

What Does High-Quality Spatial Genomics Data Look Like?

Like detective work, detailed understanding of the complexities of biological systems requires scientists to answer questions about details at every level: what, when, why, how...and where. Spatial genomics, [Nature's 2020 Method of the Year¹](#), can provide valuable information about where transcripts from hundreds of different genes are expressed across a biological tissue with single-cell resolution.

While early spatial biology methods were unable to characterize individual cells while measuring sufficient biomolecules to identify the variety of cell types present in a biological tissue, recent advances have achieved single-cell resolution with significantly increased multiplexing capacity, discovering previously hidden biology. Powerful examples of this include an atlas of the mouse primary motor cortex and maps of tumor microenvironments.

The goal of spatial genomics is to precisely map the subtle variations in cellular phenotyping *in situ* across an intact biological tissue, but not all spatial genomics platforms are created equal. Highly accurate and sensitive spatial biology measurements yield more biological insight.

Read on to explore key considerations that can help you optimize data quality for your spatial genomics research.

Establishing Data Quality Standards

In a still-emerging field like spatial genomics, it is not always obvious how scientists should distinguish high-quality data from the rest. Researchers are still working to standardize measurements and performance benchmarks across technologies. But spatial genomics research continues to accelerate, meaning that scientists have an immediate need for ways to help ensure that their tools and techniques are producing maximally reliable, usable data.

Early spatial genomics technologies presented limitations in the scale, resolution, sensitivity, and multiplexing capabilities of experiments, as well as the range of sample types scientists could process. In recent years, advanced next generation spatial technologies, such as MERFISH, have overcome many of these barriers, allowing researchers to hold an accurate magnifying glass to directly observe the molecular and spatial architecture of cells and tissues.

However, to take full advantage of expanded possibilities for spatial genomics experiments based on single molecule detection, scientists need to know which metrics to pay attention to—and what they can expect high-quality data to look like. Here we describe four key questions that you can use to assess quality when comparing spatial genomics technologies.

Key Quality Metric	Sensitivity	Specificity	Information Density	Effective Multiplexing Capacity
Central Question	How many transcripts of each targeted RNA species are detected?	How many transcripts are identified correctly?	How many transcripts can the technology identify in a cellular volume?	How many different genes can the technology accurately profile at once?
Biological Significance	Many biologically relevant transcripts are expressed at low copy numbers.	Lower noise allows more transcriptomic variations to be discovered and stand out above the noise.	Characterizing subtle variations between individual cells requires sufficient information about each cell.	Measuring more biomarkers better characterizes the complexity inherent to a biological system.

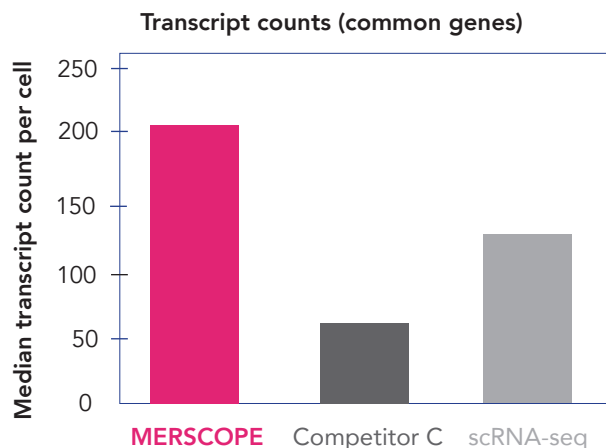
MERFISH, which stands for multiplexed error robust fluorescence *in situ* hybridization (FISH), is a technology invented by Xiaowei's Zhuang's research group at Harvard, originally published in 2015. MERFISH greatly expands the multiplexing capacity of single molecule FISH. By introducing an error robust encoding scheme, MERFISH has enabled researchers to spatially map more than 10,000 genes with single molecule resolution.

Developing individual systems for MERFISH experiments is a cumbersome and lengthy process, so Vizgen spun out of Harvard in 2019 to build the MERSCOPE™ Platform. This solution was designed to streamline researchers' ability to apply MERFISH technology to advance understanding of complex biological systems. With years of experience in highly multiplexed spatial transcriptomics, the Vizgen team focused on ensuring MERSCOPE provides the highest quality data, by integrating cutting-edge optical technology, fluidics, automation, and data processing. MERSCOPE produces spatial genomics data at the unparalleled level of quality scientists have come to expect from MERFISH. As we explore each key quality metric in greater detail, you'll see details of MERSCOPE performance in that area.

