



浙江大学
ZHEJIANG UNIVERSITY



X



Multiscale Microenvironment Vector Space Projection for Uncovering Diverse Pathological Biomarker

Zixiu Ding* · Wenxuan Hao* · Liya Ding · Yuan Dai
Jiazhen Yang · Shuwen Han · Lingxiang Jia · Zunlei Feng[†]

Zhejiang University · Central South University
Sir Run Run Shaw Hospital · Tarim University

* Equal contribution

[†] Corresponding author

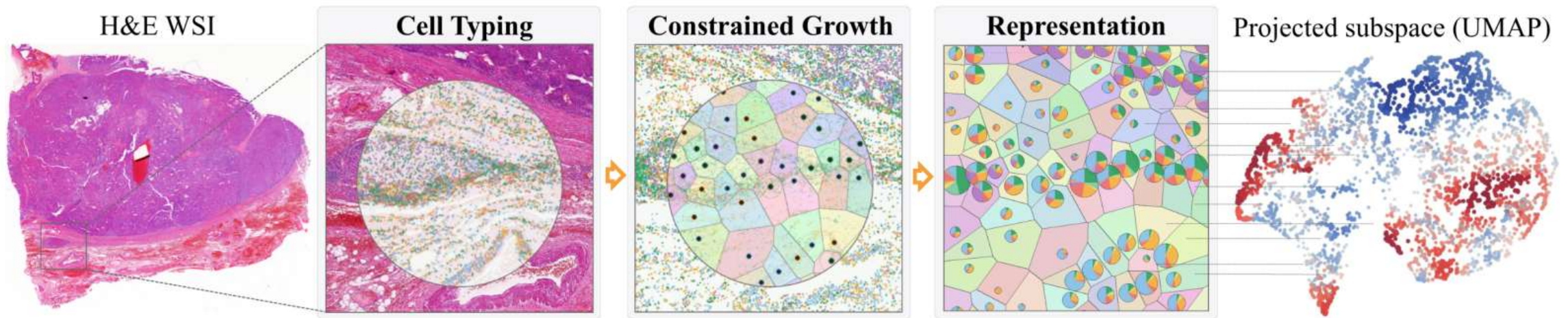
NeurIPS 2026, Sydney, Australia



The challenge: tissue organization varies in scale

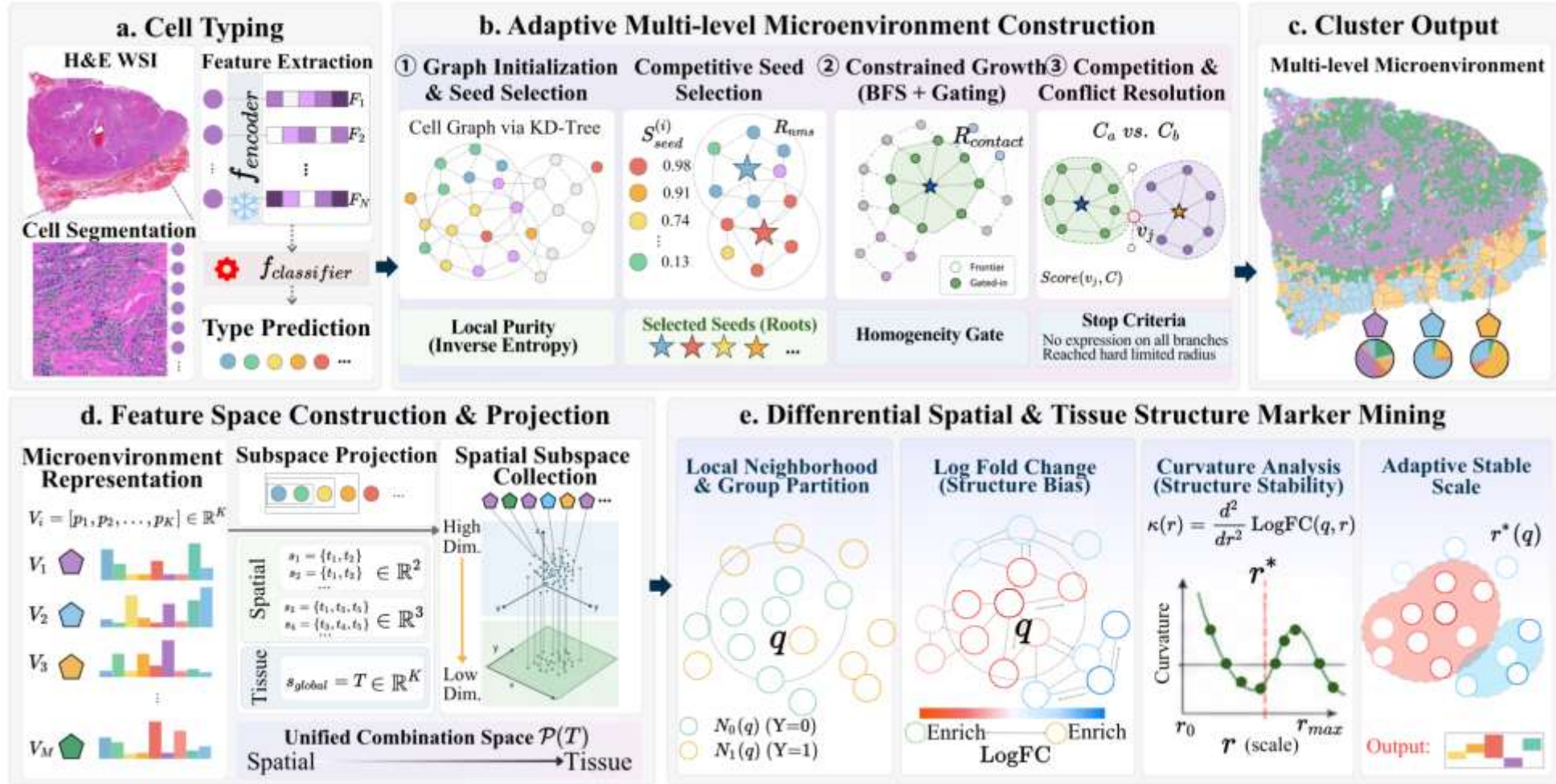
Fixed spatial windows can mix distinct cellular communities

We seek variable-size niches that follow local phenotype and physical adjacency.

