Interpreting Cell Ranger multi Web Summary Files for Fixed RNA Profiling

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 Fixed RNA Profiling (Single Cell Gene Expression Flex) assay. This Technical Note presents an overview of web summary file interpretation, including the expected metrics and characteristic plots for Gene Expression libraries generated using this Fixed RNA Profiling (FRP) assay. Metrics associated with product extensions (i.e., Feature Barcode technology) are not included in this document.

Web Summary Organization

Representative web summary files for Fixed RNA Profiling libraries and other Cell Ranger output files are available for download on the 10x Genomics Support website. For Fixed RNA Profiling experiments in which multiple cell or nuclei samples are multiplexed and loaded on a 10x Genomics 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.

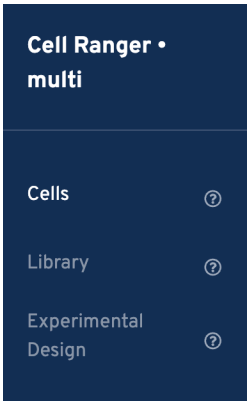


Figure 1. Three views in a cellranger multi web summary file: Cells, Library and Experimental Design.

Cells View

The Cells view contains information about cell-associated barcodes detected in the sample. For multiplexing experiments, only reads with a valid 10x GEM and Probe Barcode are assigned to a sample. The Cells view contains various metrics and plots to highlight aspects of Gene Expression data for reads and cells associated with the particular sample or Probe Barcode. These include: cell metrics, mapping metrics, t-SNE projections, and median genes per cell plot. For multiplexing experiments, the Cells view will also include cell multiplexing and genomic DNA metrics and a barcode rank plot for the particular sample.

Library View

The Library view contains information about the library. In a multiplex experiment, the library contains all cells detected in different samples or Probe Barcodes. This tab contains several metrics and plots to highlight aspects of Gene Expression data for all reads and cells in the library. These include: sequencing metrics, mapping metrics, barcode rank plot, sequencing saturation plot, and median genes per cell plot. Poor sequencing and/or mapping metrics at the library level usually indicate compromised read/library quality (please see Table 1). Poor library metrics impact the interpretation of metrics at the sample level (Cells view).

Experimental Design View

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

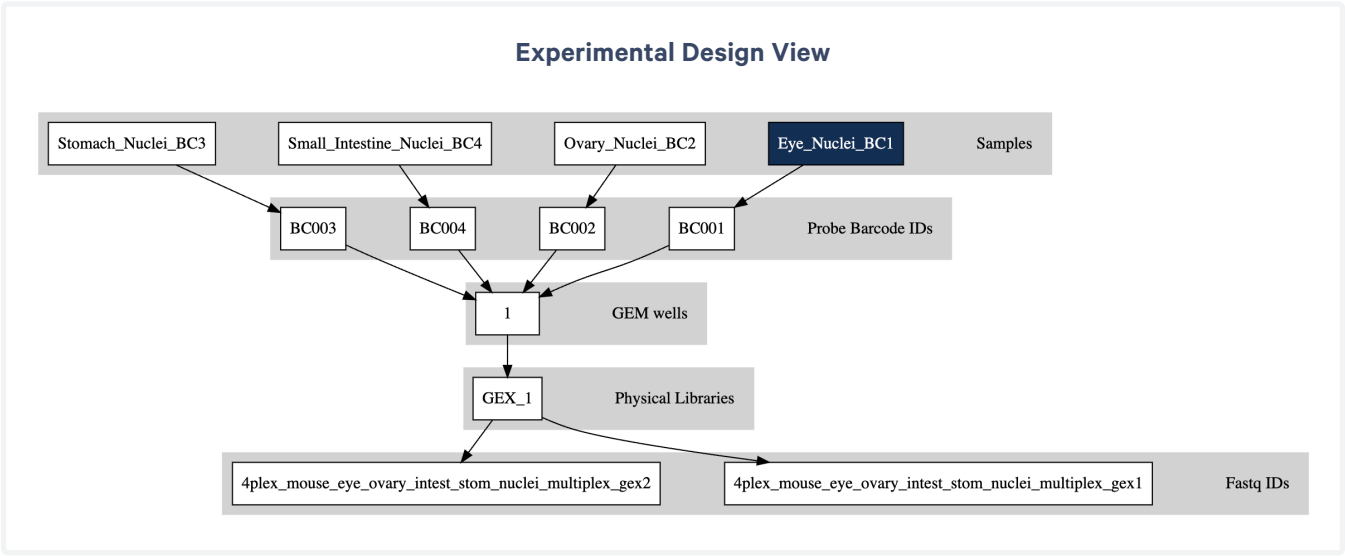


Figure 2. The Experimental Design view contains information about the number of samples, Probe Barcode IDs, GEM wells, Physical Libraries, and FASTQ IDs that help verify the data were analyzed based on the experimental design. This view is helpful for tracking complex experiments.

