TECHNICAL NOTE

Interpreting Cell Ranger ATAC Web Summary Files for Single Cell ATAC Assay

Introduction

The web summary file in the output folder of the Cell Ranger ATAC analysis software is the initial point of reference for determining sample performance in the Single Cell ATAC assay. This Technical Note presents an overview of web summary file interpretation, including the expected metrics and characteristic plots for libraries generated using the Single Cell ATAC workflow.

Interpreting Web Summary File Metrics

Representative summary files for Chromium Single Cell ATAC libraries and other Cell Ranger ATAC output files are available for download on the 10x Genomics Support website. The top of the web summary file displays key metrics (Figure 1). Green text indicates that the key metrics are in the expected range while red/yellow text indicates errors/warnings. Descriptions of the metrics can also be found by clicking the icon ⓘ next to the section header. The summary tab reports various metrics including sample, sequencing, cells, cell clustering, insert sizes, targeting, and library complexity.

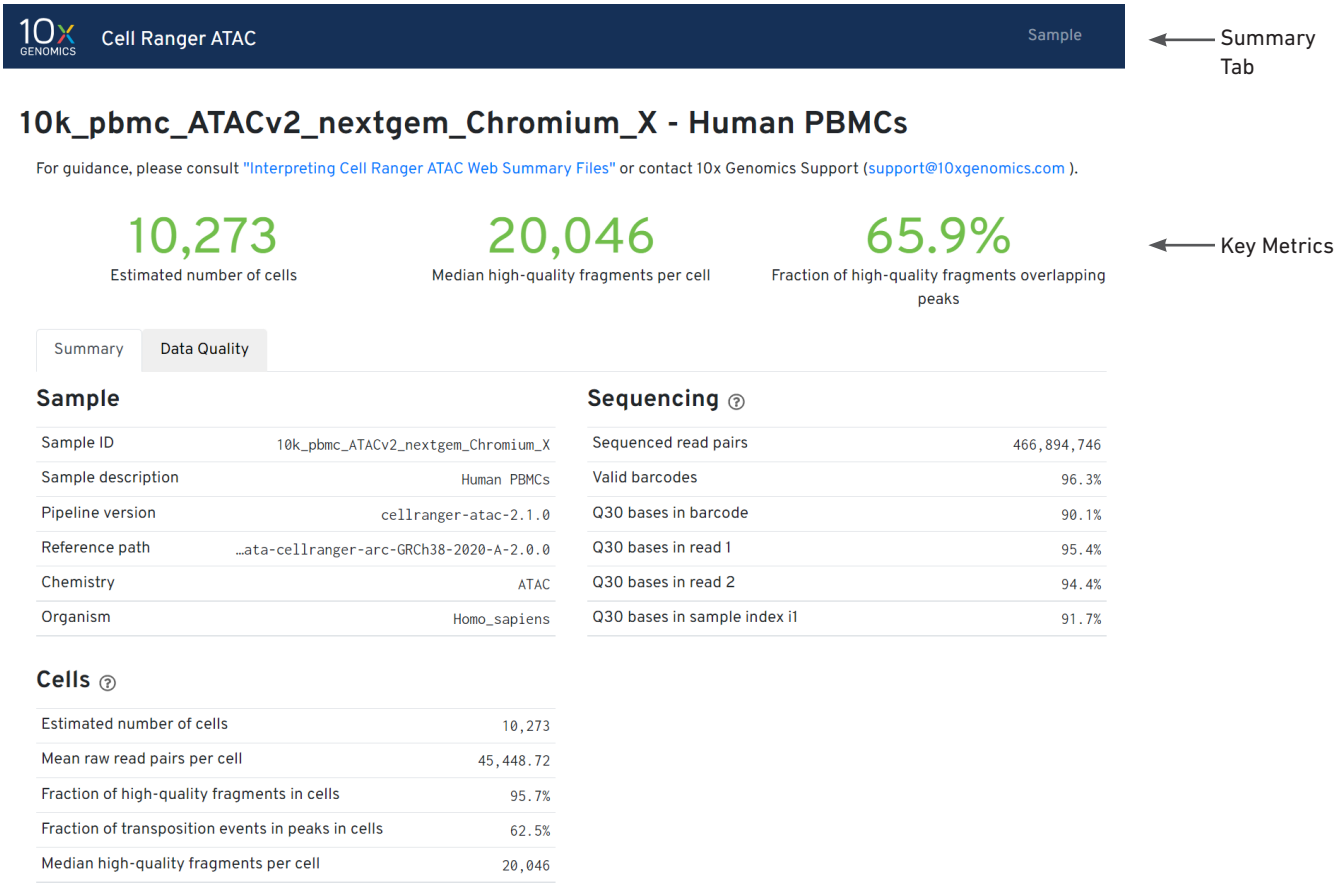


Figure 1. A representative web summary file (top section) for a PBMC sample targeting 10,000 nuclei.

Table 1. Metrics in the ATAC web summary file.

Metrics	Definition	Expected Value	Notes
Sequencing Metrics			
Sequenced read pairs	Total number of sequenced read pairs assigned to the sample	User defined	Suggested sequencing depth of 25,000 read pairs per cell.
Valid barcodes	Fraction of read pairs with barcodes that match the whitelist with error correction	>75%	Low valid barcodes may indicate sequencing related problem or issues with library preparation.
Q30 bases in barcode	Fraction of barcode read (i2) bases with Q-score ≥30	Sequencing platform dependent (ideally >65%)	Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration of the library.
Q30 bases in read 1	Fraction of read 1 bases with Q-score ≥30	Sequencing platform dependent (ideally >65%)	Expected to be higher than Q30 Bases in barcode (i5 read) or Sample Index (i7 read) and is sequencing platform dependent. Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration of the library.
Q30 bases in read 2	Fraction of read 2 bases with Q-score ≥30	Sequencing platform dependent (ideally >65%)	Expected to be higher than Q30 bases in barcode (i5 read) or Sample Index (i7 read) and is sequencing platform dependent. Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration of the library.
Q30 bases in sample index i1	Fraction of sample index read (i1) bases with Q-score ≥30	Sequencing platform dependent (ideally >90%)	Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration of the library.
Cell Metrics			
Estimated number of cells	The total number of barcodes identified as cells	500-10,000	±20% expected value is acceptable. Higher or lower values outside of this range may indicate inaccurate nuclei count, nuclei lysis or failures during GEM generation.
Mean raw read pairs per cell	Total number of read pairs divided by the number of cell barcodes	Dependent on sequencing depth	-
