TRANSFAC Getting Started Guide
Version 2016.3

TRANSFAC is a database containing published data on eukaryotic transcription factors and miRNAs, their experimentally-proven binding sites, and regulated genes. The extensive compilation of binding sites forms the basis of derived positional weight matrices which can be used with the included Match tool to search DNA sequences for putative transcription factor binding sites – a method that is crucial for filling in gaps in the published literature. Integrated transcription factor ChIP-Seq, DNase hypersensitivity and histone methylated intervals from the ENCODE project complement the curated and predicted binding site data.

This guide will walk you through some of the most common uses of TRANSFAC. For additional help, please contact our support team at mailto:support@genexplain.com

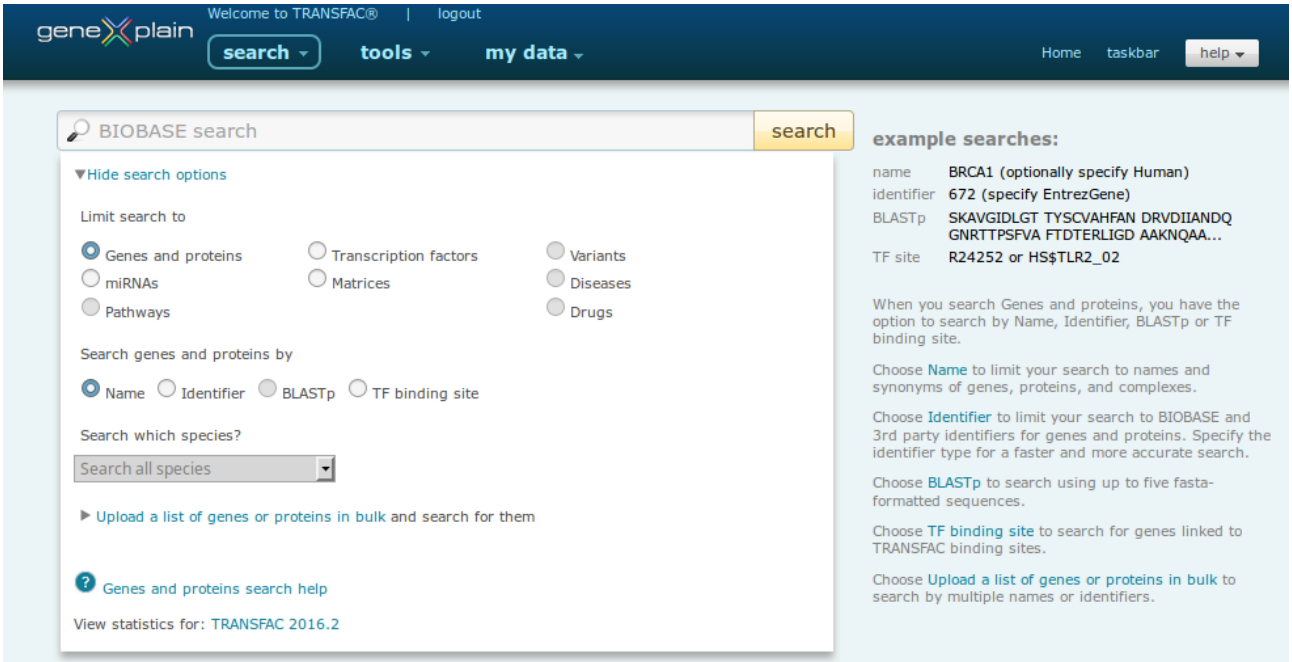
Quick links to sections covered:

Identify experimentally demonstrated regulators for a gene of interest

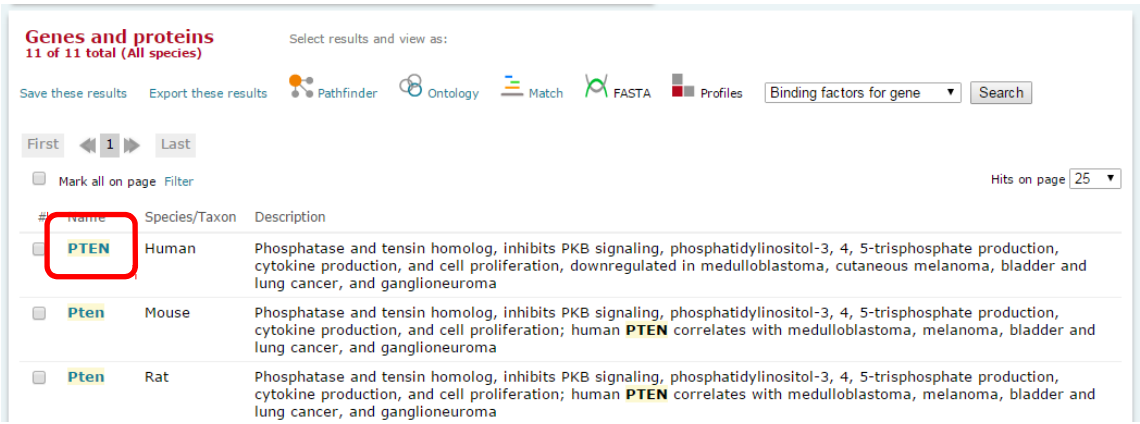
TRANSFAC Locus Reports provide information about genes, including the transcription factors which have been experimentally demonstrated to bind to the gene's sequence.

To navigate to the Locus Report for a gene of interest:

- Select the "Genes and proteins" radio button
- Select the "Name" radio button
- Enter the name of your gene in the search box (PTEN in this example)
- Click the "search" button



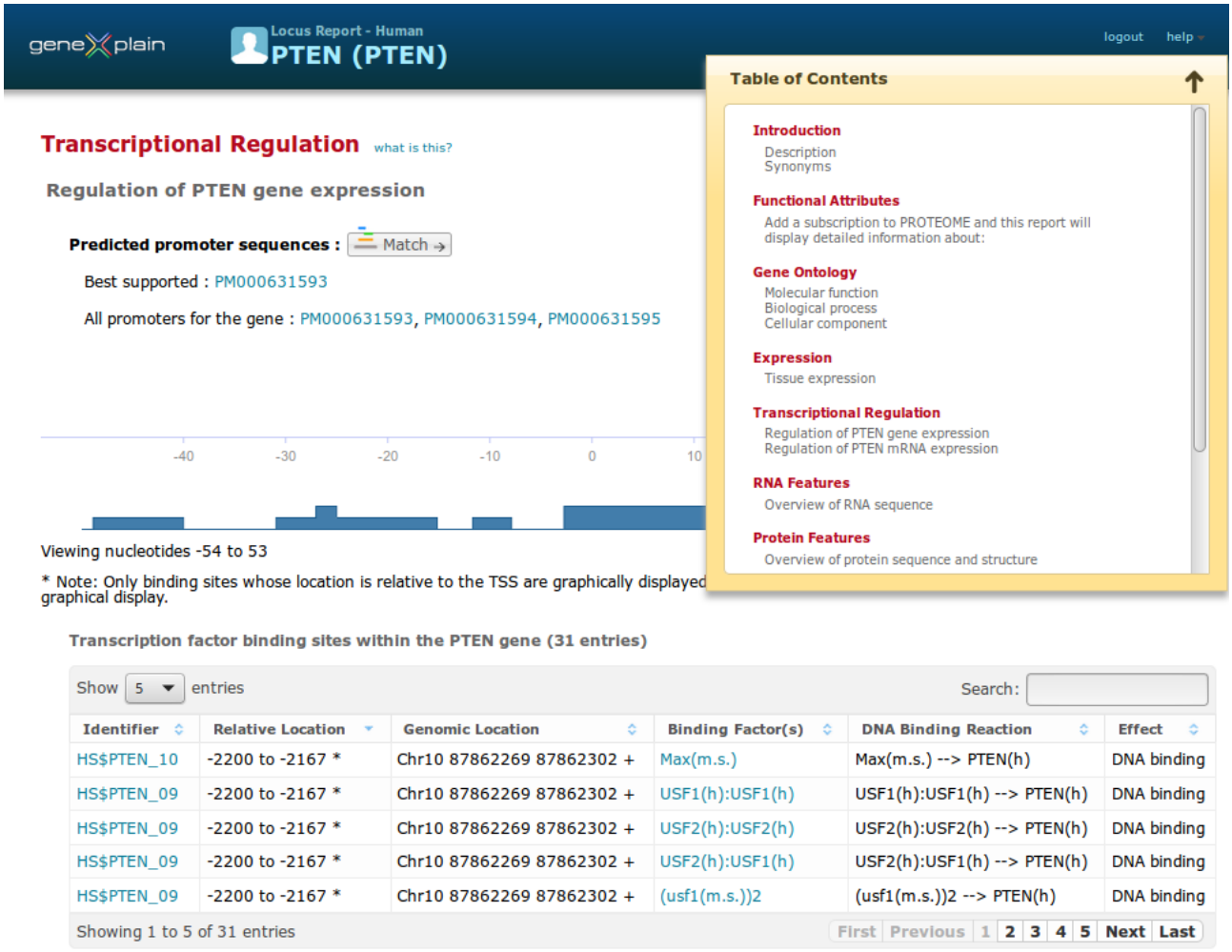
The list of matched genes is returned:



Click the human “PTEN” link to navigate to the Locus Report.

Scroll down the page or use the “Table of Contents” section to jump to the “Transcriptional Regulation” section:

A summary of all curated, experimental binding sites is given in the table. For more detailed information about a specific binding site, click the link in the “Identifier” column to navigate to the Site Report which contains detailed supporting information about the method(s) used to measure binding, the experimental system and more, including links to the abstract of the reference from which the information was curated.



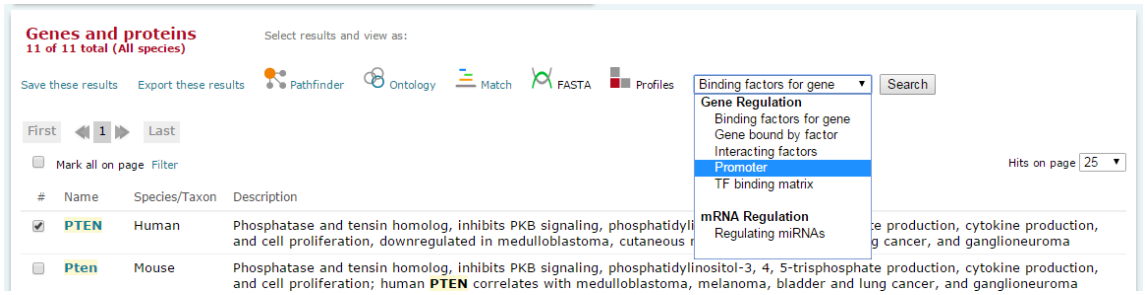
Visualize a gene's promoter and associated regulatory features

TRANSFAC contains promoter sequences for human, mouse, rat, Arabidopsis, rice and soybean species. Promoter sequences are 11,000 nucleotides in length, spanning from 10,000 nucleotides upstream of the transcription start site (TSS) to 1,000 nucleotides downstream of a virtual TSS that is calculated based on clusters of experimental TSSs. An individual gene may have more than one associated promoter. When this is the case one of the promoters will be selected as the best-supported, which indicates that it is the promoter with the greatest fraction of clustered TSSs.

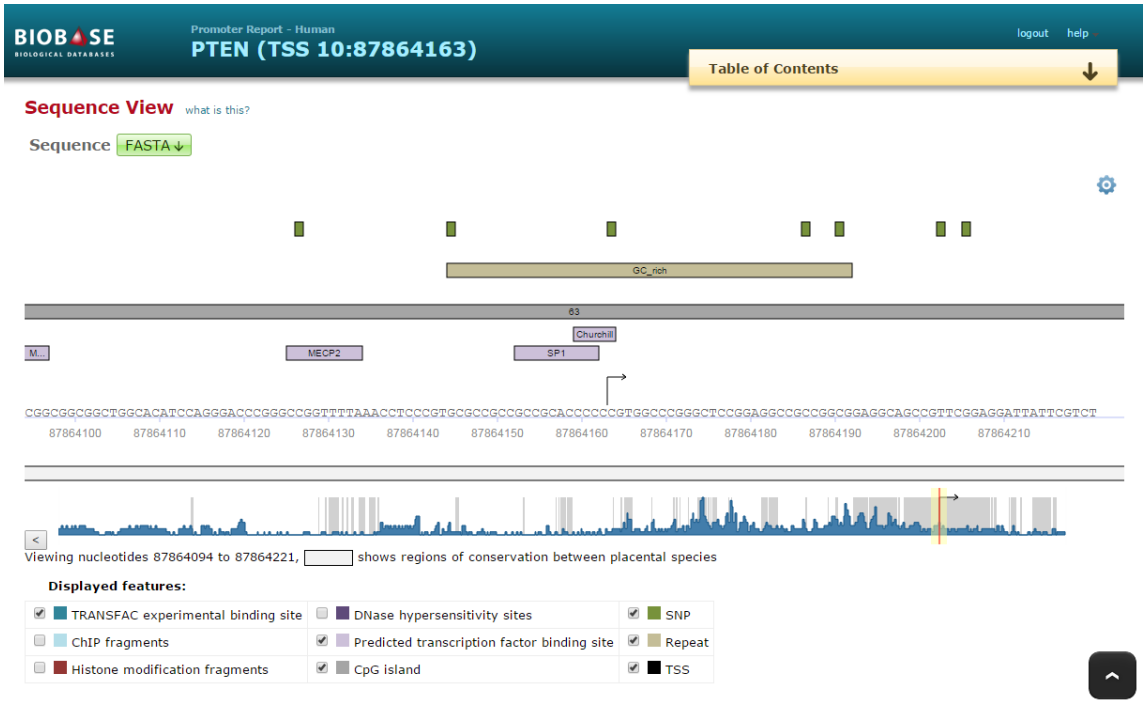
There are two ways to graphically visualize the promoter sequence plus associated regulatory features which have been mapped to the promoter, including experimentally demonstrated factor binding sites.

The first way is to navigate to the "Transcriptional Regulation" section of the desired gene (see Identify experimentally demonstrated regulators for a gene of interest) and then click one of the "predicted promoter sequences" links:

The second way is to select the desired gene in the search results and perform a second search for its associated promoter(s):



Once the promoter search results load, click the link in the “Accession” column to open the Promoter Report. The Promoter Report provides a graphical view of the sequence:



Use the “Displayed features” options shown above to add or remove different types of features. The red orientation point and flanking yellow region in the frequency bar below the nucleotide sequence indicates the portion of the sequence that is displayed. Drag the red orientation point along the frequency bar to inspect different regions of the promoter at the nucleotide level.

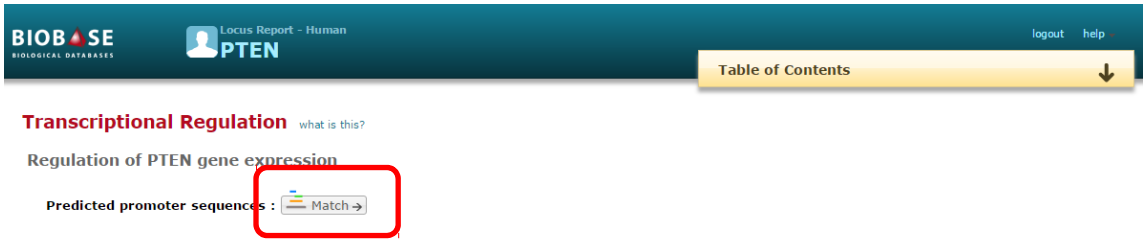
The graphic is followed by a tabular view of the different features

Identify predicted binding sites for a gene of interest

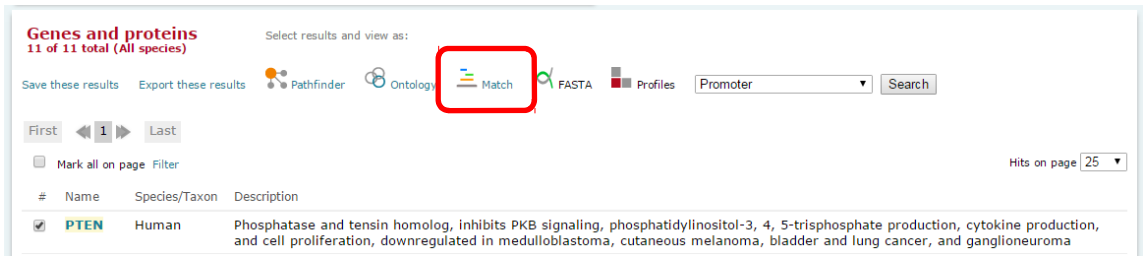
While the Locus Report provides an overview of all curated, experimental binding sites for a gene (see Identify experimentally demonstrated regulators for a gene of interest) and the Promoter Report provides a graphical view of those experimental binding site plus additional regulatory features located within the gene’s promoter (see Visualize a gene’s promoter and associated regulatory features), the Predict TF binding sites tool and the Match algorithm that it employs enables you to scan the gene’s promoter for additional predicted binding sites using TRANSFAC’s comprehensive positional weight matrix library.

