


## 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.

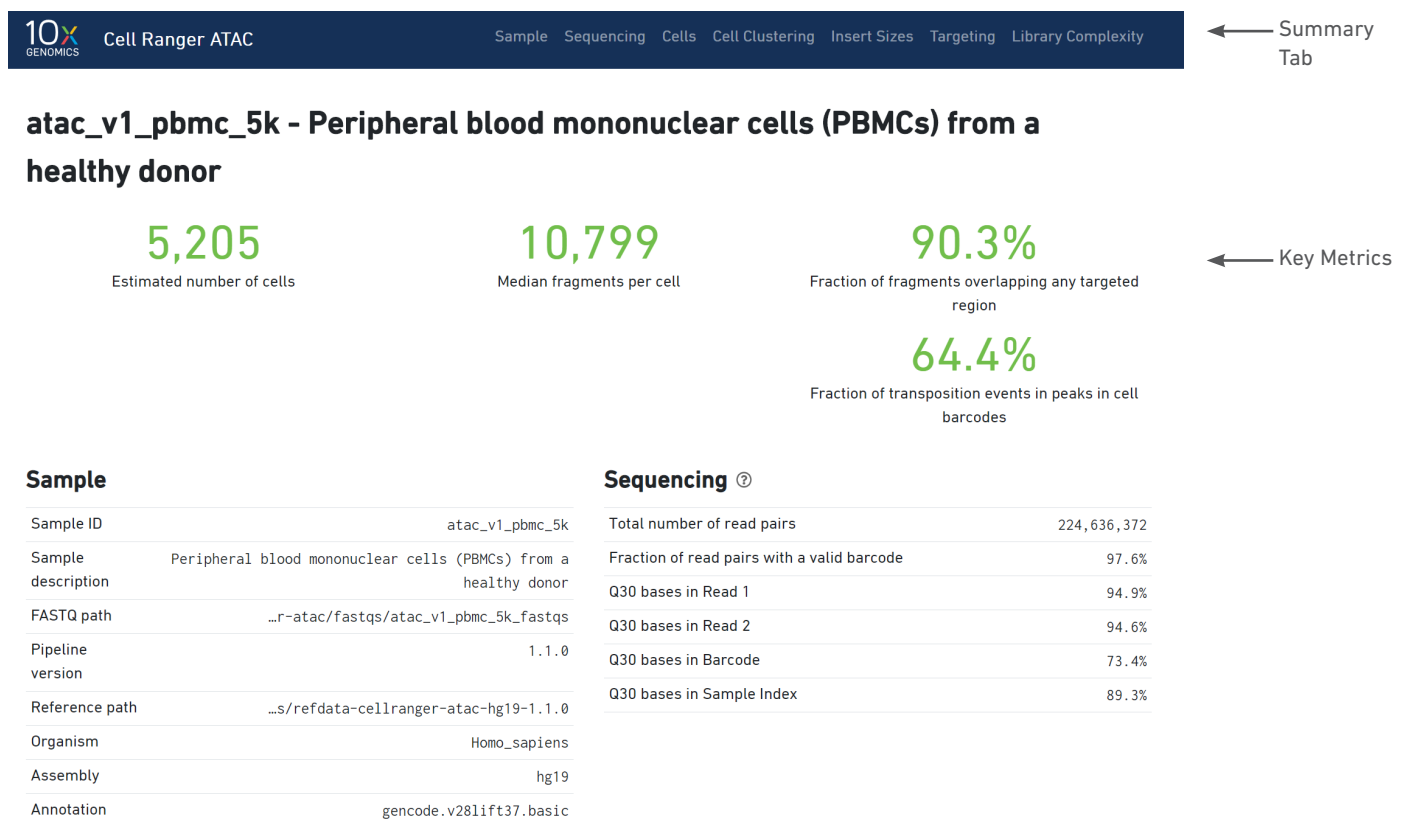


Figure 1. A representative web summary file (top section) for a PBMCs sample targeting 5,000 nuclei.

