

Report on gene regulation analysis

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 20/08/2026; Run on 20/08/2026; Report generated on 21/08/2026
MATCH Suite release 3.3 (TRANSFAC® 2.0 release 2026.1)

Summary

The MATCH Suite software package was applied to the fruit fly input gene list [demo-gene-list](#) that contained 314 genes. The goal of the analysis was to identify the transcription factors that regulate the studied set of genes through their -500/+100 promoter region.

The most promising transcription factors regulating the input gene set appeared to be factors: **brk**, **ZIPIC**, **M1BP**, **DHR3**, **Hr46**, **REPTOR-BP**, **cg**, **hng1**, **Mad**, **ewg**. The selection of these factors was based on a complex criteria that includes: (1) the statistically significant enrichment of the binding sites for these factors in the promoters of the studied set of genes and (2) combinatorial effect of several TFs acting together binding to co-localized binding sites in these promoters.

At least 249 genes from your input set are regulated by these factors.

Results overview

Transcription Factors Identified

In the promoters of the analyzed genes, potential transcription factor binding sites were identified and checked for (a) statistical enrichment and (b) for overrepresented combinations (see [Methods](#)).

From these analyses, the transcription factors presented in the Table 1 were identified as the most probable regulators of the studied gene set.

Table 1. Key transcription factors identified as the potential regulators of the analyzed gene set.