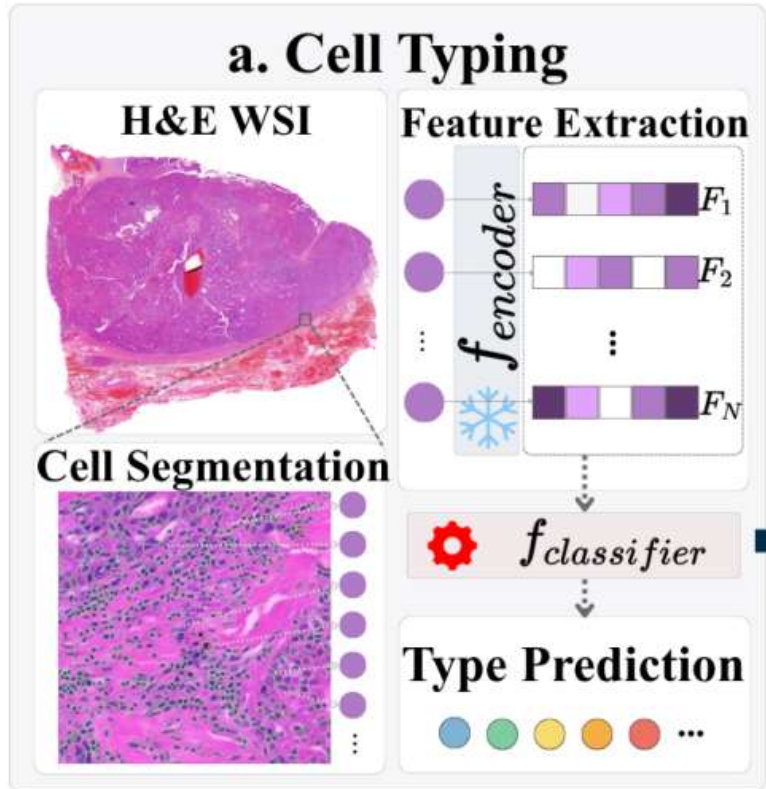


H&E → cell types → coherent niches → comparable composition vectors

Framework: tissue niches to differential biomarkers



1. Cell typing and local phenotypic representation



Predict a semantic cell type

$$T_i = f_{\text{classifier}}(F_i), \quad T_i \in \{T_1, \dots, T_K\}, \quad (1)$$

$F_i = \text{fencoder}(v_i)$: morphology features of cell i

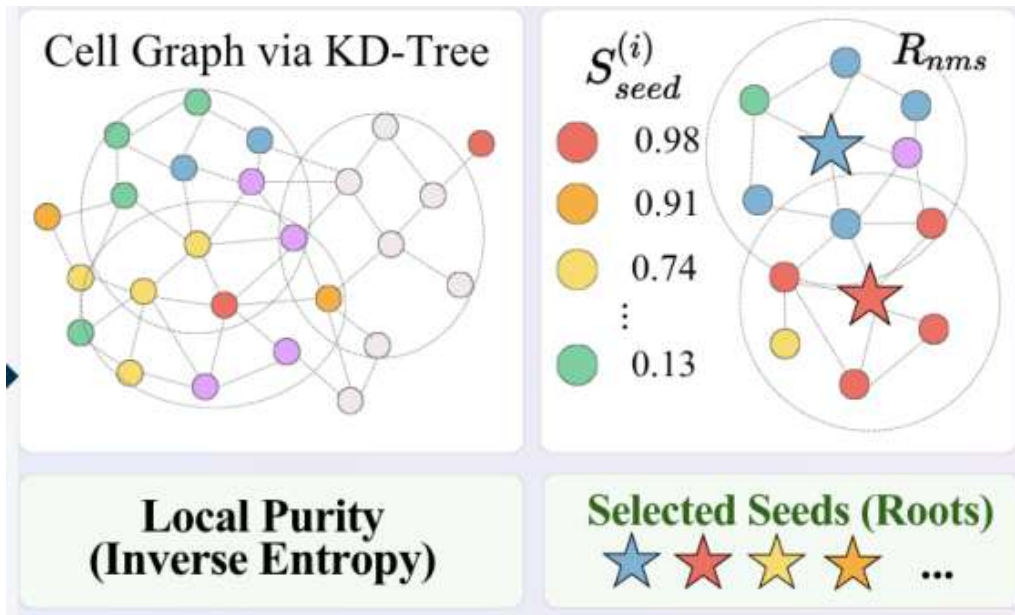
Build a physical neighborhood G_i with a KD-Tree

$$P_i = [p_1, p_2, \dots, p_K]$$

p_k : fraction of cell type k in the local neighborhood

Local composition connects cell predictions to spatial niche construction.

