

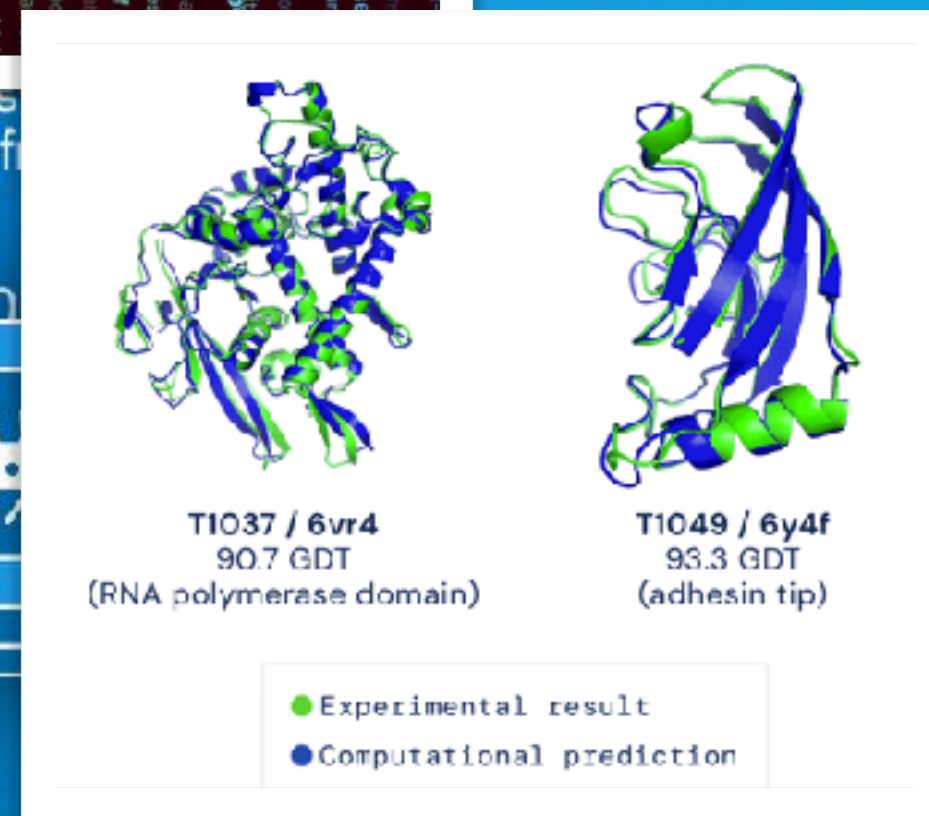
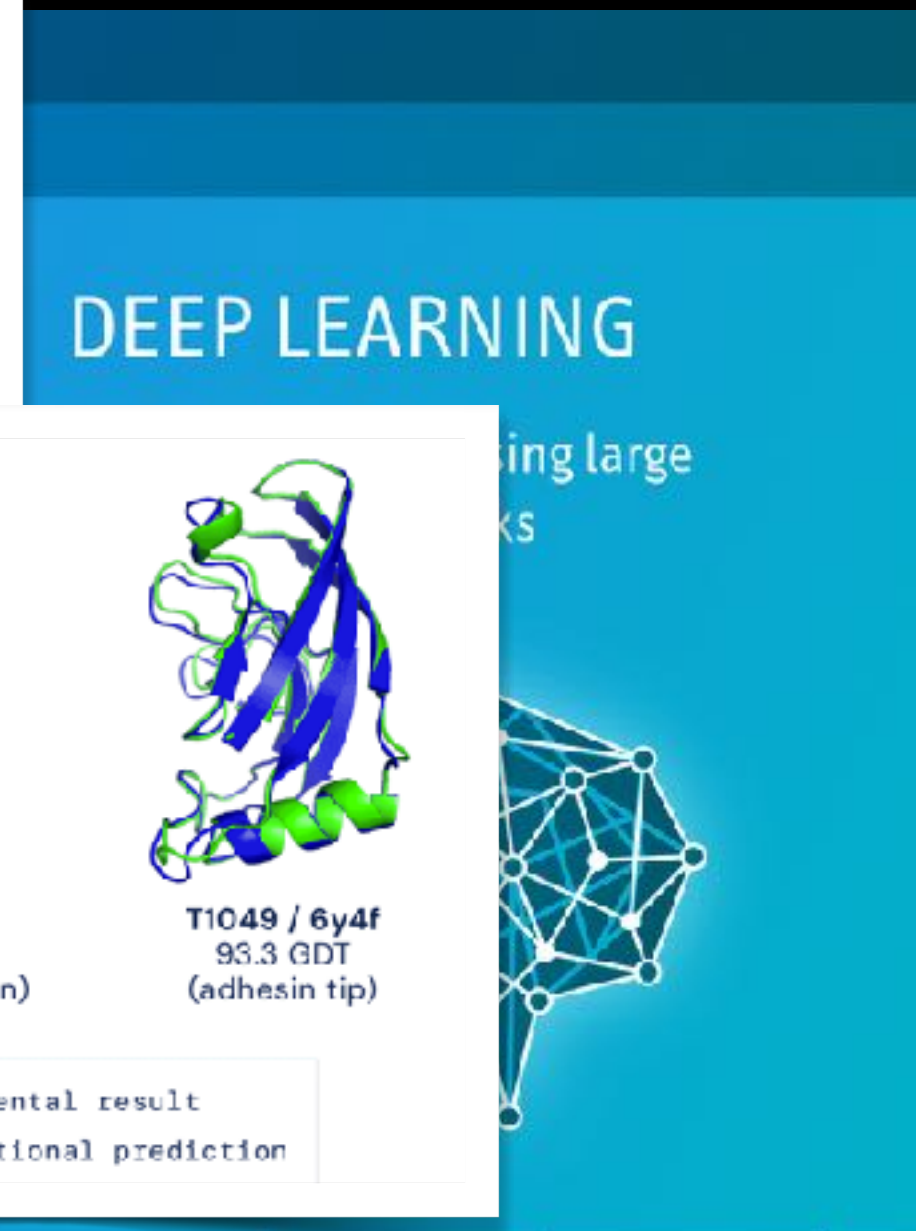
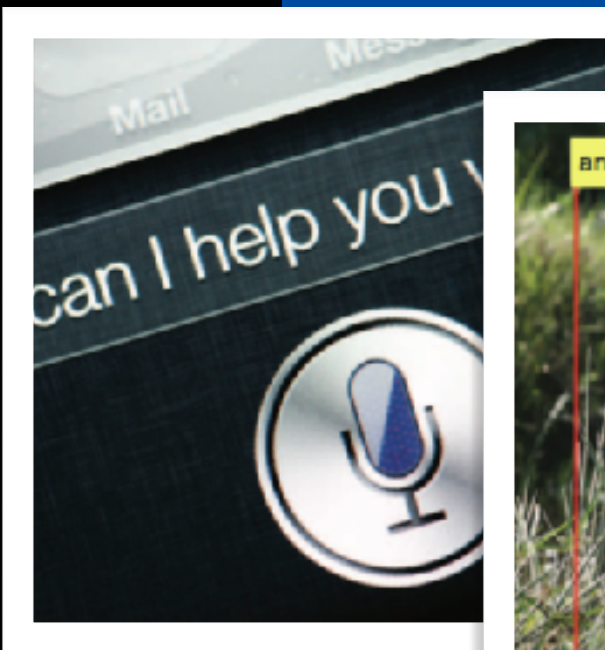
Artificial intelligence in biomedicine: how we enable a cellular view of human health with big data

Fabian Theis
Helmholtz Munich & TU Munich & Wellcome Sanger Institute

www.comp.bio



[@fabian_theis](https://twitter.com/fabian_theis)



1950

1960

1970

1980

1990

2000

2010

2020

BIG DATA

HELMHOLTZ

RESEARCH FOR GRAND CHALLENGES

ENERGY



EARTH AND
ENVIRONMENT



HEALTH



AERONAUTICS,
SPACE AND
TRANSPORT



MATTER



INFORMATION



BIG DATA

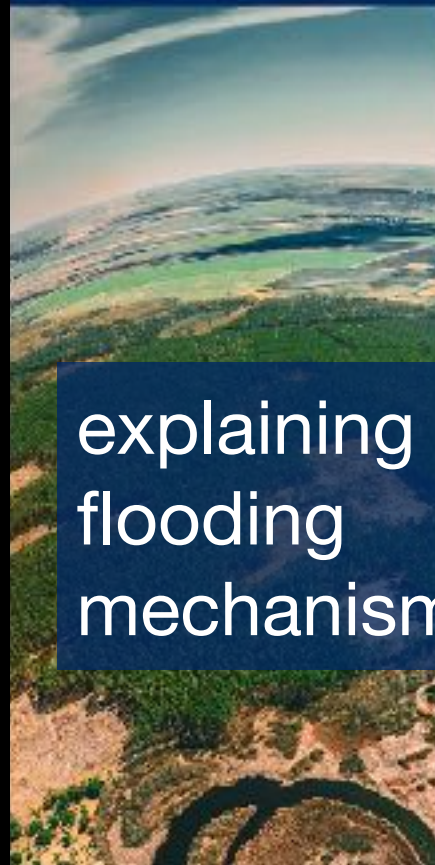
HELMHOLTZAI

ENERGY



detecting
thermal
bridges

EARTH AND
ENVIRONMENT



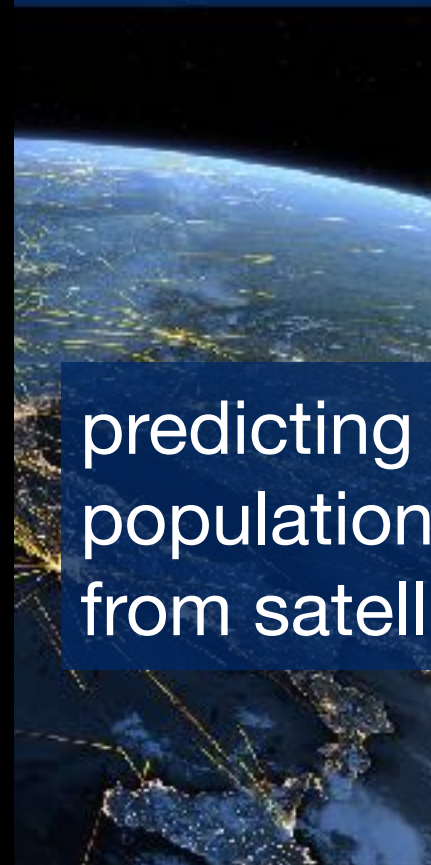
explaining
flooding
mechanisms

HEALTH



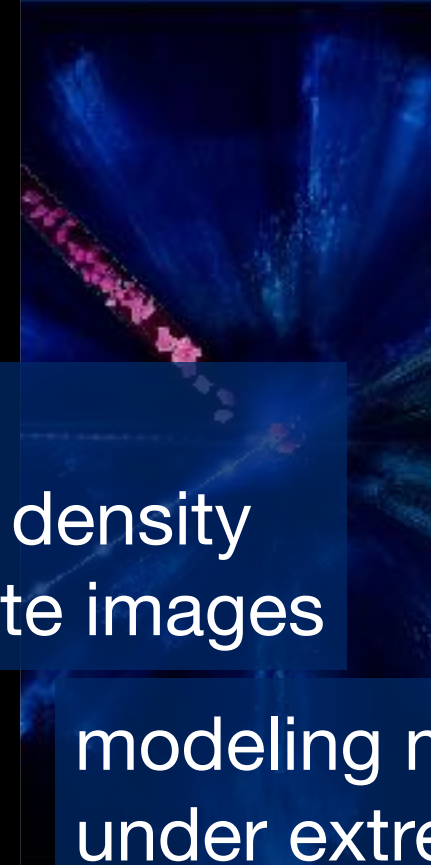
learning
cellular
atlases

AERONAUTICS,
SPACE AND
TRANSPORT



predicting
population density
from satellite images

MATTER



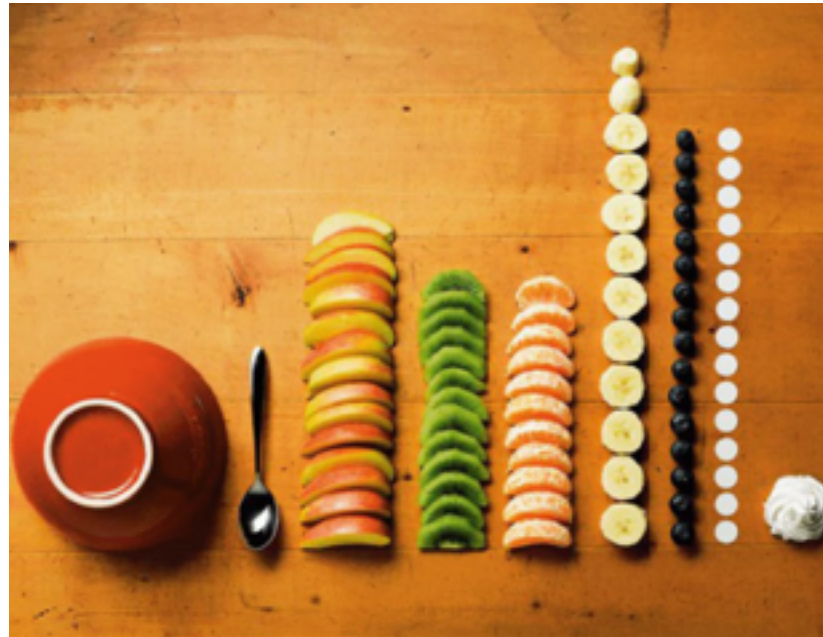
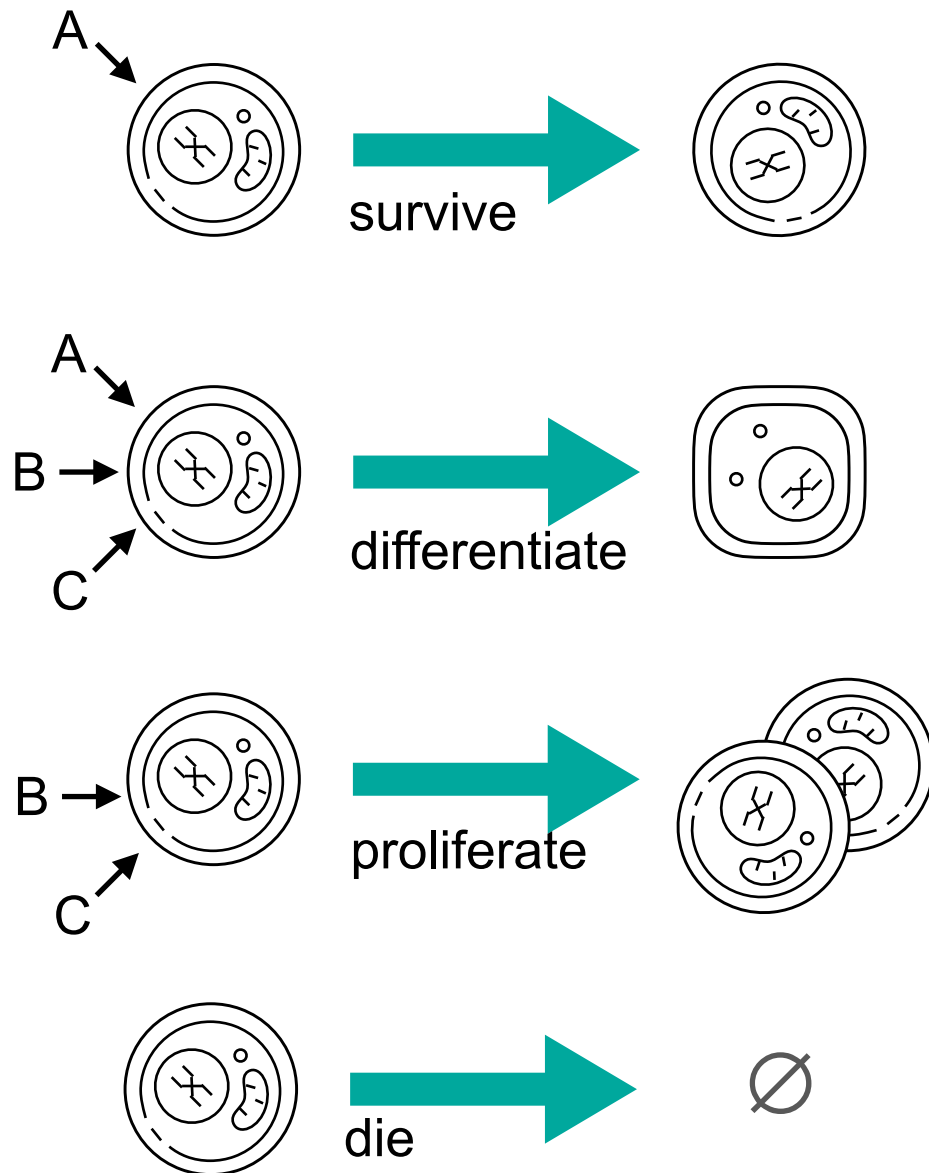
modeling matter
under extreme
conditions

