Spatial Transcriptomics Analysis of Breast Cancer and Lymph Node Metastases

This repository presents a reproducible, biologically informed spatial transcriptomics analysis using breast cancer samples. It integrates spatial gene expression, single-cell-derived signatures, and region-level annotation to explore tumor–immune–stromal organization across matched primary tumors and lymph node metastases (LNMT).

Dataset Summary

- **Tissue Types**: Primary breast tumors and matched LNMT tissue sections.
- **Samples**:
  - 4 Visium slides (2 primary tumor, 2 LNMT)
  - Signature scores derived from prior scRNA-seq data.
- **Modalities Integrated**:
  - Spatial transcriptomics (10X Visium)
  - Signature scoring (epithelial, immune, stroma, proliferation)
  - Region-based annotation and masking
  - Tissue classification (primary vs LNMT)
  - Comparative analysis (composition, DE genes, spatial exclusion)
  - Summary gene tables (top DE genes per group)
  - Region–region adjacency and spatial inference

Analysis Overview

1. Sample Loading and QC

- Four spatial datasets loaded into AnnData.
- Tissue boundaries visualized with H&E overlay.
- Normalization and filtering performed.

2. Signature-Based Program Scoring

- Spot-wise scoring of four gene programs:
  - Epithelial (tumor-like)
  - Stromal (connective/fibrotic tissue)
  - Immune (lymphoid infiltration)
  - Proliferation (cell cycle activity)
- Spatial plots reveal program localization and tissue heterogeneity.

3. Region-Based Annotation and Spatial Characterization

3.1 Region Masking via Signature Thresholds

- Each spot was scored for epithelial, immune, stromal, and proliferative signatures.
- A mask was applied to the top 20% of spots per score to define high-activity regions.

3.2 Combined Region Labeling

- A hierarchical rule-based approach assigned a single region label per spot:
  - Priority order: Tumor > Immune > Proliferative > Stroma > Unassigned.
- This enabled mutually exclusive, interpretable region classification across the tissue.

3.3 Compact Region Plots

- Each sample was visualized using a compact panel layout with a refined color palette.
- This provided an at-a-glance spatial overview of region organization across all slides.

3.4 Region Composition Summary

- Tables summarized the percentage of spots per region and per tissue type (primary vs LNMT).
- Primary tumor sections showed higher tumor and stromal proportions.
- LNMT samples exhibited greater immune and proliferative content.

3.5 Mean Signature Scores by Region

- Scores were averaged across all samples, grouped by assigned region.
- This confirmed biological specificity:
  - Tumor regions were enriched in epithelial scores.
  - Immune regions showed strong immune signal.
  - Proliferative regions were dominated by cell cycle genes.
  - Stroma had elevated stromal expression and lower activity otherwise.

3.6 Region–Region Spatial Adjacency

- Region adjacency graphs were constructed from spatial coordinates.
- This enabled neighborhood-level inference and spatial exclusion patterns.
- Tumor and immune regions frequently appeared adjacent but showed minimal overlap, consistent with partial exclusion and compartmentalization.

4. Spatial Pattern Quantification

- Region proportions summarized per sample.
- Compact tables show compositional variation and mutual exclusion trends.

5. Tissue-Type Comparison

- Samples manually annotated by tissue type (primary or LNMT).
- Region prevalence compared between tissue groups.
- Immune and proliferative regions enriched in LNMT, while tumor and stromal regions dominate primary tumors.
- Spatial organization trends (e.g., compartmentalization, immune–tumor separation) are highlighted by region plots.

6. Differential Expression Analysis

- Differential gene expression computed between primary and LNMT samples.
- Full ranked table generated (saved as CSV and Excel).
- Summary tables of top upregulated genes per tissue type provided.
- DE results interpreted cautiously as complementary to region-based analysis.

Main Findings

- Epithelial, stromal, immune, and proliferative programs exhibit distinct spatial distributions.
- Tumor–immune compartmentalization and stromal peripheries are conserved across tissues.
- Region assignment allows clear and interpretable labeling for downstream analysis.
- Signature-based masking is effective for identifying functional tissue zones without cell segmentation.
- Immune infiltration and proliferation are more prominent in LNMT.
- DE analysis identifies candidate genes but is secondary to region-centric spatial inference.
- Region–region adjacency analysis reveals functional boundaries and spatial relationships.
- Simple region-based annotations can reveal biologically meaningful tissue organization that complements raw expression maps.

Tools and Technologies

- Python (Scanpy, matplotlib, pandas, NumPy)
- Jupyter Notebooks
- Anndata structure
- Gene program scoring via `sc.tl.score_genes`
- Modular and interpretable visualizations

Results demonstrate how scRNA-seq-derived programs and spatial localization can reveal tissue-level coordination in cancer samples.

-