

Nextflow pipeline: nf-atacco

Transcription factor activity inference

Yuan Li

Yuan.li@nbis.se

Team meeting: 2026-01-29

Introduction & Goals

➤ Project 8271 (bulk ATAC + bulk RNAseq)

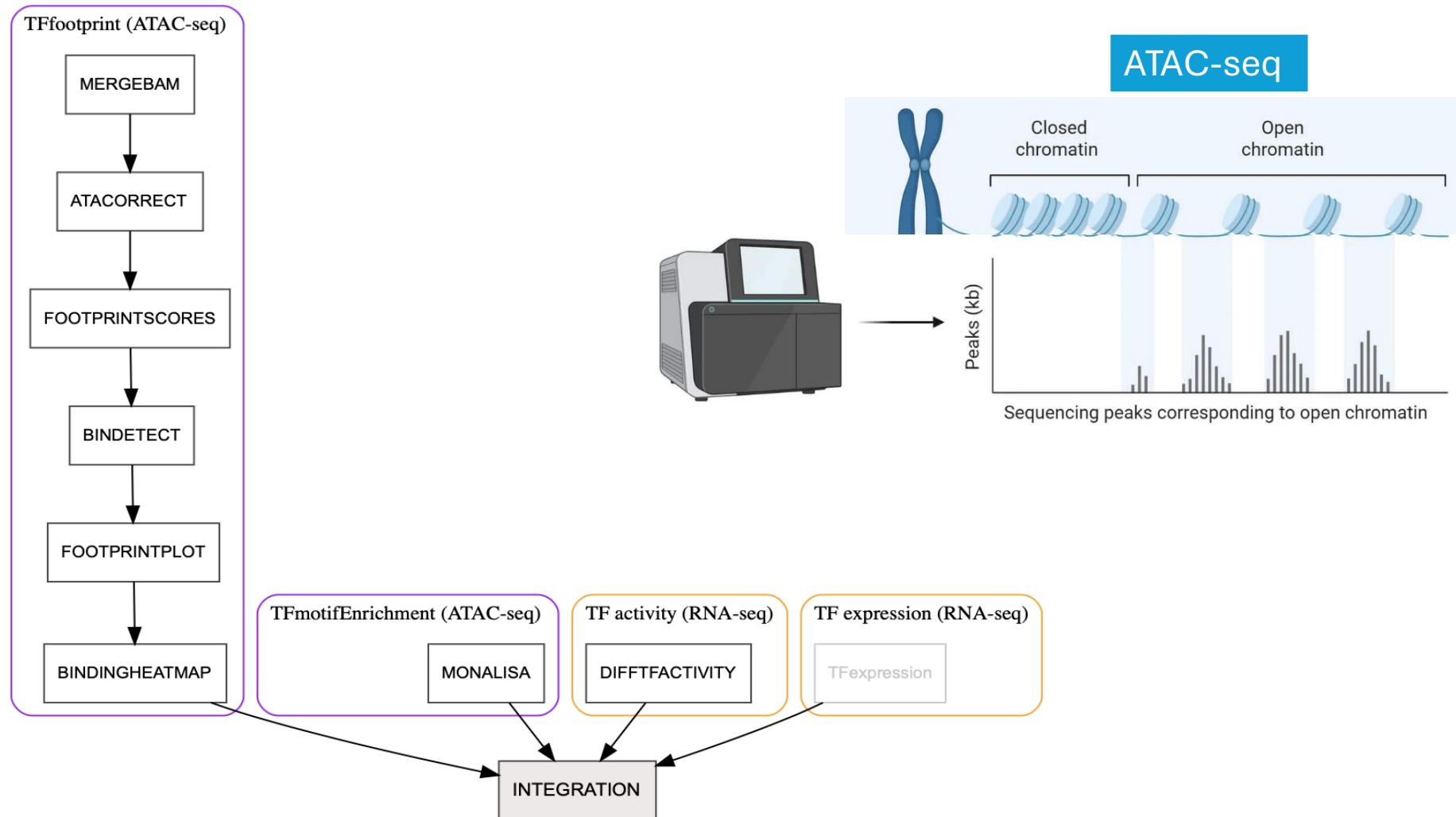
<https://github.com/addityea/nf-atac>

<https://github.com/Maj18/nf-atacco>

➤ Transcription factor/TF activity inference

- Gene expression is tightly regulated
- TFs: key mediators
- TF disorder: diseases

Pipeline Overview (Diagram)

