Genome Network Project

Cis-element analysis system

Cis-Finder

User Manual

contents

1. About Cis-Finder

Cis-Finder is a tool that predicts candidate motifs (called Cis Elements candidates) from DNA sequences in genetic promoter region.

Cis-Finder will prevent omission on detecting motifs. Since it executes analyzing with combined representative algorithms for DNA sequence prediction, it achieves to detect characteristic patterns of motifs utilizing predicted data from those algorithms.

4 algorithms/tools (Consensus, MEME, Gibbs Sampler, MDScan) are used in combination.

You can set parameters in each algorithms.

Followings are the main use of Cis-Finder.

1. To predict sequences that exist

① predict the sequence information that exists in common within the transcription control region of functionally relevant gene(transcript factor binding site).

② To predict common sequence present in the transcription control region of genes in different species evolved from a common ancestral gene (orthologous gene).

Figure1 Top Page of Cis-Finder

2. Aanalysis Execution

2.1 Loading query sequence

For analysis, you are expected to prepare query sequence data in multi-FASTA format. To load data, select one of those procedures below.

①copy & paste ・・・

input the query sequence by copy-and-paste.

②file upload ・・・

upload data file from a local disk.

③Biomart ・・・

Obtain the query sequence from Biomart.

(reference[2.2 Obtaining query sequence using BioMart.])

Figure 2 Query sequence specified Cis-Finder

2.2 Obtaining query sequence using BioMart.

Cis-Finder holds search funciton of BioMart and is able to get ortholog gene sequences of human and other species registered in the Ensembl database.

(1) BioMart keyword search & species specification. (Figure3)

When you click the link as ③ of Figure2 points, BioMart Keyword search screen appears (Figure3)。

① Keyword ・・・

- If you want to search by keywords, select the "Gene Symbol" radio button, and enter the keyword (Keyword is expected to be the exact match).
- If you want to search by the Ensembl ID, select the "Ensembl Gene ID" radio button, and then enter the ID. (The ID is expected to be the human Ensembl Gene ID)

② Ortholog ・・・ specify the species.

③ When you press the "Get Gene ID-List" button, gene & region setting screen shows up (Figure4).

(2) BioMart gene search: gene & region settings (Figure4)

① Sequence Type ・・・ specify the type of sequence.

② Sequence Region ・・・ If you choose "_flank" at the Sequence Type, select upstream or downstream of sequence and specify the search range in number.

③ When you press the "Get Sequence" button, you will get sequence with customized condition you made.

Figure 3 BioMart keyword search & species specification.

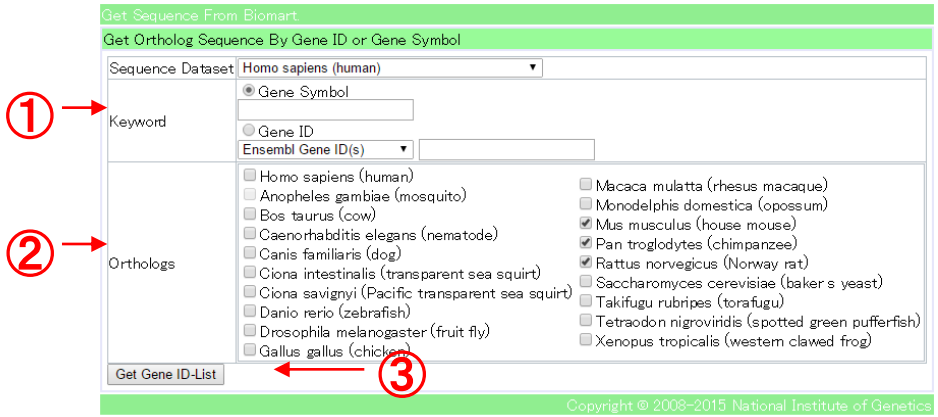
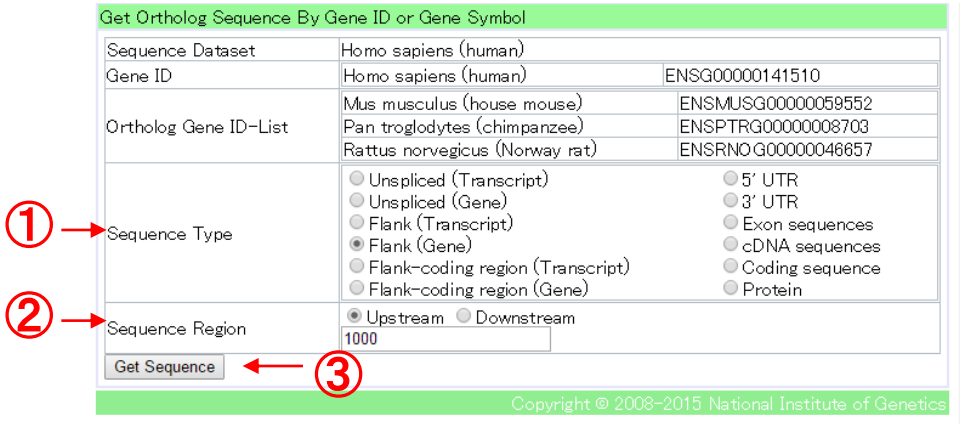


