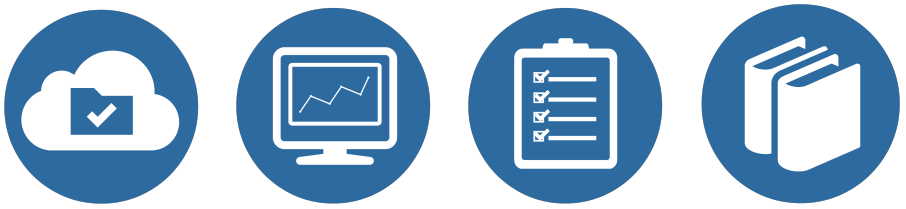
MIC-1(h), IGF-2(h), IGF-2-isoform2(h) and LRRK2(h) are promising master regulators that control activity of RELA, FOS and RXRA transcription factors that are potentially involved in the regulation of the differentially expressed genes

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 28/11/2025 ; Run on 09/12/2025 ; Report generated on 09/12/2025

Pathway Omics Suite release 3.6 (TRANSFAC®, TRANSPATH®release 2025.1)



Abstract

In the present study we applied the software package "Pathway Omics Suite" to a multiomics data set that contains *transcriptomics and epigenomics* data. The goal of this pipeline is to identify potential drug targets in the molecular network that governs the biological process under study. In the first step of analysis pipeline discovers transcription factors (TFs) that regulate genes activities under study. 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 biological process under study. 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: RELA, FOS, RUNX3, RXRA, SOX9 and PARP1.
The subsequent network analysis suggested MIC-1(h), IGF-2(h), IGF-2-isoform2(h) and LRRK2(h) as the most promising molecular targets for further research on the basis of identified molecular mechanism of the the biological process under study.

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 deviced 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 "Pathway Omics Suite".

2. Data

For this study the following experimental data was used:

Table 1. Experimental datasets used in the study

File name	Data type
Counts from GSE237606	Transcriptomics
GSM7634670_COHP_43566_1_UN_Whole_C1_NTWGS_L36501_H3HMFDSX2_GAGAATGG	Epigenomics
GSM7634671_COHP_43567_1_UN_Whole_C1_NTWGS_L36502_H3HMFDSX2_AACCATAG	Epigenomics
GSM7634690_COHP_43598_1_UN_Whole_C1_NTWGS_L36533_H3HMFDSX2_TATGCCTT	Epigenomics
GSM7634691_COHP_43599_1_UN_Whole_C1_NTWGS_L36534_H3HMFDSX2_TCAGAAGG	Epigenomics

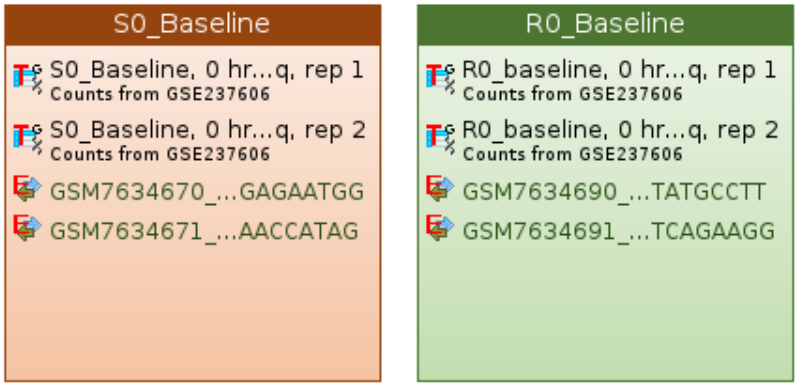


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

3. Results

We have compared the following conditions: R0_Baseline *versus* S0_Baseline.

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 edgeR tool (R/Bioconductor package integrated into our pipeline) and compared gene expression in the following sets: "R0_Baseline" with "S0_Baseline". edgeR 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 4372 upregulated genes (LogFC>0.5) out of which 2091 genes were found as significantly upregulated (p-value<0.01) and 4476 downregulated genes (LogFC<-0.5) out of which 1998 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 R0_Baseline vs. S0_Baseline.*
See full table →

ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000167244	IGF2	insulin like growth factor 2	11.12	6.04	0	0
ENSG00000107159	CA9	carbonic anhydrase 9	10.47	2.25	4.41E-96	7.65E-95
ENSG00000214548	MEG3	maternally expressed 3	9.71	1.51	1.78E-58	1.72E-57
ENSG00000131771	PPP1R1B	protein phosphatase 1 regulatory inhibitor subunit 1B	8.84	8.38	0	0
ENSG00000112559	MDFI	MyoD family inhibitor	8.64	3.57	3.72E-222	2.24E-220
ENSG00000130600	H19	H19 imprinted maternally expressed transcript	8.52	10.59	0	0
ENSG00000189058	APOD	apolipoprotein D	8.45	3.38	1.37E-193	6.68E-192
ENSG00000187037	GPR141	G protein-coupled receptor 141	8.22	2.31	5.53E-96	9.59E-95
ENSG00000185559	DLK1	delta like non-canonical Notch ligand 1	8.04	2.98	1.25E-149	3.94E-148
ENSG00000155849	ELMO1	engulfment and cell motility 1	7.08	3.89	1.87E-262	1.45E-260

*Table 3. Top ten significant **down-regulated** genes in R0_Baseline vs. S0_Baseline.*
See full table →

ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000145824	CXCL14	C-X-C motif chemokine ligand 14	-12.74	4.51	0	0
ENSG00000102287	GABRE	gamma-aminobutyric acid type A receptor subunit epsilon	-12.43	4.2	4.79E-282	4.11E-280
ENSG00000180964	TCEAL8	transcription elongation factor A like 8	-11.22	3	1.15E-156	3.96E-155
ENSG00000165169	DYNLT3	dynein light chain Tctex-type 3	-10.8	2.59	7.43E-120	1.7E-118
ENSG00000157502	PWWP3B	PWWP domain containing 3B	-10.71	2.5	3.84E-112	8.04E-111
ENSG00000146938	NLGN4X	neuroligin 4 X-linked	-10.66	2.45	1.29E-108	2.59E-107
ENSG00000268089	GABRQ	gamma-aminobutyric acid type A receptor subunit theta	-10.44	2.23	1.93E-94	3.28E-93
ENSG00000187601	MAGEH1	MAGE family member H1	-10.19	1.99	1.15E-77	1.55E-76
ENSG00000173210	ABLIM3	actin binding LIM protein family member 3	-9.95	1.76	9.76E-69	1.13E-67
ENSG00000214652	ZNF727	zinc finger protein 727	-9.83	1.64	1.18E-63	1.24E-62