INFORMATION



towards
energy-
efficient AI

Understand the cellular underpinning of health & disease



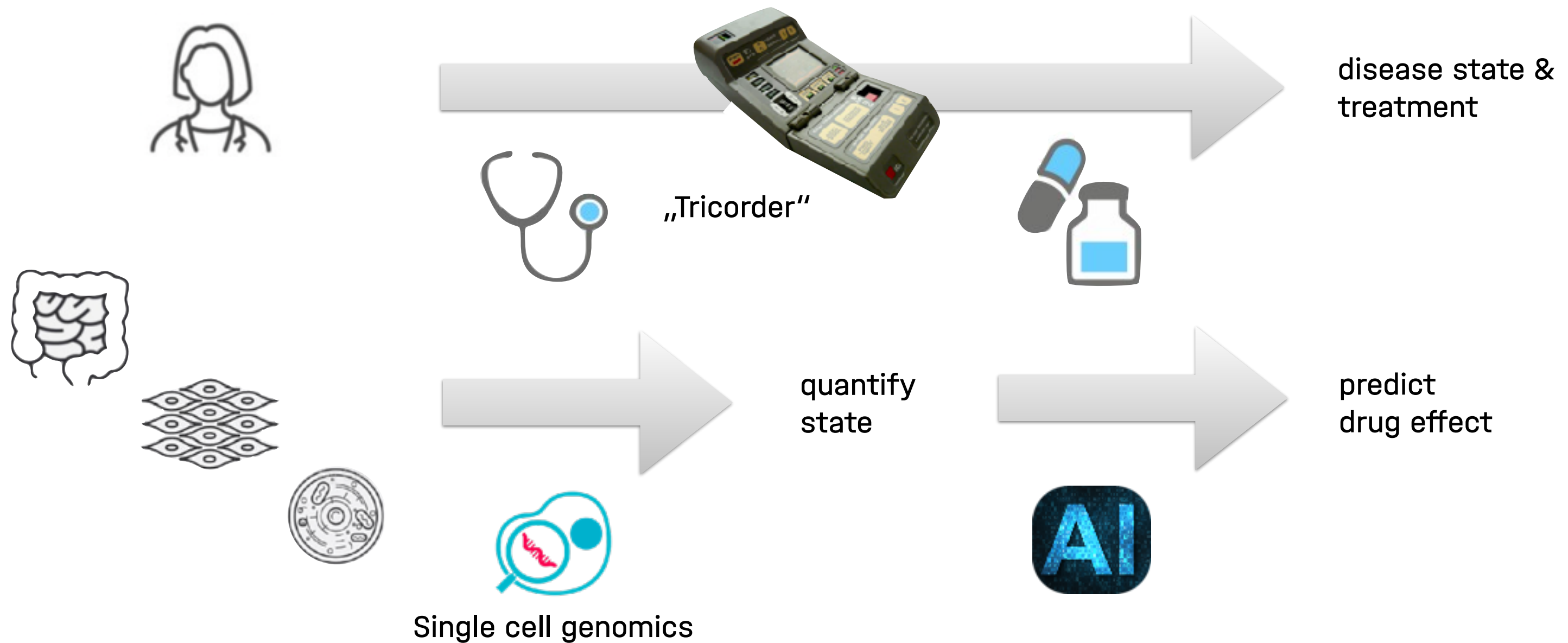
single-cell genomics

$$X_{new} = f(X_{old}, p)$$

model state change using
machine learning

Can we understand a system if
we can predict its behavior?

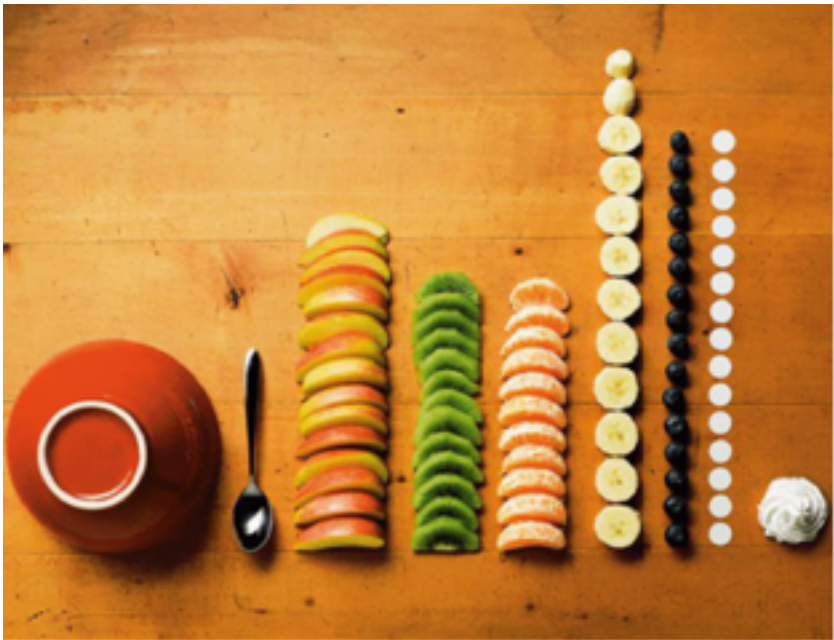
Vision - learn optimal perturbations



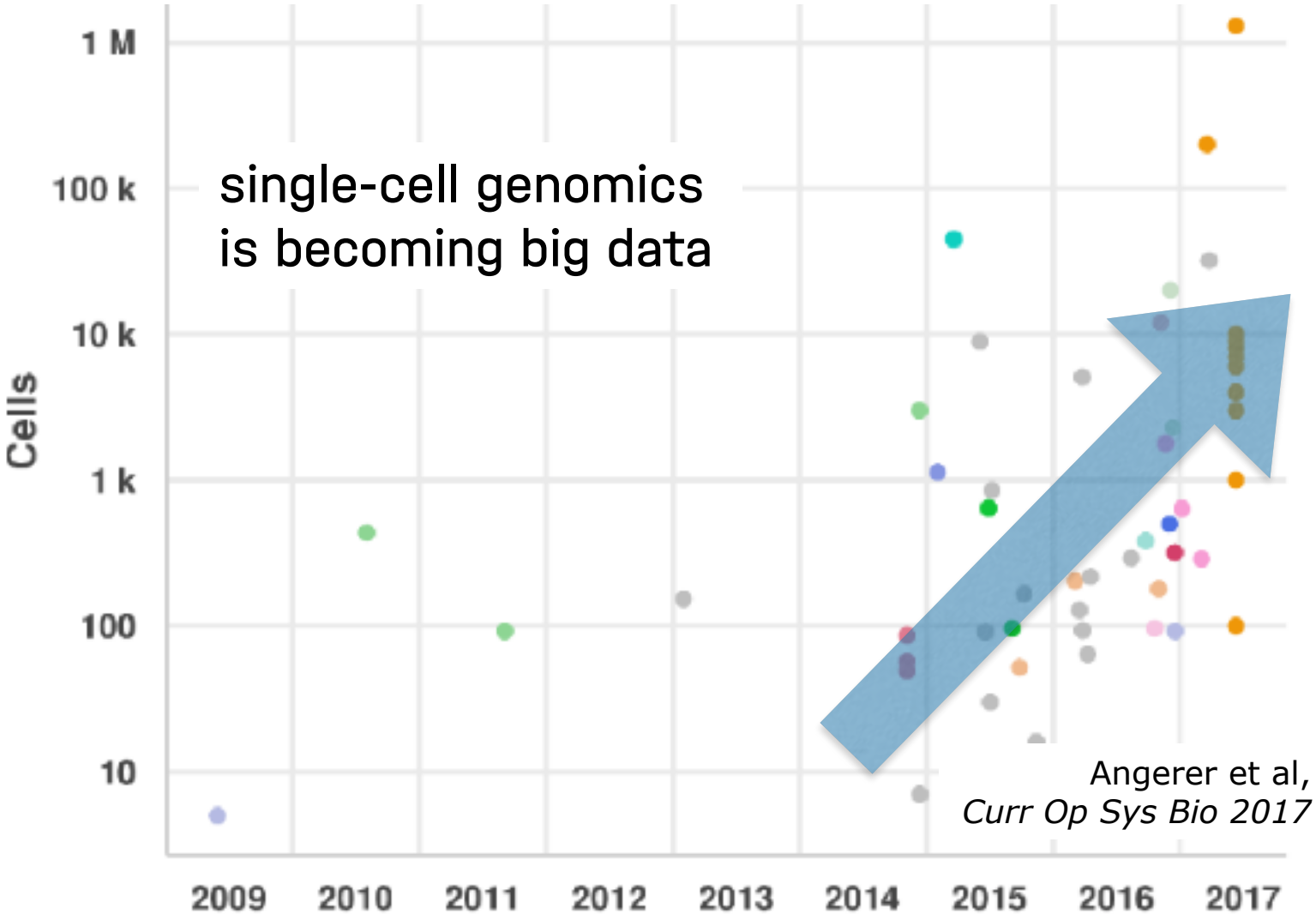
Unbiased description of cellular state



bulk genomics



single-cell genomics



gel beads

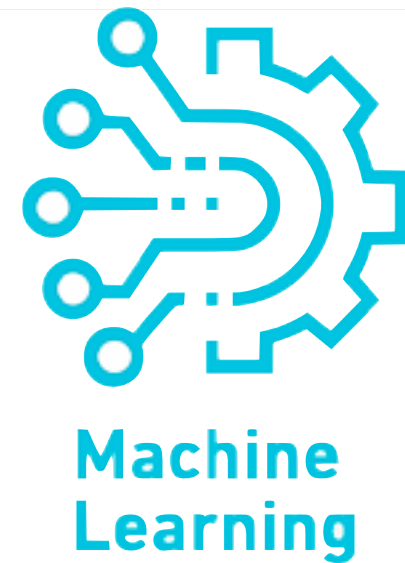
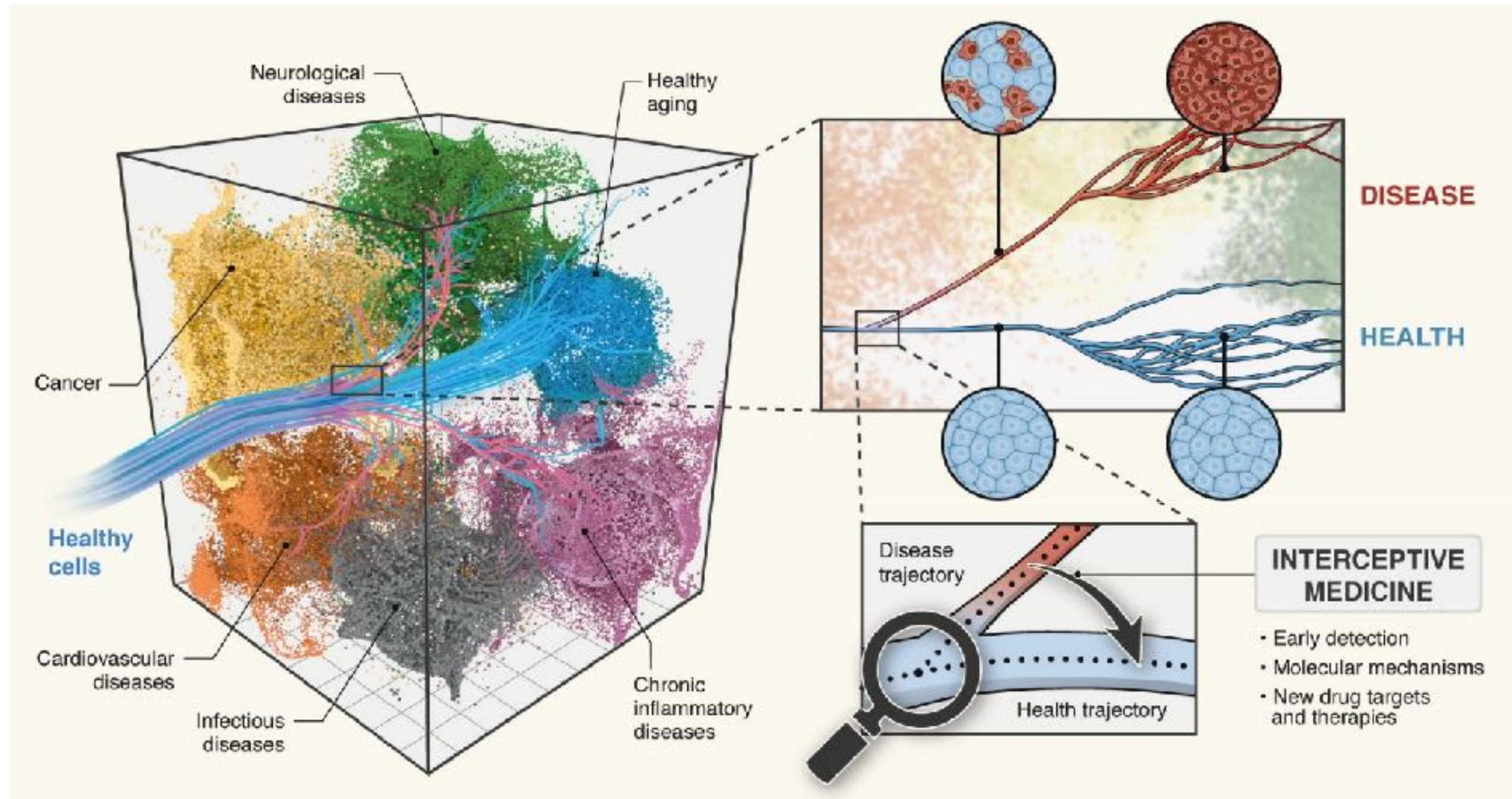


