Chromium Single Cell Gene Expression with Feature Barcoding Analysis

Software Training

CG000200 Rev A
Objectives

• Learn how to run Cell Ranger for Single Cell Gene Expression profiling with Feature Barcoding technology

• Understand key steps in Cell Ranger Count pipeline

• Know 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
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: Feature Barcoding data analysis for Cell Surface Protein or CRISPR Screening
  – Cell Ranger “count” pipeline for Feature Barcoding data
  – Outputs
  – QC metrics
Workflow Review

Single Cell 3’ Gene Expression with Feature Barcoding technology
Chromium Single Cell Gene Expression Solution with Feature Barcoding technology

**Solution Features**

- Unbiased single cell transcriptome 3’ gene expression that enables discovery research
- Simultaneously assess perturbation phenotypes, protein abundance and gene expression from the same cell
- Ready-to-use, robust workflow, including demonstrated protocols for various sample types such as cell lines, primary cells, dissociated fresh tissue
- Compatible with whole cells and nuclei
- Latest improvements increase sensitivity enabling the detection of more unique transcripts per cell, potentially decreasing sequencing requirements
- Easy-to-use and convenient software with Cell Ranger Analysis Pipeline and Loupe Cell Browser visualization tools
Single Cell Gene Expression with Feature Barcoding technology

Biochemistry 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
# Read Structure and Sequencing

## Single Cell 3’ Gene Expression Library

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

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Read 1</th>
<th>i7 Index</th>
<th>i5 Index</th>
<th>Read 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>10x Barcode &amp; UMI</td>
<td>Sample Index</td>
<td>N/A</td>
<td>Insert (Transcript)</td>
</tr>
<tr>
<td>Length</td>
<td>28*</td>
<td>8</td>
<td>0</td>
<td>91**</td>
</tr>
</tbody>
</table>

* If Pooled with Single Cell 3’ Gene Expression v2 Libraries, ensure that the Read 1 length is adjusted to 28 bp

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

<table>
<thead>
<tr>
<th></th>
<th>Read 1</th>
<th>i7 Index</th>
<th>i5 Index</th>
<th>Read 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>10x Barcode &amp; UMI</td>
<td>Sample Index</td>
<td>N/A</td>
<td>Insert (Feature Barcode)</td>
</tr>
<tr>
<td>Length</td>
<td>28</td>
<td>8</td>
<td>0</td>
<td>91*</td>
</tr>
</tbody>
</table>

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

*Single Cell 3’ CRISPR Screening Library*

Minimum 5,000 read pairs per cell is recommended for CRISPR Screening libraries

<table>
<thead>
<tr>
<th></th>
<th>Read 1</th>
<th>i7 Index</th>
<th>i5 Index</th>
<th>Read 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>10x Barcode &amp; UMI</td>
<td>Sample Index</td>
<td>N/A</td>
<td>Insert (Feature Barcode)</td>
</tr>
<tr>
<td><strong>Length</strong></td>
<td>28</td>
<td>8</td>
<td>0</td>
<td>91*</td>
</tr>
</tbody>
</table>

* Single Cell 3’ CRISPR Screening Libraries are typically pooled with Single Cell Gene Expression Libraries and sequenced using these parameters.
Analyzing Gene Expression Data

Laboratory and Analysis Workflow Overview

Cells → Chromium → Libraries → Sequencing → Cell Ranger → Loupe Cell Browser

- BCL
- mkfastq
- FASTQ
- count
- Feature Barcode Matrix
- BAM
- Run Summary
Workflow Review

Single Cell Immune Profiling with Feature Barcoding technology
Chromium Single Cell Immune Profiling Solution

Solution Features

- Whole transcriptome single cell 5’ gene expression that enables cell discovery and phenotypic analysis
- Enrichment for paired, full length human or mouse V(D)J sequences from single T and B cells
- Simultaneously characterize TCR transcripts, B Cell Ig transcripts, and gene expression from any sample
- Profile hundreds to tens of thousands of cells to detect rare clonotypes and cell type sub-populations
- Additional analyses enabled by Feature Barcoding technology
  - Link full-length, paired TCR α and β chain sequences and TCR-pMHC specificity
  - Simultaneously measure cell surface protein expression with gene expression
The Biochemistry Behind the Single Cell Immune Profiling Solution with Feature Barcoding technology

**Input:** Labeled single cells in suspension + 10x Gel Beads and Reagents

**Output:** Full-length paired V(D)J sequences and digital gene expression and cell surface protein expression profiles from single cells

Gene Expression, TCR/Ig Clonotype, and Surface Marker profiling of individual cells

<table>
<thead>
<tr>
<th>Cell 1...</th>
<th>Gene 1</th>
<th>Gene 2...</th>
<th>Gene 2,000</th>
<th>Feature 1...</th>
<th>Feature 100</th>
<th>TCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell 5,000</td>
<td>Gene 1</td>
<td>Gene 2...</td>
<td>Gene 2,000</td>
<td>Feature 1...</td>
<td>Feature 100</td>
<td>Ig</td>
</tr>
</tbody>
</table>
Libraries Compatible with Illumina Sequencers

Chromium Single Cell 5’
Gene Expression Library
*Minimum 20,000 read pairs per cell*

Chromium Single Cell 5’
Cell Surface Protein Library
*Minimum 5,000 read pairs per cell*

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Read 1</th>
<th>i7 Index</th>
<th>i5 Index</th>
<th>Read 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10x Barcode &amp; UMI</td>
<td>Sample Index</td>
<td>N/A</td>
<td>Insert (Transcript* and Feature Barcode)</td>
</tr>
<tr>
<td>Length</td>
<td>26</td>
<td>8</td>
<td>0</td>
<td>91**</td>
</tr>
</tbody>
</table>

* Single Cell 5’ Cell Surface Protein Libraries are typically pooled with Single Cell 5’ Gene Expression Libraries and sequenced using these parameters.
** User controlled trade-off between read length and transcript insert mapping rate.
Analyzing Immune Profiling Data

Laboratory and Analysis Workflow Overview

Cells → Chromium → Libraries → Sequencing → Cell Ranger → Output → Integrated Analysis

- Cells
- Chromium
- Libraries
- Sequencing
- Cell Ranger
- Output
- Integrated Analysis

Diagram showing the workflow from Cells to Chromium, then Libraries, Sequencing, followed by Cell Ranger, and finally leading to Output and Integrated Analysis.
Cell Ranger Overview

Single Cell Gene Expression with Feature Barcoding technology
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 Barcoding 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
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

- 7865 cells, 35433 reads/cell
- ~8 hours wall-time at minimum compute resources
Cell Ranger

Easy to Download and Install

• Download Cell Ranger and 10x reference data from:
  http://support.10xgenomics.com

• Simple Installation:
  $ tar -xzvf cellranger-x.y.x.tar.gz
  $ tar -xzvf refdata-cellranger-x.y.z.tar.gz
  $ export PATH=/opt/cellranger-x.y.z:$PATH

No Make. No Compile.

