SPP1 and CLK4 are promising druggable targets for treating Parkinson Disease that control activity of ESR1, SMAD2 and RAD21 transcription factors on promoters of differentially expressed genes

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 02/07/2020 ; Run on 09/11/2022 ; Report generated on 10/11/2022

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: ESR1, SMAD2, SMAD3, RAD21, BRCA2 and NR3C1. The subsequent network analysis suggested

- CLK4
- SNCA
- FIGF
- SNCA
- Osteopontin

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 2,5,7-Trihydroxynaphthoquinone .

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] 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 [11-13]. 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

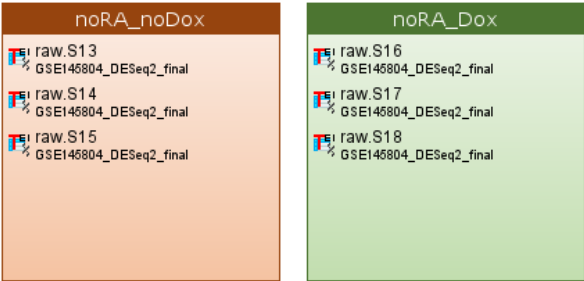| File name	Data type
GSE145804_DESeq2_final	Transcriptomics

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 5393 upregulated genes (LogFC>0.1) out of which 578 genes were found as significantly upregulated (p-value<0.1) and 6588 downregulated genes (LogFC<-0.1) out of which 726 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	10.36	2.71E-88	4.59E-84
ENSG00000118785	SPP1	secreted phosphoprotein 1	2.98	-0.28	6.23E-7	2.11E-3
ENSG00000169282	KCNAB1	potassium voltage-gated channel subfamily A member regulatory beta subunit 1	2.91	0.49	1.32E-7	5.61E-4
ENSG00000162692	VCAM1	vascular cell adhesion molecule 1	2.67	-4.05E-2	1.1E-5	2.34E-2
ENSG00000237280		novel transcript	2.15	0.2	1.42E-4	0.16
ENSG00000214892	USP8P1	USP8 pseudogene 1	1.81	0.19	4.4E-3	0.83
ENSG00000224837	GCSHP5	glycine cleavage system protein H pseudogene 5	1.72	-0.1	1.84E-3	0.6
ENSG00000243300			1.65	0.44	2.8E-3	0.68
ENSG00000229474	PATL2	PAT1 homolog 2	1.53	-0.14	7.97E-3	0.98
ENSG00000236813	BTF3P8	basic transcription factor 3 pseudogene 8	1.53	-0.48	9.95E-3	0.98

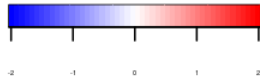*Table 3. Top ten significant **down-regulated** genes in noRA_Dox vs. noRA_noDox.*
See full table →

ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000186081	KRT5	keratin 5	-10.36	2.57	9.35E-7	2.64E-3
ENSG00000257594	GALNT4	polypeptide N-acetylgalactosaminyltransferase 4	-2.37	1.39E-2	4.31E-5	6.08E-2
ENSG00000167244	IGF2	insulin like growth factor 2	-2.13	-0.49	1.23E-3	0.47
ENSG00000255115		family with sequence similarity 162, member A (FAM162A) pseudogene	-1.97	-0.38	2.22E-3	0.61
ENSG00000134955	SLC37A2	solute carrier family 37 member 2	-1.9	-0.42	9.67E-3	0.98
ENSG00000111679	PTPN6	protein tyrosine phosphatase non-receptor type 6	-1.9	-0.13	4.43E-3	0.83
ENSG00000179846	NKPD1	NTPase KAP family P-loop domain containing 1	-1.87	-0.33	3.5E-3	0.79
ENSG00000137801	THBS1	thrombospondin 1	-1.83	0.49	4.4E-3	0.83
ENSG00000269054	ZNF497-AS1	ZNF497 antisense RNA 1	-1.82	-9.51E-2	1.69E-3	0.56
ENSG00000224886		novel pseudogene	-1.82	-0.35	6.76E-3	0.98

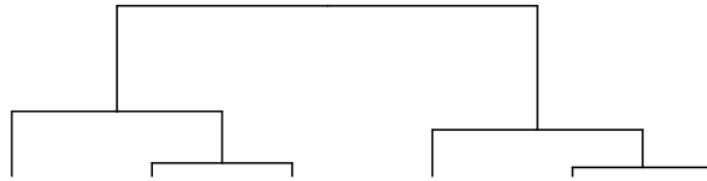## *3.2. Functional classification of genes*

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.

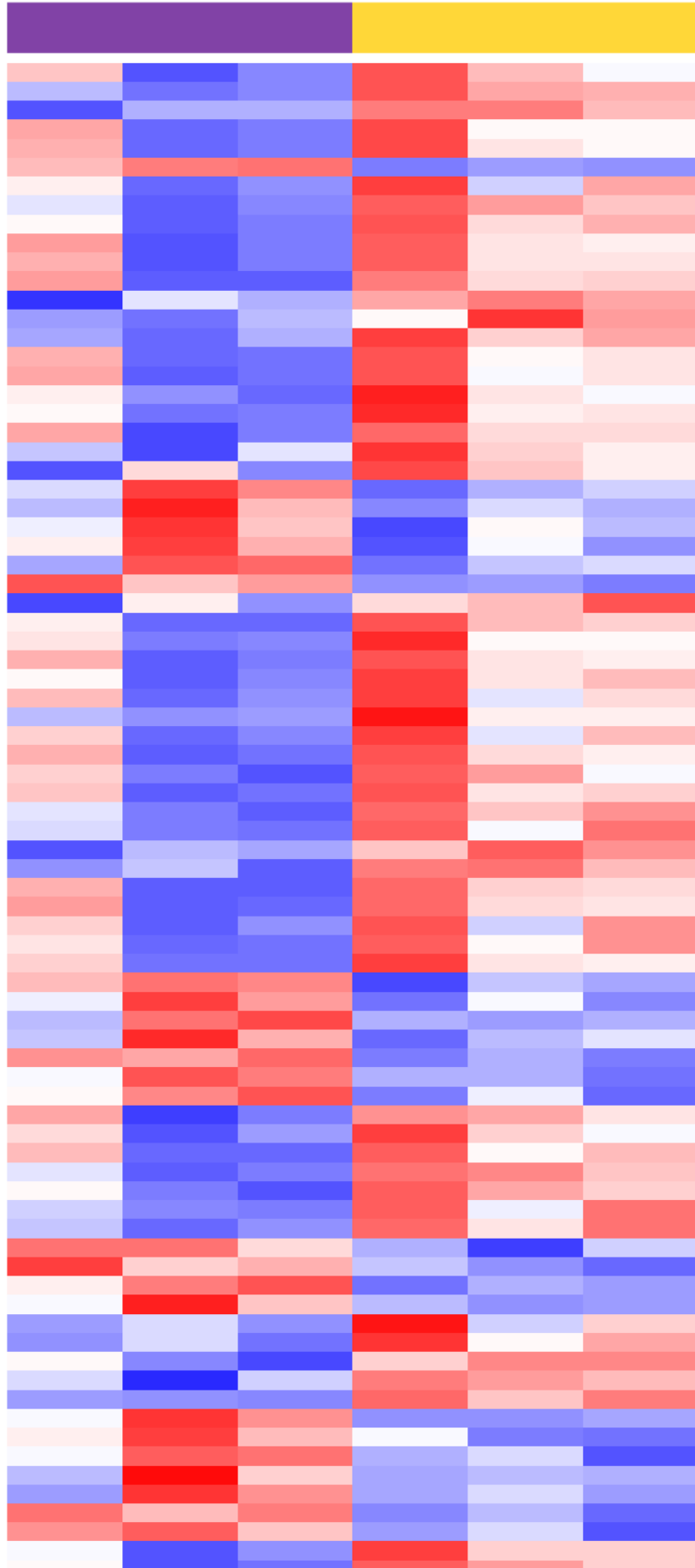## Heatmap of differentially expressed genes in noRA_Dox vs. noRA_noDox

