

Considerations for analysing single cell ATAC-seq data

ATAC-seq course 2024

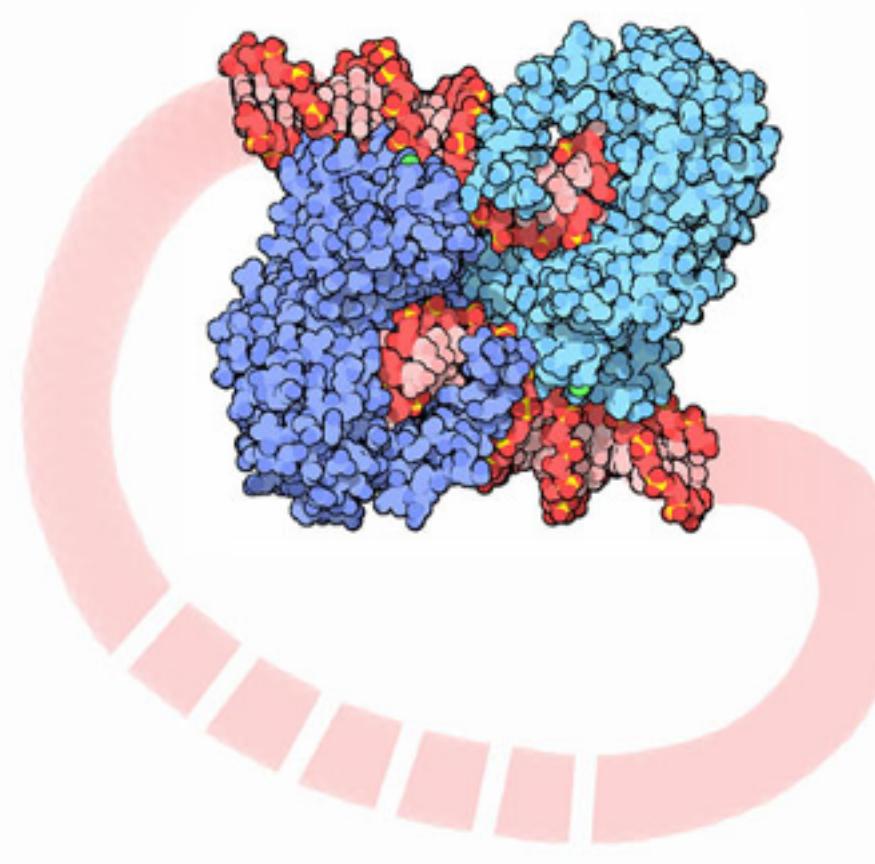
Judith Zaugg, April 2024

Outline

- **Interpretation of ATAC-seq data**
- Opportunities and challenges of single cell ATAC-seq
- Data integration
 - What data do you have?
 - What do you want to learn from the integration?
- Gene regulatory networks - example from our work

What is measured by ATAC-seq?

Tn5 homodimer (blue)
complexed with DNA



The DNA gets cut and pasted
DNA ligase fixes the backbones

Tn5 homodimer (blue) with adapter
loaded as paste-ready DNA

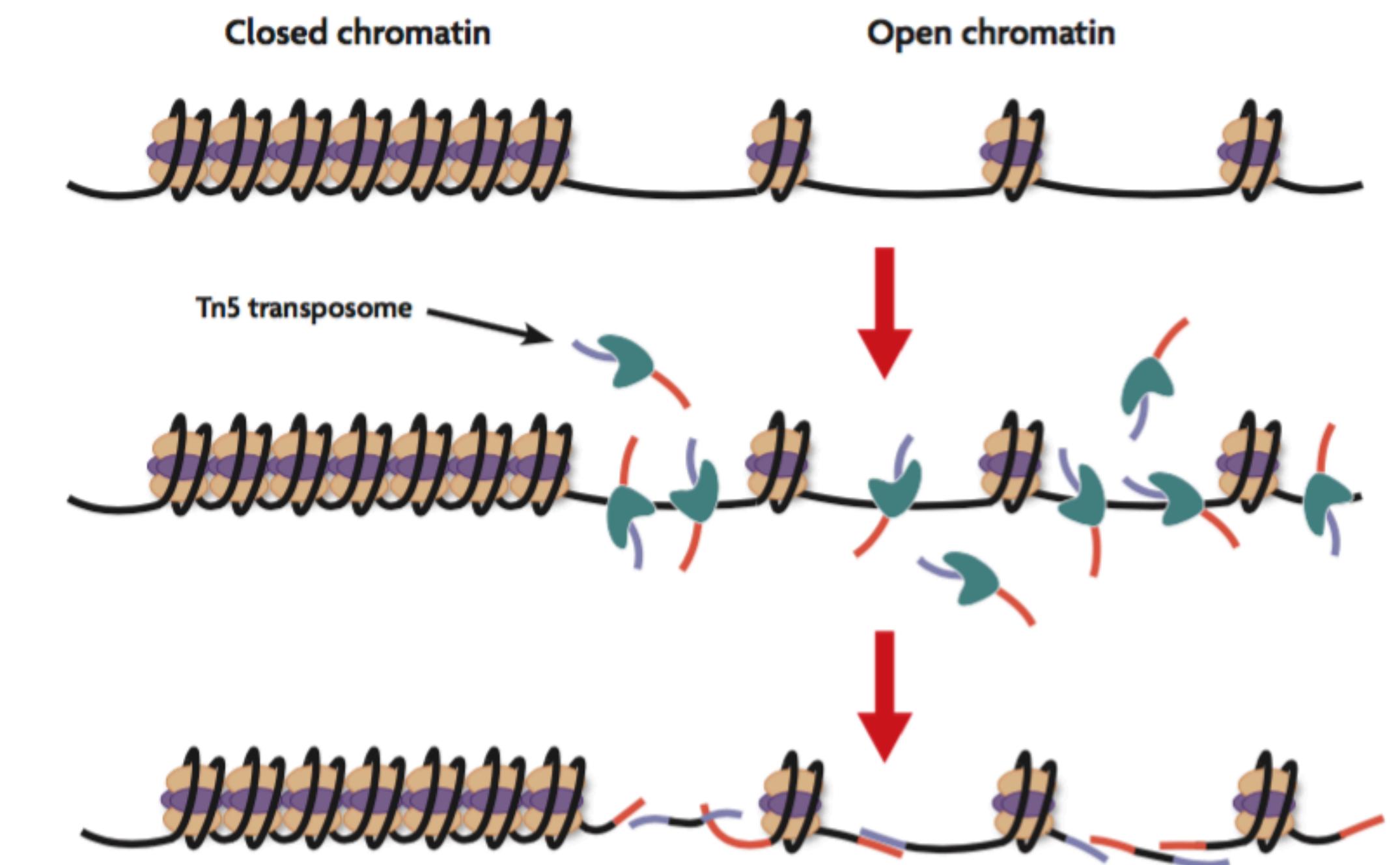
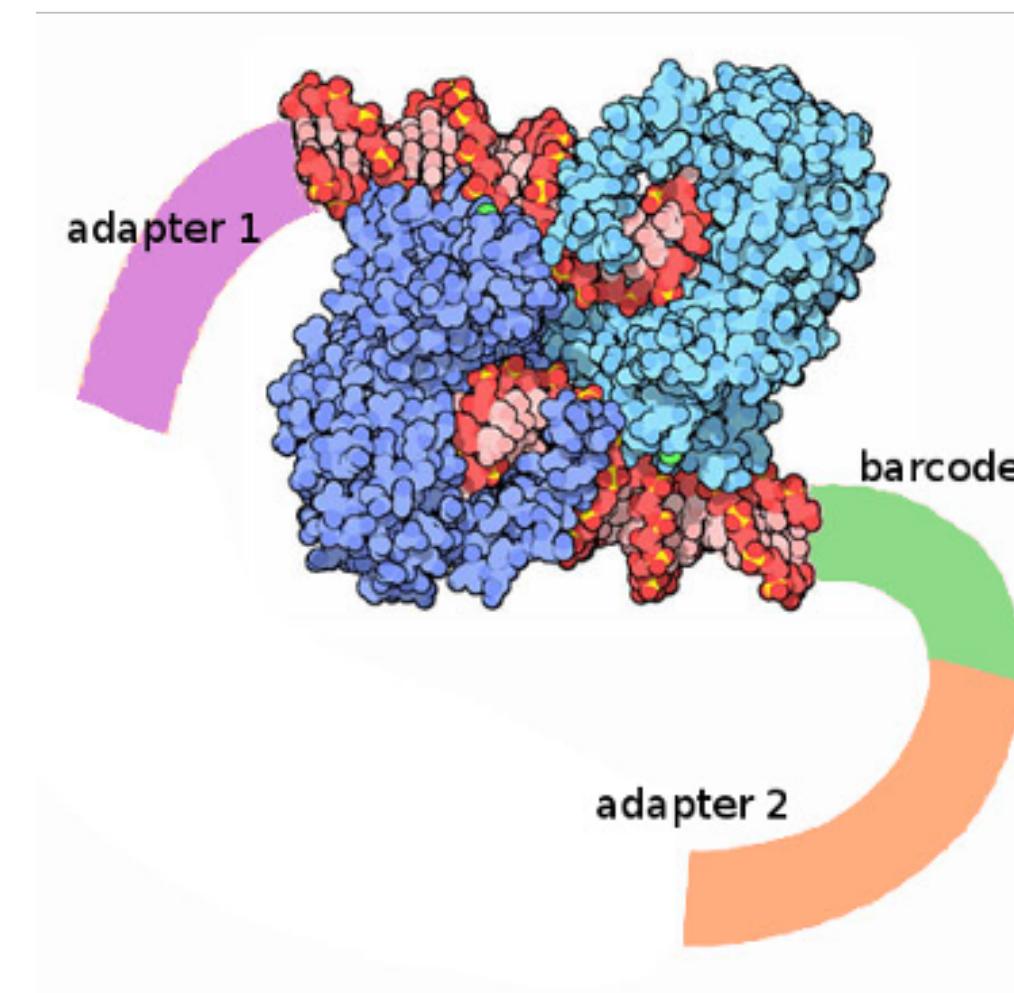
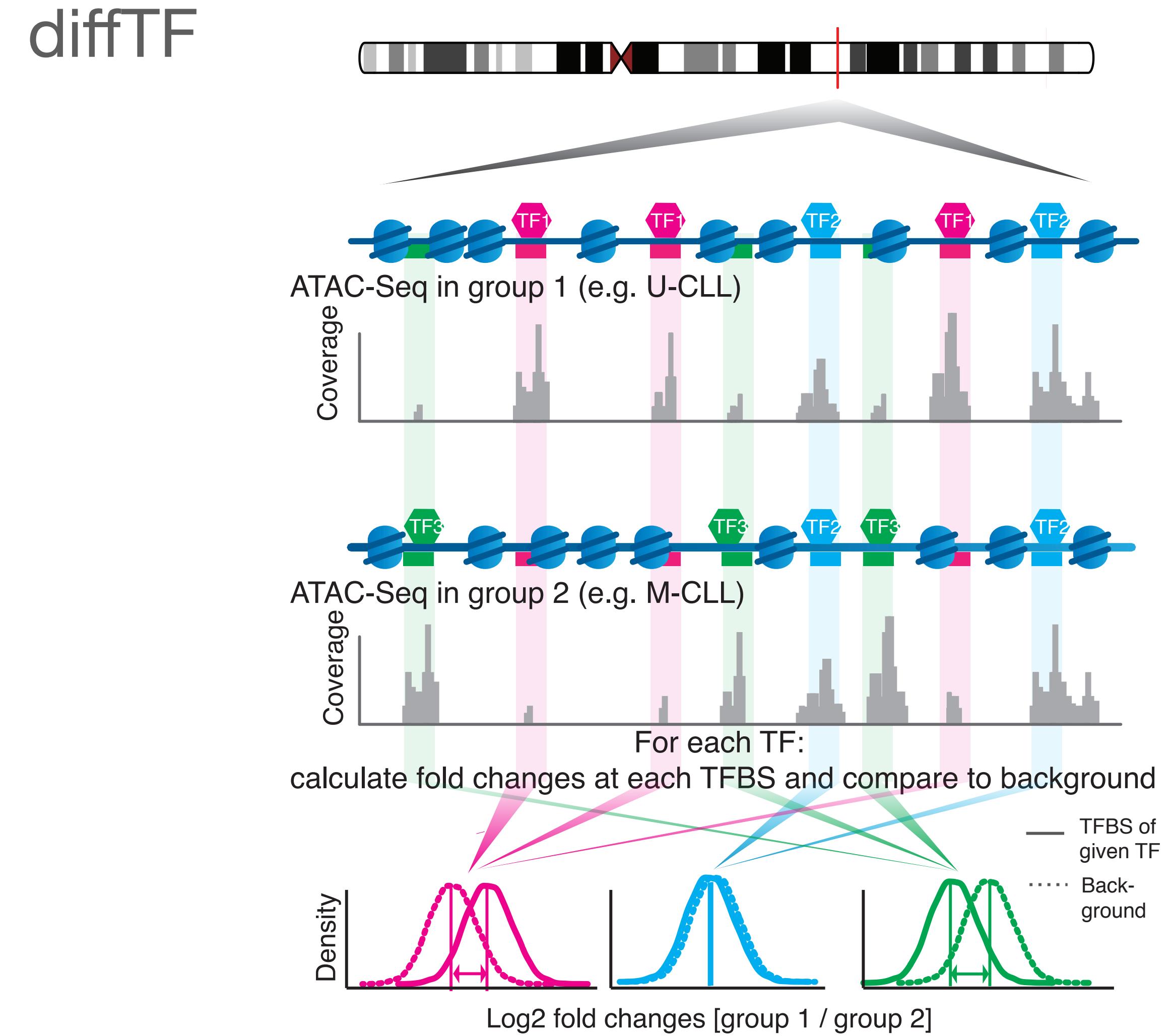
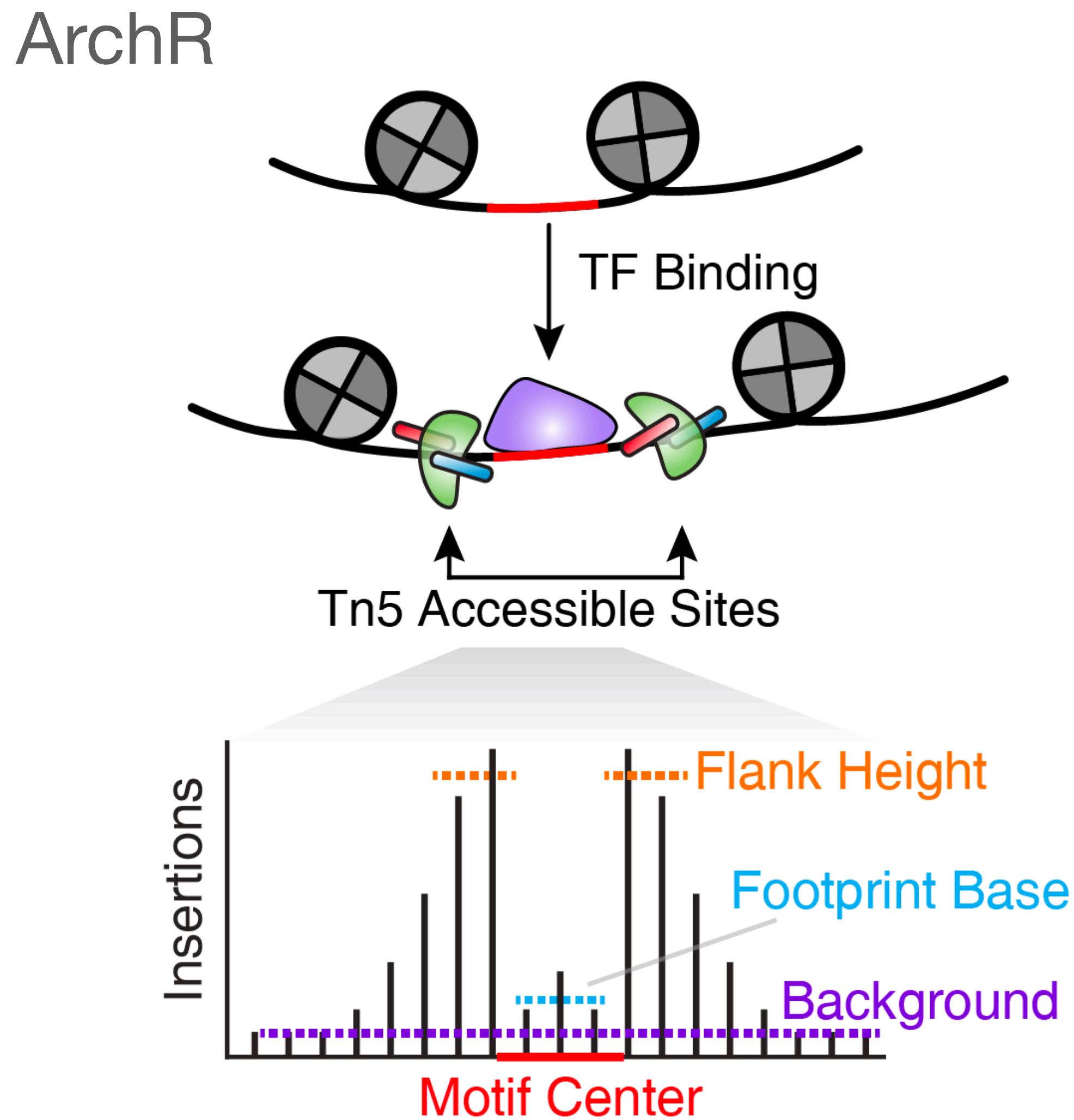


Figure from: <https://www.activemotif.com/blog-atac-seq>

Buenrostro et al., Nature Methods 2013

ATAC-seq can measure transcription factor binding

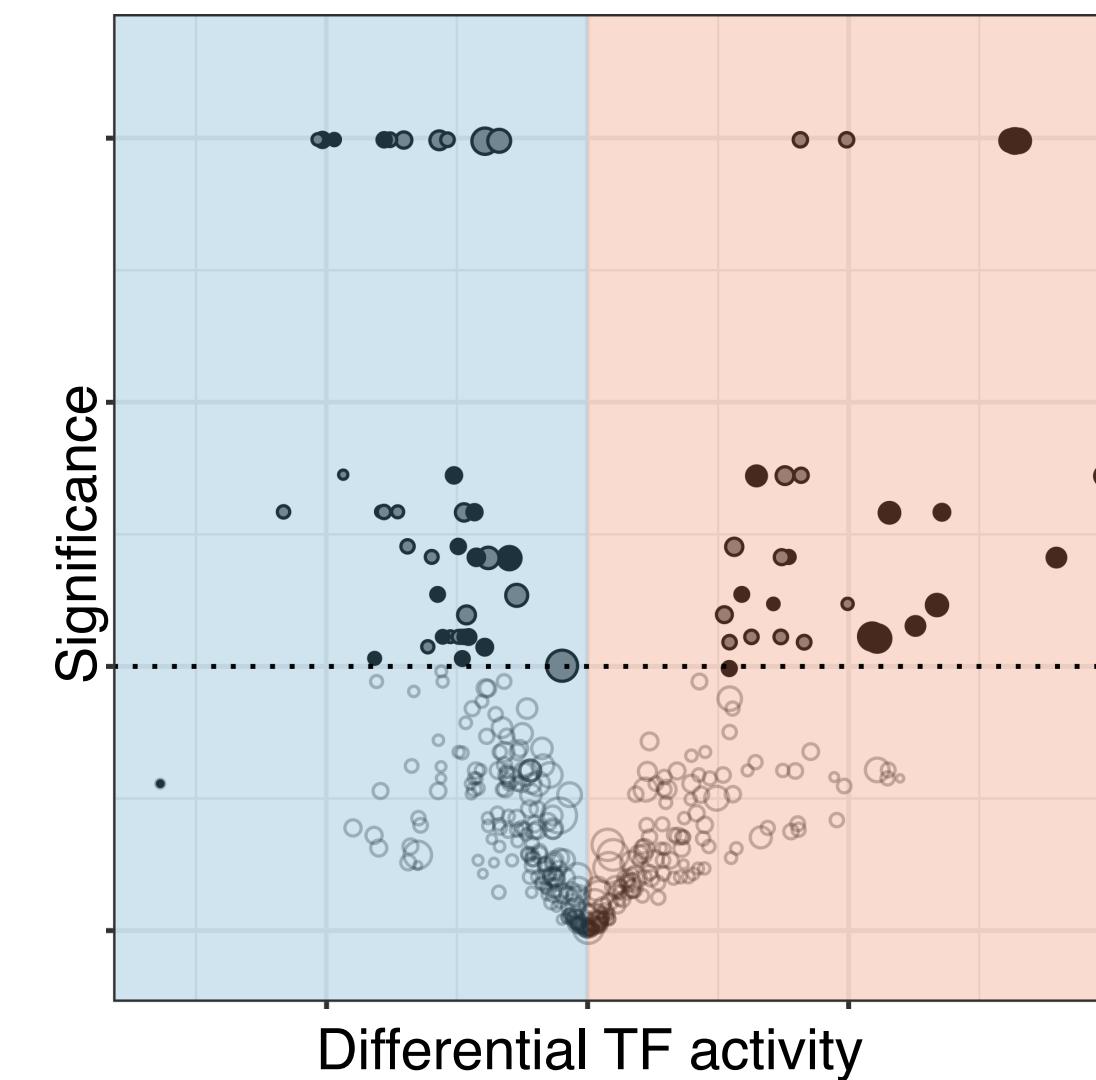
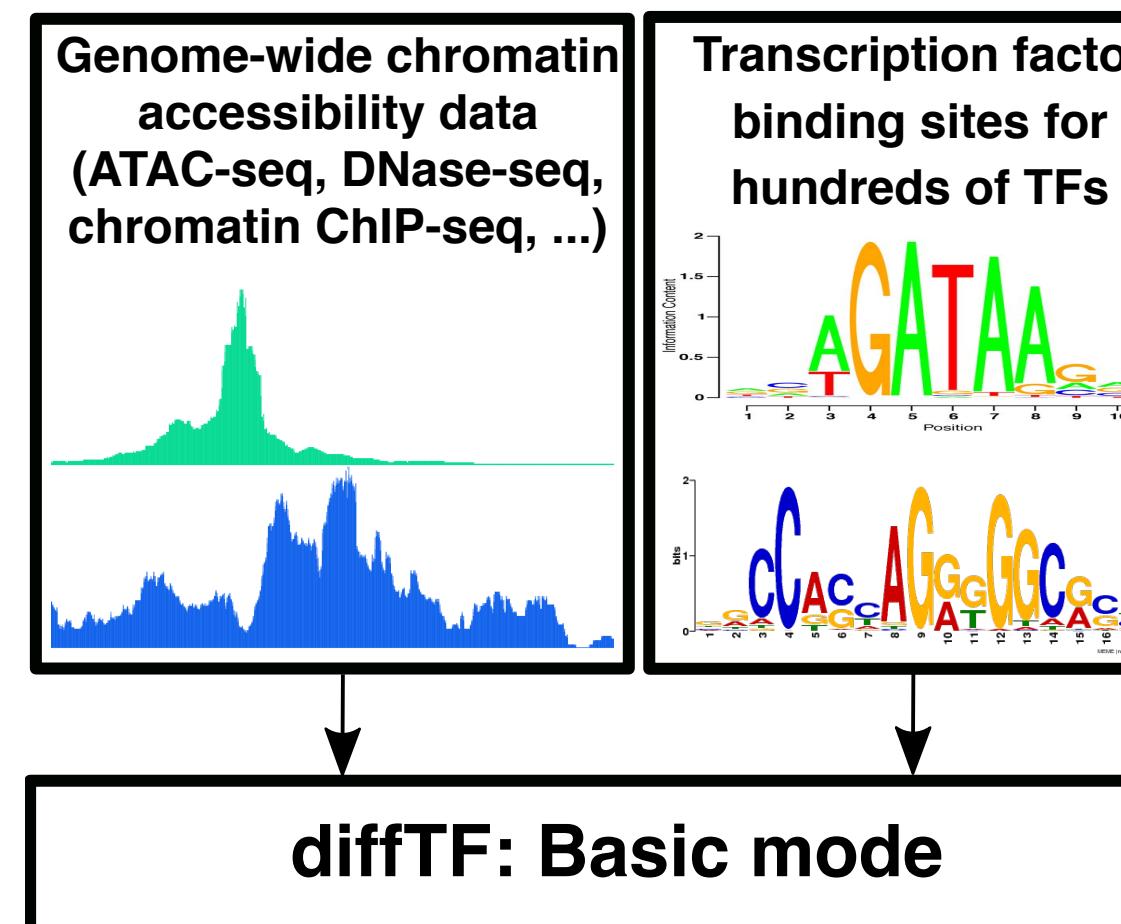




Ivan Berest

Christian Arnold

diffTF - tool to quantify differential TF activity



Ranking reprogramming factors for cell differentiation

Jennifer Hammelman^{1,2}, Tulsi Patel^{3,4,5}, Michael Closser^{3,4,5}, Hynek Wichterle^{3,4,5} and David Gifford^{1,2,6,7}



"We provide evidence that AME and diffTF are optimal methods [...] for systematic prioritization of transcription factor candidates [in reprogramming]"

<https://git.embl.de/grp-zaugg/diffTF>

What is NOT measured by ATAC-seq?

