

UNIVERSITY OF
CAMBRIDGE



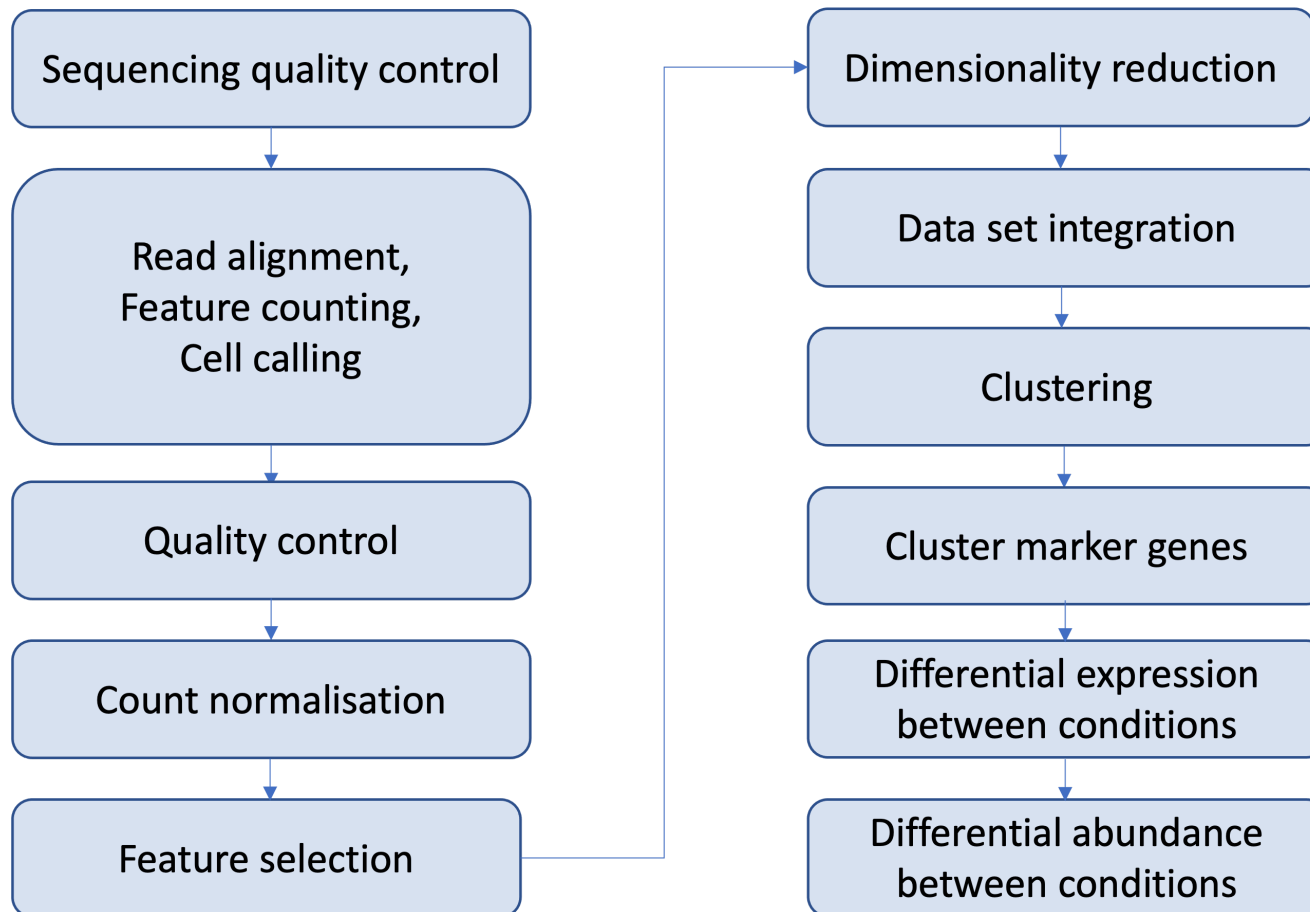
CANCER
RESEARCH
UK

Cambridge
Institute

Alignment and feature counting

April 2026

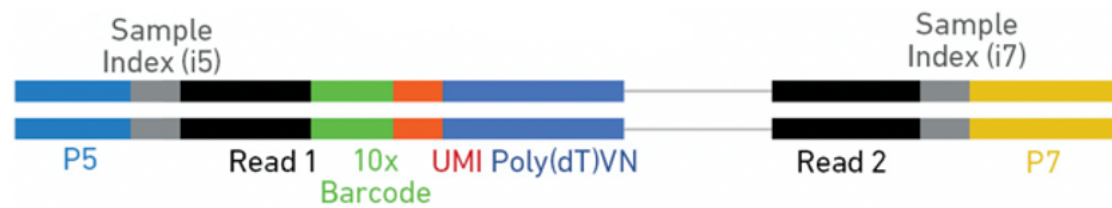
Single Cell RNAseq Analysis Workflow



10x library file structure

The 10x library contains four pieces of information, in the form of DNA sequences, for each “read”.

