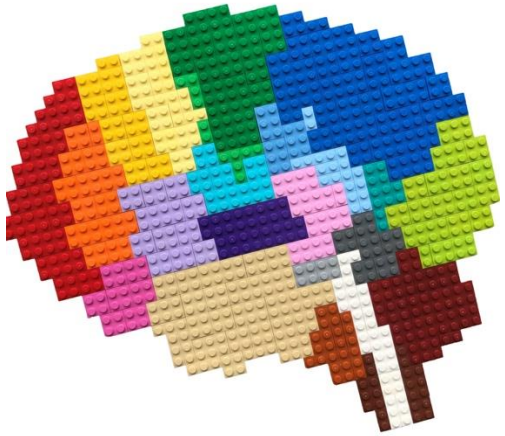
Github page for workshop:

https://github.com/margaretc-ho/BCBB_STx_workshop_2024



Spatial Transcriptomics Workshop Part 2: Hands On Tutorial with R and Seurat to analyze Visium dataset

Margaret Ho, PhD

Bioinformatics and Computational Biosciences Branch (BCBB) at NIAID

margaret.ho@nih.gov

Moderator: Colton McNinch, PhD (BCBB/NIAID)

Image credit to Bo Xia

Learning Objectives of Part 2

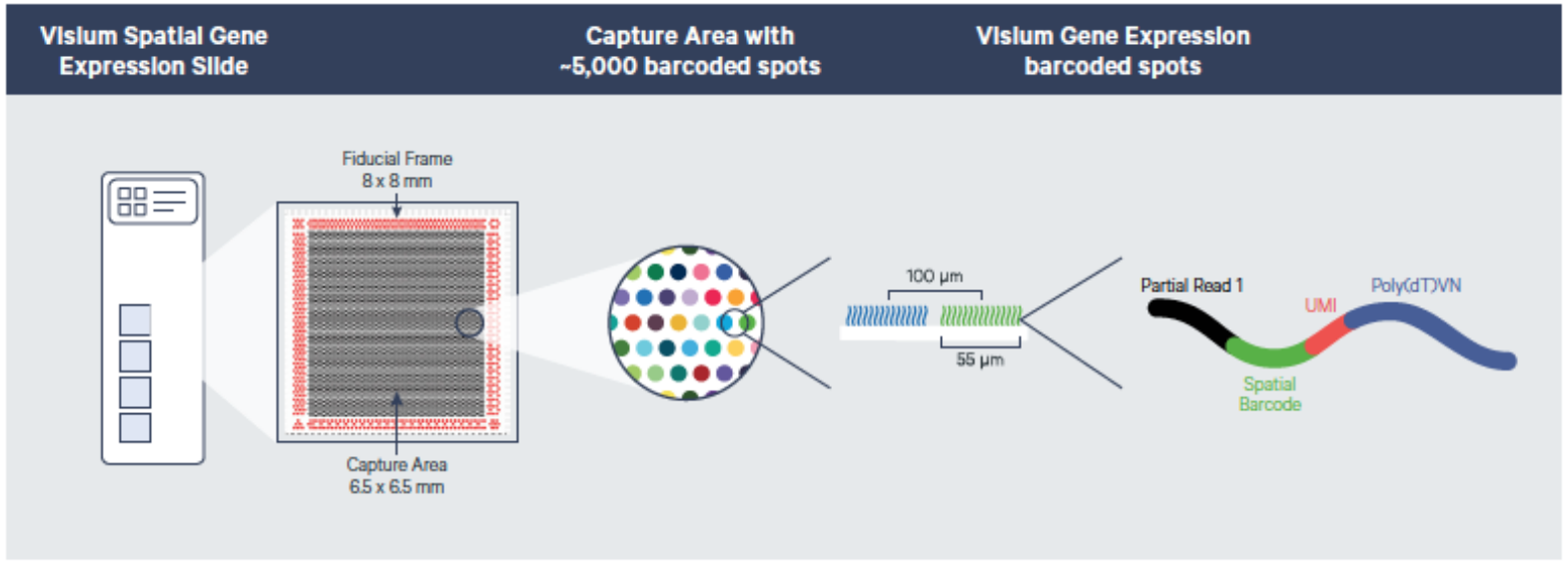
- Introduce techniques for processing and analyzing Visium STx data with R and Seurat
  - Understand hierarchical structure within Seurat object to store spatial dataset
  - Be familiar with count data overdispersion (Poisson, Negative Binomial)
  - Filtering data, and normalization using scTransform
  - Learn steps of unsupervised analysis: Dim Reduce, Clustering, Visualize with UMAP, DE Markers
  - Be able to plot counts, genes, and clusters for STx: VlnPlot, SpatialDimPlot, SpatialFeaturePlot, ImageDimPlot and ImageFeaturePlot
  - Identify Spatially Variable Genes with Moran's I
  - Annotate cell identities with a scRNA-seq reference dataset using RCTD

