



The toolbox of TRANSFAC 2.0

User guide release 3.1

Table of contents

Getting started	2
Welcome to MATCH Suite	2
Basic interface	2
Tissue expression based analysis	3
Gene set analysis	3
Launching the analysis	3
Selecting the input gene list	3
Specifying the launch parameters	8
Optimizing the input gene set by GO functional classification (optional step)	10
Confirming the launch parameters and starting the analysis	11
Viewing the results	12
Operating in the <i>Launch history</i>	12
Results visualization	14
Single gene analysis	25
Launching the analysis	25
Specifying the input gene	25
Specifying the launch parameters	28
Confirming the launch parameters and starting the analysis	29
Viewing the results	30

Operating in the Launch history	30
Results visualization	32
GO categorization-based gene set analysis	40
Launching the analysis	40
Selecting the input gene list	40
Composing an input gene list in the TRANSFAC® database	40
Selecting input gene list from the MATCH Suite interface	42
Specifying the launch parameters	46
Filtering the input gene set by GO functional classification (optional step)	48
Confirming the launch parameters and starting the analysis	49
Viewing the results	50
Operating in the Launch history	50
Results visualization	52
Factor view table	54
Matrix view table	55
Genes table	57
Genome Browser	57
Heatmaps	60
Switching to geneXplain platform	63

Getting started

Welcome to MATCH Suite

Welcome to MATCH Suite – the toolbox of TRANSFAC 2.0, which comprehensively addresses the syntax and semantics of gene regulation and allows you to identify the transcription factors regulating the genes(s) of your interest.

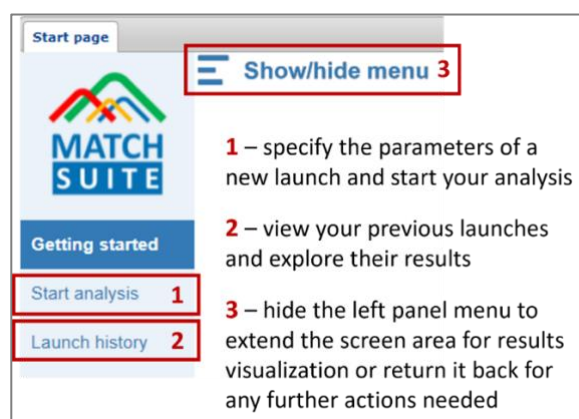
MATCH Suite can be launched on an individual Human gene or on a list of Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast genes. Depending on your selection, different analyses will be ran to identify the activation patterns of your input (see Methods document linked to the analysis report of each run for further details).

Explore the functional enrichment of your gene set, find only transcription factors (TFs) expressed in the tissue of your interest or those belonging to the functional categories in focus, and much more! Navigate to *Start analysis* to immediately launch the comprehensive and fully automatized gene regulation study.

Basic interface

The MATCH Suite interface is based on two main sections:

- (1) **Start analysis** – navigate to this section to launch a new analysis
- (2) **Launch history** – navigate to this section to view the history of your previous launches and open their results



The **Show/hide** menu button at the top of the screen (3) allows you to hide the left panel menu for extending the useful screen area when this is needed (this function is of particular importance when visualizing the analysis results).

Tissue expression based analysis

Please note that tissue expression based analysis is available for human genes only.

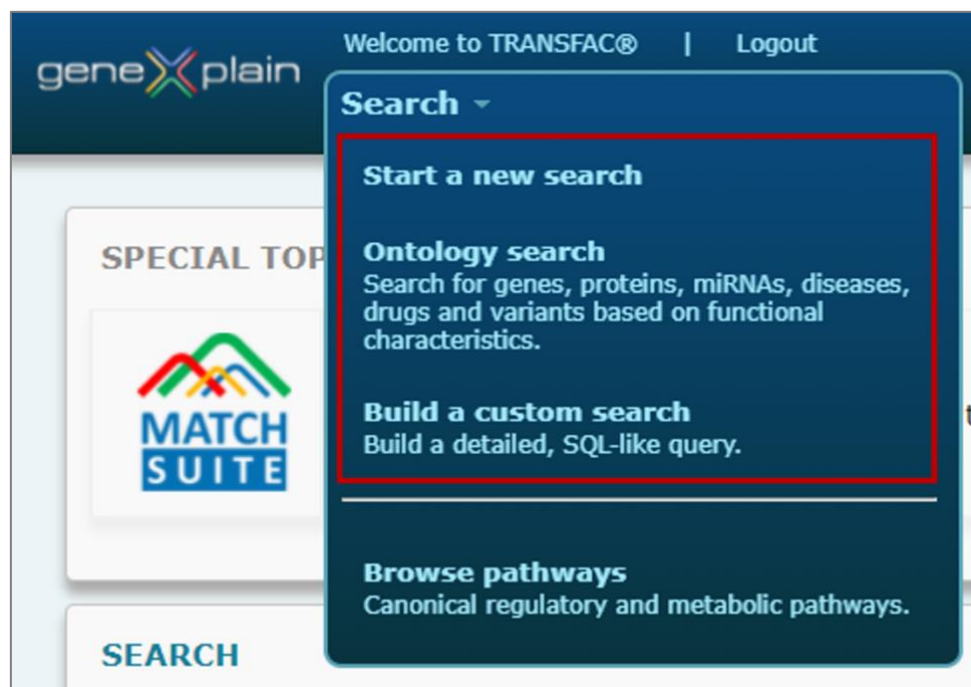
Gene set analysis

Launching the analysis

Selecting the input gene list

Composing an input gene list in the TRANSFAC® database

A particularly convenient way to compose a gene list for analysis is to run the MATCH Suite on your search results in the TRANSFAC® database. For instance, when you search for a certain disease in the standard Search field or select genes belonging to a certain Gene Ontology (GO) category (Search > Ontology search), you end up with a list of genes.



In the search result inside TRANSFAC® database you can view all found entries on one page (select "All" in the "Hits on page" list) and then select the search results

of your interest to launch the MATCH Suite gene set analysis on a respective gene list (you can use the *Mark all on page* option and then only human genes from your selection will be automatically taken to the MATCH Suite analysis upon launching it).

geneXplain Welcome to TRANSFAC® + PROTEOME™ | Logout

SEARCH MATCH Suite Tools My Data Home Taskbar Help

SEARCH **Guided Tour**

search

View more search options if you want to do a more specific search

Found 280 results for "brc" in 6 categories:

Search results

114 Genes and proteins	16 Transcription factors	0 Variants	Still searching? More search options Ontology search Custom search
0 miRNAs	15 Matrices	116 Diseases	
5 Pathways		14 Drugs	

Genes and proteins
114 of 114 total

Select results and forward them to one of the tools (Guided Tour):

Save these results Export these results Pathfinder Ontology Match Match Suite FASTA Profiles Binding factors for gene Search Filter

First 1 Last

☒ Mark all on page (To select the complete result to be forwarded to the tools, mark none.)

Hits on page: All

#	Name	Type	Species/Taxon	Description
☒	br		Drosophila melanogaster	broad [matched: 'brc' (GENES_AND_PROTEINS, MATRICES)]
☒	Arid4b		Mus musculus	Protein with high similarity to human ARID4B, which inhibits transcription and cell proliferation, and is associated with lung neoplasm, member of the RNA binding activity-knot of a chromodomain containing family, contains an RBB1NT (NUC162) domain [matched: 'brc' (GENES_AND_PROTEINS)]
☒	brn		Drosophila melanogaster	brainiac [matched: 'brc' (GENES_AND_PROTEINS)]
☒	BRCA1		Homo sapiens	Breast cancer 1 early onset, a transcription regulator that acts in double-strand break repair, apoptosis, protein folding, and mRNA cleavage, upregulated in breast, ovary, and prostatic neoplasms; mRNA is downregulated in melanoma [matched: 'brc' (GENES_AND_PROTEINS, MATRICES)]
☒	BRCC3		Homo sapiens	BRCA1-BRCA2 containing complex subunit 3, may be involved in regulation of transcription; gene mutations are associated with moyamoya syndrome, hemophilia A, and T-cell leukemia associated with ataxia telangiectasia

When you have selected the search results of your interest, press the button "Match Suite" and specify that you are studying Human genes and that it is a gene set analysis:

MATCH Suite options

? Select species: Human

? Select analysis type: Gene set

Click on *Submit* and you will be automatically transferred to the MATCH Suite tool, where you will be asked to assign a name to your gene list prior to proceeding to the next step of the analysis launch wizard:

Please name the selected gene list

Gene list name:

Click on *Next* to proceed to the parameters specification step.

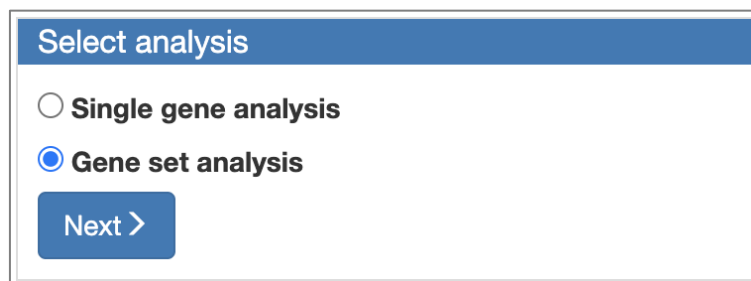
Selecting input gene list from the MATCH Suite interface

When you click on the *Start analysis* section, the system immediately navigates you to the analysis launch wizard. At the very first step of this wizard you will be asked to select the type of the model organism that you are studying:



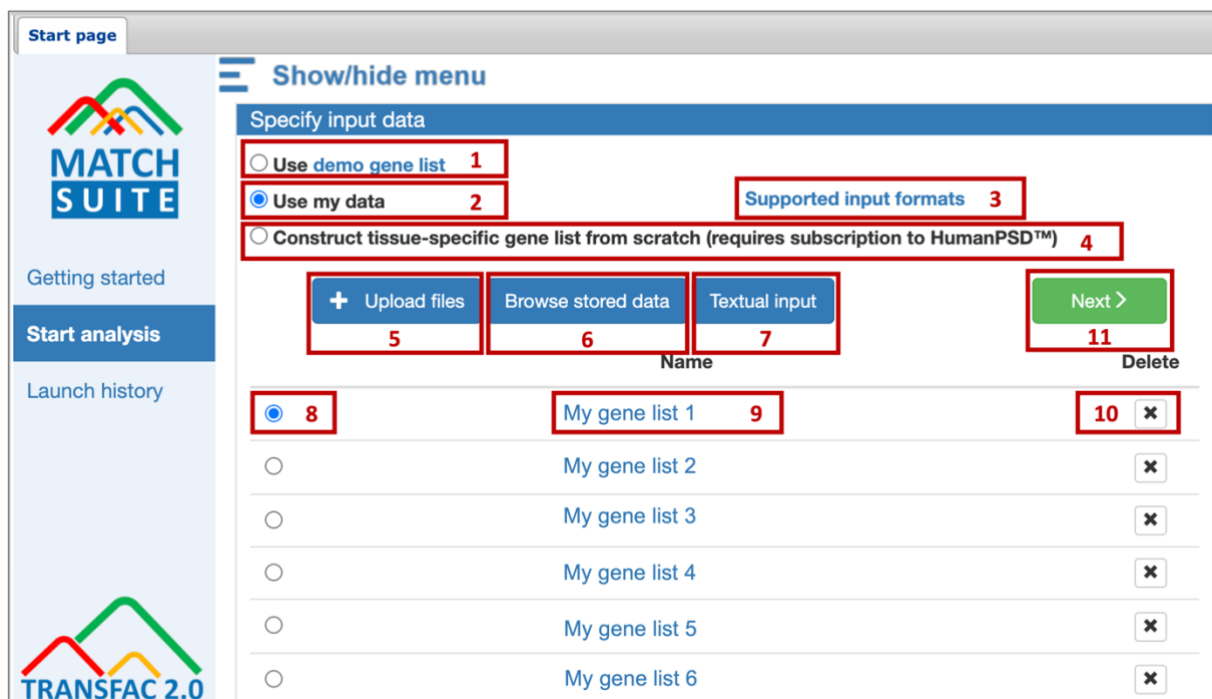
A dialog box titled "Select species" with a dropdown menu showing "Human" and a "Next >" button.

Select Human and click on Next. After that specify the analysis type you would like to perform:



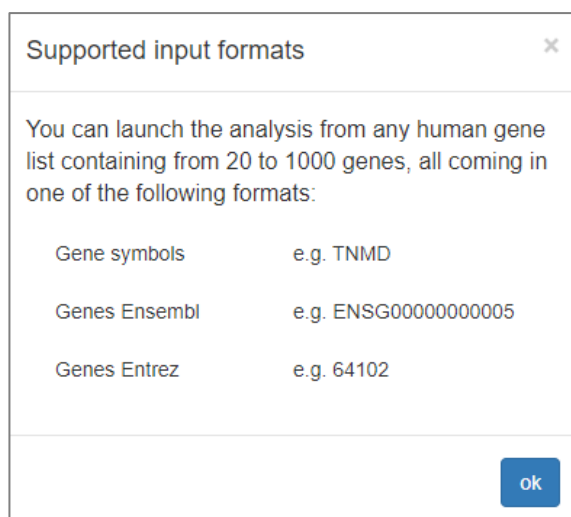
A dialog box titled "Select analysis" with two radio buttons: "Single gene analysis" and "Gene set analysis" (which is selected). There is a "Next >" button at the bottom.

If you will select the *Gene set analysis* option, you will be taken to the next step of the wizard, where you will be asked to select the gene list which should be used for the analysis.



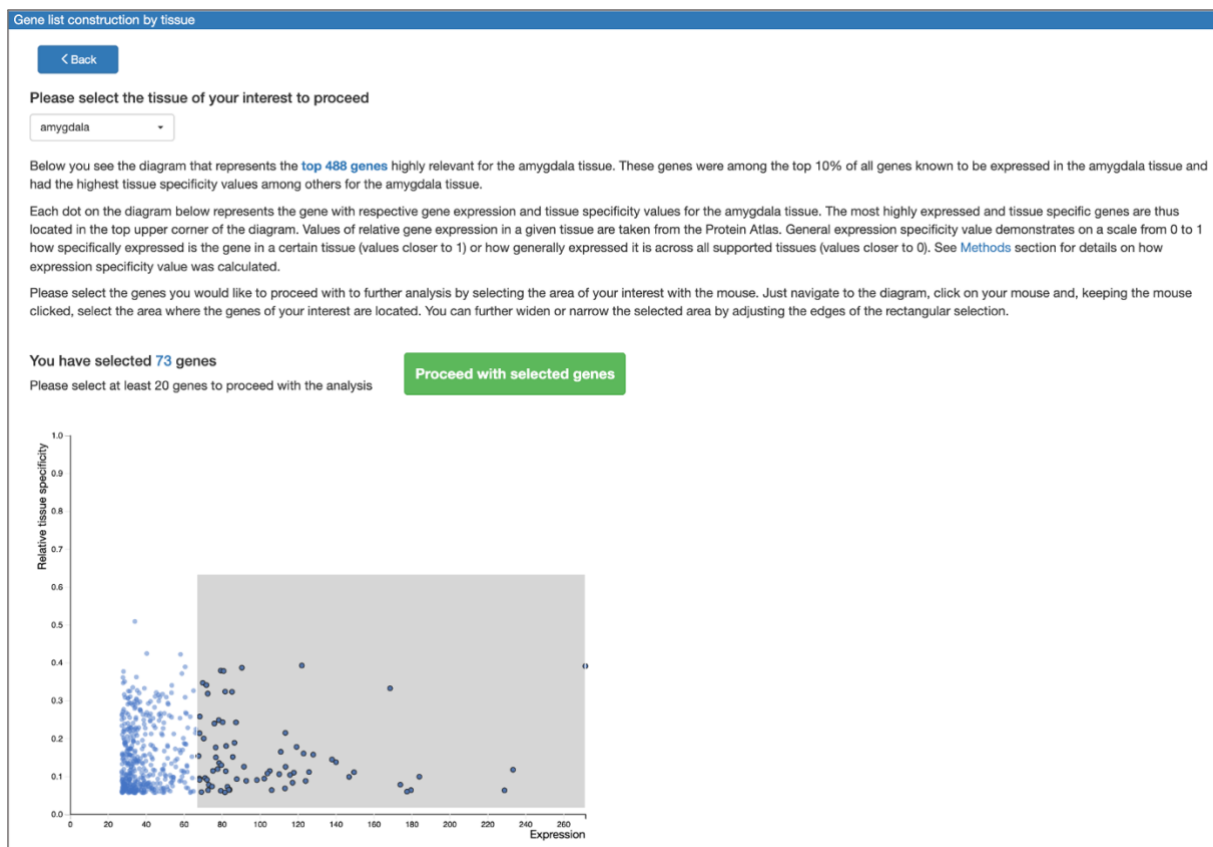
The MATCH Suite main interface. On the left is a sidebar with the MATCH SUITE logo, "Getting started", "Start analysis" (highlighted), and "Launch history". The main area is titled "Specify input data" and contains several options: "Use demo gene list" (1), "Use my data" (2), "Supported input formats" (3), and "Construct tissue-specific gene list from scratch (requires subscription to HumanPSD™)" (4). Below these are three buttons: "Upload files" (5), "Browse stored data" (6), and "Textual input" (7). A "Next >" button (11) is on the right. Below these buttons is a table with a "Name" column and a "Delete" column. The first row is selected (8) and shows "My gene list 1" (9) with a delete icon (10). Below it are five more rows labeled "My gene list 2" through "My gene list 6", each with a delete icon.

