Interpreting Cell Ranger Multi Web Summary Files for Single Cell 3' Gene Expression

with Feature Barcode technology for Cell Multiplexing

Introduction

The web summary file (web_summary.html) output by the cellranger multi pipeline is the initial point of reference for determining sample performance in the Chromium Single Cell 3' Gene Expression with Feature Barcode technology for Cell Multiplexing assay. This Technical Note presents an overview of web summary file interpretation, including the expected metrics and characteristic plots for libraries generated using this assay.

Web Summary Organization

Representative web summary files for Chromium Single Cell 3' Gene Expression and Cell Multiplexing libraries and other Cell Ranger output files are available for download on the 10x Genomics Support website. For Single Cell 3' Cell Multiplexing experiments in which multiple cell or nuclei samples are pooled prior to loading on a 10x Genomics chip GEM (Gel Bead-in-emulsion) well, one web summary file will be generated per sample.

The web summary is organized into three views (Figure 1). Each view contains important information for assessing the success of an experiment.

Sample View:
Contains information about cell-associated barcodes assigned to the sample. For cell multiplexing experiments, only cell-associated barcodes assigned exactly one CMO are assigned to a sample. The Sample View contains various metrics to highlight aspects of the Gene Expression data for cells assigned to the particular sample. These metrics include: cell metrics, mapping metrics, t-SNE projections, sequencing saturation plot, and median genes per cell plot.

Library View - Gene Expression:
The Gene Expression tab within the Library View contains information about the physical Gene Expression library. This tab contains several metrics to highlight aspects of the Gene Expression data for all cells in the library. These metrics include: sequencing metrics, mapping metrics, barcode rank plot, sequencing saturation plot, and median genes per cell plot.

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

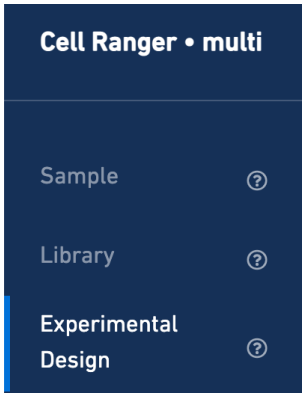


Figure 1. Three views in a cellranger multi web summary file: Sample, Library and Experimental Design.

Library View - Multiplexing:

The Multiplexing tab within the Library View contains information about the physical Cell Multiplexing library. This tab contains various metrics to highlight aspects of Cell Multiplexing data (Table 1-5).

The Experimental Design View:

Contains information about the experimental setup for the dataset (Figure 2) and also includes the input Multi Config CSV.

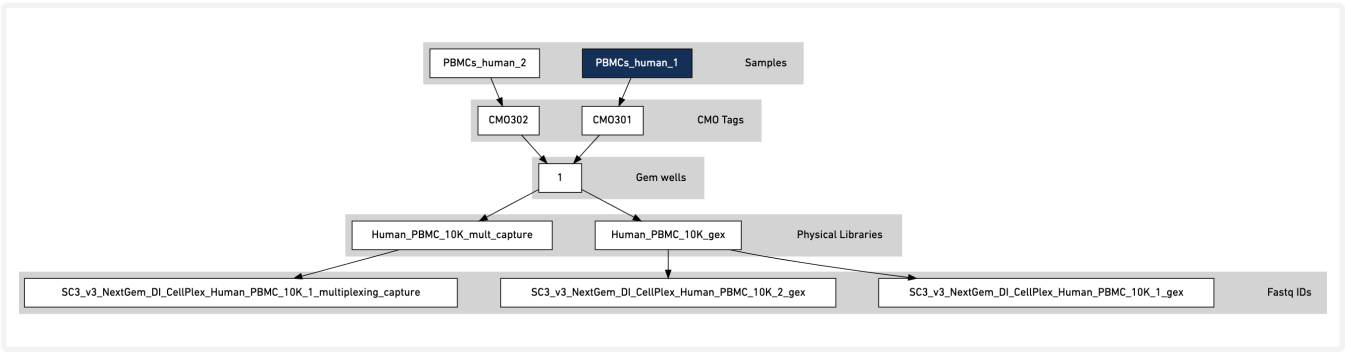


Figure 2. The Experimental Design view contains information about the number of samples, CMO tags, GEM wells, physical libraries, and FASTQ files that help verify the data were analyzed properly with respect to experimental design and also helps track complex experiments.