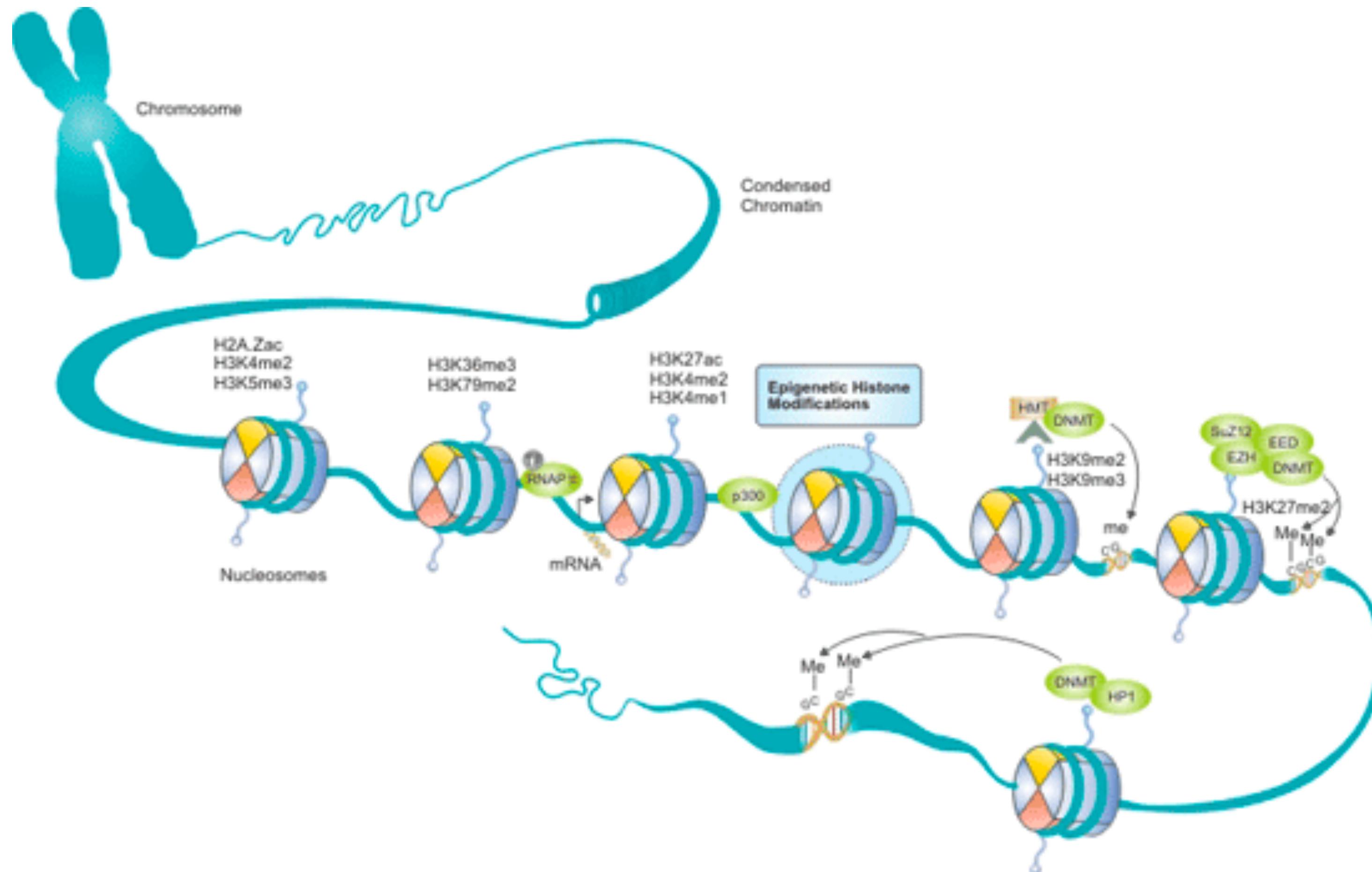


Figure from: <https://www.cusabio.com/c-20825.html>

Interpretation of ATAC-seq data

Discuss with
your neighbour

- What do you expect to see from a TF with long residency time? What from a TF with very dynamic binding and short residency time?
- What do you expect from an activating TF? What from a repressing TF?
- How would you expect an active/inactive/bivalent promoter to look like?
- Where do you expect to be able to see nucleosome patterns?

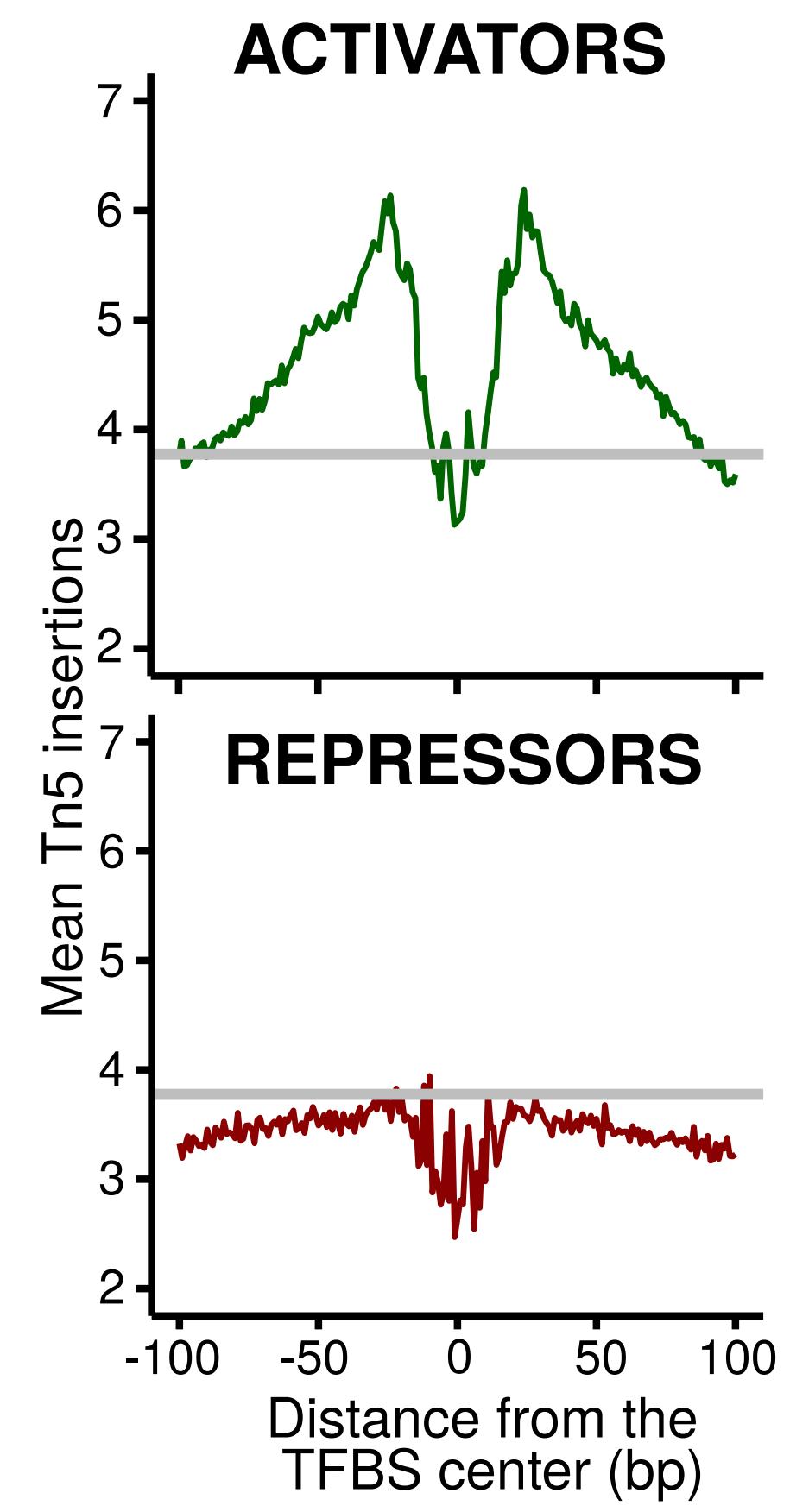
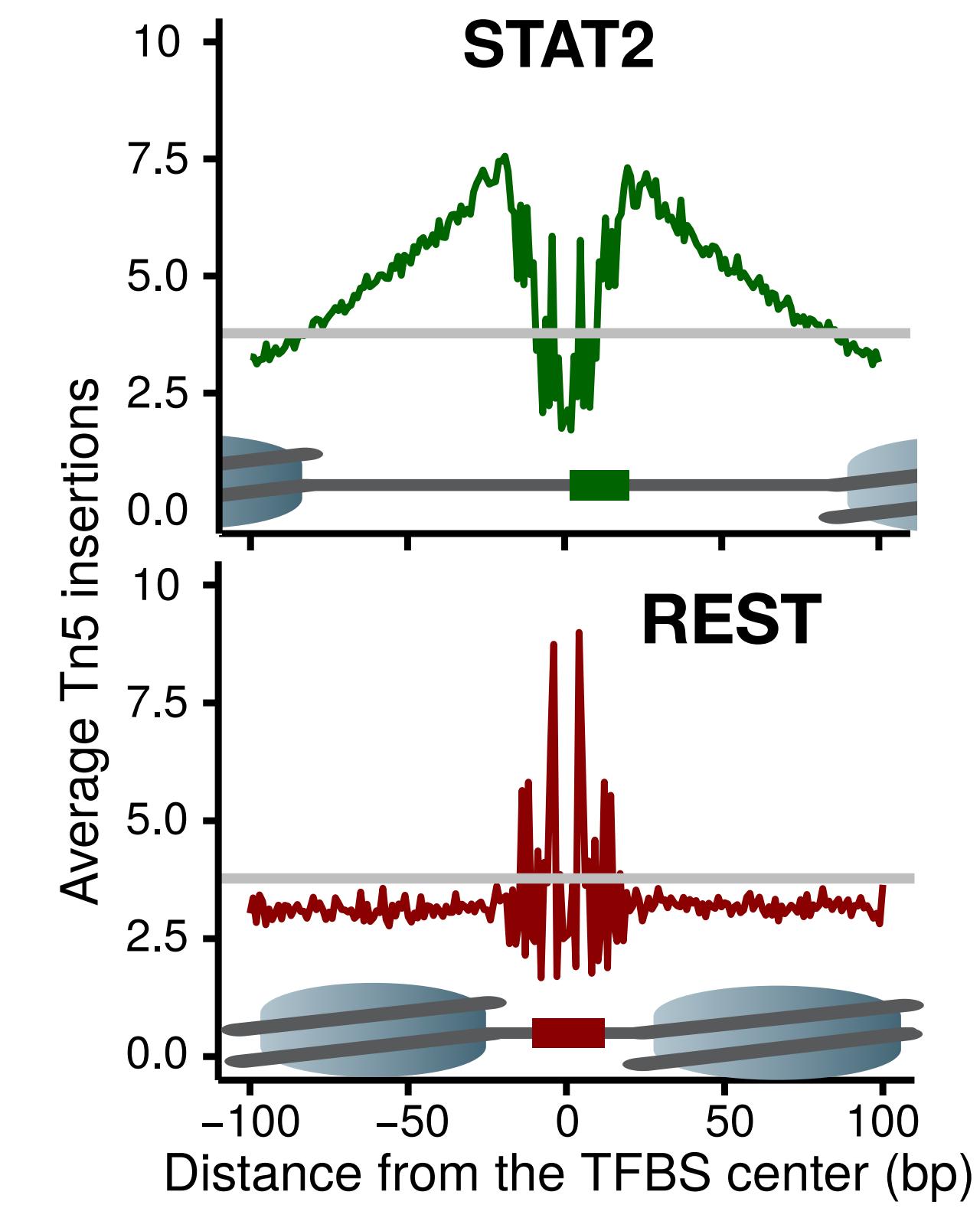
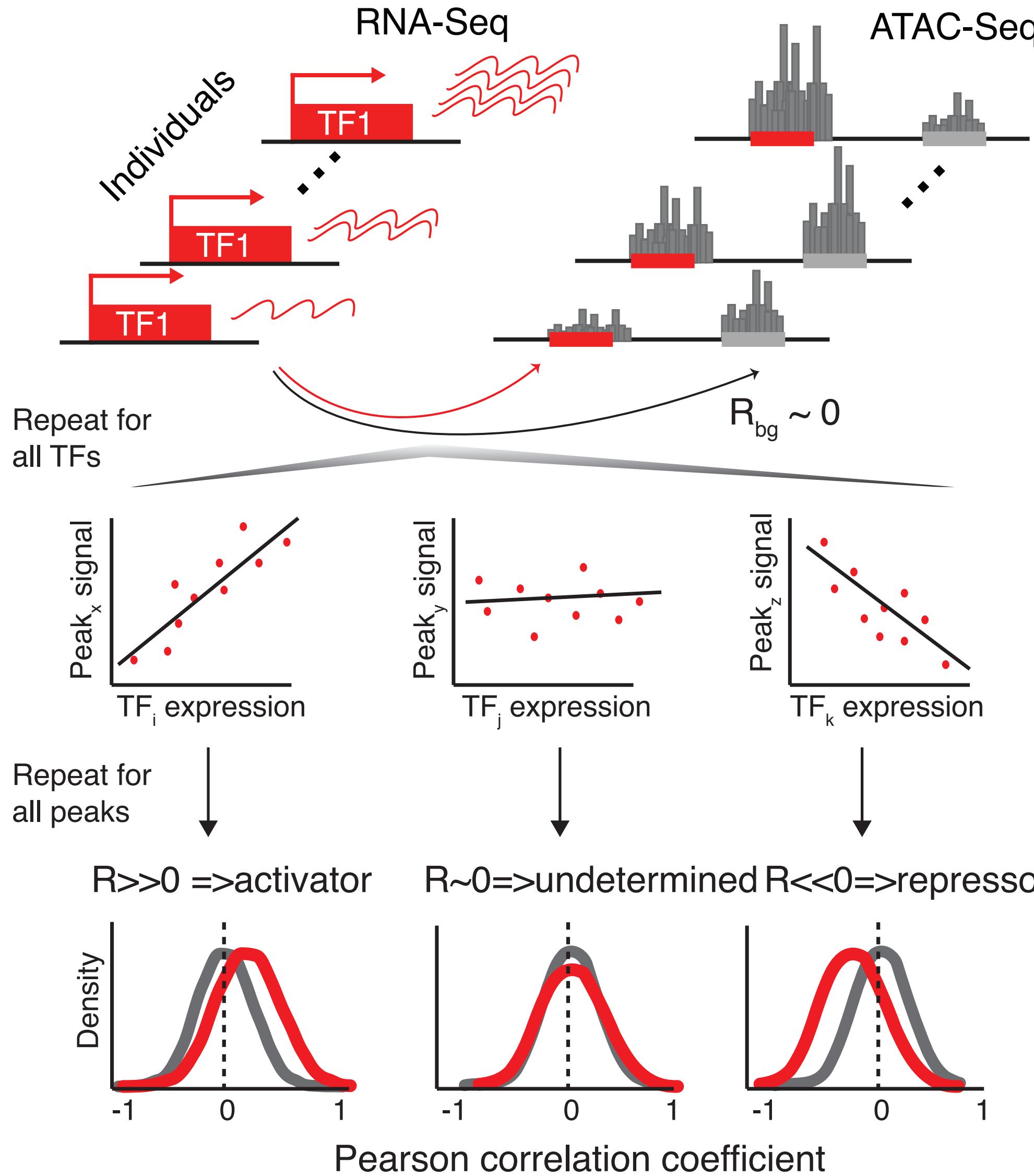


Ivan Berest



Christian Arnold

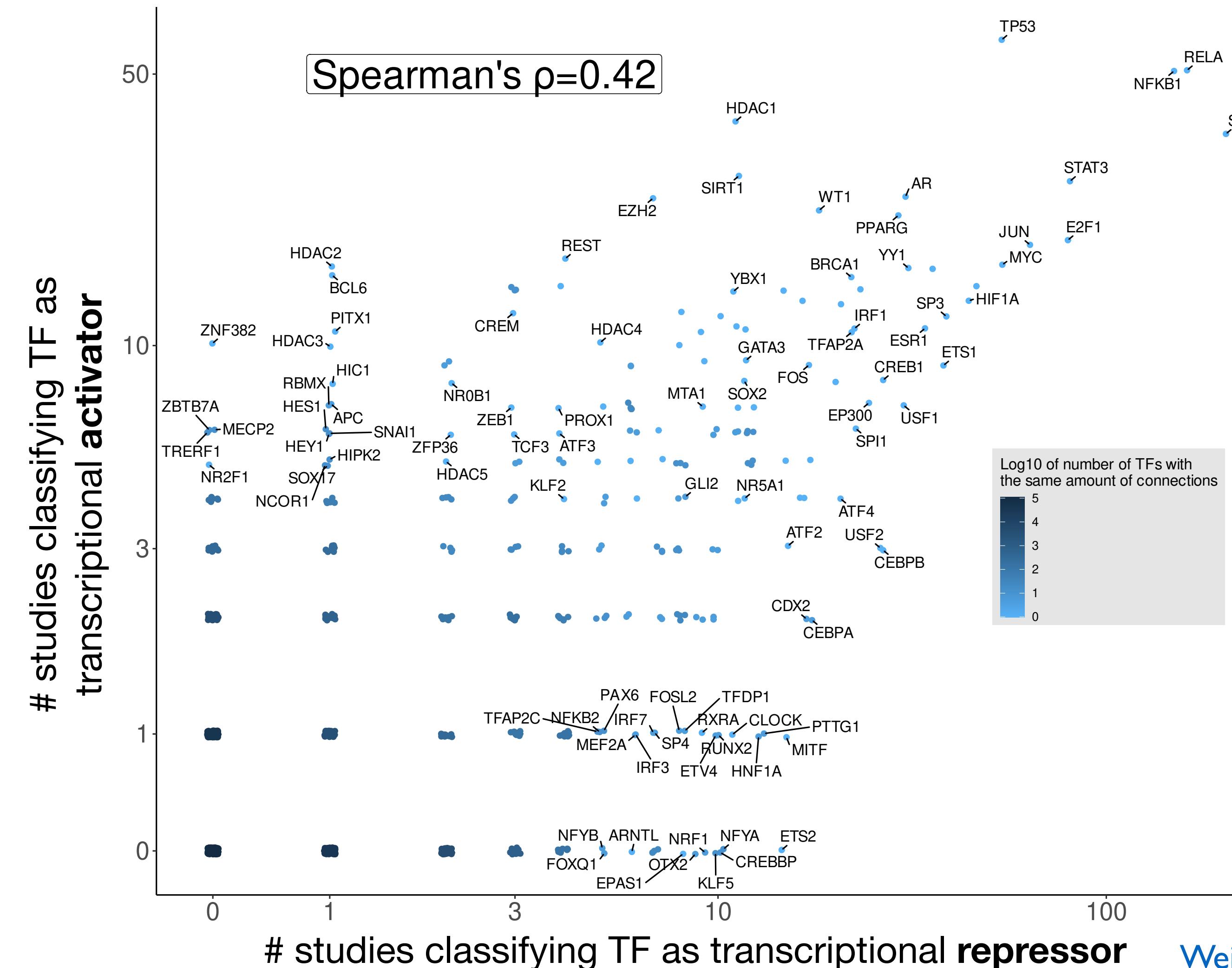
Activator vs repressor signatures in ATAC-seq



Many TFs are classified as activator and repressor equally often in literature



Maksim Kholmatov



Combining RNA and ATAC-seq offers data-driven classification of TFs (e.g. with diffTF)

Outline

- Interpretation of ATAC-seq data
- **Opportunities and challenges of single cell ATAC-seq**
- Data integration
 - What data do you have?
 - What do you want to learn from the integration?
- Gene regulatory networks - example from our work

What to expect from single cell ATAC-seq?

Opportunities and Challenges

What to expect from single cell ATAC-seq?

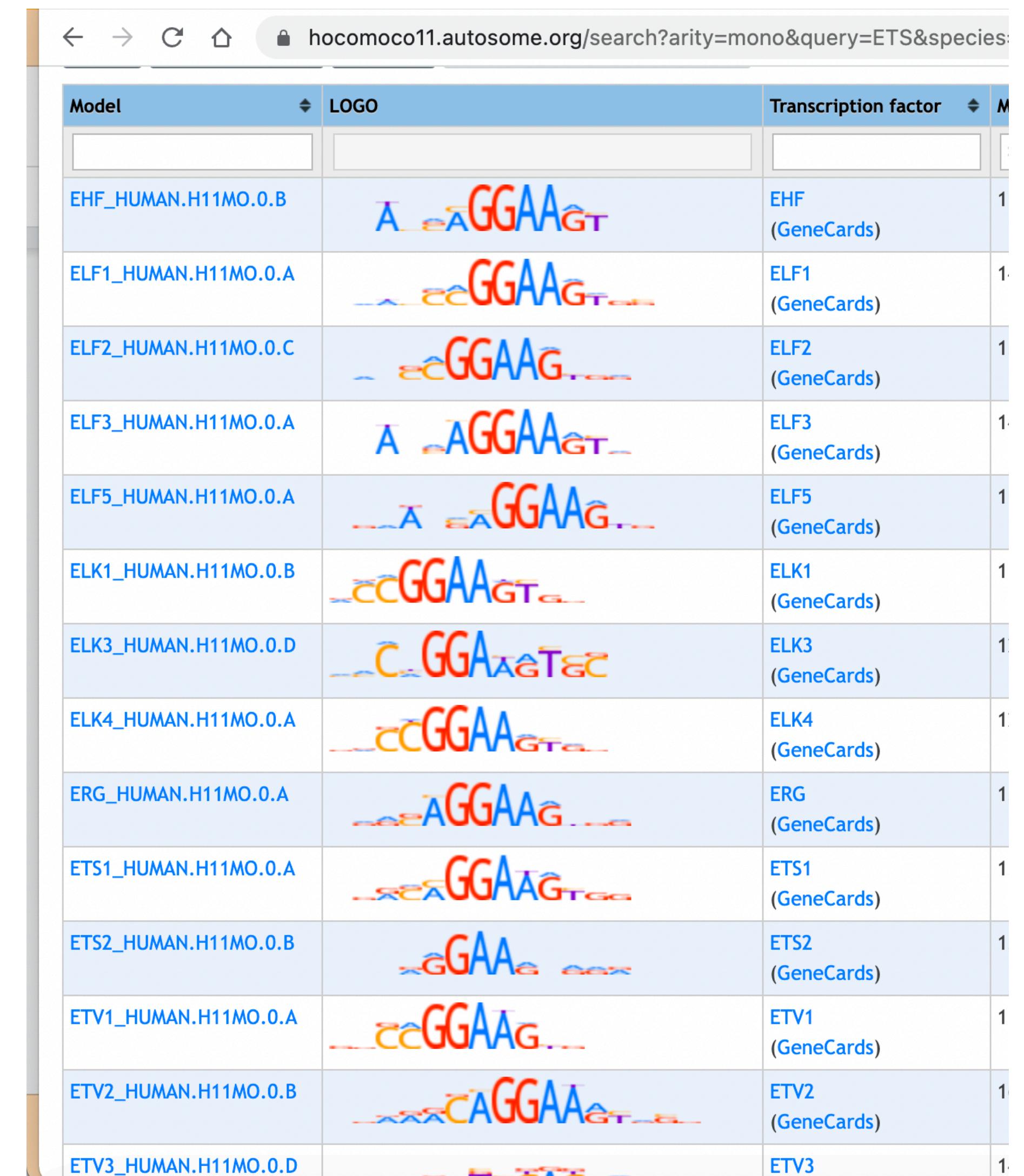
Opportunities

- Cell-type specific accessible chromatin regions => interpretation of disease-associated genetic variants
- Cell-type specific TF activity
- Cell-type specific enhancers
- Potential to generate cell-type specific gene regulatory networks (when integrated with RNA)

What to expect from single cell ATAC-seq?

