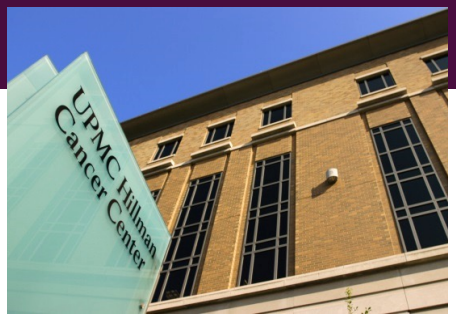


Computational Approaches for Identifying Context-Specific Transcription Factors using Single-Cell Multi-Omics Datasets

Hatice Ülkü Osmanbeyoğlu

Assistant Professor

Department of Biomedical Informatics

https://www.osmanbeyoglulab.com/

osmanbeyogluhu@pitt.edu

ISMB Tutorial July 9, 2024

UPMC | HILLMAN CANCER CENTER

NCI Designated Comprehensive Cancer Center

University of Pittsburgh

Who are we?

- Hatice Ulku Osmanbeyoglu, Assistant Professor, University of Pittsburgh, USA
- Linan Zhang, Assistant Professor, Ningbo University, China
- Parham Hadikhani, Postdoctoral fellow, University of Pittsburgh, USA
- Merve Sahin, Computational Biologist, Memorial Sloan Kettering Cancer Center, USA

Practical information

- Tutorial materials:
  https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/

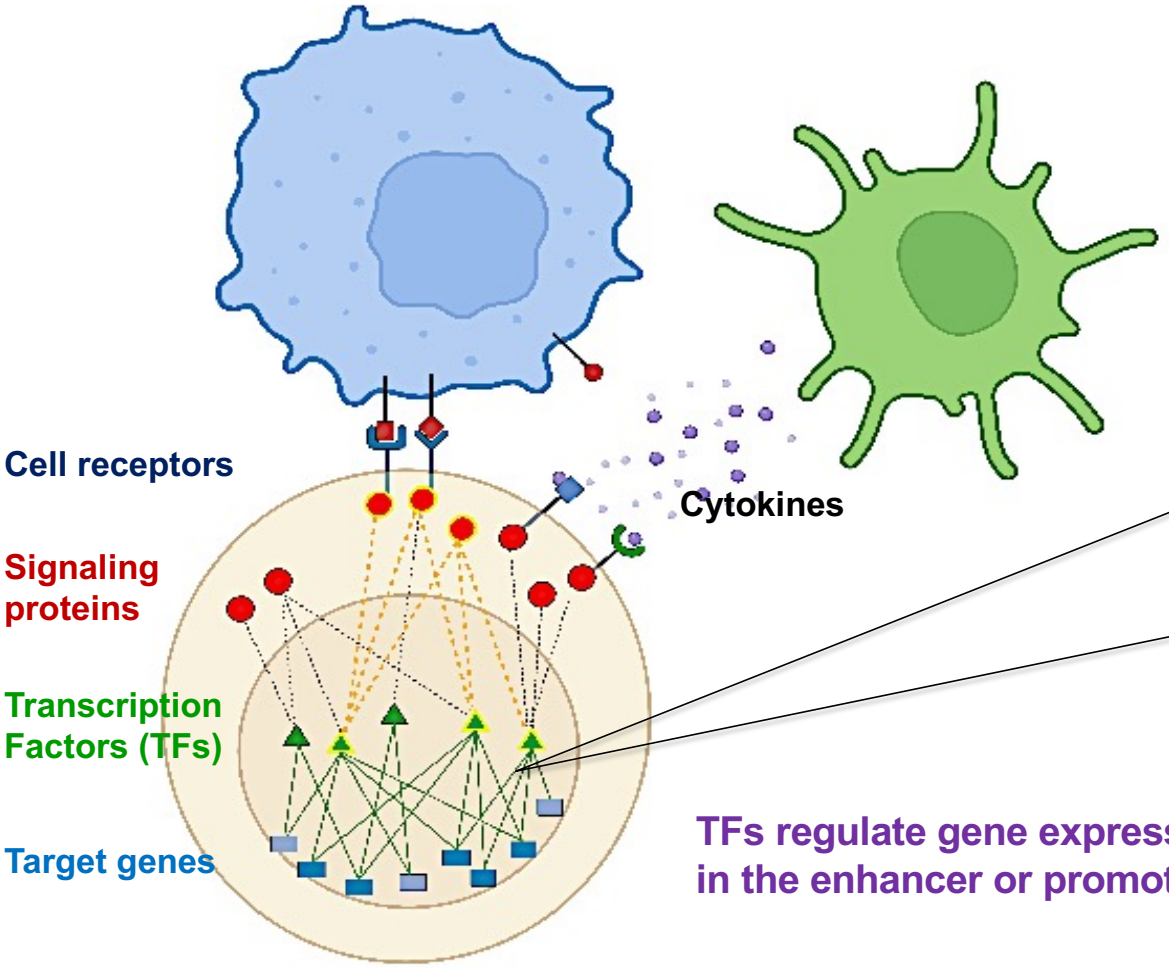
- ISMB feedback link

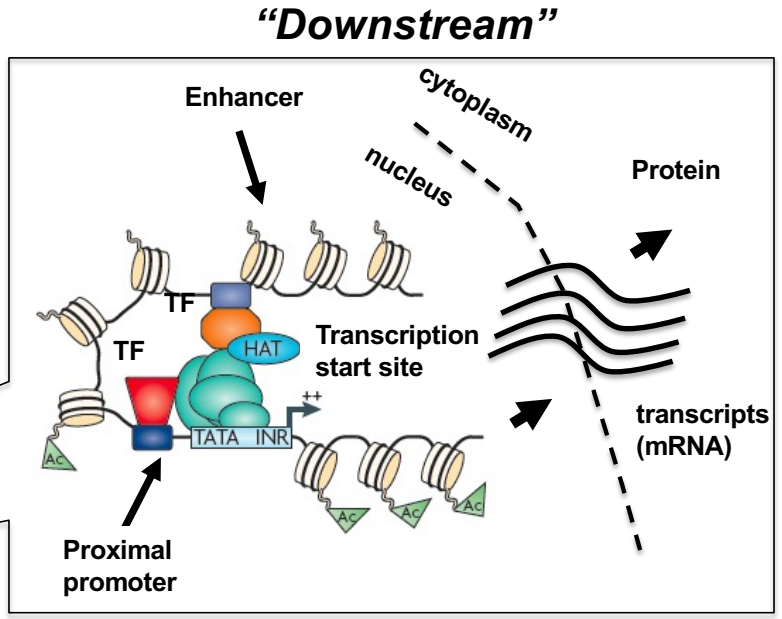
https://docs.google.com/forms/d/e/1FAIpQLSfrjbxbGdE45OFPQgM58JDt59B--gonwjuKw1HkHrR4FLvhhw/viewform

Outline

14:00	Welcome remarks and tutorial overview
14:05	Basic principles behind TF activity inference methods * Overview of the importance of context-specific TF regulation in biological systems * Significance of TF dynamics in health and disease * Single-cell multi-omics technologies for TF activity inference (scRNA-seq, spatial transcriptomics, scATAC-seq, Multiome, CITE-seq) Overview of computational TF inference methods based on single cell omics
15:45	Break
16:00	Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public scRNA-seq and spatial transcriptomics
16:45	Hands-on experience in applying tools and interpreting results using TF activity inference methods using public CITE-seq
17:30	Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public scATAC-seq and multiome
17:55	Discuss current bottlenecks, gaps in the field, and opportunities for future work

Transcription factors (TFs) are important modulators of cell types, fate and functional states



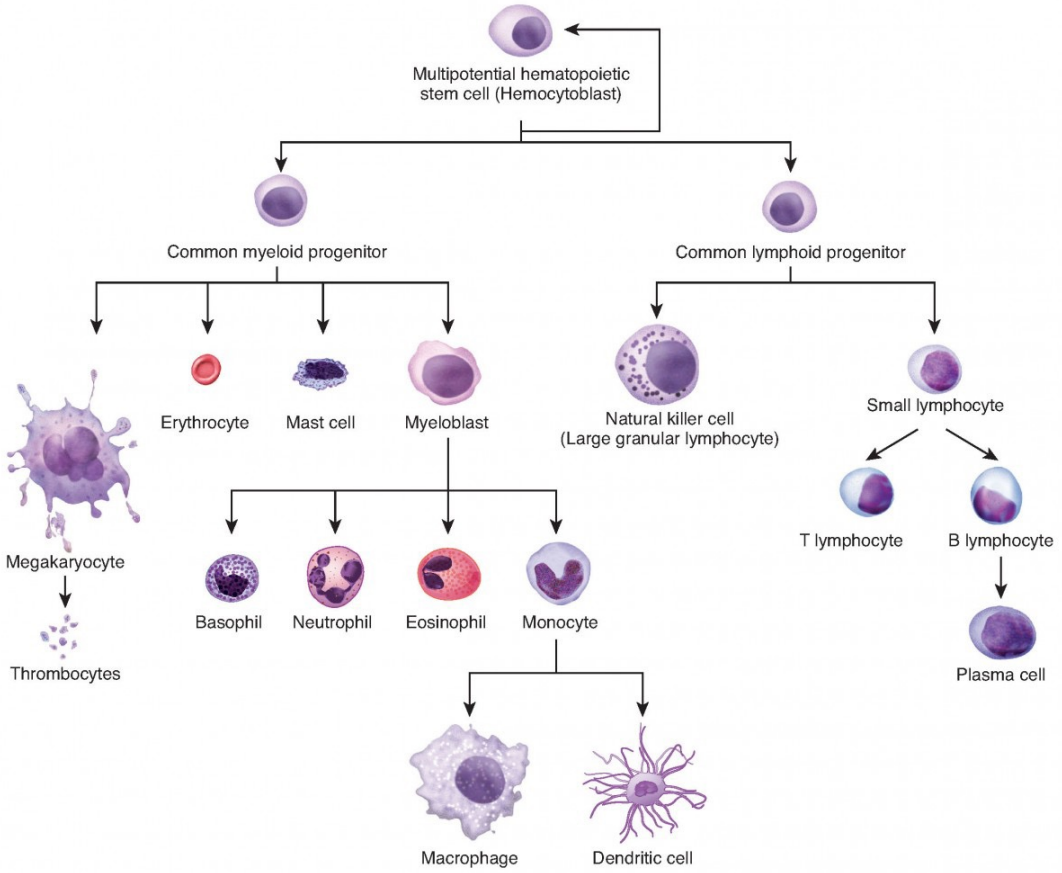


Farnham et al, *Nat Rev Genet.,* 2009

TFs regulate gene expression by binding specific DNA sequences in the enhancer or promoter regions of genes

Transcription factors activate/inhibit genes to affect cell transition from one state to another

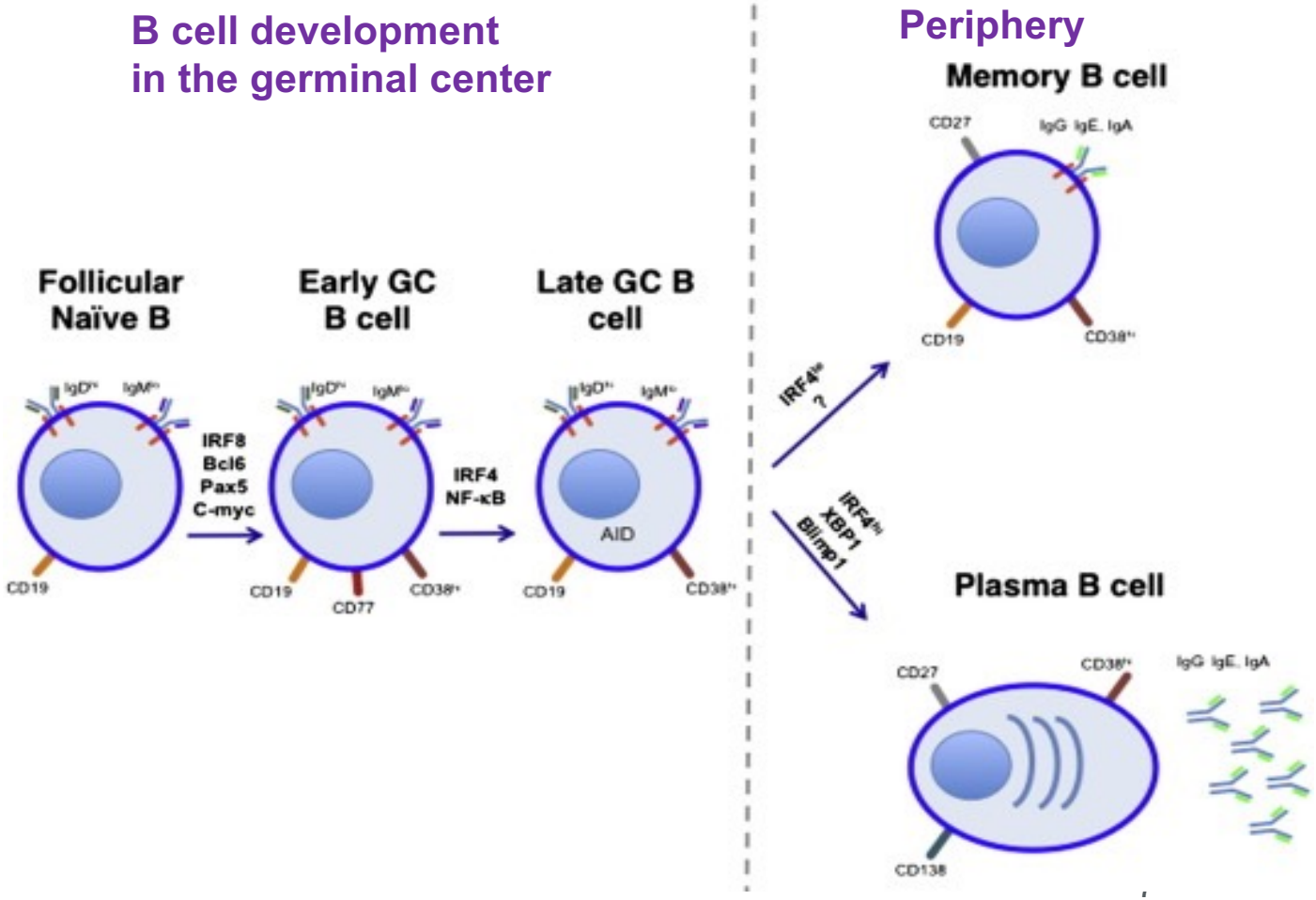
- **Hematopoiesis** involves the differentiation of multipotent cells into blood and immune cells.
- The multipotent hematopoietic stem cells give rise to many different cell types, including the cells of the immune system and red blood cells.
- Cells in different states express different sets of genes.
- Cells move from one “state” to another.



Barnes et al, *Clinical Immunology*, 2014

Example: TFs regulating B cell functional states

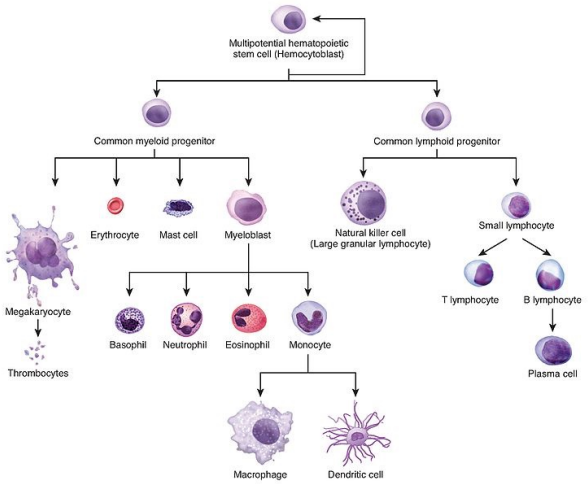
- B cells are the immune system's secret agents, producing antibodies and keeping a record of past enemies to protect body from future threats
- The interplay between signaling proteins, transcription factors and their downstream targets allows B cells to adapt and transition between various functional states (e.g. memory, plasma B cells) depending on the specific immune challenge and context

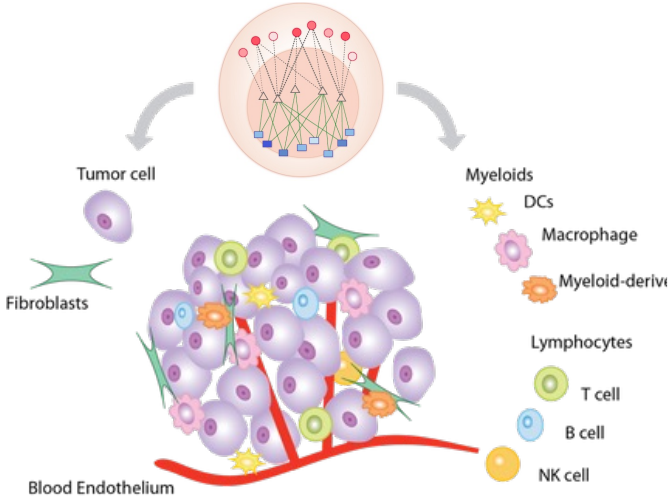


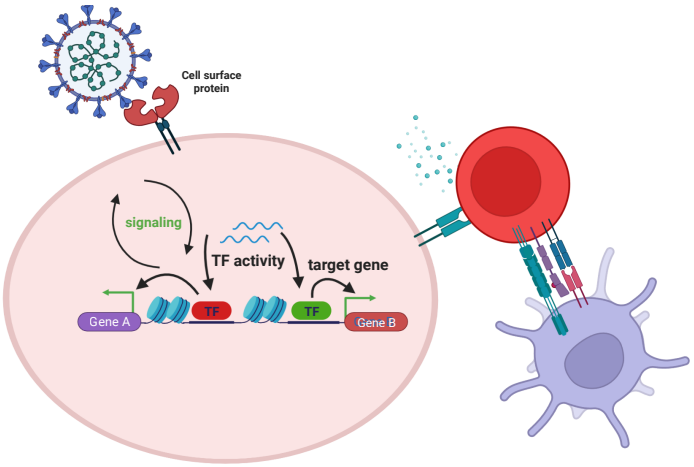
Barnes et al, *Clinical Immunology*, 2014

Significance of TF dynamics in health and disease

- What are TFs underlying cellular identities (e.g. naïve vs memory T cells)?
- What are different or common TFs given a cell type across tissues (e.g. T cells in blood vs lung)?
- What are commonalties as well as differences of cell-type/state specific TFs across healthy individuals and those manifesting a disease?





