

HUMAN
CELL
ATLAS

Benchmarking of Single Cell RNA Sequencing Protocols for Cell Atlas Projects

EuroBioc2019, Brussels
Dec 10, 2019

cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico



Elisabetta Mereu, PhD
Single Cell Genomics Group
National Centre for Genomic Analysis
Centre for Genomic Regulation (CNAG-CRG)
Barcelona, Spain
Email: elisabetta.mereu@cnag.crg.eu
Web: <https://elisabettamereu.netlify.com>



@elisabetta1412



<https://github.com/elimereu>

Background

Cell Types



Dendritic cells



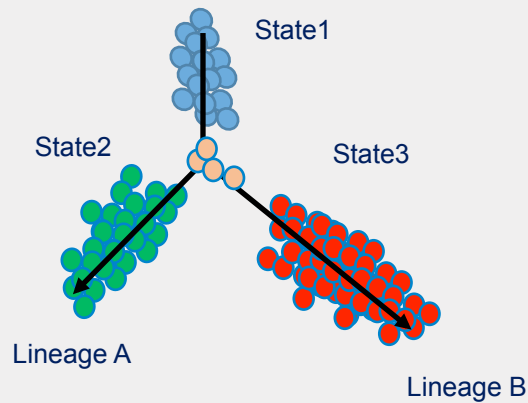
B cells



Endothelial cells

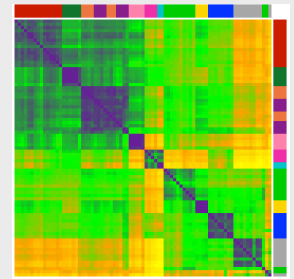


Cell States and lineages

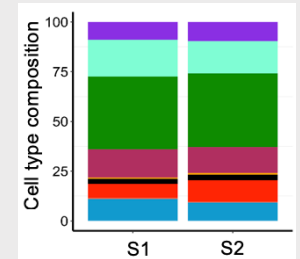


Evaluation

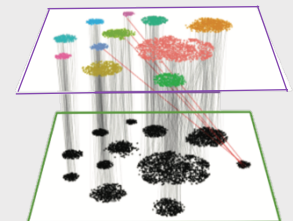
Reproducibility



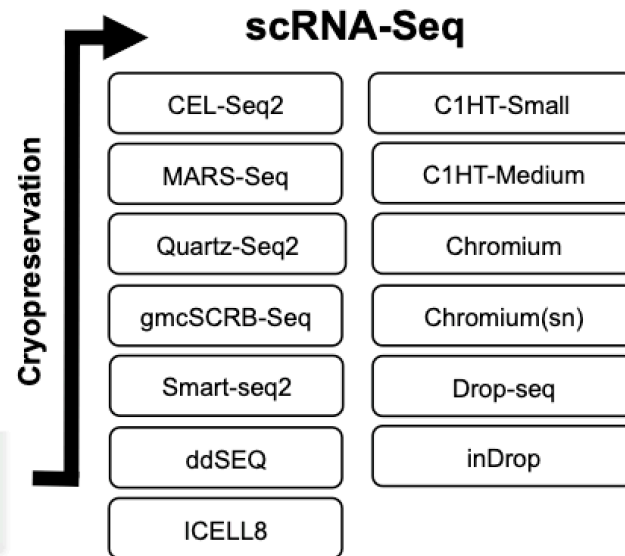
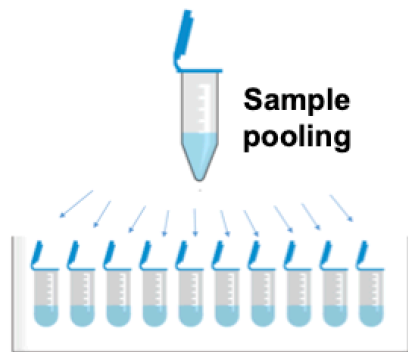
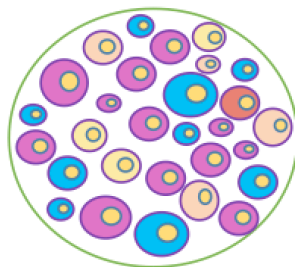
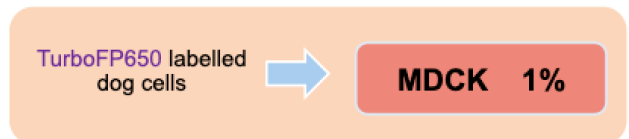
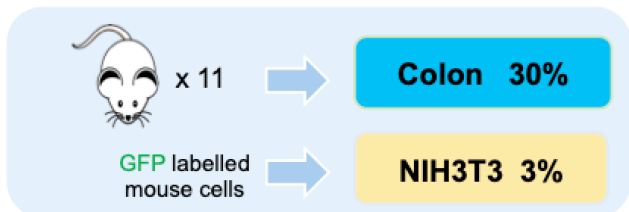
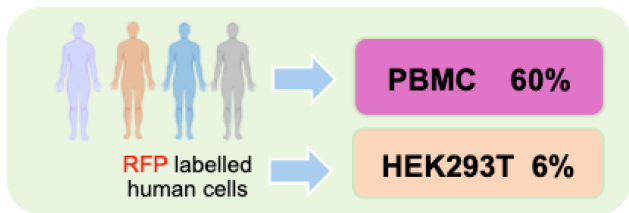
Integrity



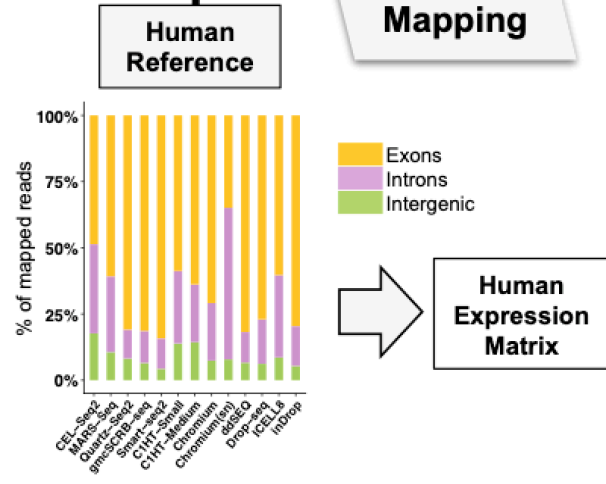
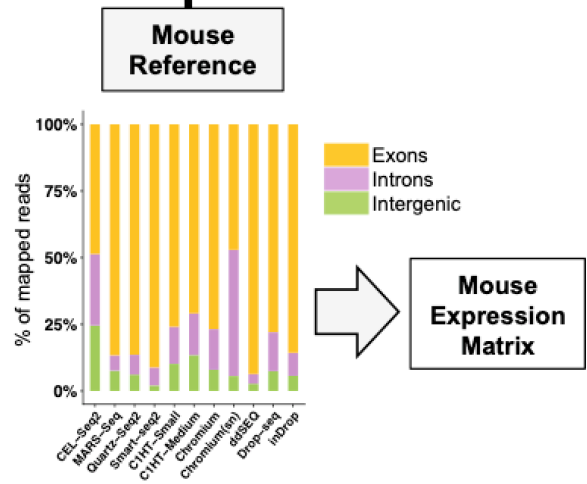
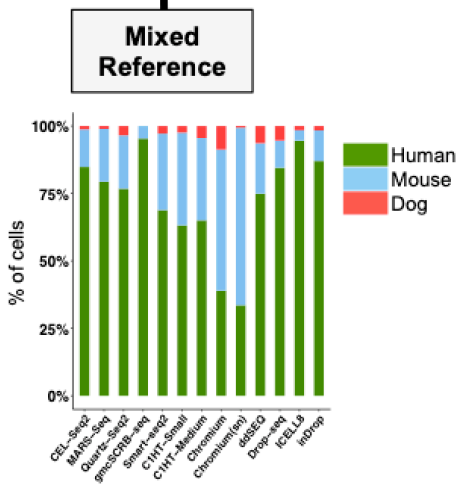
Predictive values



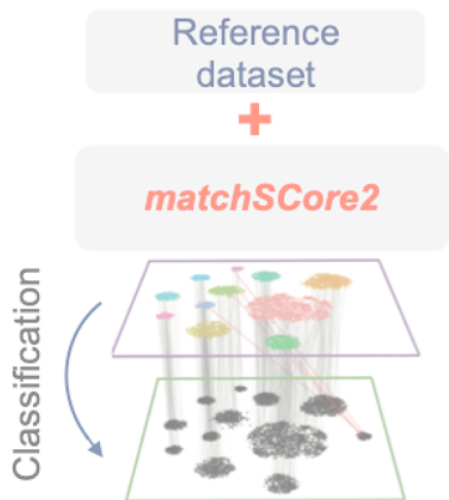
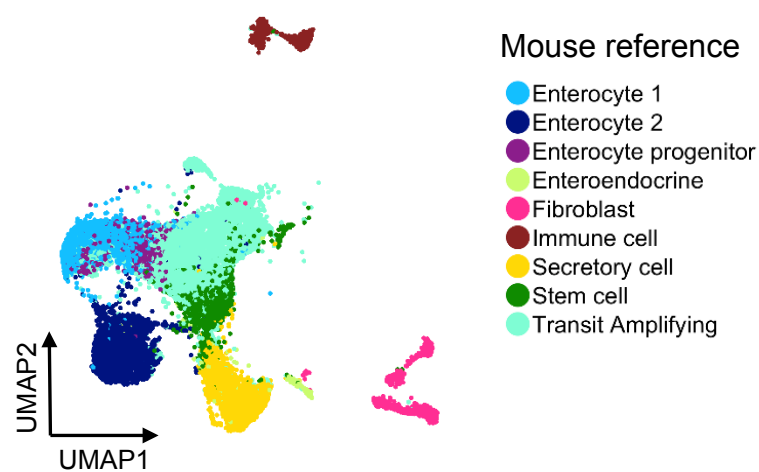
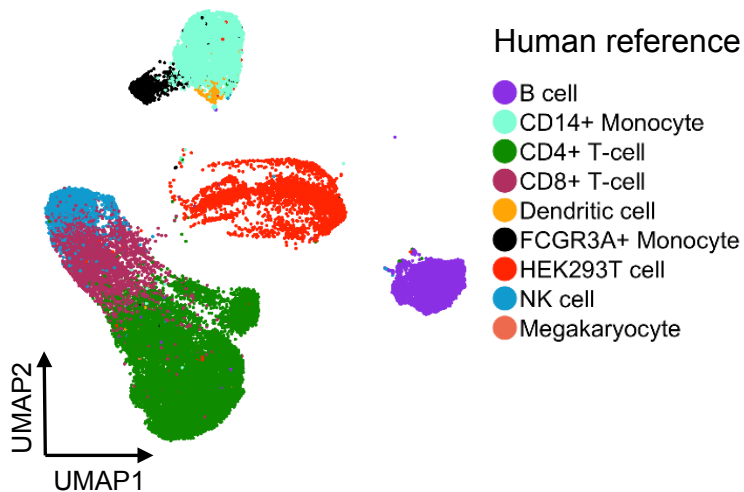
Reference sample



