The toolbox of TRANSFAC 2.0

User guide release 3.0

Table of contents

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.

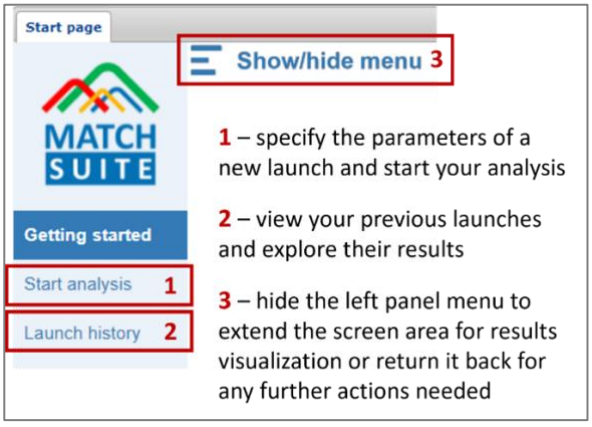
MATCH Suite can be launched on an individual human gene or on a list of Human, Mouse, or Rat genes. Depending on your selection, different analyses will be ran to identify the activation patterns of your input (see Methods document linked to the analysis report of each run for further details).

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.

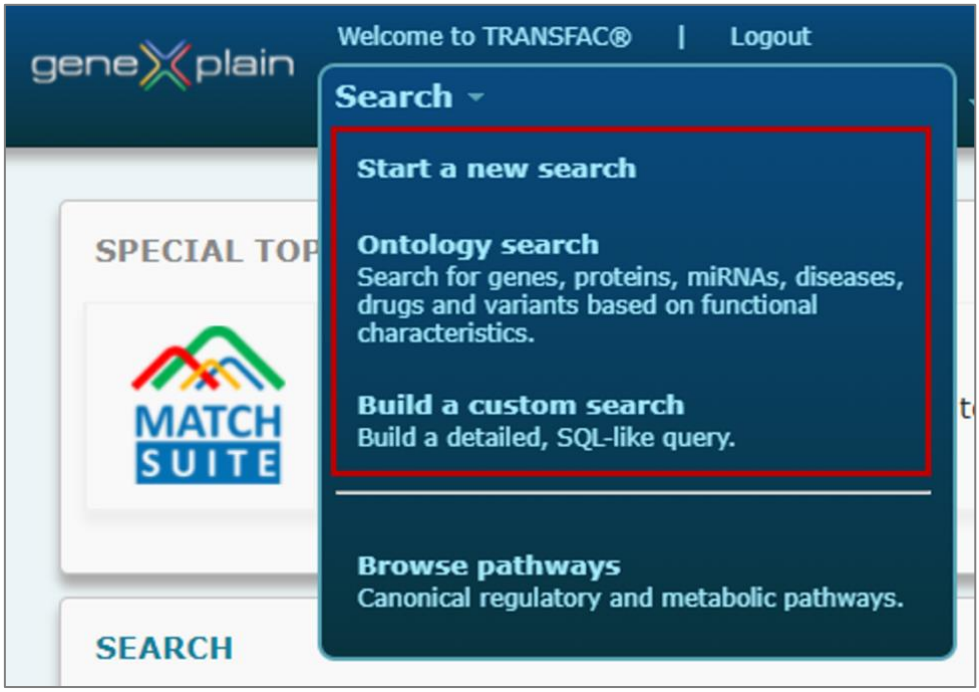
Gene set analysis

Launching the analysis

Selecting the input gene list

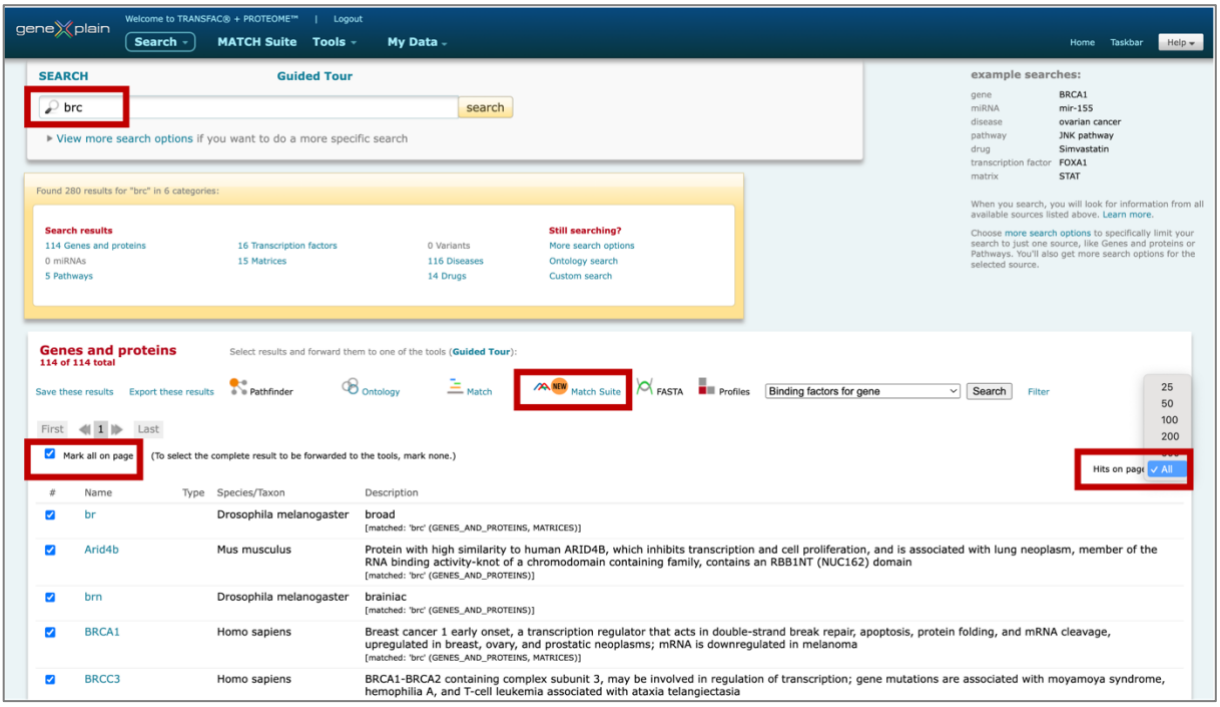
Composing an input gene list in the TRANSFAC® database

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.

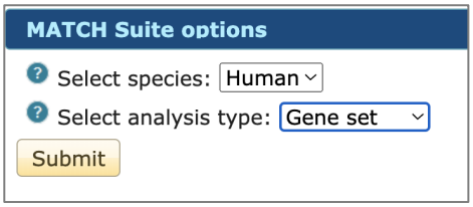


In the search result inside TRANSFAC® database you can view all found entries on one page (select "All" in the "Hits on page" list) and then select the search results

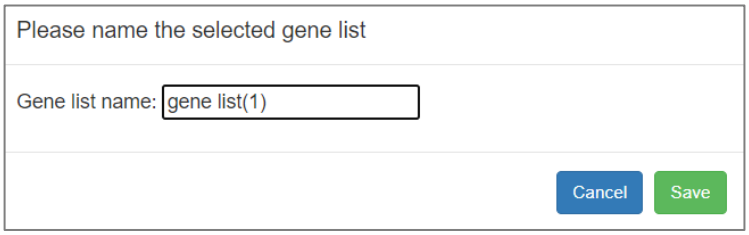
of your interest to launch the MATCH Suite gene set analysis on a respective gene list (you can use the *Mark all on page* option and then only human genes from your selection will be automatically taken to the MATCH Suite analysis upon launching it).



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:



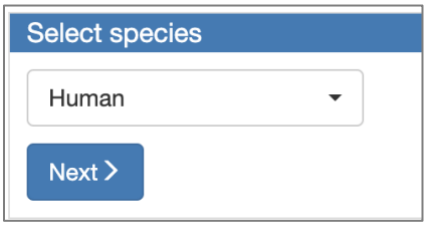
Click on *Submit* and you will be automatically transferred to the MATCH Suite tool, where you will be asked to assign a name to your gene list prior to proceeding to the next step of the analysis launch wizard:



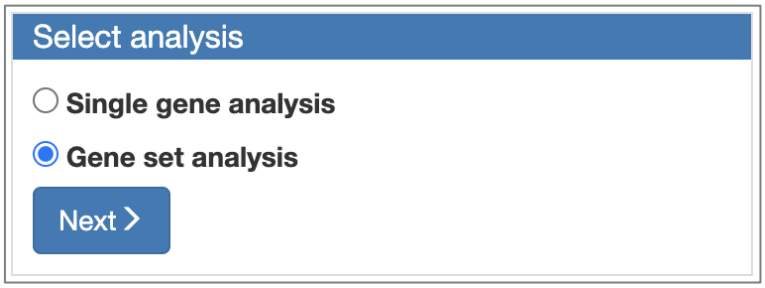
Click on *Next* to proceed to the parameters specification step.

Selecting input gene list from the MATCH Suite interface

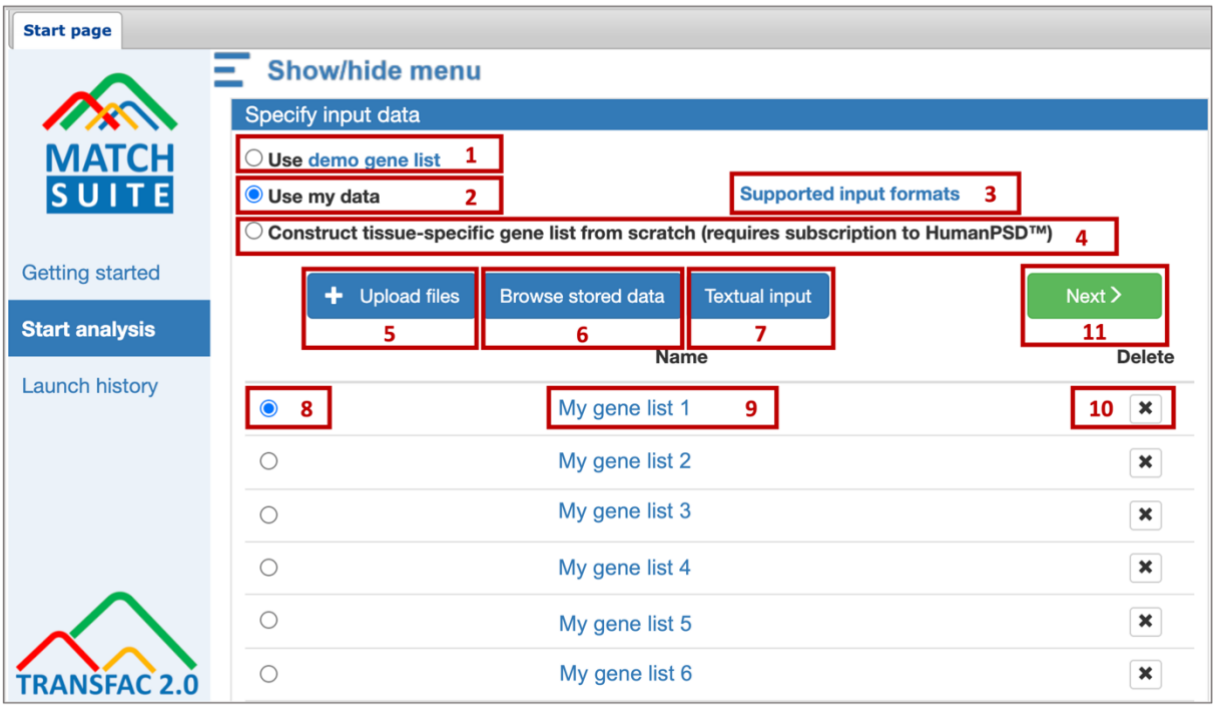
When you click on the *Start analysis* section, the system immediately navigates you to the analysis launch wizard. At the very first step of this wizard you will be asked to select the type of the model organism that you are studying:



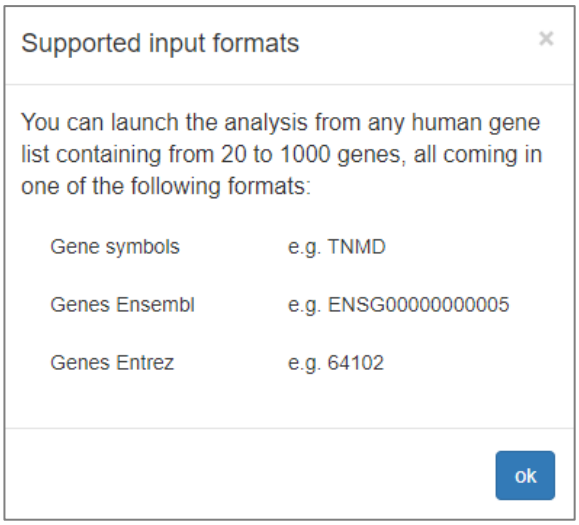
Select Human and click on Next. After that specify the analysis type you would like to perform:



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.

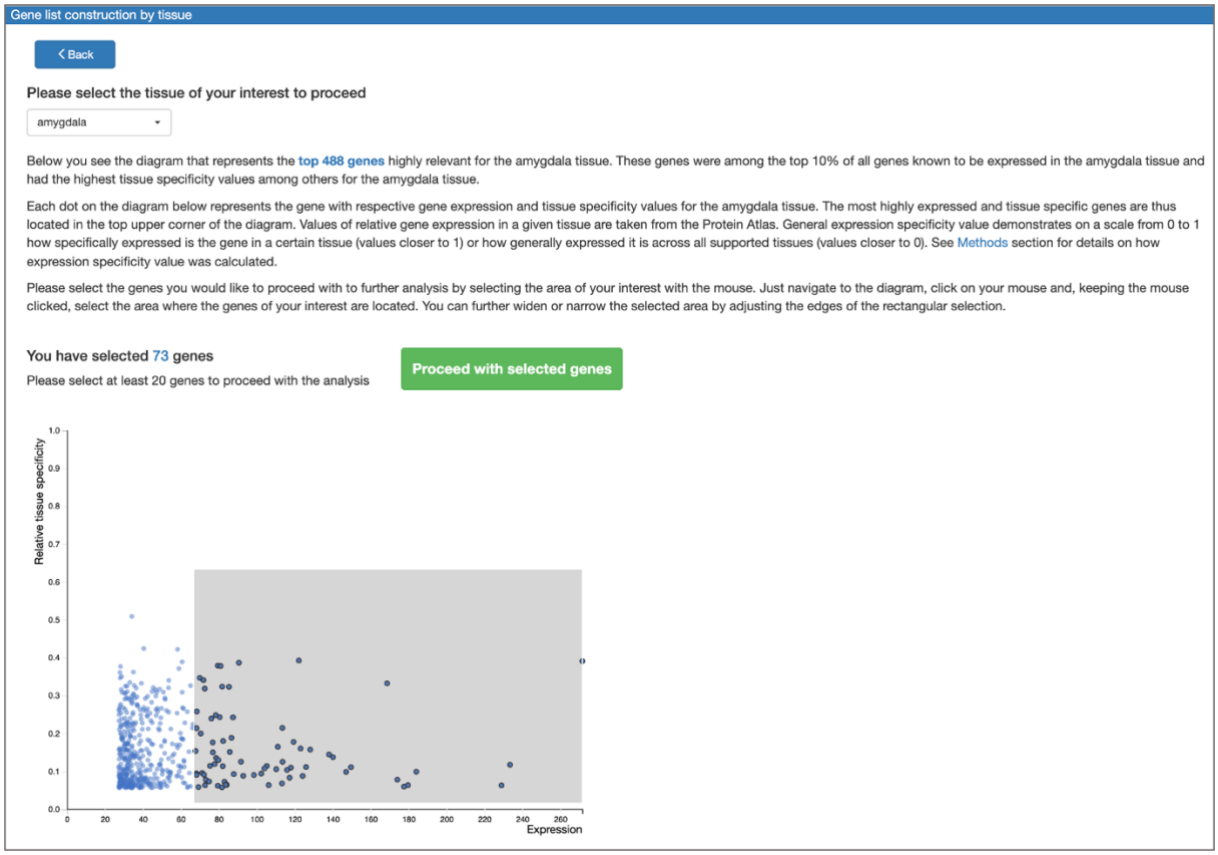


You can try launching the analysis using the **Use demo gene list** (1) option, or you can specify your own gene list by selecting the **Use my data** (2). The demo gene list contains 50 Ensembl genes, you can view them by clicking on the *demo gene list* blue link. The **Supported input formats** (3) 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 3.0:



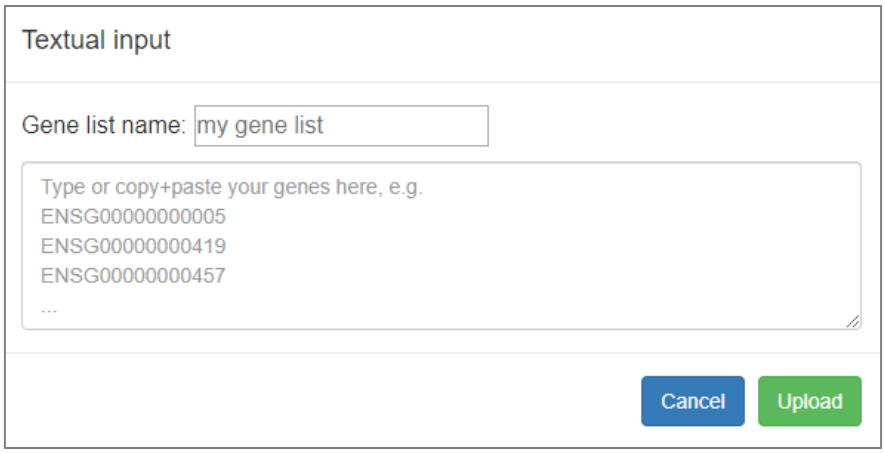
Lists of 20 to 2000 genes coming in Ensembl ID, Entrez ID or Gene symbols format are accepted by the MATCH Suite 3.0. The system will select automatically the human genes from your list, only these will be subject to the further analysis.

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:



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)

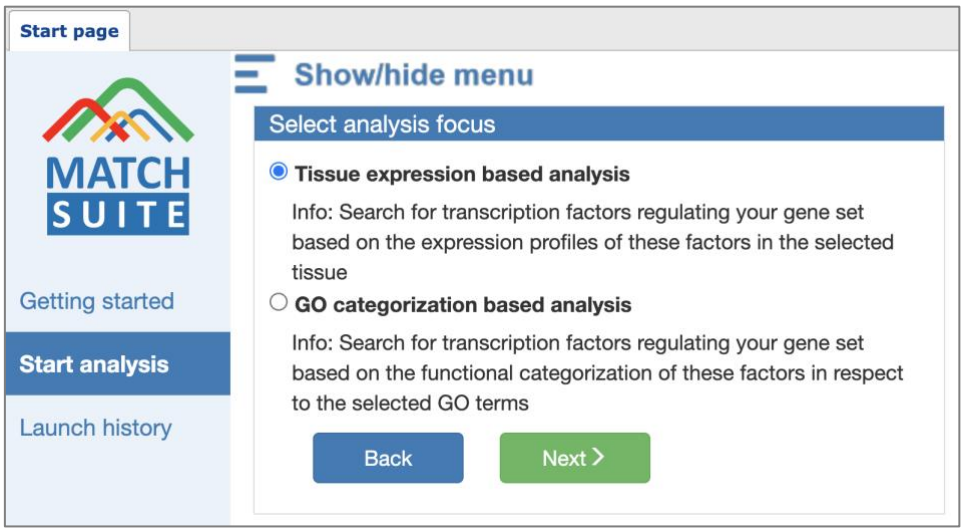
To select the gene list for the current analysis run, mark it with the radio button in the list of all currently available inputs (8) (any newly uploaded gene list will automatically appear in this list).

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.

