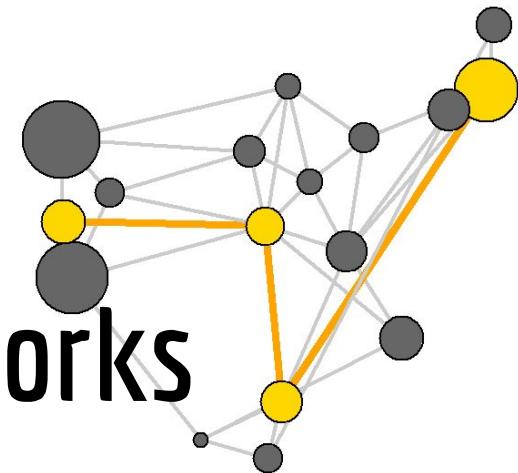


Gene Regulatory Networks in Evo-Devo

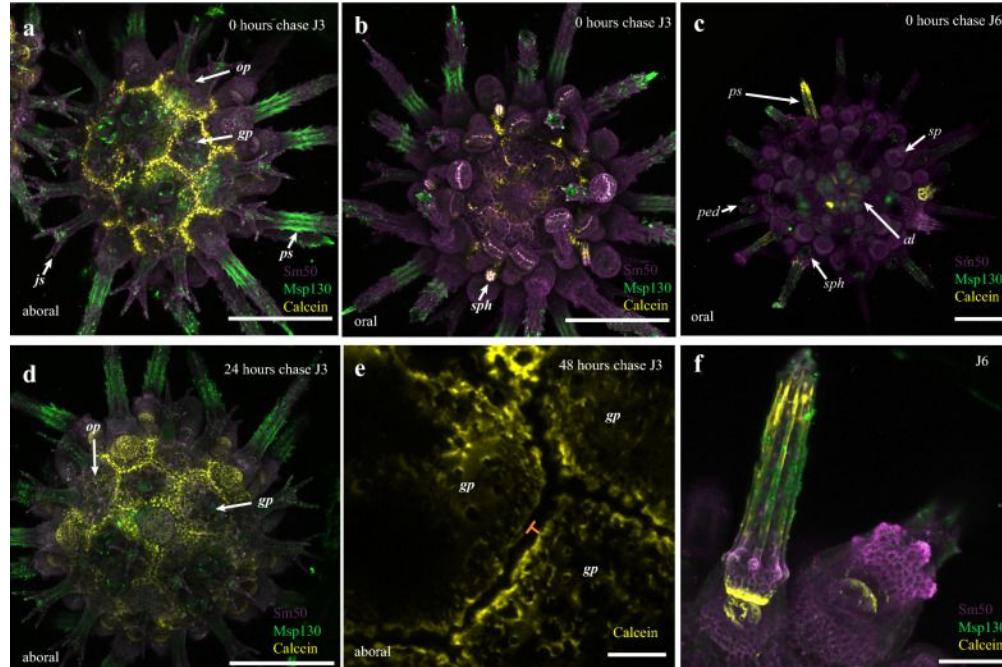


Marta Portela Martínez
UB
2025

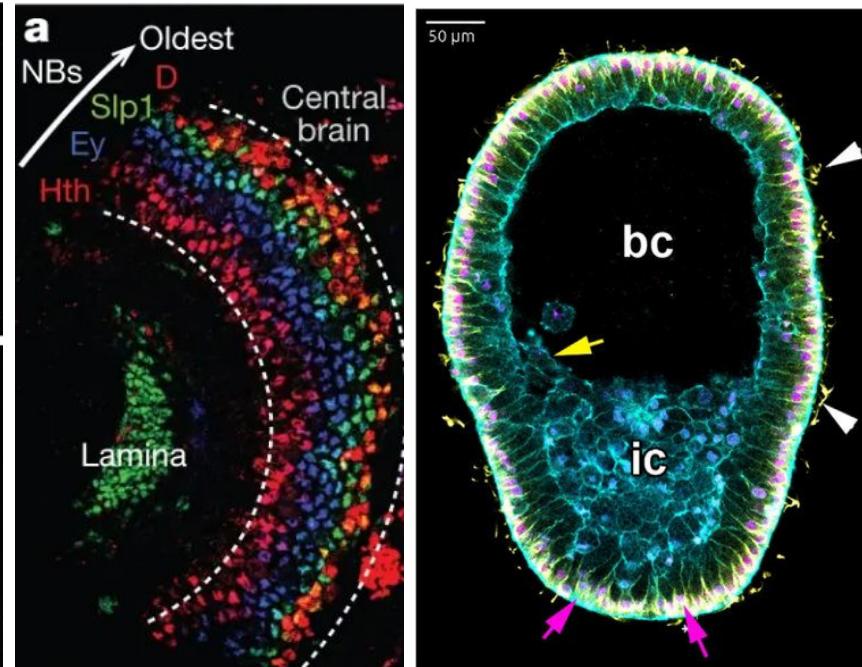
NEW TECHNOLOGIES FOR DEVELOPMENTAL EVOLUTIONARY BIOLOGY STUDIES



Development



Thompson et al (2021)



Li et al (2013)

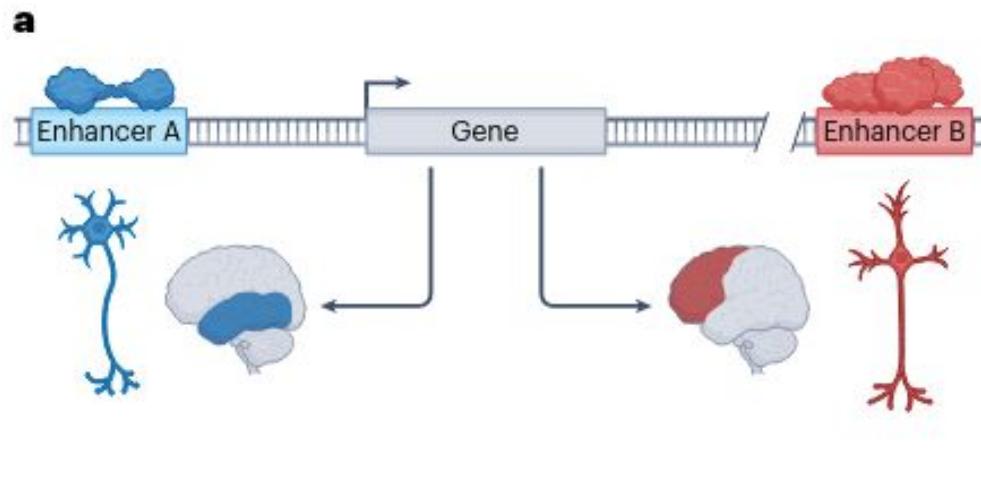
van der Sande et al (2021)

Gene regulation

only one genome → countless possibilities

Gene regulation

Spatio-Temporal *specificity*



Coyle and King, 2025

Historical perspective: 1960

J. Mol. Biol. (1961) **3**, 318–356

REVIEW ARTICLE

Genetic Regulatory Mechanisms in the Synthesis of Proteins †

FRANÇOIS JACOB AND JACQUES MONOD

*Services de Génétique Microbienne et de Biochimie Cellulaire,
Institut Pasteur, Paris*

(Received 28 December 1960)

Gene Regulation for Higher Cells: A Theory

New facts regarding the organization of the genome provide clues to the nature of gene regulation.

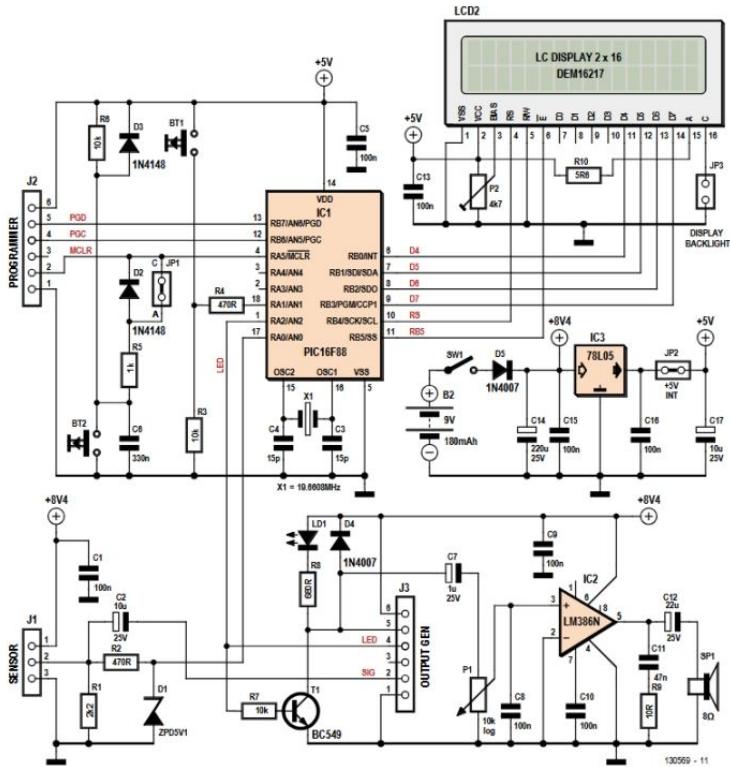
Roy J. Britten and Eric H. Davidson

Sillicon Valley

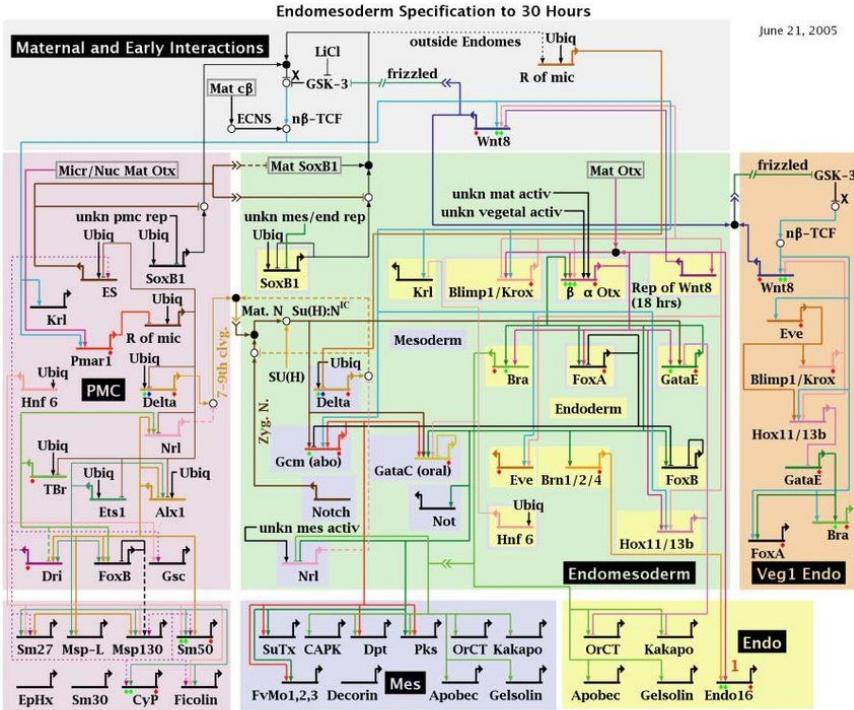


Software engineer mindset

PIC radiation meter



Sea Urchin Endomesoderm Specification



Copyright © 2001–2005, Hamid Bolouri and Eric Davidson.