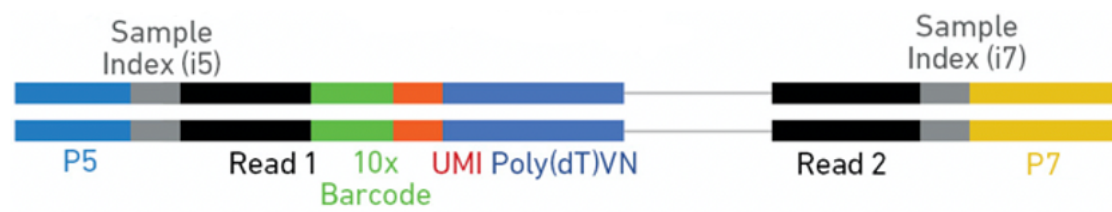
- **sample index** - identifies the library, with one or two indexes per sample
- **10x barcode** - identifies the droplet in the library
- **UMI** - identifies the transcript molecule within a cell and gene
- **insert** - the transcript molecule



Raw fastq files

The sequences for any given fragment will generally be delivered in 3 or 4 files:

- I1: I7 sample index
- I2: I5 sample index if present (dual indexing only)
- R1: 10x barcode + UMI
- R2: insert sequence



QC of Raw Reads - FASTQC

FastQC Report

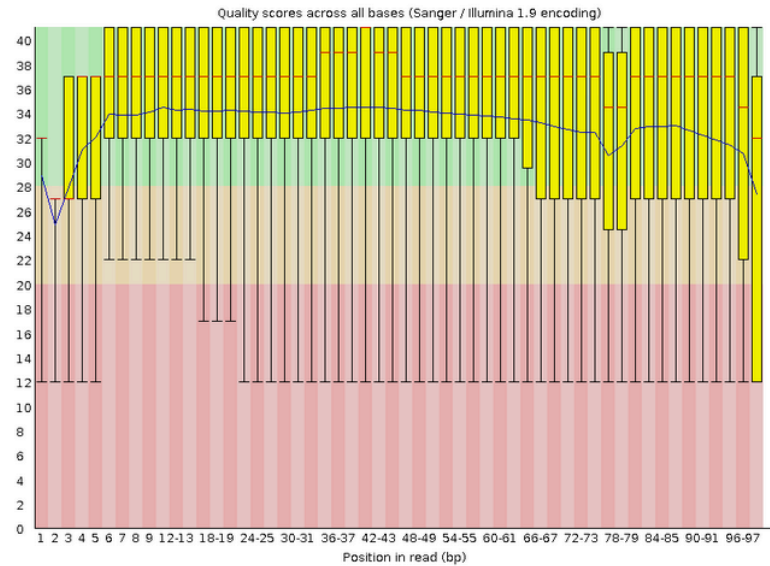
Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✗ Per tile sequence quality
- ✓ Per sequence quality scores
- ! Per base sequence content
- ! Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✓ Overrepresented sequences
- ✓ Adapter Content

