Interpreting Cell Ranger Web Summary Files for Chromium Single Cell Immune Profiling

Introduction

The web summary file output (web_summary. html) by the cellranger multi pipeline is the initial point of reference for determining sample performance in the Chromium Single Cell Immune Profiling assay. This Technical Note presents an overview of web summary file interpretation, including the expected metrics and characteristic plots for V(D)J libraries generated using this assay.

For interpreting gene expression metrics, refer to the Technical Note: Interpreting Cell Ranger Web Summary Files for Single Cell Gene Expression Assay (Document CG000329).

Web Summary Organization

Representative web summary files and other Cell Ranger output files are available for download on the 10x Genomics Support website. The web summary is organized into three views (Figure 1). Each view contains important information for assessing the success of an experiment.

- The Cells View contains information about cells called for the sample.
- The Library View contains information about each library.
- The Experimental Design View contains information about the experimental setup for the dataset and also includes the input Multi Config CSV.

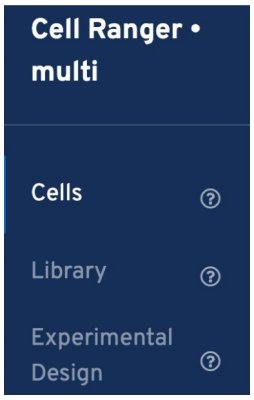


Figure 1. Three views in a cellranger multi file web summary file: Cells, Library, and Experimental Design.

The Cells View and Library View each contain a tab for each library type generated in the experiment.

The VDJ-B and/or VDJ-T tabs contain information about the B or T cell receptor V(D)J libraries, which are the focus of this Technical Note.

Library view — VDJ-T or VDJ-B tab

Metrics

Table 1. V(D)J metrics in the Library tab.

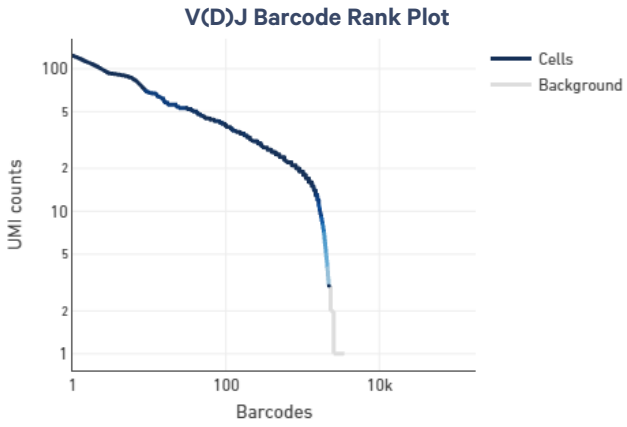
Metrics	Definition	Expected Value	Notes
Cell Statistics			
Estimated number of cells	The number of barcodes estimated to be associated with cells that express targeted V(D)J transcripts	Dependent on the number of cells loaded and the fraction of those cells that express V(D)J transcripts	Lower or higher than expected V(D)J cell calling may be due to inaccurate cell counting, poor T/B cell enrichment, poor sample quality, poor library quality, or low sequencing depth
Mean reads per cell	Number of input read pairs divided by the estimated number of cells	Sequencing output dependent	Recommended minimum sequencing depth is 5,000 reads per cell. Lower sequencing depth may lead to inaccurate V(D)J cell calling.
Enrichment			
Reads mapped to any V(D)J gene	Fraction of reads with valid barcodes that partially or wholly map to any germline V(D)J gene segment	Ideal > 50%. Acceptable >40%	Lower than expected values may be due to a low fraction of B or T cells in the sample, poor sample quality, poor library quality, or incorrect reference genome
Reads mapped to TRA/TRB or IGH/IGK/IGL	Fraction of reads with valid barcodes that map partially or wholly to a germline TRA/TRB or IGH/IGK/IGL gene segment	Variable; Fraction mapped to TRA is typically lower than fraction mapped to TRB, due to lower TRA expression levels	
Sequencing metrics			
Number of reads	Total number of read pairs sequenced during this run	Sequencing output dependent	Lower than expected values may indicate poor sequencing run (over-clustering, under-clustering, low % passing filter) or incorrect library pooling ratios
Number of short reads skipped	Total number of read pairs that were ignored by the pipeline because they do not satisfy the minimum length requirements (for example Read-1 less than 26 bases in Single Cell 5' assays)	Ideal 0	Higher than expected values may indicate that reads were sequenced or trimmed below the minimum length requirement
Q30 barcodes/ Q30 UMI/ Q30 RNA read	Fraction of cell barcode/UMI/ Read 2 bases with Q-score ≥30, excluding very low quality/ no-call (Q ≤2) bases from the denominator	Sequencing platform dependent (ideally >80%). Refer to Document CG000401 for expected sequencing metrics on various Illumina sequencing platforms.	Lower values may indicate sequencing issues such as sub-optimal loading concentration of the library

