

Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Supported by



Creative Commons

This page is available in the following languages:

Afrikaans Български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macdonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски српски (latinica) Sotho svenska
中文 華語 (台灣) isiZulu



Attribution-Share Alike 2.5 Canada

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

[Disclaimer](#)

Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
[English](#) [French](#)

[Learn how to distribute your work using this licence](#)

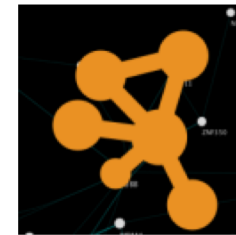
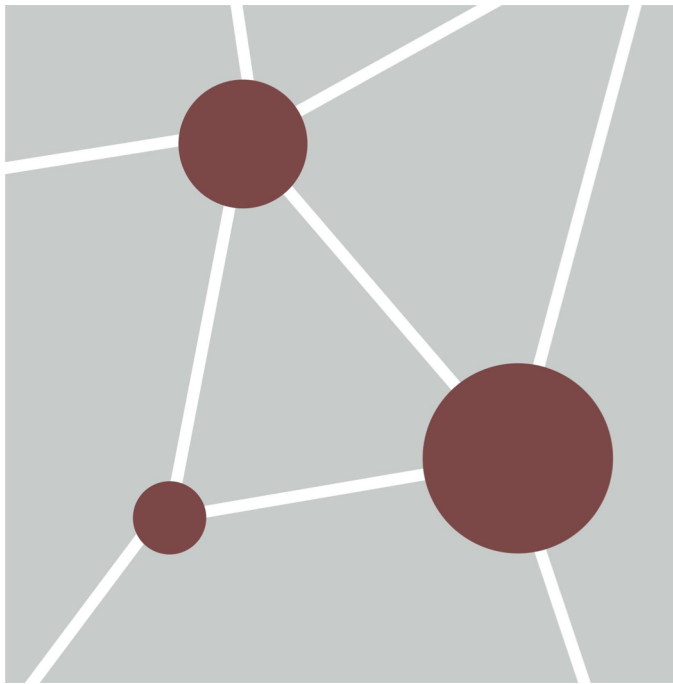
Gene Regulation and Motif Analysis Practical Lab



Veronique Voisin

Pathway and Network Analysis of -omics Data

July 27-29, 2020





iRegulon



<http://iregulon.aertslab.org/>

- Cytoscape app
- Developed to find transcription factor targets in a set of co-regulated genes
- Use precomputed results:
 - * precomputed occurrence of transcription motifs and chip-Seq peaks in promoter of all genes in genome
 - * rank all genes in genomes based on this result for each known transcription factor

Learning Objectives

This practical lab focuses on finding potential transcription factors that would regulate our gene list.

We are using 2 tools

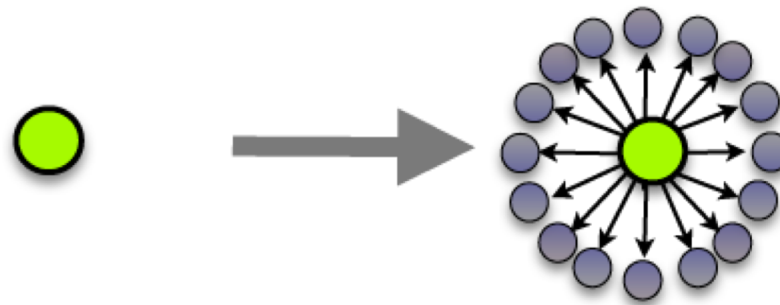
- EnrichR which is a web app doing enrichment analysis.
- IRegulon which is a Cytoscape app applying a GSEA like algorithm.



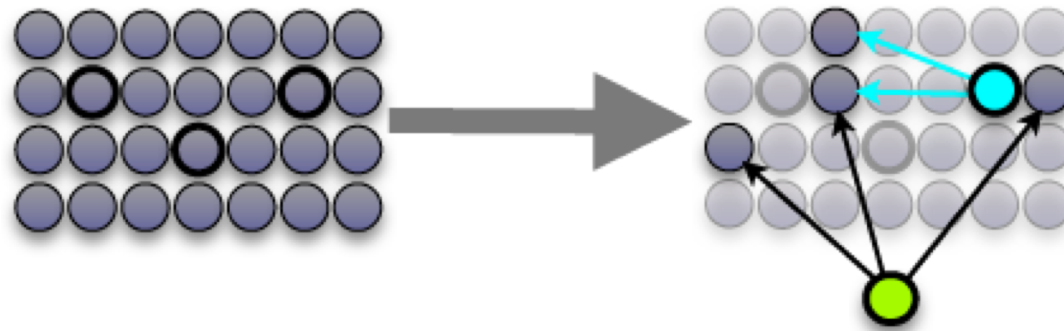
We are using iRegulon to answer 2 questions:



- 1) Find the targets of a transcription factor of interest
 - Advantages: Targets are displayed as a network, we can overlay extra information of the created network like node colors.



