



workshop

Zeynep Kalender Atak & Hana Imrichova



CENTER FOR HUMAN GENETICS



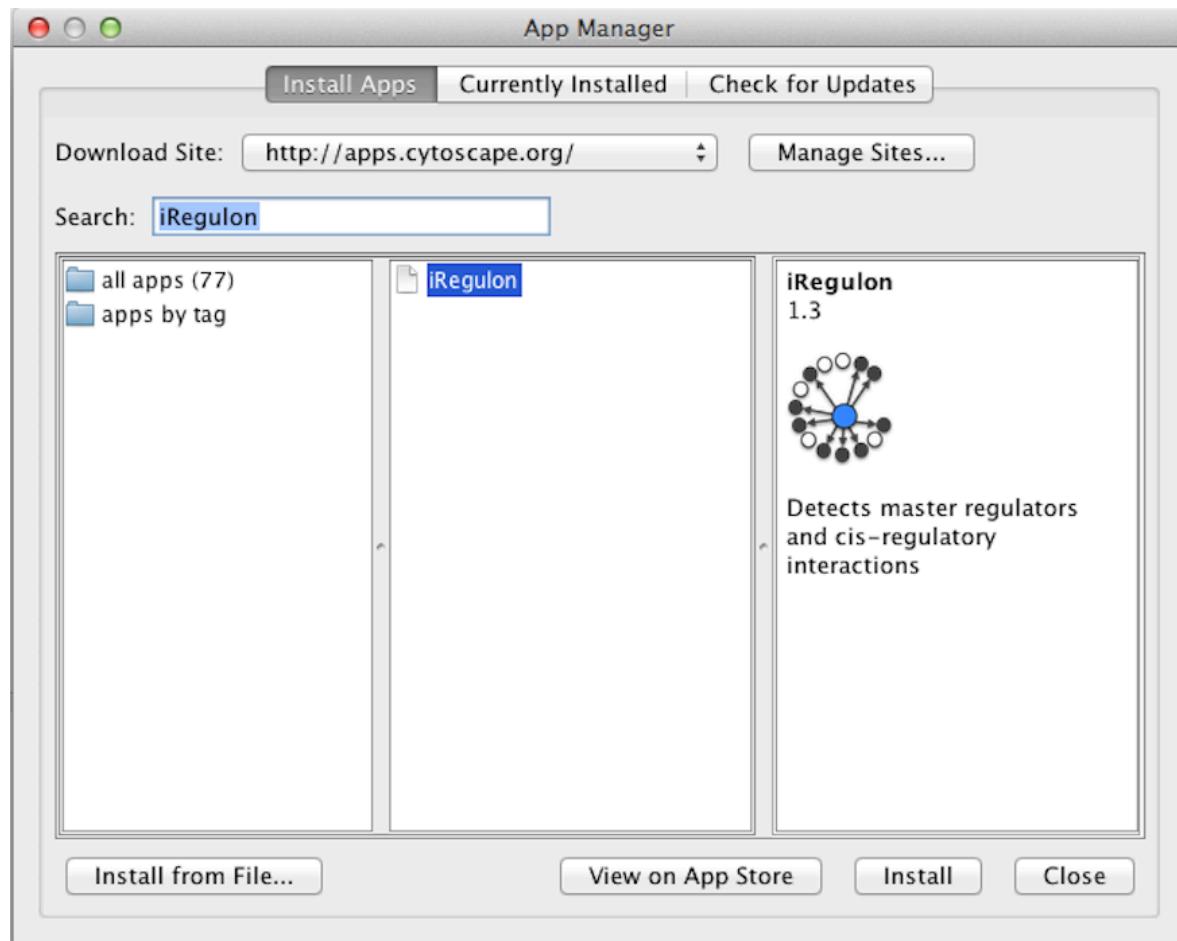
<http://aertslab.org>
<http://iregulon.aertslab.org>

 @aertslab

Analyzing gene expression data using iRegulon

Analysis of a ZEB1 perturbation data set

Installing iRegulon



- You need to have Cytoscape 3.0 installed
- Launch Cytoscape (ignore the dialog box)
- Go to Apps > Apps manager
- Search for iRegulon and install
- iRegulon can now be found under the Apps menu

Getting the Data from MsigDB

- Go to the MsigDB website (www.broadinstitute.org/gsea/msigdb)
- Click the search option
- Login with your email or register first (free)

Getting the Data from MsigDB

The screenshot shows the MSigDB search interface. In the search bar, the keyword "ZEB1" is entered. Red arrows point to the search filters for collection and organism. The collection filter includes options like hallmark gene sets, curated gene sets, and chemical and genetic perturbations. The organism filter includes all organisms, Homo sapiens, and other primates like Macaca mulatta and Mus musculus. The contributor filter lists various institutions and individuals. Below the filters, a table displays search results for "ZEB1". The first result, "AIGNER_ZEB1_TARGETS", is highlighted with a red border. It contains 35 genes and is described as genes up-regulated in MDA-MB-231 cells (breast cancer) after knockdown of ZEB1 [GeneID=6935] by RNAi.

name	# genes	description
AIGNER_ZEB1_TARGETS	35	Genes up-regulated in MDA-MB-231 cells (breast cancer) after knockdown of ZEB1 [GeneID=6935] by RNAi.
ALFANO_MYC_TARGETS	239	Genes up-regulated hT-RPE cells (immortalized retinal pigment epithelium) by MYC [GeneID=4609].
AMIT_EGF_RESPONSE_240_HELA	60	Genes whose expression peaked at 240 min after stimulation of HeLa cells with EGF [GeneID=1950].

- Type ZEB1 in to search box
- Put the following filters
 - C2-CGP
 - Homo Sapiens
- Pick the first dataset:
AIGNER_ZEB1_TARGETS

