WHITEPAPER

Evaluating the Technical Performance of Single-Cell Spatial Molecular Imaging Technologies

Joseph M Beechem, Patrick Danaher*, Hiromi Sato, Claire Williams, and Michael Rhodes*

NanoString Technologies, Inc., Seattle, WA 98109
MK5986 | September 2023

*Corresponding authors: pdanaher@nanostring.com, jbeechem@nanostring.com

nanoString®

Single-Cell Spatial Molecular Imaging Technologies

Executive Summary

- Maximizing the multiplex capabilities (a.k.a. "plexity") of spatial assays yields the highest sensitivity (number of transcripts/cell) and greatest power in understanding all the spatial biology (number of genes/cell) of a tissue sample.
- Cell segmentation errors are by far the largest source of error in molecular imaging.
- Coupling multi-omic analysis (simultaneous imaging of protein and RNA on a single-slide) with advanced machine-learning software can minimize cell segmentation errors while providing unique insights into spatial biology not provided by single-omic analysis.

Introduction

Named *Nature* Journals Method of the Year in 2020, high-plex spatial transcriptomics continues to advance, currently able to perform at single-cell and sub-cellular resolution (e.g., He et al.[1]). As spatial technologies revolutionize almost every area of life sciences and translational sciences, nearly all medical and basic biology researchers are educating themselves on the critical criteria for selecting the appropriate model of this powerful new class of single-cell spatial molecular imager to purchase for their labs and research centers. As a means of guiding those decisions, this article provides a framework for selecting the best spatial imagers, emphasizing the technical performance specifications that most directly influence the net information content obtained from this new technology.

In order to make appropriate recommendations, it is first necessary to qualify the type of research question being addressed. This review is for any researcher entering the field of spatial biology with a desire to measure as many biological processes in tissues and engineered cells as possible (e.g., signaling pathways, ligand-receptor interactions, biomarkers of response to therapy, developmental organogenesis, understanding heterogeneity in tissue, etc.) on a spatial scale that spans enough tissue to capture the true heterogeneity of biological samples (e.g., over ~100,000 cells at a time per sample). To that end, this review focuses on the more finely detailed technical specifications important for selecting a spatial imager. Also included is a table of the specifications of molecular imagers **(Table A1 in Appendix 1)** which includes parameters requiring less detailed technical analysis that are still necessary for making an instrument selection. This article therefore focuses on the finer technical details critical to the system's overall performance capability.

Key Technical Specification Parameters

The two key parameters for evaluating the performance of a spatial imaging platform are its:

1. **Multiplexing Capability:** This parameter determines the breadth of biological information you can spatially resolve. The higher the plex of a spatial assay, the greater the ability to observe multiple biological processes operating in the tissue under study. Higher plex capabilities also yield a higher number of total measured transcripts per cell (**sensitivity,** defined below) than lower plex assays.

2. **Data Accuracy:** How reliable is the data collected by a spatial imaging platform? To answer this question, it is important to understand the parameters for assessing data accuracy. Although there are multiple contributing sources of error in molecular imaging, it is critical to rank them in terms of their impact on data accuracy. The most critical factor affecting data accuracy is cell segmentation error **(Table 1 and Fig. 3)**, which dominates all other error terms due to misassignment of gene transcript localization. Other dominant sources of error, namely rates of false positive and false negative RNA, have a much smaller (nearly negligible) impact on most spatial analyses, especially when compared to the greater magnitude of segmentation errors. Appropriate assignment of cell segmentation is therefore the key factor for generating accurate data.

Performance Metrics

The performance of a spatial imager can be evaluated on two axes: multiplexing capability and accuracy **(Table 1).** Multiplexing capability results from both the abundance and diversity of data that is measured by evaluating hundreds of genes simultaneously, which provides a breadth of biological insights for exploring a wide range of scientific questions with significant statistical power. Accuracy refers to the reliability of the generated data by minimizing the influence of experimental errors, thus providing faithful biological content. The CosMx Spatial Molecular Imager (SMI) is a high-plex in situ imaging platform that provides spatial multiomics at cellular and subcellular resolution.

Single-cell spatial performance = Multiplexing capability and accuracy

TABLE 1: Key single-cell spatial performance metrics.

Category	Parameter	Metric	Impact
Multiplexing Capability	Sensitivity	Mean number of transcripts detected above background per cell (or unit cellular area)	**Accuracy of biological analyses. The ability to understand the biological state of individual cells (e.g. activation, signaling)**
	Genomic breadth	Number of genes detected per cell (and per sample)	**Broader biological insights, such as biomarker discovery, ligand-receptor interaction, and pathway analysis**
Accuracy	Cell segmentation error	"Gold standard" segmentation error is still quantified using visual examination. Mean counts of misassigned transcripts[1] minus mean counts of negative probes is introduced here as an exploratory automatable surrogate.	**Misleading results due to spatial bias in gene expression (Highest impact on data quality)**
	False positive	Mean counts of negative control probes	**Noise leads to reduced statistical power (low impact[2] compared to segmentation error)**
	False negative	See Note[3]	**Noise leads to reduced statistical power (low impact[2] compared to segmentation error)**

[1] Cell segmentation errors predominately derived from incorrectly segmented neighboring cells.
[2] Spatial single-cell data involve the examination of spatial correlations amongst 100's-of-thousands to millions of single cells. The impact of false positives and false negatives, that have negligible spatial correlation, really eliminates these as a major complicating factor in spatial analysis. There are exceptions to this general rule, but keep in mind the key two questions this review addresses (listed above).
[3] Note that a false negative occurs when transcripts present in the cell fail to be detected. False negatives for RNA are an intrinsic part of all single-cell data sets, as evidenced by the fact that no high-plex technology currently invented detects the number of transcripts commonly believed to exist in a typical cell, an average of 100,000 to 1,000,000 transcripts per cell (Islam S. et al.[2]). Single-cell RNA-seq (scRNA-seq) solves this issue primarily by cluster analysis of related cells while spatial single-cell technology primarily solves it by direct spatial correlation of gene-expression data (see discussion below). There are exceptions to this general rule, but keep in mind the key two questions this review addresses (listed above).

