Transcription factors

Hands-on-training

info@genexplain.com

Part 1: Mapping a gene list with TF classification

Part 2: Converting Matrices to molecules

Part 3: Analyzing ChIP-seq peaks

Part 1

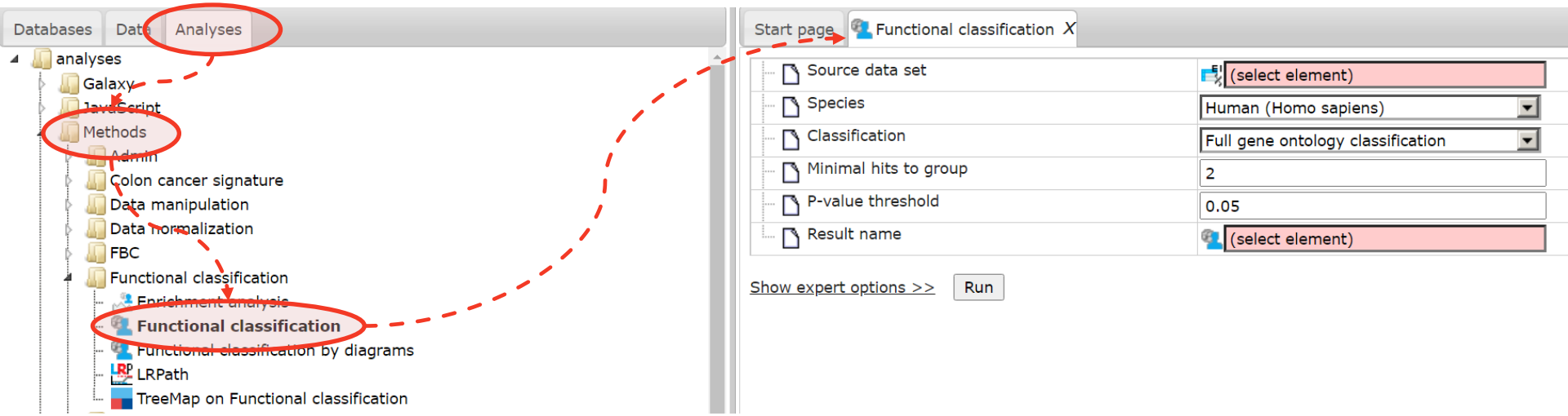
TFClass is a classification of eukaryotic transcription factors based on the characteristics of their DNA-binding domains (DBDs).

We will use a list of genes and identify the most prominent transcription factor class with the functional classification tool and underlying statistics.

This analysis allows you to classify a set of genes into TF classification groups.

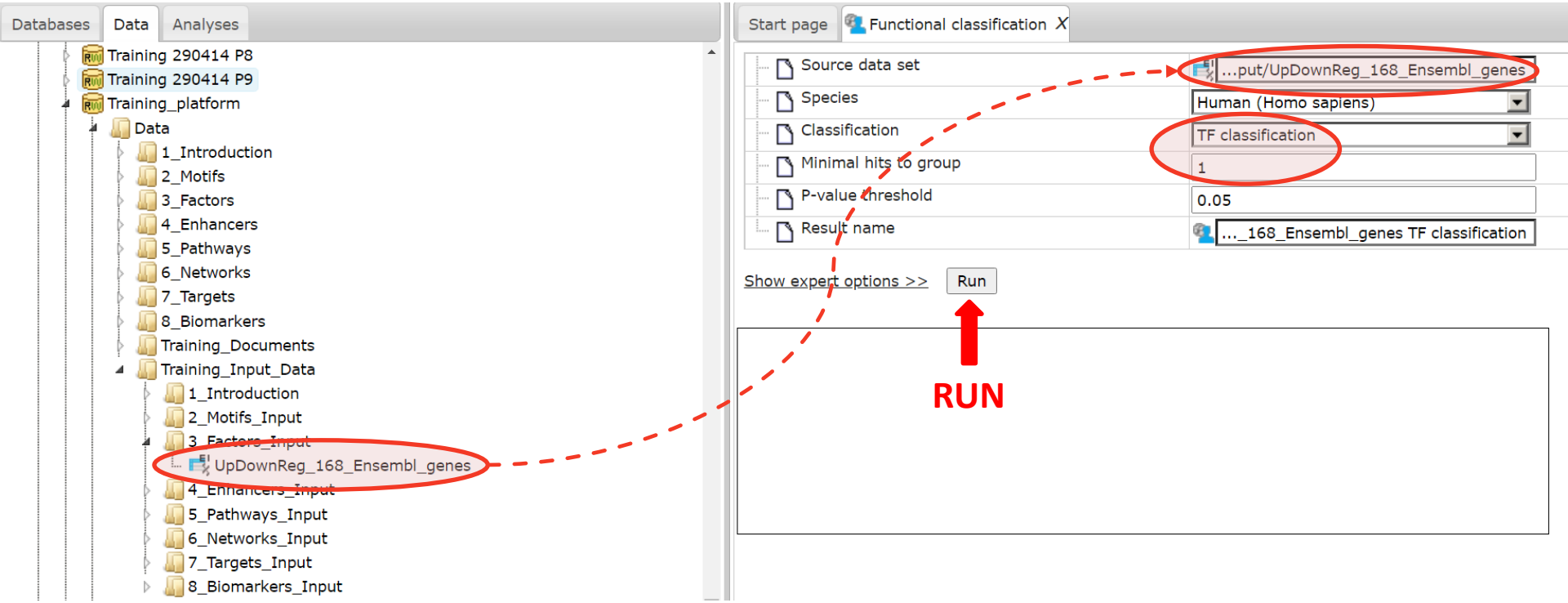
Part 1 – Classify a set of genes into TF classification groups