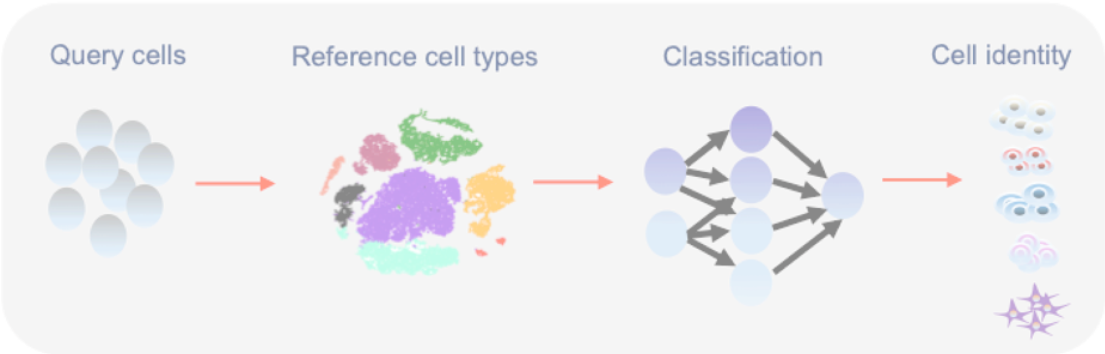
Cryopreservation



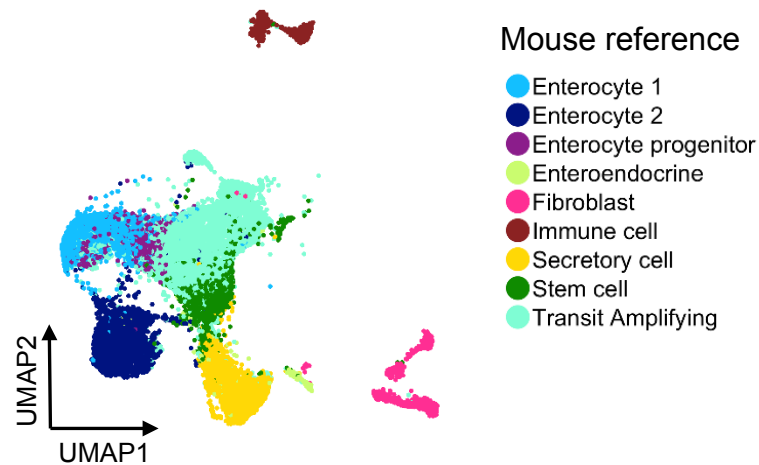
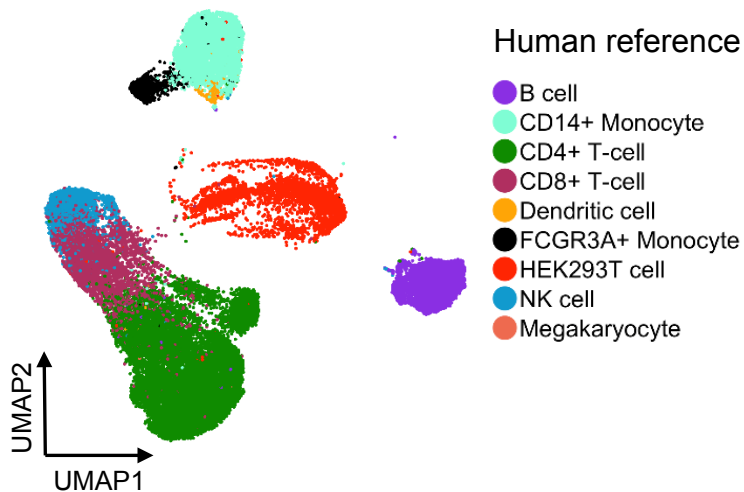
Reference datasets



matchScore2 allows the fast annotation of unknown cell types using a reference dataset



Reference datasets

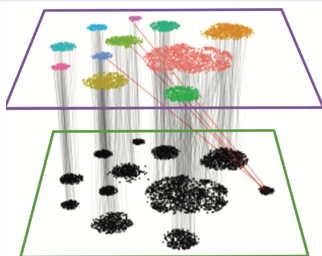


Reference dataset



matchScore2

Classification



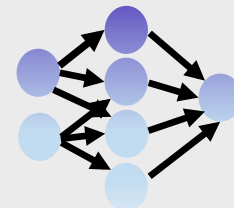
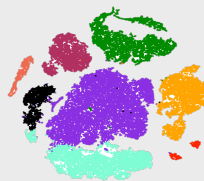
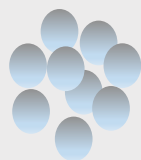
matchScore2 allows the fast annotation of unknown cell types using a reference dataset

Query cells

Reference cell types

Classification

Cell identity



cnag

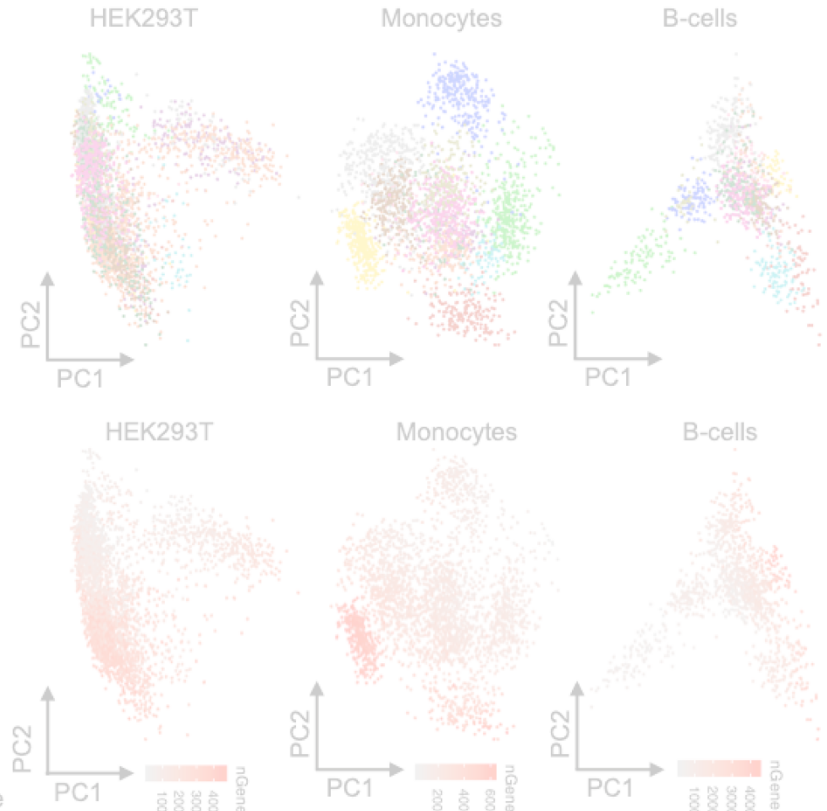
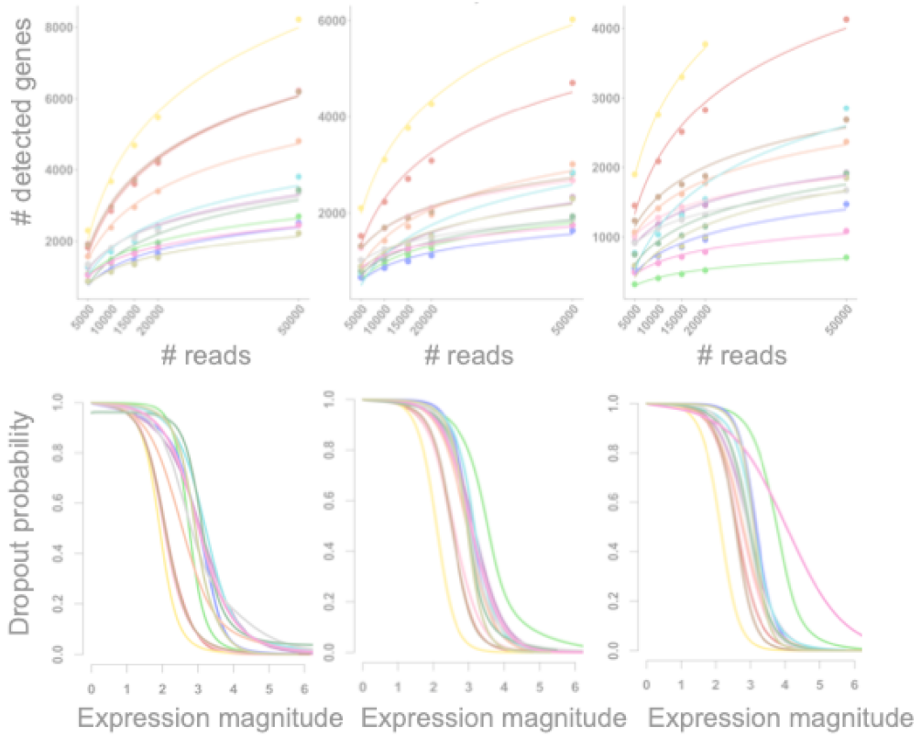
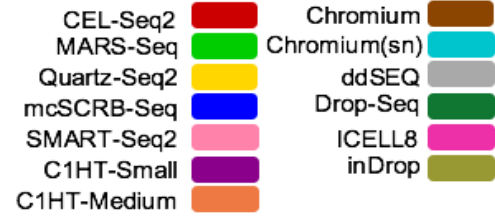
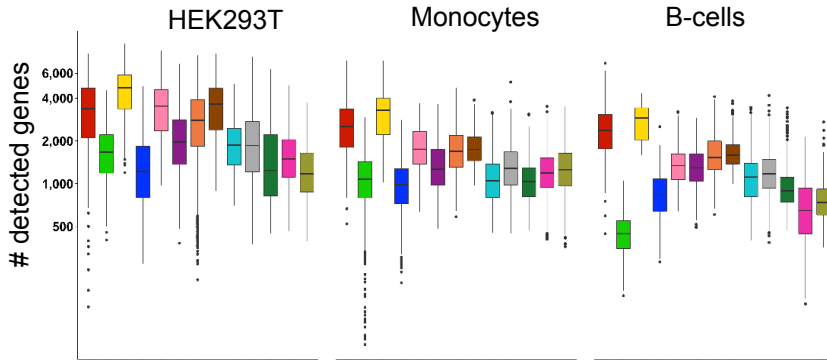
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico



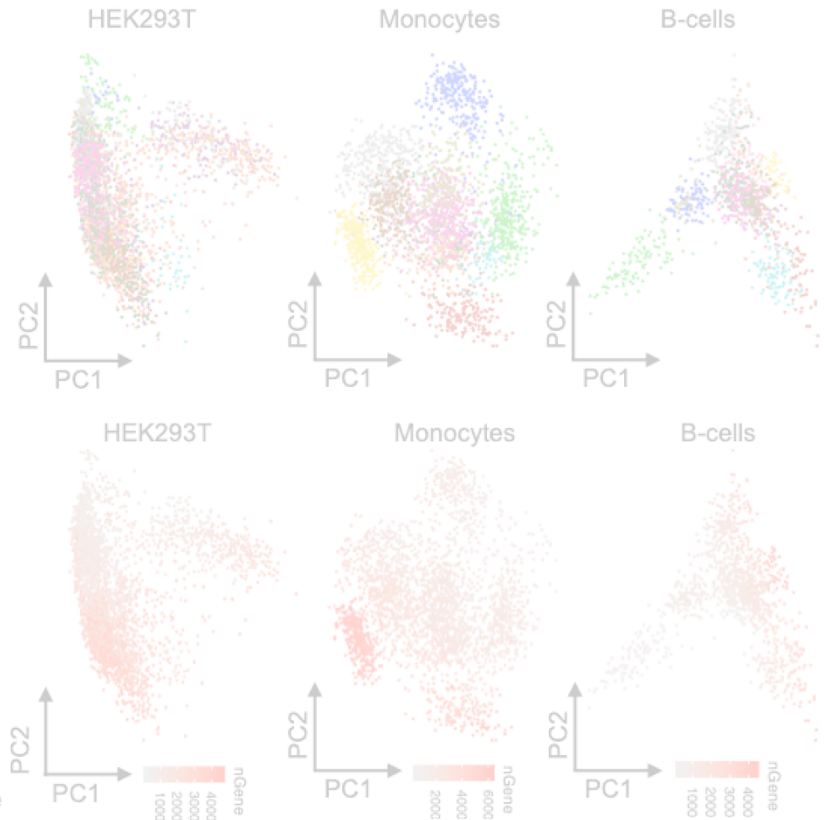
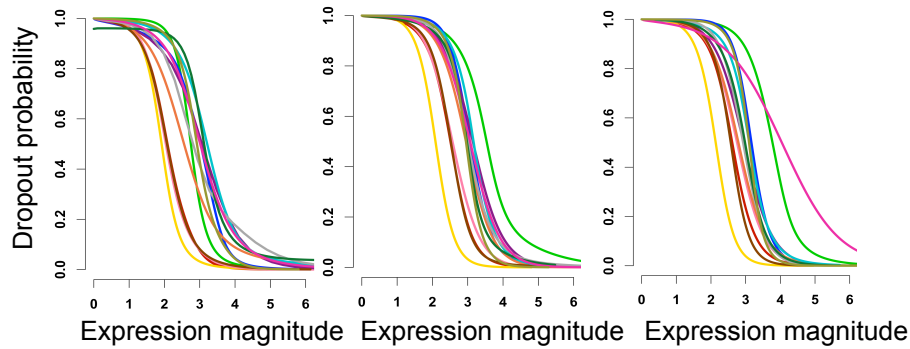
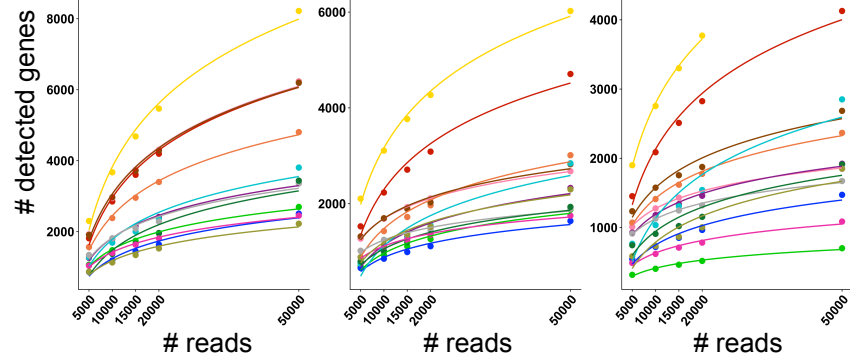
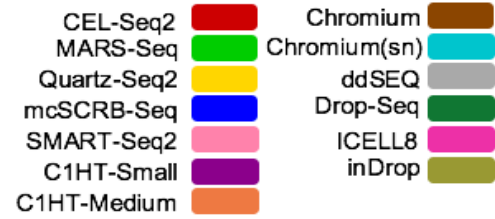
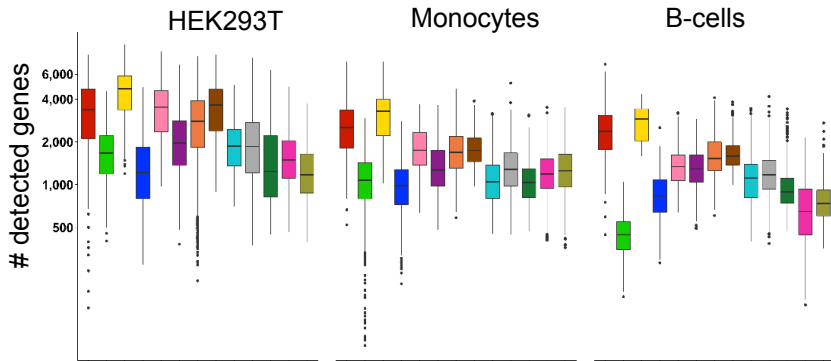
HUMAN
CELL
ATLAS



