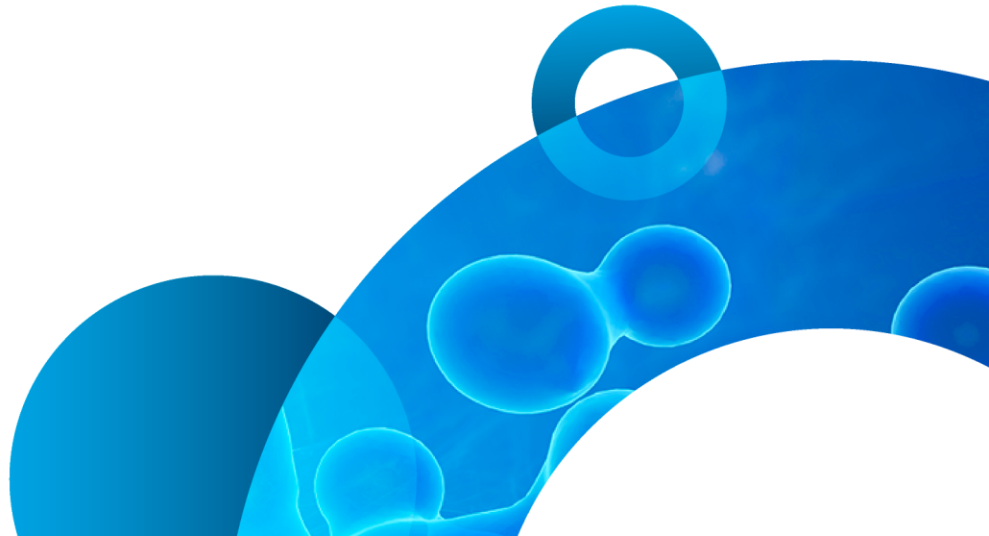
Chromium Single Cell Gene Expression with Feature Barcode Analysis

Software Training



CG000200 Rev D

Objectives

- Overview of Cell Ranger for Single Cell analysis including
  - Whole transcriptome gene expression
  - Targeted gene expression
  - Feature Barcode technology
- Understand key steps in Cell Ranger Count pipeline
- Overview of key output files generated by the pipeline
- Understand the run metrics and know what to look for to quickly gauge the quality of the run

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Outline

- Workflow Review
- Cell Ranger Overview
- Demultiplexing with “mkfastq”
- Cell Ranger “count” Pipeline for Gene Expression Analysis
- Gene Expression Analysis Outputs
- Gene Expression Analysis QC Metrics
- Additional Cell Ranger Pipelines: “aggregate” and “reanalyze”
- Optional: Targeted Gene Expression analysis
- Optional: Feature Barcode data analysis for Cell Surface Protein or CRISPR Screening

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

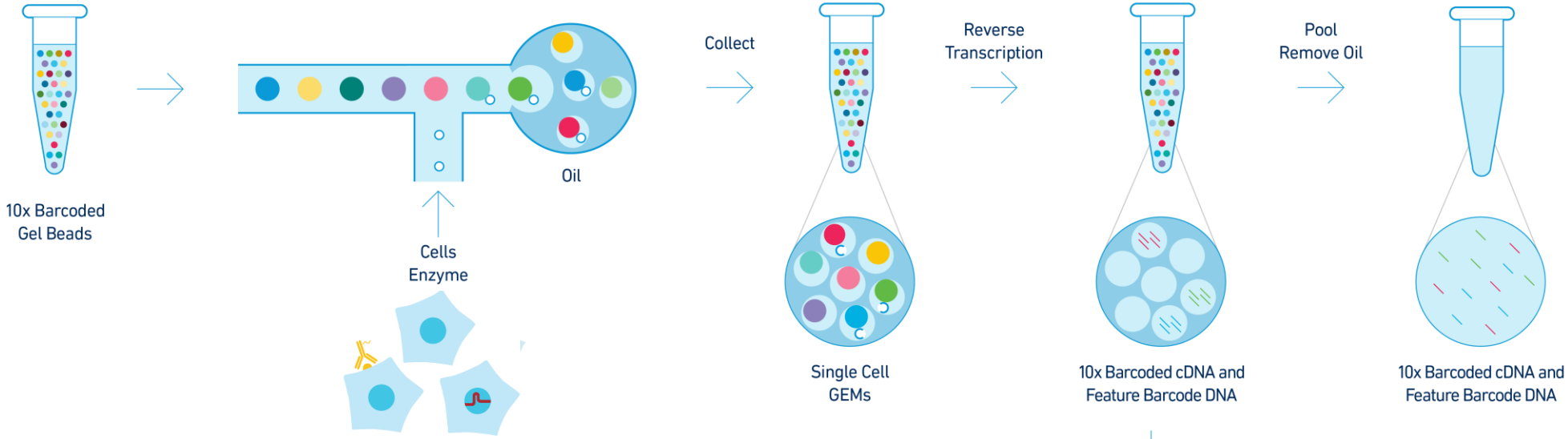
Workflow review

Single Cell 3’ Gene Expression with
Feature Barcode technology

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Single Cell Gene Expression with Feature Barcode technology

Workflow overview

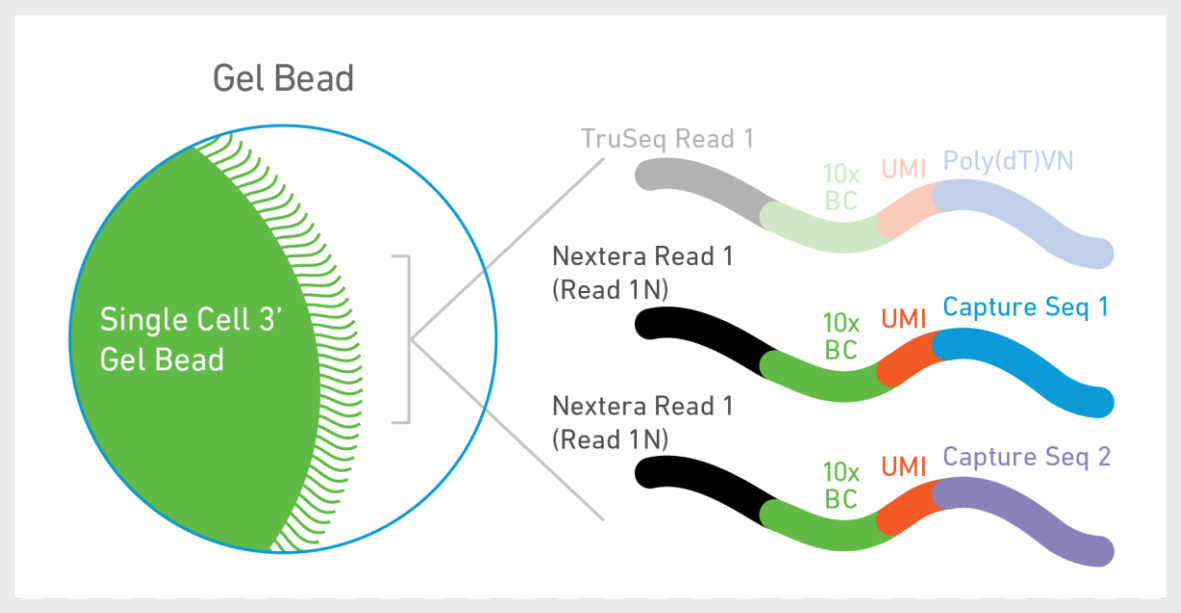


Inputs:
10x Gel Beads, Reagents and single cells in suspension
Outputs:
Digital gene expression and cell surface protein expression or CRISPR perturbation profiles from every partitioned cell

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Single Cell Gene Expression with Feature Barcode technology

Feature Barcode technology enabled via Single Cell 3′ Gel Beads



i. **Nextera Read 1 (Read 1N)**

22 nt Partial Illumina Nextera Read 1 sequence (Enables selective enrichment of the Feature Barcode construct)

ii. **10x Barcode**

16 nt 10x Barcode
~3.6 M defined barcode sequences

iii. **UMI**

12 nt Unique Molecular Identifier

iv. **Capture Sequence 1** or **2**

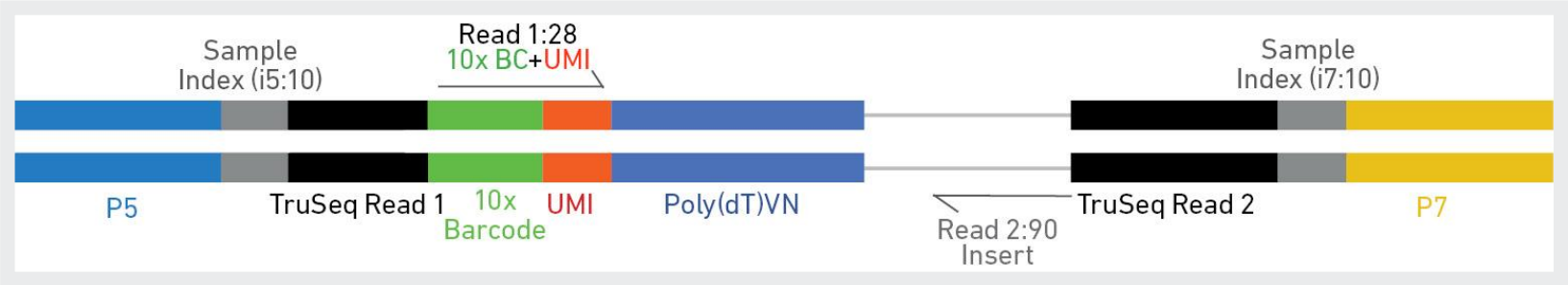
22 nt sequence that is the reverse complement of the sequence inserted into the DNA (Antibody) or RNA (sgRNA) based sequence

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Read structure and sequencing

Single Cell 3’ Gene Expression library



*Minimum 20,000 read pairs per cell is recommended for 3’ Gene Expression libraries**

	Read 1	i7 Index	i5 Index	Read 2
Purpose	10x Barcode & UMI	Sample Index	Sample Index	Transcript
Length	28	10	10	90**

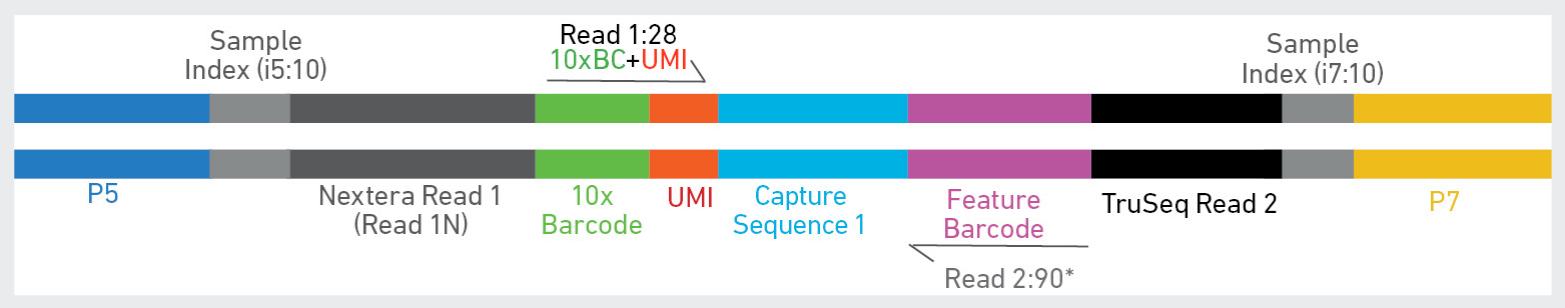
* 10-20% of depth of Whole Transcriptome library is recommended for Targeted libraries with pre-designed panels.

** User controlled trade-off between read length and mapping rate

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Read structure and sequencing

Single Cell 3’ cell surface protein library



Minimum 5,000 read pairs per cell is recommended for Cell Surface Protein libraries

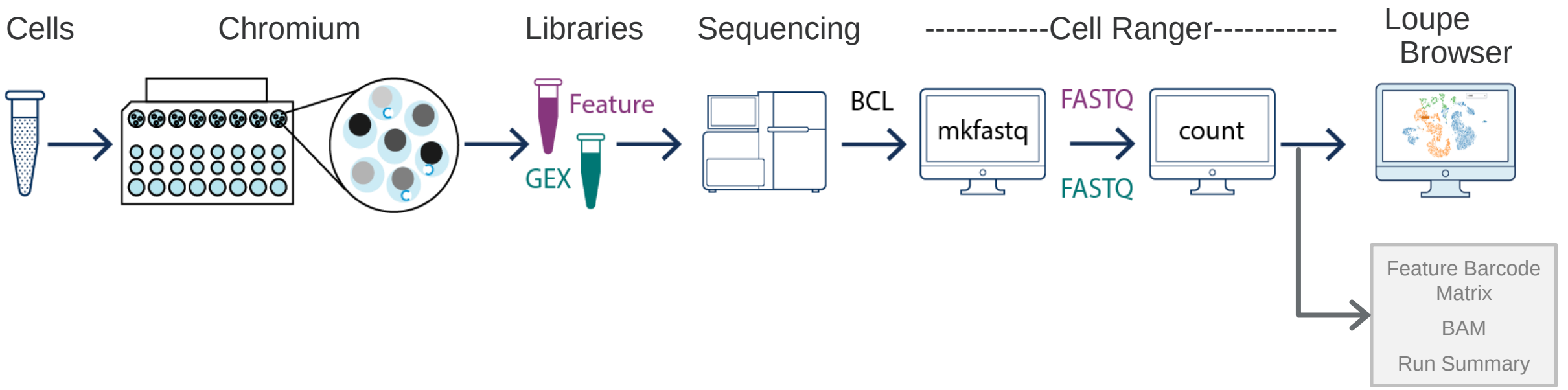
	Read 1	i7 Index	i5 Index	Read 2
Purpose	10x Barcode & UMI	Sample Index	Sample Index	Insert (Feature Barcode)
Length	28	10	10	90*

*Single Cell 3’ Cell Surface Protein Libraries are typically pooled with Single Cell Gene Expression Libraries and sequenced using these parameters.

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Analyzing gene expression data

Laboratory and analysis workflow overview



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger overview

Single Cell Gene Expression with
Feature Barcode technology

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger

Introduction

- A set of analysis pipelines that process Chromium Single Cell RNAseq data
- Capable of analyzing Gene Expression data with or without additional Feature Barcode data
- Contains various pipelines for:
  - Demultiplexing (mkfastq)
  - Single sample analysis (count)
  - Combining data from multiple samples into an experiment-wide analysis (aggr)
  - Reanalyzing data with custom parameters (reanalyze)
- Easy to download and run on Linux
  - http://support.10xgenomics.com

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger

System requirements





Local Mode

- Run on single, standalone Linux system
- 64-bit CentOS/RedHat 6.0 or Ubuntu 12.04
- 8 cores, 64GB RAM (minimum)

Cluster Mode

- Run on SGE and LSF
- Each node must have 8+ cores and 8GB+ RAM/core
- Shared filesystem between nodes (e.g. NFS)

Runtime

- GEX only
  - ~8000 cells, 35k reads/cell
  - ~4 hours wall-time with 16 cores, 64 GB RAM
- Antibody only
  - ~8000 cells, 16k reads/cell
  - ~1 hours wall-time at 16 cores, 64 GB RAM

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Demultiplexing

Generating FASTQs

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Executing `cellranger mkfastq`

Demultiplexing

- Run the `cellranger mkfastq` pipeline to demultiplex an Illumina sequencing run folder into FASTQ files

```
$cellranger mkfastq --id=run1 \
                    --csv=my_samples.csv \
                    --run=/mnt/hiseq/run1_bcl
```

Output directory name

Sample sheet

Path to bcls

- Output

```
$ls -l /opt/bar/run1/outs/fastq_path
drwxr-xr-x 3 jdoe jdoe           3 Sep 24 19:19 HAWT7ADXX
drwxr-xr-x 3 jdoe jdoe           3 Sep 24 19:17 Reports
drwxr-xr-x 2 jdoe jdoe          10 Sep 24 19:17 Stats
-rw-r--r-- 1 jdoe jdoe    27971547 Sep 24 18:07 Undetermined_S0_L001_I1_001.fastq.gz
-rw-r--r-- 1 jdoe jdoe    44432222 Sep 24 18:07 Undetermined_S0_L001_R1_001.fastq.gz
-rw-r--r-- 1 jdoe jdoe   912086076 Sep 24 18:07 Undetermined_S0_L001_R2_001.fastq.gz

$tree opt/bar/run1/outs/fastq_path/HAWT7ADXX/sample1/
opt/bar/sample1/outs/fastq_path/HAWT7ADXX/sample1
|── sample1_S1_L001_I1_001.fastq.gz
|── sample1_S1_L001_R1_001.fastq.gz
|── sample1_S1_L002_R2_001.fastq.gz
```

