



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

ISMB Tutorial July 9, 2024

UPMC | HILLMAN
CANCER CENTER



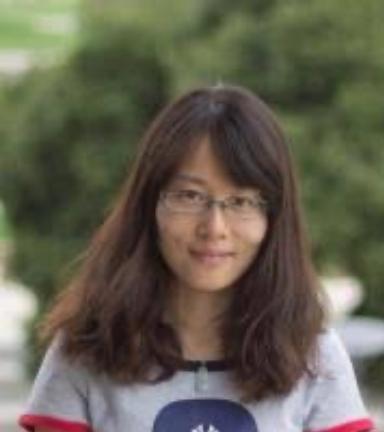
University of
Pittsburgh

Who are we?

- Hatice Ulku Osmanbeyoglu, Assistant Professor, University of Pittsburgh, USA
- Linan Zhang, Assistant Professor, Ningbo University, China
- Xiaojun Ma, Software engineer, University of Pittsburgh, USA
- Merve Sahin, Computational Biologist, Memorial Sloan Kettering Cancer Center, USA



Hatrice



Linan



Xiaojun



Merve

Practical information

- Tutorial materials:
<https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/>
- ISMB feedback link
<https://docs.google.com/forms/d/e/1FAIpQLSfrjbxGdE45OFPQgM58JDt59B--gonwjuKw1HkHrR4FLvhhw/viewform>

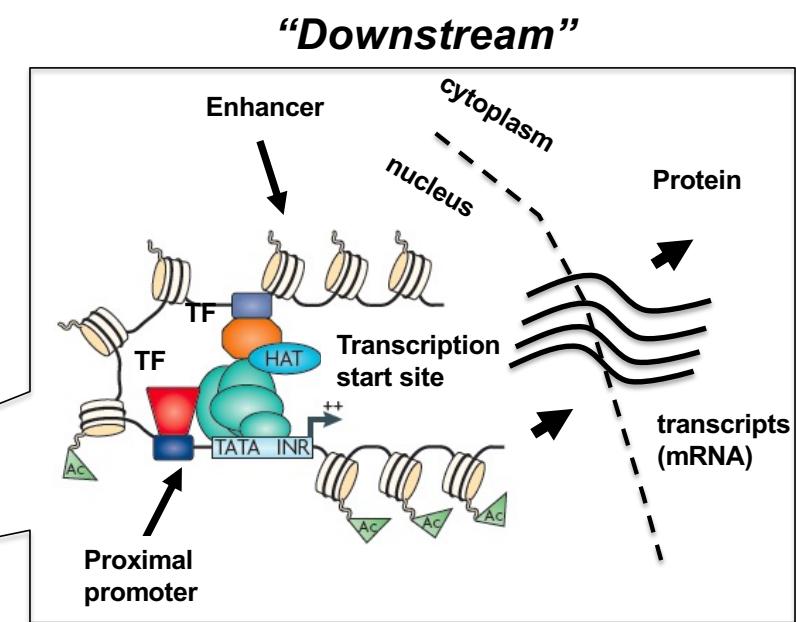
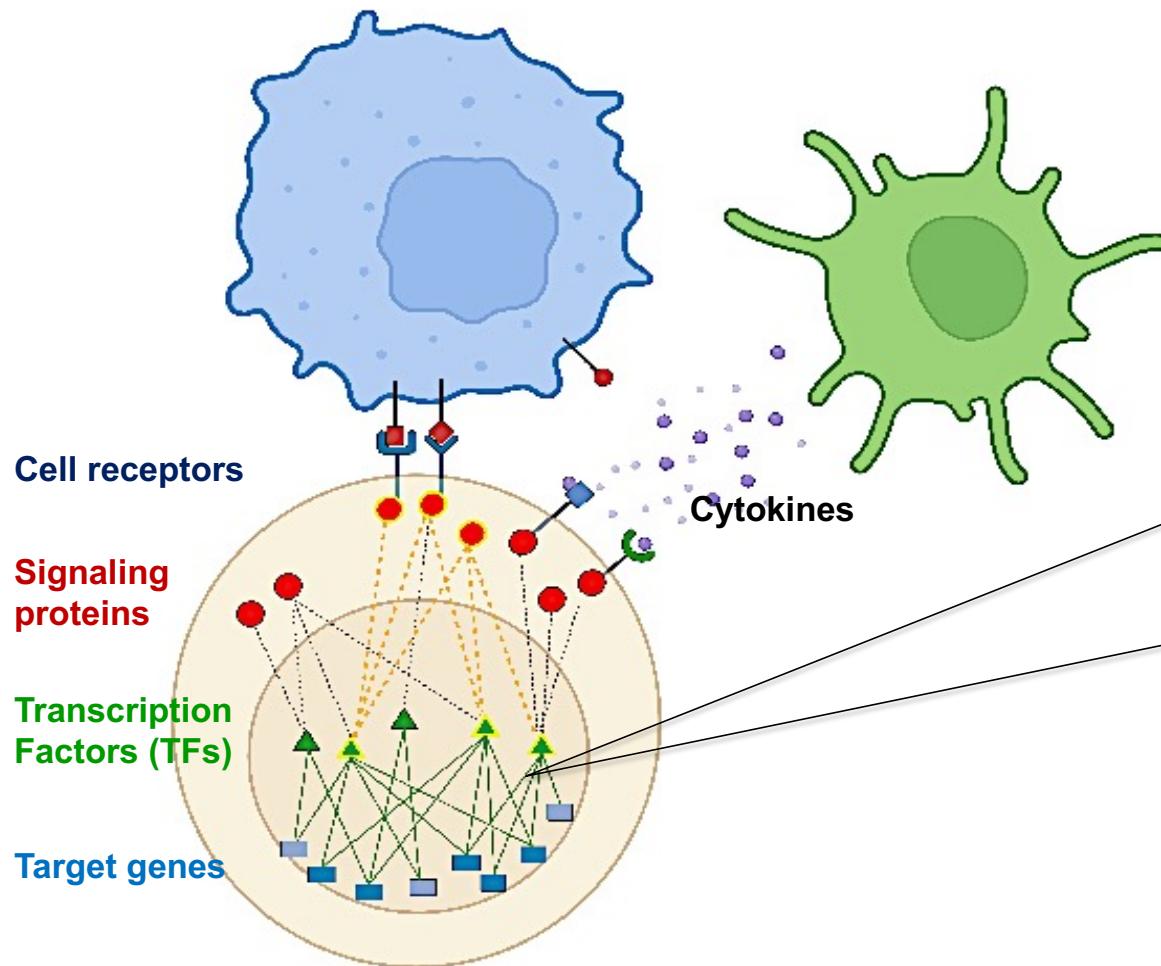
Tutorial overview

- This tutorial is focused on basic concepts and recent advances in transcription factor 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

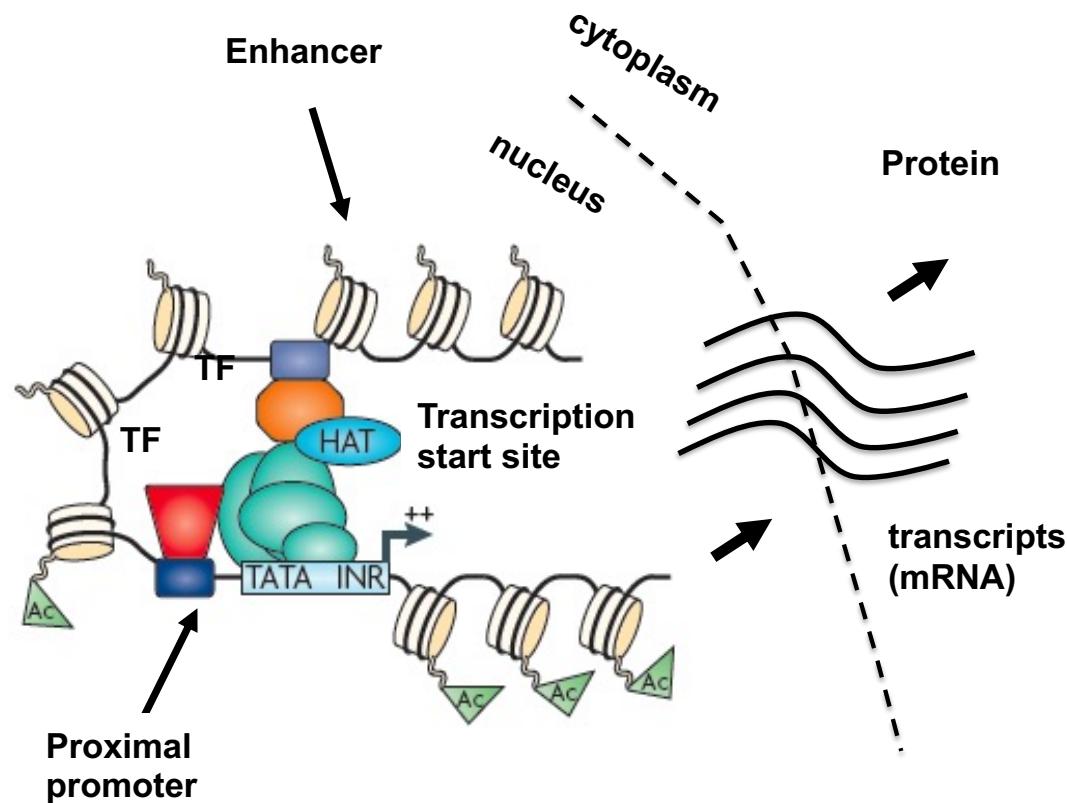
14:00	Welcome remarks and tutorial overview
14:05	Basic principles behind TF activity inference methods (Hatice Ulku Osmanbeyoglu) * Overview of the importance of context-specific TF regulation in biological systems * Significance of TF dynamics in health and disease * Single-cell multi-omics technologies for TF activity inference (scRNA-seq, spatial transcriptomics, CITE-seq, scATAC-seq, multiome)
14:45	Questions/Break
15:00	*Overview of computational TF inference methods based on single cell omics
15:45	Questions/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 (Linan Zhang)
16:30	Questions/Break
16:40	Hands-on experience in applying tools and interpreting results using TF activity inference methods using public CITE-seq (Xiaojun Ma)
17:00	Questions/Break
17:20	Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public multiome (Merve Sahin)
17:55	Final remarks

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



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

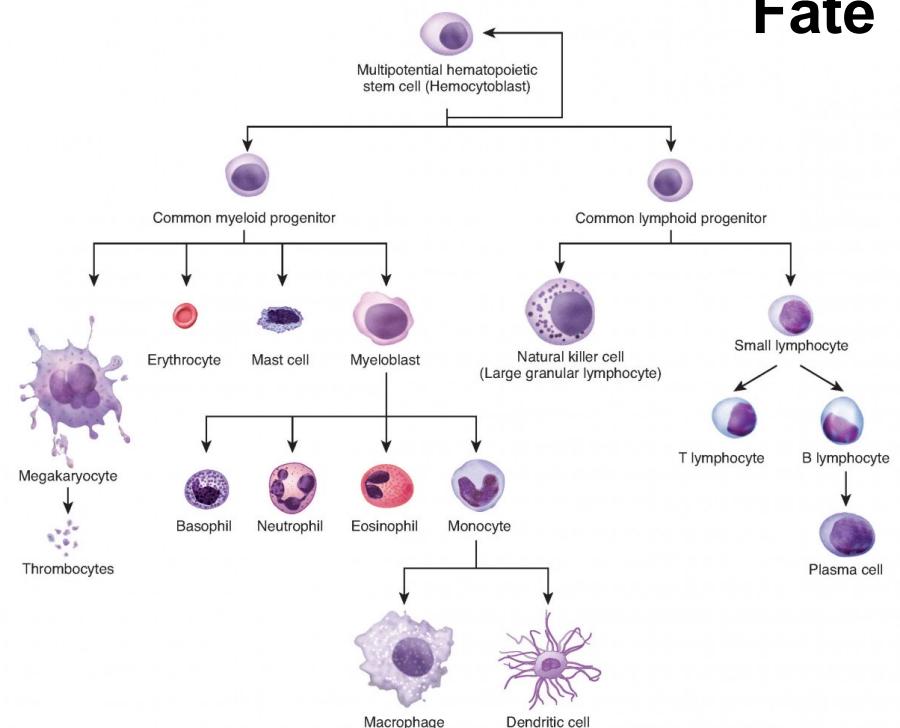
TFs regulate gene expression by binding specific DNA sequences in the enhancer or promoter regions of genes



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

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

- Cells in different states express different sets of genes



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

- Hematopoiesis involves the differentiation of multipotent cells into blood and immune cells

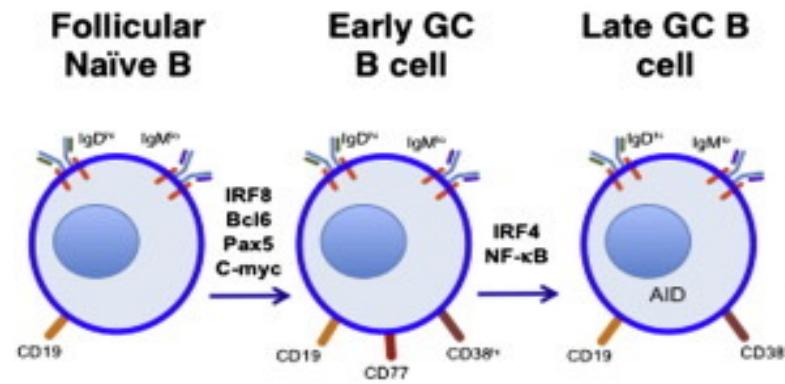
Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013 Mar;152(6):1237-1251. DOI: 10.1016/j.cell.2013.02.014.

Barnes et al, *Clinical Immunology*, 2014

Example: TFs regulating B cell functional states

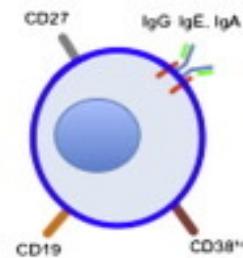
- B cells are the immune system's secret agents, producing antibodies and keeping a record of past enemies to protect body from future threats
- The interplay between signaling proteins, transcription factors and their downstream targets allows B cells to adapt and transition between various functional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

B cell development in the germinal center



Periphery

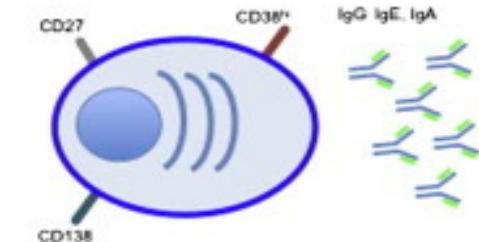
Memory B cell



IRF4^{hi}

IRF4^{hi}
XBP1
Blimp1

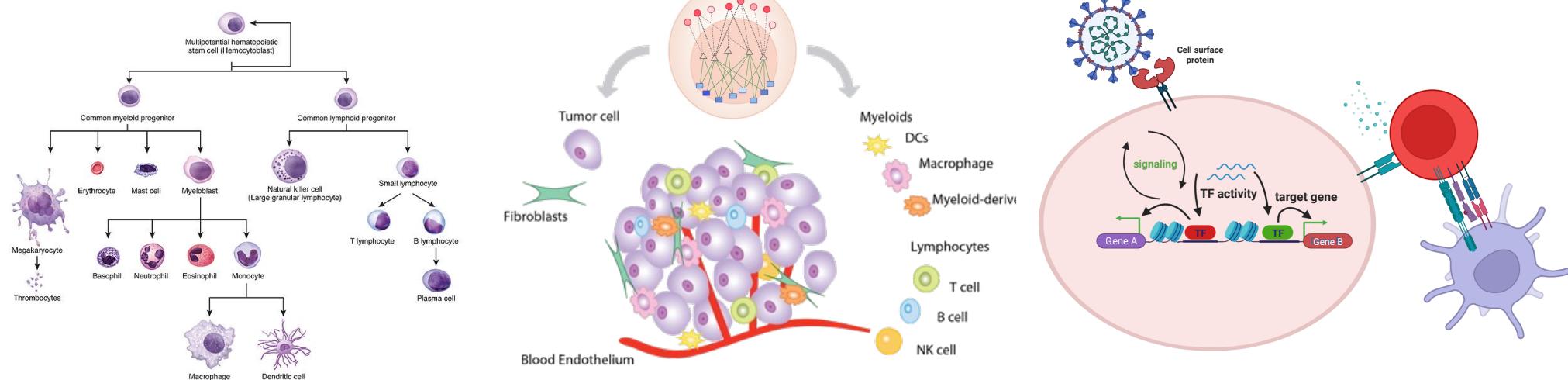
Plasma B cell



Barnes et al, *Clinical Immunology*, 2014

Significance of TF dynamics in health and disease

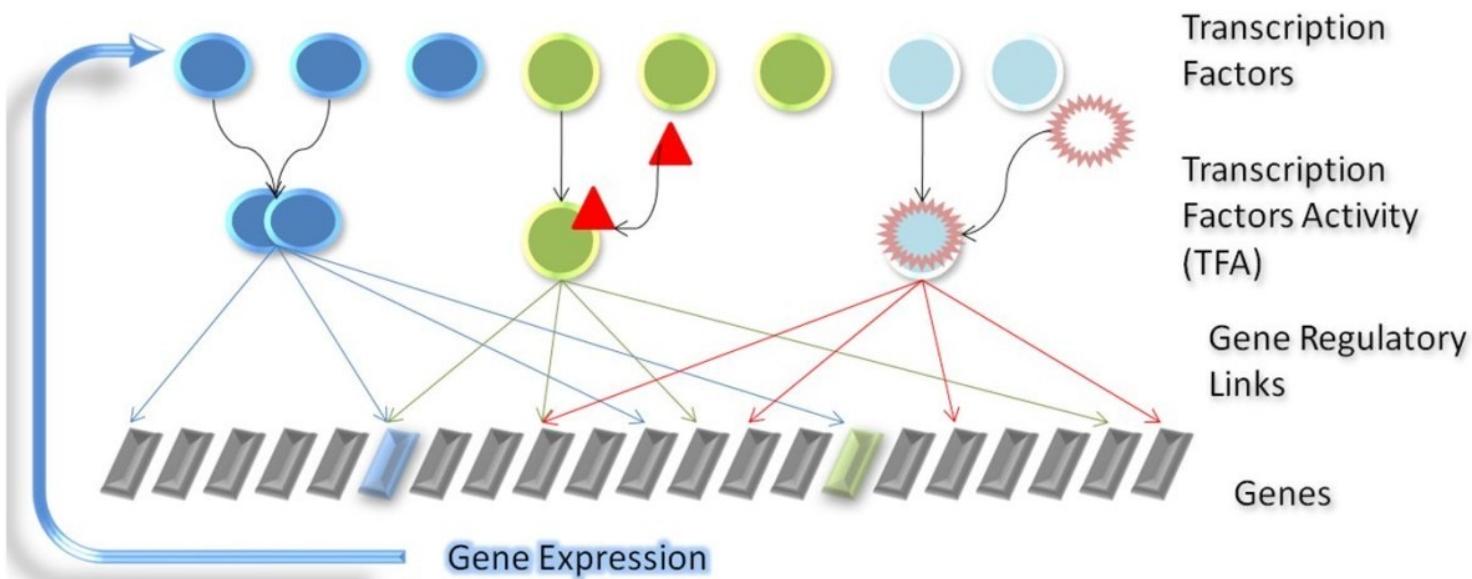
- What are TFs underlying cellular identities (e.g. naïve vs memory B cells)?
- What are different or common TFs given a cell type across tissues (e.g. B cells in spleen vs lung)?
- What are commonalities as well as differences of cell-type/state specific TFs across healthy individuals and those manifesting a disease (e.g., developmental changes and those associated with aging)?