- 2) Given a set of genes, found if these genes are co-regulated by some transcription factors listed in the iRegulon database.





iRegulon uses:



- Position Weight Matrix (PWM) (+transcription factor predicted binding site)



and

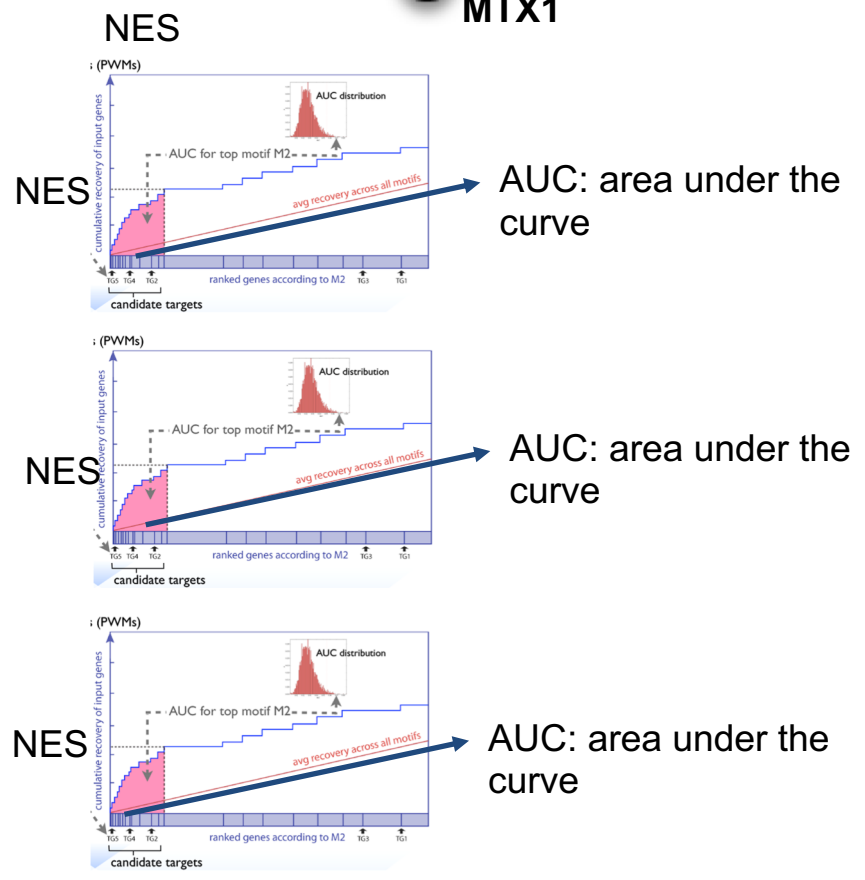
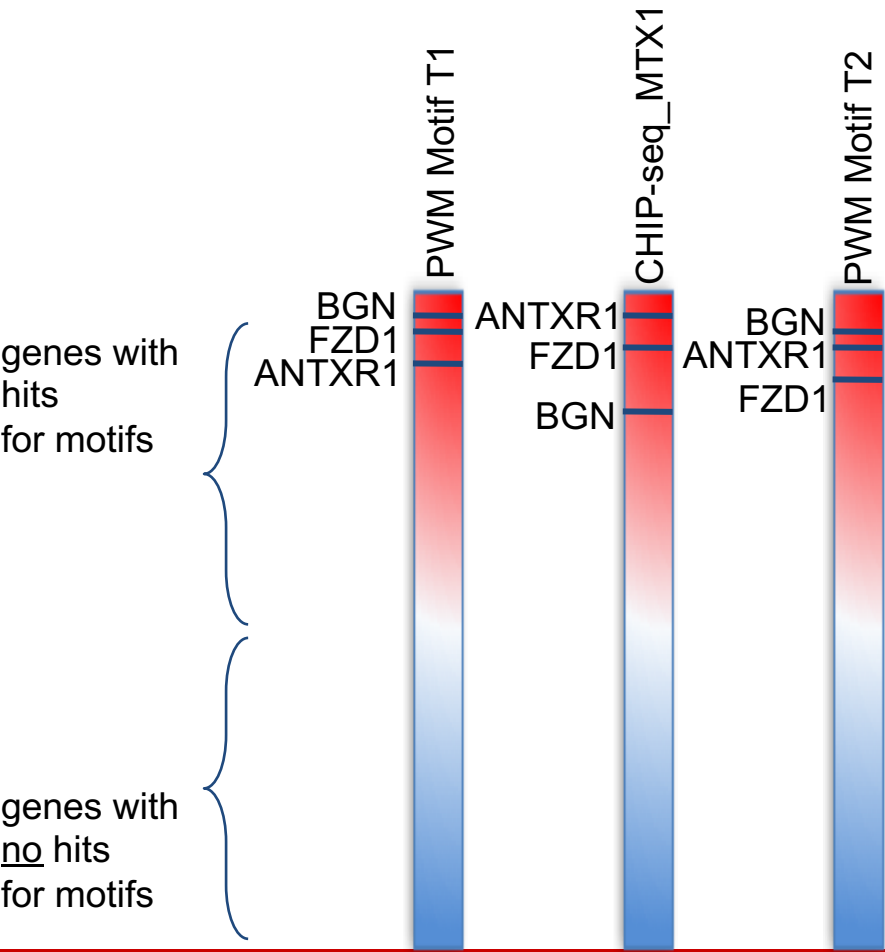
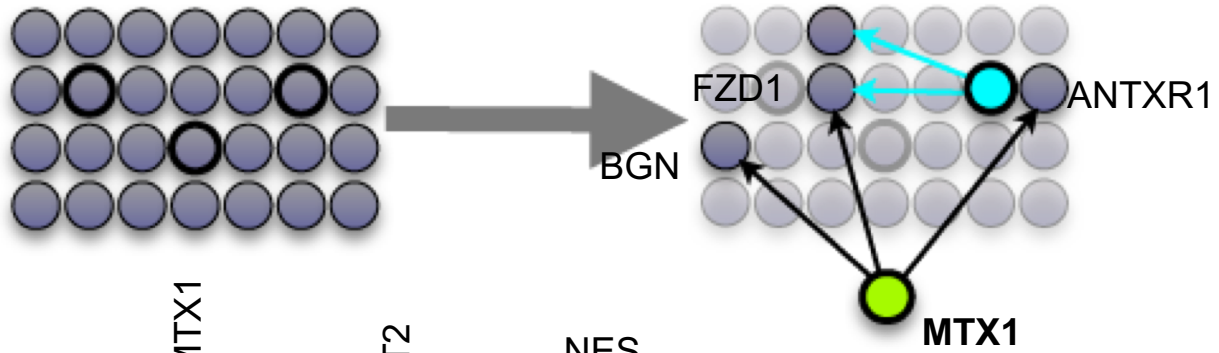
- CHIP-Seq data (immunoprecipitation of chromatin using an antibody specific to a TF of interest + DNA sequencing)