2. Select pure, spatially separated niche seeds



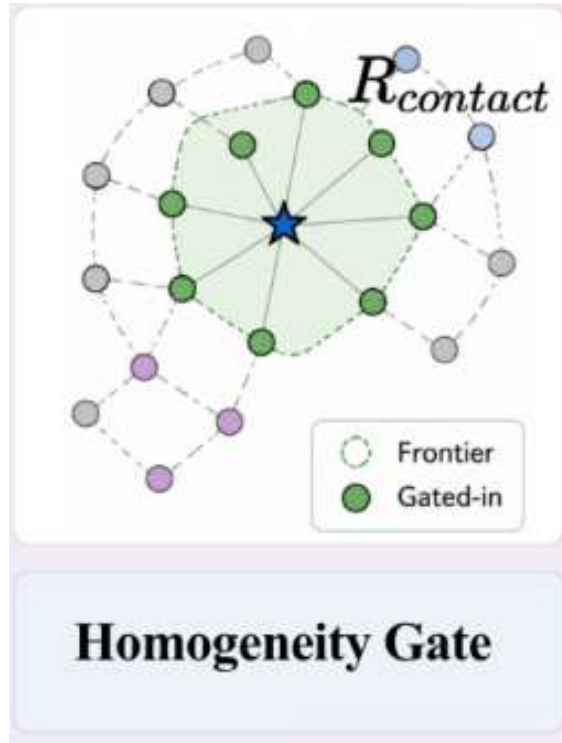
$$S_{seed}^{(i)} = \frac{1}{H(P_i)}, \quad H(P_i) = - \sum_{k=1}^K p_k \log p_k. \quad (2)$$

Lower entropy $\rightarrow$ higher seed priority

- 1 Rank cells by local purity
- 2 Select the highest-scoring unmasked cell
- 3 Suppress neighbors within R_{nms}

Initialize each niche centroid with the selected root composition.

3. Grow niches with an adaptive similarity gate



P_j : local composition
 n : accepted cell count

Accept a neighbor

$$\cos(\theta) = \frac{P_j \cdot \bar{P}_C^n}{|P_j| |\bar{P}_C^n|} > \mathcal{T}_C(n), \quad (3)$$

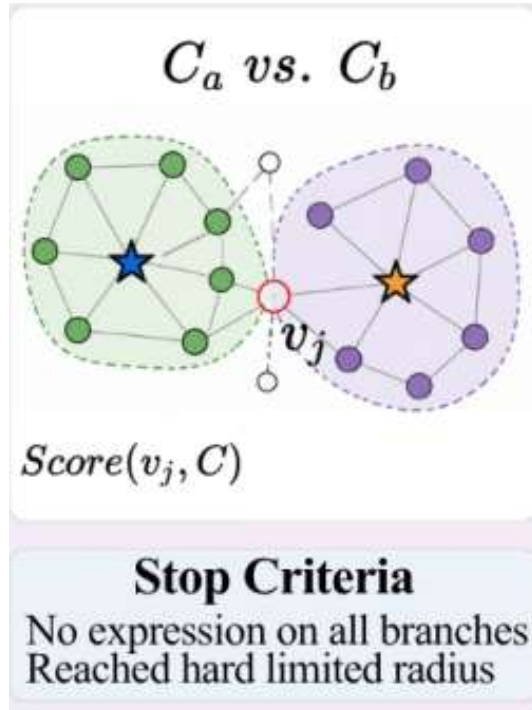
Update the centroid

$$\bar{P}_C^{n+1} = \frac{n \cdot \bar{P}_C^n + P_j}{n + 1}. \quad (4)$$

Tighten the threshold as the niche grows

$$\mathcal{T}_C(n) = \mathcal{T}_{\text{base}} + \sigma \log(n), \quad (5)$$

4. Resolve collisions using phenotype and distance



$$Score(v_j, C) = \omega D_M(v_j, C) + (1 - \omega) \frac{\|\mathbf{xy}_j - \mathbf{xy}_{top, C}\|_2}{R_{current, C}}, \quad (6)$$

