SPATIAL TRANSCRIPTOMICS

Spatial Transcriptomics: From Tissue Architecture to Tumor Ecosystems

A practical introduction through an ovarian cancer case study

What you should be able to do after this seminar

01 **Explain what spatial transcriptomics measures—and what it does not**

02 Choose an analysis strategy that matches platform resolution

03 Interpret spatial clusters, deconvolution, subclones and ligand–receptor claims

04 Design a reproducible reanalysis starting from GEO

Methods and study: supplied paper and GEO record.

The central problem

Biology is organized in space, but conventional RNA-seq erases that organization.

Spatial transcriptomics preserves a coordinate system so expression can be interpreted as tissue architecture, neighborhoods and local interactions.

Three coordinate systems must be aligned

Molecular

Genes × capture units
Counts, UMIs or molecules
Technical depth and sparsity

What was measured?

Spatial

x–y positions
Spot or cell boundaries
Neighborhood graph

Where was it measured?

Histologic

H&E / IF image
Tissue compartments
Pathologist annotations

What tissue context?

Platforms trade breadth, resolution and sensitivity

Sequencing-based

Visium, Slide-seq, Stereo-seq
Broad transcriptome coverage
Spot/bin may mix cells

Discovery-oriented

Imaging-based

CosMx, MERFISH, Xenium
Single-cell or subcellular localization
Targeted gene panels

Localization-oriented

Hybrid designs

Discovery with whole transcriptome
Validation at single-cell resolution
Integrate with scRNA-seq

Best of both—at added cost

Background map | Spatial transcriptomics connects tissue to therapy

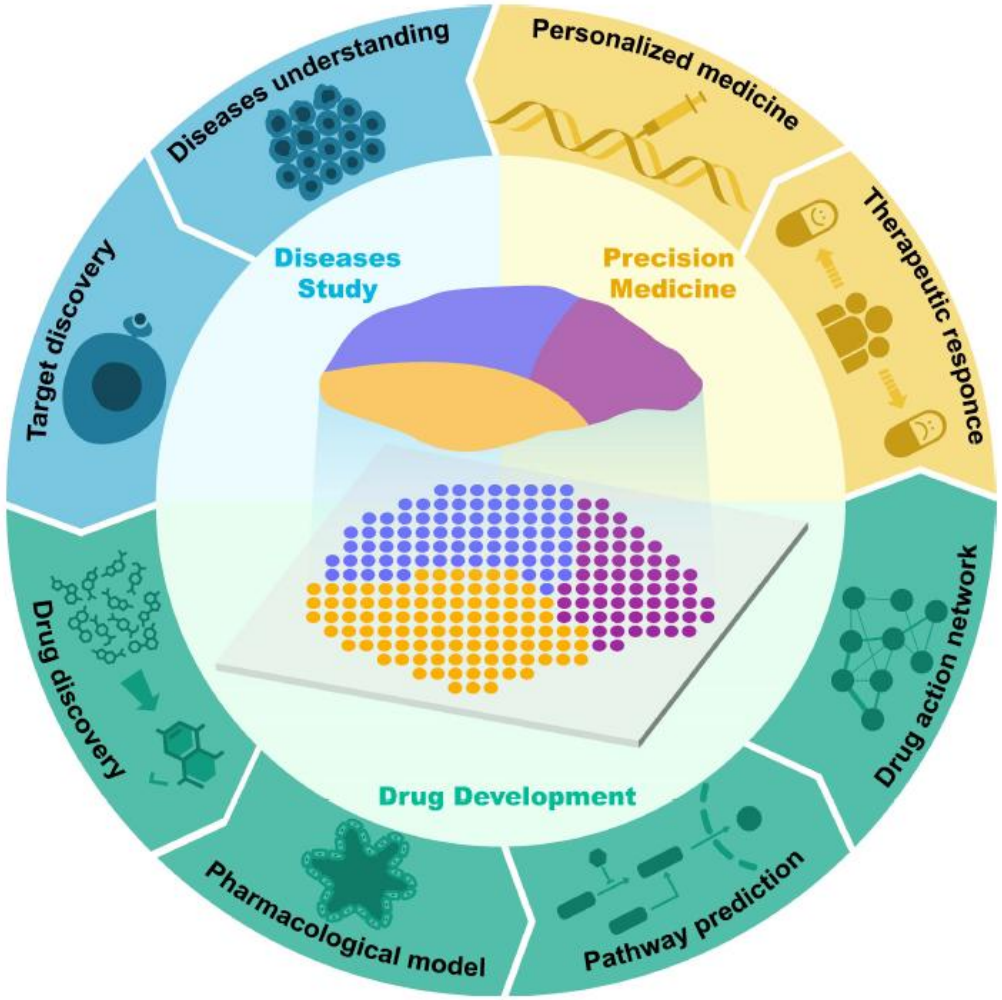


Figure 1. Application of spatial transcriptomics in drug research and development.

WHY IT MATTERS

The review frames spatial transcriptomics as a bridge from tissue organization to disease understanding, target discovery and drug development.

Use the diagram to connect technology, biology and analysis.