To estimate transcription factor binding on promoter of genes of interest

Description of the motif and track collections used.

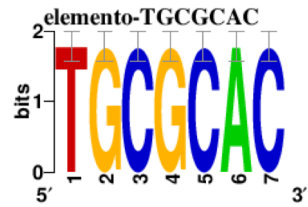
Source	Organism(s)	Type of motif	# motifs "6K"	# motifs "10K"	# tracks "1K ChIP"
Elemento [73]	Drosophila	Predicted (conserved) ^a	371	371	-
FlyFactorSurvey [75]	Drosophila	B-1H, others (e.g., FlyReg)	614	652	-
hPDI [77]	Human	Experimental	437	437	-
Jaspar [21]	Multiple species	Curated	1315	1315	-
SelexConsensus [76]	Drosophila	Curated (FlyReg)	38	38	-
Stark [74]	Drosophila	Predicted (conserved) ^a	228	228	-
Tiffin [76]	Drosophila	Predicted (gene sets) ^a	120	120	-
TRANSFAC PUBLIC [5]	Multiple species	Curated, ChIP-chip	398	398	-
TRANSFAC PRO [5]	Multiple species	Curated, ChIP-chip	1153	1850	-
YetFasco [78]	Yeast	Uniprobe, Curated, ChIP-chip	1709	1709	-
ENCODE [79]	Human	Predicted (from DHS) ^a	-	683	-
Factorbook [46]	Human	ENCODE ChIP-Seq motifs	-	79	-
Taipale [132]	Human, Mouse	HT-Selex	-	820	-
iDMMPMM [133]	Human	footprints, Selex, b1h, peaks	-	39	-
SwissRegulon [134]	Human	Curated	-	190	-
Wolfe [135]	Drosophila	ZFP motifs	-	36	-
HOMER [116]	Multiple species	ChIP-Seq Motifs, others (e.g. ENCODE)	-	1865	-
Dimers [136]	Human	Predicted dimers	-	603	-
ENCODE ChIP-Seq [23]	Human	-	-	-	999
Taipale ChIP-Seq [24]	Human	-	-	-	117
p53 and control ChIP-Seq (this study)	Human	-	-	-	2
Total			6383	11611 (9713 nr)	1118

How does iRegulon find if a set of co-expressed genes is the target of a TF?