There are two ways to interrogate the gene’s promoter.

The first way is to navigate to the “Transcriptional Regulation” section of the Locus Report (see Identify experimentally demonstrated regulators for a gene of interest) and then click the Match icon to launch the Match algorithm:

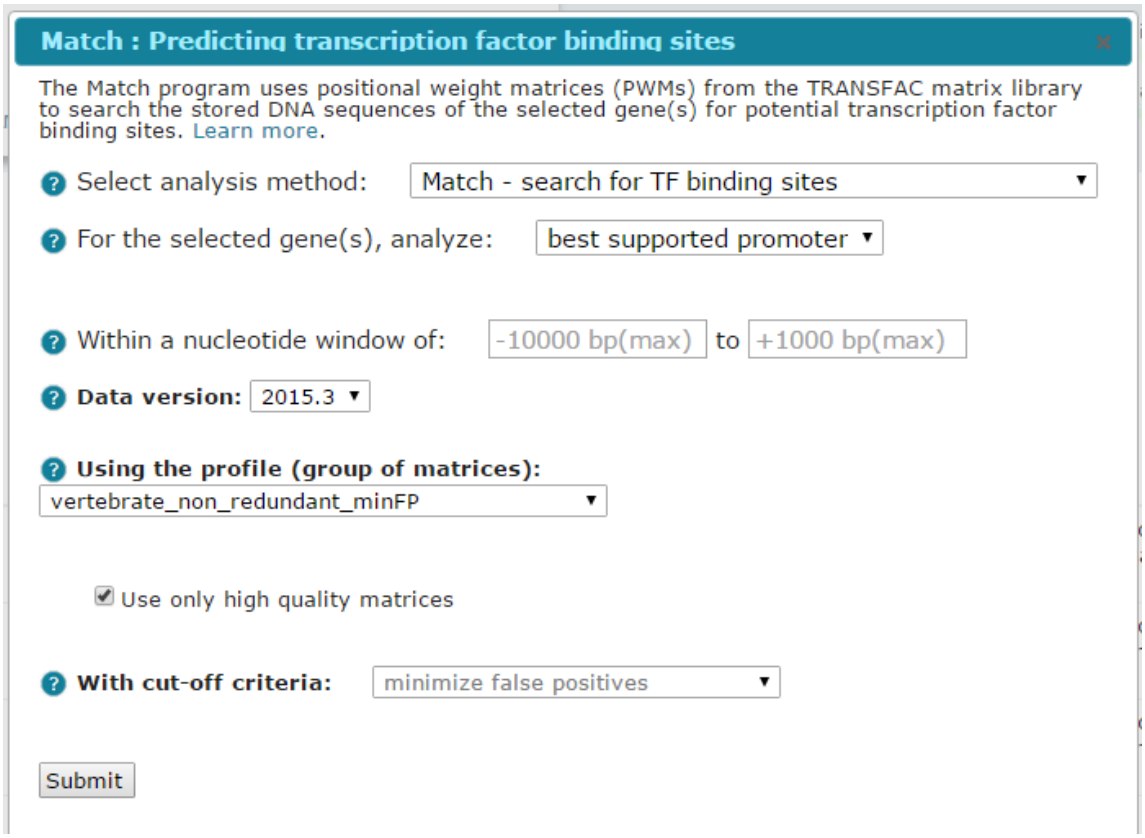


The second way is to perform a search for the gene of interest (see Identify experimentally demonstrated regulators for a gene of interest), and then launch the Match algorithm directly from the search results by selecting the gene of interest and clicking the Match icon:

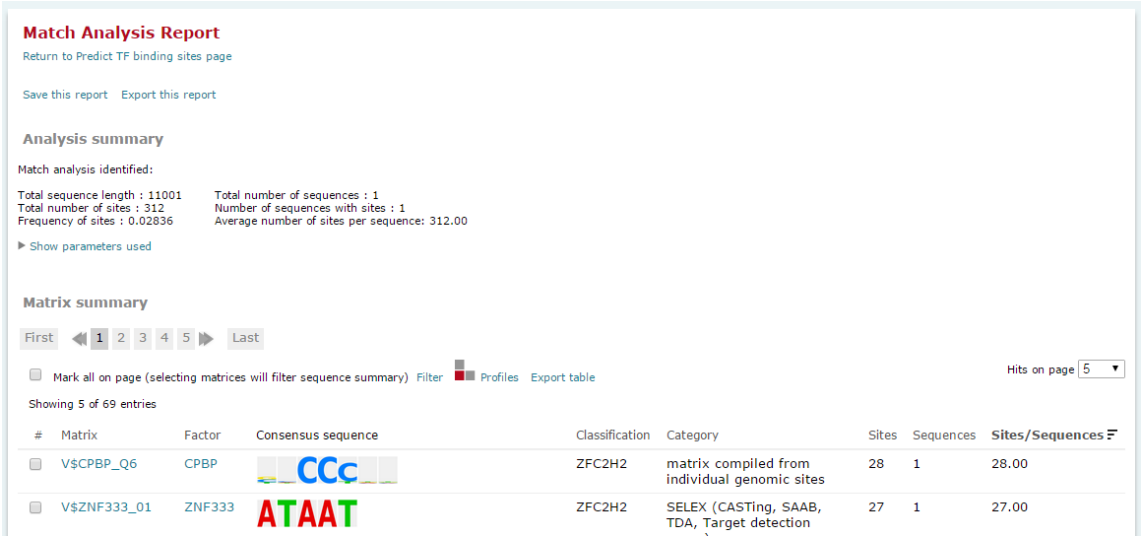


In both cases an intermediate window will open where you are asked to:

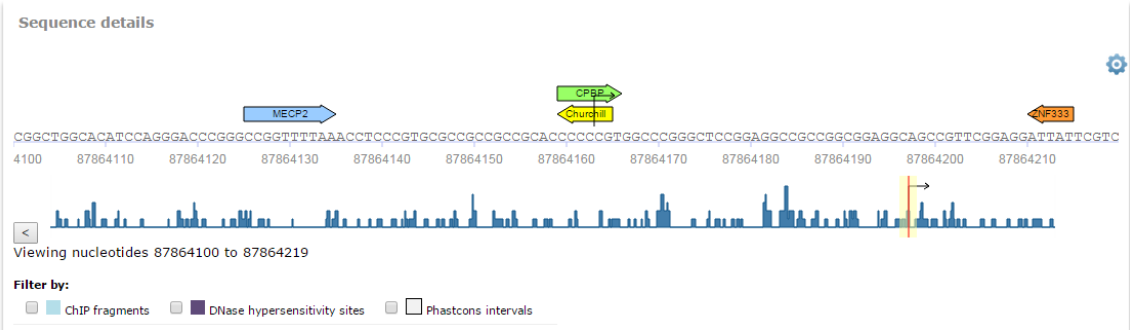
- Select the analysis method (use the default “Match”)
- Select the promoter to be analyzed (use the default “best supported promoter” if you want to analyze the promoter with the greatest number of clustered TSSs, or change to “all promoters” if you would like to analyze all promoters for comparison)
- Specify the region of the promoter to analyze (use the default “-10,000 bp to +1,000 bp” to analyze the entire promoter, or select a smaller region within those bounds)
- Specify the data version (use the default, current version)
- Specify the profile (group of matrices) to be used (use the default “vertebrate_non_redundant_minFP” which contains representative matrices for distinctive binding motifs unless you are working with non-vertebrate species or want to limit your analysis to factors that are active in a particular tissue)
- Specify the cut-off criteria (use the default “minimize false positives” which is automatically assigned with the “vertebrate_non_redundant_minFP” profile)



When you click the Submit button your analysis will be forwarded to the taskbar. Once the analysis completes, the Match Analysis Report will automatically load:



The “Matrix summary” section provides a tabular overview of all matrices that were found to have binding sites within the promoter. Scrolling down to the “Sequence details” section a graphical view of the results is provided:

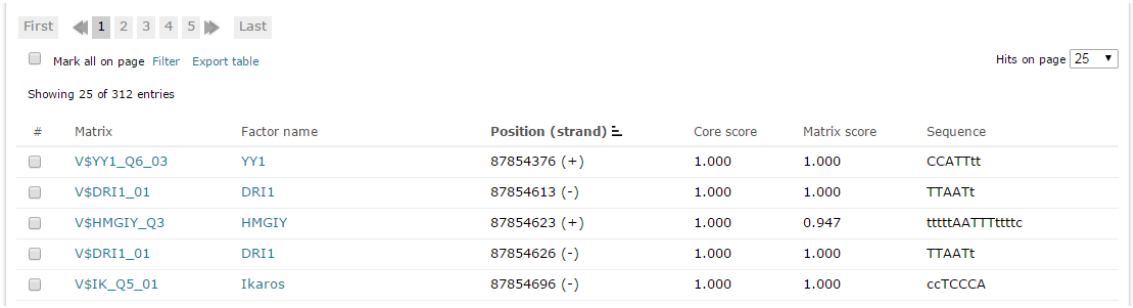


A frequency bar displays the distribution of sites along the entire sequence, with the red orientation point and flanking yellow highlighting indicating the portion of the sequence that is displayed at the nucleotide level above. Slide the red orientation point along the frequency bar to view other portions of the sequence in detail.

Optionally use the ChIP fragments, DNase hypersensitivity sites and Phastcons intervals filters to apply additional biological context to your results. Activating these filters removes any sites that do not overlap with regions of sequence that are suggested to be available to the transcriptional machinery based on being bound by transcription factors (ChIP fragments), being accessible to DNase enzymes (DNase hypersensitivity) or conserved between human and mouse (Phastcons).

Please note that if “all promoters” was selected for the analysis it will be necessary to click the Sequence name in the “Sequence summary” section to display the nucleotide level view.

The table below the graphical view lists all identified binding sites with supporting details:



First 1 2 3 4 5 Last

Mark all on page Filter Export table

Hits on page 25

Showing 25 of 312 entries

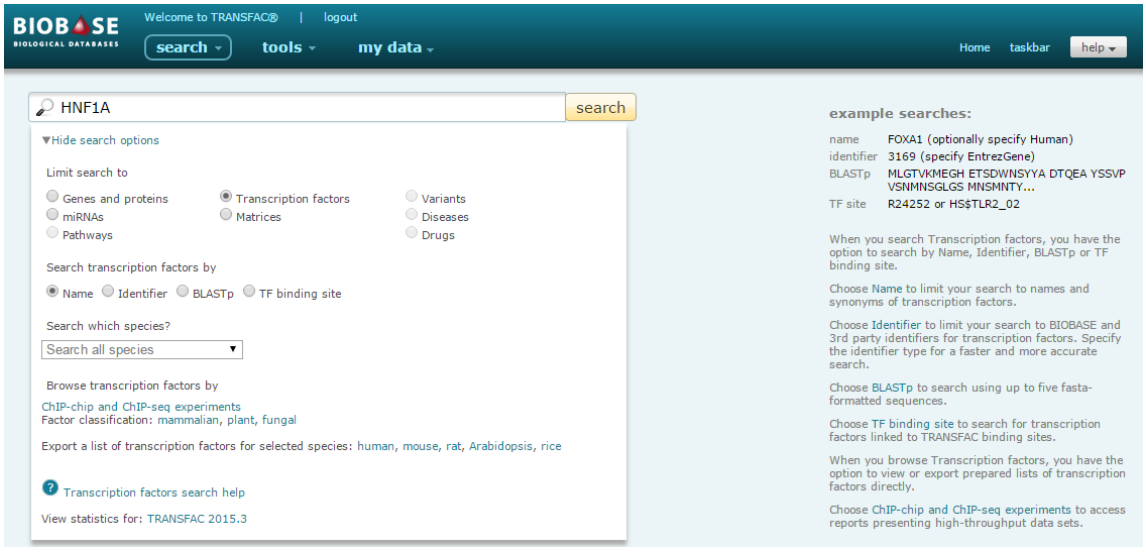
#	Matrix	Factor name	Position (strand)	Core score	Matrix score	Sequence
	V$YY1_Q6_03	YY1	87854376 (+)	1.000	1.000	CCATTtt
	V$DRI1_01	DRI1	87854613 (-)	1.000	1.000	TTAATt
	V$HMGIY_Q3	HMGIY	87854623 (+)	1.000	0.947	tttttAATTTttttc
	V$DRI1_01	DRI1	87854626 (-)	1.000	1.000	TTAATt
	V$IK_Q5_01	Ikaros	87854696 (-)	1.000	1.000	ccTCCCA

Identify experimentally demonstrated targets for a transcription factor of interest

There are two ways to identify the experimentally determined targets for a transcription factor. The first way is to navigate to the Locus Report for the factor and view the information provided in the “Transcriptional Regulation” section.

To navigate to the Locus Report for a transcription factor of interest:

- Select the “Transcription factors” radio button
- Select the “Name” radio button
- Enter the name of your gene in the search box (HNF1A in this example)
- Click the “search” button



BIOBASE BIOLOGICAL DATABASES

Welcome to TRANSFAC® | logout

search tools my data Home taskbar help

HNF1A search

Hide search options

Limit search to

Genes and proteins Transcription factors Variants
miRNAs Matrices Diseases
Pathways Drugs

Search transcription factors by

Name Identifier BLASTp TF binding site

Search which species?

Search all species

Browse transcription factors by

ChIP-chip and ChIP-seq experiments
Factor classification: mammalian, plant, fungal

Export a list of transcription factors for selected species: human, mouse, rat, Arabidopsis, rice

Transcription factors search help

View statistics for: TRANSFAC 2015.3

example searches:

name FOXA1 (optionally specify Human)
identifier 3169 (specify EntrezGene)
BLASTp MLGTVKMEGH ETSDWNSYYA DTQEA YSSVP VSNMNSGLGS MNSMNTY...
TF site R24252 or HS$TLR2_02

When you search Transcription factors, you have the option to search by Name, Identifier, BLASTp or TF binding site.

Choose Name to limit your search to names and synonyms of transcription factors.

Choose Identifier to limit your search to BIOBASE and 3rd party identifiers for transcription factors. Specify the identifier type for a faster and more accurate search.

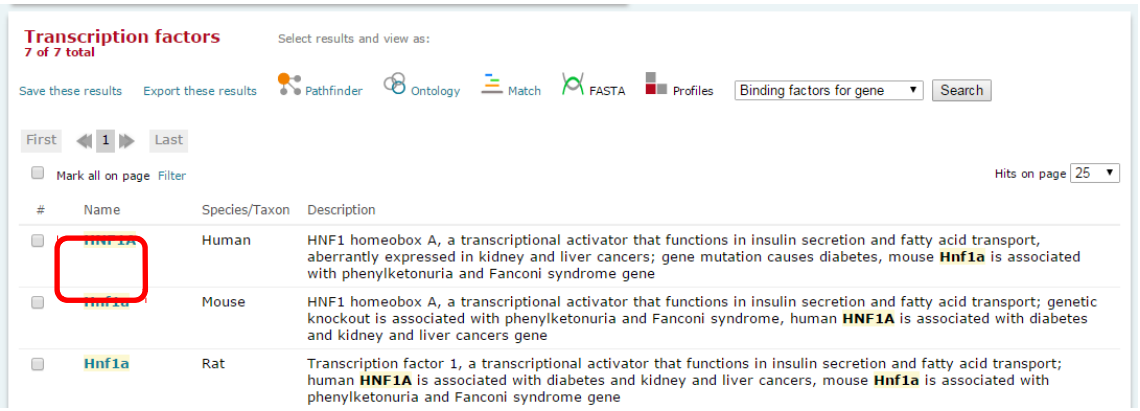
Choose BLASTp to search using up to five fasta-formatted sequences.

Choose TF binding site to search for transcription factors linked to TRANSFAC binding sites.

When you browse Transcription factors, you have the option to view or export prepared lists of transcription factors directly.

