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

info@genexplain.com

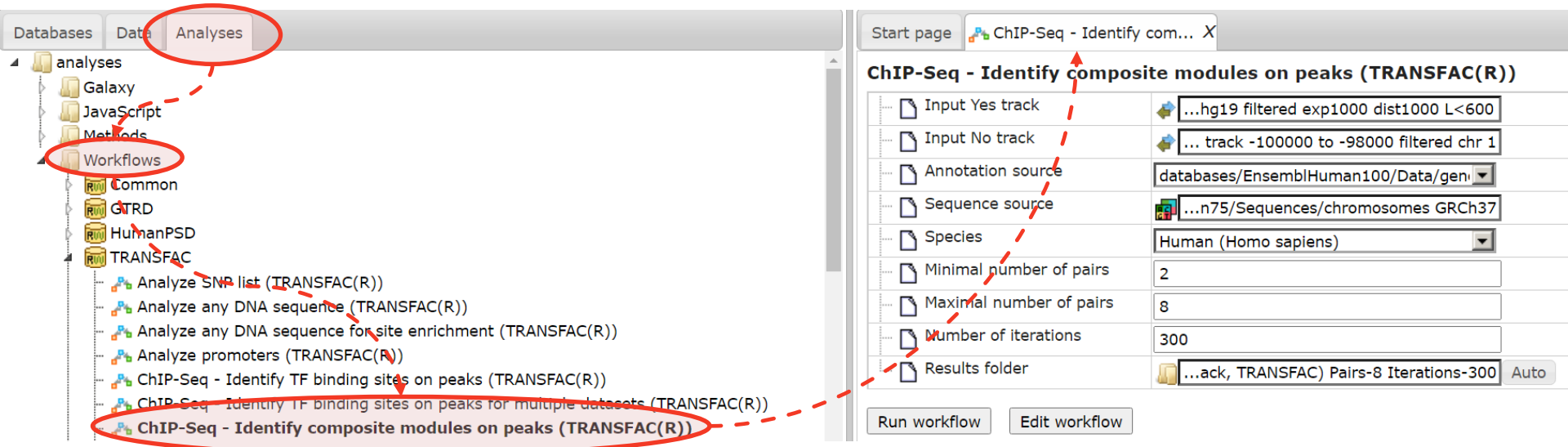
Part 1

Composite Module Analysis (CMA)

info@genexplain.com

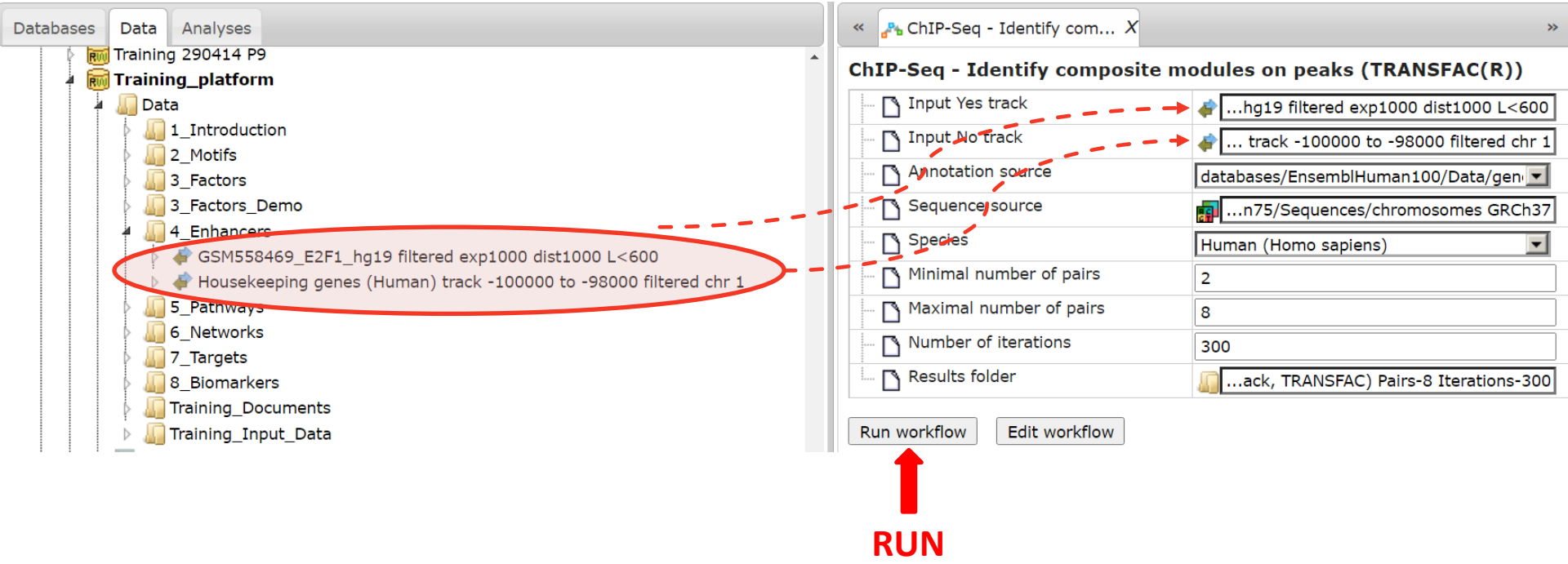
Part 1 – Composite Module Analysis (CMA)