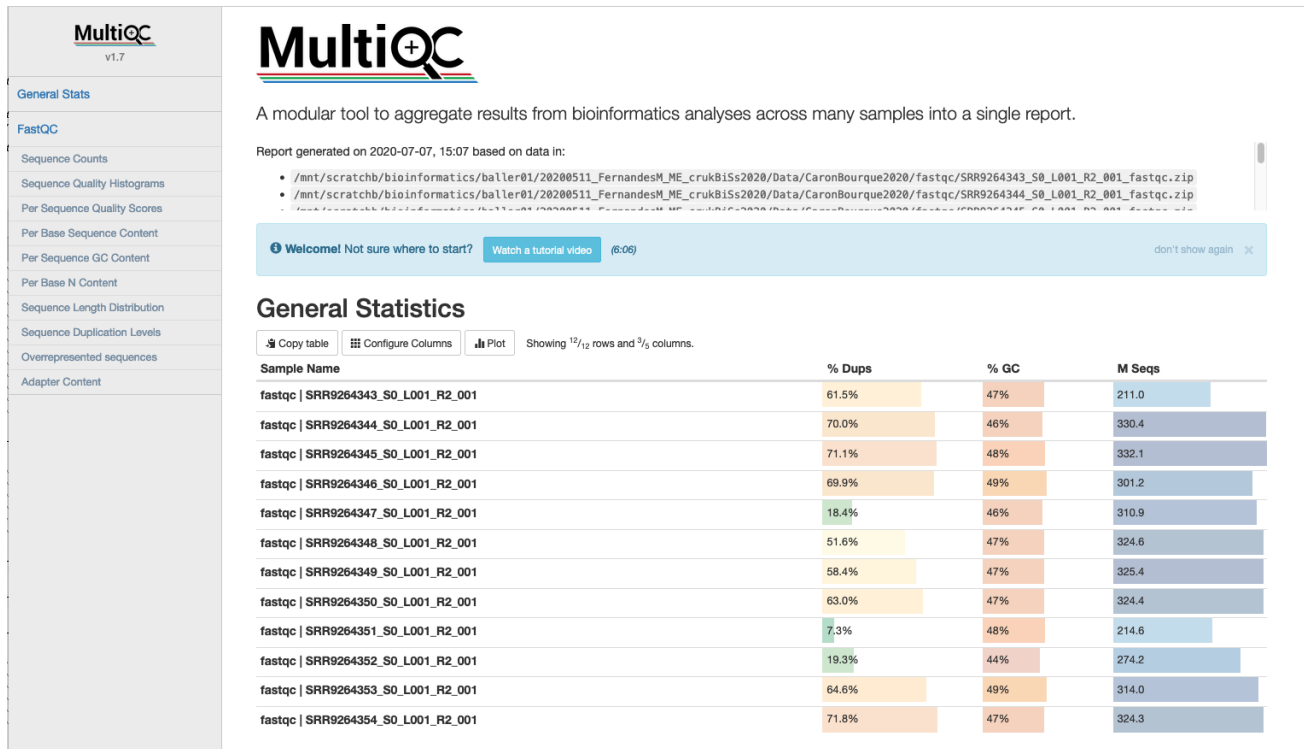
Basic Statistics

| Measure | Value |
|-----------------------------------|------------------------------------|
| Filename | SRR9264344_S0_L001_R2_001.fastq.gz |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 330404706 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 98 |
| %GC | 46 |

