

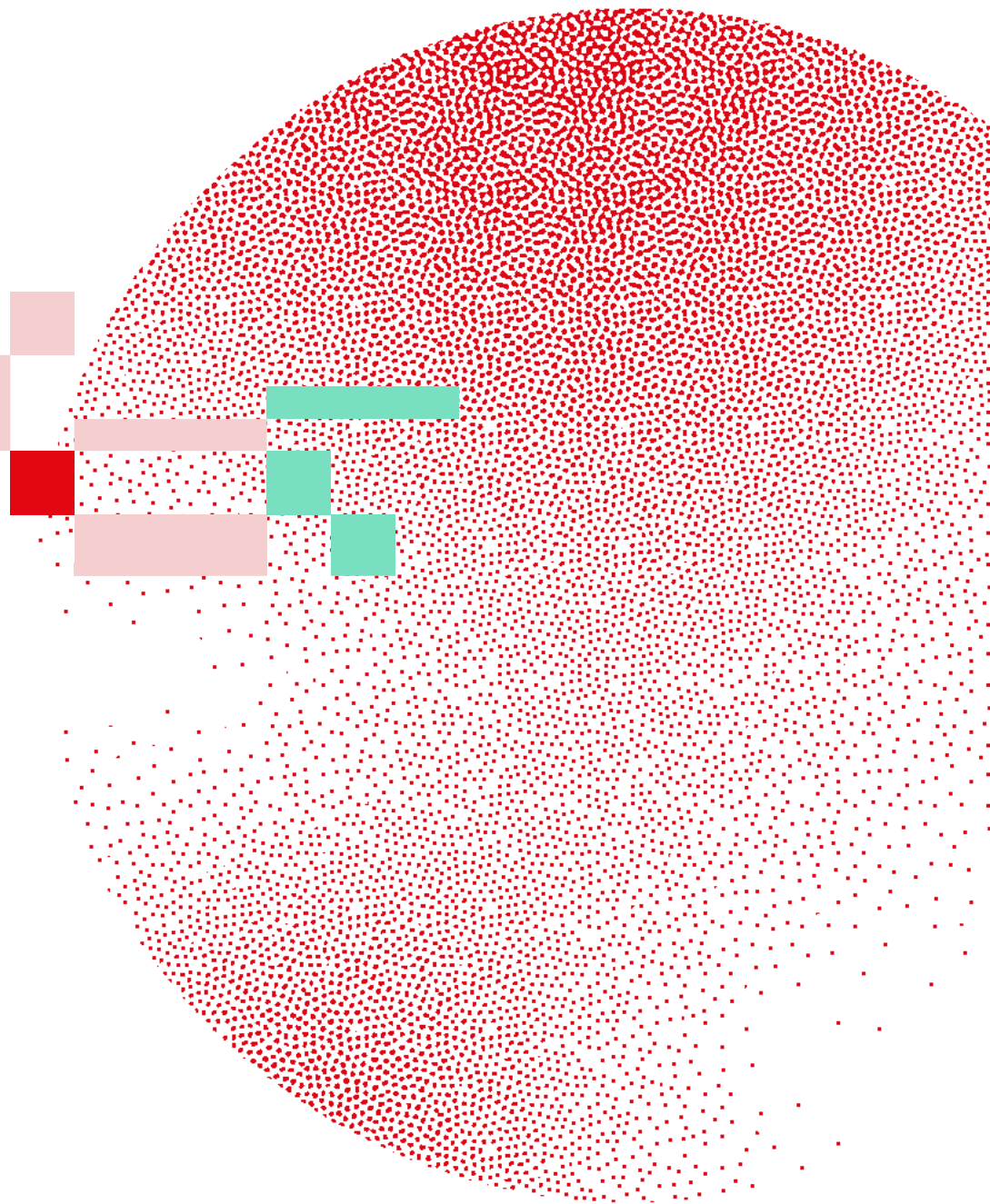


Swiss Institute of
Bioinformatics

INTRODUCTION TO SPATIAL TRANSCRIPTOMICS DATA ANALYSIS

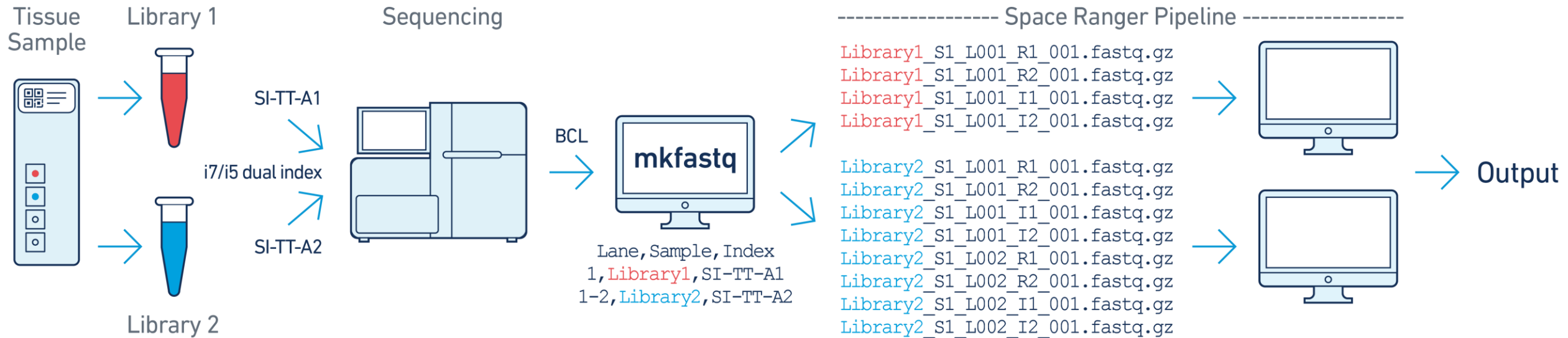
Preprocessing & Data Structures

Deepak Tanwar
June 03-05, 2026



**What is the pipeline (with tools) to
obtain counts by features metrics
for single-cell transcriptomics
data?**

Workflow of obtaining read counts



What does this sample name
mean?

`Library1_S1_L001_R1_001.fastq.gz`

Library1_S1_L001_R1_001.fastq.gz

Component	Meaning
Library1	Library name from tissue sample
S1	Sample number
L001	Lane number
R1/R2	Read direction (forward/reverse)
I1/I2	Index reads (i7/i5)
001	Chunk number
.fastq.gz	FASTQ format, gzip compressed

Space-ranger

binned_output :

└ square_002um

└ square_008um

└ square_016um

└── analysis

└── cloupe.cloupe

└── filtered_feature_bc_matrix

└── filtered_feature_bc_matrix.h5

└── spatial

| └── LV123

| └── scalefactors_json.json

| └── tissue_hires_image.png

└── web_summary.html

Space-ranger output

Component	Explanation
square_002um / square_008um / square_016um	Binned gene expression matrices at different resolutions (2 μm , 8 μm , 16 μm)
analysis/	Standard downstream analysis results (clustering, PCA, t-SNE, UMAP, differential expression)
cloupe.cloupe	Interactive visualization file for Loupe Browser
filtered_feature_bc_matrix/	Filtered gene-barcode (spot) count matrix in Matrix Market format
filtered_feature_bc_matrix.h5	Same filtered count matrix in HDF5 format
spatial/	Spatial information: tissue image, scale factors, alignment files
web_summary.html	HTML report with quality control metrics and summary statistics

Space-ranger output

Component	Explanation
<code>square_002um / square_008um / square_016um</code>	Binned gene expression matrices at different resolutions (2 μm, 8 μm, 16 μm)
<code>analysis/</code>	Standard downstream analysis results (clustering, PCA, t-SNE, UMAP, differential expression)
<code>cloupe.cloupe</code>	Interactive visualization file for Loupe Browser
<code>filtered_feature_bc_matrix/</code>	Filtered gene-barcode (spot) count matrix in Matrix Market format
<code>filtered_feature_bc_matrix.h5</code>	Same filtered count matrix in HDF5 format
<code>spatial/</code>	Spatial information: tissue image, scale factors, alignment files
<code>web_summary.html</code>	HTML report with quality control metrics and summary statistics

Visium HD: binning strategy

square_002um

1 native capture spot

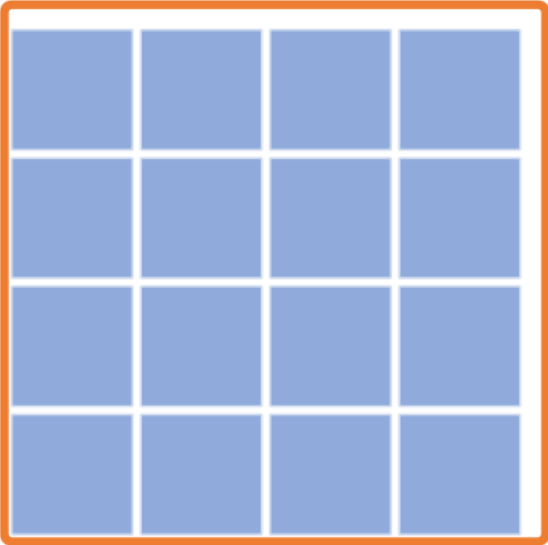


2 μm × 2 μm
highest resolution
most sparse



square_008um

pool of 16 spots (4×4)

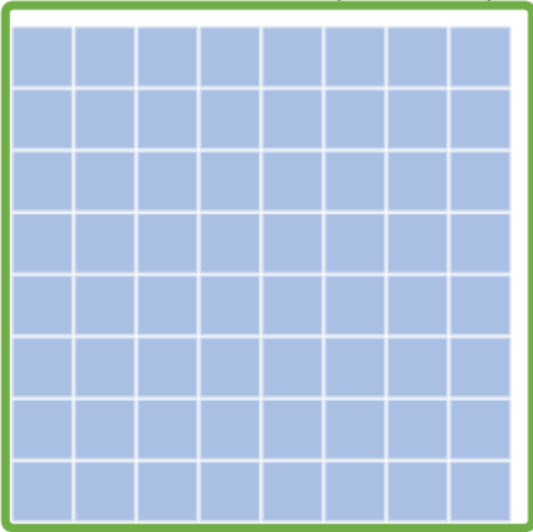


8 μm × 8 μm
~cell-scale
balanced depth



square_016um

pool of 64 spots (8×8)



16 μm × 16 μm
multi-cell, deeper counts
used in this course ✓

Which bin size should you use?

2 μm – sub-cellular resolution, very sparse counts. Use only if sub-cellular structure matters.