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



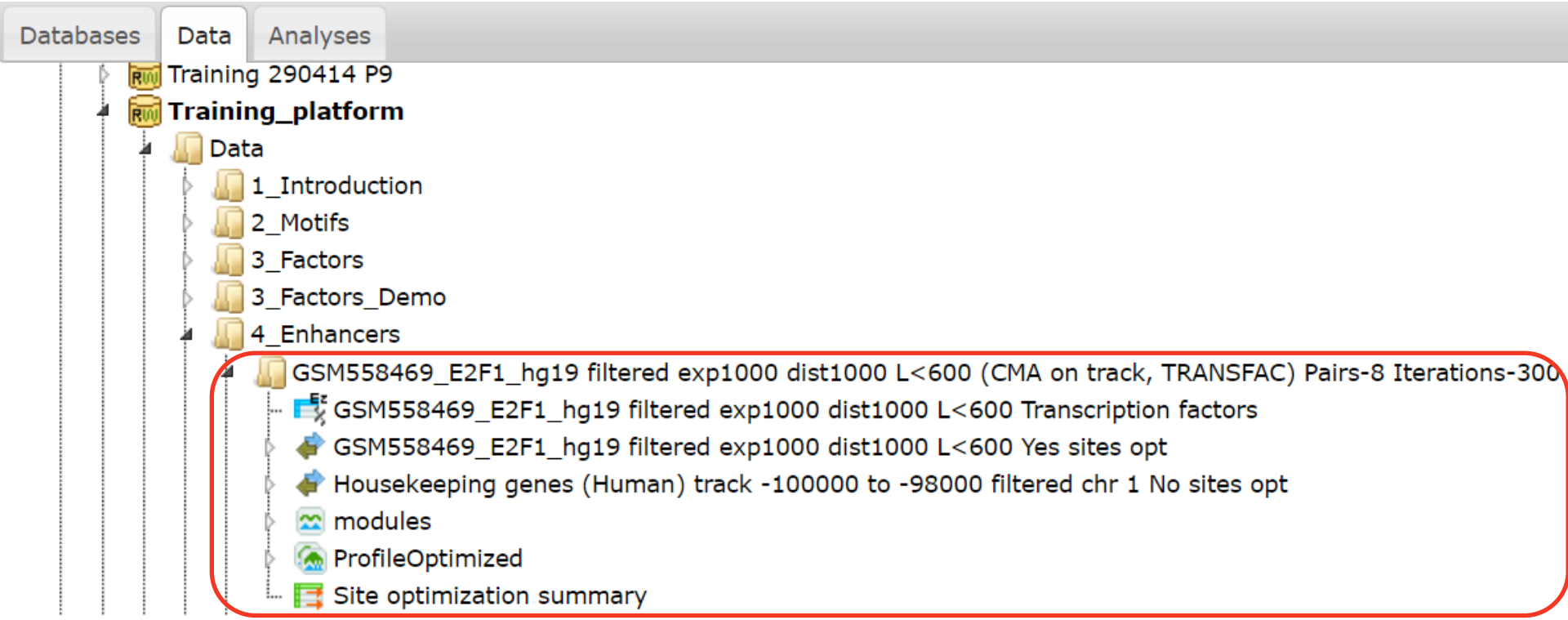
Part 1 – Composite Module Analysis (CMA)

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.



