

PTGS2 and PLK3 are promising druggable targets for treating Alzheimer Disease that control activity of CUX1, IRF3 and FOXO1 transcription factors on promoters of differentially expressed genes

Alexander Kel
geneXplain GmbH
alexander.kel2@gmail.com

Data received on 02/11/2020 ; Run on 25/02/2026 ; Report generated on 25/02/2026

Genome Enhancer release 3.7 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2025.2)



Abstract

In the present study we applied the software package "Genome Enhancer" to a data set that contains *transcriptomics* data. The study is done in the context of *Alzheimer Disease*. The goal of this pipeline is to identify potential drug targets in the molecular network that governs the studied pathological process. In the first step of analysis pipeline discovers transcription factors (TFs) that regulate genes activities in the pathological state. The activities of these TFs are controlled by so-called master regulators, which are identified in the second step of analysis. After a subsequent druggability checkup, the most promising master regulators are chosen as potential drug targets for the analyzed pathology. At the end the pipeline comes up with (a) a list of known drugs and (b) investigational active chemical compounds with the potential to interact with selected drug targets.

From the data set analyzed in this study, we found the following TFs to be potentially involved in the regulation of the differentially expressed genes: CUX1, IRF3, SOX4, FOXO1, PPARG and CEBPB. The subsequent network analysis suggested

- COX2
- plk3

as the most promising molecular targets for further research, drug development and drug repurposing initiatives on the basis of identified molecular mechanism of the studied pathology.

1. Introduction

Recording "-omics" data to measure gene activities, protein expression or metabolic events is becoming a standard approach to characterize the pathological state of an affected organism or tissue. Increasingly, several of these methods are applied in a combined approach leading to large "multiomics" datasets. Still the challenge remains how to reveal the

underlying molecular mechanisms that render a given pathological state different from the norm. The disease-causing mechanism can be described by a re-wiring of the cellular regulatory network, for instance as a result of a genetic or epigenetic alterations influencing the activity of relevant genes. Reconstruction of the disease-specific regulatory networks can help identify potential master regulators of the respective pathological process. Knowledge about these master regulators can point to ways how to block a pathological regulatory cascade. Suppression of certain molecular targets as components of these cascades may stop the pathological process and cure the disease.

Conventional approaches of statistical "-omics" data analysis provide only very limited information about the causes of the observed phenomena and therefore contribute little to the understanding of the pathological molecular mechanism. In contrast, the "upstream analysis" method [1-4] applied here has been devised to provide a casual interpretation of the data obtained for a pathology state. This approach comprises two major steps: (1) analysing promoters and enhancers of differentially expressed genes for the transcription factors (TFs) involved in their regulation and, thus, important for the process under study; (2) re-constructing the signaling pathways that activate these TFs and identifying master regulators at the top of such pathways. For the first step, the database TRANSFAC® [5] is employed together with the TF binding site identification algorithms Match [6] and CMA [7]. The second step involves the signal transduction database TRANSPATH® [8] and special graph search algorithms [10-11] implemented in the software "Genome Enhancer".

The "upstream analysis" approach has now been extended by a third step that reveals known drugs suitable to inhibit (or activate) the identified molecular targets in the context of the disease under study. This step is performed by using information from HumanPSD™ database [11]. In addition, some known drugs and investigational active chemical compounds are subsequently predicted as potential ligands for the revealed molecular targets. They are predicted using a pre-computed database of spectra of biological activities of chemical compounds of a library of 2245 known drugs and investigational chemical compounds from HumanPSD™ database. The spectra of biological activities for these compounds are computed using the program PASS on the basis of a (Q)SAR approach [12-14]. These predictions can be used for the research purposes - for further drug development and drug repurposing initiatives.

2. Data

For this study the following experimental data was used:

Table 1. Experimental datasets used in the study

File name	Data type
GSE144194_MC65_exons_raw_count	Transcriptomics

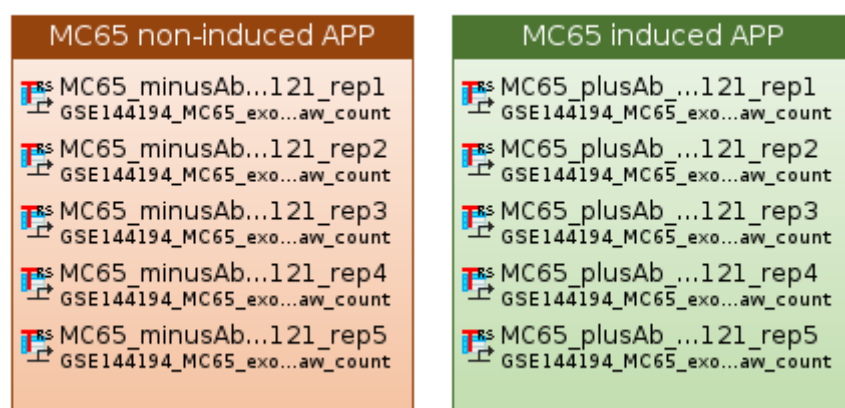


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

3. Results

We have compared the following conditions: MC65 induced APP versus MC65 non-induced APP.

3.1. Identification of target genes

In the first step of the analysis **target genes** were identified from the uploaded experimental data. We applied the Limma tool (R/Bioconductor package integrated into our pipeline) and compared gene expression in the following sets: "MC65 induced APP" with "MC65 non-induced APP". Limma calculated the LogFC (the logarithm to the base 2 of the fold change between different conditions), the p-value and the adjusted p-value (corrected for multiple testing) of the observed fold change. As a result, we detected 1424 upregulated genes (LogFC>0.5) out of which 108 genes were found as significantly upregulated (p-value<0.01) and 1641 downregulated genes (LogFC<-0.5) out of which 31 genes were significantly downregulated (p-value<0.01). See tables below for the top significantly up- and downregulated genes. Below we call **target genes** the full list of up- and downregulated genes revealed in our analysis (see tables in [Supplementary section](#)).

Table 2. Top ten significant **up-regulated** genes in MC65 induced APP vs. MC65 non-induced APP.

[See full table](#) →

ID	Gene symbol	Gene description	logFC	P.Value	adj.P.Val
ENSG00000169429	CXCL8	C-X-C motif chemokine ligand 8	3.06	3.75E-8	6.54E-6
ENSG00000100292	HMOX1	heme oxygenase 1	2.95	2.11E-9	6.08E-7
ENSG00000073756	PTGS2	prostaglandin-endoperoxide synthase 2	2.83	2.94E-14	1.4E-10
ENSG00000120738	EGR1	early growth response 1	2.7	8.28E-13	1.96E-9
ENSG00000139318	DUSP6	dual specificity phosphatase 6	2.48	1.64E-12	2.59E-9
ENSG00000244405	ETV5	ETS variant transcription factor 5	2.42	3.75E-15	3.56E-11
ENSG00000175832	ETV4	ETS variant transcription factor 4	1.9	2.27E-7	2.51E-5
ENSG00000170345	FOS	Fos proto-oncogene, AP-1 transcription factor subunit	1.87	1.18E-11	1.12E-8
ENSG00000106366	SERPINE1	serpin family E member 1	1.82	5.68E-10	2.12E-7
ENSG00000177606	JUN	Jun proto-oncogene, AP-1 transcription factor subunit	1.76	3.09E-12	4.18E-9

Table 3. Top ten significant **down-regulated** genes in MC65 induced APP vs. MC65 non-induced APP.

[See full table](#) →

ID	Gene symbol	Gene description	logFC	P.Value	adj.P.Val
ENSG00000117318	ID3	inhibitor of DNA binding 3	-1.2	6.2E-10	2.12E-7
ENSG00000133110	POSTN	periostin	-0.98	7.27E-7	6.05E-5
ENSG00000129757	CDKN1C	cyclin dependent kinase inhibitor 1C	-0.83	6.84E-6	3.51E-4
ENSG00000149212	SESN3	sestrin 3	-0.72	1.39E-7	1.67E-5
ENSG00000100427	MLC1	modulator of VRAC current 1	-0.72	1.58E-7	1.85E-5
ENSG00000007237	GAS7	growth arrest specific 7	-0.71	1.72E-3	1.61E-2
ENSG00000108984	MAP2K6	mitogen-activated protein kinase kinase 6	-0.71	6.16E-7	5.31E-5
ENSG00000115461	IGFBP5	insulin like growth factor binding protein 5	-0.71	2.74E-8	5.19E-6
ENSG00000185432	TMT1A	thiol methyltransferase 1A	-0.66	6.58E-6	3.45E-4
ENSG00000104081	BMF	Bcl2 modifying factor	-0.65	1.63E-8	3.22E-6

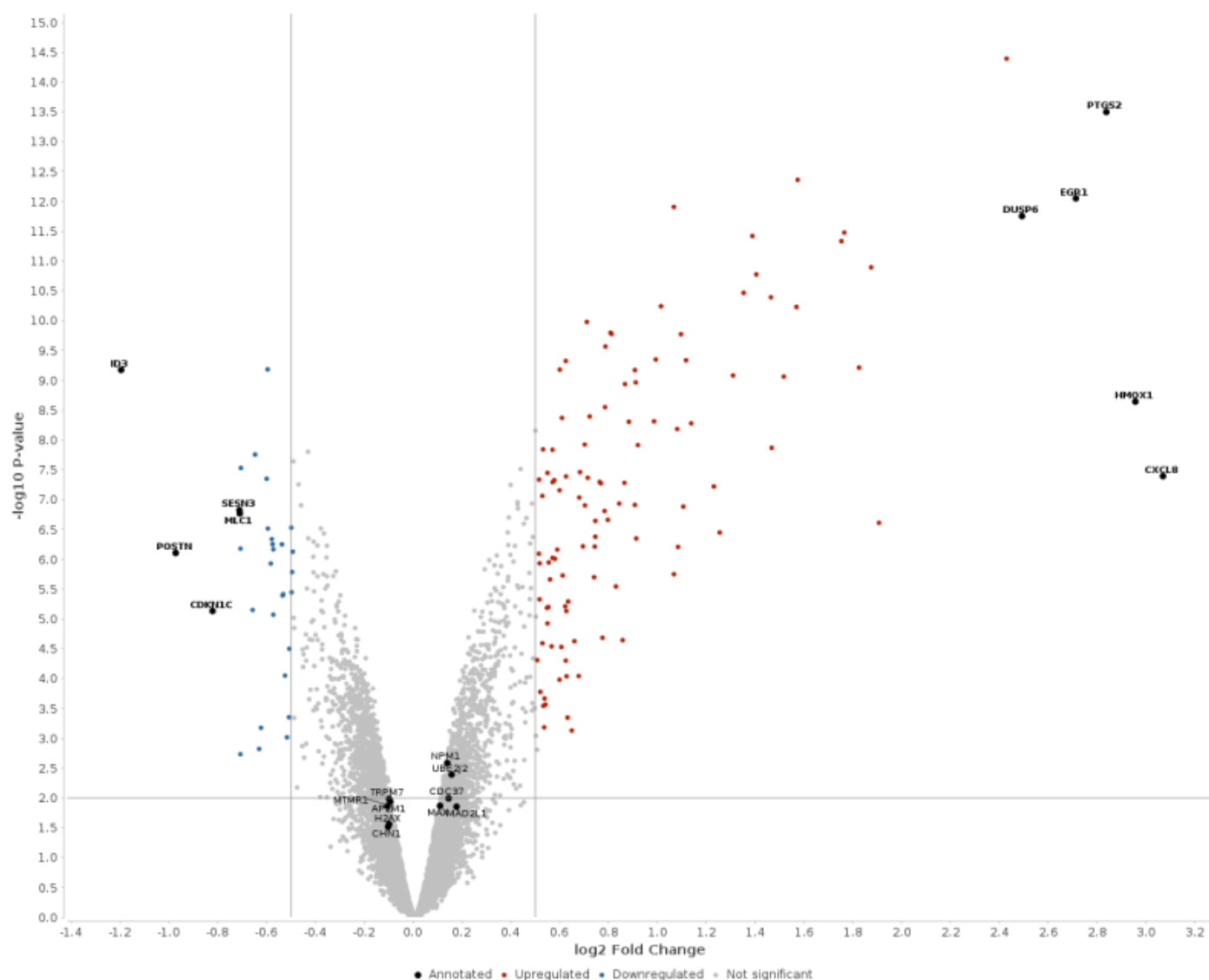


Figure 3. Volcano plot. Each dot represents one gene. Gray dots represent genes with no significant expression between conditions analyzed, the blue dots represent downregulated genes, and the red dots represent upregulated genes. Annotated are the top DEGs (bold) and the genes that encode top 10 up- and down-regulated master-regulators identified in the current study.

3.2. Functional classification of genes

A functional analysis of differentially expressed genes was done by mapping the significant up-regulated and significant down-regulated genes to several known ontologies, such as Gene Ontology (GO), disease ontology (based on HumanPSD™ database) and the ontology of signal transduction and metabolic pathways from the [TRANSPATH®](#) database. Statistical significance was computed using a binomial test.