A heatmap of all differentially expressed genes playing a potential regulatory role in the system (enriched in TRANSPATH® pathways) is presented in Figure 2.
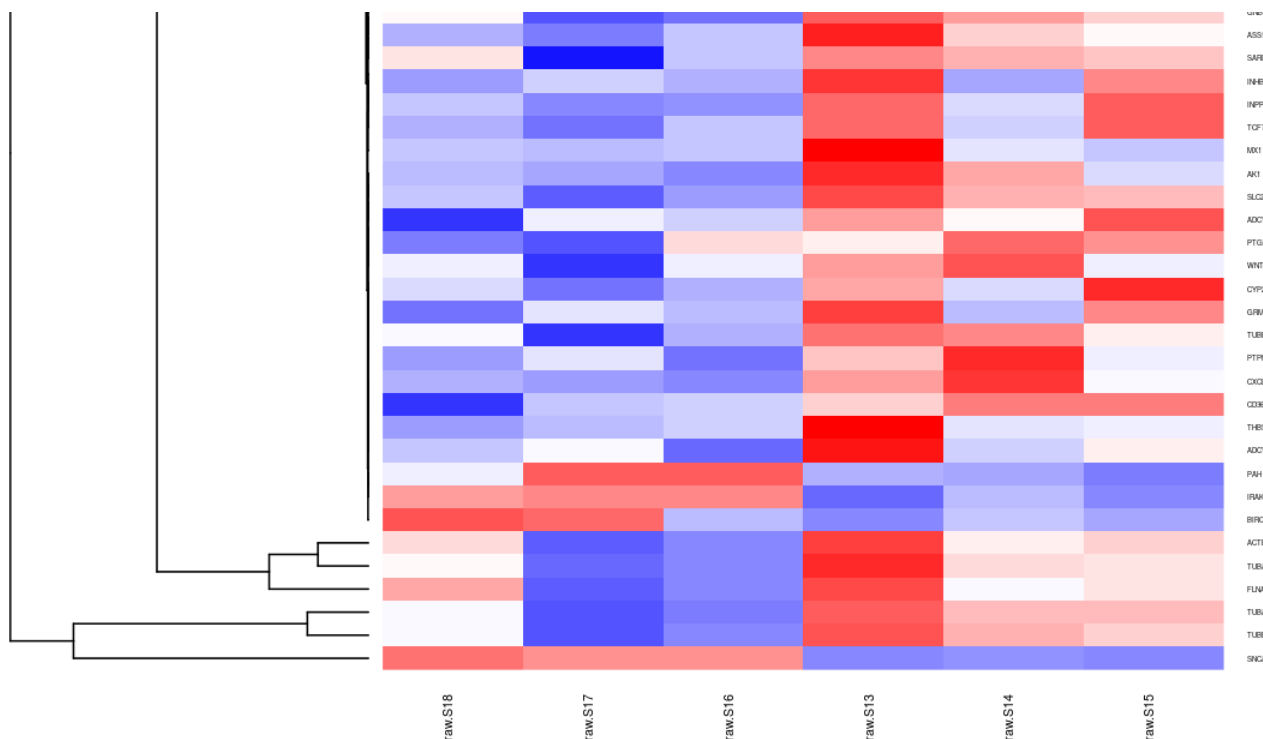
Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.
See full diagram →

Up-regulated genes in noRA_Dox vs. noRA_noDox:

578 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 →
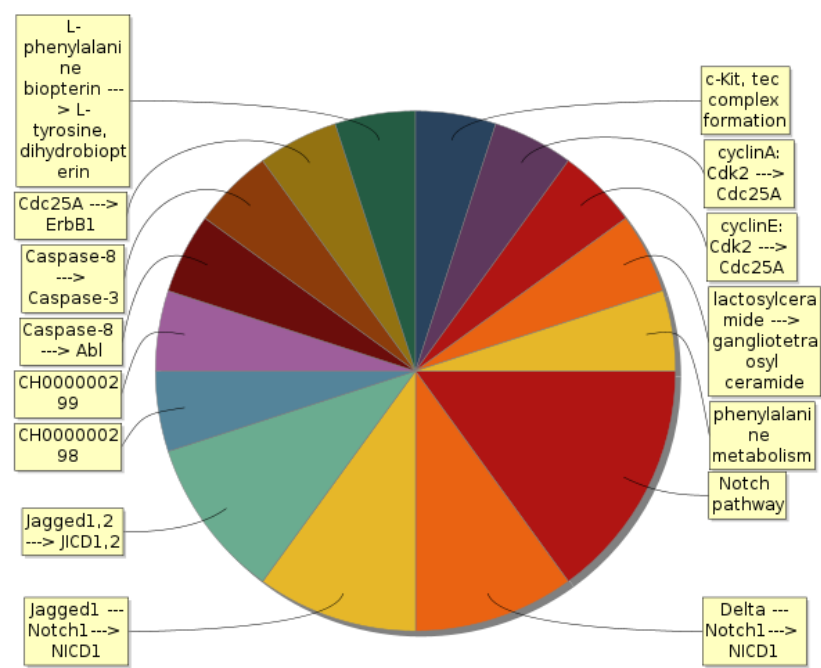
TRANSPATH® Pathways (2022.2)

Figure 4. Enriched TRANSPATH® Pathways (2022.2) of up-regulated genes in noRA_Dox vs. noRA_noDox.
Full classification →

HumanPSD(TM) disease (2022.2)

■ Central Nervous System Diseases ■ Brain Diseases ■ Dry Eye Syndromes ■ Eosinophilia
■ Lacrimal Apparatus Diseases ■ AIDS-Related Opportunistic Infections
■ Digestive System Abnormalities ■ Opportunistic Infections ■ Fibromuscular Dysplasia
■ Giant Cell Tumor of Tendon Sheath ■ Synovitis ■ Synovitis, Pigmented Villonodular
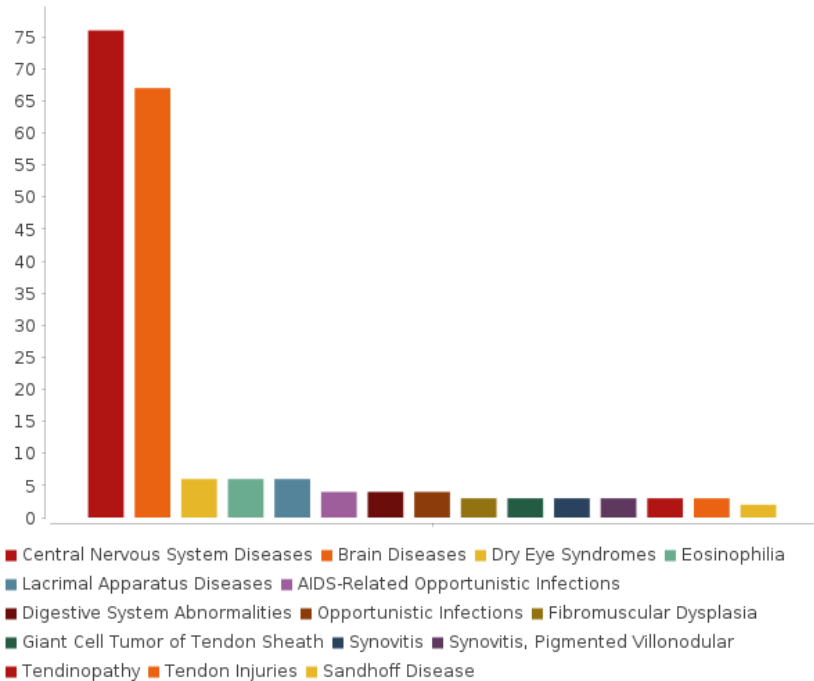■ Tendinopathy ■ Tendon Injuries ■ Sandhoff Disease

Figure 5. Enriched HumanPSD(TM) disease (2022.2) 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:

726 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.
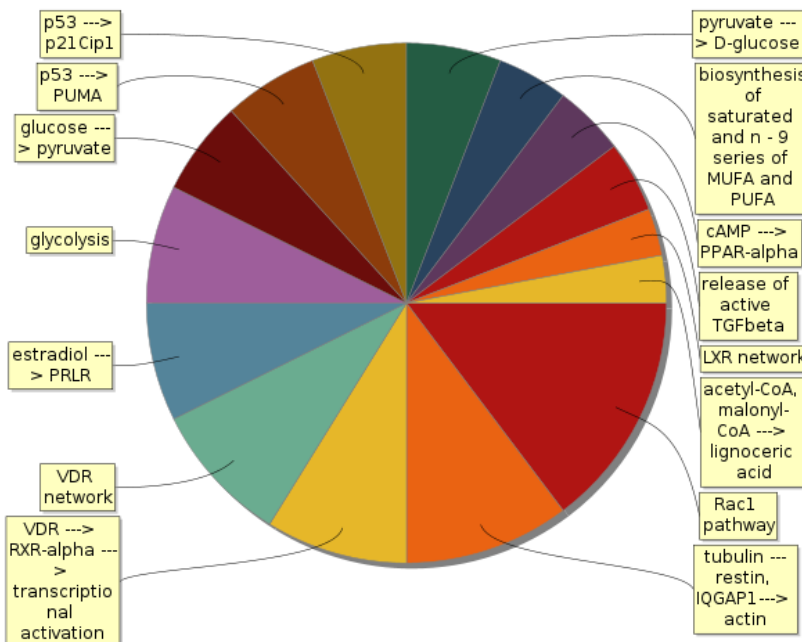**Full classification →**

TRANSPATH® Pathways (2022.2)