One Dependency. Illumina’s bcl2fastq installed and added to PATH variable
Demultiplexing

Generating FASTQs
Executing `cellranger mkfastq`

**Demultiplexing**

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

```
$cellranger mkfastq --id=sample1 \\
   --csv=my_samples.csv \\
   --run=/mnt/hiseq/sample1_bcl
```

- Output

```
$ls -l /opt/bar/sample1/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/sample1/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
```
Advanced: Demultiplexing

*Sample Indices*

- 10x sample indices are supplied on a 96 well plate
- Each well/10x sample index is composed of 4 oligos
- Example:
  - 10x index: SI-GA-A1
  - 4 Oligos: "GGTTTACT", "CTAAACGG", "TCGGCGTC", "AACCGTAA"
- 10x software tools can recognize oligos or 10x sample indices.
Cell Ranger “count” Pipeline for Gene Expression Analysis

Command line and Pipeline Steps
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_R1_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
Advanced: Cell Ranger Count Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>--id</td>
<td>A unique run ID string: e.g. sample345</td>
</tr>
<tr>
<td>--fastqs</td>
<td>Path of folder created by mkfastq or bcl2fastq. Path of the FASTQ folder generated by cellranger mkfastq. e.g. /home/jdoe/runs/HAWT7ADXX/outs/fastq_path Can take multiple comma-separated paths, which is helpful if the same library was sequenced on multiple flowcells.</td>
</tr>
<tr>
<td>--sample</td>
<td>Sample name as specified in the sample sheet supplied to mkfastq. Can take multiple comma-separated values, which is helpful if the sample was sequenced on multiple flowcells and the sample name used (and therefore fastq file prefix) is not identical between them. Doing this will treat all reads from the library, across flowcells, as one sample.</td>
</tr>
<tr>
<td>--transcriptome</td>
<td>Path of folder containing 10x-compatible reference.</td>
</tr>
<tr>
<td>--description</td>
<td>Sample description to embed in output files.</td>
</tr>
<tr>
<td>--libraries</td>
<td>CSV file declaring input library data sources.</td>
</tr>
<tr>
<td>--expect-cells</td>
<td>Expected number of recovered cells.</td>
</tr>
<tr>
<td>--force-cells</td>
<td>Force pipeline to use this number of cells, bypassing the cell detection algorithm.</td>
</tr>
<tr>
<td>--feature-ref</td>
<td>Feature reference CSV file, declaring feature-barcode constructs and associated barcodes.</td>
</tr>
<tr>
<td>--nosecondary</td>
<td>Disable secondary analysis, e.g. clustering. Optional.</td>
</tr>
<tr>
<td>--r1-length</td>
<td>Hard trim the input Read 1 to this length before analysis</td>
</tr>
<tr>
<td>--r2-length</td>
<td>Hard trim the input Read 2 to this length before analysis</td>
</tr>
<tr>
<td>--chemistry</td>
<td>Assay configuration. NOTE: by default the assay configuration is detected automatically, which is the recommended mode. You usually will not need to specify a chemistry. Options are: 'auto' for autodetection, 'threeprime' for Single Cell 3', 'fiveprime' for Single Cell 5', 'SC3Pv1' or 'SC3Pv2' or 'SC3Pv3' for Single Cell 3' v1/v2/v3, 'SC5P-PE' or 'SCSP-R2' for Single Cell 5' paired-end/R2-only. Default: auto.</td>
</tr>
<tr>
<td>--lanes</td>
<td>Comma-separated lane numbers.</td>
</tr>
<tr>
<td>--indices</td>
<td>Comma-separated sample index set &quot;SI-001&quot; or sequences.</td>
</tr>
<tr>
<td>--project</td>
<td>Name of the project folder within a mkfastq or `bcl2fastq-generated folder to pick FASTQs from.</td>
</tr>
</tbody>
</table>

Use “cellranger count --help” for a complete list of options
Reference Transcriptome

- 10x annotation uses ENSEMBL genomes and gene annotations

- 10x pre-built references
  - Human (hg19 and GRCh38)
  - Mouse (GRCm38)
  - Human and Mouse (GRCh38 + GRCm38)
  - ERCC

- Custom reference
  - Bundled `mkref` utility generates a 10x reference package from any FASTA and GTF gene file
  - Supports any reference compatible with STAR
Cell Ranger “count” Pipeline

*Single Cell 3’ Gene Expression Libraries*

Extract Barcode, UMI

Align

Correct Barcodes

Filter UMIs

Count UMIs

Call Cells

Secondary Analysis
Cell Ranger “count” Pipeline

*Single Cell Immune Profiling 5’ Gene Expression Libraries*

**Extract Barcode, UMI**
- Align
- Correct Barcodes
- Filter UMI
- Count UMI
- Call Cells
- Secondary Analysis

---

![Diagram of the Cell Ranger “count” Pipeline](image-url)
Cell Ranger “count” Pipeline: Align

In this step, the reads are aligned to the reference transcriptome.
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
Advanced: Cell Ranger “count” Pipeline

*Effect of Read Length on Mapping Rate*

- Higher read lengths yield better mapping rates / sensitivity
Cell Ranger “count” pipeline: Correct Barcodes

- 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).

```
Extract Barcode, UMI
Align
Correct Barcodes
Filter UMI
Count UMI
Call Cells
Secondary Analysis
```

Whitelist

Reads

1 base mismatch at low quality position
Cell Ranger “count” pipeline: Filter UMIs

- **Filter UMIs**
  - Must not be a homopolymer, e.g. AAAAAAAAAAA
  - 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.

---

**Diagram:**

- **UMI**
  - **Barcode1**: TCTGGAAAGTACC
  - **Barcode1**: TCTGGAAAGTACC
  - **Barcode1**: TCTGGAAAGTACC
  - **Barcode1**: TCTGGAAAGTACC

**Post Correction**

- **UMI**
  - **Barcode1**: TCTGGAAAGTACC
  - **Barcode1**: TCTGGAAAGTACC
  - **Barcode1**: TCTGGAAAGTACC
  - **Barcode1**: TCTGGAAAGTACC
Cell Ranger “count” pipeline: Count UMIs

- 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
Cell Ranger “count” pipeline: Cell Calling

*Why?*

- 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
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.

• Step2: 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
Advanced: Cell Calling

**Determining UMI Threshold for Step 1**

- Compute a **threshold based on total UMI counts per cell** to call cells with high RNA content.

- Rank-sort valid barcodes based on total UMI count per cell.

- Compute 99th percentile of the UMI counts among the top N barcodes. N= expected number of cells

- 10% of the 99th percentile value is the UMI threshold used for calling first mode of cells.
Advanced: Cell Calling: Potential Side Effects

Including Mitochondrially Enriched Cells

Cell Ranger 3.0

(6586 cells)
Cell Ranger “count” pipeline: Secondary Analysis

- Secondary analysis done only on cells
Gene Expression Analysis Outputs
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 file with t-SNE 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 Cell Browser
## Sources of Information

<table>
<thead>
<tr>
<th>Information</th>
<th>Web Summary</th>
<th>Loupe Cell Browser</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC Metrics &amp; Plots</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Barnyard Results</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>UMI Counts</td>
<td>✔️</td>
<td>Heatmap</td>
</tr>
<tr>
<td>t-SNE Plots</td>
<td>✔️</td>
<td>✔️ (Preferred)</td>
</tr>
<tr>
<td>Significant Genes</td>
<td>Pre-computed</td>
<td>Pre-computed &amp; dynamic</td>
</tr>
</tbody>
</table>
Gene Expression Analysis
QC Metrics

web_summary.html
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
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
Cell Calling

Why

• 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
QC Metrics: Per cell metrics

- Cell handling
  - Estimated Number of Cells: 7,077
  - Fraction Reads in Cells: 95.7%
- Depth
  - Mean Reads per Cell: 53,133
  - Median Genes per Cell: 1,505
- Data yield
  - Total Genes Detected: 23,036
  - Median UMI Counts per Cell: 4,492
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.