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

TF activity

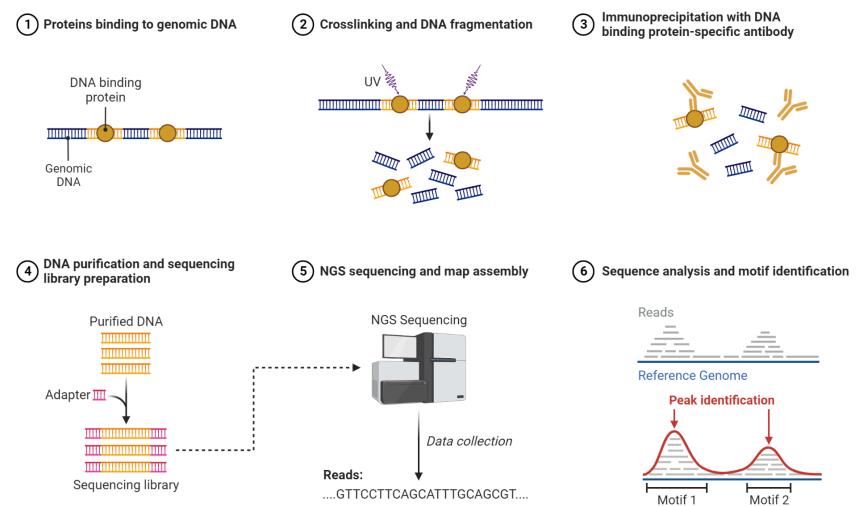
- Cellular processes depend on the expression levels (and activities) of proteins, notably TFs, which can be distinct from mRNA levels.
- TF activity (TFA) defines the regulatory impact exerted by a transcription factor on its target genes.
 - Examples include activation, repression, and effects like alternative splicing



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>

Experimental techniques for studying context specific transcription factors (1)

- **ChIP-Seq (Chromatin Immunoprecipitation followed by Sequencing):** To identify genomic regions where a transcription factor binds.
 - Cross-link DNA and proteins, immunoprecipitate the transcription factor-DNA complexes, sequence the DNA, and analyze the binding sites.
- **Reporter Assays:** To measure transcriptional activity of a specific transcription factor.
 - 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 transcription factor activity.



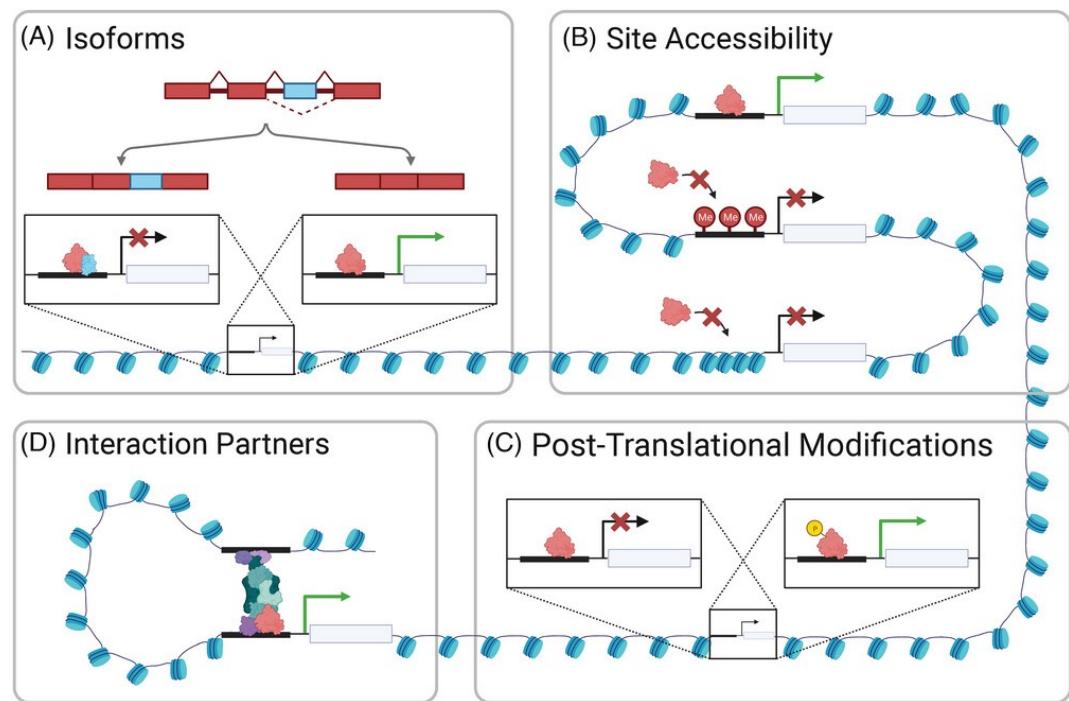
<https://www.biorender.com/template/chip-sequencing>

Experimental techniques for studying context specific transcription factors (2)

- **Gene Knockdown or Overexpression:** To manipulate transcription factor expression and observe downstream effects.
 - Use techniques like siRNA, shRNA, CRISPR-Cas9 for knockdown, or overexpress transcription factors using expression vectors.
- **Co-immunoprecipitation (Co-IP):** To identify protein-protein interactions involving transcription factors.
 - Immunoprecipitate a transcription factor of interest and identify associated proteins by mass spectrometry or western blotting.
- **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 transcription factor manipulation.

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)

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



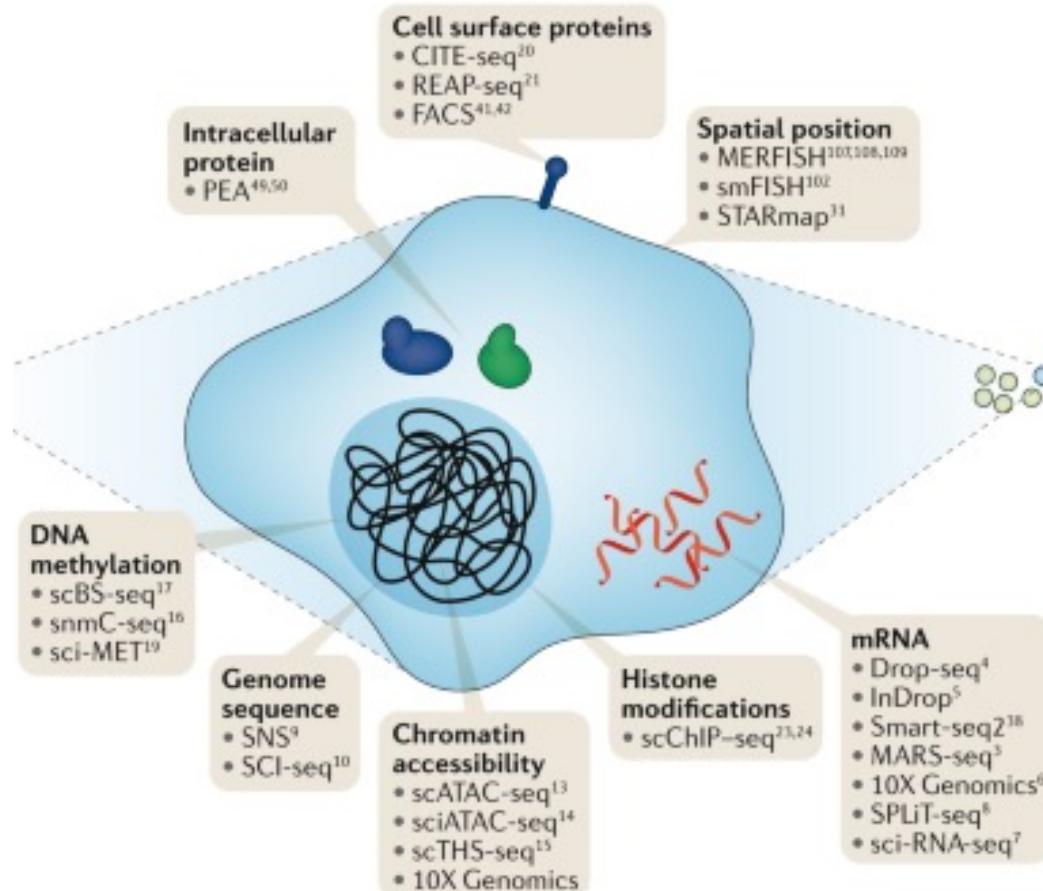
Single cell analysis retain single cell information but lose spatial organization



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



Single-cell/spatial omics



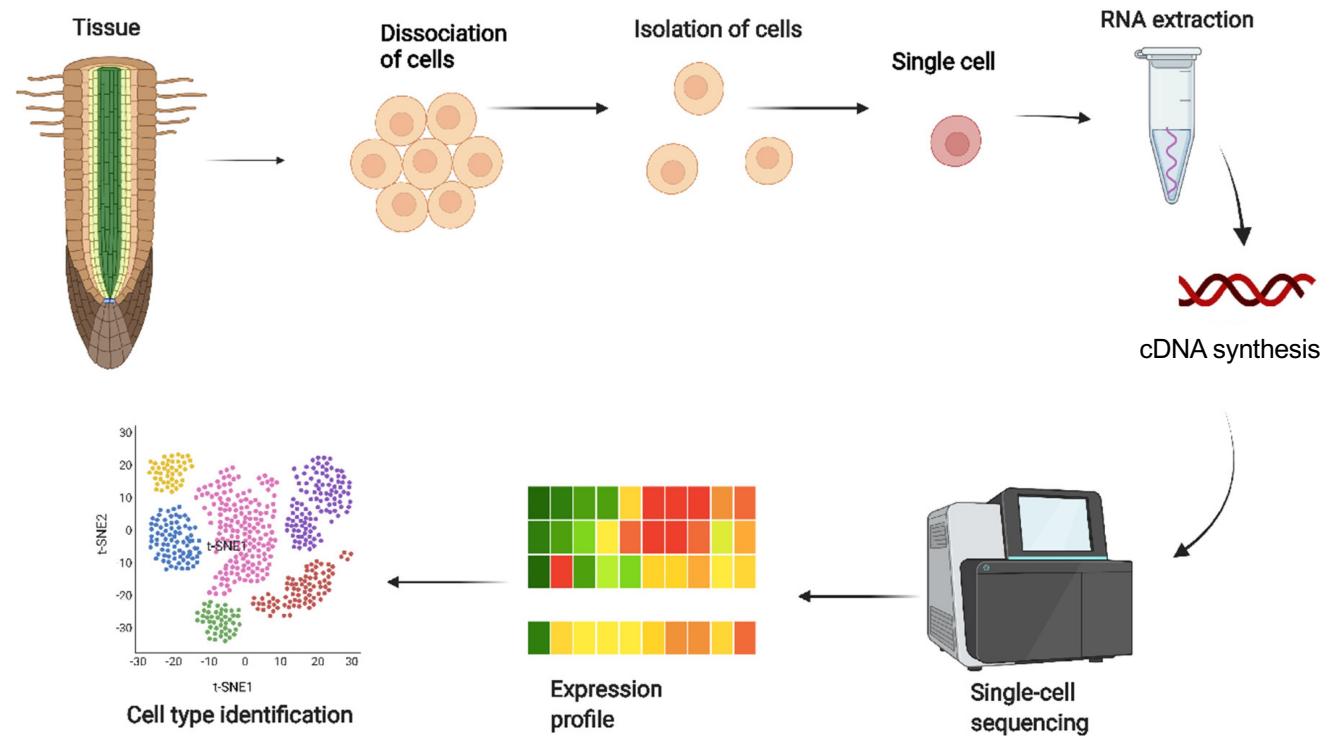
Stuart et al..(2019) "Integrative single-cell..." Nat Rev Genet 20, 257–272.

Single-cell/spatial omics technologies for TF activity inference

- Single cell RNA sequencing (scRNA-seq)
- Spatial transcriptomics
- 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)

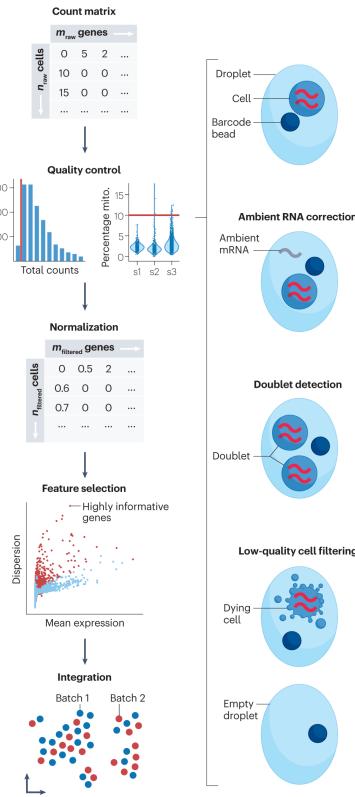
Single-cell gene expression (scRNA-seq)

- 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)



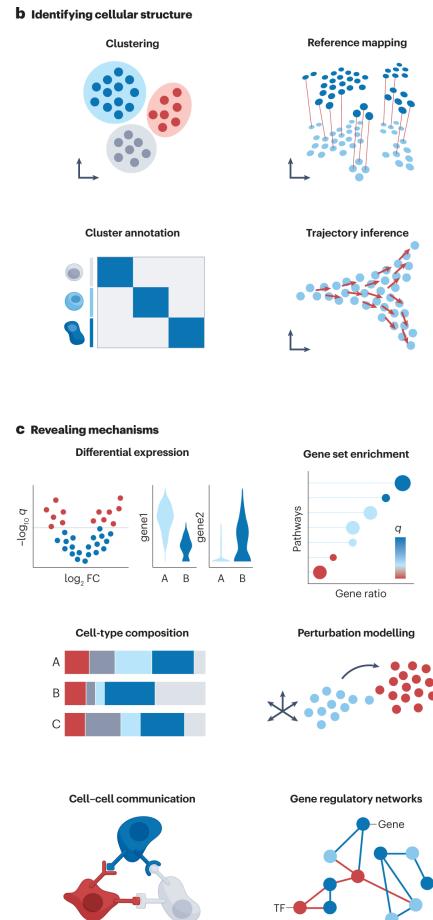
Data Processing and Quality Control:

- 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. 19

Single cell gene expression data analysis (2)



Cluster Analysis and Annotation:

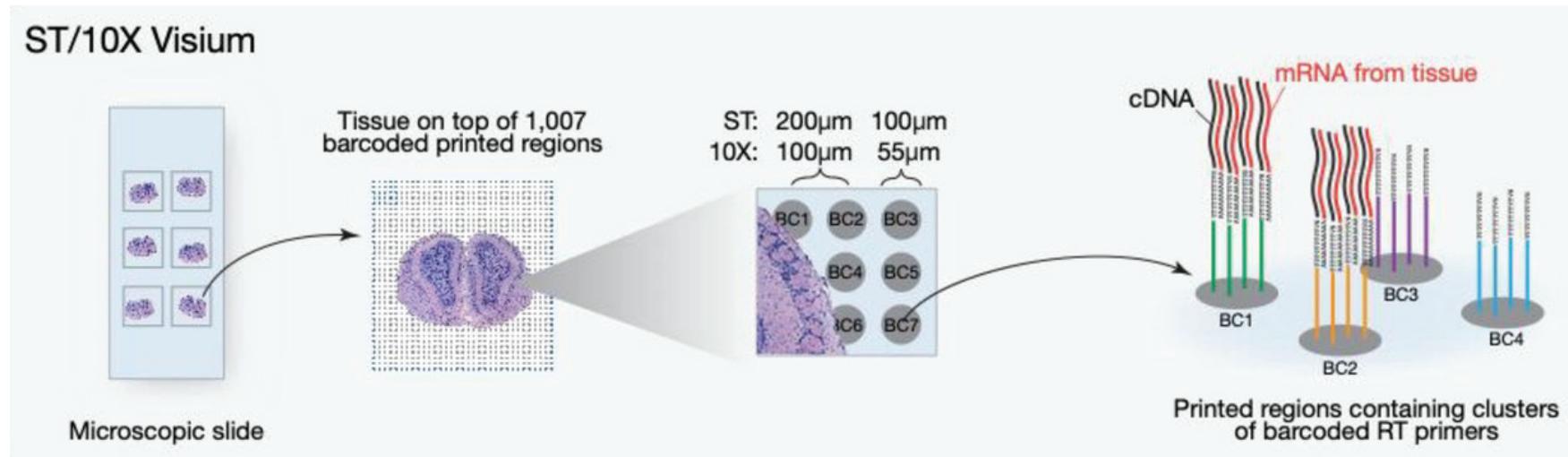
- Data is clustered based on gene expression profiles.
- Clusters are annotated with cell type labels manually or automatically.
- Continuous processes like cell differentiation transitions are inferred.