Multiplexing Capability

The most important question about a spatial imager is: How much biology can researchers see with this instrument? Will the spatial data provide only a cell-type map, or will it support more advanced analyses?

Single-cell spatial imagers are designed to visualize and quantify molecular information within individual cells in a tissue. Imagers that can deliver a more comprehensive view of the biological state (i.e. higher plex) provide both broader and deeper spatial biological insights to researchers. The multiplexing capability metric is quantified by two parameters: sensitivity (number of transcripts detected per cell) and genomic breadth (number of genes detected per cell; **Table 1**).

Sensitivity (total transcripts measured per cell): In designing highly multiplexed assays, a fundamental question is: maximum multiplexing of what? The CosMx SMI high-plex assays are designed to measure as many protein-coding mRNAs as possible. While the CosMx SMI has the flexibility to spike-in additional RNA types into its high-plex assays (non-coding RNAs, splice isoforms, circular RNAs, etc.), the fundamental panel is designed to maximize the detection of protein-coding genes. Based on that design objective, optimizing the analytical sensitivity of the assay requires an ability to measure single unique transcripts in as small a sample as can be accurately measured (Saah & Hoover[3]). In the case of protein-coding transcripts, the samples consist of tissues or cells that can be analyzed for quantitation of the number of transcripts per cell, with the smallest sample size, due to biological heterogeneity, being thousands of cells. Hence, a spatial biology imager with maximal sensitivity can detect the highest number of total transcripts measured per cell over a spatial domain of at least 100,000 cells.

Another way to think about sensitivity is with the equation from Yerushalmy et al[4]:

Sensitivity of GeneX = (1 – false negative rate of GeneX)

However, this equation for calculating sensitivity is for a 1-plex assay, whereas the revolution in spatial biology has been driven by increasing the multiplexity of spatial assays. A better estimate of high-plex spatial sensitivity may therefore require changing the frame-of-reference of the sensitivity equation from 1 gene to 1 spatial multiplexed assay, as so:

Sensitivity (Spatial Assay) = 1 – false negative rate (Spatial Assay)

For the CosMx SMI, the Spatial Assay is equivalent to the measurement of as many protein-coding gene transcripts as possible (which currently number 1,000, and will reach 6,000 in Q1 of 2024). Defining sensitivity at the assay level as an overall sum of protein-coding transcripts allows high-plex spatial imagers to follow the exact same sensitivity specifications as scRNA-seq, in which the number of transcripts detected per single cell (sensitivity) and the number of genes per single cell (genomic breadth, discussed below) are maximized. The greater the number of RNA molecules detected per cell, and the larger the diversity of genes detected (per cell and per sample), the deeper the insights into the biology of each cell. Background counts should be estimated with negative cell-type marker genes as described below, as well as with the external RNA control consortium counts (ERCC)[5] when provided in the assay.

RNA assays performed in high-plex spatial imagers (> 1,000-plex) such as the CosMx SMI excel their sensitivity from their ability to detect a high number of transcripts above background per cell in various tissue types. To highlight what is possible with the CosMx SMI, a dataset generated using the CosMx Mouse Neuroscience Panel on FFPE mouse brain tissue is publicly available (**https://nanostring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/**). This dataset showcases a top level of sensitivity, with an average of more than one thousand transcripts per cell detected **(Figure 1)**.

Multiplexing Capability (Genomic Breadth): The multiplexing capability of a spatial imaging platform gives genomic breadth to the data derived from its assays by detecting a larger number of expressed genes than previously possible. Genomic breadth on a multiplex scale reveals the wide range of genes in a cell at one time, thus providing clues to the diversity of the biological phenomena at work in specific cell types and guiding which biological hypotheses are worth characterizing. Greater genomic breadth (measured as the number of genes with total counts study-wide > 2 Standard Deviation (SD) above the mean total counts from negative control probes) therefore means more potentially relevant hypotheses to test. CosMx RNA assays are designed to provide genomic breadth with their ultra-high-plex RNA panels that can detect high numbers of genes per sample. The CosMx mouse neuroscience dataset generated from brain tissue demonstrated the genomic breadth performance of the CosMx SMI, with 854 and 903 genes detected above the negative control probe background in its two samples (transcript-count data shown in **Figure 1**).

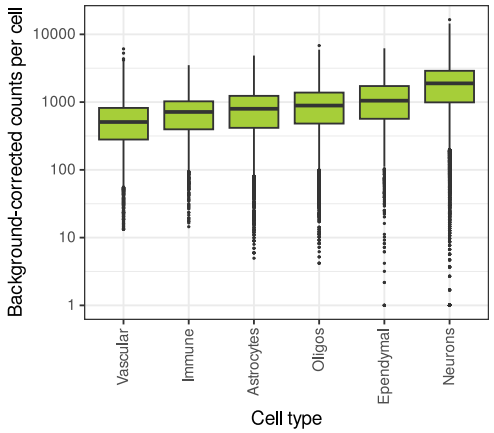


Figure 1. Box plot showing the number of unique transcripts detected per cell, above the negative control probe background level, from FFPE mouse brain sample assayed with the CosMx Mouse Neuroscience Panel (https://nanostring.com/products/cosmx-spatial-molecular-imager/cosmx-panels-assays/cosmx-mouse-neuroscience-panel/).

NOTE: The net counts/cell will vary widely depending on the sample studied and panel used, especially with human FFPE samples. The range of counts/cell for a 1,000-plex panel using a 5 µm thick FFPE sample will range anywhere from 100 transcripts/cell to as many as 1,500 transcripts/cell, depending mainly on pre-analytical variables (especially with human samples) that are challenging to control for (e.g., sample fixation time, excisional biopsy surgical ischemic time, sample storage temperature, sample exposure to oxygen/humidity, etc.). We have found that if low transcripts per cell are measured using a high-plex panel due to pre-analytical variables, that same sample will show significantly lower counts when using a lower plex panel, proportional to the degree of plex in the assay.