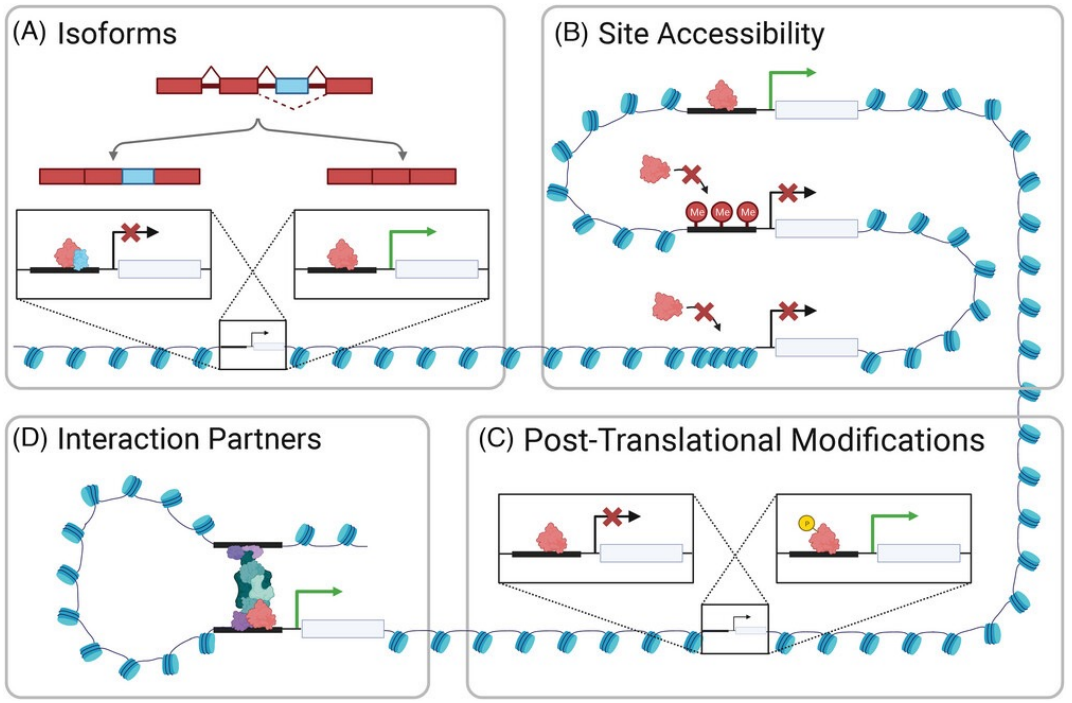


https://commons.wikimedia.org/wiki/File:0337_Hematopoiesis_new.jpg

The complexity of TF activity inference

- TF activity (TFA) defines the regulatory impact exerted by a transcription factor on its target genes.
  - Examples include activation, repression, and effects like alternative splicing

- TFA can be influenced by: epigenetic modifications, post-transcriptional regulation, post-translational modifications, protein–protein interactions, presence of cofactors localization, DNA structural changes



Proteomics, Volume: 23, Issue: 23-24, First published: 14 September 2023, DOI: (10.1002/pmic.202200462)

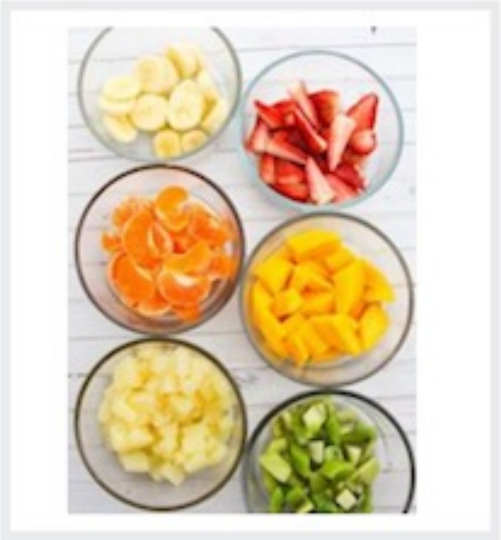
Due to the complexity and cost of experimental approaches, computational tools are essential.

Single cell and spatial omics allow the investigation of Identifying transcription factors driving cell types, fate and functional states

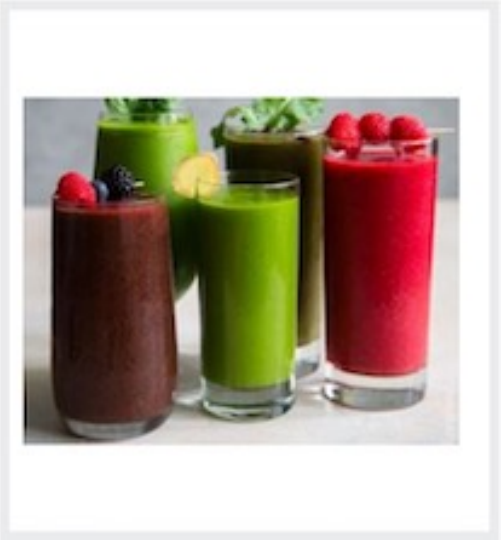
Technological advance: Bulk vs single-cell vs spatial resolution

Spatial analysis provides molecular information with spatial organization

Single cell analysis retain single cell information but lose spatial organization

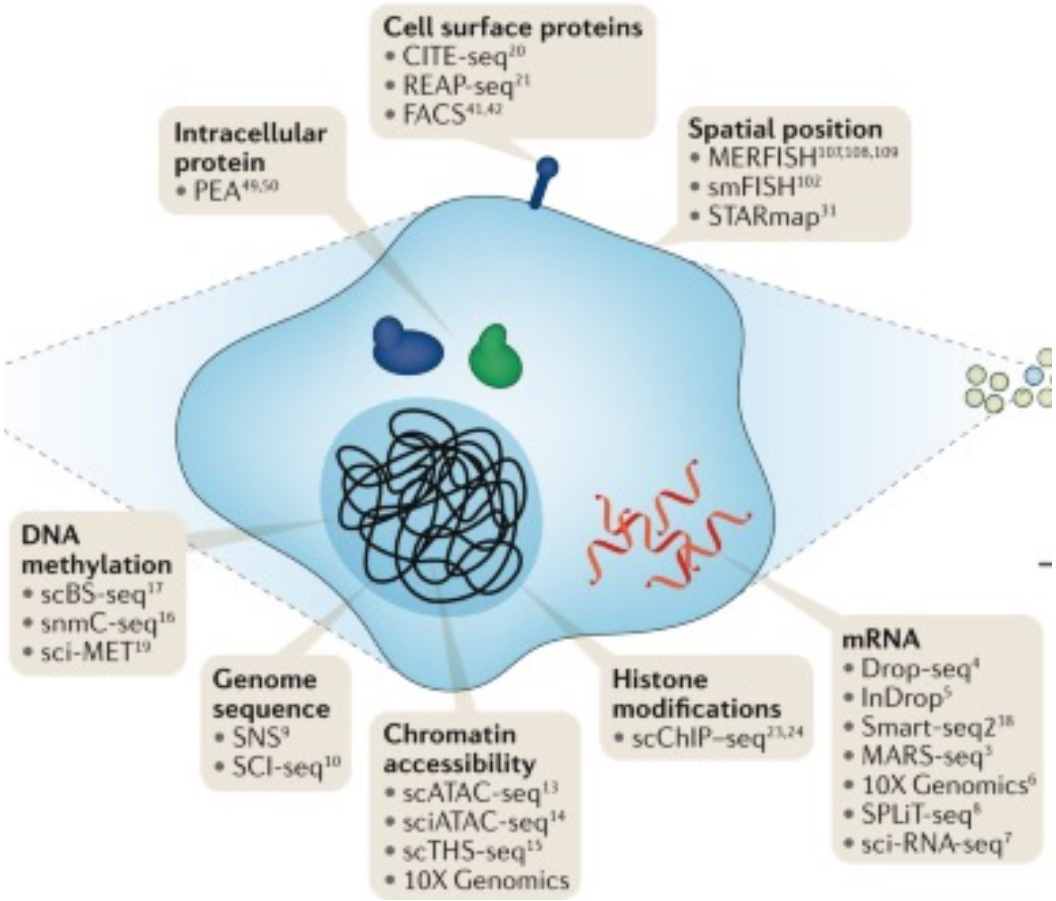


Bulk ("population of cells") analysis lose spatial and single cell resolution



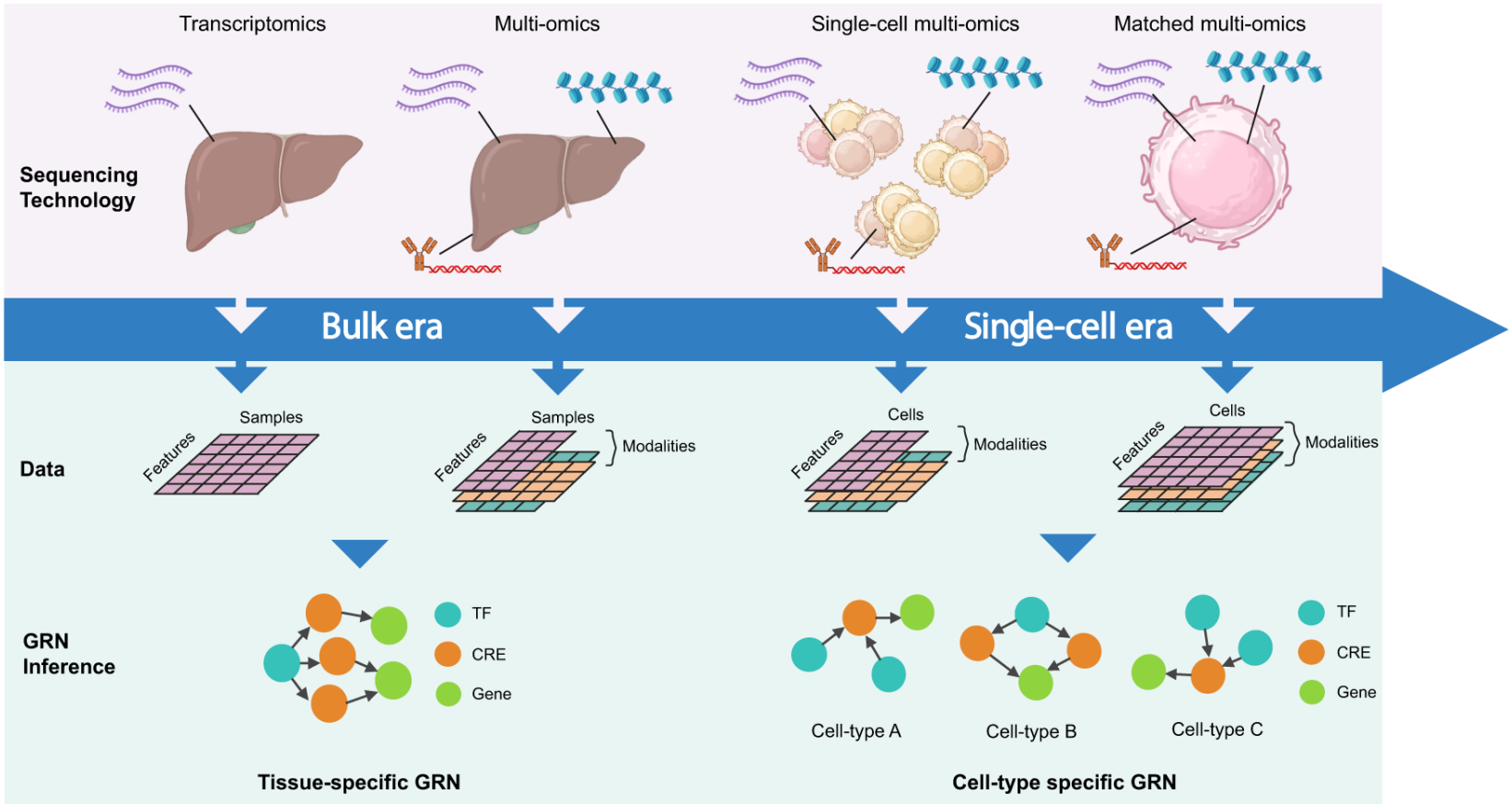
https://biotech.illinois.edu/htdna/applications

What can we study at single-cell/spatial level?



Stuart et al..(2019) “Integrative single-cell...” Nat Rev Genet 20, 257–272.

How do we infer activity of TFs using omic data-driven computational techniques?

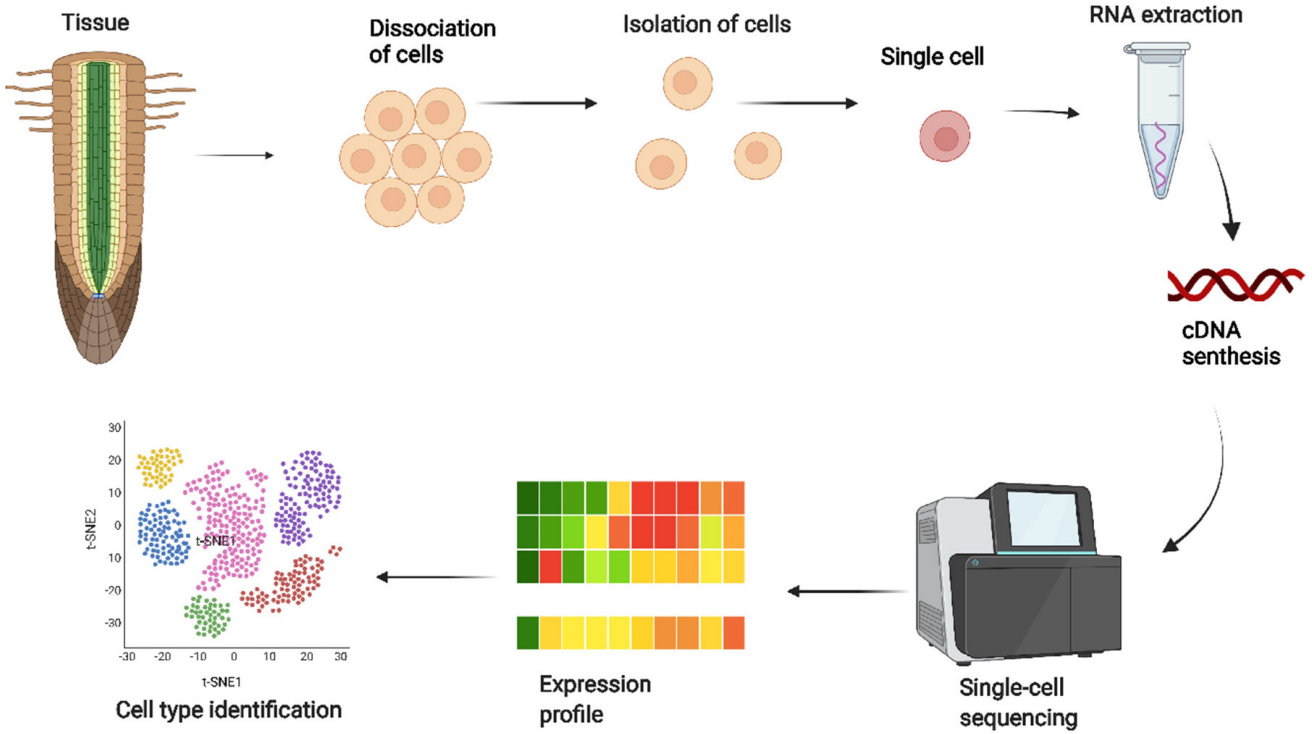


Kim, Daniel et al. *NPJ systems biology and applications,* 2023

Single-cell multi-omics technologies for TF activity inference

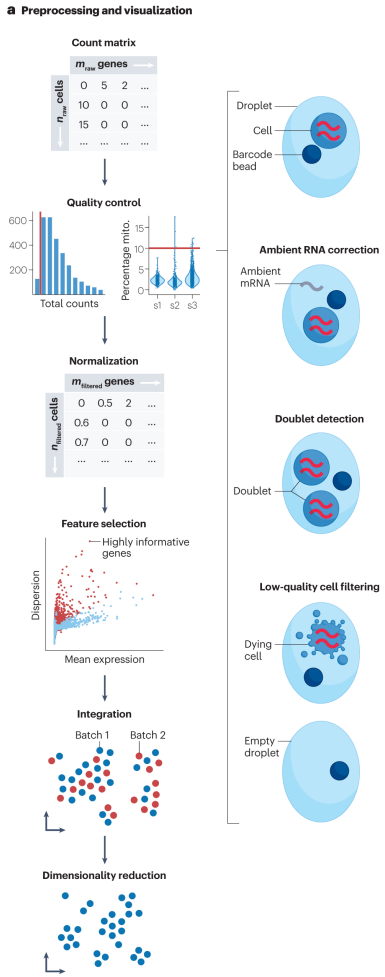
Single-cell gene expression

- Generate numerous independent gene expression measurements
- Full length: Fluidigm C1, Smart-seq; Single end: Drop-seq, 10X Genomics, Microwell, SPLIT-seq
- **Allow for comprehensive analysis of TF and their target gene expression profiles across individual cells within a heterogeneous population**



Bawa, G et al. . *Int. J. Mol. Sci.* 2022, *23*, 4497.

Single cell gene expression data analysis (1)



Data Processing and Quality Control:

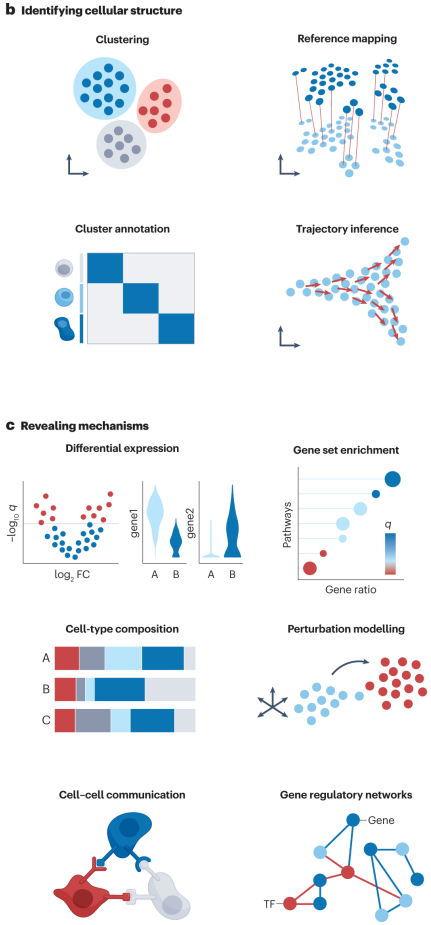
- Obtain count matrices from raw data.
- Correct for cell-free RNA and filter for doublets, low-quality, or dying cells.
- Apply quality control metrics (count depth, genes per barcode, % mito.) and normalize data to ensure accurate gene abundance comparisons.

Analysis and Visualization:

- Select most variably expressed genes from datasets up to 30,000 genes.
- Integrate data batches and employ dimensionality reduction techniques for visualization.

Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).

Single cell gene expression data analysis (2)



Cluster Analysis and Annotation:

- Data is clustered based on gene expression profiles.
- Clusters are annotated with cell type labels manually or automatically.
- Continuous processes like cell differentiation transitions are inferred.

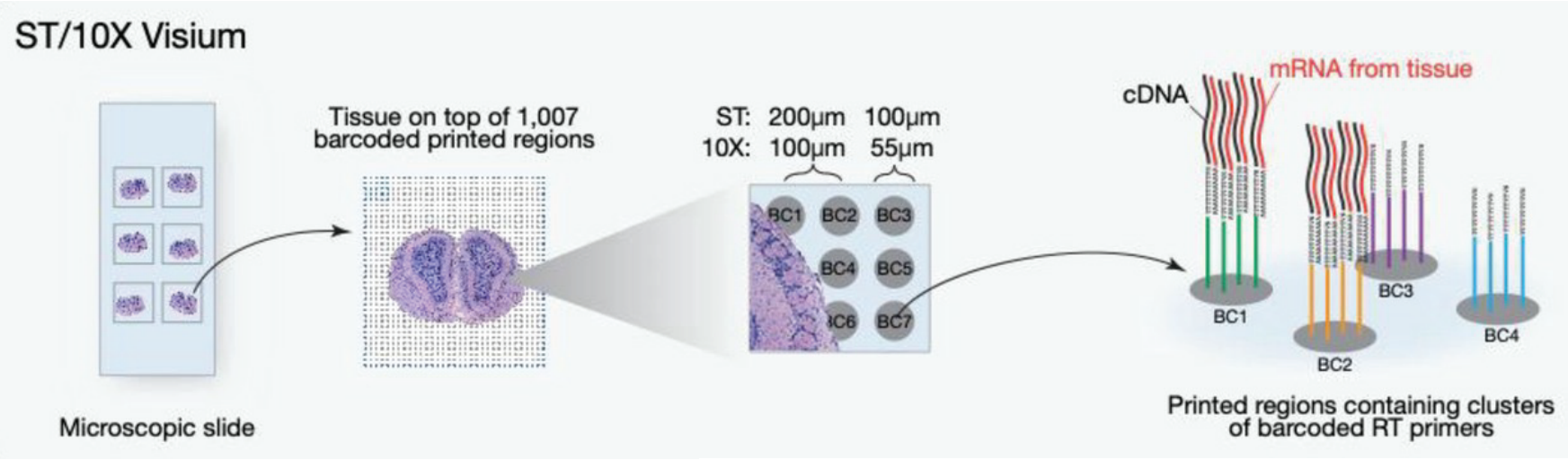
Interpretation and Analysis of scRNA-seq Data:

- Differential expression analysis identifies upregulated/downregulated genes.
- Gene set enrichment analysis assesses pathway effects.
- Changes in cell-type composition are examined.
- Perturbation modeling predicts effects of induced or unmeasured perturbations.
- Analysis of ligand-receptor expression reveals altered cell-cell communication.
- Gene regulatory networks are inferred from transcriptomics data.