Interpretation and Analysis of scRNA-seq Data:

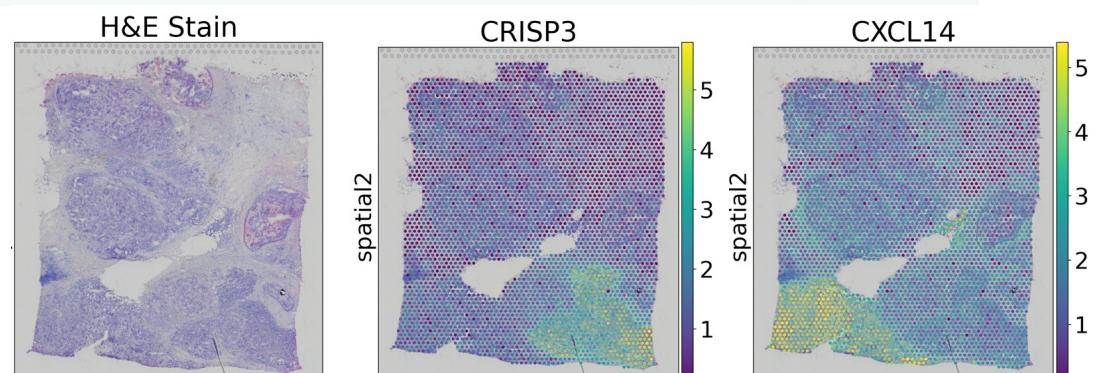
- Differential expression analysis identifies upregulated/downregulated genes.
- Gene set enrichment analysis assesses pathway effects.
- Changes in cell-type composition are examined.
- Perturbation modeling predicts effects of induced or unmeasured perturbations.
- Analysis of ligand-receptor expression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

Spatially resolved transcriptomics (genome-wide) data

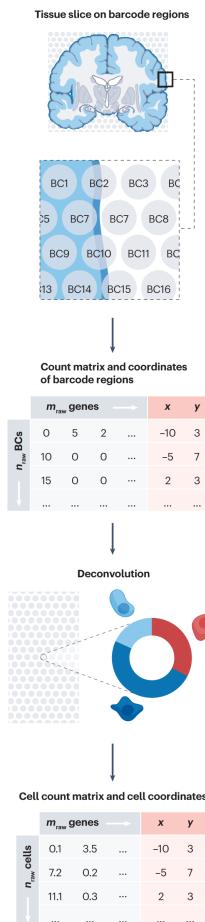
- Spatial transcriptomics measures mRNA across a tissue slice while preserving regional information



- 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



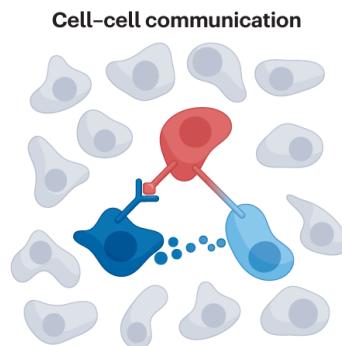
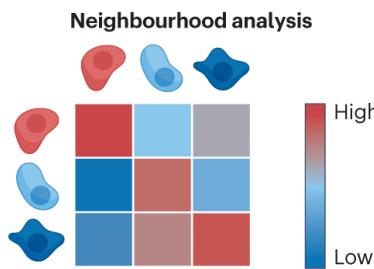
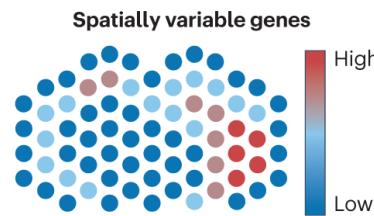
Spatial gene expression data analysis (1)



Array-based Spatial Transcriptomics:

- Quantifies gene expression in predefined barcoded (BC) regions spanning 10 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.

Spatial gene expression data analysis (2)

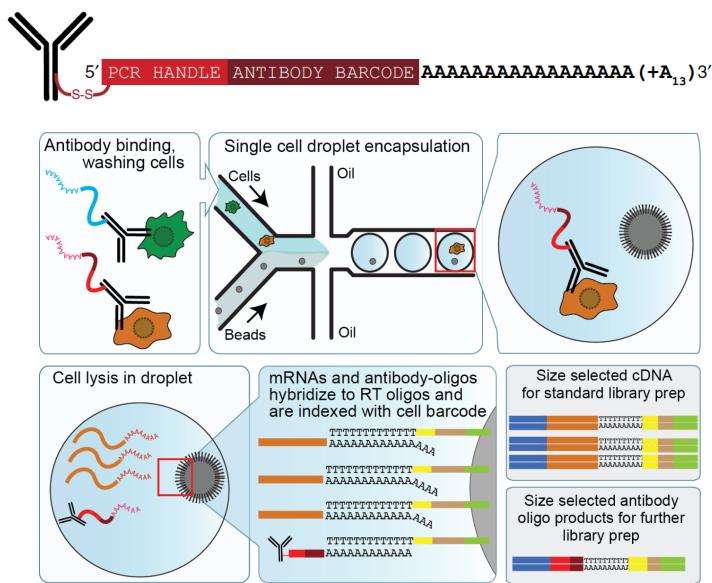


- Analyzes mechanisms via spatial positions, identifying spatially varying genes.
- Examines cell neighborhoods and infers communication events (receptors, ligands, tight junctions, mechanical and indirect mechanisms).

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

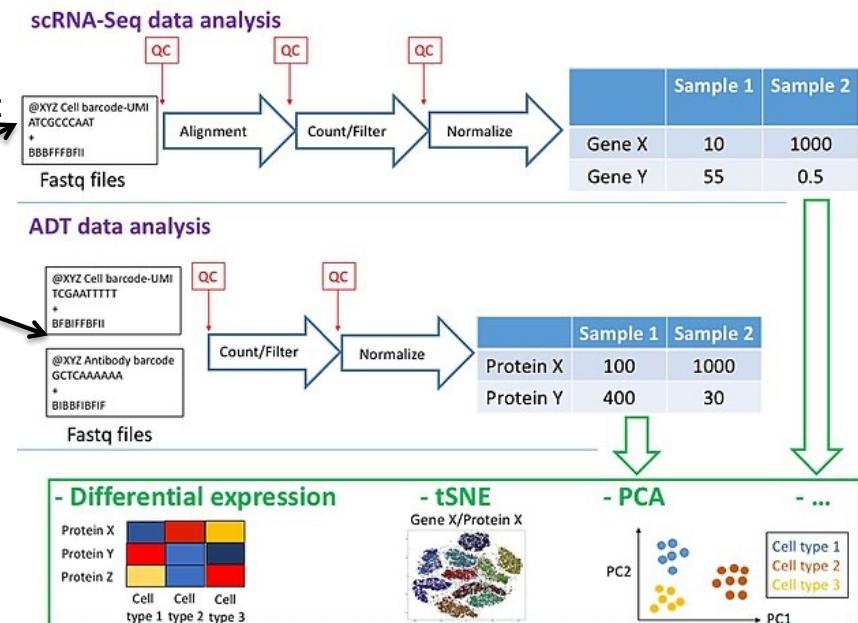
- Uses DNA-barcoded antibodies with polyA tail to convert detection of proteins into a quantitative, sequenceable readout
- Antibody-derived tags (ADTs) are antibody clones with unique barcodes attached to poly(A) sequences and a PCR handle.
- They bind to surface proteins, and their sequenced counts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



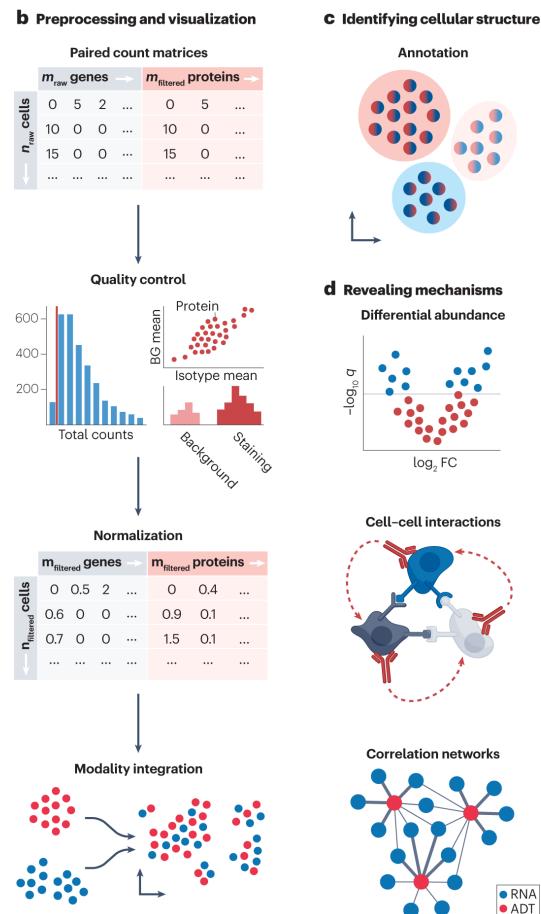
<https://cite-seq.com>

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:

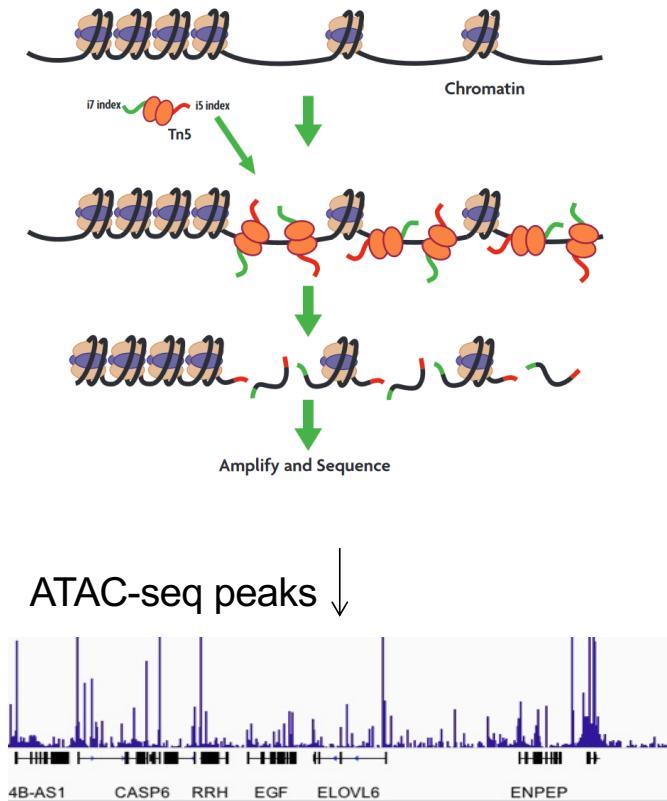
- Each dataset undergoes individual quality control and normalization.
- Data can be visualized individually or jointly to reveal relationships.
- Annotation can be based on transcriptomics data, ADT data, or both.
- Clusters are matched to marker genes and ADTs to annotate cell types accurately.

Biological Mechanisms and Analysis:

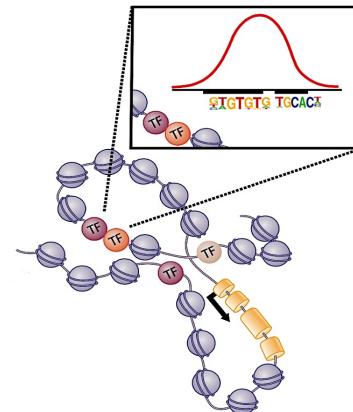
- ADT data is analyzed for differential abundance to understand protein expression changes.
- Cell-cell communication is inferred based on ADT profiles.
- Correlation networks integrate RNA and ADT information to elucidate biological interactions.

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



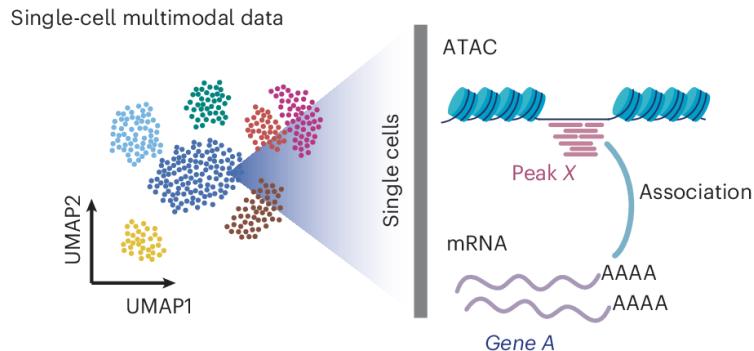
TF motif prediction in ATAC-seq peak regions



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

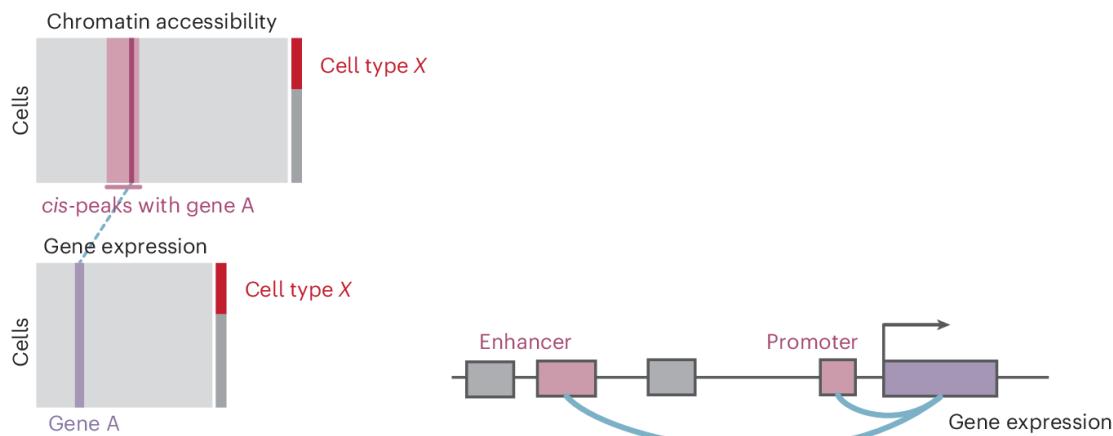
Single-cell multiome

a

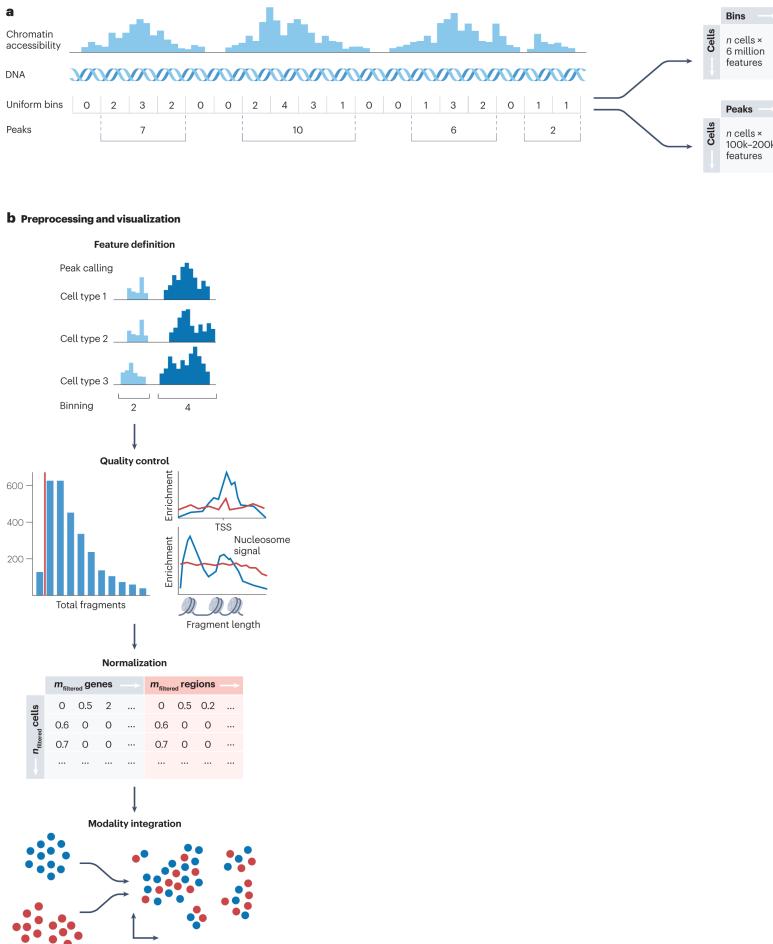


- Combines single-cell gene expression analysis with single-cell open chromatin mapping to provide genome-wide mapping of both the transcriptional and epigenetic landscapes at the single-cell level

b



scATAC-seq analysis steps (1)

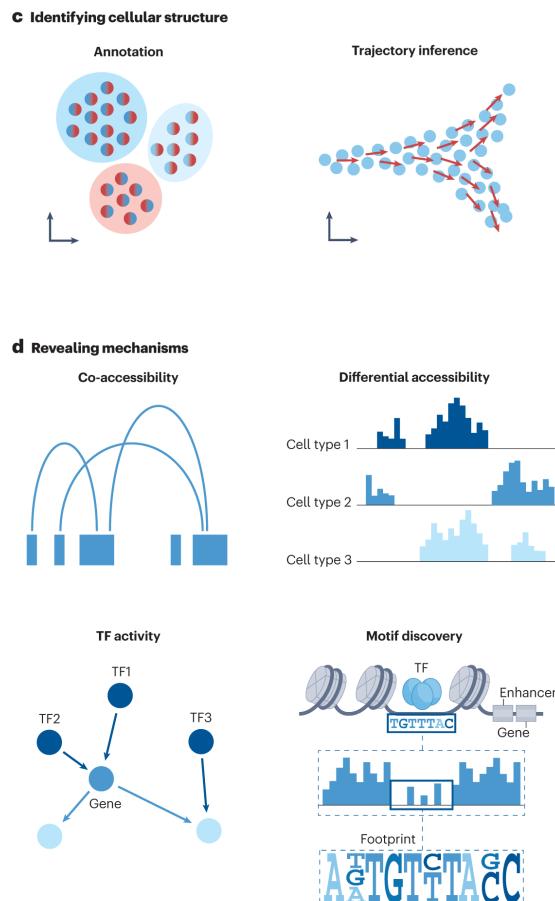


Quality Control and Preprocessing:

- 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)