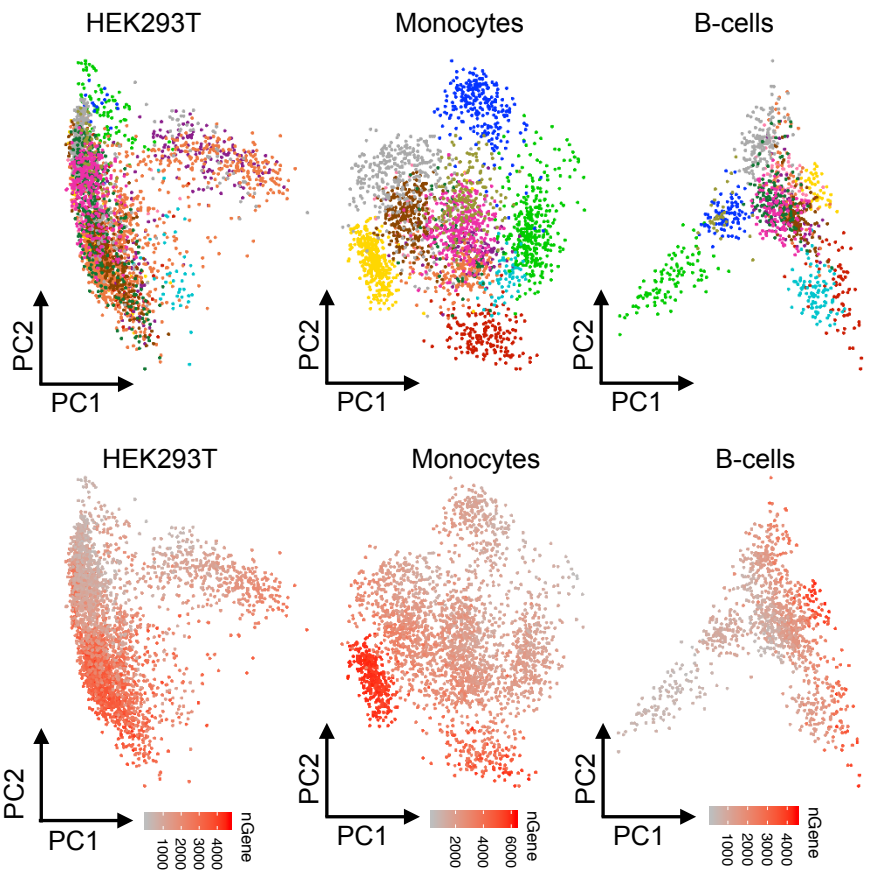
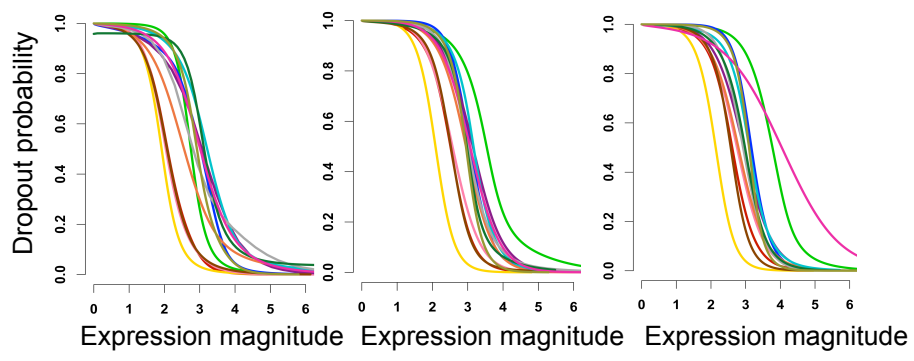
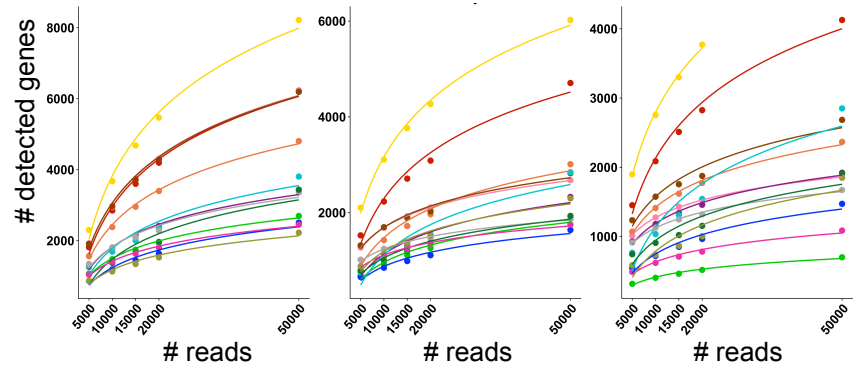
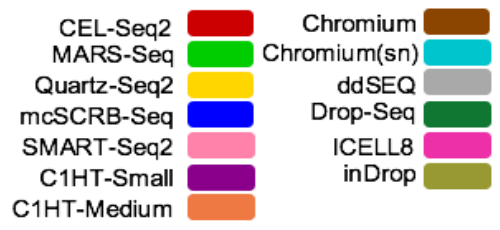
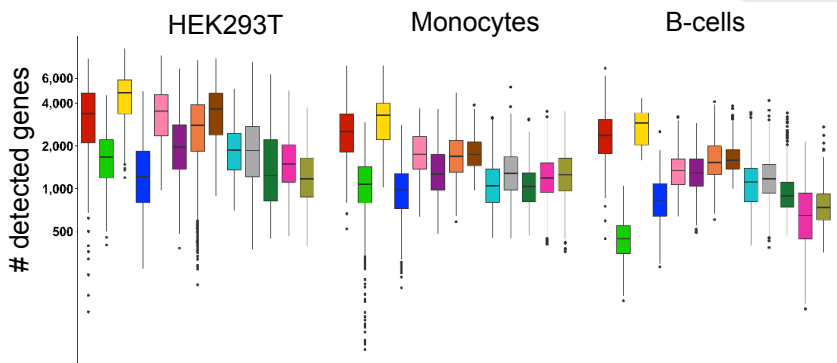
Gene Detection



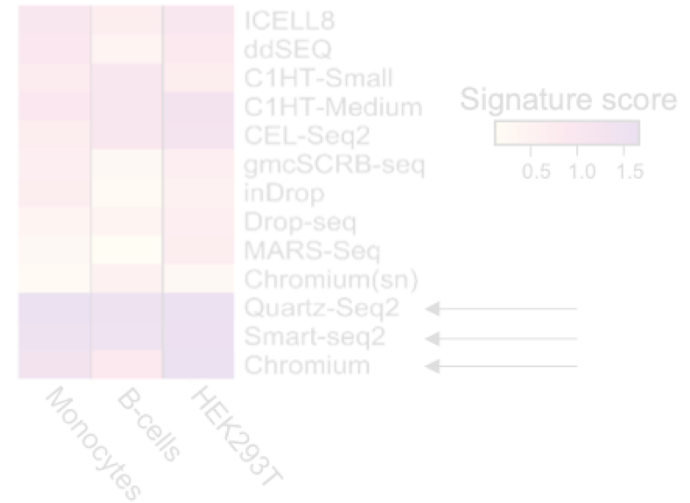
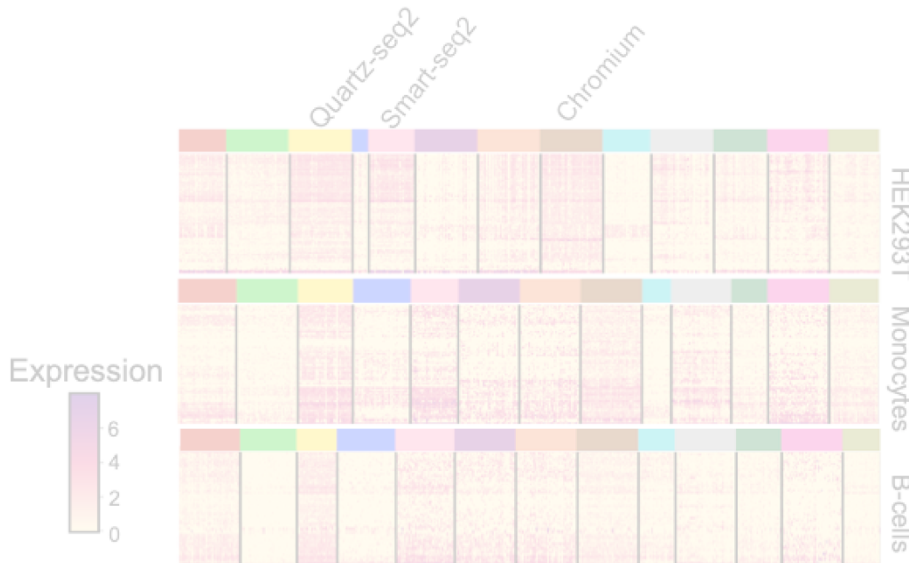
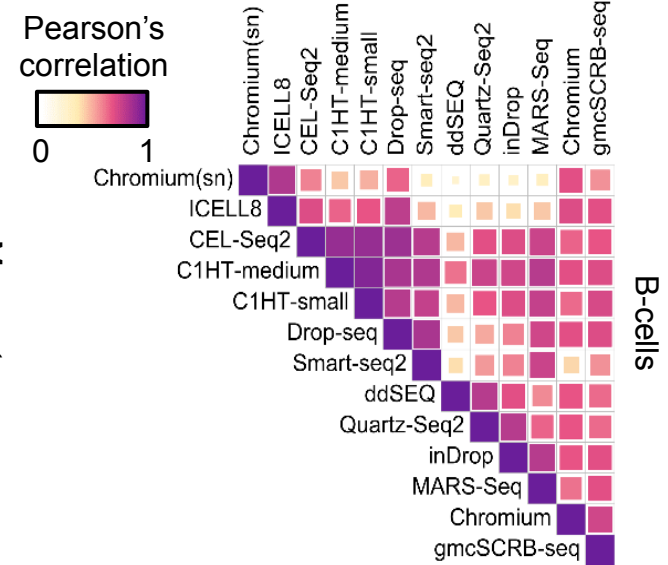
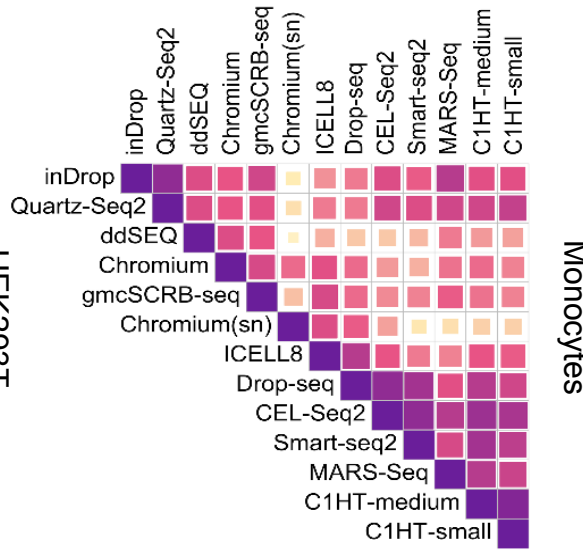
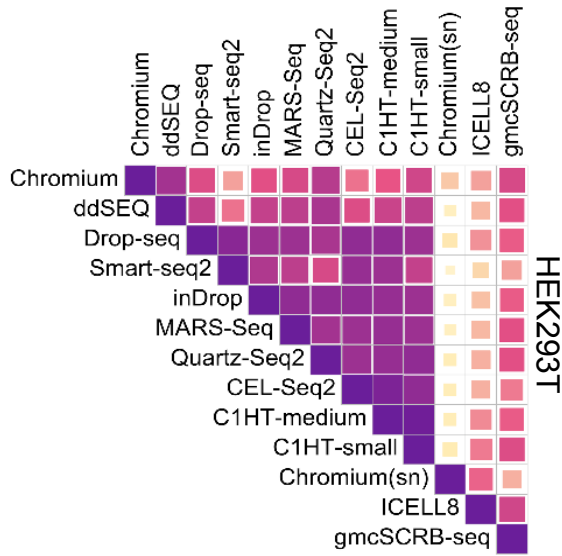
Gene Detection



