Report on gene regulation analysis

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 11/04/2022; Run on 11/04/2022; Report generated on 11/04/2022
MATCH Suite release 2.0 (TRANSFAC® 2.0 release 2022.1)

Summary

The MATCH Suite software package was applied to the input gene BRCA1 (ENSG00000012048). The study was performed in the scope of testis tissue. The goal of the analysis was to identify the transcription factors that regulate the studied gene through binding to the regulatory regions of this gene containing its promoter and associated enhancers and silencers. We analyzed the most active promoter in the range of [−500,+100] bp from TSS and all known enhancers and silencers associated with this gene and identified in the testis tissue.

The most promising transcription factors regulating the input gene appeared to be factors: **RFX4, CREM, RFX2, RFX3, CREB, JARID1B, KLF17, ARID2, RFX1, ETV2**. The selection of these factors was based on a complex criteria that includes: (1) estimation of cumulative binding affinity of the transcription factors to the analyzed regulatory regions, and (2) high as well as specific expression of these factors in the testis tissue.

Results overview

Transcription Factors Identified

In the regulatory regions of the input gene, potential transcription factor binding sites were identified and checked for cumulative binding affinity (see Methods).

From these analyses, the transcription factors presented in the Table 1 were identified as the most probable regulators of the BRCA1 (ENSG00000012048) gene.

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

Factor name	Gene symbol	Class name and TF classification	Site model	-log(affinity p-value)	Factor rank	Testis: factor expression	Testis: expression difference (rank)
RFX4	RFX4	Fork head / winged helix factors 3.3.3.0.4	V$RFX3_11	5.45	1	70.5	65.1 1/62
CREM	CREM	Basic leucine zipper factors (bZIP) 1.1.7.1.3	V$CREM_01	3.61	2	74.9	60.4 1/62
RFX2	RFX2	Fork head / winged helix factors 3.3.3.0.2	V$RFX3_11	5.45	3	51.5	44.7 1/62
RFX3	RFX3	Fork head / winged helix factors 3.3.3.0.3	V$RFX3_11	5.45	4	36.5	27.9 1/62
CREB	CREB1	Basic leucine zipper factors (bZIP) 1.1.7.1.1	V$CREM_01	3.61	5	39.2	20.6 1/62
JARID1B	KDM5B	ARID domain factors 3.7.1.6.2	V$ARID5A_04	3.16	6	74.0	63.8 1/62
KLF17	KLF17	C2H2 zinc finger factors 2.3.1.2.17	V$TIEG1_Q6	2.48	7	31.5	30.7 1/62
ARID2	ARID2	ARID domain factors 3.7.1.2.1	V$ARID5A_04	3.16	8	30.9	20.4 2/62
RFX1	RFX1	Fork head / winged helix factors 3.3.3.0.1	V$RFX3_11	5.45	9	21.0	11.7 2/62
ETV2	ETV2	Tryptophan cluster factors 3.5.2.1.3	V$GABPA_Q4	2.07	10	40.3	37.4 1/62

View full table →

Table 1 shows the transcription factors with cumulative binding affinity to the studied regulatory regions. The input gene BRCA1 (ENSG00000012048) had 0 enhancer(s) and 0 silencer(s), that are known to be active in the testis tissue. Only the promoter region of the input gene was analyzed. Together with the TF name, its gene symbol, its numerical identifier in the TF classification [1], and the description of the class it belongs to, Table 1 also shows the best matrix that refers to the respective factor in the 'Site model' column. The '-log(affinity p-value)' column shows the respective value based on the minimum affinity p-value of the best matrix of the factor. The length of the bar is proportional to the value itself and the color is green in case the Match Site score of the best factor's matrix is higher than 0,9 and blue otherwise (see Methods for details).

The ranking of factors is based on the summed rank that includes ranking by -log(affinity p-value), ranking by factor's expression level in the testis tissue, and ranking by its expression in the testis tissue compared to all other supported tissues.

