Welcome to our hands-on tour of the single cell analysis journey

From Cloud to Loupe to community-developed tools

Date

Location

CG000725

Rev A

Goals and objectives

Primary goal

- **Demystify the data analysis process with a hands-on, active learning approach.**

Target audience

- Those who have or will soon have 10X data and have limited to no bioinformatics experience

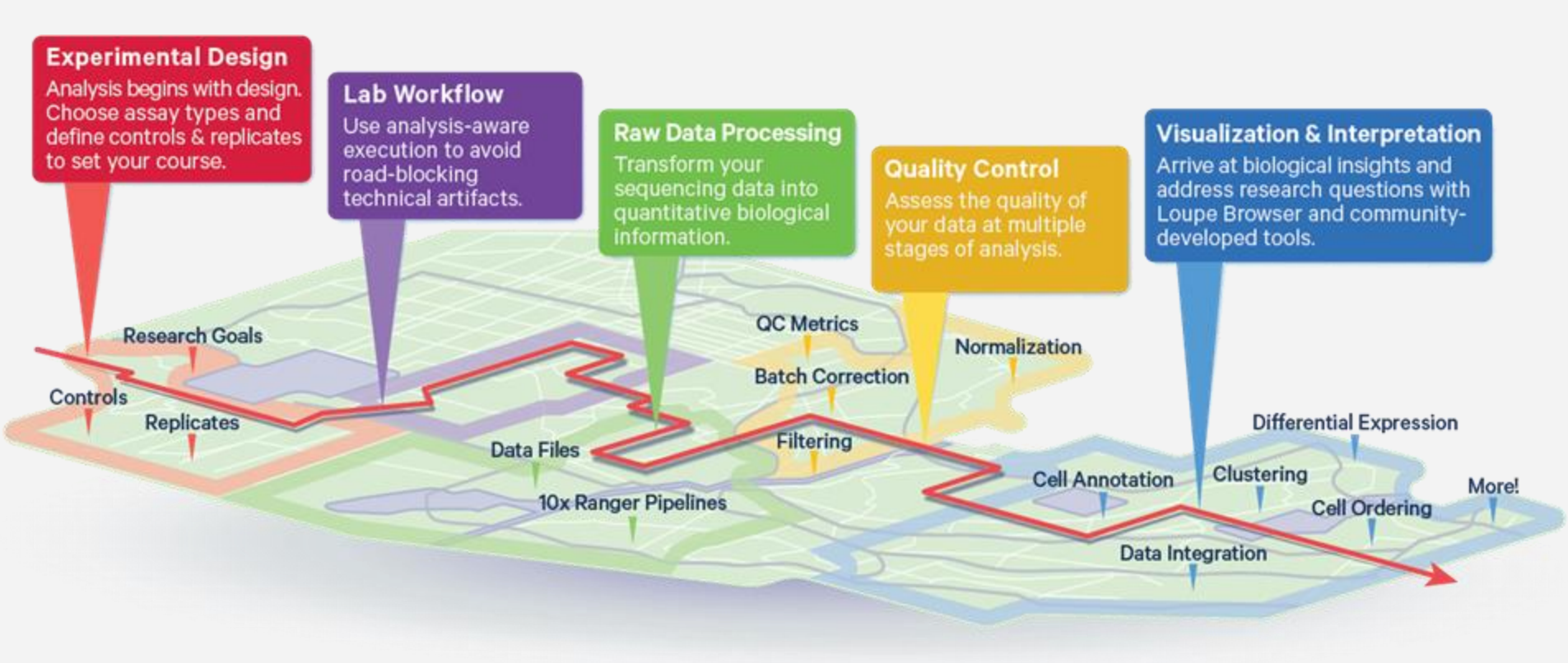
Goals and objectives

Secondary goals

- Help you see a path for you through the analysis journey
- Bring awareness to useful analysis resources, such as software documentation and analysis guides
- Get you started in raw data processing using 10x cloud analysis
- Show the expanded functionality of our Loupe Browser
- Introduce the amazing ecosystem of community developed tools for analyzing 10X data
- Fostering local analysis community

TL;DR: Help you move forward in analysis

Analysis Journey: Orientation



10x GENOMICS

Introduction to the FFPE tumor dataset: analysis begins at design

Hands on tour of the single cell analysis journey

Outline

- Analysis begins at design
- Purpose of the experiment
- FFPE challenges
- The data we are using today
- Analysis plan

Analysis begins at design

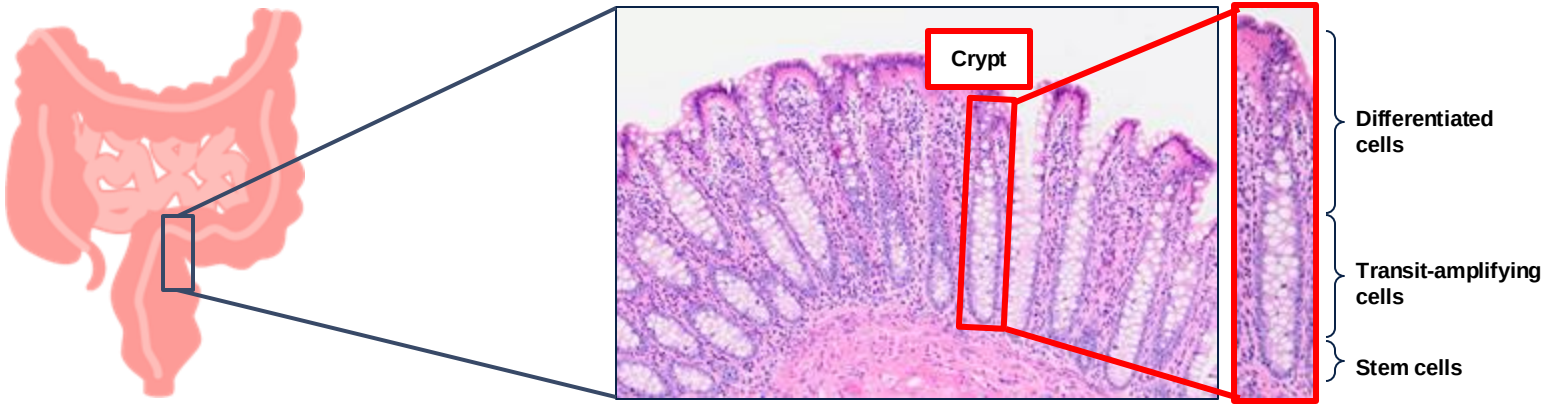
- The design enables analysis
- What samples, conditions, and data do you need?
- Planning is the cheapest part of the experiment

https://imgflip.com/memegenerator/One-Does-Not-Simply

Background of the experiment

What we know about **colorectal cancer (CRC)**:

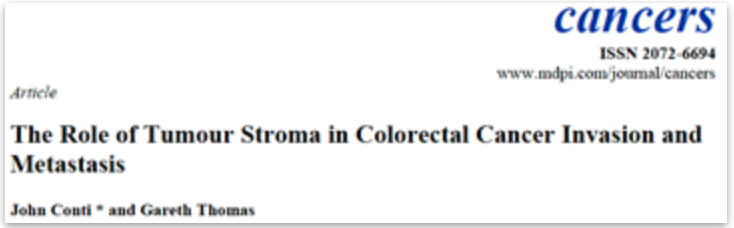
- One of the most common types of cancer
- Occurs in the epithelial cells of the colon
- Mutations in signaling pathways may increase gene expression or activity



Background of the experiment

What we know about **colorectal cancer (CRC)**:

- Nearby tumor stromal cells support cancerous cells and play a critical role in CRC tumor development & metastases
- The composition of the tumor microenvironment may act as a prognostic marker
    - Some infiltrating immune cells are associated with higher survival
        - Others may actually *promote* tumor development



Cancers 2011, 3, 2160-2168; doi:10.3390/cancers3022160

Nat Rev Cancer. 2022 Apr 7;22(7):414–430. doi: 10.1038/s41568-022-00466-1

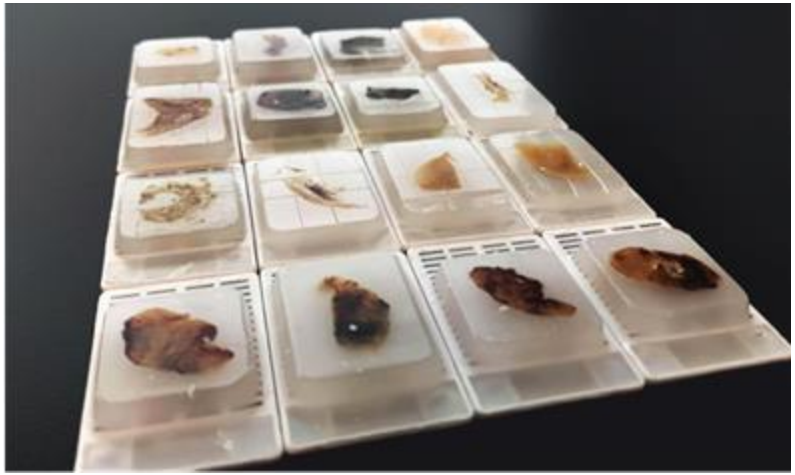
Immunology 14 August 2022 https://doi.org/10.1111/imm.13568

Goal of the experiment

- Analyze clinical CRC tumor sample (FFPE-preserved)
- Identify cancerous epithelial population and supporting tumor stroma
- Profile tumor stroma and identify infiltrating immune cells
- Characterize gene expression profile of immune cells for downstream clinical use

Tumor microenvironment

Challenges of working with FFPE samples

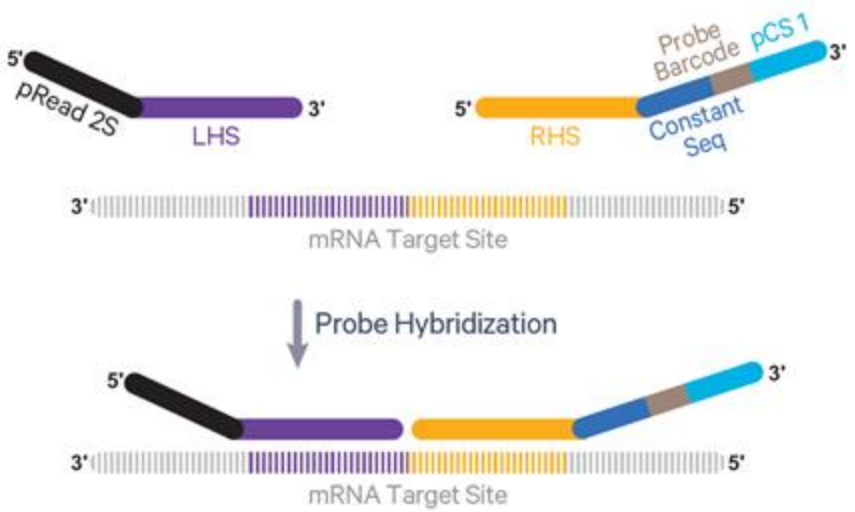


- Formalin Fixation and Paraffin Embedding (FFPE) is the primary method of clinical sample preservation
  - Excellent preservation and tissue stability
  - In many cases, FFPE blocks may be the only sample available
- FFPE leads to nucleic acid degradation
  - RNA-seq analysis is particularly challenging
- FFPE preservation is not standardized
  - Some blocks work better for molecular experiments