Fraction of high-quality fragments in cells	Fraction of high-quality fragments with a valid barcode that are associated with cell-containing partitions. High-quality fragments are defined as read pairs with a valid barcode that map to the nuclear genome with mapping quality (map Q) ≥30, are not chimeric, and not duplicate.	>40%	-
Fraction of transposition events in peaks in cells	Fraction of transposition events that are associated with cell-containing partitions and fall within peaks Transposition events are located at both ends of all high-quality fragments. This metric measures the percentage of such events that overlap with peaks.	>15%	-
Median high-quality fragments per cell	The median number of high-quality fragments per cell barcode	Dependent on cell type & sequencing depth	-

Table 1 contd. Metrics in the ATAC web summary file.

Metrics	Definition	Expected Value	Notes
Library Complexity Metric			
Percent duplicates	Fraction of high-quality read pairs that are deemed to be PCR duplicates. This is the fraction of high-quality fragments with a valid barcode that align to the same genomic position as another read pair in the library.	≥30%	A high-quality read-pair is one with mapping quality (mapQ) ≥30, that is not chimeric and maps to nuclear contigs. This metric is a measure of sequencing saturation and is a function of library complexity and sequencing depth.
Mapping Metrics			
Confidently mapped read pairs	Fraction of sequenced read pairs with mapping quality (mapQ) ≥30	>80%	-
Unmapped read pairs	Fraction of sequenced read pairs that have a valid barcode but could not be mapped to the genome	<5%	-
Non-nuclear read pairs	Fraction of sequenced read pairs that have a valid barcode and map to non-nuclear genome contigs, including mitochondria, with mapping quality (mapQ) ≥30	<20%	-
Fragments in nucleosome-free regions	Fraction of fragments passing all filters with a size smaller than 124 basepairs	>40%	Expected to be the highest proportion as compared to mononucleosome and dinucleosome fragment.
Fragments flanking a single nucleosome	Fraction of fragments passing all filters with a size between 124 and 296 basepair	Dependent on sample type	An increased proportion of mononucleosome fragments may indicate dead/dying cells or granulocyte contamination.
Targeting Metrics			
Number of peaks	Total number of peaks on primary contigs either detected by the pipeline or input by the user	>45,000	-
Fraction of genome in peaks	Fraction of bases in primary contigs that are defined as peaks	>2% and <20%	-
TSS enrichment score	Maximum value of the transcription-start-site (TSS) profile. The TSS profile is the summed accessibility signal (defined as number of cut sites per base) in a window of 2,000 bases around all the annotated TSSs, normalized by the minimum signal in the window.	>5	-
Fraction of high-quality fragments overlapping TSS	Fraction of high-quality fragments in cell barcodes that overlap transcription start sites (TSS)	>15%	-
Fraction of high-quality fragments overlapping peaks	Fraction of high-quality fragments in cell barcodes that overlap called peaks	>15%	Low percentage indicates that fragments are not coming from called peaks but rather from random regions of the genome. Causes include dead cells, or very low sequencing depth.

