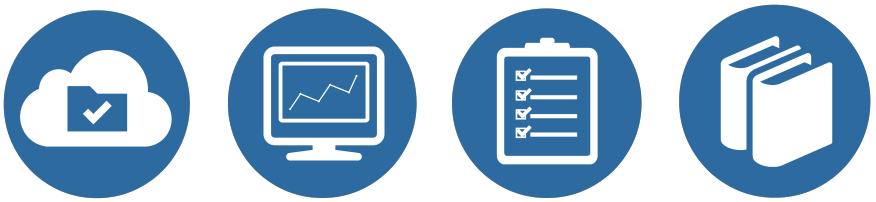
CASP8 and LRRK2 are promising druggable targets for treating Parkinson Disease that control activity of AR, FOXO1 and RELA transcription factors on promoters of differentially expressed genes

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 02/07/2020 ; Run on 24/11/2024 ; Report generated on 24/11/2024

Genome Enhancer release 3.5 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2024.2)



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 *Parkinson 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: AR, FOXO1, POU2F1, RELA, THRA and MAFG. The subsequent network analysis suggested

- Caspase-8
- LRRK2
- txnip
- Caspase-8

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. Having checked the actual druggability potential of the full list of identified targets, both, via information available in medical literature and via cheminformatics analysis of drug compounds, we have identified the following drugs as the most promising treatment candidates for the studied pathology: Sirolimus, 1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea, Lipoic Acid and Edotecarin.

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® [6] is employed together with the TF binding site identification algorithms Match [7] and CMA [8]. The second step involves the signal transduction database TRANSPATH® [9] 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 [5]. 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
GSE145804_DESeq2_final	Transcriptomics

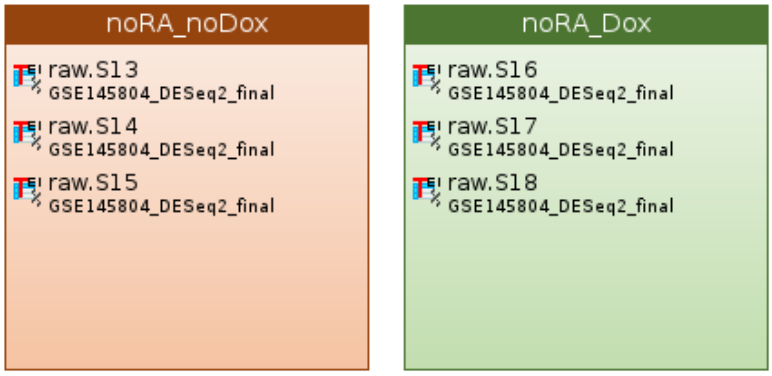


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

3. Results

We have compared the following conditions: noRA_Dox *versus* noRA_noDox.

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: "noRA_Dox" with "noRA_noDox". 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 3391 upregulated genes (LogFC>0.1) out of which 333 genes were found as significantly upregulated (p-value<0.1) and 4391 downregulated genes (LogFC<-0.1) out of which 344 genes were significantly downregulated (p-value<0.1). 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 noRA_Dox vs. noRA_noDox.*
See full table →

ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000145335	SNCA	synuclein alpha	4.38	10.37	5.54E-90	6.59E-86
ENSG00000206651	Y_RNA	Y RNA	1.27	2.08	8.17E-2	0.96
ENSG00000108231	LGI1	leucine rich glioma inactivated 1	1.04	3.42	4.04E-3	0.96
ENSG00000205403	CFI	complement factor I	1.03	3.11	1.25E-4	0.17
ENSG00000275140	SEC22B3P	SEC22 homolog B3, pseudogene	1.03	2.97	8.86E-3	0.96
ENSG00000189057	FAM111B	FAM111 trypsin like peptidase B	0.95	3.47	1.18E-3	0.56
ENSG00000231043		IK cytokine, down-regulator of HLA II (IK) pseudogene	0.95	2.41	2.56E-2	0.96
ENSG00000108691	CCL2	C-C motif chemokine ligand 2	0.94	4.19	5.39E-6	1.07E-2
ENSG00000086300	SNX10	sorting nexin 10	0.94	3.01	4.34E-4	0.36
ENSG00000265972	TXNIP	thioredoxin interacting protein	0.87	8.05	7.68E-10	3.04E-6

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

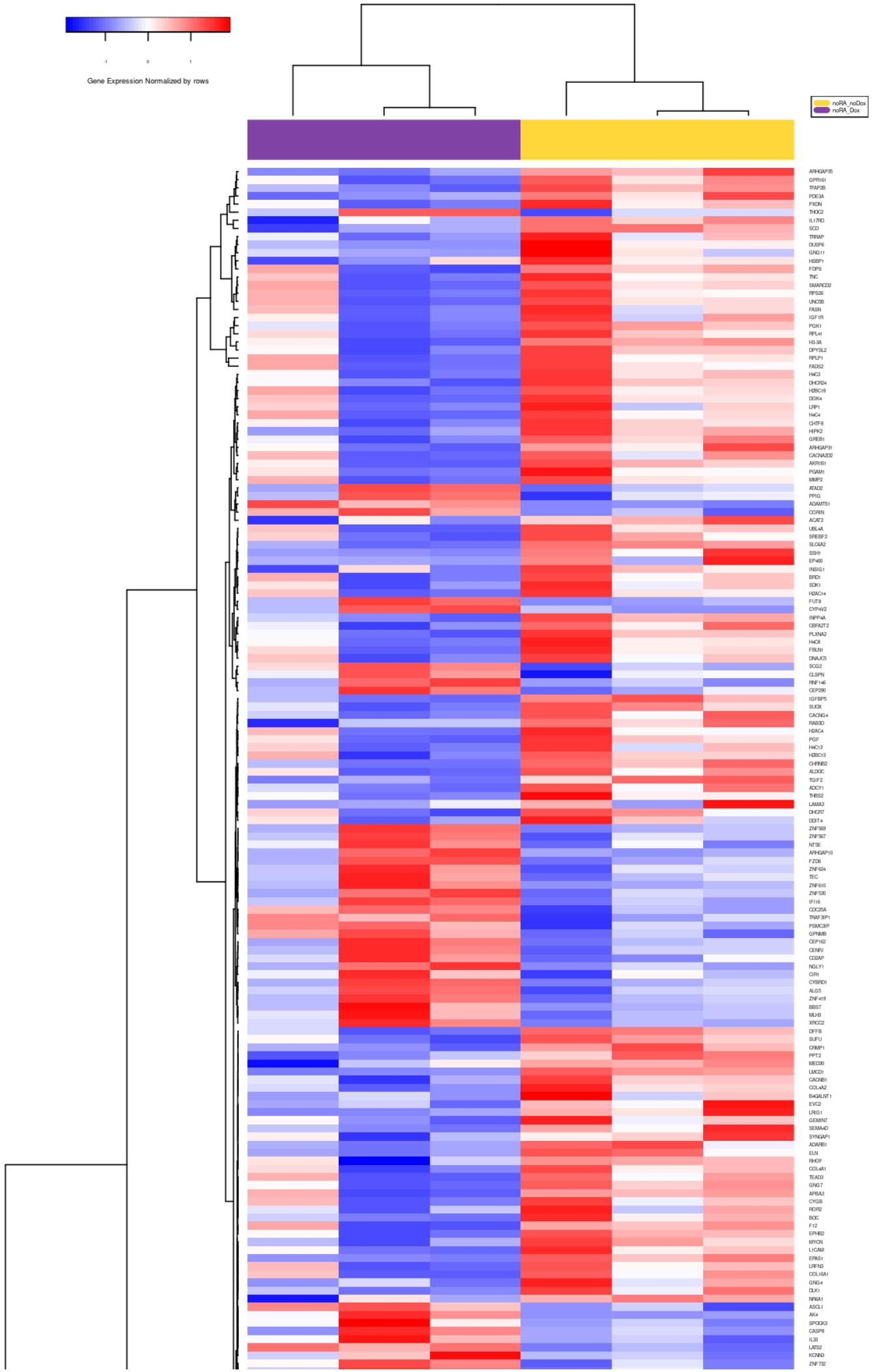
ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000186081	KRT5	keratin 5	-10.37	2.59	4.32E-6	1.03E-2
ENSG00000147256	ARHGAP36	Rho GTPase activating protein 36	-1.67	5.58	3.6E-16	2.14E-12
ENSG00000157601	MX1	MX dynamin like GTPase 1	-1.64	1.75	1.44E-2	0.96
ENSG00000123454	DBH	dopamine beta-hydroxylase	-1.48	3.54	4.57E-5	6.78E-2
ENSG00000254656	RTL1	retrotransposon Gag like 1	-1.36	3.09	2.62E-4	0.31
ENSG00000116016	EPAS1	endothelial PAS domain protein 1	-1.29	2.34	2.42E-6	7.19E-3
ENSG00000185559	DLK1	delta like non-canonical Notch ligand 1	-1.15	2.69	8.96E-4	0.48
ENSG00000165912	PACSIN3	protein kinase C and casein kinase substrate in neurons 3	-1.05	3.21	1.55E-2	0.96
ENSG00000148357	HMCN2	hemicentin 2	-1.03	2.42	1.01E-2	0.96
ENSG00000049540	ELN	elastin	-0.95	2.92	5.79E-4	0.36

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 3-8 show the most significant categories.