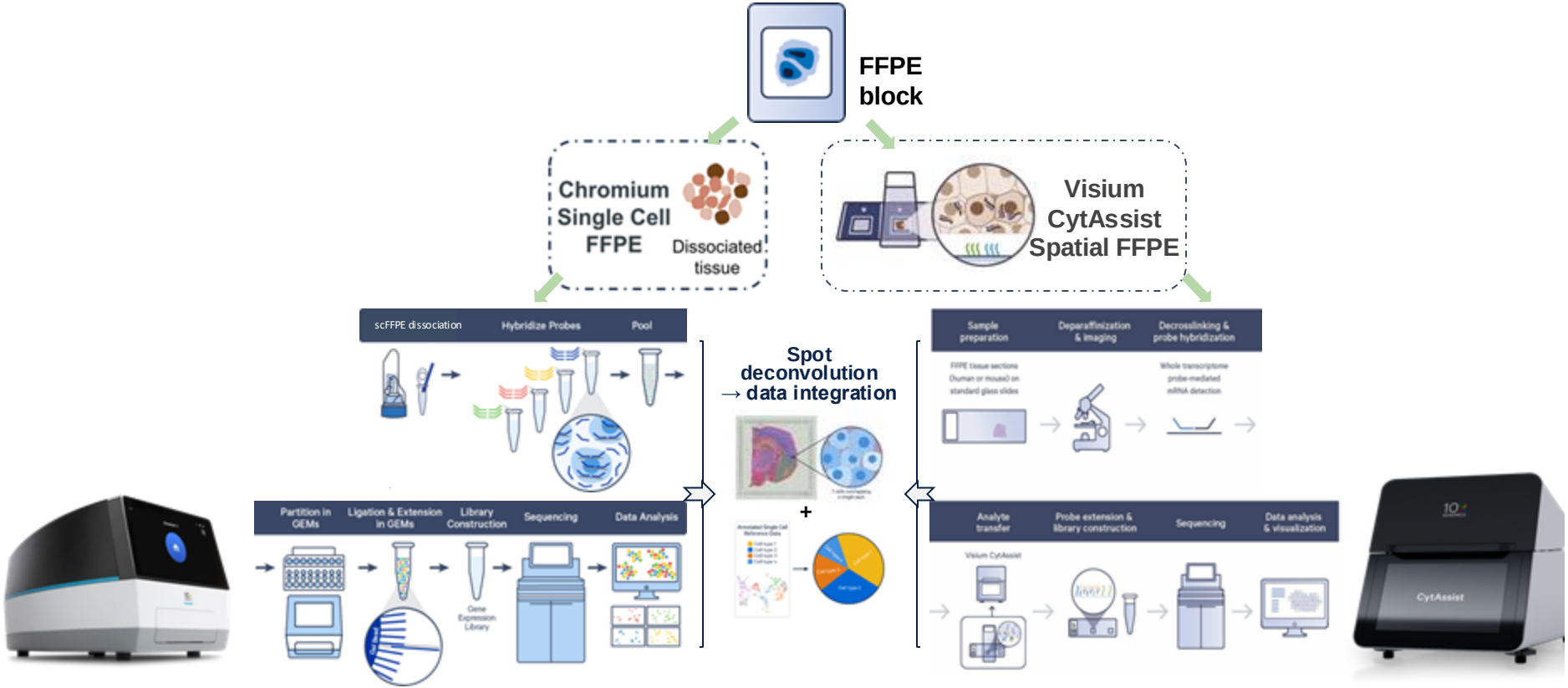
Benefits of Chromium Single-Cell Flex

Probe-based chemistry enables capture of fixed RNA

- Compatible with RNA from FFPE samples
- Does not rely on an intact poly-A tail and only requires 50nt transcript sequence
- Maximum sensitivity with 3 probe pairs per gene
- Built-in multiplexing
- Does not include:
  - Intronic, intergenic, or antisense reads
  - Variable regions such as BCR/TCR sequences

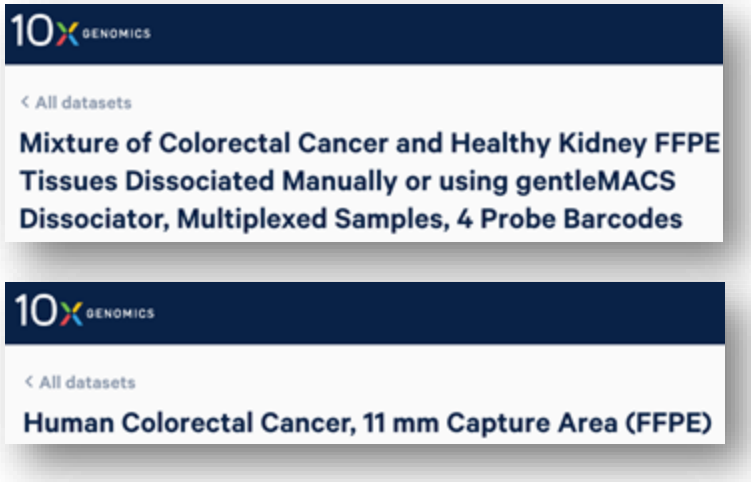


Single-cell and spatial data from the same FFPE sample



The data we are using today

- Formalin-fixed paraffin-embedded (FFPE) specimen from Discovery Life Sciences
  - Diagnosis: Colorectal cancer, Adenocarcinoma
  - Stage (AJCC): II-A
  - TNM system: T3
  - FFPE Block Age: 2 years
  - DV200: 67
- Paired Single-Cell Gene Expression Flex and Visium Spatial Gene Expression assay
  - Multiplexed with 3 other samples for single-cell assay

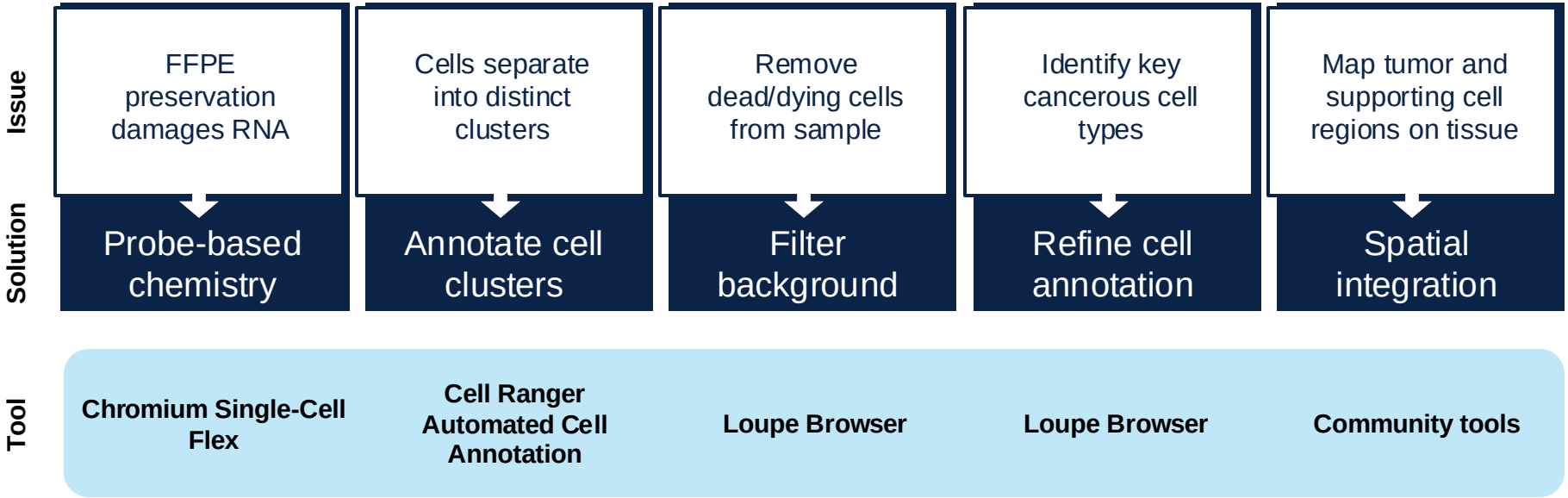


Single-cell:
www.10xgenomics.com/datasets/mixture-of-colorectal-cancer-and-healthy-kidney-ffpe-tissues-dissociated-manually-or-using-gentlemacs-dissociator-multiplexed-samples-4-probe-barcodes-1-standard

Spatial:
www.10xgenomics.com/datasets/human-colorectal-cancer-11-mm-capture-area-ffpe-2-standard

Clinical CRC sample analysis plan

Our journey through analysis



10X GENOMICS

10X Cloud: from FASTQs to quantitative biological information

Hands on tour of the single cell analysis journey

Learning objectives

We will use the 10x Genomics Cloud Analysis platform to perform the raw data processing for a FFPE sample with the intent to explore tumor cells and microenvironment.

We will demonstrate the following:

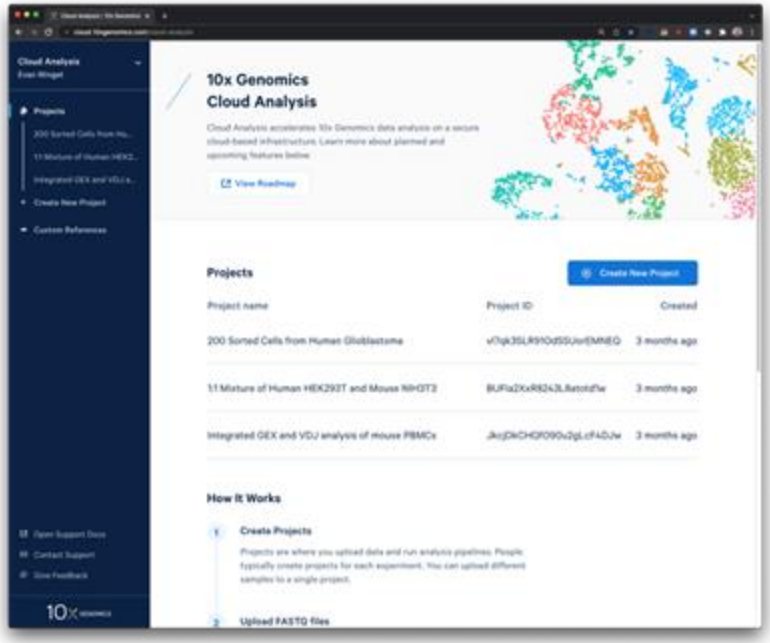
1. Signing in and creating a project
2. Using the command-line interface FASTQ uploader
3. Running initial data analysis
4. Perform automated cell type annotation
5. Downloading output files