Figure 7. Enriched TRANSPATH® Pathways (2022.2) of down-regulated genes in noRA_Dox vs. noRA_noDox.
Full classification →
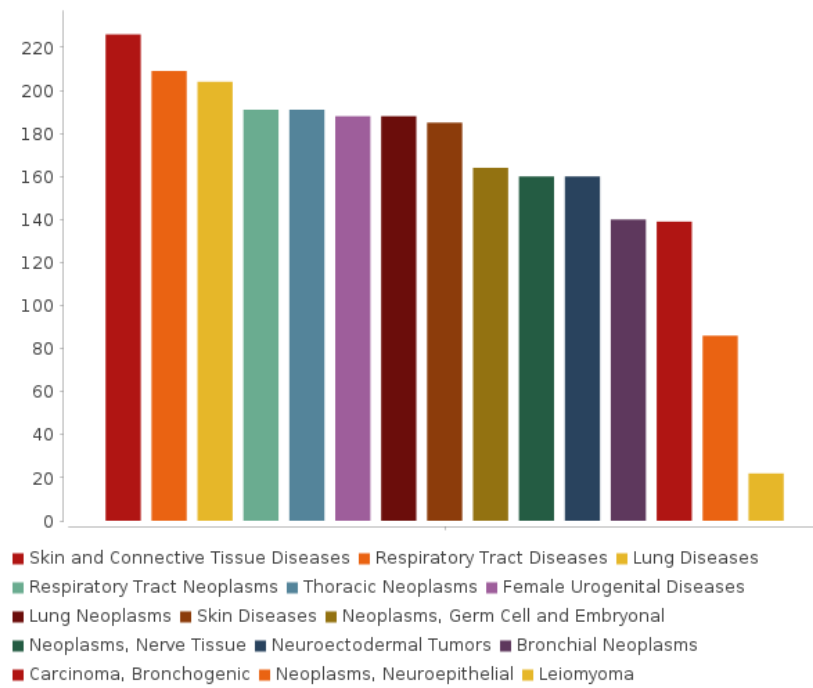
HumanPSD(TM) disease (2022.2)

Figure 8. Enriched HumanPSD(TM) disease (2022.2) 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):
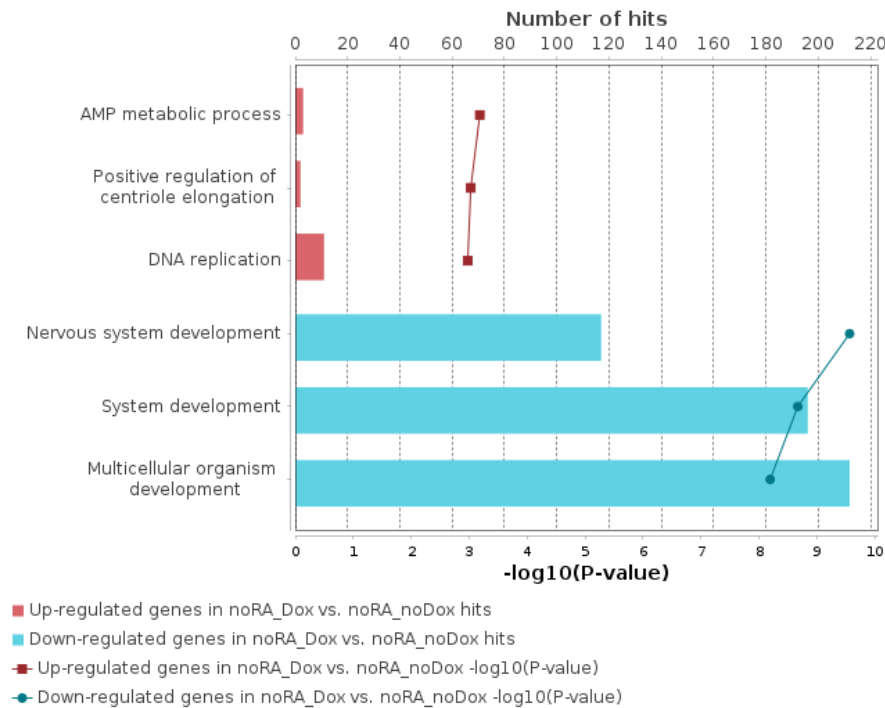
3.3. Analysis of enriched transcription factor binding sites and composite modules

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].
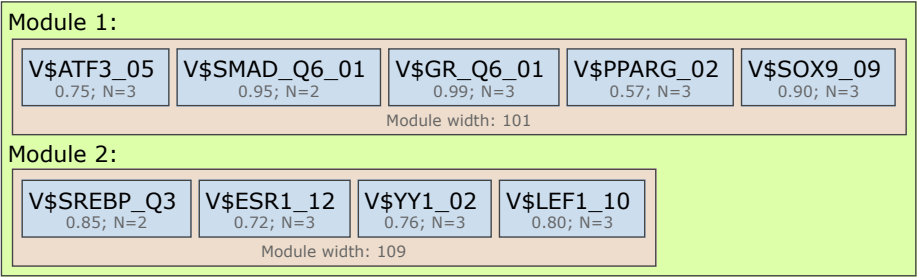
We applied the Composite Module Analyst (CMA) [8] method to detect such potential enhancers, as targets of multiple TFs bound in a cooperative manner to the regulatory regions of the genes of interest. CMA applies a genetic algorithm to construct a generalized model of the enhancers by specifying combinations 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$ATF3_05	V$SMAD_Q6_01	V$GR_Q6_01	V$PPARG_02	V$SOX9_09
0.75; N=3	0.95; N=2	0.99; N=3	0.57; N=3	0.90; N=3

Module width: 101

Module 2:

V$SREBP_Q3	V$ESR1_12	V$YY1_02	V$LEF1_10
0.85; N=2	0.72; N=3	0.76; N=3	0.80; N=3

Module width: 109

Model score (-p*log10(pval)): 14.29
Wilcoxon p-value (pval): 2.42e-28
Penalty (p): 0.517
Average yes-set score: 7.65
Average no-set score: 6.26
AUC: 0.73
Separation point: 7.16
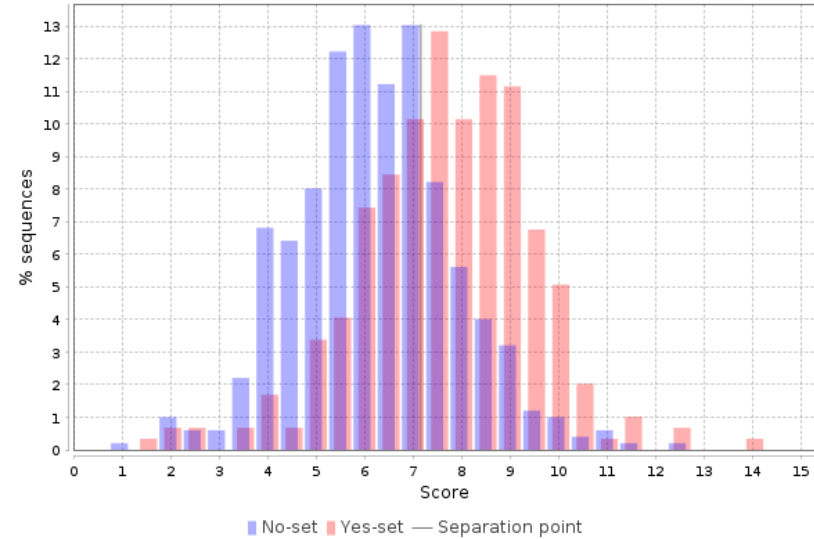**False-positive:** 25.85%
False-negative: 35.81%

