TECHNICAL NOTE

Targeted Chromium Single Cell 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 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 cell 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 WTA library.

Method

Chromium Single Cell 3' Gene Expression (v3, v3.1) and 5' Gene Expression (v1, v2) libraries that were previously sequenced for whole transcriptome analysis (WTA) were used to generate Targeted Gene Expression libraries (Figure 1). Targeted libraries were generated using the reagents and protocol described in the Targeted Gene Expression - Single Cell User Guide (Document CG000293) 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 samples including cell lines, peripheral blood mononuclear cells (PBMCs), and dissociated tumor cells. 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 4. Cell calling concordance (Figure 5) was analyzed by comparing the parent and targeted library cell calls generated by the count pipeline in Cell Ranger 4.0.

To confirm the absence of PCR artifacts like chimeric molecules, the chimeric index rate during the target enrichment workflow was measured using dual index sequencing libraries. Four Single Cell 5' Gene Expression libraries with unique, dual i5 and i7 index reads were pooled for target enrichment using the same pre-designed panel (Figure 2). Separately, the same libraries were target enriched alone (Figure 3). Library pools were sequenced on one Illumina MiSeq flowcell. Single enriched libraries were pooled and sequenced on a separate MiSeq flowcell to control for index hopping.

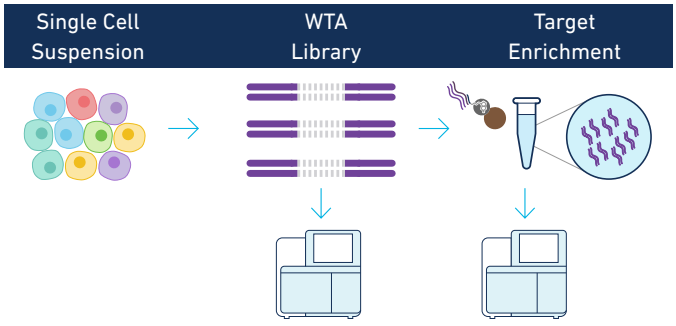

Figure 1. Targeted Gene Expression workflow overview. Whole Transcriptome Analysis (WTA) libraries are generated from single cell suspensions. WTA libraries are incubated with biotinylated baits to generate targeted libraries. WTA and targeted libraries are both sequenced. These data are used for metrics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

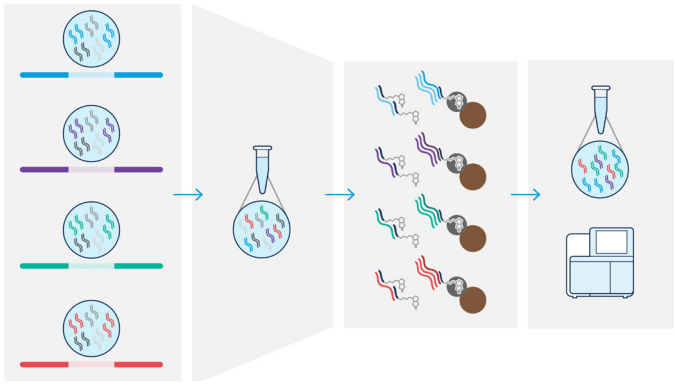

Figure 2. Chimeric index rate experiment - library pool enrichment workflow. Chromium Single Cell 5' Gene Expression libraries were combined prior to target enrichment.

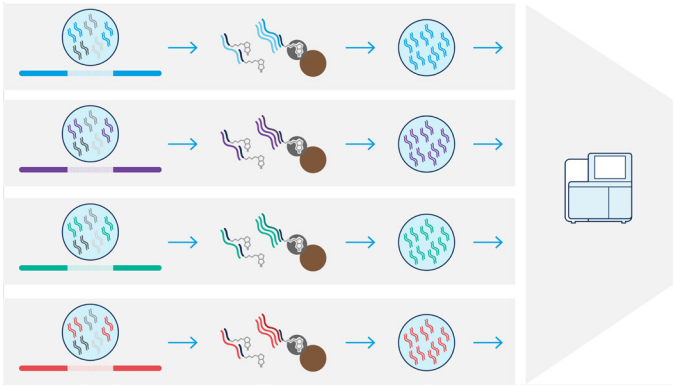

Figure 3. Chimeric index rate experiment - single library enrichment workflow. Chromium Single Cell 5' Gene Expression libraries were target enriched alone before pooling for sequencing.

Sequencing Depth

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

Parent WTA Libraries: 50,000 read pairs per cell.

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

Example: WTA library with 6% on target reads = 6,000 read pairs per cell (0.06 x 50,000 x 2) for the associated target library. Sequencing libraries were computationally downsampled and compared at the read numbers described in the calculation above.