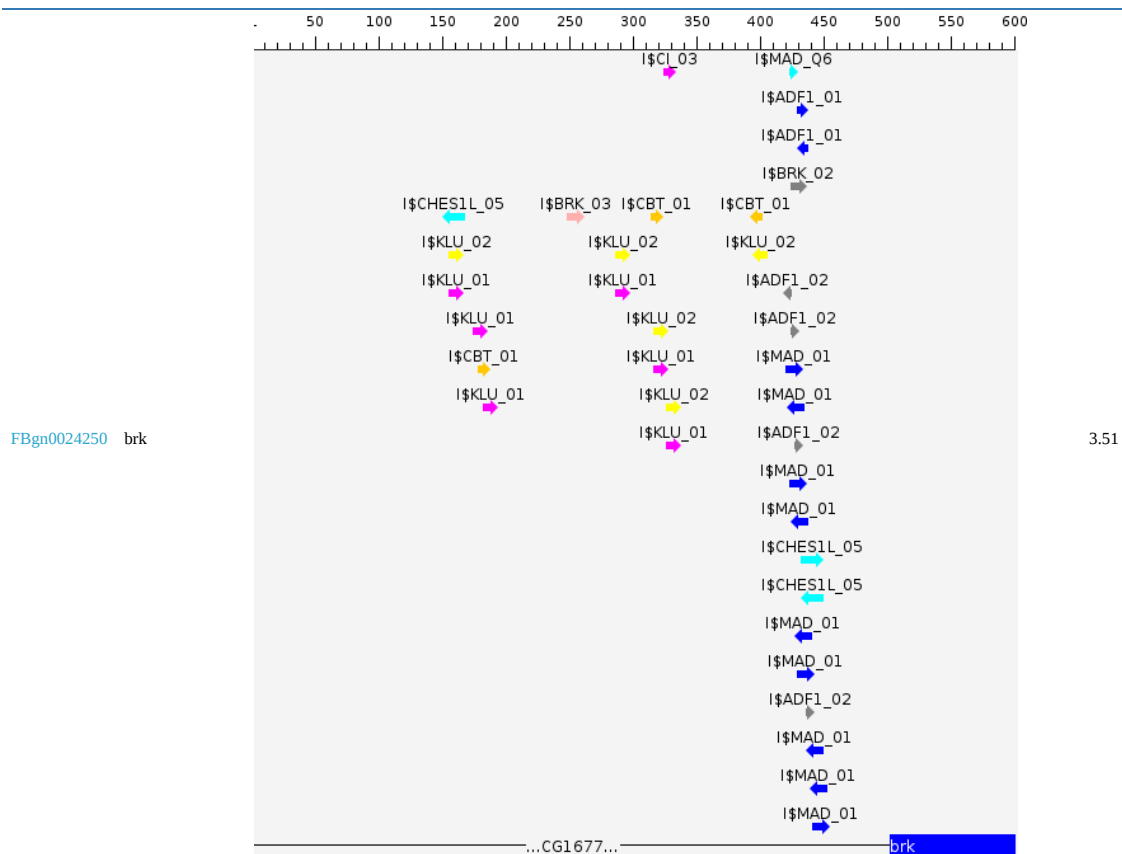
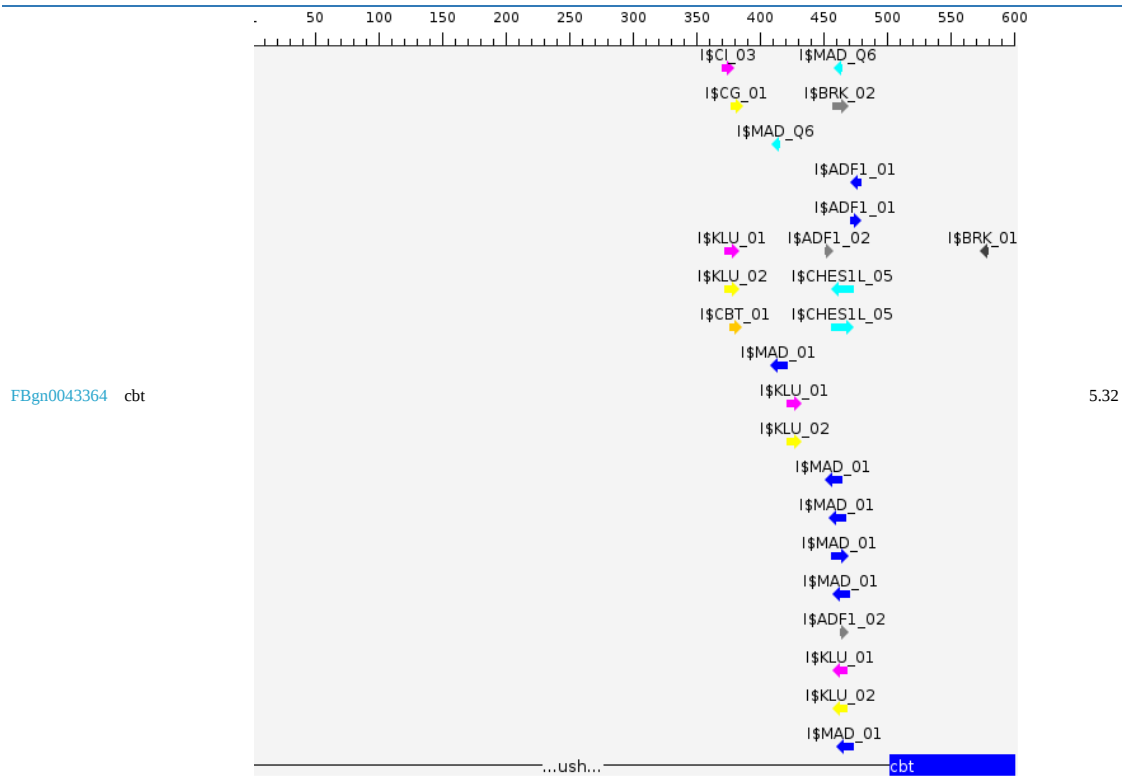
Factor name	Enrichment analysis ⓘ	Combinatorial analysis ⓘ
brk	●	✓
ZIPIC	●	✓
M1BP	●	✓
DHR3	●	✓
Hr46	●	✓
REPTOR-BP	●	–
cg	●	✓
hng1	●	–
Mad	●	✓
ewg	●	–
fd59A	●	✓
Atf3	●	–
Adf1	●	✓
DREF	●	✓
BEAF-32	●	–
pnr	●	–
Ci	●	✓
Sry-delta	●	–
klu	●	–
checkpoint suppressor homologue	●	–

[View full table](#) →

Table 1 shows the transcription factors the binding sites of which are enriched in the analyzed promoters. The sites that have sequence enrichment FDR less than 0.05 are marked with green color in the column 'Enrichment analysis', the rest are marked with blue. Those factors that additionally are part of a combinatory module are marked by a tick in the 'Combinatorial Analysis' column.