Algorithm has trouble discerning cells from the background.
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.
QC Metrics

**Mapping**

- Alert seen when Reads confidently mapped to transcriptome (<30%)
  - Reads mapped to wrong genome or different strain
  - Too short of read length
  - Custom reference contains overlapping genes

### Mapping

<table>
<thead>
<tr>
<th>Mapping</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reads Mapped to Genome</td>
<td>95.5%</td>
</tr>
<tr>
<td>Reads Mapped Confidently to Genome</td>
<td>92.5%</td>
</tr>
<tr>
<td>Reads Mapped Confidently to Intergenic Regions</td>
<td>5.0%</td>
</tr>
<tr>
<td>Reads Mapped Confidently to Intronic Regions</td>
<td>34.7%</td>
</tr>
<tr>
<td>Reads Mapped Confidently to Exonic Regions</td>
<td>52.7%</td>
</tr>
<tr>
<td>Reads Mapped Confidently to Transcriptome</td>
<td>49.7%</td>
</tr>
<tr>
<td>Reads Mapped Antisense to Gene</td>
<td>1.3%</td>
</tr>
</tbody>
</table>
QC Metrics

**Sequencing**

- Alert seen in web_summary.html when Q30 metrics are < 70%

- RNA Read Q30 metrics are low
  - Sequencing problems
  - Suboptimal loading concentration on sequencer
  - Note: RNA Read Q30 (R2) tends to be lower than the other reads.

### Sequencing

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Reads</td>
<td>638,901,019</td>
</tr>
<tr>
<td>Valid Barcodes</td>
<td>97.4%</td>
</tr>
<tr>
<td>Sequencing Saturation</td>
<td>68.2%</td>
</tr>
<tr>
<td>Q30 Bases in Barcode</td>
<td>93.7%</td>
</tr>
<tr>
<td>Q30 Bases in RNA Read</td>
<td>90.1%</td>
</tr>
<tr>
<td>Q30 Bases in Sample Index</td>
<td>90.1%</td>
</tr>
<tr>
<td>Q30 Bases in UMI</td>
<td>92.4%</td>
</tr>
</tbody>
</table>
• 20,000 read pairs per cell is the recommended minimum sequencing depth for typical samples

• Given variability in cell counting and loading, extra sequencing may be required if the cell count is higher than anticipated
Additional Cell Ranger Pipelines

Aggregate and Reanalyze
# Cell Ranger Pipelines for Gene Expression

## Overview

<table>
<thead>
<tr>
<th>Pipeline</th>
<th>Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>cellranger mkfastq</td>
<td>Barcode-aware demultiplexing from BCL to FASTQ</td>
</tr>
</tbody>
</table>
| 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 combined expression analysis and static visualizations  
• Produces .cloupe file for Loupe Cell Browser                                                                 |
| cellranger reanalyze| • Re-run the expression analysis with custom parameters  
• Produces .cloupe file for Loupe Cell Browser                                                                 |
Combining Sample Data with Cell Ranger “aggr” Pipeline

Laboratory and Analysis Workflow Overview
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.

<table>
<thead>
<tr>
<th>library_id</th>
<th>molecule_h5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMLNormal1</td>
<td>/opt/runs/LV123/outs/molecule_info.h5</td>
</tr>
<tr>
<td>AMLNormal2</td>
<td>/opt/runs/LB456/outs/molecule_info.h5</td>
</tr>
<tr>
<td>AMLPatient</td>
<td>/opt/runs/LP789/outs/molecule_info.h5</td>
</tr>
</tbody>
</table>

```$ cellranger aggr --sample=experiment1 --csv=library_sheet.csv$```
Combining Multiple Samples with Cell Ranger “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)
Aggregating Single Cell Gene Expression 3’ v2 and v3 chemistry data

Incorporating Chemistry Batch Correction Reduces Chemistry-Specific Clustering

10000 PBMCs – No Chemistry Correction

10000 PBMCs – With Chemistry Correction
Advanced: Aggregating Single Cell Gene Expression 3’ v2 and v3 chemistry data

Incorporating Chemistry Batch Correction Improves Subpopulation Clustering

10000 PBMCs – No Chemistry Correction

10000 PBMCs – With Chemistry Correction
Advanced: Aggregating V2 and V3 chemistry data

**Differential Gene Expression**

- Chemistry batch correction algorithm in “aggr” works in PCA space and affects t-SNE and cell clustering
- The UMI counts in feature barcode matrix are not affected by batch correction
- Differential gene expression uses these raw UMI counts
Advanced: Aggregating V2 and V3 chemistry data

Chemistry Batch Correction

10000 PBMCs – No Chemistry Correction

10000 PBMCs – With Chemistry Correction
# Cell Ranger “aggr” Pipeline

## Guidance on Combining Data Across Chemistries

This table shows various chemistry combinations and their compatibility in Cell Ranger’s “aggr” pipeline.

### Notes:

1. Data analyzed using Cell Ranger versions prior to 1.2 are not compatible with “aggr” pipeline.

2. The chemistry batch correction has been validated for combining 3’ v2 and 3’ v3 chemistry datasets. Using the correction for combining data from other chemistries could improve the mixing of the batches. However we advise careful validation of the results.

<table>
<thead>
<tr>
<th>Assay Chemistry</th>
<th>3’ v1</th>
<th>3’ v2</th>
<th>3’ v3</th>
<th>5’</th>
</tr>
</thead>
<tbody>
<tr>
<td>3’ v1</td>
<td>Supported. Do not use chemistry batch correction.</td>
<td>Possible. Can use chemistry batch correction with caution (Note 2).</td>
<td>Possible. Can use chemistry batch correction with caution (Note 2).</td>
<td>Not supported.</td>
</tr>
<tr>
<td>3’ v2</td>
<td></td>
<td>Supported. Do not use chemistry batch correction.</td>
<td>Supported. Use chemistry batch correction.</td>
<td></td>
</tr>
<tr>
<td>3’ v3</td>
<td></td>
<td></td>
<td>Supported. Do not use batch correction.</td>
<td></td>
</tr>
<tr>
<td>5’</td>
<td></td>
<td></td>
<td></td>
<td>Supported. Do not use batch correction.</td>
</tr>
</tbody>
</table>
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 parameters
  - Create 3-D t-SNE plots
  - User specified max K and graph based cluster detection values
Analyzing Multi-species Samples

Estimating Multiplet Rate and UMI Count Purity
Sequencing depth for human-mouse mix

• Goal: Establish functionality of the system

• Mix ~600 human and ~600 mouse cells

• 5,000 reads per cell is sufficient sequencing depth for the human-mouse mixture and for library QC.

• Cell Ranger reports specific metrics when given a mouse-human mixture that cannot be obtained otherwise:
  – Multiplet rate
  – UMI count purity
The human-mouse mixing experiment

Cell UMI Counts

mm10 UMI counts

hg19 UMI counts

hg19
mm10
Multiplet

Multiplets
Multiplet Rate

- Count GEMs that are more likely to be multiplets than singletons
- Infer the number of unobserved multiplets
  - In a 1:1 mixing ratio, this should be equal to the number of observed multiplets
- The reported multiplet rate will be approximately twice the number of “Multiplet” dots seen on the plot.

![Diagram of GEM with Gel Bead, Human cell, and Mouse cell]
Count Purity

- Among single-cell GEMs, compute the fraction of UMI counts coming from the less-abundant species.

[Diagram: GEM with Gel Bead. Human cell (>98% of UMI counts with this barcode) and Mouse RNA (<2% of UMI counts w/ this barcode).]
The human-mouse mixing experiment

Estimated Number of Cells
787

Mean Reads per Cell
58,111

GEM Partitions

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEMs with &gt;0 Cells</td>
<td>787</td>
</tr>
<tr>
<td>GEMs with &gt;1 Cell</td>
<td>4</td>
</tr>
<tr>
<td>Fraction GEMs with &gt;1 Cell</td>
<td>0.5%</td>
</tr>
<tr>
<td>Mean UMI Count Purity</td>
<td>99.1%</td>
</tr>
</tbody>
</table>

*Inferred. Usually this will be 2x the number of observed multiplets because we do not directly observe human-human and mouse-mouse GEMs.
Low Multiplet Rate Maintained with the Single Cell Gene Expression Solution v3

Mouse cells → Human cells

1,125 Cells: Multiplet Rate 0.5%
5,292 Cells: Multiplet Rate 3.6%
10,062 Cells: Multiplet Rate 5.4%
Cell Ranger “count” Pipeline for Feature Libraries

How to Run the Pipeline
Analyzing Gene Expression with Feature Barcoding Data

Laboratory and Analysis Workflow Overview

---

Cell RANGER

Loupe Cell Browser

---

Cells → Chromium → GEM Well → Libraries → Sequencing → Demultiplex → Data Analysis → Visualization

Feature Barcoding Data

mkfastq → count

FASTQ
 Executing `cellranger mkfastq`

**Demultiplexing**

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