Sensitivity

Definition: The relative number of expressed transcripts a technology is capable of detecting compared to another technology; also referred to as detection efficiency when compared to the total number of expressed transcripts.

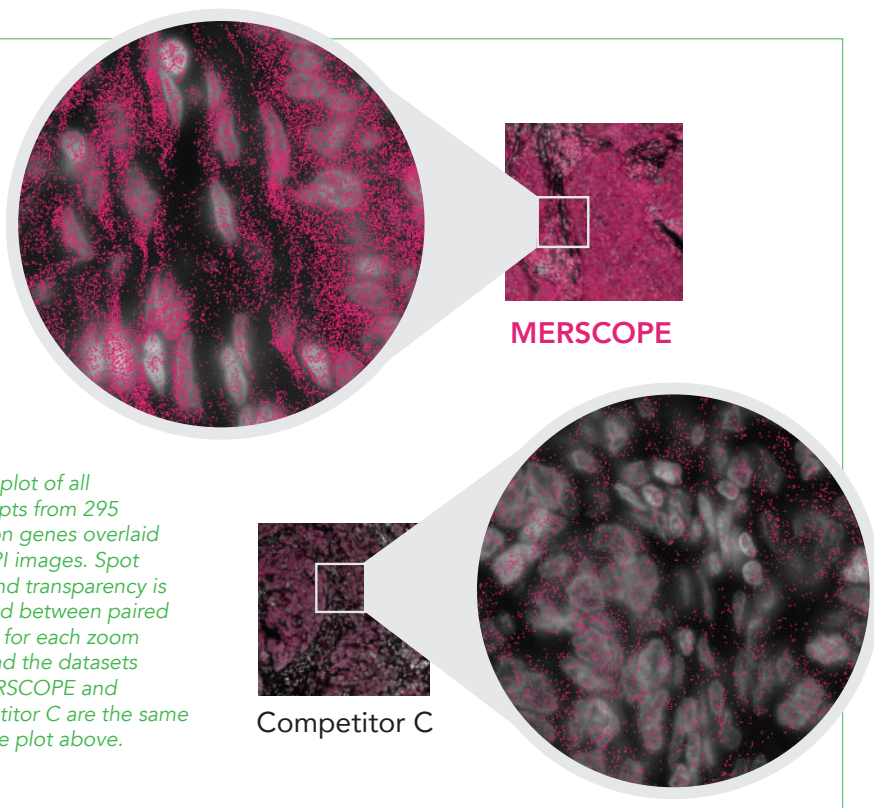
Each cell has a finite number of copies of each expressed RNA, and, often, even functionally relevant genes can be expressed at relatively low transcript copy number. Over the past decade, scientists have been measuring transcriptomics at a single-cell level without spatial context to understand cell types and states using single cell sequencing technologies. From these extensive scientific endeavors, we know sensitivity comparable to or better than single-cell RNA sequencing should provide detailed insight into the cellular composition of a sample. Sensitivity can be characterized by comparing the number of transcripts detected per cell for common genes measured between technologies. MERSCOPE has a higher sensitivity compared to both Competitor C and single-cell sequencing.



Median number of transcripts detected per cell in lung cancer samples between MERSCOPE, Competitor C, and scRNA-seq for 295 genes common between all three technologies. For MERSCOPE and Competitor C, the sample was a 5um thick FFPE sample and the scRNA-seq data is from a fresh frozen sample. The MERSCOPE data is from Vizgen's data release HumanLungCancerPatient2, the Competitor C data is from Nanostring's data release Lung9_2, and the scRNA-seq data is from GSE131907.

Why Sensitivity Matters

The promise of spatial biology to be able to map cellular composition across tissues relies on the ability to detect a sufficient number of transcripts to finely characterize variations in cell type and state; with limited sensitivity, this would not be possible. Many functionally relevant genes are expressed at less than 10 transcript copies per cell, so spatial tools with low sensitivity might fail to measure these and therefore leave important biology hidden. Better transcript detection improves biological relevance across application areas and the ability to detect more subtle variations between cells.