After logging into the geneXplain platform, open the **Functional classification** tool as shown below.



Part 1 – Classify a set of genes into TF classification groups

Navigate to the *UpDownReg_168_Ensembl_genes* table and drag-and-drop it to the **Source data set** field of the tool. The field **Classification** is a source directory whose information is used for the comparison (mapping) and select **TF classification**. Change **Minimal hits to group** to **1**, because our input list is quit small and we would like to have an overrepresentation from the first hit already. Click **Run** when parameters are set as shown below.



Part 1 – Classify a set of genes into TF classification groups

Sorting the resulting table with lowest **Adjusted P-value** on top gives four most prominent TF classes in the red box.

Start page | Functional classification X | UpDownReg_168_Ense... X

Edit | Apply | Cancel | Select all | Select page

First | Previous | Page 1 of 1 | Next | Last

Showing 1 to 35 of 35 entries

Show 50 entries

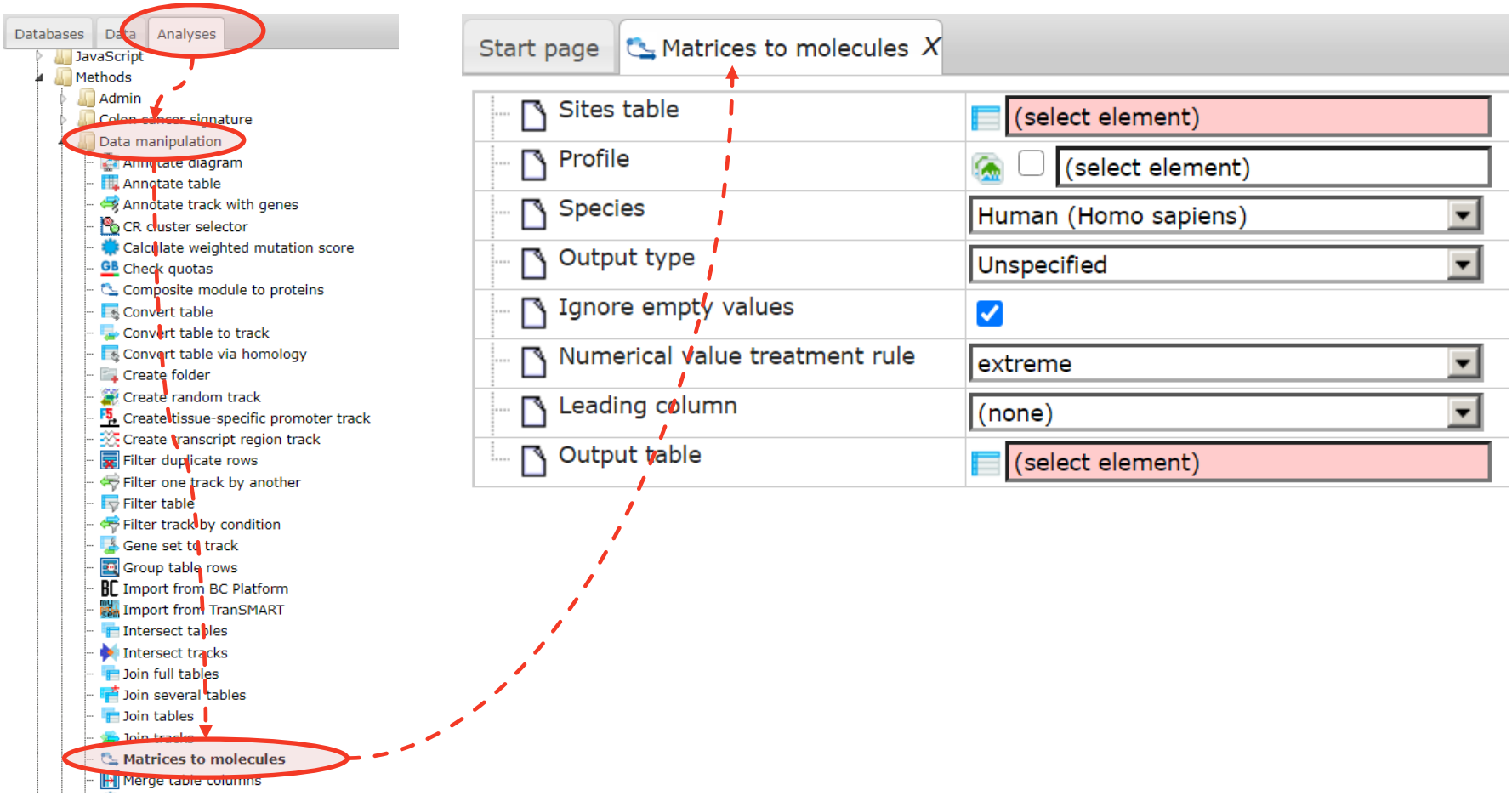
ID	Title	Number of hits	Group size	Expected hits	P-value	Adjusted P-value	Hit names
6.1	Rel homology region (RHR) factors	4	30	0.35225	3.0202E-4	0.00944	BCL3, NFKBIA, NFKBIE, RELB
6.1.2	Ankyrin domain-only factors	3	14	0.16438	4.5699E-4	0.00944	BCL3, NFKBIA, NFKBIE
6.1.2.1	IκB-related factors	3	13	0.15264	3.6172E-4	0.00944	BCL3, NFKBIA, NFKBIE
6	Immunoglobulin fold	5	67	0.78669	7.5582E-4	0.01172	BCL3, NFKBIA, NFKBIE, RELB, STAT6
1.1.1.1.2	JunB	1	1	0.01174	0.01174	0.02912	JUNB
1.1.2.2.1	ATF-3	1	1	0.01174	0.01174	0.02912	ATF3
2.1.3.5.2	COUP-TFII (NR2F2)	1	1	0.01174	0.01174	0.02912	NR2F2
2.2.1.1.3	GATA-3	1	1	0.01174	0.01174	0.02912	GATA3
2.3.4.5.2	HIV-EP2 [2+2] (MBP-2, ZNF40B)	1	1	0.01174	0.01174	0.02912	HIVEP2
3.1.2.11.1	MSX-1 (HOX7)	1	1	0.01174	0.01174	0.02912	MSX1
3.3.2.1.8	E2F-8	1	1	0.01174	0.01174	0.02912	E2F8
3.5.2.1.2	c-Ets-2	1	1	0.01174	0.01174	0.02912	ETS2
3.5.3	Interferon-regulatory factors	2	9	0.10568	0.00447	0.02912	IRF1, IRF2
3.5.3.0.1	IRF-1	1	1	0.01174	0.01174	0.02912	IRF1
3.5.3.0.2	IRF-2	1	1	0.01174	0.01174	0.02912	IRF2
5.1.1.1.3	MEF-2C	1	1	0.01174	0.01174	0.02912	MEF2C
6.1.1.2.2	RelB (I-Rel)	1	1	0.01174	0.01174	0.02912	RELB