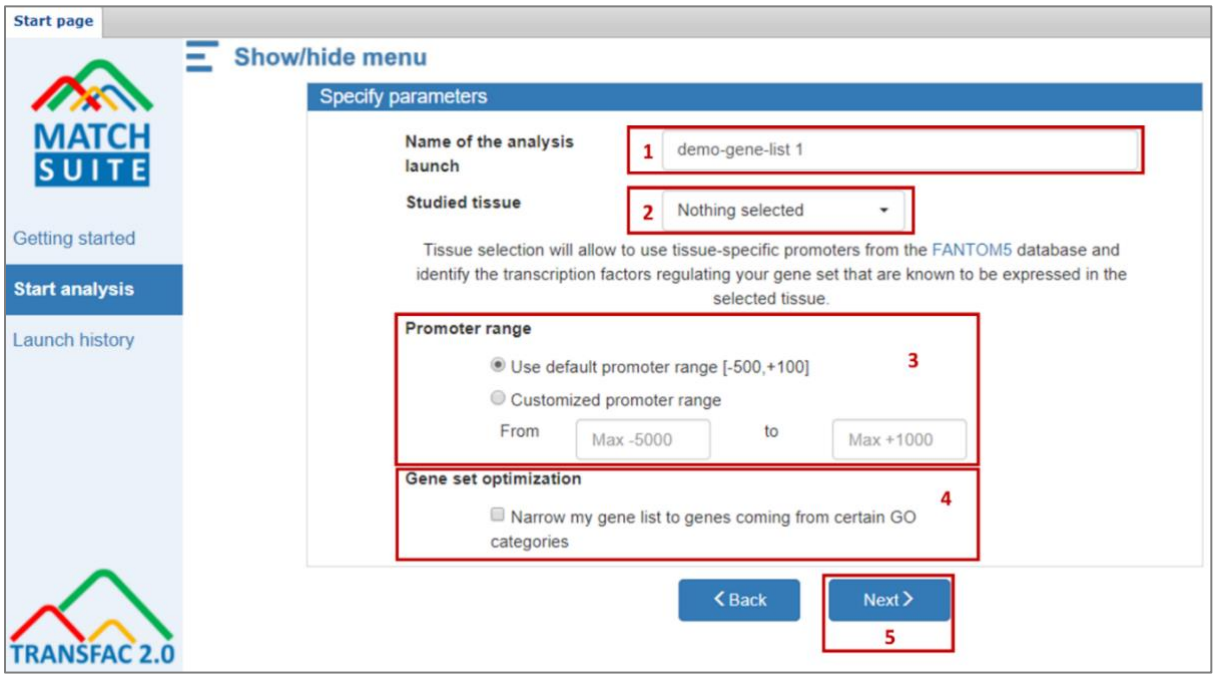
Specifying the launch parameters

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:



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:



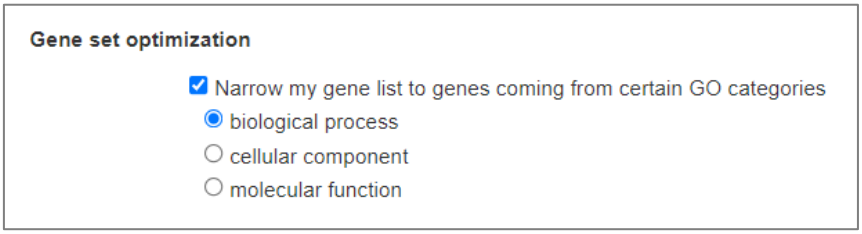
You can give a name to your analysis launch or keep the default name suggested by the system in the **Name of the analysis launch** field (1). This name will be further used in the launches history for easy selection of the run of your interest.

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:

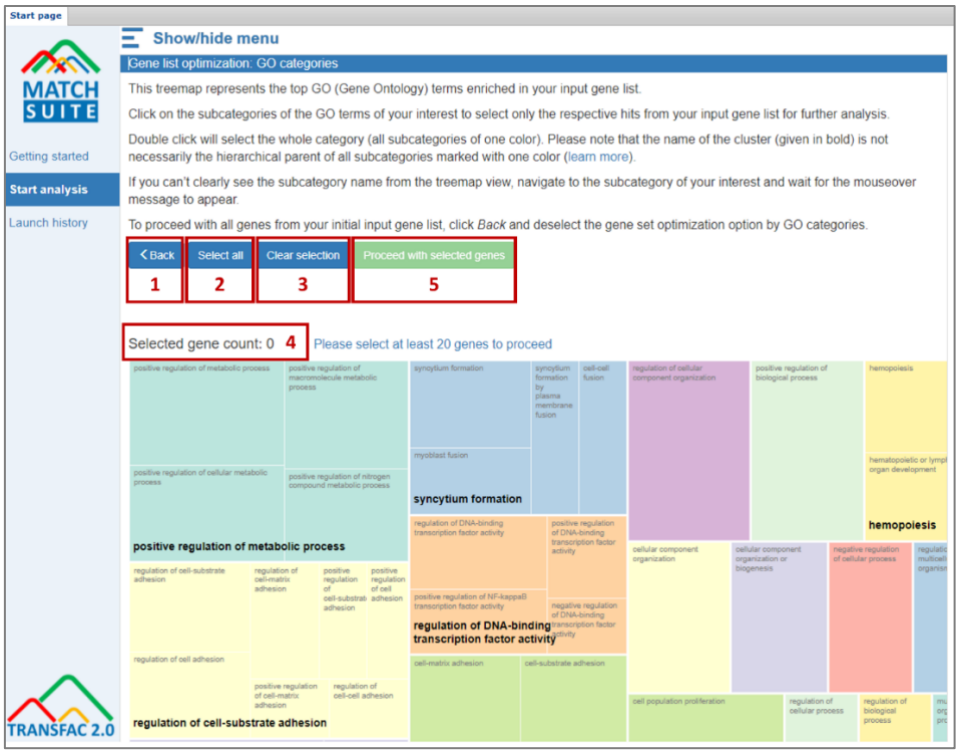


Depending on your selection, the tree map visualization of functional classification of your input gene set will be built either on biological processes, or on cellular components, or on molecular functions. If you do not wish to optimize your input gene set by functional categories, simply do not tick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

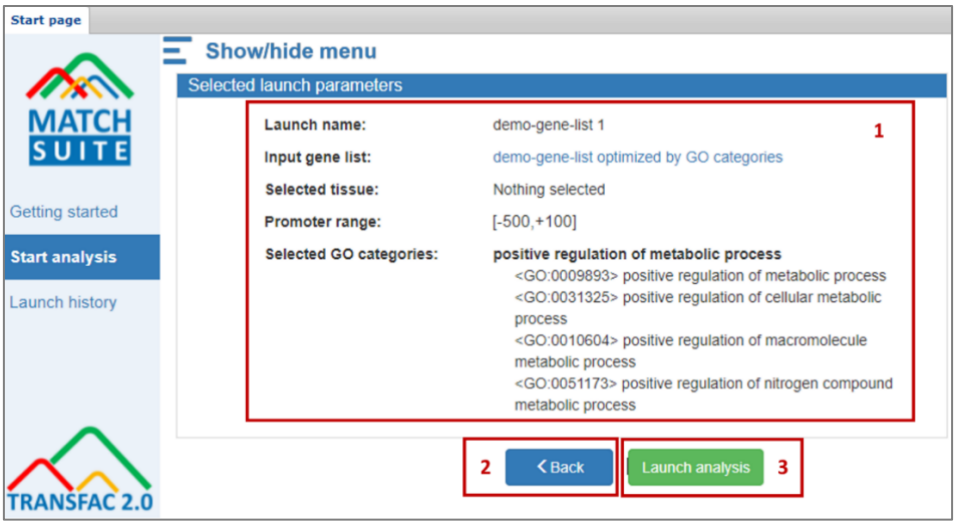


You can click **Back** (1) to return to the parameters specification step and cancel the GO optimization of your input gene list by unticking the *Gene set optimization* option. Otherwise you are requested to select the GO categories of your interest, following the instructions provided at the top of 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:



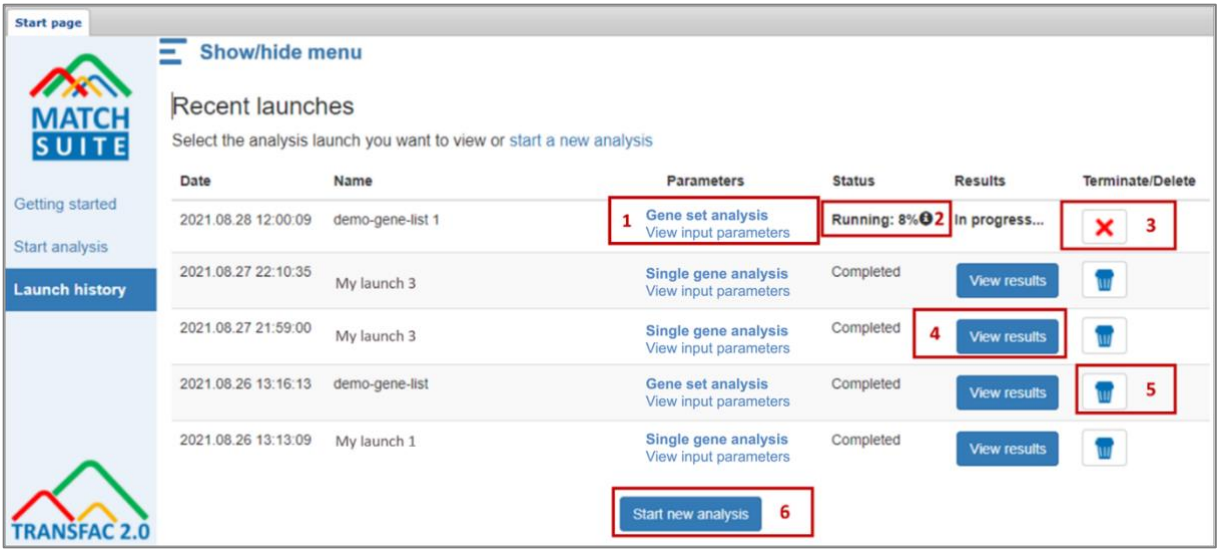
All parameters that you selected for the current launch will be shown on the screen (1). You can check the genes that were eventually selected for the analysis launch after GO functional classification (if it was applied) by clicking on the name of the input gene list.

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).

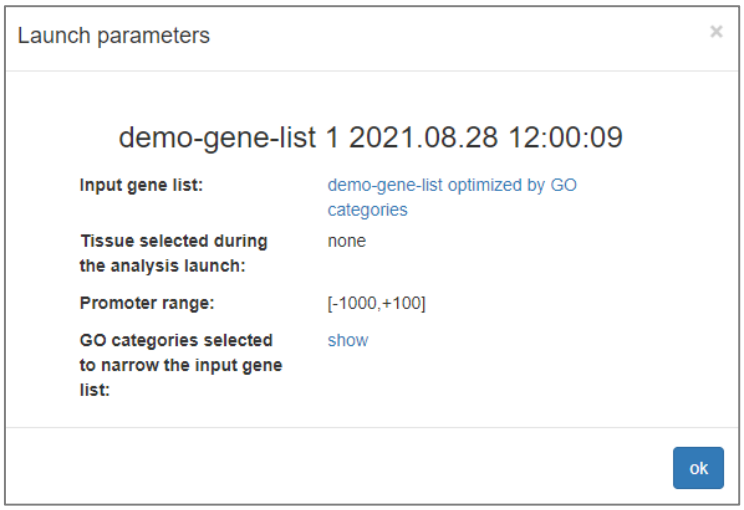
Viewing the results

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.



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:



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.

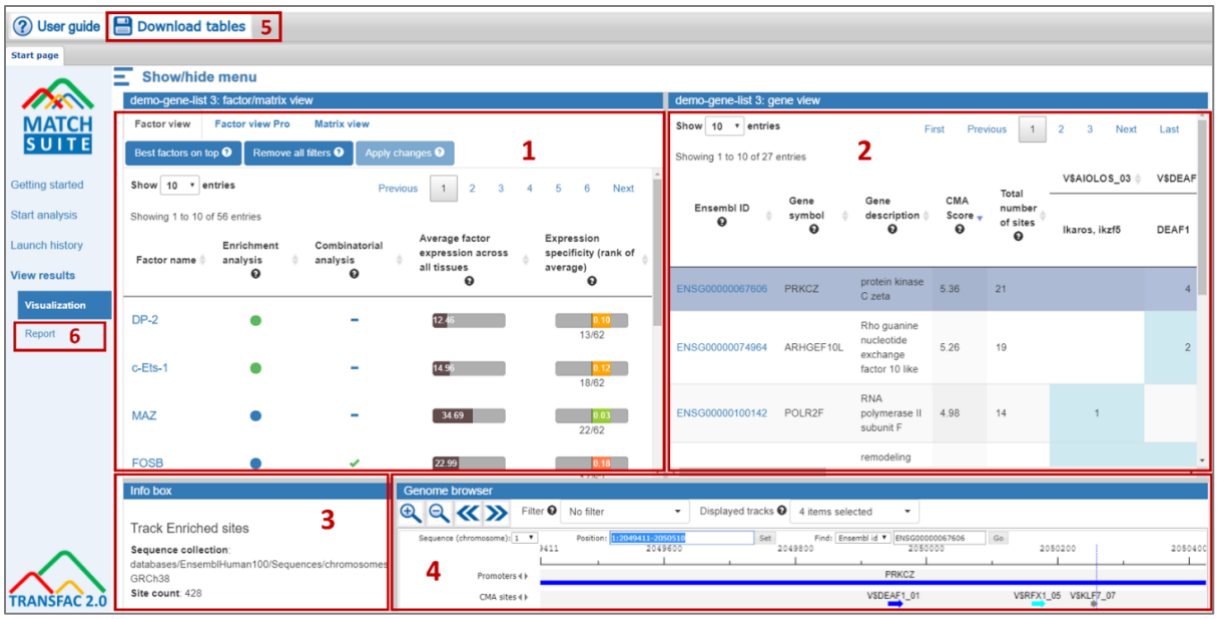
In the results visualization mode the screen will be divided into four different segments with the following functions:

(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

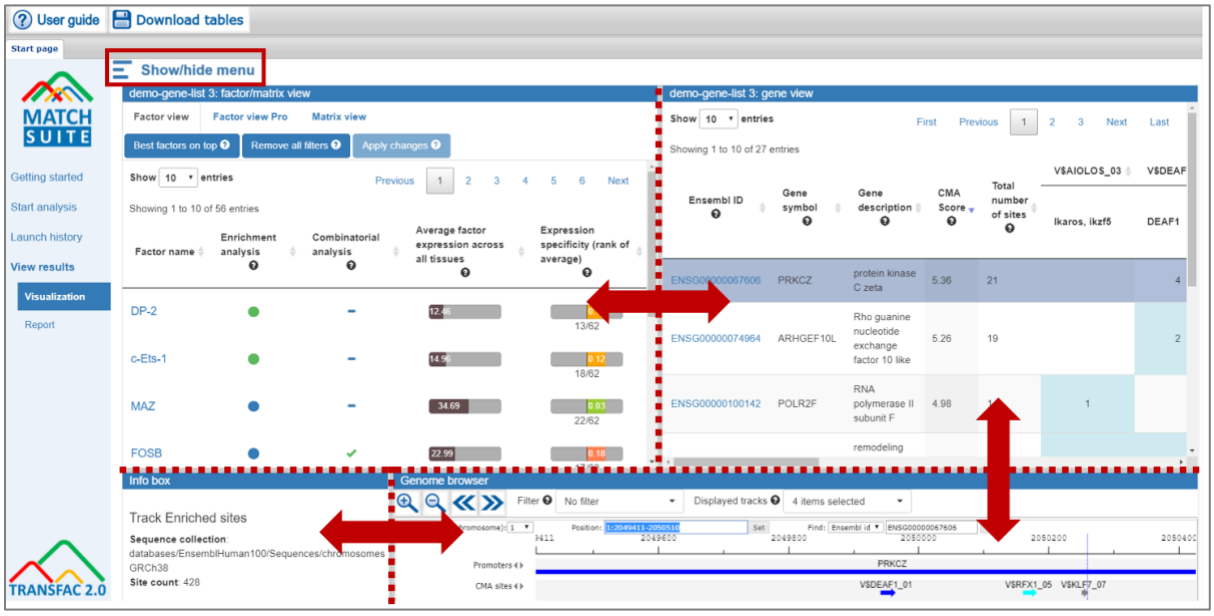
(4) The **genome browser**, allowing to visualize the tracks of found sites and additional annotation tracks for further results interpretation



The 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.



TRANSFAC 2.0

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

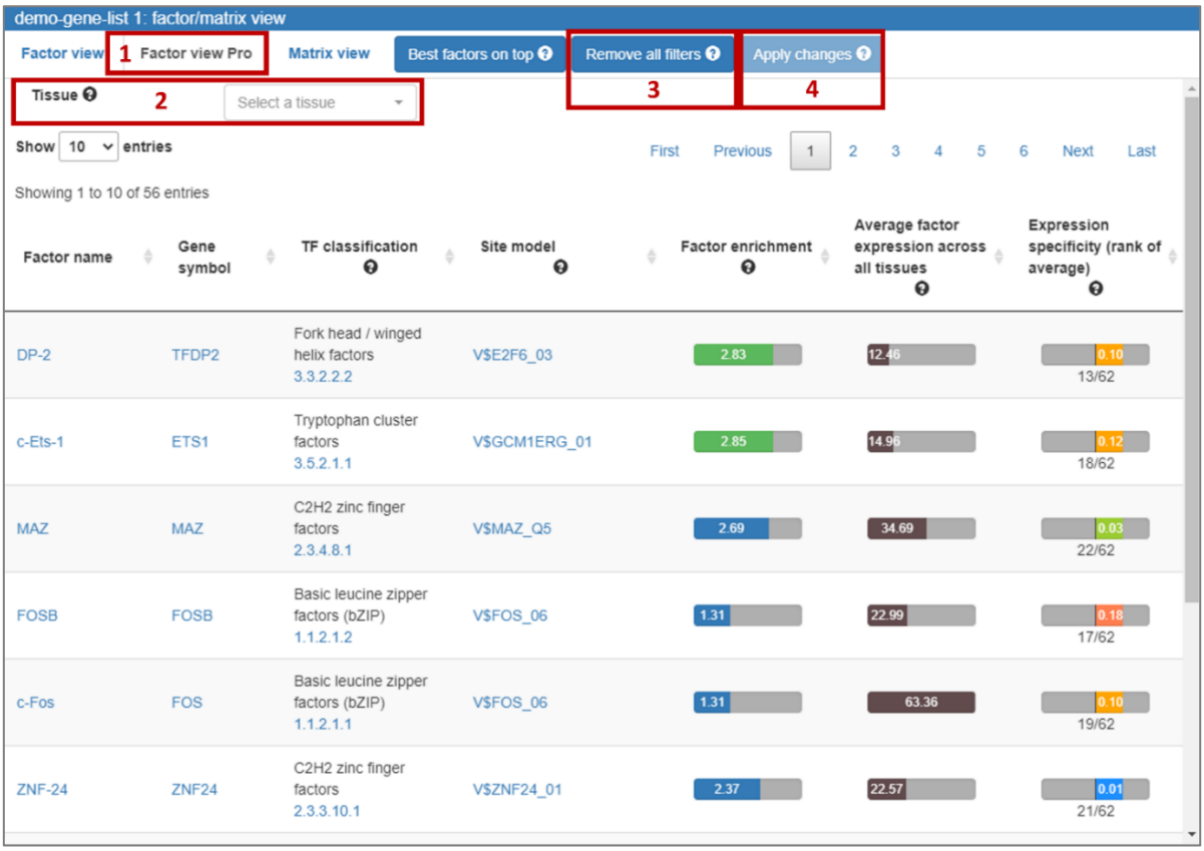
In the **Factor view** (1) table you will see a simplified summary of the transcription factors predicted to be regulating your input gene set. By default top 10 factors will be shown, this can be changed in **Show <number> entries** (2) field. The total number of factors identified will be displayed 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.

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.



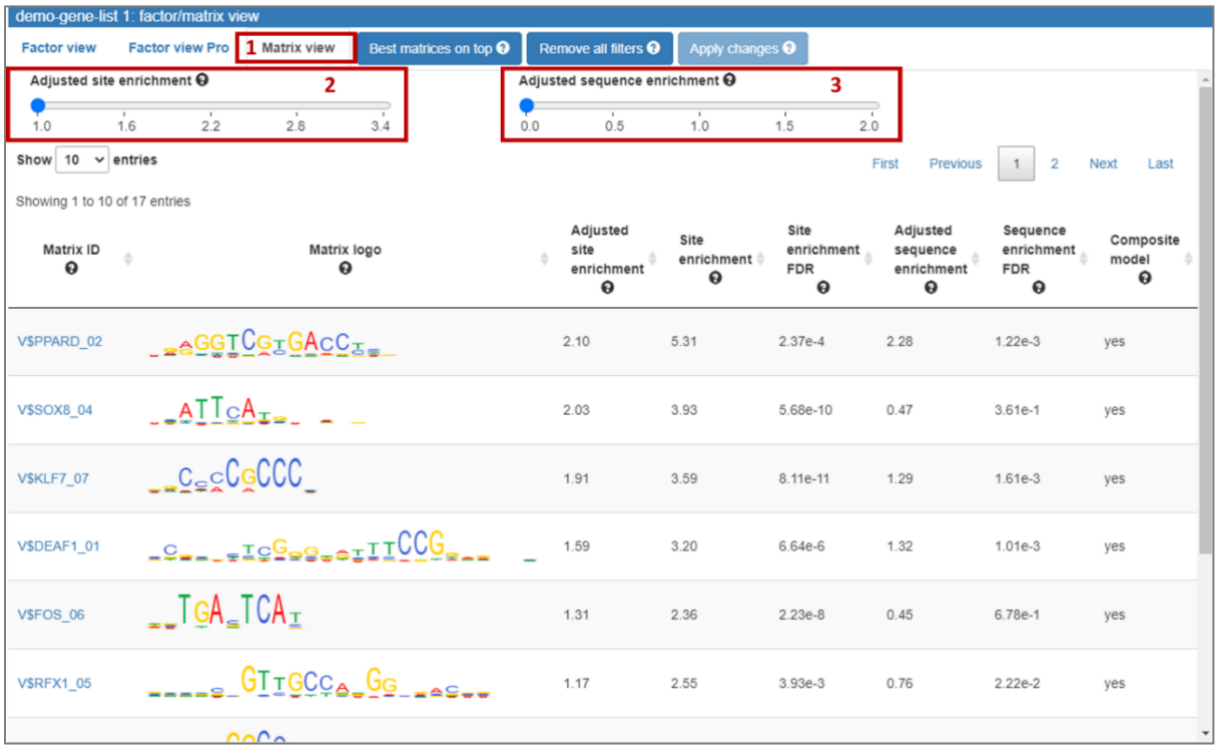
The *Factor view Pro* table allows you to apply the **Tissue** filter (2). You can select the tissues(s) of your interest from the dropdown list in (2) and click on the **Apply changes** button (4) to recalculate the factors and genes tables, as well as the track of found sites in the genome browser, 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.



