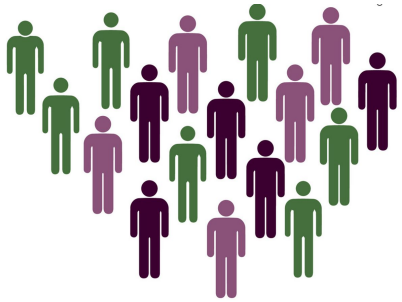
Bridging AI and Biomedical Research

October 27, 2026

Rui Chen, Ph.D.
Brunson Center for Translational Vision Research
Gavin Herbert Eye Institute
Department of Ophthalmology and Visual Sciences
University of California, Irvine

The Multiscale Nature of Biomedical Data

Population



- Biobanks & cohorts
- GWAS & epidemiology
- Disease registries

Cohort • Biobank

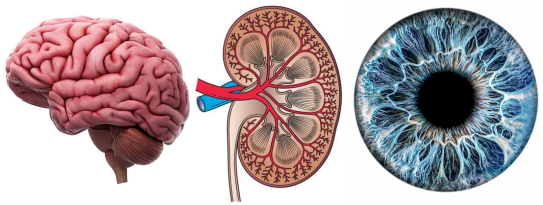
Individual



- Electronic health records
- Clinical tests & diagnoses
- Wearables & genomics

EHR • Lab results

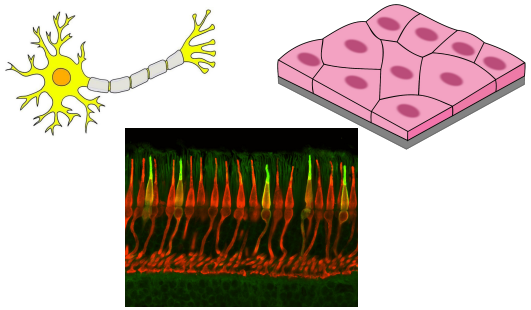
Tissue / Organ



- Medical imaging
- Histology & pathology
- Spatial measurements

Imaging • Physiology • Histology

Individual Cell



- Gene expression
- Epigenomics & proteins
- Cellular phenotypes

Morphology • Omics • Pathway

NIH Common Fund Ecosystem



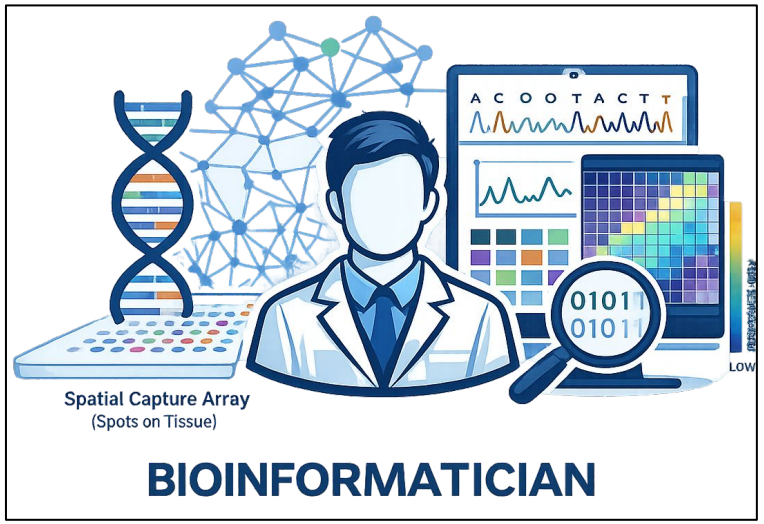
Challenges in Biomedical Big Data Analysis

- **Data complexity and diversity**: Keeping pace with the rapidly growing volume and heterogeneity of biomedical datasets.
- **Specialized expertise**: Analyzing complex datasets requires advanced computational and bioinformatics skills.
- **Rapidly evolving analytical tools**: The expanding landscape of computational tools makes it challenging to select and apply the most appropriate methods.
- **Lack of standardized workflows**: Integrating and analyzing data across multiple modalities remains complex and often requires customized workflows.
- **Computational infrastructure**: Large-scale data analysis often requires substantial computing resources and specialized infrastructure.