Library view: Interpreting metrics for the whole library

Table 1. Metrics under the Library view, Gene Expression tab

Metrics	Definition	Expected Value	Notes
Cell Statistics			
Estimated number of cells	Number of barcodes identified as cells by the cell calling algorithm Barcodes removed by Protein Aggregate Detection and Filtering or High Occupancy GEM Filtering are not counted	500-10,000 per Probe Barcode. Up to 128,000 cells per 16 plex library	• Higher or lower than expected values may indicate inaccurate cell count, high amount of debris, poor sample quality with extremely low RNA content, or failures during GEM generation. • Cell recovery with FFPE samples can be variable or lower compared to other sample types. FFPE samples often exhibit low complexity, characterized by low RNA content or RNA expression. Therefore, when working with cells isolated from FFPE blocks, aim for a recovery rate of over 50%, based on the 'Estimated Number of Cells' as a fraction of the targeted number of cells. • Notably, Cell Ranger's cell calling algorithm discards barcodes with <500 UMIs, potentially leading to the exclusion of a few 'true' FFPE cells. This may be addressed by inspecting the data to determine the appropriate cell count and using force-cells in the multi config CSV file for Cell Ranger.
Mean reads per cell	Total number of sequenced read pairs divided by the number of cell associated barcodes	User defined; 10,000 reads/cell minimum recommended per library	• The necessary sequencing depth per cell depends on the cell type (high or low RNA content) and the desired analysis.
Fraction of initial cell barcodes passing high occupancy GEM filtering	Fraction of cell associated barcodes from initial cell calls that are retained after high occupancy GEM filtering. Cell calling is performed for each sample All barcodes associated with GEMs that have significantly higher Probe Barcodes per GEM (than expected from optimal chip loading) are removed. This mitigates higher than expected barcode collision rates	Ideally >95% based on 10x internal testing A web summary alert appears if this metric is <90%	• This metric is only calculated for multiplexing experiments. • Values may be lower than expected if there is a partial clog, wetting failure, cell clumping, or if there were significant deviations from the recommended chip loading protocol.
Sequencing Metrics			
Number of reads	Total number of read pairs sequenced during this run	Sequencing output dependent	• Lower values may indicate poor sequencing run (over clustering, under clustering, low % passing filter).
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 (e.g., Read 1 less than 26 bases in Fixed RNA Profiling)	Ideally 0	• Higher values may indicate that reads were sequenced or trimmed below the minimum length requirement.