Currently, panels for the CosMx SMI all focus on measuring protein-coding genes, and do so at the 1,000-plex level, which translates to a comprehensive genomic breadth of approximately 5%, assuming approximately 20,000 protein-coding genes in humans (1,000/20,000). This will increase to approximately a 30% comprehensive genomic breadth when the 6,000-plex CosMx assay launches in 2024, with further increases expected in the near future through ongoing efforts to develop spatial assays with higher plex levels. Additionally, a complete range of panels is being developed at a variety of levels of protein-coding gene coverage. Fortunately, the efficiency of multiplex spatial imaging technology prevents dramatic increases in assay run-times as one increases plex capacity. For example, the increase in run-time of the CosMx 6,000-plex panel is only about 1.68-fold greater than the run-time of the CosMx 1,000-plex assay even though transcript coverage increases six-fold.

To examine whether increases in genomic breadth impact data analyses proportionally with increased plexity, a complete 950-plex CosMx dataset was compared to a 250-plex subset randomly selected from those same 950 genes **(Figure 2).** The result of this analysis suggests that an assay capable of identifying a maximum of 250 targets yields fewer testable hypotheses than an assay with 950 targets. For example, the experiment shown in **Figure 2** detected

894 genes expressed from the 950 gene panel compared to 239 from the randomly selected 250 gene panel (top row of **Figure 2**). Simple cell-type gene expression identification such as this can be equated to a simple hypothesis phrased as "Gene X expression plays a role in the function of cell Y" for each gene identified. The larger plex panel thus generated nearly four times as many hypotheses as the smaller plex panel. Such hypotheses can then be further refined by deeper analysis of the expressed genes via such relationships as Reactome pathways **(row 2 of Figure 2)**, spatial correlation **(row 3 of Figure 2)**, and ligand-receptor pairs **(row 4 of Figure 2)** to identify biologically relevant hypotheses embedded in the data. Reactome pathway analysis, for example, provides the starkest contrast between the 950-plex panel and the 250-plex panel, yielding 34 identified pathways compared to zero, respectively. Hypotheses generated from the biological information in data uncovered due to the genomic breadth of a 950-plex panel (soon to be higher) will accelerate studies of gene expression changes across space and cell type, driving biomarker discovery, pathway analysis, and ligand-receptor interaction, in addition to the cell-typing analysis.**(Figure 2)**.

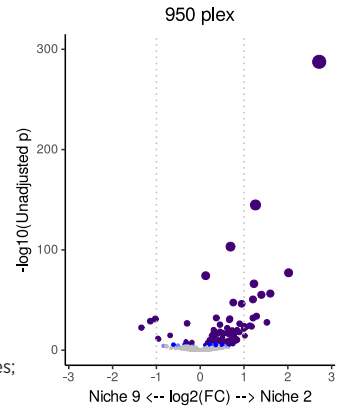
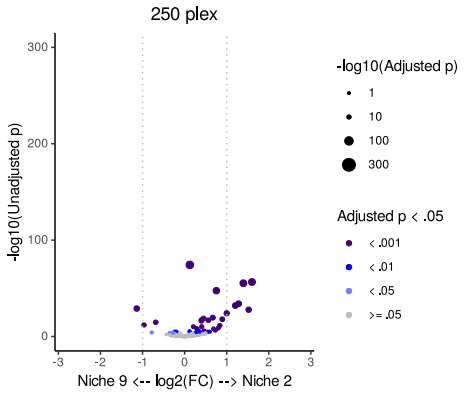
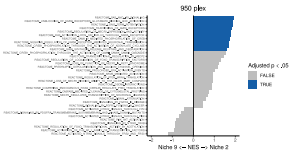
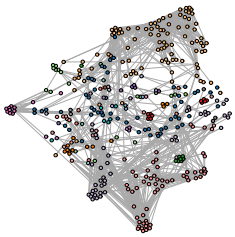
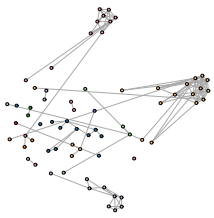
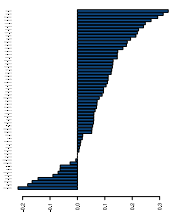
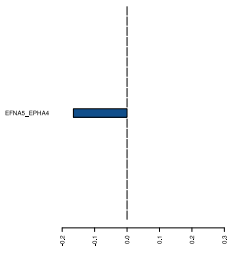
Question	**Examples**	**950-plex results** (894 genes detected)	**250-plex results** (239 genes detected)
How does a cell type change across space? e.g. How do microglia near tau plaques change behavior? How do T-cells near tumors change expression?	Differential, expression, SpatialDE	894 hypotheses; 68 "hits" 	239 hypotheses; 23 "hits" 
What pathways participate in these changes?	GSEA, ssGSEA	34 Reactome pathways with >50% coverage; 10 are significant 	0 Reactome pathways with >50% coverage 
What genes are spatially correlated?	Lee's L, spatialDE	894^2 = 720,801 hypotheses, 1,662 spatially correlated gene pairs 	239^2 = 57,121 hypotheses, 140 spatially correlated gene pairs 
What ligand-receptor pairs are spatially interacting?	COMMOT, Lee's L	58 LR pairs 	1 LR pairs 

Figure 2. Impact of genomic breadth on data analyses as seen with two different plex-level assays. Analyses were performed on a complete 950-plex CosMx dataset and a randomly selected 250-plex subset of the same data. **1st row** Differential expression results comparing gene expression in astrocytes between two spatial niches. **2nd row** Gene Set Enrichment Analysis (GSEA) for Reactome Pathways. Only pathways with ≥50% of genes in the panel and ≥5 genes were considered. **3rd row** Spatial correlation analysis (>0.5) of gene pairs. **4th row** Bars show a spatial correlation between ligand-receptor pairs defined by CellChatDB (see: cellchat.org)

Accuracy

Evaluating the performance accuracy of spatial imagers is most simply assessed by characterizing how well the measured gene expression profiles reflect the true expression profiles of the cells in question. Assessing accuracy of a spatial imager requires quantifying the three major types of errors that reduce accuracy of the spatial imaging data: inaccurate cell segmentation, which leads to the assignment of RNA transcripts to the wrong cells; false positives, the detection of non-specifically bound RNA probes; and false negatives, not detecting the existing RNA molecules. Of these three error modes, cell segmentation error has the greatest impact on spatial analyses (see **Figure 4**). Each of these error modes is discussed in detail below along with proposed metrics for each error type.

