TECHNICAL NOTE

Interpreting Cell Ranger Web Summary Files for Single Cell Expression Assay

Introduction

The web summary file in the output folder of the Cell Ranger analysis software is the initial point of reference for determining sample performance in the Single Cell Gene Expression 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 Gene Expression Solution.

Interpreting Web Summary File Metrics

Representative summary files for Chromium Single Cell Gene Expression libraries and other Cell Ranger output files are available for download on the 10x Genomics Support website. Several Metrics in the web summary file can be used to assess the overall success of an experiment, including sequencing, mapping, and cell metrics. A representative web summary file for a Chromium Single Cell 3' Gene Expression library is shown below (Figure 1).

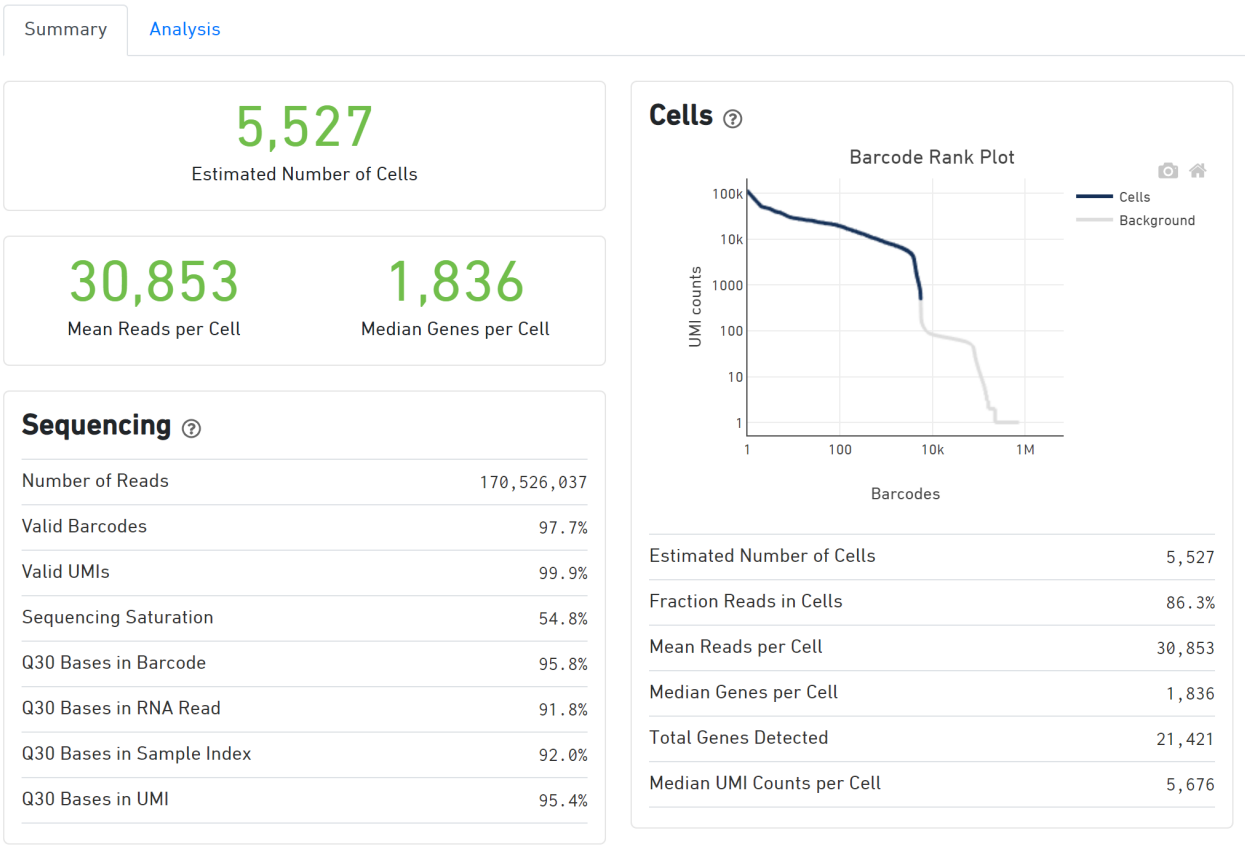


Figure 1. A representative web summary file (top section) from PBMCs sample targeting 5,000 cells from a healthy donor. The summary tab reports metrics that can be used to assess assay performance. Green text indicates that the key metrics are in the expected range while red/yellow text indicates errors/warnings. Descriptions of the metrics can also be found by clicking the icon ⓘ next to the section header.