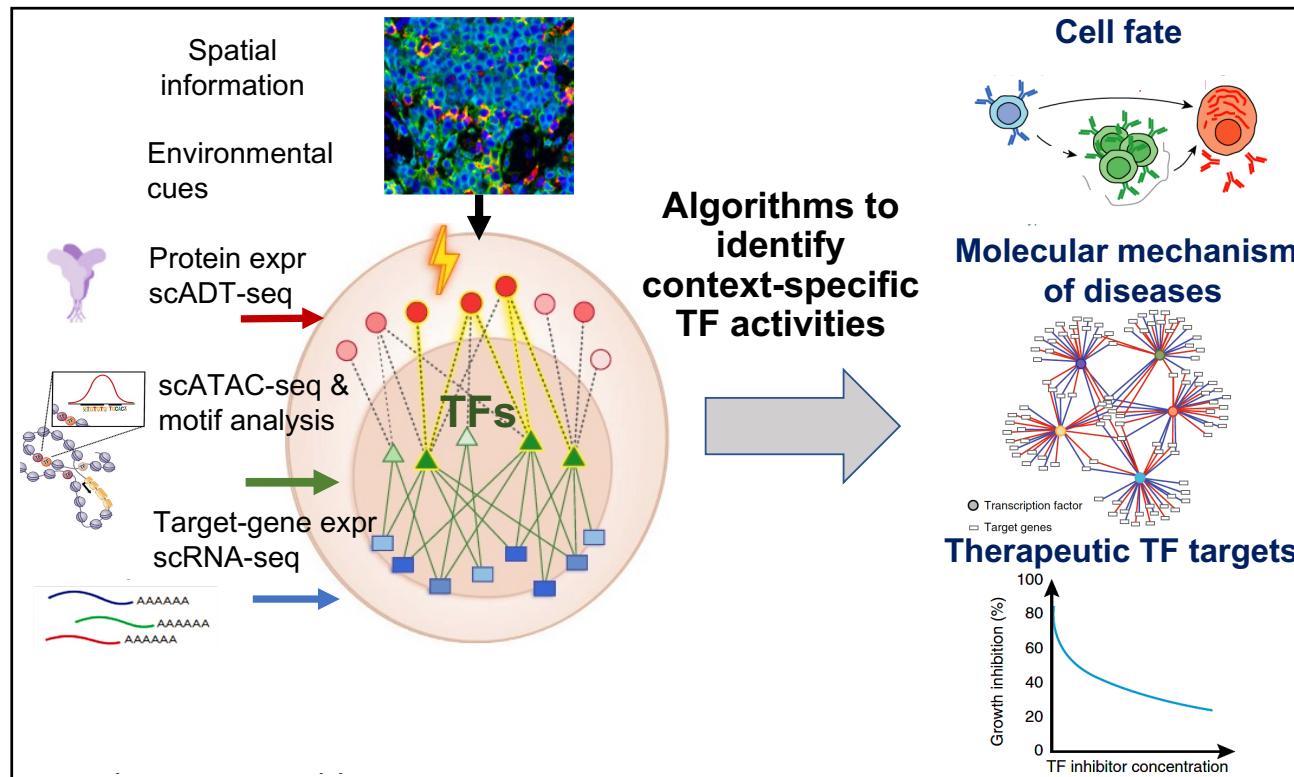
Annotation, Cell Type Identification ana Data Analysis:

- 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 omics

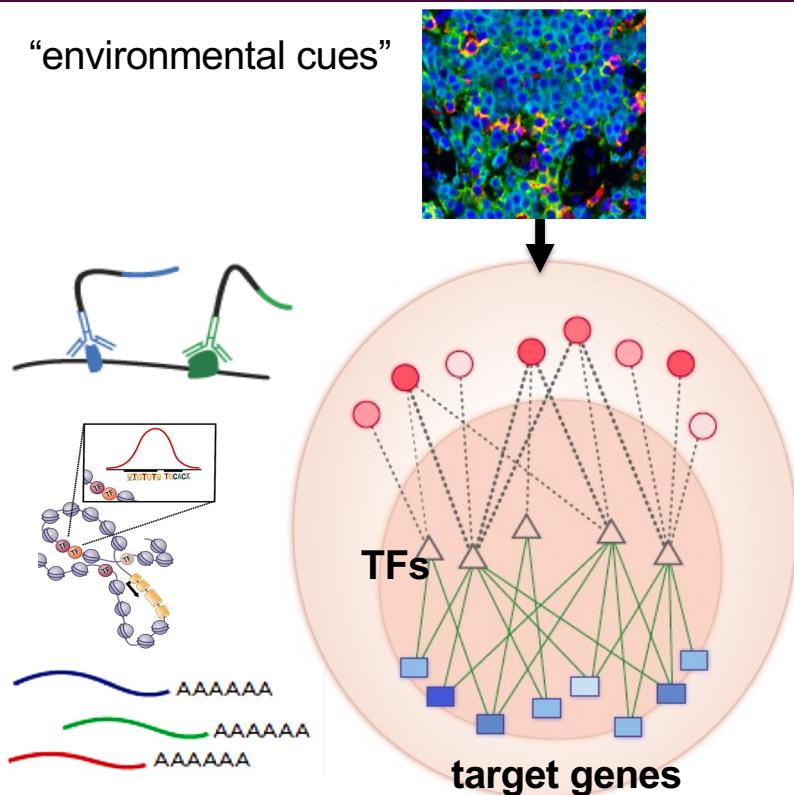
Computational methods for modeling context-specific transcription factor activities using single-cell and spatial omics approaches

- 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

“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

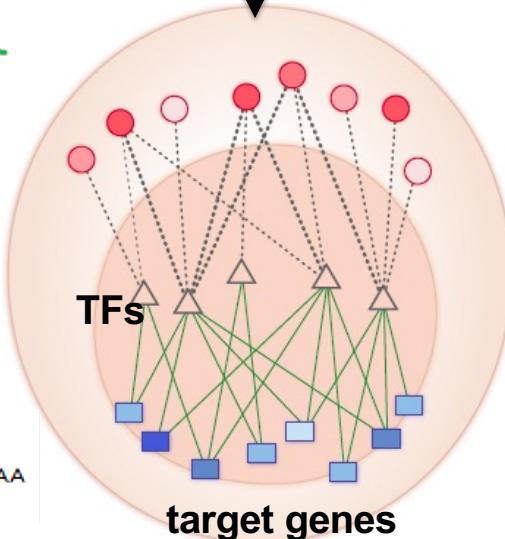
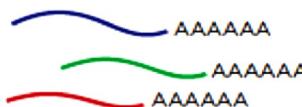
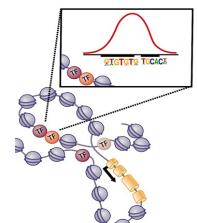
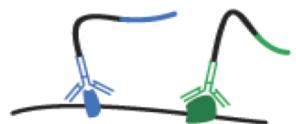
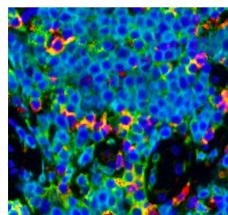
The major classes of methods for TFA inference



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

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

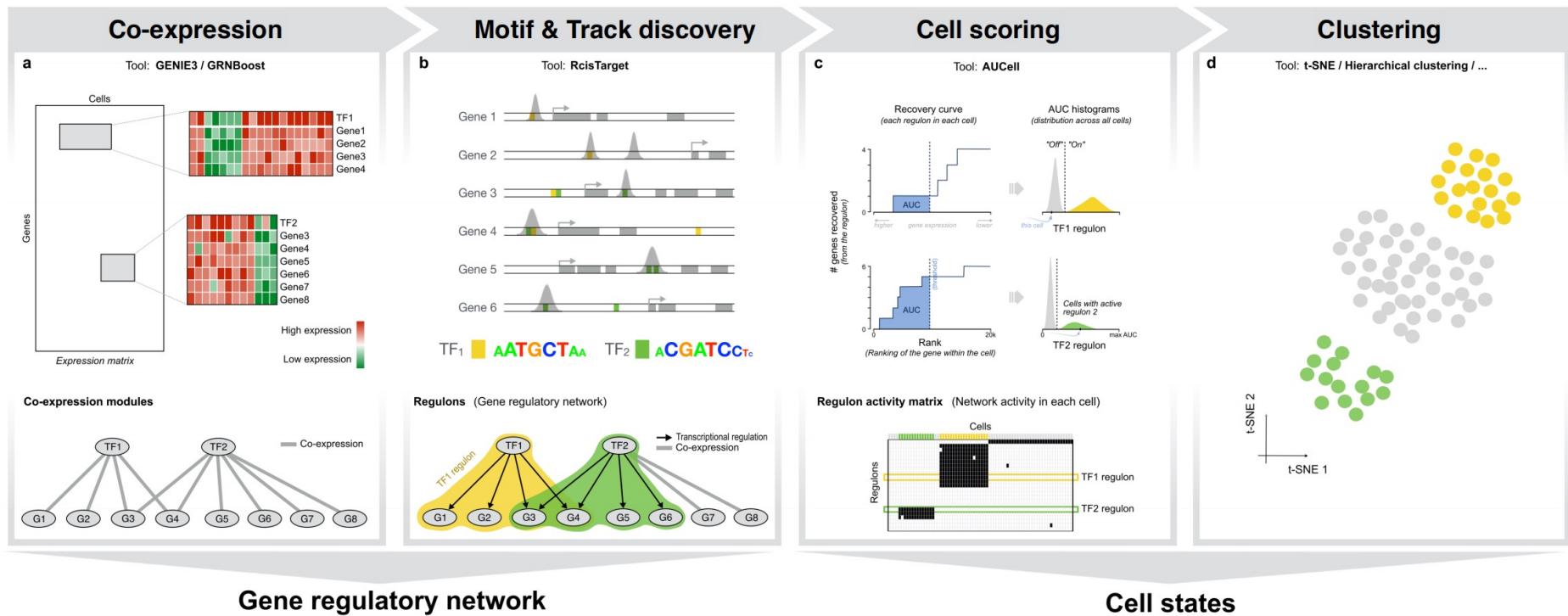
e.g. BITFAM, chromVAR, scBAsset, scFAN

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)



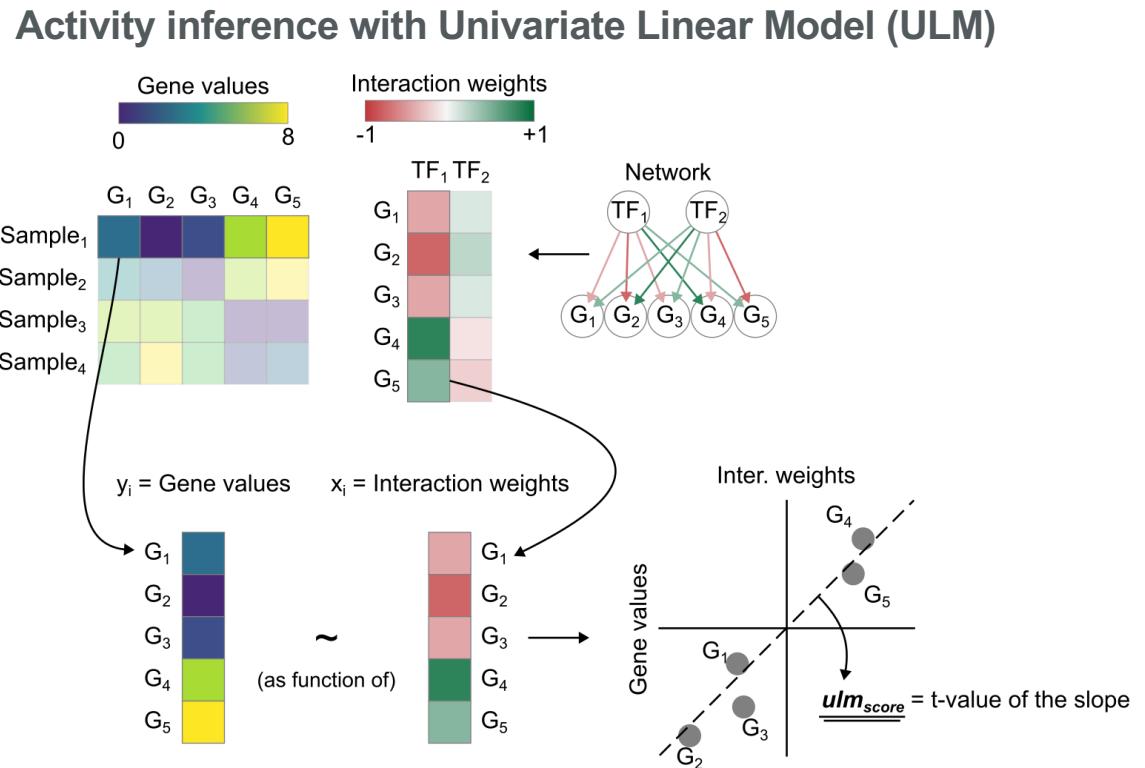
<https://github.com/aertslab/SCENIC>

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

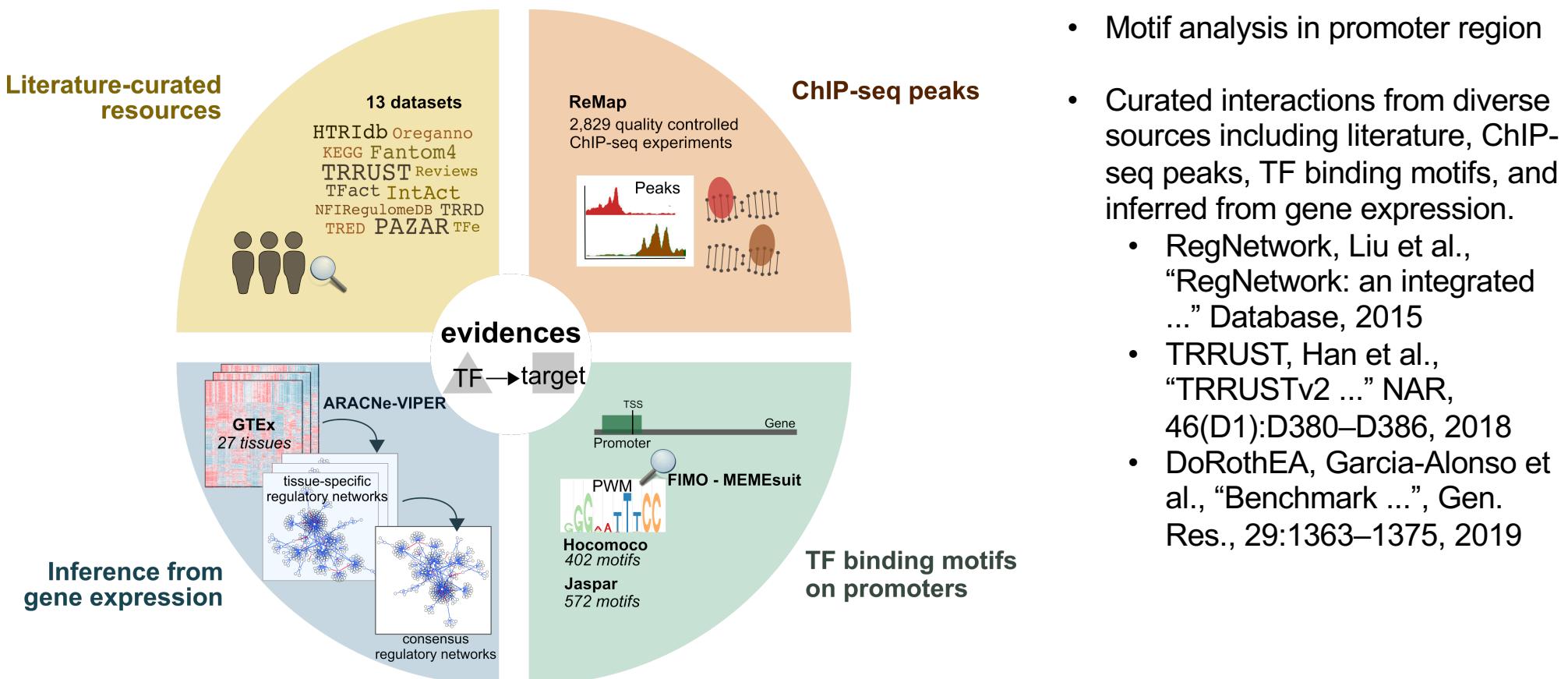
Nature Methods 14, 1083 (2017)