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline for gene expression analysis

Command line and pipeline steps

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Executing “cellranger count”

- This pipeline performs read alignment, UMI counting and secondary analysis for single sample.
- Typical command line
  - If FASTQ files are in a directory as below

```
opt/bar/run1
|— sample1_S1_L001_I1_001.fastq.gz
|— sample1_S1_L001_R1_001.fastq.gz
|— sample1_S1_L002_R2_001.fastq.gz
```

  - Command run as below

```
cellranger count --id=mysample \
                 --fastqs=/opt/bar/run1 \
                 --sample=sample1 \
                 --transcriptome=refdata-cellranger/GRCh38
```

Output directory name

Path to fastq files

Sample name in fastq files

Reference

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Reference transcriptome

- 10x Genomics pre-built references
  - Human (GRCh38)
  - Mouse (GRCm38)
  - Human and Mouse (GRCh38 + GRCm38)

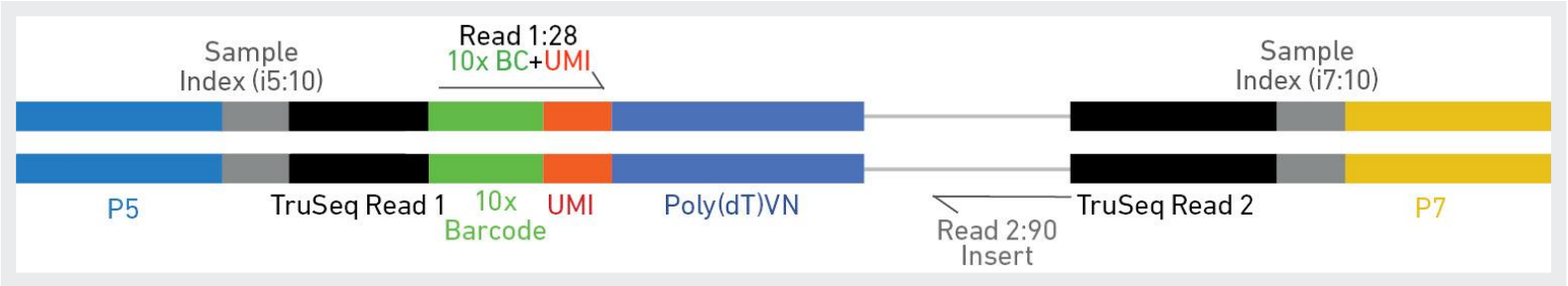
- Custom reference
  - Bundled mkref utility generates a 10x reference package from any FASTA and GTF gene file
  - Supports any reference compatible with STAR

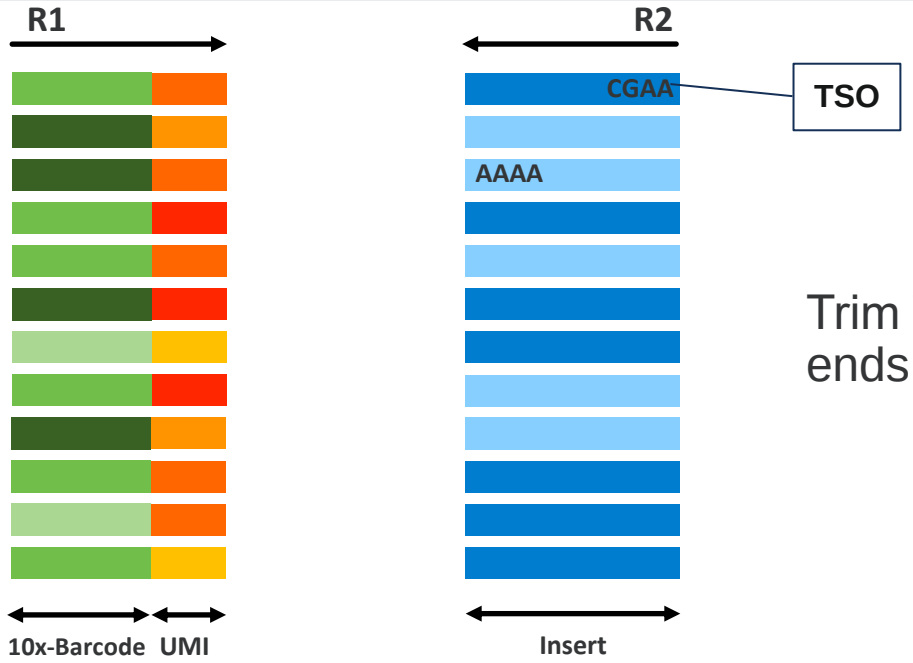
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline

Single Cell 3’ Gene Expression libraries

- **Read Processing**
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Cells
- Secondary Analysis





Trim TSO from 5’ and polyA from 3’ ends of R2 read

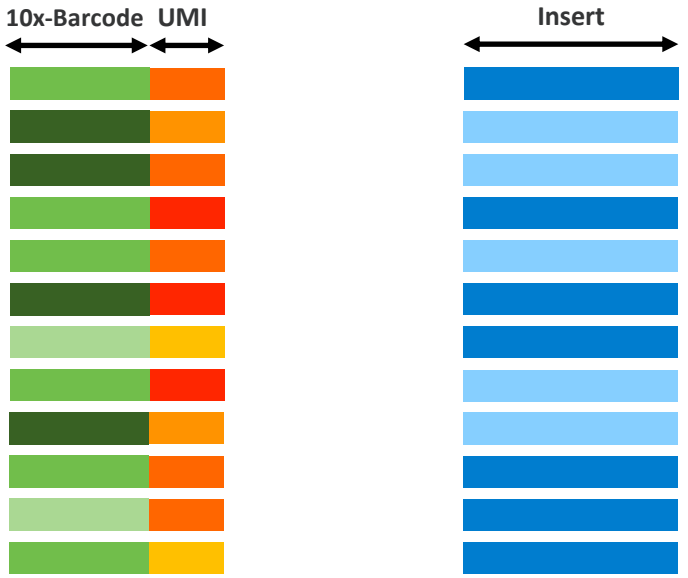
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline: Align

- Read Processing
- **Align**
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Cells
- Secondary Analysis



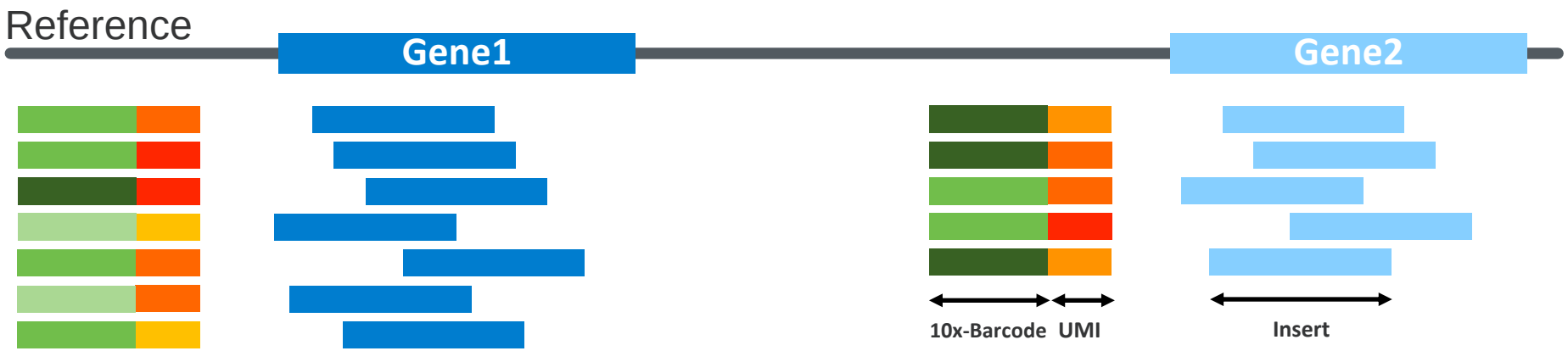
In this step, the reads are aligned to the reference transcriptome



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline: Align

- Read Processing
- **Align**
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Cells
- Secondary Analysis



Alignment done via STAR (Spliced Transcripts Alignment to a Reference)

- Robust, open-source, junction-aware RNA-seq aligner
- Aligns reads to the genome and transcriptome simultaneously

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline: Correct barcodes

Read Processing

Align

Correct Barcodes

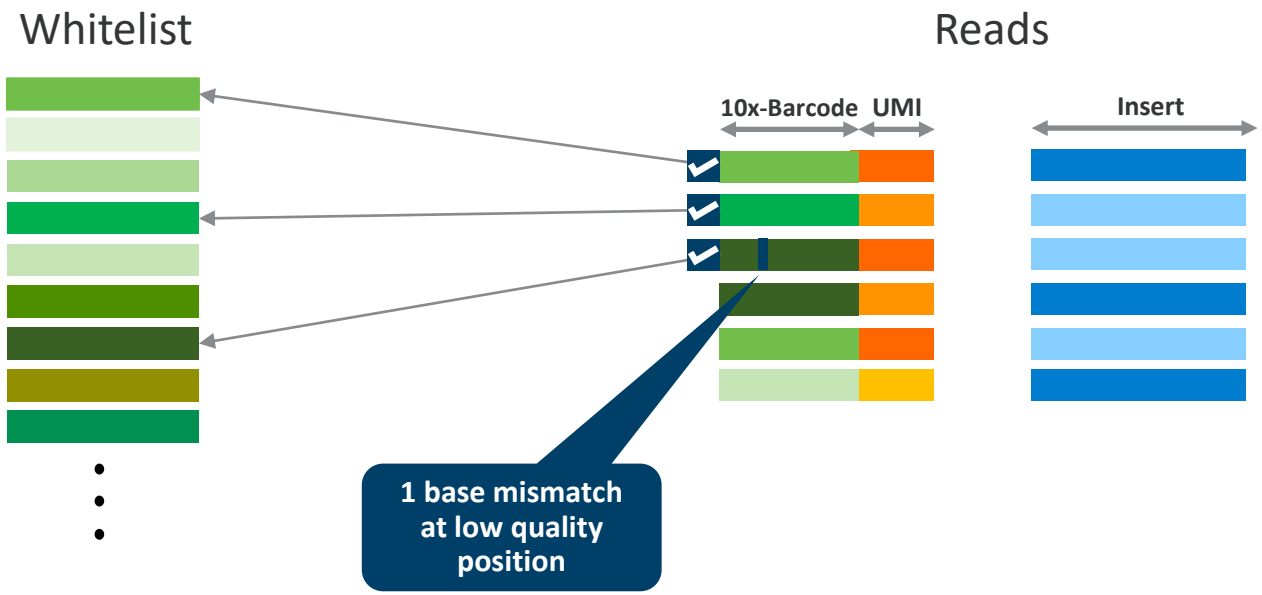
Filter UMIs

Count UMIs

Call Cells

Secondary Analysis

- Cell barcodes
  - Must be on a list of known cell barcode sequences
  - May allow 1 bp mismatch if the mismatch occurs at a low-quality position (the barcode is then corrected).



10X GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline: Filter UMIs

Read Processing

Align

Correct Barcodes

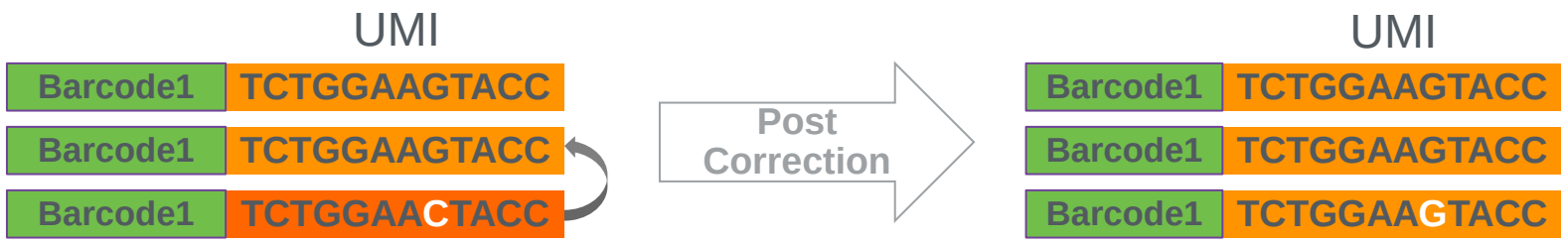
Filter UMIs

Count UMIs

Call Cells

Secondary Analysis

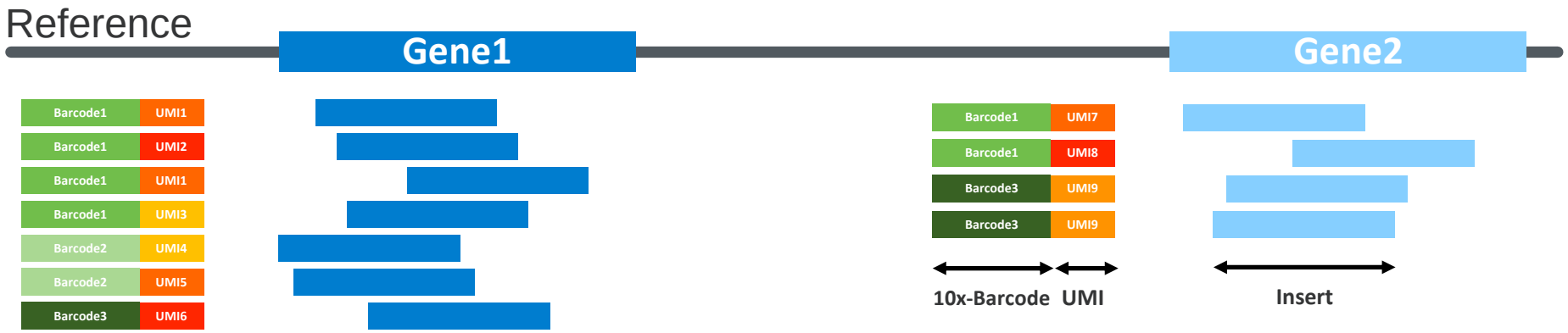
- Filter UMIs
  - Must not be a homopolymer, e.g. AAAAAAAAAA
  - Must not contain N
  - Must not contain bases with base quality < 10
- Correct UMIs
  - UMIs that are 1 mismatch away from a higher count UMI are corrected to that UMI if they share a cell barcode and gene.



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline: Count UMIs

- Read Processing
- Align
- Correct Barcodes
- Filter UMIs
- **Count UMIs**
- Call Cells
- Secondary Analysis



Raw feature barcode matrix

Genes

Barcodes	Gene1	Gene2
Barcode1	3	2
Barcode2	2	0
Barcode3	1	1

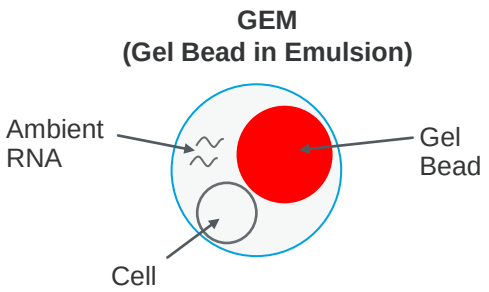
- Only the confidently mapped reads with valid barcodes and UMIs are used for counting
- Counting is done for all observed barcodes generating an unfiltered feature barcode matrix

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline: Cell calling

Why ?

- Read Processing
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- **Call Cells**
- Secondary Analysis



- One Chromium Single Cell Chip A or Chip B channel generates approximately 100,000 GEMs
- What is in the ~100,000 GEMs?
  - A fraction of GEMs contains 1 cell each



  - Typically very small number (depending of cell load) contain more than one cells



  - A large fraction of GEMs are considered “empty,” without any cells, but do contain ambient RNA. These are called background barcodes.