Per base sequence quality



QC of Raw Reads - MultiQC - General Statistics



QC of Raw Reads - MultiQC - Sequence Quality Histograms

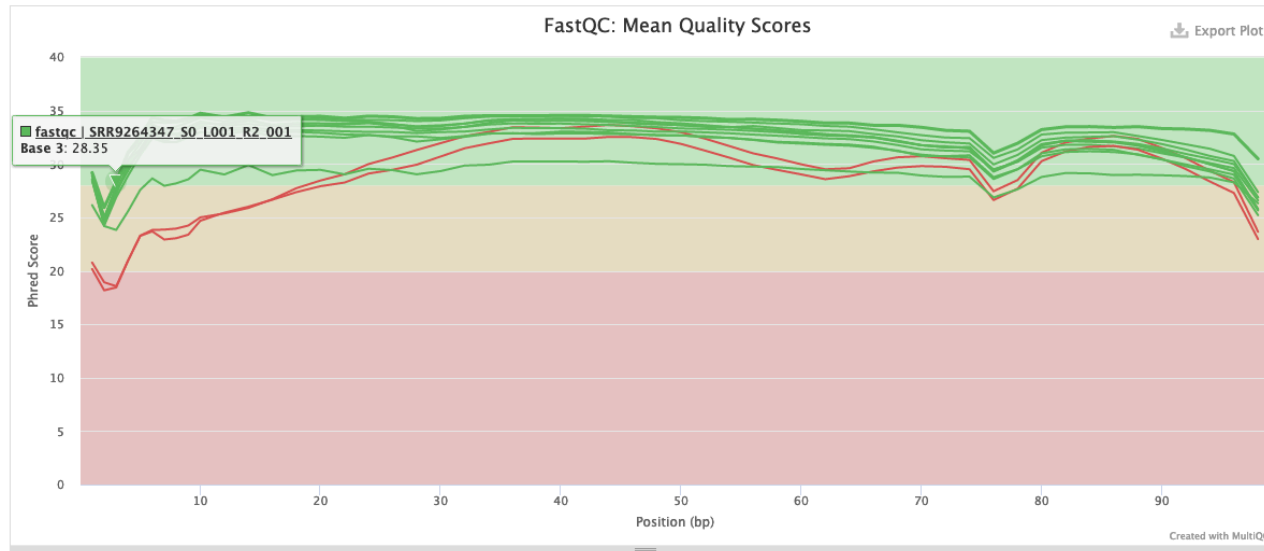
Sequence Quality Histograms

10 2

The mean quality value across each base position in the read.

Help

Y-Limits: on



Alignment and counting

The first steps in the analysis of single cell RNAseq data:

- Align reads to genome
- Annotate reads with feature (gene)
- Quantify gene expression

