



# 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/>

[osmanbeyoglu@pitt.edu](mailto:osmanbeyoglu@pitt.edu)

ECCB2024 Tutorial

September 17, 2024

**UPMC | HILLMAN  
CANCER CENTER**



University of  
Pittsburgh

## Tutorial overview

- This tutorial is focused on basic concepts and recent advances in transcription factor (TF) inference methods for single-cell/spatial omics datasets including scRNA-seq, spatial transcriptomics, CITE-seq, scATAC-seq and multiome.
- This tutorial is intended for an audience with genomics/computational background.
- TF activity inference methods that are recently developed in the field will be introduced from a high-level perspective.

## Schedule

---

9.00-10.30	Welcome remarks and tutorial overview  Basic principles behind TF activity inference methods <ul style="list-style-type: none"><li>• Overview of the importance of context-specific TF regulation in biological systems</li><li>• Significance of TF dynamics in health and disease</li><li>• Single-cell multi-omics and spatial transcriptomics technologies for TF activity inference (scRNA-seq, scATAC-seq, Multiome and CITE-seq)</li></ul> Overview of computational TF inference methods based on single cell omics
10.30-11.00	Coffee
11.00-12.00	Hands-on experience applying tools and interpreting results with transcription factor (TF) activity inference methods, utilizing public single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics data
12.00-13.00	Lunch break

# Practical information

- Tutorial materials (presentation slides + hands-on tutorials):  
<https://github.com/osmanbeyoglulab/ECCB-2024-Tutorial/>
- Environmental setup



## Installation of Conda

Download the installer by choosing the proper installer for your machine. 2. Verify your installer hashes using [SHA-256 checksums](#). 3. Install the installer: - Windows: Double-click the .exe file. Follow the instructions on the screen. For a detailed reference, please read [this page](#). - macOS: double-click the .pkg file. Follow the instructions on the screen. For a detailed reference, please read [this page](#). - Linux: In your terminal window, run: `bash Anaconda-latest-Linux-x86_64.sh`. Follow the prompts on the screen. For a detailed reference, please read [this page](#).

## Managing Environment

With Conda, you can create, export, list, and update environments that have different versions of Python and/or packages installed in them. The JupyterLab, which can run in Conda environment, is a web application for computational documents so that our code can produce rich, interactive output.

Below we will demonstrate how to create a Conda environment and install JupyterLab and packages for each tutorial session on macOS/Linux. You need to create a separate Conda environment for each session.

Use the **terminal** for the following steps. For a detailed reference, please read [this page](#).

1. Create an environment with python 3.

```
conda create --name stan python=3.12 -y
```

2. Activate the environment you just created:

```
conda activate stan
```

3. Install JupyterLab:

```
pip install jupyterlab
```

4. Install required packages

```
git clone https://github.com/osmanbeyoglulab/ECCB-2024-Tutorial.git
```

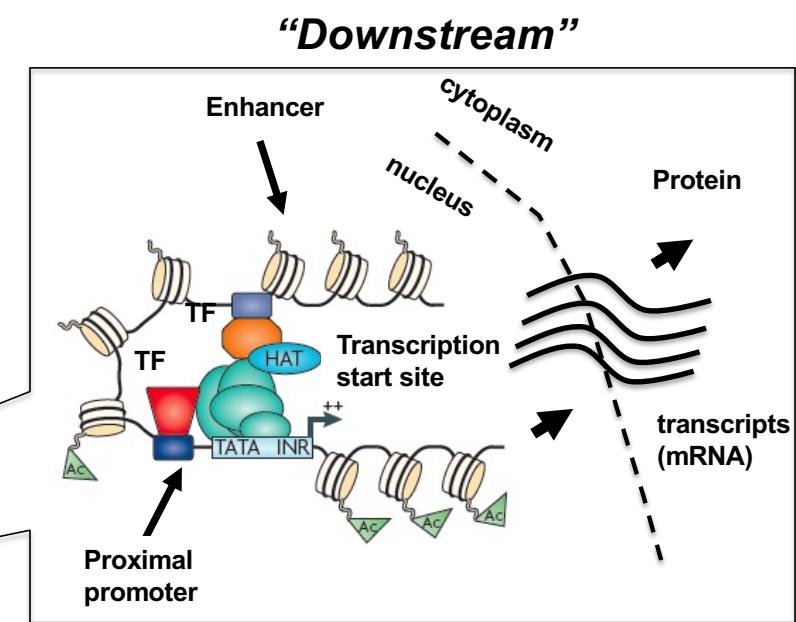
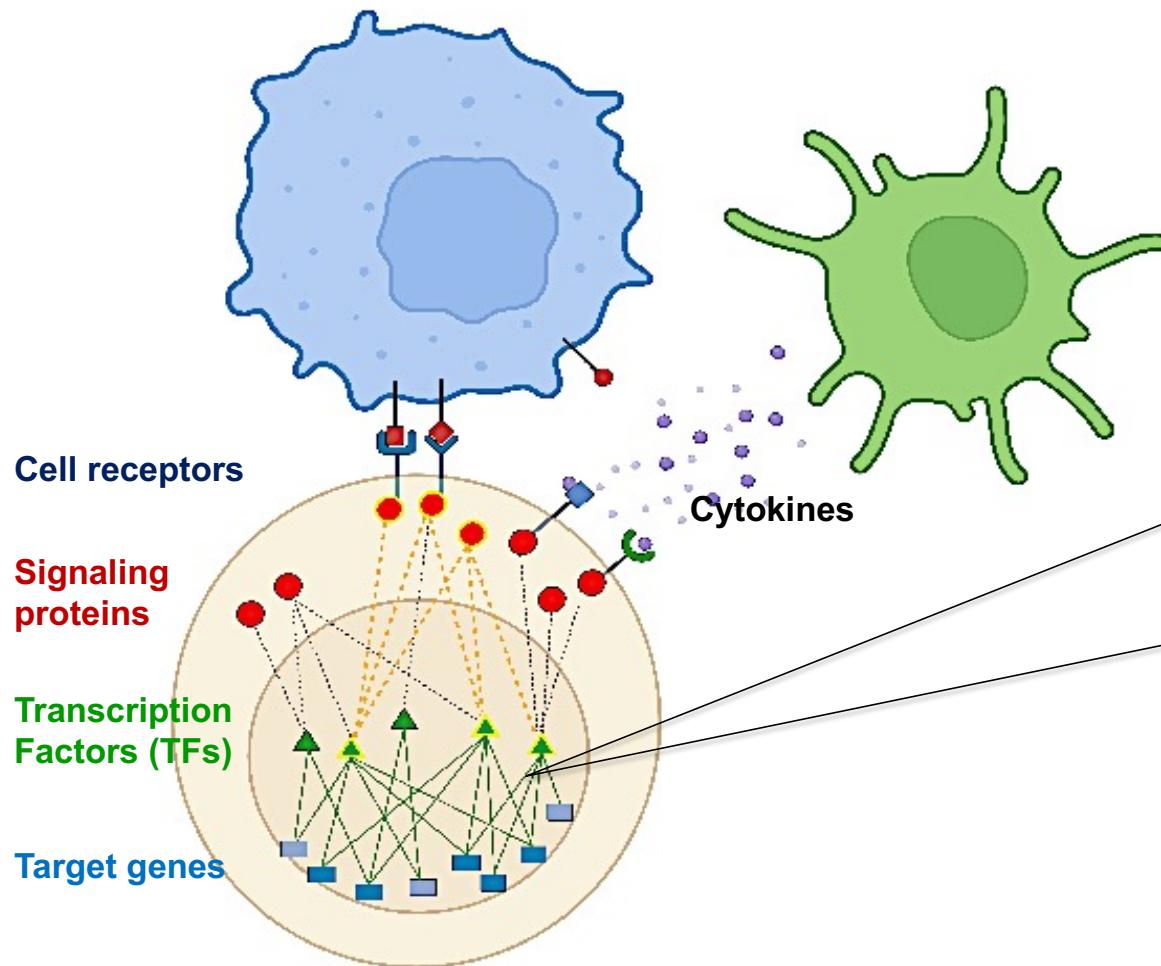
```
cd ECCB-2024-Tutorial
```

```
pip install -r requirements.txt
```

After installing the required packages for the tutorial, launch JupyterLab and open the Jupyter notebook in each session subfolder to start the tutorial. To launch JupyterLab, enter the following command in your terminal:

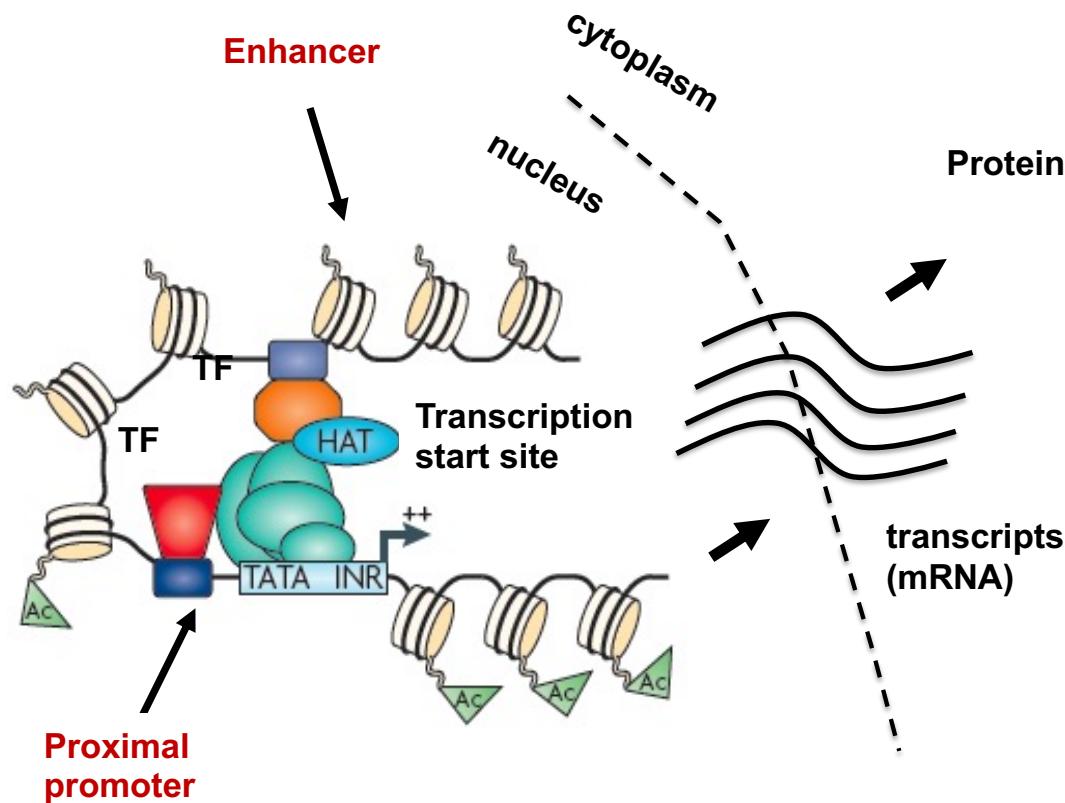
```
Jupyter lab
```

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



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

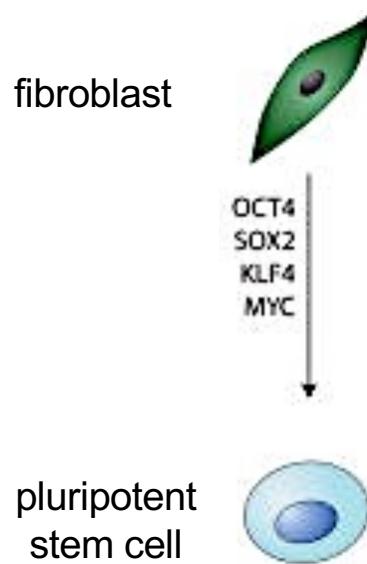
TFs regulate gene expression by binding to specific DNA sequences known as cis-regulatory elements (CREs), such as enhancers or promoters



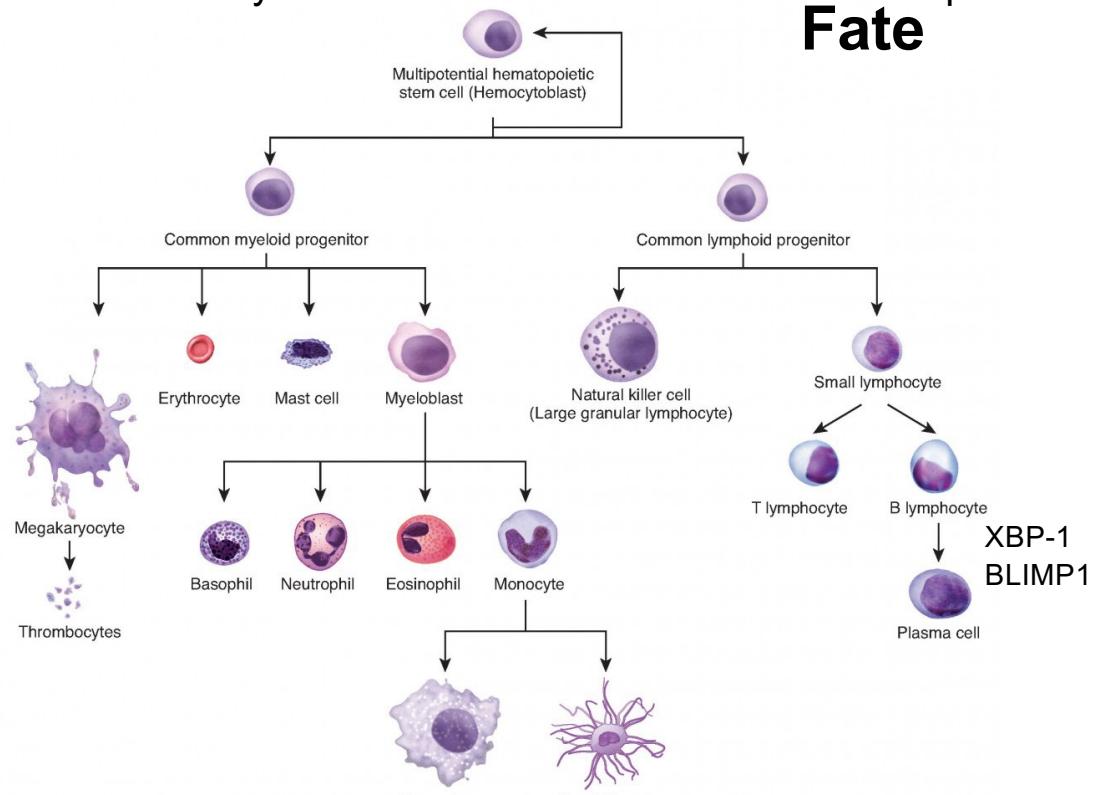
Farnham et al, *Nat Rev Genet.*, 2009

## TFs activate/inhibit genes to affect cell transition from one state to another

- Ectopic expression of the TF factors Oct4, Sox2, Klf4 and c-Myc are sufficient to reprogram fibroblasts into induced pluripotent stem cells



- Hematopoiesis: differentiation of multipotent cells into blood and immune cells
- Governed by a network of TFs that direct cell development

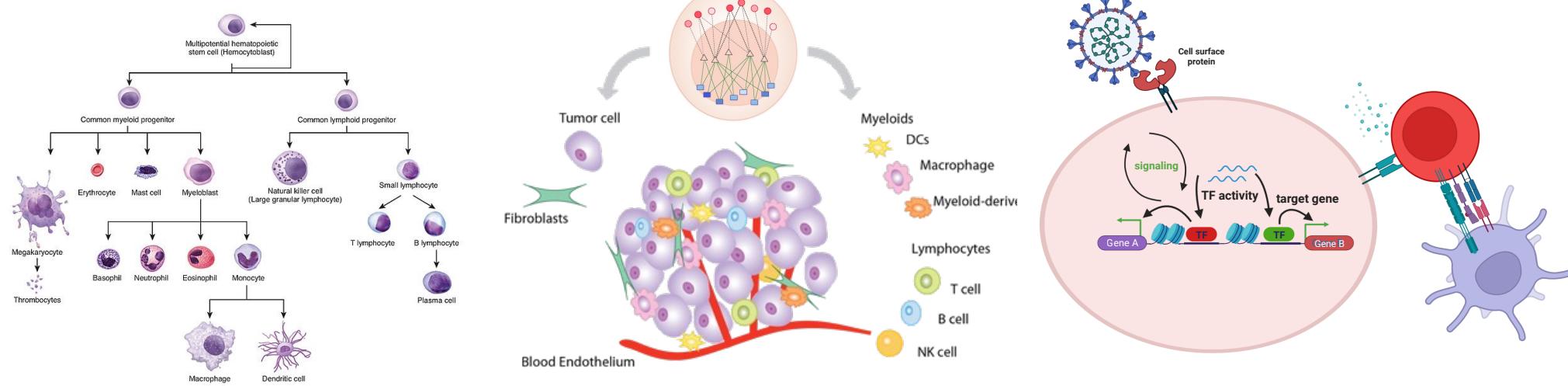


Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013 Mar;152(6):1237-1251. DOI: 10.1016/j.cell.2013.02.014.

Barnes et al, *Clinical Immunology*, 2014

# Significance of TF dynamics in health and disease

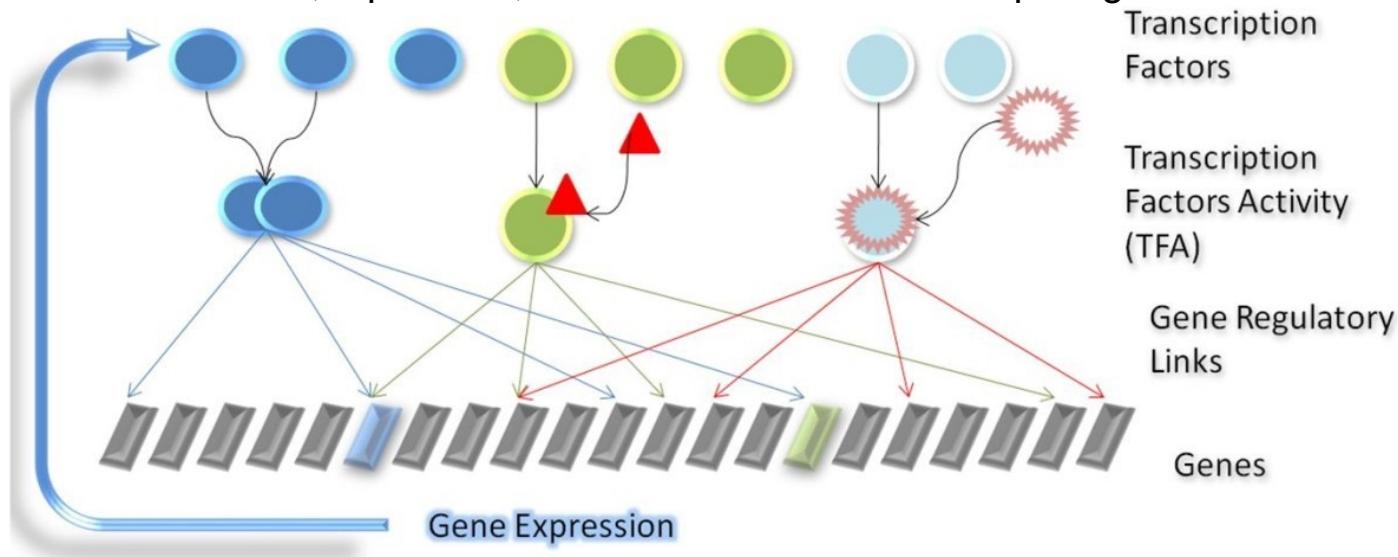
- TFs are essential in defining cellular identity, tissue specificity, disease response, and are influenced by spatial context.
- Understanding context specific TFs aids in developing targeted therapies and biomarkers.



[https://commons.wikimedia.org/wiki/File:0337\\_Hematopoiesis\\_new.jpg](https://commons.wikimedia.org/wiki/File:0337_Hematopoiesis_new.jpg)

## TF activity (TFA)

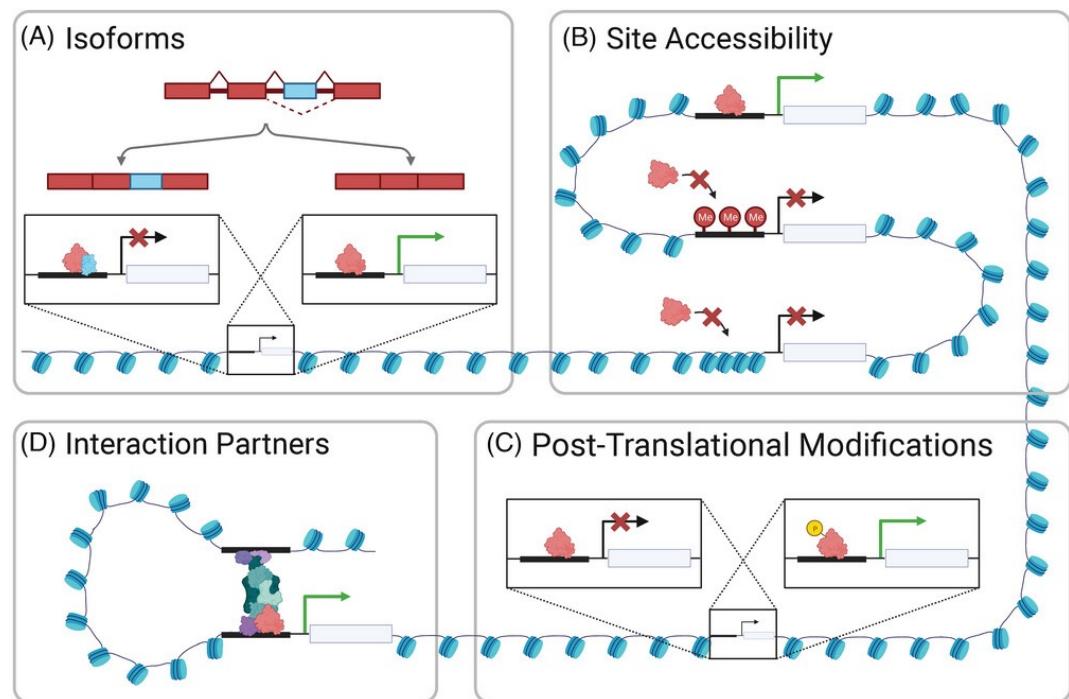
- TFs are expressed in any cell type, although only a small number of them are thought to be sufficient for establishing the cell type-defining gene expression programs
- Cellular processes depend on the expression levels (and activities) of proteins, notably TFs, which can be distinct from mRNA levels.
- TF activity defines the regulatory impact exerted by a TF on its target genes
  - Examples include activation, repression, and effects like alternative splicing



Fu, Y., Jarboe, L.R. & Dickerson, J.A. Reconstructing genome-wide regulatory network of *E. coli* using transcriptome data and predicted transcription factor activities. *BMC Bioinformatics* 12, 233 (2011). <https://doi.org/10.1186/1471-2105-12-233>

# The complexity of TF activity inference

- TF activity 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)

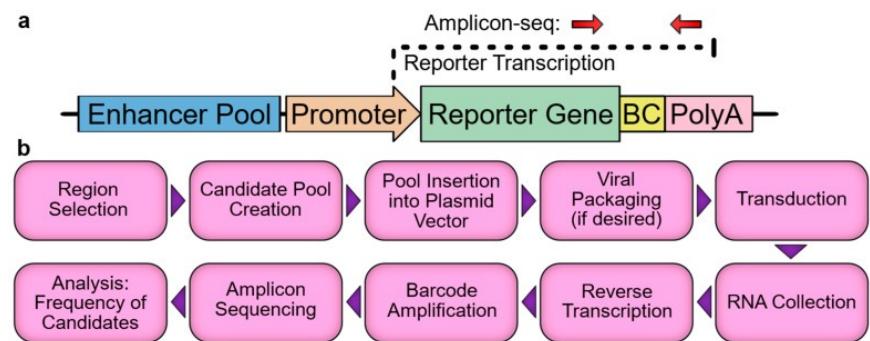
## Experimental techniques for studying context specific transcription factors (1)

- **Gene Knockdown or Overexpression:** To manipulate transcription factor expression and observe downstream effects.
  - Use techniques like siRNA, shRNA, CRISPR-Cas9 for knockdown, or overexpress TFs using expression vectors.
- **Gene Expression Analysis (qRT-PCR, RNA-Seq):** To assess changes in target gene expression in response to altered transcription factor activity.
  - Quantify mRNA levels of target genes upon TF manipulation.
- **Co-immunoprecipitation (Co-IP):** To identify protein-protein interactions involving TFs
  - Immunoprecipitate a TF of interest and identify associated proteins by mass spectrometry or western blotting.
- **ChIP-Seq (Chromatin Immunoprecipitation followed by Sequencing):** To identify genomic regions where a TF binds.
  - Cross-link DNA and proteins, immunoprecipitate the TF-DNA complexes, sequence the DNA, and analyze the binding sites.