8 μm – roughly cell-scale, moderate depth. Good for detailed analysis.

16 μm – multi-cell bins, densest counts. Best for first-pass, QC

→ In this course we use `bin_size = "016"` for speed and count depth.

Space-ranger output

Component	Explanation
square_002um / square_008um / square_016um	Binned gene expression matrices at different resolutions (2 μm , 8 μm , 16 μm)
analysis/	Standard downstream analysis results (clustering, PCA, t-SNE, UMAP, differential expression)
cloupe.cloupe	Interactive visualization file for Loupe Browser
filtered_feature_bc_matrix/	Filtered gene-barcode (spot) count matrix in Matrix Market format
filtered_feature_bc_matrix.h5	Same filtered count matrix in HDF5 format
spatial/	Spatial information: tissue image, scale factors, alignment files
web_summary.html	HTML report with quality control metrics and summary statistics

Space-ranger: [web_summary.html](#)

Key Metrics

393,543

Number of 8 μ m binned squares under tissue

849.2

Mean reads per 8 μ m bin

226.4

Mean UMIs per 8 μ m bin

18,991

Total genes detected

Sequencing metrics

Q30 bases in RNA read

Sequencing saturation

Number of reads

Mapping metrics

Reads mapped to genome

Reads mapped confidently to transcriptome

Reads mapped to intronic regions

Spot/bin metrics

Number of spots under tissue

Median genes per spot

Median UMI counts per spot

Tissue detection QC

Tissue image + spot overlay

Clustering (UMAP/t-SNE) preview

Mapping ?

Reads Mapped to Genome	85.1%
------------------------	-------

Reads Mapped Confidently to Genome	81.6%
------------------------------------	-------

Reads Mapped Confidently to Intergenic Regions	5.9%
------------------------------------------------	------

Reads Mapped Confidently to Intronic Regions	1.6%
----------------------------------------------	------

Reads Mapped Confidently to Exonic Regions	74.1%
--------------------------------------------	-------

Reads Mapped Confidently to Transcriptome	72.7%
-------------------------------------------	-------

Reads Mapped Antisense to Gene	0.6%
--------------------------------	------

Space-ranger with segmentation

binned_output :
├ square_002um

.

.

├── spatial
└── web_summary.html

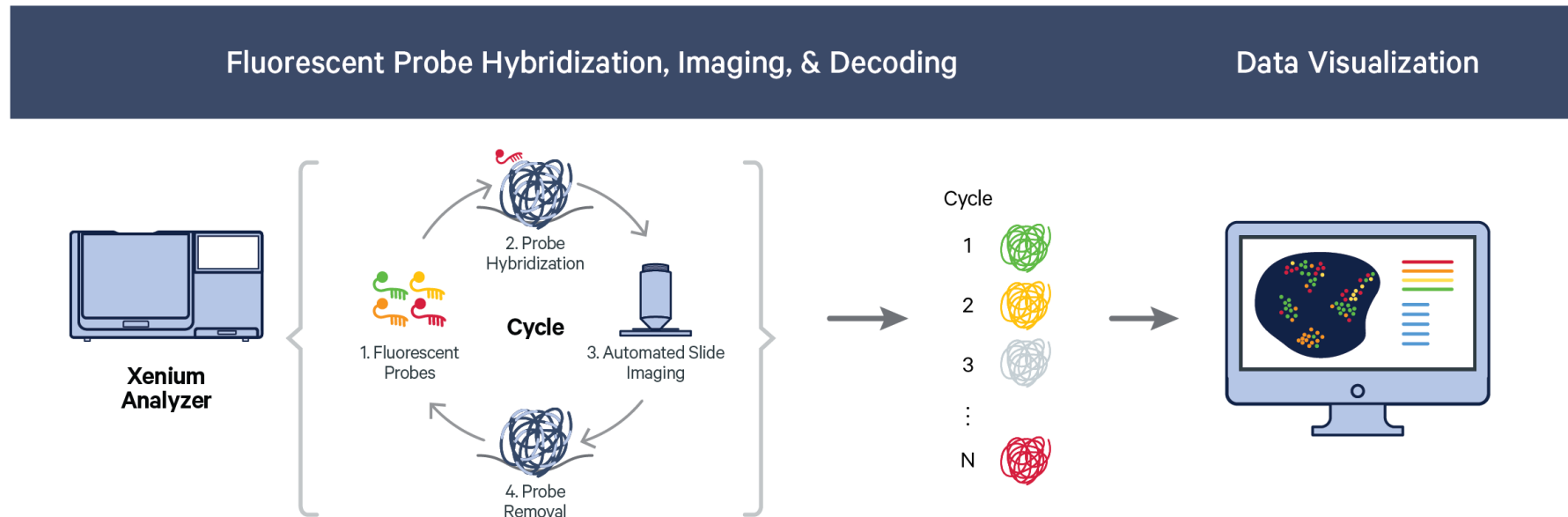
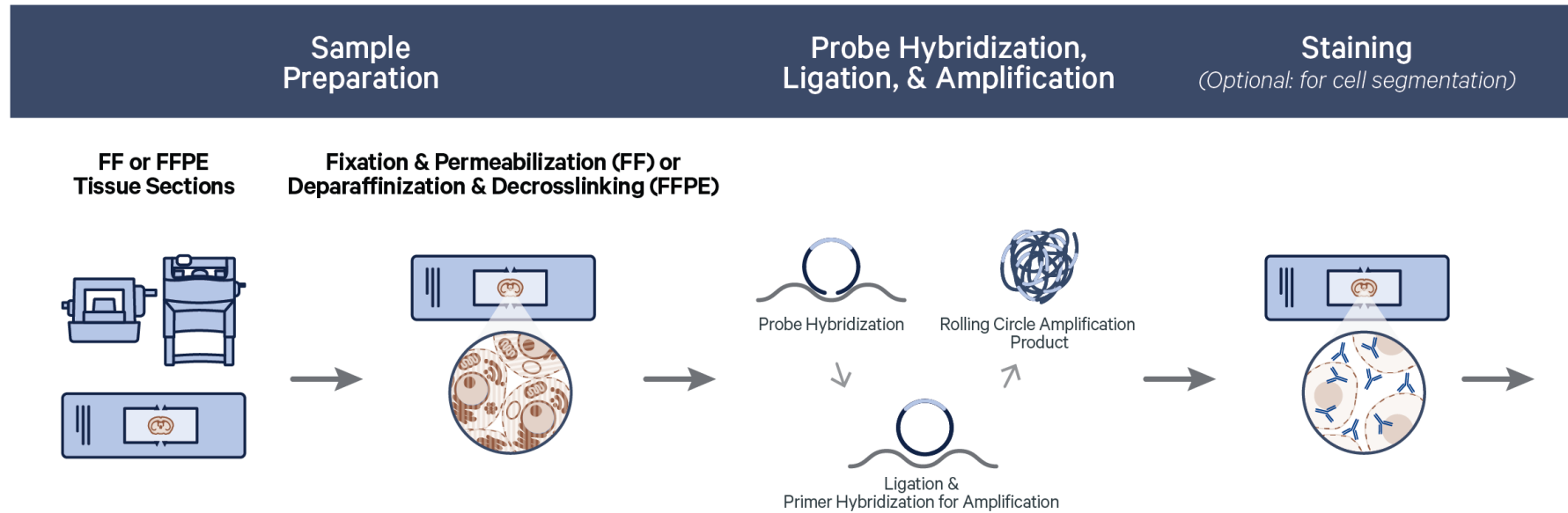
+ segmented_outputs : ← Cell segmentation output

├── analysis
├── **cell_segmentations.geojson**
├── cloupe.cloupe
├── filtered_feature_cell_matrix
├── filtered_feature_cell_matrix.h5
├── graphclust_annotated_cell_segmentations.geojson
├── nucleus_segmentations.geojson
├── raw_feature_cell_matrix
├── raw_feature_cell_matrix.h5
├── spatial
└── Visium_HD_Human_Colon_Cancer_barcode_mappings.parquet

Space-ranger segmentation output

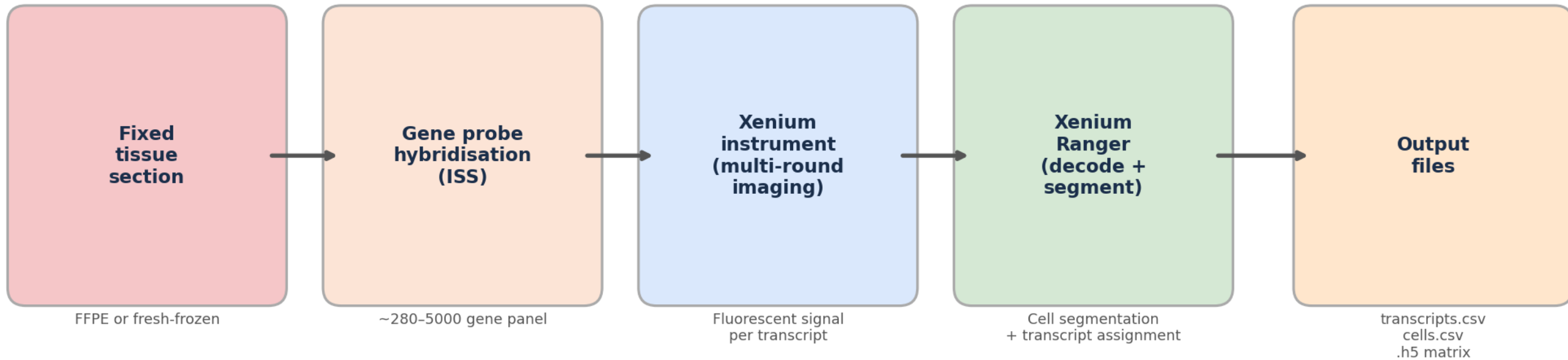
Component	Explanation
cell_segmentations.geojson	Polygon boundaries of detected cells
nucleus_segmentations.geojson	Polygon boundaries of detected nuclei
graphclust_annotated_cell_segmentations.geojson	Cell polygons with graph-based clustering annotations
graphclust_annotated_nucleus_segmentations.geojson	Nucleus polygons with clustering annotations
filtered_feature_cell_matrix/	Gene expression matrix at single-cell resolution (filtered)
filtered_feature_cell_matrix.h5	Same cell-level count matrix in HDF5 format
raw_feature_cell_matrix/	Unfiltered (raw) gene expression matrix at cell level
Visium_HD_Human_Colon_Cancer_barcode_mappings.parquet	Critical mapping file linking binned spots to segmented cells
spatial/	Spatial files specific to the segmented data
analysis/	Analysis results at cell level

