MANCHESTER 1824
The University of Manchester

# NexuST · Hierarchical Foundation Model for Spatial Transcriptomics

**Hongpeng Zhou**

Dame Kathleen Ollerenshaw Fellow (Lecturer) · Machine Learning and Robotics Group

Department of Computer Science, The University of Manchester

**ABCP 2026, 28 Aug 2026**

# Spatial transcriptomics (ST) measures gene expression and cell position

The University of Manchester

## What one ST slide gives you

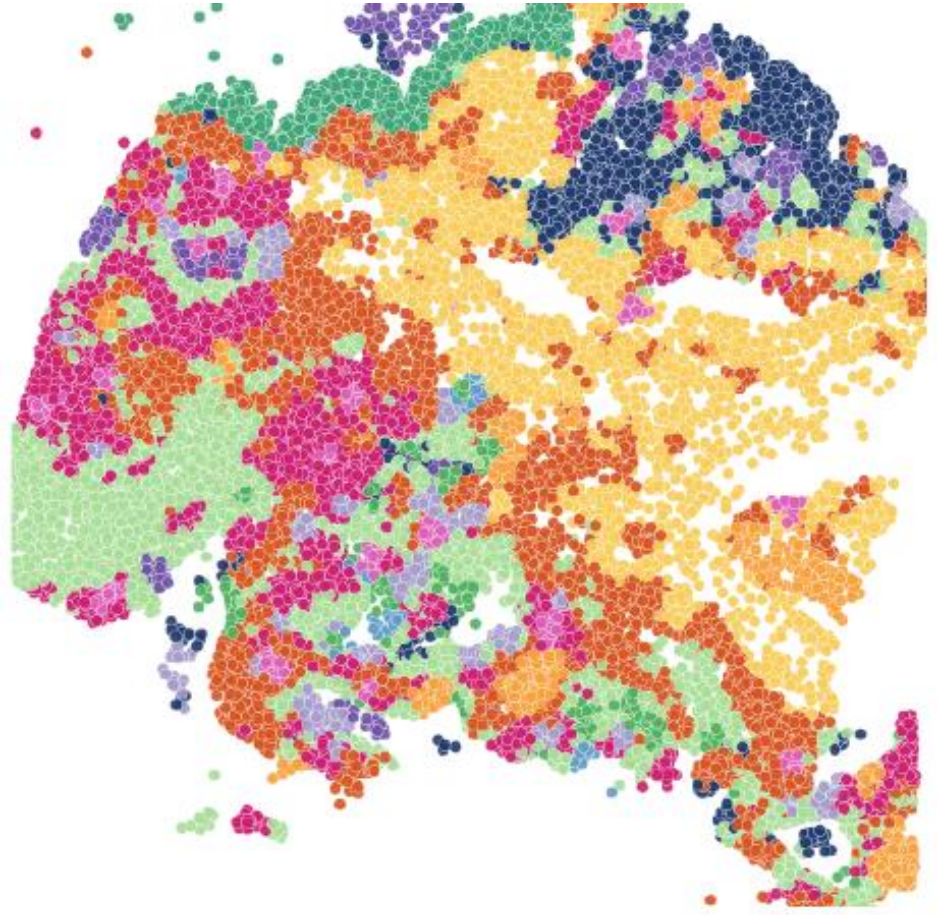
For every cell in an intact section: counts for a fixed gene panel, plus its (x, y) position. Imaging-based platforms — MERFISH, Xenium, CosMx — resolve this at single-cell level.

## ST Features:

**High dimensional:** hundreds to thousands of measured genes within cells.

**Fixed but varied panel:** A few hundred to ~1,000 genes, chosen before the experiment. Different slides measure different panels,which brings difficulty on data integration.

**Massive scale:** A single physical slide holds 5.5 thousand to 1.85 million cells. Tissue-wide modelling has to be tractable at that size.

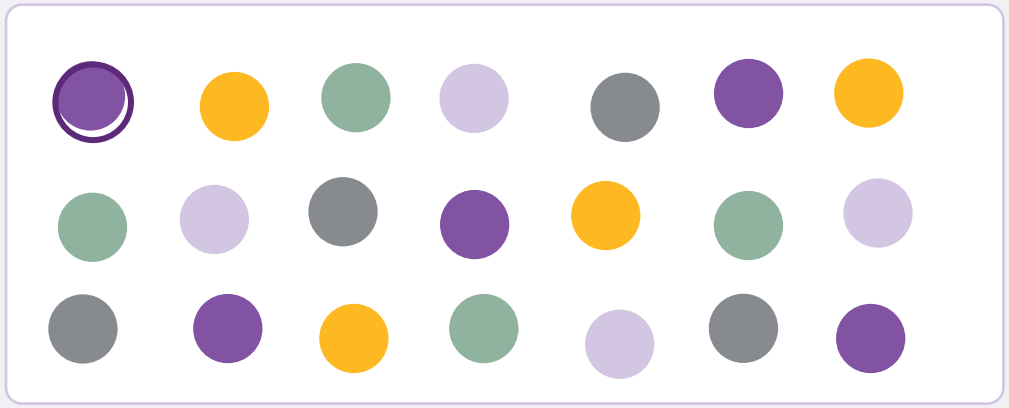


*One Xenium lung fibrosis section — every dot is a cell, coloured by its annotated histological region*

# The data is organised at two scales — and they are coupled

The University of Manchester

**Level 2 — cells within a spatial neighbourhood**



**Level 1 — genes within one cell**


## Within a cell

Gene–gene dependencies define the transcriptional programme. This is what scRNA-seq foundation models already model well.