Data Integration and Interpretation Require Domain and Informatics Expertise



Wet Lab Researcher





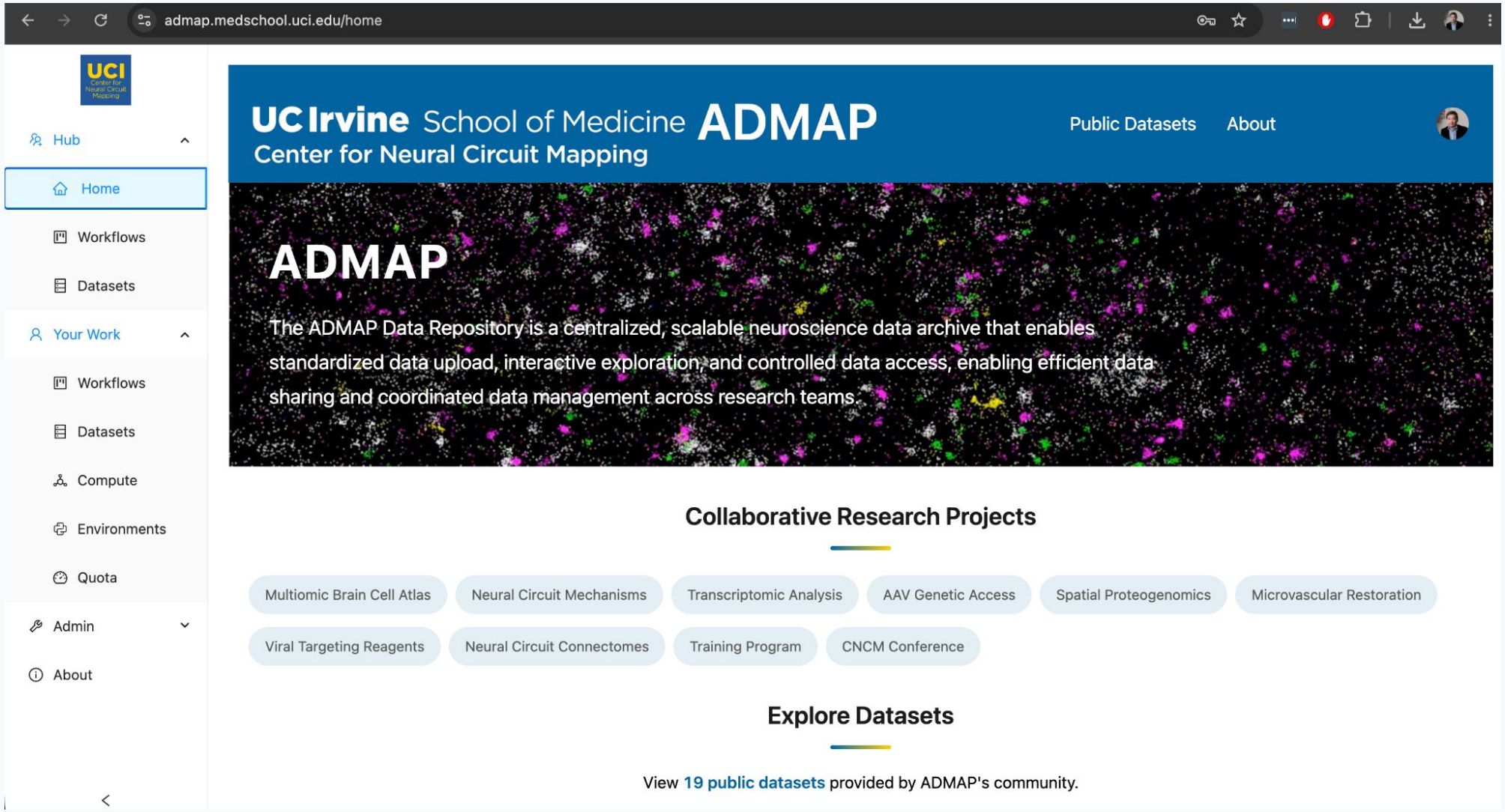
Low efficiency, Lost information

AI Technologies Empowers Individual Scientist



A Texera-Based Platform for Neuroscience and Alzheimer's Disease Research

https://admap.medschool.uci.edu

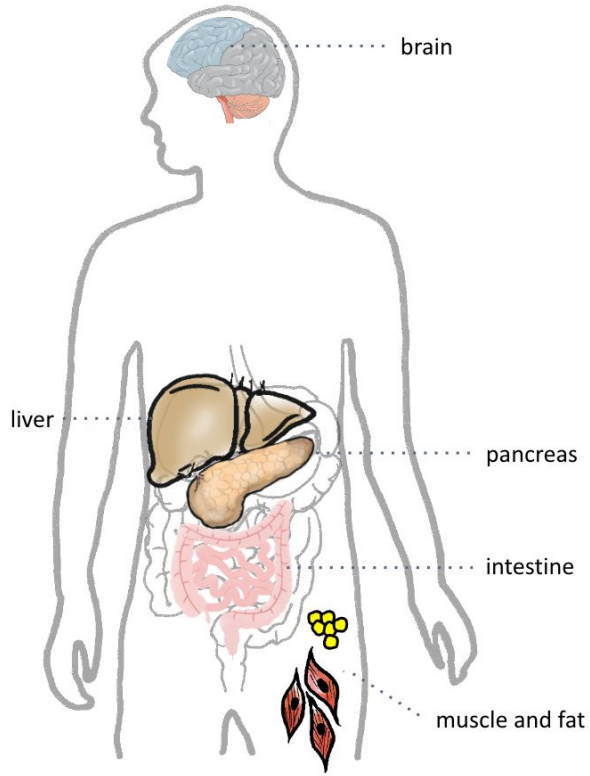


NSF BRIDGE · Biomedical research domain

Metabolic Traits

Finding how DNA differences shape metabolic traits, one cell at a time

Stephen C.J. Parker
Ha Vu
Jonathan Poisson (U-M Advanced Research Computing)
University of Michigan, with Chen Li's Texera team at UC Irvine



VisionKB: Making Advanced AI Accessible to Vision Researchers

- **Integrate diverse datasets**: Perform meta-analysis of diverse datasets to build reference. Perform integrative analysis of heterogeneous data across multiple modalities.
- **Leverage cutting-edge AI technologies**: Make advanced AI models and analytical tools readily accessible to the vision research community.
- **Develop customized analytical workflows**: Provide user-friendly, scalable, and reproducible workflows tailored to specific biomedical research needs.
- **Democratize access to advanced computational tools**: Enable biomedical researchers to perform sophisticated data analyses without requiring extensive computational expertise.



To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.

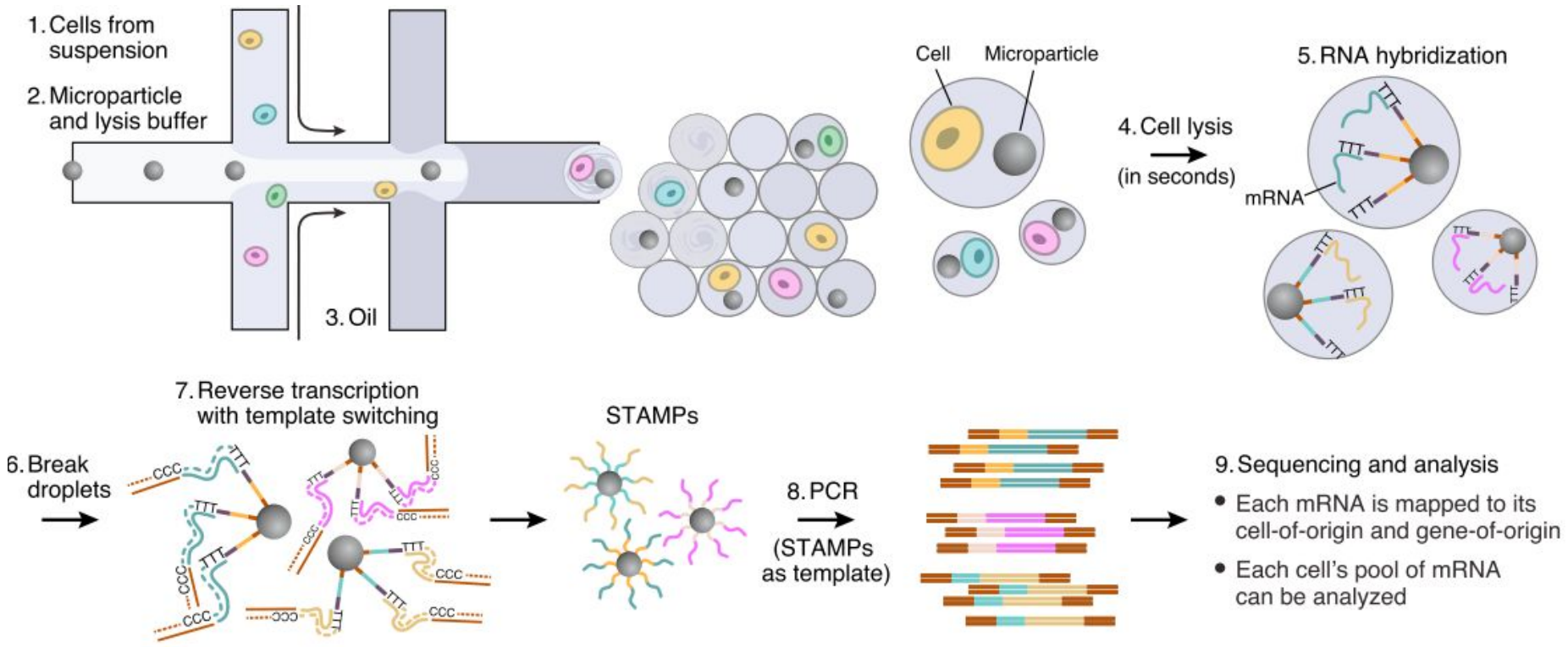
https://www.humancellatlas.org

HUMAN CELL ATLAS

More about individual networks

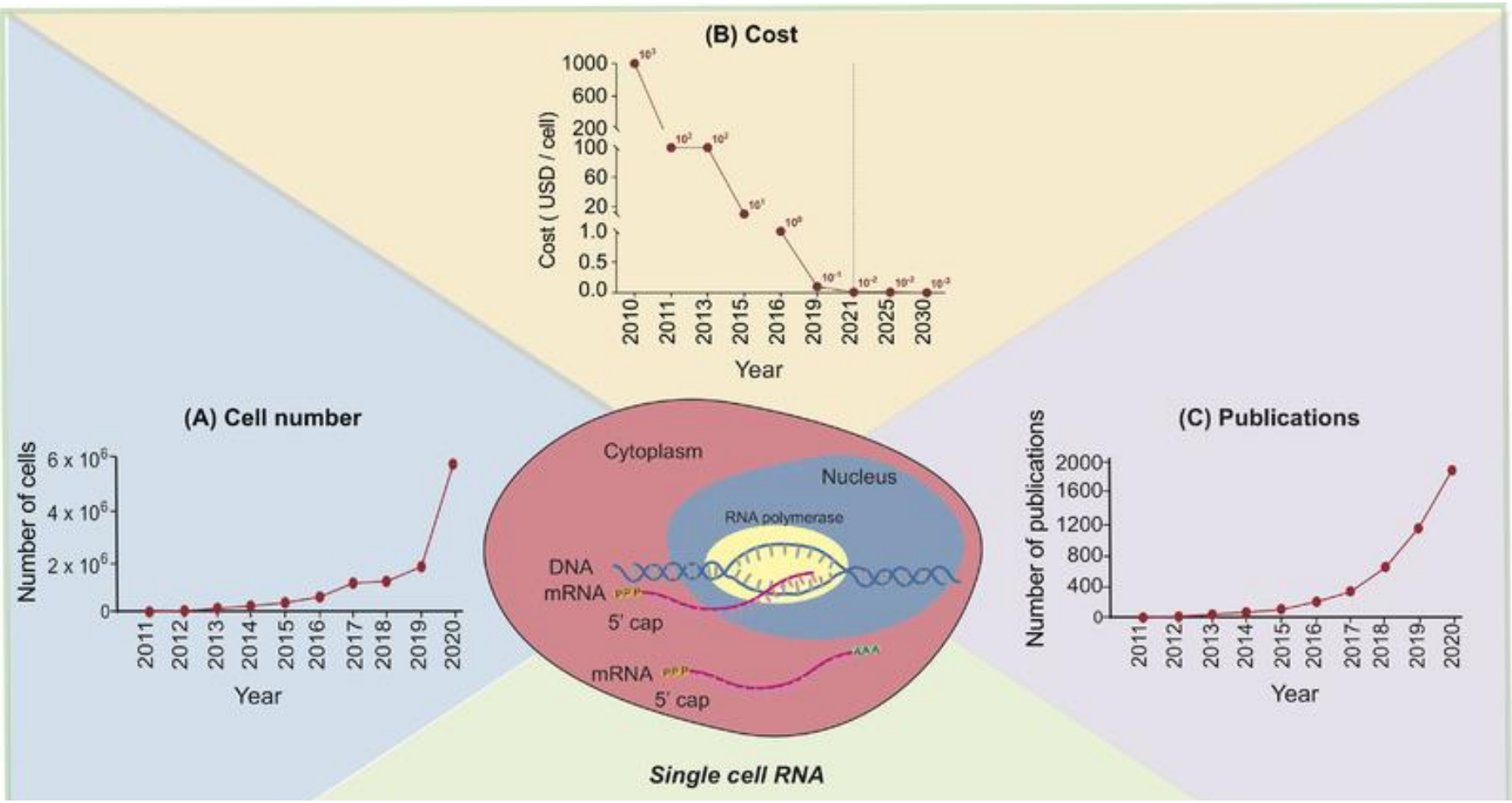
- Adipose
- Breast
- Development
- Eye
- Genetic Diversity
- Gut
- Heart
- Immune
- Kidney
- Liver
- Lung
- Musculoskeletal
- Nervous System
- Oral & Craniofacial
- Organoid
- Pancreas
- Reproduction
- Skin