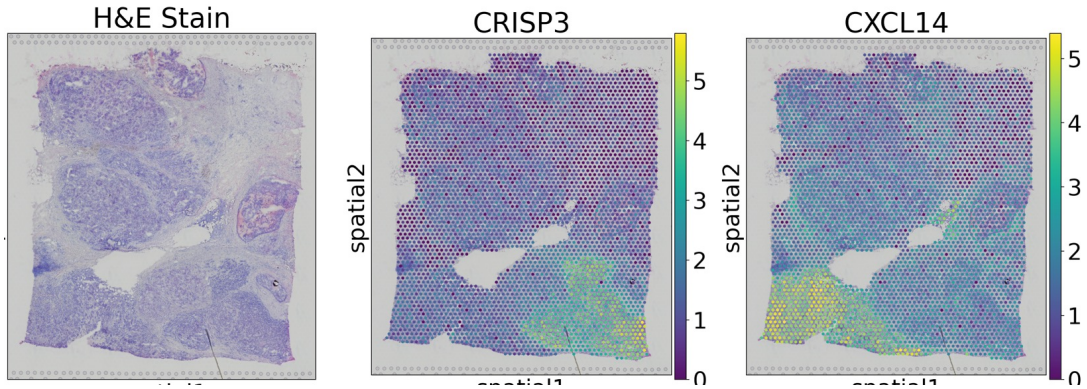
Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).

Spatially resolved transcriptomics (genome-wide) data

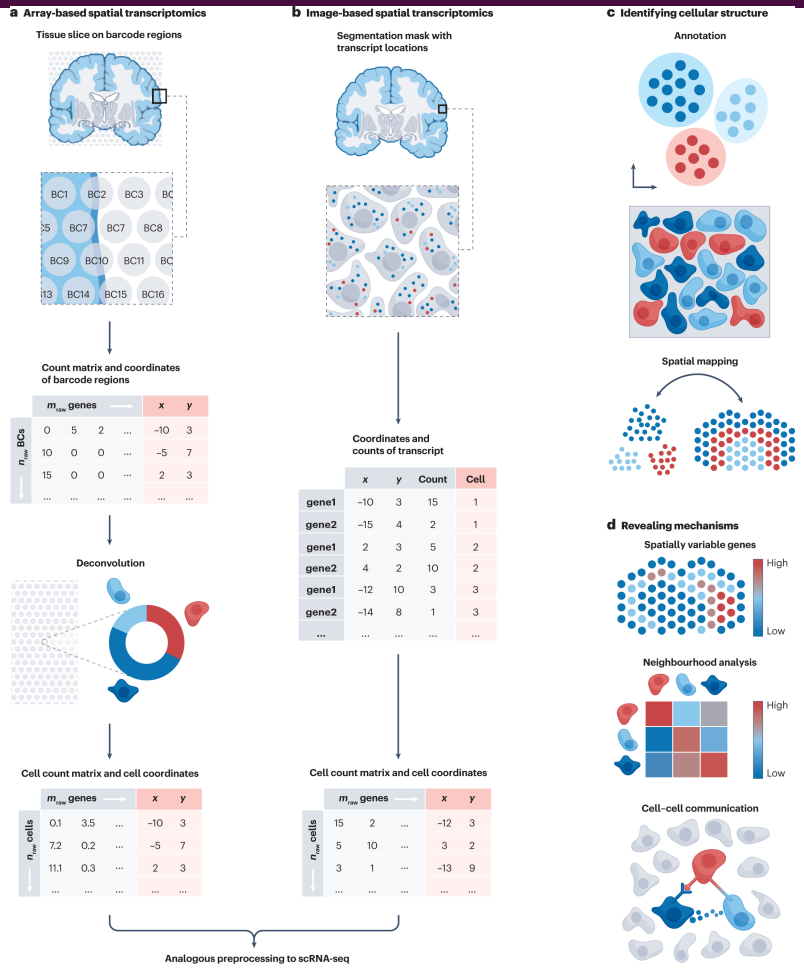
- Spatial transcriptomics measures mRNA across a tissue slice while preserving regional information



- 10x Genomics Visium Protocol
  - Samples 4078 ‘spots’ (55µm, 1-10 cells/spot)
  - Each spot is assigned a barcode
  - Sparser than scRNA-seq data
  - Not single cell resolution



Spatial gene expression data analysis (1)



Array-based Spatial Transcriptomics:

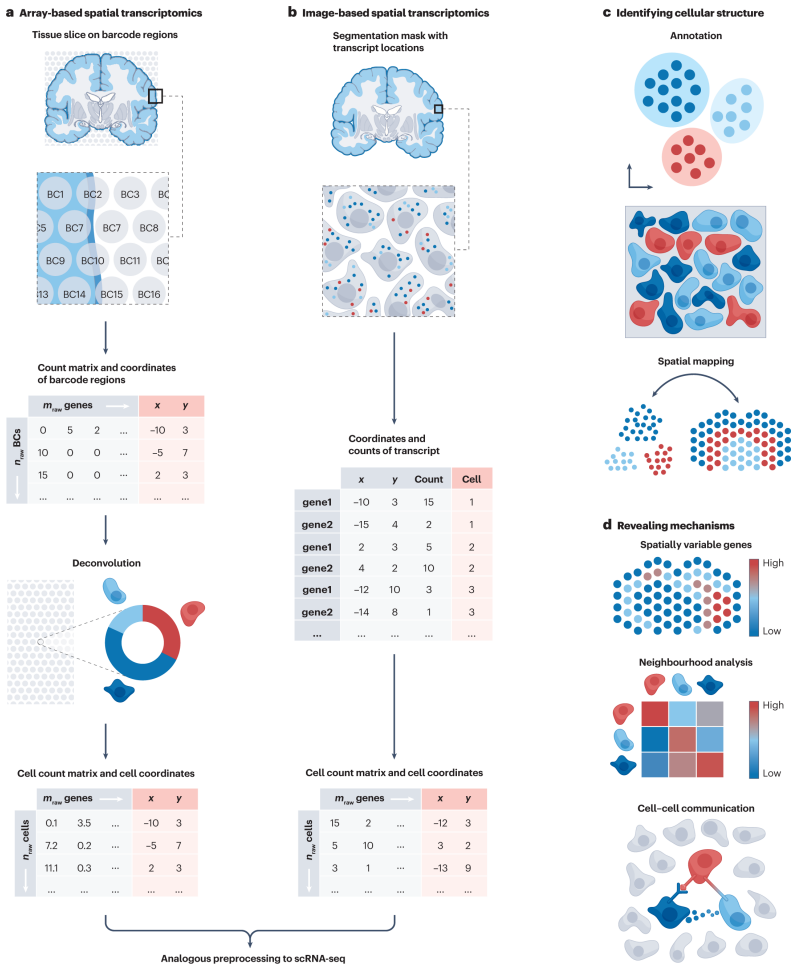
- Quantifies gene expression in predefined barcoded (BC) regions spanning 10 µm to 200 µm.
- BC regions aggregate measurements from multiple cells into count matrices with spatial coordinates.
- Cell-type deconvolution methods decompose BC regions to obtain single-cell count matrices and spatial coordinates.
- Preprocessing parallels analysis of scRNA-seq datasets.

Image-based Spatial Transcriptomics:

- Technologies like FISH and ISS capture transcript locations through sequential hybridization rounds.
- Aggregate transcript locations into count matrices and spatial coordinates at the single-cell level.
- Subsequent processing aligns with scRNA-seq methodologies.

Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).

Spatial gene expression data analysis (2)



Resolution and Spatial Mapping:

- Identifies cellular structures at single-cell or BC region resolution.
- Addresses limitations of small feature space in image-based methods through spatial mapping, imputing unmeasured transcripts onto coordinates.

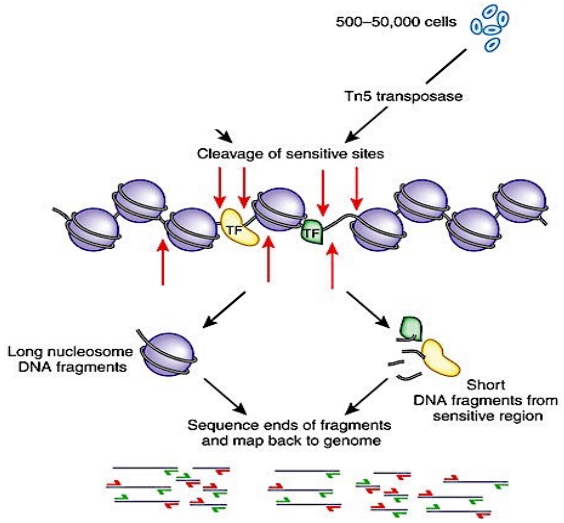
Analysis and Mechanisms:

- Analyzes mechanisms via spatial positions, identifying spatially varying genes.
- Examines cell neighborhoods and infers communication events (receptors, ligands, tight junctions, mechanical and indirect mechanisms).

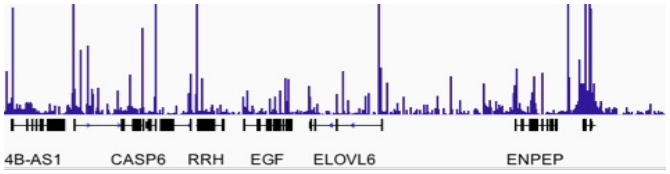
Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).

Single cell chromatin accessibility

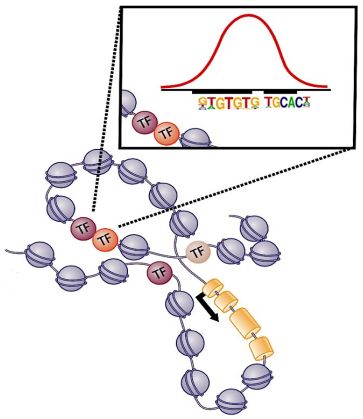
- ATAC: Assay for Transposon-Accessible Chromatin
- scATAC-seq, sci-ATAC-seq, sc-THS-seq
- Map regions of open chromatin = held open by DNA-binding proteins, like TFs



ATAC-seq peaks ↓

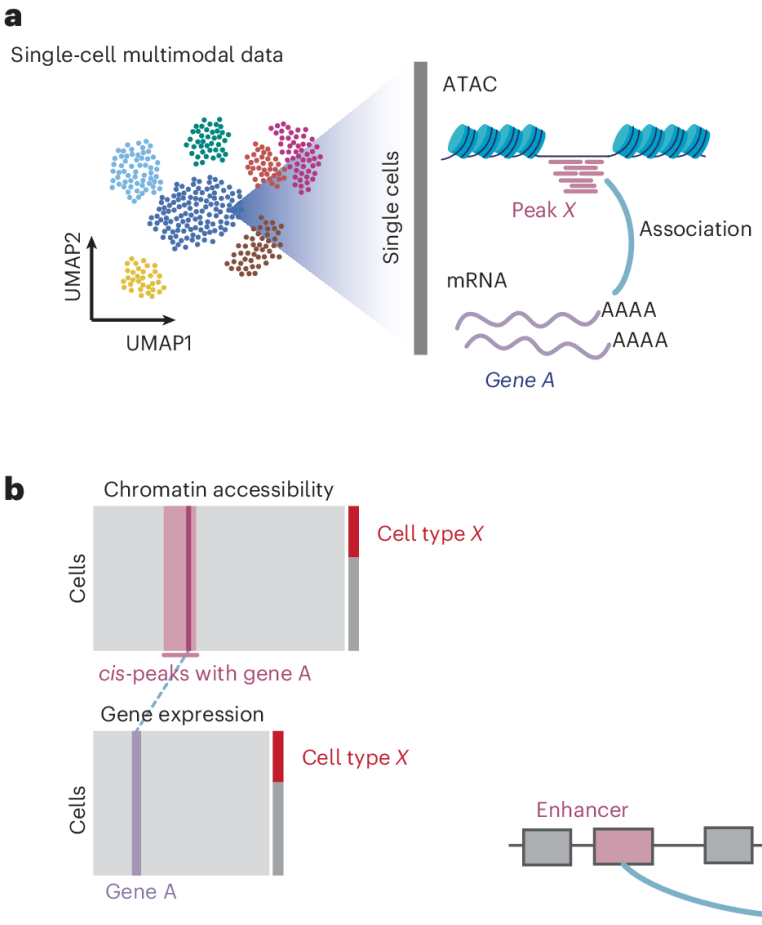


TF motif prediction in ATAC-seq peak regions



An open TF motif does not necessarily represent a binding event, and the same motifs can be shared by many different TFs

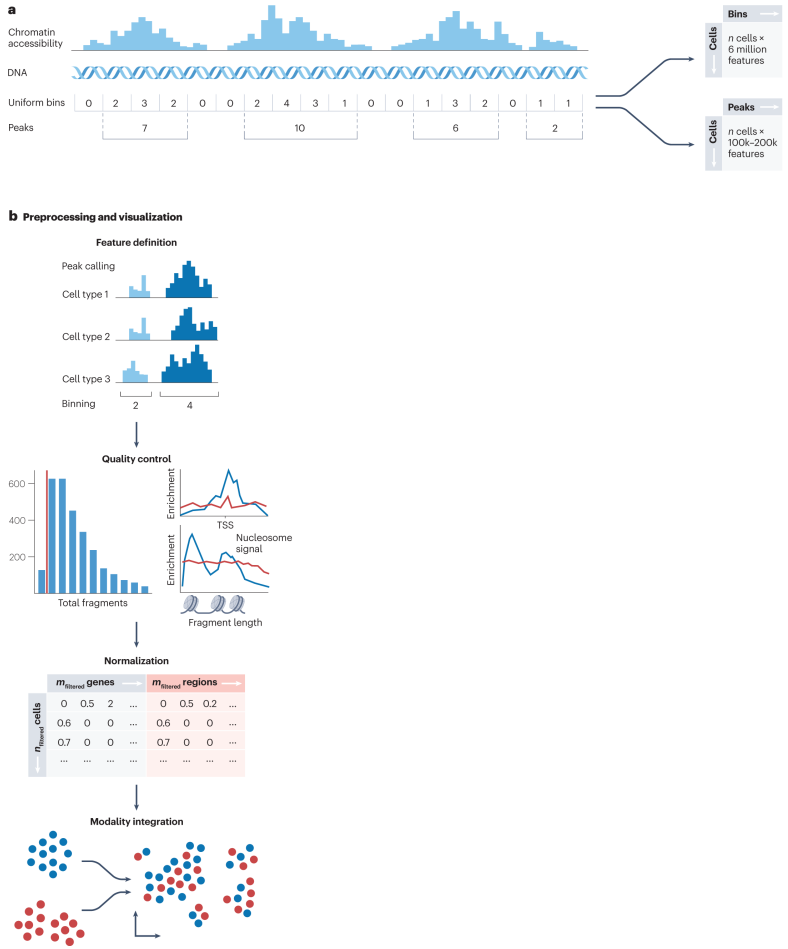
Single-Cell Multiome



- Combines single-cell gene expression analysis with single-cell open chromatin mapping to provide genome-wide mapping of both the transcriptional and epigenetic landscapes at the single-cell level

Sakaue, S., Weinand, K., Isaac, S. *et al.* Tissue-specific enhancer–gene maps from multimodal single-cell data identify causal disease alleles. *Nat Genet* 56, 615–626 (2024).

scATAC-seq analysis steps (1)

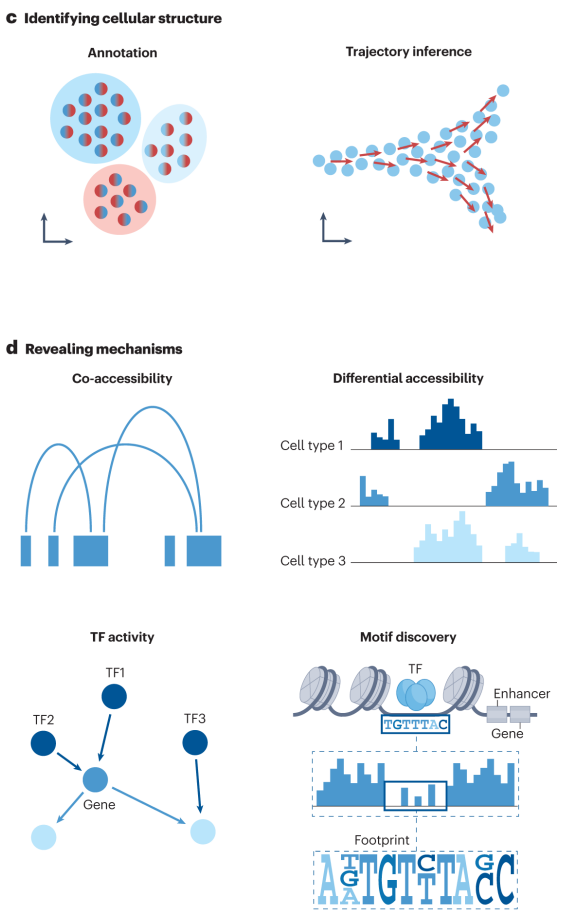


Quality Control and Preprocessing:

- Data represented as cell-by-peak or cell-by-bin matrices.
  - Peak-calling identifies regions of high accessibility
  - Binning captures Tn5 transposition events in equally sized bins.
- Quality control includes assessing cellular sequencing depth (total fragments per cell), non-zero peak counts, TSS enrichment score, nucleosome signal, and artifact signals.
- Normalize sparsely distributed scATAC-seq features.

Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).

scATAC-seq analysis steps (2)



Annotation, Cell Type Identification ana Data Analysis:

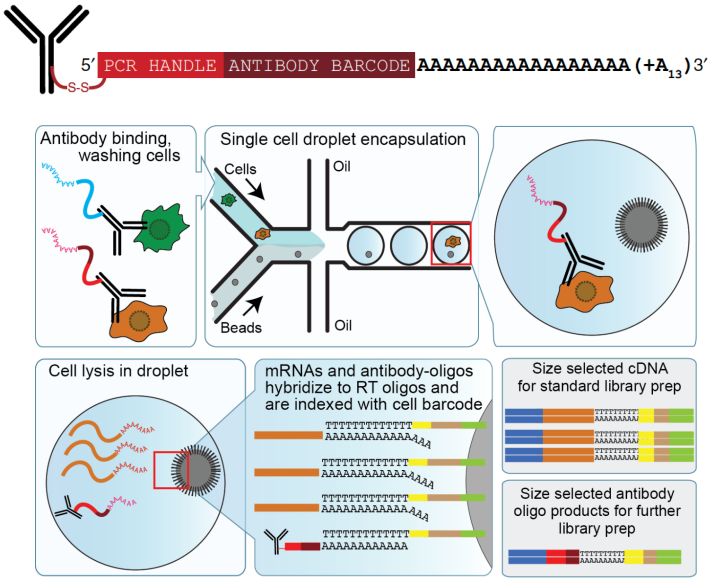
- Annotate scATAC-seq data with cell types based on differentially accessible regions.
- Use annotated cells for trajectory inference to analyze continuous processes.
- Investigate co-accessibility to identify cis-regulatory interactions.
- Identify differentially accessible regions to understand changes between conditions.
- Assess transcription factor (TF) activity and discover DNA sequence motifs for TF binding sites.

Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).

Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq)

CITE-seq: assay that can quantify protein and mRNA levels simultaneously, at the single-cell level

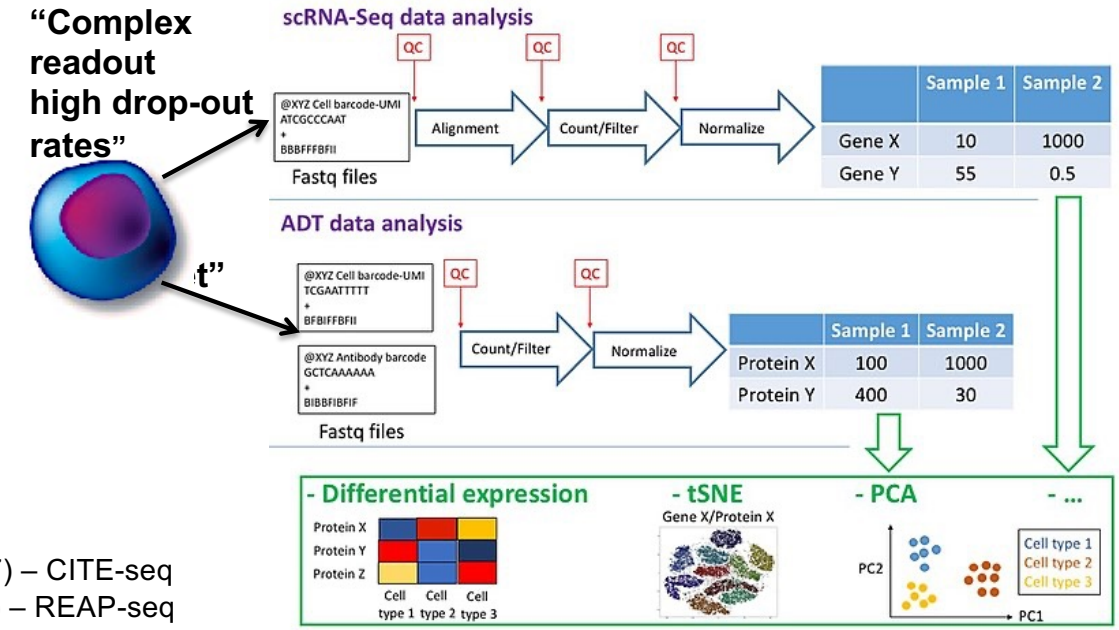
- Uses DNA-barcoded antibodies with polyA tail to convert detection of proteins into a quantitative, sequenceable readout
- Antibody-derived tags (ADTs) are antibody clones with unique barcodes attached to poly(A) sequences and a PCR handle.
- They bind to surface proteins, and their sequenced counts indicate protein expression levels.
- More comprehensive than scRNA-seq data



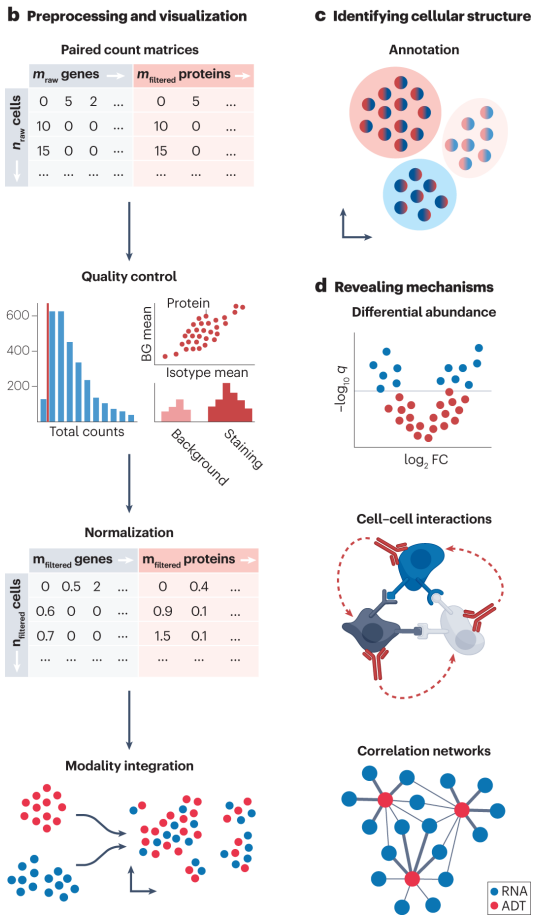
https://cite-seq.com

Stoeckius et al. Nature Methods (2017) – CITE-seq
Peterson et al., Nat Biotechnol. (2017) – REAP-seq

CITE-Seq data analysis



CITE-seq analysis steps



Integration of ADT and Gene Expression Data:

- Each dataset undergoes individual quality control and normalization.
- Data can be visualized individually or jointly to reveal relationships.
- Annotation can be based on transcriptomics data, ADT data, or both.
- Clusters are matched to marker genes and ADTs to annotate cell types accurately.

Biological Mechanisms and Analysis:

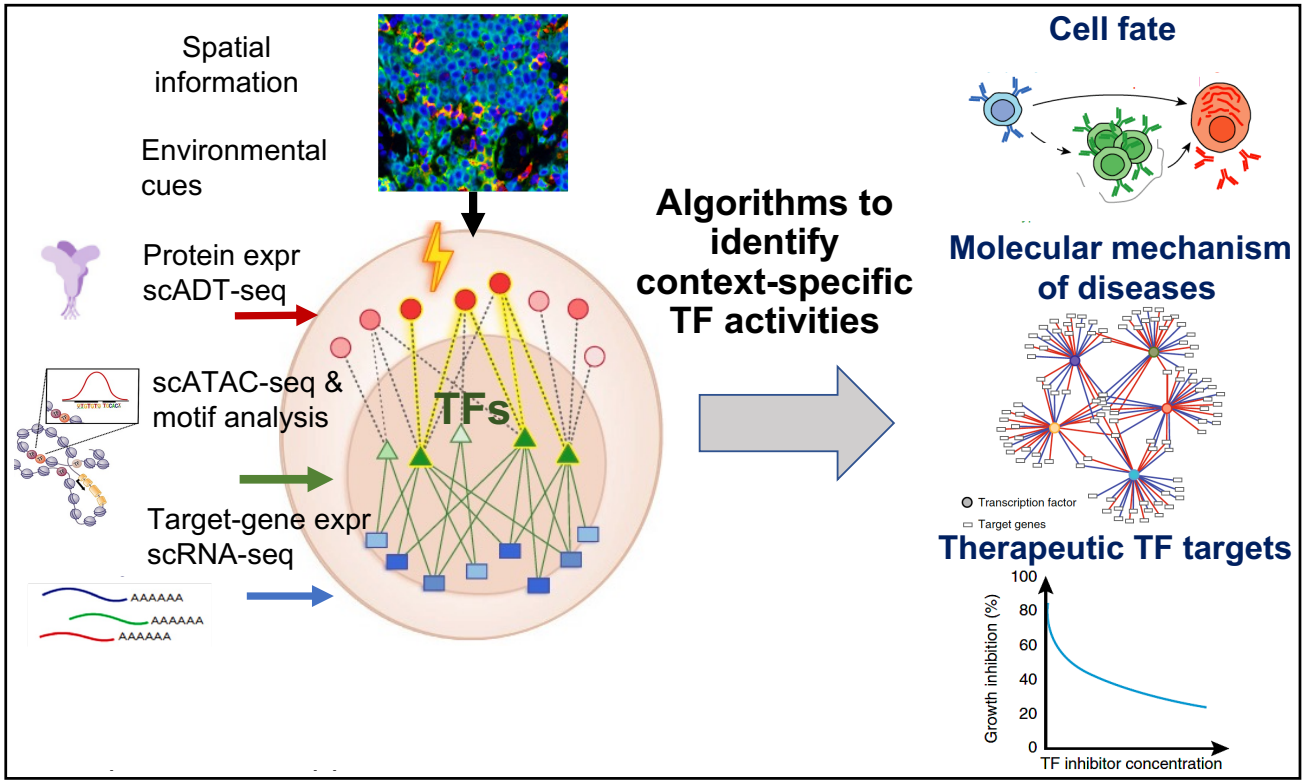
- ADT data is analyzed for differential abundance to understand protein expression changes.
- Cell-cell communication is inferred based on ADT profiles.
- Correlation networks integrate RNA and ADT information to elucidate biological interactions.

Heumos, L., Schaar, A.C., Lance, C. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).

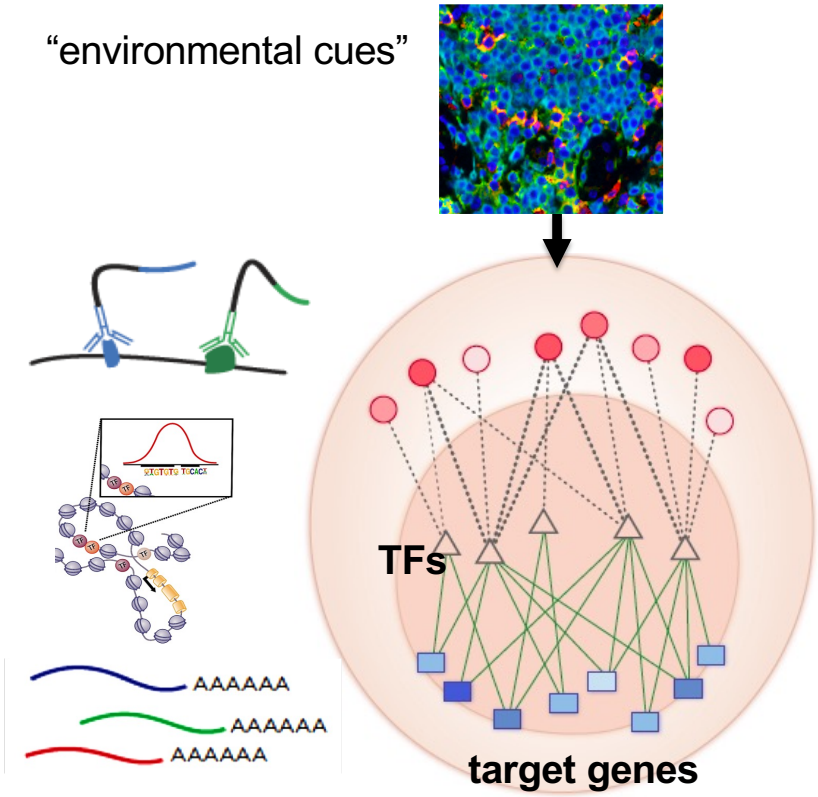
Overview of computational TF inference methods based on single cell omics

Computational methods for modeling context-specific transcription factor activities using single-cell and spatial omics approaches

- Sparse data or poor coverage can be a significant challenge in single-cell and spatial omics experiments
- Machine learning algorithms recover biology from noisy and high-dimensional single/spatial omics data



Computational methods for TF activity inference based on single cell omics



TF activity inference from spatial transcriptomics — e.g. STAN

TF activity inference from single-cell proteomics and transcriptomics data — e.g. SPaRTAN

TF activity inference from single-cell epigenomic and transcriptomics data — e.g. SCENIC+

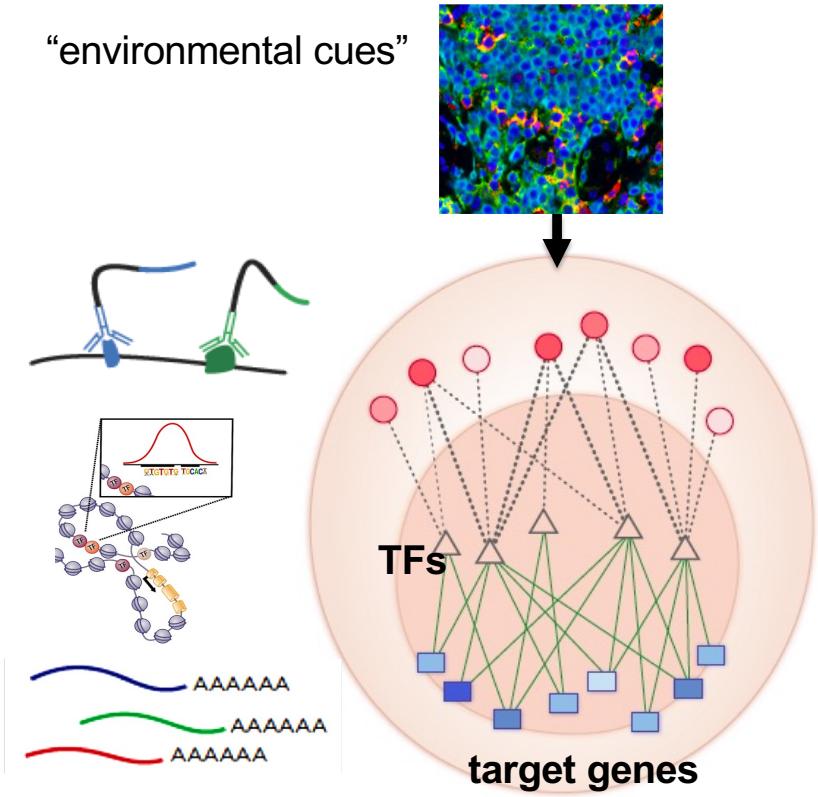
TF activity inference from single-cell epigenomic data — e.g. BITFAM, chromVAR, scBAsset, scFAN

TF activity inference from single-cell gene expression data — e.g. SCENIC, BITFAM, metaVIPER, INFERELATOR 3.0

The major classes of methods for TFA inference

Kim, Daniel et al. *NPJ systems biology and applications,* 2023

Computational methods for TF activity inference based on single cell omics



TF activity inference from spatial transcriptomics — e.g. STAN

TF activity inference from single-cell proteomics and transcriptomics data — e.g. SPaRTAN

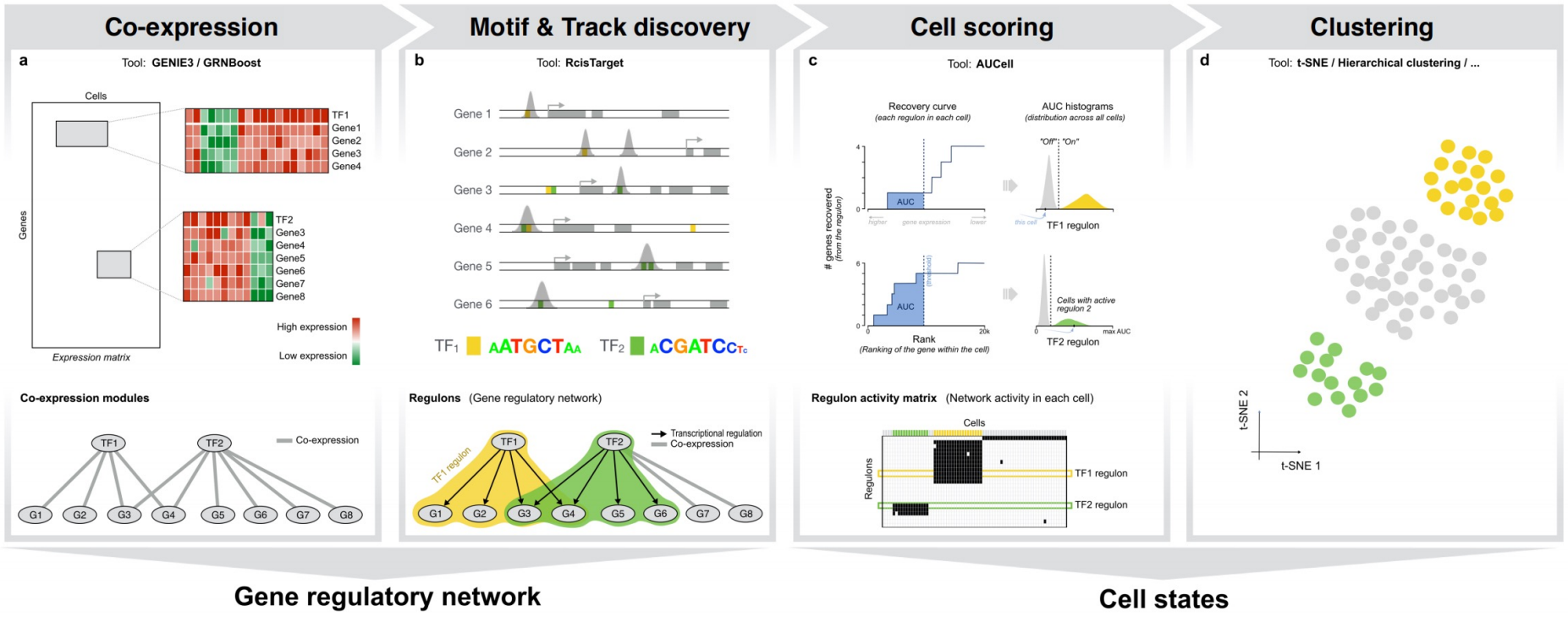
TF activity inference from single-cell epigenomic and transcriptomics data — e.g. SCENIC+

TF activity inference from single-cell epigenomic data — e.g. BITFAM, chromVAR, scBAsset, scFAN

TF activity inference from single-cell gene expression data — e.g. SCENIC, BITFAM, metaVIPER, INFERELATOR 3.0

TF activity inference from single-cell gene expression data (1)

- SCENIC (Single Cell rEgulatory Network Inference and Clustering)



https://github.com/aertslab/SCENIC

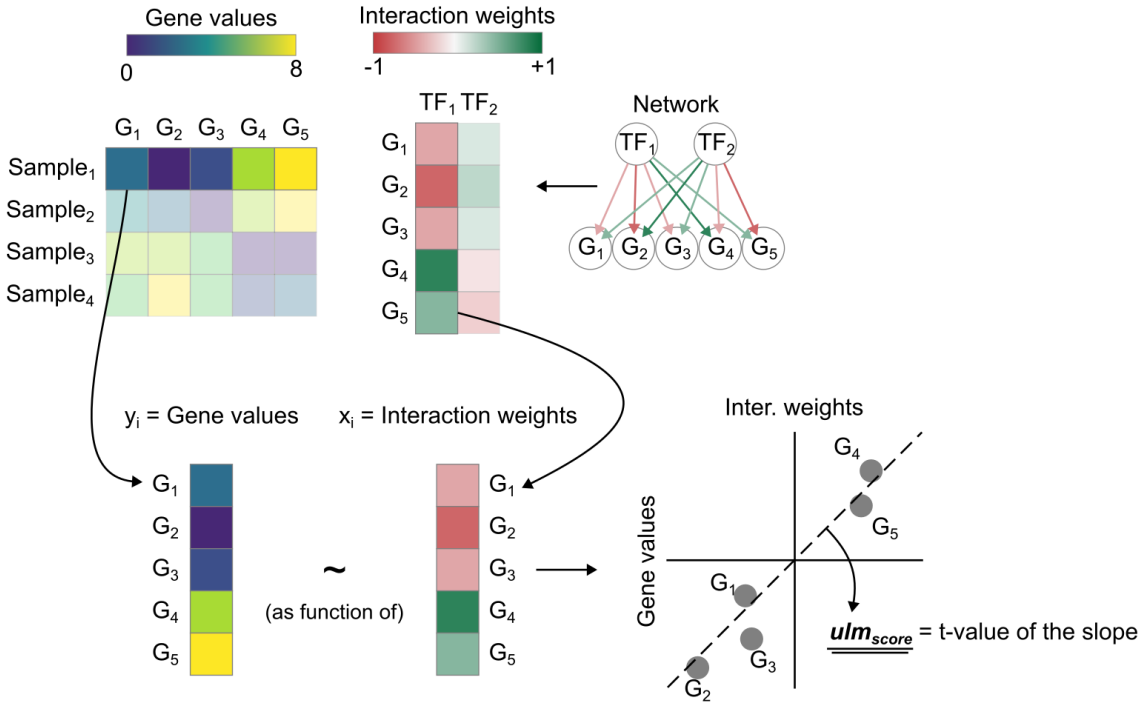
https://github.com/aertslab/SCENICprotocol

Nature Methods 14, 1083 (2017)

TF activity inference from single-cell gene expression data (2)

- Decoupler
  - This method fits a linear model that predicts observed gene expression using the TF's TF-Gene interaction weights. The resulting t-value of the slope serves as the score: a positive score indicates active TF involvement while a negative score suggests inactivity.
  - https://decoupler-py.readthedocs.io/en/1.1.0/notebooks/orothea.html