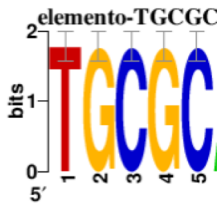


From Motifs, Tracks to Transcription Factor

Motif2Fmapping

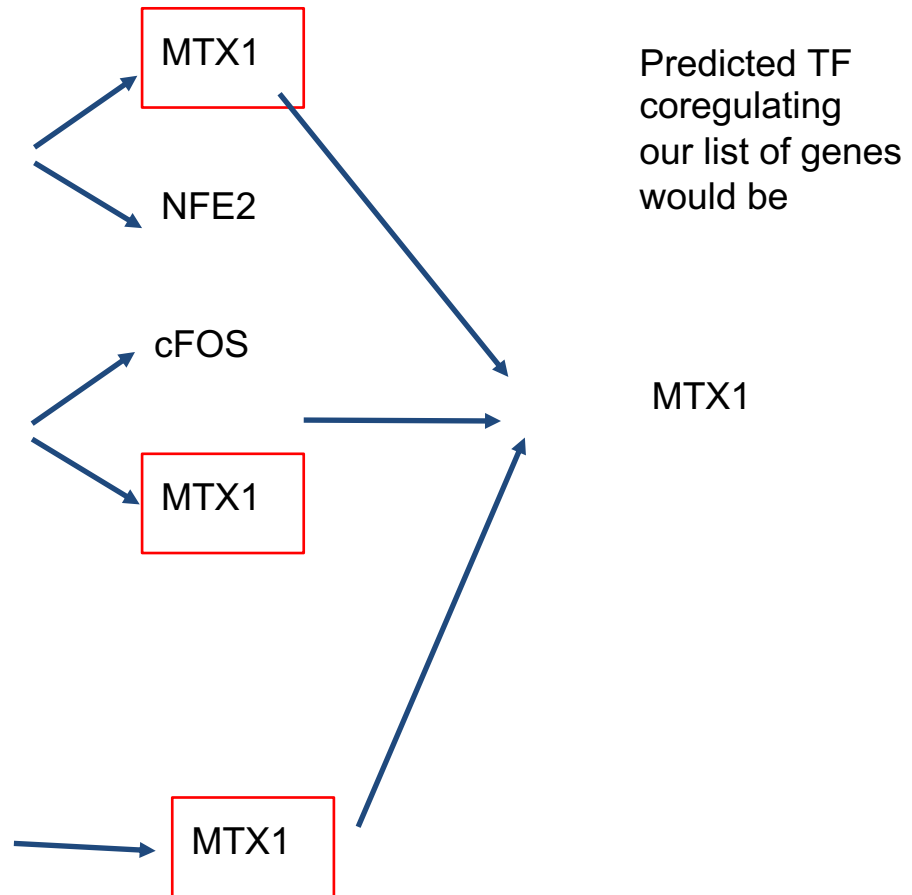


PWM Motif T1

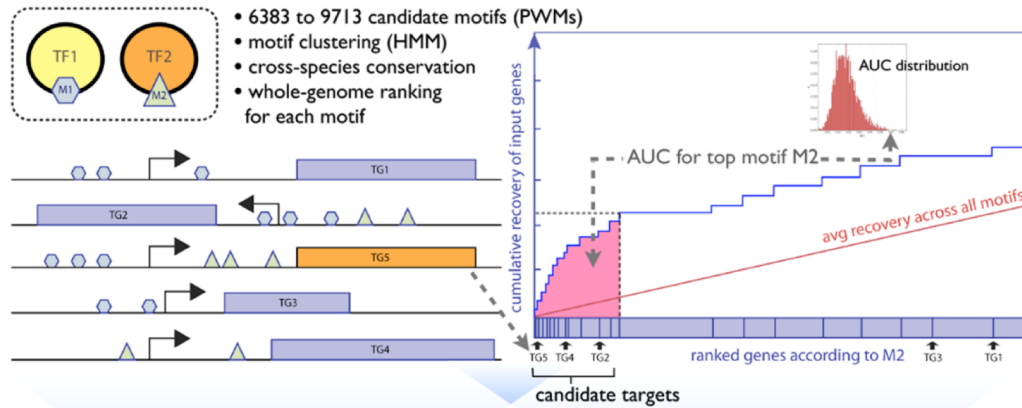


PWM Motif T2

TRACK CHIP-seq_MTX1

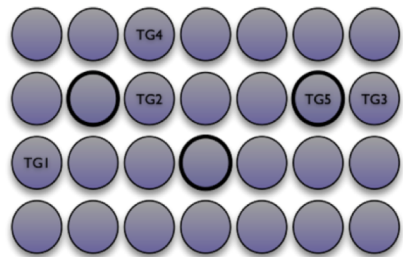


Motif2Fmapping



INPUT

Set of genes (HGNC Symbol)