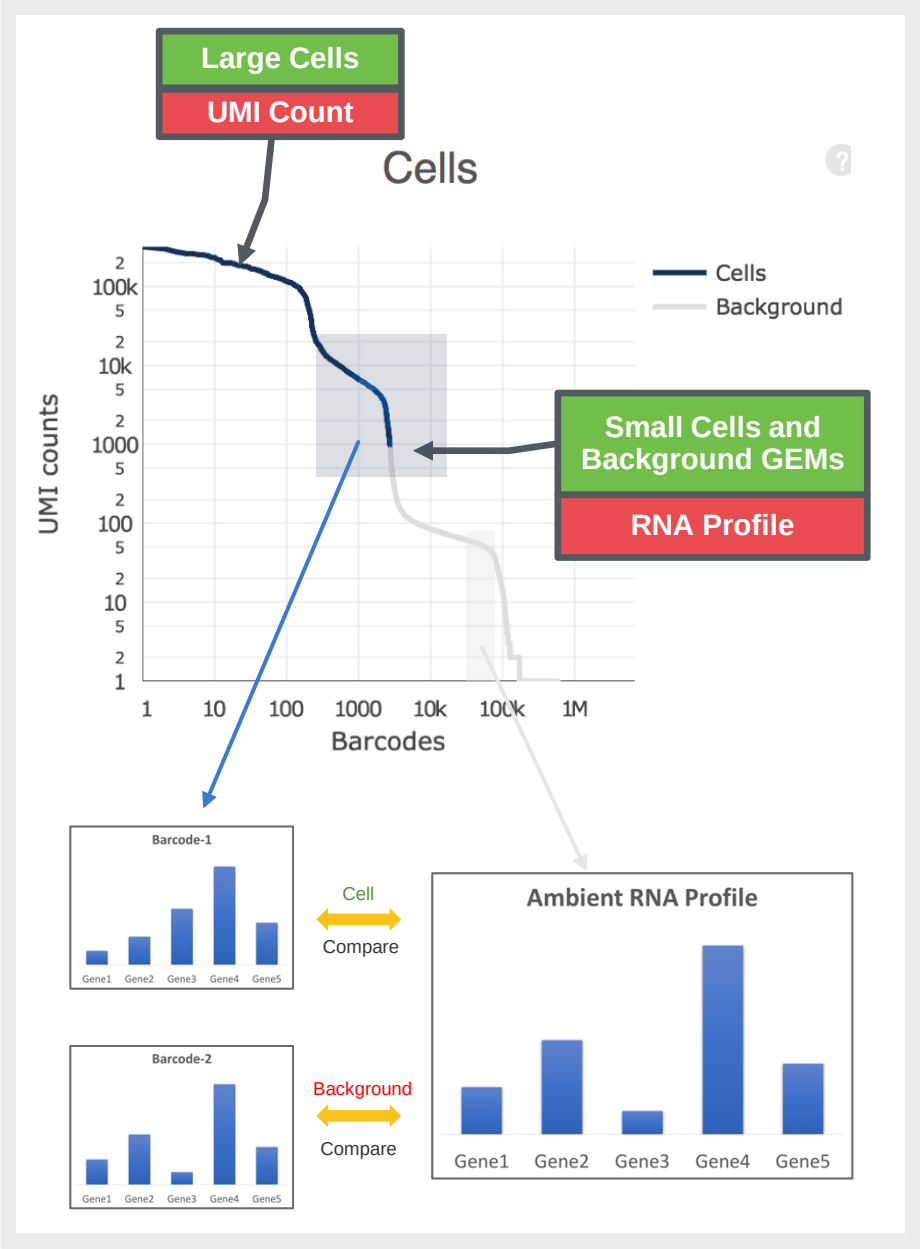
- Need to distinguish GEMs/barcodes containing cells from background barcodes

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Calling
Algorithm

- Based on Empty Drops method (Lun et al., 2018)
  https://www.biorxiv.org/content/early/2018/04/04/234872
- Step1: Identify first mode of high RNA content cells.
  - Filter based on UMI threshold.
  - Threshold is 10% of a robust estimate of maximum UMI counts per cell.
- Step 2: Find additional cells based on their RNA profile.
  - Select a set of low UMI barcodes to represent background GEMs.
  - Generate the ambient RNA profile from the selected barcodes.
  - Take top 20000 remaining barcode
  - Discard barcodes with total UMI count < 500 or < median(initial_cell_umis) * 1%
  - Compare the RNA profile of selected barcodes with ambient profile.
  - Call barcodes with RNA profile significantly different from ambient as cells.
- Generate “filtered feature barcode matrix” with barcodes called as cells

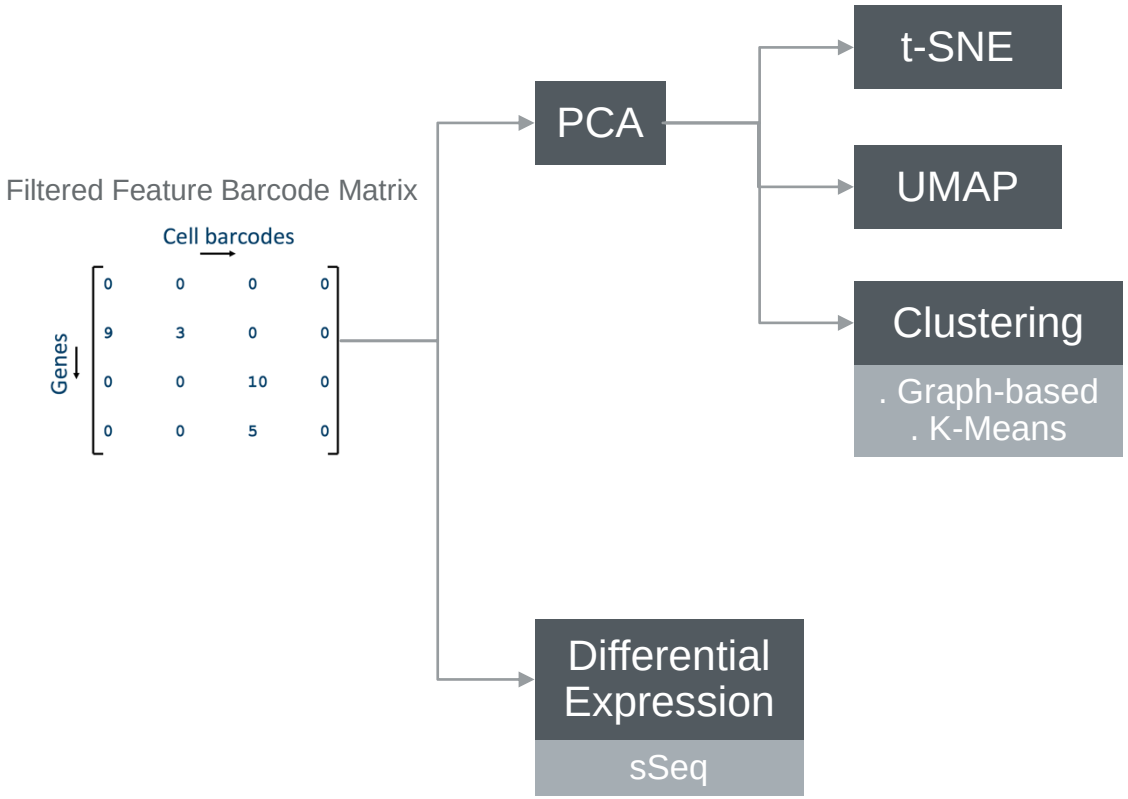


© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” pipeline: Secondary analysis

- Read Processing
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Cells
- Secondary Analysis

- Secondary analysis done only on cells





© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Gene expression analysis outputs

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” output

- QC metrics: web_summary.html, metrics_summary.csv
- Feature barcode matrices (MEX and HDF5 formats)
- Position-sorted and indexed BAM file
- Secondary analysis
  - CSV file with PCA analysis
  - CSV files with t-SNE and UMAP projections
  - Cluster assignments for each cell are in clusters.csv for both K-means and graph-based clustering
  - CSV file indicating which genes are differentially expressed in each cluster relative to all other clusters
- .cloupe file for visualization in Loupe Browser

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Gene Expression analysis QC metrics

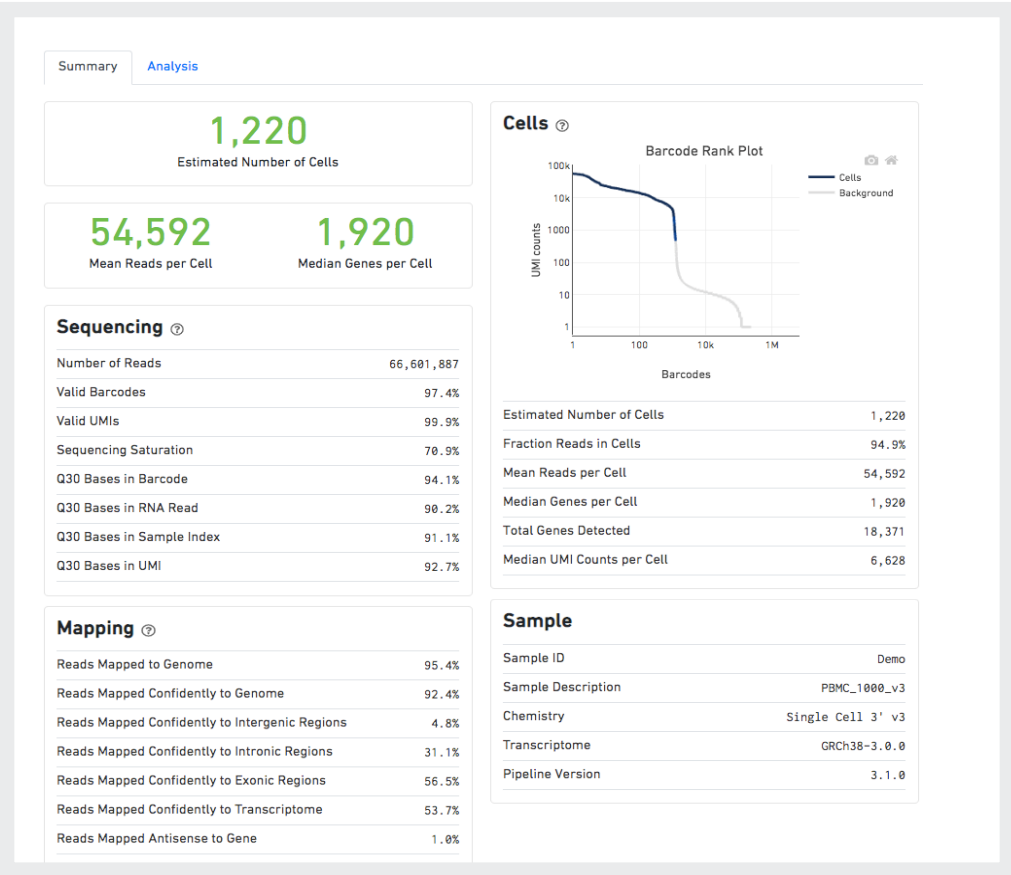
web_summary.html

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

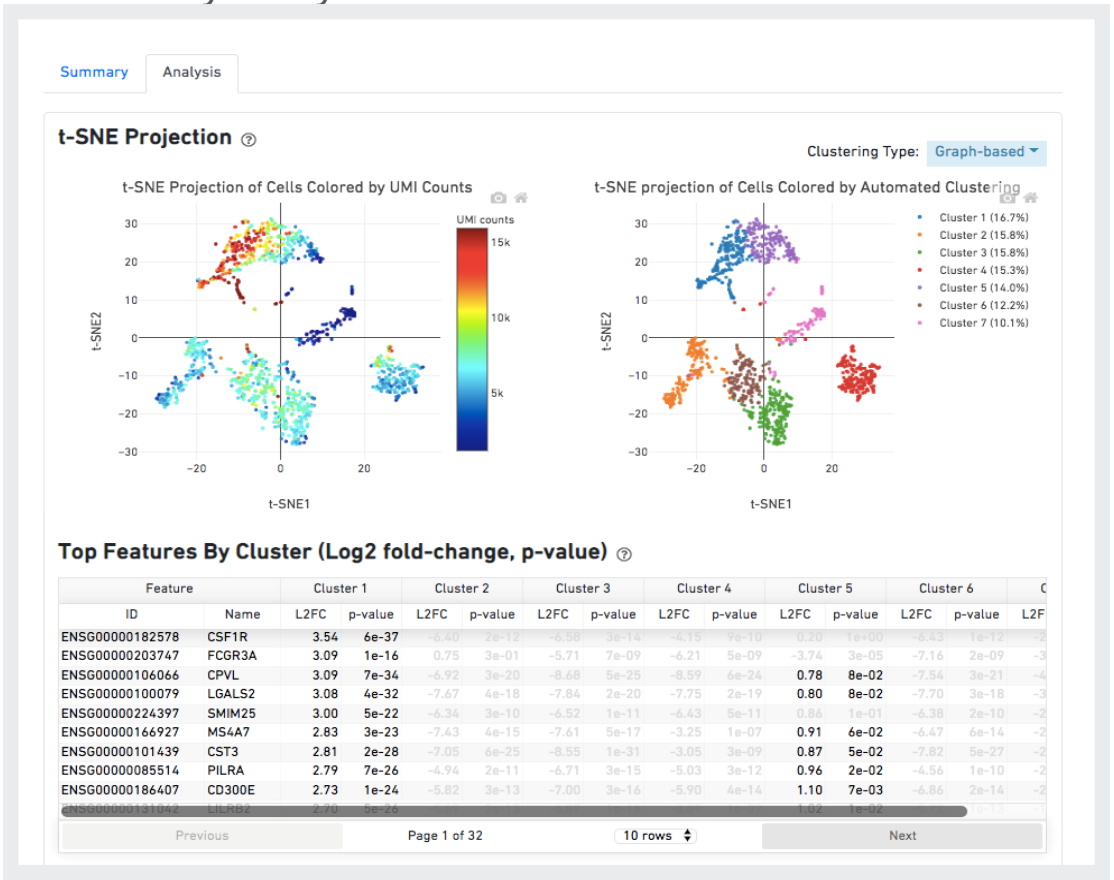
QC metrics: Gene expression

web_summary.html file

The summary metrics describe sequencing quality and various characteristics of the detected cells.



Analysis tab contains results from automated secondary analysis

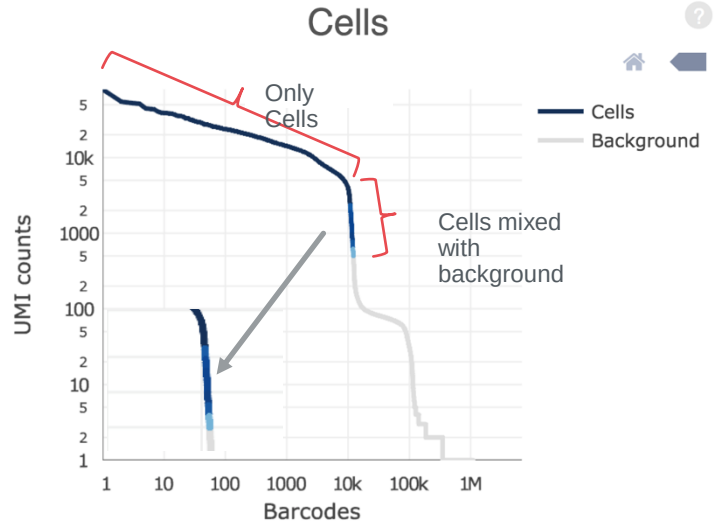


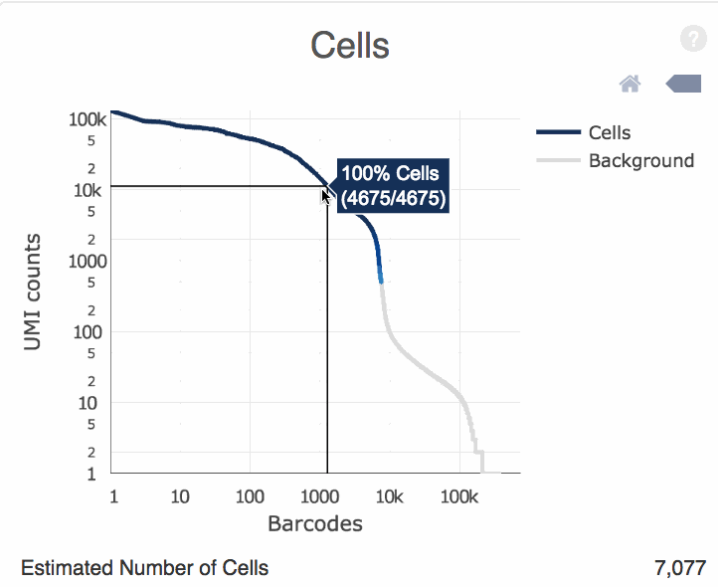
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC Metrics: Cell calling

Barcode rank plot

- Barcode rank plot shows distribution of barcode counts and which barcodes were inferred to be associated with cells.
- Cells are colored in “blue”
- Region on the plot where cells and background barcodes have similar UMI counts
- Blue color’s gradient is proportional to the fraction of cells in a given subset of barcodes
- Tool tip that displays the fraction of cells versus total cells in a given region of the plot





