Immune Exclusion vs Infiltration (IEvI) — Tutorial Overview

Aim

This notebook demonstrates how the **SpatialMMKPNN framework** can be applied to investigate immune exclusion versus infiltration in tumor tissues. The workflow is designed as a **step-by-step, tutorial-style pipeline** that begins with raw Visium-like inputs and produces interpretable, audit-ready results.

The specific focus is on quantifying **ligand–receptor (LR) signaling axes** at tumor boundaries and interiors, providing biological insights into exclusion mechanisms such as stromal barriers, chemokine gradients, and angiogenic interfaces.

What This Notebook Shows

- End-to-end reconstruction of IEvI from raw inputs to interpretable results.
- Explicit construction of LR edges constrained by spatial graphs.
- Quantification of tumor rim enrichment with transparent statistics.
- Robustness analysis to confirm conclusions are not dependent on a single parameter choice.
- Practical troubleshooting notes for reproducibility.

Key Design Choices (the “why”)

- **Mechanism-first unit:** explicit ligand→receptor edges in space, not just cell proximity.
- **Tumor rim baseline:** each axis’s boundary share is compared against that slide’s own geometry (p0), avoiding artifacts from shape differences.
- **Transparent stats:** per-axis counts, fractions, Wilson confidence intervals, and one-sided tests vs baseline.
- **Robustness:** rim thickness and detection thresholds varied to ensure stable calls.

Method (step-by-step)

Preprocessing

- Load 10x matrices (`.h5`, `matrix.mtx` + barcodes/features) or `.h5ad`.
- Keep in-tissue spots and attach pixel coordinates.
- Normalize counts to CPM → log1p.
- Apply `gene_aliases.yaml` and collapse duplicates.
- **Output:** `preproc_summary.csv`

Spatial Graph

- Build a k-nearest neighbor graph (k=8) on pixel coordinates.
- **Output:** `graph_summary.csv`

Axes (biology tested)

- Fixed set for lEvl analysis:
  - VEGFA→KDR
  - CXCL12→CXCR4
  - TGFB1→TGFBR*
  - SPP1→ITG*

Roles (tumor/stroma/immune)

- Preferred: use `configs/role_map.csv`.
- Otherwise: derive from marker genes, smoothed by kNN.

Tumor Rim (boundary vs interior)

- Rasterize tumor spots, blur silhouette, fill holes.
- Compute distance to boundary, classify rim vs interior.
- **Outputs:** `region_summary.csv`, `region_summary_tumor.csv`

Edge Accounting

- For each directed graph edge, check ligand at source and receptor at target.
- Count LR edges, assign to rim or interior.

- **Outputs:** `edges_counts_by_axis.csv`, `edges_counts_by_axis_tumor.csv`

Statistics (boundary enrichment)

- Calculate `boundary_share = boundary_edges / total_edges`.
- Compare against baseline p0 (rim fraction).
- Report Wilson CIs, one-sided enrichment/depletion calls.
- **Outputs:**
  - `axis_boundary_stats.csv`, `axis_boundary_stats_tumor.csv`
  - `axis_boundary_calls.csv`, `axis_boundary_calls_TUMOR.csv`
  - Pooled versions (`*_pooled.csv`)
  - Visualizations in `plots/` and `plots_tumor/`

Robustness

- Re-run enrichment under rim ±25% and CPM≥1 threshold.
- Record flips and sensitivity.
- **Output:** `robustness_calls.csv`

Reproducibility Note

- All intermediate outputs are recomputed for visibility.
- Checkpoints optional; not required to reproduce results.

Results (lEvl Summary)

- **VEGFA→KDR and SPP1→integrins:** consistently rim-enriched across slides, indicating angiogenic and adhesion/matrix interfaces concentrate at tumor edges (exclusion-like).
  See enrichment calls in `axis_boundary_calls_TUMOR.csv` and visualizations in `plots_tumor/axis_VEGFA-KDR_boundary_share.png` and `plots_tumor/axis_SPP1-ITG_boundary_share.png`.
- **TGFB1→TGFBR*:** mostly mild rim enrichment or neutral, with slide-dependent variability.
  See `axis_boundary_stats_tumor.csv` for statistics and corresponding plots in `plots_tumor/axis_TGFB1-TGFBR_boundary_share.png`.

- **CXCL12→CXCR4:** context-dependent; interior-skewed on some slides, rim-enriched on others (consistent with chemotaxis biology).
  See per-slide calls in `axis_boundary_calls_TUMOR.csv` and corresponding figures in `plots_tumor/axis_CXCL12-CXCR4_boundary_share.png`.

- **Robustness:** calls remained stable under rim ±25% and CPM≥1, with only a few borderline flips flagged as low-confidence.
  See robustness analysis in `robustness_calls.csv`.

- **Caveat:** one slide with few tumor spots (when roles were derived from markers) showed weaker rim definition; providing a curated `role_map.csv` would improve accuracy.

Limitations & Recommendations

- **Role inference variability:** derived roles can be noisy; weak tumor labeling occurs without curated `role_map.csv`.

- **Geometry dependence:** rim definition relies on distance transforms; results should be interpreted as relative enrichments.

- **Context-dependence:** some axes (e.g., TGFB1→TGFBR*, CXCL12→CXCR4) shift differently across slides, limiting generalization.

- **Low tumor coverage:** slides with few tumor spots yield unstable rim definitions and weaker enrichment calls.

Recommendations:

- Use curated role annotations (`role_map.csv`) when available.

- Always report robustness checks (rim ±25%, CPM thresholds).

- Interpret enrichment as **relative trends** across slides, not absolute metrics.

- Validate context-dependent axes in multiple slides or cohorts before biological interpretation.

Troubleshooting

- **“No spatial” error:** check `spatial/tissue_positions*`; adjust blur sigma if rim too thin.

- **Few tumor spots:** derived roles too strict; provide `role_map.csv`.

- **Counts too small:** raise k in neighbor graph (e.g., k=10) or relax detection, but document changes.

- **Rim too thick/thin:** adjust distance quantile and re-run; report both settings.

Interpretation

This application shows how SpatialMMKPNN can **distinguish immune exclusion vs infiltration motifs** in tumor slides. Rim-enriched VEGF and SPP1 signals highlight stromal and vascular barriers, while variable CXCL12 gradients reveal heterogeneity in immune cell recruitment.

The tutorial emphasizes **auditability** (all outputs in CSVs/plots), **robustness checks**, and **mechanistic clarity** (tracking LR edges, not just proximity). It demonstrates how immune access to tumors can be systematically studied in a reproducible way.