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

Metrics	Definition	Expected Value	Notes
<b>Sequencing Metrics</b>			
Total number of read pairs	Total number of read pairs that were assigned to this library after demultiplexing	User defined	Suggested sequencing depth of 25,000 read pairs per cell.
Fraction of read pairs with a valid barcode	Fraction of reads with barcodes that match the whitelist after barcode correction	>75%	Low valid barcodes may indicate sequencing related problem or issues with library preparation.
Q30 bases in R1	Fraction of insert read bases (R1N) with Q-score $\geq 30$	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 R2	Fraction of insert read bases (R2N) with Q-score $\geq 30$	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 barcode	Fraction of barcode read bases (i5) with Q-score $\geq 30$	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 Sample Index	Fraction of sample index read bases (i7) with Q-score $\geq 30$	Sequencing platform dependent (ideally >90%)	Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration of the library.
<b>Cell Metrics</b>			
Estimated Number of Cells	Total number of barcodes associated with cell-containing partitions estimated by ATAC Cell Ranger	500-10,000	$\pm 20\%$ expected value is acceptable. Higher or lower values outside of this range may indicate inaccurate nuclei count, nuclei lysis or failures during GEM generation.
Lower threshold on the number of fragments overlapping peaks per barcode to annotate barcode as cell	Minimum number of fragments overlapping peaks needed to be called as a cell	Dependent on cell type and sequencing depth	Number is determined through dynamic modeling for each sample (refer to " <a href="#">Cell Ranger ATAC Algorithms Overview</a> ").
Median fragments per cell barcode	Median number of transposase accessible fragments being assigned to cell barcodes	>500	Dependent on cell type and sequencing depth, will vary from sample to sample.
Median fragments per non-cell barcode	Median number of transposase accessible fragments being assigned to non-cell barcodes	Dependent on cell type and sequencing depth	Dependent on cell type and sequencing depth. A high number of fragments in non-cell barcodes may indicate high background and low sample quality.
<b>Insert Size Metrics</b>			
Fragments in nucleosome-free regions	Fraction of fragments passing all filters with a size smaller than 147 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 147 and 294 basepair	Dependent on sample type	An increased proportion of mononucleosome fragments may indicate dead/dying cells or granulocyte contamination.

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

Metrics	Definition	Expected Value	Notes
<b>Targeting Metrics</b>			
Fraction of fragments overlapping any targeted region	Fragments overlapping any TSS, DNase HS, enhancer, or promoter regions	>55%	Dependent on quality of genome annotation.
Fraction of total read pairs mapped confidently to genome (>30 mapq)	Fraction of reads pairs with a 99.9% chance of mapping correctly	>80%	Reference quality and sequencing configuration (shorter than recommended cycles on Read 2) can impact mapping. May indicate the use of the wrong reference (wrong species, an improper build of custom reference, etc).
Fraction of total read pairs in mitochondria and in cell barcodes	Percent total read pairs that map to mitochondrial reads and are assigned as cell-associated barcodes	<40%	Using an alternative lysis buffer other than that indicated in the <a href="#">Demonstrated Protocols for isolating nuclei for ATAC Sequencing</a> may increase this metric.
Fraction of transposition events in peaks in cell barcodes	Measures the abundance of transposition events in called peaks from cell-associated barcodes	>25%	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.
Enrichment score of transcription start sites	The summed 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. This metric reports the maximum value in the profile	>5%	Low score may indicate sample quality issues such as loss of chromatin structure or improper cell lysis.

## 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	
<b>Barcode Rank Plot:</b> The Barcode Rank (or knee plot) for fragments overlapping peaks marks the barcodes that were inferred to be associated with cells.	
<b>Example</b>	
<p><b>Ideal Sample:</b> A steep drop-off is indicative of good separation between the cell-associated barcodes and the barcodes associated with empty GEMs.</p>	<p><b>Compromised Sample:</b> Round curve and lack of steep drop-off may indicate low sample quality or loss of single-cell behavior.</p>



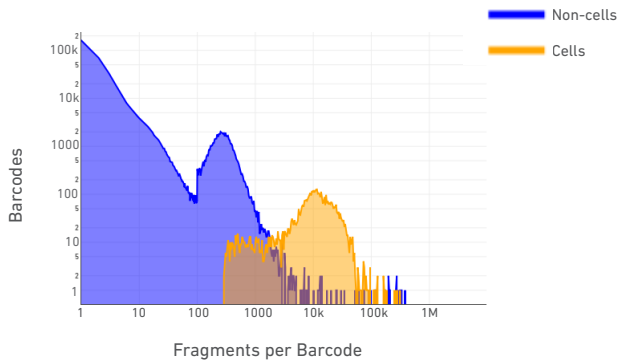
Table 2 contd. Plots in the ATAC web summary file.

