



Welcome to our hands-on tour of the Visium HD spatial data analysis journey

From 10x Cloud to Loupe to Community-Developed Tools

CG000807 RevA

Goals and objectives

Primary goal

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



Secondary goals

TL;DR: Helping you move forward with your analysis

- Help you navigate your analysis journey
- Highlight valuable resources like documentation and analysis guides
- Get you started with raw data processing using 10x Cloud Analysis
- Showcase Loupe Browser's expanded functionality
- Introduce the incredible ecosystem of community-developed tools for 10x data analysis
- Foster a local analysis community

Agenda

Morning	Raw data processing and quality control	<ul style="list-style-type: none">• Introduction• Raw data processing with Cloud Analysis• Break• Overview of Visium spatial data analysis• Quality assessment• Manual image alignment
Noon	Lunch	
Afternoon	Data visualization and downstream analysis	<ul style="list-style-type: none">• Data visualization and exploration with Loupe Browser• Break• Introducing community-developed tools for Visium data analysis• Further analysis using community tools: GSEA as an example• Wrap up and optional chat / office hour with the 10x team

Introduction to the Experiment: Analysis Begins at Design

Outline

- Analysis journey begins at design
- Background on the sample we are working with today
- Introducing spatial transcriptomics technology
- Purpose of the experiment
- Analysis plan

Introducing analysis journey

Experimental Design

Analysis begins with design. Choose assay types and define controls & replicates to set your course.

Lab Workflow

Use analysis-aware execution to avoid road-blocking technical artifacts.

Raw Data Processing

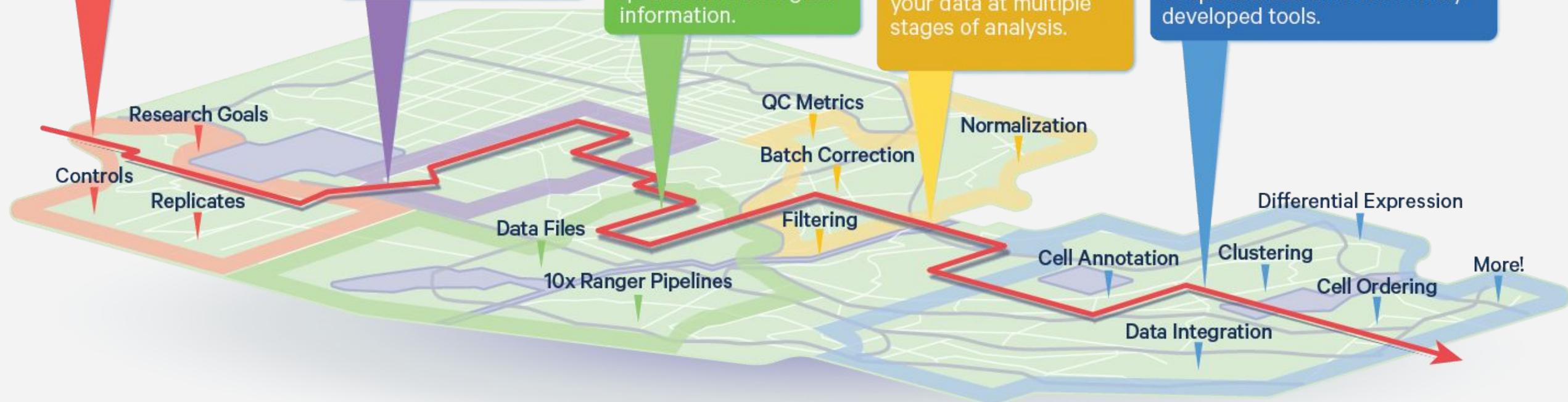
Transform your sequencing data into quantitative biological information.

Quality Control

Assess the quality of your data at multiple stages of analysis.

Visualization & Interpretation

Arrive at biological insights and address research questions with Loupe Browser and community-developed tools.



Analysis begins at design

- The design enables the analysis
- What samples, conditions, and data are needed to answer your research question?
- The planning call is the most cost-effective part of the experiment.

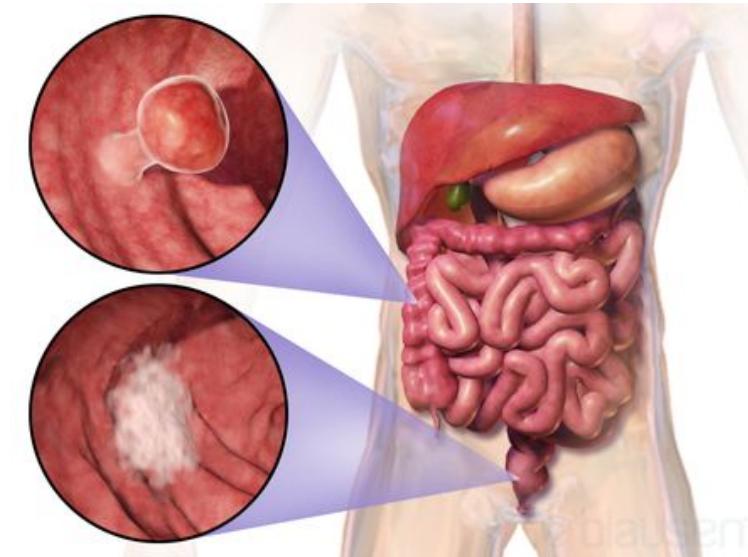


<https://imgflip.com/memegenerator/One-Does-Not-Simply>

Background on the sample we are working with today

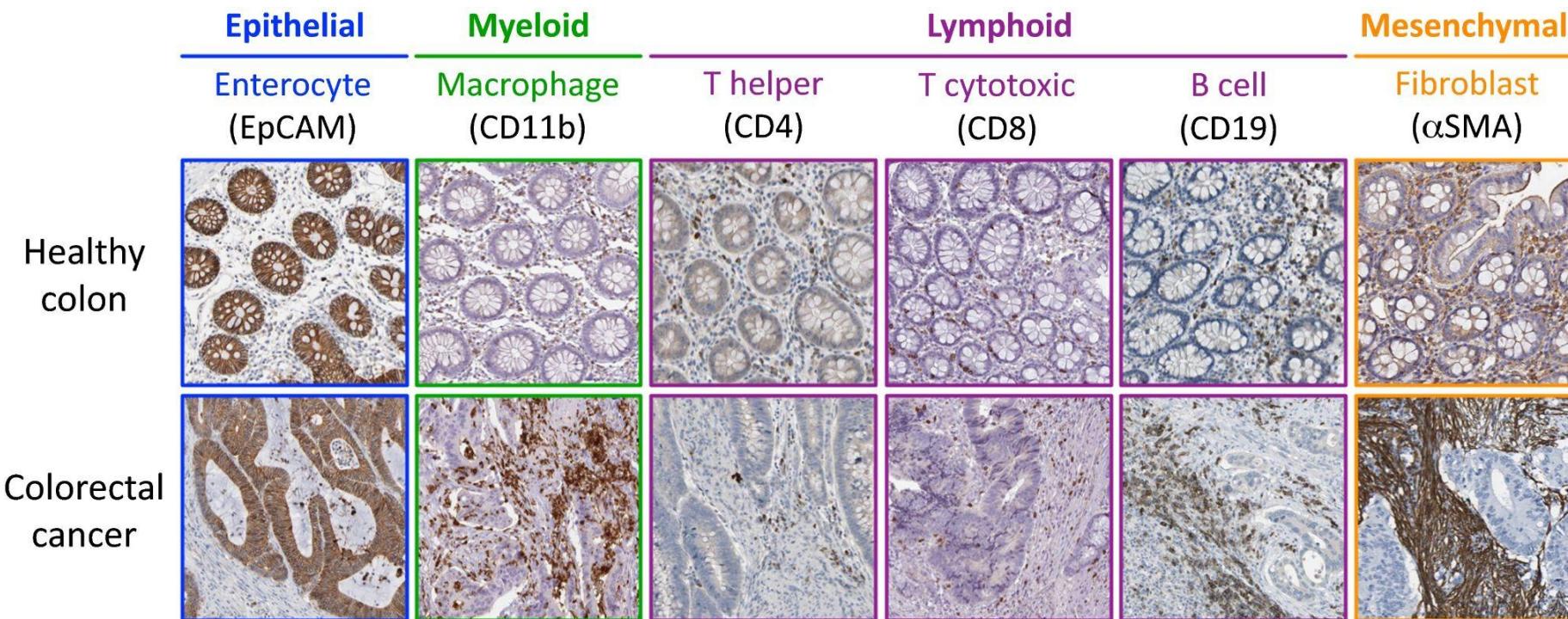
Colorectal cancer (CRC)

- CRC is the second-deadliest cancer in the world, accounted for 0.9 million deaths in 2020.
- Its global incidence is predicted to double by 2035.
- Poor overall 5-year survival rate.
- Though single-cell profiling has revealed much about the transcriptional profiles of these tumors, **the complexities of the tumor microenvironment remain poorly understood.**



Two example colorectal tumors
(https://en.wikipedia.org/wiki/Colorectal_cancer)

Colorectal cancer is a heterocellular system



<https://www.cell.com/trends/cancer/fulltext/S2405-8033%2816%2930208-4>

Trends in Cancer

Better understanding of the tumor microenvironment have the potential to inform better clinical intervention strategies.

Spatial location matters

Bulk
RNA-seq



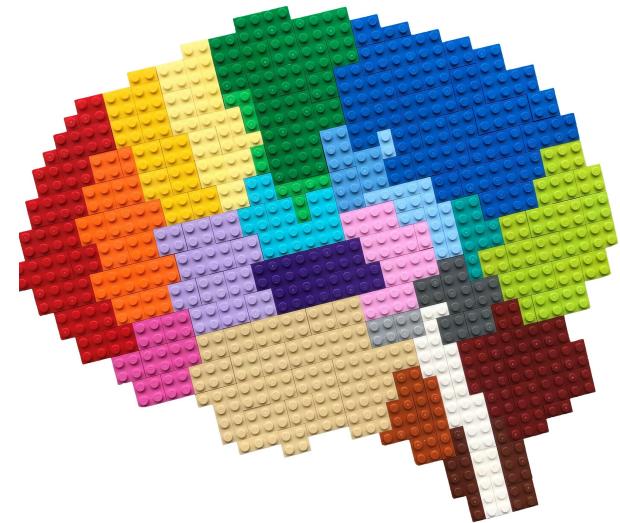
Get a general idea about gene expression but lacks cellular or spatial resolution

Single Cell Gene Expression



Characterize the cell types in a tissue

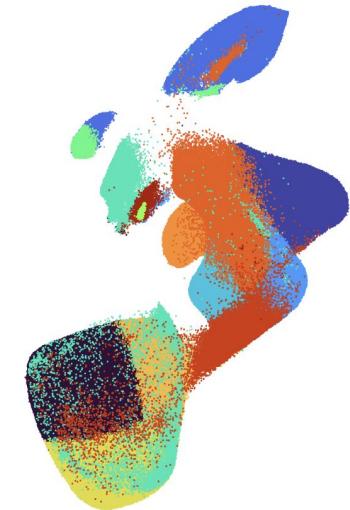
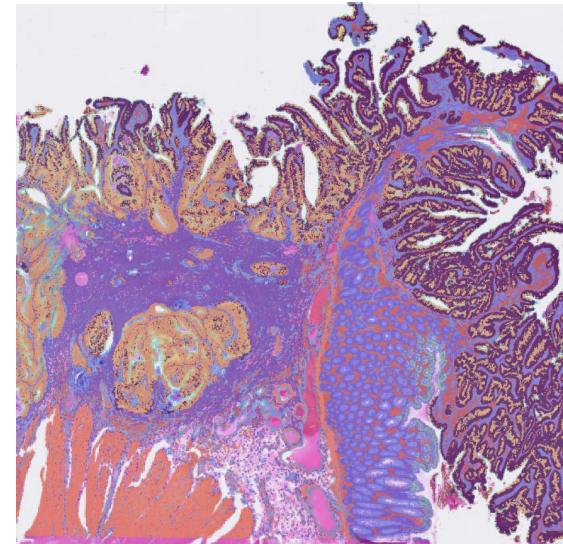
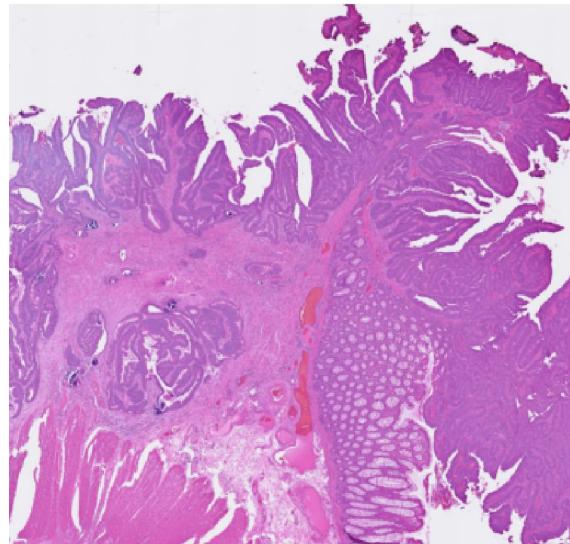
Spatial Gene Expression



Map the spatial patterns of gene expression

Two types of information make up Visium data

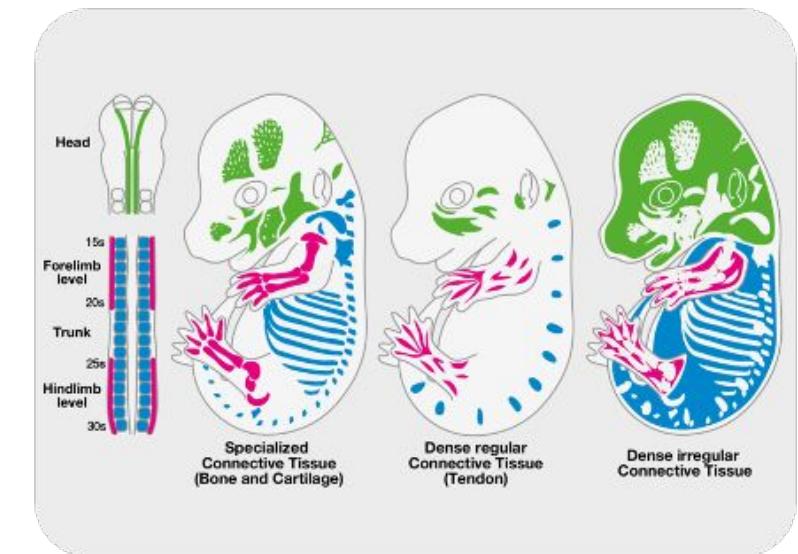
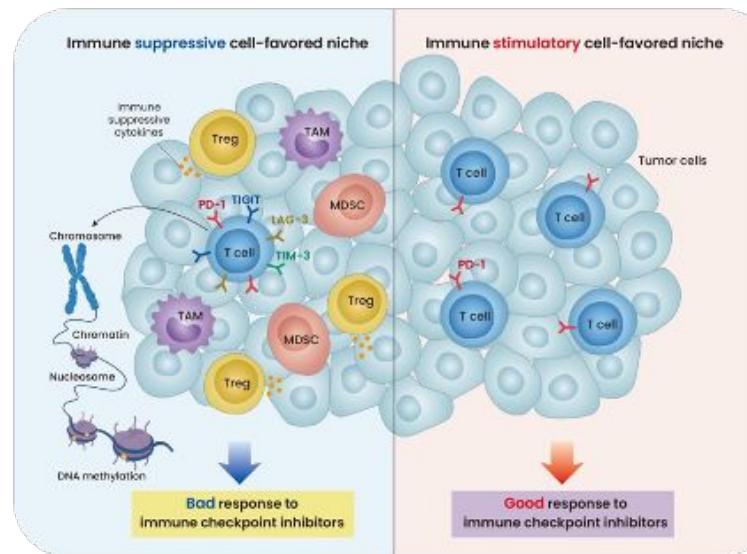
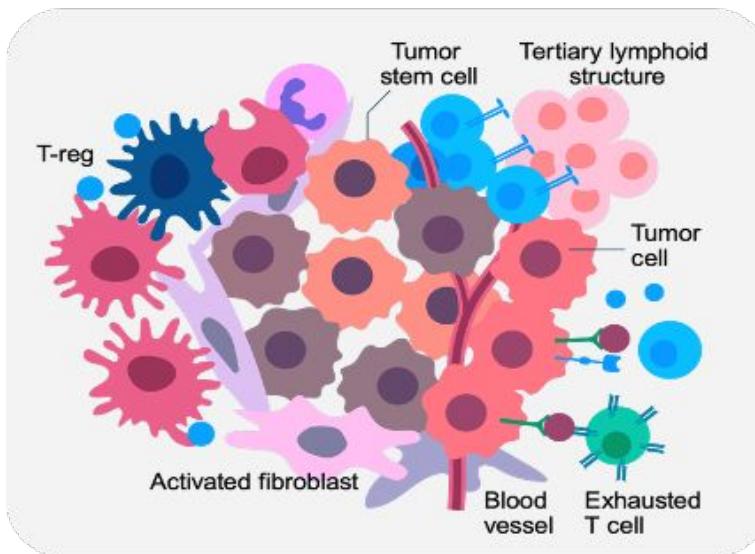
Tissue image and gene expression



The 10x Visium assays combine spatial context with gene expression data, providing researchers with a comprehensive method to study complex models

Some questions that 10x spatial technology can answer

Combining spatial and molecular information has tremendous promise for improving human health



How cell–cell interactions impact the microenvironment

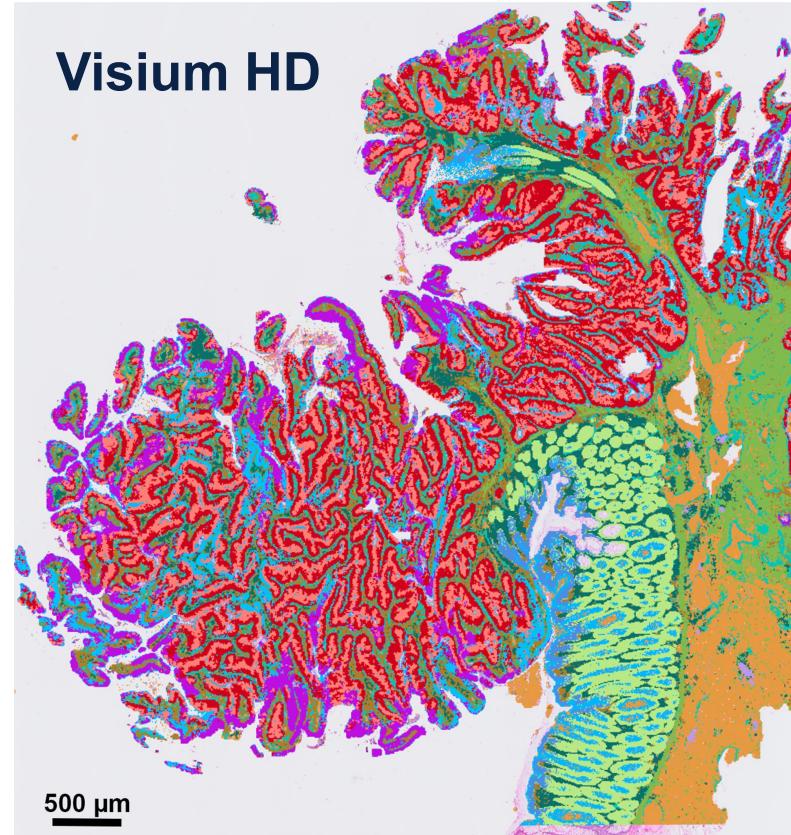
How immunological conditions are created and maintained

How spatial context influences development

Experimental goal

Goal of this experiment:

- Define cell types in a tissue section from a colon cancer patient.
- Explore the molecular functions of the cells surrounding a tumor region.



Visium HD is perfect for this unbiased spatial discovery study

Dataset

<https://www.10xgenomics.com/datasets/visium-hd-cytassist-gene-expression-libraries-of-human-crc>

New Results

 [Follow this preprint](#)

Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling

Michelli F. Oliveira, Juan P. Romero, Meii Chung, Stephen Williams, Andrew D. Gottscho, Anushka Gupta, Susan E. Pilipauskas, Syrus Mohabbat, Nandhini Raman, David Sukovich, David Patterson, Visium HD Development Team,  Sarah E. B. Taylor

doi: <https://doi.org/10.1101/2024.06.04.597233>

10X GENOMICS Products Resources Support Company Store Q Search

◀ All datasets

Visium HD Spatial Gene Expression Library, Human Colorectal Cancer (FFPE)

HD Spatial Gene Expression dataset analyzed using Space Ranger 3.0.0

Mapped 95.7% 

Assess data quality View summary metrics to assess data quality and more. [View summary](#)

 **Visualize and explore data** Discover differentially expressed genes, visualize your favorite genes, and explore your data with our visualization software. [Explore data](#)

 [Dataset overview](#)  [Output and supplemental files](#)  [Input files](#)  [Learn about Visium analysis](#)

Overview
This dataset is associated with the tutorial [Mapping the Tumor Microenvironment with Visium HD and Loupe Browser](#). Following this tutorial, you can learn how to:

- Review unsupervised clustering results
- Use goblet cell markers and the microscope image to visually assess transcript localization
- Map marker genes for tumor cells, cancer-associated fibroblasts, macrophages, and neutrophils to get a closer look at the tumor microenvironment
- Create a new group using known marker genes and perform differential gene expression to discover new markers
- Plot co-expression of tumor and fibroblast genes using Loupe's new feature

This dataset was generated using late stage in-development protocols, reagents, consumables, and instrument software. However, limited changes were made between this version and the final product, and we expect final data quality to be equivalent.

Biomaterials
A human colon cancer sample (FFPE) was obtained from Discovery Life Sciences.