Heatmap of differentially expressed genes in noRA_Dox vs. noRA_noDox

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


Gene Expression Normalized by rows
noRA_noDox
noRA_Dox
ARHGAP35
GPR161
TRAP3B
PDE3A
PXDN
THOC2
IL17RD
SCD
TRRAP
DUSP6
GNG11
HSBP1
FDPS
TNC
SMARCD2
RPS26
UNC5B
FASN
IGF1R
PGK1
RPL41
H3-3A
DPYSL2
RPLP1
FADS2
H4C3
DHCR24
H2BC18
DDR4
LRP1
H4C4
CHTF8
HIPK2
GREB1
ARHGAP31
CACNA2D2
AKR1B1
PGAM1
MMP2
ATAD2
PPIG
ADAMTS1
CORIN
ACAT2
UBL4A
SREBF2
SLC6A2
SSH1
EP400
INSIG1
BRD1
SDK1
H2AC14
FUT8
CYP4V2
INPP4A
CBFA2T2
PLXNA2
H4C8
FBLN1
DNAJC5
SCG2
CLSPN
RNF146
CEP290
IGFBP5
SUCK
CACNG4
RAB3D
H2AC4
PGF
H4C13
H2BC13
CHRNB2
ALDOC
TGIF2
ADCY1
THBS2
LAMA3
DHCR7
DDIT4
ZNF569
ZNF567
NT5E
ARHGAP10
FZD6
ZNF624
TEC
ZNF610
ZNF530
IFI16
CDC25A
TRAF3IP1
PSMC3IP
GPNMB
CEP162
CENPJ
CD2AP
NGLY1
CIR1
CYBRD1
ALG5
ZNF419
BBS7
MLH3
XRCC2
DFFB
SUFU
CRMP1
PPT2
MED30
LMCD1
CACNB1
COL4A2
B4GALNT1
EVC2
LRIG1
GEMIN7
SEMA4D
SYNGAP1
ADARB1
ELN
RHOF
COL4A1
TEAD3
GNG7
AP3A3
CYGB
ROR2
BOC
F12
EPHB2
MYCN
L1CAM
EPAS1
LRFN3
COL16A1
GNG4
DLK1
NR6A1
ASCL1
AK4
SPOCK3
CASP8
IL33
LATS2
KCNN3
ZNF732

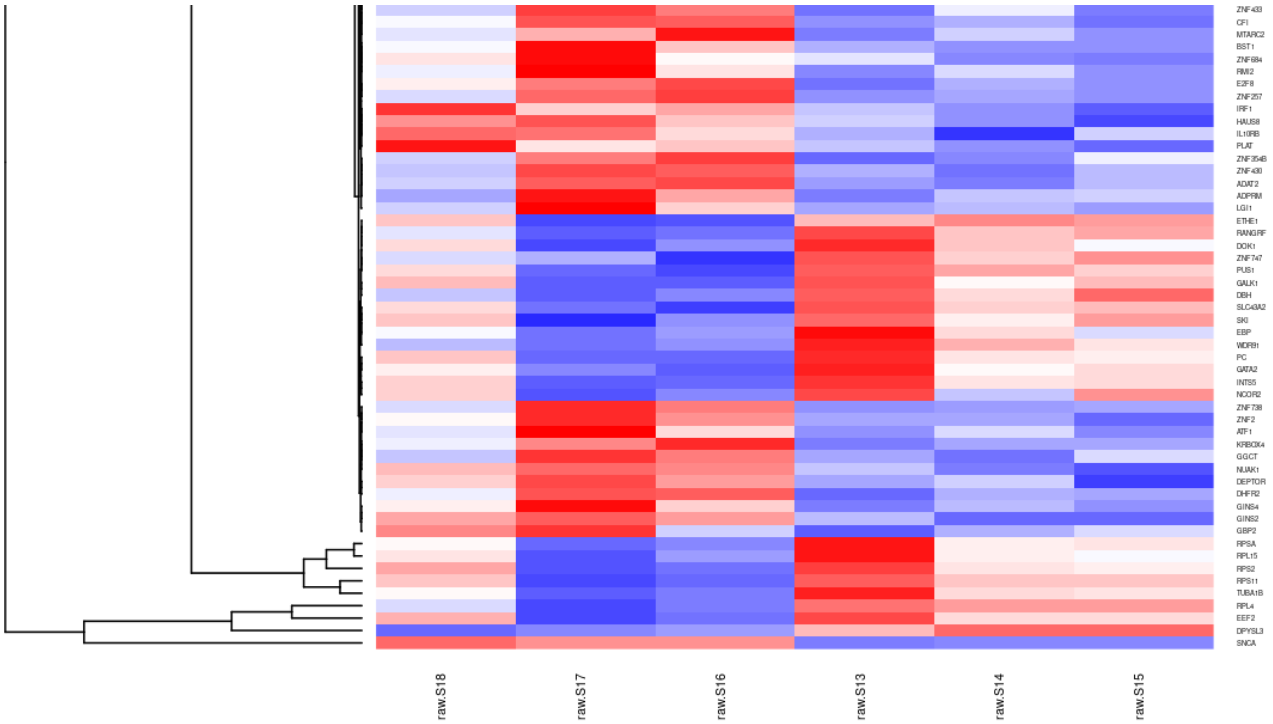


Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.

See full diagram →

Up-regulated genes in noRA_Dox vs. noRA_noDox:

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

GO (biological process)

Figure 3. Enriched GO (biological process) of up-regulated genes in noRA_Dox vs. noRA_noDox.

Full classification →

TRANSPATH® Pathways (2024.2)

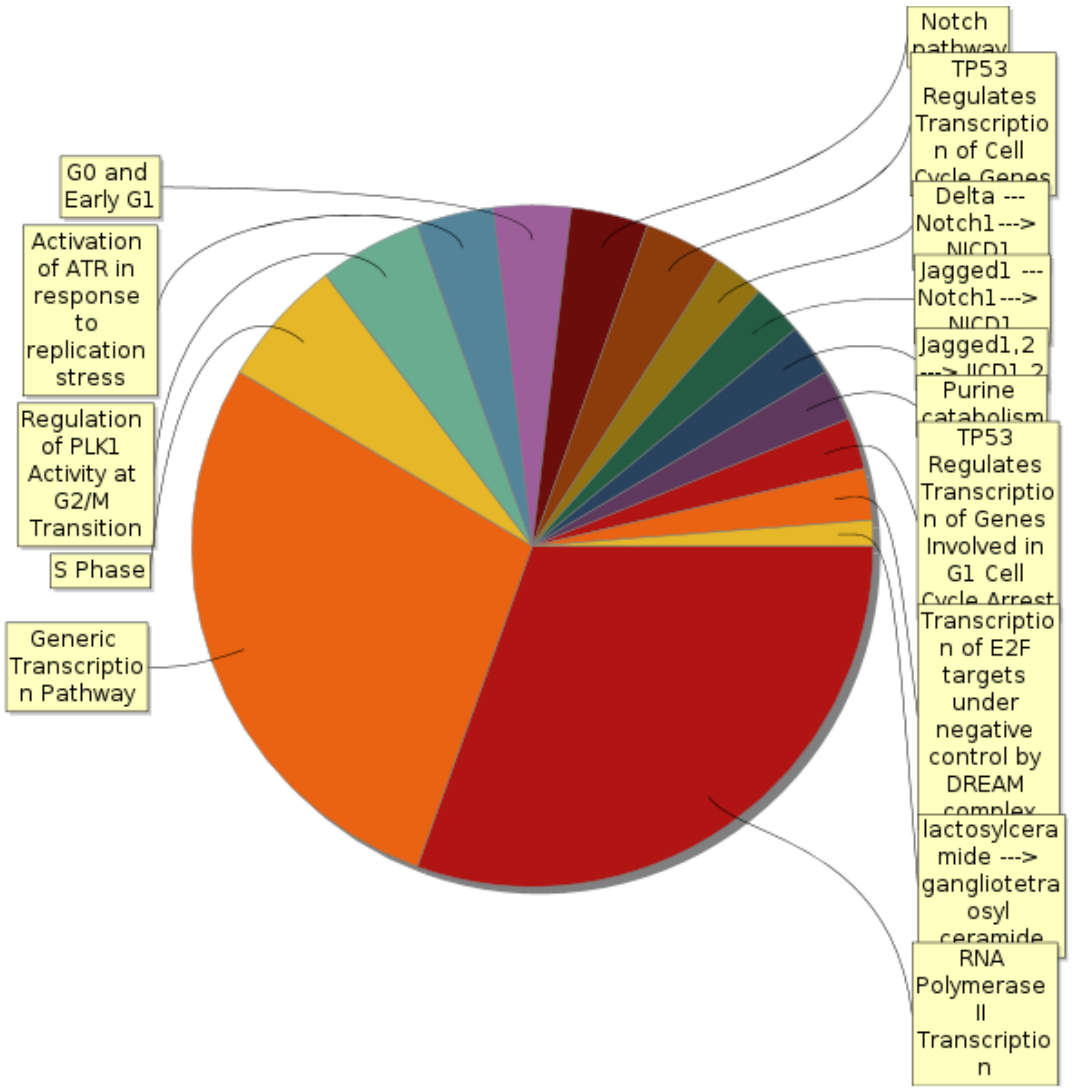


Figure 4. Enriched TRANSPATH® Pathways (2024.2) of up-regulated genes in noRA_Dox vs. noRA_noDox.