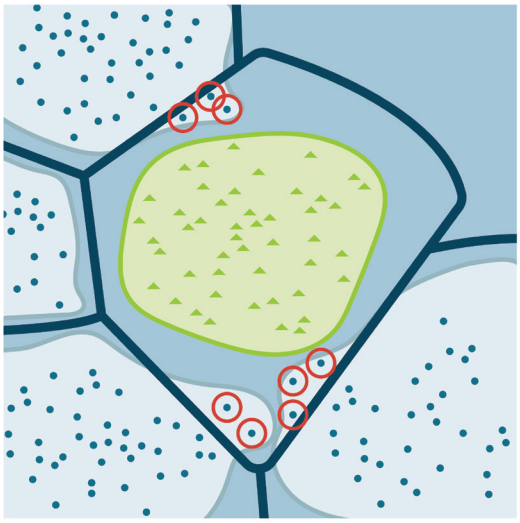
Cell segmentation error: The most fundamental function of a single-cell spatial imager is the ability to accurately assign molecules to individual cells in a sample. Assignment of detected transcripts requires the imager to identify the proper coordinates of the molecule within the cellular boundaries inferred by a segmentation algorithm. Cellular boundary assignment, also known as cell segmentation, needs to be highly accurate for properly assigning detected RNA molecules to the correct cells. Inaccurate cell segmentation, in extreme cases, can create spatial "doublets" where one cell represents a hybrid of adjacent cell types. Assignment of an RNA transcript to the wrong cell type or a hybrid of adjacent cell types creates inaccurate data analysis by grouping cells with their spatial neighbors, leading to incorrect results that appear statistically significant, critically affecting spatial imager performance. Unfortunately, cell segmentation errors are still relatively abundant (best case = 5% to 10% of cells)[6,1], even though segmentation algorithms continue to improve. An example of the impact of segmentation errors on spatial imaging data is illustrated by considering T cells in a cancer study. Poor cell segmentation assignments can cause T cells in close proximity to tumor cells to appear to falsely express tumor genes **(Figure 3A).** Similarly, poorly defined T cells in the stroma will falsely appear to express genes from stromal populations such as fibroblasts. A study comparing T cells in tumor tissue to T cells in the stroma would, in this case, make it hard to discern genuine differences in T cell expression across different spatial contexts since genes from tumor cells and fibroblasts are among the most statistically significant genes. These types of errors cause artifacts in downstream data analysis such as clustering, differential gene expression, and pathway analysis, leading to inappropriate conclusions about results.

To date, there are no simple metrics for measuring cell segmentation errors. However, systematic visual examination is often sufficient to distinguish between appropriate and inappropriate cell segmentation performance, but has yet to be automated. An alternative to estimate segmentation error instead of visual inspection is analysis of negative cell type marker genes. Using existing scRNA-seq datasets to identify genes known to be absent in a given cell type, negative cell type marker analysis **(Appendix 2)** calculates the average negative control probe background-subtracted expression of each negative cell type marker gene in the cells of a tissue sample. This surrogate assay has been used to estimate, in orders of magnitude, levels of segmentation error compared to other error sources for examined samples **(Figure 4).** One caveat: these negative marker genes are generally low expressers across all cell types, but segmentation errors will drive more background in genes with high expression in other cell types. This approach likely underestimates background from segmentation error. However, due to the direction of this bias, this analysis suffices to demonstrate the important contribution of segmentation errors to total background. As can be seen, the negative marker counts are larger than the false positive codes, which are larger than the False Codes. Calculating the ratio of this effect per cell-type shows that segmentation error is greater than false positive codes (negative code in graph) by ~ 3 times. Choosing negative marker genes is a difficult task; our selected negative markers are focused on low expressers and may underestimate the average rate of transcript misassignment. While imperfect, particularly regarding its sensitivity to gene choice and cell typing accuracy, we believe negative cell type marker gene analysis clearly indicates that segmentation error is the key factor in spatial imager data accuracy, deserving a focused effort by the stakeholding research community. Although more research is needed to derive an objective and robust measure of mis-segmentation rates, at this point, the best method for comparing two segmentation algorithms is still visual examination.

NanoString Technologies has a long-standing research initiative in this important area. The CosMx SMI utilizes sophisticated cell segmentation based on an enhanced CellPose algorithm (Stringer et al.[6]) that combines information from cell membrane proteins, nuclei, and RNA. This algorithm has been trained and tested on hundreds of different tissue-type specimens measured on the CosMx SMI, to date representing over 50 million single cells. As shown in **Figures 3B-D**, the CosMx SMI effectively segments complex cells, outperforming other platforms employing nuclear expansion segmentation strategies, especially when cells are dispersed or have complex shapes.

Additional advances in complex-shaped cell segmentation algorithms have recently been made by NanoString Technologies and can be found in the CosMx SMI abstracts, posters, and talks. In brief, high-plex protein and high-plex RNA assays combined with the plasma membrane (and other key cell features) detection on a same slide have enabled the segmentation of complex-shaped cells (e.g. brain cells). These information-rich raw-data are analyzed using a combination of machine-learning/artificial-intelligence software and classic methods to generate a highly detailed cell