```
$ cellranger mkfastq --id=sample1 \ 
  --csv=my_samples.csv \ 
  --run=/mnt/hiseq/sample1_bcl
```

- Output

```
$ls -l sample1/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 sample1/outs/fastq_path/HAWT7ADXX/sample1/
  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
```
Executing “cellranger count”

Gene Expression + Feature Barcoding 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
```

- 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!
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
```

<table>
<thead>
<tr>
<th>fastqs</th>
<th>sample</th>
<th>library_type*</th>
</tr>
</thead>
<tbody>
<tr>
<td>/opt/bar/run1</td>
<td>sample1_gex</td>
<td>Gene Expression</td>
</tr>
<tr>
<td>/opt/bar/run1</td>
<td>sample1_fbc</td>
<td>Antibody Capture</td>
</tr>
</tbody>
</table>

*Values used in library_type field needs to match one of “Gene Expression”, “Antibody Capture”, “CRISPR Guide Capture”, “Custom”
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_R1_001.fastq.gz
|-- sample1_gex_S1_L001_R2_001.fastq.gz
|-- sample1_fbc_S1_L001_R1_001.fastq.gz
|-- sample1_fbc_S1_L001_R2_001.fastq.gz
```

<table>
<thead>
<tr>
<th>fastqs</th>
<th>sample</th>
<th>library_type</th>
</tr>
</thead>
<tbody>
<tr>
<td>/opt/bar/run1</td>
<td>sample1_gex</td>
<td>Gene Expression</td>
</tr>
<tr>
<td>/opt/bar/run1</td>
<td>sample1_fbc</td>
<td>CRISPR Guide Capture</td>
</tr>
</tbody>
</table>

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

Diagram:

- Library Definition
- Transcriptome Reference
- Feature Definition

Output:

- cellranger count
- Matrix, etc.
Feature Definition File

Cell Surface Protein Libraries

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

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3</td>
<td>CD3_Custom</td>
<td>R2</td>
<td>P5(BC)</td>
<td>AACAAAGACCCTTGAG</td>
<td>Antibody Capture</td>
</tr>
<tr>
<td>CD4</td>
<td>CD4_Custom</td>
<td>R2</td>
<td>P5(BC)</td>
<td>TACCCGTAATAGCGT</td>
<td>Antibody Capture</td>
</tr>
</tbody>
</table>

Antibody panels from BioLegend are designed as (for example)

NNNNNNNNNAACAAGACCCTTGAGNNNNNNNNNN

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3</td>
<td>CD3_TotalSeqB</td>
<td>R2</td>
<td>P5NNNNNNNNNNNN(BC)</td>
<td>AACAAAGACCCTTGAG</td>
<td>Antibody Capture</td>
</tr>
<tr>
<td>CD4</td>
<td>CD4_TotalSeqB</td>
<td>R2</td>
<td>P5NNNNNNNNNNNN(BC)</td>
<td>TACCCGTAATAGCGT</td>
<td>Antibody Capture</td>
</tr>
</tbody>
</table>

Example files on [https://support.10xgenomics.com/](https://support.10xgenomics.com/)
Advanced: Read Structure

**CRISPR Screening Libraries**

- **Standard CRISPRi sgRNA**
  - Protospacer
  - Feature Barcode
  - Capture Sequence
  - hairpin
  - Nextera Read 1
  - 10x Barcode
  - Capture Sequence
  - Protospacer TSO
  - Read 2
  - P7
  - Sample Index
  - UMI

**Assay Workflow**
### Feature Definition File

**CRISPR Screening Libraries**

#### Standard CRISPRi sgRNA

![Diagram of CRISPRi sgRNA](image)

#### Example files on [https://support.10xgenomics.com/](https://support.10xgenomics.com/)

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
<th>target_gene_id</th>
<th>target_gene_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>GFP</td>
<td>R2</td>
<td>(BC)GTTTAAGAGCTAAGCTGGAA</td>
<td>GACCAGGATGGGCACCACCC</td>
<td>CRISPR Guide Capture</td>
<td>GFP</td>
<td>GFP</td>
</tr>
<tr>
<td>Ctrl1</td>
<td>Control1</td>
<td>R2</td>
<td>(BC)GTTTAAGAGCTAAGCTGGAA</td>
<td>ACGTACGTACGTACGTACGT</td>
<td>CRISPR Guide Capture</td>
<td>Non-Targeting</td>
<td>Non-Targeting</td>
</tr>
</tbody>
</table>

The presence of Non-Targeting guides is required to trigger CRISPR specific differential expression analysis.
Cell Ranger “count” for Cell Surface Protein

Pipeline Steps and Outputs
Cell Ranger “count” for Cell Surface Protein

GEX
- Extract Barcode, UMI
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Cells
- Secondary Analysis

Feature
- Extract Barcode, UMI, fBC
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Filter Aggregates
- Secondary Analysis

Sample Index
- P5
- Read 1N
- 10x Barcode
- Capture Seq 1
- Feature Barcode
- Read 2
- P7

R1
- Read 1
- 10x Barcode
- Capture Seq 1

R2
- Read 2
Cell Ranger “count” for Cell Surface Protein

- Extract Barcode, UMI
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Cells
- Secondary Analysis

GEX

Extract Barcode, UMI
Align
Correct Barcodes
Filter UMIs
Count UMIs
Call Cells
Secondary Analysis

Feature Extract Barcode, UMI, fBC
Align
Correct Barcodes
Filter UMIs
Count UMIs
Filter Aggregates
Secondary Analysis

Sample Index N

P5 Read 1 10x BC UMI Capture Seq Feature Barcode Read 2N P7

R1
R2

© 10X Genomics, Inc. 2018
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.
Cell Ranger “count” for Cell Surface Protein

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3</td>
<td>CD3_TotSeqB</td>
<td>R2</td>
<td>P5NNNNNNNNNN(BC)</td>
<td>AACAAGACCCCTTGAG</td>
<td>Antibody Capture</td>
</tr>
<tr>
<td>CD8a</td>
<td>CD8a_TotSeqB</td>
<td>R2</td>
<td>P5NNNNNNNNNN(BC)</td>
<td>ATGGCACTCAGATG</td>
<td>Antibody Capture</td>
</tr>
</tbody>
</table>

Feature Reference -->

- CD3
- CD8a
Cell Ranger “count” for Cell Surface Protein

<table>
<thead>
<tr>
<th>Feature Reference -&gt;</th>
<th>CD3</th>
<th>CD8a</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x-Barcode UMI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature Barcode</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 base mismatch at low quality position is corrected

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3</td>
<td>CD3_TotSeqB</td>
<td>R2</td>
<td>P5NNNNNNNNNN(BC)</td>
<td>AACAAGACCTTGGAG</td>
<td>Antibody Capture</td>
</tr>
<tr>
<td>CD8a</td>
<td>CD8a_TotSeqB</td>
<td>R2</td>
<td>P5NNNNNNNNNN(BC)</td>
<td>ATTGGCACTCAGATG</td>
<td>Antibody Capture</td>
</tr>
</tbody>
</table>
Cell Ranger “count” for Cell Surface Protein

**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).
Cell Ranger “count” for Cell Surface Protein

- Filter UMIs
  - Must not be a homopolymer, e.g. AAAAAAAAA
  - 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.