Table 1. Metrics in the Cell Ranger summary file.

Metrics	Definition	Expected Value	Notes
Sequencing Metrics			
Number of reads	Total number of read pairs that were assigned to this library	Sequencing output dependent	Lower than expected may indicate poor sequencing run (over-clustering, under-clustering, low % passing filter).
Valid barcodes	Fraction of reads with barcodes that match the whitelist fraction of reads originating from an already observed UMI	>75%	Low valid barcodes may indicate sequencing issues (such as low Read 1 Q30 score).
Valid UMIs	Fraction of reads with valid UMIs; i.e. UMI sequences that do not contain Ns and that are not homopolymers	>75%	Low valid UMIs may indicate issues with sequencing or library quality.
Sequencing saturation	The fraction of reads originating from an already-observed UMI. This is a function of library complexity and sequencing depth	Dependent upon sequencing depth and sample complexity	Dependent on library complexity, sequencing depth, and experiment analysis goals. Lower sequencing saturation indicates a high proportion of the library complexity has not been captured by sequencing.
Q30 bases in barcode, Sample Index, or UMI	Fraction of cell barcode, Sample Index, or UMI bases with Q-score ≥30, excluding very low quality/ no call (Q-score ≤2) bases from the denominator	Sequencing platform dependent	Low Q30 base percentages could indicate sequencing issue such as sub-optimal loading concentration.
Q30 bases in RNA read	Fraction of RNA read bases with Q-score ≥30, excluding very low quality/no-call (Q-score ≤ 2) bases from the denominator	Sequencing platform dependent, ideally >65%	Expected to be lower than Q30 Bases in Barcode or UMI (Read 1) or Sample Index (i7 read) and is sequencing platform dependent. Consult Technical Note – Chromium Single Cell 3' v2 Libraries – Sequencing Metrics for Illumina Sequencers (v2 Chemistry) Document CG000089 for more information. Low Q30 Base percentages could indicate sequencing issue such as sub-optimal loading concentration.
Cell Metrics			
Estimated number of cells	The number of barcodes associated with at least one cell	500-10,000	Higher or lower than expected values may indicate inaccurate cell count, cell lysis, or failures during GEM generation.
Fraction reads in cells	The fraction of reads that contain a valid barcode, are confidently mapped to the transcriptome and are associated with a barcode that is called as a cell	>70%	Lower percentages indicate that a high level of ambient RNA partitioned into all (cell-containing and non-cell-containing) GEMs.
Median reads per cell	The total number of sequenced reads divided by the number of barcodes associated with cell-containing partitions	User defined; 20,000 reads/ cell minimum recommended	The necessary sequencing depth per cell depends on the cell type (high or low RNA) and the desired analysis.
Median genes per cell	The median number of genes detected per cell-associated barcode. Detection is defined as the presence of at least 1 UMI count	Dependent on cell type and sequencing depth	Lower than expected 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 any cell	Dependent on cell type and sequencing depth	Lower than expected could be a result of shallower sequencing depth and/or sample/library quality.
Median UMI counts per cell	The median number of UMI dependent on cell counts per cell-associated type barcode	Dependent on cell type and sequencing depth	Lower than expected could be a result of shallower sequencing depth and/or sample/library quality.

Table 1 contd. Metrics in the Cell Ranger summary file.

Metrics	Definition	Expected Value	Notes
Mapping Metrics			
Reads mapped to genome	Fraction of reads that are mapped to the genome	Variable	Dependent on the quality of genome annotation. Lower than expected values could be an indication of incorrect reference selection or library quality.
Reads mapped confidently to genome	Fraction of reads that mapped uniquely to a genome. A gene mapped to exonic loci from a single gene and also to non-exonic loci is considered uniquely mapped to one of the exonic loci	Variable	Lower than expected values could be indicative of low library quality or reference quality.
Reads mapped confidently to intergenic regions	Fraction of reads that mapped uniquely to an intergenic region of the genome	Variable	May vary based on sample type and genome annotation.
Reads mapped confidently to intronic regions	Fraction of reads that mapped uniquely to an intronic region of the genome	Variable	Sample types with low RNA content (e.g. PBMCs, nuclei) or samples with suboptimal health may have a higher fraction of reads mapping to intronic regions.
Reads mapped confidently to exonic regions	Fraction of reads that mapped uniquely to an exonic region of the genome	Variable	There is a balance between exonic and intronic reads. A sample with higher exonic reads will have lower intronic reads, and vice versa. This is highly dependent upon sample type.
Reads mapped confidently to transcriptome	Fraction of reads that mapped to a unique gene in the transcriptome. The read must be consistent with annotated splice junctions. These reads are considered for UMI counting	Variable, ideally >30%	Reference quality and sequencing configuration (shorter than recommended cycles on Read 2) can impact mapping. Lower than expected values may indicate the use of the wrong reference transcriptome.
Reads mapped antisense to gene	Fraction of reads confidently mapped to the transcriptome, but on the opposite strand of their annotated gene. A read is counted as antisense if it has any alignments that are consistent with an exon of a transcript but antisense to it, and has no sense alignments	Ideal <10%	These values may be higher if using a pre-mRNA reference or may indicate incorrect Gel Bead chemistry.