## Experimental techniques for studying context specific transcription factors (2)

- **Reporter Assays:** To measure transcriptional activity of a specific TF
  - Fuse the promoter region of a gene of interest to a reporter gene (e.g., luciferase), introduce this construct into cells, and measure reporter gene expression as an indicator of TF activity

### Massively Parallel Reporter Assays



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10146671/>

- Since TFs act cell type-specific and each TF has its own characteristic, untangling their regulatory interactions from an experimental point of view is labor-intensive and challenging
- Due to the complexity and cost of experimental approaches, computational tools are essential

## Single cell and spatial omics allow the investigation of identifying TFs driving cell types, fate and functional states

### Technological advance: Bulk vs single-cell vs spatial resolution

**Bulk (“population of cells”)**  
analysis lose spatial and  
single cell resolution



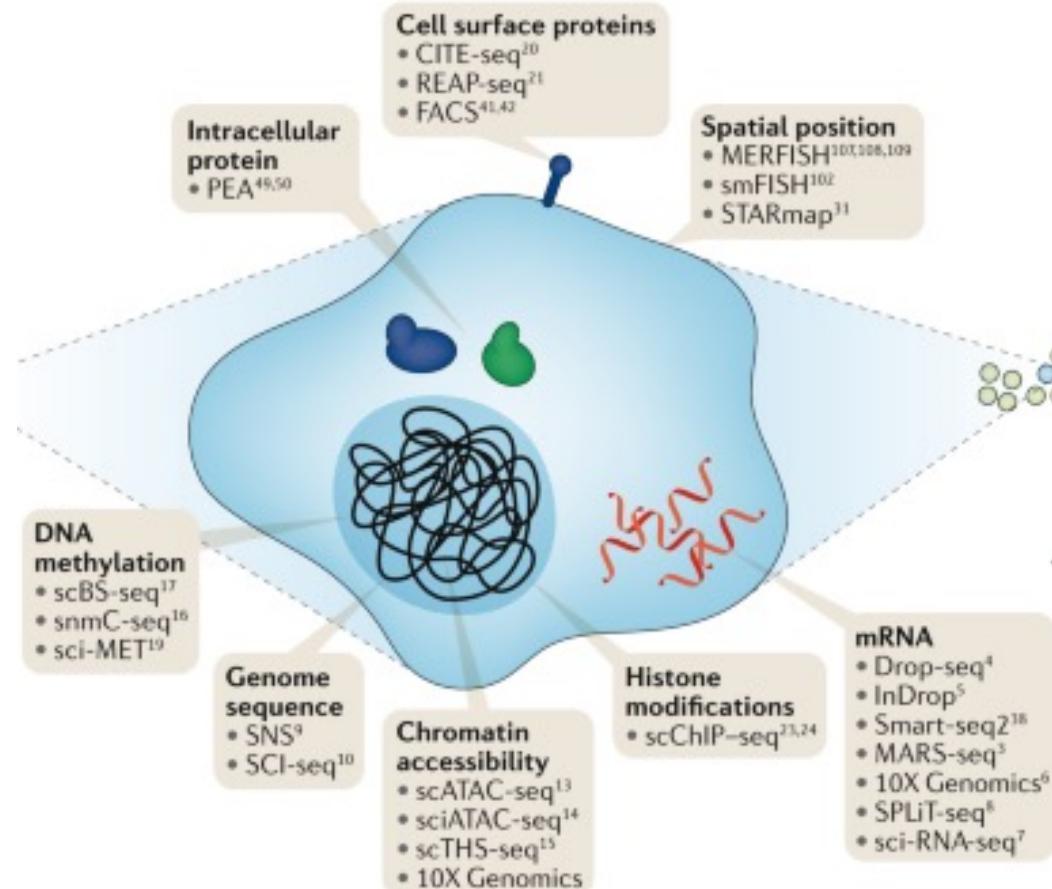
**Single cell analysis**  
retain single cell  
information but lose  
spatial organization



**Spatial analysis**  
provides molecular  
information with spatial  
organization



## Advancement of single/spatial omics technologies enables profiling of molecular features in gene regulation



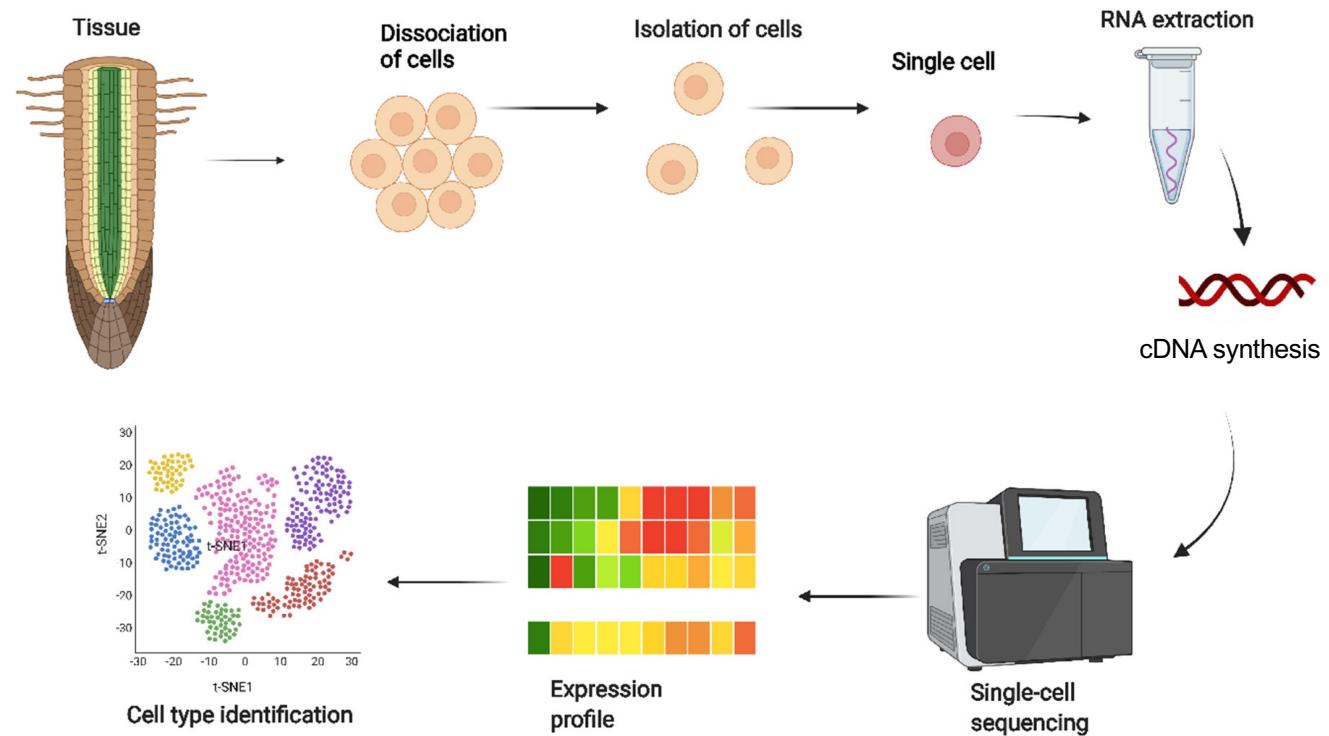
Stuart et al..(2019) “Integrative single-cell...” Nat Rev Genet 20, 257–272.

# **Single-cell/spatial omics technologies for TF activity inference**

- Single cell RNA sequencing (scRNA-seq)
- Spatial transcriptomics (ST)
- Cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)
- Single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq)
- Multiome (scRNA-seq/scATAC-seq)

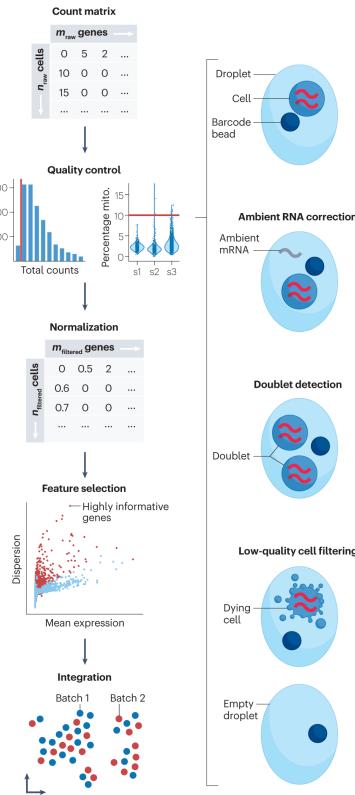
## Single-cell gene expression (scRNA-seq)

- Enable unbiased, high-throughput, and high-resolution transcriptomic analysis of individual cells.
  - 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)



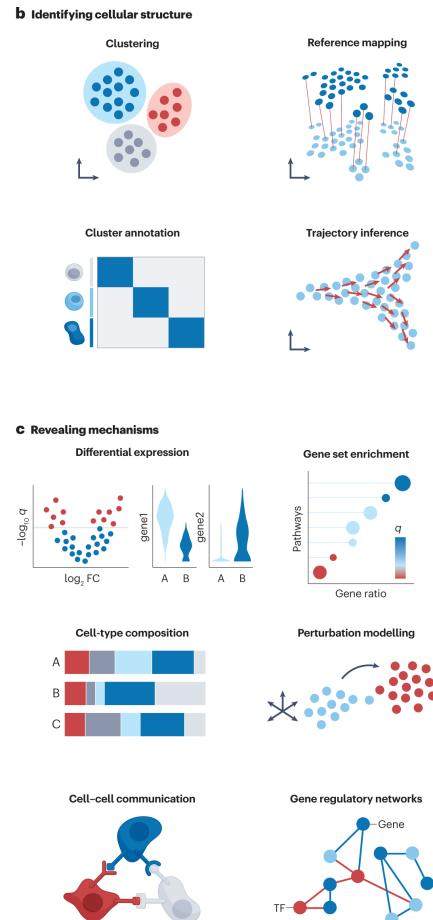
## Data Processing and Quality Control:

- 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.
- 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).

Luecken, M. D. and F. J. Theis (2019). "Current best practices in single-cell RNA-seq analysis: a tutorial." *Mol Syst Biol* 15(6): e8746. 17

# Single cell gene expression data analysis (2)



## Cluster Analysis and Annotation:

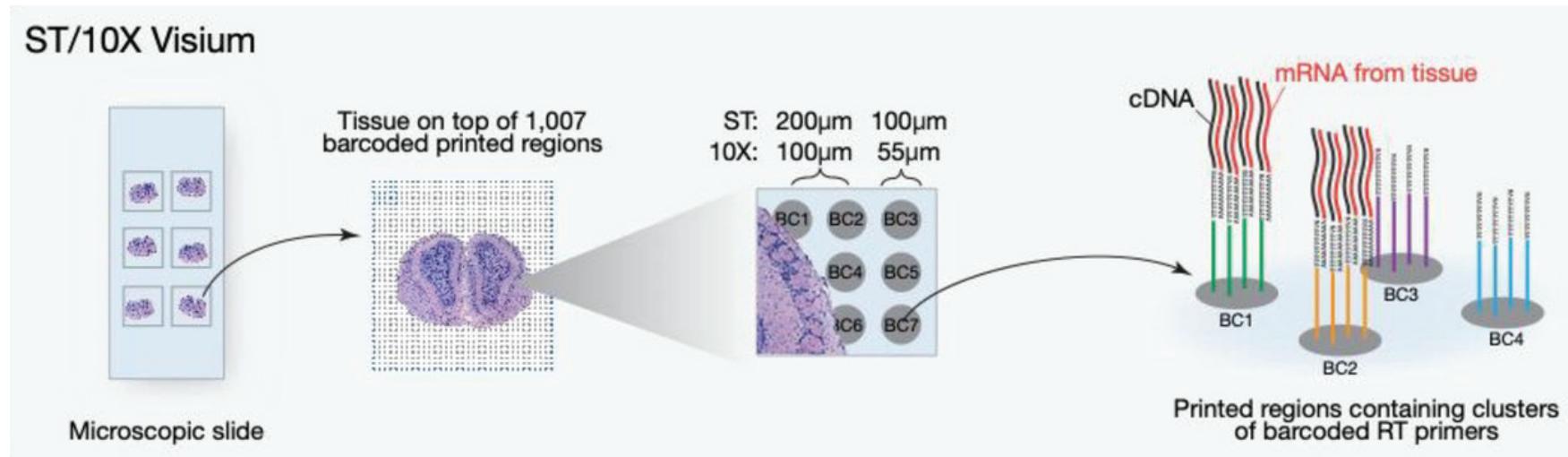
- Data is clustered based on gene expression profiles.
- Clusters are annotated with cell type labels manually or automatically.
- Continuous processes like cell differentiation transitions are inferred.

## Interpretation and Analysis of scRNA-seq Data:

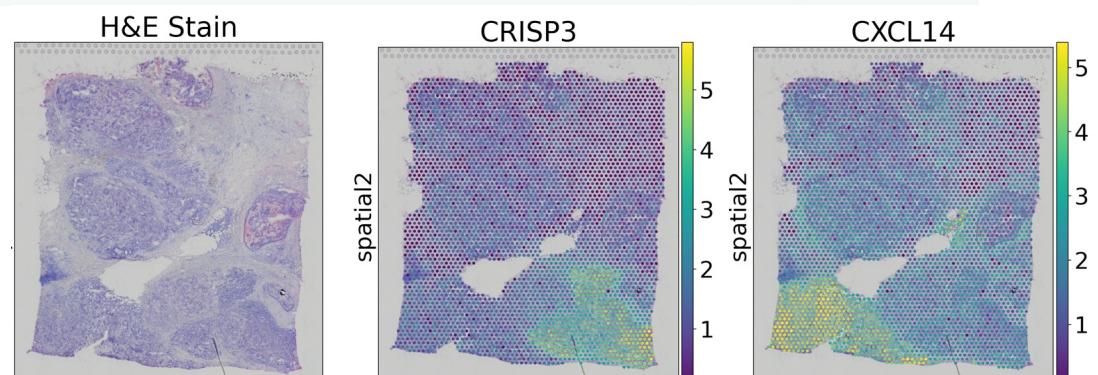
- Differential expression analysis identifies upregulated/downregulated genes.
- Gene set enrichment analysis assesses pathway effects.
- Changes in cell-type composition are examined.
- Perturbation modeling predicts effects of induced or unmeasured perturbations.
- Analysis of ligand-receptor expression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

## Spatially resolved transcriptomics (genome-wide) data

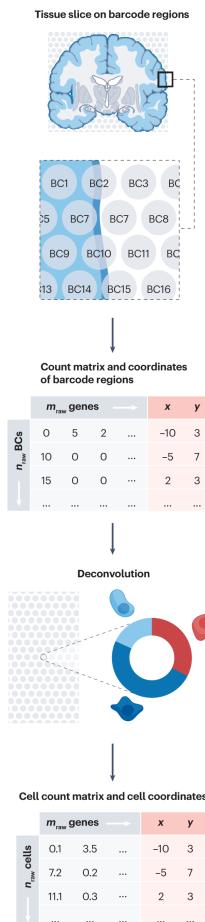
- Spatial transcriptomics measures mRNA across a tissue slice while preserving regional information



- 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



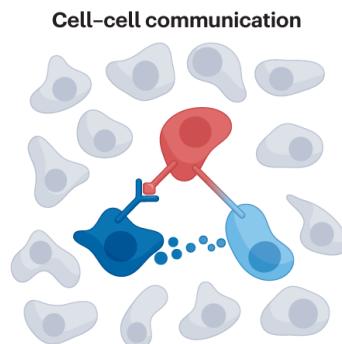
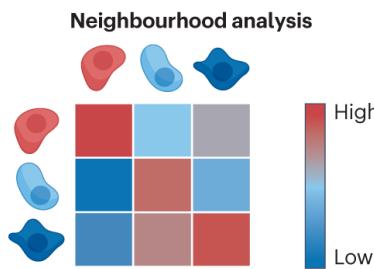
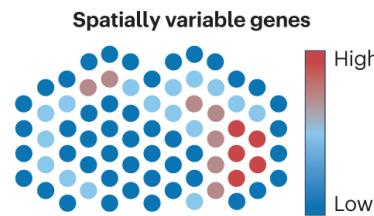
# Spatial gene expression data analysis (1)



## Array-based Spatial Transcriptomics:

- Quantifies gene expression in predefined barcoded (BC) regions spanning 10  $\mu\text{m}$  to 200  $\mu\text{m}$ .
- 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.

## Spatial gene expression data analysis (2)

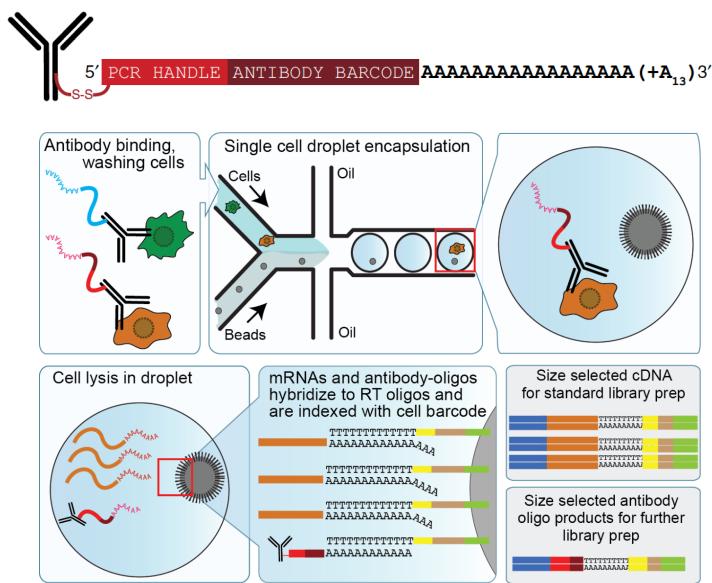


- Analyzes mechanisms via spatial positions, identifying spatially varying genes.
- Examines cell neighborhoods and infers communication events (receptors, ligands, tight junctions, mechanical and indirect mechanisms).