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

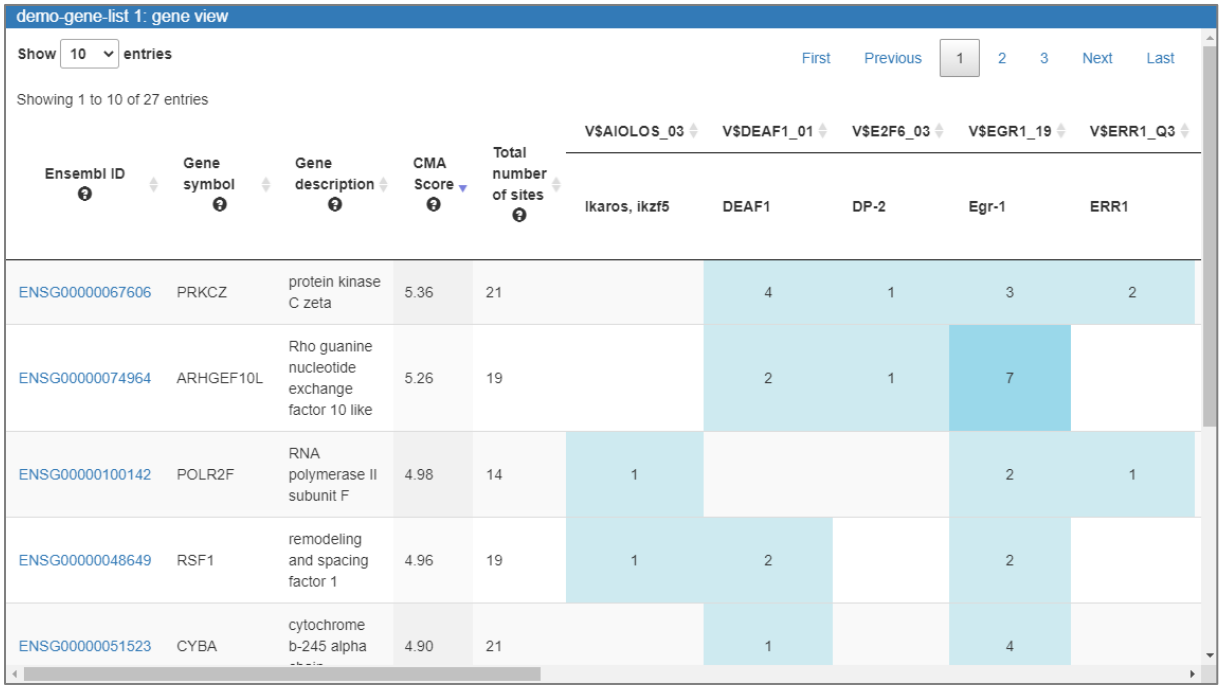
genome browser, only with matrices that were left after applying the filter, you should click on the *Apply changes* button.

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.

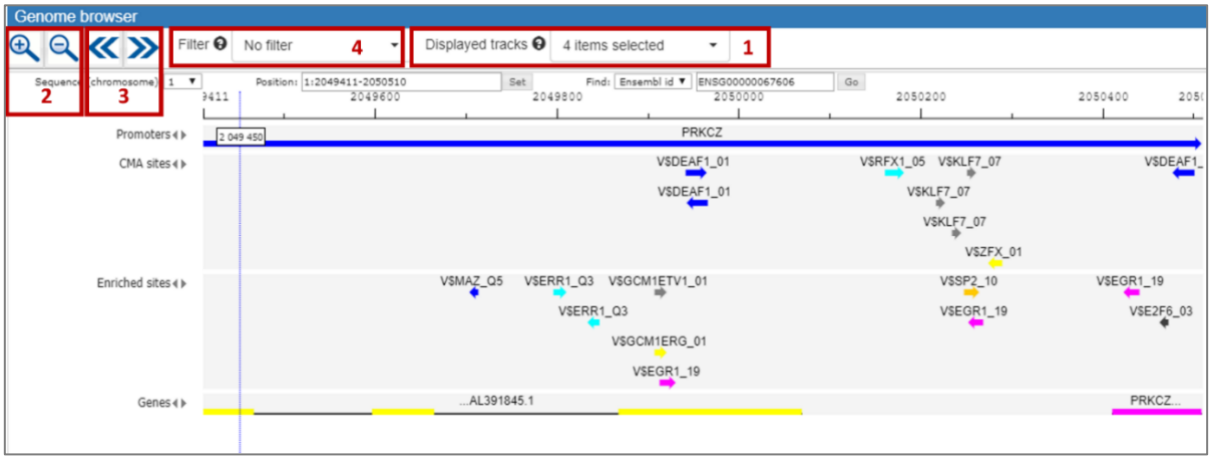


To visualize in genome browser the sites found in the promoter of any gene, just click on the line with the gene of your interest inside the genes table and refer to the genome browser below to explore the predicted regulation model of the respective gene.

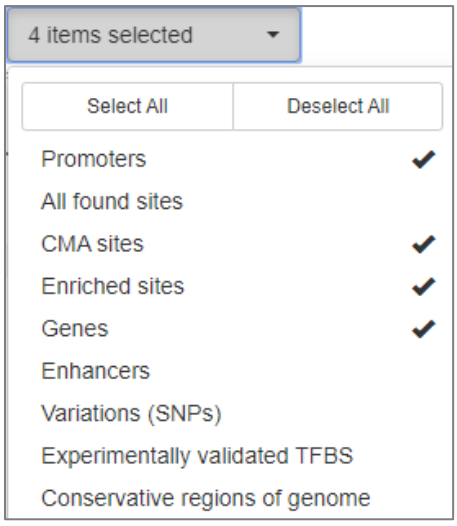
Genome Browser

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



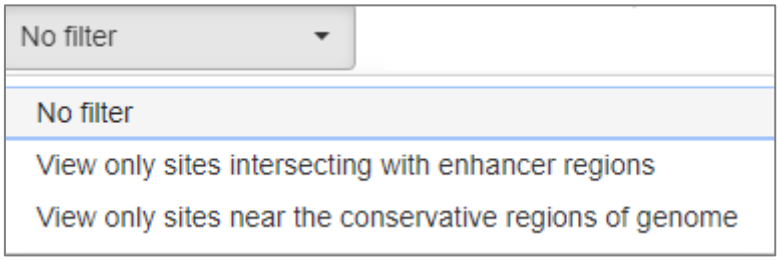
Displayed tracks can be customized using the **Displayed tracks** dropdown list (1)



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.

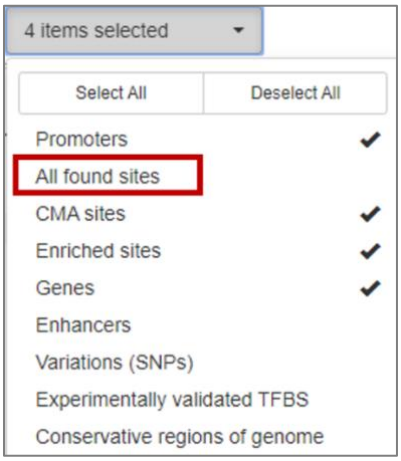
Navigation through the track is supported by the **Shift to the left** and **Shift to the right** buttons (3)

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



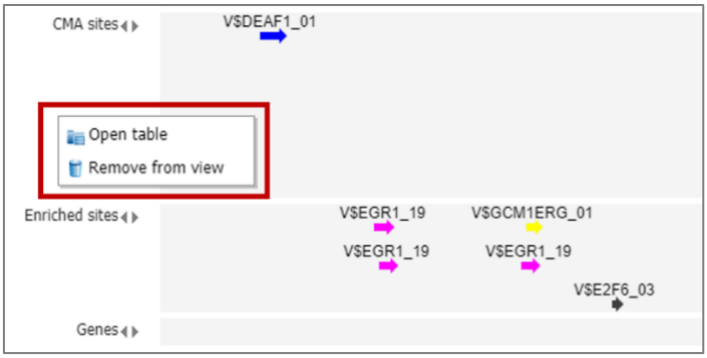
Depending on your selection, only sites intersecting with the conservative regions of the genome or those intersecting with the enhancer regions will be displayed. The genes table will be automatically recalculated leaving only the hits for the sites left after applying the respective filter. To cancel 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:

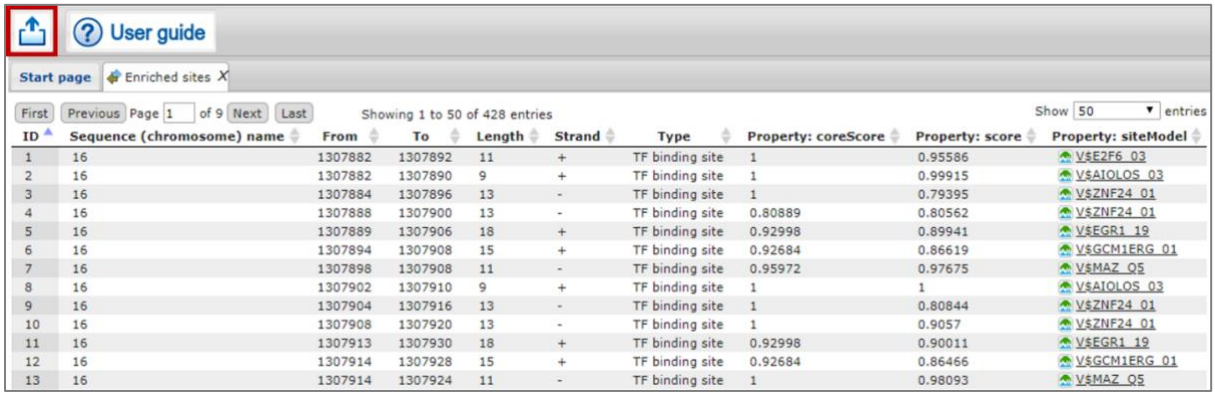


For receiving additional information about any of the found sites, you can click on the site of your interest and explore the contents displayed in the *Info box* section:

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):

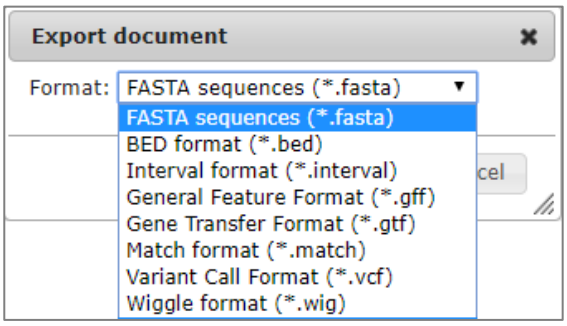


Tracks opened as table can be exported using the *Export* button at the top menu panel or used for further analysis from the geneXplain platform perspective.



ID	Sequence (chromosome) name	From	To	Length	Strand	Type	Property: coreScore	Property: score	Property: siteModel
1	16	1307882	1307892	11	+	TF binding site	1	0.95586	V$E2F6_03
2	16	1307882	1307890	9	+	TF binding site	1	0.99915	V$AIOLOS_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$MAZ_Q5
8	16	1307902	1307910	9	+	TF binding site	1	1	V$AIOLOS_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:



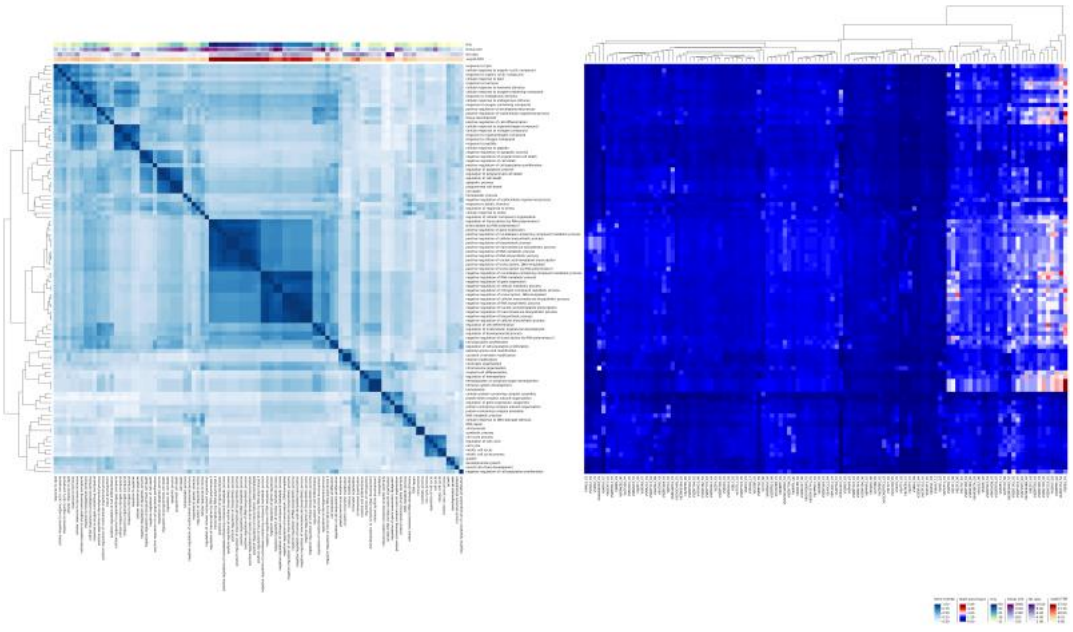
Heatmaps

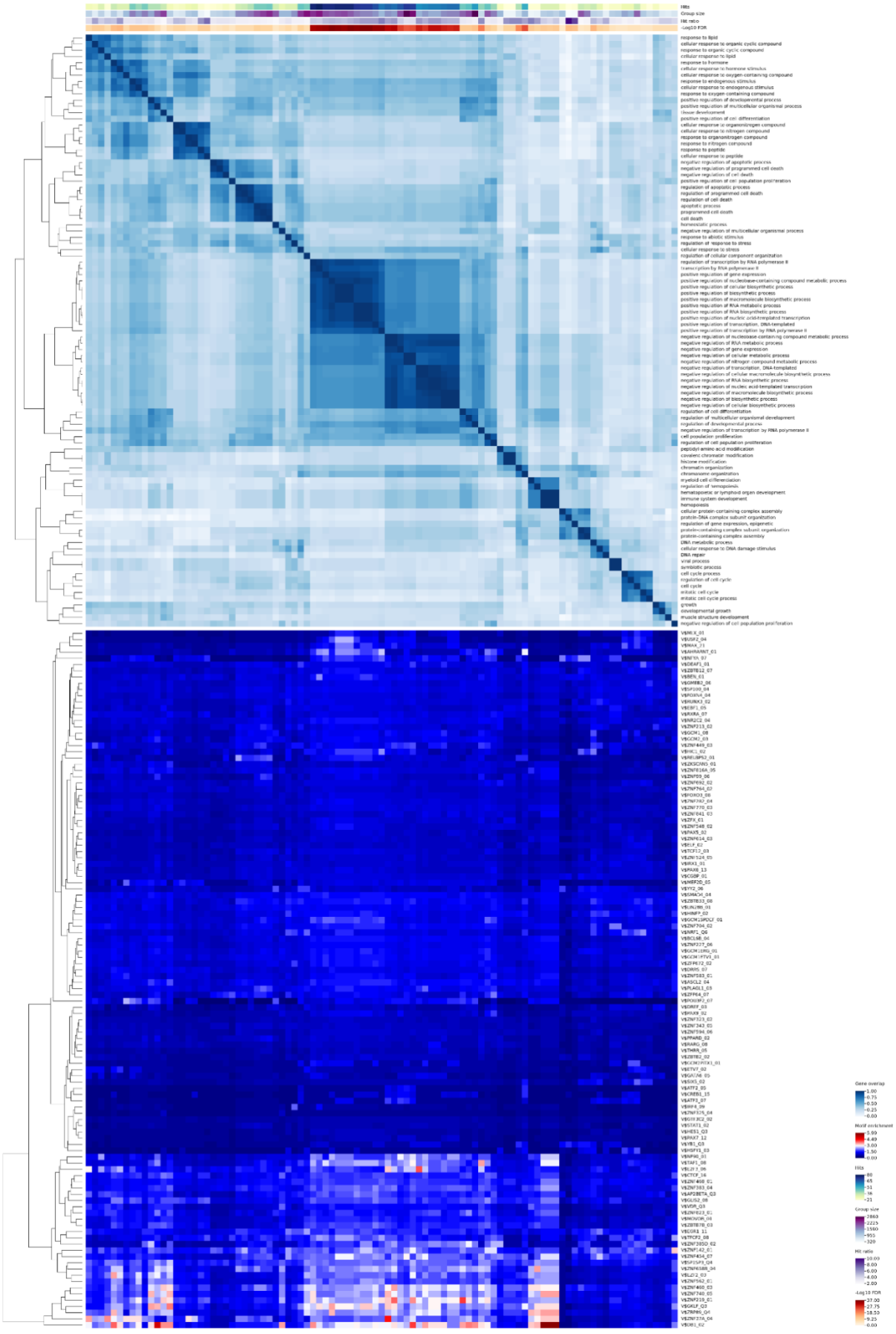
If the functional classification of your input gene set by GO terms produced reliable results (by p-values and by the number of genes in each functional category), heatmaps for the respective GO categorization will be included in the optional section *Functional Analysis of Gene Regulation* of the analysis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




Hits
Group size
Hit ratio
-Log10 FDR
response to lipid
cellular response to organic cyclic compound
response to organic cyclic compound
cellular response to lipid
response to hormone
cellular response to hormone stimulus
cellular response to oxygen-containing compound
response to endogenous stimulus
cellular response to endogenous stimulus
response to oxygen-containing compound
positive regulation of developmental process
positive regulation of multicellular organismal process
tissue development
positive regulation of cell differentiation
cellular response to organonitrogen compound
cellular response to nitrogen compound
response to organonitrogen compound
response to nitrogen compound
response to peptide
cellular response to peptide
negative regulation of apoptotic process
negative regulation of programmed cell death
negative regulation of cell death
positive regulation of cell population proliferation
regulation of apoptotic process
regulation of programmed cell death
regulation of cell death
apoptotic process
programmed cell death
cell death
homeostatic process
negative regulation of multicellular organismal process
response to abiotic stimulus
regulation of response to stress
cellular response to stress
regulation of cellular component organization
regulation of transcription by RNA polymerase II
transcription by RNA polymerase II
positive regulation of gene expression
positive regulation of nucleobase-containing compound metabolic process
positive regulation of cellular biosynthetic process
positive regulation of biosynthetic process
positive regulation of macromolecule biosynthetic process
positive regulation of RNA metabolic process
positive regulation of RNA biosynthetic process
positive regulation of nucleic acid-templated transcription
positive regulation of transcription, DNA-templated
positive regulation of transcription by RNA polymerase II
negative regulation of nucleobase-containing compound metabolic process
negative regulation of RNA metabolic process
negative regulation of gene expression
negative regulation of cellular metabolic process
negative regulation of nitrogen compound metabolic process
negative regulation of transcription, DNA-templated
negative regulation of cellular macromolecule biosynthetic process
negative regulation of RNA biosynthetic process
negative regulation of nucleic acid-templated transcription
negative regulation of macromolecule biosynthetic process
negative regulation of biosynthetic process
negative regulation of cellular biosynthetic process
regulation of cell differentiation
regulation of multicellular organismal development
regulation of developmental process
negative regulation of transcription by RNA polymerase II
cell population proliferation
regulation of cell population proliferation
peptidyl-amino acid modification
covalent chromatin modification
histone modification
chromatin organization
chromosome organization
myeloid cell differentiation
regulation of hemopoiesis
hematopoietic or lymphoid organ development
immune system development
hemopoiesis
cellular protein-containing complex assembly
protein-DNA complex subunit organization
regulation of gene expression, epigenetic
protein-containing complex subunit organization
protein-containing complex assembly
DNA metabolic process
cellular response to DNA damage stimulus
DNA repair
viral process
symbiotic process
cell cycle process
regulation of cell cycle
cell cycle
mitotic cell cycle
mitotic cell cycle process
growth
developmental growth
muscle structure development
negative regulation of cell population proliferation
Gene overlap
Motif enrichment
Hits
Group size
Hit ratio
-Log10 FDR

