Single-cell Multi-Omics integration
SinCellTE 2022

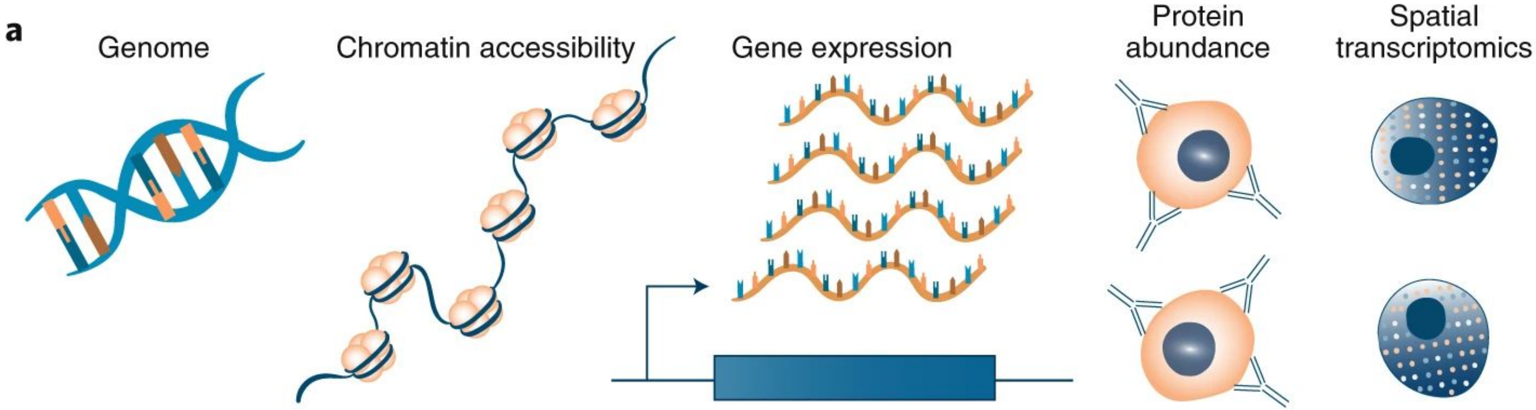
Andres Quintero
Carl Herrmann

Health Data Science Unit
Heidelberg University

13/01/2022
Roscoff, France

Content

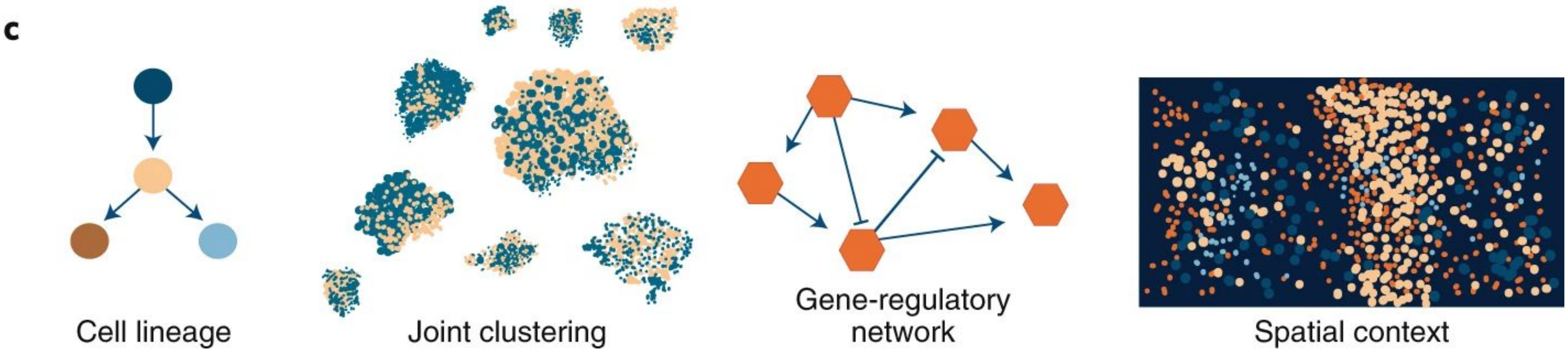
Single-cell multi-omics



To disentangle the complexity of gene-regulatory and cell–cell communication networks that drive cell functions and responses, measuring multiple modalities of the multivariate phenotypic and genetic cellular state is extremely powerful.

Efremova et al. 2020. Nature Methods.
DOI: 10.1038/s41592-019-0692-4

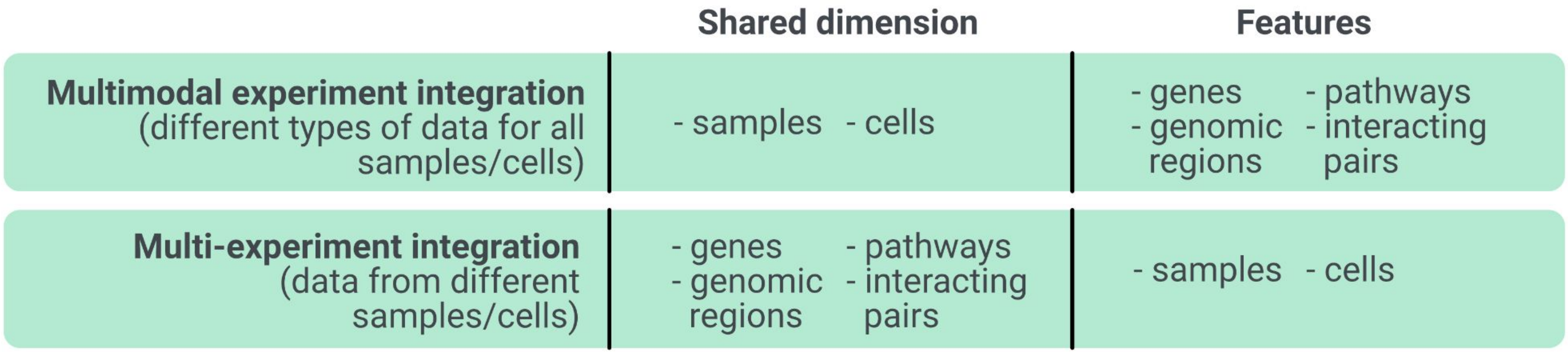
Single-cell multi-omics



To disentangle the complexity of gene-regulatory and cell–cell communication networks that drive cell functions and responses, measuring multiple modalities of the multivariate phenotypic and genetic cellular state is extremely powerful.

Efremova et al. 2020. Nature Methods.
DOI: 10.1038/s41592-019-0692-4

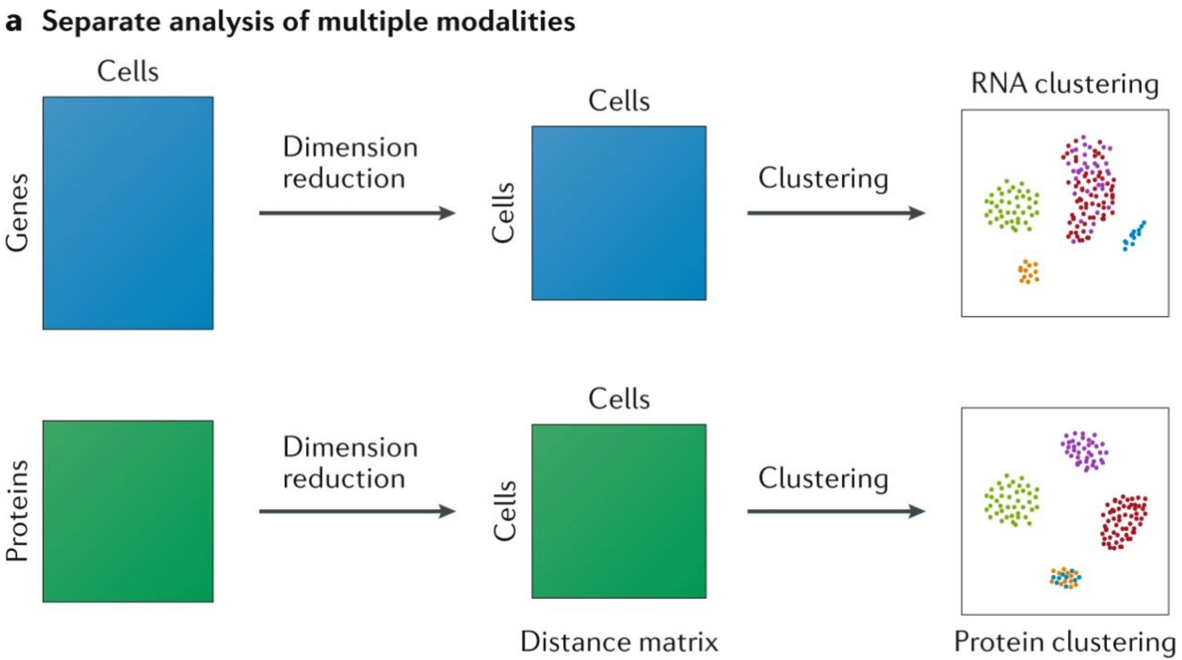
Single-cell multi-omics



	Shared dimension	Features
Multimodal experiment integration (different types of data for all samples/cells)	- samples - cells	- genes - genomic regions - pathways - interacting pairs
Multi-experiment integration (data from different samples/cells)	- genes - genomic regions - pathways - interacting pairs	- samples - cells

The integrative analysis of multi-omics data can be done in two different ways, integrating over cells for multimodal datasets or over molecular features.

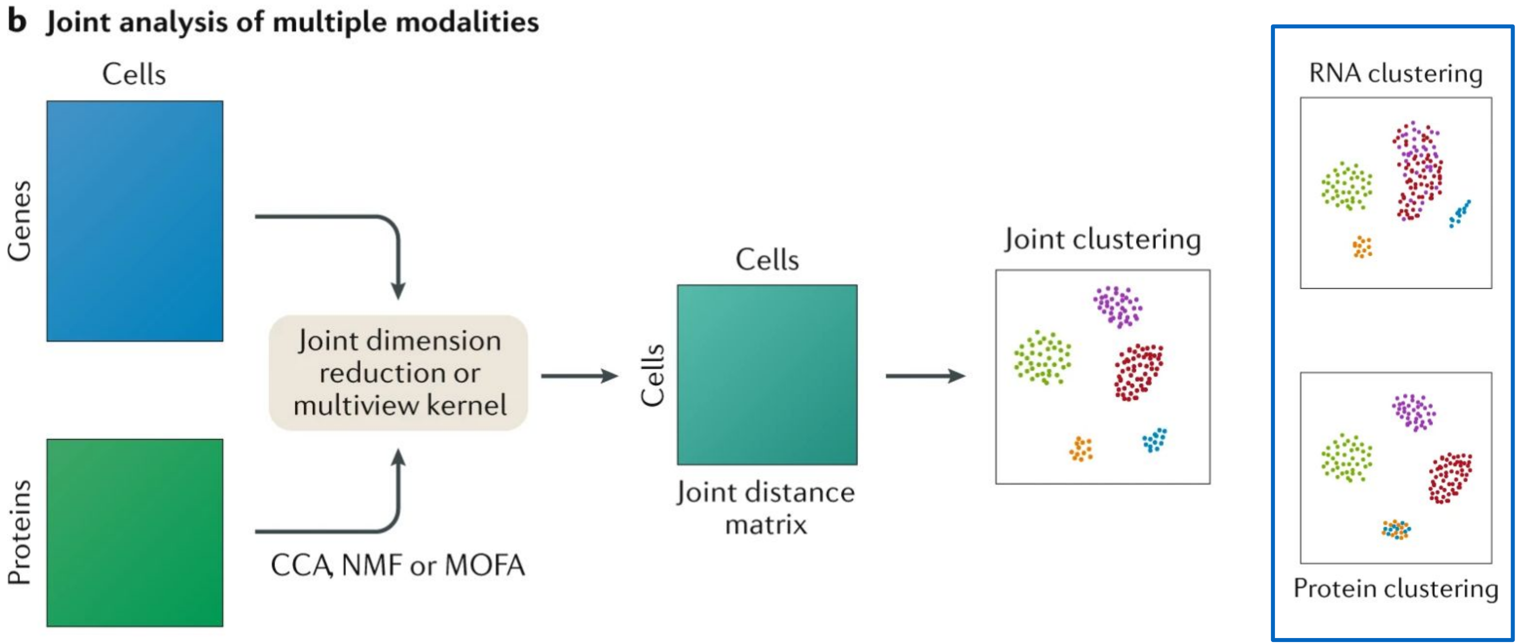
Single-cell multi-omics



Independent analysis of multiple modalities measured from the same group of cells can lead to conflicting identification of clusters of cells.