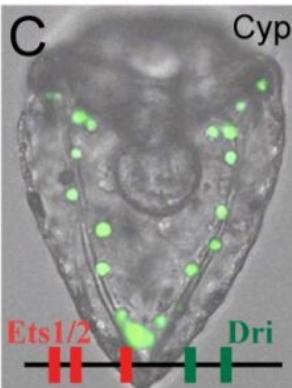
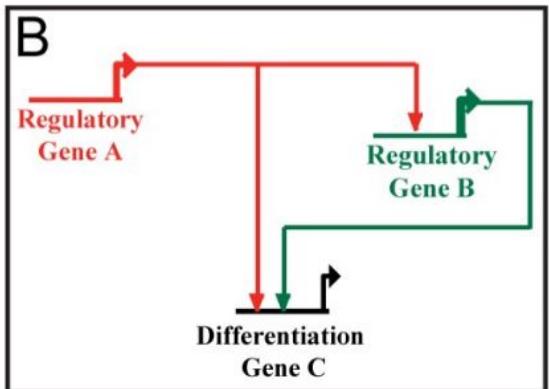
RAND TR310-6.3

Experimental evolution

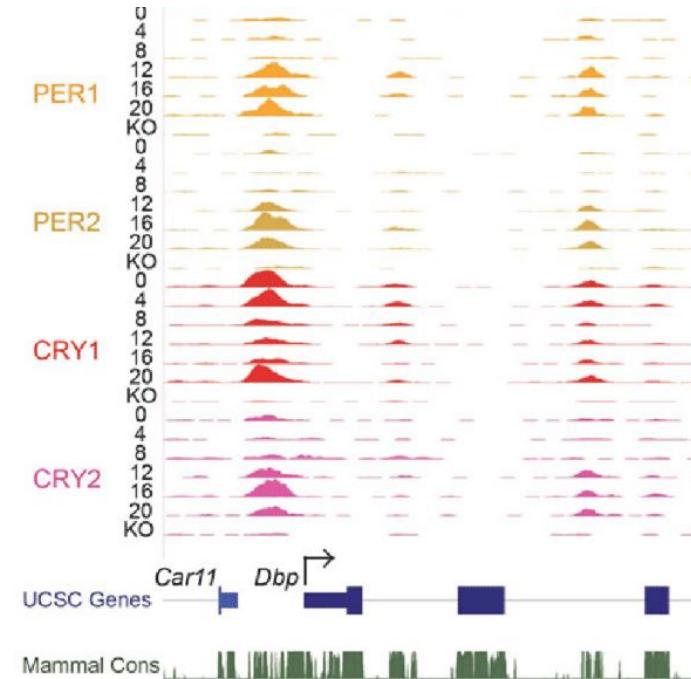
RNA probe → Multiomics

Countless players // Gradual information

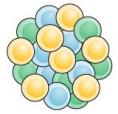
Master regulators // On-Off



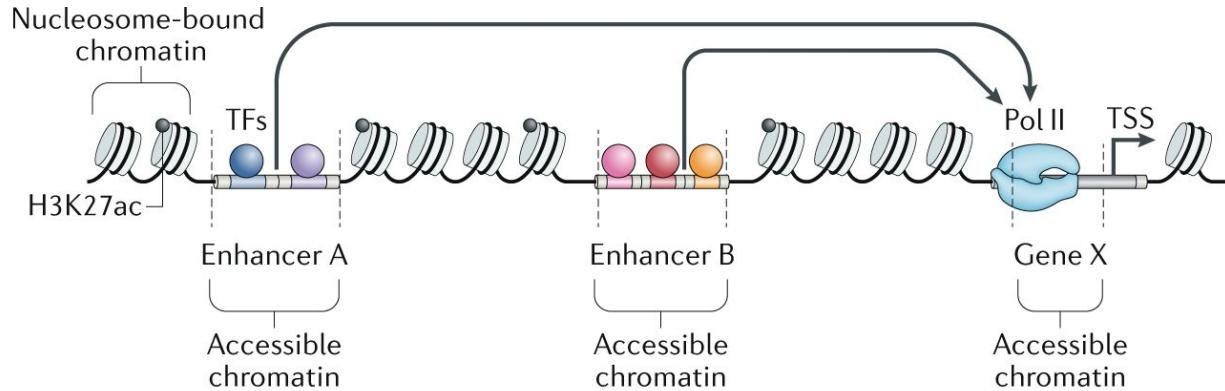
Olivieri, GERN 2025



Koike et al. 2012



Gene regulation



Minoye et al 2021

Chromatin and epigenetic state

- Histone marks (H3K27ac, H3K4me3, etc.)
- DNA methylation
- Chromatin accessibility (ATAC-seq signal)
- Nucleosome positioning
- Polycomb / Trithorax domains

DNA regulatory elements (*cis*)

- Core promoters (TATA box, Inr, BRE, DPE)
- Enhancers
- Insulators / boundary elements
- Locus control regions (LCRs)
- ...

Transcription Factors (seq-specific regulators)

- Activator
- Repressors
- Pioneer
- ...

Cofactors and Chromatin regulators

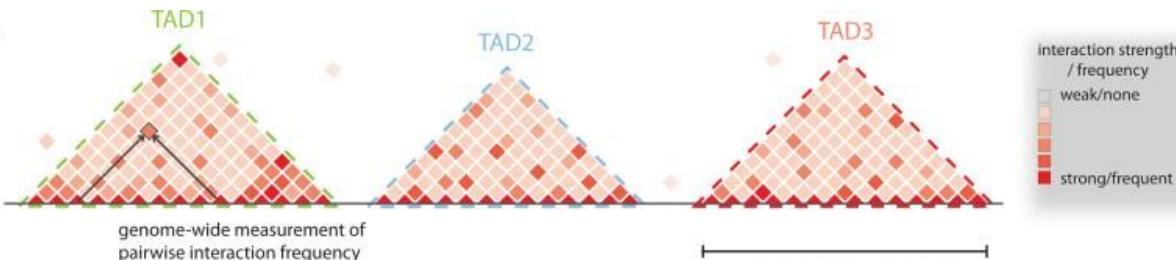
- Co-activators
- Co-repressors
- Chromatin remodelers
- Histone modifiers
- Histone variants

RNA-based regulatory elements

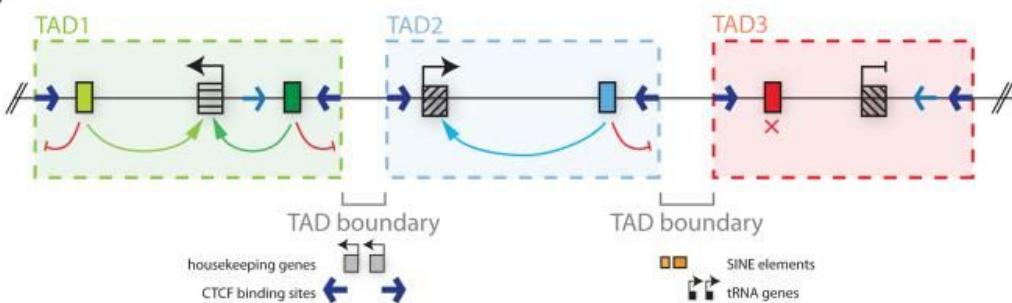
Structural proteins
 CTCF
 Cohesin
 WAPL
 YY1
 ...

3D gene regulation

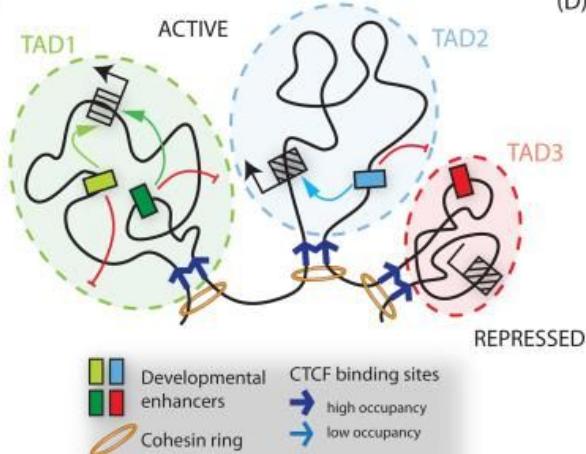
(A)



(B)



(C)

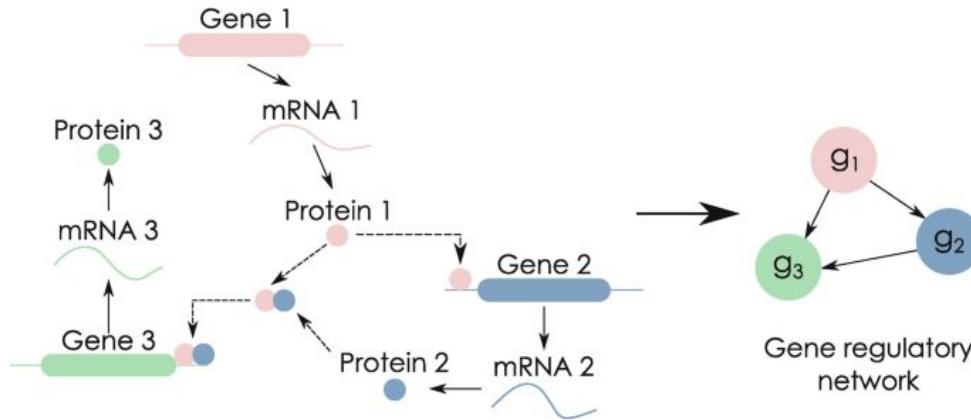


(D)

Gene Regulatory Networks integrate information

GRN

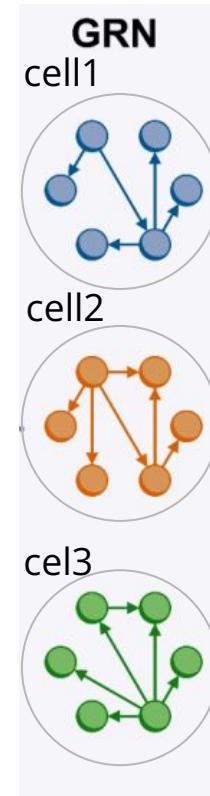
Abstract collection of regulatory interactions between genes



Huynh-Thu et al. 2019

Novel biological insights from GRN

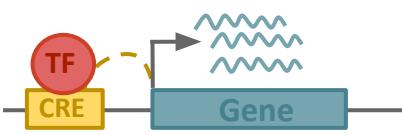
- **Topological analysis:** identification of relevant players, effectors...
- TF's **regulon** behaviour
- **Comparison** across cell types, conditions, species→multiple perspectives
- Dimensionality reduction



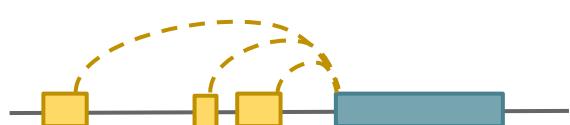
Today's plan

1. CRE-gene association
2. GRN based on RNAseq expression
3. TF-CREs association
4. Data integration in a “multi-supported” GRN
5. Visualization

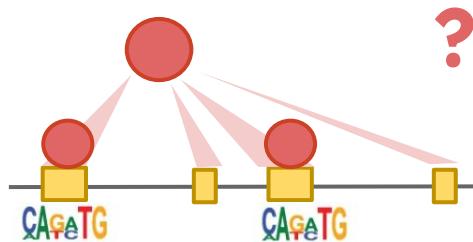
GRN



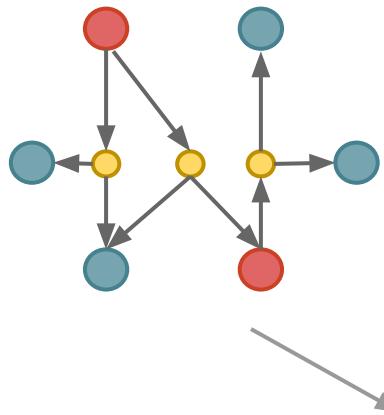
1. CRE-gene association



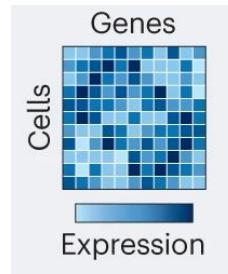
3. TF - CRE association



3. TF-CRE-Gene Network

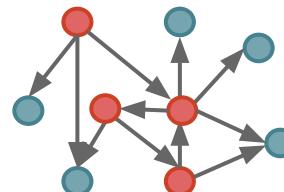


2. GRN based on RNAseq expression



Tree ensemble

4. Data integration

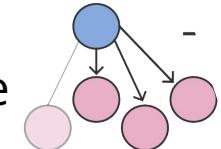


Regulon:
TF + targets

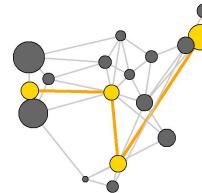


Final outputs from a GRN

1. **Regulons** (TF + target genes) → Regulome



2. **Topology**



A way for BigData exploitation

GRN modularity

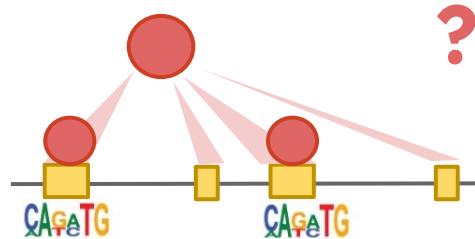


GRN

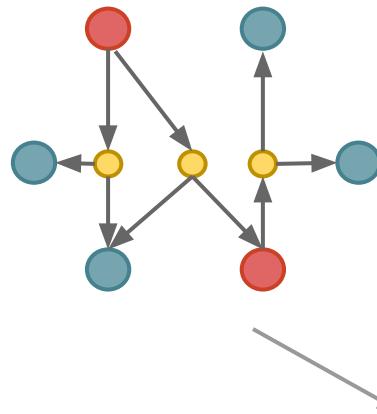
1. Cis-regulome



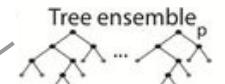
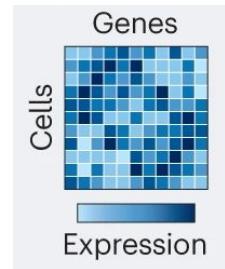
2. TF - CRE association