Figures 4-9 show the most significant categories.

Heatmap of differentially expressed genes in MC65 induced APP vs. MC65 non-induced APP

A heatmap of all differentially expressed genes playing a potential regulatory role in the system (enriched in [TRANSPATH®](#) pathways) is presented in Figure 2.

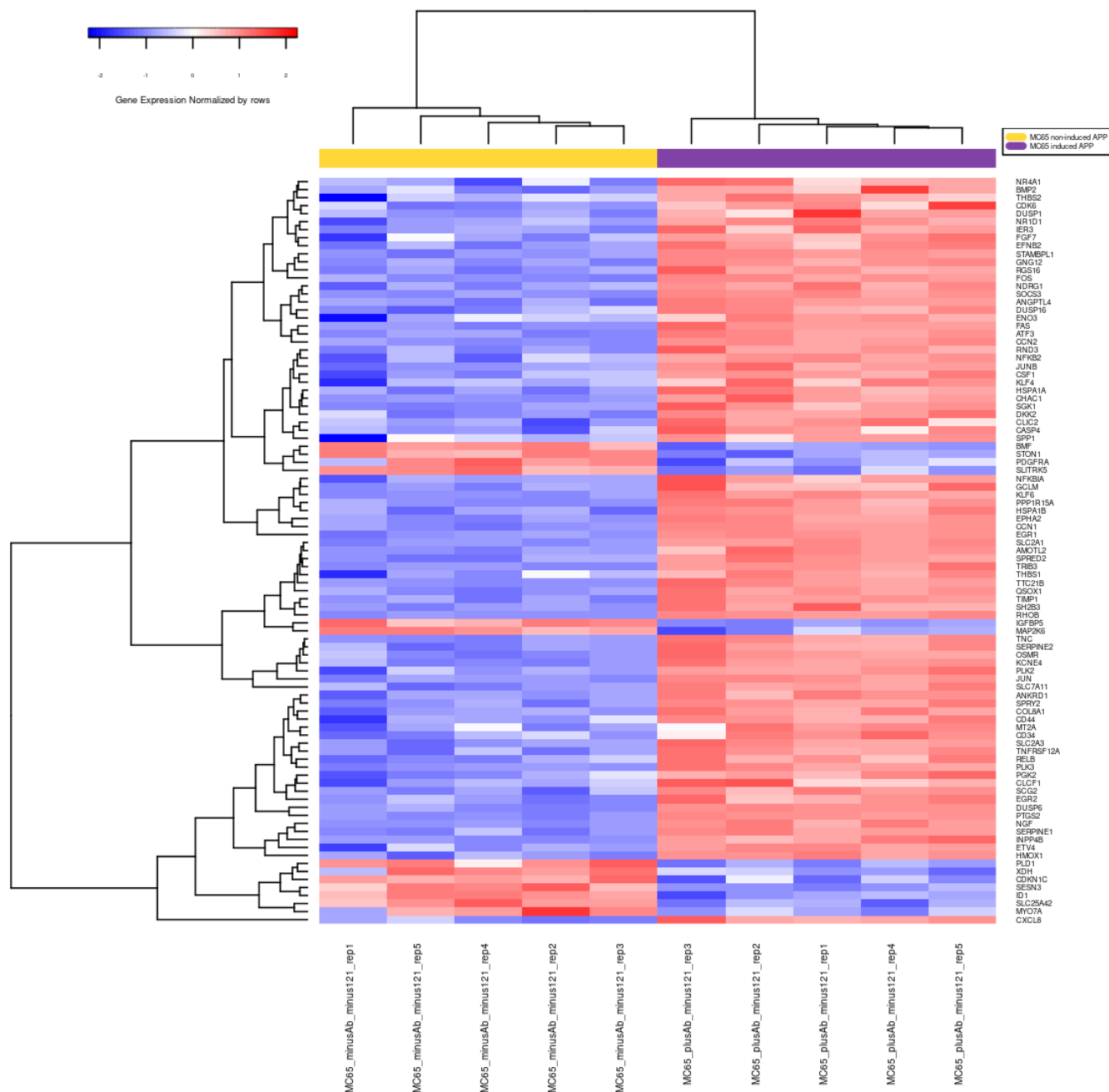


Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.

[See full diagram →](#)

Up-regulated genes in MC65 induced APP vs. MC65 non-induced APP:

108 significant up-regulated genes were taken for the mapping.

GO (biological process)



Figure 4. Enriched GO (biological process) of up-regulated genes in MC65 induced APP vs. MC65 non-induced APP.
Full classification →

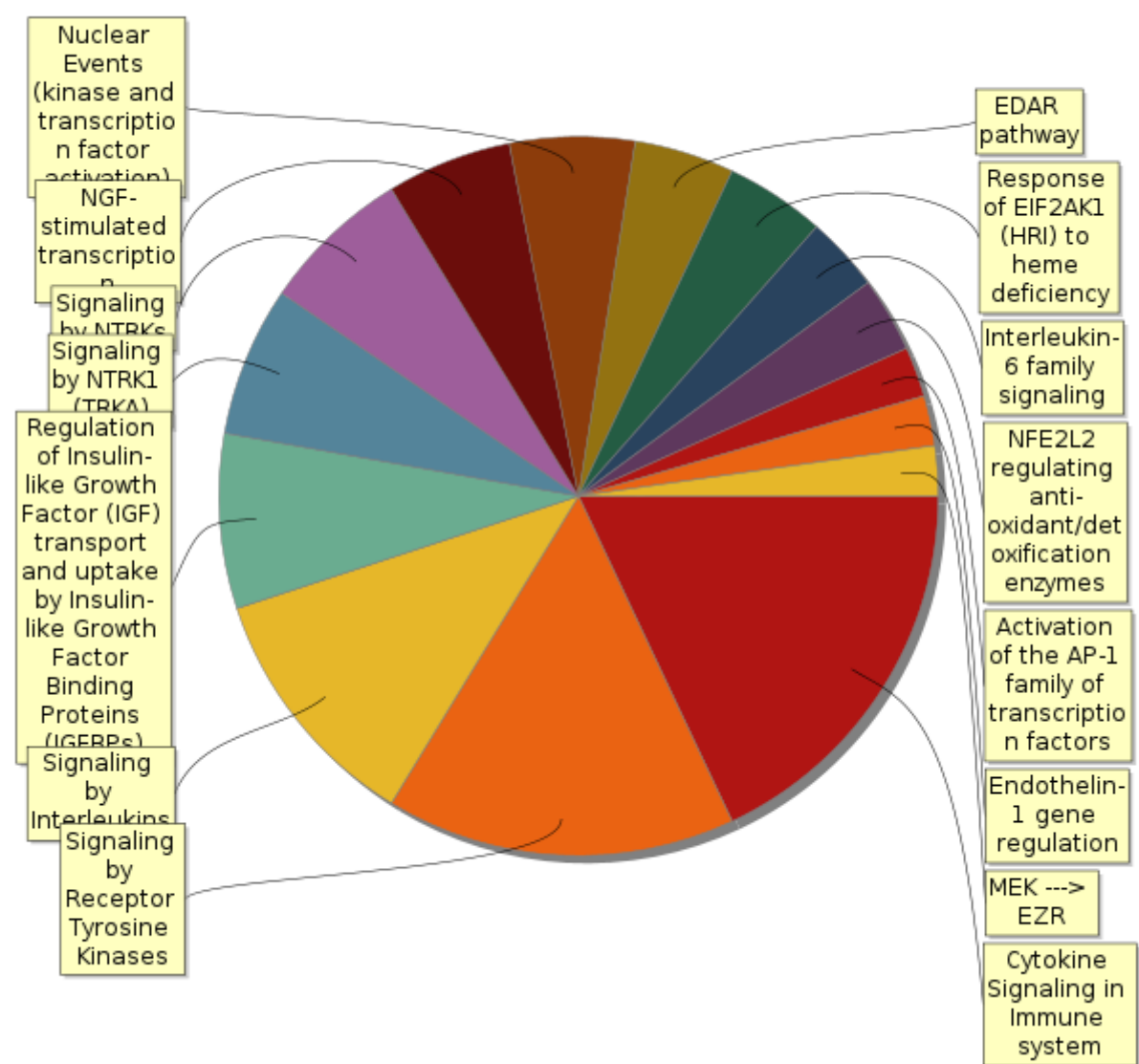


Figure 5. Enriched TRANSPATH® Pathways (2025.2) of up-regulated genes in MC65 induced APP vs. MC65 non-induced APP.
[Full classification →](#)

HumanPSD(TM) disease (2025.2)

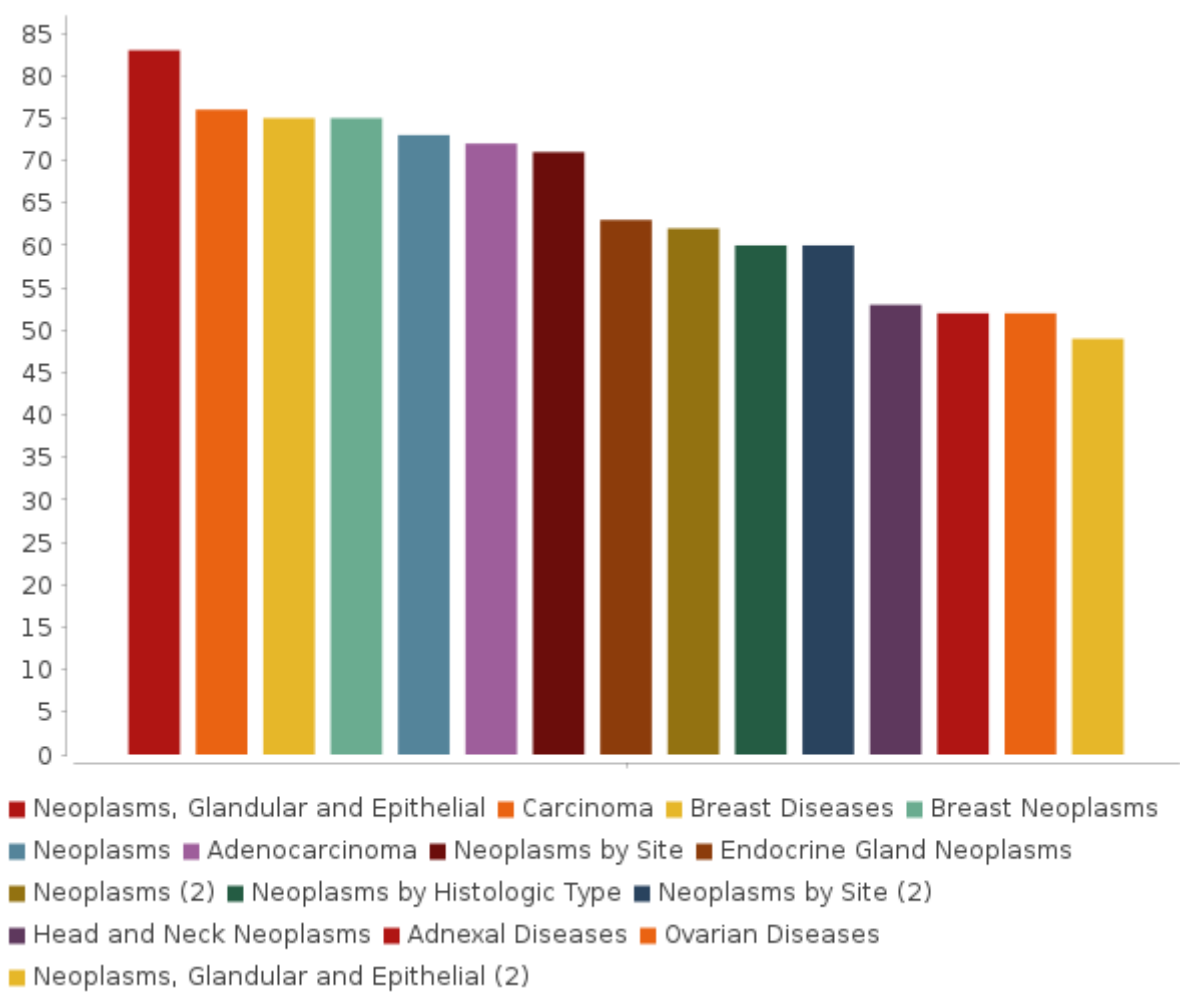


Figure 6. Enriched HumanPSD(TM) disease (2025.2) of up-regulated genes in MC65 induced APP vs. MC65 non-induced APP. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

[Full classification](#) →

Down-regulated genes in MC65 induced APP vs. MC65 non-induced APP:

31 significant down-regulated genes were taken for the mapping.