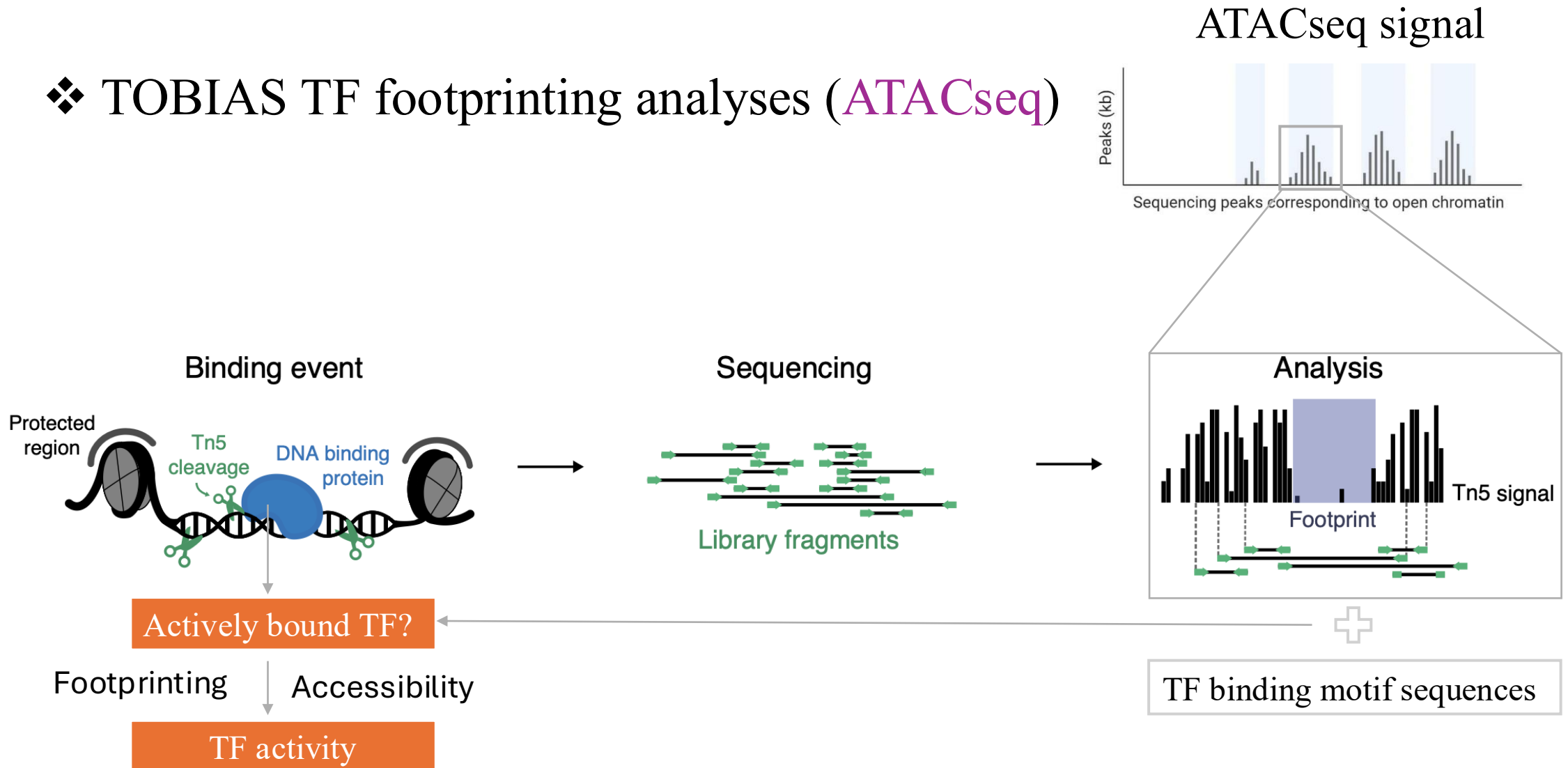


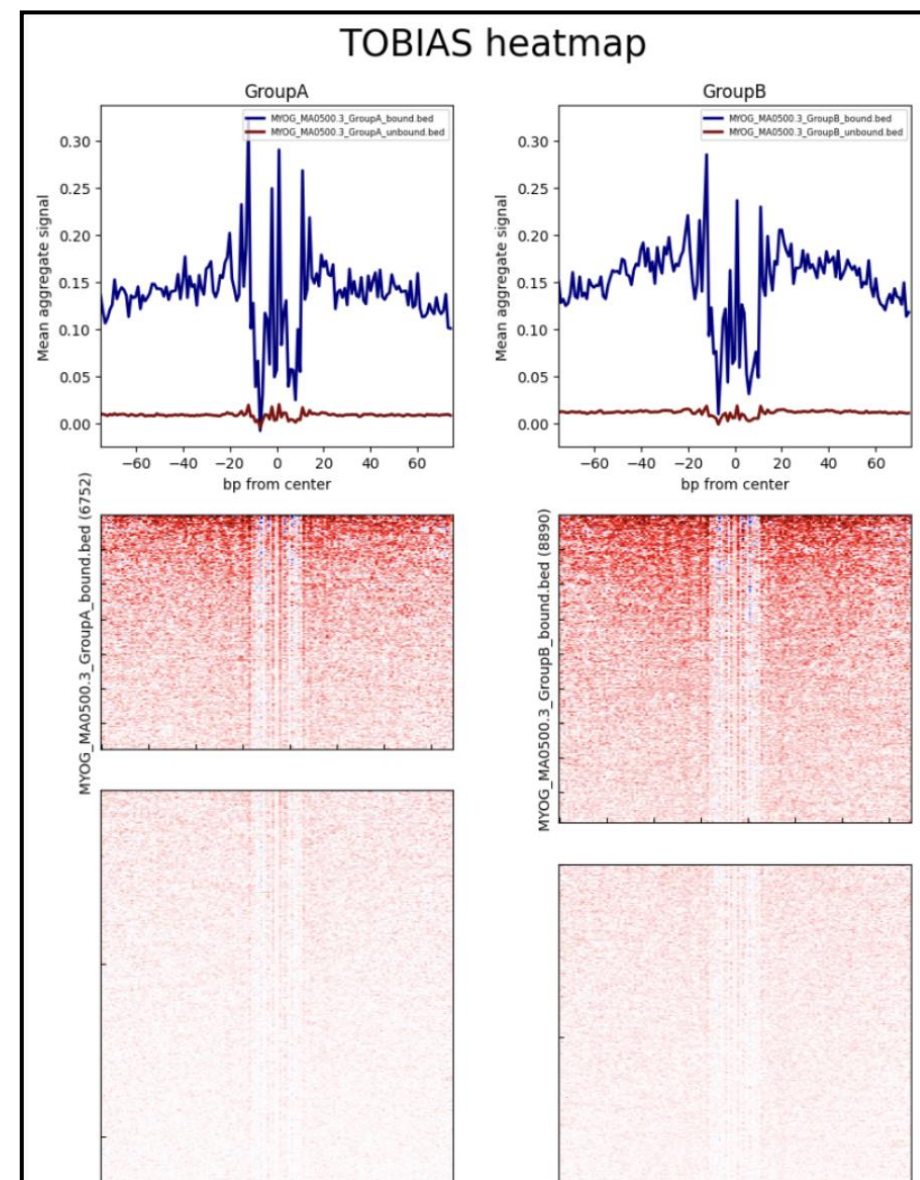
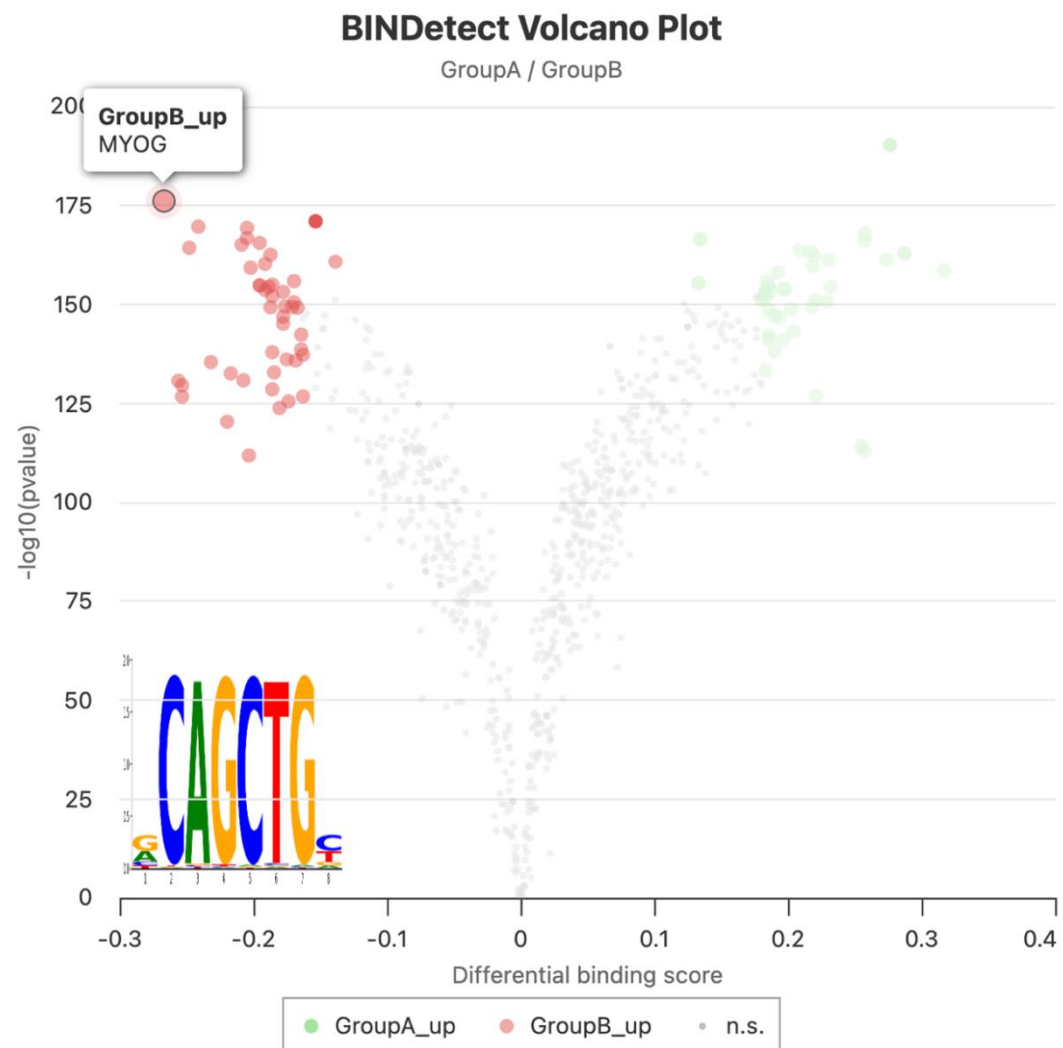
Key Subworkflows & Tools

- ❖ TOBIAS TF footprinting analyses (ATACseq)
 - TF binding footprint + TF motif sequence: infer TF binding
- ❖ MONALISA binned TF motif enrichment analysis (ATACseq)
 - TF motif sequence: infer TF with potential to bind
- ❖ VIPER+Dorothea differential TF activity (RNAseq)
 - TF target expression: infer downstream transcriptional output
- ❖ TF expression (not in the pipeline)
 - TF expression: TF availability
- ❖ Integrative visualization