Interpreting the Web Summary File Plots

The summary file also contains a Barcode Rank plot and t-SNE Projection plots (Table 2).

Table 2. Plots in the web summary file.

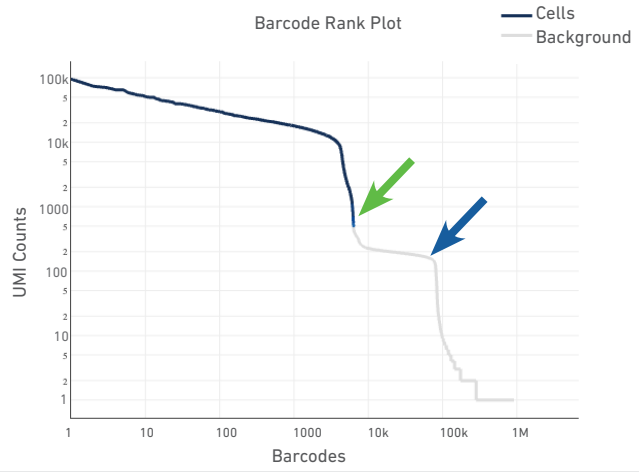
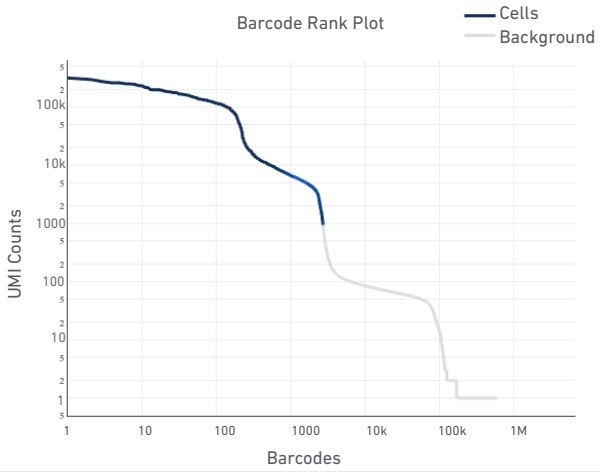
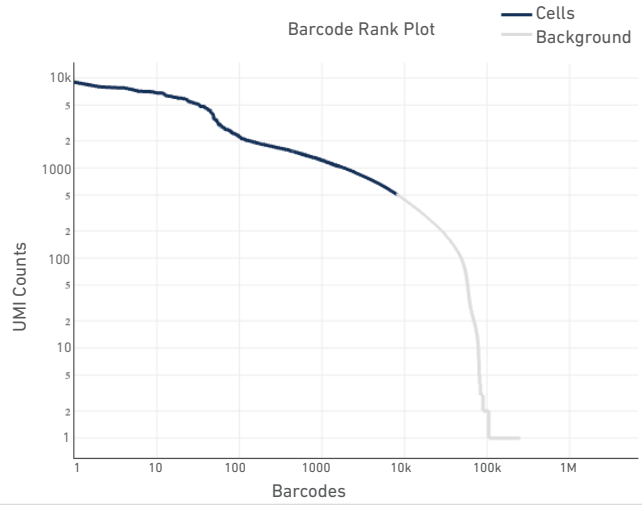
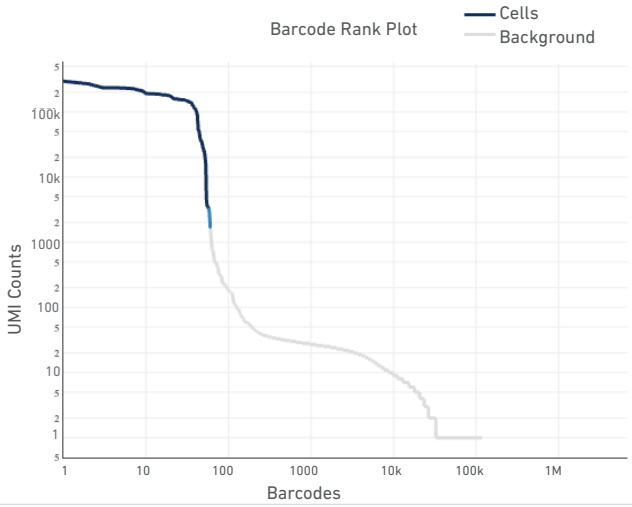
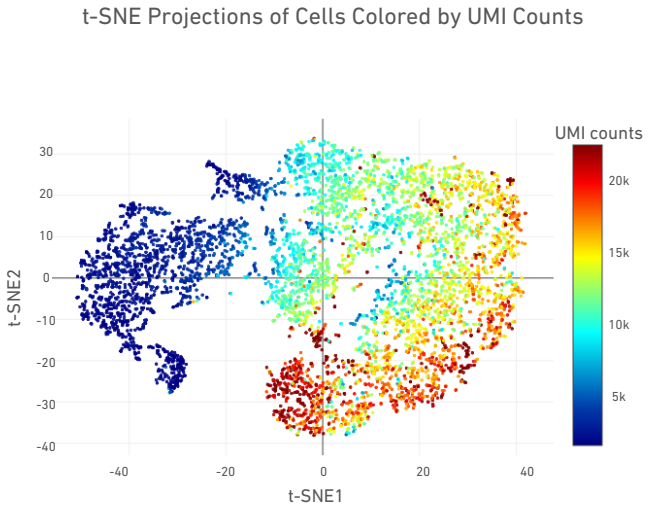
Plot & Interpretation
Barcode Rank Plot: All 10x Barcodes detected during sequencing (~100k) are plotted in decreasing order of the number of UMIs associated with that particular barcode. The number of UMIs detected in each GEM is used by Cell Ranger to determine which GEMs likely contain a cell. GEMs containing cells are expected to have a greater number of transcripts (and thus UMIs) associated with them than non-cell containing GEMs.
Examples
Typical Sample: A steep drop-off is indicative of good separation between the cell-associated barcodes and the barcodes associated with empty GEMs. An ideal Barcode Rank plot has a distinctive shape, which is referred to as a "cliff and knee". The blue-to-gray transition (green arrow) is referred to as the cliff; the solid gray is referred to as the knee (blue arrow).