Technical challenges

- Very low number of fragments per cell (even more sparse than scRNA-seq)
- Challenging to define cell types (current gold-standard: accessibility across marker genes)
- TF motif based analyses suffer from motif similarities



Standard processing of scATAC-seq data

- Read mapping
- peak calling/counts per peak
- (binarizing)
- clustering, cell type annotations (e.g. based on “marker gene” accessibility)
- Lots of downstream analyses!
- Extensive analysis workflow comprising many of the most recent tools **signac**:
<https://satijalab.org/signac/index.html>

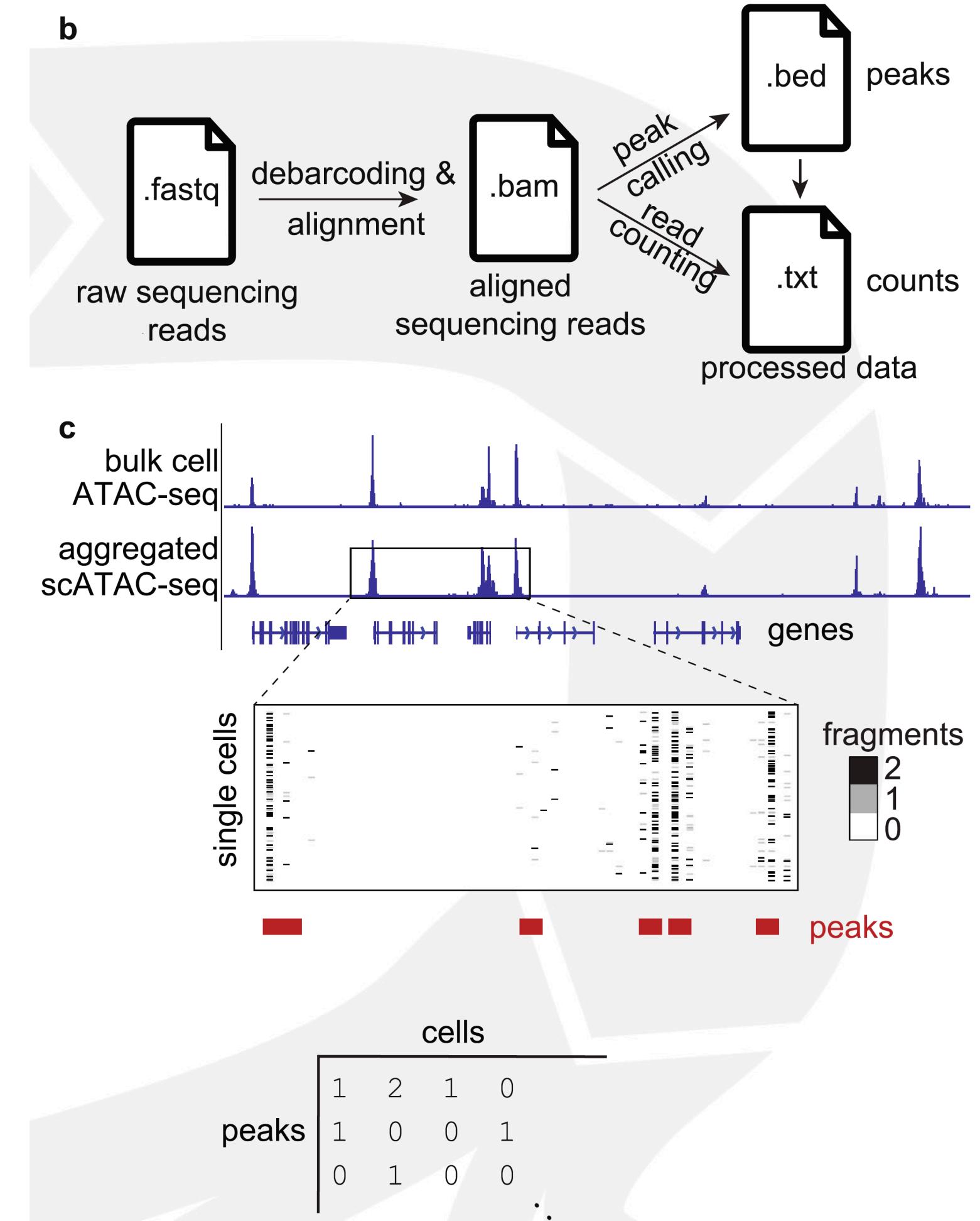
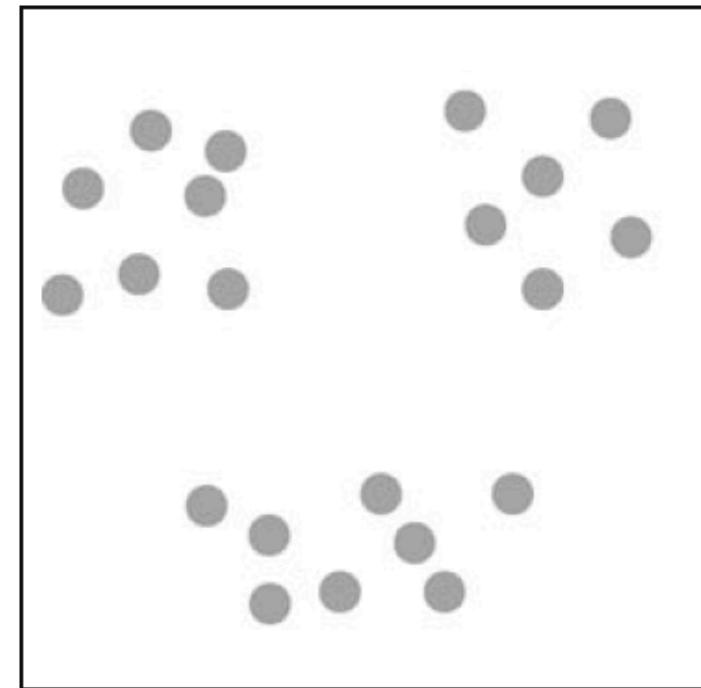


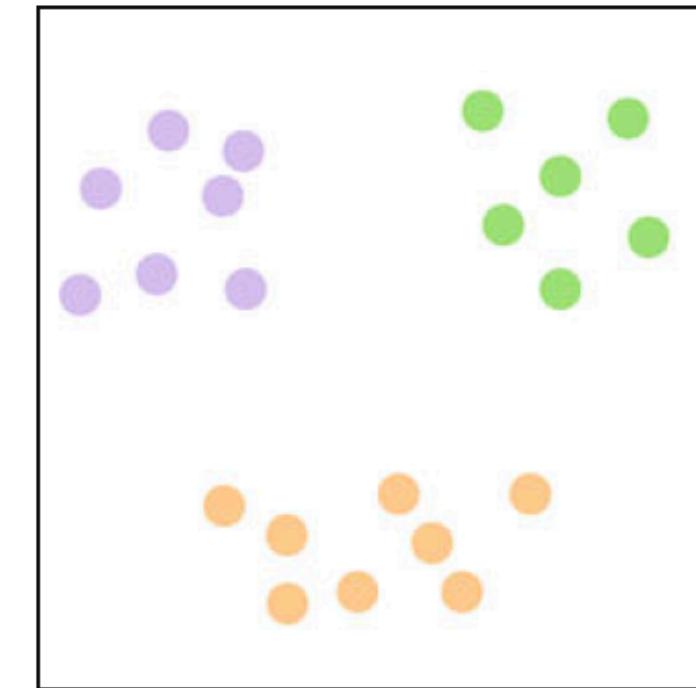
Figure from: Chen et al. *Genome Biology* (2019) 20:241

A selection of downstream analyses

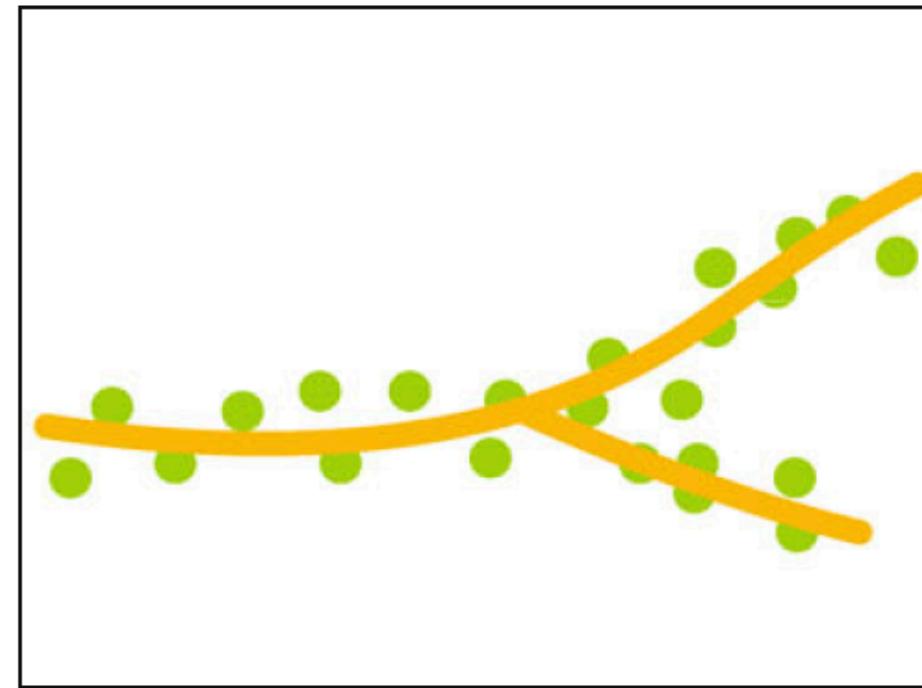
visualization



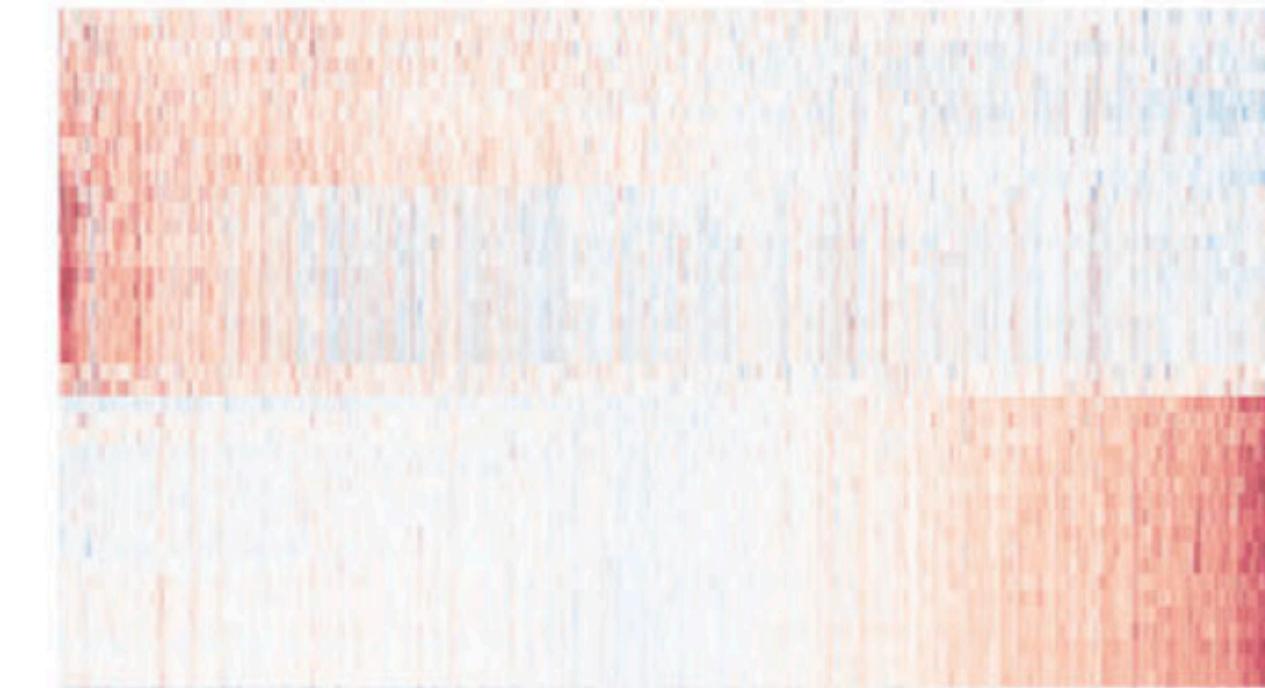
clustering



trajectory inference



differential accessibility



cis-regulatory network



Benchmarking scATAC-seq methods

- <https://github.com/pinellolab/scATAC-benchmarking/>

Figure from: Chen et al. *Genome Biology* (2019) 20:241

Challenges of scATAC-seq data interpretation

- Peak-by-peak (enhancers / promoters)
 - Promoter-peak: what does accessibility mean?
 - Enhancer-peak: what gene does the “enhancer” regulate?
- TF activities (aggregate signal across TF binding sites / kmers)
 - how do you distinguish TFs of the same TF family?
 - how can we infer TF activity and what does it mean?

Promoter /enhancer peaks (bulk)



Daria Bunina

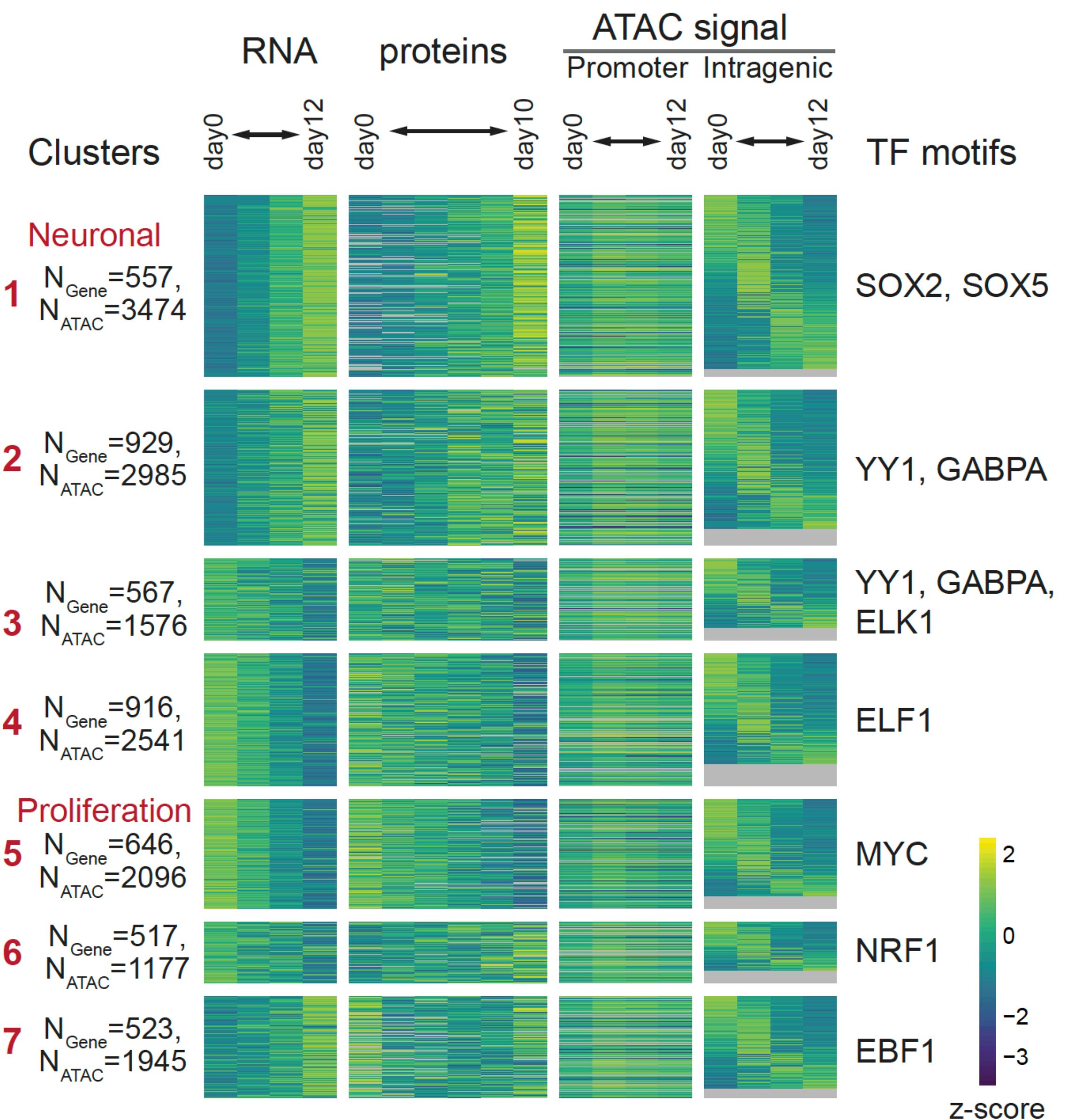
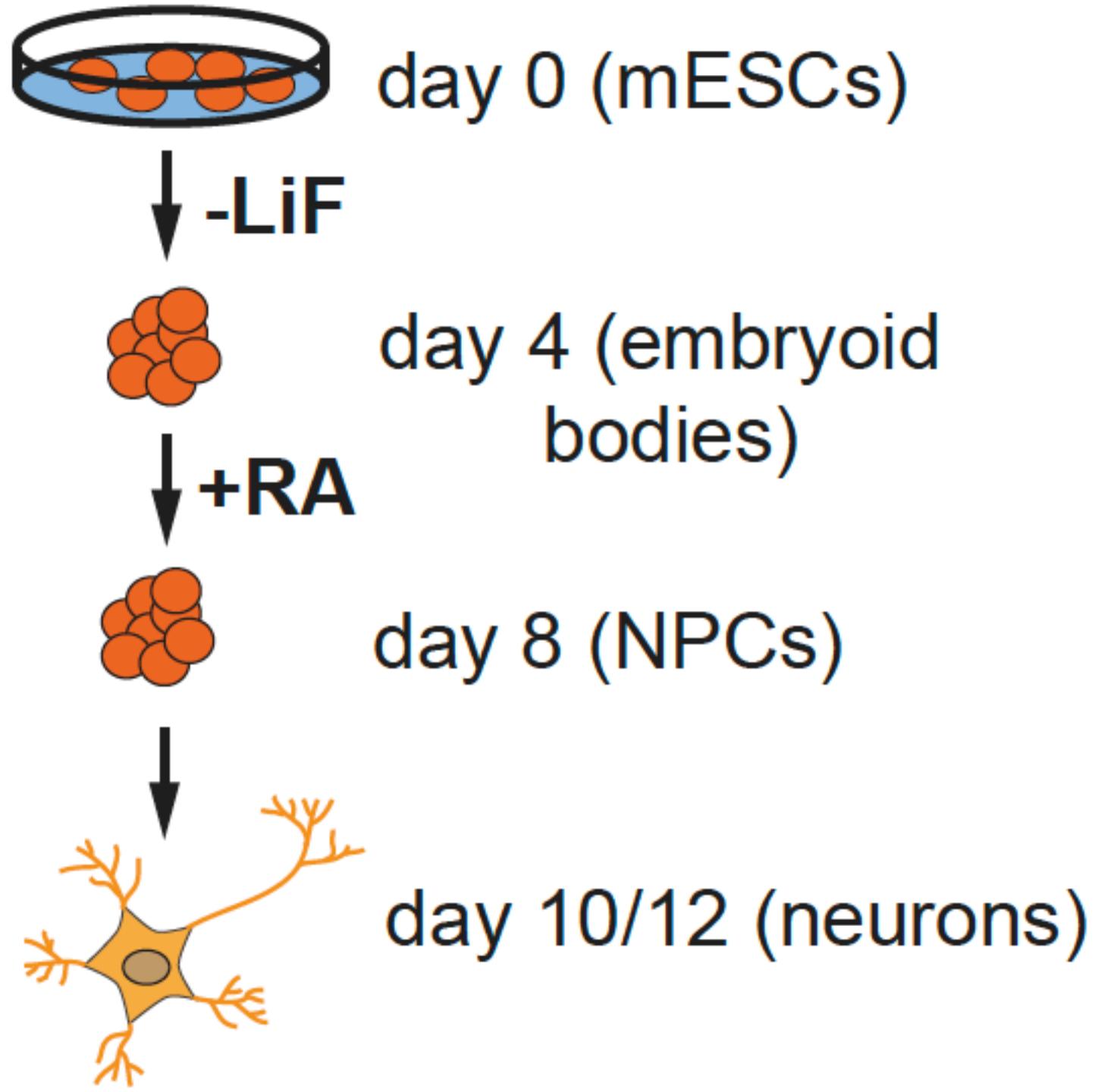
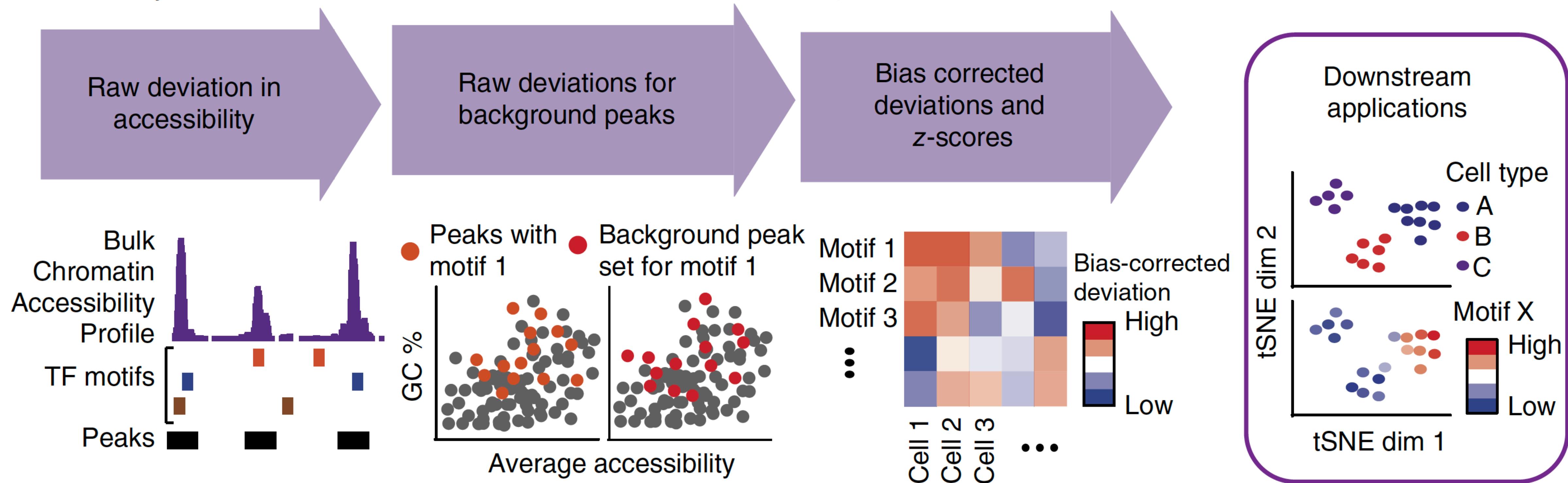