You can try launching the analysis using the **Use demo gene list** (1) option, or you can specify your own gene list by selecting the **Use my data** (2). The demo gene list contains 50 Ensembl genes, you can view them by clicking on the *demo gene list* blue link. The **Supported input formats** (3) link also is clickable, it provides you with the information about the requested format of the gene list, which can be analyzed with the MATCH Suite 3.1:



Lists of 20 to 2000 genes coming in Ensembl ID, Entrez ID or Gene symbols format are accepted by the MATCH Suite 3.1. The system will select automatically the human genes from your list, only these will be subject to the further analysis.

A tissue-specific gene list can be automatically constructed for further analysis using the **Construct tissue-specific gene list from scratch** (4) option.



This option provides you with the ability to select any tissue among the 61 supported tissues and visualize the most tissue-specific genes among the top 10% of all genes known to be expressed in the selected tissue according to the information about relative gene expression levels provided by the Protein Atlas. You can further select the genes of your interest on the interactive diagram to either proceed with them to the analysis launch or just save the selected genes into the inputs of your project (both options managed by the *Proceed with selected genes* button). Please note that tissue-specific gene list construction option is available only to the HumanPSD™ database subscription owners.

The gene list for your analysis can be uploaded from your local computer using the **Upload files** (5) option, it can be also selected from the data stored in any geneXplain platform project accessible to you by using the **Browse stored data** (6) option. In such case the selected gene list will be copied to the MATCH Suite project and taken for further analysis. The next option for input gene list specification is **Textual input** (7). This function provides you with an ability to simply copy and paste any gene list of your interest (in Ensembl, Entrez or Gene symbols format) to the dedicated textual input form:

Textual input

Gene list name:

Type or copy+paste your genes here, e.g.

ENSG000000000005

ENSG000000000419

ENSG000000000457

...

Cancel

Upload

The last option to specify an input gene list is to select it from the gene lists that were previously used by you as inputs for the MATCH Suite analysis (8). The names of the previously used gene lists are clickable and the respective gene list will open upon the click on its name (9). You can manage the gene lists stored in your project and permanently delete the unnecessary ones using the **Delete** option (10). Please note that the maximum size of your MATCH Suite project is 2 GB. You can free up space by deleting the unnecessary gene lists and analysis results if that would be needed (see *Operating in the Launch history* section below for further info)

To select the gene list for the current analysis run, mark it with the radio button in the list of all currently available inputs (8) (any newly uploaded gene list will automatically appear in this list).

Please note that regardless of the input gene list source, the selected gene list will be checked by the MATCH Suite for correspondence to the input gene lists requirements and only human gene lists containing from 20 to 2000 human genes in supported formats will be accepted for further analysis.

After selecting the input, click **Next** (11) to proceed.

Specifying the launch parameters

In the next step of the analysis launch wizard the MATCH Suite will ask you to select the type of the analysis you want to perform:

Start page

MATCH SUITE

Getting started

Start analysis

Launch history

Show/hide menu

Select analysis focus

☒ **Tissue expression based analysis**
Info: Search for transcription factors regulating your gene set based on the expression profiles of these factors in the selected tissue

☐ **GO categorization based analysis**
Info: Search for transcription factors regulating your gene set based on the functional categorization of these factors in respect to the selected GO terms

Back **Next >**

Select the *Tissue expression based analysis* option and click on *Next* to proceed to the parameters specification step.

At the parameters specification step fill in the following form:

Start page

MATCH SUITE

Getting started

Start analysis

Launch history

Show/hide menu

Specify parameters

Name of the analysis launch 1 demo-gene-list 1

Studied tissue 2 Nothing selected

Tissue selection will allow to use tissue-specific promoters from the [FANTOM5](#) database and identify the transcription factors regulating your gene set that are known to be expressed in the selected tissue.

Promoter range 3

☒ Use default promoter range [-500,+100]

☐ Customized promoter range

From Max -5000 to Max +1000

Gene set optimization 4

☐ Narrow my gene list to genes coming from certain GO categories

< Back **Next >** 5

TRANSFAC 2.0

You can give a name to your analysis launch or keep the default name suggested by the system in the **Name of the analysis launch** field (1). This name will be further used in the launches history for easy selection of the run of your interest.

Optionally you can specify the tissue of your interest from the dropdown list of supported tissues in the **Studied tissue** field (2). The provided tissues are the tissues from [FANTOM5](#) database with specific coordinates for the transcription start sites (TSSs) that allow to select tissue-specific promoters for the subsequent

analysis. The selected tissue is also used further for identification of transcription factors that are known to be expressed in the tissue of your interest. Please refer to the [Methods](#) document for further details.

The promoter range that will be used for the search of TFBS in your analysis run is specified by the **Promoter range** parameter (3). By default the promoter range used by the MATCH Suite is [-500,100] relative to the TSS. You can specify a customized promoter range limited to the maximum of -5000 and +1000 from TSS.

Your input gene set can be optimized by functional enrichment using the **Gene set optimization** option (4). Having clicked on it, you will be offered to select between the supported GO categories:

Gene set optimization

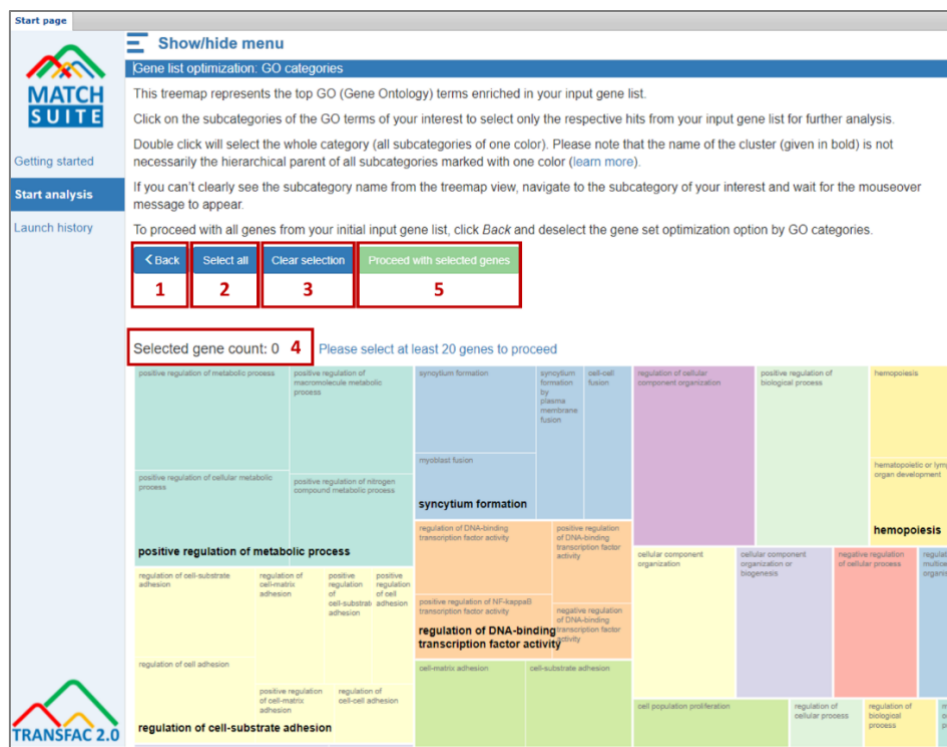
- ☒ Narrow my gene list to genes coming from certain GO categories
 - ☒ biological process
 - ☐ cellular component
 - ☐ molecular function

Depending on your selection, the tree map visualization of functional classification of your input gene set will be built either on biological processes, or on cellular components, or on molecular functions. If you do not wish to optimize your input gene set by functional categories, simply do not tick the *Gene set optimization* option.

When done with parameters selection, click on **Next** (5) to proceed either to GO functional optimization of your input gene set or skip this step and start the analysis immediately by confirming the selected analysis launch parameters.

Optimizing the input gene set by GO functional classification (optional step)

In case you have selected to optimize your input gene list by GO functional classification at the parameters specification step, the tree map of GO biological processes, cellular components or molecular functions will be constructed for your input gene set depending on your selection.



You can click **Back** (1) to return to the parameters specification step and cancel the GO optimization of your input gene list by unticking the *Gene set optimization* option. Otherwise you are requested to select the GO categories of your interest, following the instructions provided at the top of the screen. Only genes belonging to the selected GO categories will be taken by the MATCH Suite for further analysis.

The **Select all** (2) button will allow you to select for further analysis all genes, belonging to all GO categories enriched in your input gene set (genes belonging to all GO categories visualized on the tree map). The **Clear selection** (3) button will deselect all previously selected GO categories and will reset the gene count to 0. The number you will see next to the **Selected gene count** (4) will show you the total amount of genes underlying the currently selected GO categories. Once this number will reach the minimum of 20 genes that are requested for launching the analysis, the **Proceed with selected genes** (5) button will become active and you will be able to launch the analysis using the optimized gene set.

Confirming the launch parameters and starting the analysis

After specifying all parameters for your launch, the MATCH Suite wizard will ask you to confirm your selection:

Start page

MATCH SUITE

Getting started

Start analysis

Launch history

TRANSFAC 2.0

Show/hide menu

Selected launch parameters

Launch name: demo-gene-list 1

Input gene list: demo-gene-list optimized by GO categories

Selected tissue: Nothing selected

Promoter range: [-500,+100]

Selected GO categories: positive regulation of metabolic process
 <GO:0009893> positive regulation of metabolic process
 <GO:0031325> positive regulation of cellular metabolic process
 <GO:0010604> positive regulation of macromolecule metabolic process
 <GO:0051173> positive regulation of nitrogen compound metabolic process

2 < Back

Launch analysis 3

All parameters that you selected for the current launch will be shown on the screen (1). You can check the genes that were eventually selected for the analysis launch after GO functional classification (if it was applied) by clicking on the name of the input gene list.

If you want to apply any changes to the specified parameters of the launch, click on **Back** (2), otherwise you are ready to start your analysis by clicking on the **Launch analysis** button (3).

Viewing the results

Operating in the *Launch history*

Once your analysis was launched, you will be redirected to the *Launch history* section, also accessible by the direct link at the left menu panel. The Launch history allows you to view the results of all your previous analysis runs and to follow the progress of the currently running analyses.

Start page

Show/hide menu

Recent launches

Select the analysis launch you want to view or start a new analysis

Date	Name	Parameters	Status	Results	Terminate/Delete
2021.08.28 12:00:09	demo-gene-list 1	1 Gene set analysis View input parameters	Running: 8% 2 In progress...		3
2021.08.27 22:10:35	My launch 3	Single gene analysis View input parameters	Completed	View results	
2021.08.27 21:59:00	My launch 3	Single gene analysis View input parameters	Completed	4 View results	
2021.08.26 13:16:13	demo-gene-list	Gene set analysis View input parameters	Completed	View results	5
2021.08.26 13:13:09	My launch 1	Single gene analysis View input parameters	Completed	View results	

[Start new analysis](#) **6**

You can check the parameters used for the launch by clicking on the **View input parameters** link (1). The following pop-up form will appear:

Launch parameters

demo-gene-list 1 2021.08.28 12:00:09

Input gene list:

demo-gene-list optimized by GO categories

Tissue selected during the analysis launch:

none

Promoter range:

[-1000,+100]

GO categories selected to narrow the input gene list:

show

ok

The gene set used for the current run can be viewed by clicking on the respective link in the *Input gene list* field. In case you have selected to optimize your input gene list by certain GO categories, you can view them by clicking on the *show* link next to the *Categories selected to narrow the input gene list field*. If no GO functional categories optimization was done, the respective field will display *none*.

The progress of the currently running analysis will be displayed in the **Status** column (2). Please note that the progress is displayed in percentage of the finished steps of the underlying workflow and it has no direct correlation with the time left for the analysis to finish. Commonly one analysis run will take several hours, but this time interval is highly dependent on the input gene set and other parameters selected for the respective analysis launch.

It is not recommended to have several analysis launches running in parallel. Analysis runs will finish faster when launched consecutively one after another.

If you wish to terminate the launched analysis, you can click on the **Terminate** button (3).

To view the results of a finished analysis please click on the **View results** button (4).

You can manage the stored analysis results and delete the unnecessary data by clicking on the **Delete** icon (5). This action will delete all results of the respective analysis launch. The input gene list used for this analysis launch will still be accessible in the available list of inputs, which is viewed at the very first step of the start analysis wizard (see the *Selecting the input gene list* section of this document). By default, your MATCH Suite account is equipped with 2 GB disk space for storing your analysis results and input gene lists used. You can extend this volume by contacting us via info@genexplain.com with a respective request.

You can launch a new analysis run directly from the *Launch history* section by clicking on the **Start new analysis** (6) button.

Results visualization

Having selected in the *Launch history* section the analysis result which you want to view, it will open in the results visualization mode on your screen.

In the results visualization mode the screen will be divided into four different segments with the following functions:

- (1) The **identified transcription factors** regulating your gene set and respective **matrices tables**
- (2) The **table of your input genes** and found site hits in their promoters
- (3) The **info box** displaying the information about the currently selected object
- (4) The **genome browser**, allowing to visualize the tracks of found sites and additional annotation tracks for further results interpretation

The screenshot displays the MATCH SUITE web application interface. The top navigation bar includes a 'User guide' link and a 'Download tables' button (labeled 5). The left sidebar contains a 'Show/hide menu' button (labeled 2) and a 'Report' button (labeled 6). The main content area is divided into two panels: 'demo-gene-list 3: factor/matrix view' and 'demo-gene-list 3: gene view'. The 'factor/matrix view' panel shows a table of factors with columns for 'Factor name', 'Enrichment analysis', 'Combinatorial analysis', 'Average factor expression across all tissues', and 'Expression specificity (rank of average)'. The 'gene view' panel shows a table of genes with columns for 'Ensembl ID', 'Gene symbol', 'Gene description', 'CMA Score', 'Total number of sites', and 'VSAIOLOS_03'. The 'Genome browser' section (labeled 4) at the bottom shows a genomic track with various annotations.

The Factor, Matrix and Gene tables that you see on the screen can be exported with the applied filters (see filtering instructions below) using the **Download tables button** (5). The archive containing these three tables in tab-separated text format will be downloaded to your local computer.

You can open the comprehensive analysis report about the respective run by clicking on the **Report** (6) link in the left menu panel. The self-explaining report will contain the factor, matrix and gene tables you see in the results visualization section along with supplementary tables and analysis steps description.

You can extend or narrow certain segments of the screen by moving the splitter lines. At this point you might want to use the *Show/hide menu* button, which will hide the left panel menu from view and will broaden the results area of the screen.

This screenshot shows the same MATCH SUITE interface as the previous one, but with red arrows indicating the movement of the splitter lines. The arrows show how the width of the 'factor/matrix view' and 'gene view' panels can be adjusted by dragging the vertical lines between them. The 'Show/hide menu' button (labeled 2) is also visible in the top bar.

In the screen segment, which visualizes the tables of predicted factors and respective matrices, you will find three accessible tables: Factor view, Factor view Pro and Matrix view.

Factor view table

In the **Factor view** (1) table you will see a simplified summary of the transcription factors predicted to be regulating your input gene set. By default top 10 factors will be shown, this can be changed in **Show <number> entries** (2) field. The total number of factors identified will be displayed in (3). You can navigate through the predicted factors using the pages in (4). The columns of the *Factor view* table are fully matching the *Table 1* given in the analysis report of the respective run. You can refer to the analysis report for denominations of column names and their contents or use the info hints provided in the results visualization interface as a mouseover message that will appear upon hovering above the ? sign (5) available at multiple places in the interface.

demo-gene-list 1: factor/matrix view

Factor view **1** Factor view Pro Matrix view Best factors on top **7** Remove all filters **?** Apply changes **?**

Show 10 **2** entries **2**

Showing 1 to 10 of 56 entries **3**

4 Previous 1 2 3 4 5 6 Next

Factor name	Enrichment analysis ?	Combinatorial analysis ?	Average factor expression across all tissues ?	Expression specificity (rank of average) ?
DP-2	●	—	12.46	0.10 13/62
c-Ets-1	●	—	14.96	0.12 18/62
MAZ	●	—	34.69	0.03 22/62
FOSB	●	✓	22.99	0.18 17/62
c-Fos	●	✓	63.36	0.10 19/62
ZNF-24	●	—	22.57	0.01 21/62
BTEB2	●	✓	15.23	0.20 21/62
Sox-10	●	✓	10.14	0.24 16/62

The columns of tables have in-built sorting option (6) which allows to sort the values within one column by ascending or descending order or by alphabetical order in case of textual contents. Simply click on the gray arrows for the sorting to be applied. By default, best factors (or matrices) are brought to the top (please see the analysis report and the [Methods](#) document for explanations on the factors

and matrices ranking procedures). If you want to return to the original order of factors, click on the **Best factors on top** button (7).