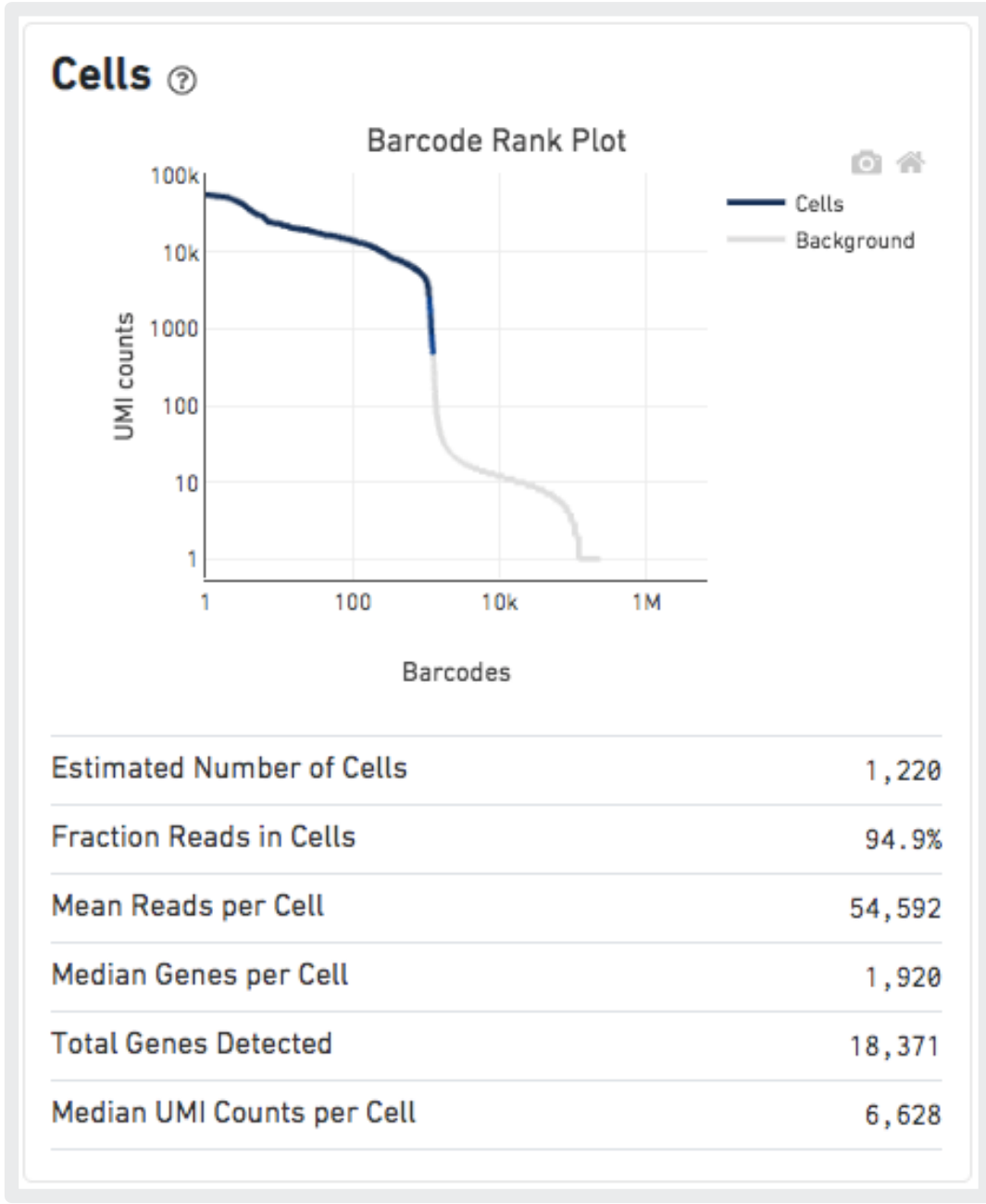
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC Metrics: Per cell metrics

Cell handling

Depth

Data yield



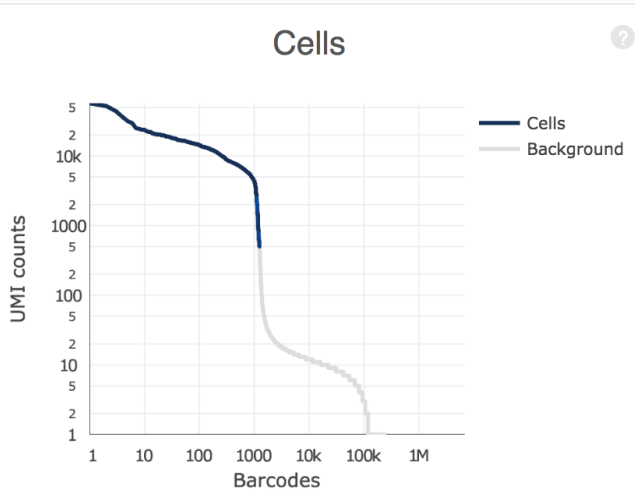
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC Metrics: Cell calling

Wetting failure

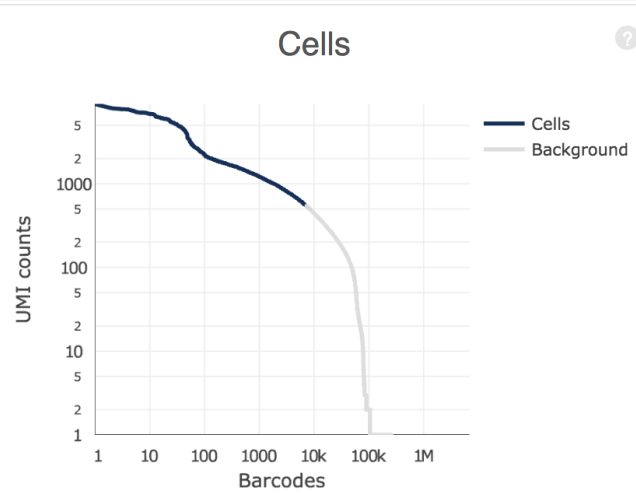
- A comparison of cell calling between a typical sample and a sample with wetting failure.
- Estimated number of cells is unreliable in such failure modes.

Typical Sample



Estimated Number of Cells	1,222
Fraction Reads in Cells	94.9%
Mean Reads per Cell	54,502
Median Genes per Cell	1,919
Total Genes Detected	18,391
Median UMI Counts per Cell	6,628

Loss of Single Cell Behavior (eg. Wetting failure)



Estimated Number of Cells	7,307
Fraction Reads in Cells	37.5%
Mean Reads per Cell	4,677
Median Genes per Cell	436
Total Genes Detected	15,426
Median UMI Counts per Cell	762

Algorithm has trouble discerning cells from the background

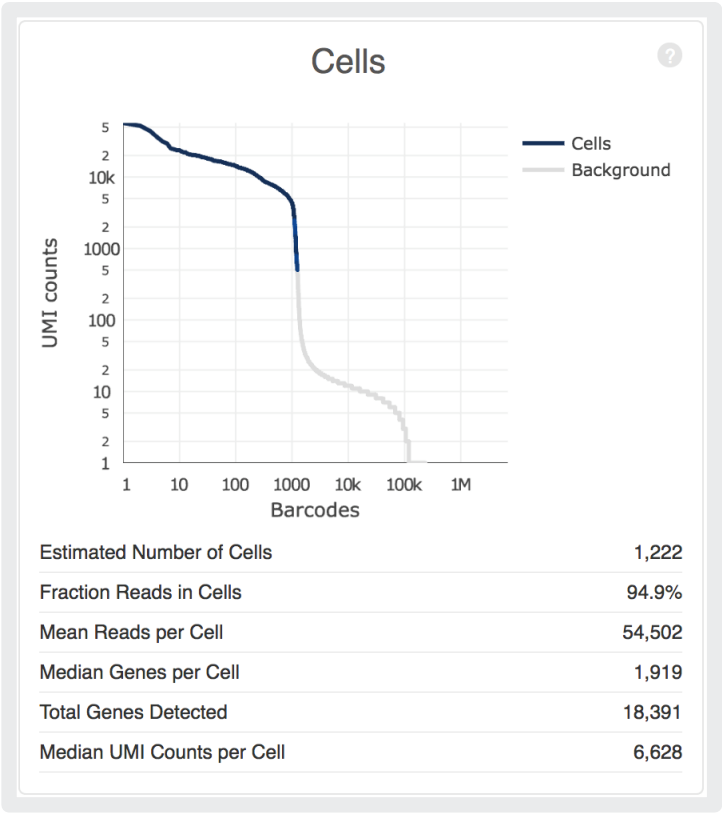
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics: Cell calling

Clog

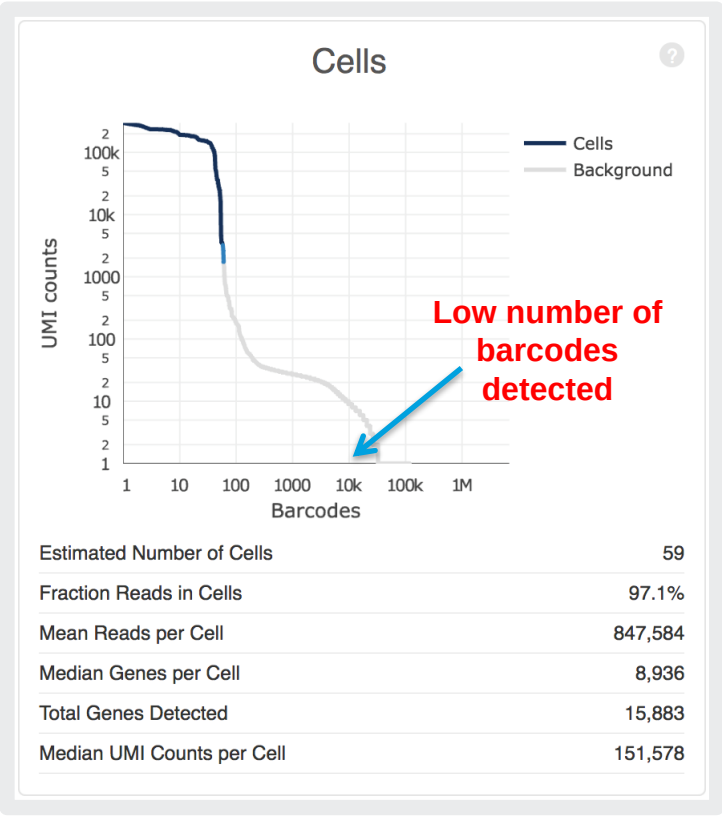
- A comparison of cell calling between a typical sample and a sample with a clog.
- Similar barcode rank profile can be seen for example for samples with clogs, low sequencing depth.
- The number of cells recovered will likely be lower than the targeted cell numbers. But the data for the cells called is usable.

Typical Sample



Estimated Number of Cells	1,222
Fraction Reads in Cells	94.9%
Mean Reads per Cell	54,502
Median Genes per Cell	1,919
Total Genes Detected	18,391
Median UMI Counts per Cell	6,628

Low Barcode counts (eg. Clog, low-depth)



Estimated Number of Cells	59
Fraction Reads in Cells	97.1%
Mean Reads per Cell	847,584
Median Genes per Cell	8,936
Total Genes Detected	15,883
Median UMI Counts per Cell	151,578

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics

Mapping

- Alert seen when Reads confidently mapped to transcriptome (<30%)
- Possible causes for low mapping rates:
  - Reads mapped to wrong genome or different strain
  - Very short RNA read length
  - Large stretches of no-calls in the reads
  - Custom reference contains overlapping genes

Mapping

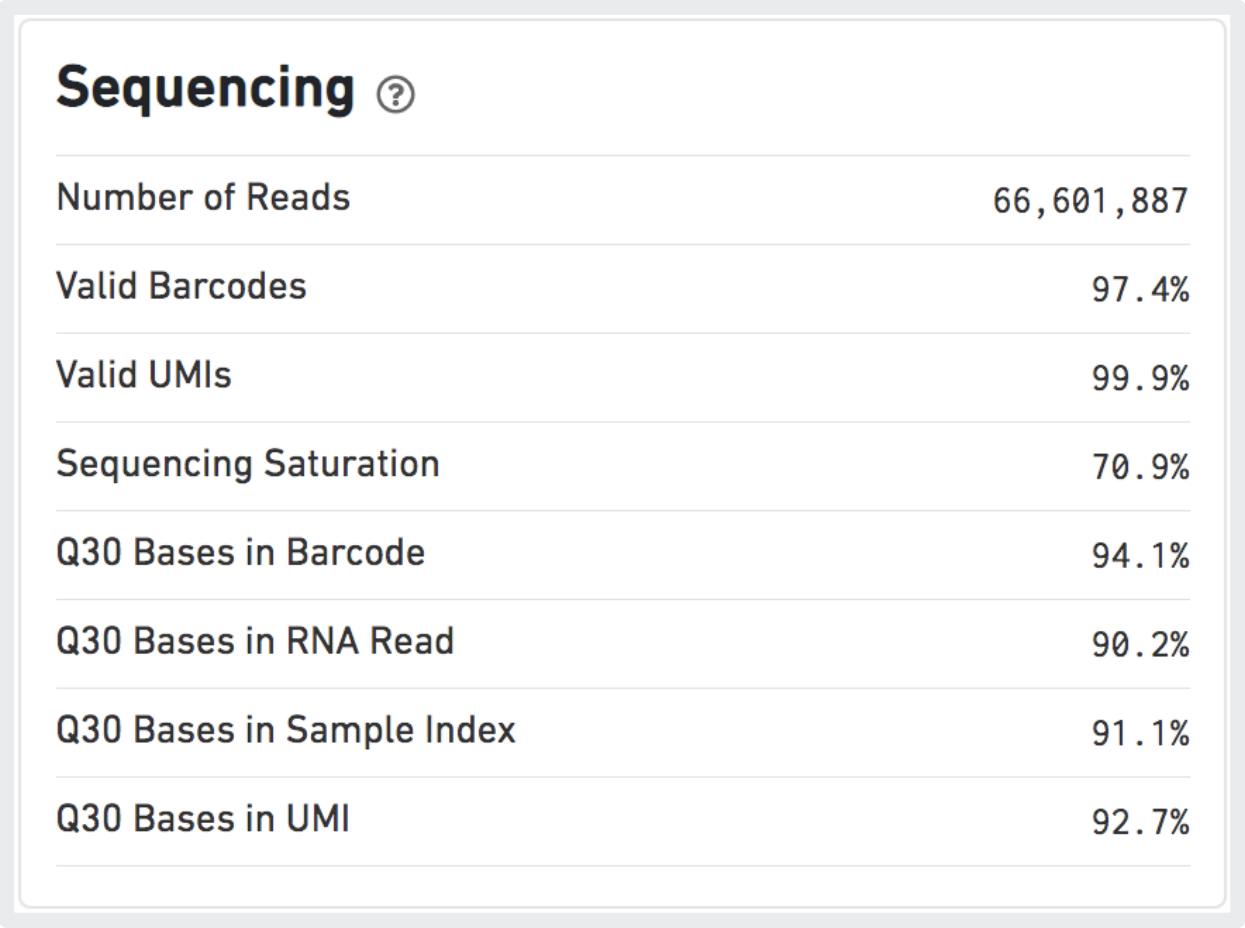
Reads Mapped to Genome	95.4%
Reads Mapped Confidently to Genome	92.4%
Reads Mapped Confidently to Intergenic Regions	4.8%
Reads Mapped Confidently to Intronic Regions	31.1%
Reads Mapped Confidently to Exonic Regions	56.5%
Reads Mapped Confidently to Transcriptome	53.7%
Reads Mapped Antisense to Gene	1.0%

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics

Sequencing

- Alert seen in web_summary.html when Q30 metrics are < 70%
- Possible causes for low RNA Read Q30 metrics:
  - Sequencing problems
  - Suboptimal loading concentration on sequencer
  - Note: RNA Read Q30 (R2) tends to be lower than the other reads.



