TECHNICAL NOTE

Targeted Visium Spatial Gene Expression: Pre-designed Panel Performance Metrics

Introduction

The Targeted Gene Expression product is a modular enrichment kit designed to enrich libraries for relevant genes, while decreasing sequencing requirements by up to 90%. Pre-designed panels from 10x Genomics are used in the Targeted Gene Expression workflow to target over 1,000 genes per panel. When used to enrich whole transcriptome Visium Spatial Gene Expression libraries, these panels enable significantly reduced sequencing costs while accurately reflecting gene expression information from the corresponding whole transcriptome analysis (WTA) parent library for targeted genes. This Technical Note describes the performance metrics for pre-designed panels as tested across several human tissue types. This document shows that the use of pre-designed panels with targeted libraries results in even enrichment and accurate representation of on target molecules relevant to the corresponding Visium library.

Method

Visium Spatial Gene Expression libraries were previously prepared as described with the Visium Spatial Gene Expression Reagent Kits User Guide (CG000239, Rev D) and sequenced for whole transcriptome analysis (WTA). The libraries were then used to generate Targeted Gene Expression libraries (Figure 1) using the reagents and protocol described in the Targeted Gene Expression - Spatial User Guide (Document CG000377) with the following pre-designed human panels: Human Gene Signature (PN-1000245), Human Pan-Cancer Panel (PN-1000247), Human Immunology Panel (PN-1000246), and Human Neuroscience Panel (PN-1000278). Panels were tested across 21 human tissue types. Six tissue types were fixed and stained using the Methanol Fixation, Immunofluorescence Staining, & Imaging Demonstrated Protocol (Document CG000160), while the remaining tissues were fixed and stained using the Methanol Fixation, H&E Staining, & Imaging Demonstrated Protocol (CG000312). Twelve tissue types had an average cDNA length of less than 700 bp and were processed according to the protocol modifications outlined in the Targeted Gene Expression User Guide. The remaining tissue types were processed using the standard (long cDNA) protocol. Data were analyzed by comparing parent and targeted data on a matched (parent derived) barcode set, at the sequencing depths described below, to compute the metrics in Figure 2.

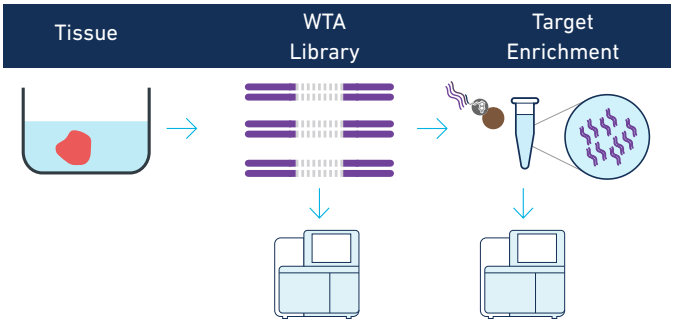

Figure 1. Targeted Gene Expression workflow overview. Whole Transcriptome Analysis (WTA) libraries are generated from tissue. WTA libraries are incubated with biotinylated baits to generate targeted libraries. WTA and targeted libraries are both sequenced. These data are used for metrics comparisons.

Sequencing Depth

To calculate parent-targeted comparison metrics, both libraries were computationally downsampled and compared at the following depths:

Parent WTA Libraries: 30,000 read pairs per tissue-covered spot.

Targeted Gene Expression Libraries: two-fold the number of on target panel reads in the WTA parent library.

Example: WTA library with 5% on target reads = 3,000 read pairs per tissue-covered spot (0.05 x 30,000 x 2) for the associated target library. Sequencing libraries were computationally downsampled and compared at the read numbers described in the calculation above. This ranged from 1,136-8,472 raw read pairs per tissue-covered spot.

Metrics Analyzed

Targeting Metrics:

- **Reads mapped confidently to the targeted transcriptome**: Fraction of reads that mapped to a unique and targeted gene in the transcriptome. The read must be consistent with annotated splice junctions. These reads are considered for UMI counting. This metric can be calculated for the parent WTA library, the targeted library, or both, and does not require a matched WTA library control.
- **Targeted total UMI recovery**: Fraction of spot-associated, on target UMIs recovered (pseudo-bulk) at the sequencing depth described above. Requires a matched WTA library control. This metric is calculated on matched cell barcodes across WTA and targeted libraries.
- **Fraction targeted genes with ≥80% UMI recovery**: Fraction of observed (defined as ≥ 10 UMI in WTA library) panel genes in a WTA library control for which ≥80% of UMIs are recovered in the matched targeted library at the sequencing depth described above. Requires a matched WTA library control.
- **Pseudo-bulk targeted UMI R^2**: Pearson correlation coefficient of cell-associated panel UMIs across WTA and targeted libraries at the sequencing depth described above. Requires a matched WTA library control.