TRANSFAC 2.0

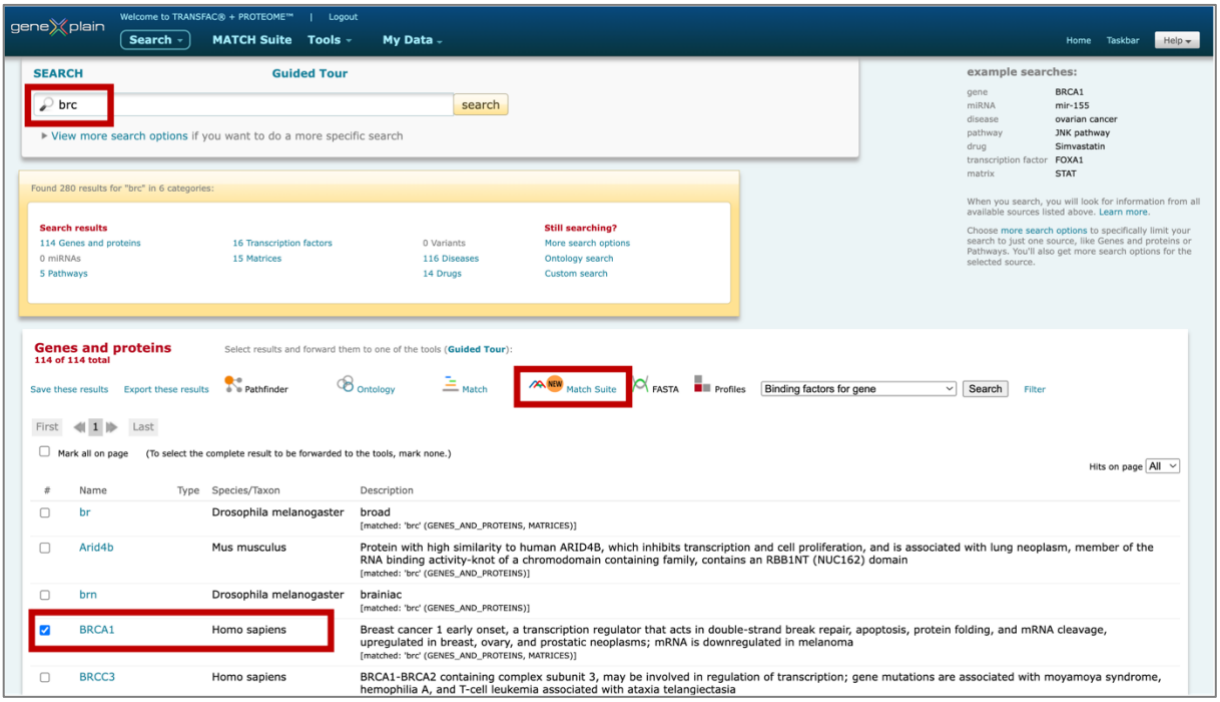
Single gene analysis

Launching the analysis

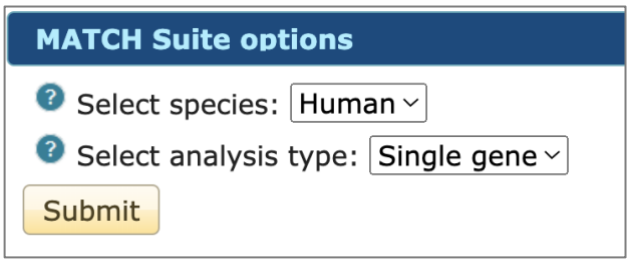
Specifying the input gene

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:

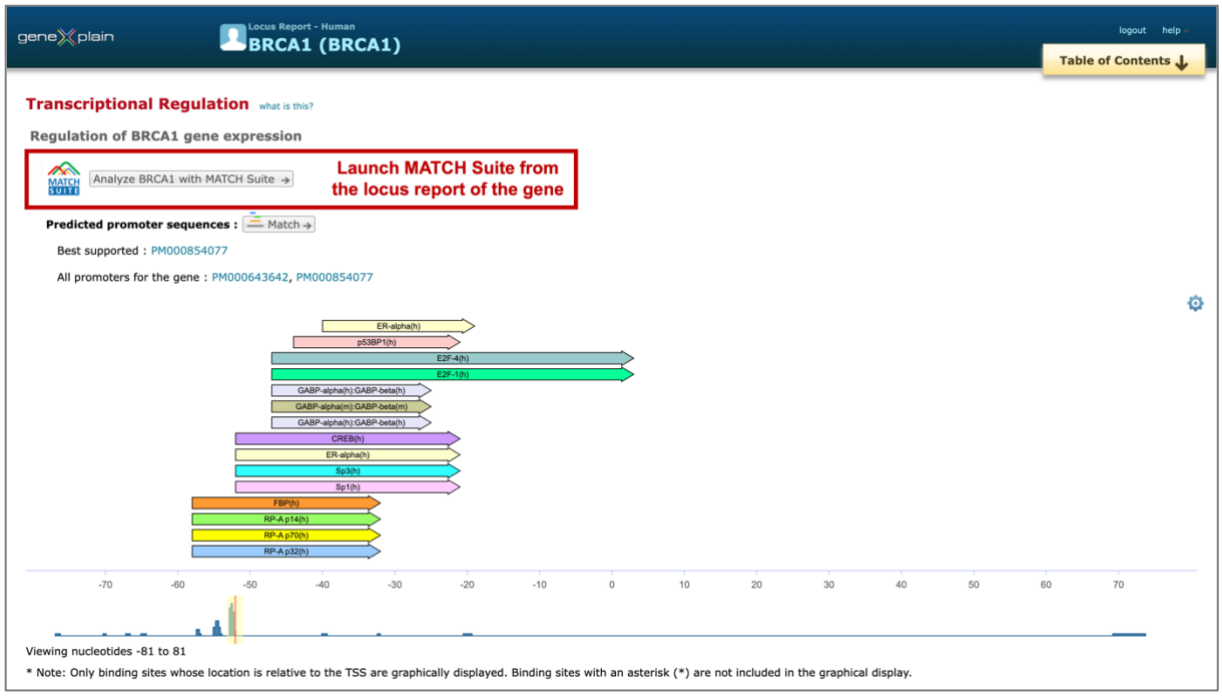


Once you have clicked on the Match Suite icon, indicate that you aim to perform human single gene analysis:



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

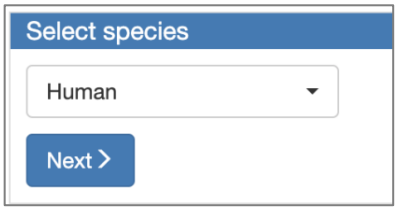
Alternatively, you can start the MATCH Suite single gene analysis from a locus report of the gene of your interest:



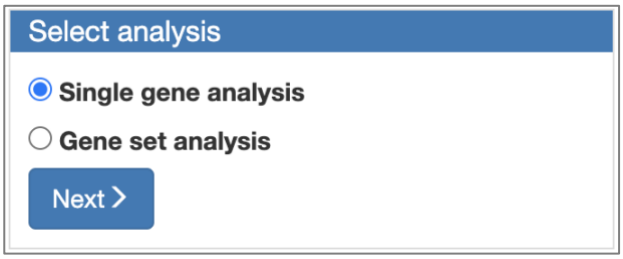
When you have selected the gene of your interest, press the button "Match Suite single gene" in the search results or click on the "Analyze <gene name> with MATCH Suite" button in the *Transcriptional regulation* section of the locus report of the selected gene, and you will be automatically transferred 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

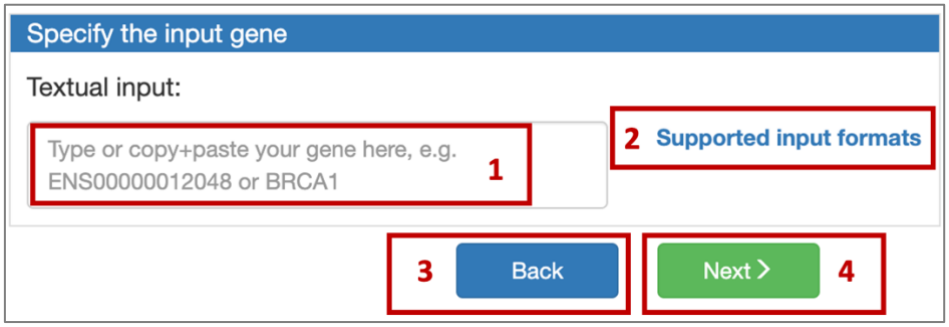
When you click on the *Start analysis* section, the system immediately navigates you to the analysis launch wizard. At the very first step of this wizard, you will be asked to select the species you are working with:



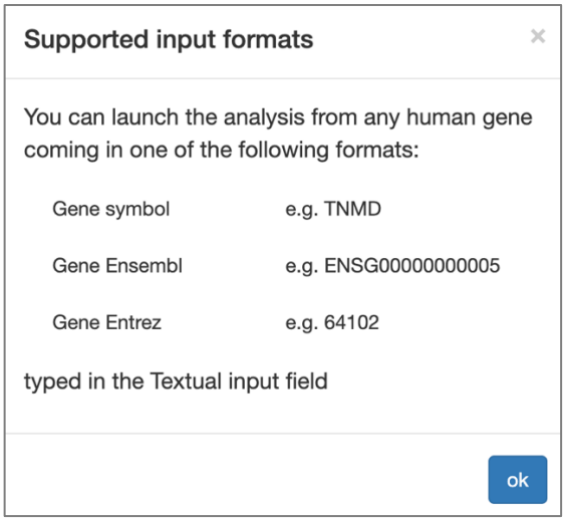
And type of the analysis you would like to perform:



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:



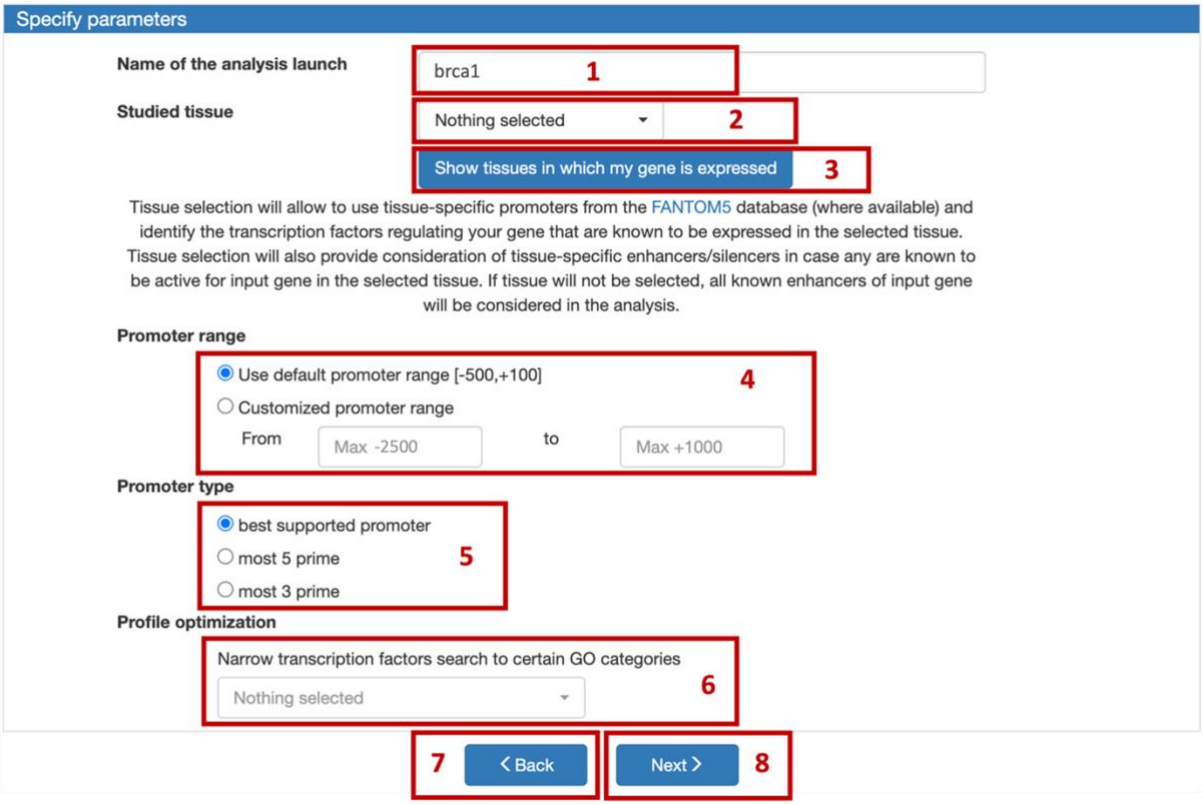
You can launch the analysis by typing any gene symbol, Entrez gene ID or Ensembl gene ID into the **textual input field** (1). The **Supported input formats** (2) will assist you in providing the input gene in the correct format:



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

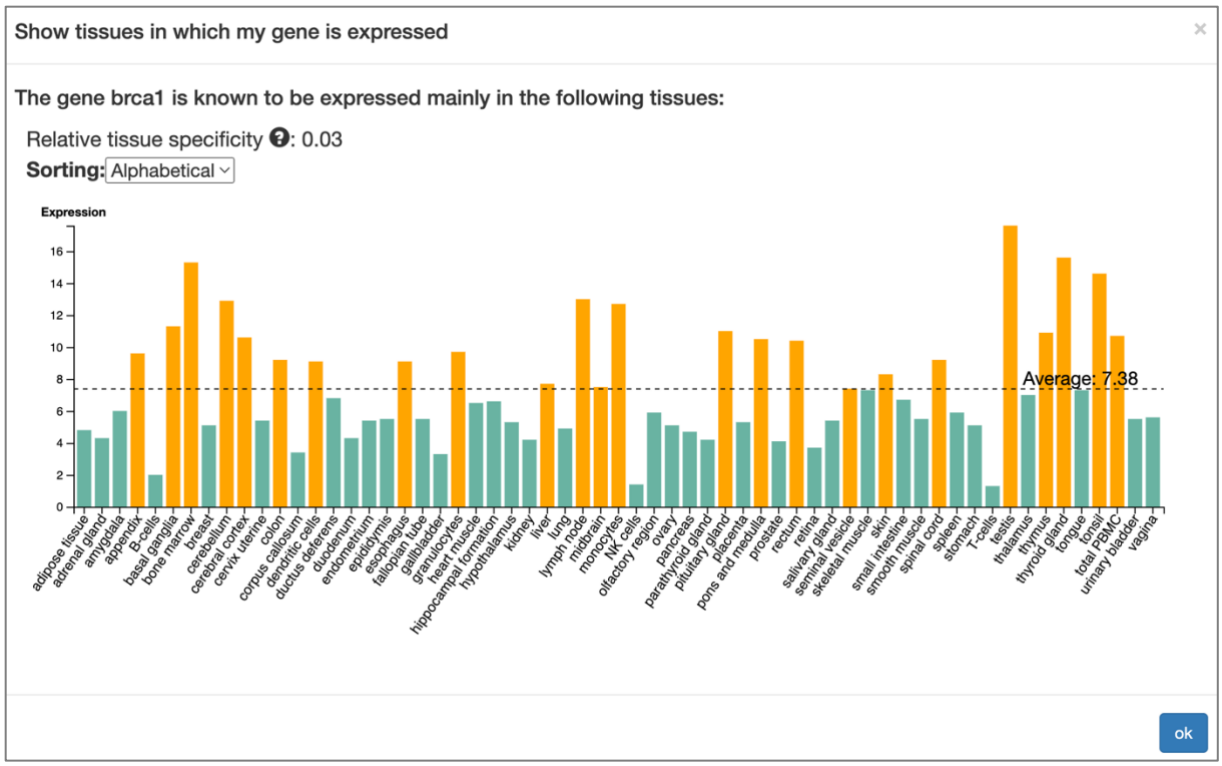
In the next step of the analysis launch wizard the MATCH Suite will ask you to fill in the following form:



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:



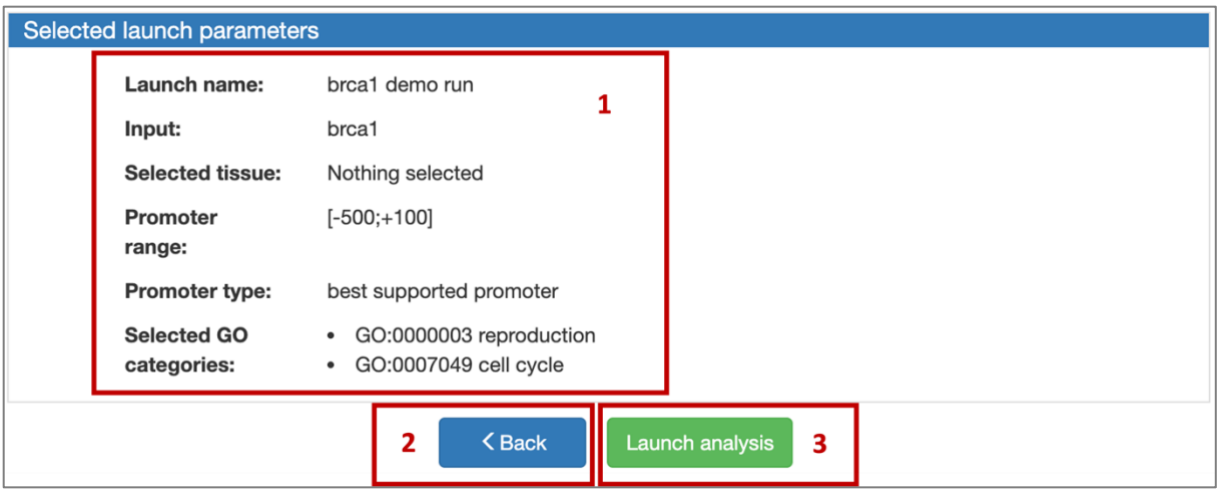
The promoter range that will be used for the search of TFBS in your analysis run is specified by the **Promoter range** parameter (4). By default the promoter range used by the MATCH Suite is [-500,100] relative to the TSS. You can specify a customized promoter range within the interval of -2500 and +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:

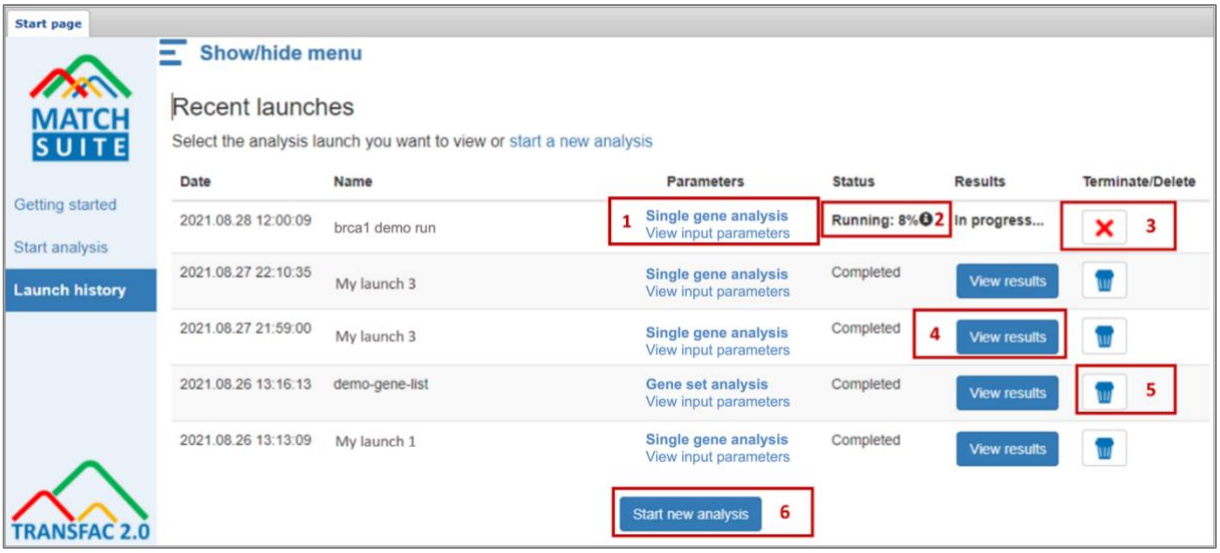


All parameters that you selected for the current launch will be shown on the screen (1). If you want to apply any changes to the specified parameters of the launch, click on **Back** (2), otherwise you are ready to start your analysis by clicking on the **Launch analysis** button (3).