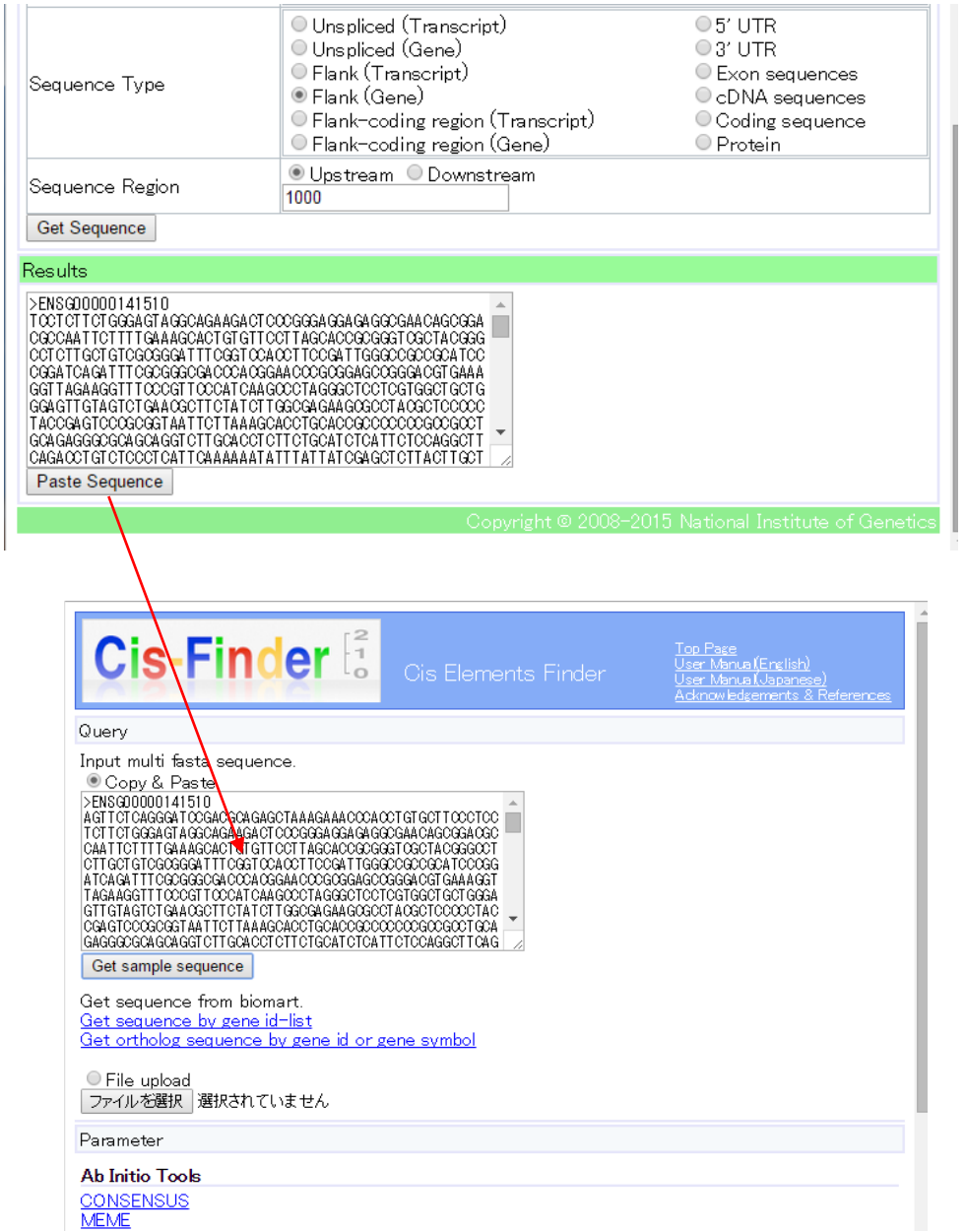
Figure 4 BioMart gene search: gene & region settings



(3) Obtaining sequence data (Figure 5)

- When the sequence data successfully being acquired, it appears in the "Results" area as Figure 5 above.
- When you click the "Paste Sequence" button, the sequence data which shows in the results area will be copied and pasted to the query pane as Figure 5 below indicates.