Droplet Single Cell RNA-Seq (Drop-Seq)



Macosko et al, Cell 2015

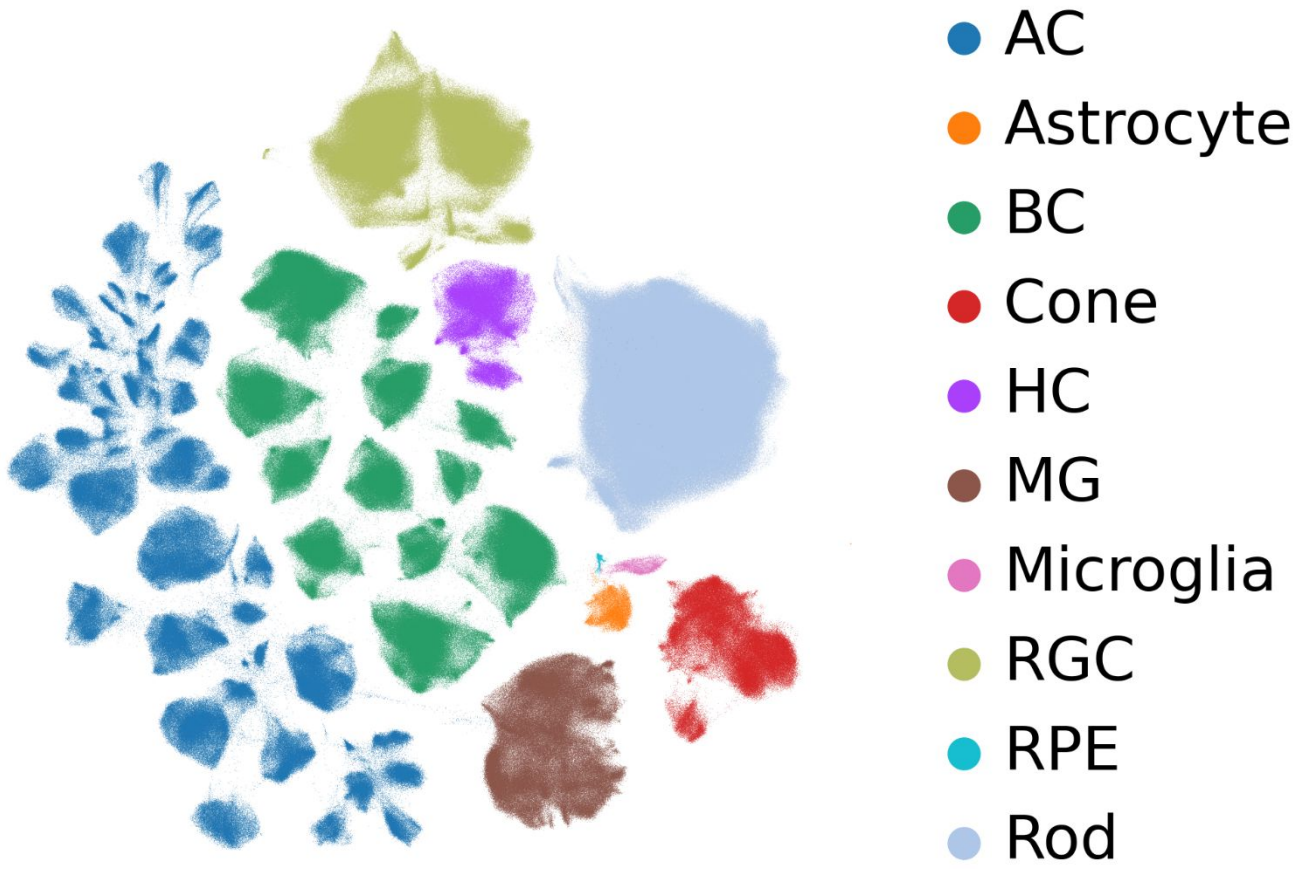
Rapid Development of Single Cell Transcriptomics Technologies



Clin Transl Med. 2022 Mar; 12(3): e694. Jovic et. al.

HUMAN CELL ATLAS

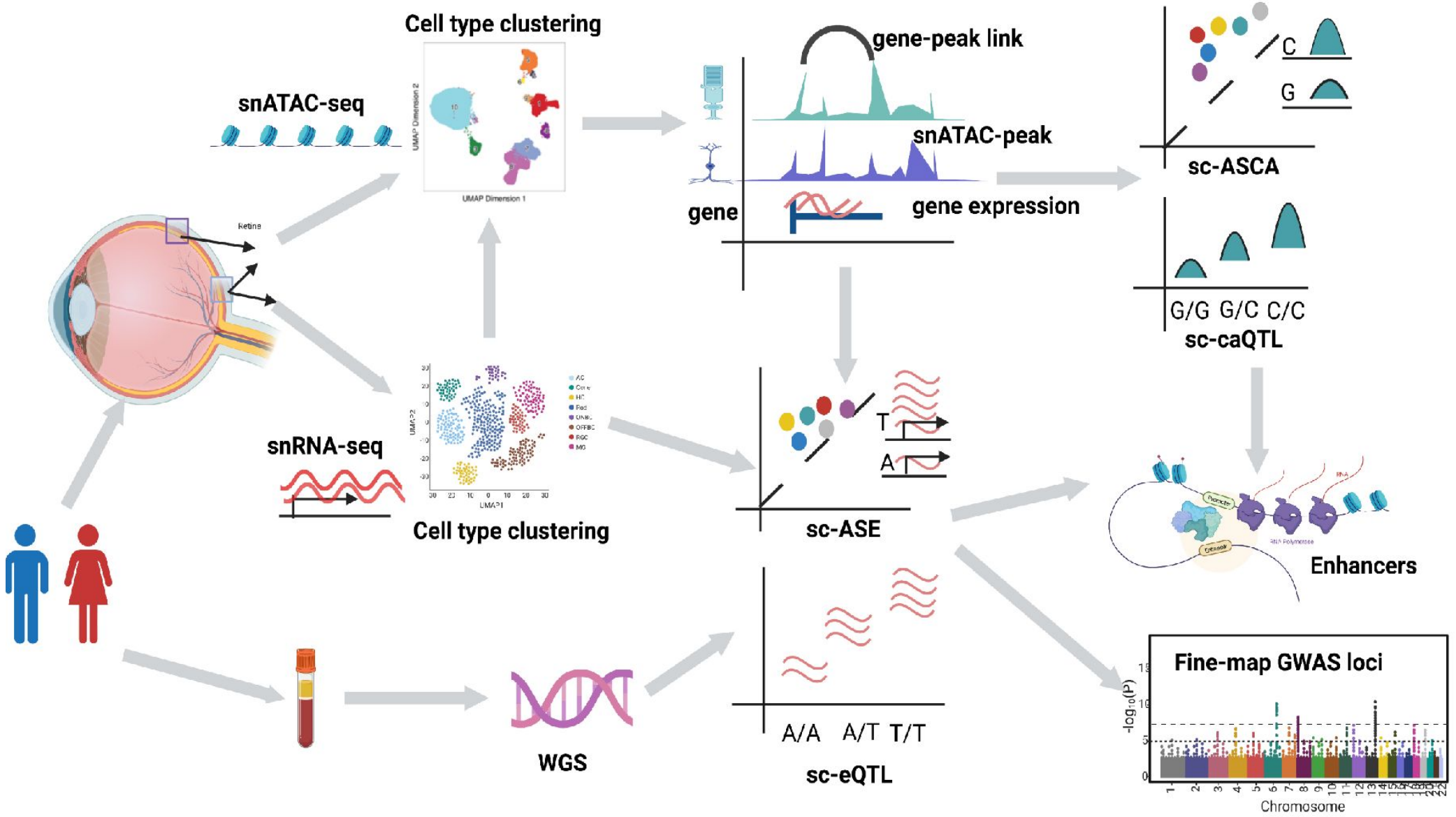
Single Cell Atlas References of the Human Retina