Stuart et al. 2019. Nat Rev Genet.
DOI: 10.1038/s41576-019-0093-7

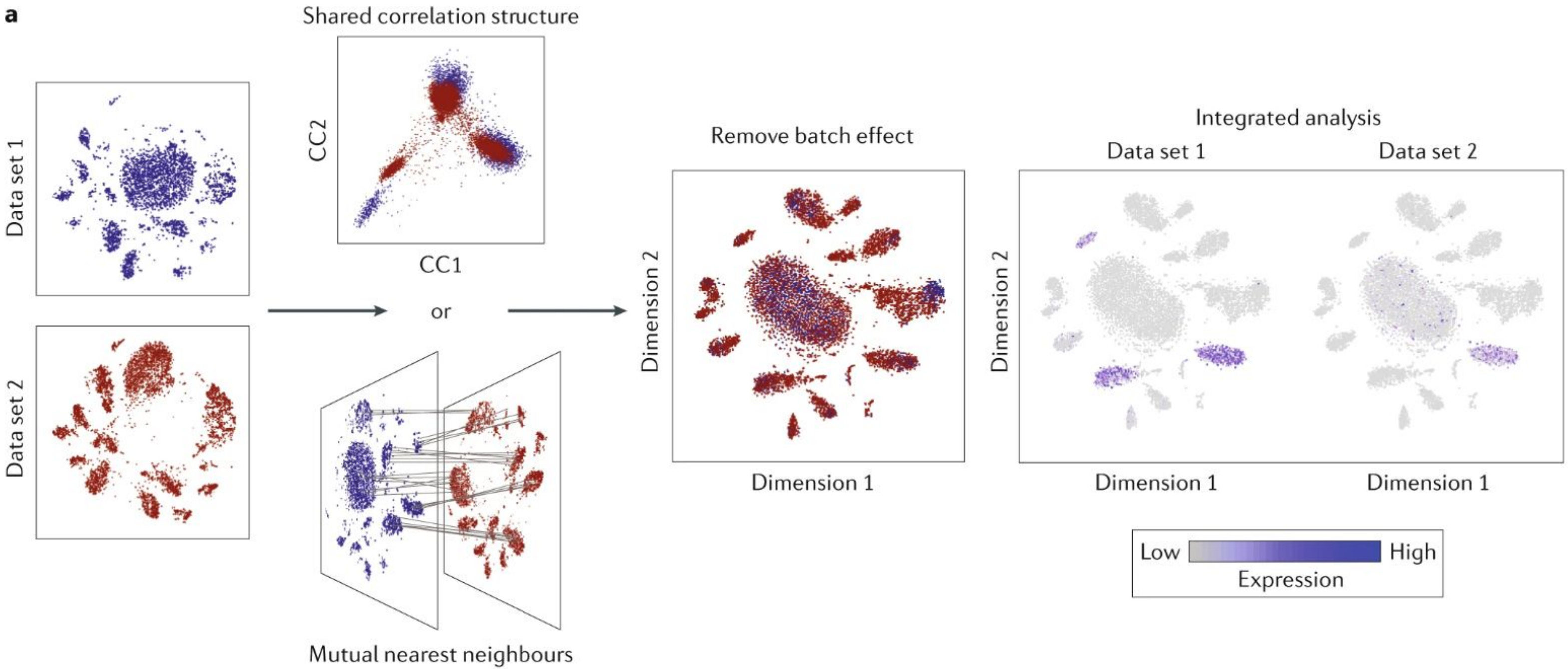
Single-cell multi-omics



Joint analysis of multiple modalities measured from the same group of cells can have greater power to identify unique cell states.

Stuart et al. 2019. Nat Rev Genet.
DOI: 10.1038/s41576-019-0093-7

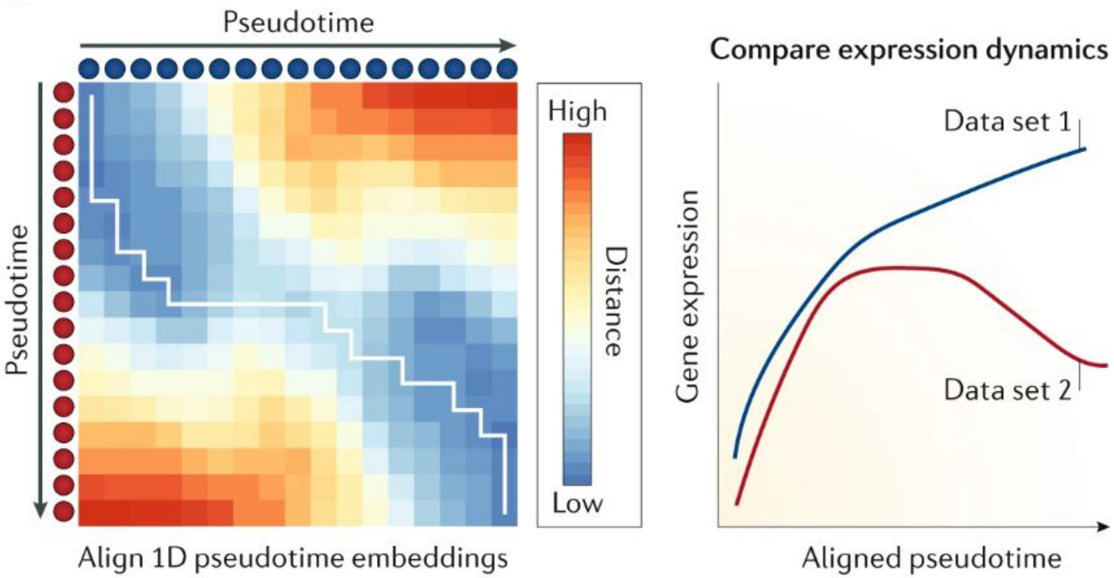
Integrating multiple single-cell datasets



Multiple data sets can be integrated computationally to facilitate downstream comparative analysis. Shared correlation structure can be detected using canonical correlation analysis (CCA) or mutual nearest neighbours (MNNs) identified.

Stuart et al. 2019. Nat Rev Genet.
DOI: 10.1038/s41576-019-0093-7

Integrating multiple single-cell datasets



The alignment of pseudotime vectors can allow temporal differences in cell trajectories to be removed and equivalent points in the trajectory across two data sets to be identified.

Gene expression can then be directly compared across the corrected pseudotime trajectories.

Stuart et al. 2019. Nat Rev Genet.
DOI: 10.1038/s41576-019-0093-7

Integrating multiple single-cell datasets

Example: Projection of bulk and single-cell data onto a reference embedding

nature cancer

Moritz Gartlgruber*, Ashwini Kumar Sharma*, Andrés Quintero*, Daniel Dreidax*, Selina Jansky, Young-Gyu Park, Sina Kreth, Johanna Meder, Daria Doncevic, Paul Saary, Umut H. Toprak, Naveed Ishaque, Elena Afanasyeva, Elisa Wecht, Jan Koster, Rogier Versteeg, Thomas G. P. Grünewald, David T. W. Jones, Stefan M. Pfister, Kai-Oliver Henrich, Johan van Nes, Carl Herrmann*, Frank Westermann*. **Super enhancers define regulatory subtypes and cell identity in neuroblastoma. *Nature Cancer.* (December 2020).**

Division of Neuroblastoma Genomics

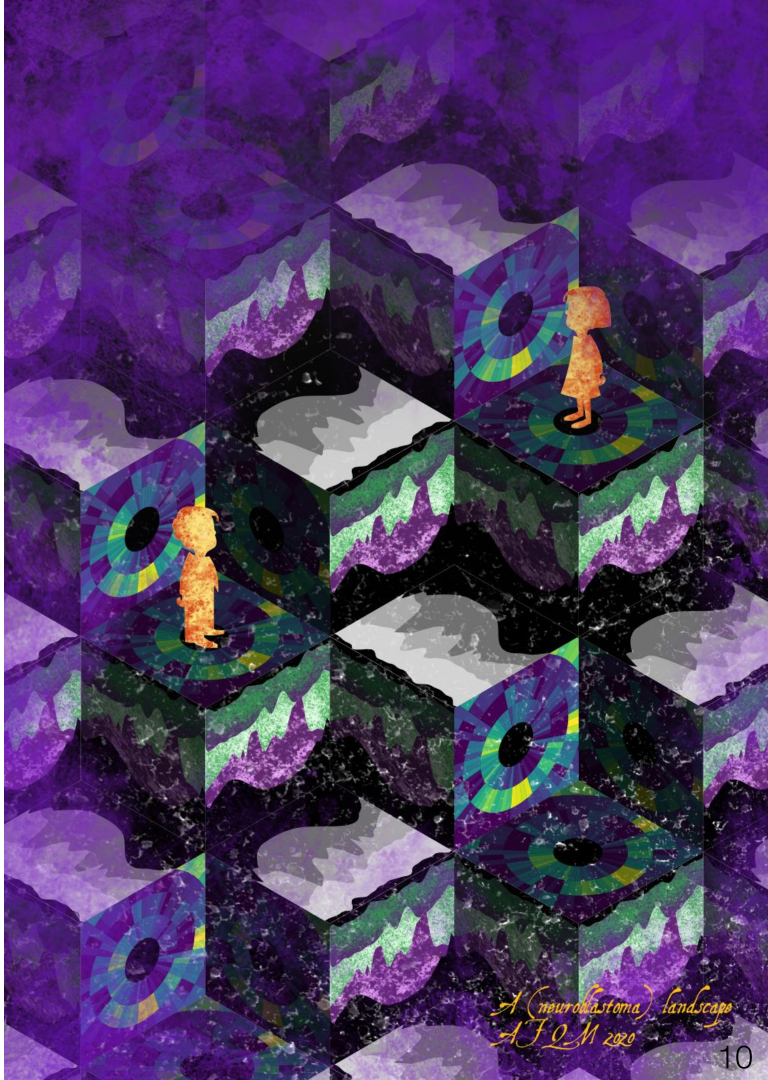
dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION
Research for a Life without Cancer