28 genes in my gene list

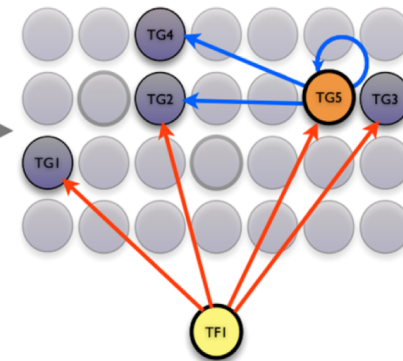
A. RANKING

B. RECOVERY

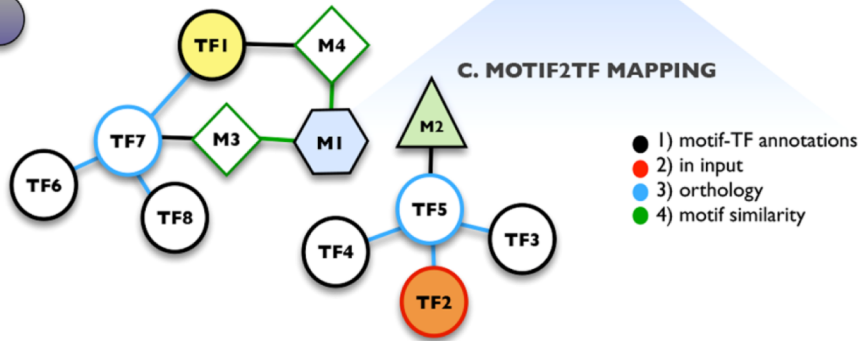
Regulon

OUTPUT

Regulons

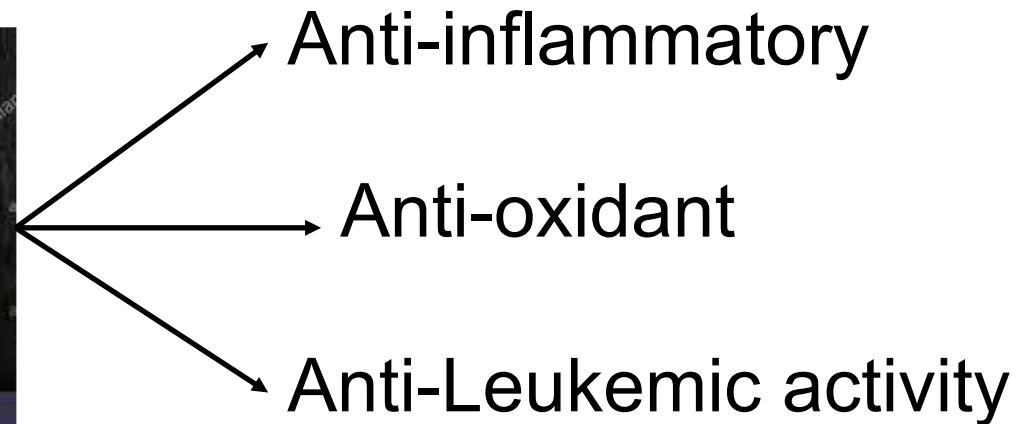


C. MOTIF2TF MAPPING



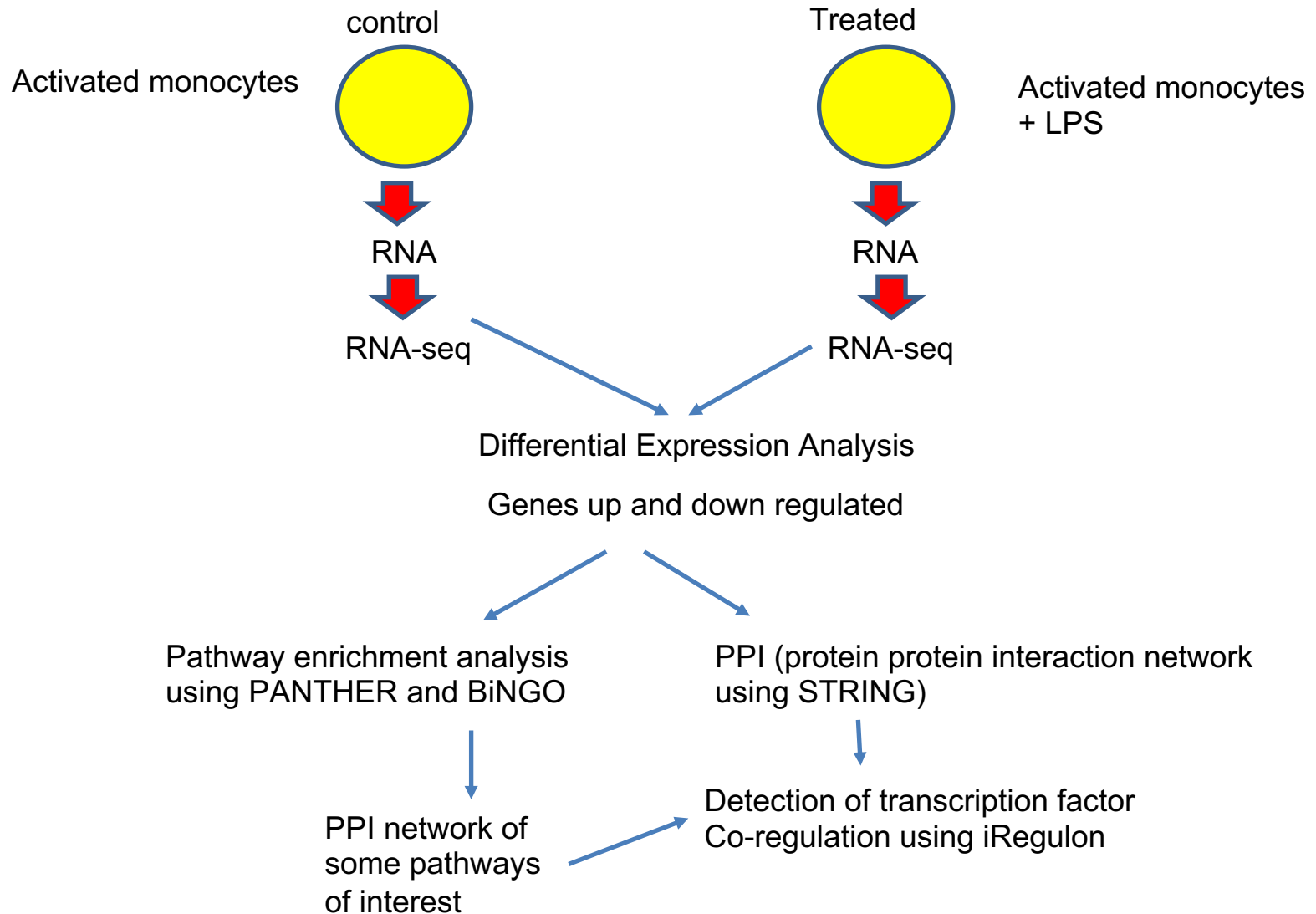
- Example of a published paper using iRegulon

