

Scalable single-cell multiomics and AI bridge cellular complexity to clinics

JIANG Lan 蒋岚

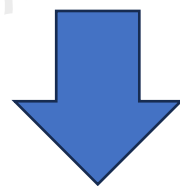
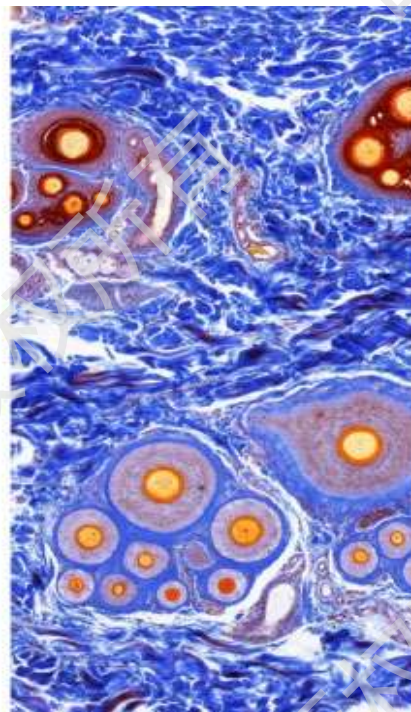
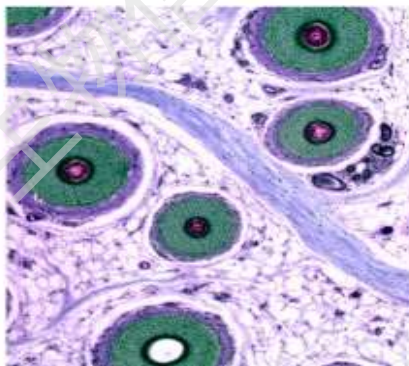
Email: jiangl@cncb.ac.cn

China National Center For Bioinformation (CNCB)

Beijing Institute of Genomics, Chinese Academy of Sciences

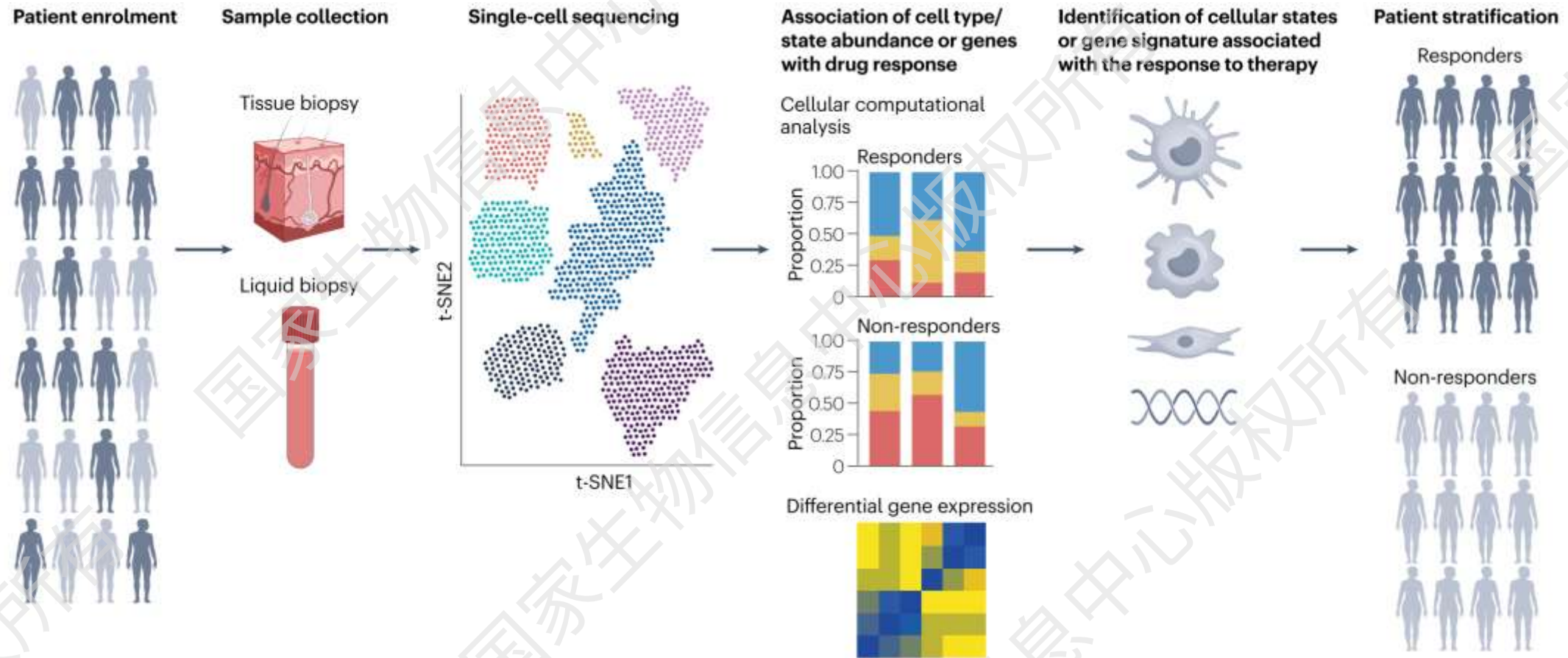


HUMAN CELL ATLAS



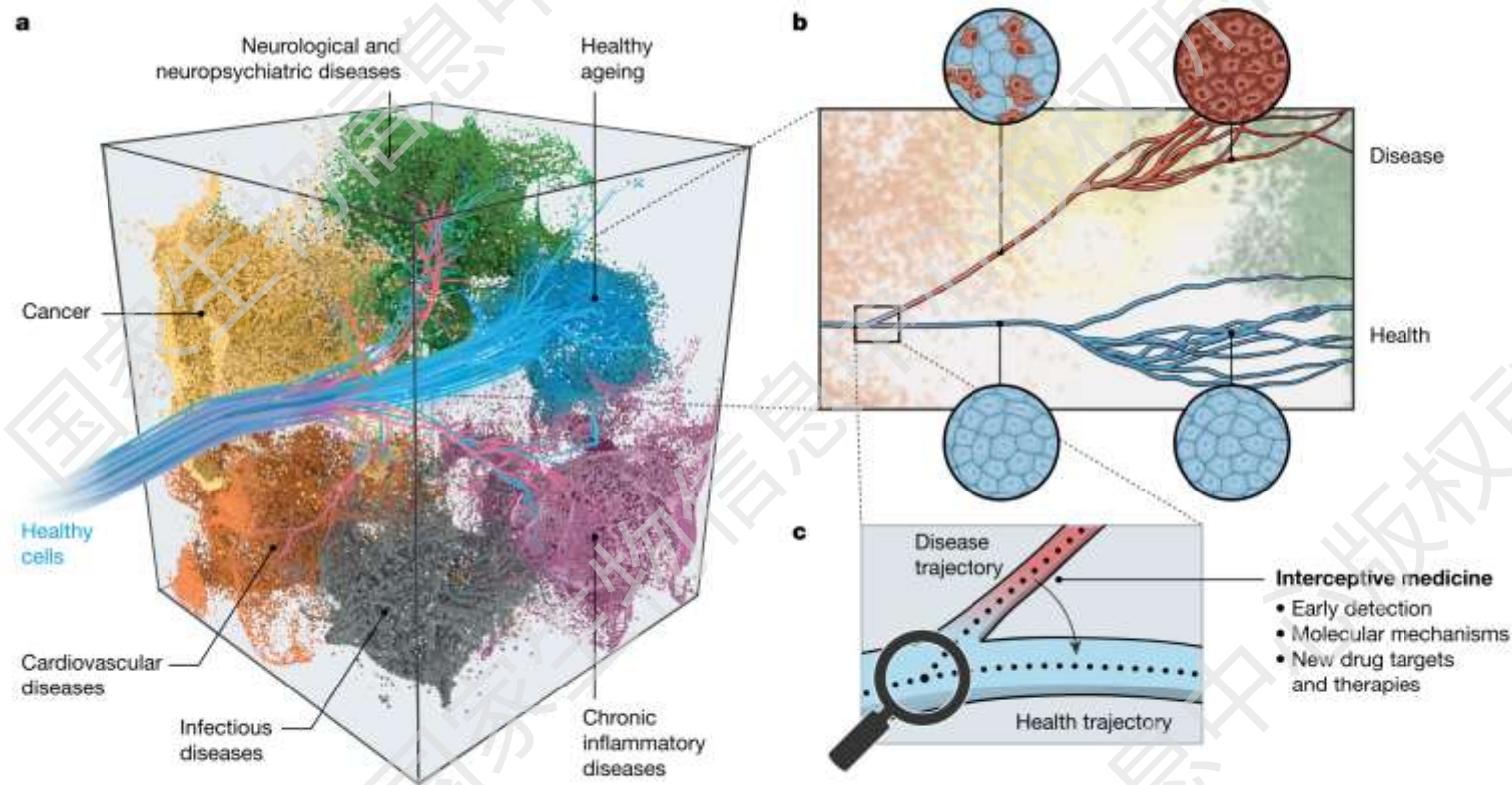
Personalized Cell Atlas 人人细胞图谱

Patient stratification based on subpopulations/niches



Van de Sande B, Lee JS, Mutasa-Gottgens E, Naughton B, Bacon W, Manning J, Wang Y, Pollard J, Mendez M, Hill J, Kumar N, Cao X, Chen X, Khaladkar M, Wen J, Leach A, Ferran E: *Applications of single-cell RNA sequencing in drug discovery and development*. *Nature Reviews Drug Discovery* 2023, **22**:496–520.

Investigating the Origins of Disease and Promoting Early Detection and Intervention



LifeTime

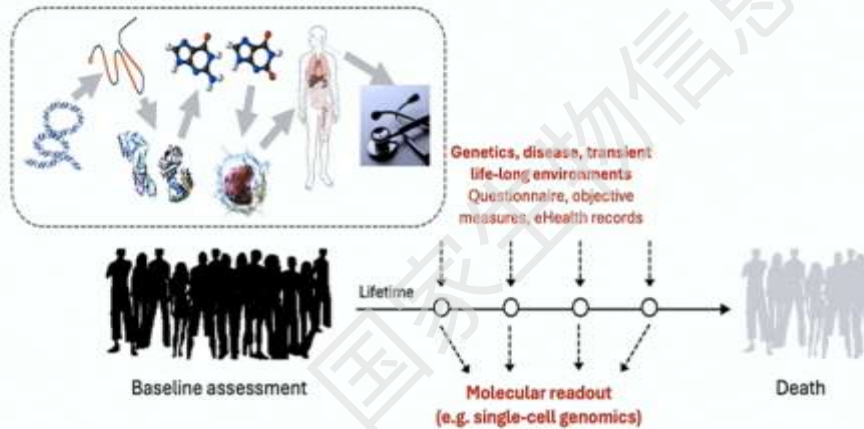
Fig. 1 | Early disease detection and interception by understanding and targeting cellular trajectories through time. a, Cells are programmed to

(red cells). At this point, cells, tissues and organs have undergone extensive and often irreversible molecular and physiological changes since the initial events

Rajewsky N, Almouzni G, Gorski SA, Aerts S, Amit I, Bertero MG, Bock C, Bredenoord AL, Cavalli G, Chiocca S, Clevers H, De Strooper B, Eggert A, Ellenberg J, Fernández XM, Figlerowicz M, Gasser SM, Hubner N, Kjems J, Knoblich JA, Krabbe G, Lichter P, Linnarsson S, Marine JC, Marioni JC, Marti-Renom MA, Netea MG, Nickel D, Nollmann M, Novak HR, et al.: *LifeTime and improving European healthcare through cell-based interceptive medicine*. Nature 2020, **587**:377–386.

New trends: Single-Cell Sequencing to Population Levels

Value of deep phenotyping in population cohorts

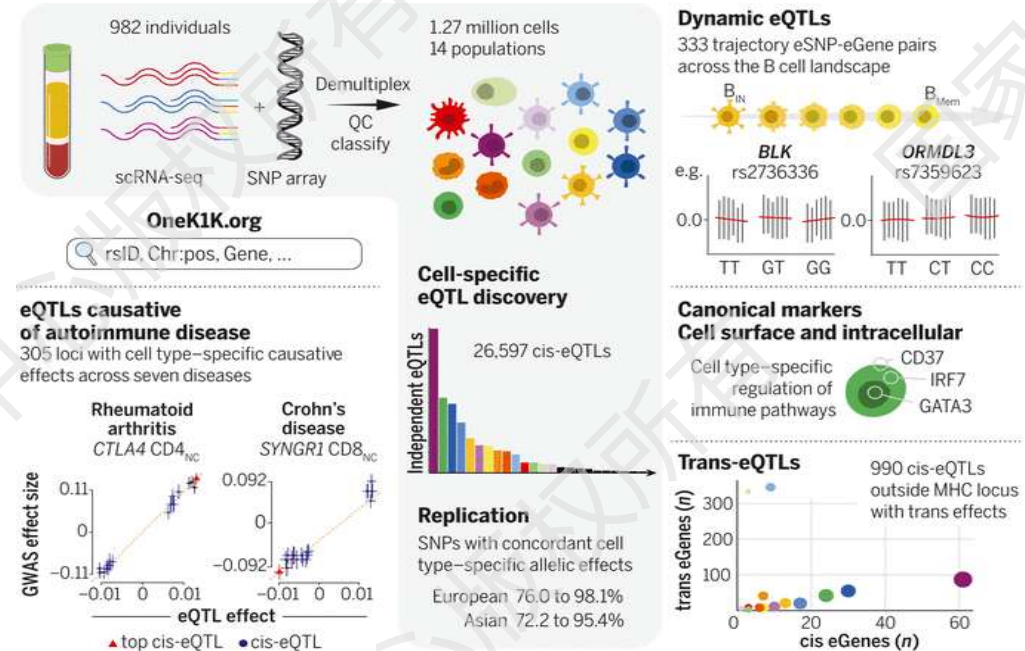


1. Locus discovery and genetic effect triangulation
2. Prediction of long-term disease risk from baseline data
3. Assess perturbation from population-level data



UK BioBank, Genes&Health cohorts
~10000 individuals.

Nicole Soranzo
Human Technopole

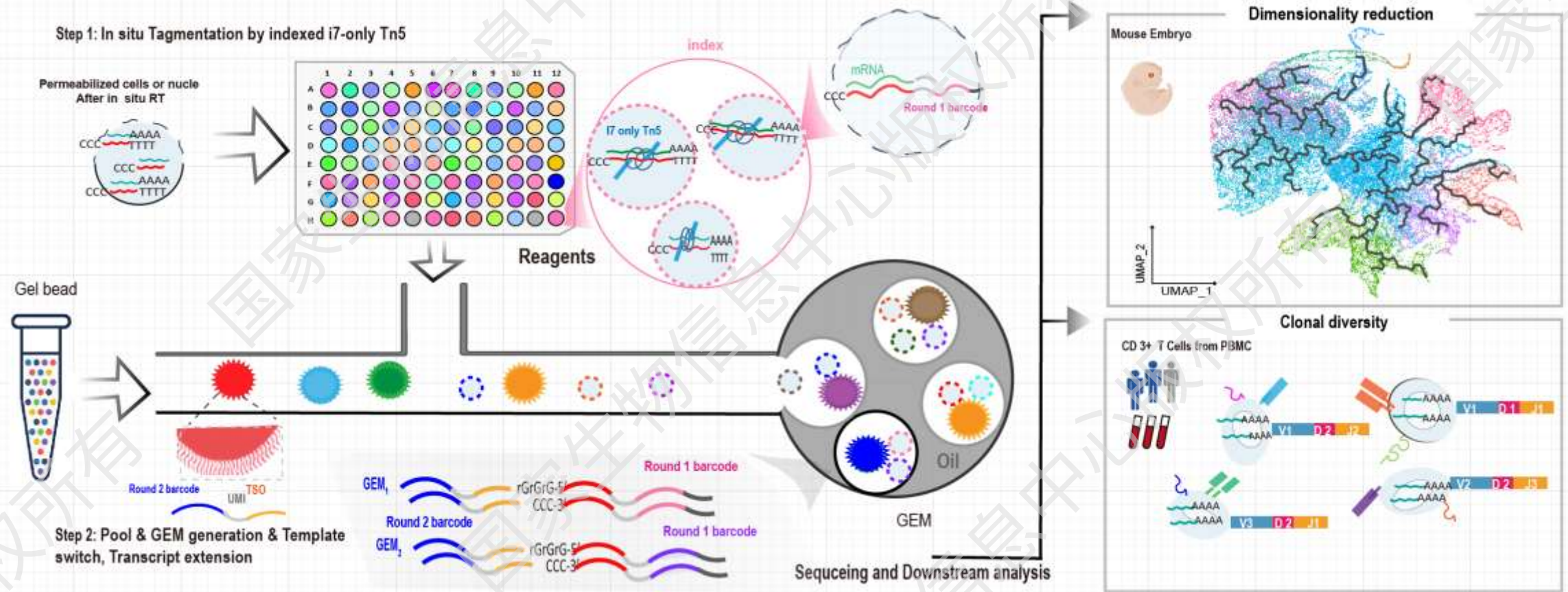


TenK10K
\$27 million

Joseph Powell
Garvan-Weizmann Centre

Droplet microfluidics+ Combinatorial Indexing

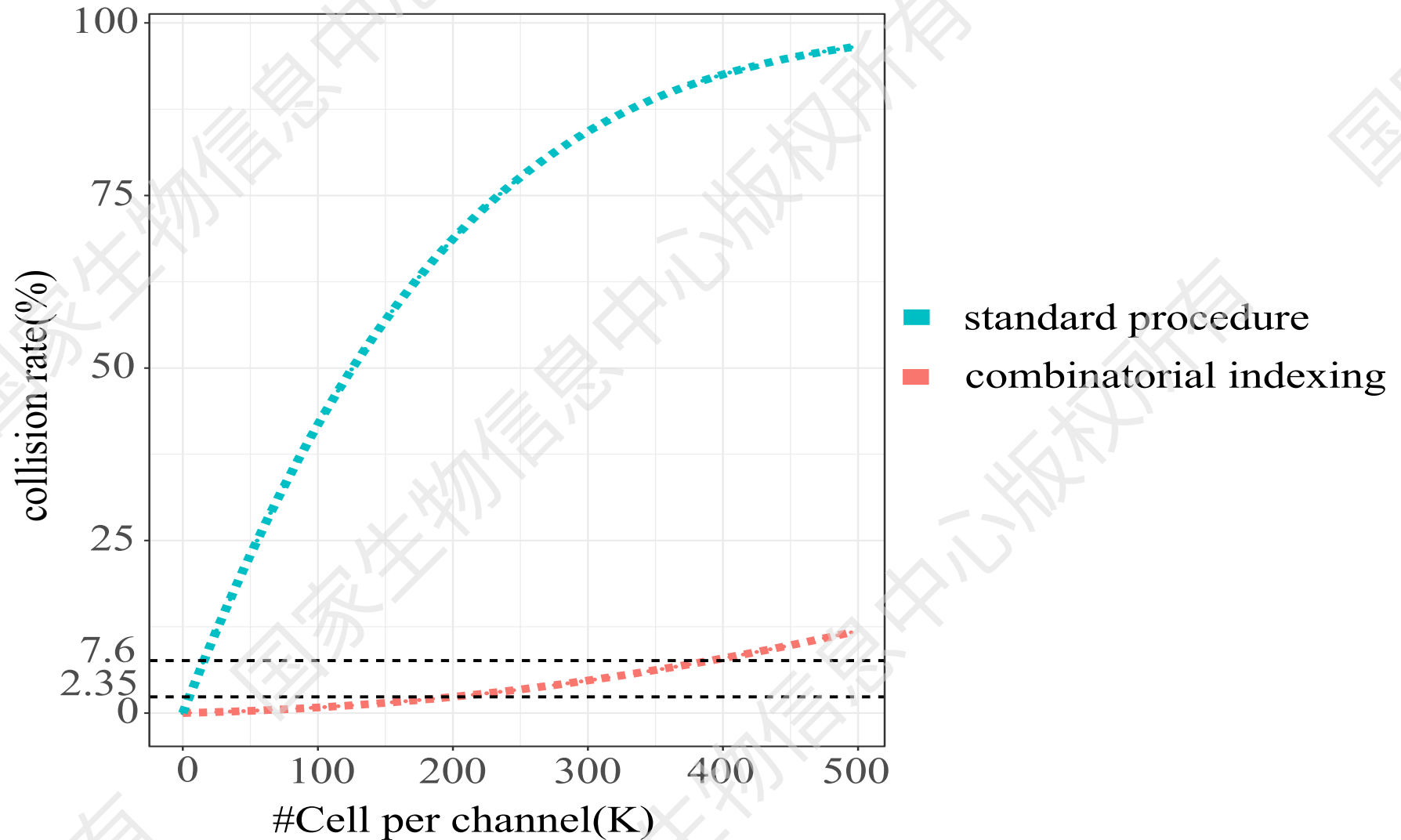
FIPRESCI Five PPrime End Single-cell Combinatorial Indexing **RNA** sequencing