The 'testis: factor expression' column shows the relative expression value of the factor in testis tissue as provided by the Protein Atlas. The 'testis: expression difference (rank)' column shows the difference between factor expression in the testis tissue and the average value of factor expression among all tissues. The color of the bar refers to the factor's general expression specificity level and uses the following color code:

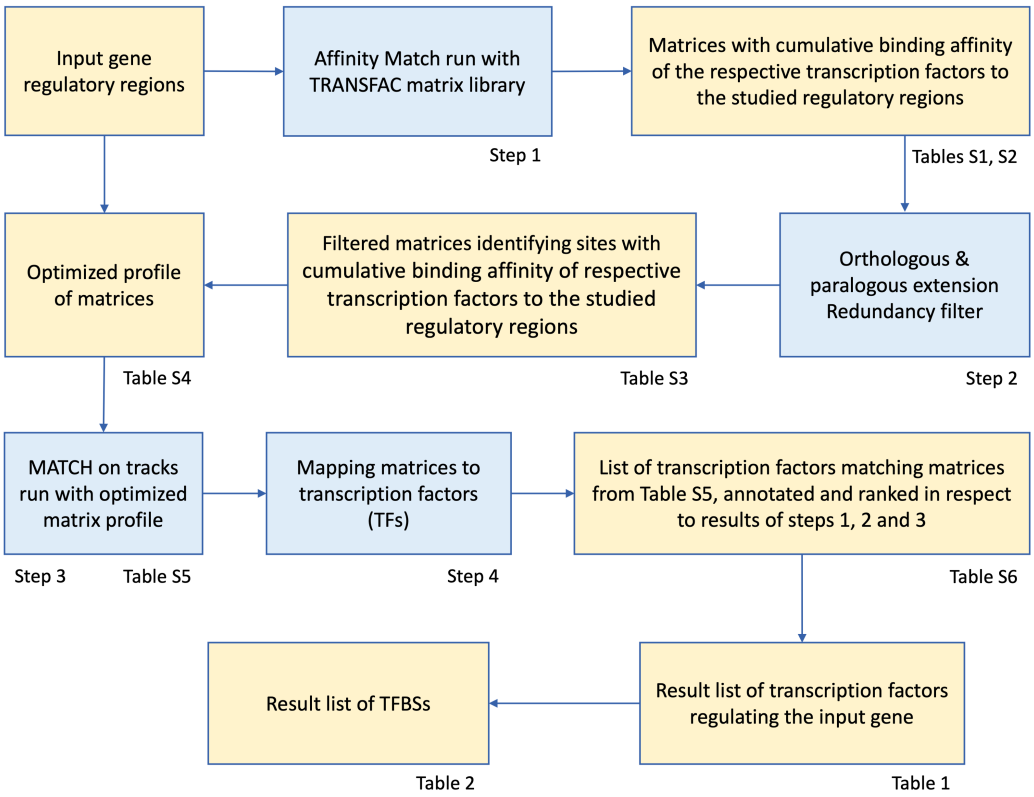
≤0.014 ≤0.019 ≤0.025 ≤0.031 ≤0.045 ≤0.07 ≤0.13 ≤0.23 ≤0.4 ≤1.0

The rank below represents the rank of factor expression in testis tissue out of the number of supported tissues for each factor and factor's average expression value in addition to them. The maximum number of currently supported tissues for one factor is 61.

Results details

The analysis workflow of the MATCH Suite comprises 4 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 single gene analysis workflow



Step 1

The promoter of the input gene was extracted using the FANTOM5 database promoters for the testis tissue. The promoter regions used were taken as −500 bp upstream of transcription start site (TSS) and +100 bp downstream the TSS. There were no enhancers or silencers found for the gene BRCA1 (ENSG00000012048) to be active in the testis tissue and thus the track of regulatory regions containing only the extracted promoter was then used for running the Affinity Match analysis [2] (Affinity Match™ for tracks).
Affinity Match analysis was ran using the matrix profile consisting of 5643 TRANSFAC® vertebrate matrices [3,4].

Supplementary table 1 (Table S1) comprises the results of this analysis. It represents the Affinity scores and respective p-values calculated for each of the matrices from the used profile for each of the examined regulatory regions (promoter). By default the sorting of the table is done according to the affinity score values. See Methods for a detailed explanation of the contents of this table.

For each matrix from Table S1 the best affinity p-value among all examined regulatory regions is chosen. The respective affinity score value is taken accordingly. Supplementary table 2 (Table S2) includes these best affinity p-values and respective affinity score values for each of the matrices.

Step 2

The list of matrices from Table S2 is then extended to matrices associated with TFs that are orthologs and paralogs to the ones already associated with the listed matrices. For this purpose, factor clusters have been defined based on geneXplain's expert knowledge; the corresponding table can be found here. See Methods for a detailed explanation of the orthologous and paralogous extension applied.