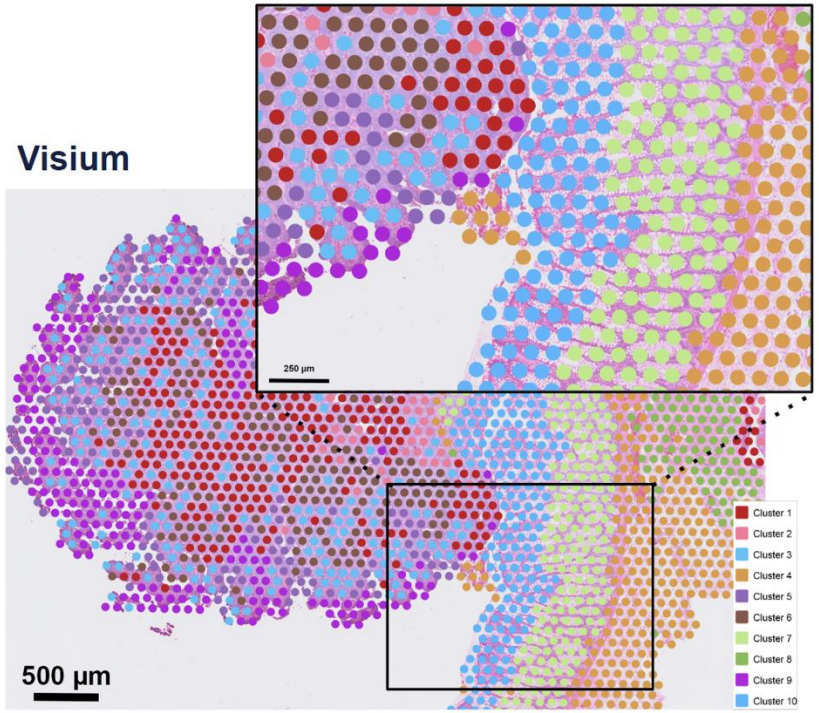
Coming up: 15 slides of intro // Seurat commands, then we'll get to Rstudio!

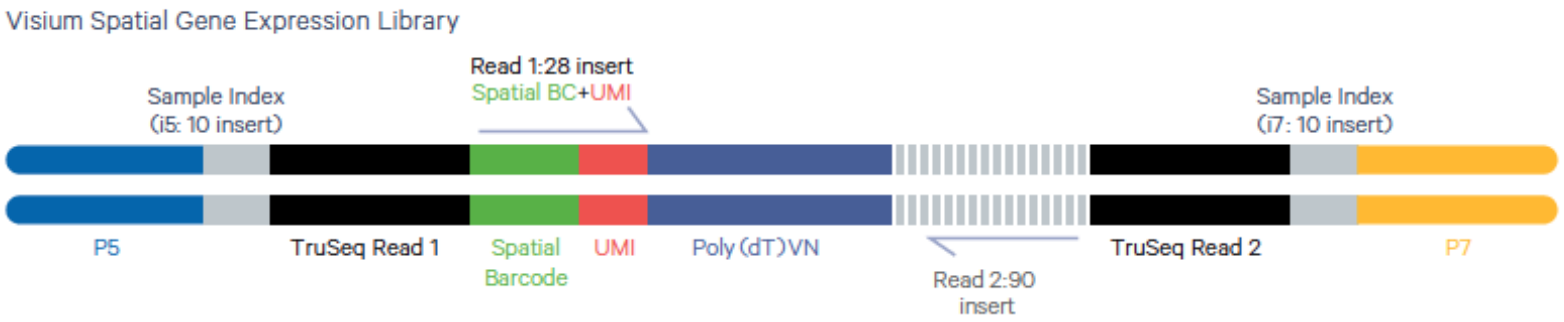
Github page for workshop
https://github.com/margaretc-ho/BCBB_STx_workshop_2024

Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
10X Visium	Whole Transcriptome	100um	(6.5 mm × 6.5 mm)x4	Few hours

10X Visium
Spot array-based spatial barcoding







4992 total spots per each of 4 capture areas
Sequence to at least 15k read pairs per spot

1-10 mammalian cells per spot
depending on tissue type

Fresh Frozen Only

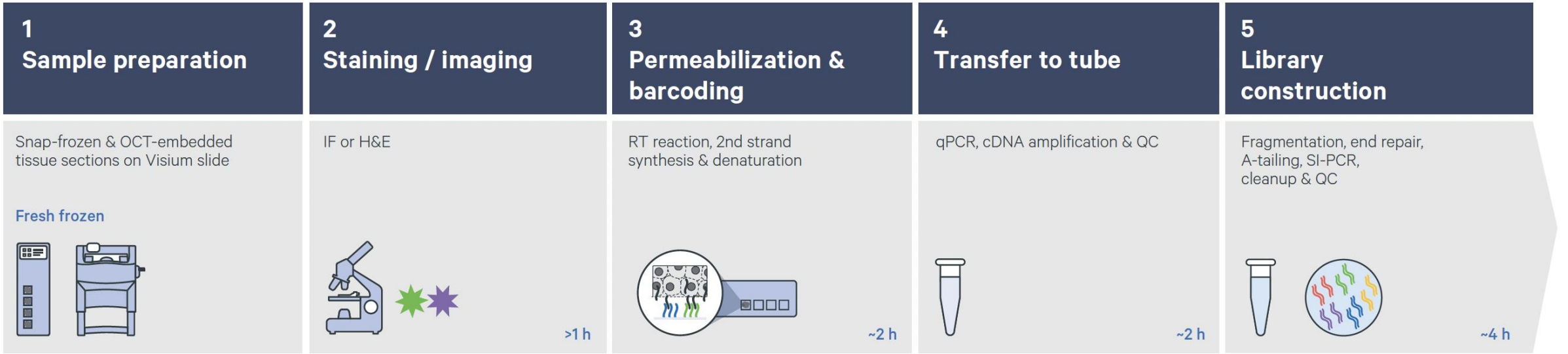
https://kb.10xgenomics.com/hc/en-us/categories/360002304851-Visium