Part 2

Now we are using a tool to convert a list of transcription factor binding sites into a list of corresponding transcription factors.

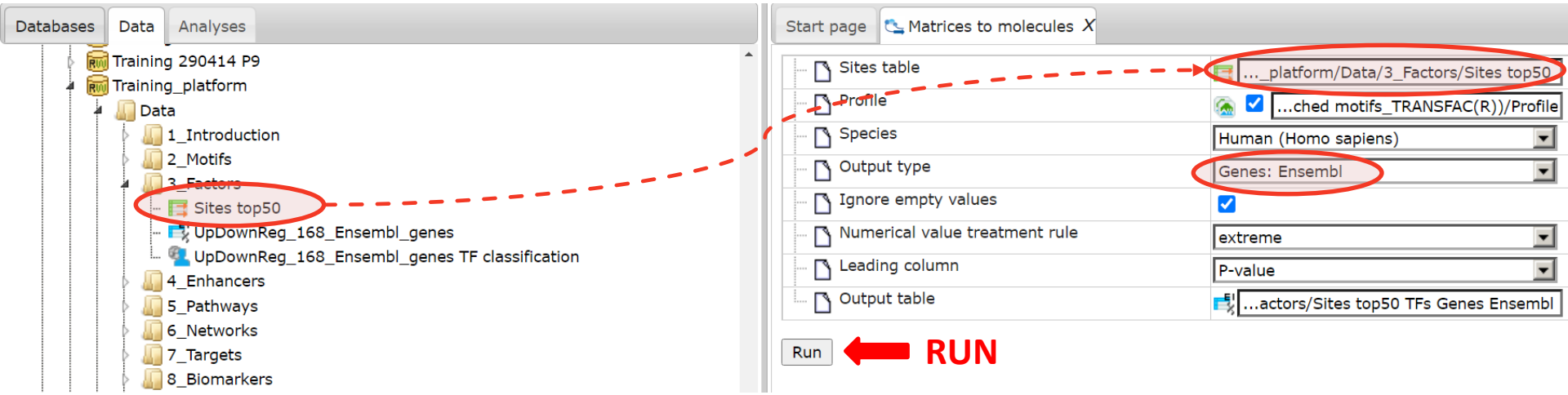
Part 2 – Convert a list of transcription factor binding sites (TFBSs) into a list of transcription factors (TFs)