Key Subworkflows & Tools

❖ TOBIAS TF footprinting analyses (ATACseq)





Key Subworkflows & Tools

MONALISA binned TF motif enrichment analysis (ATACseq)

Differential accessibility analysis



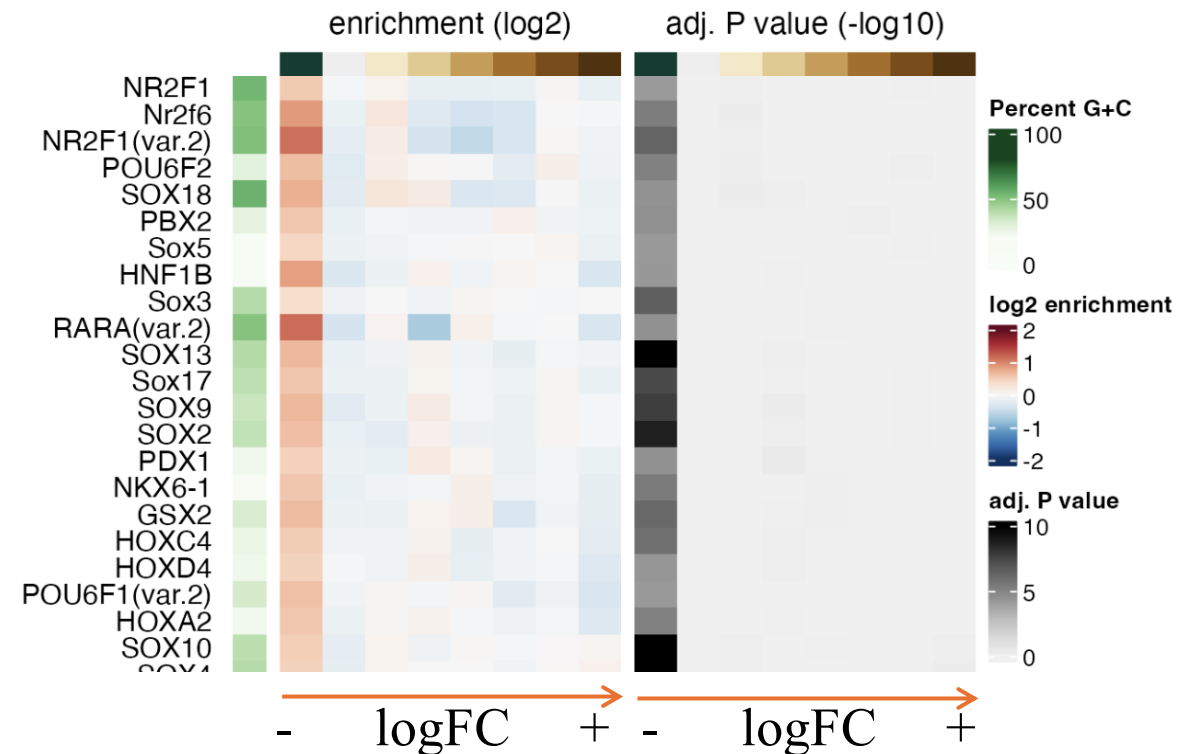
logFC bins (regions=seqs)