Platform	# of genes profiled	Spatial Resolution	Imaging Area	Time Required
10X Visium	Whole Transcriptome	100um	(6.5 mm × 6.5 mm)x4	Few hours

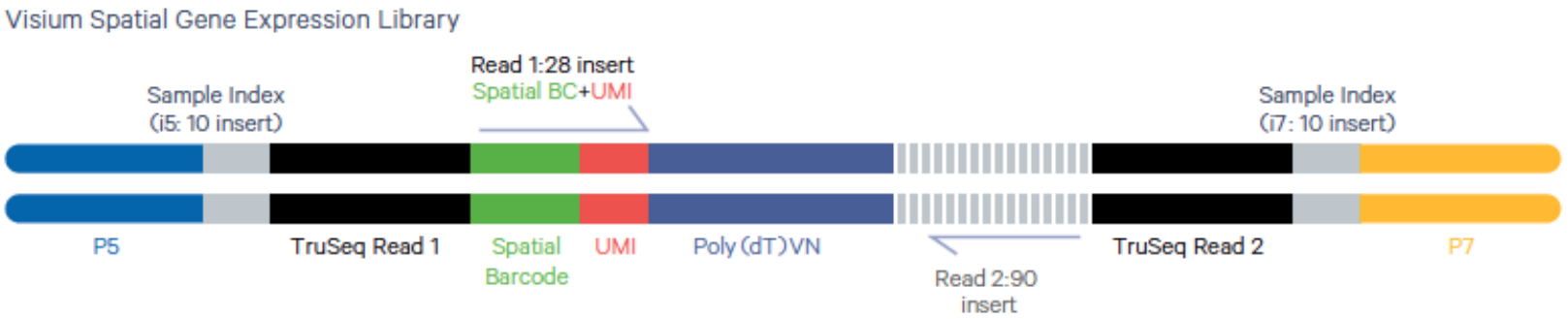
10X Visium

Spot array-based spatial barcoding

Fresh frozen



→ **Standard NGS**



Wang, Y., Liu, B., Zhao, G., Lee, Y., Buzdin, A., Mu, X., Zhao, J., Chen, H., & Li, X. (2023). Spatial transcriptomics: Technologies, applications and experimental considerations. *Genomics*, *115*(5), 110671.

https://kb.10xgenomics.com/hc/en-us/categories/360002304851-Visium

STxBrain Example Visium Dataset

- Fresh frozen sagittal mouse brain slice data generated using Visium v1 chemistry.
- There are two serial anterior sections, and two (matched) serial posterior sections – we will be analyzing anterior slice 1

Web summary: https://cf.10xgenomics.com/samples/spatial-exp/1.0.0/V1_Mouse_Brain_Sagittal_Posterior/V1_Mouse_Brain_Sagittal_Posterior_web_summary.html

3,353	Number of Spots Under Tissue
87,180	Mean Reads per Spot
4,564	Median Genes per Spot

Sequencing

Number of Reads	292,313,529
Valid Barcodes	97.6%
Valid UMIs	100.0%
Sequencing Saturation	73.8%
Q30 Bases in Barcode	97.2%
Q30 Bases in RNA Read	94.5%
Q30 Bases in UMI	97.2%

Mapping

Reads Mapped to Genome	94.0%
Reads Mapped Confidently to Genome	90.8%
Reads Mapped Confidently to Intergenic Regions	4.4%
Reads Mapped Confidently to Intronic Regions	1.6%
Reads Mapped Confidently to Exonic Regions	84.8%
Reads Mapped Confidently to Transcriptome	82.3%
Reads Mapped Antisense to Gene	0.8%

Mouse Brain Serial Section 1 (Sagittal-Posterior)

Spatial Gene Expression Dataset by Space Ranger 1.0.0

10x Genomics obtained fresh frozen mouse brain tissue (Strain C57BL/6) from BioIVT Asterand. The tissue was embedded and cryosectioned as described in Visium Spatial Protocols - Tissue Preparation Guide (Demonstrated Protocol CG000240). Tissue sections of 10 µm thickness from a sagittal slice of the posterior were placed on Visium Gene Expression Slides.

The H&E image acquired using a Nikon Ti2-E microscope with the following settings:

- Color camera
- 10X objective
- Numerical Aperture: 0.45
- Exposure: 10 ms
- Gain: 4.5X

Visium Spatial Gene Expression Library

The library (T1T2-F3) was prepared following the Visium Spatial Gene Expression Reagent Kits User Guide (CG000239) and was sequenced on an Illumina NovaSeq 6000.

- Sequencing Depth: 87,180 read pairs per spot
- Sequencing Configuration: 28 x 120 bp
- Sequencing Coverage: Read 1 - 28 bp (includes 16 bp Spatial Barcode, 12 bp UMI); Read 2 - 120 bp (transcript); i7 sample index - 10 bp; i5 sample index - 10 bp.
- Slide: V19L29-035
- Area: A1