Xenium: pipeline overview



Xenium: pipeline overview

Imaging-based — no sequencing required



Xenium: output files

Key difference vs. Visium HD

No spot grid — coordinates are in **physical microns**

No spatial/ folder or scalefactors

Alignment tool: **Xenium Ranger / Explorer**

transcripts.csv.gz

One row per detected transcript
x_location, y_location (microns)

cells.csv.gz

Cell metadata: centroid, area, QC metrics

cell_feature_matrix.h5

Cells x genes count matrix (HDF5)

nucleus_boundaries.csv.gz

Polygon vertices per nucleus

morphology.ome.tif

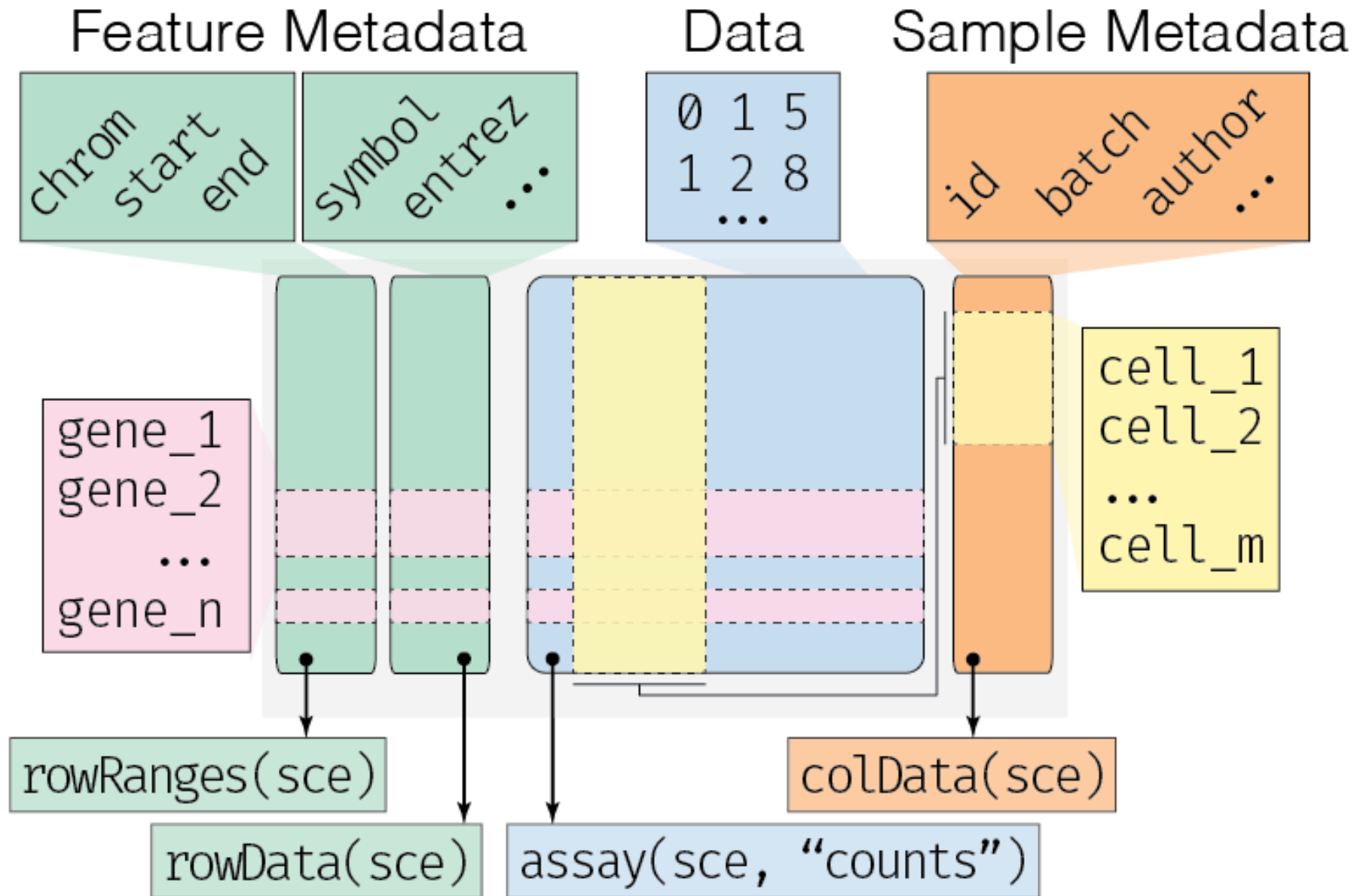
DAPI / morphology image (full resolution)

**Which framework have you used
for single-cell transcriptomics
data analysis?**

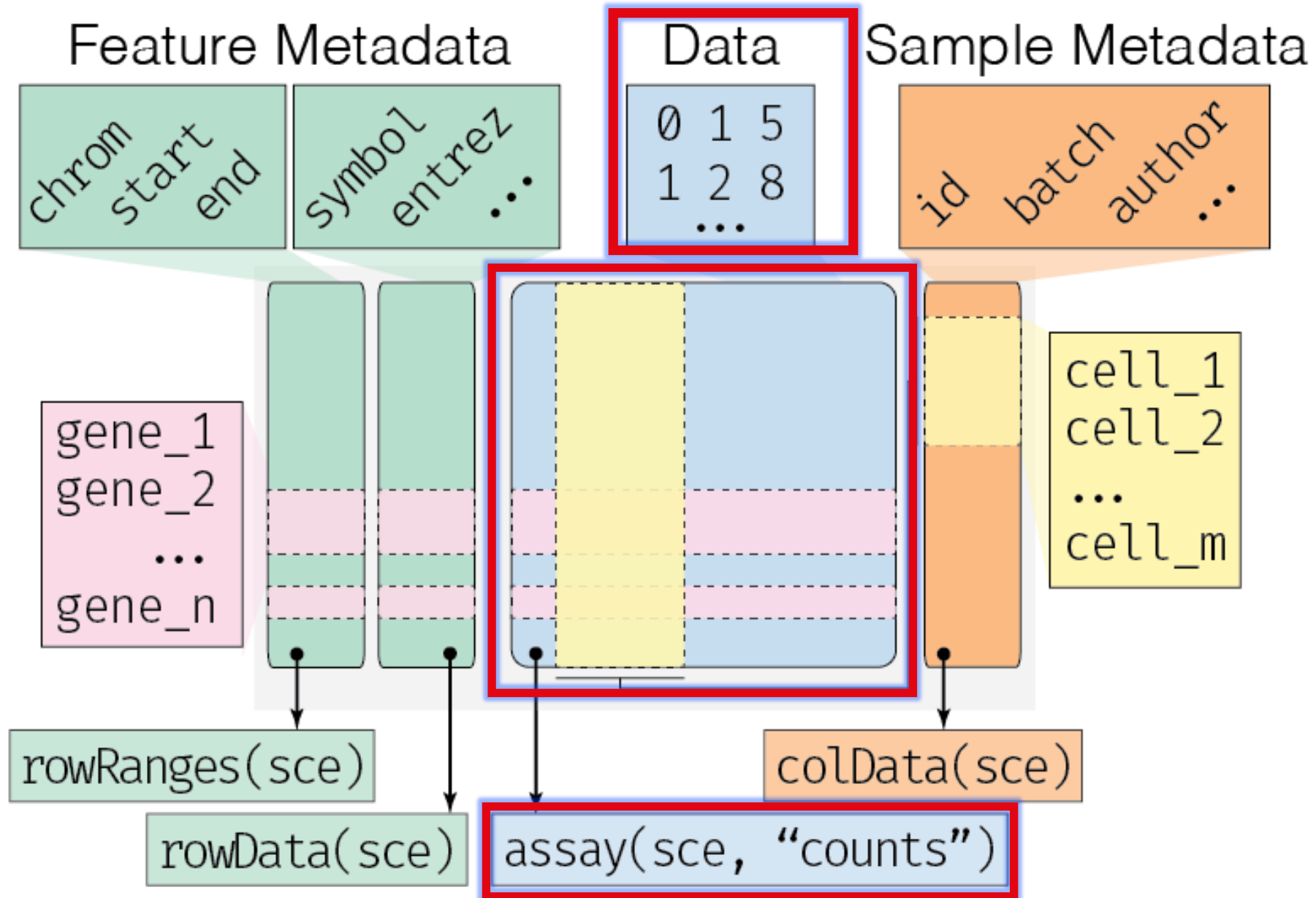
**Which framework have you used
for single-cell transcriptomics
data analysis?**

Seurat?

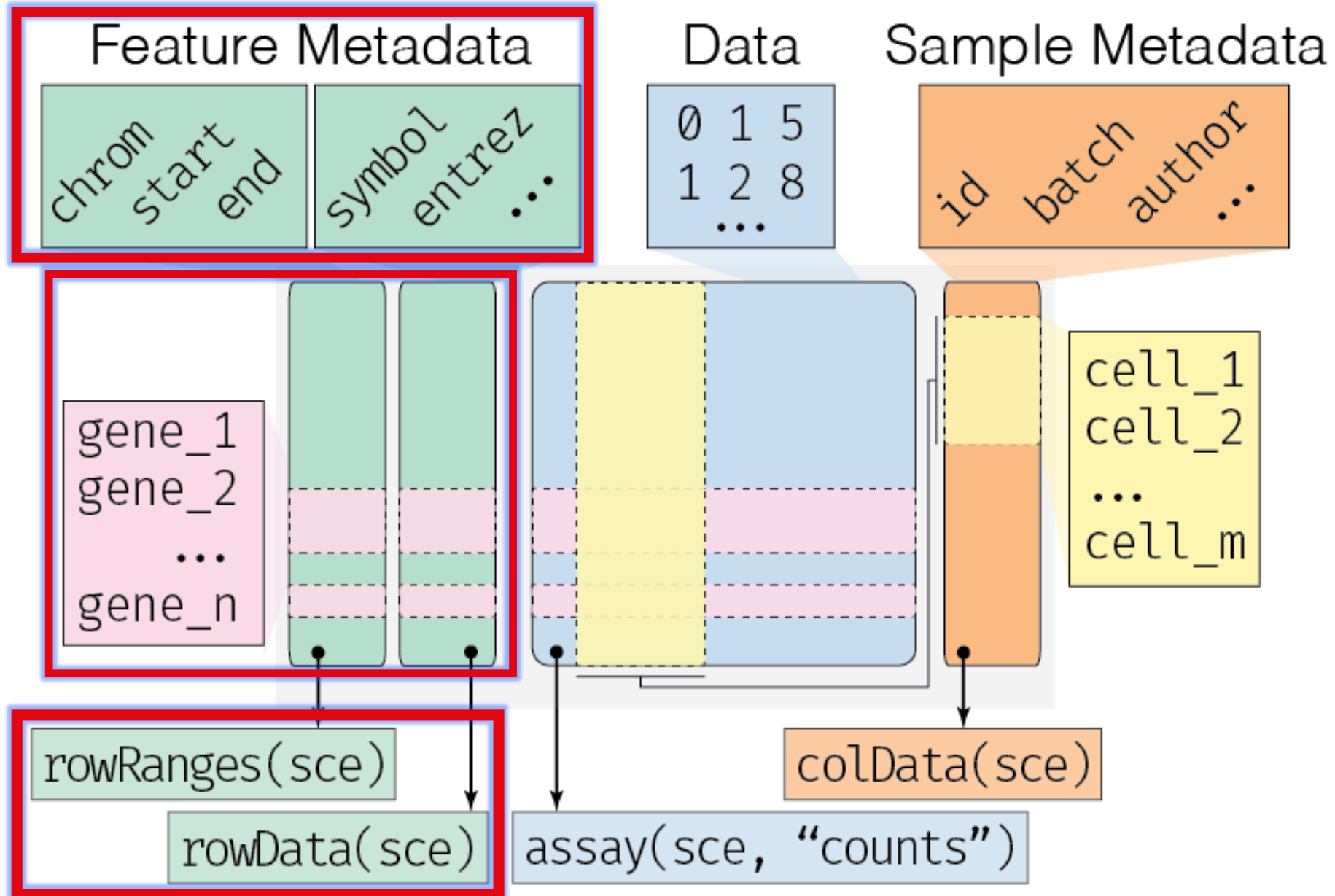
SingleCellExperiment (SCE)