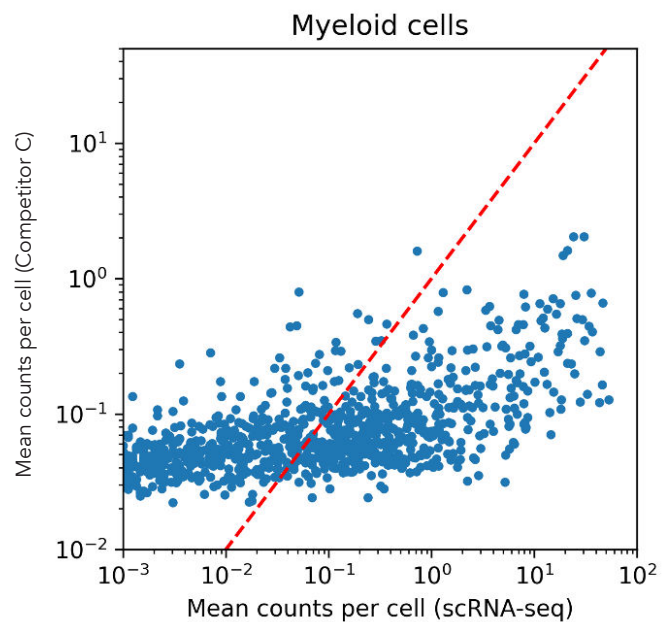
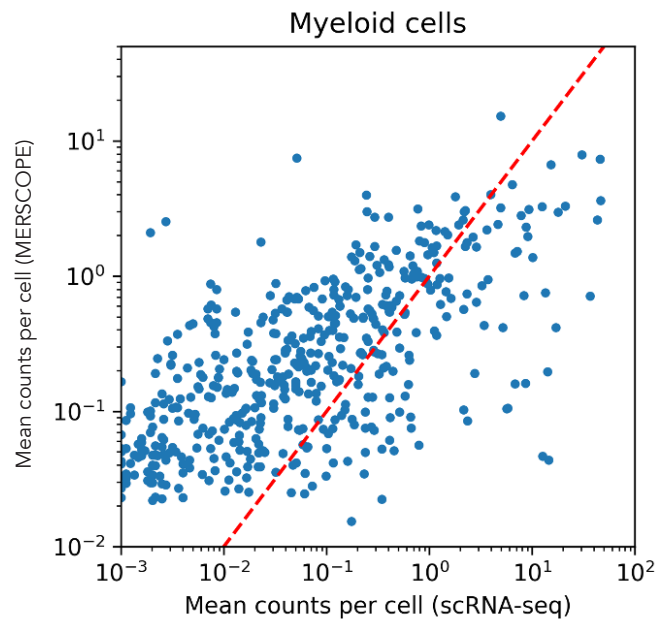


