CellTypist and CellHint: towards automated annotation and integration of single-cell data

Chuan Xu

05/03/2025

TEICHMANN LAB
Be Bold. Be Brilliant. Be Kind

Cambridge Stem Cell Institute

UNIVERSITY OF CAMBRIDGE

What Is the First Step Towards Understanding Organs?

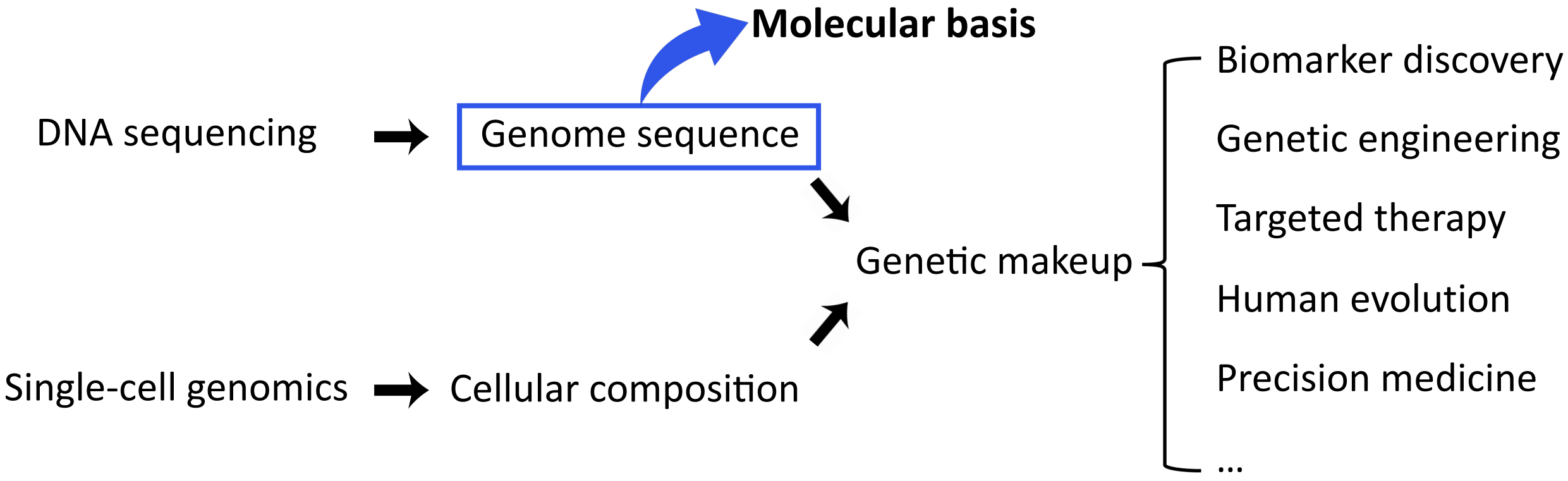
DNA sequencing → Genome sequence

Single-cell genomics → Cellular composition

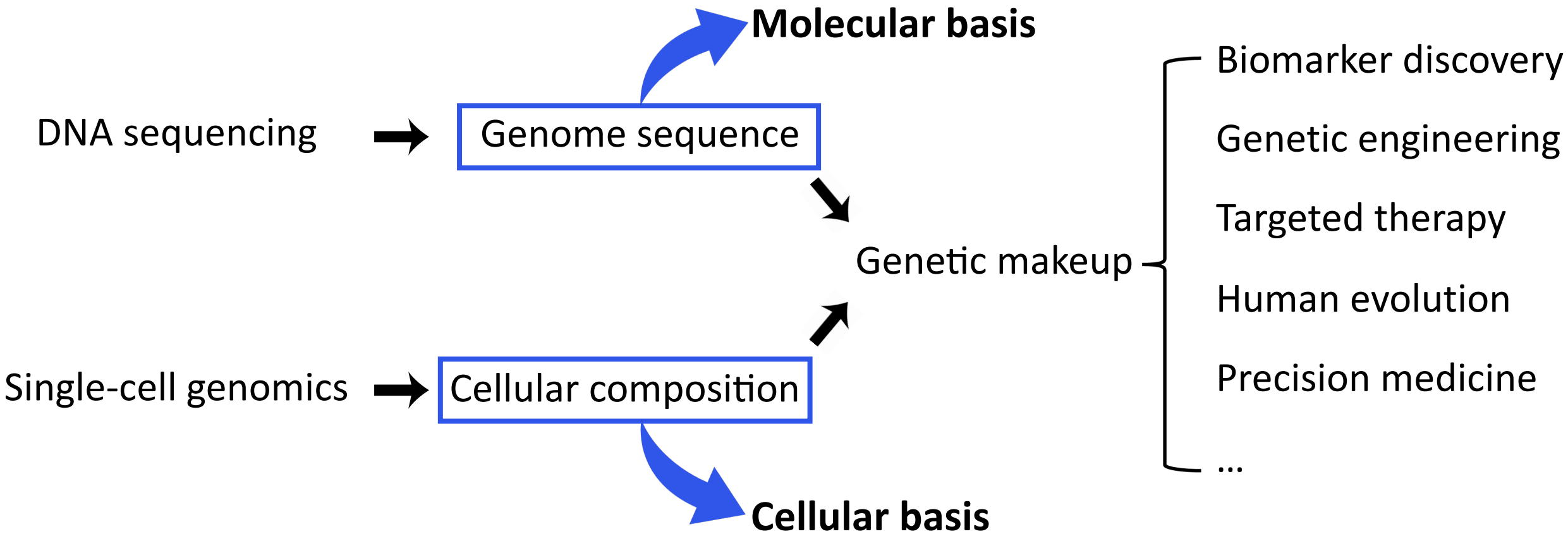
Genome sequence ↘ Genetic makeup ↗ Cellular composition

Genetic makeup:
- Biomarker discovery
- Genetic engineering
- Targeted therapy
- Human evolution
- Precision medicine
- …

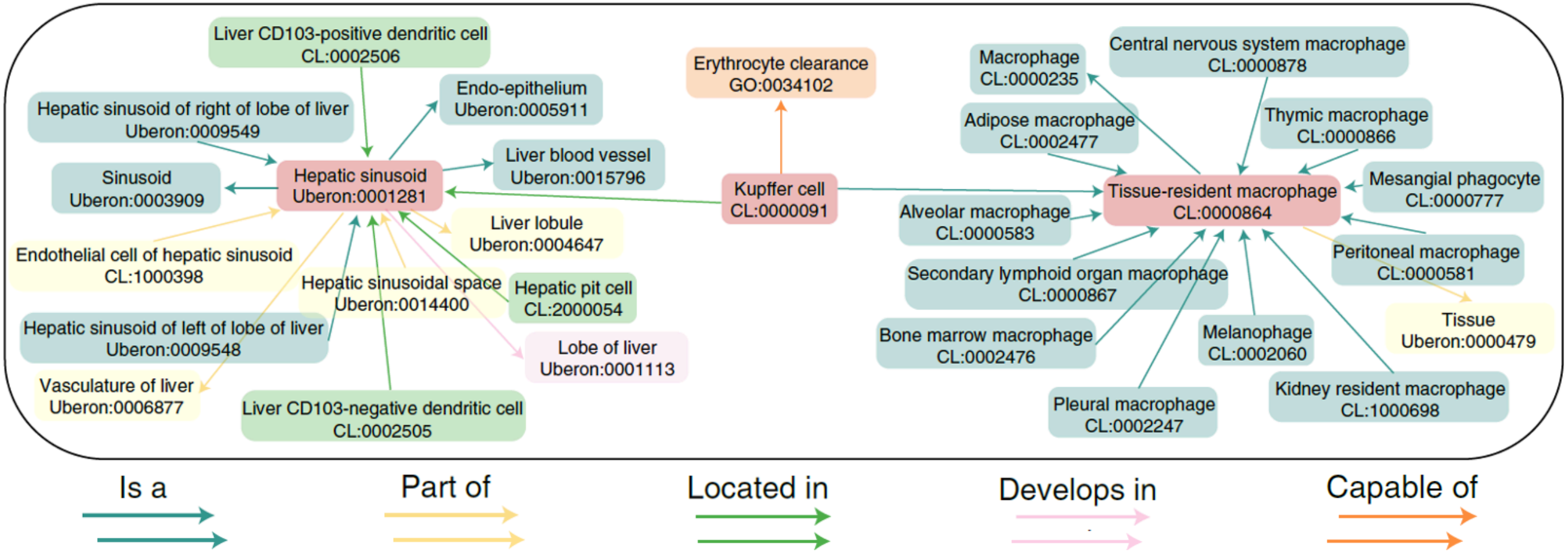
What Is the First Step Towards Understanding Organs?



What Is the First Step Towards Understanding Organs?



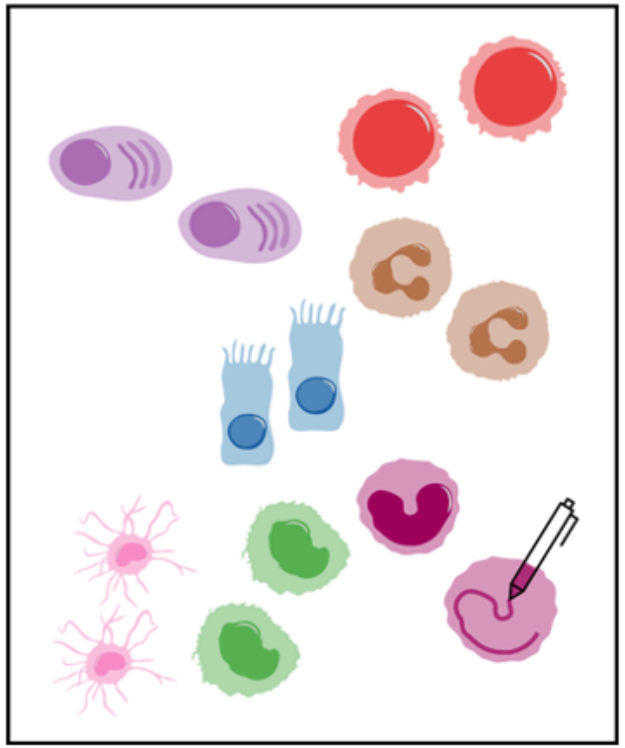
More Than Just a Catalog



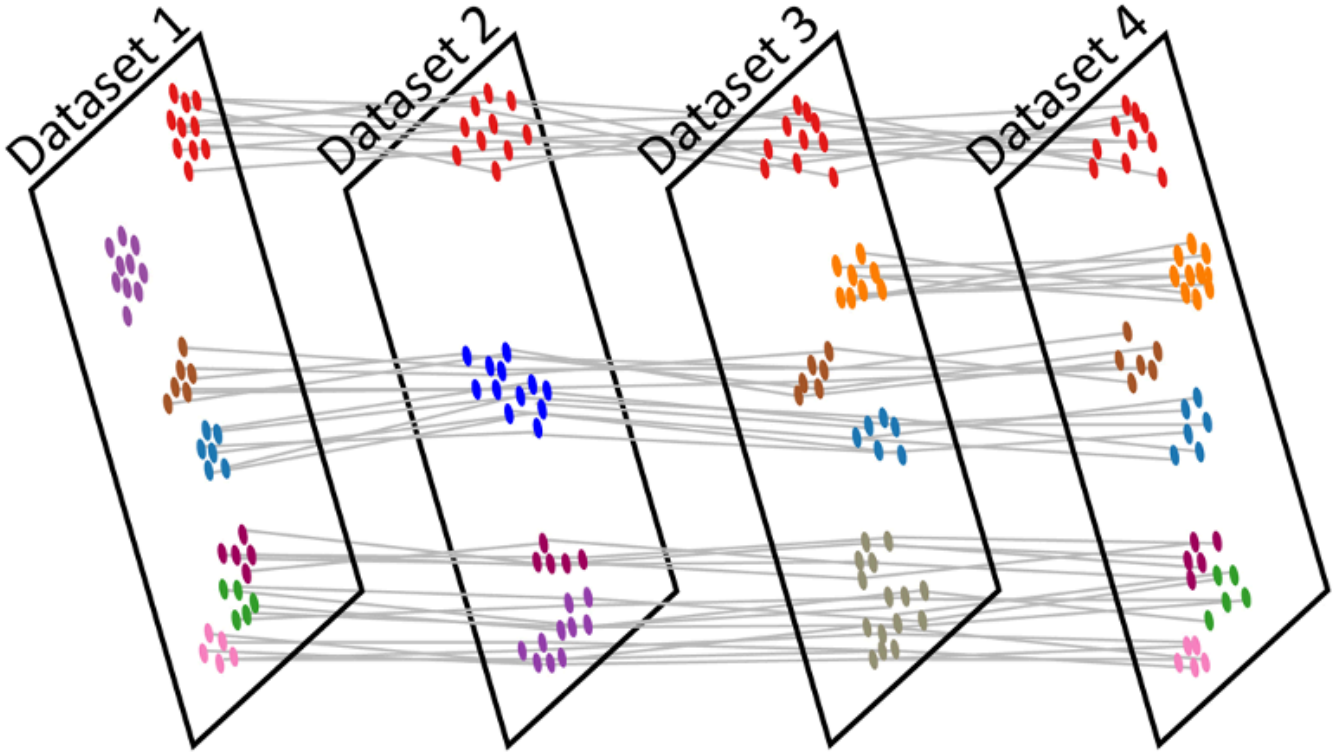
Sutherland, Xu, et al., Nature Cell Biology (review), 2021.

What Problems To Solve?

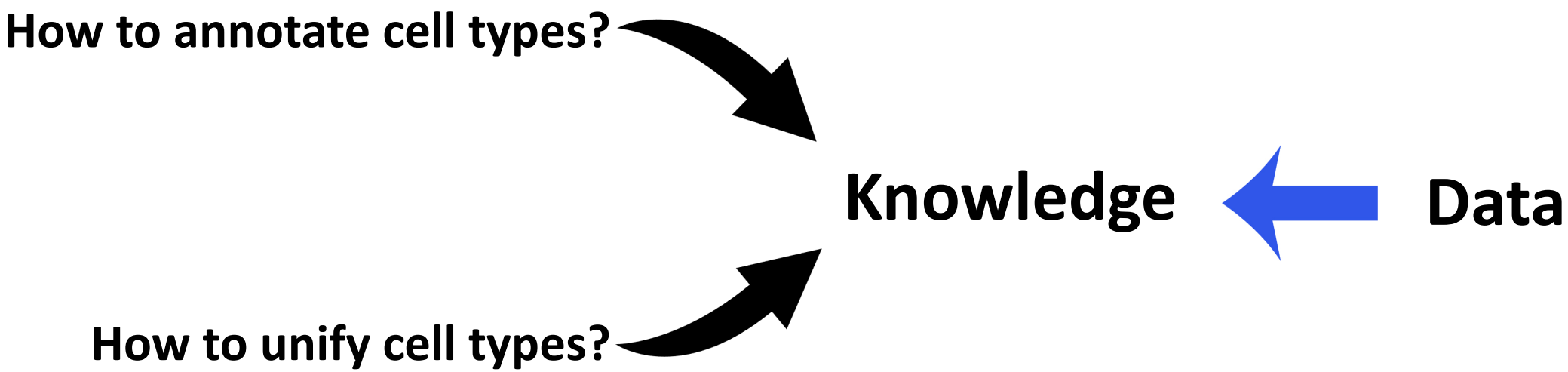
How to annotate cell types?



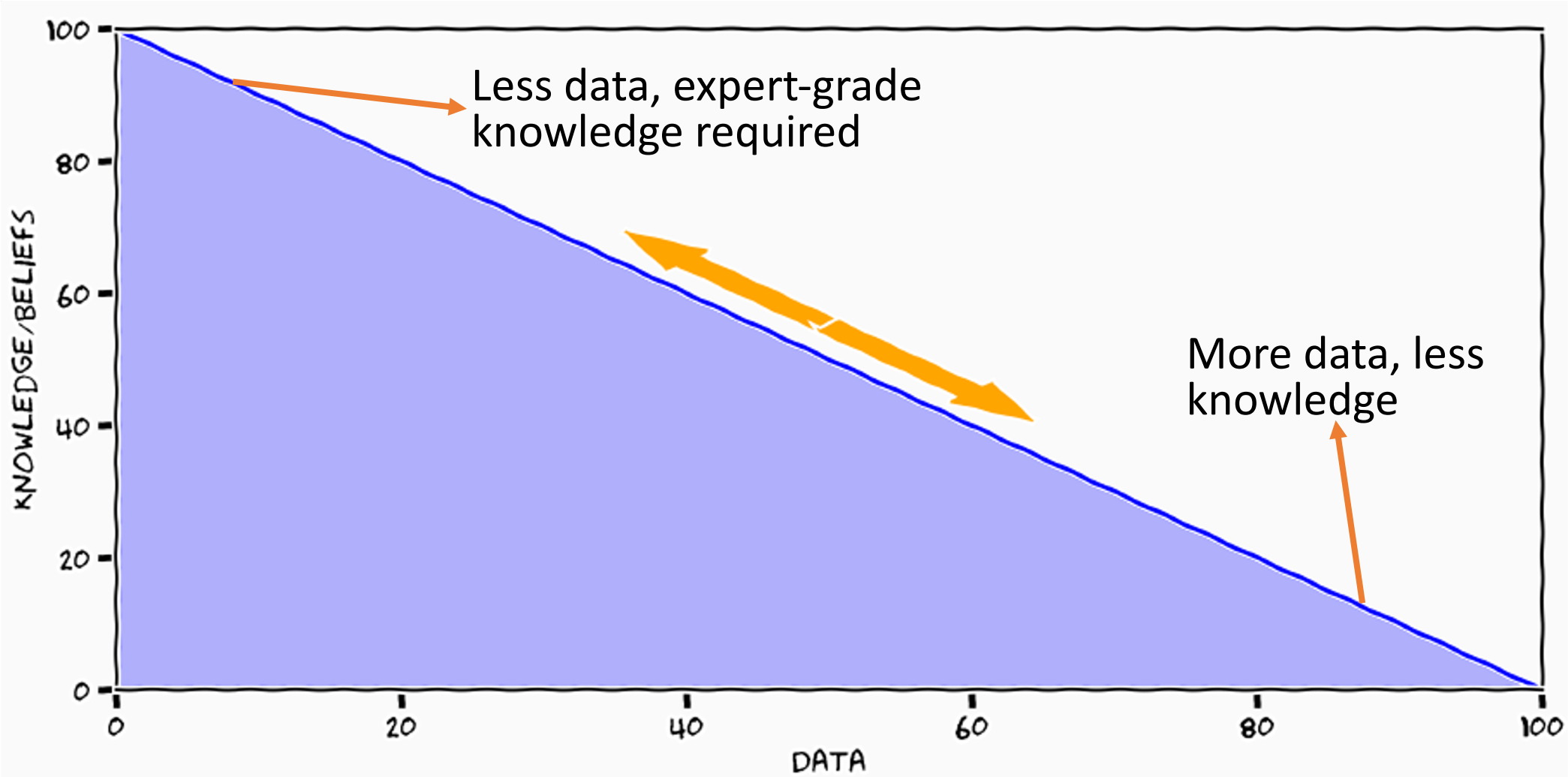
How to unify cell types?



How To Solve?


How to annotate cell types?
Knowledge
Data
How to unify cell types?

Why Big Data?



Adapted from Carl Henrik Ek, Accelerate Science, 2021.

How To Annotate Cell Types?

CellTypist

Website: celltypist.org

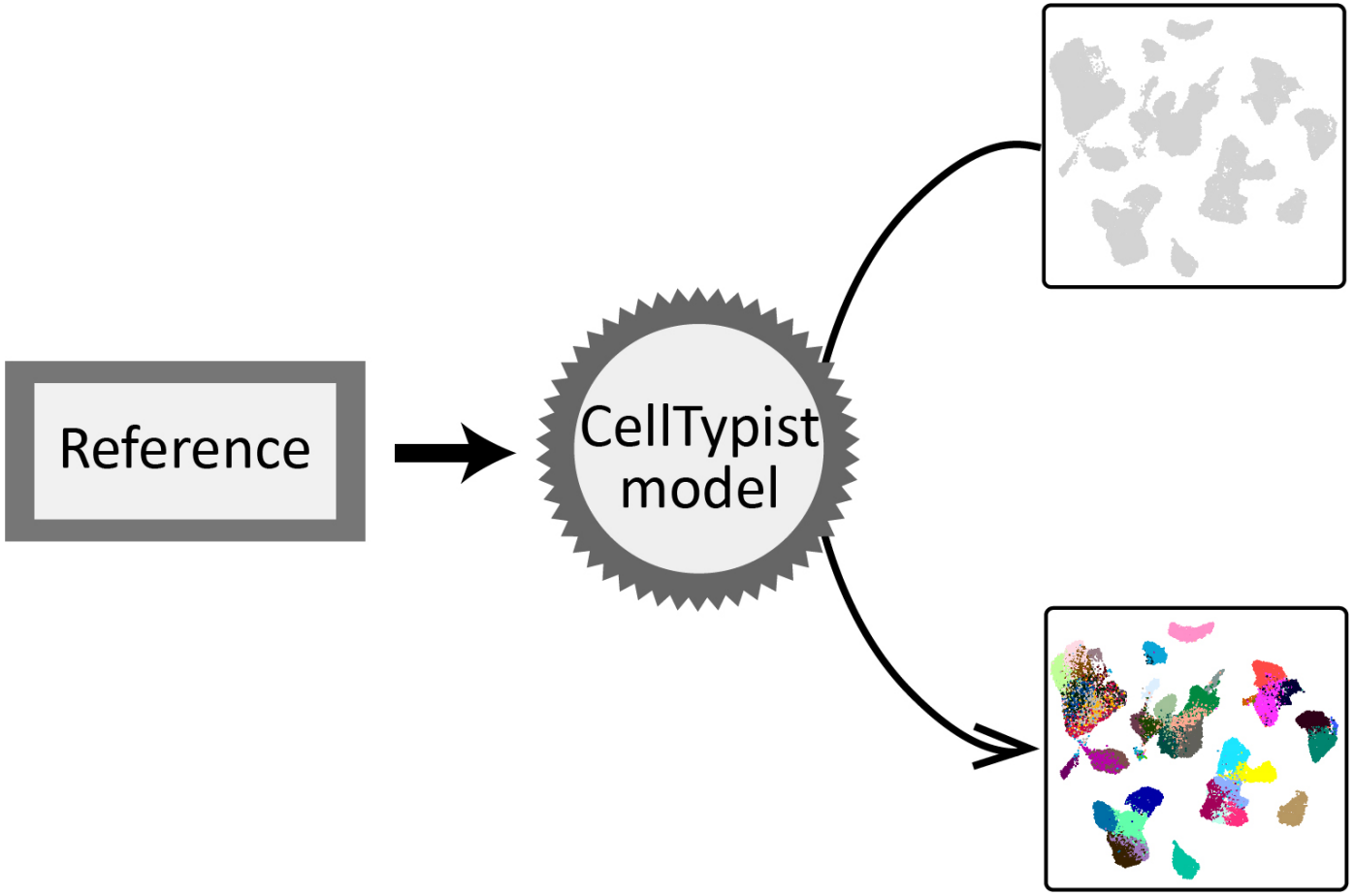
GitHub: Teichlab/celltypist

Tutorial: celltypist.readthedocs.io

Domínguez Conde, Xu*, et al., Science, 2022.*

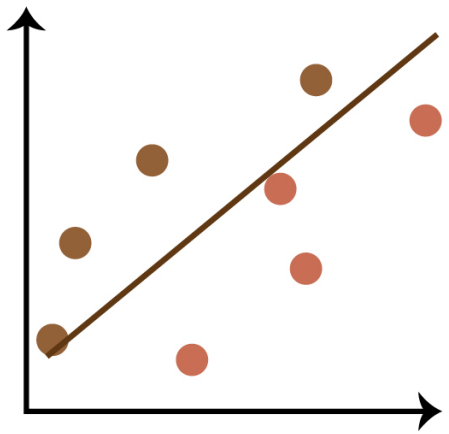
What Is CellTypist?