3.4M total number of single cell/nuclei from 122 donors;

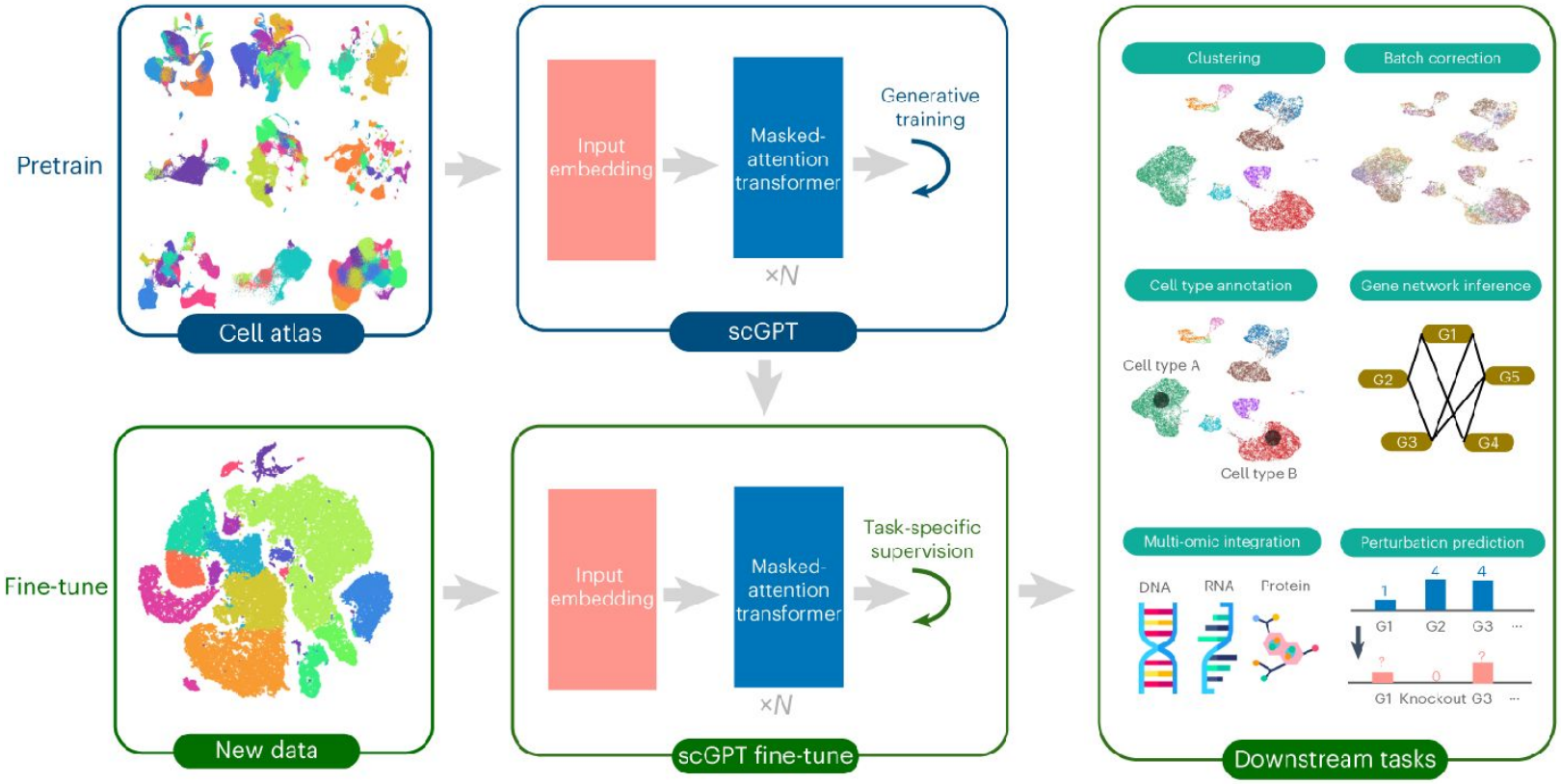
Over 120 cell types identified

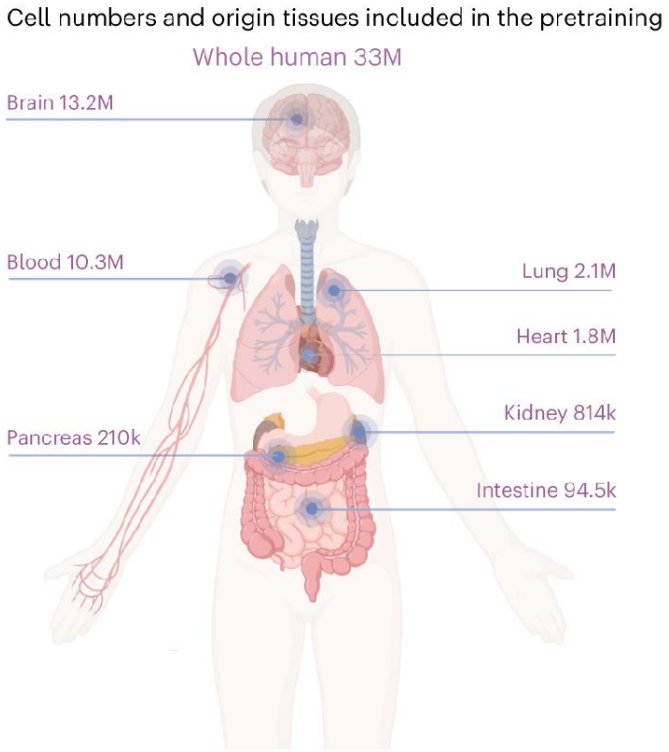
Integrative analysis by combining genomics and single cell multi-omics



Single-cell GPT (scGPT)

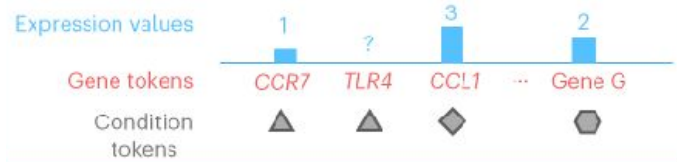
- Pre-training: over 33 million normal human cells
- Fine-tuning: for downstream tasks