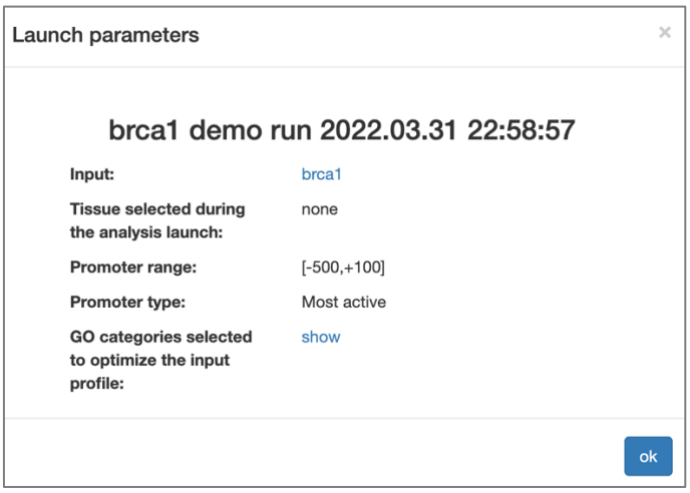
Viewing the results

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.



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:



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.

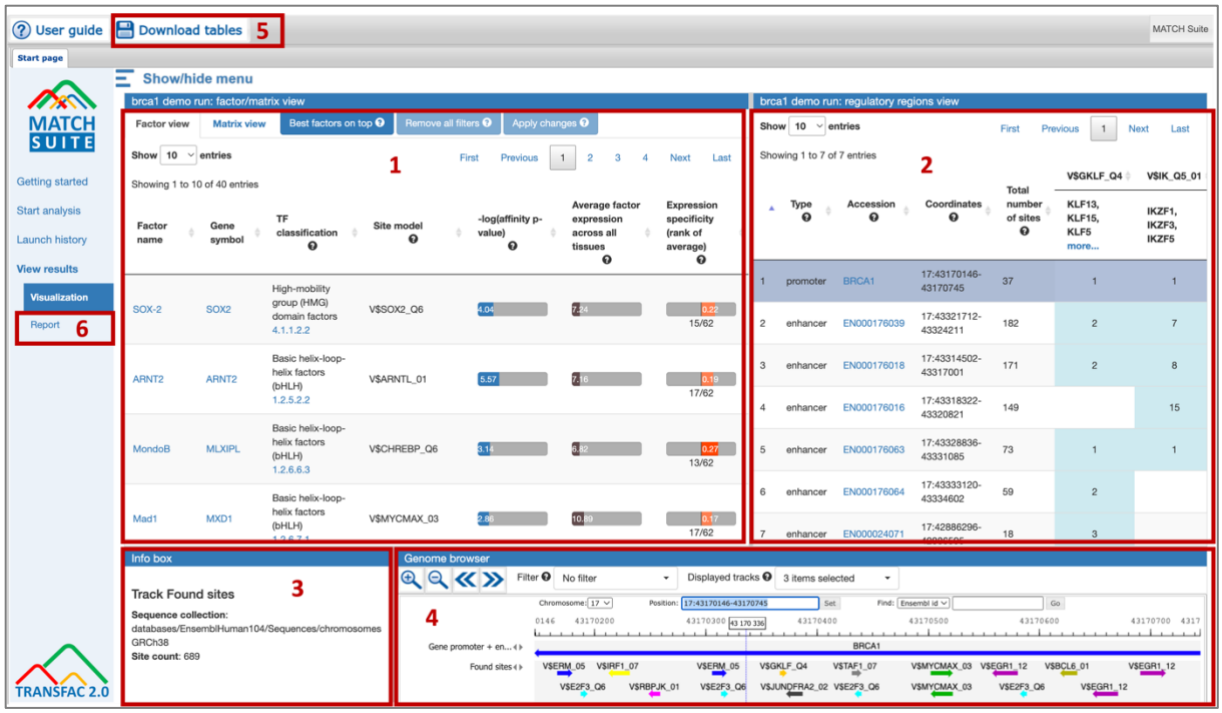
In the results visualization mode the screen will be divided into four different segments with the following functions:

(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

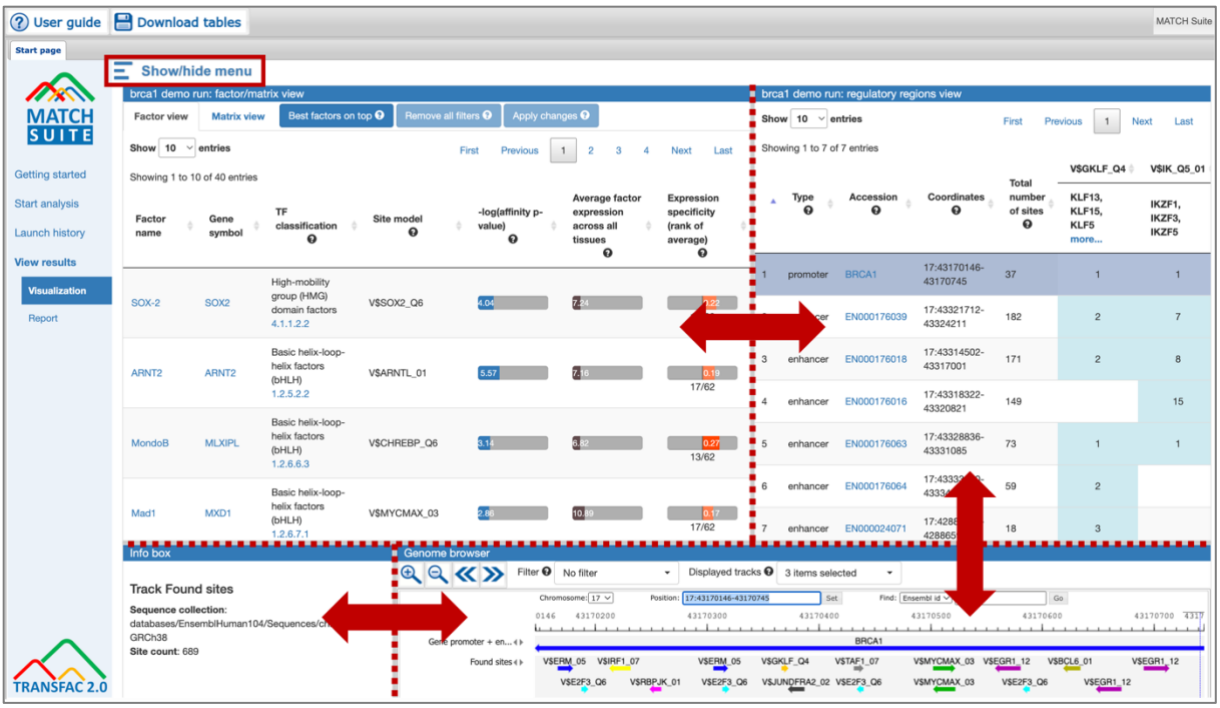
(4) The **genome browser**, allowing to visualize the tracks of found sites and additional annotation tracks for further results interpretation



The 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.

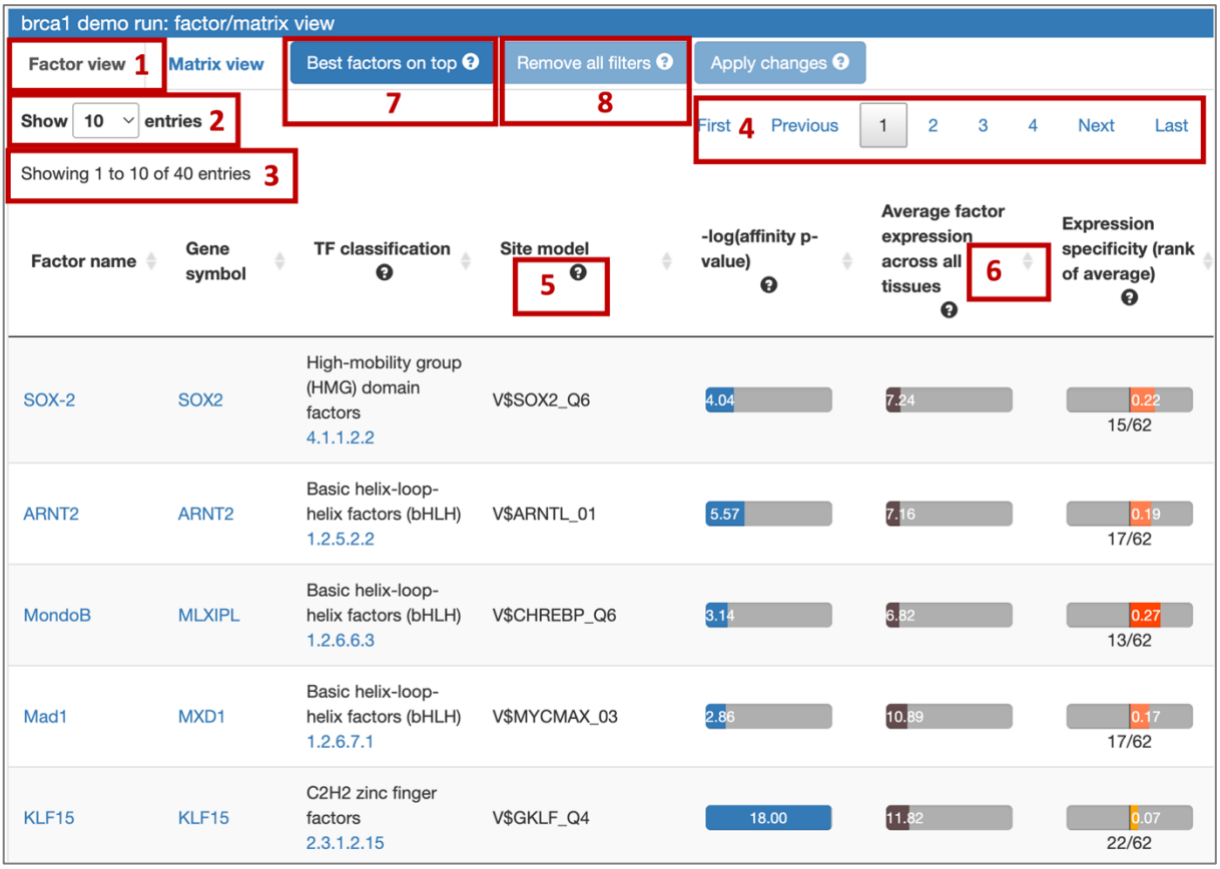


In the screen segment, which visualizes the tables of predicted factors and respective matrices, you will find two accessible tables: Factor view and Matrix view.

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.

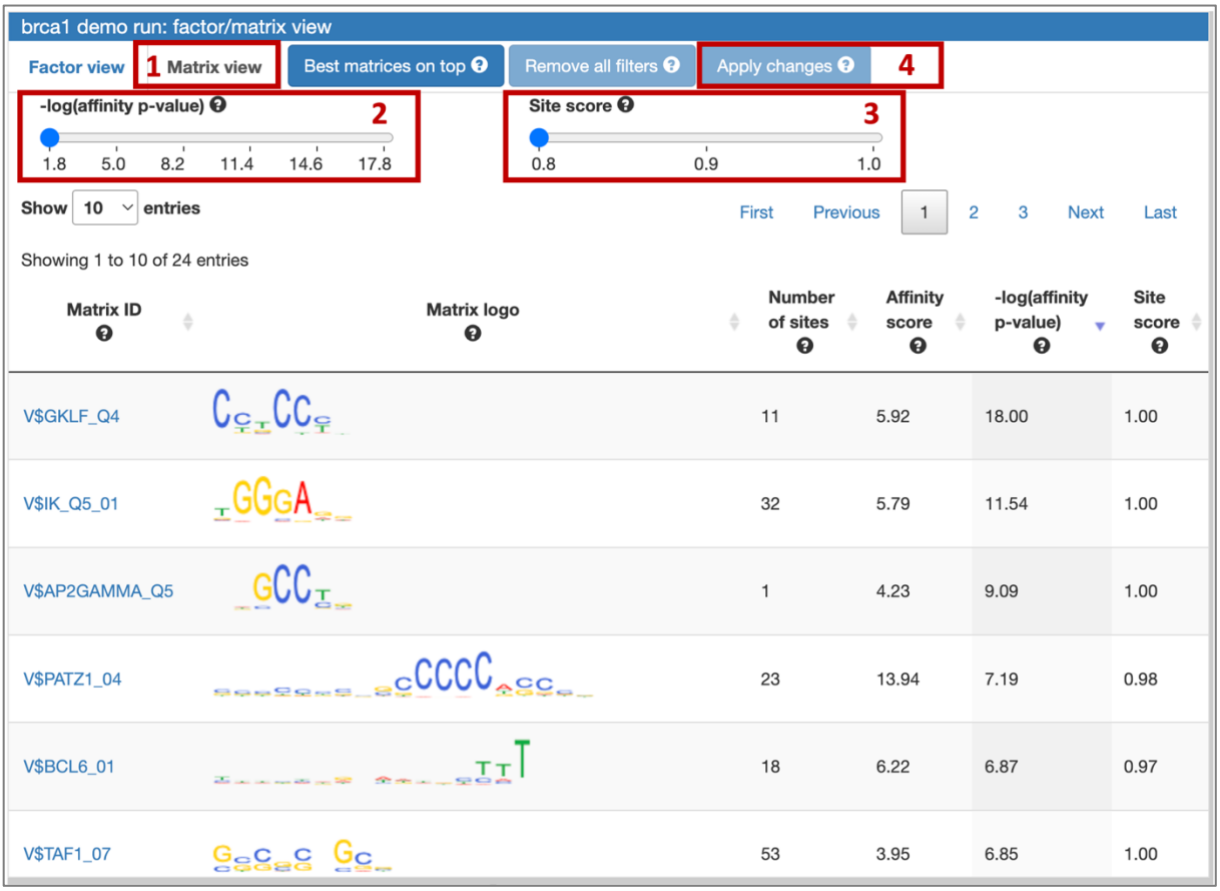


The columns of tables have in-built sorting option (6) which allows to sort the values within one column by ascending or descending order or by alphabetical order in case of textual contents. Simply click on the gray arrows for the sorting to be applied. By default, best factors (or matrices) are brought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



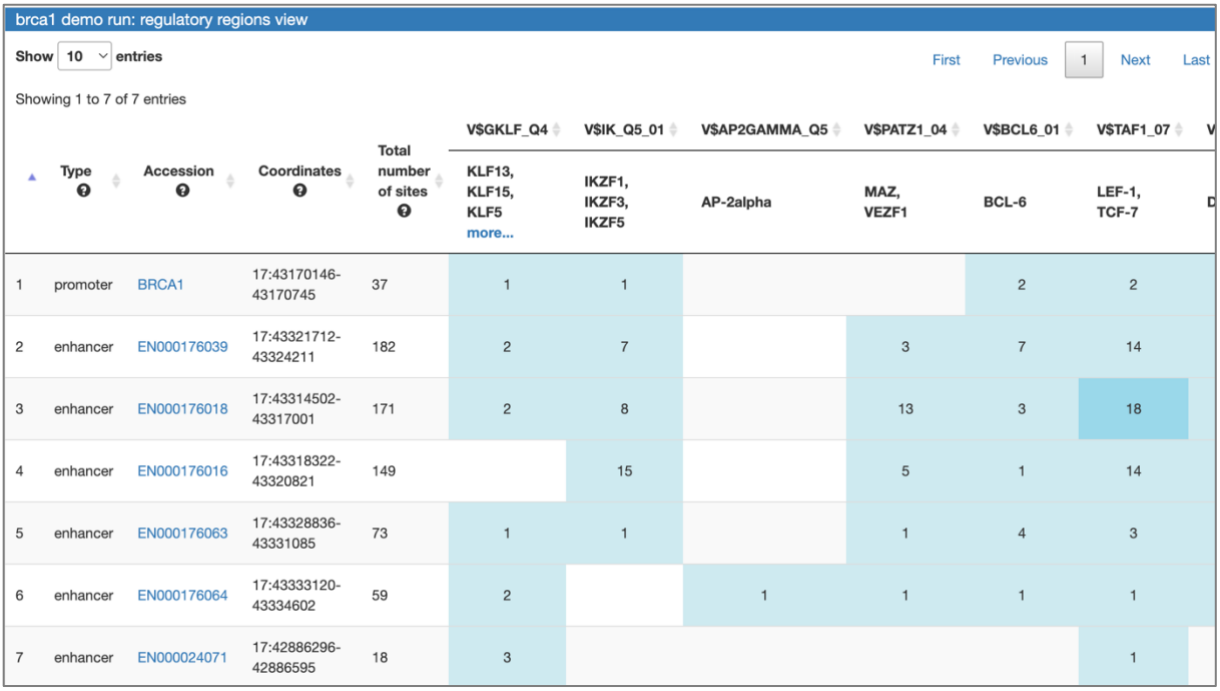
The **-log(affinity p-value) filter** (2) allows to leave only matrices with -log(affinity p-value) values higher than the threshold specified by the filter. The set filter will be first applied exclusively to the *Matrix view* table. To recalculate the factors and the regulatory regions tables, as well 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 — First Previous 1 Next Last

Showing 1 to 7 of 7 entries

	Type	Accession	Coordinates	Total number of sites	V$GKLF_Q4 KLF13, KLF15, KLF5 more...	V$IK_Q5_01 IKZF1, IKZF3, IKZF5	V$AP2GAMMA_Q5 AP-2alpha	V$PATZ1_04 MAZ, VEZF1	V$BCL6_01 BCL-6	V$TAF1_07 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

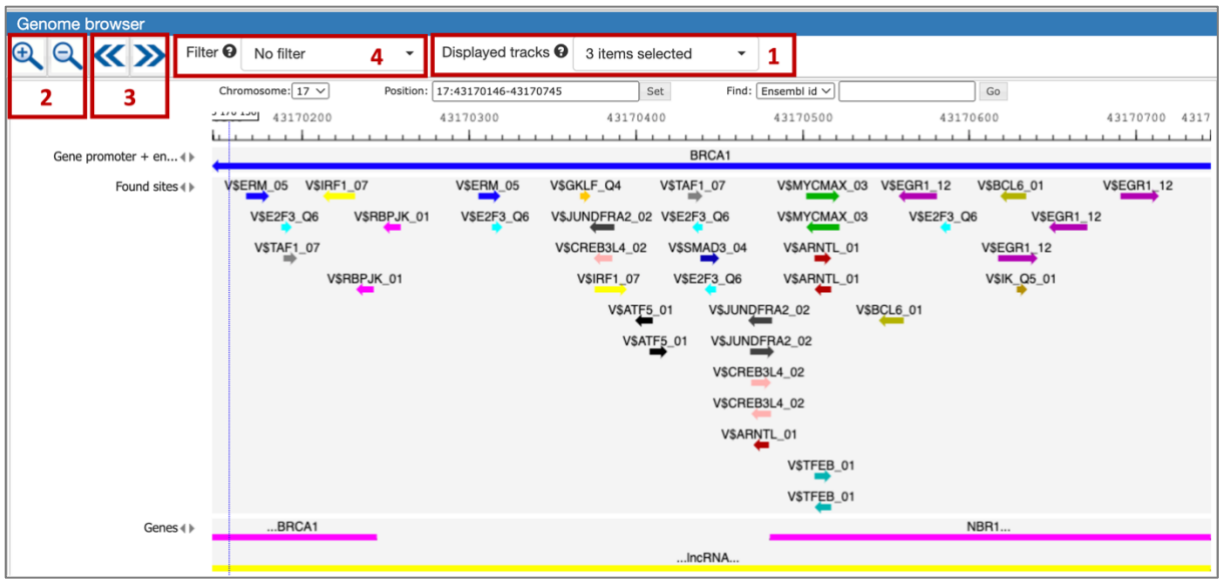
To visualize in genome browser the sites found in the certain regulatory region (promoter / enhancer / silencer), just click on the line with the regulatory region of your interest inside the regulatory regions table and refer to the genome browser below to explore the predicted regulation model of the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