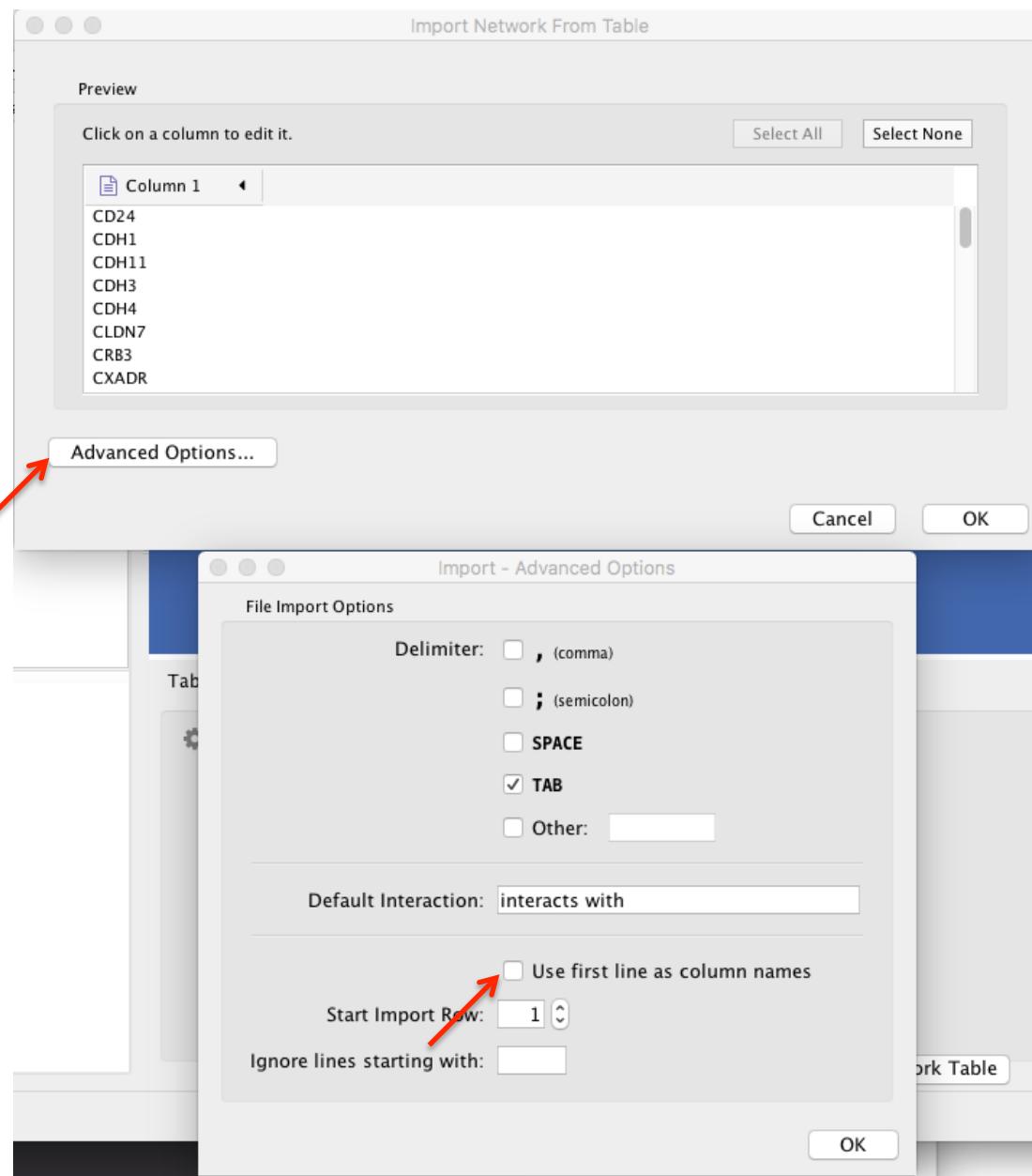
Getting the Data from MsigDB

Gene Set: AIGNER_ZEB1_TARGETS

Standard name	AIGNER_ZEB1_TARGETS
Systematic name	M14590
Brief description	Genes up-regulated in MDA-MB-231 cells (breast cancer) after knockdown of ZEB1 [GeneID=6935] by RNAi.
Full description or abstract	Epithelial to mesenchymal transition (EMT) is implicated in the progression of primary tumours towards metastasis and is likely caused by a pathological activation of transcription factors regulating EMT in embryonic development. To analyse EMT-causing pathways in tumorigenesis, we identified transcriptional targets of the E-cadherin repressor ZEB1 in invasive human cancer cells. We show that ZEB1 repressed multiple key determinants of epithelial differentiation and cell-cell adhesion, including the cell polarity genes Crumbs3, HUGL2 and Pals1-associated tight junction protein. ZEB1 associated with their endogenous promoters <i>in vivo</i> , and strongly repressed promoter activities in reporter assays. ZEB1 downregulation in undifferentiated cancer cells by RNA interference was sufficient to upregulate expression of these cell polarity genes on the RNA and protein level, to re-establish epithelial features and to impair cell motility <i>in vitro</i> . In human colorectal cancer, ZEB1 expression was limited to the tumour-host interface and was accompanied by loss of intercellular adhesion and tumour cell invasion. In invasive ductal and lobular breast cancer, upregulation of ZEB1 was stringently coupled to cancer cell dedifferentiation. Our data show that ZEB1 represents a key player in pathologic EMTs associated with tumour progression.
Collection	C2: curated gene sets CGP: chemical and genetic perturbations
Source publication	Pubmed 17486063 Authors: Aigner K,Dampier B,Descovich L,Mikula M,Sultan A,Schreiber M,Mikults W,Brabietz T,Strand D,Obrist P,Sommergruber W,Schweifer N,Wernitznig A,Beug H,Foisner R,Eger A
Exact source	Table 1
Related gene sets	(show 14 gene sets from the same authors)
External links	
Organism	Homo sapiens
Contributed by	Leona Saunders (Broad Institute)
Source platform	HUMAN_GENE_SYMBOL
Dataset references	
Download gene set	format: grp text gmt gmx xml
Compute overlaps 	(show collections to investigate for overlap with this gene set)
Compendia expression profiles 	Human Tissue compendium (Novartis) Global Cancer Map (Broad Institute) NCI60 cell lines (National Cancer Institute)
Advanced query	Further investigate these 35 genes
Gene families 	Categorize these 35 genes by gene family
Show members	(show 37 members mapped to 35 genes)
Version history	3.0: First introduced

- Click on ‘text’ to get a text file with the names of the target genes
- Save the file in your preferred directory
- Remove the first two lines in a text editor

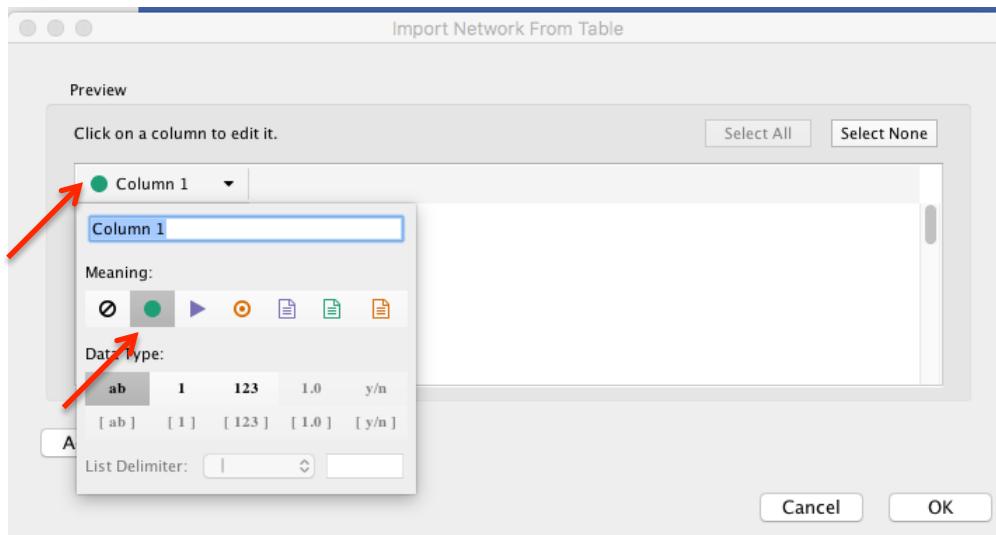
Loading the Data in Cytoscape



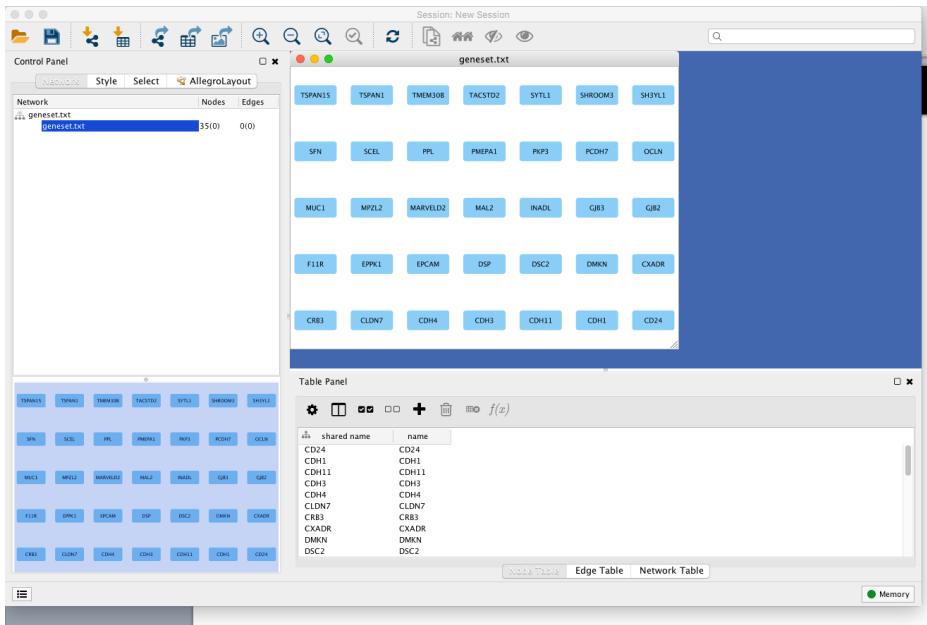
- Launch Cytoscape
- Ignore the initial dialogue box
- Go to File menu > Import > network > file
- Select your input text file
- Go to Advanced options
- Un-tick the box “Use first line as column names”

Loading the Data in Cytoscape

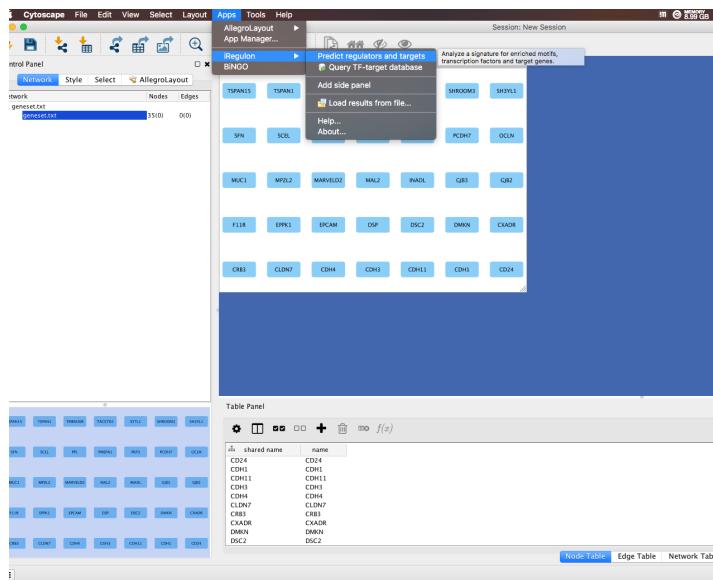
- Click on Column 1 and select the green dot
- Click OK
- Click Yes on the next dialog box



Starting up an iRegulon analysis



- You will get a network without connections
- Each node is one of the imported genes



- Select all the nodes (CMD + A)
- Go to Apps menu > iRegulon > Predict regulators and targets

Selecting the parameters

Predict regulators and targets

Name for analysis: ZEB1_genes.txt

Species and gene nomenclature: Homo sapiens, HGNC symbols

Node information

Node attribute that corresponds to geneID: name

Number of selected genes (nodes): 35

Ranking

Type of search space: gene-based

Motif collection: 10K (9713 PWMs)

Track collection: 1120 ChIP-seq tracks (ENCODE raw signals)

Putative regulatory region: 20kb centered around TSS

Motif rankings database: 20kb centered around TSS (7 species)

Track rankings database: 20kb centered around TSS (ChIP-seq-derived)

Region-based specific parameters

Overlap fraction: 0.4

Upstream region: 5000

Downstream region: 5000

Recovery

Enrichment score threshold: 3.0

ROC threshold for AUC calculation: 0.03

Rank threshold: 5000

TF prediction

Minimum identity between orthologous genes: 0.0

Maximum false discovery rate (FDR) on motif similarity: 0.001

Cancel Submit

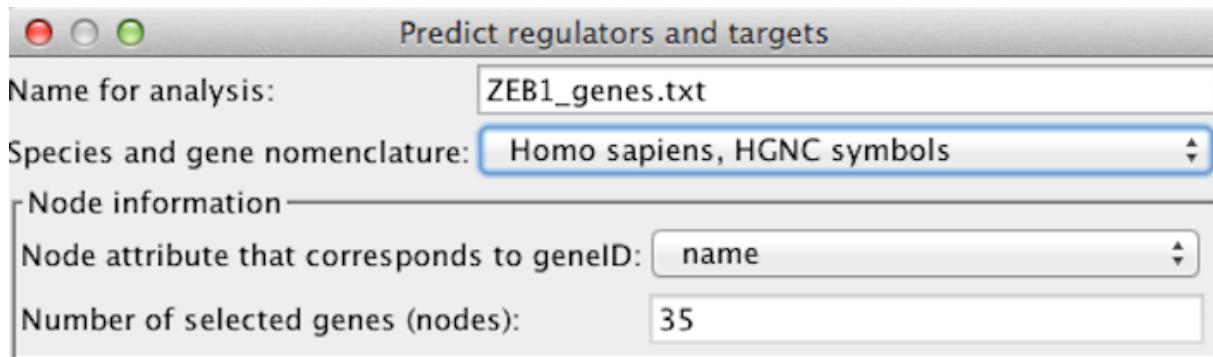
General parameters

Ranking parameters

Region-based parameters

Recovery and TF prediction parameters

Selecting the parameters: general parameters



- Give your analysis a name
- Select the species corresponding to the imported genes, in this case human
- The node attribute that is selected as the gene identification is by default the name.
- Number of nodes reflects the number of genes you've selected for the analysis