Library View - Multiplexing: Interpreting Metrics

Table 1. Multiplexing sample assignment metrics

Metrics	Definition	Expected Value	Notes
Multiplexing Metrics			
Estimated number of cell-associated barcodes	Number of cell-associated barcodes called as containing one or more cells.	Dependent on the number of cells loaded. Refer to the cell loading tables in the relevant User Guide.	Higher or lower than expected values may indicate inaccurate cell counting, cell lysis, or failures during GEM generation.
Samples assigned at least one cell	Number of samples to which at least one cell was assigned. Only cell-associated barcodes assigned exactly one CMO were assigned to a sample.	Expected to be equal to the number of cell/nuclei samples pooled prior to loading on the 10x Genomics chip GEM well.	Higher or lower than expected values may indicate inaccurate CMO tag assignment.* (See page 6)
Cells assigned to a sample	Number of cells assigned to a sample. Only cell-associated barcodes assigned exactly one CMO were assigned to a sample.	Expected to be less than the number of cell-associated barcodes, given that cell-associated barcodes classified as multiplets will not be assigned to a sample. Varies based on the expected multiplet rate at a given cell load, and based on the CMO tag number and sample pooling ratios, which dictate what fraction of multiplets will be detectable in Cell Ranger. Refer to Document CG000383 for further details on expected multiplet rates and multiplet detection.	
Singlet capture ratio	Ratio between the number of singlets (i.e. cell-associated barcodes assigned exactly one CMO) obtained and the number of singlets expected in this experiment according to Poisson statistics.	Values above 0.85 are acceptable, with values approaching 1.0 in optimal datasets.	
Cell-associated barcodes identified as multiplets	Fraction of cell-associated barcodes that were assigned more than one CMO and hence identified as multiplets.	Varies based on the expected multiplet rate at a given cell load, and based on the CMO tag number and sample pooling ratios, which dictate what fraction of multiplets will be detectable in Cell Ranger. Refer to Document CG000383 for further details on expected multiplet rates and multiplet detection.	
Median CMO UMIs per cell	Median number of CMO UMIs captured per cell-associated barcode assigned exactly one CMO.	Dependent on cell type and sequencing depth.	Lower than expected values could be due to shallow sequencing depths and/or poor sample/library quality.

Table 2. Multiplexing sample assignment metrics.

Metrics	Definition	Expected Value	Notes
Multiplexing Sample Assignments			
Estimated number of cell-associated barcodes	Refer to the indicated metrics in Table 1		
Mean reads per cell-associated barcode	The total number of sequenced read pairs divided by the number of cell-associated barcodes.	Sequencing output dependent	Sequencing depths below 5,000 read pairs per cell-associated barcode may indicate lower value leading to inaccurate CMO tag assignment.* (See page 6)
Samples assigned at least one cell	Refer to the indicated metrics in Table 1		
Cells assigned to a sample			
Cell-associated barcodes identified as multiplets			
Cell-associated barcodes not assigned any CMOs	Cell-associated barcodes that either (i) did not have enough CMO molecules above background or (ii) could not be confidently assigned to a singlet or multiplet state.	Ideal <10%	Higher values may indicate inaccurate CMO tag assignment.* (See page 6)
Median CMO UMIs per cell-associated barcode	Median number of CMO UMIs captured per cell-associated barcode.	Dependent on cell type and sequencing depth	Lower than expected values could be due to shallow sequencing depths and/or poor sample/library quality.

Table 3. Multiplexing sequencing metrics.

Metrics	Definition	Expected Value	Notes
Multiplexing 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 3' (v2 /v3) or 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 >65%). Refer to Document CG000374 for expected sequencing metrics on various Illumina sequencing platforms.	Lower values may indicate sequencing issues such as sub-optimal loading concentration of the library. Lower values for Read 2 may also arise if Cell Multiplexing libraries are sequenced alone (not recommended) instead of pooling with Gene Expression libraries.