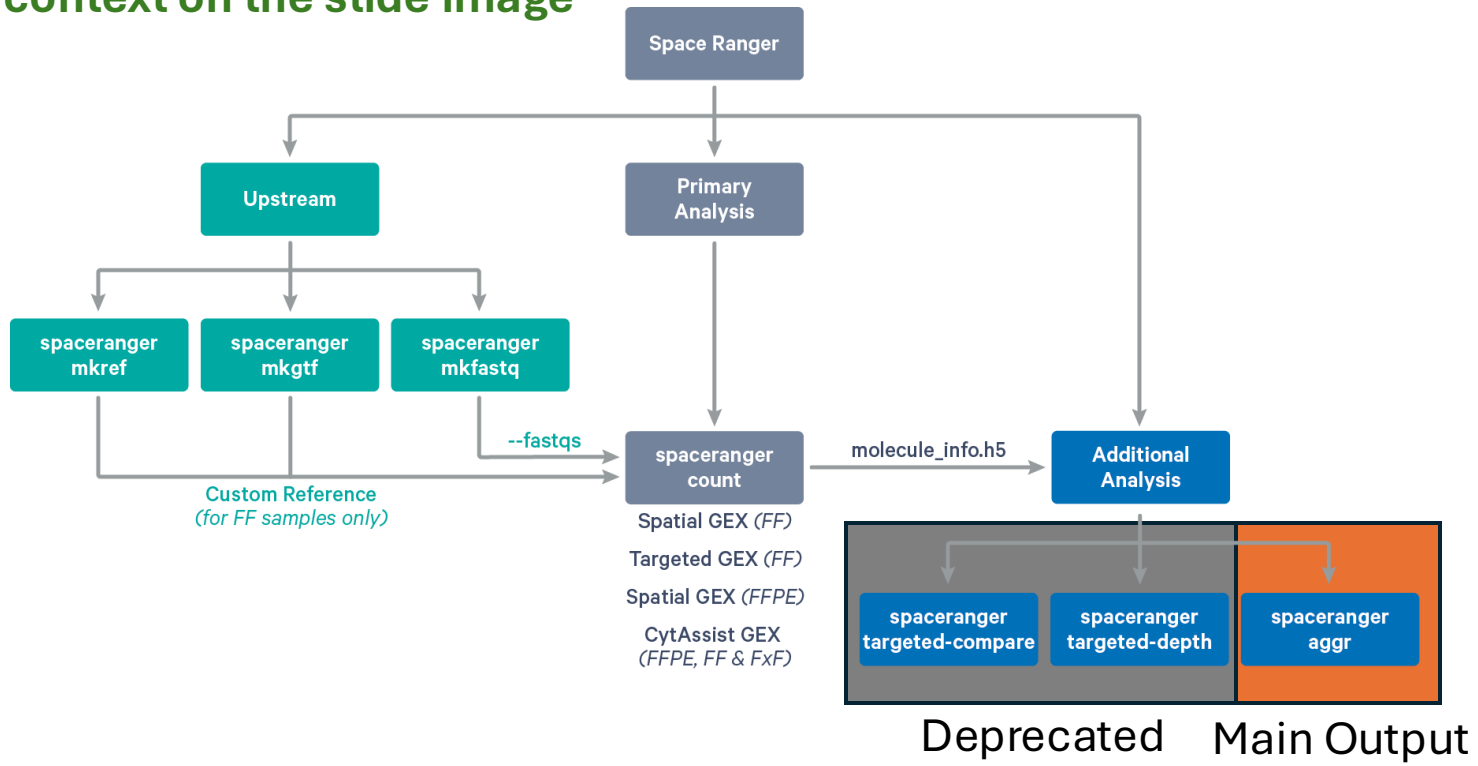
Key cell metrics were:

- Spots detected under tissue - 3,353
- Median UMI counts per spot - 14,864
- Median genes per spot - 4,564

https://support.10xgenomics.com/spatial-gene-expression/datasets/1.0.0/V1_Mouse_Brain_Sagittal_Posterior?

Space Ranger
for Visium/Visium HD data

Space Ranger detects tissue, aligns reads, **generate feature-spot matrices**, perform clustering and gene expression analysis, and **place spots in spatial context on the slide image**



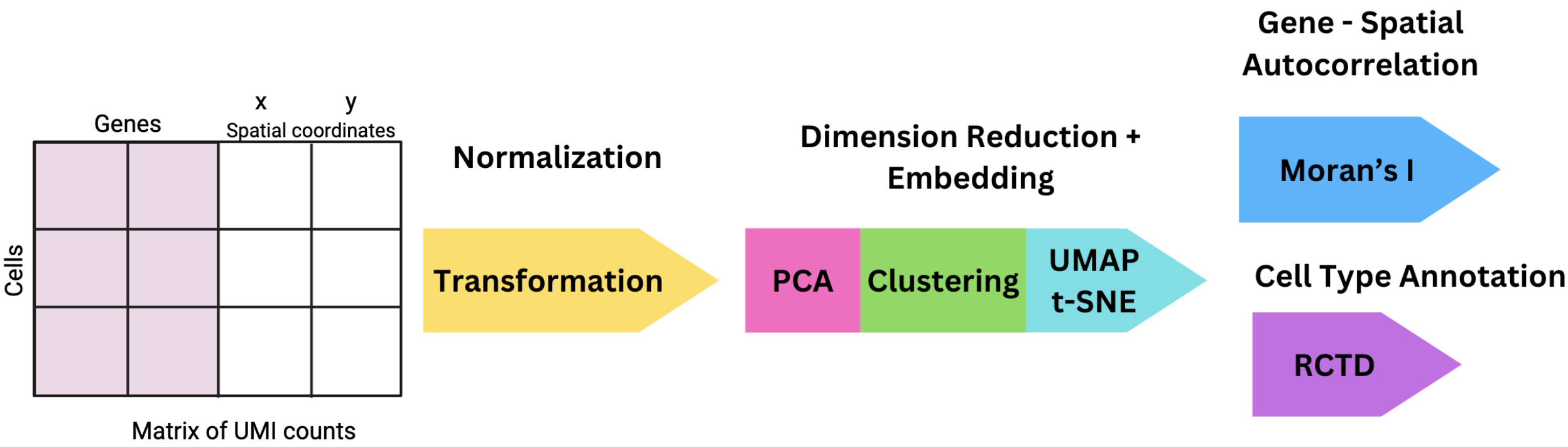
Spaceranger aggr output files:

```
outs
├── aggregation.csv
├── aggr_tissue_positions.csv
├── analysis
│   ├── clustering
│   ├── diffexp
│   ├── pca
│   ├── tsne
│   └── umap
├── cloupe.cloupe
├── filtered_feature_bc_matrix
│   ├── barcodes.tsv.gz
│   ├── features.tsv.gz
│   └── matrix.mtx.gz
├── filtered_feature_bc_matrix.h5
├── spatial
│   ├── LV123
│   │   ├── scalefactors_json.json
│   │   ├── tissue_hires_image.png
│   │   └── tissue_lowres_image.png
│   ├── LB456
│   │   ├── scalefactors_json.json
│   │   ├── tissue_hires_image.png
│   │   └── tissue_lowres_image.png
│   └── LP789
│       ├── scalefactors_json.json
│       ├── tissue_hires_image.png
│       └── tissue_lowres_image.png
├── summary.json
└── web_summary.html
```