Selecting the parameters: Ranking parameters

Ranking

Type of search space: gene-based

Motif collection: 10K (9713 PWMs)

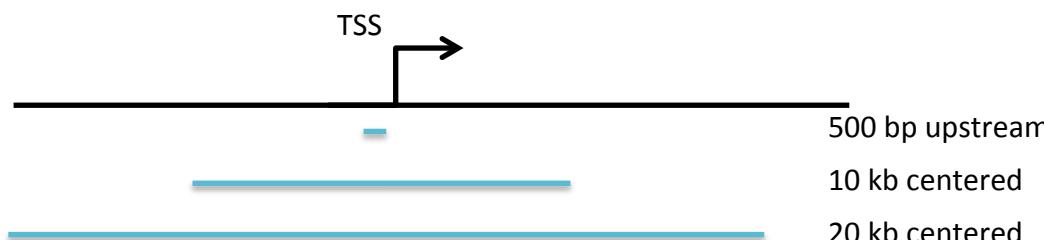
Track collection: 1120 ChIP-seq tracks (ENCODE raw signals)

Putative regulatory region: 10kb centered around TSS

Motif rankings database: 10kb centered around TSS (7 species)

Track rankings database: 10kb centered around TSS (ChIP-seq-derived)

Schematic of the iRegulon search space



- iRegulon only allows you to do gene-based analysis
- Choose the motif and track collections. Both have larger and smaller versions
- Select the size of your search space around the TSS of your input genes
- Select the motif ranking database to determine the conservation level
(7 species = only mammals)

Selecting the parameters: Recovery parameters

Recovery	
Enrichment score threshold:	<input type="text" value="3.0"/>
ROC threshold for AUC calculation:	<input type="text" value="0.03"/>
Rank threshold:	<input type="text" value="5000"/>
TF prediction	
Minimum identity between orthologous genes:	<input type="text" value="0.0"/>
Maximum false discovery rate (FDR) on motif similarity:	<input type="text" value="0.001"/>

- Each parameter in this section are set by default on optimized numbers.
- enrichment score determines the minimal NES a motif or track needs to have to be called
- Rank threshold will determine the maximal ranked position of a predicted target gene that will be reported for an enriched factor. Eg if a target gene is ranked 5001 then it will no longer be reported

When you are happy with the parameters you've selected click the submit button to run iRegulon

Exploring the results panel



- The iRegulon analysis will generate a result panel containing 3 different views
 - Motif view
 - Track view
 - Transcription view

Motifs view

Motifs Tracks Transcription Factors

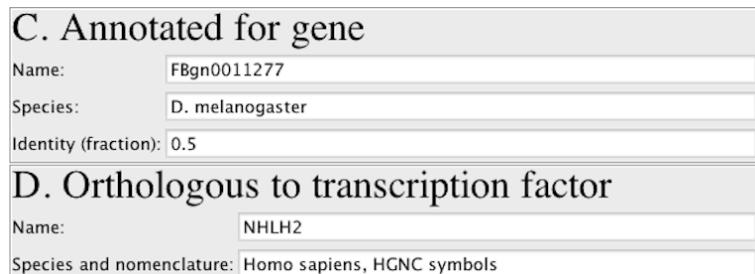
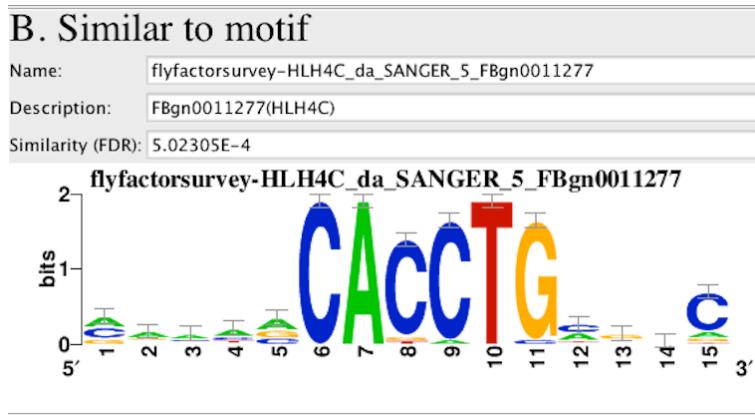
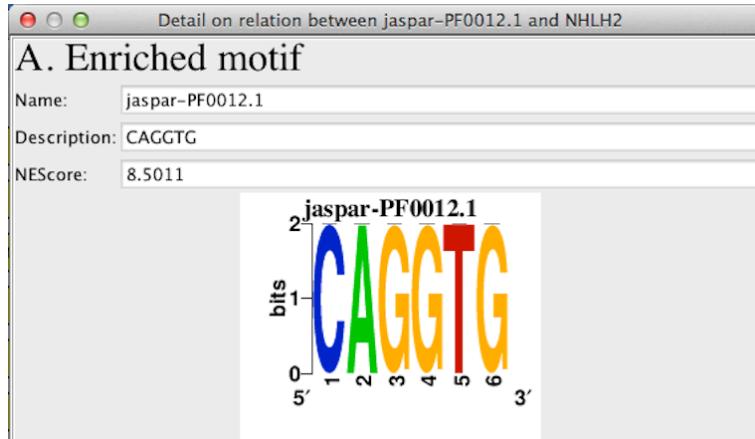
Filtered	Rank	Enriched Motif ID	NES	AUC	Clust...	#Targets	#TFS
✓	1	jaspar-PF0012.1	8.501	0.224	M1	15	1
✓	2	swissregulon-SNAI1..3.p2	8.430	0.222	M1	15	4
✓	3	homer-M00033	8.387	0.221	M1	17	10
✓	4	elemento-CAGGTGA	7.871	0.209	M1	8	0
✓	5	jaspar-MA0103.1	7.794	0.207	M2	17	1
✓	6	elemento-AGGTCCG	7.607	0.203	M3	8	0
✓	7	flyfactorsurvey-wor_SANGER_2.5_FBgn0...	7.554	0.202	M1	20	9
✓	8	jaspar-MA0086.1	7.514	0.201	M1	11	5
✓	9	swissregulon-ZEB1.p2	7.230	0.194	M1	20	4
✓	10	transfac_pro-M03565	7.054	0.190	M1	19	1
✓	11	flyfactorsurvey-wor_SOLEXA_2.5_FBgn0...	6.918	0.187	M1	9	8

Transcripti...	Orthologou...	Motif Simila...	Rank	Target Name
NHLH2	SE-1	5.023E-4	2	CLDN7
CAGGTG			9	CRB3
			12	TMEM30B
			87	SH3YL1
			89	INADL
			105	OCLN
			161	CDH1
			233	MAL2
			279	PKP3
			513	PPL
			644	SHROOM3
			948	GJB3
			1030	EPPK1
			1182	CDH3
			1246	MUC1

Transcription Fact... Orthologous Ident... Motif Similarity (F...
 NHLH2 SE-1 5.023E-4
[Detail...](#)

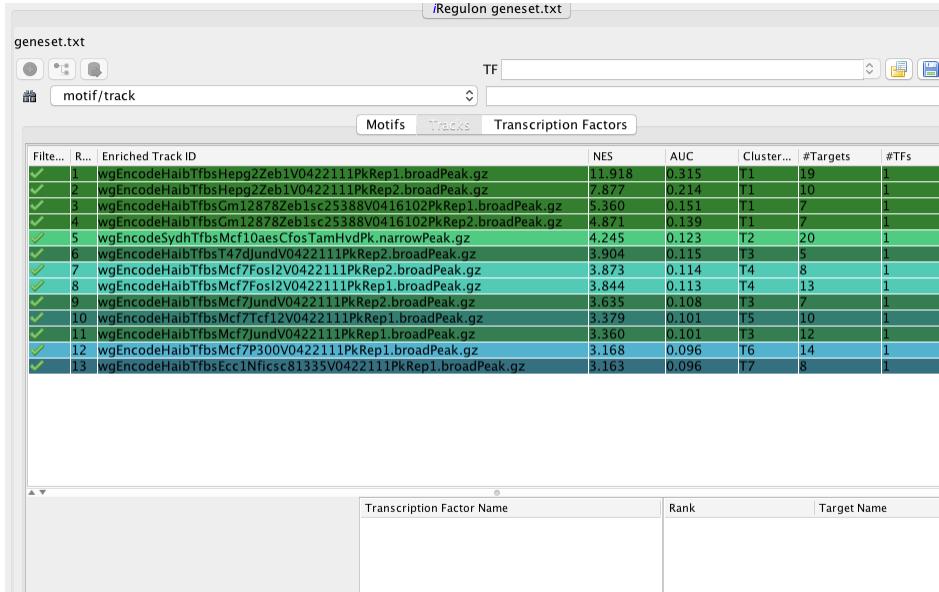
- Each motif found enriched is shown with following info
 - NES, the default ranking for the motifs
 - AUC
 - # predicted targets
 - # linked TF
- Similar motifs are clustered, shown by the cluster numbers (e.g. M1)
- When clicking a motif more info is shown in the panel below
- At the left are the predicted target genes (in red)
- At the right are the logo and name of the motif
- The middle shows you the TF potentially linked to the motif. When right clicking on a TF you can see in more detail the way TF and motif are linked

Intermezzo: Motif to TF determination



- The first detail panel will give you more information on the motif itself
 - If the motif is directly linked to a TF then this will be the only panel.
- If no direct TF exists then the most similar motif will be calculated and reported
 - This can be a motif from any of the other species
- The TF directly linked to the most similar motif is reported
- The ortholog of the annotated TF for human is reported in D using an orthologous scoring method.