Sequencing

Number of Reads	66,601,887
Valid Barcodes	97.4%
Valid UMIs	99.9%
Sequencing Saturation	70.9%
Q30 Bases in Barcode	94.1%
Q30 Bases in RNA Read	90.2%
Q30 Bases in Sample Index	91.1%
Q30 Bases in UMI	92.7%

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Additional Cell Ranger pipelines

Aggregate and reanalyze

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger Pipelines for Gene Expression

Overview

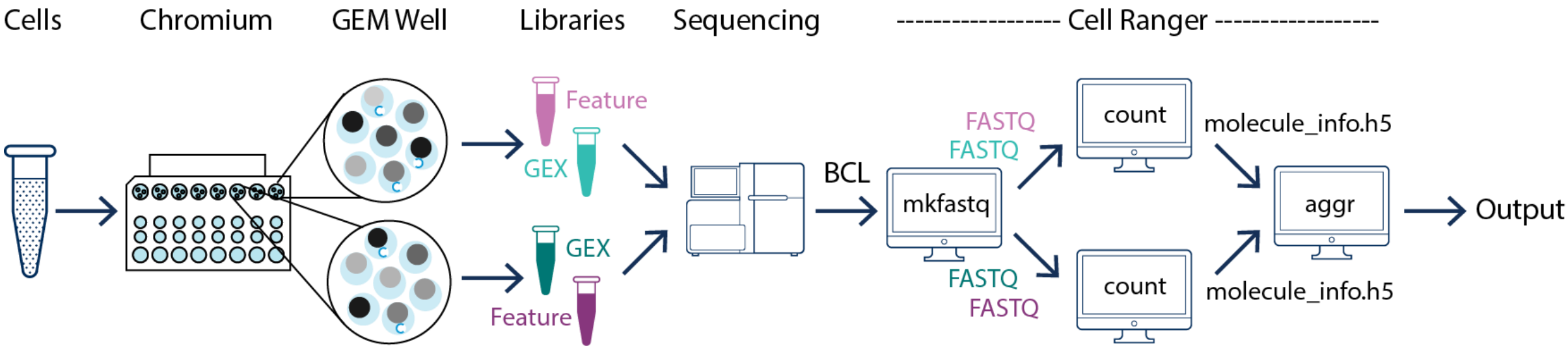
Pipeline	Functionality
cellranger mkfastq	Barcode-aware demultiplexing from BCL to FASTQ
cellranger count	• Read-level analysis of a single library • Produces feature/barcode matrix • Produces expression analysis and static visualizations • Produces .cloupe file for Loupe Cell Browser
cellranger aggr	• Aggregate `cellranger count` results • Combine data from multiple libraries and normalize depth • Optional chemistry correction for combining data between chemistries • Produces combined feature barcode matrix • Produces .cloupe file for Loupe Cell Browser
cellranger reanalyze	• Regenerate filtered matrix with custom parameters • Re-run the secondary analysis with custom parameters • Produces .cloupe file for Loupe Cell Browser

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Combining samples with “aggr” pipeline

Laboratory and analysis workflow overview



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Executing `cellranger aggr`

- Combines and normalizes libraries previously run with `cellranger count`
- Construct a Library Sheet pointing to the molecule_info.h5 output of `cellranger count`. Save as CSV file.

library_id	molecule_h5
AMLNormal1	/opt/runs/LV123/outs/molecule_info.h5
AMLNormal2	/opt/runs/LB456/outs/molecule_info.h5
AMLPatient	/opt/runs/LP789/outs/molecule_info.h5

```
$ cellranger aggr --sample=experiment1 --csv=library_sheet.csv
```

10x GENOMICS

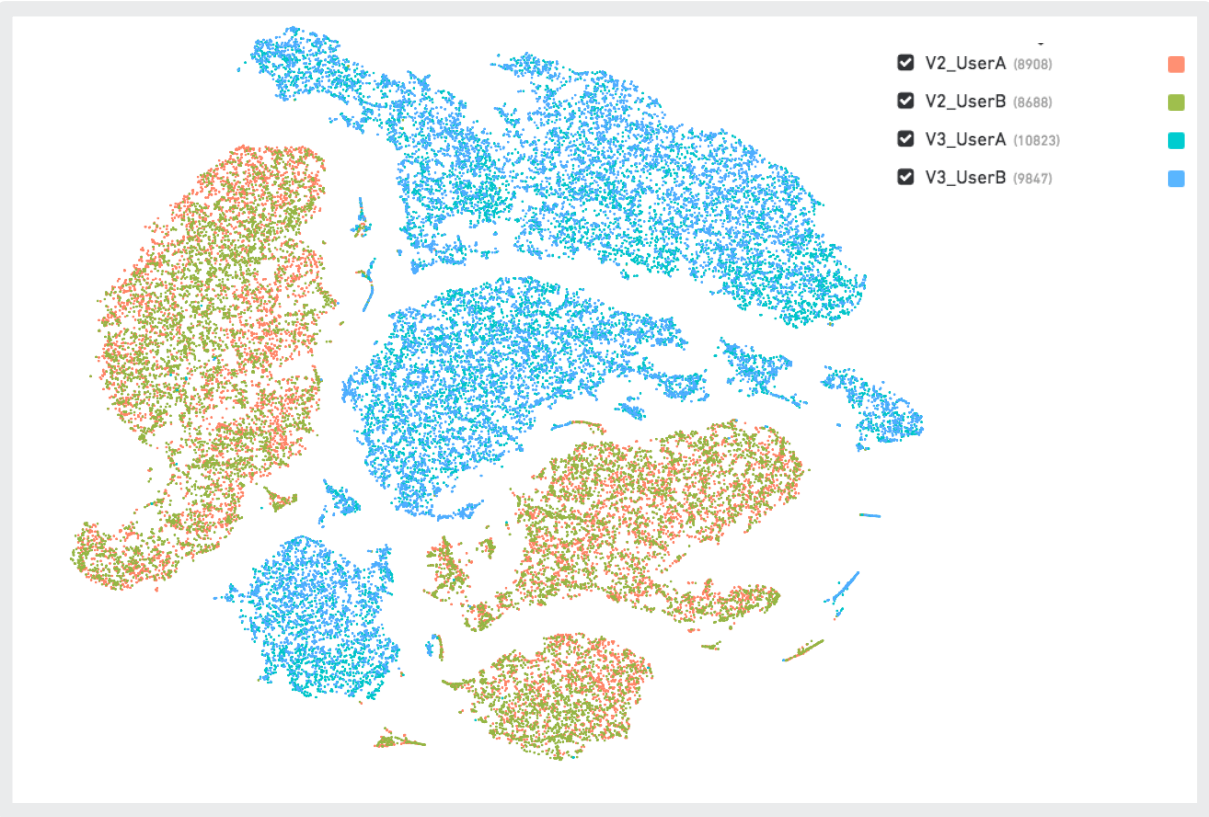
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Combining samples with “aggr” pipeline

Optional **Chemistry** batch effects correction

- With Cell Ranger 3.0 “aggr” pipeline, users can aggregate Single Cell 3’ Gene Expression v2 and v3 chemistry data
- Chemistry batch correction algorithm in “aggr” works in PCA space and affects t-SNE and cell clustering
- The UMI counts are not affected by chemistry batch correction
- Based on ideas from multiple published methods:
  - Mutual Nearest Neighbor (Haghverdi et al., 2018)
  - Scanorama (Hie et al., 2018)
  - Batch-Balanced kNN (Park et al., 2018)

10000 PBMCs – No Chemistry Correction

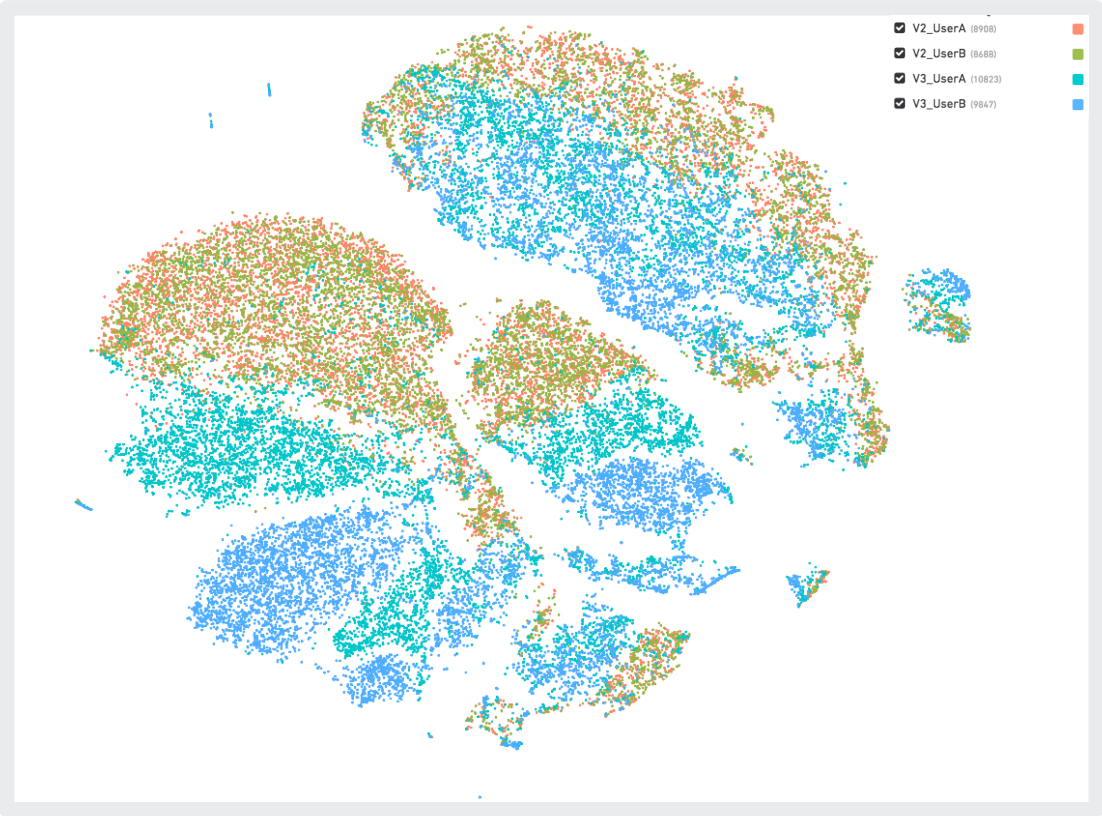


© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

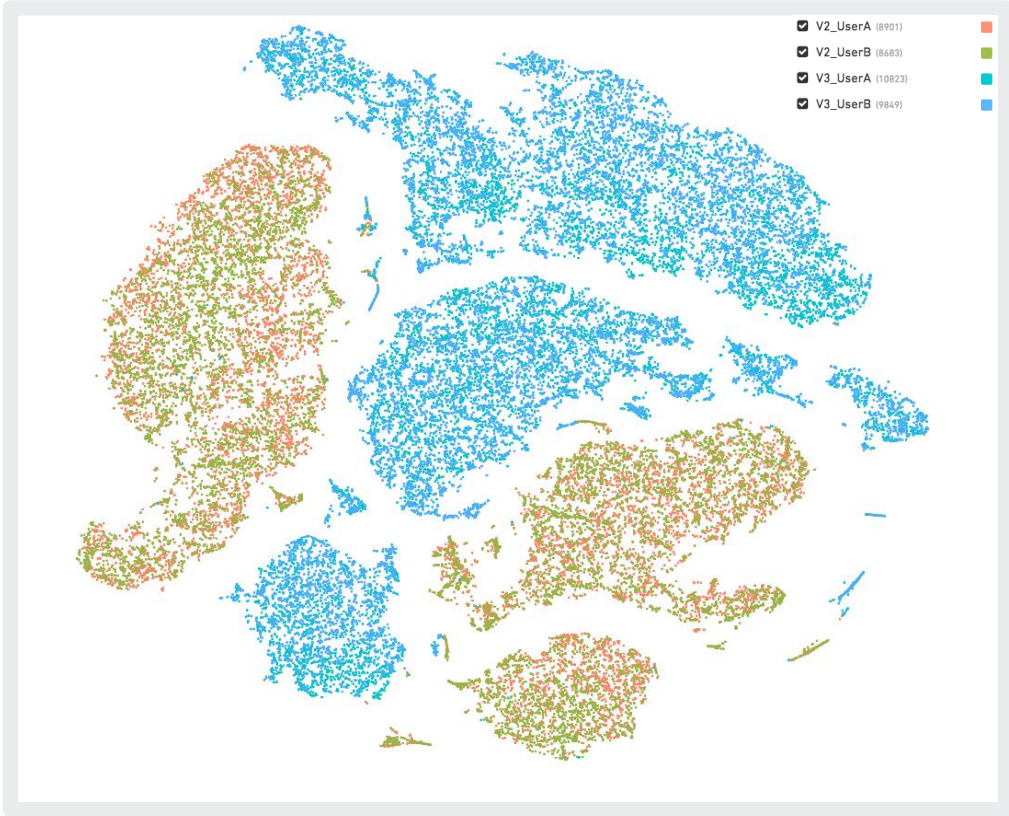
Aggregating 3’ v2 and v3 chemistry data

Incorporating chemistry batch correction reduces chemistry-specific clustering

10000 PBMCs – No Chemistry Correction



10000 PBMCs – With Chemistry Correction



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Executing `cellranger reanalyze`

- Allows users to customize advanced data analysis parameters executed by Cell Ranger
- Example use cases
  - Run secondary analysis on a subset of genes or barcodes
  - Increase number of principle components used for analysis
  - Change default t-SNE/UMAP parameters
  - Create 3-D t-SNE plots
  - User specified max K and graph-based cluster detection values

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

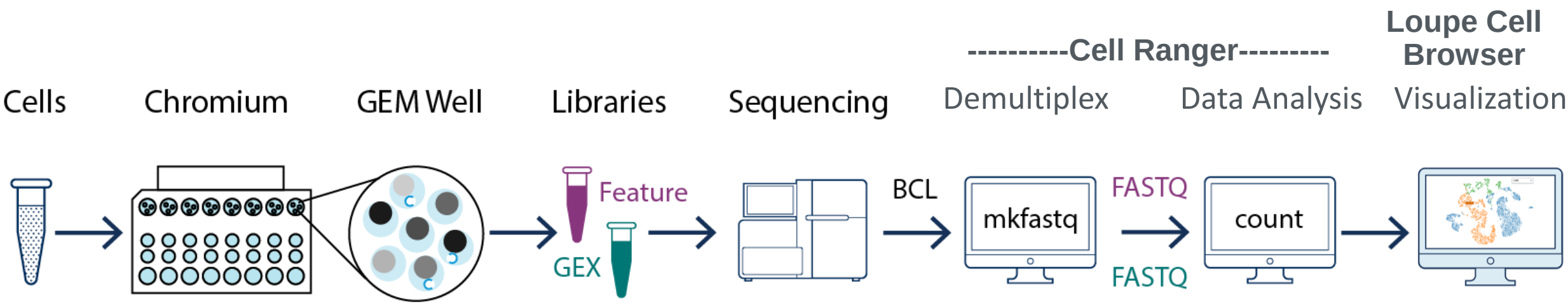
Cell Ranger “count” pipeline for feature libraries

How to run the pipeline

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Analyzing Gene Expression with Feature Barcode technology data

Laboratory and analysis workflow overview



Executing `cellranger mkfastq`

Demultiplexing

- Run the `cellranger mkfastq` pipeline to demultiplex an Illumina sequencing run folder into FASTQ files

```
$cellranger mkfastq --id=run1 \
                    --csv=my_samples.csv \
                    --run=/mnt/hiseq/run1_bcl
```

Output directory name

Sample sheet

Path to bcls

- Output

```
$ls -l /opt/bar/run1/outs/fastq_path
drwxr-xr-x 3 jdoe jdoe           3 Sep 24 19:19 HAWT7ADXX
drwxr-xr-x 3 jdoe jdoe           3 Sep 24 19:17 Reports
drwxr-xr-x 2 jdoe jdoe          10 Sep 24 19:17 Stats
-rw-r--r-- 1 jdoe jdoe    27971547 Sep 24 18:07 Undetermined_S0_L001_I1_001.fastq.gz
-rw-r--r-- 1 jdoe jdoe    44432222 Sep 24 18:07 Undetermined_S0_L001_R1_001.fastq.gz
-rw-r--r-- 1 jdoe jdoe   912086076 Sep 24 18:07 Undetermined_S0_L001_R2_001.fastq.gz

$tree opt/bar/run1/outs/fastq_path/HAWT7ADXX/sample1/
opt/bar/sample1/outs/fastq_path/HAWT7ADXX/sample1
|── sample1_S1_L001_I1_001.fastq.gz
|── sample1_S1_L001_R1_001.fastq.gz
|── sample1_S1_L002_R2_001.fastq.gz
```

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Executing “cellranger count”

Gene Expression + Feature Barcode technology libraries

- Using the FASTQ folder from `cellranger mkfastq`:

```
$ cellranger count --id=sample1 \
                   --libraries=library_def.csv \
                   --transcriptome=refdata-cellranger/GRCh38 \
                   --feature-ref=my_features.csv
```

Output directory name

Input fastq files

Transcriptome reference

Feature reference

- A successful run concludes with a message similar to this:

```
Outputs:
- Run summary HTML:                         /opt/sample1/outs/web_summary.html
- Run summary CSV:                          /opt/sample1/outs/metrics_summary.csv
- BAM:                                      /opt/sample1/outs/possorted_genome_bam.bam
- BAM index:                                /opt/sample1/outs/possorted_genome_bam.bam.bai
- Filtered feature-barcode matrices MEX:    /opt/sample1/outs/filtered_feature_bc_matrix
- Filtered feature-barcode matrices HDF5:   /opt/sample1/outs/filtered_feature_bc_matrix.h5
- Unfiltered feature-barcode matrices MEX:  /opt/sample1/outs/raw_feature_bc_matrix
- Unfiltered feature-barcode matrices HDF5: /opt/sample1/outs/raw_feature_bc_matrix.h5
- Secondary analysis output CSV:            /opt/sample1/outs/analysis
- Per-molecule read information:            /opt/sample1/outs/molecule_info.h5
- CRISPR-specific analysis:                 null
- Loupe Cell Browser file:                  /opt/sample1/outs/cloupe.cloupe

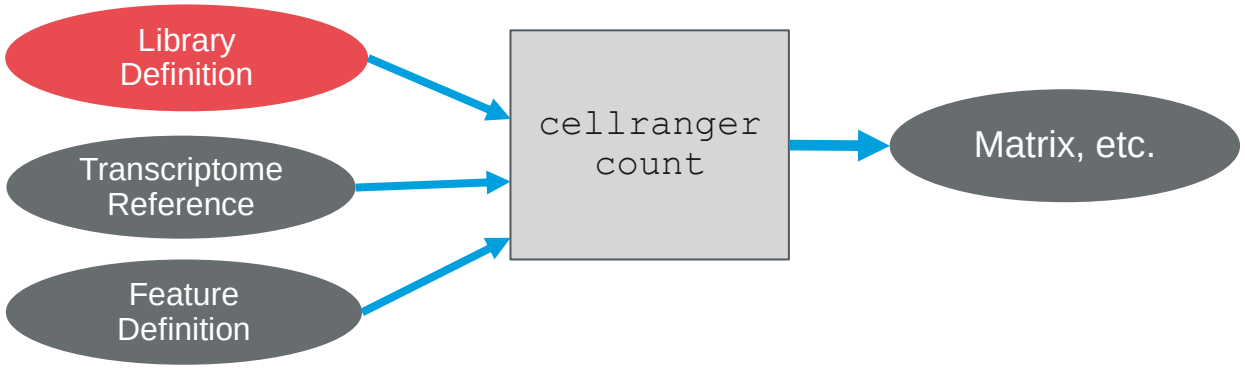
Waiting 6 seconds for UI to do final refresh.
Pipestance completed successfully!
```

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Input fastqs

Library definition file



- CSV file with paths to FASTQs and column specifying whether the library is GEX or Feature Barcode.