Model-based label transfer

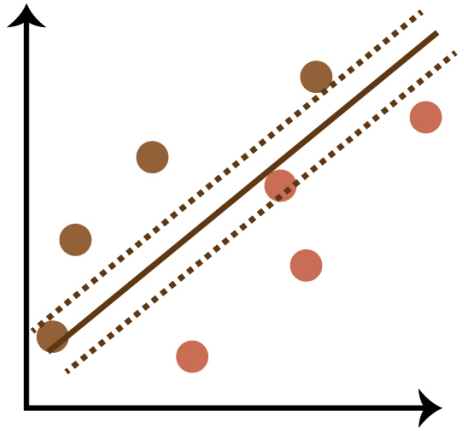


What Algorithm To Rely On?

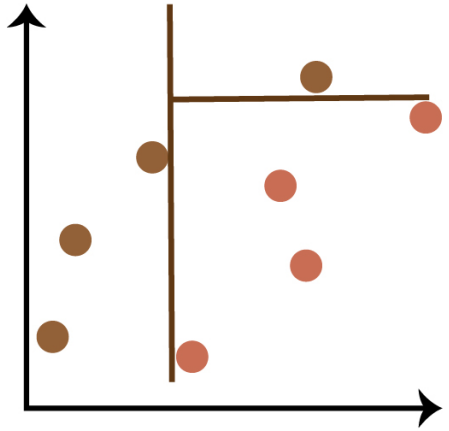
Logistic regression



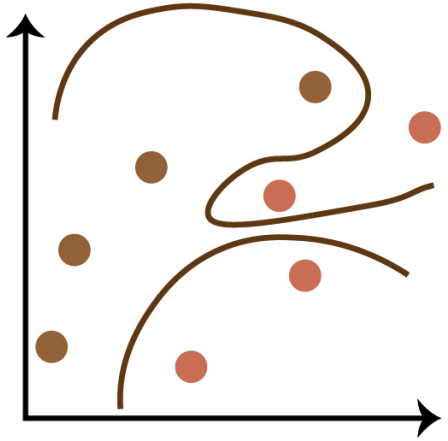
Linear support vector machine



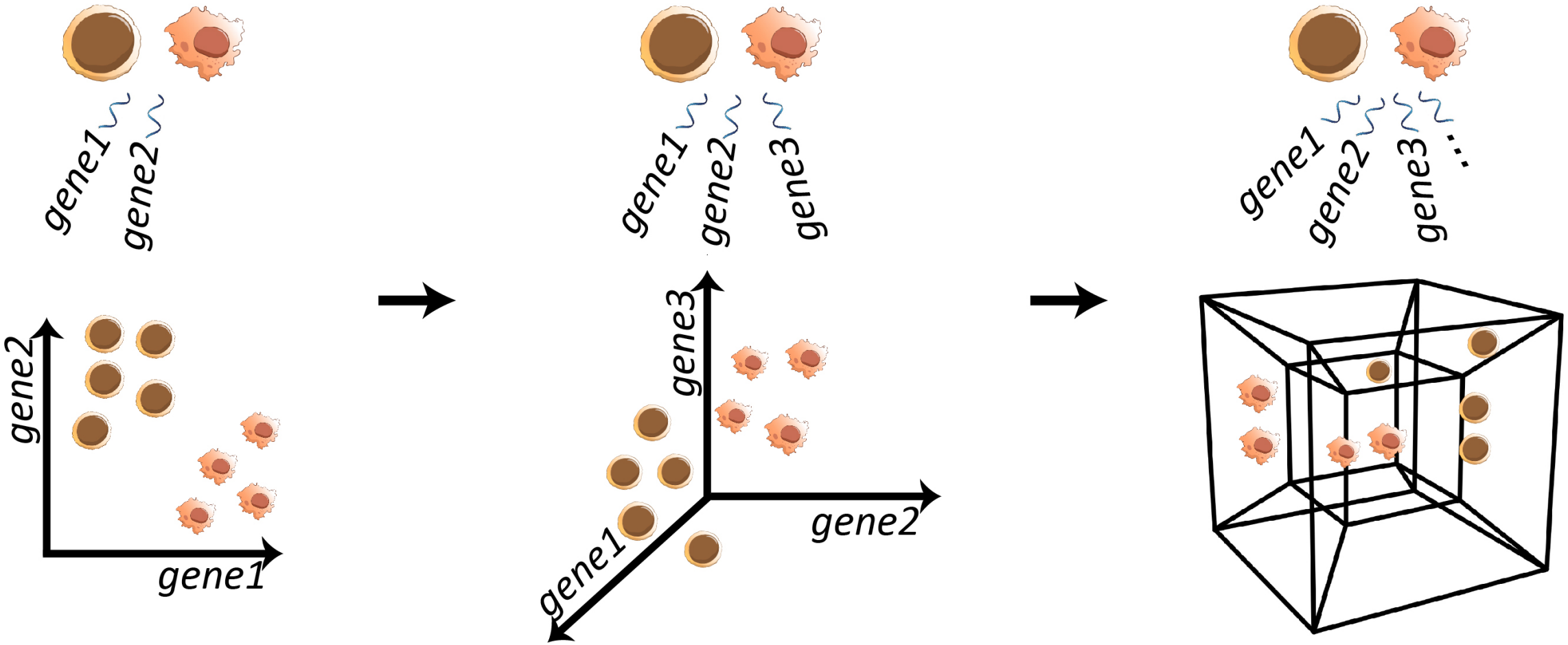
Decision tree & Random forest



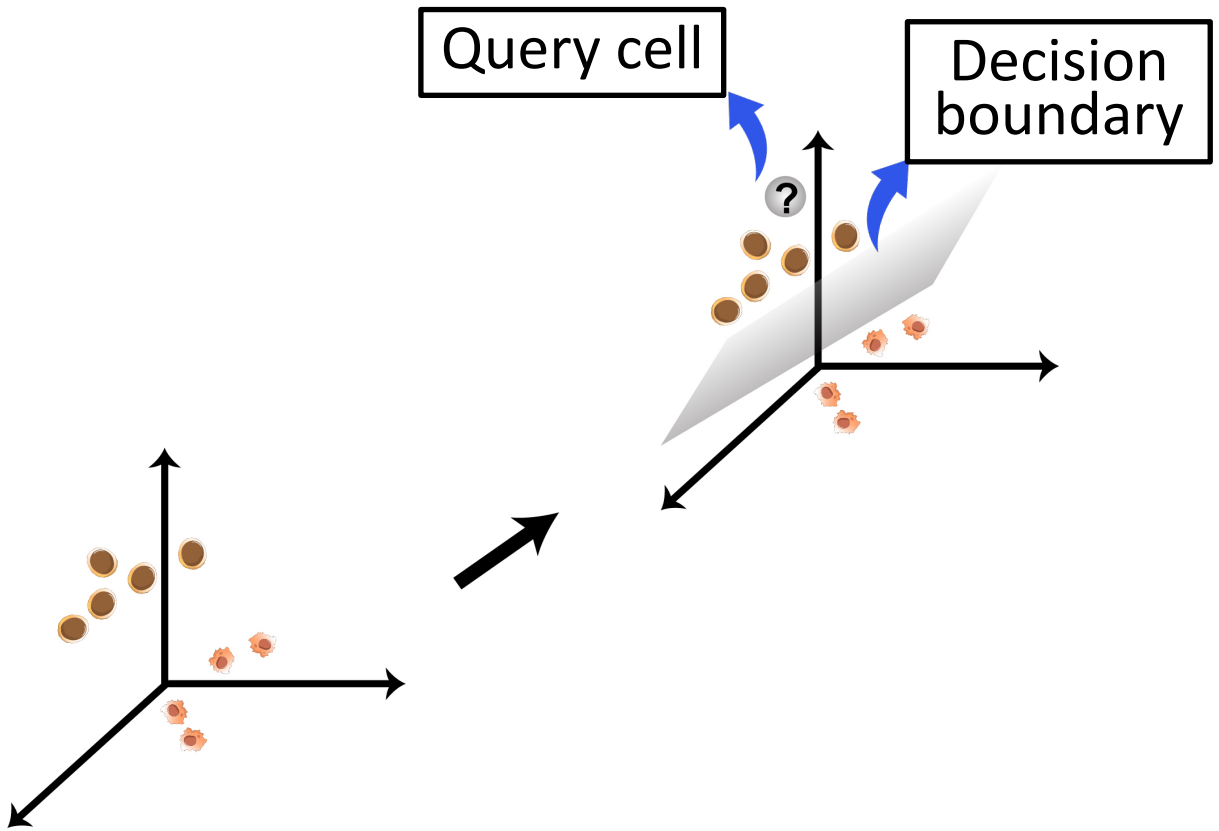
Neural network



“The curse of dimensionality”

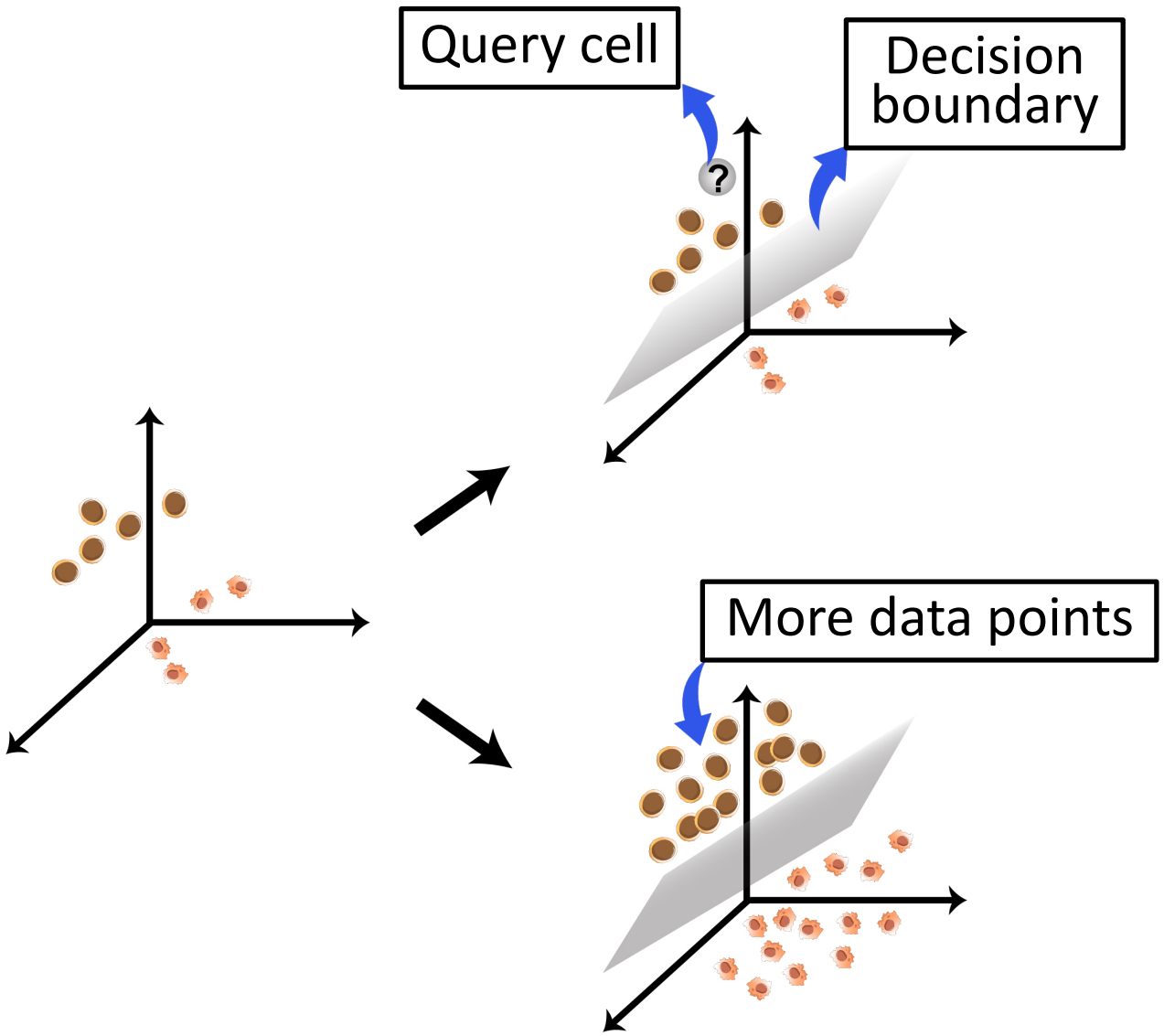


Interpretability



- Interpretable hyperplane
- Interpretable probability

Interpretability



- Interpretable hyperplane
- Interpretable probability

- Feature space coverage
- Robustness

Example: Cross-Tissue Immune Cells

Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy

scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation

Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation

Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis

Spatiotemporal immune zonation of the human kidney

Decoding human fetal liver haematopoiesis

Single-cell transcriptomics of the human retinal pigment epithelium and choroid in health and macular degeneration

A cell atlas of human thymic development defines T cell repertoire formation

Memory CD4+ T cells are generated in the human fetal intestine

Distinct microbial and immune niches of the human colon

Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease

A cellular census of human lungs identifies novel cell states in health and in asthma

Lineage tracking reveals dynamic relationships of T cells in colorectal cancer

Massively parallel digital transcriptional profiling of single cells

Single-cell reconstruction of the early maternal-fetal interface in humans

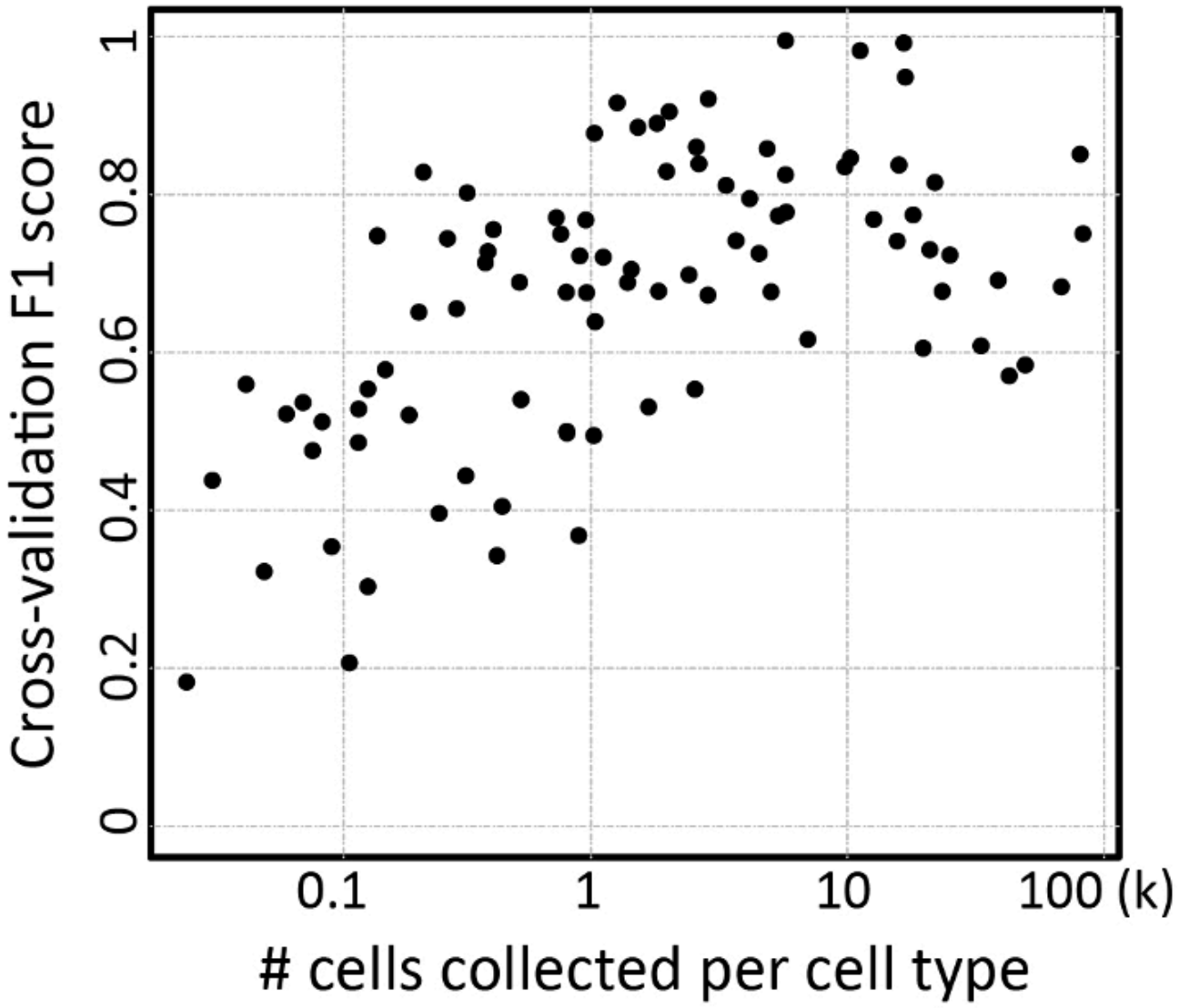
The heterogeneity of human CD127+ innate lymphoid cells revealed by single-cell RNA sequencing

***In Vitro* and *In Vivo* Development of the Human Airway at Single-Cell Resolution**

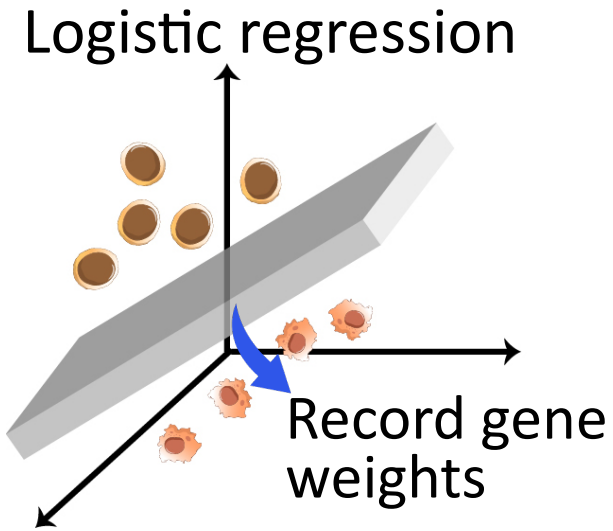
A single cell immune cell atlas of human hematopoietic system

Lipid-Associated Macrophages Control Metabolic Homeostasis in a Trem2-Dependent Manner

