ANEMONE User Guide

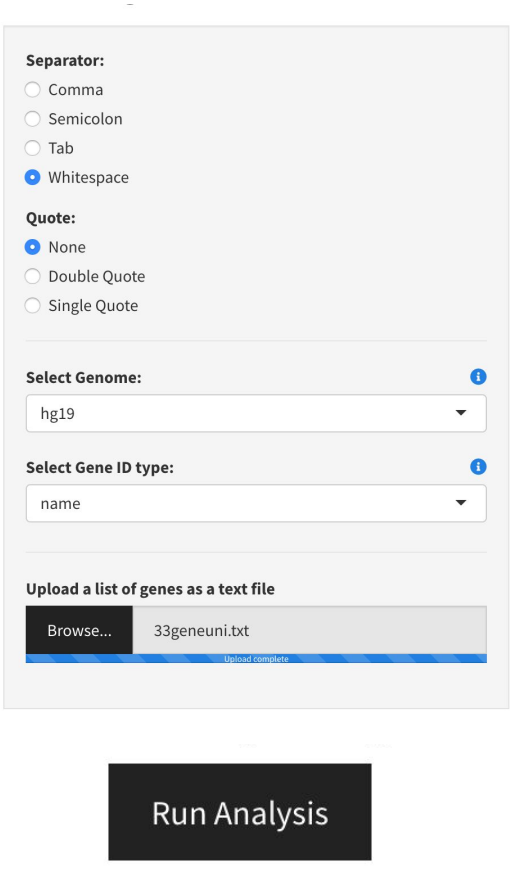
Charlene Miciano | Salk Institute | Integrative Genomics Core

About Anemone

ANEMONE is an intuitive interactive web tool that gives users the ability to explore differentially expressed genes by their regulators. ANEMONE has two functions: gene clustering and transcription factor-gene visualization. The first function clusters genes based on motif similarity and provides visualizations of those clusterings. The second function outputs a regulatory network graph of genes and the transcription factors that regulate them.

Inputting Genes

Genes must be uploaded as a text file with each line containing one gene ID. The gene IDs supported are entrez, ensembl, Unigene, RefSeq, and gene name. Note how your gene IDs are separated (comma, semicolon, tab whitespace), and whether your gene IDs have quotes (none, double quote, single quote). Use the dropdown bar to select the Gene ID type your gene symbols are formatted as. Genes will be converted to gene symbols for further processing.



When you upload your text file, a message will prompt you to navigate to the “Input Genes” tab. Here you can check to see if your converted gene names appear in the “Converted” table. Once that you have seen your genes are converted, click “Run Analysis” to cluster the genes. The clustering may take around 15 seconds to run.

Inputted

original
NPY1R
RLN2
FAM198B
PRSS23
TMEM87B
SGK1

Converted

name
ACLY
ADCY1
AGR3
ATRNL1
CA12
CCDC68

Motif Gene Clustering Tool

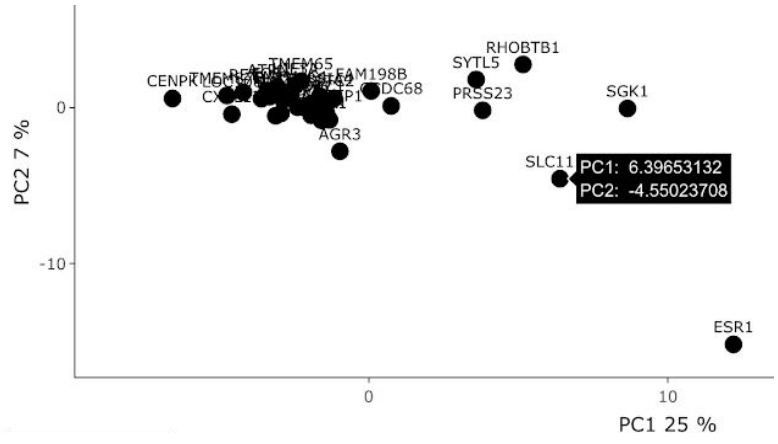
The motif gene clustering tool provides a complete search of motif sequences across promoters that enables users to explore novel relationships between genes of interest and find genes that are potentially co-regulated. After running the analysis, the user can then browse the tabs of the tool to explore their clustering results.

Motif/Gene Table

The “Motif/Gene table” tab displays a table of genes and their associated motifs. The count denotes the number of particular motif occurrences in the promoter (-1000 to 100bp) of a gene. The higher the count, the more likely a transcription factor binding to that motif sequence has a role in regulating the gene. The table is available for download as a text file.