Factor view Pro table

The **Factor view Pro** table (1) provides a deeper look into the transcription factors identified to be regulating the input gene set. It fully corresponds to the *Table 3* of the analysis report, where you can find the denominations of its column names and their contents description. Respective info is also summarized in the **?** hints available as mouseover messages upon navigating on them, similar to the *Factor view* table.

demo-gene-list 1: factor/matrix view

Factor view **1** Factor view Pro Matrix view Best factors on top **3** Remove all filters **4** Apply changes **4**

Tissue **2** Select a tissue

Show 10 entries First Previous 1 2 3 4 5 6 Next Last

Showing 1 to 10 of 56 entries

Factor name	Gene symbol	TF classification	Site model	Factor enrichment	Average factor expression across all tissues	Expression specificity (rank of average)
DP-2	TFDP2	Fork head / winged helix factors 3.3.2.2.2	V\$E2F6_03	2.83	12.46	0.10 13/62
c-Ets-1	ETS1	Tryptophan cluster factors 3.5.2.1.1	V\$GCM1ERG_01	2.85	14.96	0.12 18/62
MAZ	MAZ	C2H2 zinc finger factors 2.3.4.8.1	V\$MAZ_Q5	2.69	34.69	0.03 22/62
FOSB	FOSB	Basic leucine zipper factors (bZIP) 1.1.2.1.2	V\$FOS_06	1.31	22.99	0.18 17/62
c-Fos	FOS	Basic leucine zipper factors (bZIP) 1.1.2.1.1	V\$FOS_06	1.31	63.36	0.10 19/62
ZNF-24	ZNF24	C2H2 zinc finger factors 2.3.3.10.1	V\$ZNF24_01	2.37	22.57	0.01 21/62

The *Factor view Pro* table allows you to apply the **Tissue** filter (2). You can select the tissue(s) of your interest from the dropdown list in (2) and click on the **Apply changes** button (4) to recalculate the factors and genes tables, as well as the track of found sites in the genome browser, leaving only those factors and their sites, which are known to be expressed in the tissue(s) of your selection. Respective columns with expression values of factors in the selected tissue(s) will be added to the table. You can refer to the [Methods](#) document for information on the expression values origin. To remove the filtering of factors by selected tissues, click on the **Remove all filters** button (3).

Please note that the tissue expression filter will be also auto applied to all tables and tracks when you will switch from the Factor view Pro table to either of the Factor view or Matrix view tables. To cancel the filters, click on the *Remove all filters* button.

Matrix view table

The **Matrix view** table (1) shows the PWMs (positional weight matrices) of the TRANSFAC® library, the respective sites of which were identified in the promoters of the studied gene set.

Similar to the factors table, the *Best matrices on top* button allows to bring the best matrices to the top of the matrix table in case their order was changed while sorting the values in individual columns. This table fully corresponds to the *Table 4* of the analysis report, where you can find the denominations of its column names and their contents description. Respective info is also summarized in the ? hints available as mouseover messages upon hovering above them, similar to the *Factor view* and *Factor view Pro* tables.

Matrix ID	Matrix logo	Adjusted site enrichment	Site enrichment	Site enrichment FDR	Adjusted sequence enrichment	Sequence enrichment FDR	Composite model
VSPARD_02		2.10	5.31	2.37e-4	2.28	1.22e-3	yes
VSSOX8_04		2.03	3.93	5.68e-10	0.47	3.61e-1	yes
VSKLF7_07		1.91	3.59	8.11e-11	1.29	1.61e-3	yes
VSDEAF1_01		1.59	3.20	6.64e-6	1.32	1.01e-3	yes
VSFOS_06		1.31	2.36	2.23e-8	0.45	6.78e-1	yes
VSRFX1_05		1.17	2.55	3.93e-3	0.76	2.22e-2	yes

The **Adjusted site enrichment filter** (2) allows to leave only matrices with the adjusted site enrichment values higher than the threshold specified by the filter. The set filter will be first applied exclusively to the *Matrix view* table. To recalculate the factors and the genes tables, as well as the track of the found sites in the

genome browser, only with matrices that were left after applying the filter, you should click on the *Apply changes* button.

The **Adjusted sequence enrichment filter** (3) allows to leave only matrices with the adjusted sequence enrichment values higher than the threshold specified by the filter. The set filter will be first applied exclusively to the *Matrix view* table. To recalculate the factors and the genes tables, as well as the track of the found sites in the genome browser, only with matrices that were left after applying the filter, you should click on the *Apply changes* button.

Filters will be also auto applied to all tables and tracks when you will switch from matrix table to either of the factor tables. To cancel the filters, click on the *Remove all filters* button.

Genes table

The genes table presented in the results visualization section fully corresponds to the gene table provided in the *Table 5* of the analysis report, where you can find the denominations of its column names and their contents description. Respective info is also summarized in the **?** hints available as mouseover messages upon navigating on them, similar to the *Factor view*, *Factor view Pro* and *Matrix view* tables.

demo-gene-list 1: gene view

Show 10 entries

First Previous 1 2 3 Next Last

Showing 1 to 10 of 27 entries

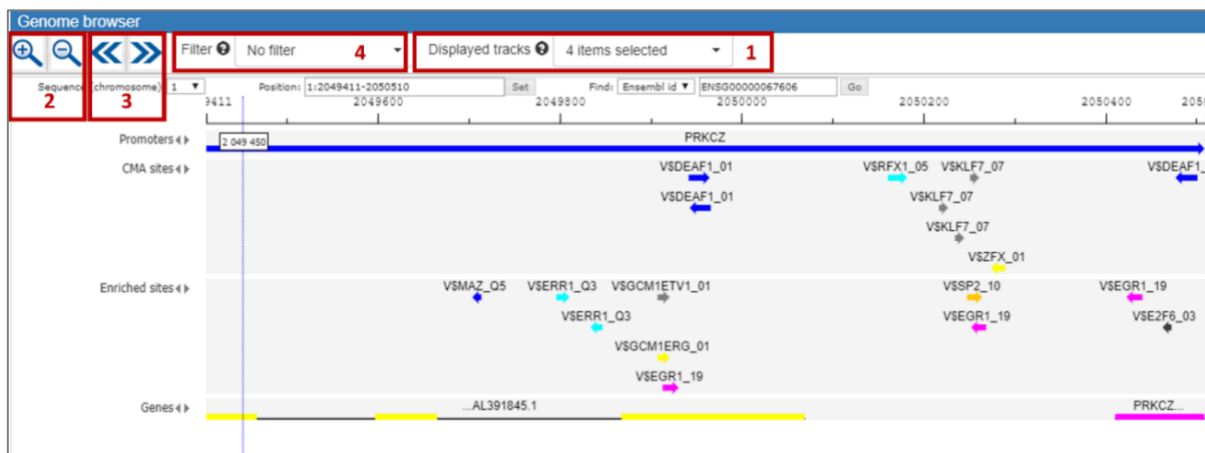
Ensembl ID	Gene symbol	Gene description	CMA Score	Total number of sites	V\$AIOLOS_03	V\$DEAF1_01	V\$E2F6_03	V\$EGR1_19	V\$ERR1_Q3
					Ikaros, ikzf5	DEAF1	DP-2	Egr-1	ERR1
ENSG00000067606	PRKCZ	protein kinase C zeta	5.36	21		4	1	3	2
ENSG00000074964	ARHGEF10L	Rho guanine nucleotide exchange factor 10 like	5.26	19		2	1	7	
ENSG00000100142	POLR2F	RNA polymerase II subunit F	4.98	14	1			2	1
ENSG00000048649	RSF1	remodelling and spacing factor 1	4.96	19	1	2		2	
ENSG00000051523	CYBA	cytochrome b-245 alpha chain	4.90	21		1		4	

To visualize in genome browser the sites found in the promoter of any gene, just click on the line with the gene of your interest inside the genes table and refer to the genome browser below to explore the predicted regulation model of the respective gene.

Genome Browser

The genome browser provides you with visualizations of the predicted gene regulation models for each gene from the input gene set. Having clicked on any gene from the *Genes table*, the promoter model of the respective gene will be automatically opened in the genome browser with visualization of the sites found within the respective promoter. By default the displayed tracks include:

- **Promoters** - the track of all promoters that were used for the performed analysis
- **CMA sites** - the track of sites belonging to the combinatorial matrices of the constructed CMA model (see [Methods](#) for further info)
- **Enriched sites** - the track of the remaining sites found to be enriched in the promoters of the studied genes
- **Genes** - the track with all Ensembl genes



Displayed tracks can be customized using the **Displayed tracks** dropdown list (1)

4 items selected

Select All

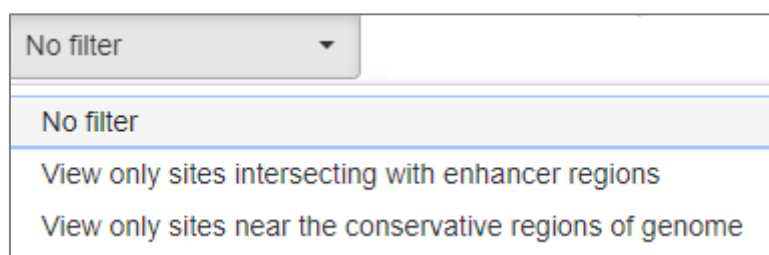
Deselect All

- Promoters ✓
- All found sites
- CMA sites ✓
- Enriched sites ✓
- Genes ✓
- Enhancers
- Variations (SNPs)
- Experimentally validated TFBS
- Conservative regions of genome

You can zoom in and zoom out for a more generalized or a more detailed view of the promoter model using the **zoom in** and **zoom out** buttons (2). In case navigation in genome browser is lost due to multiple zoom clicks, you can always return to the promoter of your interest by selecting the respective gene in the genes table once again.

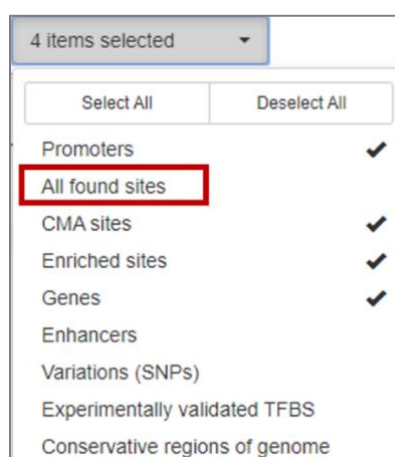
Navigation through the track is supported by the **Shift to the left** and **Shift to the right** buttons (3)

Filtering can be applied to the displayed sites using the **Filter** option (4):



Depending on your selection, only sites intersecting with the conservative regions of the genome or those intersecting with the enhancer regions will be displayed. The genes table will be automatically recalculated leaving only the hits for the sites left after applying the respective filter. To cancel the applied filter, select the *No filter* option in the dropdown list of filters.

Please note that the *CMA sites* track shows only the best sites relevant to the identified CMA combinatorial model. The genes table will contain the full counts for all sites identified, both enriched and combinatorial. To view the respective full track of all found sites, you can add the *All found sites* track to the visualization:



For receiving additional information about any of the found sites, you can click on the site of your interest and explore the contents displayed in the *Info box* section:

Info box
Site ID: 278
Type: TF binding site
Sequence name: 1
Sequence: AGGCGGGAAGT
Position: 17539416 - 17539426 (11)
Properties:

- coreScore: 1.0
- score: 0.991601
- siteModel: V\$E2F6_03

Model V\$E2F6_03
Binding element: E2F-6
Threshold: 0.9538317285805532
Matrix: V\$E2F6_03
Matrix length: 11

Genome browser
Filter: No filter
Displayed tracks: 4 items selected
Sequence (chromosome): 1
Position: 17538698-17539797
Find: Ensembl id: ENSG00000074964
Go: 17539400
Promoters: ARHGEF10L
CMA sites: V\$DEAF1_01
Enriched sites: V\$EGR1_19, V\$GCM1ERG_01, V\$EGR1_19, V\$EGR1_19, V\$E2F6_03
Genes:

A right click on the track name in genome browser will open a context menu allowing you to remove the track from the visualization or open it as a table (function supported only for tracks of predicted sites or promoter track):

CMA sites: V\$DEAF1_01
Enriched sites: V\$EGR1_19, V\$GCM1ERG_01, V\$EGR1_19, V\$EGR1_19, V\$E2F6_03
Genes:

Open table
Remove from view

Tracks opened as table can be exported using the *Export* button at the top menu panel or used for further analysis from the geneXplain platform perspective.

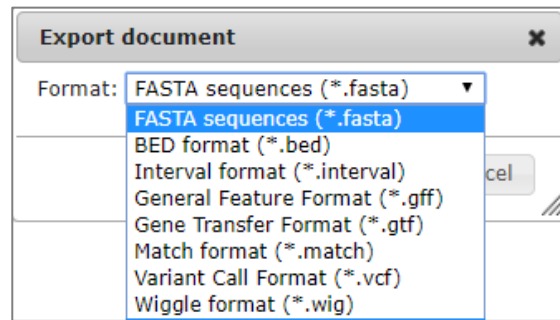
User guide

Start page
Enriched sites X

First Previous Page 1 of 9 Next Last Showing 1 to 50 of 428 entries Show 50 entries

ID	Sequence (chromosome) name	From	To	Length	Strand	Type	Property: coreScore	Property: score	Property: siteModel
1	16	1307882	1307892	11	+	TF binding site	1	0.95586	V\$E2F6_03
2	16	1307882	1307890	9	+	TF binding site	1	0.99915	V\$AIOLQ5_03
3	16	1307884	1307896	13	-	TF binding site	1	0.79395	V\$ZNF24_01
4	16	1307888	1307900	13	-	TF binding site	0.80889	0.80562	V\$ZNF24_01
5	16	1307889	1307906	18	+	TF binding site	0.92998	0.89941	V\$EGR1_19
6	16	1307894	1307908	15	+	TF binding site	0.92684	0.86619	V\$GCM1ERG_01
7	16	1307898	1307908	11	-	TF binding site	0.95972	0.97675	V\$MAZ_Q5
8	16	1307902	1307910	9	+	TF binding site	1	1	V\$AIOLQ5_03
9	16	1307904	1307916	13	-	TF binding site	1	0.80844	V\$ZNF24_01
10	16	1307908	1307920	13	-	TF binding site	1	0.9057	V\$ZNF24_01
11	16	1307913	1307930	18	+	TF binding site	0.92998	0.90011	V\$EGR1_19
12	16	1307914	1307928	15	+	TF binding site	0.92684	0.86466	V\$GCM1ERG_01
13	16	1307914	1307924	11	-	TF binding site	1	0.98093	V\$MAZ_Q5

The list of available export formats is as follows:



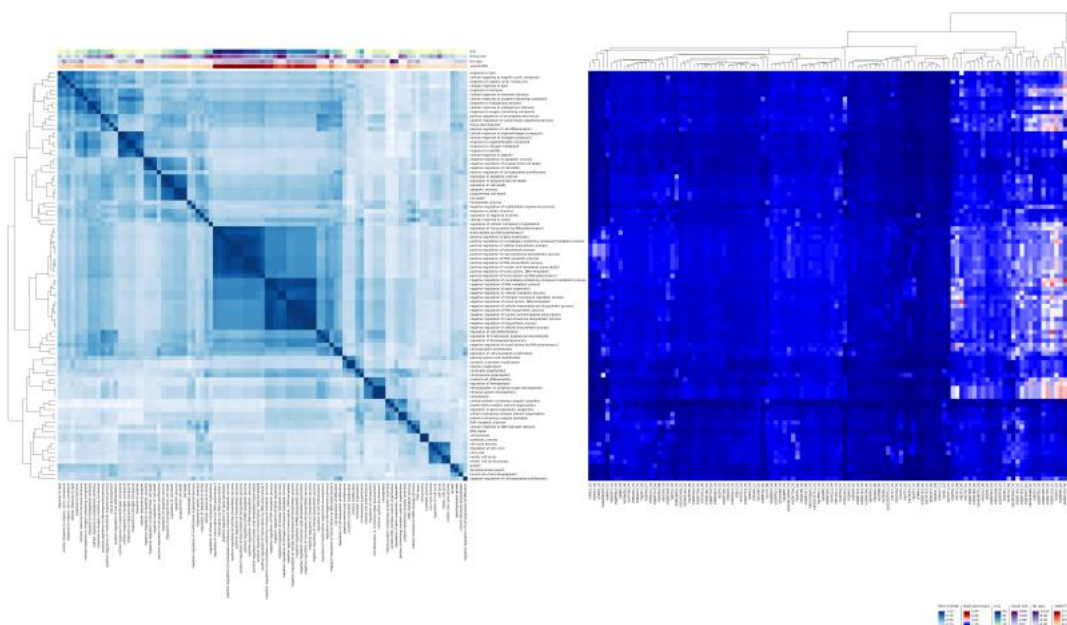
Heatmaps

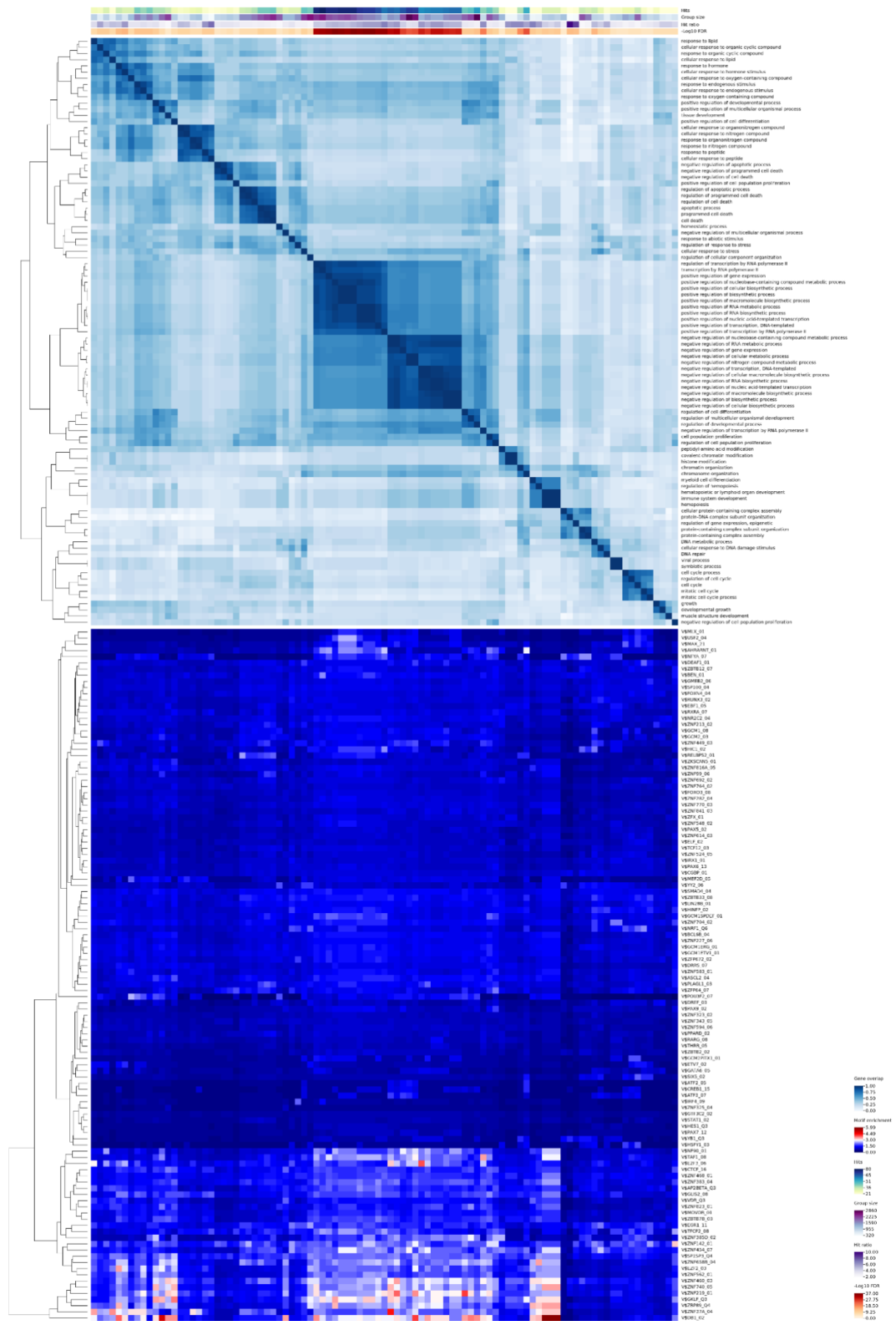
If the functional classification of your input gene set by GO terms produced reliable results (by p-values and by the number of genes in each functional category), heatmaps for the respective GO categorization will be included in the optional section *Functional Analysis of Gene Regulation* of the analysis report.

The constructed image will contain two parts:

- (1) The heatmap of GO to GO terms mapping for the GO terms overrepresented among the studied gene set;
- (2) The heatmap visualizing how enriched motifs are associated with the respective GO categories.

The visualization examples of such heatmaps is given below. Depending on your preference, you can use either of the heatmaps: the one with the horizontal layout (provided in the analysis report as a clickable image), or the one with the vertical layout (provided under the link to alternative view in the report text):





Single gene analysis

Launching the analysis

Specifying the input gene

Specifying the input gene in the TRANSFAC® database

A particularly convenient way to launch the MATCH Suite single gene analysis is to select the gene of your interest from the search results in the TRANSFAC® database. For instance, you can select any human gene in the results of your search in TRANSFAC® database and launch the MATCH Suite single gene analysis on a respective gene:

The screenshot shows the TRANSFAC database search interface. At the top, there's a navigation bar with 'geneXplain', 'Welcome to TRANSFAC® + PROTEOME™', 'Logout', and 'My Data'. Below this is a search bar with 'brc' entered and a 'search' button. To the right, there are 'example searches' for gene, miRNA, disease, pathway, drug, transcription factor, and matrix, with corresponding results like BRCA1, miR-155, ovarian cancer, JNK pathway, Simvastatin, FOXA1, and STAT. Below the search bar, it says 'Found 280 results for "brc" in 6 categories:'. A table shows search results: 114 Genes and proteins, 0 miRNAs, 5 Pathways, 16 Transcription factors, 15 Matrices, 0 Variants, 116 Diseases, and 14 Drugs. To the right of this table is a 'Still searching?' section with links for 'More search options', 'Ontology search', and 'Custom search'. Below the search results, there's a section titled 'Genes and proteins' with '114 of 114 total'. It includes a 'Select results and forward them to one of the tools (Guided Tour):' section with icons for Pathfinder, Ontology, Match, Match Suite (highlighted with a red box), Fasta, Profiles, and Binding factors for gene. Below this is a table of search results. The table has columns: #, Name, Type, Species/Taxon, and Description. The first row is 'br' (Drosophila melanogaster, broad). The second row is 'Arid4b' (Mus musculus, Protein with high similarity to human ARID4B...). The third row is 'brn' (Drosophila melanogaster, brainiac). The fourth row is 'BRCA1' (Homo sapiens, Breast cancer 1 early onset, a transcription regulator that acts in double-strand break repair, apoptosis, protein folding, and mRNA cleavage, upregulated in breast, ovary, and prostatic neoplasms; mRNA is downregulated in melanoma) - this row is highlighted with a red box. The fifth row is 'BRCC3' (Homo sapiens, BRCA1-BRCA2 containing complex subunit 3, may be involved in regulation of transcription; gene mutations are associated with moyamoya syndrome, hemophilia A, and T-cell leukemia associated with ataxia telangiectasia).

Once you have clicked on the Match Suite icon, indicate that you aim to perform human single gene analysis:

MATCH Suite options

?

Select species:

Human

?

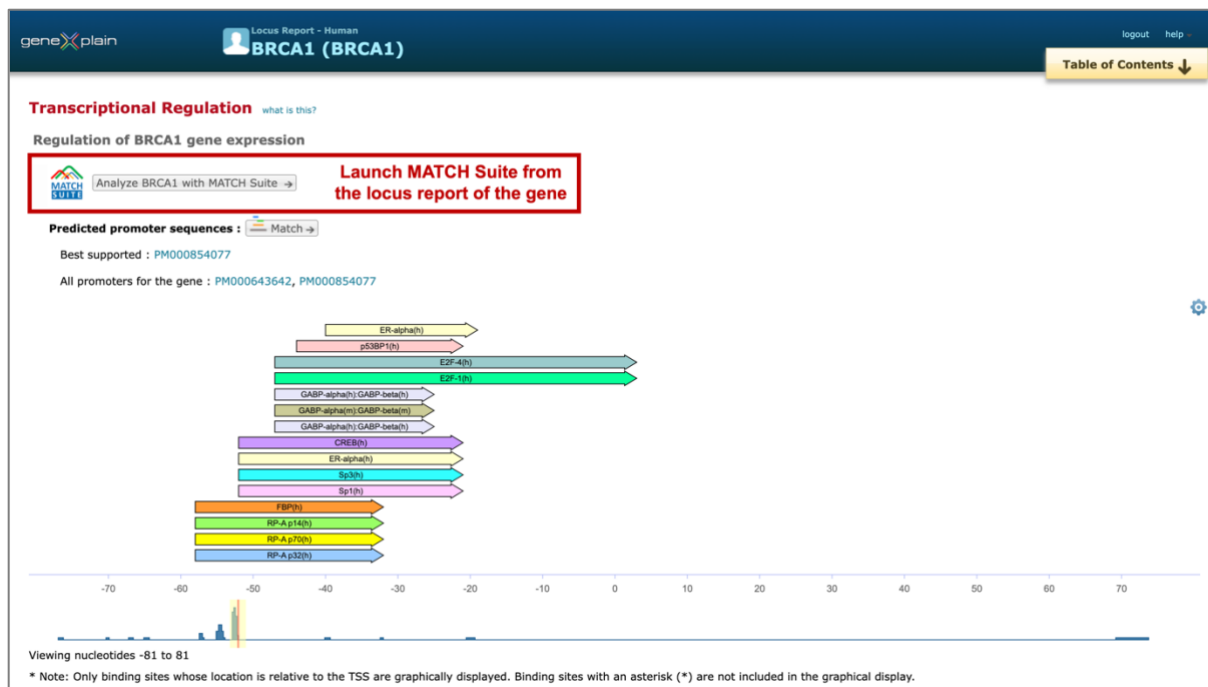
Select analysis type:

Single gene

Submit

And click on Submit to proceed to the next step of the wizard.

Alternatively, you can start the MATCH Suite single gene analysis from a locus report of the gene of your interest:



When you have selected the gene of your interest, press the button “Match Suite single gene” in the search results or click on the “Analyze <gene name> with MATCH Suite” button in the *Transcriptional regulation* section of the locus report of the selected gene, and you will be automatically transferred to the MATCH Suite tool. The wizard will automatically save and process the gene you have selected and in several seconds you will be taken to the *Specifying the launch parameters* step of the MATCH Suite single gene analysis wizard.

Specifying the input gene from the MATCH Suite interface

When you click on the *Start analysis* section, the system immediately navigates you to the analysis launch wizard. At the very first step of this wizard, you will be asked to select the species you are working with:

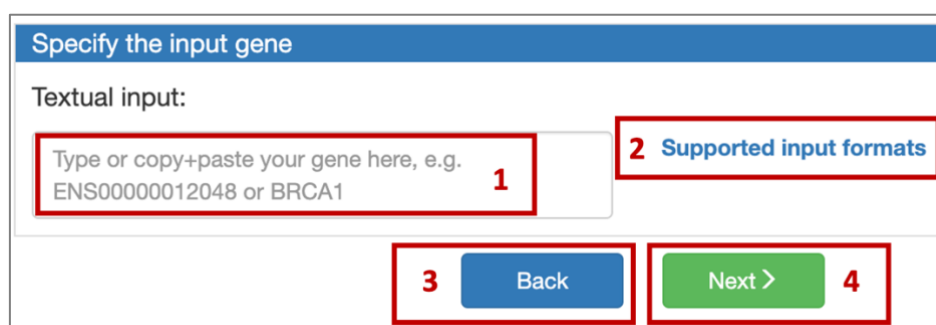
The screenshot shows a 'Select species' dialog box. It has a blue header with the text 'Select species'. Below the header is a dropdown menu with 'Human' selected. At the bottom of the dialog is a blue button with the text 'Next >'.

And type of the analysis you would like to perform:



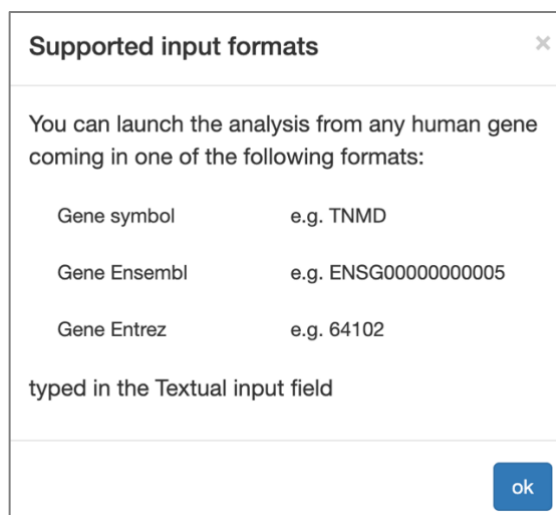
A dialog box titled "Select analysis" with a blue header. It contains two radio button options: "Single gene analysis" (selected) and "Gene set analysis". Below the options is a blue button labeled "Next >".

If you will select the *Single gene analysis* option, you will be taken to the next step of the wizard, where you will be asked to specify the gene you want to use for the analysis:



A dialog box titled "Specify the input gene" with a blue header. It contains a "Textual input:" label and a text input field. The input field has placeholder text: "Type or copy+paste your gene here, e.g. ENS00000012048 or BRCA1". To the right of the input field is a box labeled "2 Supported input formats". Below the input field are two buttons: a blue "Back" button (labeled 3) and a green "Next >" button (labeled 4).

You can launch the analysis by typing any gene symbol, Entrez gene ID or Ensembl gene ID into the **textual input field** (1). The **Supported input formats** (2) will assist you in providing the input gene in the correct format:



A dialog box titled "Supported input formats" with a close button (X) in the top right corner. It contains the text: "You can launch the analysis from any human gene coming in one of the following formats:". Below this is a table with three rows:

Gene symbol	e.g. TNMD
Gene Ensembl	e.g. ENSG000000000005
Gene Entrez	e.g. 64102

typed in the Textual input field

At the bottom right is a blue button labeled "ok".

You can return to the MATCH Suite analysis type selection using the **Back** (3) button or click on **Next** (4) to proceed with specifying the launch parameters of the MATCH Suite single gene analysis.

Specifying the launch parameters

In the next step of the analysis launch wizard the MATCH Suite will ask you to fill in the following form:

Specify parameters

Name of the analysis launch: 1

Studied tissue: 2

3

Tissue selection will allow to use tissue-specific promoters from the [FANTOM5](#) database (where available) and identify the transcription factors regulating your gene that are known to be expressed in the selected tissue. Tissue selection will also provide consideration of tissue-specific enhancers/silencers in case any are known to be active for input gene in the selected tissue. If tissue will not be selected, all known enhancers of input gene will be considered in the analysis.