single cells



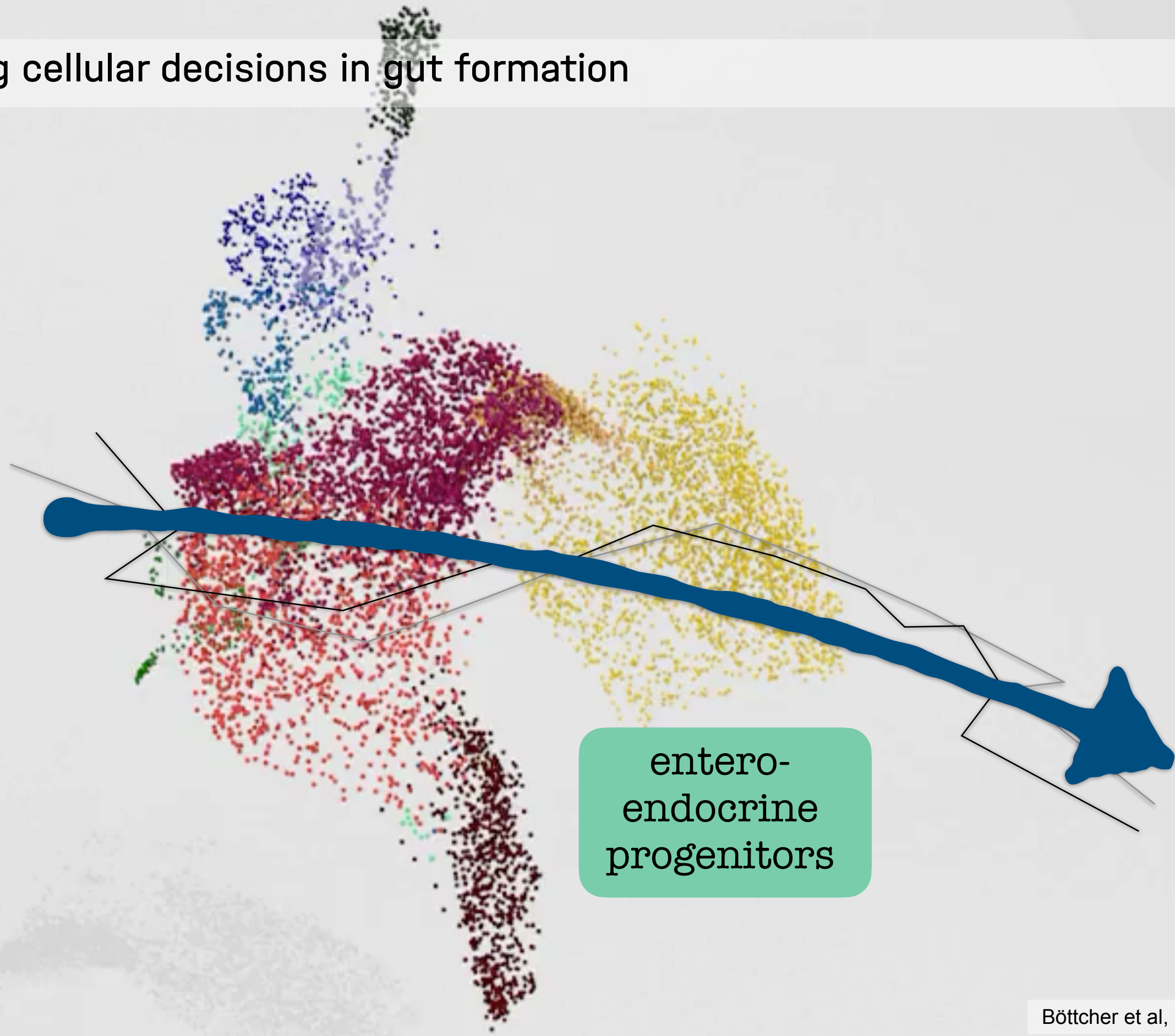
oil

Single cell analysis for understanding cell fate in health & disease



Example: learning cellular decisions in gut formation

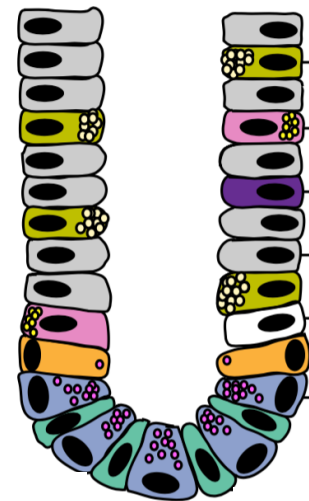
intestinal
stem cells



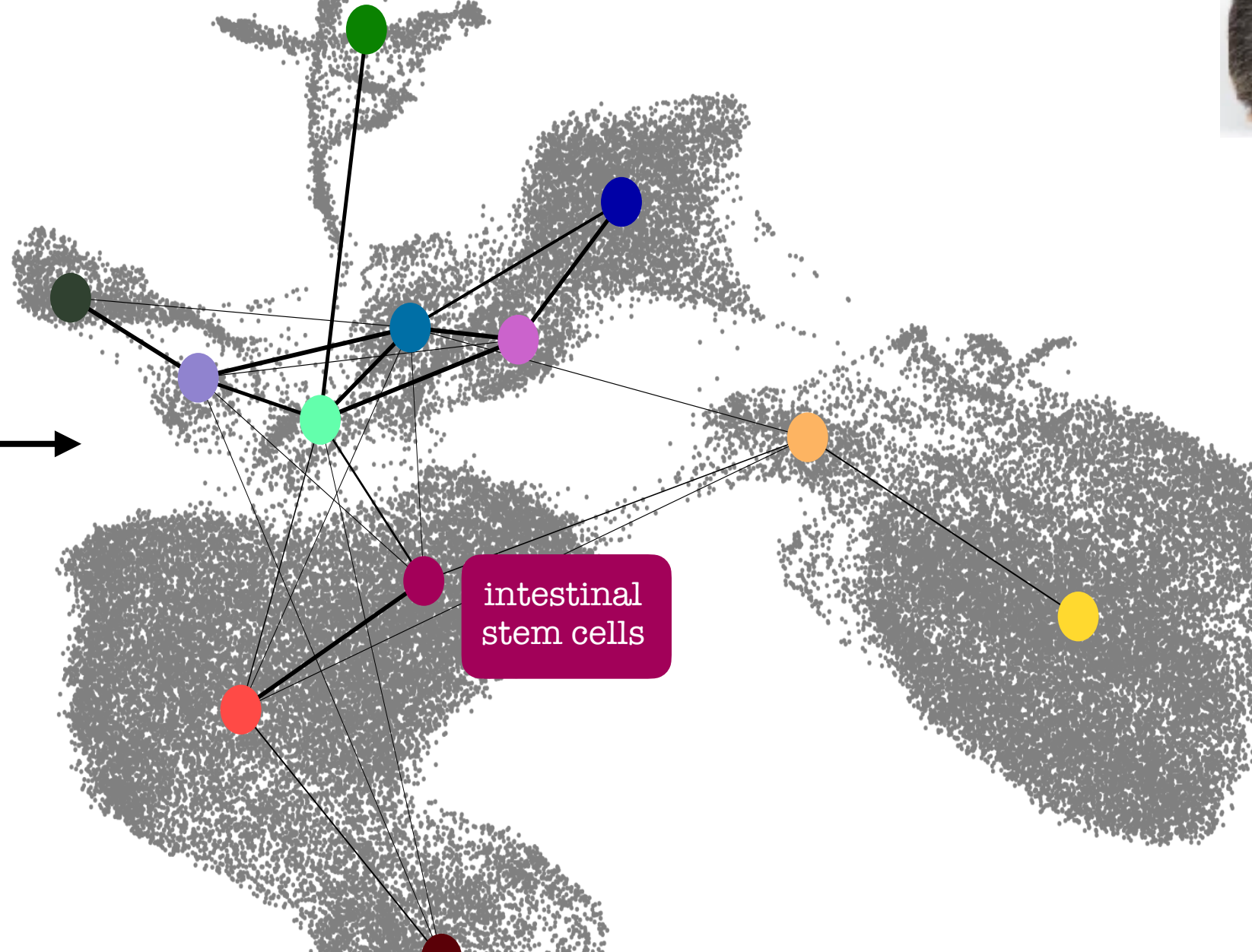
entero-
endocrine
progenitors

entero-
endocrine
cells

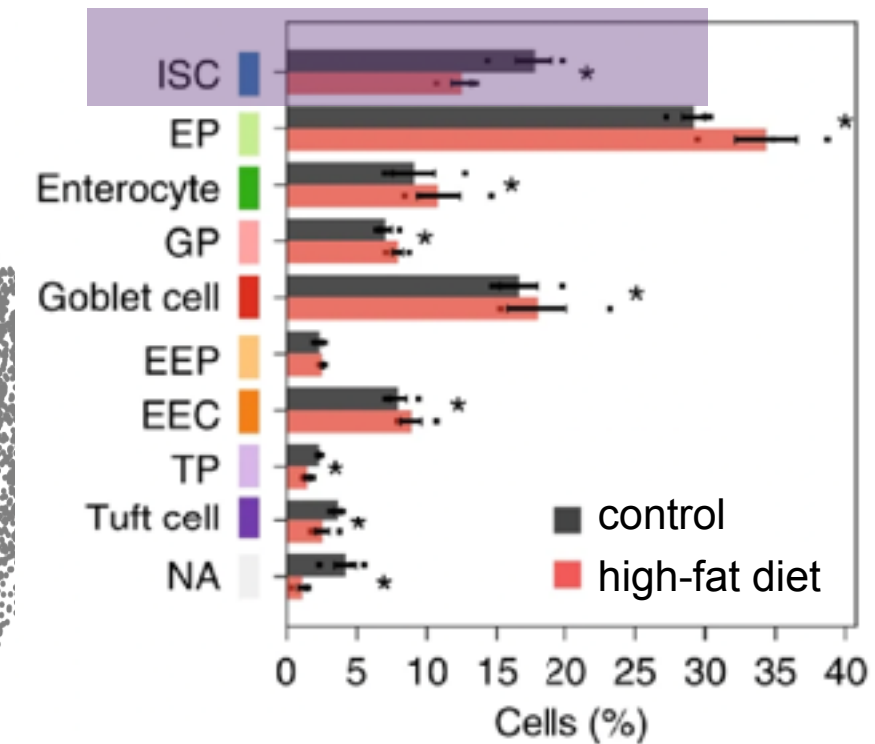
Application: intestinal lineage formation in diabetes



scRNA-seq



less stem cells in high-fat diet



Observation: Diseases such as diabetes alters stem cell decision making
Treatment aim: Manipulate this by adequate drugs

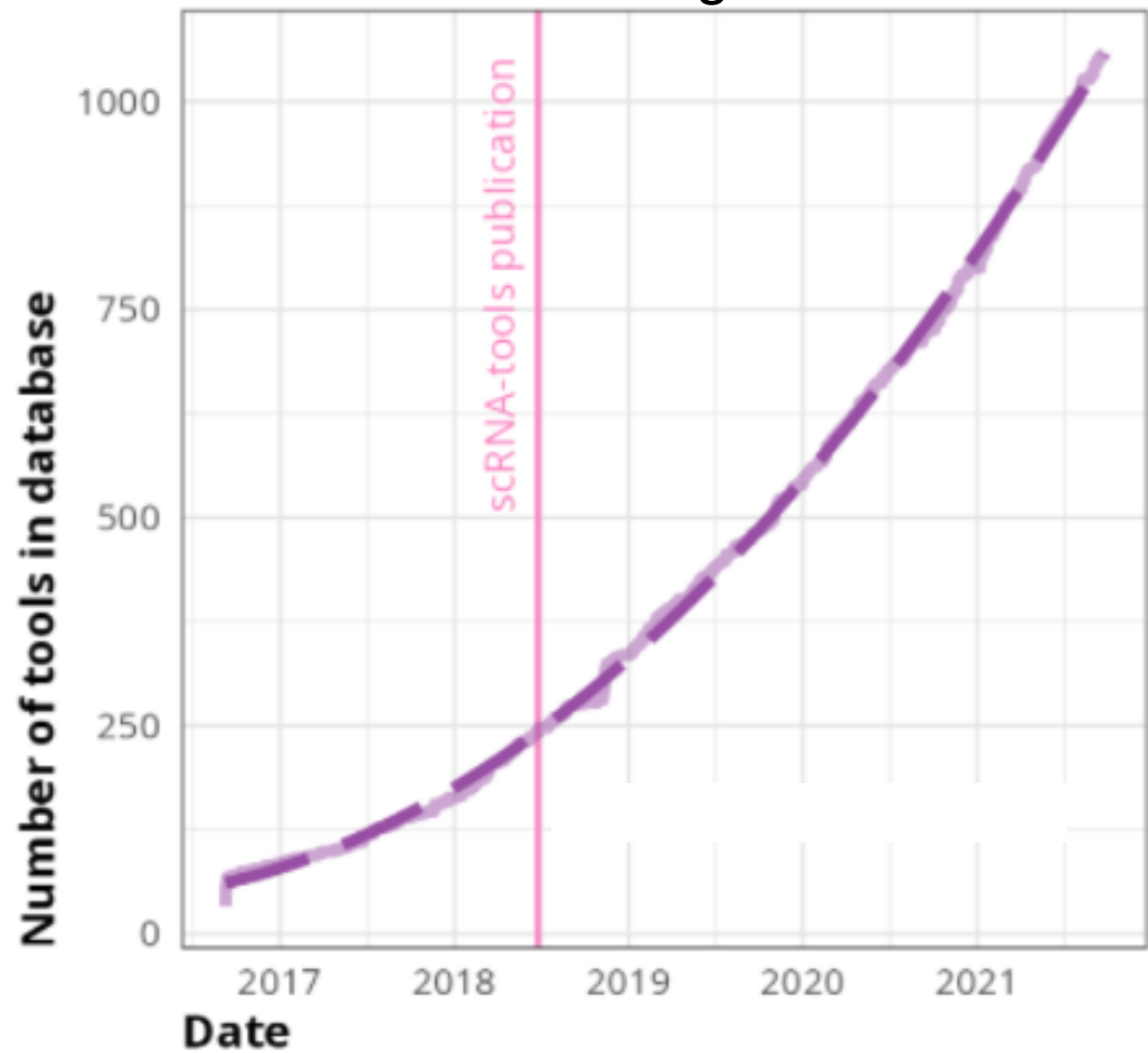
Software matters: computational tools in single cell genomics