KiTZ Hopp-Kindertumorzentrum Heidelberg

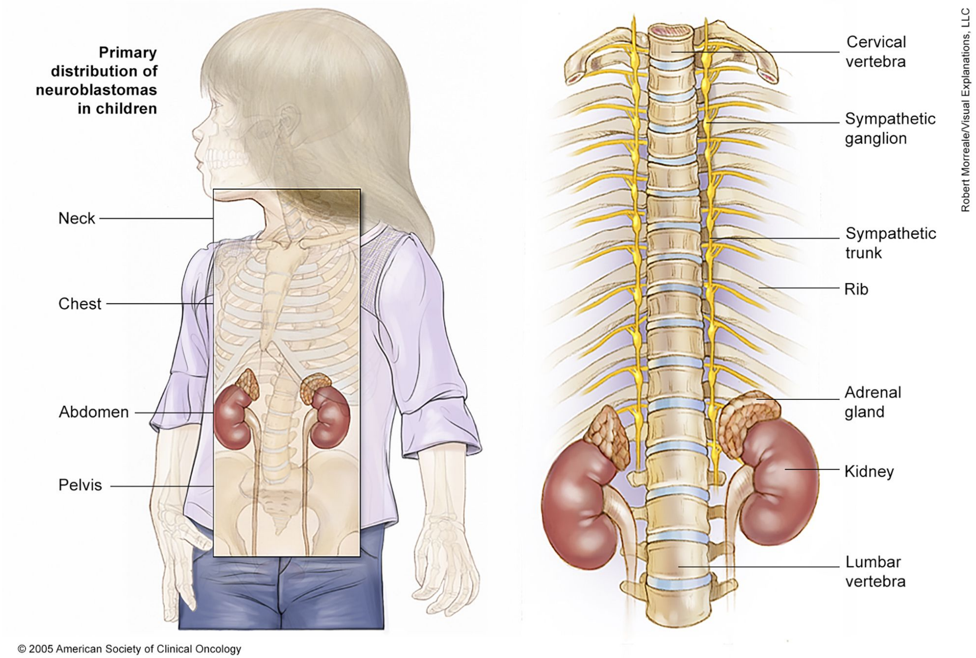
BioQuant MODEL base of LIFE

healthdata scienceunit

HEIDELBERG FACULTY OF MEDICINE



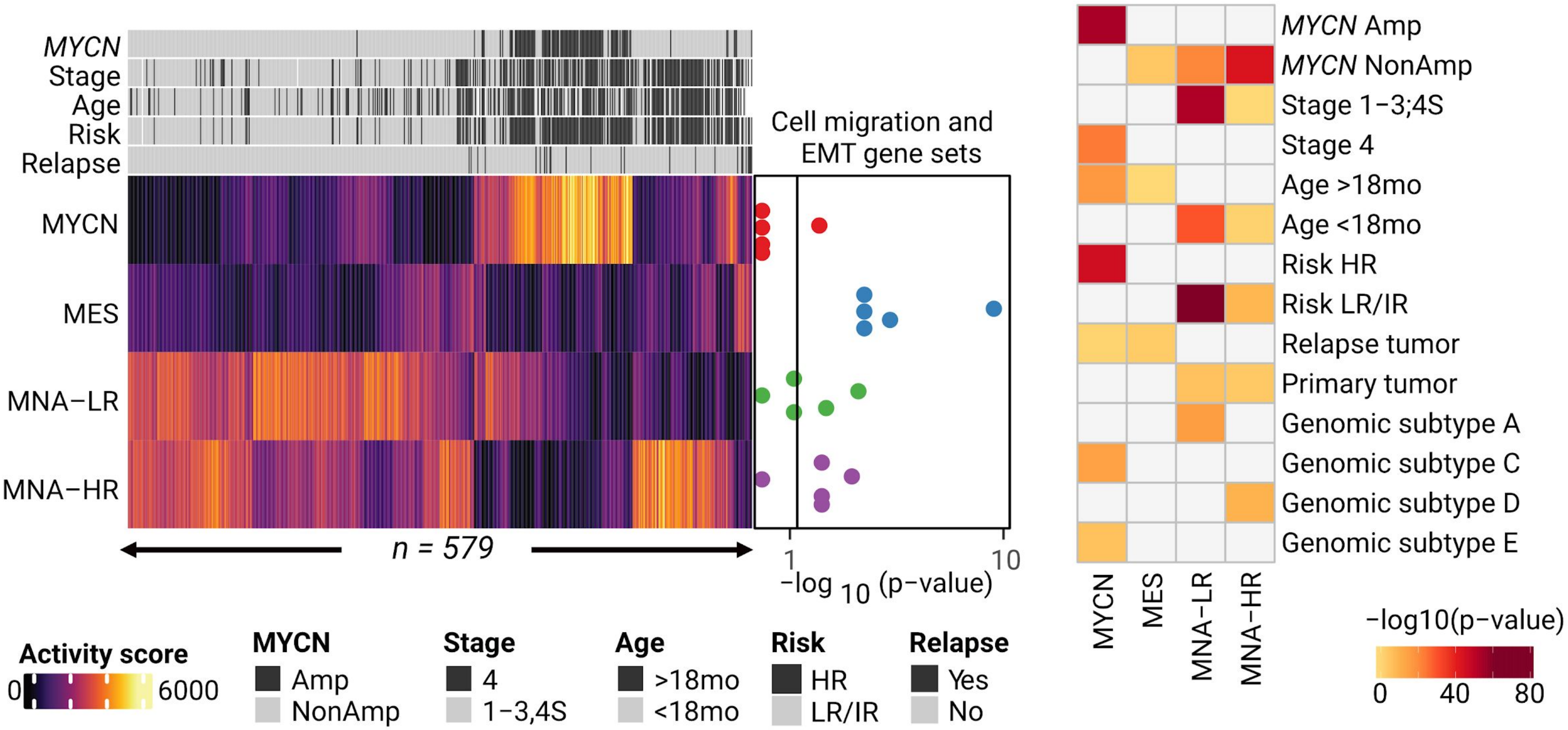
NMF to Reveal Regulatory Subtypes in Neuroblastoma



Taken from:
https://www.cancer.net/cancer-types/neuroblastoma-childhood/medical-illustrations
2005 American Society of Clinical Ontology

- Pediatric solid tumor of the sympathetic nervous system (SNS).
- Common place of origin is the adrenal gland.
- Is thought to develop from neural crest-derived precursor cells of the SNS.

Neuroblastoma transcriptomic signatures

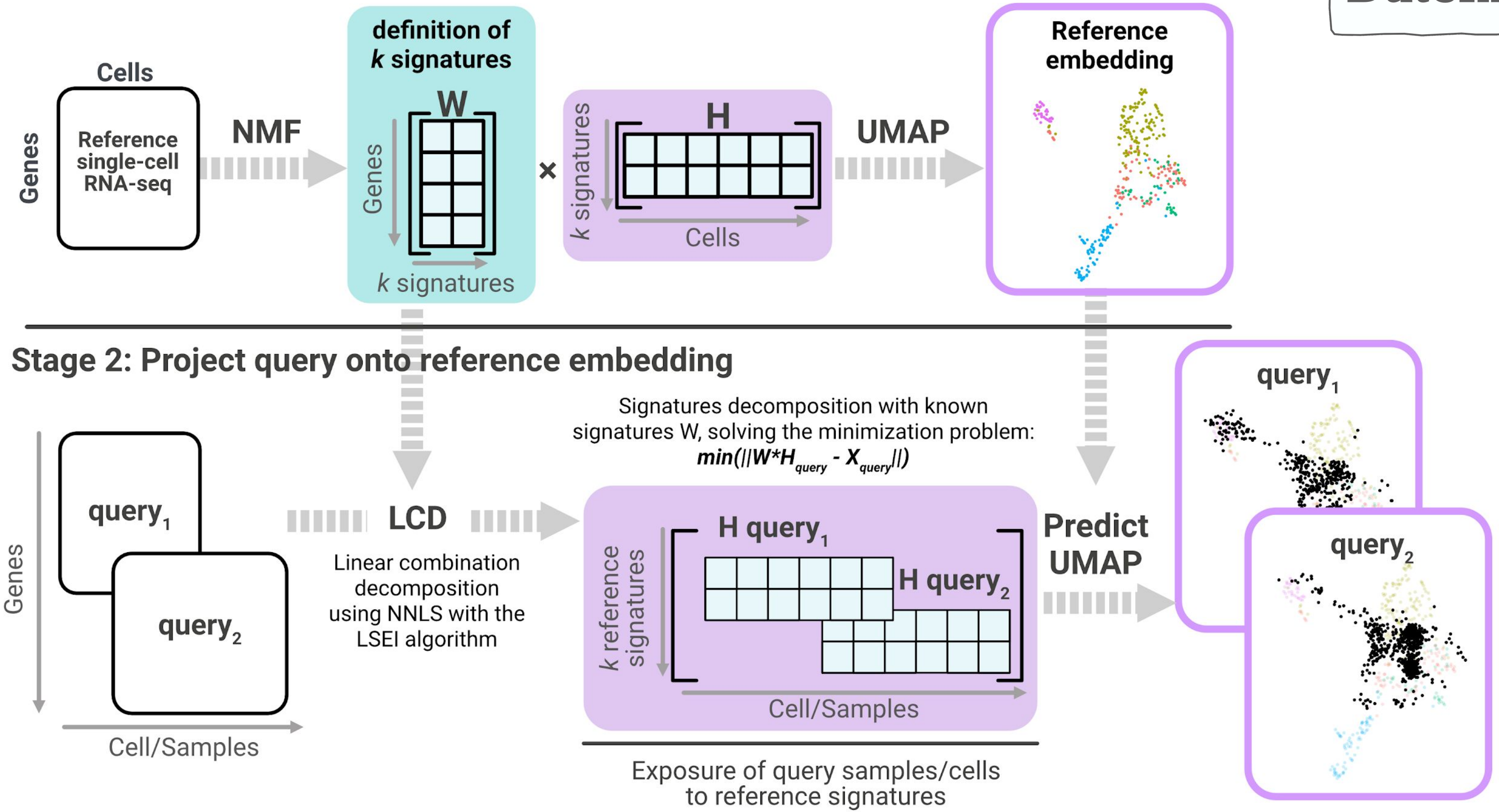


Gartlgruber et al. 2020. Nature Cancer.
DOI: 10.1038/s43018-020-00145-w

Projection onto reference dataset

ButchR

Stage 1: Build reference embedding



Stage 2: Project query onto reference embedding

Linear combination decomposition using NNLS with the LSEI algorithm

Signatures decomposition with known signatures W, solving the minimization problem:
$min(||W*H_{query} - X_{query}||)$