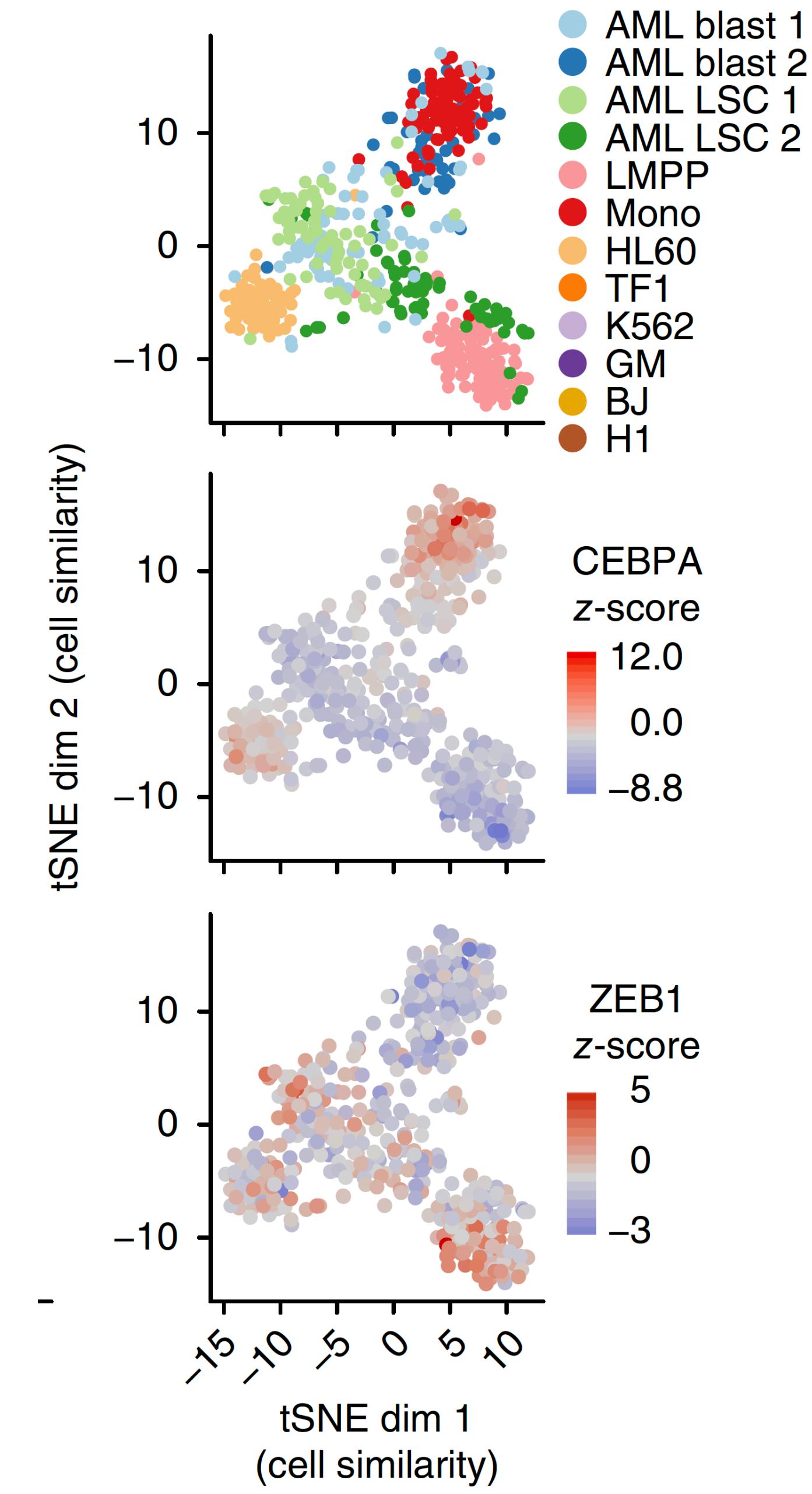
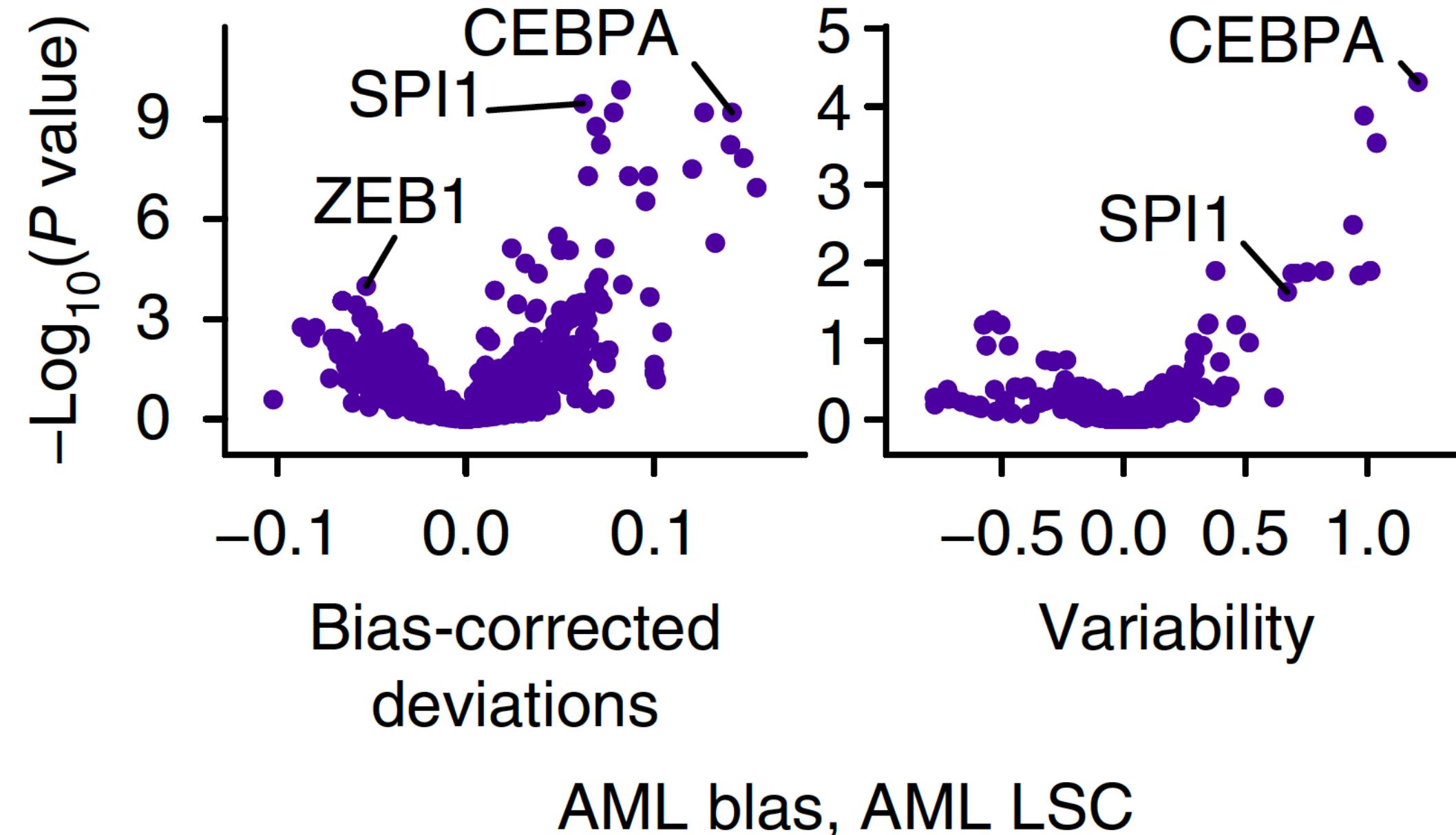
Figure from Bunina et al, Cell Systems 2020

Transcription factor readout from scATAC-seq (e.g. ChromVar)

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

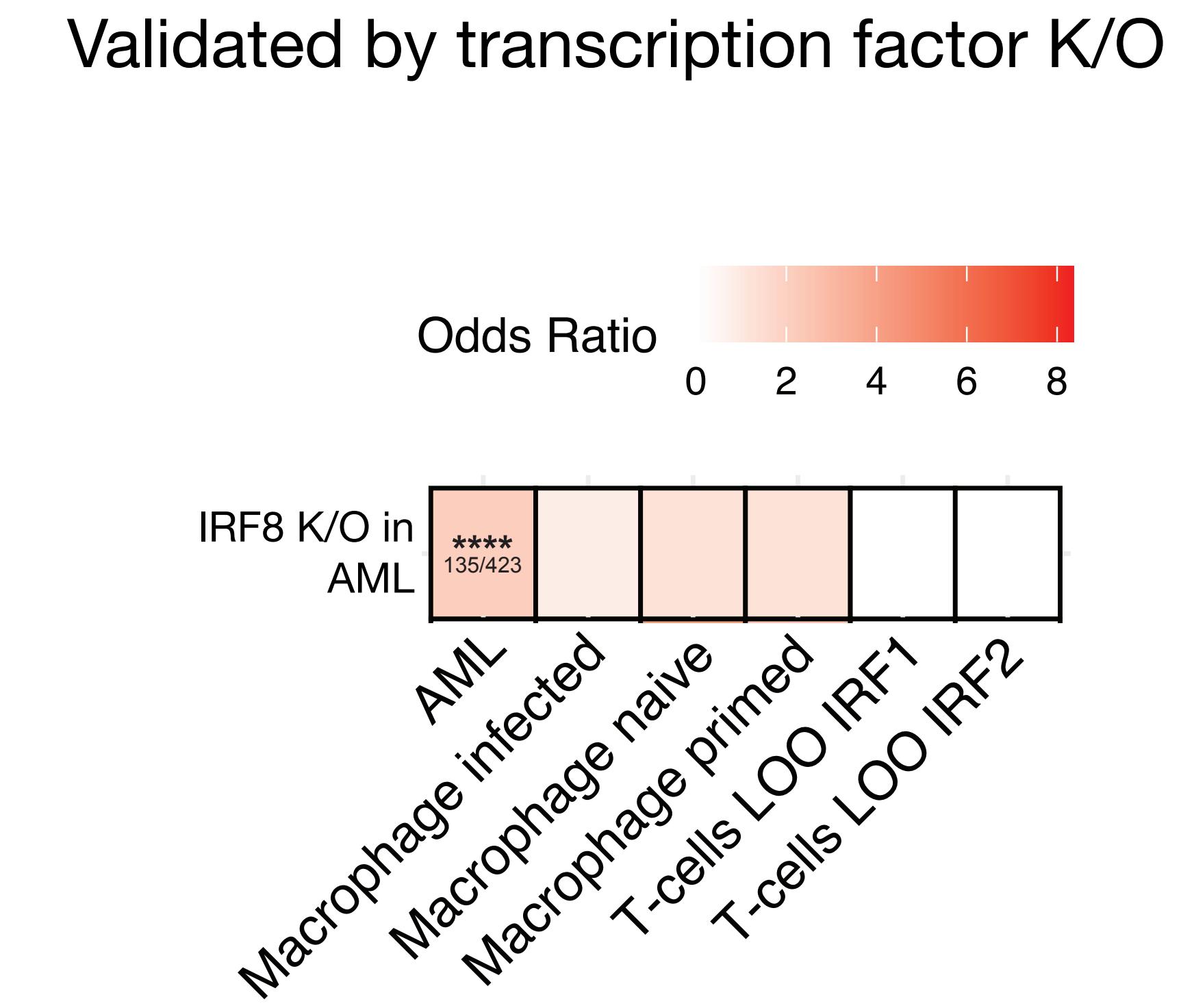
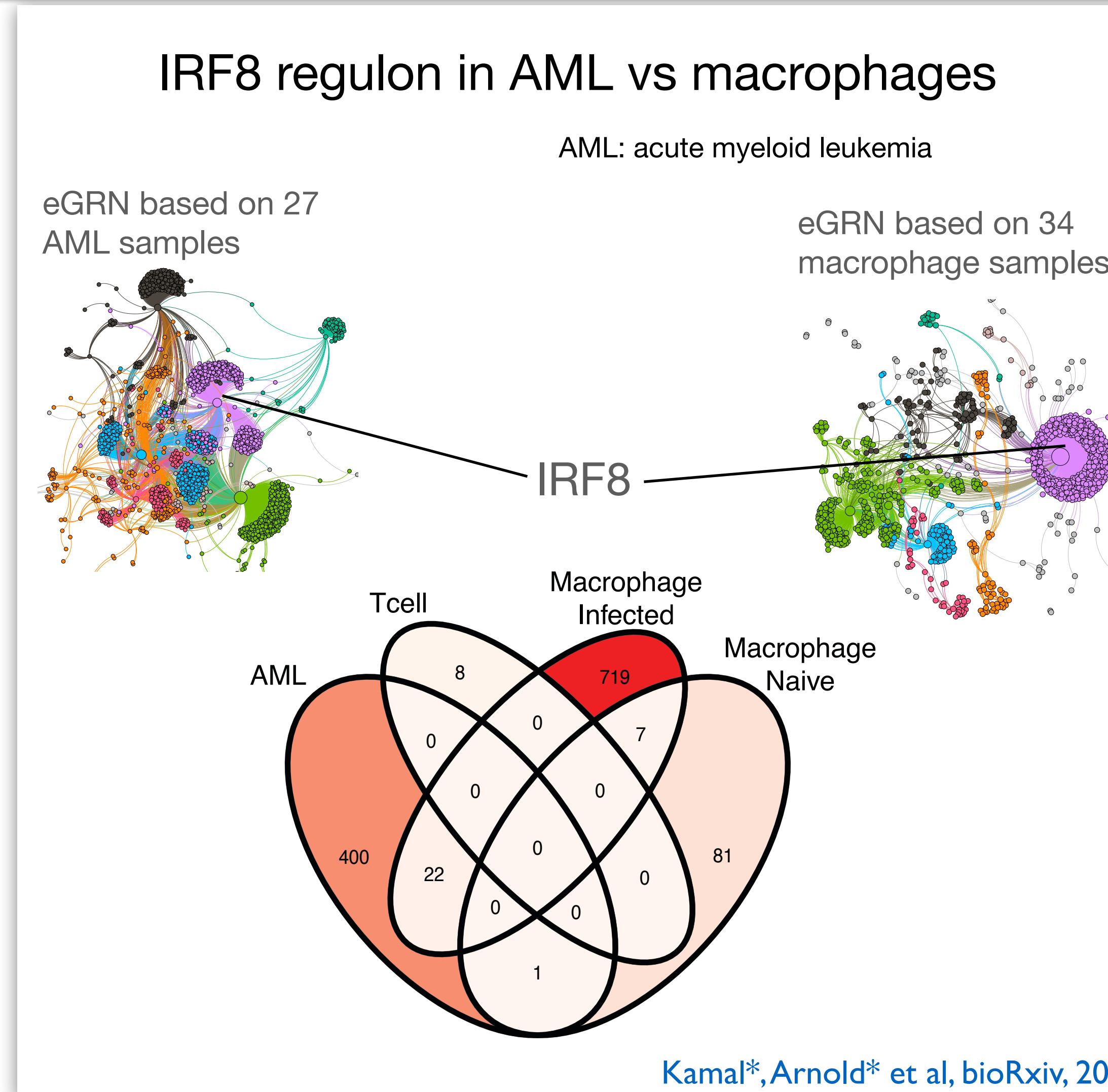


ChromVar - cell-type specific TF activities



but... activity of a TF can have a very different effect in different cell types

Transcription factor regulons are cell-type specific - e.g. IRF8



IRF8 K/O: Liss et al, Cancers (Basel). 2021 Feb 12; 13(4)

Need for cell-type specific regulatory networks to study TF function

Interpretation of single cell ATAC-seq data

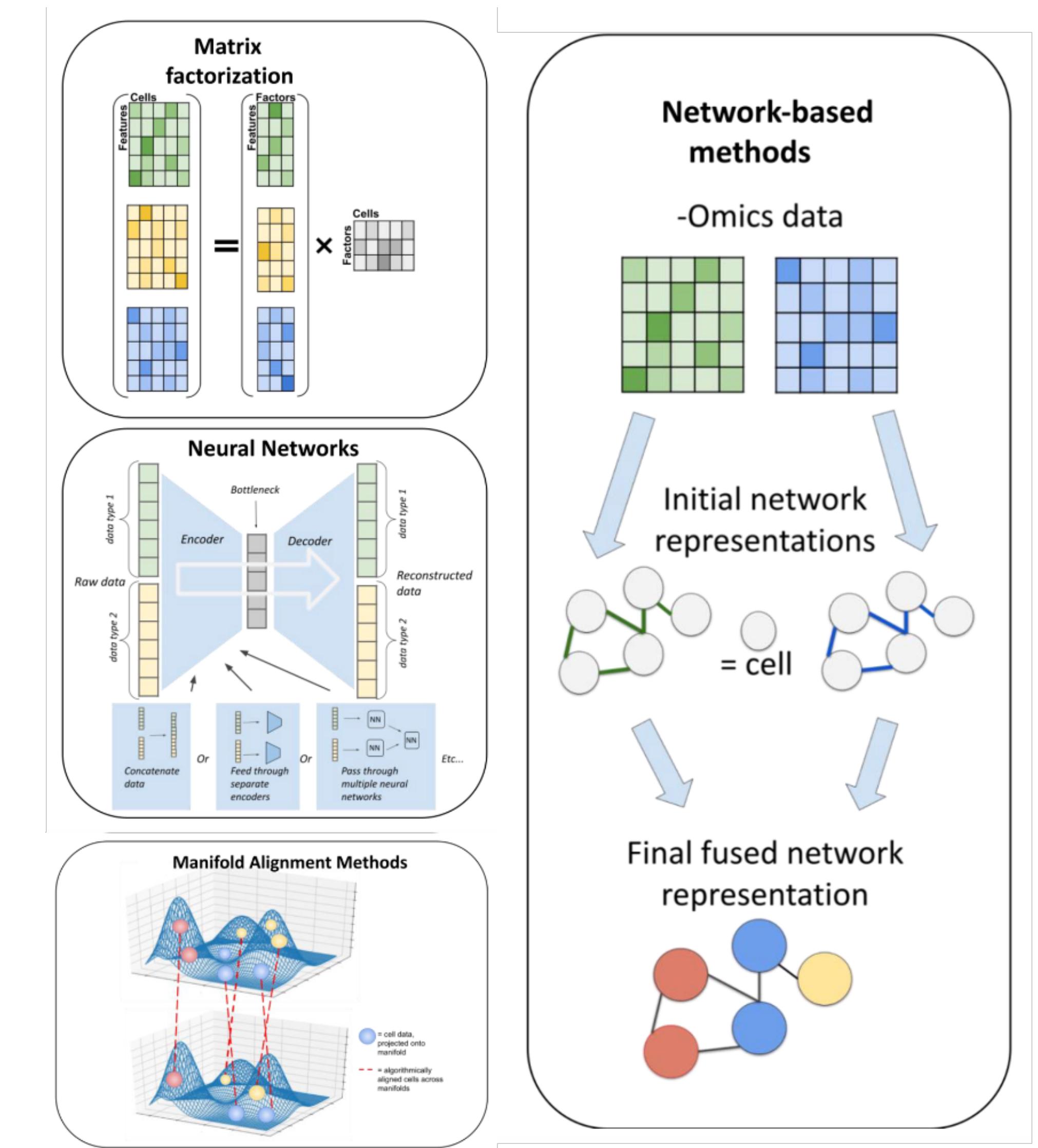
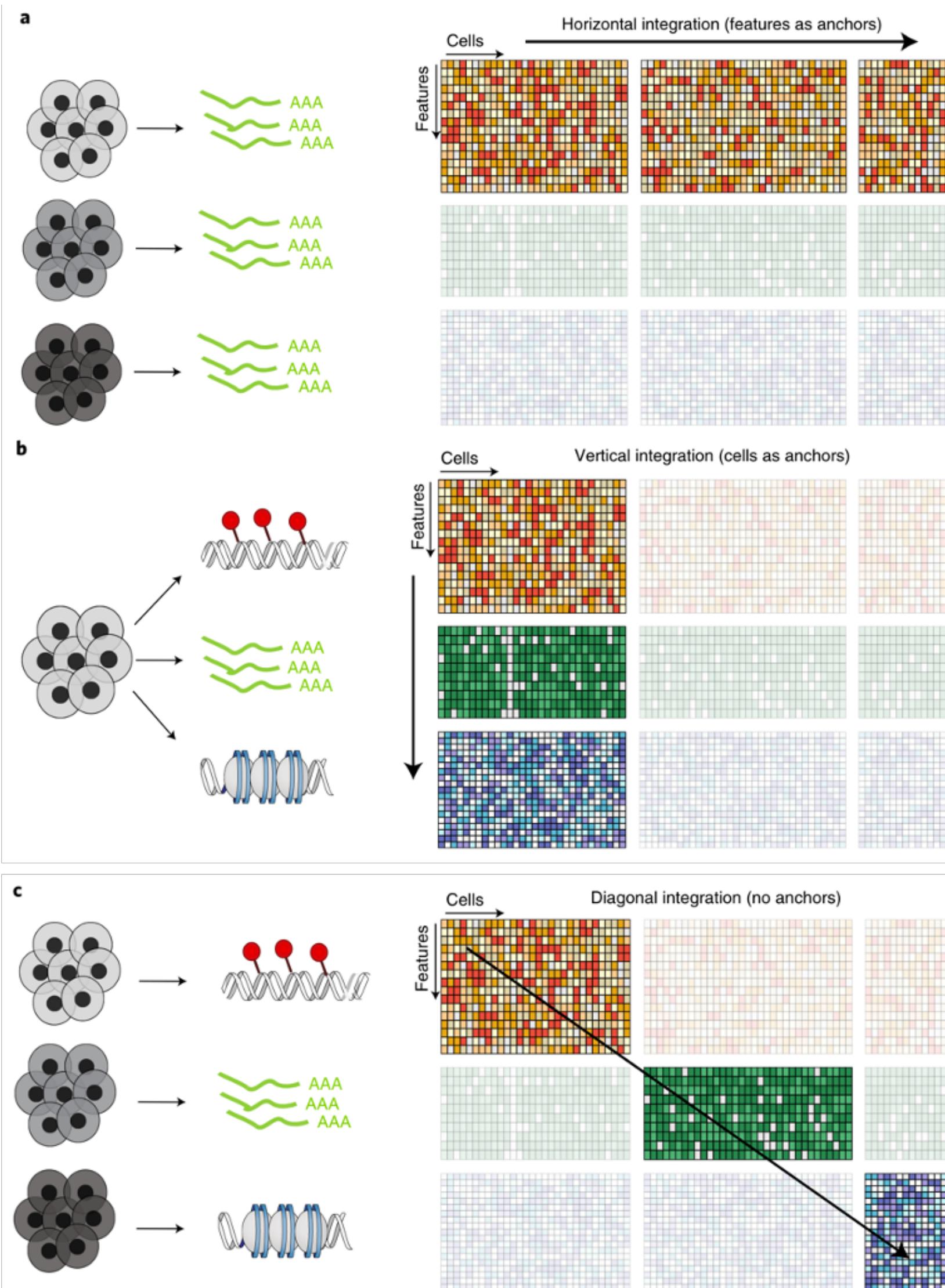
Discuss with
your neighbour

- What do you expect to see from a cell type-specific TF? What do you expect from a house-keeping TF?
- What do you expect to see from a cell-type specific enhancer? What do you expect to see from a house-keeping enhancer?
- How would you interpret a locus where you find strong accessibility at the promoter of a gene in all cell types but varying accessibility in an upstream regulatory element (e.g. 50kb away from the TSS)?

Outline

- Interpretation of ATAC-seq data
- Opportunities and challenges of single cell ATAC-seq
- **Data integration**
 - **What data do you have?**
 - **What do you want to learn from the integration?**
- Gene regulatory networks - example from our work

What data do you want to integrate? And what do you want to learn?



What do you want to learn from data integration?