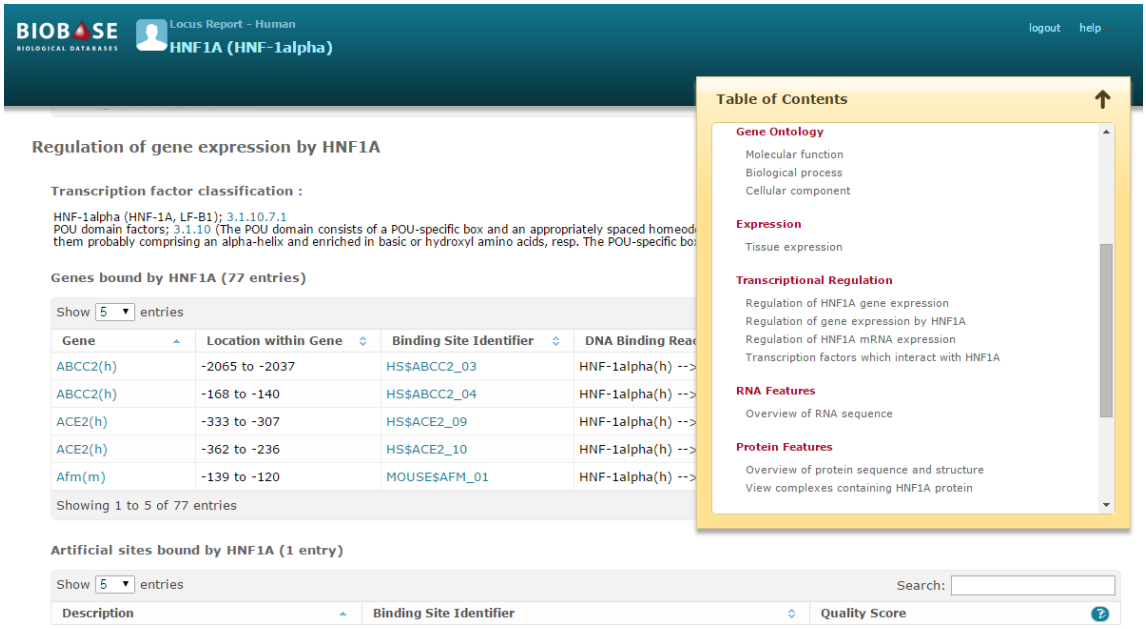
Choose ChIP-chip and ChIP-seq experiments to access reports presenting high-throughput data sets.

The list of matched factors is returned:



Click the human “HNF1A” link to navigate to the Locus Report.

Scroll down the page or use the “Table of Contents” section to jump to the “Transcriptional Regulation” section and scroll to the “Regulation of gene expression by HNF1A” sub-section:

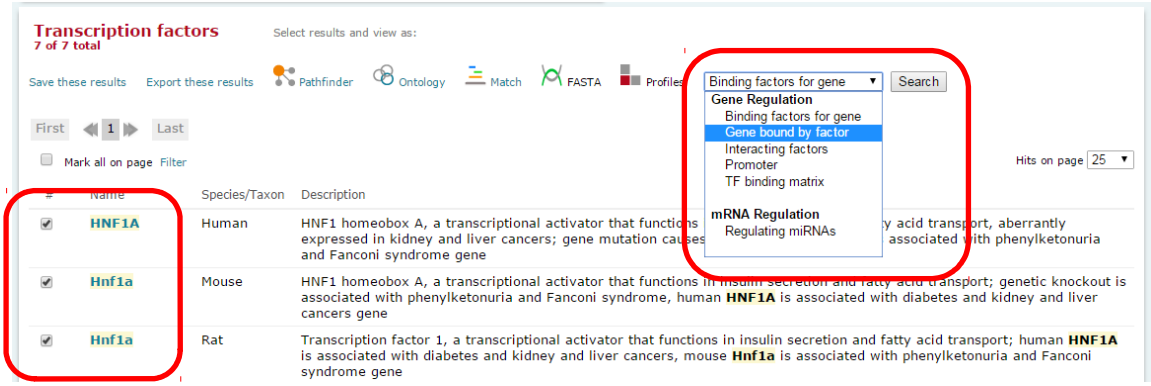


A summary of all curated, experimental binding sites for human HNF1A is given in the table. For more detailed information about a specific binding site, click the link in the “Identifier” column to navigate to the Site Report which contains detailed supporting information about the method(s) used to measure binding, the experimental system and more, including links to the abstract of the reference from which the information was curated.

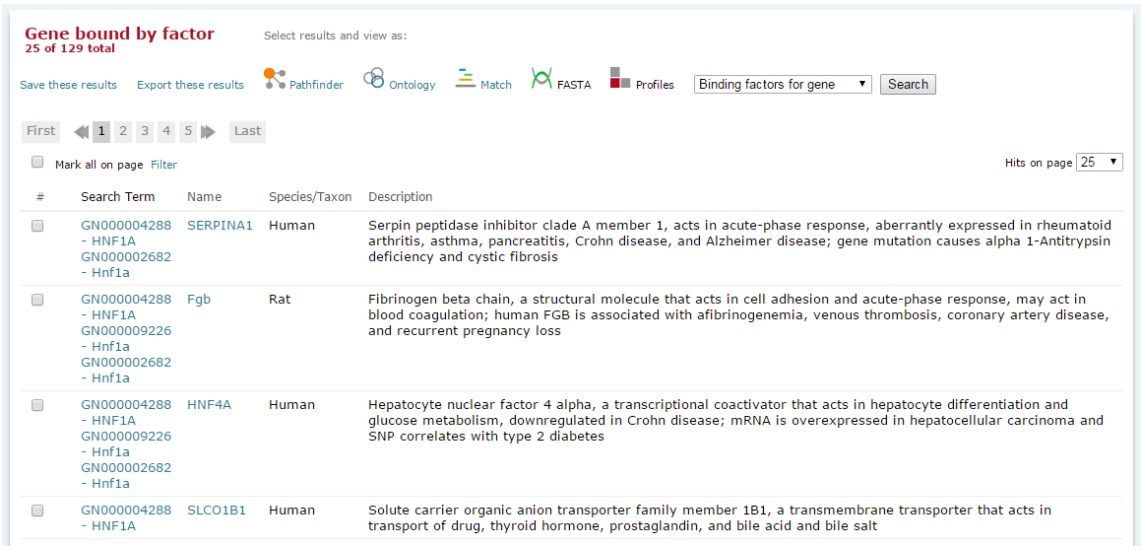
It is important to note that because curation is species specific, this Locus Report will only provide information for results of experiments where human HNF1A was the focus. Results

from experiments using closely related mouse and rat Hnf1a factors will only be covered on the respective mouse and rate Locus Reports.

The second way to identify the experimentally determined binding sites for a transcription factor of interest avoids this separation of information by species. In this case perform the same “Transcription factors” search for HNF1A. Then when the results are returned, select the human, mouse and rat entries, select the “Gene bound by factor” option in the pulldown menu in the search results header and click the Search button next to the pulldown menu:



All three factors are used as input to the search. The results are a combined list of the genes that each have been shown to bind experimentally:

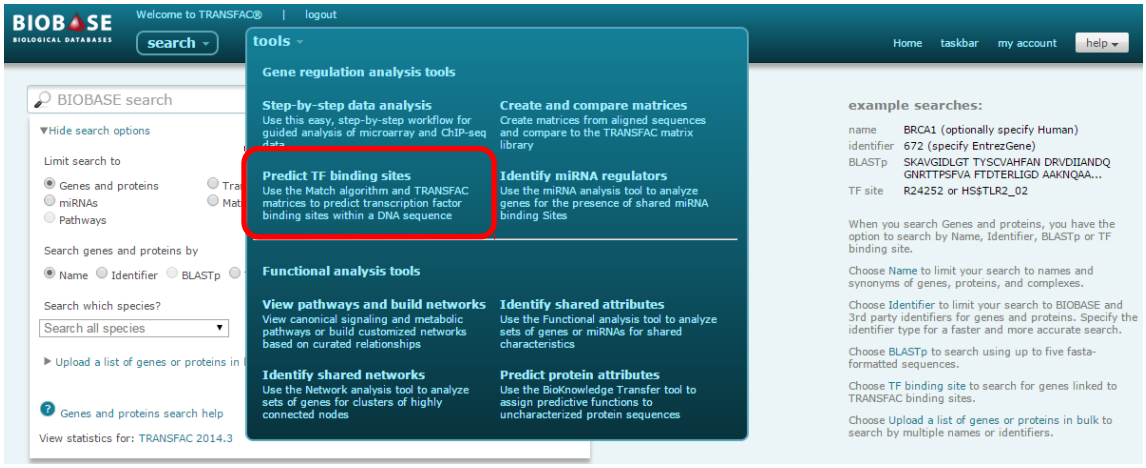


Following the link to the Locus Report of the transcription factor or the gene and navigating to the “Transcriptional Regulation” section will allow you to access the same experimental details described above.

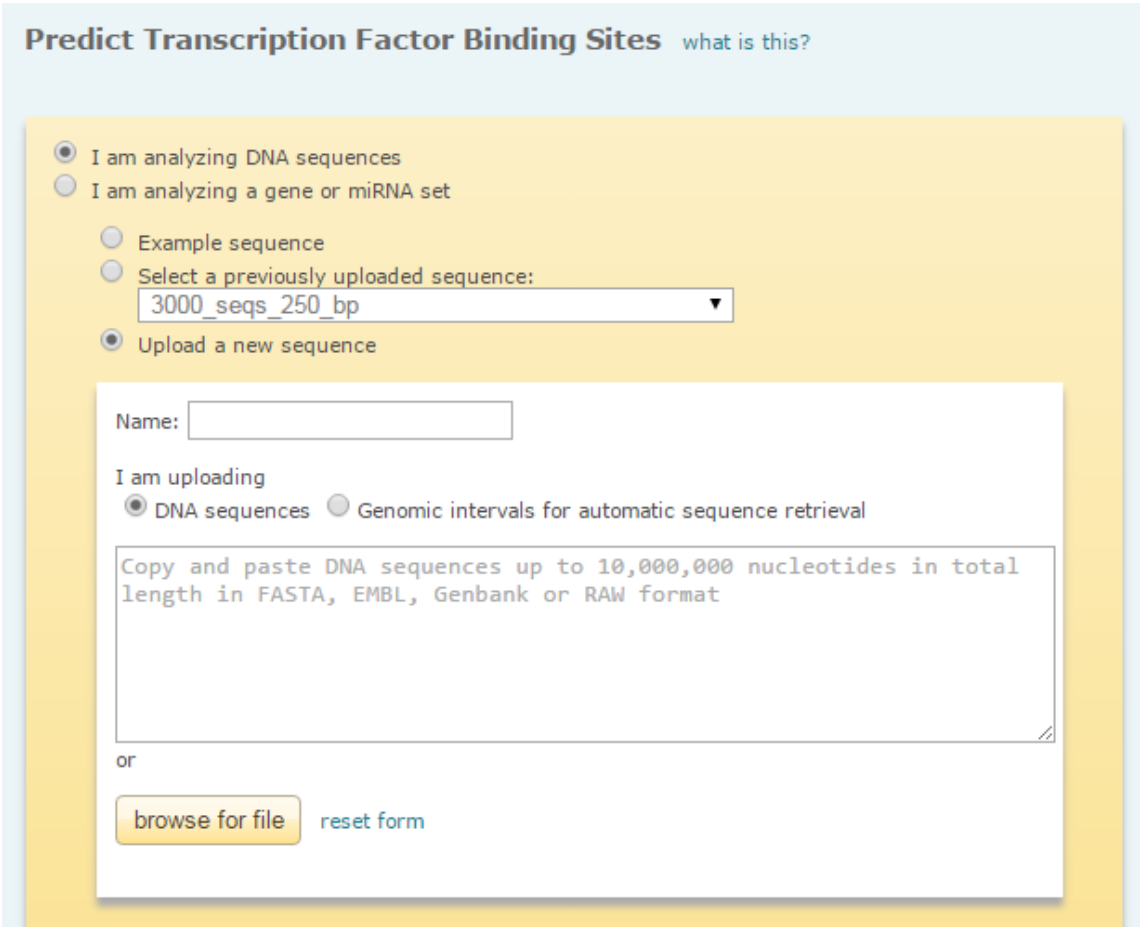
Upload your own sequence and identify the predicted binding sites within

TRANSFAC provides preloaded promoter sequences for human, mouse, rat, Arabidopsis, rice and soybean genes but you may upload your own sequence or set of sequences for these or any other species and then analyze the sequence(s) for predicted transcription factor binding sites.

Select the “Predict TF binding sites” option from the tools menu:



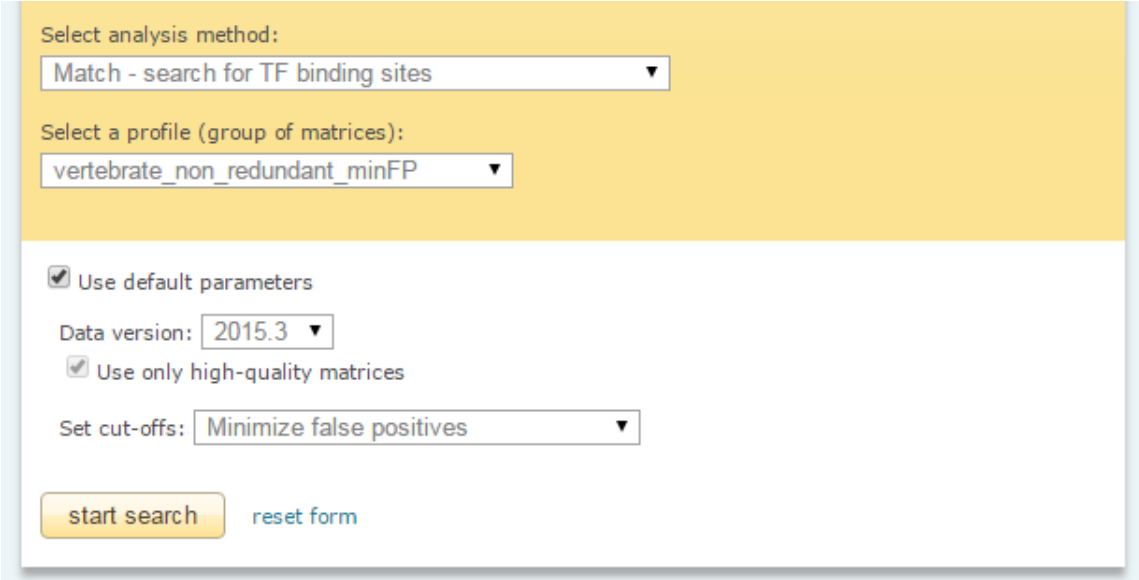
When the tool loads, click the “I am analyzing DNA sequences” radio button followed by the “Upload a new sequence” radio button:



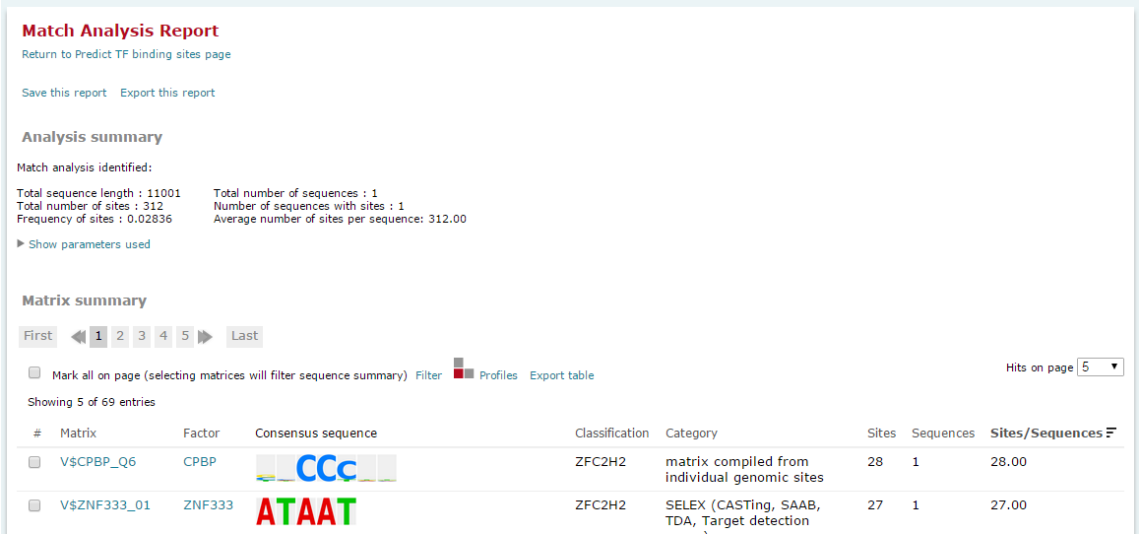
You will be presented with the option to upload your data as a set of DNA sequences in FASTA, EMBL, Genbank or RAW format or as a set of genomic intervals in .bed format (note that the genomic intervals option is only supported for human, mouse and rat species).