GO (biological process)



Figure 7. Enriched GO (biological process) of down-regulated genes in MC65 induced APP vs. MC65 non-induced APP.
[Full classification](#) →

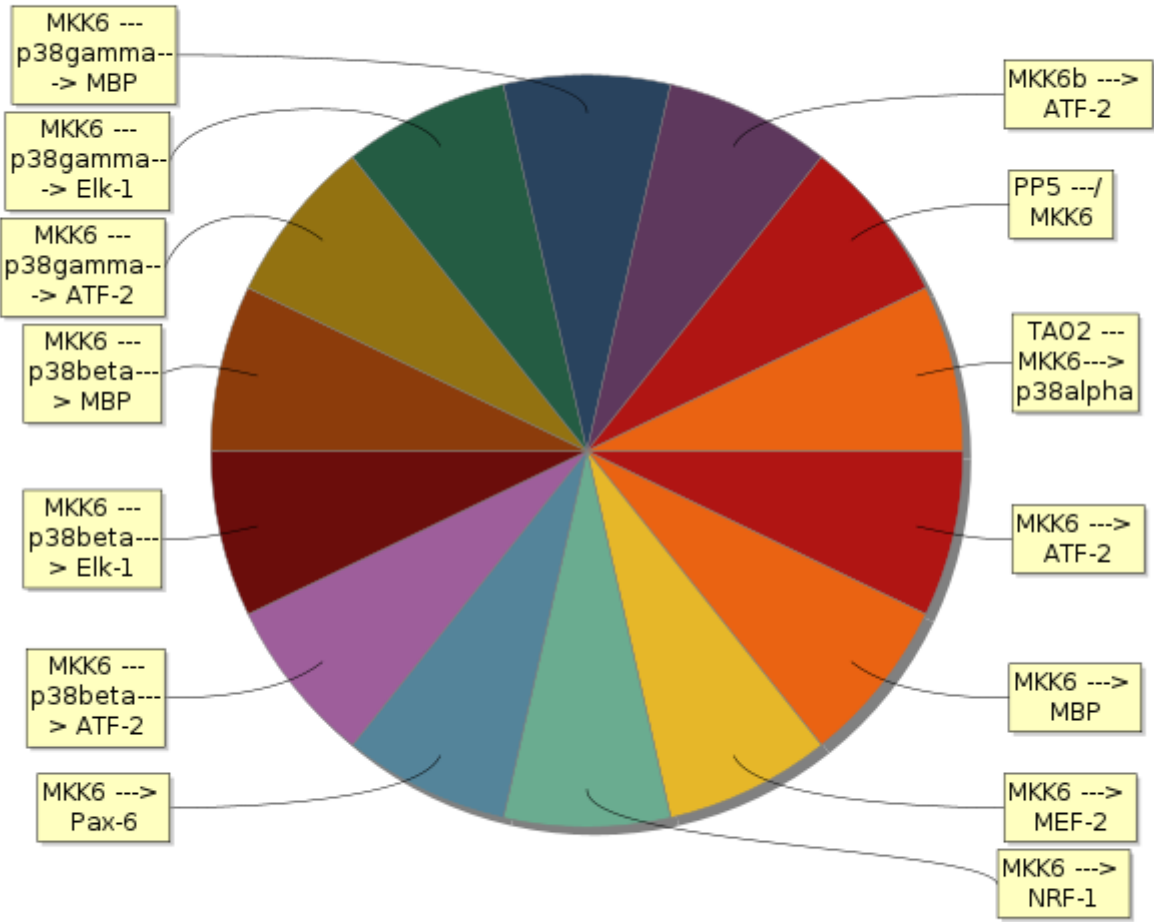


Figure 8. Enriched TRANSPATH® Pathways (2025.2) of down-regulated genes in MC65 induced APP vs. MC65 non-induced APP.
[Full classification →](#)

HumanPSD(TM) disease (2025.2)

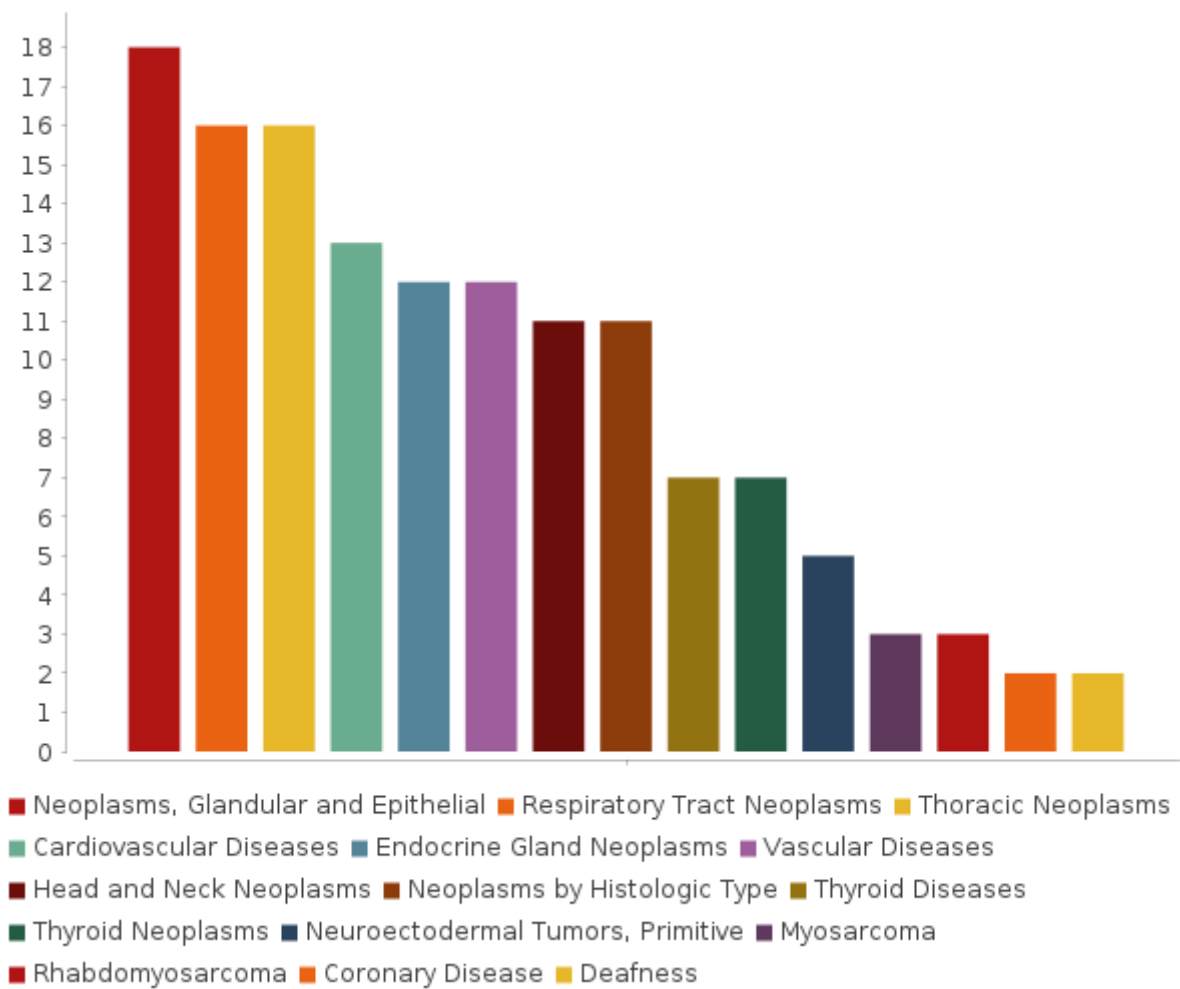
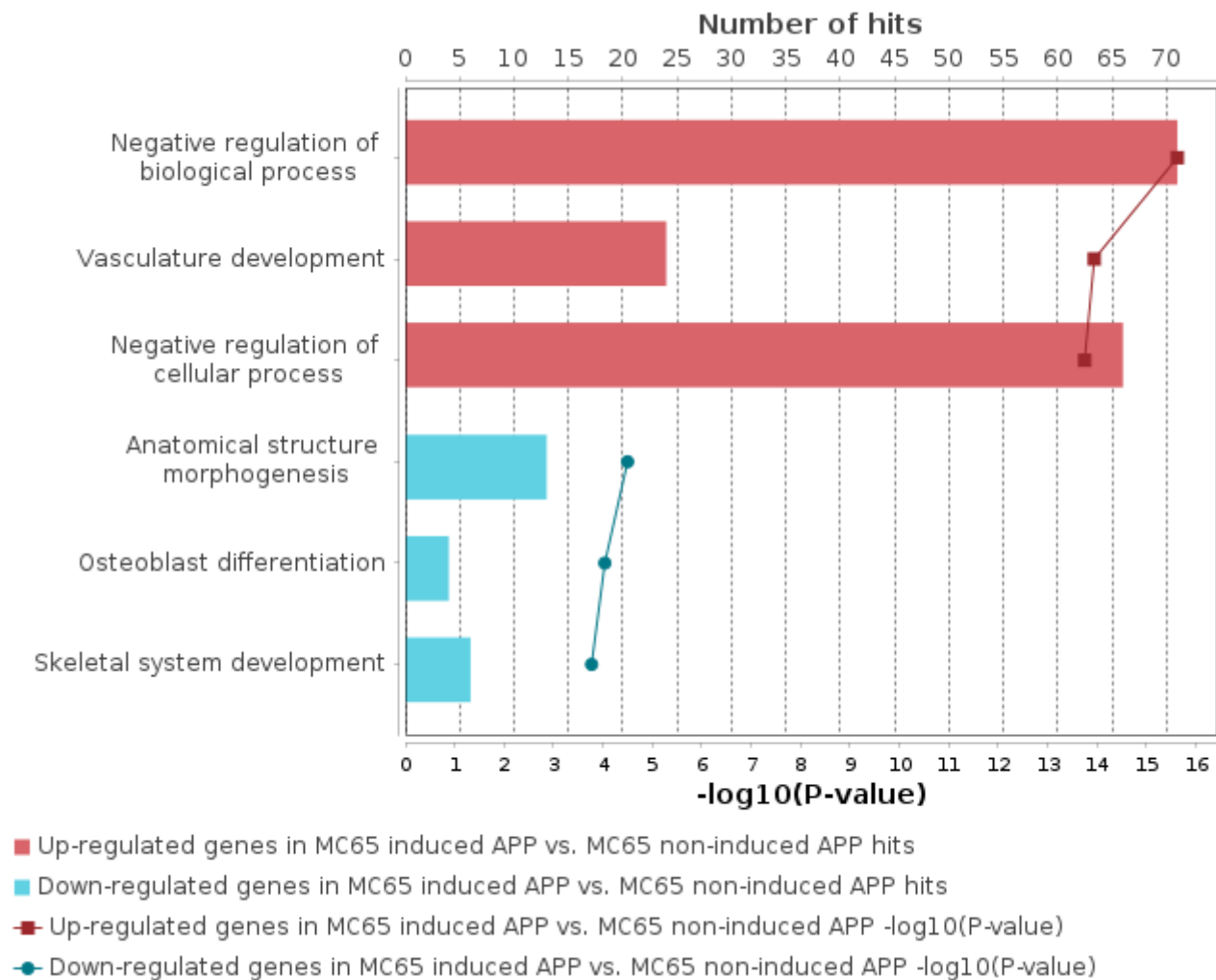


Figure 9. Enriched HumanPSD(TM) disease (2025.2) of down-regulated genes in MC65 induced APP vs. MC65 non-induced APP. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

[Full classification →](#)

The result of overall Gene Ontology (GO) analysis of the differentially expressed genes of the studied pathology can be summarized by the following diagram, revealing the most significant functional categories overrepresented among the observed (differentially expressed genes):



3.3. Analysis of enriched transcription factor binding sites and composite modules

In the next step a search for transcription factors binding sites (TFBS) was performed in the regulatory regions of the **target genes** by using the TF binding motif library of the [TRANSFAC®](#) database. We searched for so called **composite modules** that act as potential condition-specific **enhancers** of the **target genes** in their upstream regulatory regions (-1000 bp upstream of transcription start site (TSS)) and identify transcription factors regulating activity of the genes through such **enhancers**.

Classically, **enhancers** are defined as regions in the genome that increase transcription of one or several genes when inserted in either orientation at various distances upstream or downstream of the gene [7]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [8].

We applied the Composite Module Analyst (CMA) [7] method to detect such potential enhancers, as targets of multiple TFs bound in a cooperative manner to the regulatory regions of the genes of interest. CMA applies a genetic algorithm to construct a generalized model of the enhancers by specifying combinations of TF motifs (from [TRANSFAC®](#)) whose sites are most frequently clustered together in the regulatory regions of the studied genes. CMA identifies the transcription factors that through their cooperation provide a synergistic effect and thus have a great influence on the gene regulation process.

Enhancer model potentially involved in regulation of target genes (up-regulated genes in MC65 induced APP vs. MC65 non-induced APP).

To build the most specific composite modules we choose all significant up-regulated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all up-regulated genes in MC65 induced APP vs. MC65 non-induced APP.