Finally, redundancy elimination is done by selecting just one matrix for each factor cluster - the one that minimizes the affinity p-value. The resulting list of matrices after the orthologous and paralogous extension and redundancy filtering is shown in Supplementary table 3 (Table S3). It comprises 139 matrices. A detailed explanation of the values shown in Table S3 is given in the Methods section.

The filtered list of matrices (Table S3) is used to construct a new matrix profile that will be used for the next analysis. The cut-offs for the profile are taken from the MATCH Suite standard profile. The constructed profile is shown in Supplementary table 4 (Table S4).

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 regulatory regions of the studied gene. The method uses the MATCH on tracks algorithm (see [5] and Methods for

further info) and generates the track of found sites. As a result, Supplementary table 5 (Table S5) is generated that in addition to previously calculated affinity p-value and affinity score values contains the columns: 'Site score' and the 'Number of sites'. A detailed explanation of the values shown in Table S5 is given in the Methods section.

Step 4

The transcription factors associated with the matrices from Table S5 are given in Supplementary table 6 (Table S6). Only factors that are known to be expressed in the testis tissue were considered. The sorting of the factors is based on the sum of three different ranks: ranking by the best -log(affinity p-value) of the factor's matrices, ranking by factor expression in the testis tissue, and ranking by its expression in the testis tissue compared to all other supported tissues. Table 1 contains the top 40 factors from Table S6, these factors are predicted to be regulating the BRCA1 (ENSG00000012048) gene in the selected conditions.

The matrices corresponding to the factors from Table 1 are listed in Table 2. These are the most relevant matrices (site models) that determine the regulation of the analyzed gene.

Table 2. Resulting matrices (site models) table

ID	Matrix logo	Number of sites	Affinity score	-log(affinity p-value)	Site score
V$RFX3_11		6	12.95	5.45	0.94
V$CREM_01		2	8.79	3.61	1
V$ZKSCAN1_05		2	10.38	3.19	0.82
V$ARID5A_04		1	6.61	3.16	0.98
V$NFATC4_02		2	9.98	3.05	0.91
V$ZNF740_06		4	10.91	2.95	0.91
V$TIEG1_Q6		2	5.52	2.48	1
V$CNOT3_01		3	5.98	2.45	0.96
V$E2F3_06		3	7.38	2.35	0.93
V$ZFP583_01		3	0.43	2.34	0.85
V$ZNF385D_02		8	4.18	2.29	0.82
V$GMEB2_04		3	5.39	2.15	0.97
V$ELF1_Q6		2	9.36	2.1	0.98
V$GABPA_Q4		1	5.33	2.07	1
V$STAT5A_03		2	4.01	2.01	1
V$ZBTB33_08		3	4.28	1.86	0.87
V$SOX6_02		1	6.65	1.74	0.95
V$CTCF_08		2	7.56	1.65	0.92
V$ZNF692_02		1	6.16	1.65	0.87
V$CGBP_02		4	5.03	1.61	0.96

View full table →

The list of all regulatory regions of BRCA1 (ENSG00000012048) gene together with the number of hits identified within each of these regions with the matrices from Table 2 are presented in Table 3 Underneath each matrix name, the TFs referring to it are given in the order of expression significance in the testis tissue.

Table 3. Regulatory regions table with the identified transcription factors and their site models regulating the input gene

Type	Accession	Coordinates	Total number of sites	V$RFX3_11 RFX4 RFX2 RFX3 more	V$CREM_01 CREM CREB	V$ZKSCAN1_05 ZSCAN9 ZNF323	V$ARID5A_04 JARID1B ARID2 ARID3B more	V$NFATC4_02 NFATc3	V$ZNF740_06 ZNF511
promoter	BRCA1	17:43125265-43125864	62	6	2	2	1	2	4

View full table →

References

[1] 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.

[2] Lloyd, K., Papoutsopoulou, S., Smith, E., Stegmaier, P., Bergey, F., Morris, L., Kittner, M., England, H., Spiller, D., White, M.H. and Duckworth, C.A. (2020) Using systems medicine to identify a therapeutic agent with potential for repurposing in inflammatory bowel disease. Disease models & mechanisms. 13(11), p.dmm044040. 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.

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 2.0 (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