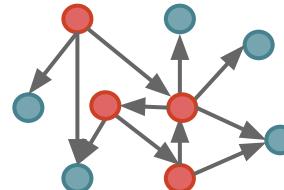
3. TF-CRE-Gene Network



4. Coexpression modules



5. Gene Regulatory Network



Regulon:
TF + targets



Good reviews on GRN inference

Review article

Check for updates

Gene regulatory network inference in the era of single-cell multi-omics

Pau Badia-i-Mompel , Lorna Wessels , Sophia Müller-Dott , Rémi Trimbour , Ricardo O. Ramirez Flores , Ricard Argelaguet , & Julio Saez-Rodriguez

npg systems biology and applications

www.nature.com/npgsba

REVIEW ARTICLE

OPEN

Check for updates

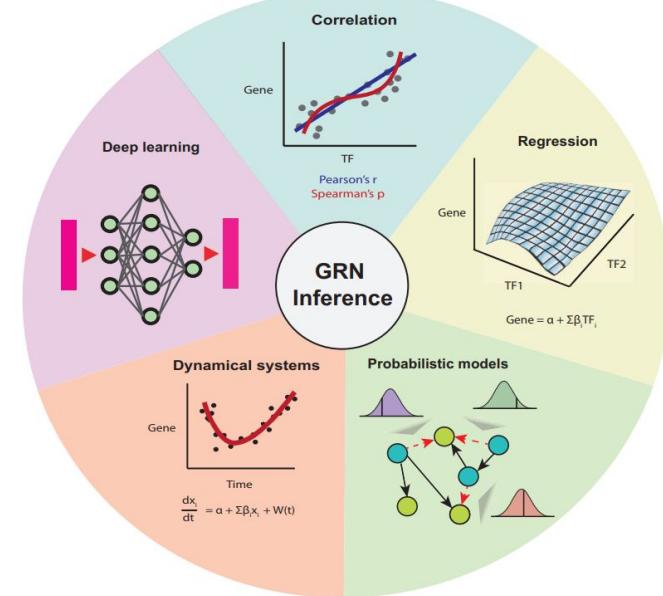
Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data

Daniel Kim ^{1,2,3,5}, Andy Tran ^{1,3,4,5}, Hani Jeun Kim ^{2,3}, Yingxin Lin ^{1,3,4}, Jean Yee Hwa Yang ^{1,3,4} and Pengyi Yang ^{1,2,3,4}

Comparison and evaluation of methods to infer gene regulatory networks from multimodal single-cell data

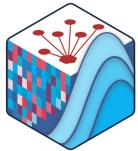
Pau Badia-i-Mompel, Roger Casals-Franch, Lorna Wessels, Sophia Müller-Dott, Rémi Trimbour, Yunxiao Yang, Ricardo O. Ramirez Flores, Julio Saez-Rodriguez

doi: <https://doi.org/10.1101/2024.12.20.629764>

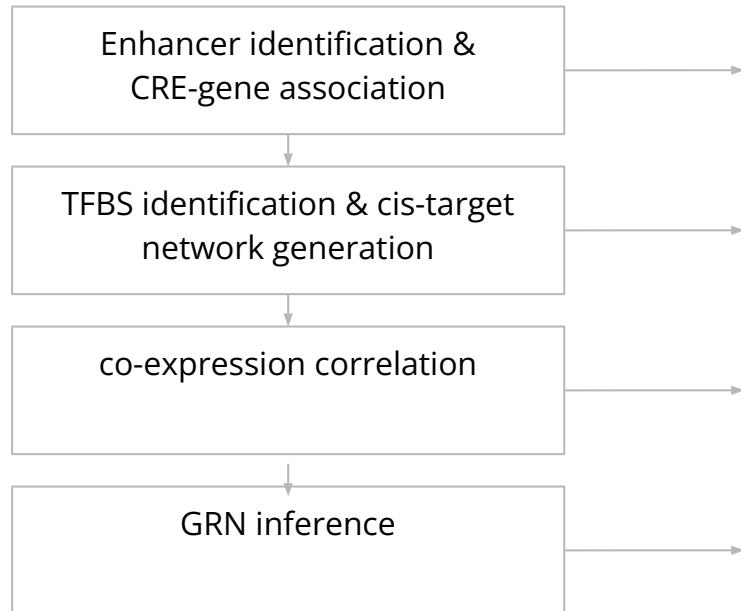


Method		Prog.	Cell type	Metacell	Gene Regulatory Network			Inputs		
					TF-CRE	CRE-Gene	TF-Gene	RNA	ATAC	Paired
Correlation	scMEGA				Motif enrichment	Pearson's correlation	Pearson's correlation			
	FigR				Motif enrichment Spearman's correlation	Spearman's correlation	Spearman's correlation			
	STREAM				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				

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			
D.L.	SCENIC+				Motif enrichment	Gradient boosting	Gradient boosting			
	scMTNI				Motif enrichment	N/A	Bayesian inference			
	Dictys				Motif enrichment	N/A	Stochastic diff. eq.			
	DeepMAPS				Motif enrichment	Graph autoencoder	Regulon construction			
	MTLRank				TF activity score		Multilayer neural network			
	LINGER				Motif enrichment Pearson's correlation	Multilayer neural network				



SCENIC+
Single-cell enhancer-gene
regulatory networks



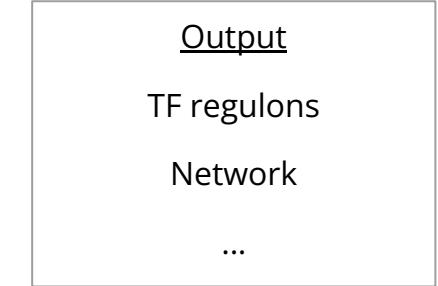
SCENIC+

MACs2, LDA,
GREAT

cluster-buster

GENIE3

RcisTarget +AUCell



Why GRN are not to be run in R



Speed & Scalability

Speed advantage for intensive calculations when **data fits in RAM**.

Modularity & Development

Primarily designed as a **functional language**; modularity is achieved through **packages**.

Interoperability

Strong **specialized ecosystem** (statistics, Bioconductor). Efficient interoperability with Python and web deployment.

Highly efficient for vectorized and matrix operations, with **greater inherent scalability**.

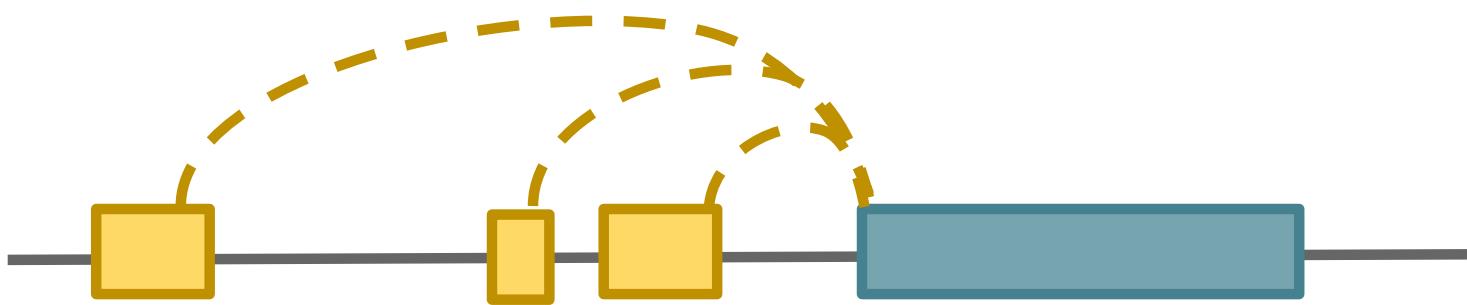
Object-Oriented, which encourages the development of **modular**, reusable code and **software integration**.

Leading in interoperability with systems (Linux, Windows), languages (C, Java), and production/web frameworks.

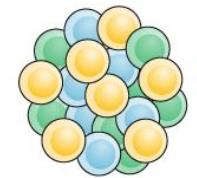
Today's plan

1. CRE-gene association
2. GRN based on RNAseq expression
3. TF-CREs association
4. Data integration in a “multi-supported” GRN
5. Visualization