Metrics	Definition	Expected Value	Notes
Metrics per physical library			
Number of reads	Total number of read pairs that were assigned to this library	Dependent on sequencing output and number of sequencing runs	Lower than expected values may indicate poor sequencing run (over-clustering, under-clustering, low % passing filter) or incorrect library pooling ratios
Valid barcodes	Fraction of reads with barcodes that are present in the whitelist after barcode correction	Ideal >85%, acceptable >75%	Lower values may indicate issues with sequencing/library quality
Mean reads per cell	The total number of sequenced read pairs divided by the number of cell-associated barcodes	Sequencing output dependent	Recommended minimum sequencing depth is 5,000 reads per cell. Lower sequencing depth may lead to inaccurate V(D)J cell calling.
Mean used reads per cell	Mean number of read pairs used in assembly per cell-associated barcode. These reads must have a cell-associated barcode, map to a V(D)J gene, and have a UMI with sufficient read support	Sequencing output dependent	Lower fraction of used reads may indicate issues with sample quality, library quality or sequencing quality
Fraction reads in cells	Number of reads with cell-associated barcodes divided by the number of reads with valid barcodes	Ideal >50%	Lower values may indicate poor sample quality or failures during GEM generation

Interpreting the Web Summary File Plots

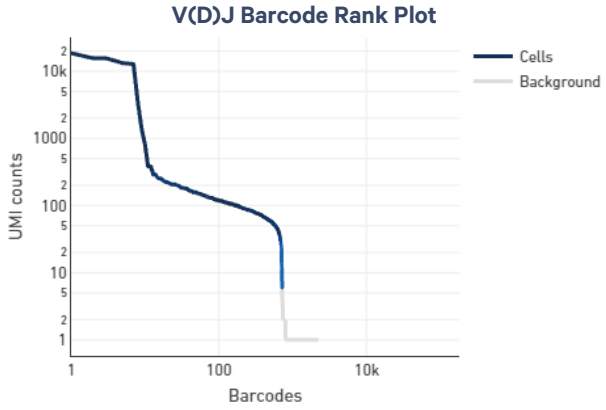
Table 2. Plots in the Library view - Antigen tab

Plots & Interpretation
V(D)J Barcode Rank Plot: The plot shows the count of filtered UMIs mapped to each barcode. A barcode must have a contig that aligns to a V segment to be identified as a targeted cell. In the denovo case, the only requirement is a contig's presence. There must also be at least three filtered UMIs with at least two read pairs each. It is possible that a barcode with at least as many filtered UMIs as another cell-associated barcode is not identified as a targeted cell. The color of the graph is based on the local density of cell-associated barcodes.
Examples
Typical sample A steep drop-off is indicative of good separation between the cell-associated barcodes and background

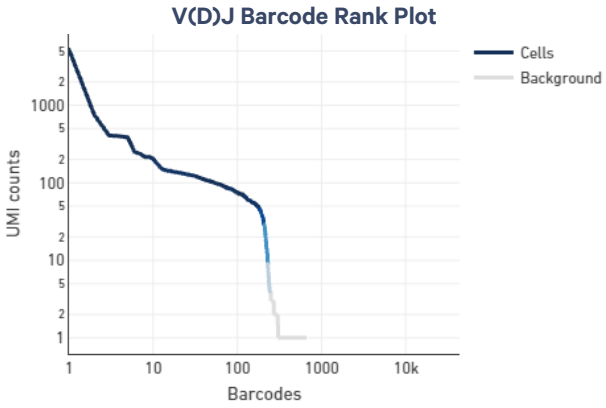


Plots & Interpretation

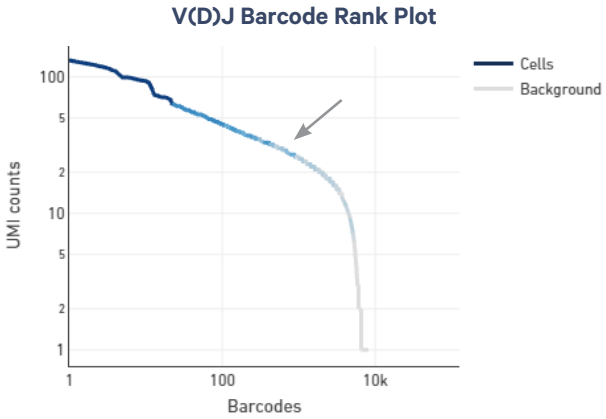
Typical sample
High-expressing plasma cells present in VDJ-B data can lead to a set of barcodes with high UMI counts



Compromised sample
Low number of cells called. May be due to a low fraction of B or T cells in the sample, inaccurate cell counting, poor sample quality, or poor library quality



Compromised sample
Undercalling of cells in the V(D)J library is evident by the trailing light blue line. May be due to poor sample quality or poor library quality.