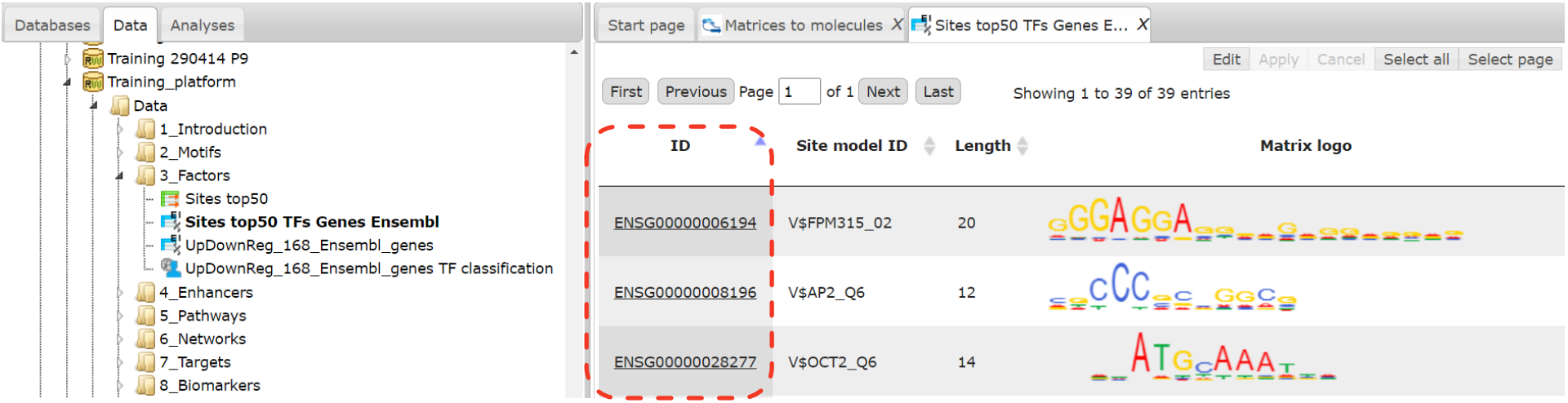
Open the **Matrices to molecules** tool as shown below.



Part 2 – Convert a list of transcription factor binding sites (TFBSs) into a list of transcription factors (TFs)

Navigate to the *Enriched motifs top50* table and drag-and-drop it to the **Sites table** field of the tool. The field **Output type** defines the hub, which is used for the conversion mapping and defines the output type and ID. Click **Run** when parameters are set as shown below. The resulting table has Ensembl IDs added to the original Sites table.





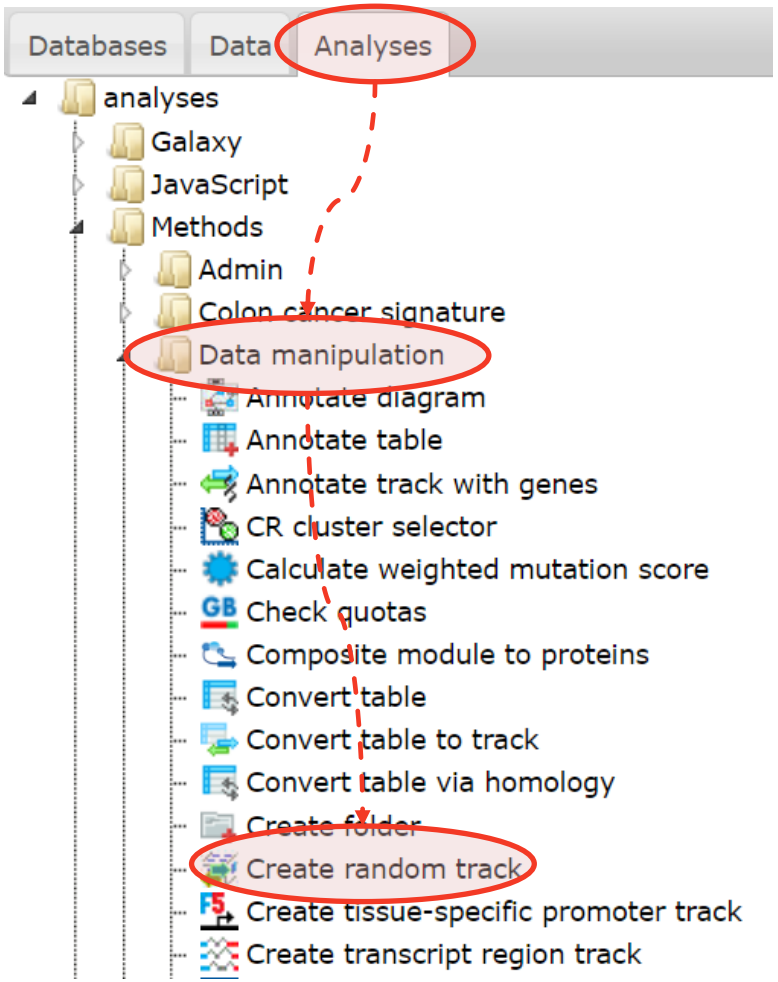
Part 3

This part demonstrates the analysis of ChIP-seq data in the geneXplain platform. We will show how to work with genomic intervals loaded from BED files. The ChIP-seq peaks from Jurkat cells were bound by TAL1 in a study be Palii et al. [1]. In the first part, we will analyze binding sites in TAL1-bound regions in Jurkat cells.