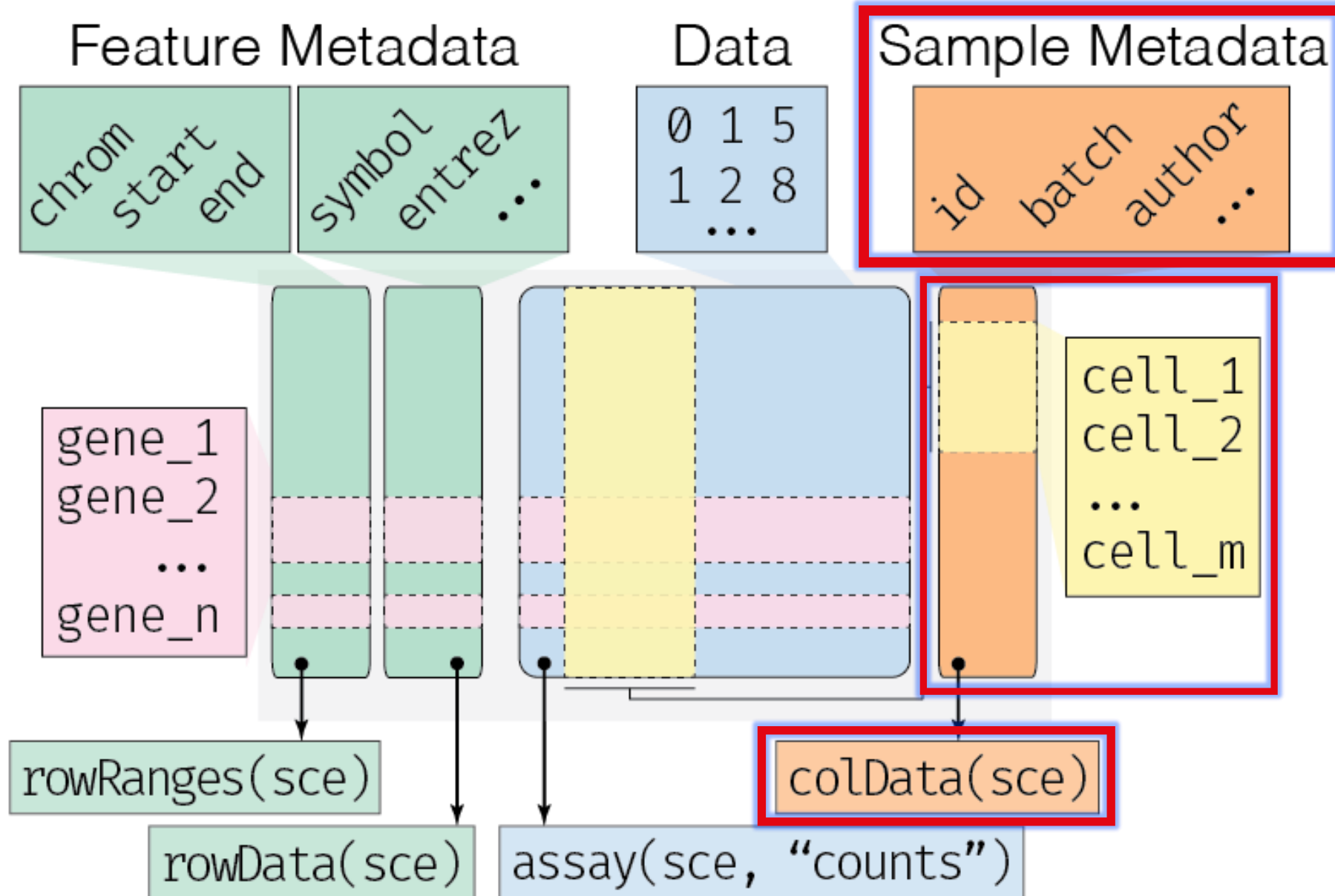
SingleCellExperiment



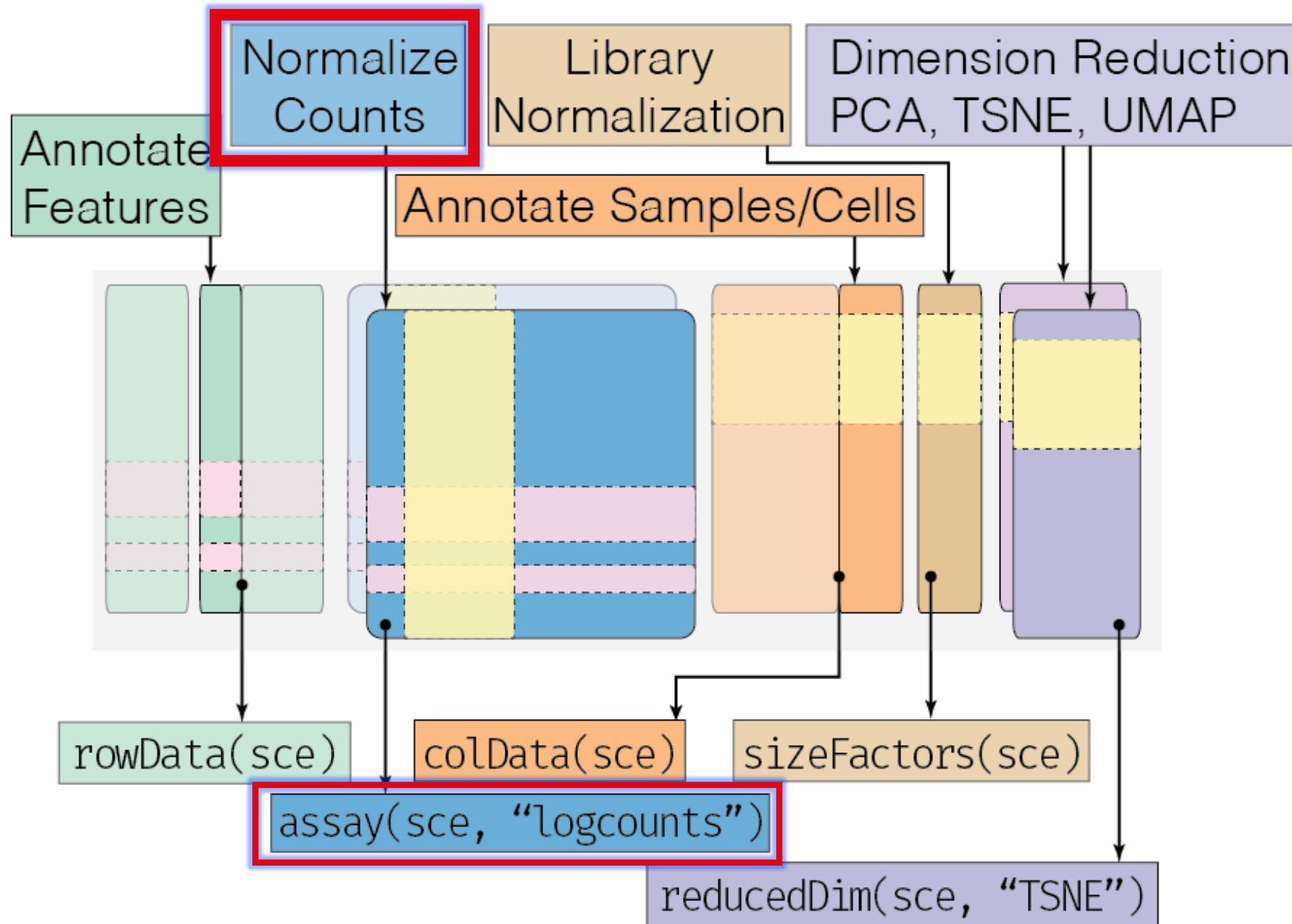
SingleCellExperiment



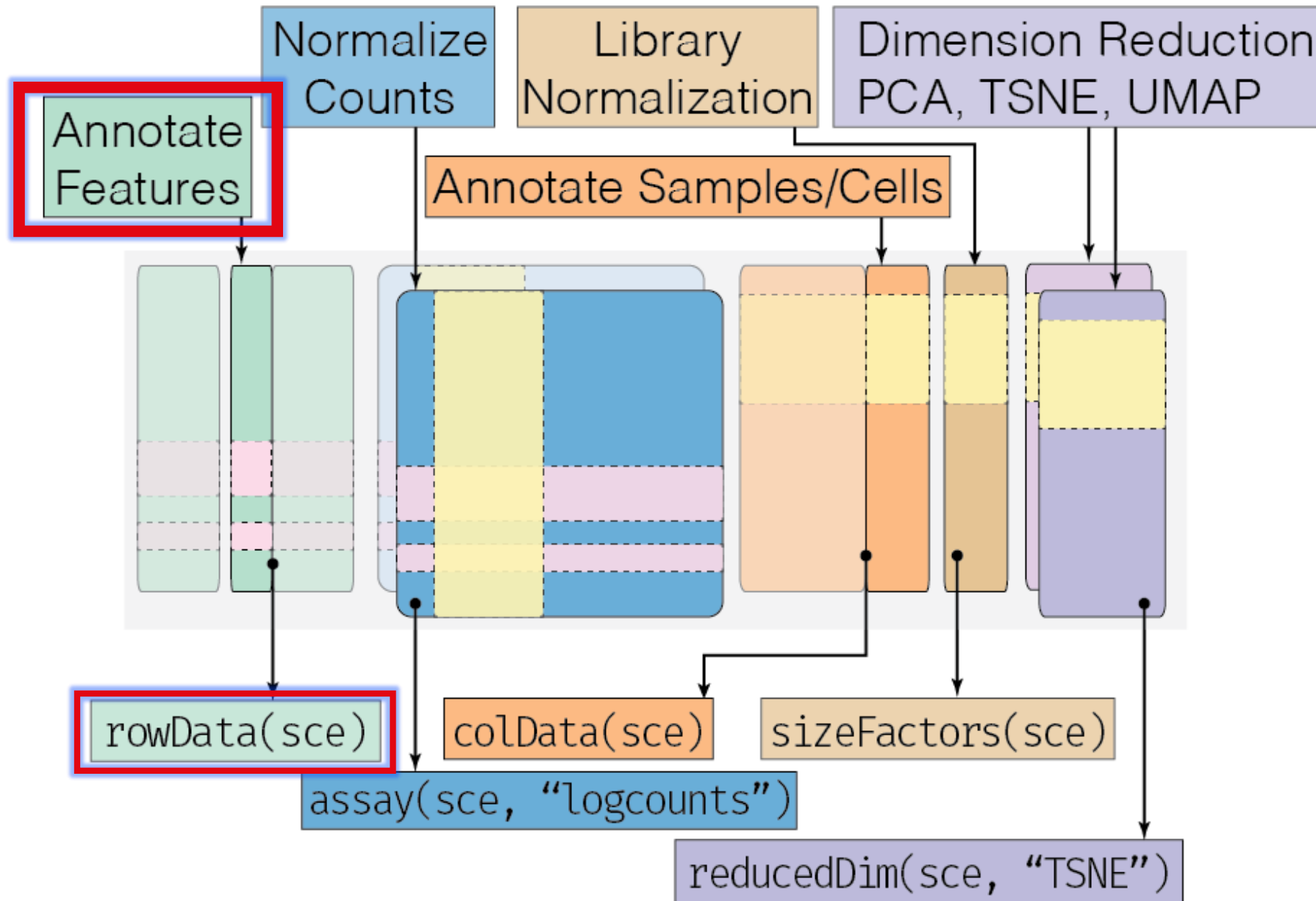
SingleCellExperiment



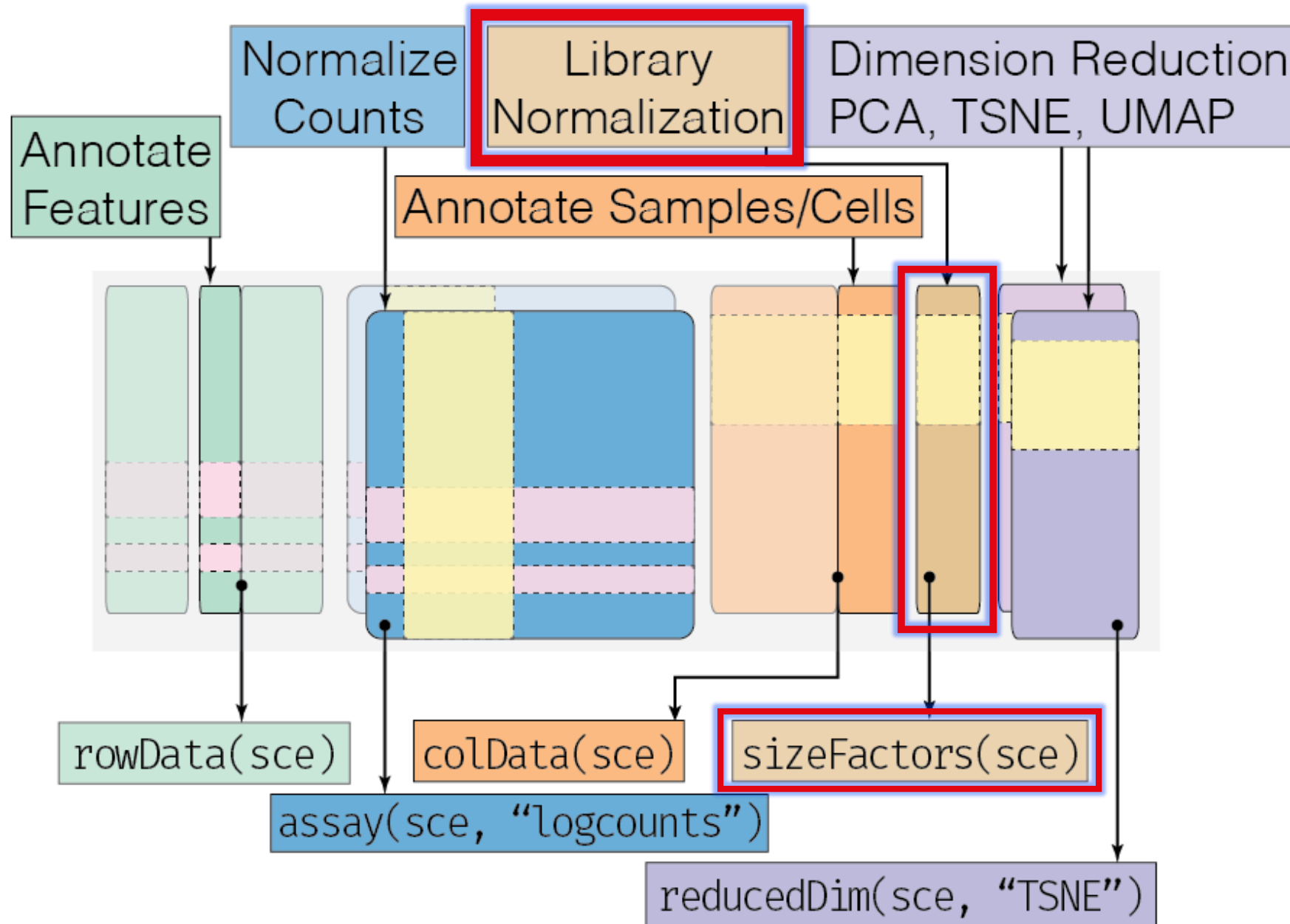
Extending SingleCellExperiment



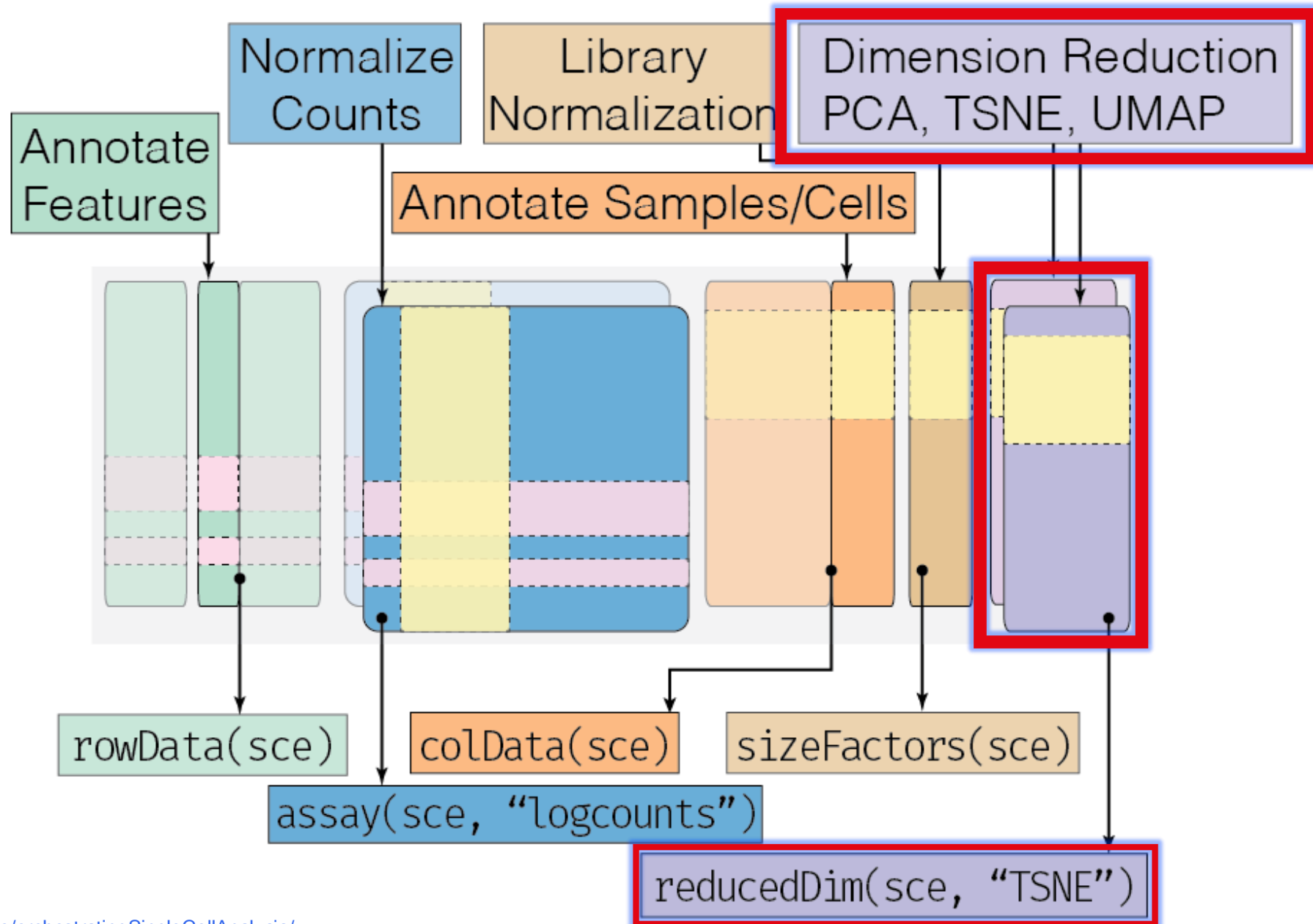
Extending SingleCellExperiment



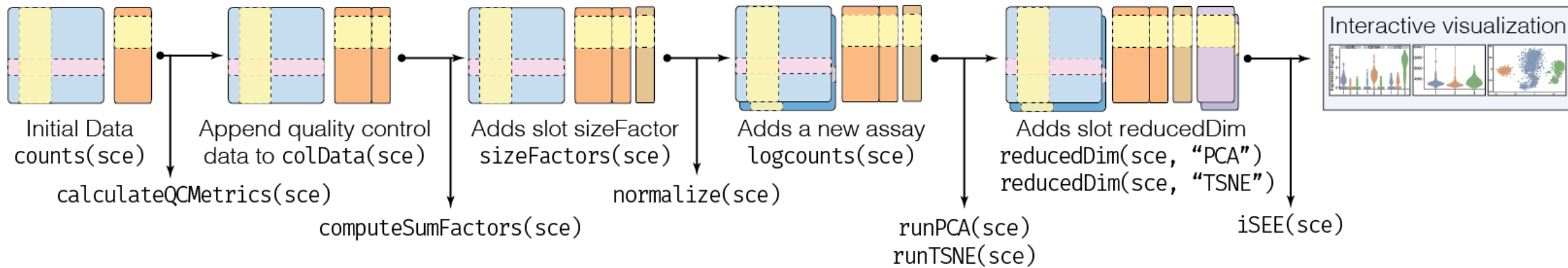