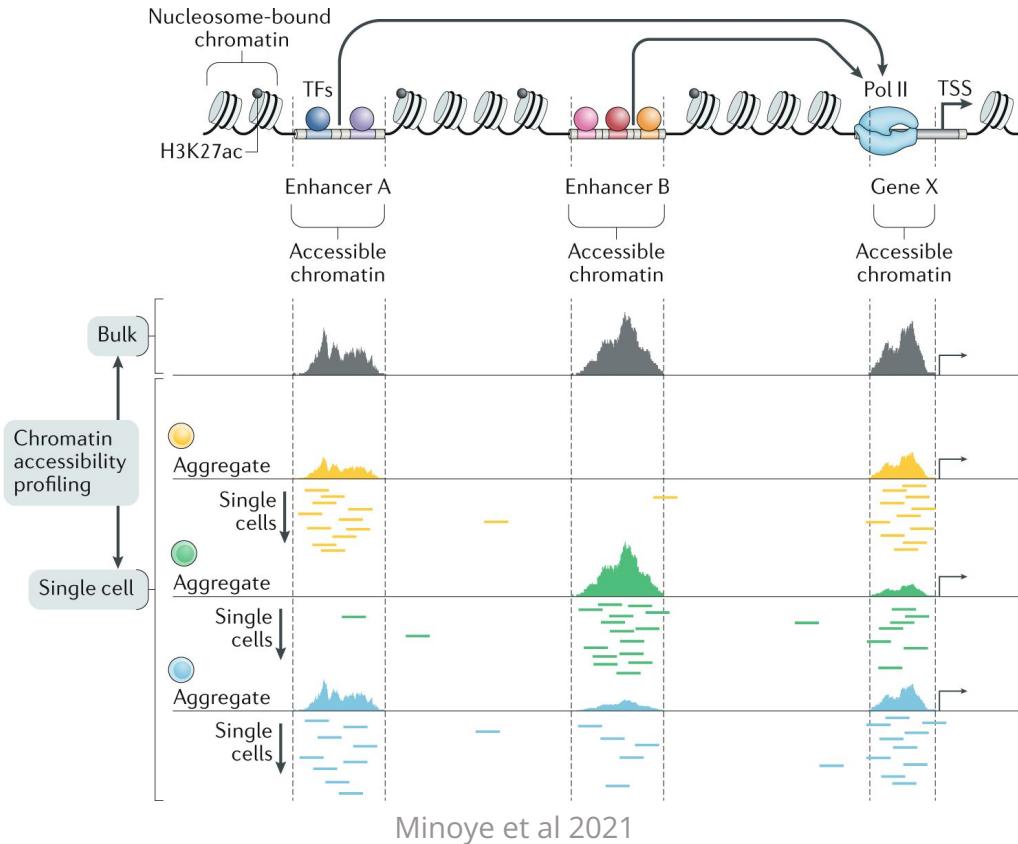
1. CRE-gene association



cis-regulation



Heterogeneous sample



Accessibility

ATACseq

**Protein
binding**

ChIP /
Cut&Run

Analysis of TF binding / ATAC data

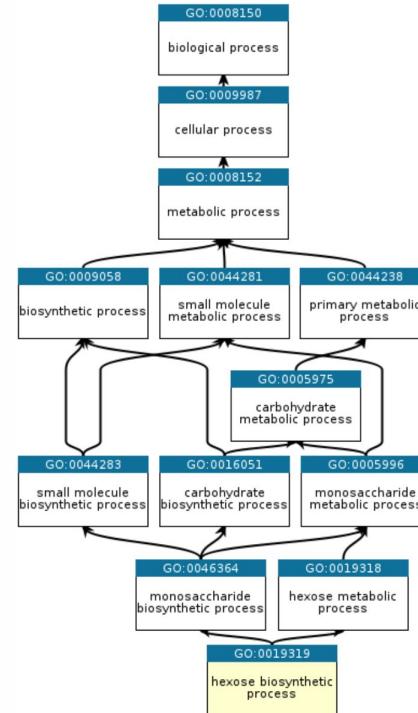
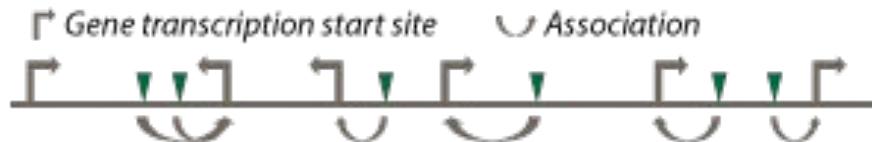
- Peak calling MACS2, SICER, SEACR
- Peak annotation ChIPseeker
- motif discovery MEME-ChIP, DREME, STREME, GimmeMotifs
- Co-binding/co-association bedtools intersect, deeptools
- Differential binding DESeq2, bedtools intersect
- Conservation across species liftOver, phastCons
- **Gene association** GREAT, ChIPseeker, ABC model



Gene Ontology

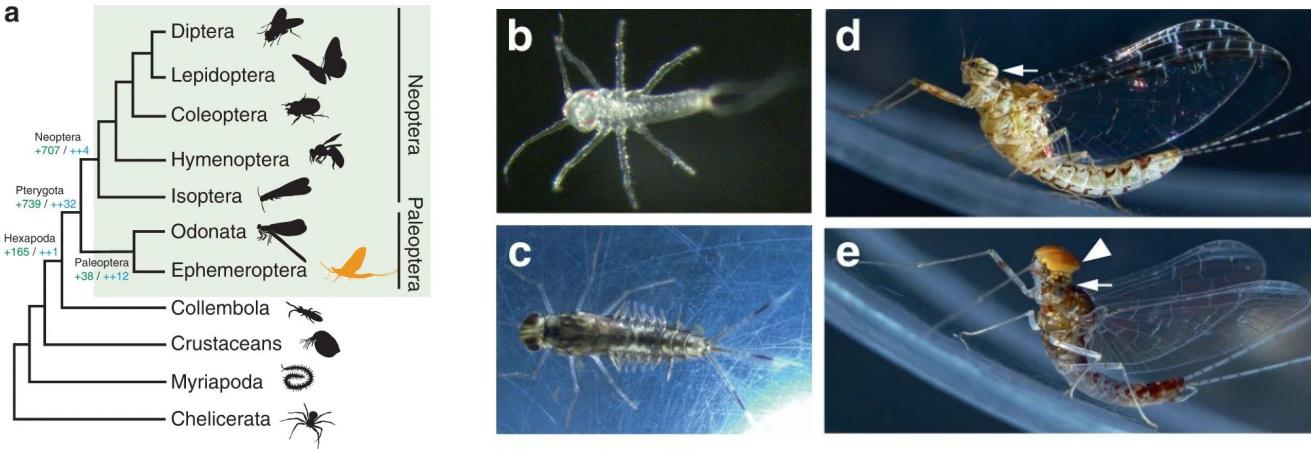
CRE-gene association

Example: SRF ChIP-Seq called peaks

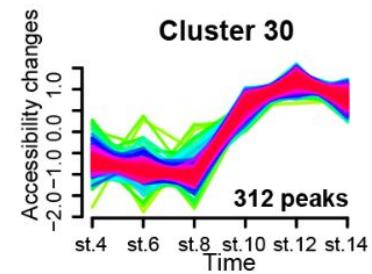
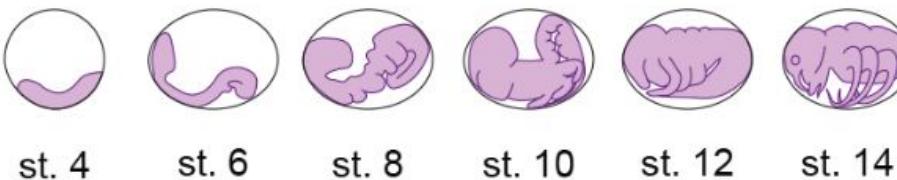




Mayfly development



DOI: 10.1038/s41467-020-16284-8



Why mayfly?

Because it is a ...

**non-model
organism**



GitHub and...get started!

Scripts: https://github.com/mportela95/EvoDevo_2025

Data: <https://drive.google.com/drive/folders/1ZvYrkHUGwts61ZRXEdSUNnvrk3BwQr0X?usp=sharing>

All data is available at: <https://zenodo.org/records/17350362>

Explore the data with the UCSC browser (*C. dipterum*: GCA_902829235.1)

cis-regulation through GREAT

Marta Portela

2025-11-14

1. GREAT

GREAT (Genomic Regions Enrichment of Annotations Tool) is a type of functional enrichment analysis directly performed on genomic regions. This tool allows to predict **functions of cis-regulatory regions** through a linear **association** of regions (potential CREs) to genes and a later **gene ontology** analysis.

Scientific background In this exercise, we will use GREAT to explore the regulatory landscape of late embryogenesis in the mayfly *Cloeon dipterum*, a **non-model insect** that has emerged as an informative system from evolutionary, regulatory and developmental perspectives.

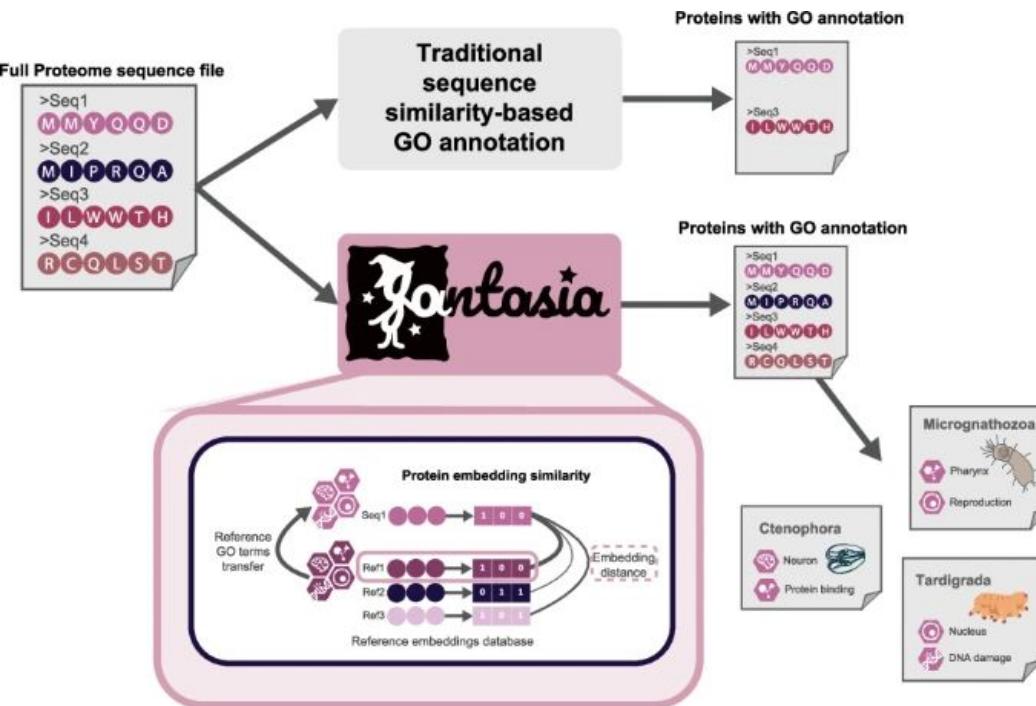
Late embryogenesis is a period marked by the activation of gene regulatory programmes that shape key insect traits, including tissue differentiation and the final establishment of body plan features. By analysing **ATAC-seq** clusters that open and remain accessible during these stages, we capture regulatory regions likely involved in driving these developmental transitions. This provides an ideal context for applying GREAT, as it allows us to infer the biological processes potentially controlled by cis-regulatory elements during a **morphogenetically dynamic period**.

We will take advantage of a previous **M-fuzz** analysis performed over the ATAC peaks in development and choose the ATAC regions that get opened and maintained in late-embryonic development (<https://doi.org/10.1242/dev.203017>).



Figure 1: Mayfly embryonic development

GO in nm-organisms



- **Protein sequence → embedding**

- Uses pre-trained protein language models (e.g. ProtT5, ESM2)
- Each protein is converted into a high-dimensional numerical vector (embedding)

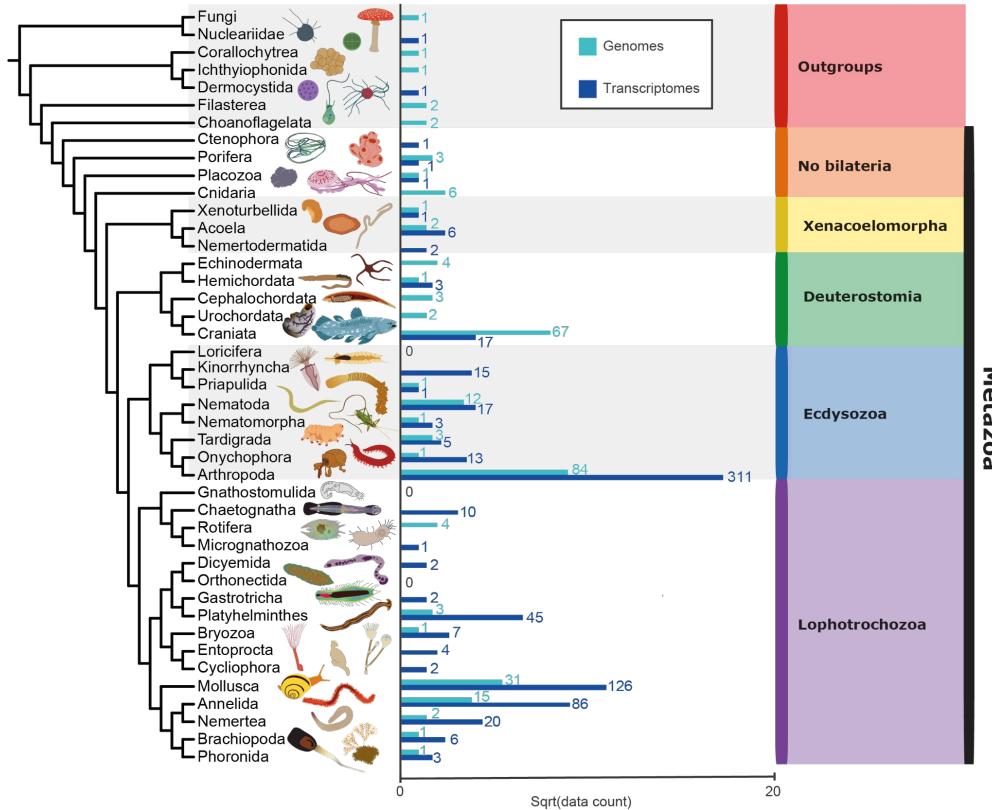
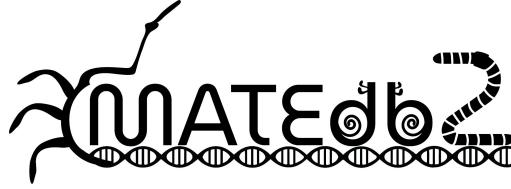
- **Embedding-based similarity search**