Exposure of query samples/cells to reference signatures

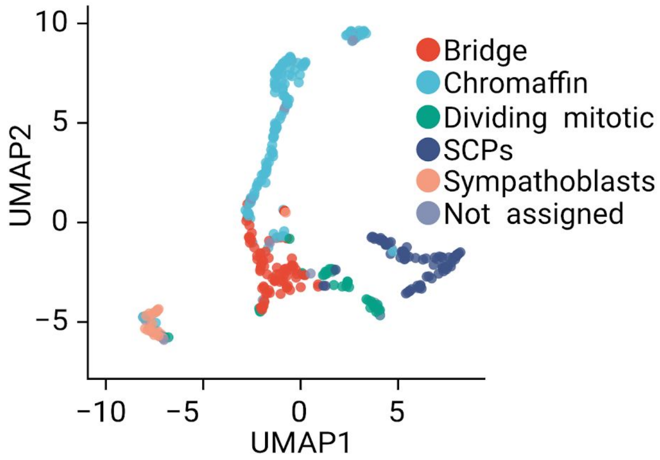
BioQuant

healthdata scienceunit

HEIDELBERG FACULTY OF MEDICINE

Projection of neuroblastomas onto mouse-developing adrenal sympathetic system

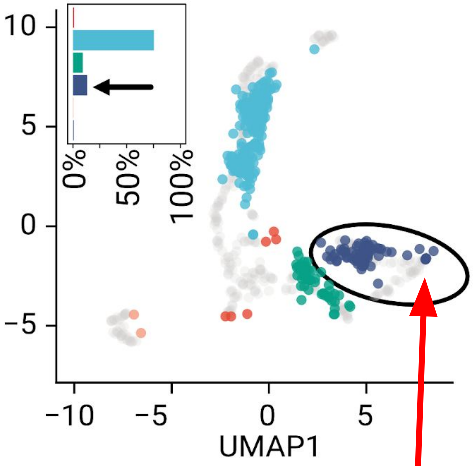
Mouse E12.5 developing adrenal medulla (Furlan *et al.* 2017)



Every point represents one single-cell

NB RNA-seq cohort projection

n = 573 projected samples

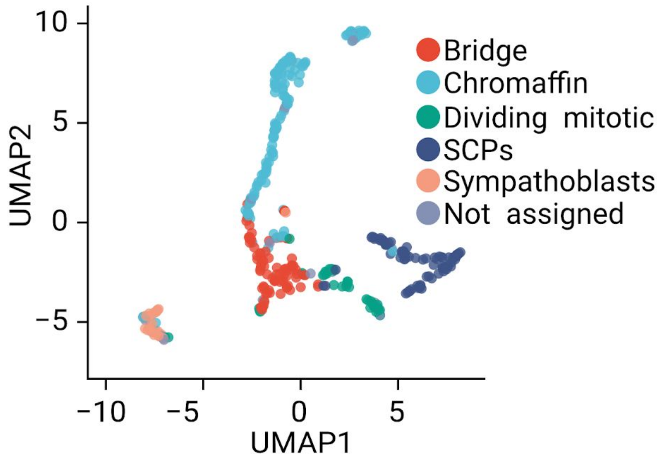


Every colored point represents one bulk sample

Gartlgruber et al. 2020. Nature Cancer.
DOI: 10.1038/s43018-020-00145-w

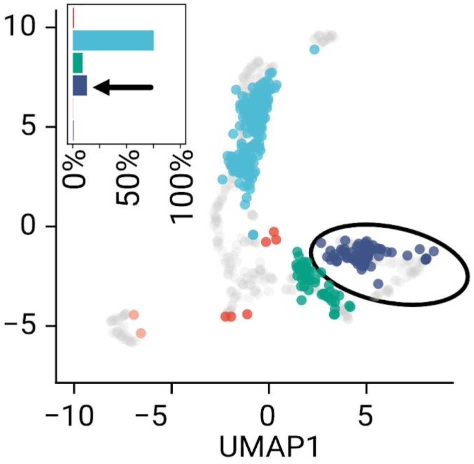
Projection of neuroblastomas onto mouse-developing adrenal sympathetic system

Mouse E12.5 developing adrenal medulla (Furlan *et al.* 2017)



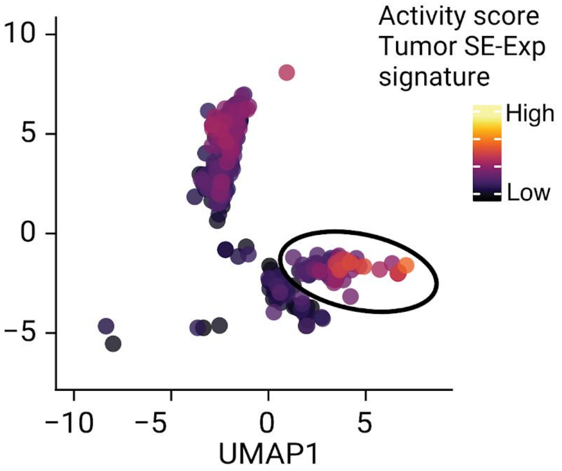
NB RNA-seq cohort projection

n = 573 projected samples



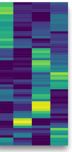
NB RNA-seq cohort projection

MES score

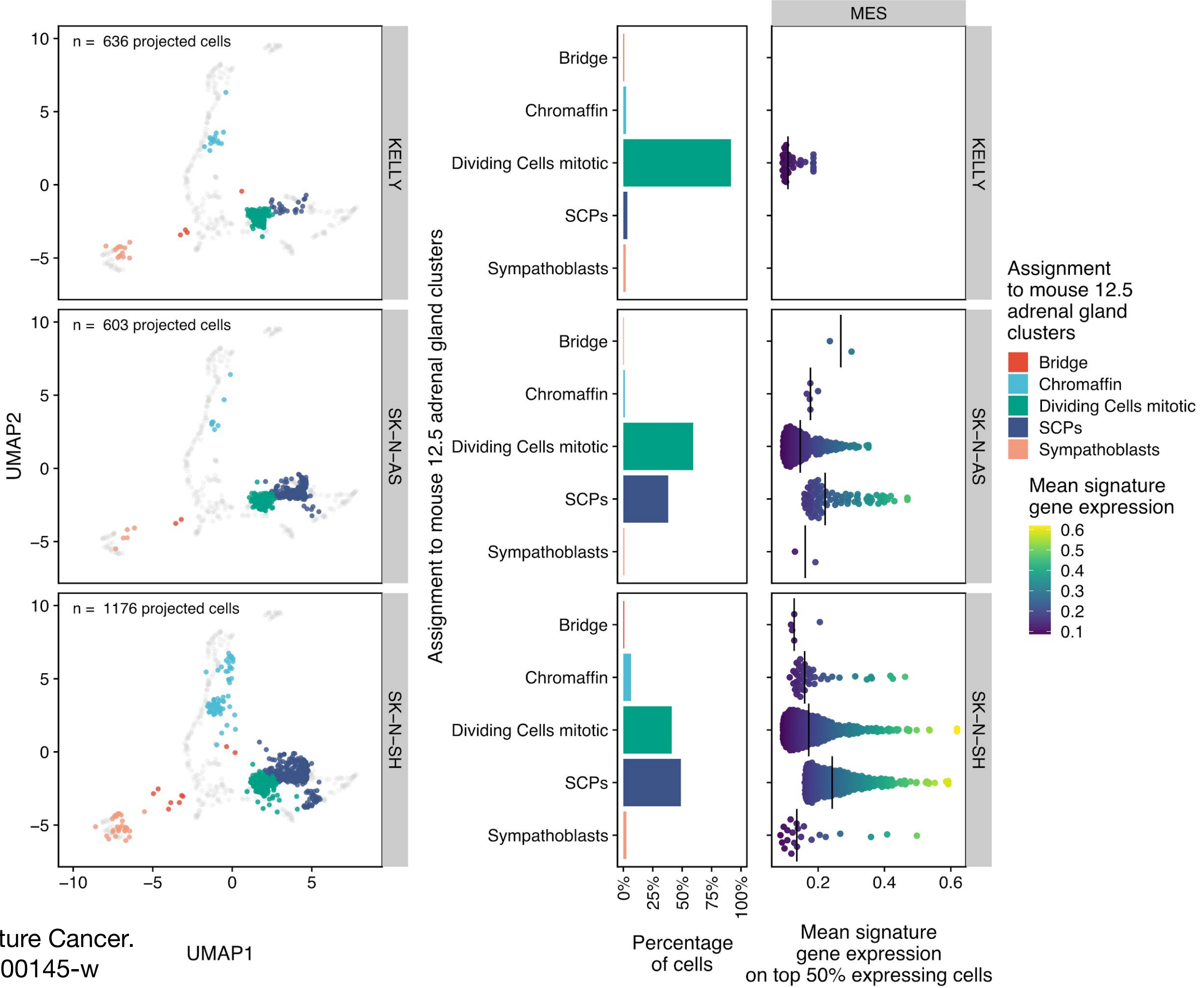


MES subtype associated with multipotent **Schwann cell precursors**

Gartlgruber et al. 2020. Nature Cancer.
DOI: 10.1038/s43018-020-00145-w



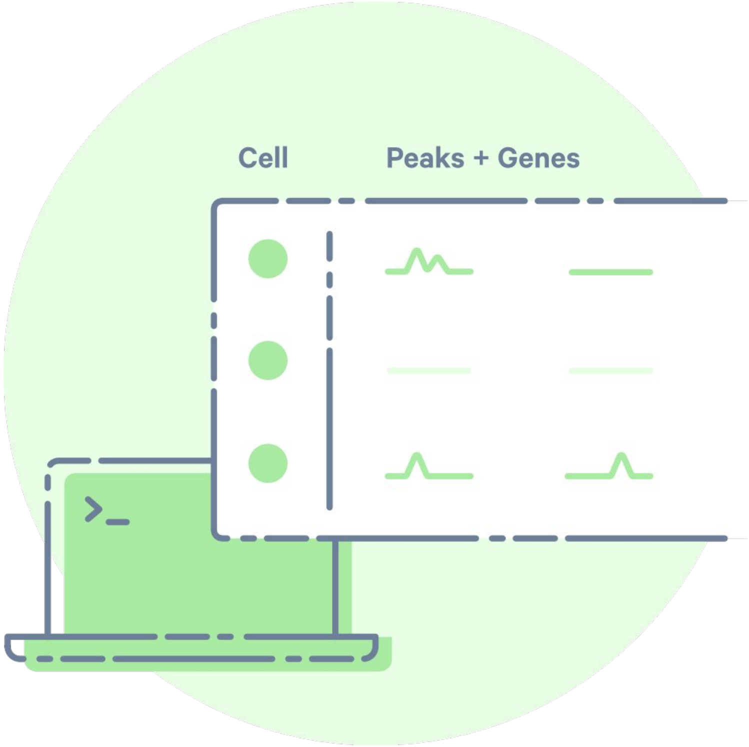
Projection of neuroblastoma cell lines onto mouse-developing adrenal sympathetic system