## Across cells

Cell–cell dependencies define the microenvironment: which neighbours a cell has, and how they are arranged.

## And the two are coupled

Neighbouring cells influence a cell's expression; a cell's programme shapes how it interacts with the tissue around it. Modelling the levels separately, or one after the other, breaks that loop.

**Biology is not only inside a cell - it is also between cells.**

# Foundation models (FMs) — and why spatial data needs one

FMs are large scale deep neural network models trained by self-supervision on large-scale datasets, and can learn learn general representations adapt to diverse downstream tasks via transfer learning.

**Large unlabelled ST corpus** → **Self-supervised pretraining** → **One general cell representation**

**+ a small task head**

annotation · regions · gene recovery · niche composition

**1**

### Labels are the bottleneck

Cell-type and region annotations are expensive, expert-dependent and inconsistent between studies.

**2**

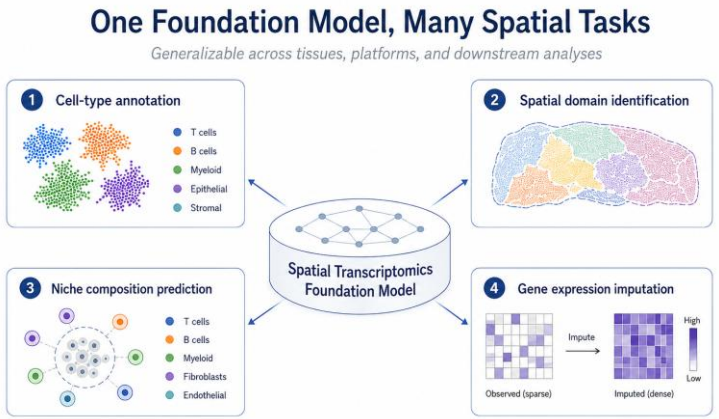
### Abundant data

Public high-resolution human ST data is growing quickly across organs, platforms and disease states.

**3**

### Generalization ability

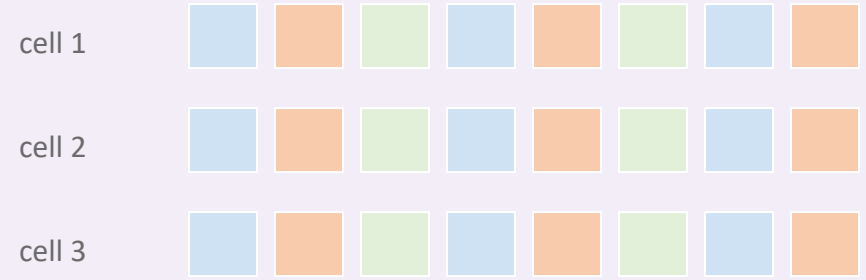
Existing spatial methods are built per dataset and per task, so every new question relearns the spatial representation from scratch



# FM model design: ST creates a gene-cell scalability tension

**The input must preserve gene relationships within cells and spatial relationships across cells.**

## A Concatenate all gene tokens




**Gains:** Keeps fine-grained gene–gene structure across the whole neighbourhood.

**Costs:** 512 cells × 300 genes ≈ 150,000 tokens in one sequence, with quadratic attention. Not tractable.

## B Compress each cell to one token first


**Gains:** Cross-cell modelling becomes cheap and scales to large neighbourhoods.

**Costs:** Explicit gene-level structure is discarded before spatial context arrives — so it can never feed back.

**The trade-off is not only computational.** Gene state and microenvironment are biologically coupled, so an architecture should let them refine one another throughout training — not model one, freeze it, and then model the other.

# How existing ST foundation models resolve it

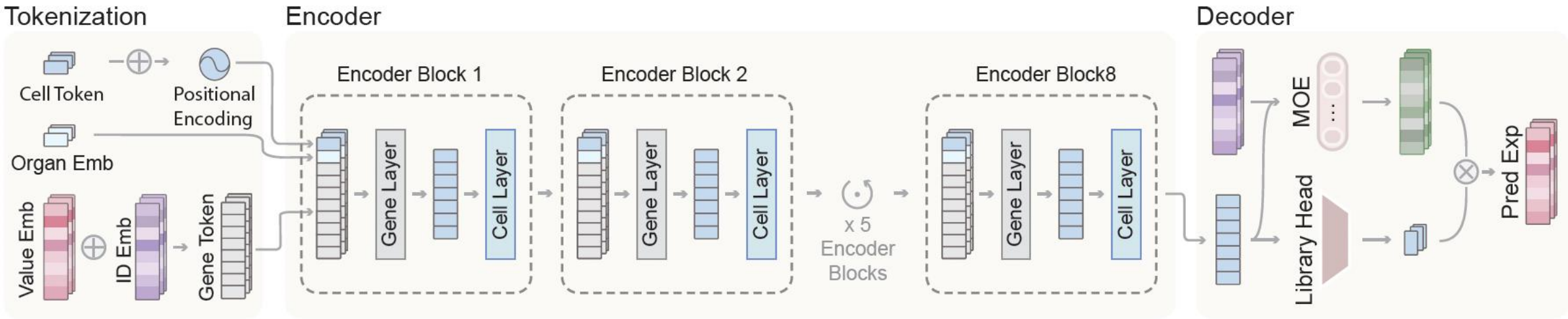
MANCHESTER 1824
The University of Manchester