Cloud Analysis

Fast and free analysis for every 10x Genomics sample

10xgenomics.com/cloud

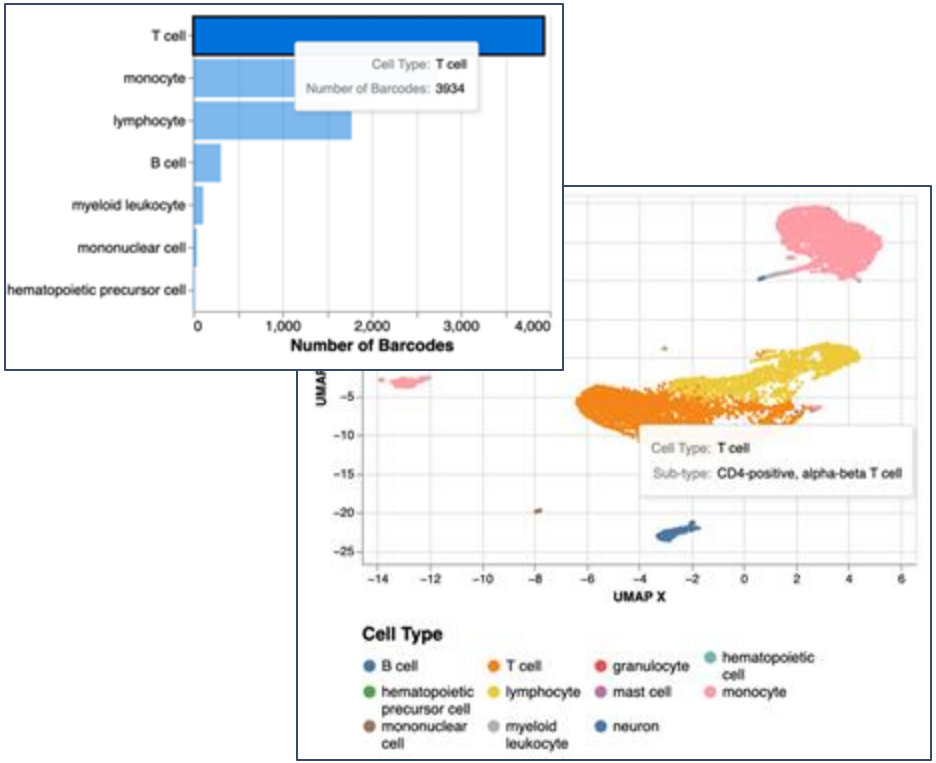
- Our recommended method for primary analysis of single cell Gene Expression and Multiome data for most new customers
- Lets you easily run the analysis from your web browser
- Get your results **quickly** with our fast and scalable cloud platform
- Receive free storage, analysis, and file downloads for each uploaded 10x dataset*



* See 10x Genomics Cloud Terms of Use for restrictions and details

Automated cell annotation now available in Cloud Analysis

- High level cell typing (e.g. T cell) creates an easy starting point for data analysis
- Instant visualization in Loupe Browser without manual steps
- More accurate insights with annotations from CELLxGENE database
- Model co-developed by 10x Genomics and the Cellarium AI Lab at the Data Sciences Platform of the Broad Institute



Datasets

Tiny FASTQ dataset

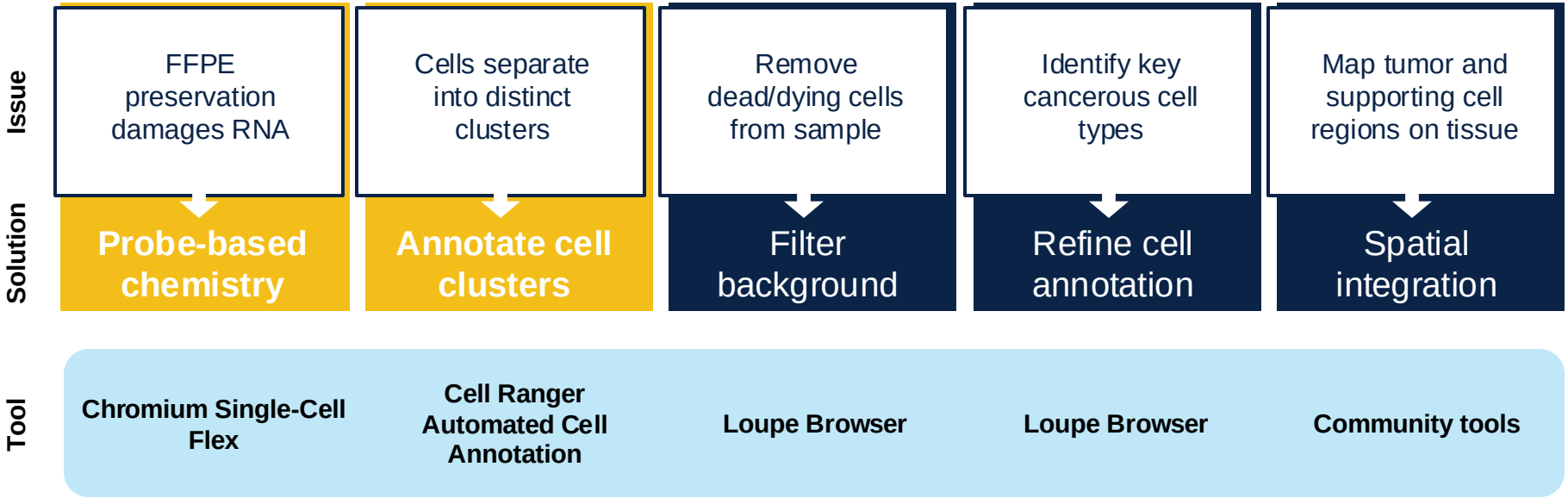
- Purpose:
  - Practice creating project
  - Practice uploading FASTQ
- Download link can be found in your welcome email or on the 10x Genomics Support site
  - cf.10xgenomics.com/supp/cell-exp/workshop_tutorial/tiny_gex.zip

Colorectal cancer dataset

- Purpose:
  - Run the analysis
  - Demultiplex datasets
  - Annotate cell clusters
  - Outputs will be used in later sections
- Files are in the project transferred to you.
  - Can also be accessed from 10x dataset page: www.10xgenomics.com/datasets/mixture-of-colorectal-cancer-and-healthy-kidney-ffpe-tissues-dissociated-manually-or-using-gentlemacs-dissociator-multiplexed-samples-4-probe-barcodes-1-standard

Clinical CRC sample analysis plan

Our journey through analysis



Interactive session

10x GENOMICS

Introduction to scRNA-seq data analysis

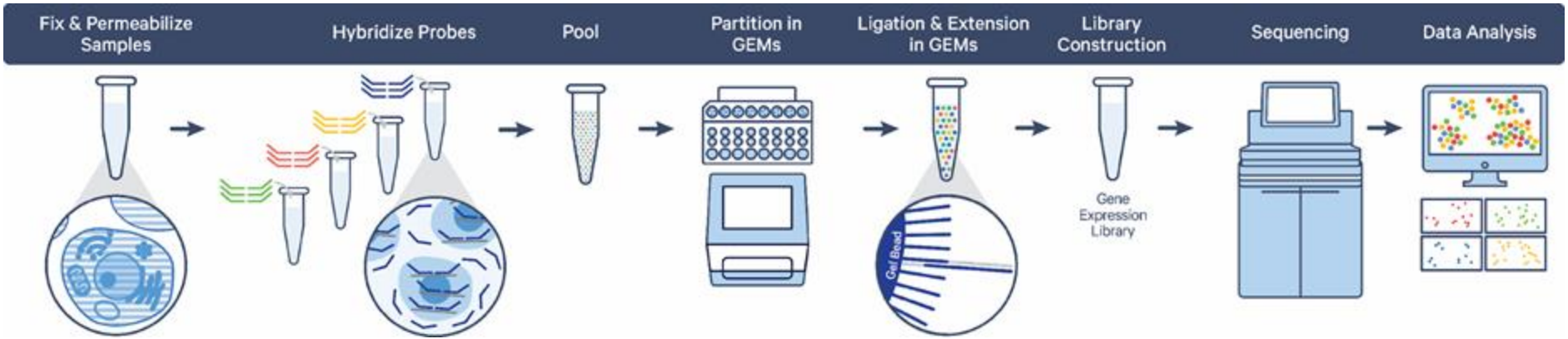
Hands on tour of the single cell analysis journey

Outline

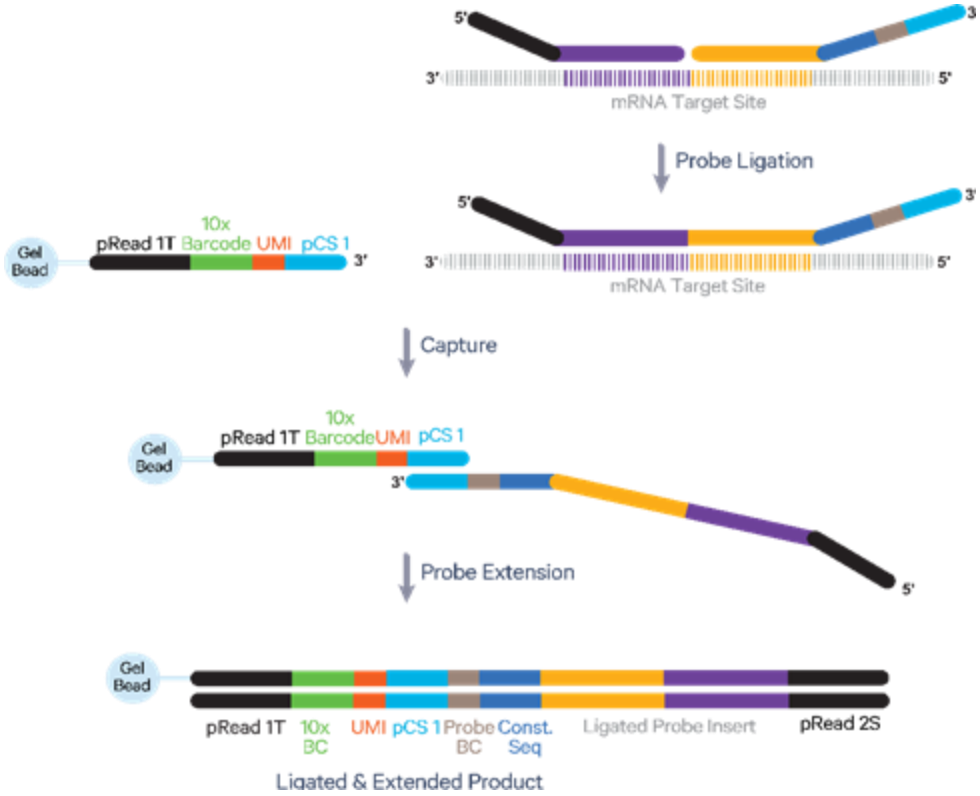
- Introduction to scRNA-seq
- Overview of Cell Ranger
- Where are in our CRC FFPE data set analysis