Track view



Transcription Factor Name	Rank	Target Name
ZEB1	8	EPCAM
	10	MAL2
	12	CRB3
	20	F11R
	24	INADL
	43	CDH1
	65	MARVELD2
	106	PPL
ZEB1_(SC-25388)/HepG2/None	120	OCLN
	143	PKP3
	177	DSP
	363	SH3YL1
	474	SYTL1
	475	DSC2
	639	SFN
	769	CXADR
	884	CDH3
	916	DMKN
	1151	TACSTD2

- Each track found enriched is shown with following info
 - NES, the default ranking for the motifs
 - AUC
 - # predicted targets
 - # linked TF
- Similar tracks are clustered, shown by the cluster numbers (e.g. T1)
- When clicking a Track more info is shown in the panel below
 - At the left are the predicted target genes
 - At the right is the full track name and antibody and cell line information
 - The middle shows you the TF for which the track was generated. There is always only one TF possible

Transcription factor view

F...	TF	NES	#Targets	#Motifs/Tracks
✓	T1 ZEB1	11.918	20	4
✓	M2 SNAI2	7.796	28	24
✓	M3 IRF5	6.405	8	2
✓	M4 ZEB1	6.373	15	2
✓	M7 BACH2	4.562	24	11
✓	M8 CRHL1	4.433	18	3
✓	T2 FOS	4.245	20	1
✓	M9 DMRT1	4.057	5	1
✓	M.. CEBPG	4.016	10	6
✓	T3 JUND	3.904	15	3
✓	M.. MAFK	3.887	4	1
✓	T4 FOSL2	3.872	12	2



R...	Enriched Track ID	NES	AUC	C...	#Targets	#TFs	Filter	Transcription Factor Name	#Tracks	Rank	#Motifs	Target Name
1	wgEncodeHaibTfbsHegp22eb1v0422...	12.901	0.306	T1 21	1	1	✓	ZEB1	4	3	4	CRB3
2	wgEncodeHaibTfbsHegp22eb1v0422...	7.707	0.189	T1 8	1	1		SNAI2	3	5	2	EPCAM
3	wgEncodeHaibTfbsGm12878Zeb1sc2...	6.193	0.155	T1 7	1	1			7	2		MAL2
4	wgEncodeHaibTfbsGm12878Zeb1sc2...	5.689	0.143	T1 11	1	1			15	4		FIIR

R...	Enriched Motif ID	NES	AUC	C...	#Targets	#TFs	Filter	Transcription ...	#Motifs	Ortholo...	Motif Si...	Rank	#Motifs	Target Name
1	jaspar-PF0012_1	8.501	0.224	M1 15	4	1	✓	NHLH2	4	SE-1	5.023E-4	1	33	CLDN7
2	swissregulon-SNAI1..3.p2	8.430	0.222	M1 15	4	1		SNAI1	1	N/A	Direct	1	37	CRB3
3	homter-M00033	8.387	0.221	M1 17	10	1		SNAI2	3	N/A	Direct	3	30	TMEM30B
4	elemento-CAGGTGA	7.871	0.209	M1 8	0	1		SNAI3	1	N/A	Direct	8	33	PKP3
7	flyfactorsurvey-wor_SANGER_2.5_F...	7.554	0.202	M1 20	9	1		TCF3	1	N/A	Direct	11	17	INADL
8	jaspar-MA0086.1	7.514	0.201	M1 11	5	1		TCF4	1	N/A	1.918E-6	12	24	SHROOM3
9	swissregulon-ZEB1.p2	7.230	0.194	M1 20	4	1		MESP1	1	N/A	1.148E-5	19	11	SFN
10	transfac_pro-M03565	7.054	0.190	M1 19	1	1		ID4	1	N/A	1.64E-5	20	37	OCLN
11	flyfactorsurvey-wor_SOLEXA_2.5_F...	6.918	0.187	M1 9	8	1		LMO2	2	N/A	1.377E-4	46	29	MAL2
12	stark-CAGGTG	6.867	0.186	M1 17	4	1		MYO1	1	N/A	5.448E-4	59	19	SHSYL1
13	flyfactorsurvey-_1_sc_da_SANGER...	6.665	0.181	M1 11	3	1		FIGLA	1	N/A	5.46E-4	75	17	PMEPA1

- If you want a more general overview you can use the transcription factor view
- Here, similar tracks or motifs belonging to the same cluster will be given as one output with the following info
 - the highest NES of the cluster
 - total number of unique targets combined from all single features of the cluster
 - The total number of 'single features' within each cluster
- When clicking a cluster you will get more information below similar to the track or motif view e.g. detailed names of each event of a cluster
- For motifs it will also give you all TFs each motif of that cluster was linked to. When clicking each motif in a cluster it will indicate the corresponding TFs with a green tick

Search function in the results panel

ZEB1_genes.txt

TF TCF12

motif/track transcription factor target

F...	C...	TF	NES	#Targets	#Motifs/Tracks
✓	T1	ZEB1	12.901	23	4
✓	M1	NHLH2	8.501	26	37
✓	M2	ZEB1	7.794	27	4
✓	M5	IRF5	5.265	13	4
✓	M6	ZNF655	5.023	20	4
✓	M7	NANOG	4.609	11	3
✓	M8	ZNF423	4.325	19	4
✓	M9	ESO	4.177	9	2
✗	T2	TCF12	3.966	17	2
✗	M...	DMRT1	3.946	8	2
✗	M...	CEBPE	3.909	12	8

ZEB1_genes.txt

TF ZEB1

transcription factor

Motifs Tracks Transcription Factors

F...	C...	TF	NES	#Targets	#Motifs/Tracks
✓	T1	ZEB1	12.901	23	4
✗	M1	NHLH2	8.501	26	37
✓	M2	ZEB1	7.794	27	4
✗	M5	IRF5	5.265	13	4
✗	M6	ZNF655	5.023	20	4
✗	M7	NANOG	4.609	11	3
✗	M8	ZNF423	4.325	19	4
✗	M9	ESO	4.177	9	2
✗	T2	TCF12	3.966	17	2
✗	M...	DMRT1	3.946	8	2
✗	M...	CEBPE	3.909	12	8

ZEB1_genes.txt

TF CDH1

target

Motifs Tracks Transcription Factors

F...	C...	TF	NES	#Targets	#Motifs/Tracks
✓	T1	ZEB1	12.901	23	4
✓	M1	NHLH2	8.501	26	37
✓	M2	ZEB1	7.794	27	4
✗	M5	IRF5	5.265	13	4
✓	M6	ZNF655	5.023	20	4
✓	M7	NANOG	4.609	11	3
✓	M8	ZNF423	4.325	19	4
✓	M9	ESO	4.177	9	2
✗	T2	TCF12	3.966	17	2
✓	M...	DMRT1	3.946	8	2
✗	M...	CEBPE	3.909	12	8

- You can search for a motif or track, a transcription factor or a target

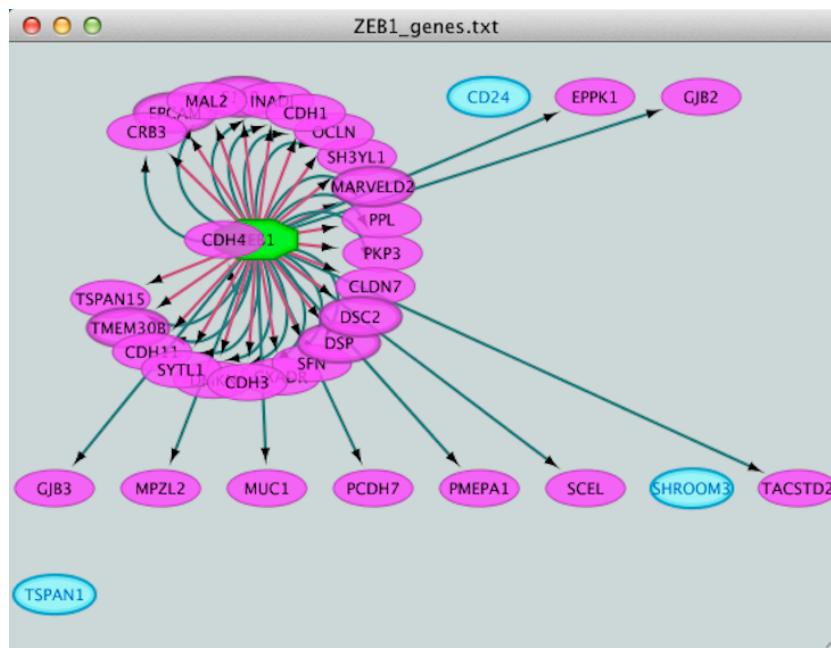
- Each feature that contains your search function will be indicated by a green tick

Making a network

The screenshot shows a software interface titled "Regulon geneset.txt". At the top, there are buttons for "geneset.txt" (with a red arrow pointing to it), "motif/track" (with a red arrow pointing to it), and "TF TBX5". Below this is a dropdown menu set to "motif/track". There are three tabs at the bottom: "Motifs", "Tracks", and "Transcription Factors" (which is selected). The main area is a table with the following columns: "...", "Clust...", "TF", "NES", "#Targets", and "#Motifs/Tracks". The data rows are:

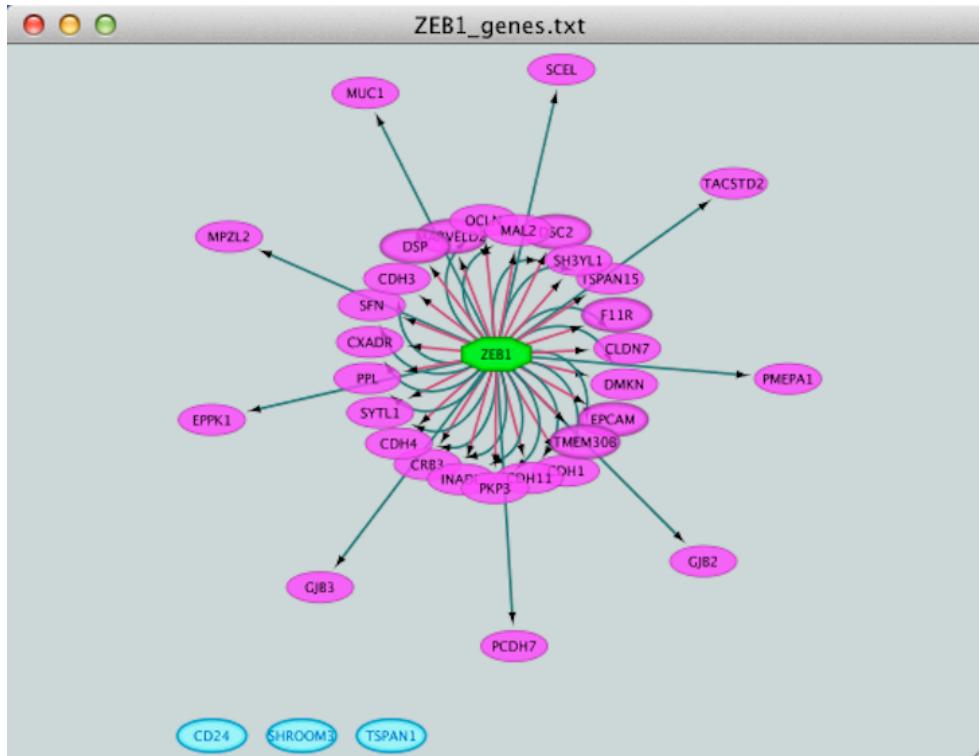
...	Clust...	TF	NES	#Targets	#Motifs/Tracks
✓	T1	ZEB1	11.918	20	4
✓	M2	SNAI2	7.796	28	24
✓	M3	IRF5	6.405	8	2
✓	M4	ZEB1	6.373	15	2
✓	M7	BACH2	4.562	24	11
✓	M8	GRHL1	4.433	18	3
✓	T2	FOS	4.245	20	1
✓	M9	DMRT1	4.057	5	1
✓	M10	CEBPG	4.016	10	6
✓	T3	JUND	3.904	15	3
✓	M11	MAFK	3.887	4	1

- The icon of the green button with the white cross inside will allow you to add the predicted regulator and draw a network with its predicted target genes in the existing Cytoscape network
- If you want to make a new network in Cytoscape by pressing the icon next to it (red arrow)



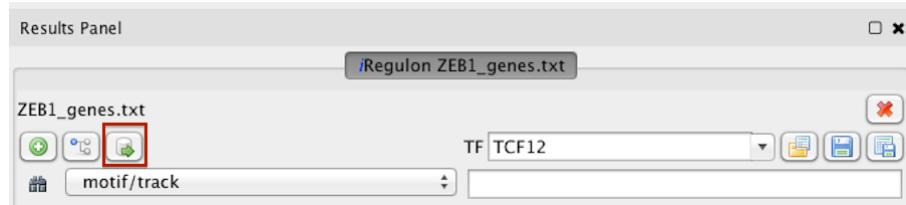
- For instance you can select cluster T1 with TF ZEB1 and click the button to generate a network
- You can add other results to your network, for instance cluster M4
- the edge color in the network will reflect the cluster the target was predicted as target
- Notice that in our example, three input genes are not predicted as target by either T1 or M4

Making a network



- You can give your network a layout using Cytoscape
- Go to layout > edge-weighted Spring Embedded Layout > all nodes > cluster color

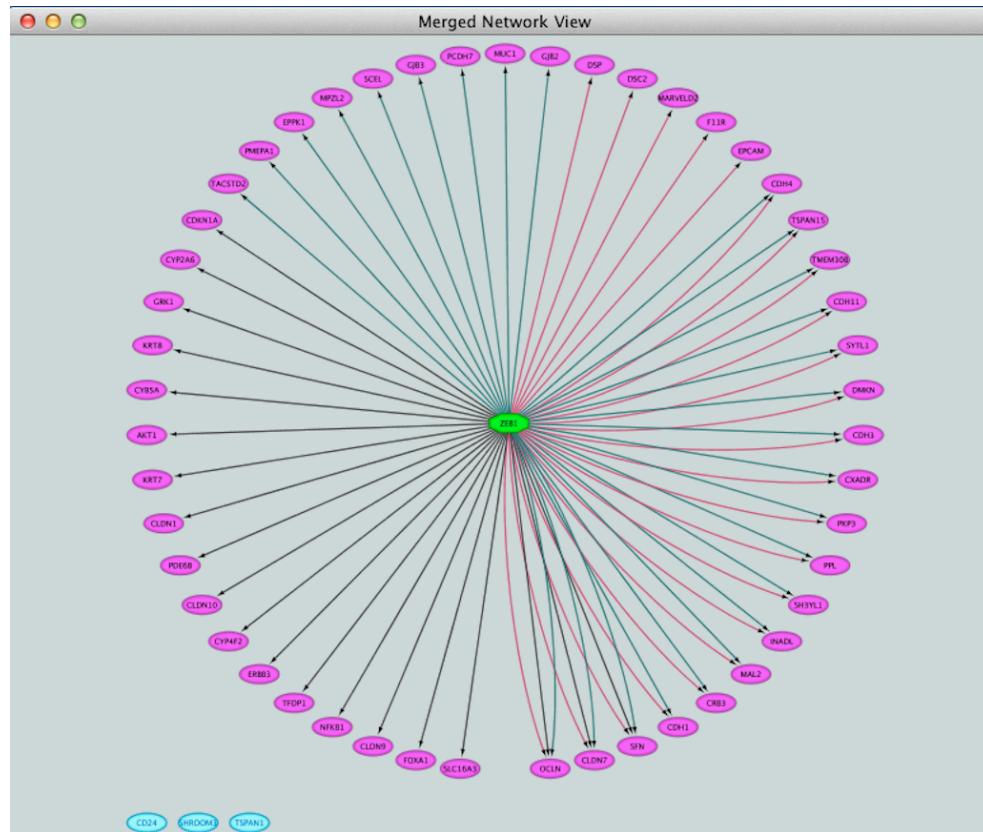
Metatargetome

A screenshot of a dialog box titled "Query TF-target database for a factor". It has several input fields and dropdown menus:

- "Transcription Factor:" dropdown set to "ZEB1"
- "Species and Gene nomenclature:" dropdown set to "Homo sapiens, HGNC symbols"
- "Database" dropdown menu with three options: "MSigDB", "GeneSigDB", and "Ganesh Clusters", with "MSigDB" selected.
- "Databases:" dropdown menu with the same three options, also with "MSigDB" selected.
- "Occurrence count threshold:" input field set to "5"
- "Network" section:
 - "Number nodes (approx.):" input field set to "20"
 - A checked checkbox labeled "Create new network"
 - "Attribute name:" dropdown menu set to "name"
- "Cancel" and "Submit" buttons at the bottom

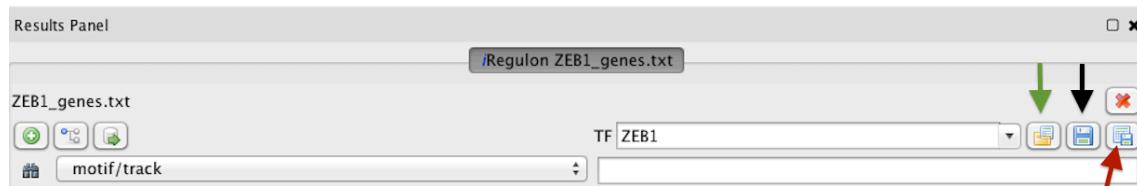
- Go to Apps > iRegulon > query TF-target database
- Or click the icon depicting a cilinder with a green arrow
- Select the TF that you want to query
- Indicate which of the three databases you want to include in the analysis
- You can adapt the number of times a target is found in a signature. Default is set at 5. Lowering the threshold heightens the chance of false positives.
- You can determine the number of reported targets. Default is set at 20
- Indicate whether a new network needs to be created. If you don't, only those targets that are in your existing network will be reported back

Metatargetome



- This is a network obtained by taking the union of the previously created network and the metatargetome for ZEB1.
- Only three genes overlap
- 17 are metatargetome predicted only
- 29 are predicted by cluster T1 and/or M2

Exporting and saving an analysis



- You can export the results of your iRegulon analysis as a tab delimited table (icon at red arrow)
- You can also save your iRegulon analysis as a session. This will generate an .irf file (icon at the black arrow)
- If you want to reload previous results click the icon indicated by the green arrow

Important! Saving the iRegulon results will not save the networks you made in cytoscape. To do so you need to save the cytoscape session (.cys). Likewise, the saved cytoscape session will not save your iRegulon results.

BONUS SECTION

Analyzing gene expression data using *i-cisTarget*

Analysis of a ZEB1 perturbation data set

Loading the data in i-cisTarget

Go to: Help | Examples | New Features | Input Formats | Old version

i-cisTarget

An integrative genomics method for the prediction of regulatory features and cis-regulatory modules in Human, Mouse, and Fly.

2. Paste list of gene IDs or peak locations
Please do not enter a single gene nor peak here, but a list of related genes or peaks.
An explanation on supported input formats can be found [here](#).

Or choose a file to upload:

Choose File No file chosen

2. Paste list of gene IDs or peak locations
Please do not enter a single gene nor peak here, but a list of related genes or peaks.
An explanation on supported input formats can be found [here](#).

Or choose a file to upload:

Choose File ZEB1_genes.txt

- Go to the i-cisTarget website
[https://gbimed.kuleuven.be/apps/lcb/
i-cisTarget-NAR/](https://gbimed.kuleuven.be/apps/lcb/i-cisTarget-NAR/)
- Go to section 2
- Paste a list of region coordinates or gene symbols
- Or choose a file with your gene region coordinates or gene symbols

Setting the parameters

3. For which species?

Homo sapiens (hg19) ▾

- Select your species

4. Input type

Please make sure that the selected input type corresponds to the data pasted in the step 2.