Once you have specified the sequence(s) to be uploaded, accept or adapt in the remaining parameters:

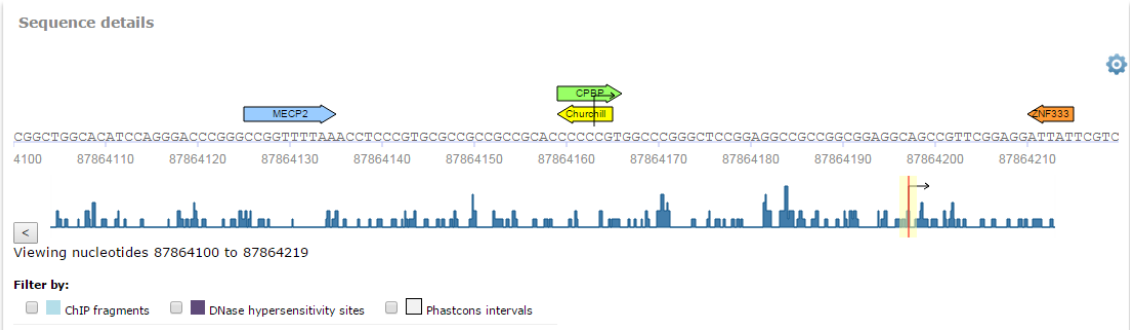
- Select the analysis method (use the default “Match”)
- Specify the profile (group of matrices) to be used (use the default “vertebrate_non_redundant_minFP” which contains representative matrices for distinctive binding motifs unless you are working with non-vertebrate species or want to limit your analysis to factors that are active in a particular tissue)
- Keep the “Use default parameters” option enabled



When you click the start search button your analysis will be forwarded to the taskbar. Once the analysis completes, the Match Analysis Report will automatically load:



The “Matrix summary” section provides a tabular overview of all matrices that were found to have binding sites within the uploaded sequence. Scrolling down to the “Sequence details” section a graphical view of the results is provided:

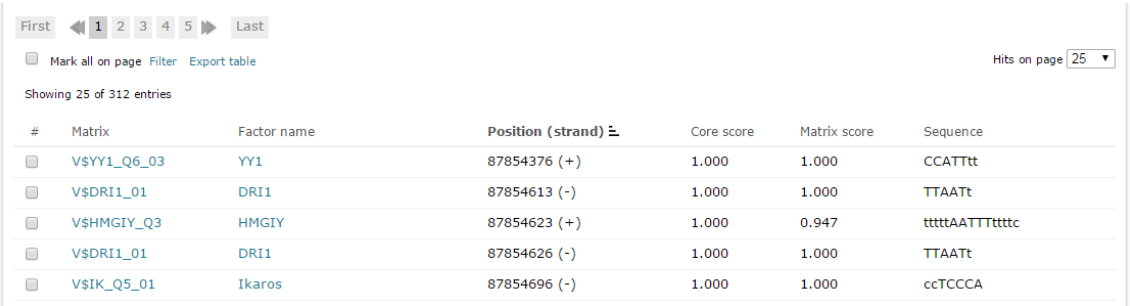


A frequency bar displays the distribution of sites along the entire sequence, with the red orientation point and flanking yellow highlighting indicating the portion of the sequence that is displayed at the nucleotide level above. Slide the red orientation point along the frequency bar to view other portions of the sequence in detail.

If you uploaded your sequence in genomic coordinates, you may optionally use the ChIP fragments, DNase hypersensitivity sites and Phastcons intervals filters to apply additional biological context to your results. Activating these filters removes any sites that do not overlap with regions of sequence that are suggested to be available to the transcriptional machinery based on being bound by transcription factors (ChIP fragments), being accessible to DNase enzymes (DNase hypersensitivity) or conserved between human and mouse (Phastcons).

Please note that if more than one sequences was submitted for analysis it will be necessary to click the Sequence name in the “Sequence summary” section to display the nucleotide level view.

The table below the graphical view lists all identified binding sites with supporting details:



#	Matrix	Factor name	Position (strand)	Core score	Matrix score	Sequence
	V$YY1_Q6_03	YY1	87854376 (+)	1.000	1.000	CCATTtt
	V$DRI1_01	DRI1	87854613 (-)	1.000	1.000	TTAATt
	V$HMGIY_Q3	HMGIY	87854623 (+)	1.000	0.947	tttttAATTTttttc
	V$DRI1_01	DRI1	87854626 (-)	1.000	1.000	TTAATt
	V$IK_Q5_01	Ikaros	87854696 (-)	1.000	1.000	ccTCCCA

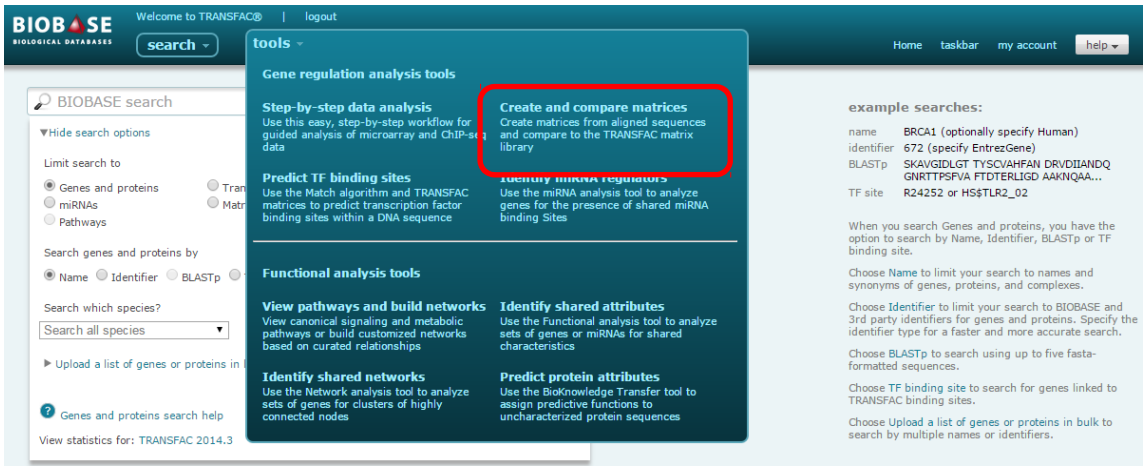
Your analysis will be preserved in the taskbar for 7 days. To preserve it indefinitely, click the “Save this report” link at the top of the report.

Import your own custom matrix for use in binding site predictions

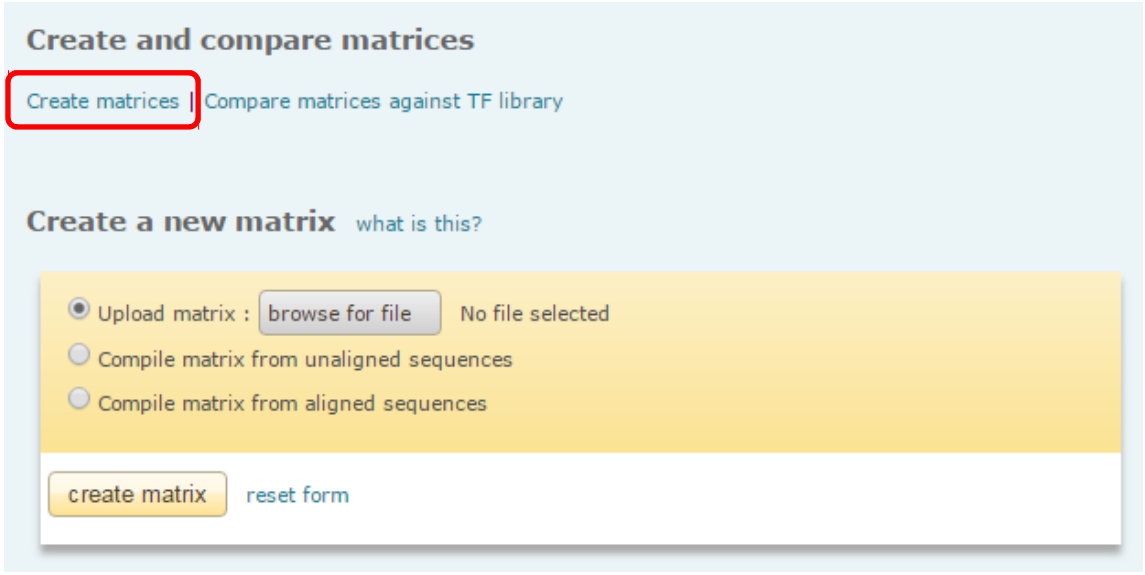
TRANSFAC’s matrix library is based on the published literature and published data sets. If

you have your own matrix, or set of sequences that you would like to compile into a matrix, you can upload it into TRANSFAC which will make it available to use in all analyses.

Select the “Create and compare matrices” option from the tools menu:



When the tool loads, click the “Create matrices” option to refresh the view if needed:



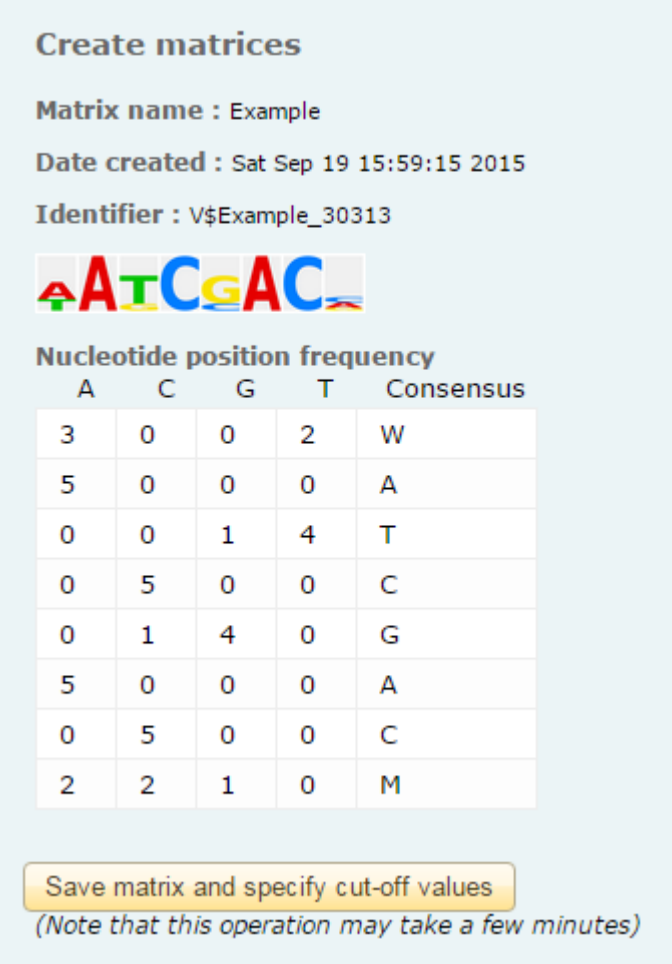
You have the option to:

- Upload a matrix in TRANSFAC file format
- Compile a matrix from unaligned sequences
- Compile a matrix from aligned sequences

To compile a matrix from unaligned sequences, you need to provide a set of positive sequences (those that are expected to contain the motif) and a set of negative sequences (those that are not expected to contain the motif).

To compile a matrix from aligned sequences, you need to provide a minimum of five sequences that are expected to contain the motif and are of the exact same length.

Once you have chosen the relevant method and supplied the requested information, click the “create matrix” button. The consensus sequence and nucleotide position frequency is displayed:



Create matrices

Matrix name : Example

Date created : Sat Sep 19 15:59:15 2015

Identifier : V$Example_30313

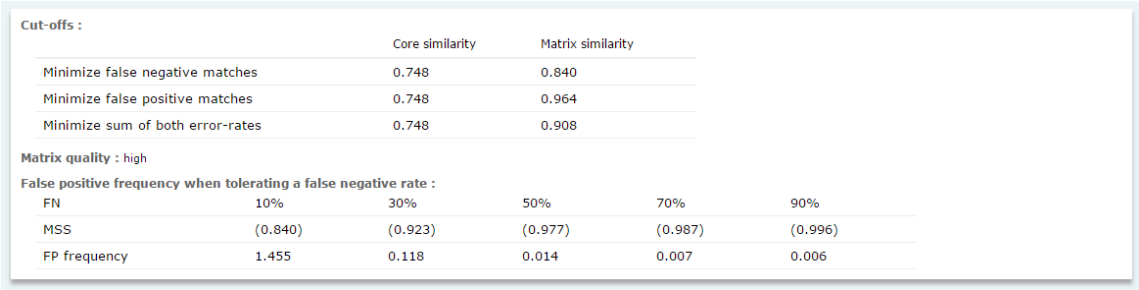
Nucleotide position frequency

A	C	G	T	Consensus
3	0	0	2	W
5	0	0	0	A
0	0	1	4	T
0	5	0	0	C
0	1	4	0	G
5	0	0	0	A
0	5	0	0	C
2	2	1	0	M

Save matrix and specify cut-off values

(Note that this operation may take a few minutes)

Click the “Save matrix and specify cut-off values” button to complete the process of calculating the cut-offs. It may take a few minutes, but once the calculation is complete the cut-offs will be displayed and you will be notified that your matrix has been saved:



Cut-offs :

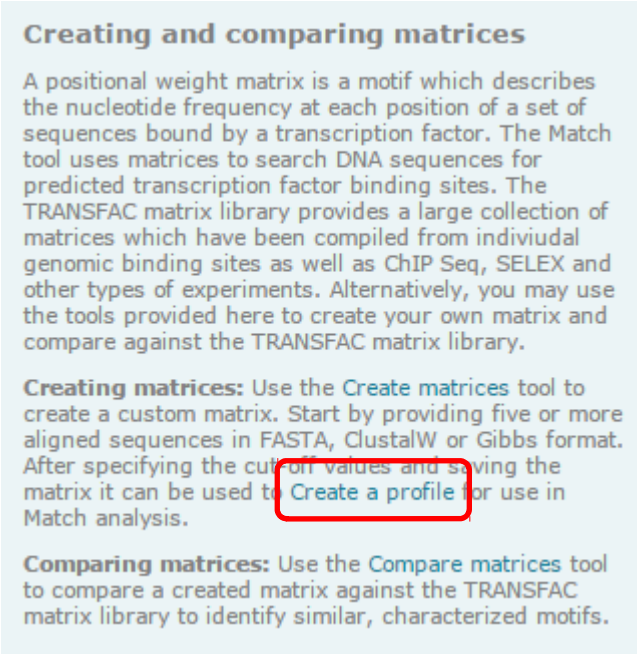
	Core similarity	Matrix similarity
Minimize false negative matches	0.748	0.840
Minimize false positive matches	0.748	0.964
Minimize sum of both error-rates	0.748	0.908

Matrix quality : high

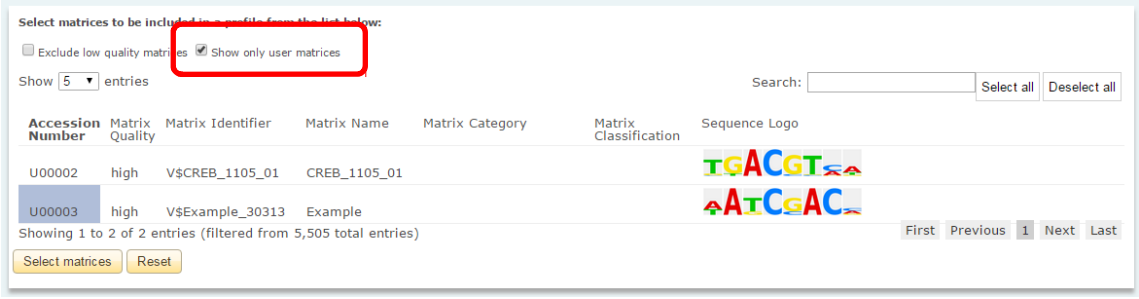
False positive frequency when tolerating a false negative rate :

FN	10%	30%	50%	70%	90%
MSS	(0.840)	(0.923)	(0.977)	(0.987)	(0.996)
FP frequency	1.455	0.118	0.014	0.007	0.006

In order to use the matrix in an analysis you will need to take the additional step of creating a profile that contains the matrix. From the help menu on the right hand side of the screen, click the “Create a profile” link:

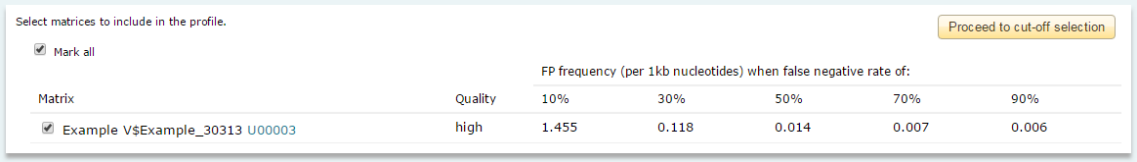


When the Create profile tool loads you can quickly find your matrix by clicking the “Show only user matrices” option:

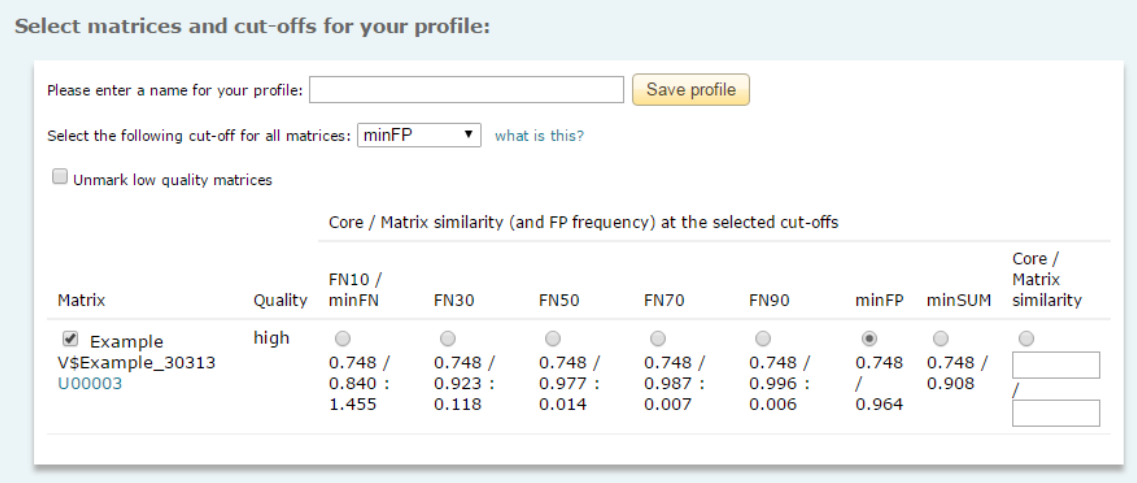


Click on your matrix to select it and then click the “Select matrices” button. It will be

displayed below with the calculated cut-offs:



Click the “Proceed to cut-off selection” button. A new window loads where you are able to specify a name for your profile as well as set the default cut-off to be used in analyses:

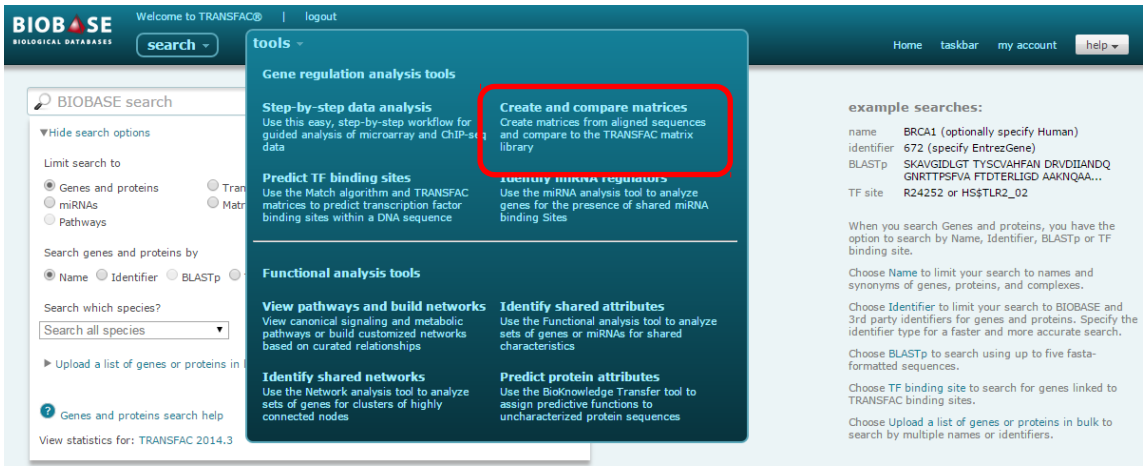


The suggested default is minFP. Click the “Save profile” button to save the profile and make it available in the matrices pull-down menu of all analyses.

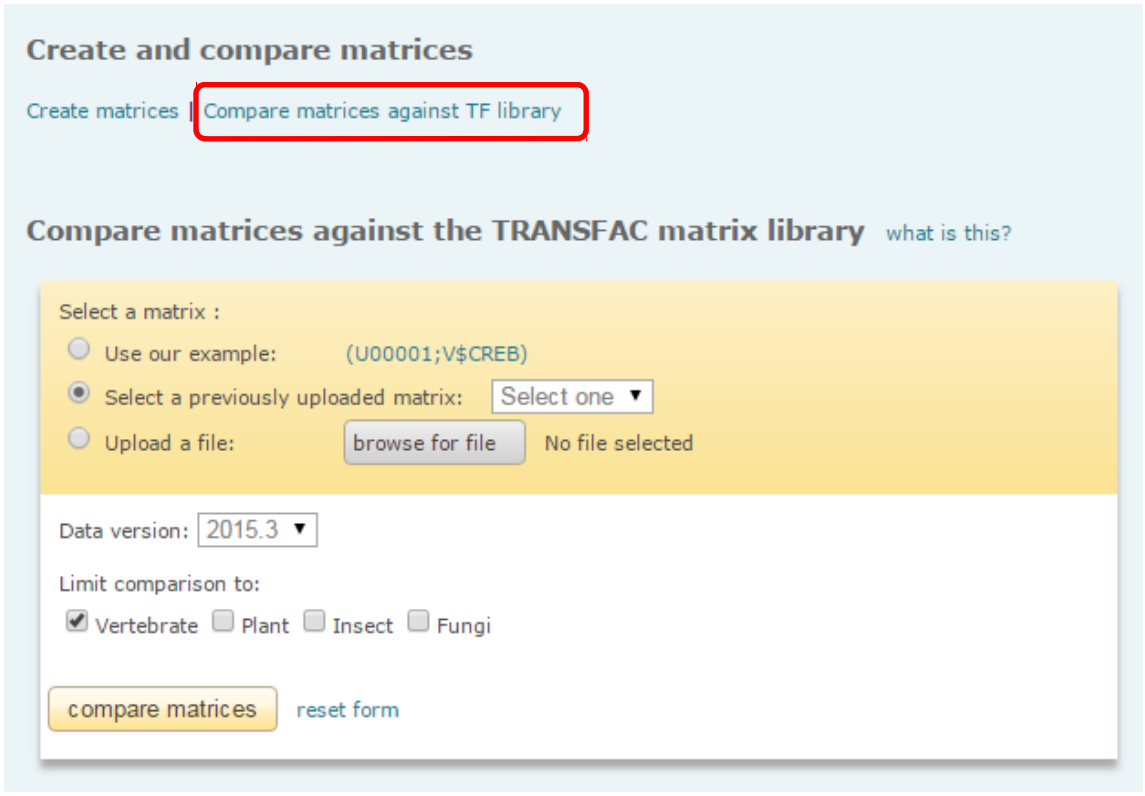
Compare your own custom matrix against TRANFAC’s matrix library

If you have uploaded a custom matrix (see Import your own custom matrix for use in binding site predictions) you can use the Compare matrices tool to investigate whether your matrix is similar to other matrices within the TRANSFAC matrix library.

Select the “Create and compare matrices” option from the tools menu:



When the tool loads, click the “Compare matrices against TF library” option to refresh the view if needed:



Select your previously uploaded matrix from the pull-down menu, select the data version (use the default current version) and select whether you want to limit your comparison to vertebrate, plant, insect or fungal matrices. Click the “compare matrices” button. The screen will be refreshed and the results of the comparison will be provided in a table:

Result what is this?

Create and compare matrices

Create matrices | Compare matrices against TF library

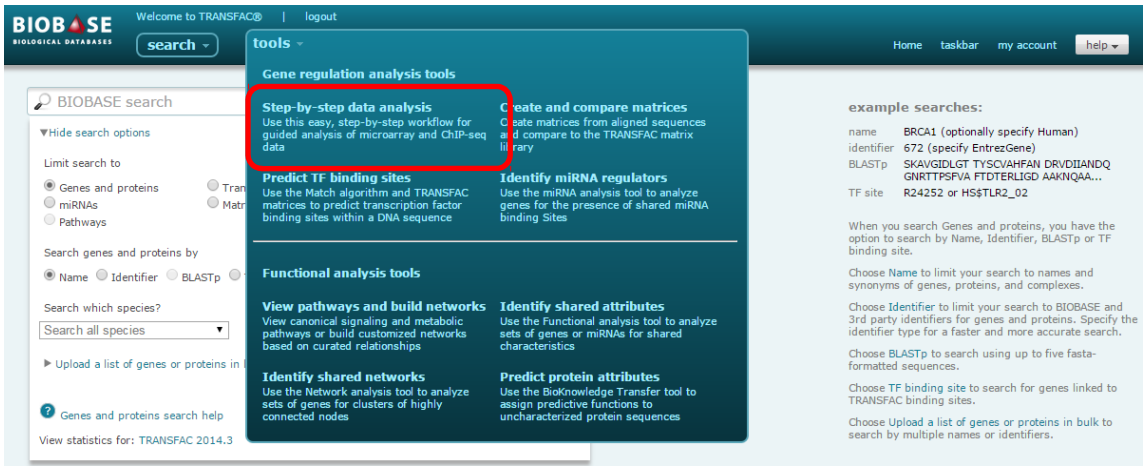
Query Matrix	Transfac Id	Matrix accession	Matrix class	Alignment	P-value	FDR
U00001	V$CREB_01	M00039	BZIP	View	1.1913E-10	5.2096E-7
U00001	V$ATF_B	M00338	BZIP	View	6.6155E-8	1.2175E-4
U00001	V$CREBP1_Q2	M00179	BZIP	View	8.3561E-8	1.2175E-4
U00001	V$ATF2_Q6	M07312	BZIP	View	1.5165E-7	1.6576E-4
U00001	V$CREB1_01	M02279	BZIP	View	4.8909E-7	3.8012E-4
U00001	V$CREB_Q4	M00178	BZIP	View	5.2214E-7	3.8012E-4
U00001	V$CREB1_Q6	M03544	BZIP	View	1.1678E-6	5.0606E-4
U00001	V$CREB_Q4_01	M00917	BZIP	View	1.2472E-6	5.0606E-4
U00001	V$ATF2_Q5	M01862	BZIP	View	1.2598E-6	5.0606E-4
U00001	V$ATF1_03	M02738	BZIP	View	1.3240E-6	5.0606E-4

Up to ten best matching matrices will be listed. Click the “View” link in the Alignment column to see a graphical alignment of the motifs:

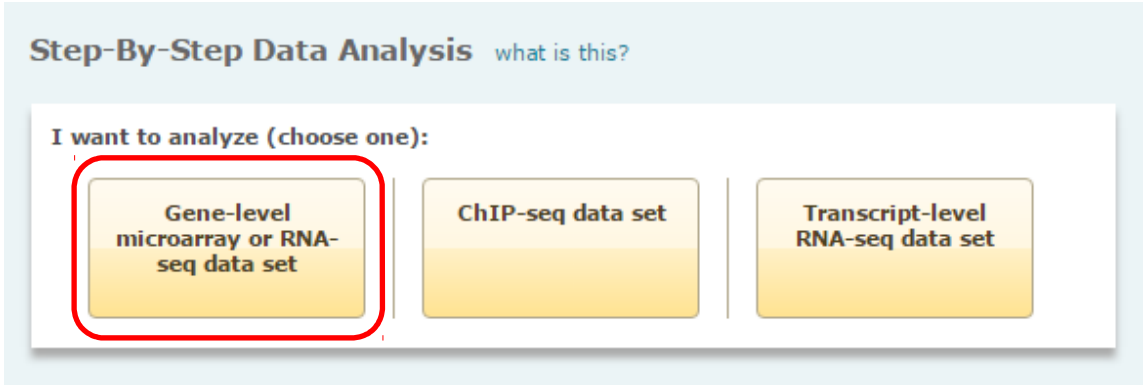
Analyze a microarray or gene-based RNA-seq data set

Use TRANSFAC’s step-by-step analysis workflow for gene-level microarray and RNA-seq data sets to identify the transcription factors that are most likely to be responsible for the differential expression pattern observed in your experimental data as compared to background data.

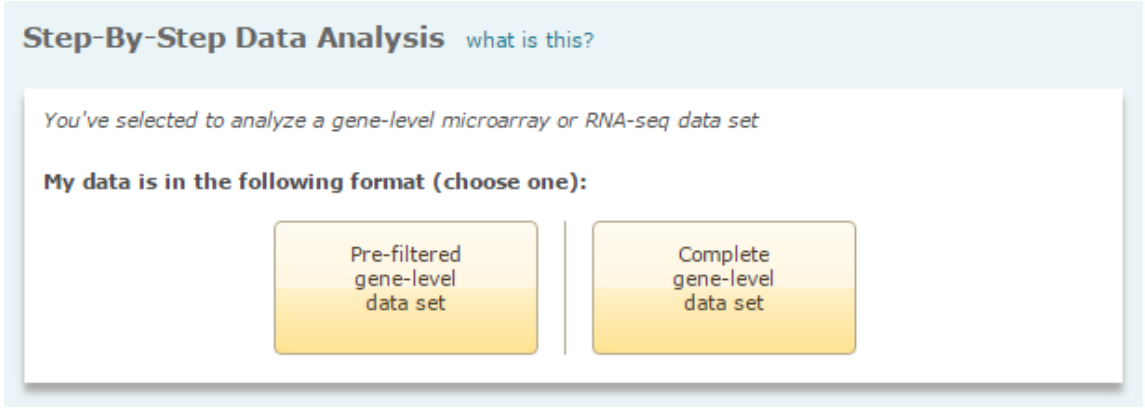
Select the “Step-by-step data analysis” options from the tools menu:



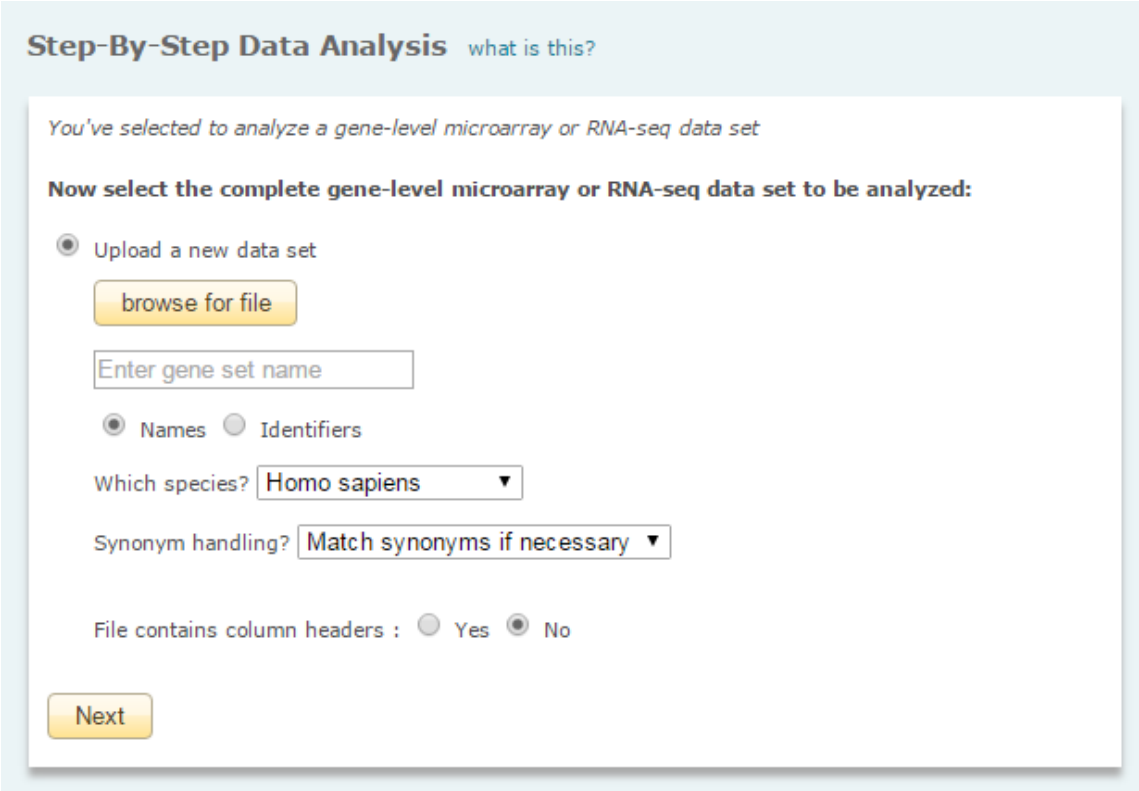
Then select the Gene-level microarray or RNA-seq data set option:



Specify whether your data is in the form of a pre-filtered list (only the subset of genes that were identified as up- or down-regulated is included) or whether you are uploading the complete data set including genes that showed no change in expression:

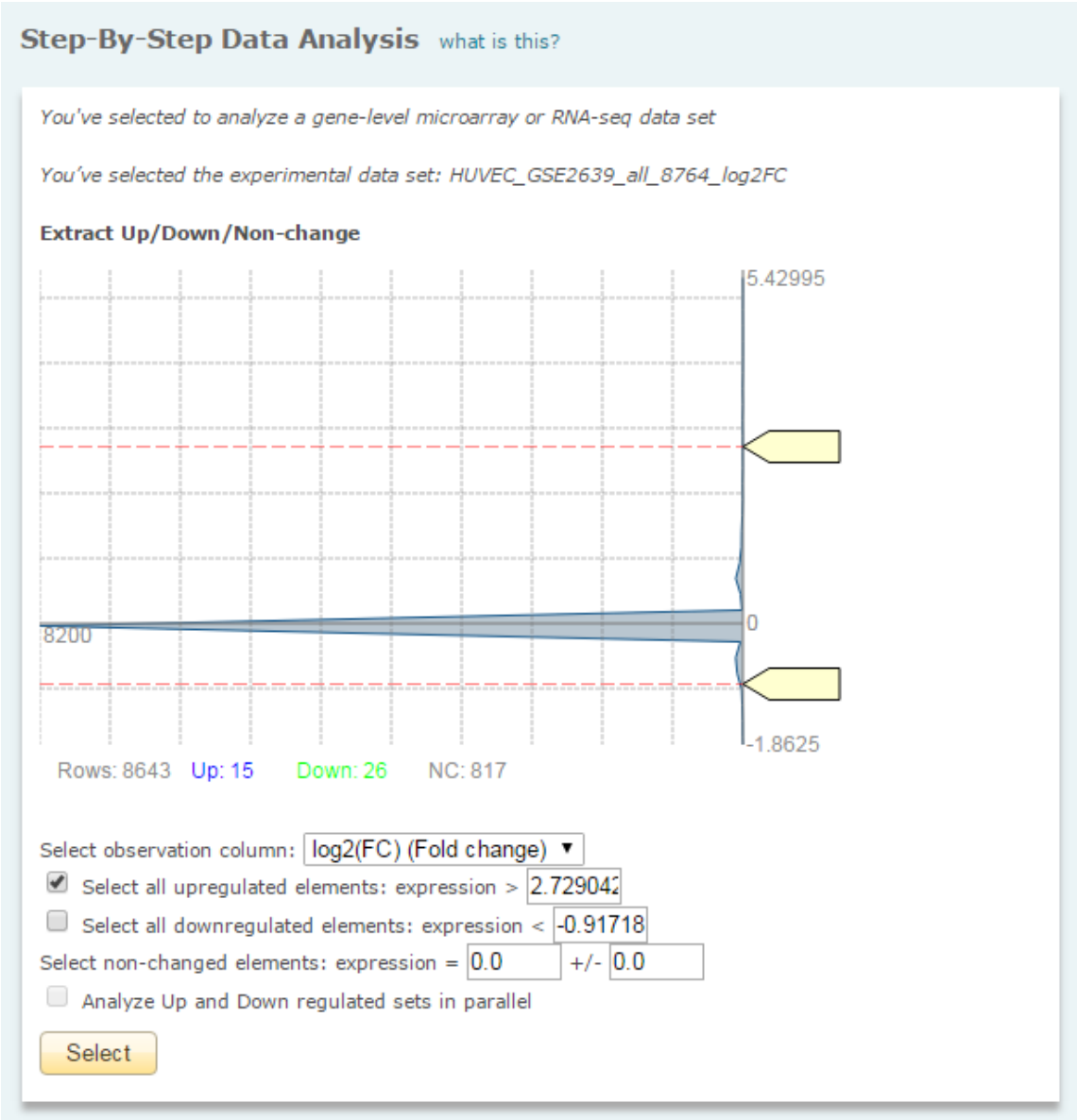


Upload your data set (in this example we are uploading the complete data set):



Note that once you've run at least one analysis you will have the additional option to select a previously uploaded data set instead of uploading a new data set.

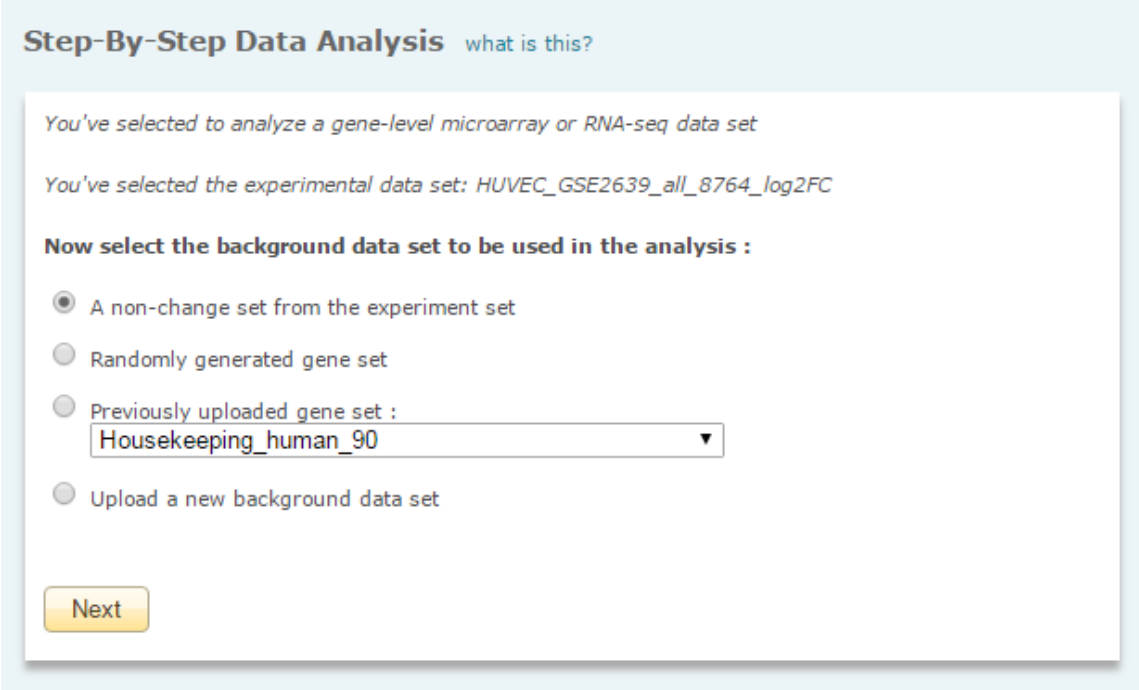
You will be given a preview of the list of matched genes and then will be forwarded to a screen where you will specify the criteria for selecting the up-regulated, down-regulated and non-changed gene sets:



In this example, we have chosen to select all up-regulated genes which show a log2 fold change of >1 as our experimental set and to select all genes which show a log2 fold change of 0 as the non-changed set.

Note that if your data is uploaded in the form of a pre-filtered list this step will automatically be skipped.

Next you will specify the background set to use:

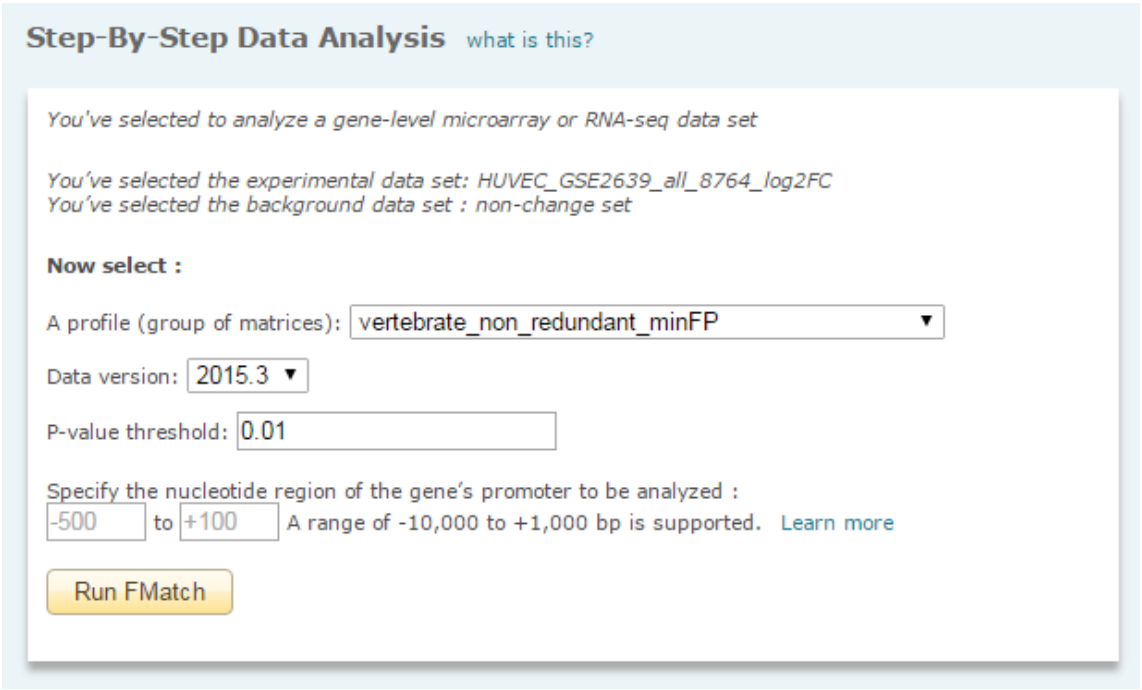


You have multiple options for the background set including:

- Using the non-change genes from the data set that you just uploaded, (the default, recommended option)
- Using a randomly generated gene set
- Using a previously uploaded gene set
- Uploading a new gene set

Note that if your data is in the form of a pre-filtered list you will not have the option to directly use the non-change genes from the data set as they were not included, you will only have the latter three options.

Next you will accept or change the default analysis parameters:



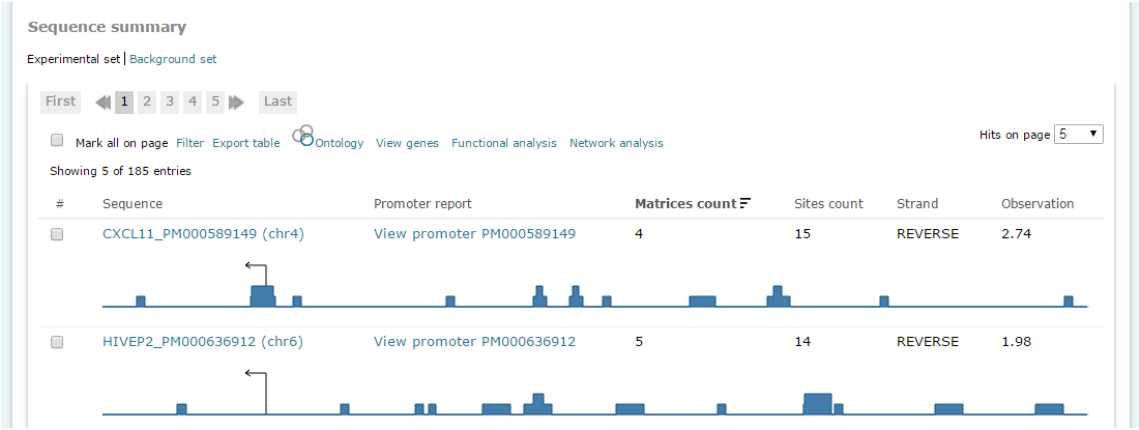
You will:

- Specify the profile (group of matrices) to be used (use the default "vertebrate_non_redundant_minFP" which contains representative matrices for distinctive binding motifs unless you are working with non-vertebrate species or want to limit your analysis to factors that are active in a particular tissue)
- Specify the data version (use the default, current version)
- Specify the P-value threshold (0.01 is the default)
- Specify the region of the gene's promoter to be considered in the analysis (a range of -10,000 bp to +1,000 bp relative to the TSS is supported, but the default range of -500 bp to +100 bp relative to the TSS is recommended for most analyses)

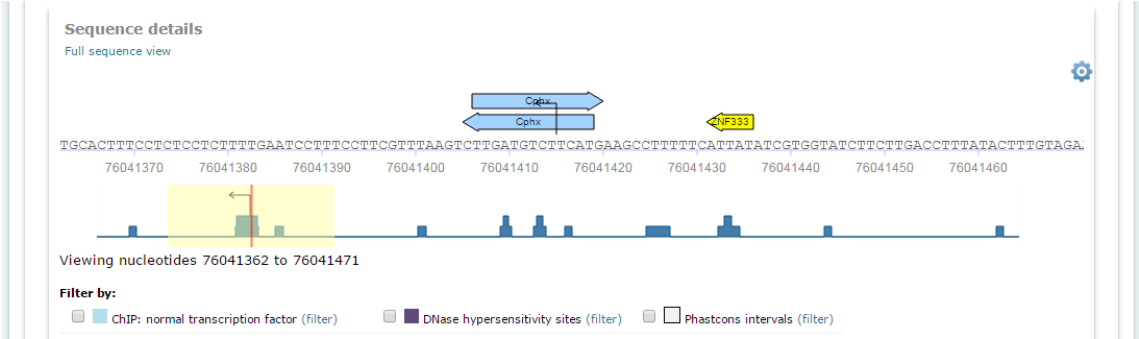
When you click the "Run FMatch" button your analysis will be forwarded to the taskbar. Once the analysis completes, the FMatch Analysis Report will automatically load.

The "Matrix summary" section provides a tabular overview of all matrices that were found to have significantly more binding sites within your experimental data set than were found in the background data set:

Scrolling down to the “Sequence details” section a graphical view of the results is provided:



Click the Sequence name to display the nucleotide level view for a specific promoter:



A frequency bar displays the distribution of sites along the entire sequence, with the red orientation point and flanking yellow highlighting indicating the portion of the sequence that is displayed at the nucleotide level above. Slide the red orientation point along the frequency bar to view other portions of the sequence in detail.

Optionally use the ChIP fragments, DNase hypersensitivity sites and Phastcons intervals filters to apply additional biological context to your results. Activating these filters removes any sites that do not overlap with regions of sequence that are suggested to be available

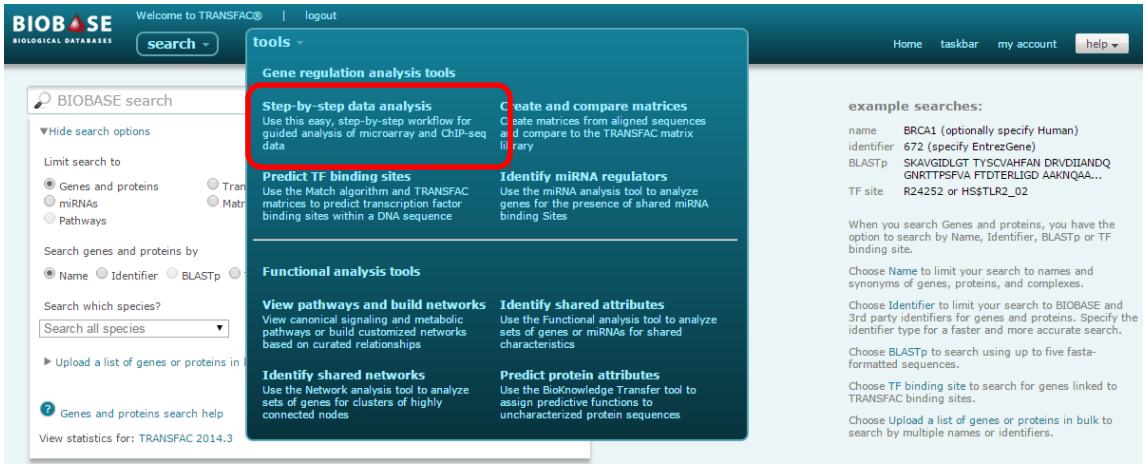
to the transcriptional machinery based on being bound by transcription factors (ChIP fragments), being accessible to DNase enzymes (DNase hypersensitivity) or conserved between human and mouse (Phastcons).

Your analysis will be preserved in the taskbar for 7 days. To preserve it indefinitely, click the “Save this report” link at the top of the report.

Analyze a ChIP-seq data set

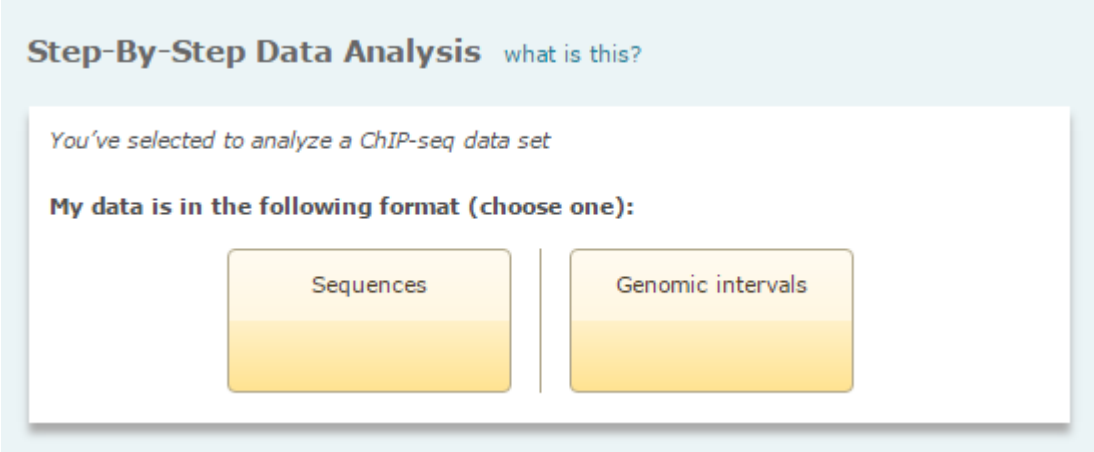
Use TRANSFAC’s step-by-step analysis workflow for ChIP-seq data sets to identify the transcription factors that are most likely to be responsible for the differential expression pattern observed in your experimental data as compared to background data.