Representative spatial foundation models differ in where they compress information and when they couple the two levels.

| Design | Examples | Strength | Remaining gap |
|---|---|---|---|
| **Gene-token** | Nicheformer<br>scGPT-spatial | Preserves within-cell gene dependencies | Limited direct cross-cell contextualisation |
| **Cell-token** | CellPLM<br>STPath | Scales to many cells and neighbourhoods | Explicit gene-gene structure is compressed away |
| **Coupled, constrained** | HEIST<br>BrainBeacon · SToFM | Connects gene and cell levels | Fixed graphs before training (i.e., HEIST) or sequential training (i.e., BrainBeacon, SToFM ) restrict mutual refinement |
| **NexuST** | This work | End-to-end mutual refinement | Hierarchical factorisation keeps the computation tractable |

**The missing capability is joint, scalable, end-to-end refinement of genes and cells.**

# NexuST: couple the two levels inside every encoder block




**Input construction =** 1 cell token + 1 organ token + up to 300 gene tokens

**Encoder block: gene-level attention** summarises expression within each cell, then **cell-level attention** contextualises 512 cells in tissue space to capture cell–cell interactions.

**The cell token is the interface for cross-level interaction.** A cell token is a learnable [CLS] token that summarises the cell's genes, is contextualised by neighbouring cells, then written back into its own gene sequence for the next block.

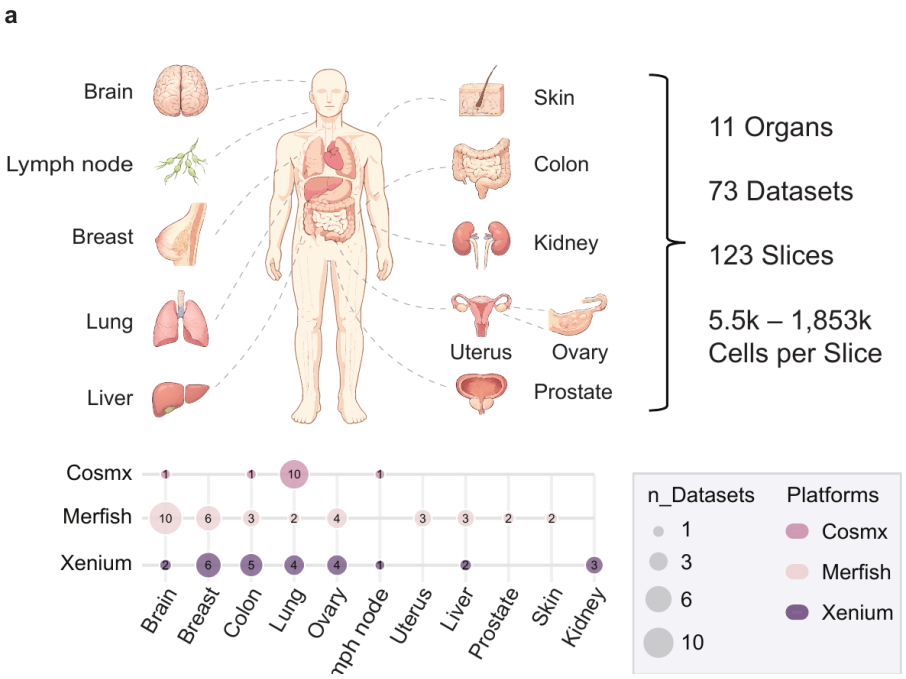
**8 hierarchical blocks · 65M parameters**

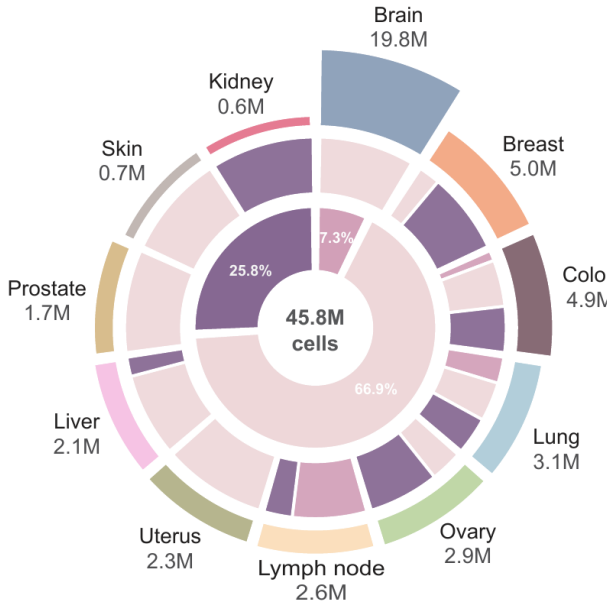
**Mutual refinement: Spatial context modulates within-cell gene representations; gene representations shape the next round of cross-cell attention.**