Assign each contested cell to the lowest-score niche

$D_M(v_j, C)$: Mahalanobis distance in phenotypic space

$\mathbf{xy}_{top, C}$: position of the selected root

$R_{current, C}$: current maximum niche extent

ω : phenotype–distance weighting

5. Project niches into cell-composition subspaces

Select a subset of cell types

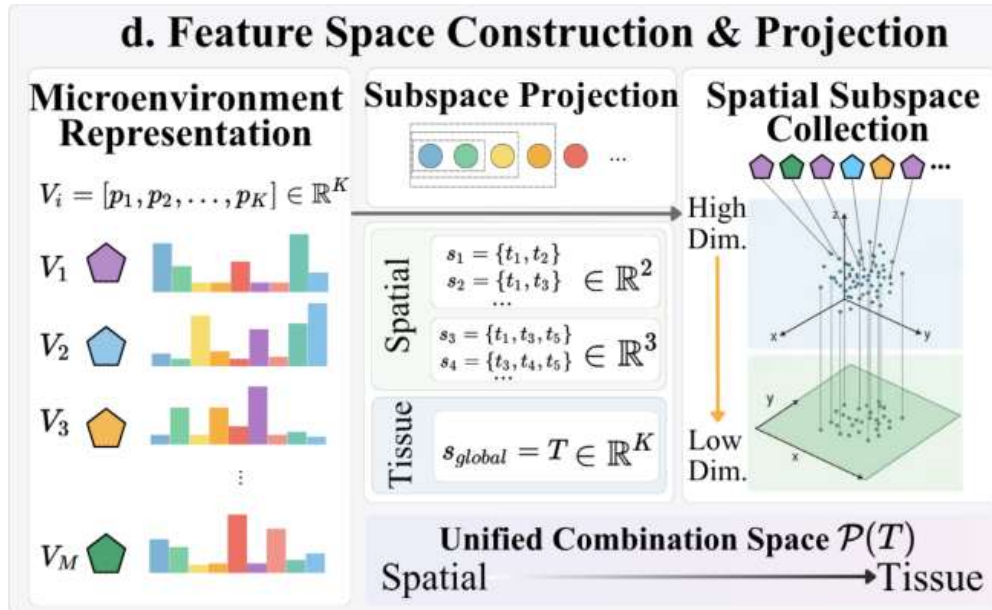
$$V_{m,s} = \Pi_s(V_m), \quad (7)$$

Collect complementary subspaces

$$\mathcal{V}_m = \{V_{m,s} \mid s \in \mathcal{P}(T)\}, \quad (8)$$

$s \subset T$: spatial co-occurrence patterns

$s = T$: full tissue-unit composition



Different-sized niches become comparable through composition.

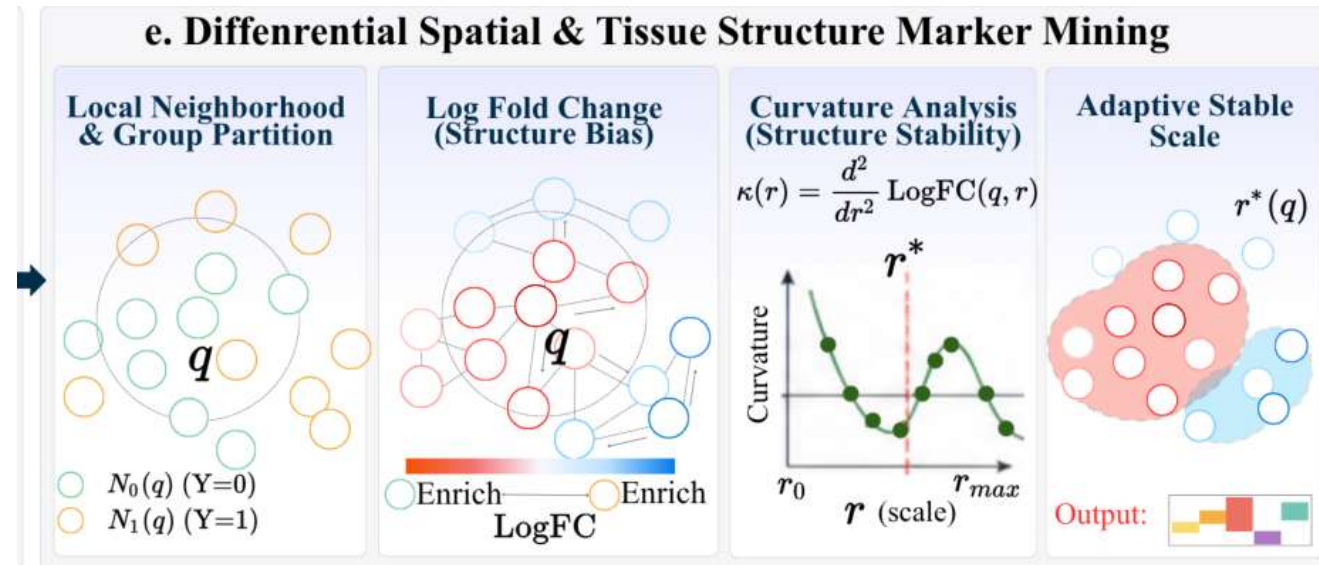
6. Compare local distributions across outcome groups

Retrieve similar compositions

Measure outcome-group enrichment

$$N(q, r) = \{z \mid \|z - q\|_2 \leq r\}, \quad (9)$$

$$\text{LogFC}(q, r) = \log_2 \frac{|N_0(q, r)| + \epsilon}{|N_1(q, r)| + \epsilon}, \quad (10)$$



7. Select a stable phenotypic scale

Track curvature of the differential trajectory

$$\frac{d^2}{dr^2} \text{LogFC}(q, r). \quad (11)$$

r : distance in composition space

q : query niche

ε : count smoothing constant

Select the radius with minimum absolute curvature

$$r^*(q) = \arg \min_r \left| \frac{d^2}{dr^2} \text{LogFC}(q, r) \right|, \quad (12)$$

Output: a stable local composition interval with its group-enrichment direction.