- Compares query to a reference set of proteins with known functions
- Similarity measured in **embedding space** (not alignment-based):

- **Function transfer**

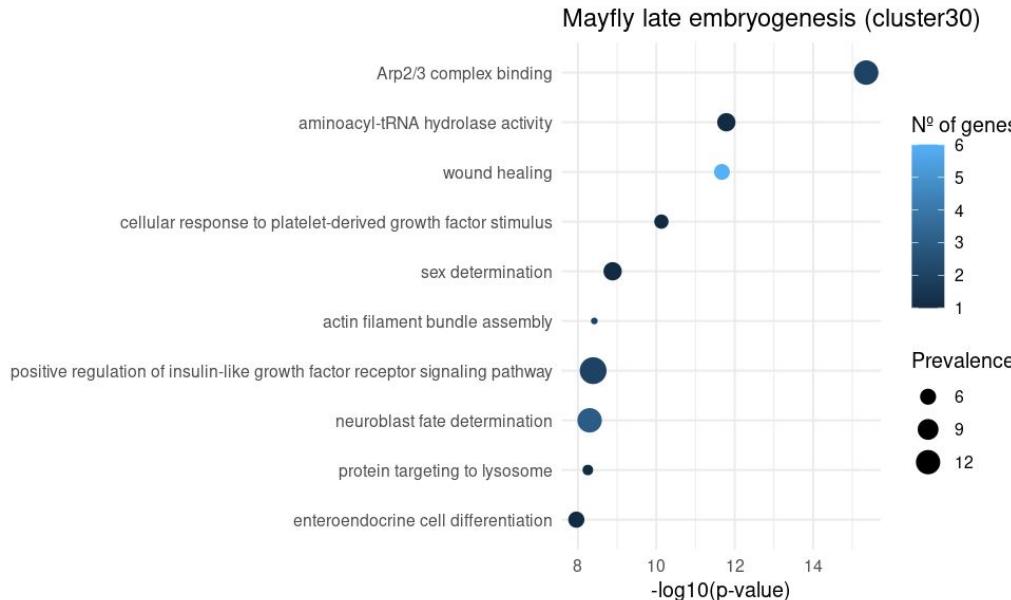
- Assigns **Gene Ontology (GO) terms** based on nearest neighbors
- Includes confidence scores for predicted annotations

Existing GO db



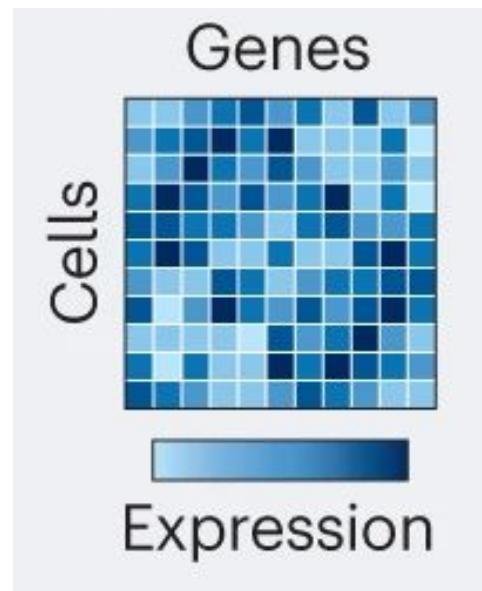


Results & Questions



1. Does it make sense to use this with all the peaks from an ATACseq file? What will we see?
2. What biological signal do the enriched terms reflect?
3. What limitations of GREAT you can identify?
4. How would the results change if:
 - you used a promoter-only assignment (no extension)
 - you used 1 Mb assignment
 - you used gene deserts / intergenic peaks?

2. GRN from expression data



GRN from expression data

Explanation of the observed variability of gene expression based on the expression of other genes.

Weighted gene co-expression network analysis (WGCNA)

- pairwise correlations across the whole transcriptome
- identify modules of co-expressed genes
- False positives

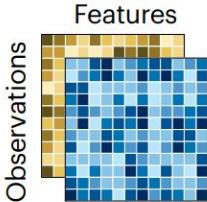
GENIE3 or GRNboost2

- distinguish TFs from target genes (reported activity): reduction of interactions
- random forest inference method
- causal relationships (directional)

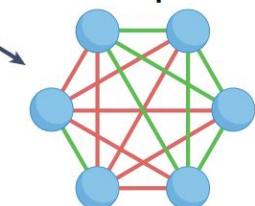


Which TF are relevant for predicting each gene expression

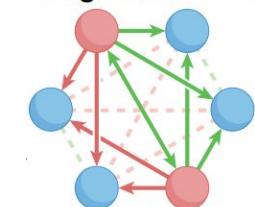
Observable data



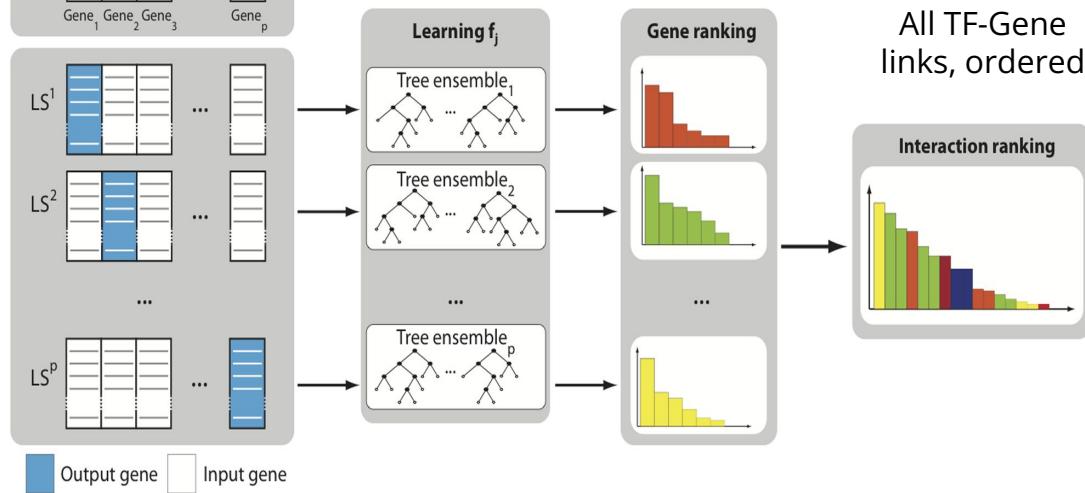
Gene coexpression



Assignment of TFs

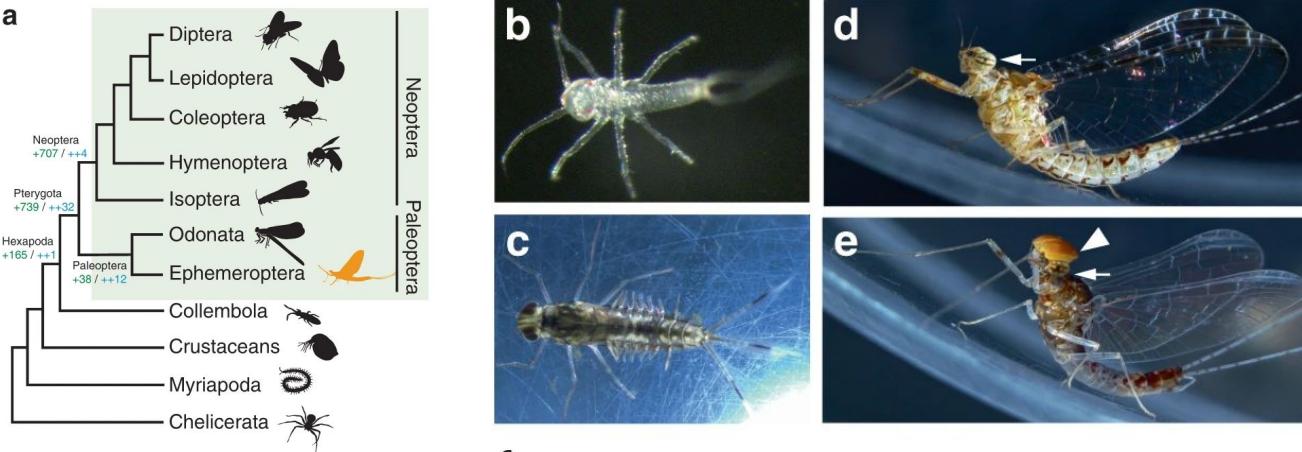


Each Gene links to TF

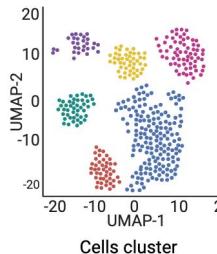




Mayfly development



Almudi et al. 2020

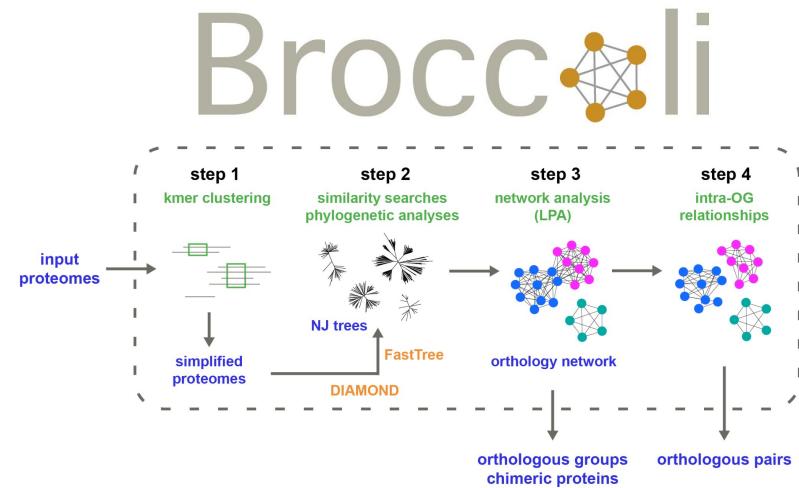
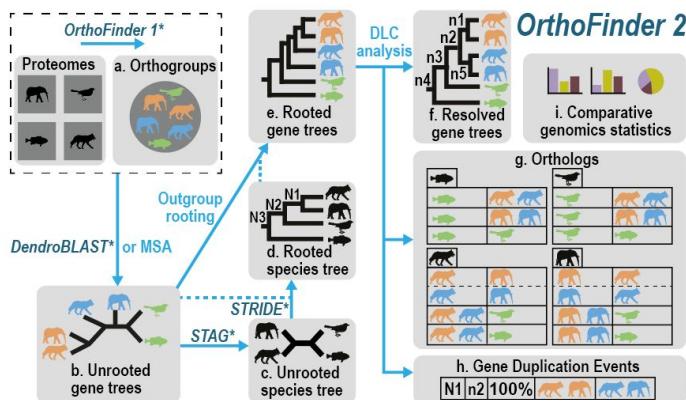


scRNAseq



TF list: Orthology

Orthogroups represent the **complete set of genes descended** from a single ancestral gene in the last common ancestor of the analyzed species.

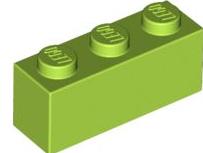


GitHub and...get started!