# HumanST-46M: the pretraining corpus

The University of Manchester

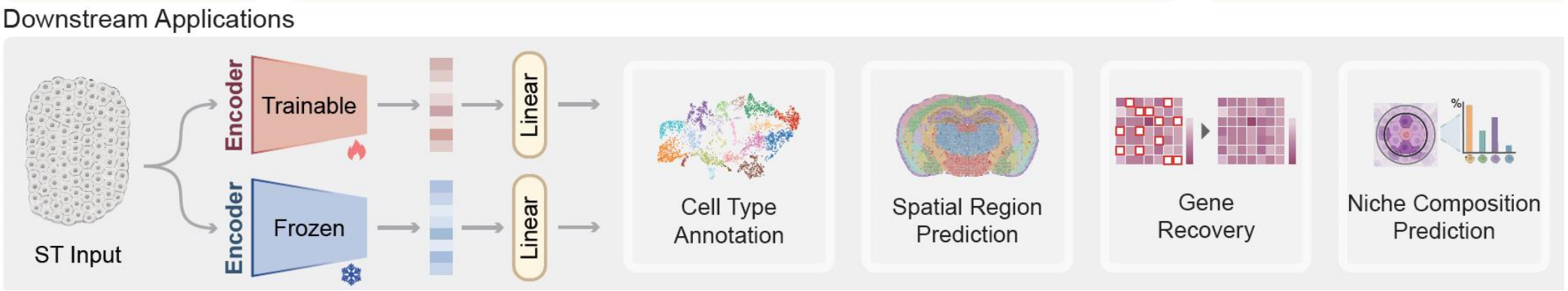
| 45.7M | 72 | 123 | 11 | 3 |
|---|---|---|---|---|
| cells | datasets | slides | organs | platforms |




**HumanST-46M supplies scale across tissues and platforms,**
**one of the largest high-resolution human spatial corpora assembled so far.**

# Held-out tasks test the pretrained representation




**≈2.6M held-out cells · 4 datasets · 3 platforms**

**Frozen linear probe**

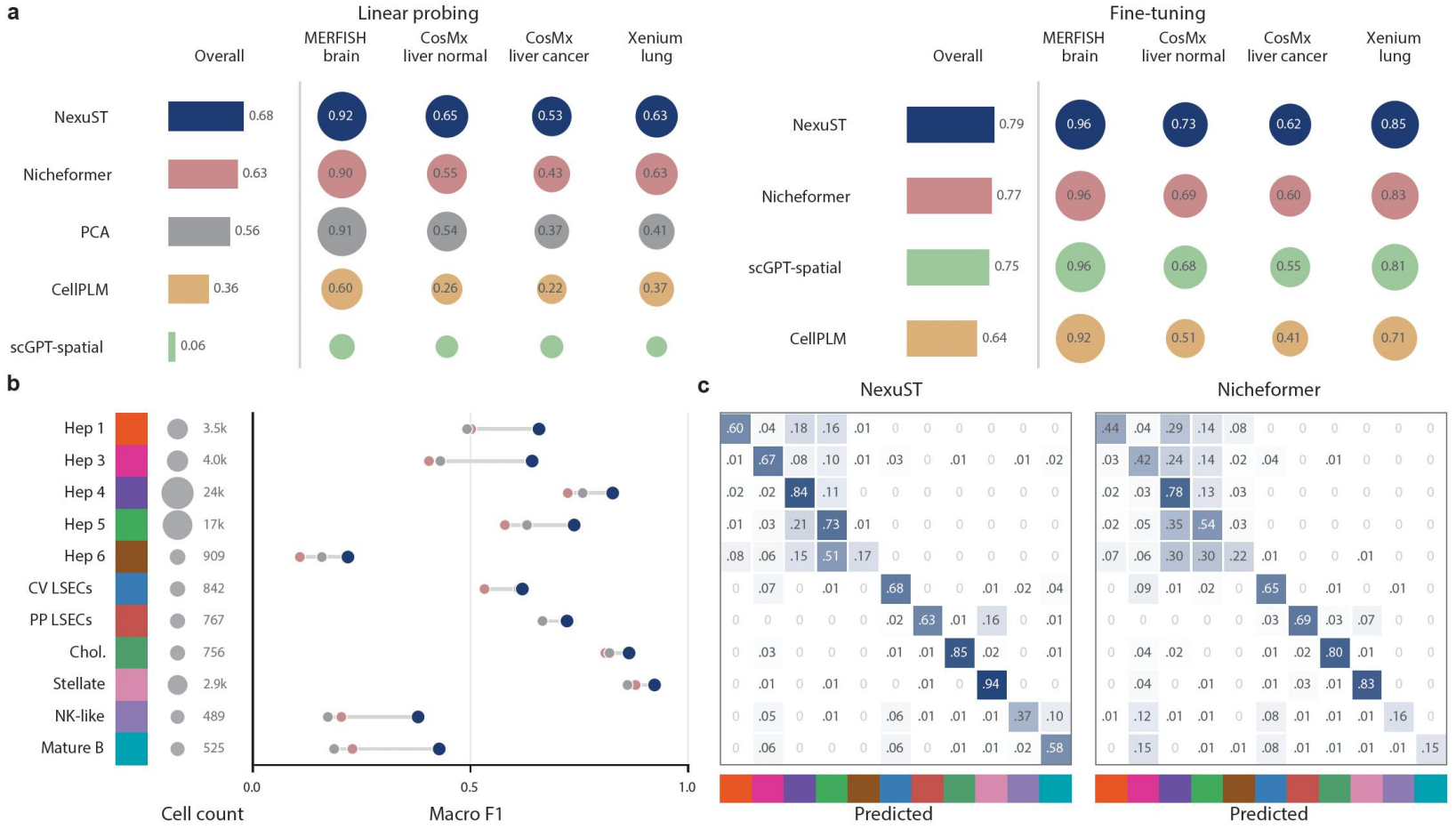
Representation quality: only the task head learns.