Metrics Analyzed

Chimeric Index Rate: The fraction of sequencing reads on dual index libraries that have a mismatched i5 and i7 index. The chimeric index rate is calculated on libraries that are enriched alone and pooled for sequencing and compared with libraries that are enriched as a library pool.

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 cell-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.

Cell calling concordance: A comparison of the number of cells detected, calculated as the number of cells detected in the targeted library divided by the number detected in the WTA library.

Results

Chimeric Index Rate:

Table 1 below compares the chimeric index rate between individually enriched samples and samples pooled upstream of enrichment. These two sample types had the same rate of chimeric indices detected; thus, the enrichment PCR does not result in appreciable levels of chimeric index molecules in the final library.

An analogous experiment was performed using Single Cell 3'v3.1 libraries which yielded similar results (data not shown). Libraries enriched alone and pooled before sequencing had the same chimeric index rate as libraries that were pooled upstream of enrichment.

Chromium Single Cell 5' Gene Expression Library Chimeric Index Rates		
Sample ID	Single Library Enrichment	Library Pool Enrichment
Sample 1	0.77%	0.68%
Sample 2	0.80%	0.70%
Sample 3	1.17%	1.02%
Sample 4	0.83%	0.74%

Table 1. Chimeric index rates for Dual Index Chromium Single Cell 5' Gene Expression Libraries.