# 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

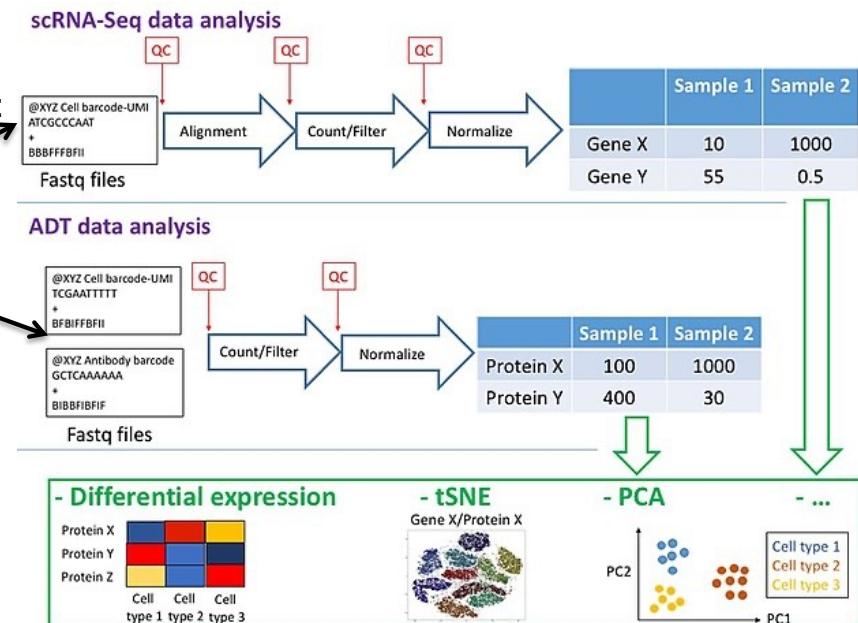
- Uses DNA-barcoded antibodies with polyA tail to convert detection of proteins into a quantitative, sequenceable readout
- Antibody-derived tags (ADTs) are antibody clones with unique barcodes attached to poly(A) sequences and a PCR handle.
- They bind to surface proteins, and their sequenced counts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



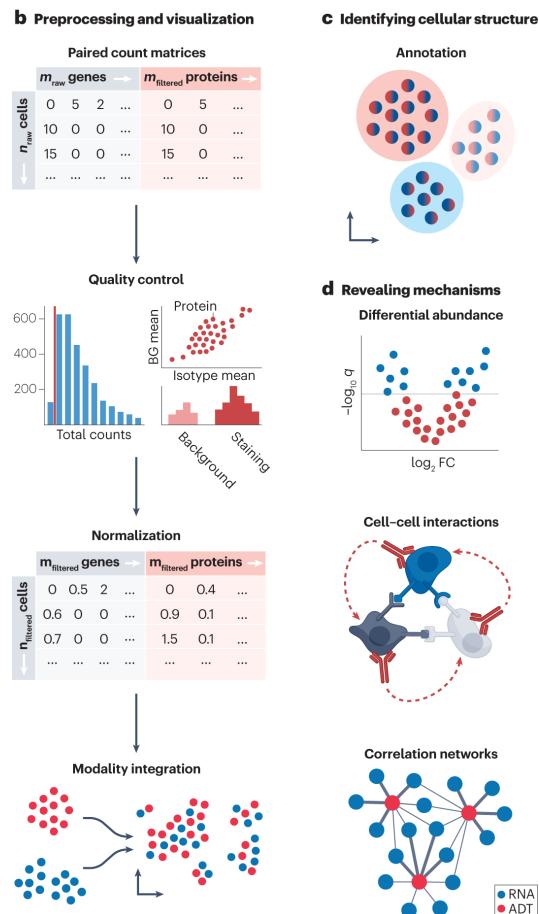
<https://cite-seq.com>

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

“Complex readout high drop-out rates”  
“limited set”



# CITE-seq analysis steps



## Integration of ADT and Gene Expression Data:

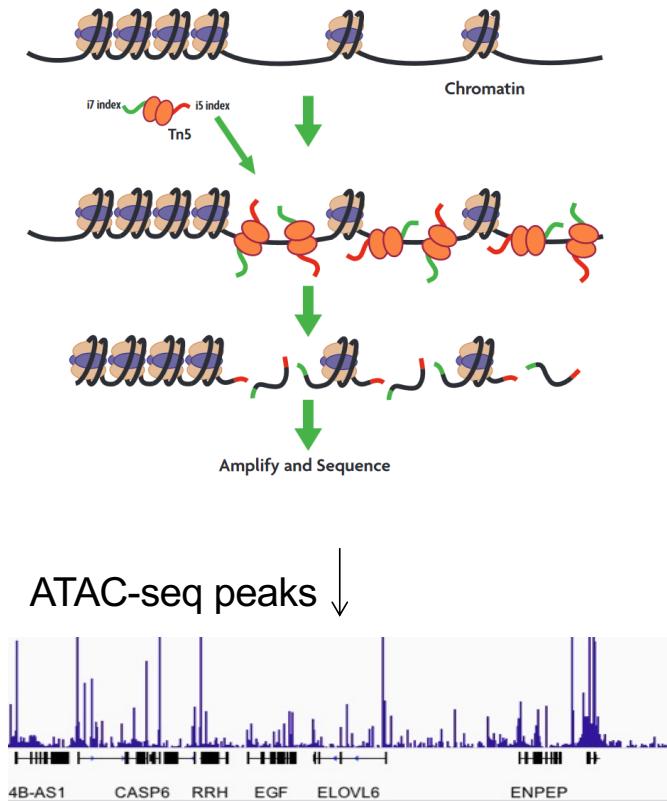
- Each dataset undergoes individual quality control and normalization.
- Data can be visualized individually or jointly to reveal relationships.
- Annotation can be based on transcriptomics data, ADT data, or both.
- Clusters are matched to marker genes and ADTs to annotate cell types accurately.

## Biological Mechanisms and Analysis:

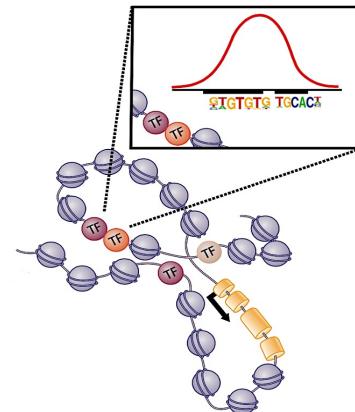
- ADT data is analyzed for differential abundance to understand protein expression changes.
- Cell-cell communication is inferred based on ADT profiles.
- Correlation networks integrate RNA and ADT information to elucidate biological interactions.

## Single cell chromatin accessibility

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



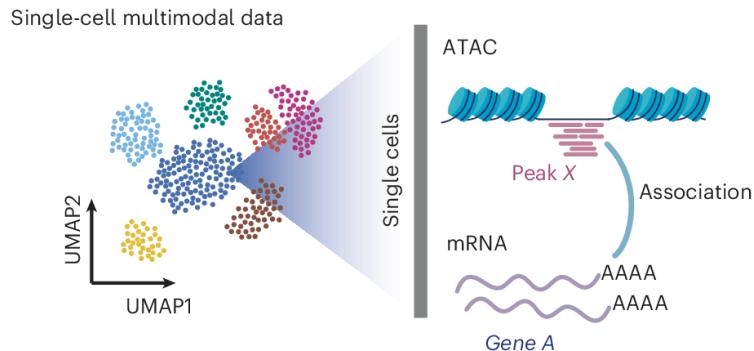
TF motif prediction in ATAC-seq peak regions



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

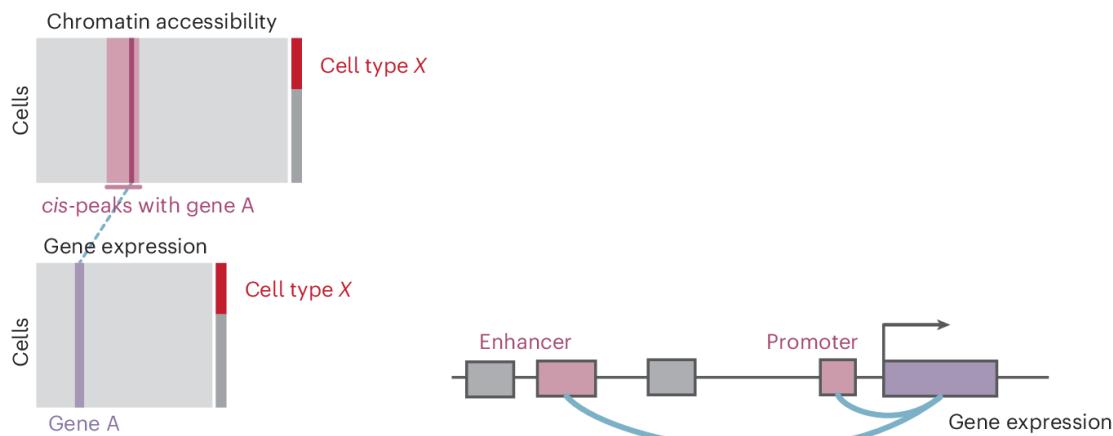
# Single-cell multiome

a



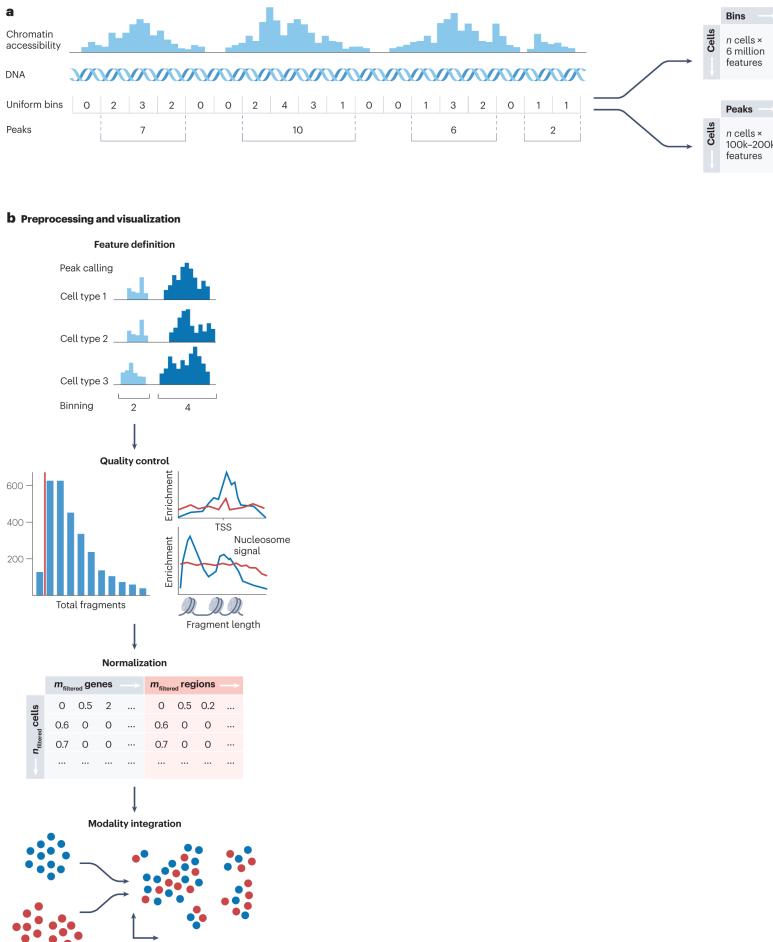
- Combines single-cell gene expression analysis with single-cell open chromatin mapping to provide genome-wide mapping of both the transcriptional and epigenetic landscapes at the single-cell level

b



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)

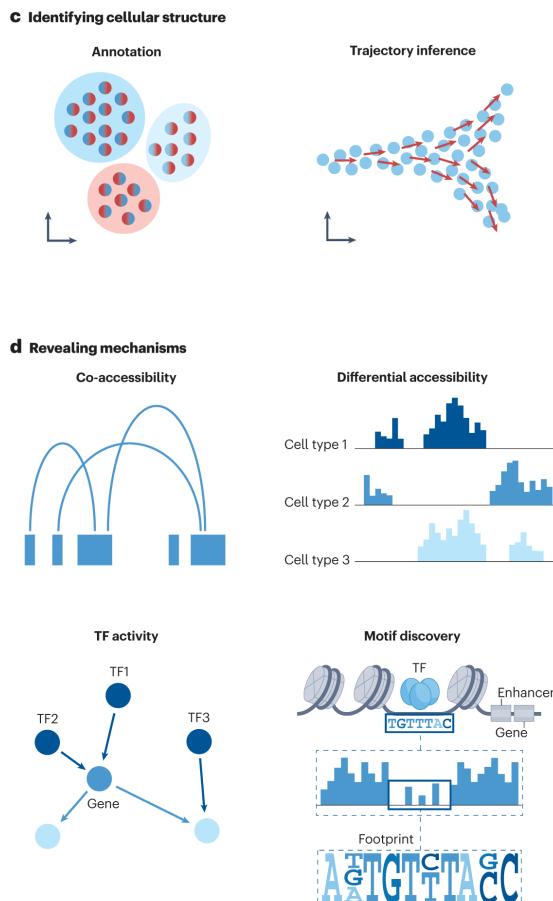


### Quality Control and Preprocessing:

- 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)



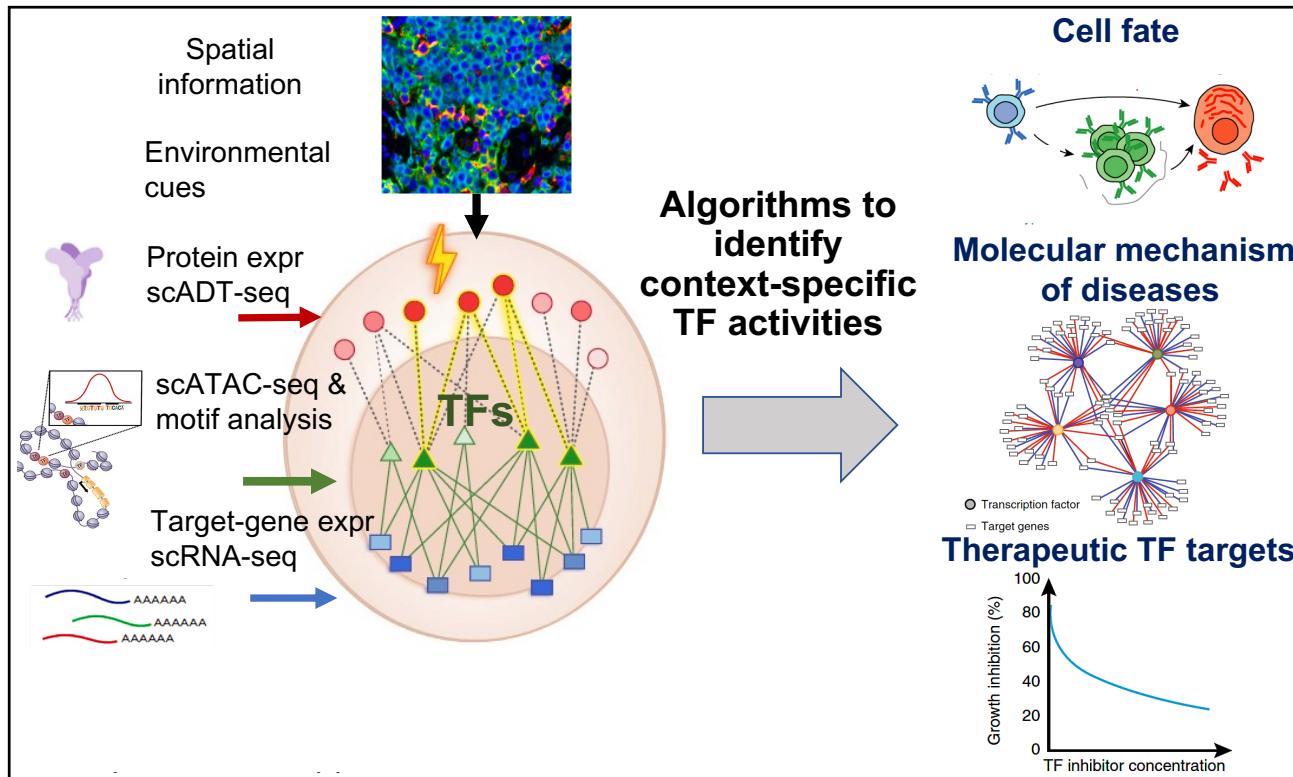
### Annotation, Cell Type Identification ana Data Analysis:

- 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.

# **Overview of computational TF inference methods based on single cell/spatial omics**

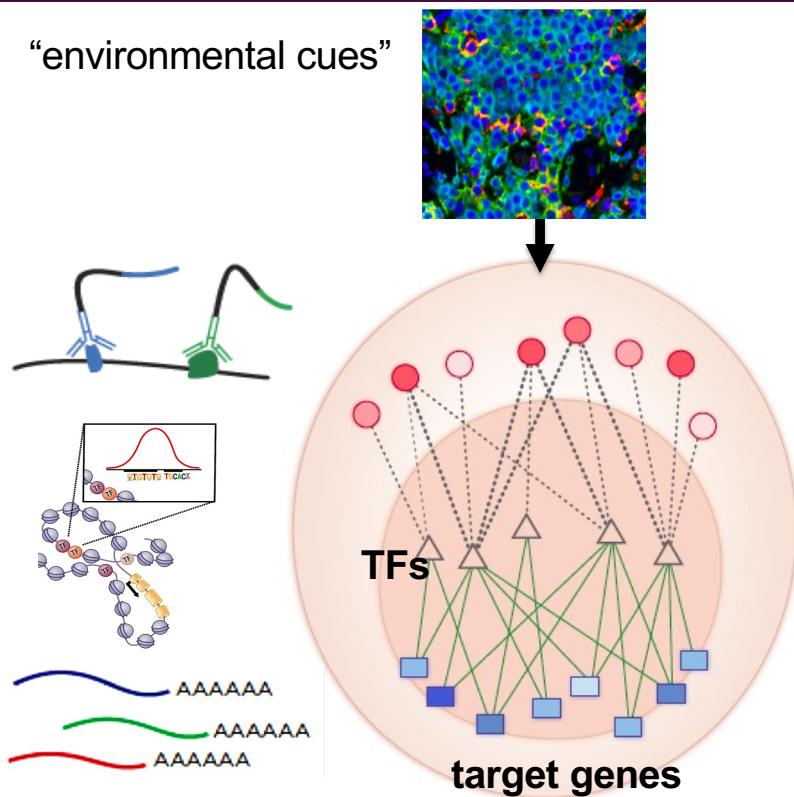
# Computational methods for modeling context-specific transcription factor activities using single-cell and spatial omics approaches

- Advancement of single/spatial omics technologies enables profiling of molecular features in gene regulation
- 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

“environmental cues”



**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+

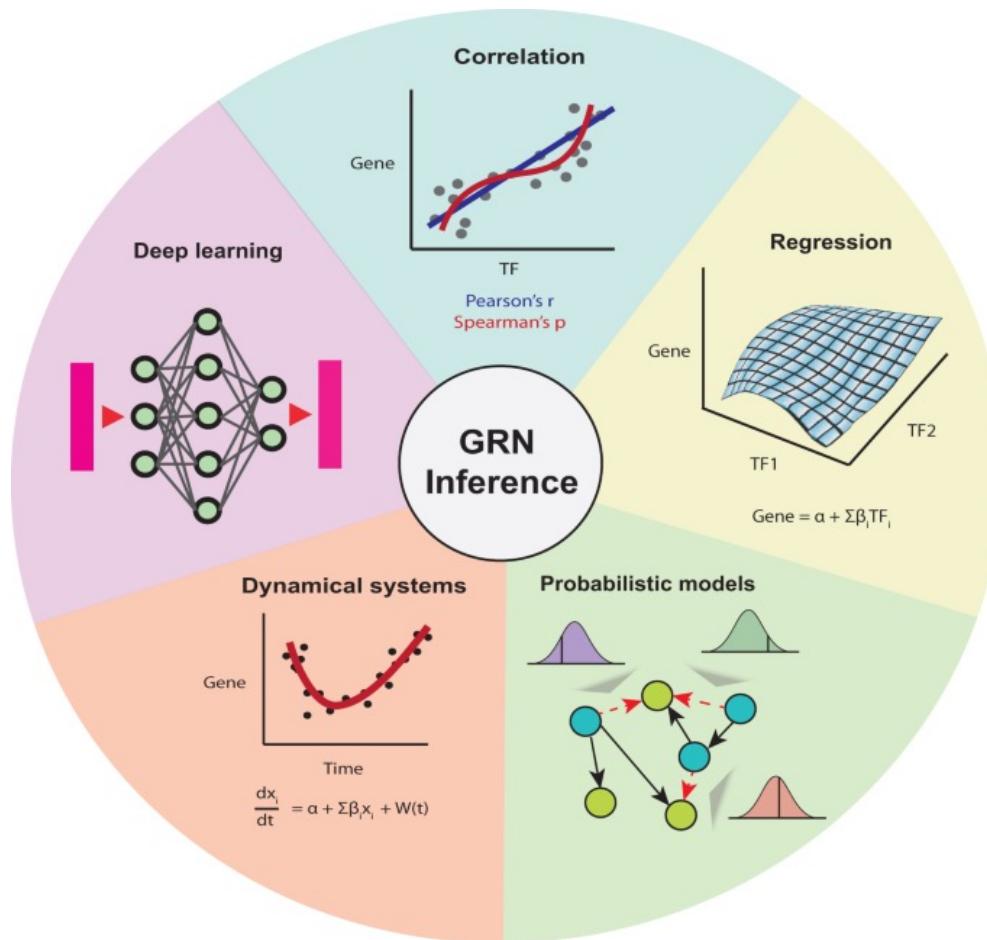
**TF activity inference from single-cell epigenomic data**

e.g. BITFAM, chromVAR, scBAsset, scFAN

**TF activity inference from single-cell gene expression data**

e.g. SCENIC, BITFAM, metaVIPER, INFERELATOR 3.0

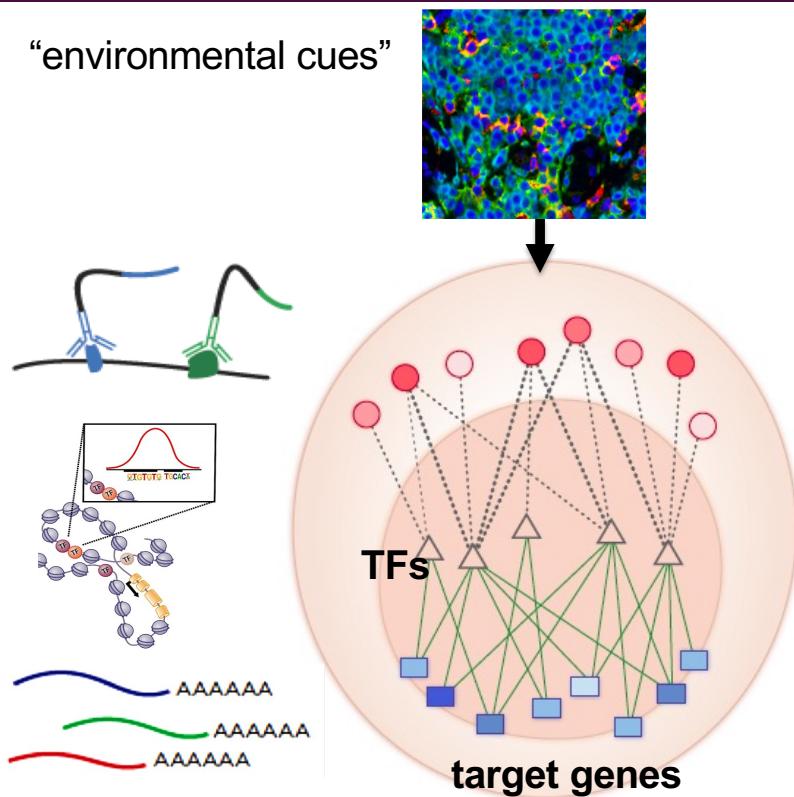
## The major classes of methods for TF activity inference



- Correlation-Based Methods:
  - identify pairs of variables (e.g., TF expression, gene expression, CRE accessibility) that vary together.
- Regression-Based Approaches:
  - Model gene expression using multiple predictors (e.g., TF expression, CRE accessibility).
- Probabilistic Models:
  - Determine the most likely regulators for a gene.
- Dynamical Systems-Based Approaches:
  - Model gene expression changes based on biological factors (e.g., TF expression, cell cycle stage, stochasticity).
- Deep Learning-Based Approaches:
  - Use neural networks to uncover complex relationships between TFs, CREs, genes, and cells.