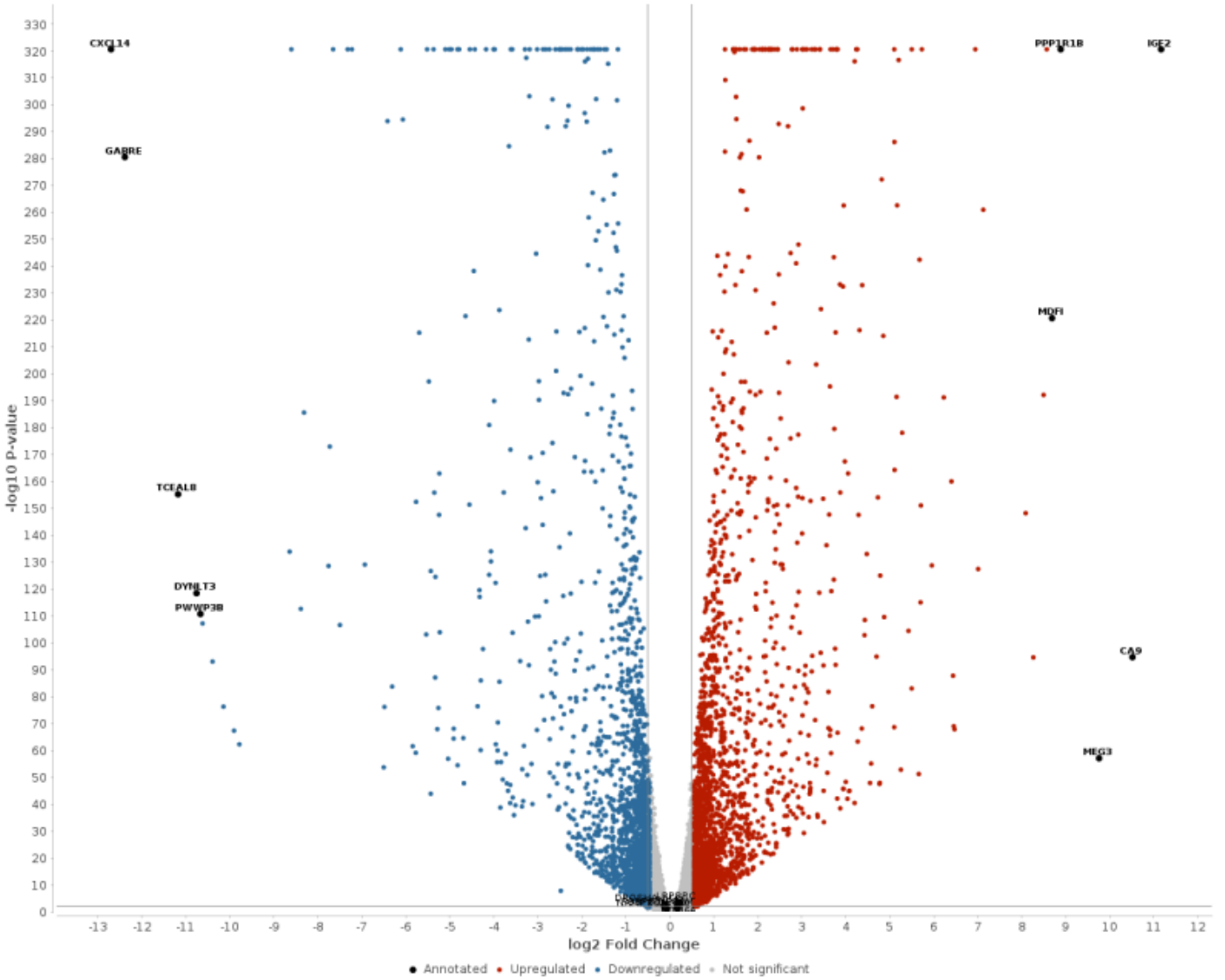


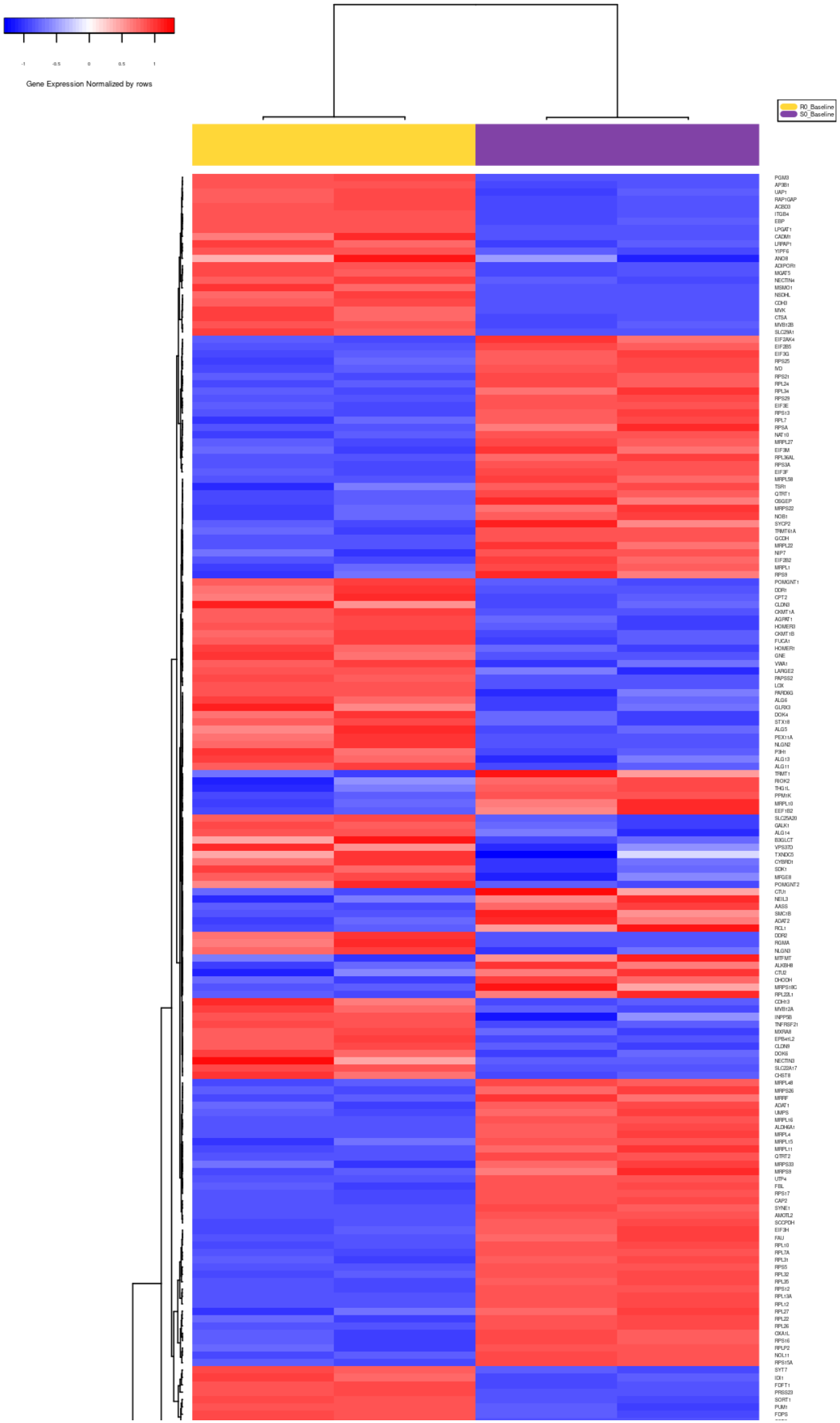
Figure 3. Volcano plot. Each dot represents one gene. Gray dots represent genes with no significant expression between conditions analyzed, the blue dots represent downregulated genes, and the red dots represent upregulated genes. Annotated are the top DEGs (bold) and the genes that encode top 10 up- 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 4-7 show the most significant categories.