Promoter range:

☒ Use default promoter range [-500,+100] 4

☐ Customized promoter range

From to

Promoter type:

☒ best supported promoter 5

☐ most 5 prime

☐ most 3 prime

Profile optimization:

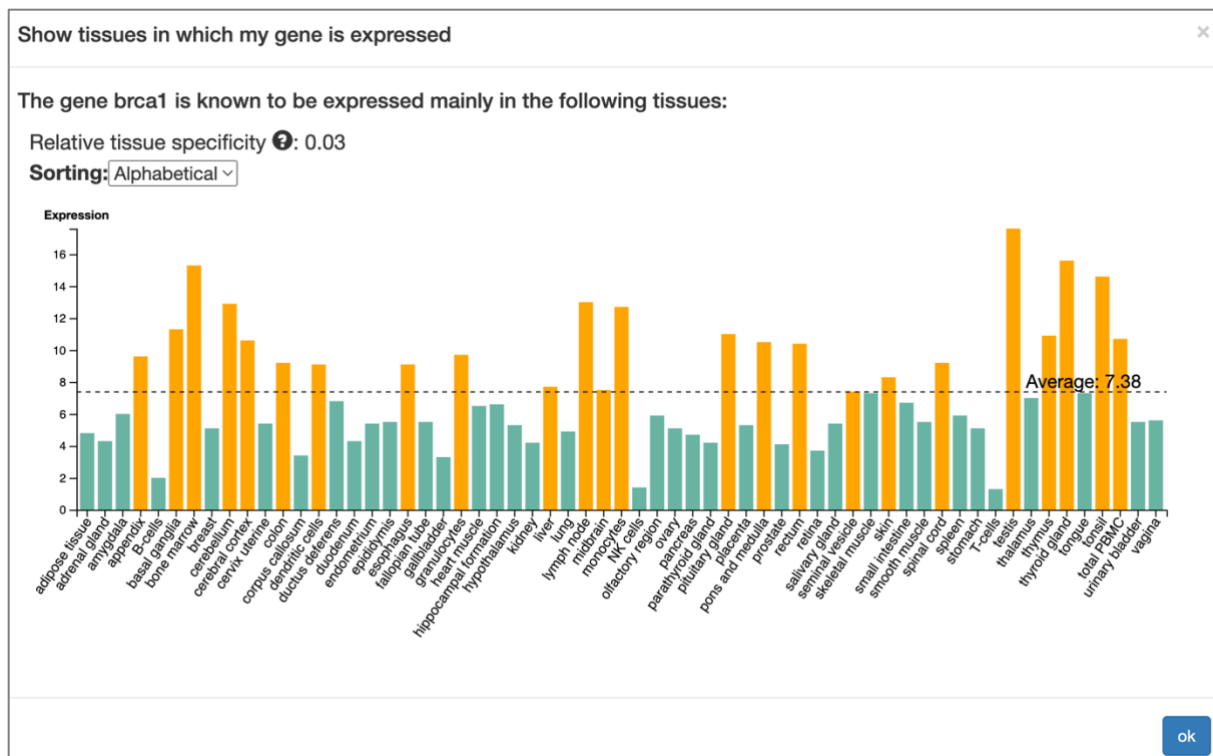
Narrow transcription factors search to certain GO categories

6

You can give a name to your analysis launch or keep the default name suggested by the system based on the gene you have submitted for the analysis in the **Name of the analysis launch** field (1). This name will be further used in the launches history for easy selection of the run of your interest.

Optionally you can specify the tissue of your interest from the dropdown list of supported tissues in the **Studied tissue** field (2). The provided tissues are the tissues from [FANTOM5](#) database with specific coordinates for the transcription start sites (TSSs) that allow to select tissue-specific promoters for the subsequent analysis. Tissue selection will also provide consideration of tissue-specific enhancers/silencers in case any are known to be active for input gene in the selected tissue. If tissue will not be selected, all known enhancers of input gene will be considered in the analysis. The selected tissue is also used further for identification of transcription factors that are known to be expressed in the tissue of your interest. Please refer to the [Methods](#) document for further details.

You can view the tissues in which the input gene is known to be mainly expressed by clicking on the **Show tissues in which my gene is expressed** (3) button:



The promoter range that will be used for the search of TFBS in your analysis run is specified by the **Promoter range** parameter (4). By default the promoter range used by the MATCH Suite is [-500,100] relative to the TSS. You can specify a customized promoter range within the interval of -2500 and +1000 from TSS. The total length of the promoter range should not exceed 2500 bp. The **Promoter type** (5) parameter allows you to specify which type of the promoter should be used in the analysis: best supported / most 5 prime / most 3 prime.

The search for transcription factors that are regulating the input gene can be narrowed to transcription factors that are encoded by genes belonging to the certain Gene Ontology (GO) categories. For this you can specify the GO categories of your interest from the provided dropdown list in the **Profile optimization** (6) field.

When done with parameters selection, click on **Next** (8) to proceed to the analysis launch confirmation or click on **Back** (7) to return to the input gene specification.

Confirming the launch parameters and starting the analysis

After specifying all parameters for your launch, the MATCH Suite wizard will ask you to confirm your selection:

Selected launch parameters

Launch name:

brca1 demo run

Input:

brca1

Selected tissue:

Nothing selected

Promoter range:

[-500;+100]

Promoter type:

best supported promoter

Selected GO categories:

- GO:0000003 reproduction
- GO:0007049 cell cycle

2

< Back

Launch analysis

3

All parameters that you selected for the current launch will be shown on the screen (1). If you want to apply any changes to the specified parameters of the launch, click on **Back** (2), otherwise you are ready to start your analysis by clicking on the **Launch analysis** button (3).

Viewing the results

Operating in the Launch history

Once your analysis was launched, you will be redirected to the *Launch history* section, also accessible by the direct link at the left menu panel. The Launch history allows you to view the results of all your previous analysis runs and to follow the progress of the currently running analyses.

Start page

Getting started

Start analysis

Launch history

Show/hide menu

Recent launches

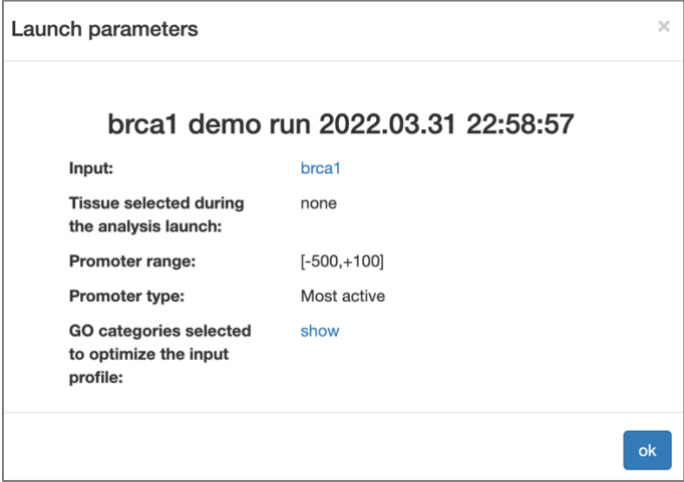
Select the analysis launch you want to view or start a new analysis

Date	Name	Parameters	Status	Results	Terminate/Delete
2021.08.28 12:00:09	brca1 demo run	1 Single gene analysis View input parameters	Running: 8% 2 In progress...		3
2021.08.27 22:10:35	My launch 3	Single gene analysis View input parameters	Completed	View results	
2021.08.27 21:59:00	My launch 3	Single gene analysis View input parameters	Completed	4 View results	
2021.08.26 13:16:13	demo-gene-list	Gene set analysis View input parameters	Completed	View results	5
2021.08.26 13:13:09	My launch 1	Single gene analysis View input parameters	Completed	View results	

Start new analysis

6

You can check the parameters used for the launch by clicking on the **View input parameters** link (1). The following pop-up form will appear:



A screenshot of a web-based pop-up form titled "Launch parameters" with a close button (X) in the top right corner. The form displays the following information:

- brca1 demo run 2022.03.31 22:58:57**
- Input:** [brca1](#)
- Tissue selected during the analysis launch:** none
- Promoter range:** [-500,+100]
- Promoter type:** Most active
- GO categories selected to optimize the input profile:** [show](#)

An "ok" button is located at the bottom right of the form.

In case you have selected to optimize the profile used for the site search by certain GO categories, you can view them by clicking on the *show* link next to the *GO categories selected to optimize the input profile*. If no GO categories were selected, respective field will display *none*.

The progress of the currently running analysis will be displayed in the **Status** column (2). Please note that the progress is displayed in percentage of the finished steps of the underlying workflow and it has no direct correlation with the time left for the analysis to finish. Commonly one analysis run will take several hours, but this time interval is highly dependent on the input gene set and other parameters selected for the respective analysis launch.

It is not recommended to have several analysis launches running in parallel. Analysis runs will finish faster when launched consecutively one after another.

If you wish to terminate the launched analysis, you can click on the **Terminate** button (3).

To view the results of a finished analysis please click on the **View results** button (4).

You can manage the stored analysis results and delete the unnecessary data by clicking on the **Delete** icon (5). This action will delete all results of the respective analysis launch. By default, your MATCH Suite account is equipped with 2 GB disk space for storing your analysis results and inputs. You can extend this volume by contacting us via info@genexplain.com with a respective request.

You can launch a new analysis run directly from the *Launch history* section by clicking on the **Start new analysis** (6) button.

Results visualization

Having selected in the *Launch history* section the analysis result which you want to view, it will open in the results visualization mode on your screen.

In the results visualization mode the screen will be divided into four different segments with the following functions:

- (1) The **identified transcription factors** regulating your gene and respective **matrices tables**
- (2) The **table of analyzed regulatory regions of the input gene** and site hits found within them
- (3) The **info box** displaying the information about the currently selected object
- (4) The **genome browser**, allowing to visualize the tracks of found sites and additional annotation tracks for further results interpretation

The screenshot displays the MATCH Suite results visualization interface. The interface is divided into four main sections:

- Factor view (1):** A table showing identified transcription factors. Columns include Factor name, Gene symbol, TF classification, Site model, -log10(affinity p-value), Average factor expression, and Expression specificity. The table lists factors like SOX-2, ARNT2, MondoB, and Mad1.
- Regulatory regions view (2):** A table showing analyzed regulatory regions. Columns include Type, Accession, Coordinates, Total number of sites, and a list of transcription factors. The table lists regions like 17:43170146-43170745, 17:43321712-43324211, etc.
- Info box (3):** A section displaying track found sites and sequence collection information. It shows the sequence collection: databases/EnsemblHuman104/Sequences/chromosomes GRCh38 and Site count: 689.
- Genome browser (4):** A genomic track visualization showing found sites and additional annotation tracks. It includes a chromosome browser with a position slider and a list of found sites.

The Factor, Matrix and Regulatory regions tables that you see on the screen can be exported with the applied filters (see filtering instructions below) using the **Download tables button** (5). The archive containing these three tables in tab-separated text format will be downloaded to your local computer.

You can open the comprehensive analysis report about the respective run by clicking on the **Report** (6) link in the left menu panel. The self-explaining report will contain the factor, matrix and regulatory regions tables you see in the results visualization section along with supplementary tables and analysis steps description.

You can extend or narrow certain segments of the screen by moving the splitter lines. At this point you might want to use the *Show/hide menu* button, which will hide the left panel menu from view and will broaden the results area of the screen.

The screenshot displays the MATCH Suite interface. On the left, a sidebar contains navigation links: 'Start page', 'Getting started', 'Start analysis', 'Launch history', 'View results', 'Visualization', and 'Report'. The 'Report' link is highlighted. The main area is split into two panels. The left panel, titled 'brca1 demo run: factor/matrix view', shows a table of transcription factors with columns for Factor name, Gene symbol, TF classification, Site model, -log(affinity p-value), Average factor expression across all tissues, and Expression specificity (rank of average). The right panel, titled 'brca1 demo run: regulatory regions view', shows a table of regulatory regions with columns for Type, Accession, Coordinates, Total number of sites, and a list of transcription factors (VSGKLF_Q4, VSIK_Q5_01). A 'Show/hide menu' button is located at the top left of the main area. Red arrows indicate the splitter lines used to adjust the width of the panels.

In the screen segment, which visualizes the tables of predicted factors and respective matrices, you will find two accessible tables: Factor view and Matrix view.

Factor view table

The **Factor view** table (1) provides the information on transcription factors identified to be regulating the input gene in the specified conditions. By default top 10 factors will be shown, this can be changed in **Show <number> entries** (2) field. The total number of factors identified will be displayed in (3). For single gene analysis in MATCH Suite this number will be always equal to 40, see [Methods](#) document for details. You can navigate through the predicted factors using the pages in (4). The columns of the *Factor view* table are fully matching the *Table 1* given in the analysis report of the respective run. You can refer to the analysis

report for denominations of column names and their contents or use the info hints provided in the results visualization interface as a mouseover message that will appear upon hovering above the ? sign (5) available at multiple places in the interface.

brca1 demo run: factor/matrix view

Factor view **1** Matrix view Best factors on top **7** Remove all filters **8** Apply changes **?**

Show 10 entries **2** First **4** Previous 1 2 3 4 Next Last

Showing 1 to 10 of 40 entries **3**

Factor name ?	Gene symbol ?	TF classification ?	Site model 5 ?	-log(affinity p-value) ?	Average factor expression across all tissues 6 ?	Expression specificity (rank of average) ?
SOX-2	SOX2	High-mobility group (HMG) domain factors 4.1.1.2.2.2	V\$SOX2_Q6	4.04	7.24	0.22 15/62
ARNT2	ARNT2	Basic helix-loop-helix factors (bHLH) 1.2.5.2.2	V\$ARNTL_Q1	5.57	7.16	0.19 17/62
MondoB	MLXIPL	Basic helix-loop-helix factors (bHLH) 1.2.6.6.3	V\$CHREBP_Q6	3.14	6.82	0.27 13/62
Mad1	MXD1	Basic helix-loop-helix factors (bHLH) 1.2.6.7.1	V\$MYCMAX_Q3	2.86	10.89	0.17 17/62
KLF15	KLF15	C2H2 zinc finger factors 2.3.1.2.15	V\$GKLF_Q4	18.00	11.82	0.07 22/62

The columns of tables have in-built sorting option (6) which allows to sort the values within one column by ascending or descending order or by alphabetical order in case of textual contents. Simply click on the gray arrows for the sorting to be applied. By default, best factors (or matrices) are brought to the top (please see the analysis report and the [Methods](#) document for explanations on the factors and matrices ranking procedures). If you want to return to the original order of factors, click on the **Best factors on top** button (7). If filters were applied using the matrix table or the genome browser visualization, you can cancel the applied filtering by clicking on the **Remove all filters** button (8).

Matrix view table

The **Matrix view** table (1) shows the PWMs (positional weight matrices) of the TRANSFAC® library, the respective sites of which were identified in the studied regulatory regions of the input gene.

Similar to the factors table, the *Best matrices on top* button allows to bring the best matrices to the top of the matrix table in case their order was changed while sorting the values in individual columns. This table fully corresponds to the *Table 2* of the analysis report, where you can find the denominations of its column names and their contents description. Respective info is also summarized in the **?** hints available as mouseover messages upon hovering above them, similar to the *Factor view* table.

brca1 demo run: factor/matrix view

Factor view **1** Matrix view Best matrices on top **?** Remove all filters **?** Apply changes **?** **4**

-log(affinity p-value) **? 2**

Site score **? 3**

Show 10 entries First Previous 1 2 3 Next Last

Showing 1 to 10 of 24 entries

Matrix ID ?	Matrix logo ?	Number of sites ?	Affinity score ?	-log(affinity p-value) ?	Site score ?
V\$GKLF_Q4		11	5.92	18.00	1.00
V\$IK_Q5_01		32	5.79	11.54	1.00
V\$AP2GAMMA_Q5		1	4.23	9.09	1.00
V\$PATZ1_04		23	13.94	7.19	0.98
V\$BCL6_01		18	6.22	6.87	0.97
V\$TAF1_07		53	3.95	6.85	1.00

The **-log(affinity p-value) filter (2)** allows to leave only matrices with -log(affinity p-value) values higher than the threshold specified by the filter. The set filter will be first applied exclusively to the *Matrix view* table. To recalculate the factors and the regulatory regions tables, as well as the track of the found sites in the genome browser, only with matrices that were left after applying the filter, you should click on the *Apply changes* button.

The **Site score filter (3)** allows to leave only matrices with site score values higher than the threshold specified by the filter. The set filter will be first applied exclusively to the *Matrix view* table. To recalculate the factors and the regulatory regions tables, as well as the track of the found sites in the genome browser, only with matrices that were left after applying the filter, you should click on the **Apply changes (4)** button.

Filters will be also auto applied to all tables and tracks when you will switch from matrix table to the factor table or vice versa. To cancel the filters, click on the *Remove all filters* button.