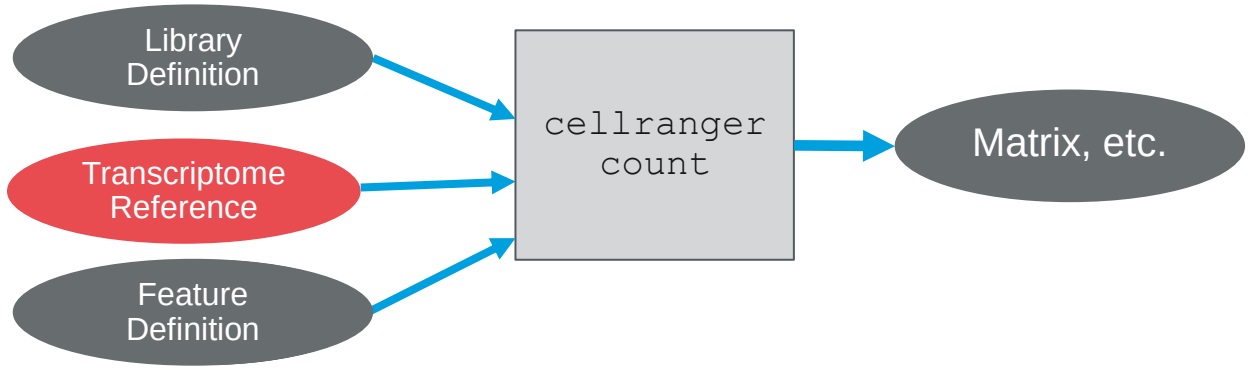
```
opt/bar/run1/
|── sample1_gex_S1_L001_I1_001.fastq.gz
|── sample1_gex_S1_L001_R1_001.fastq.gz
|── sample1_gex_S1_L001_R2_001.fastq.gz
|── sample1_fbc_S1_L001_I1_001.fastq.gz
|── sample1_fbc_S1_L001_R1_001.fastq.gz
|── sample1_fbc_S1_L001_R2_001.fastq.gz
```

fastqs	sample	library_type*
/opt/bar/run1	sample1_gex	Gene Expression
/opt/bar/run1	sample1_fbc	Antibody Capture

*Values used in library_type field needs to match one of “Gene Expression”, “Antibody Capture”, “CRISPR Guide Capture”, “Custom”

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

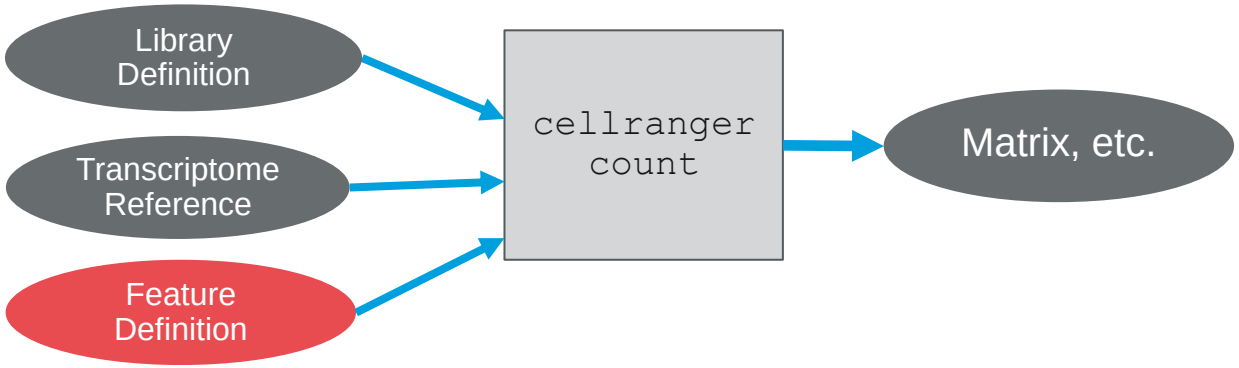
Reference files



- Transcriptome Reference
  - Reference used for gene expression analysis
  - 10x pre-built references available for human and mouse genomes
  - Custom references can be built using ”mkref” utility

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Reference files

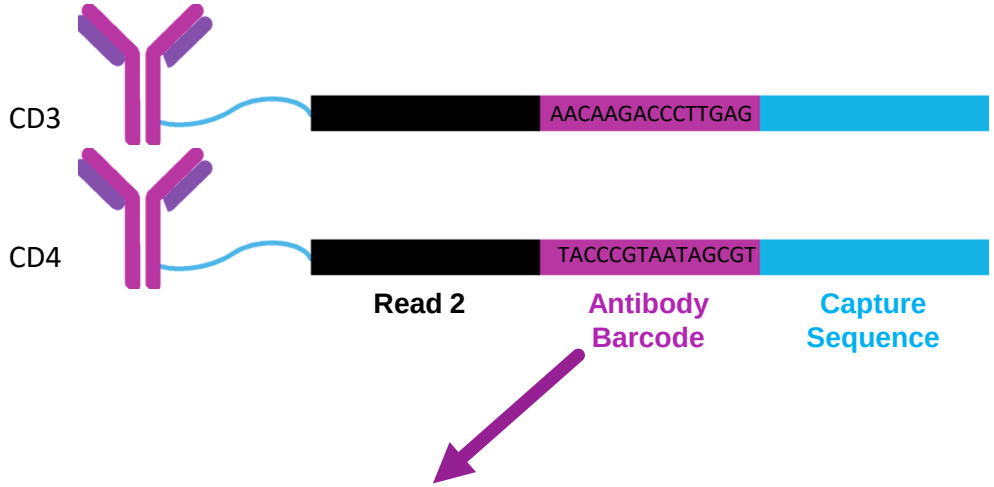


- Feature Definition File
  - CSV file that specifies the feature reference
  - Which feature barcodes were used
  - Where to find the feature barcodes on the reads

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Feature definition file

Cell surface protein libraries



A ‘minimal’ custom oligonucleotide for antibody conjugation may have this pattern:

id	name	read	pattern	sequence	feature_type
CD3	CD3_Custom	R2	5P(BC)	AACAAGACCCTTGAG	Antibody Capture
CD4	CD4_Custom	R2	5P(BC)	TACCCGTAATAGCGT	Antibody Capture

Antibody panels from BioLegend are designed as (for example)

NNNNNNNNNNAACAAGACCCTTGAGNNNNNNNNN

id	name	read	pattern	sequence	feature_type
CD3	CD3_TotalSeq	R2	5PNNNNNNNNNN(BC)	AACAAGACCCTTGAG	Antibody Capture
CD4	CD4_TotalSeq	R2	5PNNNNNNNNNN(BC)	TACCCGTAATAGCGT	Antibody Capture

Example files on https://support.10xgenomics.com/

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

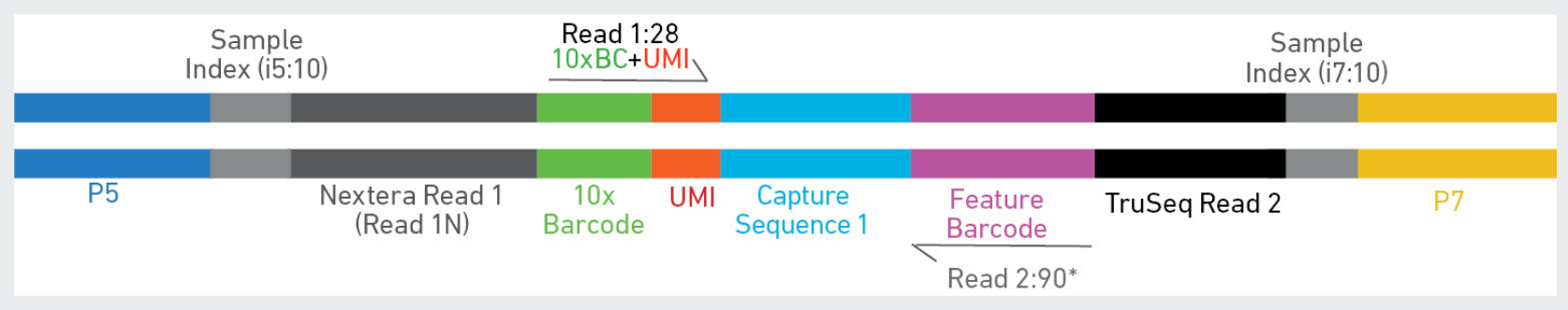
Pipeline steps and outputs

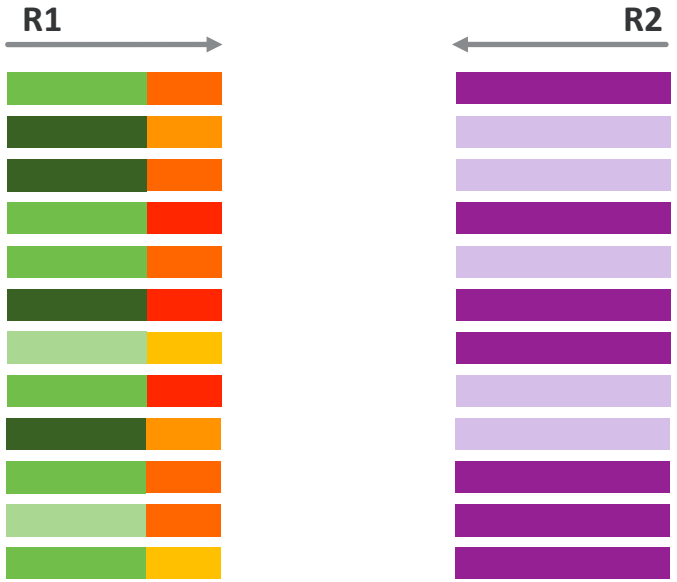
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

Pipeline

- **Read Processing**
- Align
- Filter Barcodes, UMIs
- Count UMIs
- Filter Aggregates
- Call Cells
- Secondary Analysis





© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

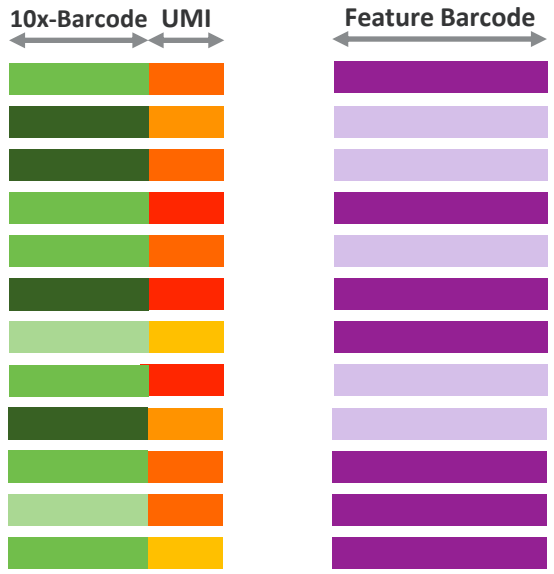
Cell Ranger “count” for cell surface protein

Pipeline

- Read Processing
- **Align**
- Filter Barcodes, UMIs
- Count UMIs
- Filter Aggregates
- Call Cells
- Secondary Analysis

id	name	read	pattern	sequence	feature_type
CD3	CD3_TotSeqB	R2	P5NNNNNNNNNN(BC)	AACAAGACCCTTGAG	Antibody Capture
CD8a	CD8a_TotSeqB	R2	P5NNNNNNNNNN(BC)	ATTGGCACTCAGATG	Antibody Capture

Feature Reference --> CD3 CD8a



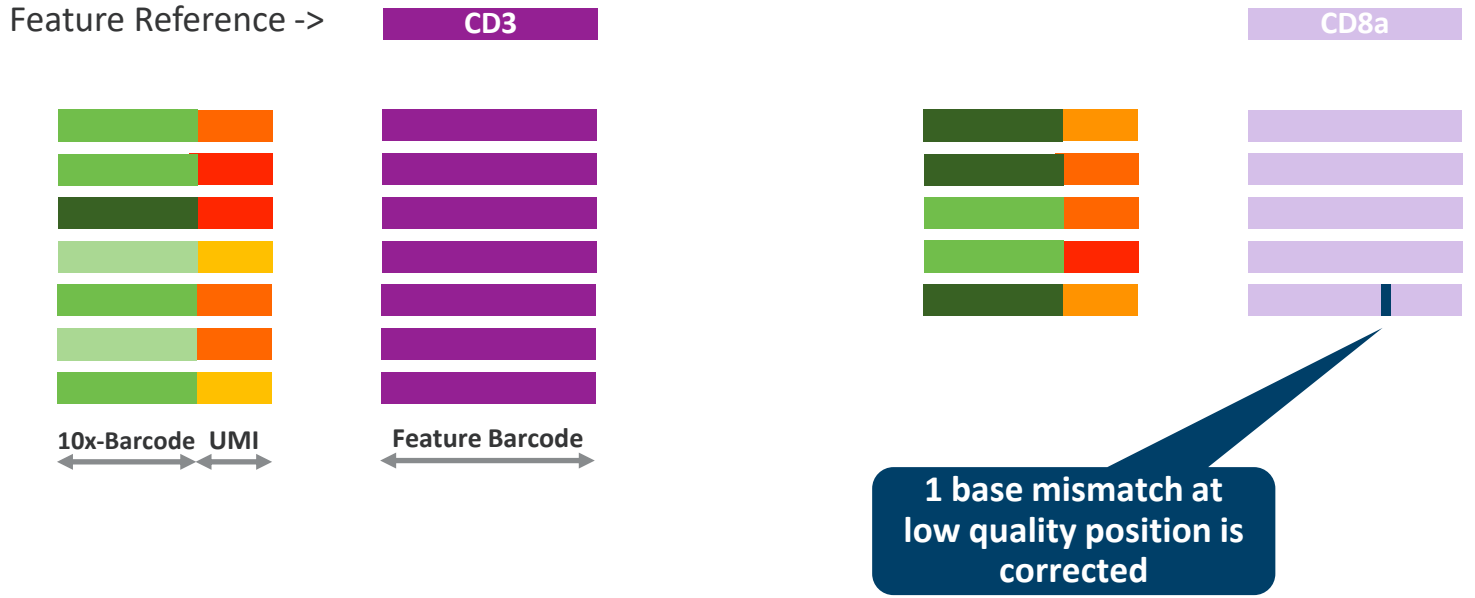
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