Introduction to Flex scRNA-seq data analysis

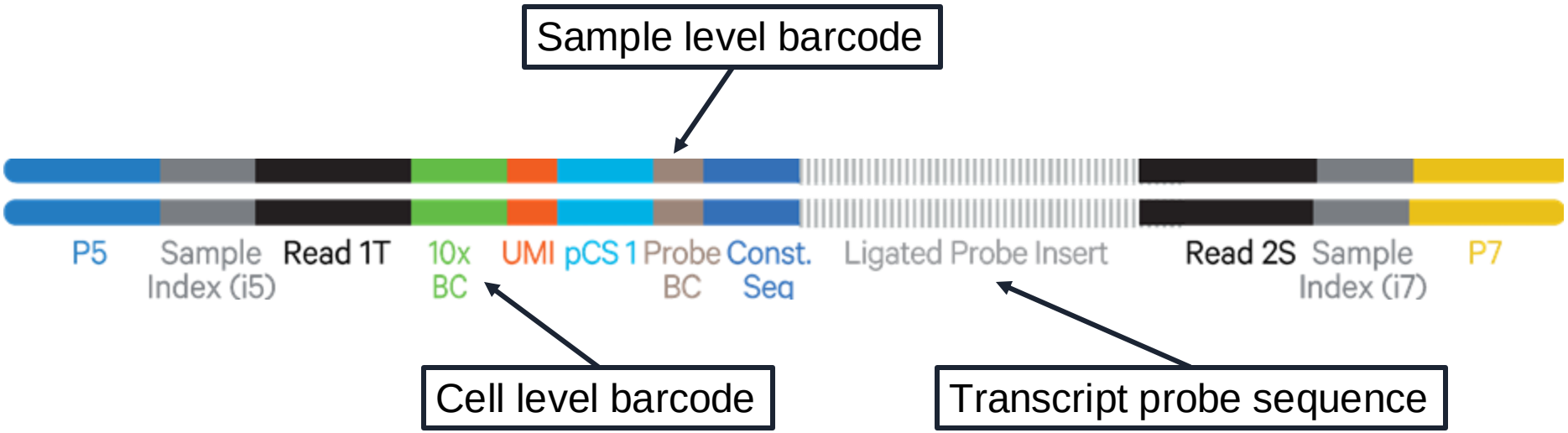
10x Genomics Flex Single Cell Gene Expression Assay



Fixed RNA probe-based capture

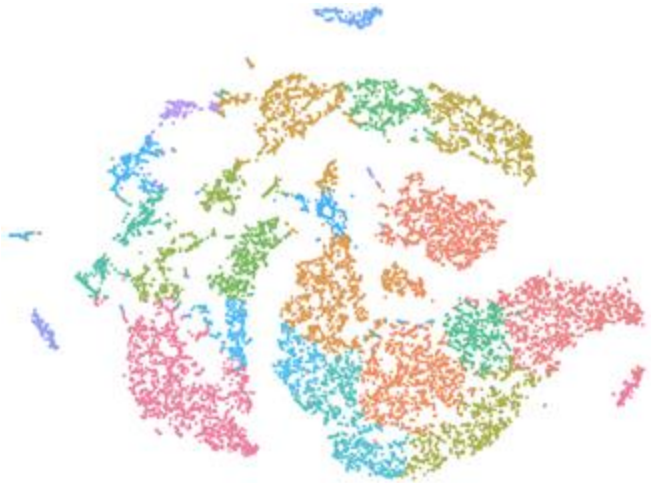


Flex Gene Expression library read structure



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

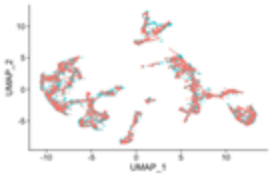
Single cell gene expression - data analysis flow

FASTQ files

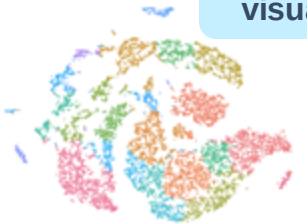
Read processing & quantification

QC & filtering

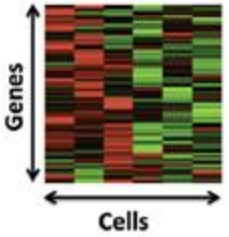
Data integration



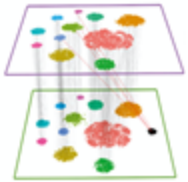
Clustering & visualization



Differential gene expression analysis



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

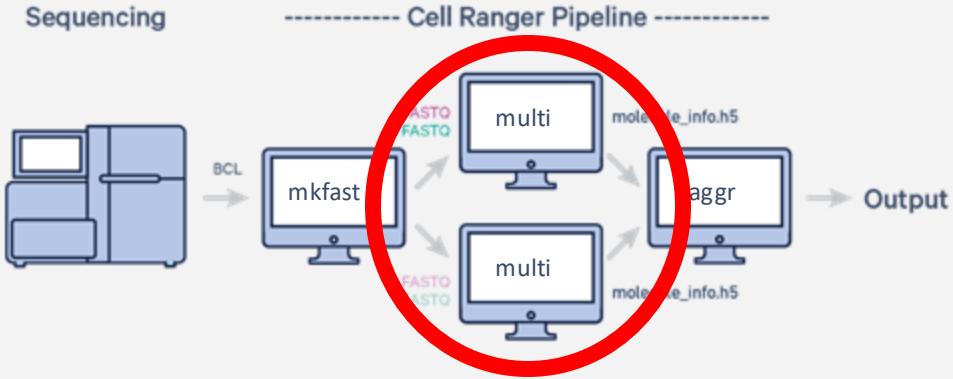
- Primarily programming libraries with some stand-alone tools (e.g. Seurat)
- Developed by the broader research community
- Not officially supported by 10x Genomics

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**)
  - Multiple sample analysis (**multi**)
    - Singleplex and Multiplex FLEX data is analyzed using multi
  - 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 experimental design

On a Linux system (i.e. Wynton Cluster)

Required skills:
- Understand 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

Running *cellranger multi* on the 10x Cloud

Command line

Example ‘multi’ command*

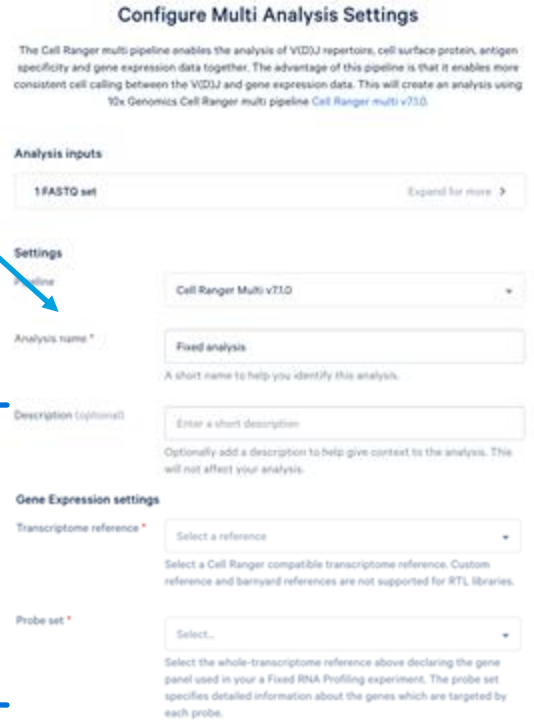
```
cellranger multi --id=OutputDirectory \
                 --csv=experiment.csv
```

Output directory / analysis name

Experiment details

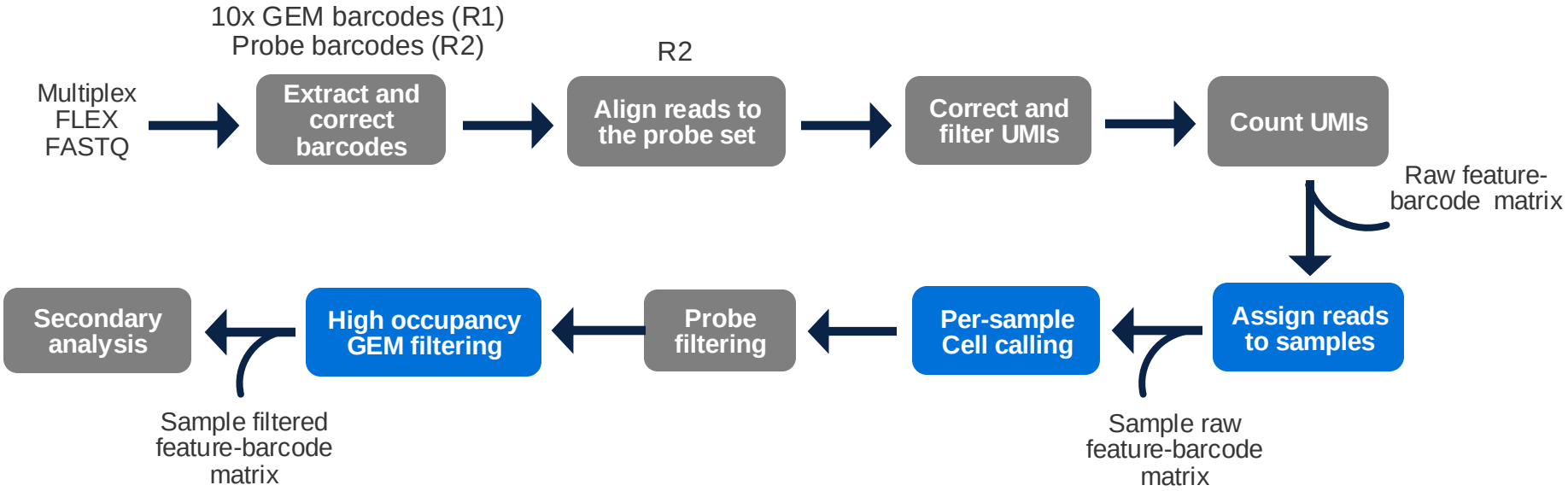
Run ‘cellranger multi --help’ for complete list of options

Cloud Analysis

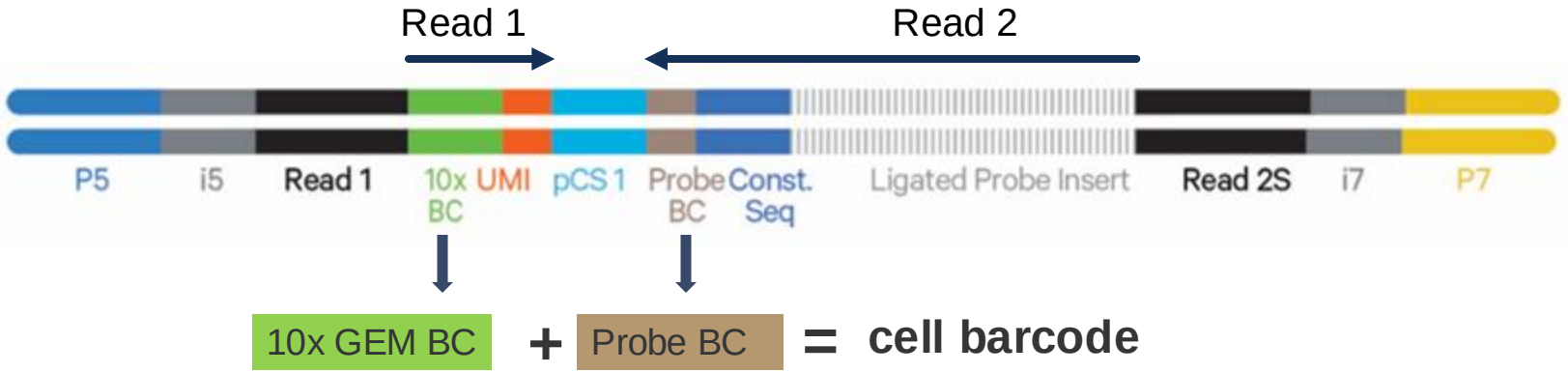


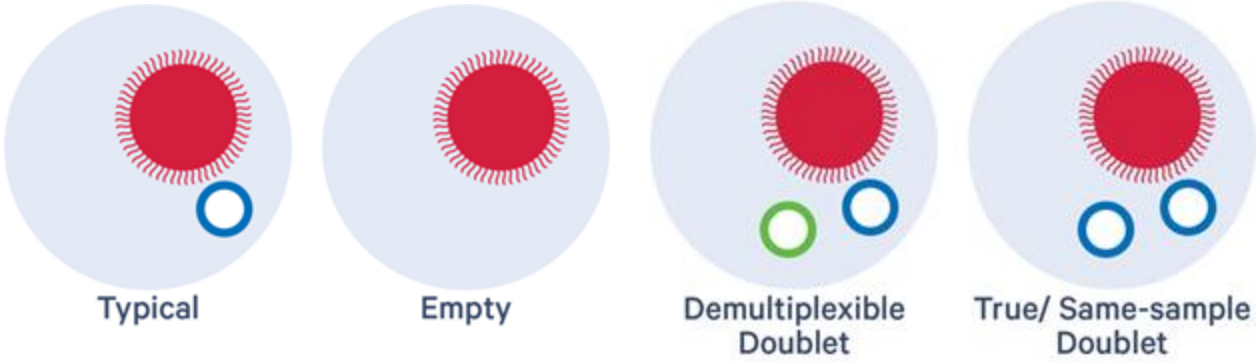
Overview of pipeline steps for multiplex Flex analysis