**Full fine-tuning**

Adaptability: encoder and task head learn together.

**4 baselines methods: 3 Foundation models** - Nicheformer, scGPT-spatial, CellPLM**, and PCA**

# Task 1 · NexuST improves cell-type annotation by resolving spatially contextual cell identities

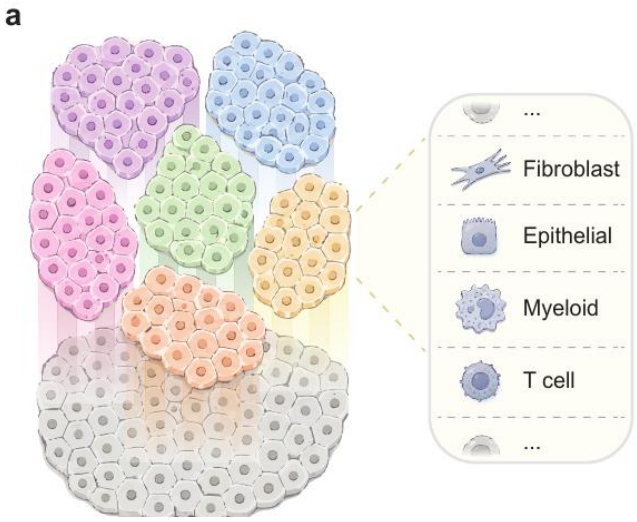



- **Highest macro F1 on all four datasets, both protocols.**
- **But the average hides the story.** On MERFISH brain, NexuST, Nicheformer and PCA are indistinguishable (all > 0.9). The margin lives in the hard cases — on CosMx liver cancer NexuST leads Nicheformer by 0.10 and PCA by 0.16.

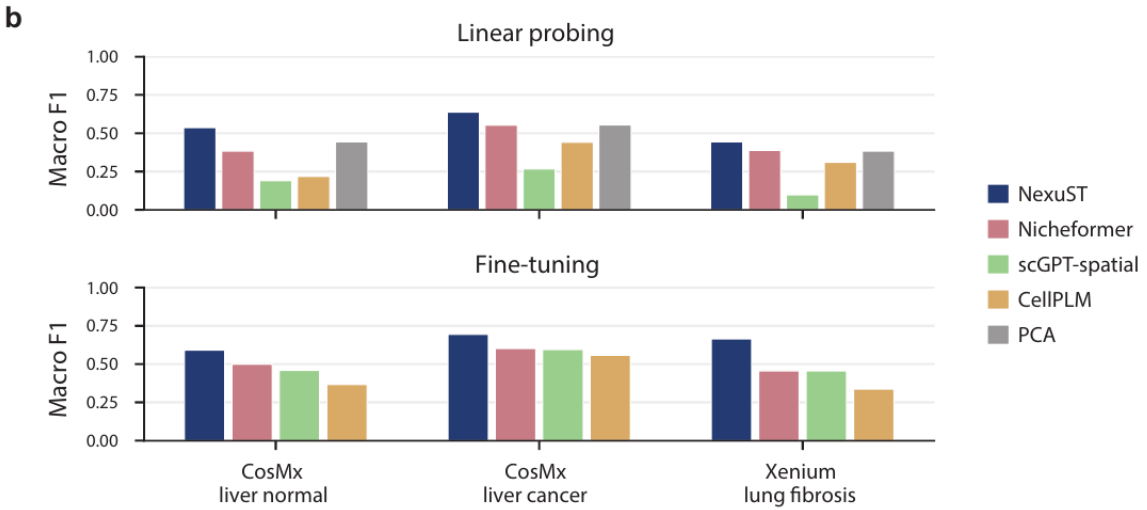
**The largest gains appear in context-dependent cell populations, including zonated hepatocyte subtypes (i.e., Hep1-6).**

# Task 2 · Region prediction: a region is context, not identity

**Why this task matters:** A tissue region is defined by the composition and arrangement of multiple cell populations - not by one dominant cell type. Region prediction requires spatial context and answers which region each cell belongs to.