*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.*
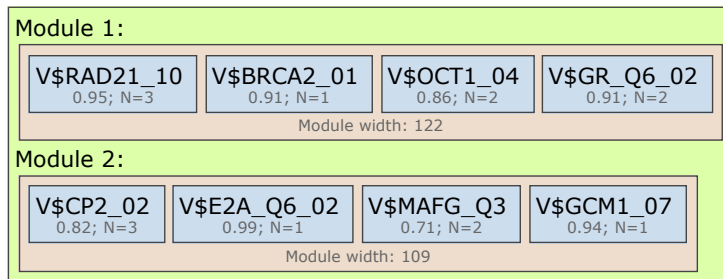
See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000154822	PLCL2	phospholipase C like 2	13.78	LEF-1(h), YY1(h), SREBP-1(h),SREBP-2(h), ER-alpha(h), SMAD1(h),SMAD2(h),SMAD3(h),SMAD4(h),SMAD5(h),SMAD6(h),SMAD7(h),SMAD9(h), PPARgamma(h), ATF-3(h)...
ENSG00000171951	SCG2	secretogranin II	13.77	ER-alpha(h), LEF-1(h), ATF-3(h), SOX-9(h), PPARgamma(h), SMAD1(h),SMAD2(h),SMAD3(h),SMAD4(h),SMAD5(h),SMAD6(h),SMAD7(h),SMAD9(h), YY1(h)...
ENSG00000171428	NAT1	N-acetyltransferase 1	13.13	ER-alpha(h), YY1(h), LEF-1(h), SREBP-1(h),SREBP-2(h), GR(h), SOX-9(h), PPARgamma(h)...
ENSG00000230793	SMARCE1P5	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1 pseudogene 5	12.73	PPARgamma(h), ATF-3(h), SOX-9(h), ER-alpha(h), SMAD1(h),SMAD2(h),SMAD3(h),SMAD4(h),SMAD5(h),SMAD6(h),SMAD7(h),SMAD9(h), LEF-1(h), YY1(h)...
ENSG00000225912	RPL26P36	ribosomal protein L26 pseudogene 36	12.6	ATF-3(h), LEF-1(h), SREBP-1(h),SREBP-2(h), ER-alpha(h), YY1(h), GR(h), SOX-9(h)...
ENSG00000108691	CCL2	C-C motif chemokine ligand 2	12.49	SMAD1(h),SMAD2(h),SMAD3(h),SMAD4(h),SMAD5(h),SMAD6(h),SMAD7(h),SMAD9(h), GR(h), PPARgamma(h), ATF-3(h), LEF-1(h), SOX-9(h), ER-alpha(h)...
ENSG00000215068		novel transcript, antisense to ANXA2R	12.44	SREBP-1(h),SREBP-2(h), SMAD1(h),SMAD2(h),SMAD3(h),SMAD4(h),SMAD5(h),SMAD6(h),SMAD7(h),SMAD9(h), LEF-1(h), SOX-9(h), PPARgamma(h), ER-alpha(h), GR(h)...
ENSG00000139323	POC1B	POC1 centriolar protein B	12.36	YY1(h), ER-alpha(h), SREBP-1(h),SREBP-2(h), LEF-1(h), ATF-3(h), SOX-9(h), PPARgamma(h)...
ENSG00000206688	SNORD116-18	small nucleolar RNA, C/D box 116-18	12.11	SMAD1(h),SMAD2(h),SMAD3(h),SMAD4(h),SMAD5(h),SMAD6(h),SMAD7(h),SMAD9(h), GR(h), PPARgamma(h), SOX-9(h), ATF-3(h), ER-alpha(h), YY1(h)...
ENSG00000172575	RASGRP1	RAS guanyl releasing protein 1	12.05	SOX-9(h), SREBP-1(h),SREBP-2(h), PPARgamma(h), ER-alpha(h), SMAD1(h),SMAD2(h),SMAD3(h),SMAD4(h),SMAD5(h),SMAD6(h),SMAD7(h),SMAD9(h), ATF-3(h), YY1(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$RAD21_10	V$BRCA2_01	V$OCT1_04	V$GR_Q6_02
0.95; N=3	0.91; N=1	0.86; N=2	0.91; N=2

Module width: 122

Module 2:

V$CP2_02	V$E2A_Q6_02	V$MAFG_Q3	V$GCM1_07
0.82; N=3	0.99; N=1	0.71; N=2	0.94; N=1

Module width: 109

Model score (-p*log10(pval)): 18.98
Wilcoxon p-value (pval): 3.81e-36
Penalty (p): 0.536
Average yes-set score: 6.70
Average no-set score: 5.35
AUC: 0.77
Separation point: 6.27
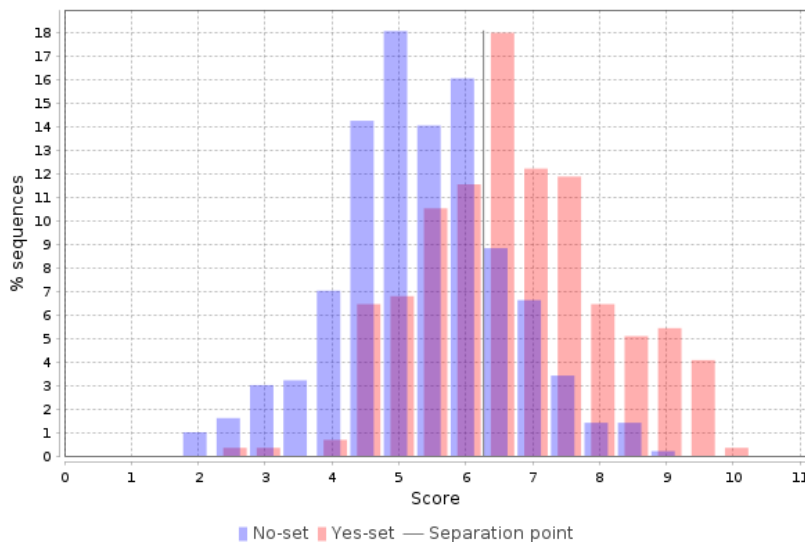**False-positive:** 21.44%
False-negative: 36.61%

*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
ENSG00000133519	ZDHHC8P1	ZDHHC8 pseudogene 1	10.91	hHR21(h), BRCA2(h), GR(h), CP2(h), MafG(h), GCMa(h), E2A(h)
ENSG00000178404	CEP295NL	CEP295 N-terminal like	10.69	hHR21(h), CP2(h), MafG(h), GR(h), BRCA2(h), GCMa(h), E2A(h)
ENSG00000262528		novel transcript, antisense to WDR90	10.67	MafG(h), CP2(h), E2A(h), GCMa(h), GR(h), hHR21(h), BRCA2(h)
ENSG00000065802	ASB1	ankyrin repeat and SOCS box containing 1	10.54	GR(h), hHR21(h), BRCA2(h), E2A(h), GCMa(h), CP2(h), MafG(h)
ENSG00000103066	PLA2G15	phospholipase A2 group XV	10.5	hHR21(h), GR(h), BRCA2(h), MafG(h), CP2(h), GCMa(h), E2A(h)
ENSG00000167470	MIDN	midnolin	10.47	MafG(h), hHR21(h), CP2(h), GCMa(h), E2A(h), BRCA2(h), GR(h)
ENSG00000225746	MEG8	maternally expressed 8, small nucleolar RNA host gene	10.4	GCMa(h), MafG(h), CP2(h), E2A(h), hHR21(h), POU2F1(h), GR(h)...
ENSG00000229427	ANKRD26P4	ankyrin repeat domain 26 pseudogene 4	10.27	hHR21(h), BRCA2(h), GR(h), POU2F1(h), CP2(h), GCMa(h), MafG(h)...
ENSG00000122136	OBP2A	odorant binding protein 2A	10.24	hHR21(h), BRCA2(h), GR(h), POU2F1(h), CP2(h), GCMa(h), MafG(h)...
ENSG00000154358	OBSCN	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF	10.23	MafG(h), CP2(h), GCMa(h), E2A(h), hHR21(h), GR(h), BRCA2(h)

On the basis of the enhancer models we identified transcription factors potentially regulating the **target genes** of our interest. We found 17 and 8 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
MO000019431	ESR1	estrogen receptor 1	2.76	1.37
MO000057829	SMAD2	SMAD family member 2	2.7	1.28
MO000057832	SMAD3	SMAD family member 3	2.67	1.4
MO000020635	SMAD5	SMAD family member 5	2.37	1.28
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	2.24	1.3
MO000033565	PPARG	peroxisome proliferator activated receptor gamma	2.24	1.61
MO000020634	SMAD7	SMAD family member 7	2.22	1.28
MO000090056	ATF3	activating transcription factor 3	2.14	1.27
MO000019609	SMAD1	SMAD family member 1	2.13	1.29
MO000020402	SMAD4	SMAD family member 4	2.13	1.3

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
MO000042938	RAD21	RAD21 cohesin complex component	3.09	1.33
MO000045487	BRCA2	BRCA2 DNA repair associated	3.06	1.32
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	3.01	4.17
MO000032492	TCF3	transcription factor 3	2.98	2.54
MO000025003	POU2F1	POU class 2 homeobox 1	2.97	1.92
MO000117988	TFCP2	transcription factor CP2	2.77	5.07
MO000026306	GCM1	glial cells missing transcription factor 1	2.36	5.5
MO000028667	MAFG	MAF bZIP transcription factor G	2.32	20.3

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: ESR1, SMAD2, SMAD3, RAD21, BRCA2 and NR3C1.
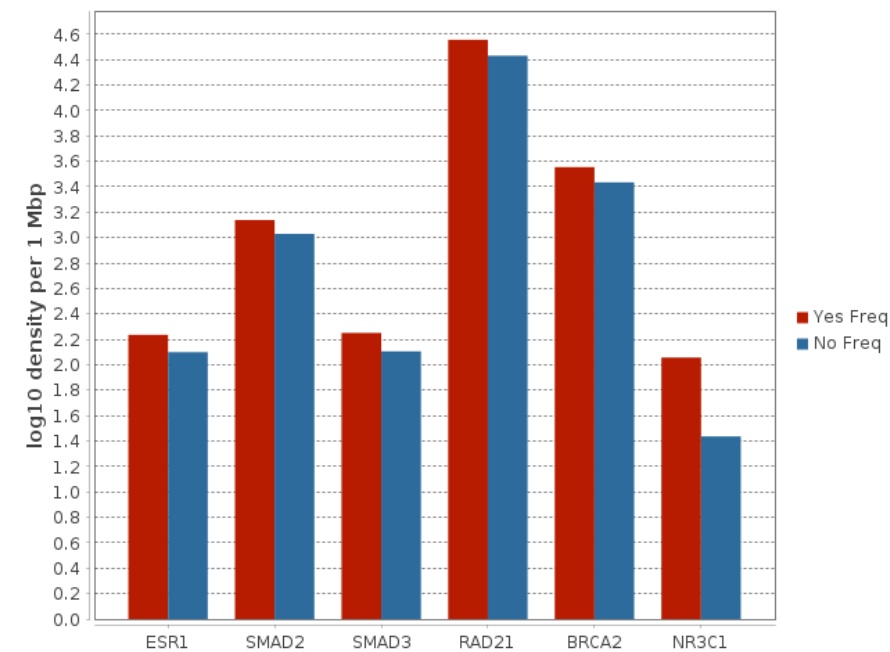
3.4. Finding master regulators in networks