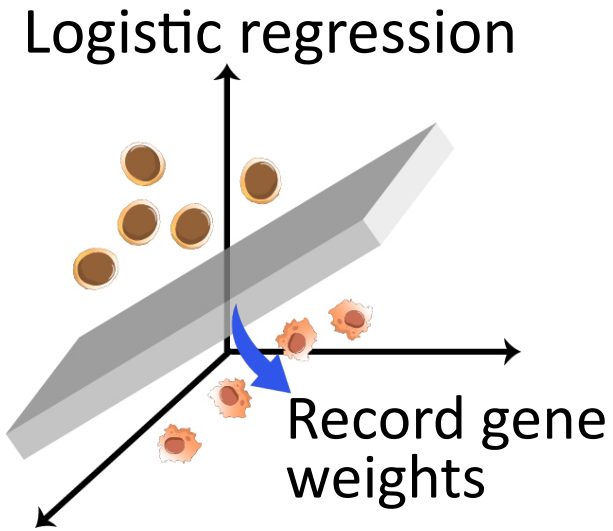
Power of Big Data

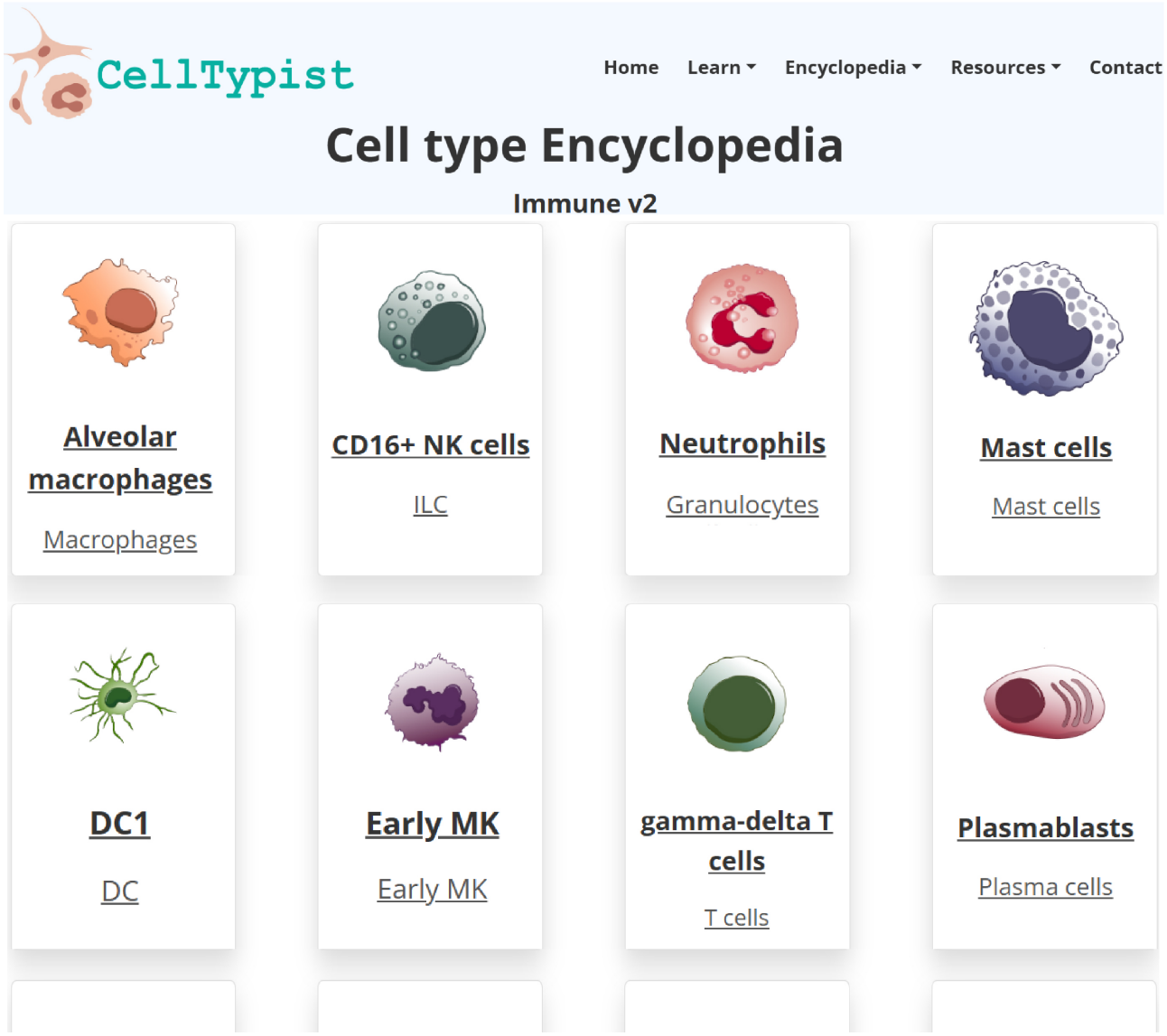


Lightweight & Shareability



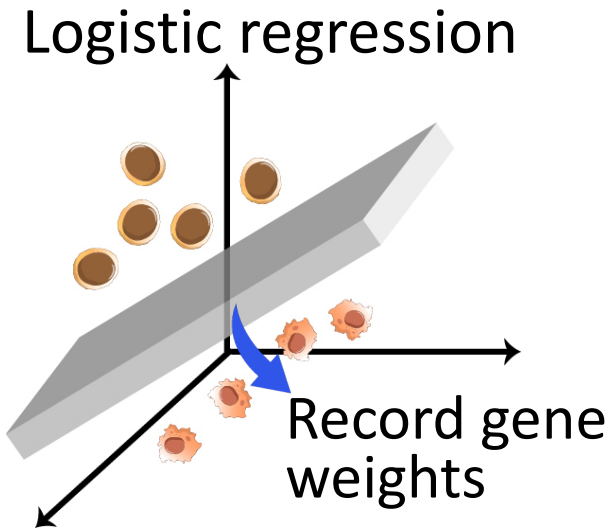
Cell Type “Encyclopedia”

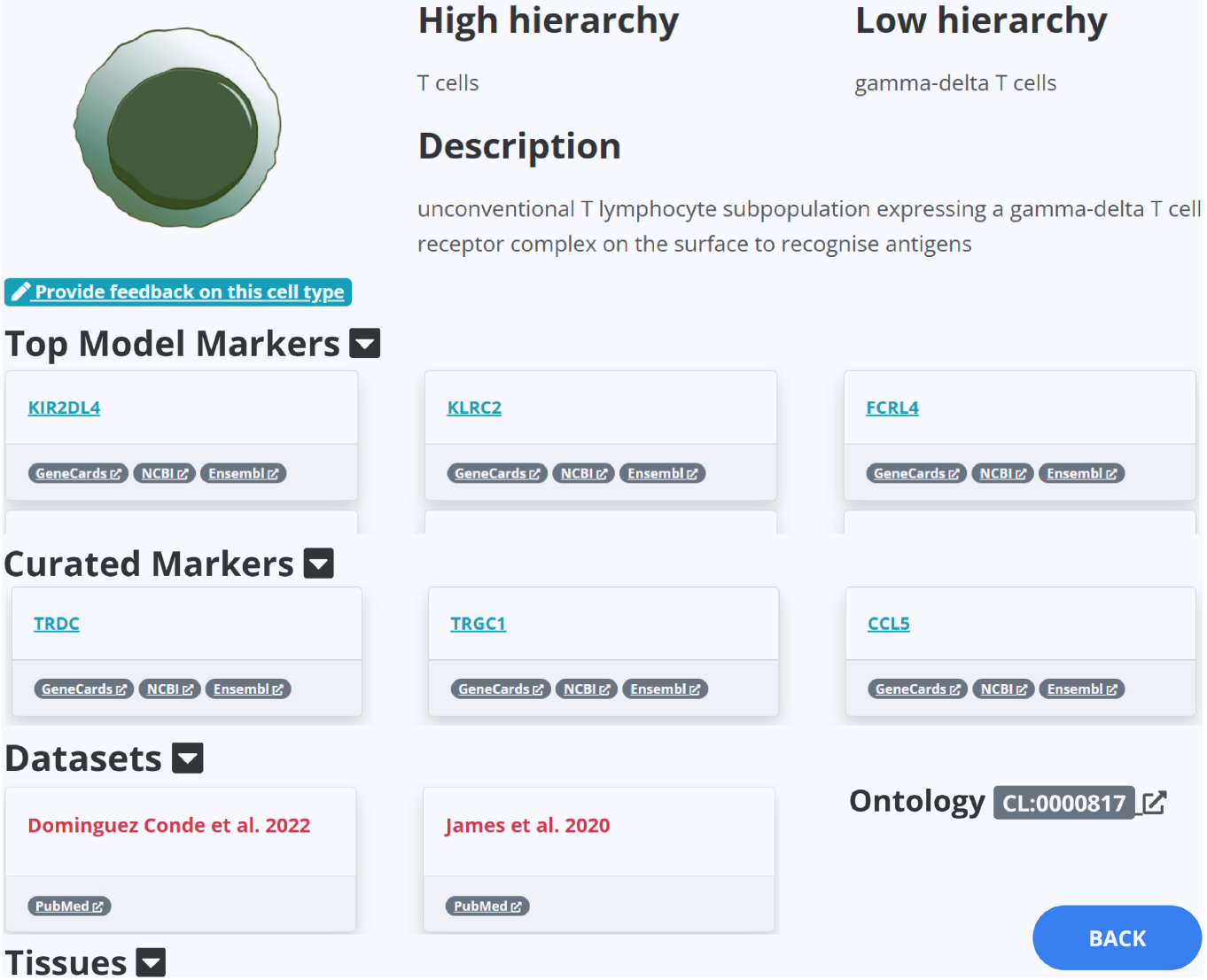




www.celltypist.org/encyclopedia/Immune

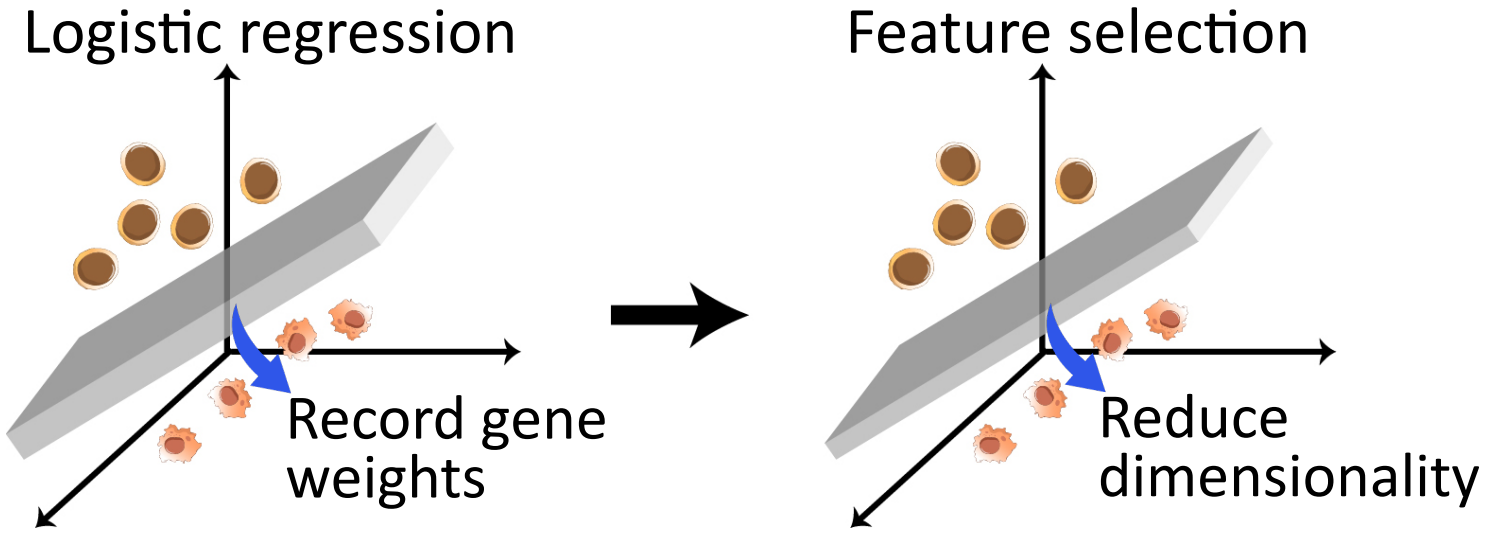
Cell Type “Encyclopedia”



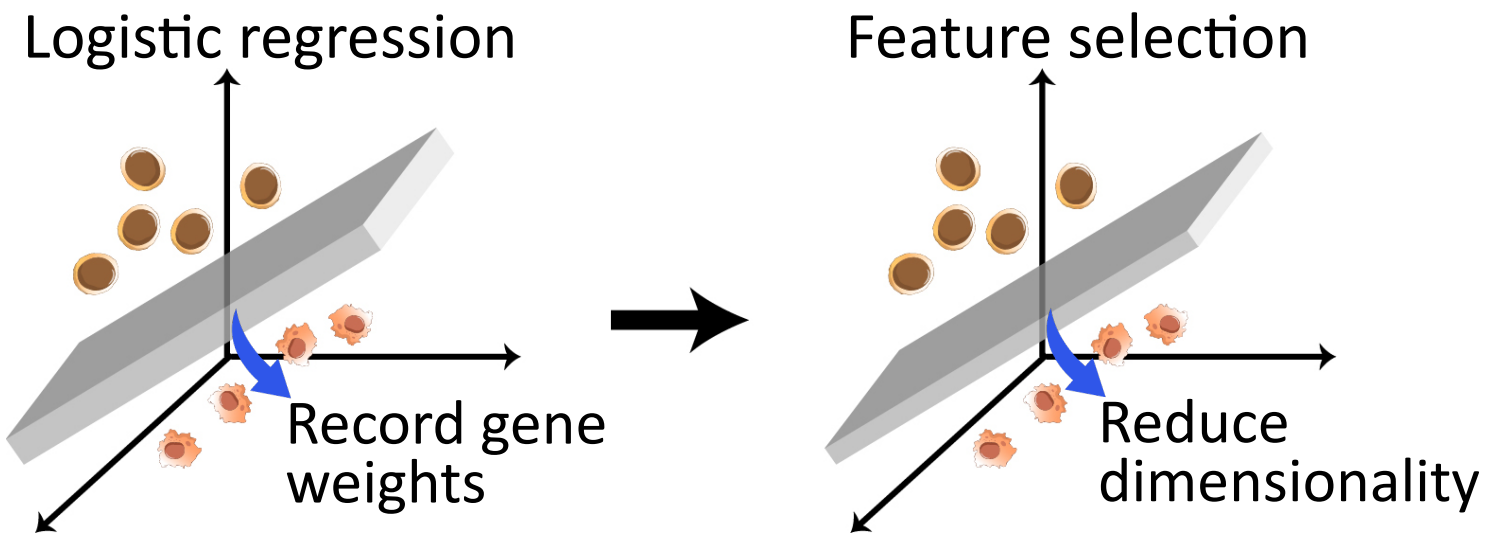


www.celltypist.org/encyclopedia/Immune

Lightweight & Shareability



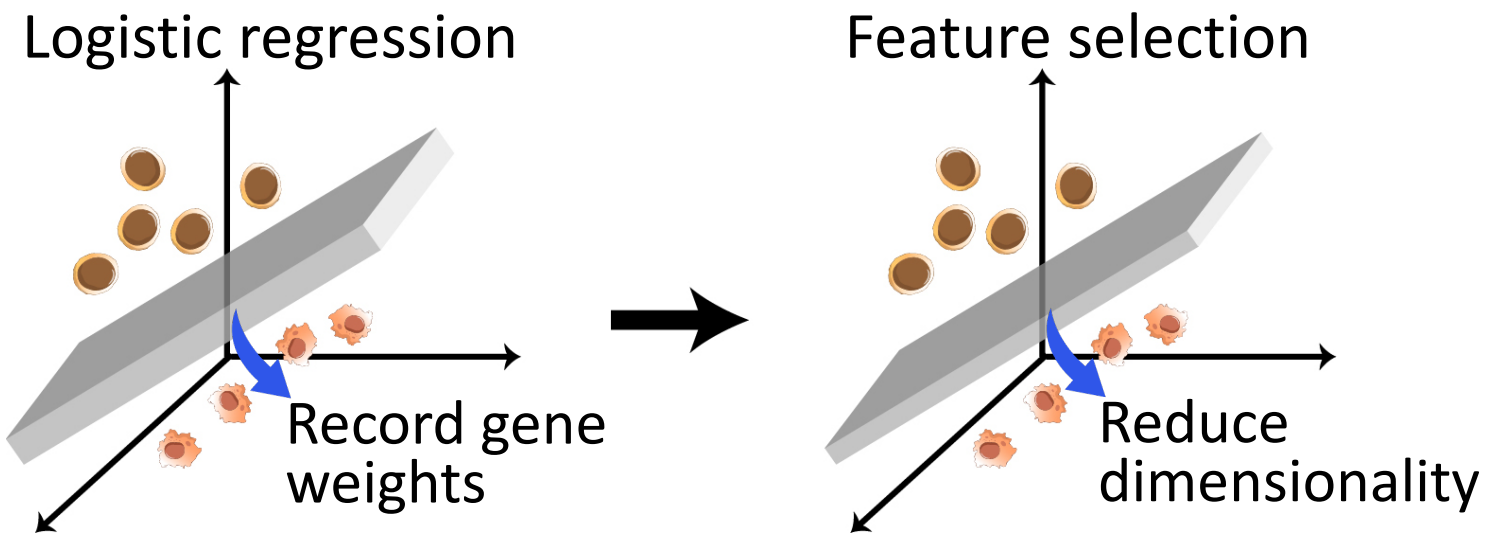
CellTypist Models



Model	Details	# types	Ver.	Size
Immune_All_Low	Immune sub-populations combined from 20 tissues of 18 studies	98	v2	2.7M
Healthy_Adult_Heart	Cell types from eight anatomical regions of the healthy adult human heart	75	v1	1.3M
Human_Lung_Atlas	Integrated human lung cell atlas combining multiple datasets of the healthy respiratory system	61	v2	1.4M
Healthy_COVID19_PBMC	Peripheral blood mononuclear cell types from healthy and COVID-19 individuals	51	v1	798K
Developing_Mouse_Brain	Cell types from the embryonic mouse brain between gastrulation and birth	174	v1	5.2M

For the full list of CellTypist models, check www.celltypist.org/models

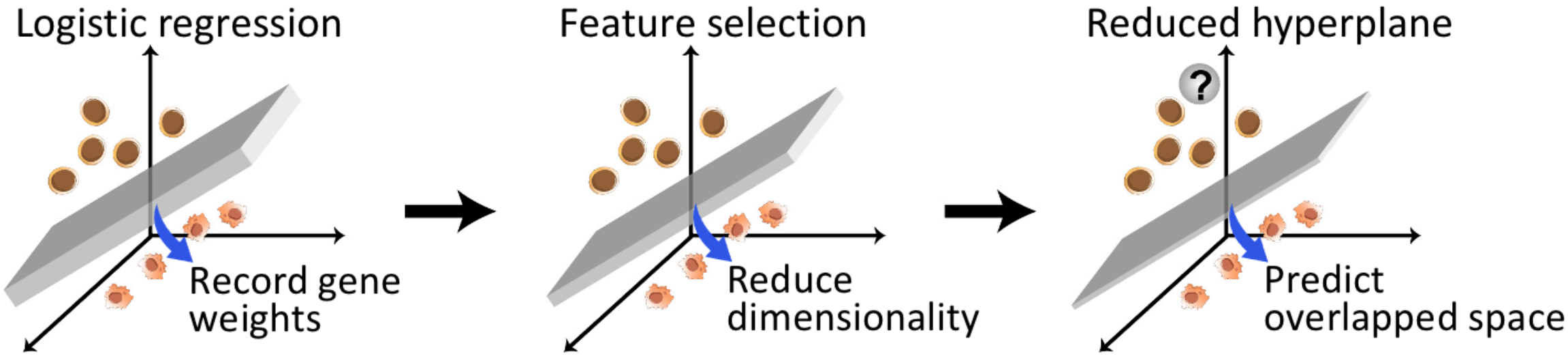
CellTypist Models



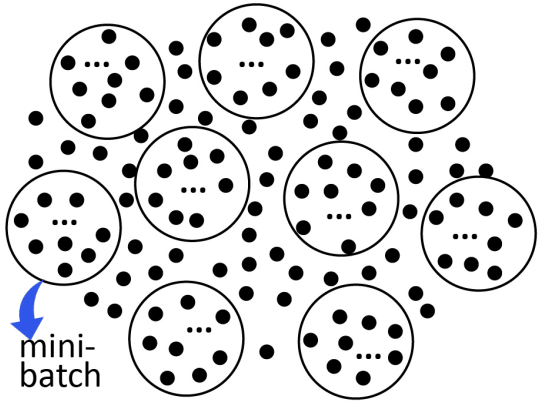
Model	Details	# types	Ver.	Size
Immune_All_Low	Immune sub-populations combined from 20 tissues of 18 studies	98	v2	2.7M
Healthy_Adult_Heart	Cell types from eight anatomical regions of the healthy adult human heart	75	v1	1.3M
Human_Lung_Atlas	Integrated human lung cell atlas combining multiple datasets of the healthy respiratory system	61	v2	1.4M
Healthy_COVID19_PBMC	Peripheral blood mononuclear cell types from healthy and COVID-19 individuals	51	v1	798K
Developing_Mouse_Brain	Cell types from the embryonic mouse brain between gastrulation and birth	174	v1	5.2M