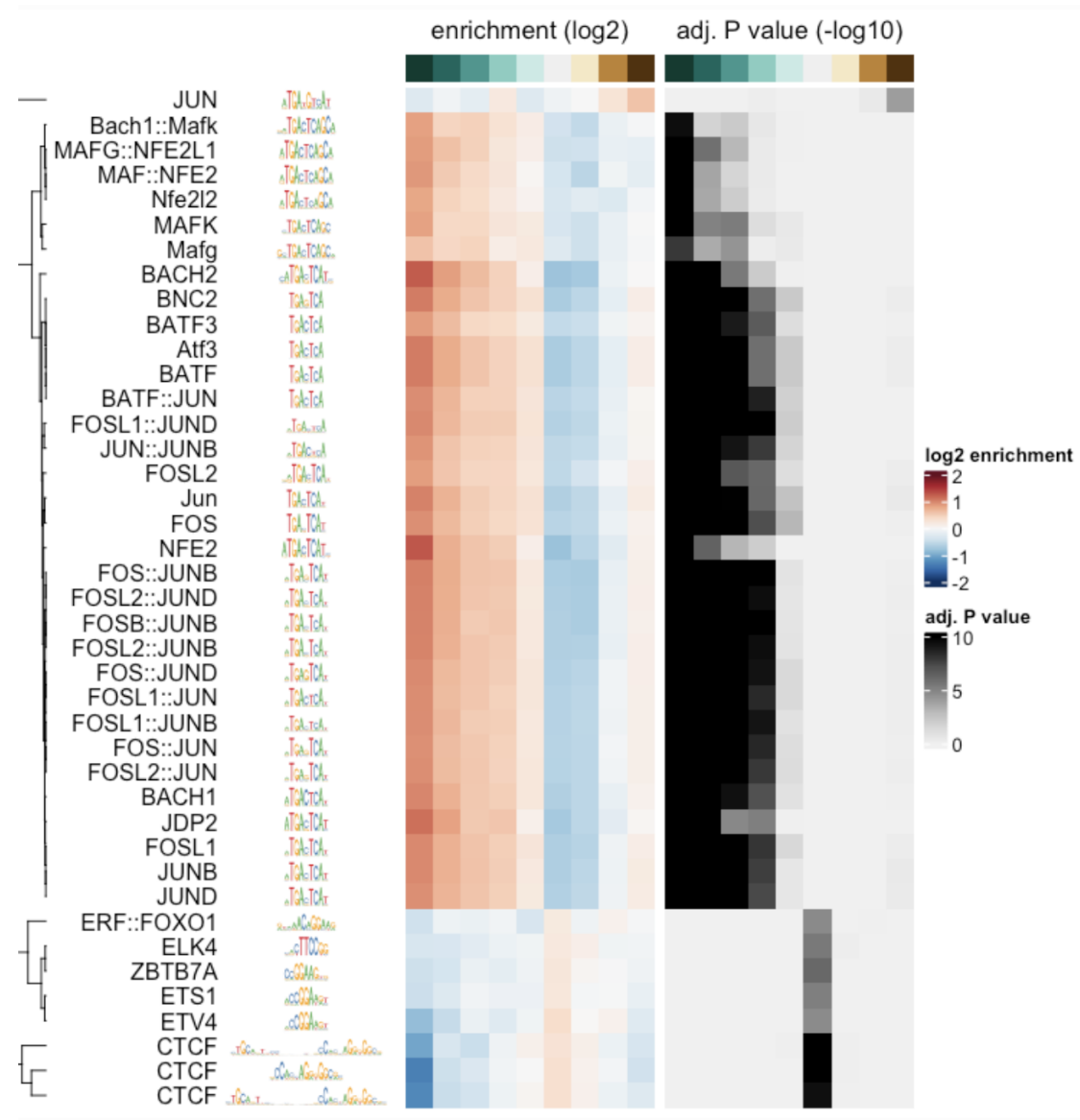


Detect TF motif within bins



Motif enrichment bin vs other

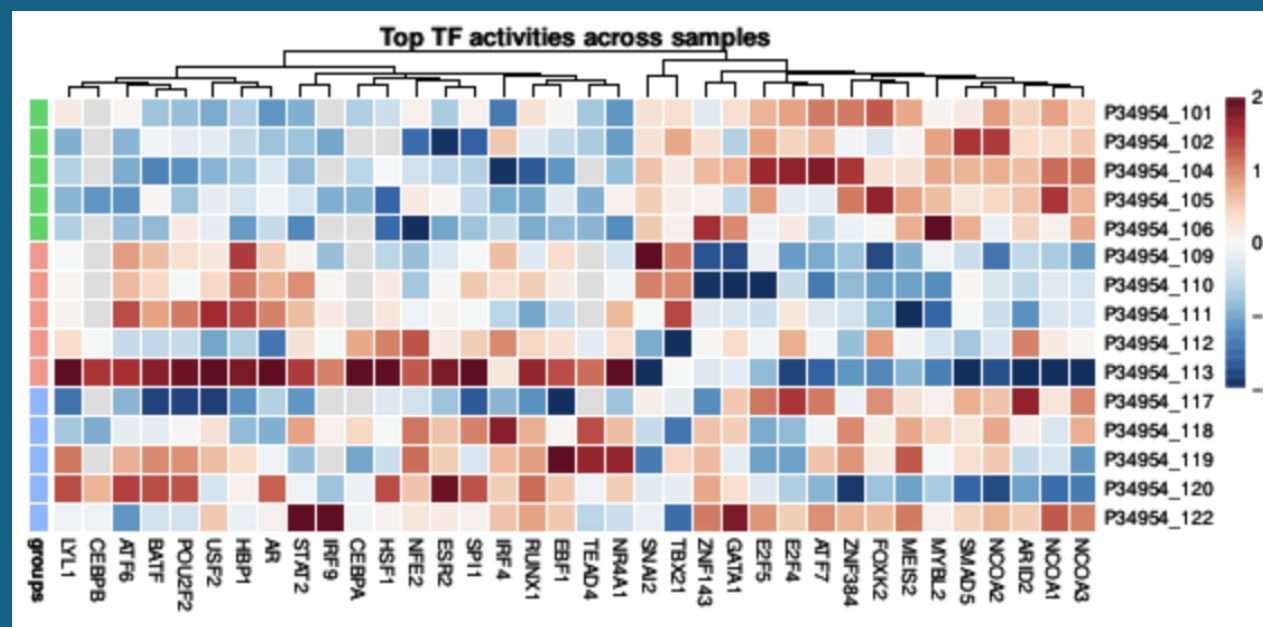
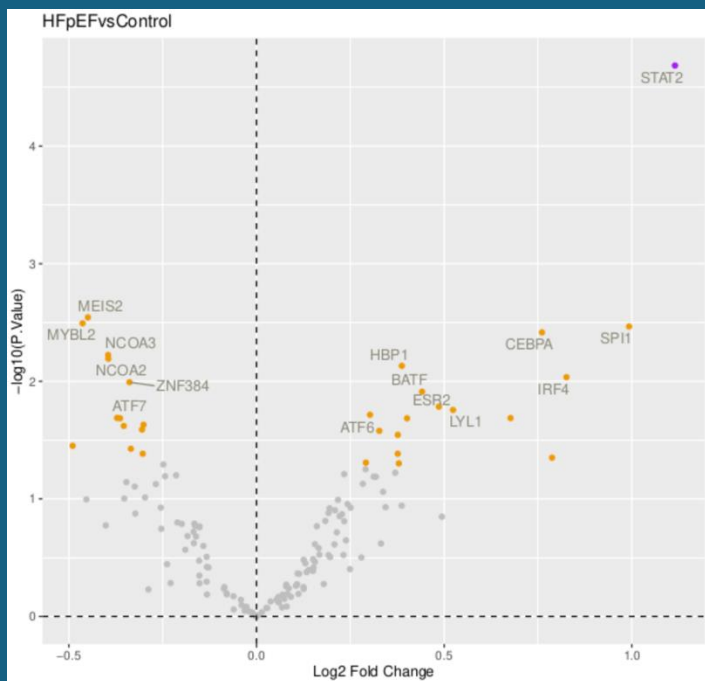