Evaluation design and operating parameters

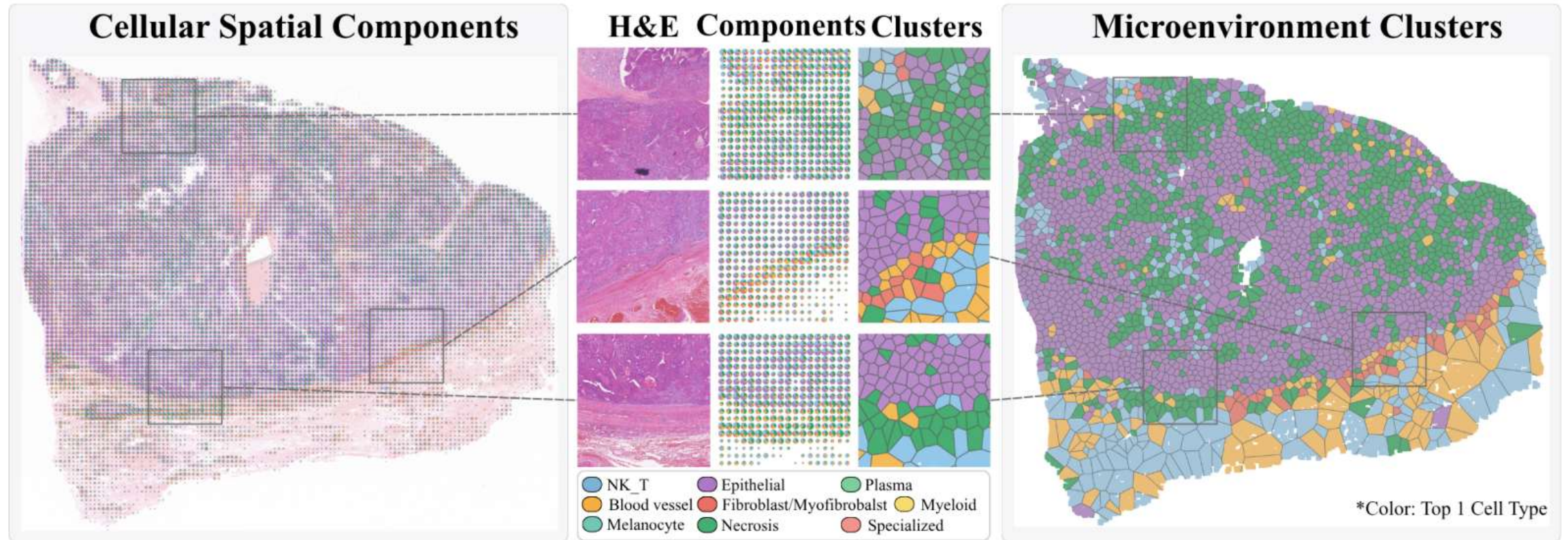
Component	Setting
Cell typing	STHELAR $\approx$ 11M cells; ConvNeXt Tiny
Cancer cohorts	LIHC, BLCA and BRCA
Independent cohort	Private LIHC discovery $\rightarrow$ TCGA
Molecular reference	5 paired H&E–Xenium slides

Niche construction: $R_{nms} = 500$; $R_{contact} = 80$; $R_{hard-limit} = 1500$

Growth gate: $T_{base} = 0.90$; $\sigma = 0.01$

Discovery: $|\text{LogFC}| \geq 0.6$; nominal $p < 0.05$; balanced group sampling

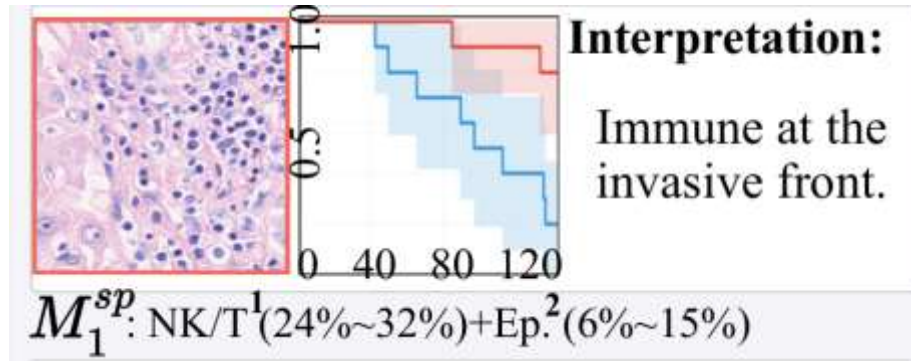
Adaptive niches recover coherent tissue structures



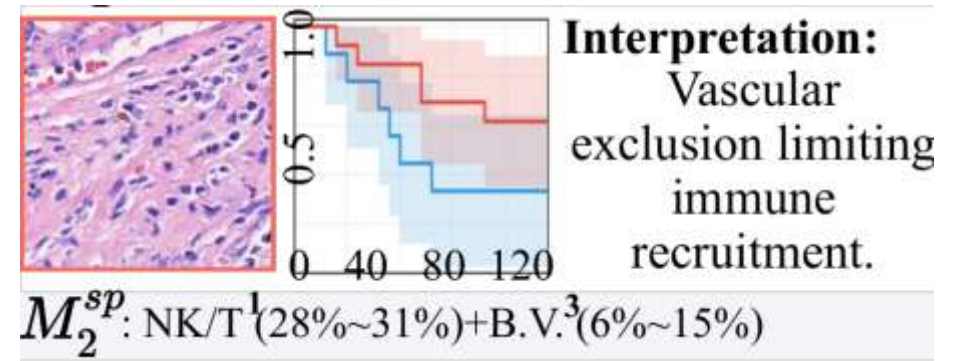
Local phenotype, spatial continuity and histological boundaries remain visible.