Analysis of Flex data is supported with Cell Ranger v7.0+ using ***cellranger multi***

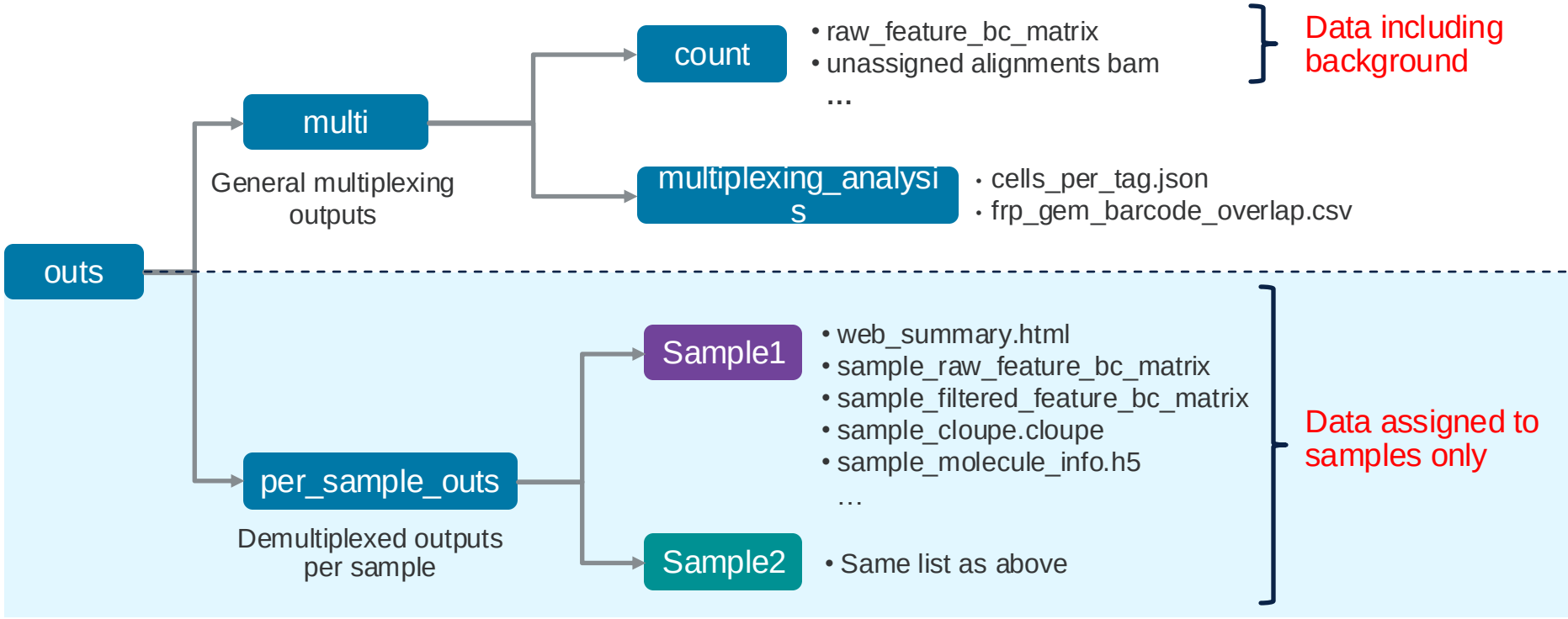


A cell is defined by the 10x GEM barcode + probe barcode





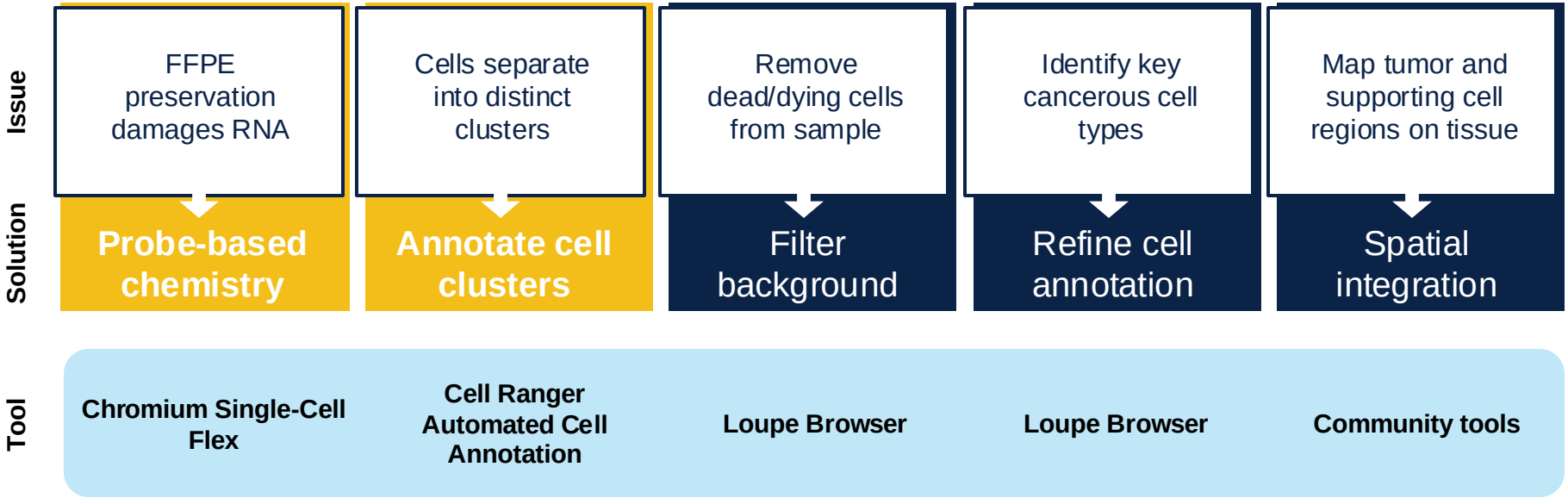
Output structure for multiplex Flex analysis



Where are we in our data analysis

Clinical CRC sample analysis plan

Our journey through analysis



Quality assessment: interpreting the Cell Ranger web summary

Hands on tour of the single cell analysis journey

Outline:

Slides presentation:

- Cell calling deep dive with barcode rank plot

Interactive demo:

- Navigating the web summary file
- Assessing metrics:
  - key metrics, sequencing metrics, mapping metrics, barcode rank plots, cell metrics

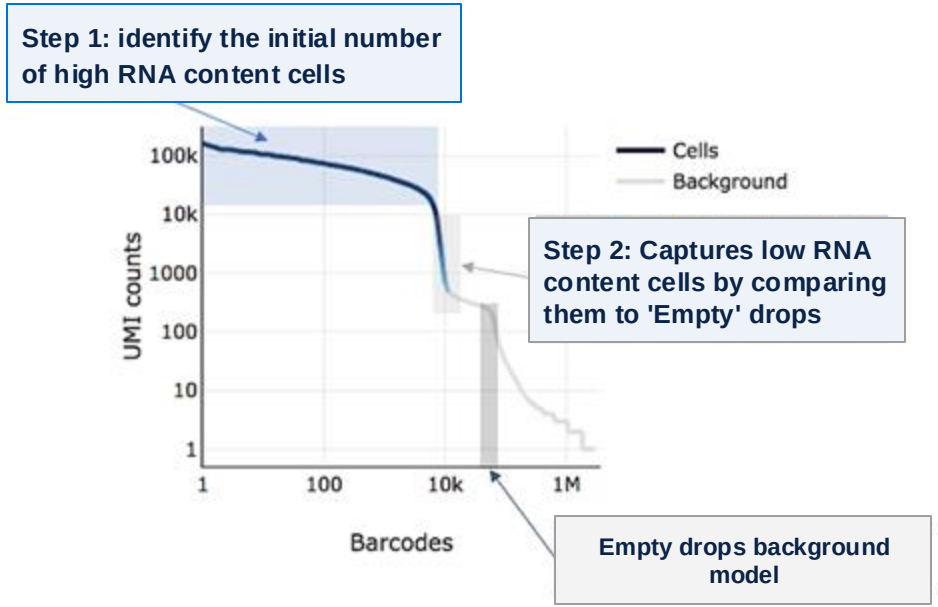
Interactive demo:

- Comparing results from FFPE/fixed cells vs. live cell capture

Summary of cell calling steps using the barcode rank plot

The cell calling algorithm can be broadly divided into two major steps:

- **Step 1:** Identify barcodes/GEMs that are likely to contain an intact cell based on the expected cell number and UMI counts.
- **Step 2:** Distinguish low RNA content cells from empty droplets based on the expression profiles using the EmptyDrops method.



Barcode Rank Plot

Step 2: Captures low RNA content cells

- Generate the ambient RNA profile from representative barcodes.
- Discard barcodes with total UMI count < 500 or < max of UMI count observed in the ambient background range 80-160k
- Call barcodes with RNA profile significantly different from ambient as cells.