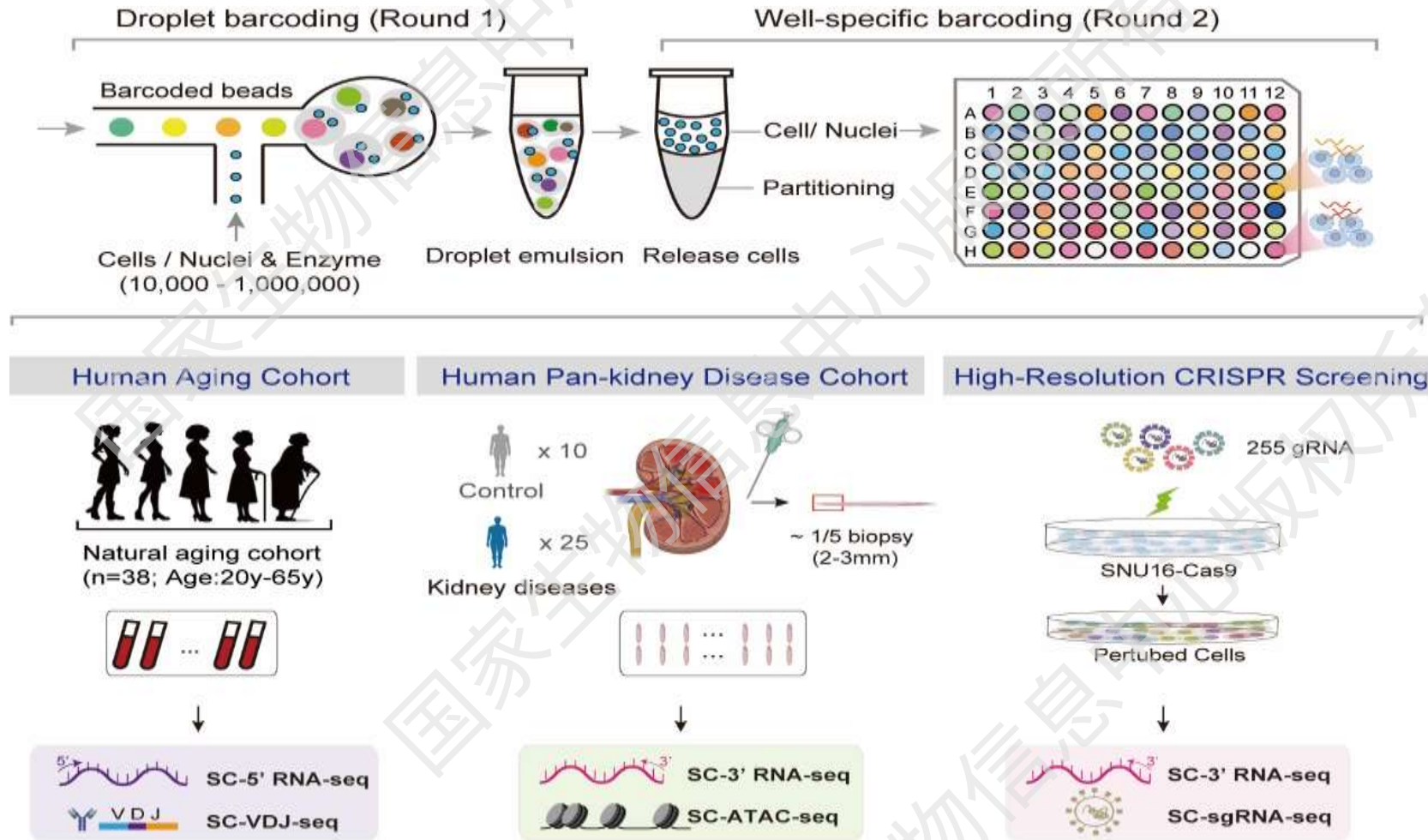
100 K cell/nuclei per channel

Y Li, ... , L, Jiang, Genome Biology, 2023

Decoupling throughput and error rate via dual-round combinatorial indexing



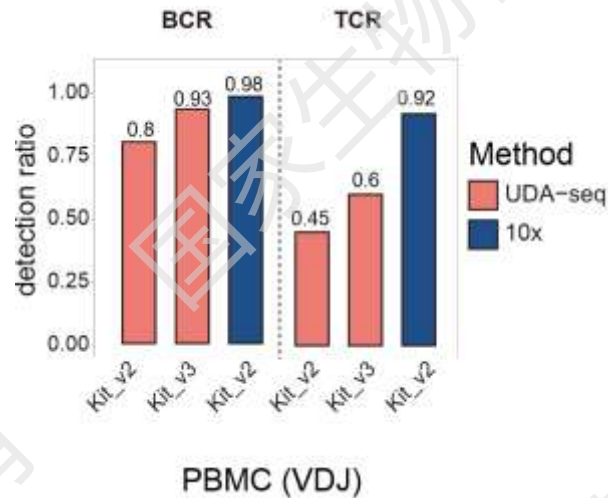
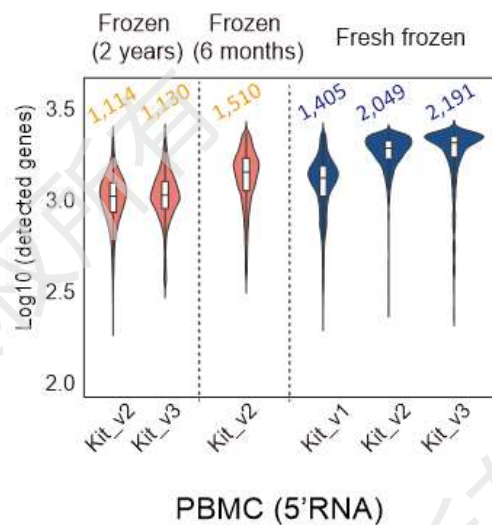
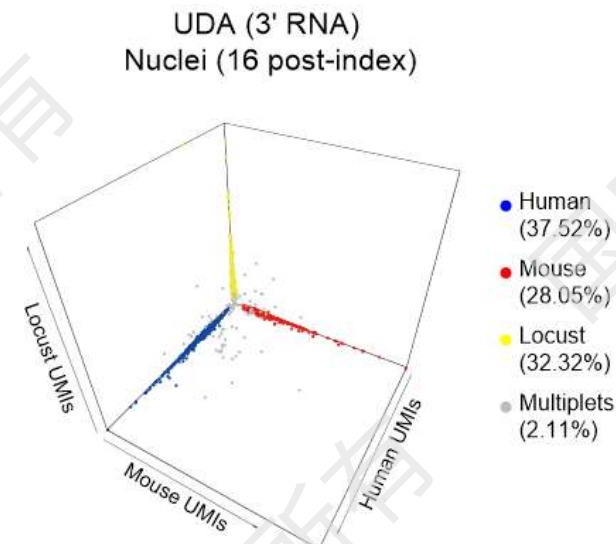
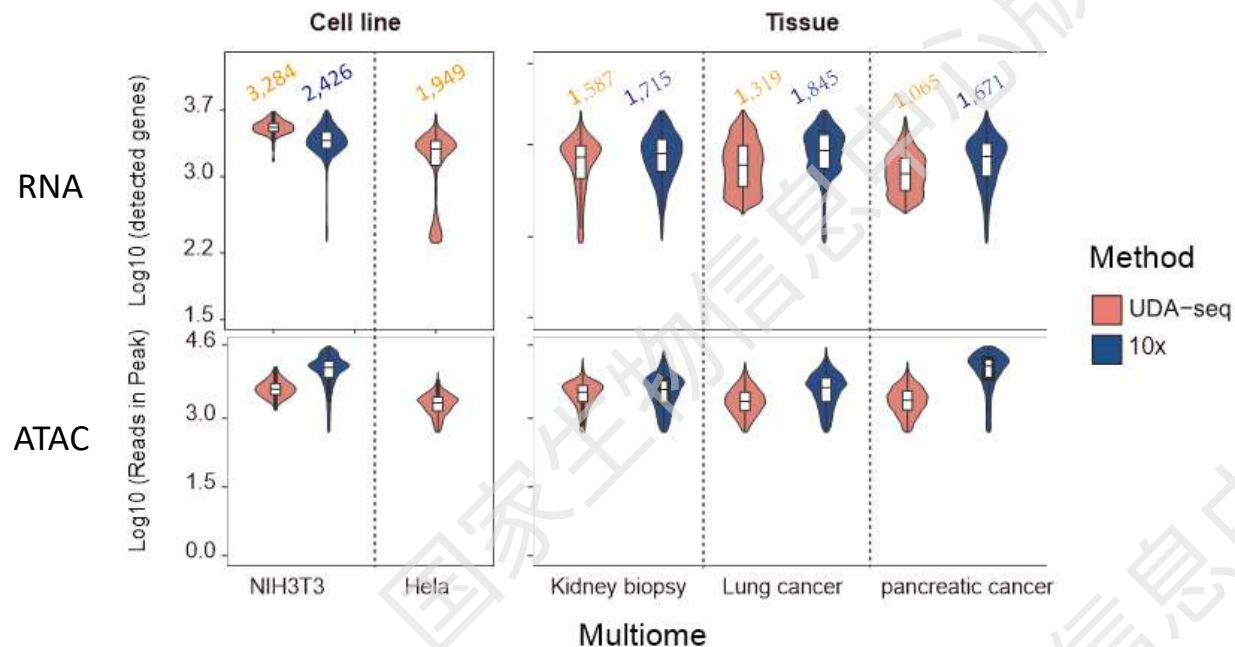
UDA-seq: universal droplet microfluidics based combinatorial indexing for massive-scale single-cell **multimodal** sequencing



100 K cell/nuclei per channel

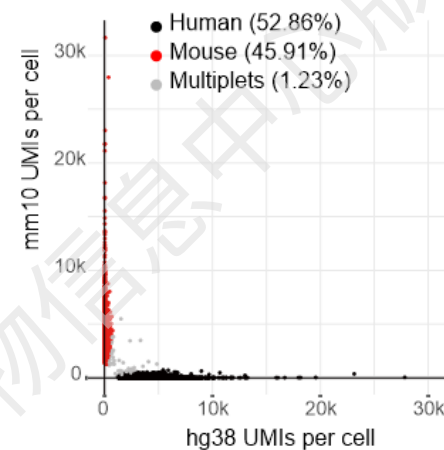
Y Li, ..., L, Jiang, *Nature Methods*, 2025 8

UDA-seq generate high quality single-cell data



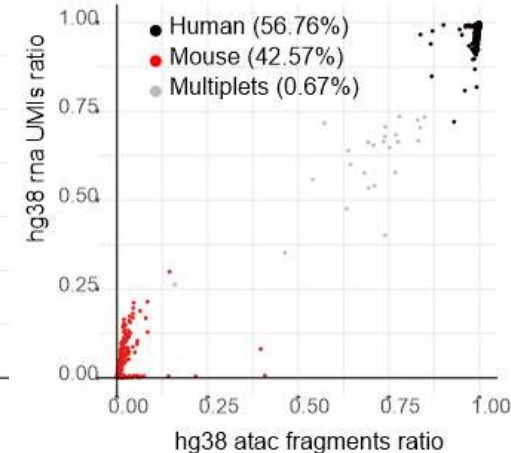
UDA (3' RNA)

Whole cell (96 post-index)

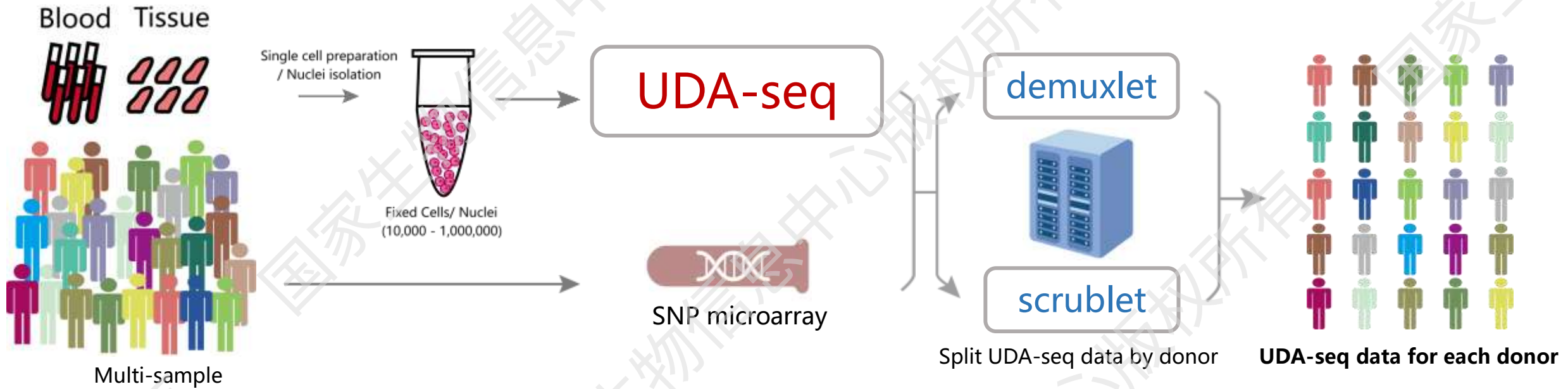


UDA (Multiome)

Nuclei (16 post-index)

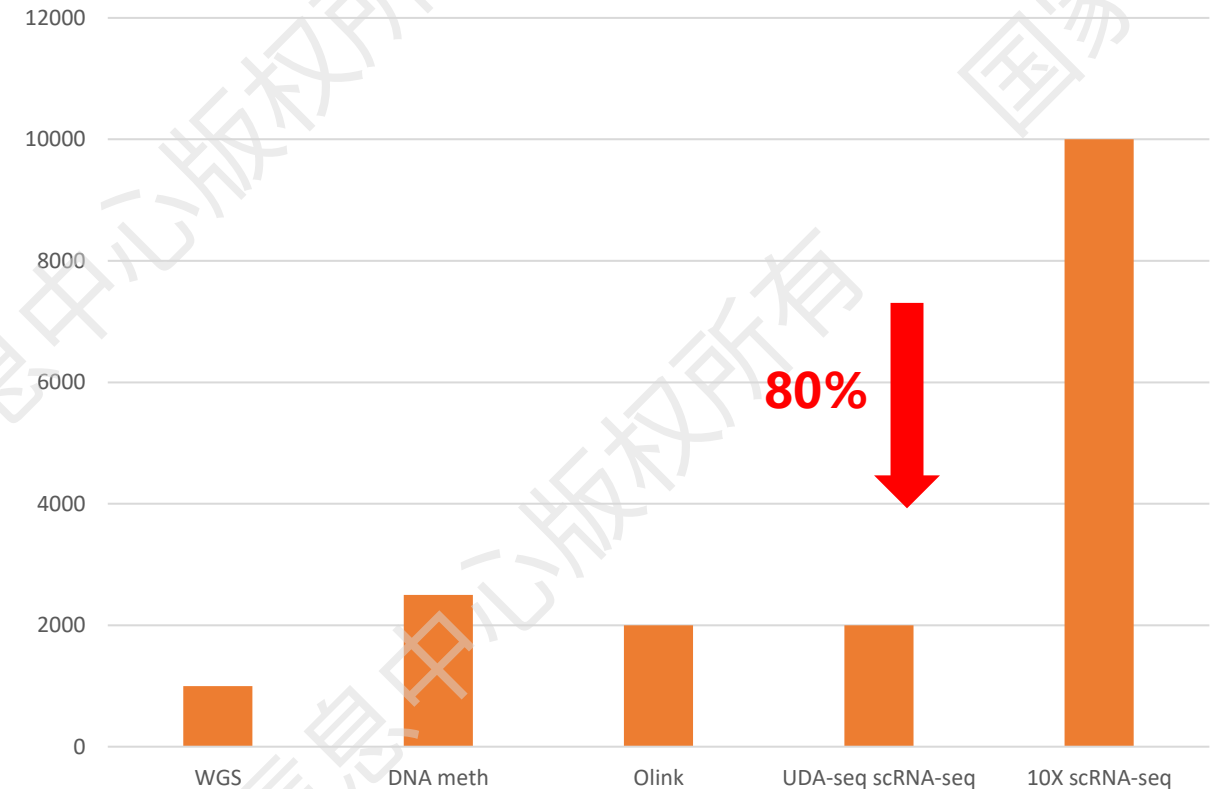
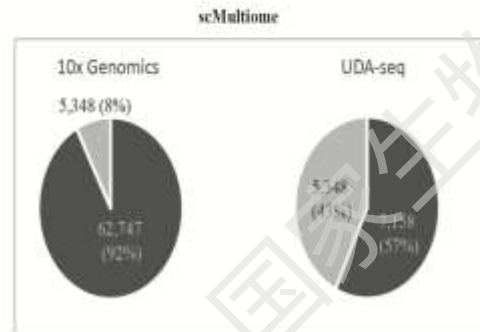
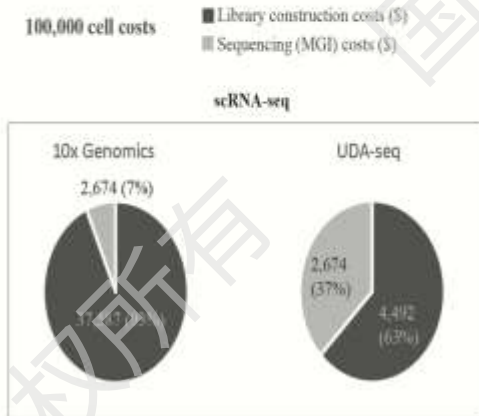
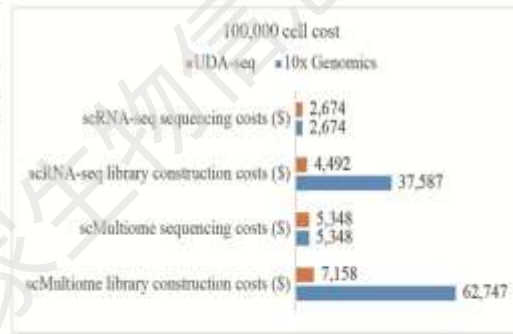