Specificity

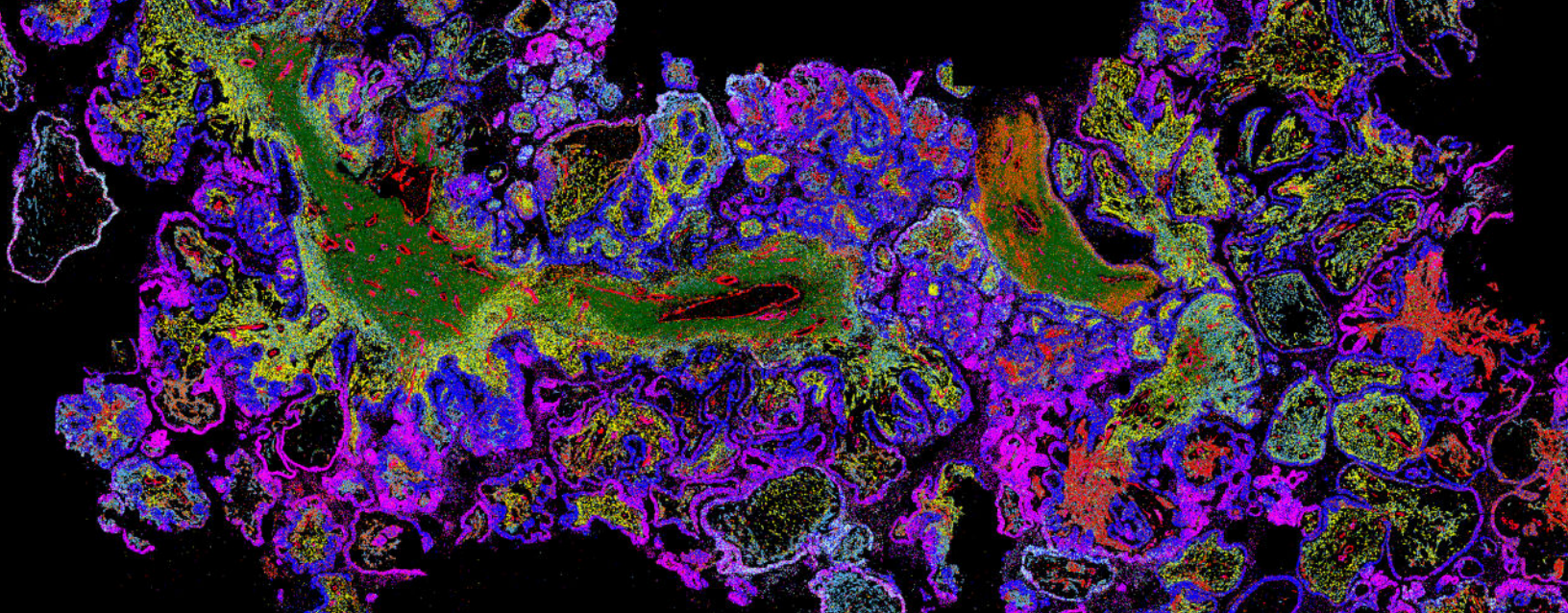
Definition: The fraction of reported transcripts that correspond to a true transcript within the biological sample.

For spatial genomics technologies with single molecule resolution, misidentified transcripts contribute noise that can obscure true biology. A variety of factors can affect the accuracy of spatial genomics, including errors from stray autofluorescence in the tissue, incomplete probe binding, or molecular crowding. MERFISH uses an error-robust [barcoding system](#)³ to reduce transcript detection error, allowing for computational error correction when noise arises that is inherent to fluorescence imaging of biological tissues. Some spatial genomics technologies are at risk of non-specific RNA probe binding, but this too can be overcome. MERFISH technology uses [matrix embedding and clearing](#)⁴ to reduce aberrant RNA signal.

Specificity can be characterized by comparing the output from a spatial genomics technology to a well-established predecessor technology, such as bulk RNA sequencing or scRNA-seq. Comparing to scRNA-seq is helpful to deconvolve variations in the number of cells of each cell type in the sample with the measured gene expression profiles. Comparing the gene expression per cell between MERSCOPE and other single-cell spatial technologies, there are distinct differences in the plots. MERSCOPE is able to maintain strong correlation with single-cell RNA sequencing, even for the most lowly expressed transcripts, while Competitor C exhibits a strong noise floor—for any transcript expressed at fewer than ~3 copies per cell, the signal is overwhelmed by noise. A change in gene expression of nearly 4 orders of magnitude measured by single-cell RNA sequencing corresponds with only a ~2-fold change as measured by Competitor C within myeloid cells.