Luke Zappia

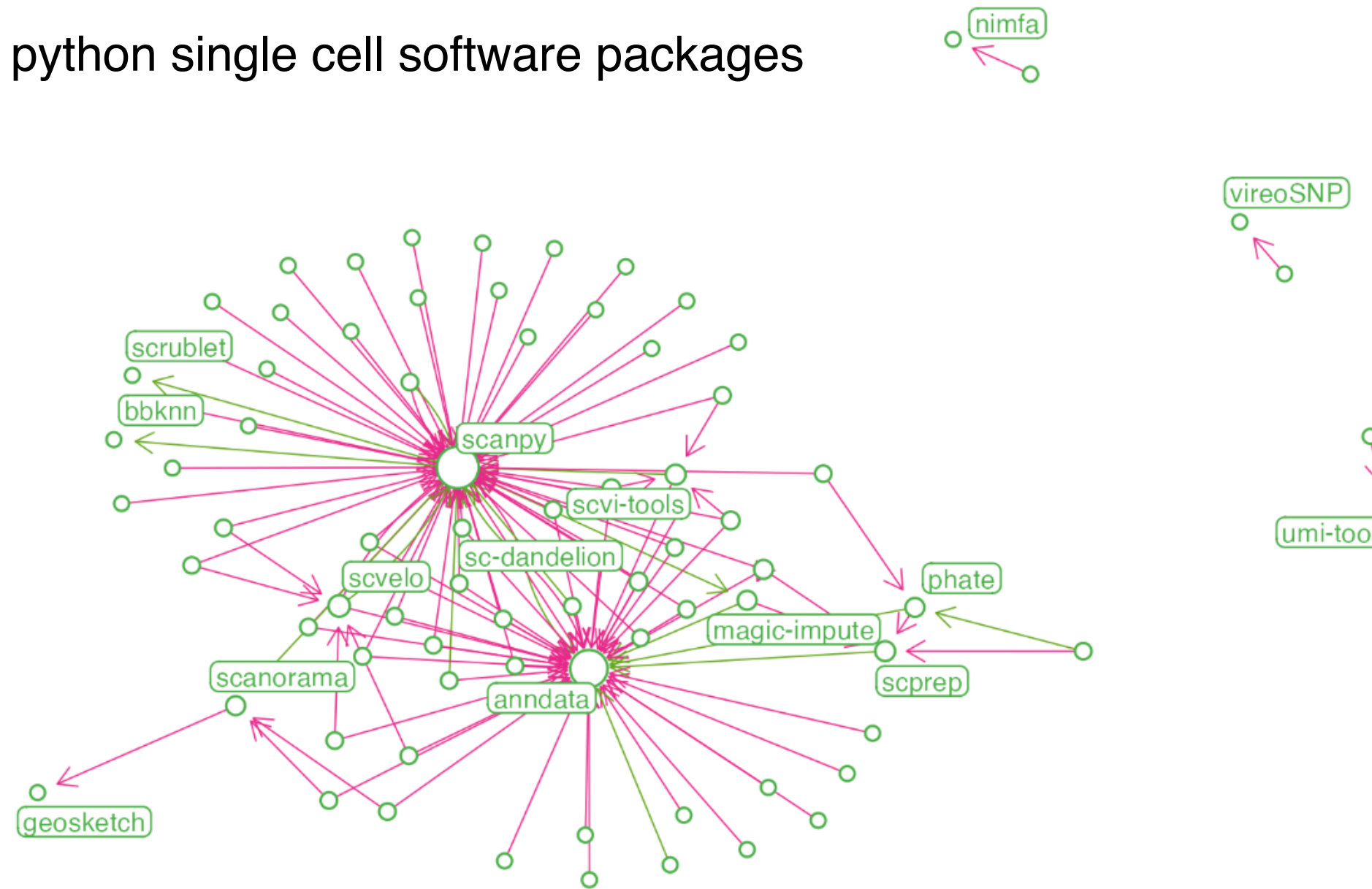
trend: increasing machine learning questions asked to single cell data

www.scrna-tools.org



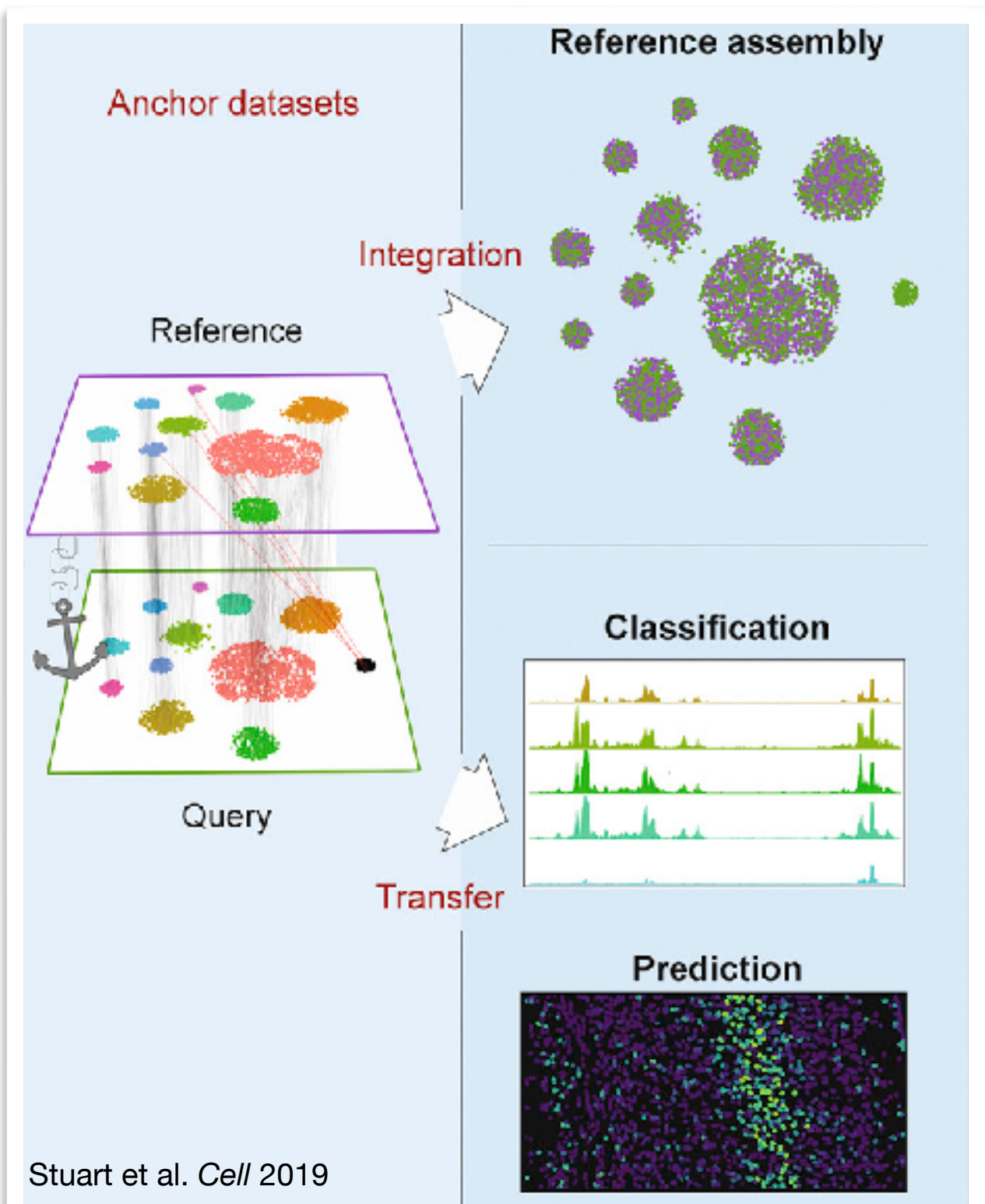
HELMHOLTZ MUNICH

python single cell software packages



Building a cell atlas?

→ Learning an integrated cell atlas



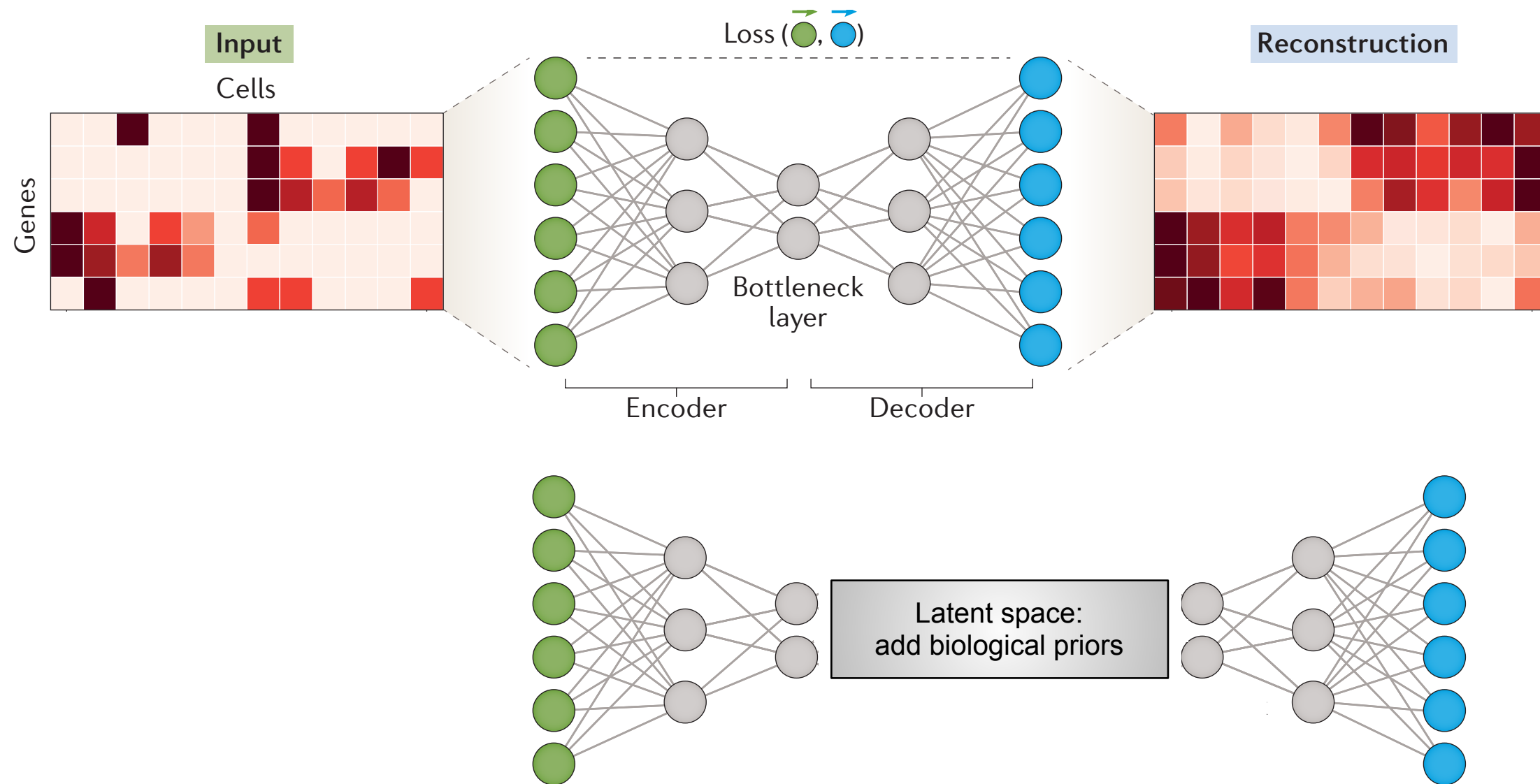
atlas would be a **collection of cellular reference maps**,
h of the thousands of cell types in the human body and
nd.

Benchmarking single cell data integration

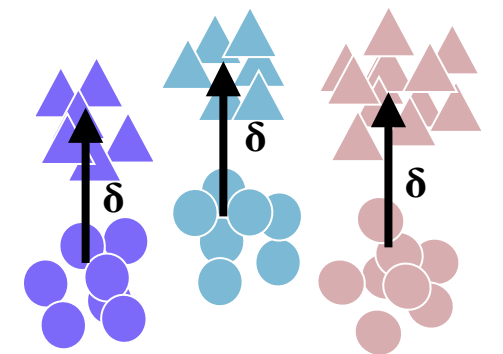
	Output	Features	Scaling	Overall Sc	Batch Con	PCR batch	Batch ASV	graph iLIs	graph com	Bio conservation	NMI cluster/label	ARI cluster/label	Cell type ASW	isolated label F1	isolated label silhouette	graph cLISI	CC conservation	HVG conservation
Scanvi		HVG	-															
Fastmnn		HVG	-															
Scanvi		FULL	-															
BBKNN		HVG	-															
Seuratpca		HVG	-															
Scanorama		HVG	+															
BBKNN		FULL	-															
Fastmnn		HVG	-															
ComBat		FULL	-															
ComBat		HVG	-															
Seuratpca		FULL	-															
scVI		HVG	-															
scVI		FULL	-															
Scanorama		HVG	-															
ComBat		HVG	+															
Scanorama		FULL	-															
Conos		HVG	-															
Harmony		HVG	+															
ComBat		FULL	+															

Luecken et al. Nat Methods 2022

Deep learning for latent space modeling & integration



Eraslan et al, *Nat Rev Gen* 2019



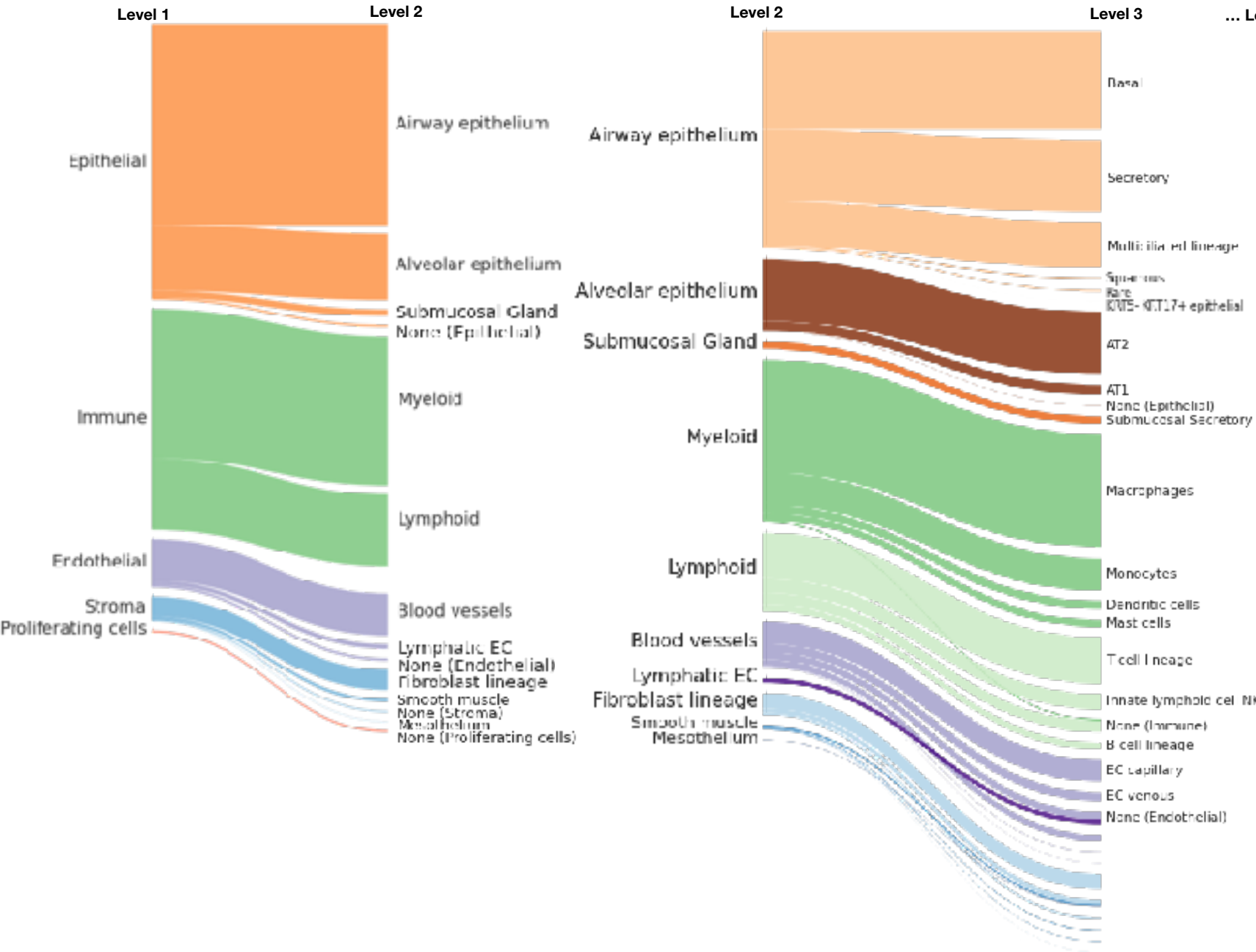
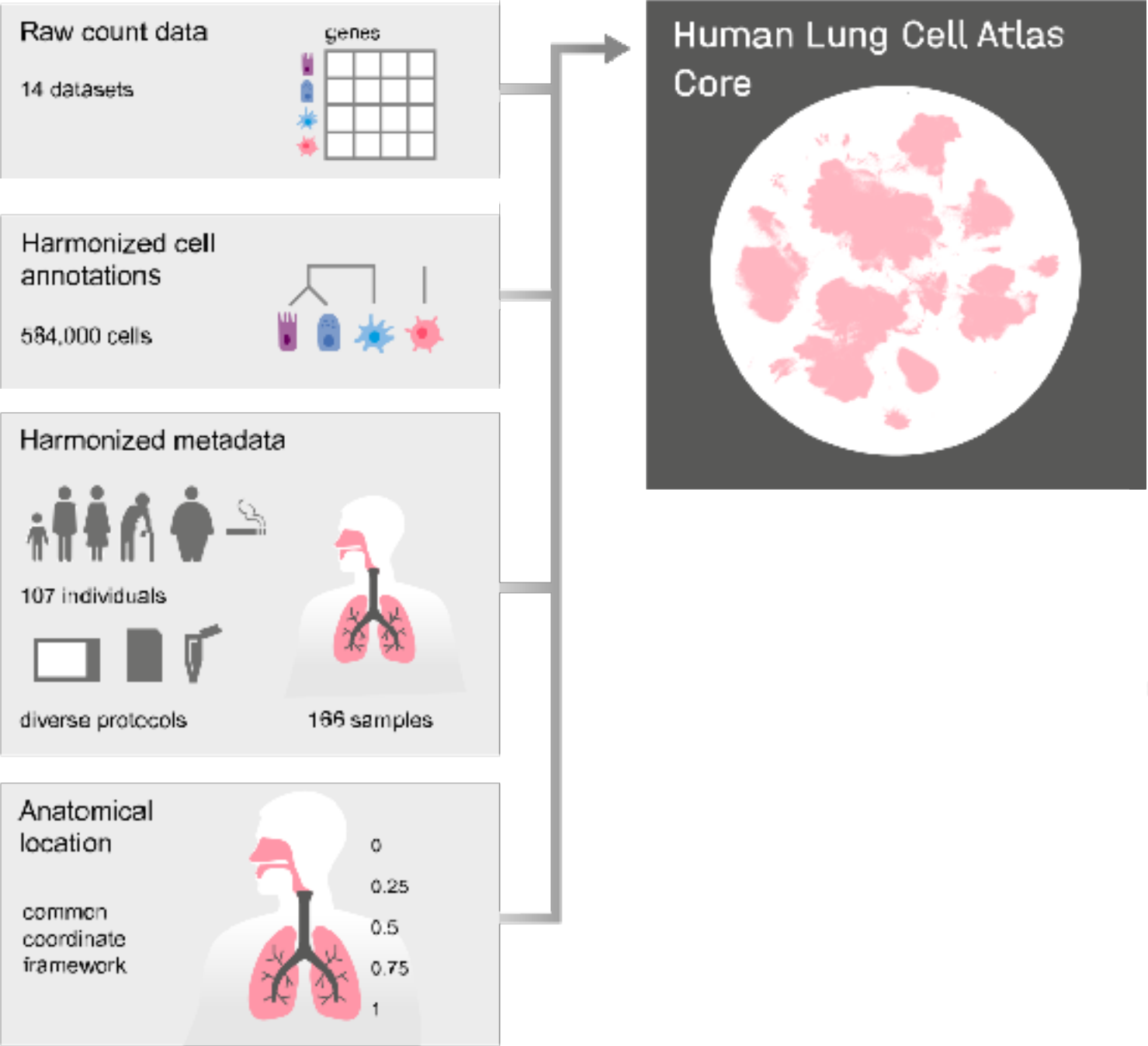
example: linear perturbation effects
Lotfollahi et al, *Nat Methods* 2019