Gartlgruber et al. 2020. Nature Cancer.
DOI: 10.1038/s43018-020-00145-w

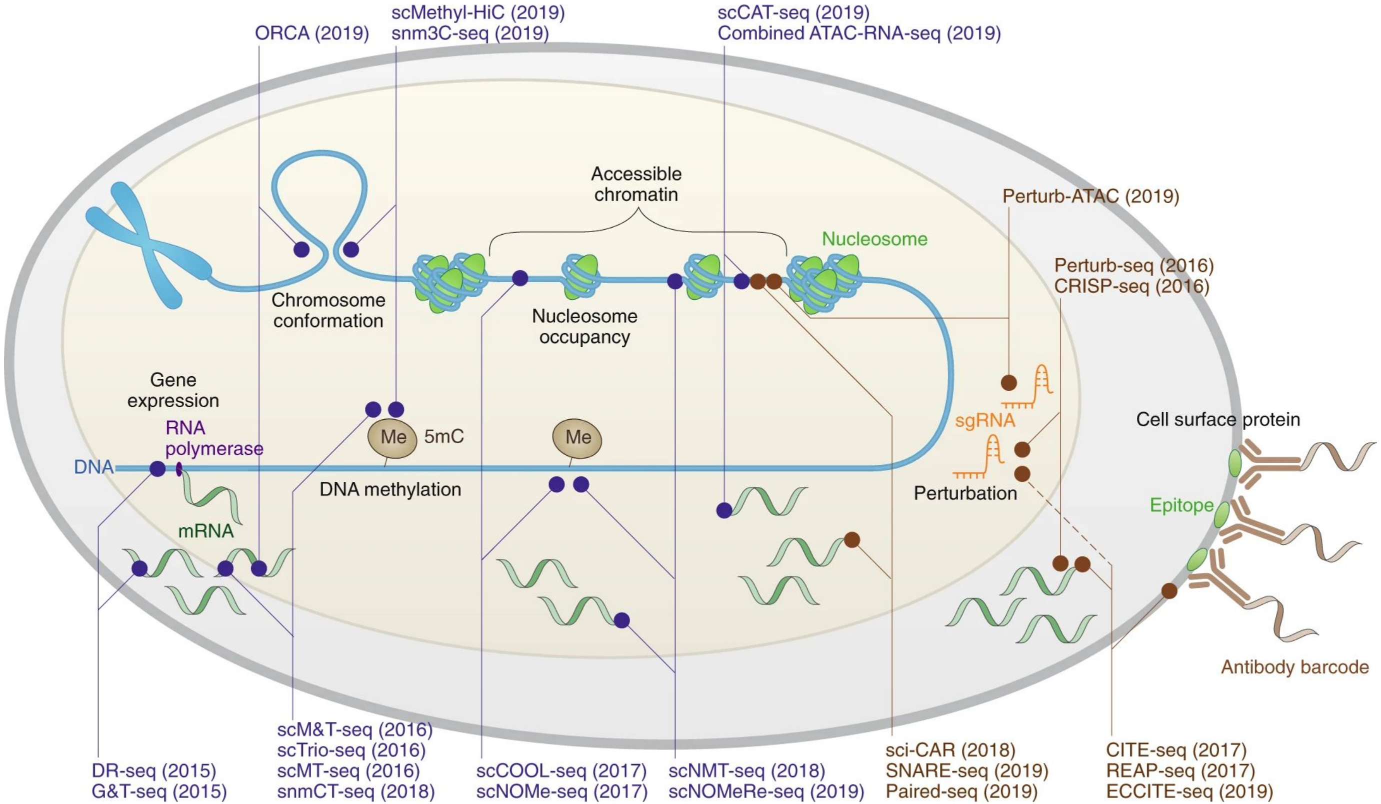
Andres Quintero

Single-cell multimodal integration



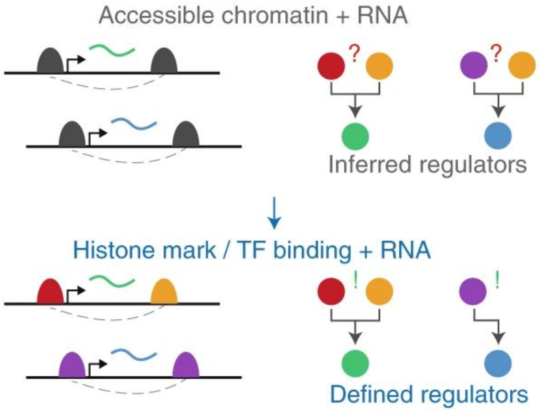
https://www.10xgenomics.com/products/single-cell-multiome-atac-plus-gene-expression

Single-cell multimodal profiling

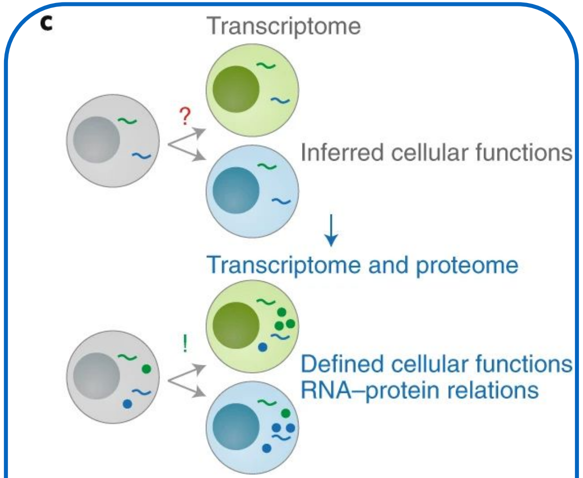


Zhu et al. 2020. Nature Methods.
DOI: 10.1038/s41592-019-0691-5

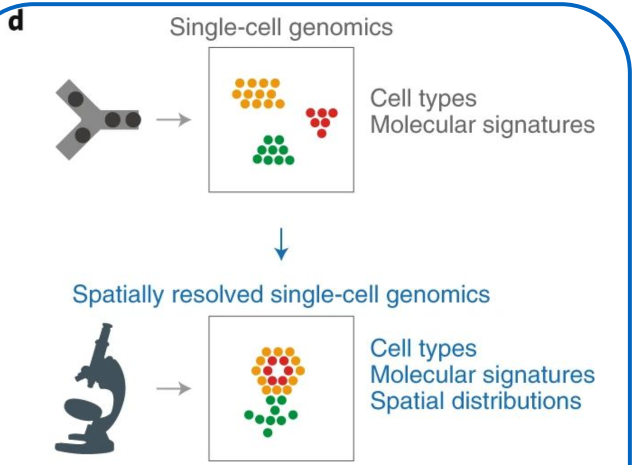
Single-cell multimodal profiling



Co-assays of histone marks or transcription-factor binding and gene expression will refine the understanding of gene regulatory networks.



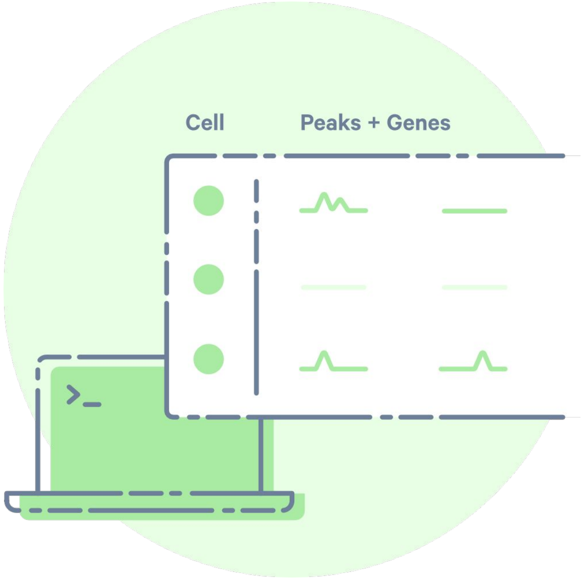
Joint analysis of transcripts and proteins in single cells will reveal the dynamic relationships of transcripts and protein abundance



Extension of the spatial transcriptomics toolbox to epigenome analysis will facilitate the dissection of molecular and cellular networks in complex tissues.

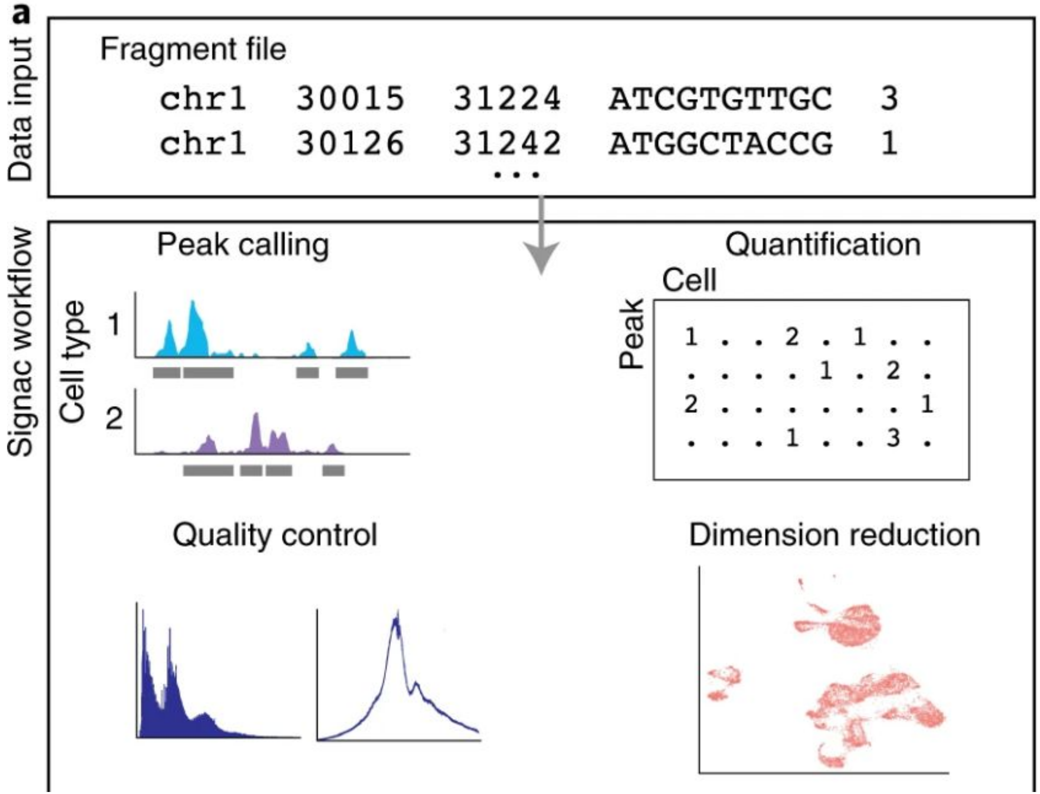
Zhu et al. 2020. Nature Methods.
DOI: 10.1038/s41592-019-0691-5