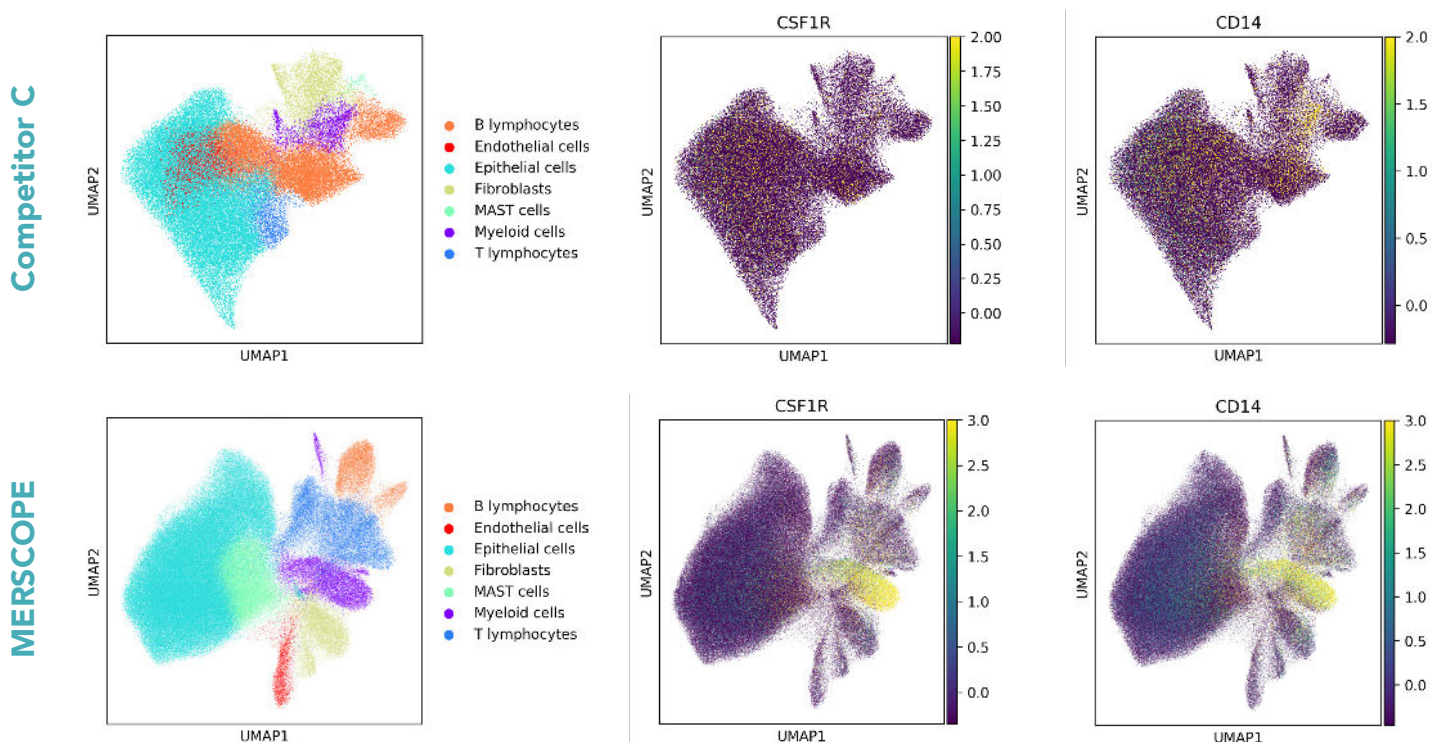
Correlation between the mean gene expression measured by Competitor C or MERSCOPE compared to scRNA-seq within myeloid cells for each gene. The dashed red line indicates $y=x$, or equal detection efficiency between the spatial technology and scRNA-seq. Data comparison between Vizgen public dataset (vizgen.com) sample HuLuCaPt2 and NanoString public dataset (nanosttring.com) sample Lung9_2.



Why Specificity Matters

Spatial genomics data must accurately recapitulate well-established biology to provide a strong foundation for discovering new biology. Inaccurate data might lead to inaccurate results and conclusions since transcripts may be detected in cells or parts of the tissue where the gene isn't truly expressed.

The consequences of the accuracy differences between MERSCOPE and Competitor C become apparent when characterizing gene expression at the single-cell level. With MERSCOPE data, the expression of key myeloid cell type markers CSF1R and CD14 is highly enriched in the myeloid cell population as depicted in the UMAP representation below. However, with Competitor C, the enrichment of these key myeloid cell markers is faint or non-existent. With high accuracy, MERSCOPE recapitulates known biology with high fidelity, providing a foundation for expanding our biological understanding.



UMAP plots depicting cells in dimensionality reduced gene expression space colored by identified cell type and colored by z-scored expression of CSF1R and CD14. CSF1R and CD14 are well characterized cell type markers for myeloid cells and expected to be enriched in the population of cells identified as myeloid cells. Data comparison between Vizgen public dataset (vizgen.com) sample HuLuCaPt2 and NanoString public dataset (nanosttring.com) sample Lung9_2.