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
MO000480224	SNCA(h){pS129}	SNCA	synuclein alpha	4.38	79
MO000044264	SNCA(h)	SNCA	synuclein alpha	4.38	80
MO000103362	SNCA-isoform3(h)	SNCA	synuclein alpha	4.38	81
MO000044265	SNCA-isoform1(h)	SNCA	synuclein alpha	4.38	82
MO000103359	SNCA-isoform2(h)	SNCA	synuclein alpha	4.38	83
MO000125587	Osteopontin(h)	SPP1	secreted phosphoprotein 1	2.98	89
MO000044272	SNCA(h){gly}	SNCA	synuclein alpha	4.38	93
MO000038316	LPS:lbp:CD14:TLR4:MD-2:TIRAP:IRAK-2	CD14, IRAK2, LBP, LY96, TIRAP, TLR4	CD14 molecule, TIR domain containing adaptor protein, interleukin 1 receptor associated kinase 2, li...	0.86	141
MO000032073	cIAP-2(h)	BIRC3	baculoviral IAP repeat containing 3	0.62	160
MO000125585	Osteopontin-isoform1(h)	SPP1	secreted phosphoprotein 1	2.98	182

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
MO000032657	DUSP8(h)	DUSP8	dual specificity phosphatase 8	-0.65	100
MO000004685	ERK4(h)	MAPK4	mitogen-activated protein kinase 4	-0.63	132
MO000281381	(angiotensin II)2:(AT2 receptor)2:(ATIP-isoform3)2:SHP-1	AGT, AGTR2, MTUS1, PTPN6	angiotensin II receptor type 2, angiotensinogen, microtubule associated scaffold protein 1, protein ...	-1.9	134
MO000104136	cyclosome(h):Fzr1(h)	ANAPC1, ANAPC10, ANAPC11, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC26, CDC27, FZR1	anaphase promoting complex subunit 1, anaphase promoting complex subunit 10, anaphase promoting comp...	-0.54	143
MO000041952	calpain-1(h)	CAPN1	calpain 1	-0.48	191
MO000017291	integrins	ITGA1, ITGA2B, ITGA3, ITGA4, ITGA5, ITGA6, ITGA8, ITGA9, ITGAL, ITGAV, ITGB1, ITGB2, ITGB3, ITGB4, I...	integrin subunit alpha 1, integrin subunit alpha 2b, integrin subunit alpha 3, integrin subunit alph...	-1.12	229
MO000044269	cyclosome(h):Fzr(h)	ANAPC1, ANAPC10, ANAPC11, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC26, CDC27, FZR1	anaphase promoting complex subunit 1, anaphase promoting complex subunit 10, anaphase promoting comp...	-0.54	244
MO000043414	cyclosome(h)	ANAPC1, ANAPC10, ANAPC11, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC26, CDC27	anaphase promoting complex subunit 1, anaphase promoting complex subunit 10, anaphase promoting comp...	-0.54	279
MO000031189	PKCdelta(h)	PRKCD	protein kinase C delta	-0.34	298
MO000186472	CDK5R1p35(h):Cdk5-isoform1(h)	CDK5, CDK5R1	cyclin dependent kinase 5, cyclin dependent kinase 5 regulatory subunit 1	-0.53	302

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.
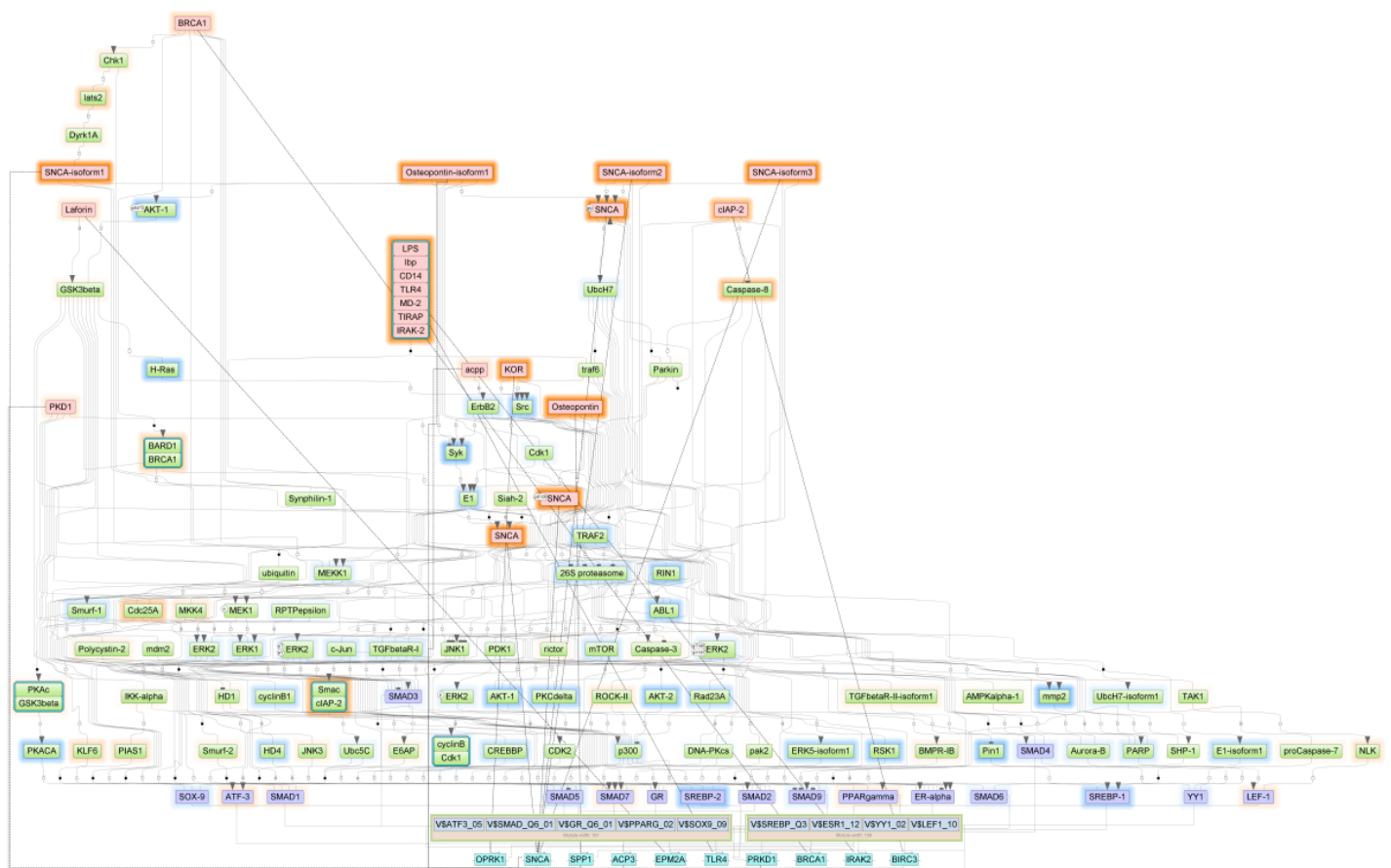
Figure 9. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in noRA_Dox vs. noRA_noDox. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during the search for master regulators from selected TFs. Orange and blue frames highlight molecules that are encoded by up- and downregulated genes, resp.
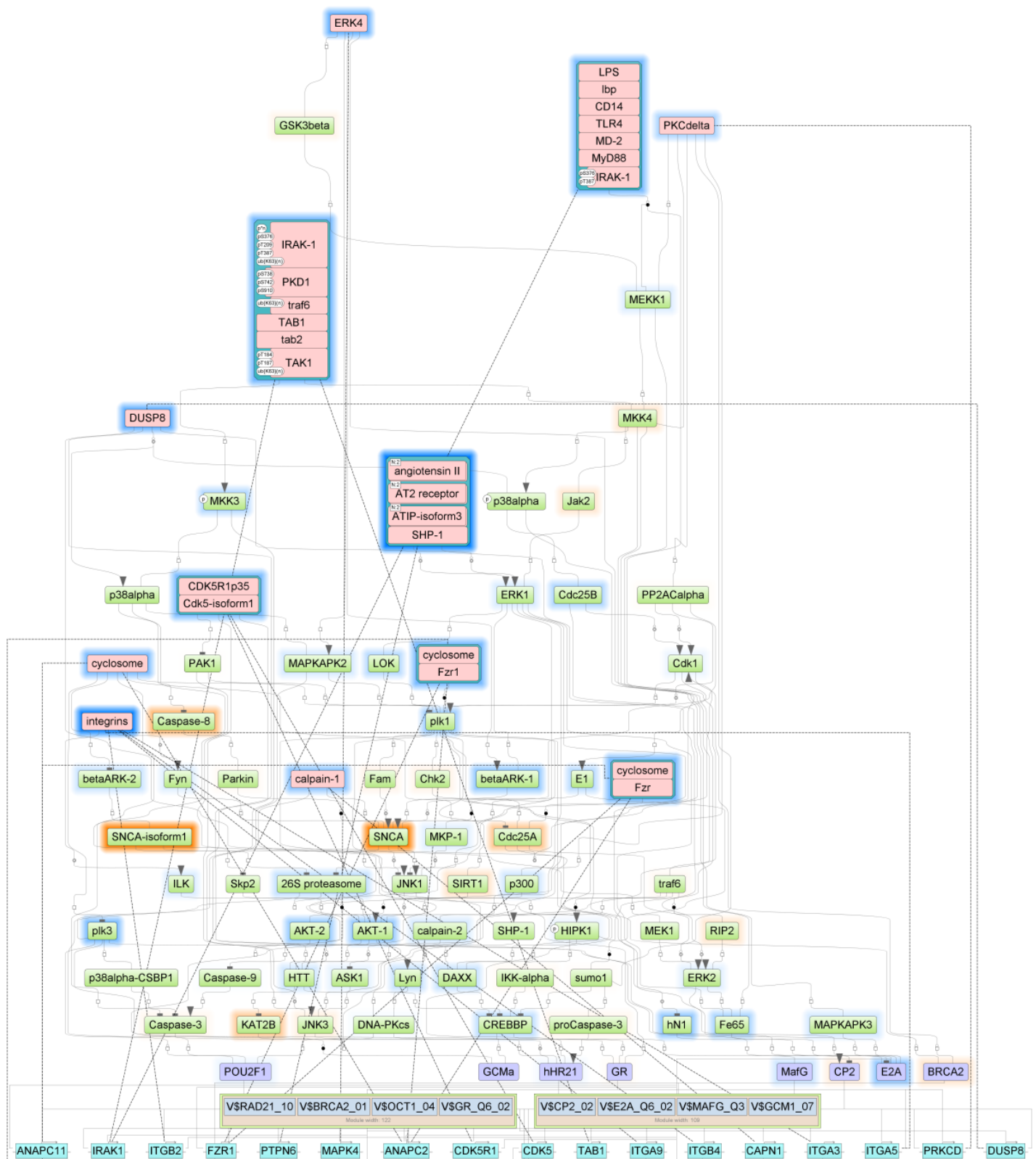
See full diagram →