Table 4. Metrics per physical 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 >75%	Lower values may indicate issues with sequencing/library quality.
Valid UMIs	Fraction of reads with valid UMIs; i.e. UMI sequences that do not contain Ns and that are not homopolymers.	Ideal >75%	
Fraction reads in cell-associated barcodes	The fraction of valid-barcode, valid-UMI, recognized multiplexing-barcode reads with cell-associated barcodes.	Ideal >40%	Lower values may indicate insufficient washing to remove unbound CMOs after CMO labeling, letting CMO labeled cells/nuclei sit for too long before chip loading, or may indicate poor sample quality, cell lysis, or failures during GEM generation.
Mean reads per cell-associated barcode	Refer to the indicated metrics in Table 1		
Fraction CMO reads	Fraction of reads that contain a recognized CMO sequence.	Ideal >95%	Lower values may indicate poor library/sequencing quality or errors in the CMO or sample definitions specified in the Config CSV.
Fraction CMO reads usable	Fraction of read pairs that contain a recognized CMO sequence, a valid UMI, and a cell-associated barcode.	Ideal >35%	Lower values may indicate insufficient washing to remove unbound CMOs after CMO labeling, letting CMO labeled cells/nuclei sit for too long before chip loading, poor library or sequencing quality, poor sample quality, cell lysis, or failures during GEM generation.
Fraction unrecognized CMO	Fraction of read pairs with an unrecognized CMO sequence.	Ideal <5%	Higher values may indicate poor library/sequencing quality or errors in the CMO or sample definitions specified in the Config CSV.
Fraction reads from multiplets	Amongst all sequenced read pairs, fraction with a cell-barcode identified as a multiplet.	Varies based on the expected multiplet rate at a given cell load, and based on the CMO tag number and sample pooling ratios, which dictate what fraction of multiplets will be detectable in Cell Ranger. Refer to Document CG000383 for further details on expected multiplet rates and multiplet detection.	Higher or lower values may indicate inaccurate CMO tag assignment.* (See page 6)

Table 5. Metrics per CMO.

Metrics	Definition	Expected Value	Notes
Metrics per CMO			
Fraction reads in cell-associated barcodes	Amongst all reads with a valid barcode, valid UMI, and this particular CMO sequence, fraction arising from cell-containing partitions.	Ideal >40%	Lower values may indicate insufficient washing to remove unbound CMOs after CMO labeling, letting CMO labeled cells/nuclei sit for too long before chip loading, or may indicate poor sample quality, cell lysis, or failures during GEM generation.
Cells assigned to CMO	Number of cells assigned this particular CMO (and only this CMO).	Varies based on sample pooling ratios.	Lower or higher than expected values may indicate inaccurate cell counting/pooling, or inaccurate CMO tag assignment.*
CMO signal-to-noise ratio	Computed as the difference between labeled and unlabeled mean CMO counts (log scale) divided by the variance.	Ideal >3.5	Lower values may indicate insufficient washing to remove unbound CMOs after CMO labeling, letting CMO labeled cells/nuclei sit for too long before chip loading, or may indicate poor sample quality, cell lysis, or failures during GEM generation.

* Inaccurate CMO tag assignment due to:

- Poor sample quality (eg. cell death, cell lysis, over-lysis in nuclei sample, clumping, high debris)
- Cells/nuclei maintained at room temperature (instead of 4°C) after CMO labeling
- Delays after CMO labeling (recommend loading cells/nuclei onto the 10x Genomics chip within 30 min of CMO labeling and pooling)
- Insufficient washing to remove unbound CMOs after labeling
- Inefficient CMO labeling (eg. due to low CMO volume for labeling)
- Failures during GEM generation (eg. wetting failure or clog)
- Poor library quality
- Poor sequence quality
- Insufficient sequencing depth in the Cell Multiplexing library (recommend at least 5,000 reads/cell)
- Inaccurate cell calling (eg. due to low sequencing depth of the Gene Expression library)
- Pooling cells at ratios beyond 5%-95%
- Pooling CMO-labeled and unlabeled samples
- Errors in the CMO or sample definitions specified in the Config CSV for Cell Ranger Multi

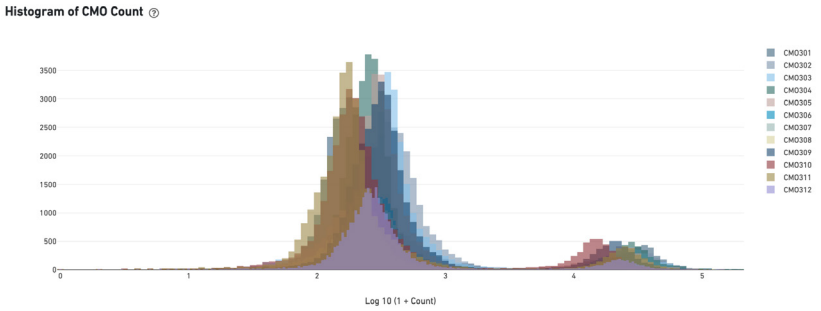
Contact support@10xgenomics.com for troubleshooting assistance.