Single-cell Multimodal analysis



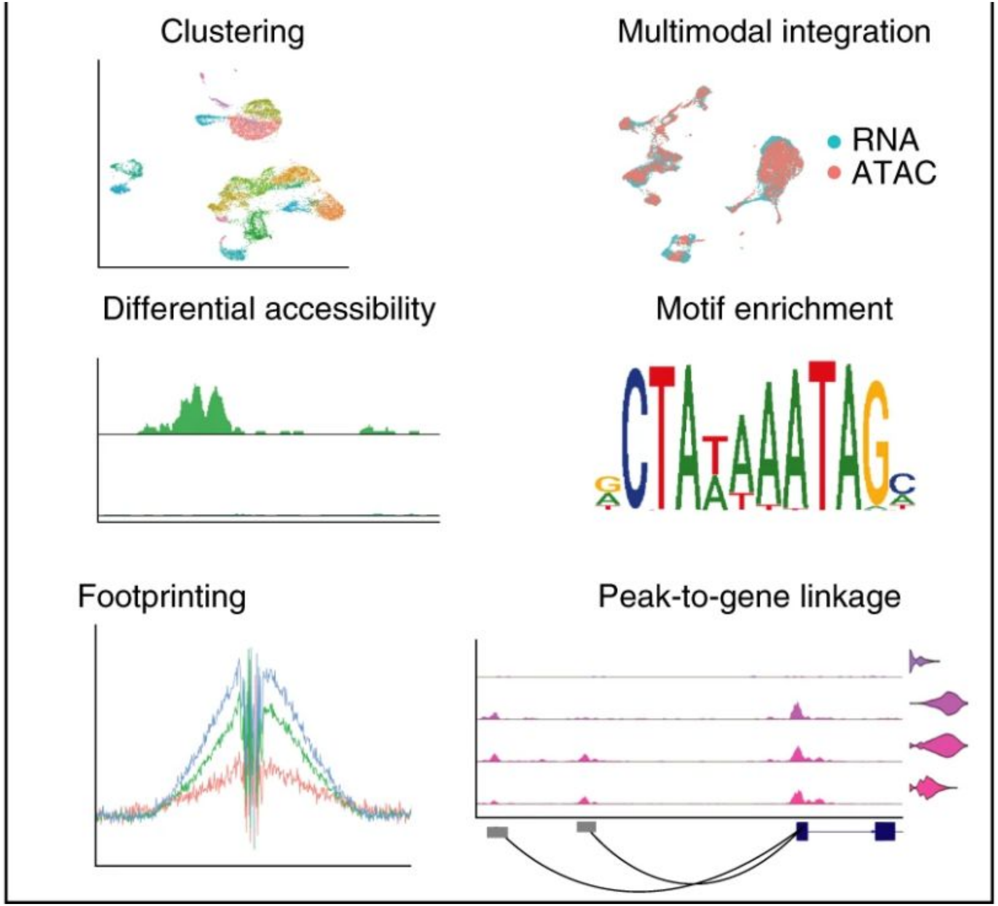
https://www.10xgenomics.com/products/single-cell-multiome-atac-plus-gene-expression

Single-cell multimodal analysis steps



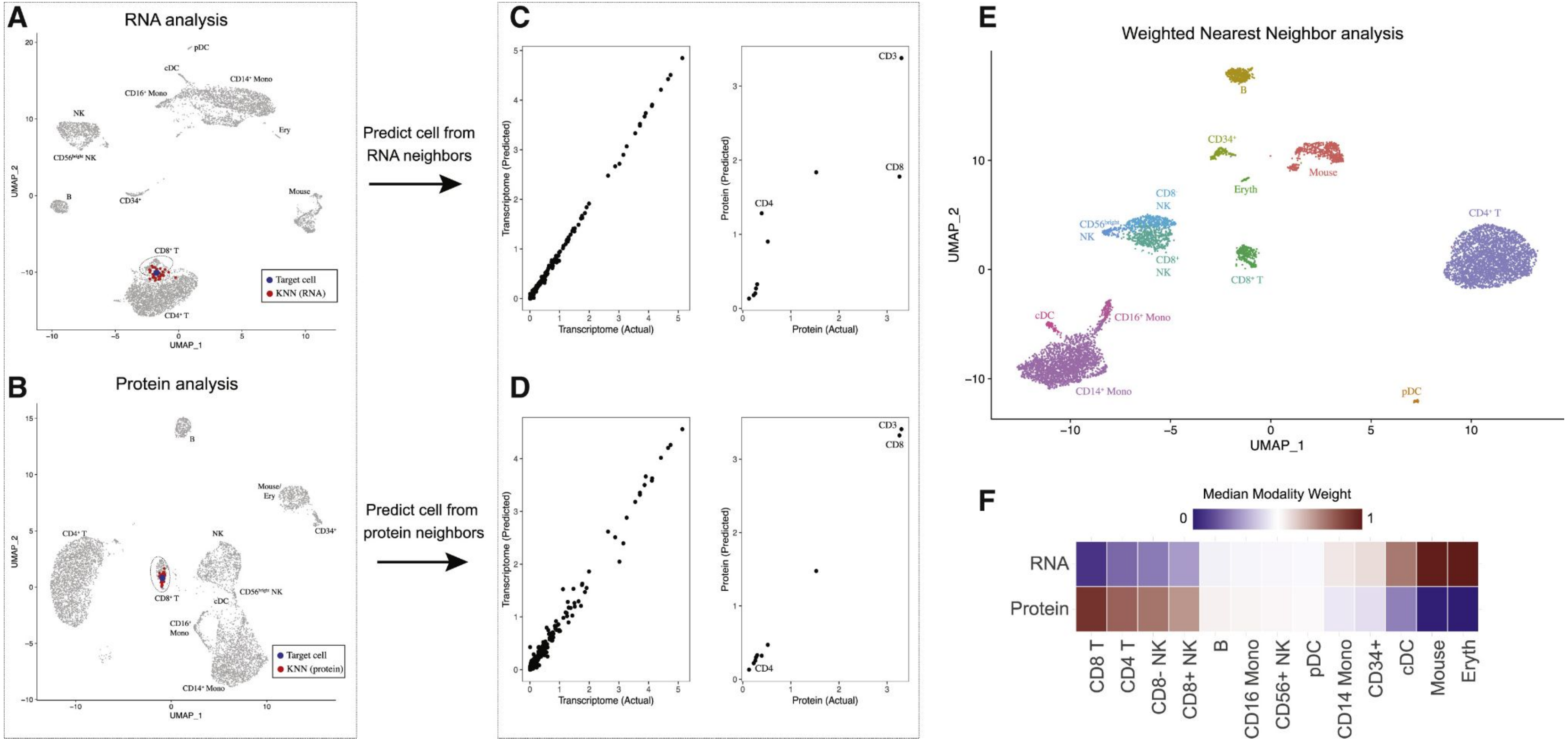
Stuart et al. 2021. Nature Methods
DOI: 10.1038/s41592-021-01282-5

Single-cell multimodal analysis steps



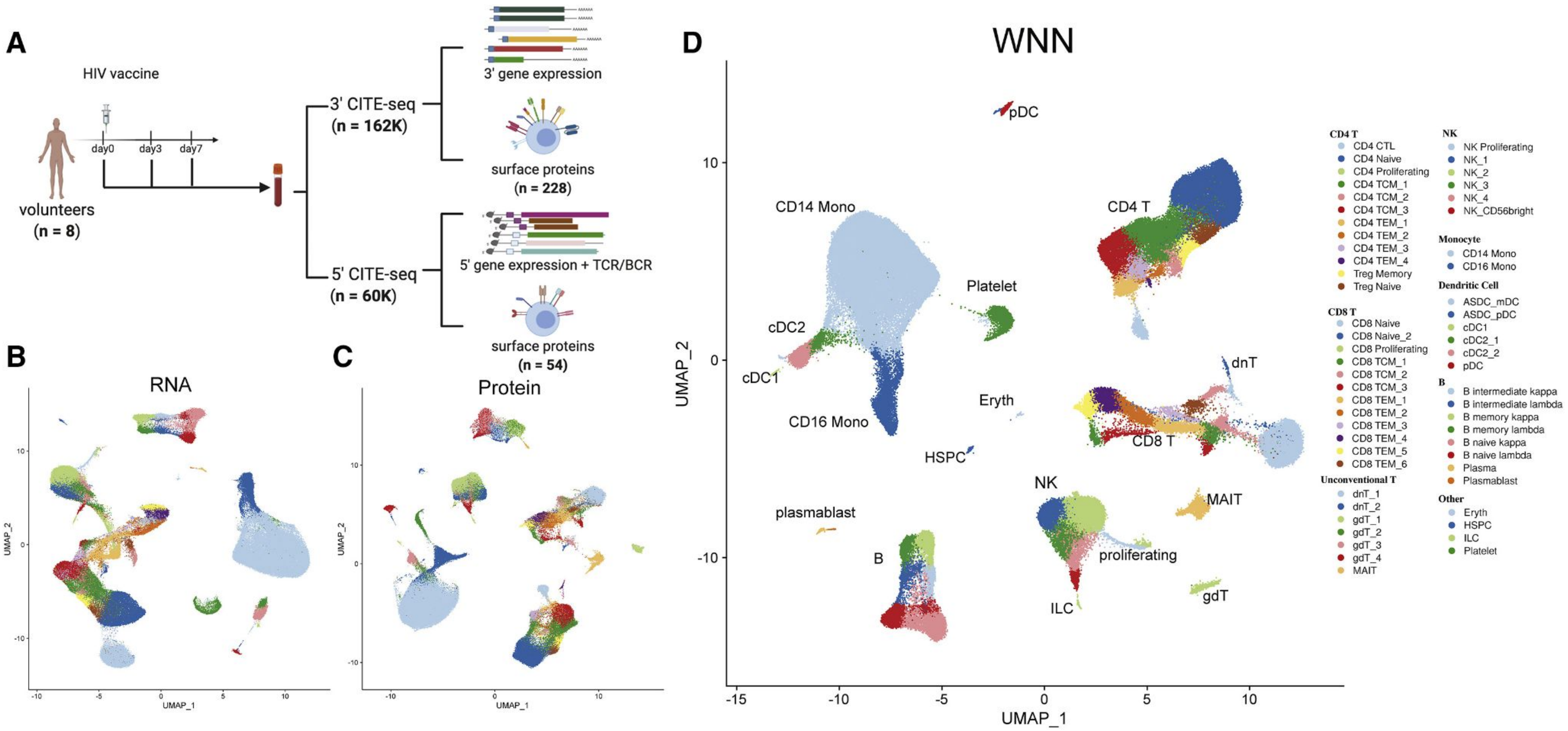
Stuart et al. 2021. Nature Methods
DOI: 10.1038/s41592-021-01282-5

Single-cell multimodal analysis Weighted nearest neighbors



Hao et al. 2021. Cell
DOI: 10.1016/j.cell.2021.04.048

Weighted nearest neighbors



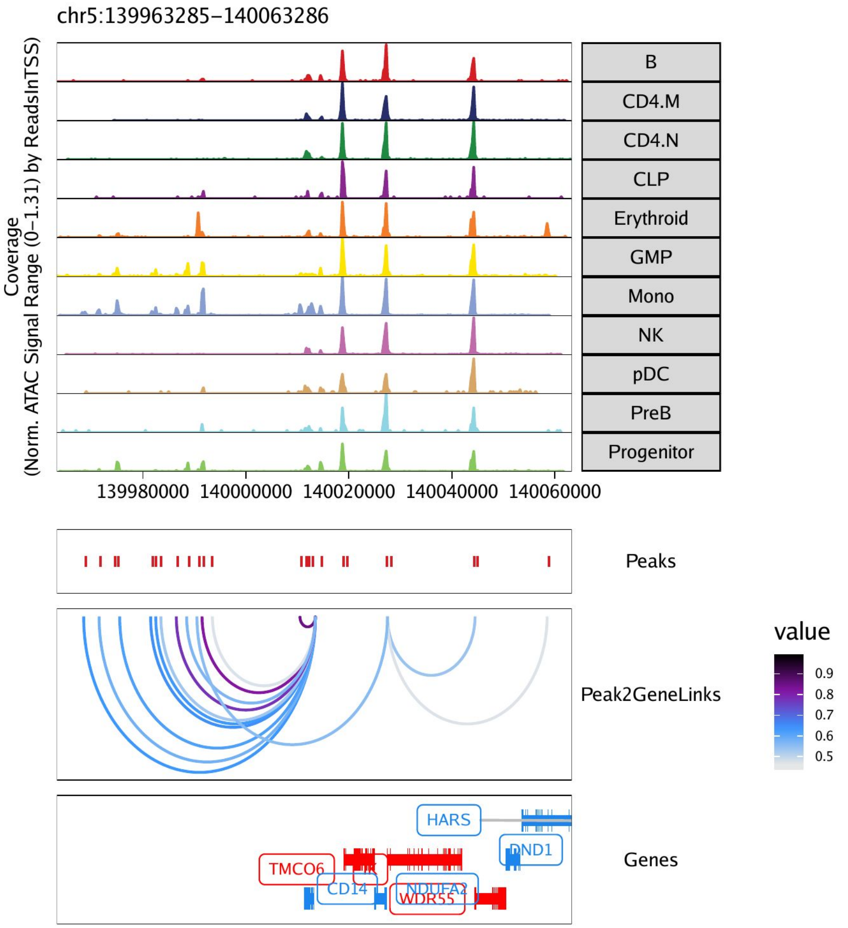
A multimodal atlas of human PBMC with CITE-seq data: 161,764 cells

Hao et al. 2021. Cell
DOI: 10.1016/j.cell.2021.04.048

Andres Quintero

Finding peak-to-gene links

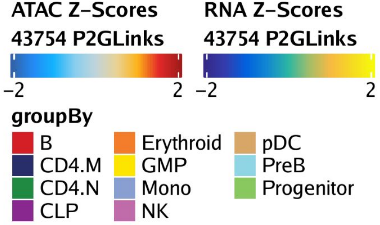
Peak-to-gene linkage leverages integrated scRNA-seq data to look for correlations between peak accessibility and gene expression.