# Computational methods for TF activity inference based on single cell omics

“environmental cues”



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

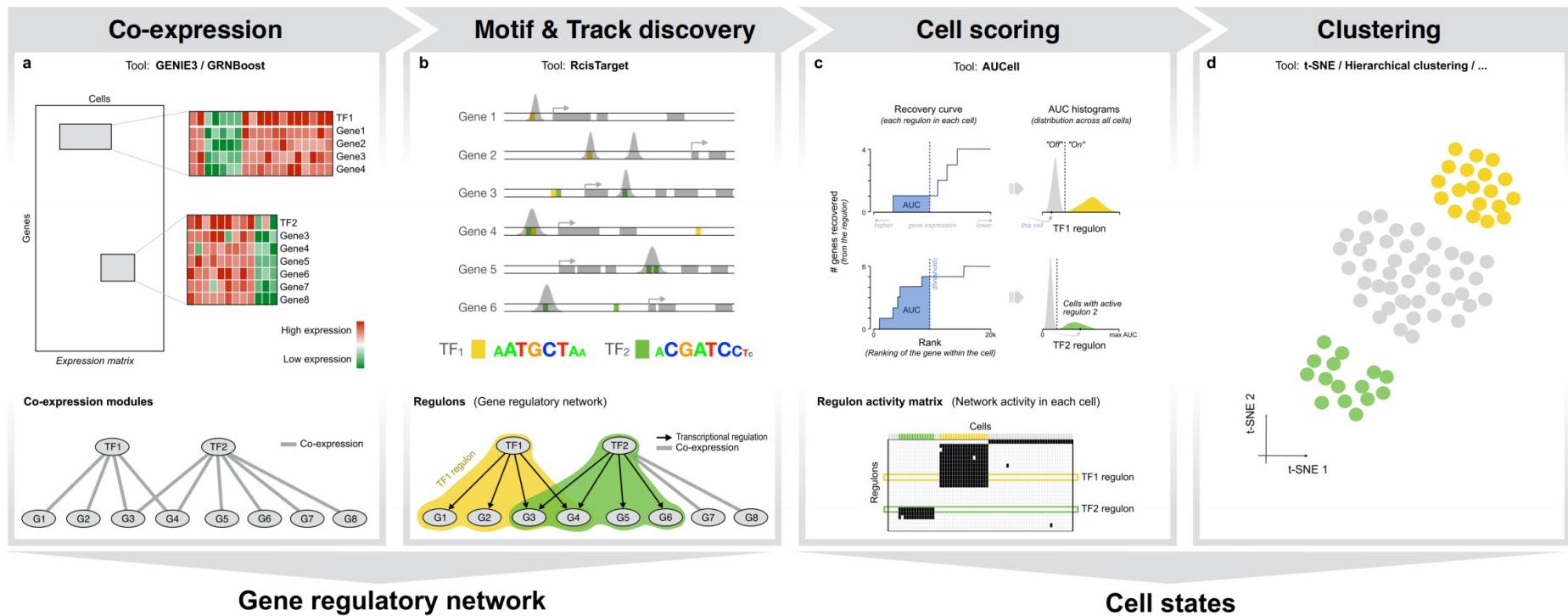
e.g. BITFAM, chromVAR, scBAsset, scFAN

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)

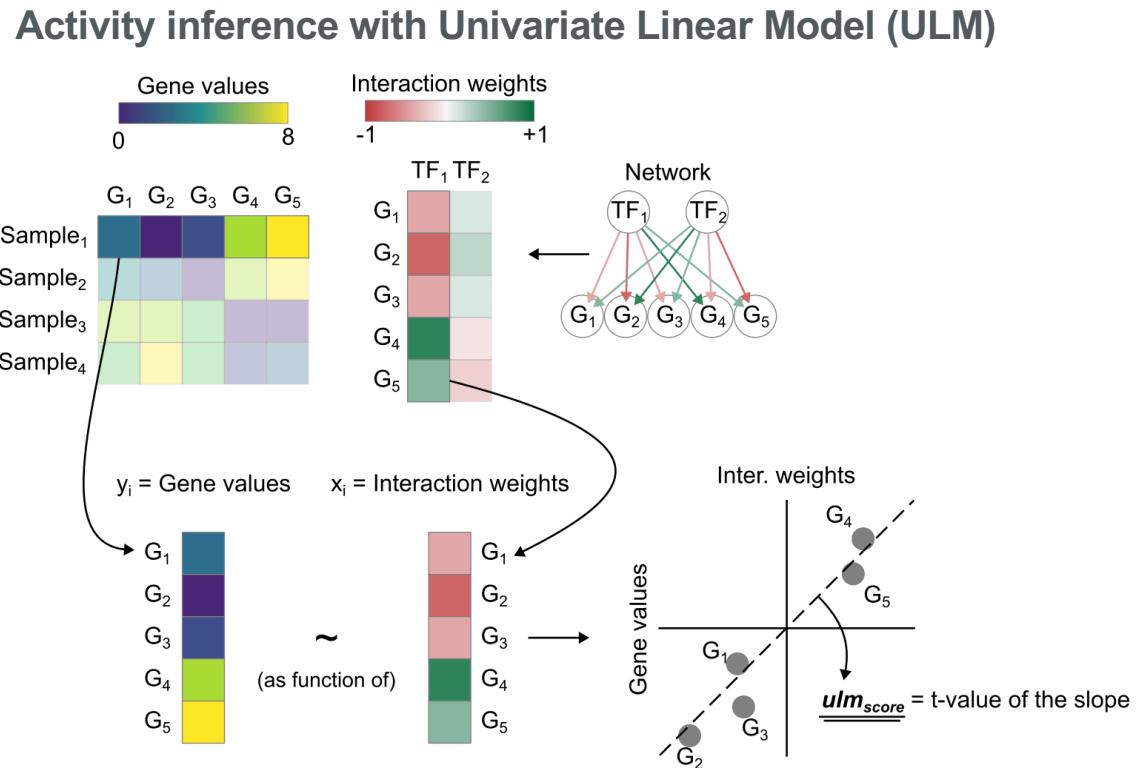


<https://github.com/aertslab/SCENIC>  
<https://github.com/aertslab/SCENICprotocol>

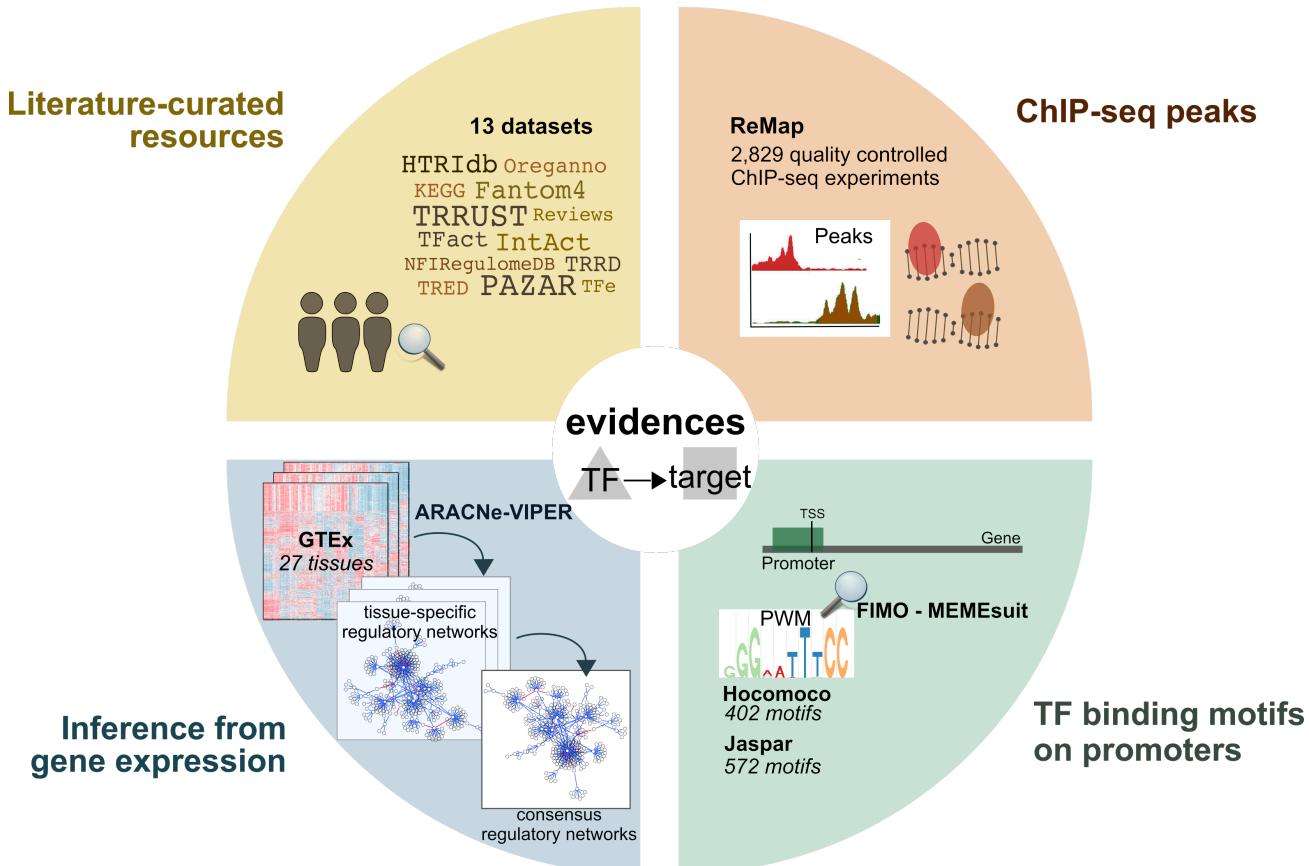
Nature Methods 14, 1083 (2017)

## 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>



## TF – target gene priors



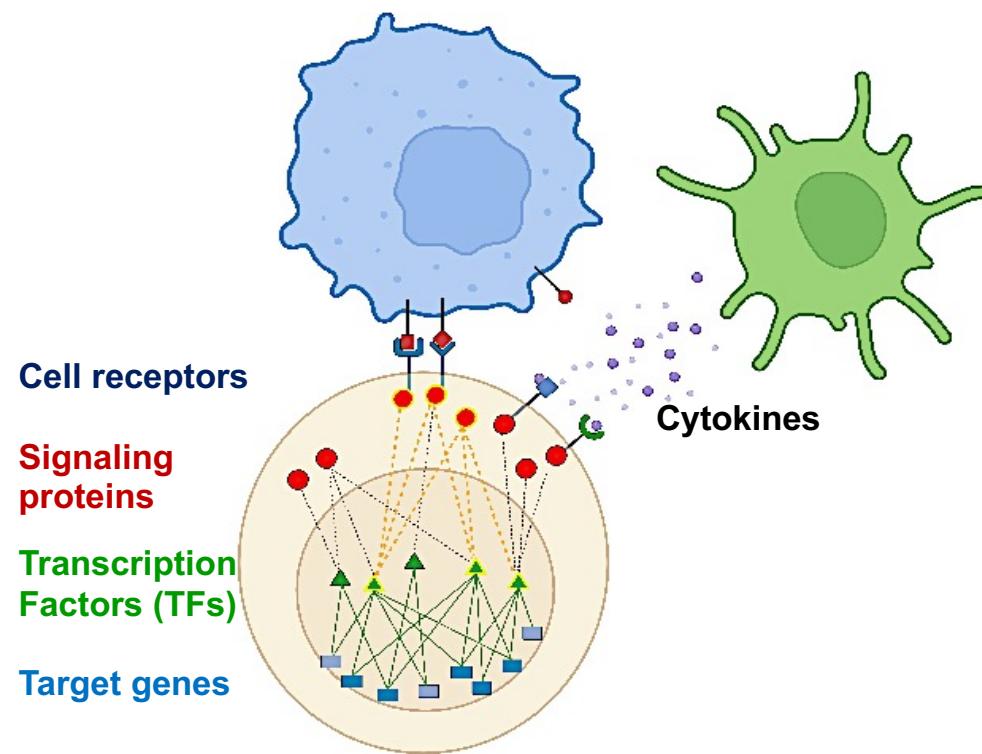
<https://saezlab.github.io/dorothea/>

- 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

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

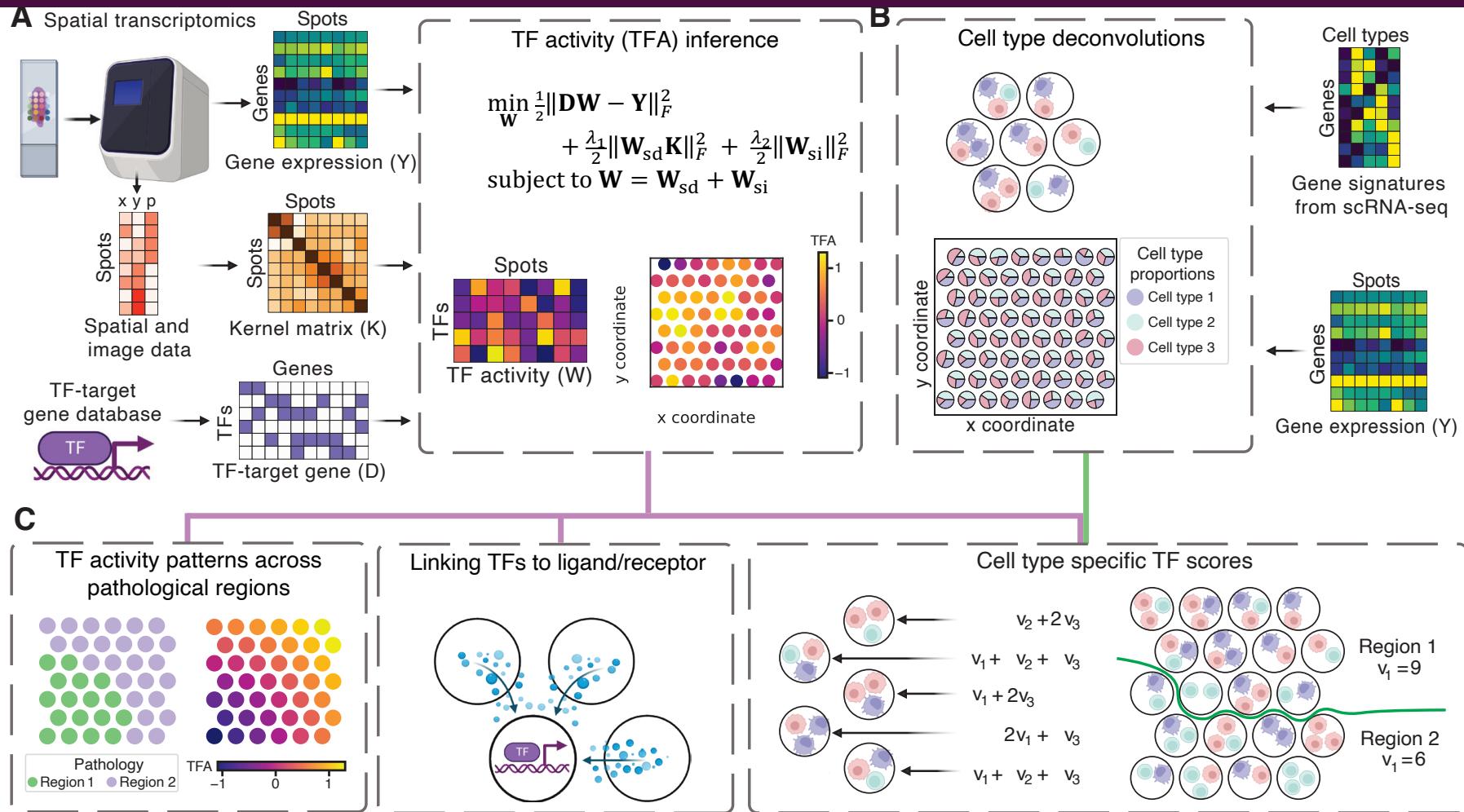
- INFERELATOR 3.0
  - <https://inferelator.readthedocs.io/en/latest/tutorial.html>
- BITFAM: The Bayesian inference transcription factor activity model
  - <https://github.com/jaleesr/BITFAM>
- metaVIPER
  - <https://github.com/califano-lab/single-cell-pipeline>

## TF activity inference from spatial transcriptomics

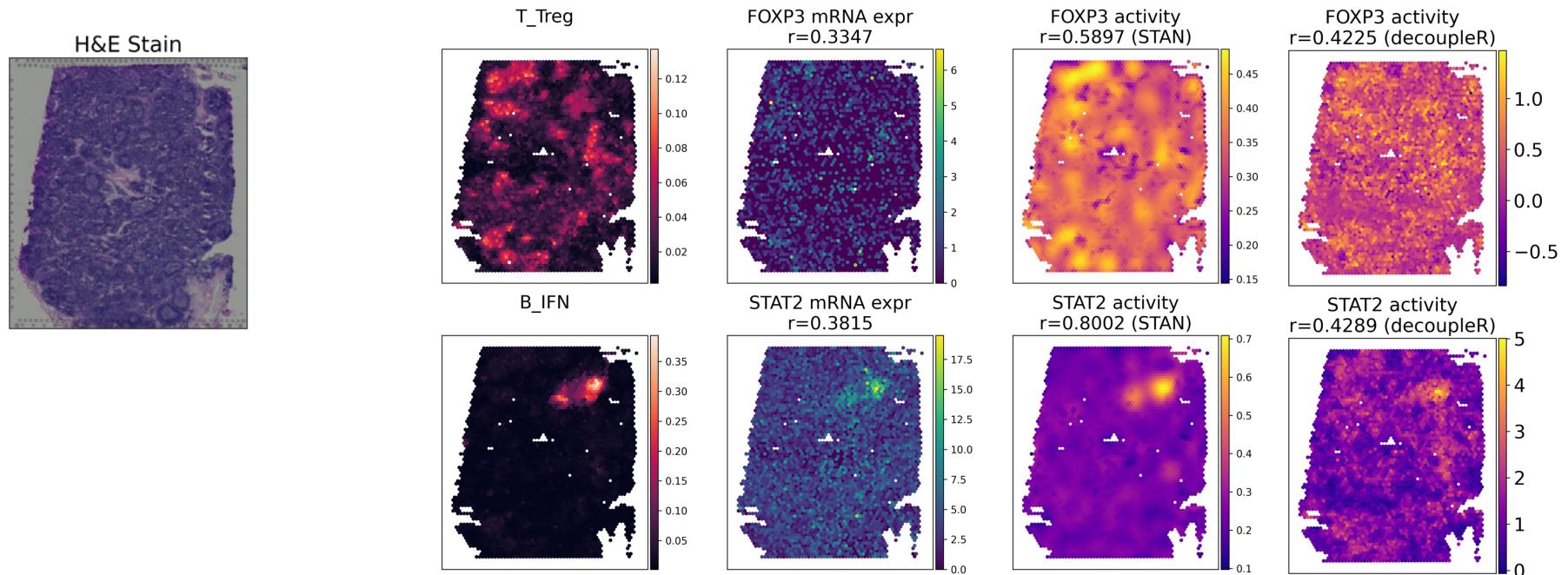


Nearby cells often affect cell context-specific TF activities through receptor-ligand mediated cell-cell communication or secreting molecules like cytokines

# Spatially informed Transcription factor Activity Network (STAN)

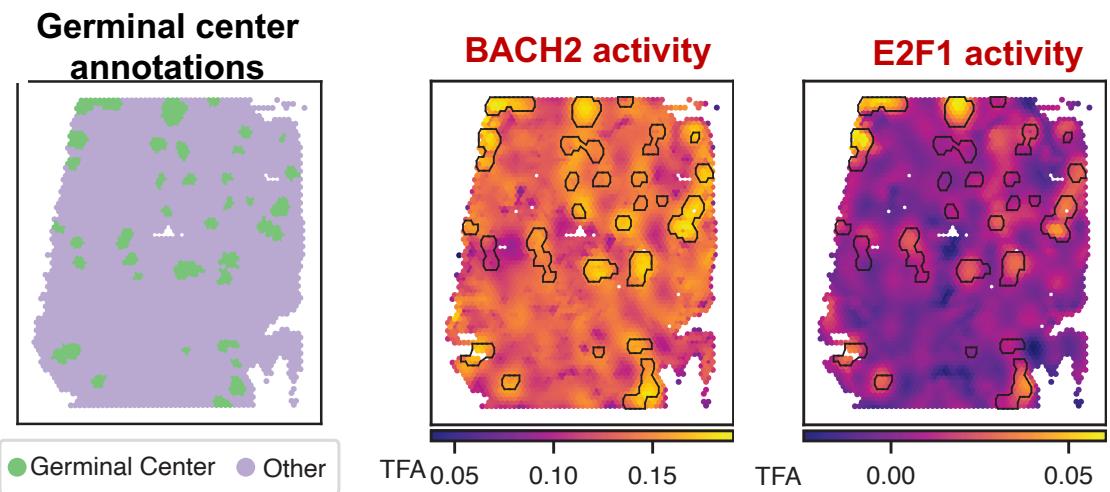
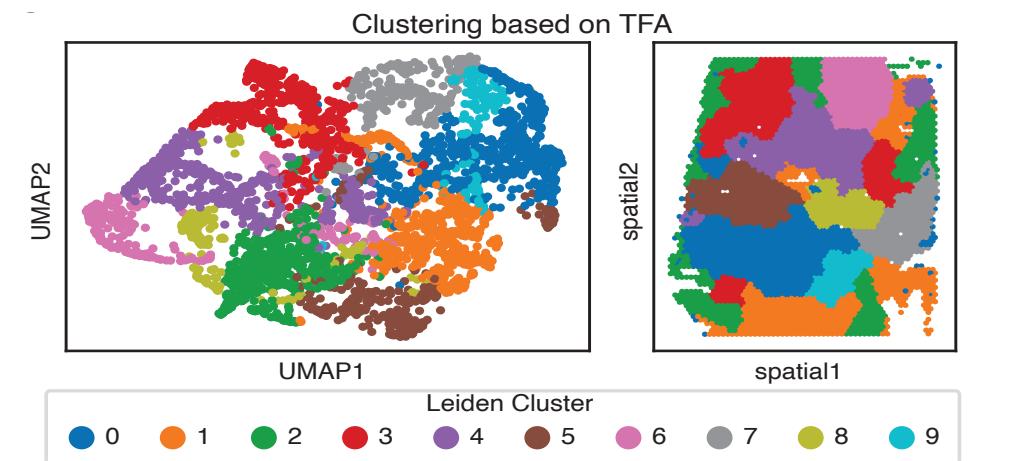
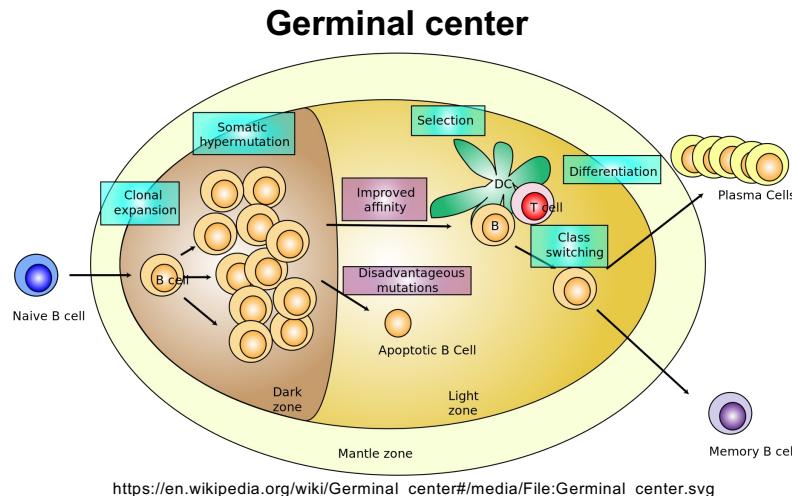