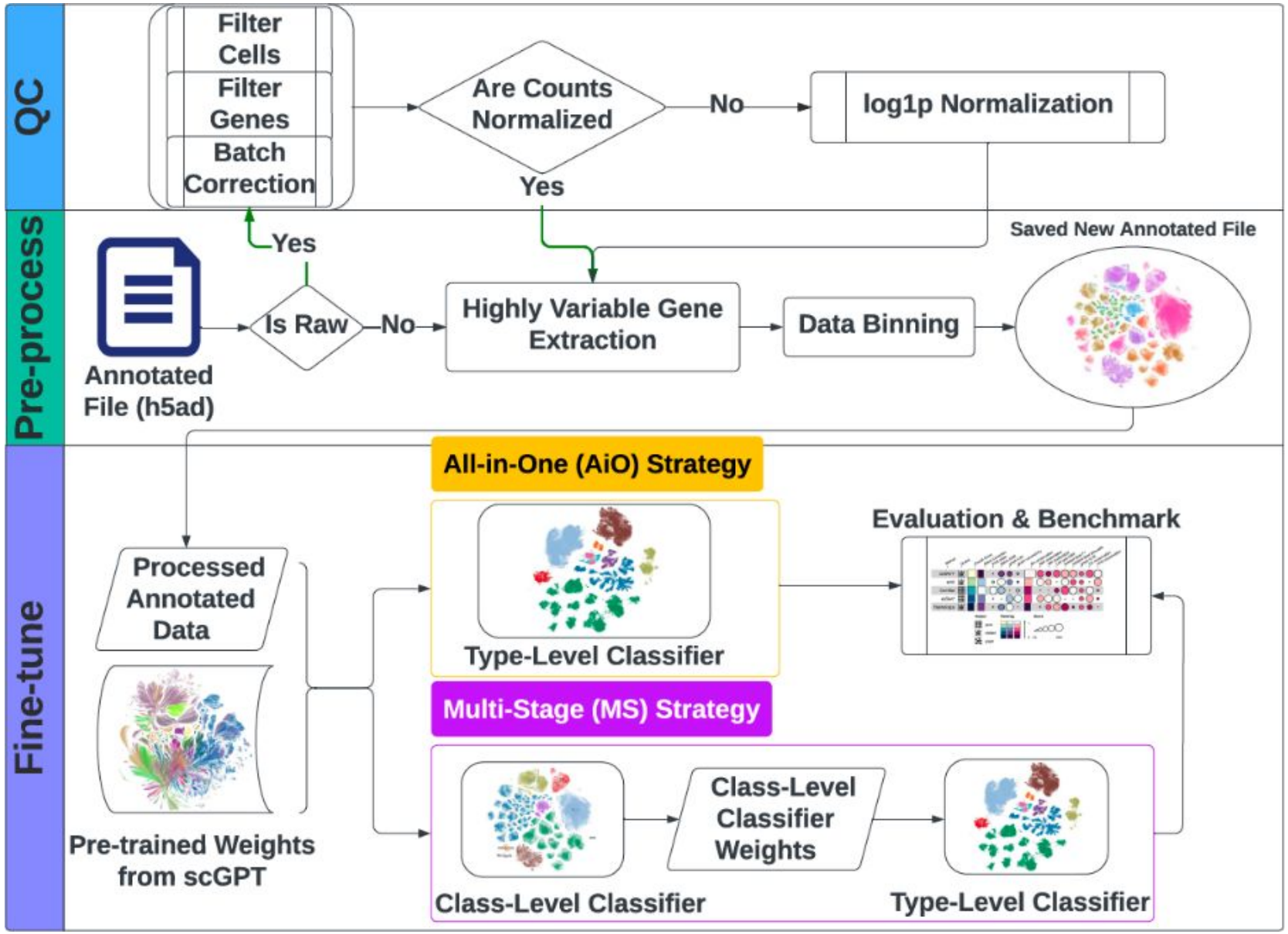


51 organs/tissues from the CZ CellxGene collection

Input components: (1) gene as token
(2) expression values
(3) condition tokens (modality, batch, perturbation conditions, etc.)



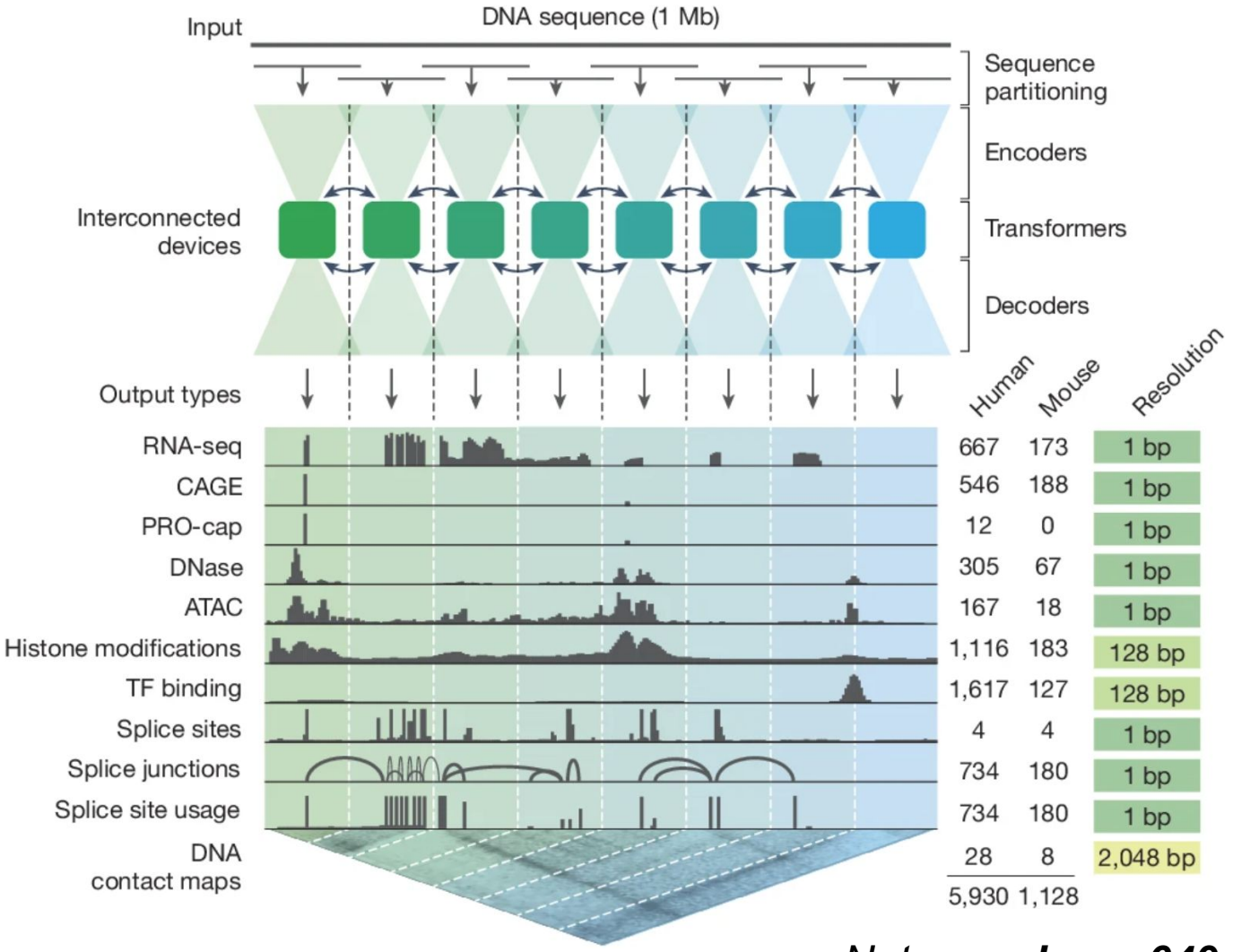
Overall Flowchart of Generating RetinaGPT via Fine Tuning