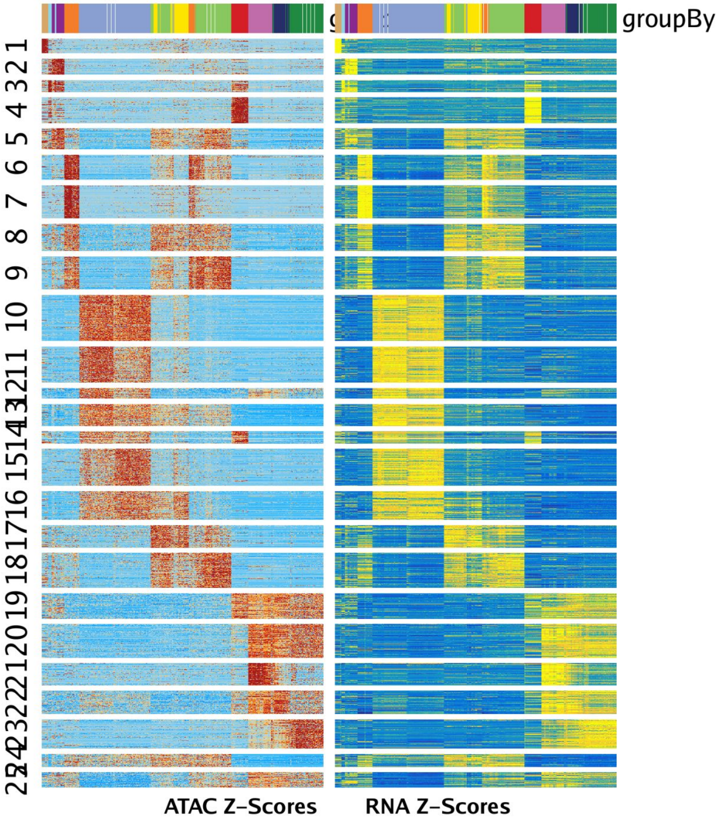


https://www.archrproject.com/bookdown/peak2genelinkage-with-archr.html

Finding peak-to-gene links

Peak-to-gene linkage leverages integrated scRNA-seq data to look for correlations between peak accessibility and gene expression.





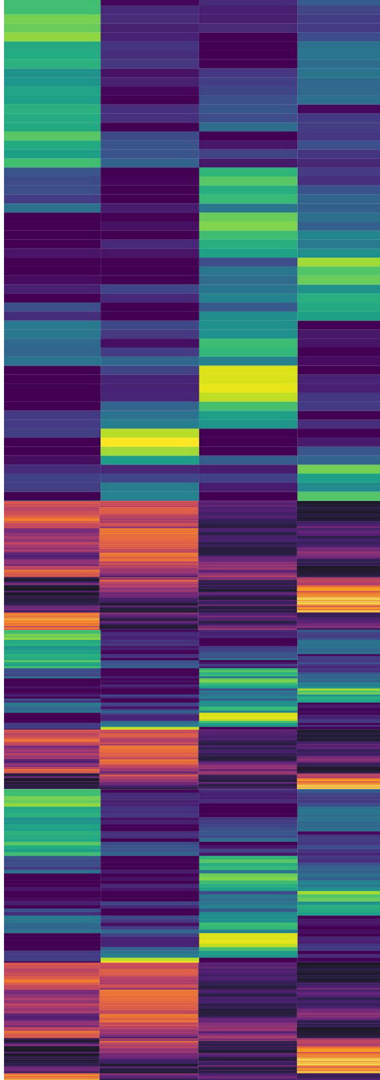
https://www.archrproject.com/bookdown/peak2genelinkage-with-archr.html

Single-cell multimodal analysis

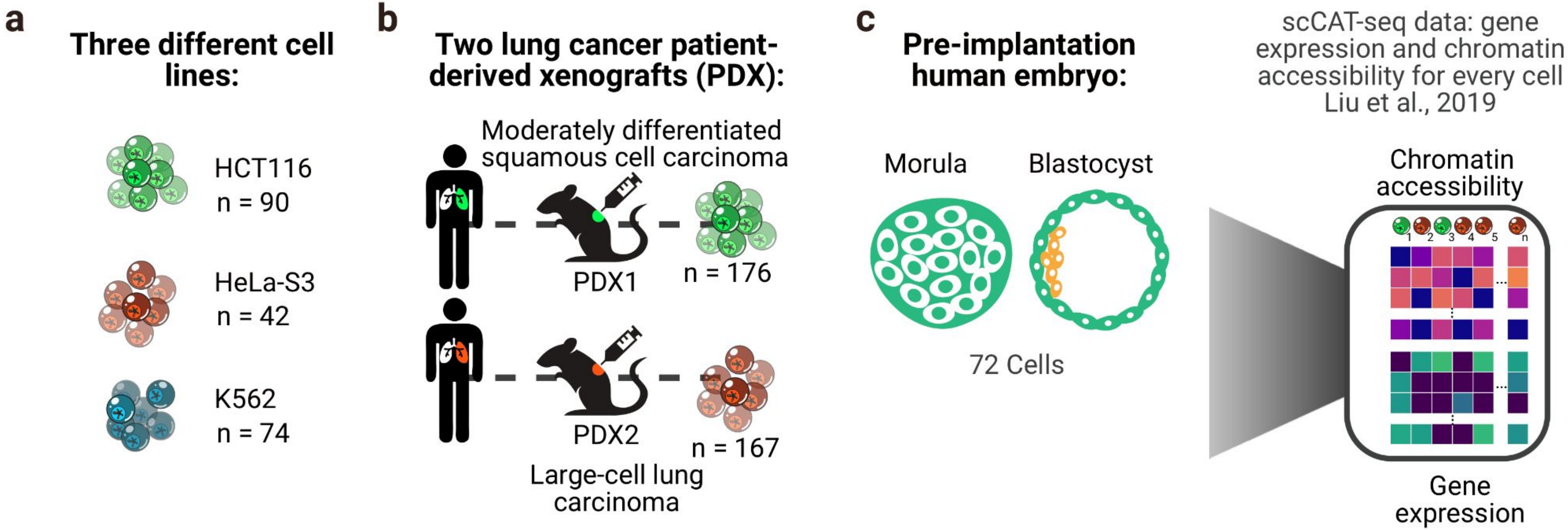
Example: Prediction of gene regulatory networks using GRANet

nature communications

Longqi Liu*, Chuanyu Liu*, Andrés Quintero*, Liang Wu*, Yue Yuan*, Mingyue Wang, Mengnan Cheng, Lizhi Leng, Liqin Xu, Guoyi Dong, Rui Li, Yang Liu, Xiaoyu Wei, Jiangshan Xu, Xiaowei Chen, Haorong Lu, Dongsheng Chen, Quanlei Wang, Qing Zhou, Xinxin Lin, Guibo Li, Shiping Liu, Qi Wang, Hongru Wang, J. Lynn Fink, Zhengliang Gao, Xin Liu, Yong Hou, Shida Zhu, Huanming Yang, Yunming Ye, Ge Lin, Fang Chen, Carl Herrmann, Roland Eils, Zhouchun Shang & Xun Xu. **Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity.** ***Nature Communications*** 10, 470 **(2019)**