Liver cancer (LIHC): spatial biomarker patterns

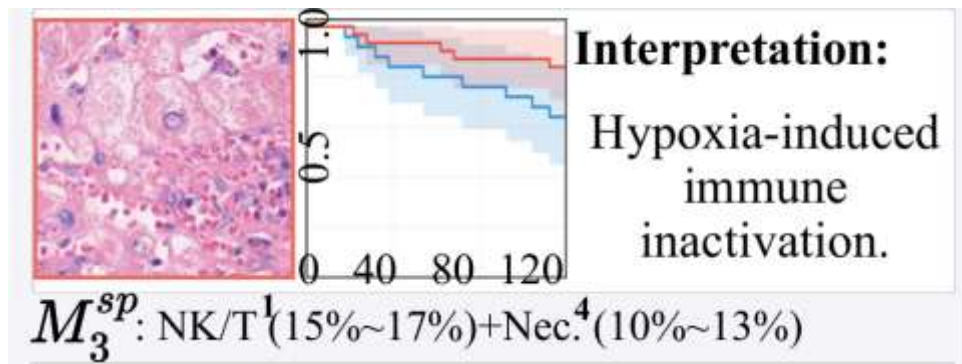
M1 · Poor enrichment · $p = 0.00477$



M2 · Poor enrichment · $p = 0.00000459$



M3 · Poor enrichment · $p = 0.0261$



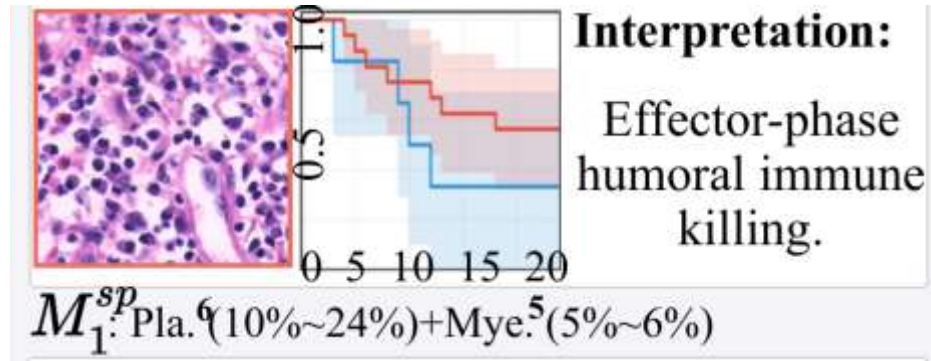
Three distinct NK/T-associated spatial configurations

Blue border: favorable; red: poor prognosis
KM: survival probability vs months

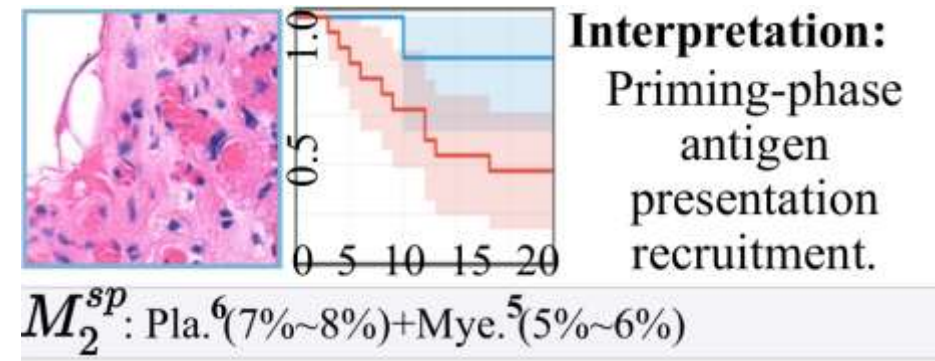
Nominal associations; source pathological interpretations are hypotheses.

Bladder cancer (BLCA): spatial biomarker patterns

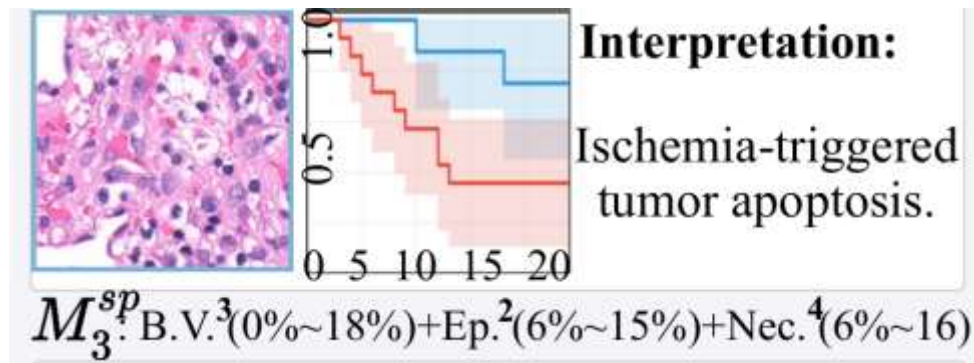
M1 · Poor enrichment · $p = 0.0133$



M2 · Favorable enrichment · $p = 0.0141$



M3 · Favorable enrichment · $p = 0.000366$



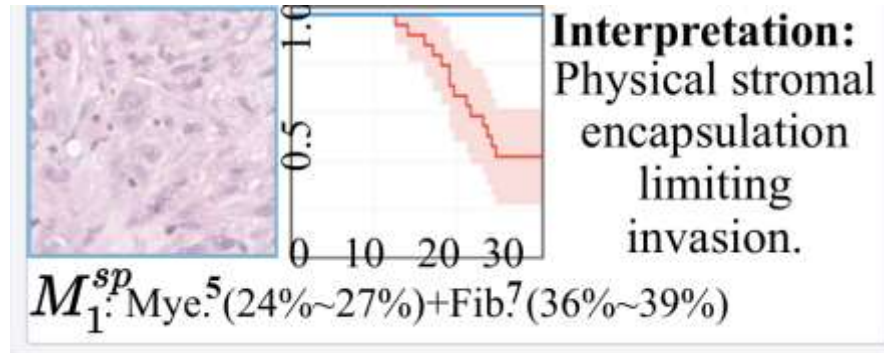
Closely related plasma–myeloid compositions
have different enrichment directions