A



Tumor mRNA

T-cell mRNA

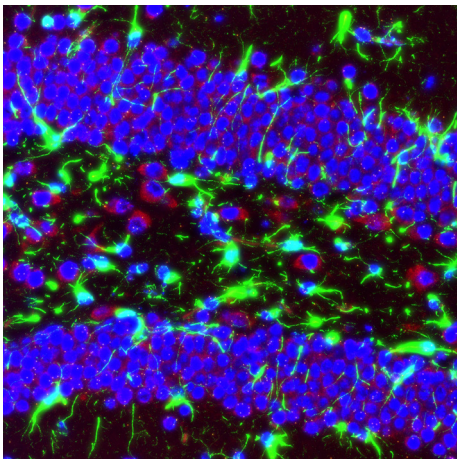
Cell membrane

Segmentation boundaries

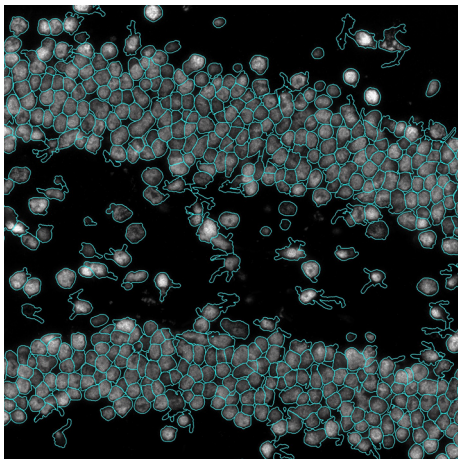
Tumor mRNA falsely attributed to T-cell after mild segmentation errors.

Figure 3. Transcript assignment quality. A) The cartoon depicts the impact of minor segmentation errors, leading to the assignment of incorrect genes to a cell, subsequently giving rise to artifacts in downstream data analysis such as clustering, differential gene expression, and pathway analysis. (B and C) The CosMx assay uses antibody-based protein detection morphology markers (cell membrane) and nuclear staining for cell segmentation, whereas D) other spatial imagers often utilize simple nuclear expansion (~15 µm) for cell segmentation as shown here with the same image using simplified polygons for ease of visualization.

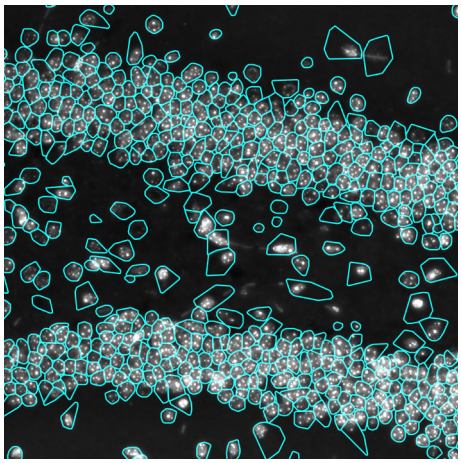
B



C



D



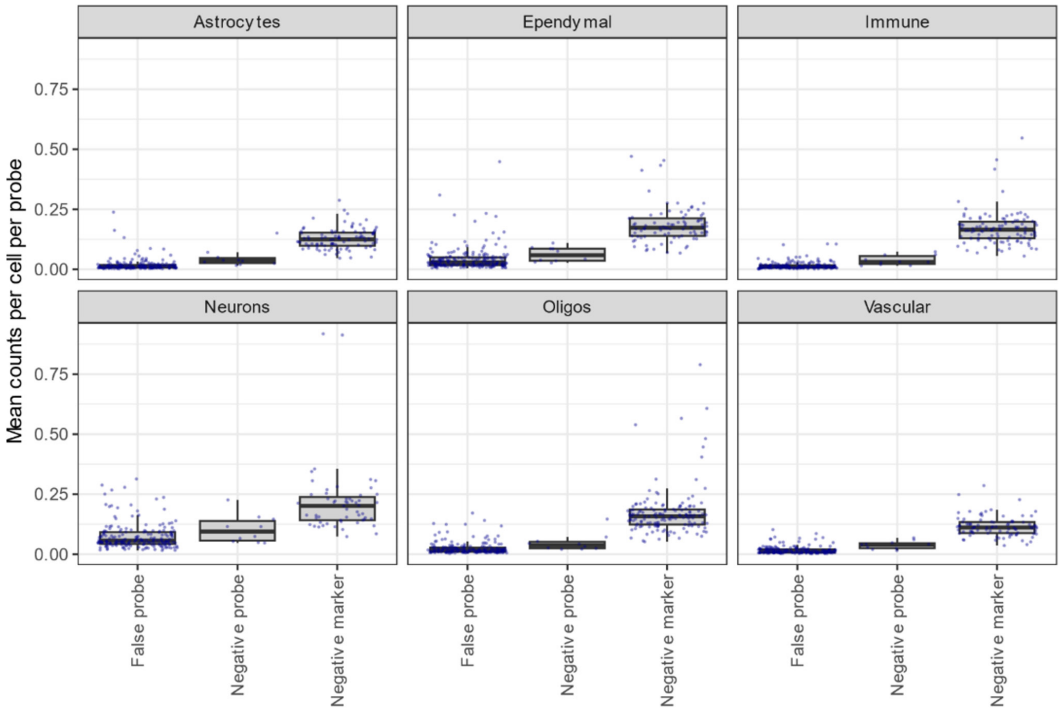


Figure 4. Background noise from different false positive sources. Mean counts of FalseCodes, negative control probes (ERCCs), and negative cell-type-marker genes are shown for distinct cell types in the CosMx mouse brain dataset. The cell segmentation error (counts of negative cell-type-marker genes) is the major source of false counts. Despite our approach's tendency to underestimate segmentation error, the ratio of segmentation error over negative probe error ranges from 2.1 to 4.6, with a mean of 3.2 times larger segmentation error than negative probe error (false positives). CosMx SMI's segmentation capability with lower error rates is currently best-in-class. For very simplistic high-plex segmentation methods (e.g., 15 µm expansion beyond a DAPI nuclear signal to estimate cell boundaries) used in other platforms, a larger effect on segmentation error is expected. Note: Choosing negative marker genes is a difficult task; the above selected negative markers are biased towards low expressers; thus, they underestimate the average rate of transcript misassignment.