<table>
<thead>
<tr>
<th>GEX</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract Barcode, UMI</td>
<td>Extract Barcode, UMI, fBC</td>
</tr>
<tr>
<td>Align</td>
<td>Align</td>
</tr>
<tr>
<td>Correct Barcodes</td>
<td>Correct Barcodes</td>
</tr>
<tr>
<td>Filter UMIs</td>
<td>Filter UMIs</td>
</tr>
<tr>
<td>Count UMIs</td>
<td>Count UMIs</td>
</tr>
<tr>
<td>Call Cells</td>
<td>Filter Aggregates</td>
</tr>
<tr>
<td>Secondary Analysis</td>
<td>Secondary Analysis</td>
</tr>
</tbody>
</table>

UMI before correction: TCTGGAAGTACC
UMI after correction: TCTGGAAGTACC

Post Correction:
Barcode1 TCTGGAAGTACC
Barcode1 TCTGGAAGTACC
Barcode1 TCTGGAAGTACC
Barcode1 TCTGGAAGTACC
Cell Ranger “count” for Cell Surface Protein

<table>
<thead>
<tr>
<th>GEX</th>
<th>Feature</th>
<th>Feature Reference</th>
<th>CD3</th>
<th>CD8a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract Barcode, UMI</td>
<td>Extract Barcode, UMI, fBC</td>
<td>10x-Barcode UMI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Align</td>
<td>Align</td>
<td>Feature Barcode</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Barcodes</td>
<td>Correct Barcodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Filter UMIs</td>
<td>Filter UMIs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count UMIs</td>
<td>Count UMIs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call Cells</td>
<td>Filter Aggregates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary Analysis</td>
<td>Secondary Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Feature Barcode Matrix**

<table>
<thead>
<tr>
<th>Barcode</th>
<th>CD3</th>
<th>CD8a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Barcode2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Barcode3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Cell Ranger “count” for Cell Surface Protein

- 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
Advanced: Cell Ranger “count” for Cell Surface Protein

• GEMs with protein aggregates results in:
  – High UMI counts
  – Can exhaust UMI space in these GEMs
  – Cause UMI collision and high UMI correction rate

• If UMI were 2 bp, there are 16 possible UMI oligos

• Normally a small subset of UMIs are used. Low UMI collision rate

• Barcode with protein aggregates has high molecules numbers
### Advanced: Cell Ranger “count” for Cell Surface Protein

<table>
<thead>
<tr>
<th>GEX</th>
<th>Feature</th>
<th>UMI Observed</th>
<th>UMI Post correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract Barcode, UMI</td>
<td>Extract Barcode, UMI, fBC</td>
<td>Barcode1 ACGACTCGTCAA</td>
<td>Barcode1 ACGACTCGTCAA</td>
</tr>
<tr>
<td>Align</td>
<td>Align</td>
<td>Barcode1 ACGACTCG</td>
<td>Barcode1 ACGACTCG</td>
</tr>
<tr>
<td>Correct Barcodes</td>
<td>Correct Barcodes</td>
<td>Barcode1 ACGACTCGA CAA</td>
<td>Barcode1 ACGACTCGTCAA</td>
</tr>
<tr>
<td>Filter UMIs</td>
<td>Filter UMIs</td>
<td>Barcode1 ACGACTCGTCAA</td>
<td>Barcode1 ACGACTCGTCAA</td>
</tr>
<tr>
<td>Count UMIs</td>
<td>Count UMIs</td>
<td>Barcode1 ACGACTCGTCAA</td>
<td>Barcode1 ACGACTCGTCAA</td>
</tr>
<tr>
<td>Call Cells</td>
<td>Filter Aggregates</td>
<td>Barcode1 ACGACTCGTCAA</td>
<td>Barcode1 ACGACTCGTCAA</td>
</tr>
<tr>
<td>Secondary Analysis</td>
<td>Secondary Analysis</td>
<td>Barcode1 ACGACTCGTCAA</td>
<td>Barcode1 ACGACTCGTCAA</td>
</tr>
</tbody>
</table>

- Filter out barcodes with >50% corrected reads and 10k reads

**Not-so-bad sample**
- Filtered barcodes

**Very bad sample**
- Filtered barcodes

---

© 10X Genomics, Inc. 2018
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.
Cell Ranger “count” for Cell Surface Protein

GEX
- Extract Barcode, UMI
- Align
- Correct Barcodes
- Filter UMI
- Count UMI
- Call Cells
- Secondary Analysis

Secondary Analysis

Feature
- Extract Barcode, UMI, fBC
- Align
- Correct Barcodes
- Filter UMI
- Count UMI
- Filter Aggregates
- Feature Barcode Matrix

Feature Barcode Matrix

Genes
Cell barcodes
0 0 0 0 0
0 0 0 0 0
0 0 0 5 0
9 3 0 0 0
0 0 10 0 0

→ clustering
→ t-SNE

Differential Expression

Feature Barcode Matrix

Cell barcodes
0 0 0 0 0
0 0 0 0 0
0 0 0 0 0
0 0 0 2 0
0 0 0 2 0
0 0 0 10 0
0 0 0 0 0

→ t-SNE

© 10X Genomics, Inc. 2018
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.
Outputs for Cell Surface Protein
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 on raw antibody counts to visualize cells in a 2-D space
  – This t-SNE is viewable in Loupe Cell Browser
QC metrics for Cell Surface Protein

web_summary.html
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
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
Filtering Reads: Cell Surface Protein

Diagram:
- All Reads
- Valid Barcodes
- Cells
- Aligned
- Valid UMIs

The diagram illustrates the process of filtering reads, focusing on valid UMIs and aligning them for analysis.
QC metrics – Cell Surface Protein

“Antibody Application” Metrics

- Antibody: Antibody Reads in Cells

  - 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**

| Antibody: Fraction Antibody Reads             | 90.2% |
| Antibody: Fraction Antibody Reads Usable     | 56.4% |
| Antibody: Antibody Reads Usable per Cell     | 9,974 |
| Antibody: Fraction Reads in Barcodes with High UMI Counts | 0.0%  |
| Antibody: Fraction Unrecognized Antibody     | 9.8%  |
| Antibody: Antibody Reads in Cells            | 63.1% |
| Antibody: Median UMIs per Cell (summed over all recognized antibody barcodes) | 4,404 |
QC metrics – Cell Surface Protein

“Antibody Application” Metrics

- Antibody: Fraction Antibody Reads Usable

- What percent of all input reads are used for counting in filtered feature barcode matrix

- > 20% or 1000 usable reads per cell is fine
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%

<table>
<thead>
<tr>
<th>Antibody Application</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibody: Fraction Antibody Reads</td>
<td>90.2%</td>
</tr>
<tr>
<td>Antibody: Fraction Antibody Reads Usable</td>
<td>56.4%</td>
</tr>
<tr>
<td>Antibody: Antibody Reads Usable per Cell</td>
<td>9,974</td>
</tr>
<tr>
<td>Antibody: Fraction Reads in Barcodes with High UMI Counts</td>
<td>0.0%</td>
</tr>
<tr>
<td>Antibody: Fraction Unrecognized Antibody</td>
<td>9.8%</td>
</tr>
<tr>
<td>Antibody: Antibody Reads in Cells</td>
<td>63.1%</td>
</tr>
<tr>
<td>Antibody: Median UMIs per Cell (summed over all recognized antibody barcodes)</td>
<td>4,404</td>
</tr>
</tbody>
</table>
Cell Ranger “count” for CRISPR Screening

