Designing Custom Xenium Panels

Introduction

Xenium In Situ measures gene expression in tissue sections derived from either formalin fixed & paraffin embedded (FFPE) or fresh frozen (FF) tissue samples placed on Xenium slides. Gene expression is measured via the use of pre-designed panels of probes that target transcripts of interest. 10x Genomics pre-designed Xenium panels are designed to cover all major cell types found in a tissue. However, researchers may also be interested in selecting markers for specific subpopulations of cells, or genes expressed under specific conditions (e.g., disease).

This document provides specific guidance for selecting genes to include on Xenium add-on panels. 10x Genomics recommends the use of relevant single cell RNA-seq data (ideally, matched single cell data obtained from the same sample or condition) to identify such marker genes and to estimate overall panel performance. Once genes are selected, customers will work with 10x Genomics to finalize details of the panel.

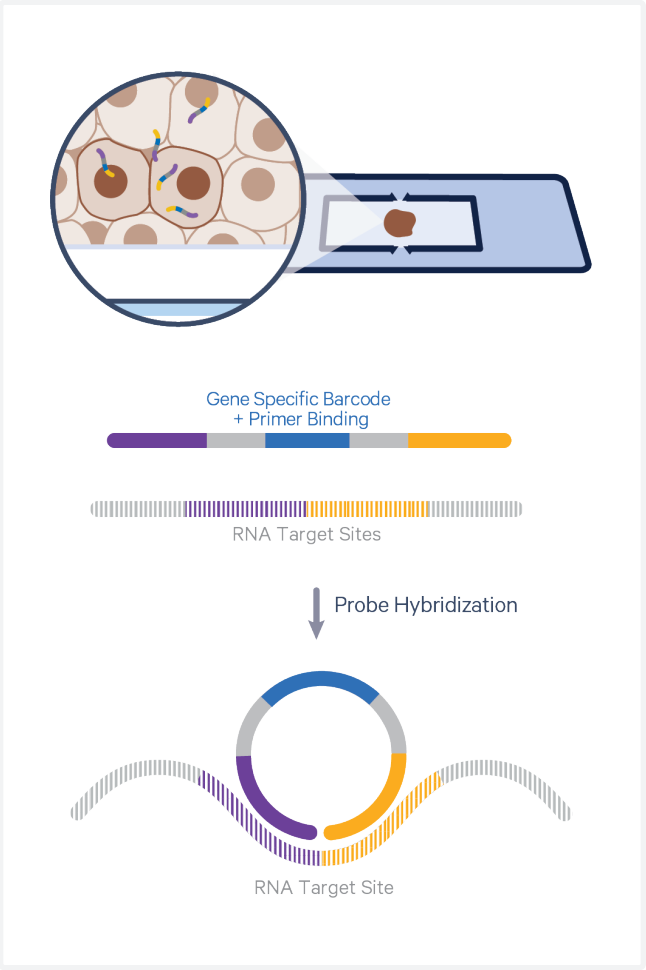


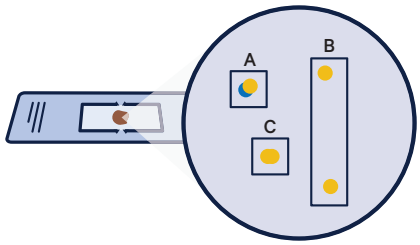
Figure 1. Probe hybridization during the Xenium In Situ workflow. Custom probes designed bind to RNA target sites in the sample. After probe hybridization and ligation, rolling circle amplification products are generated, which are detected by the Xenium instrument.

Background

Xenium add-on panels allow researchers to design customized Xenium panels by choosing up to 100 genes to add to a pre-designed Xenium panel. Genes present on the pre-designed panel are referred to as "base panel genes", whereas the genes selected by the researcher are referred to as "add-on panel genes". Researchers can select specific add-on panel genes based on their specific needs. This document provides guidance on selecting add-on panel genes that maximize insights gained from Xenium In Situ data, additional background on recommended expression levels for candidate genes, and how to use single cell RNA-seq data to assess the expression levels of individual candidate genes.

In Situ Detection Budget

Xenium In Situ uses fluorescent microscopy to detect individual transcripts in cells. Like all imaging-based technologies, only a finite number of fluorescent signals can be distinguished within a given area or volume, such as a cell. With Xenium, each transcript is labeled with fluorescent dyes in a subset of the cycles and fluorescent channels.