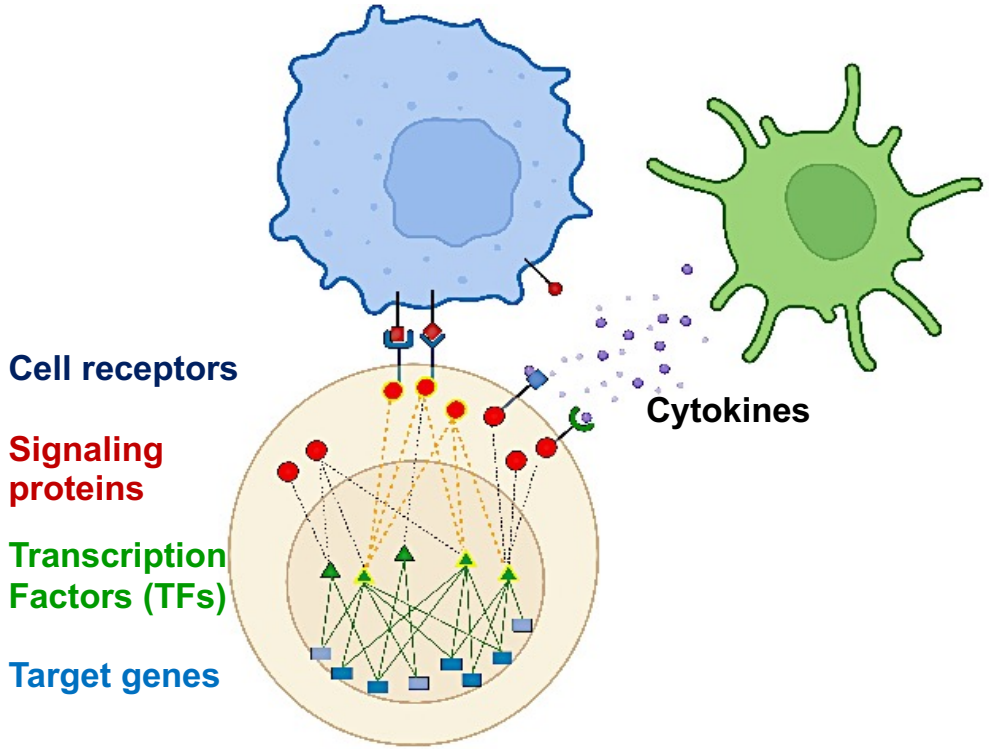
Activity inference with Univariate Linear Model (ULM)



TF activity inference from single-cell gene expression data (3)

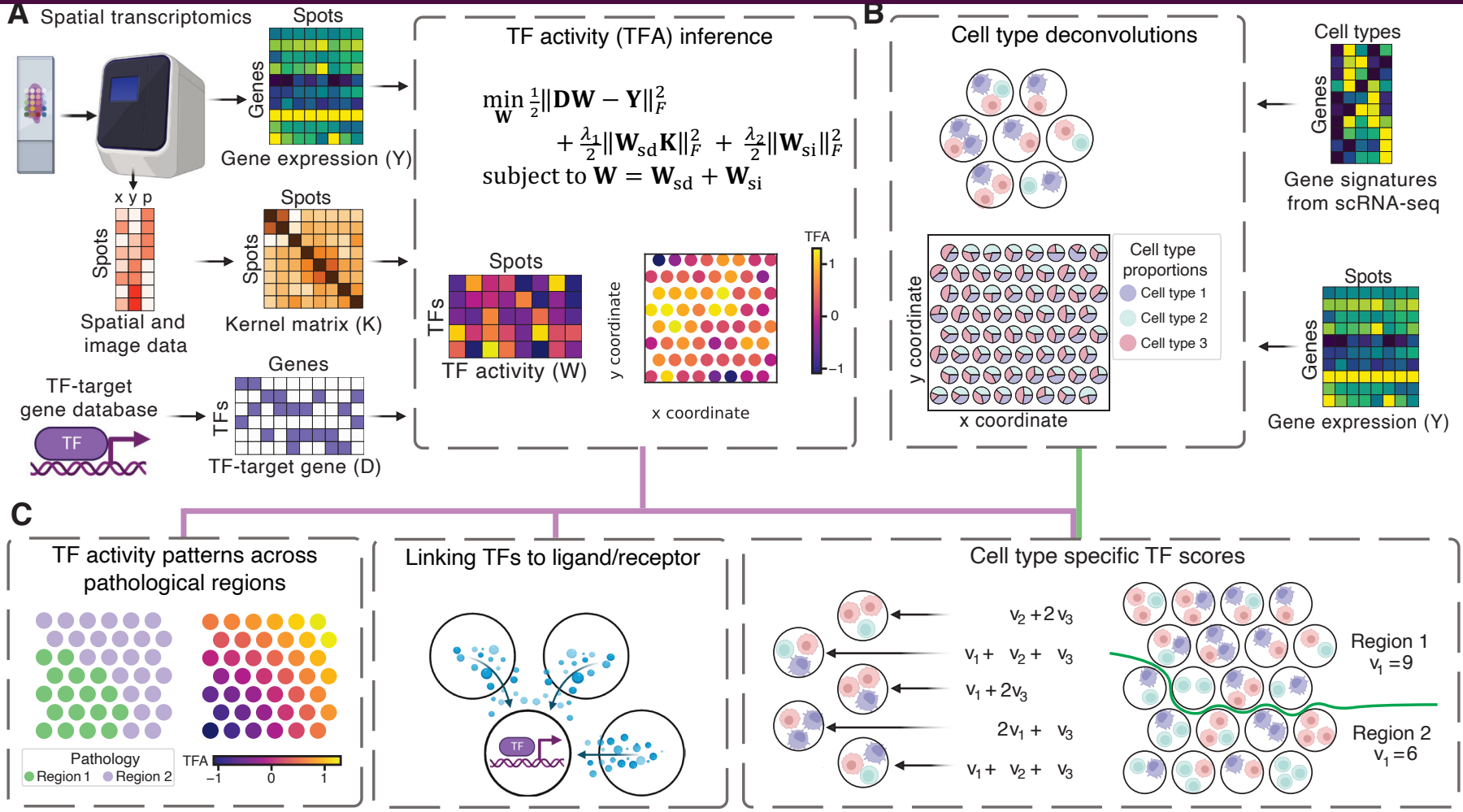
- INFERELATOR 3.0
  - https://inferelator.readthedocs.io/en/latest/tutorial.html
- BITFAM: The Bayesian inference transcription factor activity model
  - https://github.com/jaleesr/BITFAM
- metaVIPER
  - https://github.com/califano-lab/single-cell-pipeline

TF activity inference from spatial transcriptomics



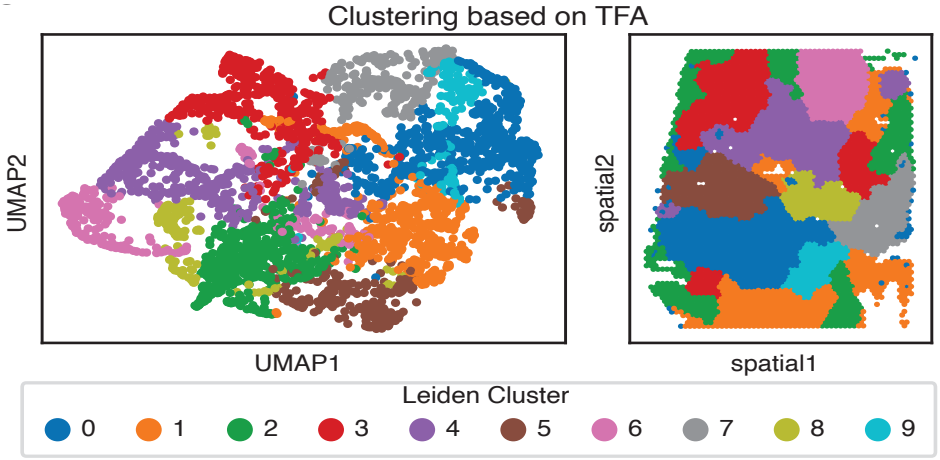
Nearby cells often affect cell context-specific TF activities through receptor-ligand mediated cell-cell communication or secreting molecules like cytokines

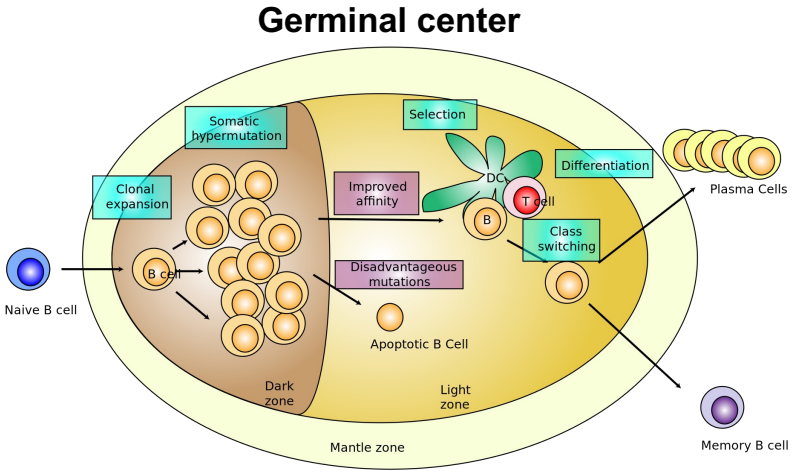
Spatially informed Transcription factor Activity Network (STAN)



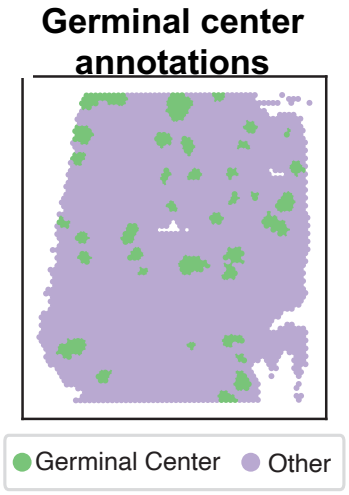
Clustering on STAN predicted TF activities (TFAs) yields spatially coherent clustering in lymph node

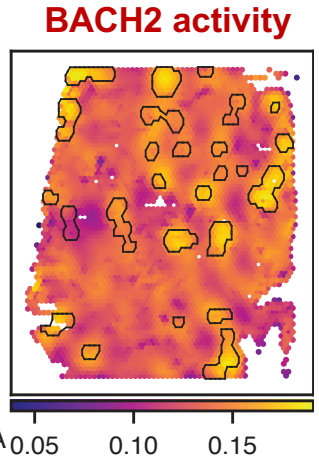
- The human lymph node is characterized by dynamic microenvironments with many spatially interlaced cell populations such as germinal centers (GCs)
- Germinal centers develop in the B cell follicles of secondary lymphoid tissues during T cell-dependent antibody responses

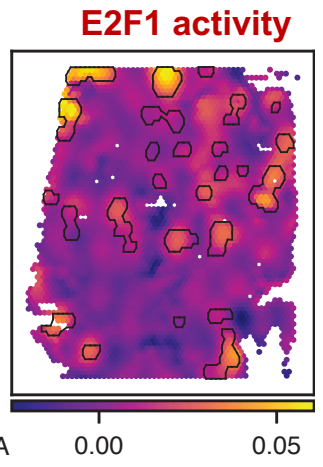




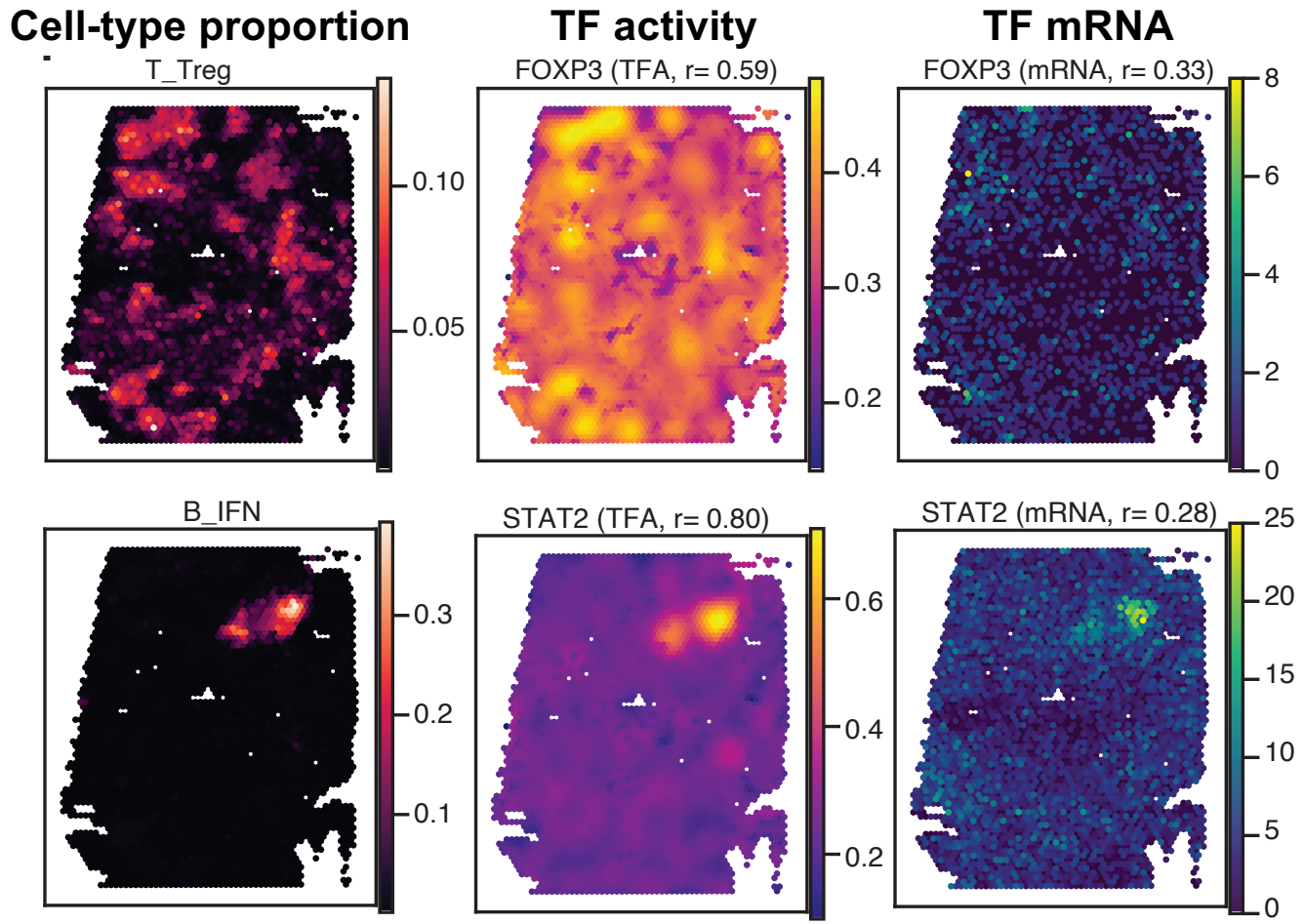
https://en.wikipedia.org/wiki/Germinal_center#/media/File:Germinal_center.svg



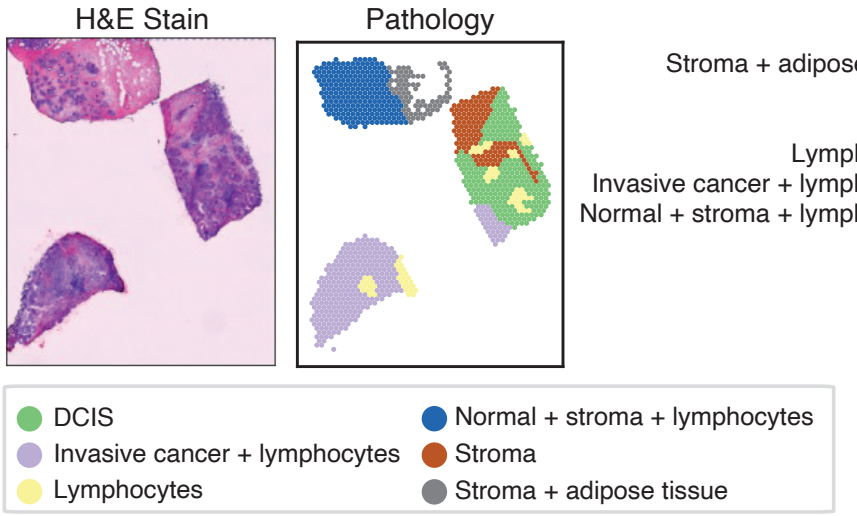


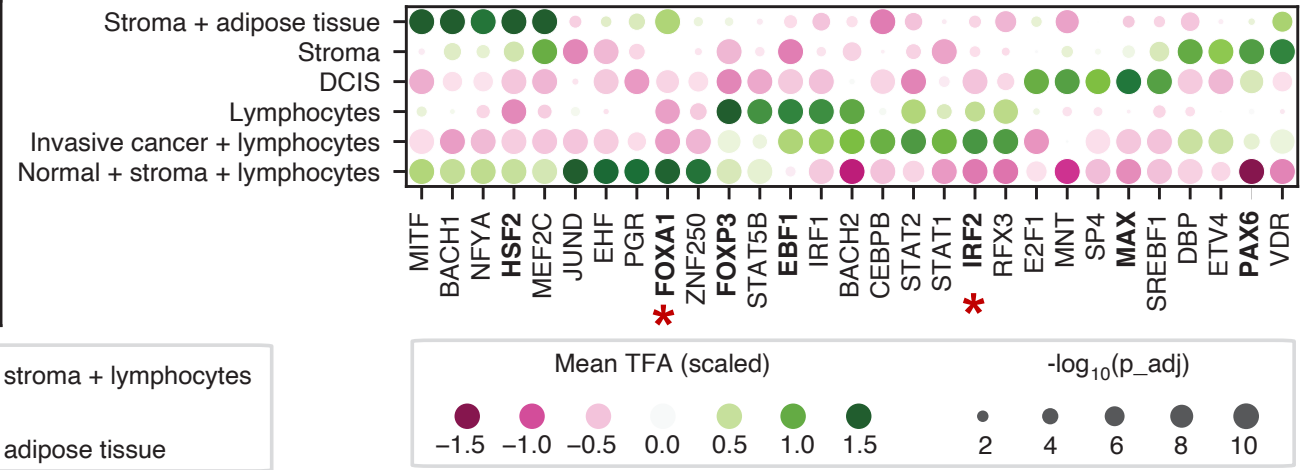


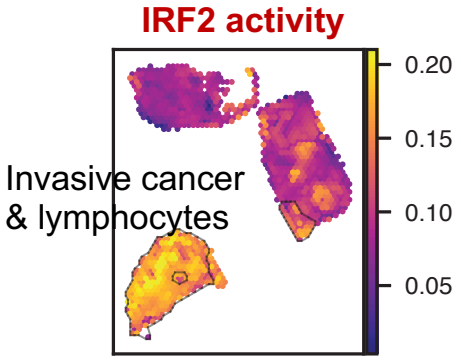
STAN identifies cell-type specific TFs in lymph node

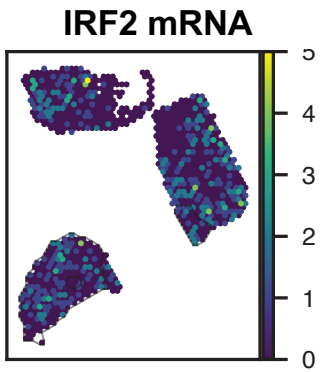


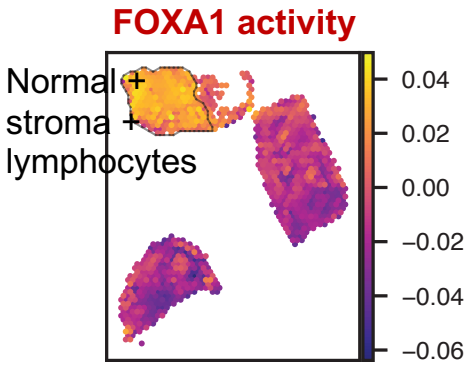
Differential TF activity associated with pathological regions – triple negative breast cancer (TNBC)