## STAN identifies cell-type specific TFs – lymph node

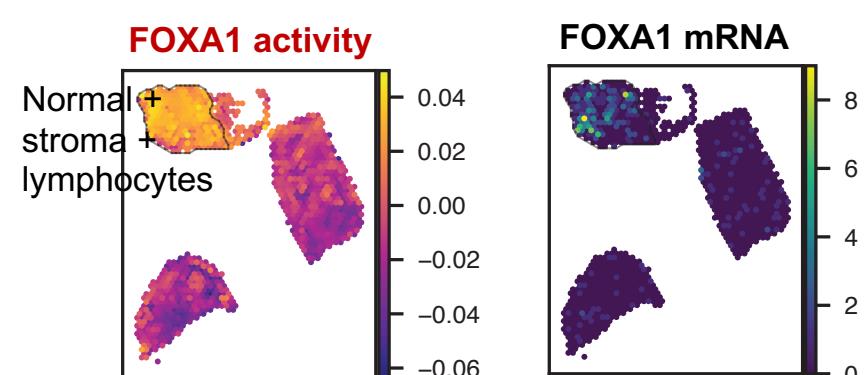
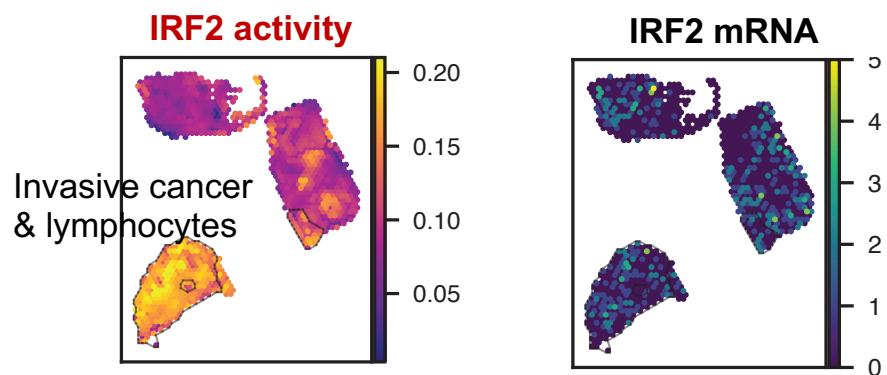
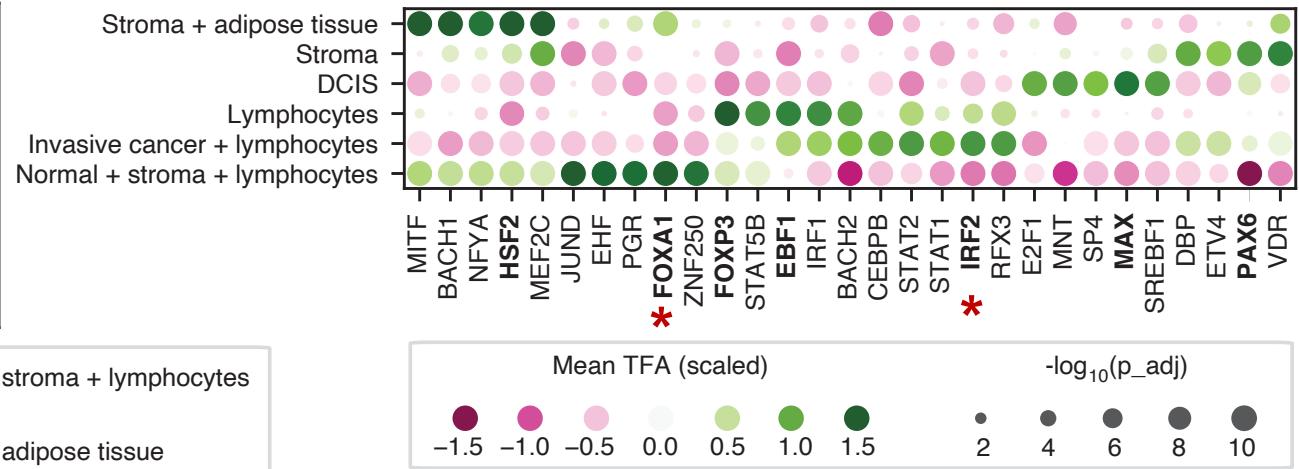
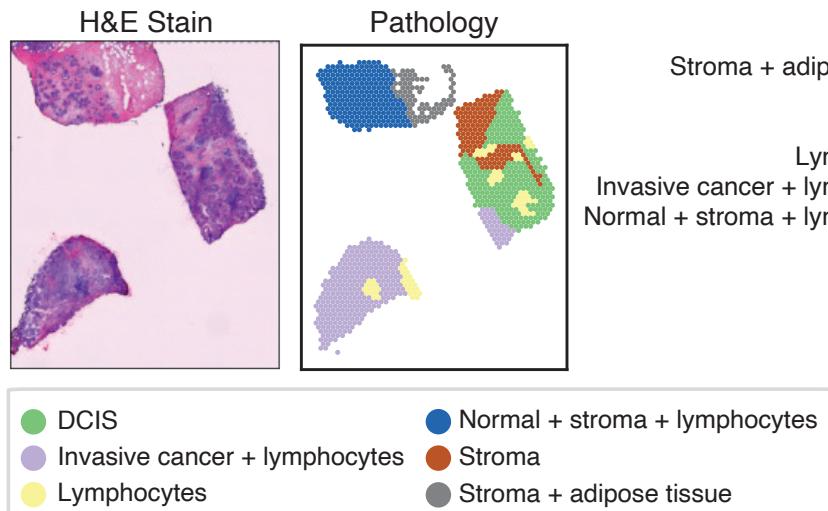


# Clustering on STAN predicted TF activities (TFAs) yields spatially coherent clustering in lymph node

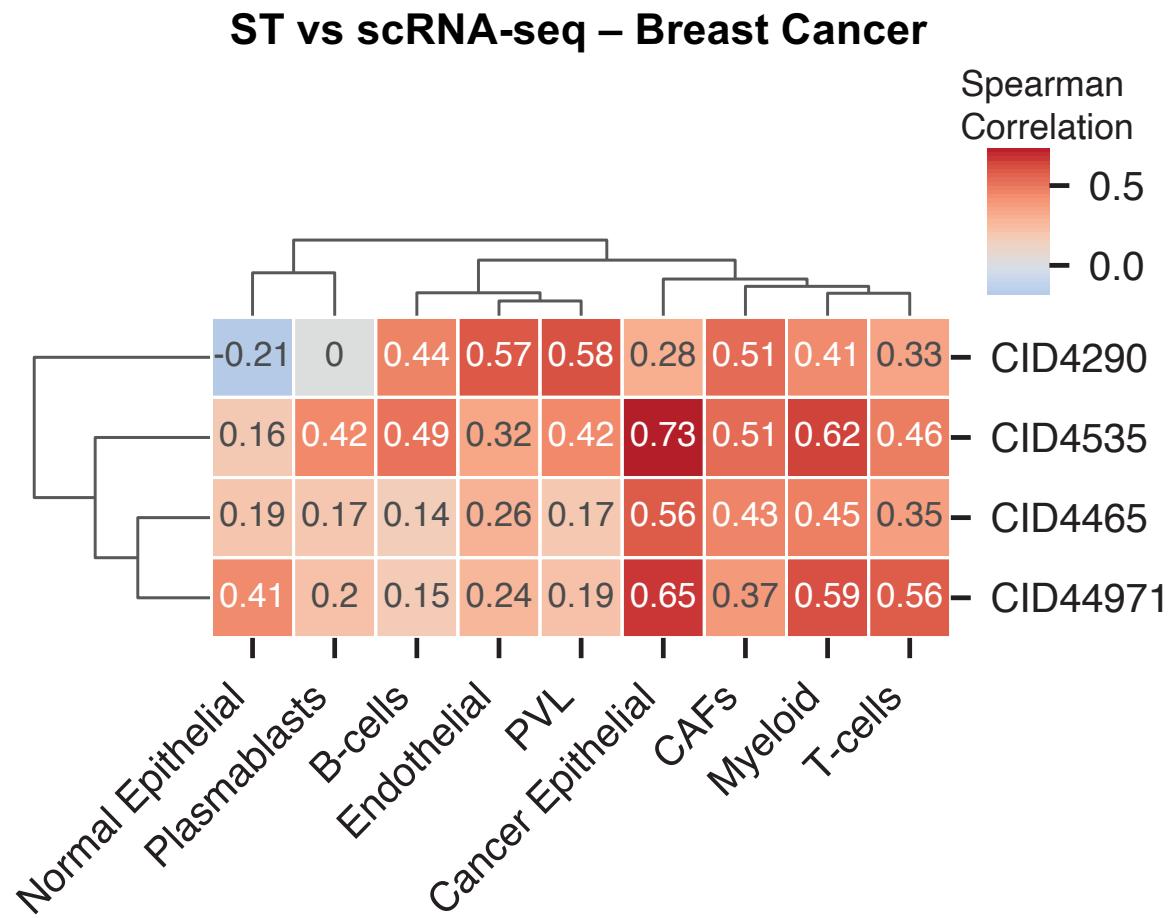
- The human lymph node is characterized by dynamic microenvironments with many spatially interlaced cell populations such as germinal centers (GCs)
- Germinal centers develop in the B cell follicles of secondary lymphoid tissues during T cell-dependent antibody responses



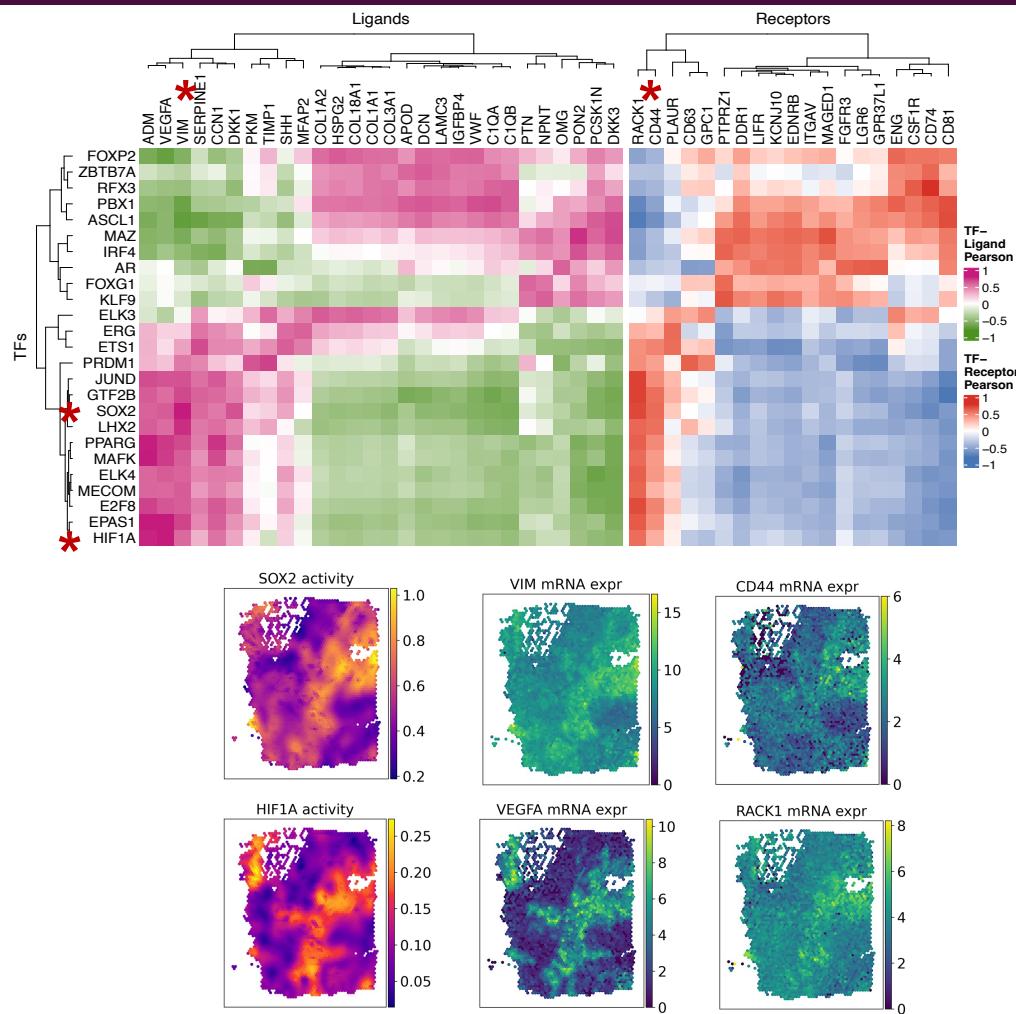
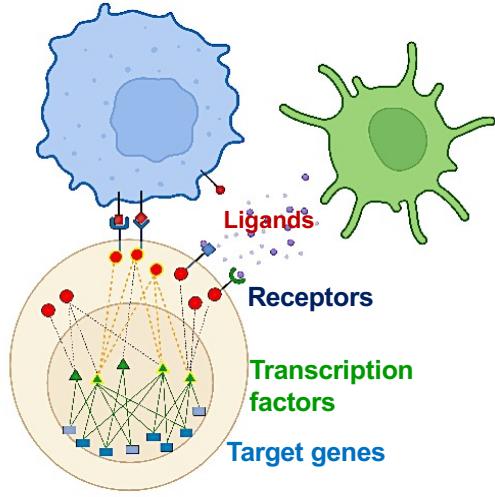
# Inferred STAN TF activity associated with pathological regions – triple negative breast cancer (TNBC)



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

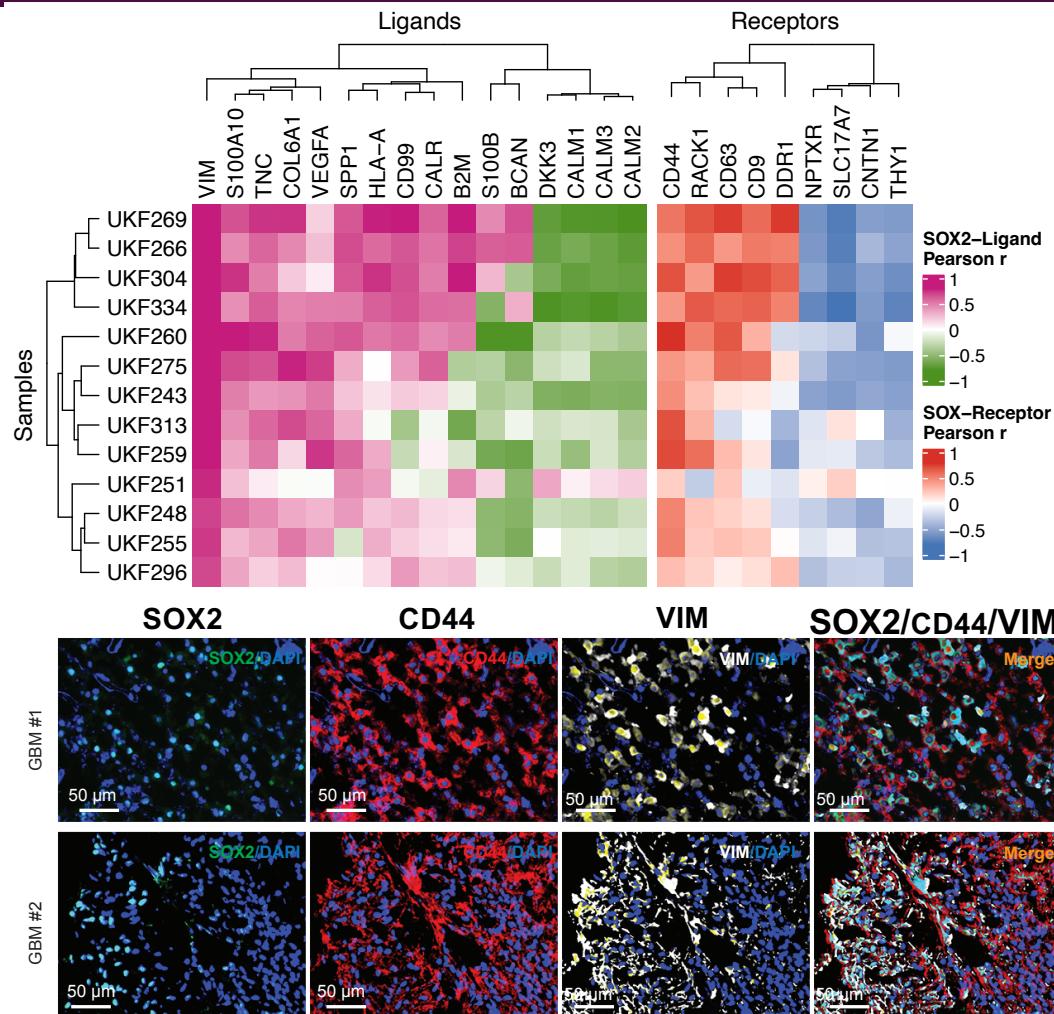


## Linking ligands and receptors to TFs in glioblastoma



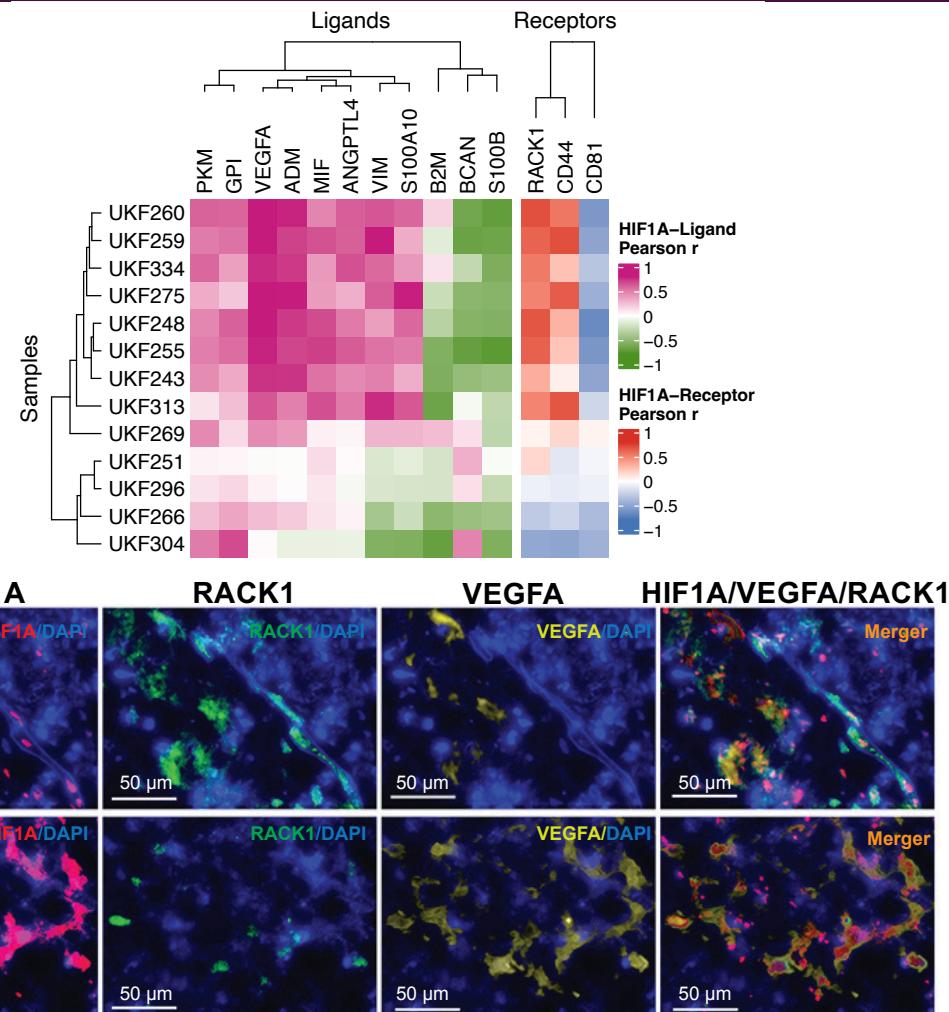
## Correlation between inferred SOX2 activity and ligand/receptor expression across glioblastoma samples

SOX2 has been extensively implicated in promoting malignancy in glioblastoma



# Correlation between inferred HIF1A activity and ligand/receptor expression across glioblastoma samples

The activity of HIF1A correlates with increased tumor grade and reduced overall survival in glioblastoma patients



## 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>

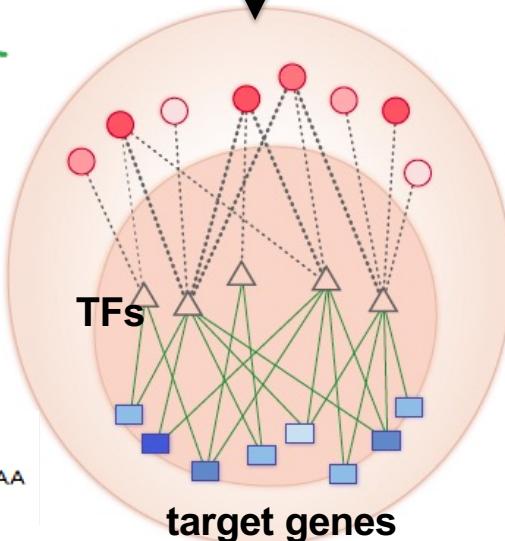
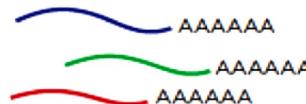
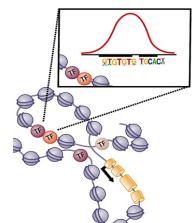
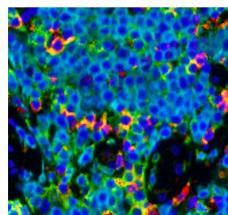
Zhang L, Sagan A, Qin B, Hu B, Osmanbeyoglu HU, STAN, a computational framework for inferring spatially informed transcription factor activity across cellular contexts, bioRxiv 2024.06.26.600782; doi: <https://doi.org/10.1101/2024.06.26.600782>

## Hands-on session

- Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public spatial transcriptomics and scRNA-seq (optional)
- STAN, Decoupler (optional), SCENIC (optional),
- [https://github.com/osmanbeyoglulab/ECCB-2024-Tutorial/tree/main/hands-on\\_tutorial](https://github.com/osmanbeyoglulab/ECCB-2024-Tutorial/tree/main/hands-on_tutorial)

# Computational methods for TF activity inference based on single cell omics

“environmental cues”