	Detection Cycle/ Channel	Distance of Events	Events *(Detected/ Present)*	Optical Crowding
A	Different	Close	2/2	**No**
B	Same	Far	2/2	**No**
C	Same	Close	1/2	**Yes**

Figure 2. Visualization of optical crowding. Optical crowding occurs when a detection event occurs in the same detection cycle and channel and if the distance between events is small.

However, once the total number of transcripts targeted in a cell type exceeds a certain limit, the number of fluorescent signals that are too close to each other will increase significantly. This effect is referred to as optical crowding (Figure 2). In certain cases, optical crowding can be compensated for using decoding algorithms. This restricts the effect of optical crowding to a small number of highly expressed genes, thus not significantly affecting data quality. In other cases, crowding can lower the sensitivity for most or all genes on a panel and introduce technical artifacts, resulting in poor data quality. The goal of successful panel design is to select genes that avoid exceeding the finite detection budget, thus avoiding crowding. 10x Genomics pre-designed panels leave unused detection budget, enabling the inclusion of add-on panels (Figure 3).

Ideal genes for Xenium profiling have highly cell type-specific, biologically-informative expression patterns. These genes only consume budget in a few cell types. In contrast, "housekeeping" genes with broad or ubiquitous expression (e.g., ACTB,

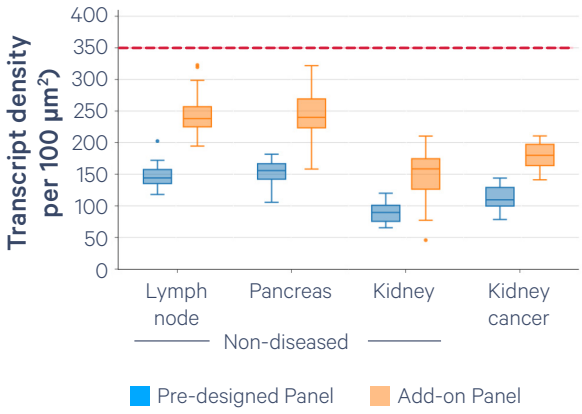


Figure 3. Per-cluster transcript densities observed with the pre-designed Human Multi-Tissue and Cancer Panel and an add-on panel. Densities are based on zero-expanded segmentation, where transcripts are only assigned within the nucleus as determined by the algorithm based on DAPI staining. Pre-designed panel leaves significant unused detection budget to allow for the inclusion of Add-on Panel. Red dotted line indicates recommended detection budget to limit sensitivity loss due to crowding.

GAPDH) are poor candidates for Xenium profiling, as they take up a disproportionate fraction of the available detection budget in most or all cells.

Extremely Highly Expressed Genes

Sometimes, genes that are of biological interest are also extremely highly expressed. Unlike housekeeping genes, these extremely highly expressed genes (EHEGs) are typically only expressed in specific cell types. While EHEGs can result in the generation of a large number of puncta in individual cells, Xenium In Situ was designed to enable robust performance in the presence of such genes. Therefore, a limited number of EHEGs can be included on Xenium Custom Panels without compromising overall data quality.

To understand how EHEGs can affect transcript detection, consider the example of the Insulin gene (INS). INS is only highly expressed in pancreatic beta cells. Assume that in a given pancreatic beta cell, rolling circle amplification products (RCPs) are generated for 300 INS transcripts and all get detected in the same cycle, causing a significant amount of optical crowding (Figure 4) . As a result, only 100 of the 300 transcripts are confidently detected. Though detection sensitivity has been reduced to 33% relative to the overall assay sensitivity, 100 transcripts per cell are more than enough to identify this cell as a pancreatic beta cell, since it is not expressed at high levels in any other pancreatic cell type. Any gene that is not expressed in beta cells will be unaffected, because INS only causes optical crowding in those cells. More generally, an EHEG will not affect any gene that is not co-expressed with it.