Key Subworkflows & Tools

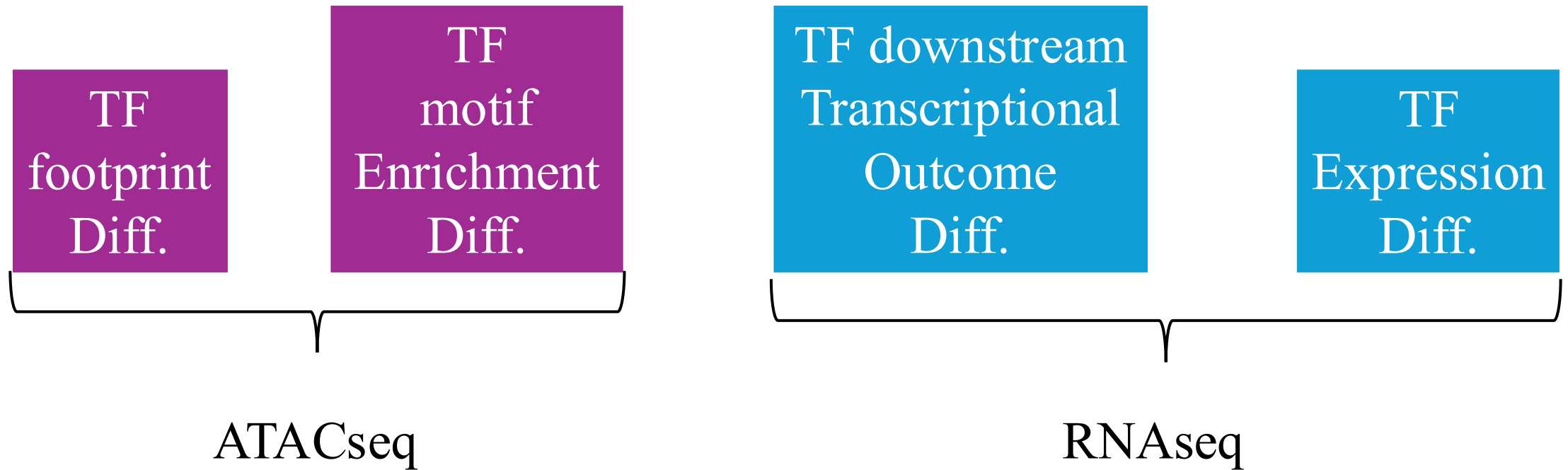
❖ VIPER+Dorothea: downstream transcriptional outcome (RNAseq)