Performance of Fine Tuning

Macro Metrics	ZS scGPT	FT scGPT	FT scArches	FT Geneformer
Class-Level Predictions (10 Cell Classes)				
Accuracy	98.1	99.9	99.8	99.9
Precision	98.9	99.9	99.9	99.8
Recall	85.5	99.9	99.8	99.9
Macro F1	90.5	99.9	99.8	99.9
Inference Time (s)	~ 110	~ 82	~ 1090	~ 834
Type-Level Predictions (14 BC Types)				
Accuracy	38.8	99.6	99.5	99.6
Precision	60.0	99.2	87.0	99.7
Recall	28.1	99.0	85.6	98.9
Macro F1	38.2	99.1	86.3	99.3
Inference Time (s)	~87	~ 80	~ 1051	~ 830

AlphaGenome: Latest foundation model for predicting DNA sequence function



Nature ***volume 649****, pages1206–1218 (2026)*

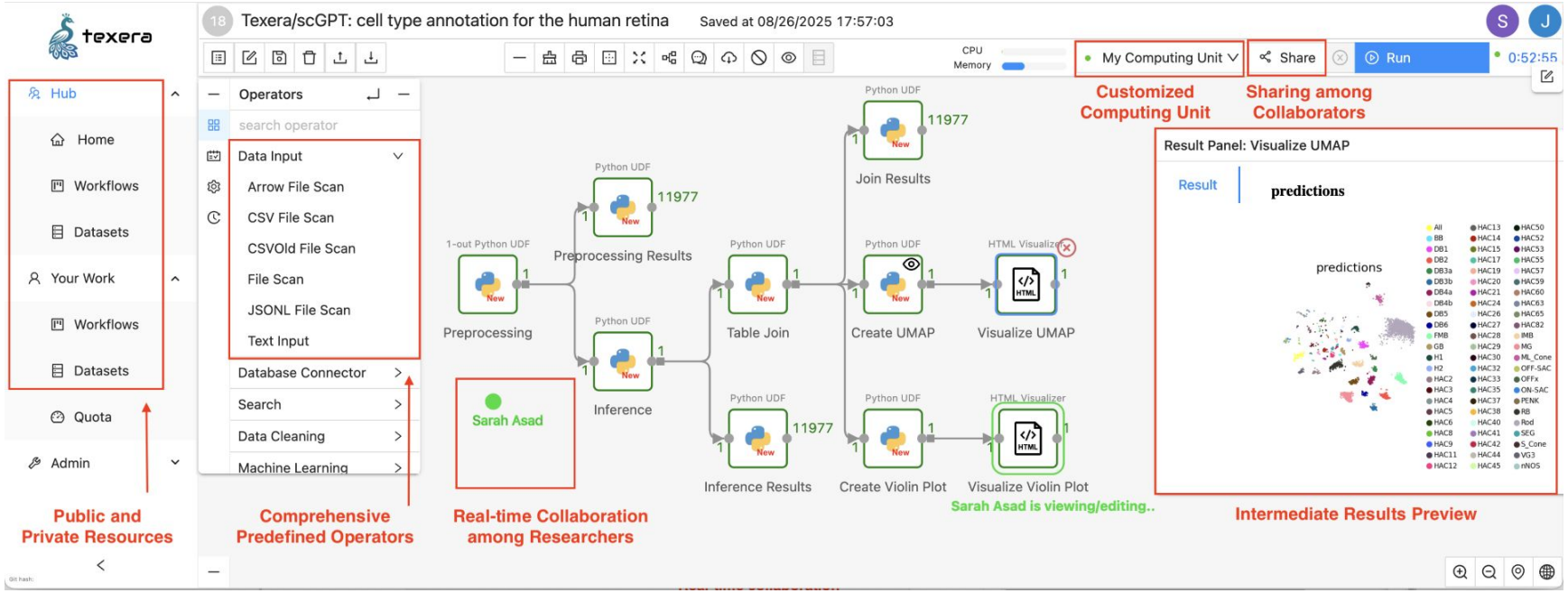
Fine Tuning of AlphaGenome for Chromatin Accessibility in RPE

	r 1 bp	r 128 bp	AUPRC	RPE-spec. ρ
zero-shot	0.200	0.492	0.249	0.33
head only	0.552	0.810	0.424	0.29
head only, 10 % data	0.556	0.810	0.427	0.29
decoder unfrozen	0.589	0.845	**0.464**	**0.41**
blocks 6–8 + decoder	**0.595**	**0.846**	0.456	0.37
shuffled labels (control)	0.110	0.149	0.061	0.03

Develop customized analytical workflows



Wet Lab Researcher



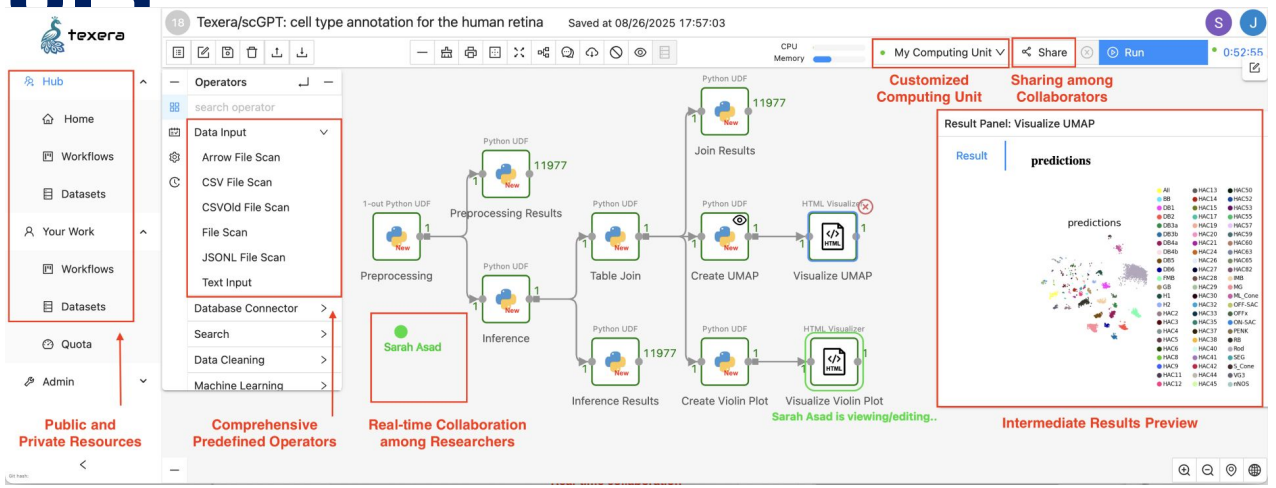
Develop workflows based on community needs.

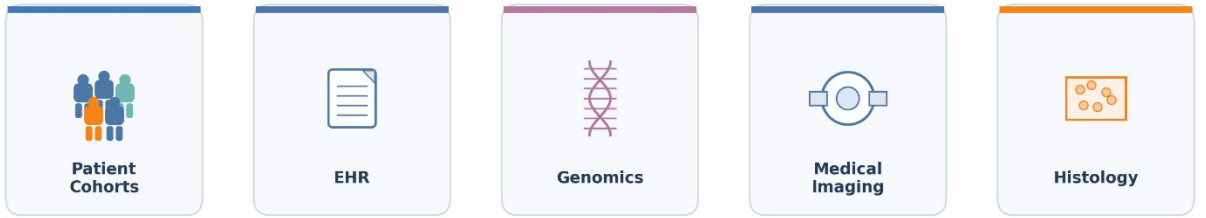
Democratize access to advanced computational tools



Wet Lab Researcher

AI AGENT





Agent based custom workflow development for integrative analysis

Diverse data modalities

Collaboration among biomedical domains with the BRIDGE program

Shared Tools, Workflows, and Functionalities: Develop reusable tools, standardized workflows, and common functionalities across projects.

Integration of AI technologies: New analysis powered AI tools and foundation models.

Shared AI Agent Framework: Establish a unified framework for designing, developing, and deploying AI agents across different biomedical applications.