Sample multiplexing with Genetic variant



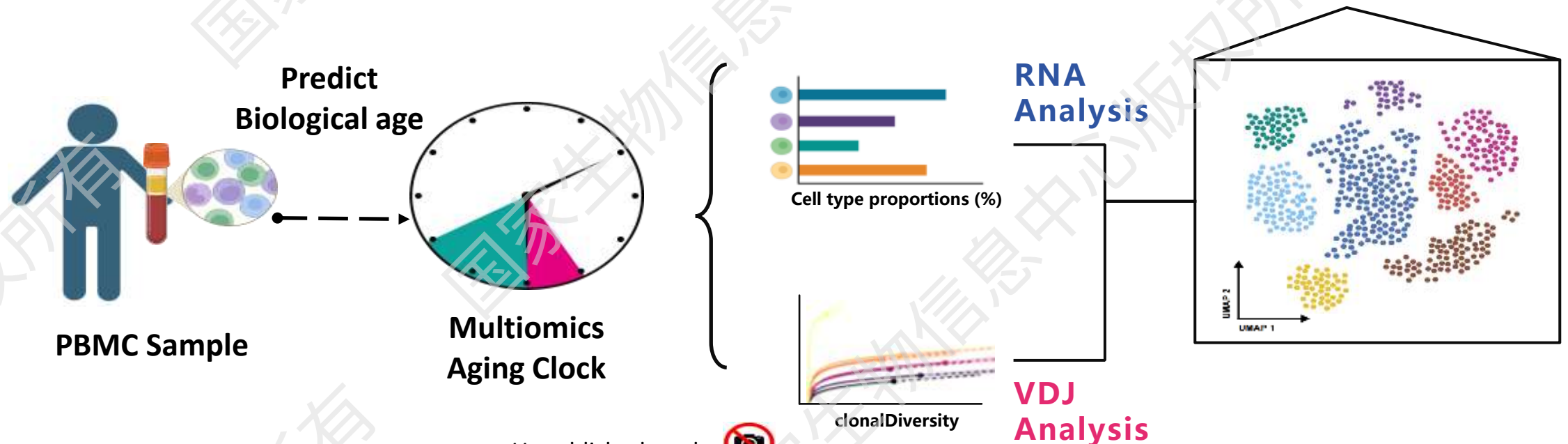
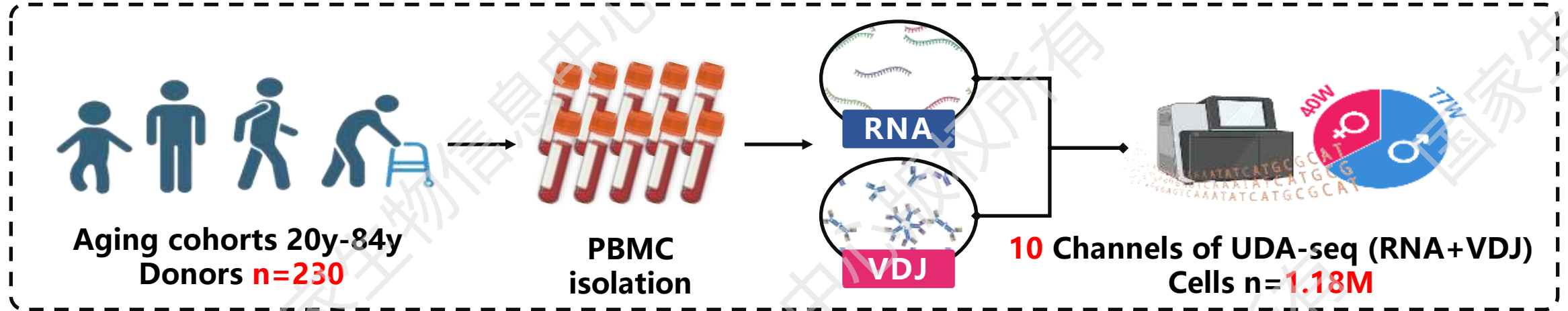
Disruptive technology dramatically slashes costs for large-scale projects.

Summary (100,000 cell cost)		
	10x Genomics	UDA-seq
scRNA-seq library construction costs (\$)	37,587	4,492
scRNA-seq sequencing costs (\$)	2,674	2,674
scMultiome library construction costs (\$)	62,747	7,158
scMultiome sequencing costs (\$)	5,348	5,348



Multimodality, 10K individuals, 50M Cells, 3M \$

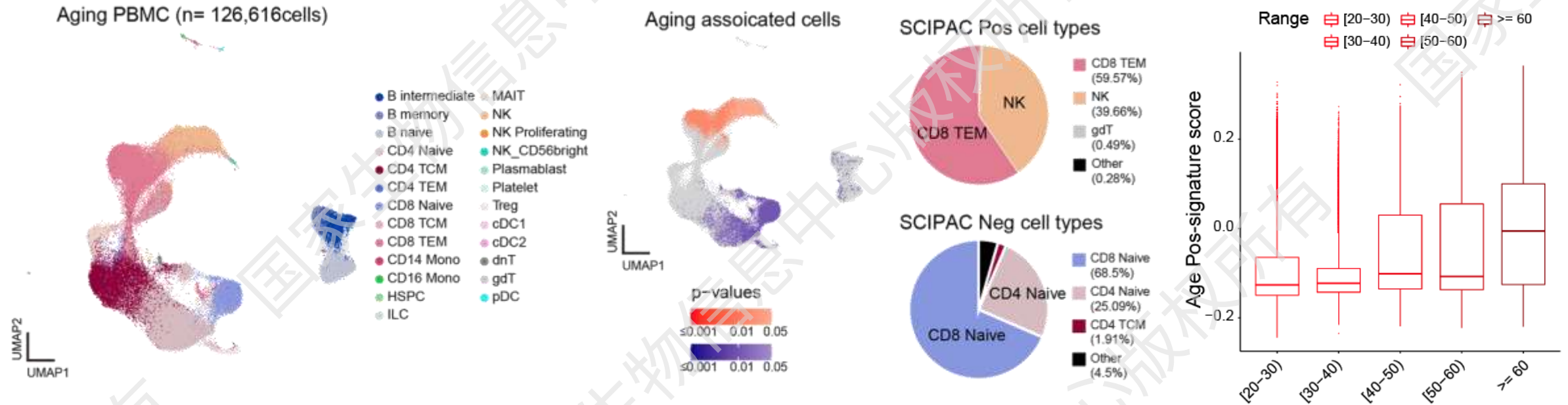
Aging Immunity Single Cell Atlas Integrating RNA and VDJ Profiling



Unpublished work.

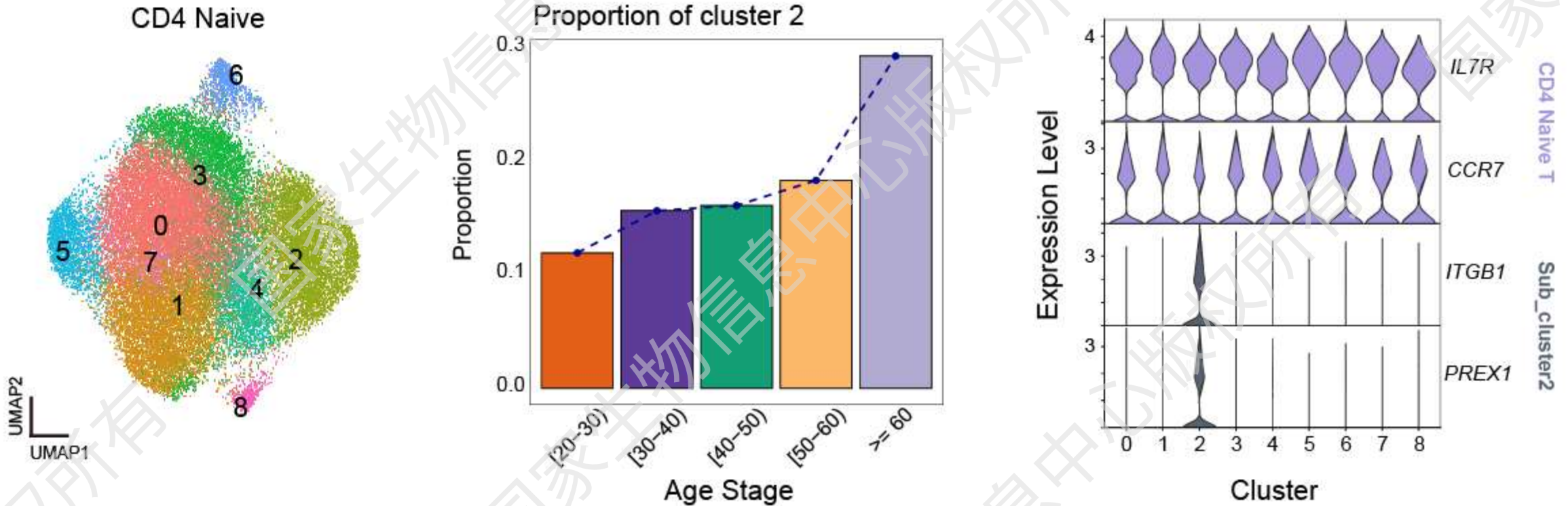


Aging associated cell subpopulations



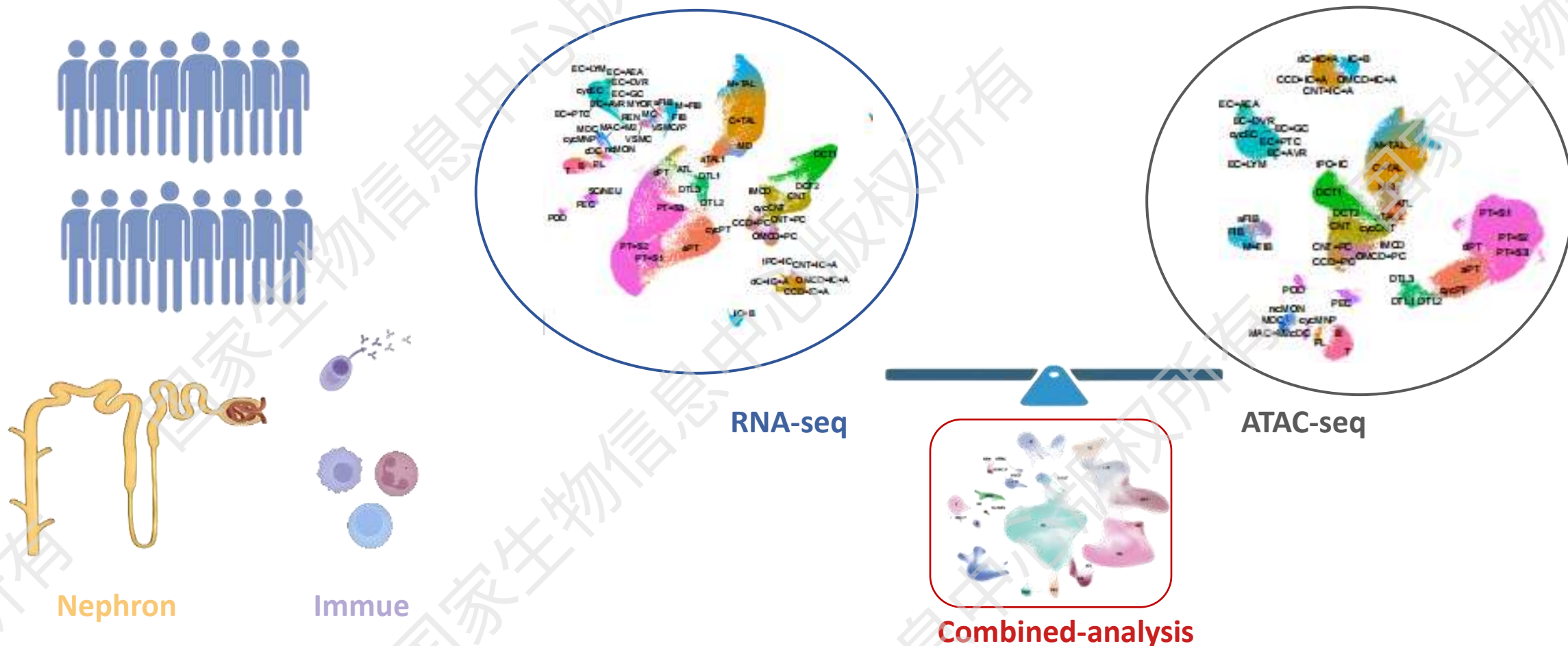
In collocations with Profs. Weiqi Zhang, Guang-hui Liu, et al

ITGB1+PREX1Naive CD4+T subpopulation



Highlight: discovery of maker of key subpopulation

Multimodal Single-Cell Atlas of Pan-Renal Pathologies

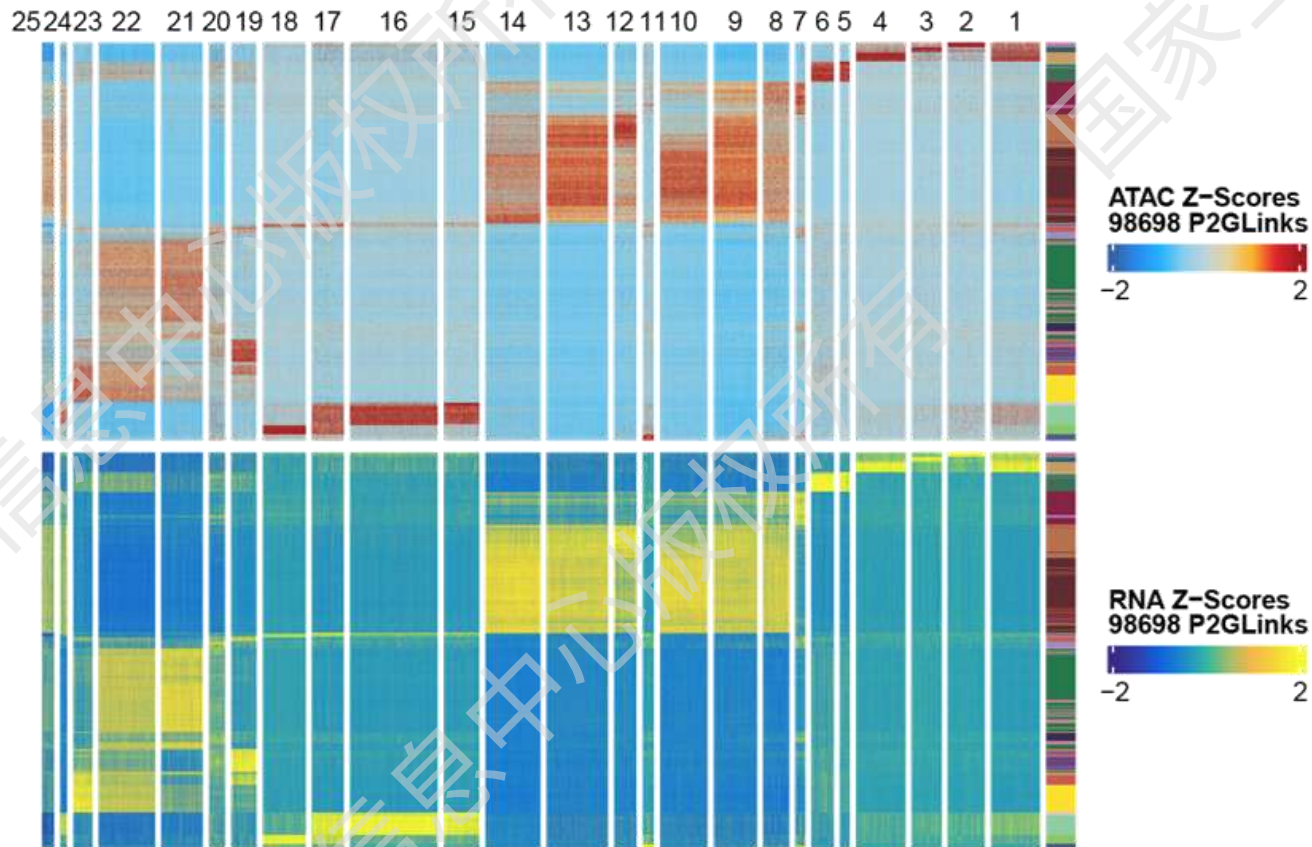
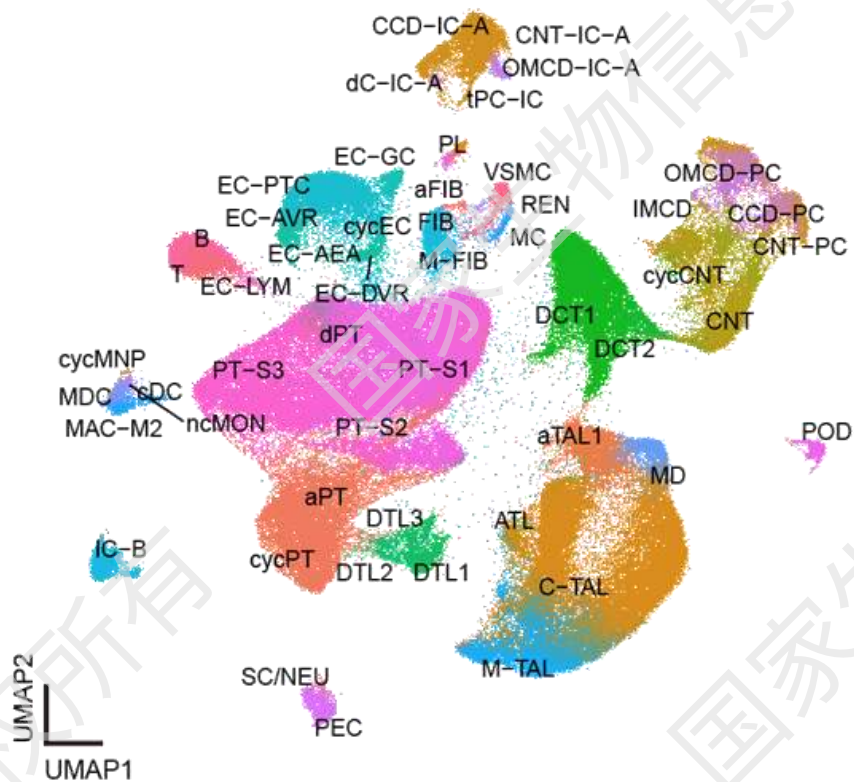


Diagnostic Categories n=16
Donors n=213

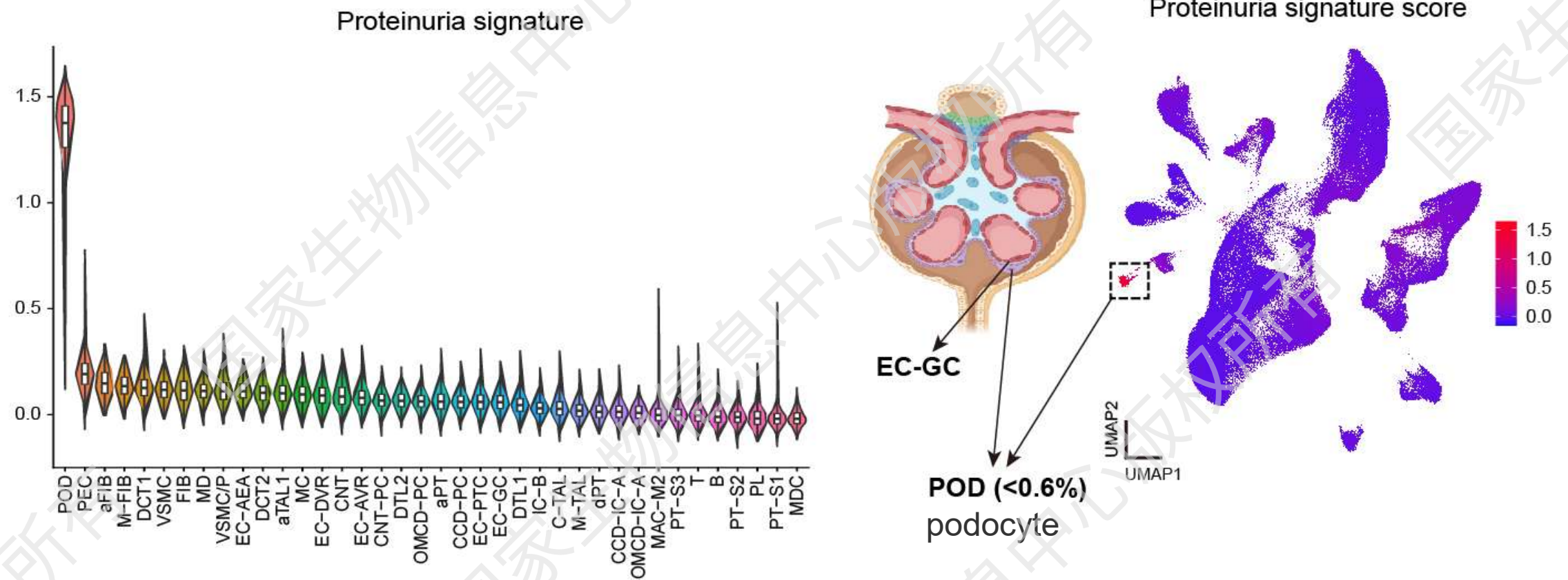
10 Channels of UDA-multiome
~5000 cells/donor, cells n=1M