Heatmap of differentially expressed genes in R0_Baseline vs. S0_Baseline

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


-1
-0.5
0
0.5
1
Gene Expression Normalized by rows
R0_Baseline
S0_Baseline
PGM3
AP3B1
UAP1
RAP1GAP
ACBD3
ITGB4
EBP
LPGAT1
CADM1
LRPAP1
YIPF6
ANO8
ADIPOR1
MGAT5
NECTIN4
MSMO1
NSDHL
CDH3
MVK
CTSA
MVB12B
SLC29A1
EIF2AK4
EIF2B5
EIF3G
RPS25
IVD
RPS21
RPL24
RPL34
RPS29
EIF3E
RPS13
RPL7
RPSA
NAT10
MRPL27
EIF3M
RPL36AL
RPS3A
EIF3F
MRPL58
TSR1
QTRT1
OSGEP
MRPS22
NOB1
SYCP2
TRMT61A
GCDH
MRPL22
NIP7
EIF2B2
MRPL1
RPS9
POMGNT1
DDR1
CPT2
CLDN9
CKMT1A
AGPAT1
HOMER3
CKMT1B
FUCA1
HOMER1
GNE
VWA1
LARGE2
PAPSS2
LOX
PARD6G
ALG6
GLRX3
DGK4
STX18
ALG5
PEX11A
NLGN2
P3H1
ALG13
ALG11
TRMT1
RIOK2
THG1L
PPM1K
MRPL10
EEF1B2
SLC25A20
GALK1
ALG14
B3GLCT
VPS37D
TXNDC5
CYBRD1
SDK1
MFGE8
POMGNT2
CTU1
NEIL3
AASS
SMC1B
ADAT2
RGL1
DDR2
RGMA
NLGN3
MTFMT
ALKBH8
CTU2
DHODH
MRPS18C
RPL22L1
CDH13
MVB12A
INPP5B
TNFRSF21
MXRA8
EPB41L2
CLDN8
DGKB
NECTIN3
SLC22A17
CHST8
MRPL48
MRPS26
MRPF
ADAT1
UMPS
MRPL16
ALDH6A1
MRPL4
MRPL15
MRPL11
QTRT2
MRPS33
MRPS9
UTP4
FBL
RPS17
CAP2
SYNE1
ANKTL2
SCCPDH
EIF3H
FAU
RPL10
RPL7A
RPL31
RPS5
RPL32
RPL35
RPS12
RPL13A
RPL12
RPL27
RPL22
RPL26
CKAP4L
RPS16
RPLP2
NOL11
RPS15A
SYT7
IDI1
FDFT1
PRSS23
SORT1
PUM1
FDPS

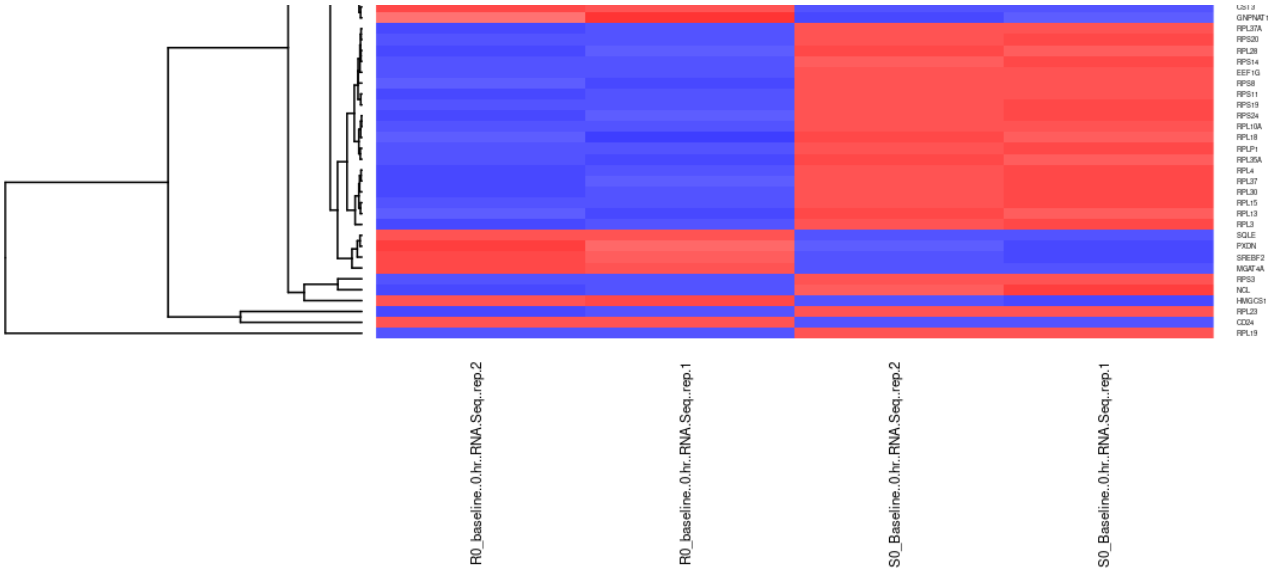


Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.

See full diagram →

Up-regulated genes in R0_Baseline vs. S0_Baseline:

2091 significant up-regulated genes were taken for the mapping.

GO (biological process)

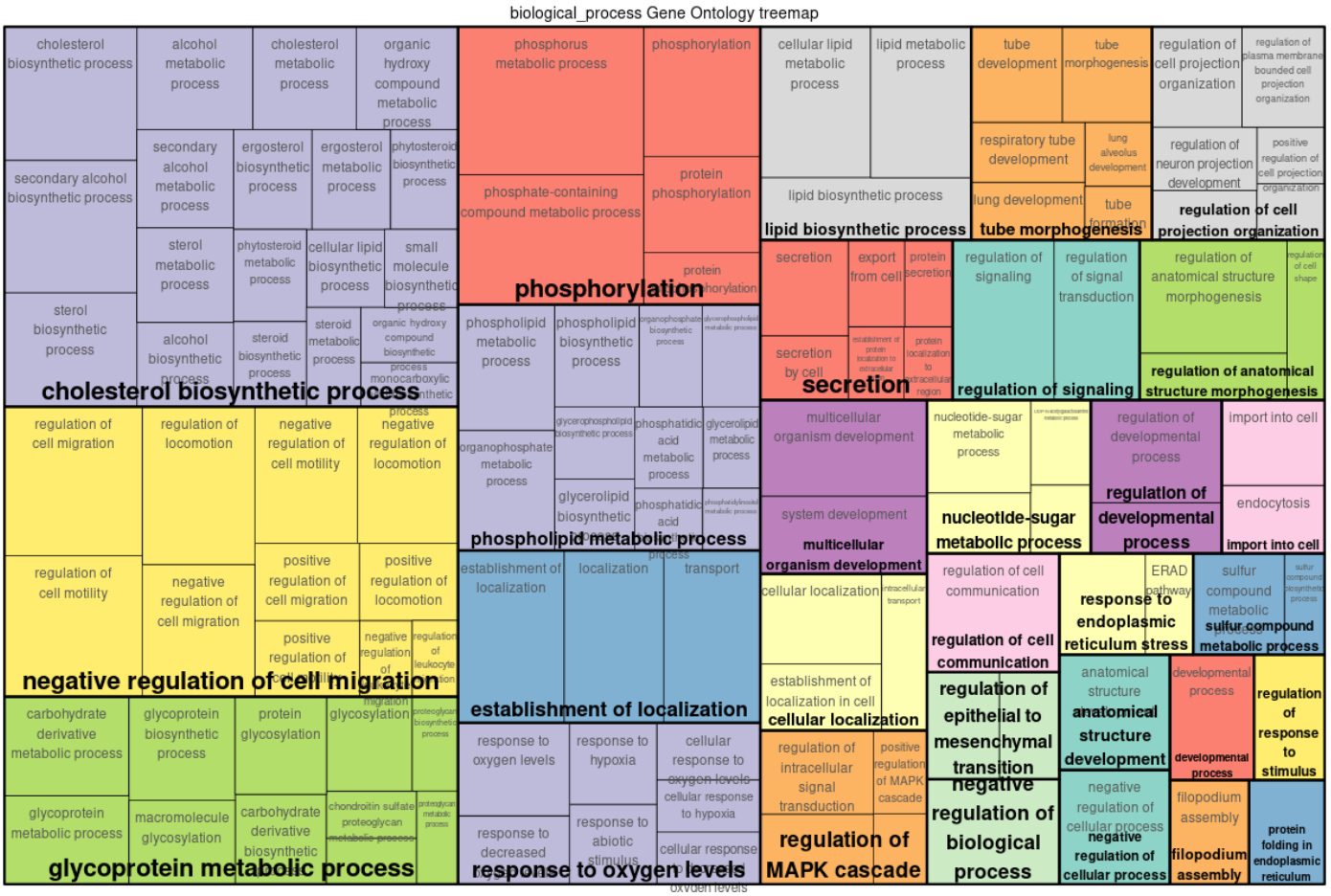


Figure 4. Enriched GO (biological process) of up-regulated genes in R0_Baseline vs. S0_Baseline.