Technology map | In situ capture and image-based methods

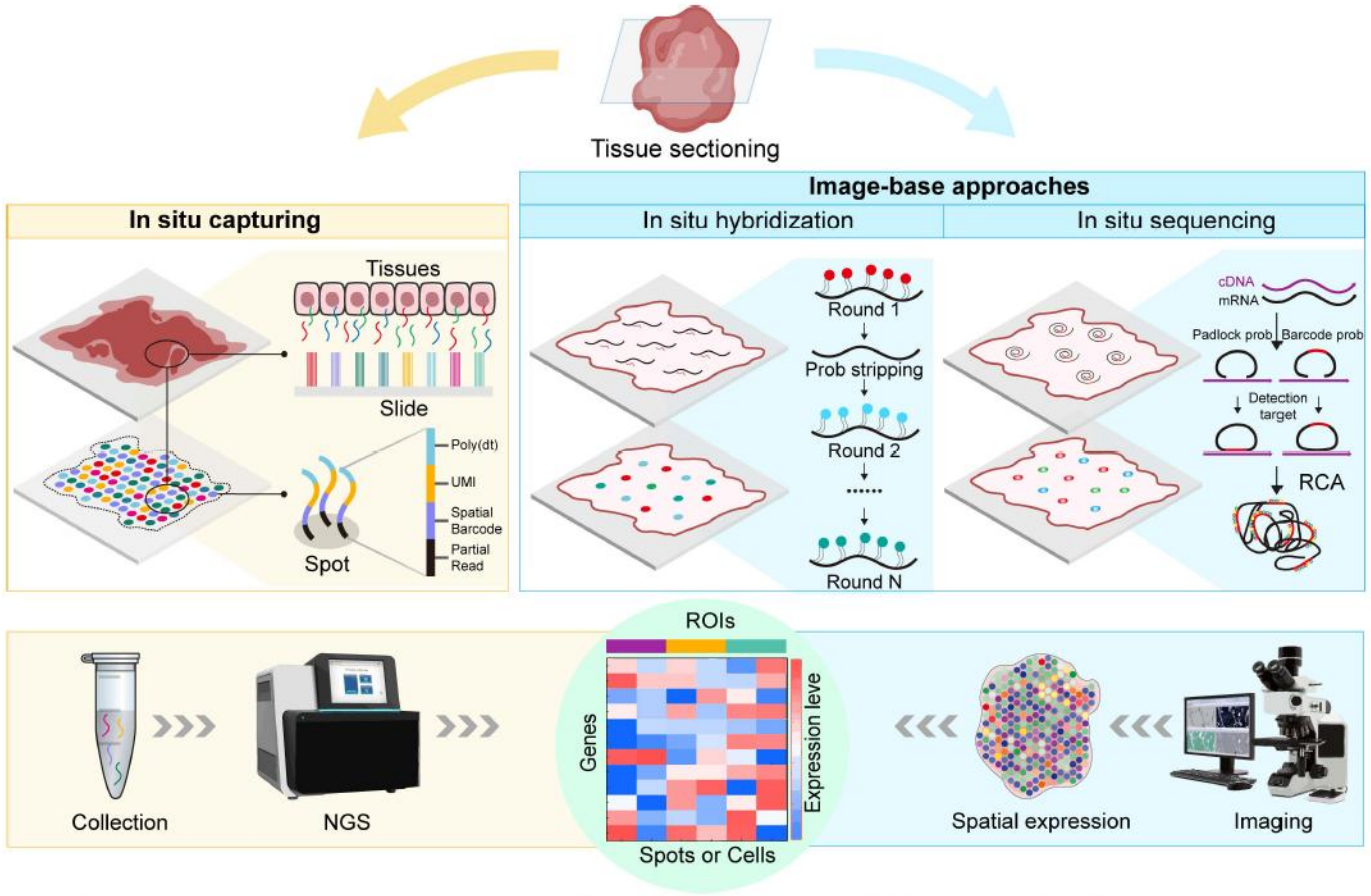


Figure 2. Current mainstream spatial transcriptomics methods. ISC involves the specific binding of mRNA from corresponding cellular locations with spots or beads on slides carrying capture probes, followed by further NGS of the captured RNA molecules. ISC offers high sensitivity and specificity, enabling the detection of rare transcripts and precise spatial localization, while achieving single-cell resolution with this technology is currently challenging. Image-based methods, including ISH and ISS. ISH involves the specific hybridization of RNA in tissues with multiple fluorescently labelled short oligonucleotide probes, reflecting the abundance and spatial localization of specific gene transcripts. ISH detects low-abundance transcripts with high sensitivity and relatively high resolution, while the limited signal-to-noise ratio during imaging necessitates high-magnification imaging systems, limiting the area of detection. ISS involves the hybridization of padlock probes with cDNA, followed by RCA to form RCP. ISS allows for the sequencing of RNA molecules directly on tissue sections, providing higher-resolution data, while it has lower sensitivity, and padlock probes may introduce significant probe-specific biases, increasing the complexity of image processing.

WHY IT MATTERS

The key split is how spatial identity is recorded: capture barcodes followed by sequencing, or optical detection in the tissue.

Use the diagram to connect technology, biology and analysis.

Background map | From a tissue section to a target

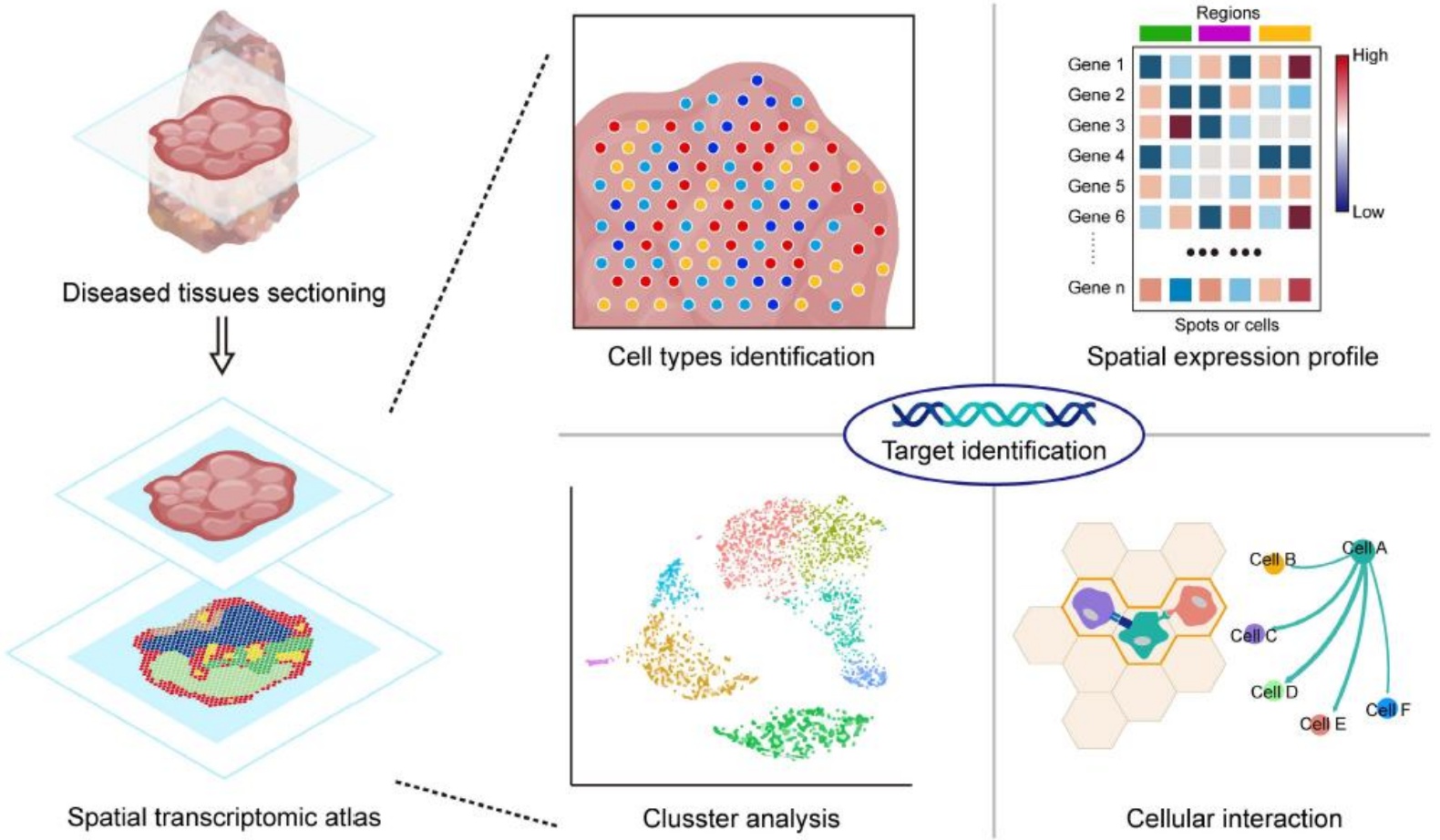


Figure 3. Application of spatial transcriptomics in target identification. By identifying cell types through spatial ST, it is possible to reveal the distribution of specific cell types within tissues and understand their function and state. The spatial expression profile depicts the expression of mRNAs in spatial dimensions, aiding in the identification of characteristic DEGs involved in disease processes. Clustering analysis helps in discovering specific cell populations associated with diseases that may be closely linked to the occurrence, progression, or prognosis of the disease. By analysing interactions between cells, key interactions involved in cell communication and signal transduction can be identified. Integrating this information can provide insights into the process of disease occurrence and reveal potential therapeutic targets.