Figure 5　Obtaining sequence data



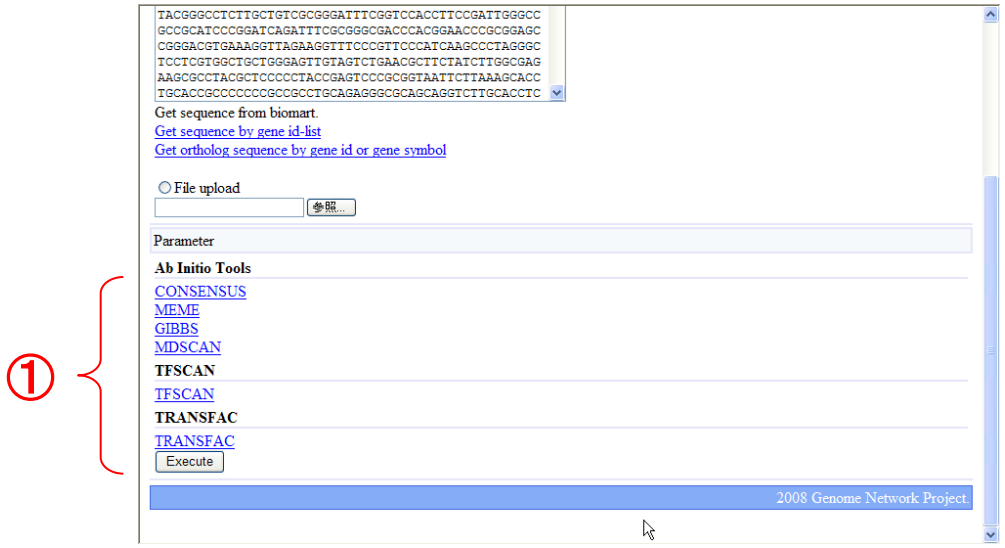
2.3 Parameter Settings

As shown in Figure 6, you can specify various parameters for the cis-element analysis.

① Parameter settings for the representative algorithms.

You can specify parameters for each algorithm of Consensus,MEME, Gibbs sampler,MDScan and TFSCAN. (There is no parameter to TRANSFAC)

Figure 6 Parameter Settings



2.4 Parameter settings for each algorithm

The parameter setting screen of each algorithm is shown in Figures 7 to 10. To return to the default value, click the "set default" button at the top of a setting table.