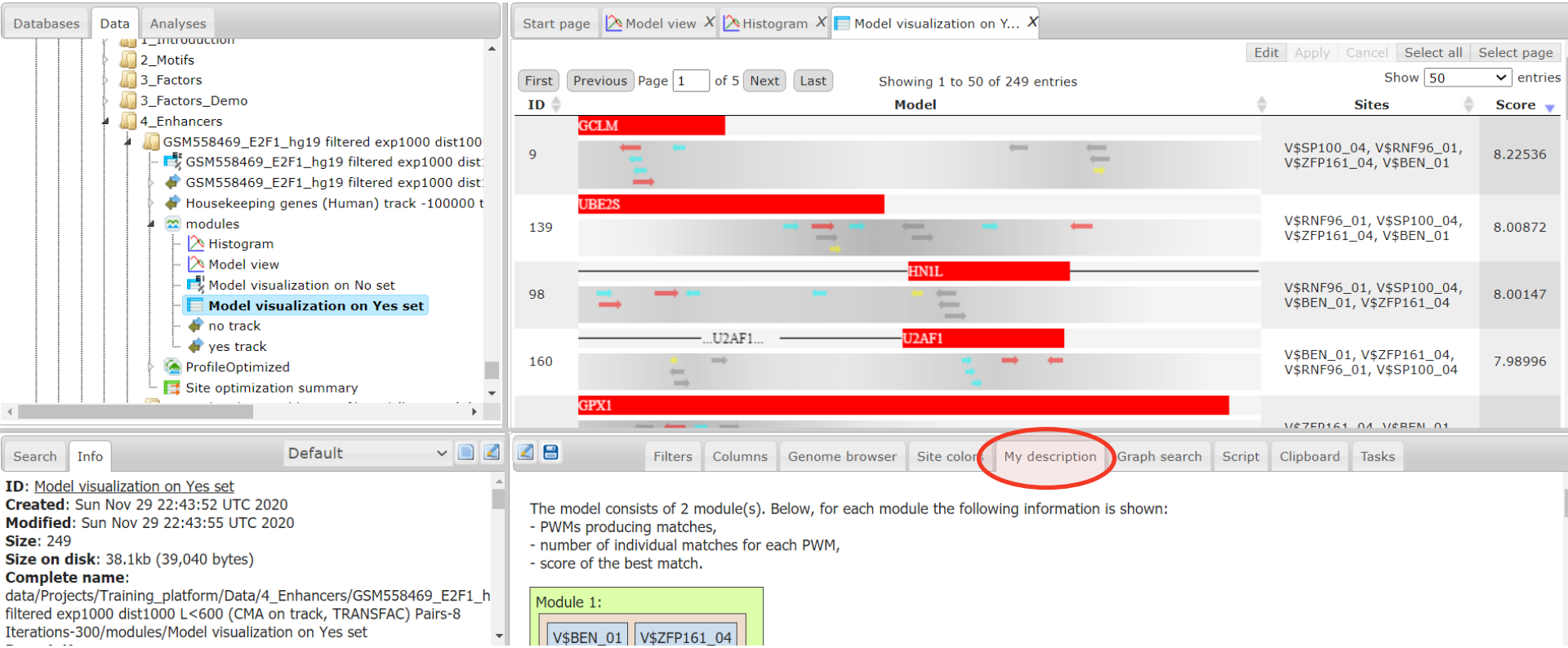
Part 1 – Composite Module Analysis (CMA)

Have a look to the results folder and corresponding files.



Part 1 – Composite Module Analysis (CMA)

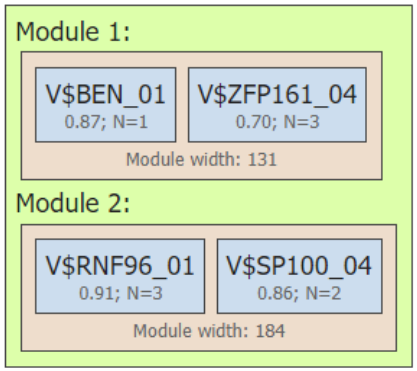
Open under **modules** the file **Model visualization on Yes set** and you will see the composite module, which was found with the workflow for all input sequences (track). While having this file open, more information can be found in the tab **My description** below.