WHY IT MATTERS

A spatial analysis links cell-type identification, spatial expression, clustering and cellular interaction to target identification.

Use the diagram to connect technology, biology and analysis.

Future map | From drug discovery to precision medicine

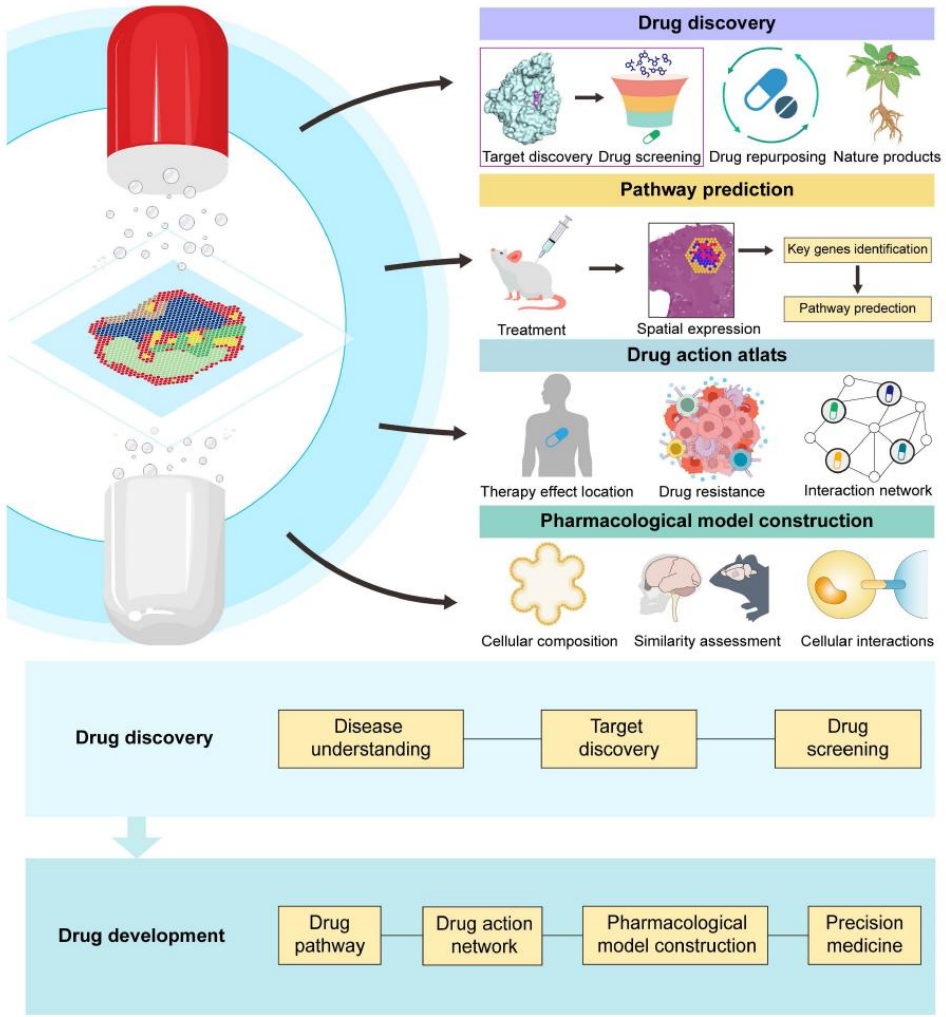


Figure 4. Application of spatial transcriptomics in drug discovery and drug development. ST aids in identifying new targets, facilitating drug screening. ST holds immense potential in drug repurposing and natural product drug discovery. ST utilizes spatial expression maps after drug treatment to identify key genes and predict treatment pathways. ST constructs a drug action atlas, enabling the localization of drug effects and the study of drug resistance and interactions. By analysing the spatial cell composition of pharmacological models, assessing model similarity to cases, and studying intercellular interactions, ST assists in building pharmacological research models. ST is being applied across various stages of drug discovery and development, and these applications are enhancing the probability of drug discovery success and improving therapeutic outcomes.

WHY IT MATTERS

Spatial maps can connect target discovery, treatment response, drug resistance, interaction networks and precision medicine.

Use the diagram to connect technology, biology and analysis.

A practical technology map: ask four questions

Whole transcriptome?

Discovery platforms capture many genes without a predefined panel.

Single-cell resolution?

Imaging or very small bins resolve cells, but may trade sensitivity or breadth.

Fresh or FFPE?

Sample preservation and chemistry constrain what can be measured.

Scale and cost?

Large tissue areas, depth and imaging time affect study design.

Seurat v5 supports both sequencing- and imaging-based spatial data

Sequencing-based

Load a spatial assay with image and coordinates
Normalize, cluster and map features
Use anchor/RCTD-style integration

Imaging-based

Represent cells, transcripts and segmentation
Use ImageFeaturePlot and molecule-level views
Analyze spatial niches

Shared grammar

Assay → metadata → reductions → spatial image
The object keeps biology and geometry together

Seurat workflow: from Space Ranger output to a spatial object

01 **Load10X_Spatial() reads the matrix and image-linked coordinates**

02 Filter spots/cells using spatially mapped QC metrics

03 Normalize with SCTransform or log normalization

04 Find variable features, PCA, neighbors and clusters

05 Visualize with SpatialFeaturePlot and SpatialDimPlot

Seurat sequencing vignette: https://satijalab.org/seurat/articles/spatial_vignette.html

Seurat QC: quality metrics must be read in tissue context

Core metrics

nCount_Spatial: molecular depth per spot
nFeature_Spatial: detected genes per spot
percent.mt: mitochondrial fraction, if gene symbols support it

Start broad

Spatial check

Use VlnPlot for distribution
Use SpatialFeaturePlot to see whether depth follows anatomy, tissue edges or damage

Then map it

Filtering logic

Avoid one universal cutoff
Record excluded spots
Check whether tumor/stroma regions are selectively removed

Document choices

Seurat normalization: choose the model before interpreting clusters

Why normalize?

Spatial spots differ in sequencing depth and cell density.
Normalization reduces depth-driven structure before PCA.

Official example

The Seurat vignette proceeds with SCTransform on assay = "Spatial" after visualizing nCount_Spatial.

Practical caution

SCTransform and log normalization can change variable features, PCs and clusters.
Report the choice and test sensitivity.

Seurat dimensional reduction and clustering

1. **PCA** — RunPCA on the normalized assay to summarize major expression axes.
2. **Neighbors** — FindNeighbors builds a graph from selected PCs, commonly dims 1:30 in the vignette.
3. **Clusters** — FindClusters identifies graph communities; resolution changes granularity.
4. **UMAP** — RunUMAP gives a 2D transcriptional embedding, not physical tissue coordinates.