Plot & Interpretation

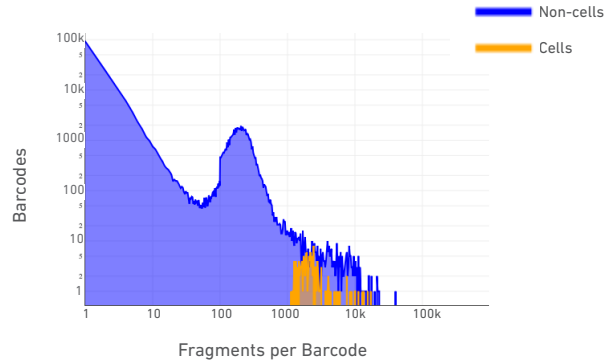
**Fragment Distribution Plot:** The distribution of the number of fragments per barcode for the non-cell and cell groups is displayed in the Fragment Distribution plot.

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.

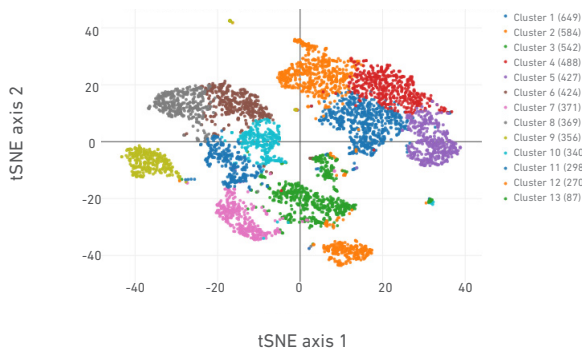


Plot & Interpretation

**Cell Clustering Scatter Plot:** The 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 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.

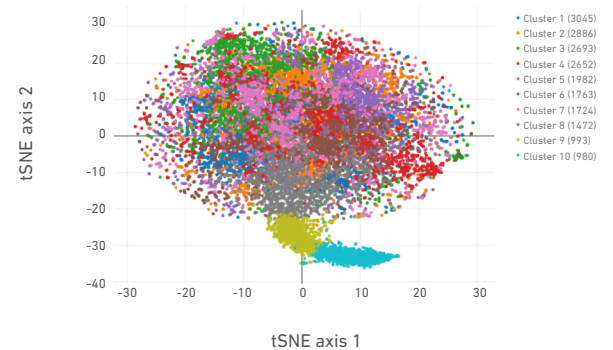


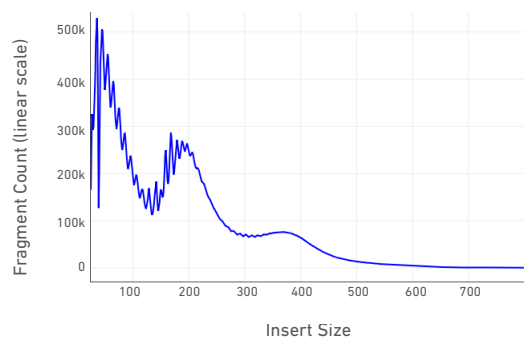
Table 2 contd. Plots in the ATAC web summary file.

### Plot & Interpretation

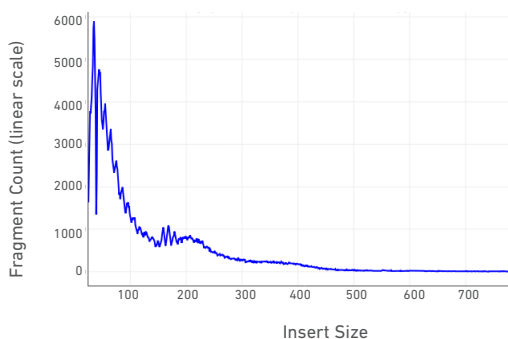
**Insert Size Distribution Plot:** Insert size distribution of transposase accessible fragments sequenced is displayed in the Insert Size Distribution plot.