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.

Module 1:

V\$MZF1_02 0.86; N=3	V\$TEAD4_01 0.92; N=2	V\$CDPCR3HD_01 0.94; N=3	V\$MEIS1BHOXA9_02 0.74; N=2	V\$SOX4_05 0.87; N=1
-------------------------	--------------------------	-----------------------------	--------------------------------	-------------------------

Module width: 153

Module 2:

V\$FOXP2_03 0.92; N=2	V\$OSX_Q3 0.99; N=2	V\$IRF3_11 0.80; N=2	V\$PEA3_03 0.97; N=3	V\$TEF3_01 0.97; N=1
--------------------------	------------------------	-------------------------	-------------------------	-------------------------

Module width: 100

Model score ($-p \cdot \log_{10}(pval)$): 13.84

Wilcoxon p-value (pval): 2.46e-28

Penalty (p): 0.501

Average yes-set score: 5.35

Average no-set score: 3.48

AUC: 0.84

Separation point: 3.96

False-positive: 34.40%

False-negative: 8.33%

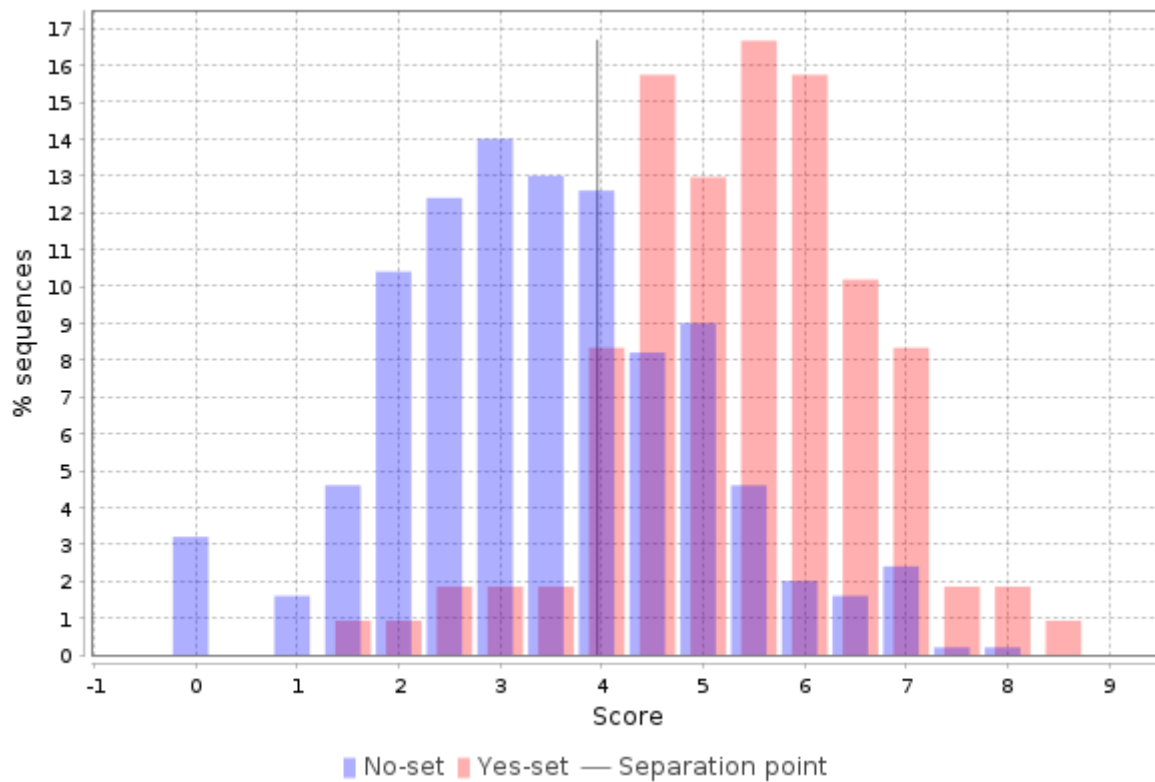


Table 4. List of top ten up-regulated genes in MC65 induced APP vs. MC65 non-induced APP with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.

[See full table](#) →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000169908	TM4SF1	transmembrane 4 L six family member 1	8.33	TEF-3(h), CUX-1(h), Hox-A9(h), Meis1(h), SOX-4(h), MZF1(h), Sp7(h), ETV4(h)
ENSG00000119820	YIPF4	Yip1 domain family member 4	8.19	CUX-1(h), MZF1(h), SOX-4(h), FOXP2(h), TEF-3(h), Hox-A9(h), Meis1(h), ETV4(h)
ENSG00000109320	NFKB1	nuclear factor kappa B subunit 1	8.18	MZF1(h), TEF-3(h), CUX-1(h), Hox-A9(h), Meis1(h), ETV4(h), Sp7(h)
ENSG00000198492	YTHDF2	YTH N6-methyladenosine RNA binding protein F2	8.15	SOX-4(h), TEF-3(h), MZF1(h), Hox-A9(h), Meis1(h), IRF-3(h), ETV4(h), Sp7(h)
ENSG00000182752	PAPPA	pappalysin 1	8.1	MZF1(h), SOX-4(h), TEF-3(h), Hox-A9(h), Meis1(h), Sp7(h), CUX-1(h), FOXP2(h)
ENSG00000197905	TEAD4	TEA domain transcription factor 4	8.09	SOX-4(h), Hox-A9(h), Meis1(h), ETV4(h), Sp7(h), MZF1(h), TEF-3(h)
ENSG00000107672	NSMCE4A	NSE4 homolog A, SMC5-SMC6 complex component	7.96	ETV4(h), Sp7(h), FOXP2(h), Hox-A9(h), Meis1(h), TEF-3(h), SOX-4(h), MZF1(h)...
ENSG00000163660	CCNL1	cyclin L1	7.94	Hox-A9(h), Meis1(h), ETV4(h), TEF-3(h), SOX-4(h), CUX-1(h), MZF1(h)
ENSG00000006468	ETV1	ETS variant transcription factor 1	7.94	MZF1(h), Sp7(h), FOXP2(h), Hox-A9(h), Meis1(h), TEF-3(h), SOX-4(h), CUX-1(h)
ENSG00000182199	SHMT2	serine hydroxymethyltransferase 2	7.81	Hox-A9(h), Meis1(h), CUX-1(h), MZF1(h), TEF-3(h), SOX-4(h), ETV4(h), Sp7(h)

Enhancer model potentially involved in regulation of target genes (down-regulated genes in MC65 induced APP vs. MC65 non-induced APP).

To build the most specific composite modules we choose all significant down-regulated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all down-regulated genes in MC65 induced APP vs. MC65 non-induced APP.

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.

Module 1:

V\$ZBTB33_10 0.99; N=2	V\$IRF8_Q6 0.97; N=2	V\$CEBPB_03 0.85; N=2	V\$PPARG_09 0.81; N=2	V\$SOX10_07 0.91; N=2
---------------------------	-------------------------	--------------------------	--------------------------	--------------------------

Module width: 154

Module 2:

V\$PBX1_17 0.93; N=2	V\$CPEB1_01 0.91; N=2	V\$WT1_Q6 0.99; N=2	V\$FOXO1_09 0.97; N=3	V\$IRF7_01 0.87; N=1
-------------------------	--------------------------	------------------------	--------------------------	-------------------------

Module width: 145

Model score ($-\log_{10}(pval)$): 7.86

Wilcoxon p-value (pval): 2.05e-16

Penalty (p): 0.501

Average yes-set score: 6.95

Average no-set score: 4.15

AUC: 0.94

Separation point: 5.39

False-positive: 21.40%

False-negative: 3.23%

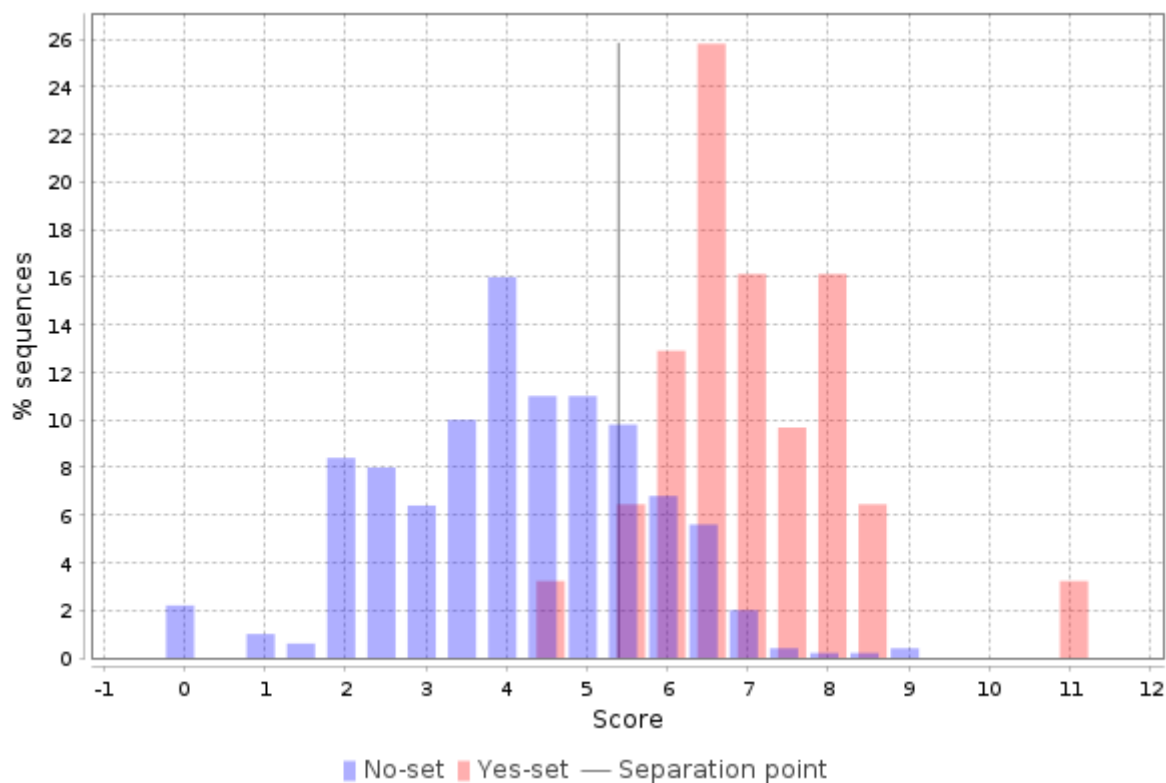


Table 5. List of top ten down-regulated genes in MC65 induced APP vs. MC65 non-induced APP with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.

[See full table](#) →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000117318	ID3	inhibitor of DNA binding 3	10.8	PBX-1(h), ZBTB33(h), IRF-8(h), SOX-10(h), FOXO1(h), C/EBPbeta(h), PPARGgamma(h)...
ENSG00000166073	GPR176	G protein-coupled receptor 176	10.04	PBX-1(h), IRF-7(h), CPE-BP1(h), ZBTB33(h), FOXO1(h), IRF-8(h), PPARGgamma(h)...
ENSG00000138468	SEN7	SUMO specific peptidase 7	9.55	IRF-8(h), ZBTB33(h), WT1(h), PPARGgamma(h), C/EBPbeta(h), PBX-1(h), SOX-10(h)...
ENSG00000257365	FNTB	farnesyltransferase, CAAX box, beta	9.21	IRF-7(h), PBX-1(h), WT1(h), IRF-8(h), SOX-10(h), C/EBPbeta(h), ZBTB33(h)...
ENSG00000113083	LOX	lysyl oxidase	9.06	ZBTB33(h), PBX-1(h), WT1(h), PPARGgamma(h), IRF-8(h), CPE-BP1(h), C/EBPbeta(h)...
ENSG00000114805	PLCH1	phospholipase C eta 1	9	IRF-7(h), WT1(h), PPARGgamma(h), CPE-BP1(h), IRF-8(h), C/EBPbeta(h), SOX-10(h)
ENSG00000057019	DCBLD2	discoidin, CUB and LCCL domain containing 2	8.92	PPARGgamma(h), ZBTB33(h), CPE-BP1(h), FOXO1(h), WT1(h), PBX-1(h)
ENSG00000142449	FBN3	fibrillin 3	8.81	PBX-1(h), WT1(h), IRF-8(h), SOX-10(h), C/EBPbeta(h)
ENSG00000092847	AGO1	argonaute RISC component 1	8.7	FOXO1(h), SOX-10(h), C/EBPbeta(h), PPARGgamma(h), CPE-BP1(h), PBX-1(h), IRF-8(h)
ENSG00000115687	PASK	PAS domain containing serine/threonine kinase	8.7	C/EBPbeta(h), WT1(h), IRF-8(h), ZBTB33(h), PBX-1(h), IRF-7(h), SOX-10(h)...