UMAP and tissue maps answer different questions

DimPlot / UMAP

Shows transcriptional similarity.
Useful for cluster separation, gradients and outliers.
No physical distance interpretation.

SpatialDimPlot

Overlays clusters or labels on tissue.
Useful for anatomical restriction, boundaries and regional mixing.

Read both together

A clean UMAP cluster may be spatially scattered.
A tissue domain may contain multiple expression states.

A Seurat code block for Visium-style spatial RNA-seq

Load + QC

```
obj <- Load10X_Spatial("outs")
obj$percent.mt <-
PercentageFeatureSet(obj, pattern =
"^MT-")
VlnPlot(obj, features =
c("nCount_Spatial",
"nFeature_Spatial"))
```

Map QC

```
SpatialFeaturePlot(obj, features =
"nCount_Spatial")
# inspect anatomy, edges, tissue
damage
# choose sample-aware filters
```

Model

```
obj <- SCTransform(obj, assay =
"Spatial", verbose = FALSE)
obj <- RunPCA(obj, assay = "SCT",
verbose = FALSE)
obj <- FindNeighbors(obj, reduction =
"pca", dims = 1:30)
obj <- FindClusters(obj)
obj <- RunUMAP(obj, reduction =
"pca", dims = 1:30)
```

Seurat visualizations answer different spatial questions

SpatialFeaturePlot

Where is a gene or program expressed?
Use continuous scales and inspect tissue context.

SpatialDimPlot

Where are clusters or labels?
Compare boundaries with H&E.

ImageDimPlot / molecule view

Which cells or transcripts are visible?
Most useful for imaging-based platforms.

Seurat + scRNA-seq: label transfer is a model, not a stain

1	**Reference**	Create a quality-controlled scRNA-seq reference with cell-type labels.
2	**Anchors**	Find shared features and anchors between reference and spatial query.
3	**Transfer**	Map labels or prediction scores to spots/cells.
4	**Check**	Plot transferred labels, marker genes and uncertainty in tissue.

RCTD and label transfer solve related but different problems

Label transfer

Assigns a likely label or score to each spot/cell.
Good for annotation and visualization.

Deconvolution

Estimates proportions of multiple cell types in a mixed spot.
Good for composition-aware modeling.

Niche analysis

Summarizes spatially adjacent cell-type composition.
Good for tissue ecology.

Applying the Seurat grammar to GSE211956

01 **Create one Seurat object per Visium section; keep patient and section IDs**

02 Plot nCount_Spatial, nFeature_Spatial and tissue boundaries before filtering

03 Cluster each section first; integrate only when the scientific question requires it

04 Use the scRNA reference for label transfer or RCTD, then compare with H&E

05 Export a patient-aware table of clusters, cell-type weights and spatial programs

GEO GSE211956: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211956; Seurat spatial vignette: https://satijalab.org/seurat/articles/spatial_vignette.html

The computational bottleneck is no longer only clustering

Spatial analysis must model geometry, uncertainty and scale.

As resolution increases, datasets become larger—and naïve spot-by-spot inference becomes less defensible.

Resolution is not only a pixel size

A spot is a measurement unit—not necessarily a cell.

For Visium, one spot commonly aggregates transcripts from multiple cells; cell-type proportions and expression programs are therefore entangled.

A sequencing-based experiment has five linked stages

1	**Prepare**	Preserve morphology and RNA quality in a tissue section.
2	**Capture**	Hybridize transcripts to spatially barcoded probes.
3	**Sequence**	Generate reads carrying gene identity, UMI and spatial barcode.
4	**Reconstruct**	Align reads, image and capture positions into a spatial object.
5	**Model**	Quantify composition, states, neighborhoods and interactions.

The spatial object contains four inseparable layers

Counts

Raw gene-by-spot matrix
Sparse integer counts
Required for principled modeling

Coordinates

Array row/column
Pixel x/y
In-tissue flag

Image

High/low-resolution H&E
Scale factors
Tissue boundaries

Metadata

Patient, section, batch
Pathology, treatment
QC and annotations

Quality control begins with tissue-aware patterns

01 **Confirm tissue detection and image alignment**

02 Inspect UMIs, detected genes and mitochondrial fraction across space

03 Identify edge effects, folds, tears and low-RNA regions

04 Compare distributions by patient and section—not only pooled

Typical QC metrics; supplied paper Results.

Normalization answers a technical question—not a biological one

Library-size scaling

Corrects different sequencing depth
Useful for visualization
Can distort composition-heavy tissue

Variance stabilization

Models mean–variance relation
Useful for PCA and clustering
Must respect counts and batch

Model-based analysis

Use raw counts with offsets/covariates
Preferred for inference
Aggregate or model patients

Spatial clustering is a hypothesis generator

01 **Expression-only clustering finds transcriptional similarity**

02 Spatially aware clustering adds neighborhood continuity

03 Histology should be compared—not forced to agree

04 Cluster stability matters more than a single resolution

Supplied paper Fig. 1 and Results.

Annotation separates composition from state

Marker scoring

Fast and interpretable
Confounded by mixed spots
Sensitive to marker specificity

Deconvolution

Estimates cell-type weights
Requires a reference atlas
Outputs are model-dependent

Histology

Independent tissue evidence
Can reveal tumor, stroma, necrosis
Resolution differs from RNA

Deconvolution inherits every bias in its reference

01 **Cell types absent from the reference cannot be recovered**

02 Closely related states may be exchanged or merged

03 Tumor programs can resemble normal epithelial programs

04 Weights are estimates—not direct cell counts

Supplied paper Supplementary analysis and RCTD reference.

Spatially variable genes require a spatial null model

Global SVG

Any non-random spatial pattern
May reflect morphology or depth

Domain markers

Differential expression among regions
Depends on domain definition

Local autocorrelation

Moran's I / Geary's C
Tests neighbor similarity
Requires multiple-testing control

ANALYSIS WORKFLOW

Neighborhood analysis turns coordinates into tissue ecology

1. **Build a graph** — Define neighbors by physical distance, kNN or tissue topology.
2. **Summarize context** — Count cell types, states or programs around each unit.
3. **Test enrichment** — Compare observed contacts with label-permuted or patient-aware nulls.
4. **Link to state** — Ask whether neighborhood predicts expression after composition adjustment.

Integrating scRNA-seq provides labels, not ground truth

Reference mapping

Transfer cell-type labels or signatures
Check dataset and treatment match

Joint embedding

Align shared variation
May remove real spatial differences

Deconvolution

Estimate mixture proportions
Validate with orthogonal markers

Copy-number patterns can reveal tumor subclones

01 **Infer broad CNAs from coordinated expression along chromosomes**

02 Use non-malignant cells as a reference when possible

03 Cluster CNA profiles to nominate subclones

04 Validate with DNA-based data or orthogonal spatial evidence

Supplied paper inferCNV workflow.

The most common statistical error

Thousands of spots do not equal thousands of independent patients.

Use patient/section as the replication unit; aggregate counts, fit mixed models or test effects consistently across patients.