Figure 7 Parameter settings of Consensus

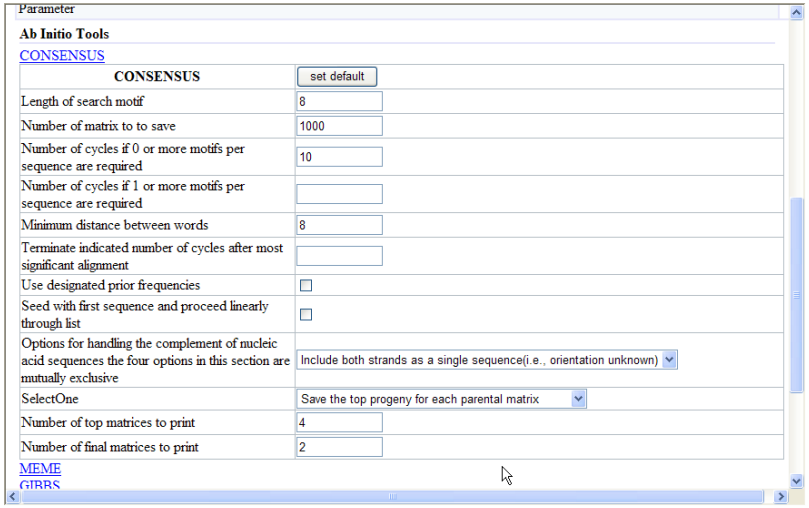


Figure 8 Parameter settings of MEME

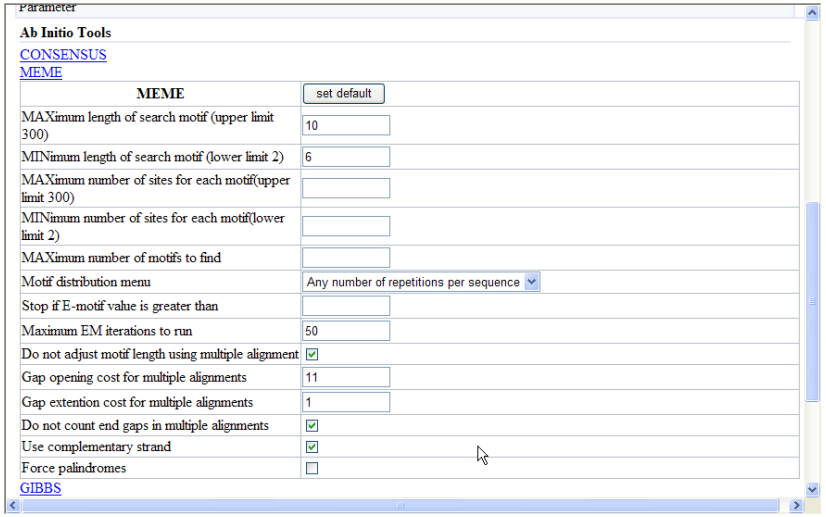


Figure 9 Parameter settings of Gibbs sampler

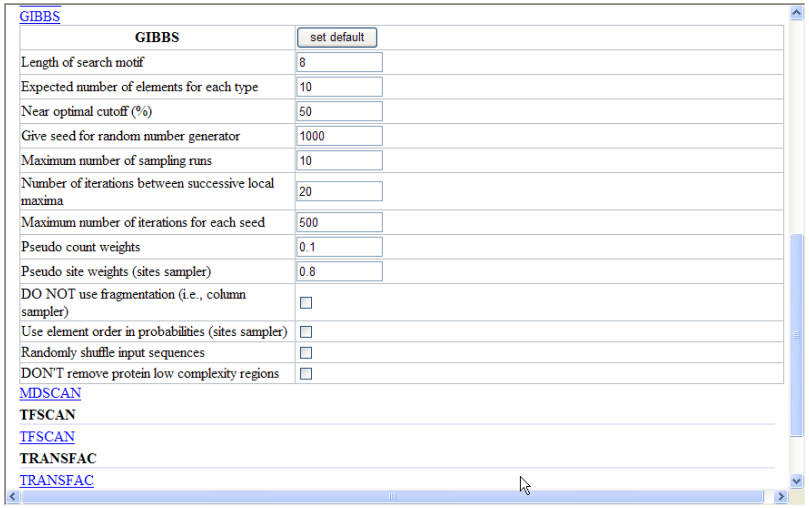
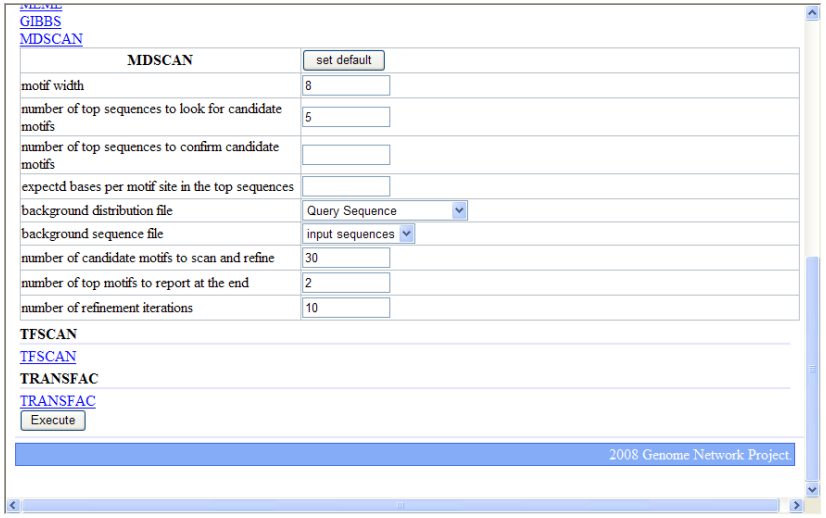


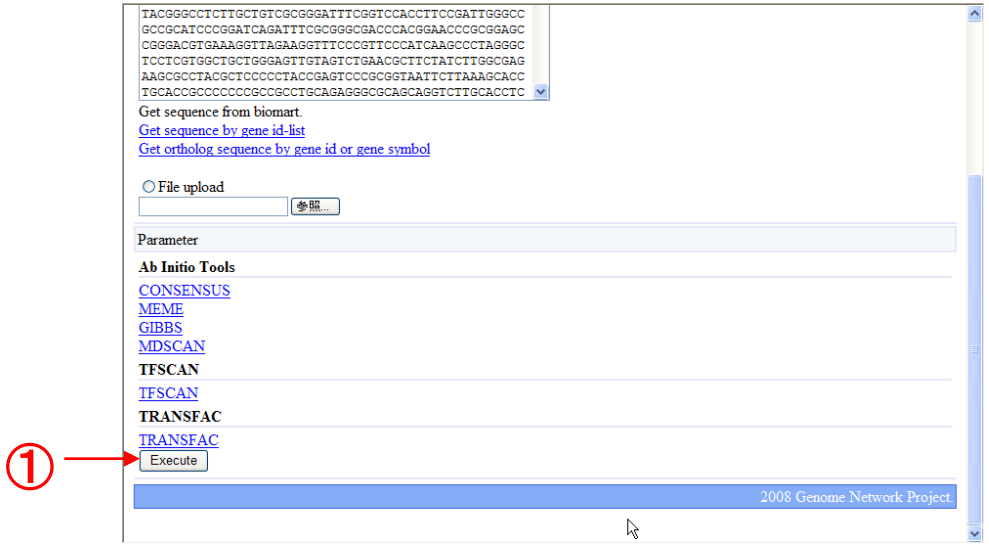
Figure 10 Parameter settings of MDScan



2.5 Analysis execution

① presses the “Execute” button to start the analysis.