ID	AP.2gamma.AP2.	AR.halfsite.NR.	Arnt.Ahr.bHLH.	Atf4.bZIP.	Atf7.bZIP.
ACLY	1	2	0	0	0
ADCY1	5	4	0	0	1
AGR3	0	2	1	0	0
ATRNL1	0	0	0	1	0
CA12	0	3	1	0	0
CCDC68	2	4	0	0	0
CENPK	0	0	0	0	0
CXCL12	1	2	0	0	0
DKC1	0	3	2	0	0
ESR1	2	9	5	1	0
FAM198B	0	5	1	0	0
FOS	3	3	0	0	0
KIF3A	0	3	1	0	0

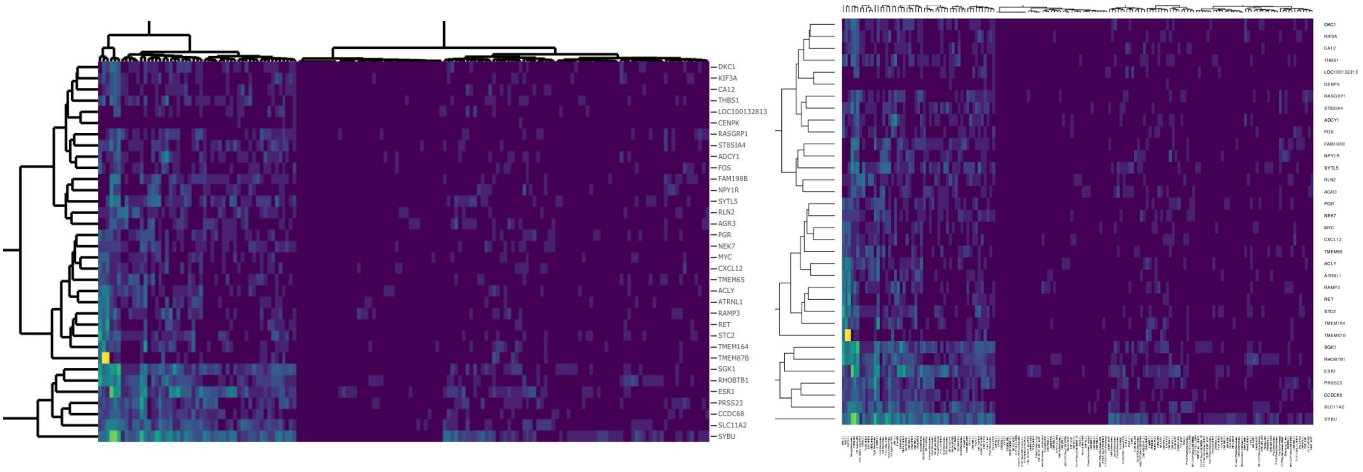
PCA Plot



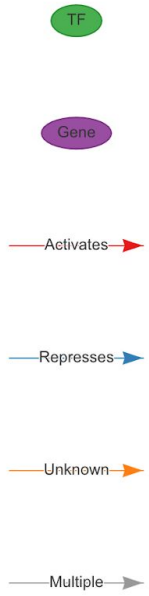
The “PCA” tab displays an interactive PCA plot of genes based on their motif similarity in the promoter region. The user can hover over a point to see the gene name and its PCA coordinates. They can also zoom into a region of interest by dragging a rectangle over the area. Close proximity of points indicates higher motif similarity in the promoter region. The user can download the plot in html.

Heatmaps

The “Heatmaps” tab provides heatmaps depicting the hierarchical clustering of the user’s genes of interest by motif similarity. Clustering was performed using Ward’s method on Euclidean distance. Cells with a more intense value corresponds to a higher number of motif occurrences within the promoter region of its matching gene. A static pdf version of the heatmap is available for viewing and download. The user can search for genes of interest within the pdf. The user can also view an interactive version of the heatmap. The user can inspect a cell by hovering over it, and can zoom into a region of the heatmap by dragging a rectangle around the area. An interactive version of the heatmap is also available for the user to download as an html file.