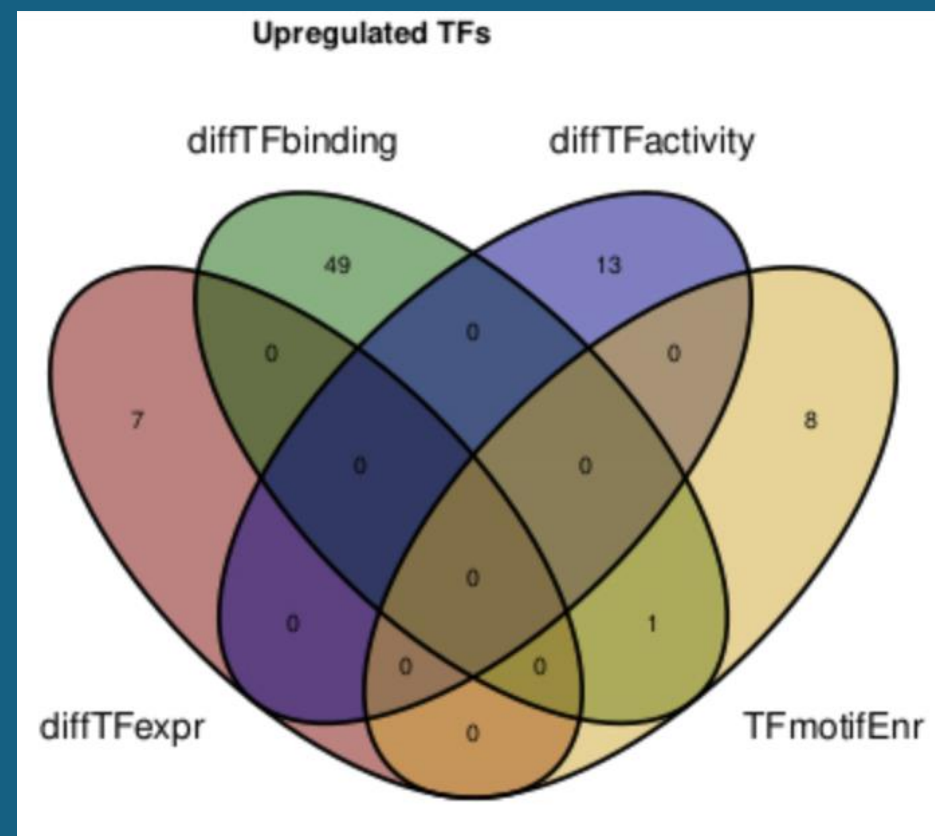
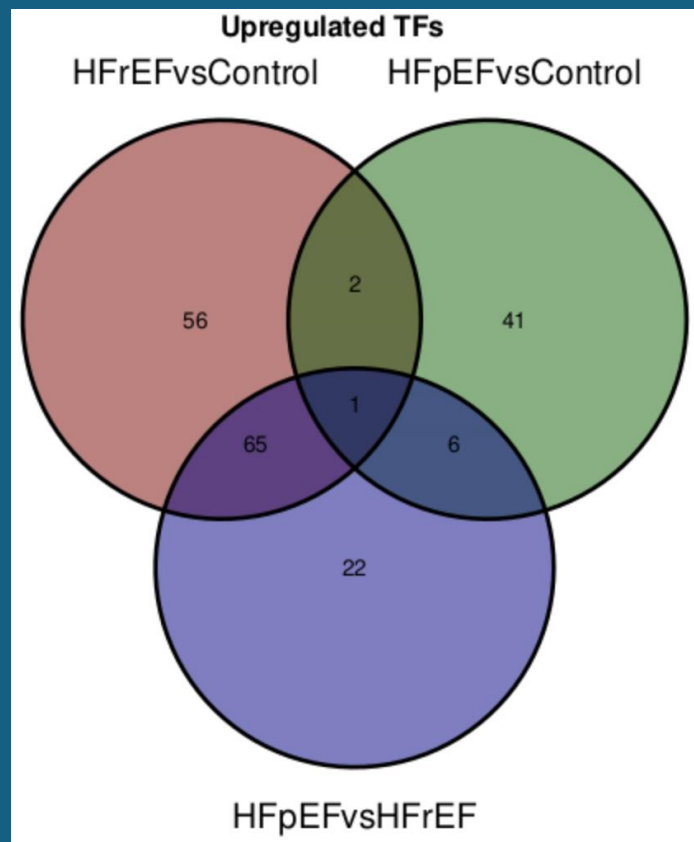
- Dorothea: A curated database of TF-target interactions (confidence: A+B)
- VIPER: Rank-based enrichment scores, weighted by the effect size and the TF-target interaction confidence.
- Limma differential analysis of TF activity scores.