Select the “Step-by-step data analysis” options from the tools menu:

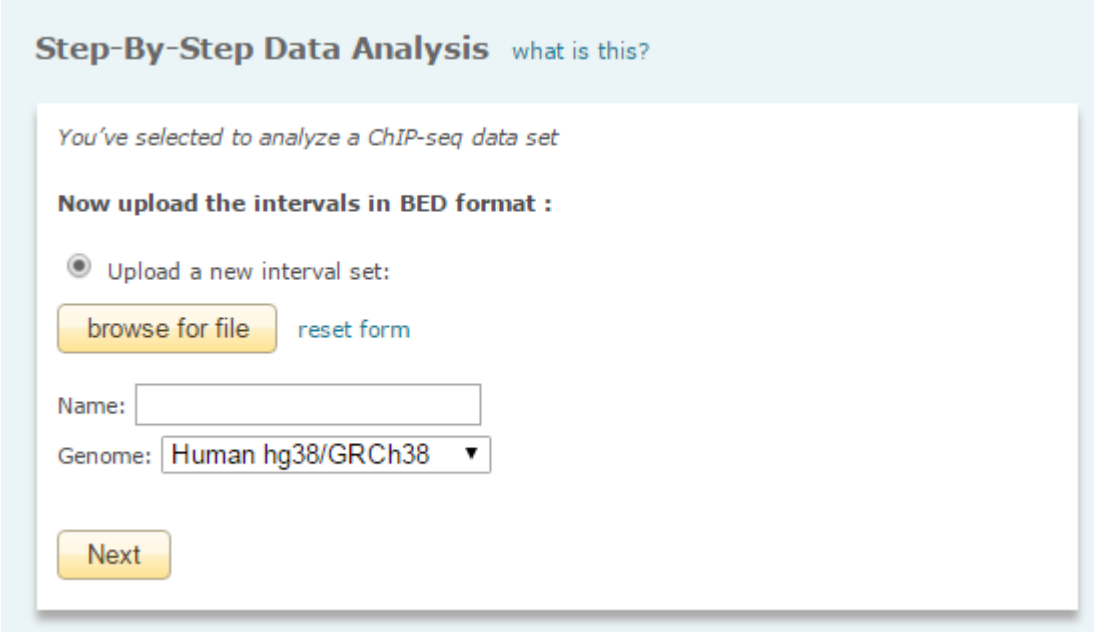


Then select the ChIP-seq data set option:

Specify whether your data is in the format of sequences (FASTA, EMBL, Genbank and RAW formats are supported) or .bed formatted genomic intervals:

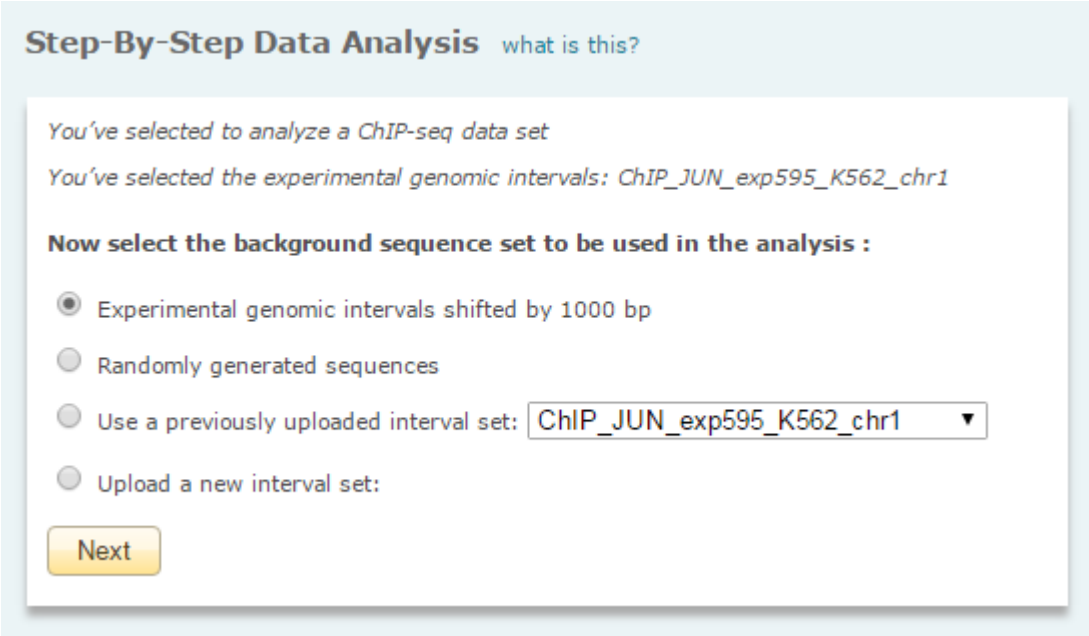


Upload your data set (in this example we are uploading a set of genomic intervals):



Note that once you've run at least one analysis you will have the additional option to select a previously uploaded data set instead of uploading a new data set.

Next you will specify the background set to use:

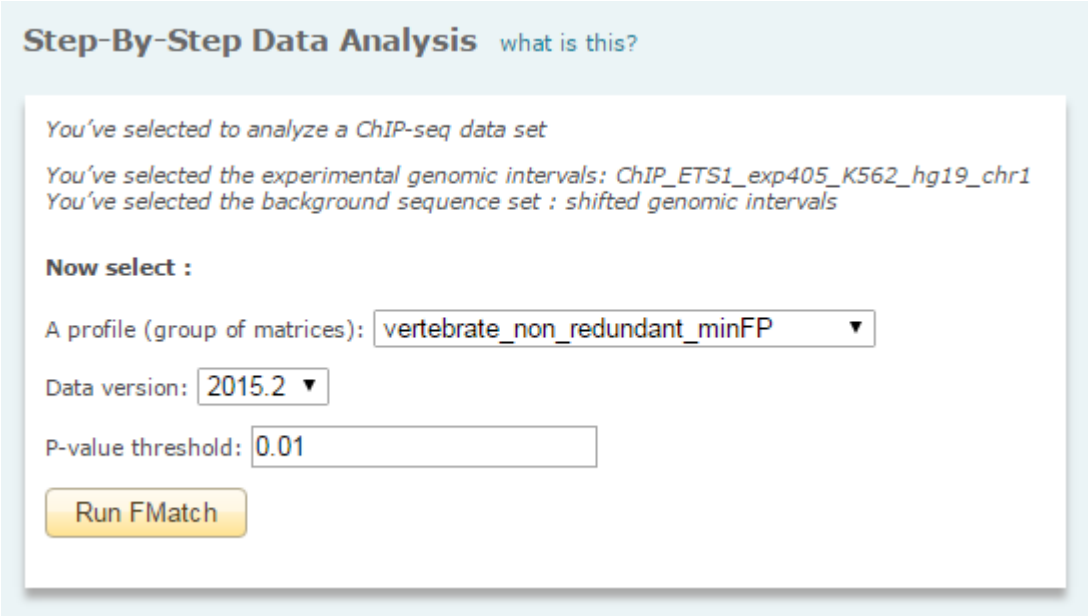


You have multiple options for the background set including:

- Using the data set that you just uploaded, shifted by 1,000 bp (the default, recommended option)
- Using a randomly generated sequence set
- Using a previously uploaded interval set
- Uploading a new interval set

Note that if your data is uploaded as sequences you will not have the option to shift the data set by 1,000 bp, you will only the latter three options.

Next you will accept or change the default analysis parameters:



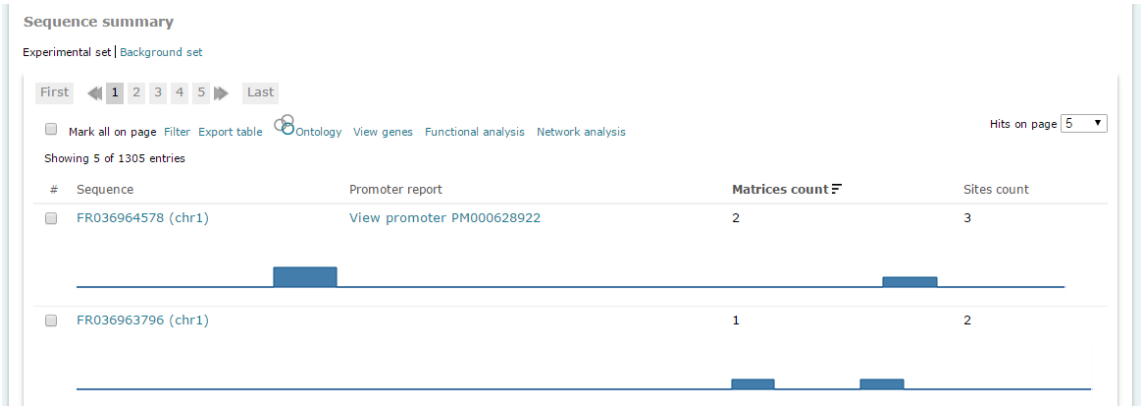
You will:

- Specify the profile (group of matrices) to be used (use the default "vertebrate_non_redundant_minFP" which contains representative matrices for distinctive binding motifs unless you are working with non-vertebrate species or want to limit your analysis to factors that are active in a particular tissue)
- Specify the data version (use the default, current version)
- Specify the P-value threshold (0.01 is the default)

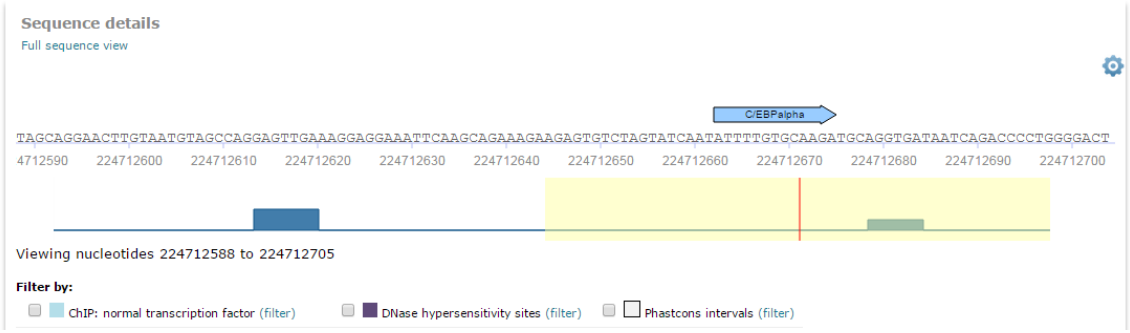
When you click the "Run analysis" button your analysis will be forwarded to the taskbar. Once the analysis completes, the FMatch Analysis Report will automatically load.

The "Matrix summary" section provides a tabular overview of all matrices that were found to have significantly more binding sites within your uploaded data set than were found in the background data set:

Scrolling down to the "Sequence details" section a graphical view of the results is provided:



Click the Sequence name to display the nucleotide level view for a specific sequence:



A frequency bar displays the distribution of sites along the entire sequence, with the red orientation point and flanking yellow highlighting indicating the portion of the sequence that is displayed at the nucleotide level above. Slide the red orientation point along the frequency bar to view other portions of the sequence in detail.

Optionally use the ChIP fragments, DNase hypersensitivity sites and Phastcons intervals filters to apply additional biological context to your results. Activating these filters removes any sites that do not overlap with regions of sequence that are suggested to be available to the transcriptional machinery based on being bound by transcription factors (ChIP fragments), being accessible to DNase enzymes (DNase hypersensitivity) or conserved between human and mouse (Phastcons). Note that this option will not be available if you uploaded your data as a set of FASTA or other formatted sequences.

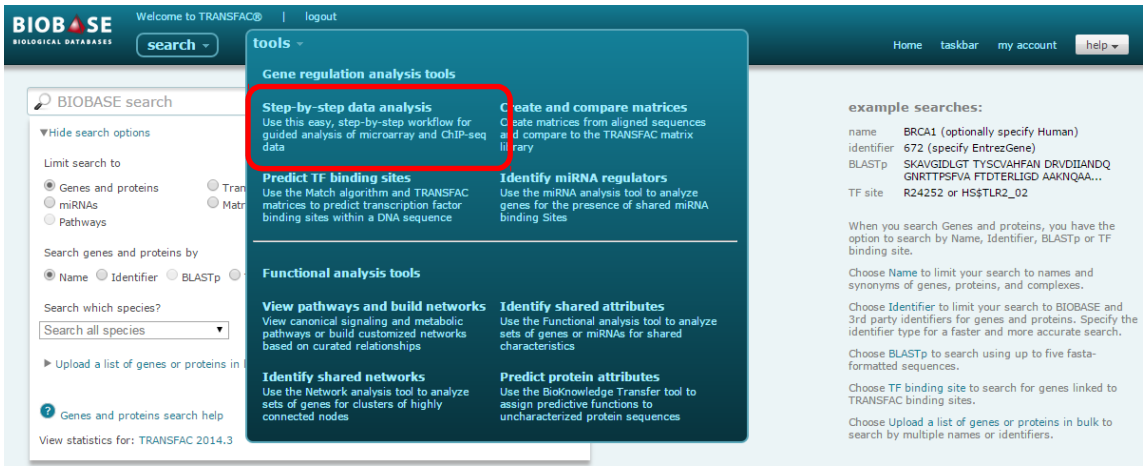
Your analysis will be preserved in the taskbar for 7 days. To preserve it indefinitely, click the “Save this report” link at the top of the report.

Analyze a transcript-based RNA-seq data set

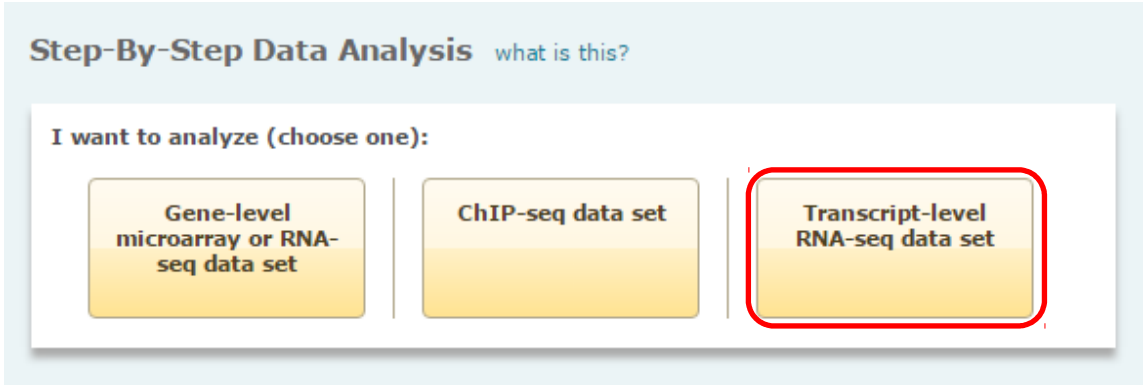
Use TRANSFAC’s step-by-step analysis workflow for transcript-level RNA-seq data sets to

identify the transcription factors that are most likely to be responsible for the differential expression pattern observed in your experimental data as compared to background data.