Extending SingleCellExperiment



Extending SingleCellExperiment



Example workflow of SingleCellExperiment extension



Can we use
SingleCellExperiment
for spatial transcriptomics data?

SpatialExperiment (SPE)

SpatialExperiment

SingleCellExperiment

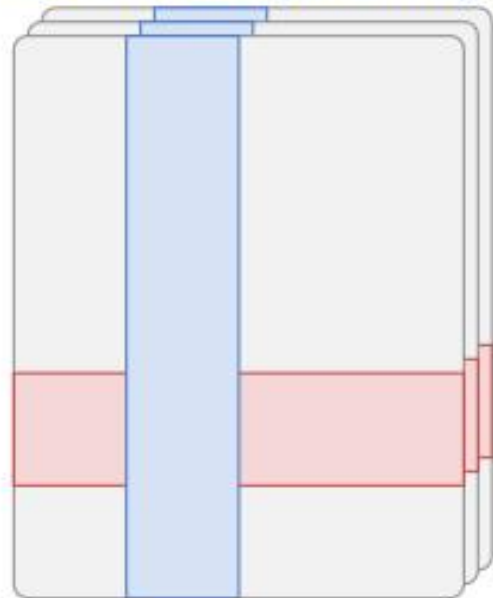
observations (columns)

features (rows)



rowData