Table 1 contd. Metrics under the Library view, Gene Expression tab

Metrics	Definition	Expected Value	Notes
Q30 barcodes/ Q30 GEM barcodes	Fraction of cell barcode or GEM barcode bases with Q-score >=30, excluding very low quality/no-call (Q <=2) bases from the denominator. If the data are from multi-sample Fixed RNA Profiling, the cell barcode is a combination of the GEM barcode and the Probe Barcode	Ideally >90% based on 10x internal datasets A web summary alert appears if this metric is <55%	• Low Q30 base percentages could indicate sequencing issues such as suboptimal loading concentration. • Lower values for Q30 barcodes or Probe Barcodes in multiplex Fixed RNA Profiling data may also arise when libraries were sequenced using a lower than recommended concentration of PhiX. • Q30 Probe Barcode metric is only relevant for multiplex Fixed RNA Profiling.
Q30 probe barcodes	Fraction of Probe Barcode bases with Q-score >=30, excluding very low quality/ no-call (Q <=2) bases from the denominator	Ideally >80%	
Q30 UMI	Fraction of UMI bases with Q-score >=30, excluding very low quality/no-call (Q <=2) bases from the denominator	Ideally >75%	• Low Q30 base percentages could indicate sequencing issues such as suboptimal loading concentration.
Q30 RNA read	Fraction of Read 2 probe bases with Q-score >=30, excluding very low quality/no-call (Q <=2) bases from the denominator	Ideally >90% based on 10x internal datasets A web summary alert appears if this metric is <65%	
Mapping Metrics (All reads in library)			
Number of reads in the library	The total number of read pairs in the library	Dependent on sequencing output and number of sequencing runs	• Lower value may indicate poor sequencing run (over clustering, under clustering, low % passing filter).
Reads half-mapped to probe set	Fraction of reads that mapped to unpaired ligation products	Ideally <10% based on 10x internal tests A web summary alert appears if this metric is >20%	• Higher value will result in a low fraction of reads confidently mapped to the probe set (see below). This may be due to low RNA content in the sample, poor washing after probe hybridization, deviation from recommended protocol during probe hybridization, or suboptimal sample preparation.
Reads split-mapped to probe set	Fraction of reads that mapped to mispaired ligation products		
Reads mapped to probe set	Fraction of reads that mapped to the probe set	Ideally >80% based on 10x internal tests A web summary alert appears when this metric is <50%	• Lower values can indicate low total expression, use of the wrong probe set, suboptimal sample preparation, high expression of genes removed by filtering, or the use of input FASTQs from products other than Fixed RNA Profiling. For more information on probe filtering please visit the support website.
Reads confidently mapped to probe set	Fraction of reads that mapped uniquely to a probe in the probe set	Ideally >50%	
Reads confidently mapped to filtered probe set	Fraction of reads from probes that map to a unique gene. These reads are considered for UMI counting		
Metrics Per Physical Library			

Table 1 contd. Metrics under the Library view, Gene Expression tab

Metrics	Definition	Expected Value	Notes
Number of reads	Total number of read pairs that were assigned to this library in demultiplexing	Dependent on sequencing output and number of sequencing runs	• Lower numbers may indicate poor sequencing run (over clustering, under clustering, low % passing filter).
Valid barcodes	Fraction of reads with barcodes that match the whitelist after barcode correction For Multiplexed Fixed RNA Profiling samples, a cell barcode is the combination of the GEM barcode and the Probe Barcode, and both parts must be valid following correction	Ideally >90% based on 10x internal datasets A web summary alert appears if this value is <75%	• Lower values may indicate sequencing issues (such as low Q30 scores for Read 1 or Read 2) or compromised library quality.
Valid GEM barcodes	Fraction of reads with GEM barcodes that are present in the GEM barcode whitelist after barcode correction		
Valid probe barcodes	Fraction of reads with Probe Barcodes that are present in the Probe Barcode whitelist after barcode correction		
Valid UMIs	Fraction of reads with valid UMIs; i.e. UMI sequences that do not contain Ns and that are not homopolymers		
Sequencing saturation	The fraction of reads originating from an already-observed UMI. This is the fraction of confidently mapped, valid cell-barcode, valid UMI reads that had a non-unique combination of cell-barcode, UMI, and gene	Dependent on sequencing depth and sample complexity	• Dependent on library complexity, sequencing depth, and analysis goals. • To simply cluster cells into populations, it may not be necessary to detect every unique transcript (UMI count) in each cell. In this case, a lower sequencing saturation is sufficient. • Higher sequencing saturation may be required to detect and recover lowly expressed transcripts.
Confidently mapped reads in cells	The fraction of valid-barcode, valid-UMI, confidently-mapped-to-probe-set reads with cell-associated barcodes.	Ideally >=70%	• Low value could be caused by high levels of ambient RNA partitioned into all GEMs. • Since the algorithm discards barcodes with low UMI counts, this metric may be inaccurately low in cases where a substantial population of cells exhibits low RNA content. Inspect the data to determine the appropriate cell count and use force-cells in Cell Ranger's multi config CSV.
Mean reads per cell	The total number of sequenced read pairs divided by the number of cell associated barcodes.	User defined; 10,000 reads/cell minimum recommended per library	• The necessary sequencing depth per cell depends on the cell type (high or low RNA content) and the desired analysis.