Regulatory regions table

The regulatory regions table presented in the results visualization section fully corresponds to the regulatory regions table provided in the *Table 3* of the analysis report, where you can find the denominations of its column names and their contents description. Respective info is also summarized in the **?** hints available as mouseover messages upon navigating on them, similar to the *Factor view* and *Matrix view* tables.

brca1 demo run: regulatory regions view										
Show 10 entries				FirstPrevious 1 NextLast						
Showing 1 to 7 of 7 entries										
	Type	Accession	Coordinates	Total number of sites	V\$GKLF_Q4	V\$IK_Q5_01	V\$AP2GAMMA_Q5	V\$PATZ1_04	V\$BCL6_01	V\$TAF1_07
					KLF13, KLF15, KLF5 more...	IKZF1, IKZF3, IKZF5	AP-2alpha	MAZ, VEZF1	BCL-6	LEF-1, TCF-7
1	promoter	BRCA1	17:43170146-43170745	37	1	1			2	2
2	enhancer	EN000176039	17:43321712-43324211	182	2	7		3	7	14
3	enhancer	EN000176018	17:43314502-43317001	171	2	8		13	3	18
4	enhancer	EN000176016	17:43318322-43320821	149		15		5	1	14
5	enhancer	EN000176063	17:43328836-43331085	73	1	1		1	4	3
6	enhancer	EN000176064	17:43333120-43334602	59	2		1	1	1	1
7	enhancer	EN000024071	17:42886296-42886595	18	3					1

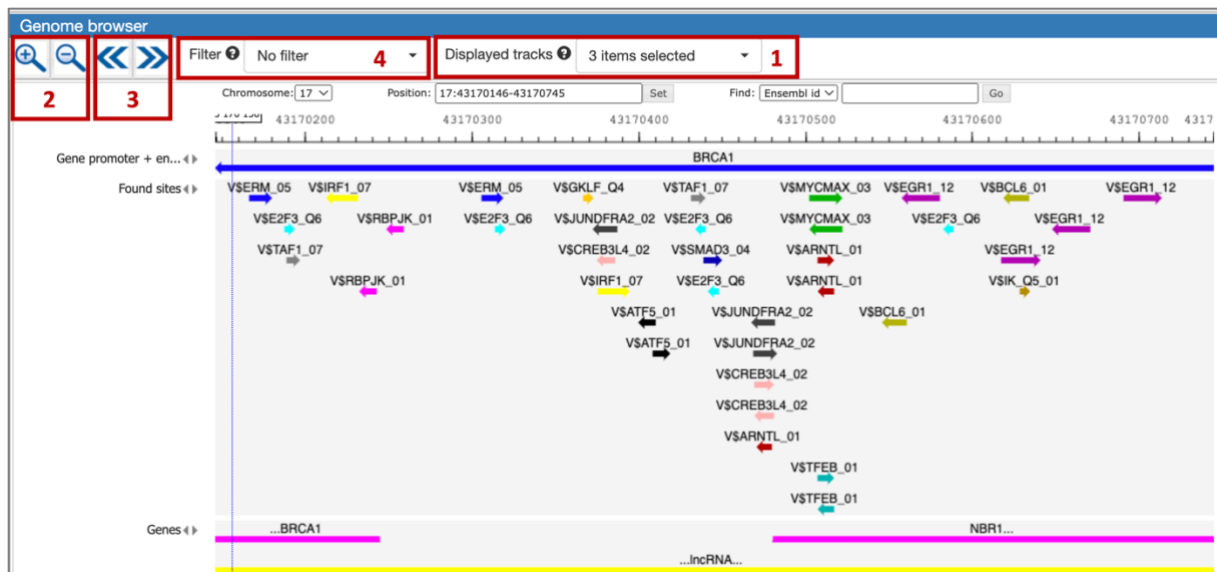
To visualize in genome browser the sites found in the certain regulatory region (promoter / enhancer / silencer), just click on the line with the regulatory region of your interest inside the regulatory regions table and refer to the genome browser below to explore the predicted regulation model of the respective region.

Genome Browser

The genome browser provides you with visualizations of the predicted regulation models for each of the studied regulatory regions of the input. Having clicked on any regulatory region from the *Regulatory regions table*, the model of the

respective regulatory regions will be automatically opened in the genome browser with visualization of the sites found within the respective regions. By default the displayed tracks include:

- **Gene promoter + enhancers** - the track of all regulatory regions that were used for the performed analysis
- **Found sites** - the track of all sites found to have cumulative binding affinity to the studied regulatory regions (see [Methods](#) for further info)
- **Genes** - the track with all Ensembl genes



Displayed tracks can be customized using the **Displayed tracks** dropdown list (1)

The screenshot shows the 'Displayed tracks' dropdown menu. It has a header '3 items selected' and two buttons: 'Select All' and 'Deselect All'. Below these are seven tracks with checkboxes:

Track	Selected
Found sites	✓
Gene promoter + enhancers	✓
Genes	✓
All enhancers	
Variations (SNPs)	
Experimentally validated TFBS	
Conservative regions of genome	

You can zoom in and zoom out for a more generalized or a more detailed view of the regulatory region model using the **zoom in** and **zoom out** buttons (2). In case navigation in genome browser is lost due to multiple zoom clicks, you can always

return to the regulatory region of your interest by selecting the respective regulatory regions entry in the table of regulatory regions.

Navigation through the track is supported by the **Shift to the left** and **Shift to the right** buttons (3)

Filtering can be applied to the displayed sites using the **Filter** option (4):

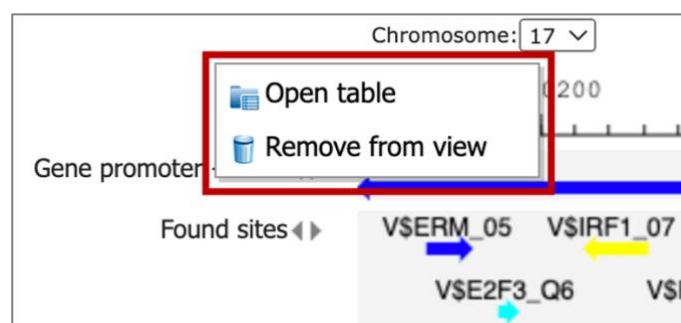


Depending on your selection, only sites intersecting with the conservative regions of the genome can be displayed. The regulatory regions table will be automatically recalculated leaving only the hits for the sites left after applying the respective filter. To cancel the applied filter, select the *No filter* option in the dropdown list of filters.

For receiving additional information about any of the found sites, you can click on the site of your interest and explore the contents displayed in the *Info box* section:

The image shows the TRANSFAC 2.0 interface. On the left is the 'Info box' for site ID 681, which is a TF binding site. It displays the sequence 'ctcagggtgc', position '43170508 - 43170517 (10)', and properties including a core score of 0.807483 and a site model 'VSARNTL_01'. The binding element is shown as 'GT_CACGTGAc'. On the right is the 'Genome browser' showing a genomic track for chromosome 17. The position 17:43170146-43170745 is highlighted. The track shows various regulatory regions and genes, including BRCA1 and NBR1. A red box highlights the 'VSARNTL_01' site in the track.

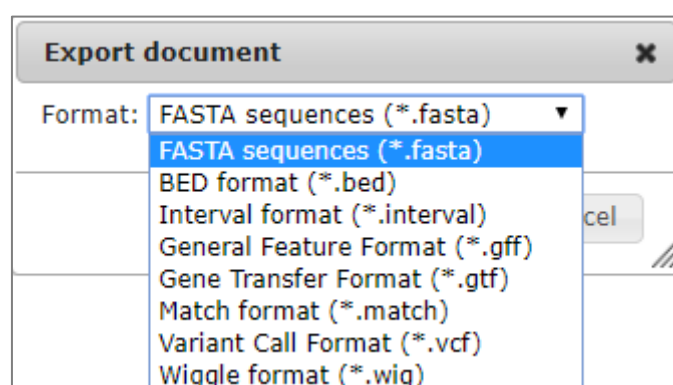
A right click on the track name in genome browser will open a context menu allowing you to remove the track from the visualization or open it as a table (function supported only for track of found sites or regulatory regions track):



Tracks opened as table can be exported using the *Export* button at the top menu panel or used for further analysis from the geneXplain platform perspective.

ID	Sequence (chromosome) name	From	To	Length	Strand	Type	Property: coreScore	Property: score	Property: siteModel
1	17	42886310	42886324	15	+	TF binding site	0.8181	0.85072	V\$JUNDFRA2_02
2	17	42886311	42886322	12	+	TF binding site	1	0.96061	V\$NFE2L1_Q5
3	17	42886311	42886322	12	+	TF binding site	0.74828	0.81669	V\$CREB3L4_02
4	17	42886391	42886401	11	+	TF binding site	0.94133	0.86479	V\$ATF5_01
5	17	42886394	42886400	7	+	TF binding site	1	1	V\$GKLF_Q4
6	17	42886411	42886421	11	-	TF binding site	1	0.92022	V\$RBPJK_01
7	17	42886426	42886434	9	+	TF binding site	0.92074	0.94714	V\$TAF1_07
8	17	42886444	42886457	14	-	TF binding site	0.91624	0.81753	V\$SERM_05

The list of available export formats is as follows:



GO categorization-based gene set analysis

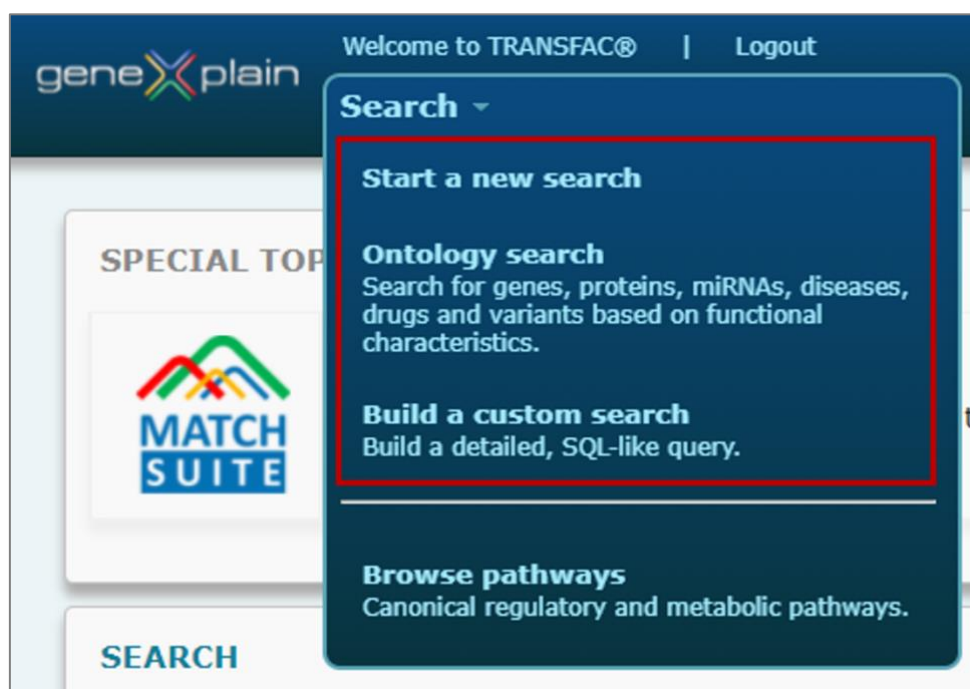
Please note that GO categorization-based analysis is available for sets of Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast genes.

Launching the analysis

Selecting the input gene list

Composing an input gene list in the TRANSFAC® database

A particularly convenient way to compose a gene list for analysis is to run the MATCH Suite on your search results in the TRANSFAC® database. For instance, when you search for a certain disease in the standard Search field or select genes belonging to a certain Gene Ontology (GO) category (Search > Ontology search), you end up with a list of genes.



In the search results inside TRANSFAC® database you can view all found entries on one page (select "All" in the "Hits on page" list) and then select the search results of your interest to launch the MATCH Suite gene set analysis on a respective gene list (you can use the *Mark all on page* option and then specify the organism of your interest (see below), then only genes of the selected organism will be taken to further analysis from your search results).

The screenshot shows the TRANSFAC database search results page. The top navigation bar includes 'geneXplain', 'Welcome to TRANSFAC® + PROTEOME™', 'Logout', 'Search', 'MATCH Suite', 'Tools', and 'My Data'. The search criteria section shows 'genes' that match 'GO biological process' is 'regulation of behavior'. The 'Categories' section shows a list of categories with counts. The 'Genes and proteins' section shows 348 of 348 total genes. A red box highlights the 'Launch MATCH Suite on the search results' button. Another red box highlights the 'Hits on page' dropdown menu, which is set to 'All'.

When you have selected the search results of your interest, click on the "Match Suite" tool and specify the model organism of your interest (Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast):

MATCH Suite options

? Select species: Mouse

? Select analysis type: Gene set

Submit

and you will be automatically transferred to the MATCH Suite tool, where you will be asked to assign a name to your gene list prior to proceeding to the next step of the analysis launch wizard:

Please name the selected gene list

Gene list name:

Give the name to your gene list and click on *Save* to proceed to the next step of the wizard.

Selecting input gene list from the MATCH Suite interface

When you click on the *Start analysis* section, the system immediately navigates you to the analysis launch wizard. At the very first step of this wizard you will be asked to select the type of the model organism you want to work with:

Select species

Human ▼

Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish, Baker's yeast species are supported. If you will select Human, at the next step of the wizard you should select the *Gene set analysis*:

Select analysis

☐ Single gene analysis

☒ Gene set analysis

In case of Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast gene set analysis, you will immediately be taken to the next step of the input gene list specification:

Start page

Show/hide menu

MATCH SUITE

Getting started

Start analysis

Launch history

TRANSFAC 2.0

Specify input data

☐ Use demo gene list 1

☒ Use my data 2

Supported input formats 3

☐ Construct tissue-specific gene list from scratch (requires subscription to HumanPSD™) 4

+ Upload files 5

Browse stored data 6

Textual input 7

Next > 11

	Name	Delete
☒ 8	My gene list 1 9	10 x
☐	My gene list 2	x
☐	My gene list 3	x
☐	My gene list 4	x
☐	My gene list 5	x
☐	My gene list 6	x

You can try launching the analysis using the **Use demo gene list** (1) option, or you can specify your own gene list by selecting the **Use my data** (2). The demo gene list can be viewed by clicking on the *demo gene list* blue link, it will contain Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast genes depending on the type of analysis you have selected. The **Supported input formats** (3) link also is clickable, it provides you with the information about the requested format of the gene list, which can be analyzed with the MATCH Suite 3.1:

Supported input formats

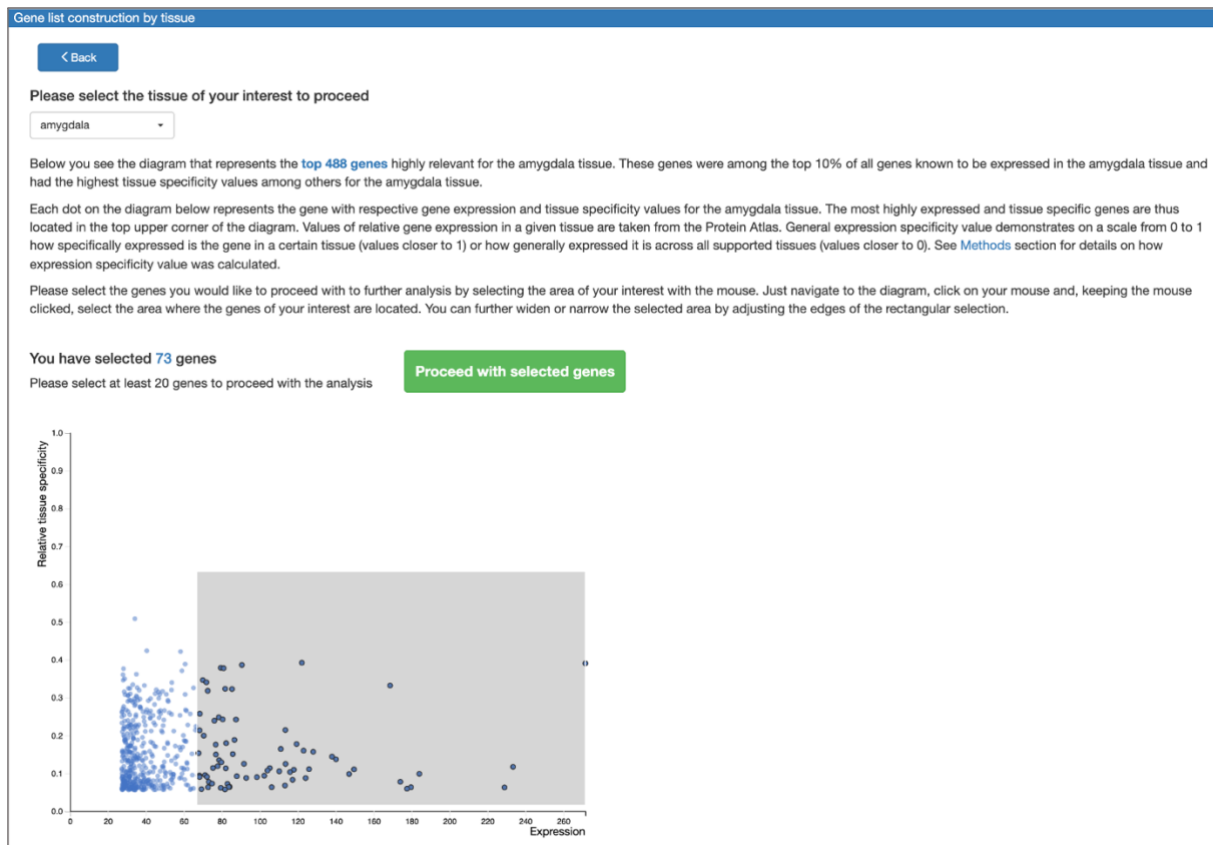
You can launch the analysis from any Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast gene list containing from 20 to 2000 genes, all coming in one of the following formats:

Gene symbols	e.g. TNMD
Genes Ensembl	e.g. ENSG00000000005
Genes Entrez	e.g. 64102

ok

Lists of 20 to 2000 genes coming in Ensembl ID, Entrez ID or Gene symbols format are accepted by the MATCH Suite 3.1. The system will select automatically the Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast genes from your input gene list, based on the selected type of the model organism, and only those genes will be subject to the further analysis.