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

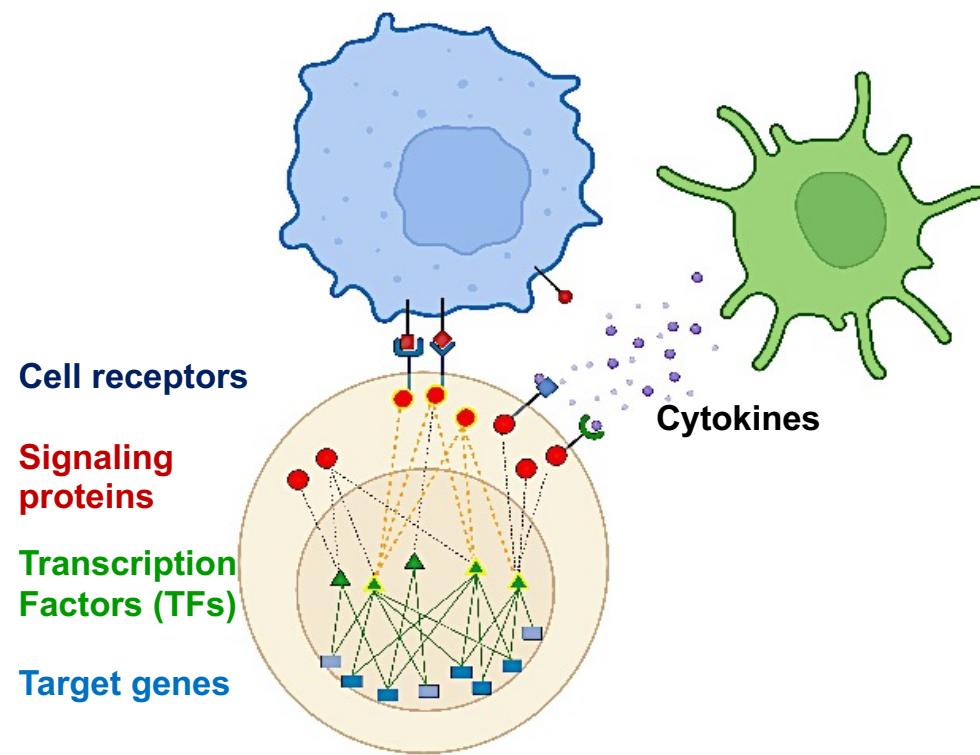


<https://saezlab.github.io/dorothea/>

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

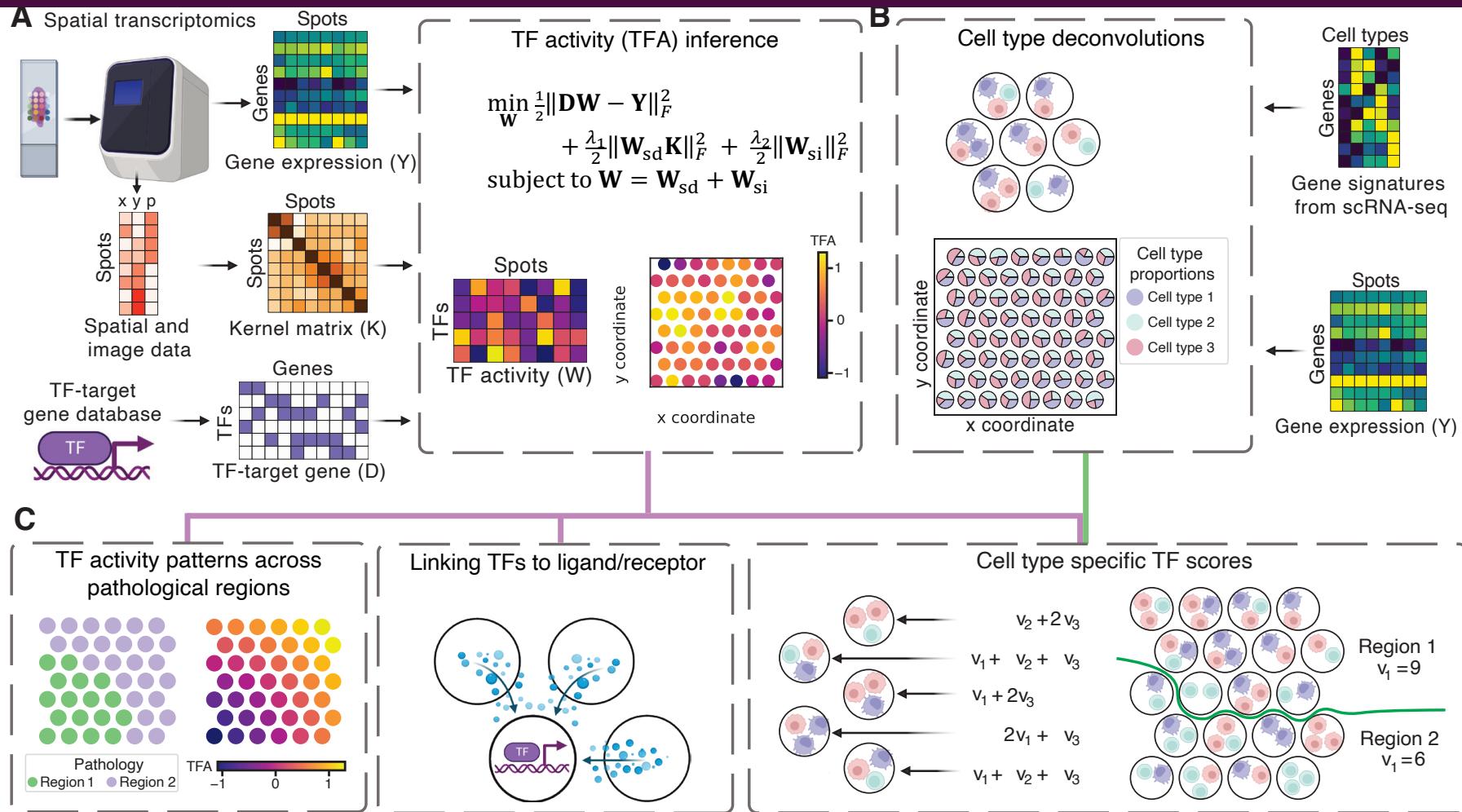
- INFERELATOR 3.0
 - <https://inferelator.readthedocs.io/en/latest/tutorial.html>
- BITFAM: The Bayesian inference transcription factor activity model
 - <https://github.com/jaleesr/BITFAM>
- metaVIPER
 - <https://github.com/califano-lab/single-cell-pipeline>

TF activity inference from spatial transcriptomics



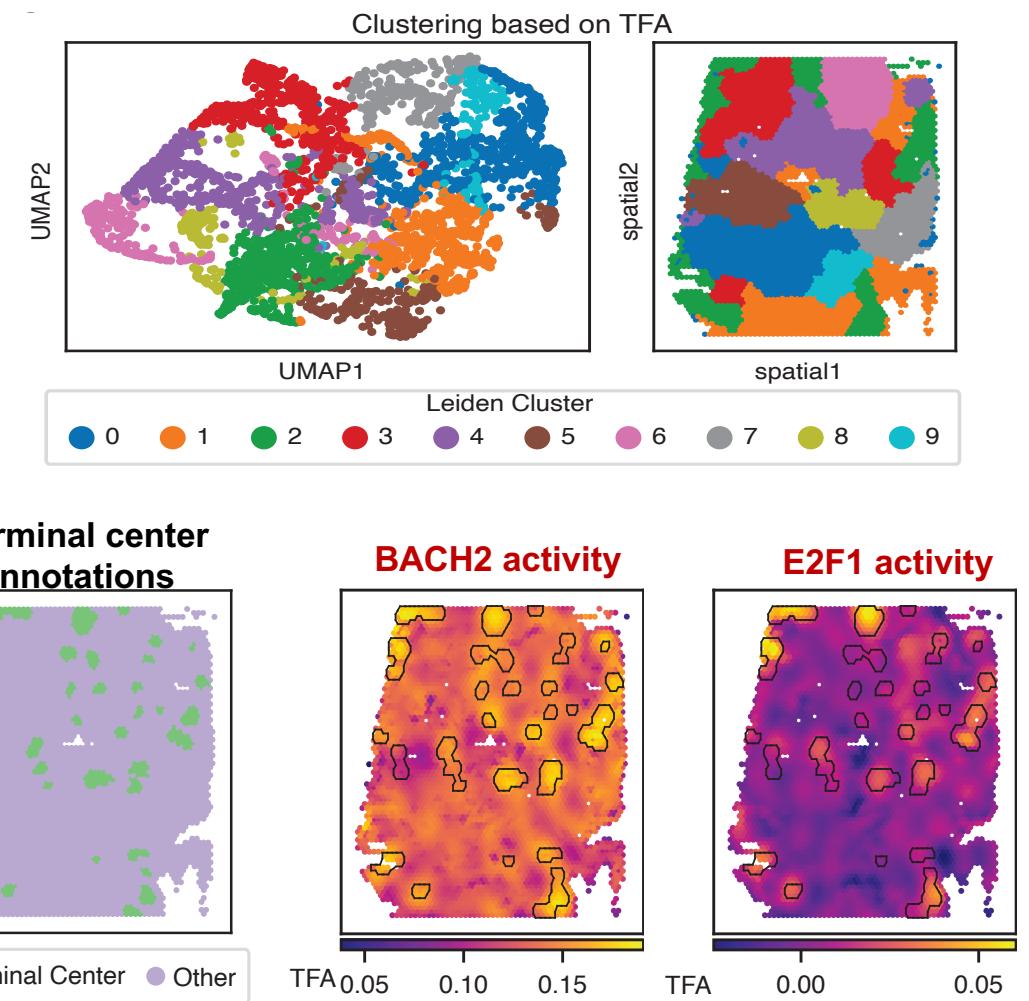
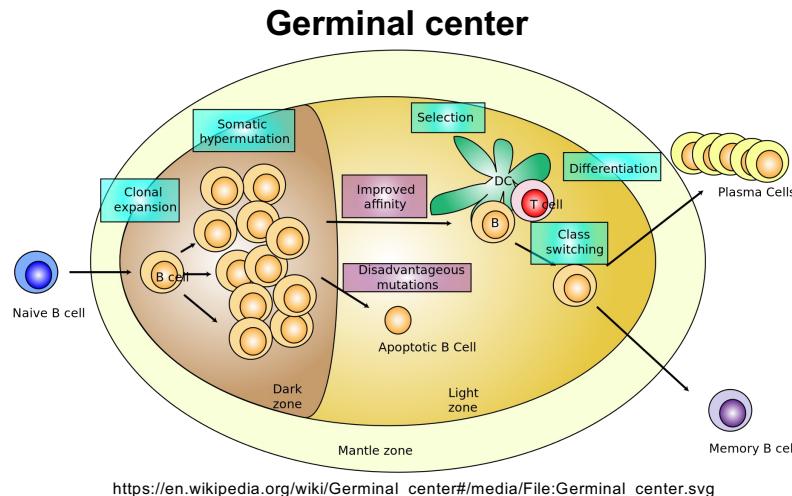
Nearby cells often affect cell context-specific TF activities through receptor-ligand mediated cell-cell communication or secreting molecules like cytokines

Spatially informed Transcription factor Activity Network (STAN)

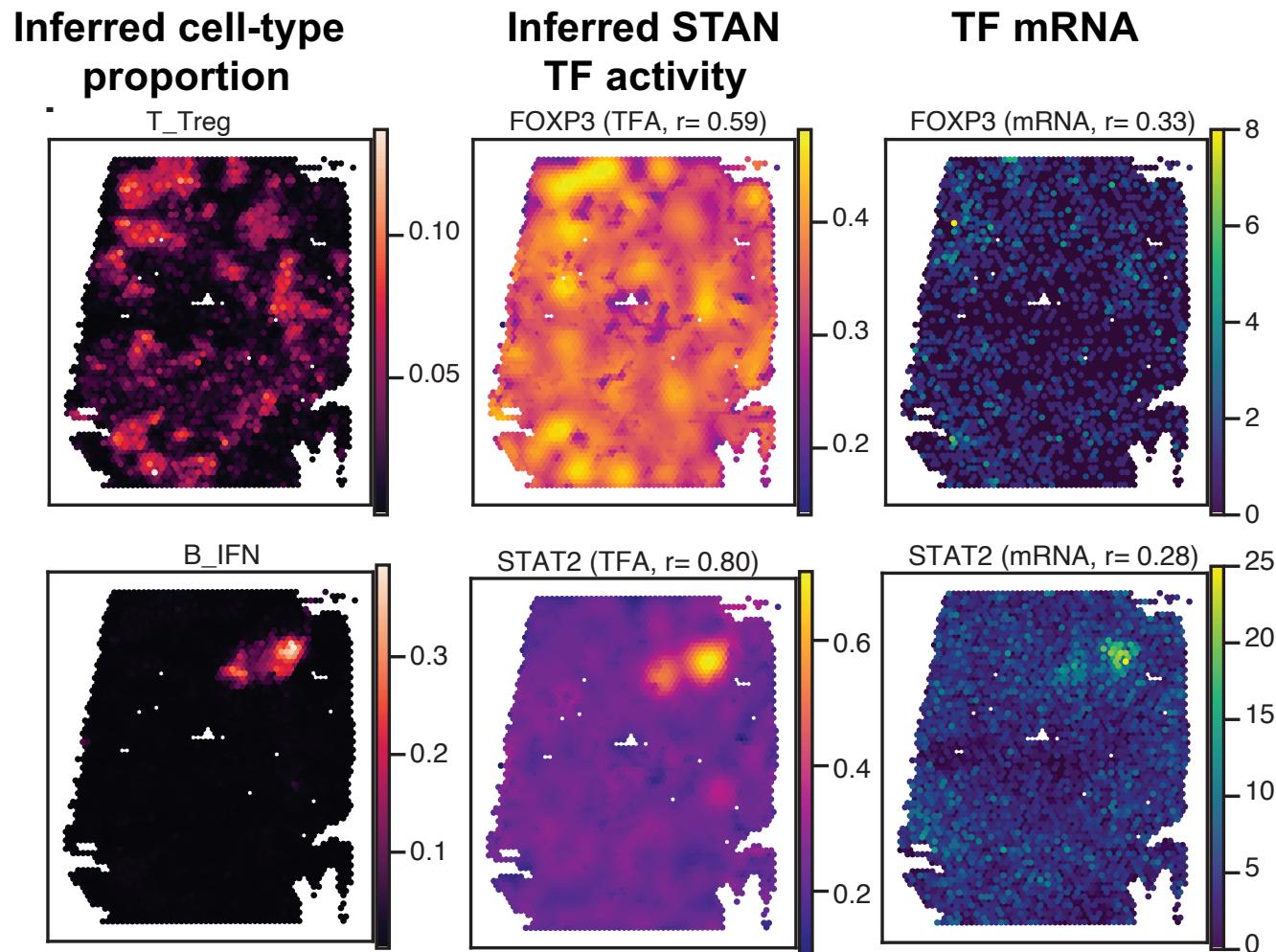


Clustering on STAN predicted TF activities (TFAs) yields spatially coherent clustering in lymph node

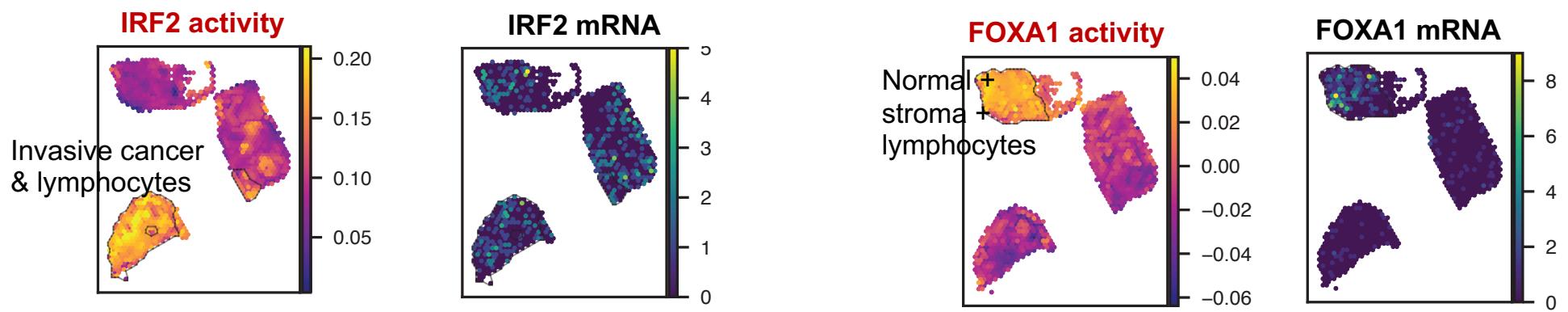
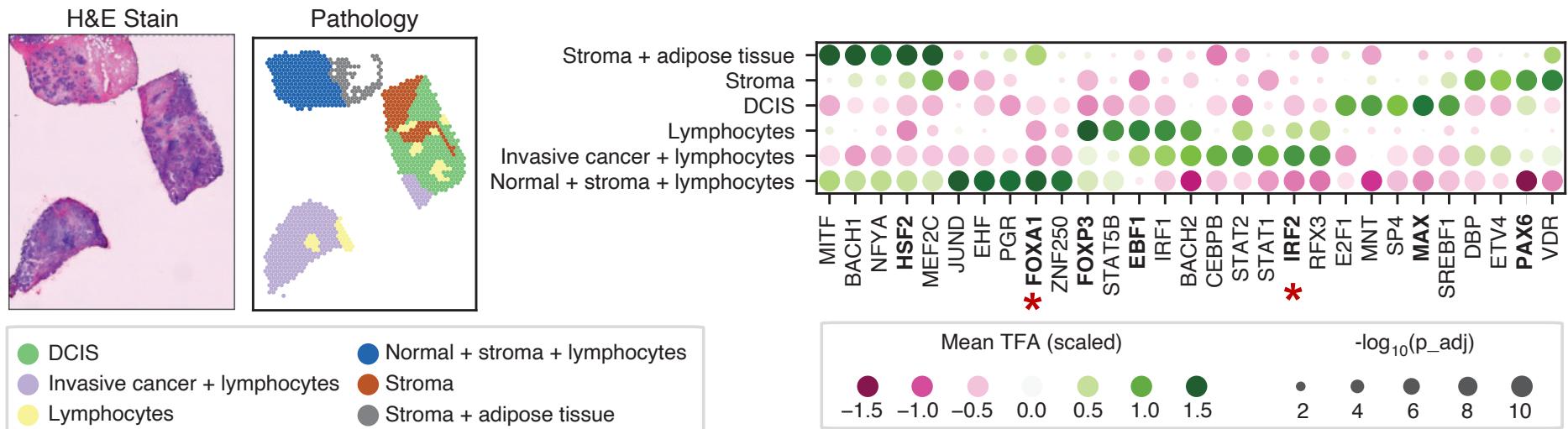
- The human lymph node is characterized by dynamic microenvironments with many spatially interlaced cell populations such as germinal centers (GCs)
- Germinal centers develop in the B cell follicles of secondary lymphoid tissues during T cell-dependent antibody responses