Table 1 contd. Metrics under the Library view, Gene Expression tab

Metrics	Definition	Expected Value	Notes
Metrics Per Probe Barcode (applicable to multiplex Fixed RNA Profiling only)			
UMIs per probe barcode	Number and fraction of UMIs for this Probe Barcode amongst all UMIs in the raw feature barcode matrix.	Dependent upon how different samples were pooled to generate the library	• The values can be impacted by variations in sample complexity and distribution of reads within the pool.
Cells per probe barcode	Number and fraction of cells per Probe Barcode amongst all cells detected in this GEM well	Dependent on how many cells are targeted for each sample/Probe Barcode. Normally <10,000	• Higher or lower values may indicate inaccurate cell count, large amounts of debris, poor sample quality with extremely low RNA content, or failures during GEM generation.

Cells view: Interpreting metrics for each sample

Table 2. Metrics under the Cells view, Gene Expression tab

Metrics	Definition	Expected Value	Notes
Cell Metrics			
Cells	Number of cells called from this sample	**[singleplex FRP]** 500-10,000 **[multiplex FRP, one Probe Barcode per sample]** 500-10,000 (4 plex); 500-8,000 (>4 plex) **[multiplex FRP, multiple Probe Barcodes used for one sample]** 1000-128,000	• Higher or lower than expected values may indicate inaccurate cell count, high amount of debris, poor sample quality with extremely low RNA content, or failures during GEM generation. • Cell recovery with FFPE samples can be variable or lower compared to other sample types. FFPE samples often exhibit low complexity, characterized by low RNA content or RNA expression. Therefore, when working with cells isolated from FFPE blocks, aim for a recovery rate of over 50%, based on the 'Estimated Number of Cells' as a fraction of the targeted number of cells. • Notably, Cell Ranger's cell calling algorithm discards barcodes with <500 UMIs, potentially leading to the exclusion of a few 'true' FFPE cells. This may be addressed by inspecting the data to determine the appropriate cell count and using force-cells in the multi config CSV file for Cell Ranger.

Table 2 contd. Metrics under the Cells view, Gene Expression tab

Metrics	Definition	Expected Value	Notes
Median reads per cell	Median number of read pairs associated with the cells called from this sample	Variable, dependent on sample complexity, cell types, and library sequencing depth	• This metric is calculated based only on reads associated with "cells" in the particular sample. Do not compare this metric to the sequencing recommendation (10,000 mean reads per cell). The recommended sequencing depth is calculated at the library level (mean reads per cell), not the per-sample level. • The necessary number of reads per cell for each sample depends on the cell type (high or low RNA content) and the desired analysis.
Median genes per cell	The median number of genes detected per cell called from this sample. Detection is defined as the presence of at least 1 UMI count		• Low median genes per cell may be biological (low transcriptional diversity) or may indicate low sequencing depth or library complexity.
Total genes detected	The number of genes with at least one UMI count in the cells in this sample		• A lower than expected number could be the result of shallow sequencing depth and/or compromised sample/library quality.
Median UMI counts per cell	Median number of UMIs obtained from the cells called from this sample		
Confidently mapped reads in cells	The fraction of valid-barcode, valid-UMI, confidently-mapped-to-probe-set reads with cell-associated barcodes	Ideal >70%	• Low value could be caused by high levels of ambient RNA partitioned into all GEMs. • Since the algorithm discards barcodes with low UMI counts, this metric may be inaccurately low in cases where a substantial population of cells exhibits low RNA content. Inspect the data to determine the appropriate cell count and use force-cells in Cell Ranger's multi config CSV.
Cell Multiplexing			
Cells detected in this sample	Number and fraction of cells detected in this sample among all cells detected in this GEM well	Dependent on the number of cells loaded and targeted for each sample.	
Cells detected in other samples	Number and fraction of cells detected in other samples among all cells detected in this GEM well		