On the basis of the enhancer models we identified transcription factors potentially regulating the **target genes** of our interest. We found 10 and 10 transcription factors controlling expression of up- and down-regulated genes respectively (see Tables 6-7).

Table 6. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (up-regulated genes in MC65 induced APP vs. MC65 non-induced APP). **Yes-No ratio** is the ratio between frequencies of the sites in Yes sequences versus No sequences. It describes the level of the enrichment of binding sites for the indicated TF in the regulatory target regions. **Regulatory score** is the measure of involvement of the given TF in the controlling of expression of genes that encode master regulators presented below (through positive feedback loops).

[See full table](#) →

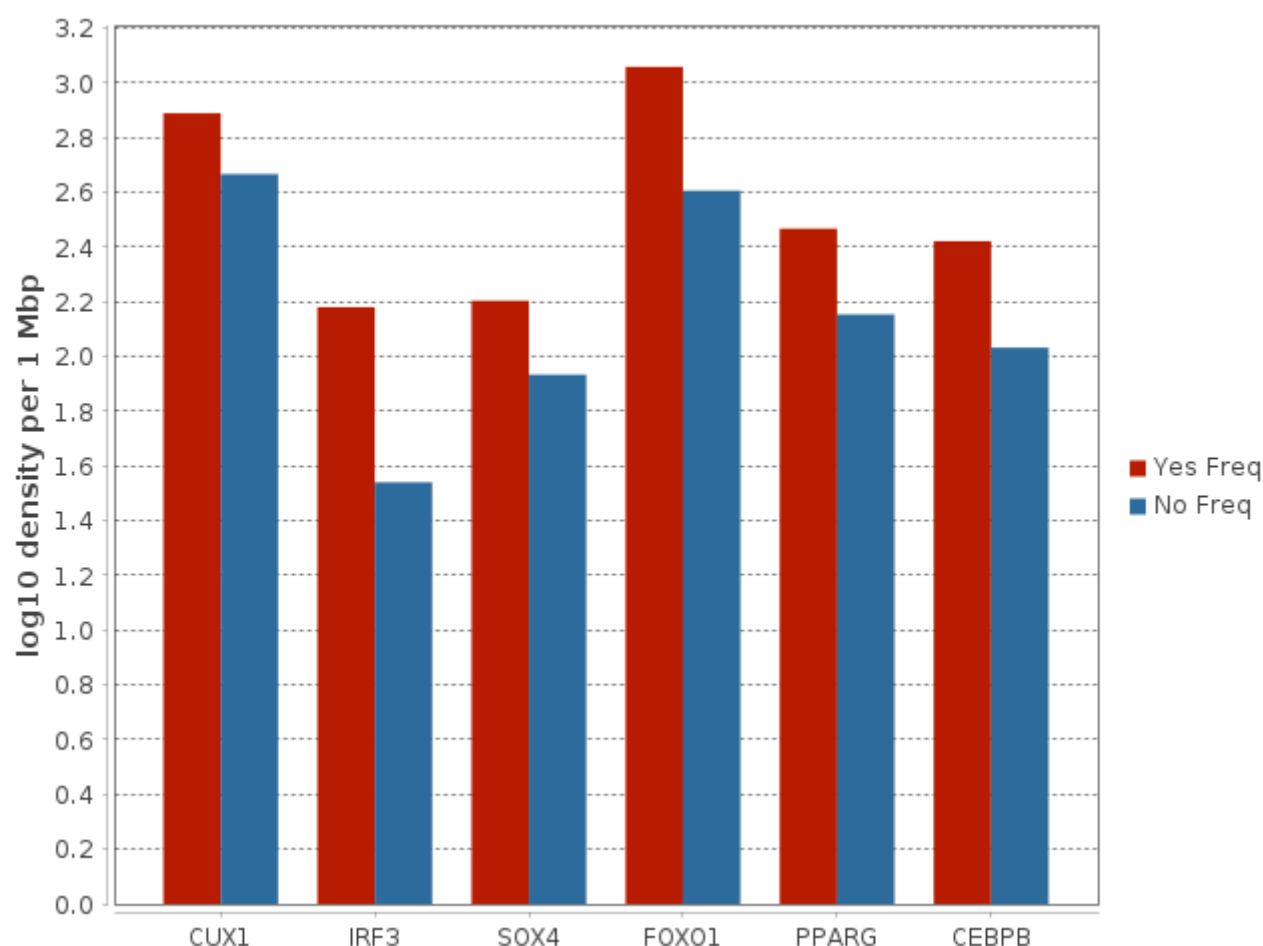
ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000024708	CUX1	cut like homeobox 1	2.61	1.67
MO000285816	IRF3	interferon regulatory factor 3	2.51	4.37
MO000025546	SOX4	SRY-box transcription factor 4	2.32	1.86
MO000028723	TEAD4	TEA domain transcription factor 4	2.31	13.82
MO000119037	HOXA9	homeobox A9	2.21	1.29
MO000046009	ETV4	ETS variant transcription factor 4	1.9	1.51
MO000028623	FOXP2	forkhead box P2	1.72	1.45
MO000176655	SP7	Sp7 transcription factor	1.37	1.31
MO000027355	MEIS1	Meis homeobox 1	0.42	6.91
MO000084468	MZF1	myeloid zinc finger 1	0.33	1.38

Table 7. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (down-regulated genes in MC65 induced APP vs. MC65 non-induced APP). **Yes-No ratio** is the ratio between frequencies of the sites in Yes sequences versus No sequences. It describes the level of the enrichment of binding sites for the indicated TF in the regulatory target regions. **Regulatory score** is the measure of involvement of the given TF in the controlling of expression of genes that encode master regulators presented below (through positive feedback loops).

[See full table](#) →

ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000034454	FOXO1	forkhead box O1	1.56	2.83
MO000033565	PPARG	peroxisome proliferator activated receptor gamma	1.49	2.06
MO000019381	CEBPB	CCAAT enhancer binding protein beta	1.4	2.45
MO000056592	ZBTB33	zinc finger and BTB domain containing 33	1.33	1.43
MO000007703	IRF7	interferon regulatory factor 7	1.07	1.71
MO000102040	WT1	WT1 transcription factor	1.04	5.35
MO000023424	IRF8	interferon regulatory factor 8	0.98	1.71
MO000082851	CPEB1	cytoplasmic polyadenylation element binding protein 1	0.95	1.53
MO000042273	PBX1	PBX homeobox 1	0.57	1.63
MO000028699	SOX10	SRY-box transcription factor 10	0	4.1

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: CUX1, IRF3, SOX4, FOXO1, PPARG and CEBPB.



3.4. Finding master regulators in networks

In the second step of the upstream analysis common regulators of the revealed TFs were identified. These master regulators appear to be the key candidates for therapeutic targets as they have a master effect on regulation of

intracellular pathways that activate the pathological process of our study. The identified master regulators are shown in Tables 8-9.

Table 8. Master regulators that may govern the regulation of **up-regulated** genes in MC65 induced APP vs. MC65 non-induced APP. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics data.

[See full table](#) →

ID	Master molecule name	Gene symbol	Gene description	logFC	Total rank
MO000023462	COX2(h)	PTGS2	prostaglandin-endoperoxide synthase 2	2.83	41
MO000096389	COX2(h)	PTGS2	prostaglandin-endoperoxide synthase 2	2.83	45
MO000031101	plk3(h)	PLK3	polo like kinase 3	1.46	56
MO000138699	plk3(h)	PLK3	polo like kinase 3	1.46	59
MO000019469	c-Jun(h)	JUN	Jun proto-oncogene, AP-1 transcription factor subunit	1.76	74
MO000059638	Sprouty2(h)	SPRY2	sprouty RTK signaling antagonist 2	1.11	81
MO000329204	Cdk6(h):cyclinD3-isoform1(h)	CCND3, CDK6	cyclin D3, cyclin dependent kinase 6	0.54	114
MO000033735	Sprouty2(h)	SPRY2	sprouty RTK signaling antagonist 2	1.11	115
MO000038172	plk2(h)	PLK2	polo like kinase 2	0.79	115
MO000154924	plk2(h)	PLK2	polo like kinase 2	0.79	117

Table 9. Master regulators that may govern the regulation of **down-regulated** genes in MC65 induced APP vs. MC65 non-induced APP. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics data.

[See full table](#) →

ID	Master molecule name	Gene symbol	Gene description	logFC	Total rank
MO000010977	PDGFRalpha(h)	PDGFRA	platelet derived growth factor receptor alpha	-0.52	49
MO000045378	txnip(h)	TXNIP	thioredoxin interacting protein	-0.37	70
MO000178122	txnip-isoform1(h)	TXNIP	thioredoxin interacting protein	-0.37	128
MO000256763	PDGFRalpha-isoform2(h)	PDGFRA	platelet derived growth factor receptor alpha	-0.52	163
MO000026702	PPARalpha(h)	PPARA	peroxisome proliferator activated receptor alpha	-0.28	169
MO000022211	MKK6(h)	MAP2K6	mitogen-activated protein kinase kinase 6	-0.71	181
MO000032712	MKP-4(h)	DUSP9	dual specificity phosphatase 9	-0.21	192
MO000035235	calcineurin(h)	PPP3CA, PPP3R1	protein phosphatase 3 catalytic subunit alpha, protein phosphatase 3 regulatory subunit B, alpha	-0.19	192
MO000033253	EVI-1(h)	MECOM	MDS1 and EVI1 complex locus	-0.43	193
MO000017049	beta-catenin(h)	CTNNB1	catenin beta 1	-0.19	194

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figures 10 and 11. These diagrams display the connections between identified transcription factors, which play important roles in the regulation of differentially expressed genes, and selected master regulators, which are responsible for the regulation of these TFs.

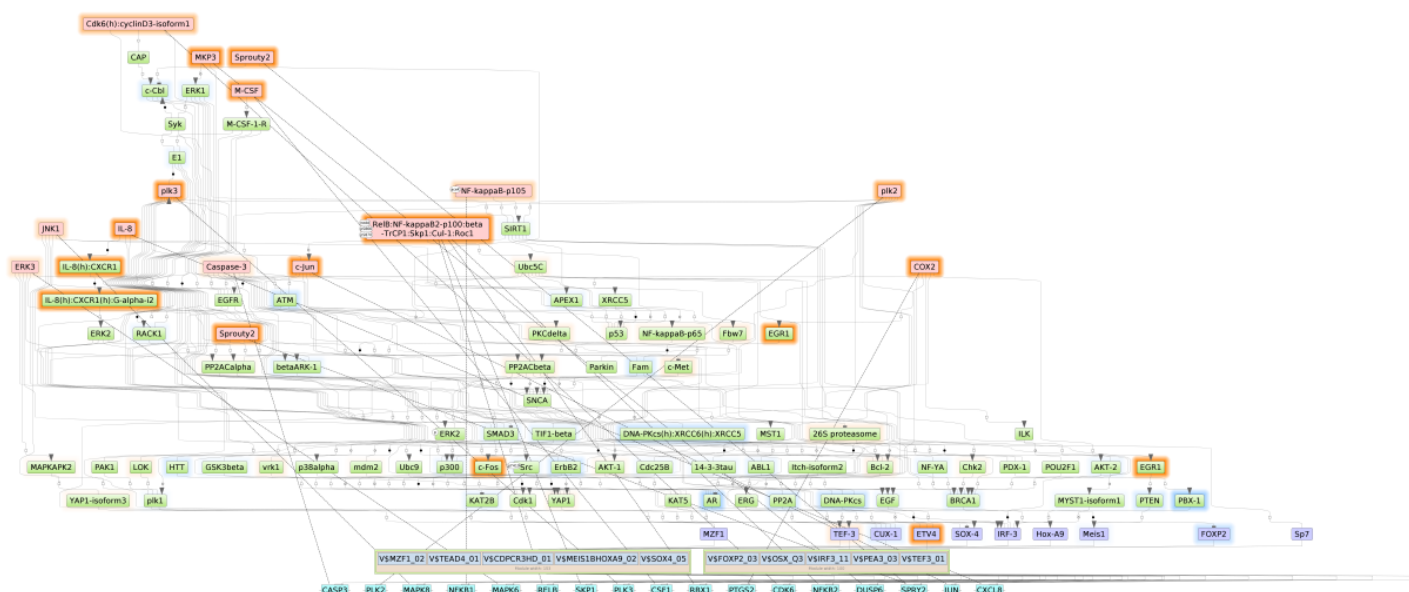


Figure 10. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in MC65 induced APP vs. MC65 non-induced APP. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during the search for master regulators from selected TFs. Orange and blue frames highlight molecules that are encoded by up- and downregulated genes, resp.