Figure 11 Analysis execution



3. Analysis Result

3.1. Displaying analysis result

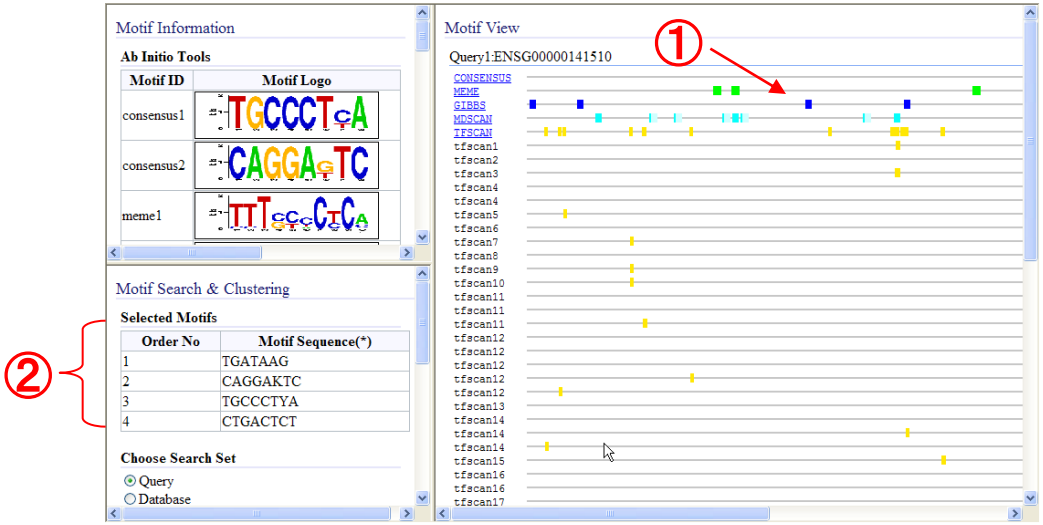
After the analysis being completed, you can view detailed result for each sequence as shown in Figure 12.

Figure 12 Displaying analysis result

3.2 Motif search and clustering execution.

If you click on the colored box (①) which inplies a motif, its sequence and location will be shown on the left side of the window (②).

図13 selection of motif



① Selection of Database ・・・

Select the database you want to search with.

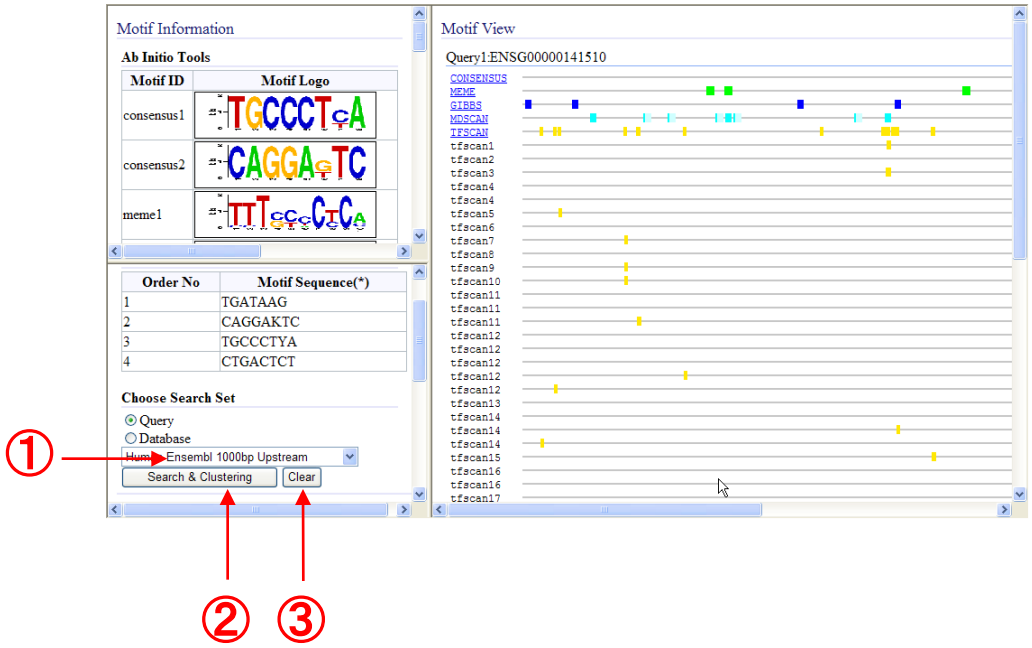
・Ensembl ・・・ Sequences upstream or downstream 1kb of Ensembl genes of human and mouse

・CAGE ・・・ sequence of unknown transcription start point

②To start searching, please click the ″Search & Clustering″ button.

④ To clear motif selections, please press ″Clear″ button.

図14 seach motif and execute clustering



3.3 Motif search and clustering result.

When the motif search being completed, you will see the screen shown in Figure 15.

① Selectetd Motifs ・・・

Displays sequence pattern and number of hits of the selected motif.