Single-Cell Transcriptome and Chromatin Accessibility in Human Kidney Diseases

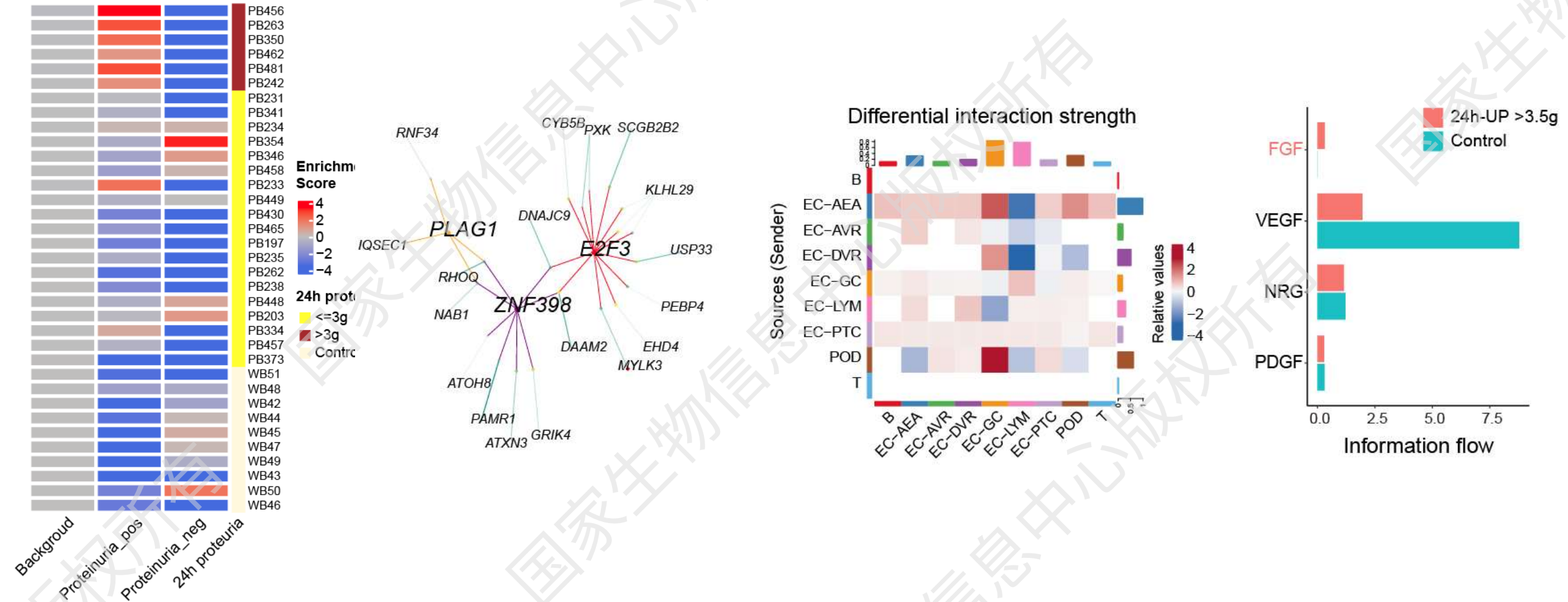
ATAC+RNA profile (207,789 cells)



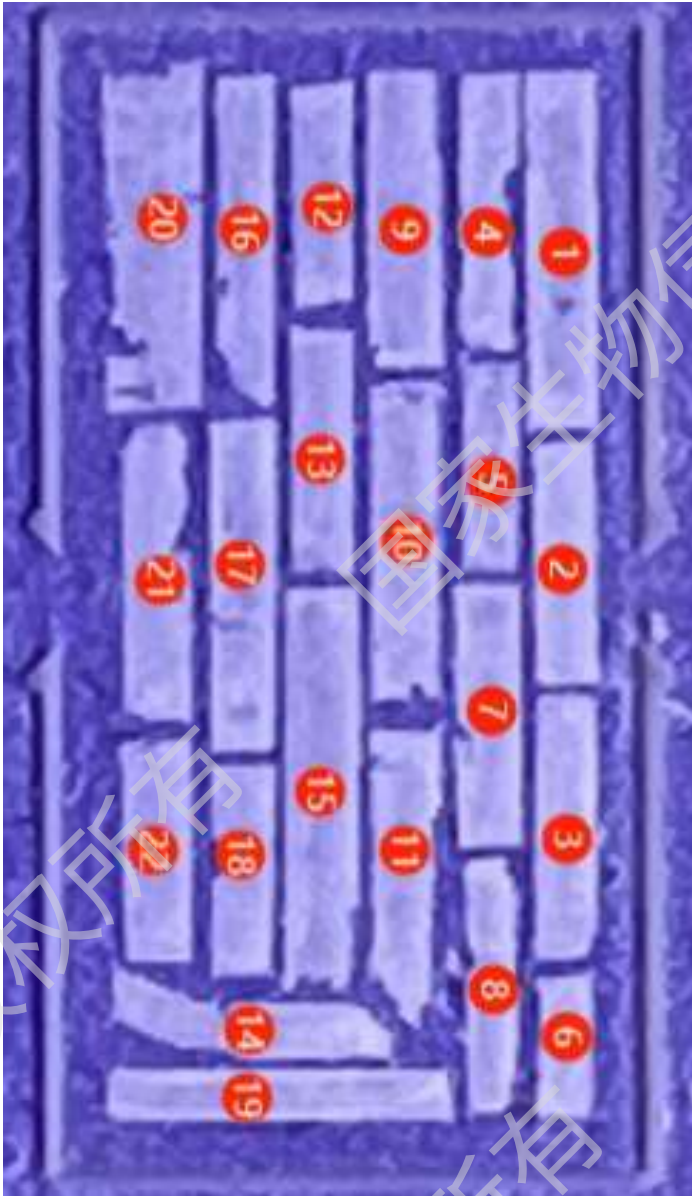
Phenotype associated key subpopulation



Key transcription regulator and target



Sample multiplexing by spatial location

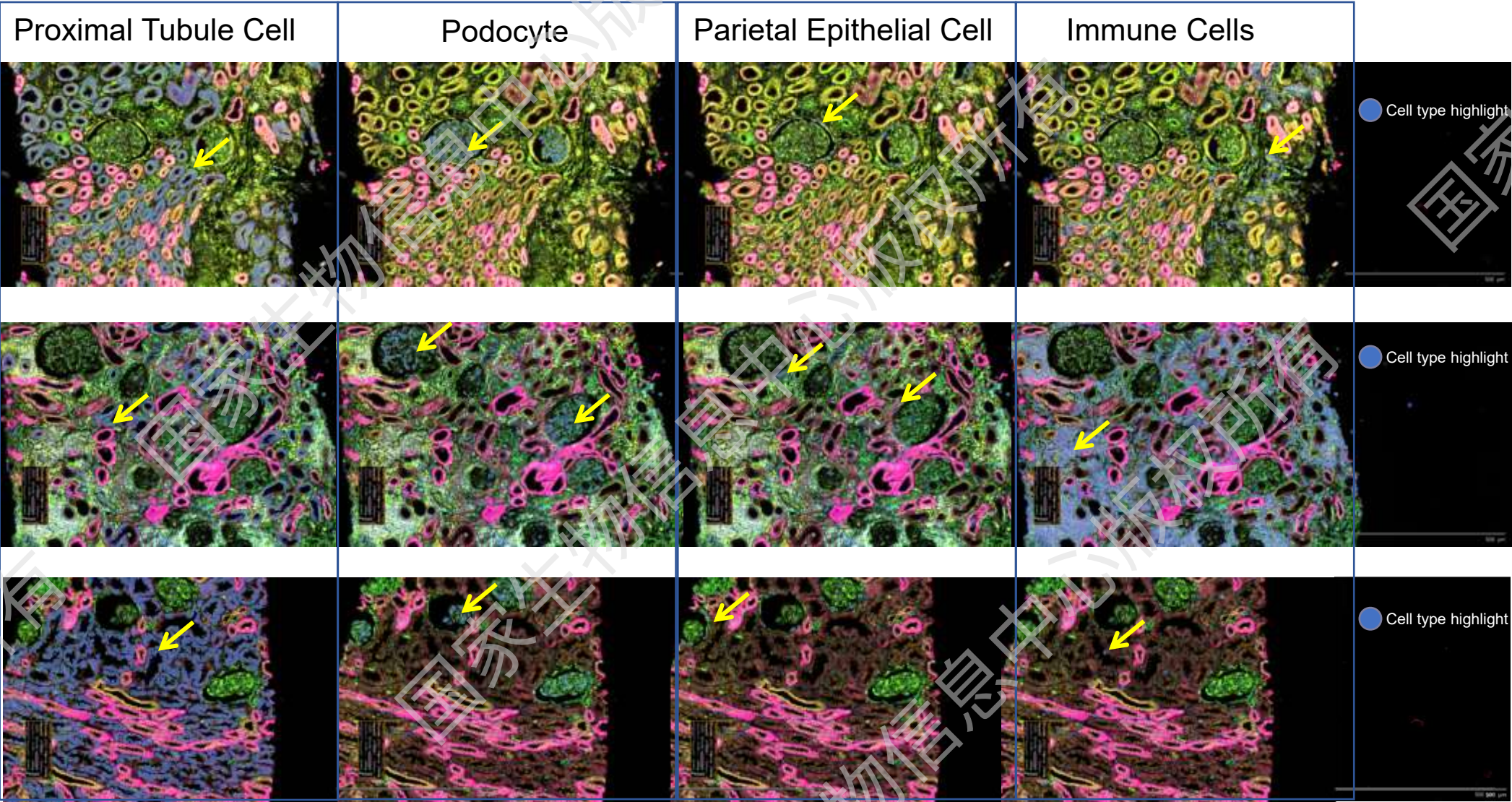


Xenuim image

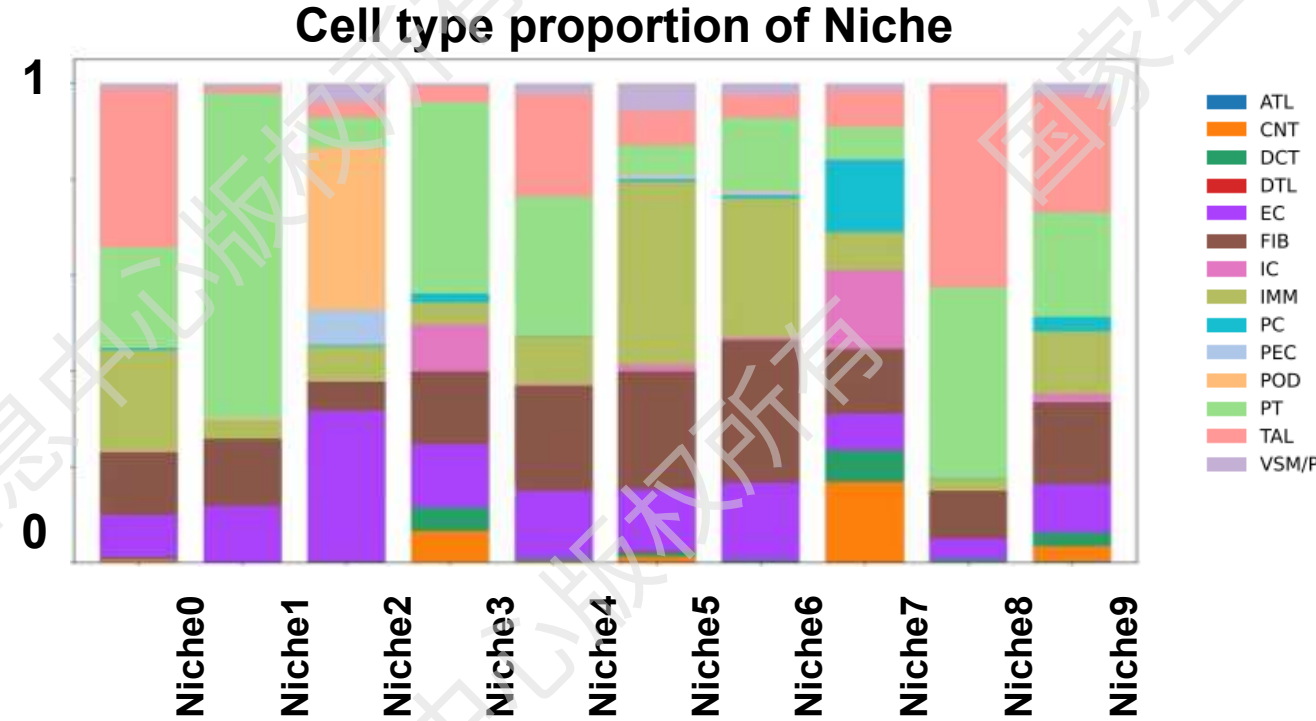
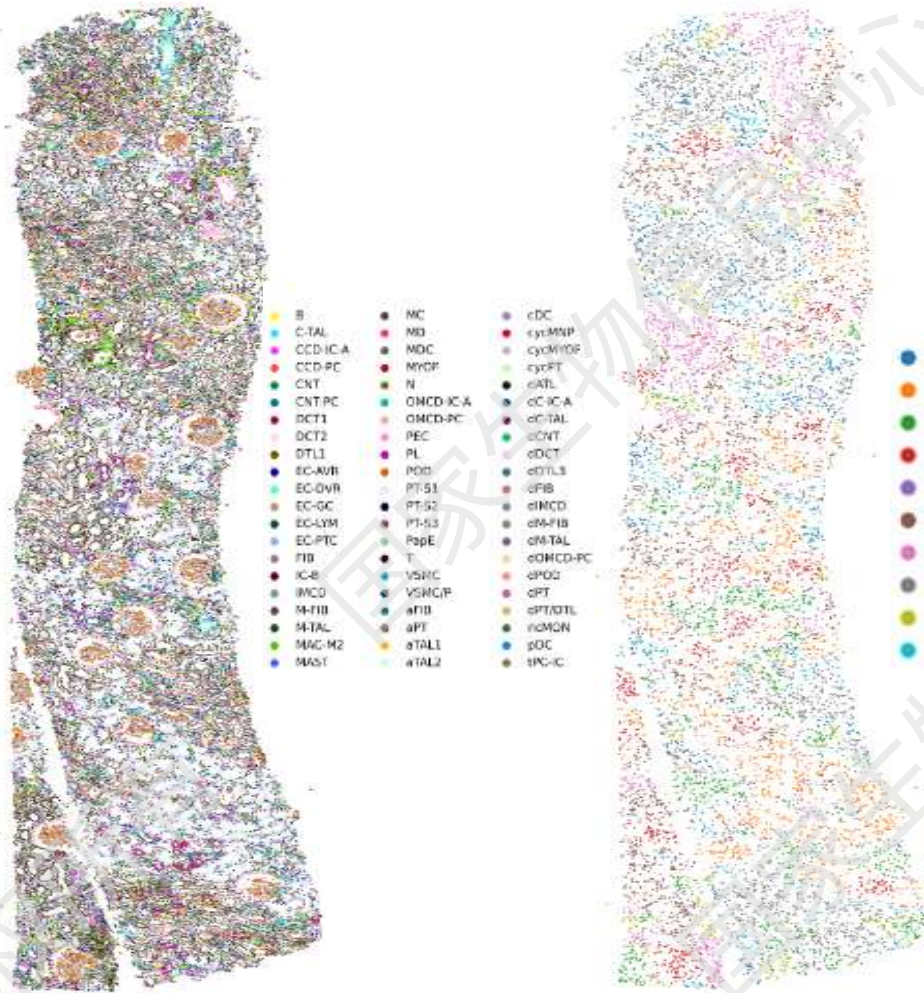
22 biopsy FFPE

Sample ID	Renal Disease
1	Control
2	Control
3	T1-salt losing nephropathy
4	T1-ATN
5	T1-MM
6	T1-Fanconi
7	T1-RTA, SS
8	T1-AIN
9	T1-Fanconi
10	T1- sarcoidosis, granulomatosis T1
11	T1-AIN
12	T1-TINU
13	T1-NPHP
14	T1-SS, RTA, Fanconi
15	T1-SS, RTA, Fanconi
16	T1,irAE, arteriosclerosis
17	T1-SS
18	T1-irAE
19	T1-SS
20	T1-SS
21	T1-irAE, endocytopathy
22	AIN, arteriosclerosis, Liddle syndrome

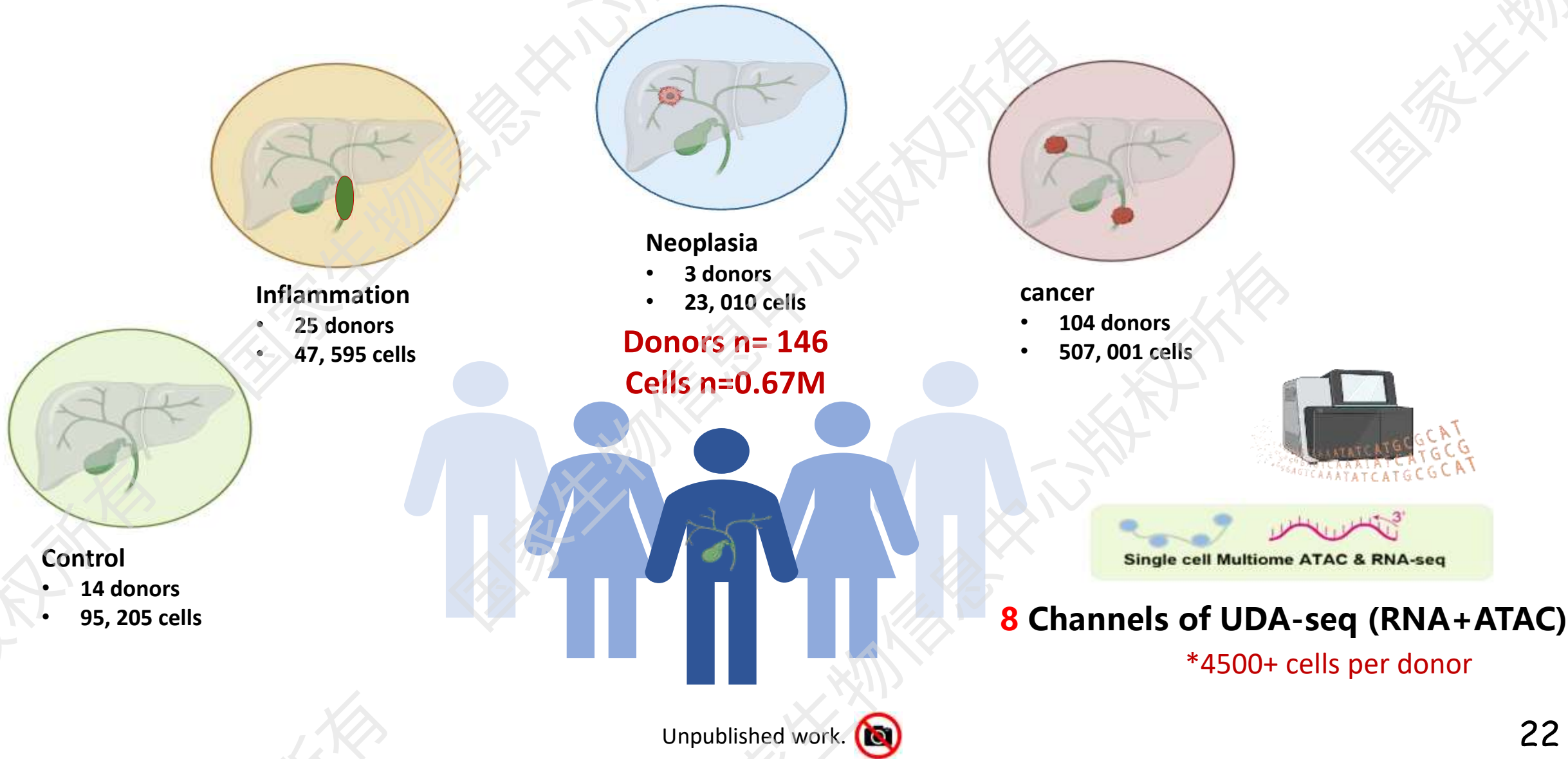
Spatial distribution of Renal important cell type