Full classification →

HumanPSD(TM) disease (2024.2)

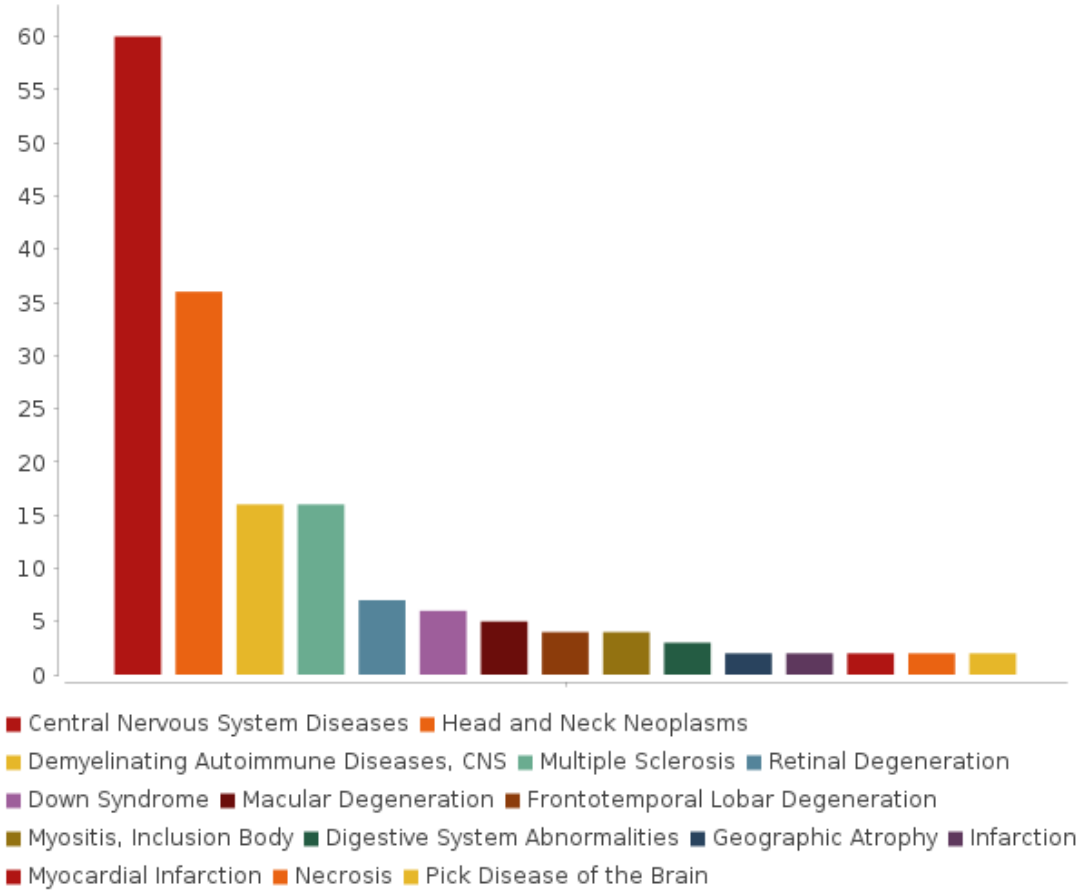


Figure 5. Enriched HumanPSD(TM) disease (2024.2) of up-regulated genes in noRA_Dox vs. noRA_noDox. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

Full classification →

Down-regulated genes in noRA_Dox vs. noRA_noDox:

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

GO (biological process)

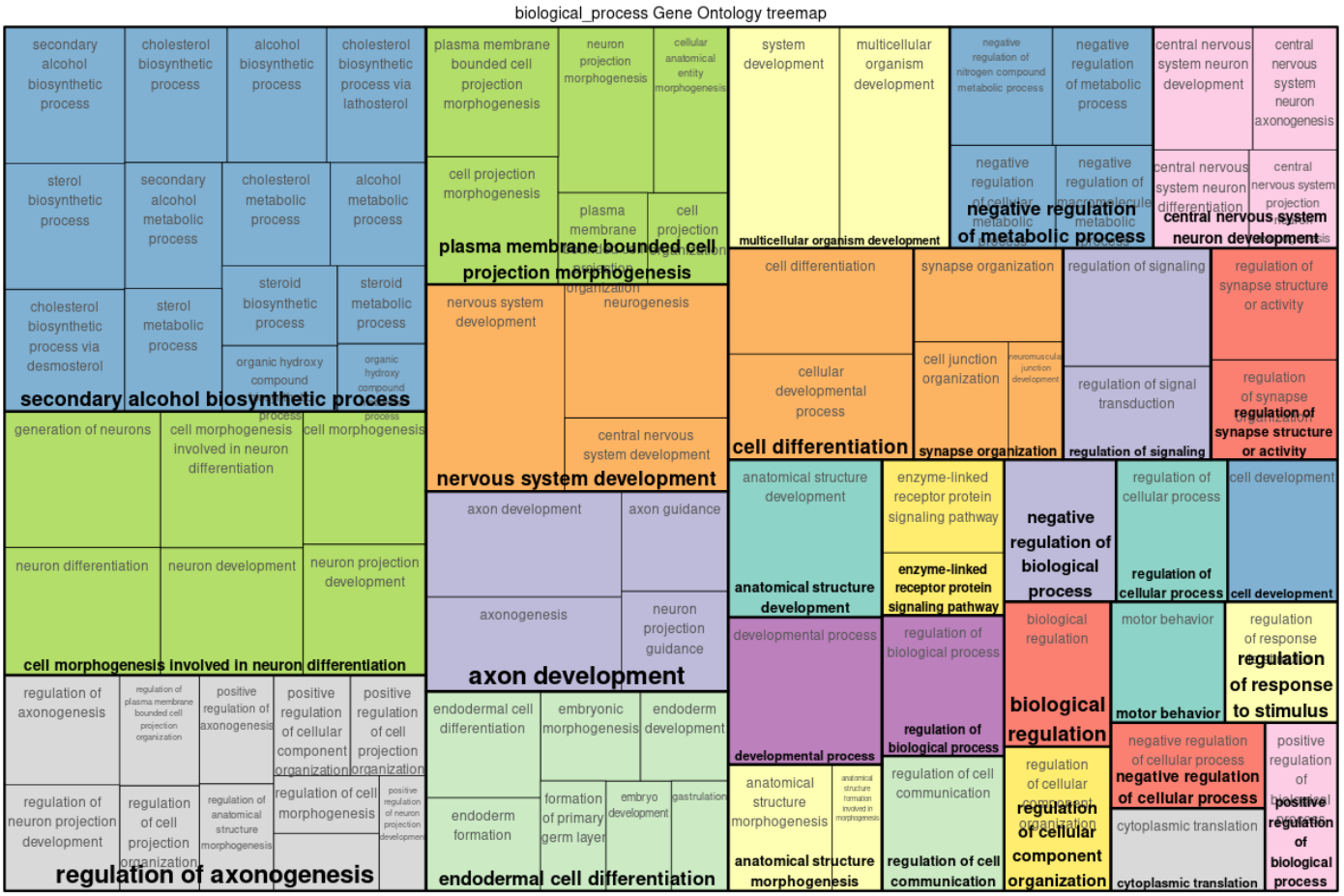


Figure 6. Enriched GO (biological process) of down-regulated genes in noRA_Dox vs. noRA_noDox.