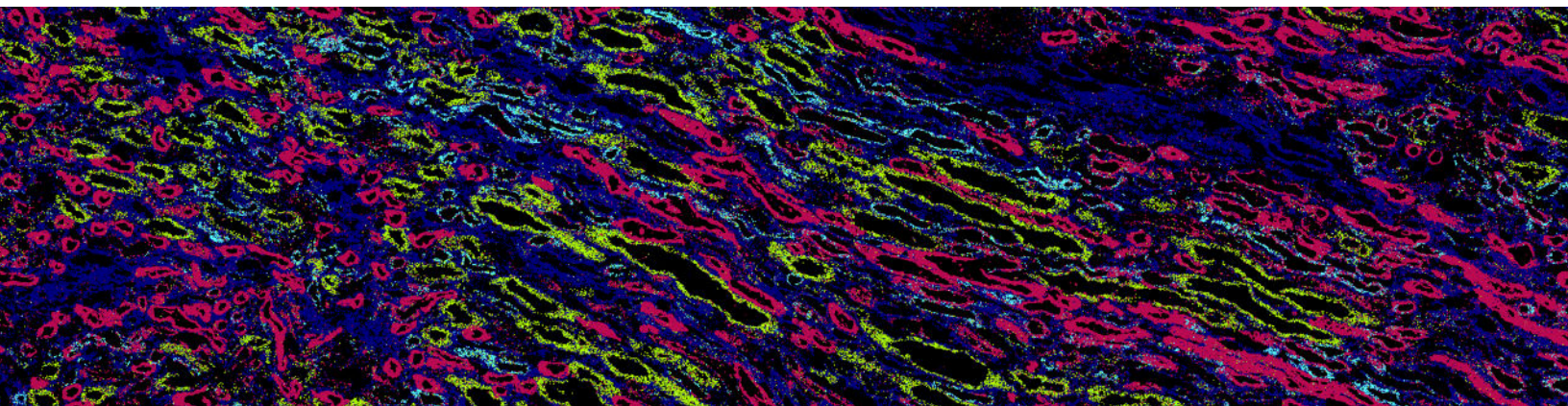
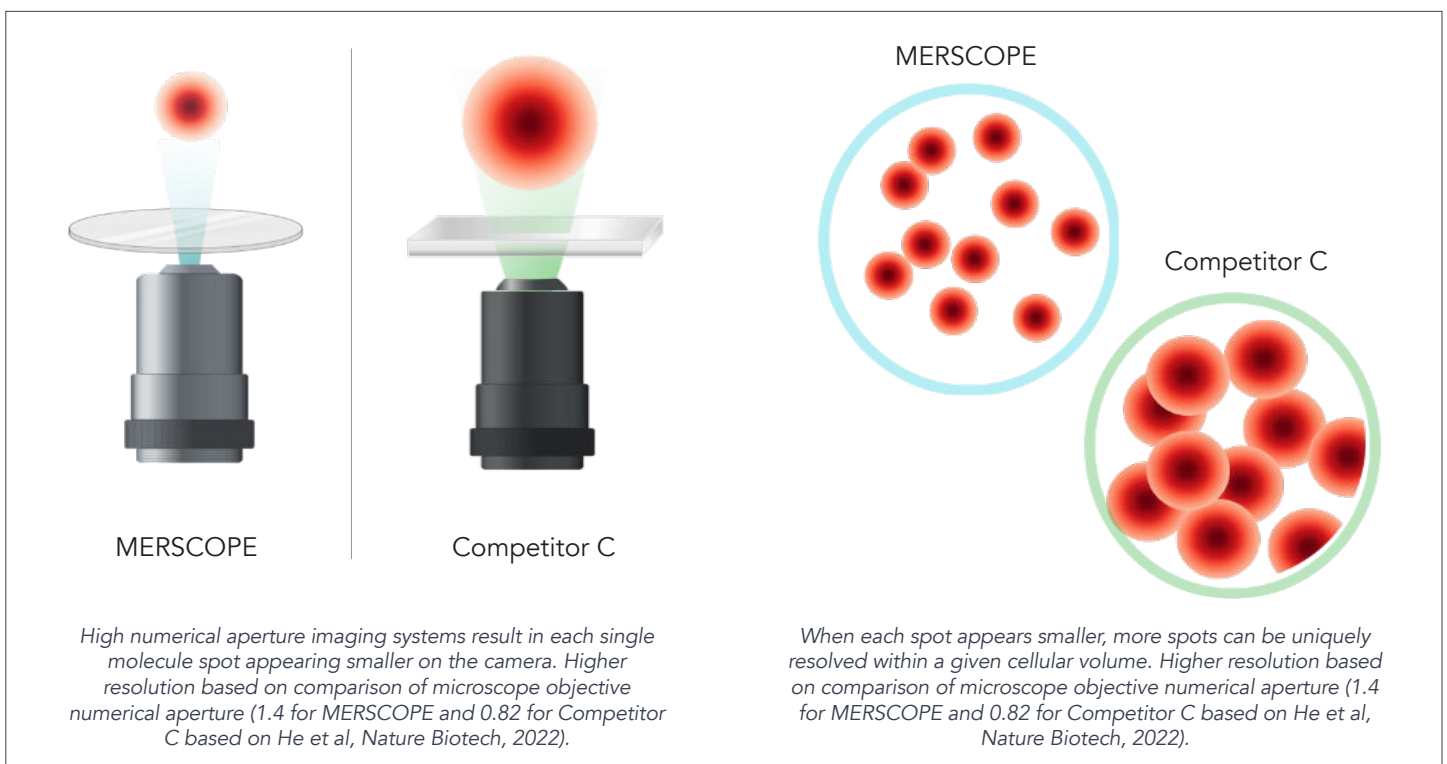
Information Density