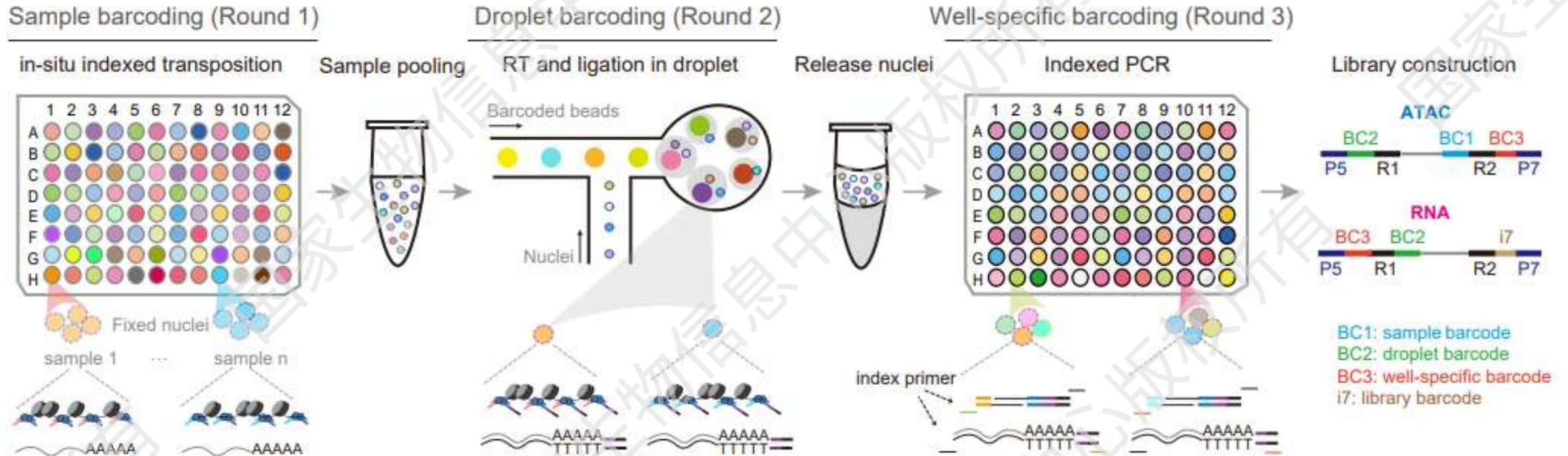
Niche identification by scNiche



Single-Cell Multiomics Decoding Inflammation-Driven Oncogenic Trajectory in Cholangiocarcinoma

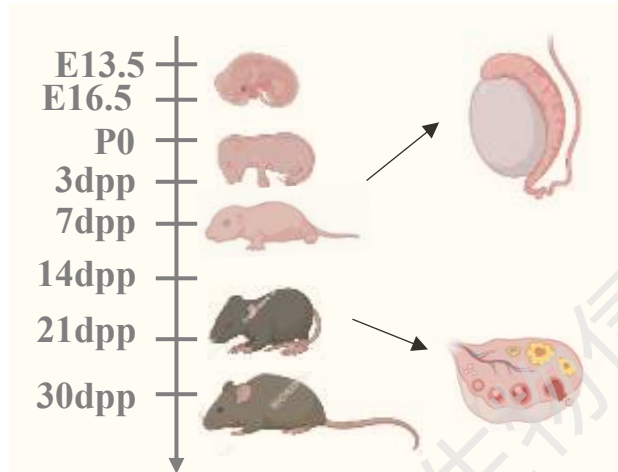


UDA-Seq2: multiplexed UDA-seq (RNA+ATAC)



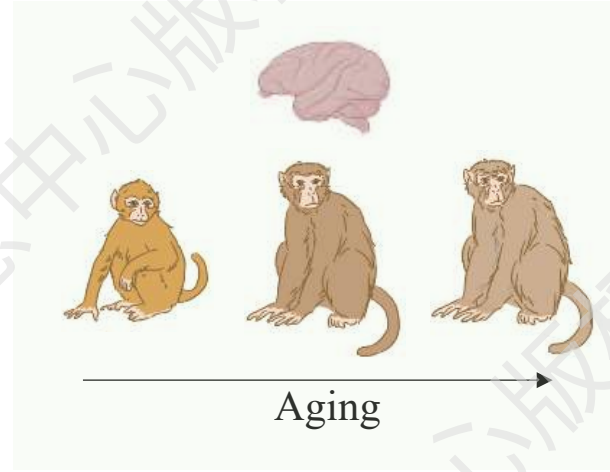
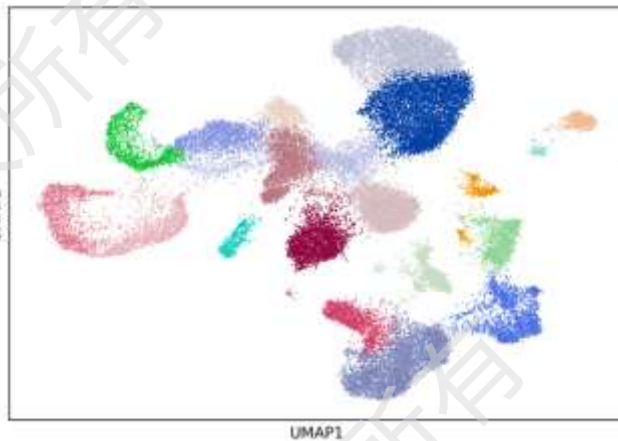
300K cells per channel, sample multiplexing

Application of UDA-Seq2 (RNA+ATAC)



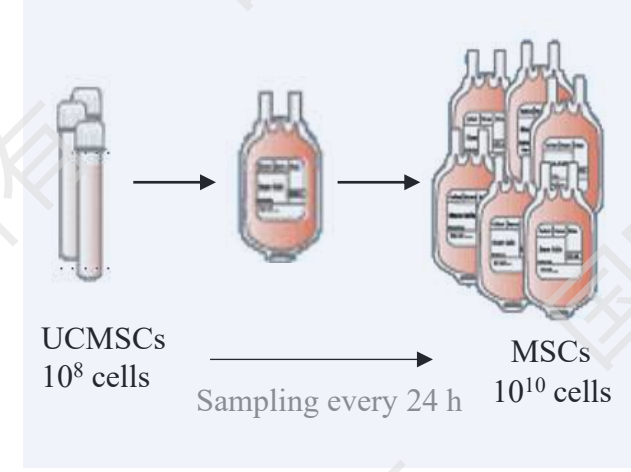
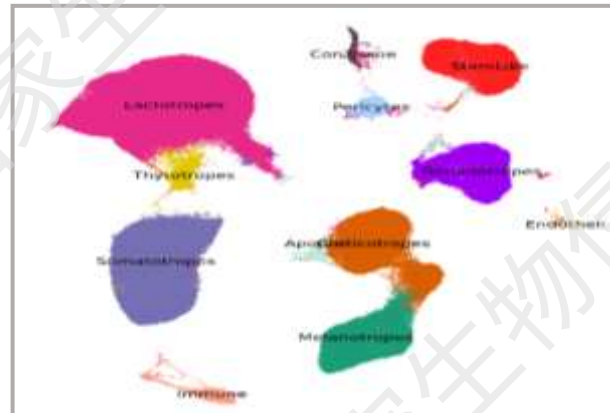
Mouse gonad development

1 channel 16 samples
Capture ~300,000 cells
Median gene (~800)
Median fragments (~7000)



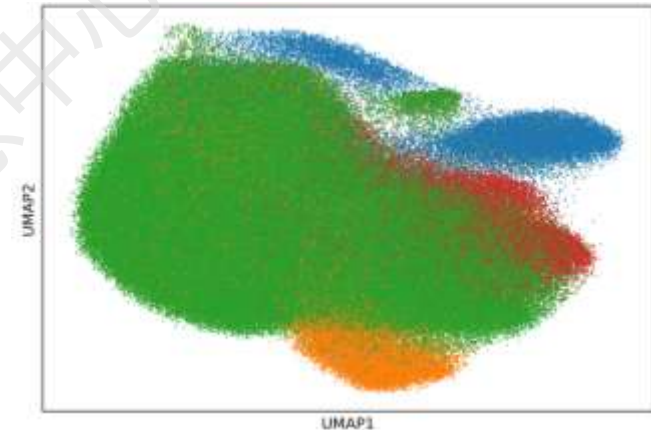
Hypothalamic pituitary aging

3 channel 23 samples
Capture ~450,000 cells
Median gene (~1100)
Median fragments (~1300)

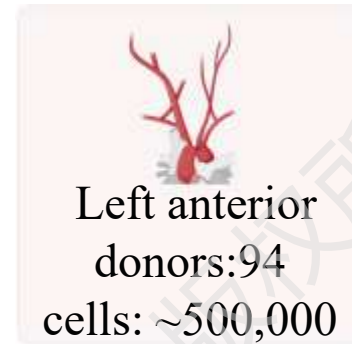
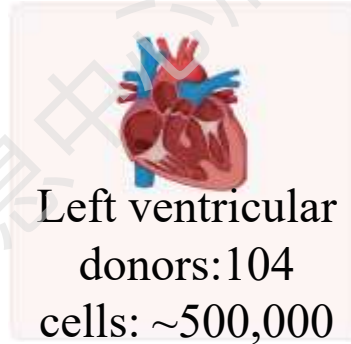
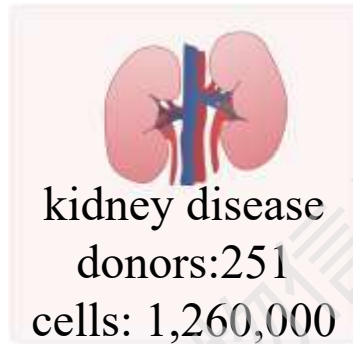
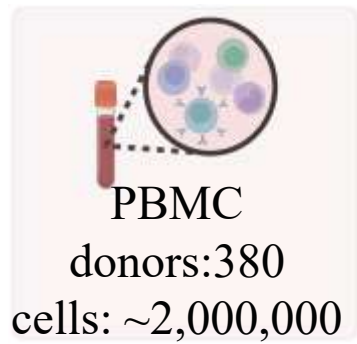


Quality control analysis of MSCs in IND applications

3 channel 51 samples
Capture ~360,000 cells
Median gene (~1315)
Median fragments (~1300)

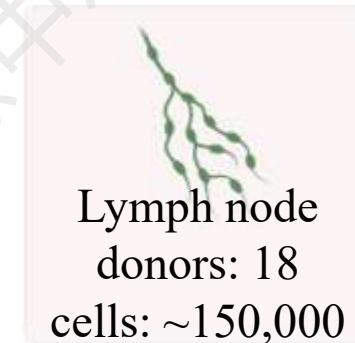
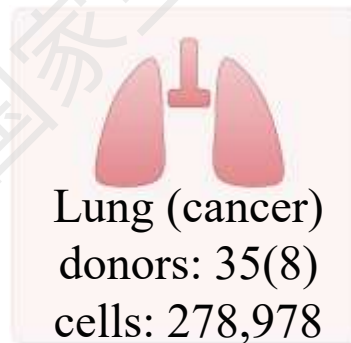
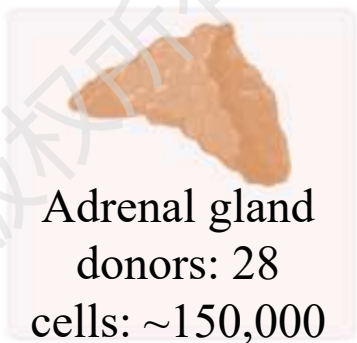
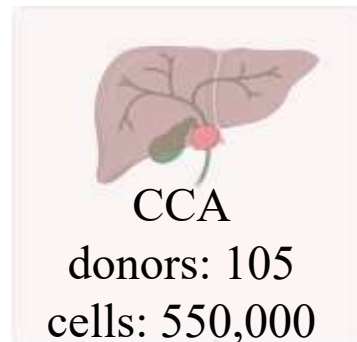


UDA-seq applied in real-world large scale studies



> 6.5M cells to date

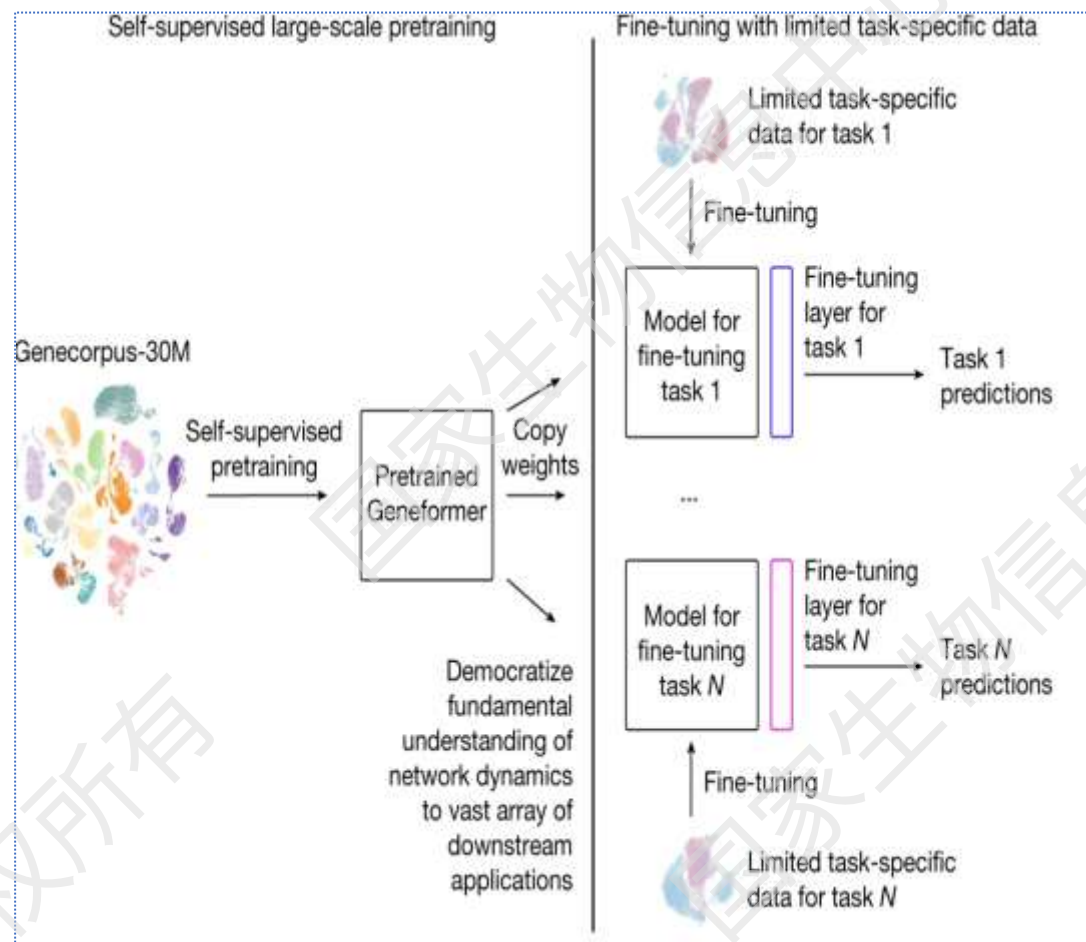
> 16 primary tissues, 1150 donors



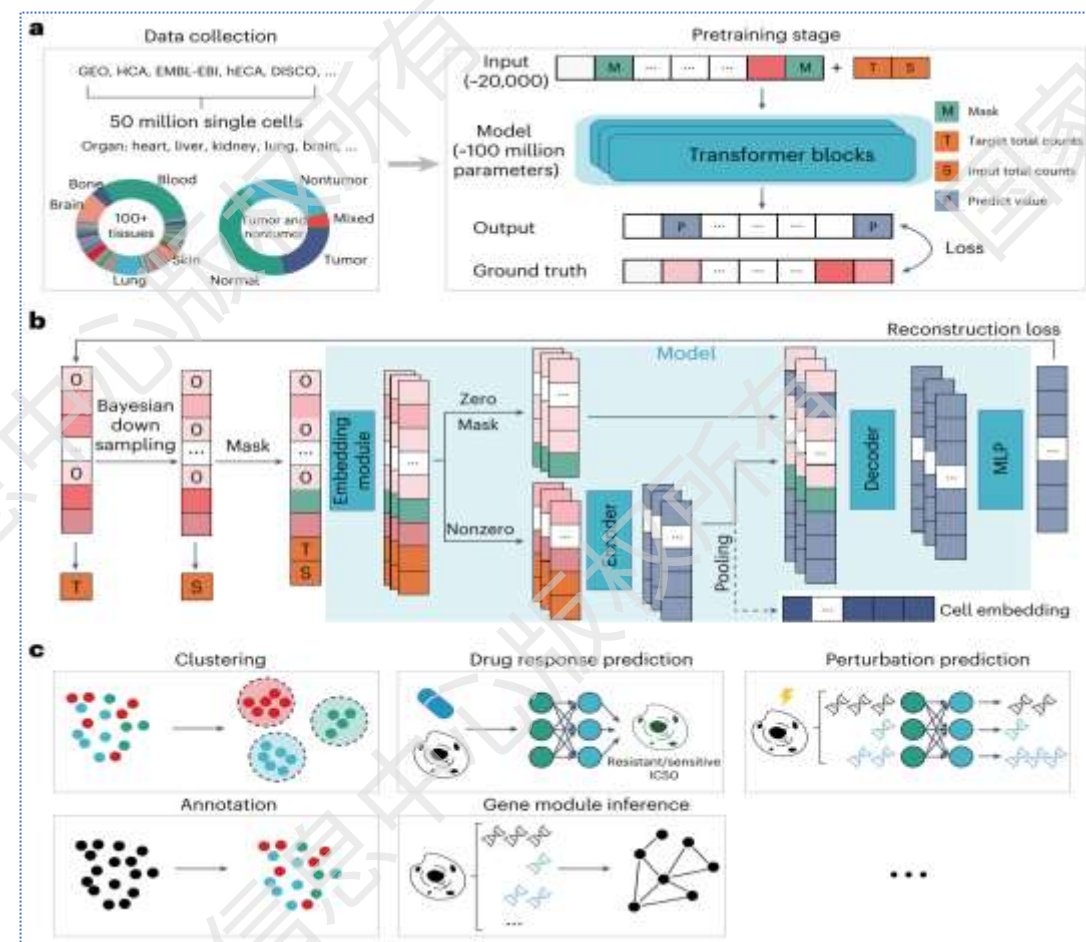
Summary of UDA-seq

- ✓ **UDA-seq Integrated Droplet Microfluidics to achieve ultra-high throughput multiomics**
 - 10-fold increase throughput ($10^4 \rightarrow 10^5$ cells/channel)
 - 20–40 scMultiome samples/channel
 - Gene median: 1k–3k (matches conventional), error rate <2%
 - Compatible modality: (RNA+VDJ, RNA+ATAC, RNA+CRISPR)
- ✓ **Clinical compatibility: Frozen needle biopsies & PBMCs**
- ✓ **Cohorts' studies: Aging, renal diseases, cancer.**
- ✓ **Pheno-linked cell subpopulation discovered:**
 - ITGB1⁺PREX1⁺CD4⁺ Naive T, POD cells, EC-GCs