Scripts and data: https://github.com/mportela95/EvoDevo_2025

Data: <https://drive.google.com/drive/folders/1ZvYrkHUGwts61ZRXEdSUNnvrk3BwQr0X?usp=sharing>

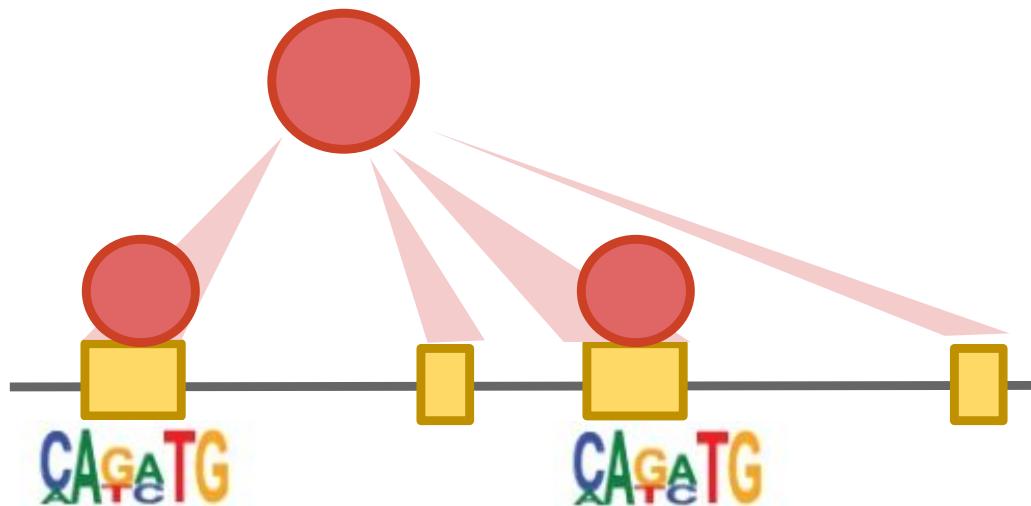
Explore the data with the UCSC browser (*C. dipterum*: GCA_902829235.1)



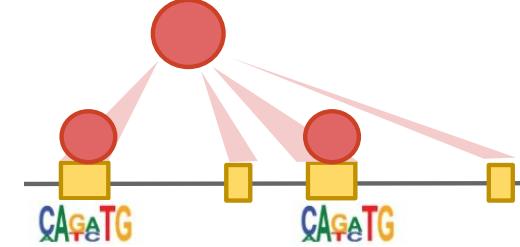
Results & Questions

- We detect causality but ... do we distinguish repression from activation?
- Our results are going to be robust?
- Debate about the thresholds: error type 1 (false positives) or type 2 (false negative)

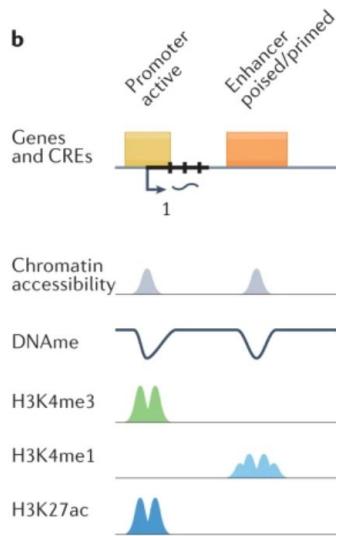
3. TF - CRE association



TF - CRE association



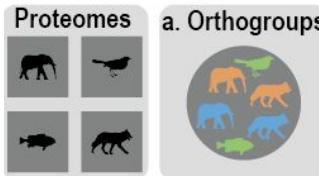
1. Sequences (bed2fasta)



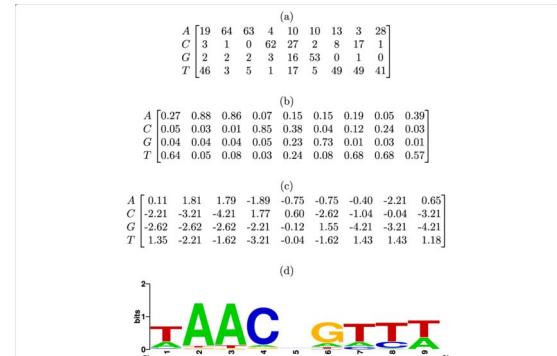
2. TF list in our specie

Bcl6b
Zscan26
Mtf1
Klf9
Zic5
Zfp410
Zfp3
...

*OrthoFinder 1**



3. PWM matrices



Softwares for TF - CRE association

Name	URL	Refs.
Binding motif databases		
CIS-BP	http://cisbp.ccbr.utoronto.ca/	209
cisTarget databases	https://resources.aertslab.org/cistarget/databases/	67
ENCODE	https://www.encodeproject.org/software/encode-motifs/	210
HOCOMOCO	https://hocomoco11.autosome.org/	211
JASPAR	https://jaspar.genereg.net/	212
TRANSFAC	https://genexplain.com/transfac/	213
UniPROBE	http://thebrain.bwh.harvard.edu/uniprobe/	214
Motif matcher algorithms		
FIMO	https://snystrom.github.io/memes-manual/	215
GimmeMotifs	https://gimmemotifs.readthedocs.io/	216
HOMER	http://homer.ucsd.edu/homer/motif/	68
MOODs (as implemented in motifmatchr)	https://github.com/jhkorhonen/MOODs https://github.com/GreenleafLab/motifmatchr	217, 218
motifanalysis (as implemented in reg-hint)	https://reg-gen.readthedocs.io/	219
PIQ toolkit	https://bitbucket.org/thashim/piq-single/src/master/	220
PWMScan	https://ccg.epfl.ch/pwmtools/pwmscan.php	221
pycistarget	https://pycistarget.readthedocs.io/	67

weng-lab/cluster-buster

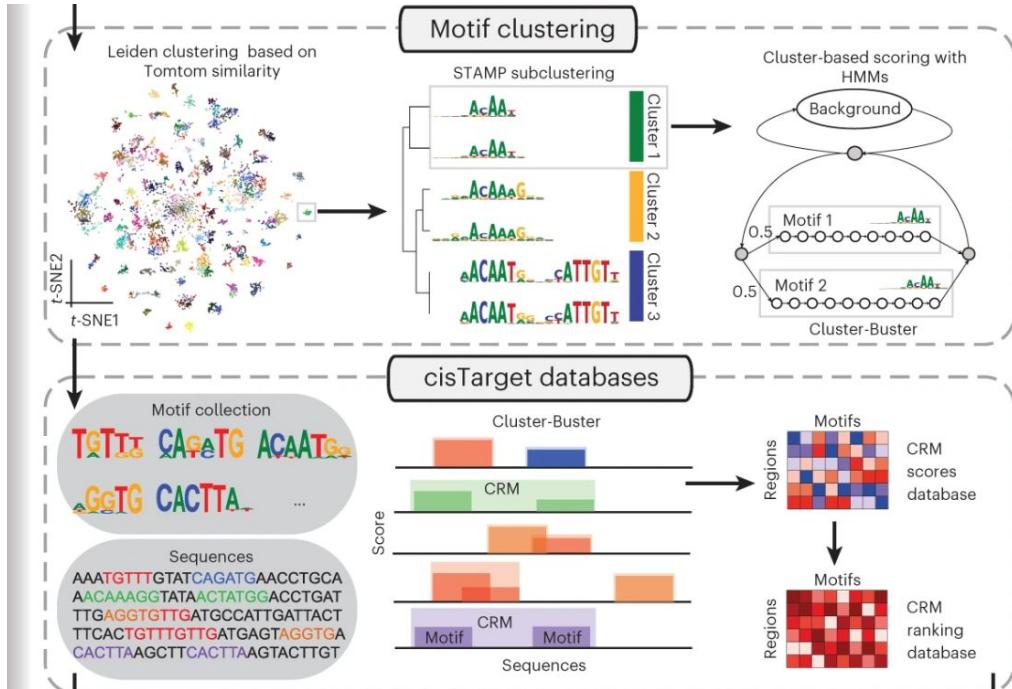


<https://bu.wenglab.org/cluster-buster/index.html>



Cluster-Buster

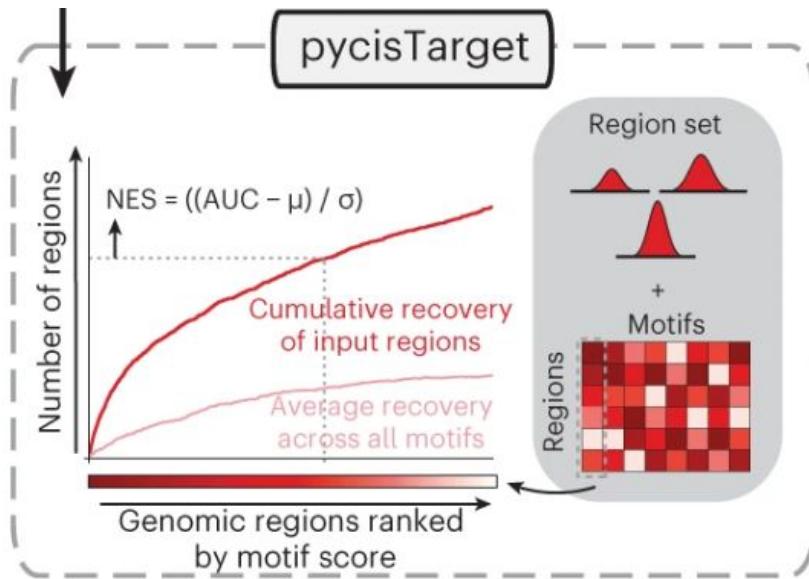
Regulatory regions are enriched not in single motifs, but in **clusters of motifs**.



The final **score** to generate a TF cistrome:

- matrix similarity
- density of motifs from its family in a region
- lineal proximity to other motifs
- position above cut-off in the *global ranking* in our interest regions

Area under the curve: global measure



TF - CRE association

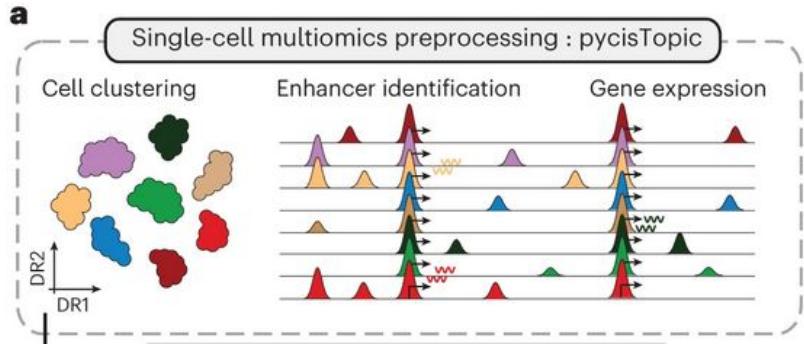
Motif Rankings in mayfly nymph 3

motifs <chr>	clodip_v4_1 <int>	clodip_v4_10 <int>	clodip_v4_100 <int>	clodip_v4_1000 <int>
bergman__Adf1	13740	12933	16580	14861
bergman__Aef1	8844	16666	12794	16565
bergman__Hr46	12741	3009	16239	2298
bergman__Kr	9893	4867	7136	8275
bergman__Su_H_	14055	7715	15593	4633
bergman__TFAM	7264	9270	8248	390

TF - CRE association

Is it cell-type /condition specific?

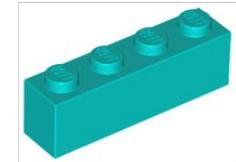
Enhancer identification at single-cell



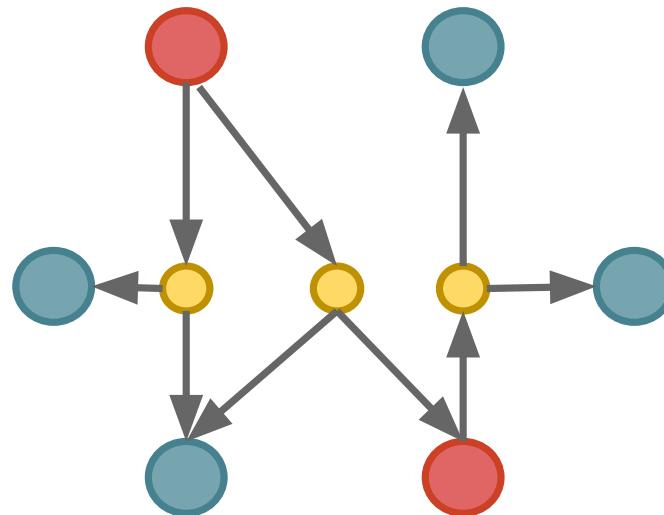
- **Low coverage**
- Topics
- metacells



TF - CRE -Gene association

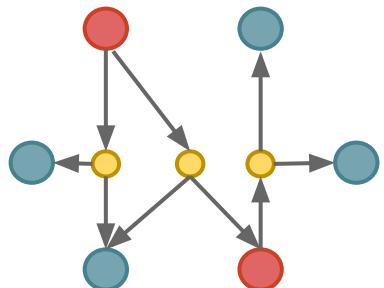


Thanks to the fasta and the motifAnnotations we can generate the cis-regulome

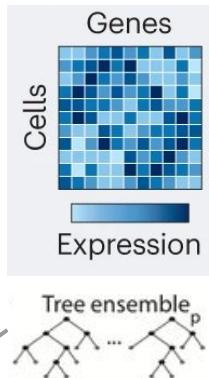


4. GRNs integration

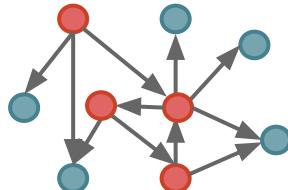
TF-CRE-Gene Network



Coexpression modules



Gene Regulatory Network



Regulon:
TF + targets

GRNs integration

1. Motif enrichment

Identifies which **motifs** are over-represented in a **gene set**:

cis-regulatory support for co-expressed genes.