Genome Browser

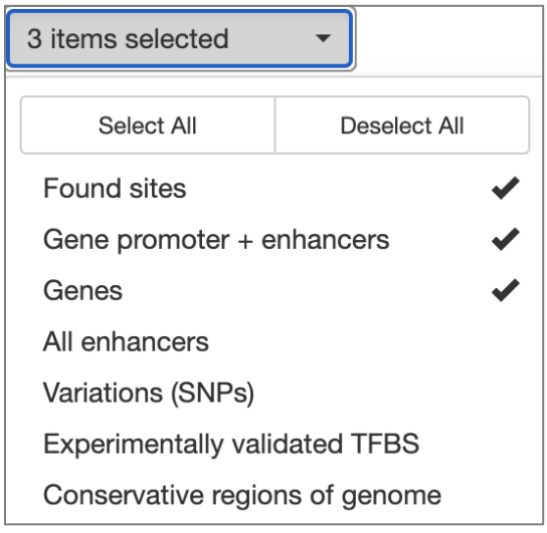
The genome browser provides you with visualizations of the predicted regulation models for each of the studied regulatory regions of the input. Having clicked on any regulatory region from the *Regulatory regions table*, the model of the

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)

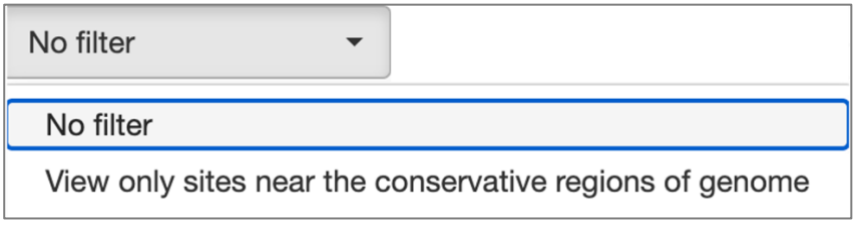


You can zoom in and zoom out for a more generalized or a more detailed view of the regulatory region model using the **zoom in** and **zoom out** buttons (2). In case navigation in genome browser is lost due to multiple zoom clicks, you can always

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)

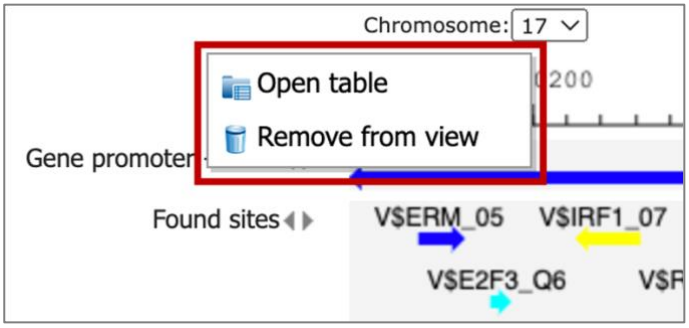
Filtering can be applied to the displayed sites using the **Filter** option (4):



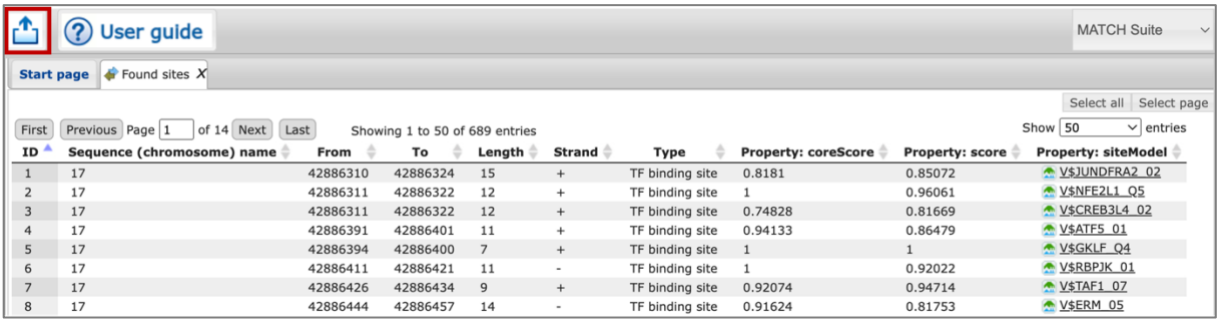
Depending on your selection, only sites intersecting with the conservative regions of the genome can be displayed. The regulatory regions table will be automatically recalculated leaving only the hits for the sites left after applying the respective filter. To cancel the applied filter, select the *No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

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):

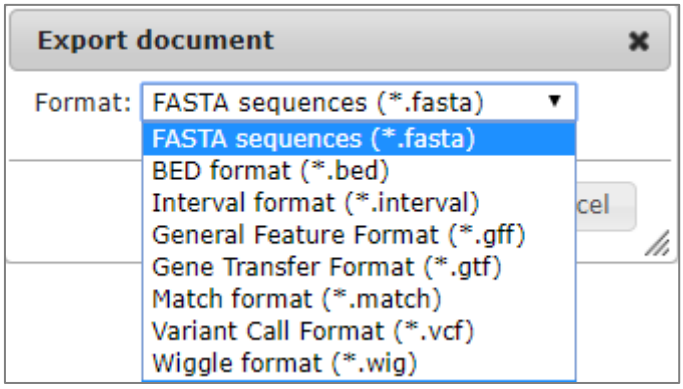


Tracks opened as table can be exported using the *Export* button at the top menu panel or used for further analysis from the geneXplain platform perspective.



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$TAF1_07
8	17	42886444	42886457	14	-	TF binding site	0.91624	0.81753	V$ERM_05

The list of available export formats is as follows:



GO categorization-based gene set analysis

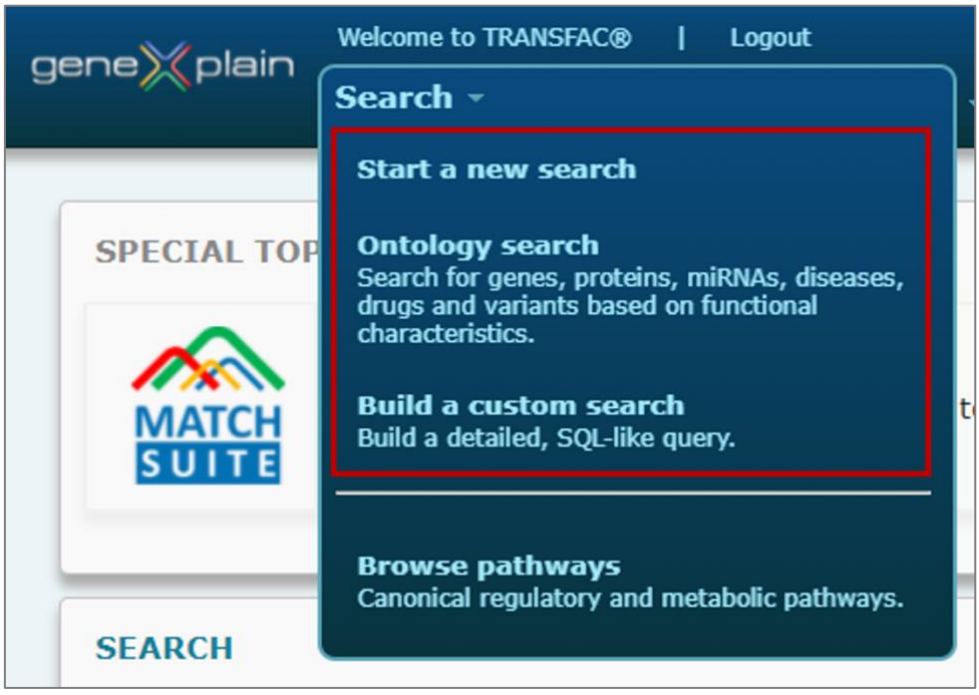
Please note that GO categorization-based analysis is available for sets of Human, Mouse, and Rat genes.

Launching the analysis

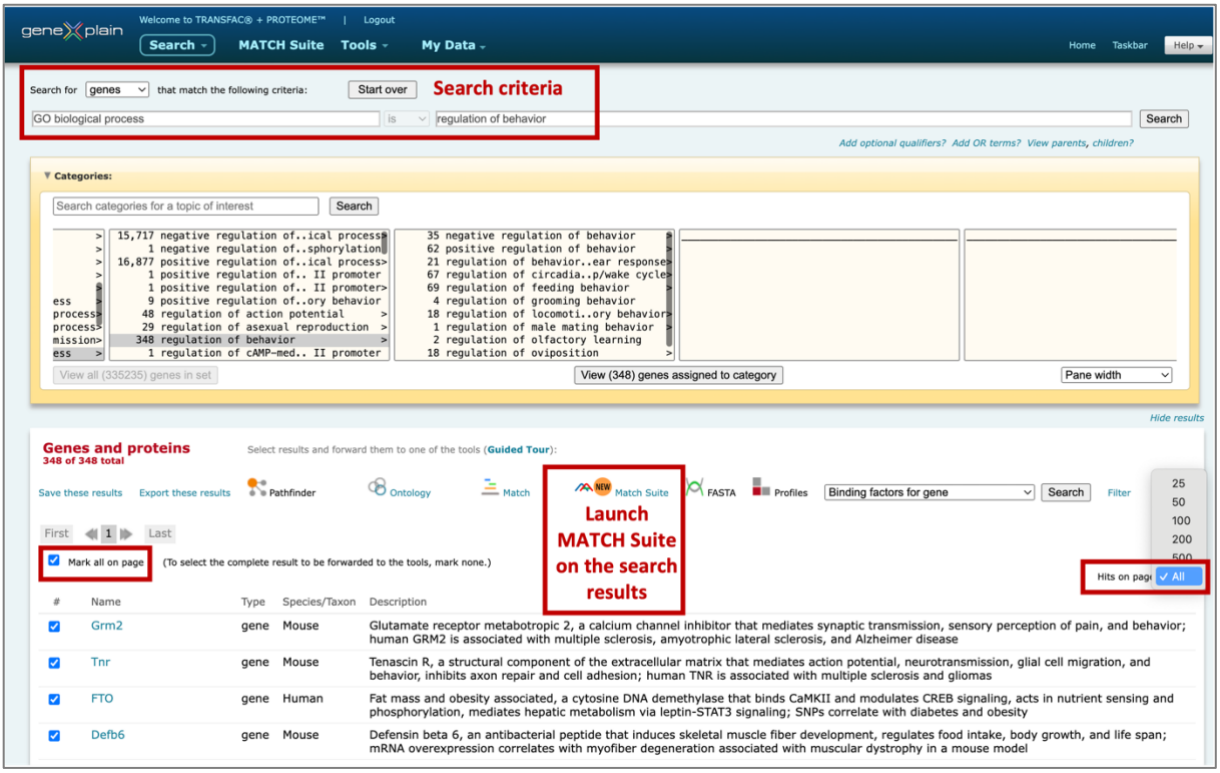
Selecting the input gene list

Composing an input gene list in the TRANSFAC® database

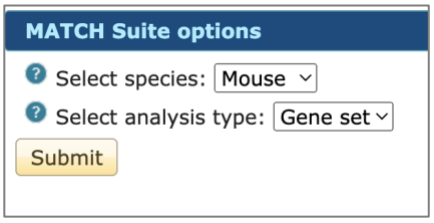
A particularly convenient way to compose a gene list for analysis is to run the MATCH Suite on your search results in the TRANSFAC® database. For instance, when you search for a certain disease in the standard Search field or select genes belonging to a certain Gene Ontology (GO) category (Search > Ontology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



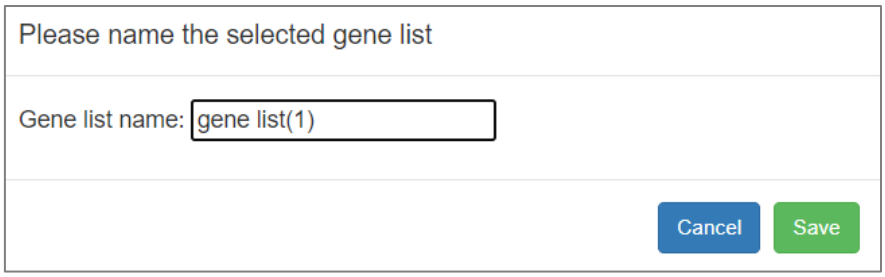
In the search results inside TRANSFAC® database you can view all found entries on one page (select "All" in the "Hits on page" list) and then select the search results of your interest to launch the MATCH Suite gene set analysis on a respective gene list (you can use the *Mark all on page* option and 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).



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, or Rat):



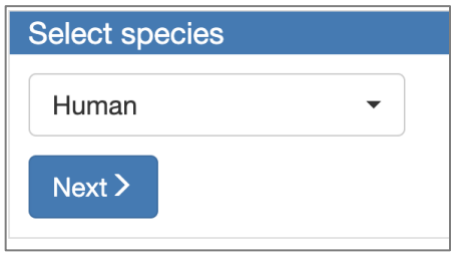
and you will be automatically transferred to the MATCH Suite tool, where you will be asked to assign a name to your gene list prior to proceeding to the next step of the analysis launch wizard:



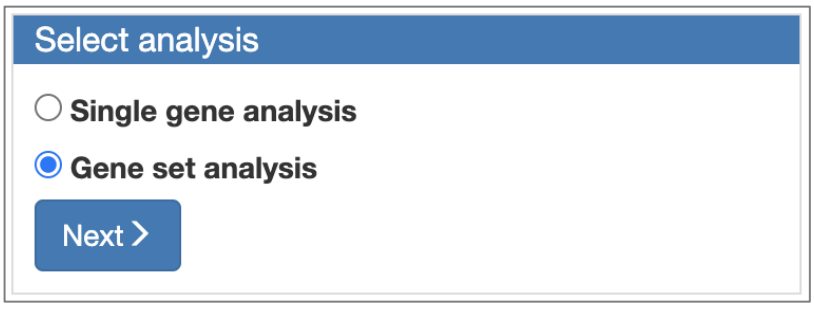
Give the name to your gene list and click on *Save* to proceed to the next step of the wizard.

Selecting input gene list from the MATCH Suite interface

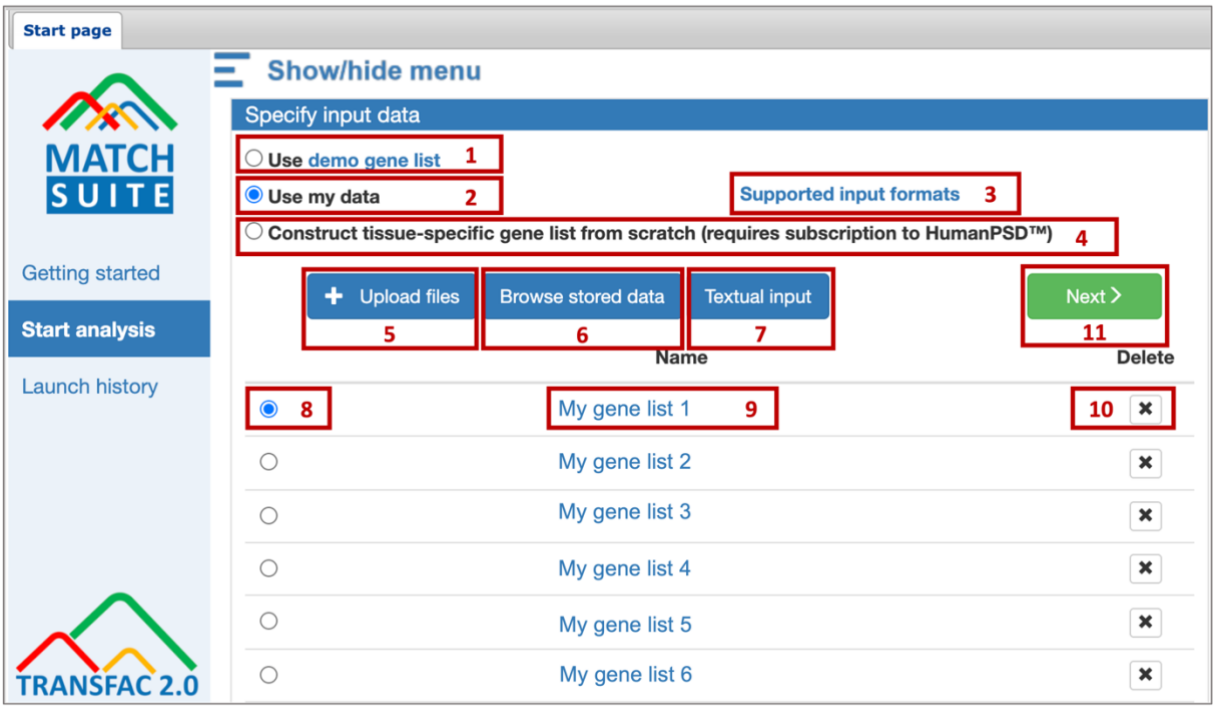
When you click on the *Start analysis* section, the system immediately navigates you to the analysis launch wizard. At the very first step of this wizard you will be asked to select the type of the model organism you want to work with:



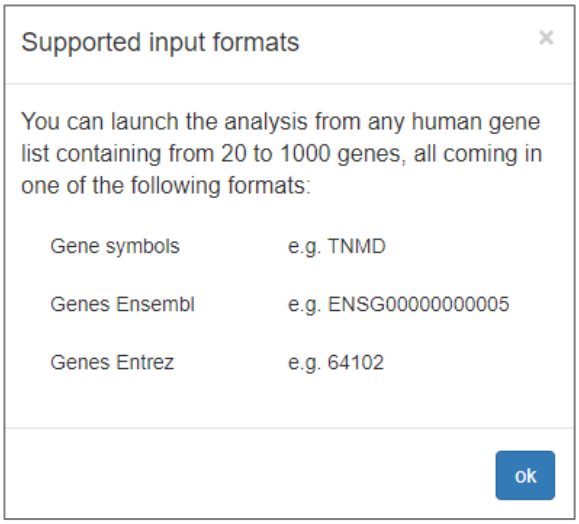
Human, Mouse, and Rat species are supported. If you will select Human, at the next step of the wizard you should select the *Gene set analysis*:



In case of Mouse or Rat gene set analysis, you will immediately be taken to the next step of the input gene list specification:

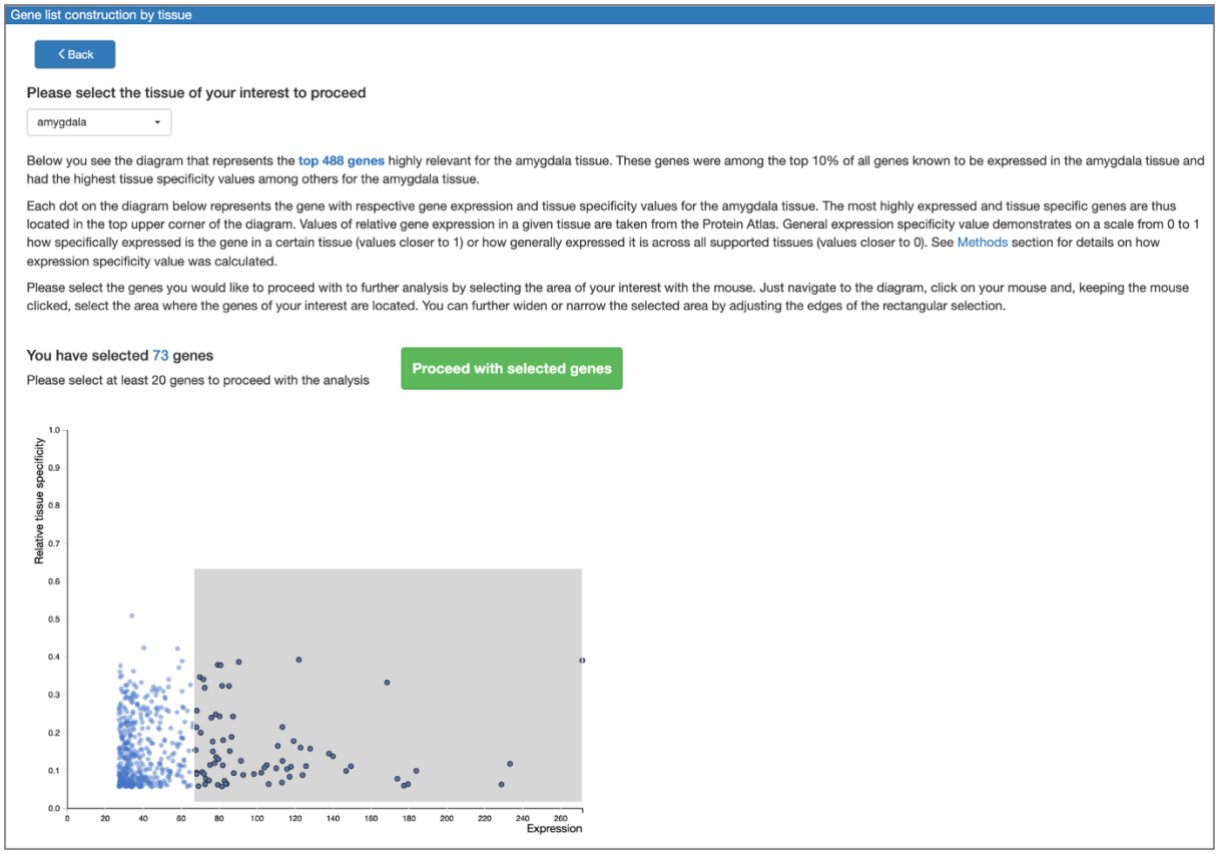


You can try launching the analysis using the **Use demo gene list** (1) option, or you can specify your own gene list by selecting the **Use my data** (2). The demo gene list can be viewed by clicking on the *demo gene list* blue link, it will contain Human, Mouse, or Rat genes, depending on the type 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 3.0:



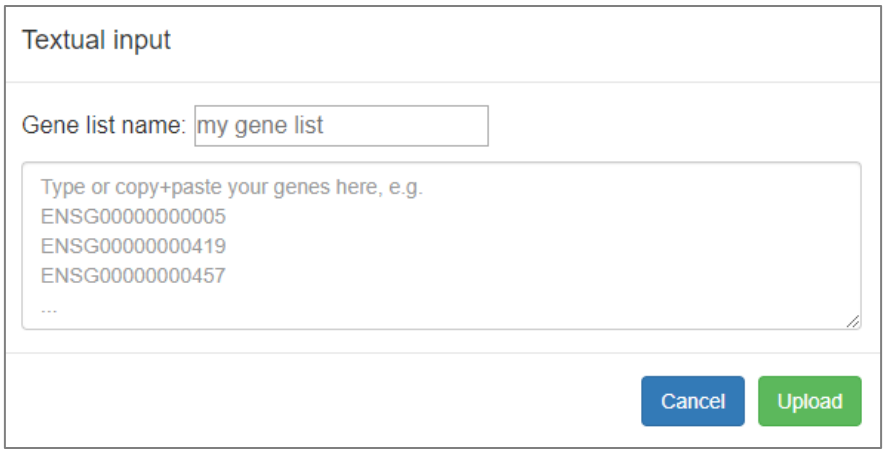
Lists of 20 to 2000 genes coming in Ensembl ID, Entrez ID or Gene symbols format are accepted by the MATCH Suite 3.0. The system will select automatically the Human, Mouse, or Rat genes from your input gene list, based on the selected type of the model organism, and only those genes will be subject to 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)

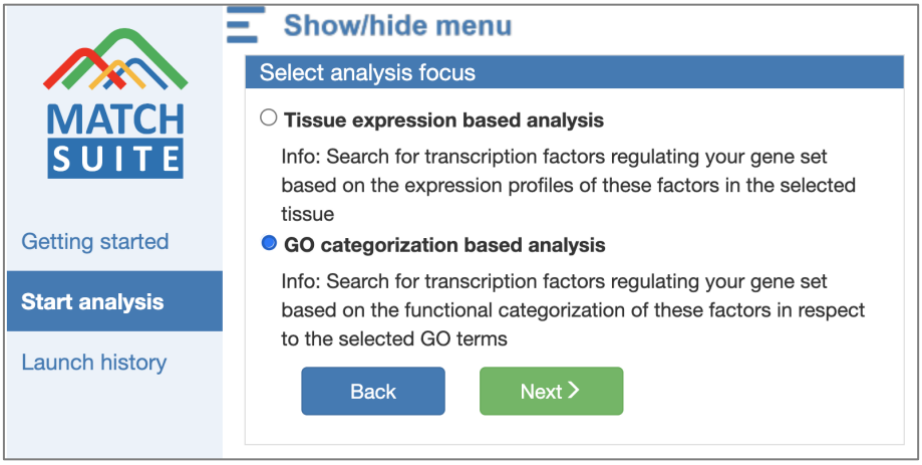
To select the gene list for the current analysis run, mark it with the radio button in the list of all currently available inputs (8) (any newly uploaded gene list will automatically appear in this list).

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, or Rat 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.

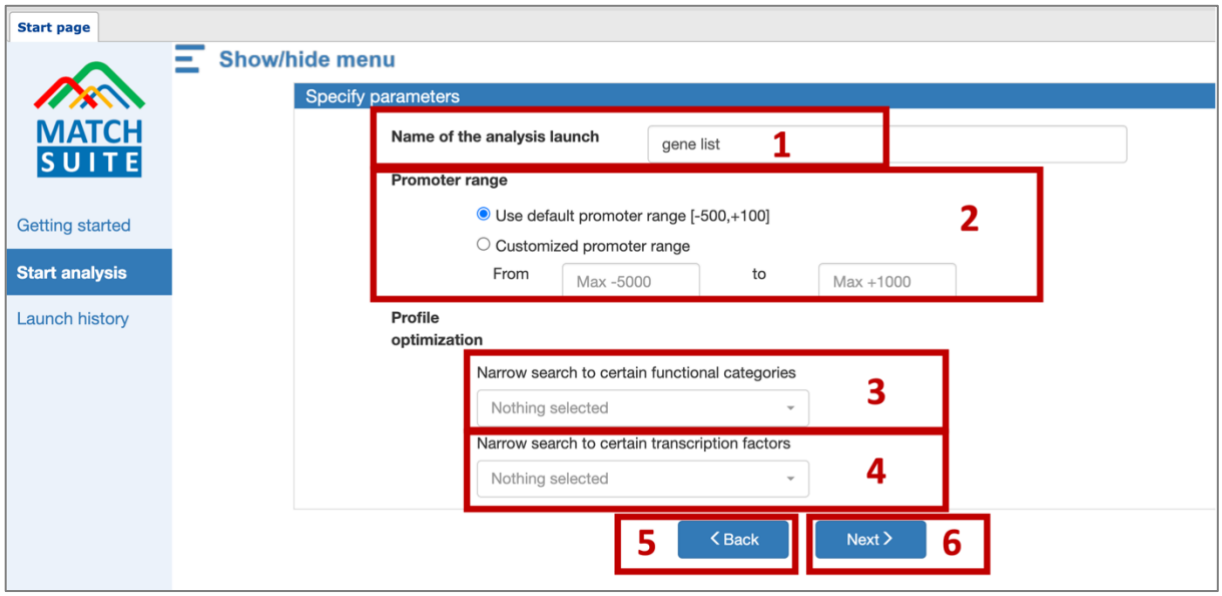
Specifying the launch parameters

After giving a name to your gene list, the wizard will take you to the next step: in case of Rat or Mouse 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:



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:



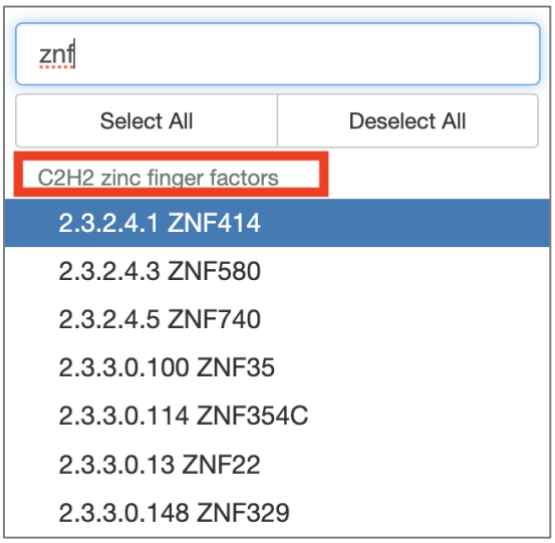
You can give a name to your analysis launch or keep the default name suggested by the system in the **Name of the analysis launch** field (1). This name will be further used in the launches history for easy selection of the run of your interest.

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.

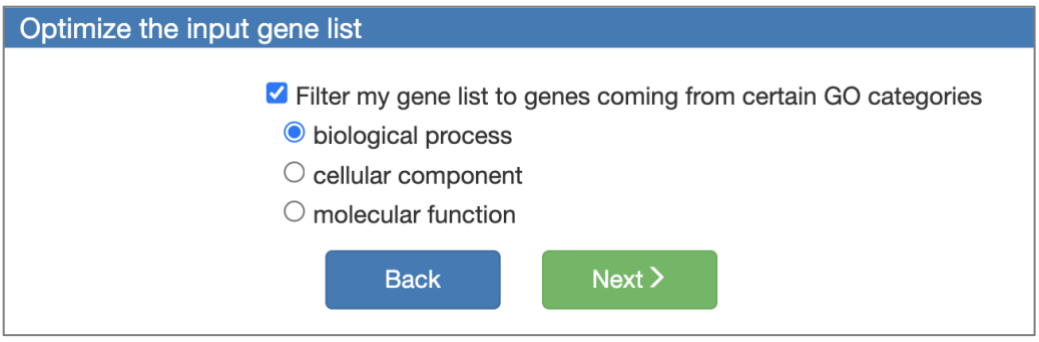
When selecting the TF, you can use the textual search. Tap on the factors of your interest to select them, or click on the name of the whole family (highlighted with red on the screenshot below) to select all factors of the respective family:



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:

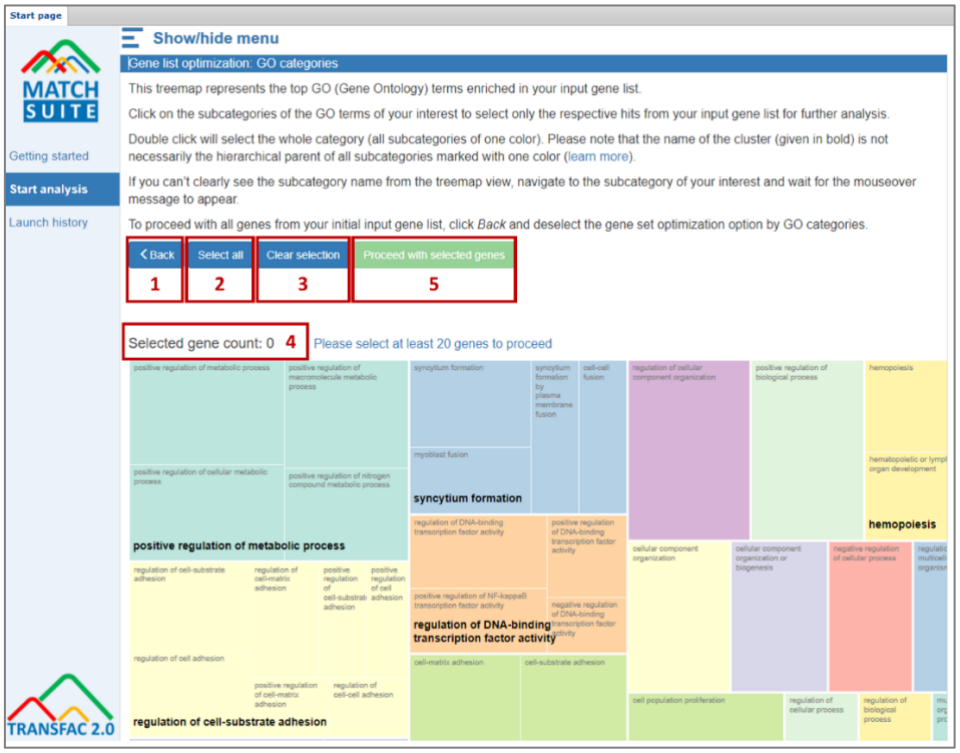


Depending on your selection, the tree map visualization of functional classification of your input gene set will be built either on biological processes, or on cellular components, or on molecular functions.

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.

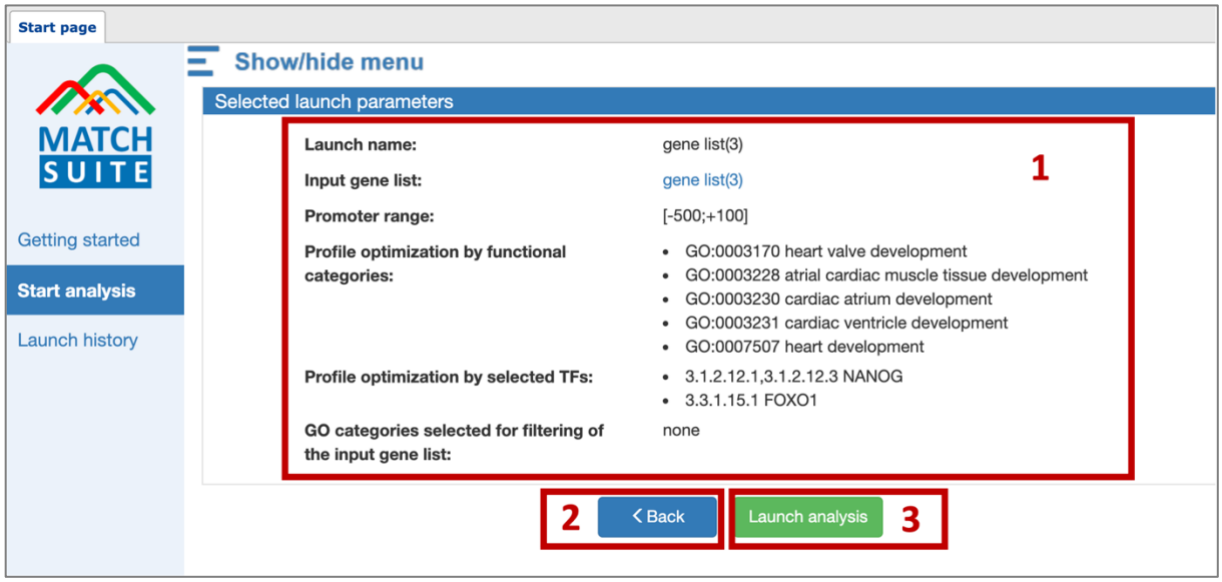


You can click **Back** (1) to return to the parameters specification step and cancel the GO optimization of your input gene list by unticking the *Filter my gene list to genes coming from certain GO categories*. Otherwise, you are requested to select the GO categories of your interest, following the 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:



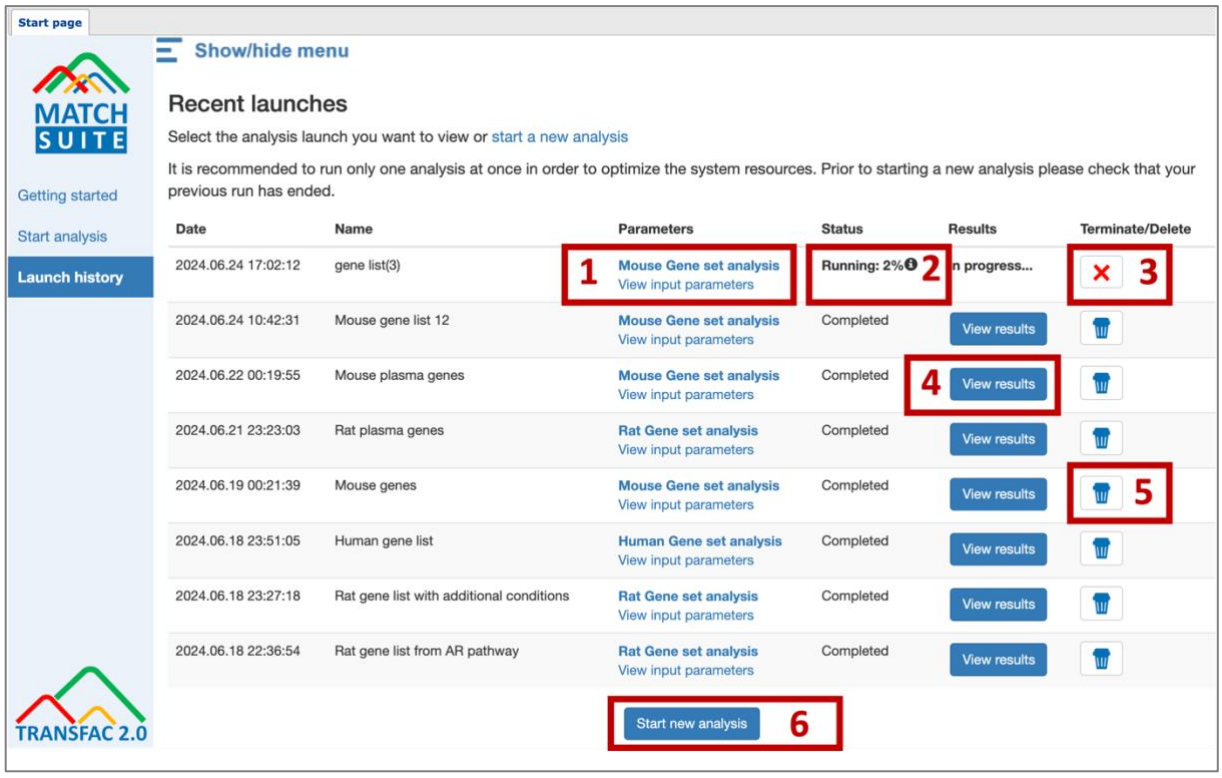
All parameters that you selected for the current launch will be shown on the screen (1). You can check the genes that were eventually selected for the analysis launch after GO functional classification (if it was applied) by clicking on the name of the input gene list.

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).

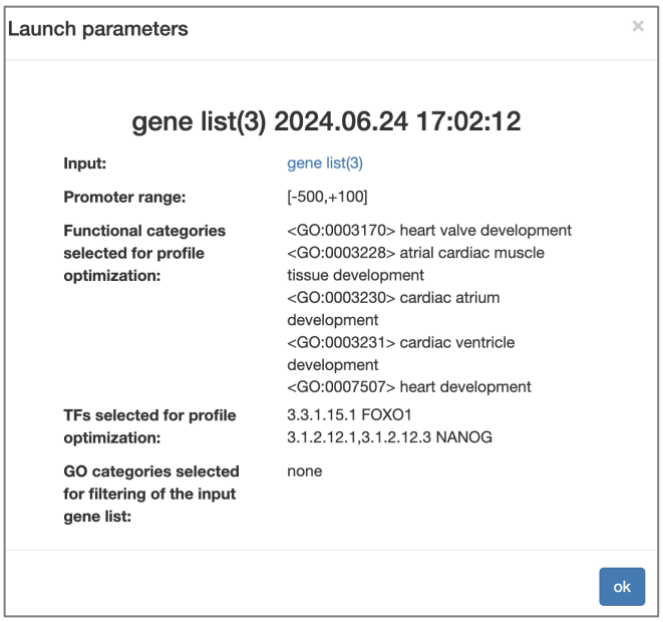
Viewing the results

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.



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:



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.