feature
metadata



assays

primary and
transformed data

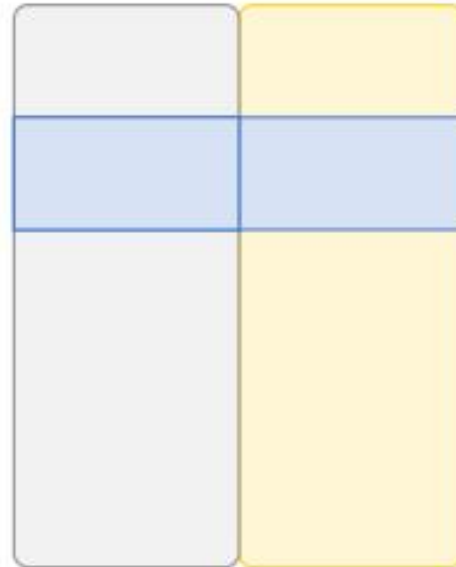


reducedDims

reduced
dimensions

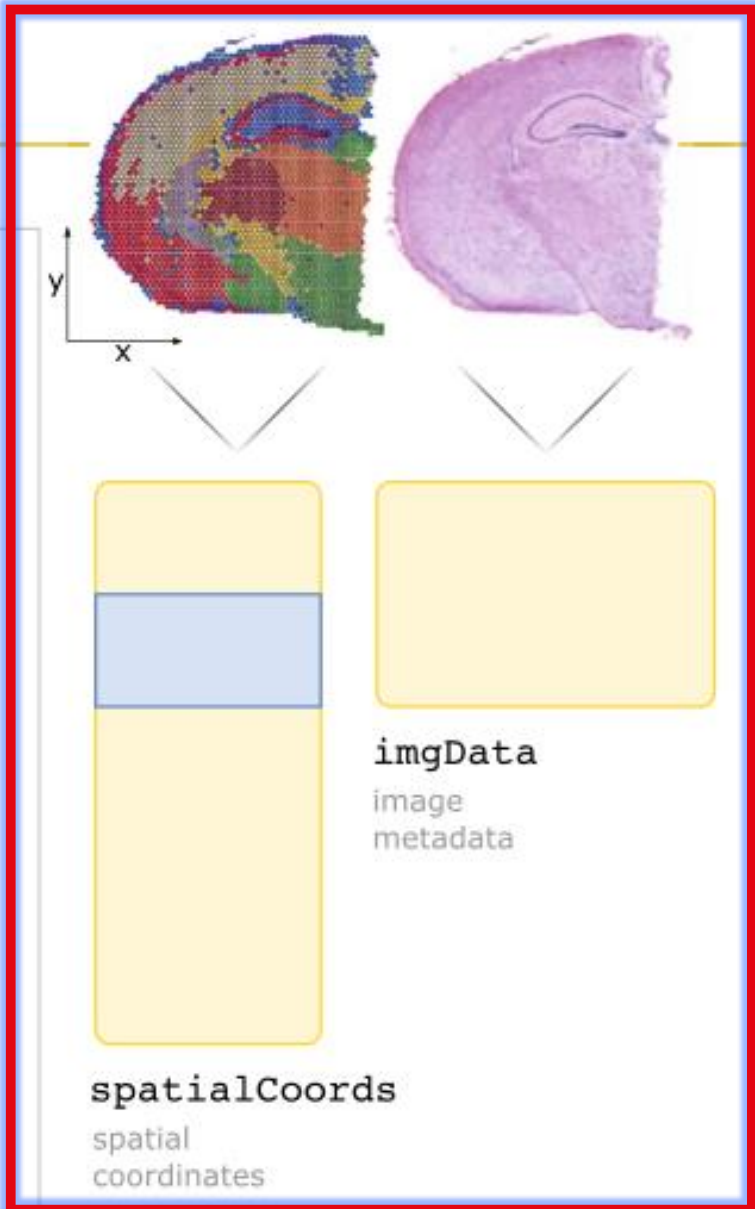
sample_id
condition
cell_type
...

barcode
in_tissue
array_row
...

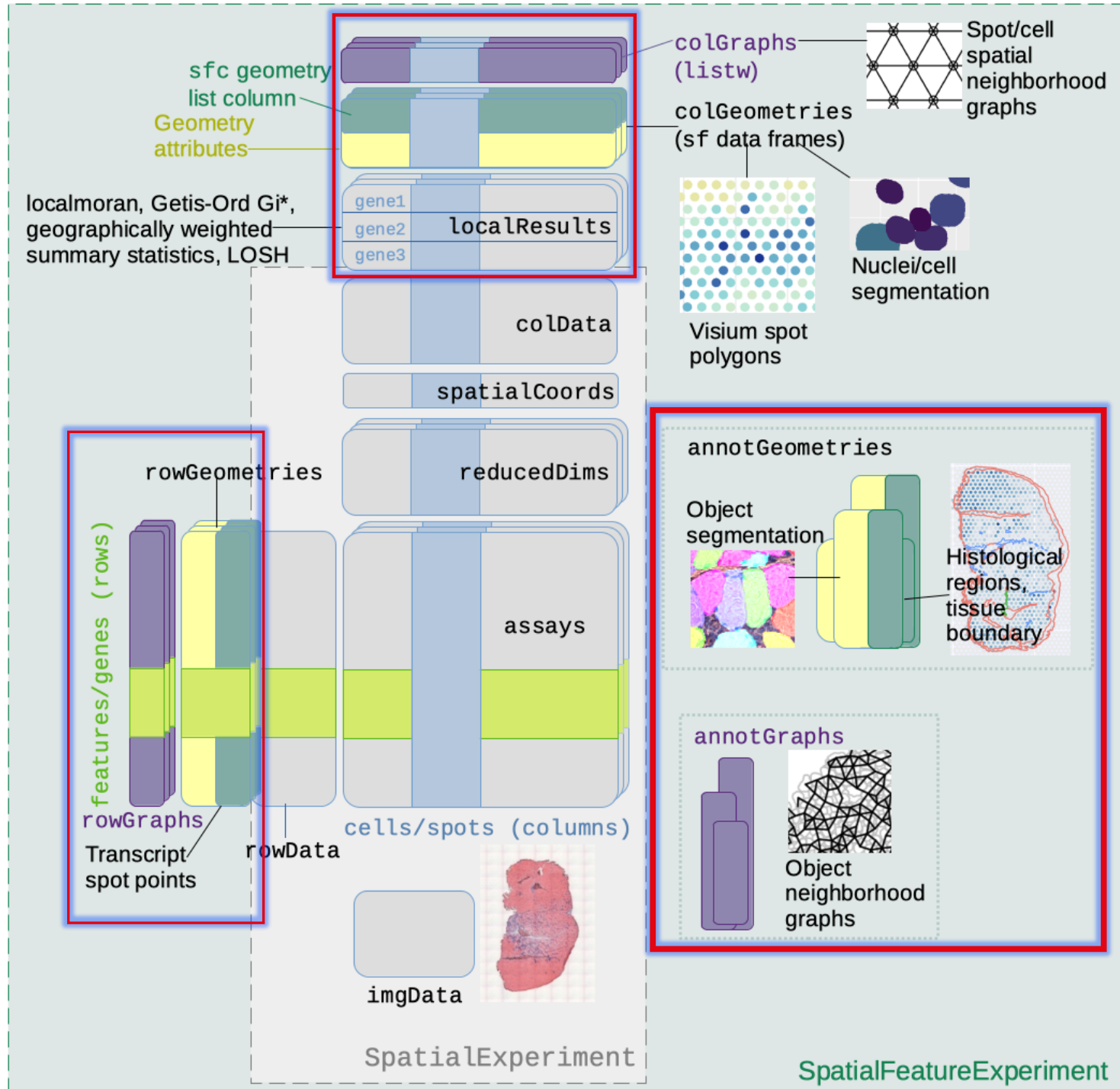


colData

observation
metadata spatial
metadata



SpatialFeature Experiment (SFE)