1. Palii CG, Perez-Iratxeta C, Yao Z, Cao Y, Dai F *et al.* (2011) Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. EMBO J. 30:494-509.

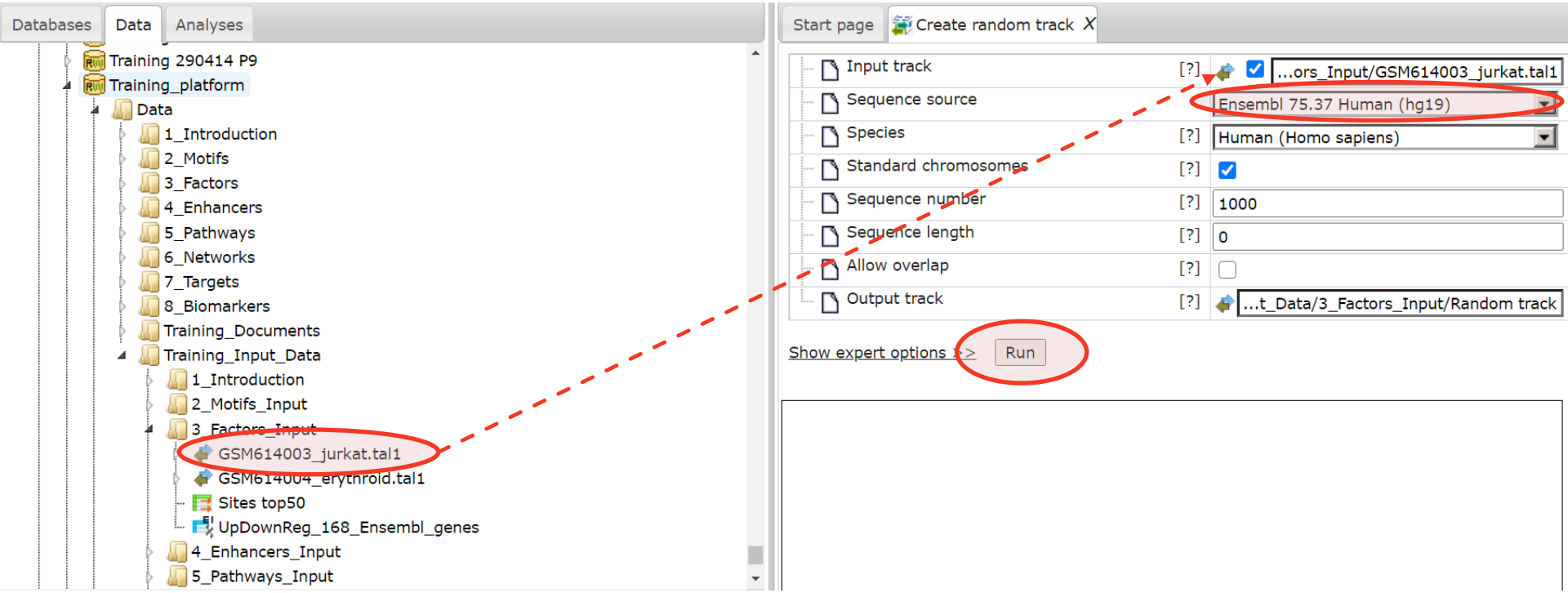
Part 3 – Enriched motifs in ChIP-seq peaks

Open the **Create random track tool**. This tool randomly samples upstream regions of genes that do not overlap with and has the same distribution of lengths as a specified set of intervals. The random intervals will present our background set of genomic sequences for subsequent binding site analysis.