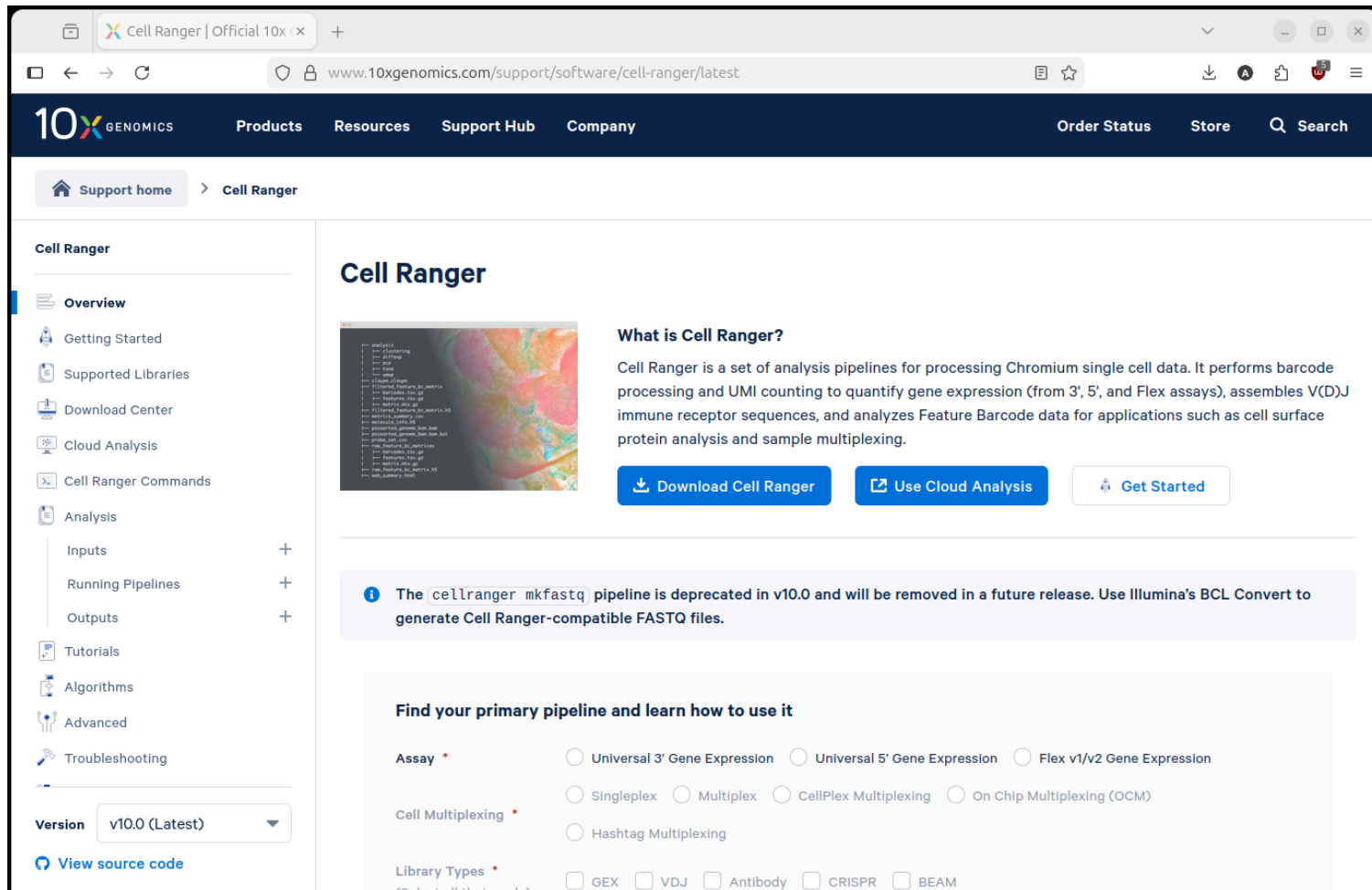
Cell Ranger

- 10x Cell Ranger - This not only carries out the alignment and feature counting, but will also:
 - Call cells
 - Generate a summary report in html format
 - Generate a “cloupe” file

Alternative methods include:

- STAR solo:
 - Generates outputs very similar to CellRanger minus the cloupe file and the QC report
 - Will run with lower memory requirements in a shorter time than Cell Ranger
- Alevin:
 - Based on the popular Salmon tool for bulk RNAseq feature counting
 - Alevin supports both 10x-Chromium and Drop-seq derived data

Obtaining Cell Ranger



The screenshot shows the Cell Ranger website on a browser. The browser address bar displays `www.10xgenomics.com/support/software/cell-ranger/latest`. The website header includes the 10x Genomics logo and navigation links for Products, Resources, Support Hub, and Company. A dark blue navigation bar contains links for Order Status, Store, and Search.

The main content area is titled "Cell Ranger" and features a sidebar on the left with a navigation menu. The sidebar includes sections for "Cell Ranger" (Overview, Getting Started, Supported Libraries, Download Center, Cloud Analysis, Cell Ranger Commands, Analysis) and "Inputs" (Running Pipelines, Outputs). A "Version" dropdown menu is set to "v10.0 (Latest)", and a "View source code" link is present.

The main content area includes a "What is Cell Ranger?" section with a description: "Cell Ranger is a set of analysis pipelines for processing Chromium single cell data. It performs barcode processing and UMI counting to quantify gene expression (from 3', 5', and Flex assays), assembles V(D)J immune receptor sequences, and analyzes Feature Barcode data for applications such as cell surface protein analysis and sample multiplexing." Below this text are three buttons: "Download Cell Ranger", "Use Cloud Analysis", and "Get Started".

A deprecation notice is displayed: "The `cellranger mkfastq` pipeline is deprecated in v10.0 and will be removed in a future release. Use Illumina's BCL Convert to generate Cell Ranger-compatible FASTQ files."

At the bottom, a section titled "Find your primary pipeline and learn how to use it" contains a form with radio buttons for "Assay" (Universal 3' Gene Expression, Universal 5' Gene Expression, Flex v1/v2 Gene Expression), "Cell Multiplexing" (Singleplex, Multiplex, CellPlex Multiplexing, On Chip Multiplexing (OCM), Hashtag Multiplexing), and "Library Types" (GEX, VDJ, Antibody, CRISPR, BEAM).

Cell Ranger tools

Cell Ranger includes a number of different tools for analysing scRNAseq data, including:

- `cellranger mkref` - for making custom references
- `cellranger count` - for aligning reads and generating a count matrix
- `cellranger aggr` - for combining multiple samples and normalising the counts

Preparing the raw fastq files

Cell Ranger requires the fastq file names to follow a convention:

```
<SampleName>_S<SampleNumber>_L00<Lane>_<Read>_001.fastq.gz
```

e.g. for a single sample in the Caron data set we have:

```
SRR9264343_S0_L001_I1_001.fastq.gz
```

```
SRR9264343_S0_L001_R1_001.fastq.gz
```

```
SRR9264343_S0_L001_R2_001.fastq.gz
```

Genome/Transcriptome Reference

As with other aligners Cell Ranger requires the information about the genome and transcriptome of interest to be provided in a specific format.