For the full list of CellTypist models, check www.celltypist.org/models

Lightweight & Shareability

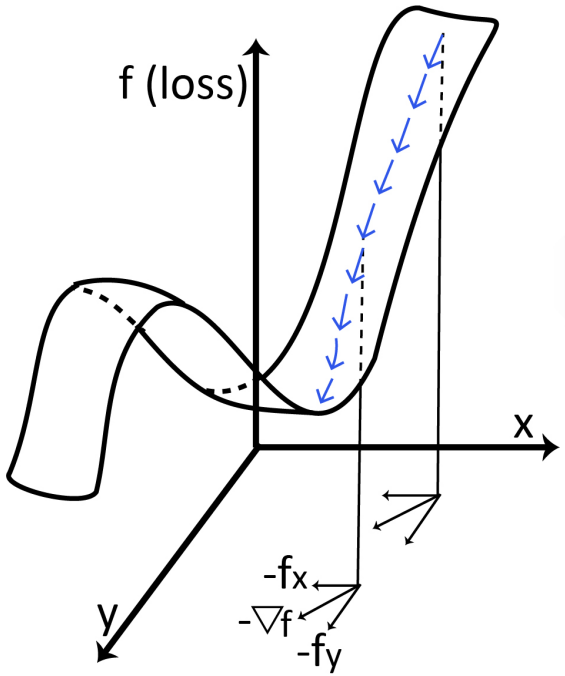


Efficiency



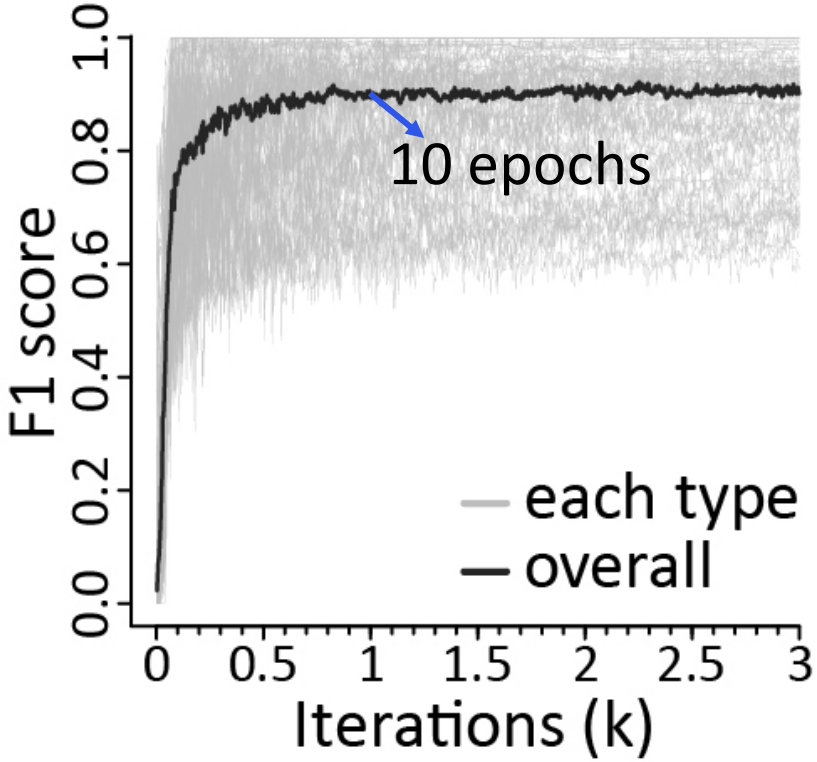
Mini-batch training

- 1,000 cells per batch
- 100 batches per epoch

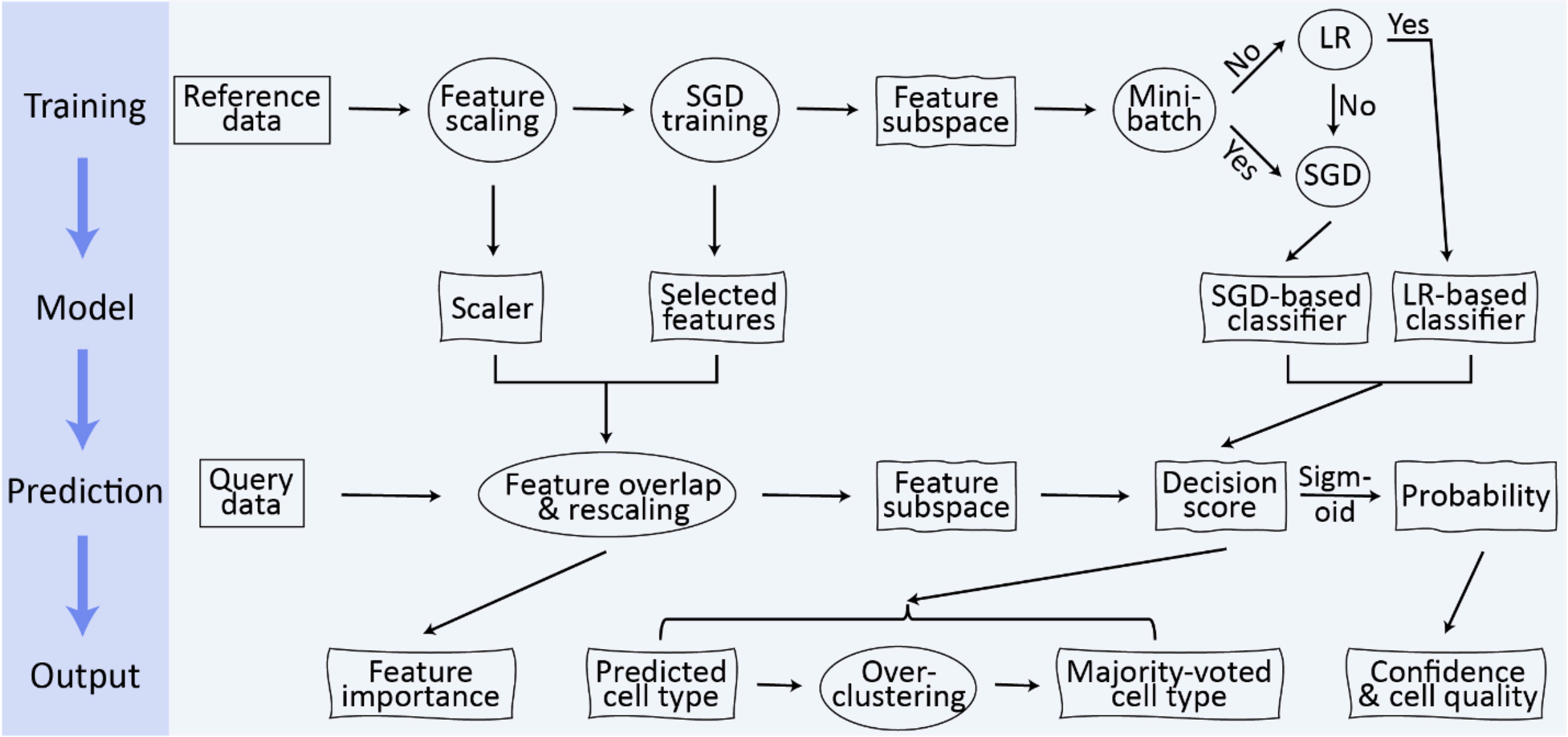


Stochastic gradient descent

- Iterative optimization
- Online training

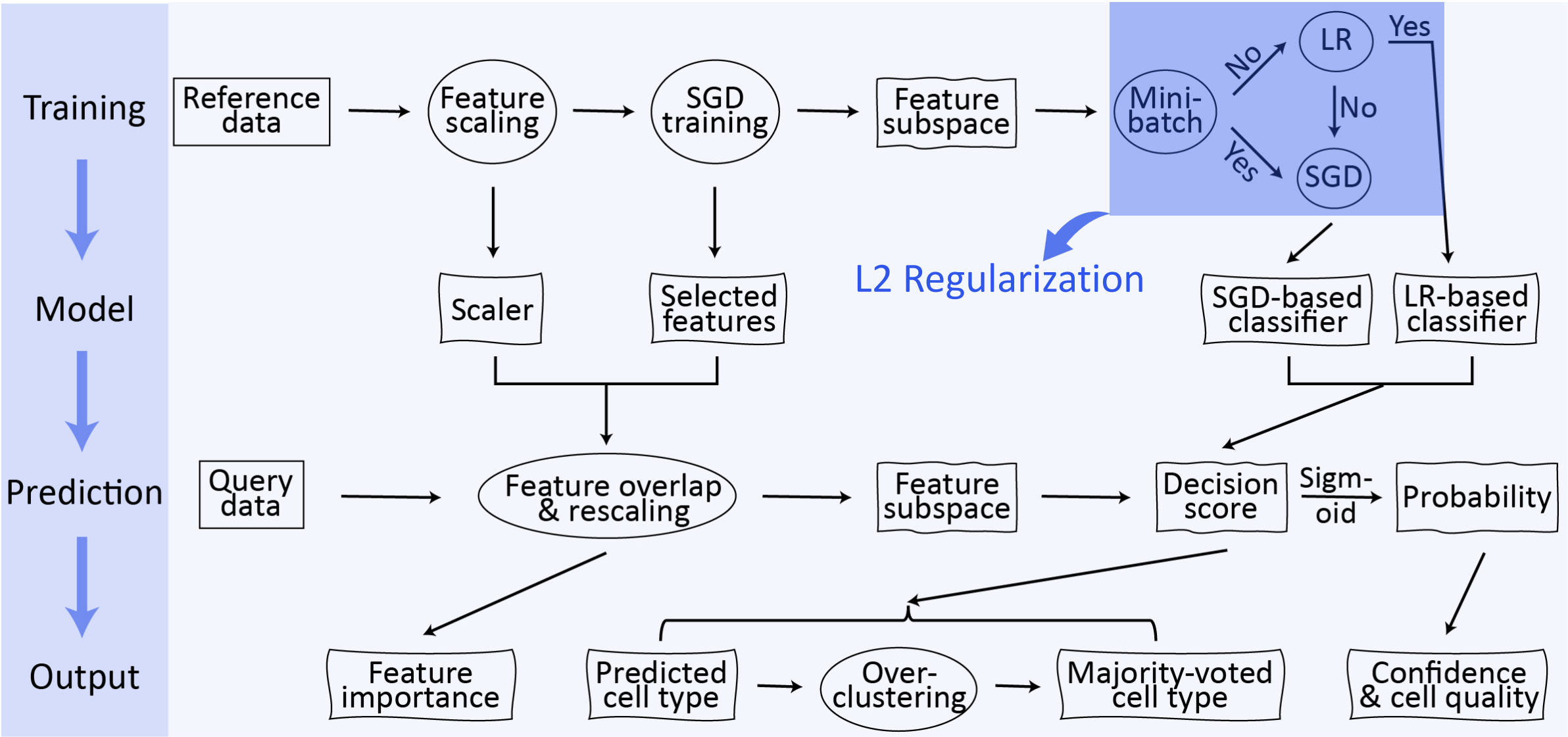


Accuracy



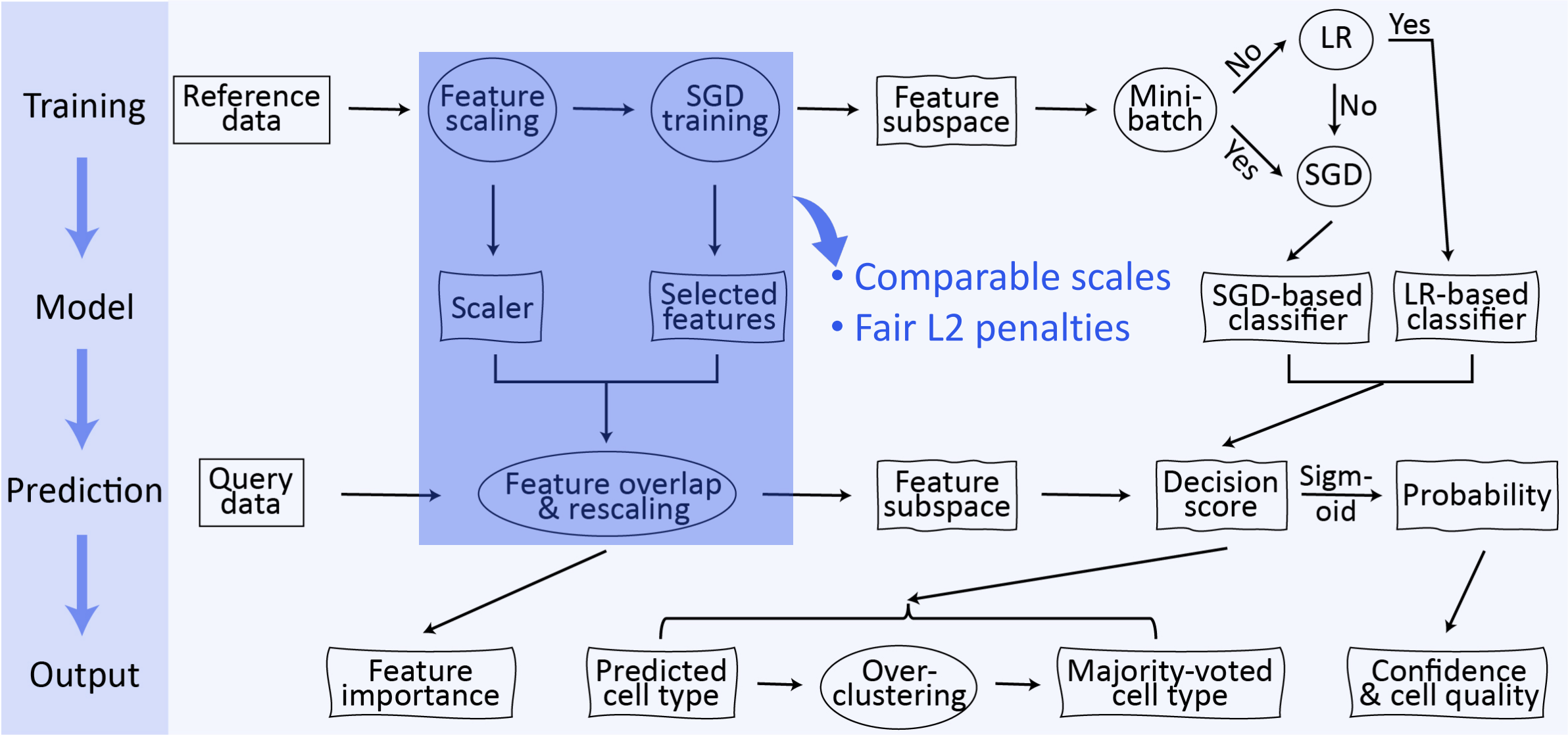
(SGD: stochastic gradient descent; LR: logistic regression)

Accuracy



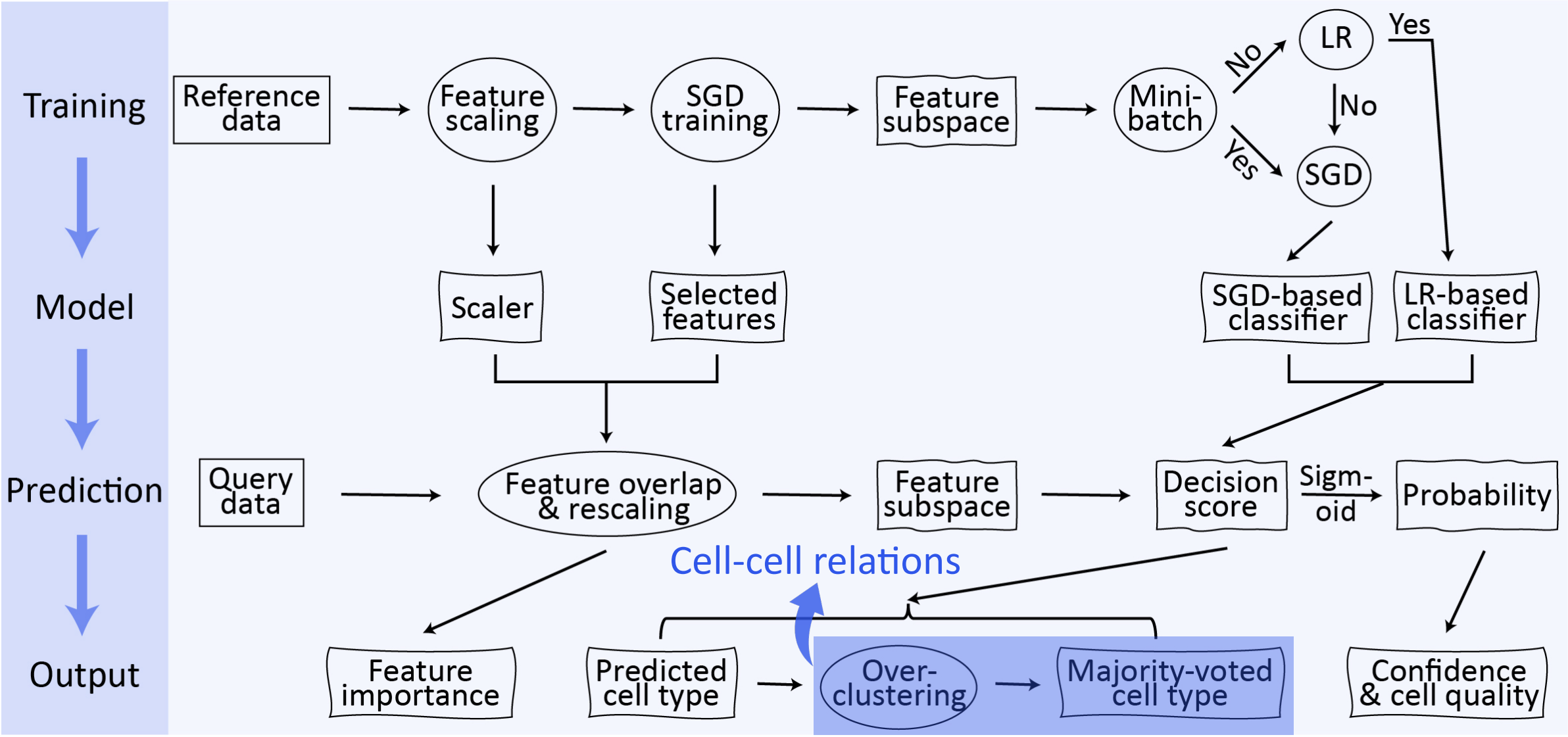
(SGD: stochastic gradient descent; LR: logistic regression)

Accuracy



(SGD: stochastic gradient descent; LR: logistic regression)

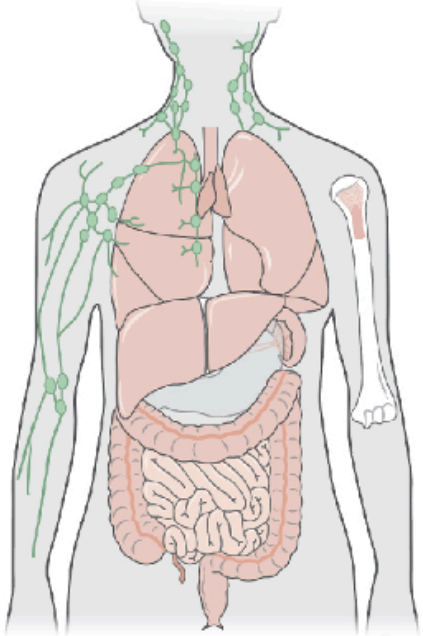
Accuracy



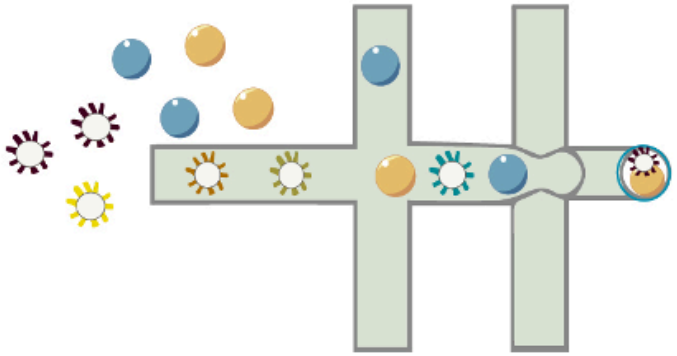
(SGD: stochastic gradient descent; LR: logistic regression)

Practical Application

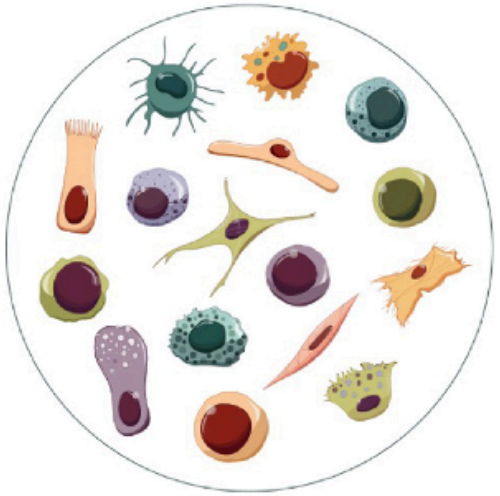
12 donors, 12 tissues



Dissociation & scRNA-seq



330k immune cells
(*PTPRC*+)



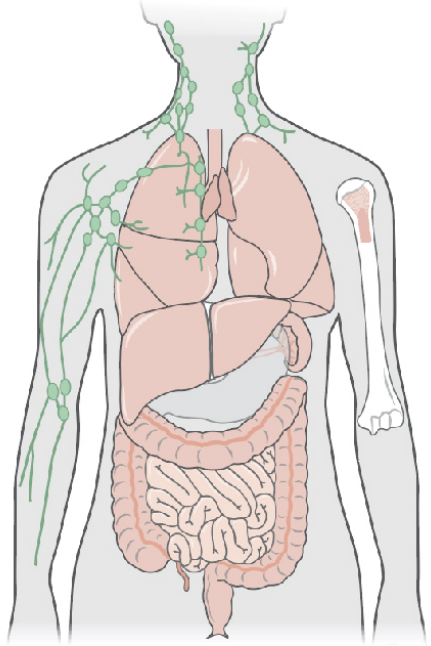
Cecilia Domínguez Conde

Lorna Jarvis

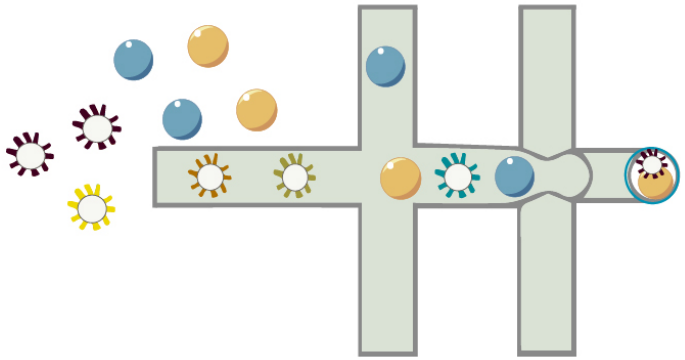
Daniel Rainbow

Practical Application

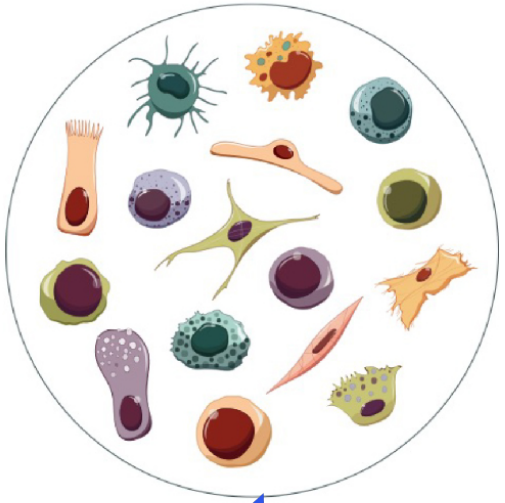
12 donors, 12 tissues



Dissociation & scRNA-seq

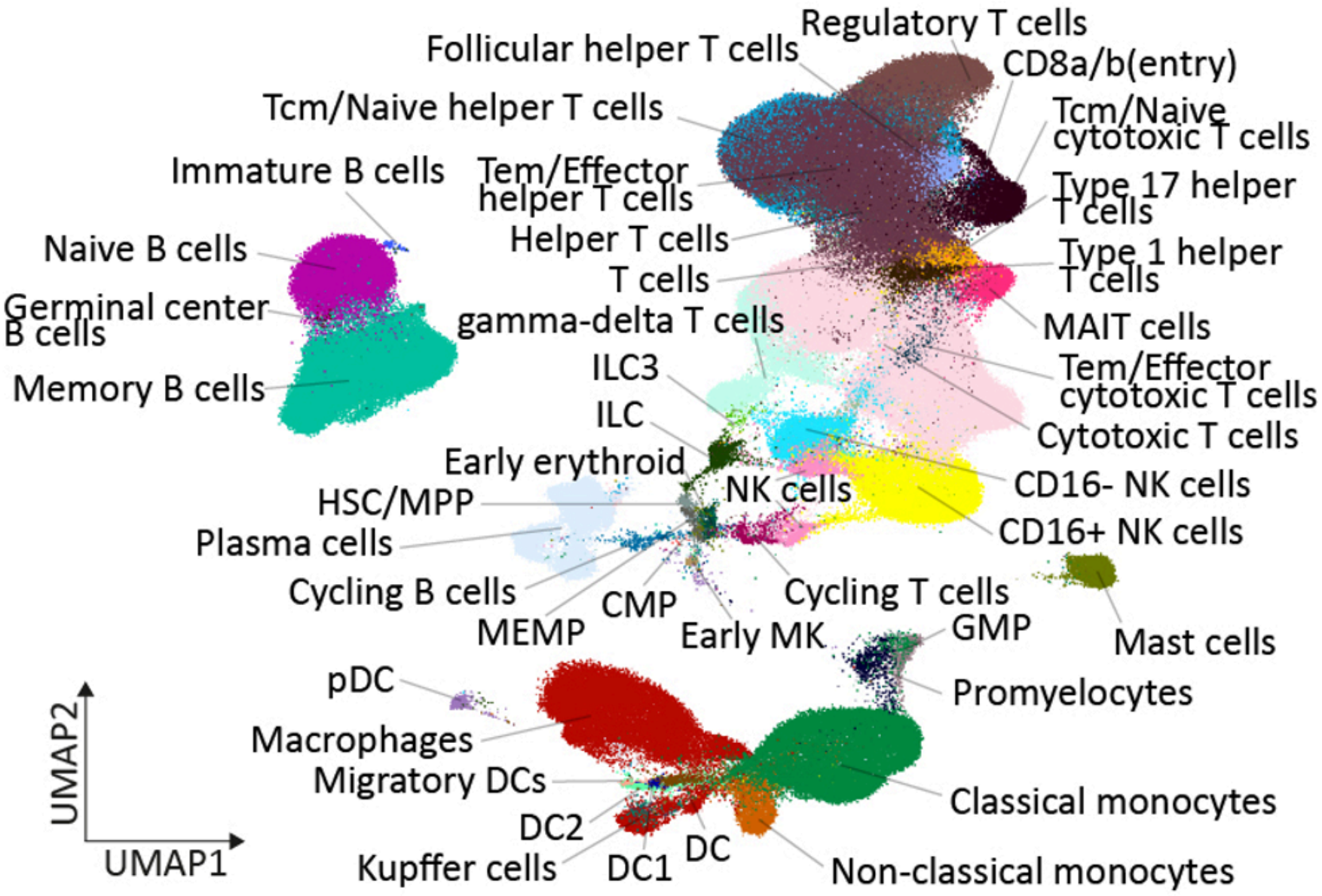


330k immune cells
(*PTPRC*+)