ADMAP Data Resource Platform: Overview and Existing Features

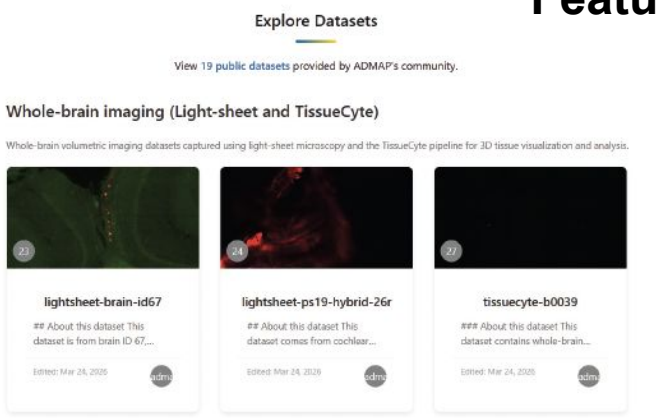


Fig 2. The ADMAP front page showcases various dataset modalities, with accompanying metadata information. Each entry can be clicked on and interacted with for downloading and visualization.



Fig 3. A visual image gallery integrated into ADMAP's data repository, depicting U24 rat histology sections, injected with rabies virus and AAV helpers. The application supports displaying rich metadata content and showcases controls to manipulate color channels.

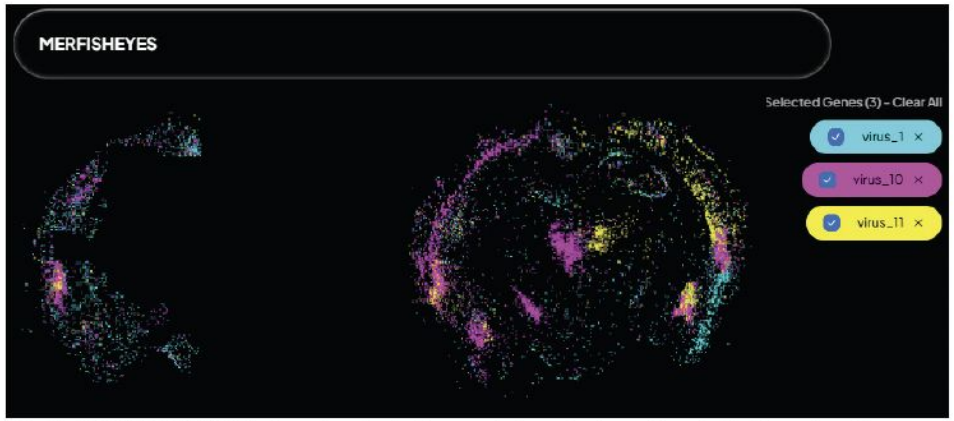


Fig 4. MERFISHEYES integration for high-resolution single-cell and single-molecule spatial transcriptomics visualization is shown for a sample MERFISH dataset.

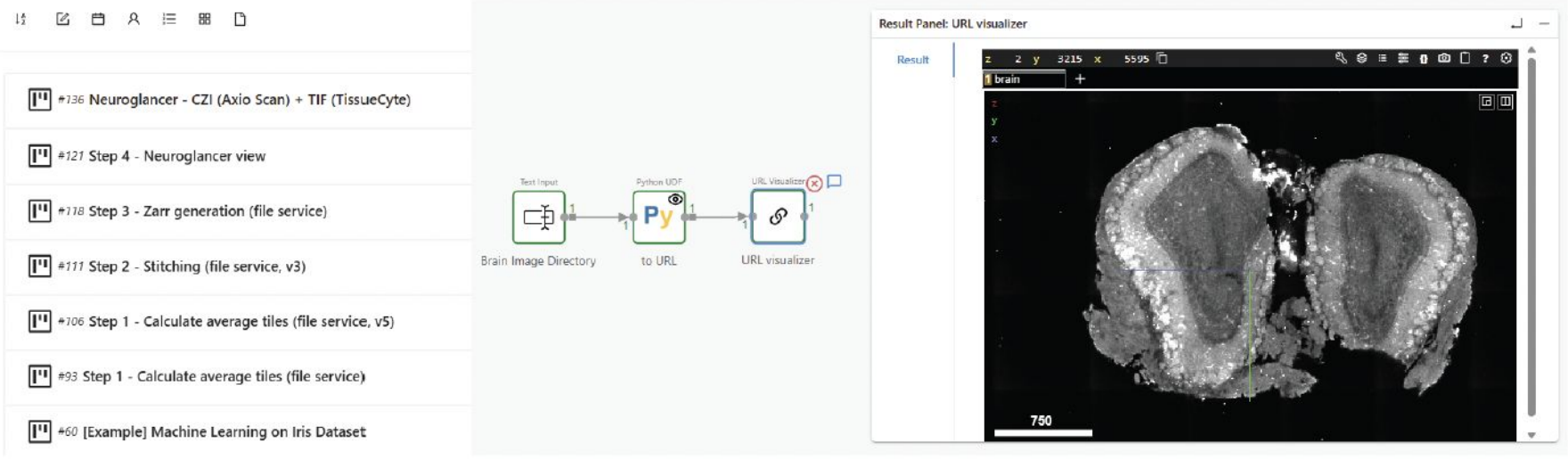


Fig 5. Texera workflows are integrated into ADMAP with dedicated computing resources for fast and parallelized data processing tasks. Users can select from a list of premade workflows, copy them to their account, customize the workflows to their data, and run them. Depicted is a sample visualization workflow to visualize a TissueCyte mouse brain section in Neuroglancer.

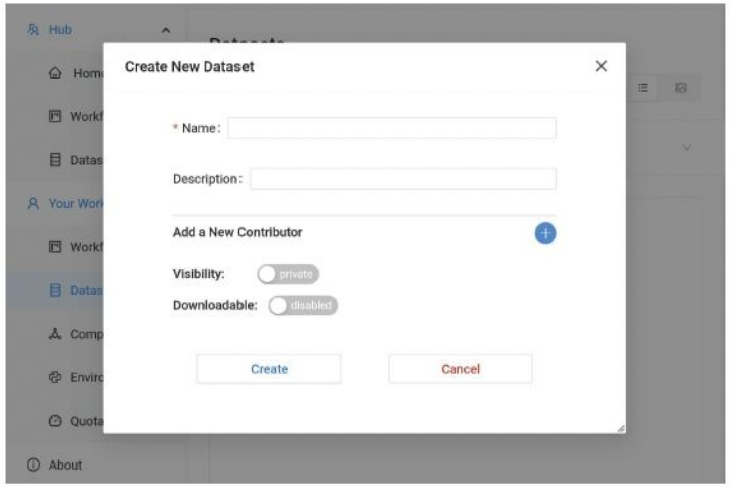


Fig 6. ADMAP allows for registered contributors to create, upload, and share their datasets and corresponding metadata on the platform.

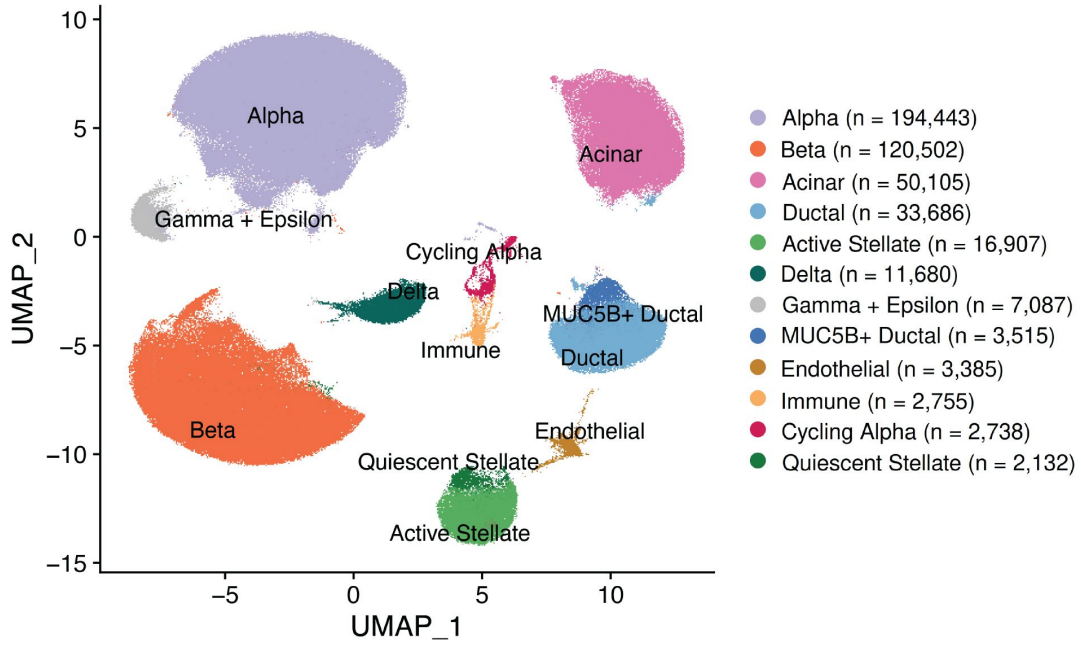
ADAMP + BRIDGE

Our goal is to develop and establish a Texera-based cloud computational platform that enables the neuroscience and Alzheimer's disease research community to efficiently share their data science and AI knowledge, including datasets, analysis workflows, and ML models, and use cloud resources to run computational tasks.