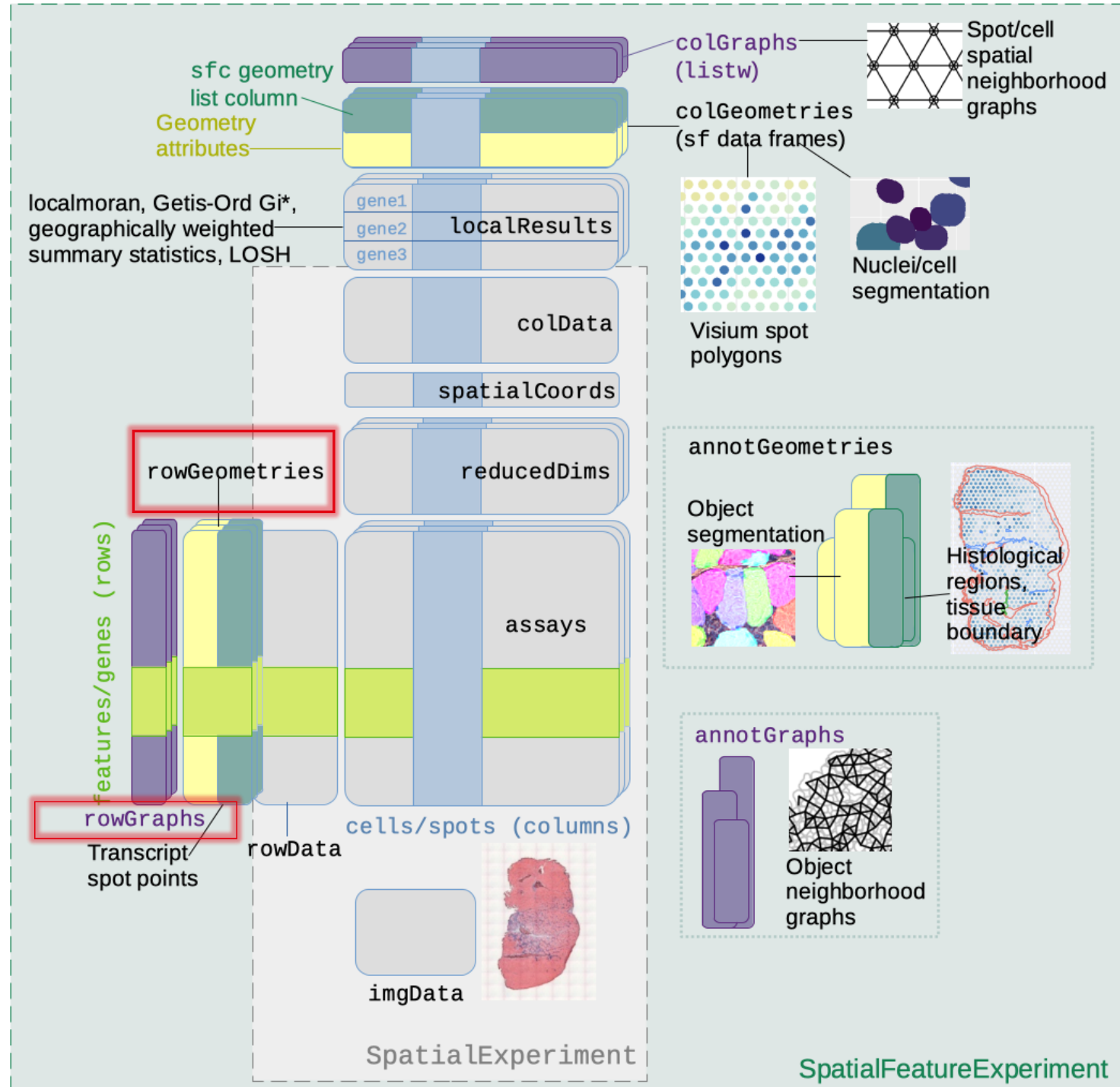
SpatialFeature Experiment

rowGeometries: transcript location points per gene

The exact spatial coordinates (X, Y) of individual detected RNA molecules.

rowGraphs: transcript spatial graphs

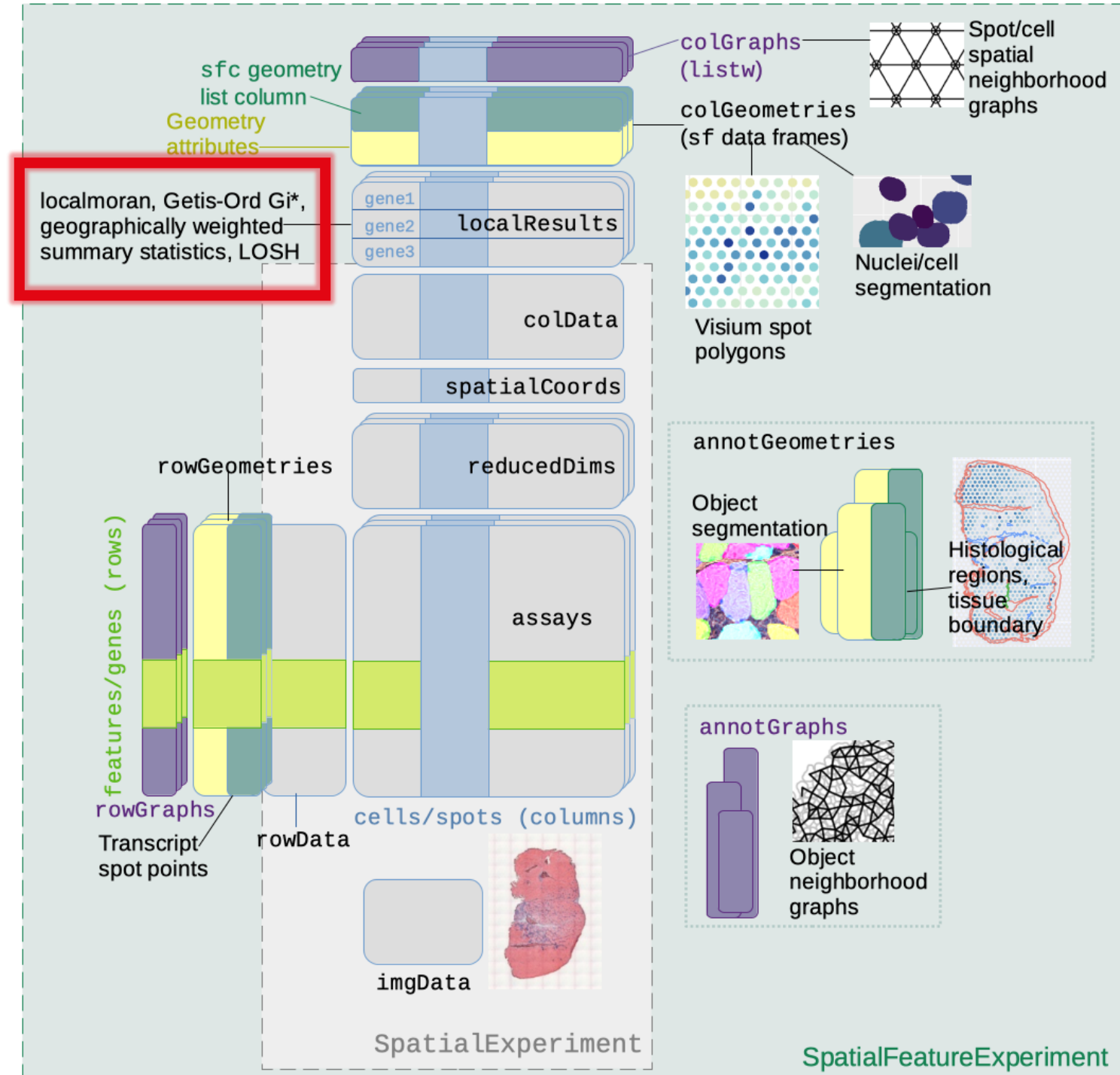
Network connections between individual transcript points to analyze molecule clustering.



SpatialFeature Experiment

localResults: per-gene local spatial statistics (Moran's I, Getis-Ord G_i^* , LOSH...)

Mathematical scores showing exactly where specific genes are significantly clustered or isolated.