Full classification →

TRANSPATH® Pathways (2024.2)

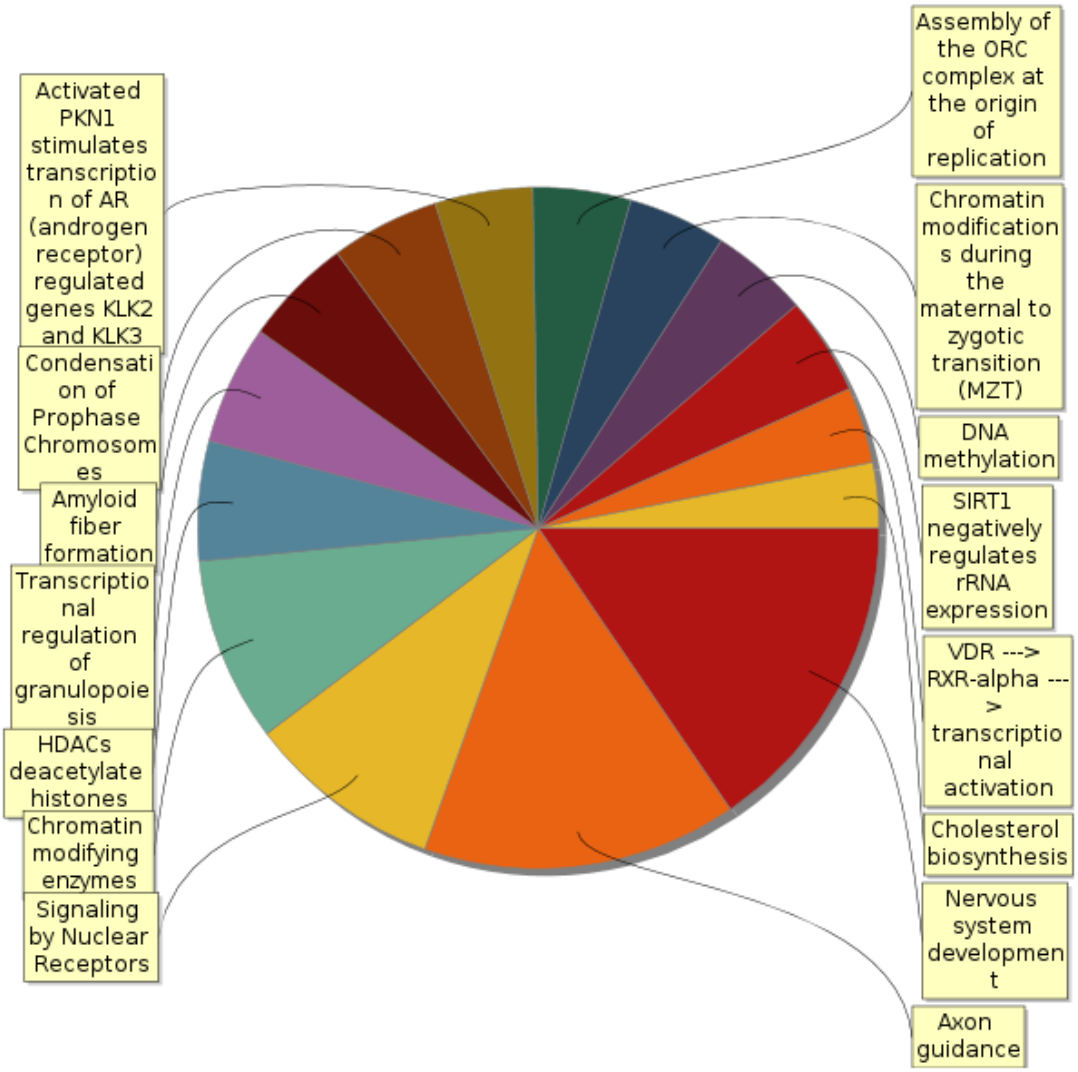


Figure 7. Enriched TRANSPATH® Pathways (2024.2) of down-regulated genes in noRA_Dox vs. noRA_noDox.
Full classification →

HumanPSD(TM) disease (2024.2)

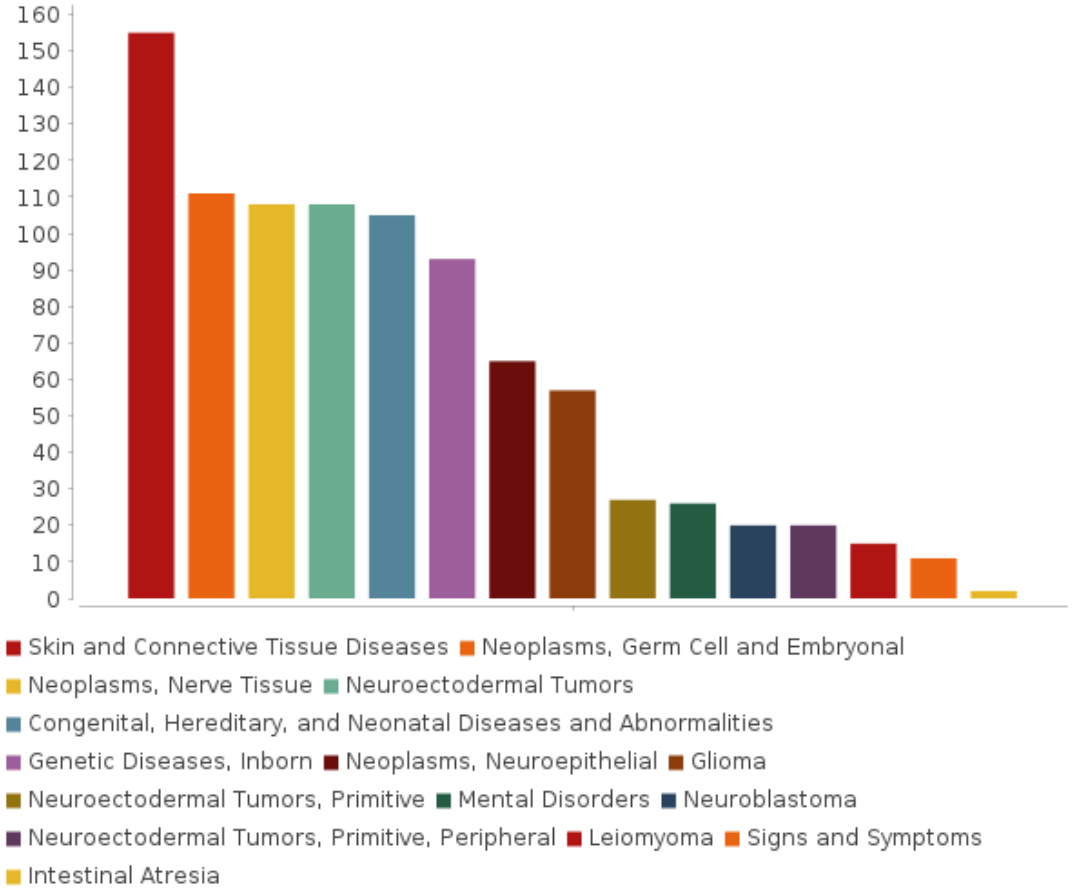
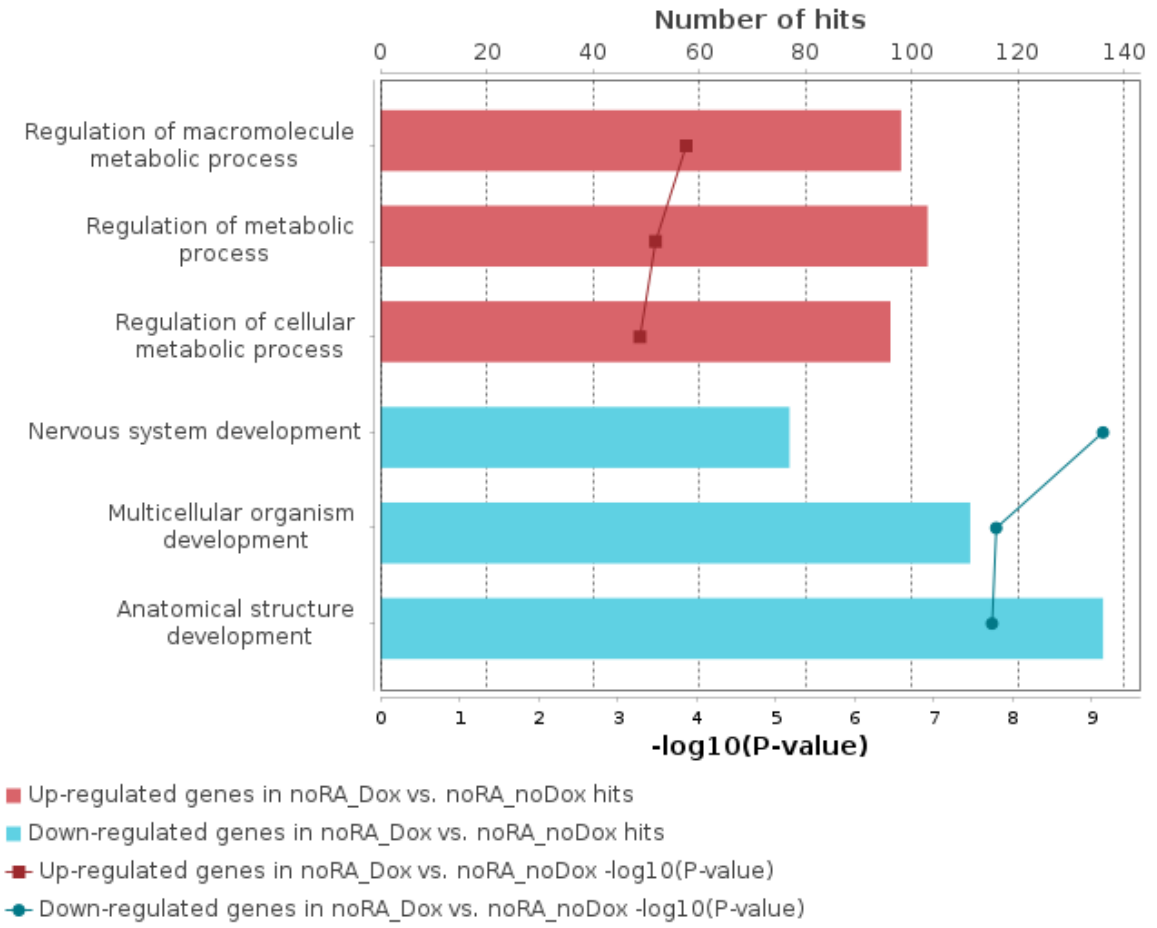


Figure 8. Enriched HumanPSD(TM) disease (2024.2) of down-regulated genes in noRA_Dox vs. noRA_noDox. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

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 [8]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [9].