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

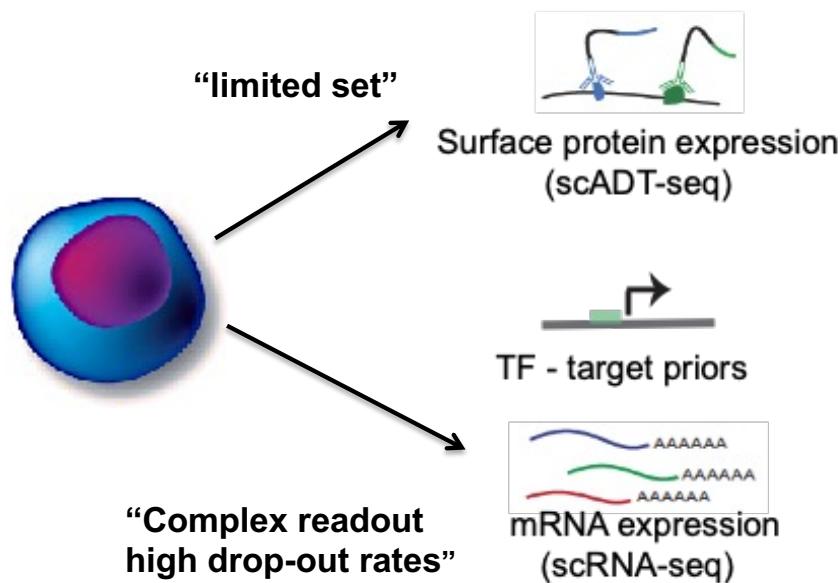
TF activity inference from single-cell gene expression data

e.g. SCENIC, BITFAM, metaVIPER, INFERELATOR 3.0

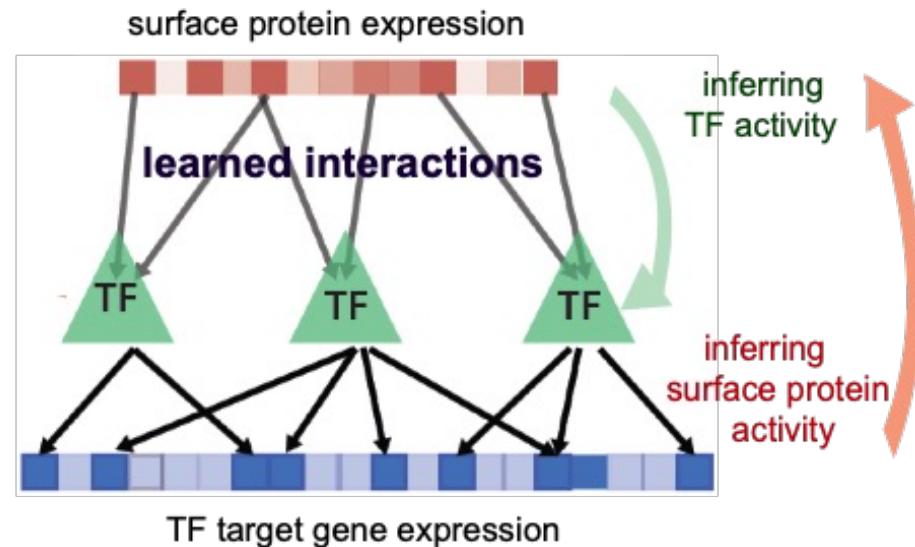
# TF activity inference from single-cell proteomics and transcriptomics data

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

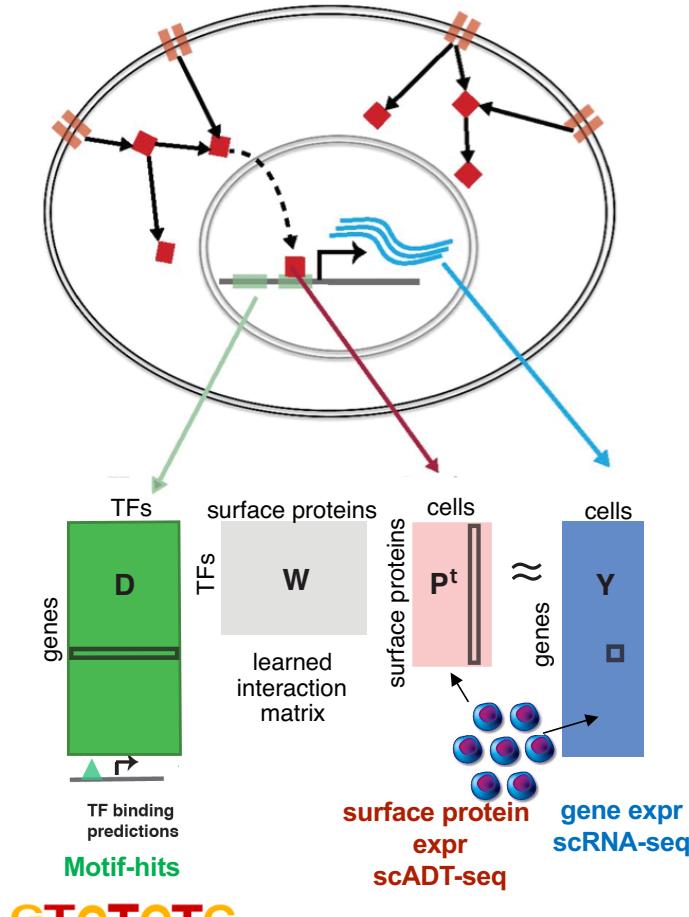
### Single cell multi-omics profiling CITE-seq



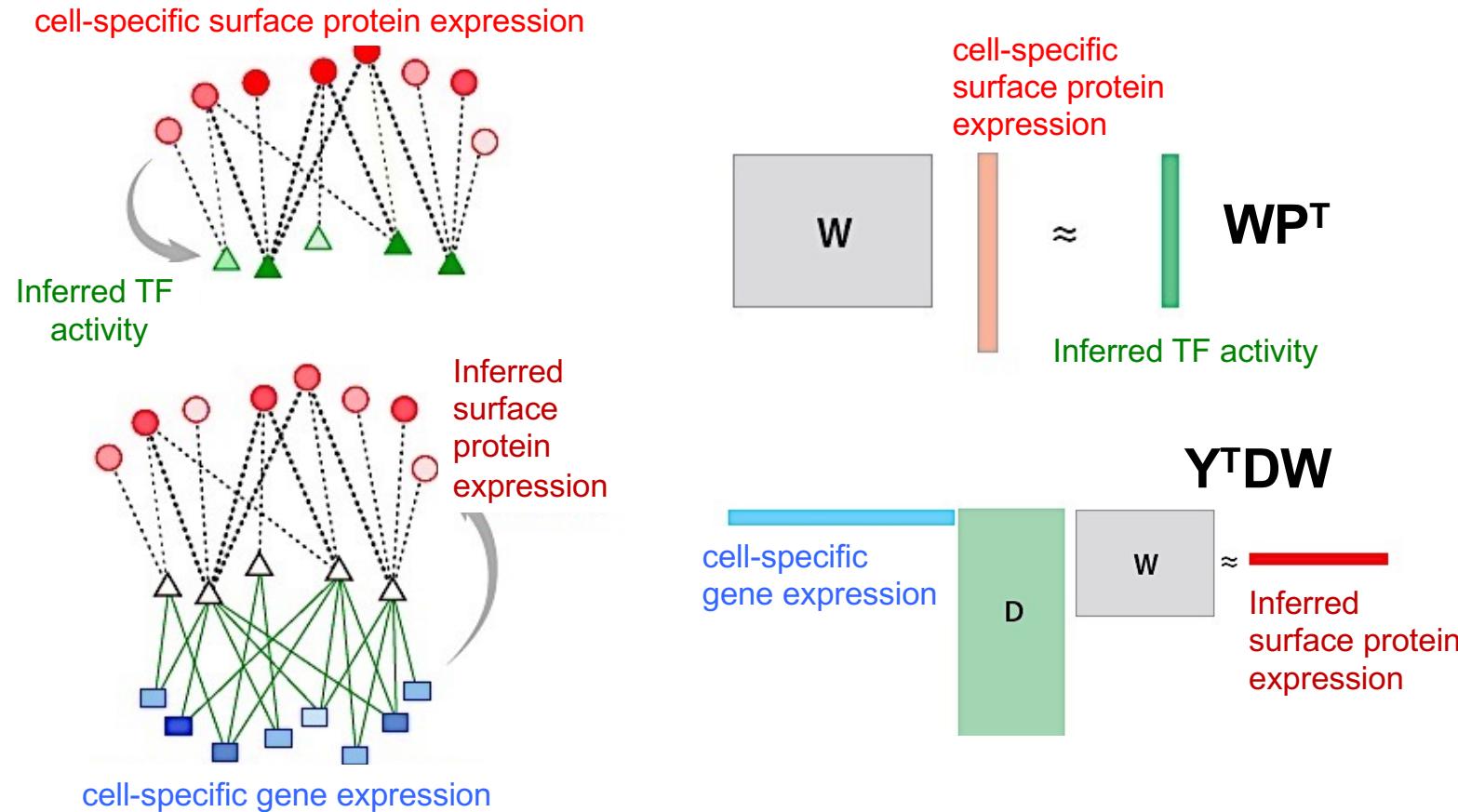
### Linking surface proteins to transcriptional regulators



## SPaRTAN - Linking surface proteins to TFs

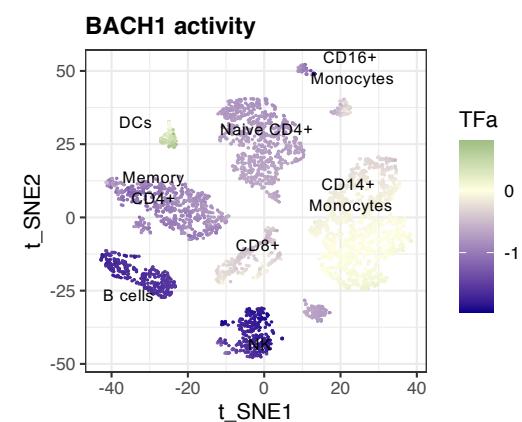
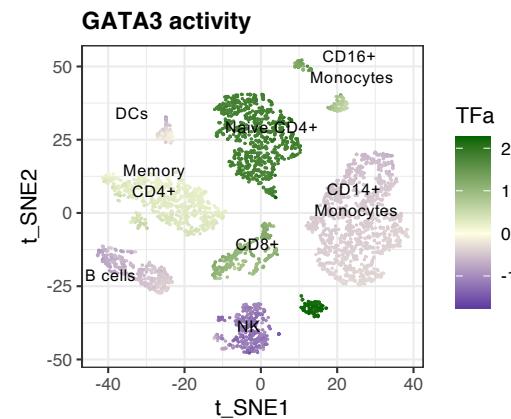
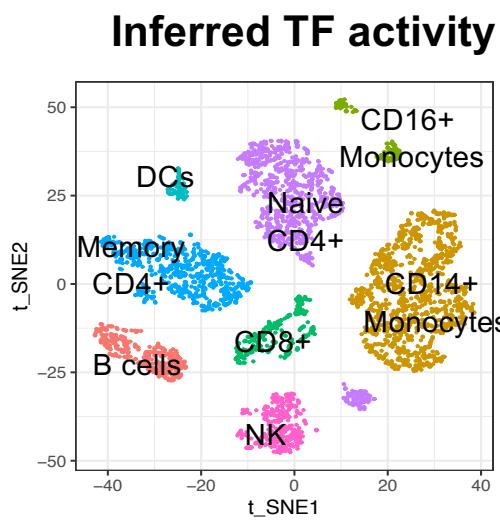
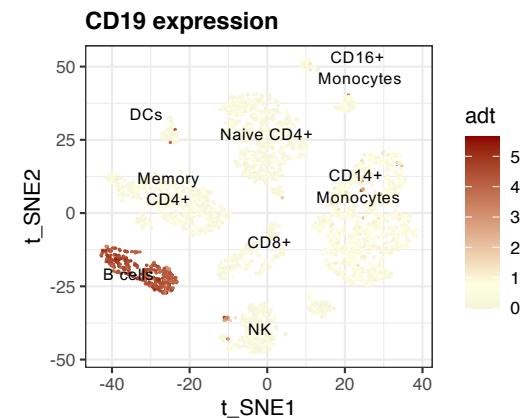
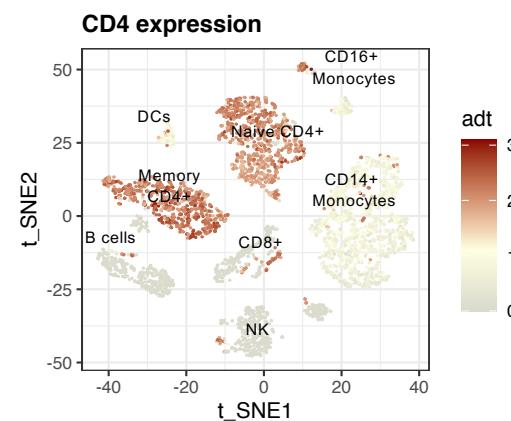
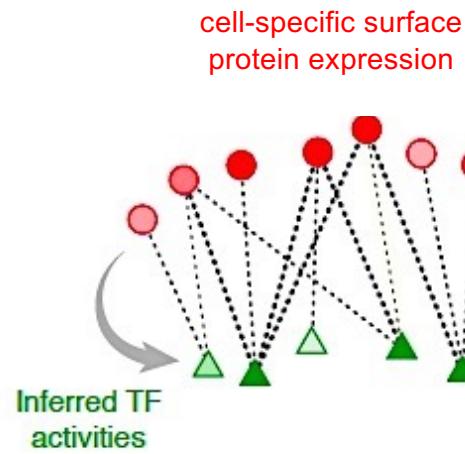
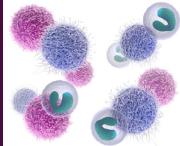


## Inferring cell-specific TF activity and surface protein expression

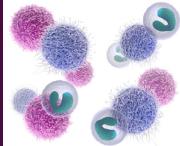


Reduced dimension in terms of TF activities and surface protein expression make data more tractable and reduce noise in single cell data while preserving the often intrinsically low dimensional signal of interest

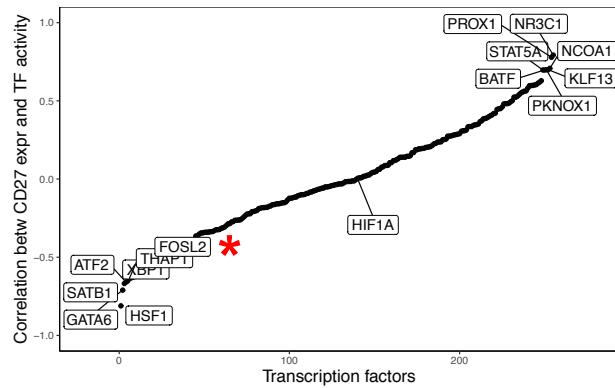
# SPaRTAN identifies both known and novel cell type-specific TFs



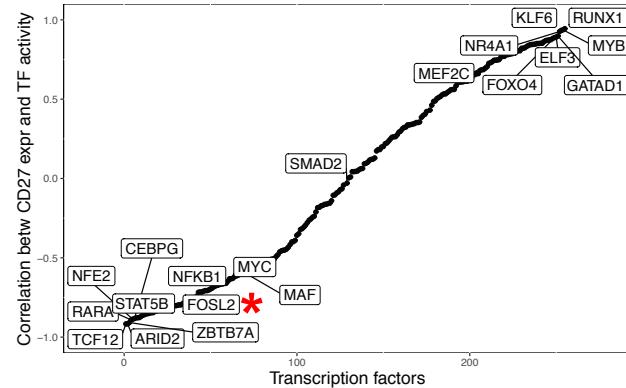
# Common and cell-type specific surface protein-TF correlation (PBMC)



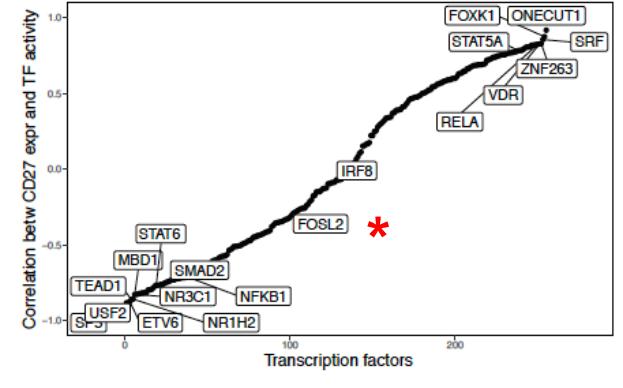
**CD4<sup>+</sup> memory T cells**



**CD8<sup>+</sup> T cells**



**B cells**

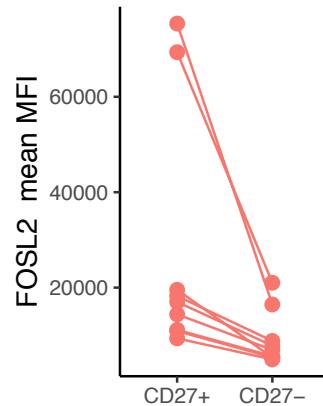


Flow validation:

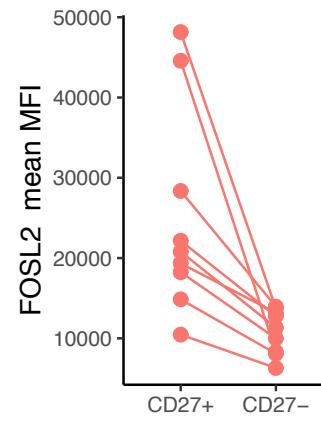
Paired analysis of FOSL2 expression  
in CD27<sup>+</sup> vs CD27<sup>-</sup>

CD4<sup>+</sup> memory T, CD8<sup>+</sup> T and B cells

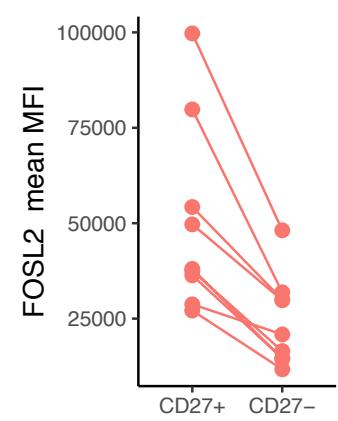
P=0.002



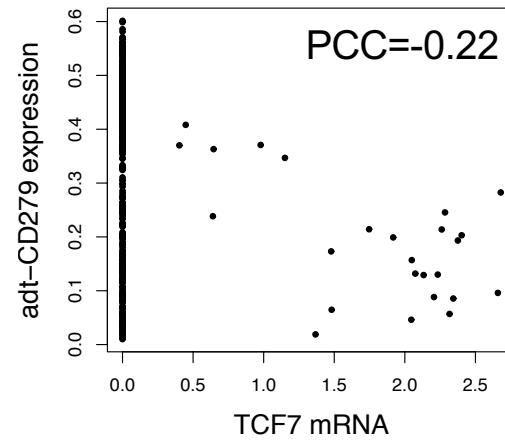
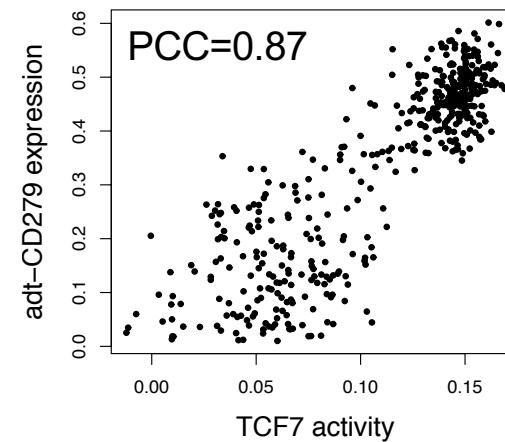
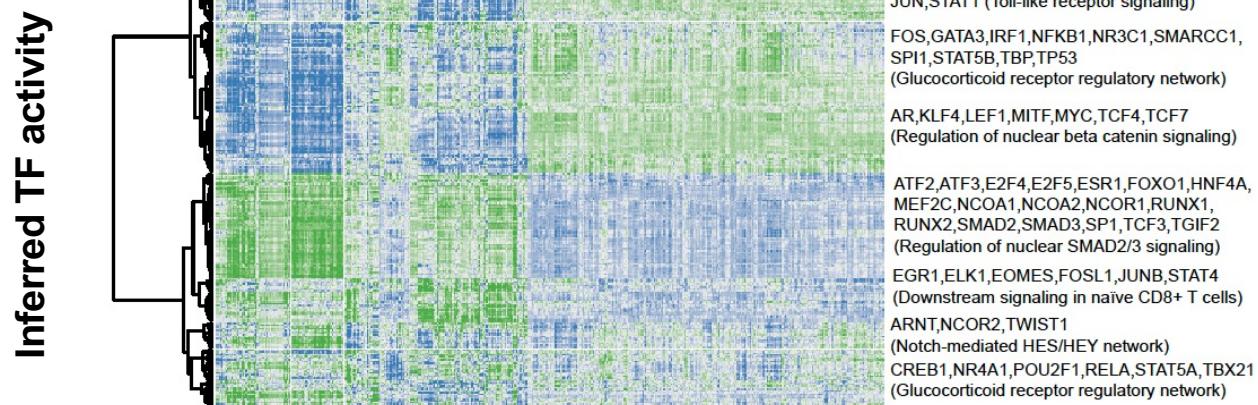
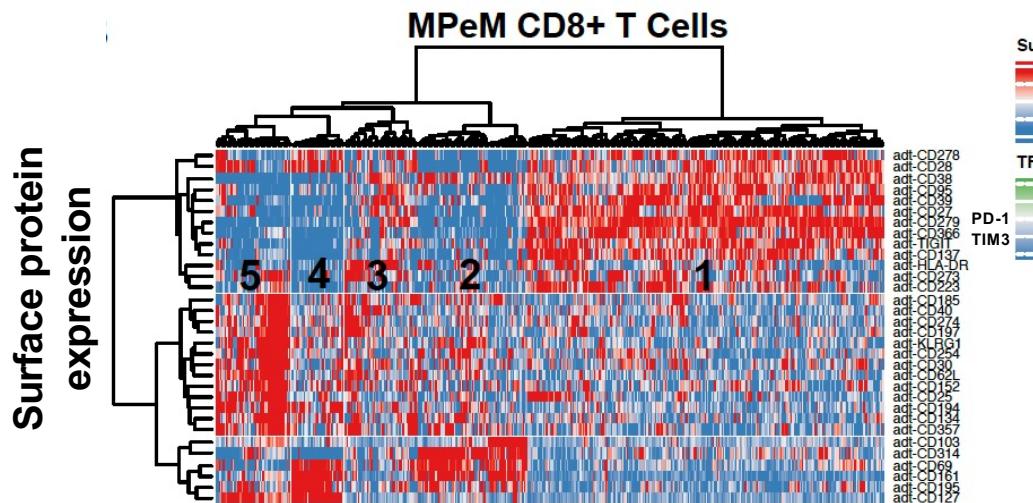
P=0.002