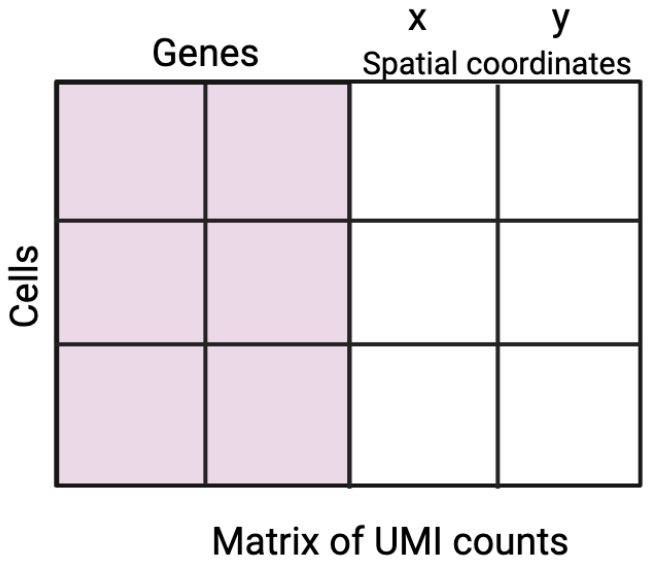
Adapted from https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/what-is-space-ranger

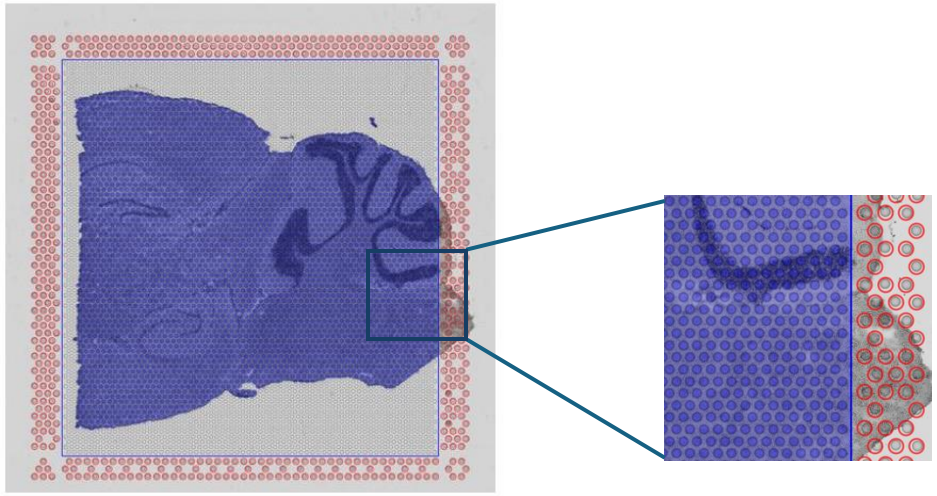
SpaceRanger on Biowulf: https://hpc.nih.gov/apps/spaceranger.html

R/Seurat Workflow for STx Analysis



Seurat Commands: Load Visium Data



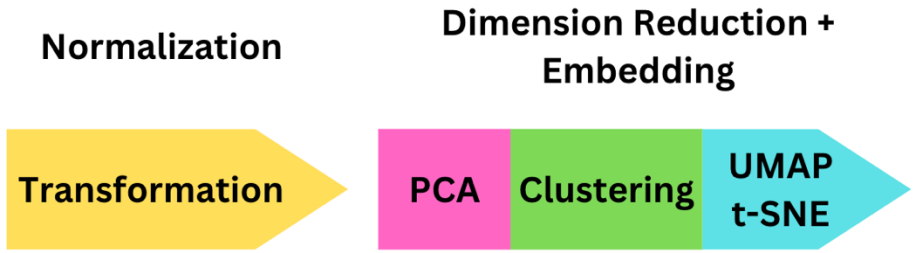


- **Load10x_Spatial**(dir="/data/dir/", filename = "**filtered_feature_bc_matrix.h5**", assay = "**Spatial**", slice = "slice1")