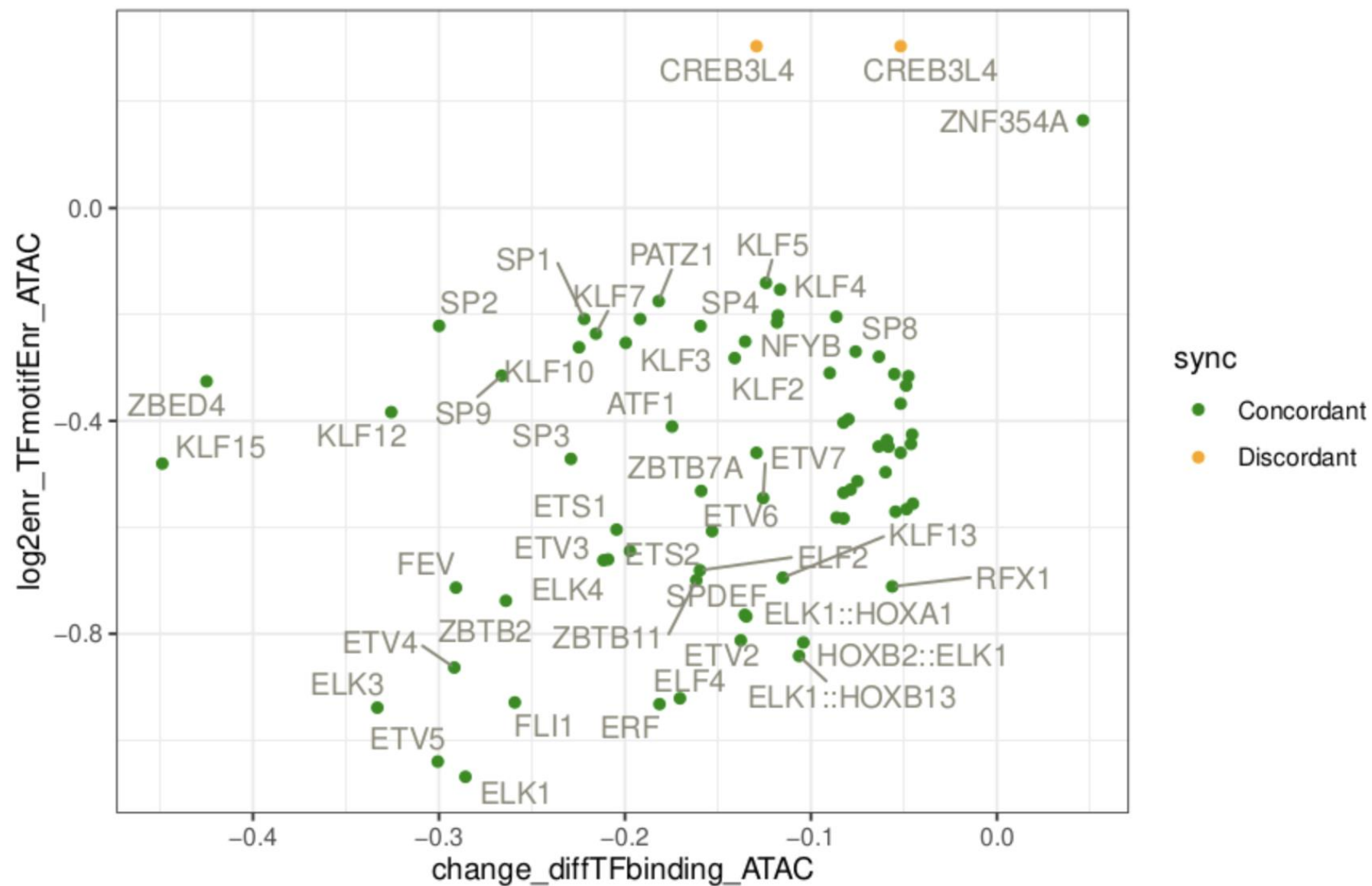


Key Subworkflows & Tools

Integration







Integration interpretation

- TOBIAS TF binding:
 - Reflect DNA occupation, not downstream output
 - Often transient: bind to DNA for short periods of time
 - Continuously associate and dissociate from binding sites
 - Not all active TFs produce strong and detectable footprints
 - Global/aggregated signal
- MONALISA TF motif enrichment:
 - Only based on motif sequences
 - all TFs with potential to bind
 - Global/aggregated signal
- TF expression:
 - TF availability
 - TF mRNA levels often don't correlate well with TF activity
- VIPER TF activity:
 - Downstream transcriptional outcome



- Integrative visualization

Reproducibility & Portability



Containers (singularity, docker)



Bianca/Dardel/Rackham



Version (4 tags)

<https://github.com/Maj18/nf-atacco/tree/main>

Scalability & Resource Management

- ❖ Task (process)-based execution
 - ❖ Parallelization: automatically run multiple tasks in parallel when possible
 - ❖ Resource specification (process, label)
 - ❖ Save processes, -resume
 - ❖ Submit each task as a job to SLURM under the hood
 - ❖ Configuring the executor
 - ❖ ENTER points
-

Configuration priority

Command-line parameters (--param value)



Explicit config files (-c file.conf)



Profile config (inside nextflow.config or .conf)



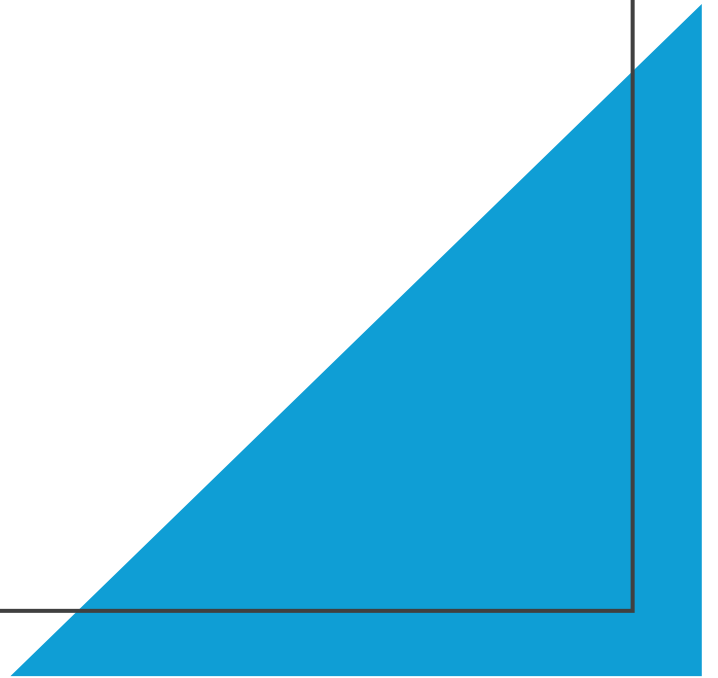
nextflow.config (base/parameter defaults)



Defaults in main.nf (parameters)

Execution Demo

- <https://github.com/Maj18/nf-atacco/tree/main>



Next Steps & Feedback

- Connect all entries
- Provide integration of different TF inference combinations
- Reduce parameters for simplicity
- Add parameters for flexibility

THANK YOU

- Title: nf-atacco: A nextflow pipeline for transcription factor activity inference
- Short description: **nf-atacco** is a bioinformatics pipeline built using **Nextflow** for transcription factor inference by integrating processed ATAC-seq and RNA-seq data through multiple approaches, including TF footprinting, motif enrichment, and downstream transcriptional output analysis. The pipeline also supports integrative visualization of these results, including TF expression. In addition, it provides BAM-to-bedgraph conversion and genome track visualization for ATAC-seq downstream analysis.