Heterogeneous Sample: Occasionally and based on sample type, there can be heterogeneous populations of cells in a sample that may result in a bimodal plot. In these situations, the cell-associated 10x Barcodes will have two "cliff and knee" distributions. However, there should still be clear separation between the barcodes called as 'cells' and barcodes called as 'background'.

Compromised Sample: Round curve and lack of steep cliff may indicate low sample quality or loss of single-cell behavior. This can be due to a wetting failure, premature cell lysis, or low cell viability.

Compromised Sample: Defined cliff and knee, but the total number of barcodes detected may be lower than expected. This can be caused by a sample clog or inaccurate cell count.


Table 2 contd. Plots in the web summary file.

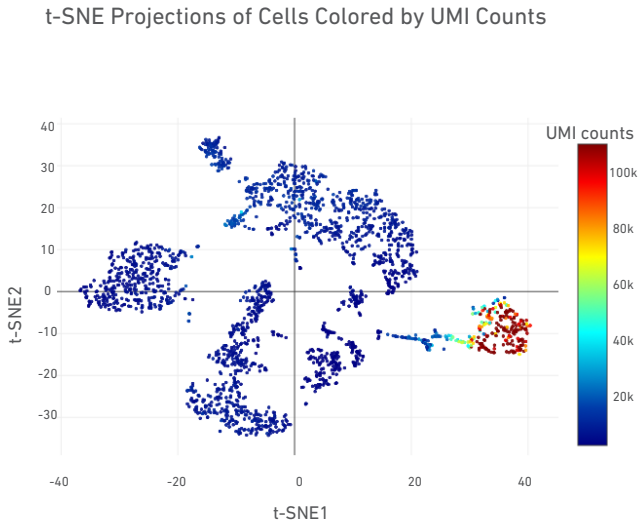
Plot & Interpretation
t-SNE Projection of Cells Colored by UMI Counts: The Cell Clustering plot shows the cell-associated barcodes in a 2-D t-SNE projection. Each cell is color-coded according to how many UMIs it contains. Cells with greater UMI counts likely have higher RNA content than cells with fewer UMI counts.