Pipeline Steps
Cell Ranger “count” pipeline: CRISPR

<table>
<thead>
<tr>
<th>GEX</th>
<th>CRISPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract Barcode, UMI</td>
<td>Extract Barcode, UMI, fBC</td>
</tr>
<tr>
<td>Align</td>
<td>Align</td>
</tr>
<tr>
<td>Correct Barcodes</td>
<td>Correct Barcodes</td>
</tr>
<tr>
<td>Filter UMI s</td>
<td>Filter UMI s</td>
</tr>
<tr>
<td>Count UMI s</td>
<td>Count UMI s</td>
</tr>
<tr>
<td>Call Cells</td>
<td>Call Protospacer</td>
</tr>
<tr>
<td>Secondary Analysis</td>
<td>Secondary Analysis</td>
</tr>
</tbody>
</table>

“Protospacer”

Sample Index

P5  Read 1 N  10x UMI Capture Seq 1  Feature Barcode  TSO  Read 2  P7
Cell Ranger “count” pipeline: CRISPR

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
<th>target_gene_id</th>
<th>target_gene_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>GFP</td>
<td>R2</td>
<td>(BC)GTTTAAGAGCTAAGCTGGA</td>
<td>GACCAGGATGGGGCACCACCC</td>
<td>CRISPR Guide Capture</td>
<td>GFP</td>
<td>GFP</td>
</tr>
<tr>
<td>Ctrl</td>
<td>Control1</td>
<td>R2</td>
<td>(BC)GTTTAAGAGCTAAGCTGGA</td>
<td>ACGTACGTACGTACGTACGT</td>
<td>CRISPR Guide Capture</td>
<td>Non-Targeting</td>
<td>Non-Targeting</td>
</tr>
</tbody>
</table>

Reads are first searched for the constant sequence in pattern column.
Cell Ranger “count” pipeline: CRISPR

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
<th>target_gene_id</th>
<th>target_gene_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>GFP</td>
<td>R2</td>
<td>(BC)GTTTAAGAGCTAAGCTGGAA</td>
<td>GACCAGGATGGGCACCACCC</td>
<td>CRISPR Guide Capture</td>
<td>GFP</td>
<td>GFP</td>
</tr>
<tr>
<td>Ctrl1</td>
<td>Control1</td>
<td>R2</td>
<td>(BC)GTTTAAGAGCTAAGCTGGAA</td>
<td>ACGTACGTACGTACGTACGT</td>
<td>CRISPR Guide Capture</td>
<td>Non-Targeting</td>
<td>Non-Targeting</td>
</tr>
</tbody>
</table>

1 base mismatch at low quality position is corrected
Cell Ranger “count” pipeline: CRISPR

- **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).

---

**Whitelist**

![Whitelist Diagram]

**Reads**

![Reads Diagram]

1 base mismatch at low quality position

© 10X Genomics, Inc. 2018

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.
# Cell Ranger “count” pipeline: CRISPR

<table>
<thead>
<tr>
<th>GEX</th>
<th>CRISPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract Barcode, UMI</td>
<td>Extract Barcode, UMI, fBC</td>
</tr>
<tr>
<td>Align</td>
<td>Align</td>
</tr>
<tr>
<td>Correct Barcodes</td>
<td>Correct Barcodes</td>
</tr>
<tr>
<td>Filter UMIs</td>
<td>Filter UMIs</td>
</tr>
<tr>
<td>Count UMIs</td>
<td>Count UMIs</td>
</tr>
<tr>
<td>Call Cells</td>
<td>Call Protosporacer</td>
</tr>
<tr>
<td>Secondary Analysis</td>
<td>Secondary Analysis</td>
</tr>
</tbody>
</table>

### 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.

---

**UMI**
- TCTGGAA GTACC
- TCTGGAA GTACC
- TCTGGAA GTACC

**Post Correction**
- TCTGGAA GTACC
- TCTGGAA GTACC
- TCTGGAA GTACC
Cell Ranger “count” pipeline: CRISPR

<table>
<thead>
<tr>
<th>GEX</th>
<th>CRISPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract Barcode, UMI</td>
<td>Extract Barcode, UMI, fBC</td>
</tr>
<tr>
<td>Align</td>
<td>Align</td>
</tr>
<tr>
<td>Correct Barcodes</td>
<td>Correct Barcodes</td>
</tr>
<tr>
<td>Filter UMIs</td>
<td>Filter UMIs</td>
</tr>
<tr>
<td>Count UMIs</td>
<td>Count UMIs</td>
</tr>
<tr>
<td>Call Cells</td>
<td>Call Protospacer</td>
</tr>
<tr>
<td>Secondary Analysis</td>
<td>Secondary Analysis</td>
</tr>
</tbody>
</table>

**10x Barcode, UMI, fBC**

- **Feature Barcode**
  - 10x-Barcode
  - UMI

**GFP**

- Control 1

---

**Feature Barcode Matrix**

<table>
<thead>
<tr>
<th>Barcode</th>
<th>GFP</th>
<th>Control1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcode1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Barcode2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Barcode3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
• Protospacer calling is done to identify the CRISPR sgRNA associated with each cell barcode, and to distinguish true signal from background protospacer UMIs
  – Background protospacer in cells may be present due to
    • Ambient protospacer transcripts
    • PCR chimeras

• Protospacer-calling is done per sgRNA, independent of all other guide RNAs

• Algorithm
  – For each sgRNA,
  – Fit a 2-component Gaussian Mixture Model to the log(UMI counts) across all cell-associated barcodes.
  – Cells that fall in second mode called as having the sgRNA present.
Cell Ranger “count” pipeline: CRISPR

**GEX**
- Extract Barcode, UMI
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Cells
- Secondary Analysis

**CRISPR**
- Extract Barcode, UMI, fBC
- Align
- Correct Barcodes
- Filter UMIs
- Count UMIs
- Call Protospacer
- Secondary Analysis

---

**Feature Barcode Matrix**

<table>
<thead>
<tr>
<th>Cell barcodes</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0 0 0</td>
<td>0 3 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0</td>
<td>0 0 10 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0</td>
<td>0 0 5 0 0</td>
</tr>
</tbody>
</table>

→ clustering
→ t-SNE
→ Differential Expression

**Feature Barcode Matrix**

<table>
<thead>
<tr>
<th>Cell barcodes</th>
<th>Protospacer</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0 0 0</td>
<td>0 2 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0</td>
<td>0 0 100 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0</td>
<td>0 0 0 0 0</td>
</tr>
</tbody>
</table>

→ t-SNE

---

If control guides are provided

- **CRISPR analysis**
  - perturbation_effects_by_feature
top_perturbed_genes.csv
transcriptome_analysis.csv
perturbation_effects_by_target
top_perturbed_genes.csv
transcriptome_analysis.csv
perturbation_efficiencies_by_feature.csv
perturbation_efficiencies_by_target.csv
Outputs for CRISPR Screening
Cell Ranger “count” Output Files for CRISPR Screening Analysis

• Feature-Barcode Matrix
  – No additional files are produced – all CRISPR feature counts are output alongside gene counts in the feature-barcode matrix

• CRISPR Screening library sequencing and application metrics will be displayed in a single combined web summary with Gene Expression library metrics
  – “CRISPR Sequencing” and “CRISPR Application” tabs