Library View - Multiplexing: Interpreting Plots

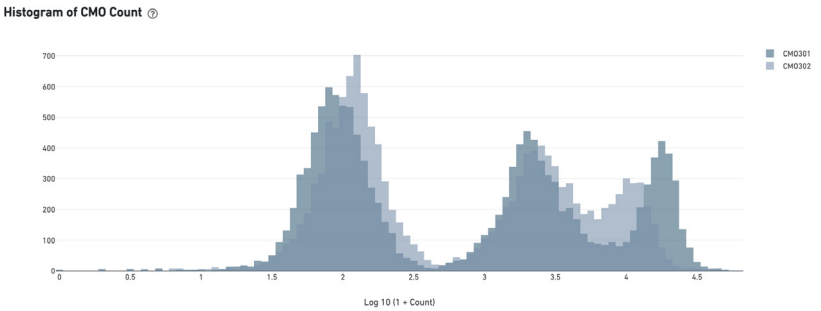
Plot & Interpretation
Histogram of CMO Count: Histogram of CMO UMI counts per cell for each CMO. Clicking the CMO tag legend on the right of the histogram can turn the overlap on/off.

Examples

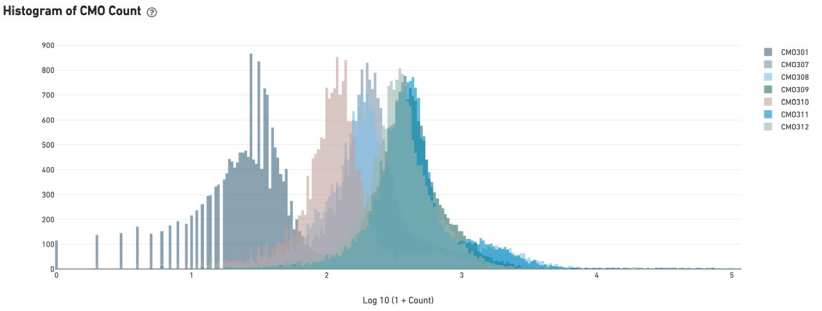
Typical sample (12 CMO tags):
Clear separation between background/noise (left peaks), and foreground/signal (right peaks).



Heterogeneous sample (2 CMO tags):
The foreground population may appear multimodal in heterogeneous samples such as PBMCs. Clear separation between background (left peak) and foreground (right peaks) is still observed.



Compromised sample (7 CMO tags):
A single peak is observed for each CMO, indicating that there is no separation between signal and noise. This may be caused by poor sample quality, failures during GEM generation, or high background from unbound CMOs due to insufficient washing or letting cells sit too long after CMO labeling.

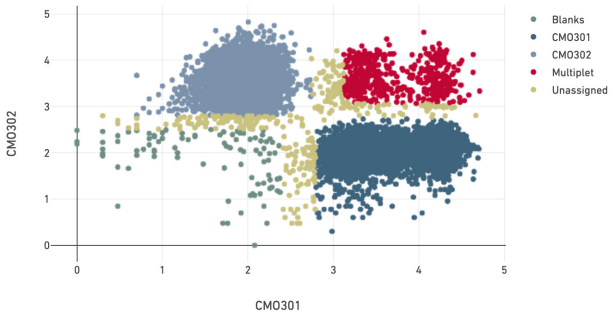


Plot & Interpretation

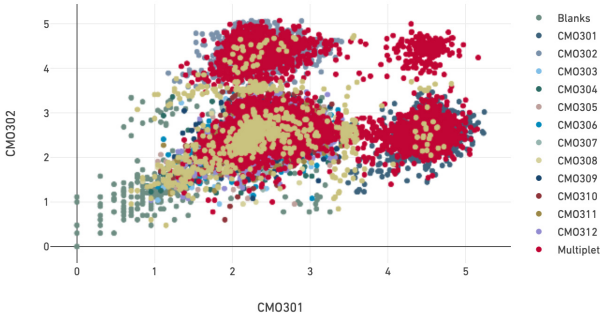
Biplots of CMO Count: Plot showing relationships between CMO UMI counts for cells. Each point is a cell and the X and Y axes are UMI counts for a given CMO. The CMOs on the axes can be changed with the selector. Cells which are not confidently assigned to any CMO are indicated.