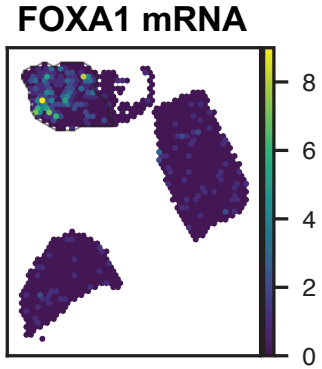




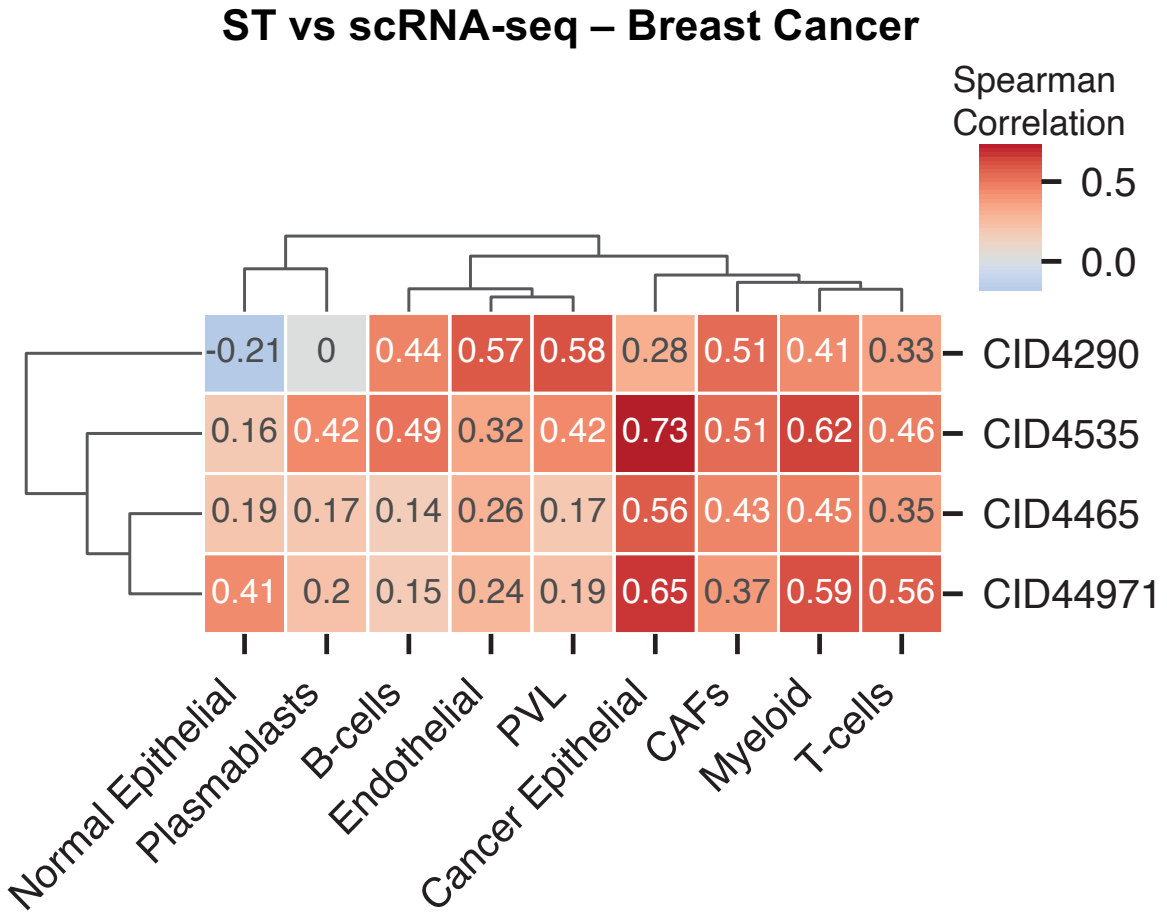




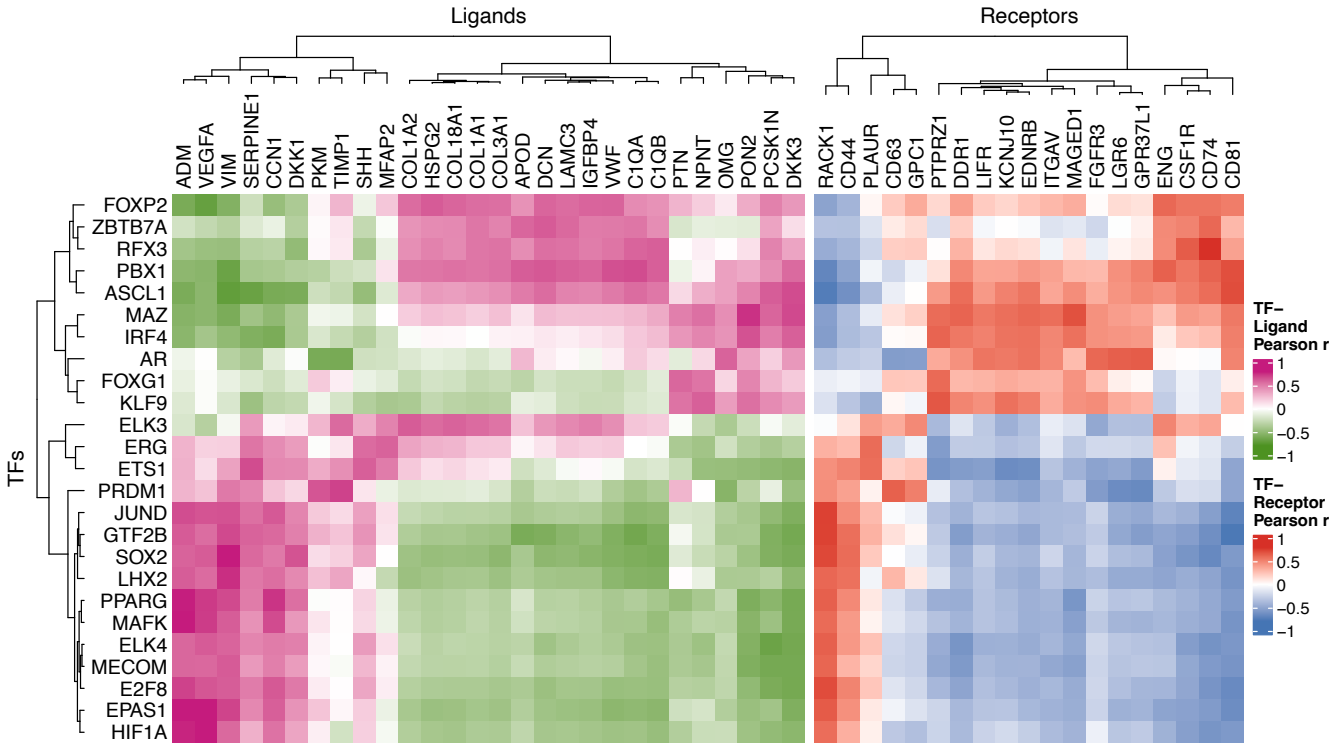




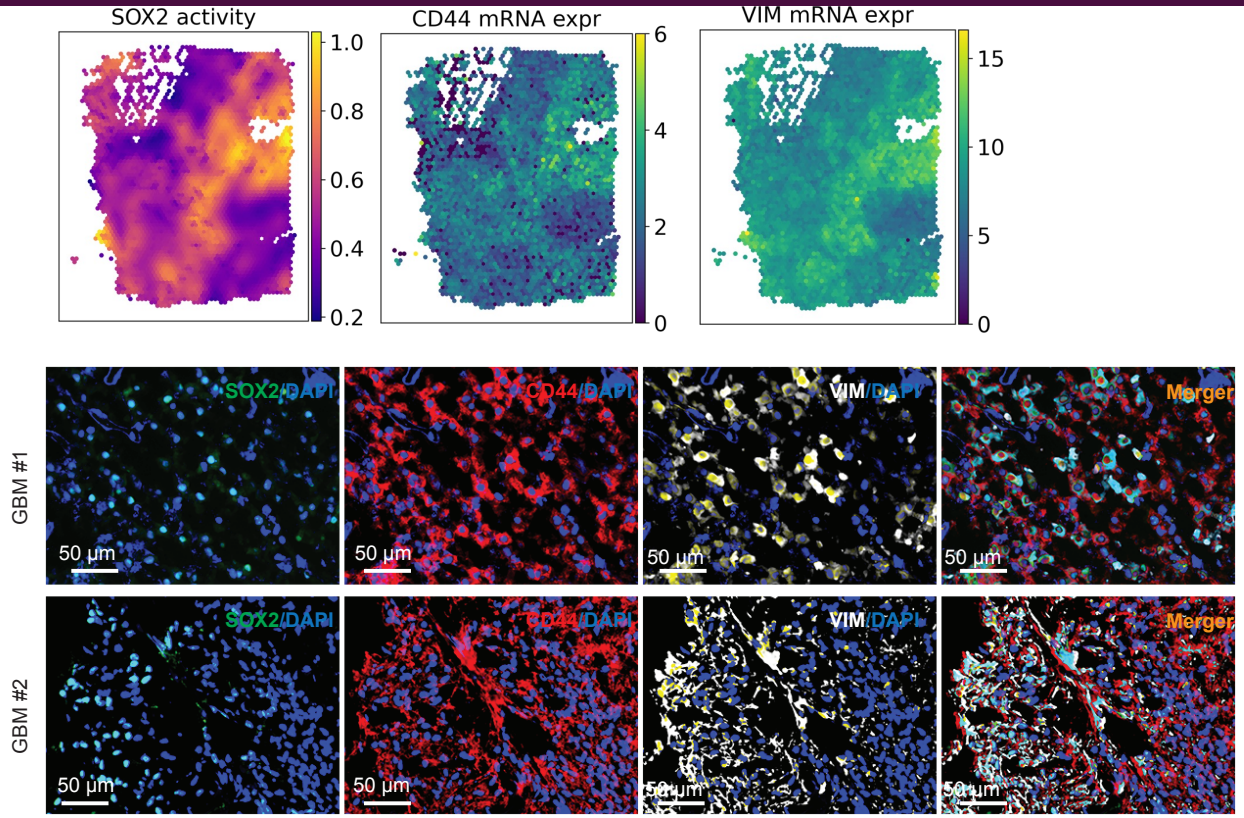
Cell type specific TF scores based ST and scRNA-seq are mostly correlated



Linking ligands and receptors to TFs in glioblastoma



Linking ligands and receptors to TFs in glioblastoma



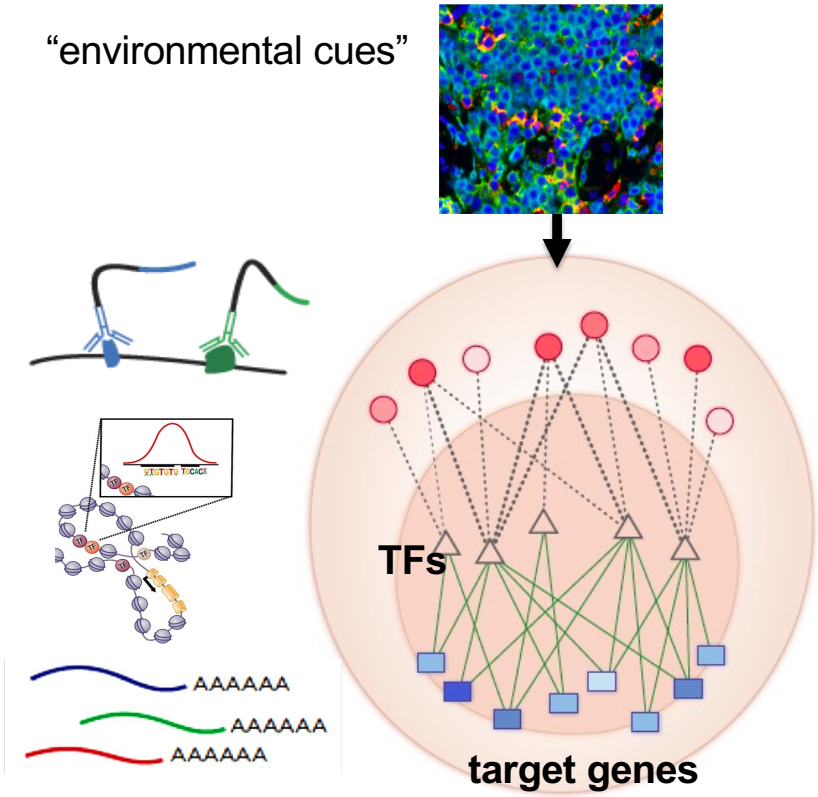
Summary: STAN

- STAN (Spatially informed Transcription Factor Activity Network), a computational method to predict spot-specific TF activities by utilizing spatial transcriptomics datasets and cis-regulatory information
- Applying STAN to ST datasets
  - Decipher critical TF regulators underlying cell identities, spatial domains (e.g., GCs), and pathological regions (e.g., stroma vs. tumor);
  - Determine whether a given pathological region/spatial domain has different and/or common regulators across disease subtypes (e.g., stroma in TNBC vs. ER+ breast cancer);
  - Identify similar and/or different TFs associated with spatial domains or cell types across healthy individuals and those manifesting a disease
  - Link ligands and receptors to TFs for elucidating potential signaling pathways and regulatory networks involved in cellular communication and tissue microenvironment interactions.

https://github.com/osmanbeyoglulab/STAN

Zhang L, Sagan A, Qin B, Hu B, Osmanbeyoglu HU, STAN, a computational framework for inferring spatially informed transcription factor activity across cellular contexts, BioXviv

Computational methods for TF activity inference based on single cell omics



TF activity inference from spatial transcriptomics — e.g. STAN

TF activity inference from single-cell proteomics and transcriptomics data — **e.g. SPaRTAN**

TF activity inference from single-cell epigenomic and transcriptomics data — e.g. SCENIC+

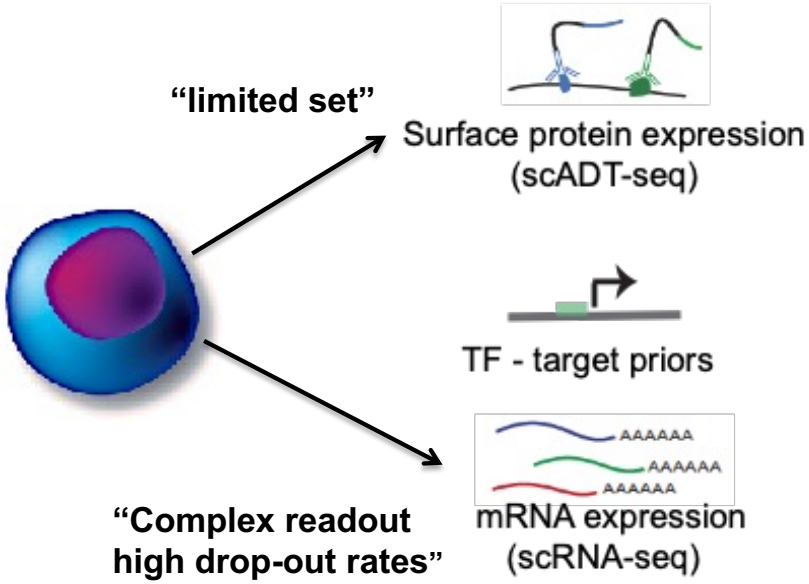
TF activity inference from single-cell epigenomic data — e.g. BITFAM, chromVAR, scBAsset, scFAN

TF activity inference from single-cell gene expression data — e.g. SCENIC, BITFAM, metaVIPER, INFERELATOR 3.0

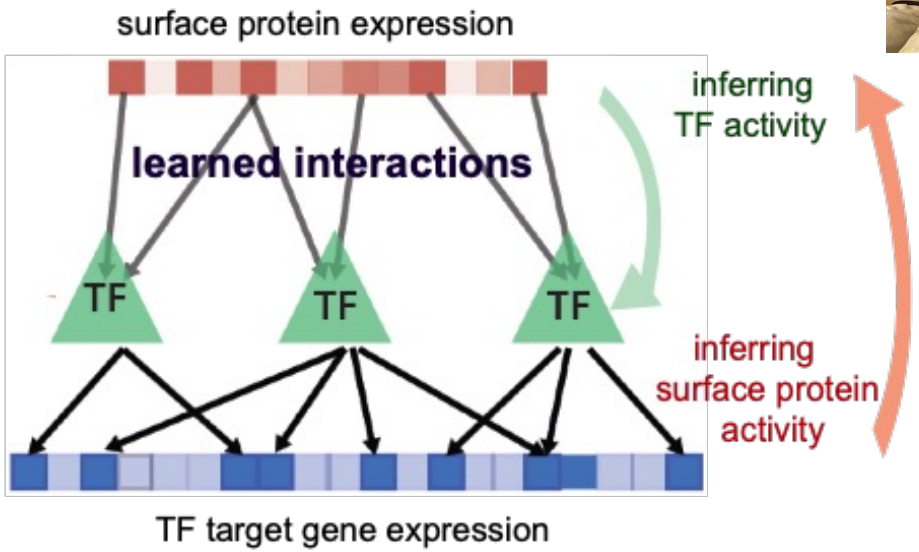
TF activity inference from single-cell proteomics and transcriptomics data

SPaRTAN (Single cell Proteomic And RNA based Transcription factor Activity Network)

Single cell multi-omics profiling CITE-seq

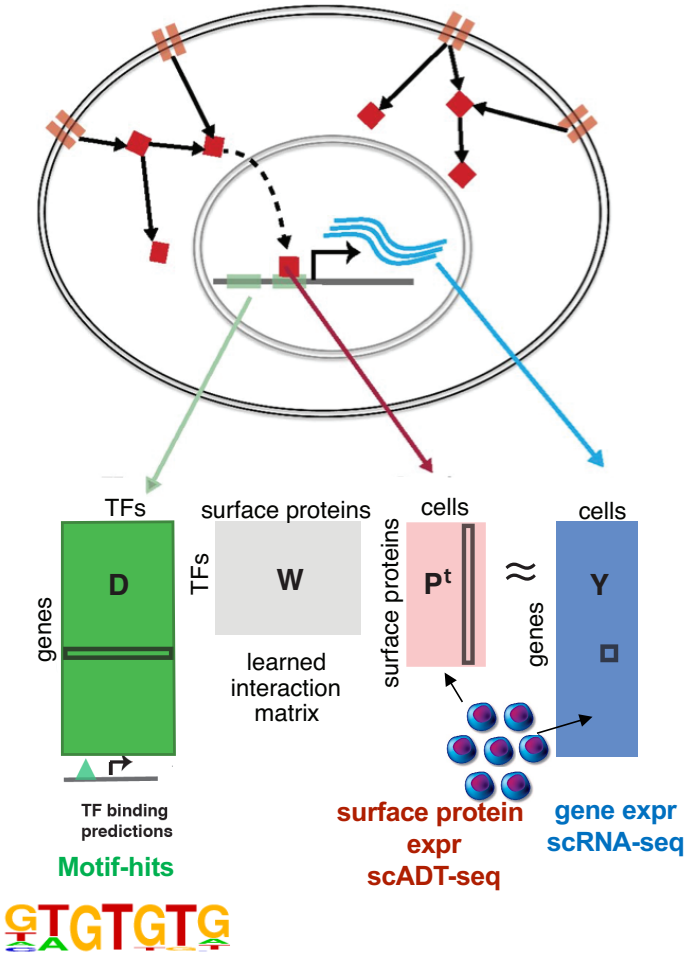


Linking surface proteins to transcriptional regulators



Ma et al. *Nucleic Acids Research,* 2021

SPaRTAN - Linking surface proteins to TFs



genes ~ 5000
cells ~ 1000
TFs ~ 300
surface proteins ~ 30

Regularized bilinear regression

$$\mathbf{DWP}^t + \varepsilon = \mathbf{Y}$$

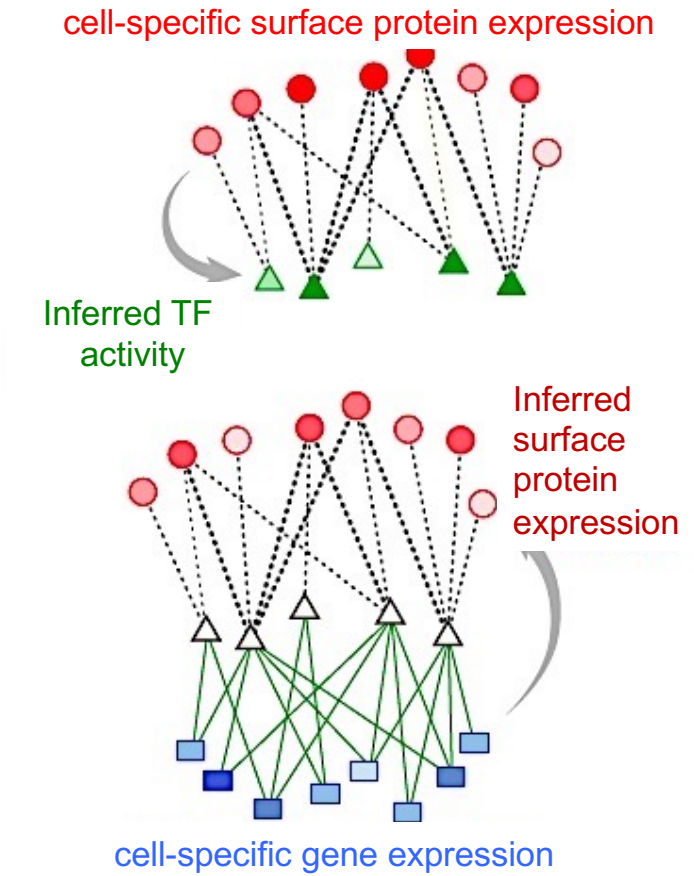
$$\mathbf{Y}^t\mathbf{DWP}^t \approx \mathbf{Y}^t\mathbf{Y}$$