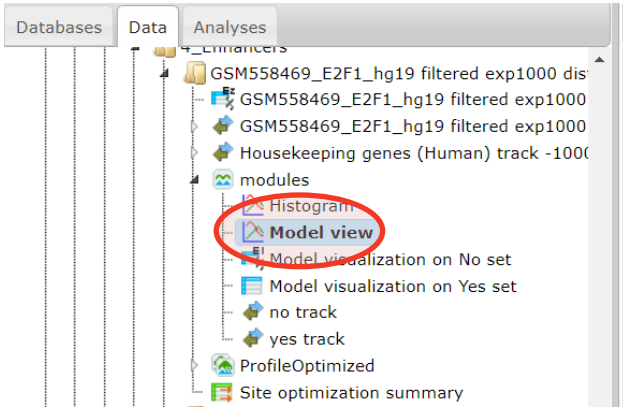


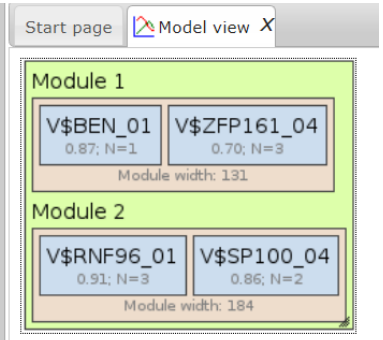
Part 1 – Composite Module Analysis (CMA)

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.







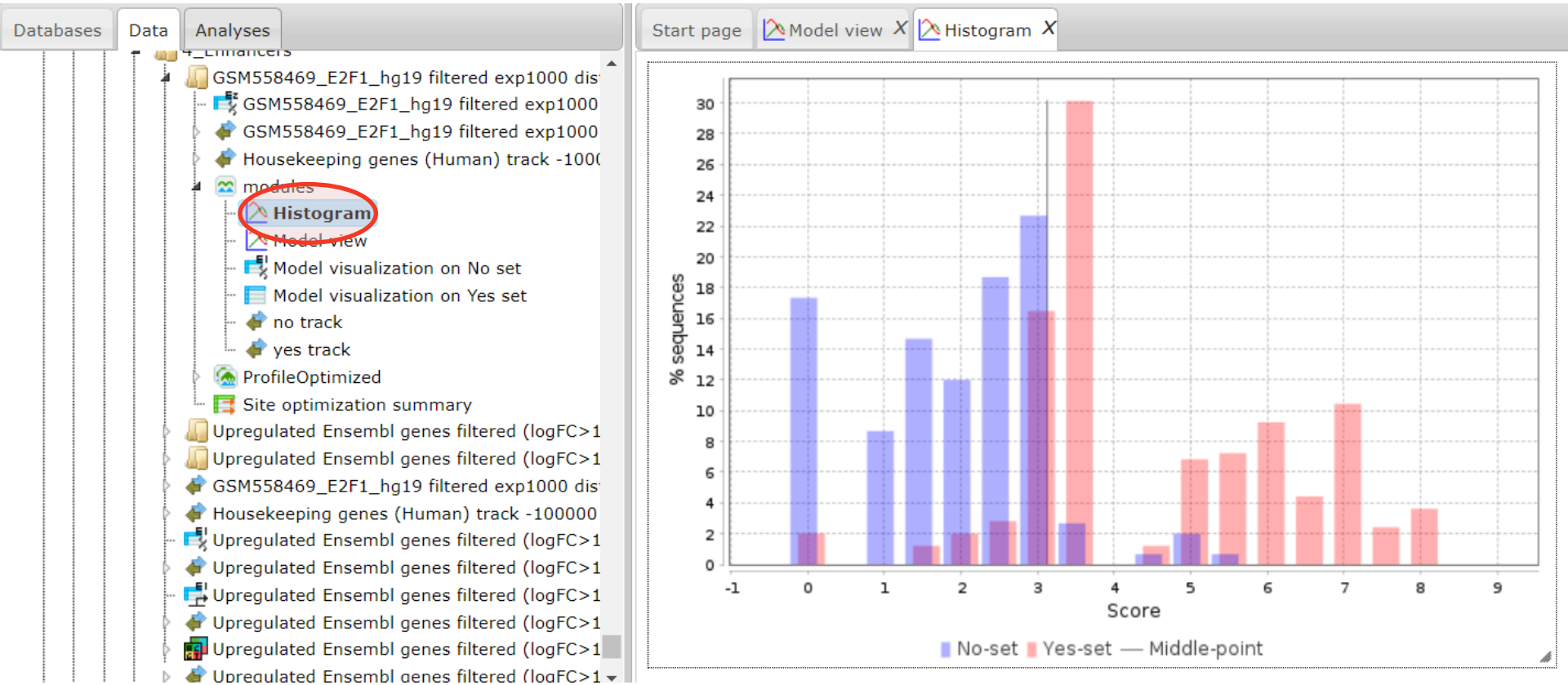
Model score (-p*log10(pval)): 29.16
Wilcoxon p-value (pval): 6.30e-45
Penalty (p): 0.660
Average yes-set score: 4.50
Average no-set score: 1.96
AUC: 0.92
Middle-point: 3.13
False-positive: 9.33%
False-negative: 13.65%