NeurIPS data integration competition (>280 scientists):
all winning models use neural networks

Building an integrated lung cell atlas



Lisa Sikkema Malte Lücken



Acknowledgements

Helmholtz (Germany):

Malte Lücken

Lisa Sikkema

Daniel Strobl

Luke Zappia

Ciro Ramirez -Suastegui

Thomas Walzthoeni

Leander Dony

Elmar Spiegel

David Fischer

Lukas Heumos

Carlo de Donno

Carlos Talavera-Lopez

Ignacio Ibarra

Mohammad Lotfollahi

Isaac Virshup

Herbert Schiller

Meshal Ansari

Christoph Mayr

Janine Schniering

Yuexin Chen

Ahmed Agami

Baharak H. Kashani

Tgen (US):

Nicholas Banovich

Austin Gutierrez

Lance Peter

Mei-i Chung

Linh Bui

Northwestern (US):

Alexander Misharin

Nikolay Markov

Manu Jain

Sanger (UK):

Kerstin Meyer

Sarah Teichmann

Elo Madissoon

Anna Wilbrey-Clark

Amanda Oliver

Ni Huang

Chuan Xu

CNRS (France):

Pascal Barbry

Sylvie Leroy

Laure-Emmanuelle Zaragosi

Marie-Jeanne Arguel

Christophe Bécavin

Antoine Collin

Univ. Bonn (Germany):

Joachim Schultze

Evgeny Chichelnitskiy

Univ. Hannover (Germany):

Christine Falk

Theodore Capellos

UMCG (Netherlands):

Martijn Nawijn

Marijn Berg

Leonie Appertoo

Aurore Gay

Tessa Kole

Maarten v.d. Berge

UCSD (US):

Elizabeth Duong

Dinh Diep

Blue Lake

Kun Zhang

Xin Sun

Vanderbilt (US):

Jonathan Kropski

Chase Taylor

Lorraine Ware

Stanford (US):

Mark Krasnow

Kyle Travaglini

Pittsburgh (US):

Robert Lafyatis

Mauricio Rojas

Tracy Tabib

Oliver Eickelberg

CommaSoft (Germany):

Mitch von Papen

LungMAP consortium

and Biorepository (US):

Gloria Pryhuber

Heidie Huyck

Cory Poole

UCL (UK):

Marko Nikolic

Yoshida Masahiro

Kaylee Worlock

Duke (US):

Purushothama Tata Rao

Preetish Murthy

MSKCC (US):

Dana Pe'er

NJH (US):

Max Seibold

Nathan Jackson

Yale (US):

Naftali Kaminski

Broad (US):

Orit Rozenblatt-Rosen

Arizona State (US):

Douglas Shepherd

Stanford (US):

Tushar Desai

Mount Sinai (US):

Alexander Tsankov

Garvan Institute (AUS):

Joseph Powell

SciLife Lab (SWE):

Christos Samakovlis

Joakim Lundeborg

Univ. Helsinki (FIN):

Peter Horvath

Côte d'azur (FRA):

Charles-Hugo Marquette

MGH (US):

Jayaraj Rajagopal

UNC (US):

Jim Hagood

UCSF (US):

Dean Sheppard

CCHMC (US):

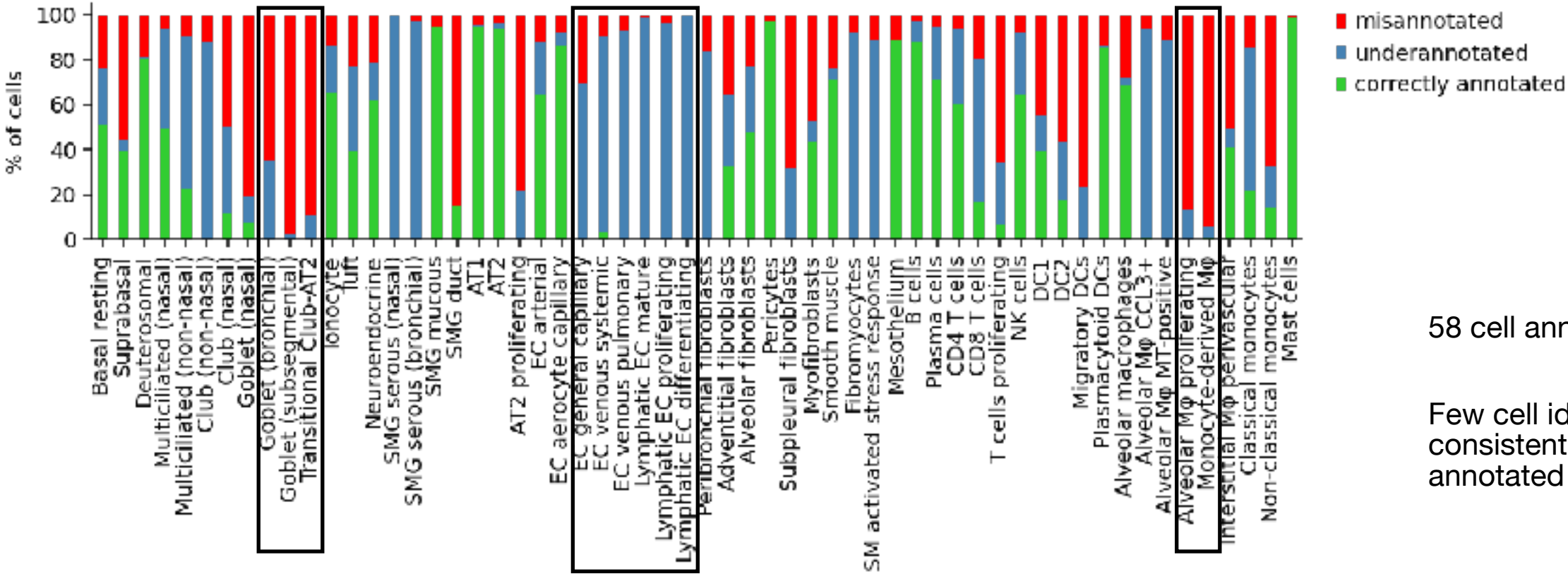
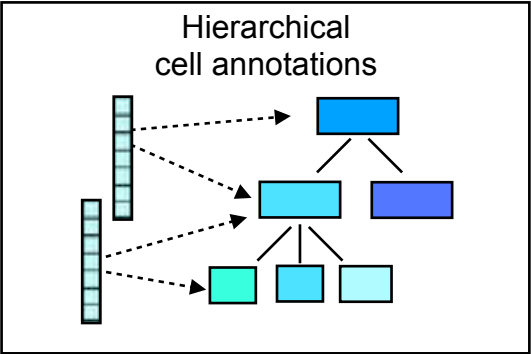
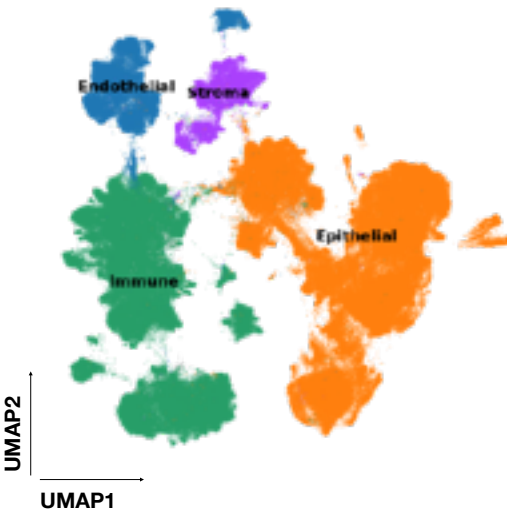
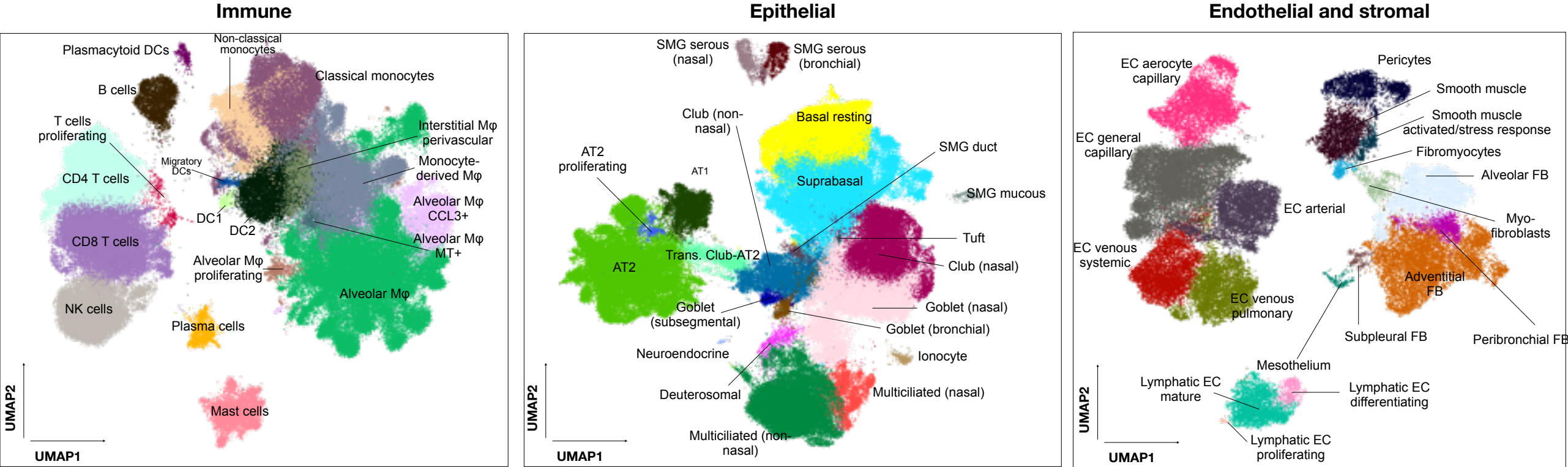
Yan Xu