Cell view – VDJ-T or VDJ-B tab

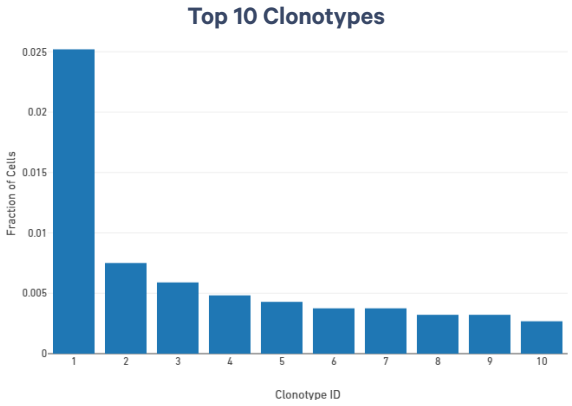
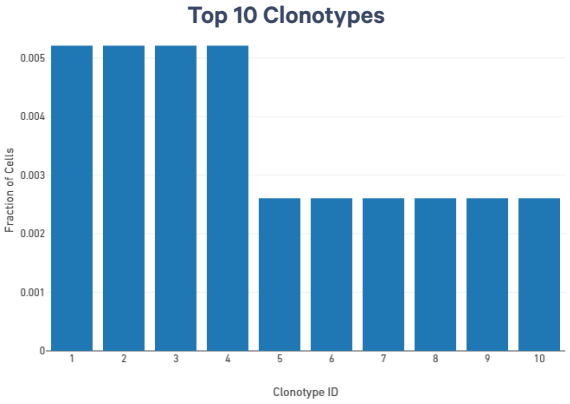
Metrics

Table 3. V(D)J metrics in the VDJ tab.

Metrics	Definition	Expected Value	Notes
B or T Cell Expression			
Estimated number of cells	The number of barcodes estimated to be associated with cells that express targeted V(D)J transcripts	Dependent on the number of cells loaded and the fraction of those cells that express V(D)J transcripts	Low or higher than expected V(D)J cell calling may be due to inaccurate cell counting, poor T/B cell enrichment, poor sample quality, poor library quality, or low sequencing depth
Number of cells with productive V-J spanning pair	Number of cell barcodes for which at least 1 full-length productive sequence was found for each chain of the (TRA, TRB) receptor pair		
Median TRA/TRB or IGH/IGK/IGL UMIs per Cell	Median number of UMIs assigned to a IGH contig per cell	Dependent on sample type and sequencing depth	Lower than expected values may be due to low sequencing depth, poor sample quality, or poor library quality
V(D)J Annotation			
Cells with productive V-J spanning pair	Fraction of cell-associated barcodes with at least one productive contig for each chain of the receptor pair. A productive contig satisfies the following conditions: the contig annotations span the 5' end of the V region to the 3' end of the J region of the chain, a start codon was found in the expected part of the V sequence, an in-frame CDR3 amino acid motif was found, and no stop codons were found in the aligned V-J region	Ideal >30%. Acceptable >20%	Lower than expected values may be to due a low fraction of B or T cells in the sample, poor sample quality, poor library quality, or low sequencing depth
Cells with productive V-J spanning (IGK, IGH) pair	Fraction of cell-associated barcodes with at least one productive contig for each chain of the (IGK, IGH) receptor pair, or (IGL, IGH) receptor pair	Sample type dependent	For B cell datasets, depends on the fraction of B cells expressing the kappa immunoglobulin light chain (IGK) or lambda immunoglobulin light chain (IGL)
Cells with productive V-J (IGL, IGH) spanning pair			
Cells with productive TRA/TRB or IGH/IGK/IGL contig	Fraction of cell-associated barcodes with at least one contig that spans the 5' end of the V region to the 3' end of the J region for TRA/TRB or IGH/IGK/IGL, has a start codon in the expected part of the V sequence, has an in-frame CDR3, and has no stop codons in the aligned V-J region		
Paired clonotype diversity	Effective diversity of the paired clonotypes, computed as the Inverse Simpson Index of the clonotype frequencies. A value of 1 indicates a minimally diverse sample - only one distinct clonotype was detected. A value equal to the estimated number of cells indicates a maximally diverse sample.	Sample type dependent	Lower than expected values may be to due a low fraction of B or T cells in the sample, poor sample quality, poor library quality, or low sequencing depth

Interpreting the Web Summary File Plots

Table 4. Plots in the Cell view – Antigen tab

Plots & Interpretation	
Top 10 Clonotypes: The histogram displays the fraction of cells (percentage of cells) occupied by the 10 most abundant clonotypes in this sample.	
Examples	
Typical sample In a sample that contains expanded clonotype(s), the expanded clonotype(s) will be present in a higher fraction of cells	
Typical sample In a sample that does not contain expanded clonotype(s), each clonotype will be observed in a similar fraction of cells	

References

- Interpreting Cell Ranger Web Summary Files for Single Cell Gene Expression Assays (Document CG000329)
- Sequencing Metrics & Base Composition of Single Cell 5' v2 Dual Index Libraries (Document CG000401)

Document Revision Summary

Document Number	CG000719
Title	Interpreting Cell Ranger Web Summary Files for Chromium Single Cell Immune Profiling
Revision	Rev A
Revision Date	December 2023

Specific Changes:

N/A

General Changes:

N/A

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