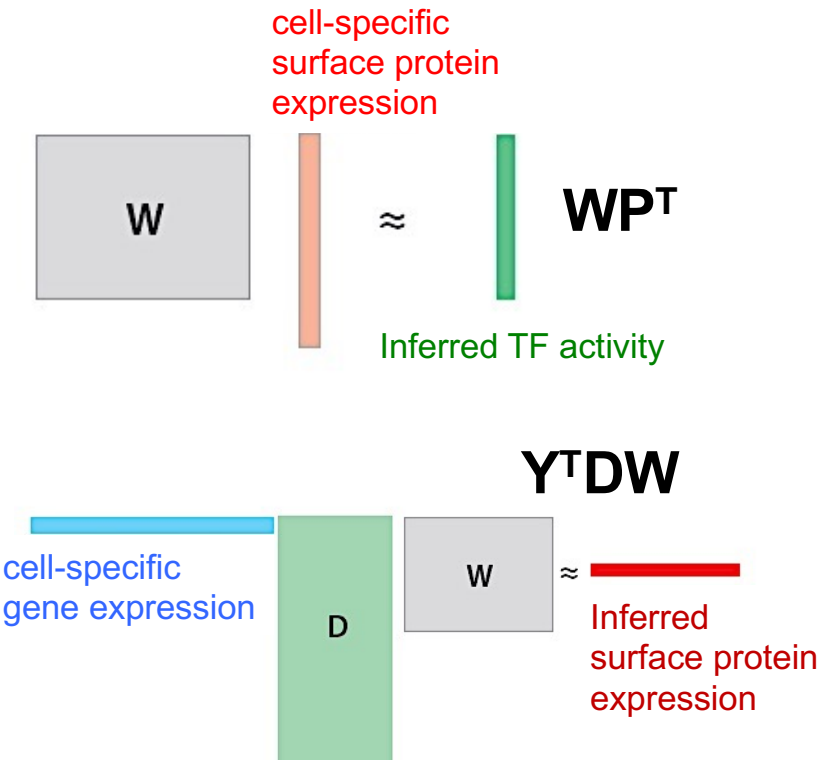
$$\mathbf{P} \otimes \mathbf{Y}^t\mathbf{D}\,\mathrm{vec}(\mathbf{W}) \approx \mathrm{vec}(\mathbf{Y}^t\mathbf{Y})$$

$$\arg\min_{\mathbf{W}} \left\| \mathrm{vec}(\mathbf{Y}^t\mathbf{Y}) - \mathbf{P} \otimes \mathbf{Y}^t\mathbf{D}\,\mathrm{vec}(\mathbf{W}) \right\|_2^2 + \lambda \|\mathbf{W}\|_2^2$$

Osmanbeyoglu et al., *Genome Res.,* 2014
Osmanbeyoglu et al., *Nat.Commun.,* 2017
Hmeljak et al., *Cancer Discovery,* 2018
Ma et al. *NAR*, 2021
Ramjattun*, Ma* et al., *JMV*, 2023

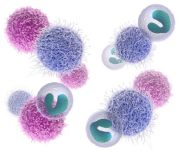
Inferring cell-specific TF activity and surface protein expression



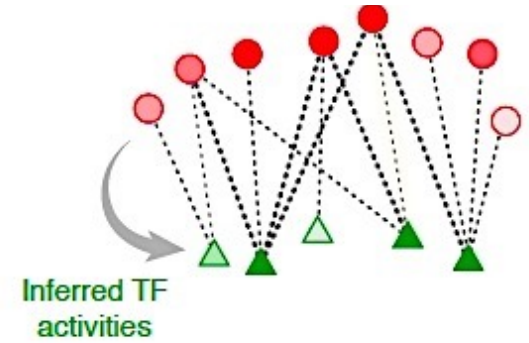


Reduced dimension in terms of TF activities and surface protein expression make data more tractable and reduce noise in single cell data while preserving the often intrinsically low dimensional signal of interest

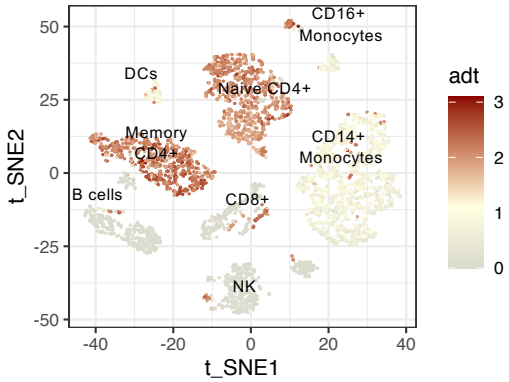
SPaRTAN identifies both known and novel cell type-specific TFs



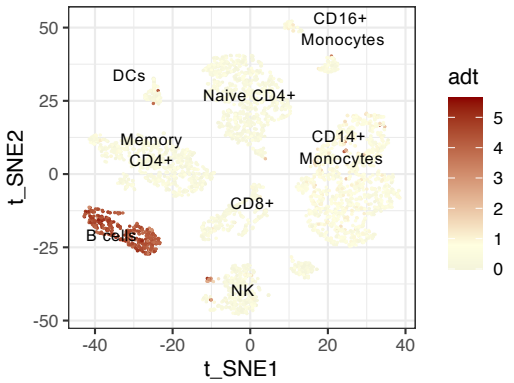
cell-specific surface protein expression



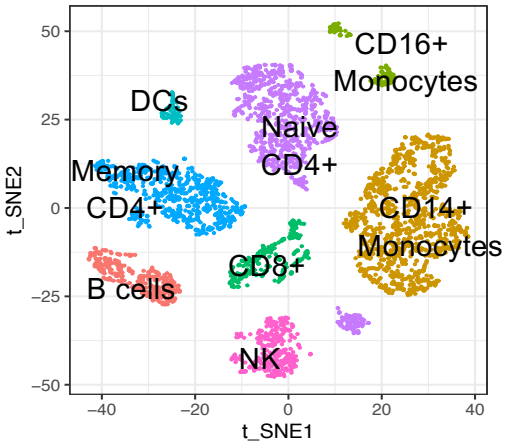
CD4 expression



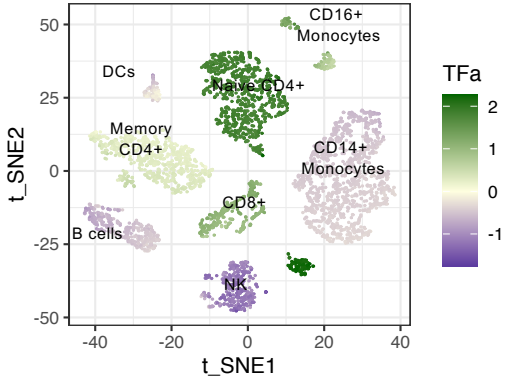
CD19 expression



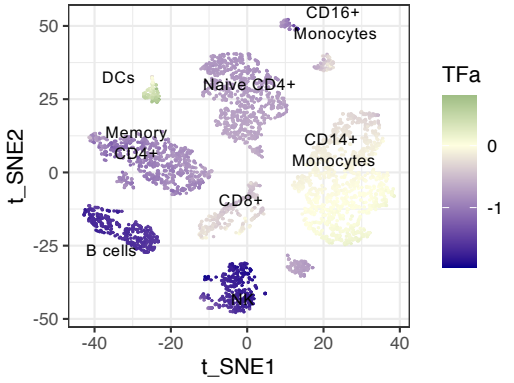
Inferred TF activity



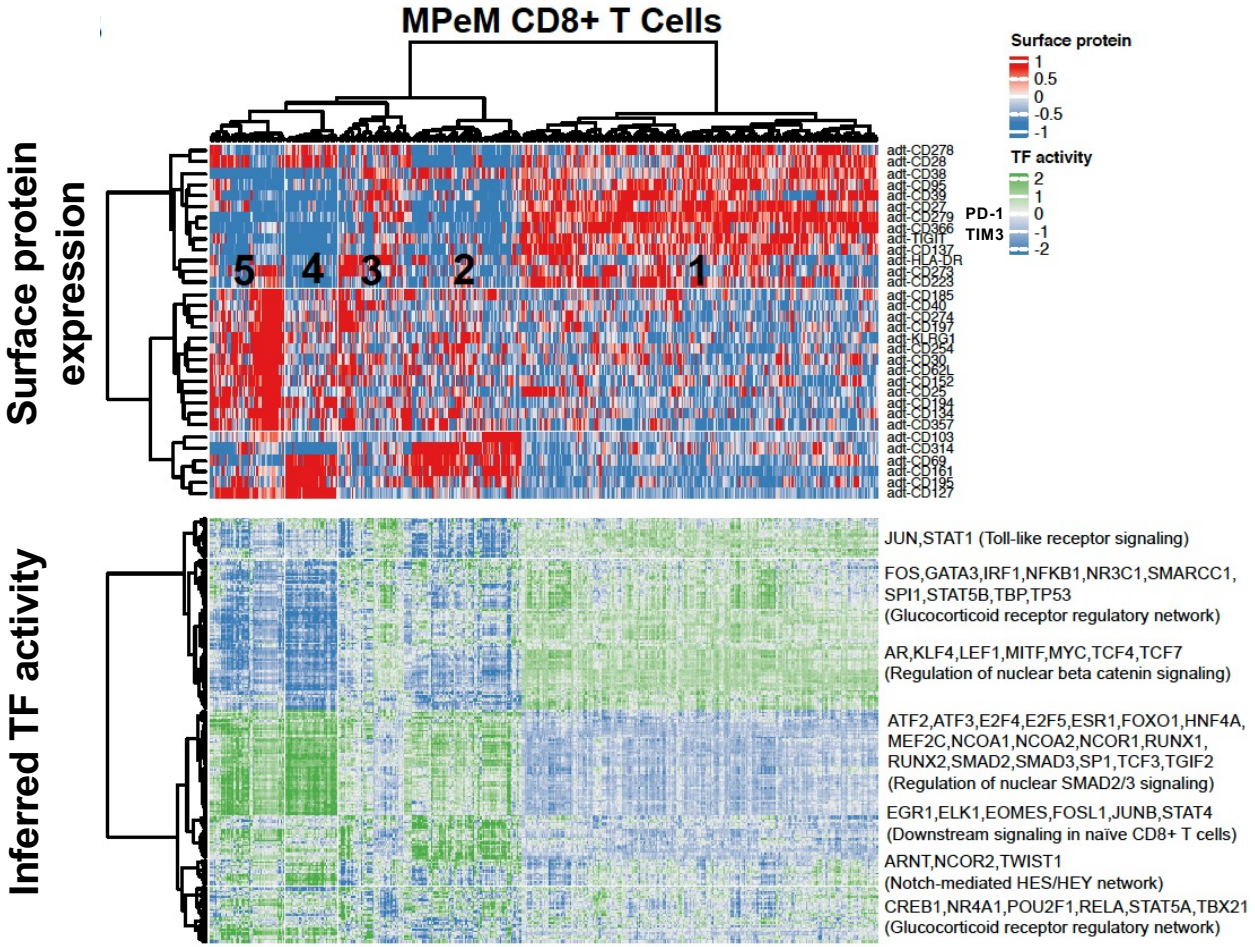
GATA3 activity

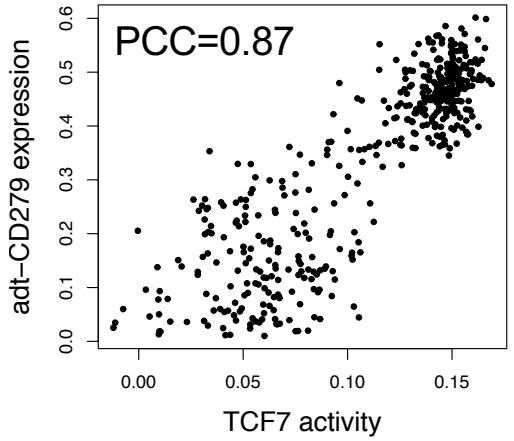


BACH1 activity



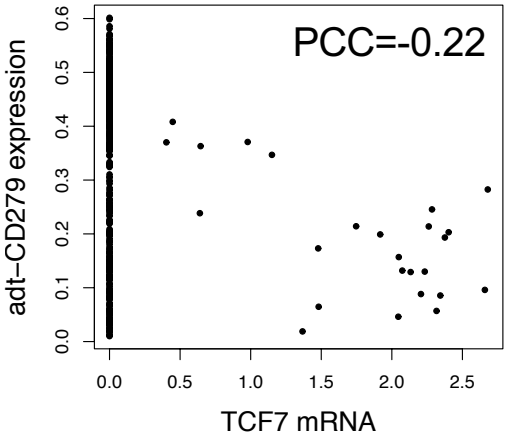
Relating inferred SPaRTAN TF activity, surface protein expression and MPeM CD8^{+} T cell subsets (n=466)





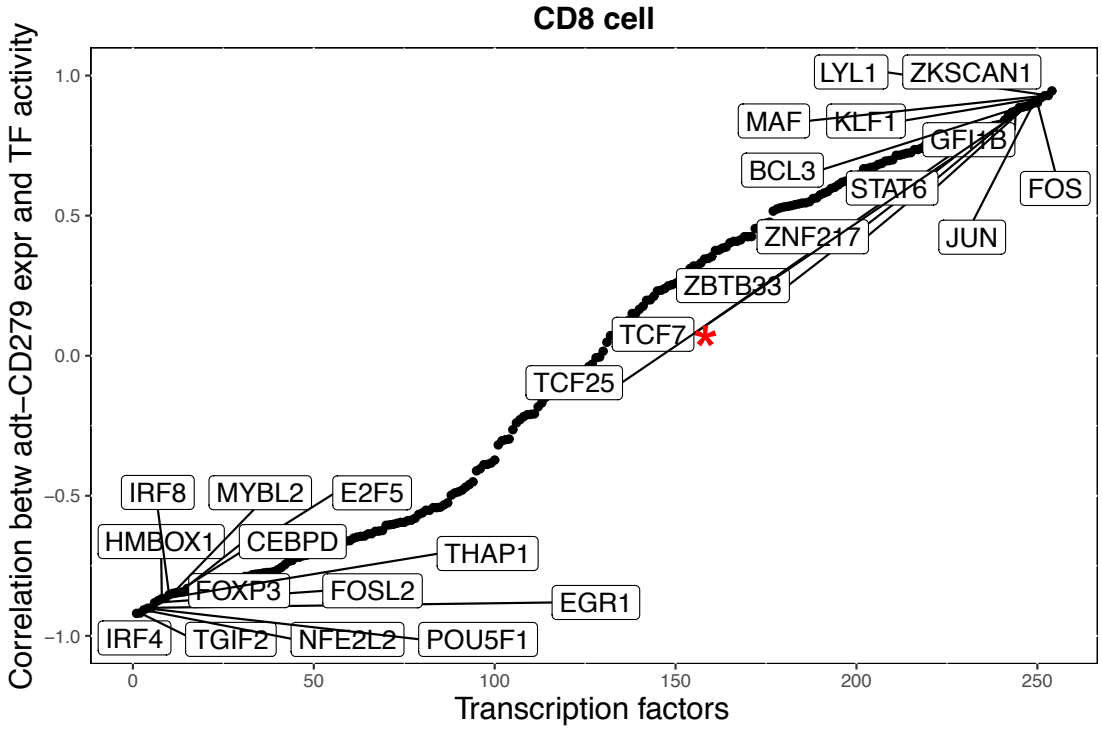
TCF7 is required for memory cells and efficacy of immunotherapies

Siddiqui et al. Immunity 2019
Kurtulus et al., Immunity 2019

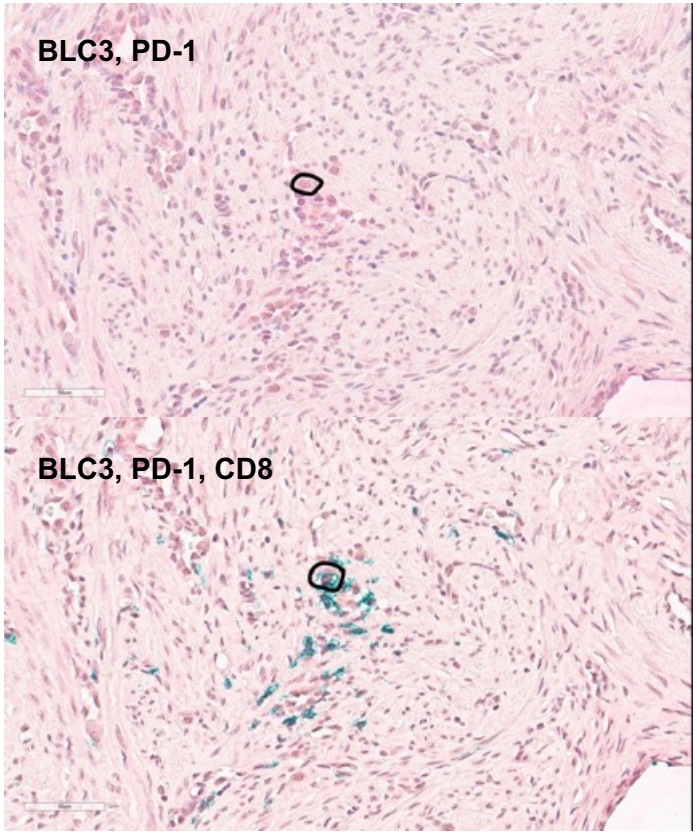


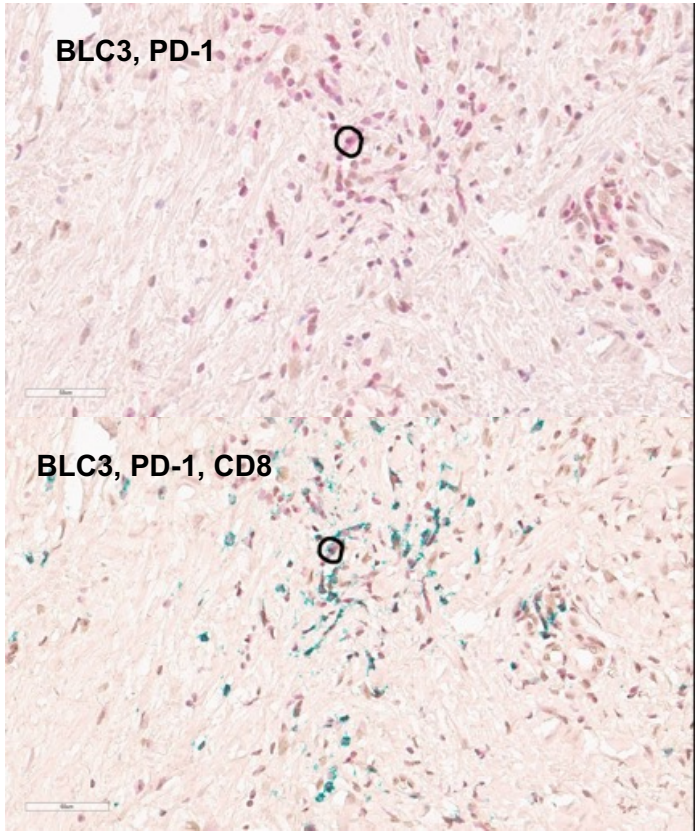
We do not observe those differences at the gene expression level