Blue border: favorable; red: poor prognosis
KM: survival probability vs months

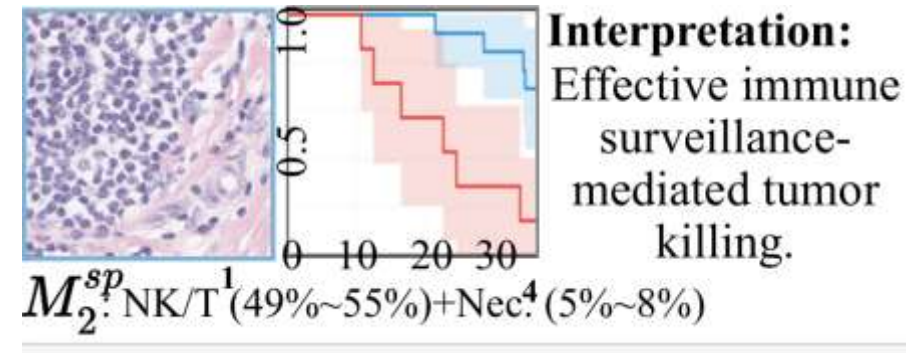
Nominal associations; source pathological interpretations are hypotheses.

Breast cancer (BRCA): spatial biomarker patterns

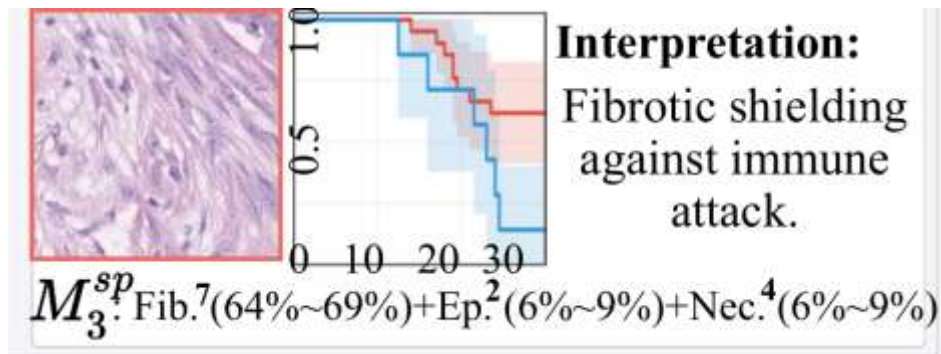
M1 · Favorable enrichment · $p = 0.0153$



M2 · Favorable enrichment · $p = 0.0349$



M3 · Poor enrichment · $p = 0.0175$



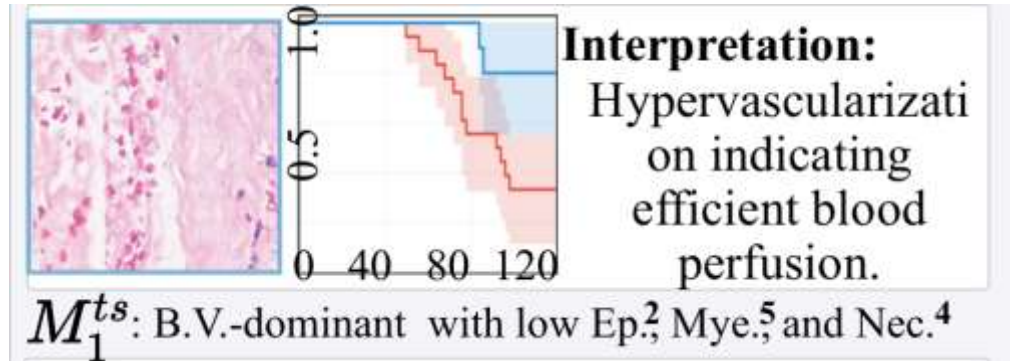
Stromal, immune and necrotic combinations distinguish outcome groups

Blue border: favorable; red: poor prognosis
KM: survival probability vs months

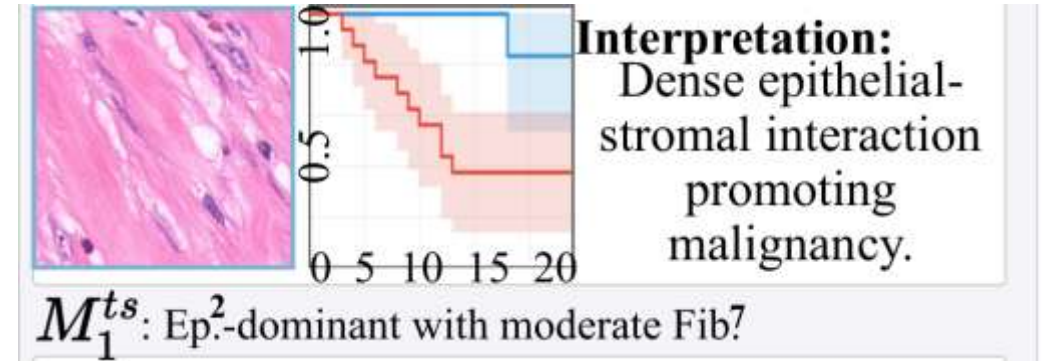
Nominal associations; source pathological interpretations are hypotheses.