TABLE 2: Parameters impacting the background of spatial imagers

Source of background	How to measure	Abundance	Impact
False positive	Negative control probes	**Low-moderate**	**LOW** Noise, loss of statistical power when analyzing few cells
Cell segmentation error	Negative cell-type-marker genes	**Moderate-high** Depending on segmentation accuracy	**HIGH** Bias, spurious statistical significance

segmentation pattern. Where most imagers typically discard half of total measured transcripts in brain samples, due to an inability to map the complex-shaped cells back to any particular single cell, NanoString now has the ability to accurately segment human brain cells up to 1 millimeter in size. These advances will greatly reduce the large fraction of discarded "orphan" transcripts in brain images, and simultaneously increase our understanding of the spatially resolved intracellular regulation of gene expression and biological reactions. This advanced segmentation approach is also being applied to non-brain tissue, with the expectation of minimizing nearly all imaging-based errors associated with the cell segmentation process. However, there may be an insurmountable lower limit of segmentation error due to the fraction of cells in tissues spatially intertwined and physically overlapping which may need to be deconvolved mathematically from each other.

This discussion of cell segmentation error is intended to highlight a dominant error term in selecting a spatial imager. More than just a software problem, good segmentation requires good input data that maximizes plasma-membrane signals from as many targets as possible. While some spatial molecular imaging systems leave all segmentation problems for open-source software to solve, CosMx SMI technology has focused on the critical issue of segmentation from the very beginning of instrument and chemistry design, developing instruments that use machine learning-guided segmentation based on imaging analysis of labeled protein morphology markers on the same slide as high-plex RNA assays. Future CosMx SMI cell segmentation data will be further enhanced by enabling ultra-high-plex protein (~ 100-plex) and ultra-high-plex RNA (~ 6,000-plex) analysis on the same slide. This high-plex multi-omic data will enable the highest quality automated cell segmentation to date.

False positive errors: False positives occur when a spatial imager detects a signal for non-specific probe binding. False positives are typically caused by two factors: either optical crowding, causing a misreading of genes in the absence of any probe binding, resulting in a barcode reading error; or hybridization probes non-specifically bind within the tissue. To overcome optical crowding, the CosMx SMI uses System Controls which represent target barcodes not present in the assay ("FalseCodes") to quantify this phenomenon, so that even at high-density transcripts per cell (over 1,000/cell), optical crowding is not a common source of errors. Non-specific binding can be controlled by using multiple negative control probes, typically artificial mRNA sequences not expressed by any known organisms, selected from the ERCC[5]. We recommend estimating false positive counts by calculating the mean counts per negative ERCC control probe per cell to define the combined impact of barcode reading error and non-specific binding.

While it is important to control for false positives during any imaging assay, it is not, as stated above, the main source of error affecting data accuracy for spatial molecular imagers **(Table 2).** As seen in **Figure 4,** negative control counts for false positives account for significantly less data errors than do those for negative cell-type-marker genes, the surrogate measure for cell segmentation errors. Therefore, controls for cell segmentation errors are the most critical for accurate data analysis. A true understanding of the dataset accuracy requires estimating background signals using biological controls in combination with negative cell-type-marker genes in addition to relying on false positive counts of ERCC controls. This approach is one way to estimate this error term, but more work needs to be done in this area to establish objective, automatable estimators of cell segmentation errors.

False negative errors: The failure of a technology to detect a present molecule is a false negative. Such events are common in single-cell profiling platforms which tend to produce sparse datasets where genes are frequently measured with 0 counts in most cells. As such false negatives can be assessed both at a single-analyte level and across the platform as a whole. When validating new technologies such as spatial molecular imagers, it is necessary to compare its sensitivity against a "Gold

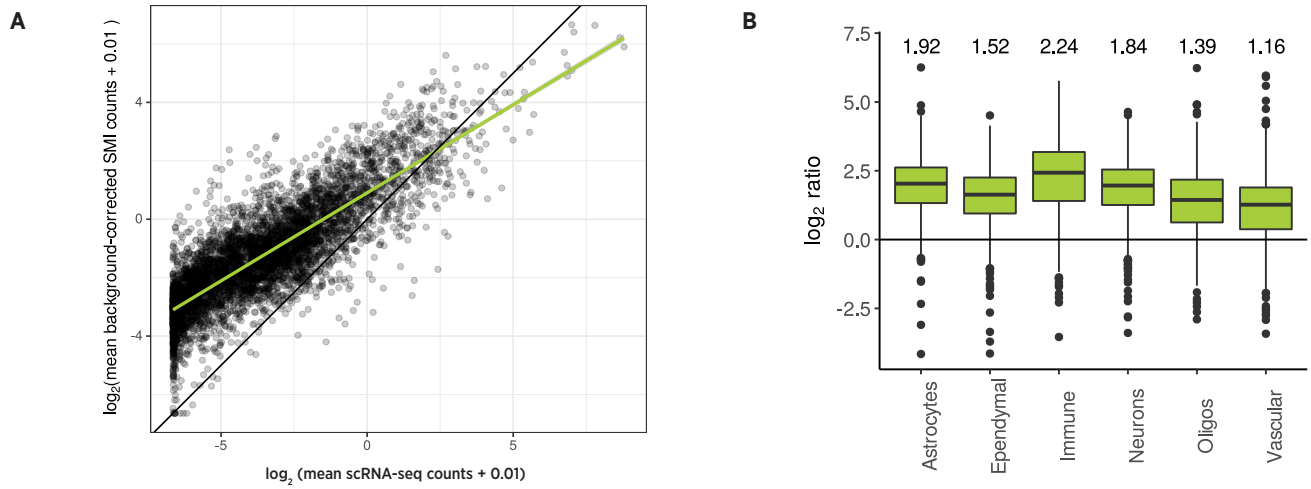


Figure 5 A) The single-cell RNA-seq versus CosMx SMI gene-by-gene sensitivity (and correlation) is measured by comparing mean raw transcripts counts per major cell type per gene (negative probe background subtracted). Here, the CosMx SMI mouse brain dataset is compared with scRNA-seq data (Zeisel et al.[7]). Each data point is a specific gene in a specific cell type. B) Box plot showing genes' mean sensitivity ratio (in log_2) for the CosMx dataset over scRNA-seq data for each cell type for targets expressed in the top 60% of single-cell sequencing data. Log_2 mean ratios shown above each cell type. On average, the CosMx SMI has ~4.6X greater sensitivity than scRNA-seq in this comparison, though this may be influenced by sample quality and sequencing depth of scRNA-seq.