- Colon region: Sigmoid
- Stage: Unknown
- Sex: Male
- Age: 60

Sample preparation

Product
HD Spatial Gene Expression

Software
Space Ranger

Subpipeline
spaceranger count

10x Instrument(s)
Visium CytAssist

Species
Human

Anatomical Entity
Colon

Preservation Method
FFPE

Disease State
Colorectal cancer

Biomaterial Type
Specimen from Organism, Imaged Specimen

Donor Count
1

Analysis plan

Our journey through analysis today

Process Raw Data	Assess Sample Quality	Map Cell Types	Recluster Region of Interest	Explore Enriched Pathways
Analyze Visium HD raw data with Space Ranger on 10x Cloud Analysis	Evaluate sample quality with the summary HTML file	Manually map specific cell types using Loupe Browser	Focused analysis in tumor and surrounding regions	Explore the molecular pathways in tumor microenvironment using GSEA



10x Cloud Analysis: From Raw Data to Quantitative Biological Information

Analysis plan

Our journey through analysis today

Process Raw Data	Assess Sample Quality	Map Cell Types	Recluster Region of Interest	Explore Enriched Pathways
Analyze Visium HD raw data with Space Ranger on 10x Cloud Analysis	Evaluate sample quality with the summary HTML file	Manually map specific cell types using Loupe Browser	Focused analysis in tumor and surrounding regions	Explore the molecular pathways in tumor microenvironment using GSEA

Learning objectives

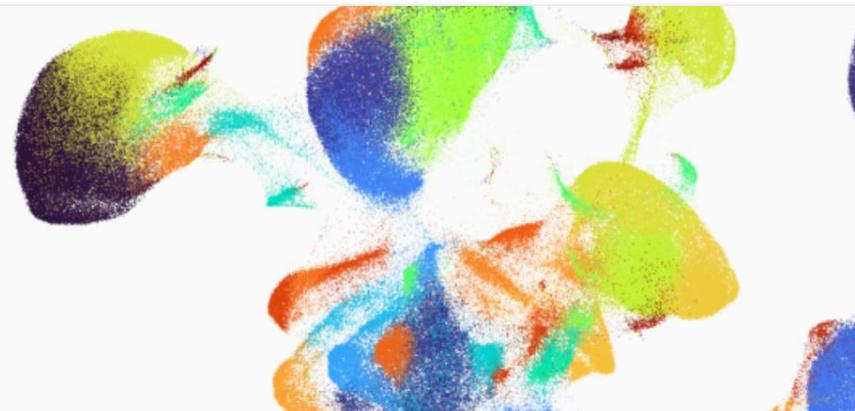
We will use the 10x Cloud platform to perform the raw data processing for a a human colorectal cancer Visium HD dataset with the intent to explore tumor microenvironment.

We will demonstrate the following:

1. Signing in and creating a project
2. Using the web based FASTQ and image uploader
3. Run data analysis with different parameter settings
4. Downloading output files

**10x Genomics
Cloud Analysis**

Upload FASTQ or image files after sequencing a 10x Genomics library, analyze the data using 10x analysis pipelines, and download the results.

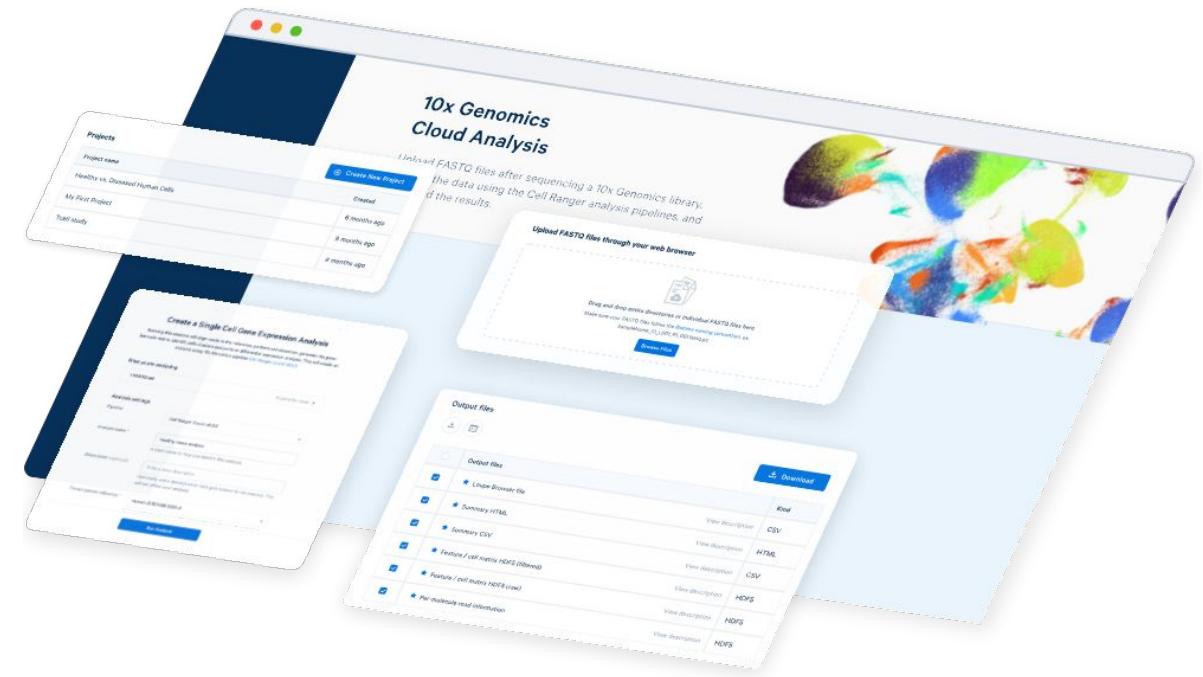


Cloud Analysis

Fast and free analysis for every 10x Genomics sample

- Our recommended method for primary analysis of Single cell Gene Expression, Multiome, and Visium HD 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.*

10xgenomics.com/cloud



* See [10x Genomics Cloud Terms of Use](#) for restrictions and details

How it works

Follow an easy 3-step workflow

Upload input files with a simple web interface or via CLI

Upload FASTQ files through your web browser

Drag and drop entire directories or individual FASTQ files here

Make sure the names of your FASTQ files do not contain special characters and follow the Illumina naming convention, ex. SampleName_S1_L001_R1_001.fastq.gz

Browse Files

4 files queued for upload

Total upload size: 3.29 GB Total estimated time: 19 minutes

File Name	Size	Action
1k_mouse_kidney_CNIK_3pv3_S1_L004_I1_001.fastq.gz	261.26 MB	Delete
1k_mouse_kidney_CNIK_3pv3_S1_L004_I2_001.fastq.gz	329.95 MB	Delete
1k_mouse_kidney_CNIK_3pv3_S1_L004_R1_001.fastq.gz	839.5 MB	Delete
1k_mouse_kidney_CNIK_3pv3_S1_L004_R2_001.fastq.gz	1.89 GB	Delete

Start Upload

[Input File Upload Guide](#)

Run Cell Ranger and Space Ranger with a guided web interface

Configure Multi Analysis Settings

The Cell Ranger multi pipeline supports the analysis of cell multiplexed data (Cell Ranger 6.0 and Space Ranger 1.0) and can run both analyses in parallel.

Create a Single Cell Gene Expression Analysis

Running this analysis will align reads to the reference, perform cell detection, generate the gene-barcode matrix, identify cells clusters and perform differential expression analysis. This will create an analysis using 10x Genomics pipeline Cell Ranger count v7.0.

Analysis inputs

1 FASTQ set

Settings

Pipeline: Cell Ranger count v 7.0

Analysis name: Enter a name

Description (optional): Enter a short description

[Analysis Creation Guide](#)

Download output for further analysis and interpretation

Upload FASTQ files through your web browser

Drag and drop entire directories or individual FASTQ files here

Make sure the names of your FASTQ files do not contain special characters and follow the Illumina naming convention, ex. SampleName_S1_L001_R1_001.fastq.gz

Browse Files

4 files queued for upload

Total upload size: 3.29 GB Total estimated time: 19 minutes

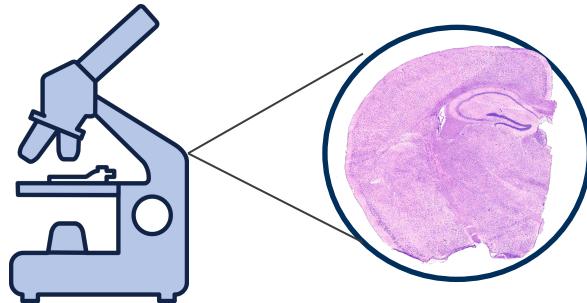
File Name	Size	Action
1k_mouse_kidney_CNIK_3pv3_S1_L004_I1_001.fastq.gz	261.26 MB	Delete
1k_mouse_kidney_CNIK_3pv3_S1_L004_I2_001.fastq.gz	329.95 MB	Delete
1k_mouse_kidney_CNIK_3pv3_S1_L004_R1_001.fastq.gz	839.5 MB	Delete
1k_mouse_kidney_CNIK_3pv3_S1_L004_R2_001.fastq.gz	1.89 GB	Delete

Start Upload

[File Download Guide](#)

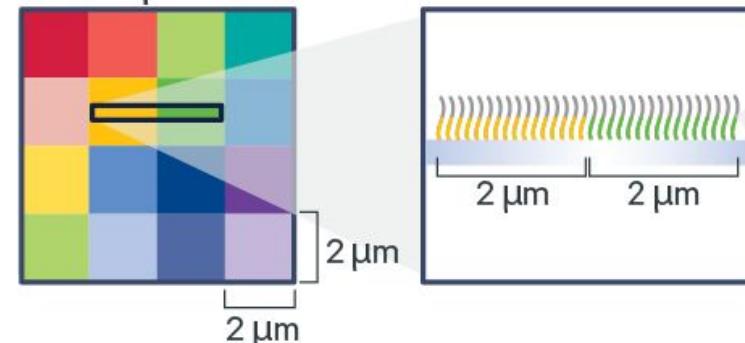
With Visium, we have two general types of data.

Image Data



Images and slide (grid) layout
(CytAssist and Microscope)

Transcript Data



Expression with the location encoded,
probe sequences and a reference genome

[Visium HD Spatial Applications Imaging Guidelines](#)

Datasets

Tiny FASTQ dataset

- Purpose:
 - Practice creating a project.
 - Practice uploading input FASTQ and CytAssist image files.
 - Get familiar with the 10x cloud.
- Download link can be found in your welcome email or on the agenda page.

Human colon dataset

- Purpose:
 - Run the analysis.
 - Explore different parameters and options for image alignment.
 - Outputs will be used in later sessions.
- Files are in the project transferred to you.
 - Can also be accessed from 10x dataset page:
<https://www.10xgenomics.com/datasets/visual-hd-cytassist-gene-expression-libraries-of-human-crc>

Interactive session

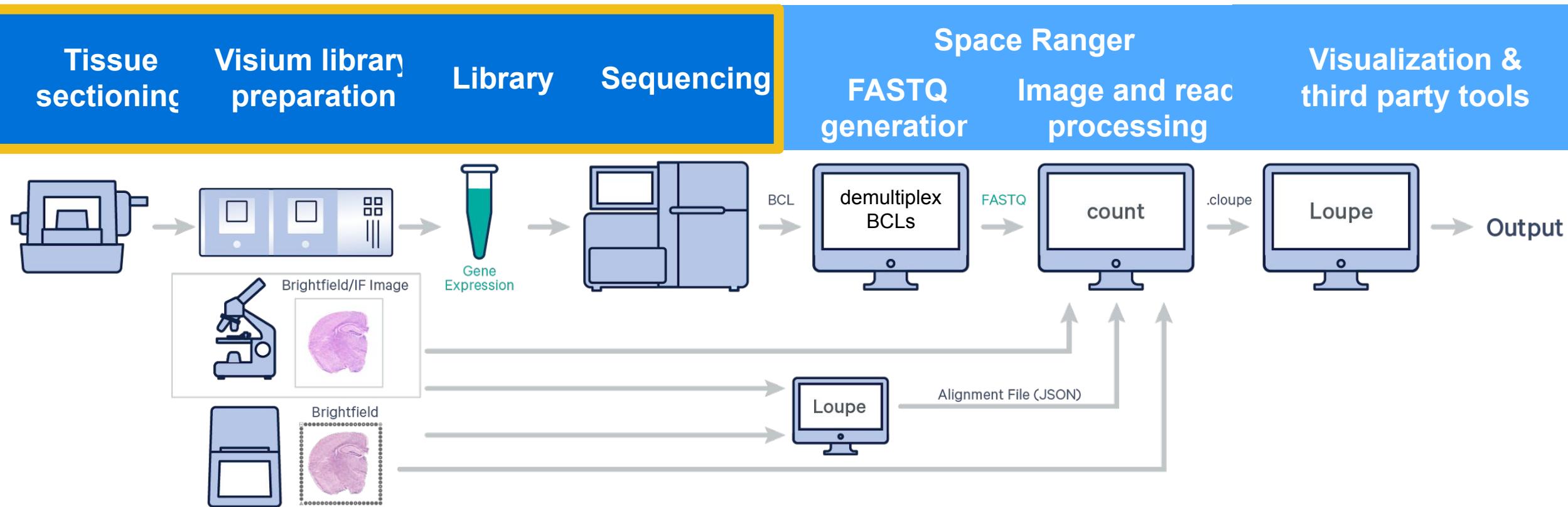


Overview of Visium Spatial Data Analysis

Outline

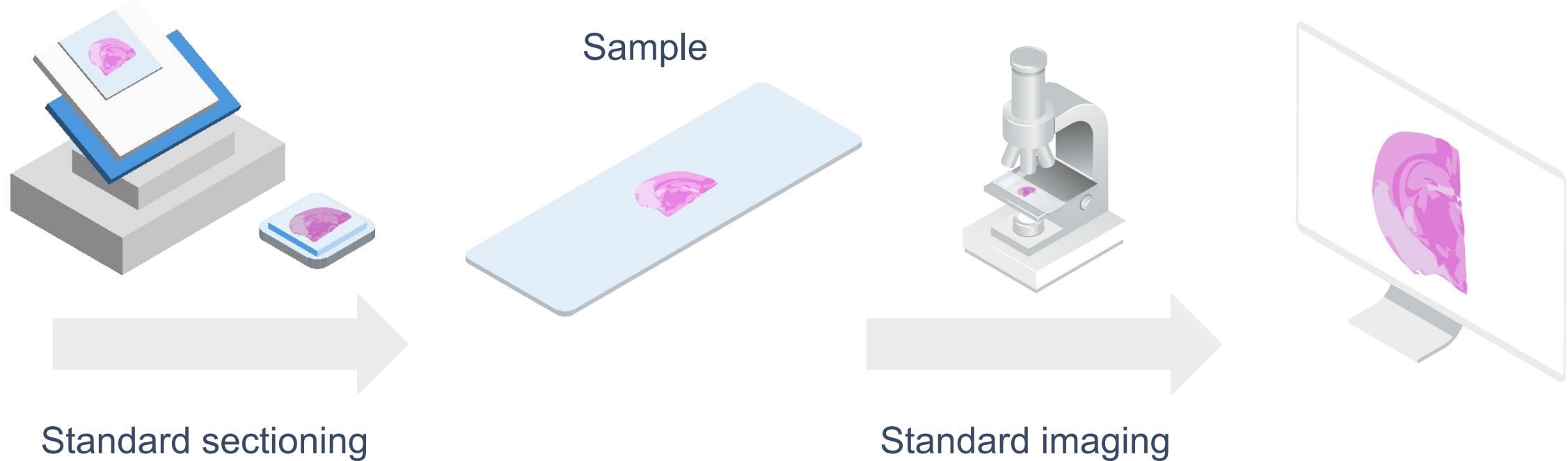
- Introduction to Visium workflow
- Overview of Space Ranger
 - Inputs
 - Pipeline overview
 - Outputs

Complete Visium workflow



Visium workflow – starts with standard glass slides

Tissue sectioning, Imaging



Input file to Space Ranger: High resolution microscope image (optional but recommended)

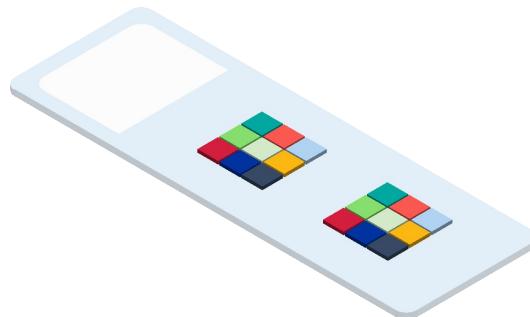
Visium workflow – library preparation

Visium library preparation with CytAssist workflow

Sample slides



Visium HD slide
(barcoding slide)



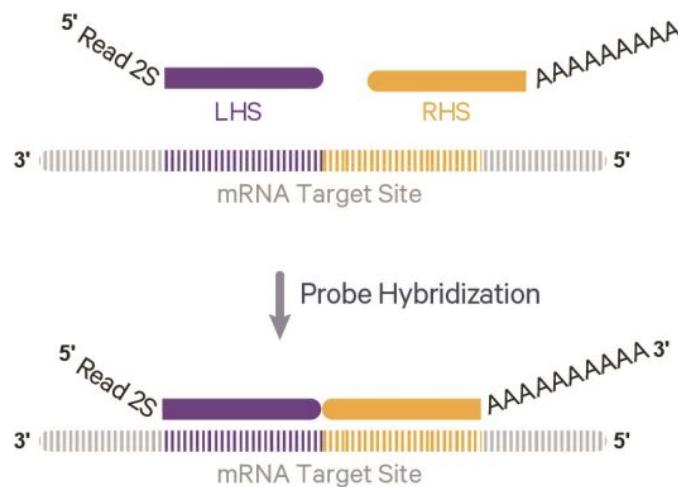
Automated barcoding
workflow

Target molecules are now
captured on the HD slide

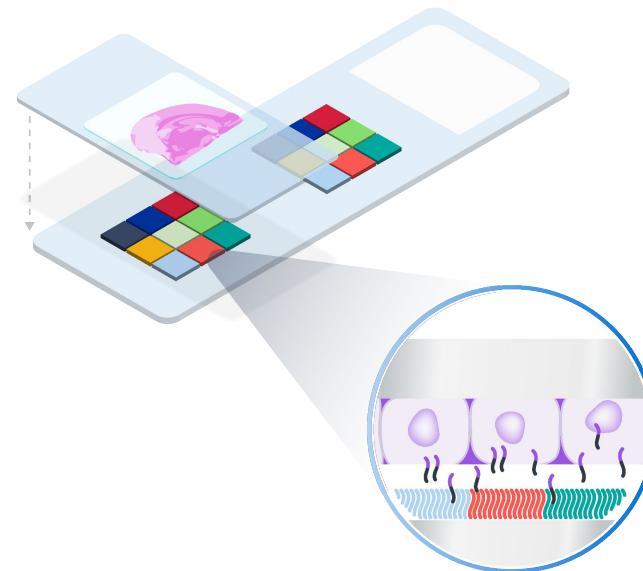
Input file to Space Ranger: CytAssist image, required

Visium workflow – library preparation

Probe hybridization & ligation



Probe capture (on CytAssist)

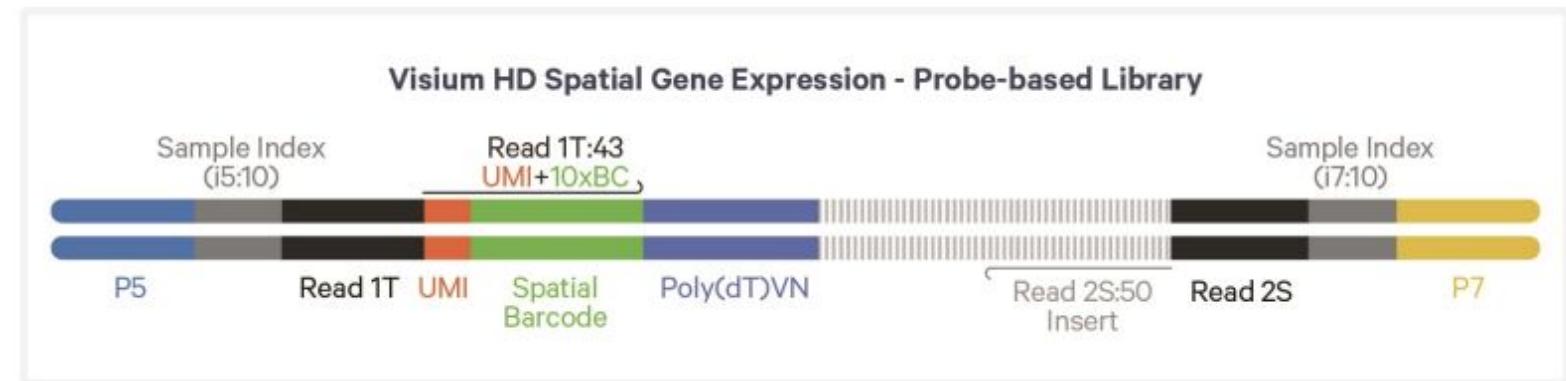


Input file to Space Ranger: FASTQ files from a compatible sequencer, required

Visium workflow – understanding the read structure

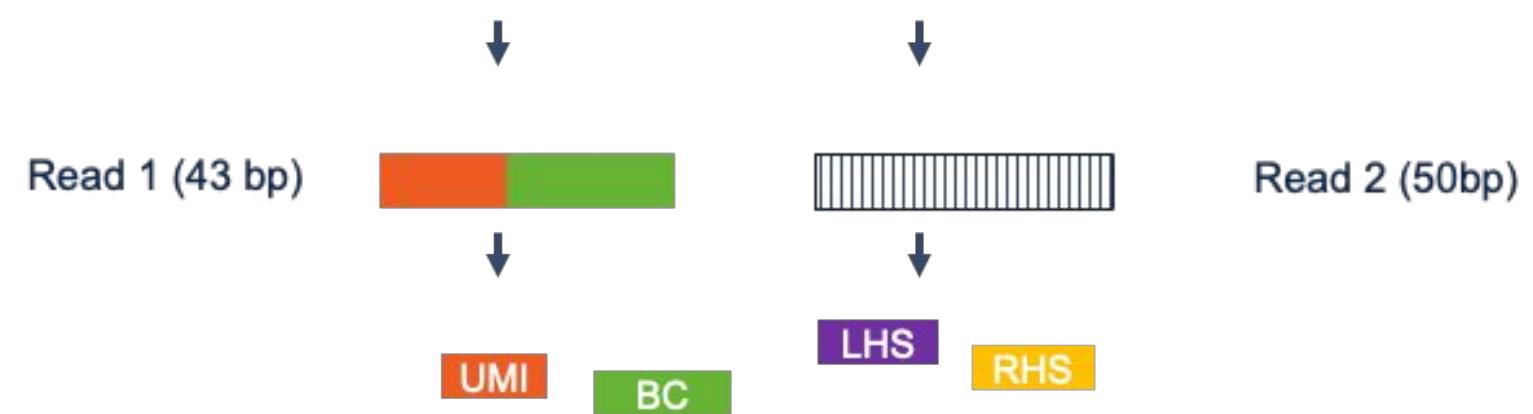
Sequencing and read structure

Read 1: UMI (for counts) + BC (spot position)