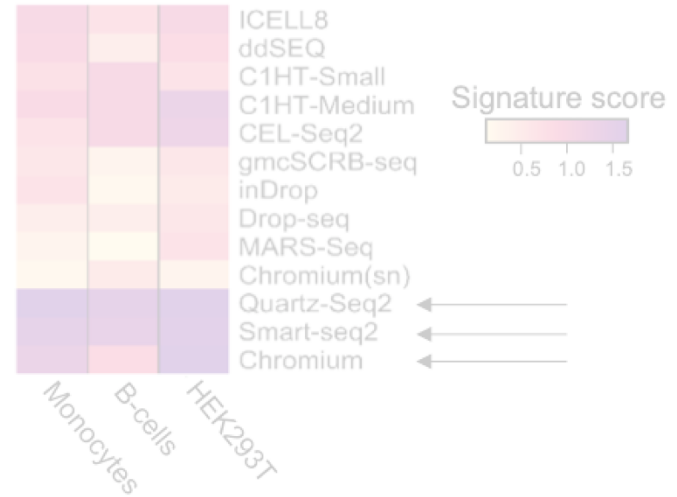
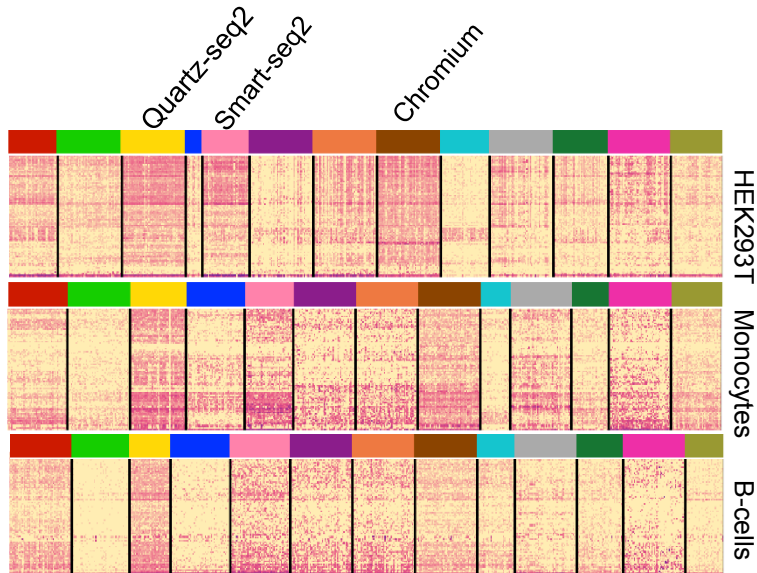
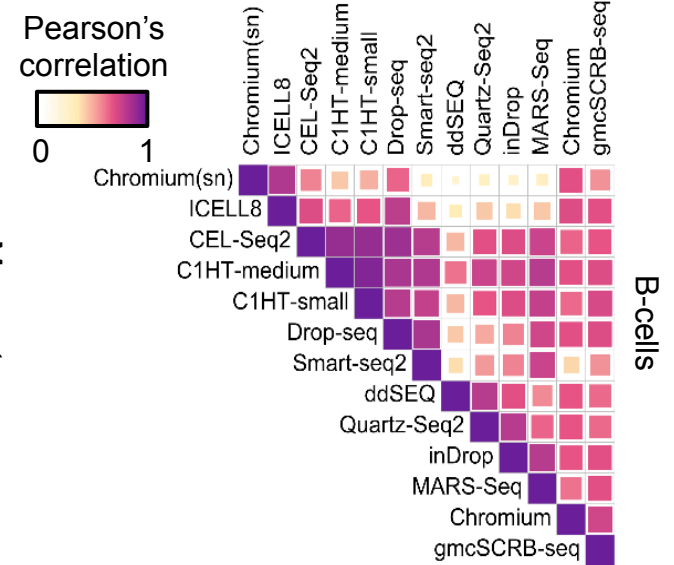
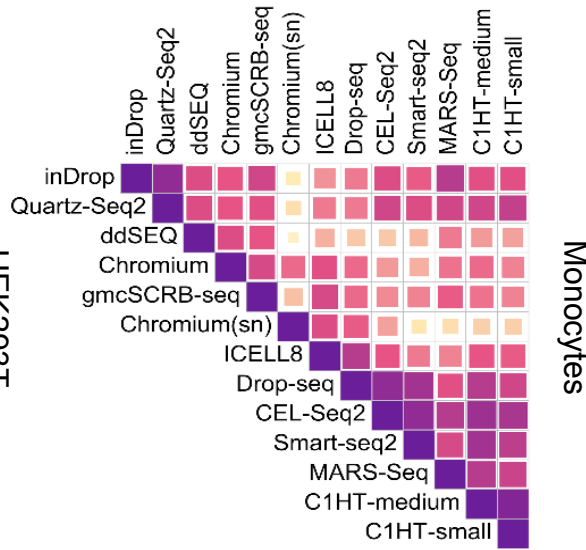
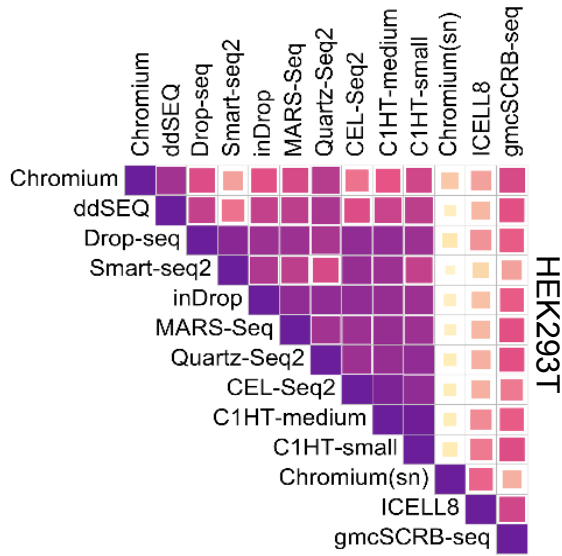
Gene Detection



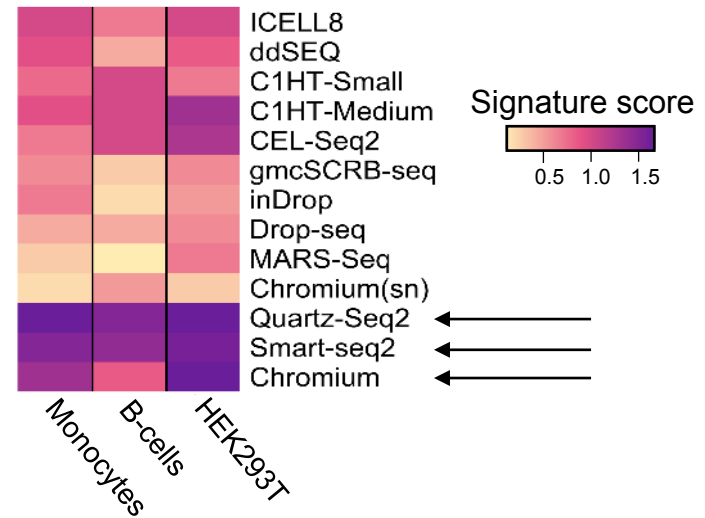
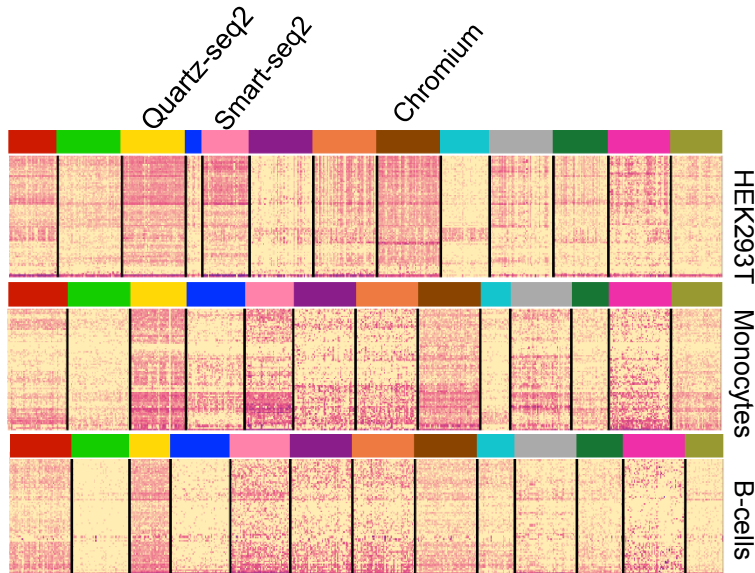
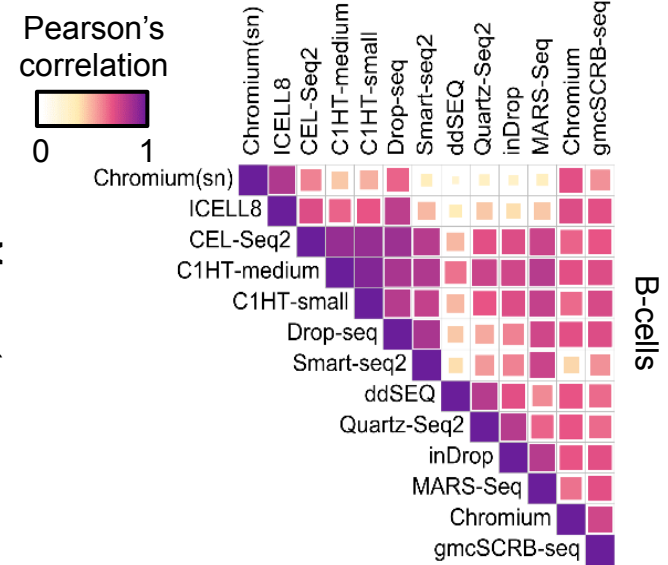
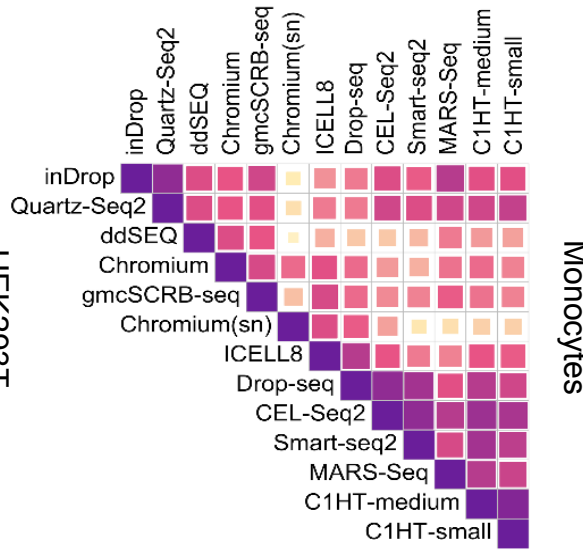
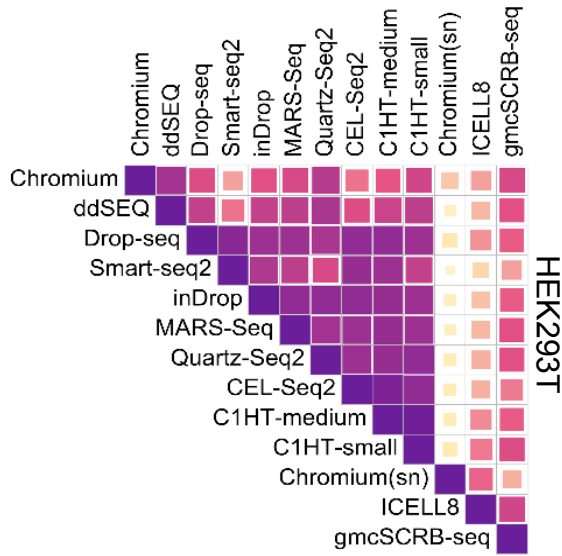
Correlation of gene expression levels