A reproducible analysis records the entire coordinate chain

01 **Raw counts, image, positions and scale factors**

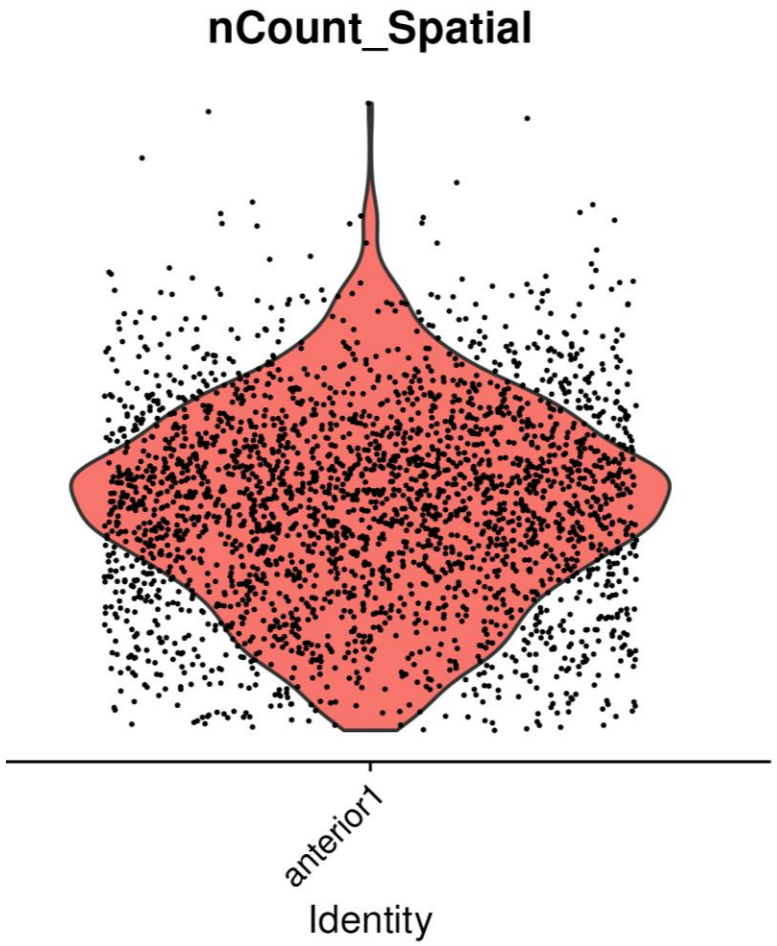
02 Software versions and parameter choices

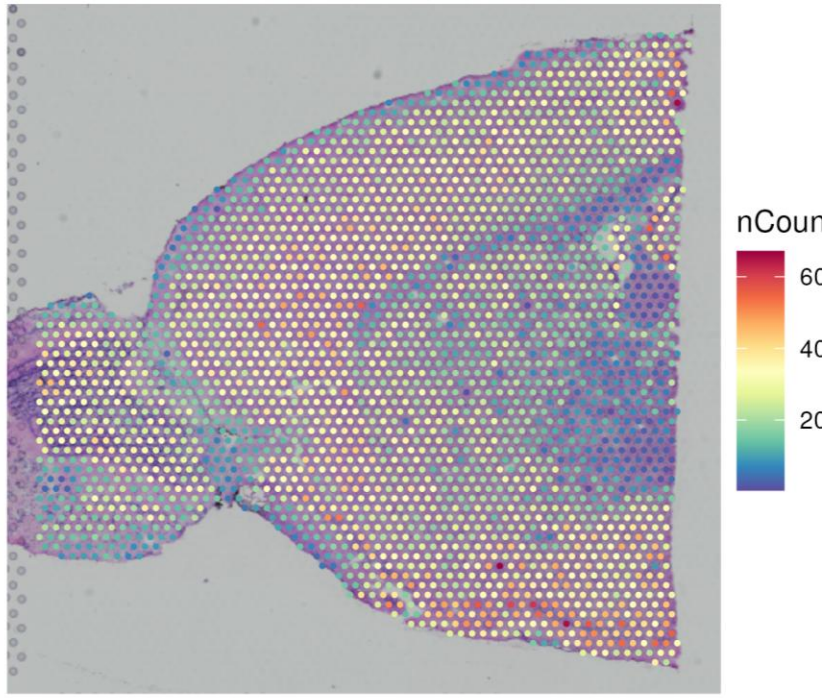
03 Sample-level metadata and exclusion rules

04 Seeds, reference atlases and gene identifiers

Reproducible research principles.

Seurat vignette figure | QC must be mapped back to tissue





WHY IT MATTERS

The official Seurat vignette pairs VlnPlot with SpatialFeaturePlot for nCount_Spatial, showing why spot depth must be interpreted with anatomy.

Use the diagram to connect technology, biology and analysis.

Seurat principle 1: the assay is molecular, the image is contextual

1. **Assay** — The Spatial assay stores spot-level expression values.
2. **Image** — The image slot stores tissue pictures and coordinate mapping.
3. **Metadata** — Sample, QC and annotation fields travel with each spot.
4. **Reduction** — PCA and UMAP store expression-derived embeddings.

Seurat principle 2: QC is a biological decision under technical constraints

Do not filter blindly

Low counts may reflect tissue edge, necrosis, low cell density or failed capture.

Use paired views

Inspect metric distributions and spatial overlays before setting thresholds.

Keep a decision log

Record cutoffs, excluded spot counts and whether exclusion is region-specific.

Seurat principle 3: normalization defines the expression scale

Depth correction

NormalizeData or SCTransform reduce sequencing-depth effects.