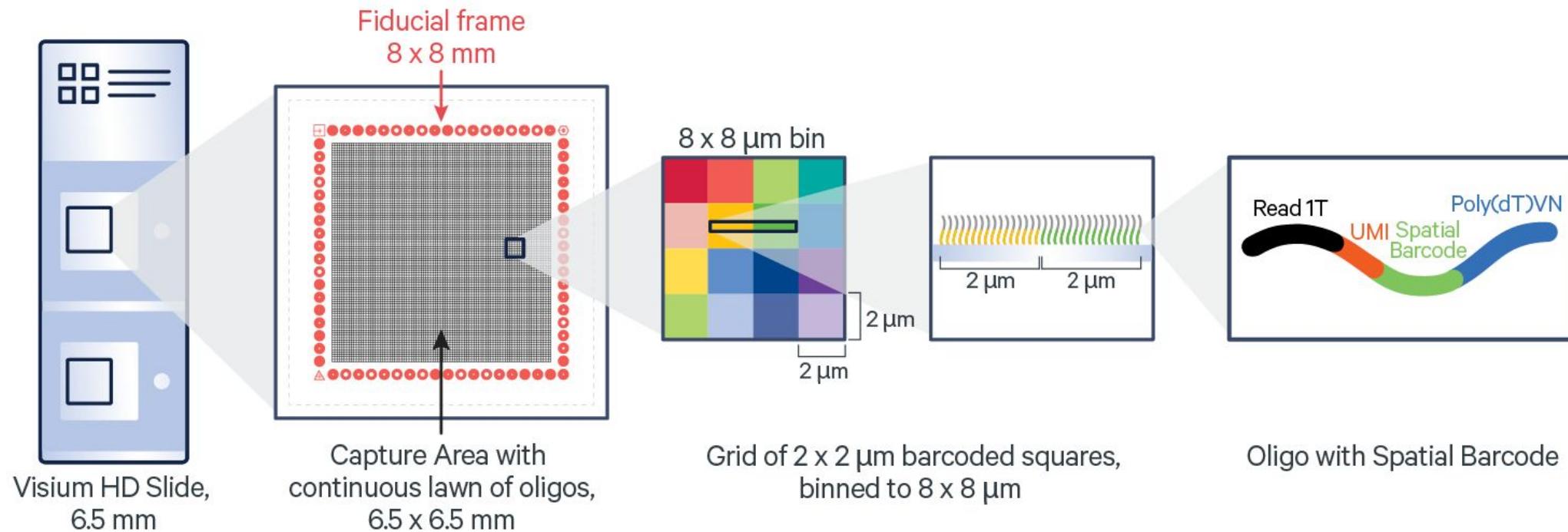
Read 2: probe sequences

- left-hand sequence (LHS)
- right-hand sequence (RHS)



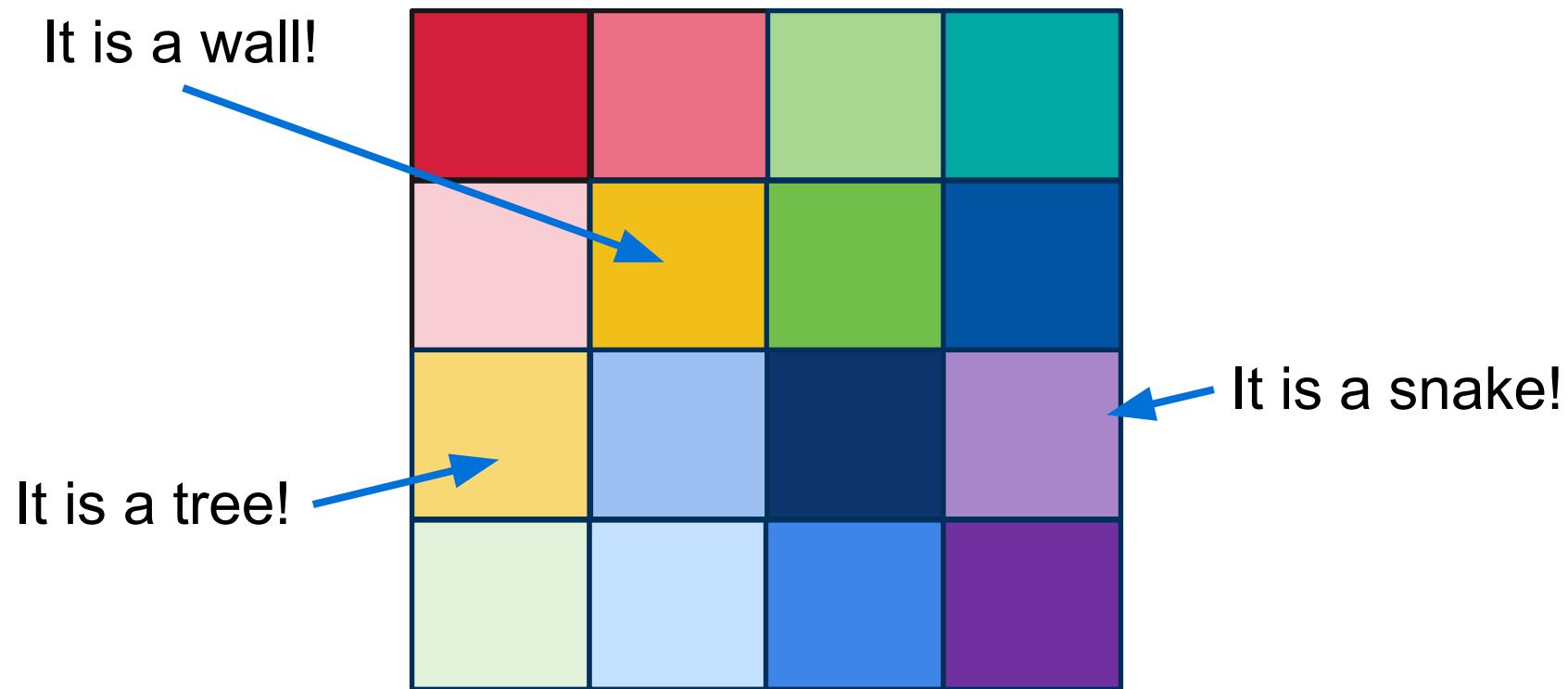
Visium workflow – understanding the slide architecture

Slide architecture enables single cell–scale resolution

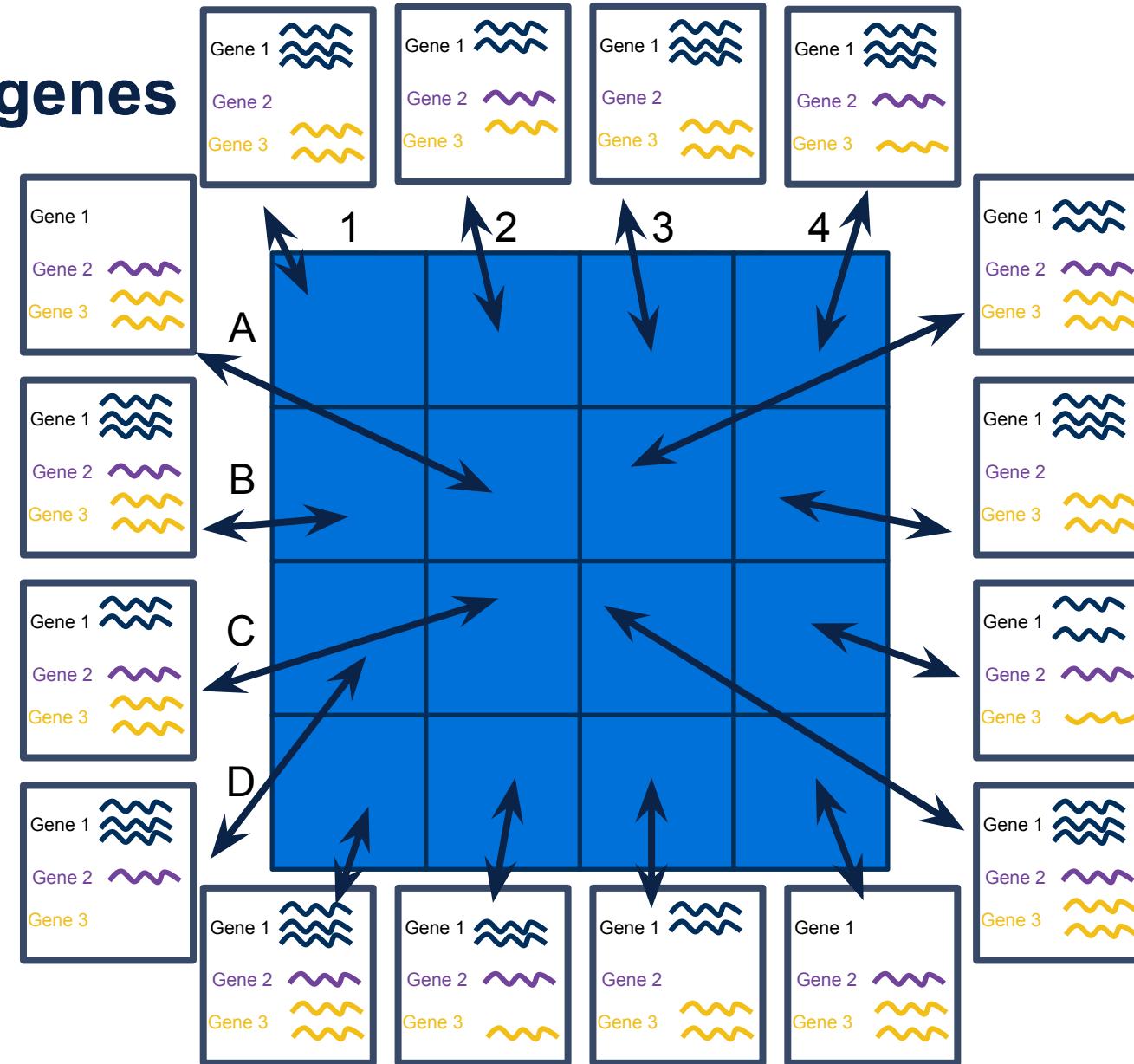


Why do we bin Visium HD gene expression data?

This is done because cells can span multiple squares, and binning improves analysis because UMI recovery per 2 µm square can be limiting.

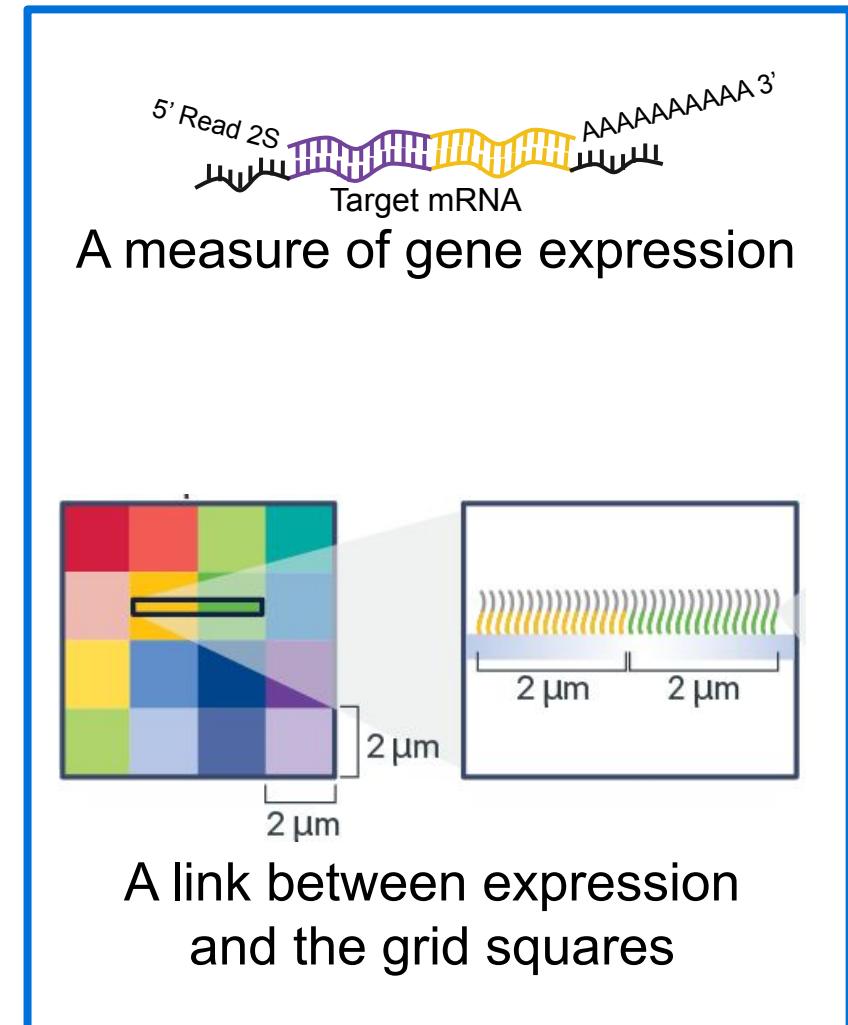
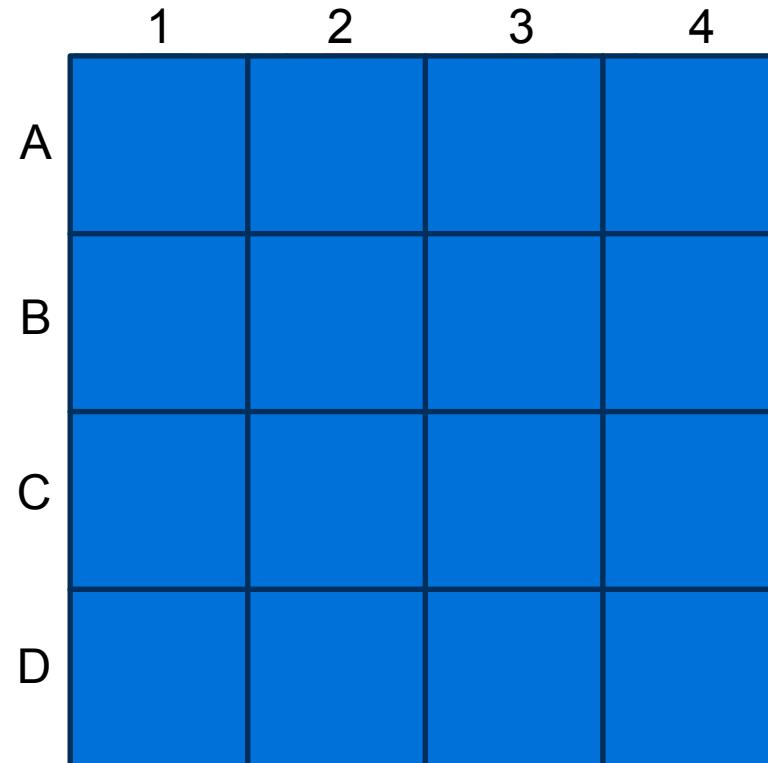
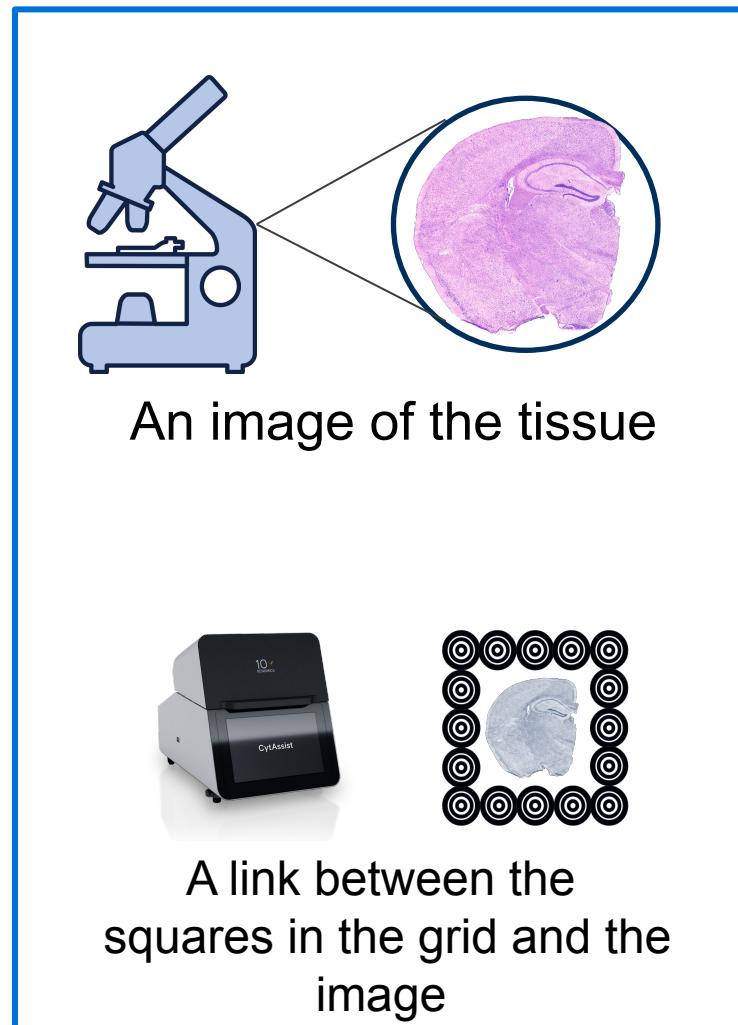


We want the expression of genes in that square.

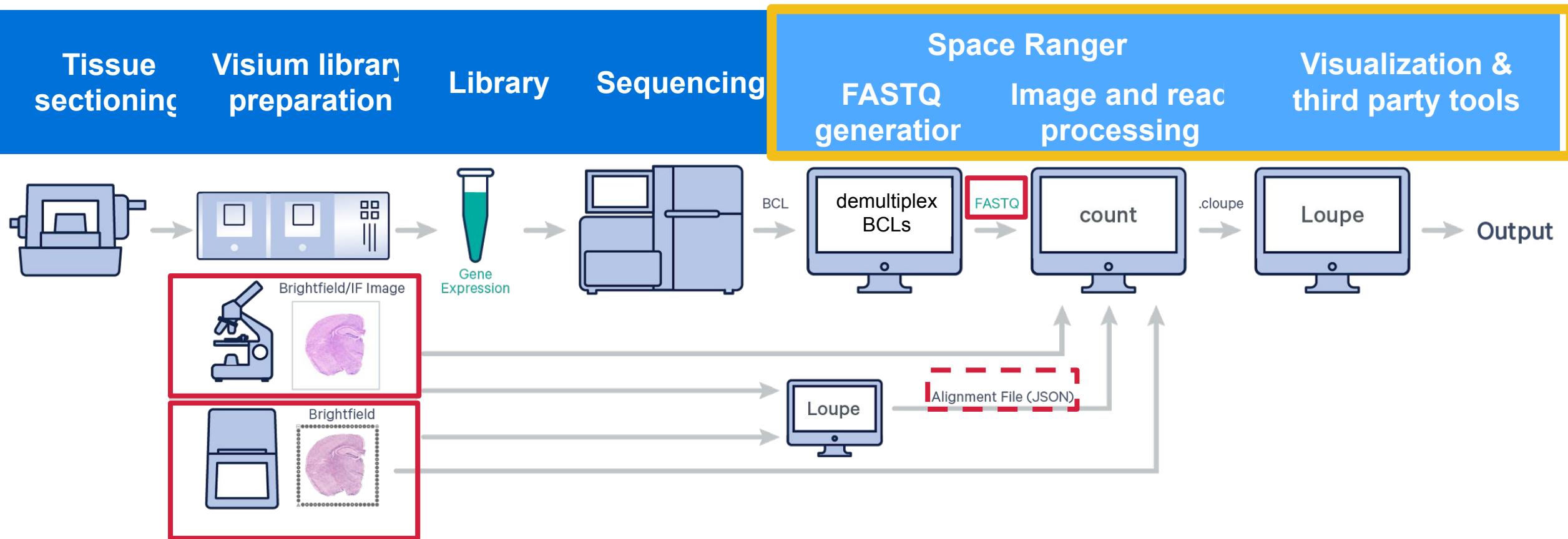


...and we want it
for each square
in the grid.

What information do we need to do this?