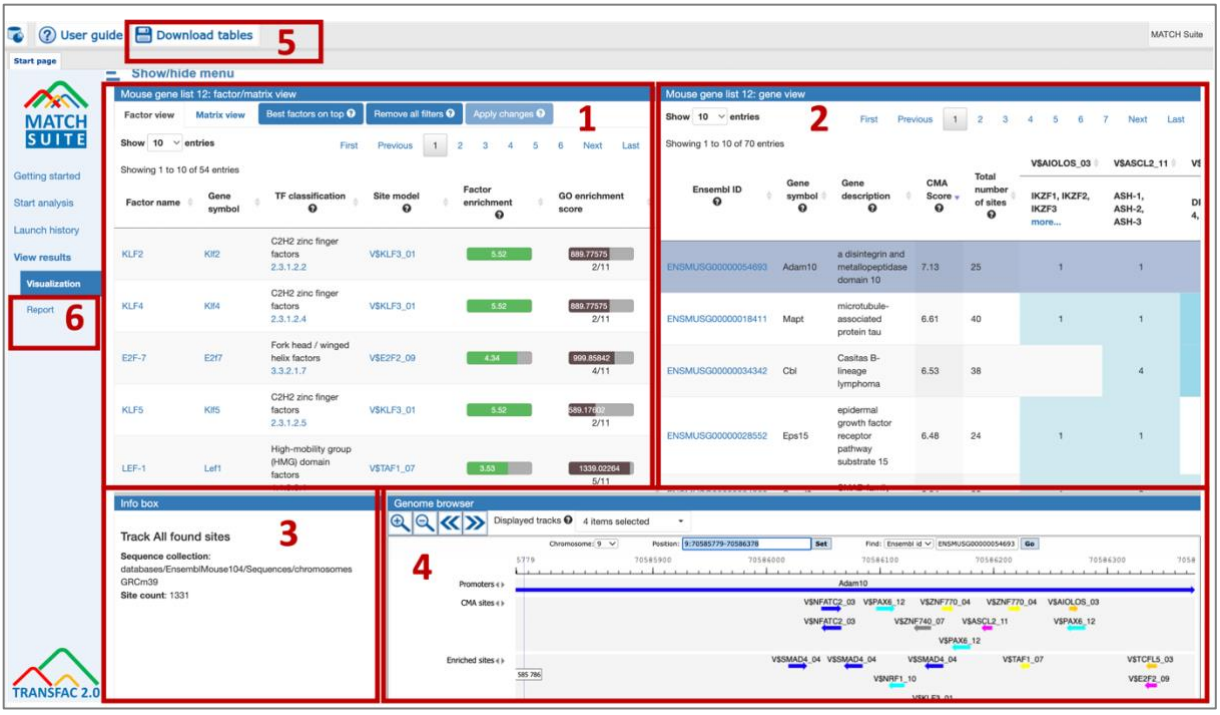
In the results visualization mode the screen will be divided into four different segments with the following functions:

(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

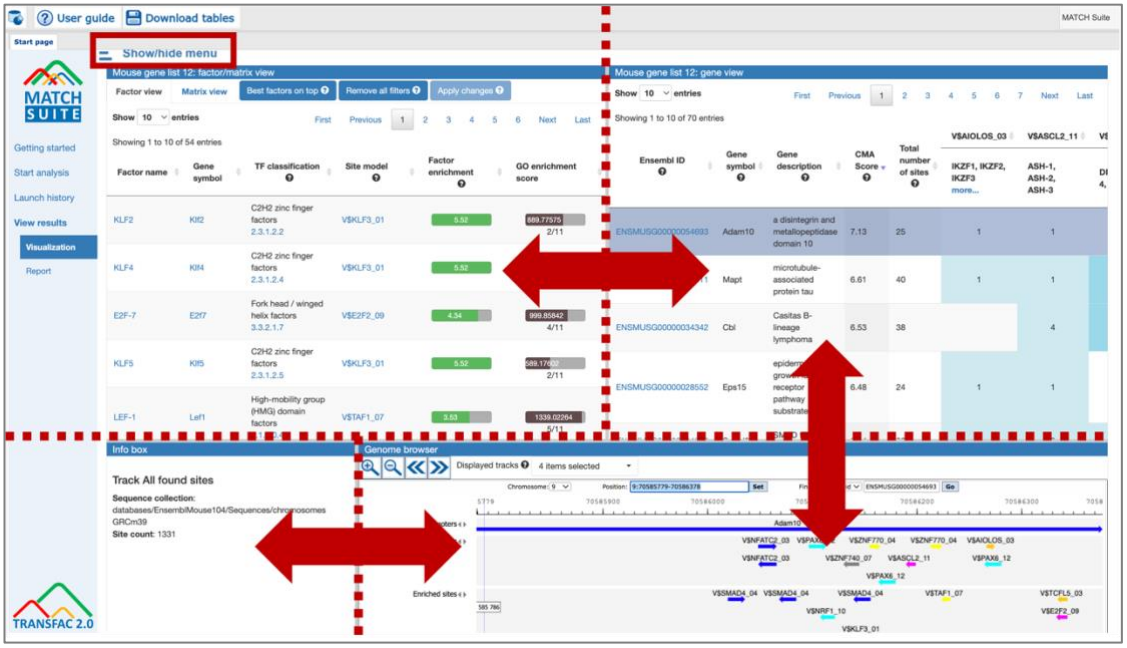
(4) The **genome browser**, allowing to visualize the tracks of found sites and additional annotation tracks for further results interpretation



The 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.

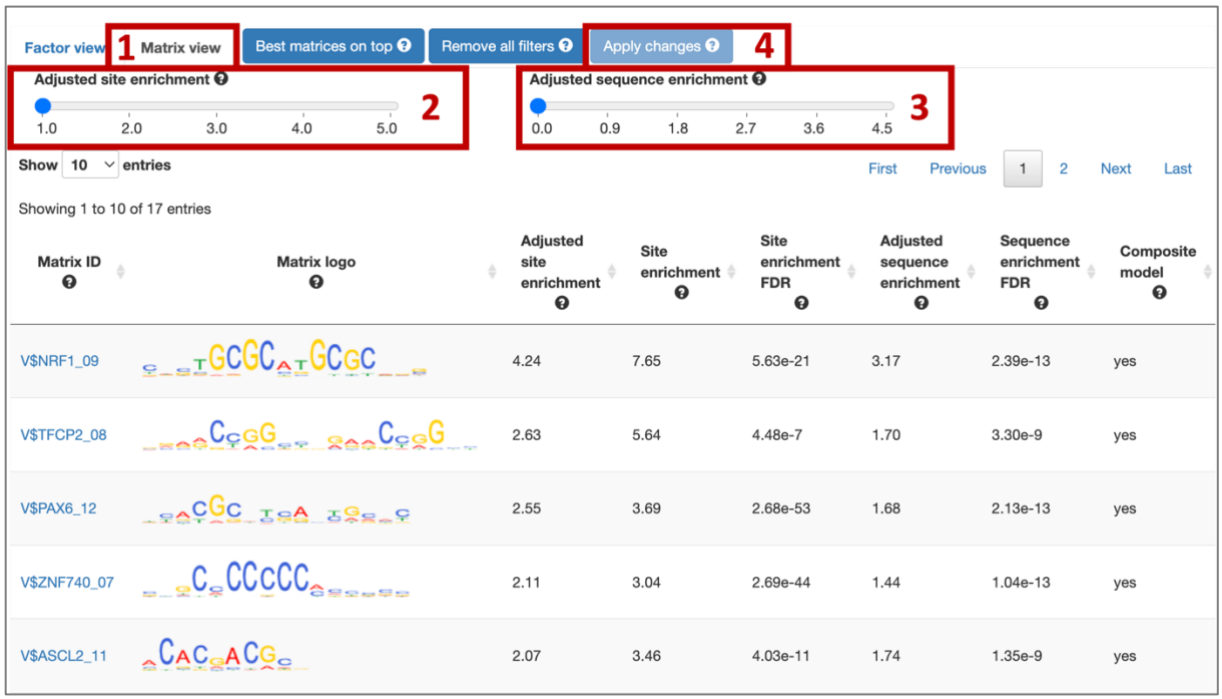
The columns of tables have in-built sorting option (6) which allows to sort the values within one column by ascending or descending order or by alphabetical order in case of textual contents. Simply click on the gray arrows for the sorting to be applied. By default, best factors (or matrices) are brought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



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.

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$AIOLOS_03 IKZF1, IKZF2, IKZF3 more...	V$ASCL2_11 ASH-1, ASH-2, ASH-3	V$ D 4,
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	

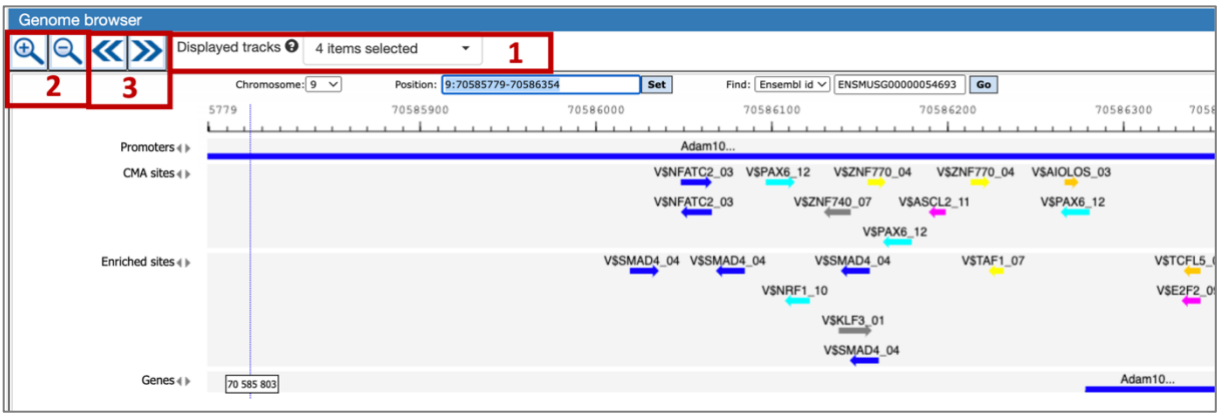
To visualize in genome browser the sites found in the promoter of any gene, just click on the line with the gene of your interest inside the genes table and refer to the genome browser below to explore the predicted regulation model of the respective gene.

Genome Browser

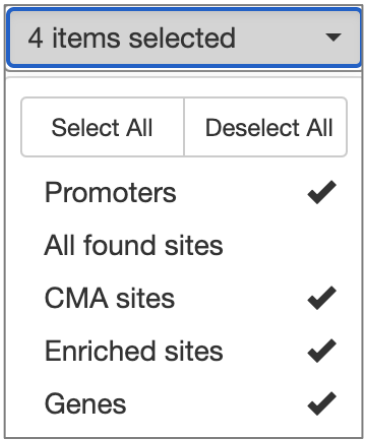
The genome browser provides you with visualizations of the predicted gene regulation models for each gene from the input gene set. Having clicked on any gene from the *Genes table*, the promoter model of the respective gene will be automatically opened in the genome browser with visualization of the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



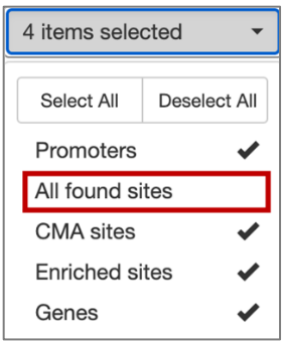
Displayed tracks can be customized using the **Displayed tracks** dropdown list (1)



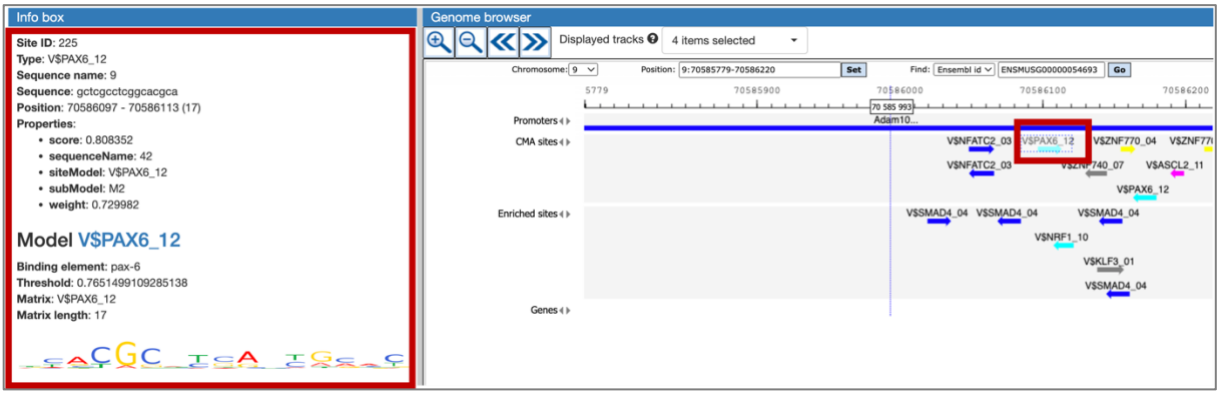
You can zoom in and zoom out for a more generalized or a more detailed view of the promoter model using the **zoom in** and **zoom out** buttons (2). In case navigation in genome browser is lost due to multiple zoom clicks, you can always return to the promoter of your interest by selecting the respective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



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:



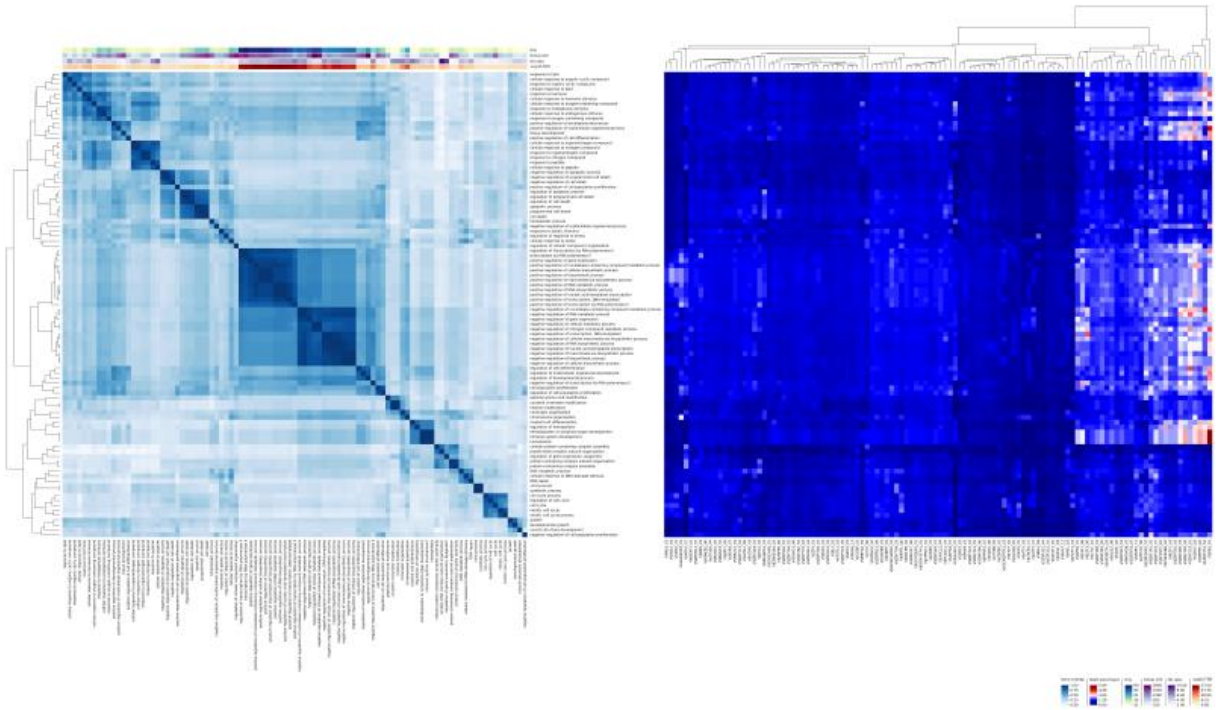
Heatmaps

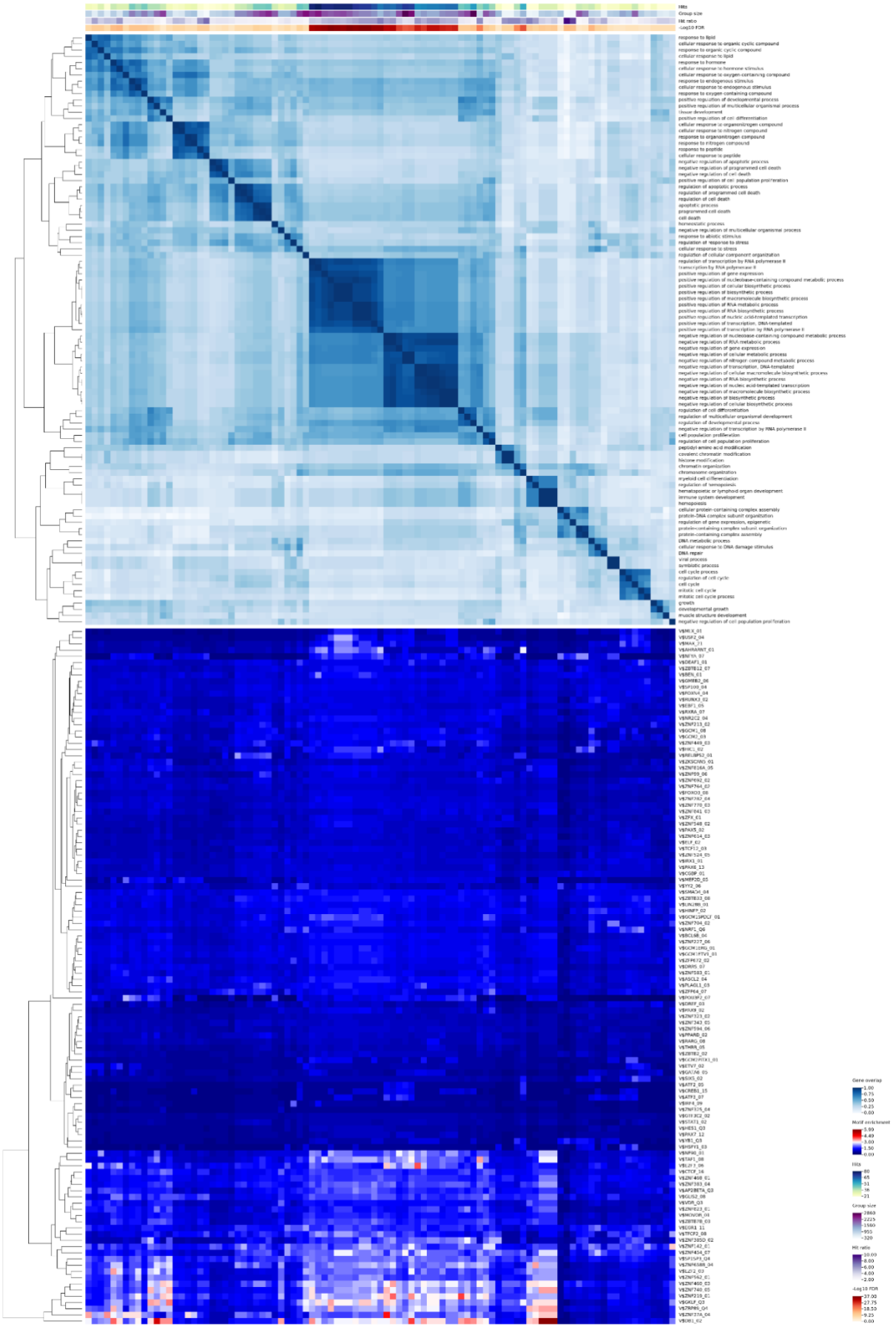
If the functional classification of your input gene set by GO terms produced reliable results (by p-values and by the number of genes in each functional category), heatmaps for the respective GO categorization will be included in the optional section *Functional Analysis of Gene Regulation* of the analysis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



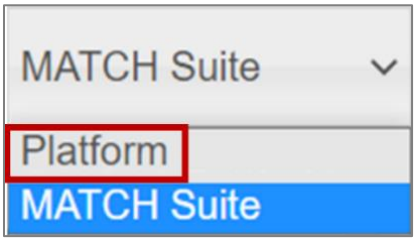

Hits
Group size
Hit ratio
-Log10 FDR
response to lipid
cellular response to organic cyclic compound
response to organic cyclic compound
cellular response to lipid
response to hormone
cellular response to hormone stimulus
cellular response to oxygen-containing compound
response to endogenous stimulus
cellular response to endogenous stimulus
response to oxygen-containing compound
positive regulation of developmental process
positive regulation of multicellular organismal process
tissue development
positive regulation of cell differentiation
cellular response to organonitrogen compound
cellular response to nitrogen compound
response to organonitrogen compound
response to nitrogen compound
response to peptide
cellular response to peptide
negative regulation of apoptotic process
negative regulation of programmed cell death
negative regulation of cell death
positive regulation of cell population proliferation
regulation of apoptotic process
regulation of programmed cell death
regulation of cell death
apoptotic process
programmed cell death
cell death
homeostatic process
negative regulation of multicellular organismal process
response to abiotic stimulus
regulation of response to stress
cellular response to stress
regulation of cellular component organization
regulation of transcription by RNA polymerase II
transcription by RNA polymerase II
positive regulation of gene expression
positive regulation of nucleobase-containing compound metabolic process
positive regulation of cellular biosynthetic process
positive regulation of biosynthetic process
positive regulation of macromolecule biosynthetic process
positive regulation of RNA metabolic process
positive regulation of RNA biosynthetic process
positive regulation of nucleic acid-templated transcription
positive regulation of transcription, DNA-templated
positive regulation of transcription by RNA polymerase II
negative regulation of nucleobase-containing compound metabolic process
negative regulation of RNA metabolic process
negative regulation of gene expression
negative regulation of cellular metabolic process
negative regulation of nitrogen compound metabolic process
negative regulation of transcription, DNA-templated
negative regulation of cellular macromolecule biosynthetic process
negative regulation of RNA biosynthetic process
negative regulation of nucleic acid-templated transcription
negative regulation of macromolecule biosynthetic process
negative regulation of biosynthetic process
negative regulation of cellular biosynthetic process
regulation of cell differentiation
regulation of multicellular organismal development
regulation of developmental process
negative regulation of transcription by RNA polymerase II
cell population proliferation
regulation of cell population proliferation
peptidyl-amino acid modification
covalent chromatin modification
histone modification
chromatin organization
chromosome organization
myeloid cell differentiation
regulation of hemopoiesis
hematopoietic or lymphoid organ development
immune system development
hemopoiesis
cellular protein-containing complex assembly
protein-DNA complex subunit organization
regulation of gene expression, epigenetic
protein-containing complex subunit organization
protein-containing complex assembly
DNA metabolic process
cellular response to DNA damage stimulus
DNA repair
viral process
symbiotic process
cell cycle process
regulation of cell cycle
cell cycle
mitotic cell cycle
mitotic cell cycle process
growth
developmental growth
muscle structure development
negative regulation of cell population proliferation
Gene overlap
Motif enrichment
Hits
Group size
Hit ratio
-Log10 FDR

TRANSFAC 2.0

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.