Results

Targeting Metrics

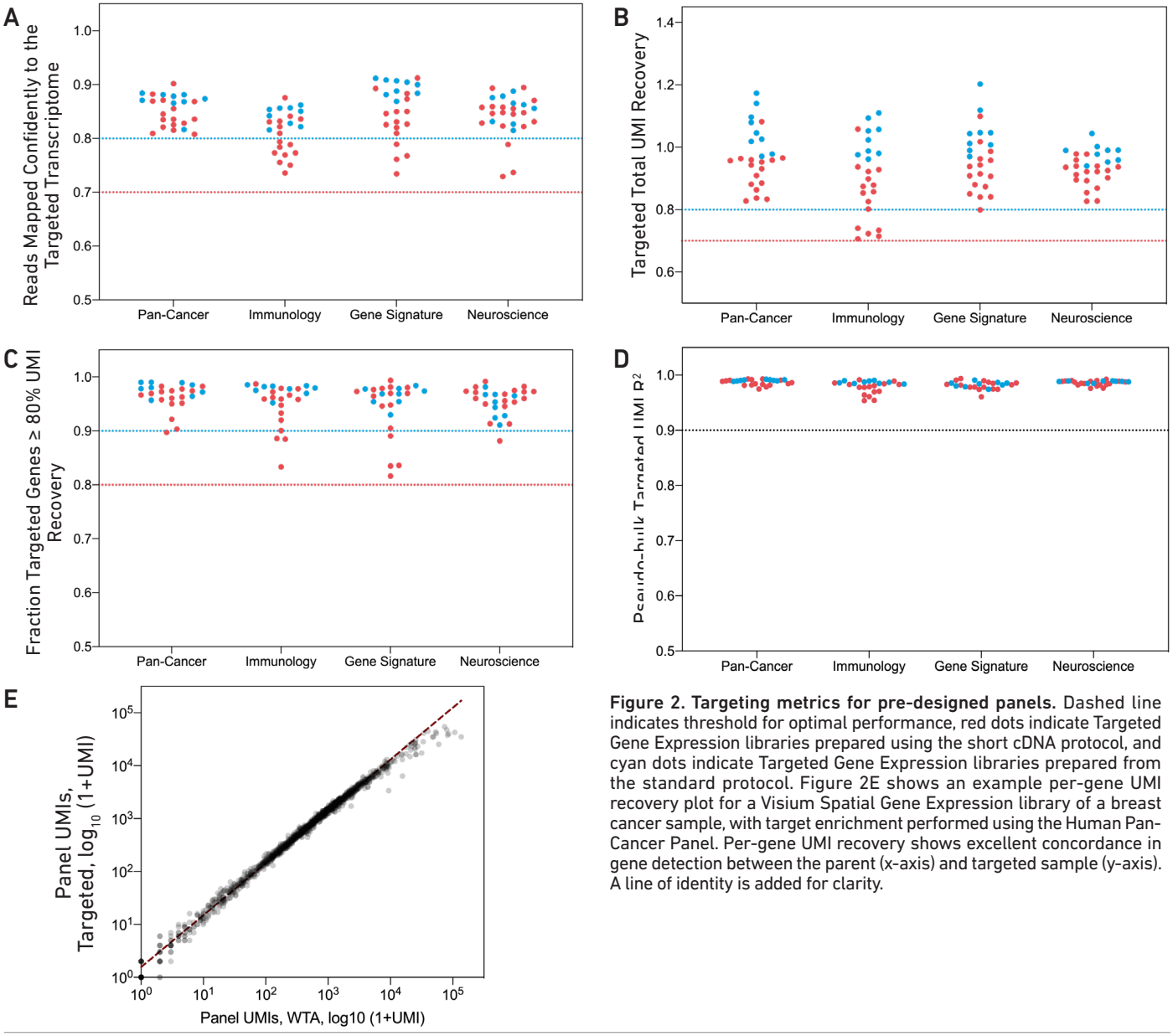

Figure 2. Targeting metrics for pre-designed panels. Dashed line indicates threshold for optimal performance, red dots indicate Targeted Gene Expression libraries prepared using the short cDNA protocol, and cyan dots indicate Targeted Gene Expression libraries prepared from the standard protocol. Figure 2E shows an example per-gene UMI recovery plot for a Visium Spatial Gene Expression library of a breast cancer sample, with target enrichment performed using the Human Pan-Cancer Panel. Per-gene UMI recovery shows excellent concordance in gene detection between the parent (x-axis) and targeted sample (y-axis). A line of identity is added for clarity.