Input files for Space Ranger



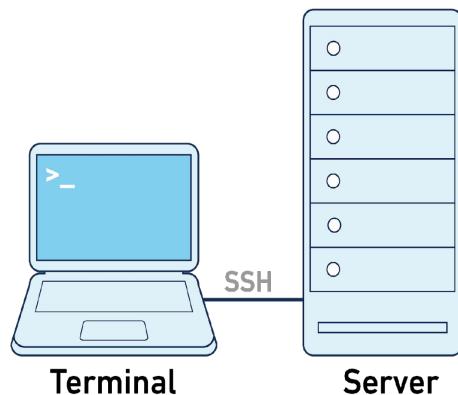
Running Space Ranger



On 10x Cloud: Friendly user interface and simple data management

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

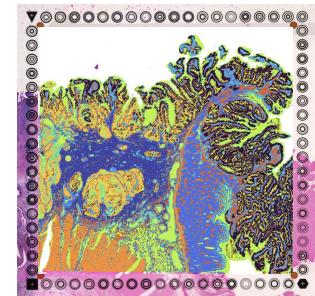
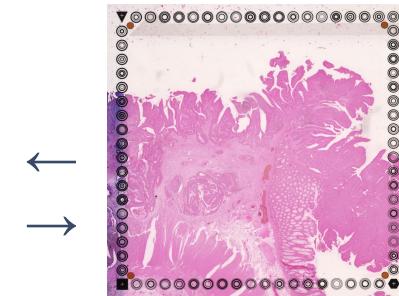
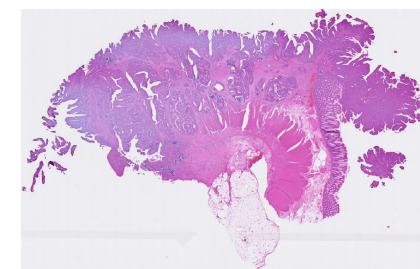
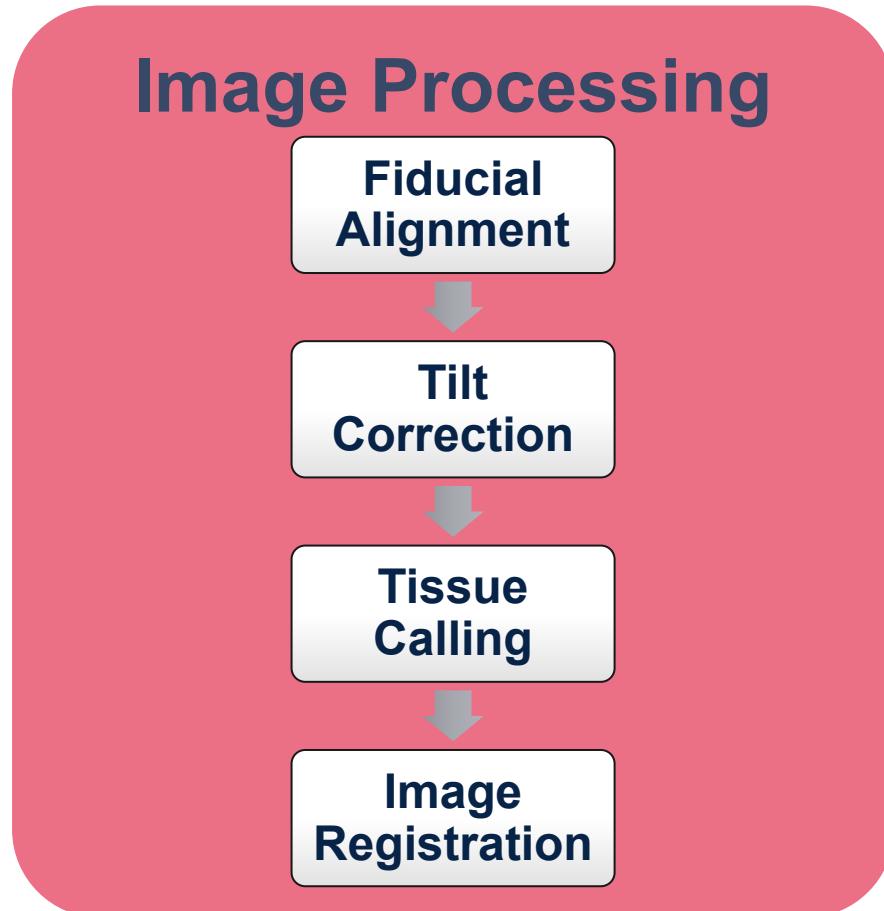
Space Ranger count

Image Processing

Read Processing

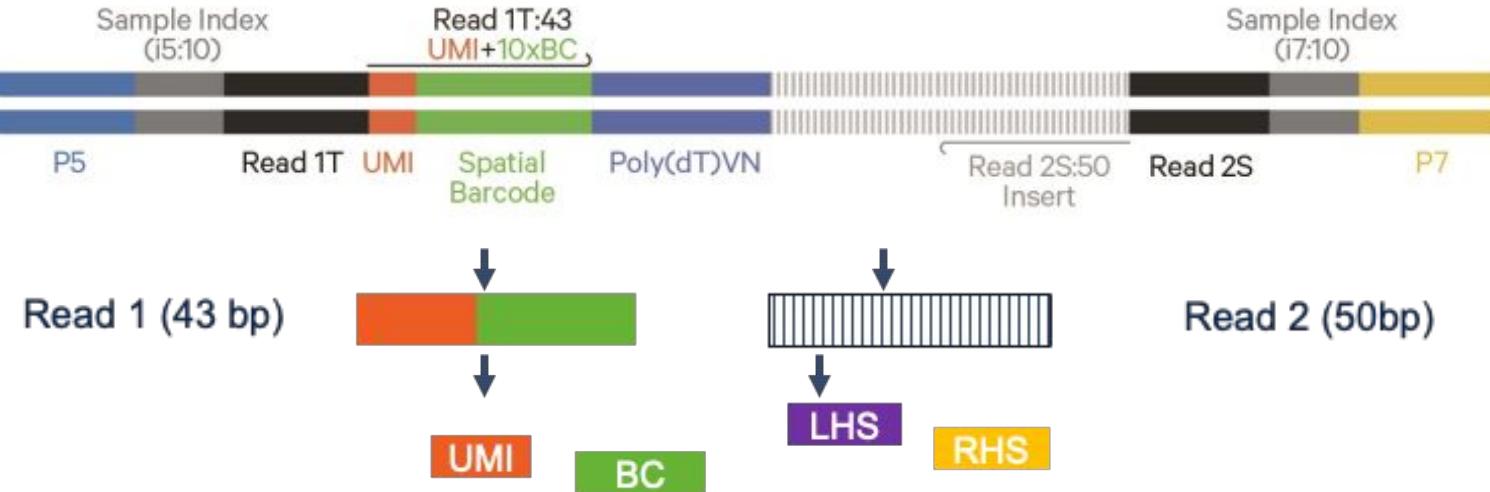
Space Ranger count

Purpose: Overlay the CytAssist image with the microscope image



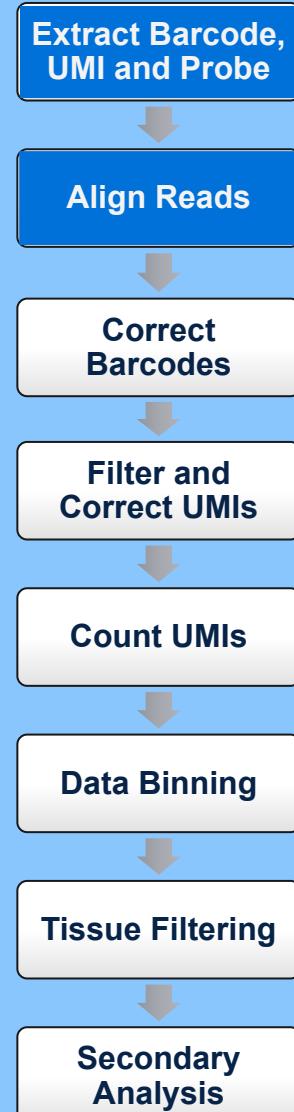
Space Ranger count

Purpose: Obtain the GEX matrix and spatial outs



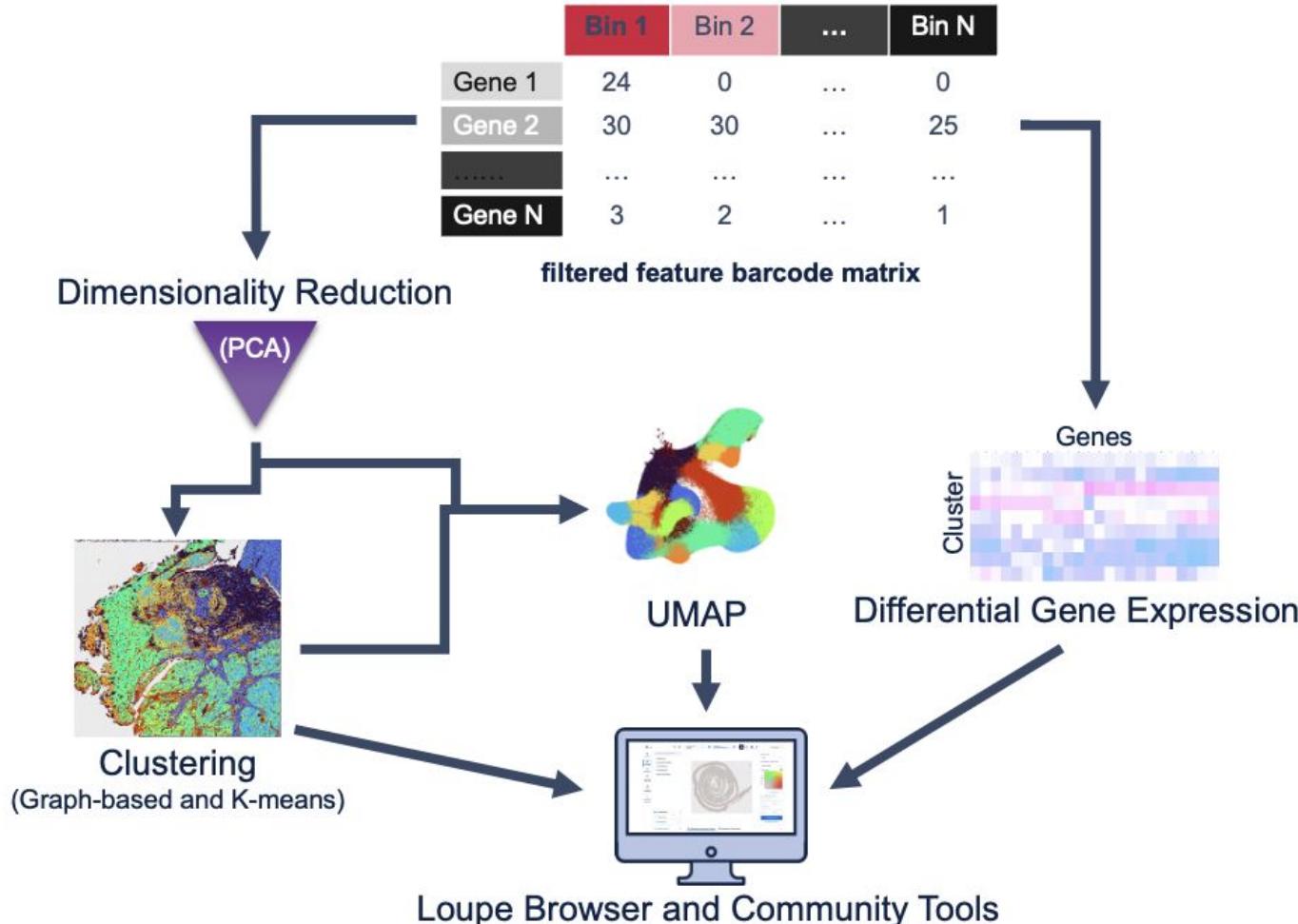
- UMI and BC are extracted from Read1 and split apart.
- The probe sequence is extracted from Read 2, and the LHS and RHS sequences are split apart.
 - Both halves need to map to the same probe in the probeset be considered confidently mapped and used for UMI counting and subsequent analysis steps.

Read Processing

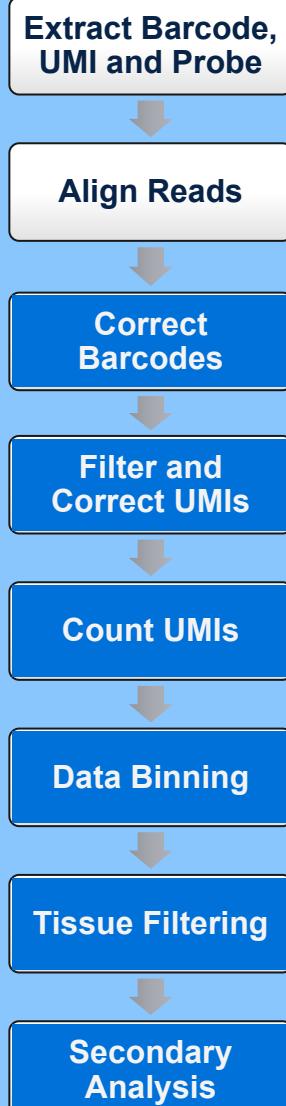


Space Ranger count

Purpose: Obtain the GEX matrix and spatial outs



Read Processing

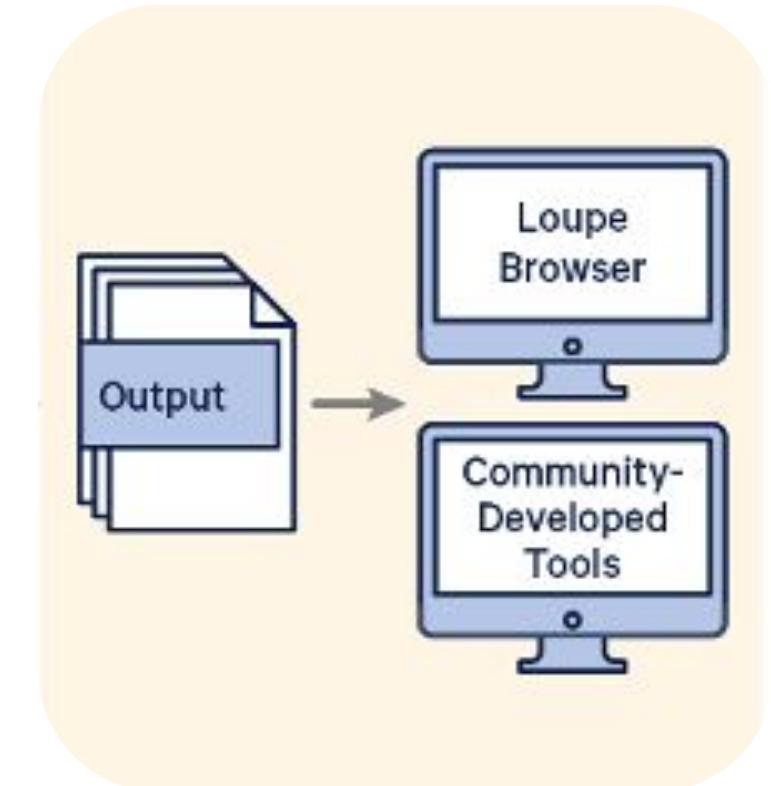


Space Ranger count output files

Structured for easy sharing and import into downstream analysis tools

- **Web summary** file for assessing data quality
- **Loupe Browser** cloupe file for downstream analysis
- Feature-barcode matrices
- Spatial outs folder for all data
- Secondary analyses

Depending on the bin levels (2 μ m, 8 μ m, 16 μ m, or custom square bins) some files may/not be generated





Quality Assessment: Interpreting the Space Ranger Web Summary

Analysis plan

Our journey through analysis today

Process Raw Data	Assess Sample Quality	Map Cell Types	Recluster Region of Interest	Explore Enriched Pathways
Analyze Visium HD raw data with Space Ranger on 10x Cloud Analysis	Evaluate sample quality with the summary HTML file	Manually map specific cell types using Loupe Browser	Focused analysis in tumor and surrounding regions	Explore the molecular pathways in tumor microenvironment using GSEA

Outline

- Where to Find the Web Summary
- Spatial Web Summary Overview
- Assessing Data Quality
- Troubleshooting

Where to Find the Web Summary

Interactive session

Where to Find the Web Summary

10x Cloud Analysis Platform

Output files

Each analysis comes with free downloads up to a limit. The amount of free downloads for each analysis is based on the total size of the outputs. For this analysis, you have 36.59 GB of free downloads remaining.

Free storage: 89 days remaining
[Manage data storage](#)

[Download in Browser](#) [Download with CLI](#)

	Output files	Kind	Size	File ID
<input checked="" type="checkbox"/>	Loupe Browser file <small>View description</small>	CLOUPE	7.14 GB	gI5Tm0PvS3W68z99...
<input checked="" type="checkbox"/>	Summary HTML <small>View description</small>	HTML	21.75 MB	YQwU902DpQyW58J...
<input type="checkbox"/>	Summary CSV <small>View description</small>	CSV	1.68 KB	ijZrlXxTQuVBuSOO6...
<input type="checkbox"/>	Per-molecule read information <small>View description</small>	HDF5	1.69 GB	mqfIwRfFTPCvYlWs...
<input type="checkbox"/>	Probe set <small>View description</small>	CSV	5.85 MB	67UNh1UxTfGX6HqE...
<input type="checkbox"/>	Binned outputs <small>View description</small>	TAR	14.75 GB	alqbLVZVTdeNVFD8...
<input type="checkbox"/>	Feature slice HDF5 <small>View description</small>	HDF5	751.57 MB	fkUZDPZ2RQmvdsYz...
<input type="checkbox"/>	Spatial outputs <small>View description</small>	TAR	59.38 MB	JlrErB71Q4uUgW3R...

Spatial Web Summary Overview and Assessing Data Quality

Interactive session

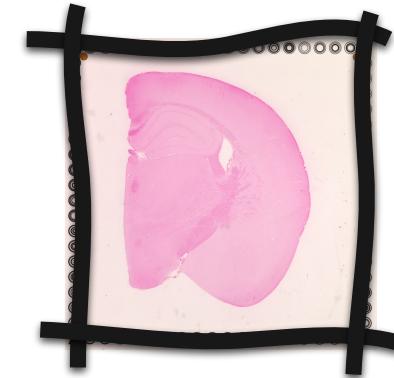
Troubleshooting

Interactive session

Common Causes of Image Processing Errors

Coverage Issues

- Entire fiducial frame covered (marker/tissue)
- Entire capture area covered by homogenous tissue



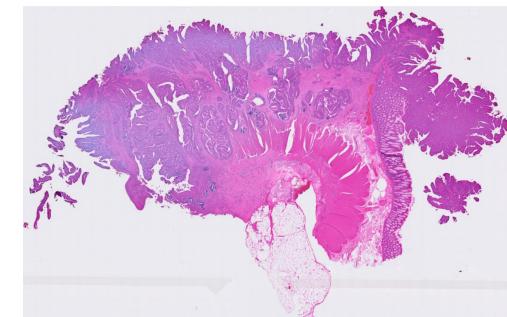
Poor Contrast with Background

- Poor CytAssist destaining
- Poor high resolution image staining
- Marker in the capture area



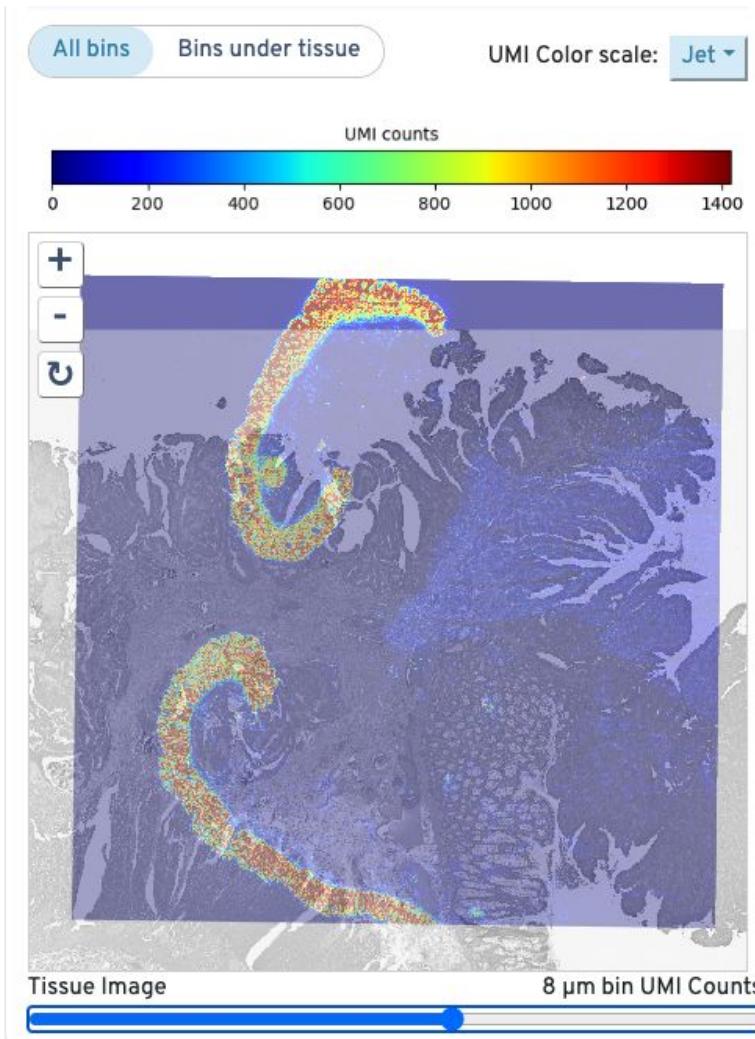
Image Disparities

- Sequential section used
- Significantly different tissue area in CytAssist or high resolution image

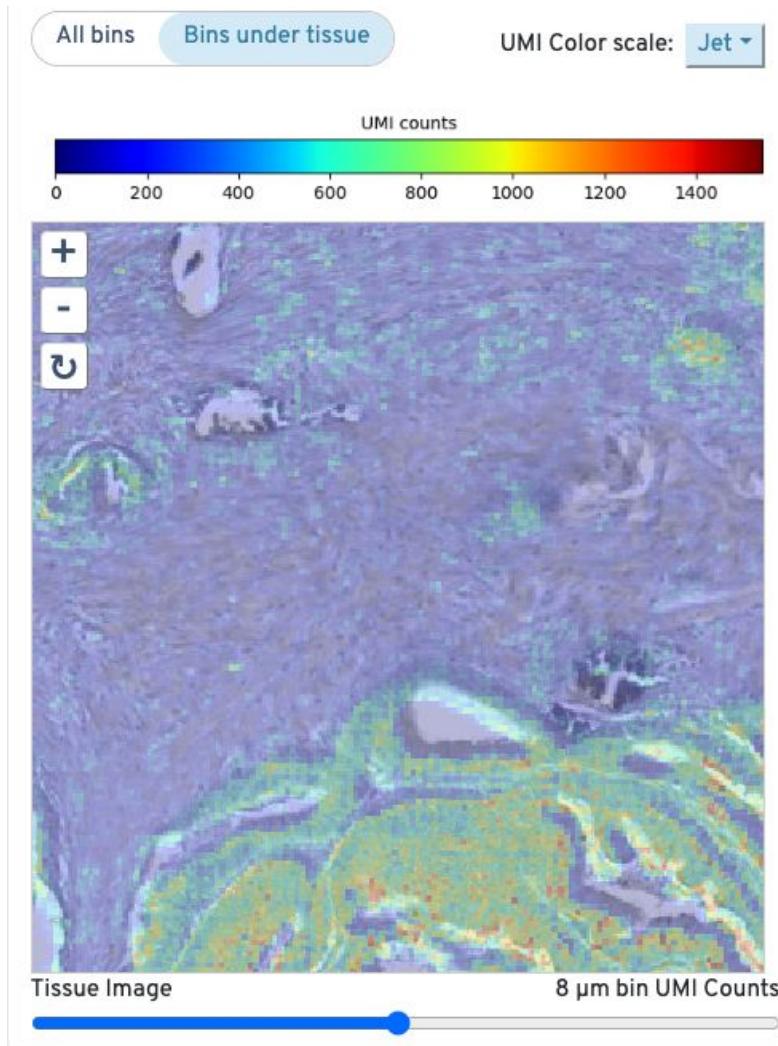


Additional Troubleshooting Examples

Sample Swap



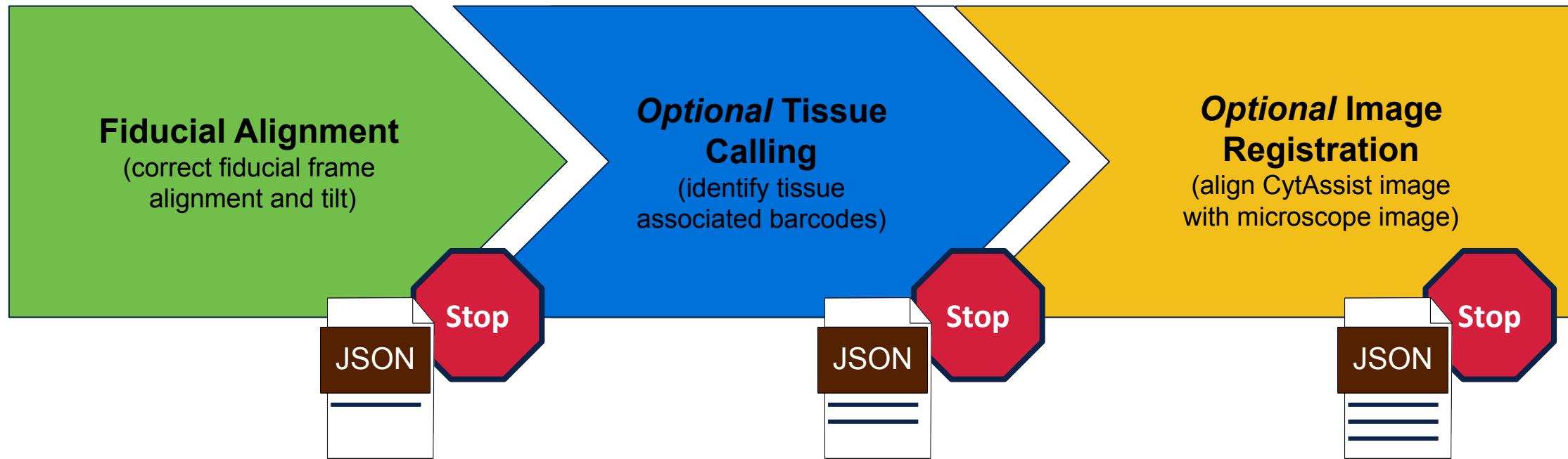
Wrong Serial Number



Manual Image Alignment

Correcting Image Processing Errors

Manual Image Processing Workflow in Loupe Browser



Interactive session

What Do You Do With the JSON File?