Full classification →

TRANSPATH® Pathways (2025.1)

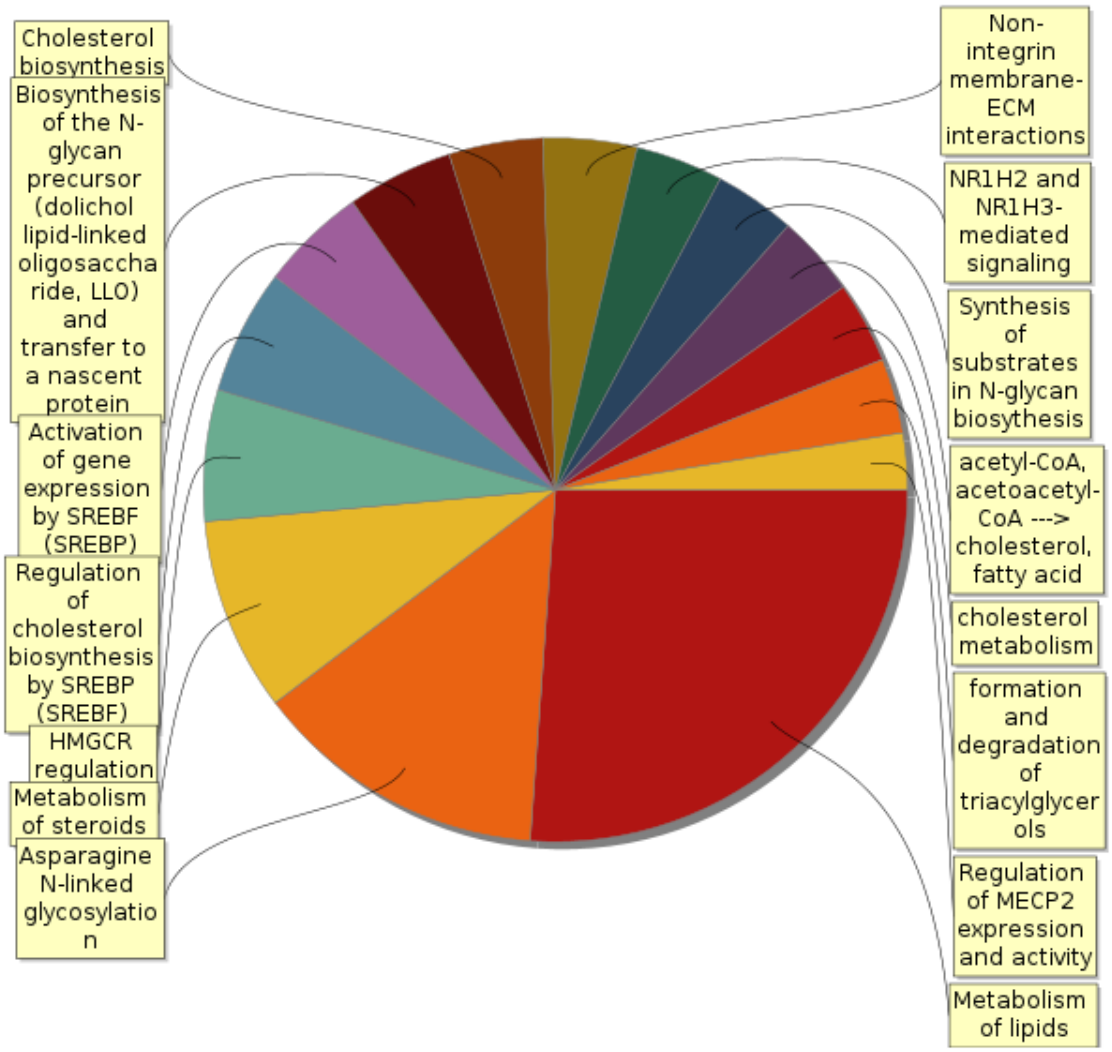


Figure 5. Enriched TRANSPATH® Pathways (2025.1) of up-regulated genes in R0_Baseline vs. S0_Baseline.
Full classification →

Down-regulated genes in R0_Baseline vs. S0_Baseline:

1998 significant down-regulated genes were taken for the mapping.

GO (biological process)