Interpreting the Web Summary File Plots

The summary file also contains multiple plots. Table 2 describes the six major plots that help in assessing assay performance.

Table 2. Plots in the ATAC web summary file.

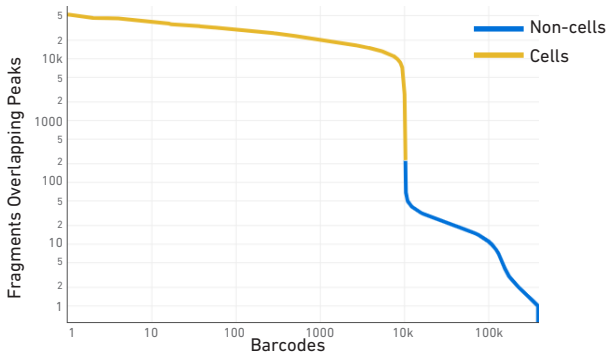
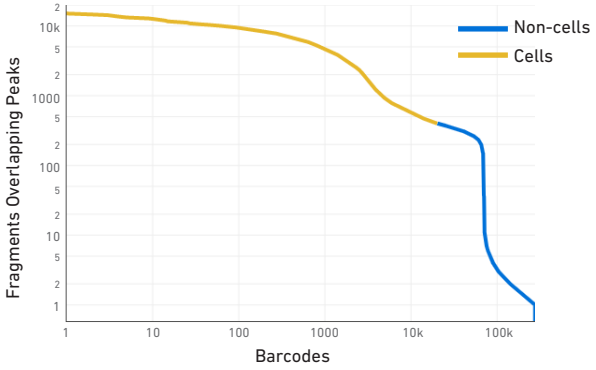
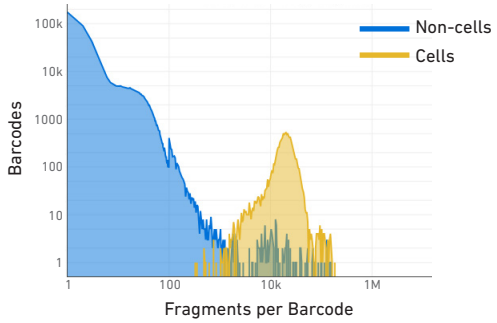
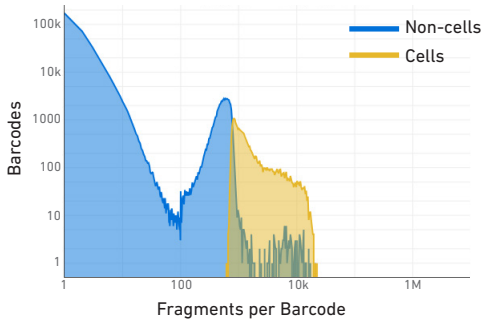
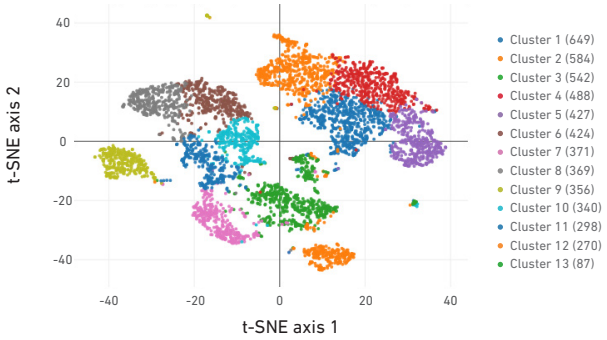
Plot & Interpretation	
Barcode Rank Plot: The Barcode Rank (or knee plot) for fragments overlapping peaks marks the barcodes that were inferred to be associated with cells.	
Example	
Ideal Sample: A steep drop-off is indicative of good separation between the cell-associated barcodes and the barcodes associated with empty GEMs.	**Compromised Sample:** Round curve and lack of steep drop-off may indicate low sample quality or loss of single-cell behavior.
	