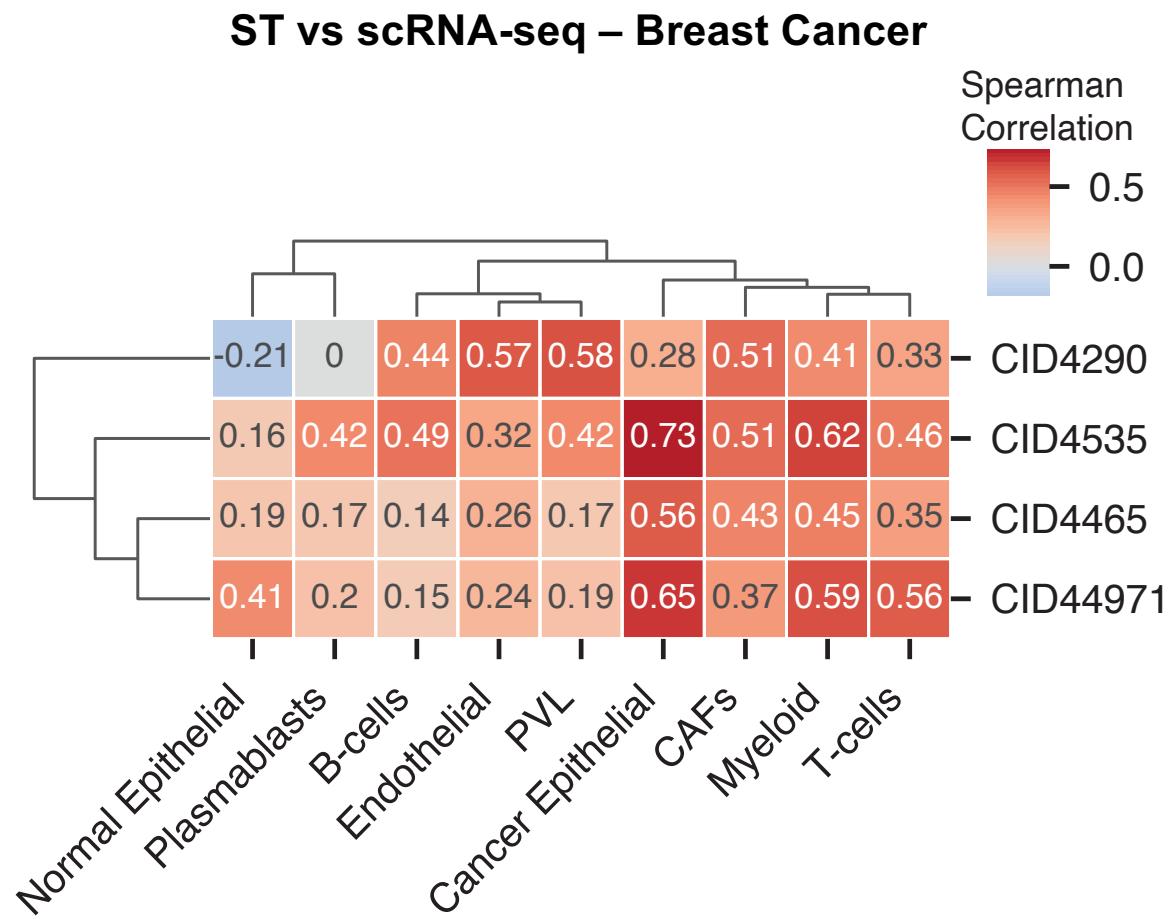
STAN identifies cell-type specific TFs – lymph node



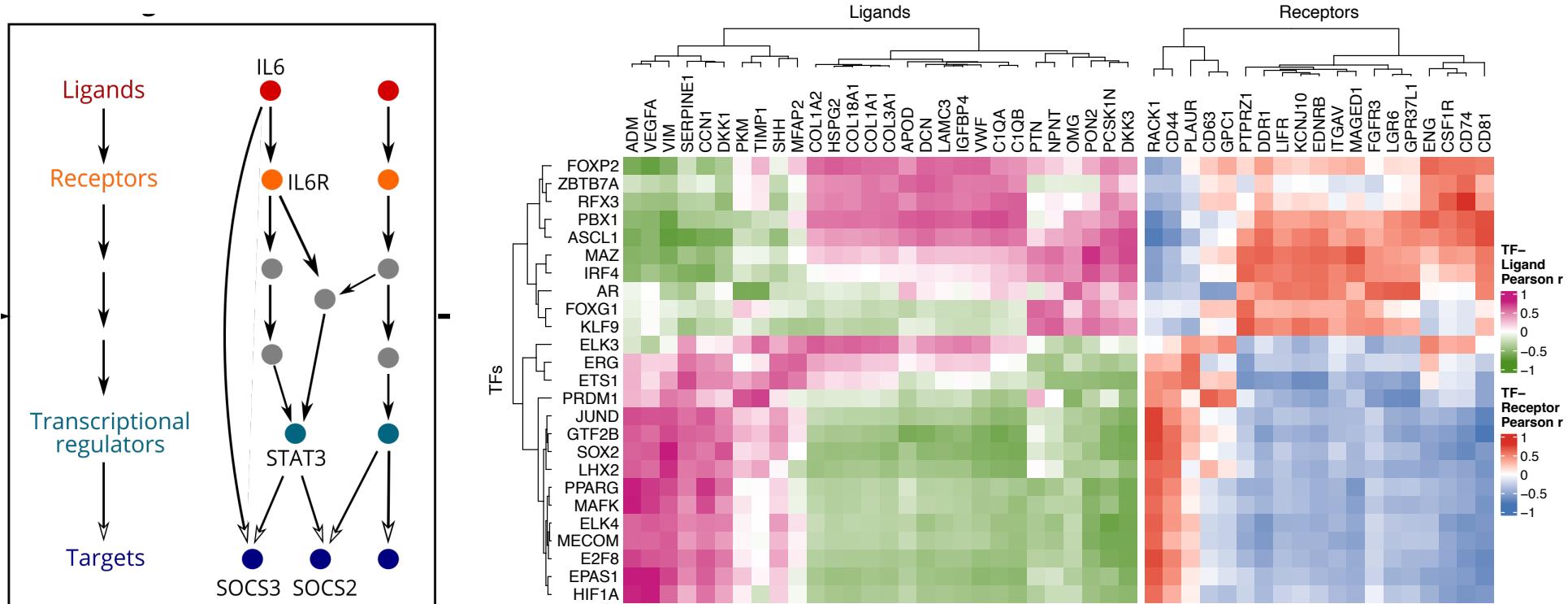
Inferred STAN TF activity associated with pathological regions – triple negative breast cancer (TNBC)



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

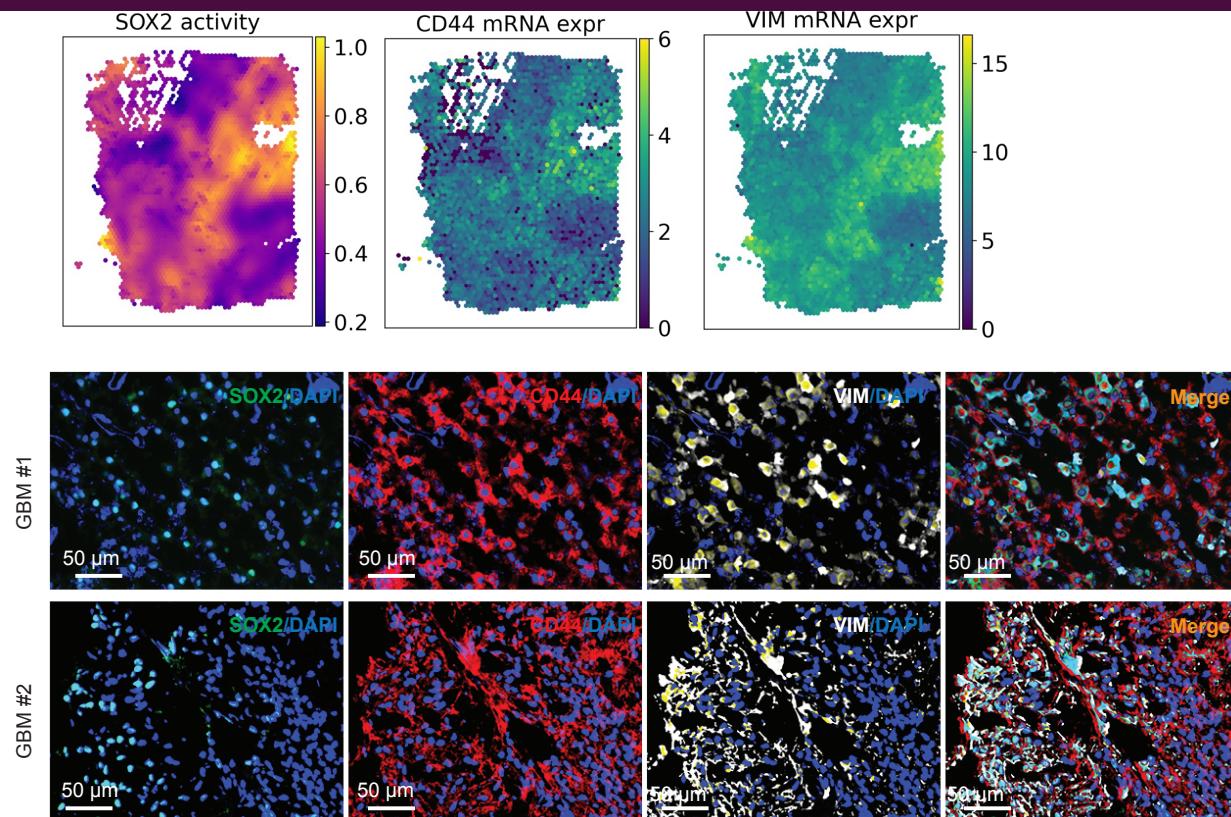


Linking ligands and receptors to TFs in glioblastoma



Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* 17, 159–162 (2020).

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>

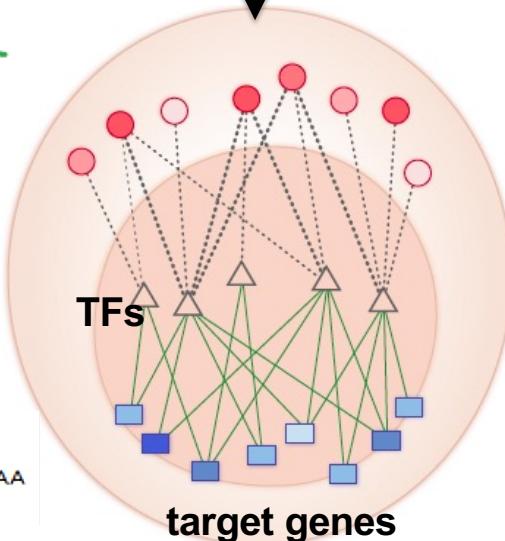
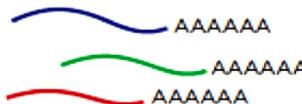
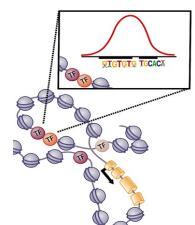
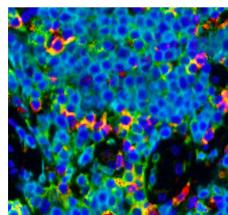
Zhang L, Sagan A, Qin B, Hu B, Osmanbeyoglu HU, STAN, a computational framework for inferring spatially informed transcription factor activity across cellular contexts, bioRxiv 2024.06.26.600782; doi: <https://doi.org/10.1101/2024.06.26.600782>

Hands-on session 1

- Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public scRNA-seq and spatial transcriptomics with Linan Zhang
- Decoupler, SCENIC (optional), STAN
- https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/tree/main/hands-on_tutorial/session-1

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

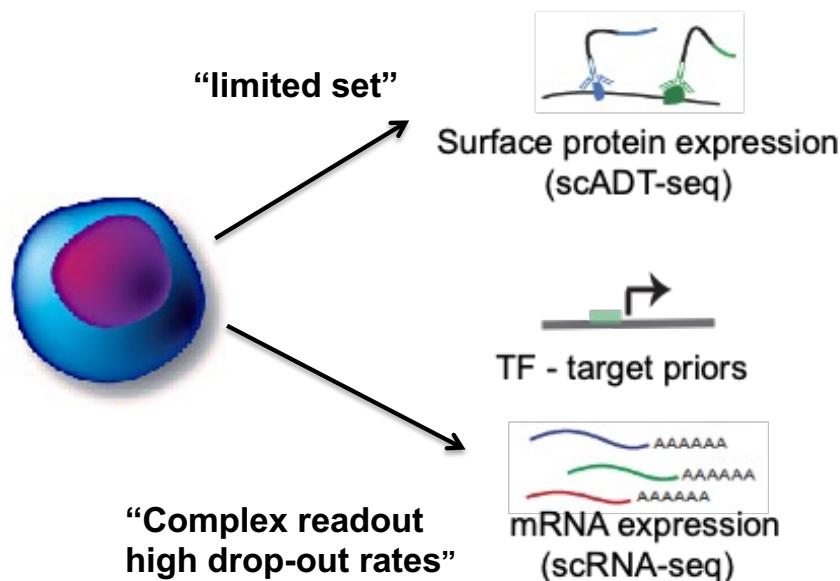
TF activity inference from single-cell gene expression data

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

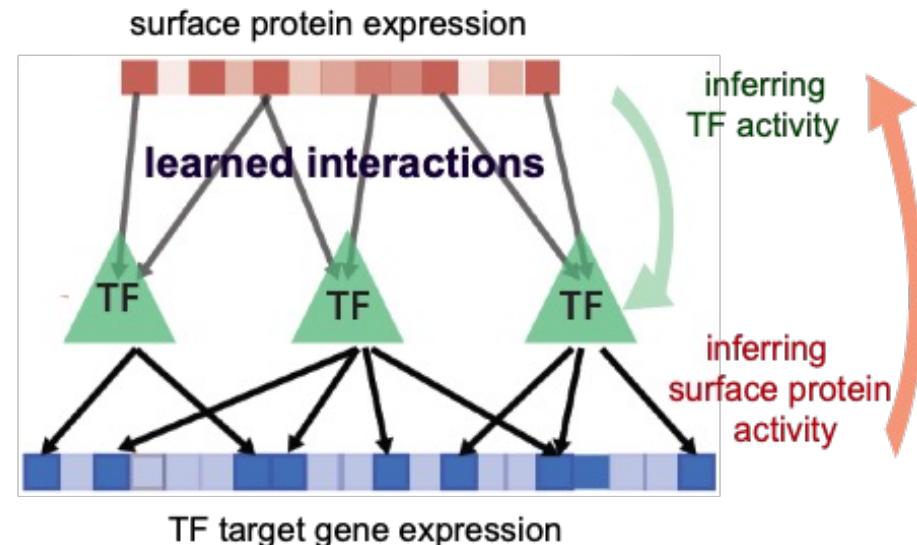
TF activity inference from single-cell proteomics and transcriptomics data

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

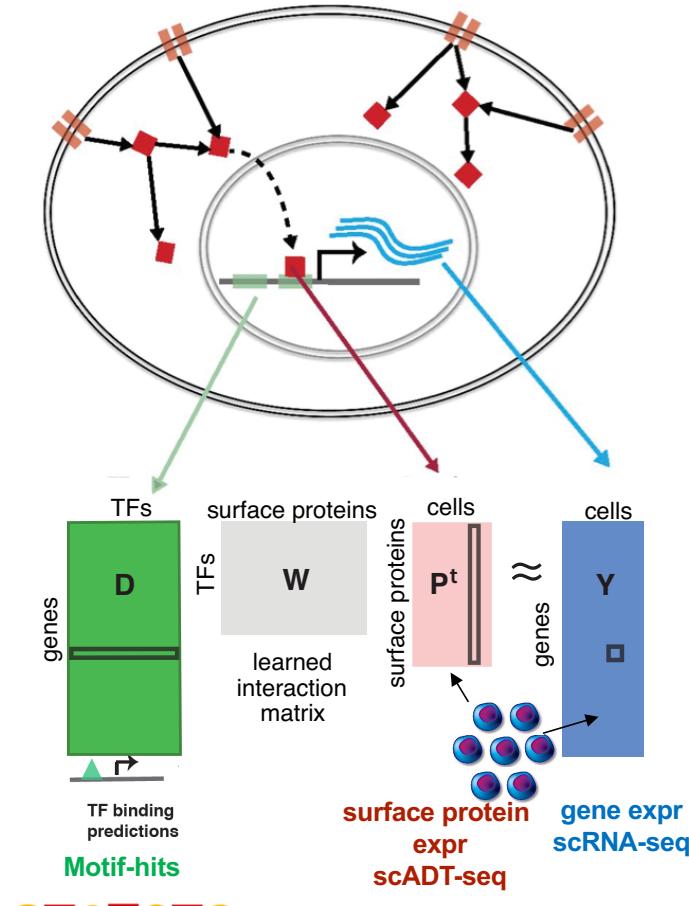
Single cell multi-omics profiling CITE-seq



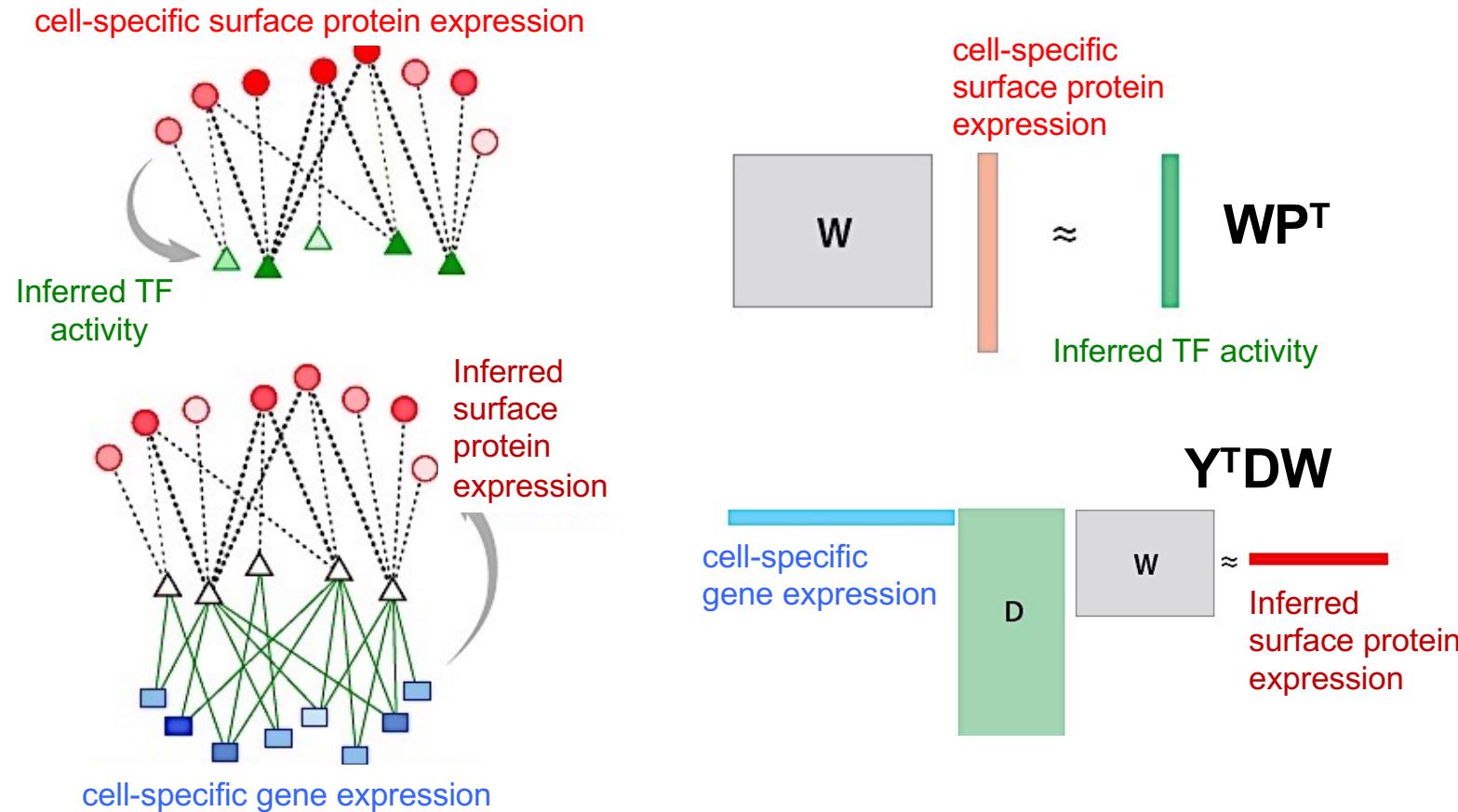
Linking surface proteins to transcriptional regulators