• .cloupe file for data visualization and interactive analysis in Loupe Cell Browser
  – CRISPR gRNA feature counts are combined with Gene Expression data into a single .cloupe file per sample

• Additional CRISPR-specific output files are generated
  – “crispr_analysis” subfolder of outputs
Example CRISPR Screening Experimental Design

- Imagine the following feature reference file for a CRISPR experiment

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
<th>target_gene_id</th>
<th>target_gene_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>g1A</td>
<td>guide1A</td>
<td>R2</td>
<td>(BC)GTTTAAGCTAAGCTGGAA</td>
<td>GACCAGGATGGGCACCACCC</td>
<td>CRISPR Guide Capture</td>
<td>Gene1</td>
<td>TargetGene1</td>
</tr>
<tr>
<td>g1B</td>
<td>guide1B</td>
<td>R2</td>
<td>(BC)GTTTAAGCTAAGCTGGAA</td>
<td>CAGGAGGATGGGCACCAACC</td>
<td>CRISPR Guide Capture</td>
<td>Gene1</td>
<td>TargetGene1</td>
</tr>
<tr>
<td>g2A</td>
<td>guide2A</td>
<td>R2</td>
<td>(BC)GTTTAAGCTAAGCTGGAA</td>
<td>AATCCCTATGTCCACACGCC</td>
<td>CRISPR Guide Capture</td>
<td>Gene2</td>
<td>TargetGene2</td>
</tr>
<tr>
<td>g2B</td>
<td>guide2B</td>
<td>R2</td>
<td>(BC)GTTTAAGCTAAGCTGGAA</td>
<td>TAGCACGATCGGAGCCATTTC</td>
<td>CRISPR Guide Capture</td>
<td>Gene2</td>
<td>TargetGene2</td>
</tr>
<tr>
<td>Ctrl1</td>
<td>Control1</td>
<td>R2</td>
<td>(BC)GTTTAAGCTAAGCTGGAA</td>
<td>ACGTACGTACGTACGTACGT</td>
<td>CRISPR Guide Capture</td>
<td>Non-Targeting</td>
<td>Non-Targeting</td>
</tr>
</tbody>
</table>
Interpreting Cell Ranger Output files for CRISPR Screening

Overview

• From a CRISPR experiment, the questions we can ask are:
  • How was my transduction?
  • How good was my CRISPR guide design?
  • How were the gene expression profiles of cells affected by each perturbation?
Interpreting Cell Ranger Output files for CRISPR Screening

Overview

• Cell Ranger “count” pipeline generates CRISPR analysis files as below

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: /opt/sample1/outs/crispr_analysis
- Loupe Cell Browser file: /opt/sample1/outs/cloupe.cloupe

Waiting 6 seconds for UI to do final refresh.
Pipestance completed successfully!
Interpreting Cell Ranger Output files for CRISPR Screening

**Protospacer calls**

- How good was my transduction?
  - For each cell, which protospacer(s) were called?
  - What were their UMI counts?

**Note:** Protospacer calls are made irrespective of “Non-Targeting” guide’s presence in feature reference
Interpreting Cell Ranger Output files for CRISPR Screening

Assessing Perturbation Efficiency

• How good was your sgRNA design?
• What is the change in expression of target genes per guide RNA with respect to control guides?

![Diagram showing fold change for genes with different sgRNAs compared to control](image-url)
### Advanced: Interpreting Cell Ranger Output files for CRISPR Screening

#### Assessing Perturbation Efficiency

<table>
<thead>
<tr>
<th>Perturbation</th>
<th>Target Guide</th>
<th>Log2 Fold Change</th>
<th>Cells with Perturbation</th>
<th>Mean UMI Count Among Cells with Perturbation</th>
<th>Cells with Non-Targeting Guides</th>
<th>Mean UMI Count Among Cells with Non-Targeting Guides</th>
</tr>
</thead>
<tbody>
<tr>
<td>g1A</td>
<td>g1A</td>
<td>-4.267695195</td>
<td>26</td>
<td>0</td>
<td>1310</td>
<td>0.672519084</td>
</tr>
<tr>
<td>g1B</td>
<td>g1B</td>
<td>-0.181897629</td>
<td>11</td>
<td>0</td>
<td>1310</td>
<td>0.1</td>
</tr>
<tr>
<td>g2A</td>
<td>g2A</td>
<td>-2.735391576</td>
<td>425</td>
<td>0.025882353</td>
<td>1310</td>
<td>0.184732824</td>
</tr>
<tr>
<td>g2B</td>
<td>g2B</td>
<td>-2.718071785</td>
<td>534</td>
<td>1.082397004</td>
<td>1310</td>
<td>6.97480916</td>
</tr>
<tr>
<td>g1A</td>
<td>g2A</td>
<td>g1A</td>
<td>-2.250533855</td>
<td>15</td>
<td>1.466666667</td>
<td>1310</td>
</tr>
</tbody>
</table>
Interpreting Cell Ranger Output files for CRISPR Screening

**Perturbation Effects**

- Differential expression for each combination of gRNAs (or target genes) versus control gRNAs

<table>
<thead>
<tr>
<th>Perturbing Gene1 vs Control</th>
<th>Perturbing Gene2 vs Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top Perturbed Genes</strong></td>
<td><strong>Top Perturbed Genes</strong></td>
</tr>
<tr>
<td><strong>Genes</strong></td>
<td><strong>Genes</strong></td>
</tr>
<tr>
<td><strong>Log2 Fold Change</strong></td>
<td><strong>Log2 Fold Change</strong></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>SULT1C2</td>
<td>PRSS2</td>
</tr>
<tr>
<td>-1.66</td>
<td>-2.94</td>
</tr>
<tr>
<td>AC016995.3</td>
<td>RP11-1072C15.4</td>
</tr>
<tr>
<td>-1.19</td>
<td>-1.92</td>
</tr>
<tr>
<td>FAM53A</td>
<td>CD24</td>
</tr>
<tr>
<td>-1.09</td>
<td>-1.73</td>
</tr>
<tr>
<td>TMEM158</td>
<td>NCR3</td>
</tr>
<tr>
<td>-1.08</td>
<td>-1.50</td>
</tr>
<tr>
<td>MNS1</td>
<td>SIGLEC8</td>
</tr>
<tr>
<td>-1.07</td>
<td>-1.35</td>
</tr>
<tr>
<td>MTCL1</td>
<td>AL049794.1</td>
</tr>
<tr>
<td>-1.04</td>
<td>-1.32</td>
</tr>
<tr>
<td>LINC01125</td>
<td>SNGC</td>
</tr>
<tr>
<td>-1.03</td>
<td>-1.26</td>
</tr>
<tr>
<td>ZNF718</td>
<td>ASS1</td>
</tr>
<tr>
<td>-1.03</td>
<td>1.25</td>
</tr>
<tr>
<td>FZD7</td>
<td>RP11-253E3.3</td>
</tr>
<tr>
<td>-1.01</td>
<td>-1.22</td>
</tr>
<tr>
<td>HRC</td>
<td>DNAJC22</td>
</tr>
<tr>
<td>-1.01</td>
<td>-1.22</td>
</tr>
<tr>
<td>AC020571.3</td>
<td>SLC4A8</td>
</tr>
<tr>
<td>-0.98</td>
<td>-1.16</td>
</tr>
<tr>
<td>RASL10A</td>
<td>ARRD4</td>
</tr>
<tr>
<td>-0.96</td>
<td>-1.13</td>
</tr>
<tr>
<td>PAIP28</td>
<td>RP11-2E11.6</td>
</tr>
<tr>
<td>-0.90</td>
<td>-1.10</td>
</tr>
<tr>
<td>GP6</td>
<td>TXNIP</td>
</tr>
<tr>
<td>-0.88</td>
<td>-1.05</td>
</tr>
</tbody>
</table>
QC metrics for CRISPR Screening

web_summary.html
QC metrics: CRISPR Screening

**Overview**

- One `web_summary.html` file
- Summary tab: includes metrics for
  - Gene expression library
  - CRISPR library
- CRISPR metrics in “Summary” tab only
QC metrics: CRISPR Screening