Gene symbols (HGNC symbols) ▾

- Indicate in step 4 which file type you used as input

note: iCisTarget region IDs are generally only obtained from a previous analysis

5. Which database version?

Version 3.0 of databases ▾

6. What features do you want to analyze?

Approx. time for the analysis using:

- only PWM database: 3-4 mins
- only TFBS ChIP database: 1 min
- all databases: 5-6 mins

PWMs (9713)
TF binding sites (1394)
Histone modifications (2003)
DHS & FAIRE (908)

- In Step 6 you can select one or more features for the analysis

New here are the histone modification tracks , DHS and FAIRE tracks

- Here we will select PWMs and TF-binding sites to best compare with our previous iRegulon analysis

- Finally you can give your job a name

- If you want to be notified when the results are ready you can add your email address

7. Give a name for your job

ZEB1_analysis

8. Your E-mail address (optional)

[]

1. Analysis type

Quick analysis (~ 20 s) - only JASPAR motif collection.

Full analysis (~ 6 mins) - whole motif collection.

Full analysis ▾

- Don't forget to choose what type of analysis you want to do in step 1. A fast analysis will only use motifs coming from the Jaspar motif collection and thus limit your results

Setting the parameters: optional settings

Optional parameters

Region mapping	10kb upstream, TSS, 10kb downstream
Fraction of overlap	0.4
Normalized enrichment score (NES) threshold	3.0
Enrichment analysis	Within each database separately
ROC threshold for AUC calculation	0.005
Threshold for visualization	20000

- There are a number of optional parameters below the submit button you can alter if wanted
- The most important to consider is whether you want the enrichment score calculated within each selected feature (motif, track, ...) or across all selected.

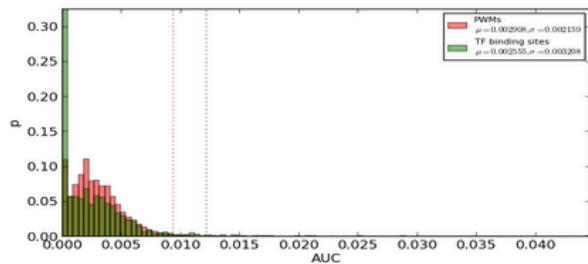
When happy with your parameters click the submit button to run the analysis

Exploring the results page

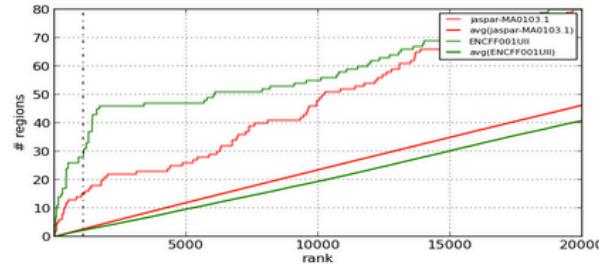
Parameters and statistics for ZEB1_analysis

Number of features	11107
Number of enriched features (NES > 3.0)	129
Total number of ranked regions	220330
Type of input query	hgnc_symbol
Fraction of mapped input IDs	0.972
Number of regions in input set	472
Normalized enrichment score (NES) threshold	3.0
AUC threshold (fraction / # of ranked regions)	0.005 (1102)
Recovery curve threshold (# of regions)	20000

AUC distribution



Recovery of best feature



- The first section shows you some statistics
- Additionally it gives you the AUC and ROC

Exploring the results page: statistics and input

Number of features	11107
Number of enriched features (NES > 3.0)	129

- Number of features will tell you how many you queried based on the databases you selected (only PWM and TF-binding sites in this example)
- Below are the enriched features

Total number of ranked regions	220330
Type of input query	hgnc_symbol
Fraction of mapped input IDs	0.972
Number of regions in input set	472

- All the regions in the human genome are counted under ‘The total number of ranked regions’
- When converting input genes into regions, 97% was mapped
- All genes from the input resulted in 472 regions in total

Normalized enrichment score (NES) threshold	3.0
AUC threshold (fraction / # of ranked regions)	0.005 (1102)
Recovery curve threshold (# of regions)	20000

- Finally some used settings are mentioned

Exploring the results page: converting your input to regions

Number of regions in input set

472

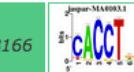
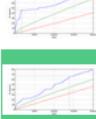
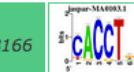
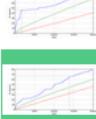
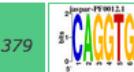
```
chr1-reg109906  
chr1-reg109907  
chr1-reg109908  
chr1-reg109909  
chr1-reg109910  
chr1-reg109911  
chr1-reg109916  
chr1-reg109917  
chr1-reg109919  
chr1-reg109920  
chr1-reg115221  
chr1-reg115223  
chr1-reg115225  
chr1-reg115227  
chr1-reg115228  
chr1-reg115230  
chr1-reg115231  
chr1-reg115232  
chr1-reg115233  
chr1-reg115235  
chr1-reg115237  
chr1-reg115239
```

- For i-cisTarget to work your input genes or regions will be converted to i-cisTarget regions.
- i-cisTarget will give you the corresponding region IDs of your converted input
- You can access and retrieve the region IDs of your input by clicking on the ‘regions in input set’
- You will get a list of region IDs
- You can save this and use this later as input in another i-cisTarget analysis (using input type ‘region ID input’)
- You can also obtain the bed file with all the i-cisTarget regions across the full genome

[Download a bed file with candidate regulatory regions](#)

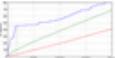
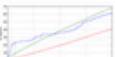
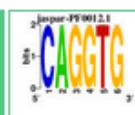
-
- [Human CRRs](#)
 - [Mouse CRRs](#)
 - [Fly CRRs](#)

Exploring the results page: enriched features

#	Feature	NES	Logo	Recovery Curve	Candidate regions targets	All regions in top 20000	Database
1	<input type="checkbox"/> ENCFF001UII Description: ZEB1 ChIP-seq protocol v042211.1 on human HepG2	13.10846			link	link	TF binding sites
2	<input type="checkbox"/> jaspar-MA0103.1 Description: ZEB1 Possible TFs: ZEB1	8.98166			link	link	PWMs
3	<input type="checkbox"/> ENCFF001UNB Description: JUND ChIP-seq protocol v042211.1 on human MCF-7	8.20466			link	link	TF binding sites
4	<input type="checkbox"/> jaspar-PF0012.1 Description: CAGGTG Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	7.61379			link	link	PWMs
5	<input type="checkbox"/> swissregulon-SNAI1..3.p2 Description: SNAI2,SNAI1,SNAI3 Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	7.45884			link	link	PWMs
6	<input type="checkbox"/> stark-CAGGTG Description: CAGGTG(sna) Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	7.27450			link	link	PWMs
7	<input type="checkbox"/> ENCFF001UIJ Description: ZEB1 ChIP-seq protocol v042211.1 on human HepG2	7.05767			link	link	TF binding sites
8	<input type="checkbox"/> transfac_pro-M03565 Description: V\$SLUG_Q6 Possible TFs: SNAI2	6.77668			link	link	PWMs
9	<input type="checkbox"/> transfac_pro-M02769 Description: V\$IRF5_03 Possible TFs: IRF3, IRF6, IRF5, IRF4	6.69208			link	link	PWMs
10	<input type="checkbox"/> jaspar-MA0086.1 Description: sna Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	6.63153			link	link	PWMs

When you scroll down the results page you will find your enriched features. They will be ordered by enrichment score with tracks and motifs mixed together.

Exploring the results page: enriched features

#	Feature	NES	Logo	Recovery Curve	Candidate targets	All regions in top 20000	Database
1	<input type="checkbox"/> ENCFF001UII Description: ZEB1 ChIP-seq protocol v042211.1 on human HepG2	13.10846			link	link	TF binding sites
2	<input type="checkbox"/> jaspar-MA0103.1 Description: ZEB1 Possible TFs: ZEB1	8.98166			link	link	PWMs
3	<input type="checkbox"/> ENCFF001UNB Description: JUND ChIP-seq protocol v042211.1 on human MCF-7	8.20466			link	link	TF binding sites
4	<input type="checkbox"/> jaspar-PF0012.1 Description: CAGGTG Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	7.61379			link	link	PWMs

- Each feature is accompanied by their
 - NES
 - a logo in case of a motif
 - a recovery curve
 - the database the feature belongs to (here TF-binding site (track) or PWMs (motif))
- Additionally, two different links are provided
 - One link will lead you to the predicted candidate target regions
 - The second link will lead you to all regions ranked/enriched in the top 20000

Exploring the results page: enriched features

#	Feature	NES	Logo	Recovery Curve	Candidate regions targets	Candidate regions in top	All Database
1	<input type="checkbox"/> ENCF001UII Description: ZEB1 ChIP-seq protocol v042211.1 on human HepG2	13.10846			link	link	TF binding sites 20000
2	<input type="checkbox"/> jaspar-MA0103.1 Description: ZEB1 Possible TFs: ZEB1	8.98166			link	link	PWMs
3	<input type="checkbox"/> ENCF001UNB Description: JUND ChIP-seq protocol v042211.1 on human MCF-7	8.20466			link	link	TF binding sites
4	<input type="checkbox"/> jaspar-PF0012.1 Description: CAGGTG Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	7.61379			link	link	PWMs