Plot & Interpretation	
Fragment Distribution Plot: The distribution of the number of fragments per barcode for the non-cell and cell groups is displayed in the Fragment Distribution plot.	
Example	
Ideal Sample: A good separation between cell and non-cell groups indicate proper distinction between cells and non-cells.	**Compromised Sample:** Large overlap between cells and non-cells may indicate issues with sample quality.
	

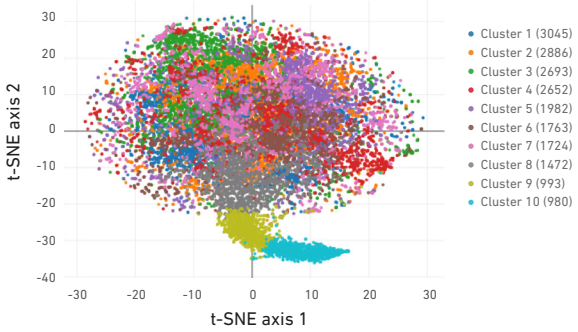
Table 2 contd. Plots in the ATAC web summary file.

Plot & Interpretation
Cell Clustering Scatter Plot: The Cell Clustering (colored by cluster) plot shows the cell-associated barcodes in a 2-D t-SNE projection, with colors showing an automated graph clustering analysis which groups together cells with similar peak profiles.
Example

Ideal Sample: Structured clusters with good separation (for a sample with expected heterogeneous cell populations).

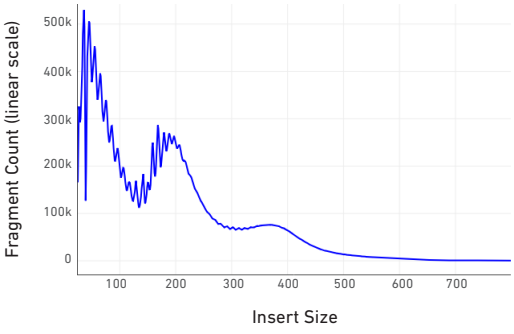


Compromised Sample: Lack of cluster structure, one large cluster or no separation (for a sample with expected heterogeneous cell populations) may indicate sample quality issue or loss of single cell behavior.



Plot & Interpretation
Insert Size Distribution Plot: Insert size distribution of transposase accessible fragments sequenced is displayed in the Insert Size Distribution plot.
Example

Ideal Sample: A periodicity of ~150 bp corresponds to the number of nucleosomes the transposase accessible fragments span (nucleosome free, mononucleosome, and dinucleosome fragments). Sawtooth pattern in fragments with insert size <200 bp corresponds to the helical pitch of DNA (~10.5 bp).



Compromised Sample: Absence of periodicity may indicate loss of chromatin structure due to low sample quality.

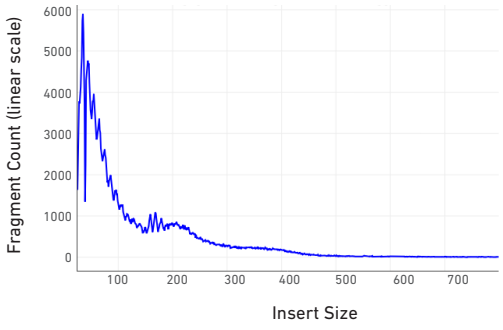
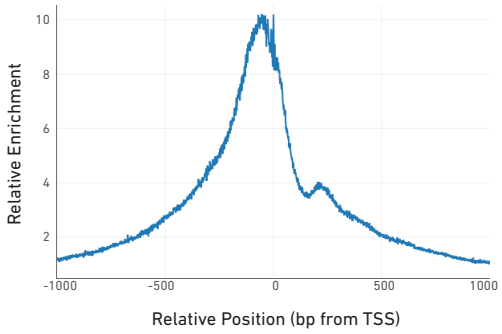