Examples

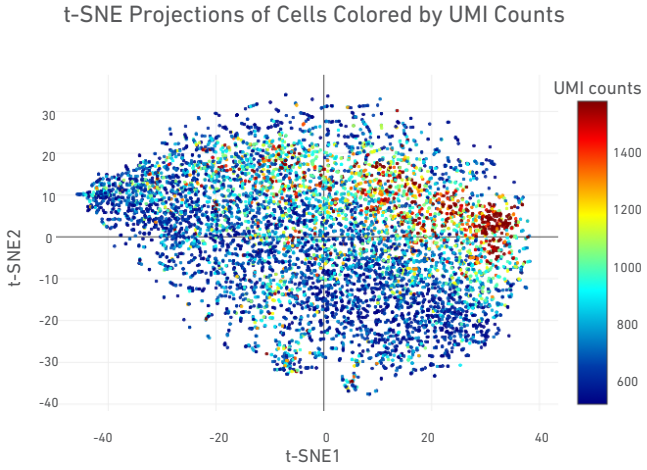
Typical Sample: Structured clusters with clear separation between high UMI and low UMI containing barcodes.



Heterogeneous Sample: Heterogeneous samples that contain high and low RNA containing cells should have visible differences in the UMIs levels between the two 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 issues or loss of single-cell behavior.



Compromised sample: Very low cell numbers may result in scattered t-SNE plots with little cluster structure.

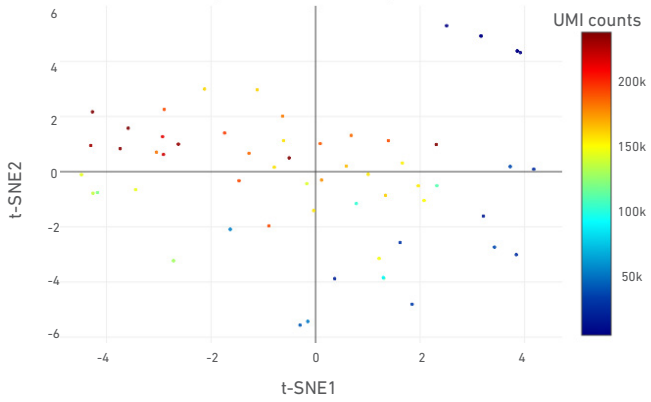
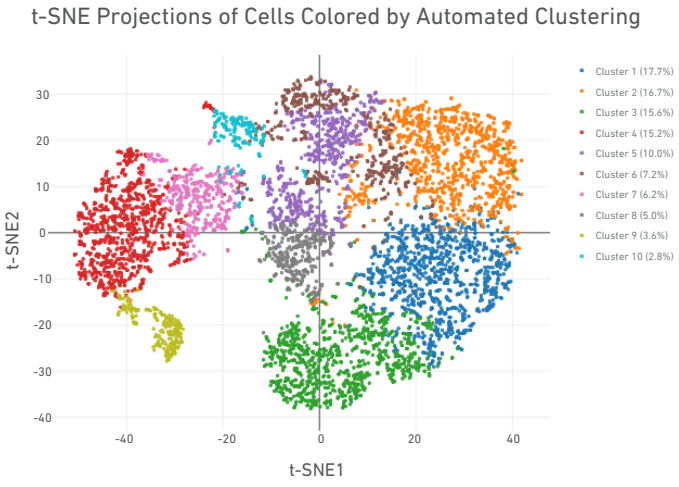


Table 2 contd. Plots in the web summary file.