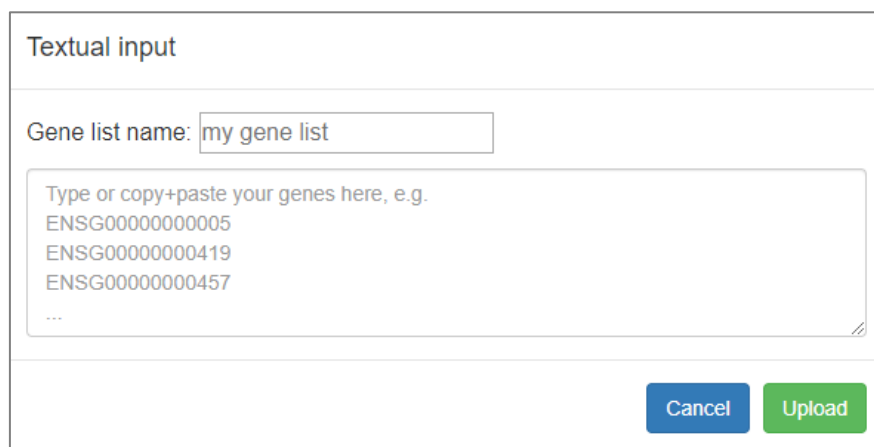
In case of human gene set analysis, a tissue-specific gene list can be automatically constructed for further analysis using the **Construct tissue-specific gene list from scratch** (4) option.



This option provides you with the ability to select any tissue among the 61 supported tissues and visualize the most tissue-specific Human genes among the top 10% of all genes known to be expressed in the selected tissue according to the information about relative gene expression levels provided by the Protein Atlas. You can further select the genes of your interest on the interactive diagram to either proceed with them to the analysis launch or just save the selected genes into the inputs of your project (both options managed by the *Proceed with selected genes* button). Please note that tissue-specific gene list construction option is available only to the HumanPSD™ database subscription owners.

The gene list for your analysis can be uploaded from your local computer using the **Upload files** (5) option, it can be also selected from the data stored in any geneXplain platform project accessible to you by using the **Browse stored data**

(6) option. In such case the selected gene list will be copied to the MATCH Suite project and taken for further analysis. The next option for input gene list specification is **Textual input** (7). This function provides you with an ability to simply copy and paste any gene list of your interest (in Ensembl, Entrez or Gene symbols format) to the dedicated textual input form:



The last option to specify an input gene list is to select it from the gene lists that were previously used by you as inputs for the MATCH Suite analysis (8). The names of the previously used gene lists are clickable and the respective gene list will open upon the click on its name (9). You can manage the gene lists stored in your project and permanently delete the unnecessary ones using the **Delete** option (10). Please note that the maximum size of your MATCH Suite project is 2 GB. You can free up space by deleting the unnecessary gene lists and analysis results if that would be needed (see *Operating in the Launch history* section below for further info)

To select the gene list for the current analysis run, mark it with the radio button in the list of all currently available inputs (8) (any newly uploaded gene list will automatically appear in this list).

Please note that regardless of the input gene list source, the selected gene list will be checked by the MATCH Suite for correspondence to the input gene lists requirements and only Human, Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast gene lists containing from 20 to 2000 genes in supported formats will be accepted for further analysis.

After selecting the input, click **Next** (11) to proceed.

Specifying the launch parameters

After giving a name to your gene list, the wizard will take you to the next step: in case of Mouse, Rat, Arabidopsis, Fruit fly, Nematoda, Zebrafish or Baker's yeast gene set analysis it will immediately be the *Specify parameters* step of the MATCH Suite wizard (described below), while in case of human gene set analysis, additional selection would have to be done to proceed:

MATCH SUITE

Getting started

Start analysis

Launch history

Show/hide menu

Select analysis focus

☐ Tissue expression based analysis

Info: Search for transcription factors regulating your gene set based on the expression profiles of these factors in the selected tissue

☒ GO categorization based analysis

Info: Search for transcription factors regulating your gene set based on the functional categorization of these factors in respect to the selected GO terms

Back Next >

Select *GO categorization based analysis* and click on *Next* to proceed to the parameters specification step.

In the next step of the analysis launch wizard the MATCH Suite will ask you to fill in the following form:

Start page

Show/hide menu

Specify parameters

Name of the analysis launch gene list 1

Promoter range 2

☒ Use default promoter range [-500,+100]

☐ Customized promoter range

From Max -5000 to Max +1000

Profile optimization

Narrow search to certain functional categories 3

Nothing selected

Narrow search to certain transcription factors 4

Nothing selected

5 < Back Next > 6

You can give a name to your analysis launch or keep the default name suggested by the system in the **Name of the analysis launch** field (1). This name will be further used in the launches history for easy selection of the run of your interest.

The promoter range that will be used for the search of TFBS in your analysis run is specified by the **Promoter range** parameter (2). By default the promoter range used by the MATCH Suite is [-500,100] relative to the TSS. You can specify a customized promoter range limited to the maximum of -5000 and +1000 from TSS.

If you want to narrow down the search for transcription factors (TFs) regulating your input genes to factors coming from certain GO functional categories, you can select the GO categories of your interest from the dropdown list in the **Narrow search to certain functional categories** field (3).

Optionally you can specify the transcription factors (TFs) of your interest from the dropdown list of all TFs of the selected model organism that have at least one TRANSFAC® matrix assigned to them in the **Narrow search to certain transcription factors** field (4). Only the matrices of the selected factors will be used for further analysis. Please refer to the [Methods](#) document for further details.

When selecting the TF, you can use the textual search. Tap on the factors of your interest to select them, or click on the name of the whole family (highlighted with red on the screenshot below) to select all factors of the respective family:

znf

Select All Deselect All

C2H2 zinc finger factors

2.3.2.4.1 ZNF414

2.3.2.4.3 ZNF580

2.3.2.4.5 ZNF740

2.3.3.0.100 ZNF35

2.3.3.0.114 ZNF354C

2.3.3.0.13 ZNF22

2.3.3.0.148 ZNF329

When done with parameters selection, click on **Next** (6) to proceed, or on **Back** (5) to return to the previous step of the wizard.

On the next step of the wizard your input gene set can be optimized by functional enrichment using the **Optimize the input gene list** option. If you tick this option, you can select the type of GO categories you are interested in. If you do not want to filter your input gene list by certain GO categories, simply do not select the checkbox and click on Next.

If you have selected to filter your input gene list by functional categories, you will be offered to select between the supported GO categories:

Optimize the input gene list

☒ Filter my gene list to genes coming from certain GO categories

☒ biological process

☐ cellular component

☐ molecular function

Back

Next >

Depending on your selection, the tree map visualization of functional classification of your input gene set will be built either on biological processes, or on cellular components, or on molecular functions.

When done with parameters selection, click on **Next** to proceed either to GO functional optimization of your input gene set or skip this step and start the analysis immediately by confirming the selected analysis launch parameters.

Filtering the input gene set by GO functional classification (optional step)

In case you have selected to optimize your input gene list by GO functional classification at the parameters specification step, the tree map of GO biological processes, cellular components or molecular functions will be constructed for your input gene set depending on your selection.

Start page

Getting started

Start analysis

Launch history

Show/hide menu

Selected launch parameters

Launch name: gene list(3) **1**

Input gene list: [gene list\(3\)](#)

Promoter range: [-500;+100]

Profile optimization by functional categories:

- GO:0003170 heart valve development
- GO:0003228 atrial cardiac muscle tissue development
- GO:0003230 cardiac atrium development
- GO:0003231 cardiac ventricle development
- GO:0007507 heart development

Profile optimization by selected TFs:

- 3.1.2.12.1,3.1.2.12.3 NANOG
- 3.3.1.15.1 FOXO1

GO categories selected for filtering of the input gene list: none

2 < Back **3** Launch analysis

All parameters that you selected for the current launch will be shown on the screen (1). You can check the genes that were eventually selected for the analysis launch after GO functional classification (if it was applied) by clicking on the name of the input gene list.

If you want to apply any changes to the specified parameters of the launch, click on **Back** (2), otherwise you are ready to start your analysis by clicking on the **Launch analysis** button (3).

Viewing the results

Operating in the *Launch history*

Once your analysis was launched, you will be redirected to the *Launch history* section, also accessible by the direct link at the left menu panel. The Launch history allows you to view the results of all your previous analysis runs and to follow the progress of the currently running analyses.

Start page

Getting started

Start analysis

Launch history

Show/hide menu

Recent launches

Select the analysis launch you want to view or [start a new analysis](#)

It is recommended to run only one analysis at once in order to optimize the system resources. Prior to starting a new analysis please check that your previous run has ended.

Date	Name	Parameters	Status	Results	Terminate/Delete
2024.06.24 17:02:12	gene list(3)	1 Mouse Gene set analysis View input parameters	Running: 2% 2 in progress...		3
2024.06.24 10:42:31	Mouse gene list 12	Mouse Gene set analysis View input parameters	Completed	4 View results	
2024.06.22 00:19:55	Mouse plasma genes	Mouse Gene set analysis View input parameters	Completed	View results	
2024.06.21 23:23:03	Rat plasma genes	Rat Gene set analysis View input parameters	Completed	View results	
2024.06.19 00:21:39	Mouse genes	Mouse Gene set analysis View input parameters	Completed	View results	5
2024.06.18 23:51:05	Human gene list	Human Gene set analysis View input parameters	Completed	View results	
2024.06.18 23:27:18	Rat gene list with additional conditions	Rat Gene set analysis View input parameters	Completed	View results	
2024.06.18 22:36:54	Rat gene list from AR pathway	Rat Gene set analysis View input parameters	Completed	View results	

[Start new analysis](#) **6**

You can check the parameters used for the launch by clicking on the **View input parameters** link (1). The following pop-up form will appear:

Launch parameters

gene list(3) 2024.06.24 17:02:12

Input:

gene list(3)

Promoter range:

[-500,+100]

Functional categories selected for profile optimization:

<GO:0003170> heart valve development
 <GO:0003228> atrial cardiac muscle tissue development
 <GO:0003230> cardiac atrium development
 <GO:0003231> cardiac ventricle development
 <GO:0007507> heart development

TFs selected for profile optimization:

3.3.1.15.1 FOXO1
 3.1.2.12.1,3.1.2.12.3 NANOG

GO categories selected for filtering of the input gene list:

none

ok

The gene set used for the current run can be viewed by clicking on the respective link in the *Input* field. You can click on *show* to view the *Functional categories selected for profile optimization* and *TFs selected for profile optimization* if these parameters were specified for the launch. In case you have selected to optimize your input gene list by certain GO categories, you can view them by clicking on the *show* link next to the *GO categories selected for filtering of the input gene list*

field. If no GO functional categories optimization was done, the respective field will display *none*.

The progress of the currently running analysis will be displayed in the **Status** column (2). Please note that the progress is displayed in percentage of the finished steps of the underlying workflow, and it has no direct correlation with the time left for the analysis to finish. Commonly one analysis run will take several hours, but this time interval is highly dependent on the input gene set and other parameters selected for the respective analysis launch.

It is not recommended to have several analysis launches running in parallel. Analysis runs will finish faster when launched consecutively one after another.

If you wish to terminate the launched analysis, you can click on the **Terminate** button (3).

To view the results of a finished analysis please click on the **View results** button (4).

You can manage the stored analysis results and delete the unnecessary data by clicking on the **Delete** icon (5). This action will delete all results of the respective analysis launch. The input gene list used for this analysis launch will still be accessible in the available list of inputs, which is viewed at the very first step of the start analysis wizard (see the *Selecting the input gene list* section of this document). By default, your MATCH Suite account is equipped with 2 GB disk space for storing your analysis results and input gene lists used. You can extend this volume by contacting us via info@genexplain.com with a respective request.

You can launch a new analysis run directly from the *Launch history* section by clicking on the **Start new analysis** (6) button.

Results visualization

Having selected in the *Launch history* section the analysis result which you want to view, it will open in the results visualization mode on your screen.

In the results visualization mode the screen will be divided into four different segments with the following functions:

- (1) The **identified transcription factors** regulating your gene set and respective **matrices tables**

- (2) The **table of your input genes** and found site hits in their promoters
- (3) The **info box** displaying the information about the currently selected object
- (4) The **genome browser**, allowing to visualize the tracks of found sites and additional annotation tracks for further results interpretation

The screenshot displays the MATCH Suite web interface with several key components highlighted by red boxes and numbers:

- Download tables (5):** A button in the top navigation bar for exporting data.
- Mouse gene list 12: factor/matrix view (1):** A table showing transcription factors and their enrichment scores.

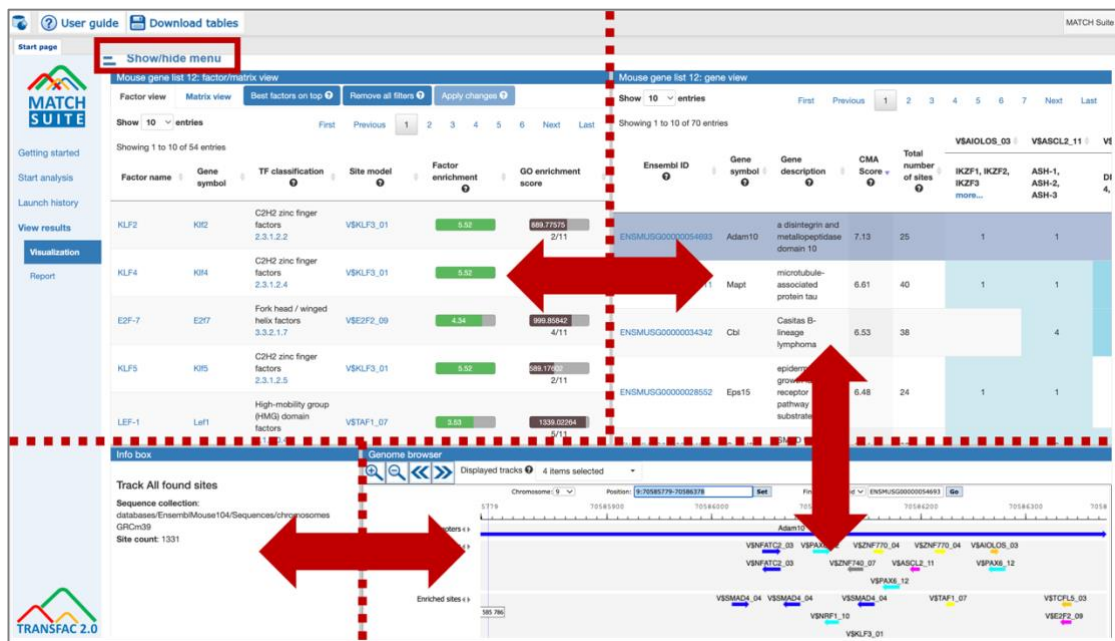
Factor name	Gene symbol	TF classification	Site model	Factor enrichment	GO enrichment score
KLF2	Klf2	C2H2 zinc finger factors 2.3.1.2.2	VSRLF3_01	5.92	889.77275 2/11
KLF4	Klf4	C2H2 zinc finger factors 2.3.1.2.4	VSRLF3_01	5.92	889.77575 2/11
E2F-7	E2f7	Fork head / winged helix factors 3.3.2.1.7	VSE2F2_09	4.94	899.85842 4/11
KLF5	Klf5	C2H2 zinc finger factors 2.3.1.2.5	VSRLF3_01	5.92	909.17697 2/11
LEF-1	Lef1	High-mobility group (HMG) domain factors	VSTAF1_07	5.93	1339.03264 5/11
- Mouse gene list 12: gene view (2):** A table showing gene details and enrichment scores.

Ensembl ID	Gene symbol	Gene description	CMA Score	Total number of sites	VSAL05_03	VSAL05_11	VSAL05_12
ENSMUSG000000054693	Adam10	a disintegrin and metalloprotease domain 10	7.13	25	1	1	
ENSMUSG00000018411	Mapt	microtubule-associated protein tau	6.61	40	1	1	
ENSMUSG000000034342	Cbl	Casitas B-lineage lymphoma	6.53	38			4
ENSMUSG000000028552	Eps15	epidermal growth factor receptor pathway substrate 15	6.48	24	1	1	
- Report (6):** A link in the left sidebar to view the analysis report.
- Info box (3):** A panel showing sequence collection details: databases/Ensembl/Mouse104/Sequences/chromosomes, GRCh38, Site count: 1331.
- Genome browser (4):** A visualization of the genomic region around position 70584300 on chromosome 9, showing various tracks including promoters, CMA sites, and enriched sites.