Minzhe Guo

Jeffrey Whitsett



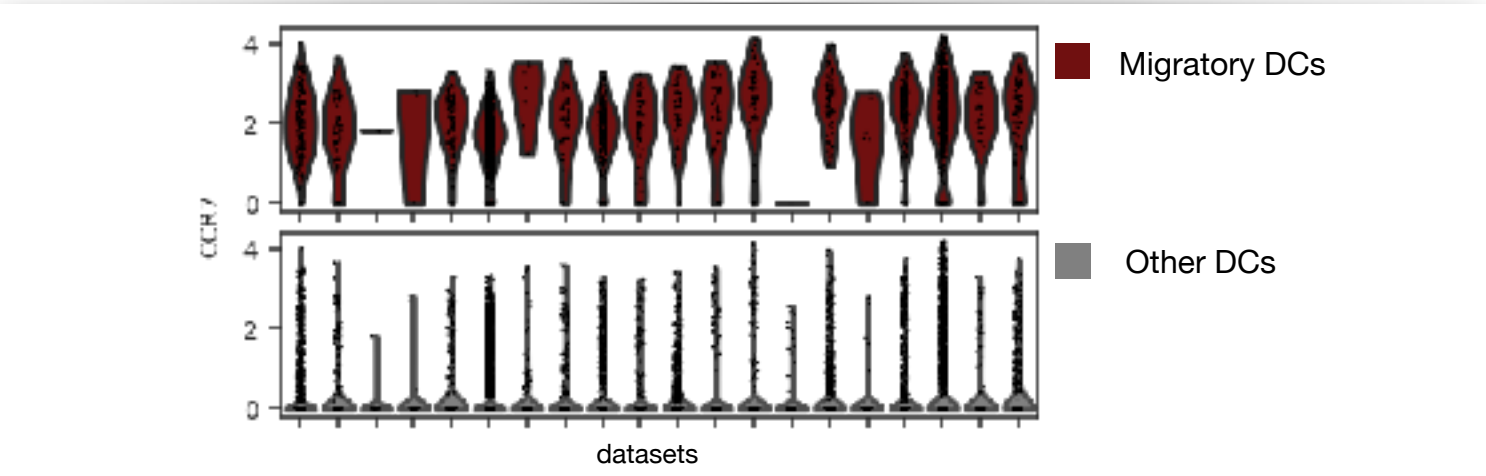
The integrated human lung cell atlas



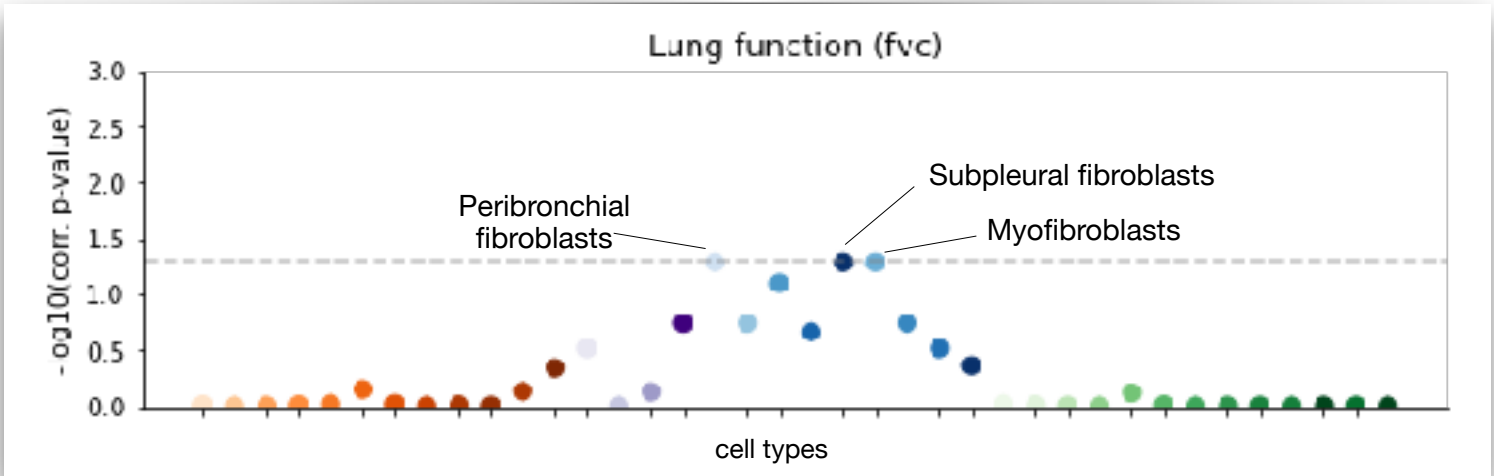
58 cell annotations

Few cell identities are consistently correctly annotated

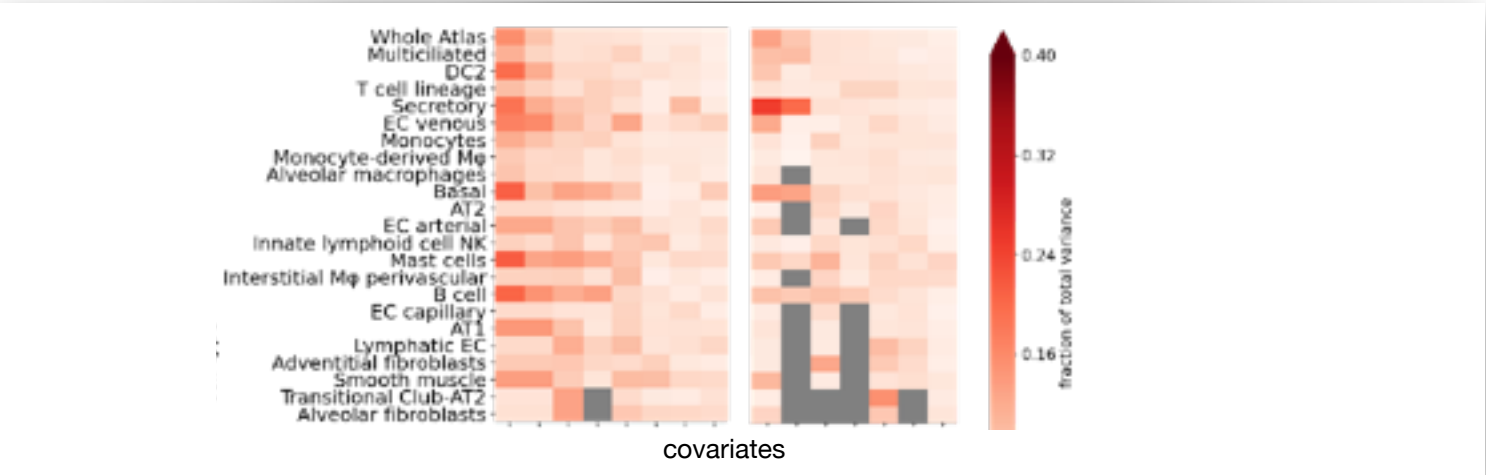
The integrated reference atlas enables ...



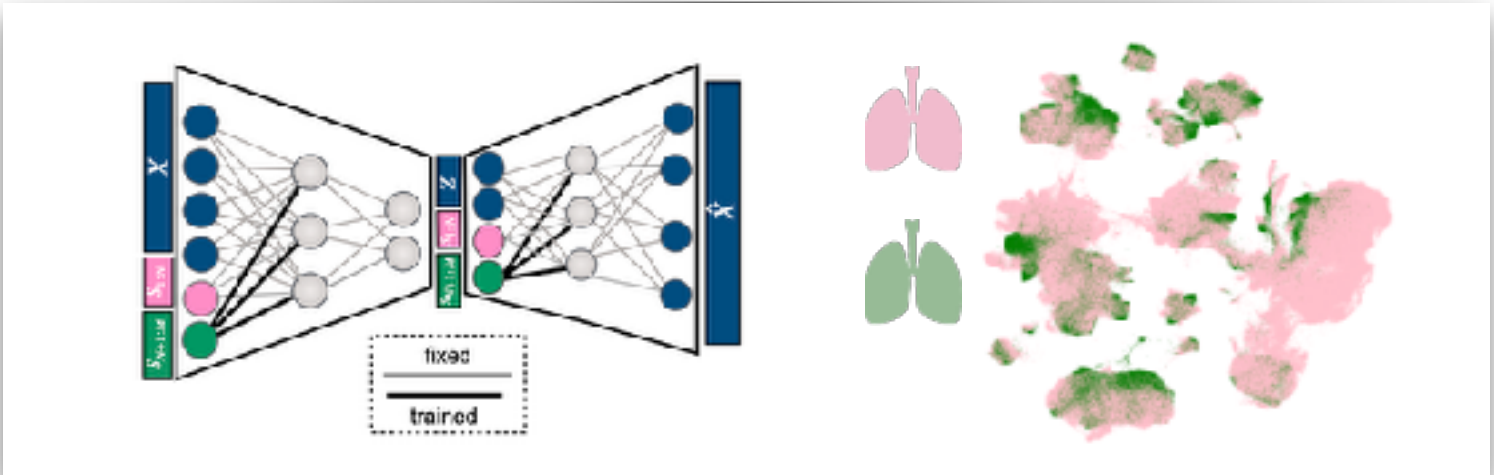
Recovery of rare cell identities



variant-cell type association

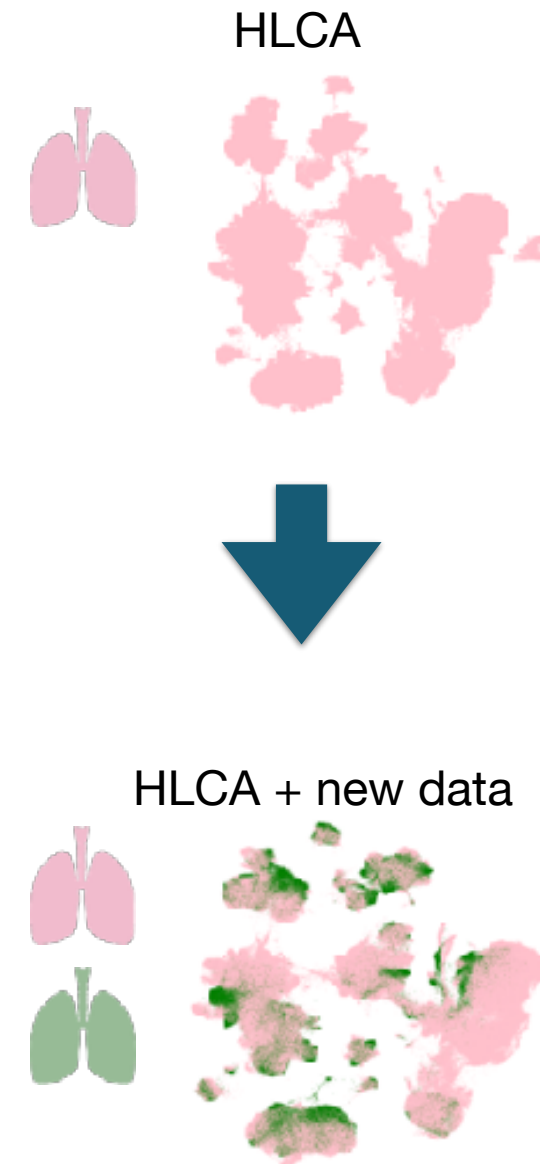
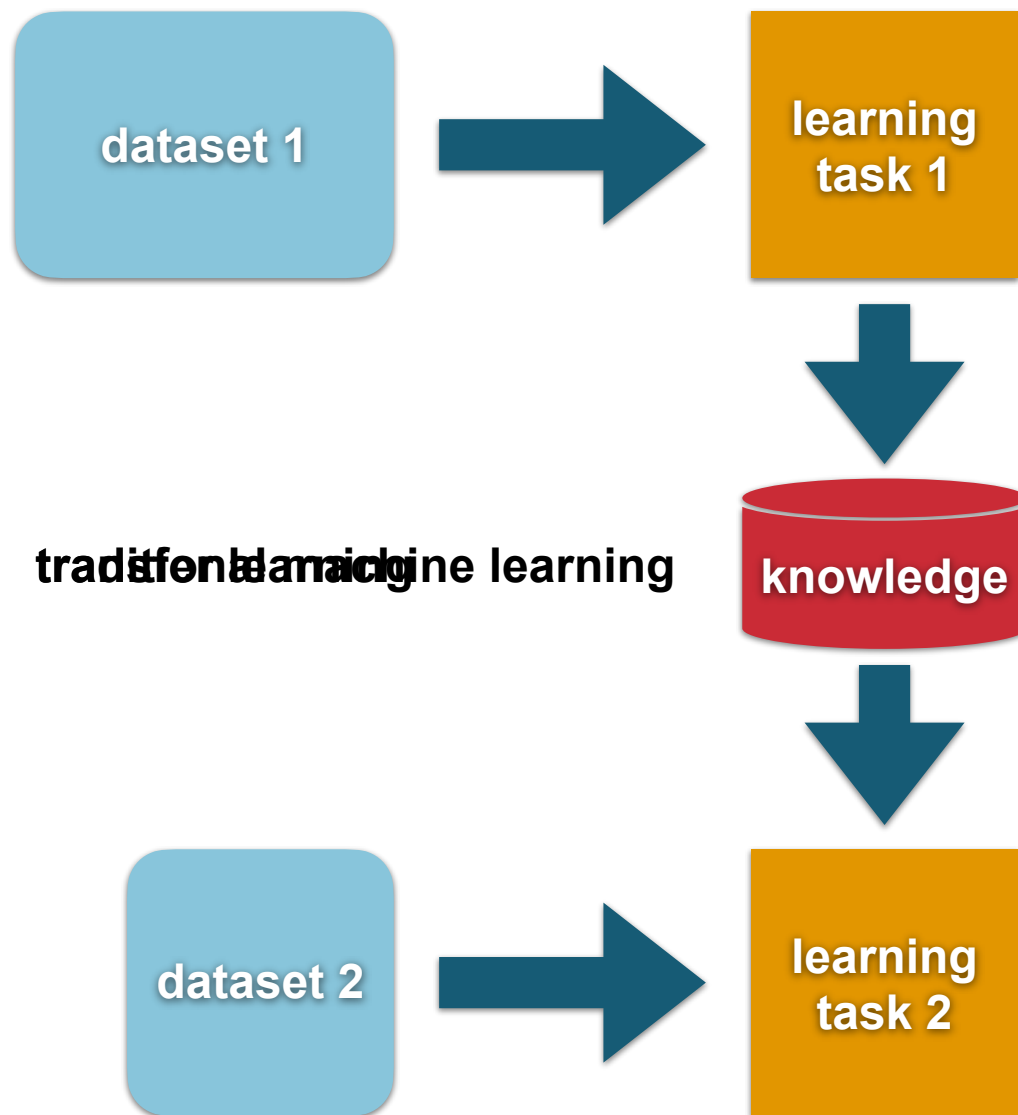


Gene module detection



Rapid analysis & contextualisation of new data

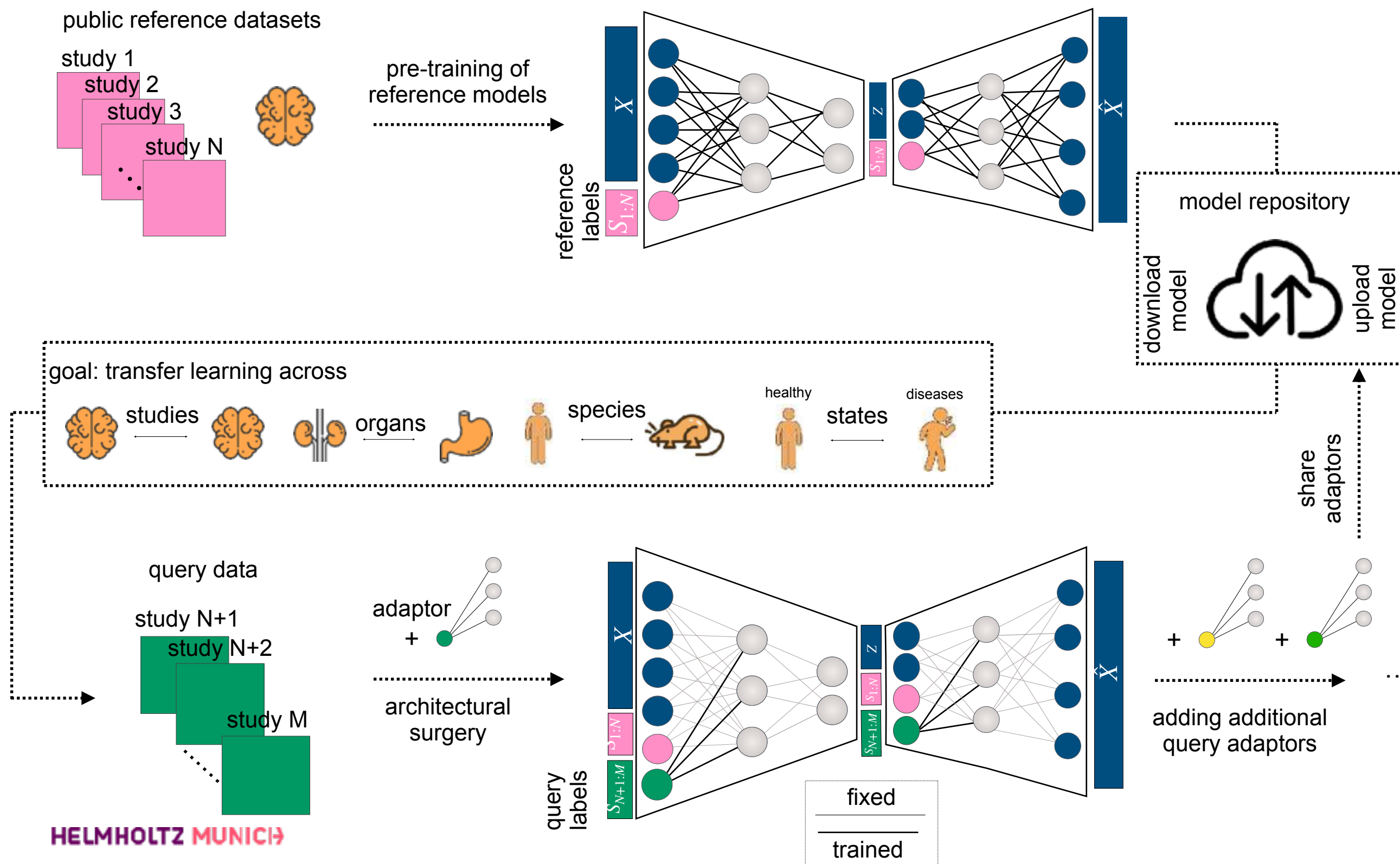
Query-to-reference data integration by transfer learning