10x Cloud Analysis Platform

Advanced settings

Unknown slide (optional) This slide's serial number and capture area are unknown
Select this option only if the slide serial number or capture area were recorded incorrectly on the CytAssist instrument and the correct values are unknown. Data may be poor quality; only select this if you are absolutely certain the serial number and capture area are lost.

Override slide ID (optional) Override the slide serial number and/or capture area
If the slide serial number or capture area were recorded incorrectly on the CytAssist instrument and the correct values are known, enable this option and provide the values below.

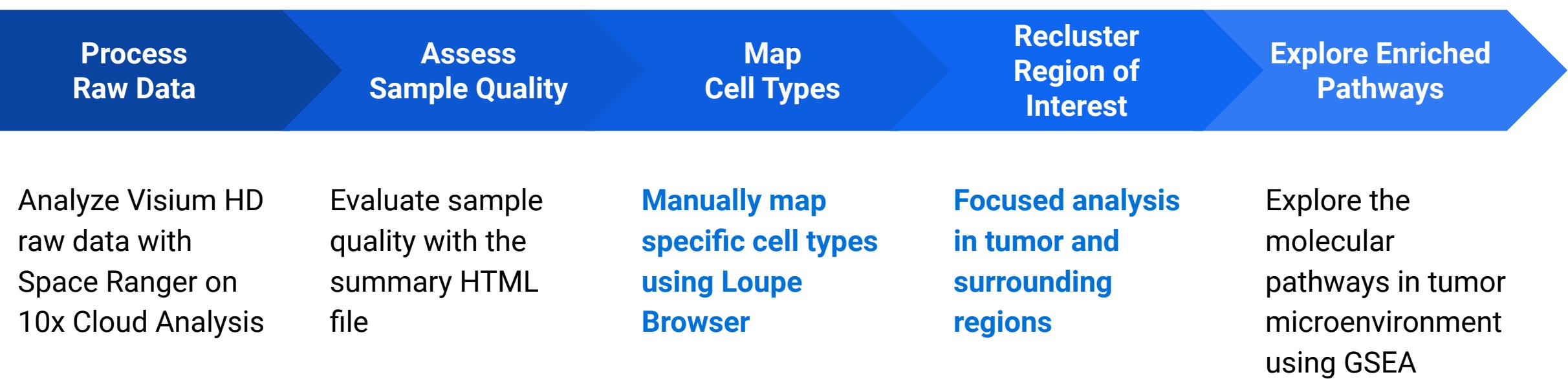
Slide serial number (optional)
Override the slide serial number recorded on the CytAssist instrument.

Slide capture area (optional)
Override the capture area recorded on the CytAssist instrument.

Alignment JSON (optional)
To bypass automated microscope image to CytAssist image alignment, provide an alignment JSON file from Loupe Browser. To learn more about manual alignment, visit our [support site](#).

Analysis plan

Our journey through analysis today





Loupe Browser: Quality Control and Rapid Data Exploration

Outline

- QC considerations that can be addressed in Loupe
- Approach to mapping cell types in our data
- Hands-on

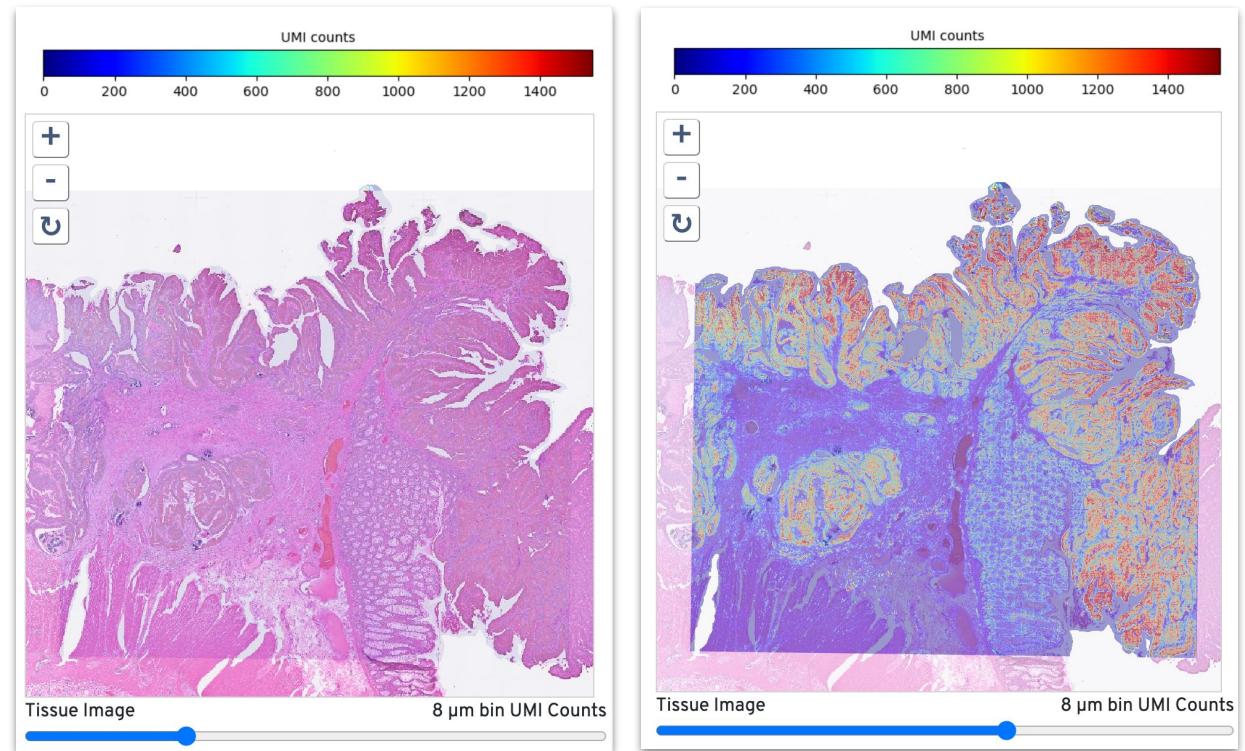
Quality control in Loupe Browser

What can we do in Loupe?

Main consideration: alignment and tissue segmentation are satisfactory

In previous section, we demo'ed how to perform manual alignment

Next, demo how to address poor tissue segmentation in Loupe without running spaceranger from the beginning

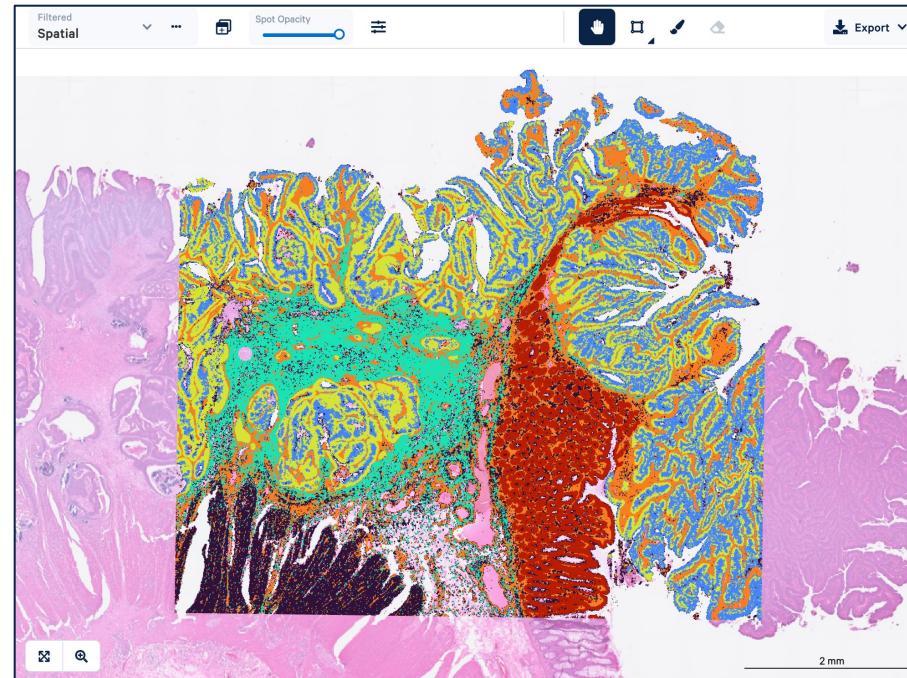


Map cell types in Loupe Browser

We will demo how to approach assigning bins to cell type categories

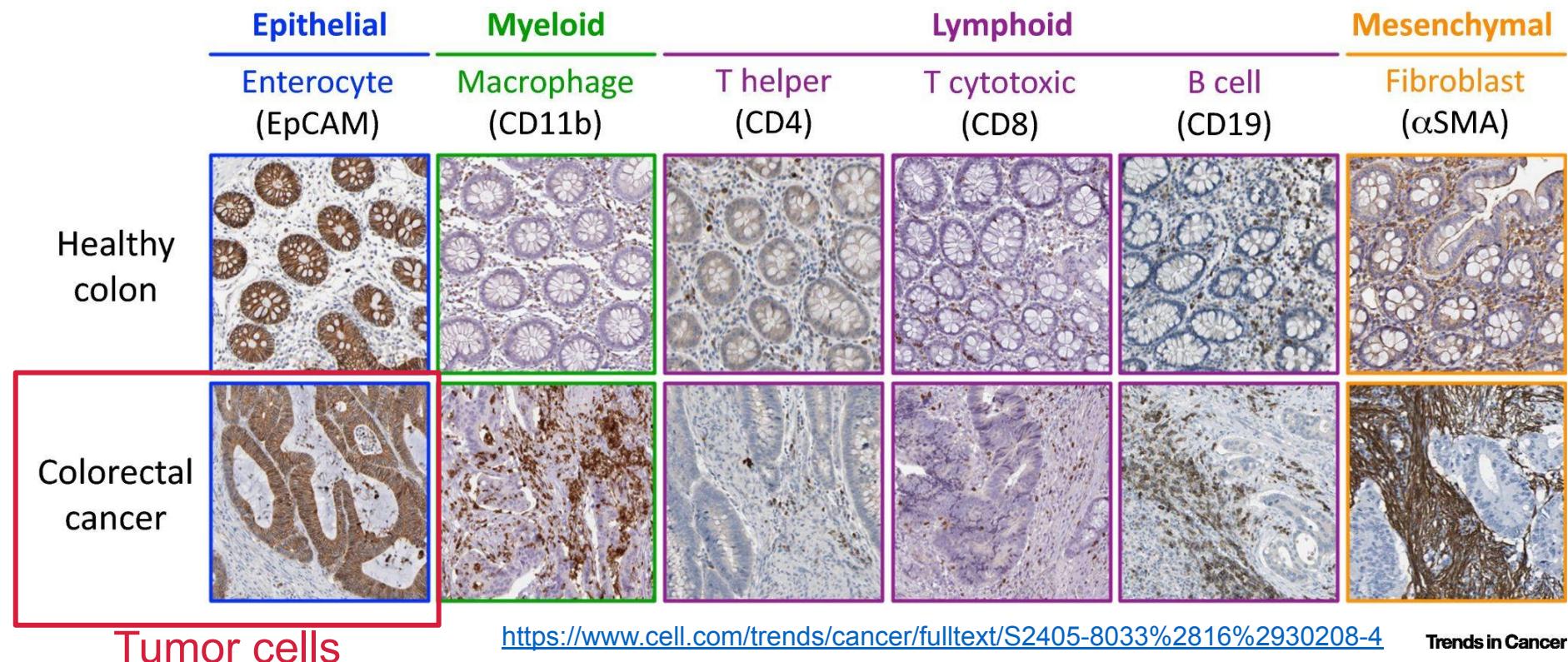
Marker genes

CEACAM6	Tumor
NKD1	Tumor
KRT23	Tumor
FCGBP	Goblet Cells
MUC2	Goblet Cells
CLCA1	Goblet Cells
COL1A1	Fibroblasts
MMP2	Fibroblasts
C1QC	Macrophages
LYZ	Macrophages
CD68	Macrophages



Spatial information

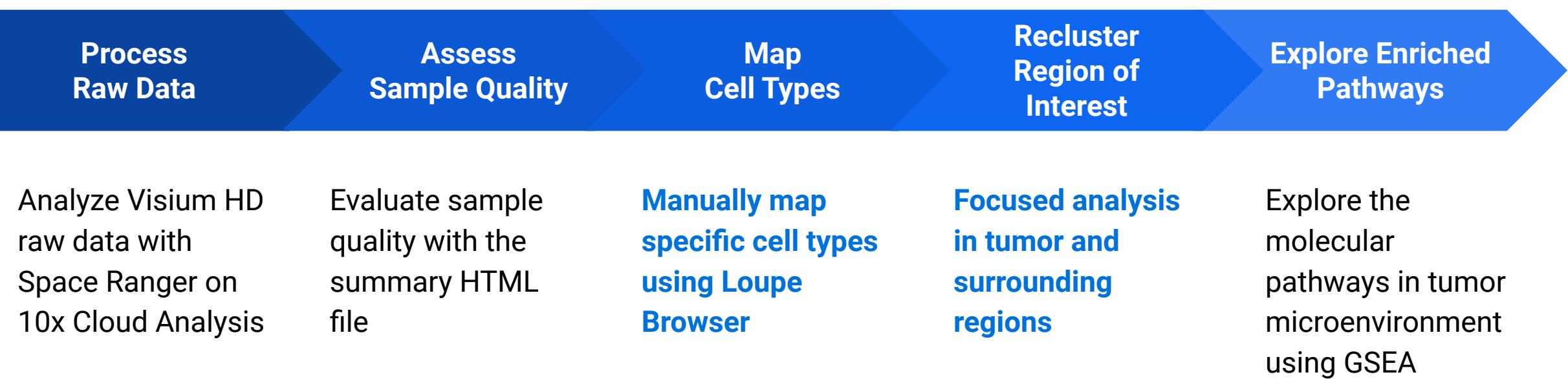
Colorectal cancer is a heterocellular system



Better understanding of the tumor microenvironment has the potential to inform better clinical intervention strategies.

Analysis plan

Our journey through analysis today



Interactive session

Going through the list one-by-one is time-consuming and tedious

How can we understand what is happening within and outside the tumor region using gene expression data?

Analysis plan

Our journey through analysis today

Process Raw Data	Assess Sample Quality	Map Cell Types	Recluster Region of Interest	Explore Enriched Pathways
Analyze Visium HD raw data with Space Ranger on 10x Cloud Analysis	Evaluate sample quality with the summary HTML file	Manually map specific cell types using Loupe Browser	Focused analysis in tumor and surrounding regions	Explore the molecular pathways in tumor microenvironment using GSEA



Introducing Community Developed Tools for Visium Data Analysis

Outline

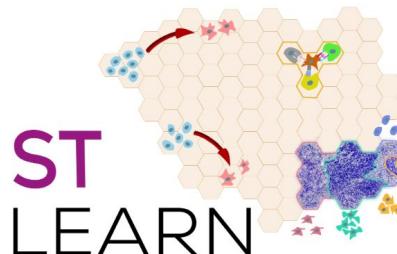
- Approaching the community developed tools
- Common downstream analyses of spatial transcriptomics data
- Recap on analysis plan
- Introduction to GSEA analysis
- Quick introduction to Jupyter Notebook and Google Colab
- Hands-on

Approaching the Spatial Data Analysis Ecosystem

- **Top challenges of spatial data analysis**
 - No one standard analysis flow applies to all projects
 - Large amount of data requires computing and storage resources
- **How do you choose tools?**
 - 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
 - Workshop



Community developed tools



General Analysis Tools



Specialized Analysis Tools

General

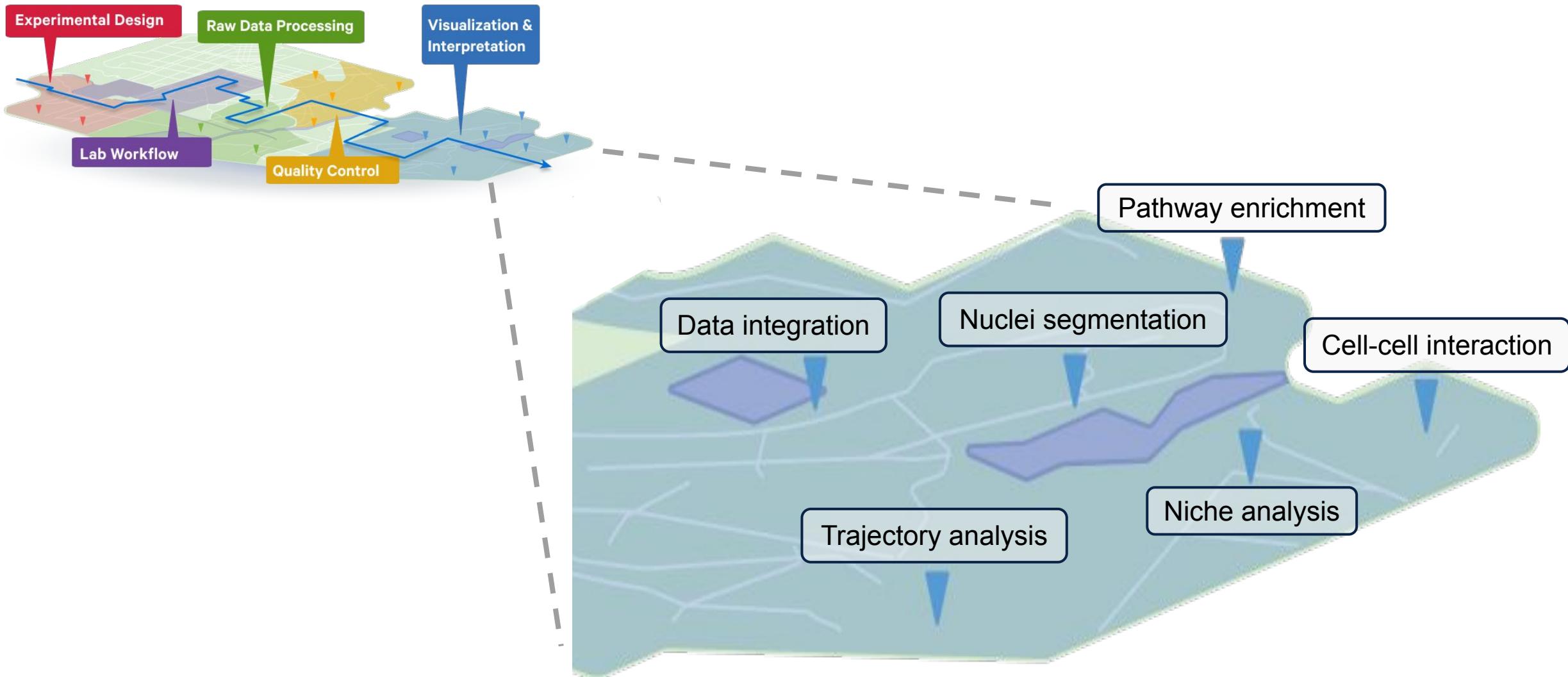
Introduction

Oct 25, 2022

Continuing Your Journey after Running Space Ranger

In this article we provide an overview of common next steps after running Space Ranger. We cover 10x and community developed tools.