P=0.002



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

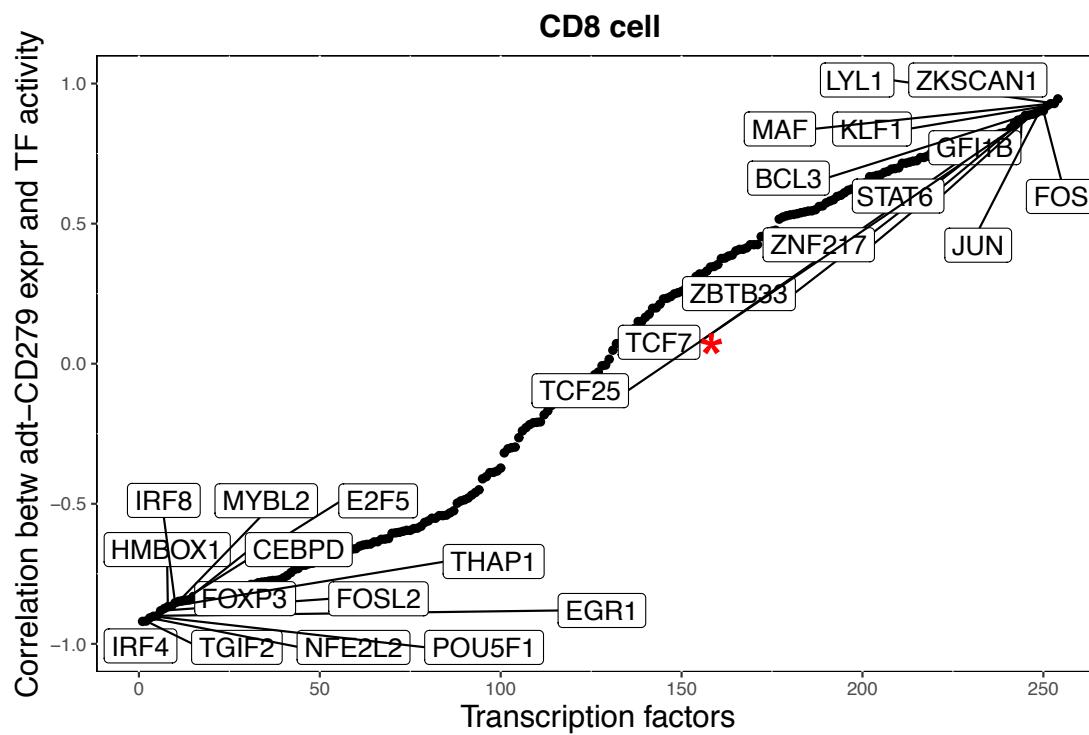


TCF7 is required for memory cells and efficacy of immunotherapies

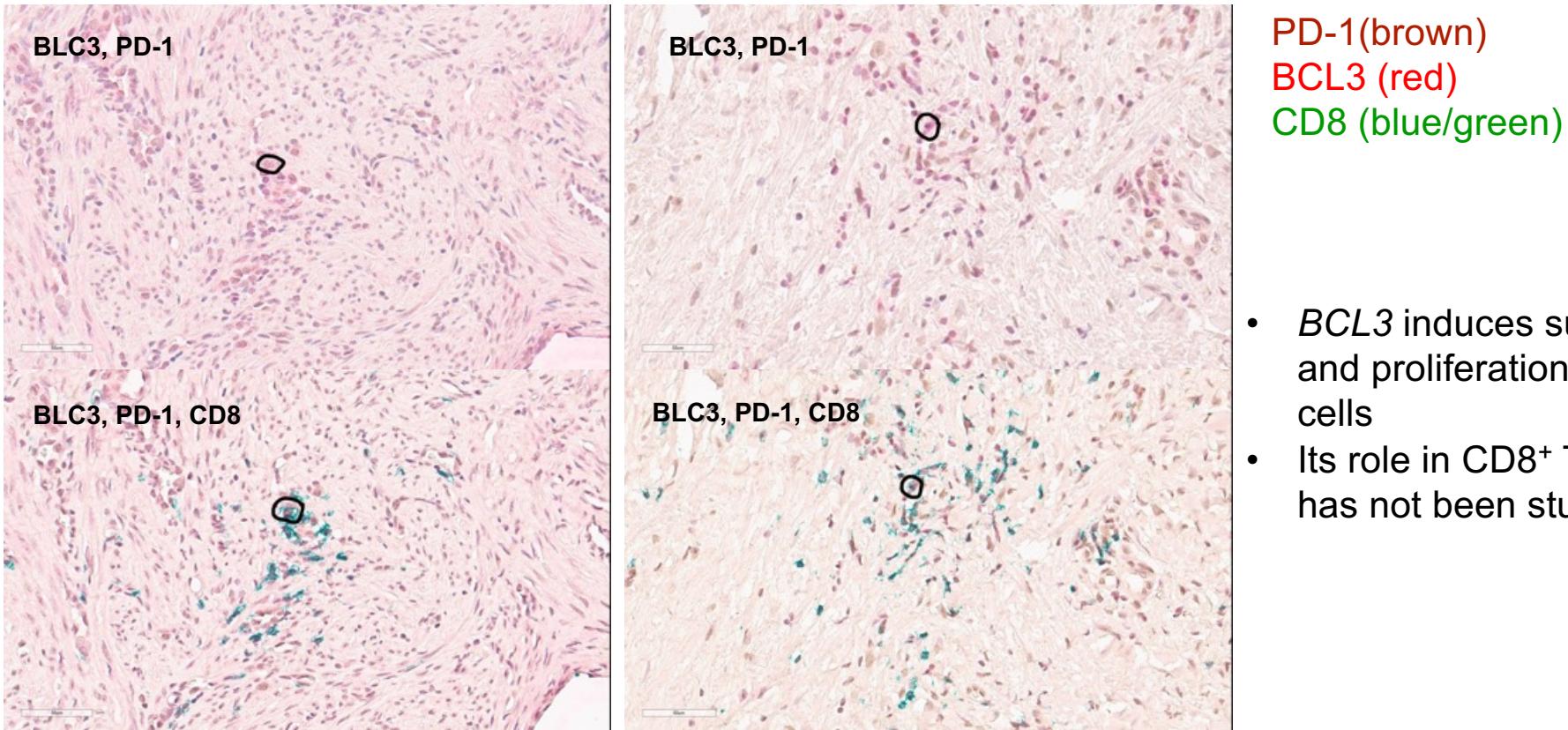
Siddiqui et al.  
Immunity 2019  
Kurtulus et al.,  
Immunity 2019

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<sup>+</sup> T cell

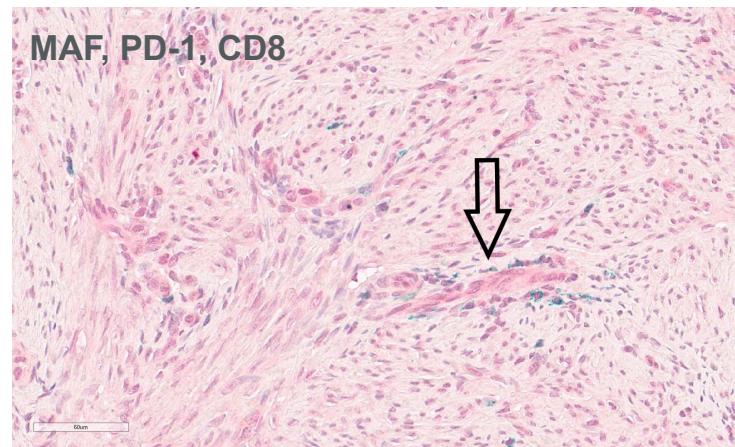
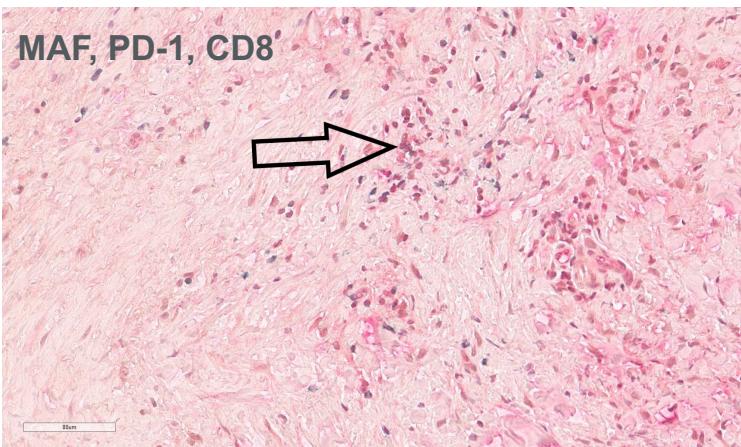
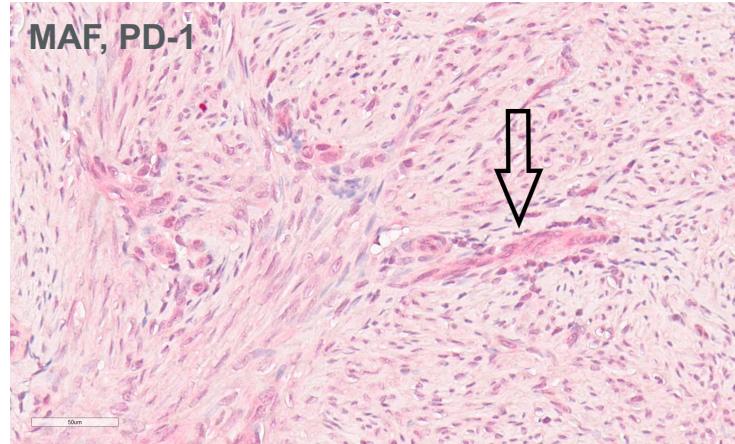
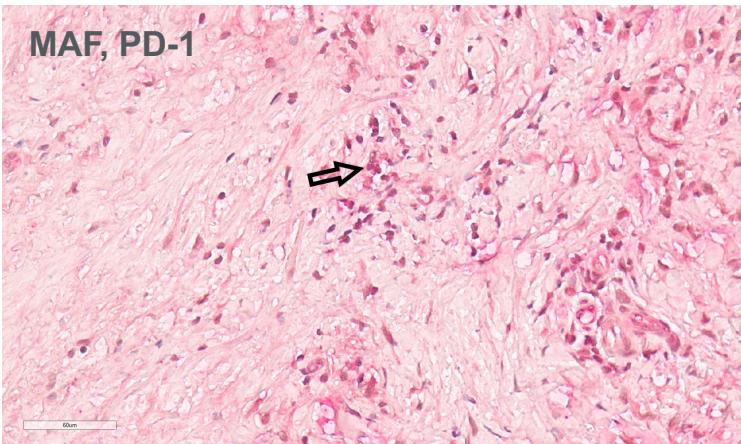


## National Mesothelioma Virtual Bnk (NMVB) Validation Cohort Co-expression of CD8, PD1, BCL3 by IHC staining



- *BCL3* induces survival and proliferation in cancer cells
- Its role in CD8<sup>+</sup> 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<sup>+</sup> 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 commonalities as well as differences of cell-specific regulatory programs across healthy individuals and those manifesting a disease.

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>

<https://github.com/osmanbeyoglulab/SPaRTAN>

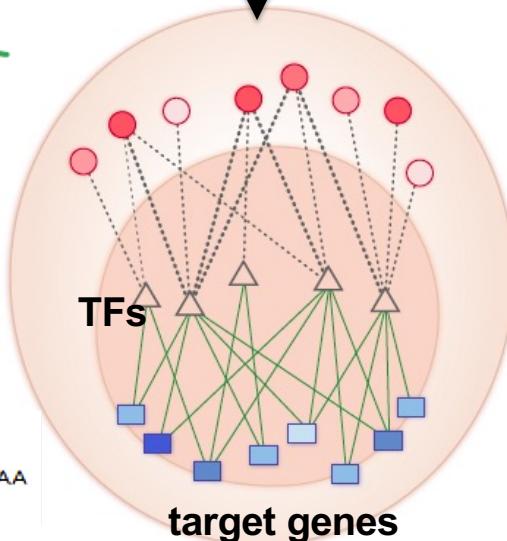
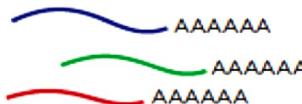
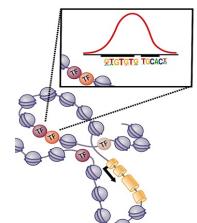
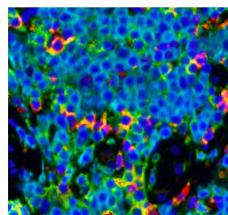
[https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/tree/main/hands-on\\_tutorial/session-2](https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/tree/main/hands-on_tutorial/session-2)

## 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

“environmental cues”



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

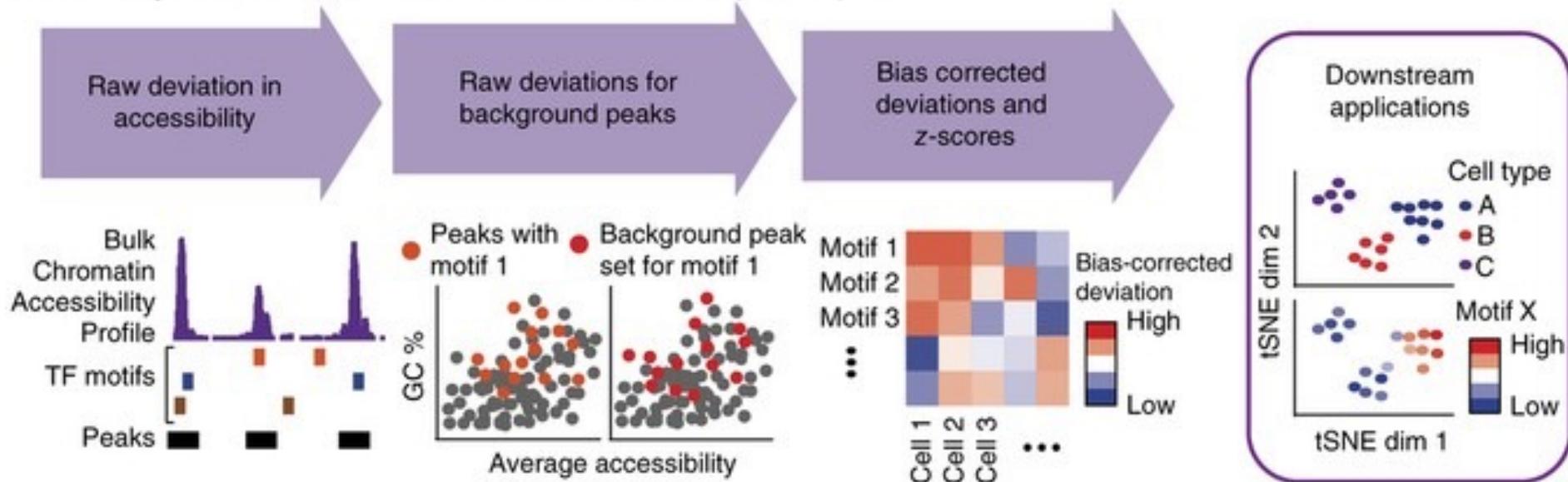
TF activity inference from single-cell gene expression data

e.g. SCENIC, BITFAM, metaVIPER, INFERELATOR 3.0

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

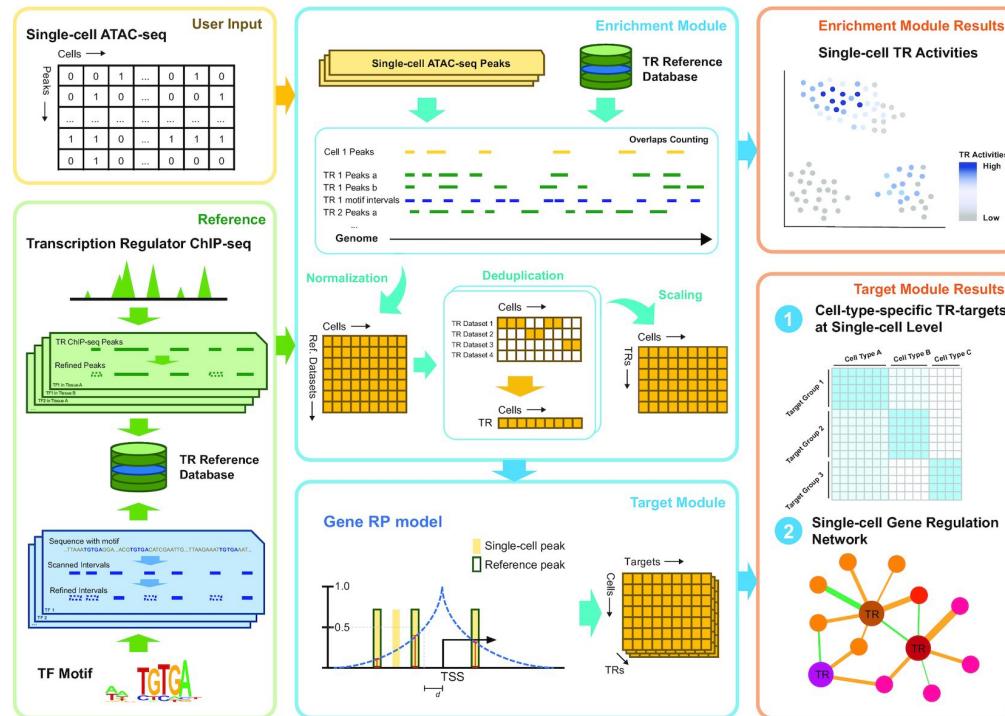
- ChromVar (Schep et al, Nature Methods 14, 975 (2017)) to find (a few hundred) TF motifs enriched in the peak count vector of each cell and visualize the results with tSNE/UMAP  
<https://bioconductor.org/packages/release/bioc/vignettes/chromVAR/inst/doc/Introduction.htm>

a For every motif, k-mer, or annotation and each cell or sample, compute:



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

- SCRIPI (Single-Cell gene Regulation network Inference using ChIP-seq and motif) (Dong et al., NAR, 2024)
- <https://github.com/wanglabtongji/SCRIPI>

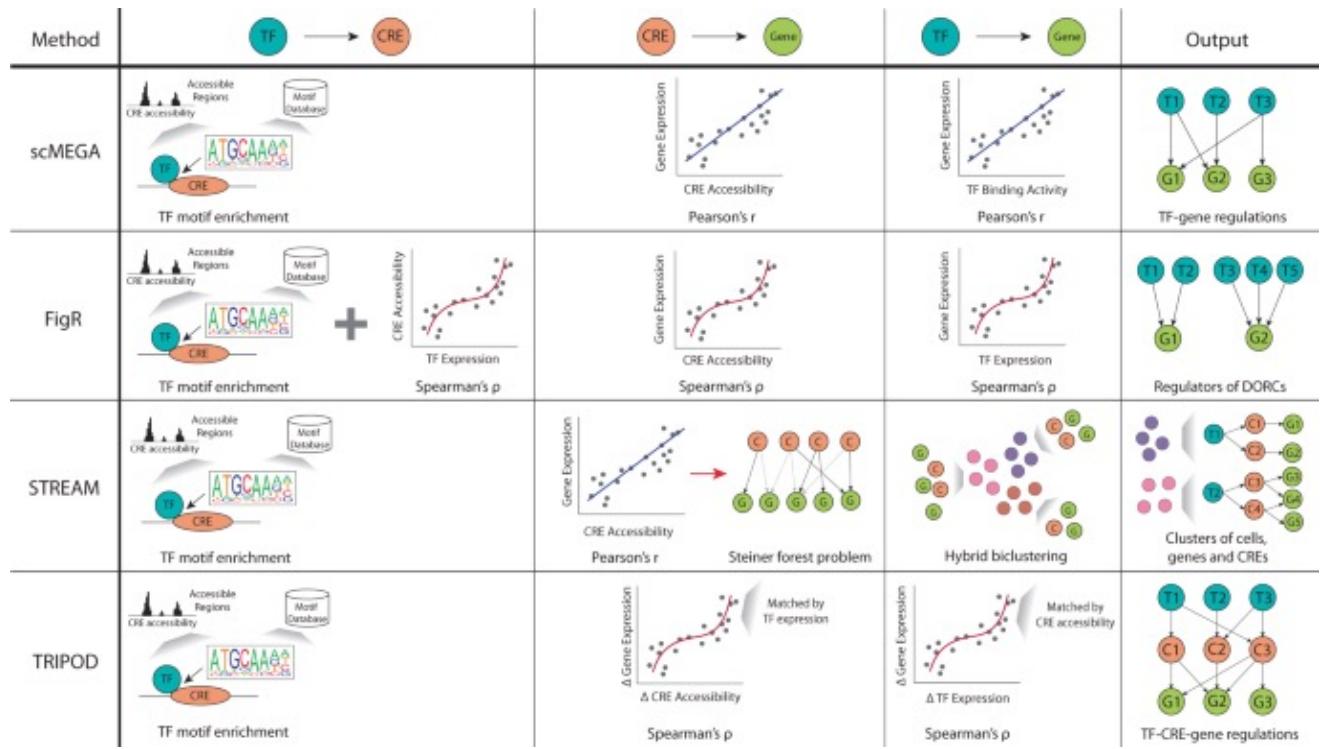


# TF activity inference from single-cell epigenomic and transcriptomics data (1)

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

# Correlation-based methods

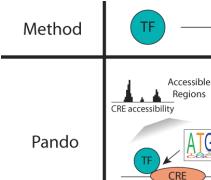
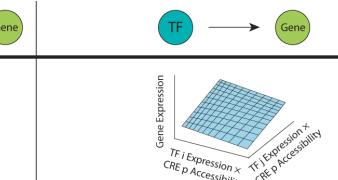
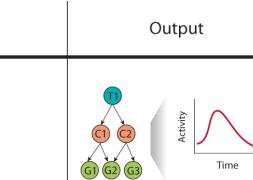
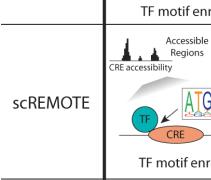
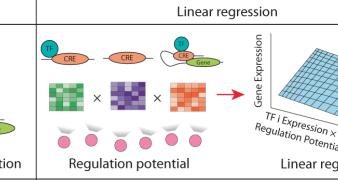
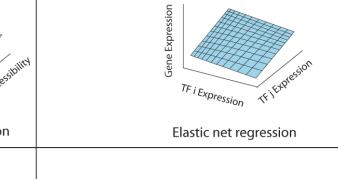
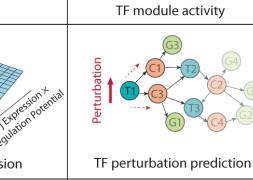
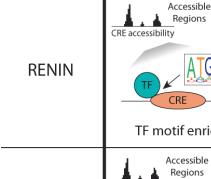
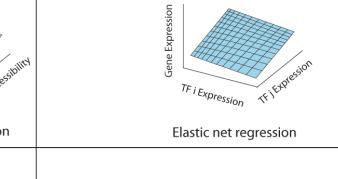
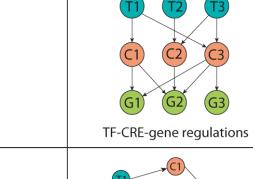
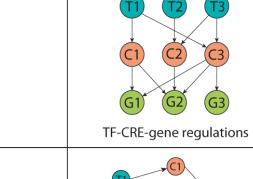
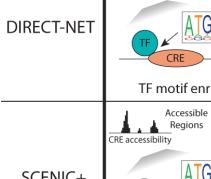
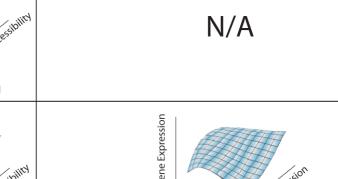
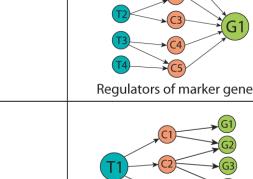
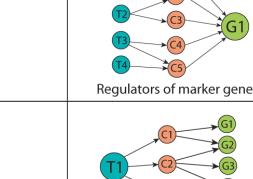
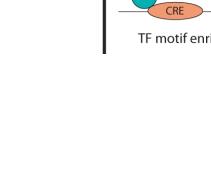
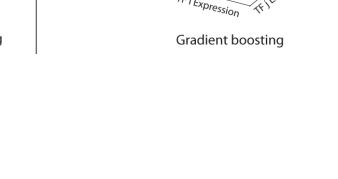
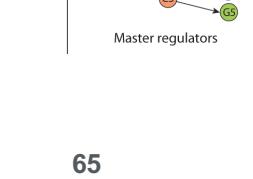
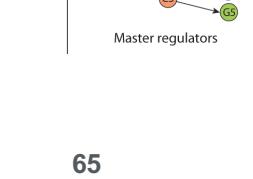
- scMEGA (Single-cell Multiomic Enhancer-based Gene regulatory network inference)  
<https://costalab.github.io/scMEGA/>
- FigR (functional inference of gene regulation)  
<https://buenrostrolab.github.io/FigR/>
- STREAM (Single-cell enhancer regulatory network inference from gene Expression And chromatin accessibility)  
<https://github.com/OSU-BMBL/STREAM>
- TRIPOD (transcription regulation interrogation through the nonparametric partial association analysis of single-cell multiomic sequencing data)  
<https://github.com/yuchaojiang/TRIPOD>



Kim, D., et al. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. *npj Syst Biol Appl* 9, 51 (2023).