Table 2 contd. Metrics under the Cells view, Gene Expression tab

Metrics	Definition	Expected Value	Notes
Mapping Metrics (Amongst reads from cells assigned to sample)			
Number of reads from cells called from this sample	The total number of reads from cells called from this sample	Refer to metrics in Mapping Metrics (All reads in Library)	
Reads half-mapped to probe set	Fraction of reads that mapped to unpaired ligation products		
Reads split-mapped to probe set	Fraction of reads that mapped to mispaired ligation products		
Reads mapped to probe set	Fraction of reads that mapped to the probe set		
Reads confidently mapped to probe set	Fraction of reads that mapped uniquely to a probe in the probe set		
Reads confidently mapped to filtered probe set	Fraction of reads from probes that map to a unique gene. These reads are considered for UMI counting		
UMIs from Genomic DNA *For **singleplex** Fixed RNA Profiling experiments, metrics located under the Library view.* *For **multiplex** Fixed RNA Profiling experiments, metrics located under the Cells view for each sample.*			
Estimated UMIs from genomic DNA	The estimated fraction of filtered UMIs derived from genomic DNA based on the discordance between probes targeting exon-junction-spanning regions and non-exon-junction-spanning regions	Usually <1%	• Values higher than expected may arise due to one of two reasons. First, samples are relatively low in complexity (low RNA content). Second, samples are de-crosslinked during an antigen retrieval step (>70°C).
Estimated UMIs from genomic DNA per unspliced probe	The estimated number of UMIs derived from genomic DNA for each probe targeting non-exon-junction-spanning regions A probe not spanning an exon junction with a total UMI count below this value has a high likelihood of its UMIs being derived primarily from hybridization to genomic DNA rather than the mRNA	Variable	• For details, visit the software support website.

Interpreting Web Summary File Plots

The summary file also contains barcode rank plots and t-SNE projection plots (Table 3).

Table 3. Plots in the web summary file.

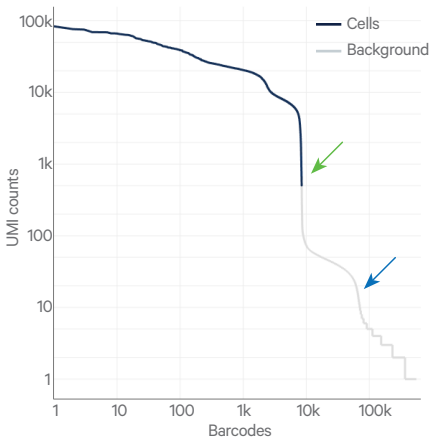
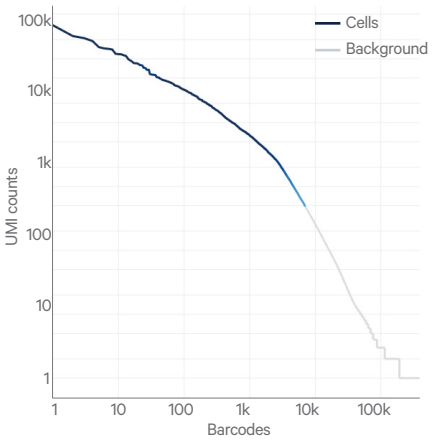
Plot and Interpretation		
Barcode rank plot: All cell barcodes detected during sequencing are plotted in decreasing order of the number of UMIs associated with that particular barcode. In singleplex Fixed RNA Profiling experiments, the cell barcode is the 10x GEM barcode, whereas in multiplex Fixed RNA Profiling experiments, the cell barcode is a combination of the 10x GEM barcode and the probe barcode. Cell Ranger uses the number of UMIs detected for each barcode to determine the barcodes that are likely to be associated with cells. Barcodes associated with cells tend to have a greater number of transcripts (and thus UMIs) associated with them than non-cell barcodes. The barcode rank plot for a singleplex experiment can be found in Library view. For a multiplex Fixed RNA Profiling experiment, examine the barcode rank plot under the Cells view for each sample to assess the quality of individual samples in the pool.		
Examples		
Typical sample	• A steep drop-off is indicative of good separation between the cell associated barcodes and the barcodes associated with empty GEMs • The barcode rank plot should have a distinctive "cliff and knee" shape. The blue-to-gray transition (green arrow) is the cliff and the solid gray is the knee (blue arrow) • The knee usually appears around the position of 50k-100k barcodes. For multiplex experiments where one sample is hybridized across multiple Probe Barcodes (sample sub-pooling), the position of the knee will be 50k-100k times the number of Probe Barcodes used • The plot is from PBMC rep 1 in this dataset	
Challenging or compromised sample	• Reviewing the barcode rank plot in combination with other metrics and t-SNE plots can help determine whether data are usable for downstream analyses. If the cluster structure in t-SNE plots and mapping metrics look as expected, data may be used for further analysis • Barcode rank plots with a round curve and lack of steep cliff may be observed in some FFPE samples, or samples with low quality or loss of single cell behavior. This may be due to a wetting failure, large amounts of debris, or poor sample quality with low RNA content • It is not unusual for single cell FFPE samples to lack the distinctive cliff and knee shape in the barcode rank plots, as FFPE samples are particularly challenging for RNA recovery. The plot is from Glioblastoma FFPE Tissues Dissociated Manually in this dataset	

Table 3 contd. Plots in the web summary file.