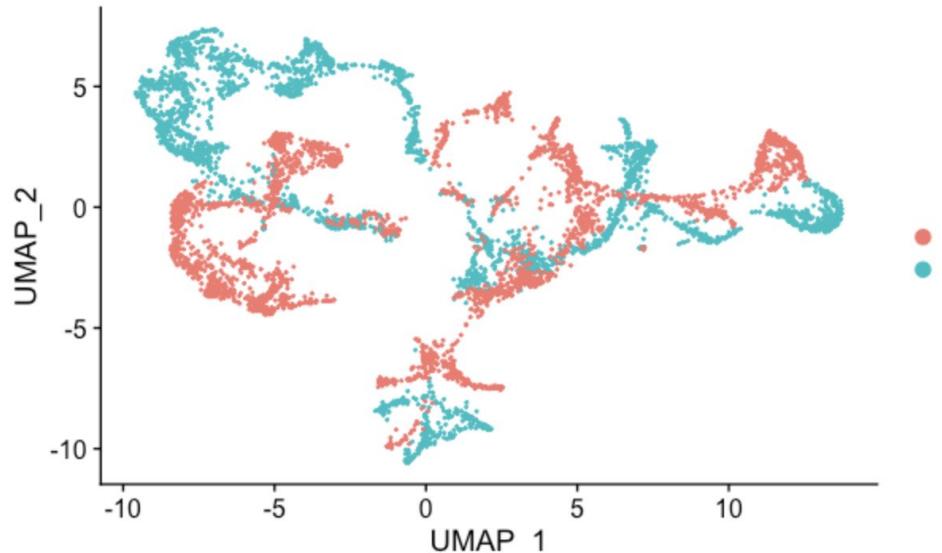
Common downstream analyses of spatial transcriptomics data



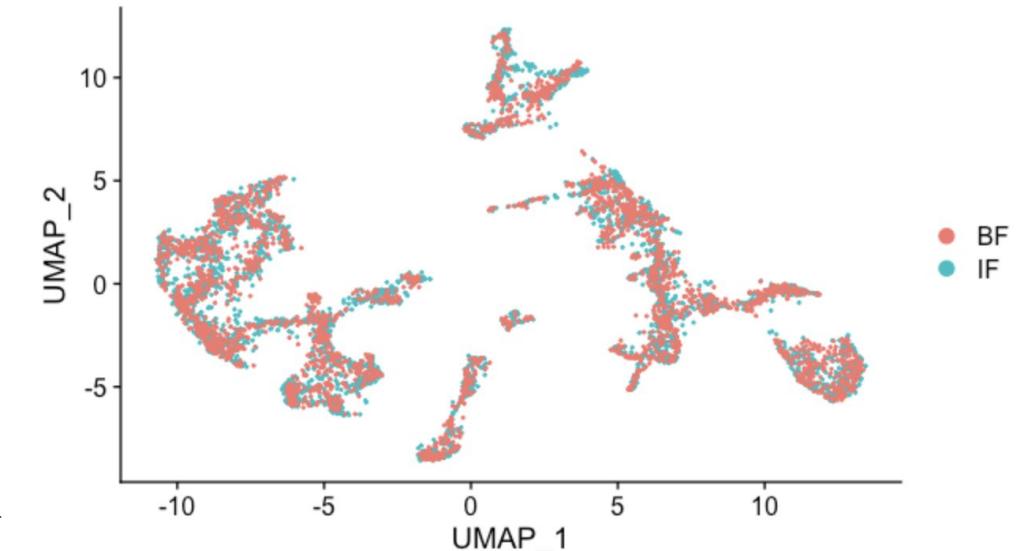
Data integration

Integrating multiple spatial transcriptomics datasets

- For comparative analysis across conditions, individuals, or serial sections
- Key aspects:
 - Normalization
 - Batch correction
 - Alignment of spatial coordinates
- Example tools: [stLearn](#), [STalign](#), [PRECAST](#), etc.
- Analysis guide: [Correcting Batch Effects in Visium Data](#)



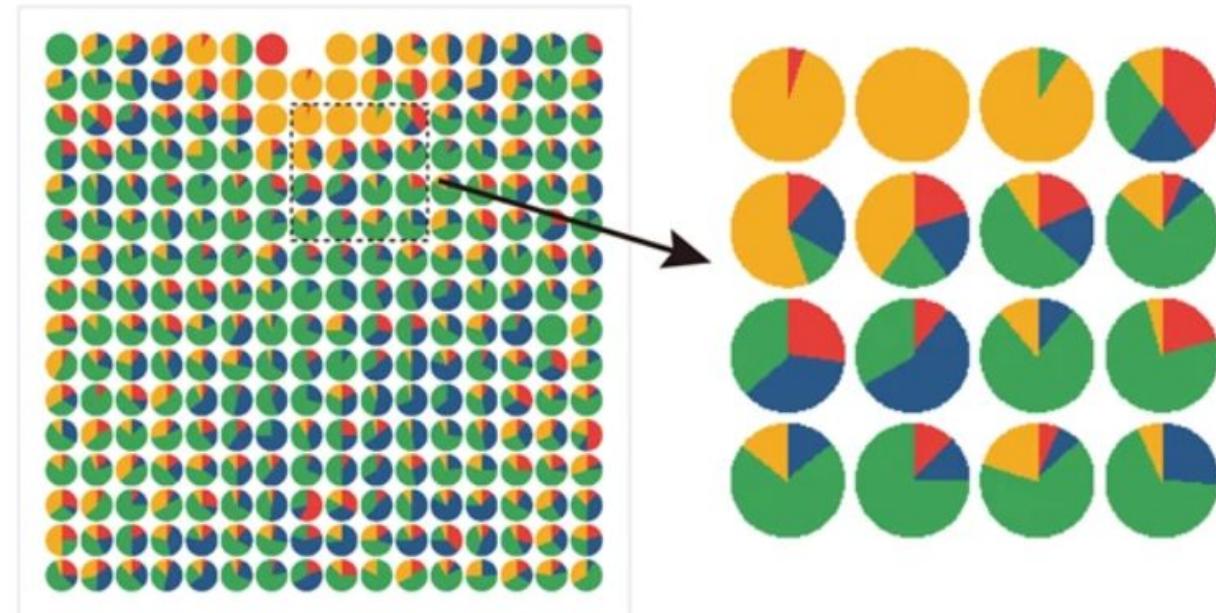
Batch
correction



Data integration

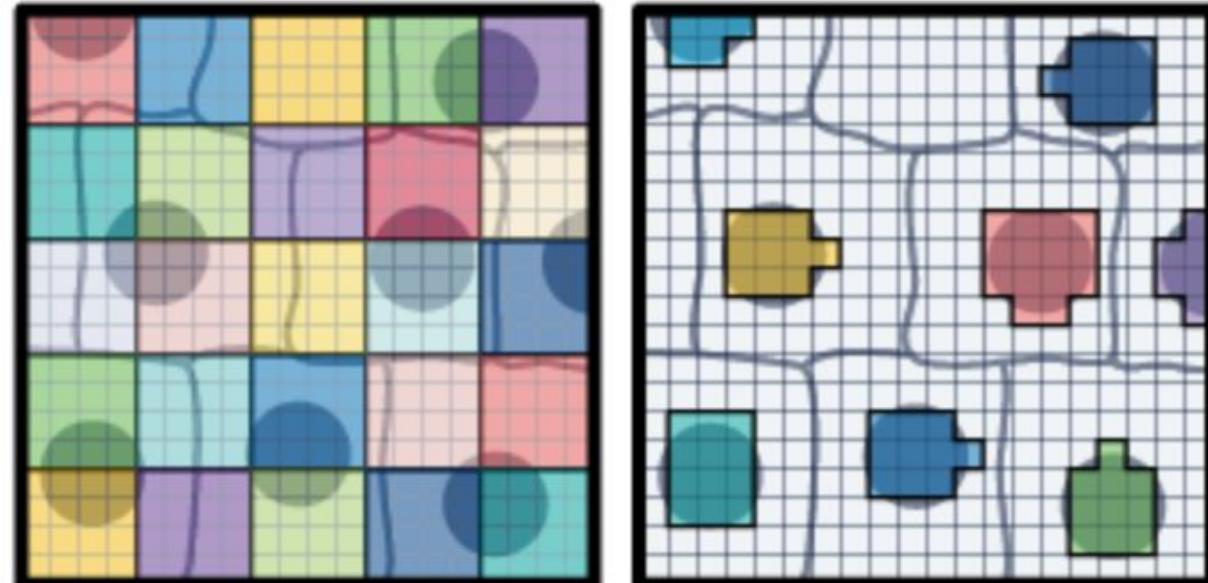
Spatial transcriptomics & single cell datasets

- Well-matched sample type is usually recommended for integration
- Spot deconvolution
 - Example tools: [Cell2location](#), [SpatialDWLS](#), [SpaceXR](#), ect.
 - Analysis guide: [Integrating 10x Visium and Chromium data with R](#)
 - Note: this can also be done without single cell data as a reference (e.g. Space Ranger)
- Annotation transfer / mapping
 - [Seurat \(v5\)](#), [scVI](#), [scArches](#)



Cell/nuclei segmentation

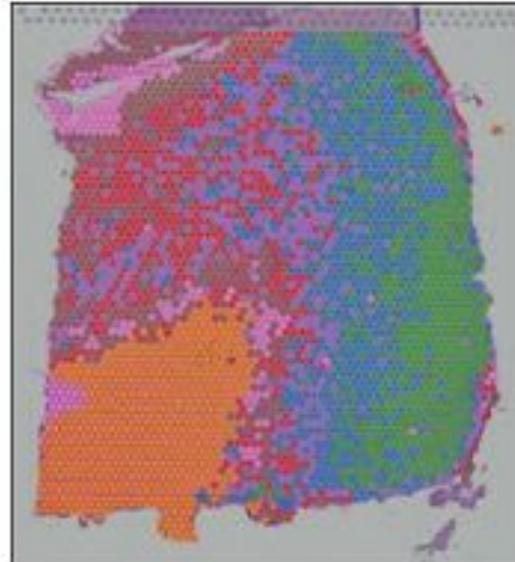
- Use information from H&E image
- Define cell/nuclei boundaries and assign transcripts to specific cells/nuclei to derive a more accurate gene expression profile
- Example tools: [StarDist](#), [Cellpose](#), [Baysor](#), etc.
- Analysis guide for visium HD data: [Nuclei Segmentation and Custom Binning of Visium HD Gene Expression Data](#)



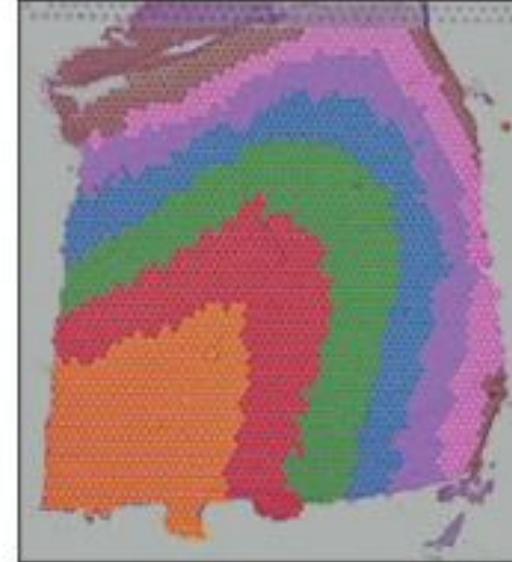
Spatially-aware clustering and niche analysis

- Identify microenvironments characterized by their cellular composition and gene expression of neighboring cells
- Provide a more accurate representation of the biological structure and function within tissues
- Example tools: [Niche-DE](#), [Seurat v5](#), [BANKSY](#), [COVET](#), etc.
- Literature: [Benchmarking spatial clustering methods with spatially resolved transcriptomics data](#)

Gene expression



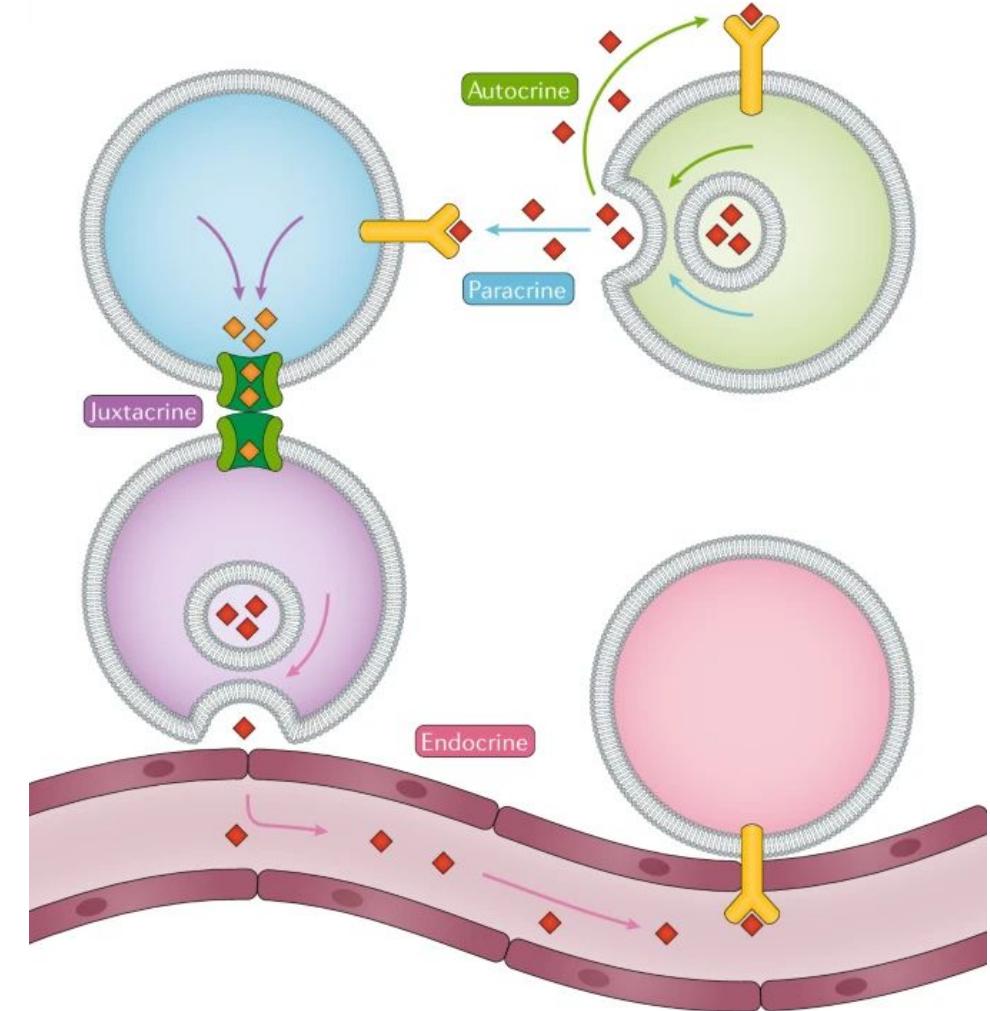
Gene expression +
Spatial information



Cell-cell interaction

Can also be performed with single-cell RNA-seq data

- Estimate interaction between cells based on expression of ligand and receptor pairs
- Spatial information helps evaluate the possibility of inferred interactions
- Example tools: [LIANA](#), [stLearn](#), [CellphoneDB](#), etc.

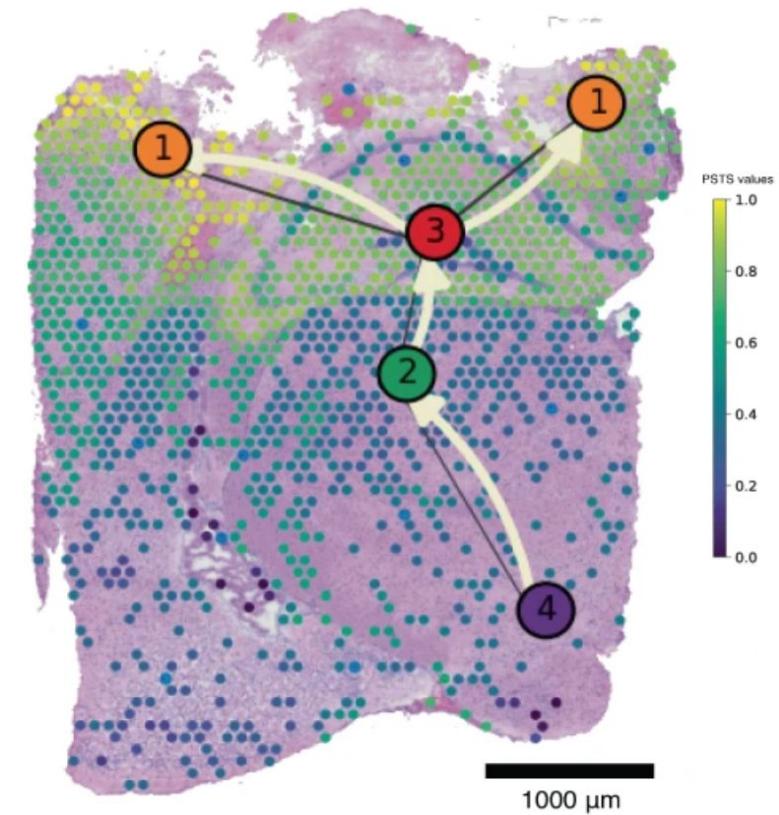


<https://www.nature.com/articles/s41576-020-00292-x>

Trajectory analysis

Can also be performed with single-cell RNA-seq data

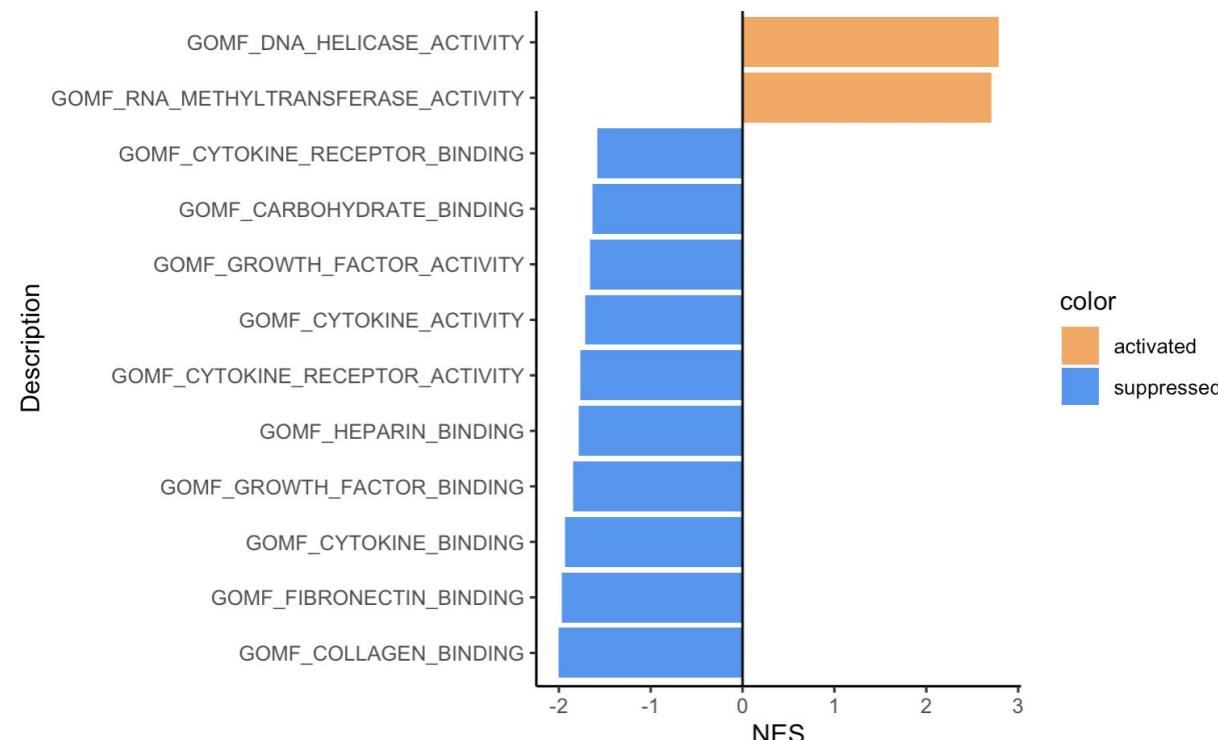
- Infer the evolution and differentiation process between cells by analyzing their spatial positions and gene expression levels within tissue sections
- Example tool: [stLearn](#), [SPATA2](#), etc.



Pathway enrichment analysis

Can also be performed with single-cell RNA-seq data

- Summarize many molecular insights into interpretable “Pathways” for specific clusters or microenvironments
- Require selection of appropriate annotation databases as the input for analysis
- Example tools: [enrichR](#), [clusterProfiler](#), [GOATOOLS](#), etc.



<https://www.gsea-msigdb.org/gsea/index.jsp>

Analysis plan

Our journey through analysis today

Process Raw Data	Assess Sample Quality	Map Cell Types	Recluster Region of Interest	Explore Enriched Pathways
Analyze Visium HD raw data with Space Ranger on 10x Cloud Analysis	Evaluate sample quality with the summary HTML file	Manually map specific cell types using Loupe Browser	Focused analysis in tumor and surrounding regions	Explore the molecular pathways in tumor microenvironment using GSEA

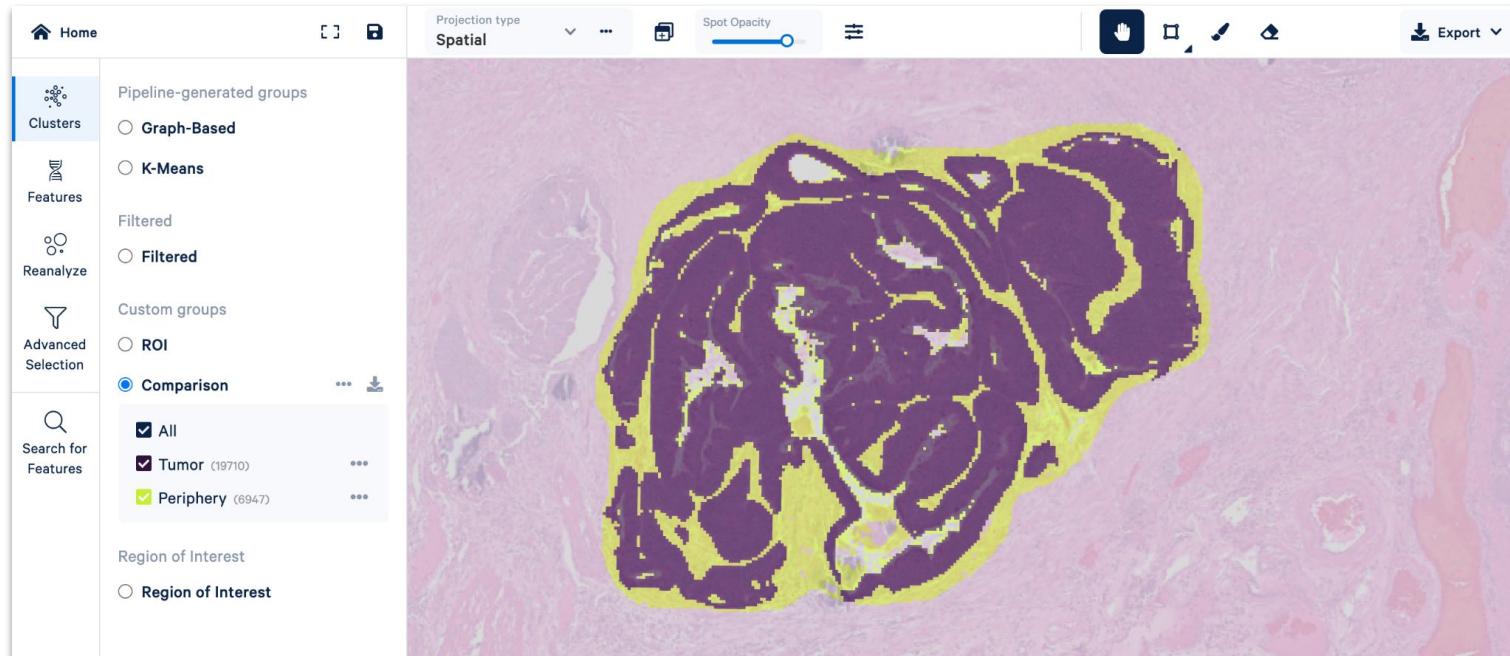
Further Analysis Using Community Tools: GSEA as an Example

GSEA introduction

Enrichment analysis of the tumor microenvironment

What is happening at the molecular level in our region of interest?

We have captured the tumor microenvironment → tumor cells + adjacent non-cancerous cells. We can learn more about what's happening on the molecular level by using Gene Set Enrichment Analysis (GSEA) on the differentially expressed genes between these two regions.