CellTypist

Practical Application



Summary of CellTypist

- Tool for rapid, precise, and automated cell annotation
- Framework for data-driven cell annotation system
- Database of immune cells, now expanded to
  - ~50 tissue models (lung, gut, etc.)
  - both embryonic and adult stages
  - >800 cell types in total

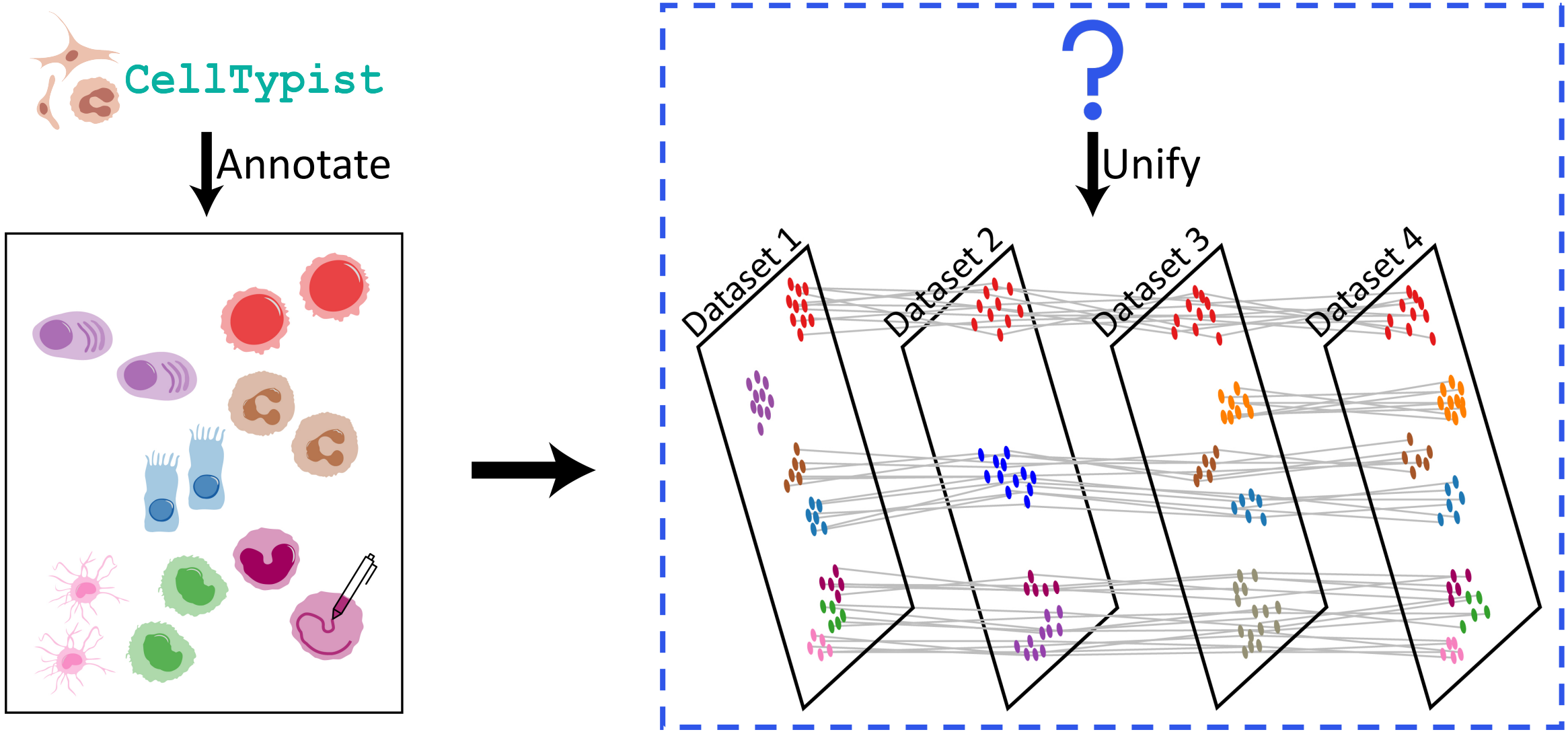
CellTypist

Website: celltypist.org

GitHub: Teichlab/celltypist

Tutorial: celltypist.readthedocs.io

How To Unify Cell Types?

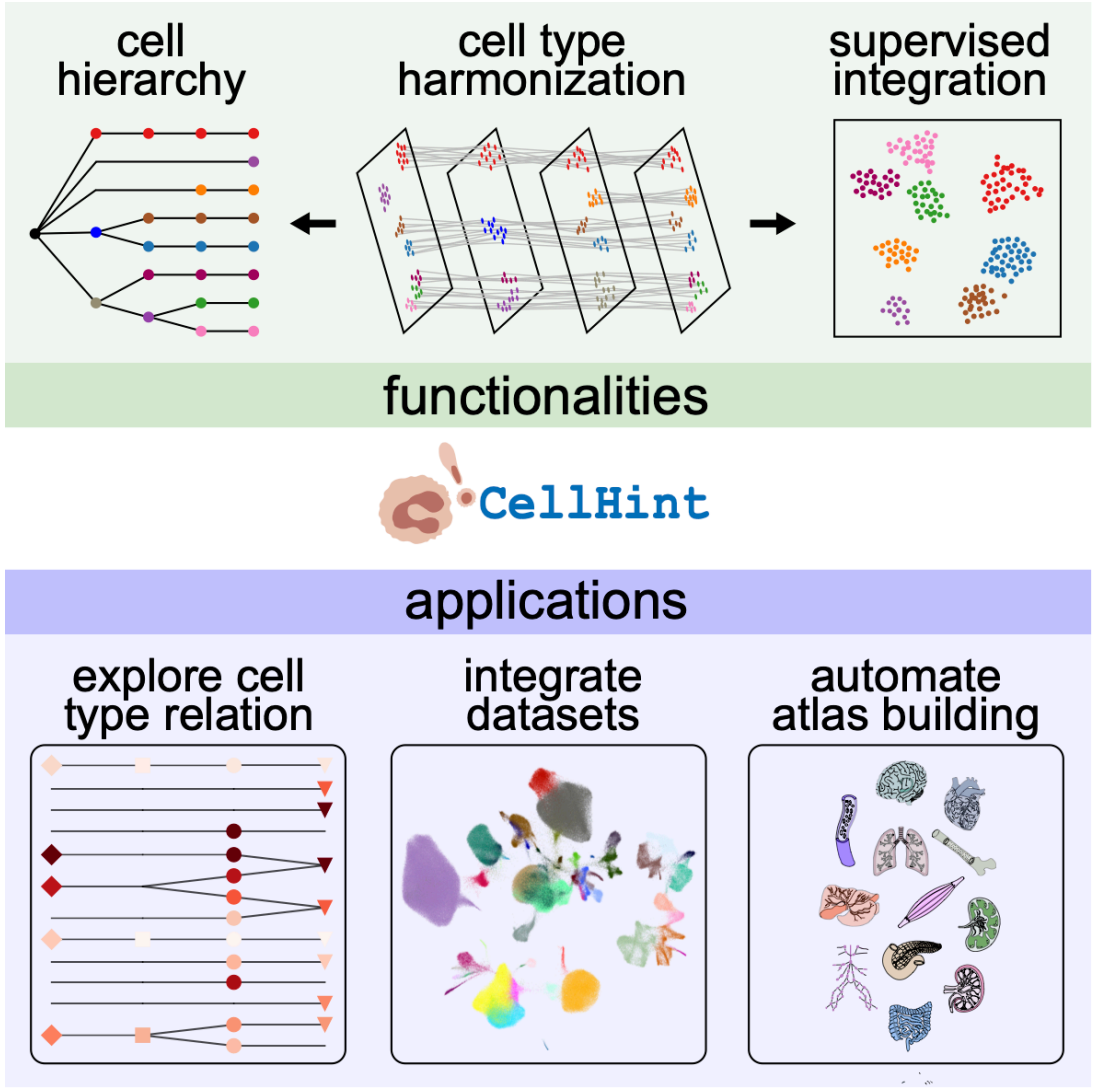


How To Unify Cell Types?

CellHint

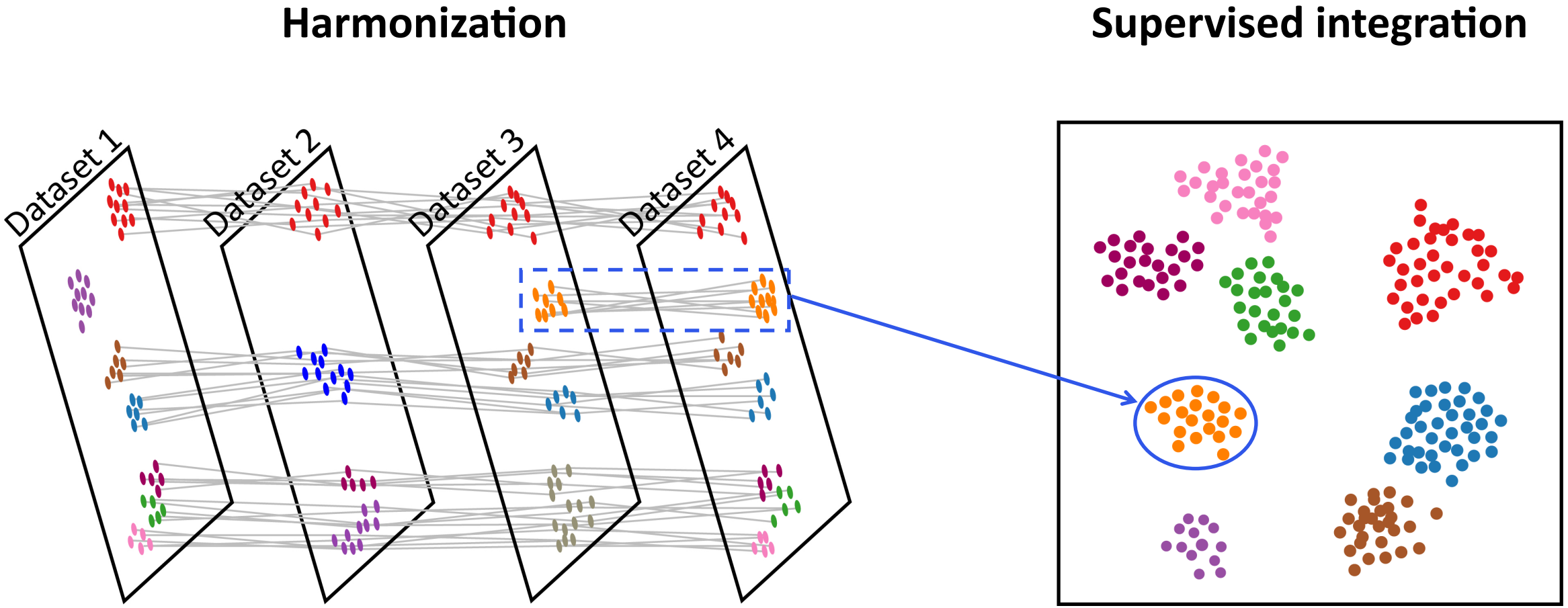
GitHub: Teichlab/cellhint

Tutorial: cellhint.readthedocs.io



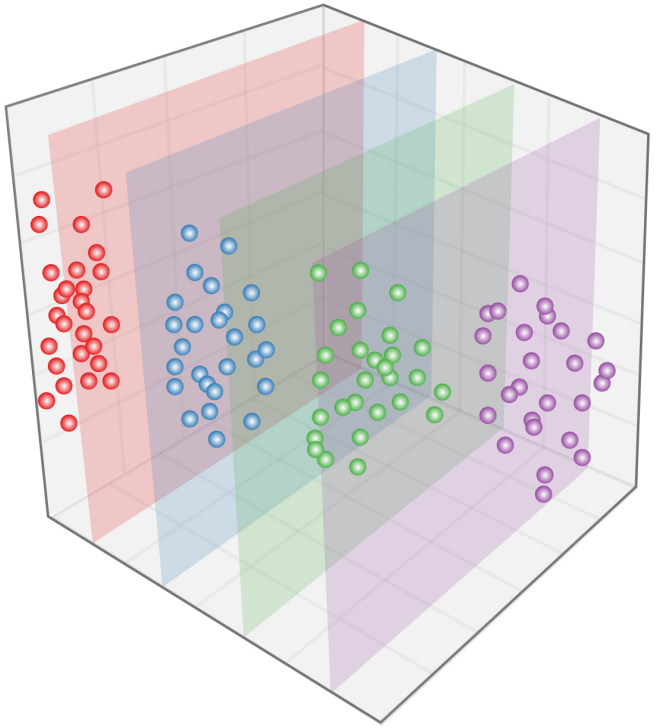
Xu et al., Cell, 2023.

What Is CellHint?

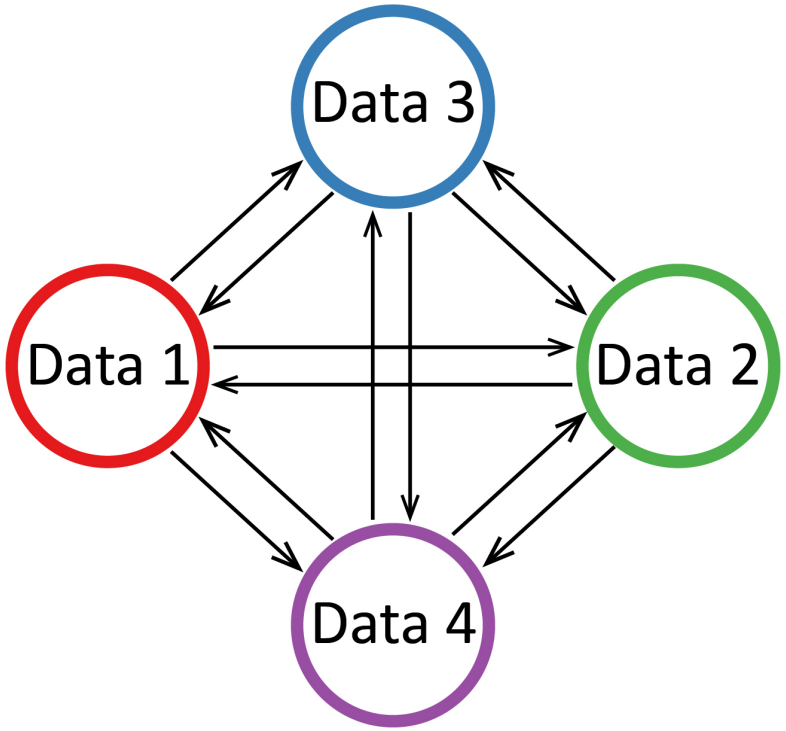


Challenges

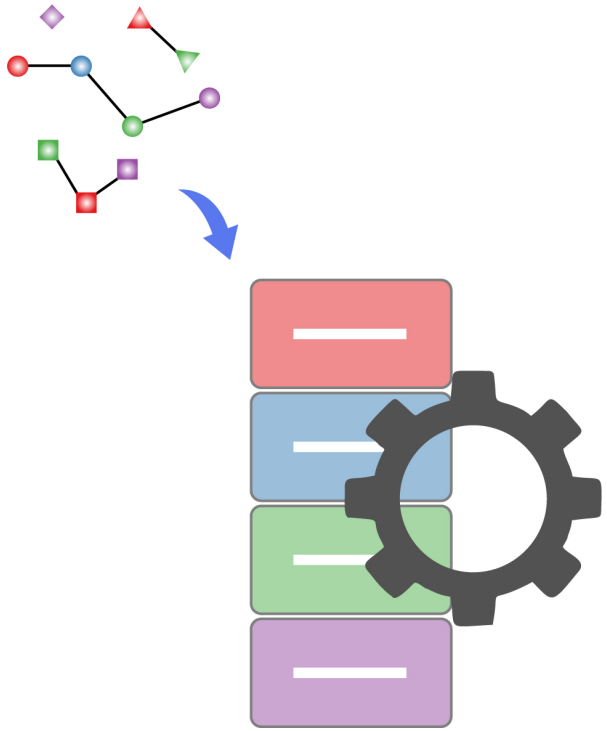
Measuring cell-cell distances



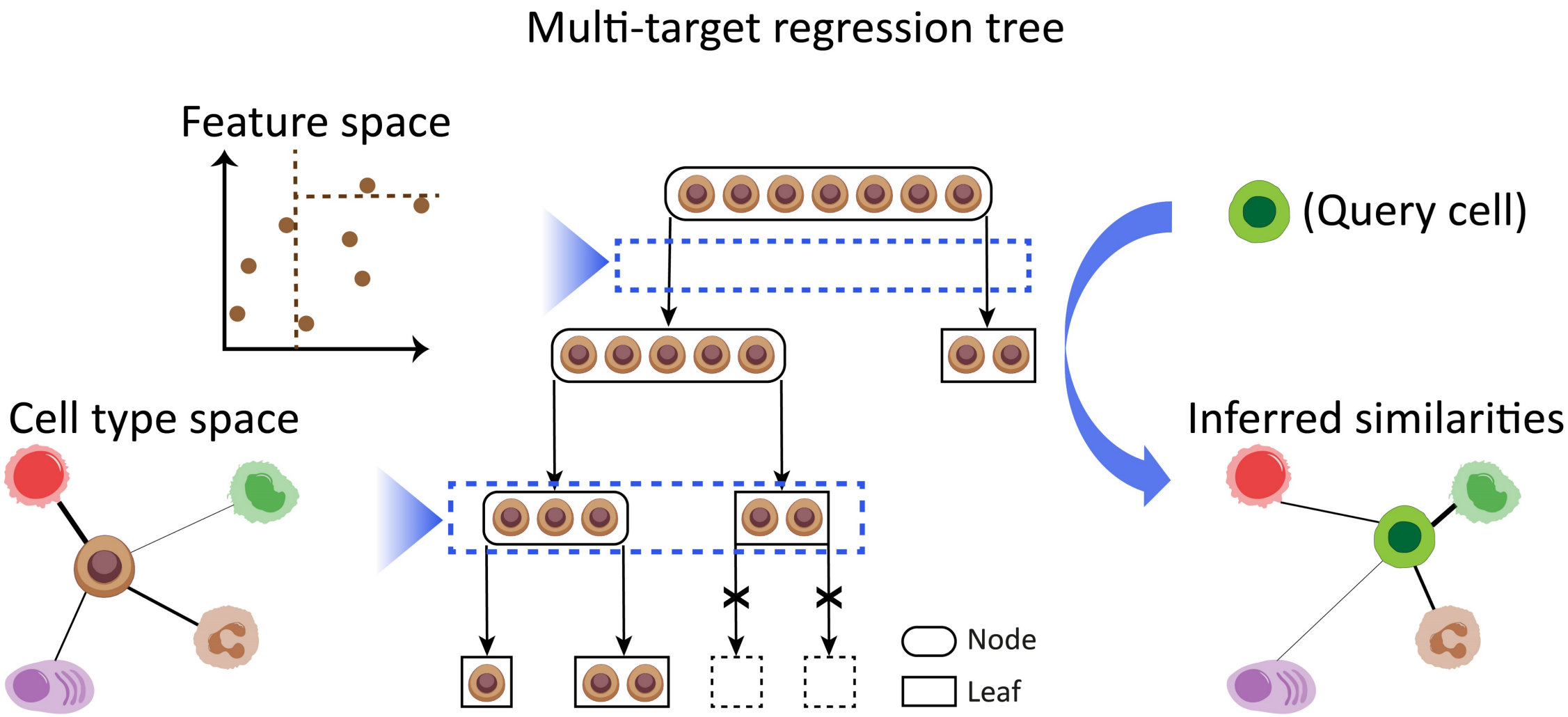
Aligning multiple datasets



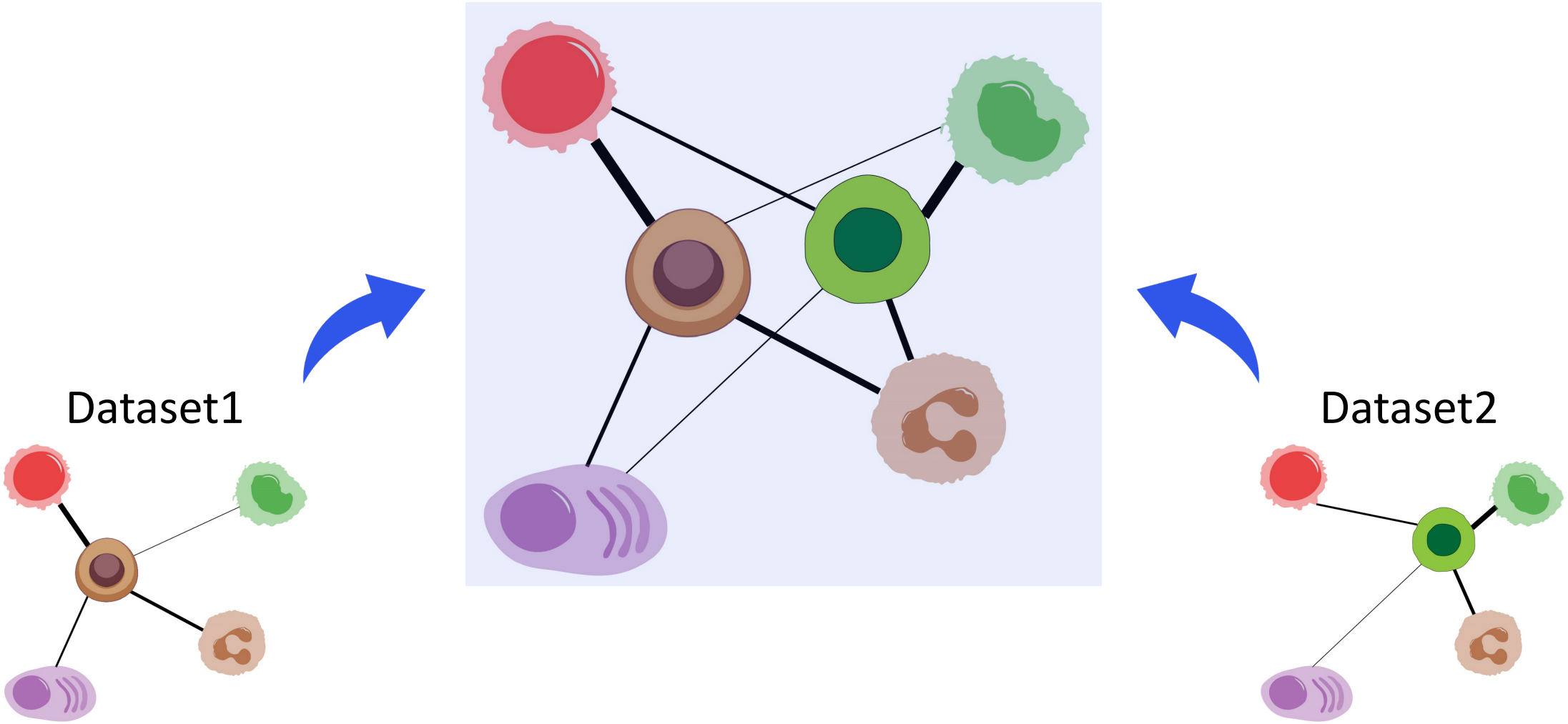
Incorporating cell relations into data integration