Large-scale foundation model

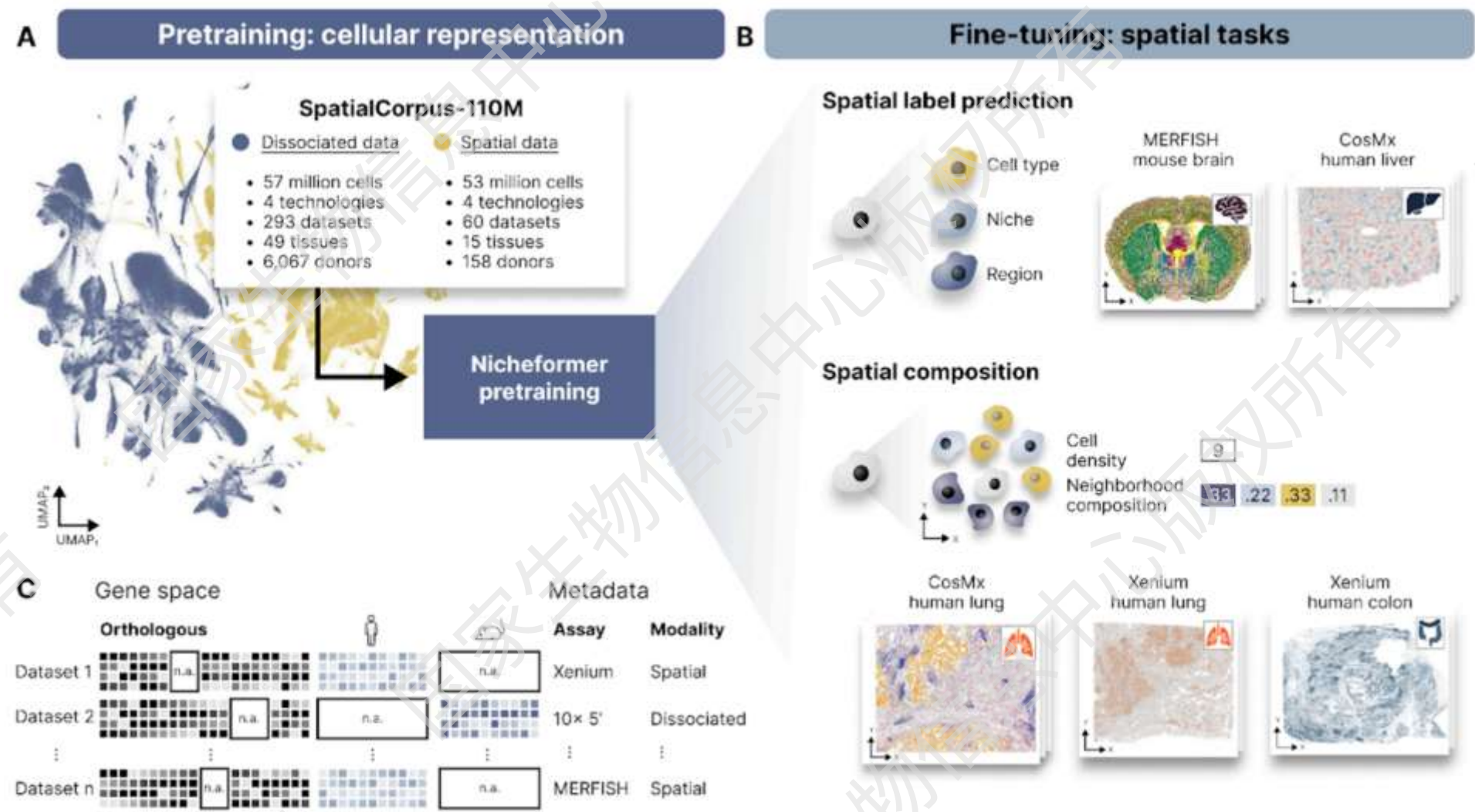


Geneformer



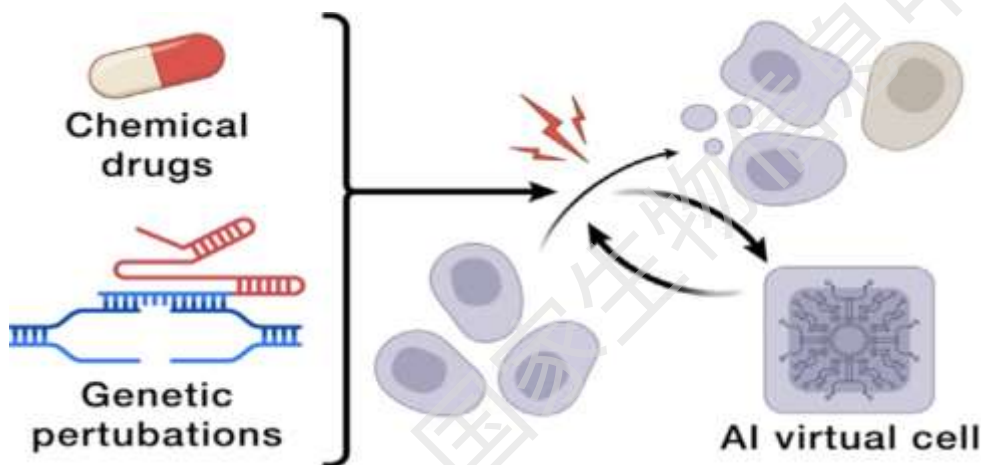
scFoundation

Nicheformer: a foundation model for single-cell and spatial omics

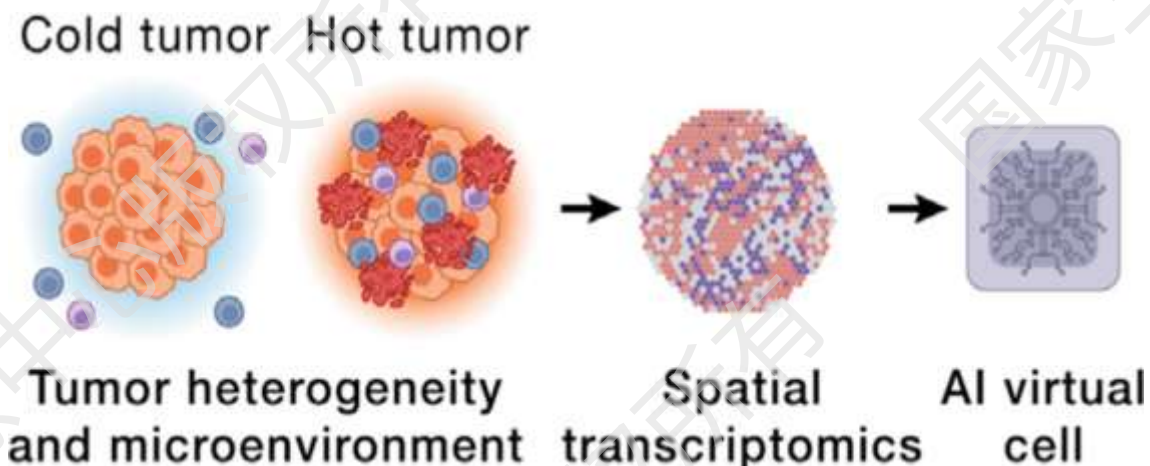


AIVC application

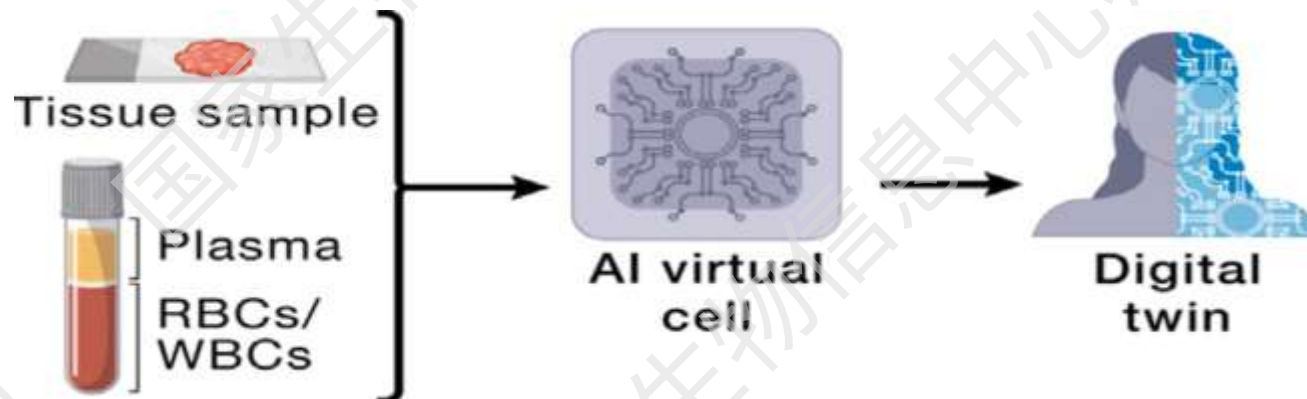
药物响应预测



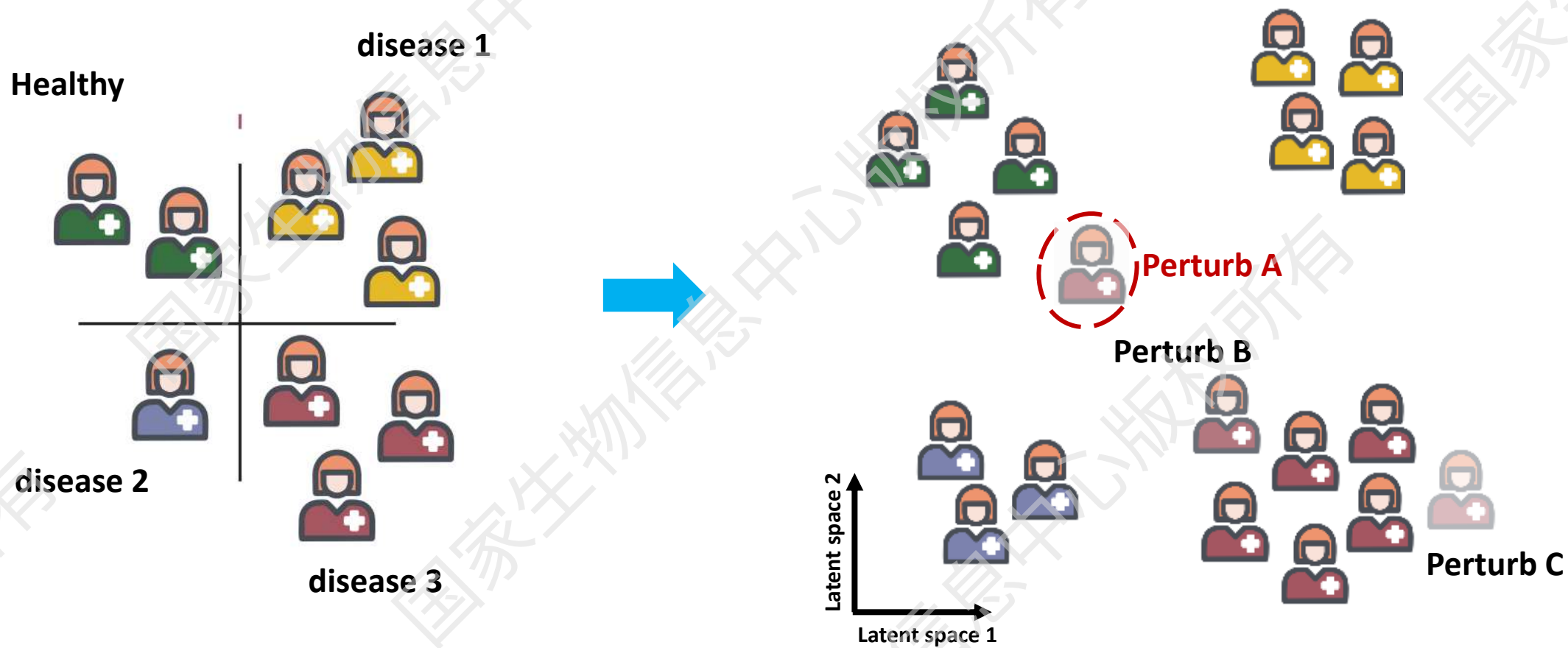
肿瘤微环境模拟



数字孪生诊断



Drug target screening based on digital representations of patients

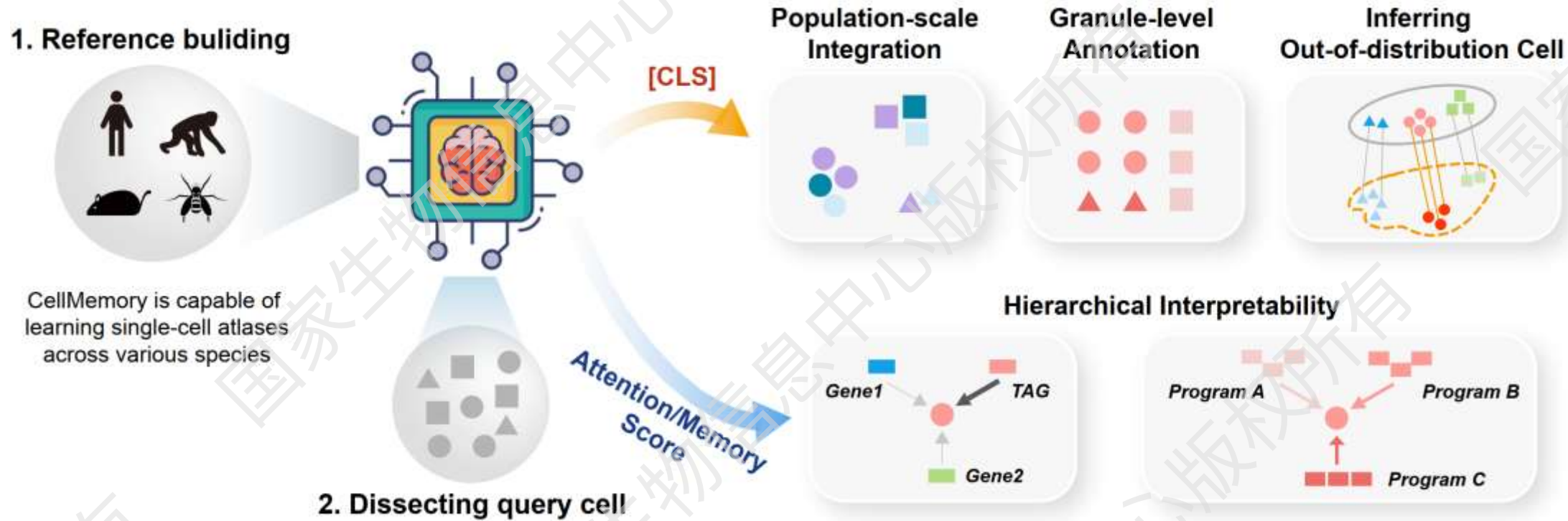


FDA Announces Plan to Phase Out Animal Testing Requirement for Monoclonal Antibodies and Other Drugs

For Immediate Release: April 10, 2025

“For too long, drug manufacturers have performed additional animal testing of drugs that have data in broad human use internationally. This initiative marks a paradigm shift in drug evaluation and holds promise to accelerate cures and meaningful treatments for Americans while reducing animal use,” said FDA Commissioner Martin A. Makary, M.D., M.P.H. “By leveraging AI-based computational modeling, human organ model-based lab testing, and real-world human data, we can get safer treatments to patients faster and more reliably, while also reducing R&D costs and drug prices. It is a win-win for public health and ethics.”

CellMemory deciphering OOD cell with hierarchical interpretation



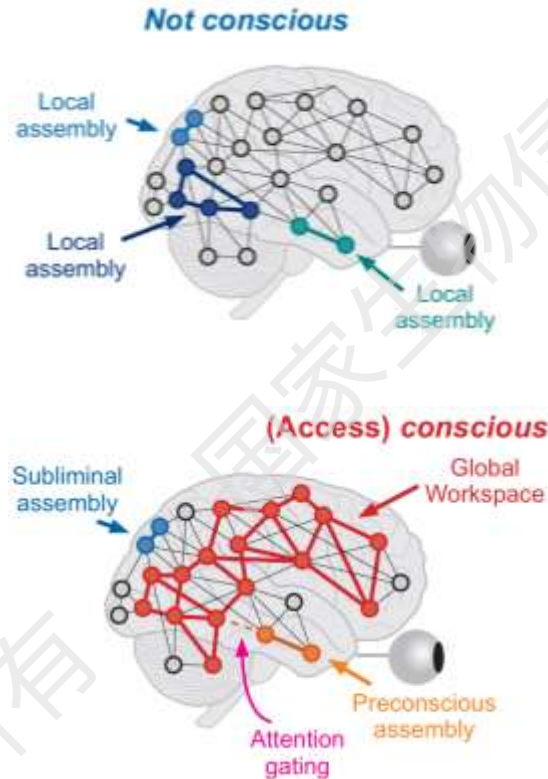
Interpretation level 1: Each cell is assigned an attention score across all features, highlighting those that are most critical for the model's understanding of that cell.

Interpretation level 2: For each cell, every memory slot within the model focuses on the subset of features (e.g., gene program), with the importance of these features represented by the memory score.

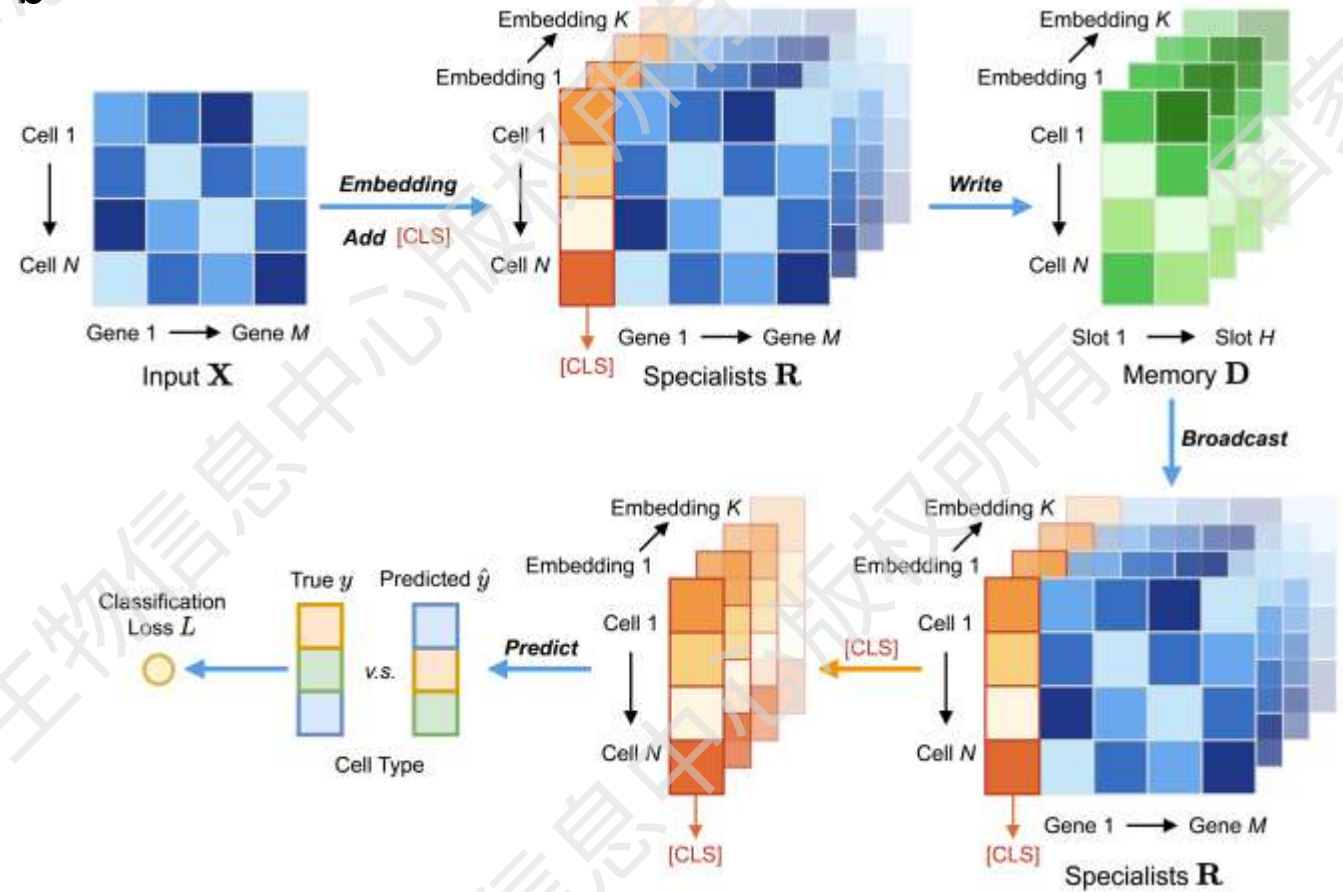
CellMemory is a Bottlenecked Transformer (Cross-Attention Architecture)