[See full diagram →](#)

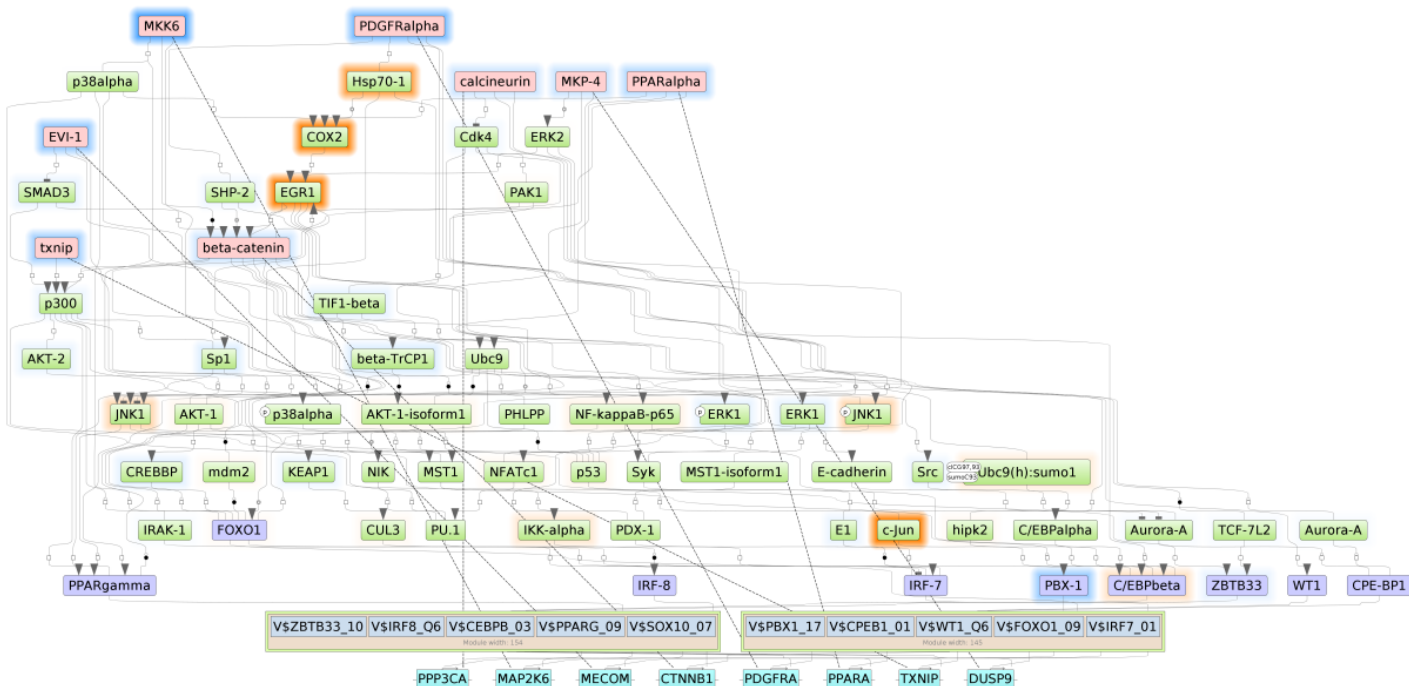


Figure 11. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in MC65 induced APP vs. MC65 non-induced APP. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during the search for master regulators from selected TFs. Orange and blue frames highlight molecules that are encoded by up- and downregulated genes, resp.

[See full diagram →](#)

4. Finding prospective drug targets

The identified master regulators that may govern pathology associated genes were checked for druggability potential using HumanPSD™ [11] database of gene-disease-drug assignments and PASS [12-14] software for prediction of biological activities of chemical compounds on the basis of a (Q)SAR approach. Respectively, for each master regulator protein we have computed two Druggability scores: HumanPSD Druggability score and PASS Druggability score. Where Druggability score represents the number of drugs that are potentially suitable for inhibition (or activation) of the corresponding target either according to the information extracted from medical literature (from HumanPSD™

database) or according to cheminformatics predictions of compounds activity against the examined target (from PASS software).

The cheminformatics druggability check is done using a pre-computed database of spectra of biological activities of chemical compounds from a library of all small molecular drugs from HumanPSD™ database, 2507 pharmaceutically active known chemical compounds in total. The spectra of biological activities has been computed using the program PASS [12-14] on the basis of a (Q)SAR approach.

If both Druggability scores were below defined thresholds (see Methods section for the details) such master regulator proteins were not used in further analysis of drug prediction.

As a result we created the following two tables of prospective drug targets (top targets are shown here):



Table 10. Prospective drug targets selected from full list of identified master regulators filtered by Druggability score from HumanPSD™ database. **Druggability score** contains the number of drugs that are potentially suitable for inhibition (or activation) of the target. The drug targets are sorted according to the **Total rank** which is the sum of three ranks computed on the basis of the three scores: keynode score, CMA score and expression change score (logFC, if present). See Methods section for details.

[See full table](#) →

Gene symbol	Gene Description	Druggability score	logFC	Total rank
PTGS2	prostaglandin-endoperoxide synthase 2	127	2.83	28
PLK3	polo like kinase 3	27	1.46	32
JUN	Jun proto-oncogene, AP-1 transcription factor subunit	12	1.76	45
SPRY2	sprouty RTK signaling antagonist 2	1	1.11	48
CDK6	cyclin dependent kinase 6	16	0.54	67
CSF1	colony stimulating factor 1	2	0.91	75



Table 11. Prospective drug targets selected from full list of identified master regulators filtered by Druggability score predicted by PASS software. Here, the **Druggability score** for master regulator proteins is computed as a sum of PASS calculated probabilities to be active as a target for various small molecular compounds. The drug targets are sorted according to the **Total rank** which is the sum of three ranks computed on the basis of the three scores: keynode score, CMA score and expression change score (logFC, if present). See Methods section for details.

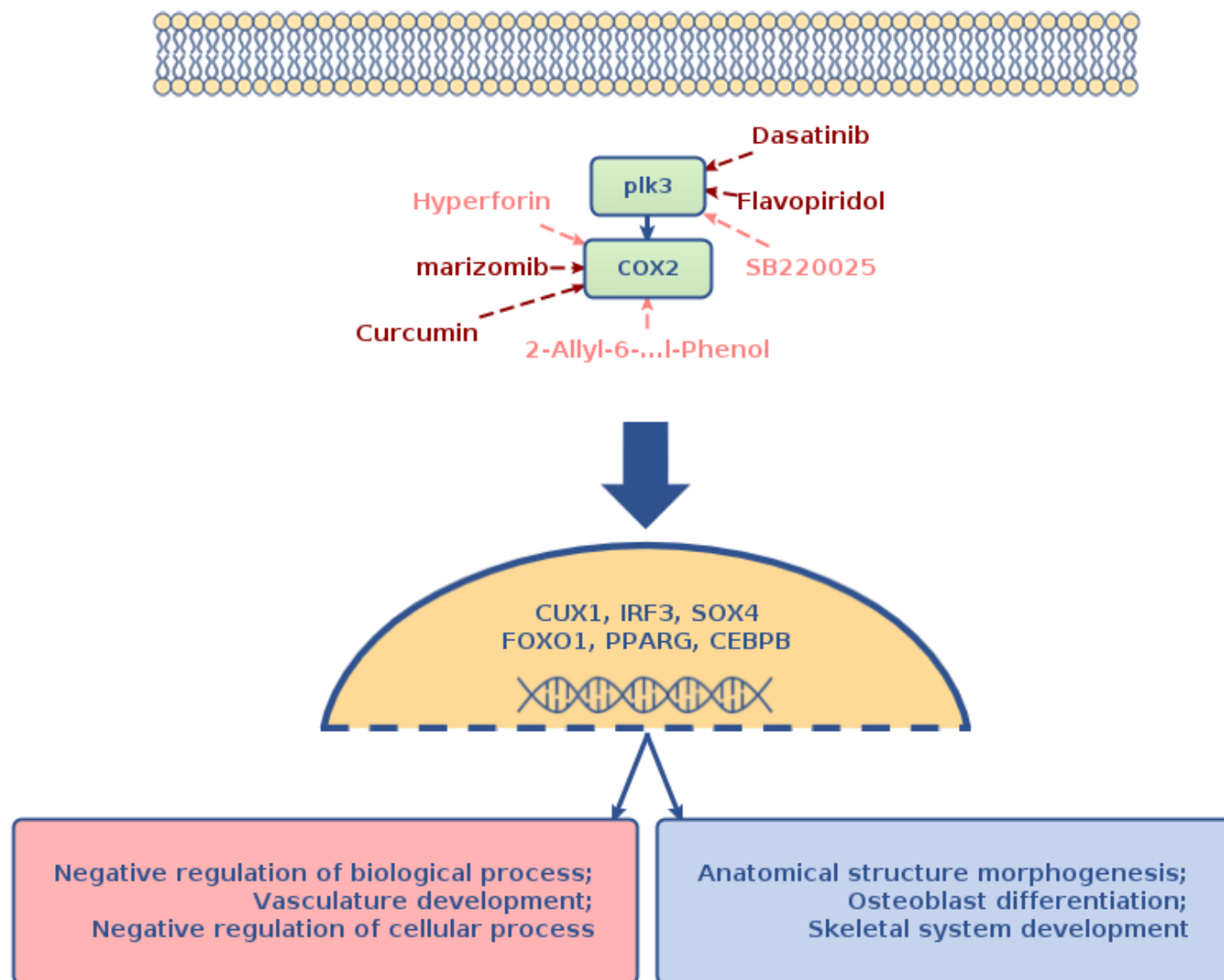
[See full table](#) →

Gene symbol	Gene Description	Druggability score	logFC	Total rank
PTGS2	prostaglandin-endoperoxide synthase 2	6.3	2.83	28
PLK3	polo like kinase 3	2.75	1.46	32
CDK6	cyclin dependent kinase 6	17.42	0.54	67
CSF1	colony stimulating factor 1	236.37	0.91	75
CXCL8	C-X-C motif chemokine ligand 8	63.76	3.06	80
NGF	nerve growth factor	14.08	1.4	81

Below we represent schematically the main mechanism of the studied pathology. In the schema we considered the top two drug targets of each of the two categories computed above. In addition we have added two top identified master regulators for which no drugs may be identified yet, but that are playing the crucial role in the molecular mechanism of the studied pathology. Thus the molecular mechanism of the studied pathology was predicted to be mainly based on the following key master regulators:

- COX2
- plk3

This result allows us to suggest the following schema of affecting the molecular mechanism of the studied pathology:



Drugs which are shown on this schema: Dasatinib, Hyperforin, Flavopiridol, SB220025, marizomib, 2-Allyl-6-Methyl-Phenol and Curcumin, should be considered as a prospective research initiative for further drug repurposing and drug development. These drugs were selected as top matching treatments to the most prospective drug targets of the studied pathology, however, these results should be considered with special caution and are to be used for research purposes only, as there is not enough clinical information for adapting these results towards immediate treatment of patients.

The drugs given in dark red color on the schema are FDA approved drugs or drugs which have gone through various phases of clinical trials as active treatments against the selected targets.

The drugs given in pink color on the schema are drugs, which were cheminformatically predicted to be active against the selected targets.

5. Identification of potential drugs

In the last step of the analysis we strived to identify known activities as well as drugs with cheminformatically predicted activities that are potentially suitable for inhibition (or activation) of the identified molecular targets in the context of specified human diseases(s).

Proposed drugs are top ranked drug candidates, that were found to be active on the identified targets and were selected from 4 categories:

1. FDA approved drugs or used in clinical trials drugs for the studied pathology;
2. Repurposing drugs used in clinical trials for other pathologies;
3. Drugs, predicted by PASS to be active against identified drug targets and against the studied pathology;
4. Drugs, predicted by PASS to be active against identified drug targets but for other pathologies.

Proposed drugs were selected on the basis of Drug rank which was computed from the ranks sum based on the individual ranks of the following scores:

- Target activity score (depends on ranks of all targets that were found for the selected drug);
- Disease activity score (weighted sum of number of clinical trials on disease(s) under study where the selected drug is known to be applied or PASS Disease activity score - cheminformatically predicted property of the compound to be active against the studied disease(s));
- Clinical validity score (applicable only for drugs predicted on the basis of literature curation in HumanPSD™ database (Tables 12 and 13), reflects the number of the highest clinical trials phase on which the drug was tested for any pathology).

You can refer to the Methods section for more details on drug ranking procedure.

Based on the Drug rank, a numerical value of Drug score was calculated, which reflects the potential activity of the respective drug on the overall molecular mechanism of the studied pathology. Drug score values belong to the range from 1 to 100 and are calculated as a quotient of maximum drug rank and the drug rank of the given drug multiplied by 100.