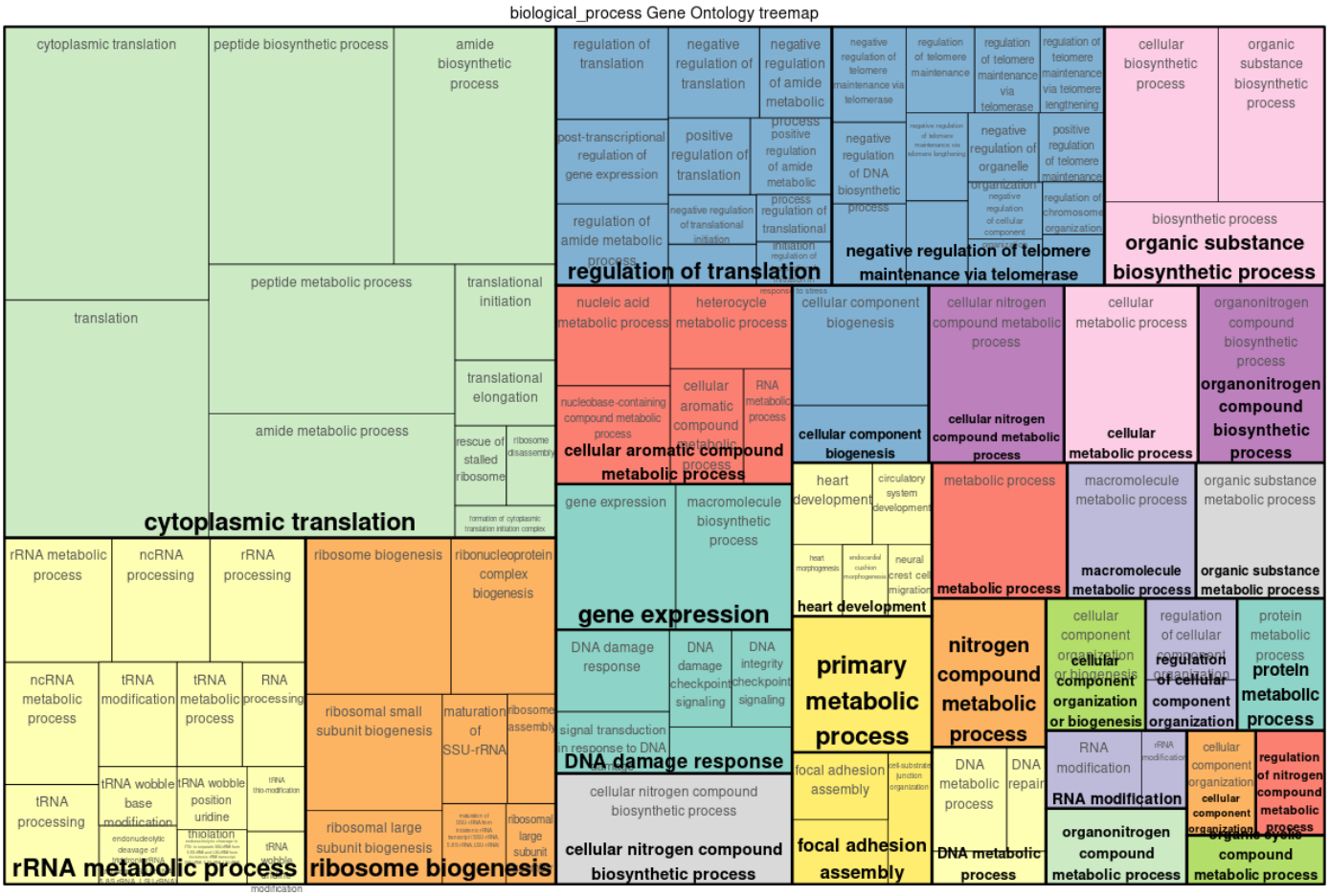


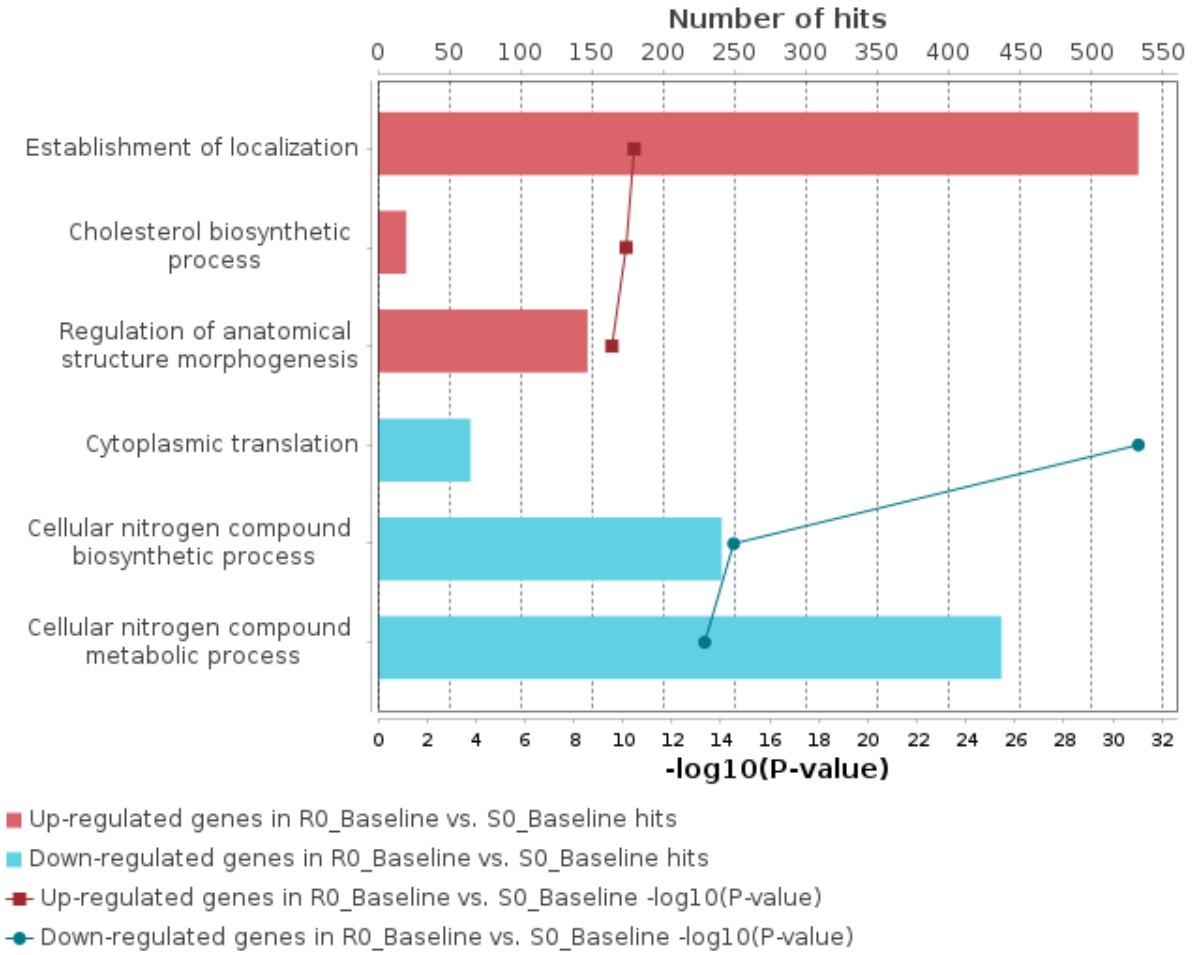
Figure 6. Enriched GO (biological process) of down-regulated genes in R0_Baseline vs. S0_Baseline.

Full classification →

TRANSPATH® Pathways (2025.1)

Figure 7. Enriched TRANSPATH® Pathways (2025.1) of down-regulated genes in R0_Baseline vs. S0_Baseline.
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].

In the current work we use the Epigenomics data from the track(s) "Subtracted peaks" to predict positions of potential ***enhancers*** regulating the differentially expressed genes revealed by comparative epigenomics analysis. We took genomic regions -550bp upstream and 550bp downstream from the middle point of each interval of the track and check if these regions are located inside the 5kb flanking areas of the differentially expressed genes (or inside the body of the genes). In such cases, these genomic regions are used for the search for potential condition-specific enhancers. In all other cases when the differentially expressed genes did not contain epigenomic peaks in their body or in the 5kb flanking regions we used the upstream regulatory regions of these genes (-1000bp upstream and 100bp downstream of TSS) for the search for condition-specific enhancers.

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 R0_Baseline vs. S0_Baseline).

To build the most specific composite modules we choose top 300 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 R0_Baseline vs. S0_Baseline.

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$DBP_Q6	V$MAFG_Q3	V$TEAD1_07	V$AML2_Q4	V$SIN3A_04	V$REST_06
0.92; N=3	0.72; N=3	0.82; N=3	0.90; N=3	0.77; N=3	0.87; N=2

Module width: 140

Module 2:

V$RXRA_16	V$JUNBFOS_01	V$RELA_03	V$ZNF462_01	V$FOXO4_06
0.86; N=3	0.81; N=3	0.95; N=3	0.91; N=3	0.87; N=2

Module width: 100

Model score (-p*log10(pval)): 26.80
Wilcoxon p-value (pval): 9.49e-56
Penalty (p): 0.487
Average yes-set score: 12.21
Average no-set score: 10.11
AUC: 0.78
Separation point: 11.46
False-positive: 22.87%
False-negative: 32.44%
The AUC of the model achieves value significantly higher than expected for a random set of regulatory regions
Z-score = 4.78