How to perform GSEA

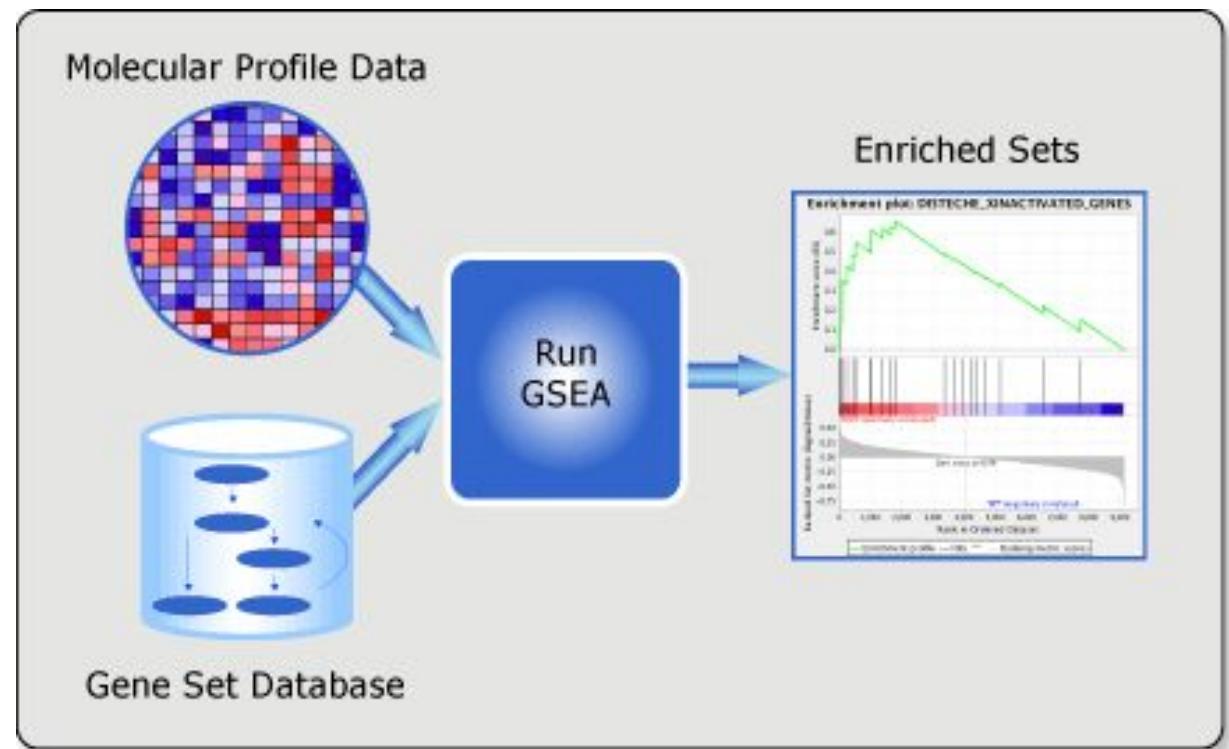
The required information

For GSEA you need two things:

- A **ranked list**
- A **set**

In single cell RNAseq or spatial gene expression analysis:

- The **ranked list** is a list of genes
- The **set** is a pathway or other biologically relevant group of genes



<https://www.gsea-msigdb.org/gsea/index.jsp>

An example: 10xer's favorite desserts

Ranked list of favorite desserts amongst 10x employees

Rank	Desert
1	Ice cream
2	Popsicle
3	Ice cream sandwich
4	Cupcakes
5	Cheesecake
6	Gelato
7	Cookies
8	Creme Brulee
9	S'mores
10	Chocolate bar
11	Tiramisu
12	Pudding
13	Apple pie
14	Macaron
15	Cherry pie
16	Brownie
17	Key lime pie
18	Pecan pie
19	Pound cake
20	Pumpkin pie

What can we learn from this favorite desserts list?

Question: What kind of desserts does 10x like the most in summer?

To answer this we will need:

- A **ranked list**: 10xers favorite desserts
- A **set**: a collection of desserts with common characteristics

Enrichment score (ES)

Quantifies the degree of overrepresentation of a set at the top or bottom of a ranked list

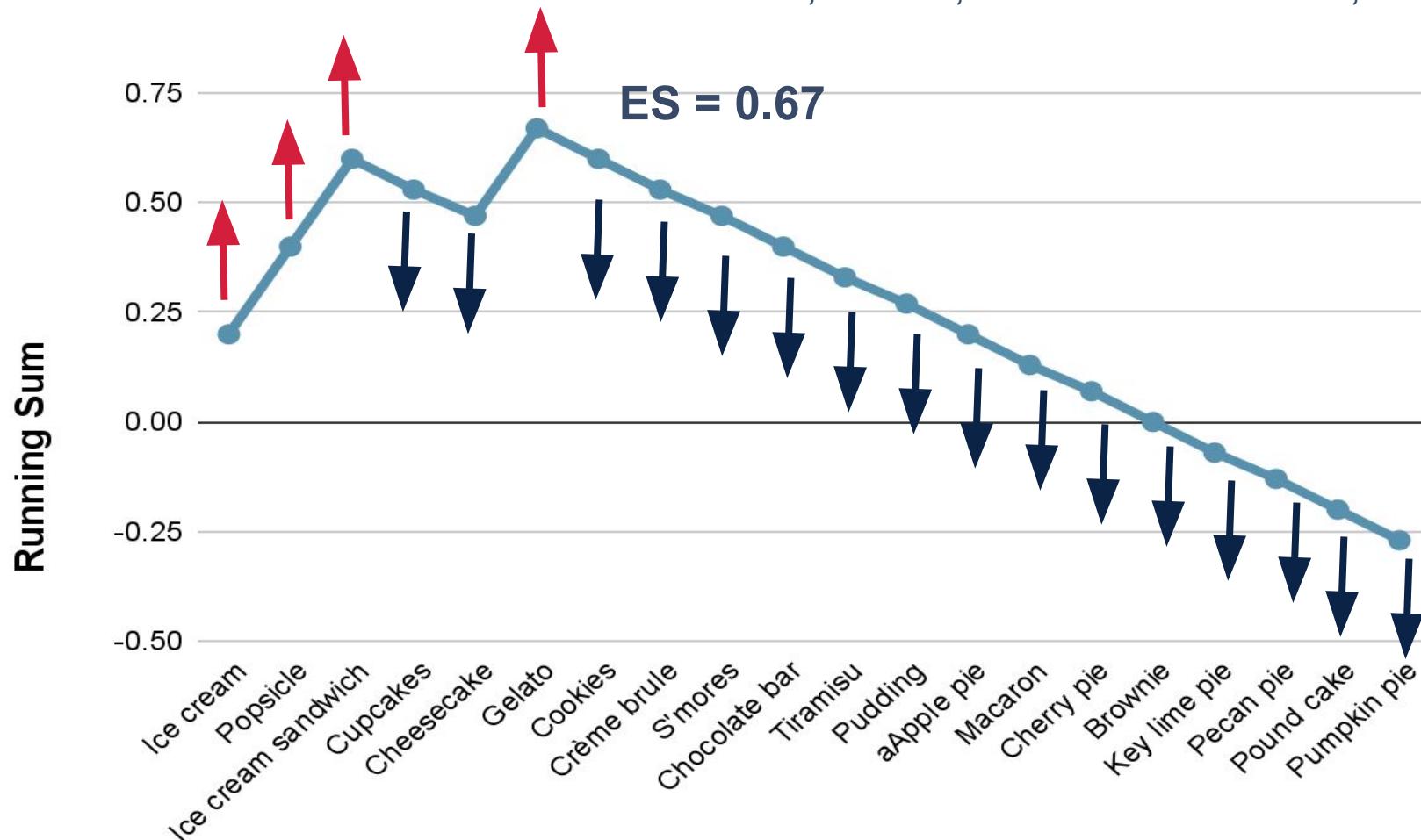
We have two sets

Frozen Desserts: Ice cream, Gelato, Ice cream sandwich, Popsicle, Frozen yogurt

Pies: Apple pie, Pumpkin pie, Pecan pie, Key lime pie, Cherry pie

Enrichment score (ES)

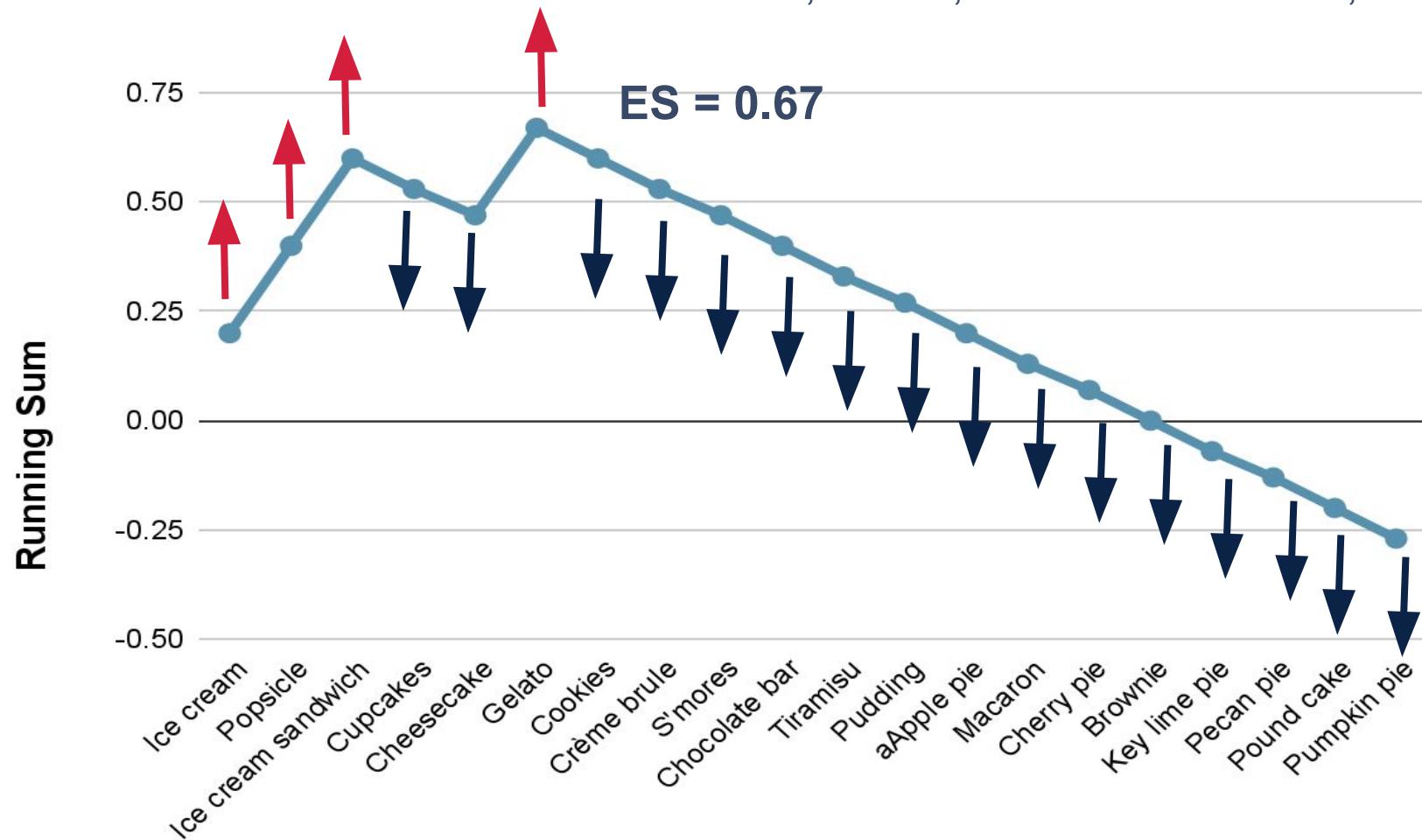
Enrichment score for **Frozen Desserts**: Ice cream, Gelato, Ice cream sandwich, Popsicle, Frozen yogurt



In the set [Y/N] Y Y Y N N Y N N N N N N N N N N N N N N N

Enrichment score (ES)

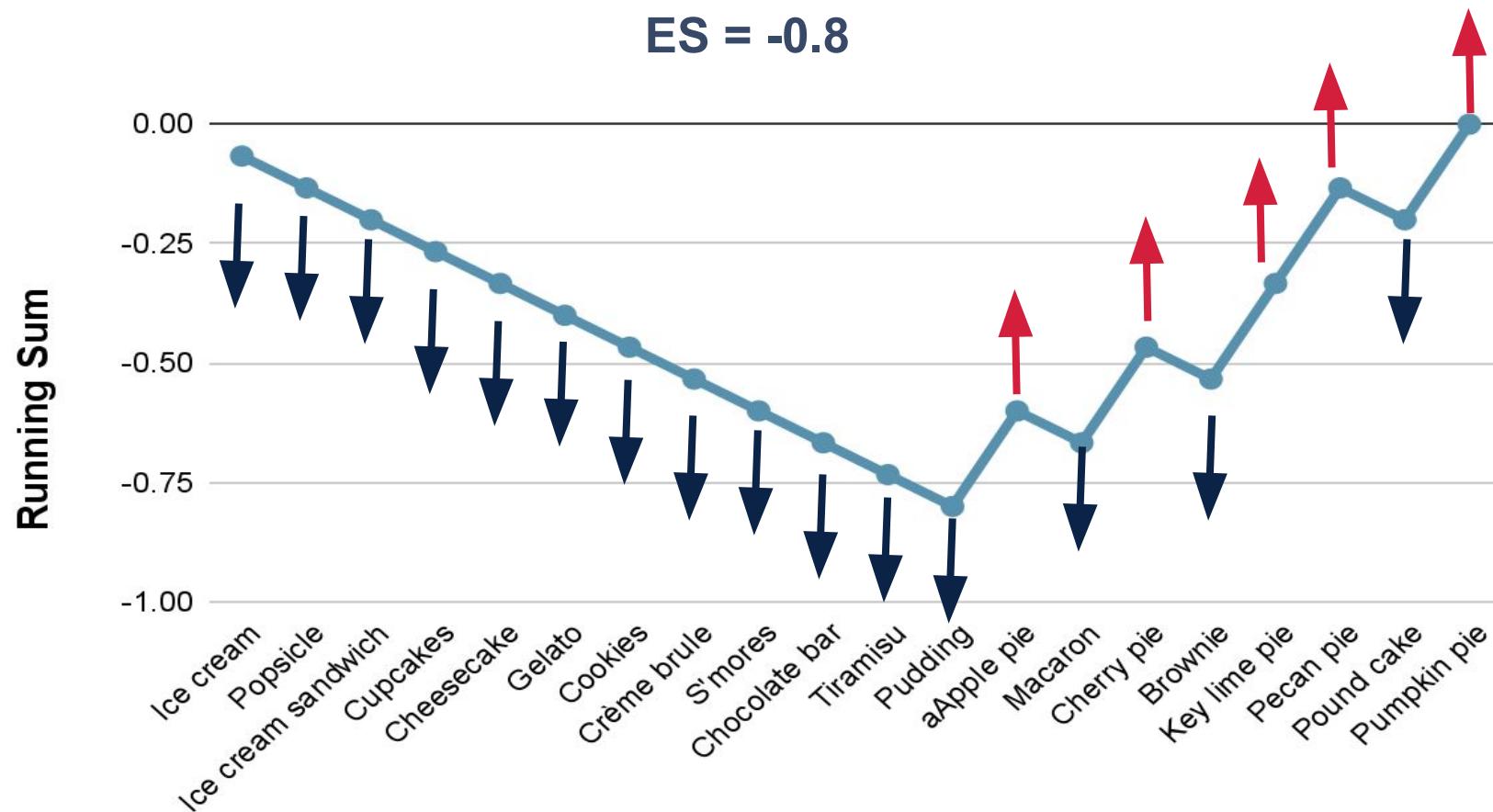
Enrichment score for **Frozen Desserts**: Ice cream, Gelato, Ice cream sandwich, Popsicle, Frozen yogurt



Frozen desserts are 10xers top favorite type of desserts

Enrichment score (ES)

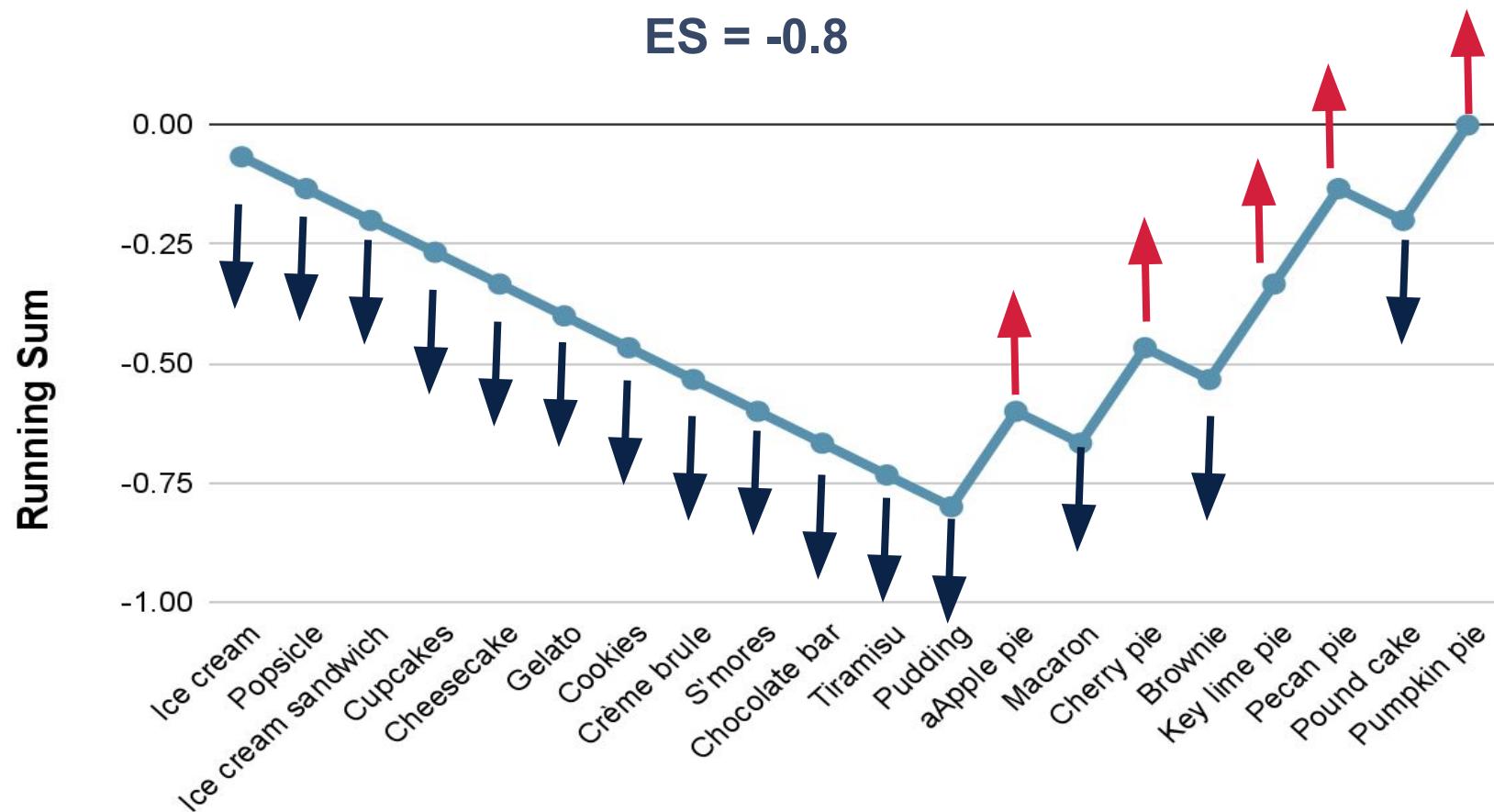
Enrichment score for **Pies**: Apple pie, Pumpkin pie, Pecan pie, Key lime pie, Cherry pie



In the set [Y/N] N N N N N N N N N N Y N Y N Y Y N Y

Enrichment score (ES)

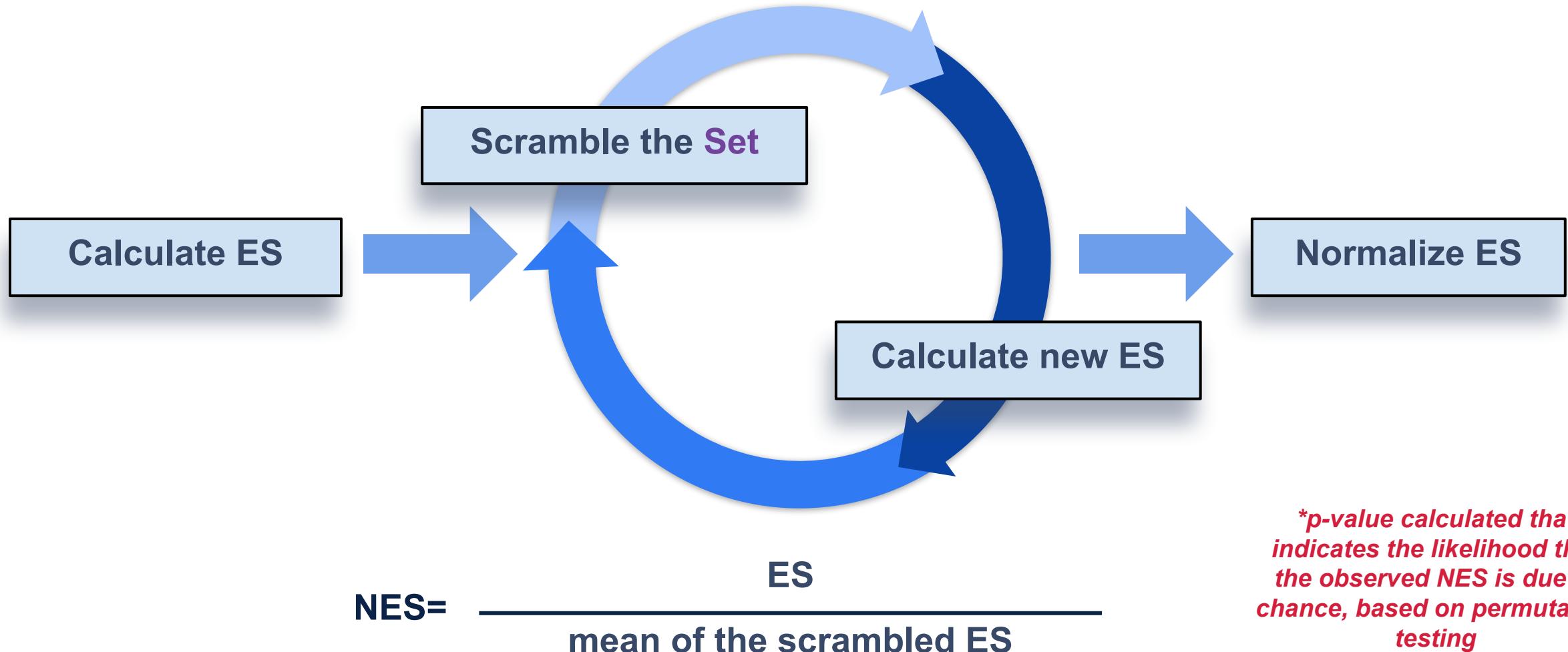
Enrichment score for **Pies**: Apple pie, Pumpkin pie, Pecan pie, Key lime pie, Cherry pie