For example, assume that "Gene B" is a gene that is also expressed in beta cells. If the codeword for Gene B does not overlap with that of INS, there are no cycles and channels in which transcripts from both INS and Gene B are fluorescently labeled. In this situation, the transcript detection of Gene B is unaffected by INS. However, if there is an overlap,

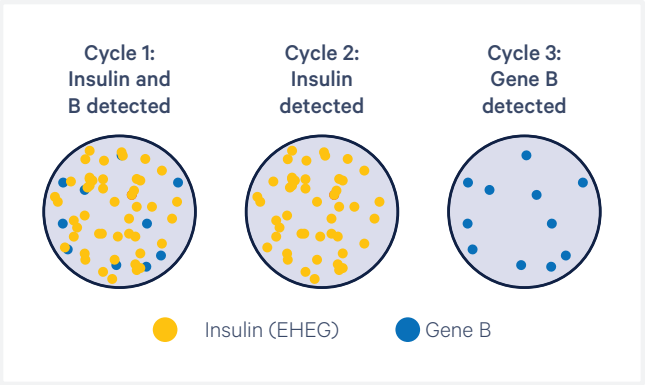


Figure 4. Visualization of gene detection across cycles. Optical crowding occurs due to an overlap in the codeword for INS and Gene B.

this means that in at least one cycle, transcripts of both genes are labeled, resulting in optical crowding (Figure 4, Cycle 1).

Xenium uses a sparse codebook, meaning that each gene is only detected in a subset of cycles and channels. This allows the panel design algorithm to try to select a codeword for Gene B that does not overlap with the INS codeword. If this fails, Gene B may share one or more cycles/channels with INS. In the affected cycles, a fraction of Gene B RCPs may fail to be detected due to optical crowding. Importantly, Gene B RCPs will still be detected in the unaffected channels.

The decoding algorithm is robust to a certain degree of optical crowding, and the exact impact of the presence of EHEGs depends on the specific codewords assigned, the number of EHEGs present, and other sample metrics. However, 10x Genomics conservatively estimate the presence of a single EHEG will have no direct impact on at least 95% of co-expressed genes, and the presence of two EHEGs will have no direct impact on at least 80% of co-expressed genes. The more EHEGs included on a panel, the higher the proportion of affected genes. 10x Genomics recommends not including more than two co-expressed EHEGs on a Xenium custom panel. Two or more EHEGs that have mutually exclusive expression patterns can be

included, as long as the panel stays within the detection budget. This will not carry a higher risk to data quality than the inclusion of a single EHEG.

Another way to mitigate crowding caused by EHEGs is to reduce the number of probesets used. By default, Xenium uses up to eight probesets to detect genes, and each contributes independently to overall detection sensitivity. If any EHEGs are in the panel, reduce the number of probesets used for those genes. Use at least two probesets per gene, as individual probesets may not perform as expected. If the degree to which the EHEG is expressed is unclear, use three probesets per EHEG.

Single Cell Gene Expression

The Xenium Custom Panel design process uses single cell expression data to calculate cell type-specific expression profiles, which represent a model of which genes are co-expressed. This model allows the panel design algorithm to assign codewords to genes in an optimal manner, making the best possible use of the available detection budget. Additionally, single cell data can be used to screen for broadly expressed "housekeeping" genes, as well as for EHEGs. For these genes, 10x Genomics will confirm that their inclusion in the Xenium Custom Panel is intentional. Single cell expression data is also used to model the expected optical density of the panel using an approach called utilization analysis, which provides a check of whether the panel exceeds the recommended detection budget. Finally, the number of probesets for any gene can be tuned to reduce the risk of that gene contributing to optical crowding in the panel.

General Recommendations for Gene Selection

High Cell Type/State Specificity

Select genes that are only expressed in specific cell types, or by cells in specific "states" (e.g., disease states, cells responding to certain signals, cell cycle states, etc.). The more specific the expression pattern of a gene, the easier it is to identify the individual cell types or cell states. Importantly, cell type-specific genes also make a more efficient use of the cell's finite detection budget (refer to Detection Budget). Assess the cell type specificity of a candidate gene using relevant single cell data, by comparing the normalized mean transcript