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

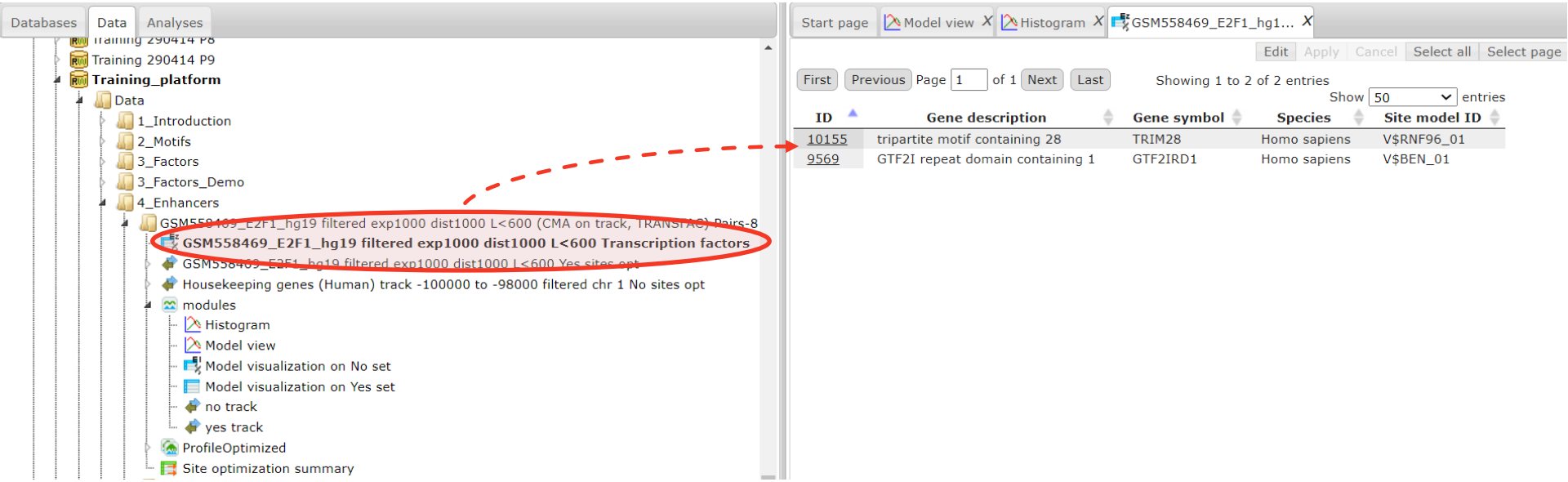
Part 1 – Composite Module Analysis (CMA)

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.



Part 1 – Composite Module Analysis (CMA)