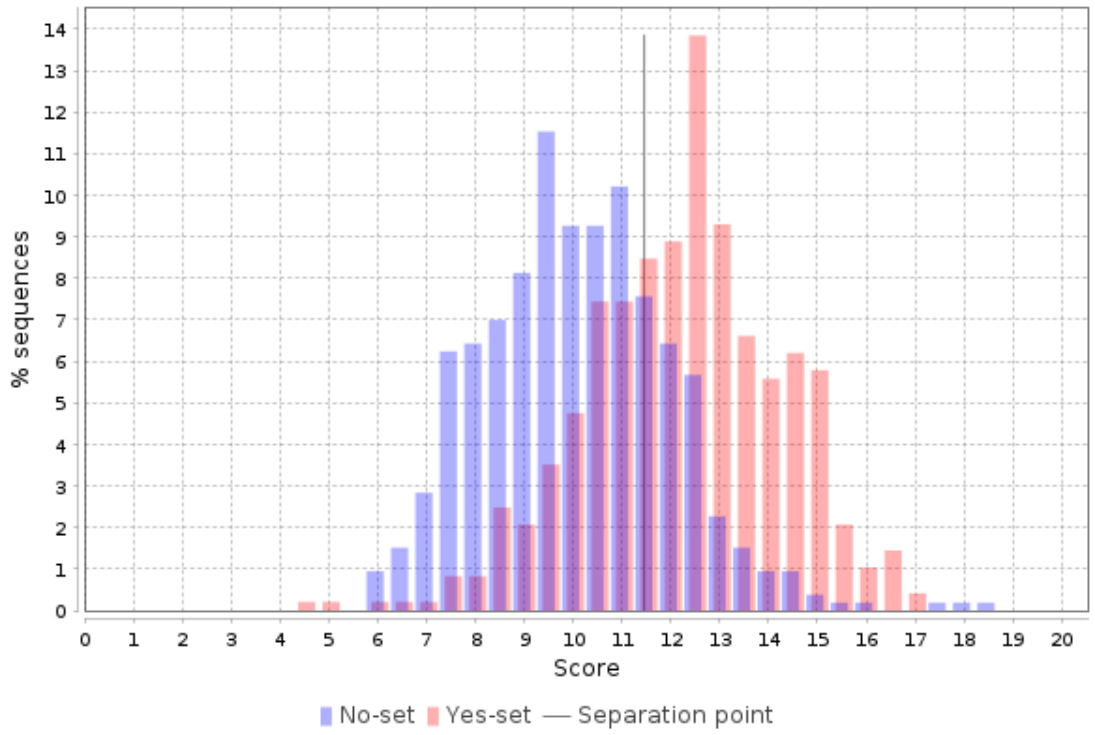


Table 4. List of top ten up-regulated genes in R0_Baseline vs. S0_Baseline with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000263528	IKBKE	inhibitor of nuclear factor kappa B kinase subunit epsilon	17.69	DBP(h), SIN3A(h), MafG(h), TEF-1(h), REST(h), JunB(h),c-Fos(h), ZNF462(h)...
ENSG00000125089	SH3TC1	SH3 domain and tetratricopeptide repeats 1	17.25	TEF-1(h), Runx3(h), foxo4(h), MafG(h), NF-kappaB-p65(h), ZNF462(h), RXRalpha(h)...
ENSG00000132952	USPL1	ubiquitin specific peptidase like 1	17.07	Runx3(h), SIN3A(h), TEF-1(h), DBP(h), ZNF462(h), REST(h), foxo4(h)...
ENSG00000100167	SEPTIN3	septin 3	16.8	REST(h), TEF-1(h), RXRalpha(h), ZNF462(h), MafG(h), JunB(h),c-Fos(h), DBP(h)...
ENSG00000196743	GM2A	ganglioside GM2 activator	16.72	NF-kappaB-p65(h), foxo4(h), ZNF462(h), RXRalpha(h), SIN3A(h), REST(h), TEF-1(h)...
ENSG00000189334	S100A14	S100 calcium binding protein A14	16.67	foxo4(h), DBP(h), SIN3A(h), MafG(h), TEF-1(h), Runx3(h), NF-kappaB-p65(h)...
ENSG00000237686	SCIRT	stem cell inhibitory RNA transcript	16.62	MafG(h), SIN3A(h), REST(h), Runx3(h), DBP(h), JunB(h),c-Fos(h), RXRalpha(h)...
ENSG00000109466	KLHL2	kelch like family member 2	16.61	NF-kappaB-p65(h), RXRalpha(h), ZNF462(h), JunB(h),c-Fos(h), SIN3A(h), foxo4(h), REST(h)...
ENSG00000130513	GDF15	growth differentiation factor 15	16.56	Runx3(h), TEF-1(h), SIN3A(h), DBP(h), MafG(h), REST(h), JunB(h),c-Fos(h)...
ENSG00000125508	SRMS	src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sites	16.46	foxo4(h), ZNF462(h), REST(h), JunB(h),c-Fos(h), NF-kappaB-p65(h), RXRalpha(h), SIN3A(h)...

Enhancer model potentially involved in regulation of target genes (down-regulated genes in R0_Baseline vs. S0_Baseline).

To build the most specific composite modules we choose top 300 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 R0_Baseline vs. S0_Baseline.

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$SOX9_02	V$TFAP4DLX3_01	V$DBP_Q6	V$AR_03
0.96; N=3	0.81; N=2	0.94; N=2	0.79; N=2

Module width: 162

Module 2:

V$GATA2_13	V$PARP_Q4	V$YY1_Q6_02	V$RXRA_21
0.87; N=1	1.00; N=3	0.92; N=2	0.68; N=3

Module width: 142

Model score (-p*log10(pval)): 17.56
Wilcoxon p-value (pval): 1.69e-33
Penalty (p): 0.536
Average yes-set score: 7.30
Average no-set score: 5.69
AUC: 0.74
Separation point: 6.67
False-positive: 30.43%
False-negative: 31.59%