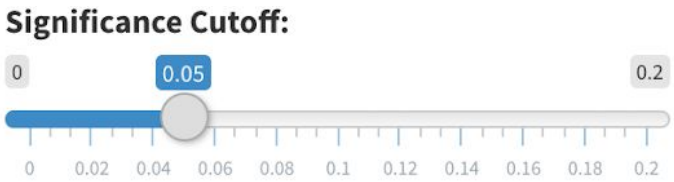


Network Graph Tool

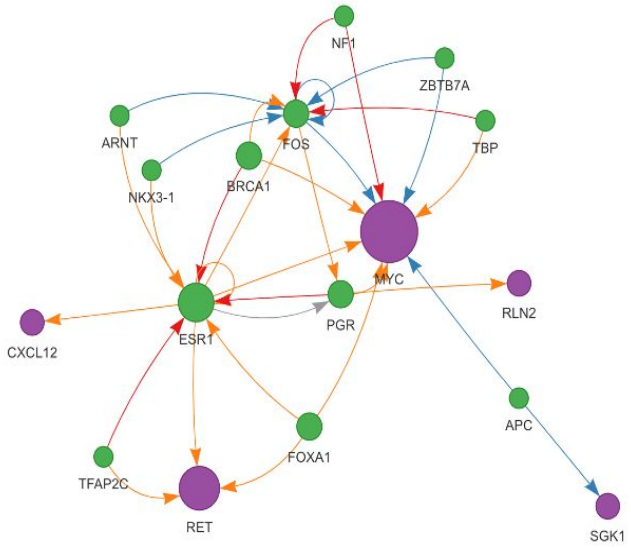


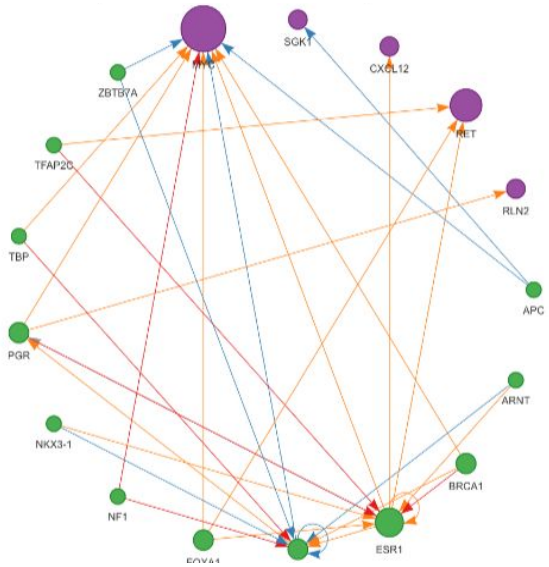
The network graph tool aims to visualize the regulatory relationships between differentially expressed genes and their transcriptional regulators. User-inputted genes are colored in purple while their transcription factors are colored in green. The arrows represent a transcription factor’s mode of regulation for a particular gene. The modes of regulation are activation, repression, unknown, and multiple.

The number of nodes can be adjusted by the significance cutoff slider. The cutoff determines how significant a transcription factor is relative to your input gene list based on Fisher's exact test. Depending on the transcription factors for your input gene list, you may have to adjust the slider to generate a network.



The graphs can be displayed in network-style or circle-style. Additionally, a table of the most enriched transcription factors, their mode of regulation, and the publication IDs describing their relationship is also viewable. The graphs are downloadable as an html file and the table is downloadable as a text file.





TF	Gene Overlap p-value	Gene Overlap adjusted p-value	# of target genes in DB	# of overlapped genes	Target Gene and Mode of Regulation
AHR	0.01	0.01	22	2	FOS_Unknown|MYC_Unknown
APC	0.00	0.01	12	2	MYC_Repression|SGK1_Repression
ARNT	0.00	0.01	20	2	ESR1_Unknown|FOS_Repression
BRCA1	0.00	0.01	57	3	ESR1_Activation|FOS_Unknown|MYC_Unknown
E2F1	0.03	0.04	134	3	MYC_Activation|RASGRP1_Activation|THBS1_Activation
ESR1	0.00	0.00	76	6	CXCL12_Unknown|ESR1_Unknown|FOS_Unknown|MYC_Unknown|PGR_Activation|PGR_Unknown|RET_Unknown
FOS	0.00	0.01	57	3	FOS_Repression|MYC_Repression|PGR_Unknown
FOXA1	0.00	0.00	21	3	ESR1_Unknown|MYC_Unknown|RET_Unknown
LEF1	0.01	0.02	29	2	ESR1_Unknown|MYC_Unknown
MTA1	0.01	0.01	24	2	ESR1_Repression|ESR1_Unknown|MYC_Unknown
NF1	0.00	0.01	14	2	FOS_Activation|MYC_Activation
NKX3-1	0.00	0.00	10	2	ESR1_Unknown|FOS_Repression
PGR	0.00	0.00	25	3	ESR1_Activation|MYC_Unknown|RLN2_Unknown
RB1	0.01	0.02	31	2	FOS_Repression|FOS_Unknown|MYC_Repression|MYC_Unknown
TBP	0.00	0.01	15	2	FOS_Activation|MYC_Unknown
TFAP2C	0.00	0.00	10	2	ESR1_Activation|RET_Unknown
TWIST1	0.01	0.02	35	2	ESR1_Repression|FOS_Activation
WT1	0.03	0.04	57	2	MYC_Repression|MYC_Unknown|THBS1_Activation|THBS1_Repression
ZBTB7A	0.00	0.00	6	2	FOS_Repression|MYC_Repression