Query-to-reference data integration by transfer learning



M Lotfallahi

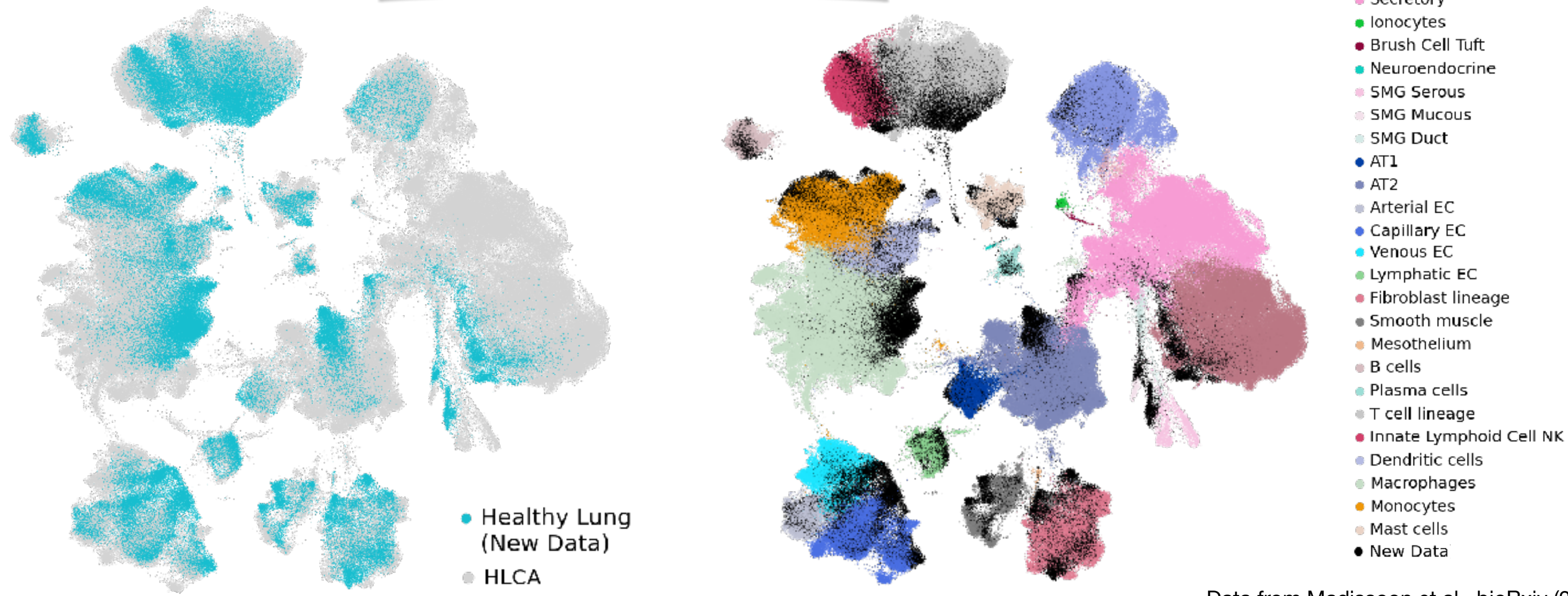


Projecting to the HLCA enables rapid analysis of new data

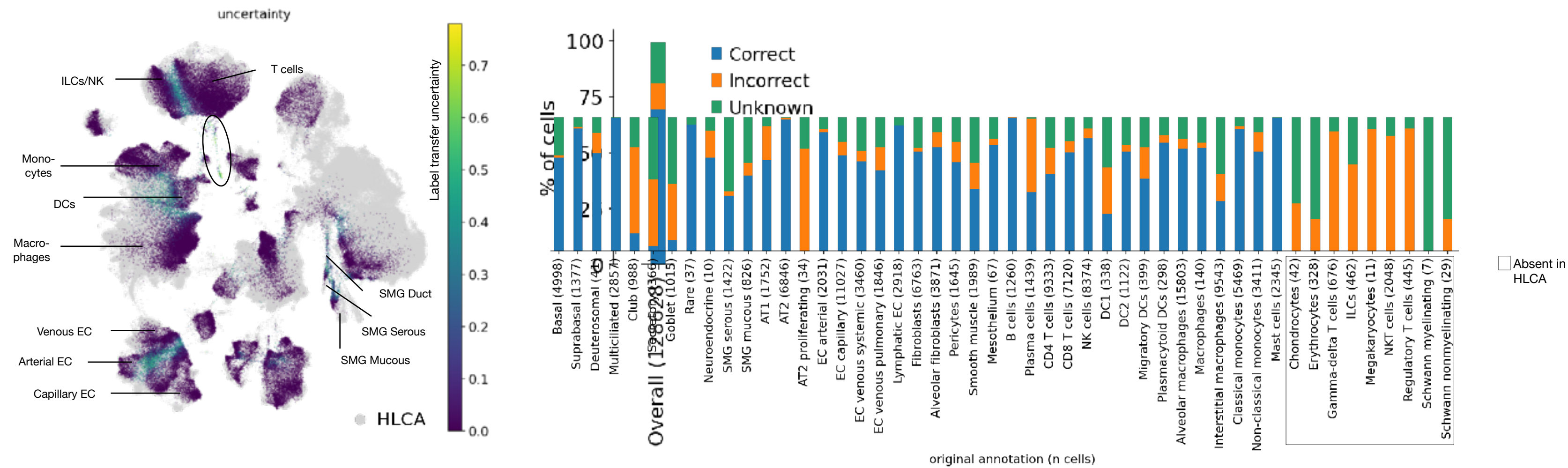
A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche

Elo Madisson, Amanda J. Oliver, Vitali Kleshchinskii, Anna Willbrey-Clark, Krzysztof Polanski, Ana Ribeiro Orsi, Lira Mamanova, Liam Bolt, Nathan Richoz, Rasa Elmentake, J. Patrick Petr, Ni Huang, Peng He, Monika Dabrowska, Sophie Pritchard, Liz Tuck, Elena Prigmore, Andrew Knights, Agnes Ossianz, Adam Hunter, Sara E. Vieira, Minal Patel, Nikitas Georgakopoulos, Krishnas Mahubani, Kouroush Saeb-Parsy, Menna Clatworthy, Omer Ali Bayraktar, Oliver Stegle, Naohiko Kumasaka, Sarah A. Teichmann, Kerstin B. Meyer

doi: <https://doi.org/10.1101/2021.11.26.470108>

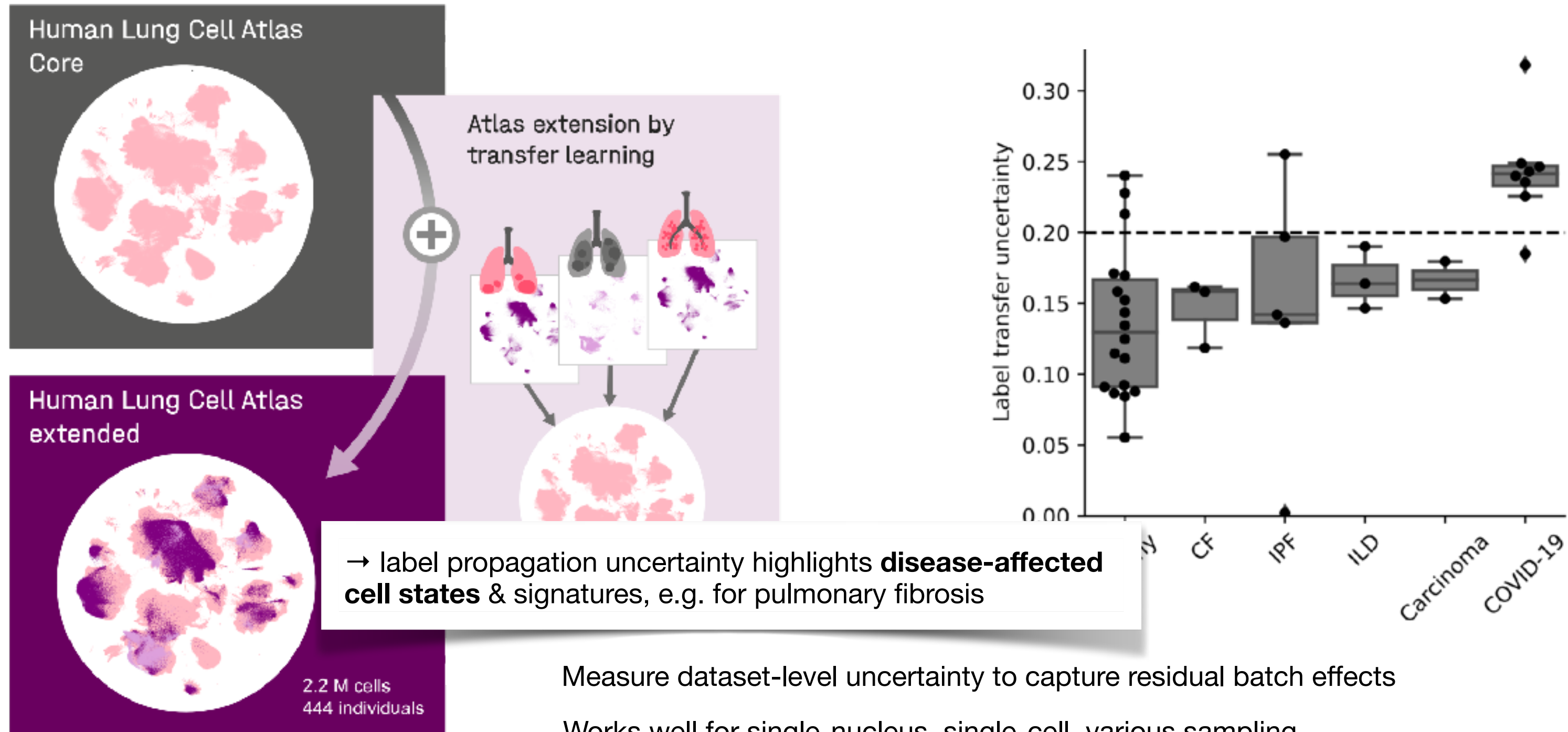


Projecting to the HLCA enables rapid analysis of new data



Label propagation performs competitively vs manual annotation

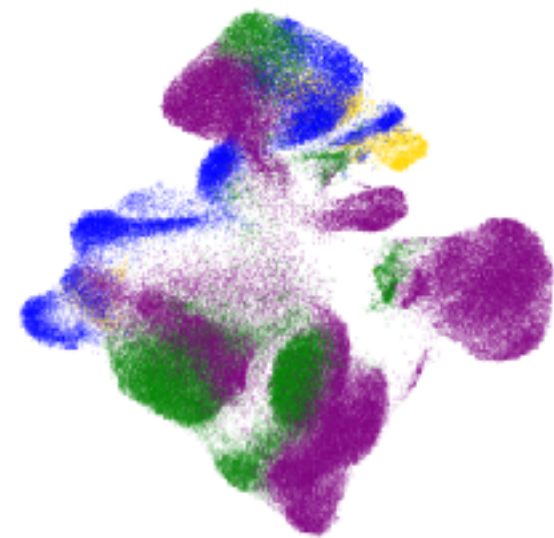
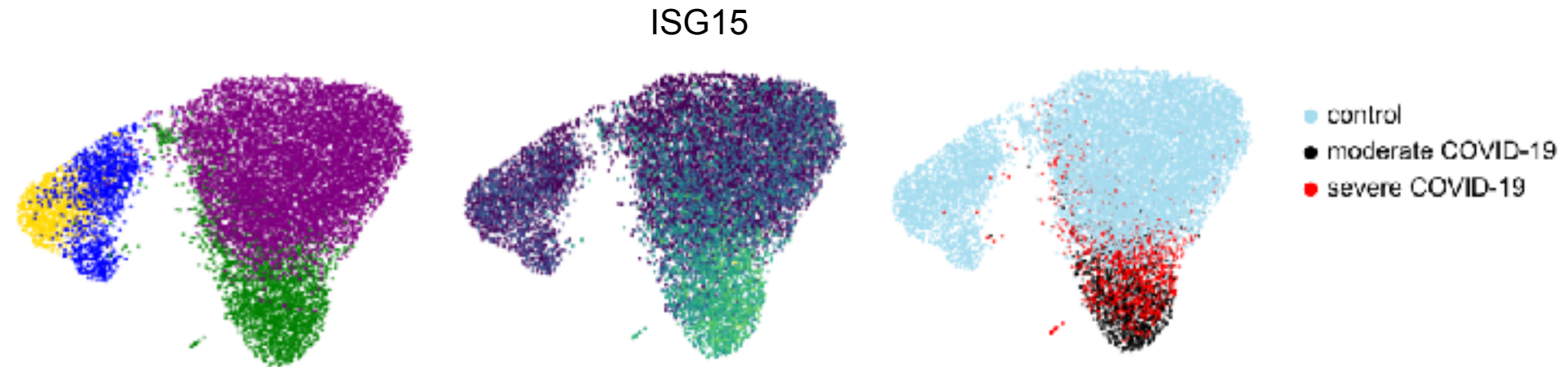
Extended Human Lung Cell Atlas: mapping of disease-affected cell states



Querying an atlas for disease: immune response for COVID19 on lung lavage samples

reference
query

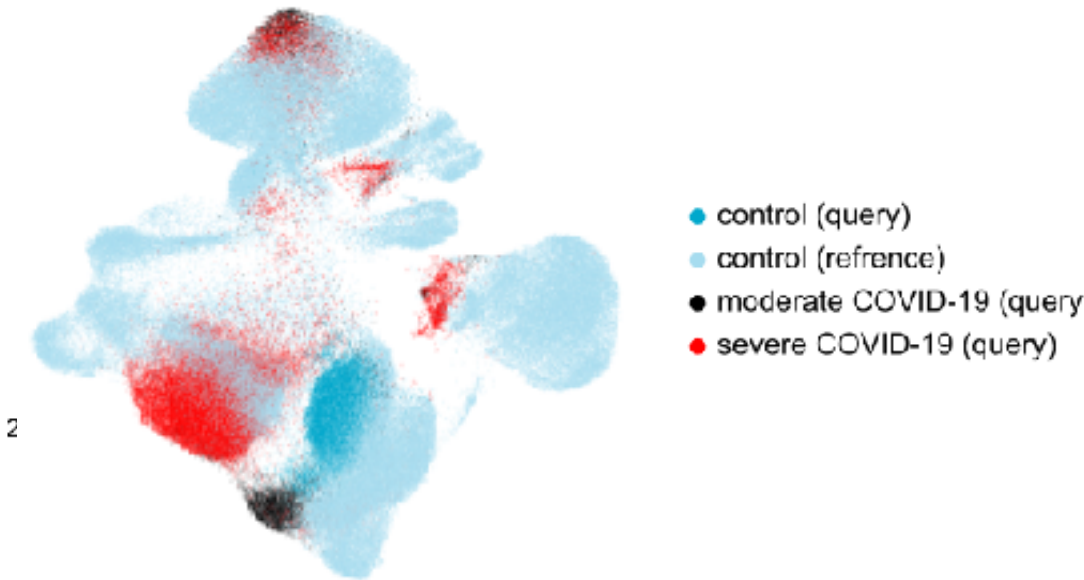
- Bone_Marrow
- Lung
- PBMCs
- BALF



CD8+ T cells

Macrophages

- AT1
- AT2
- B cell
- CD4+ T cells
- CD8+ T cells
- CD10+ B cells
- CD14+ Monocytes
- CD16+ Monocytes
- CD20+ B cells
- Ciliated
- DC
- Erythrocytes
- Erythroid progenitors
- HSPCs
- M2 Macrophages
- Macrophages
- Mast cells
- Megakaryocytes
- Monocyte progenitors
- Monocytes
- NK
- NKT cells
- Neutrophil
- Plasma
- Proliferating Macrophage
- Secretory
- Signaling Alveolar Epithelial Type 2
- T
- mDC
- pDC