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

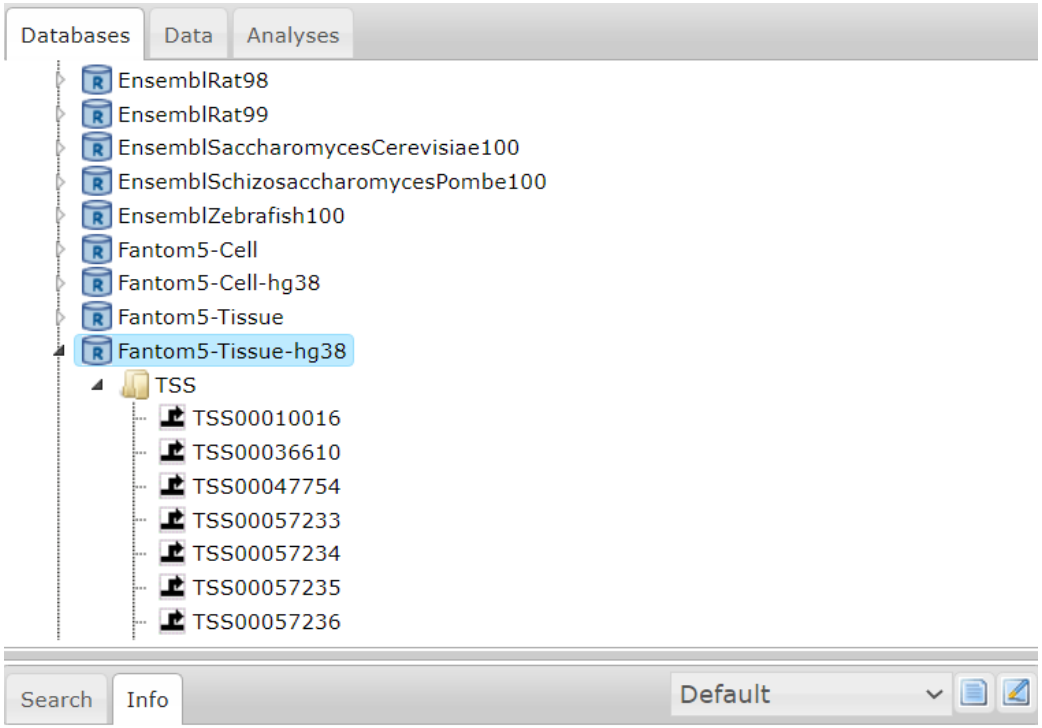


Part 2

Gene transcripts and Fantom5 database

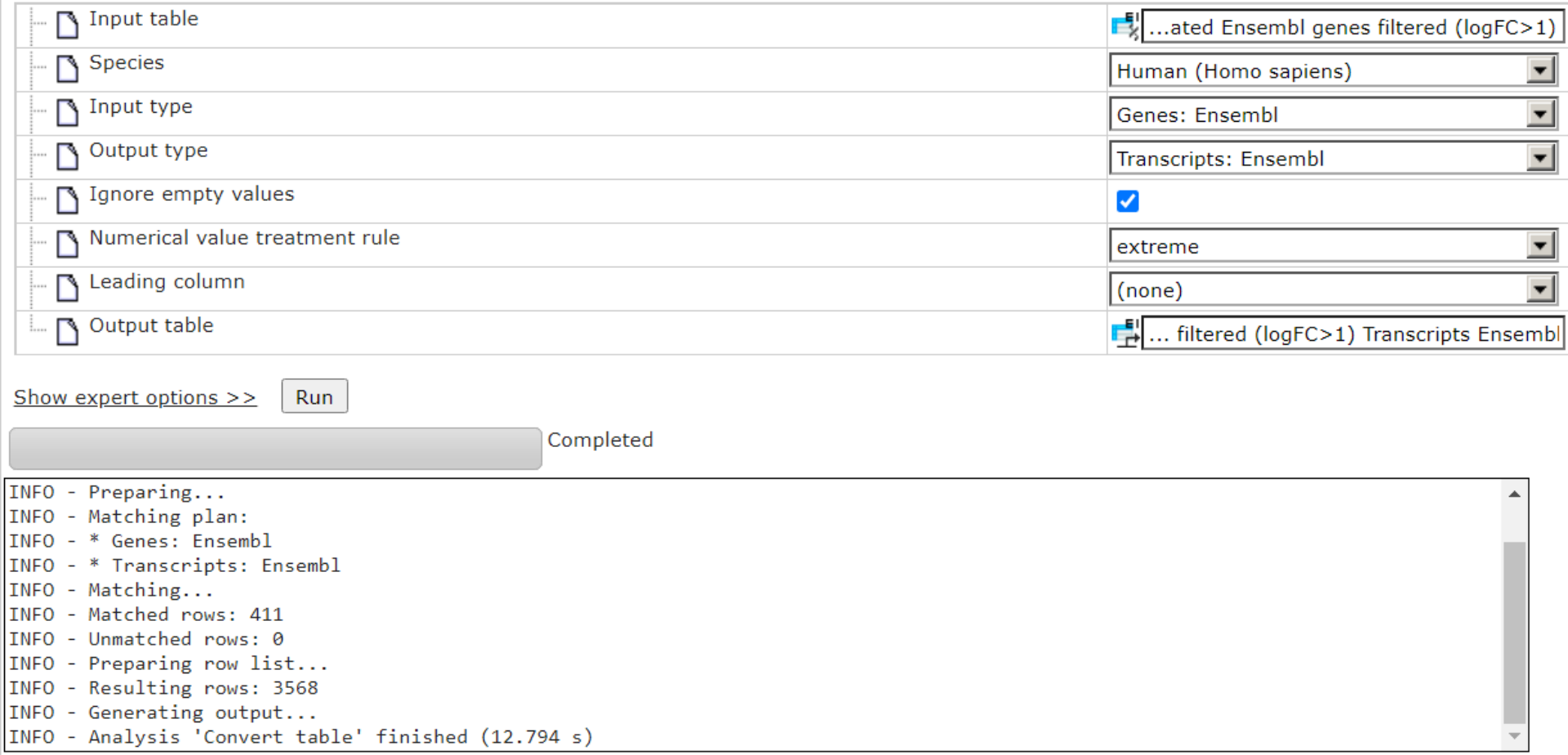
Part 2 – Gene transcripts and Fantom5 database