Correlation of gene expression levels

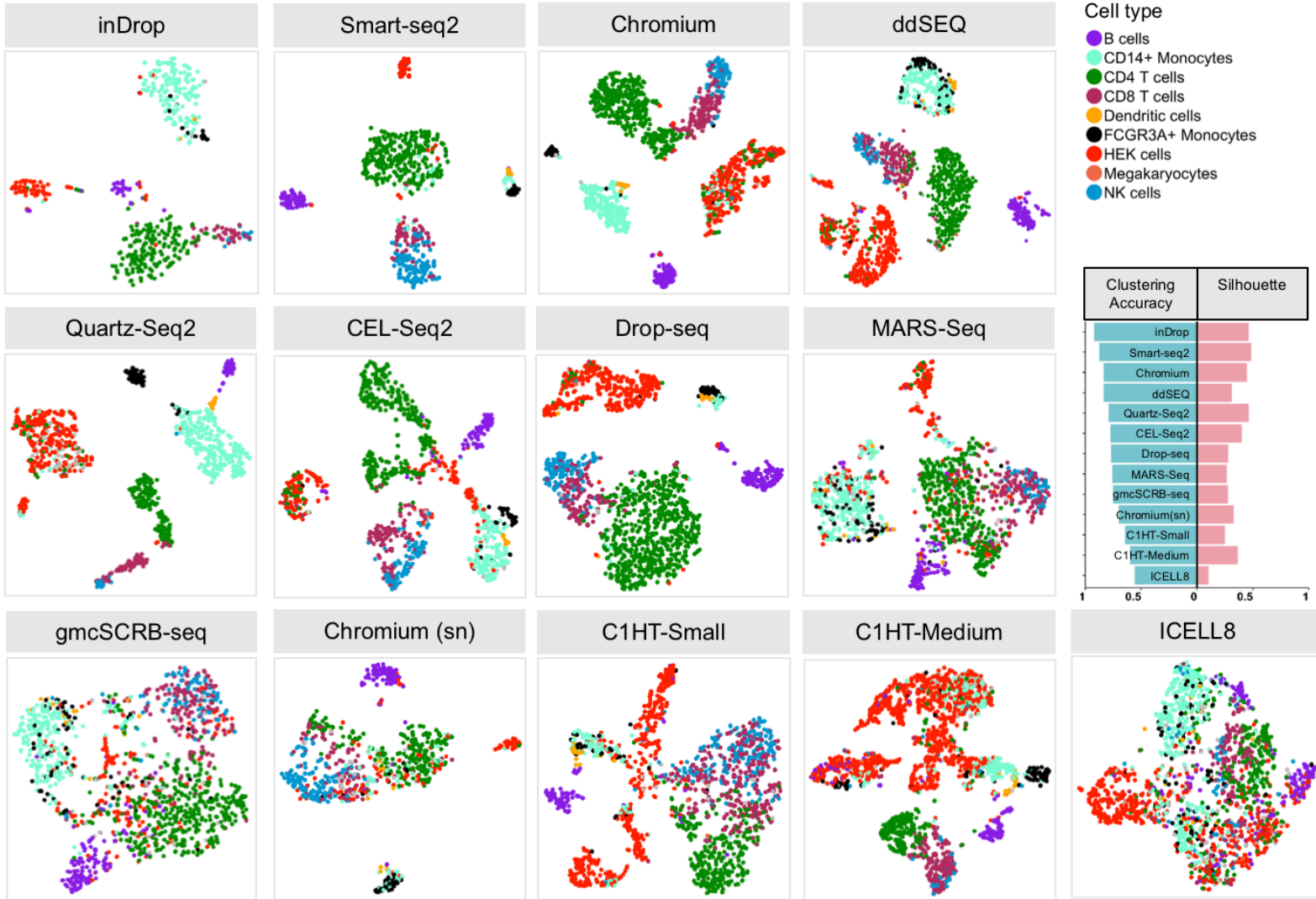


Correlation of gene expression levels

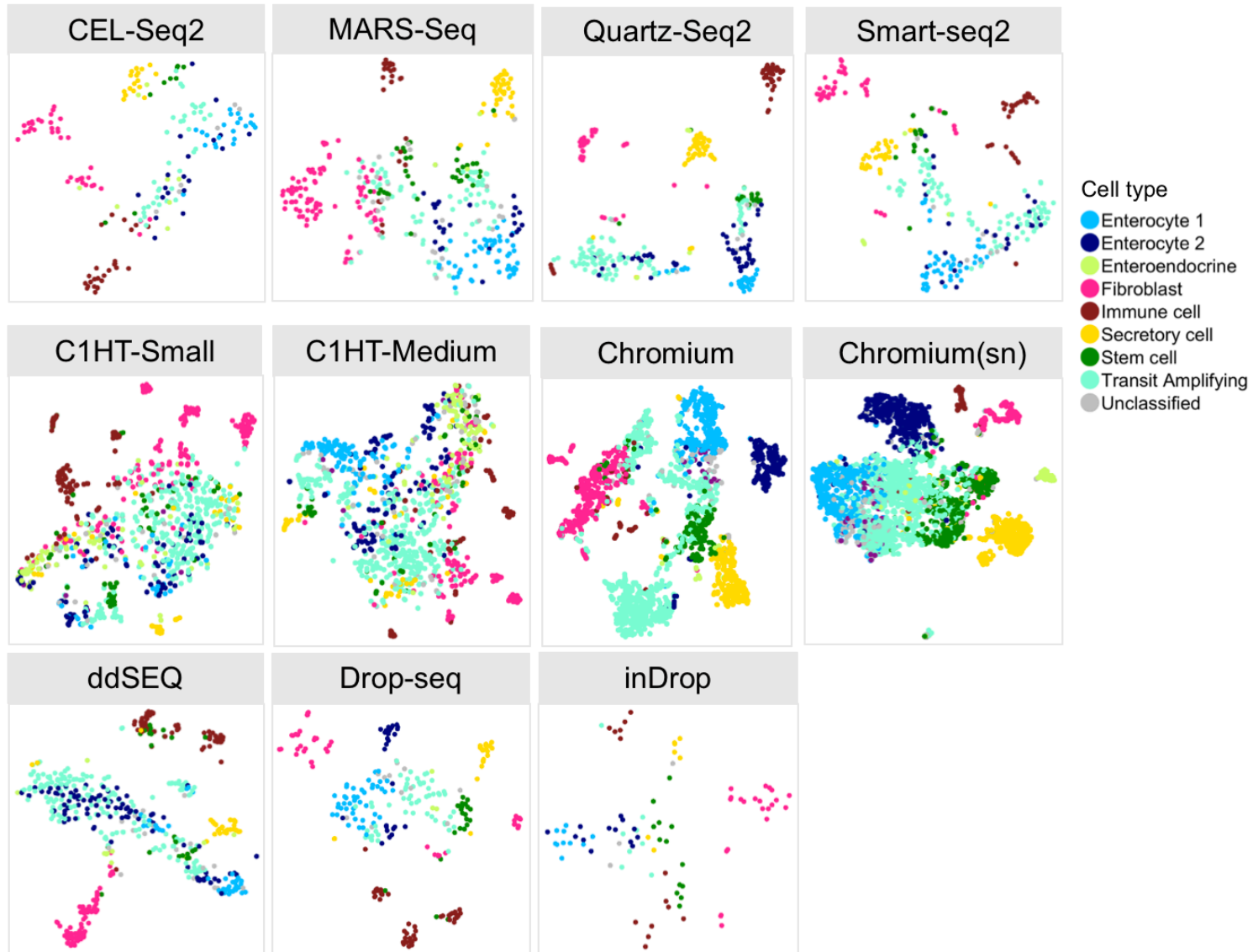


Human Clustering

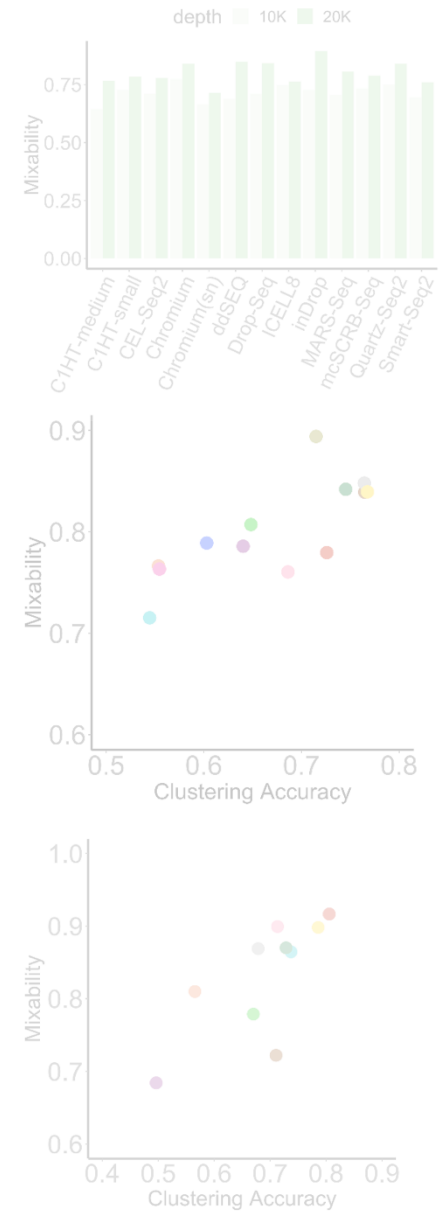
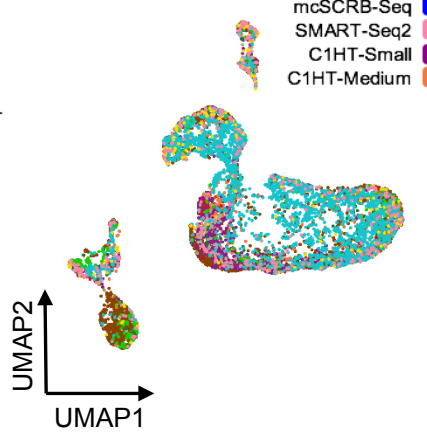
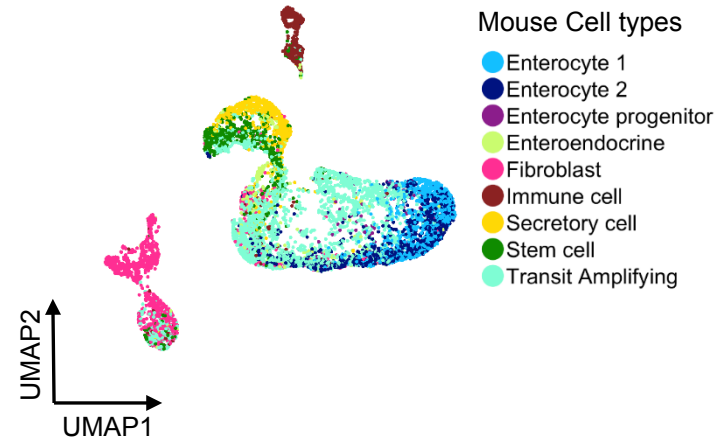
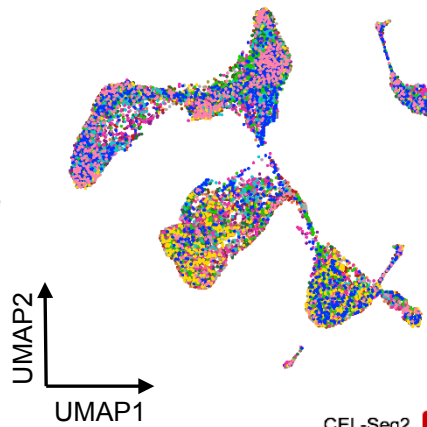
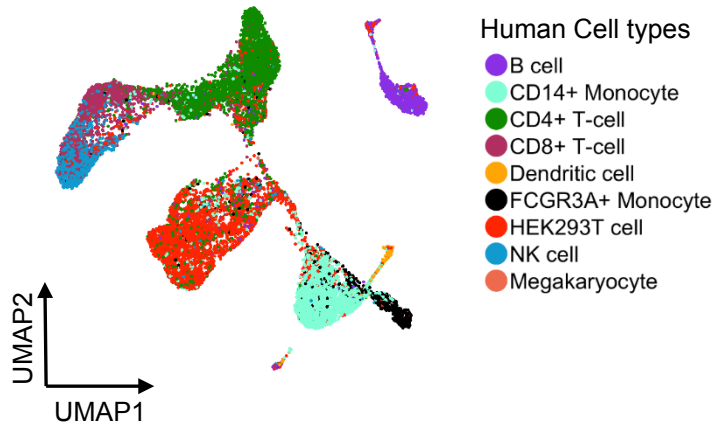
order →



