Enhancers

Hands-on-training



info@genexplain.com

Part 1: Identifying regulation specific combinations of transcription factor binding sites with the help of Composite Module Analysis (CMA)

Part 2: Using Fantom5 database to analyze tissue- or cell-type specific promoters



Part 1

Composite Module Analysis (CMA)



Open the ChIP-Seq - Identify composite modules on peaks (TRANSFAC(R)) workflow as shown below.





Navigate to the *GSM558469_E2F1_hg19 filtered exp1000 dist1000 L<600* track and drag-and-drop it to the **Input Yes track** field of the workflow input mask. Navigate to the *Housekeeping genes (Human) track -100000 to -98000 filtered chr 1* track and drag-and-drop it to the **Input No track** field of the workflow input mask. The field **Sequence source** should be set to Ensembl Human 75 genome build. Click **Run workflow** when parameters are set as shown below.





Have a look to the results folder and corresponding files.





Open under **modules** the file **Model visualization on Yes set** and you will see the composite module, which was found with the workflow for all input sequences (track). While having this file open, more information can be found in the tab **My description** below.





The model consists 2 modules (2 pairs of TFBSs) and is highly significant than expected with a random set of regulatory regions. The **Model view** file contains the visualization of the modules.

The model consists of 2 module(s). Below, for each module the following information is shown:

- PWMs producing matches,
- number of individual matches for each PWM,
- score of the best match.



Model score (-p*log10(pval)): 29.16 Wilcoxon p-value (pval): 6.30e-45 Penalty (p): 0.660 Average yes-set score: 4.50 Average no-set score: 1.96 AUC: 0.92 Middle-point: 3.13 False-positive: 9.33% False-negative: 13.65%

Start page 🖄 Model view X Databases Data Analyses ------GSM558469_E2F1_hg19 filtered exp1000 dis Module 1 SGM558469_E2F1_hg19 filtered exp1000 V\$ZFP161 04 V\$BEN 01 GSM558469_E2F1_hg19 filtered exp1000 Housekeeping genes (Human) track -100(0.87; N=1 0.70; N=3 Module width: 131 🕿 modules Histogram Module 2 💫 Model view V\$RNF96 01 V\$SP100 04 Model visualization on No set 0.91; N=3 0.86; N=2 Model visualization on Yes set Module width: 184 💣 no track 💣 yes track A ProfileOptimized 📑 Site optimization summary

The AUC of the model achieves value significantly higher than expected for a random set of regulatory regions Z-score = 7.90



info@genexplain.com

The **Histogram** shows the scores of the found modules between Yes-set sequences and No-set. The clear demarcation between red bars (Yes-set) and blue bars (No-set) described the high significance of identified modules compared with a random set of regulatory regions.





Two transcription factors were found, which are postulated to bind to the identified modules and corresponding pairs of binding sites.





Part 2

Gene transcripts and Fantom5 database



Part 2 – Gene transcripts and Fantom5 database

Get familiar with the Fantom5 database as shown below.



Fantom5 Tissue

This database contains tissue-specific transcription start sites compiled by geneXplain on the basis of Fantom5 data and mapped to the hg38 genome build.

Version: 1.2



info@genexplain.com

Part 2 – Gene transcripts and Fantom5 database

Open and run the **Convert table** tool as shown below.

Input table	🛃ated Ensembl genes filtered (logFC>1)
Species	Human (Homo sapiens)
Input type	Genes: Ensembl
Output type	Transcripts: Ensembl
Ignore empty values	
Numerical value treatment rule	extreme 💽
Leading column	(none)
Output table	📑 filtered (logFC>1) Transcripts Ensembl
Show expert options >> Run Completed	
INFO - Preparing	
INFO - Matching plan:	
INFO - * Genes: Ensembl	
INFO - * Transcripts: Ensembl	
INFO - Matching	
INFO - Matched rows: 411	
INFO - Unmatched rows: 0	
INFO - Preparing row list	
INFO - Resulting rows: 3568	
INFO - Generating output	
INFO - Analysis 'Convert table' finished (12.794 s)	*

Convert table



Part 2 – Gene transcripts and Fantom5 database

Open and run the **Create transcript region track** tool as shown below.

Input transcripts	[?]	📑 filtered (logFC>1) Transcripts Ensemb
Species	[?]	Human (Homo sapiens)
Transcript region	[?]	Promoter 🗾
Promoter start	[?]	-1000
Promoter end	[?]	100
🛄 🖸 Output path	[?]	>1) Transcripts Ensembl transcript region

Run

Completed

INFO - Analysis 'Create transcript region track' added to queue

- INFO Analysis 'Create transcript region track' started
- INFO Creating Promoter track

INFO - Analysis 'Create transcript region track' finished (42.601 s)

Create transcript region track



Open and run the Identify enriched motifs in tissue specific promoters (TRANSFAC(R)) workflow as shown below.

Identify enriched motifs in tissue specific promoters (TRANSFAC(R))

🔤 🚹 Input Yes genes	ated Ensembl genes filtered (logFC>1)
Species	Human (Homo sapiens)
Sequence source	Ensembl 100.38 Human (hg38)
Tissue_condition	51) kidney normal
TSS selction	Most active
Annotation source	databases/EnsemblHuman100/Data/gen
Profile	🕋ata/profiles/vertebrate_human_p0.001 Auto
Filter by Coefficient	0.05
Result folder	🕼 normal specific promoters, Transfac)

Identify enriched motifs in tissue specific promoters (TRANSFAC(R))