Resveratrol decreases the expression of genes involved in inflammation through transcriptional regulation

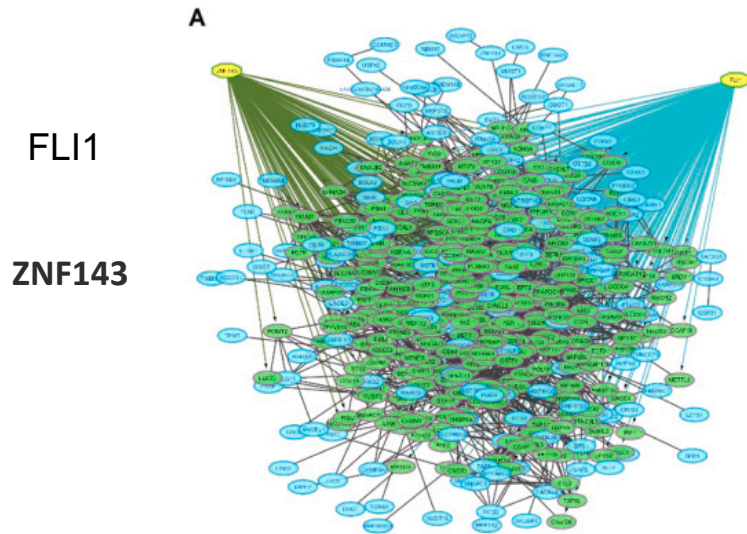


<https://www.sciencedirect.com/science/article/pii/S0891584918314886?via%3Dihub>

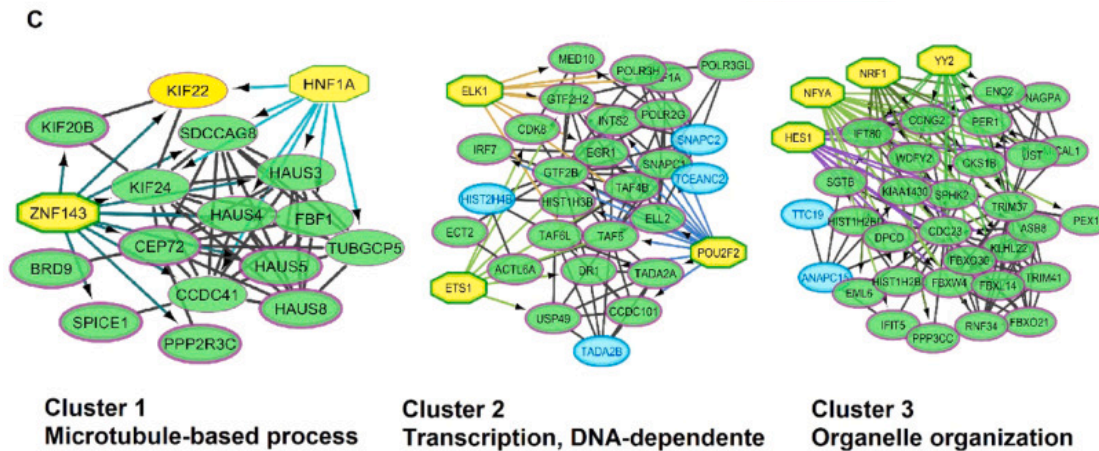
Resveratrol decreases the expression of genes involved in inflammation through transcriptional regulation