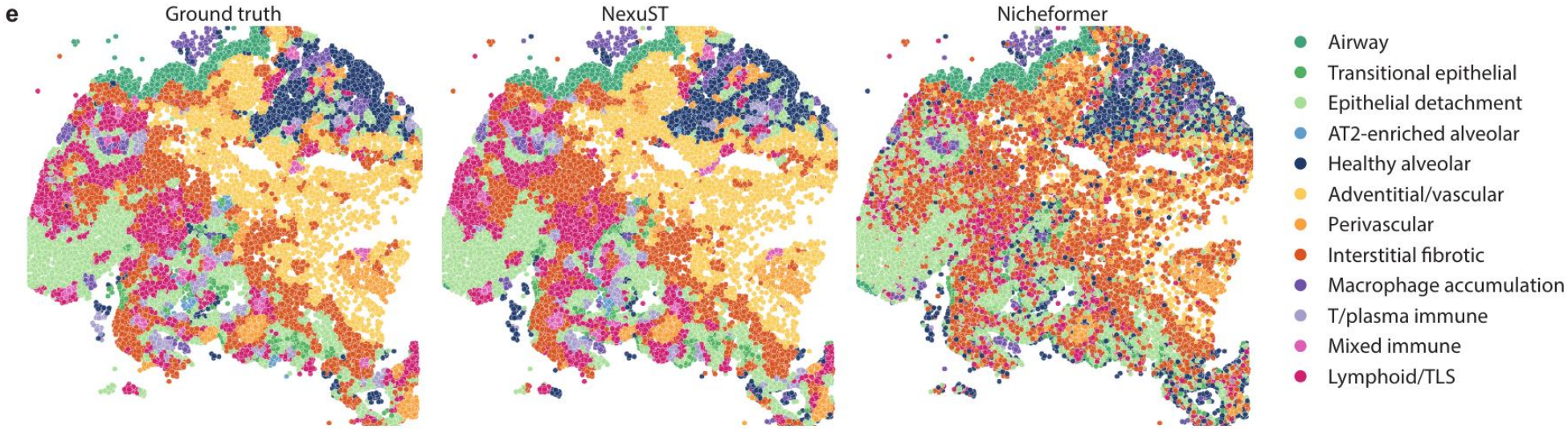



*Regions are defined by local combinations of lineages*




- **NexuST is the best under both protocols on all three annotated datasets.**
- **Frozen NexuST already exceeds every fine-tuned foundation model on both liver datasets**

# Task 2 · The hard regions are the ones defined by context




*Xenium lung fibrosis, fine-tuned: ground truth · NexuST · Nicheformer*

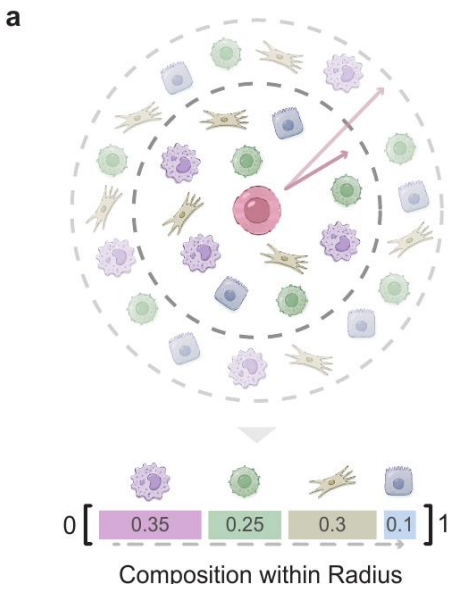
### Compositionally similar, spatially distinct