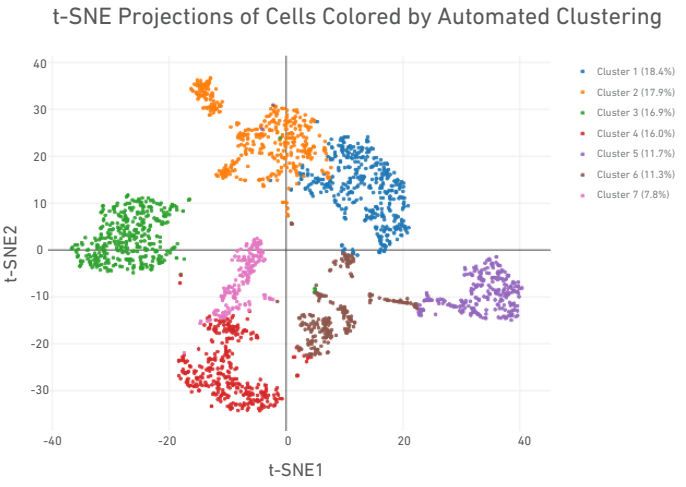
Plot & Interpretation
t-SNE Projection of Cells Colored by Clustering: The Cell Clustering (colored by cluster) plot shows the cell-associated barcodes in a 2-D tSNE projection, with colors showing an automated graph clustering analysis which groups together cells with similar gene expression profiles.

Examples

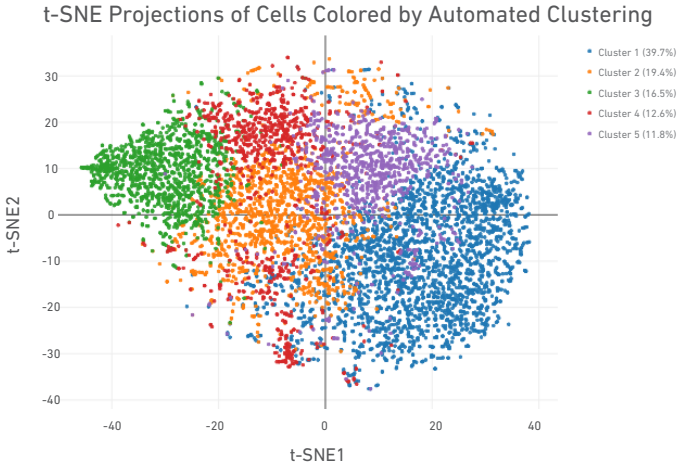
Typical Sample: Structured clusters with good separation (for a sample with expected heterogeneous cell populations).



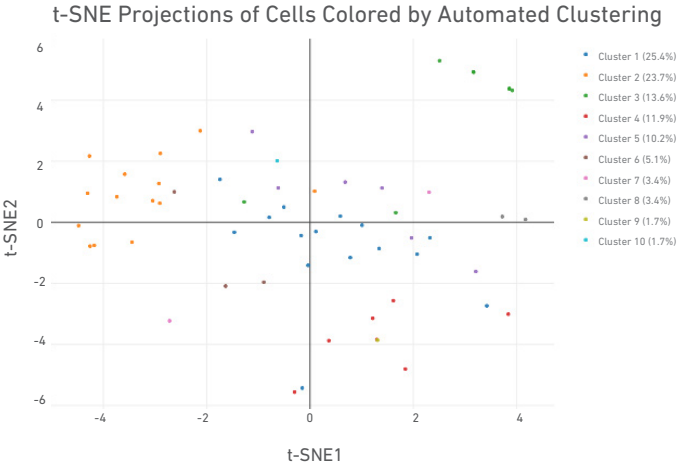
Heterogeneous sample: Structured clusters with good separation (for a sample with expected heterogeneous cell populations).



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.



Compromised sample: Very low cell numbers may result in scattered cell t-SNE plots with little cluster structure.



References

- Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide (Document CG000204)
- Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide with Feature Barcoding technology for CRISPR Screening (Document CG000205)
- Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide with Feature Barcoding technology for Cell Surface Protein (Document CG000206)
- Chromium Next GEM Single Cell V(D)J Reagent Kits v1.1 User Guide (Document CG000207)
- Chromium Next GEM Single Cell V(D)J Reagent Kits v1.1 User Guide with Feature Barcoding technology for Cell Surface Protein (Document CG000208)
- Chromium Single Cell 3' Reagent Kits v3 User Guide (Document CG000183)
- Chromium Single Cell 3' Reagent Kits v3 User Guide with Feature Barcoding technology for CRISPR Screening (Document CG000184)
- Chromium Single Cell 3' Reagent Kits v3 User Guide with Feature Barcoding technology for Cell Surface Protein (Document CG000185)
- Chromium Single Cell V(D)J Reagent Kits User Guide (Document CG000086)
- Chromium Single Cell V(D)J Reagent Kits User Guide with Feature Barcoding technology for Cell Surface Protein (Document CG000186)
- Chromium Single Cell 3' Reagent Kits v2 User Guide (Document CG00052)

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