Pies are less favored by 10xers in the summertime

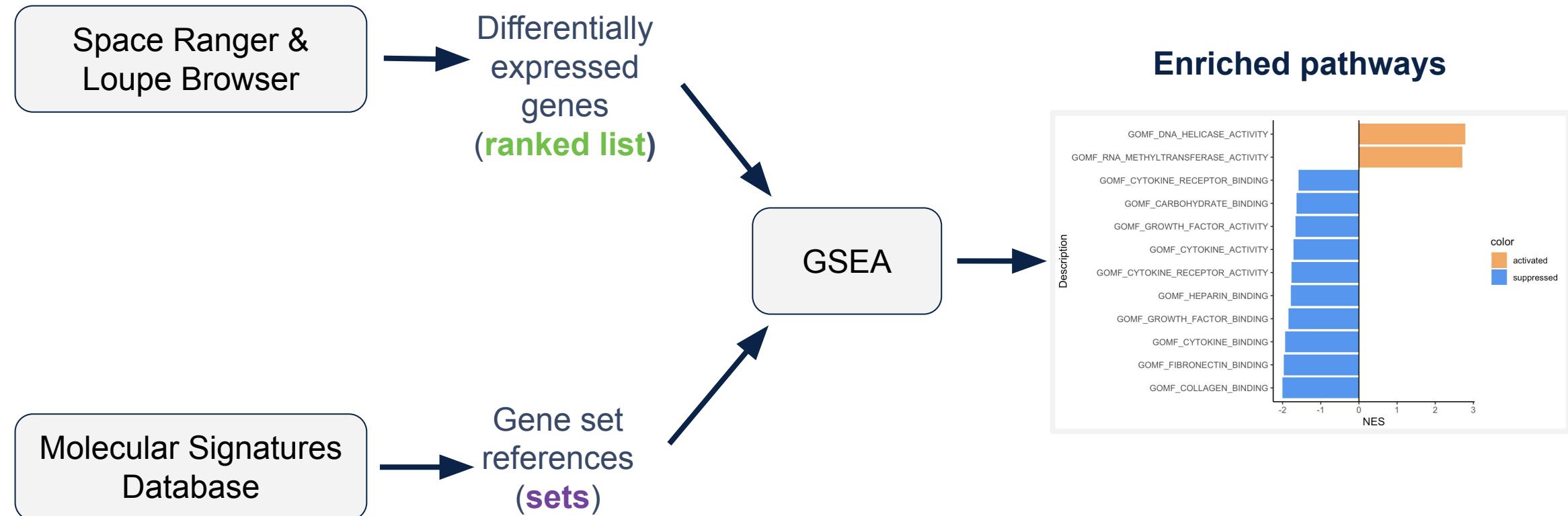
Normalized Enrichment Score (NES)

Normalizes the ES to adjust for set size and allows for comparisons between different sets



Setting up GSEA with our data

General workflow for GSEA



Inputs: gene set references (**set**)

Decide which gene sets to use for analysis

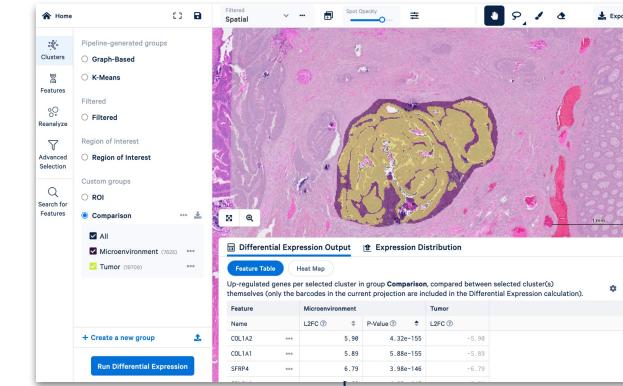
Molecular Signatures Database (MSigDB) is a resource of tens of thousands of annotated gene sets for use in GSEA. Within this larger database, we are going to utilize a collection of gene sets:

Computational Gene Sets defined by mining large collections of cancer-oriented expression data. This collection has three sub-categories - we will use one:

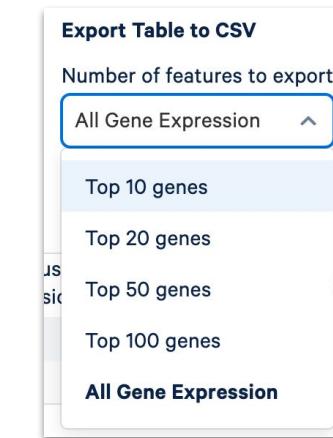
- **Cancer Gene Neighborhoods: defined by expression neighborhoods centered on 380 cancer-associated genes**

Inputs: differentially expressed genes (ranked list)

1. Set up a comparison of interest. In our case tumor versus microenvironment.



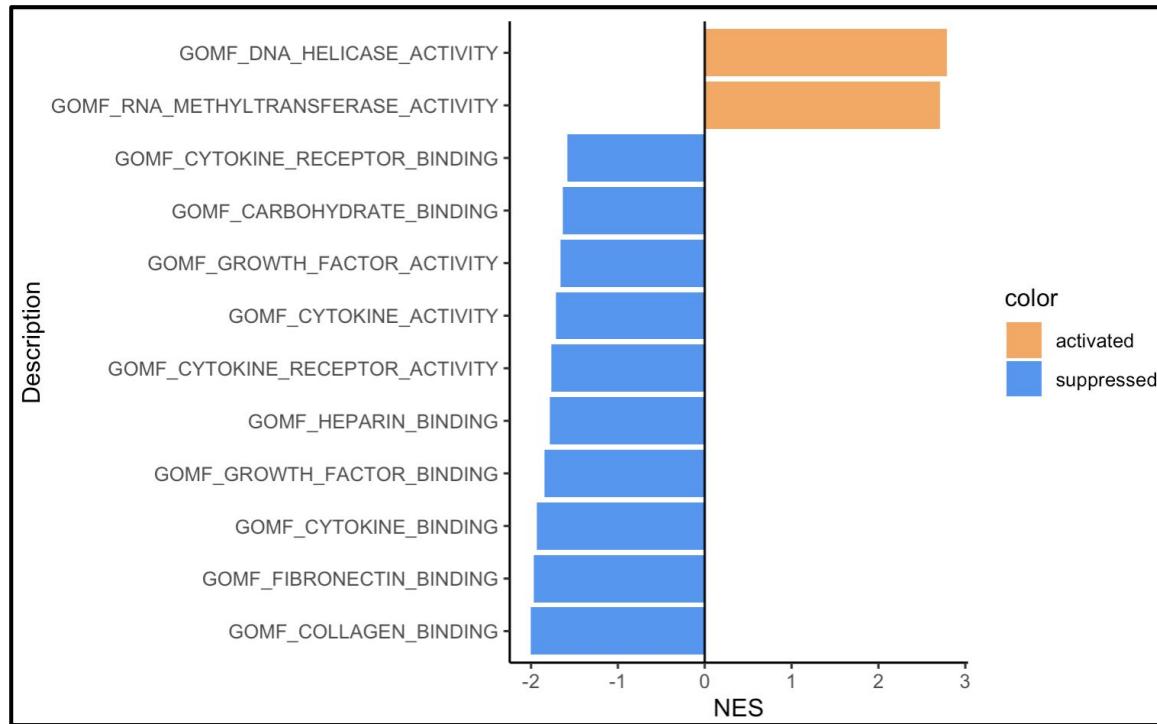
2. Export all gene expression from this comparison. This provides measures of all detected genes in terms of expression in tumor versus periphery.



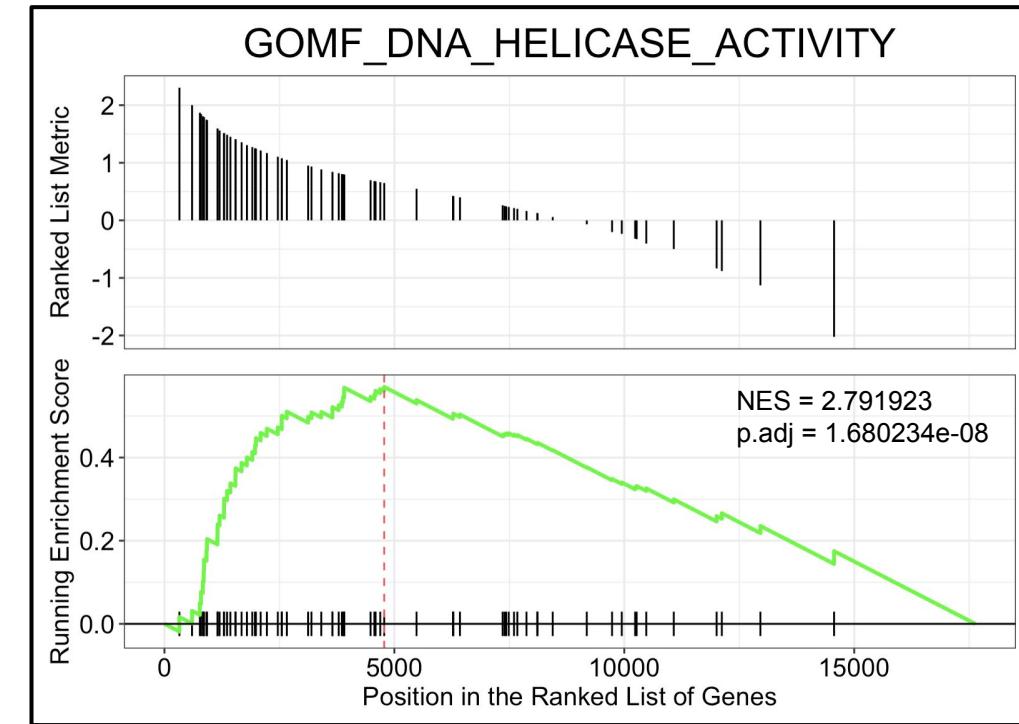
Feature	Microenvironment
Name	L2FC
COMP	***
COL11A1	***
SERpine2	***
CHIT1	***
PRDM6	***
SFRP4	***
CDH2	***

3. Rank order gene expression by Log2FC. This will be our input ranked list.

Outputs: example of GSEA results



Enriched pathways by Normalized Enrichment Score.
See the top activated and suppressed pathways in
your dataset.
Select pathways of interest for deeper exploration.



Enrichment plot. Visualize the enrichment of a specific pathway in your ranked gene list.

Quick introduction to .ipynb and Colab

R Notebook / 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



Google Colaboratory:

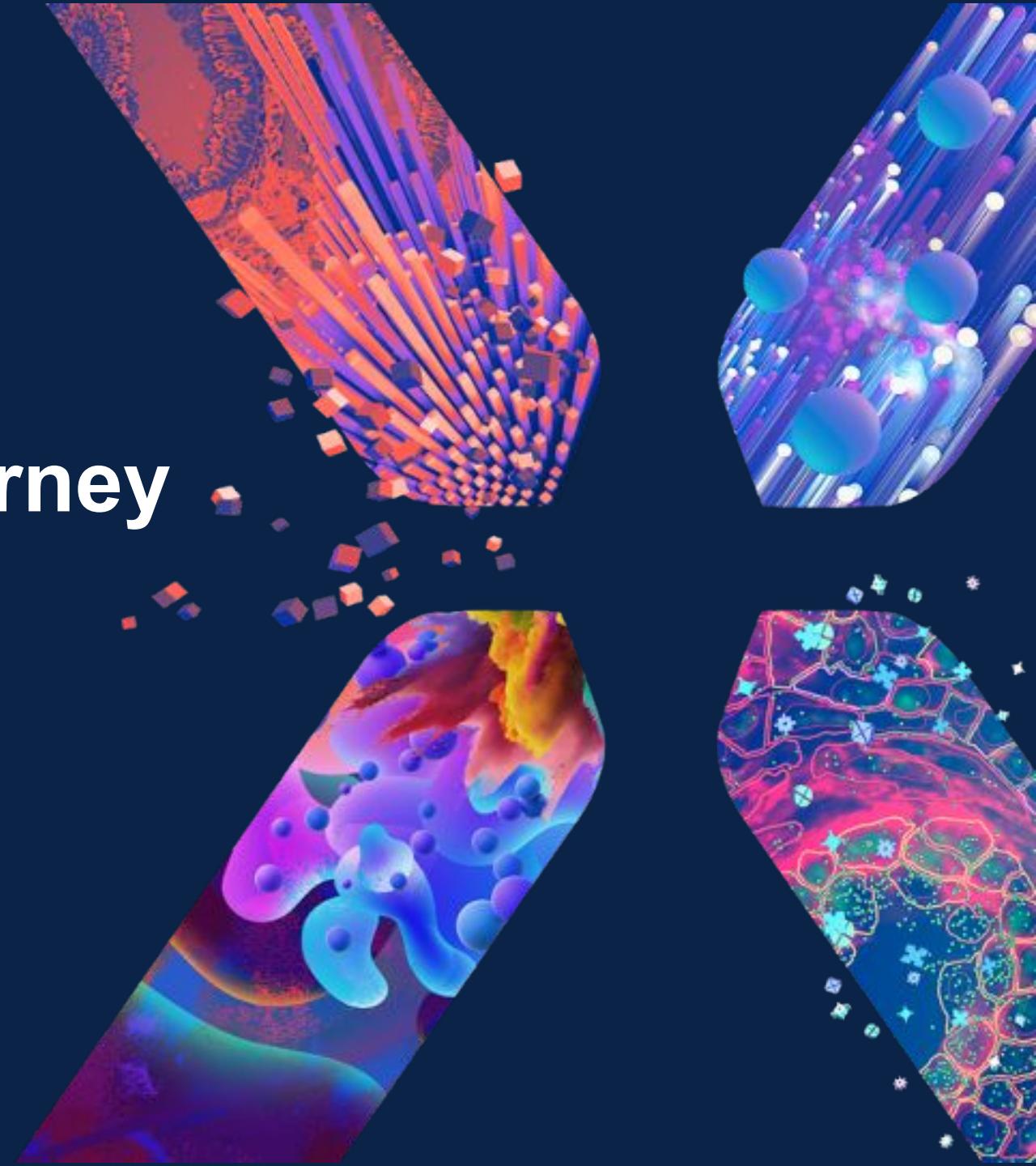
- Jupyter Notebook stored in google drive
- Requires no setup, and runs entirely (writing, running, & sharing code) on the Cloud.



Interactive session

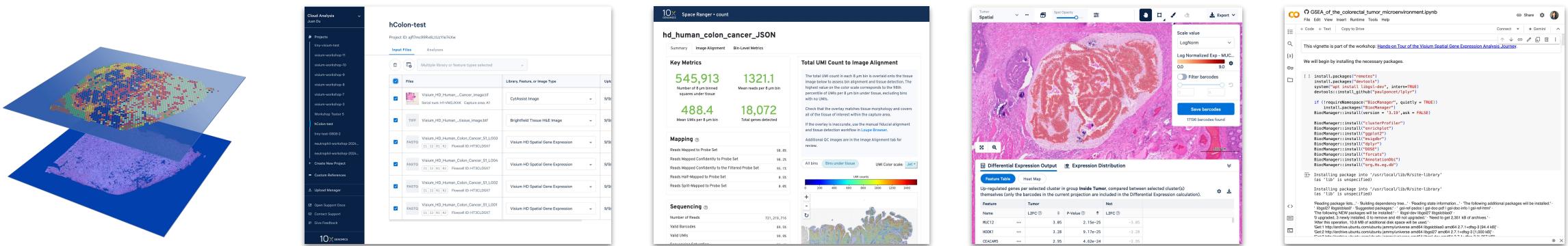
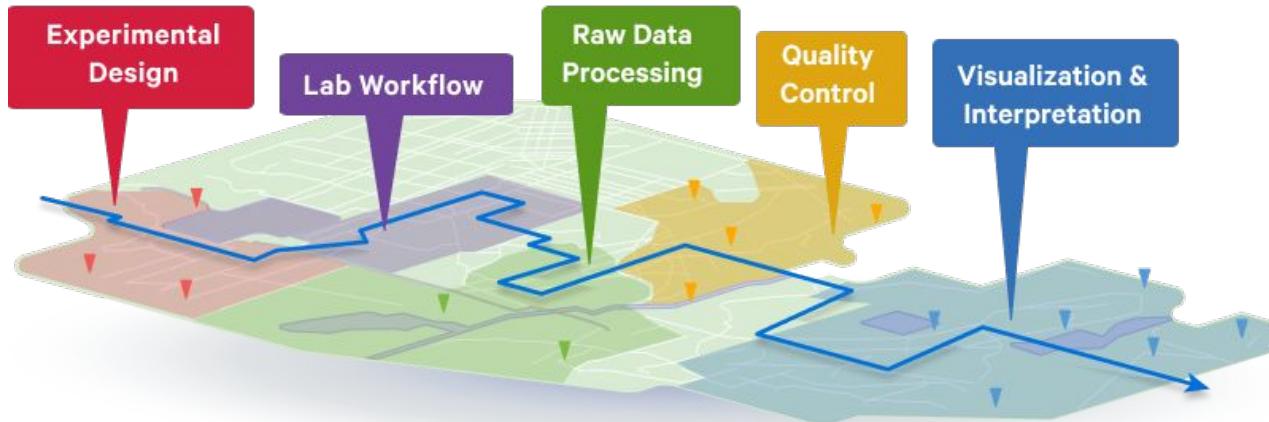


Recap & Plan Your Analysis Journey



Recap

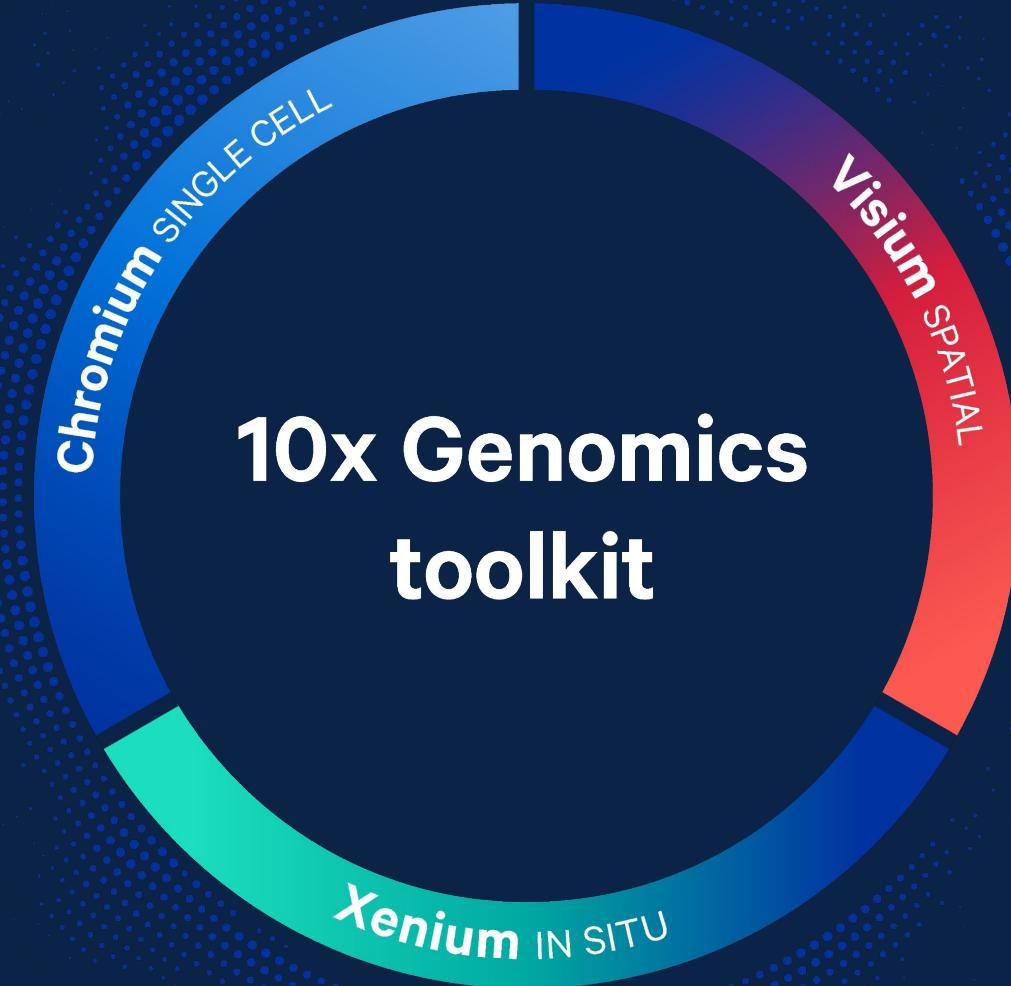
Our journey through analysis



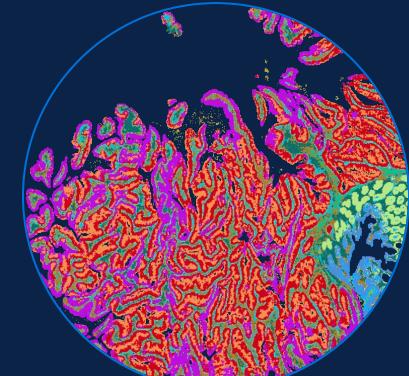
All the slides and material are available on this agenda page:

<https://10xgen.com/scworkshop>

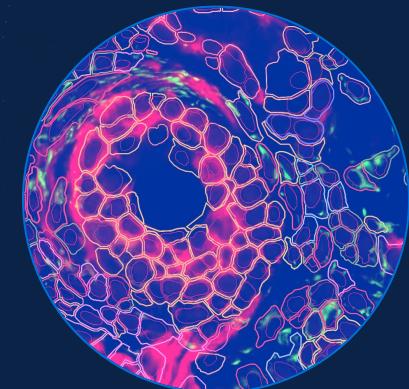
Three platforms to
resolve biology's
complexity



Chromium Single Cell



Visium Spatial



Xenium In Situ

10x Genomics Product Portfolio

Chromium



Unbiased cellular discovery

Whole transcriptome

Transcriptome, VDJ,
CRISPR gRNAs, protein, chromatin

Sequencing-based

Single cell resolution

High per-gene sensitivity and breadth

Visium



Unbiased spatial discovery

Whole transcriptome

Transcriptome, protein, histology

Sequencing-based

Single cell scale

High gene breadth

Xenium



Precise single cell spatial insights

100s–1,000s of transcripts

Targeted RNA, protein, histology

High-resolution imaging-based

Subcellular resolution

High per-gene sensitivity

We are here to support you!