Get familiar with the **Fantom5 database** as shown below.



Part 2 – Gene transcripts and Fantom5 database

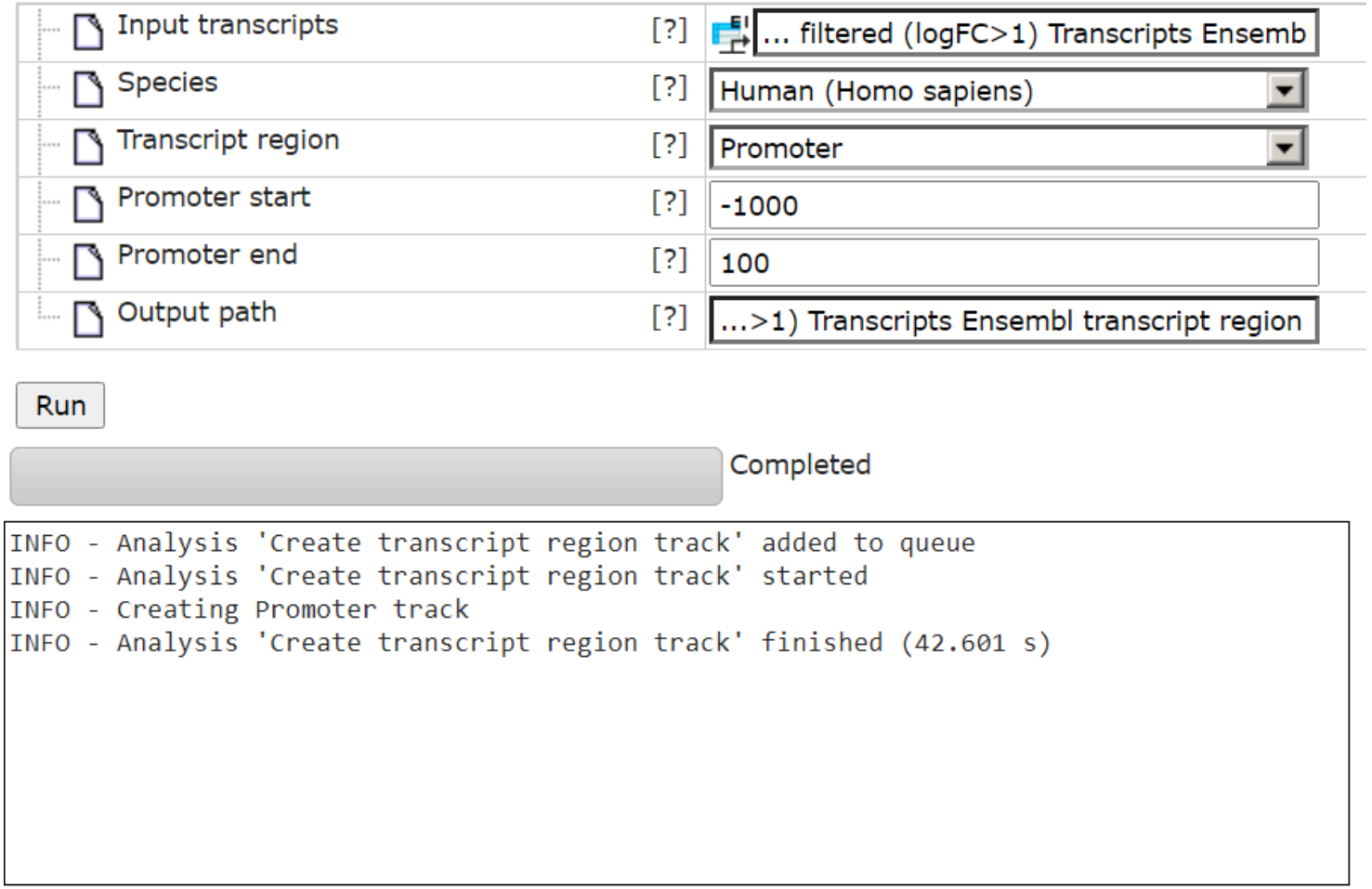
Open and run the **Convert table** tool as shown below.