Examples

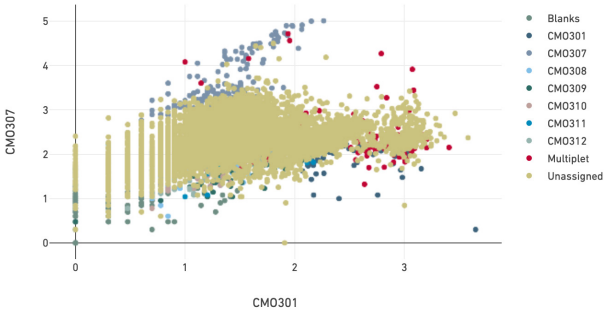
Typical sample (2 CMO tags):
When 2 CMO tags are used, clear separation into three main populations is expected. Cells assigned to a particular CMO have high UMI counts for that CMO and low UMI counts for the other CMO. Cells assigned as multiplets have high UMI counts for both CMOs. Cells that do not fall into any of these populations will not be assigned to either CMO ("blanks" or "unassigned").



Typical sample (12 CMO tags):
If more than two CMO tags are used, clear separation into four distinct populations is expected. In addition to the three populations discussed in the previous example, a fourth population of cells with low UMI counts for both selected CMOs is observed. These cells are expected to be assigned to other CMOs.



Compromised sample (7 CMO tags):
Lack of separation into distinct clusters. Many cells are unassigned. This may be caused by poor sample quality, failures during GEM generation, or high background from unbound CMOs due to insufficient washing or letting cells sit too long after CMO labeling.

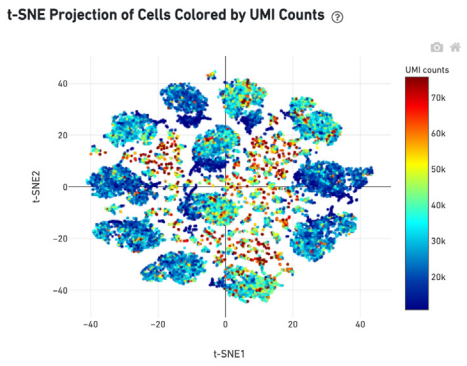


Plot & Interpretation

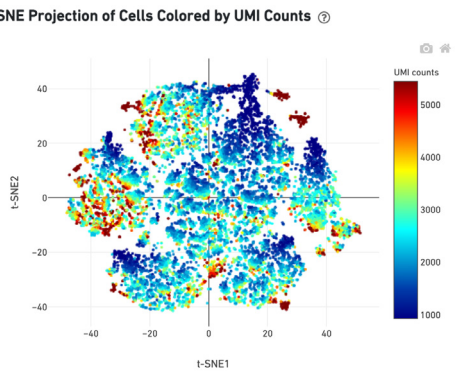
t-SNE Projection of Cells Colored by UMI Counts: Shown here are the total Multiplexing Capture UMI counts for each cell-barcode. The axes correspond to the 2-dimensional embedding produced by the t-SNE algorithm over the Multiplexing Capture features. In this space, pairs of cells that are close to each other have more similar Multiplexing Capture profiles than cells that are distant from each other. The display is limited to a random subset of cells.

Example

Typical sample (12 CMO tags): Clear separation into distinct clusters corresponding to the number of CMO tags. UMI counts within each cluster may vary in heterogeneous samples. Smaller clusters representing cell multiplets may be observed in-between the larger clusters.



Compromised sample (7 CMO tags): Lack of clear separation into distinct clusters. This may be caused by poor sample quality, failures during GEM generation, or high background from unbound CMOs due to insufficient washing or letting cells sit too long after CMO labeling.

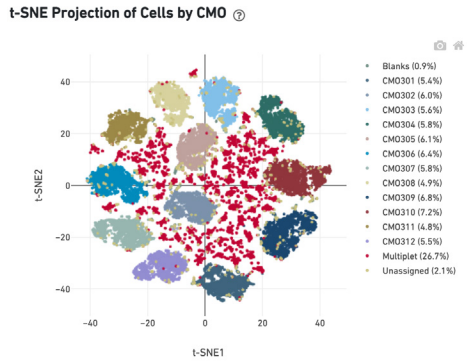


Plot & Interpretation

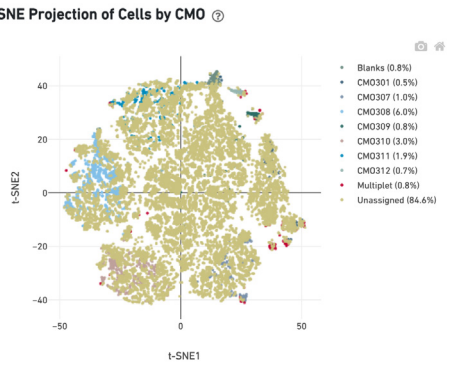
t-SNE Projection of Cells by CMO: Shown here are the CMO tag assignments for each cell-barcode. The axes correspond to the 2-dimensional embedding produced by the t-SNE algorithm over multiplexing features. In this space, pairs of cells that are close to each other have more similar Multiplexing Capture profiles than cells that are distant from each other. The display is limited to a random subset of cells.