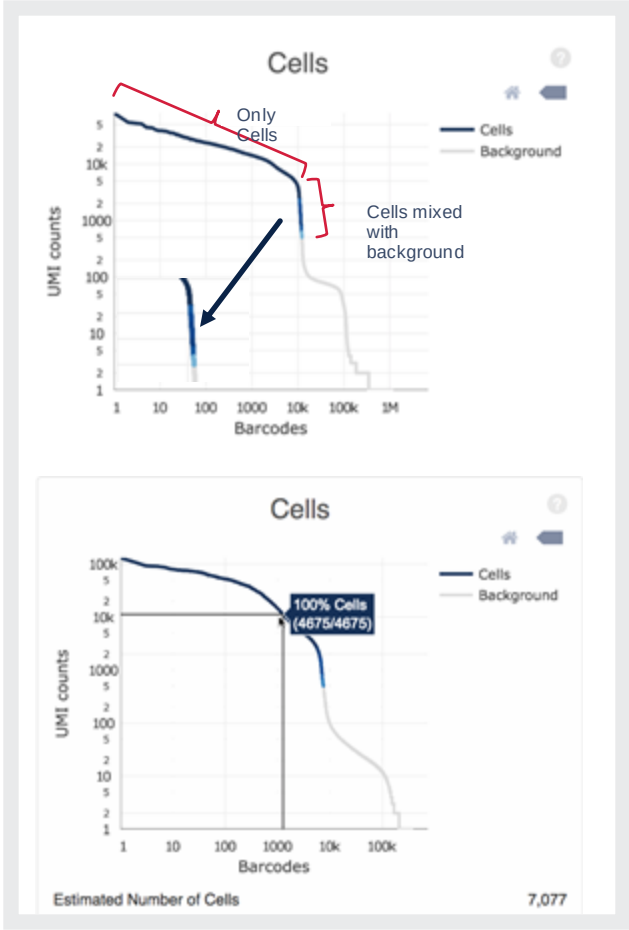
Empty drops background model range: 80-160k

Cell calling results

- Barcode rank plot shows distribution of barcode counts and which barcodes were inferred to be associated with cells.
- Blue color gradient is proportional to the fraction of cells in a given subset of barcodes

In the outputs:

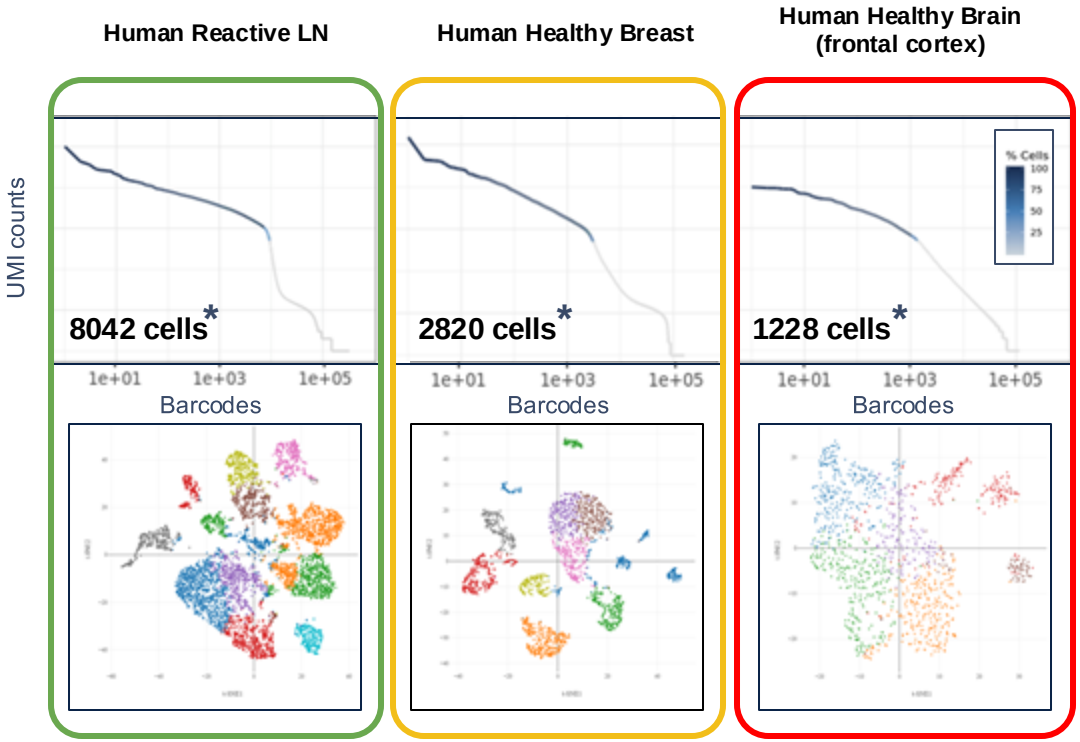
- Raw feature barcode matrix contains all barcodes in experiment
- Filtered feature barcode matrix only contains cell-associated barcodes



Barcode rank plots for FFPE samples

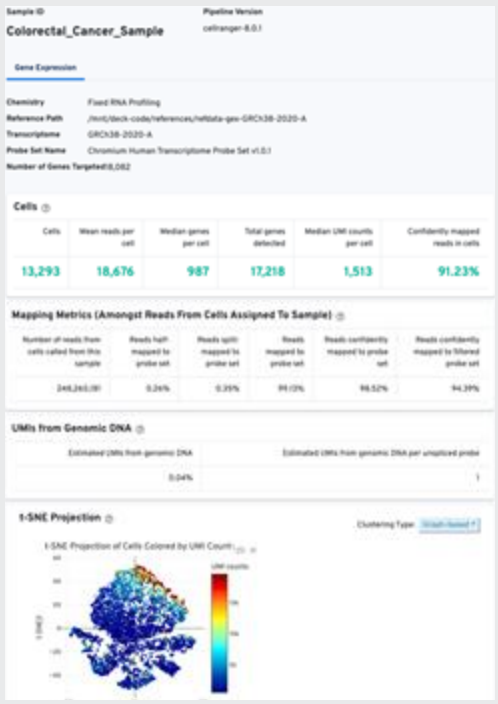
When assessing FFPE samples, consider:

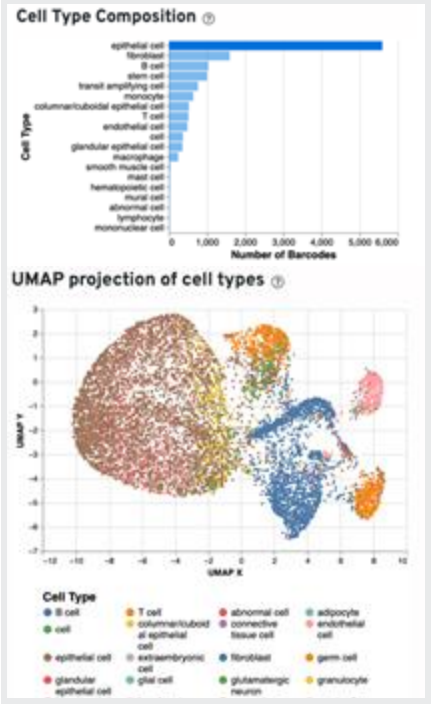
- How steep is the knee of the barcode rank plot?
- Is the cell recovery similar to expectations?
- What is the average UMI in called cell vs non-cell barcodes?
- Are there obvious clusters in the tSNE/UMAP?



* 10k cell input (Singleplex)
We recommend maximizing cell input for scFFPE workflows

Our data





Our sample is one of four samples multiplexed into one library: “Colorectal_Cancer_Sample”.

We used cellranger multi to demultiplex and analyze each sample.

Next steps:

1. Use the web summary to determine the quality of our sample and library
2. Explore called cell types from the automatic cell annotation pipeline

Considerations:

1. What was our targeted cell recovery?
2. Where does the “cliff” in the barcode rank plot taper off?
3. Does the cell composition roughly match our expectations of this tissue type?

Interactive sessions

10x GENOMICS

Loupe analysis: Quality control and rapid data exploration

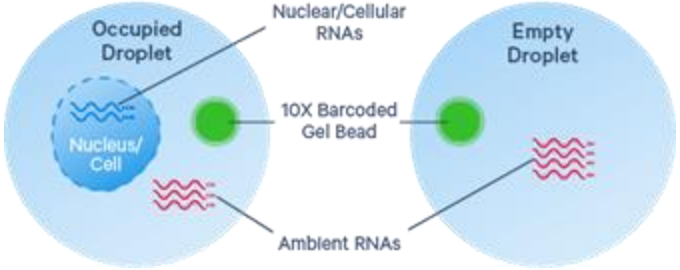
Hands on tour of the single cell analysis journey

Outline

- Overview on common QC steps and considerations
- Additional options for cell type annotation
- Hands-on

Quality control and assessment

Common QC steps for scRNA-seq



Filtering cell barcodes by UMI counts or number of features

Filtering cells by percent of mitochondrial (mt) reads

Removing empty droplets based on the expression profile

Doublet detection and removal

Removing ambient RNAs associated with barcodes

Cell Ranger, Loupe Browser

Community tools

Quality control · Introduction · Oct 26, 2022

Common Considerations for Quality Control Filters for Single Cell RNA-seq Data

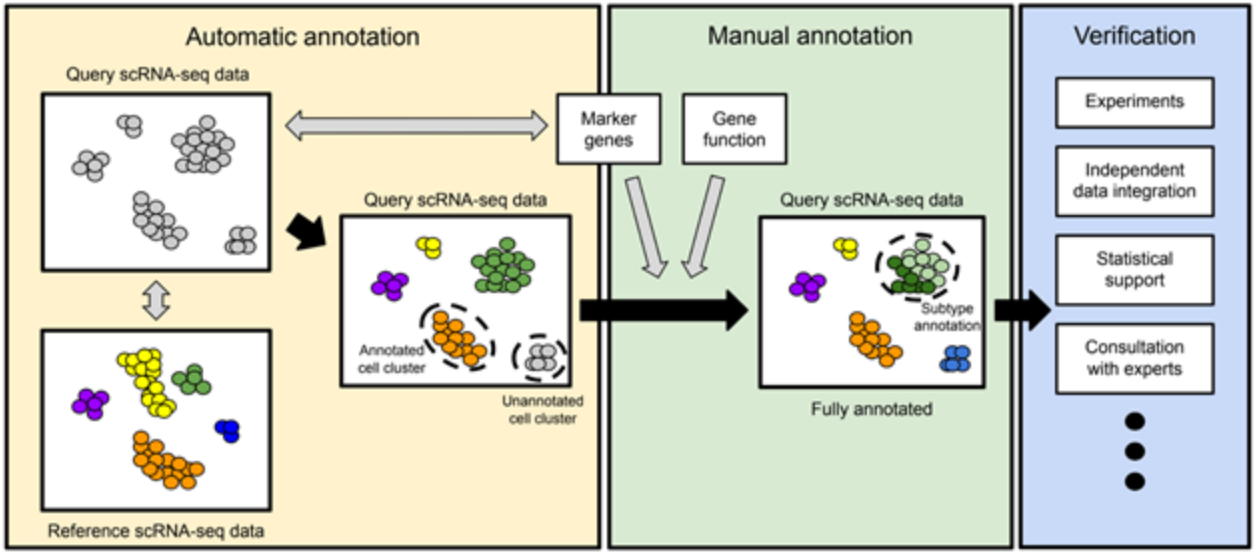
Quality control · Tutorial · Sep 19, 2022

Background Removal Guidance for Single Cell Gene Expression Datasets Using CellBender

Quality control · Introduction · Apr 24, 2023

Introduction to Ambient RNA Correction

Additional tools for cell type annotation



https://www.nature.com/articles/s41596-021-00534-0

Cell annotation Introduction Jun 28, 2023

Web Resources for Cell Type Annotation

The advantage of visualizing cell populations at single-cell resolutions has introduced us to the challenges of annotating cells. A growing number of annotation databases and tools are available to aid us in the process. In this article, we provide general guidance for a selected few to assist you with your research.