FABP4

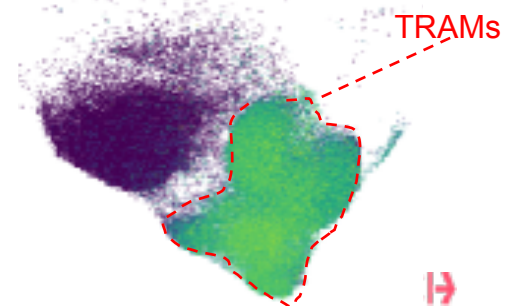
CXCL10

CCL2

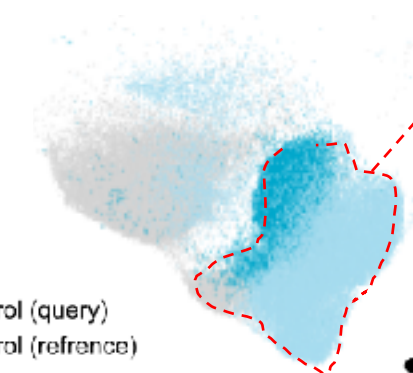
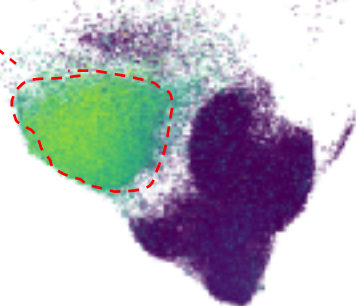
TRAMs

MoMs

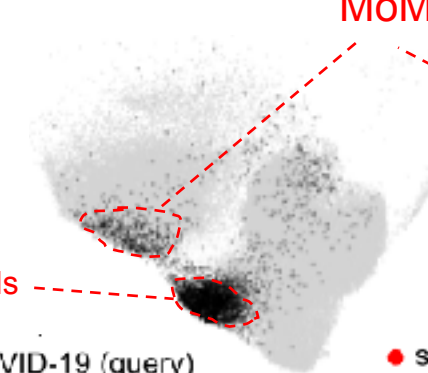
high
expression
low



MoMs

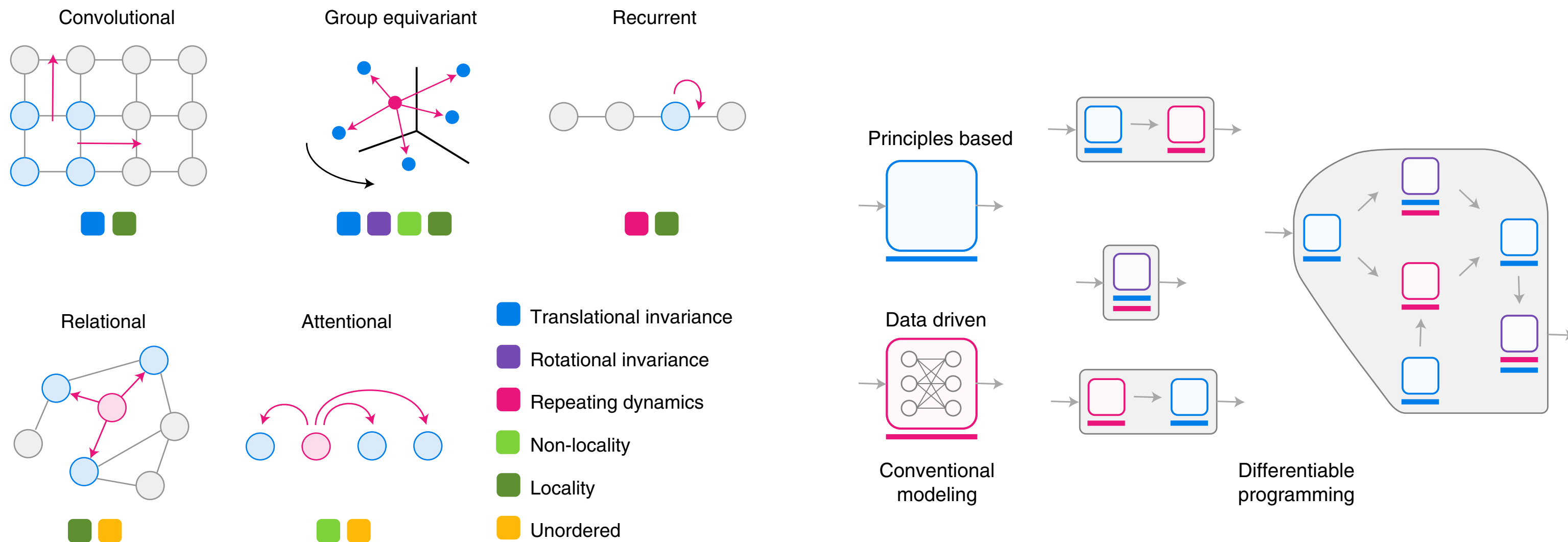


TRAMs



● moderate COVID-19 (query)
● severe COVID-19 (query)

Outlook: differential biology - deep learning for modeling molecular mechanisms



network primitives and invariances

→ questions: reuse primitives? add constraints?

Outlook: towards adaptive ML via continual learning

new models (prior information)

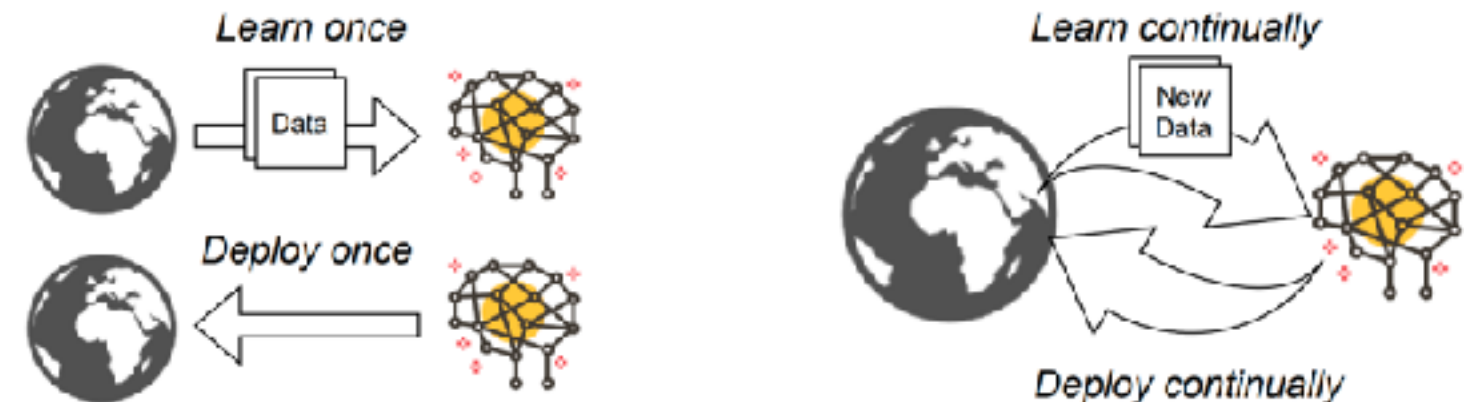
- model a particular modality/view of your problem
- then analyse & interpret
- extend with prior knowledge or constrain with mechanistic information

new features

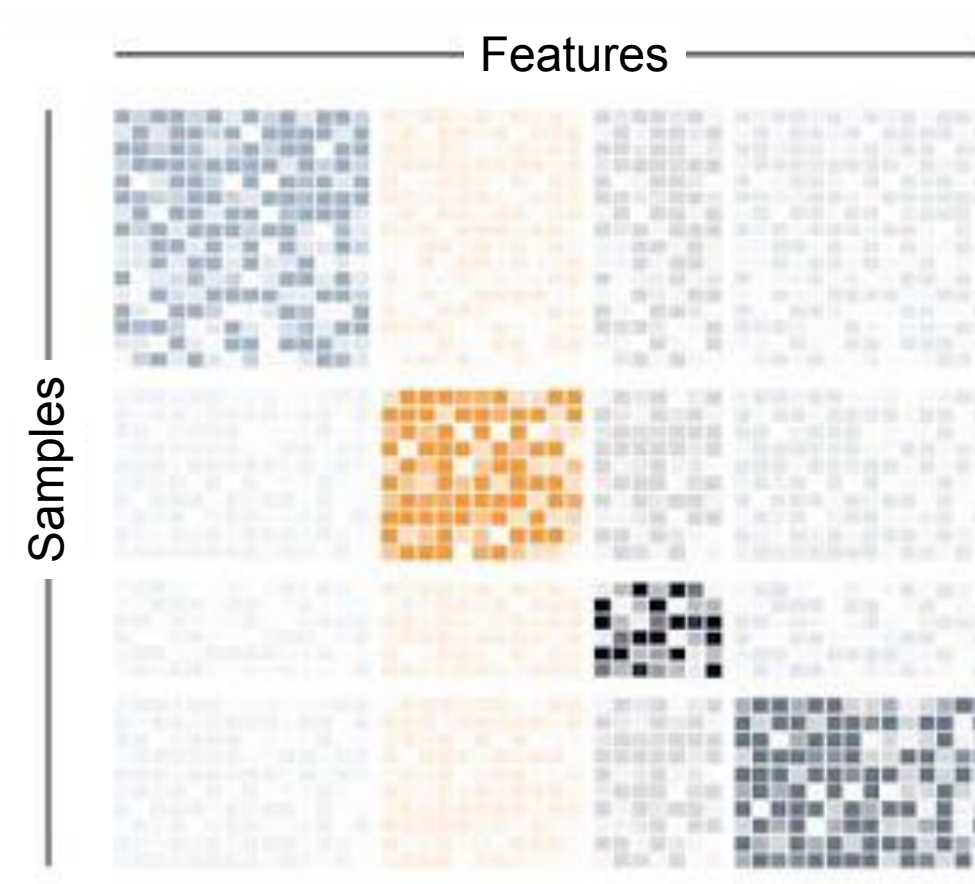
- add additional views to the model
- extend feature space, potentially link mechanistically

new data (samples)

- transfer to new domain
- reuse that model
- loop into experimental design (perturbations etc)



source: medium.com

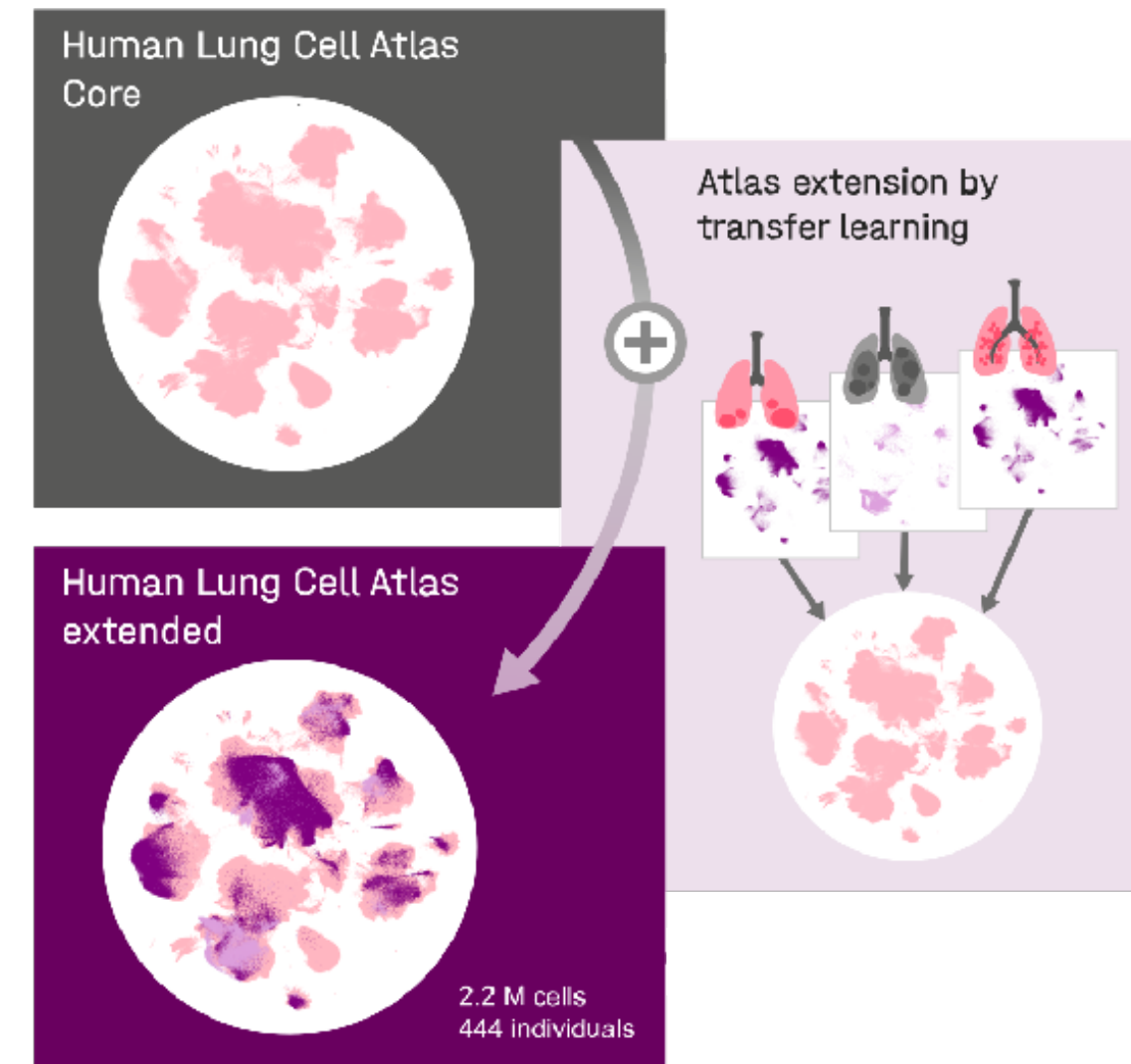


conclusion

- » building an *integrated Human Lung Cell atlas* across 44 data sets
- » *scArches* & extensions: transfer learning for efficient reference atlas reuse
- » sfaira - sc data management made simple

outlook

- » building large-scale & multimodal reference atlases
- » sfairaZero: learning a 30M-cells organism-level atlas
- » spatial transcriptomics -> towards multi-scale modeling
- » (causal) learning from perturbations & experimental design
- » towards continual learning





Bundesministerium
für Bildung
und Forschung



Deutsche
Forschungsgemeinschaft
DFG



Acknowledgements

theislab@
Helmholtz Munich Computational Health Center

twitter
@fabian_theis