Example

Typical sample (12 CMO tags): Clear separation into clusters corresponding to the number of CMO tags. Cells assigned as multiplets appear in-between the larger clusters of cells assigned to each CMO tag.



Compromised sample (7 CMO tags): Lack of clear separation into distinct clusters. Many cells are unassigned. This may be caused by poor sample quality, failures during GEM generation, or high background from unbound CMOs due to insufficient washing or letting cells sit too long after CMO labeling.

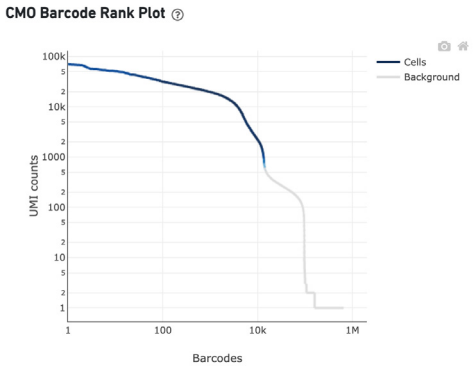


Plot & Interpretation

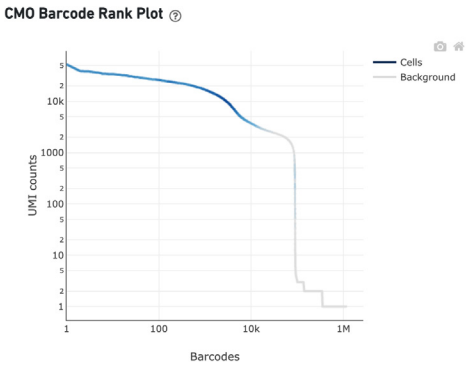
CMO Barcode Rank Plot: The plot shows the count of filtered UMIs mapped to each barcode. As barcodes are not determined to be cell-associated strictly based on their UMI count, but instead are determined by their expression profiles, some regions of the graph contain both cell-associated and background-associated barcodes. The color of the graph in these regions is based on the local density of barcodes that are cell-associated.

Examples

Typical sample: A steep drop-off is indicative of good separation between barcodes with high CMO UMI counts (cell-containing GEMs) and barcodes with low CMO UMI counts (empty GEMs).



Compromised sample: Absence of a steep cliff separating cells from background. This may be caused by poor sample quality, failures during GEM generation, or high background from unbound CMOs due to insufficient washing or letting cells sit too long after CMO labeling.



References

Technical Notes:

1. Interpreting Cell Ranger Web Summary Files for Single Cell Gene Expression Assay (Document CG000329)
2. Chromium Next GEM Single Cell 3' v3.1: Cell Multiplexing (Document CG000383)
3. Sequencing Metrics & Base Composition of Single Cell 3' v3.1 Dual Index Libraries (Document CG000374)

Demonstrated Protocols:

4. Cell Multiplexing Oligo Labeling for Single Cell RNA Sequencing Protocols (Document CG000391)
5. High Throughput Sample Preparation for Single Cell RNA Sequencing (Document CG0000426)

contd.

User Guides:

6. Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for Cell Multiplexing User Guide (Document CG000388)
7. Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for CRISPR Screening and Cell Multiplexing User Guide (Document CG000389)
8. Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for Cell Surface Protein and Cell Multiplexing User Guide (Document CG000390)
9. Chromium Next GEM Single Cell 3' HT Reagent Kits v3.1 User Guide with Feature Barcode technology for Cell Multiplexing (Document CG000419)
10. Chromium Next GEM Single Cell 3' HT Reagent Kits v3.1 User Guide with Feature Barcode technology for Cell Surface Protein and Cell Multiplexing (Document CG000420)
11. Chromium Next GEM Single Cell 3' HT Reagent Kits v3.1 User Guide with Feature Barcode technology for CRISPR Screening and Cell Multiplexing (Document CG000421)

Document Revision Summary

Document Number	CG000475
Title	Interpreting Cell Ranger Multi Web Summary Files for Single Cell 3' Gene Expression with Feature Barcode technology for Cell Multiplexing
Revision	Rev A
Revision Date	November 2021

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