Global Workspace Theory: Consciousness arises when selected information enters a limitation space and is broadcast

a

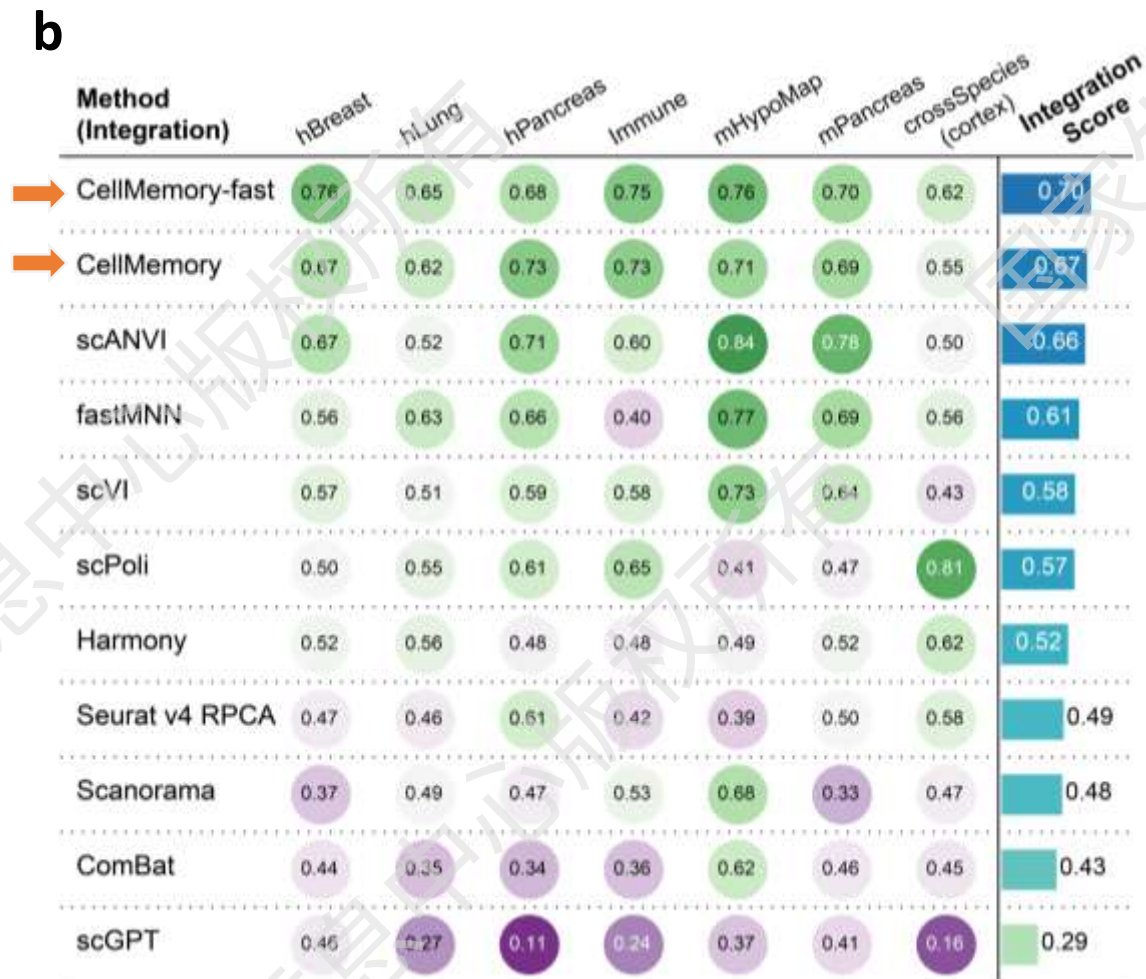
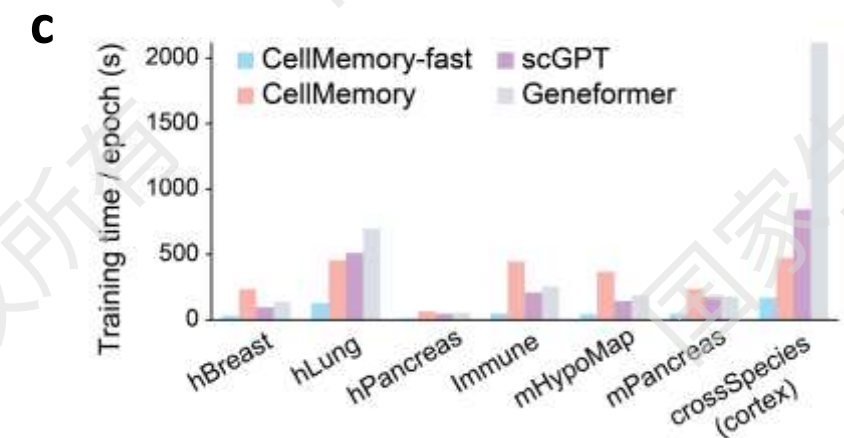
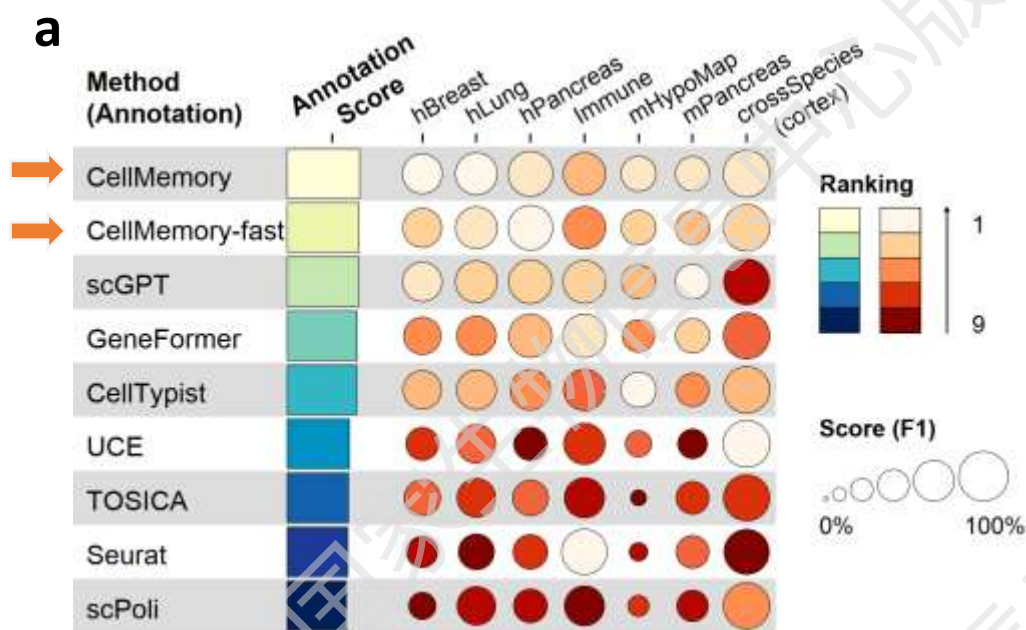


b



Bottlenecked Transformer exhibits superior
Generalization and Computational Efficiency

Reference mapping benchmark



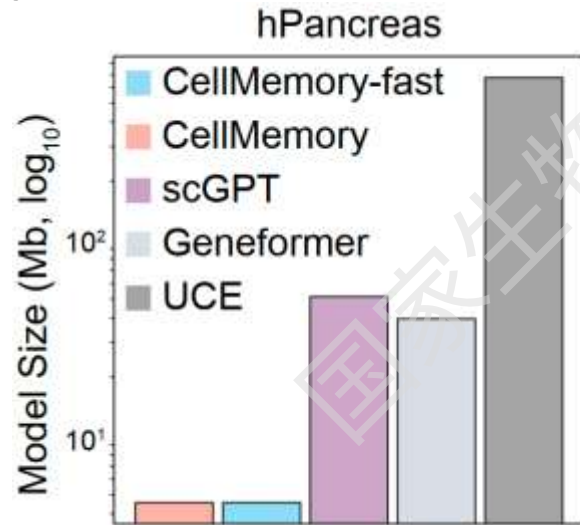
CellMemory outperforms single-cell foundation models in generalization and speed.

Scalability test

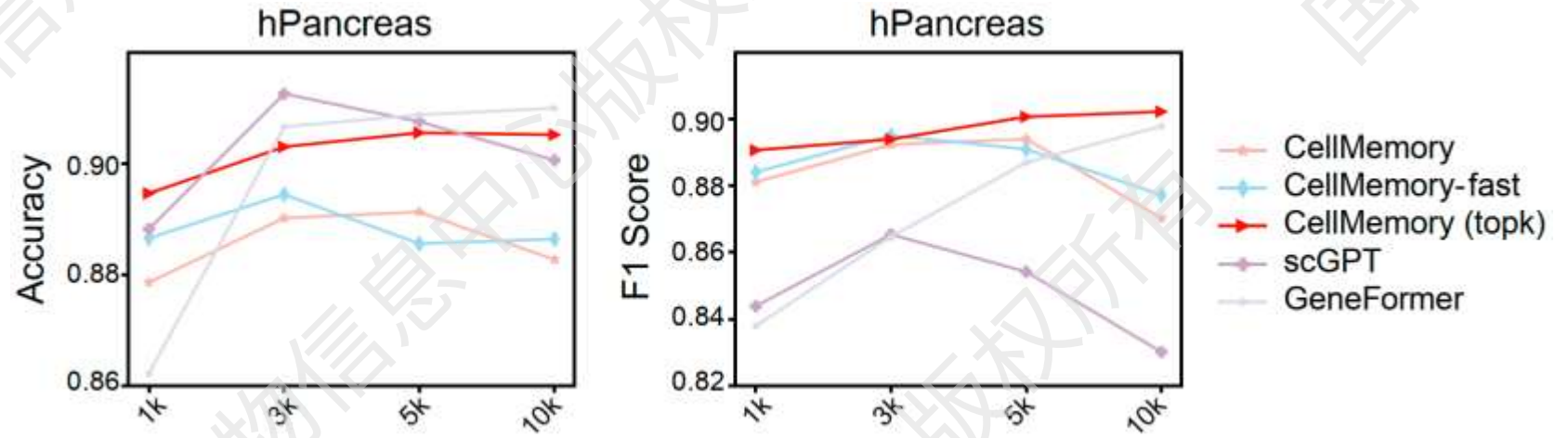
Reference: 26k cells

Query: 21k cells

a

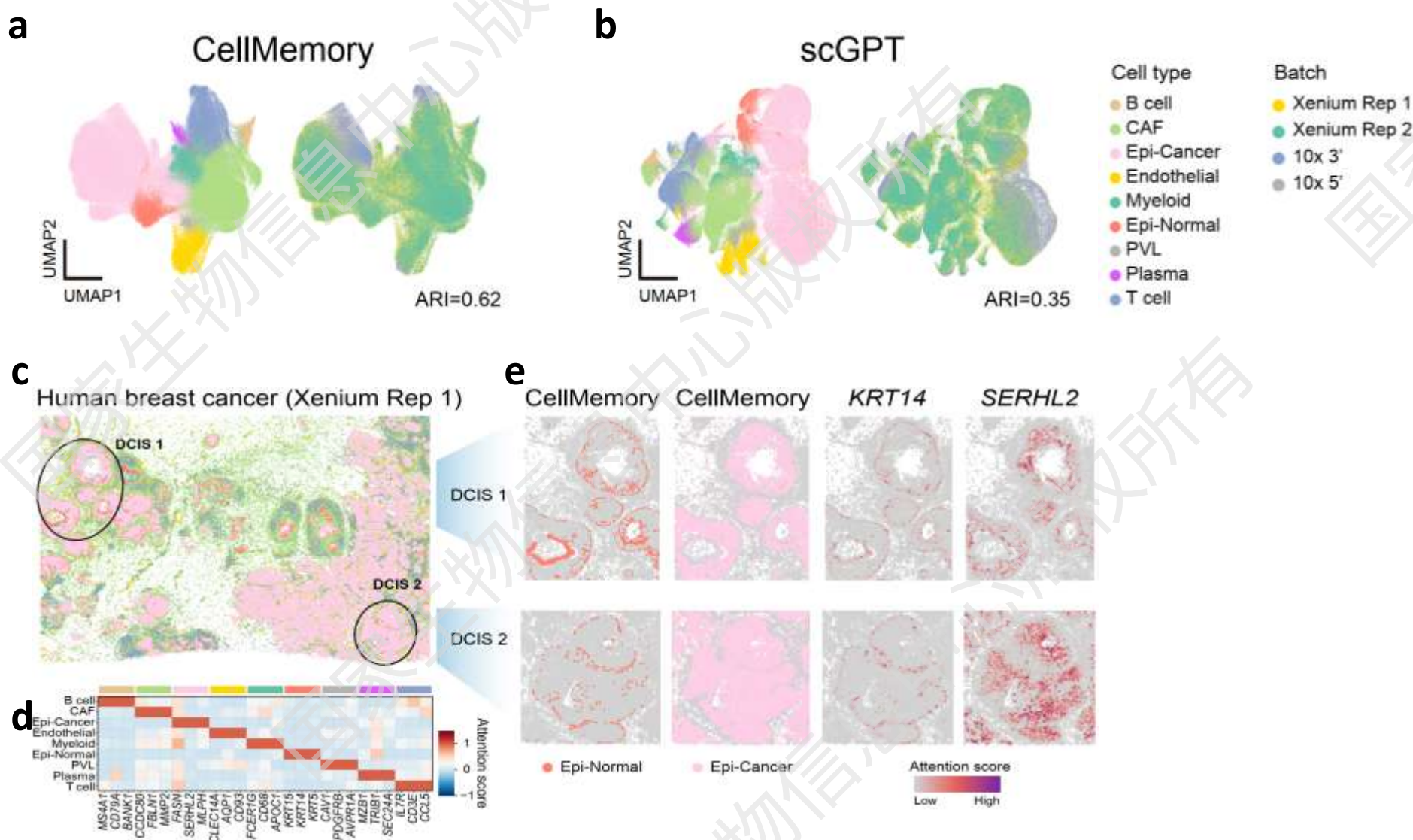


b

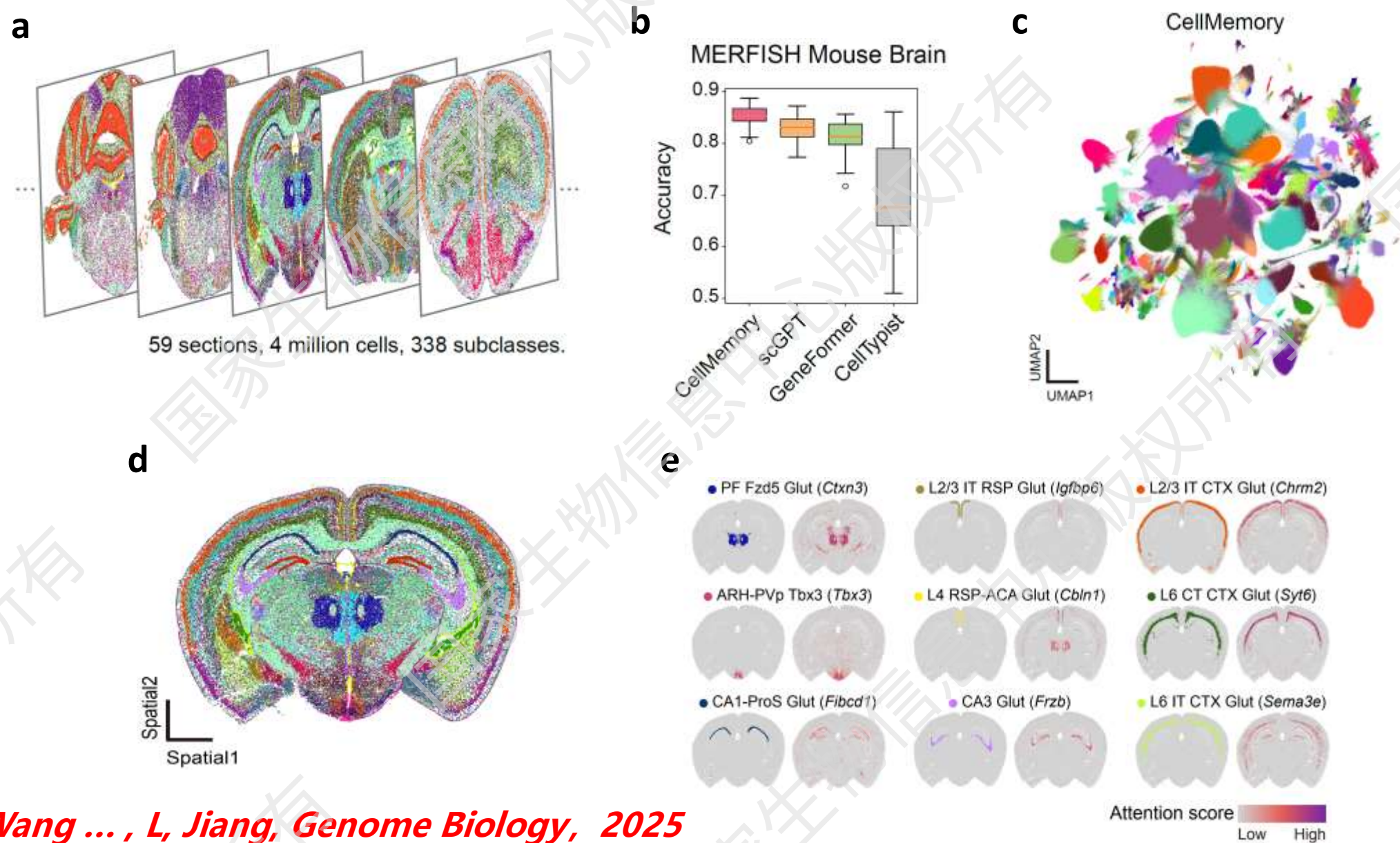


In evaluations on small dataset, sc foundation models achieved higher Accuracy. However, in scenarios characterized by imbalanced cell type proportions, CellMemory achieved superior F1-score.