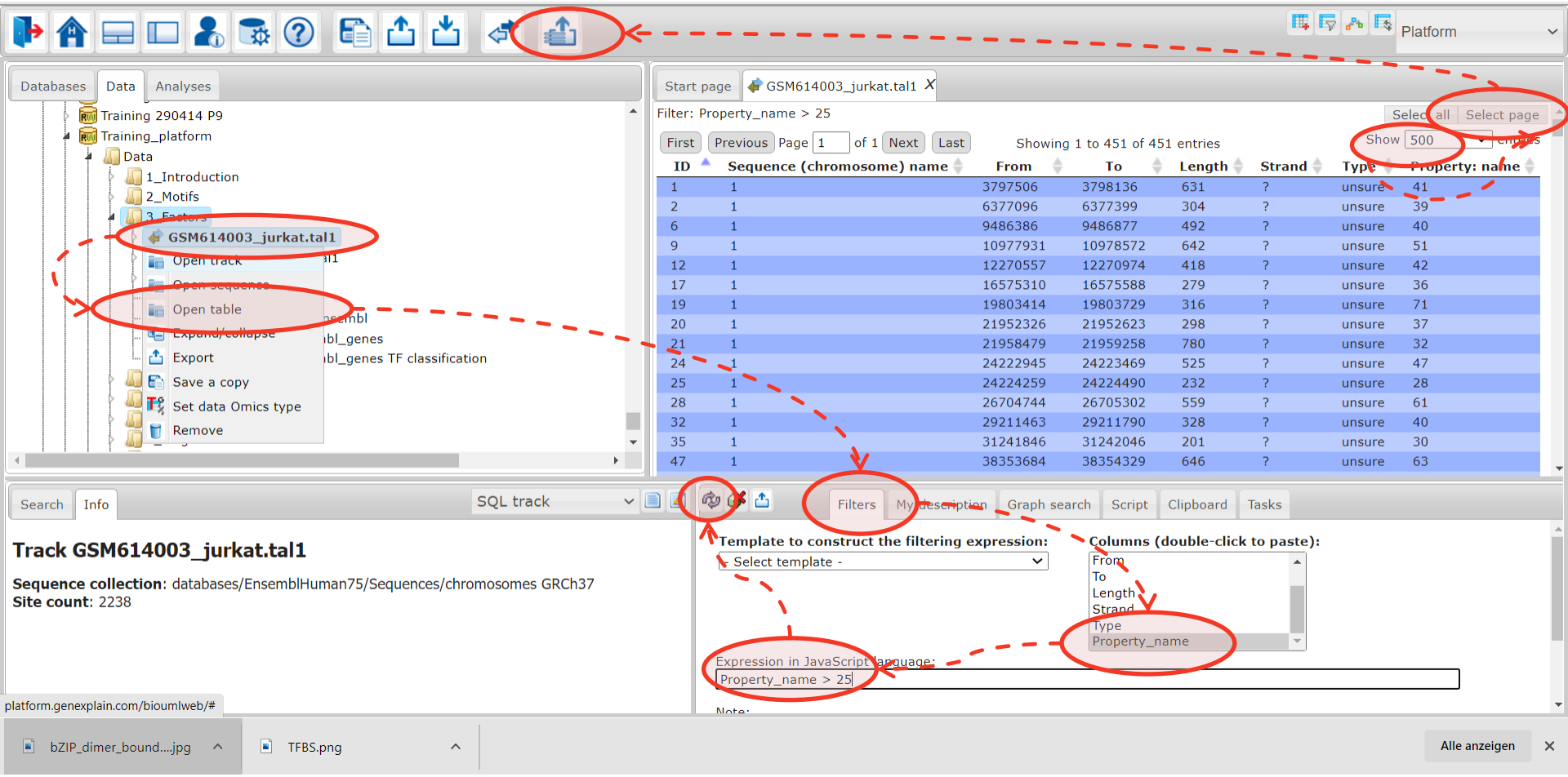
Part 3 – Enriched motifs in ChIP-seq peaks

Navigate to the *jurkat* cell track and drag-and-drop the item onto the **Input track** field of the tool. Change the Sequence source to Ensembl 75.37 Human (hg19). Click **Run** when parameters are set as shown below.