- Obtain from the 10x website for human or mouse (or both - PDX)
- Build a custom reference with `cellranger mkref`

Running cellranger count

- Computationally very intensive
- High memory requirements

```
File Edit View Search Terminal Help
%h%-$
%h%-$
%h%-$ cellranger count --id=SRR9264343 \
> --transcriptome=refdata-gex-mm10-2020-A \
> --fastqs=fastq \
> --sample=SRR9264343 \
> --localcores=8 \
> --localmem=64
```

Cell Ranger outputs

- One directory per sample

```
File Edit View Search Terminal Help
%h%-$ ..
%h%-$ ls SRR9264343/
_cmdline
_filelist
_finalstate
_invocation
_jobmode
_log
_mrosource
outs
_perf
SC_RNA_COUNTER_CS
_sitecheck
SRR9264343.mri.tgz
_tags
_timestamp
_uuid
_vdrkill
_versions
%h%-$
```

Cell Ranger outputs

```
File Edit View Search Terminal Help
_versions
%h%-$
%h%-$ ls SRR9264343/outs/
analysis
cloupe.cloupe
filtered_feature_bc_matrix
filtered_feature_bc_matrix.h5
metrics_summary.csv
molecule_info.h5
possorted_genome_bam.bam
possorted_genome_bam.bam.bai
raw_feature_bc_matrix
raw_feature_bc_matrix.h5
web_summary.html
%h%-$
```

Cell Ranger outputs

```
File Edit View Search Terminal Help
_versions
%h%-$
%h%-$ ls SRR9264343/outs/
analysis
cloupe.cloupe
filtered_feature_bc_matrix
filtered_feature_bc_matrix.h5
metrics_summary.csv
molecule_info.h5
possorted_genome_bam.bam
possorted_genome_bam.bam.bai
raw_feature_bc_matrix
raw_feature_bc_matrix.h5
web_summary.html
%h%-$
```

Cell Ranger report

SITTA6

Summary [Analysis](#)

14,668

Estimated Number of Cells

20,065

Mean Reads per Cell

1,344

Median Genes per Cell

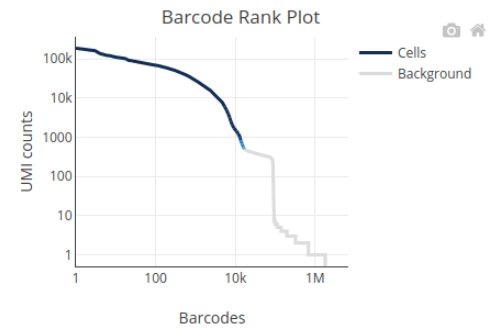
Sequencing ?

| | |
|-------------------------------|-------------|
| Number of Reads | 294,310,066 |
| Number of Short Reads Skipped | 0 |
| Valid Barcodes | 97.7% |
| Valid UMIs | 100.0% |
| Sequencing Saturation | 18.6% |
| Q30 Bases in Barcode | 96.1% |
| Q30 Bases in RNA Read | 94.6% |
| Q30 Bases in UMI | 95.7% |

Mapping ?

| | |
|------------------------------------|-------|
| Reads Mapped to Genome | 93.6% |
| Reads Mapped Confidently to Genome | 89.7% |

Cells ?



| | |
|----------------------------|--------|
| Estimated Number of Cells | 14,668 |
| Fraction Reads in Cells | 80.8% |
| Mean Reads per Cell | 20,065 |
| Median Genes per Cell | 1,344 |
| Total Genes Detected | 23,106 |
| Median UMI Counts per Cell | 2,928 |

