Introduction to scRNA-seq Data Analysis
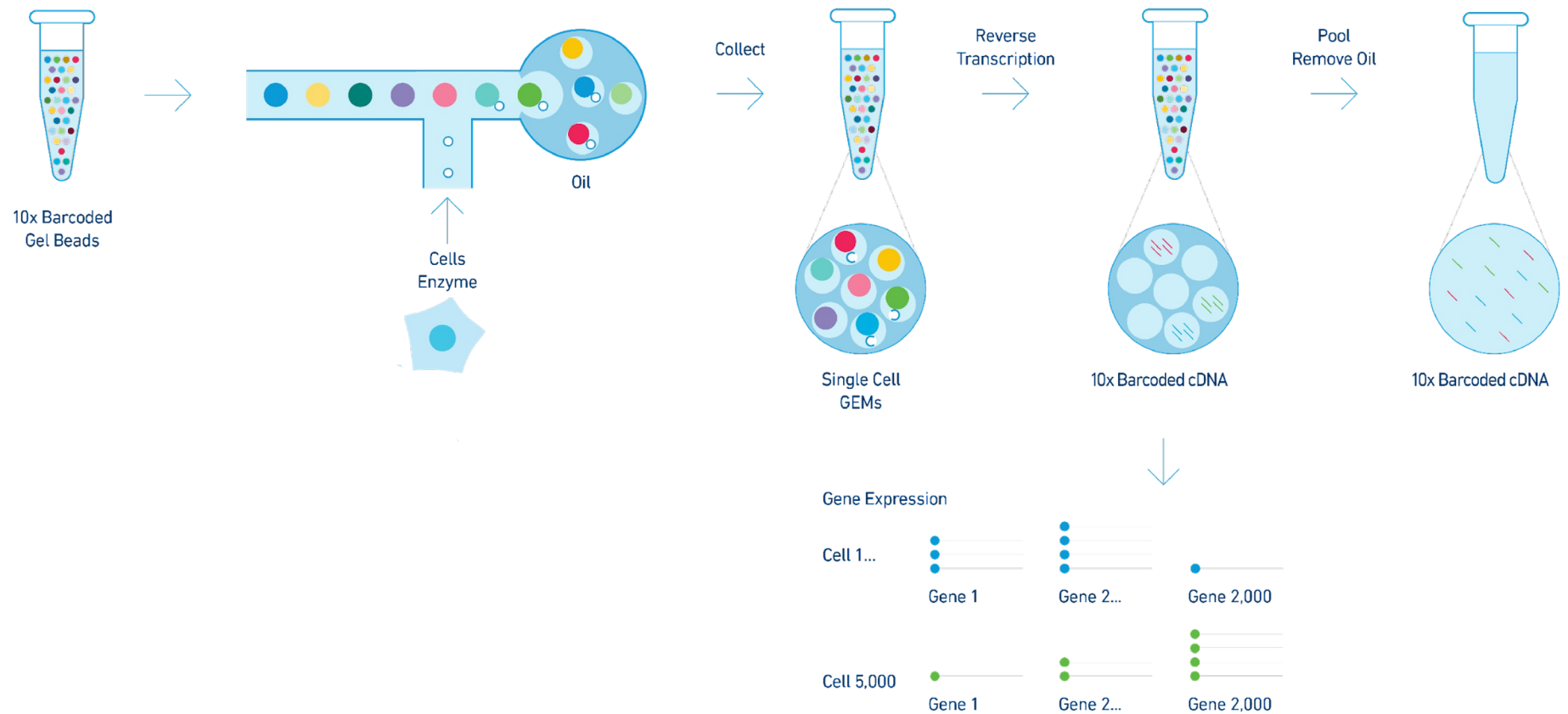
Hands on tour of the single cell analysis journey