Pipeline

- Read Processing
- **Align**
- Filter Barcodes, UMIs
- Count UMIs
- Filter Aggregates
- Call Cells
- Secondary Analysis

id	name	read	pattern	sequence	feature_type
CD3	CD3_TotSeq	R2	5PNNNNNNNNNN(BC)	AACAAGACCCTTGAG	Antibody Capture
CD8a	CD8a_TotSeq	R2	5PNNNNNNNNNN(BC)	ATTGGCACTCAGATG	Antibody Capture



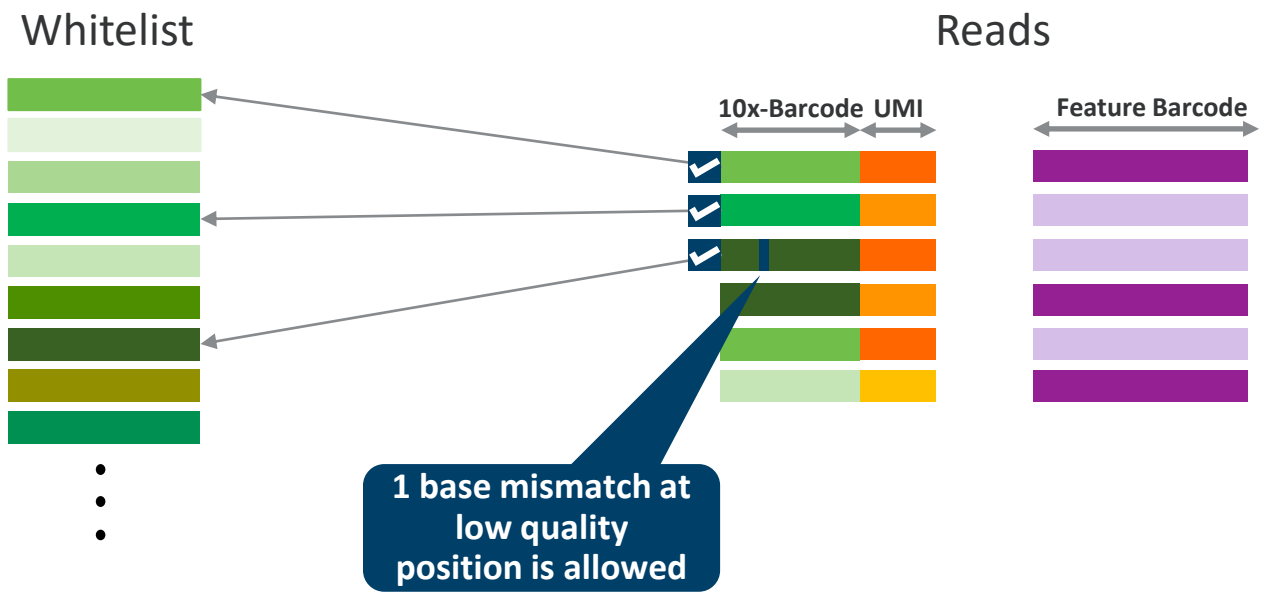
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

Pipeline

- Read Processing
- Align
- **Filter Barcodes, UMIs**
- Count UMIs
- Filter Aggregates
- Call Cells
- Secondary Analysis

- Cell barcodes
  - Must be on static list of known cell barcode sequences
  - May be 1 mismatch away from the list if the mismatch occurs at a low-quality position (the barcode is then corrected)



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

Pipeline

- Read Processing
- Align
- **Filter Barcodes, UMIs**
- Count UMIs
- Filter Aggregates
- Call Cells
- Secondary Analysis

- Filter UMIs
  - Must not be a homopolymer, e.g. AAAAAAAAAA
  - Must not contain N
  - Must not contain bases with base quality < 10

- Correct UMIs
  - UMIs that are 1 mismatch away from a higher count UMI are corrected to that UMI if they share a cell barcode and feature.



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

Pipeline

- Read Processing
- Align
- Filter Barcodes, UMIs
- **Count UMIs**
- Filter Aggregates
- Call Cells
- Secondary Analysis



Feature Barcode Matrix

	CD3	CD8a
Barcode1	3	2
Barcode2	2	0
Barcode3	1	1

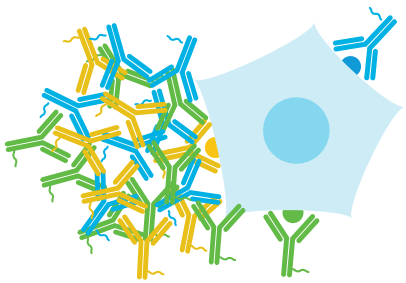
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

Pipeline

- Read Processing
- Align
- Filter Barcodes, UMIs
- Count UMIs
- **Filter Aggregates**
- Call Cells
- Secondary Analysis

- Antibody aggregates can associate with cells



- Very high UMI counts seen for these cells
- UMI counts are not really meaningful
- This step aims to filter out such barcodes

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger “count” for cell surface protein

Pipeline

- Read Processing
- Align
- Filter Barcodes, UMIs
- Count UMIs
- Filter Aggregates
- **Call Cells**
- Secondary Analysis

- When pipeline is run in Gene Expression + Cell surface protein mode:
  - Pipeline uses transcriptome data for cell calling
  - Distinguishes cell containing GEMs/barcodes from background barcodes

- When the pipeline is run with only cell surface protein data, cell calling is done
  - Pipeline uses total cell surface protein levels for cell calling
  - This step aims to identify cells stained with antibodies in the panel from all other cell types and background
  - The antibody panel used has a big effect on the number of called cells

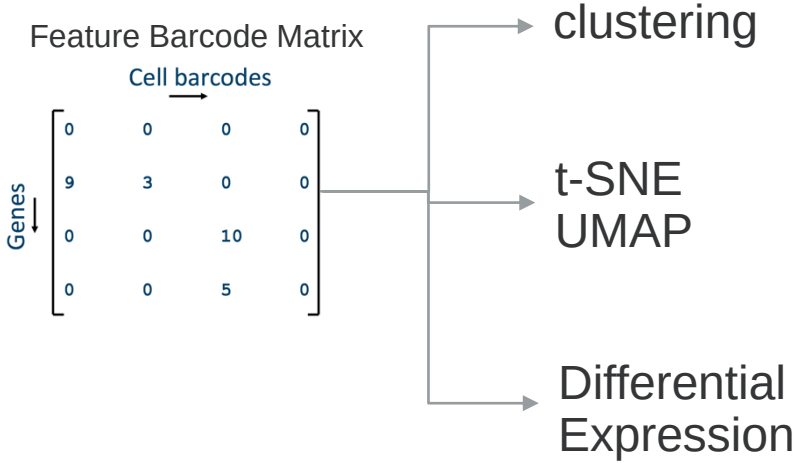
10x GENOMICS

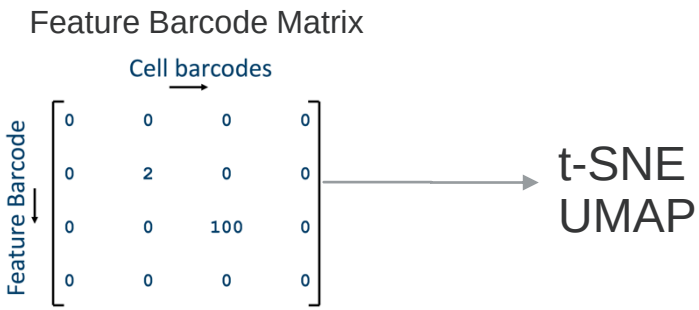
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

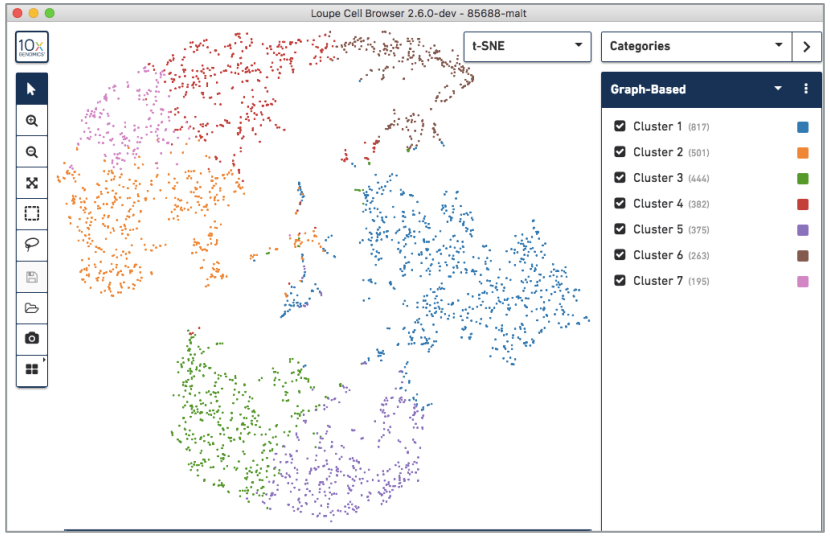
Cell Ranger “count” for cell surface protein

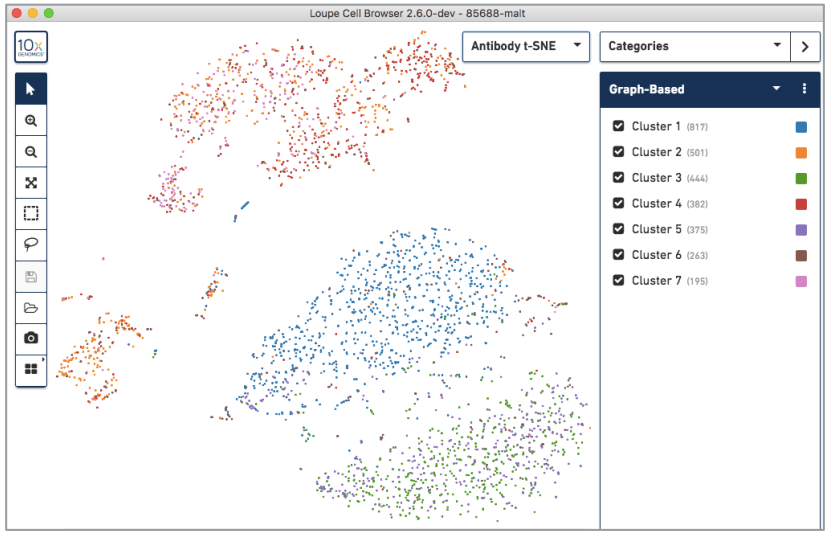
Pipeline

- Read Processing
- Align
- Filter Barcodes, UMIs
- Count UMIs
- Filter Aggregates
- Call Cells
- **Secondary Analysis**









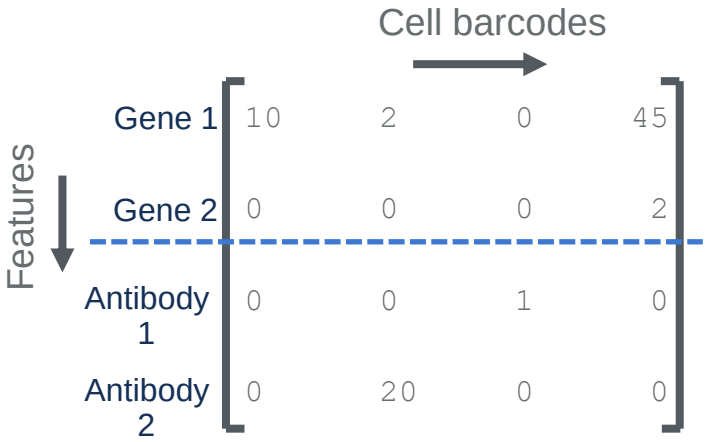
© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Outputs for cell surface protein

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell ranger “count” output files for cell surface protein analysis

- Feature Barcode Matrix
  - All Cell Surface Protein feature counts are output alongside gene counts in the feature-barcode matrix
- QC Metrics for cell surface protein library
  - Sequencing and application metrics displayed in a single combined web summary with Gene Expression library metrics
  - Referred to as “Antibody”: “Antibody Sequencing” and “Antibody Application” tabs
- .cloupe file for data visualization and interactive analysis in Loupe Cell Browser
  - Cell Surface Protein feature counts are combined with Gene Expression data into a single .cloupe file per sample
- t-SNE/UMAP on raw antibody counts to visualize cells in a 2-D space
  - The t-SNE and UMAP are viewable in Loupe Cell Browser



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics for cell surface protein

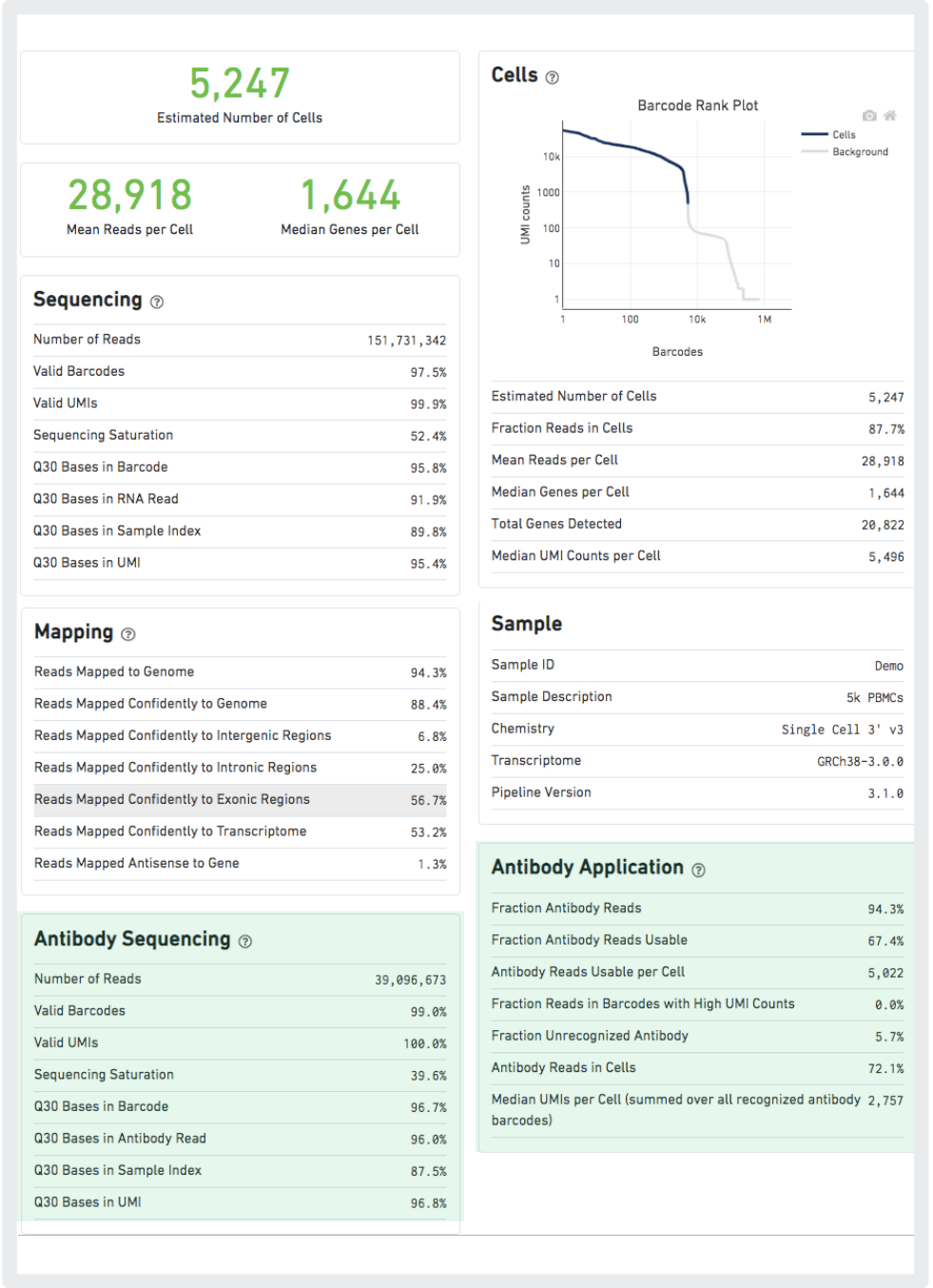
web_summary.html

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics: Cell surface protein

Overview

- One web_summary.html file
- Summary tab: includes metrics for
  - Gene expression library
  - Cell Surface Protein Feature library (Referred to as “Antibody” for metric reporting)
- Antibody metrics in “Summary” tab only