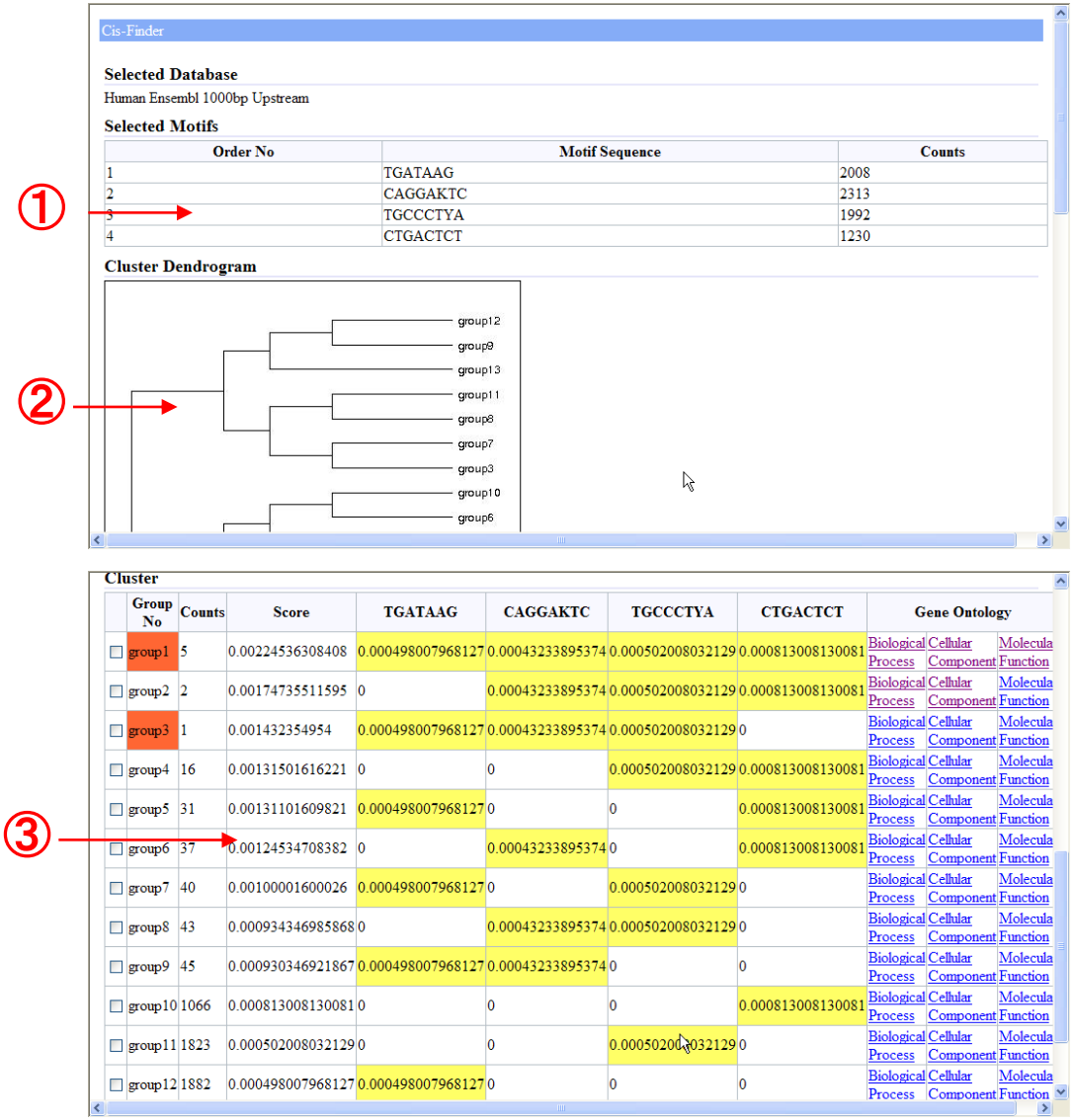
②Cluster Dendrogram・・・

shows the clustering results.

③Cluster・・・

Genes from the specified database are grouped by wheather each of those has a target motif or not. The Score row applies to inverse value of the number of hits.

Figure 15 motif search and clustering result



You can confirm detailed information for clusters by gene list and Gene Ontology information.

①Gene Ontology・・・

You can view Gene Ontology information belongs to each cluster.

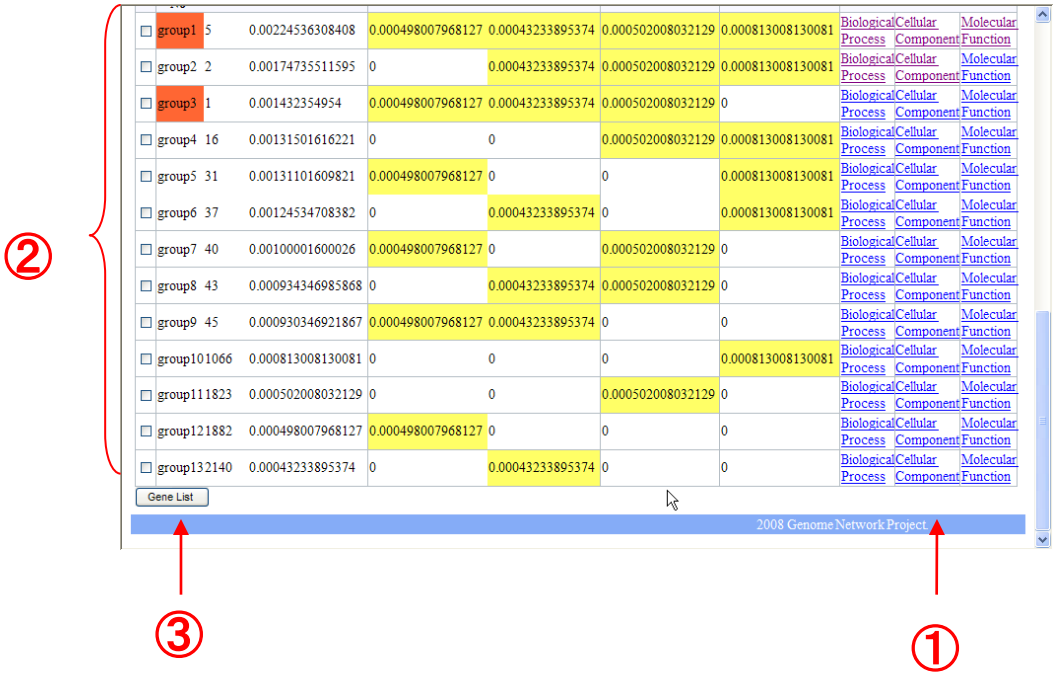
→3.4 For details, please see 3.4 Gene Ontology information.