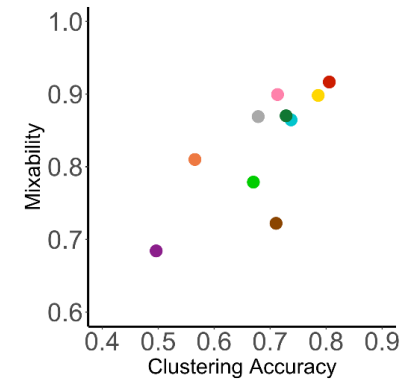
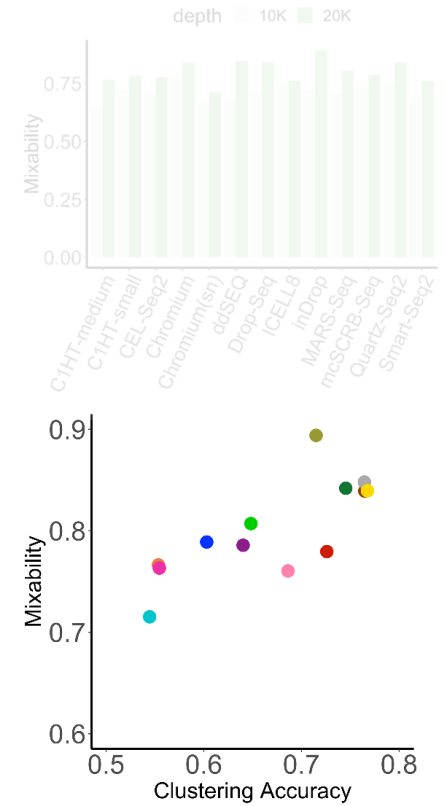
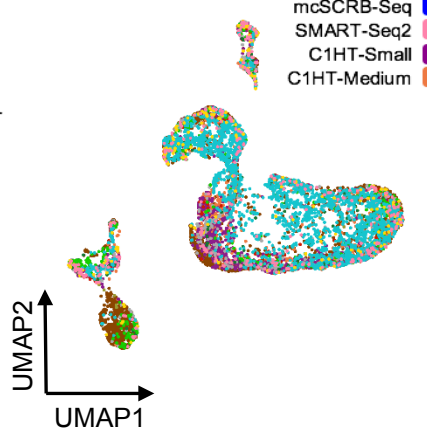
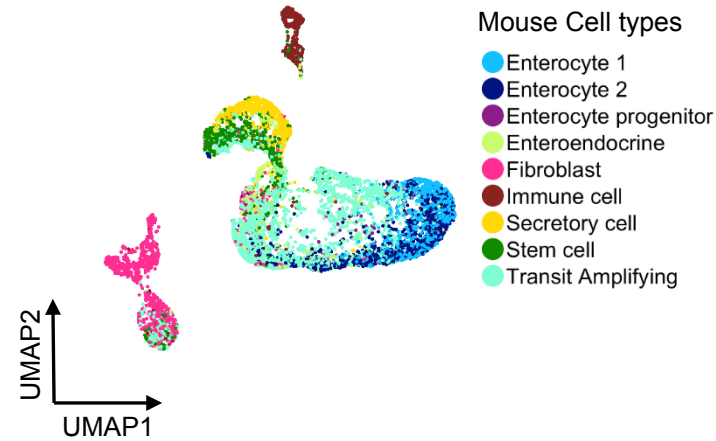
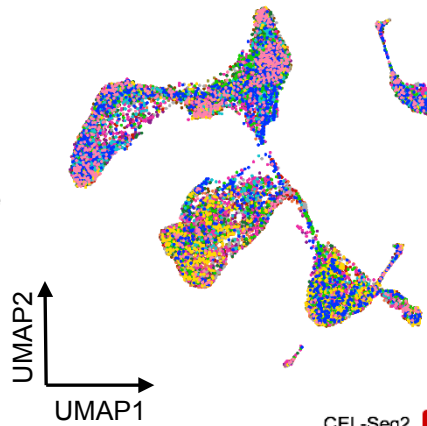
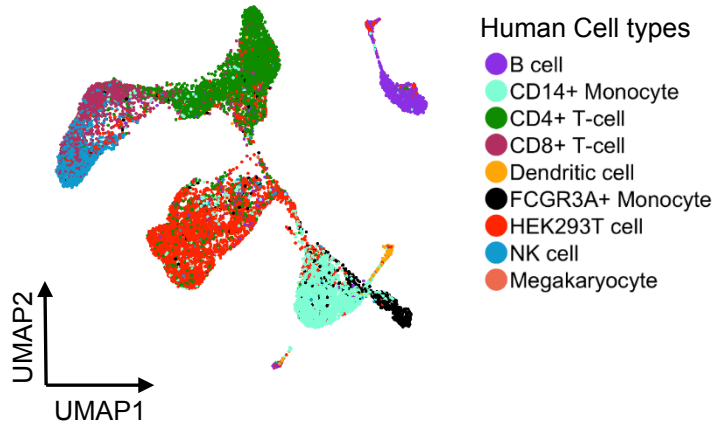
Mouse Clustering



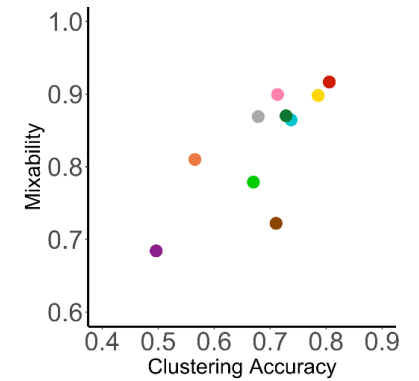
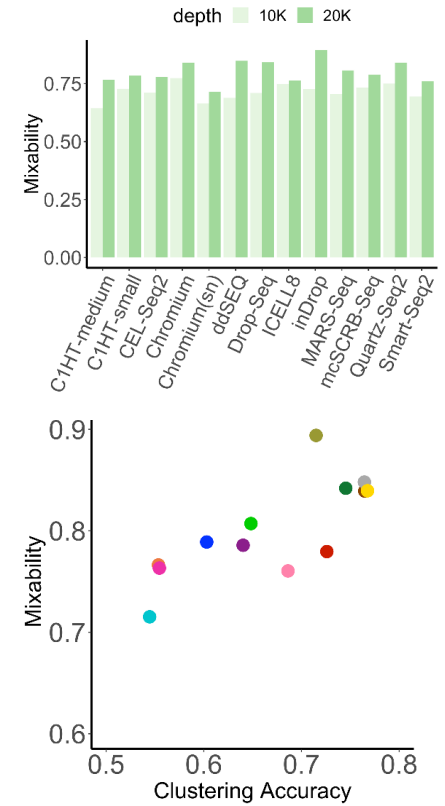
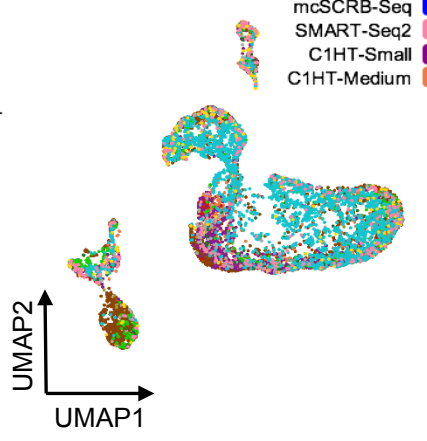
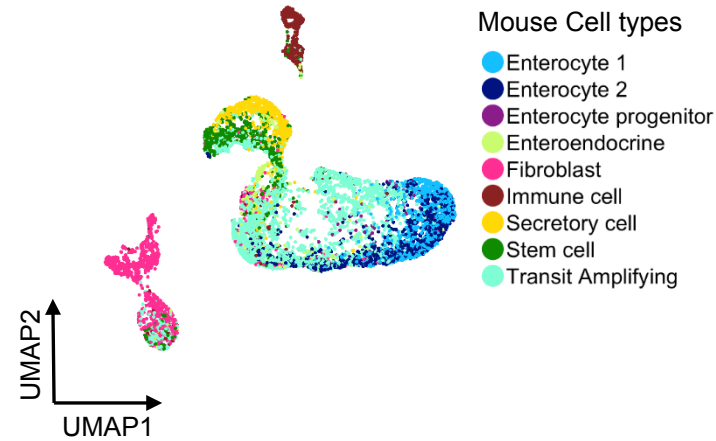
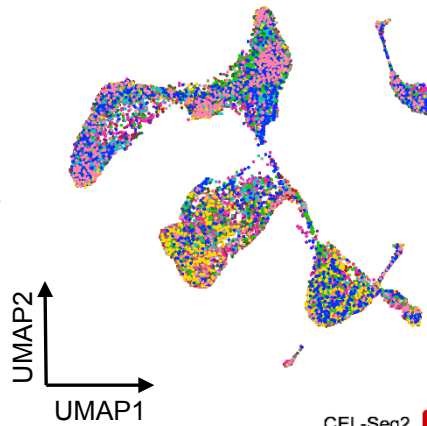
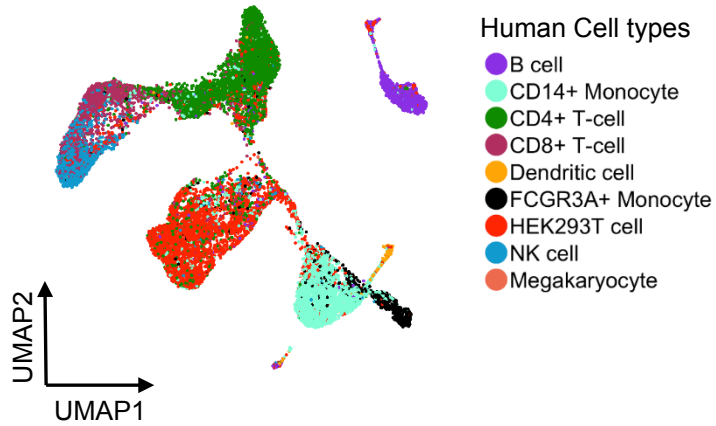
Integrability



Integrability



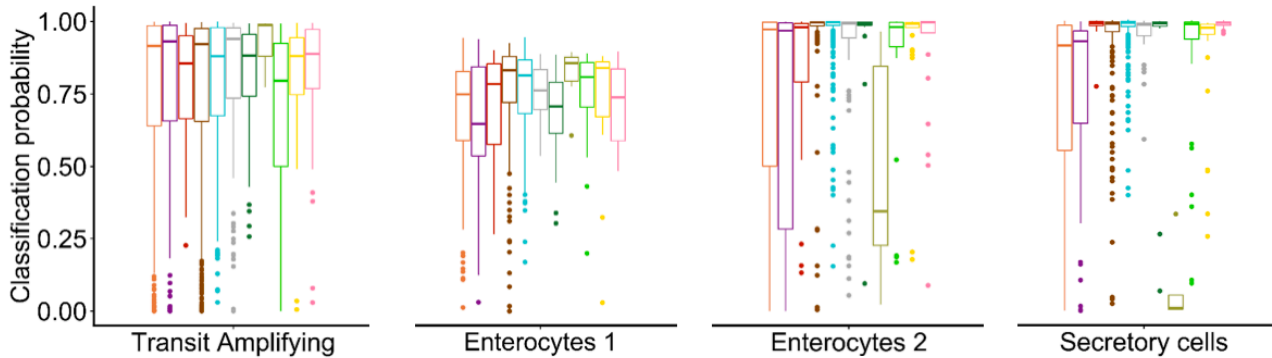
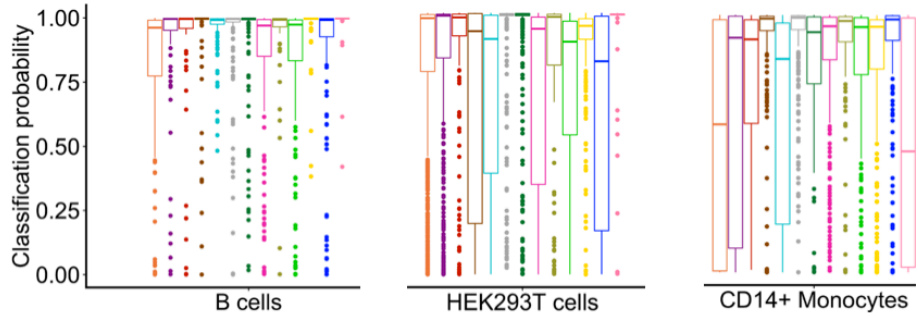
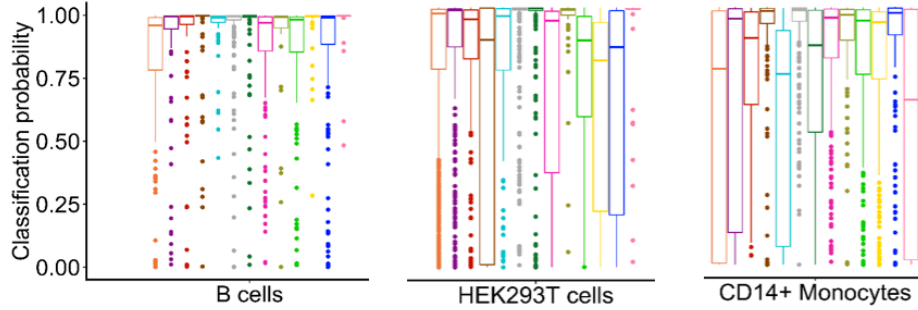
Integrability