Resveratrol decreases the expression of genes involved in inflammation through transcriptional regulation



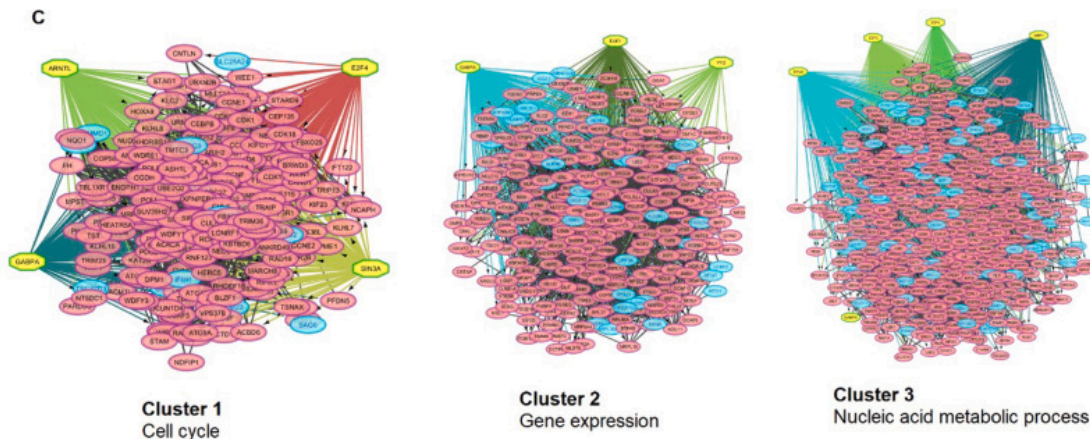
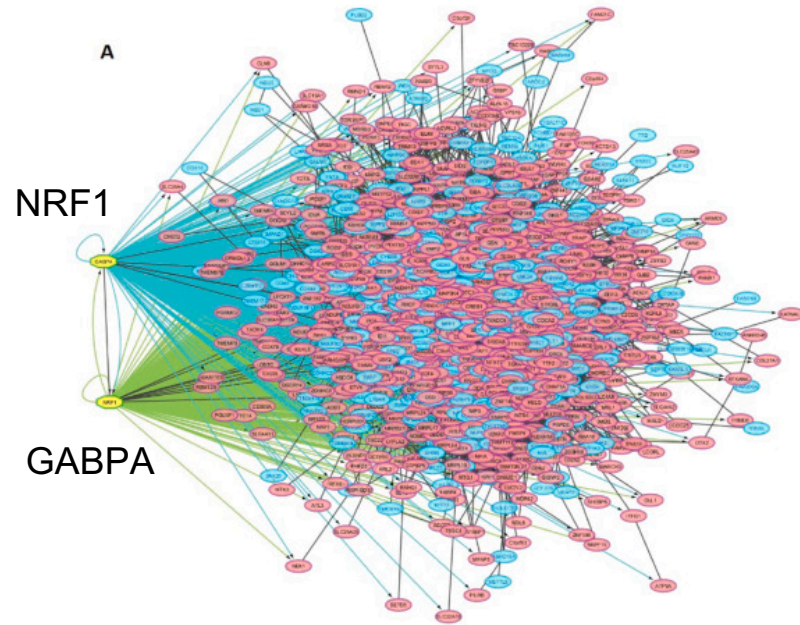
iRegulon on
GENES UP
REGULATED



Resveratrol decreases the expression of genes involved in inflammation through transcriptional regulation

iRegulon on GENES DOWN REGULATED

The prediction of master regulators using the iRegulon tool showed **nuclear respiratory factor 1 (NRF1)** and **GA-binding protein alpha subunit (GABPA)** as the main regulators of the downregulated genes





EnrichR

- Input can be a gene list or a bed file

Note: a bed file contains chromosome coordinated of the peaks identified for example in ATACseq or ChipSeq experiments. Consider filtering peaks in region proximal to associated genes like in promoter regions.

Input data

Choose an input file to upload. Either in BED format or a list of genes.

Try an example [BED file](#).

No file chosen

Paste a list of valid Entrez gene symbols on each row in the text-box below. [Try a gene set example](#).

0 gene(s) entered



Enrichr

- Test: Fisher's exact test, with a pvalue corrected for multiple hypothesis testing
- Output: a downloadable table or graphs
- No option to put a reference background: use only if you are doing a whole genome study!
- Enrichr has a vast variety of gene-set sources.
- In this practical lab, we are only going to focus on **"Transcription"**.
- Possibility to download a result of interest and create an enrichment map in Cytoscape using the Generic/gProfiler/Enrichr format



Login | Register

Transcription Pathways Ontologies Diseases/Drugs Cell Types Misc Legacy Crowd

Description Sample gene list (375 genes)