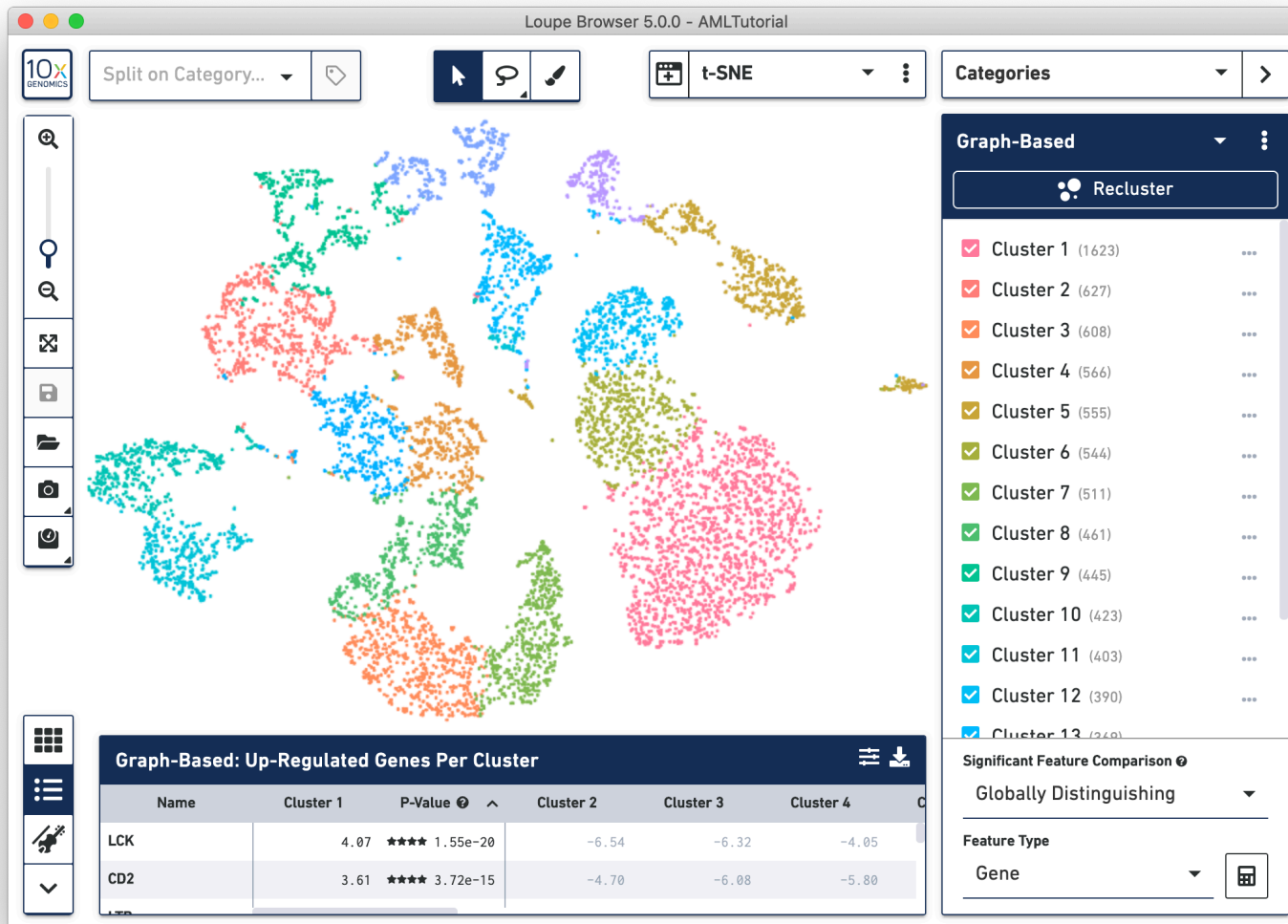
Sample

| | |
|--------------------|--------|
| Sample ID | SITTA6 |
| Sample Description | |

Cell Ranger outputs

```
File Edit View Search Terminal Help
_versions
%h%-$
%h%-$ ls SRR9264343/outs/
analysis
cLoupe.cLoupe
filtered_feature_bc_matrix
filtered_feature_bc_matrix.h5
metrics_summary.csv
molecule_info.h5
possorted_genome_bam.bam
possorted_genome_bam.bam.bai
raw_feature_bc_matrix
raw_feature_bc_matrix.h5
web_summary.html
%h%-$
```

Loupe Browser



Cell Ranger outputs

```
File Edit View Search Terminal Help
_versions
%h%-$
%h%-$ ls SRR9264343/outs/
analysis
cloupe.cloupe
filtered_feature_bc_matrix
filtered_feature_bc_matrix.h5
metrics_summary.csv
molecule_info.h5
possorted_genome_bam.bam
possorted_genome_bam.bam.bai
raw_feature_bc_matrix
raw_feature_bc_matrix.h5
web_summary.html
%h%-$
```

Cell Ranger outputs

```
File Edit View Search Terminal Help
_versions
%h%-$
%h%-$ ls SRR9264343/outs/
analysis
cloupe.cloupe
filtered_feature_bc_matrix
filtered_feature_bc_matrix.h5
metrics_summary.csv
molecule_info.h5
possorted_genome_bam.bam
possorted_genome_bam.bam.bai
raw_feature_bc_matrix
raw_feature_bc_matrix.h5
web_summary.html
%h%-$
```

