

### **10X Cell Ranger and Basic Quality Control**

## **Cell Ranger**

- Count function
- Used to process FASTQ files for 10X samples
- Generates UMI expression matrices, basic sample statistics, and interactive analysis platform
- Can only detect genes that are included in the reference genome used

Summary Analysis				
7,824 Estimated Number of Cells				0 #
10,8281,70Mean Reads per CellMedian Genesi		10k \$1000 00 WN 100		Cells Background
Sequencing @		10		
Number of Reads	84,718,756	1	100 10k 1M	_
Valid Barcodes	98.5%		Barcodes	
alid UMIs 99.9%		Estimated Number	of Cells	7,824
Sequencing Saturation	6.7%	Fraction Reads in (	Cells	83.35
Q30 Bases in Barcode	98.0%	Mean Reads per Ce	ell	10,828
Q30 Bases in RNA Read	93.3%	Median Genes per	Cell	1,700
Q30 Bases in UMI	97.7%	Total Genes Detect	ed	22,812
		Median UMI Counts	s per Cell	4,490
Mapping ③				
Reads Mapped to Genome	93.7%	Sample		
Reads Mapped Confidently to Genome	90.0%	Sample ID		N_1395BL_NextGE
Reads Mapped Confidently to Intergenic Regions	3.4%	Sample Descriptio	n	
Reads Mapped Confidently to Intronic Regions	17.9%	Chemistry		Single Cell 3' v3
Reads Mapped Confidently to Exonic Regions	68.8%	Reference Path		
Reads Mapped Confidently to Transcriptome	67.0%	Transcriptome		GRCh38-2020-/
Deads Manual Astissans to Cana	1.00			5

1.0%

Pipeline Version

Reads Mapped Antisense to Gene

Frederick

Laboratory

cellranger-4.0.0

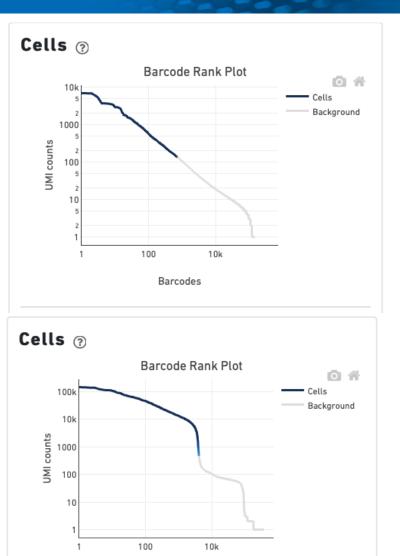
for Cancer Research

National

## **Cell Ranger**

Frederick National Laboratory for Cancer Research

- Barcode Rank Plot (Knee plot) can be used to determine sample quality
- Cell Ranger increased sensitivity for low UMI cell populations



Barcodes



- Data can contain values present due to noise, low quality cells, or doublets/multiplets
- Filtering is used to remove the excess noise to have a clean analysis
- Genes or cells with very sparse amounts of data usually excluded for downstream analysis
  - When a gene only shows up in a few cells, or a cell only contains a few genes this can be due to noise and have not enough data for analysis

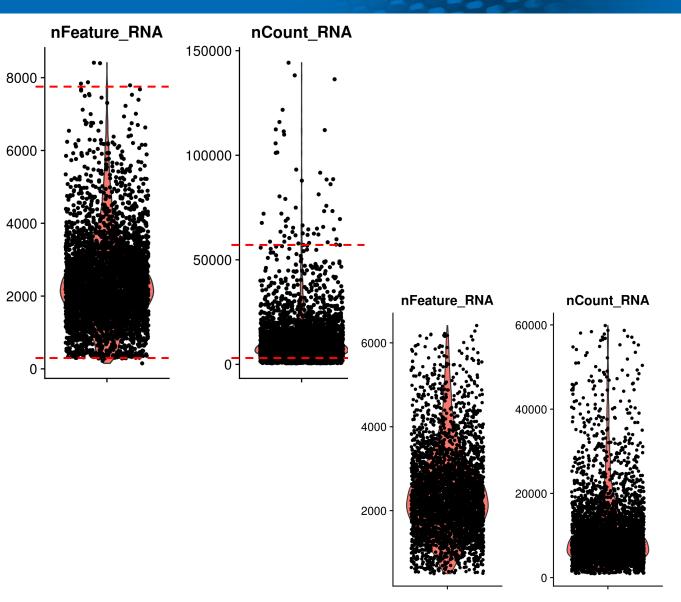


- Low quality cells include dying cells or cells with broken membranes
  - Contains lower amounts of genes
  - Has a higher expression of mitochondrial genes
- Doublet/Multiplets are when more than one cell is captured and labeled with the same cell barcode
- Stringent filters risk losing useful data
- Loose filters risk leaving in noise

#### Frederick National Laboratory for Cancer Research

# **Cell Filtering**

- Different cell types have different expression levels
- Filtering based on <u>UMI count, gene</u> <u>count</u>, and mitochondrial gene expression
- Cut-offs used for one cell type may not be appropriate for others
- Distributions and statistical methods can be used to find cut-offs



#### Frederick National Laboratory for Cancer Research

# **Cell Filtering**

- Filtering based on UMI count, gene count, and <u>mitochondrial gene</u> <u>expression</u>
- Mitochondrial fraction is linked to cell death, which may influence normalization
- Different cell types have different expression levels
- Cut-offs used for one cell type may not be appropriate for others
- Distributions and statistical methods can be used to find cut-offs