We applied the Composite Module Analyst (CMA) [8] 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 noRA_Dox vs. noRA_noDox).

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 noRA_Dox vs. noRA_noDox.

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$ZEB1_05	V$ERFFOXO1_01	V$AR_04	V$FOXO1HOXA10_01	V$POU1F1_10
0.99; N=3	0.73; N=3	0.87; N=3	0.84; N=2	0.81; N=2

Module width: 100

Module 2:

V$POU2F1FOXO6_01	V$RUSH1A_02	V$GATA6_12	V$PBX1_04	V$POU2F1_Q4	V$BCL11A_04
0.67; N=2	0.98; N=2	0.72; N=3	0.84; N=1	0.97; N=2	0.83; N=2

Module width: 65

Model score (-p*log10(pval)): 13.79
Wilcoxon p-value (pval): 4.78e-29
Penalty (p): 0.487
Average yes-set score: 6.79
Average no-set score: 5.29
AUC: 0.74
Separation point: 6.40
False-positive: 30.26%
False-negative: 31.44%

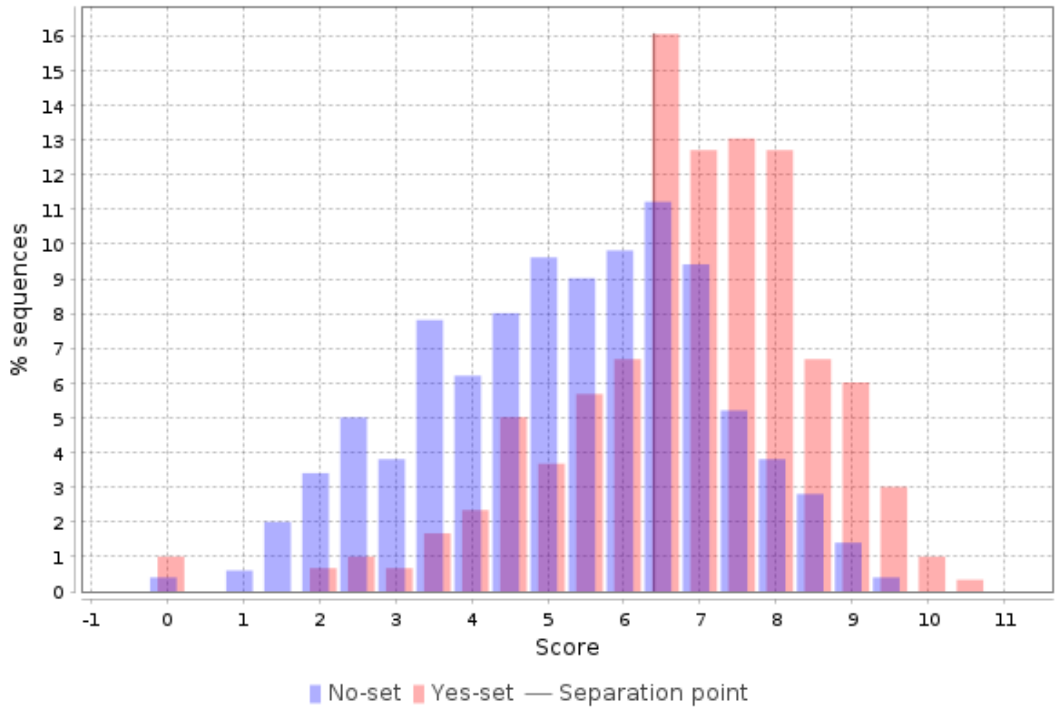


Table 4. List of top ten up-regulated genes in noRA_Dox vs. noRA_noDox with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000157214	STEAP2	STEAP2 metalloreductase	11.79	POU1F1(h), ERF(h),FOXO1(h), FOXO1(h),Hox-A10(h), BCL-11A(h), FOXO6(h),POU2F1(h), HLTF(h), PBX-1(h)...
ENSG00000204271	SPIN3	spindlin family member 3	10.89	POU1F1(h), ERF(h),FOXO1(h), FOXO1(h),Hox-A10(h), POU2F1(h), GATA-6(h), HLTF(h), FOXO6(h),POU2F1(h)...
ENSG00000133121	STARD13	StAR related lipid transfer domain containing 13	10.8	ERF(h),FOXO1(h), FOXO1(h),Hox-A10(h), POU1F1(h), PBX-1(h), GATA-6(h), POU2F1(h), FOXO6(h),POU2F1(h)...
ENSG00000249353	NPM1P27	nucleophosmin 1 pseudogene 27	10.76	POU2F1(h), HLTF(h), PBX-1(h), GATA-6(h), FOXO6(h),POU2F1(h), FOXO1(h),Hox-A10(h), ERF(h),FOXO1(h)...
ENSG00000256683	ZNF350	zinc finger protein 350	10.64	ERF(h),FOXO1(h), FOXO1(h),Hox-A10(h), GATA-6(h), ZEB1(h), PBX-1(h), FOXO6(h),POU2F1(h), HLTF(h)...
ENSG00000255291	HMGB1P40	high mobility group box 1 pseudogene 40	10.57	GATA-6(h), BCL-11A(h), PBX-1(h), FOXO6(h),POU2F1(h), HLTF(h), ERF(h),FOXO1(h), FOXO1(h),Hox-A10(h)...
ENSG00000075223	SEMA3C	semaphorin 3C	10.55	ERF(h),FOXO1(h), FOXO1(h),Hox-A10(h), GATA-6(h), FOXO6(h),POU2F1(h), PBX-1(h), POU2F1(h), HLTF(h)
ENSG00000133131	MORC4	MORC family CW-type zinc finger 4	10.51	FOXO6(h),POU2F1(h), HLTF(h), GATA-6(h), PBX-1(h), ERF(h),FOXO1(h), FOXO1(h),Hox-A10(h), POU1F1(h)
ENSG00000143753	DEGS1	delta 4-desaturase, sphingolipid 1	10.51	ZEB1(h), GATA-6(h), PBX-1(h), FOXO6(h),POU2F1(h), HLTF(h), POU2F1(h), ERF(h),FOXO1(h)...
ENSG00000249264	EEF1A1P9	eukaryotic translation elongation factor 1 alpha 1 pseudogene 9	10.49	GATA-6(h), HLTF(h), FOXO6(h),POU2F1(h), BCL-11A(h), PBX-1(h), POU2F1(h), ERF(h),FOXO1(h)...

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

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 noRA_Dox vs. noRA_noDox.

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$RELA_03	V$MAFG_Q3	V$HIC1_03	V$CMAF_03	V$GCM1_08
0.94; N=3	0.74; N=3	0.94; N=3	0.84; N=1	0.88; N=2

Module width: 109

Module 2:

V$MZF1_09	V$GLI1_Q3	V$LRH1_05	V$ZEB1_05	V$T3RALPHA_02
0.86; N=1	0.90; N=2	0.72; N=2	0.99; N=2	0.80; N=3

Module width: 129

Model score (-p*log10(pval)): 17.52
Wilcoxon p-value (pval): 1.09e-35
Penalty (p): 0.501
Average yes-set score: 6.52
Average no-set score: 4.87
AUC: 0.76
Separation point: 5.40
False-positive: 36.27%
False-negative: 19.19%