The Factor, Matrix and Gene tables that you see on the screen can be exported with the applied filters (see filtering instructions below) using the **Download tables button** (5). The archive containing these three tables in tab-separated text format will be downloaded to your local computer.

You can open the comprehensive analysis report about the respective run by clicking on the **Report** (6) link in the left menu panel. The self-explaining report will contain the factor, matrix and gene tables you see in the results visualization section along with supplementary tables and analysis steps description.

You can extend or narrow certain segments of the screen by moving the splitter lines. At this point you might want to use the *Show/hide menu* button, which will hide the left panel menu from view and will broaden the results area of the screen.



In the screen segment, which visualizes the tables of predicted factors and respective matrices, you will find two accessible tables: Factor view and Matrix view.

Factor view table

The **Factor view** (1) table reports on the transcription factors predicted to be regulating your input gene set. By default, top 10 factors will be shown, this can be changed in **Show <number> entries** (2) field. The total number of factors identified will be displayed in (3). You can navigate through the predicted factors using the pages in (4). The columns of the *Factor view* table are fully matching the *Table 3* given in the analysis report of the respective run. You can refer to the analysis report for denominations of column names and their contents or use the info hints provided in the results visualization interface as a mouseover message that will appear upon hovering above the ? sign (5) available at multiple places in the interface.

Factor view 1

Matrix view

Best factors on top 7

Remove all filters 8

Apply changes 9

Show 10 entries 2

First

Previous

1

2

3

4

5

6

Next

Last 4

Showing 1 to 10 of 54 entries 3

Factor name

Gene symbol

TF classification ?

Site model ? 5

Factor enrichment ?

GO enrichment score 6

KLF2

Klf2

C2H2 zinc finger factors
2.3.1.2.2

V\$KLF3_01

5.52

889.77575

2/11

KLF4

Klf4

C2H2 zinc finger factors
2.3.1.2.4

V\$KLF3_01

5.52

889.77575

2/11

E2F-7

E2f7

Fork head / winged helix factors
3.3.2.1.7

V\$E2F2_09

4.34

999.85842

4/11

KLF5

Klf5

C2H2 zinc finger factors
2.3.1.2.5

V\$KLF3_01

5.52

589.17602

2/11

LEF-1

Lef1

High-mobility group (HMG) domain factors
4.1.3.0.4

V\$TAF1_07

3.53

1339.02264

5/11

The columns of tables have in-built sorting option (6) which allows to sort the values within one column by ascending or descending order or by alphabetical order in case of textual contents. Simply click on the gray arrows for the sorting to be applied. By default, best factors (or matrices) are brought to the top (please see the analysis report and the [Methods](#) document for explanations on the factors and matrices ranking procedures). If you want to return to the original order of factors, click on the **Best factors on top** button (7). Filters applied from the Matrix view table (see below) can be removed using the **Remove all filters** button (8).

Matrix view table

The **Matrix view** table (1) shows the PWMs (positional weight matrices) of the TRANSFAC[®] library, the respective sites of which were identified in the promoters of the studied gene set.

Similar to the factors table, the *Best matrices on top* button allows to bring the best matrices to the top of the matrix table in case their order was changed while sorting the values in individual columns. This table fully corresponds to the *Table 4* of the analysis report, where you can find the denominations of its column names and their contents description. Respective info is also summarized in the **?** hints

available as mouseover messages upon hovering above them, similar to the *Factor view* table.

Matrix ID	Matrix logo	Adjusted site enrichment	Site enrichment	Site enrichment FDR	Adjusted sequence enrichment	Sequence enrichment FDR	Composite model
V\$NRF1_09		4.24	7.65	5.63e-21	3.17	2.39e-13	yes
V\$TFCP2_08		2.63	5.64	4.48e-7	1.70	3.30e-9	yes
V\$PAX6_12		2.55	3.69	2.68e-53	1.68	2.13e-13	yes
V\$ZNF740_07		2.11	3.04	2.69e-44	1.44	1.04e-13	yes
V\$ASCL2_11		2.07	3.46	4.03e-11	1.74	1.35e-9	yes

The **Adjusted site enrichment filter** (2) allows to leave only matrices with the adjusted site enrichment values higher than the threshold specified by the filter. The set filter will be first applied exclusively to the *Matrix view* table. To recalculate the factors and the genes tables, as well as the track of the found sites in the genome browser, only with matrices that were left after applying the filter, you should click on the *Apply changes* (4) button.

The **Adjusted sequence enrichment filter** (3) allows to leave only matrices with the adjusted sequence enrichment values higher than the threshold specified by the filter. The set filter will be first applied exclusively to the *Matrix view* table. To recalculate the factors and the genes tables, as well as the track of the found sites in the genome browser, only with matrices that were left after applying the filter, you should click on the *Apply changes* (4) button.

Filters will be also auto applied to all tables and tracks when you will switch from matrix table to the factor table. To cancel the filters, click on the *Remove all filters* button.

Please note that the **Composite model** column will be displayed in the Matrix view table only in case the combinatorial analysis of sites was performed (please see the analysis report and the Methods document for the details).

Genes table

The genes table presented in the results visualization section fully corresponds to the gene table provided in the *Table 5* of the analysis report, where you can find the denominations of its column names and their contents description. Respective info is also summarized in the **?** hints available as mouseover messages upon navigating on them, similar to the *Factor view* and *Matrix view* tables.

Show	10	entries	First	Previous	1	2	3	4	5	6	7	Next	Last
Showing 1 to 10 of 70 entries													
Ensembl ID	Gene symbol	Gene description	CMA Score	Total number of sites	V\$AIOLOS_03	V\$ASCL2_11	V\$						
					IKZF1, IKZF2, IKZF3 more...	ASH-1, ASH-2, ASH-3	D						
ENSMUSG00000054693	Adam10	a disintegrin and metallopeptidase domain 10	7.13	25	1	1							
ENSMUSG00000018411	Mapt	microtubule-associated protein tau	6.61	40	1	1							
ENSMUSG00000034342	Cbl	Casitas B-lineage lymphoma	6.53	38		4							
ENSMUSG00000028552	Eps15	epidermal growth factor receptor pathway substrate 15	6.48	24	1	1							

Please note that the **CMA Score** column will be displayed in the Genes table only in case the combinatorial analysis of sites was performed (please see the analysis report and the Methods document for the details).

To visualize in genome browser the sites found in the promoter of any gene, just click on the line with the gene of your interest inside the genes table and refer to the genome browser below to explore the predicted regulation model of the respective gene.

Genome Browser

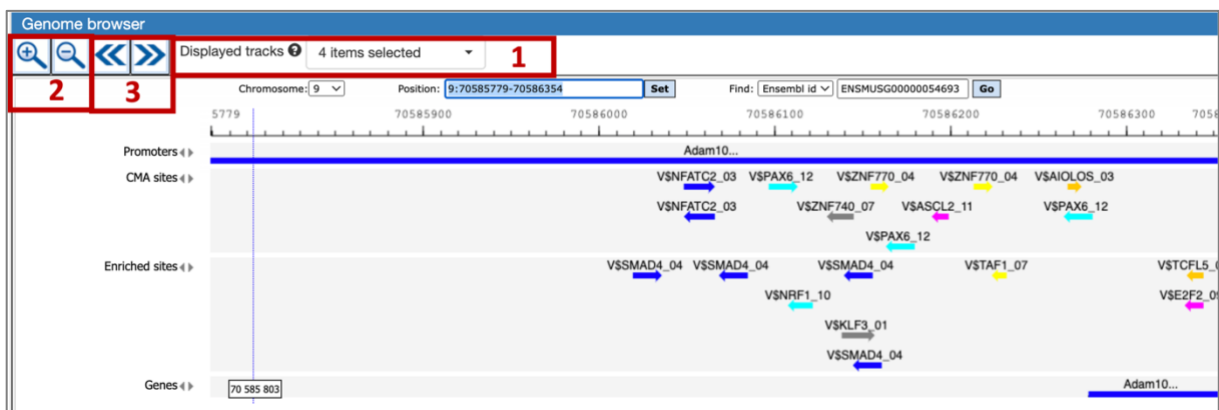
The genome browser provides you with visualizations of the predicted gene regulation models for each gene from the input gene set. Having clicked on any

gene from the *Genes table*, the promoter model of the respective gene will be automatically opened in the genome browser with visualization of the sites found within the respective promoter. In case the combinatorial analysis of sites was performed (please see the analysis report and the Methods document for the details), by default the displayed tracks will include:

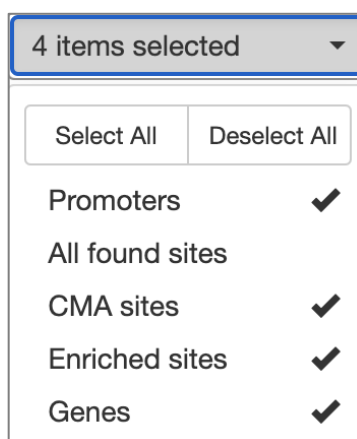
- **Promoters** - the track of all promoters that were used for the performed analysis
- **CMA sites** - the track of sites belonging to the combinatorial matrices of the constructed CMA model (see [Methods](#) for further info)
- **Enriched sites** - the track of the remaining sites found to be enriched in the promoters of the studied genes
- **Genes** - the track with all Ensembl genes

If combinatorial analysis of sites was not performed (please see the analysis report and the Methods document for the details), the displayed tracks will include:

- **Promoters** - the track of all promoters that were used for the performed analysis
- **All found sites** - the track of all sites found to be enriched in the promoters of the studied genes
- **Genes** - the track with all Ensembl genes
-



Displayed tracks can be customized using the **Displayed tracks** dropdown list (1)

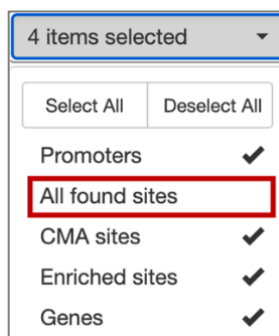


4 items selected	
Select All	Deselect All
Promoters	✓
All found sites	
CMA sites	✓
Enriched sites	✓
Genes	✓

You can zoom in and zoom out for a more generalized or a more detailed view of the promoter model using the **zoom in** and **zoom out** buttons (2). In case navigation in genome browser is lost due to multiple zoom clicks, you can always return to the promoter of your interest by selecting the respective gene in the genes table once again.

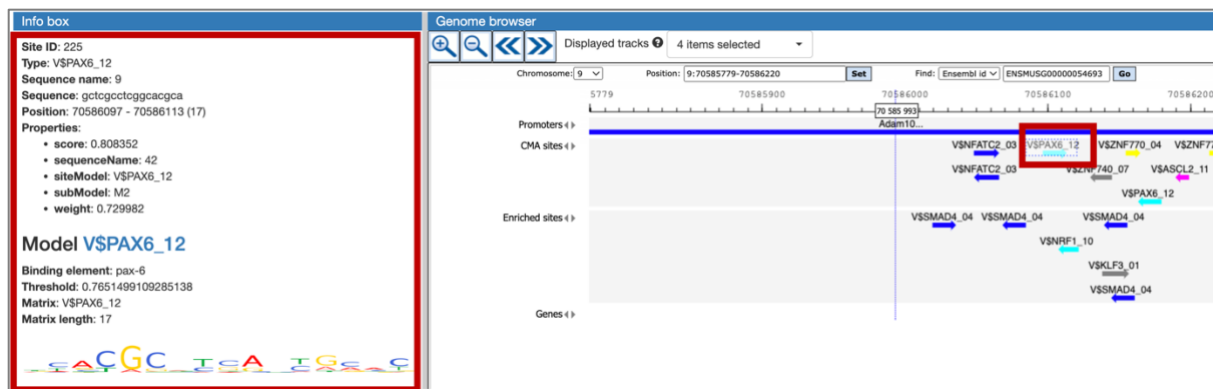
Navigation through the track is supported by the **Shift to the left** and **Shift to the right** buttons (3)

Please note that the *CMA sites* track shows only the best sites relevant to the identified CMA combinatorial model. The genes table will contain the full counts for all sites identified, both enriched and combinatorial. To view the respective full track of all found sites, you can add the *All found sites* track to the visualization:



4 items selected	
Select All	Deselect All
Promoters	✓
All found sites	
CMA sites	✓
Enriched sites	✓
Genes	✓

For receiving additional information about any of the found sites, you can click on the site of your interest and explore the contents displayed in the *Info box* section:



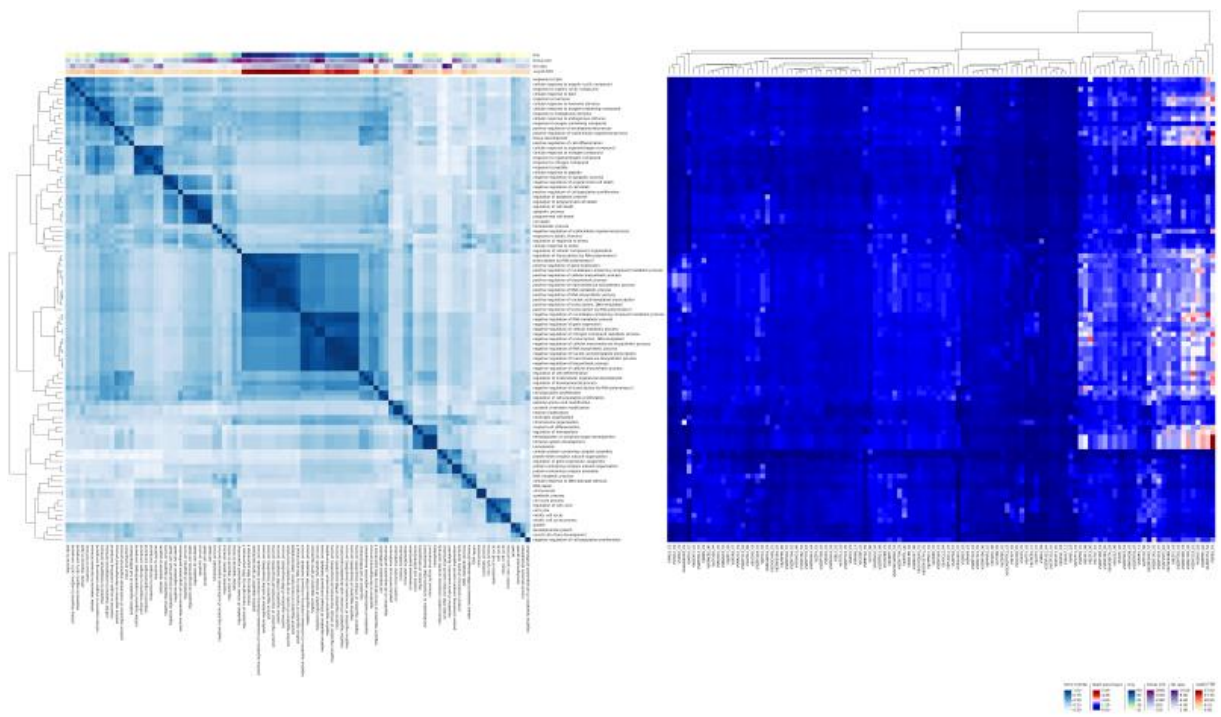
Heatmaps

If the functional classification of your input gene set by GO terms produced reliable results (by p-values and by the number of genes in each functional category), heatmaps for the respective GO categorization will be included in the optional section *Functional Analysis of Gene Regulation* of the analysis report.

The constructed image will contain two parts:

- (3) The heatmap of GO to GO terms mapping for the GO terms overrepresented among the studied gene set;
- (4) The heatmap visualizing how enriched motifs are associated with the respective GO categories.

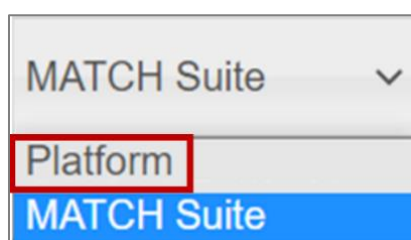
The visualization examples of such heatmaps is given below. Depending on your preference, you can use either of the heatmaps: the one with the horizontal layout (provided in the analysis report as a clickable image), or the one with the vertical layout (provided under the link to alternative view in the report text):



Switching to geneXplain platform

All results received with the MATCH Suite can be further analyzed in the geneXplain platform tool, access to which is provided to all users of the MATCH Suite. GeneXplain platform is a comprehensive online toolbox and workflow management system for a broad range of bioinformatics and systems biology applications.

For switching from the MATCH Suite perspective view to the geneXplain platform view, please use the perspectives switcher located at the top upper corner of the screen:



Please, refer to the User Guide of the geneXplain platform for detailed description of its functions.