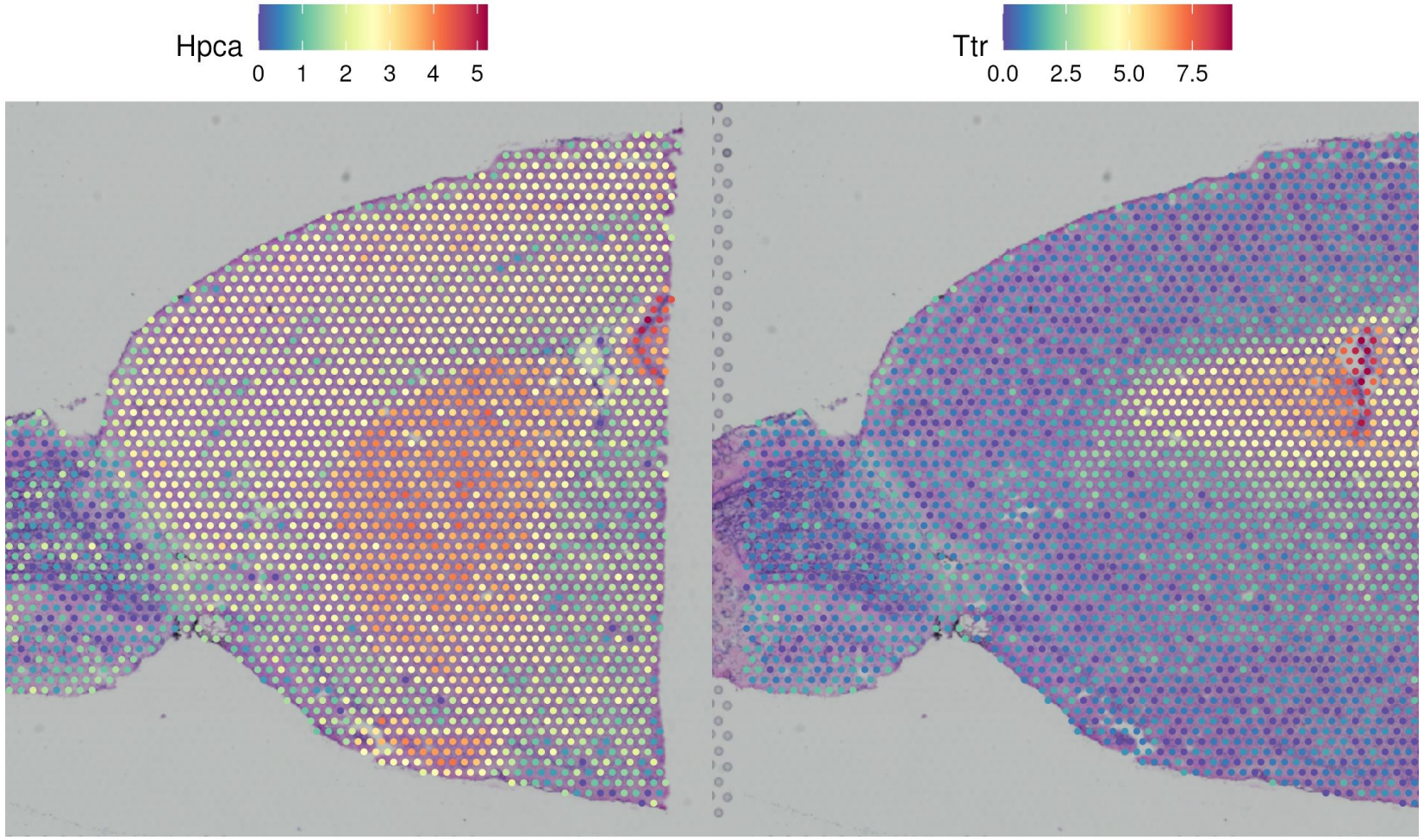
Spatial caveat

Cell density and anatomy also affect counts; normalization can remove or preserve different signals.

Report clearly

State assay, layer, normalization method, variable feature choice and Seurat version.

Seurat vignette figure | SpatialFeaturePlot overlays expression on histology



WHY IT MATTERS

SpatialFeaturePlot places expression back onto the tissue image, making anatomy part of interpretation.

Use the diagram to connect technology, biology and analysis.

Seurat principle 4: PCA is the hinge between genes and graphs

1. **Features** — Variable genes define what signal can enter PCA.
2. **PCA** — RunPCA compresses correlated expression into orthogonal axes.
3. **PC choice** — Selected dimensions determine neighbors, clusters and UMAP.
4. **Validation** — Interpret PCs with marker genes, tissue regions and technical metrics.

Seurat principle 5: clusters are graph communities, not cell types

FindNeighbors

Builds a graph from selected PCA dimensions.

FindClusters

Finds communities in that graph; resolution controls granularity.

Interpretation

A cluster becomes biological only after markers, tissue location and replicates support it.

Seurat principle 6: UMAP is a map of similarity, not a map of tissue

What UMAP shows

A two-dimensional embedding of expression similarity after PCA.

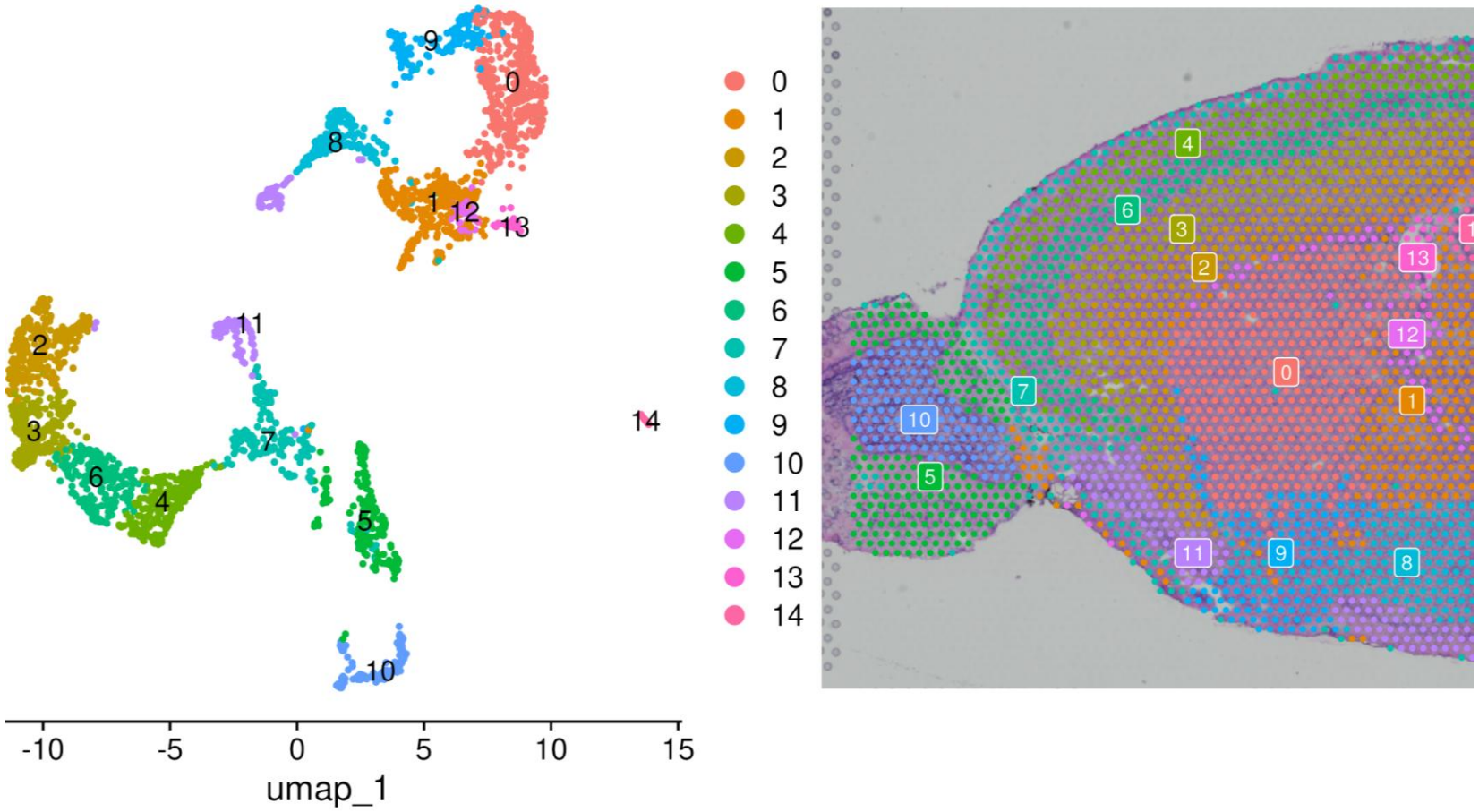
What UMAP does not show

It does not preserve physical tissue distance or histologic adjacency.

How to use it

Compare DimPlot with SpatialDimPlot; disagreement is often biologically informative.