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 [11-13] 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 [11-13] 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
SPP1	secreted phosphoprotein 1	2	2.98	182
CLK4	CDC like kinase 4	26	0.38	288
TEC	tec protein tyrosine kinase	28	0.52	407
AGT	angiotensinogen	6	0.38	417
ROCK2	Rho associated coiled-coil containing protein kinase 2	14	0.15	458
BIRC3	baculoviral IAP repeat containing 3	6	0.62	466

*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
CLK4	CDC like kinase 4	6.44	0.38	288
VEGFD	vascular endothelial growth factor D	22.32	1.27	352
CCNH	cyclin H	2.59	0.23	365
IRAK2	interleukin 1 receptor associated kinase 2	3.21	0.86	391
TEC	tec protein tyrosine kinase	7.34	0.52	407
ROCK2	Rho associated coiled-coil containing protein kinase 2	1.58	0.15	458

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:

- CLK4
- SNCA
- FIGF
- SNCA
- Osteopontin

This result allows us to suggest the following schema of affecting the molecular mechanism of the studied pathology:
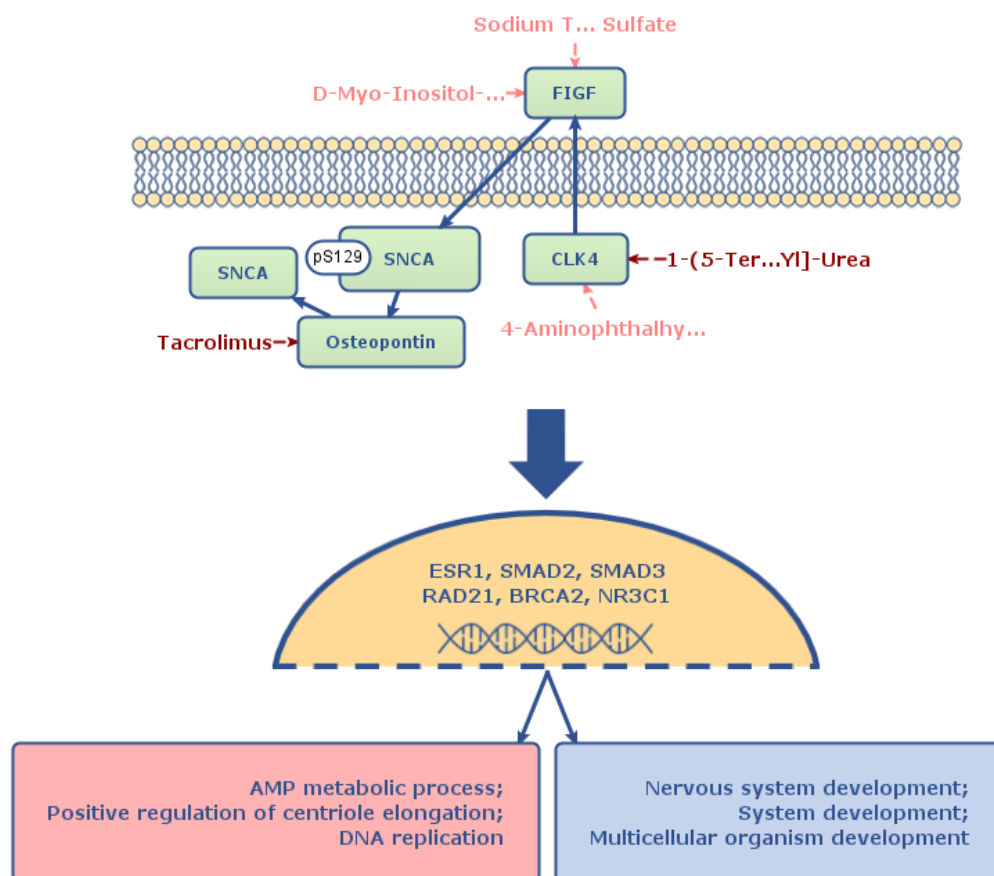
Drugs which are shown on this schema: Tacrolimus, Sodium Tetradecyl Sulfate, D-Myo-Inositol-Hexasulphate, 1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea and 4-Aminophthalhydrazide, should be considered as a prospective research initiative for further drug 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, TGM2, MAPK12, DYRK1A, CHEK1, FGF2, PRKD1, HSP90AA1, RPS6KB1	92	3	Phase 2: Parkinson Disease, Acute Disease, Acute Lung Injury, Adenocarcinoma, Adenocarcinoma of Lung, Adenocarcinoma, Clear Cell, Adenocarcinoma, Papillary, Adenoma, Adenomatous Polyposis Coli, Adenomatous Polyps, Adenomyoepithelioma, Alzheimer Disease, Amyotrophic Lateral Sclerosis, Anemia, Anemia, Aplastic, Anemia, Hemolytic, Anemia, Hemolytic, Autoimmune, Anemia, Hemolytic, Congenital, Anemia, Refractory, Anemia, Refractory, with Excess of Blasts, Anemia, Sickle Cell, Aneurysm, Angina Pectoris, Angina, Unstable, Angiofibroma, Angiomyolipoma, Angiomyoma, Aphasia, Aphasia, Primary Progressive, Arteriovenous Malformations, Arthritis, Arthritis, Rheumatoid, Arthrogryposis, Astrocytoma, Atrophy, Autistic Disorder, Autoimmune Lymphoproliferative Syndrome, Behavior, Blast Crisis, Bone Marrow Failure Disorders, Brain Abscess, Brain Neoplasms, Brain Stem Infarctions, Breast Neoplasms, Breast Neoplasms, Male, Bronchiolitis, Bronchiolitis Obliterans, Bronchitis, Burkitt Lymphoma, COVID-19, Carcinoid Tumor, Carcinoma, Carcinoma, Adenosquamous, Carcinoma, Endometrioid, Carcinoma, Hepatocellular, Carcinoma, Islet Cell, Carcinoma, Neuroendocrine, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Papillary, Carcinoma, Renal Cell, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Carcinoma, Transitional Cell, Carcinoma, Verrucous, Carcinosarcoma, Central Nervous System Neoplasms, Chondrosarcoma, Chondrosarcoma, Mesenchymal, Chordoma, Choroidal Neovascularization, Cockayne Syndrome, Cognitive Dysfunction, Colorectal Neoplasms, Congenital Abnormalities, Constriction, Pathologic, Coronary Artery Disease, Coronary Disease, Cystadenocarcinoma, Cystadenocarcinoma, Serous, Cysts, Deficiency Diseases, Dementia, Depression, Depressive Disorder, Desmoplastic Small Round Cell Tumor, Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetic Retinopathy, Dilatation, Pathologic, Disease, Dry Eye Syndromes, Edema, Endometrial Neoplasms, Epidermolysis Bullosa, Epidermolysis Bullosa Simplex, Epilepsy, Epistaxis, Erythema, Esophageal Squamous Cell Carcinoma, Eye Diseases, Fallopian Tube Neoplasms, Fanconi Anemia, Fanconi Syndrome, Fibroma, Fibromatosis, Aggressive, Fibrosarcoma, Fibrosis, Frontotemporal Dementia, Ganglion Cysts, Ganglioneuroblastoma, Gangliosidoses, Gangliosidosis, GM1, Gastrinoma, Gastrointestinal Neoplasms, Gastrointestinal Stromal Tumors, Gaucher Disease, Genetic Diseases, Inborn, Geographic Atrophy, Giant Lymph Node Hyperplasia, Glioblastoma, Glioma, Gliosarcoma, Glomerulonephritis, Glomerulonephritis, IGA, Glomerulonephritis, Membranous, Glomerulosclerosis, Focal Segmental, Glucagonoma, Glycogen Storage Disease Type VI, Gout, Graft vs Host Disease, Granuloma, Granulomatous Disease, Chronic, Graves Ophthalmopathy, Growth Disorders, HIV Infections, Hamartoma, Hamartoma Syndrome, Multiple, Head and Neck Neoplasms, Hemangioendothelioma, Hemangioma, Hemangiosarcoma, Hematologic Diseases, Hematologic Neoplasms, Hemolysis, Hodgkin