Regulation of Analyzed Genes

Not all transcription factors identified act equally on all genes in the analyzed gene set. Table 2 shows the analyzed genes with respective site models (PWMs) found in their promoters. The target genes are ranked according to the enrichment of their promoters by binding sites of the best ranked transcription factors. The visualization shows the binding sites of matrices from CMA model and all other enriched sites.



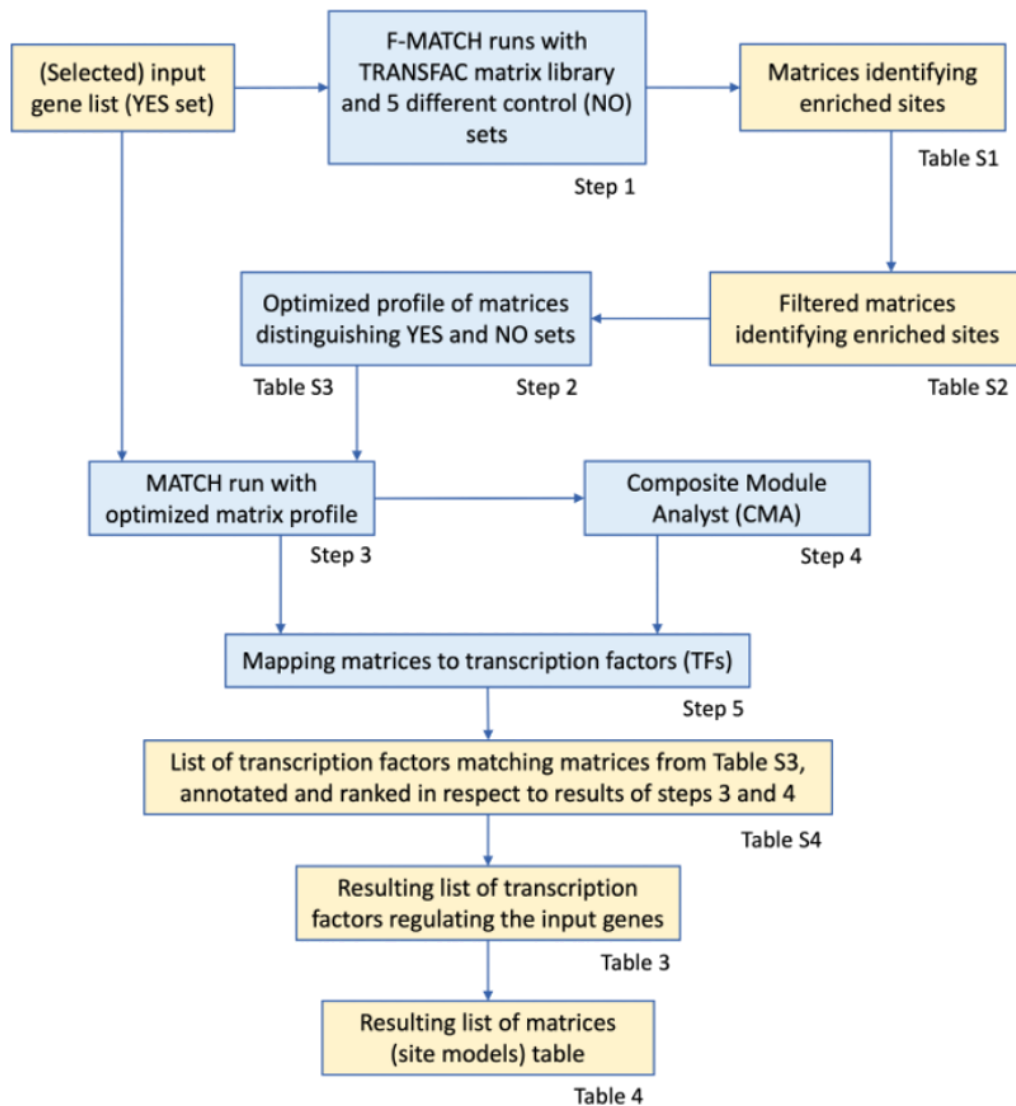


[View full table](#) →

Results details

The analysis workflow of the MATCH Suite comprises 5 steps, which are outlined in Figure 1. Each step produces a table either of intermediary or of final results, available as Supplementary table following the links given, or as part of this report, respectively.

Figure 1. Overall schema of the MATCH Suite workflow



Step 1

The promoters of the input gene set selected for specified GO categories were extracted using the Ensembl best supported promoter. The promoter regions used were taken as -500 bp upstream of transcription start site (TSS) and +100 bp downstream the TSS. The extracted promoters were compiled to one track of regulatory regions, which was then used to perform site enrichment analysis using the TRANSFAC® library of positional weight matrices (PWMs). It is done according to the PWM cut-off optimization approach published earlier [1,2]. The geneXplain platform method 'Search for enriched TFBS on tracks' [2] was used here to analyze site enrichment in the promoters of the input gene set.

The YES set (analysis set) applied in this analysis consisted of the promoters of the input genes, while five NO sets (control/ background sets), each comprising 1000 promoters, were randomly sampled from genes not belonging to your input gene set. Transcription factor binding sites (TFBSs) enriched in the YES set were identified by 5 independent F-MATCH runs, each using one of the five NO sets. The matrix profile used for identifying potential TFBSs is the collection of 667 TRANSFAC® fruit fly matrices [3, 4].