5,247
Estimated Number of Cells

28,918
Mean Reads per Cell

1,644
Median Genes per Cell

Sequencing

Number of Reads	151,731,342
Valid Barcodes	97.5%
Valid UMIs	99.9%
Sequencing Saturation	52.4%
Q30 Bases in Barcode	95.8%
Q30 Bases in RNA Read	91.9%
Q30 Bases in Sample Index	89.8%
Q30 Bases in UMI	95.4%

Mapping

Reads Mapped to Genome	94.3%
Reads Mapped Confidently to Genome	88.4%
Reads Mapped Confidently to Intergenic Regions	6.8%
Reads Mapped Confidently to Intronic Regions	25.0%
Reads Mapped Confidently to Exonic Regions	56.7%
Reads Mapped Confidently to Transcriptome	53.2%
Reads Mapped Antisense to Gene	1.3%

Antibody Sequencing

Number of Reads	39,096,673
Valid Barcodes	99.0%
Valid UMIs	100.0%
Sequencing Saturation	39.6%
Q30 Bases in Barcode	96.7%
Q30 Bases in Antibody Read	96.0%
Q30 Bases in Sample Index	87.5%
Q30 Bases in UMI	96.8%

Cells

Estimated Number of Cells	5,247
Fraction Reads in Cells	87.7%
Mean Reads per Cell	28,918
Median Genes per Cell	1,644
Total Genes Detected	20,822
Median UMI Counts per Cell	5,496

Sample

Sample ID	Demo
Sample Description	5k PBMCs
Chemistry	Single Cell 3' v3
Transcriptome	GRCh38-3.0.0
Pipeline Version	3.1.0

Antibody Application

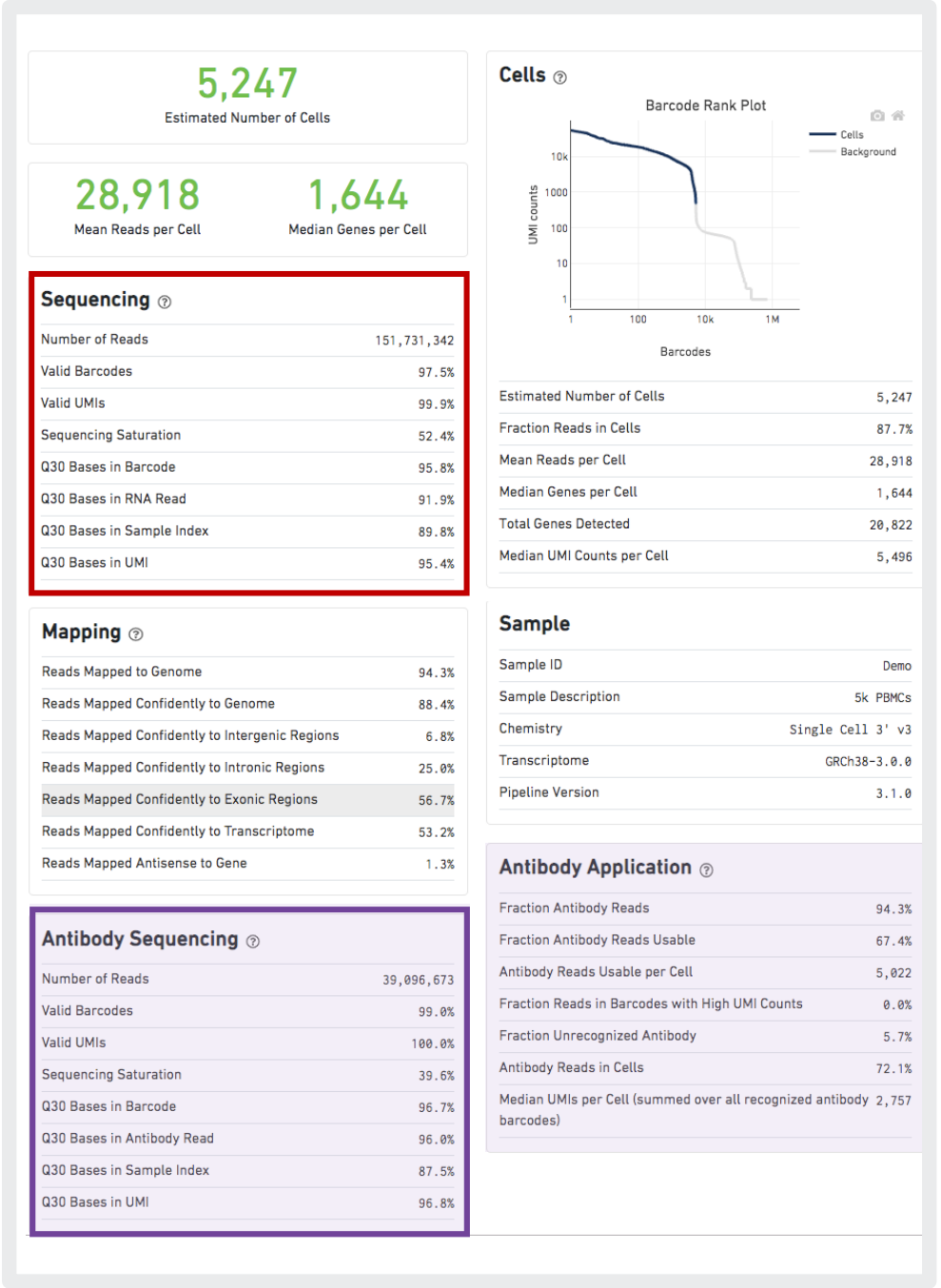
Fraction Antibody Reads	94.3%
Fraction Antibody Reads Usable	67.4%
Antibody Reads Usable per Cell	5,022
Fraction Reads in Barcodes with High UMI Counts	0.0%
Fraction Unrecognized Antibody	5.7%
Antibody Reads in Cells	72.1%
Median UMIs per Cell (summed over all recognized antibody barcodes)	2,757

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics: Cell surface protein

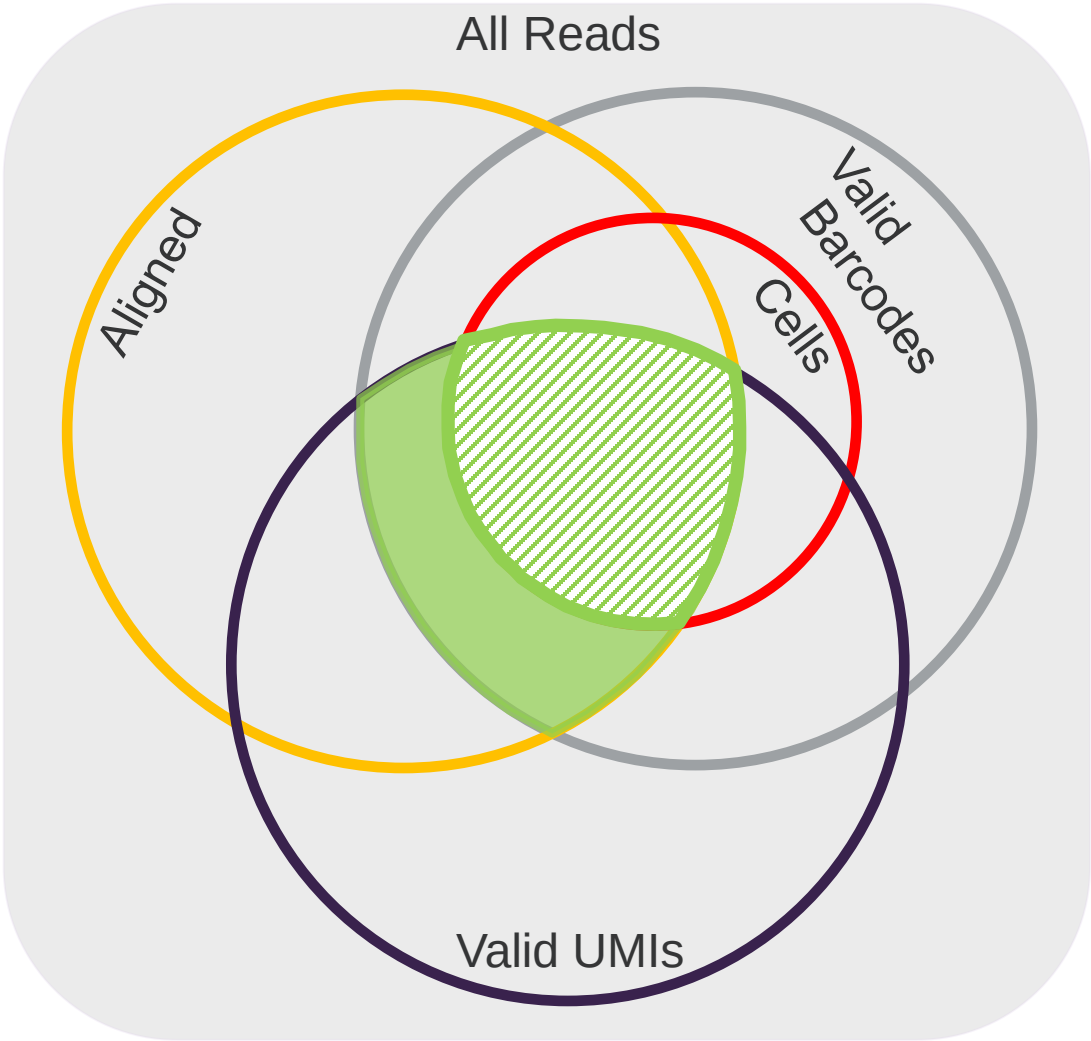
“Antibody sequencing” metrics

- Antibody: Number of Reads
  - Total number of reads for Cell Surface Protein library
- Antibody: Mean Reads per Cell
  - Mean sequencing depth per cell
- Antibody: Valid Barcodes
  - Fraction of reads with a barcode found in or corrected to one that is found in the whitelist
- Antibody: Sequencing Saturation
  - The fraction of reads originating from an already-observed UMI



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Filtering reads: Cell surface protein

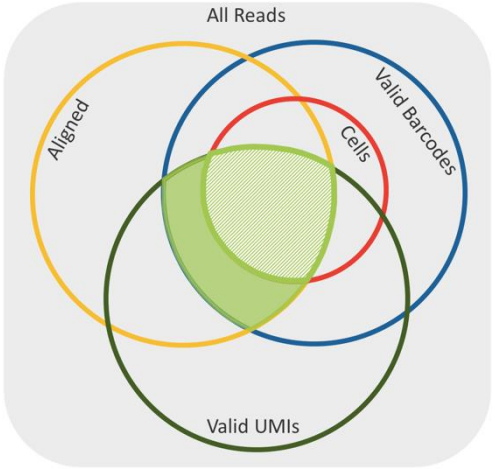


© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics: Cell surface protein

“Antibody application” metrics

- Antibody: Antibody Reads in Cells



Fraction Antibody Reads in Cells:

$$\frac{\text{Reads in}}{\text{Reads in}}$$

  - What percent of reads that meet below criteria are in cells
    - Alignment to feature reference
    - Valid 10x barcode
    - Valid UMI
  - Possible causes for low value:
    - Poor cell washing
    - Low target population, for example if rare markers were used
  - > 50% is considered okay

Antibody Application

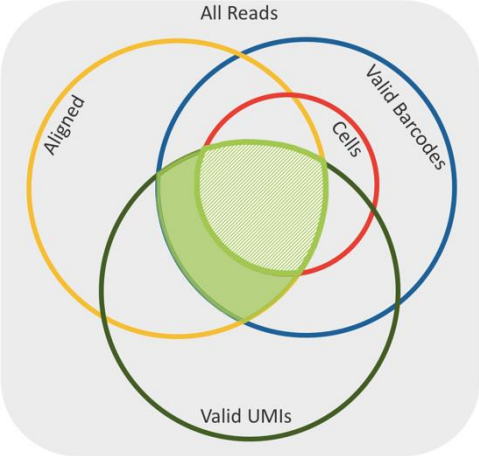
Fraction Antibody Reads	94.3%
Fraction Antibody Reads Usable	67.4%
Antibody Reads Usable per Cell	5,022
Fraction Reads in Barcodes with High UMI Counts	0.0%
Fraction Unrecognized Antibody	5.7%
Antibody Reads in Cells	**72.1%**
Median UMIs per Cell (summed over all recognized antibody barcodes)	2,757

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics: Cell surface protein

“Antibody application” metrics

- Antibody: Fraction Antibody Reads Usable



Fraction Antibody Reads Usable:

Reads in 

All Reads,i.e.,

  - What percent of all input reads are used for counting in filtered feature barcode matrix
  - > 20% or 1000 usable reads per cell is fine

Antibody Application

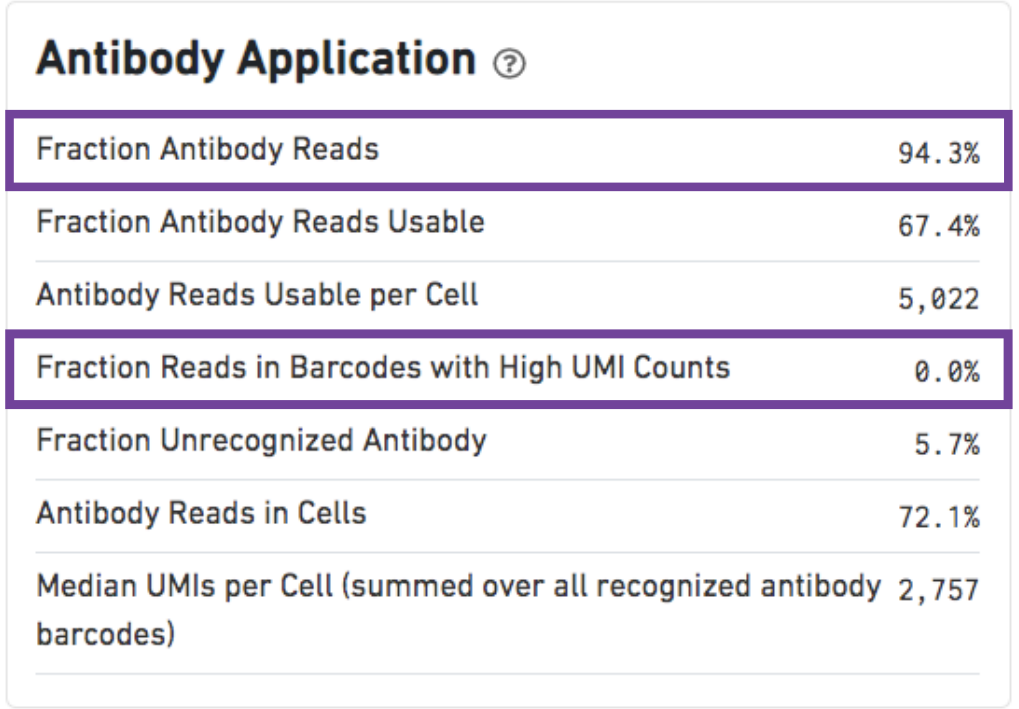
Fraction Antibody Reads	94.3%
Fraction Antibody Reads Usable	67.4%
Antibody Reads Usable per Cell	5,022
Fraction Reads in Barcodes with High UMI Counts	0.0%
Fraction Unrecognized Antibody	5.7%
Antibody Reads in Cells	72.1%
Median UMIs per Cell (summed over all recognized antibody barcodes)	2,757

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

QC metrics: Cell surface protein

“Antibody application” metrics

- Antibody: Fraction Antibody Reads
  - Percent reads aligned to feature reference
  - Poor mapping rate could be due to incorrect feature barcode sequence in feature reference
- Antibody: Fraction Reads in Barcodes with High UMI counts
  - Potential protein aggregates, antibody saturation
  - Warning at > 5%, alarm at > 50%



Antibody Application

Fraction Antibody Reads	94.3%
Fraction Antibody Reads Usable	67.4%
Antibody Reads Usable per Cell	5,022
Fraction Reads in Barcodes with High UMI Counts	0.0%
Fraction Unrecognized Antibody	5.7%
Antibody Reads in Cells	72.1%
Median UMIs per Cell (summed over all recognized antibody barcodes)	2,757

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Additional resources

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger

Troubleshooting

- Cell Ranger Pipeline automatically produces a diagnostic file

- Send us the diagnostic file for
  - Test
  - Training run
  - Troubleshooting failed runs
  - Successful runs

- Send it to us directly from command line:

```
$ cellranger upload your@email.edu SampleA/SampleA.mri.tgz
```

10x GENOMICS

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

Cell Ranger

Support

Online Documentation:

http://support.10xgenomics.com

Q&A Knowledgebase:

https://kb.10xgenomics.com/hc/en-us

Please send questions, comments, and feedback to:

support@10xgenomics.com

© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.



© 10X GENOMICS, INC. 2020 | FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.