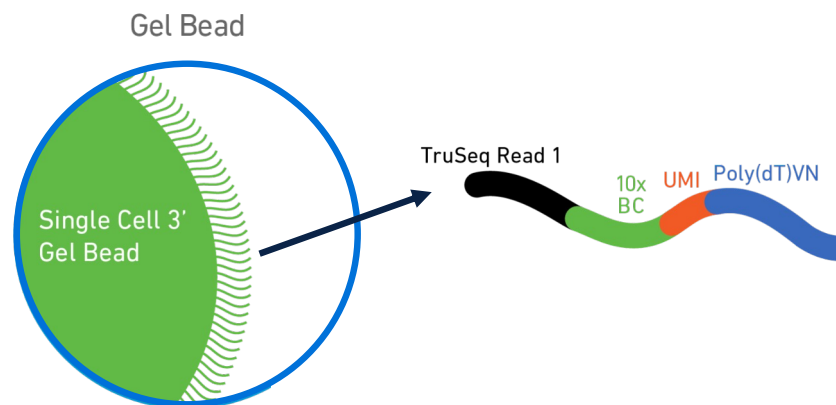# Outline

- Introduction to scRNA-seq

- Overview of Cell Ranger

- Where are we in our neutrophil analysis

Introduction to scRNA-seq data analysis
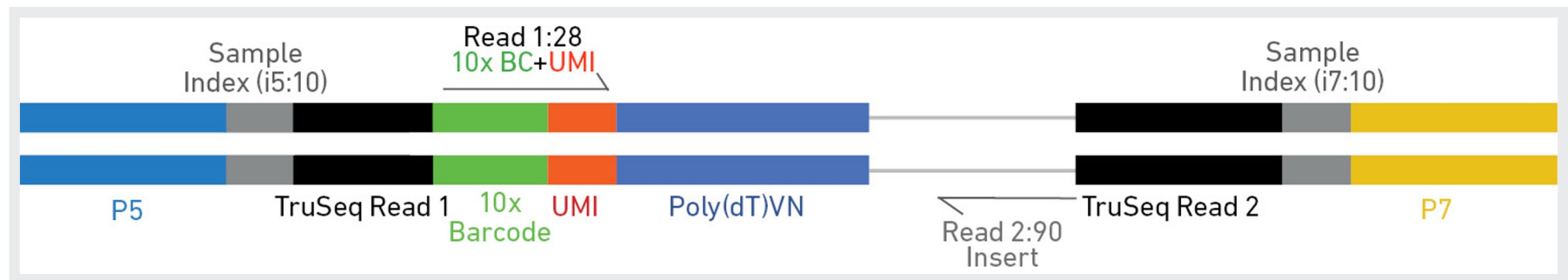
10x Genomics 3' Single Cell Gene Expression Assay

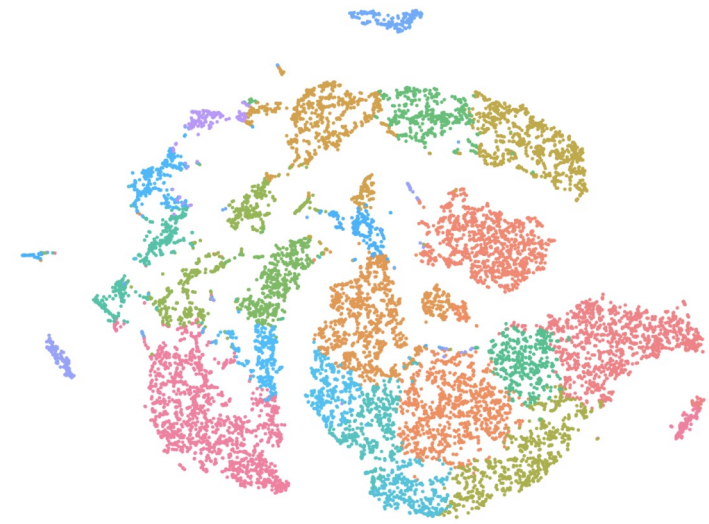# 3' Single cell gene expression

Read structure

Single Cell 3' Gene Expression library

Common questions answered with scRNA-seq

- What gene markers are expressed in different cell populations?