Standard" technique. While there are various "Gold Standard" methods available for comparison, currently only scRNA-seq has the multiplex capacity for state-of-the-art spatial molecular imager comparisons. The preferred method using scRNA-seq is to compute the ratio of the single-cell sensitivities of the two platforms and compare their signal correlations using the ratio of mean background-subtracted transcripts per cell (**Figure 5**; formulas available in Appendix A2). Background can be estimated with either negative control probes or negative markers genes, with awareness of the limitations of these respective approaches. Negative control probes do not measure segmentation error, as such, a small number of genes in the analysis will be expected to have high outlier sensitivity values arising from segmentation errors. A complete sensitivity comparison will summarize these ratios across multiple genes and multiple cell types, giving a broad picture of the sensitivity and false negative error rate for spatial molecular imaging.

In many cases, the fact that spatial imaging platforms sample cells at a much higher rate than single-cell (average study size within the HCA[87] is 124.8k cells from 19.4 donors, which represents a very small 26.9-87.5 mm^2 area collection with the CosMx SMI[9,8]) will help overcome challenges with data sparsity and false negatives. However, certain biological problems require an even higher level of sensitivity, and in such cases false negatives can become a dominant error term. When a critical cell type is rare or poorly sampled, or a transcript is low-abundance and poorly correlated with other targets on the panel, it becomes harder to use shared information content between targets to infer expression of these rare targets. These cases are beyond the scope of this white paper, but it is recommended that for biological problems for which it is essential to measure a single gene (or a small number of genes), lower-plex technologies be used since their false negative rates can be closer to zero. A good example of this type of technology is single-molecule fluorescent in situ hybridization (smFISH), which can detect 1 to 20 different transcripts at a time in a single cell, in effect trading multiplex capacity for near-zero false negative rates and extremely limited dynamic range. CosMx SMI imaging chemistry has been designed to "parallel" the strengths of scRNA-seq technology that maximize the total transcripts per cell and genes measured per cell while simultaneously maintaining the x, y, and z coordinates of each RNA that is measured.

Conclusions

The evaluation of single cell spatial imager performance requires careful consideration of the dataset accuracy, sensitivity, and multiplexing capability. The CosMx SMI platform and assay development process was designed with these dominant technical terms in mind (He et al.[1]). The best-in-class plexity of the CosMx SMI allows the highest sensitivity assay (as defined by total transcripts per cell) and broadest biological functional range (as defined by total number of different genes measured per cell), parameters that made scRNA-seq a successful technology innovation, and should therefore be used to evaluate spatial molecular imaging technologies.

Additionally, the best-in-class segmentation method utilized by the CosMx SMI is based on same-slide high-plex RNA and protein imaging coupled with advanced machine-learning-enhanced segmentation to reliably define cellular and nuclear boundaries. The CosMx SMI surpasses other molecular imaging platforms by generating the best possible raw multi-omic imaging data for use with advanced cell-segmentation software. Therefore, when evaluating high-plex spatial molecular imagers, it is best to focus on the multiplexing capability, sensitivity, and segmentation accuracy during the selection process.

Appendix 1:

Table A1: "Easy-to-understand" Operational Features of Molecular Imagers. Each feature should be considered when selecting a molecular imager instrument.

Feature	Specification	Comment
Sample-Type(s)	Formalin-Fixed Paraffin-Embedded (FFPE), Fresh-Frozen, Fixed-Frozen, Organoids, Fixed-Cells, etc.	FFPE tissue is the most challenging sample to examine compared to other tissue and cell types. The key instrument specification to look for is the minimal value of DV200 (preferably) or RIN score where samples can be measured. The lower the input requirements for FFPE the better. Real-world biobanked tissue samples often have a DV200 index below 50%. The CosMx SMI has been shown to generate data from FFPE tissue with the DV200 indexes down to ~20% (RIN score too low to measure, He et al.[1]).
Multi-Omic Capability	Molecular Imagers should have the capability to detect both RNA and Protein. At single-cell resolution, both analyte types must be measured on the same slide and tissue section.	In general, all studies benefit from multi-omic interrogation. Higher plex protein and RNA examined simultaneously will generate more comprehensive data. The CosMx SMI uses 4-plex protein plus highest plexity RNA on the same slide in 2023, extending to full high-plex protein (~100-plex protein) and 6,000-plex RNA on the same slide in 2024.
Workflow & Automation	Normal Pathology IHC workflow or typically "other"	Normal pathology is preferred for scaling up projects or using automated tissue processing instrumentation (e.g., Leica Bond RX) for sample prep. Workflows differing from a normal pathology IHC workflow should have the proposed process examined in detail. The CosMx SMI uses a standard IHC workflow that can be semi-automated on standard pathology-grade tissue processors.
Imaging Area per slide or per run (for multi-slide systems)	Larger imaging areas per instrument run yield more data. A minimum imageable area per slide should be at least 1 cm^2.	Tissues are heterogeneous, making it critical to image large areas of tissue for a fuller biological picture. The ability to examine multiple slides per run further enhances the robustness of data from a study. The CosMx SMI uses up to 4 slides per run, with 3.00 cm^2 imaging area per slide for a total imaging area of 12 cm^2 per run.
Localization Error (spatial resolution)	The localization of each RNA molecule needs to be at the single-cell level (<1 µm). Ideally, subcellular localization of RNAs should be possible (~0.1 µm).	The spatial resolution specification of an imager refers to the pure optical properties of resolving two points of light in close proximity. Ideally, the spatial imaging platform utilizes a super-resolution type of detection (implemented in cyclic imaging systems) that localizes RNAs well below the optical (diffraction-limited) spatial resolution. The CosMx SMI utilizes super-resolution imaging software to achieve <50 nm localization capability in the X and Y coordinates (He et al.[1], see Supplemental Figure S2).
Does the instrument destroy the tissue under investigation to make a measurement?	The ability to measure gene signals without destroying the tissue sample feature is highly desirable. Ideally, samples are NOT destroyed during measurement. Based on the sample type or sample quality, not every sample can be re-analyzed after a run, but many quality samples can be run multiple times (especially for protein-based work).	Methods that digest a sample onto a capture array or clear a sample before use destroy the sample under investigation and cannot be re-examined. The CosMx SMI avoids sample destruction during measurement, and often the tissue can be re-examined for additional RNA or protein content by the CosMx SMI or another technology (e.g., NGS for bulk genomic sequencing, subsequent H&E staining, etc.).
Instrument turn-around-time (TAT)	Due to their operational principles (cyclic imaging), Molecular Imagers have rather long per-slide and per-run TATs. When comparing specifications, it is critical to examine the plexity level AND imaging area of the TAT spec. Lower-plex and smaller imaging areas certainly lower TAT, but are less than ideal for resolving the maximal amount of information per sample.	TAT per slide often does not equate to TAT per sample, as systems with larger imaging areas per slide often accommodate multiple samples per slide, increasing the TAT per slide while lowering the TAT per sample. The CosMx SMI can examine from 2 to 20 slides per week depending on the plex and sample size. For CosMx primary-path samples (1 cm^2 per slide), expect approximately 2-4 slides per week.