Disease, Hyperinsulinism, Hypertension, Hyperuricemia, Hypoglycemia, Hypoxia, Immunologic Deficiency Syndromes, Infarction, Inflammatory Bowel Diseases, Insulinoma, Intestinal Neoplasms, Intracranial Aneurysm, Intracranial Arteriovenous Malformations, Ischemia, Kasabach-Merritt Syndrome, Kidney Diseases, Kidney Diseases, Cystic, Kidney Failure, Chronic, Kidney Neoplasms, Klippel-Trenaunay-Weber Syndrome, Leiomyoma, Leiomyomatosis, Leiomyosarcoma, Leukemia, Leukemia, Biphenotypic, Acute, Leukemia, Erythroblastic, Acute, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, Leukemia, Megakaryoblastic, Acute, Leukemia, Monocytic, Acute, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Leukemia, Myeloid, Accelerated Phase, Leukemia, Myeloid, Acute, Leukemia, Myeloid, Chronic, Atypical, BCR-ABL Negative, Leukemia, Myeloid, Chronic-Phase, Leukemia, Myelomonocytic, Acute, Leukemia, Myelomonocytic, Chronic, Leukemia, Myelomonocytic, Juvenile, Leukemia, Plasma Cell, Leukemia, Prolymphocytic, Leukemia, Prolymphocytic, T-Cell, Leukemia, Promyelocytic, Acute, Lipoma, Liposarcoma, Liposarcoma, Myxoid, Liver Neoplasms, Lung Injury, Lung Neoplasms, Lupus Erythematosus, Systemic, Lupus Nephritis, Lymphangioleiomyomatosis, Lymphangioma, Lymphangiomyoma, Lymphatic Abnormalities, Lymphoma, Lymphoma, B-Cell, Lymphoma, B-Cell, Marginal Zone, Lymphoma, Follicular, Lymphoma, Large B-Cell, Diffuse, Lymphoma, Large-Cell, Anaplastic, Lymphoma, Large-Cell, Immunoblastic, Lymphoma, Mantle-Cell, Lymphoma, Non-Hodgkin, Lymphoma, T-Cell, Lymphoma, T-Cell, Cutaneous, Lymphoma, T-Cell, Peripheral, Lymphoproliferative Disorders, Macular Degeneration, Macular Edema, Malignant Carcinoid Syndrome, Melanoma, Metabolic Diseases, Mixed Tumor, Mullerian, Motor Neuron Disease, Mouth Neoplasms, Multiple Myeloma, Multiple Sclerosis, Multiple Sclerosis, Relapsing-Remitting, Multiple System Atrophy, Myelodysplastic Syndromes, Myelodysplastic-Myeloproliferative Diseases, Myeloproliferative Disorders, Myocardial Infarction, Myocardial Ischemia, Myoepithelioma, Myoma, Myosarcoma, Myositis, Myositis, Inclusion Body, Nasopharyngeal Carcinoma, Nasopharyngeal Neoplasms, Neoplasm Metastasis, Neoplasm, Residual, Neoplasms, Neoplasms, Germ Cell and Embryonal, Neoplasms, Plasma Cell, Neoplasms, Squamous Cell, Neoplastic Cells, Circulating, Neovascularization, Pathologic, Nephritis, Nerve Sheath Neoplasms, Nervous System Neoplasms, Neural Tube Defects, Neurilemmoma, Neuroblastoma, Neuroectodermal Tumors, Neuroectodermal Tumors, Primitive, Neuroectodermal Tumors, Primitive, Peripheral, Neuroendocrine Tumors, Neurofibroma, Neurofibroma, Plexiform, Neurofibromatoses, Neurofibromatosis 1, Neurofibrosarcoma, Neutropenia, Oligodendroglioma, Osteosarcoma, Ovarian Diseases, Ovarian Neoplasms, Pancreatic Neoplasms, Pancytopenia, Panuveitis, Pars Planitis, Peripheral Arterial Disease, Peripheral Vascular Diseases, Peritoneal Fibrosis, Peritoneal Neoplasms, Pharyngeal Neoplasms, Pica, Pick Disease of the Brain, Pinealoma, Plasmablastic Lymphoma, Plasmacytoma, Pneumonia, Polycystic Kidney Diseases, Polycystic Kidney, Autosomal Dominant, Polyps, Precancerous Conditions, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Precursor T-Cell Lymphoblastic Leukemia-Lymphoma, Preleukemia, Primary Immunodeficiency Diseases, Primary Myelofibrosis, Progeria, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Pulmonary Fibrosis, Pulmonary Valve Stenosis, Purpura, Purpura, Thrombocytopenic, Purpura, Thrombocytopenic, Idiopathic, Rage, Rectal Neoplasms, Recurrence, Red-Cell Aplasia, Pure, Renal Artery Obstruction, Renal Insufficiency, Renal Insufficiency, Chronic, Respiratory Distress Syndrome, Respiratory Distress Syndrome, Newborn, Respiratory Insufficiency, Retinal Diseases, Retroperitoneal Fibrosis, Retroviridae Infections, Rhabdomyosarcoma, Rhabdomyosarcoma, Alveolar, Rhabdomyosarcoma, Embryonal, ST Elevation Myocardial Infarction, Salivary Gland Neoplasms, Sarcoidosis, Sarcoma, Sarcoma, Alveolar Soft Part, Sarcoma, Ewing, Sarcoma, Kaposi, Sarcoma, Myeloid, Sarcoma, Synovial, Scleritis, Scleroderma, Diffuse, Scleroderma, Systemic, Sclerosis, Self-Injurious Behavior, Severe Acute Respiratory Syndrome, Severe Combined Immunodeficiency, Shy-Drager Syndrome, Skin Neoplasms, Small Cell Lung Carcinoma, Somatostatinoma, Squamous Cell Carcinoma of Head and Neck, Stomach Neoplasms, Sturge-Weber Syndrome, Syndrome, Telangiectasia, Hereditary Hemorrhagic, Telangiectasis, Thalassemia,