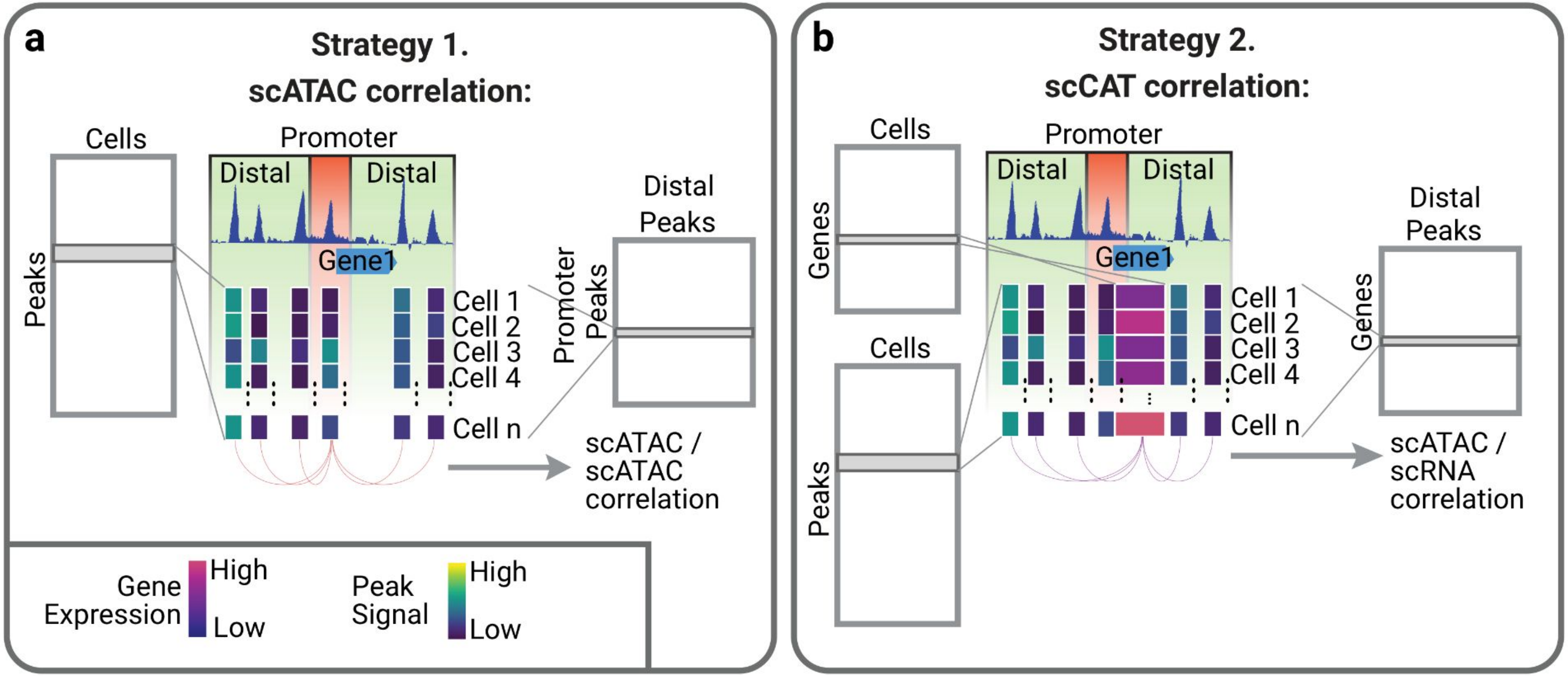


scCAT-seq multimodal data



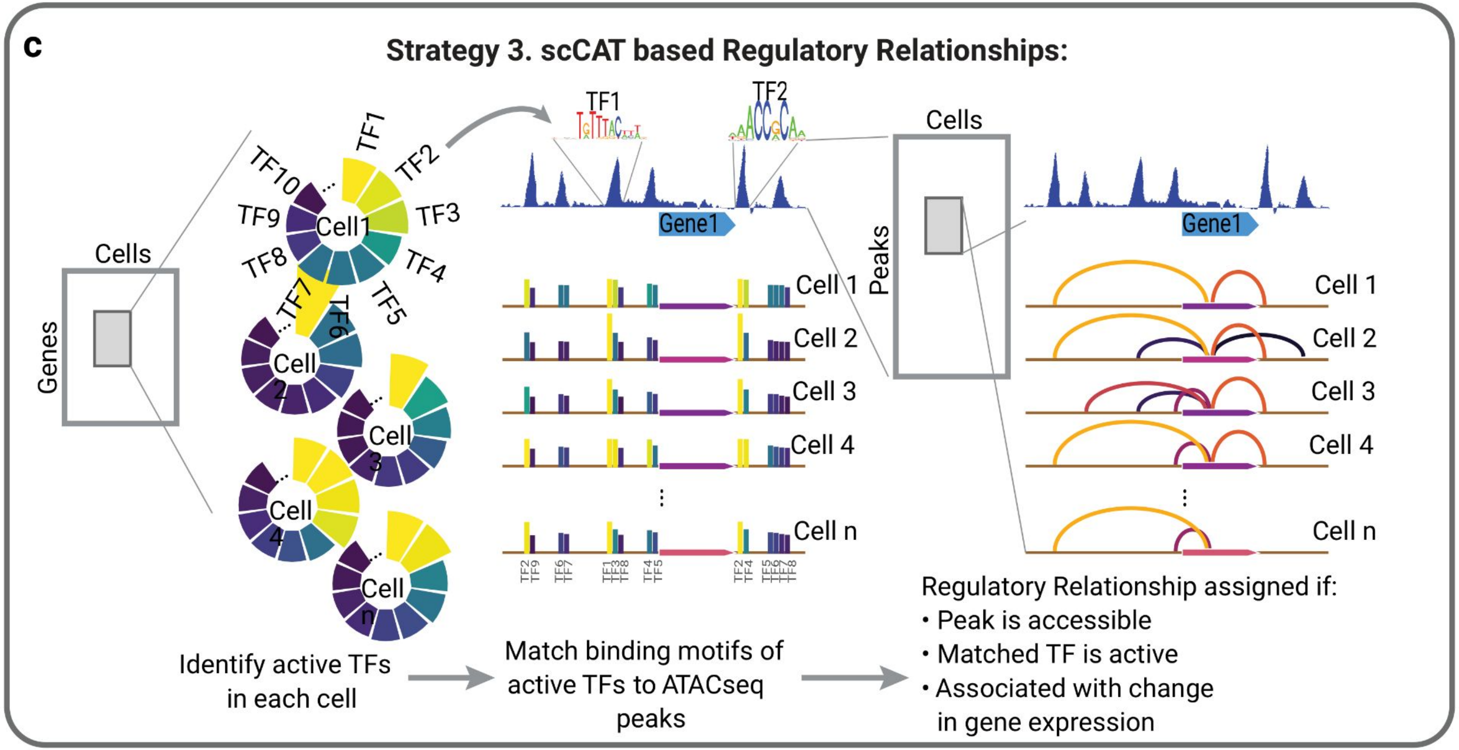
Liu et al. 2019. Nature Comm.
DOI: 10.1038/s41467-018-08205-7

Strategies to infer regulatory interactions



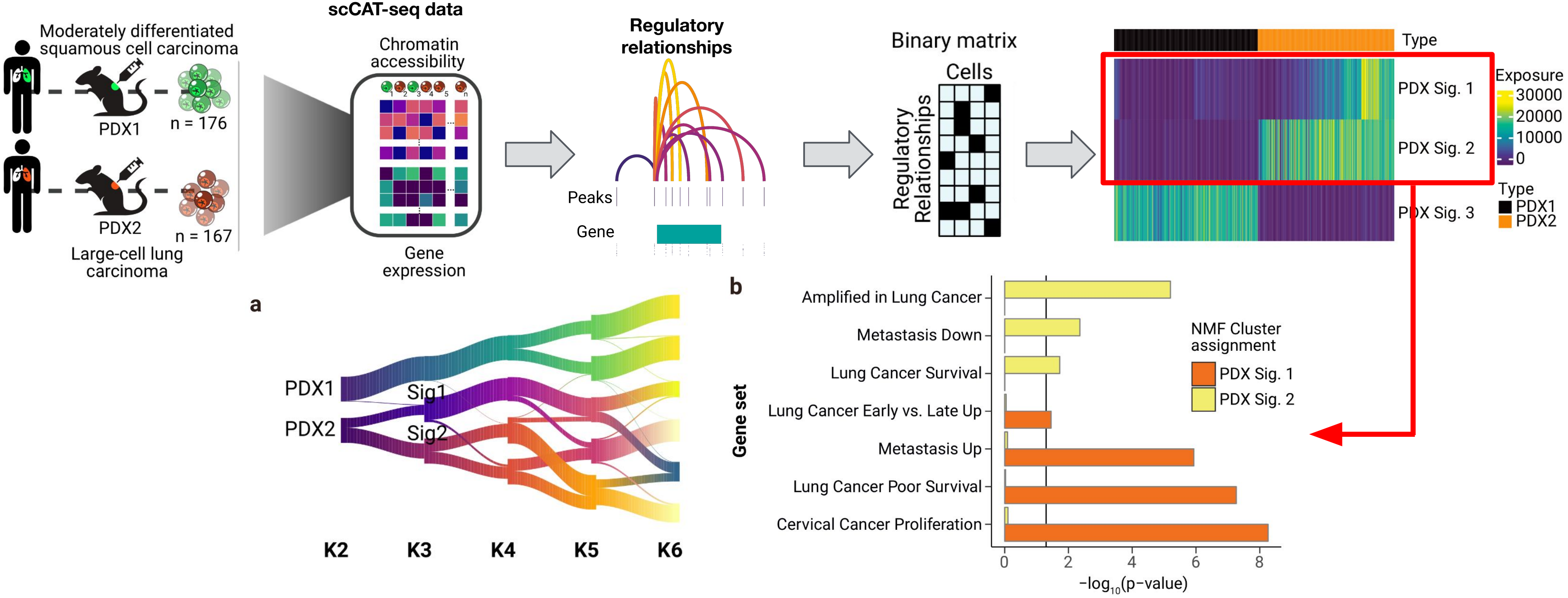
Liu et al. 2019. Nature Comm.
DOI: 10.1038/s41467-018-08205-7

Strategies to infer regulatory interactions



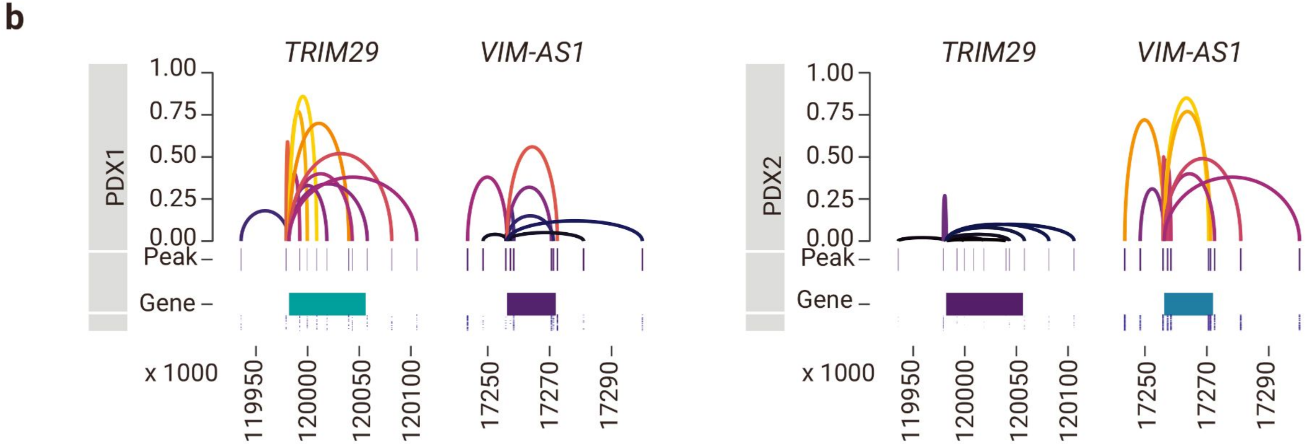
Liu et al. 2019. Nature Comm.
DOI: 10.1038/s41467-018-08205-7

Regulatory relationships from multiomics scCAT-seq data



Regulatory signatures captured the intra- and inter-tumor regulatory variability

Inter and intra-tumor regulatory variability



TRIM29: reported to mediate metastasis in lung squamous cell carcinoma via regulation of the autophagic degradation of E-cadherin (W. Xu et al. 2020).

VIM-AS1: It is a long noncoding RNA that promotes colorectal (Rezanejad bardaji, Asadi & Yaghoobi 2018) and prostate cancer (Z. Zhang et al. 2019) progression inducing EMT.

Liu et al. 2019. Nature Comm.
DOI: 10.1038/s41467-018-08205-7

Perspectives

ArchR

https://www.archrproject.com/

Signac

https://satijalab.org/signac/index.html

muon
multimodal omics
Python framework

https://github.com/scverse/muon

Cell Ranger ARC

https://support.10xgenomics.com/single-cell-multiome-atac-gex

Session 2 - Multi-omics integration

Hands-on sessions:

https://www.hdsu.org/sincellTE_2022/

RStudio server:

https://rstudio-singlecell.sb-roscoff.fr/

BioQuant
MODEL base of LIFE

healthdata scienceunit

HEIDELBERG
FACULTY OF
MEDICINE

Dr. Carl Herrmann
Dr. Carlos Ramirez
Dr. Andres Quintero
Daria Doncevic
Ana Luísa Costa
Youcheng Zhang

BioQuant
MODEL base of LIFE

healthdata
scienceunit

HEIDELBERG
FACULTY OF
MEDICINE

Thank you!