Results

Protocol Modifications for Visium Spatial Gene Expression Libraries Prepared from Short cDNA

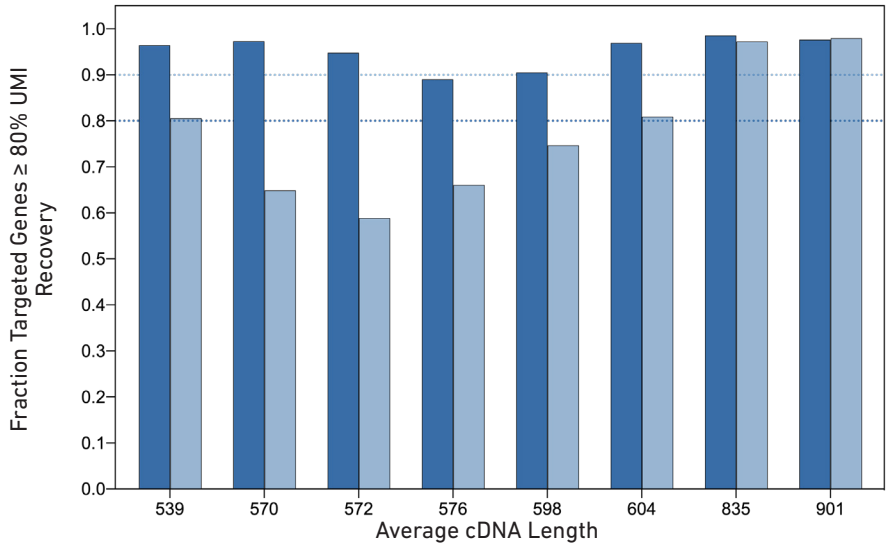

Figure 3. Effect of short cDNA protocol modifications on Fraction Targeted Genes ≥ 80% UMI Recovery. Dark blue: short cDNA protocol, Light blue: standard cDNA protocol. Paired bars indicate the same original parent library. Dashed lines represent thresholds for optimal performance for short cDNA samples (dark blue) and long cDNA samples (light blue). Average cDNA size (x-axis) was determined by BioAnalyzer.

Discussion

Targeting Metrics:

Thresholds for optimal performance were set for each calculated targeting metric and are depicted as dotted lines in Figure 2. All Visium Spatial Gene Expression libraries prepared from short cDNA (66/66) and long cDNA (36/36) passed the reads mapped confidently to the targeted transcriptome metric as shown in Figure 2A. All libraries passed metrics that measure library complexity when comparing the targeted sample to its parent WTA library: targeted total UMI recovery (Figure 2B), fraction targeted genes with ≥80% UMI recovery (Figure 2C), and pseudo-bulk targeted UMI R^2 (Figure 2D).

UMI correlations between parent and targeted samples exceeded the threshold of R^2 >0.9 for all samples tested. A breast cancer sample, target enriched using the standard protocol, showed a pseudo-bulk per-gene UMI recovery R^2 of 0.992 (Figure 2E).

Customers can compare parent and targeted libraries using the targeted-compare pipeline in Space Ranger. This will enable visualization of read enrichment, UMI recovery, and cell clustering between parent and target-enriched samples.

Targeted total UMI recovery rate ranged from 71-120% (Figure 2B). Total UMI recovery greater than 100% is common in tissue types with high RNA content. For these sample types, 30,000 read pairs per tissue-covered spot is not sufficient to detect all panel UMIs in the WTA library. The targeted library is compared using 2-fold read pairs per tissue-covered spot as compared with the number of on target reads in the parent library. These additional sequencing reads are able to detect unique molecules in the targeted library that were not detected in the parent library. Efficient target enrichment allows for deeper sequencing of genes of interest, while maintaining a >80% decrease in sequencing reads.

Library cDNA Length:

Slight performance differences are expected between Visium Spatial Gene Expression libraries prepared from short cDNA compared to libraries prepared from long cDNA, as shown by different thresholds used to assess reads mapped confidently to the targeted transcriptome (70% and 80% respectively), targeted total UMI recovery (70% and 80% respectively), and fraction targeted genes with ≥80% UMI recovery (80% and 90% respectively). Libraries prepared from short cDNA typically have shorter baited transcript length, leading to less efficient enrichment compared to libraries prepared from long cDNA.

The protocol modifications for Visium Spatial Gene Expression libraries prepared from short cDNA outlined in the Targeted Gene Expression - Spatial User Guide are necessary to achieve optimal performance when working with libraries prepared from cDNA with an average length <700 bp. Short cDNA samples with an average cDNA length from 539-604 bp showed an average improvement of 24.5% (16.5-37.9%) in fraction targeted genes with ≥80% UMI recovery when processed through the short cDNA protocol compared to the standard protocol (Figure 3). Conversely, long cDNA samples (835 bp, 901 bp) did not show a significant difference in library complexity when processed through the short cDNA protocol compared to the standard protocol and are compatible with either protocol depending on pooling strategy.

Conclusion

As shown by the metrics evaluated in this Technical Note, 10x Genomics pre-designed human panels deliver a high recovery of unique molecules and efficient targeting of panel genes over a variety of sample and tissue types.

References

- Targeted Gene Expression - Single Cell User Guide (CG000293)
- Targeted Gene Expression - Spatial User Guide (CG000377)

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