[Convert table]

Part 2 – Gene transcripts and Fantom5 database

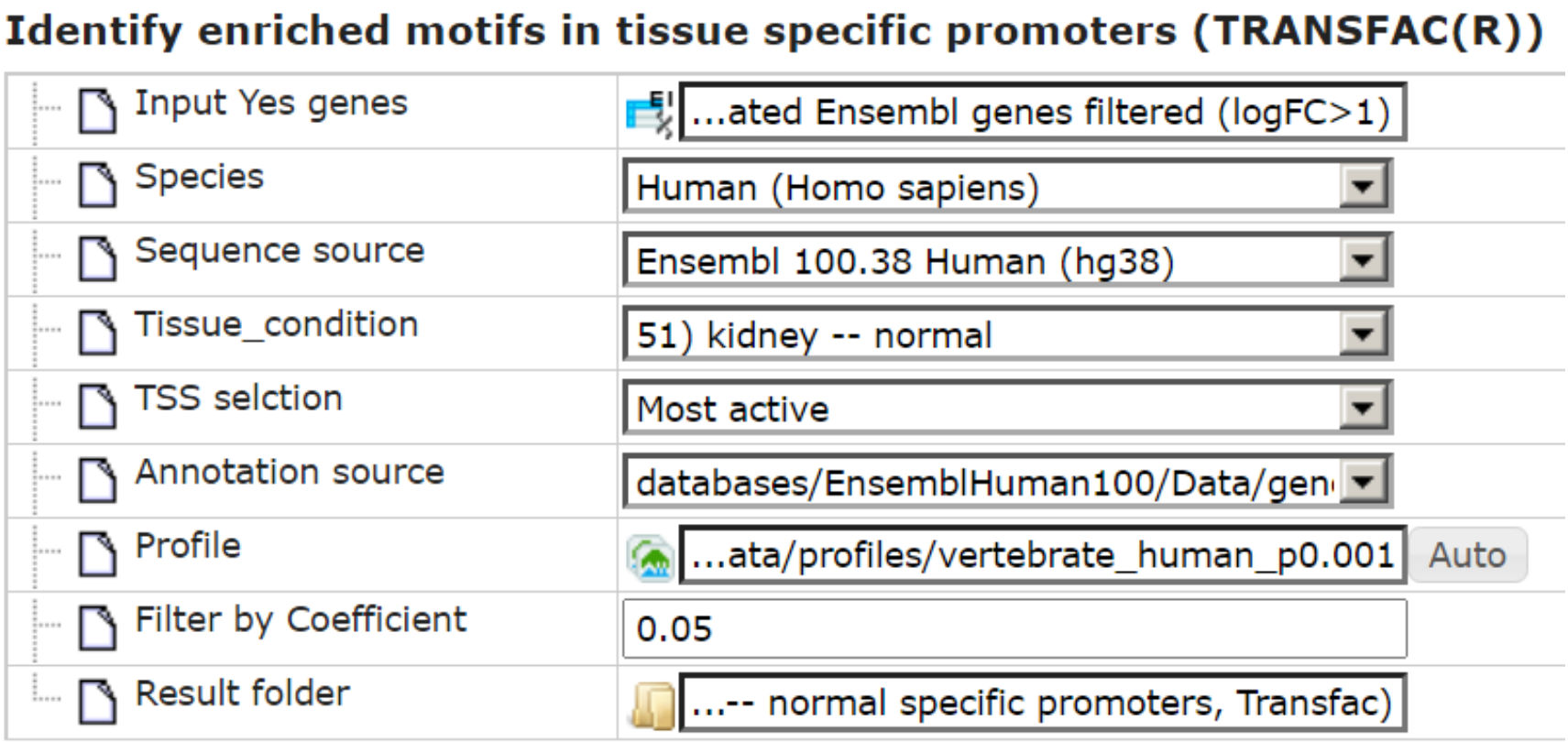
Open and run the **Create transcript region track** tool as shown below.



Create transcript region track

Part 2 – Gene transcripts and Fantom5 database

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))]