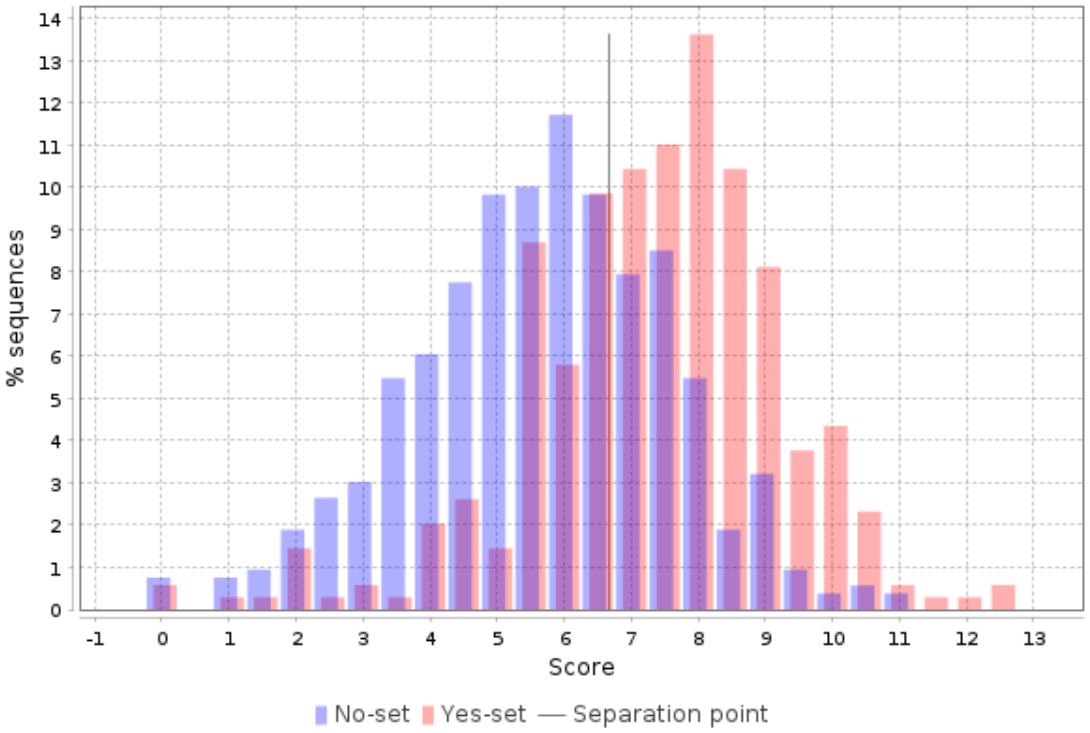


Table 5. List of top ten down-regulated genes in R0_Baseline vs. S0_Baseline with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000135048	CEMIP2	cell migration inducing hyaluronidase 2	12.21	PARP(h), YY1(h), RXRalpha(h), GATA-2(h), SOX-9(h), AP-4(h),DLX-3(h), DBP(h)...
ENSG00000237248	LINC00987	long intergenic non-protein coding RNA 987	11.99	RXRalpha(h), AR(h), SOX-9(h), PARP(h), DBP(h), GATA-2(h), YY1(h)...
ENSG00000232859	LYRM9	LYR motif containing 9	11.46	AR(h), DBP(h), RXRalpha(h), SOX-9(h), YY1(h), AP-4(h),DLX-3(h), GATA-2(h)...
ENSG00000231074	HCG18	HLA complex group 18	11.44	AP-4(h),DLX-3(h), RXRalpha(h), AR(h), PARP(h), DBP(h), SOX-9(h), YY1(h)...
ENSG00000026950	BTN3A1	butyrophilin subfamily 3 member A1	11.14	AR(h), RXRalpha(h), SOX-9(h), YY1(h), AP-4(h),DLX-3(h), DBP(h), PARP(h)...
ENSG00000108021	TASOR2	transcription activation suppressor family member 2	11.07	AP-4(h),DLX-3(h), SOX-9(h), DBP(h), YY1(h), AR(h), RXRalpha(h), PARP(h)
ENSG00000138796	HADH	hydroxyacyl-CoA dehydrogenase	11.05	RXRalpha(h), GATA-2(h), SOX-9(h), PARP(h), AP-4(h),DLX-3(h), YY1(h), DBP(h)...
ENSG00000172795	DCP2	decapping mRNA 2	10.99	PARP(h), AR(h), RXRalpha(h), YY1(h), AP-4(h),DLX-3(h), DBP(h), GATA-2(h)
ENSG00000173852	DPY19L1	dpy-19 like C-mannosyltransferase 1	10.99	GATA-2(h), SOX-9(h), PARP(h), DBP(h), YY1(h), AR(h), AP-4(h),DLX-3(h)
ENSG00000070087	PFN2	profilin 2	10.91	YY1(h), RXRalpha(h), PARP(h), GATA-2(h), AR(h), SOX-9(h), DBP(h)

On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 12 and 9 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 R0_Baseline vs. S0_Baseline). ***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
MO000079319	RELA	RELA proto-oncogene, NF-kB subunit	3.64	1.46
MO000018137	FOS	Fos proto-oncogene, AP-1 transcription factor subunit	3.42	1.67
MO000026238	RUNX3	RUNX family transcription factor 3	3.4	1.56
MO000000904	FOXO4	forkhead box O4	3.35	3.84
MO000019619	RXRA	retinoid X receptor alpha	3.01	1.65
MO000007830	JUNB	JunB proto-oncogene, AP-1 transcription factor subunit	2.99	1.51
MO000041817	REST	RE1 silencing transcription factor	2.7	6.04
MO000092587	ZNF462	zinc finger protein 462	2.68	1.33
MO000028669	DBP	D-box binding PAR bZIP transcription factor	2.67	1.21
MO000028667	MAFG	MAF bZIP transcription factor G	2.65	7.14

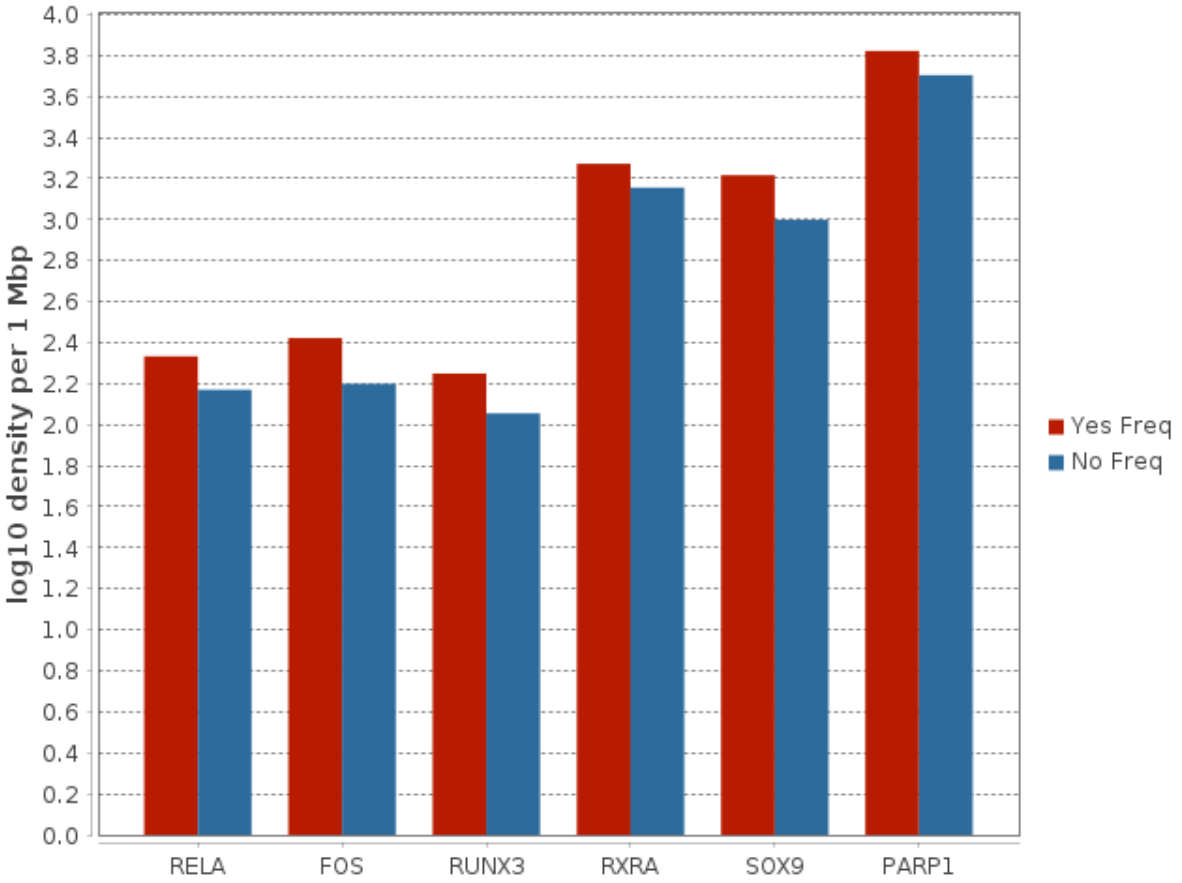
Table 7. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (down-regulated genes in R0_Baseline vs. S0_Baseline). ***Yes-No ratio*** *is the ratio between frequencies of the sites in Yes sequences versus No sequences. It describes the level 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
MO000019619	RXRA	retinoid X receptor alpha	5.18	1.31
MO000018993	SOX9	SRY-box transcription factor 9	5.16	1.65
MO000020396	PARP1	poly(ADP-ribose) polymerase 1	5.13	1.31
MO000021454	AR	androgen receptor	5.07	2.12
MO000078913	YY1	YY1 transcription factor	4.88	1.22
MO000032472	GATA2	GATA binding protein 2	4.1	1.46
MO000028669	DBP	D-box binding PAR bZIP transcription factor	3.65	1.63

MO000024660	TFAP4	transcription factor AP-4	3.38	1.51
MO000027901	DLX3	distal-less homeobox 3	0	2.63

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: RELA, FOS, RUNX3, RXRA, SOX9 and PARP1.



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 biological 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 R0_Baseline vs. S0_Baseline. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics and epigenomics data.*

See full table →

ID	Master molecule name	Gene symbol	Gene description	logFC	Total rank
MO000139403	MIC-1(h)	GDF15	growth differentiation factor 15	4.65	247
MO000007258	IGF-2(h)	IGF2	insulin like growth factor 2	11.12	262
MO000121013	IGF-2-isoform2(h)	IGF2	insulin like growth factor 2	11.12	263
MO001075092	Src(h){myrG2}{pY419}:Runx3(h)	RUNX3, SRC	RUNX family transcription factor 3, SRC proto-oncogene, non-receptor tyrosine kinase	5.65	307
MO000032653	MKP3(h)	DUSP6	dual specificity phosphatase 6	1.63	341
MO000019766	FGFR-3(h)	FGFR3	fibroblast growth factor receptor 3	3.36	359
MO001054075	Src(h){myrG2}{pY419}	SRC	SRC proto-oncogene, non-receptor tyrosine kinase	1.57	373
MO000021478	Src(h)	SRC	SRC proto-oncogene, non-receptor tyrosine kinase	1.57	391
MO000038800	VE-cadherin(h)	CDH5	cadherin 5	3.99	414
MO001075093	Src{myrG2}{pY419}:Runx3	RUNX3, SRC	RUNX family transcription factor 3, SRC proto-oncogene, non-receptor tyrosine kinase	5.65	419