SPaRTAN - Linking surface proteins to TFs

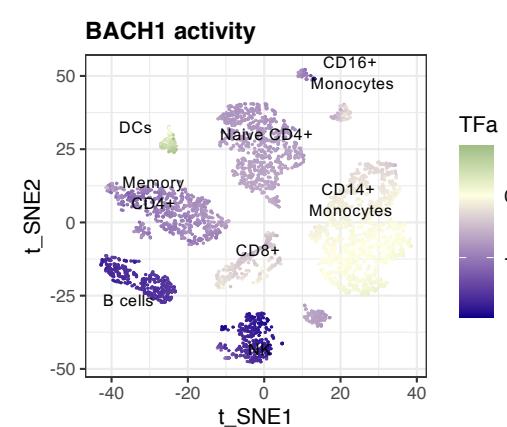
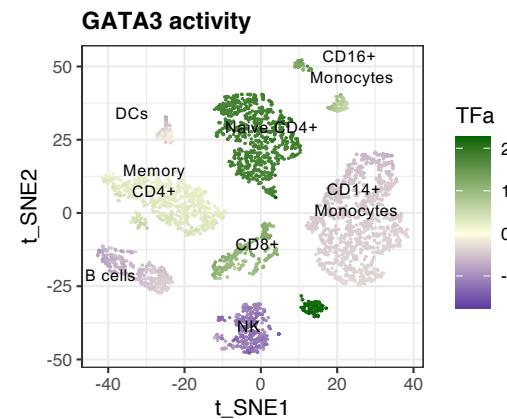
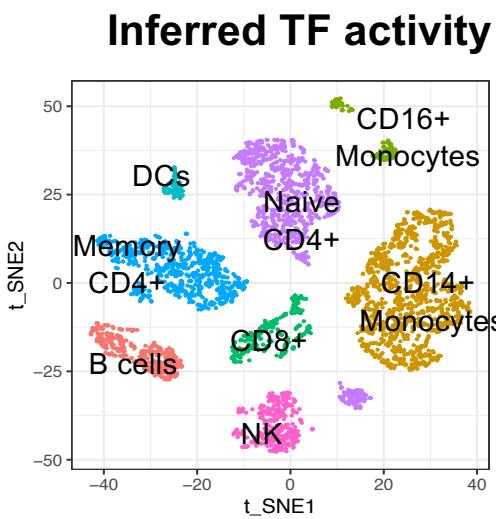
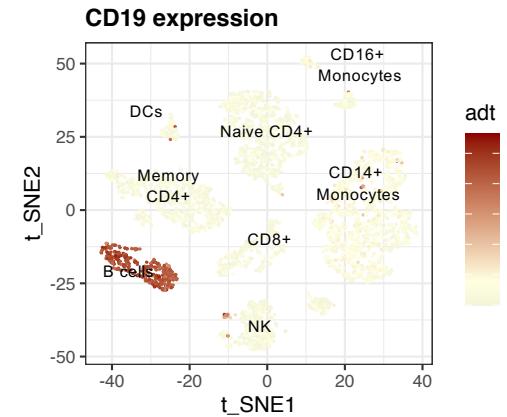
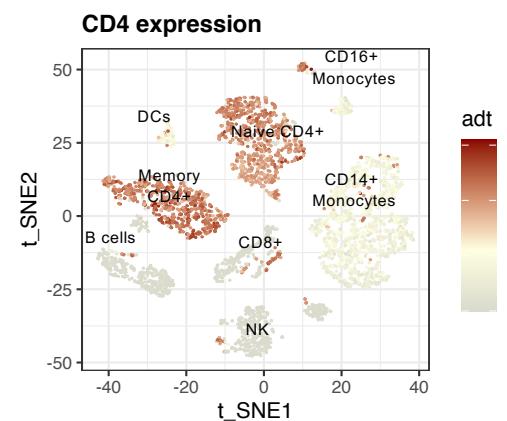
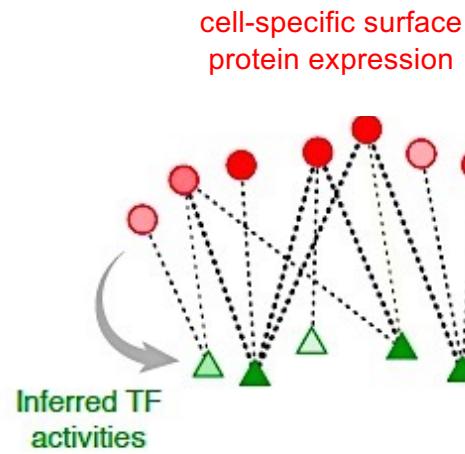
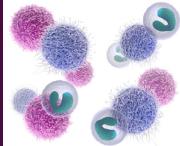


Inferring cell-specific TF activity and surface protein expression

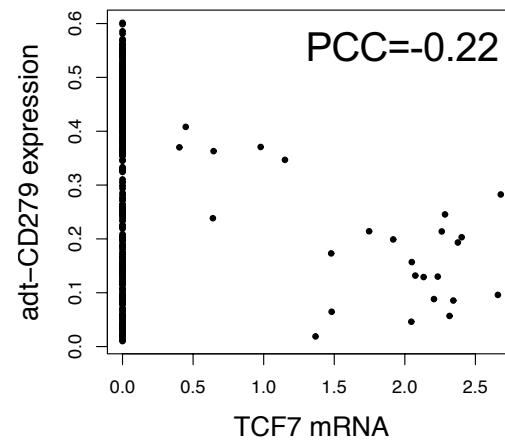
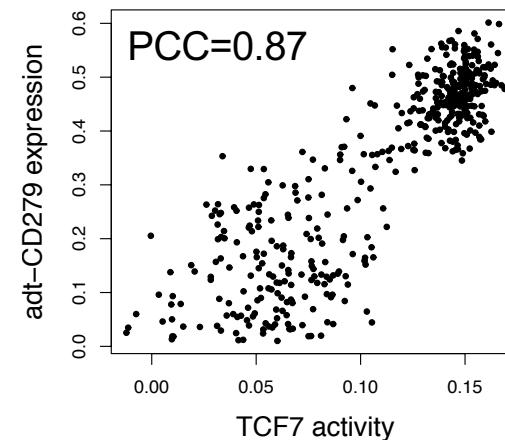
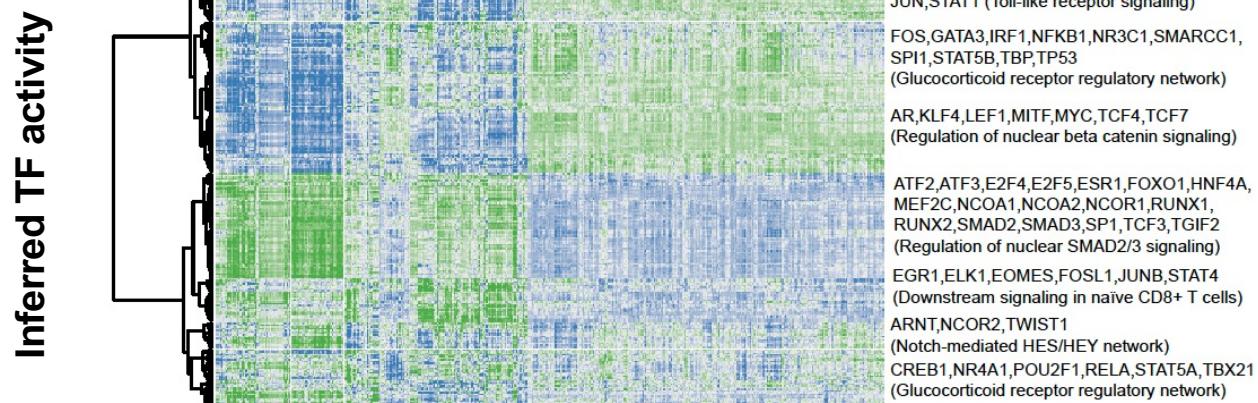
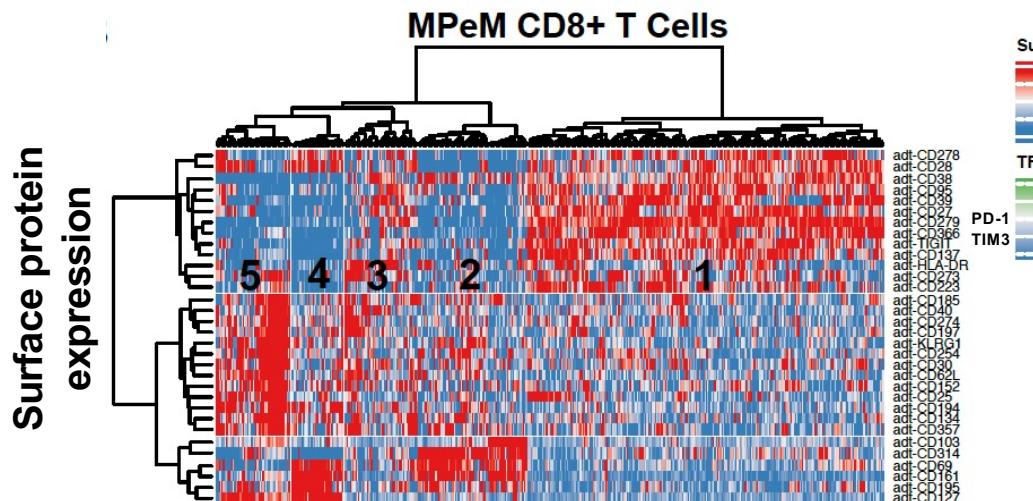


Reduced dimension in terms of TF activities and surface protein expression make data more tractable and reduce noise in single cell data while preserving the often intrinsically low dimensional signal of interest

SPaRTAN identifies both known and novel cell type-specific TFs



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

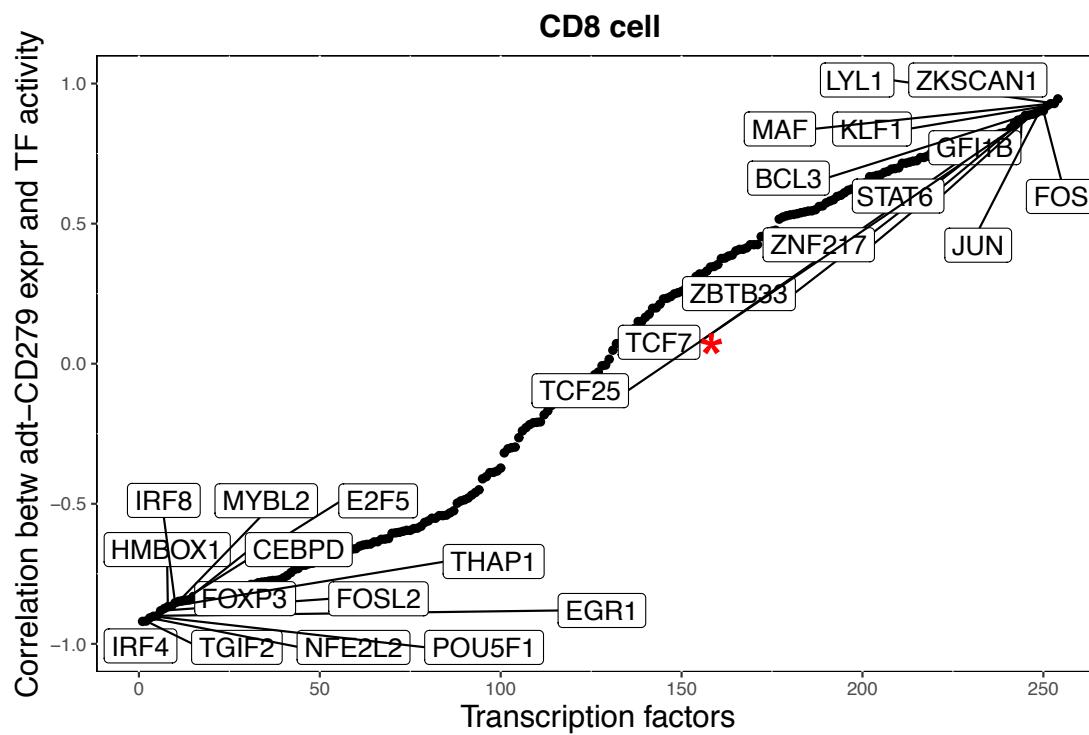


TCF7 is required for memory cells and efficacy of immunotherapies

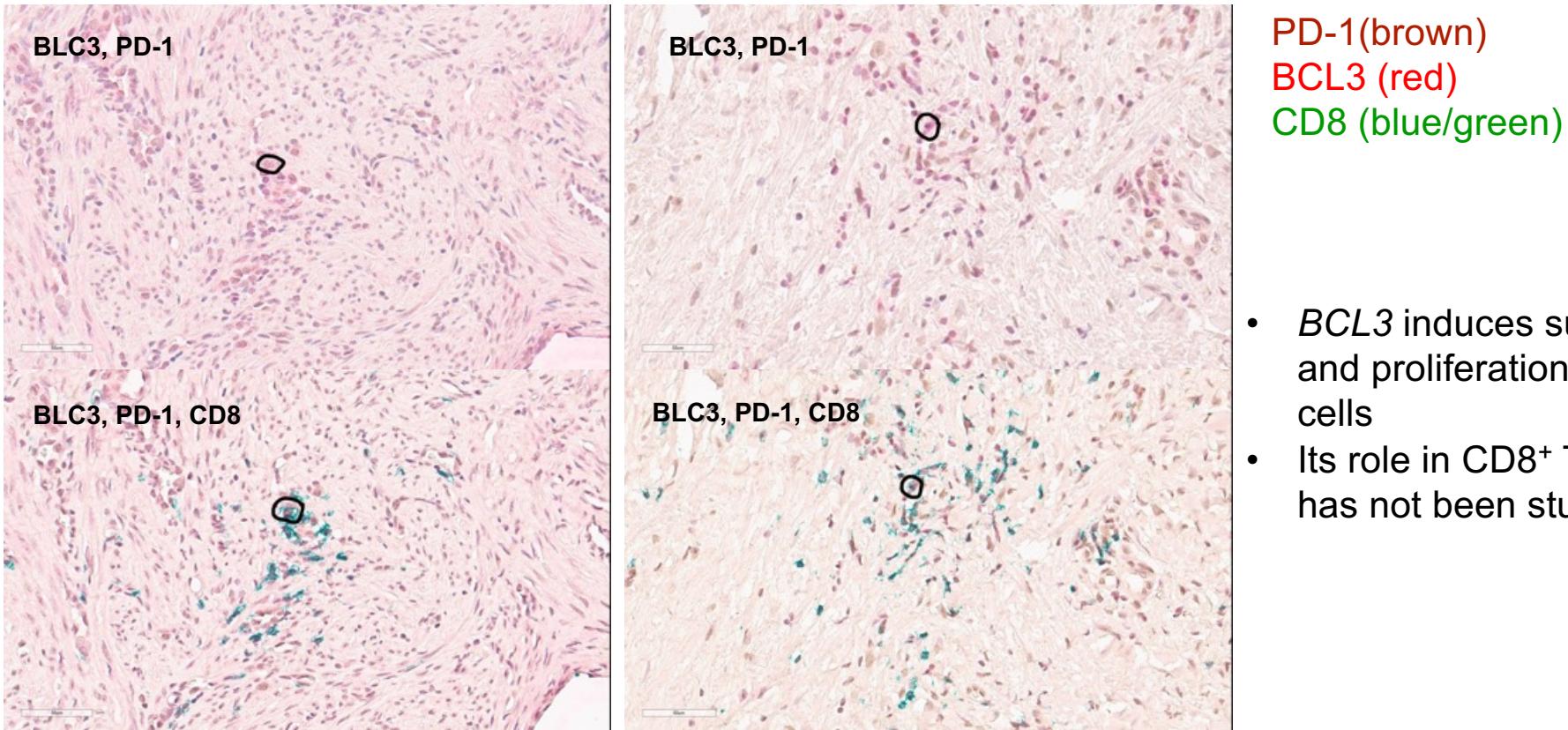
Siddiqui et al.
Immunity 2019
Kurtulus et al.,
Immunity 2019