Motifs	Tracks	Transcription Factors
Filtered Rank	Enriched Track ID	NES AUC Clust. #Targets #TFs
1	wgEncodeHaibTfbsHepg2Zeb1V042211.1...	12.901 0.306 T1 21 1
2	wgEncodeHaibTfbsHepg2Zeb1V042211.1...	7.707 0.189 T1 8 1
3	wgEncodeHaibTfbsGm12878Zeb1sc2538...	6.193 0.155 T1 7 1
4	wgEncodeHaibTfbsGm12878Zeb1sc2538...	5.689 0.143 T1 11 1
5	wgEncodeHaibTfbsMcf7Tcf12V042211P...	3.966 0.105 T2 7 1
6	wgEncodeSydhTfbsMcf1OaesCfosTamHv...	3.573 0.096 T3 4 1
7	wgEncodeHaibTfbsMcf7Fosl2V042211P...	3.513 0.094 T4 6 1
8	wgEncodeHaibTfbsH1hescTcf12Pcr1xPk...	3.412 0.092 T2 13 1
9	GSM1208642_batch1_chrom1_LoVo_KLF...	3.162 0.086 T5 16 1
10	wgEncodeSydhTfbsMcf7Gata3sc269UcdP...	3.031 0.084 T6 7 1

Motifs	Tracks	Transcription Factors
Filtered Rank	Enriched Motif ID	NES AUC Clust. #Targets #TFs
1	Jaspar-PF0012.1	8.501 0.224 M1 15 1
2	swissregulon-SNAI1..p2	8.430 0.222 M1 15 4
3	homer-M00033	8.387 0.221 M1 17 10
4	elemento-CAGGTGA	7.871 0.209 M1 8 0
5	Jaspar-MA0103.1	7.794 0.207 M2 17 1
6	elemento-AGGTCCG	7.607 0.203 M3 8 0
7	flyfactorsurvey-wor_SANGER_2.5_F8gn0...	7.554 0.202 M1 20 9
8	Jaspar-MA0086.1	7.514 0.201 M1 11 5
9	swissregulon-ZEB1.p2	7.230 0.194 M1 20 4
10	transfac_pro-M03565	7.054 0.190 M1 19 1
11	flyfactorsurvey-wor_SOLEXA_2.5_F8gn0...	6.918 0.187 M1 9 8

- The first ranked track was also the first ranked track we found with iRegulon

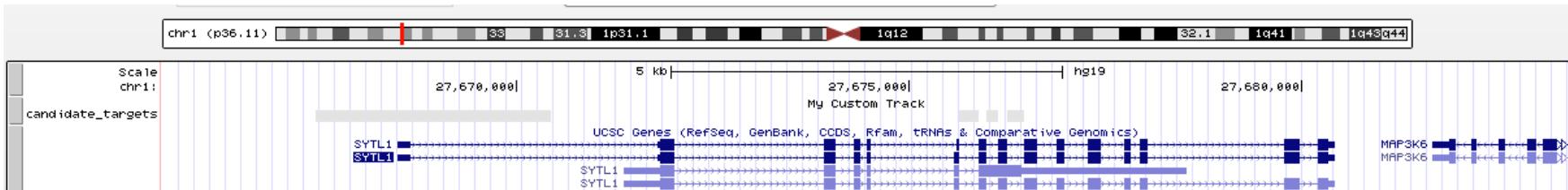
- The First and second ranked motifs were found 5th and 1st in iRegulon

Exploring the results page: enriched features



Rank	Region ID	Associated genes
8	chr2-reg39368	EPCAM
9	chr2-reg39369	EPCAM
10	chr19-reg7778	CRB3
12	chr19-reg7777	CRB3
39	chr8-reg74725	MAL2
40	chr8-reg74727	MAL2
48	chr1-reg62431	INADL
71	chr1-reg115232	F11R
72	chr1-reg115230	F11R
73	chr1-reg115231	F11R
92	chr5-reg40143	MARVELD2
145	chr16-reg46514	CDH1
146	chr16-reg46513	CDH1
147	chr16-reg46515	CDH1
254	chr6-reg7603	DSP
302	chr11-reg292	PKP3
303	chr11-reg291	PKP3
437	chr1-reg29919	SYTL1
438	chr1-reg29922	SYTL1
439	chr1-reg29920	SYTL1
440	chr1-reg29921	SYTL1
443	chr16-reg6224	PPL
444	chr16-reg6222	PPL
445	chr16-reg6223	PPL

- If you click the link to the predicted target regions for the first track you will find 46 target regions. However these regions can be associated to 18 unique gene names.
- If you scroll down you find a link to get the target regions as a bed file and to view them in the UCSC browser



Exploring the results page: enriched features

Feature

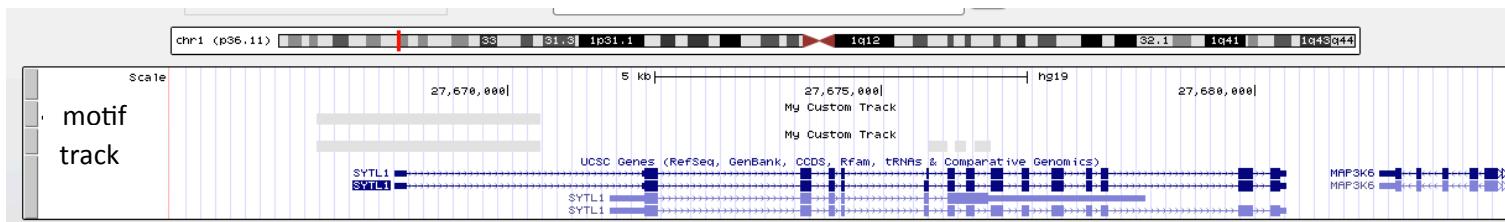
2	<input type="checkbox"/> jaspar-MA0103.1 Description: ZEB1 Possible TFs: ZEB1
---	---

Rank	Region ID	Associated genes
6	chr14-reg24448	TMEM30B
27	chr17-reg7862	CLDN7
79	chr2-reg192	SH3YL1
80	chr5-reg40200	OCLN
93	chr1-reg62431	INADL
156	chr1-reg29909	SYTL1
285	chr1-reg109908	MUC1
288	chr8-reg74725	MAL2
349	chr19-reg30388	DMDK
358	chr16-reg6223	PPL
415	chr8-reg74727	MAL2
444	chr1-reg115231	F11R
519	chr8-reg96174	EPPK1
828	chr13-reg36065	SCEL
1020	chr8-reg96176	EPPK1
1147	chr11-reg291	PKP3
1299	chr16-reg46396	CDH3
1389	chr5-reg40143	MARVELD2
1768	chr20-reg48074	PMEPA1
1779	chr1-reg49519	TSPAN1
1887	chr11-reg299	PKP3
2000	chr18-reg17905	DSC2

3102	chr20-reg48080	PMEPA1
4290	chr5-reg40193	OCLN
4377	chr6-reg7585	DSP
4845	chr19-reg7780	CRB3
5240	chr4-reg44289	SHROOM3
5409	chr11-reg82414	MPZL2
5856	chr16-reg6229	PPL
6151	chr1-reg109917	MUC1
6274	chr6-reg7603	DSP
6306	chr10-reg44975	TSPAN15
6728	chr10-reg44972	TSPAN15
6731	chr20-reg48062	PMEPA1
6896	chr16-reg6222	PPL
6950	chr16-reg42573	CDH11
7245	chr16-reg6217	PPL
7364	chr20-reg48091	PMEPA1
7371	chr16-reg6235	PPL
7582	chr1-reg29415	SFN
8078	chr20-reg48060	PMEPA1
9316	chr16-reg42571	CDH11
9444	chr16-reg46513	CDH1
9554	chr13-reg782	GJB2



- If you click on the link to the predicted target regions for the first motif you'll find 66 target regions. However these regions can be associated to 33 unique gene names.
- If you scroll down you find a link to get the target regions as a bed file and to view them in the UCSC browser
- In the example below we can see the predicted target regions for both the first track and motif



Setting the parameters: loading ChIP peaks

Application on TF ChIP-seq and MSigDB gene sets

[See results](#)

wgEncodeSydhTfbsK562Yy1UcdPk	YY1	Peaks	1 (1)	15.12	Report	Peaks	1 (1)	9.47	Report
wgEncodeHaibTfbsGm12878Zeb1sc25388V0416102PkRep1	ZEB1	Peaks	2 (2)	9.33	Report	Peaks	57	4.18	Report
wgEncodeSydhTfbsHek293tZnf263UcdPk	ZNF263	Peaks	1 (1)	18.2	Report	Peaks	low	3.03	Report

- In order to show how to run i-cisTarget with ChIP peaks as input we will load a ZEB1 ChIP peak file
- There are many bed files already available on the website as examples

Supply regions or genes

1. Analysis type
Quick analysis (~ 20 s) - only JASPAR motif collection.
Full analysis (~ 6 mins) - whole motif collection.

2. Paste a list of region coordinates or gene symbols
Please do not enter a single region nor gene here, but a list of related peaks/regions or genes. An explanation on supported input formats can be found [here](#).