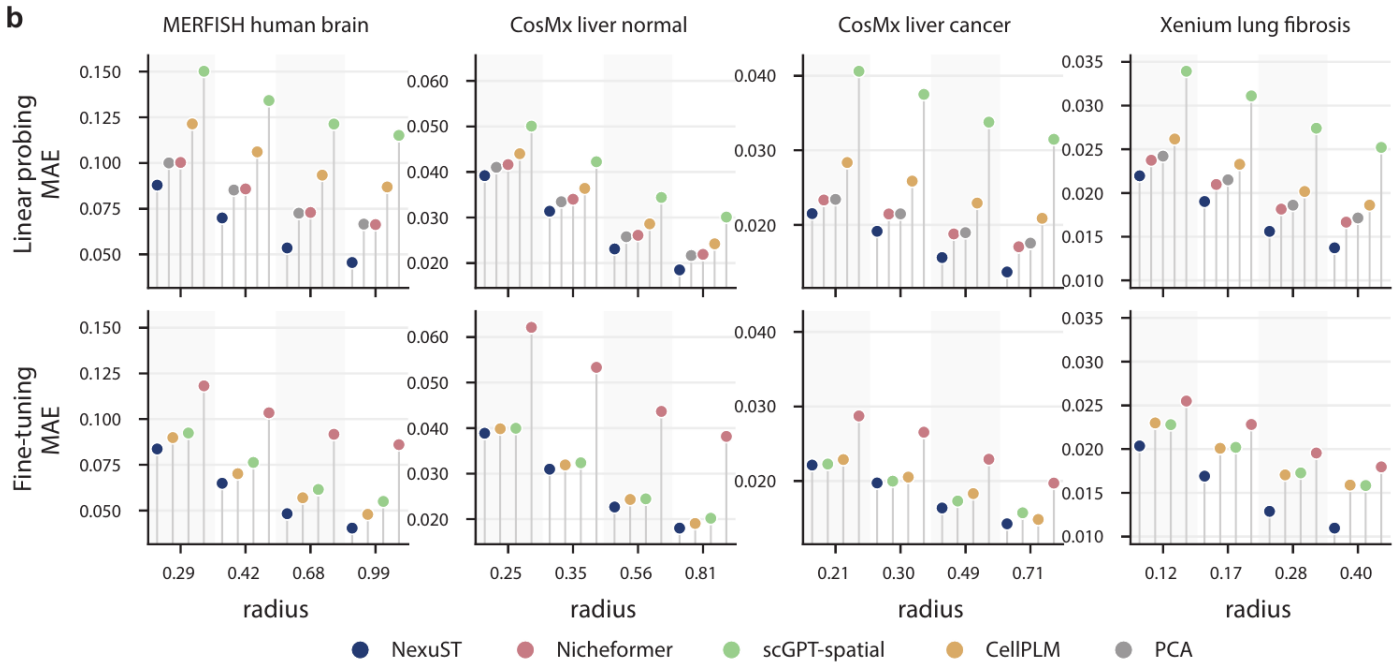
Adventitial/vascular and Interstitial fibrotic share fibroblast, endothelial and myeloid content. Nicheformer assigns adventitial/vascular to interstitial fibrotic more often than to its own label. NexuST: F1 0.27 → 0.70.

### Regions with no defining cell type

T/plasma immune and Mixed immune reflect broad infiltration, so similar immune cells appear across several regions. Baselines absorb them into surrounding compartments; NexuST recovers them, including lymphoid/TLS structures.

# Task 3 · Neighbourhood composition prediction




### The task

From one cell's representation, predict the cell-type composition of its neighbourhood within a fixed radius — four radii per dataset (≈10, 20, 50, 100 neighbours).

### The result

Lowest MAE across all four datasets, all four scales and both protocols. Under linear probing, the only model that consistently beats PCA.

### A detail

We use proportions rather than the softmax transform used previously — softmax compresses exactly the minority cell types the niche is defined by.