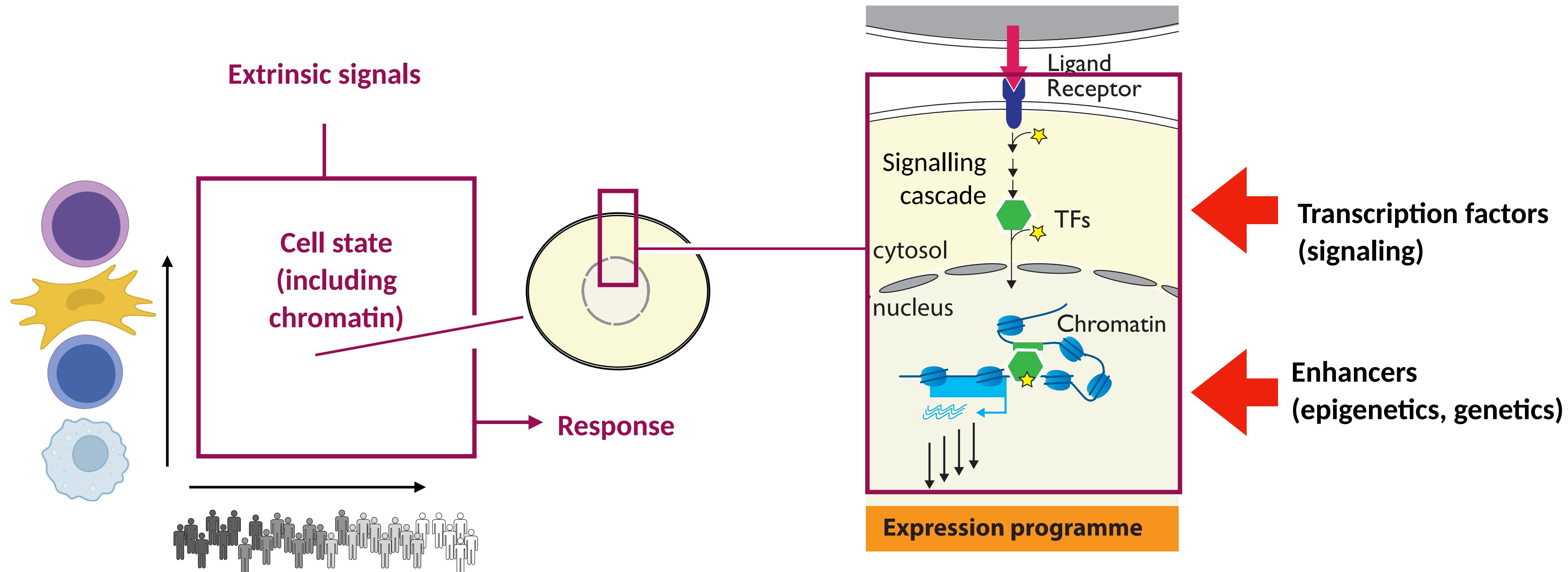
Discuss with
your neighbour

What do you want to learn from data integration?

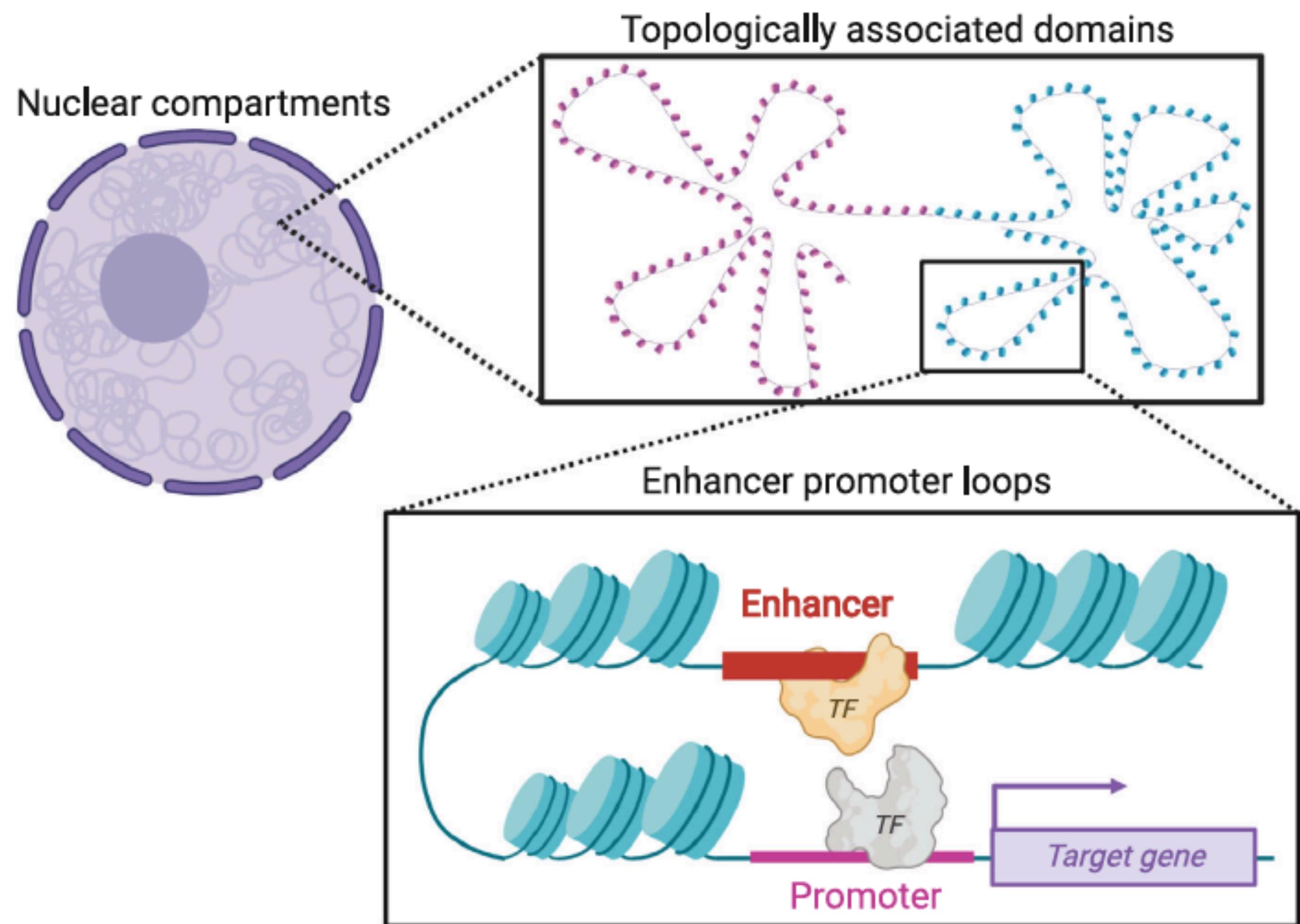
e.g. integrate RNA/ATAC for inferring TF activity

- Promoter accessibility of TF (from scATAC-seq)
 - missing values, accessibility ≠ expression
- Expression of TF (from scRNA-seq)
 - missing values, missing post-transcriptional/translational regulation
- TF-activity inferred from motif accessibility (from scATAC) - effect of TF on chromatin
 - accessibility ≠ activity
- TF-activity inferred from expression of TF targets (from scRNA) - effect of TF on transcription
 - what are the real targets of the TF?

Molecular measures as approximations of internal cell state



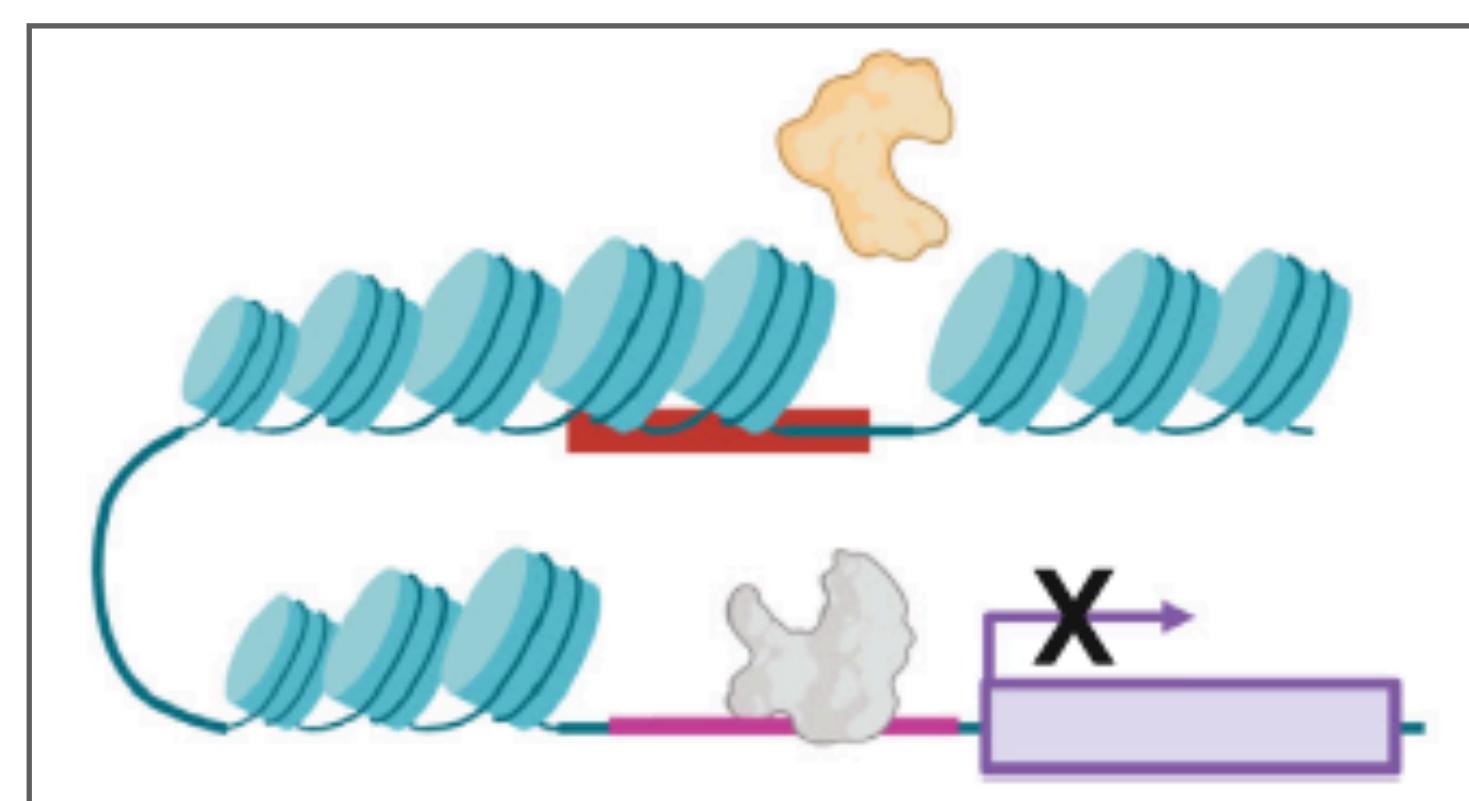
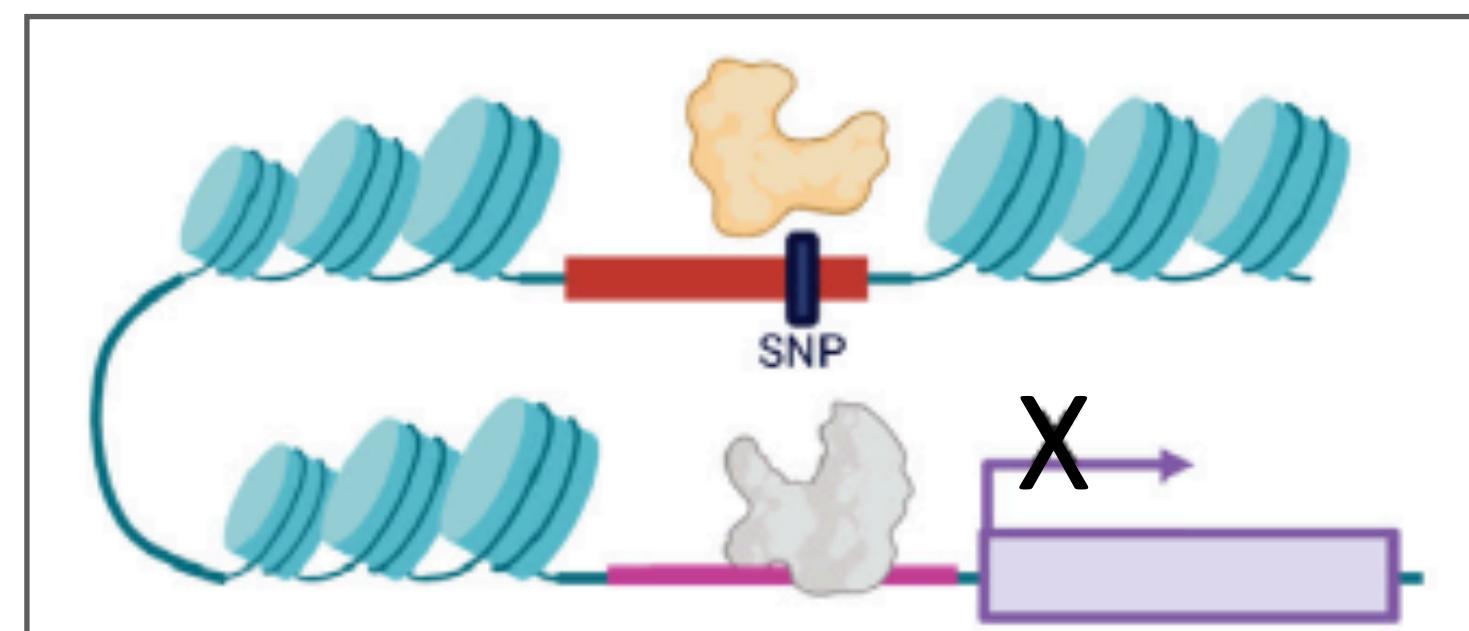
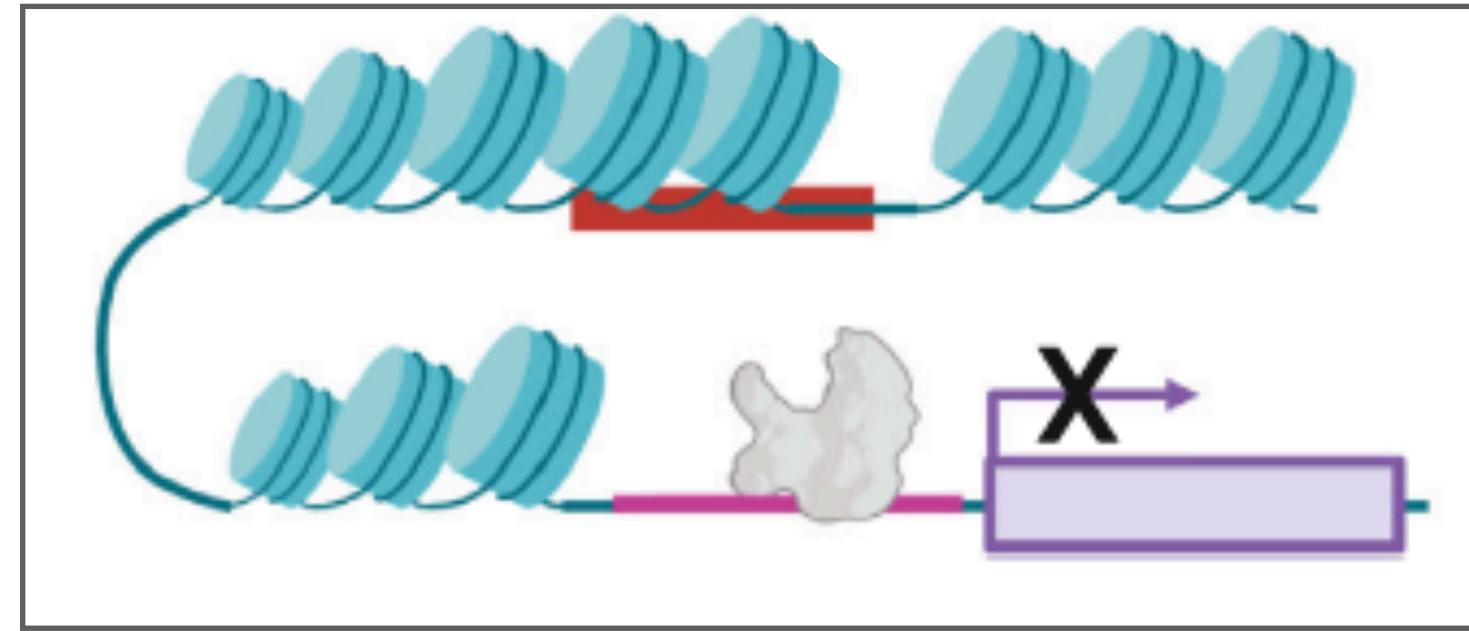
Transcription of a gene is regulated by genetic and epigenetic mechanisms



Absence of TF
=> signalling/
Trans-genetics/
cell type

Genetic variant
disrupts TF binding
=> Cis-genetics

Chromatin
inaccessibility
prevents TF binding
=> Epigenetics



Review: Claringbould & Zaugg (2021) *Trends in Molecular Medicine*

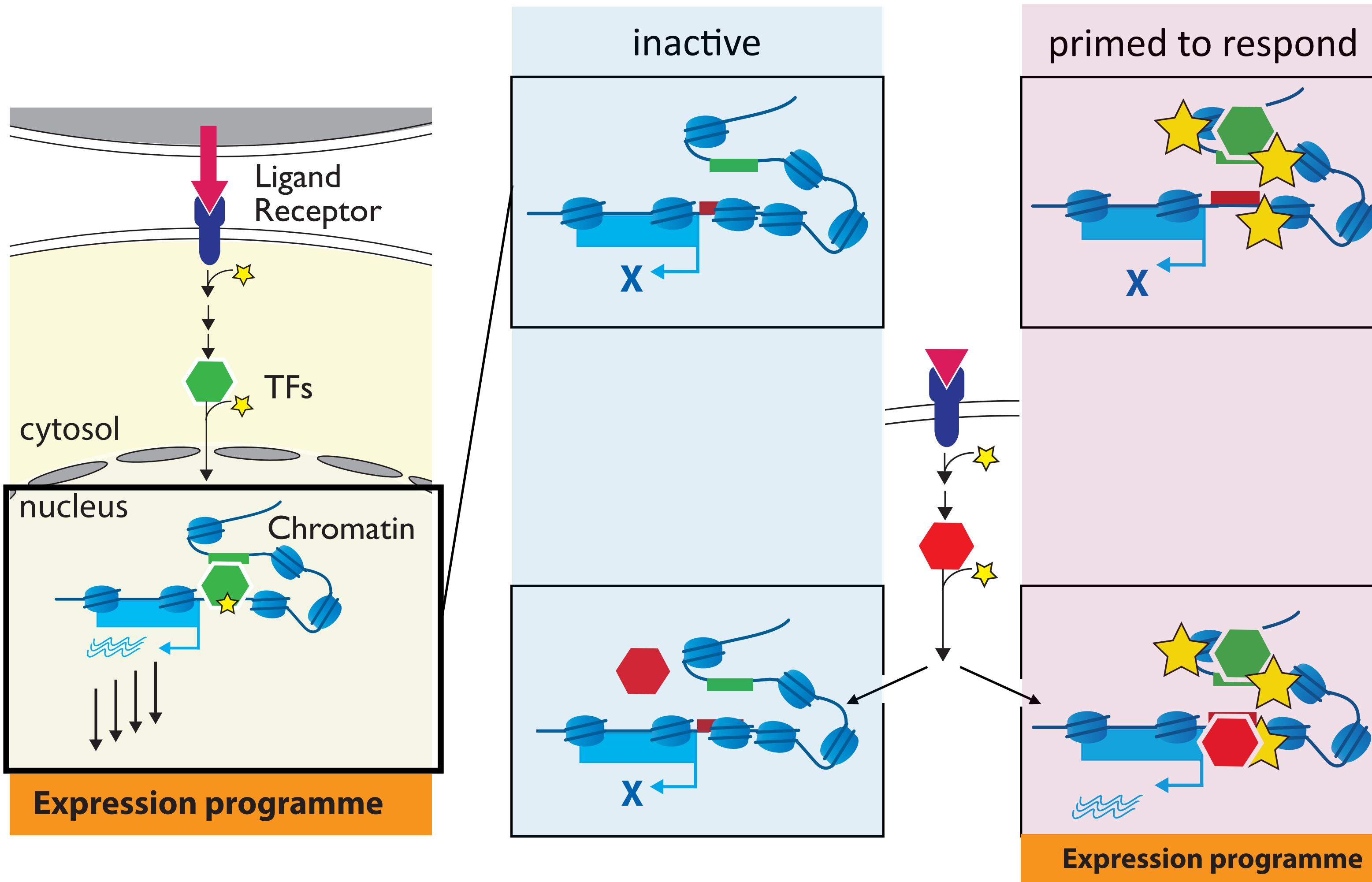
Review: Ruiz-Velasco & Zaugg (2017) *Current Opinion in Systems Biology*

Review: Kholmatov, Weidemueller, Petsalaki+, Zaugg+ (2021) *Proteins*

Perspective: Zaugg*, Anderson*, Sahlén* et al (2022) *Nature Structural Molecular Biology*

Chromatin accessibility is a readout of many regulatory processes

Enhancers may define how a cell responds to signals: "Priming"



Enhancer-priming in ageing human bone marrow stromal cells linked to immune traits

Lai,..., Zaugg (2021) *bioRxiv*, in revision

BDNF-response enhancers linked to neuro-psychiatric traits in neurons

Ibarra,..., Zaugg+, Noh+ (2022) *Molecular Systems Biology*

Enhancer-mediated lineage priming in early hematopoietic stem cells

Ranzoni,..., Zaugg, Cevjic (2021) *Cell Stem Cell*

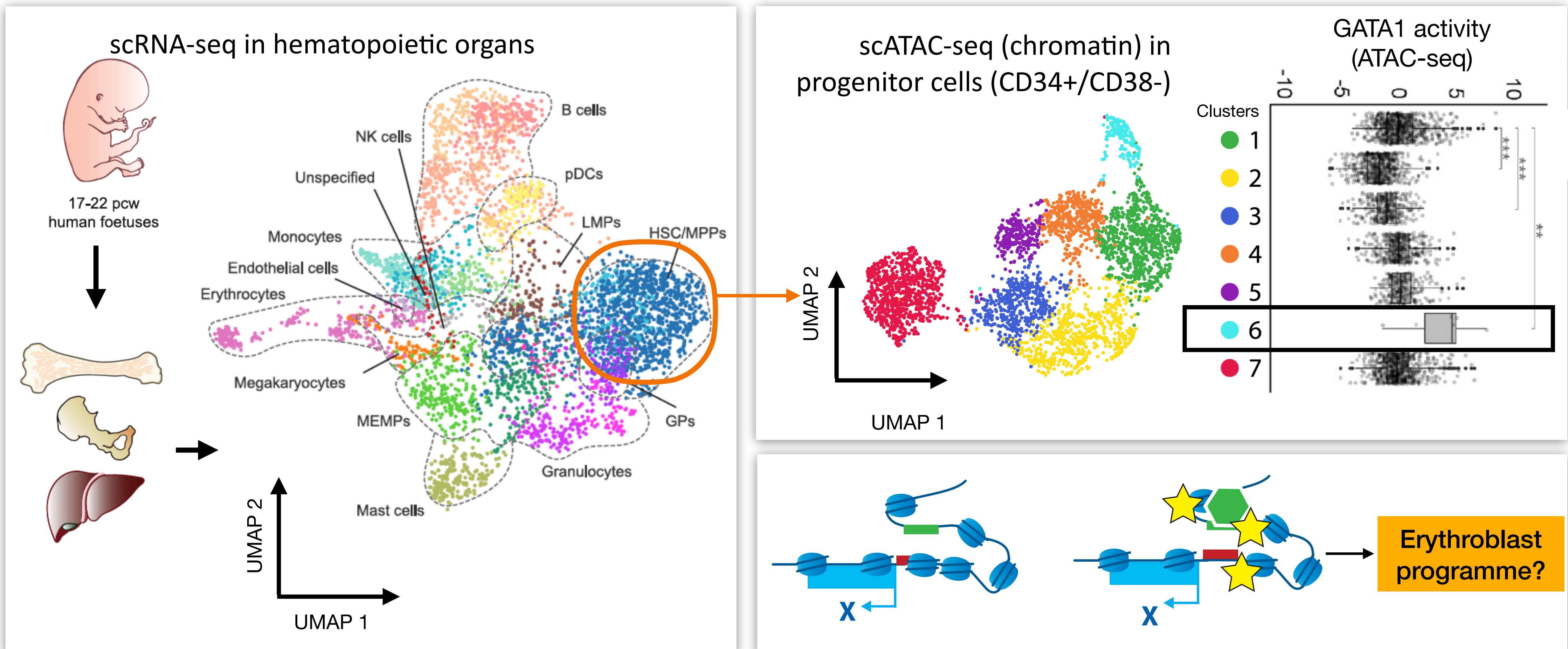
Enhancer remodeling linked to response to signaling in Pulmonary Arterial Hypertension