For each matrix of the profile, the cutoff value of the Match Site Score (MSS) is optimized so that an optimal enrichment of sites in the YES compared to the NO set is achieved. From the 5 independent runs, the median of these site cutoff values is computed [2].

Supplementary table 1 (Table S1) comprises the results of this analysis, by default sorting the matrices according to the enrichment of sites they identified. 85 transcription factor binding site models (matrices) exhibited a site enrichment higher than 1 (lower boundary of the 99% confidence interval of the site enrichment). See Methods for a detailed explanation of the contents of this table.

Step 2

The list of matrices (Table S2) is used to construct a new matrix profile that is specific for the analysis of the input gene set and will be used for the next analyses. The cut-offs for the profile are selected as the median site cutoff of matrices from Table S2. The complete profile is shown in Supplementary table 3 (Table S3).

Step 3

With the next step, the MATCH Suite workflow uses the constructed profile from Step 2 for another search for potential transcription factor binding sites (TFBS) in the set of studied genes. The method uses the MATCH algorithm (see [5] and Methods for further info) and generates tracks of sites found in the YES and NO set.

Step 4

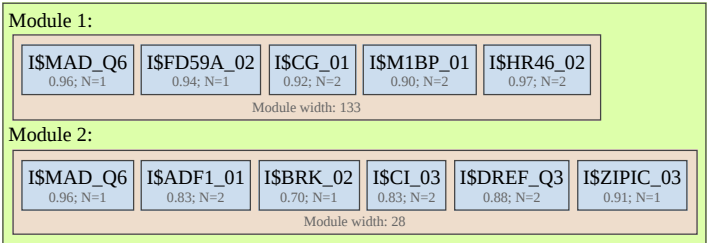
Subsequently, the MATCH Suite workflow searches for composite modules of the predicted TFBSs. Composite modules are combinations of several TFBSs that are found together in a set of regulatory sequences. Combinations of TF binding sites that are overrepresented in the regulatory regions of the genes in the YES compared to those in the NO set are identified by the Composite Module Analyst (CMA - see [6] and the [Methods](#) for further info). The genetic algorithm takes the output from the site search in Step 3 as input and comes up with a resulting composite module that differentiates the YES set from the background NO set. CMA identifies the transcription factors that through their cooperation may provide a synergistic effect and thus have a great influence on the gene regulatory process.