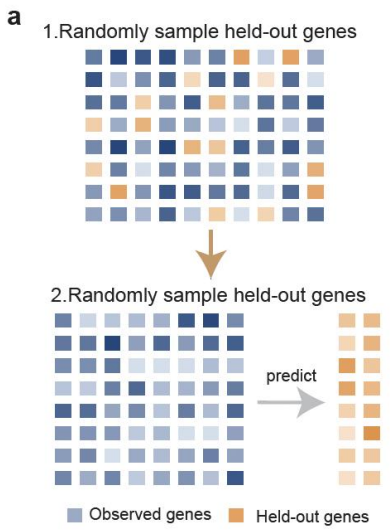
# Task 4: held-out gene recovery

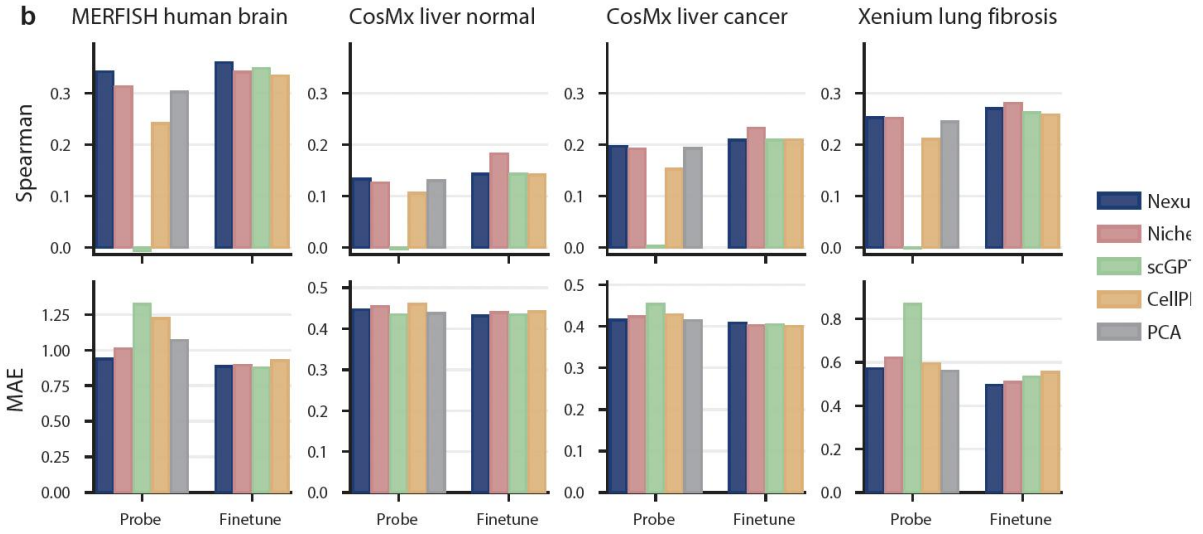
## Held-out gene recovery

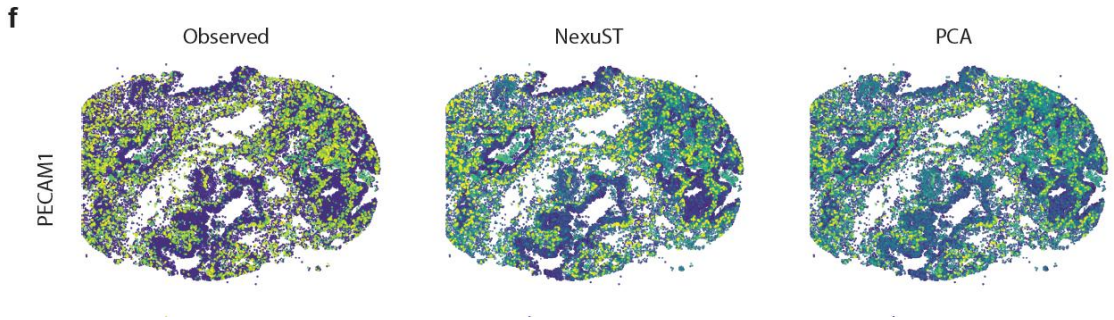
Hide 20% of the panel, predict it from the remaining 80% — panels are fixed, so relevant genes are often simply absent.

**Best Spearman and MAE on all four datasets under linear probing**, with the largest gains for spatially organised genes.

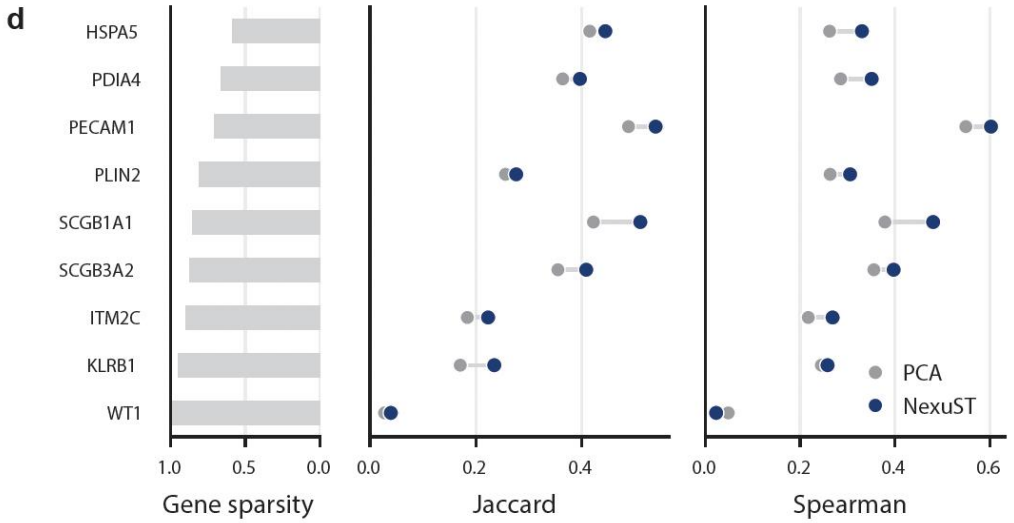
**After fine-tuning,** Nicheformer overtakes us on the two CosMx liver sets, where the targets are unusually sparse.




*PECAM1 · SCGB3A2 · SCGB1A1 — observed, NexuST, PCA*



# Take away

1

### The architecture works

Coupling gene- and cell-level representations inside every encoder block — not before or after — gives one end-to-end model that is best or competitive across four tasks and four held-out datasets, spanning three platforms and three organs.

2

### Spatial modelling pays when identity is ambiguous

$\rho = 0.85$ between the gain and how spatially dispersed a cell type is; hepatocyte zonation along the lobule; tumour-poor vs tumour-rich niches of the same cell type. A usable rule of thumb for whether a spatial model is worth its cost.

3

### PCA is strong — and that is a finding about our benchmarks

Much of the apparent spatial signal is already carried by a cell's own transcriptome. Spatial foundation models should be judged on the margin over an expression-only control, not on absolute scores.

# Thank you

**NexuST · a hierarchical foundation model for spatial transcriptomics**

Haiping Liu, Qian Zhao, Wenhao Cai, Lijing Lin, Jingyuan Sun, Zhiyong Zou, Yuxi Zhou,
Mauricio A. Álvarez, Andrew Gilmore, Magnus Rattray, Alejandro F. Frangi, Hongpeng Zhou

Computer Science · Health Sciences · Medical Sciences, The University of Manchester

**hongpeng.zhou@manchester.ac.uk**