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

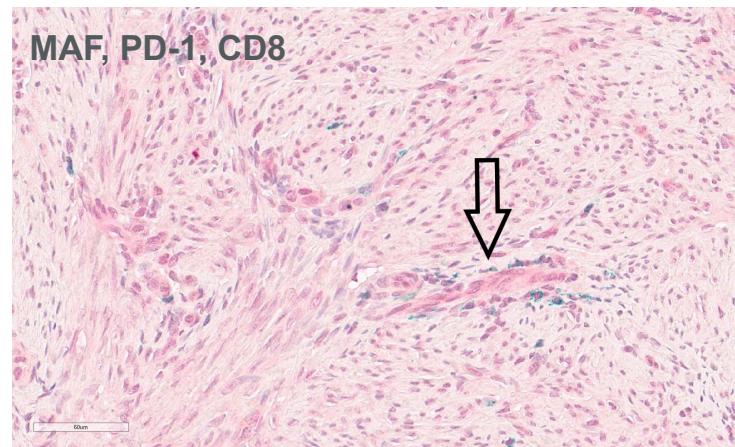
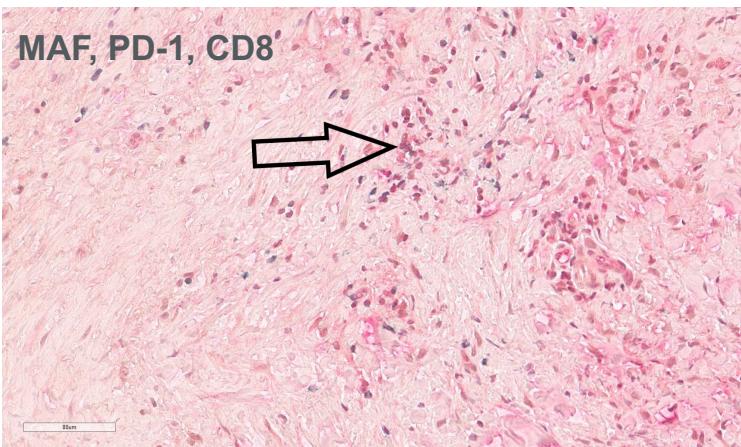
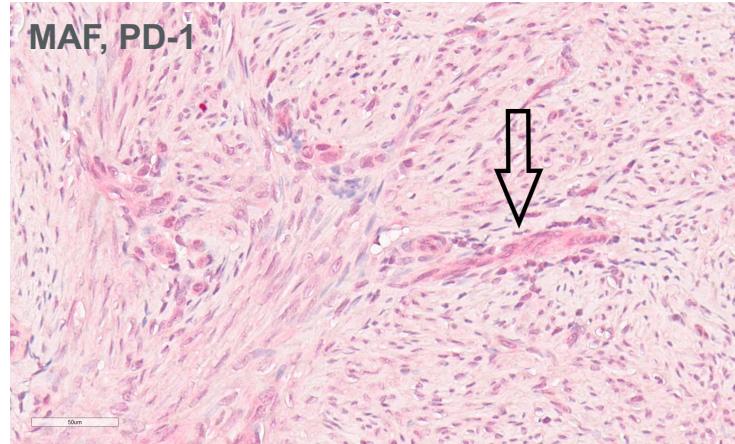
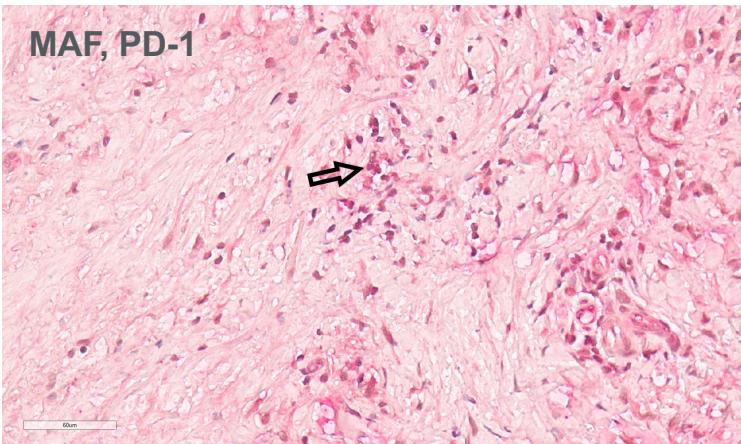


National Mesothelioma Virtual Bnk (NMVB) Validation Cohort Co-expression of CD8, PD1, BCL3 by IHC staining



- *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 commonalities 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>

Hands-on session 2

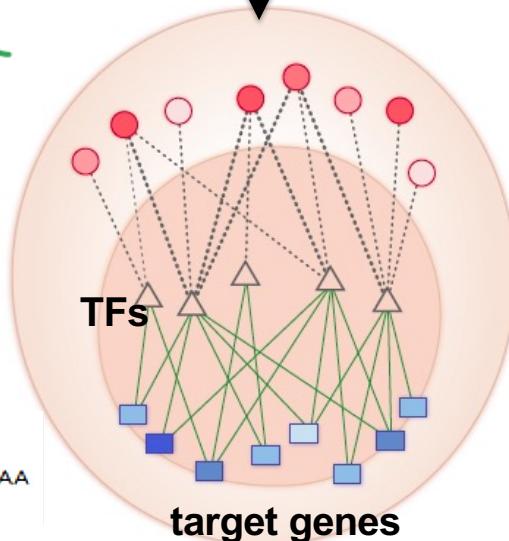
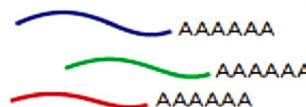
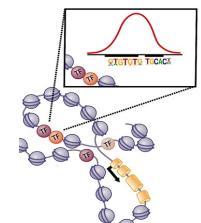
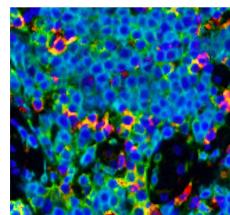
- Hands-on experience in applying tools and interpreting results using TF activity inference methods using public CITE-seq with Xiaojun Ma
- SPaRTAN
- https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/tree/main/hands-on_tutorial/session-2

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

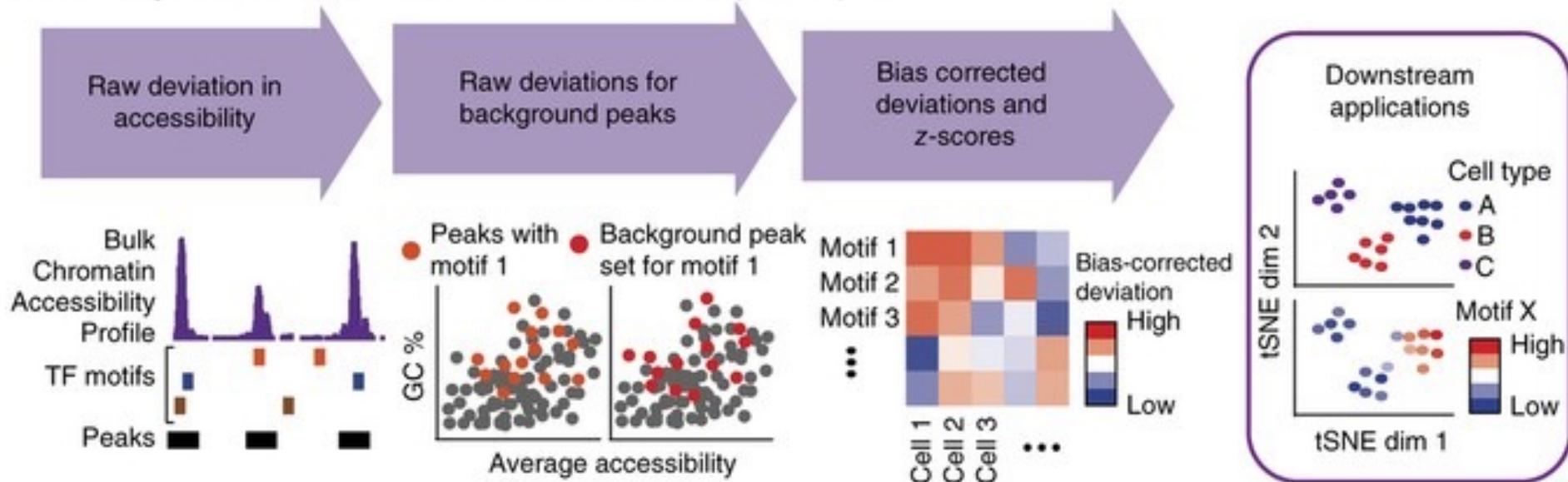
TF activity inference from single-cell gene expression data

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

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

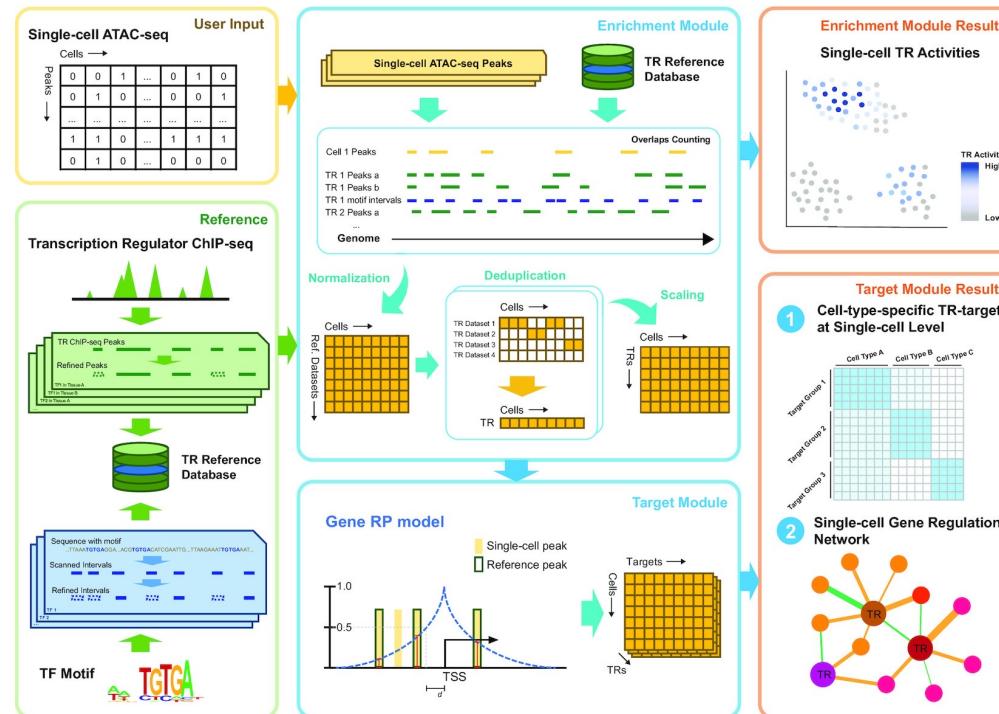
- ChromVar (Schep et al, Nature Methods 14, 975 (2017)) to find (a few hundred) TF motifs enriched in the peak count vector of each cell and visualize the results with tSNE/UMAP
<https://bioconductor.org/packages/release/bioc/vignettes/chromVAR/inst/doc/Introduction.htm>

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



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

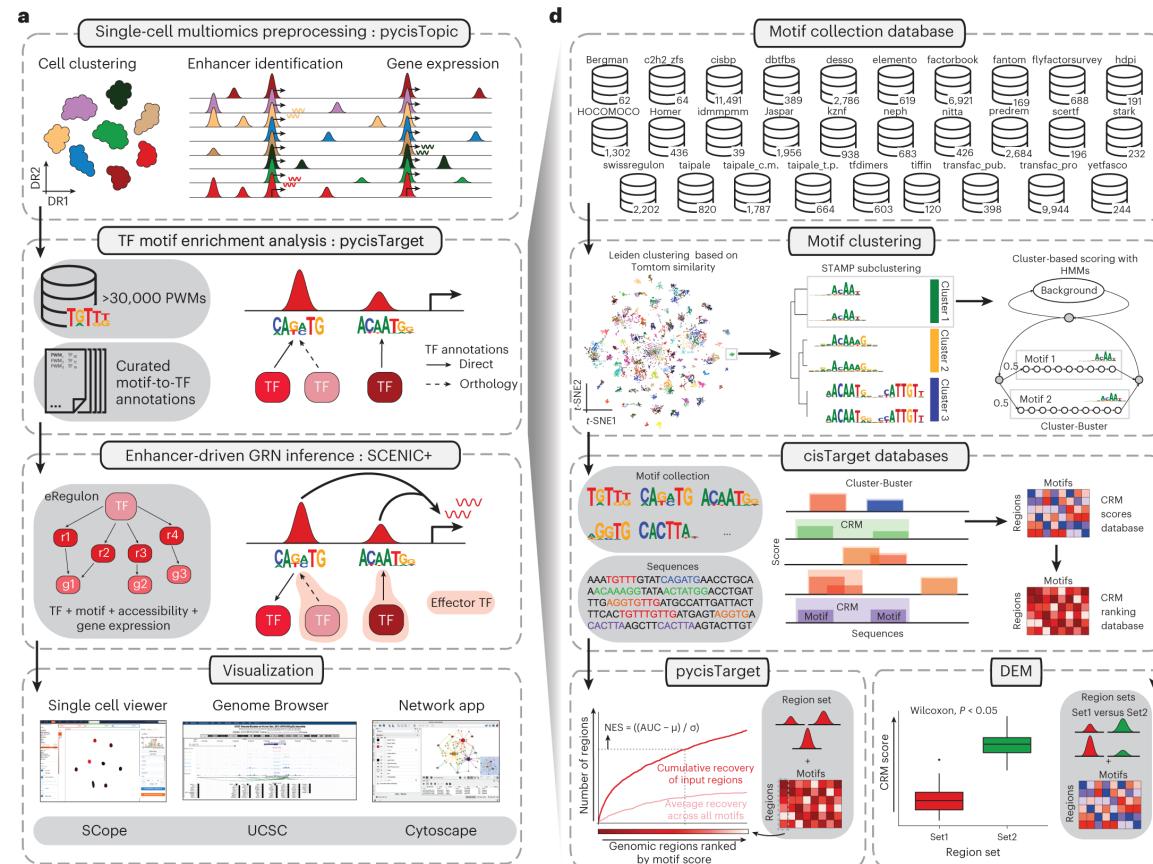
- SCRIP (Single-Cell gene Regulation network Inference using ChIP-seq and motif) (Dong et al., NAR, 2024)
- <https://github.com/wanglabtongji/SCRIP>



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/>



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				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				

Hands-on session 3

- Hands-on experience in applying tools and interpreting results using multiple TF activity inference methods using public multiome (Merve Sahin)
- SCENIC+
- https://github.com/osmanbeyoglulab/Tutorials-on-ISMB-2024/tree/main/hands-on_tutorial/session-3

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
Elena Kim
Parham Hadikhani
Haoyu Wang

<https://www.osmanbeyoglulab.com>

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



National Institutes
of Health

NIGMS R35 GM146989

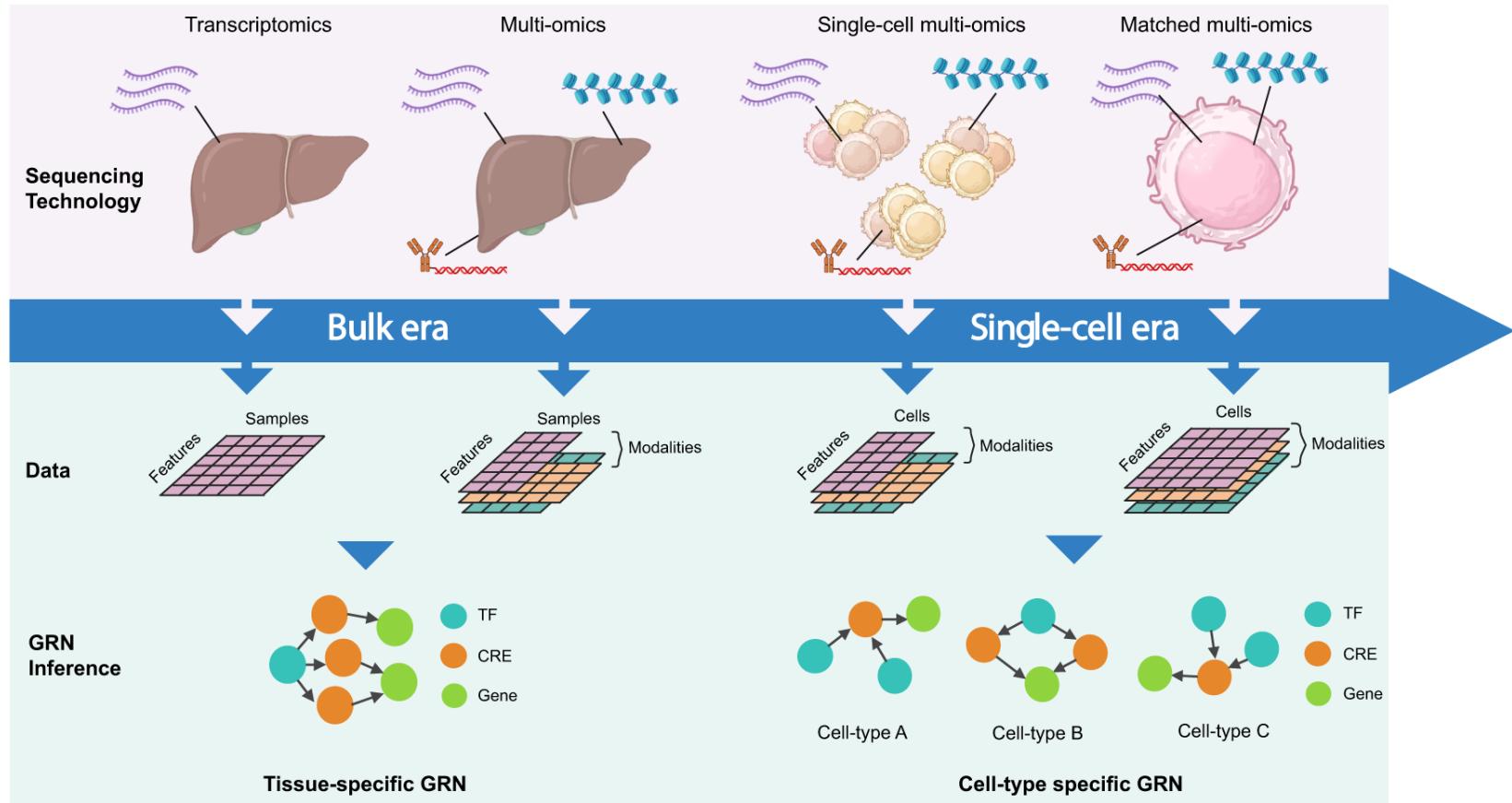
HILLMAN FELLOWS

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

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



- 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