Figure 2 shows the CMA model constructed on the basis of found YES and NO sites tracks. The obtained CMA model is then applied to compute CMA score for all of the genes from the input set.

The model consists of 2 module(s). Below, for each module the following information is shown:

- PWMs producing matches,
- number of individual matches for each PWM,
- score of the best match.

Figure 2. The constructed CMA model



Model score ($-p \cdot \log_{10}(pval)$): 12.29

Wilcoxon p-value (pval): 5.86e-26

Penalty (p): 0.487

Average yes-set score: 2.26

Average no-set score: 1.12

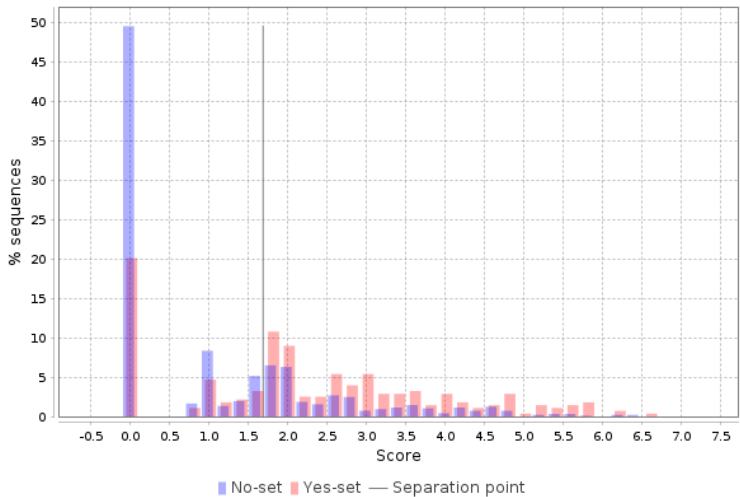
AUC: 0.71

Separation point: 1.69

False-positive: 32.61%

False-negative: 32.26%

Figure 3. Distribution of CMA scores for individual promoters



The histogram from Figure 3 shows the distribution of scores for individual promoters where the promoter score value is shown on X axis and the percentage of promoters (% sequences) having this score is shown on the Y axis. The center, a vertical gray line (separation point), corresponds to the average score value. Promoters from the NO set with a score above the separation line, i.e. blue bars to the right of the gray line, are referred to as false positives. Promoters from the YES set with a score below the separation line (red bars to the left of the gray line), are false negatives.

The YES promoters with a score above the separation line are very well separated from the NO promoters, which means that for this part of the promoters the composite model constructed is most suitable. See [Methods](#) section for further info.

Step 54

The transcription factors associated with the matrices received on steps 3 and 4 are given in [Supplementary table 4 \(Table S4\)](#). The sorting of the factors takes into account the combinatorial score and enrichment value of the respective binding sites. Table 3 contains the top 30 factors from Supplementary table 4 together with the factors corresponding to the matrices from the CMA model. These factors are predicted to be regulating the input gene set.

Table 3. Transcription factors that regulate the analyzed gene set.

Factor name	Gene symbol	Class name and TF classification	Site model	Adjusted factor enrichment	Factor rank
brk	brk		ISBRK_01, ISBRK_02 , ISBRK_03	2.22	1
ZIPIC	ZIPIC		ISZIPIC_01, ISZIPIC_02, ISZIPIC_03	1.68	2
M1BP	M1BP		ISM1BP_01	1.66	3
DHR3	Hr3		ISHR46_02	1.47	4
Hr46	Hr3		ISHR46_02	1.47	4
REPTOR-BP	REPTOR-BP		ISREPTORBP_01	1.85	6
cg	cg		ISCG_01	1.37	7
hng1	hng1		ISHNG1_01, ISHNG1_02	1.74	8
Mad	Mad	SMAD/NF-1 DNA-binding domain factors N/A	ISMAD_01, ISMAD_Q6	1.34	9
ewg	ewg		ISEWG_01	1.50	10

View full table →

In addition to the TF name, its gene symbol, its numerical identifier in the TF classification [7] and the description of the class it belongs to, Table 3 also shows all matrices from steps 3 and 4 that refer to the respective factor in the 'Site model' column. The matrices that belong to the CMA combinatorial modules are displayed in bold. The factor enrichment value is derived from the maximum enrichment value among all matrices referring to the factor. The length of the bar in the 'Factor enrichment' column is proportional to the maximum site enrichment value of the factor's matrices. The color of this bar is green if the maximum sequence enrichment FDR of the factor's matrices is less than 0.05 and blue otherwise.