Seurat vignette figure | UMAP and tissue overlays are complementary



WHY IT MATTERS

The same clusters are shown in UMAP space and tissue space; interpretation comes from comparing transcriptional similarity with anatomical localization.

Use the diagram to connect technology, biology and analysis.

Seurat principle 7: spatial plots should be diagnostic, not decorative

SpatialFeaturePlot

Use for genes, signatures, QC metrics and prediction scores.

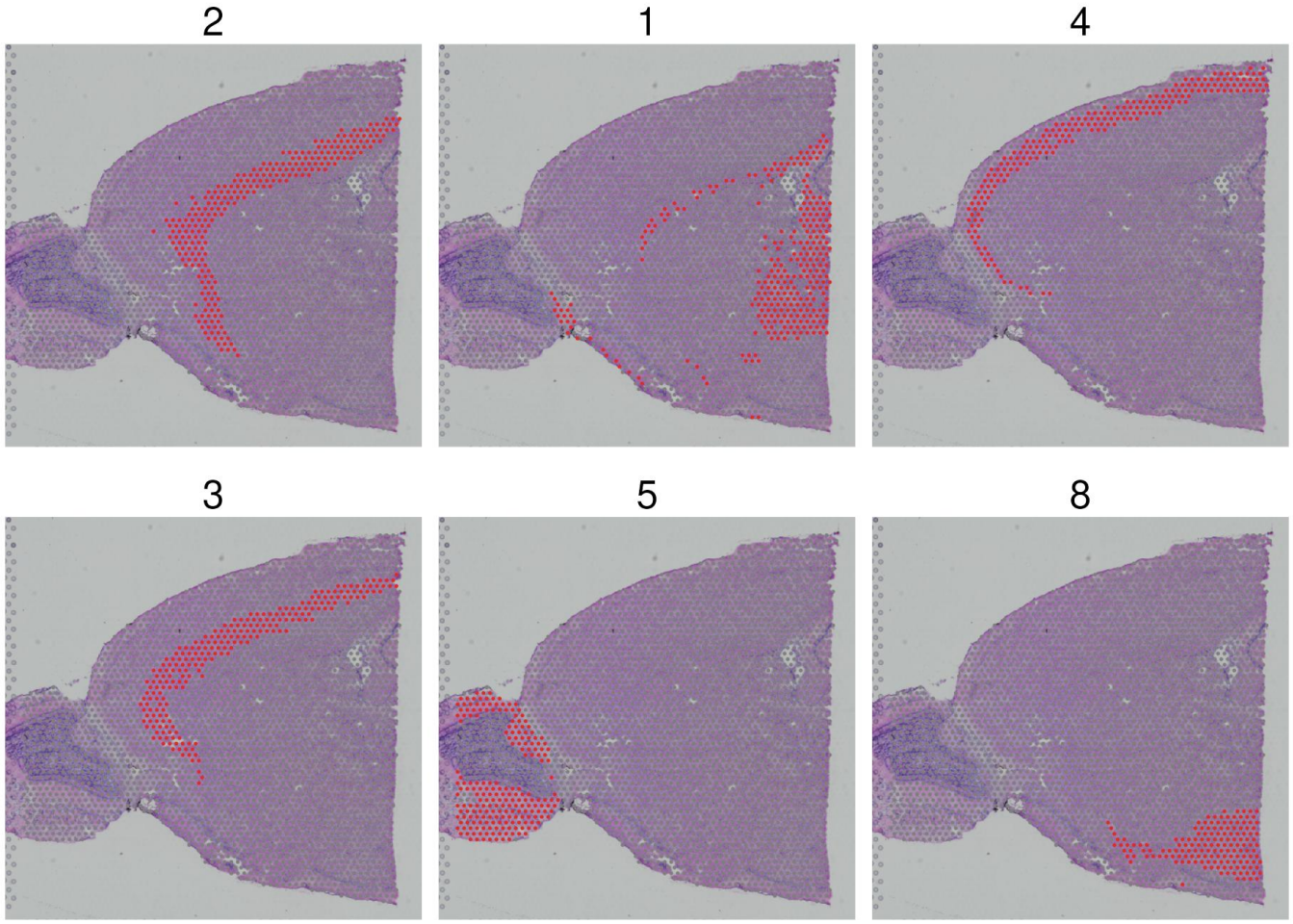
SpatialDimPlot

Use for clusters, labels and region-level annotations.

Reading rule

Always compare signal with morphology, tissue edge, depth and neighboring labels.

Seurat vignette figure | Highlighting clusters tests spatial localization



WHY IT MATTERS

Facet highlighting makes it easier to judge whether clusters form coherent tissue territories or represent scattered expression states.

Use the diagram to connect technology, biology and analysis.

Seurat principle 8: annotation is a chain of evidence

1. **Markers** — Use known markers to name clusters cautiously.
2. **Reference** — Map scRNA-seq labels when a relevant reference exists.
3. **Scores** — Inspect prediction scores, not only the maximum label.
4. **Tissue** — Confirm labels make sense in histologic context.

Seurat principle 9: mixed spots require composition-aware interpretation

Visium reality

A spot can contain multiple cells and multiple cell types.

Label transfer

Gives likely labels or prediction scores, useful but not literal staining.

Deconvolution

Estimates mixtures; conclusions depend on reference quality and cell states included.

Seurat principle 10: spatially variable genes need a spatial hypothesis

Cluster markers

FindMarkers compares predefined groups such as clusters or regions.