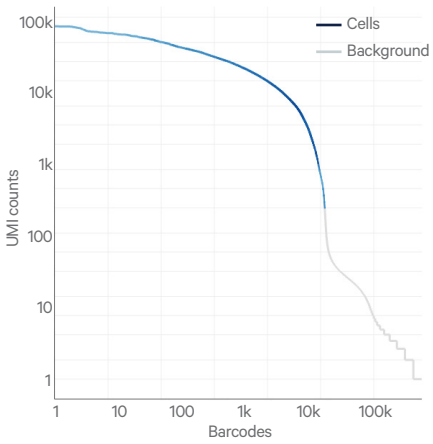
Plot and Interpretation		
Compromised sample	• Gradient/light blue region starting at the very top of the barcode rank plot. This indicates that some cell associated barcodes were filtered due to high occupancy GEM filtering, possibly due to nuclei clumping, wetting failures or issues during GEM generation, or due to cell settling from sitting in the chip for too long (>15 min) • The severity of high occupancy GEMs should be reviewed together with the metric "Fraction of initial cell barcodes passing high occupancy GEM filtering". The remaining data may still be usable. • The example plot is from a multiplex Fixed RNA Profiling experiment with nuclei samples; only a small fraction of initial cell calls could pass the high occupancy GEM filtering step (76.17%) • If the Gene Expression data is paired and analyzed with an Antibody Capture library, another cause for this phenotype in the barcode rank plot is antibody aggregates. • More information about quality assessment of Antibody Capture data is available on the software support website	

Table 3 contd. Plots in the web summary file.