Note: **LoadVizgen** (for MERFISH) and **LoadXenium** (for Xenium)

https://satijalab.org/seurat/articles/essential_commands.html , use utils::methods(class = 'Seurat')

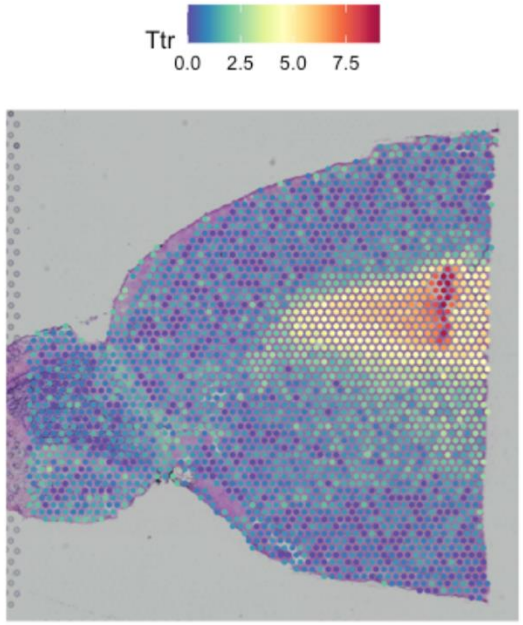
Seurat Commands: Normalize, Cluster, PCA, UMAP



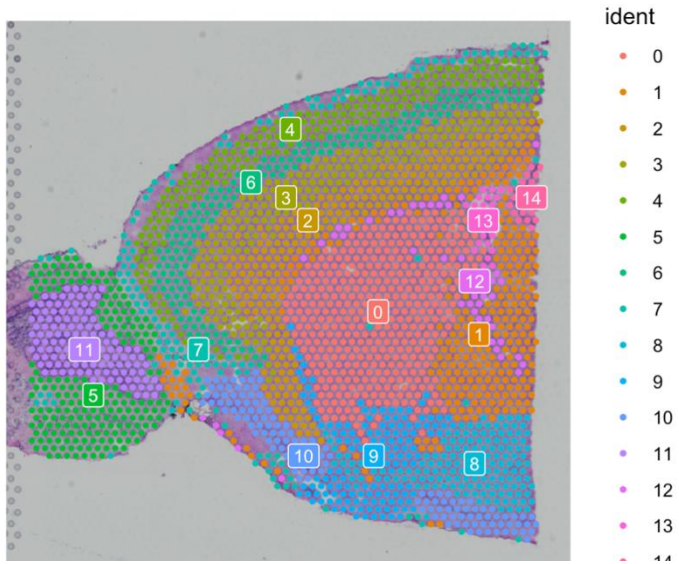
- **SCTransform** perform sctransform-based normalization
- **RunPCA**(object, assay = "SCT")
- **FindNeighbors**(object, reduction = "pca", dims = 1:30) – computes shared nearest neighbors (SNN) graph
- **FindClusters** – identify clusters of cells by SNN optimization
- **RunUMAP**(object, reduction = "pca", dims = 1:30)
- **FindMarkers** perform DE genes using Wilcoxon Rank Sum test

https://satijalab.org/seurat/articles/essential_commands.html , use utils::methods(class = 'Seurat')