Definition: The amount of information that can be measured within a given tissue volume or area.

Cells have limited volume, and the amount of information that can be detected within each cell directly impacts how much biology can be learned. Single molecule spatial transcriptomics approaches are microscopy-based and, consequently, the amount of information that can be measured per tissue volume directly relates to the quality of the optical system. MERSCOPE optics are optimized for single molecule detection. With a high numerical aperture objective lens, more fluorescence is detected from each individual transcript and the transcript appears smaller on the camera. In fact, MERSCOPE optics are more than 5-times smaller volumetrically than a transcript imaged by Competitor C. With each transcript appearing smaller, more transcripts can be detected within a given tissue volume and more information can be detected per cell.

Why Information Density Matters

Each spatial genomics platform has a limit to how many transcripts can be accurately detected per unit volume with high accuracy and specificity. Detecting more information per cell enables more finely resolving transcriptional differences between cells and identifying more subtle variations in cell type and state across a biological tissue. Going beyond the information density limit may compromise accuracy and sensitivity, resulting in less biologically consistent data.



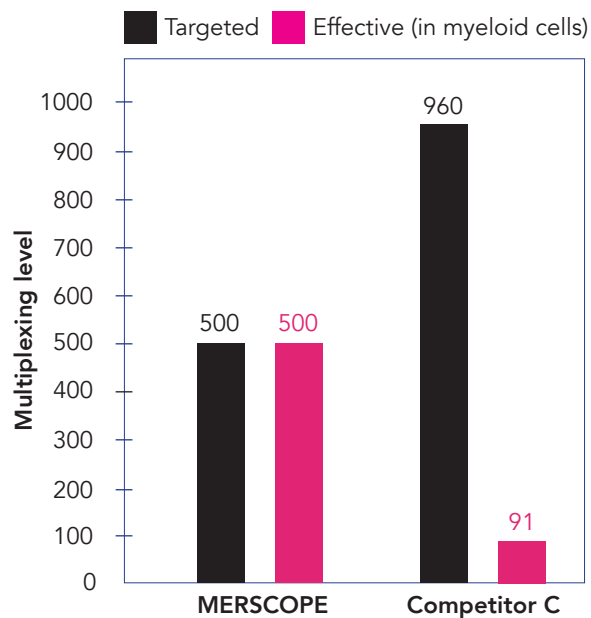
Effective Multiplexing Capacity

Definition: The number of RNAs a spatial genomics technology can detect in a single experiment.

It is important to note that true multiplexing capacity is sometimes limited not by the number of genes targeted (the targeted multiplexing capacity) but by the noise floor, or ambient background signal. Genes expressed at a level dominated by noise are not effectively measured. For a highly specific and sensitive measurement, the effective multiplexing capacity may match the targeted multiplexing capacity, where all targeted genes are effectively measured. With lower specificity and sensitivity, there may be a large gap between the targeted multiplexing capacity and the effective multiplexing capacity, with only a few genes measured accurately despite significantly more originally targeted by the measurement. Effective multiplexing, therefore, can be understood as the number of genes accurately measured above the noise floor compared to the number of genes probed.