Or choose a file to upload:

Choose File wgEncodeHaib....top500.bed

3. For which species?
Homo sapiens (hg19)

4. Input type
Please make sure that the selected input type corresponds to the data pasted in the step 2.

5. Which database version?
Version 3.0 of databases

6. What features do you want to analyze?

Approx. time for the analysis using:
• only PWM database: 3-4 mins
• only TFBS ChIP database: 1 min
• all databases: 5-6 mins

PWMs (9713)
TF binding sites (1394)
Histone modifications (2003)
DHS & FAIRE (908)

7. Give a name for your job
ZEB1_ChIP_Analysis

8. Your E-mail address (optional)

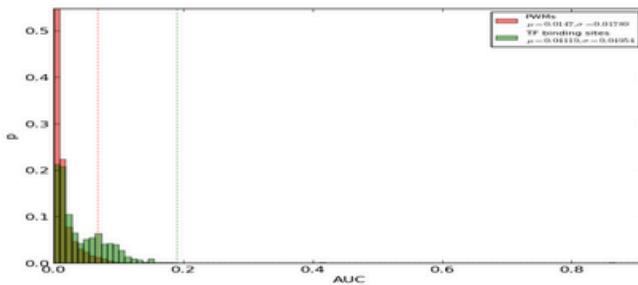
- You can download and load the ZEB1 top 500 ChIP peaks as input for a new analysis
- We will apply the same parameters as we did before, with the genes

Exploring the results page

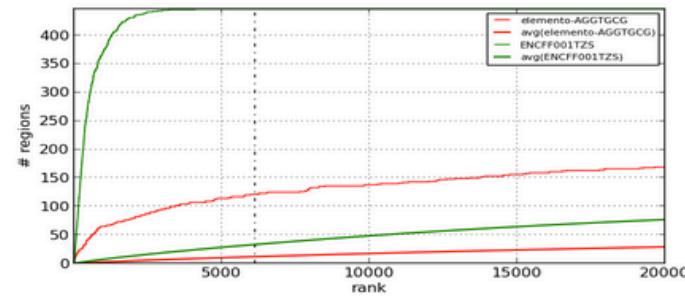
Parameters and statistics for ZEB1_ChIP_Analysis

Number of features	11107
Number of enriched features (NES > 3.0)	243
Total number of ranked regions	1223024
Type of input query	bed
Number of regions in input set	446
Normalized enrichment score (NES) threshold	3.0
AUC threshold (fraction / # of ranked regions)	0.005 (6115)
Recovery curve threshold (# of regions)	20000

AUC distribution

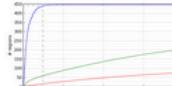
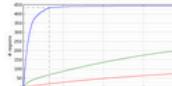
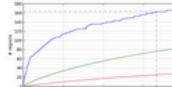
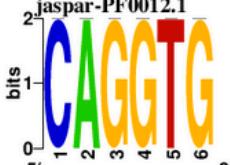
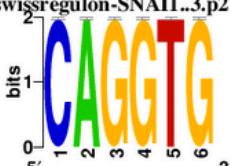


Recovery of best feature



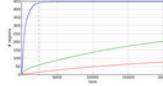
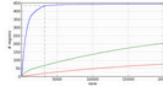
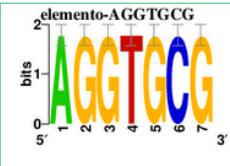
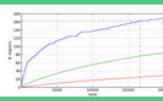
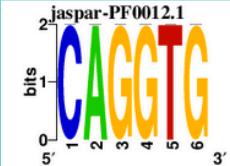
- The first section again shows you some statistics and the AUC and ROC
- The number of regions in the input set is now 446, which is less than the gene-based analysis (472)
- The number of enriched features is 243 which is higher than the gene-based analysis (129)

Exploring the results page: enriched features

#	Feature	NES	Logo	Recovery Curve	All Candidate regions targets in top 20000	Database
1	<input type="checkbox"/> ENCF001T0S Description: ZEB1 ChIP-seq protocol v041610.2 on human GM12878	17.59420			link	link TF binding sites
2	<input type="checkbox"/> ENCF001T0T Description: ZEB1 ChIP-seq protocol v041610.2 on human GM12878	16.65790			link	link TF binding sites
3	<input type="checkbox"/> elemento-AGGTGCG Description: AGGTGCG	10.30204	 bits: 2, 1, 0		link	link PWMs
4	<input type="checkbox"/> jaspar-PF0012.1 Description: CAGGTG Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	9.97555	 bits: 2, 1, 0		link	link PWMs
5	<input type="checkbox"/> swissregulon-SNAI1..3.p2 Description: SNAI2,SNAI1,SNAI3 Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	9.69897	 bits: 2, 1, 0		link	link PWMs

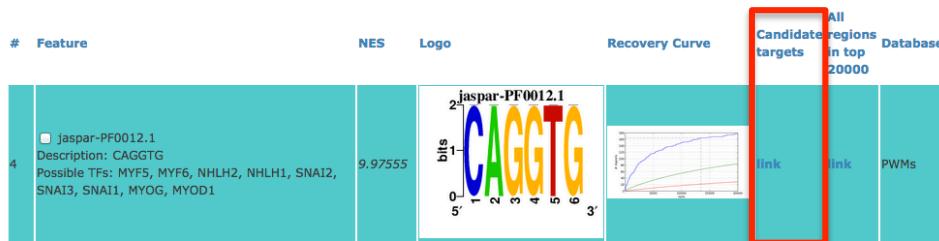
When you scroll down you will find your enriched features. There are again ordered by enrichment score with tracks and motifs mixed together.
 The layout is exactly the same as before

Exploring the results page: enriched features

#	Feature	NES	Logo	Recovery Curve	Candidate regions targets	link	Candidate regions in top 20000	Database
1	<input type="checkbox"/> ENCFF001Tzs Description: ZEB1 ChIP-seq protocol v041610.2 on human GM12878	17.59420			link	link	TF binding sites	
2	<input type="checkbox"/> ENCFF001Tzt Description: ZEB1 ChIP-seq protocol v041610.2 on human GM12878	16.65790			link	link	TF binding sites	
3	<input checked="" type="checkbox"/> elemento-AGGTGCG Description: AGGTGCG	10.30204	 bits: 0, 1, 2 5' A G G T G C G 3'		link	link	PWMs	
4	<input checked="" type="checkbox"/> jaspar-PF0012.1 Description: CAGGTG Possible TFs: MYF5, MYF6, NHLH2, NHLH1, SNAI2, SNAI3, SNAI1, MYOG, MYOD1	9.97555	 bits: 0, 1, 2 5' C A G G T G 3'		link	link	PWMs	

- The first two features are the ChIP-seq data you used as input itself
- On the fourth position (2^{nd} motif) we find the motif that we previously found 1^{st} with iRegulon and 2^{nd} with the genes as input in i-cisTarget
- You can click the candidate targets to see which are the predicted target regions and associated genes

Exploring the results page: enriched features



Rank	Region ID	Associated genes
3	chr11-reg686	DEAF1
4	chr13-reg57678	GRK1
6	chr14-reg33456	RGS6
9	chr19-reg7776	SLC25A23
14	chr17-reg7862	CLDN7
20	chr19-reg20732	CRTC1
21	chr16-reg67438	GAS8
40	chr19-reg39586	MARK4
61	chr19-reg32461	SPINT2
71	chr1-reg29523	C1orf172
72	chr16-reg62606	LOC732275
79	chr18-reg45483	CYB5A
88	chr1-reg79605	LRRC8B
100	chr1-reg3633	WDR8
125	chr12-reg96436	FBRSL1
132	chr7-reg34530	ZMIZ2
177	chr19-reg14134	HOOK2
225	chr2-reg6714	C2orf46
227	chr1-reg25686	MDS2

- You will find 165 regions predicted as target for this motif associated with 158 unique genes
- This is much more then the results retrieved in the gene-based analysis, which yielded 45 regions with 27 unique associated genes
- i-cisTarget has a feature that will allow you to actually compare both analyses

Comparing results

Comparative analysis

This tool allows comparison of motif enrichment results of 2 independent i-cisTarget analyses. In this analysis, it is possible to compare both results for the same species as well as for different species (if the same motif collection is used), e.g. enriched motifs found for active regions in mouse and drosophila heart.

The comparative analysis can be performed [here](#).



INPUT:

Input IDs of your previously run analyses that you aim to compare or the whole link to the reports. Please, make sure that your submitted analyses were performed on the same motif collection.

i-cisTarget results 1	<input type="text"/>
i-cisTarget results 2	<input type="text"/>
NES threshold	3.0 <input type="button" value=""/>

Warning: Mouse and human is not good to compare, because it is a bit circular.

- On the main page you will find a link to the comparison analysis
- You will need to input the run IDs or the full html link of your two analyses. Here we will compare

ZEB1 genes-based analysis <https://gbiomed.kuleuven.be/apps/lcb/i-cisTarget-NAR/reports/5518ef86d44b9868ada0314d37592d0506b46100/report.html>
ZEB1 ChIP peak analysis <https://gbiomed.kuleuven.be/apps/lcb/i-cisTarget-NAR/reports/66c29a9d9809dd9e77c74dc2a0935d01fffbfe7/report.html>

Comparing results

Statistics

Number of features with NES ≥ 3.0 detected in:

at least one of the analyses	316
both analyses	31
only the analysis number 1	76
only the analysis number 2	209

Feature ID	Feature description	Feature annotation for results1	Feature annotation for results2	NES for results1	NES for results2	Rank for results1	Rank for results2	Sum of ranks
jaspar-PF0012.1	CAGGTG	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1 MYOG MYOD1	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1 MYOG MYOD1	7.613792604	9.975553913	2	2	4
jaspar-MA0103.1	ZEB1	ZEB1	ZEB1	8.981663639	8.610800778	1	4	5
swissregulon-SNAI1..3.p2	SNAI2,SNAI1,SNAI3	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1 MYOG MYOD1	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1 MYOG MYOD1	7.458838463	9.698971814	3	3	6
stark-CAGGTG	CAGGTG(sna)	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1 MYOG MYOD1	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1 MYOG MYOD1	7.274496468	8.240214375	4	6	10
jaspar-MA0086.1	sna	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1	MYF5 MYF6 NHLH2 NHLH1 SNAI2 SNAI3 SNAI1	6.631525838	8.53126498	7	5	12

- For now, the comparison will only be between detected PWMs
- In this comparison you'll see that 31 motifs are overlapping (indicated in green)
- They are ranked based on the sum of their individual ranking.
- As expected, the motif that was ranked as second motif in both analyses is now ranked first

Other options in i-cisTarget

Results for ZEB1_analysis

Select features in the table below, select an operation and [proceed](#).

1. **Combine** these features via orderstatistics. Use following parameters for combined features:

Method

Combine all features into one 

Enrichment

Within each database separately 

Control

TF binding sites 

2. Use candidate target regions as **filter** and use as input for i-cisTarget again.
3. **Scan** candidate target regions of selected features for  multiple homotypic  CRMs.

This report is also available as an [archive](#).

- i-cisTarget has a number of other option you can do with your results
- Tick the features you want to use to do further analysis
- Choose one of the three options
 - Either combine the features you indicated
 - Or you can use them for a new i-cisTarget analysis
 - Perform a further scan on the selected regions for more detailed CRMs types