Seurat Commands: Spatial Plotting



SpatialFeaturePlot

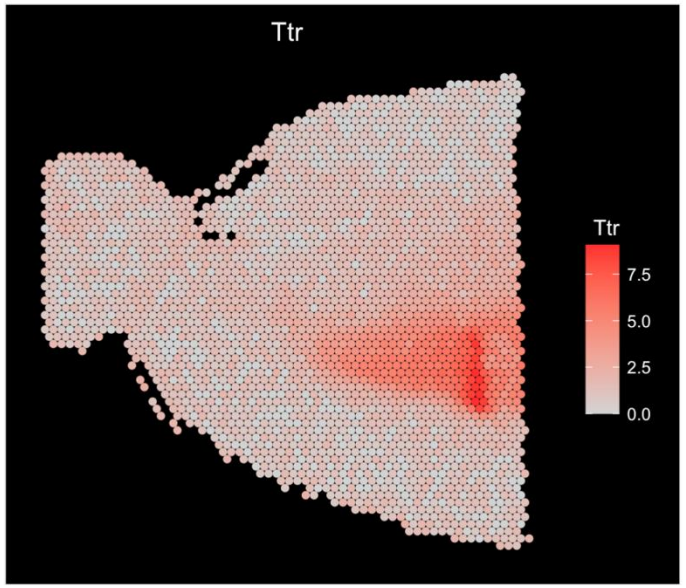


SpatialDimPlot

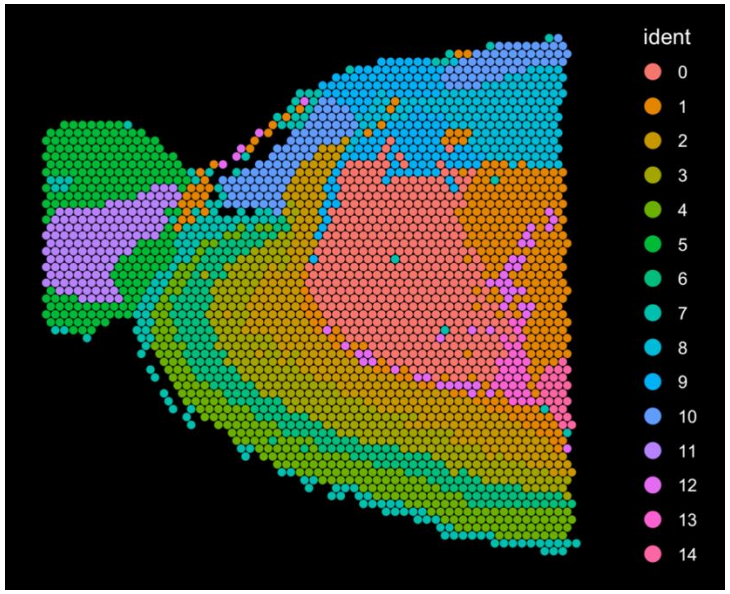
- **SpatialFeaturePlot** - plots molecular data over tissue histology image
- **SpatialDimPlot** - plots clusters or categorical groupings over tissue histology image

https://satijalab.org/seurat/articles/essential_commands.html , use utils::methods(class = 'Seurat')

Seurat Commands: Image Plotting



ImageFeaturePlot

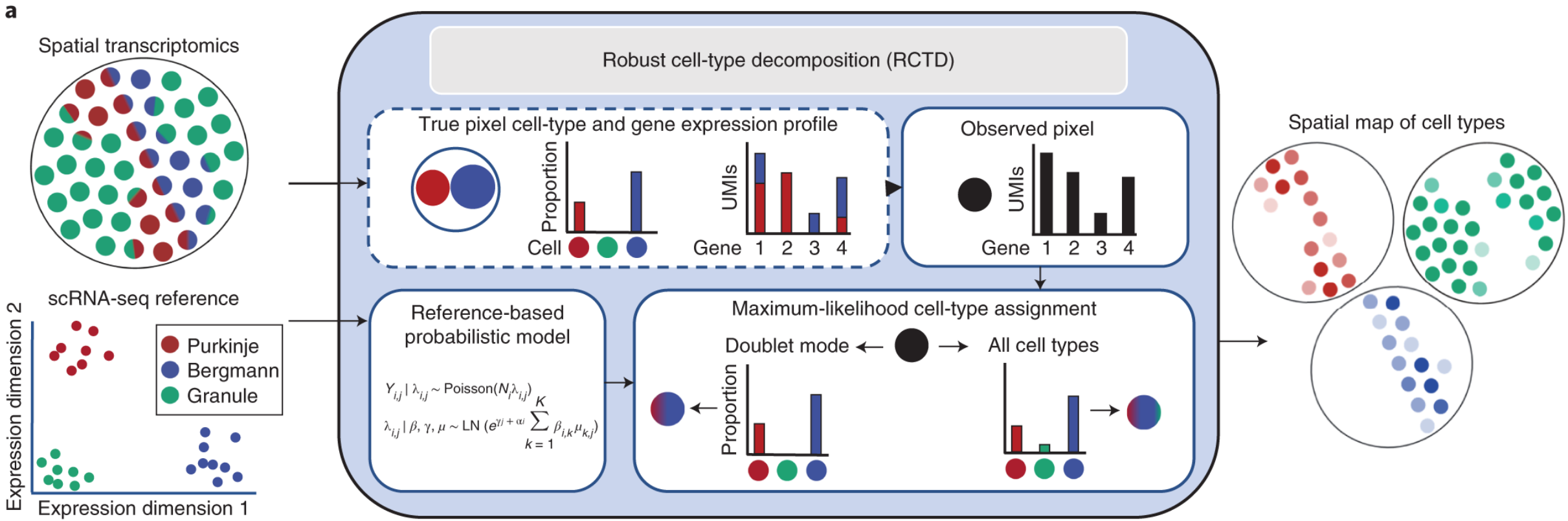


ImageDimPlot

- **ImageFeaturePlot** – plot gene/features on a spatial map
- **ImageDimPlot** – plot clusters or other categorical groupings on a spatial map

https://satijalab.org/seurat/articles/essential_commands.html , use utils::methods(class = 'Seurat')

Annotation with ref scRNA-seq data using RCTD



Cable, DM et al. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotech*, *40*(4), 517–526.

Li, B et al. (2022). Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods*, *19*(6), 662–670.
Li, Y et al. (2021). Benchmarking computational integration methods for spatial transcriptomics data. *bioRxiv* https://doi.org/10.1101/2021.08.27.457741

RCTD essential commands

```
Reference(
  counts,
  cell_types,
  nUMI = NULL,
  require_int = TRUE,
  n_max_cells = 10000,
  min_UMI = 100
)
```

Reference

```
SpatialRNA(
  coords,
  counts,
  nUMI = NULL,
  use_fake_coords = FALSE,
  require_int = TRUE
)
```

Query (spatial)

Create RCTD reference object from the scRNAseq reference counts, cluster annotations, and nUMI

reference <- **Reference**(ref_counts, ref_cluster, ref_nUMI)

Create the RCTD spatial query object

query <- **SpatialRNA**(query_coords, query_counts_hd, colSums(query_counts_hd))

Create RCTD object and run RCTD

RCTD_object <- **create.RCTD(**query_obj, reference_obj, max_cores = 28, CELL_MIN_INSTANCE=10)