Results

Targeting Metrics

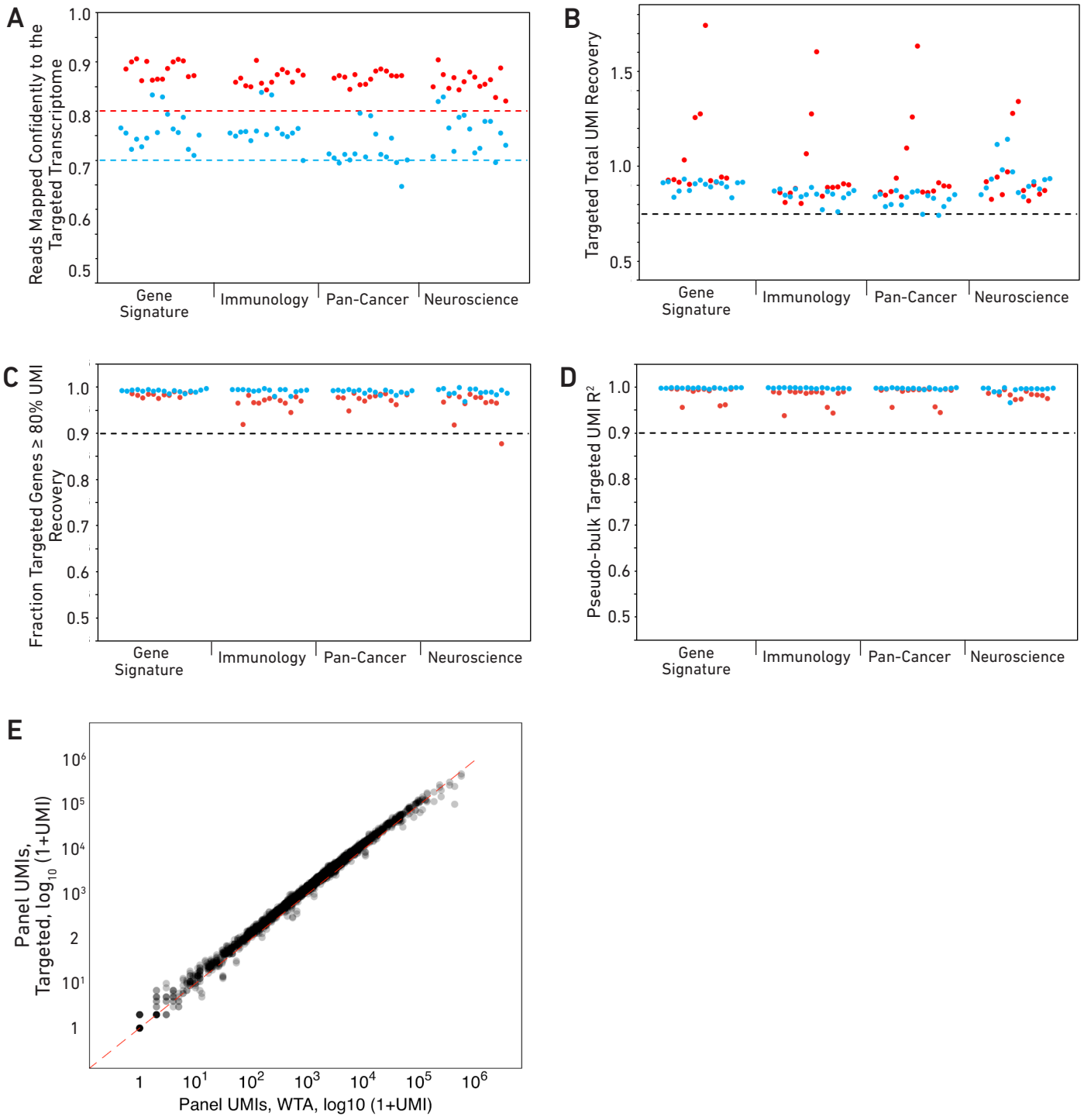

Figure 4. Targeting metrics for pre-designed panels. Dashed line indicates threshold for optimal performance, red dots indicate Single Cell 3' Gene Expression libraries, and cyan dots indicate Single Cell 5' Gene Expression libraries. Thresholds for fraction of reads on target differ for Single Cell 3' and 5' libraries for reasons described in the Discussion. Figure 4E shows an example per-gene UMI recovery plot for a Single Cell 3'v3 Gene Expression library of a human glioblastoma cell 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 red line of identity is added for clarity.

Results

Cell Calling Concordance

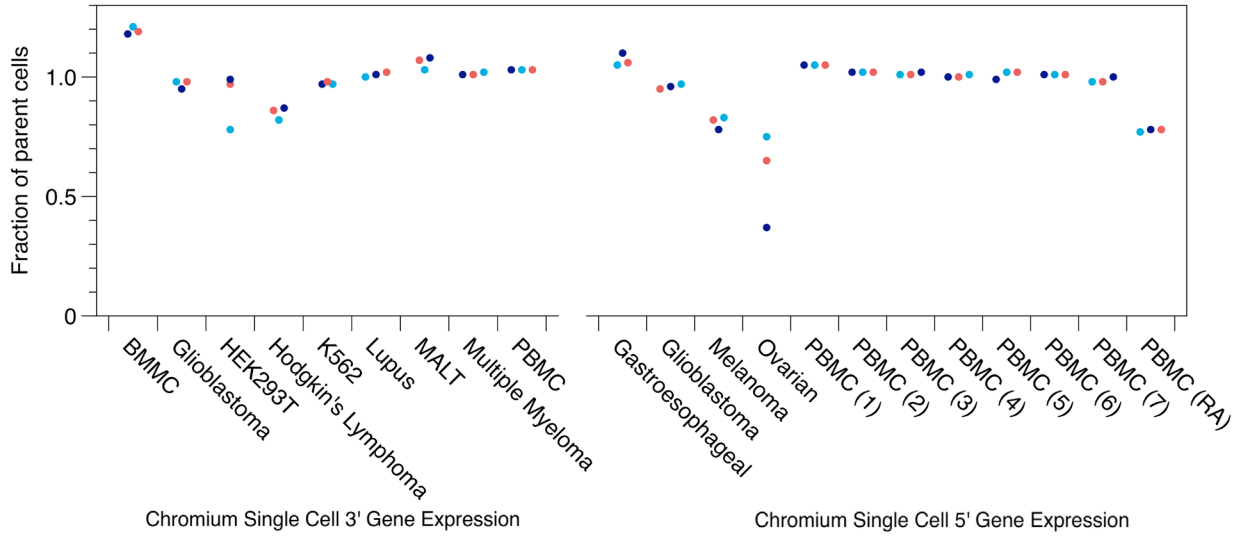

Figure 5. Cell calling concordance between WTA and targeted libraries. In most cases, cell calling concordance is close to 1.0. Cell calling concordance may vary for reasons described in the Discussion. Dark blue: Human Gene Signature Panel, Cyan: Human Immunology Panel, Red: Human Pan-Cancer Panel. MALT refers to mucosa-associated lymphoid tissue lymphoma, BMMC refers to primary bone marrow mononuclear cells, and PBMC refers to peripheral blood mononuclear cells.

Discussion

Chimeric Index Rate:

As shown in Table 1, a very low level of incorrectly paired sample indices were present in both single libraries and library pools. This indicates that no detectable index exchange occurs during post-enrichment PCR. All mispaired indices in the library pool were within the margin detected in the single library reactions, indicating that index hopping on the Illumina sequencer is the only detectable source of intra-sample index exchange. Similar experiments were executed with Single Cell 3' Gene Expression libraries, with results showing similarly low levels of incorrectly paired indices (data not shown).