Reyes-Palomares,..., Zaugg (2020) *Nature Communications*

LSD1 mutation (in rare disease) causes repression of enhancer networks

Bunina+,..., Zaugg+, Noh+ (2021) *bioRxiv*, in revision

ATAC-seq suggests early lineage-priming in hematopoietic stem and progenitor cells (not seen in RNA)

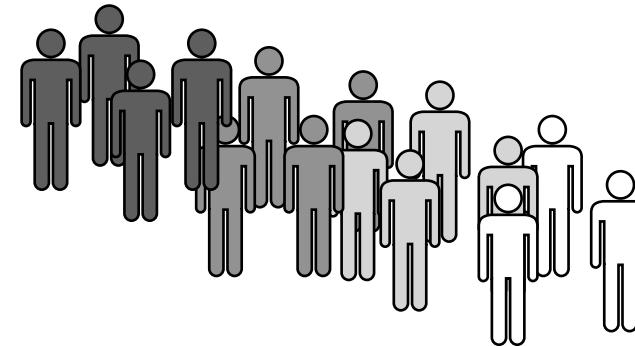


Outline

- Interpretation of ATAC-seq data
- Opportunities and challenges of single cell ATAC-seq
- Data integration
 - What data do you have?
 - What do you want to learn from the integration?
- **Gene regulatory networks - example from our work**

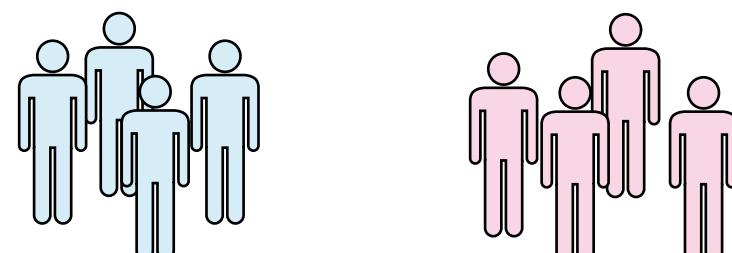
Integrate *genetics*, *epigenetics*, *signaling* to predict disease mechanisms - based on molecular variation across individuals or cells

Molecular profiling
(heterogeneous cohort)

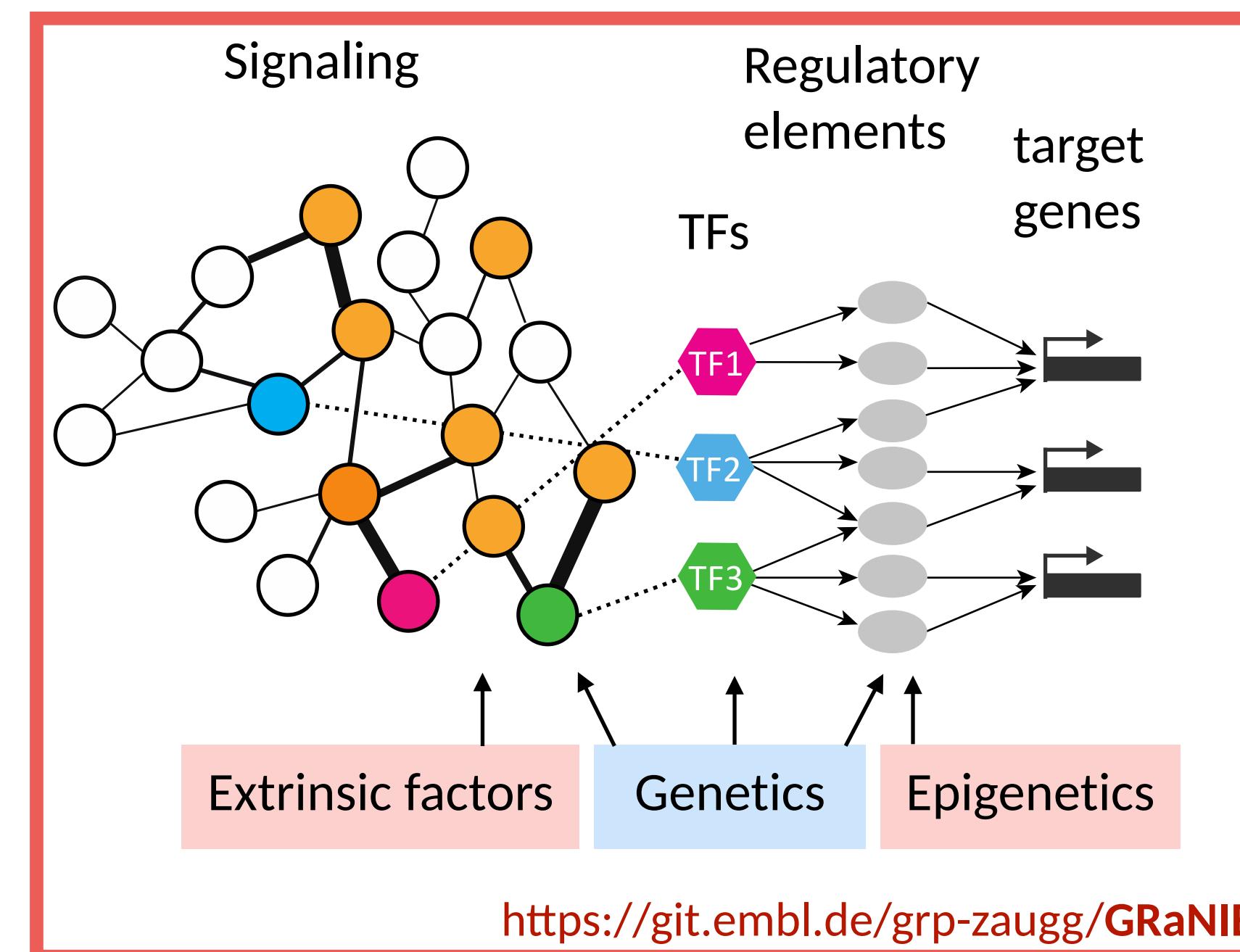


- RNA
- chromatin
- surface receptors

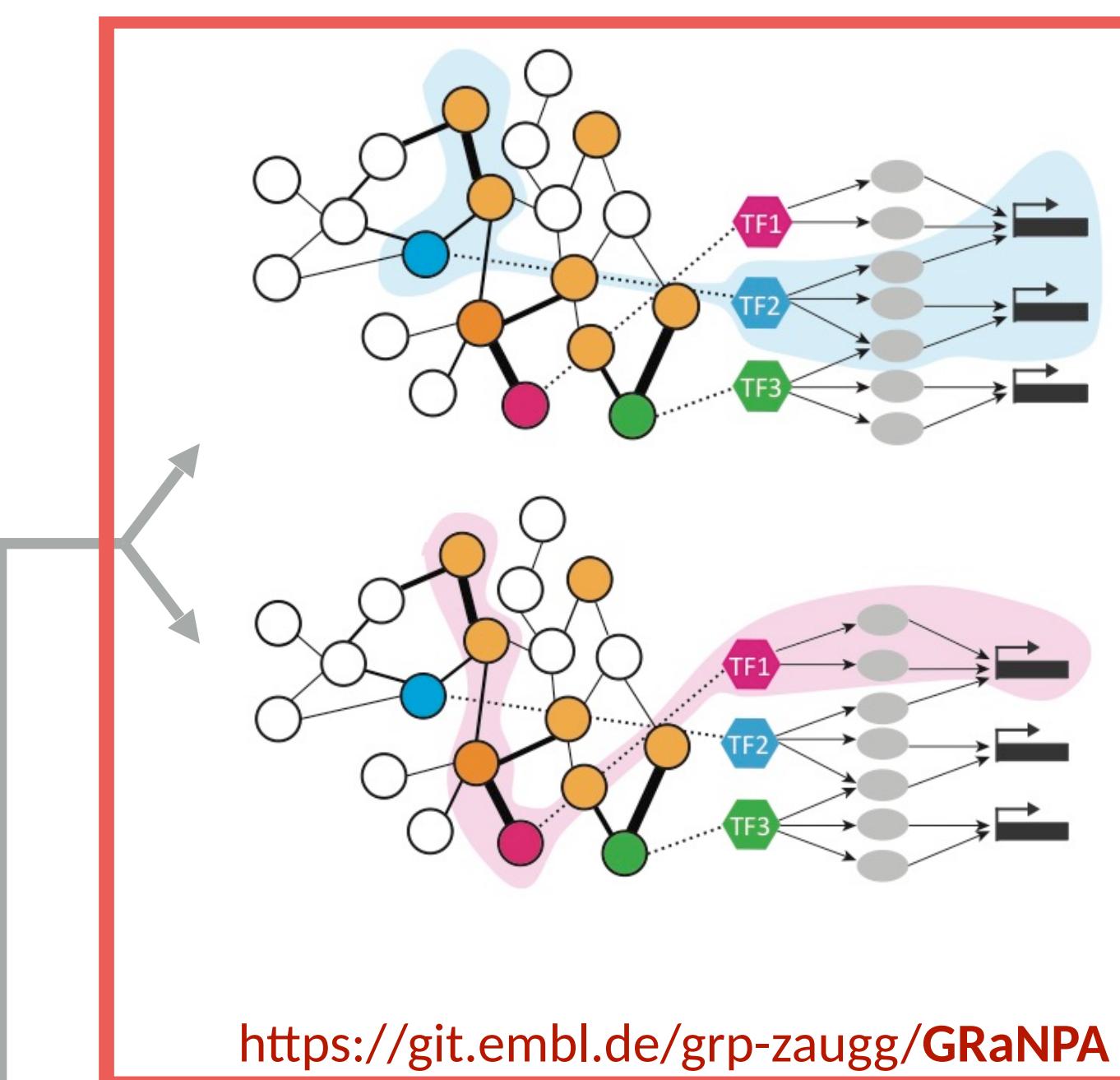
Molecular disease data
(disease cohorts)



Regulatory network models (GRaNIE)



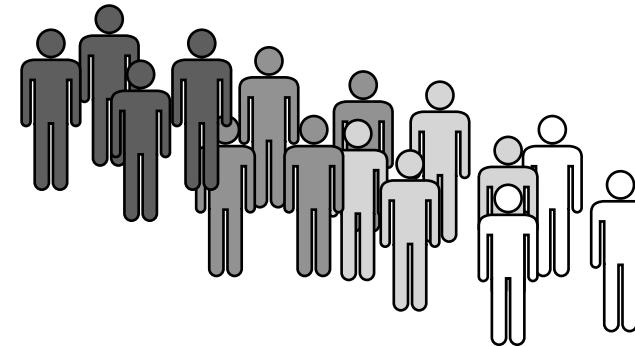
Predicted mechanisms (GRaNPA)



Kamal*, Arnold* et al (2023) Molecular Systems Biology
Reyes-Palomares (2020) Nature Communications
Berest*, Arnold* (2019) Cell Reports
Grubert (2015) Cell
Kasowski (2013) Science

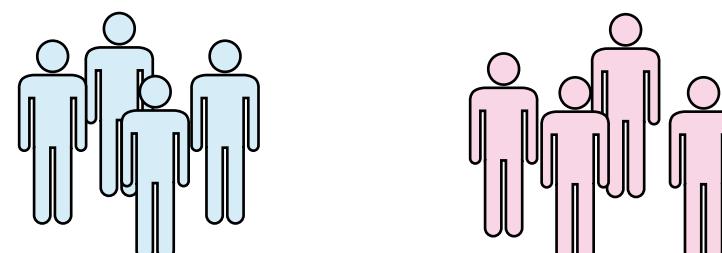
Integrate *genetics*, *epigenetics*, *signaling* to predict disease mechanisms - based on molecular variation across individuals or cells

Molecular profiling
(heterogeneous cohort)

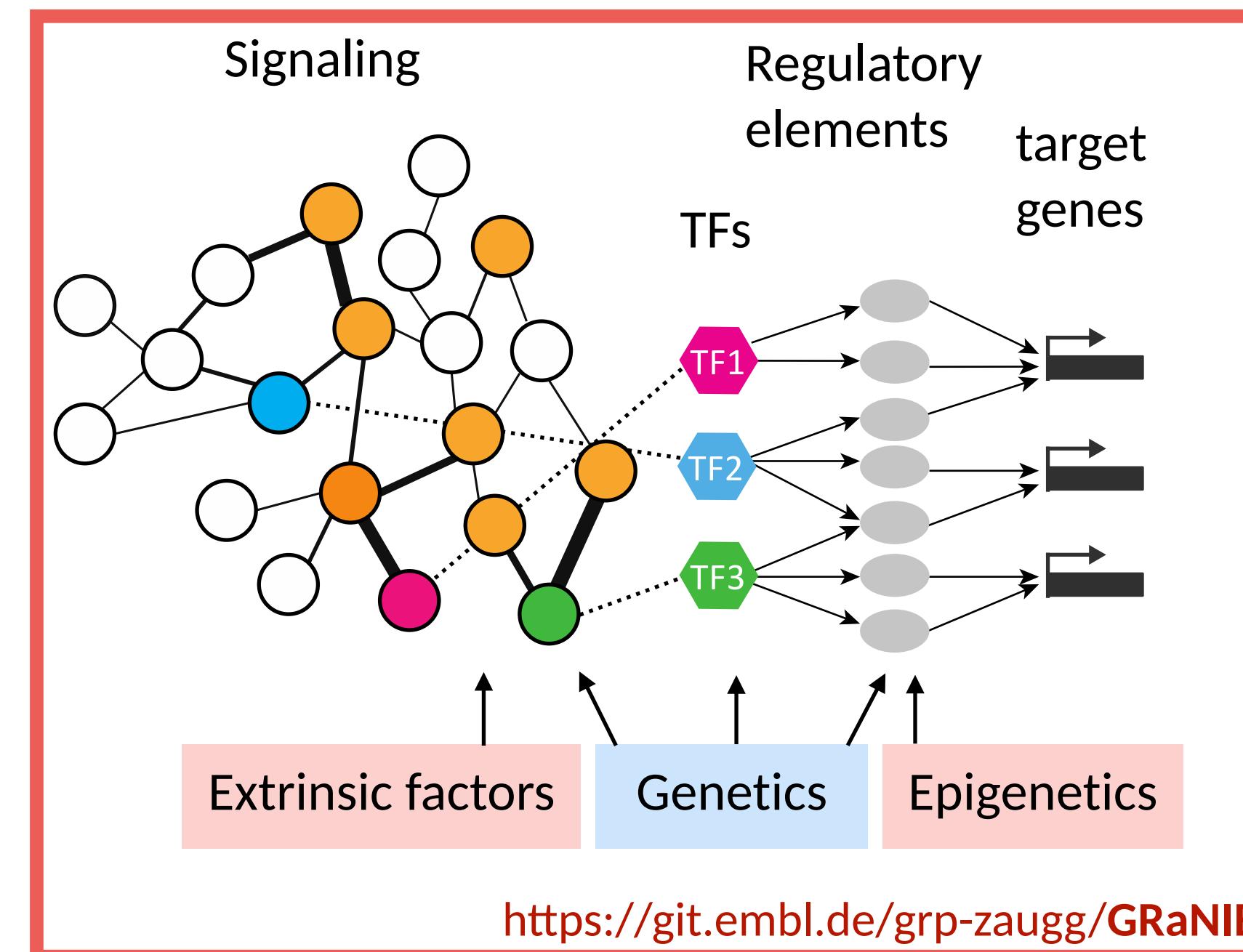


- RNA
- chromatin
- surface receptors

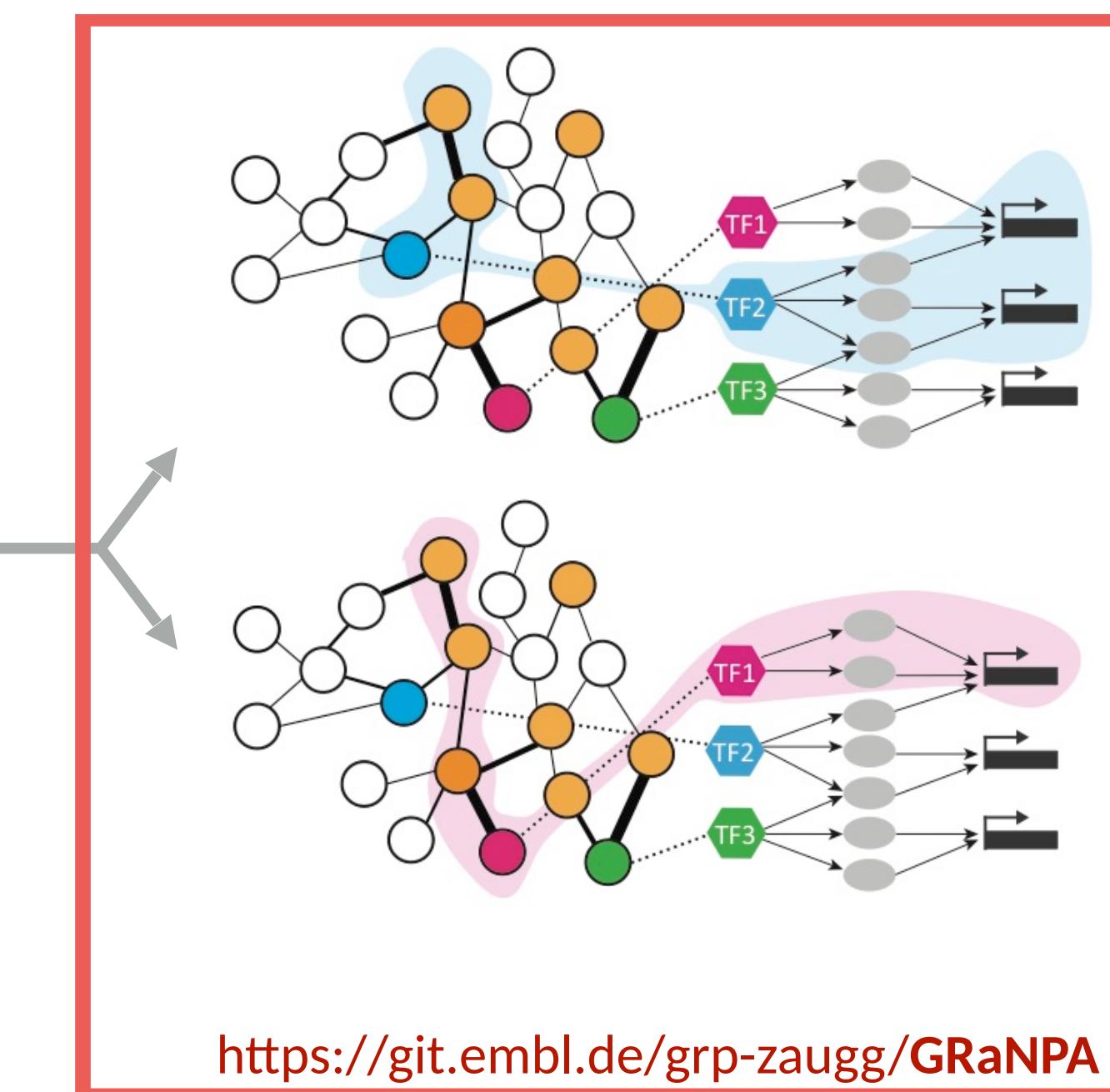
Molecular disease data
(disease cohorts)



Regulatory network models (GRaNIE)



Predicted mechanisms (GRaNPA)



Kamal*, Arnold* et al (2023) *Molecular Systems Biology*
Reyes-Palomares (2020) *Nature Communications*
Berest*, Arnold* (2019) *Cell Reports*
Grubert (2015) *Cell*
Kasowski (2013) *Science*

How can we use these networks to understand disease mechanisms?