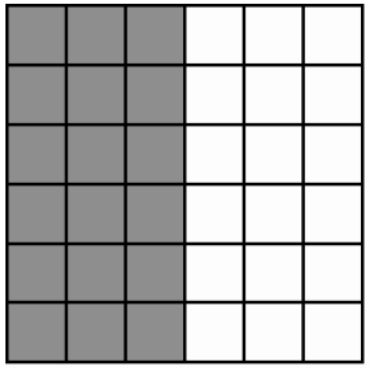
run.RCTD(RCTD_object, doublet_mode = "doublet")

```
create.RCTD(
  spatialRNA,
  reference,
  max_cores = 4,
  test_mode = FALSE,
  gene_cutoff = 0.000125,
  fc_cutoff = 0.5,
  gene_cutoff_reg = 2e-04,
  fc_cutoff_reg = 0.75,
  UMI_min = 100,
  UMI_max = 2e+07,
  counts_MIN = 10,
  UMI_min_sigma = 300,
  class_df = NULL,
  CELL_MIN_INSTANCE = 25,
  cell_type_names = NULL,
  MAX_MULTI_TYPES = 4,
  keep_reference = F,
  cell_type_profiles = NULL,
  CONFIDENCE_THRESHOLD = 5,
  DOUBLET_THRESHOLD = 20
)
```

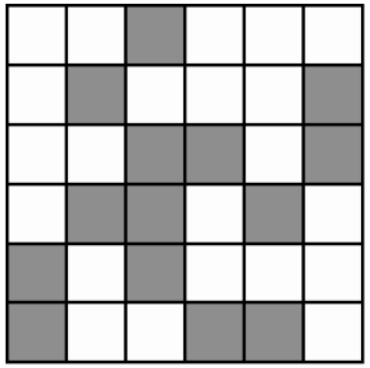
RCTD object

Identify spatially variable genes (SVGs) with Moran's I

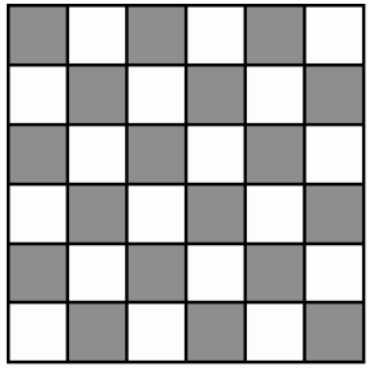
FindSpatiallyVariableFeatures to identify genes whose variability in expression can be explained to some degree by spatial location. Can use **markvariogram** or **moransi** as selection.method



Positive spatial autocorrelation



No spatial autocorrelation



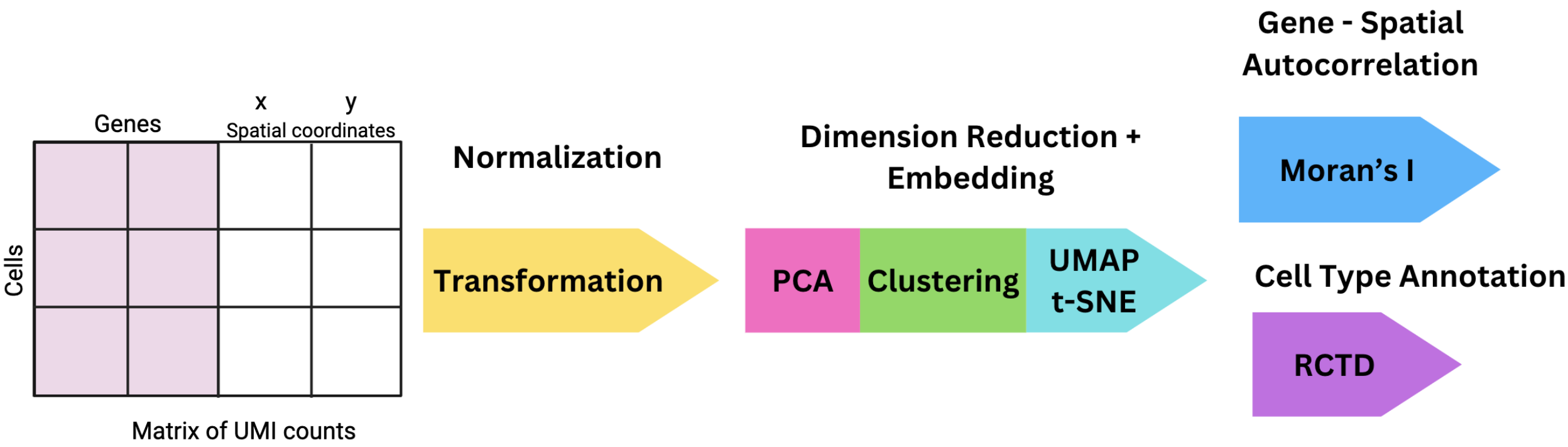
Negative spatial autocorrelation

```
FindSpatiallyVariableFeatures(
  object,
  spatial.location,
  selection.method = c("markvariogram", "moransi"),
  r.metric = 5,
  x.cuts = NULL,
  y.cuts = NULL,
  verbose = TRUE,
  ...
)
```

More info: Moran's I: https://www.youtube.com/watch?v=OJU8GNW9grc

From Radil, S. M. (2011). University of Illinois at Urbana-Champaign

R/Seurat Workflow for STx Analysis



Questions? :) Before going onto RStudio

More about BCBB:
https://www.niaid.nih.gov/research/bioinformatics-and-computational-biosciences-branch-scientific-services
Looking for bioinformatic and genomics analysis expertise? bioinformatics@niaid.nih.gov