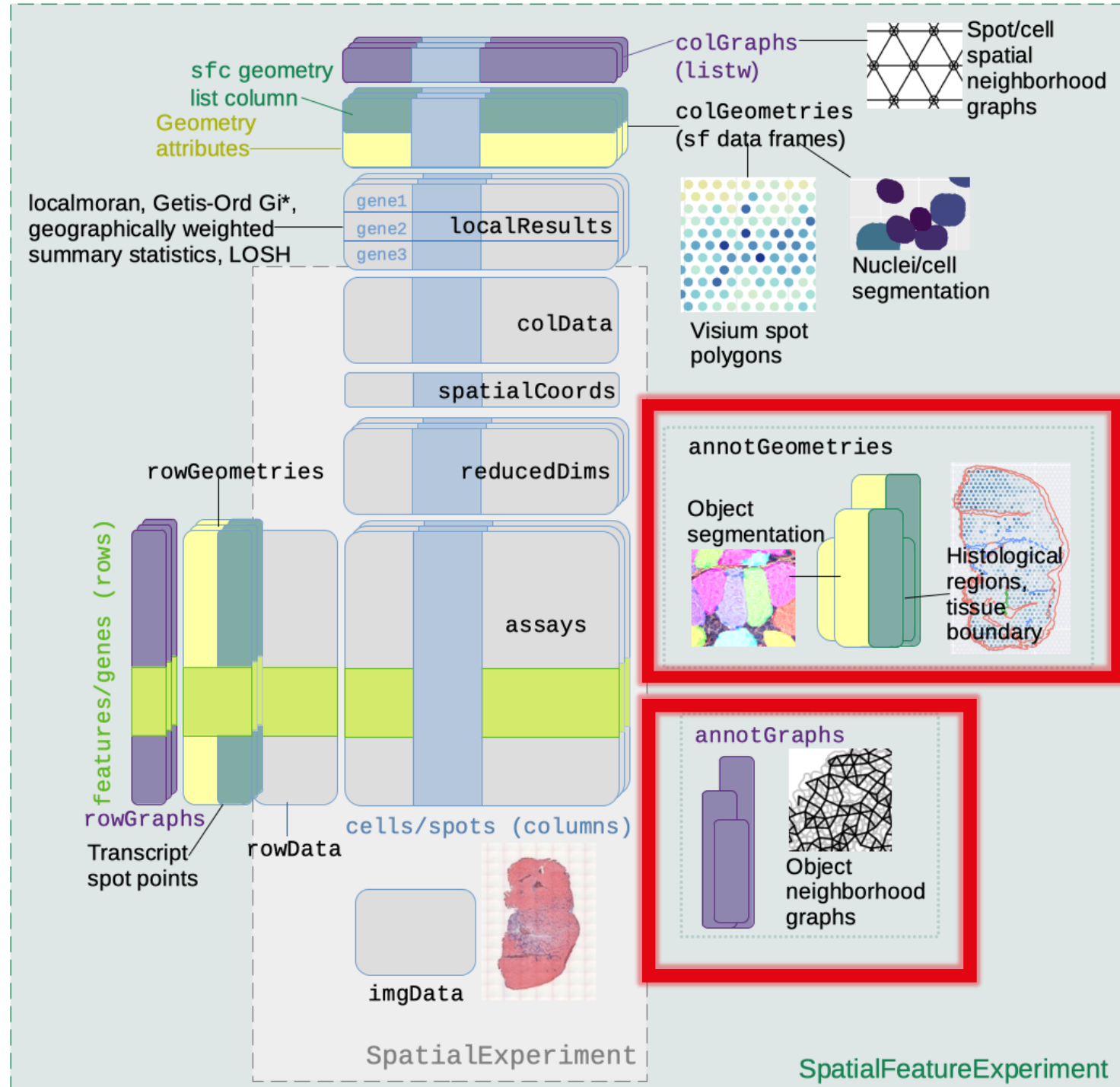
SpatialFeature Experiment

annotGeometries: tissue regions, histological boundaries

Shapes outlining specific anatomical regions, such as tumor cores, cortex layers, or artifacts.

annotGraphs: object neighborhood graphs

Spatial networks showing how different tissue regions or macro-structures border one another.



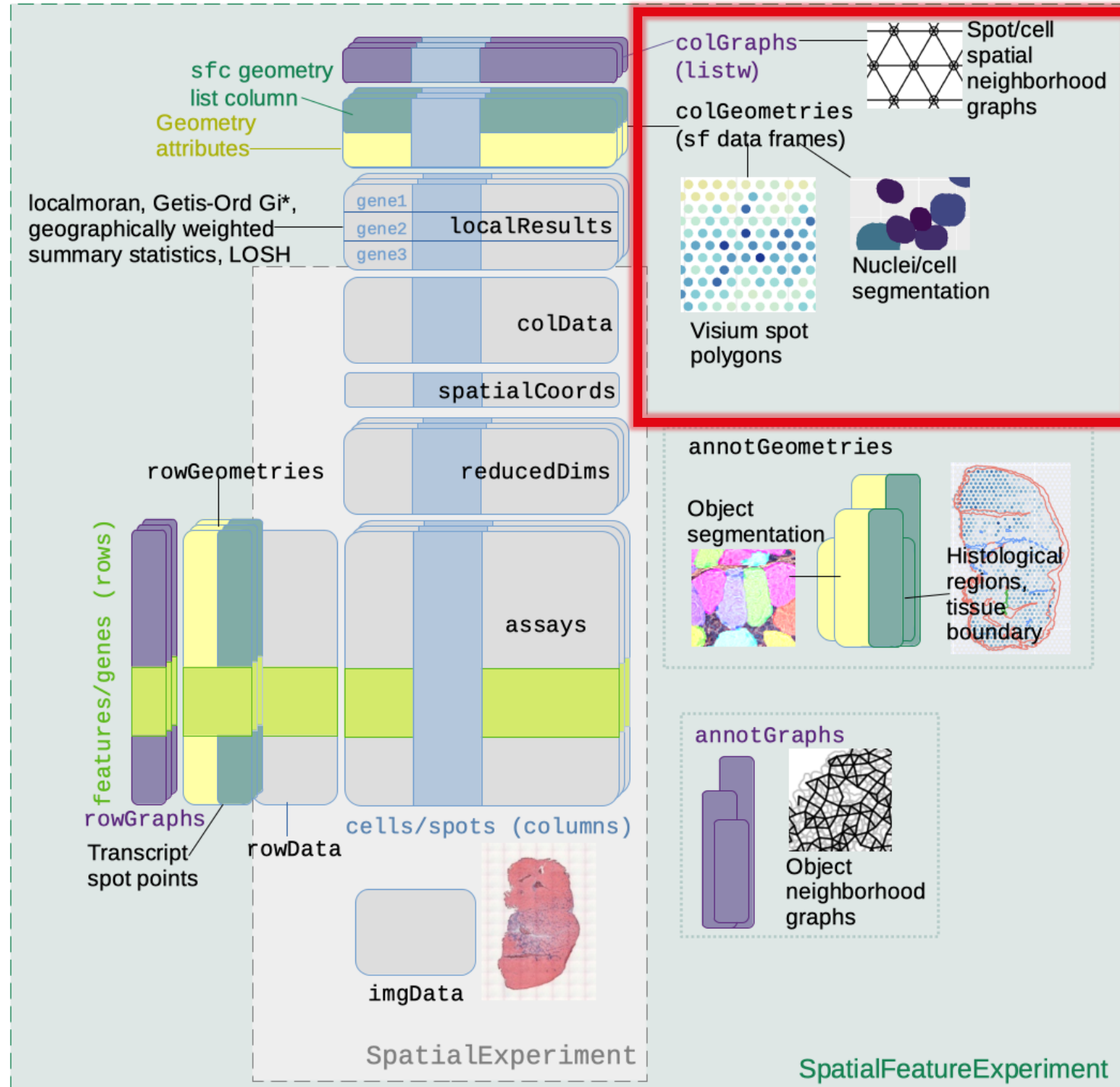
SpatialFeature Experiment

colGeometries: cell/nucleus polygon shapes

The precise boundaries and shapes of individual cells or nuclei.

colGraphs: spatial neighborhood graphs

Networks mapping which cells (or bins) are physical neighbors.



SCE → SPE → SFE

Feature
Count matrix (assays)
Cell/spot metadata (colData)
Gene metadata (rowData)
Reduced dimensions (reducedDims)
Spatial coordinates (spatialCoords)
Tissue image (imgData)
Cell/nucleus polygons (colGeometries)
Transcript locations (rowGeometries)
Spatial neighbor graphs (colGraphs)
Tissue annotations (annotGeometries)
Typical use case

SCE	SPE	SFE
✓	✓	✓
✓	✓	✓
✓	✓	✓
✓	✓	✓
—	✓	✓
—	✓	✓
—	—	✓
—	—	✓
—	—	✓
—	—	✓
scRNA-seq	Visium HD	Xenium

Which framework will you use for
single-cell data?

Which framework will you use for
single-cell data?

SingleCellExperiment

Which framework will you use for
Visium HD with segmented
output data?

Which framework will you use for
Visium HD with segmented
output data?

SpatialFeatureExperiment

Which framework will you use for
Visium HD without segmented
output data?

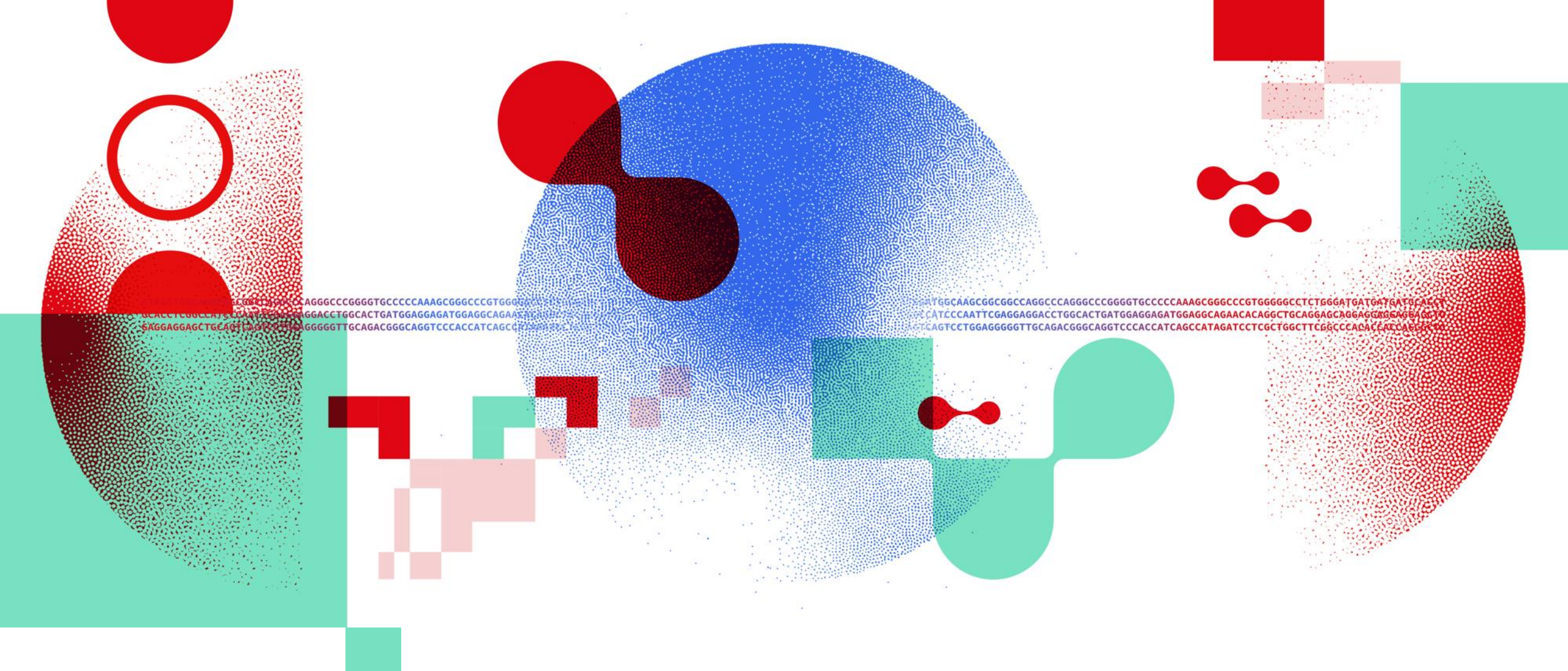
Which framework will you use for
Visium HD without segmented
output data?

SpatialExperiment

Which framework will you use for
Xenium data?

Which framework will you use for
Xenium data?

SpatialFeatureExperiment



Thank you

DATA SCIENTISTS FOR LIFE

sib.swiss