The matrices corresponding to the factors from Table 3 are listed in Table 4. These are the most relevant matrices (site models) that determine the regulation of the analyzed gene set.

Table 4. Resulting matrices (site models) table

ID	Matrix logo	Site enrichment (adjusted enrichment)	Site enrichment FDR	Adjusted sequence enrichment	Sequence enrichment FDR	Composite model	Site rank
ISZIPIC_03		2.84 (1.68)	9.77E-6	1.59	1.51E-4	yes	1
ISM1BP_01		2.54 (1.66)	1.76E-7	1.5	5.85E-5	yes	2
ISHR46_02		2.22 (1.47)	5.37E-6	1.39	2.02E-4	yes	3
ISCG_01		1.87 (1.37)	1.32E-7	1.1	5.32E-3	yes	4
ISFD59A_02		2.59 (1.24)	5.87E-3	1.09	1.81E-2	yes	5
ISADF1_01		1.60 (1.23)	2.31E-7	1.09	7.57E-4	yes	6
ISDREF_Q3		1.56 (1.20)	1.47E-6	1.12	2.35E-3	yes	7
ISBRK_02		1.49 (1.17)	5.4E-7	0.97	9.99E-3	yes	8
ISCI_03		1.68 (1.14)	2.4E-3	0.91	2.87E-2	yes	9
ISMAD_Q6		1.43 (1.11)	2.72E-5	0.97	2.88E-3	yes	10
ISBRK_03		8.32 (2.22)	3.11E-4	1.58	9.27E-3		11
ISREPTORBP_01		5.27 (1.85)	7.79E-4	1.38	1.03E-2		12
ISHNG1_02		3.20 (1.74)	2.85E-5	1.15	1.42E-2		13
ISHNG1_01		3.38 (1.67)	2.4E-4	0.83	3.54E-2		14
ISZIPIC_02		2.66 (1.57)	5.02E-5	1.51	2.77E-4		15
ISEWG_01		2.79 (1.50)	3.91E-4	1.22	4.6E-3		16
ISZIPIC_01		3.06 (1.42)	1.92E-3	1.45	3.97E-3		17
ISATF3_01		2.11 (1.34)	2.17E-4	0.74	6.13E-2		18
ISMAD_01		2.04 (1.34)	3.75E-5	0.9	2.55E-2		19
ISBEAF32_01		1.73 (1.32)	5.02E-9	1.21	9.19E-4		20

View full table →

The list of all genes from the input set, their CMA scores, the total number of identified TFBSs and the hits obtained with each site model are presented in Table 5. Underneath each matrix name, the TFs referring to it are given in the order of their ranks. The ranking of genes is done according to their CMA scores.

Table 5. Gene table with the identified transcription factors and their site models regulating the genes from the analyzed gene set