Top drugs of each category are given in the tables below:

Drugs approved in clinical trials



Table 12. FDA approved drugs or drugs used in clinical trials for the studied pathology (most promising treatment candidates selected for the identified drug targets on the basis of literature curation in [HumanPSD™](#) database)

[See full table](#) →

Name	Target names	Drug score	Disease activity score	Disease trial phase
Curcumin	PTGS2, HDAC4, CD44, CDK6, MET, HSPA5, MAPK6, HK2, LEF1, MYC, PRKAA1, NFE2L2, STAT3, SLC2A1, BAX, SNAI1, CHUK, CXCL8, JAK1, MAPK8, VEGFA, CASP7, TGFB1, APP, EGR1, NFKBIA, MYD88, SQSTM1, CASP3	97	3	Phase 2: Alzheimer Disease, Aberrant Crypt Foci, Acute Kidney Injury, Adenocarcinoma, Amyotrophic Lateral Sclerosis, Aneurysm, Anorexia, Aortic Aneurysm, Aortic Aneurysm, Abdominal, Asthma, Bipolar Disorder, Bites and Stings, Brain Abscess, Cachexia, Cognitive Dysfunction, Cytopenia, Diabetes Mellitus, Diabetes Mellitus, Type 2, Diabetic Nephropathies, Diverticulitis, Dyslipidemias, Edema, Fibrosis, Head and Neck Neoplasms, Hyperemia, Hyperlipidemias, Inflammation, Insulin Resistance, Kidney Diseases, Leukemia, Lung Diseases, Lung Neoplasms, Lupus Nephritis, Macular Edema, Mental Disorders, Monoclonal Gammopathy of Undetermined Significance, Motor Neuron Disease, Mucositis, Multiple Myeloma, Multiple Sclerosis, Myelodysplastic Syndromes, Myeloproliferative Disorders, Neoplasms, Neoplasms, Plasma Cell, Nephritis, Oral Submucous Fibrosis, Osteosarcoma, Pain, Paraproteinemias, Polycythemia, Polycythemia Vera, Preleukemia, Prostatic Neoplasms, Proteinuria, Psychotic Disorders, Renal Insufficiency, Renal Insufficiency, Chronic, Sarcoma, Schizophrenia, Sclerosis, Smoldering Multiple Myeloma, Syndrome, Thrombocythemia, Essential, Thrombocytosis, Tic Disorders, Tobacco Use, Tobacco Use Disorder, Wasting Syndrome, Wounds and Injuries
L-Sulforaphane	MAPK8, NFE2L2, STAT3, CD44, MAPK6, CASP3, BAX, JUN, CXCL8, PRKAA1, NFKBIA	93	1	N/A
Dasatinib	MAPK8, PIM2, CDK8, BMPR1A, MET, CSNK2A2, PIM3, MAPK6, LATS2, YES1, PRKAA1, CDK7, PLK3, EPHB1, MAP2K3, NEK6, EPHA2, TGFB2, DDR2, CSK, FYN, PRKD1, JAK1, STK3	90	3	Phase 2: Alzheimer Disease, Aging, Premature, Blast Crisis, Bone Neoplasms, Brain Abscess, Brain Stem Neoplasms, Breast Neoplasms, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Renal Cell, Carcinoma, Transitional Cell, Central Nervous System Neoplasms, Cholangiocarcinoma, Chondrosarcoma, Chordoma, Cognitive Dysfunction, Colonic Neoplasms, Colorectal Neoplasms, Depression, Depressive Disorder, Depressive Disorder, Treatment-Resistant, Diffuse Intrinsic Pontine Glioma, Endometrial Neoplasms, Esophageal Neoplasms, Fibrosarcoma, Frailty, Gastrointestinal Stromal Tumors, Giant Cell Tumor of Bone, Giant Cell Tumors, Glioblastoma, Glioma, Hemangiopericytoma, Hypereosinophilic Syndrome, Kidney Neoplasms, Leukemia, Leukemia, B-Cell, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Leukemia, Myeloid,

				Acute, Leukemia, Myeloid, Chronic-Phase, Leukemia, Myelomonocytic, Chronic, Liver Neoplasms, Lung Neoplasms, Lymphoma, Lymphoma, Non-Hodgkin, Mastocytosis, Melanoma, Mental Disorders, Metaplasia, Multiple Myeloma, Myelodysplastic Syndromes, Myosarcoma, Neoplasms, Neoplasms, Germ Cell and Embryonal, Neoplasms, Plasma Cell, Nerve Sheath Neoplasms, Nervous System Neoplasms, Neuroblastoma, Neuroectodermal Tumors, Neuroectodermal Tumors, Primitive, Neuroectodermal Tumors, Primitive, Peripheral, Neurofibrosarcoma, Ovarian Neoplasms, Pancreatic Neoplasms, Polycythemia, Polycythemia Vera, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Preleukemia, Primary Myelofibrosis, Prostatic Neoplasms, Psychotic Disorders, Rectal Neoplasms, Recurrence, Rhabdomyosarcoma, Rhabdomyosarcoma, Alveolar, Rhabdomyosarcoma, Embryonal, Sarcoma, Sarcoma, Alveolar Soft Part, Sarcoma, Ewing, Schizophrenia, Skin Neoplasms, Solitary Fibrous Tumors, Stomach Neoplasms, Syndrome, Thyroid Diseases, Thyroid Neoplasms, Tic Disorders, Triple Negative Breast Neoplasms, Urinary Bladder Neoplasms, Uterine Neoplasms
5-(2,6-dichlorophenyl)-2-[(2,4-difluorophenyl)sulfanyl]-6H-pyrimido[1,6-b]pyridazin-6-one	MAPK8, PIM2, CDK8, BMPR1A, MET, CSNK2A2, PIM3, MAPK6, LATS2, YES1, PRKAA1, CDK7, PLK3, EPHB1, MAP2K3, NEK6, EPHA2, TGFBR2, DDR2, CSK, FYN, PRKD1, JAK1, STK3	89	2	Phase 2: Alzheimer Disease, Encephalitis, Huntington Disease, Infectious Encephalitis, Inflammation
MPC-7869	PTGS2, APP	88	3	Phase 3: Alzheimer Disease, Dementia

The **Disease trial phase** column reflects the maximum clinical trials phase in which the drug was studied for the analyzed pathology.

Repurposing drugs



Table 13. Repurposed drugs used in clinical trials for other pathologies (prospective drugs against the identified drug targets on the basis of literature curation in [HumanPSD™](#) database)

[See full table](#) →

Name	Target names	Drug score	Maximum trial phase
Niclosamide	MAPK8, MAPK6, JUN, FOS, MYC, CYCS, NFKBIA, STAT3, CASP3, BAX, SNAI1, CHUK, CXCL8	90	Phase 3: COVID-19, Diabetic Nephropathies, Kidney Diseases, Severe Acute Respiratory Syndrome
Ingenol Mebutate	MAPK8, TGFB1, MAPK6, CASP3, SPRY2, EGR1, SNAI1, FOS, CXCL8	90	Phase 4: Keratosis, Keratosis, Actinic
marizomib	VEGFA, PTGS2, SQSTM1, CASP7, CASP3, BAX, SNAI1, MYC, CASP4, NFKBIA	90	Phase 2: Brain Abscess, Ependymoma, Multiple Myeloma, Neoplasms, Neoplasms, Plasma Cell, Recurrence
Hydroxytyrosol	NFE2L2, STAT3, SQSTM1, CASP7, CASP3, BAX, JUN, EGR1, FOS	90	NA: Coronary Artery Disease, Coronary Disease, Diabetes Mellitus, Diabetes Mellitus, Type 2, Glucose Intolerance, Inflammation, Ischemia, Metabolic Diseases, Mitochondrial Diseases, Myocardial Ischemia, Obesity, Prediabetic State, Sarcopenia
wnt-974	MAPK8, PTGS2, CDK6, MAPK6, BAX, SNAI1, MYC, CXCL8, NFKBIA	89	Phase 1: Melanoma

The **Maximum trial phase** column reflects the maximum clinical trials phase in which the drug was studied for any pathology.



Table 14. Prospective drugs, predicted by [PASS](#) software to be active against the identified drug targets with predicted activity against the studied disease(s) (drug candidates predicted with the cheminformatics tool PASS)

[See full table](#) →

Name	Target names	Drug score	Target activity score
Hyperforin	PTGS2, ADAM17, MYC	98	0.24
Lipoic Acid	DUSP26, PTGS2, CSF1, PTPN9, DUSP1, PTPN12, DUSP16, DUSP6, PTPRC	97	0.21
Physostigmine	CHKA, APP	96	0.2
Phenserine	CHKA, APP	96	0.19
(2s,3s)-Trans-Dihydroquercetin	NFKB2, MAPK8, CASP3, MAPK6, NFKB1, MYC, NGF	94	0.17



Table 15. Prospective drugs, predicted by *PASS* software to be active against the identified drug targets, though without cheminformatically predicted activity against the studied disease(s) (drug candidates predicted with the cheminformatics tool *PASS*)

[See full table](#) →

Name	Target names	Drug score	Target activity score
Bortezomib	NFKB2, PSMC5, PSMA7, CASP3, NFKB1, PSMC2, PSMD14	96	0.31
SB220025	MAPK8, PLK3, MAPK6, YES1, CXCL8, JAK1	96	0.27
1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea	MAPK8, EPHB1, MAPK6, DDR2, YES1, CXCL8	95	0.3
UBIQUINONE-1	NFKB2, VEGFA, VEGFD, CASP3, NFKB1, MYC	94	0.21
Silibinin	NFKB2, CASP3, NFKB1, NGF	94	0.17

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Curcumin, Niclosamide, Hyperforin and Bortezomib. These drugs were selected for acting on the following targets: PTGS2, JUN and NFKB2, which were predicted to be active in the molecular mechanism of the studied pathology.

The selected drugs are top ranked drug candidates from each of the four categories of drugs: (1) FDA approved drugs or used in clinical trials drugs for the studied pathology; (2) repurposing drugs used in clinical trials for other pathologies; (3) drugs, predicted by *PASS* software to be active against the studied pathology; (4) drugs, predicted by *PASS* software to be repurposed from other pathologies.

6. Conclusion

We applied the software package "Genome Enhancer" to a data set that contains *transcriptomics* data. The study is done in the context of *Alzheimer Disease*. The data were pre-processed, statistically analyzed and differentially expressed genes were identified. Also checked was the enrichment of GO or disease categories among the studied gene sets.

We propose the following drugs as most promising candidates for treating the pathology under study:



Curcumin, Niclosamide, Hyperforin and Bortezomib

These drugs were selected for acting on the following targets: PTGS2, JUN and NFKB2, which were predicted to be involved in the molecular mechanism of the pathology under study.

The identified molecular mechanism of the studied pathology was predicted to be mainly based on the following key drug targets:



COX2 and plk3

These potential drug targets should be considered as a prospective research initiative for further drug repurposing and drug development purposes. The following drugs were predicted as, matching those drug targets: Dasatinib, Hyperforin, Flavopiridol, SB220025, marizomib, 2-Allyl-6-Methyl-Phenol and Curcumin. These drugs should be considered with special caution for research purposes only.

In this study, we came up with a detailed signal transduction network regulating differentially expressed genes in the studied pathology. In this network we have revealed the following top master regulators (signaling proteins and their complexes) that play a crucial role in the molecular mechanism of the studied pathology, which can be proposed as the most promising molecular targets for further drug repurposing and drug development initiatives.

- COX2
- plk3

Potential drug compounds which can be affecting these targets can be found in the "Finding prospective drug targets" section.

7. Methods

Databases used in the study

Transcription factor binding sites in promoters and enhancers of differentially expressed genes were analyzed using known DNA-binding motifs described in the **TRANSFAC®** library, release 2025.2 (geneXplain GmbH, Wolfenbüttel, Germany) (<https://genexplain.com/transfac>).

The master regulator search uses the **TRANSPATH®** database (BIOBASE), release 2025.2 (geneXplain GmbH, Wolfenbüttel, Germany) (<https://genexplain.com/transpath>). A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in **TRANSPATH®**.

The information about drugs corresponding to identified drug targets and clinical trials references were extracted from **HumanPSD™** database, release 2025.2 (<https://genexplain.com/humanpsd>).

The Ensembl database release Human112.38 (hg38) (<http://www.ensembl.org>) was used for gene IDs representation and Gene Ontology (GO) (<http://geneontology.org>) was used for functional classification of the studied gene set.

Methods for the analysis of enriched transcription factor binding sites and composite modules

Transcription factor binding sites in promoters and enhancers of differentially expressed genes were analyzed using known DNA-binding motifs. The motifs are specified using position weight matrices (PWMs) that give weights to each nucleotide in each position of the DNA binding motif for a transcription factor or a group of them.

We search for transcription factor binding sites (TFBS) that are enriched in the promoters and enhancers under study as compared to a background sequence set such as promoters of genes that were not differentially regulated under the condition of the experiment. We denote study and background sets briefly as Yes and No sets. In the current work we used a workflow considering promoter sequences of a standard length of 1100 bp (-1000 to +100). The error rate in this part of the pipeline is controlled by estimating the adjusted p-value (using the Benjamini-Hochberg procedure) in comparison to the TFBS frequency found in randomly selected regions of the human genome (adj.p-value < 0.01).