Input:

```
# gen sets from Genie3
```

```
# Motif ranking for each gene
```

Output

```
# Motifs enriched for every gen set
```

2. Regulon pruning

Refines co-expression modules by retaining only target genes with direct cis-regulatory evidence for a given TF

Input:

```
# gen sets from Genie3
```

```
# Motifs enriched for every gen set
```

Output

```
# Pruned regulons
```

3. Regulon activity

Quantifies, for **each cell**, whether a regulon's target genes are enriched among the most highly expressed genes, yielding a cell-type-specific activity score for each regulon.

Input:

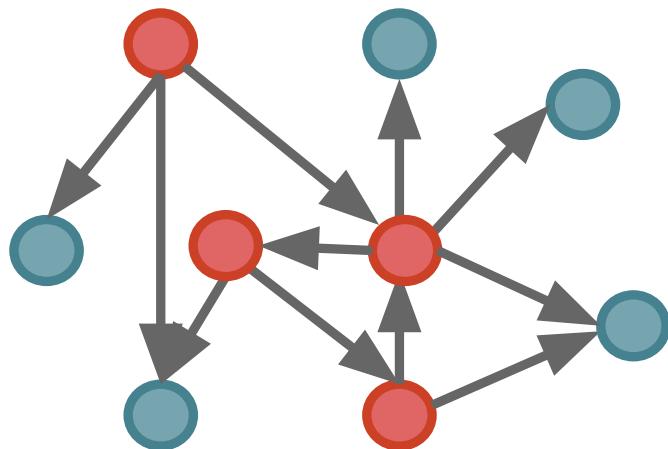
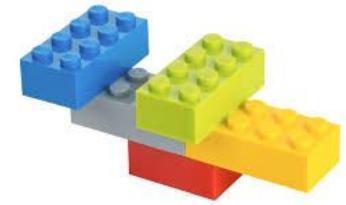
```
# Cell type specific expression data
```

```
# Pruned regulons
```

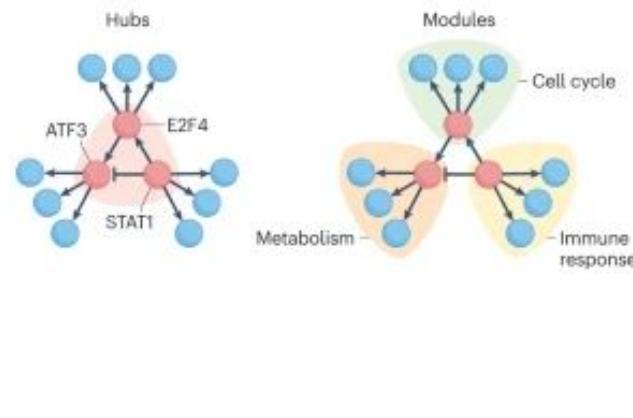
Output

```
Activity of each regulon in each cell
```

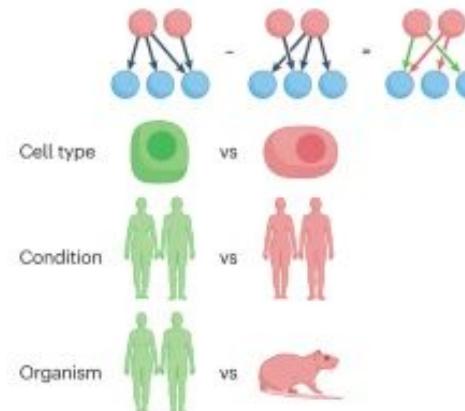
GRNs integration



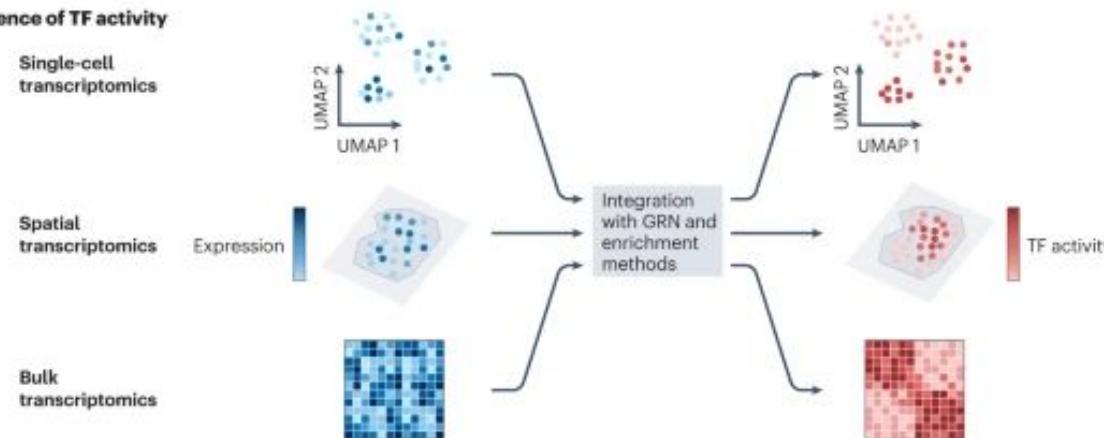
a Topological analysis



b Comparative analysis



c Inference of TF activity

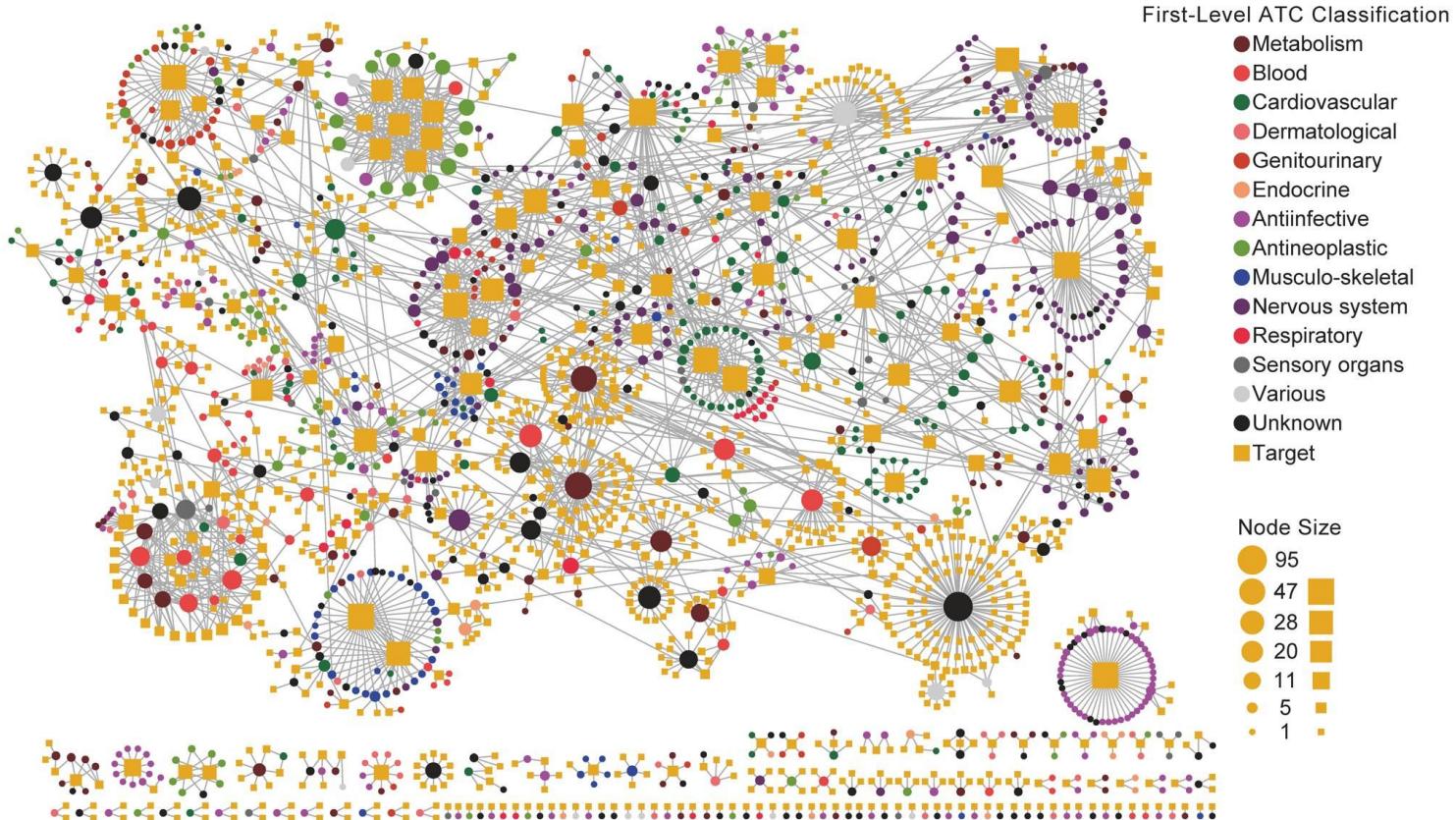


Questions

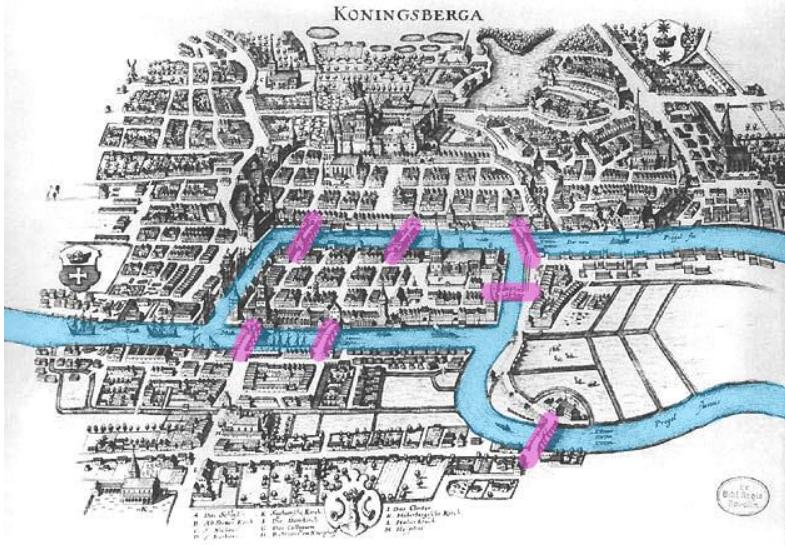
How can we check if a TF is a repressor or activator?

- TF expression vs regulon expression

Now...what?