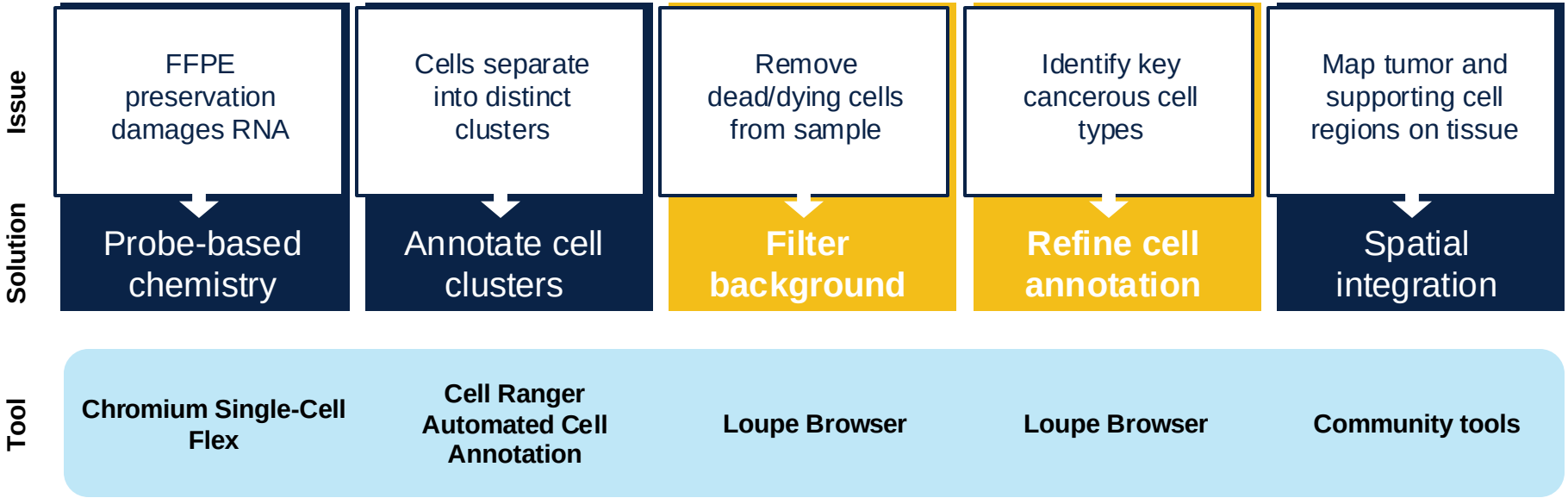
Cell annotation Tutorial Jan 25, 2024

Automated cell type annotation from R to Loupe using LoupeR

Step-by-step tutorial demonstrating how to annotate cell types using Azimuth (a reference-based annotation tool), convert the results into a .cloupe file using LoupeR, and explore the results in Loupe Browser.

Clinical CRC sample analysis plan

Our journey through analysis



Interactive sessions

10x GENOMICS

Further analysis using community developed tools: integration of single-cell and spatial data

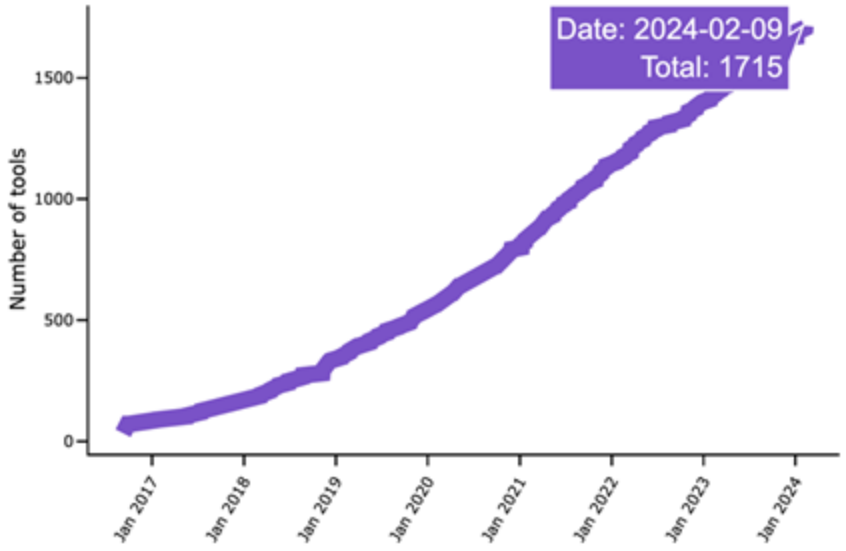
Hands on tour of the single cell analysis journey

Outline

- Approaching the community developed tools
- Recap on analysis plan
- Introduction to spatial gene expression analysis
- ?
- Hands-on?

Approaching the Analysis Ecosystem

Over 1,700 tools!



https://www.scrna-tools.org/

- How do you choose?
  - Leverage your research question
  - Look to the literature
    - Citations
    - Reviews
  - Look at GitHub sites
    - Check for regular updates
    - Check for issues and responses

- 10x Analysis Guides
  - Introductions
  - Tutorials
  - Informatics blogs

Community developed tools

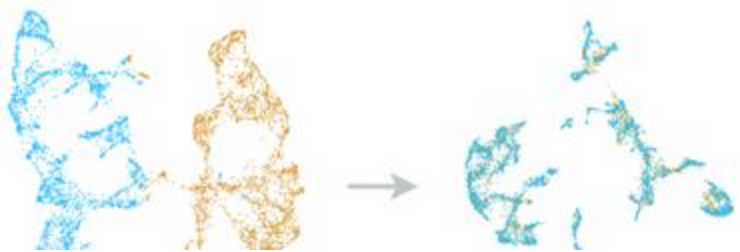
General Analysis Tools

Specialized Analysis Tools

Batch effect correction and data normalization

Community tools for quality control and assessment

Batch Effect Correction



Batch effects · Introduction · Jan 12, 2023

Batch Effect Correction

Batch effects · Blog Post · Oct 13, 2021

Publication highlight: Benchmarking scRNA-seq batch correction methods

Data normalization

- Log normalization (Seurat, Scanpy, Loupe)

$$LogNorm(feature, barcode) = ln(10000 * (\frac{feature\_count}{barcode\_count}) + 1)$$

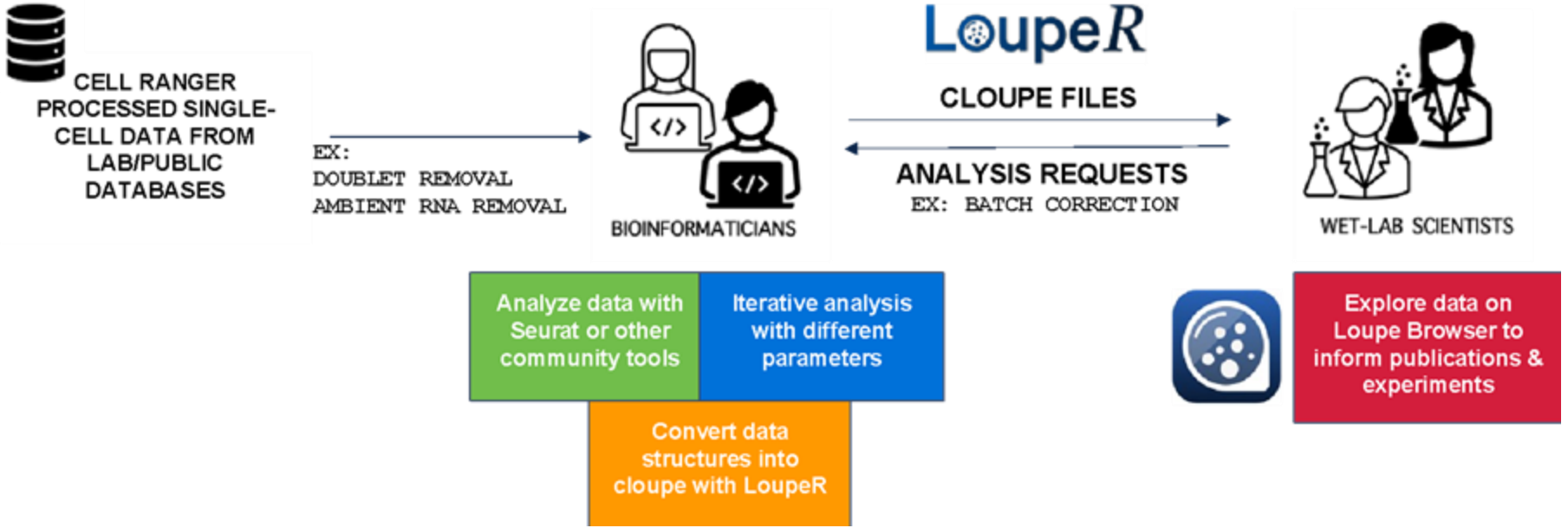
- SCTransform, etc.

Normalization · Introduction · Apr 18, 2023

Single-cell RNA-seq data normalization

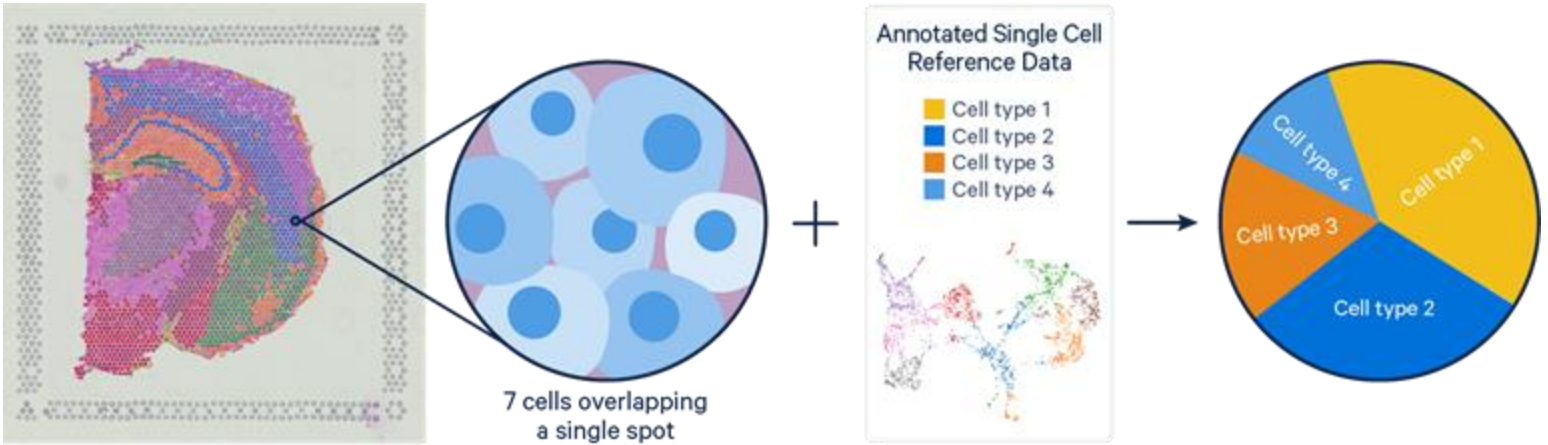
Use LoupeR to generate Loupe files from Seurat objects