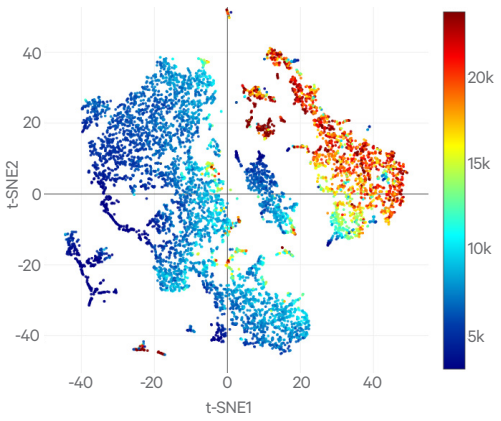
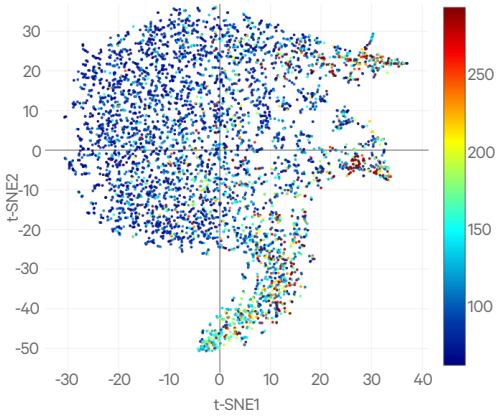
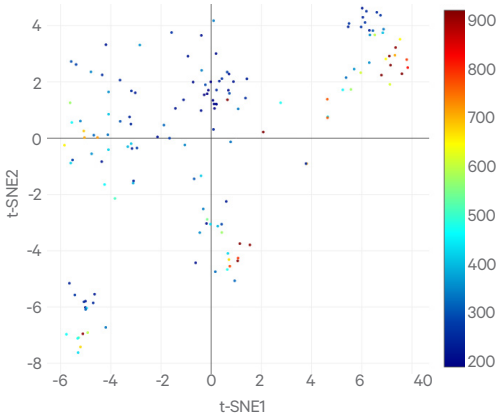
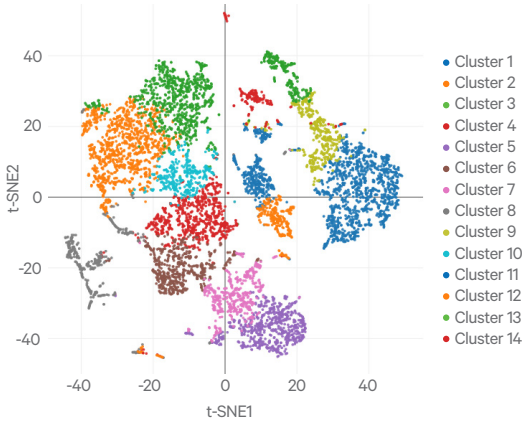
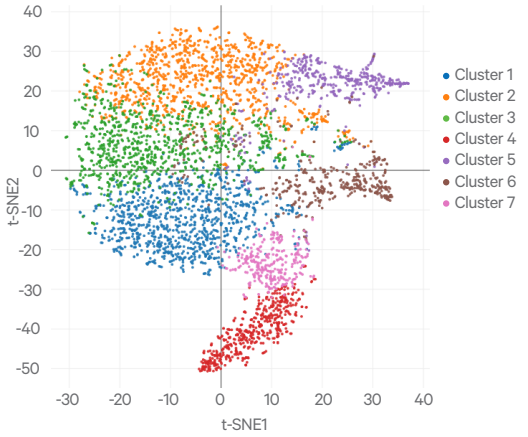
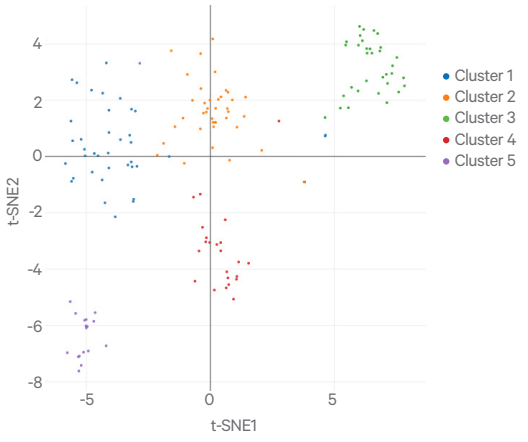
Plot and Interpretation		
t-SNE projection of cells colored by UMI counts: This cell clustering plot is located under the Cells view and shows the cell associated barcodes in a 2-D t-SNE projection. Each cell is color-coded according to how many UMIs it contains. Cells with greater UMI counts likely have higher RNA content than cells with fewer UMI counts.		
Examples		
Typical sample	• Structured clusters with clear separation between high UMI and low UMI containing barcodes. Note that the UMI counts within each cluster may vary in heterogeneous samples • The plot is from PBMC rep 1 in this dataset	
Compromised sample	• Lack of cluster structure, one large cluster, or no separation (for a sample with expected heterogeneous cell populations) may indicate sample quality issues or loss of single cell behavior • The plot is from a human thymus sample prepared from a 9-year old FFPE block with low complexity and compromised quality, with very low UMI counts overall (mostly <500). Note that if the sample is expected to be homogeneous, lack of cluster structure may be expected and may not be a sign of poor quality	
Compromised sample	• Very low cell numbers may result in scattered cell t-SNE plots with little cluster structure • The plot is from a human heart sample prepared from a 12-year old FFPE block with low complexity and compromised quality, with very low UMI counts overall (mostly <1000)	

Table 3 contd. Plots in the web summary file.

Plot and Interpretation		
t-SNE projection of cells colored by clustering: This cell clustering (colored by cluster) plot shows the cell associated barcodes in a 2-D tSNE projection, with colors showing an automated graph clustering analysis which groups cells with similar gene expression profiles.		
Examples		
Typical sample	• Structured clusters with good separation (for a sample with expected heterogeneous cell populations). Note that the number of clusters and distance separating clusters may vary in heterogeneous samples • The plot is from PBMC rep 1 in this dataset	
Compromised sample	• Lack of cluster structure, one large cluster, or no separation (for a sample with expected heterogeneous cell populations) may indicate sample quality issue or loss of single cell behavior • The plot is from a human thymus sample prepared from a 9-year old FFPE block with low complexity and compromised quality	
Compromised sample	• Very low cell numbers may result in scattered cell t-SNE plots with little cluster structure • The plot is from a human heart sample prepared from a 12-year old FFPE block with low complexity and compromised quality	

Document Revision Summary

Document Number	CG000729
Title	Interpreting Cell Ranger multi Web Summary Files for Fixed RNA Profiling
Revision	Rev A
Revision Date	January 2024

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