J., *et al.*: Probabilistic embedding, clustering, and alignment for integrating spatial transcriptomics data with precast. Nature communications **14**(1), 296 (2023)

[31] Cheng, A., Hu, G., Li, W.V.: Benchmarking cell-type clustering methods for spatially resolved transcriptomics data. Briefings in Bioinformatics **24**(1), 475 (2023)

[32] Pullin, J.M., McCarthy, D.J.: A comparison of marker gene selection methods for single-cell rna sequencing data. bioRxiv, 2022–05 (2022)

[33] Quinn, J.C., West, J.D., Hill, R.E.: Multiple functions for pax6 in mouse eye and nasal development. Genes & development **10**(4), 435–446 (1996)

[34] Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., Nie, Q.: Inference and analysis of cell-cell communication using cellchat. Nature communications **12**(1), 1088 (2021)

[35] Garcia-Alonso, L., Lorenzi, V., Mazzeo, C.I., Alves-Lopes, J.P., Roberts, K., Sancho-Serra, C., Engelbert, J., Marečková, M., Gruhn, W.H., Botting, R.A., *et al.*: Single-cell roadmap of human gonadal development. Nature **607**(7919), 540–547 (2022)

[36] Tian, L., Dong, X., Freytag, S., Lê Cao, K.-A., Su, S., JalalAbadi, A., Amann-Zalcenstein, D., Weber, T.S., Seidi, A., Jabbari, J.S., *et al.*: Benchmarking single cell rna-sequencing analysis pipelines using mixture control experiments. Nature methods **16**(6), 479–487 (2019)

[37] Chen, W., Zhao, Y., Chen, X., Yang, Z., Xu, X., Bi, Y., Chen, V., Li, J., Choi, H., Ernest, B., *et al.*: A multicenter study benchmarking single-cell rna sequencing technologies using reference samples. Nature Biotechnology **39**(9), 1103–1114

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2021)

[38] Liu, S., Iorgulescu, J.B., Li, S., Borji, M., Barrera-Lopez, I.A., Shanmugam, V., Lyu, H., Morriss, J.W., Garcia, Z.N., Murray, E., *et al.*: Spatial maps of t cell receptors and transcriptomes reveal distinct immune niches and interactions in the adaptive immune response. Immunity **55**(10), 1940–1952 (2022)

[39] Zhao, T., Chiang, Z.D., Morriss, J.W., LaFave, L.M., Murray, E.M., Del Priore, I., Meli, K., Lareau, C.A., Nadaf, N.M., Li, J., *et al.*: Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. Nature **601**(7891), 85–91 (2022)

[40] Liu, Y., DiStasio, M., Su, G., Asashima, H., Enninful, A., Qin, X., Deng, Y., Nam, J., Gao, F., Bordignon, P., et al.: High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial cite-seq. Nature Biotechnology, 1–5 (2023)

[41] Jiang, F., Zhou, X., Qian, Y., Zhu, M., Wang, L., Li, Z., Shen, Q., Wang, M., Qu, F., Cui, G., et al.: Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. Nature Methods, 1–10 (2023)

[42] Alon, S., Goodwin, D.R., Sinha, A., Wassie, A.T., Chen, F., Daugharthy, E.R., Bando, Y., Kajita, A., Xue, A.G., Marrett, K., *et al.*: Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. Science **371**(6528), 2656 (2021)

Acknowledgments

We thank members of the Tian and Liu laboratories for helpful discussions. We thank Professor Yixue Li for his support to the Tian Lab and to this project. We thank Professor Gang Cao for the helpful discussion. This project is supported by Guangzhou National Laboratory (Grant No. YW-YFYJ0301), Westlake Foundation, and the Key research and development project of the Ministry of Science and Technology (2022YFA1105700).

Author contributions

L.T. and X.L. designed and supervised this study. Y.F. curated reference tissues, wrote sectioning protocol, and performed sST experiments with help from S.J. and X.L. Y.Y. conducted data analyses with help from L.L., Z.Z., Y.X, A.S.K and Y.L. S.L. and W.R. performed padlock probe imaging experiments. F.J. and G.P. performed DBiT-seq experiments. Y.Y., Y.F, M.E.R. and L.T. wrote the manuscript with the contribution of all of the authors.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.03.569744; this version posted December 4, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Competing interests

Although not directly related to this paper, X.L. is a co-founder of iCamuno Biotherapeutics. The other authors declare no competing interests.

Appendix A Supplementary information

Supplementary figures and Supplementary tables.