- What cell types are present in my sample? Are there novel or rare cell types?

- What pathways are activated?

- How is the cell population in one sample different from the other (e.g. control vs. drug treatment)?

- How do cells transition from one state to another during development?

- How do cells communicate with each other?

Mouse intestine, ~6600 cells

Single cell gene expression - data analysis flow

FASTQ files → Read processing & quantification

Read processing & quantification → QC & filtering

QC & filtering → Data integration

Data integration → Clustering & visualization

Clustering & visualization → Differential gene expression analysis

Clustering & visualization → Cell type annotation

Differential gene expression analysis → Cell type annotation

Cell type annotation → … and more

Tools in the data analysis flow

Cell Ranger

- A collection of pipelines for processing 10x single cell data
- Developed by 10x Genomics

Loupe

- Desktop tool for analysis and visualization
- Developed by 10x Genomics

Community developed tools

- Primarily programming libraries with some stand-alone tools (e.g. Seurat)
- Developed by the broader research community
- Not officially supported by 10x Genomics

Overview of Cell Ranger

Cell Ranger introduction

- A suite of analysis pipelines that process Chromium Single Cell data

- Contains various pipelines for:
  - Demultiplexing (mkfastq)
  - Single sample analysis (count)
  - Combining data from multiple samples (aggr)
  - Reanalyzing data (reanalyze)
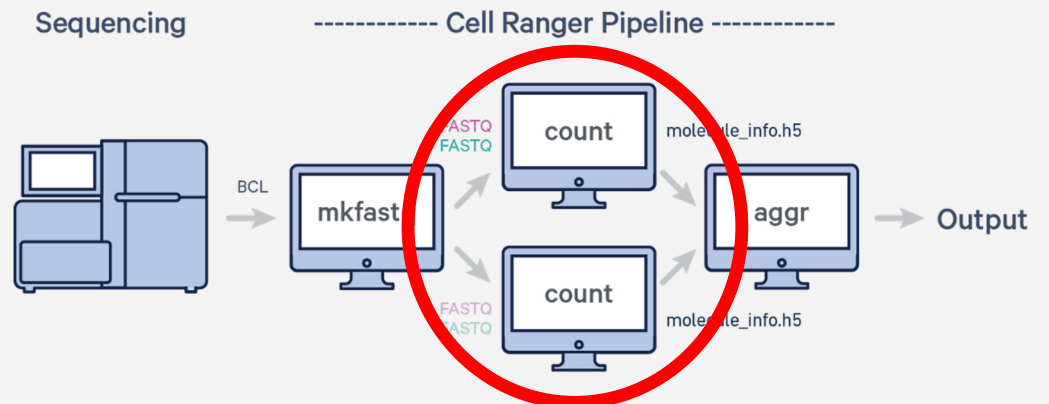
Running Cell Ranger

On 10x Cloud: Friendly user interface and simple data management (Available only in the US and Canada)

Required skill:
Understand the experimental design
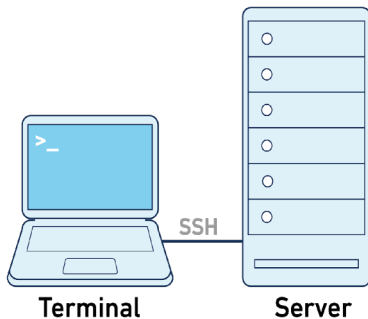
On a Linux system

Required skills:
- Understand the experimental design
- Comfortable running command line in the Linux environment
- Familiar with your organization's data management systems
- Know who to contact if there are server issues