Predictive Clustering Tree-Based Meta-Analysis

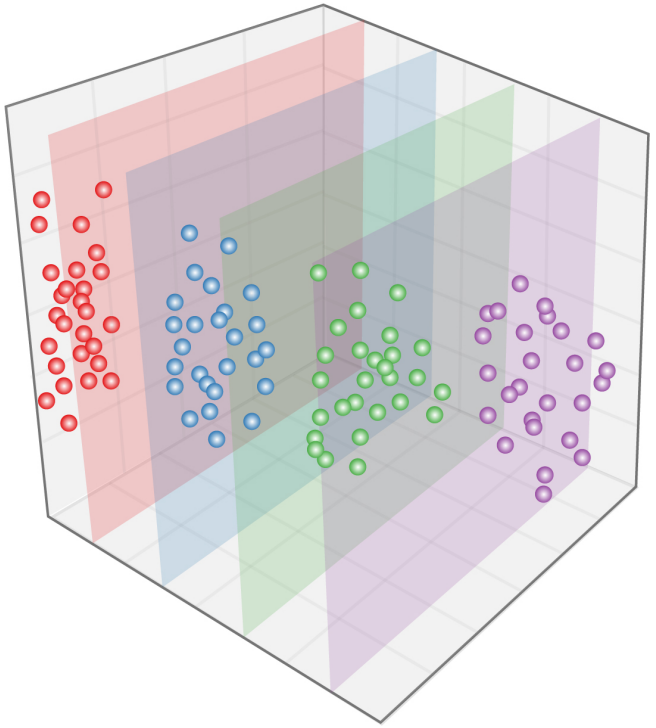


Robust Cell-Cell Distance Measure

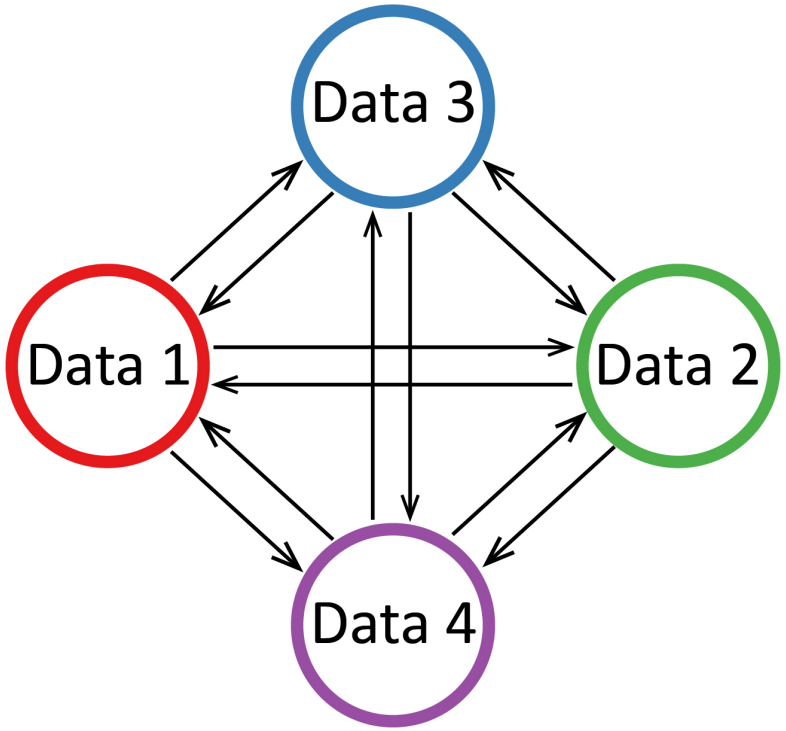


Challenges

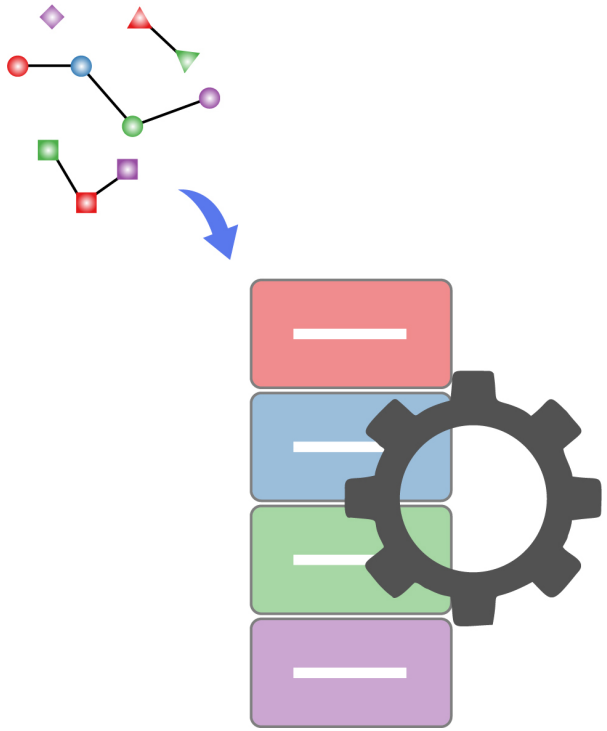
Measuring cell-cell distances



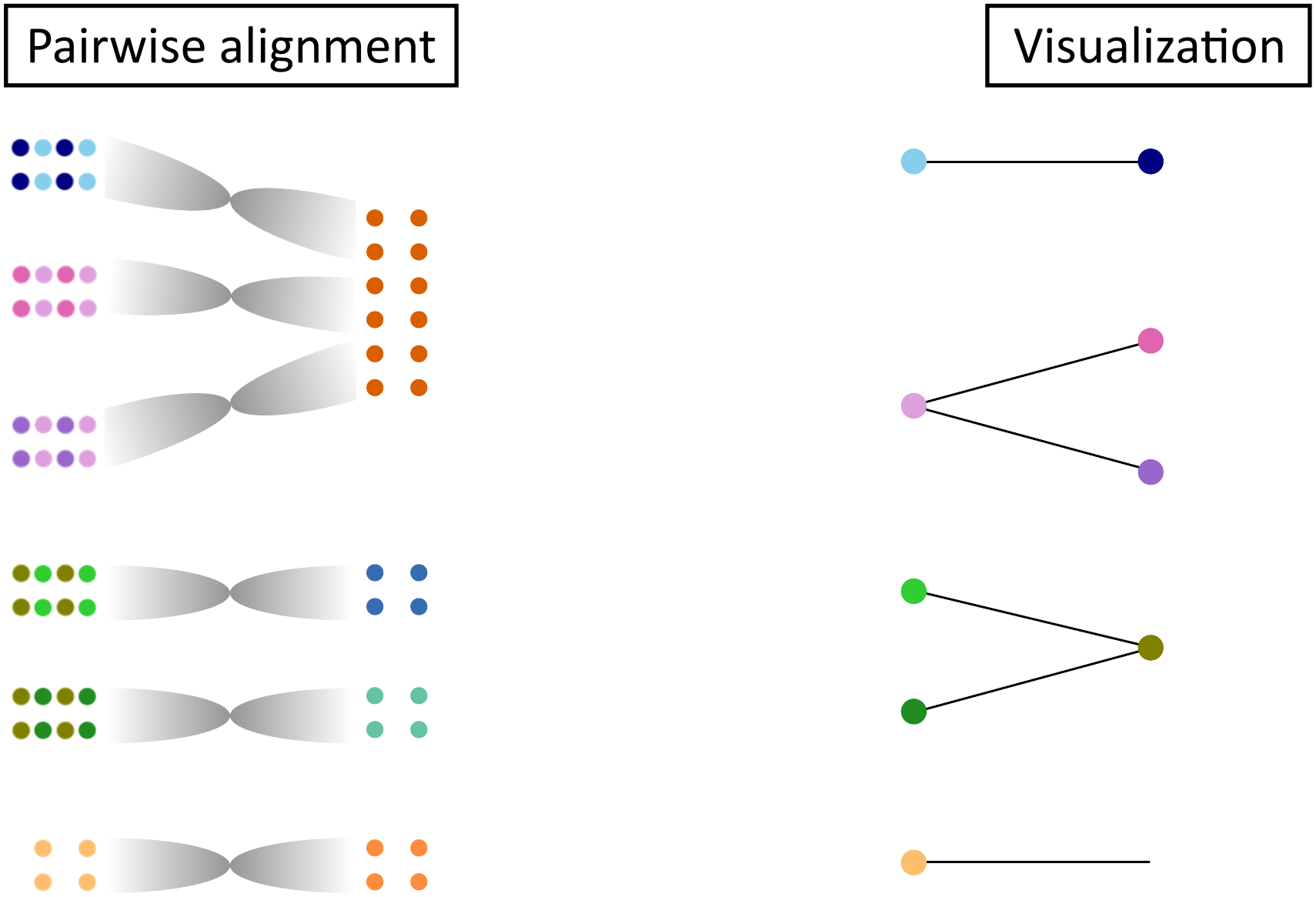
Aligning multiple datasets



Incorporating cell relations into data integration

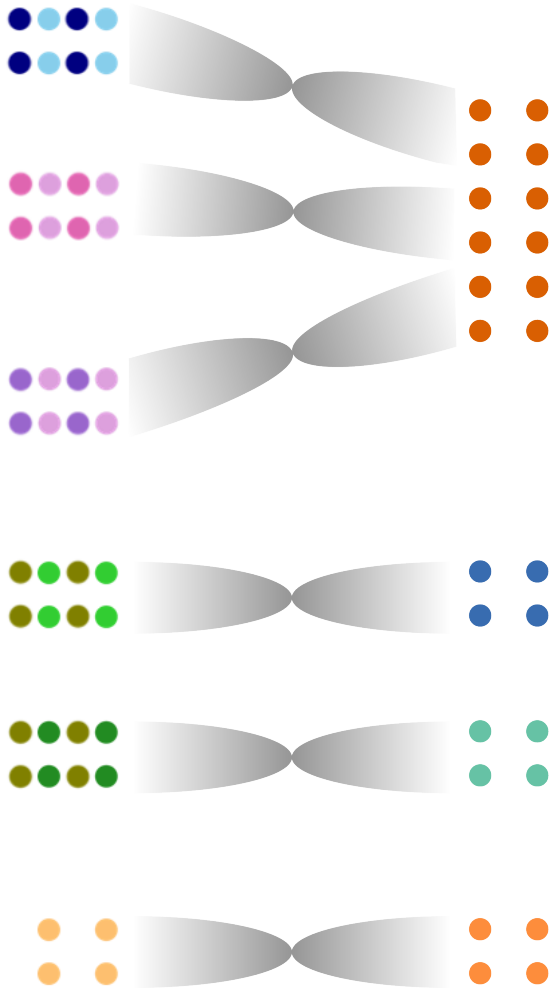


Iterative Cell Type Harmonization

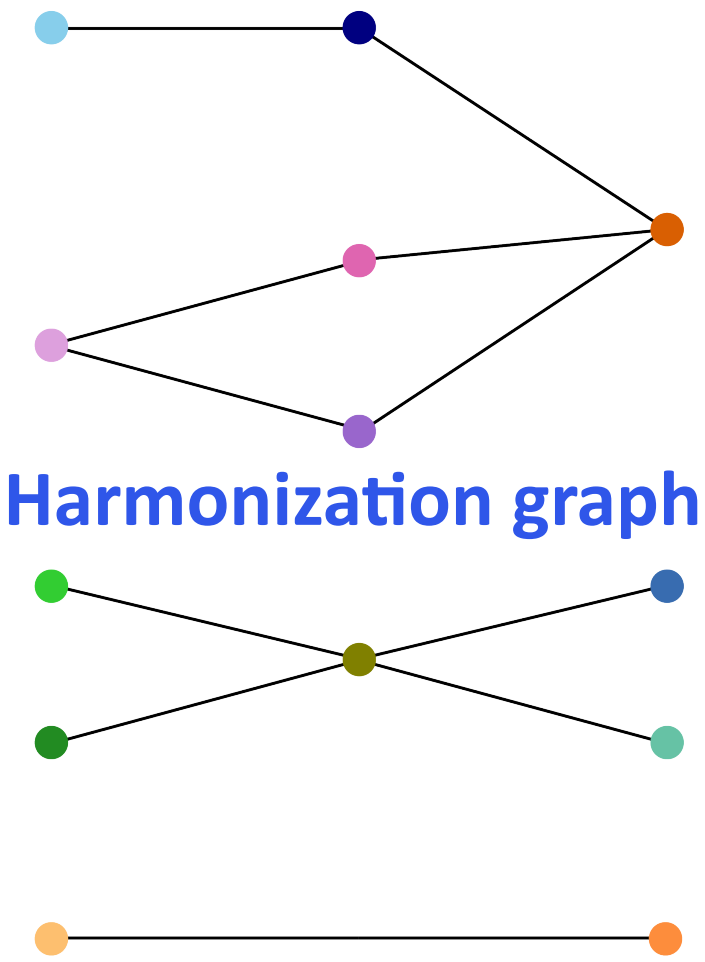


Iterative Cell Type Harmonization

Pairwise alignment

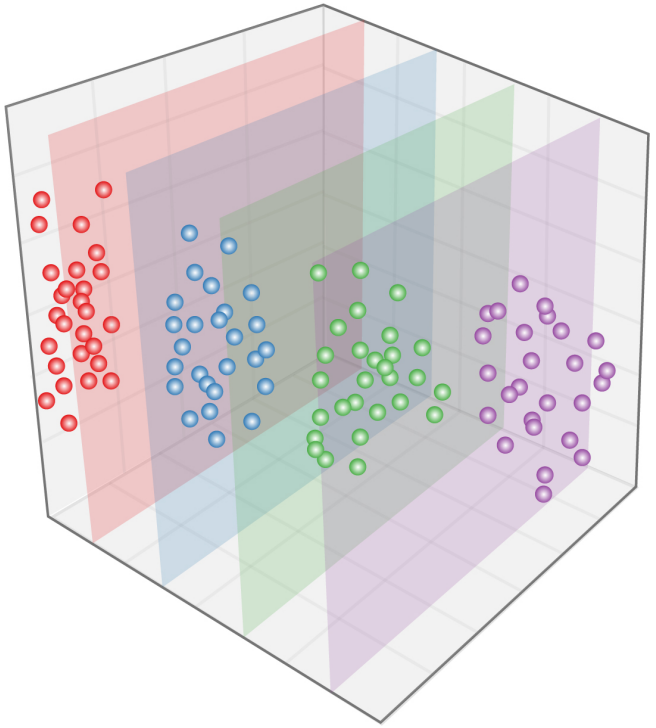


Visualization

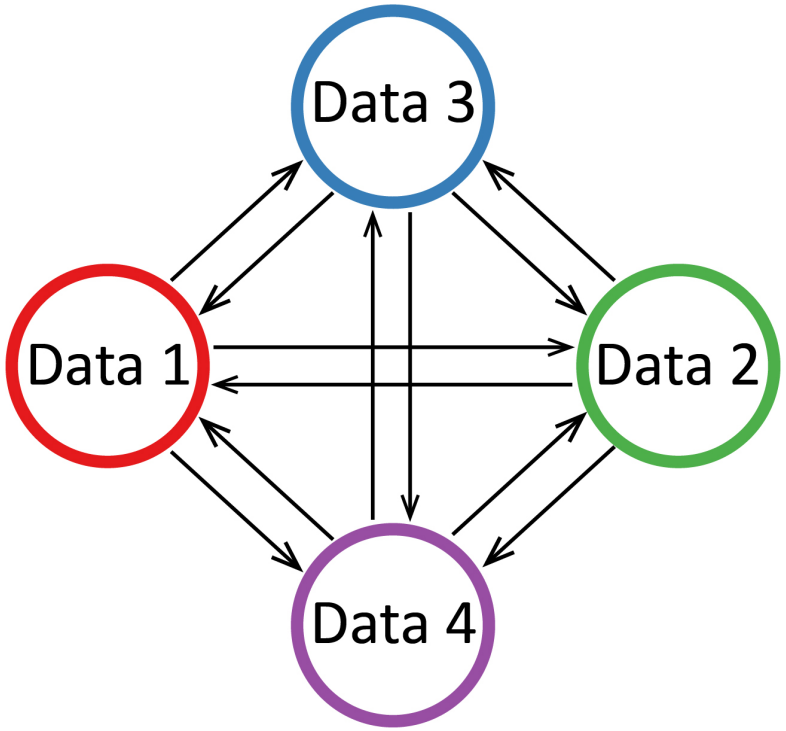


Challenges

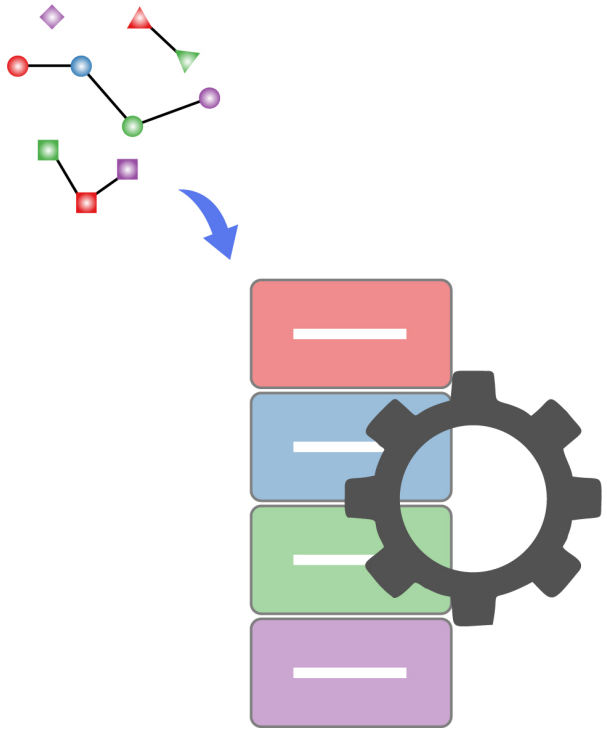
Measuring cell-cell distances



Aligning multiple datasets



Incorporating cell relations into data integration



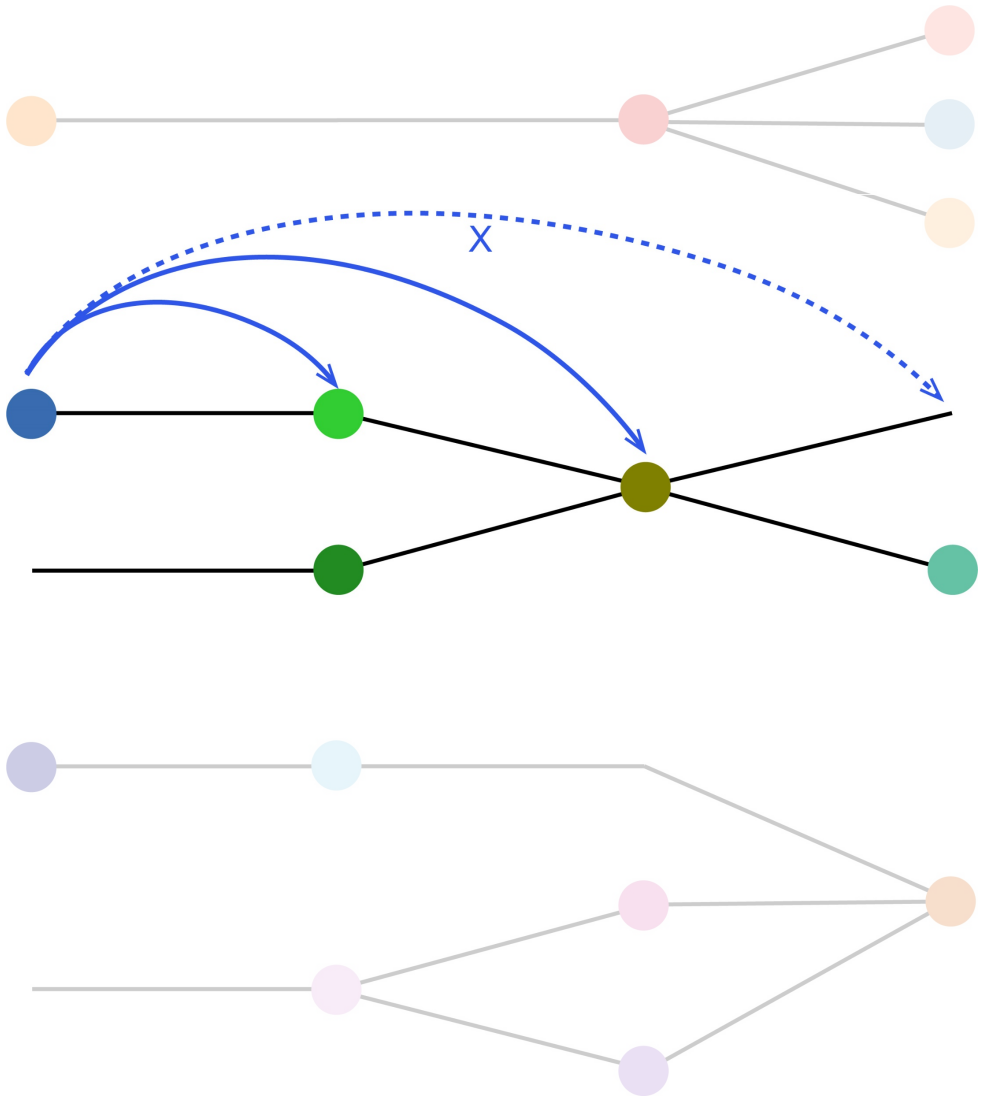
Restricted Neighborhood Search

For each cell

1. **Find its original annotation**

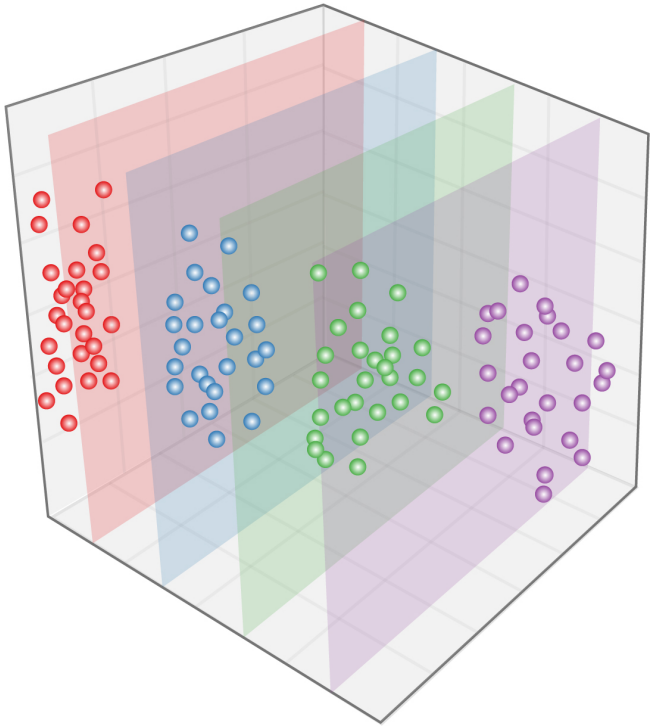
2. **Locate the branch it belongs to**

3. **Gather neighbours across datasets**

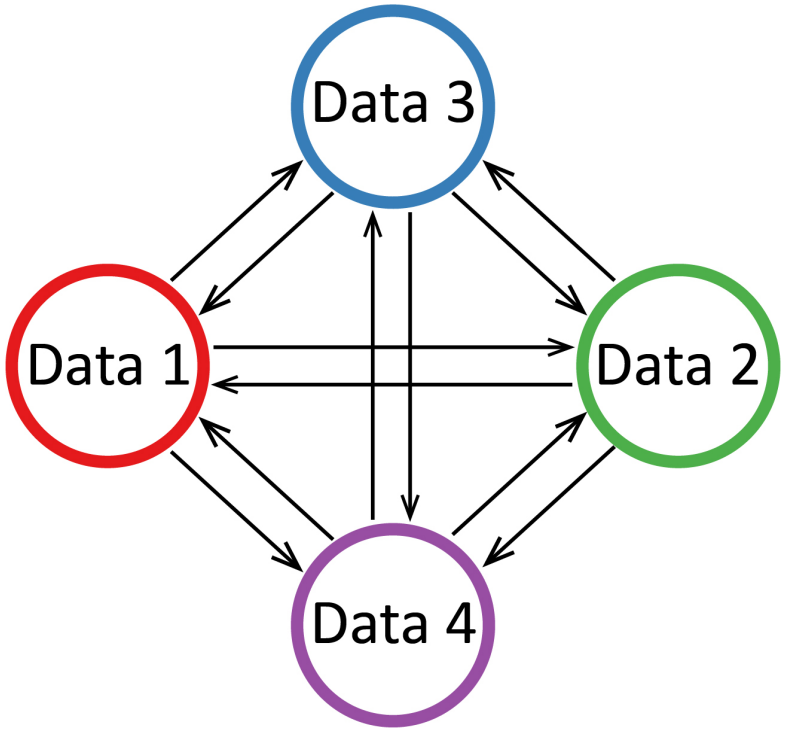


What Promise Arises?

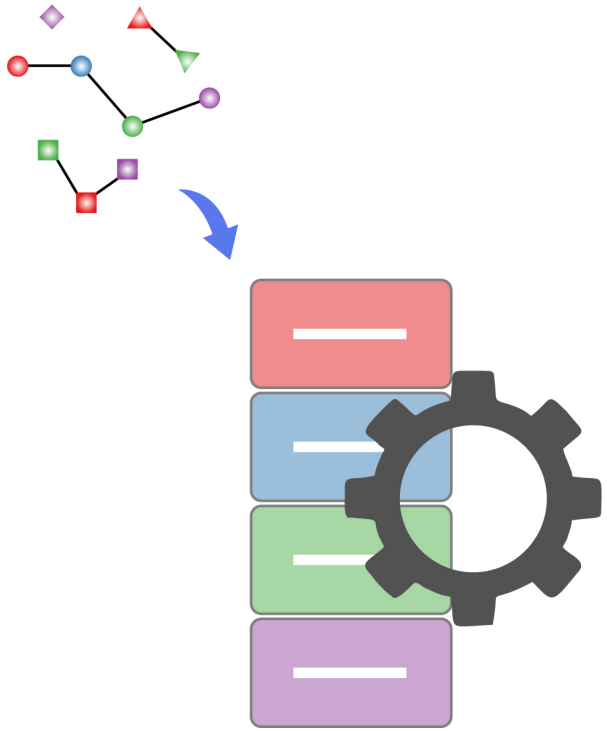