②Gene List・・・

Please mark a checkbox for the groups (②) that you want to view, and click the "Gene List" button (③).

→3.5 For details, please see 3.5 Gene List.

図16 detail of each cluster



3.4 Gene Ontology Information

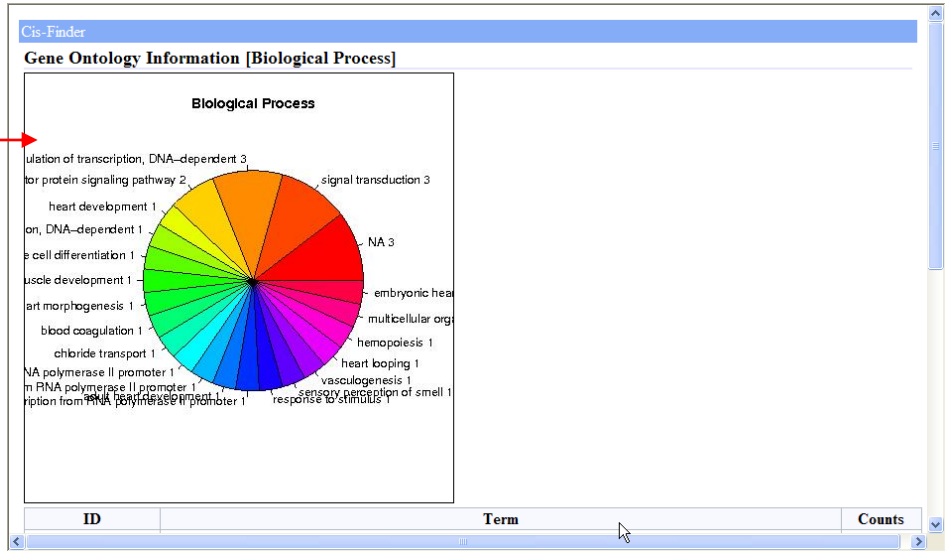
You can browse following Gene Ontology information in this pane.

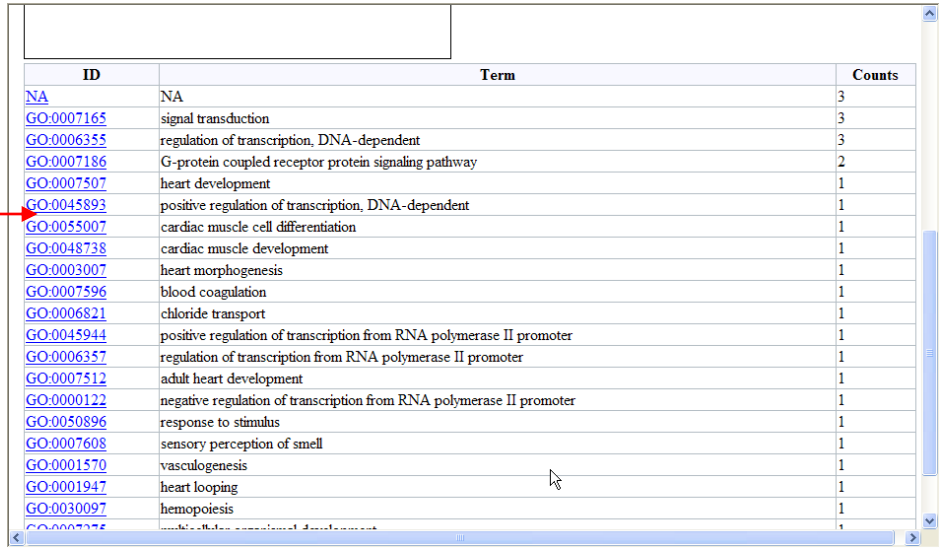
- Biological Process
- Cellular Component
- Molecular Function

① Display a pie chart which indicates ratio of Gene Ontorogy Term for genes belong to the cluster.

② Display a list of Gene Ontorogy Term.