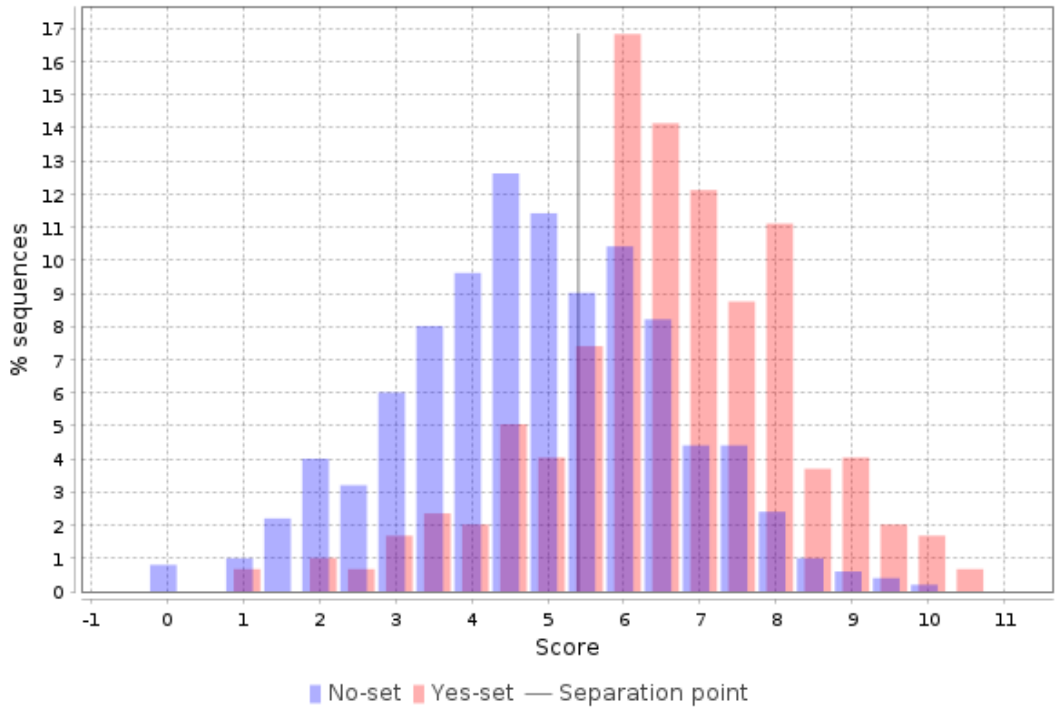


Table 5. List of top ten down-regulated genes in noRA_Dox vs. noRA_noDox with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000185347	TEDC1	tubulin epsilon and delta complex 1	11.45	T3R-alpha(h), MafG(h), NF-kappaB-p65(h), GCMa(h), c-Maf(h), LRH-1(h), GLI1(h)...
ENSG00000099940	SNAP29	synaptosome associated protein 29	11.4	MafG(h), LRH-1(h), ZEB1(h), NF-kappaB-p65(h), MZF1(h), GLI1(h), GCMa(h)...
ENSG00000197958	RPL12	ribosomal protein L12	11.27	NF-kappaB-p65(h), MafG(h), T3R-alpha(h), HIC1(h), GCMa(h), GLI1(h), c-Maf(h)...
ENSG00000214140	PRCD	photoreceptor disc component	11.22	ZEB1(h), GCMa(h), HIC1(h), NF-kappaB-p65(h), LRH-1(h), MZF1(h), GLI1(h)...
ENSG00000186716	BCR	BCR activator of RhoGEF and GTPase	11.14	NF-kappaB-p65(h), GLI1(h), LRH-1(h), MafG(h), T3R-alpha(h), c-Maf(h), ZEB1(h)...
ENSG00000076662	ICAM3	intercellular adhesion molecule 3	11.04	NF-kappaB-p65(h), GCMa(h), MafG(h), c-Maf(h), T3R-alpha(h), GLI1(h), MZF1(h)
ENSG00000006534	ALDH3B1	aldehyde dehydrogenase 3 family member B1	10.99	T3R-alpha(h), HIC1(h), MZF1(h), GLI1(h), ZEB1(h), LRH-1(h), NF-kappaB-p65(h)...
ENSG00000130702	LAMA5	laminin subunit alpha 5	10.91	GLI1(h), MZF1(h), HIC1(h), GCMa(h), MafG(h), NF-kappaB-p65(h), c-Maf(h)
ENSG00000030582	GRN	granulin precursor	10.85	LRH-1(h), T3R-alpha(h), GLI1(h), MafG(h), NF-kappaB-p65(h), GCMa(h), c-Maf(h)
ENSG00000177030	DEAF1	DEAF1 transcription factor	10.84	MafG(h), GLI1(h), NF-kappaB-p65(h), GCMa(h), LRH-1(h), MZF1(h), HIC1(h)...

On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 12 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 noRA_Dox vs. noRA_noDox). ***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 →

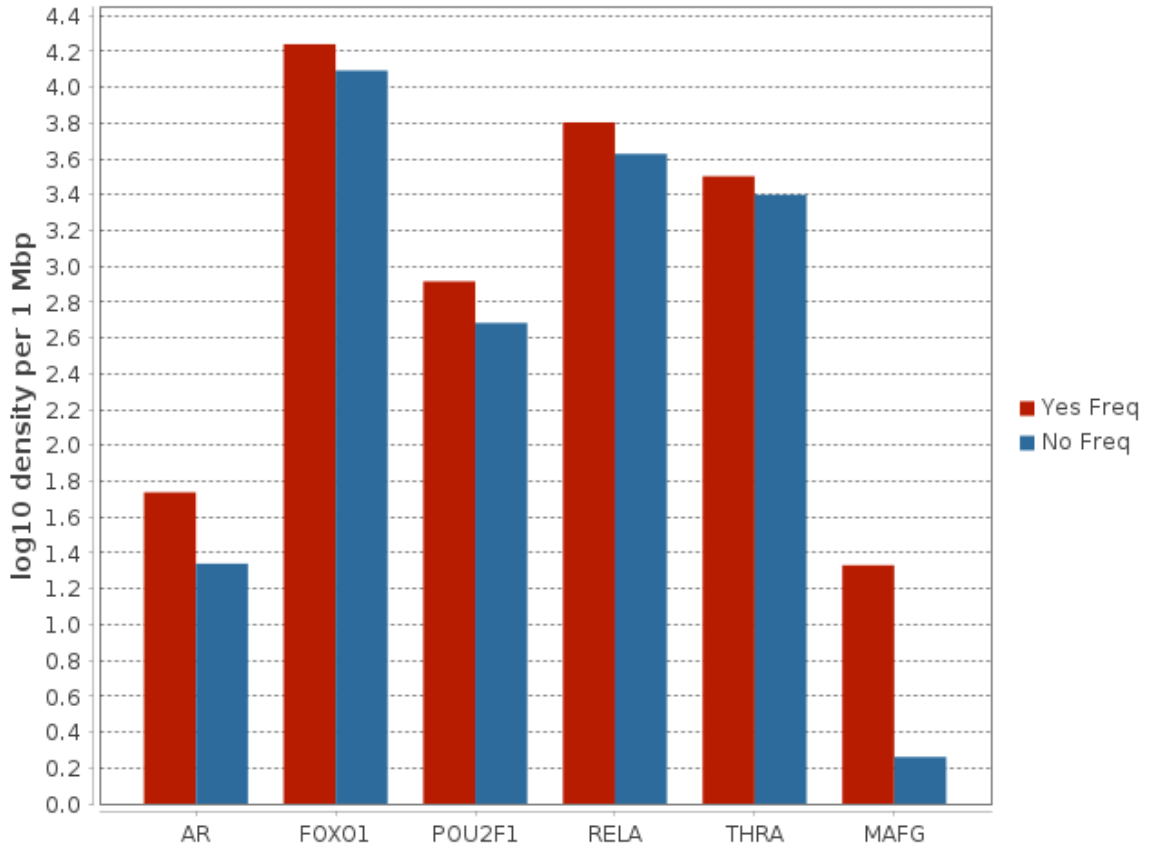
ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000021454	AR	androgen receptor	2.93	2.5
MO000034454	FOXO1	forkhead box O1	2.63	1.4
MO000025003	POU2F1	POU class 2 homeobox 1	2.58	1.71
MO000118284	HLTF	helicase like transcription factor	2.28	1.27
MO000139677	ZEB1	zinc finger E-box binding homeobox 1	2.25	4.59
MO000028675	ERF	ETS2 repressor factor	2.13	1.28
MO000177867	FOXO6	forkhead box O6	1.91	2
MO000089495	HOXA10	homeobox A10	1.87	1.52
MO000084573	POU1F1	POU class 1 homeobox 1	1.79	1.46
MO000026665	GATA6	GATA binding protein 6	0.3	1.35

Table 7. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (down-regulated genes in noRA_Dox vs. noRA_noDox). ***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	5.48	1.5
MO000025208	THRA	thyroid hormone receptor alpha	4.13	1.27
MO000028667	MAFG	MAF bZIP transcription factor G	4.05	11.75
MO000026306	GCM1	glial cells missing transcription factor 1	3.87	5.04
MO000139677	ZEB1	zinc finger E-box binding homeobox 1	3.8	2.27
MO000019117	GLI1	GLI family zinc finger 1	3.65	1.57
MO000026742	NR5A2	nuclear receptor subfamily 5 group A member 2	3.42	1.56
MO000037926	MAF	MAF bZIP transcription factor	3.31	2.24
MO000117395	HIC1	HIC ZBTB transcriptional repressor 1	3.26	1.68
MO000084468	MZF1	myeloid zinc finger 1	0	1.49

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: AR, FOXO1, POU2F1, RELA, THRA and MAFG.