Part 3 – Enriched motifs in ChIP-seq peaks

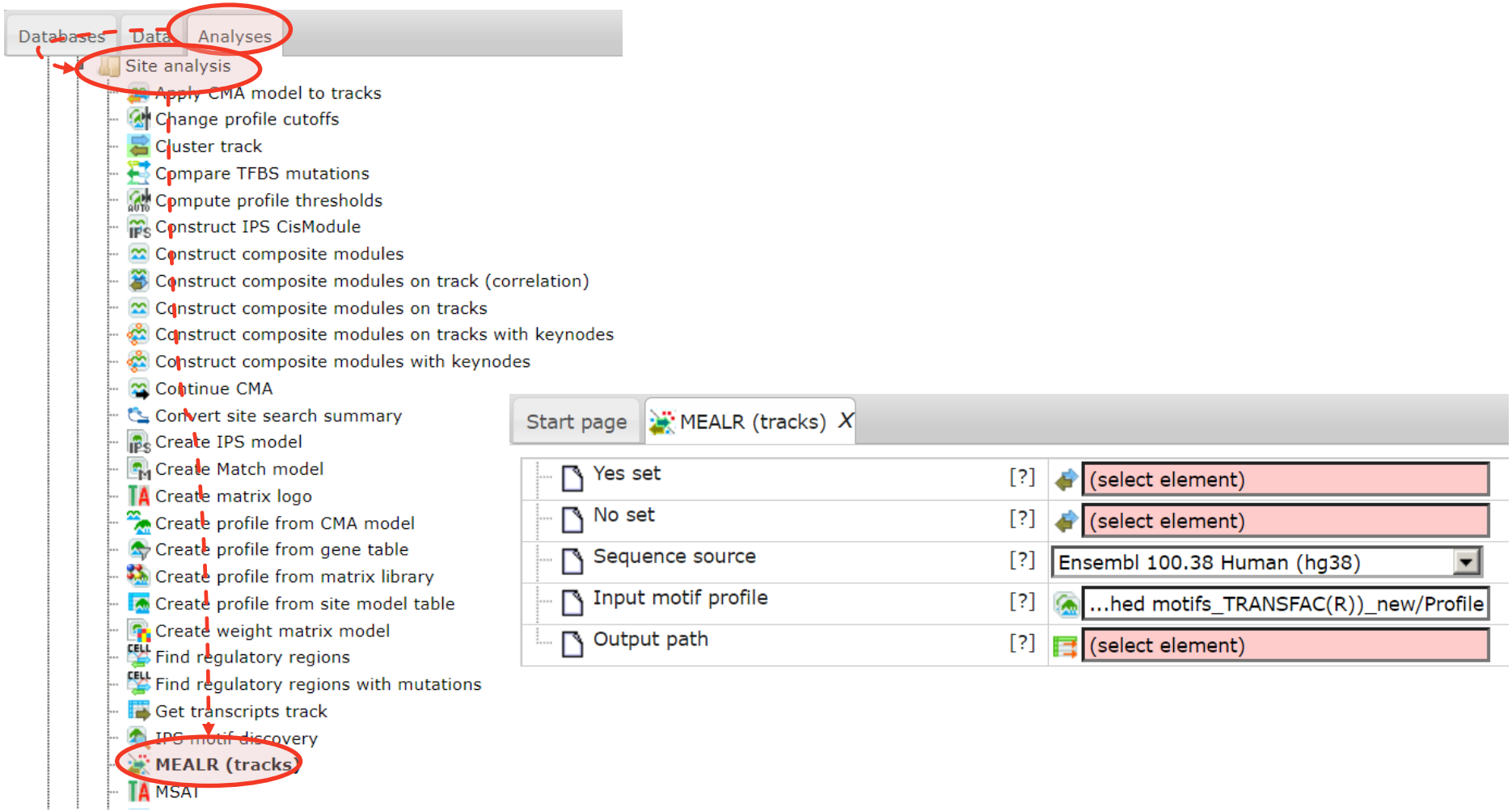
The next goal is to extract a subset of intervals, as we are not going to analyze binding sites in all peaks. Open the interval table and follow the steps described below to extract an interval subset.



Part 3 – Enriched motifs in ChIP-seq peaks

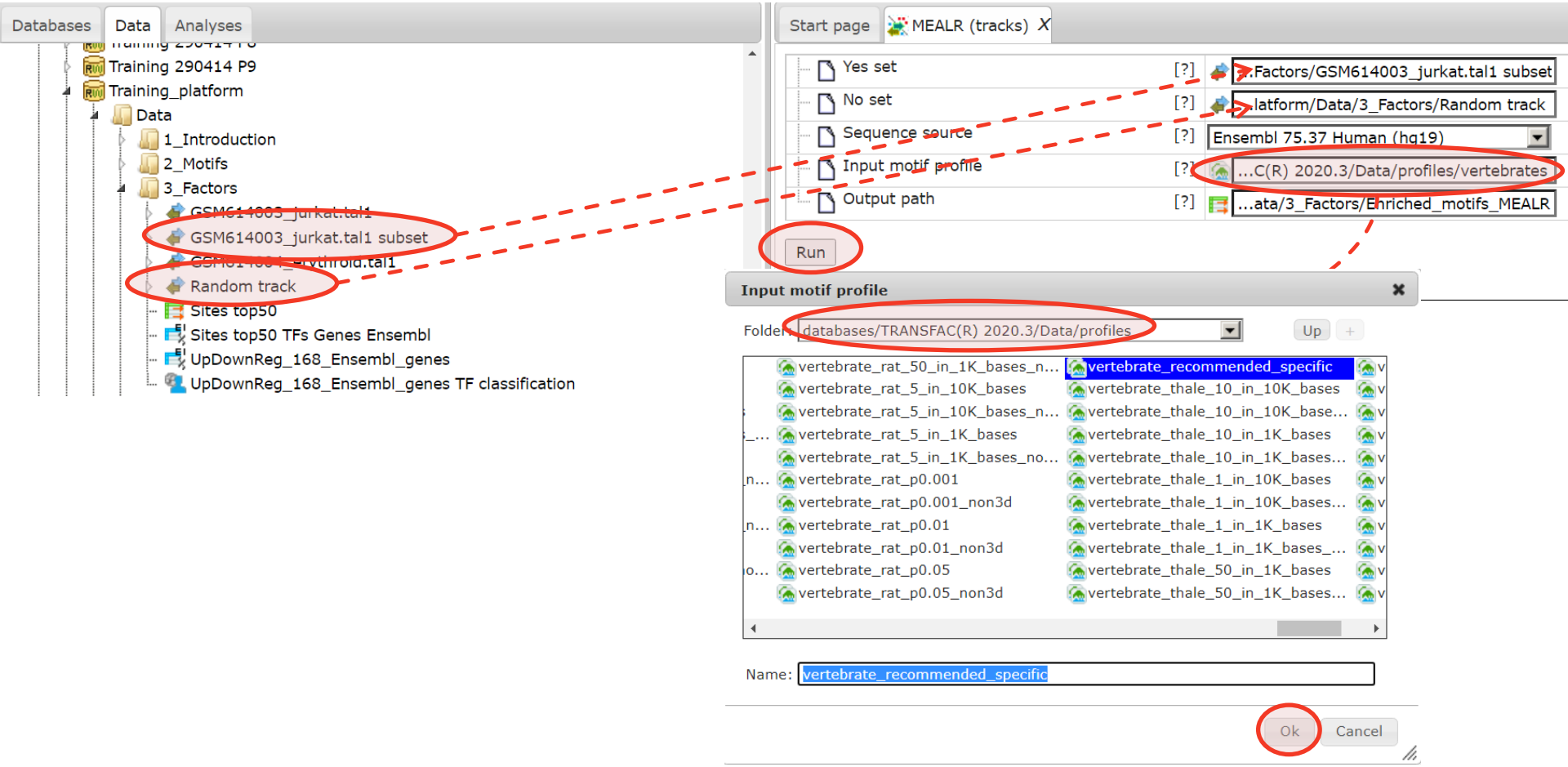
We are now ready to compare binding sites in TAL1-bound regions with those in the sampled genomic background. Here we will apply the MEALR tool which finds a weighted set of discriminating motifs using sparse logistic regression.