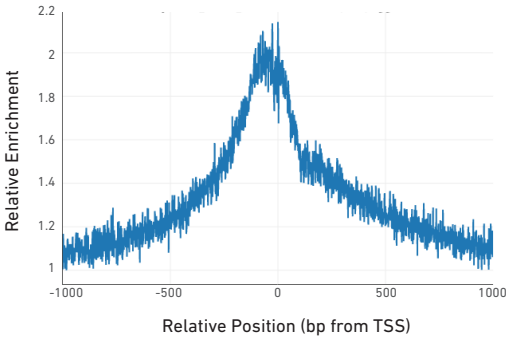
Table 2 contd. Plots in the ATAC web summary file.

Plot & Interpretation
Transition Start Site (TSS) Plot: The Transcription Start Site (TSS) profile, which is computed as the number of cut sites per base, of all the barcodes irrespective of cell versus non-cell assignment in a window of 2,000 bases around the full set of annotated TSSs is displayed in the Transcription Start Site plot. The y-axis scale is normalized by the minimum signal in the window.
Example

Ideal Sample: Large enrichment around TSS, as these regions are known to have a high degree of chromatin accessibility compared to the flanking regions.

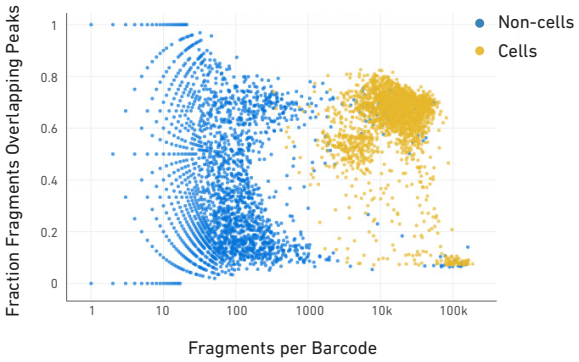


Compromised Sample: Low enrichment (<5%) around TSS sites may indicate improper lysis or loss of chromatin structure.

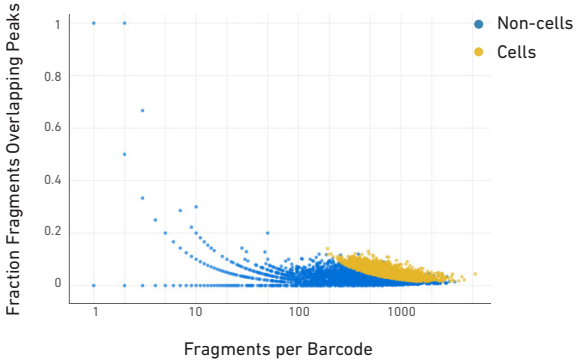


Plot & Interpretation
Single Cell Targeting Plot: A scatterplot displaying the number of fragments per barcode and the percent of fragments overlapping peaks.
Example

Ideal Sample: Cell-associated barcodes are expected to have a large number of fragments per barcode and a high percentage of fragments overlapping peaks (upper right corner). Non-cell associated barcodes are expected to have a small number of fragments per barcode and a low percentage of fragments overlapping peaks (lower left corner). An ideal sample should show good separation of cells and non-cells at the opposite ends.



Compromised Sample: Cell-associated barcodes have a low fraction of the barcode fragments overlapping peaks. Cell-associated and non-cell associated barcodes tightly concentrated in the same location may indicate issues with cell calling or sample preparation.



References

- Sequencing Metrics & Base Composition of Chromium Single Cell ATAC Libraries (Document CG000181)
- Chromium Single Cell ATAC Reagent Kits User Guide (Document CG000168)
- Chromium Next GEM Single Cell ATAC Reagent Kits v1.1 User Guide (Document CG000209)
- Chromium Next GEM Single Cell ATAC Reagent Kits v2 User Guide (Document CG000496)

Contact:
support@10xgenomics.com
10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94566 USA