Terminal   SSH   Server

10X GENOMICS

Cell Ranger Analysis Steps

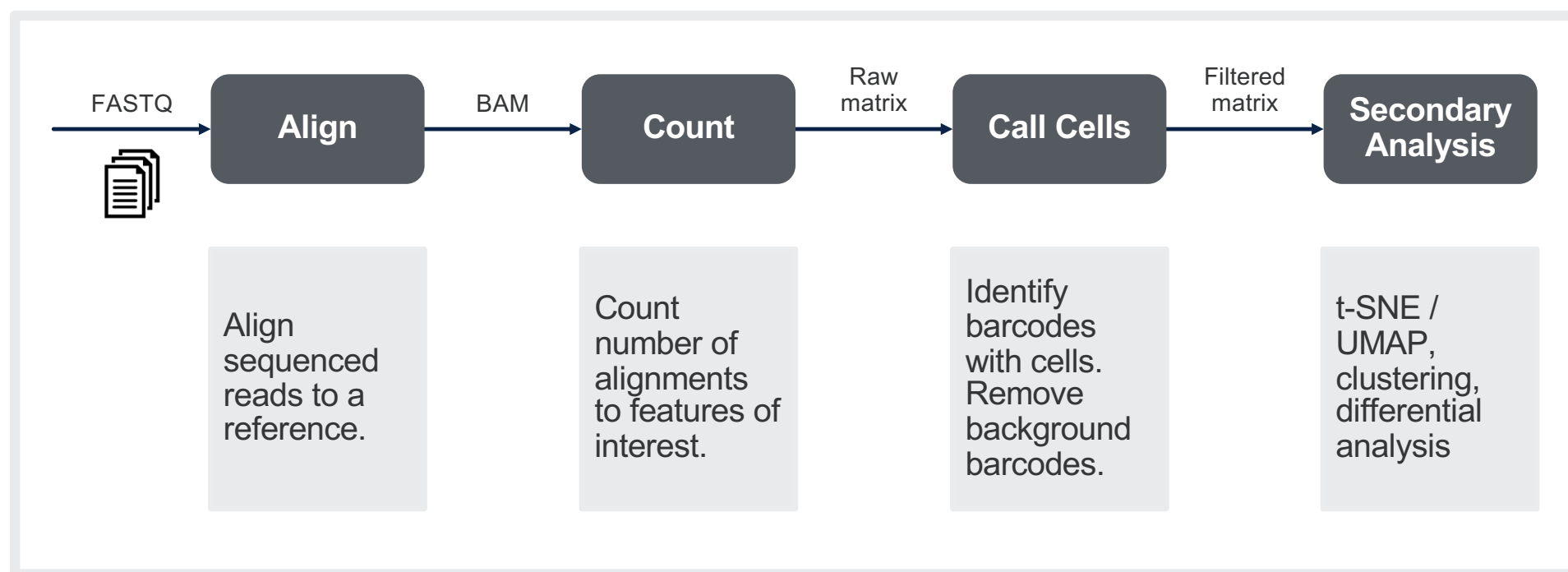FASTQ → **Align** → BAM → **Count** → Raw matrix → **Call Cells** → Filtered matrix → **Secondary Analysis**

Align	Count	Call Cells	Secondary Analysis
Align sequenced reads to a reference.	Count number of alignments to features of interest.	Identify barcodes with cells. Remove background barcodes.	t-SNE / UMAP, clustering, differential analysis

Where are we in our neutrophil analysis

10X GENOMICS

Neutrophil analysis plan

Our journey through analysis

Issue	Neutrophils have low UMI content	Neutrophils high intron retention rates	Cells separate into distinct clusters	Separate neutrophils from dead/dying cells	Neutrophils exist at different stages of maturation
Solution	Include low UMI barcodes	Map introns	Annotate cell clusters	Filter background	Velocity analysis
Tool	Cell Ranger "force-cells"	Cell Ranger "include-introns"	Loupe Browser	Loupe Browser	Community tools

Next up:

Quality Assessment using the Cell Ranger Web Summary

Topics

- Understanding your Web Summary

Next

- Sequencing metrics

- Mapping metrics

- Barcode rank plots

- Cell metrics

- Analysis Tab