Appendix 2: Procedures for calculating performance metrics

Identifying negative marker genes

Using a reference scRNA-seq dataset, identify genes that have zero or near-zero expression in a given cell type. Alternatively, use prior biological knowledge about highly specific genes. (The former approach is preferred, as it identifies more negative marker genes.) Avoid excessively fine cell typing.

Sensitivity: number of transcripts detected per cell above background

Report the mean cell's total counts minus the estimated total background counts per cell:

$$\textit{(mean total gene counts per cell)} - \textit{(mean total negprobe counts per cell)} \frac{\textit{(n genes in panel)}}{\textit{(n negprobes in panel)}}$$

Genomic breadth: number of genes above background

We define analyzable genes as those whose total signal is 2 standard deviations (SDs) above the total signal from negative control probes:

(total counts of gene) > mean(total counts per negative control) + 2 SD(total counts per negative control)

A more nuanced analysis would be to perform this calculation separately for each cell type in a study. This analysis can also be performed using negative marker genes instead of negative probes.

False positives: mean counts of negative control probes

Report the mean false positive counts per cell:

μ_f = (mean total negative control probe counts per cell) / (n negative control probes)

Mean total background

1. For each cell type k:

   a. Identify negative marker genes *(see above)*

   b. Calculate mean counts per gene per cell of the negative marker genes. Call this b_k.

2. Report the weighted average of the cell types' b_k values, with weight determined by each cell type's abundance n_k:

$$\mu_b = \frac{1}{\sum_k n_k} \sum_k n_k b_k$$

Cell segmentation error: Mean counts of negative marker genes minus negative control probes

1. Calculate the mean total background (μ_b above).

2. Calculate the mean false positive rate per cell (μ_f above).

3. Report total background minus background from false positives: $\mu_b - \mu_f$.

It is also worth examining each negative marker gene in each cell type, as in **Figure 4.** This will produce a wide range of values, giving a fuller accounting of the varying rates of contaminating counts from segmentation errors that you can expect to see.

Proportion of false counts per cell

A common metric, though one with unclear implications for data analysis, is the share of false transcripts across an entire dataset. A more informative metric looks at the share of false counts in the gene expression profiles of the cells, i.e., including cell segmentation errors. This "false count rate" can be calculated as:

(Mean total background μ_b) · (number of genes) · (number of cells) / (total gene counts in cells).

Relative sensitivity between two platforms

This calculation assumes that comparable runs are available from each platform: the same kind of tissue, in the same regions, with similar cell type composition, and similar sample quality. Using serial sections from the same tissue is ideal. It also assumes that the cell type composition of both samples is equivalent, or is a subset of a given cell type:

1. For platform i:

   a. Calculate the mean counts per negative control probe per cell. Call this $\mu_f^{(i)}$.

   b. Calculate the mean counts per cell of each gene. For gene g, call this $\mu_g^{(i)}$.

2. For all genes g shared by both platforms, calculate $(\mu_g^{(1)} - \mu_f^{(1)}) / (\mu_g^{(2)} - \mu_f^{(2)})$.

3. Report the median of $(\mu_g^{(1)} - \mu_f^{(1)}) / (\mu_g^{(2)} - \mu_f^{(2)})$ over all shared genes g.

When the cell type composition cannot be assumed to be equivalent, we recommend performing the above calculations separately for each cell type. This will produce a matrix of relative sensitivity estimates over genes x cell types. Remove low-abundance cell types (e.g., < 1,000 cells), then report the median value of the matrix. To avoid targets with virtually no observations in sparse datasets, we tested the top 60% of data in the reference dataset.

References

1. He, S. et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. Nat Biotechnol 40, 1794-1806 (2022).
2. Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 11, 163-166 (2014).
3. Saah, A.J. & Hoover, D.R. "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. Ann Intern Med 126, 91-94 (1997).
4. Yerushalmy, J. Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. Public Health Rep (1896) 62, 1432-1449 (1947).
5. Baker, S.C. et al. The External RNA Controls Consortium: a progress report. Nat Methods 2, 731-734 (2005).
6. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. Nature Methods 18, 100-106 (2021).
7. Zeisel, A. et al. Molecular Architecture of the Mouse Nervous System. Cell 174, 999-1014.e1022 (2018).
8. https://data.humancellatlas.org/explore/
9. Human Liver Normal and Cancer (https://nanostring.com/products/cosmx-spatial-molecular-imager/human-liver-rna-ffpe-dataset/) and Mouse Brain (https://nanostring.com/products/cosmx-spatial-molecular-imager/cosmx-smi-mouse-brain-ffpe-dataset/)

For more information, please visit nanostring.com

NanoString Technologies, Inc.
530 Fairview Avenue North
Seattle, Washington 98109
T (888) 358-6266
F (206) 378-6288
nanostring.com
info@nanostring.com

Sales Contacts
United States **us.sales@nanostring.com**
EMEA: **europe.sales@nanostring.com**
Asia Pacific & Japan **apac.sales@nanostring.com**
Other Regions **info@nanostring.com**

Sept 2023 MK5986