Use the steps shown below to open the MEALR tool.



Part 3 – Enriched motifs in ChIP-seq peaks

In MEALR, specify **Yes** and **No** sets by dragging the interval items of the TAL-1 peak subset and of the random track onto respective fields. Navigate to the **vertebrate recommended specific** profile of **TRANSFAC 2020.3** to specify the set of motifs (a file navigator opens when clicking on the field **Input motif profile**). Click **Run** when ready.



Part 1 – Enriched motifs in ChIP-seq peaks

Like the original study, MEALR identifies GATA, Runx-type, and ETS motifs as dominant patterns. In addition, it proposes also an MYB motif at the top of the list.

Start page | MEALR (tracks) X | Enriched_motifs_MEALR X

Edit | Apply | Cancel

First | Previous | Page 1 of 2 | Next | Last

Showing 1 to 50 of 77 entries

Show 50 entries

ID	Coefficient
V$TAL1_04	0.35132
V$AML2_03	0.16593
V$GATA1_10	0.11608
V$NGN2_02	0.11212
V$ETV2_03	0.09975
V$FLI1TCF3_01	0.09506
V$AML1_02	0.06003
V$HAND2_02	0.05756
V$COREBINDINGFACTOR_Q6	0.0532
V$GATA2_11	0.04998
V$ZNF563_04	0.04472

« Enriched_motifs_MEALR ... X

ENSG00000188227	zinc finger protein 793	ZNF793	V$ZNF79[illegible]_01	0.02102
ENSG00000141510	tumor protein p53	TP53	V$P53[illegible]3	0.02173
ENSG00000144792	zinc finger protein 660	ZNF660	V$ZN[illegible]660_01	0.02558
ENSG00000090447	transcription factor AP-4	TFAP4	V$[illegible]P4_03	0.02621
ENSG00000178403	neurogenin 2	NEUROG2	V$NEUROG2_03	0.02759
ENSG00000175691	zinc finger protein 77	ZNF77	V$ZNFPT1_02	0.02845
ENSG00000187987	zinc finger and SCAN domain containing 23	ZSCAN23	V$ZNF390_02	0.03276
ENSG00000148737	transcription factor 7 like 2	TCF7L2	V$TCF4_07	0.04458
ENSG00000188868	zinc finger protein 563	ZNF563	V$ZNF563_04	0.04472
ENSG00000164107	heart and neural crest derivatives expressed 2	HAND2	V$HAND2_02	0.05756
ENSG00000159216	RUNX family transcription factor 1	RUNX1	V$AML1_02	0.06003
ENSG00000071564	transcription factor 3	TCF3	V$FLI1TCF3_01	0.09506
ENSG00000151702	Fli-1 proto-oncogene, ETS transcription factor	FLI1	V$FLI1TCF3_01	0.09506
ENSG00000105672	ETS variant transcription factor 2	ETV2	V$ETV2_03	0.09975
ENSG00000102145	GATA binding protein 1	GATA1	V$GATA1_10	0.11608
ENSG00000020633	RUNX family transcription factor 3	RUNX3	V$AML2_03	0.16593
ENSG00000162367	TAL bHLH transcription factor 1, erythroid differentiation factor	TAL1	V$TAL1_04	0.35132

A Jurkat versus control

	Class	Motif	Score
1	Runx	VCCACA*	49
2	Gata	(C)TTATCT*	45
3	E-box(GC)	CAGCTG	40
4	Ets	CAGGAAR	28
5	Gata	AGATAA	19
6	Runx	AACCACA	17
7	?	GCAGVC	17

From Figure 6, **Palii et al.**, EMBO J. 2011, 30:494-509

Practical session completed

info@genexplain.com