#### 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.

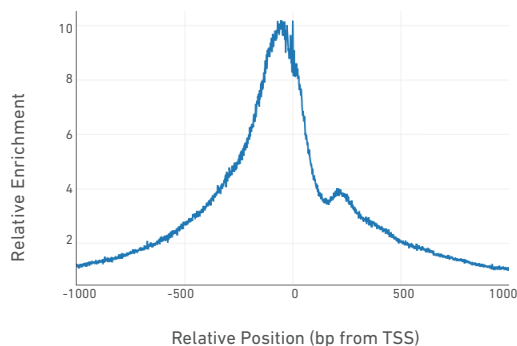


### Plot & Interpretation

**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.

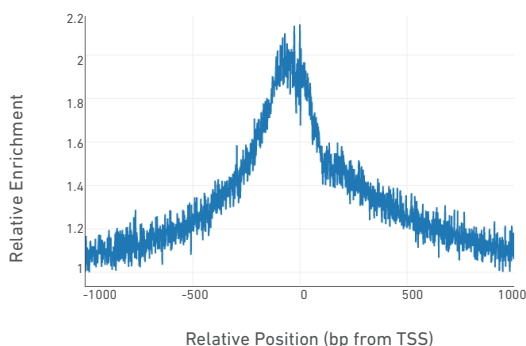


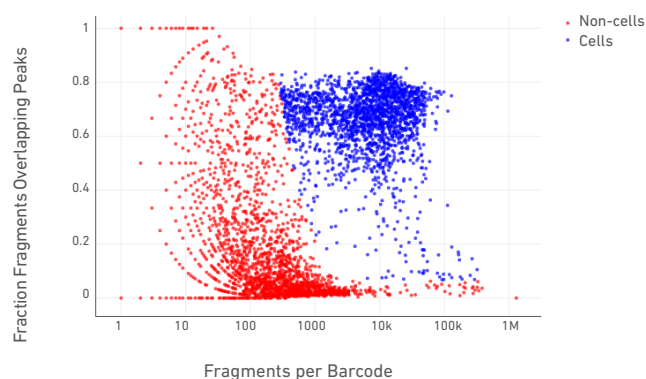
Table 2 contd. Plots in the ATAC web summary file.

### Plot & Distribution

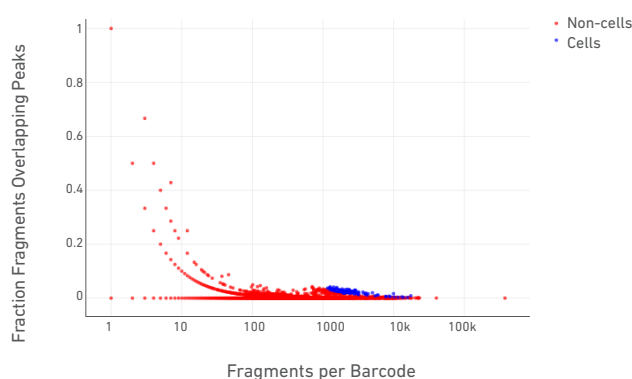
**Single Cell Targeting Plot:** A scatterplot displaying the number of fragments per barcode and the percent of fragments overlapping peaks.

### 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. Concentration of 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)

© 2019 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](http://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](http://www.10xgenomics.com/patents). The use of products described herein is subject to 10x Genomics Terms and Conditions of Sale, available at [www.10xgenomics.com/legal-notices](http://www.10xgenomics.com/legal-notices), or such other terms that have been agreed to in writing between 10x Genomics and user. 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](mailto:support@10xgenomics.com)  
10x Genomics  
7068 Koll Center Parkway  
Suite 401  
Pleasanton, CA 94566 USA