Tissue biomarkers use the complete cell composition

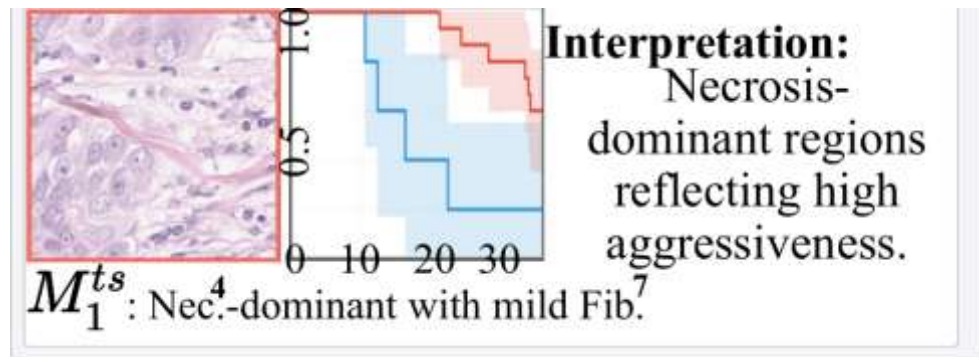
LHIC · Favorable · $p = 0.0170$



BLCA · Favorable · $p = 0.0167$



BRCA · Poor · $p = 0.00943$



Full-vector patterns capture vascular, epithelial and necrotic tissue organization.

BLCA: Table 1 reports favorable enrichment; the source interpretation describes malignancy.

Blue border: favorable; red: poor. KM: survival probability vs overall survival (months).

Method comparison: association strength and homogeneity

Method	Median $-\log_{10}(p) \uparrow$	Entropy $\downarrow$	Mean cosine
diffcyt	1.354	0.408	0.875
MELD	0.844	0.458	0.941
Squidpy	1.114	0.429	0.907
Ours	2.107	0.316	—

Higher nominal association strength; lower intra-cluster entropy

Baseline marker compositions remain similar to ours (cosine > 0.87).

Molecular-reference validation on paired slides

Measure	All labels	A-tier reference
Adjusted Rand index	0.677	—
Normalized mutual information	0.909	0.879
Composition cosine	0.769	0.776
Dominant-lineage agreement	0.706	0.748
Matched IoU	0.348	—
Boundary F1	0.315	—

Compositional correspondence is stronger than exact geometric agreement.

Direct fixed-grid comparison reveals a coverage tradeoff

Construction	Entropy ↓	Purity ↑	Coverage
Adaptive	1.534	0.500	0.590
800-pixel grid	—	0.423	≈1.000
Matched-coverage oracle	1.523	0.507	Matched

Full-grid entropy spans 1.658–1.689 across 800-, 1,200- and 1,600-pixel grids.

Median niche size: 124 cells (adaptive) vs 128 (800-pixel grid).

Selective retention of homogeneous tissue contributes to the full-grid difference.

Resolution controls cross-slide consistency

Resolution	Pairs	Mean cosine	Median cosine	Mean JSD	Median JSD
k = 8	378	0.642	0.805	0.487	0.427
k = 16	378	0.516	0.602	0.575	0.543
k = 32	378	0.373	0.353	0.665	0.654

Lower k gives more consistent cross-slide distributions

k = 16 retains intermediate granularity for downstream analysis.

Meta-clusters remain stable under refitting

k	Test	Mean NMI	Median NMI	Mean ARI	Median ARI
8	Bootstrap	0.992	0.991	0.997	0.997
8	Random seed	0.998	0.999	0.999	0.999
16	Bootstrap	0.915	0.915	0.910	0.909
16	Random seed	0.908	0.883	0.890	0.857
32	Bootstrap	0.885	0.883	0.773	0.730
32	Random seed	0.915	0.901	0.839	0.825

At $k = 16$: median NMI = 0.915 (bootstrap) and 0.883 (random seeds).

Sensitivity to cell-classification errors

$$P(a \rightarrow b) = (1 - \alpha)1[a = b] + \alpha M_{ab}$$

$\alpha \in \{0.25, 0.50, 0.75, 1.00\}$; 5 seeds per setting; geometry and candidates fixed

Summary	Observed range
Selected LIHC marker-burden Spearman correlation	0.672–0.843
Risk-group agreement	0.755–0.838
Global candidate-rank correlation	0.085–0.230

Selected patient-level summaries are more stable than fine-grained candidate ranking.

Independent LHC validation: direction and significance

Marker	Enrichment	Discovery p	External p
Spatial M1	Poor	0.00477	0.0746
Spatial M2	Poor	0.00000459	0.120
Spatial M3	Poor	0.0261	0.132
Tissue M1	Favorable	0.0170	0.0495

The tissue marker reaches nominal $p < 0.05$; spatial markers do not.

External spatial tests: M1 log-rank; M2/M3 univariable Cox.

Adaptive niches connect morphology to spatial biomarkers

Construct

Pure seeds + gated growth + spatial–phenotypic competition

Represent

Cell-type subspaces and full tissue-composition vectors

Discover

Stable group-enrichment intervals at adaptive phenotypic scales

Next: uncertainty-aware cell typing, scalable inference and independent validation



浙江大學
ZHEJIANG UNIVERSITY



Thank you!

Zixiu Ding* · Wenxuan Hao* · Liya Ding · Yuan Dai · Jiazhen Yang · Shuwen Han · Lingxiang Jia · Zunlei Feng
Zhejiang University · Central South University · Sir Run Run Shaw Hospital · Tarim University



NeurIPS 2026, Sydney, Australia