Targeting Metrics:

Thresholds for optimal performance were set for each calculated targeting metric and are depicted as dotted lines in Figure 4. All Single Cell 3' Gene Expression libraries passed the reads mapped confidently to the targeted transcriptome metric, and almost all (48/51) of Single Cell 5' Gene Expression libraries passed the same metric as shown in Figure 4A. Nearly all libraries, including those with lower-than-optimal reads on target, passed metrics that measure library complexity when comparing the targeted sample to its parent WTA library: targeted total UMI recovery (Figure 4B), fraction targeted genes with >80% UMI recovery (Figure 4C), and pseudo-bulk targeted UMI R^2 (Figure 4D). One Single Cell 3' Gene Expression library generated from a lymph node nuclei sample with the Human Neuroscience Panel did not pass the fraction targeted genes with >80% UMI recovery metric, but only marginally (~2.9%, Figure 4C). Metric performance is driven by sample type and quality.

UMI correlations between parent and targeted samples exceeded the threshold of R^2 >0.9 for all samples tested, as shown by the glioblastoma example in Figure 4E.

Customers can compare parent and targeted libraries using the targeted-compare pipeline. This will enable cell calling comparisons between parent and target-enriched samples, as well as visualize read enrichment, UMI recovery, and cell clustering.

Targeted total UMI recovery rate ranged from 81-180%. High total UMI recovery is common in high RNA cell types, such as cell lines. For these sample types, 50,000 read pairs per cell is not sufficient to detect all panel UMIs in the WTA library. The targeted library is compared using 2-fold read pairs per cell 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.

Single Cell 3' and 5' Gene Expression libraries have different thresholds used to assess fraction of reads on target (80% and 70%, respectively). Single Cell 5' Gene Expression libraries typically have a decreased Fraction of reads mapped confidently to the targeted transcriptome, due to slightly higher levels of antisense reads present in these libraries compared to Single Cell 3' Gene Expression libraries.

Due to the double stranded nature of the libraries, antisense reads are enriched as efficiently as sense reads. However, antisense reads are not mapped to the on target transcriptome which results in a slightly lower final metric.

Cell Calling Concordance:

Variability in cell calling concordance is normal and is influenced by several factors. In these experiments, >96% of samples tested with pre-designed panels detected the same number of cells as the parent library ± 25%. 75% of the targeted samples were within ± 10% of the cells detected in the parent sample. The targeted-compare pipeline presents this comparison for any parent-targeted library pair.

When using pre-designed panels, cell calling discrepancies are frequently due to populations of cells enriched in mitochondrial reads (possibly dead or dying cells) that are often filtered out in secondary analysis. Cell calling in WTA libraries may vary due to cellular RNA content, cell size, dead and dying cells, or high background RNA. For targeted libraries, Cell Ranger uses cell calling algorithms that account for their lower number of UMIs per cell. This difference, as well as differential gene expression in different cell types, can result in discrepancies in cell detection between a targeted library and its parent WTA library.

If the parent sample lacks a clear transition from cell to background in the barcode rank plot, cell definition may be unreliable in the parent sample, resulting in additional differences after target enrichment. The ovarian cancer sample in this dataset had this phenotype, resulting in poor cell calling in the targeted sample.

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)

© 2020 10x Genomics, Inc. (10x Genomics). All rights reserved. Duplication and/or reproduction of all or any portion of this document without the express written consent of 10x Genomics, is strictly forbidden. Nothing contained herein shall constitute any warranty, express or implied, as to the performance of any products described herein. Any and all warranties applicable to any products are set forth in the applicable terms and conditions of sale accompanying the purchase of such product. 10x Genomics provides no warranty and hereby disclaims any and all warranties as to the use of any third-party products or protocols described herein. The use of products described herein is subject to certain restrictions as set forth in the applicable terms and conditions of sale accompanying the purchase of such product. A non-exhaustive list of 10x Genomics' marks, many of which are registered in the United States and other countries can be viewed at: www.10xgenomics.com/trademarks. 10x Genomics may refer to the products or services offered by other companies by their brand name or company name solely for clarity, and does not claim any rights in those third-party marks or names. 10x Genomics products may be covered by one or more of the patents as indicated at: www.10xgenomics.com/patents. All products and services described herein are intended FOR RESEARCH USE ONLY and NOT FOR USE IN DIAGNOSTIC PROCEDURES.

The use of 10x Genomics products in practicing the methods set forth herein has not been validated by 10x Genomics, and such non-validated use is NOT COVERED BY 10X GENOMICS STANDARD WARRANTY, AND 10X GENOMICS HEREBY DISCLAIMS ANY AND ALL WARRANTIES FOR SUCH USE. Nothing in this document should be construed as altering, waiving or amending in any manner 10x Genomics terms and conditions of sale for the Chromium Controller or the Chromium Single Cell Controller, consumables or software, including without limitation such terms and conditions relating to certain use restrictions, limited license, warranty and limitation of liability, and nothing in this document shall be deemed to be Documentation, as that term is set forth in such terms and conditions of sale. Nothing in this document shall be construed as any representation by 10x Genomics that it currently or will at any time in the future offer or in any way support any application set forth herein.

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