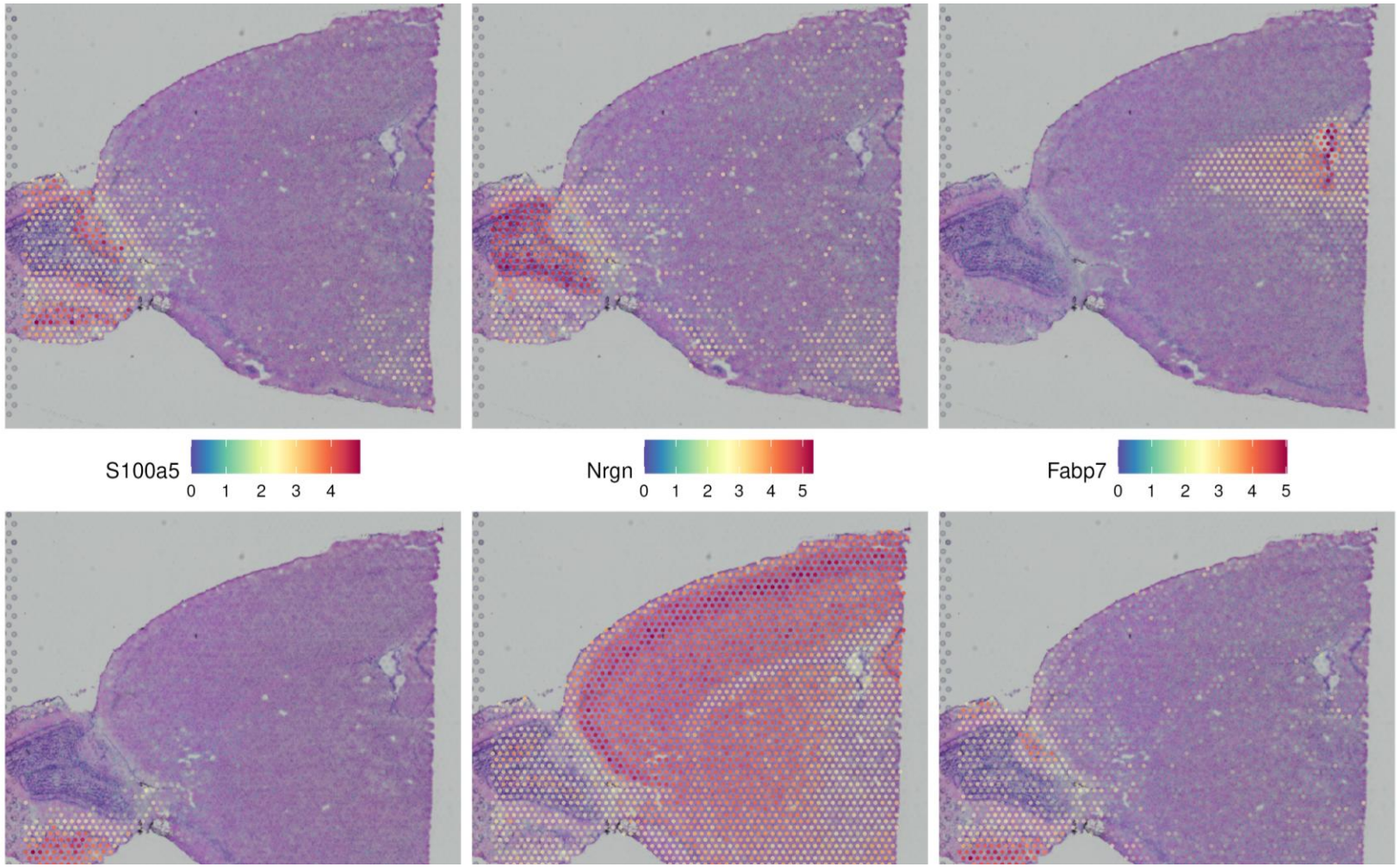
Spatial features

FindSpatiallyVariableFeatures searches for spatial patterning without pre-annotated regions.

Caution

Spatial autocorrelation can reflect anatomy, depth, cell density or true regulation.

Seurat vignette figure | Spatially variable features are pattern tests



WHY IT MATTERS

FindSpatiallyVariableFeatures ranks genes whose expression depends on spatial location; biological meaning still requires tissue and composition checks.

Use the diagram to connect technology, biology and analysis.

Seurat principle 11: multiple slices are not automatically one dataset

1	**Per-slice QC**	Check image alignment, counts and tissue detection separately.
2	**Metadata**	Keep patient, section, chemistry and processing versions explicit.
3	**Integration**	Integrate only if it answers the biological question.
4	**Statistics**	Treat patient or sample as the replicate, not spots.

Seurat analysis checklist before showing a result

01 **Can the plot be traced back to raw counts, image, coordinates and sample metadata?**

02 Were QC thresholds checked both statistically and spatially?

03 Are PCA dimensions and cluster resolution documented?

04 Does UMAP interpretation stay separate from tissue-space interpretation?

05 Are annotation and deconvolution treated as models with uncertainty?

Seurat official spatial vignette: https://satijalab.org/seurat/articles/spatial_vignette.html

Case study

Can genetically distinct ovarian cancer subclones build distinct local microenvironments?

The study combines Visium discovery with single-cell-resolved CosMx validation in post-chemotherapy HGSOC.

CASE STUDY

The study uses complementary cohorts and resolutions

Visium discovery

8 HGSOC patients
1 section per patient
5–9 expression clusters
~1,501–3,584 spots

scRNA reference

Post-NACT HGSOC cells
12 major cell types
Used for RCTD decomposition

CosMx validation

Single-cell spatial imaging
Focused validation cohort
Subclone and signaling context

CASE STUDY

What spatial adds to this ovarian cancer study

Without space

A bulk or scRNA-only result could show tumor programs and cell types, but not whether they occupy the same tissue territory.

With Visium

Expression domains and inferred CNA patterns can be mapped back to morphology and local composition.

With CosMx

Single-cell localization tests whether subclone markers and signaling partners are physically co-distributed.

What the paper establishes

HGSOC contains spatially discrete subclones embedded in distinct microenvironments.

The work links CNA-defined tumor heterogeneity to local transcriptional programs and proposes subclone-specific autocrine signaling hypotheses.

Strengths and limitations should be read together

Strengths

Multi-platform design
Patient tissue context
Genomic-like and transcriptomic integration
Single-cell spatial validation

Limitations

Small cohorts and limited sections
Post-treatment samples
InferCNV is indirect
Communication remains correlative

Five questions to ask before trusting any spatial claim

01 **Is the biological replicate the patient—not the spot or cell?**

02 Could the pattern be explained by depth, morphology or composition?

03 Does the result persist across patients and analysis choices?

04 Is there orthogonal spatial or molecular validation?

05 Does the conclusion match the strength of evidence?

General critical-appraisal framework; supplied paper.

A GEO reanalysis should proceed from raw structure to focused hypotheses

1. **Inventory** — List samples, file types, barcodes, matrices, images and metadata.
2. **Reconstruct** — Create one spatial object per section and verify image alignment.
3. **QC** — Apply tissue-aware metrics and record every exclusion.
4. **Model** — Cluster, deconvolve, infer CNA and test patient-aware contrasts.
5. **Validate** — Reproduce key paper figures, then perturb one assumption at a time.

GSE211956 contains the discovery data and supporting references

Visium

Spatial count matrices and sample-specific files
Capture positions and tissue images may be packaged separately

scRNA-seq

Reference profiles used for RCTD
Cell annotations and raw/processed matrices

Metadata

Patient, response score and sample provenance
Required for patient-aware modeling

A minimal reproducible object should pass four checks

Identity

Barcode and sample IDs are unique and traceable.

Geometry

Coordinates fall within the image and tissue mask.

Counts

Raw integers, gene IDs and feature types are preserved.

Provenance

Patient, section, platform and processing version are recorded.

A useful hands-on exercise: reproduce one figure, then challenge it

01 **Recreate tumor-cell weights and spatial clusters for one patient**

02 Repeat deconvolution with an alternative ovarian cancer reference

03 Compare inferCNV subclones across normalization choices

04 Quantify which conclusions remain unchanged

GEO GSE211956; supplied paper.

SYNTHESIS

Match the method to the biological question

1	**Where?**	Use spatial visualization, domains and pathology alignment.
2	**Who is there?**	Use markers, deconvolution or single-cell segmentation.
3	**What state?**	Use composition-adjusted programs and patient-aware contrasts.
4	**Who interacts?**	Use neighborhoods plus ligand–receptor and downstream support.
5	**What causes it?**	Move to perturbation, longitudinal sampling or functional models.

SYNTHESIS

Three take-home messages

Space is a measurement dimension, not decoration.

Separate composition from state; treat patient as the replicate; use spatial associations to prioritize—not prove—mechanisms.