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 noRA_Dox vs. noRA_noDox. **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
MO000020219	Caspase-8(h)	CASP8	caspase 8	0.55	49
MO000021036	Caspase-8(h)	CASP8	caspase 8	0.55	87
MO000320329	Optineurin(h){pS177}	OPTN	optineurin	0.53	125
MO000045378	txnip(h)	TXNIP	thioredoxin interacting protein	0.87	146
MO000043060	(Caspase-8(h))2	CASP8	caspase 8	0.55	170
MO000043221	Caspase-8a(h)	CASP8	caspase 8	0.55	181
MO000328441	txnip-isoform2(h)	TXNIP	thioredoxin interacting protein	0.87	242
MO000178122	txnip-isoform1(h)	TXNIP	thioredoxin interacting protein	0.87	243
MO001081054	Fam:SNCA{ubK}	SNCA, USP9X	synuclein alpha, ubiquitin specific peptidase 9 X-linked	4.38	249
MO000044264	SNCA(h)	SNCA	synuclein alpha	4.38	250

*Table 9. Master regulators that may govern the regulation of **down-regulated** genes in noRA_Dox vs. noRA_noDox. **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
MO000124674	EPHB2(h)	EPHB2	EPH receptor B2	-0.77	197
MO000124672	EPHB2-isoform1(h)	EPHB2	EPH receptor B2	-0.77	224
MO000124673	EPHB2-isoform2(h)	EPHB2	EPH receptor B2	-0.77	224
MO000255149	EPHB2-isoform3(h)	EPHB2	EPH receptor B2	-0.77	224
MO000138295	ZBED1(h)	ZBED1	zinc finger BED-type containing 1	-0.56	237
MO000031039	NCoR2(h)	NCOR2	nuclear receptor corepressor 2	-0.74	277
MO001087366	scmh1:BMI1,PCGF2:PHC1,PHC2,PHC3:RING1,RNF2:CBX2,CBX4,CBX8	BMI1, CBX2, CBX4, CBX8, PCGF2, PHC1, PHC2, PHC3, RING1, RNF2, SCMH1	BMI1 proto-oncogene, polycomb ring finger, Scm polycomb group protein homolog 1, chromobox 2, chromo...	-0.47	325
MO000017421	AKT-1(h)	AKT1	AKT serine/threonine kinase 1	-0.41	375
MO000044885	PP1-alpha(h)	PPP1CA	protein phosphatase 1 catalytic subunit alpha	-0.53	376
MO001091620	IL-18Ralpha(h):IL37,IL37(?-218)(h):SIGIRR(h)	IL18R1, SIGIRR	interleukin 18 receptor 1, single Ig and TIR domain containing	-0.6	384

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figures 9 and 10. 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.

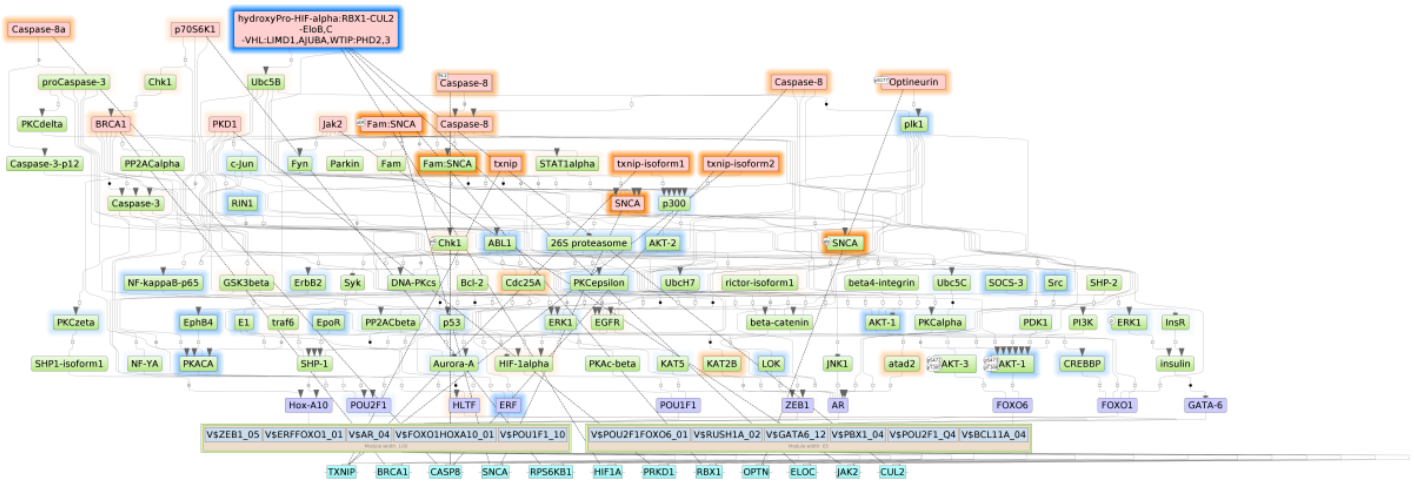


Figure 9. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in noRA_Dox vs. noRA_noDox. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network 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 →

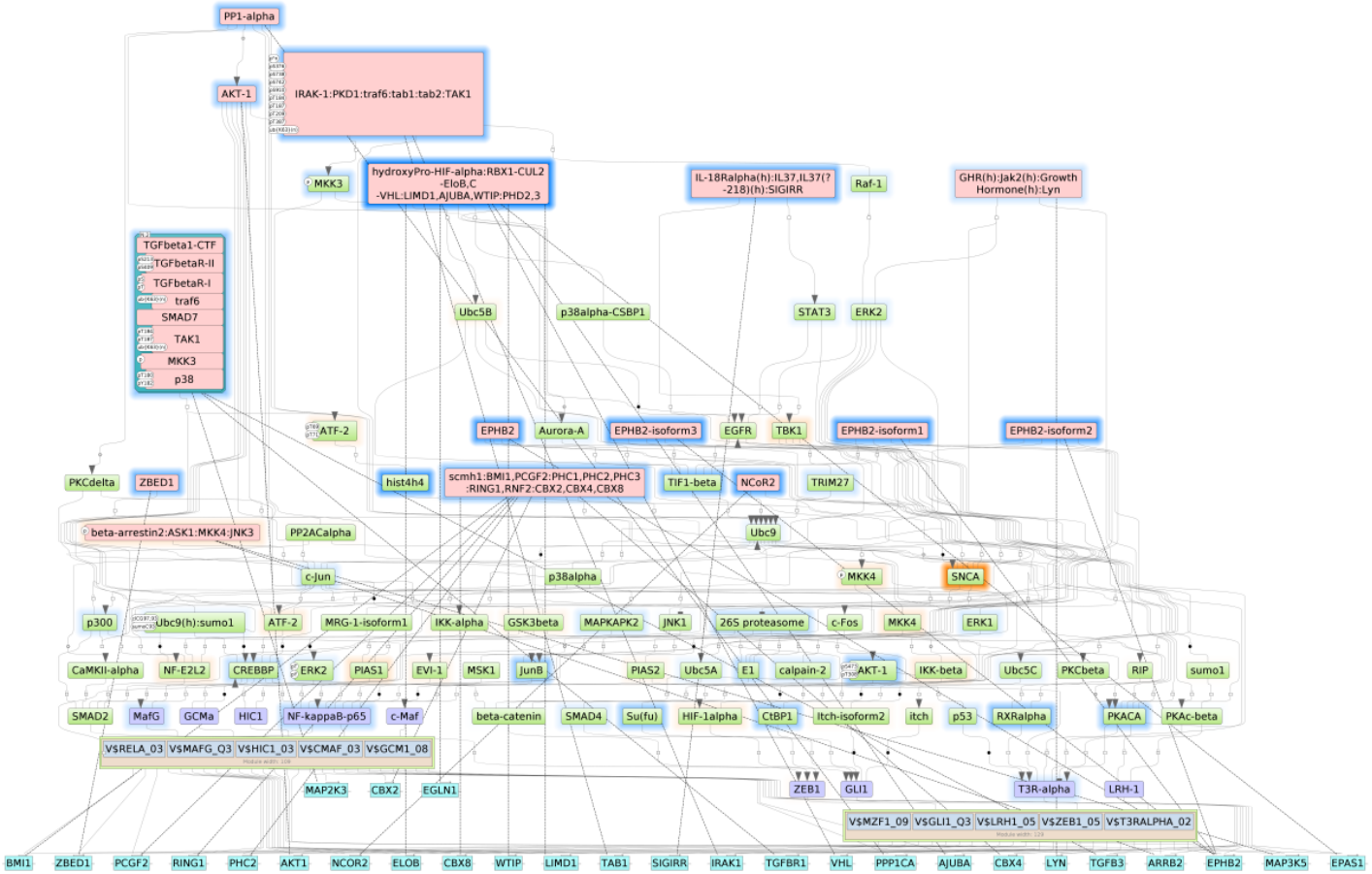


Figure 10. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in noRA_Dox vs. noRA_noDox. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network 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™ [5] 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
CASP8	caspase 8	31	0.55	49
TXNIP	thioredoxin interacting protein	1	0.87	146
SNCA	synuclein alpha	1	4.38	249
PTPN2	protein tyrosine phosphatase non-receptor type 2	5	0.3	349
TEC	tec protein tyrosine kinase	30	0.51	424
CDK7	cyclin dependent kinase 7	29	0.23	443