We have applied the CMA algorithm (Composite Module Analyst) for searching composite modules [6] in the promoters and enhancers of the Yes and No sets. We searched for a composite module consisting of a cluster of 10 TFs in a sliding window of 200-300 bp that statistically significantly separates sequences in the Yes and No sets (minimizing Wilcoxon p-value).

Methods for finding master regulators in networks

We searched for master regulator molecules in signal transduction pathways upstream of the identified transcription factors. The master regulator search uses a comprehensive signal transduction network of human cells. The main algorithm of the master regulator search has been described earlier [3,4]. The goal of the algorithm is to find nodes in the global signal transduction network that may potentially regulate the activity of a set of transcription factors found at the previous step of the analysis. Such nodes are considered as most promising drug targets, since any influence on such a node may switch the transcriptional programs of hundreds of genes that are regulated by the respective TFs. In our analysis, we have run the algorithm with a maximum radius of 12 steps upstream of each TF in the input set. The error rate of this algorithm is controlled by applying it 10000 times to randomly generated sets of input transcription factors of the same set-size. Z-score and FDR value of ranks are calculated then for each potential master regulator node on the basis of such random runs (see detailed description in [8]). We control the error rate by the FDR threshold 0.05.

To identify feed forward loops in transduction network we calculate regulatory scores for transcription factors from CMA result and keynodes analysis. The matrices from the CMA model are converted to TFs. Keynode analysis is then performed on these TFs and resulting keynodes are sorted by keynode score. Then the keynodes, which are regulated

$$S_j = \sum_{j=1}^{M_j} \frac{1}{D_{ij}}; D_{ij} = \sum_{k=1}^{C_{ij}} d_k$$

by the model, are selected.

Here:

- j – a TF (Transcription Factor),
- i – a MR (Master Regulator),
- S_j – Regulatory score of the TF,
- D_{ij} – Cumulative distance between MR i and TF j ,
- M_j – number of MRs whose signal can reach the TF j and whose gene is regulated by TF j ,
- C_{ij} – number of steps in the network from between MR i and TF j (shortest path),
- d_k distance of the step k in the shortest path between MR i and TF j .

In unweighted TRANSPATH network $d_k = 1$ for each direct signaling reaction (like $A+B-C = D$), $d_k \geq 3$ for each semantic reaction (like $A \rightarrow B$). It depends on type of the reaction, basic or orthology level of it's components, species of the components.

In weighted TRANSPATH network d_k is changed after applying the Context Algorithm as described in [4].

Methods for analysis of pharmaceutical compounds

We seek for the optimal combination of molecular targets (key elements of the regulatory network of the cell) that potentially interact with pharmaceutical compounds from a library of known drugs and biologically active chemical compounds, using information about known drugs from HumanPSD™ and predicting potential drugs using [PASS](#) program.

Method for analysis of known pharmaceutical compounds

We selected compounds from HumanPSD™ database that have at least one target. Next, we sort compounds using "Drug rank" that is the sum of the following ranks:

1. ranking by "Target activity score" ($T\text{-score}_{PSD}$),
2. ranking by "Disease activity score" ($D\text{-score}_{PSD}$),
3. ranking by "Clinical validity score".

"Target activity score" ($T\text{-score}_{PSD}$) is calculated as follows:

$$T\text{-score}_{PSD} = -\frac{|T|}{|T| + w(|AT| - |T|)} \sum_{t \in T} \log_{10} \left(\frac{rank(t)}{1 + \max Rank(T)} \right),$$

where T is set of all targets related to the compound intersected with input list, $|T|$ is number of elements in T , AT and $|AT|$ are set set of all targets related to the compound and number of elements in it, w is weight multiplier, $rank(t)$ is rank of given target, $\max Rank(T)$ equals $\max(rank(t))$ for all targets t in T .

We use following formula to calculate "Disease activity score" ($D\text{-score}_{PSD}$):

$$D\text{-score}_{PSD} = \begin{cases} \sum_{d \in D} \sum_{p \in P} phase(d, p) \\ 0, D = \emptyset \end{cases},$$

where D is the set of selected diseases, and if D is empty set, $D\text{-score}_{PSD} = 0$. P is a set of all known phases for each disease, $phase(p, d)$ equals to the phase number if there are known clinical trials for the selected disease on this phase and zero otherwise.

The clinical validity score reflects the number of the highest clinical trials phase (from 1 to 4) on which the drug was ever tested for any pathology.

Method for prediction of pharmaceutical compounds

In this study, the focus was put on compounds with high pharmacological efficiency and low toxicity. For this purpose, comprehensive library of chemical compounds and drugs was subjected to a SAR/QSAR analysis. This library contains 13040 compounds along with their pre-calculated potential pharmacological activities of those substances, their possible side and toxic effects, as well as the possible mechanisms of action. All biological activities are expressed as probability values for a substance to exert this activity (Pa).

We selected compounds that satisfied the following conditions:

1. Toxicity below a chosen toxicity threshold (defines as Pa , probability to be active as toxic substance).
2. For all predicted pharmacological effects that correspond to a set of user selected disease(s) Pa is greater than a chosen effect threshold.
3. There are at least 2 targets (corresponding to the predicted activity-mechanisms) with predicted Pa greater than a chosen target threshold.

The maximum Pa value for all toxicities corresponding to the given compound is selected as the "Toxicity score". The maximum Pa value for all activities corresponding to the selected diseases for the given compound is used as the "Disease activity score". "Target activity score" (T-score) is calculated as follows:

$$T\text{-score}(s) = \frac{|T|}{|T| + w(|AT| - |T|)} \sum_{m \in M(s)} \left(pa(m) \sum_{g \in G(m)} IAP(g) optWeight(g) \right),$$

where $M(s)$ is the set of activity-mechanisms for the given structure (which passed the chosen threshold for activity-mechanisms Pa); $G(m)$ is the set of targets (converted to genes) that corresponds to the given activity-mechanism (m) for the given compound; $pa(m)$ is the probability to be active of the activity-mechanism (m), $IAP(g)$ is the invariant accuracy of prediction for gene from $G(m)$; $optWeight(g)$ is the additional weight multiplier for gene. T is set of all targets related to the compound intersected with input list, $|T|$ is number of elements in T , AT and $|AT|$ are set set of all targets related to the compound and number of elements in it, w is weight multiplier.

"Druggability score" (D-score) is calculated as follows:

$$D\text{-score}(g) = IAP(g) \sum_{s \in S(g)} \sum_{m \in M(s,g)} pa(m),$$

where $S(g)$ is the set of structures for which target list contains given target, $M(s,g)$ is the set of activity-mechanisms (for the given structure) that corresponds to the given gene, $pa(m)$ is the probability to be active of the activity-mechanism (m), $IAP(g)$ is the invariant accuracy of prediction for the given gene.

8. References

1. Kel A, Voss N, Jauregui R, Kel-Margoulis O, Wingender E. Beyond microarrays: Finding key transcription factors controlling signal transduction pathways. *BMC Bioinformatics*. **2006**;7(S2), S13. doi:10.1186/1471-2105-7-s2-s13
2. Stegmaier P, Voss N, Meier T, Kel A, Wingender E, Borlak J. Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer. *PLoS ONE*. **2011**;6(3):e17738. doi:10.1371/journal.pone.0017738
3. Koschmann J, Bhar A, Stegmaier P, Kel A, Wingender E. "Upstream Analysis": An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays*. **2015**;4(2):270-286. doi:10.3390/microarrays4020270.
4. Kel A, Stegmaier P, Valeev T, Koschmann J, Poroikov V, Kel-Margoulis OV, and Wingender E. Multi-omics "upstream analysis" of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer. *EuPA Open Proteom*. **2016**;13:1-13. doi:10.1016/j.euprot.2016.09.002
5. Matys V, Kel-Margoulis OV, 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 AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. **2006**;34(90001):D108-D110. doi:10.1093/nar/gkj143

6. Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* **2003**;31(13):3576-3579. doi:10.1093/nar/gkg585
7. Waleev T, Shtokalo D, Konovalova T, Voss N, Cheremushkin E, Stegmaier P, Kel-Margoulis O, Wingender E, Kel A. Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res.* **2006**;34(Web Server issue):W541-5.
8. Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.* **2006**;34(90001):D546-D551. doi:10.1093/nar/gkj107
9. Boyarskikh U, Pintus S, Mandrik N, Stelmashenko D, Kiselev I, Evshin I, Sharipov R, Stegmaier P, Kolpakov F, Filipenko M, Kel A. Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3. *BMC Med Genomics.* **2018**;11(1):12. doi:10.1186/1471-2105-7-s2-s13
10. Kel, A., Boyarskikh, U., Stegmaier, P., Leskov, L.S., Sokolov, A.V., Yevshin, I., Mandrik, N., Stelmashenko, D., Koschmann, J., Kel-Margoulis, O. and Krull, M. Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer. *BMC bioinformatics.* Cambridge (UK): RSC Publishing. **2019**;20(Suppl 4):119:1-20. doi:10.1186/s12859-019-2687-7
11. Michael H, Hogan J, Kel A et al. Building a knowledge base for systems pathology. *Brief Bioinformatics.* **2008**;9(6):518-531. doi:10.1093/bib/bbn038
12. Filimonov D, Poroikov V. Probabilistic Approaches in Activity Prediction. Varnek A, Tropsha A. *Cheminformatics Approaches to Virtual Screening.* Cambridge (UK): RSC Publishing. **2008**;:182-216.
13. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal.* **2006**;50(2):66-75 (russ)
14. Filimonov D, Poroikov V, Borodina Y, Glorizova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *ChemInform.* **1999**;39(4):666-670. doi:10.1021/ci980335o

Thank you for using the Genome Enhancer!

In case of any questions please contact us at support@genexplain.com

Supplementary material

1. [Supplementary table 1 - Up-regulated genes](#)
2. [Supplementary table 2 - Down-regulated genes](#)
3. [Supplementary table 3 - Detailed report. Composite modules and master regulators \(up-regulated genes in MC65 induced APP vs. MC65 non-induced APP\).](#)
4. [Supplementary table 4 - Detailed report. Composite modules and master regulators \(down-regulated genes in MC65 induced APP vs. MC65 non-induced APP\).](#)
5. [Supplementary table 5 - Detailed report. Pharmaceutical compounds and drug targets.](#)

Disclaimer

Decisions regarding care and treatment of patients should be fully made by attending doctors. The predicted chemical compounds listed in the report are given only for doctor's consideration and they cannot be treated as prescribed medication. It is the physician's responsibility to independently decide whether any, none or all of the predicted compounds can be used solely or in combination for patient treatment purposes, taking into account all applicable information regarding FDA prescribing recommendations for any therapeutic and the patient's condition, including, but not limited to, the patient's and family's medical history, physical examinations, information from various diagnostic tests, and patient preferences in accordance with the current standard of care. Whether or not a particular patient will benefit from a selected therapy is based on many factors and can vary significantly.

The compounds predicted to be active against the identified drug targets in the report are not guaranteed to be active against any particular patient's condition. GeneXplain GmbH does not give any assurances or guarantees regarding the

treatment information and conclusions given in the report. There is no guarantee that any third party will provide a refund for any of the treatment decisions made based on these results. None of the listed compounds was checked by Genome Enhancer for adverse side-effects or even toxic effects.

The analysis report contains information about chemical drug compounds, clinical trials and disease biomarkers retrieved from the HumanPSD™ database of gene-disease assignments maintained and exclusively distributed worldwide by geneXplain GmbH. The information contained in this database is collected from scientific literature and public clinical trials resources. It is updated to the best of geneXplain's knowledge however we do not guarantee completeness and reliability of this information leaving the final checkup and consideration of the predicted therapies to the medical doctor.

The scientific analysis underlying the Genome Enhancer report employs a complex analysis pipeline which uses geneXplain's proprietary Upstream Analysis approach, integrated with TRANSFAC® and TRANSPATH® databases maintained and exclusively distributed worldwide by geneXplain GmbH. The pipeline and the databases are updated to the best of geneXplain's knowledge and belief, however, geneXplain GmbH shall not give a warranty as to the characteristics or to the content and any of the results produced by Genome Enhancer. Moreover, any warranty concerning the completeness, up-to-dateness, correctness and usability of Genome Enhancer information and results produced by it, shall be excluded.

The results produced by Genome Enhancer, including the analysis report, severely depend on the quality of input data used for the analysis. It is the responsibility of Genome Enhancer users to check the input data quality and parameters used for running the Genome Enhancer pipeline.

Note that the text given in the report is not unique and can be fully or partially repeated in other Genome Enhancer analysis reports, including reports of other users. This should be considered when publishing any results or excerpts from the report. This restriction refers only to the general description of analysis methods used for generating the report. All data and graphics referring to the concrete set of input data, including lists of mutated genes, differentially expressed genes/proteins/metabolites, functional classifications, identified transcription factors and master regulators, constructed molecular networks, lists of chemical compounds and reconstructed model of molecular mechanisms of the studied pathology are unique in respect to the used input data set and Genome Enhancer pipeline parameters used for the current run.