Measuring cell-cell distances



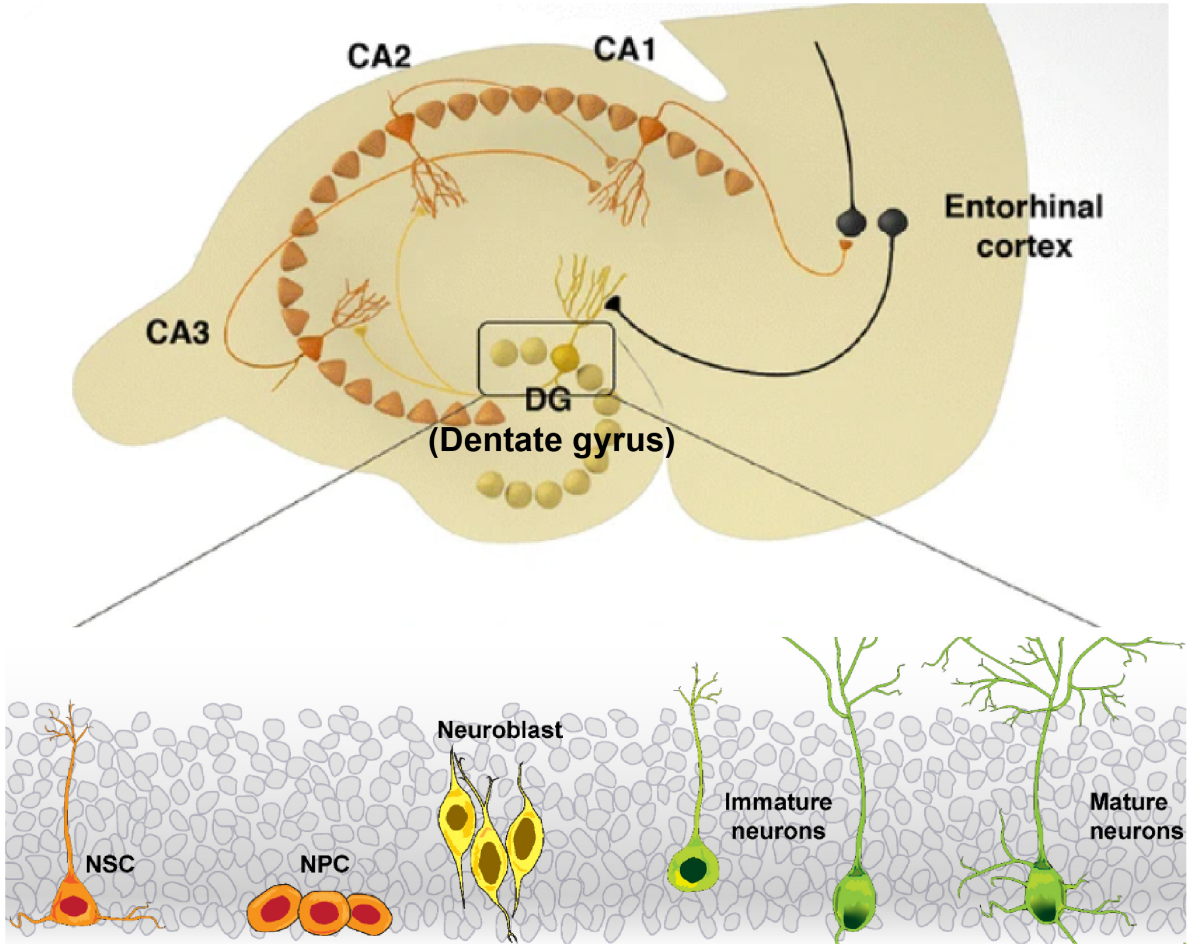
Aligning multiple datasets



Incorporating cell relations into data integration



Example: Adult Hippocampal Neurogenesis



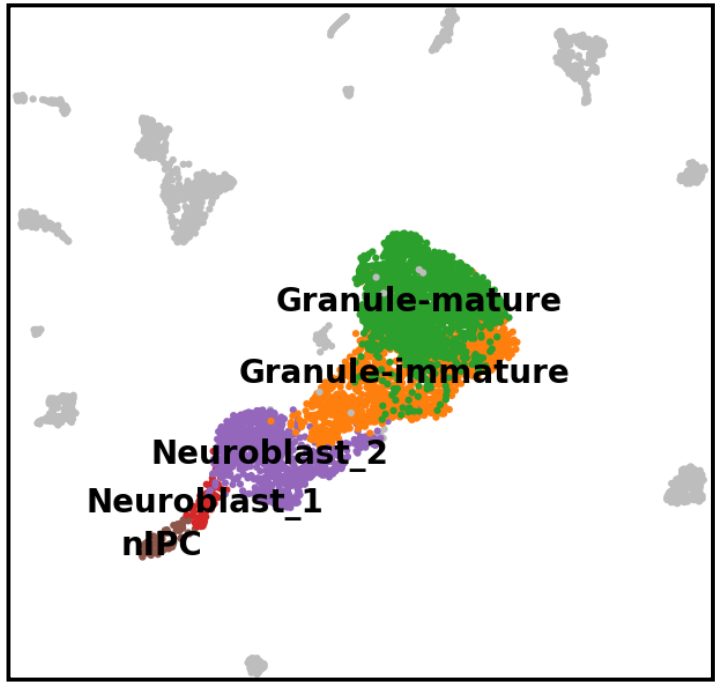
Adapted from Toda et al., Cell and Tissue Res., 2018 & Netzahualcoyotzi et al., Int. J. Mol. Sci., 2021

Adult-born neurons in the mammalian hippocampus, improving

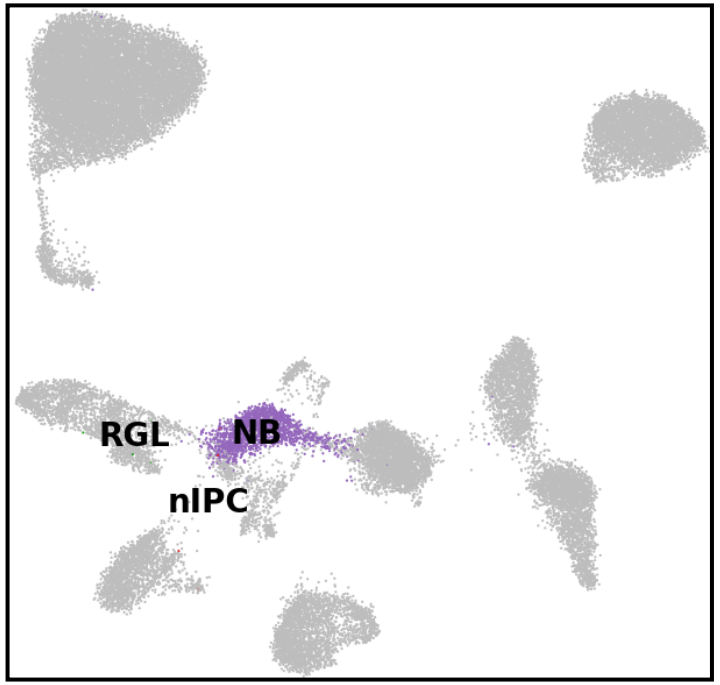
- neural circuit plasticity
- stress response
- pattern separation

Adult Hippocampal Neurogenesis
(from the perspective of single-cell transcriptomics)

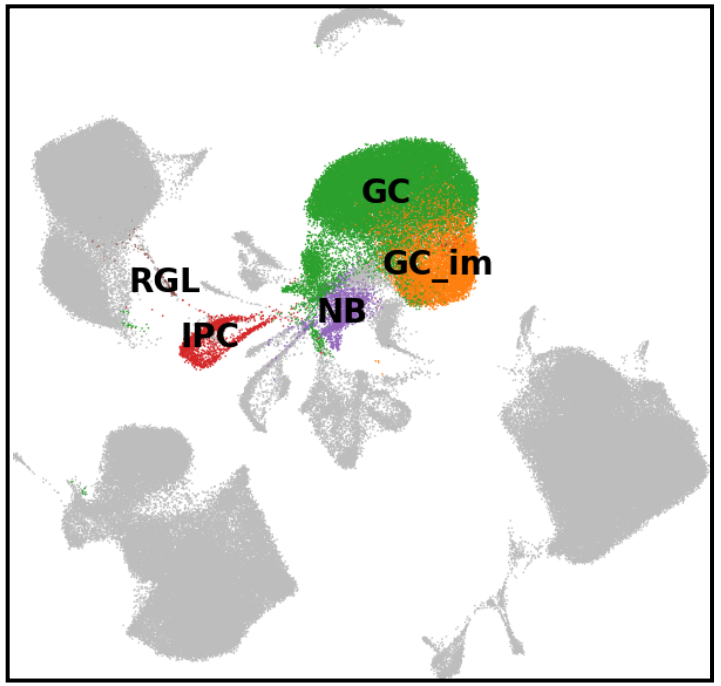
Mouse

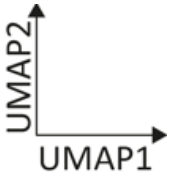


Pig



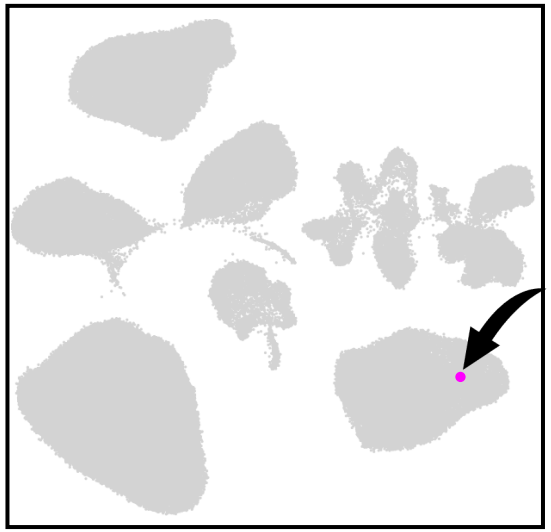
Monkey





Reanalyzed from Hochgerner et al., Nat. Neurosci., 2018 (left), Franjic et al., Neuron, 2022 (middle), and Hao et al., Nat. Neurosci., 2022 (right)

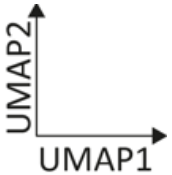
Whether Exists in Humans?



Only one neuroblast

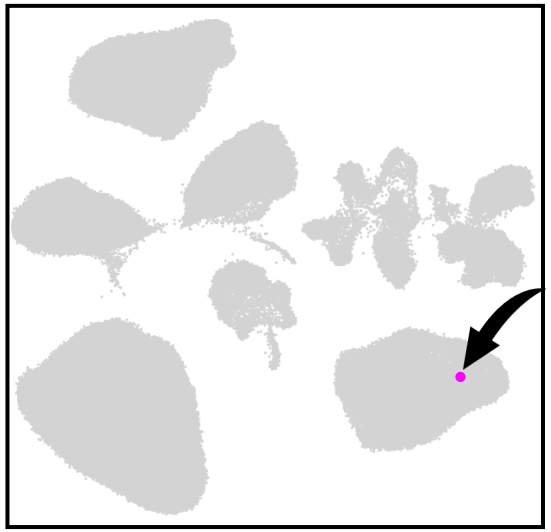
(Franjic et al., Neuron, 2022)

No neurogenic trajectory found in humans yet



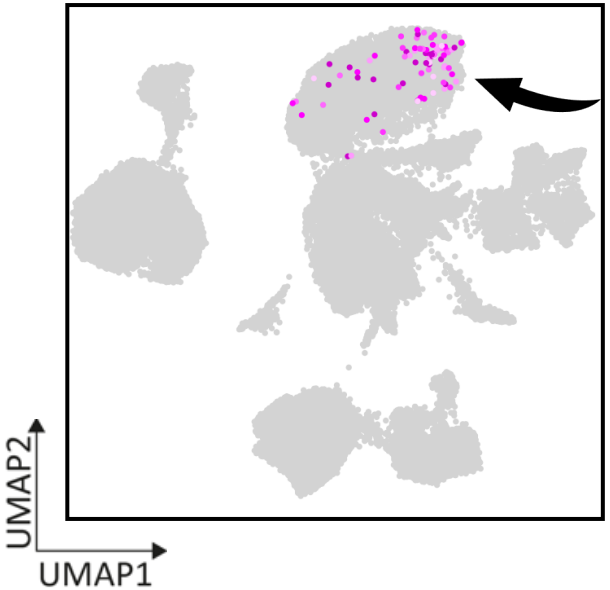
Whether Exists in Humans?