“CRISPR Sequencing” Metrics

- **CRISPR: Number of Reads**
  - Total number of reads for CRISPR library

- **CRISPR: Mean Reads per Cell**
  - Mean sequencing depth per cell

- **CRISPR: Valid Barcodes**
  - Fraction of reads with a barcode found in or corrected to one that is found in the whitelist

- **CRISPR: Sequencing Saturation**
  - The fraction of reads originating from an already-observed UMI

<table>
<thead>
<tr>
<th>QC metrics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRISPR: Number of Reads</td>
<td>28,040,666</td>
</tr>
<tr>
<td>CRISPR: Mean Reads per Cell</td>
<td>40,638</td>
</tr>
<tr>
<td>CRISPR: Valid Barcodes</td>
<td>97.4%</td>
</tr>
<tr>
<td>CRISPR: Sequencing Saturation</td>
<td>93.5%</td>
</tr>
<tr>
<td>CRISPR: Reads Mapped to Genome</td>
<td>99.5%</td>
</tr>
<tr>
<td>CRISPR: Reads Mapped to Transcripts</td>
<td>99.5%</td>
</tr>
<tr>
<td>CRISPR: Reads Mapped to Genome and Transcripts</td>
<td>99.5%</td>
</tr>
</tbody>
</table>

**CRISPR Sequencing**

- **CRISPR: Number of Reads**: 28,040,666
- **CRISPR: Mean Reads per Cell**: 40,638
- **CRISPR: Valid Barcodes**: 97.4%
- **CRISPR: Sequencing Saturation**: 93.5%
- **CRISPR: Reads Mapped to Genome**: 99.5%
- **CRISPR: Reads Mapped to Transcripts**: 99.5%
- **CRISPR: Reads Mapped to Genome and Transcripts**: 99.5%

**CRISPR Application**

- **CRISPR: Fraction Reads with Putative Protospacer Sequence**: 64.2%
- **CRISPR: Guided Reads**: 62.3%
- **CRISPR: Guided Reads Usable per Cell**: 56.9%
- **CRISPR: Fraction Protospacer Not Recognized**: 2.8%
- **CRISPR: Reads in Cells**: 91.9%
- **CRISPR: Cells with 1 or more protospacers detected**: 94.8%
- **CRISPR: Cells with 2 or more protospacers detected**: 0.0%
- **CRISPR: Median UMI per Cell (summed over all recognized protospacers)**: 850

**Sample**

- **Name**: demo
- **Description**: GFP
- **Transcriptome**: GRCm38
- **Chemistry**: Single Cell 2’-d
- **Cell Ranger Version**: 3.0
QC metrics: CRISPR Screening

“CRISPR Application” Metrics

- **Mapping metrics**
  - CRISPR: Fraction Reads with Putative Protospacer Sequence
    - Percent reads where constant sequence observed is an exact match to that in feature reference file
  - CRISPR: Fraction Guide Reads
    - Percent reads recognized as protospacers
  - Expect to see very little difference between both mapping metrics
  - Poor guide detection rates could be due to incorrect feature barcode sequence in feature reference

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>read</th>
<th>pattern</th>
<th>sequence</th>
<th>feature_type</th>
<th>target_gene_id</th>
<th>target_gene_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>GFP</td>
<td>R2</td>
<td>BCCTTTAAGAGCTAAGCTGGA</td>
<td>GACCAAGATGGGCACCCACCC</td>
<td>CRISPR Guide Capture</td>
<td>GFP</td>
<td>GFP</td>
</tr>
<tr>
<td>Ctrl1</td>
<td>Control1</td>
<td>R2</td>
<td>BCCTTTAAGAGCTAAGCTGGA</td>
<td>ACGTACGTACGTACGTACGT</td>
<td>CRISPR Guide Capture</td>
<td>Non-Targeting</td>
<td>Non-Targeting</td>
</tr>
</tbody>
</table>

CRISPR Application

<table>
<thead>
<tr>
<th>CRISPR: Fraction Reads with Putative Protospacer Sequence</th>
<th>64.2%</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRISPR: Fraction Guide Reads</td>
<td>62.3%</td>
</tr>
<tr>
<td>CRISPR: Fraction Guide Reads Usable</td>
<td>56.9%</td>
</tr>
<tr>
<td>CRISPR: Guide Reads Usable per Cell</td>
<td>23,120</td>
</tr>
<tr>
<td>CRISPR: Fraction Protospacer Not Recognized</td>
<td>2.9%</td>
</tr>
<tr>
<td>CRISPR: Guide Reads in Cells</td>
<td>91.9%</td>
</tr>
<tr>
<td>CRISPR: Cells with 1 or more protospacers detected</td>
<td>94.8%</td>
</tr>
<tr>
<td>CRISPR: Cells with 2 or more protospacers detected</td>
<td>0.0%</td>
</tr>
<tr>
<td>CRISPR: Median UMIs per Cell (summed over all recognized protospacers)</td>
<td>850</td>
</tr>
</tbody>
</table>
Filtering Reads: CRISPR Screening

Diagram showing the process of filtering reads in CRISPR screening.

- All CRISPR Reads
- Recognized Guides
- Valid Cell-Barcodes
- Cells
- Valid UMIs

The diagram illustrates the different stages of filtering, starting from all CRISPR reads and progressing through recognized guides, valid cell barcodes, and finally valid UMIs.
QC metrics: CRISPR Screening

“CRISPR Application” Metrics

- CRISPR: Fraction CRISPR Reads Usable

  - What percent of all input reads are used for counting in filtered feature barcode matrix
  - One reason for low value can be low transduction efficiency
  - > 20% or 1000 usable reads per cell is fine
QC metrics: CRISPR Screening

“CRISPR Application” Metrics

- CRISPR: Guide Reads in Cells
  - 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
    - High ambient guide RNAs due to dead cells
    - High background due to for example guide RNA contamination
  - > 50% is considered okay
QC metrics: CRISPR Screening

“CRISPR Application” Metrics

- Protospacer calling
  - CRISPR: Cells with 1 or more protospacers detected
  - CRISPR: Cell with 2 or more protospacers detected
  - Possible causes for low value
    - Poor transduction
    - Cells downregulating expression of guides

### CRISPR Application

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRISPR: Fraction Reads with Putative Protospacer Sequence</td>
<td>64.2%</td>
</tr>
<tr>
<td>CRISPR: Fraction Guide Reads</td>
<td>62.3%</td>
</tr>
<tr>
<td>CRISPR: Fraction Guide Reads Usable</td>
<td>56.9%</td>
</tr>
<tr>
<td>CRISPR: Guide Reads Usable per Cell</td>
<td>23,120</td>
</tr>
<tr>
<td>CRISPR: Fraction Protospacer Not Recognized</td>
<td>2.9%</td>
</tr>
<tr>
<td>CRISPR: Guide Reads in Cells</td>
<td>91.9%</td>
</tr>
<tr>
<td>CRISPR: Cells with 1 or more protospacers detected</td>
<td>94.8%</td>
</tr>
<tr>
<td>CRISPR: Cells with 2 or more protospacers detected</td>
<td>0.0%</td>
</tr>
<tr>
<td>CRISPR: Median UMIs per Cell (summed over all recognized protospacers)</td>
<td>850</td>
</tr>
</tbody>
</table>
Additional Resources
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
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