Figure 16 Gene Ontology information





3.5 Gene List

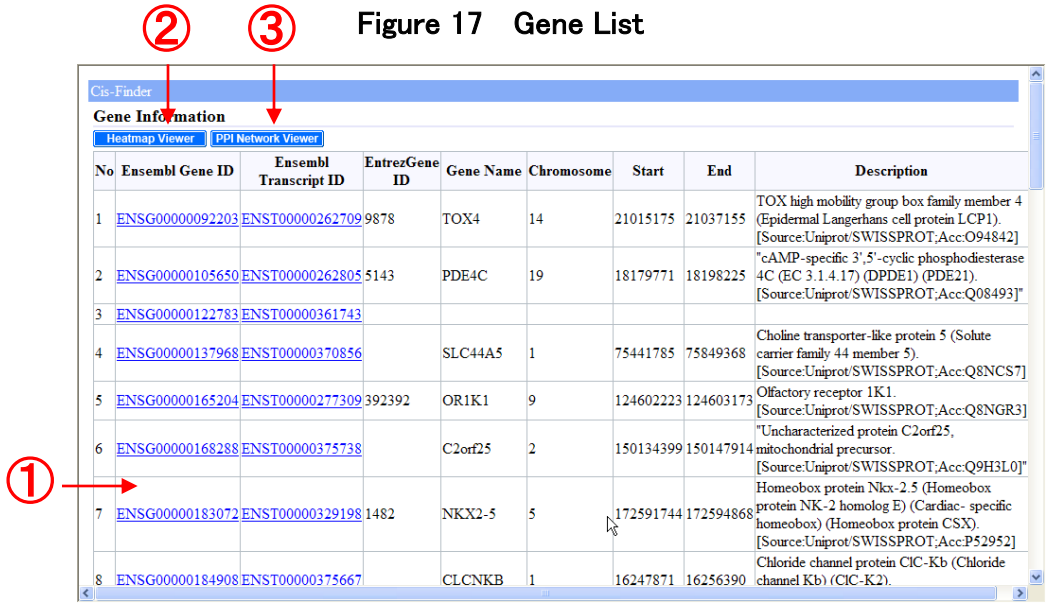
Display gene list belongs to the cluster. In addition, it shows information of gene expression and mutual interaction between protains. (Link to genome network platform)

① Display information of gene annotation.

②Heatmap Viewer ・・・ display expression informaiton for each tissues. (Figure 18)

③PPI Network Viewer ・・・ Display information of Protein-protein Interactions. (Figure 19)

Figure 17 Gene List



Gene Information

Heatmap Viewer | PPI Network Viewer

No	Ensembl Gene ID	Ensembl Transcript ID	EntrezGene ID	Gene Name	Chromosome	Start	End	Description
1	ENSG00000092203	ENST00000262709	9878	TOX4	14	21015175	21037155	TOX high mobility group box family member 4 (Epidermal Langerhans cell protein LCP1). [Source:Uniprot/SWISSPROT;Acc:O94842]
2	ENSG00000105650	ENST00000262805	5143	PDE4C	19	18179771	18198225	"cAMP-specific 3',5'-cyclic phosphodiesterase 4C (EC 3.1.4.17) (DPDE1) (PDE21). [Source:Uniprot/SWISSPROT;Acc:Q08493]"
3	ENSG00000122783	ENST00000361743						
4	ENSG00000137968	ENST00000370856		SLC44A5	1	75441785	75849368	Choline transporter-like protein 5 (Solute carrier family 44 member 5). [Source:Uniprot/SWISSPROT;Acc:Q8NCS7]
5	ENSG00000165204	ENST00000277309	392392	OR1K1	9	124602223	124603173	Olfactory receptor 1K1. [Source:Uniprot/SWISSPROT;Acc:Q8NGR3]
6	ENSG00000168288	ENST00000375738		C2orf25	2	150134399	150147914	"Uncharacterized protein C2orf25, mitochondrial precursor. [Source:Uniprot/SWISSPROT;Acc:Q9H3L0]"
7	ENSG00000183072	ENST00000329198	1482	NKX2-5	5	172591744	172594868	Homeobox protein Nkx-2.5 (Homeobox protein NK-2 homolog E) (Cardiac- specific homeobox) (Homeobox protein CSX). [Source:Uniprot/SWISSPROT;Acc:P52952]
8	ENSG00000184908	ENST00000375667		CLCNKB	1	16247871	16256390	Chloride channel protein ClC-Kb (Chloride channel Kb) (ClC-K2).

Figure 18 Heatmap Viewer

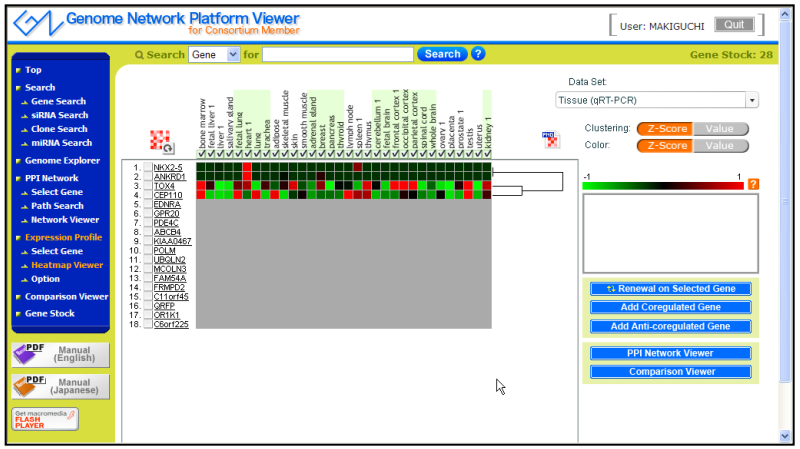


Figure 19 Information of protein-protein interactions