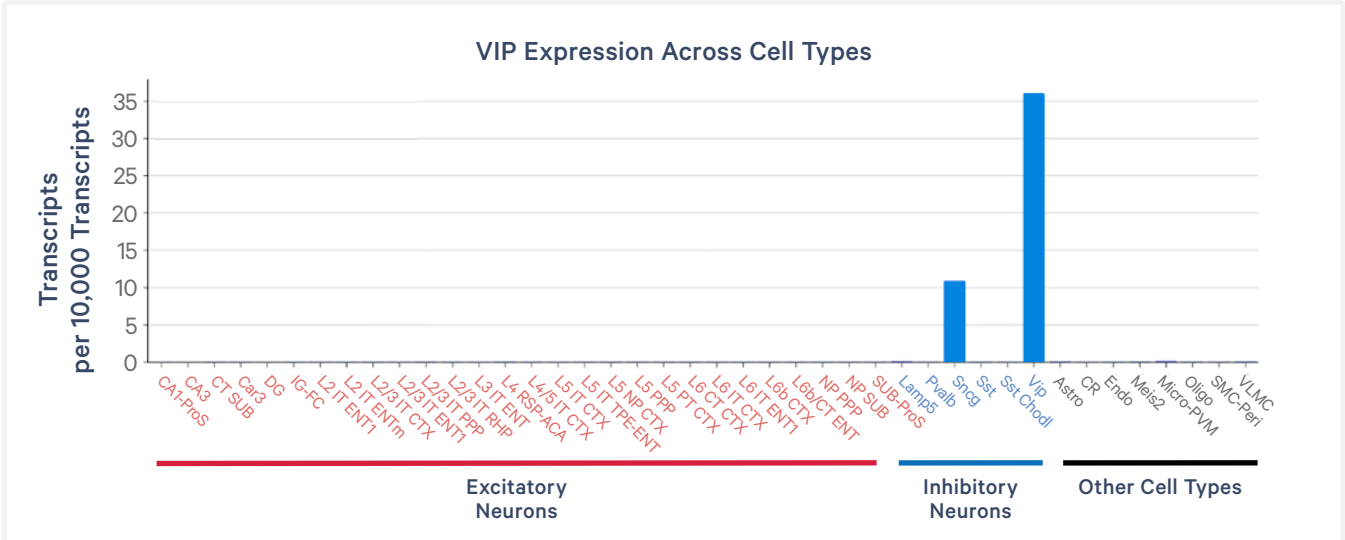


Figure 5. Single cell RNA-seq analysis of Vip, a gene with a highly cell-type specific expression pattern in the mouse brain (Yao et al., Cell 2021, doi: 10.1016/j.cell.2021.04.021). Mean transcript counts per cell type are normalized to 10,000 transcripts. Vip is only expressed in two out of the 42 cell types shown. Red labels denote excitatory neurons, blue labels denote inhibitory neurons, and black labels denote other cell types.

count of a gene across cell types. Refer to Figure 5 for an analysis of a gene (Vip) with a highly cell type-specific expression pattern.

Absolute Expression Level

The absolute expression level of a gene refers to the actual number of transcripts detected per cell. This number is an important determinant of how accurately a gene can be quantified in individual cells. If a gene is intended to serve as a cell type marker by itself, use Chromium Single Cell Gene Expression data to quantify the expression of that gene in the relevant cell type. For Chromium Next GEM Single Cell 3' v3.1 data (assuming at least 50% sequencing saturation), select genes with a mean expression of at least 4 transcripts per cell (2 transcripts per cell for Chromium Single Cell 3' v2 data).

Genes with very low absolute expression in a given cell type (<0.1 mean transcripts per cell) will likely only be detected in a very small proportion of cells; thus, including those genes may not be an efficient use of the limited number of add-on genes that can be included on a Xenium Add-on panel. In contrast, genes with extremely high absolute expression levels (>100 mean transcripts per cell in Single Cell 3' v3.1 data, or >50 mean transcripts per cell in Single Cell 3' v2 data) may be problematic as they can result in optical crowding (refer to Detection Budget). Refer to Figure 6 for an example of a gene (Plp1) with extremely high absolute expression in mouse brain oligodendrocytes. Replace such genes with co-expressed genes that have lower expression levels. If possible, use relevant single cell data to identify suitable replacement genes. For example, to replace Plp1, the single cell dataset by Yao et al. (Cell, 2021) allows the identification of several other mouse genes that are also highly specific to oligodendrocytes, but are expressed at lower levels.