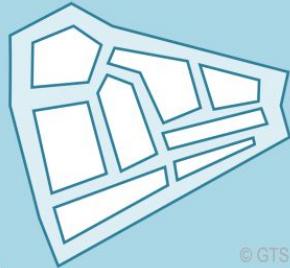


Topology

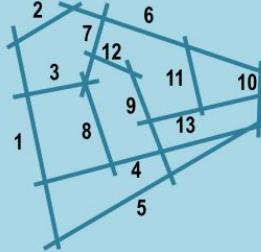


Seven Bridges of Königsberg, Euler
and the foundation of **graph theory**
(1736)

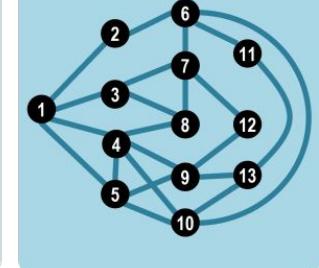
A. Urban Street Network



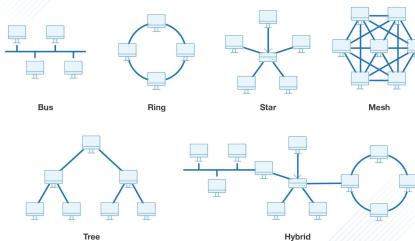
B. Axial Map



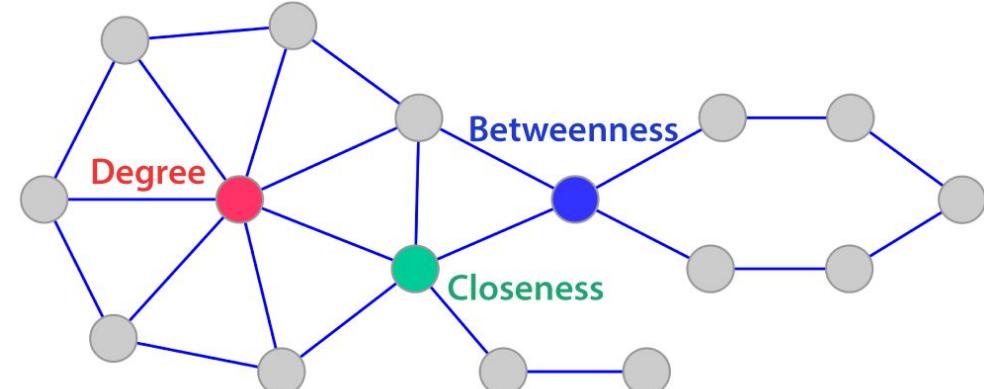
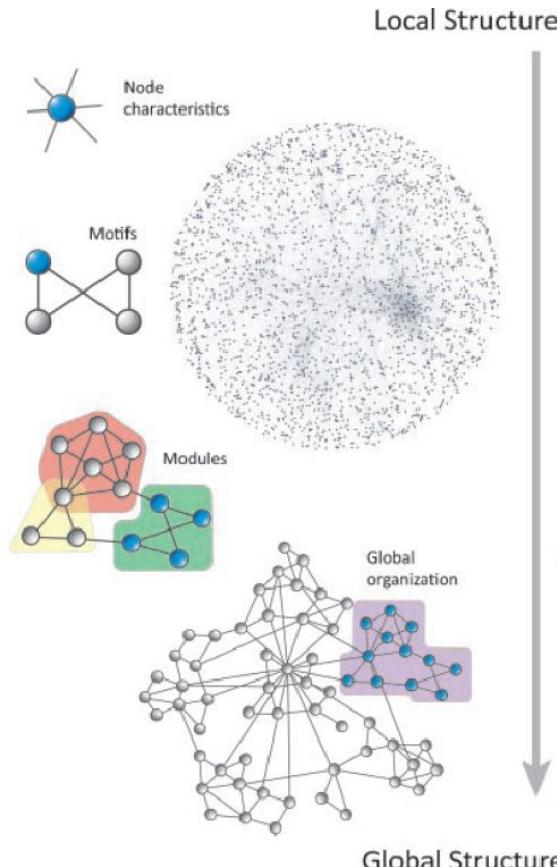
C. Dual Graph



Common Network Graph Topologies



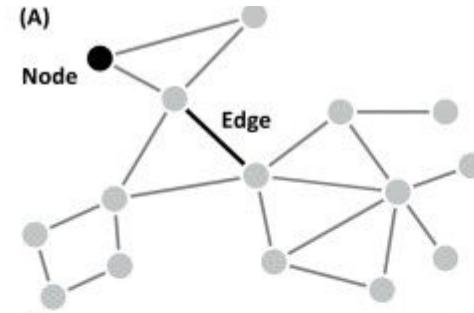
Topological approaches



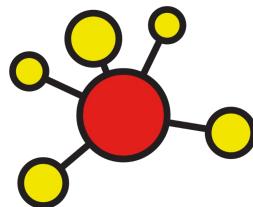
Degree	ClosenessCentrality	BetweennessCentrality
7	0.45454545	0.29047619
5	0.51724138	0.42380952
4	0.48387097	0.4952381

Information we can retrieve from topology

1. Insight into upstream and downstream relations
2. Centrality of a TF→relevance and impact ()
3. Molecular bottlenecks (bridges between processes)
4. Modularity (function)
5. Constrictions: fragile or flexible network spots
(hotspots for mutations, conservation, novelties)
6. Information direction (cascades and feedback loops)
7.



Network analysis and visualization tools



igraph

igraph: The network analysis package

igraph.org/

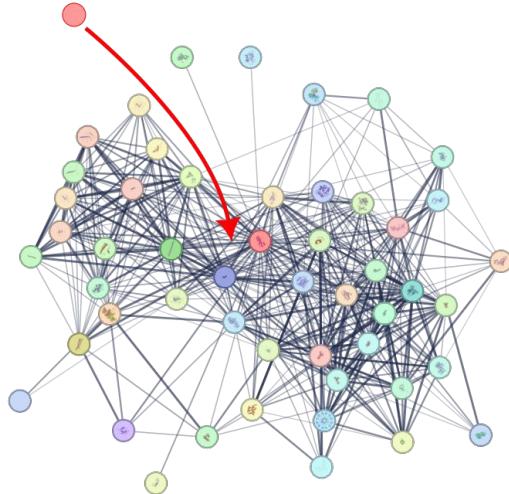


Cytoscape Web: bringing network biology to the browser

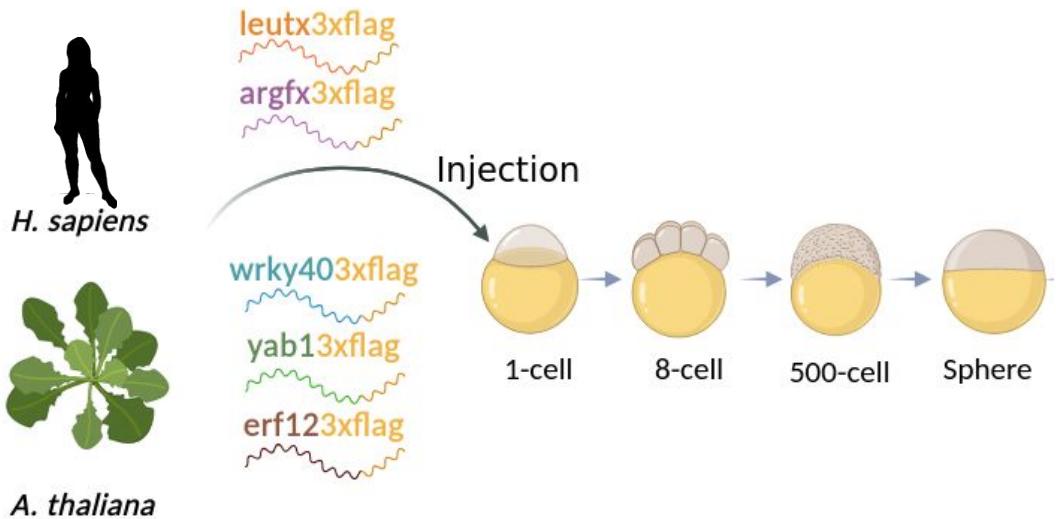
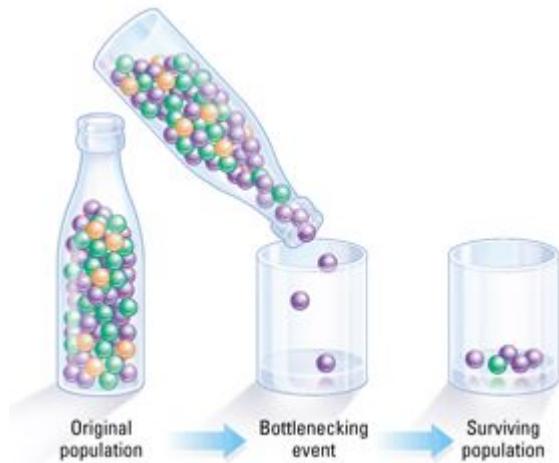
web.cytoscape.org

Our research:

What happens when a new element is added?

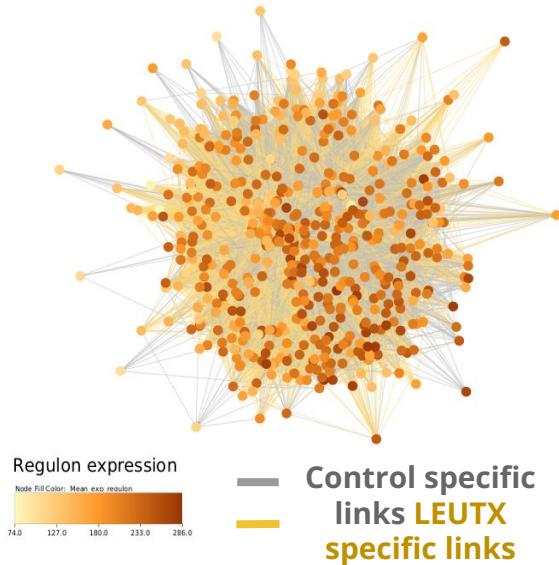


What happens when a new element is added?

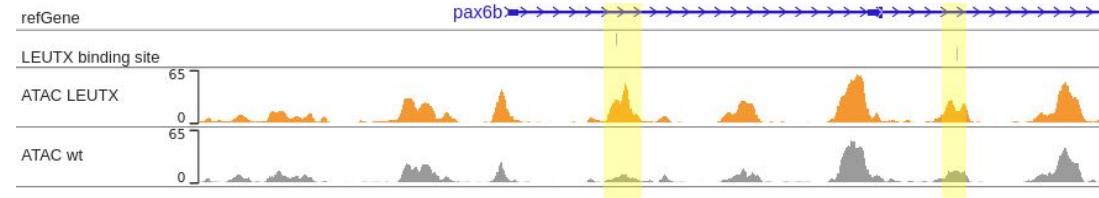
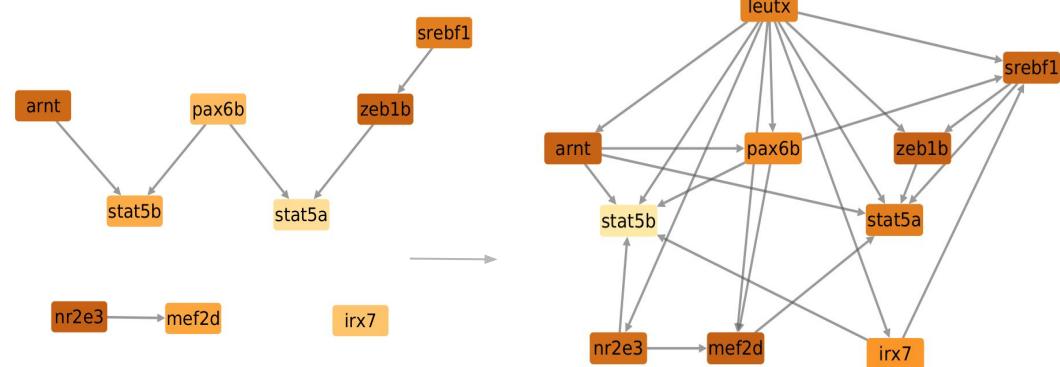


Ectopic TF impact on Gene Regulatory Networks

Differential regulation LEUTX vs Control



Rearrangement of native GRN with ectopic TF





Nacho Maeso

Alberto Pineda-Santaella

<https://nachomaesolab.wixsite.com/maesolab>



UNIVERSITAT DE
BARCELONA



Centro
Andaluz
de Biología
del Desarrollo