Interpreting human breast cancer *Xenium* data

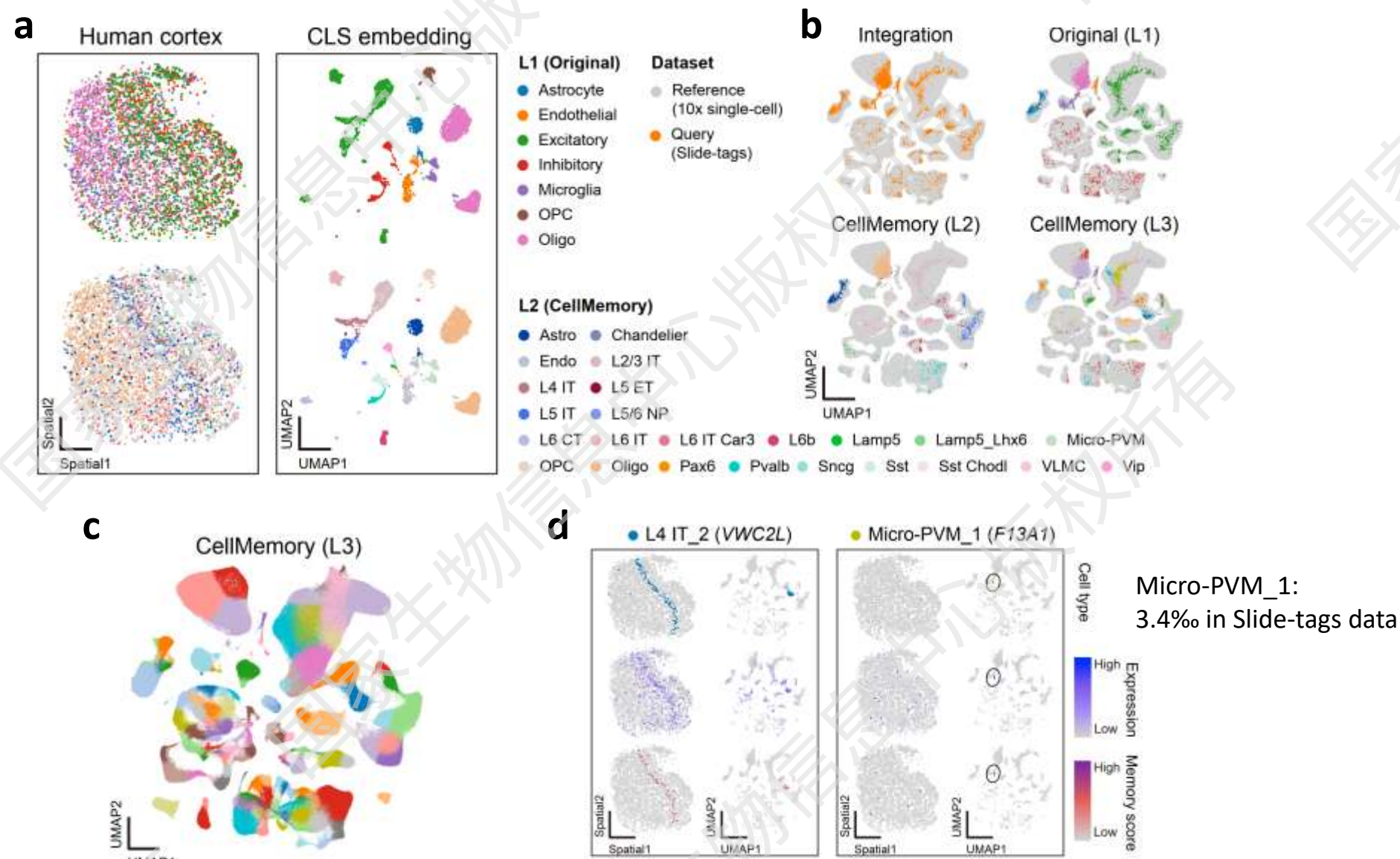


Characterizing mouse whole-brain *MERFISH* atlas at granule-level



Qifei Wang ... , L. Jiang, *Genome Biology*, 2025

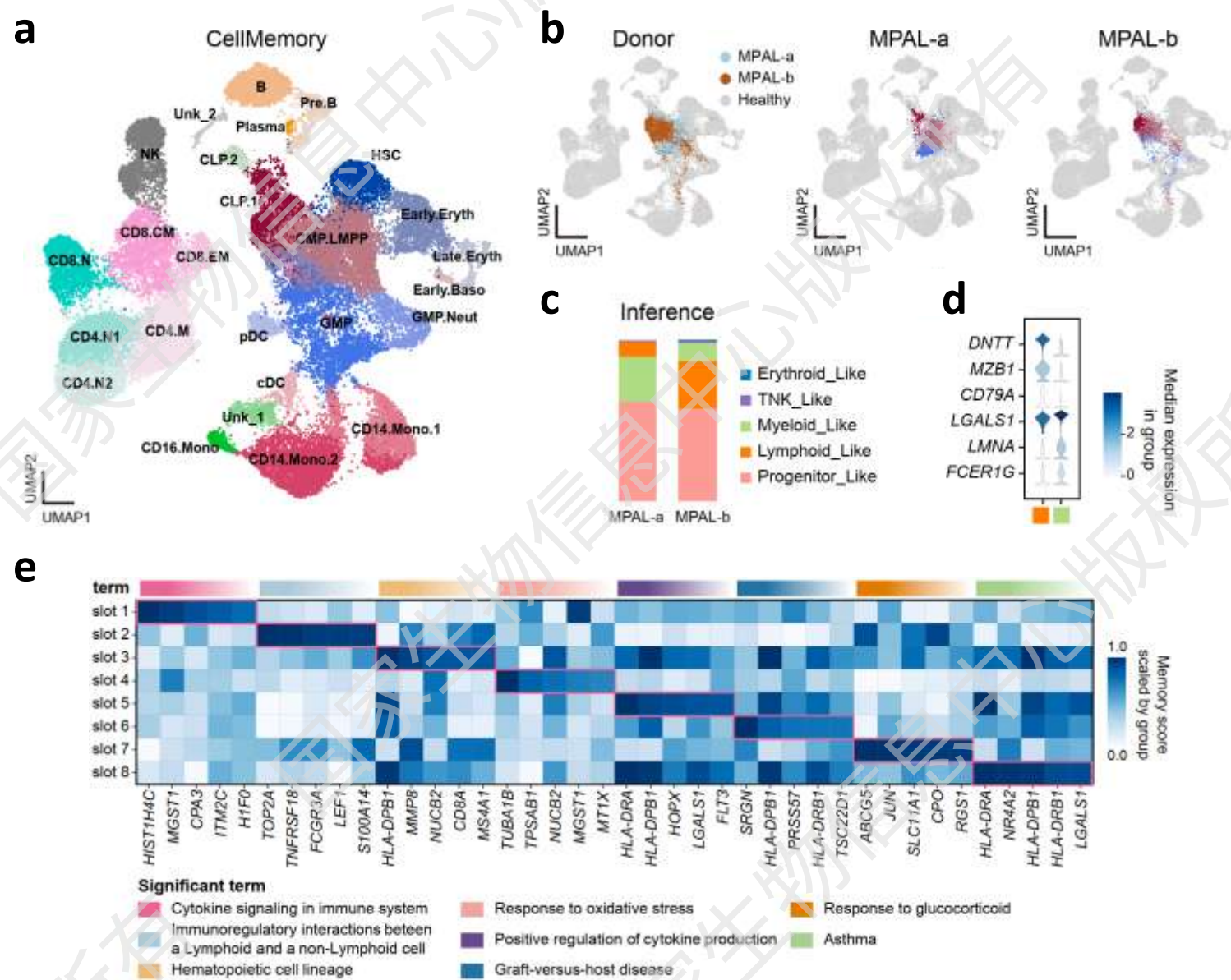
Characterizing human cortex *Slide-tags* data at granule-level



L2 (24 cell types) and L3 (131 cell states) annotated by CellMemory

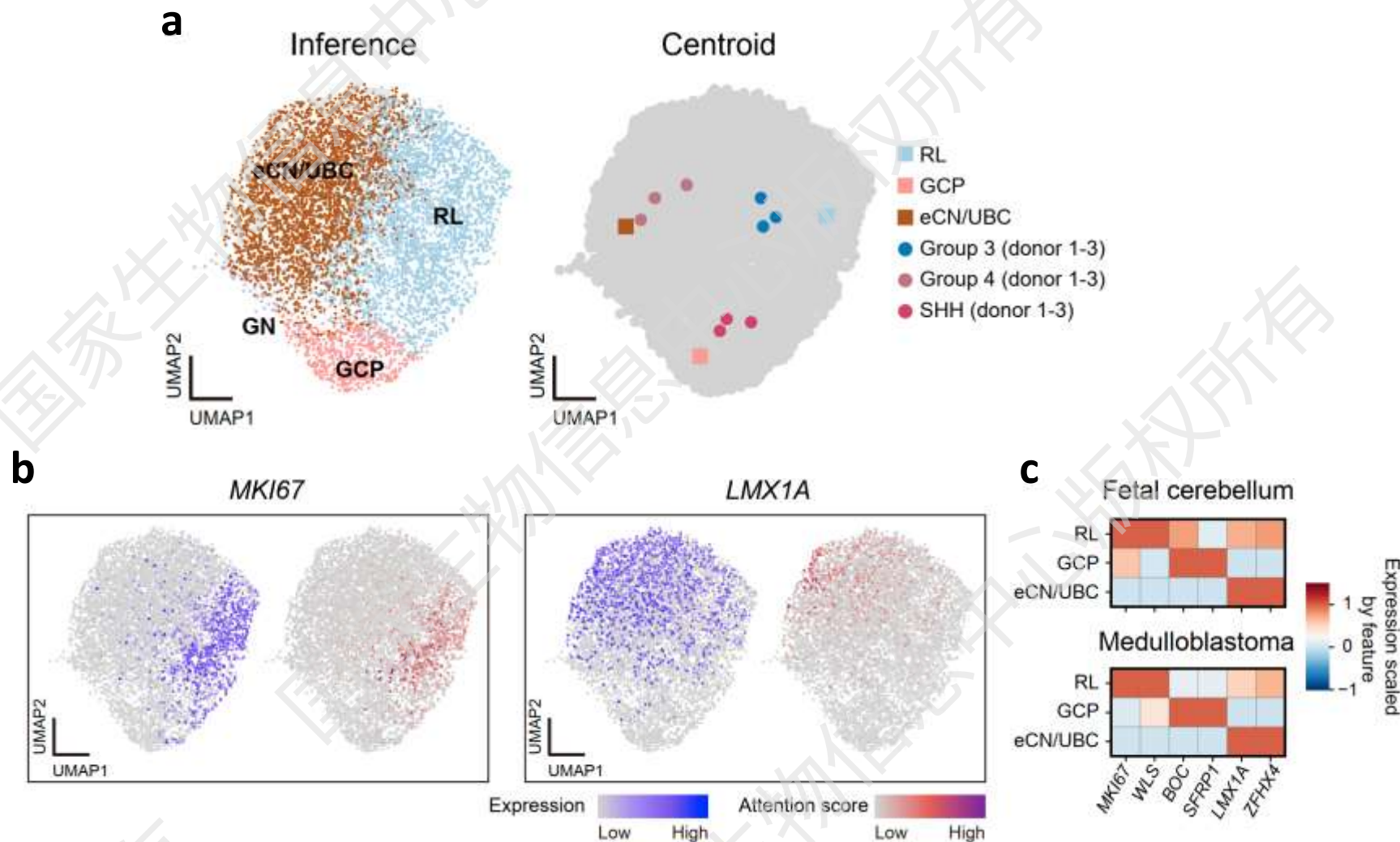
Inferring founder cell of Mixed Phenotype Acute Leukemia (MPAL)

Studying malignant cells and the heterogeneity among patients using normal cells (BMMC, PBMC) as a baseline



Inferring founder cell of Medulloblastoma (MB)

Studying malignant cells and the heterogeneity among patients using normal cells (human embryonic cerebellar glutamatergic neuronal lineages) as a baseline



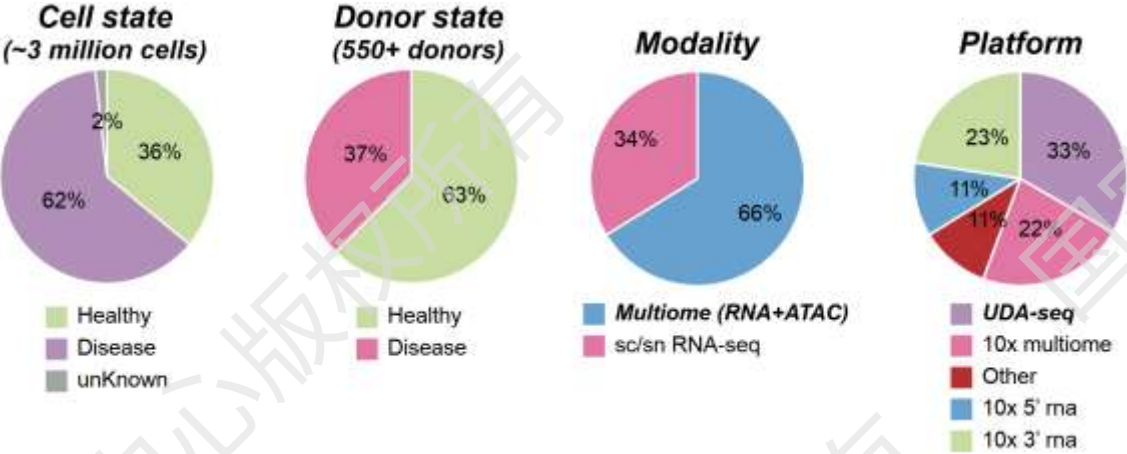
Multi-omics Foundation Model for Organ-X

a

3 million cells from organ-X

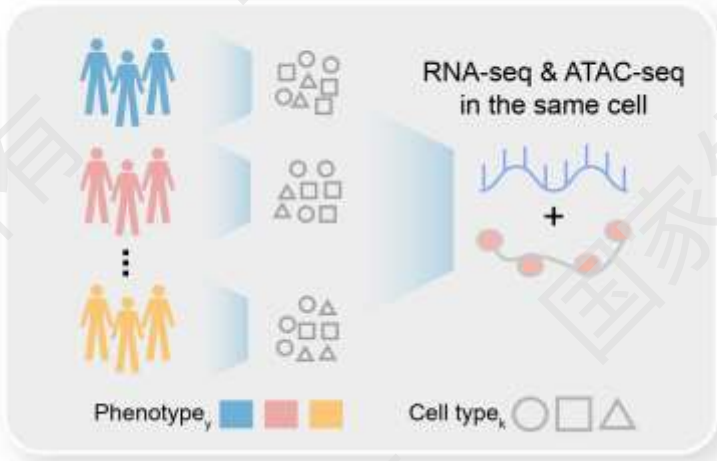


b

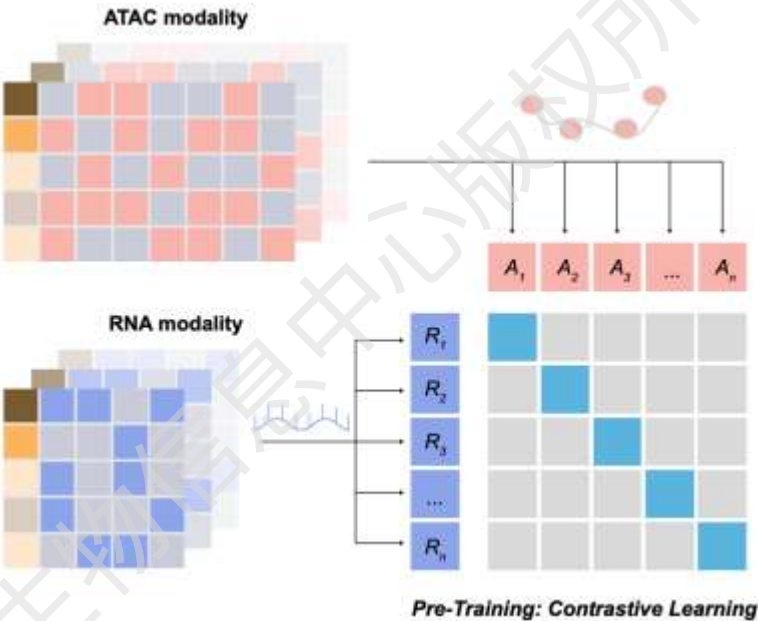


c

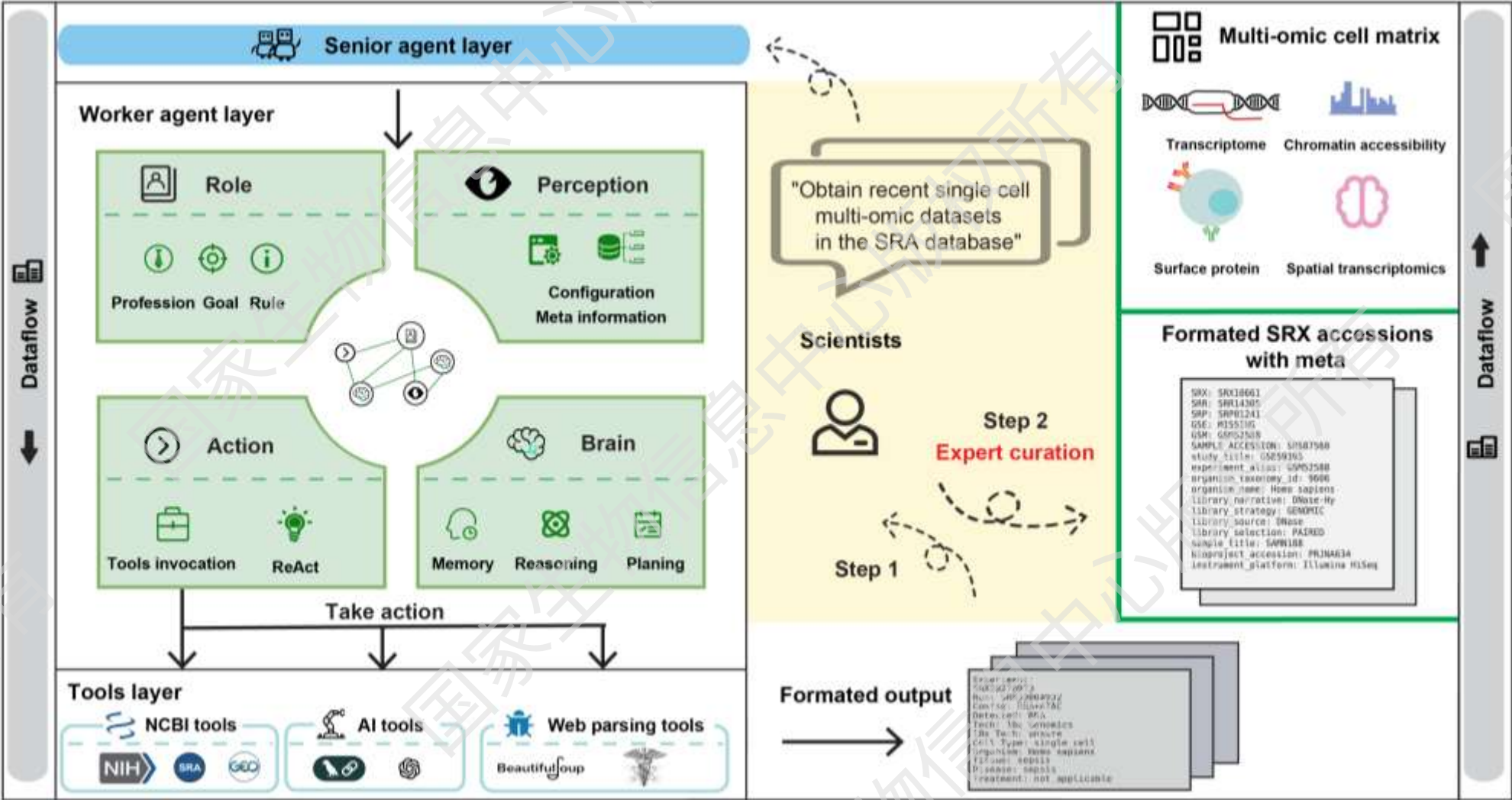
Single-cell multi-omics sequencing of population cohorts



d



AI Agent for automatic single-cell multi-omics data collection



Summary of CellMemory

- ✓ **CellMemory offers the advantages of computational efficiency, strong generalization capabilities, and interpretability**
- ✓ **It achieves state-of-the-art (SOTA) performance in cell type annotation and cross-technology platform integration**
- ✓ **Supports multi-modal data representation, enables cross-scale identification of phenotype-associated key cell subpopulations**
- ✓ **facilitates virtual target screening for cross-scale individual characterization.**

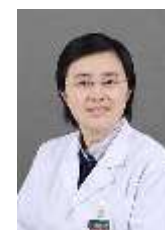
Acknowledgement



乔杰



闫丽盈



陈丽萌



毛颖



杨辉



李平



王韞芳



金烁



左先波



姚元庆



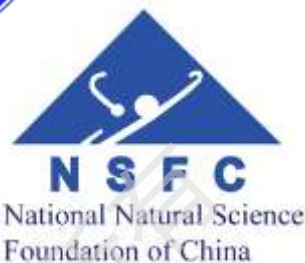
张维绮



刘金平



Funding



Thanks