Mappability

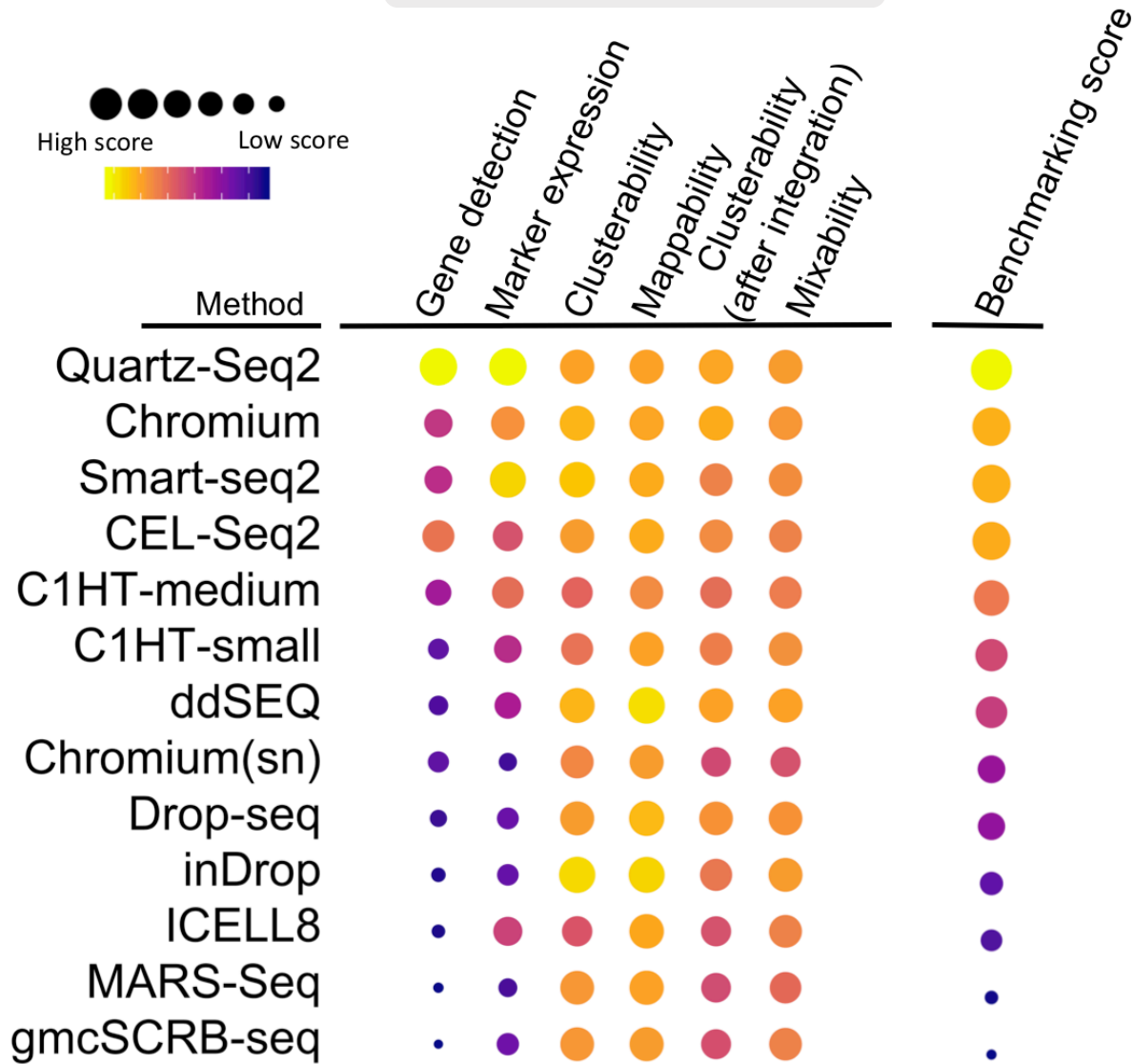
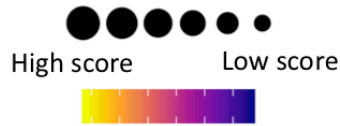
Downsampled to 20K

Downsampled to 10K

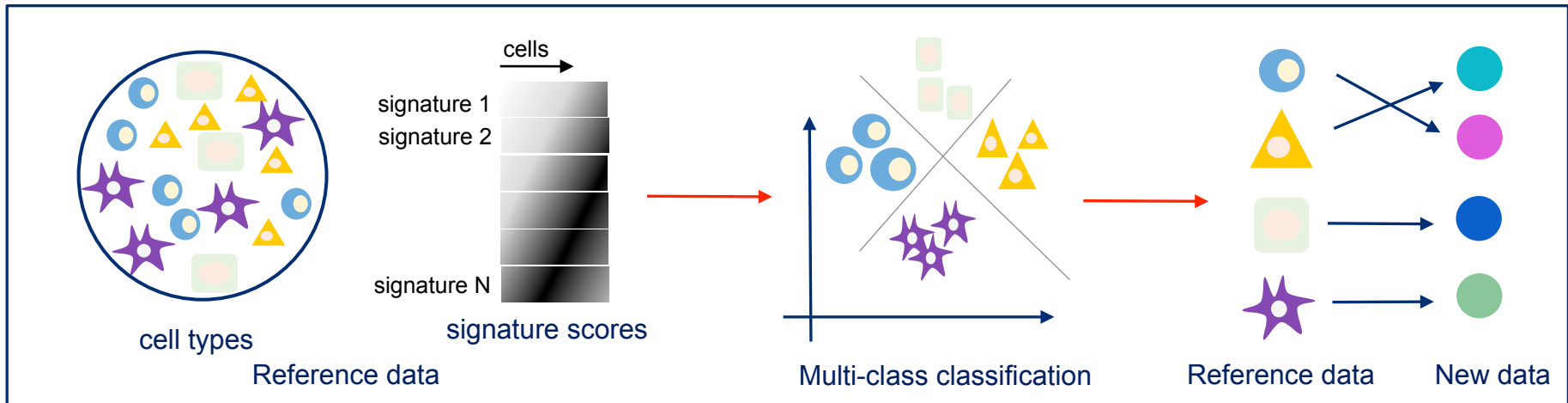


- C1HT-medium
- C1HT-small
- CEL-Seq2
- Chromium
- Chromium(sn)
- ddSEQ
- Drop-Seq
- ICELL8
- inDrop
- MARS-Seq
- Quartz-Seq2
- mcSCRB-Seq
- Smart-Seq2

Benchmarking scores



matchScore2: comparing datasets at cell and gene level

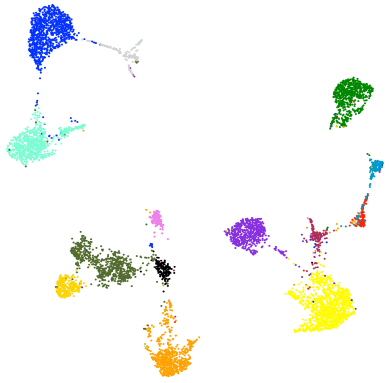


- *matchScore2* facilitates the annotation task by leveraging large-scale reference data.
- *matchScore2* trains a multinomial logistic model on the reference dataset.
- The main assumption of the model is that the number of cells N_k from each cell type and their proportions p_k are the parameters of a multinomial distribution
$$M \sim \text{multinom}(N = (N_1, \dots, N_m), p = (p_1, \dots, p_m)).$$
- The signature scores S_{jk} for the cell c_j and cell type T_k are used as predictors of the model.



matchScore2: the lung atlas

OLD

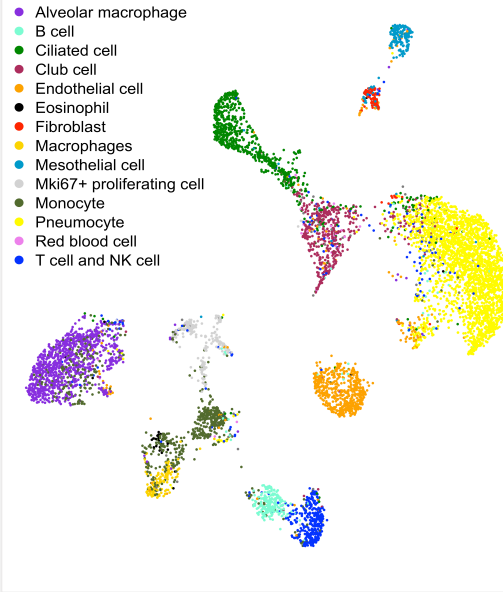


Reference cell types

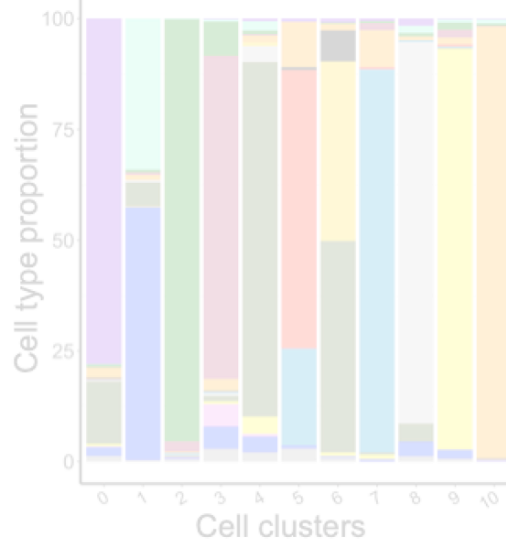
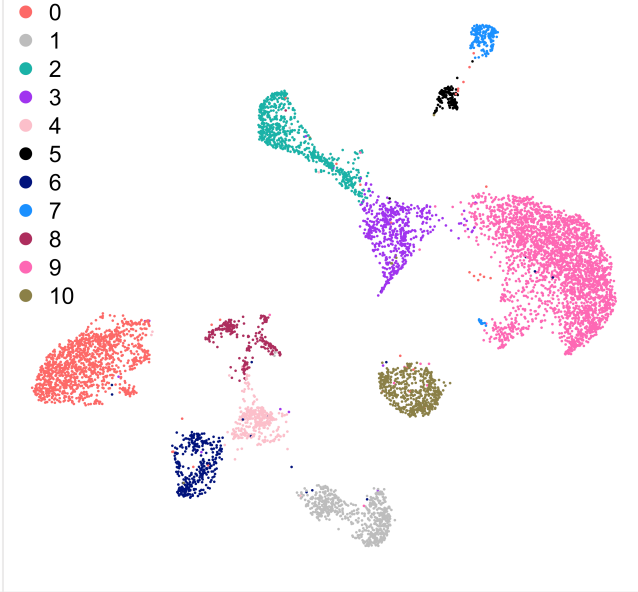
- Alveolar macrophage
- B cell
- Ciliated cell
- Club cell
- Endothelial cell
- Eosinophil
- Fibroblast
- Macrophages
- Mesothelial cell
- Mki67+ proliferating cell
- Monocyte
- Pneumocyte
- Red blood cell
- T cell and NK cell