Ensembl ID ②	Gene symbol ②	Gene description ②	CMA Score ②	Total number of sites ②	ISADF1_01 Adf1	ISADF1_02 Adf1	ISATF3_01 Atf3	ISBEAF32_01 BEAF-32	ISBEAF32_02 BEAF-32	ISBRK_01 brk	ISBRK_02 brk	ISBRK_03 brk	ISCBT_cbt
FBgn0030766	mtl1	methuselah-like 1	6.61	15	2	1	0	2	0	0	0	0	1
FBgn0043900	pygo	pygopus	6.21	26	0	0	0	2	2	0	2	0	0
FBgn0050418	nord	nord	6.13	42	3	2	0	0	0	0	7	0	3
FBgn0025800	Smox	Smad on X	5.88	20	1	0	0	2	0	1	0	0	0
FBgn0029067	Dd	Dullard	5.83	62	1	1	0	2	3	0	4	0	14
FBgn0261983	l(2)gd1	lethal (2) giant discs 1	5.8	46	1	0	2	4	4	0	4	0	0
FBgn0034012	Hr51	Hormone receptor 51	5.77	57	3	4	0	0	0	0	2	0	12
FBgn0010313	corto	corto	5.73	20	0	0	0	0	0	0	1	0	4
FBgn0004390	RasGAP1	Ras GTPase activating protein 1	5.68	13	0	0	0	0	2	0	0	0	0
FBgn0260397	Su(var)3-3	Suppressor of variegation 3-3	5.61	23	2	0	0	2	2	0	0	0	1
FBgn0003444	smo	smoothened	5.59	28	0	1	0	4	3	0	3	0	0
FBgn0000711	flw	flapwing	5.52	25	1	0	0	2	2	0	4	0	0
FBgn0003118	pnt	pointed	5.42	75	3	3	0	0	0	0	10	0	11
FBgn0043364	cbt	cabut	5.32	39	2	2	0	0	0	1	14	0	1
FBgn0285879	disco-r	disco-related	5.31	40	1	4	0	0	0	0	8	1	4
FBgn0004885	tok	tolkin	5.22	35	0	0	0	4	3	0	2	0	0
FBgn0025633	spdi	split discs	5.2	12	3	1	0	0	0	0	0	0	0
FBgn0004647	N	Notch	5.17	11	3	1	0	0	0	0	2	0	0
FBgn0035719	tow	target of wingless	5.11	49	1	1	0	4	4	0	4	0	5
FBgn0004638	drk	downstream of receptor kinase	5.09	10	1	1	0	0	0	0	2	0	0

[View full table →](#)

References

- [1] Kel A, Voss N, Jauregui R, Kel-Margoulis O, Wingender E. (2006) Beyond microarrays: find key transcription factors controlling signal transduction pathways. BMC Bioinformatics. 7 Suppl 2(Suppl 2), S13. [PubMed](#).
- [2] Koschmann J, Bhar A, Stegmaier P, Kel AE, Wingender E. “Upstream analysis”: an integrated promoter-pathway analysis approach to causal interpretation of microarray data. Microarrays. 2015;4:270–86. [PubMed](#).
- [3] Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., Wingender, E. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34:D108-D110. [PubMed](#).
- [4] Wingender, E. (2008) The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. Brief. Bioinform. 9:326-332. [PubMed](#).
- [5] Kel, A.E., Gössling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., Wingender, E. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. 31:3576-3579. [PubMed](#).
- [6] Waleev, T., Shtokalo, D., Konovalova, T., Voss, N., Cheremushkin, E., Stegmaier, P., Kel-Margoulis, O., Wingender, E., Kel, A. (2006). Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. Nucleic Acids Research, 34(suppl_2):W541-W545. [PubMed](#).
- [7] Wingender, E., Schoeps, T., Haubrock, M., Krull, M., Dönitz, J. (2018) TFClass: expanding the classification of human transcription factors to their mammalian orthologs. Nucleic Acids Res. 46:D343-D347. [PubMed](#).

How to cite

Please use the results received with the MATCH Suite in your publications or presentations with the following reference:

The results were obtained with the MATCH Suite software integrated into the TRANSFAC® 2.0 solution for gene regulation analysis release 3.3 (<https://genexplain.com/transfac>).

Please also provide reference to the following publication:

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., Wingender, E. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34:D108-D110. [PubMed](#).

Disclaimer

The results produced by the MATCH Suite, contained in any of the reports or results visualization produced by this software, are based on the best of geneXplain’s knowledge, however, we do not guarantee completeness and reliability of this information. GeneXplain GmbH does not guarantee comprehensiveness, reliability or accuracy of the information contained in the reports generated by the MATCH Suite.