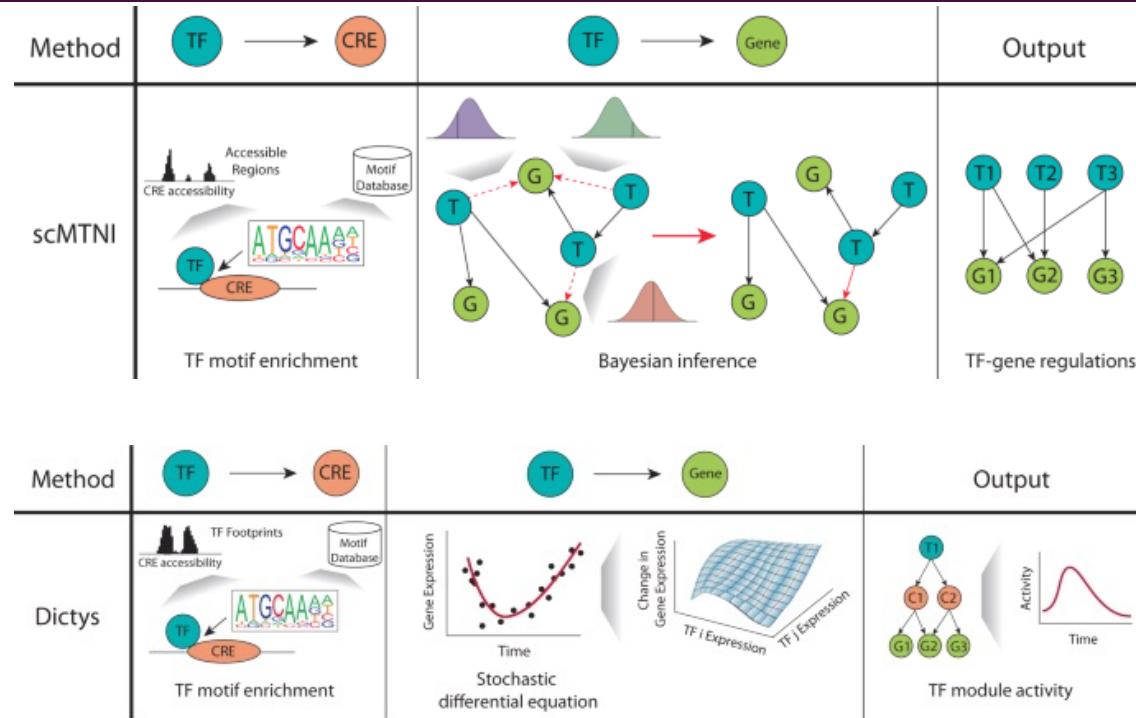
# Regression-based methods

- Pando, modeling the relationship between TF-binding site pairs with the expression of target genes, .
   
<https://github.com/quadbio/Pando>
- scREMOTE (single cell REprogramming MOdel Through cis-regulatory Elements)
   
<https://github.com/SydneyBioX/scREMOTE>
- RENIN (Regulatory Network Inference)
   
<https://github.com/nledru/renin>
- DIRECT-NET
   
<https://github.com/zhanglhbioinfor/DIRECT-NET>
- SCENIC +
   
<https://scenicplus.readthedocs.io/en/latest/>

Method				Output
Pando		N/A		
scREMOTE				
RENIN				
DIRECT-NET				
SCENIC+				
				

## Probabilistic and dynamical systems-based models

- scMTNI (single-cell Multi-Task Network Inference): a multi-task learning framework to infer the GRN for each cell type on a lineage  
<https://github.com/Roy-lab/scMTNI>
- Dictys: dynamic GRN inference and analysis method that leverages multiomic single-cell assays of chromatin accessibility and gene expression, context-specific transcription factor footprinting, stochastic process network and efficient probabilistic modeling of single-cell RNA-seq read counts  
<https://github.com/pinellolab/dictys>



# Deep learning-based methods

- DeepMAPS (Deep learning-based Multi-omics Analysis Platform for Single-cell data), a heterogeneous graph transformer framework for cell-type-specific biological network inference from scMulti-omics data.

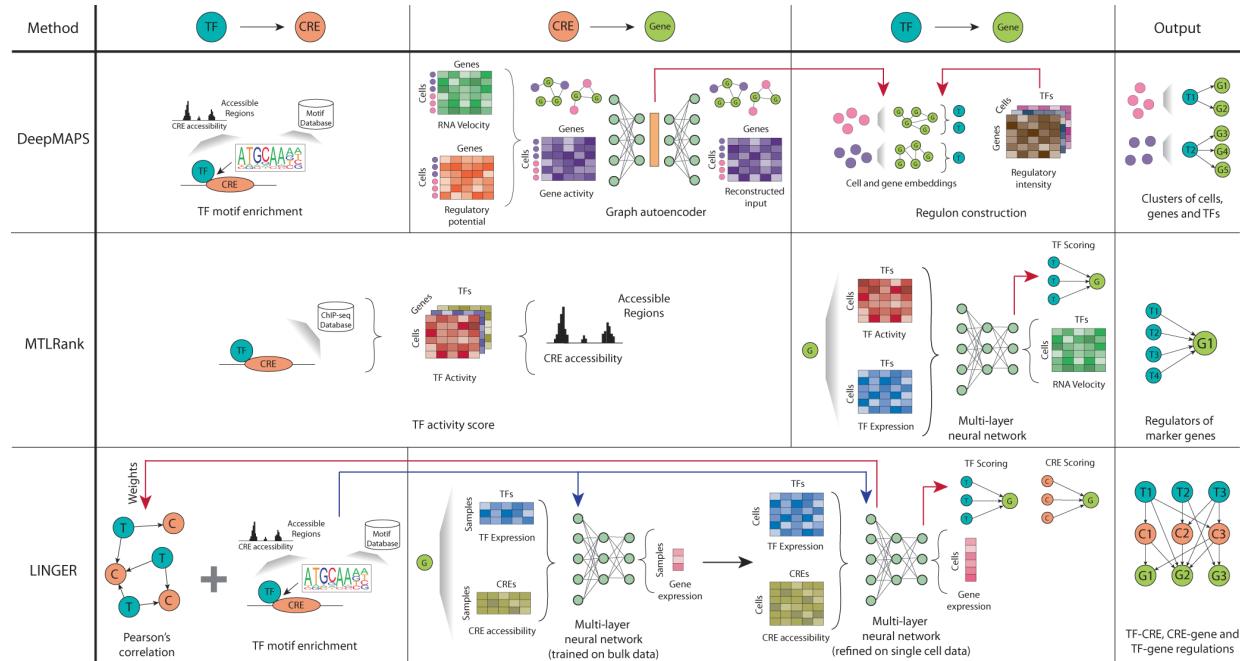
<https://github.com/OSU-BMBL/deepmaps>

- MTLRank, a multi-task learning framework that infers regulatory interactions from single-cell data by ranking TFs based on models predicting RNA velocity values for each gene

<https://github.com/alexQiSong/MTLRank>

- LINGER (Lifelong neural Network for GEne Regulation)

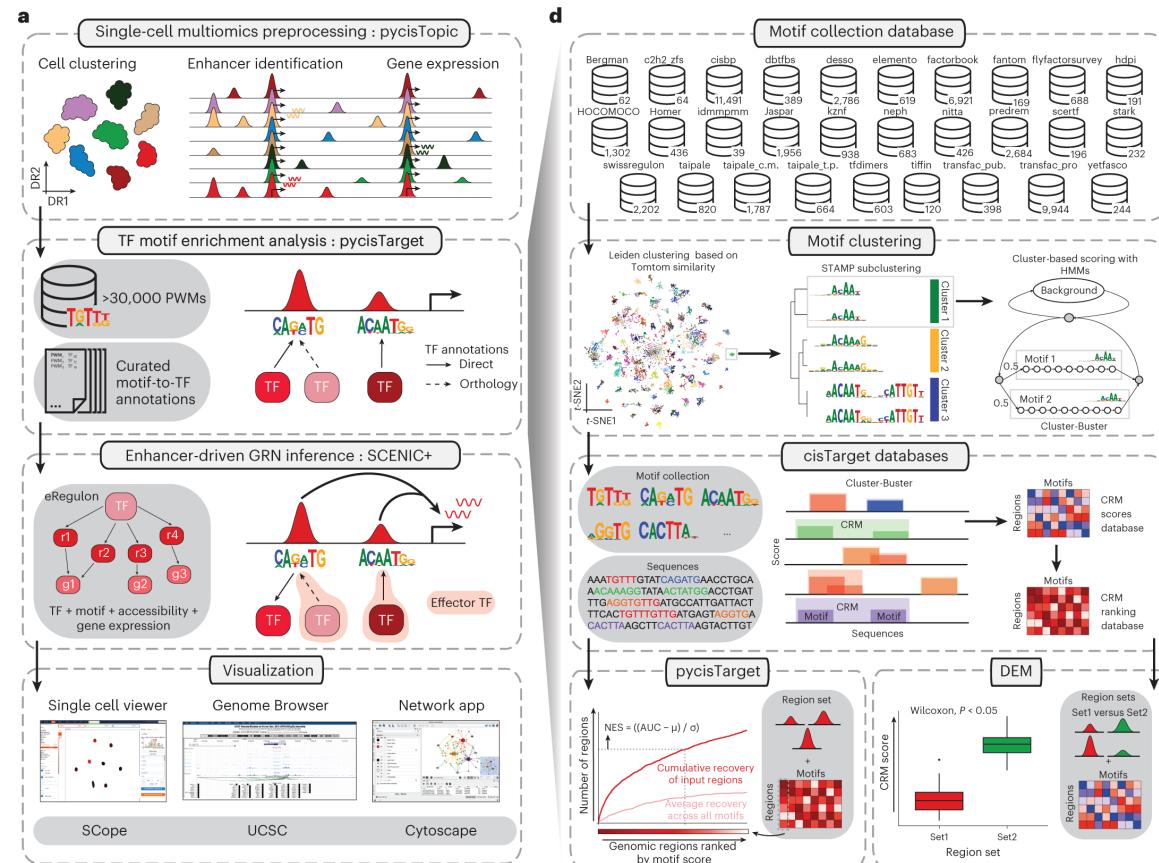
<https://github.com/Durenlab/LINGER>



# TF activity inference from single-cell epigenomic and transcriptomics data

SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks

<https://scenicplus.readthedocs.io/en/latest/>



[https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/tree/main/hands-on\\_tutorial/session-3](https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/tree/main/hands-on_tutorial/session-3)

## **Hands-on experience applying tools and interpreting results with TF activity inference methods, utilizing public spatial transcriptomics and scRNA-seq data**

- Tutorial materials:  
<https://github.com/osmanbeyoglulab/ECCB-2024-Tutorial/>



# Environmental setup



## Installation of Conda

Download the installer by choosing the proper installer for your machine. 2. Verify your installer hashes using [SHA-256 checksums](#). 3. Install the installer: - Windows: Double-click the .exe file. Follow the instructions on the screen. For a detailed reference, please read [this page](#). - macOS: double-click the .pkg file. Follow the instructions on the screen. For a detailed reference, please read [this page](#). - Linux: In your terminal window, run: `bash Anaconda-latest-Linux-x86_64.sh`. Follow the prompts on the screen. For a detailed reference, please read [this page](#).

## Managing Environment

With Conda, you can create, export, list, and update environments that have different versions of Python and/or packages installed in them. The JupyterLab, which can run in Conda environment, is a web application for computational documents so that our code can produce rich, interactive output.

Below we will demonstrate how to create a Conda environment and install JupyterLab and packages for each tutorial session on macOS/Linux. You need to create a separate Conda environment for each session.

Use the terminal for the following steps. For a detailed reference, please read [this page](#).

1. Create an environment with python 3.

```
conda create --name stan python=3.12 -y
```

2. Activate the environment you just created:

```
conda activate stan
```

3. Install JupyterLab:

```
pip install jupyterlab
```

4. Install required packages

```
git clone https://github.com/osmanbeyoglulab/ECCB-2024-Tutorial.git
cd ECCB-2024-Tutorial
pip install -r requirements.txt
```

After installing the required packages for the tutorial, launch JupyterLab and open the Jupyter notebook in each session subfolder to start the tutorial. To launch JupyterLab, enter the following command in your terminal:

```
Jupyter lab
```

[hands-on\\_tutorial/1.1-before\\_start.ipynb](#)  
[hands-on\\_tutorial/1.2-stan.ipynb](#)  
[hands-on\\_tutorial/2-pyscenic.ipynb](#)  
[hands-on\\_tutorial/3-decoupler.ipynb](#)

## Acknowledgement

### Osmanbeyoglu Lab

Linan Zhang, PhD (now at Ningbo University)  
April Sagan, PhD (now at GRAIL)  
Sanghoon Lee, PhD (now at NYU)  
Elena Kim, PhD  
Xiaojun Ma, MS  
Haoyu Wang  
Matthew Lu



NIGMS R35 GM146989

<https://www.osmanbeyoglulab.com>

**Looking for students and postdocs!**  
**Please reach out me at**  
**osmanbeyoglu@pitt.edu**

### **HILLMAN FELLOWS**

For Innovative Cancer Research Program

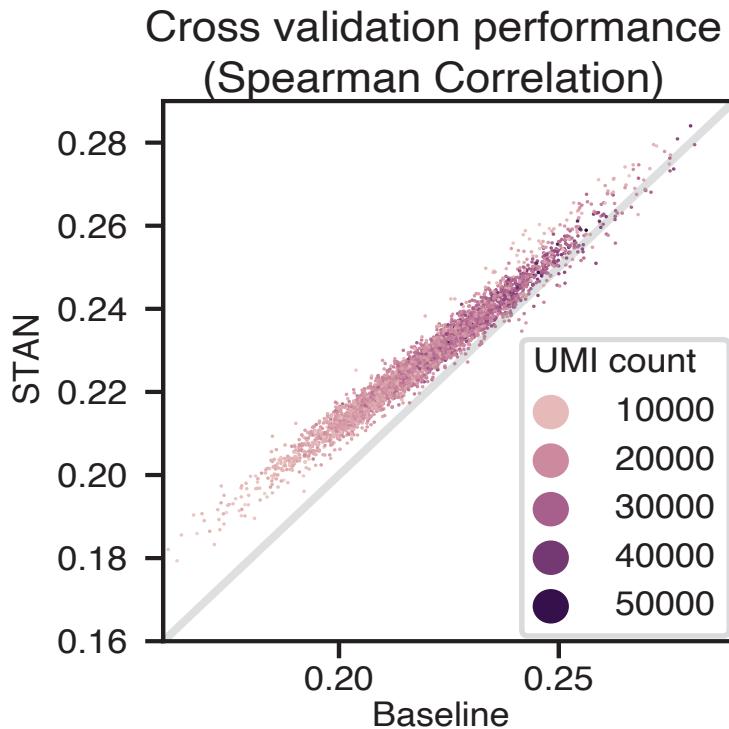


## References

- Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013 Mar;152(6):1237-1251. DOI: 10.1016/j.cell.2013.02.014.
- Kim, D., Tran, A., Kim, H.J. et al. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. *npj Syst Biol Appl* 9, 51 (2023). <https://doi.org/10.1038/s41540-023-00312-6>
- Hecker, Dennis et al. “Computational tools for inferring transcription factor activity.” *Proteomics* vol. 23,23-24 (2023): e2200462. doi:10.1002/pmic.202200462
- Heumos, L., Schaar, A.C., Lance, C. et al. Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24, 550–572 (2023).
- Linan Zhang, April Sagan, Bin Qin, Baoli Hu, Hatice Ulku Osmanbeyoglu. STAN, a computational framework for inferring spatially informed transcription factor activity across cellular contexts: bioRxiv 2024.06.26.600782; doi: <https://doi.org/10.1101/2024.06.26.600782>
- 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*, Volume 49, Issue 17, 27 September 2021, Pages 9633–9647, <https://doi.org/10.1093/nar/gkab745>
- Bravo González-Blas et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* 20, 1355–1367 (2023).

## STAN outperforms models without spatial information

10X Visium Lymph node



- Position weight matrix (PWM): need score cutoff

		Motif Matrix						
		Pos	A	C	G	T	Con	
Sites	Pos 12345678	ATGGCATG	1	0.9	0	0	0.1 A	Segment ATGCAGCT score =
	AGGGTGC G	2	0	0.1	0.2	0.7	T	$p(\text{generate ATGCAGCT from motif matrix})$
	ATCGCATG	3	0	0.1	0.7	0.2	G	$p(\text{generate ATGCAGCT from background})$
	TTGCCACG	4	0.1	0.1	0.8	0	G	$p_0A \times p_0T \times p_0G \times p_0C \times p_0A \times p_0G \times p_0C \times p_0T$
	ATGGTATT	5	0	0.7	0.1	0.2	C	
	ATTGCACG	6	0.8	0	0.2	0	A	
	AGGGCGTT	7	0	0.3	0	0.7	T	
	ATGACATG	8	0	0	0.8	0.2	G	
	ATGGCATG							
	ACTGGATG							

Shirley Liu lab

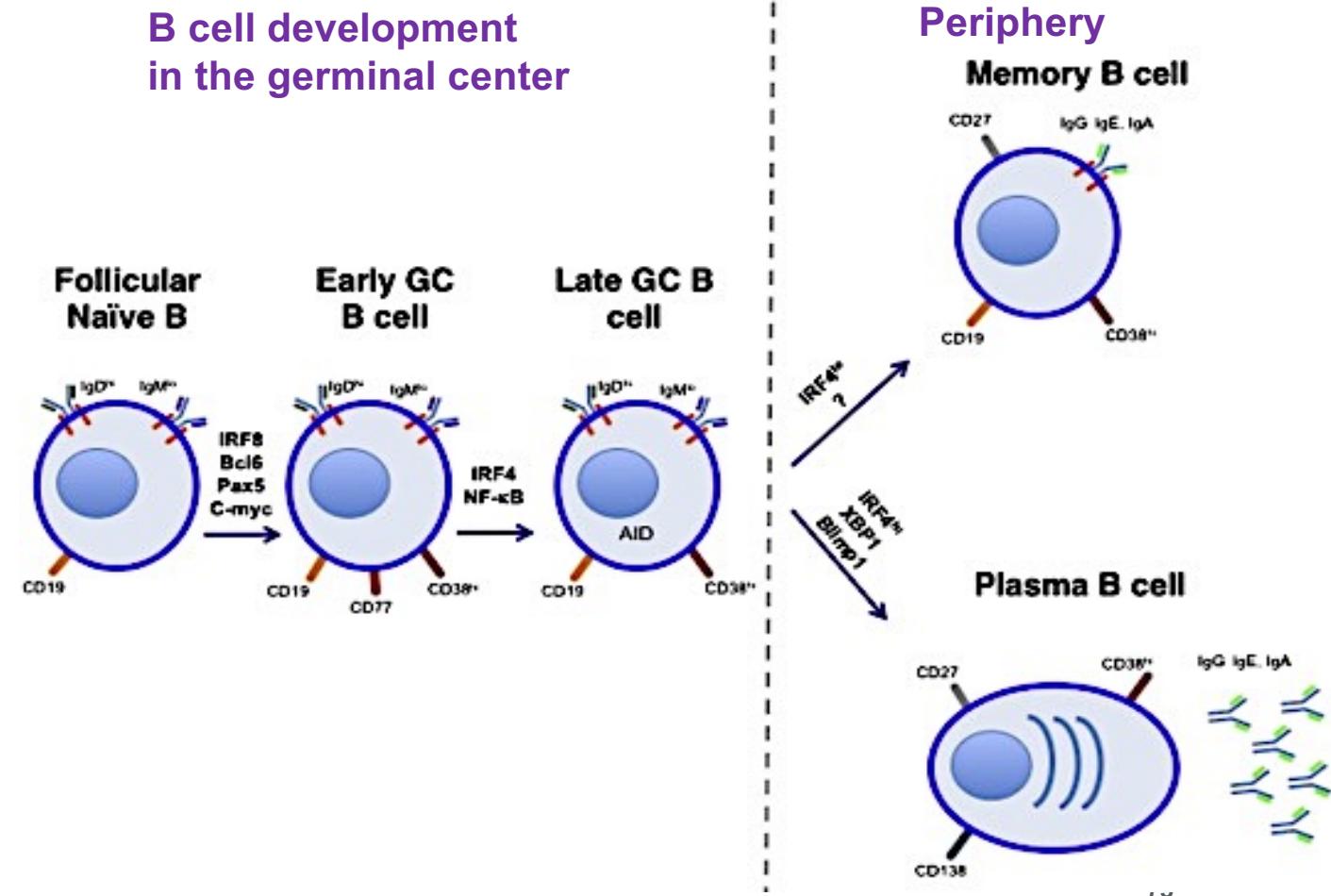


Total height: sequence conservation  
Letter height: relative frequency

- TF can have multiple motifs

## 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