No neurogenic trajectory found in humans yet



Only one neuroblast

(Franjic et al., Neuron, 2022)



Scattered immature neurons (scoring-based)

(Zhou et al., Nature, 2022)

Detect and Enrich Signals With CellHint

Dataset	Platform	Donors	# Cells post QC
Siletti et al. 2022	10x 3' v3	3	346,756
Franjic et al. 2022	10x 3' v3	6	179,674
Ayhan et al. 2021	10x 3' v2 & v3	5	111,438
Zhou et al. 2022	SPLiT-seq	10	33,553
Wang et al. 2022	10x 3' v3	4	21,512
Tran et al. 2021	10x 3' v3	3	10,050

Collect six available datasets

↓

Quality control

↓

Cell type harmonization

↓

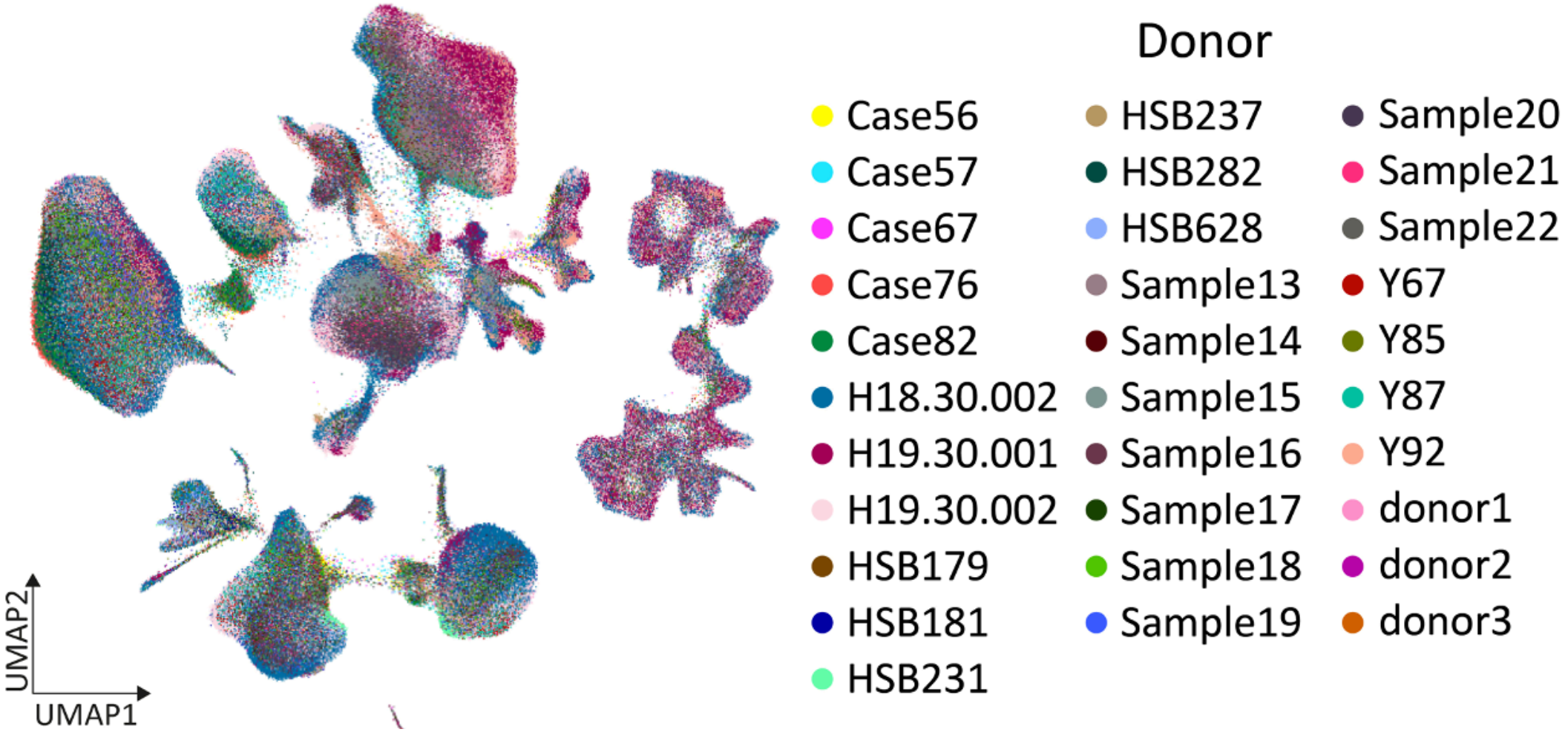
Supervised integration

↓

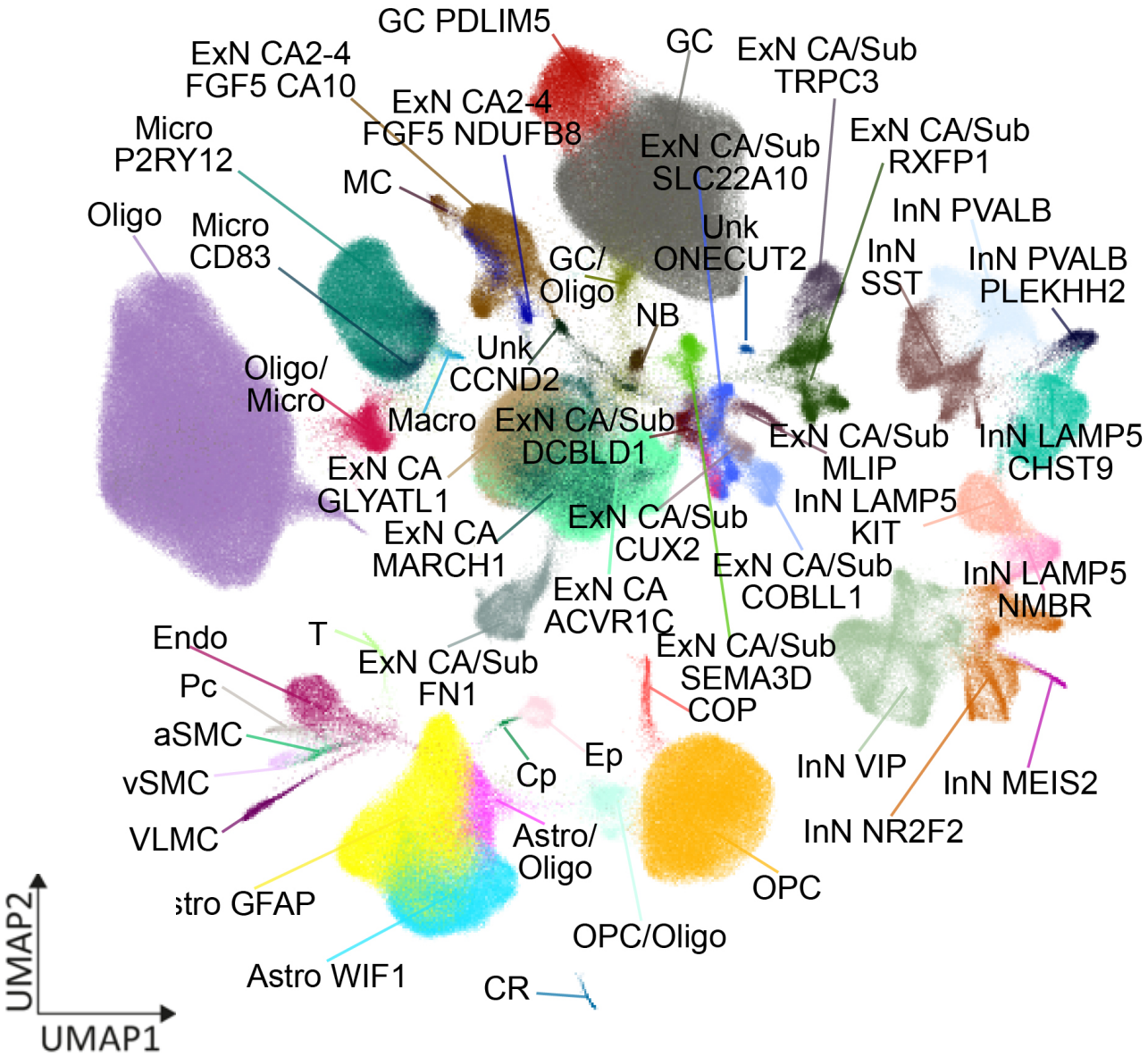
Manual inspection & revision

CellHint

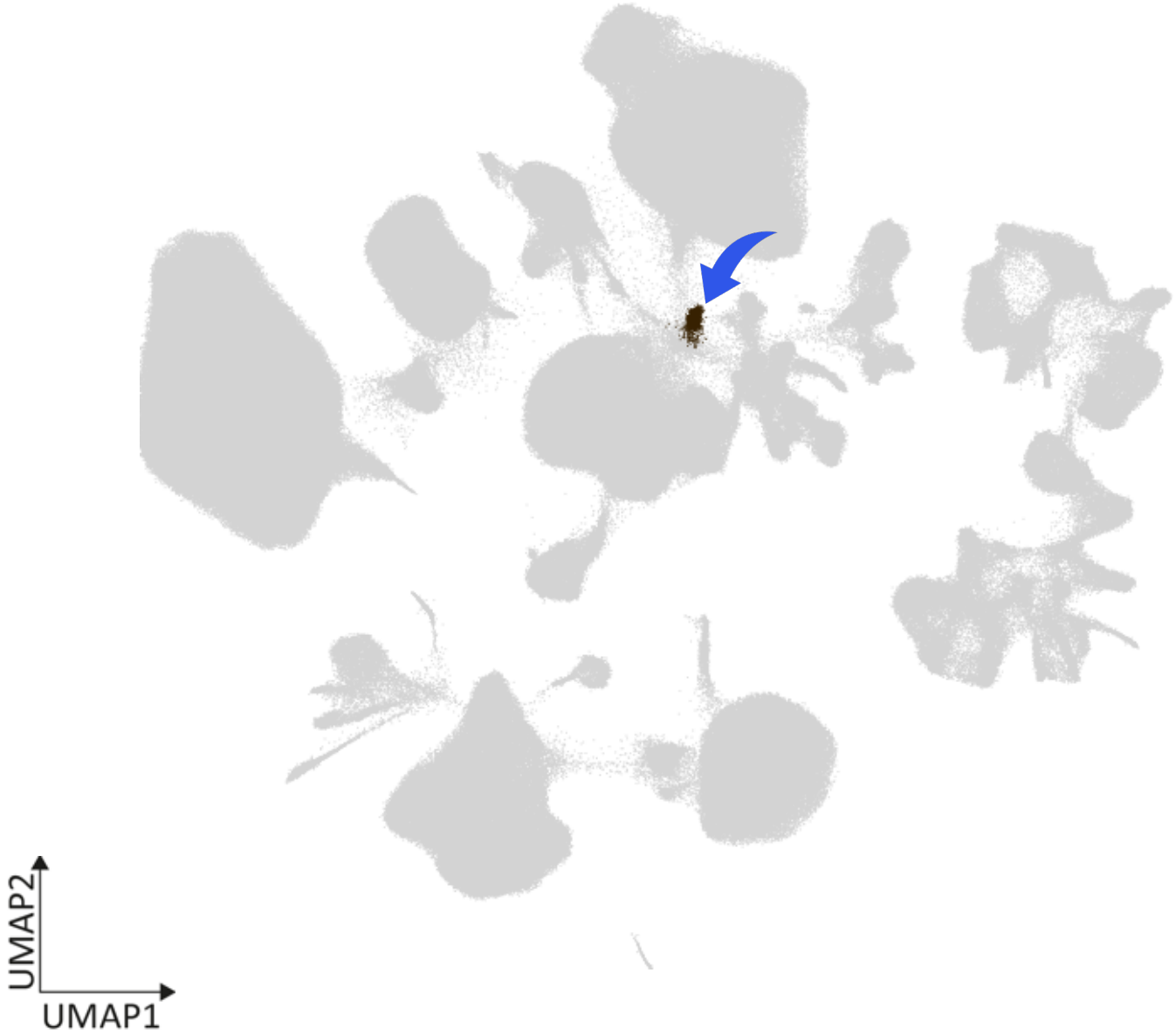
Reduced Batch Effects



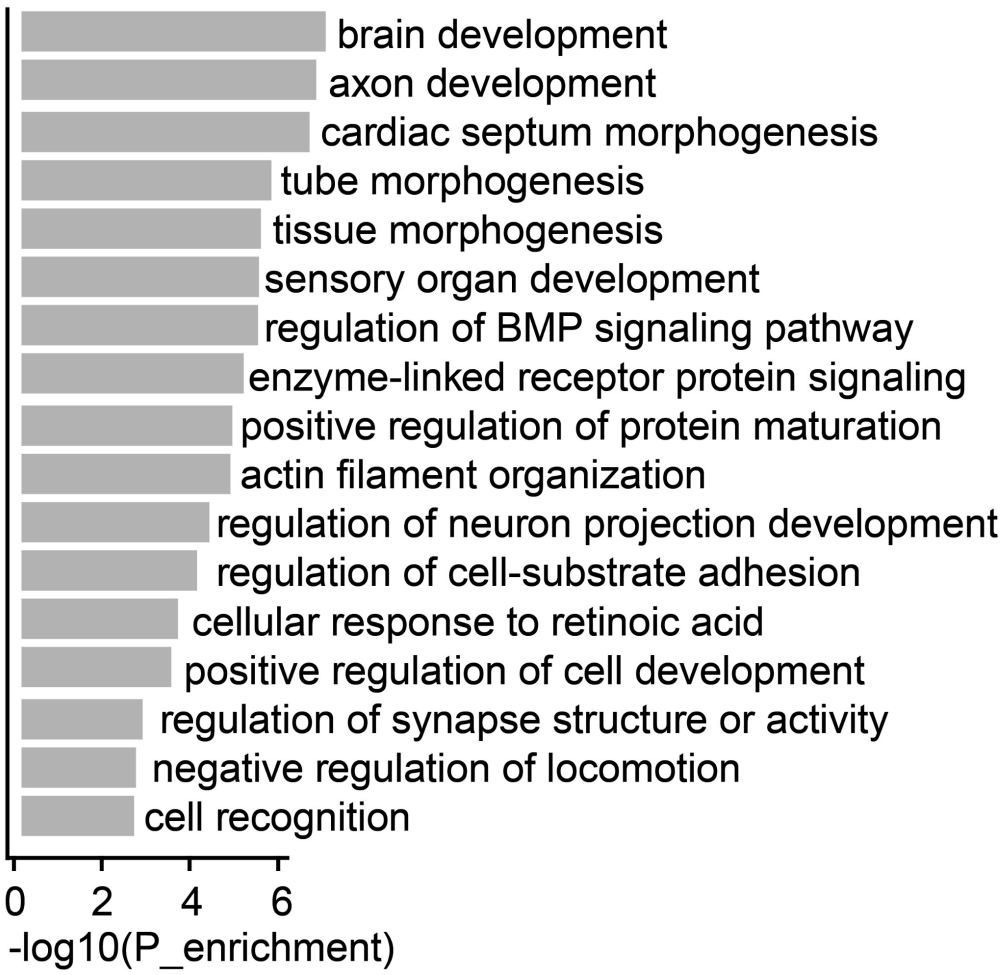
Hippocampal Cell Types



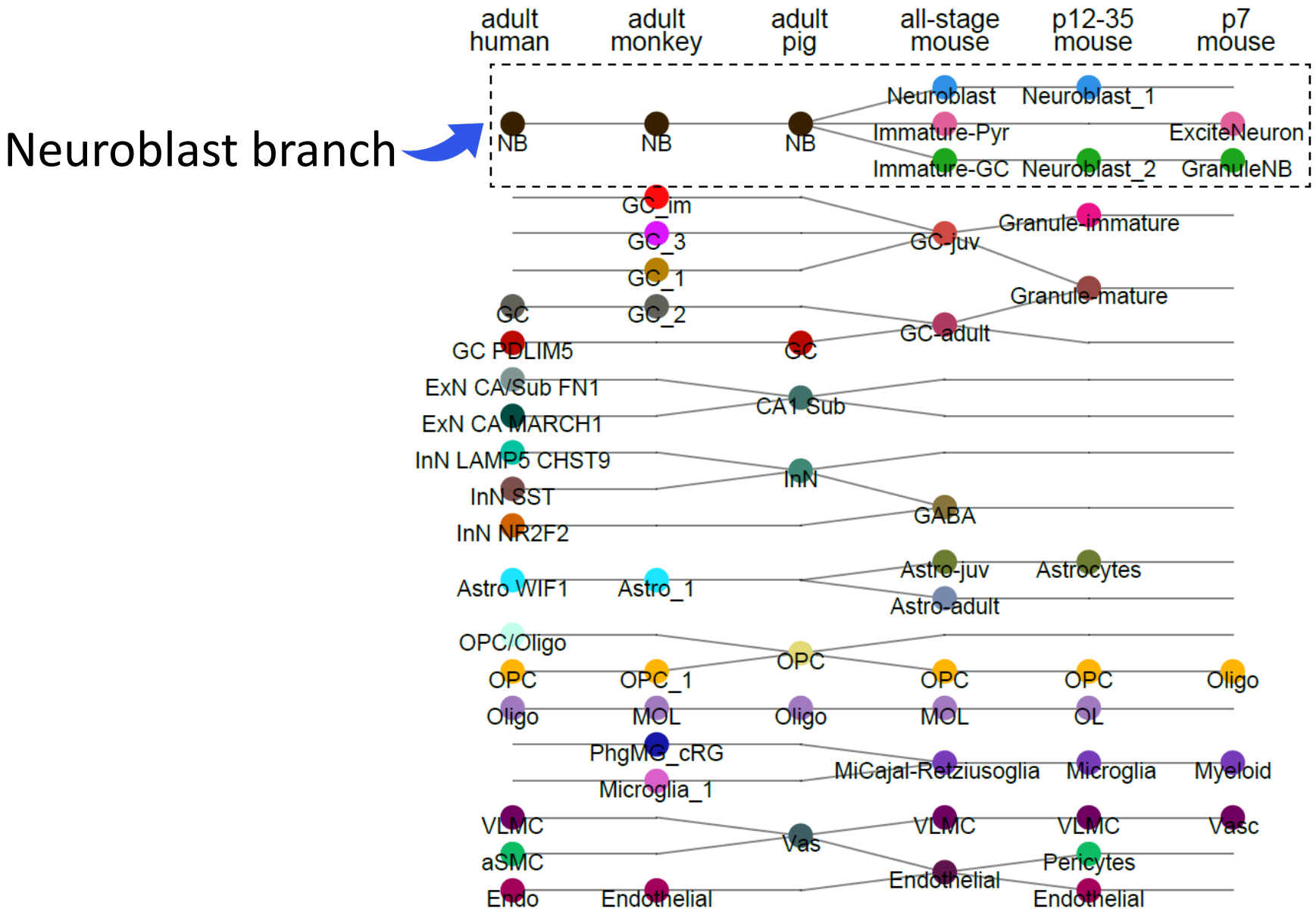
An Immature Neuron/Neuroblast Population



Evidence From Gene Signatures



Cross-Species Cell Type Harmonization



Summary of CellHint

- Tool for automated cell type harmonization and integration
- Framework for assembly of annotated cell atlases
- 12 established organ atlases

CellHint

GitHub: Teichlab/cellhint

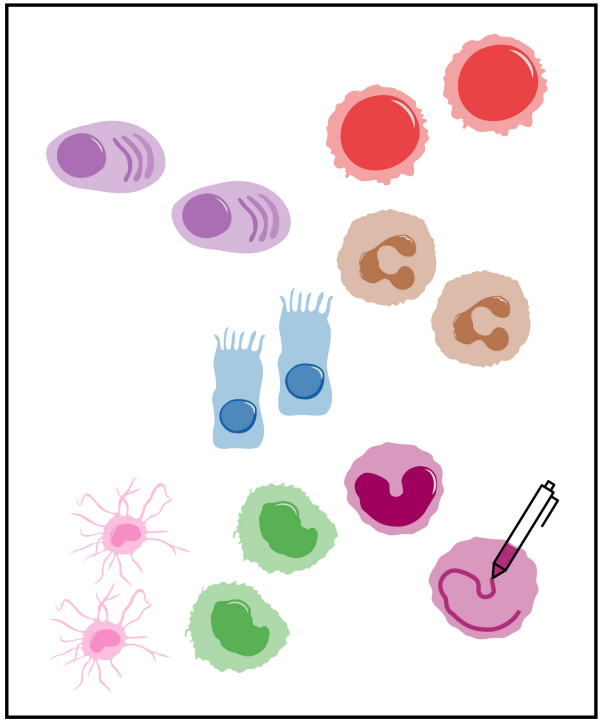
Tutorial: cellhint.readthedocs.io

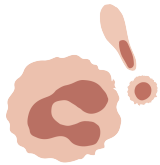
Takeaway

CellTypist

Annotate

Website: celltypist.org
GitHub: Teichlab/celltypist

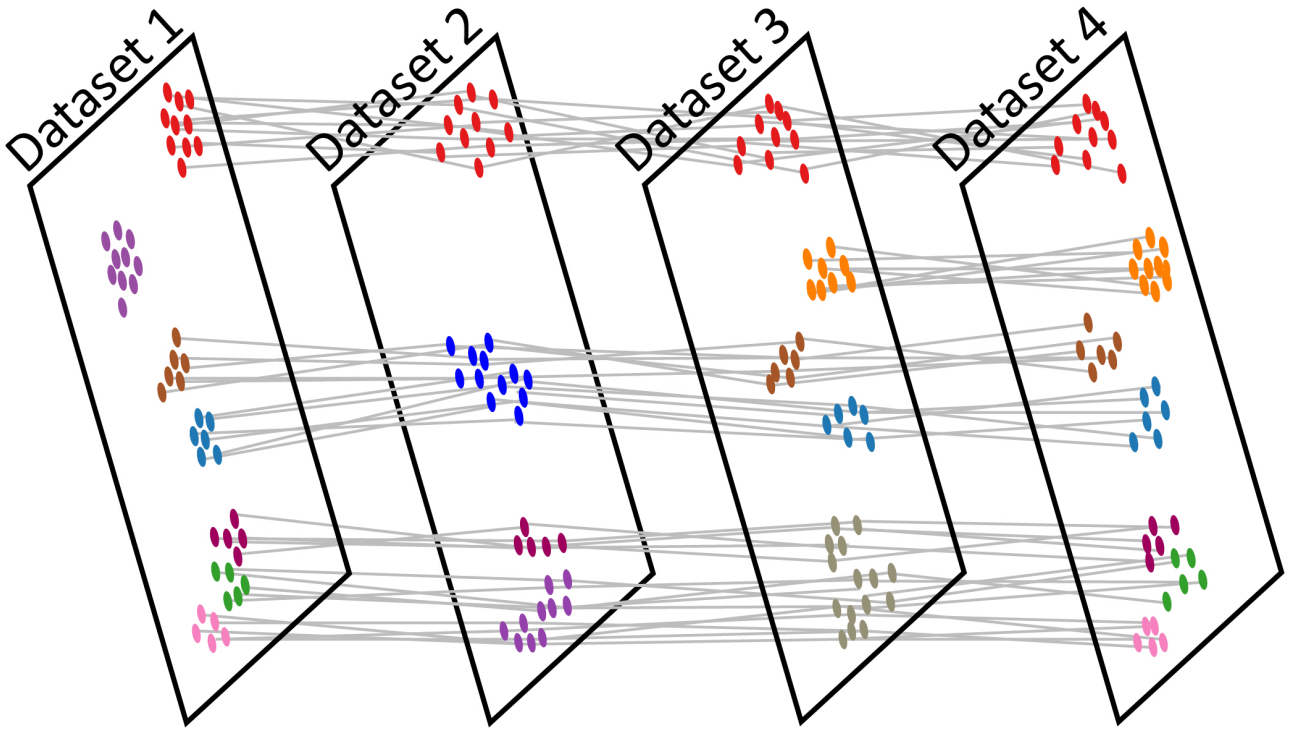




CellHint

Unify

GitHub: Teichlab/cellhint



Acknowledgements

Sarah Teichmann
Cecilia Domínguez Conde
Tomas Gomes
Kerstin Meyer
Jongeun Park
Lira Mamanova
Krzysztof Polanski
Peng He
Dinithi Sumanaweera

Martin Prete
Maria Keays

Muzlifah Haniffa
Simone Webb
Laura Jardine

Roser Vento-Tormo
Regina Hoo

Joanne Jones
Lorna Jarvis
Sarah Howlett
Dan Rainbow

Menna Clatworthy
Ondrej Suchanek
Benjamin Stewart
Kelvin Tuong

Louisa James
Hamish King

Kourosh Saeb-Parsy
Krishnaa Mahbubani

Donna Farber
Peter Sims
Steven Wells

Thank you to donors and their families!