ChEA 2016

- E2F1 18555785 ChIP-Seq MESC's Mouse
- JARID1A 20064375 ChIP-Seq MESC's Mouse
- PPARA 22158963 ChIP-Seq LIVER Mouse
- ZFX 18555785 ChIP-Seq MESC's Mouse
- NELFA 20434984 ChIP-Seq ESCs Mouse

ENCODE and ChEA Consensus TFs from

- ERG CHEA
- NR2C2 ENCODE
- GABPA ENCODE
- HNF4A ENCODE
- ZMIZ1 ENCODE

ARCHS4 TFs Coexp

- NPAS1 human tf ARCHS4 coexpression
- BBX human tf ARCHS4 coexpression
- ZBTB6 human tf ARCHS4 coexpression
- ZNF17 human tf ARCHS4 coexpression
- GABPB1 human tf ARCHS4 coexpression

TF Perturbations Followed by Expression

- GLIS2 KO MOUSE GSE6113 CREEDSID GENE
- RORA KO MOUSE GSE23736 CREEDSID GENE
- IRF9 OE HUMAN GSE50002 CREEDSID GENE
- PPARA KO MOUSE GSE18564 CREEDSID GEN
- IRF9 OE HUMAN GSE50002 CREEDSID GENE

TRRUST Transcription Factors 2019

- MTF1 human
- PTTG1 human
- PPARGC1A mouse
- NRF1 mouse
- TCF21 mouse

IncHUB lncRNA Co-Expression

- MAP4K3-DT
- SUCLG2-AS1
- LINC02776
- ITFG1-AS1
- LINC01852

Enrichr Submissions TF-Gene Cooccurrence

- ZRSR1
- ZNF830
- MYNN
- NUPL2
- PMS1

TRANSFAC and JASPAR PWMs

- CEBPE (human)
- FOXA1 (mouse)
- SOX2 (human)
- NFATC3 (human)
- FOXF2 (human)

Epigenomics Roadmap HM ChIP-seq

- H3K27ac Duodenum Smooth Muscle
- H3K79me2 H9
- H3K27ac CD4 Naive Primary Cells
- H3K27ac Mobilized CD34 Primary Cells
- H3K23ac IMR90

Enrichr output table

Fisher's exact test

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted	Odds Ratio	Combined Score	Genes
NR5A1 human	8/22	1.86E-16	1.06E-13	0	0	151.515152	5488.395484	HSD3B2;STAR;CYP11A1;CYP11B2;CYP11B1;HSD3B1;CYP19A1;CYP17A1
NR5A2 human	5/12	5.02E-11	1.43E-08	0	0	173.611111	4117.076135	HSD3B2;STAR;CYP11A1;CYP11B1;CYP19A1
NR5A1 mouse	5/23	2.09E-09	3.98E-07	0	0	90.5797101	1810.220371	STAR;CYP11A1;CYP11B2;CYP19A1;CYP17A1
NR4A1 human	4/15	3.91E-08	5.58E-06	0	0	111.111111	1895.289459	HSD3B2;STAR;CYP11A1;CYP17A1
CREB1 human	6/90	7.39E-08	8.44E-06	0	0	27.7777778	456.1178554	AR;CYP11A1;CYP11B2;CYP11B1;CYP19A1;CYP17A1
NR0B1 human	3/10	1.54E-06	1.46E-04	0	0	125	1673.092298	STAR;CYP19A1;CYP17A1
NR0B1 mouse	3/11	2.11E-06	1.72E-04	0	0	113.636364	1484.996478	STAR;CYP11A1;CYP19A1
SF1 human	3/12	2.81E-06	2.01E-04	0	0	104.166667	1331.455379	STAR;HSD3B2;CYP11A1
SF1 mouse	3/15	5.78E-06	3.67E-04	0	0	83.3333333	1005.029762	STAR;CYP11A1;CYP17A1
SP1 human	8/472	1.49E-05	8.53E-04	0	0	7.06214689	78.4727961	AR;STAR;HSD17B1;HSD17B2;HSD3B1;HSD17B11;CYP19A1;CYP17A1
NR2F1 human	2/9	2.01E-04	0.010427214	0	0	92.5925926	788.2250101	CYP19A1;CYP17A1
GATA6 human	2/9	2.01E-04	0.009558279	0	0	92.5925926	788.2250101	CYP11A1;CYP17A1
NR4A1 mouse	2/10	2.51E-04	0.011011909	0	0	83.3333333	690.9348177	AR;HSD3B1
JUN human	4/149	4.46E-04	0.018198668	0	0	11.1856823	86.2946171	AR;STAR;CYP11A1;CYP19A1

gene-set
(pathway)

Overlap:
Numerator -> # of
genes in my gene list
and tested gene-set

FDR:
Correction for multiple
hypothesis testing

Denominator ->
of genes in the
original gene-set

List of genes in the
overlap

We are on a Coffee Break & Networking Session

compute | calcul
canada | canada

Workshop Sponsors:



Canadian Centre for
Computational
Genomics

MiC:M McGill initiative in
Computational Medicine