Enable communication for complex bioinformatic analysis



[Tutorial: Automated cell type annotation from R to Loupe using LoupeR]

Integrating Single Cell and Visium Spatial Gene Expression



Tutorial: Integrating 10x Visium and Chromium data with R

Quick introduction to .ipynb and Colab

Jupyter Notebook (https://jupyter.org/)

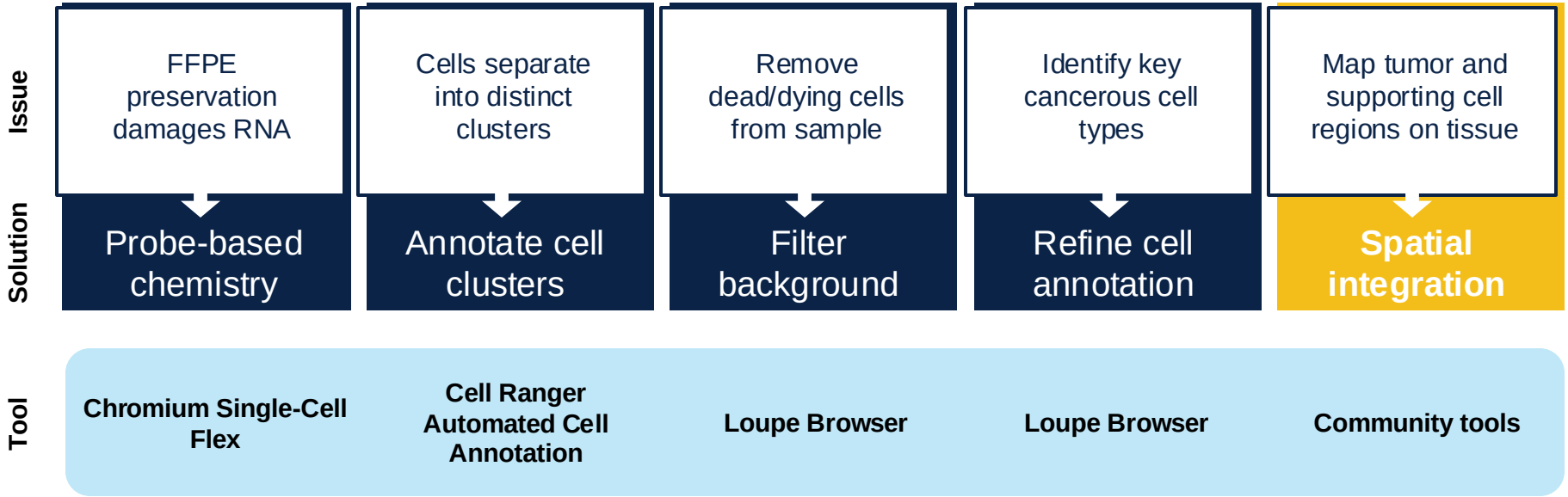
- Write and run code on a web browser
- Interactive: code, notes and outputs are all displayed “in-line”
- Supports many programming languages including R

Google Colaboratory:

- Jupyter Notebook stored in Google drive
- Requires no setup, and runs entirely (writing, running, & sharing code) on the Cloud

Clinical CRC sample analysis plan

Our journey through analysis



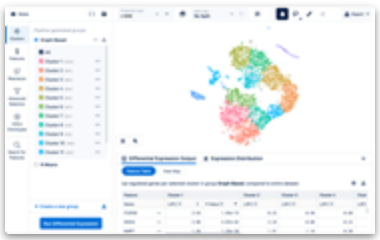
10x GENOMICS

Quick Recap

Hands on tour of the single cell analysis journey

Recap

Our journey through analysis



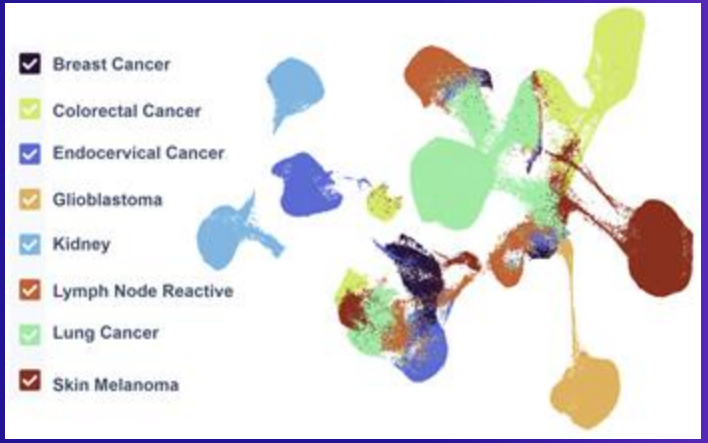
All the slides and material are available on this agenda page:
https://github.com/Lneves23/10xDataAnalysisWorkshop

Introducing GEM-X Flex

Unprecedented scale.
Ultimate flexibility.
Incredible cost savings.

GEM-X Technology: Powering the next generation of single cell

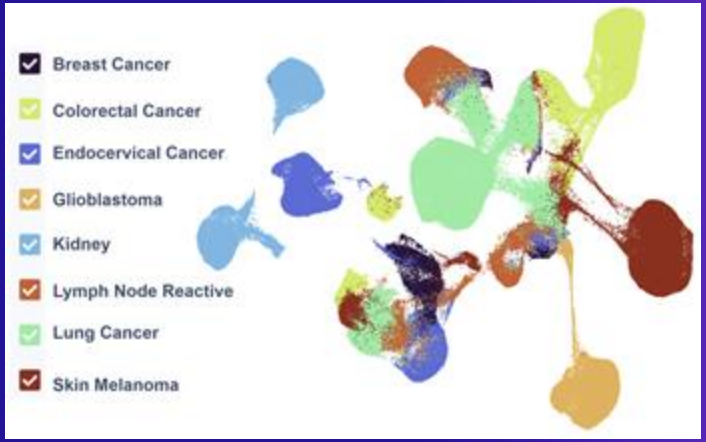
Introducing GEM-X Flex Gene Expression



- **4x lower cell input recommendation: 25K cells/sample**
- **Built to scale:** 2-fold increase in cell throughput
- **More cost effective:** >2-fold reduction in cost per cell
- **Enhanced data quality:** 2-fold reduction in multiplet rate
- **Maximum sample recovery:** during sample preparation (up to 80%) and single cell partitioning (up to 80%)
- **Improved assay robustness:** Redesigned microfluidics

High performance, low cost, mega scale

Product Specifications for new GEM-X Flex Gene Expression

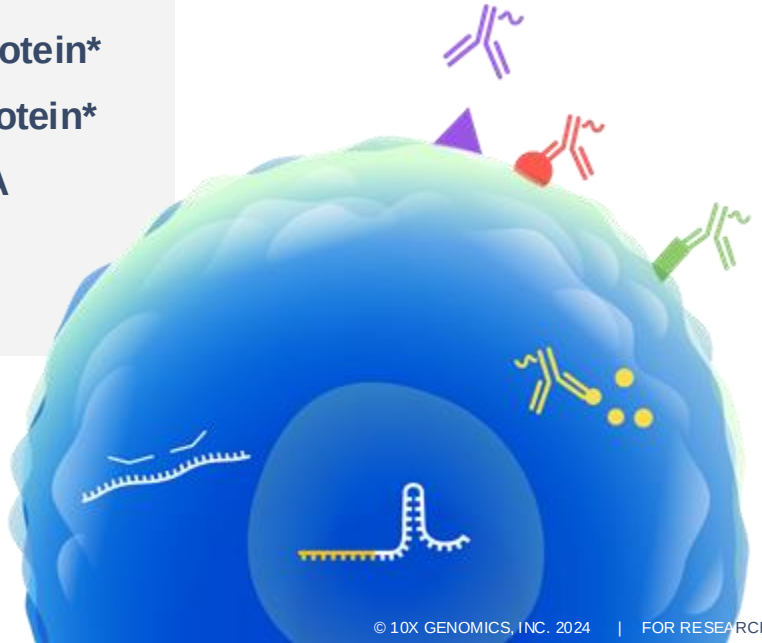


- **Efficiently partition millions of cells in less than 6 minutes**
- **Built to scale:** Up to 320,000 cells per channel, for up to 2.56M cells per run, run up to 128 samples in parallel
- **Cell size flexibility with no lower limits**
- **High cell capture rates up to 65%**
- **Low doublet rates of 0.4% per 1,000 cells**

Comprehensive multiomic analysis with GEM-X Flex

GEM-X Flex Gene Expression

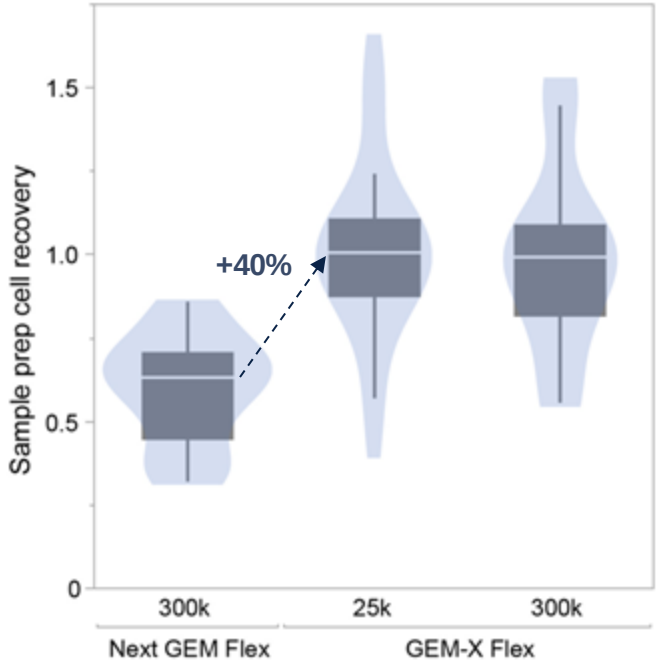
- Gene Expression
- Cell Surface Protein*
- Intracellular Protein*
- CRISPR / gRNA
  (with custom probes)



*Compatible with BioLegend TotalSeq™-C and Proteintech Genomics MultiPro™ products

Substantially improved sample preparation cell recovery

Sample preparation enhancements boost recovery, particularly with challenging samples like FFPE



- Sample preparation improvements result in **up to 40% increase** in sample prep cell recovery compared to Next GEM Flex
- Substantially improved recovery for dissociated tissue, FFPE, and nuclei

Lower cell input recommendations
(25K cells/sample)

Making fixed SC more accessible for limited samples

We are here to support you!

- Discussion with local Field Application Specialist and Science and Technology Advisor

Experimental Design + Lab Workflow

- Software training from FAS
- Support site documentation and software tutorials

Raw Data Processing

- Web Summary
- Data QC Technical Note and Analysis Guide

Data QC

- Support site documentation and software tutorials

Visualization and Interpretation

- 10x customer publications
- 10x-pert & customer webinars
- Open-source community tools
- 3rd party analysis services and products

Advanced Analysis and Results

ASK US!

support@10xgenomics.com