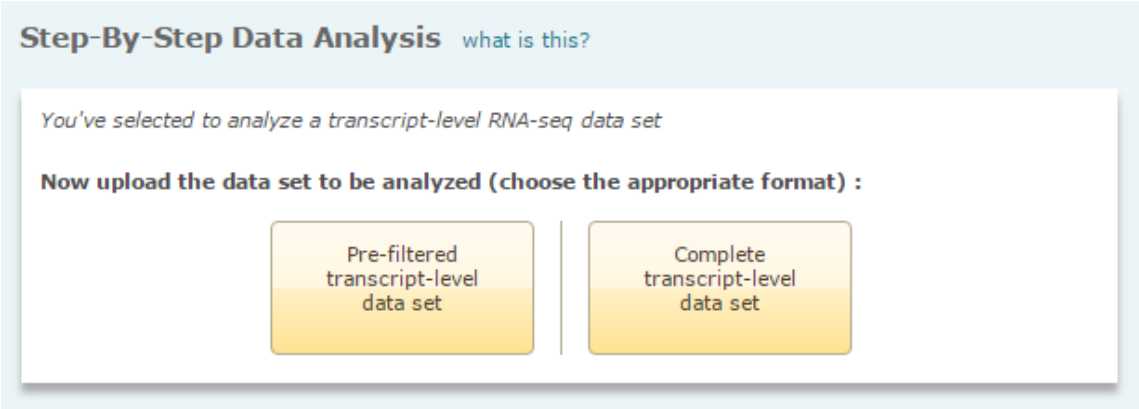
Select the “Step-by-step data analysis” options from the tools menu:



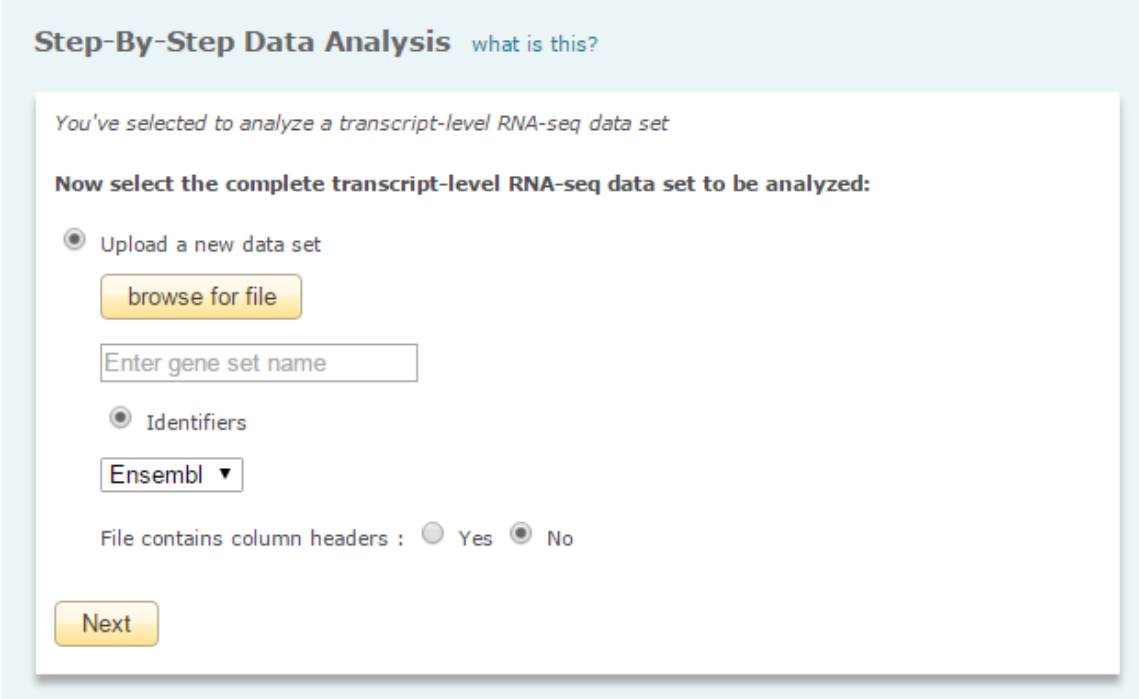
Then select the Transcript-level RNA-seq data set option:



Specify whether your data is in the form of a pre-filtered list (only the subset of transcripts that were identified as up- or down-regulated is included) or whether you are uploading the complete data set including transcripts that showed no change in expression:

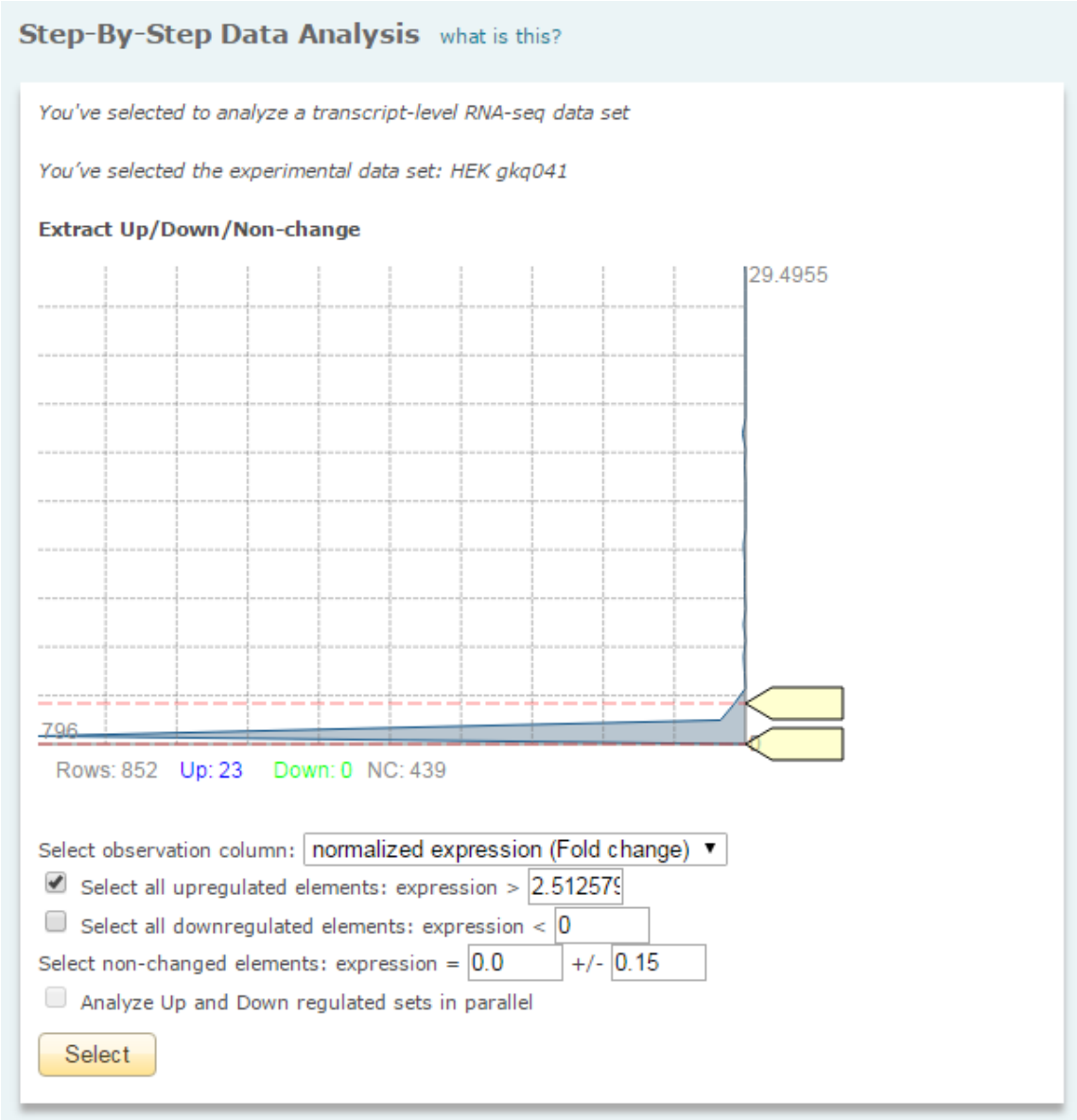


Upload your data set (in this example we are uploading the complete data set):



Note that once you've run at least one analysis you will have the additional option to select a previously uploaded data set instead of uploading a new data set.

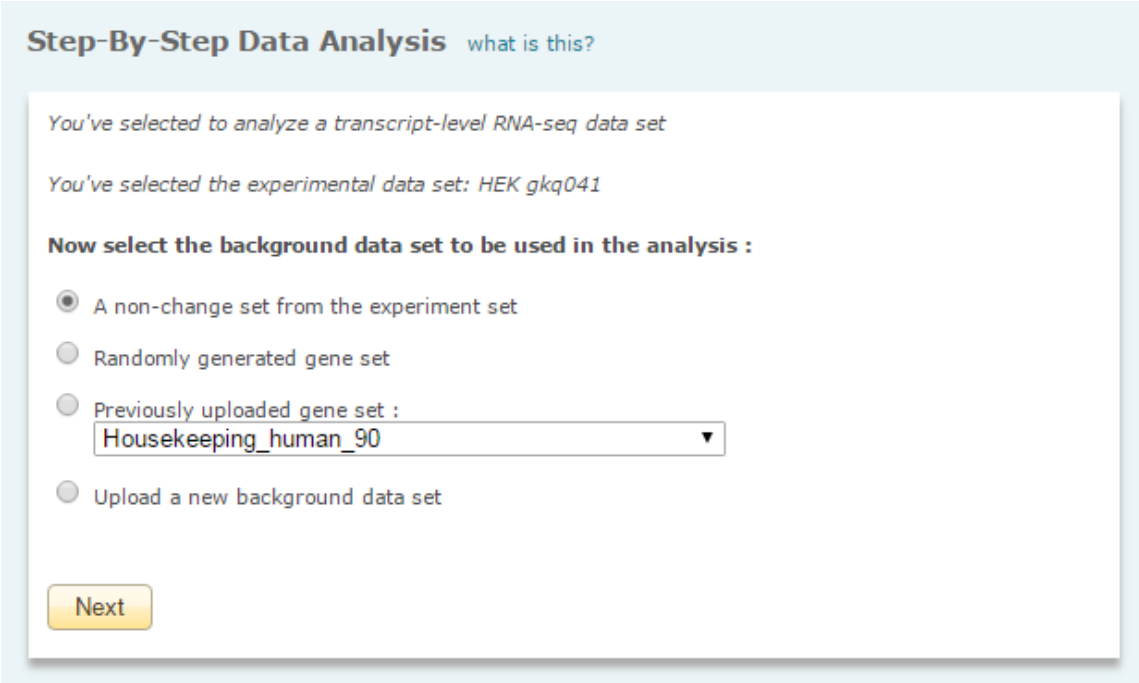
Note that with the initial release only Ensembl transcript identifiers are supported.
You will be given a preview of the list of matched transcripts and then will be forwarded to a screen where you will specify the criteria for selecting the up-regulated, down-regulated and non-changed gene sets:



In this example, we have chosen to select all up-regulated transcripts which show a normalized expression value of >2.51 as our experimental set and the set of transcripts which show a normalized expression value of 0 +/- 0.15 as our non-changed transcript set.

Note that if your data is uploaded in the form of a pre-filtered list this step will automatically be skipped.

Next you will specify the background set to use:

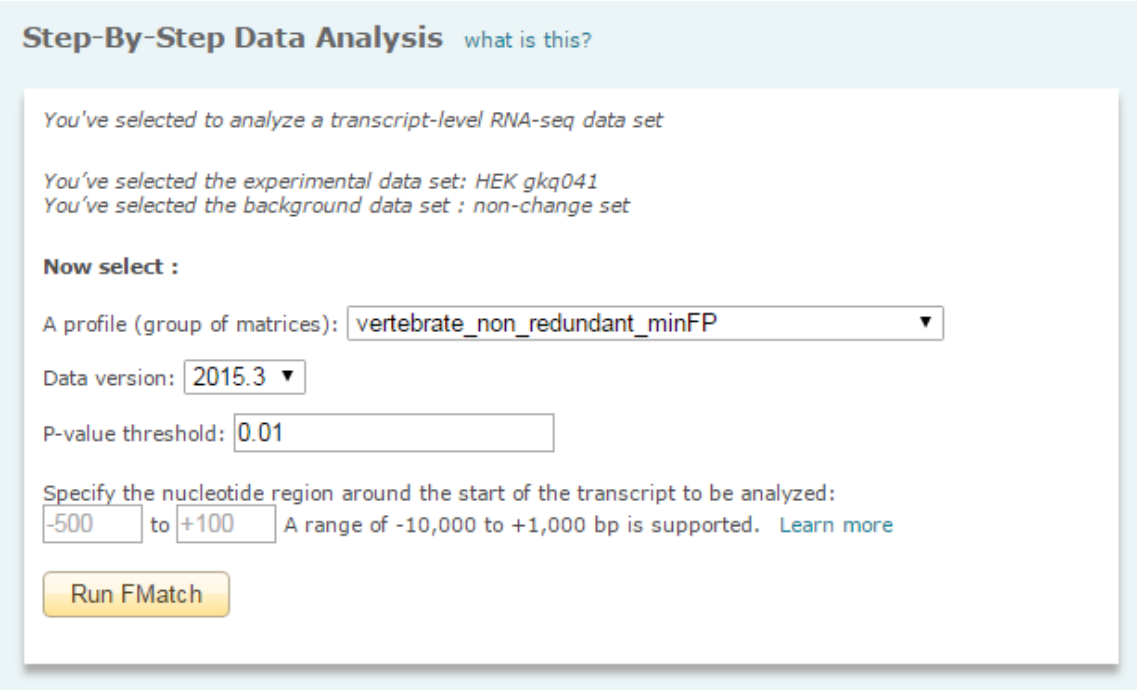


You have multiple options for the background set including:

- Using the non-change transcripts from the data set that you just uploaded, (the default, recommended option)
- Using a randomly generated transcript set
- Using a previously uploaded transcript set
- Uploading a new transcript set

Note that if your data is in the form of a pre-filtered list you will not have the option to directly use the non-change transcripts from the data set as they were not included, you will only have the latter three options.

Next you will accept or change the default analysis parameters:



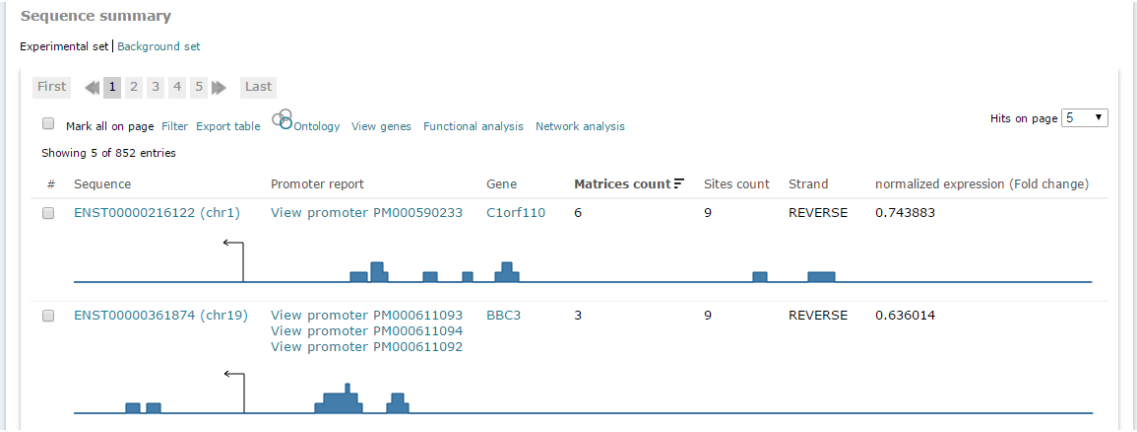
You will:

- Specify the profile (group of matrices) to be used (use the default "vertebrate_non_redundant_minFP" which contains representative matrices for distinctive binding motifs unless you are working with non-vertebrate species or want to limit your analysis to factors that are active in a particular tissue)
- Specify the data version (use the default, current version)
- Specify the P-value threshold (0.01 is the default)
- Specify the region surrounding the TSSs (determined by the Ensembl transcripts your data was mapped to) to be considered in the analysis (a range of -10,000 bp to +1,000 bp relative to the TSS is supported, but the default range of -500 bp to +100 bp relative to the TSS is recommended for most analyses)

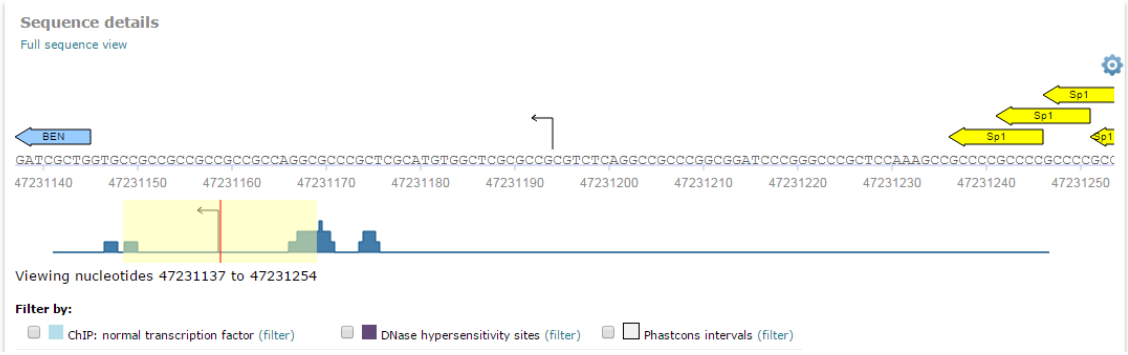
When you click the "Run FMatch" button your analysis will be forwarded to the taskbar. Once the analysis completes, the FMatch Analysis Report will automatically load.

The "Matrix summary" section provides a tabular overview of all matrices that were found to have significantly more binding sites within your experimental data set than were found in the background data set:

Scrolling down to the “Sequence details” section a graphical view of the results is provided, along with a summary of promoters and genes that are linked to the transcript:



Click the Sequence name to display the nucleotide level view for a specific transcript:



A frequency bar displays the distribution of sites along the entire sequence, with the red orientation point and flanking yellow highlighting indicating the portion of the sequence that is displayed at the nucleotide level above. Slide the red orientation point along the frequency bar to view other portions of the sequence in detail.

Optionally use the ChIP fragments, DNase hypersensitivity sites and Phastcons intervals filters to apply additional biological context to your results. Activating these filters removes

any sites that do not overlap with regions of sequence that are suggested to be available to the transcriptional machinery based on being bound by transcription factors (ChIP fragments), being accessible to DNase enzymes (DNase hypersensitivity) or conserved between human and mouse (Phastcons).

Your analysis will be preserved in the taskbar for 7 days. To preserve it indefinitely, click the “Save this report” link at the top of the report.