*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.*

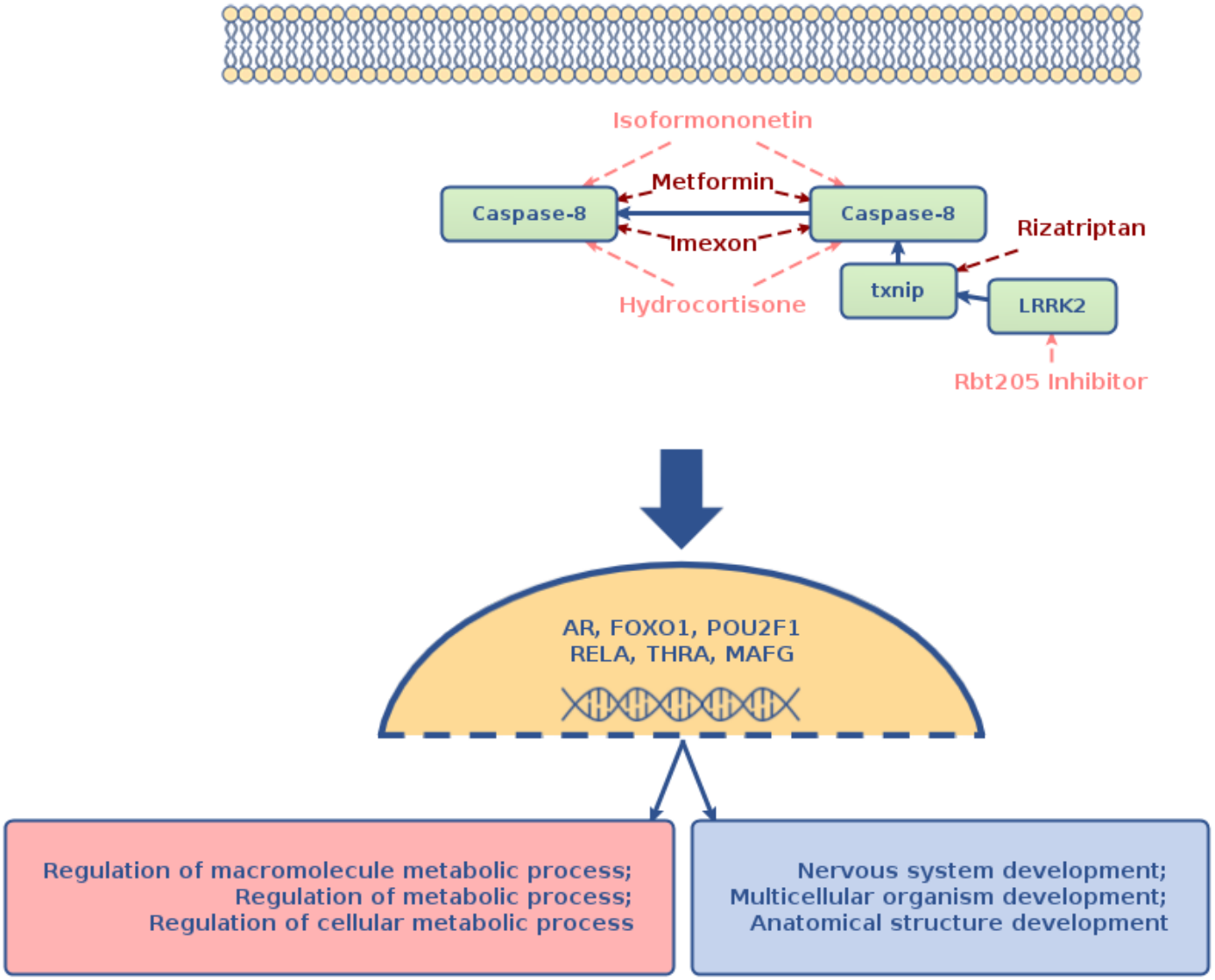
See full table →

Gene symbol	Gene Description	Druggability score	logFC	Total rank
CASP8	caspase 8	50.89	0.55	49
LRRK2	leucine rich repeat kinase 2	3.25	0.32	275
PTPN2	protein tyrosine phosphatase non-receptor type 2	18.31	0.3	349
TEC	tec protein tyrosine kinase	7.34	0.51	424
CDK7	cyclin dependent kinase 7	12.25	0.23	443
CCNH	cyclin H	2.59	0.23	443

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:

- Caspase-8
- LRRK2
- txnip
- Caspase-8

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: Imexon, Isoformononetin, Rizatriptan, Rbt205 Inhibitor, Metformin and Hydrocortisone, 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
Sirolimus	IKBKB, MAPK10, RPS6KA3, ROCK2, DYRK1A, FGF2, HSP90AA1, TGM2, NFE2L2, MAPK12, CHEK1, PRKD1, RPS6KB1	92	2	small molecule,approved,investigational
Nicotinamide	SIRT1, SIRT5	75	3	small molecule,approved
Metformin	CASP8	73	1	small molecule,approved
L-Sulforaphane	MAPK10, NFE2L2, MAPK12, CD44, CCL2	67	1	small molecule, investigational
Moxifloxacin	TOP2B	51	1	small molecule,approved,investigational

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
1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, TNIK, MERTK, LATS2, DYRK1A, NUAK1, CLK4, MAPK12, WEE1, EPHA6, TGFBR2, CLK1, NLK, PLK4	96	PHASE2: Arthritis, Arthritis, Rheumatoid, Psoriasis
pi-103	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, TNIK, MERTK, LATS2, DYRK1A, NUAK1, CLK4, MAPK12, WEE1, EPHA6, TGFBR2, CLK1, NLK, PLK4	96	N/A
seliciclib	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, TNIK, MERTK, LATS2, DYRK1A, NUAK1, CLK4, MAPK12, WEE1, EPHA6, TGFBR2, CLK1, NLK, PLK4	95	PHASE2: Cystic Fibrosis, Cysts, Fibrosis
ruboxistaurin	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, TNIK, MERTK, LATS2, DYRK1A, NUAK1, CLK4, MAPK12, WEE1, EPHA6, TGFBR2, CLK1, NLK, PLK4	95	PHASE1: Diabetes Mellitus, Diabetes Mellitus, Type 2, Heart Failure
Sorafenib	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, TNIK, MERTK, LATS2, DYRK1A, NUAK1, CLK4, MAPK12, WEE1, EPHA6, TGFBR2, CLK1, NLK, PLK4	95	EARLY_PHASE1: Carcinoma, Carcinoma, Hepatocellular, Liver Neoplasms, Neoplasms

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
Lipoic Acid	STT3A, PTPRO, PTPRJ, EPM2A, PTPN2, PTPN13, CDC25C, PTPRK, UBASH3B, CDC25A, PTPRD, PPM1D, PTPN12	95	0.51
Tiludronate	CDC25A, PTPRD, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, IL33, PTPN13, PTPN12, CDC25C, PTPRK, UBASH3B	93	0.32
[[N-(Benzyloxycarbonyl)Amino]Methyl]Phosphate	CDC25A, PTPRD, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, CDC25C, PTPRK, UBASH3B	90	0.31
Tacrolimus	PPP3CA, PPM1D, IL33	89	7.3E-2
3-(Phosphonomethyl)Pyridine-2-Carboxylic Acid	CDC25A, PTPRD, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, CDC25C, PTPRK, UBASH3B	89	0.35

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
Edotecarin	PNPT1, PRKCH, PRKCI, PRKD1	83	0.12
Isoformononetin	MAPK10, NFE2L2, MAPK12, CASP8, HIF1A, BRCA1, HSP90AA1, SENP7	83	0.21
2,5,7-Trihydroxynaphthoquinone	MAPK10, CDC25A, MAPK12, SENP6, EPM2A, CASP8, DYRK1A, SENP8, BRCA1, HSP90AA1	83	0.36
Thioproline	ITGA6, IL33, ITGB1, ITGAV, ITGA1	80	0.16
3-(4-HYDROXY-3-METHOXYPHENYL)-2-PROPENOIC ACID	MAPK10, TLR4, MAPK12, IL33, CASP8, BRCA1	79	0.1

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Sirolimus, 1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea, Lipoic Acid and Edotecarin. These drugs were selected for acting on the following targets: PRKD1, TEC and PTPN2, 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 *Parkinson 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:

Sirolimus, 1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea, Lipoic Acid and Edotecarin

These drugs were selected for acting on the following targets: PRKD1, TEC and PTPN2, 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:

Caspase-8, LRRK2, txnip and Caspase-8

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: Imexon, Isoformononetin, Rizatriptan, Rbt205 Inhibitor, Metformin and Hydrocortisone. 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.

- Caspase-8
- LRRK2
- txnip
- Caspase-8

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 2024.2 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac).
The master regulator search uses the TRANSPATH® database (BIOBASE), release 2024.2 (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 2024.2 (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 [7] 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 [9]). We control the error rate by the FDR threshold 0.05.

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-score*$_{PSD}$) is calculated as follows:

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

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, *maxRank(T)* equals *max(rank(t))* for all targets *t* in *T*.
We use following formula to calculate "Disease activity score" (*D-score*$_{PSD}$):

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

where *D* is the set of selected diseases, and if *D* is empty set, *D-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. 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
6. 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
7. 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
8. 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.
9. 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
10. 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
11. 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
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, Gloriozova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *ChemInform.* **1999**;39(4):666-670. doi:10.1002/chin.199940210

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