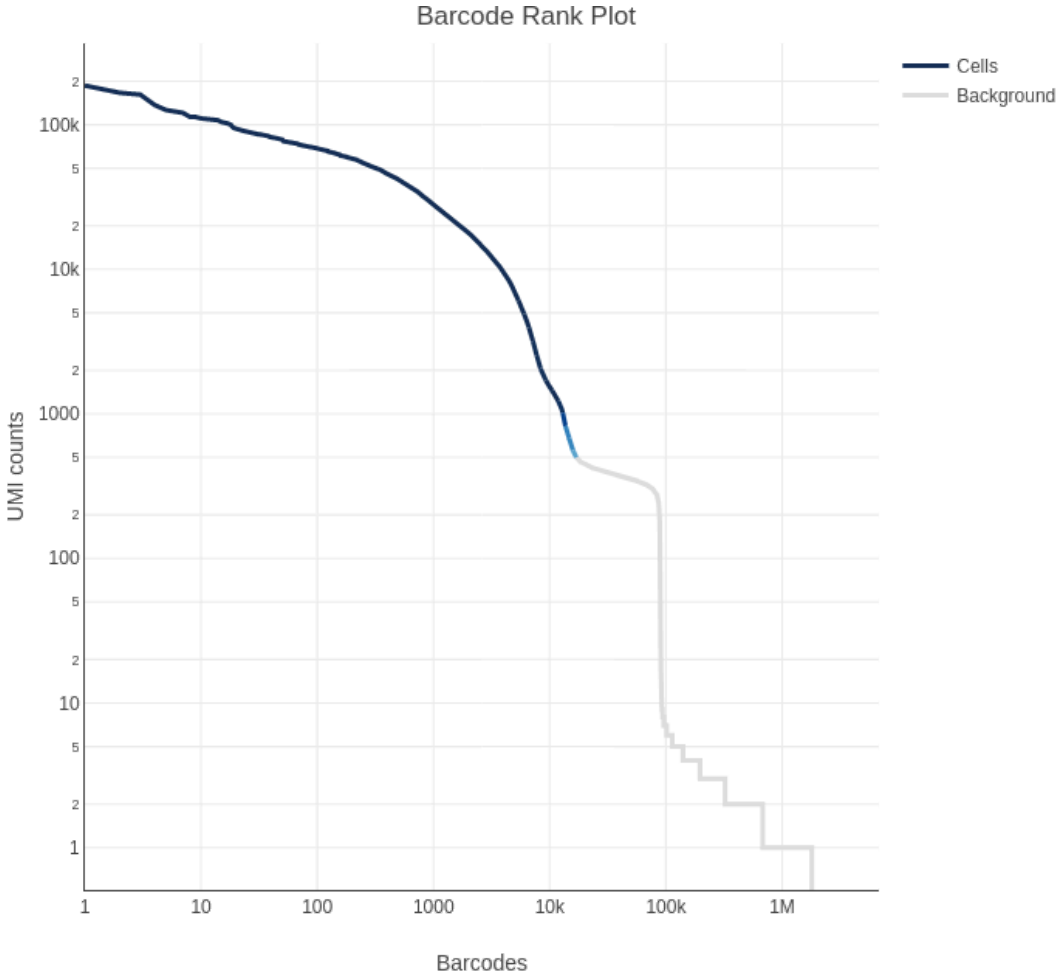
Cell Ranger outputs

```
File Edit View Search Terminal Help
_versions
%h%-$
%h%-$ ls SRR9264343/outs/
analysis
cloupe.cloupe
filtered_feature_bc_matrix
filtered_feature_bc_matrix.h5
metrics_summary.csv
molecule_info.h5
possorted_genome_bam.bam
possorted_genome_bam.bam.bai
raw_feature_bc_matrix
raw_feature_bc_matrix.h5
web_summary.html
%h%-$
%h%-$ ls SRR9264343/outs/raw_feature_bc_matrix
barcodes.tsv.gz
features.tsv.gz
matrix.mtx.gz
%h%-$
```

Cell Ranger outputs

```
File Edit View Search Terminal Help
_versions
%h%-$
%h%-$ ls SRR9264343/outs/
analysis
cloupe.cloupe
filtered_feature_bc_matrix
filtered_feature_bc_matrix.h5
metrics_summary.csv
molecule_info.h5
possorted_genome_bam.bam
possorted_genome_bam.bam.bai
raw_feature_bc_matrix
raw_feature_bc_matrix.h5
web_summary.html
%h%-$
```

Cell Ranger cell calling



Single Cell RNAseq Analysis Workflow