Building on the existing ADMAP/NeuroMAP data repository of the UCI Center for Neural Circuit Mapping (CNCM), the platform will support **the dissemination of datasets, computational analysis workflows, machine learning (ML) models, and associated metadata**, while providing scalable cloud computing resources for reproducible data analysis and collaborative research. The project will foster an open and collaborative ecosystem that accelerates AI-driven neuroscience and Alzheimer's disease research through community engagement, resource sharing, and continuous platform improvement.

This work will leverage the extensive computational and collaborative infrastructure of CNCM, which serves as a hub for technology development, data sharing, and interdisciplinary neuroscience research, providing a strong foundation for community engagement and dissemination.

Bigger data reveals hidden biology



Atlas of ~450,000 islet cells from 140 donors; each dot is one cell, colored by cell type (PanKbase; Vu et al., bioRxiv 2026)

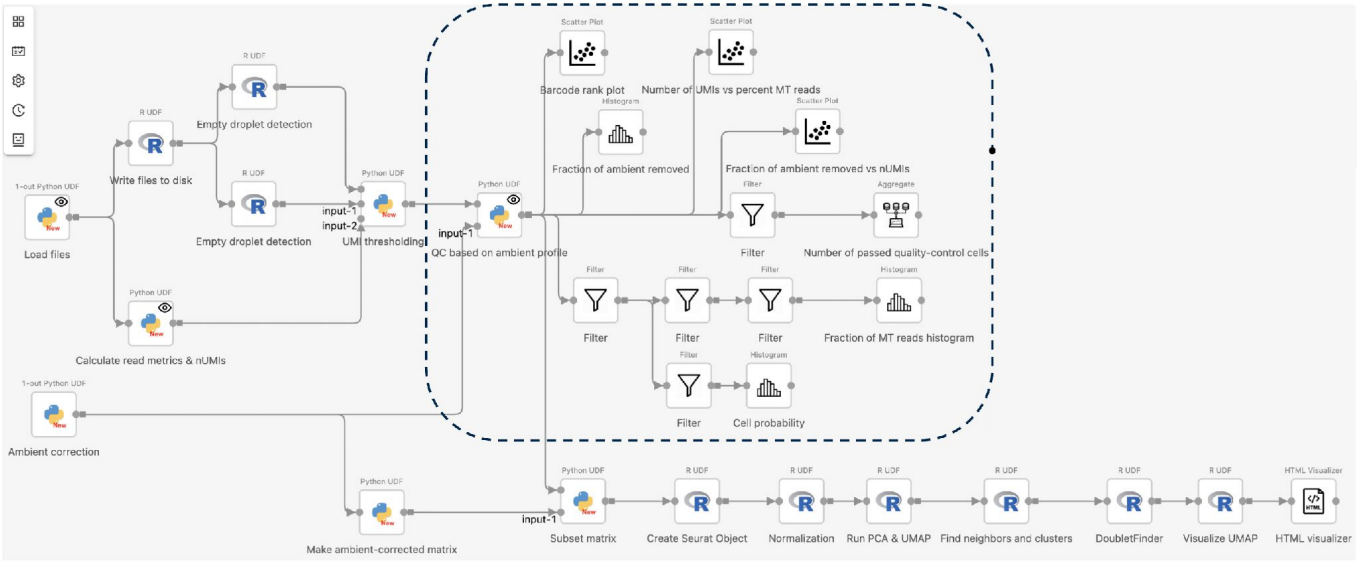
~450,000 islet cells, 140 donors
PanKbase islet atlas (Vu et al., preprint 2026)

~500,000 muscle nuclei, 287 people
Varshney et al., Nature Genetics (in press)

14,324 samples, 10,195 people
Orchard et al., Science 2026: found the 2nd, 3rd and 4th switches controlling genes and linked 10,611 genetic signals for 164 traits to genes

The bottleneck: up to 20 TB per study and dozens of steps on supercomputers. Most biologists can't run or adapt these analyses alone.

How Texera helps, and where we can collaborate



A Texera workflow built by Ha Vu: each box is one analysis step. Steps run cell quality checks on the islet scRNA atlas, then compare gene activity between groups of donors in insulin-making beta cells. Live demo at today's session.

Our goals

- Reusable workflows across tissues and data types
- Point-and-click forms so biologists run analyses without code
- Private human data kept secure on U-M HPC

Let's collaborate

- Share single-cell building blocks with the brain and vision domains
- Add AI models (e.g., scGPT, AlphaGenome) as workflow steps
- Share workflows openly with the wider research community

Thank you and questions

How do DNA differences shape metabolic traits?

Inherited

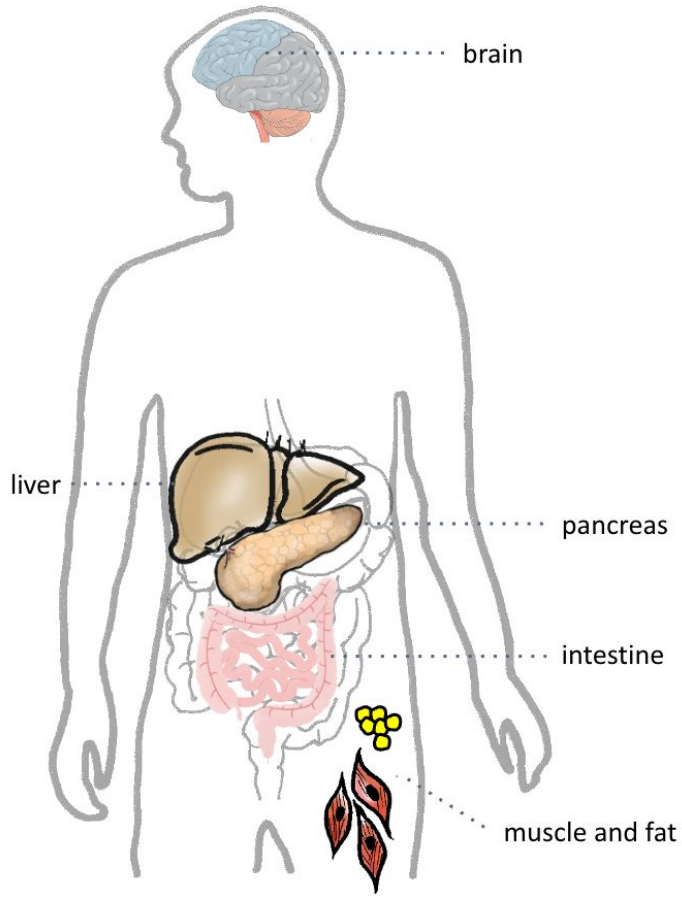
Metabolic traits, how our bodies make, store and use energy, differ between people, partly because of our DNA

1,000s

of DNA regions influence these traits, but most sit outside genes, in “switches” that turn genes up or down

Which cells?

We map these switches cell by cell in tissues that control energy use: pancreatic islets (make insulin) and muscle (uses sugar)



Functionality of the VisionKB

Build reference through meta analysis of single cell omics data for the vision domain.

Implement tools that allow easy access and explore of the reference object.

Generate fine tuned foundation models that expand functionalities.

Enable custom analysis and mining of the datasets.