*Table 9. Master regulators that may govern the regulation of **down-regulated** genes in R0_Baseline vs. S0_Baseline. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics and epigenomics data.*

See full table →

ID	Master molecule name	Gene symbol	Gene description	logFC	Total rank
MO000101469	LRRK2(h)	LRRK2	leucine rich repeat kinase 2	-1.33	439
MO000056127	Sp100(h)	SP100	SP100 nuclear antigen	-1.73	508
MO000099067	XPC(h)	XPC	XPC complex subunit, DNA damage recognition and repair factor	-0.99	527
MO000020219	Caspase-8(h)	CASP8	caspase 8	-0.84	528
MO001091316	HB-EGF(h):EGFR(h):HGFIN(h){pY525}:LINC01139:PTK6(h){pY351}:LRRK2(h)	EGFR, GPNMB, HBEGF, LRRK2, PTK6	epidermal growth factor receptor, glycoprotein nmb, heparin binding EGF like growth factor, leucine ...	-1.33	541
MO001096369	hydroxyPro-HIF-alpha:Roc1:EloC:EloB:VHL:ccdc22:COMMDs:DCUN1D1,2,4,5:Cul-2{neddK689}	CCDC22, COMMD1, COMMD10, COMMD2, COMMD3, COMMD4, COMMD5, COMMD6, COMMD7, COMMD8, COMMD9, CUL2, DCUN1...	COMM domain containing 10, COMM domain containing 2, COMM domain containing 3, COMM domain containin...	-0.83	578
MO000021036	Caspase-8(h)	CASP8	caspase 8	-0.84	589
MO000020563	p27Kip1(h)	CDKN1B	cyclin dependent kinase inhibitor 1B	-1.16	603
MO000033978	JNK3(h)	MAPK10	mitogen-activated protein kinase 10	-0.84	632
MO000201460	SIAH2(h)	SIAH2	siah E3 ubiquitin protein ligase 2	-1.55	644

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figures 8 and 9. 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.

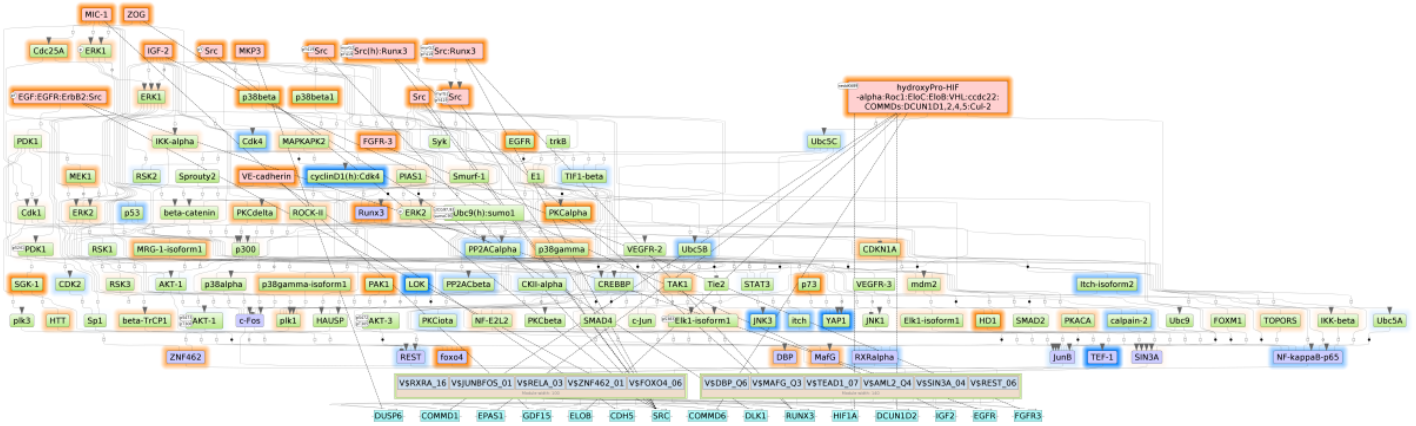


Figure 8. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in R0_Baseline vs. S0_Baseline. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the 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 →

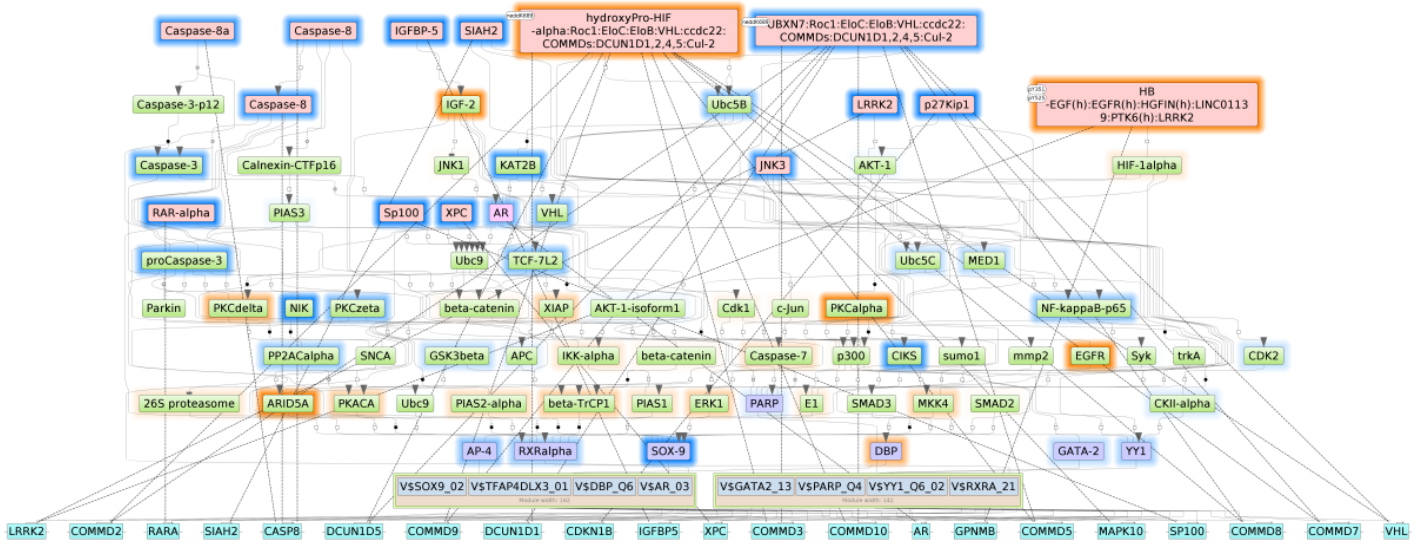


Figure 9. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in R0_Baseline vs. S0_Baseline. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the 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. 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 2025.1 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac).
The master regulator search uses the TRANSPATH® database (BIOBASE), release 2025.1 (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 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 by the model, are selected.

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

Here:

- j – a TF (Transcription Factor),
- i – a MR (Master Regulator),
- S_j – Regulatory score of the TF,
- $D_{i,j}$– 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,
- $C_{i,j}$ – 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), $d_k >= 3$ for each semantic reaction (like A → 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].

5. 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

Thank you for using the Pathway Omics Suite!

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 R0_Baseline vs. S0_Baseline).
4. Supplementary table 4 - Detailed report. Composite modules and master regulators (down-regulated genes in R0_Baseline vs. S0_Baseline).

Disclaimer

The scientific analysis underlying the Pathway Omics Suite 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 Pathway Omics Suite. Moreover, any warranty concerning the completeness, up-to-dateness, correctness and usability of Pathway Omics Suite information and results produced by it, shall be excluded.

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

Note that the text given in the report is not unique and can be fully or partially repeated in other Pathway Omics Suite 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 Pathway Omics Suite pipeline parameters used for the current run.