YOUNG

matchScore2



Clustering

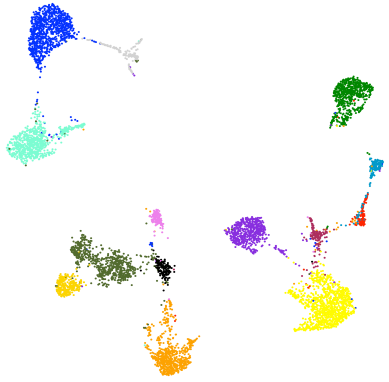


- Alveolar macrophage
- B cell
- Ciliated cell
- Club cell
- Endothelial cell
- Eosinophil
- Fibroblast
- Macrophages
- Mesothelial cell
- Mki67+ proliferating cell
- Monocyte
- Pneumocyte
- Red blood cell
- T cell and NK cell
- unclassified



matchScore2: the lung atlas

OLD

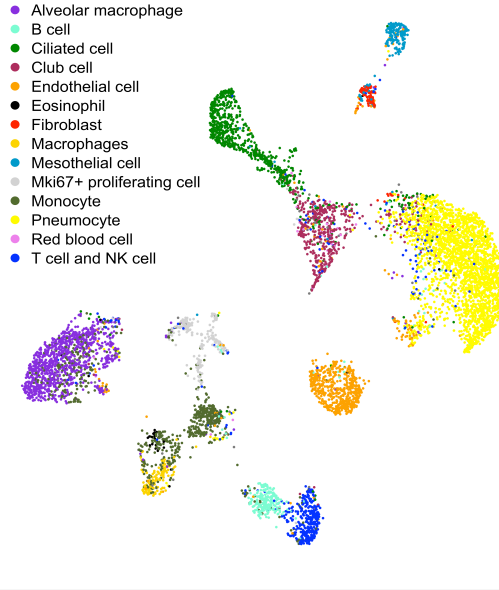


Reference cell types

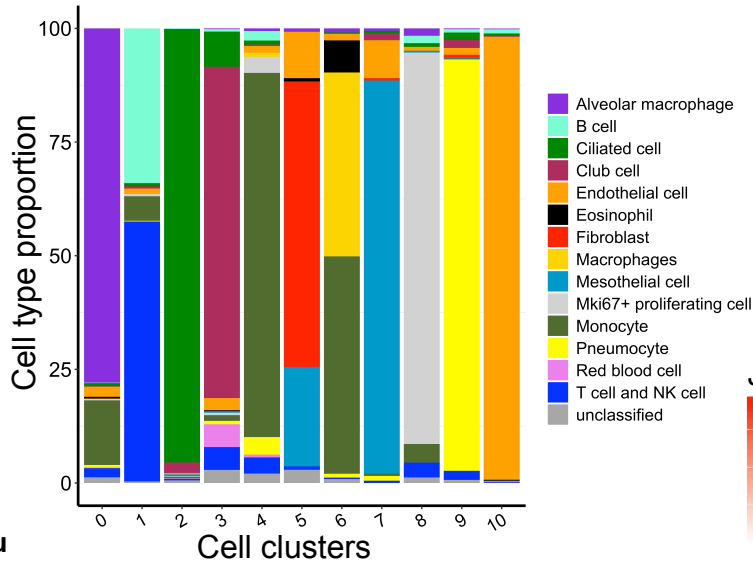
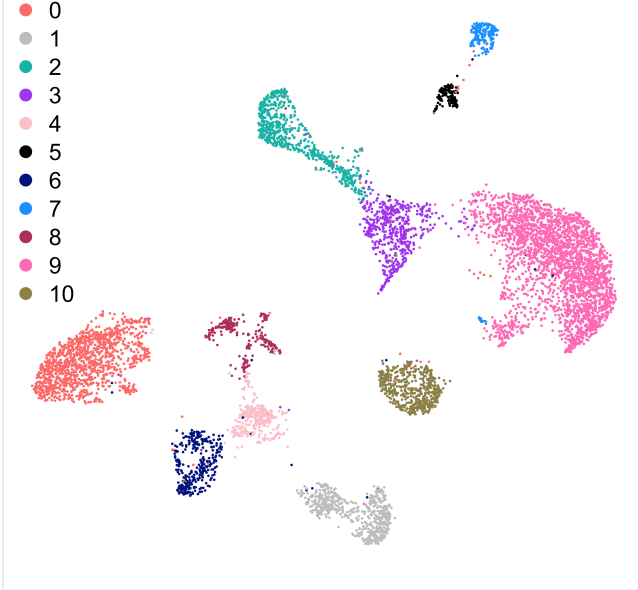
- Alveolar macrophage
- B cell
- Ciliated cell
- Club cell
- Endothelial cell
- Eosinophil
- Fibroblast
- Macrophages
- Mesothelial cell
- Mki67+ proliferating cell
- Monocyte
- Pneumocyte
- Red blood cell
- T cell and NK cell

YOUNG

matchScore2



Clustering

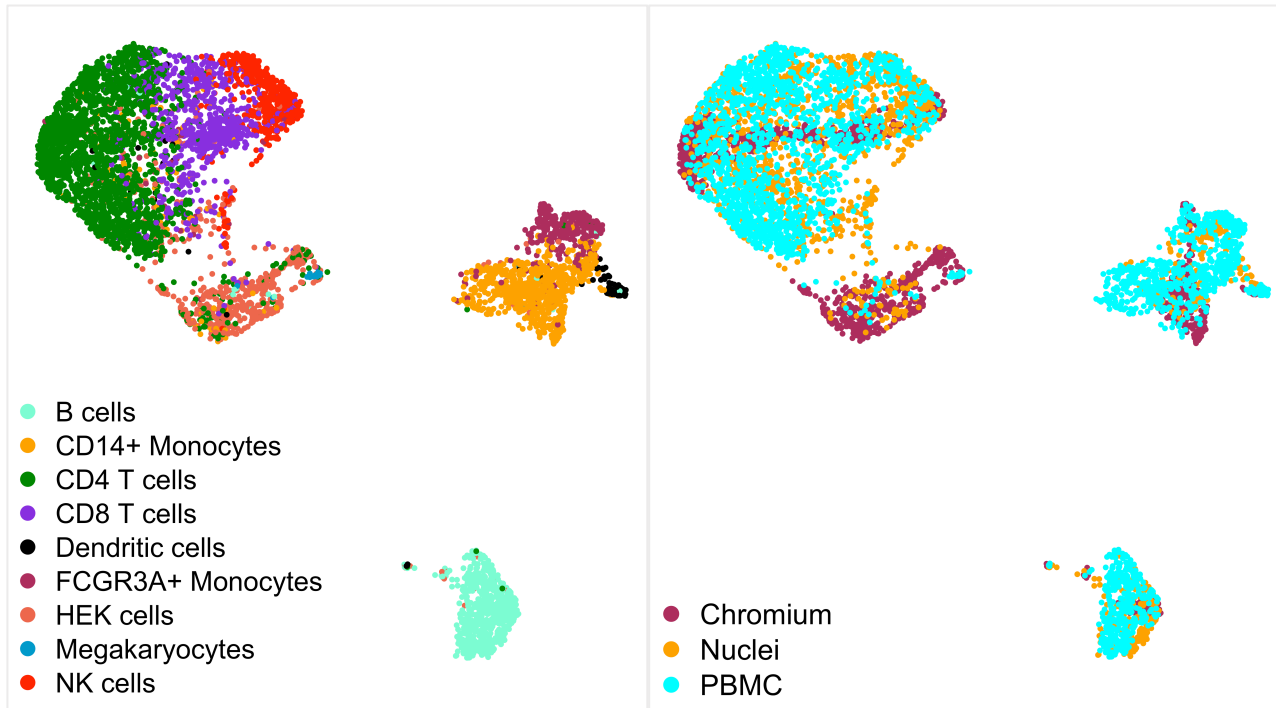


OLD (reference)	0	1	2	3	4	5	6	7	8	9	10
T cell and NK cell	0.01	0.46	0	0	0.09	0.01	0.02	0	0.01	0	0.01
Red blood cell	0	0.01	0	0.01	0.01	0	0.01	0	0	0	0
Pneumocyte	0	0	0.01	0.03	0	0	0	0	0	0	0.75
Monocyte	0.1	0.04	0	0	0.17	0.01	0.26	0.02	0	0	0.01
Mki67+ proliferating cell	0	0	0	0	0	0	0	0	0.65	0	0
Mesothelial cell	0	0	0	0.02	0.01	0.23	0.01	0.56	0	0	0.02
Macrophages	0.07	0	0	0	0.1	0.01	0.31	0.01	0	0	0
Fibroblast	0	0	0.01	0.02	0	0.48	0.01	0.16	0	0	0.01
Eosinophil	0.1	0.02	0	0.01	0.06	0.01	0.1	0.01	0.01	0.01	0.01
Endothelial cell	0.01	0	0	0.01	0.01	0.01	0.01	0.02	0	0.01	0.64
Club cell	0	0	0.02	0.58	0	0.02	0.01	0.02	0.01	0.03	0
Ciliated cell	0	0	0.68	0.03	0	0.01	0	0	0	0.01	0
B cell	0	0.27	0	0	0.1	0	0.04	0.01	0.01	0	0.01
Alveolar macrophage	0.5	0.01	0	0	0.03	0	0.07	0	0	0.01	0

JI

0.0 0.2 0.4 0.6

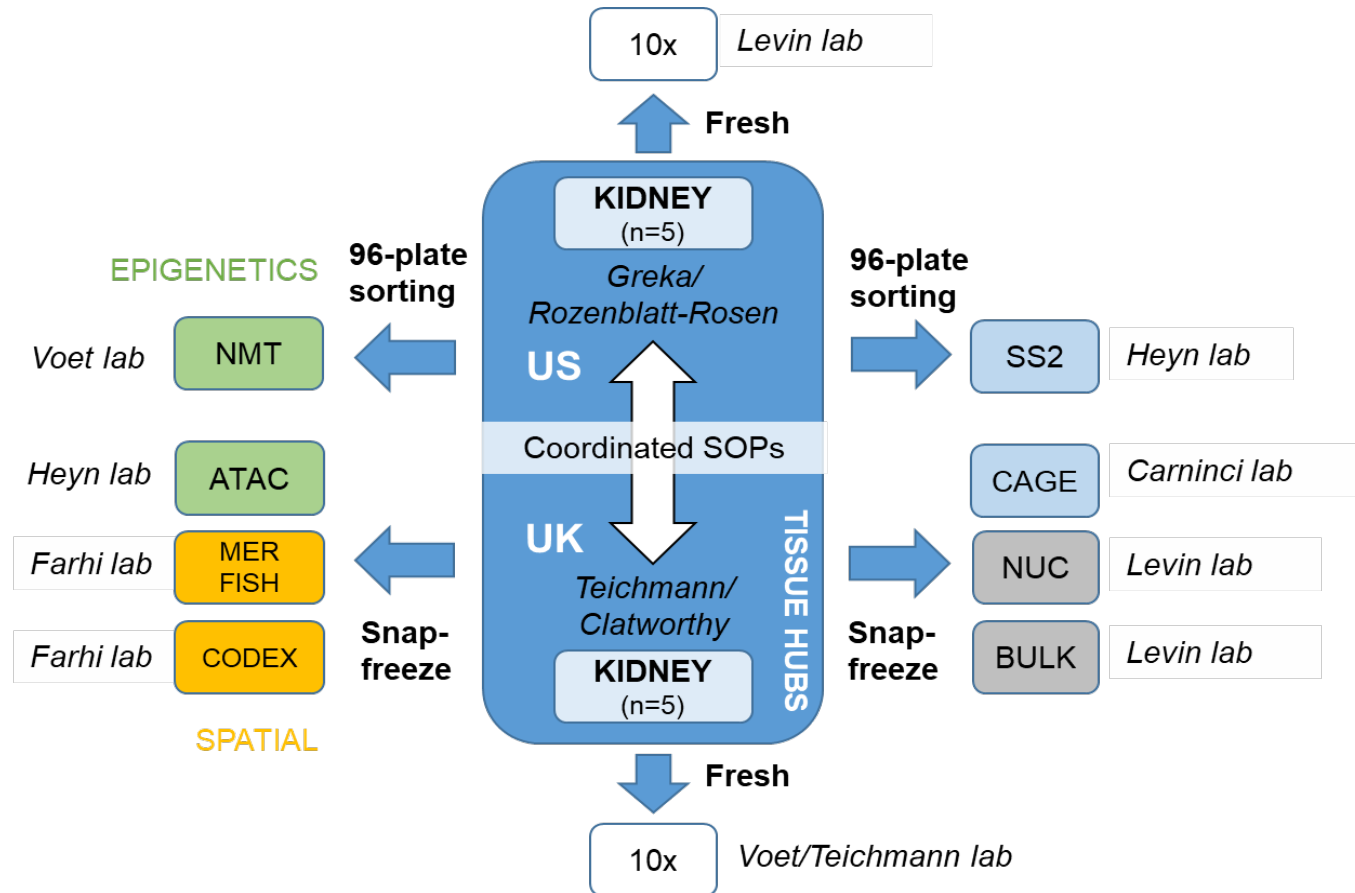




- matchScore2 combines datasets by using a SVD decomposition.
- The datasets are projected into a new common space of coordinates.
- This type of integration allows a direct comparison across cell types that are under different conditions (genotypes, treatments, diseased).



Developing *tools and standards* for the integration of **multimodal** HCA data in order to evaluate *performance, complementarity and replicability* of methods.



Acknowledgements

Single Cell Genomics Team

Atefeh Lafzi
Catia Moutinho
Holger Heyn

HCA consortium

Christoph Ziegenhain	Sascha Sauer
Davis J. McCarthy	Timo Trefzer
Adrian Alvarez	Christian Conrad
Eduard Batlle	Xian Adiconis
Sagar	Lan T. Nguyen
Dominic Grün	Aviv Regev
Julia K. Lau	Joshua Z. Levin
Stéphane C. Boutet	Swati Parekh
Chad Sanada	Aleksandar Janjic
Aik Ooi	Lucas E. Wange
Robert C. Jones	Johannes W. Bagnoli
Kelly Kaihara	Wolfgang Enard
Chris Brampton	Marta Gut
Yasha Talaga	Rickard Sandberg
Yohei Sasagawa	Itoshi Nikaido
Kaori Tanaka	Ivo Gut
Tetsutaro Hayashi	Oliver Stegle
Cornelius Fischer	



Thank you for your attention!

cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico



HUMAN
CELL
ATLAS



CRG[®]
Centre
for Genomic
Regulation