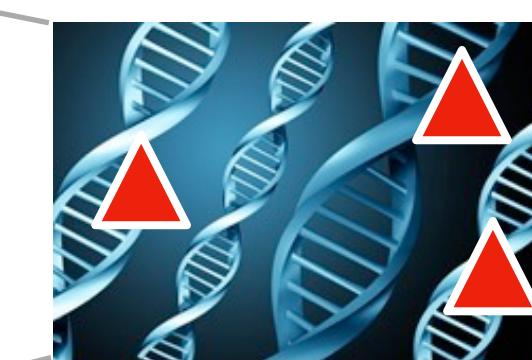
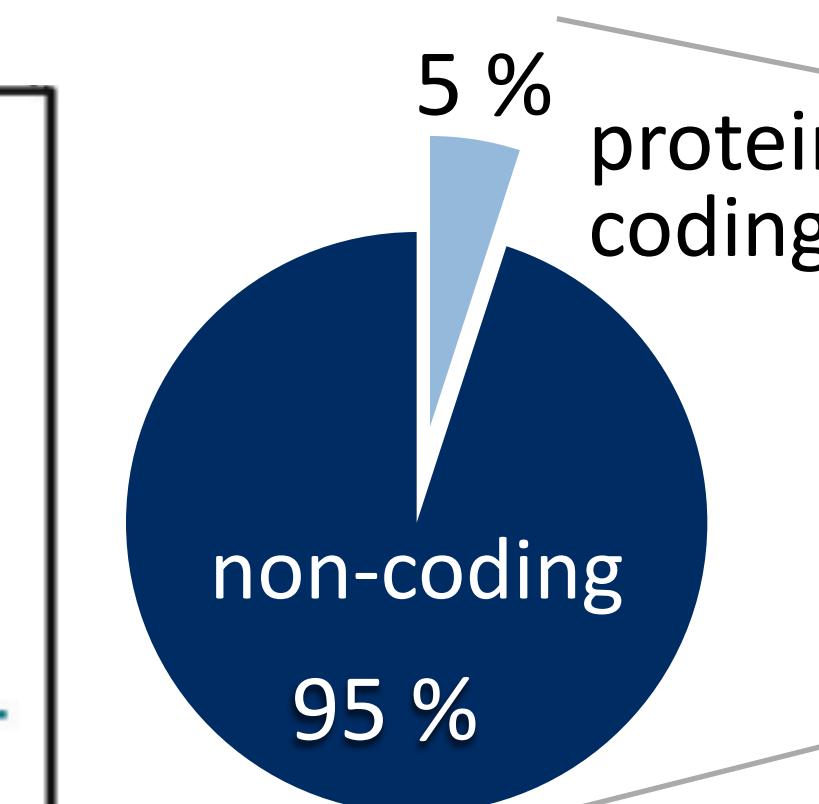
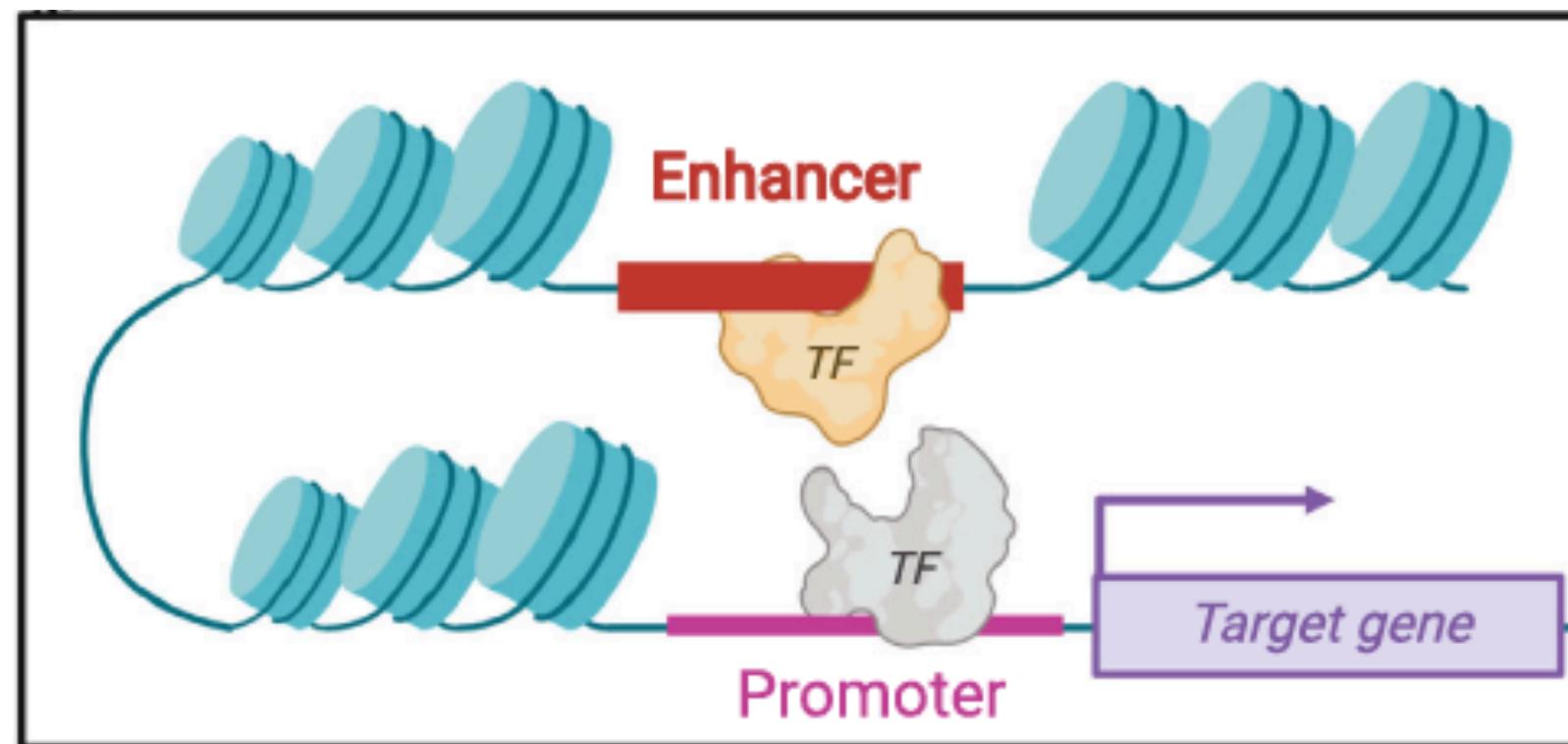
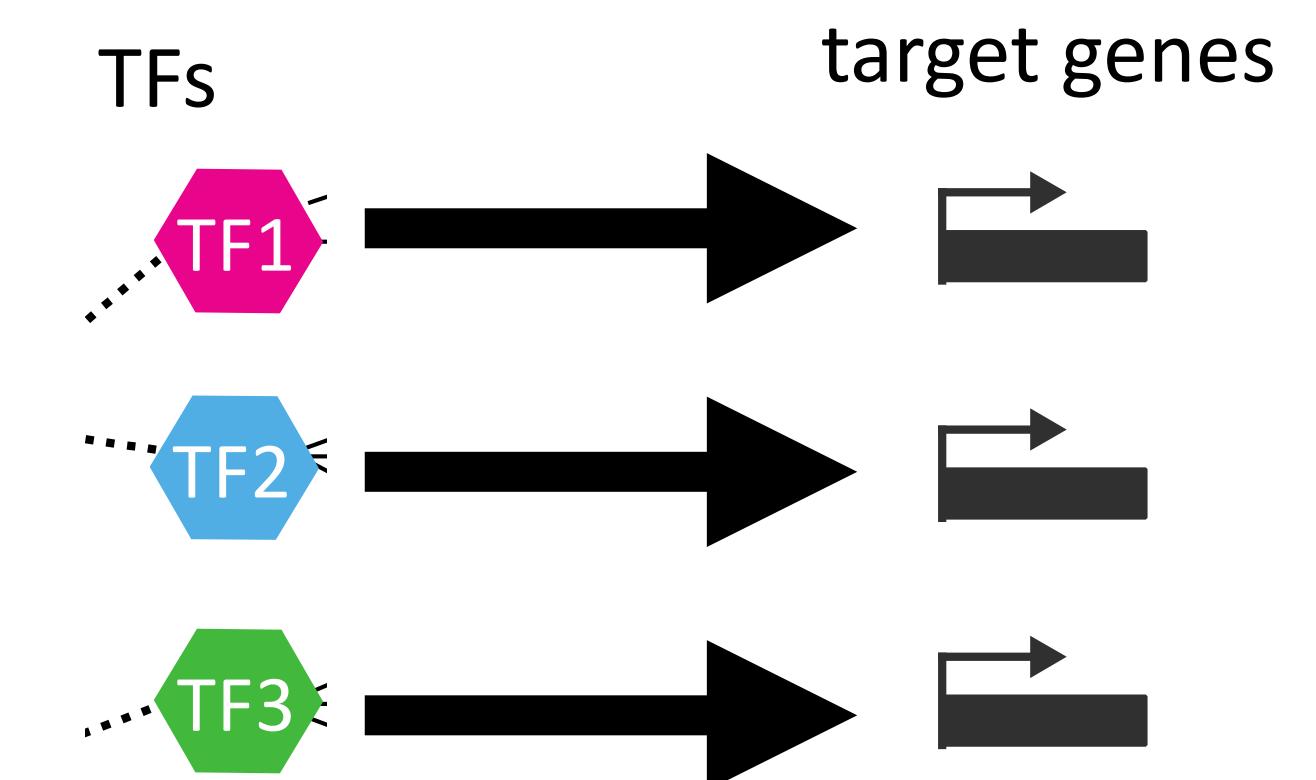
Many approaches for inferring transcription factor target genes

- most ignore enhancers

- Co-expression networks
e.g. WGCNA (Langfelder & Horvath, 2008)
- Time course data
e.g. Huynh Thu and Geurts 2018,
- Based on TF knockdown studies
e.g. KnockTF database - Feng et al, 2020
- Based on TF motifs in promoters
e.g. Luisier et al, NAR 2014

- Curated/integrated databases

e.g. DoROTHEA: Garcia-Alonso et al, 2019
e.g. TRRUST: Han et al, 2018



SNPs
~200'000

Genetics
trait associated
genetic variants

Need to include enhancers

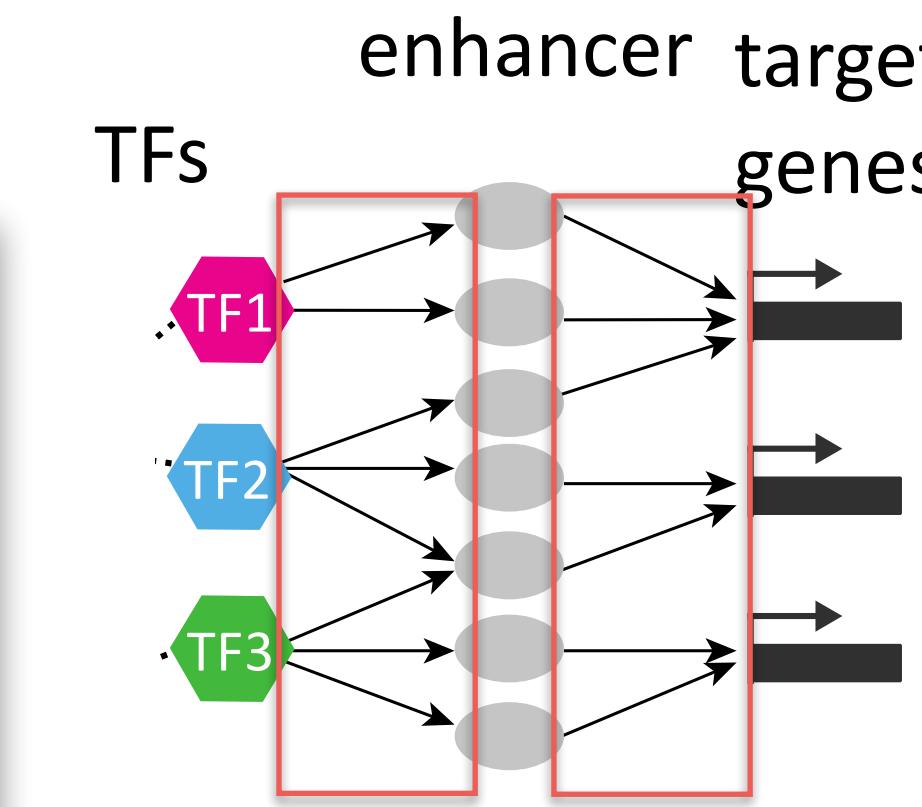
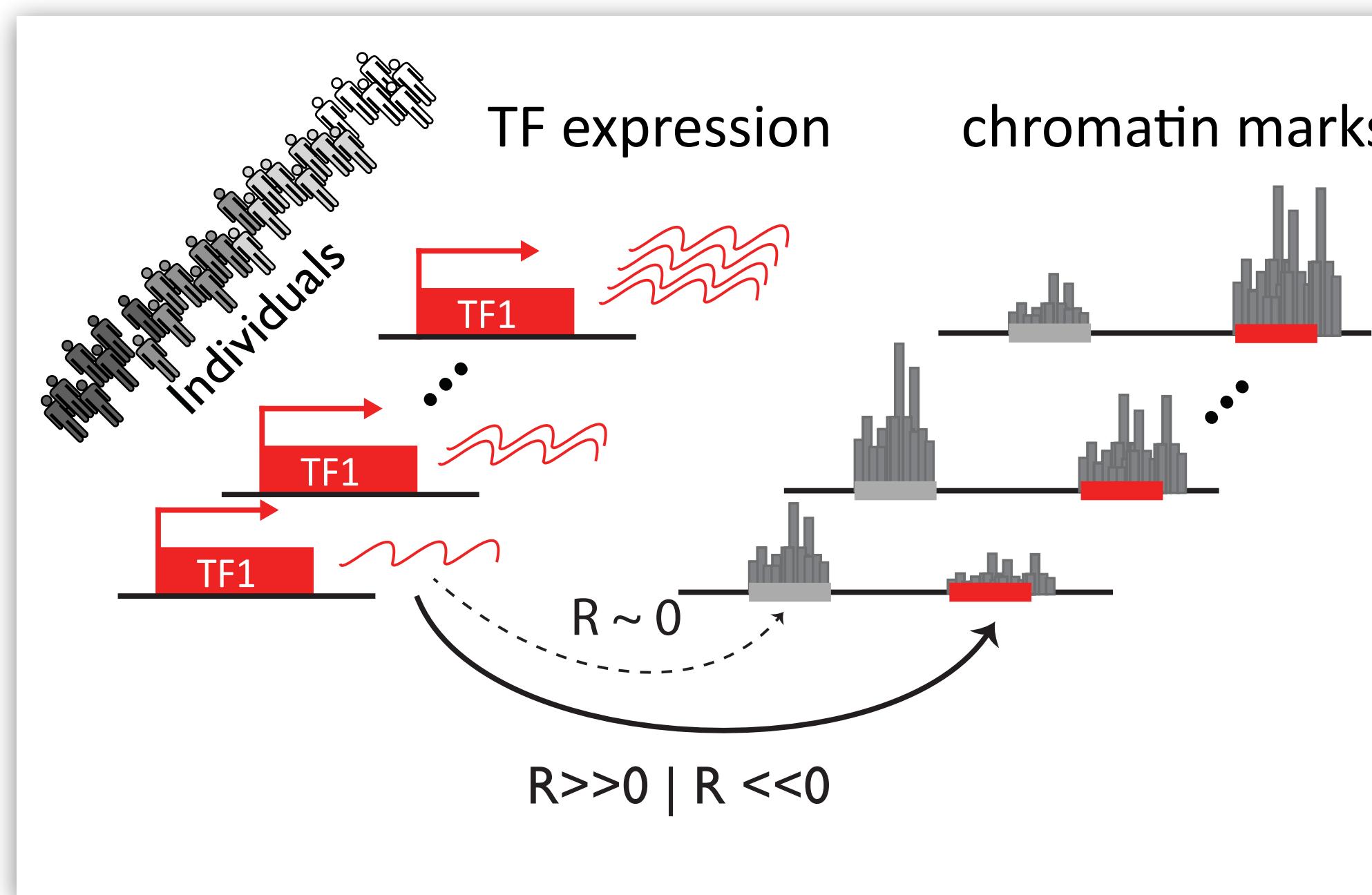


Enhancer-mediated gene regulatory networks (eGRNs) inferred from epigenetic/genetic variation across individuals (or cells)

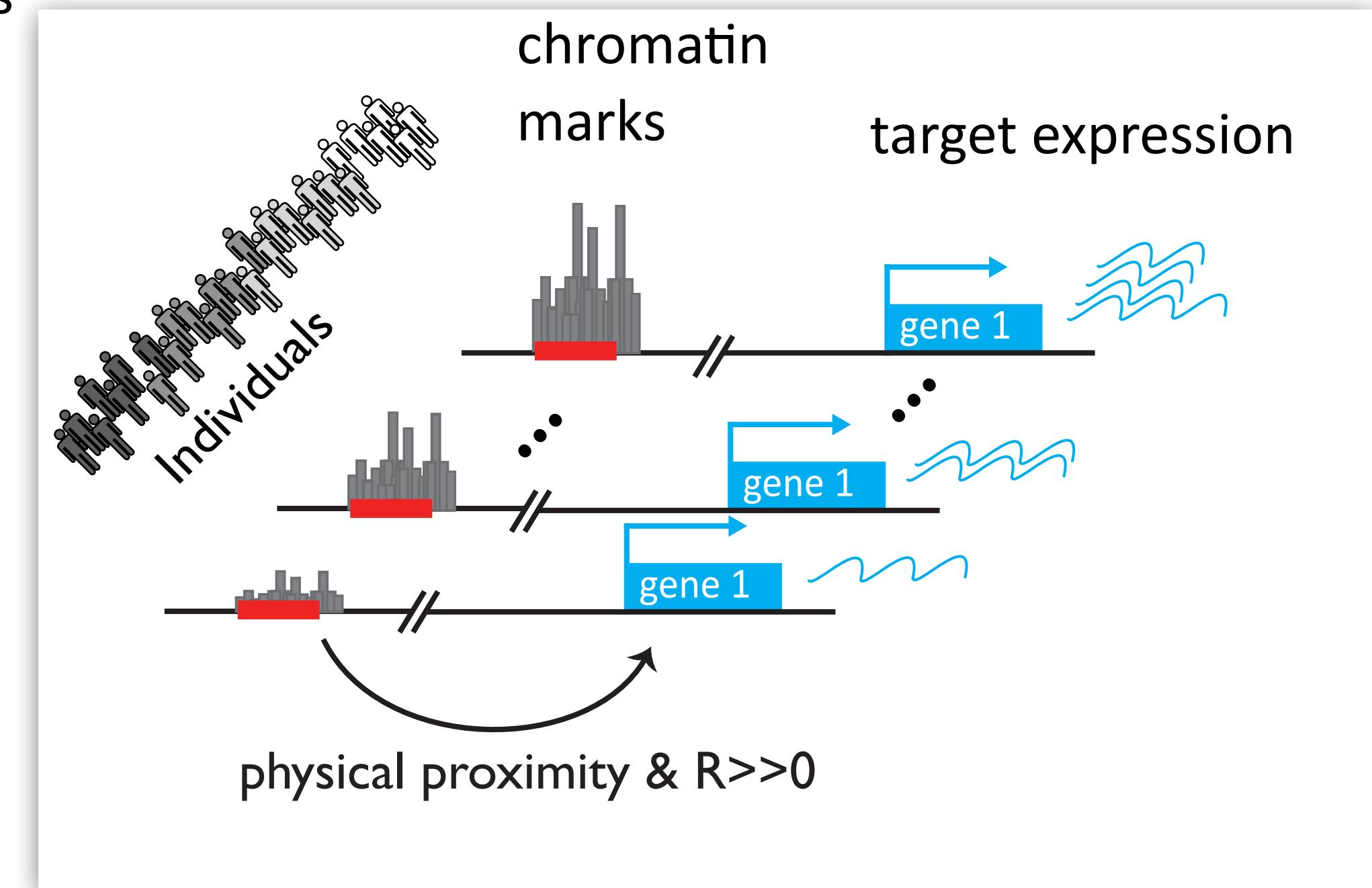
Aryan Kamal

Christian Arnold

TF => regulatory element



Enhancer => gene

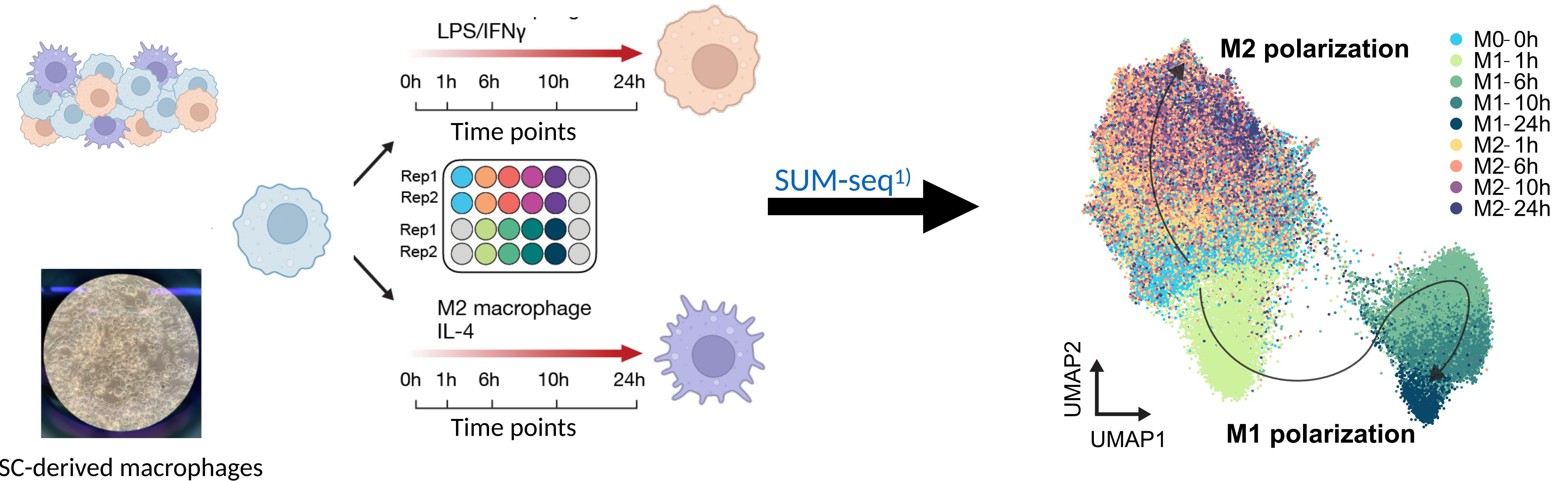


GRaNIE

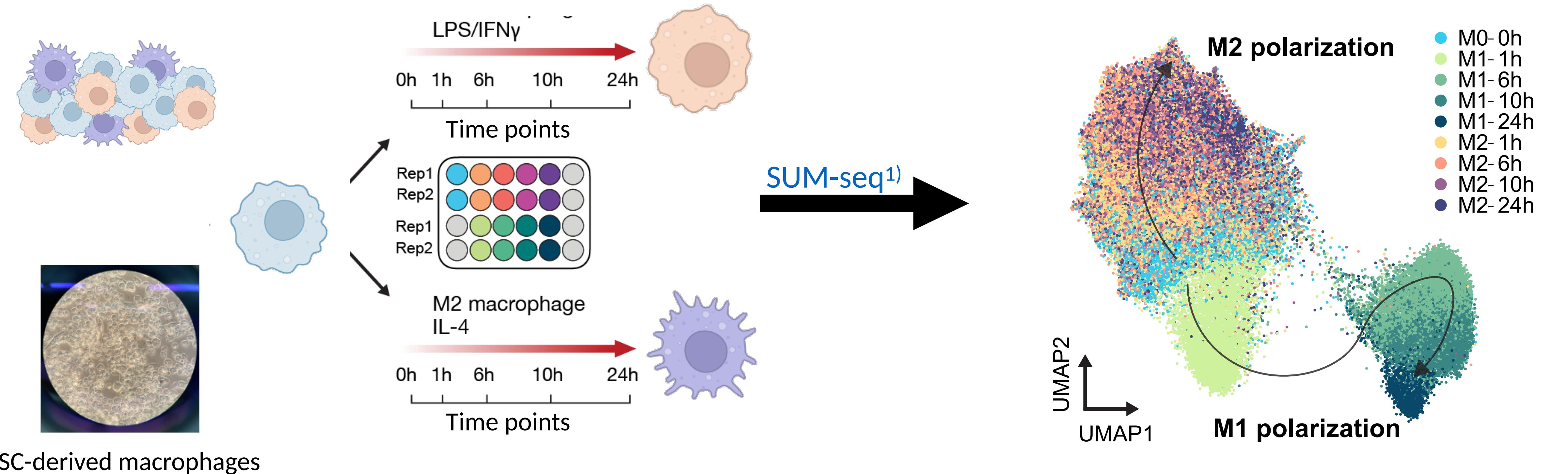
Gene Regulatory Network Inference using Enhancers: [10.18129/B9.bioc.GRaNIE](https://doi.org/10.18129/B9.bioc.GRaNIE)

Similar concept: SCENIC+ (Aerts), FigR (Buonrostro), Pando (Treutlein),..