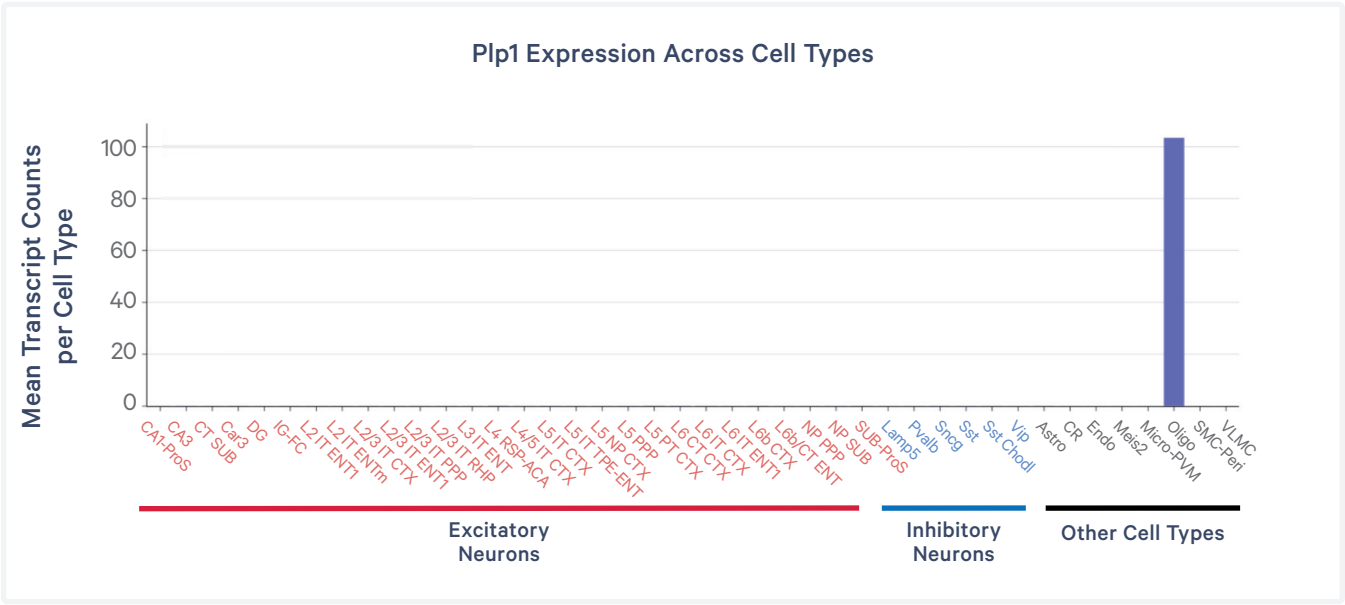


Figure 6. Single cell RNA-seq analysis of Plp1, a gene with an extremely high absolute expression level in mouse brain (Yao et al., Cell 2021, doi: 10.1016/j.cell.2021.04.021). Shown are the mean transcript counts per cell type. Plp1 has approximately 100 mean transcripts per cell detected in this Chromium Single Cell 3' v2 dataset, double the recommended expression limit of 50 mean transcripts per cell (100 mean transcripts per cell for v3.1 data).

Chemistry	Low Expression Threshold (Mean Transcripts per Cell)	High Expression Threshold (Mean Transcripts per Cell)
Chromium Single Cell 3' v2	<0.1	>50
Chromium Single Cell 3' v3.1		>100

Table 1. Low and high expression thresholds for ideal Xenium add-on panel genes.

Conclusions

This Technical Note provides guidance on the identification of genes for generating Xenium add-on panels.

Genes added onto Xenium add-on panels should:

- Be cell type or cell state-specific.
- Not be below the low expression threshold (<0.1 mean transcripts per cell)
- Not exceed the high expression threshold (>100 mean transcripts per cell)

A well-designed panel is key to accurately quantifying the expression of genes of interest in a given sample. Understanding the specificity of a gene with respect to cell type or state and its expression level will allow the design of a more robust and representative Xenium add-on panel.

References

1. Yao, Z., van Velthoven, C., Nguyen, T. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell 184,12 (2021). https:// doi: 10.1016/j.cell.2021.04.021

Document Revision Summary

Document Number	CG000643
Title	Xenium Add-on Panel Design Technical Note
Revision	Rev A to Rev B
Revision Date	September 2023

Specific Changes:

Updated information regarding optical crowding, including expanded text and additional images

General Changes:

Updated for general minor consistency of language and terms throughout.

Contact:
support@10xgenomics.com
10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94588 USA