Correlation of inferred TF activities with PD-1 (CD279) protein expression in MPeM CD8^{+} T cell



National Mesothelioma Virtual Bnk (NMVB) Validation Cohort
Co-expression of CD8, PD1, BCL3 by IHC staining

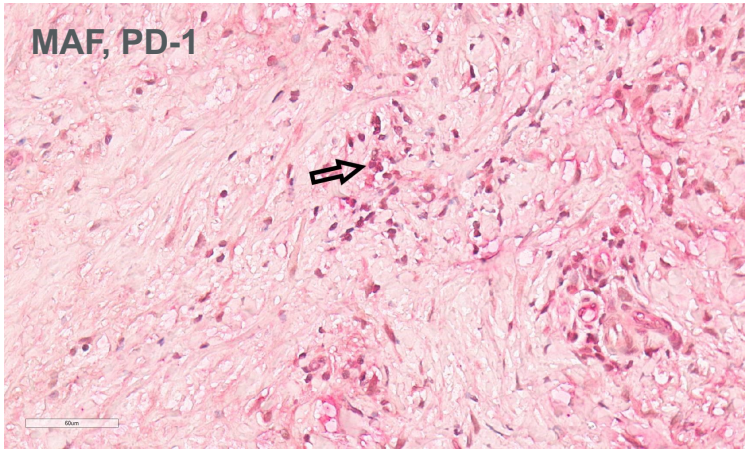


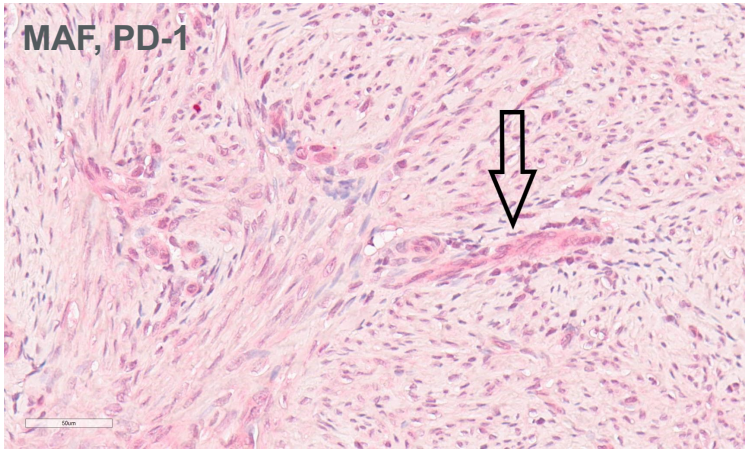


PD-1(brown)
BCL3 (red)
CD8 (blue/green)

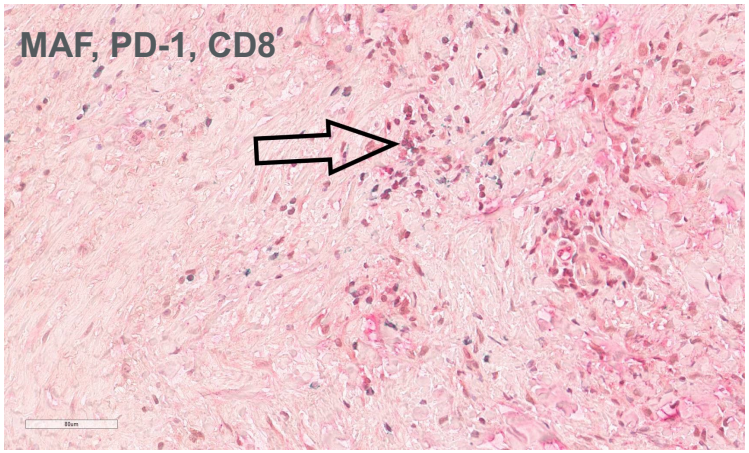
- *BCL3* induces survival and proliferation in cancer cells
- Its role in $CD8^+$ T cells has not been studied

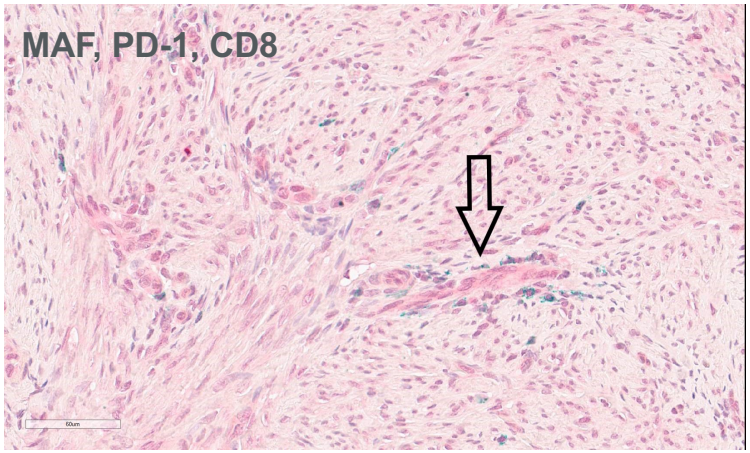
NMVB Validation Cohort
Co-expression of CD8, PD1, MAF by IHC staining





PD-1(brown)
MAF (red)
CD8 (blue/green)





MAF drives CD8^{+} T-cell exhaustion in melanoma

Giordano et al., EMBO. J. 2015

Summary: SPaRTAN

- SPaRTAN links expression of cell-surface receptors with transcription factors by utilizing paired single-cell proteomes and transcriptomes
- Application of SPaRTAN to CITE-seq datasets helps to
  - decipher critical regulators (e.g. TFs, surface receptors) underlying cellular identities (e.g. naïve versus memory T cells);
  - determine whether given cell types have different or common regulators across tissues (e.g. B cells in spleen versus lung);
  - determine commonalties as well as differences of cell-specific regulatory programs across healthy individuals and those manifesting a disease.

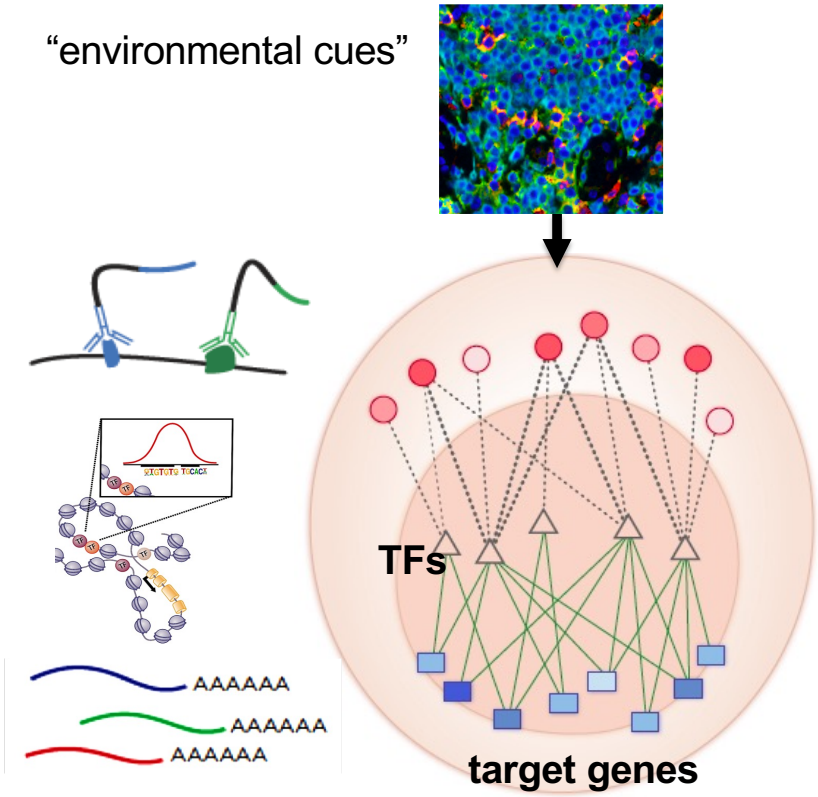
https://github.com/osmanbeyoglulab/SPaRTAN

Xiaojun Ma, Ashwin Somasundaram, Zengbiao Qi, Douglas J Hartman, Harinder Singh, Hatice Ulku Osmanbeyoglu, SPaRTAN, a computational framework for linking cell-surface receptors to transcriptional regulators, *Nucleic Acids Research*, 2021, Pages 9633–9647, https://doi.org/10.1093/nar/gkab745

Limitations related to methods based on single cell/spatial gene and/or protein expression

- Curated TF-gene interactions
  - Motif analysis in promoter region
  - Curated interactions from diverse sources including literature, ChIP-seq peaks, TF binding motifs, and inferred from gene expression.
    - RegNetwork, Liu et al., “RegNetwork: an integrated ...” Database, 2015
    - TRRUST, Han et al., “TRRUSTv2 ...” NAR, 46(D1):D380–D386, 2018
    - DoRothEA, Garcia-Alonso et al., “Benchmark ...”, Gen. Res., 29:1363–1375, 2019

Computational methods for TF activity inference based on single cell omics



TF activity inference from spatial transcriptomics — e.g. STAN

TF activity inference from single-cell proteomics and transcriptomics data — e.g. SPaRTAN

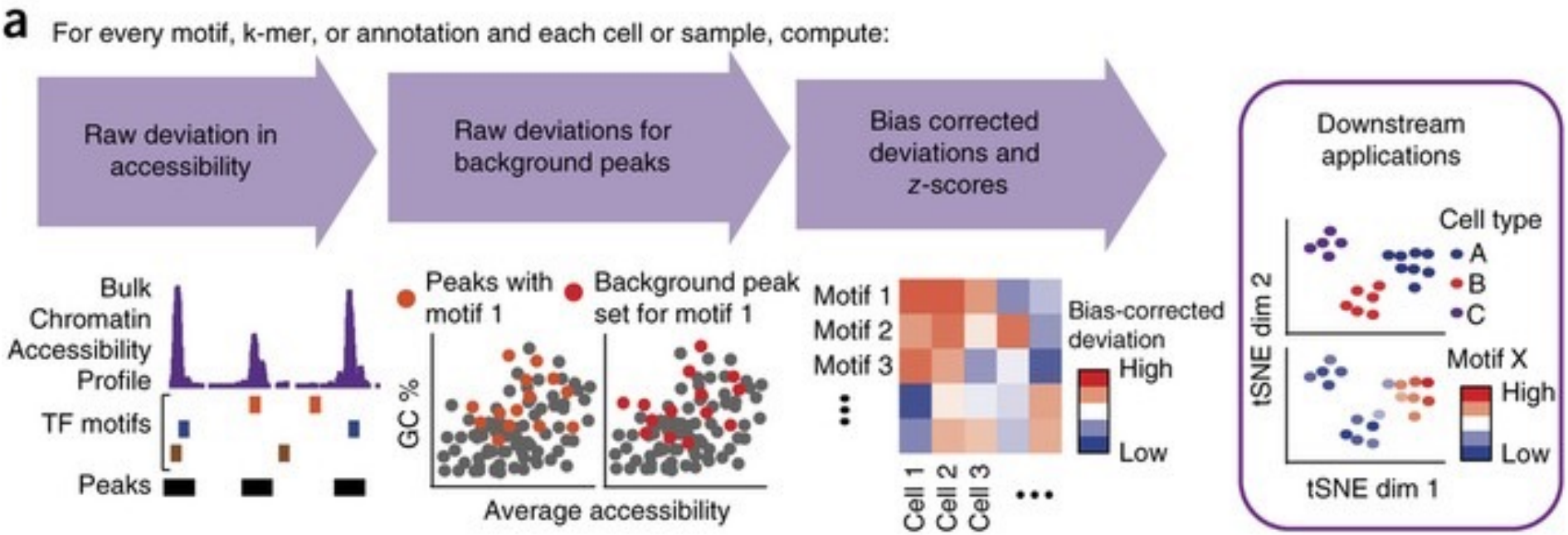
TF activity inference from single-cell epigenomic and transcriptomics data — **e.g. SCENIC+**

TF activity inference from single-cell epigenomic data — e.g. BITFAM, chromVAR, scBAsset, scFAN

TF activity inference from single-cell gene expression data — e.g. SCENIC, BITFAM, metaVIPER, INFERELATOR 3.0

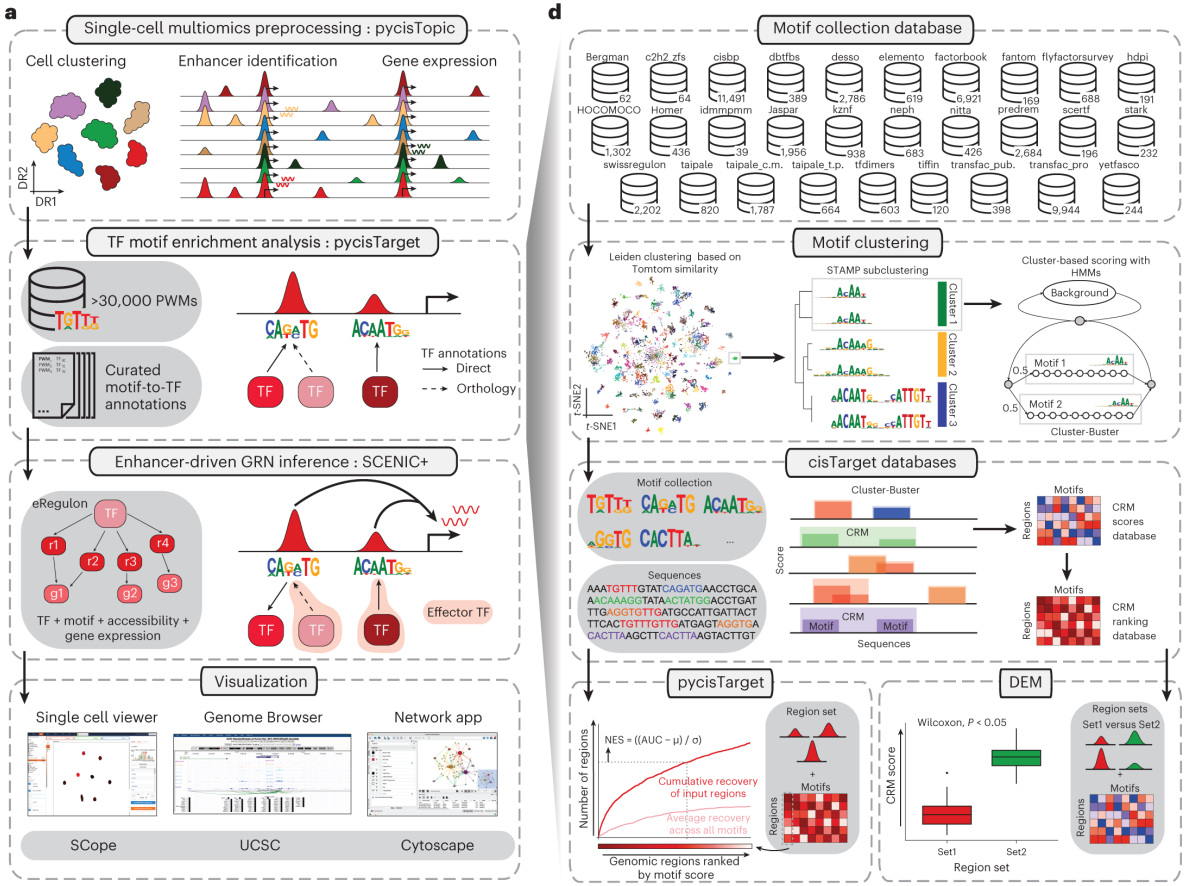
TF activity inference from single-cell epigenomic data

- ChromVar: From peaks to motifs (Nature Methods 14, 975 (2017))
  https://bioconductor.org/packages/release/bioc/vignettes/chromVAR/inst/doc/Introduction.htm



TF activity inference from single-cell epigenomic and transcriptomics data (1)

SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks
https://scenicplus.readthedocs.io/en/latest/



TF activity inference from single-cell epigenomic and transcriptomics data (2)

	Method	Prog.	Cell type	Metacell	Gene Regulatory Network			Inputs		
					TF-CRE	CRE-Gene	TF-Gene	RNA	ATAC	Paired
Correlation	scMEGA	R	✓	✗	Motif enrichment	Pearson's correlation	Pearson's correlation	✓	✓	✓
	FigR	R	✗	✗	Motif enrichment Spearman's correlation	Spearman's correlation	Spearman's correlation	✗	✗	✓
	STREAM	R	✗	✗	Motif enrichment	Pearson's correlation	Hybrid biclustering	✗	✗	✓
	TRIPOD	R	✓	✓	Motif enrichment	Spearman's correlation	Spearman's correlation	✗	✗	✓
Regression	Pando	R	✗	✓	Motif enrichment	N/A	Linear regression	✗	✗	✓
	scREMOTE	R	✗	✗	Motif enrichment	Chromatin conformation	Linear regression	✗	✗	✓
	RENIN	R	✓	✓	Motif enrichment	Elastic net regression	Elastic net regression	✗	✗	✓
	DIRECT-NET	R	✓	✓	Motif enrichment	Gradient boosting	N/A	✗	✓	✓
	SCENIC+	Python	✓	✓	Motif enrichment	Gradient boosting	Gradient boosting	✓	✓	✓
Prob.	scMTNI	C++	✓	✗	Motif enrichment	N/A	Bayesian inference	✗	✗	✓
D.S	Dictys	Python	✓	✓	Motif enrichment	N/A	Stochastic diff. eq.	✓	✓	✓
D.L	DeepMAPS	Python	✓	✗	Motif enrichment	Graph autoencoder	Regulon construction	✗	✗	✓
	MTLRank	Python	✓	✗	TF activity score		Multilayer neural network	✗	✗	✓
	LINGER	Python	✓	✗	Motif enrichment Pearson's correlation	Multilayer neural network		✗	✗	✓

Hands-on experience

- Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public scRNA-seq and spatial transcriptomics

- Hands-on experience in applying tools and interpreting results using TF activity inference methods using public CITE-seq

- Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public scATAC-seq and multiome
-

Acknowledgement

Osmanbeyoglu Lab
Linan Zhang, PhD (now at Ningbo University)
April Sagan, PhD (now at GRAIL)
Xiaojun Ma, MS
Parham Hadikhani
Haoyu Wang

ICI INNOVATION IN CANCER INFORMATICS

NIH National Institutes of Health

https://www.osmanbeyoglulab.com

NCI R00 CA207871
NIGMS R35 GM146989

HILLMAN FELLOWS
For Innovative Cancer Research Program