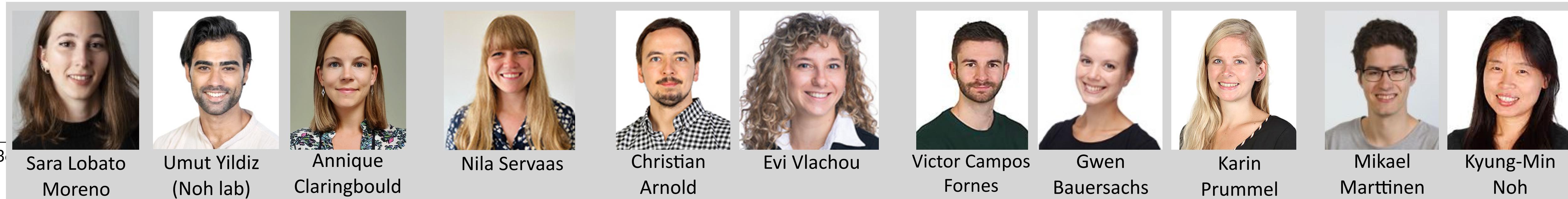
Single cell RNA / chromatin accessibility map of macrophage polarisation - profiled by SUM-seq¹⁾



Single cell RNA / chromatin accessibility map of macrophage polarisation - profiled by SUM-seq¹⁾

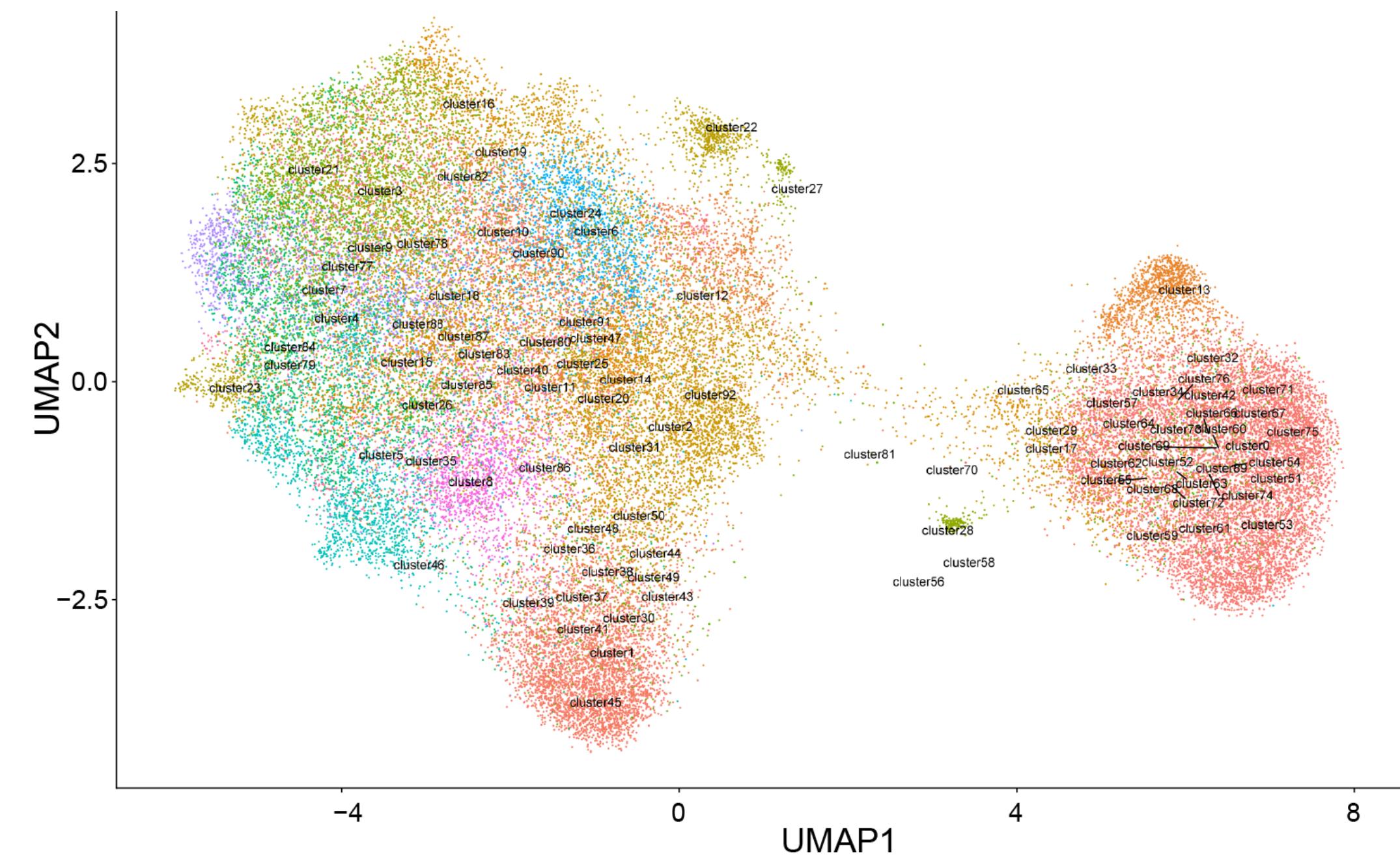


iPSC-derived macrophages



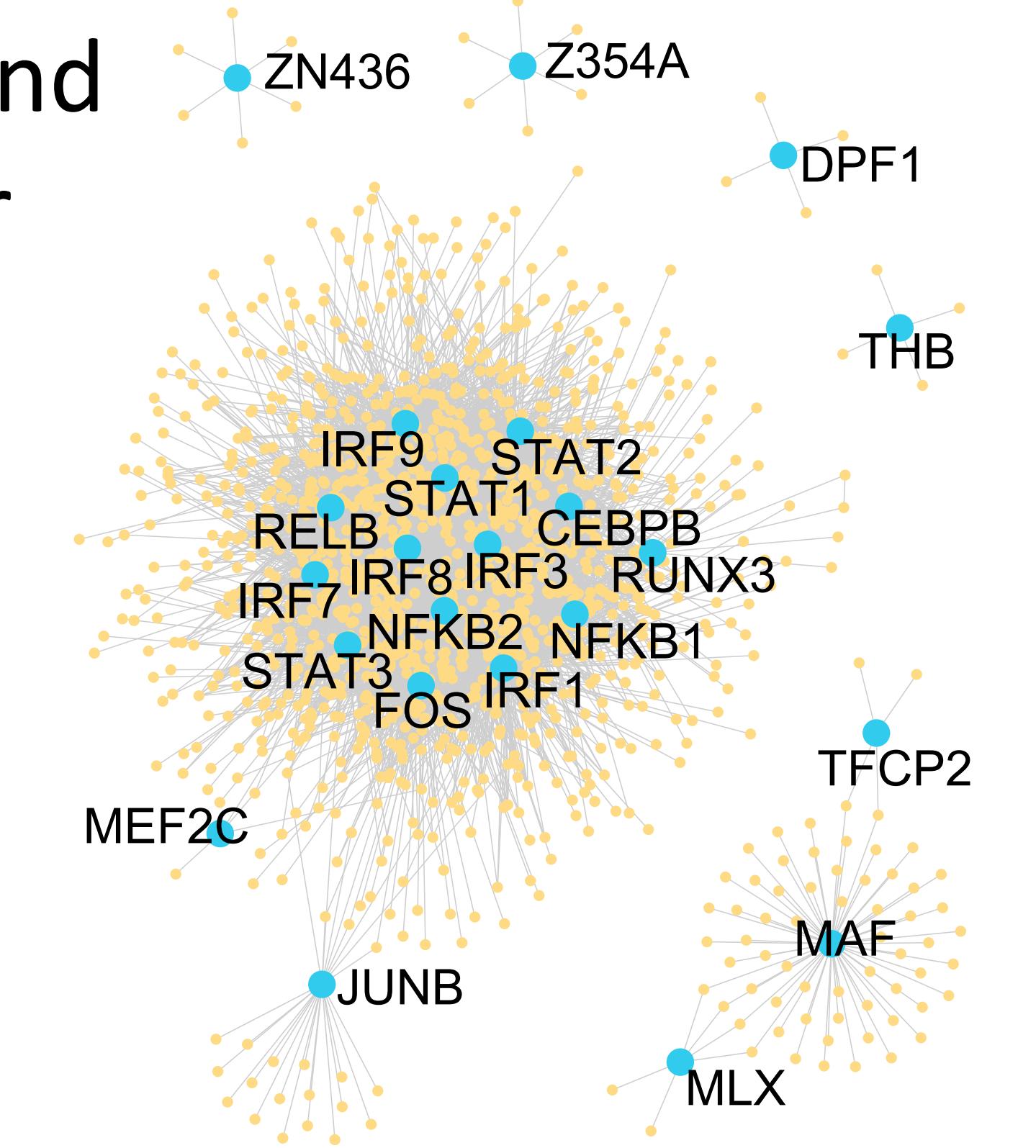
Inferring GRN from single cell multiome data along a time-course

- WNN to integrate modalities
- Overcluster (31 clusters)

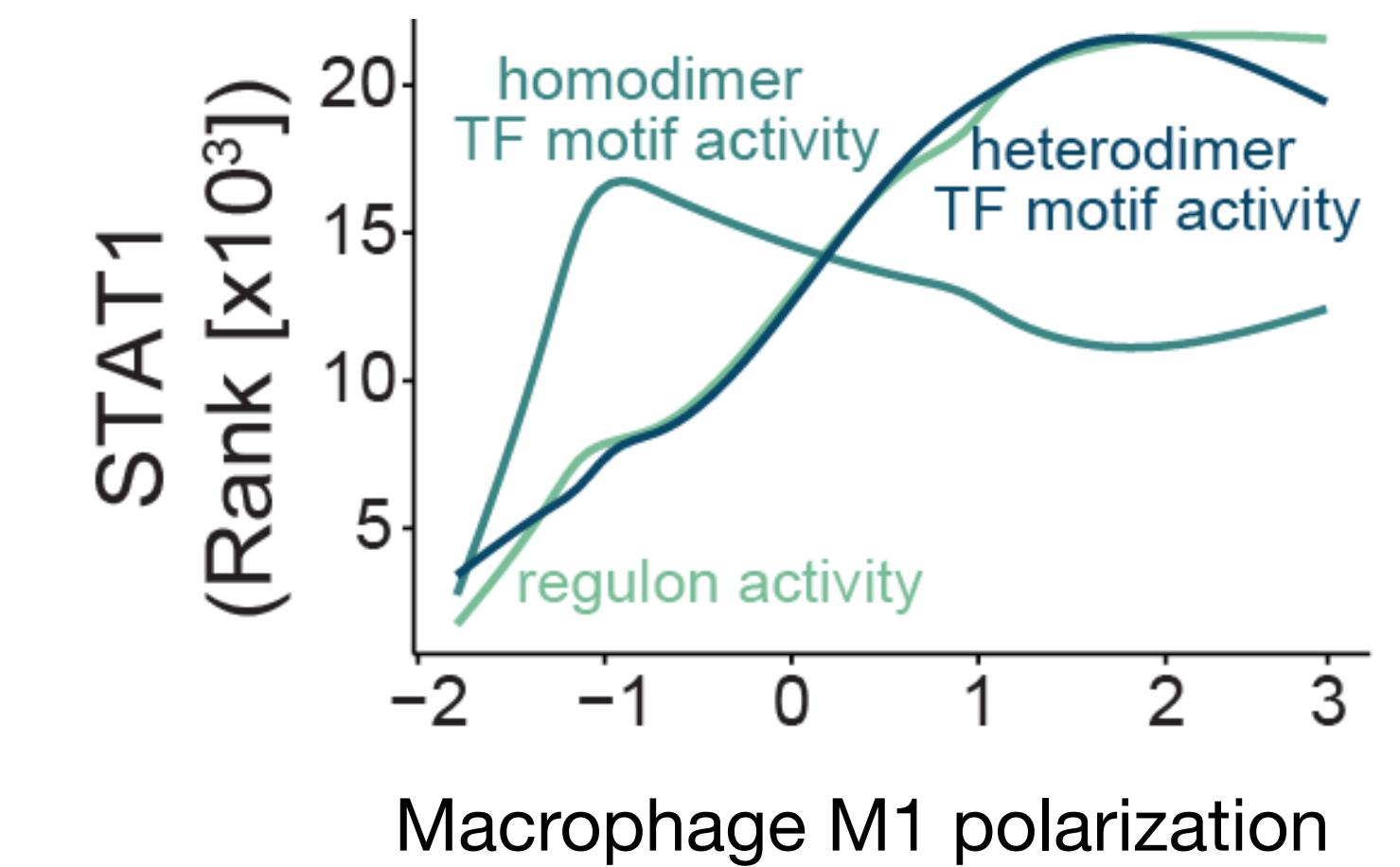
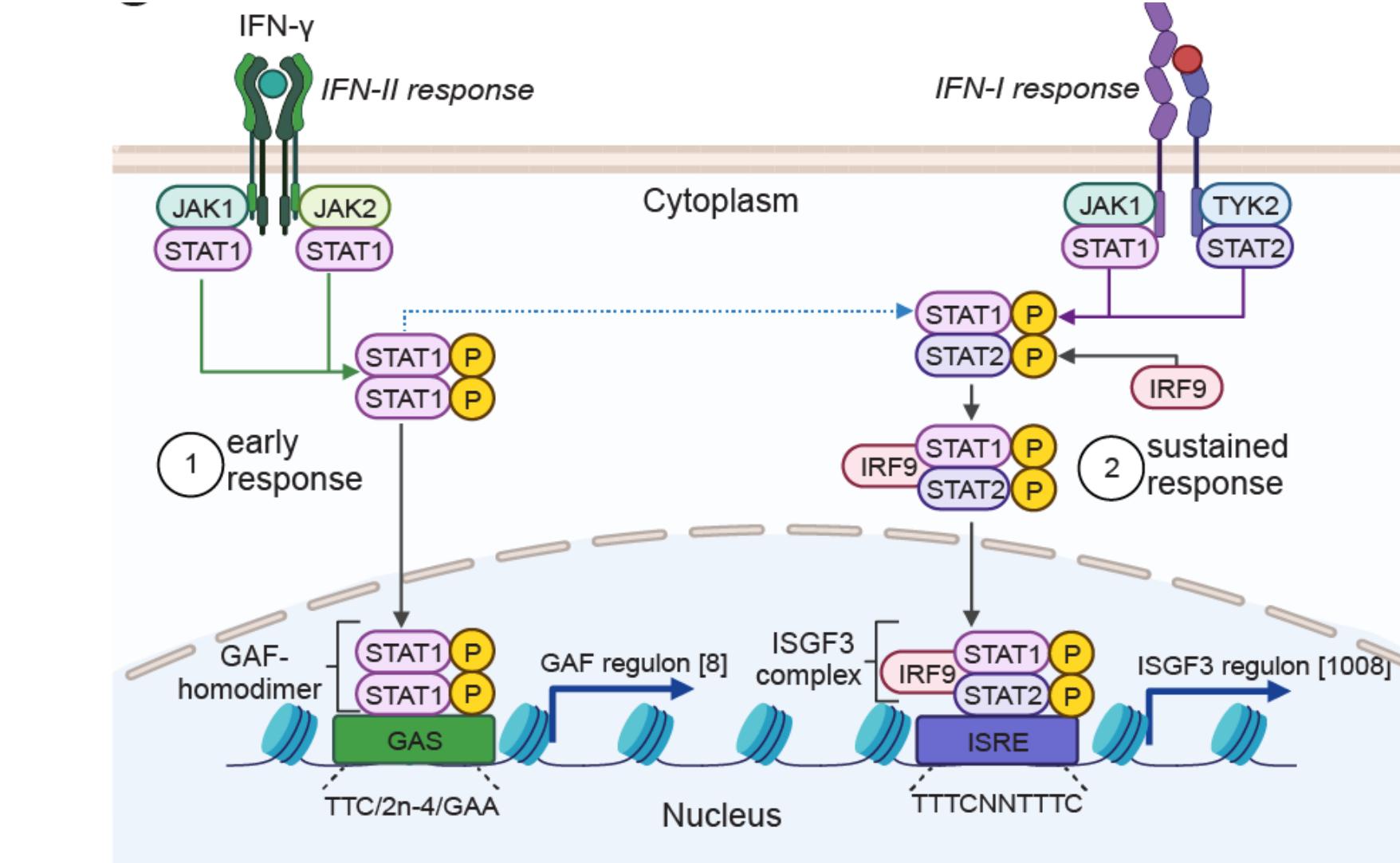
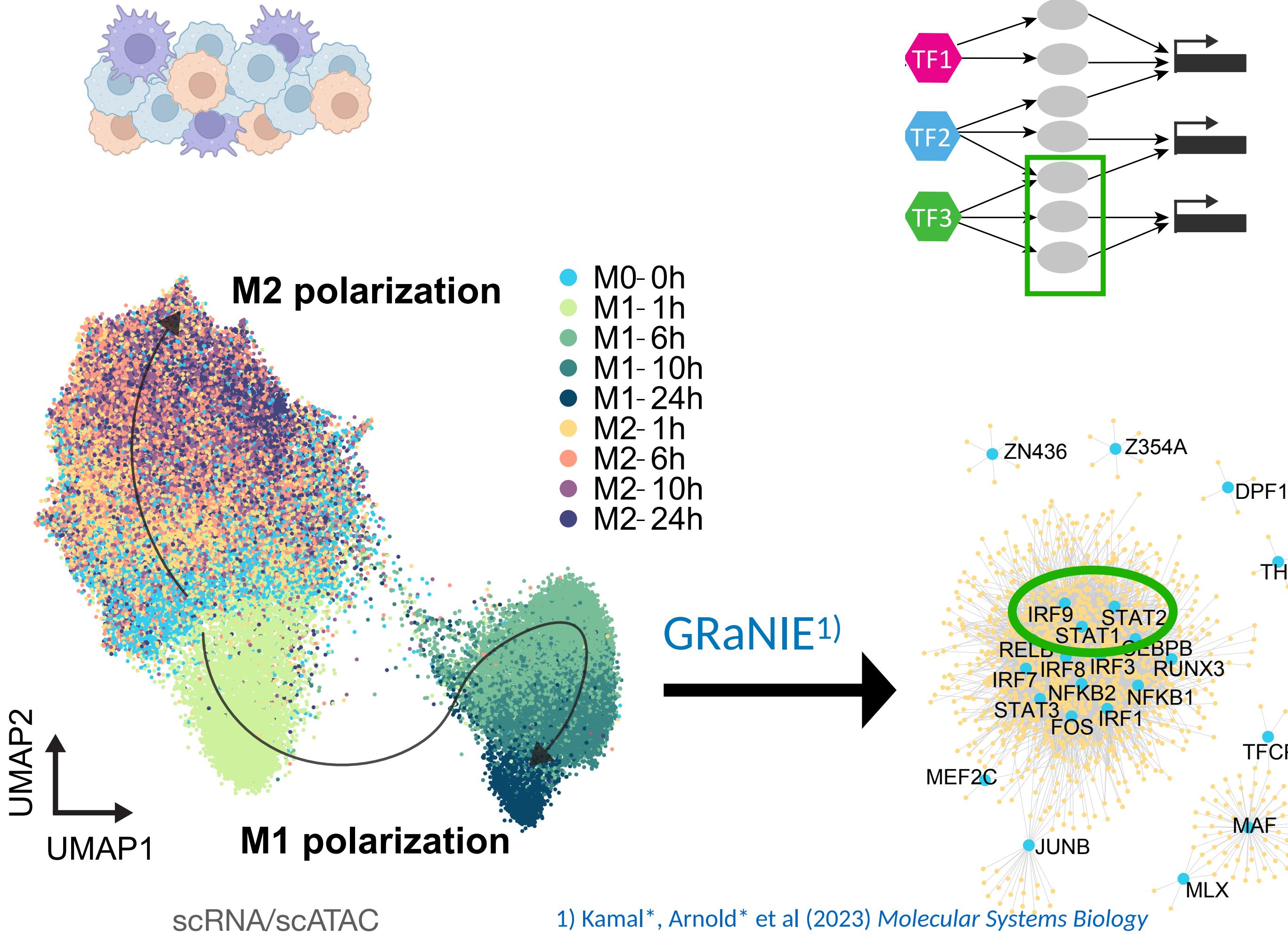


- pseudobulk ATAC and RNA in each cluster
- run GRaNIE

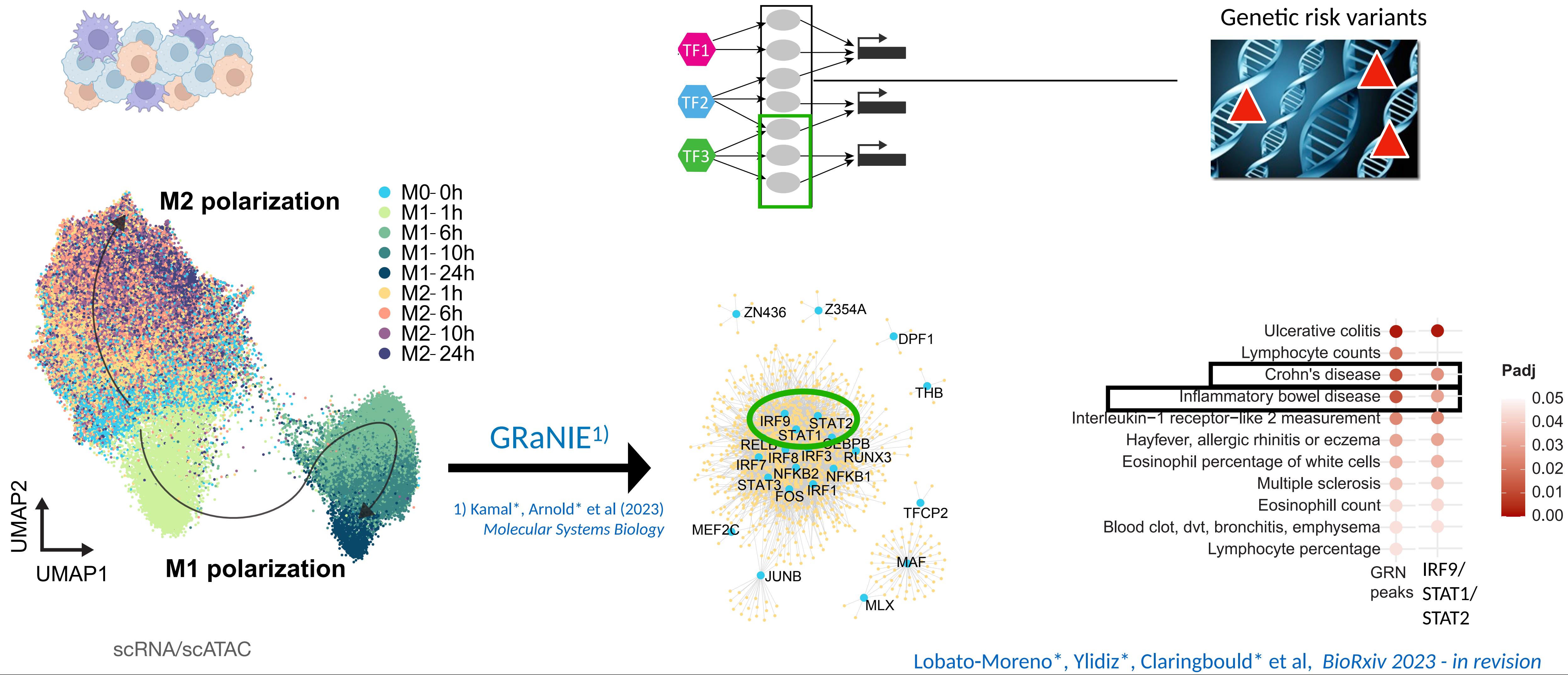
GRaNIE¹



Data integration recapitulates biology



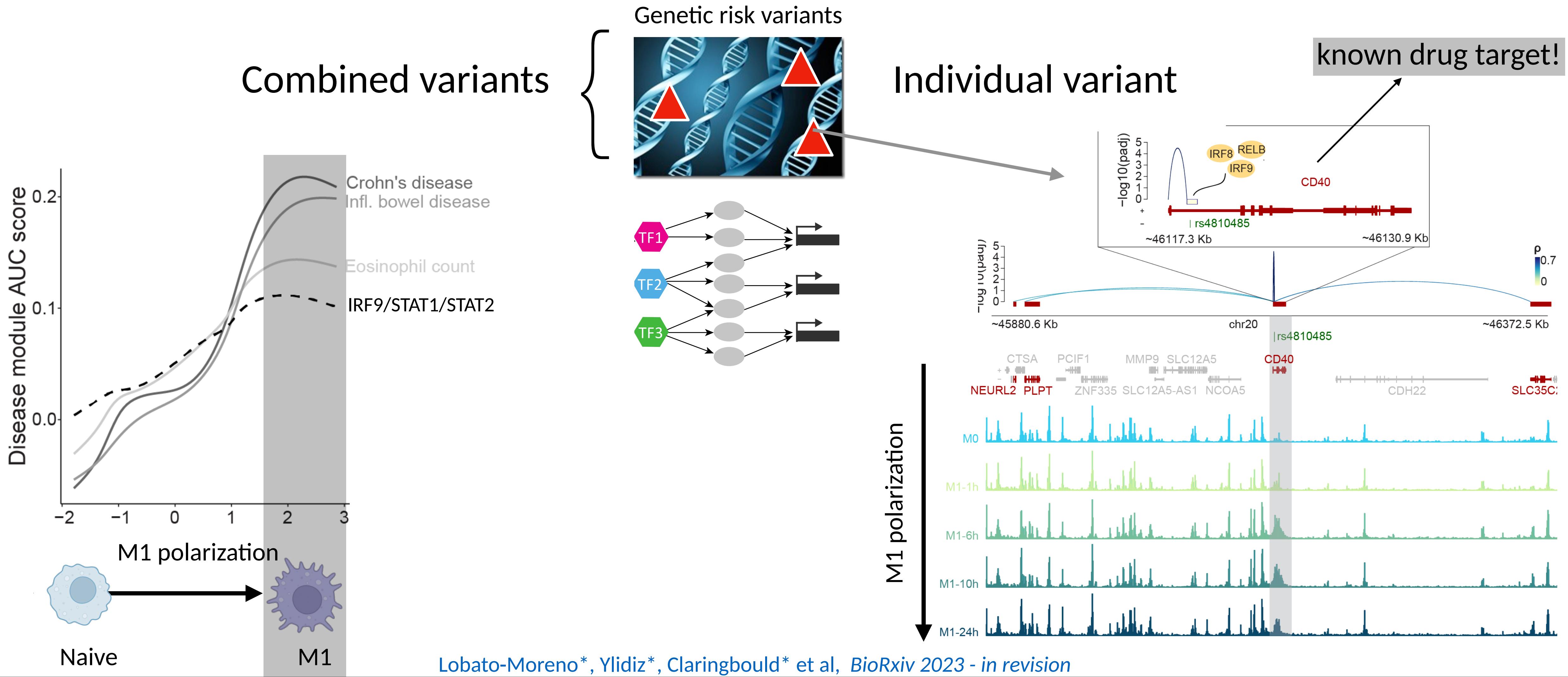
Inferred gene regulatory networks to integrate Genetic risk variants



Genetic risk variants enriched in network - what can we now learn about the mechanism?

Combined genetic variants map to M1 cell state

- individual variants map to genes including known drug targets



Gene regulatory networks can highlight disease mechanisms at cellular and molecular scales

Take-home message

1. Think about what any type of data can tell you in the best case scenario
2. Think about the question you want to ask
3. Use the technology that is best suited to address your question!