Why Effective Multiplexing Matters

Highly multiplexed measurements may not be useful if only a small percentage of genes can be reliably detected and identified. Different cell types within a tissue express different RNAs. To obtain complete and precise spatial context, researchers need accurate detection of hundreds of RNAs within the same sample.

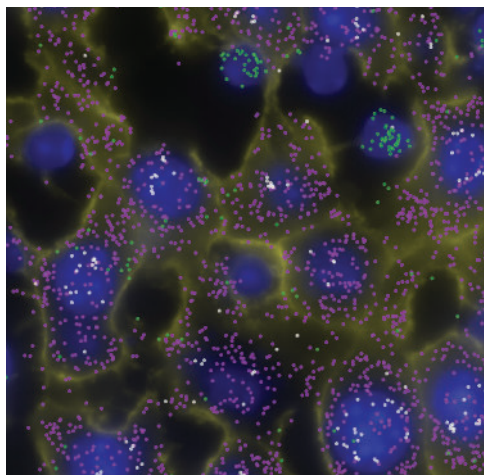


Targeted and effective multiplexing levels for MERSCOPE and Competitor C in myeloid cells. The effective multiplexing for competitor C was determined using a binary change point algorithm on mean counts for each gene detected by Competitor C sorted by the corresponding abundance of the gene in scRNA-seq to identify the transition from low correlation with scRNA-seq to high correlation. Data comparison between Vizgen public dataset (vizgen.com) sample HuLuCaPt2 and NanoString public dataset (nanosttring.com) sample Lung9_2.

MERSCOPE: Designed to Enable Biological Discovery

There has been an explosion of academic approaches to spatial genomics, but many are difficult to set up, inconsistent, and require extensive troubleshooting. Commercial platforms are slowly emerging, but as of fall 2022, Vizgen has released the largest data sets. Systematic benchmarking across methods is required to draw apples-to-apples conclusions about the respective capture performance of the different methods.

However, by prioritizing critical quality attributes, it is possible for laboratories to select tools that will deliver the best sensitivity, specificity, and biologic consistency in spatial genomics measurements.



MERFISH is a FISH based technology, involving direct hybridization to the target transcripts with no need for enzymatic amplification. Some alternatives use rolling circle amplification to improve signal intensity.

Limitations of rolling circle amplification-based spatial techniques include:

- Low detection efficiency: not all probes are ligated and amplified
- Biased RNA detection: ligation and amplification efficiency may depend on the specific target sequence
- Expensive and more sensitive sample prep: enzymes drive up the cost of reagents and are more susceptible to degradation than DNA oligos

As demonstrated in examples throughout this resource, data quality metrics are interconnected—for instance, low accuracy can decrease true detection efficiency and dilute multiplexing capabilities. MERSCOPE provides error-robust barcode correction and matrix embedding and clearing, providing a platform that doesn't limit your experimental options or ask you to sacrifice data quality in one area in favor of another.

Key Quality Metric	Sensitivity	Specificity	Information Density	Effective Multiplexing Capacity
Summary of MERSCOPE key metrics	More transcripts of targeted genes detected than scRNA-seq.	Specific detection with strong correlation with bulk sequencing across all abundances.	Optimized optics with 1.4NA.	Equal to targeted multiplexing capacity (up to 500 genes).

MERSCOPE also offers an expanded range of capabilities while retaining the same high data quality:

- Custom panel design – select the genes most relevant for your biological question
- Cell boundary staining – accurate cell segmentation and single-cell analysis
- Multiomics – measure RNA and protein within the same sample

With high-quality data, researchers like you can uncover biological complexity at every level, from subcellular expression to cell type differentiation within tumors, organs, and important tissues. Report results with confidence to advance your field and make a lasting impact.



Visit www.Vizgen.com/product to explore MERSCOPE, the end-to-end platform for getting high-quality data from MERFISH experiments.

References:

1. Method of the year 2020: Spatially resolved transcriptomics. *Nature Methods* **18**, (2021).
2. Zhang, M., Eichhorn, S.W., Zingg, B., et al. Spatially resolved cell atlas of the mouse primary motor cortex by merfish. *Nature* **598**, 137–143 (2021).
3. Moffitt, J. R. & Zhuang, X. RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). *Methods in Enzymology* **572**, 1–49 (2016).
4. Moffitt, J. R., Hao, J., Bambah-Mukku, D., et al. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *PNAS* **113**, 14456–14461 (2016).