				Thrombocytopenia, Thrombosis, Thyroid Diseases, Thyroid Neoplasms, Tic Disorders, Tobacco Use, Tobacco Use Disorder, Tongue Neoplasms, Triple Negative Breast Neoplasms, Tuberous Sclerosis, Ureteral Neoplasms, Urethral Neoplasms, Urinary Bladder Neoplasms, Uterine Cervical Neoplasms, Uveitis, Uveitis, Intermediate, Uveitis, Posterior, Vascular Diseases, Vascular Malformations, Vitiligo, Waldenstrom Macroglobulinemia, Wet Macular Degeneration, Wounds and Injuries, alpha-Thalassemia, beta-Thalassemia
nalmefene	OPRK1	80	2	Phase 2: Parkinson Disease, Alcohol Drinking, Alcohol-Induced Disorders, Alcoholism, Conduct Disorder, Dermatitis, Dermatitis, Atopic, Disease, Disruptive, Impulse Control, and Conduct Disorders, Eczema, Fatty Liver, Fatty Liver, Alcoholic, Gambling, Liver Diseases, Non-alcoholic Fatty Liver Disease, Pruritus, Smoking Cessation
Remifentanil	OPRK1	78	1	N/A
Oxycodone	OPRK1	71	5	Phase 3: Parkinson Disease, Ankle Injuries, Ankylosis, Anterior Cruciate Ligament Injuries, Arthritis, Brain Abscess, Calculi, Constipation, Diabetic Neuropathies, Genital Neoplasms, Female, Hallux Valgus, Head and Neck Neoplasms, Hypospadias, Joint Diseases, Kidney Calculi, Lithiasis, Low Back Pain, Mucositis, Multiple Myeloma, Nausea, Neoplasm Metastasis, Neoplasms, Nephrolithiasis, Nervous System Diseases, Neuralgia, Opioid-Induced Constipation, Osteoarthritis, Osteoarthritis, Hip, Osteoarthritis, Knee, Pain, Peripheral Nervous System Diseases, Polyneuropathies, Postoperative Complications, Postoperative Nausea and Vomiting, Pruritus, Restless Legs Syndrome, Sinusitis, Substance Withdrawal Syndrome, Tibial Meniscus Injuries, Vomiting, Wounds and Injuries
Acetylsalicylic acid	IKBKB, PTGS1, FAS, FGF2	70	4	Phase 4: Parkinson Disease, Acute Coronary Syndrome, Aneurysm, Aneurysm, Dissecting, Angina Pectoris, Angina, Unstable, Aortic Valve Disease, Aortic Valve Insufficiency, Aortic Valve Stenosis, Arterial Occlusive Diseases, Arteriosclerosis, Arthritis, Arthritis, Rheumatoid, Asthma, Asthma, Aspirin-Induced, Atherosclerosis, Atrial Fibrillation, Barrett Esophagus, Brain Abscess, Brain Concussion, Brain Infarction, Brain Injuries, Brain Injuries, Traumatic, Brain Ischemia, Brain Neoplasms, Carcinoma, Ovarian Epithelial, Cardiomyopathies, Cardiovascular Diseases, Carotid Artery Diseases, Carotid Stenosis, Cerebral Hemorrhage, Cerebral Infarction, Cerebral Small Vessel Diseases, Chagas Disease, Cognitive Dysfunction, Common Cold, Communicable Diseases, Constriction, Pathologic, Coronary Aneurysm, Coronary Artery Disease, Coronary Disease, Coronary Restenosis, Coronary Stenosis, Coronary Vessel Anomalies, Craniocerebral Trauma, Dementia, Dementia, Vascular, Dental Caries, Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetes Mellitus, Type 2, Diabetic Nephropathies, Disease, Drug-Related Side Effects and Adverse Reactions, Dyslipidemias, Eclampsia, Embolic Stroke, Embolism, Endometriosis, Epilepsy, Erythema, Esophagitis, Esophagitis, Peptic, Essential Hypertension, Fetal Growth Retardation, Fibrosis, Flushing, Foramen Ovale, Patent, Gastroesophageal Reflux, Gastrointestinal Hemorrhage, Glucosephosphate Dehydrogenase Deficiency, Hate, Headache, Heart Diseases, Heart Failure, Heart Valve Diseases, Hematoma, Hematoma, Subdural, Hematoma, Subdural, Chronic, Hematuria, Hemorrhage, Hyperlipidemias, Hypersplenism, Hypertension, Hypertension, Portal, Hypertension, Pregnancy-Induced, Infarction, Infections, Inflammation, Insulin Resistance, Intermittent Claudication, Intracranial Aneurysm, Intracranial Embolism, Intracranial Hemorrhages, Ischemia, Ischemic Attack, Transient, Ischemic Stroke, Kidney Diseases, Kidney Failure, Chronic, Low Back Pain, Lupus Erythematosus, Systemic, Microvascular Angina, Migraine Disorders, Mood Disorders, Mucocutaneous Lymph Node Syndrome, Multiple Sclerosis, Multiple Sclerosis, Relapsing-Remitting, Muscular Diseases, Myocardial Infarction, Myocardial Ischemia, Myocardial Stunning, Myocarditis, Neoplasm Metastasis, Neoplasms, Neural Tube Defects, Opioid-Related Disorders, Ovarian Hyperstimulation Syndrome, Ovarian Neoplasms, Pain, Peptic Ulcer, Pericarditis, Periodontitis, Peripheral Arterial Disease, Peripheral Nervous System Diseases, Peripheral Vascular Diseases, Pharyngitis, Placenta Diseases, Placental Insufficiency, Pre-Eclampsia, Premature Birth, Prostatic Neoplasms, Proteinuria, Pulmonary Embolism, Recurrence, Renal Artery Obstruction, Renal Insufficiency, Renal Insufficiency, Chronic, Respiration Disorders, Respiratory Tract Diseases, Respiratory Tract Infections, Restless Legs Syndrome, ST Elevation Myocardial Infarction, Sarcoma, Scleroderma, Diffuse, Scleroderma, Systemic, Sclerosis, Stomach Neoplasms, Stroke, Stroke, Lacunar, Syndrome, Tachycardia, Tachycardia, Ventricular, Tension-Type Headache, Thromboembolism, Thrombophilia, Thrombosis, Tonsillitis, Ulcer, Vascular Diseases, Venous Thromboembolism, Venous Thrombosis, Ventricular Premature Complexes, Vertebral Artery Dissection, Wounds and Injuries

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, STK38L, MERTK, LATS2, DYRK1A, MAP4K1, NUAK1, CLK4, MAPK12, WEE1, TGFBR2, CLK1, PLK4	94	Phase 2: Arthritis, Arthritis, Rheumatoid, Psoriasis
seliciclib	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, STK38L, MERTK, LATS2, DYRK1A, MAP4K1, NUAK1, CLK4, MAPK12, WEE1, TGFBR2, CLK1, PLK4	93	Phase 2: ACTH-Secreting Pituitary Adenoma, Adenoma, Carcinoma, Non-Small-Cell Lung, Cystic Fibrosis, Cysts, Fibrosis, Pituitary ACTH Hypersecretion, Pituitary Neoplasms
ruboxistaurin	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, STK38L, MERTK, LATS2, DYRK1A, MAP4K1, NUAK1, CLK4, MAPK12, WEE1, TGFBR2, CLK1, PLK4	93	Phase 3: Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetes Mellitus, Type 2, Diabetic Neuropathies, Diabetic Retinopathy, Edema, Macular Edema, Nervous System Diseases, Peripheral Nervous System Diseases, Retinal Diseases
Sorafenib	IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, STK38L, MERTK, LATS2, DYRK1A, MAP4K1, NUAK1, CLK4, MAPK12, WEE1, TGFBR2, CLK1, PLK4	93	Phase 4: Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Renal Cell, Hand-Foot Syndrome, Liver Neoplasms, Neoplasm Metastasis, Neoplasms, Thrombosis
Tofacitinib	MAPK10, RPS6KA3, TEC, ROCK2, TXK, MAP4K4, PRKCQ, TTK, LATS1, SLK, MAP2K4, CDK7, PRKCH, ACVR2A, EIF2AK2, PRKG1, CHEK1, JAK2, MAP3K20, PRKD1, RPS6KB1, RIPK2, STK38L, MERTK, LATS2, MAP4K1, NUAK1, CLK4, MAPK12, WEE1, TGFBR2, CLK1, PLK4	93	Phase 4: Alopecia, Alopecia Areata, Aortic Arch Syndromes, Arteritis, Arthritis, Arthritis, Psoriatic, Arthritis, Rheumatoid, COVID-19, Colitis, Colitis, Ulcerative, Disease, Embolism, Granuloma, Granulomatosis with Polyangiitis, Infections, Lung Diseases, Lung Diseases, Interstitial, Necrosis, Rheumatic Fever, ST Elevation Myocardial Infarction, Spondylarthritis, Spondylitis, Spondylitis, Ankylosing, Systemic Vasculitis, Takayasu Arteritis, Thromboembolism, Ulcer, Vasculitis

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	CDC25A, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTGS1, PTPN12, CDC25C, PTPRK, UBASH3B	94	0.41
Tiludronate	CDC25A, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, ACPP, CDC25C, PTPRK, UBASH3B	92	0.24
Terlipressin	ITGA6, ITGB1, PTGS1, ITGAV, ITGA2, ITGA1	91	0.32
3-(Phosphonomethyl)Pyridine-2-Carboxylic Acid	CDC25A, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, CDC25C, ACPP, PTPRK, UBASH3B	91	0.37
Halcinonide	TLR4, TLR2, HIF1A	91	0.14

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
2,5,7-Trihydroxynaphthoquinone	MAPK10, CDC25A, MAPK12, SENP6, EPM2A, CASP8, DYRK1A, SENP8, BRCA1, HSP90AA1	84	0.36
Lanreotide	ITGA6, ITGB1, ITGAV, ITGA2, ITGA1	83	0.25
Ibudilast	TLR4, TLR2	82	0.17
Thioproline	ITGA6, ITGB1, ITGAV, ITGA2, ITGA1	82	0.17
Apomorphine	MAPK10, MAPK12, OPRK1, HIF1A	81	0.15

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 2,5,7-Trihydroxynaphthoquinone. These drugs were selected for acting on the following targets: ROCK2, CLK4, PTPN2 and SENP8, 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 2,5,7-Trihydroxynaphthoquinone

These drugs were selected for acting on the following targets: ROCK2, CLK4, PTPN2 and SENP8, 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:

CLK4, SNCA, FIGF, SNCA and Osteopontin

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: Tacrolimus, Sodium Tetradecyl Sulfate, D-Myo-Inositol-Hexasulphate, 1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea and 4-Aminophthalhydrazide. 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.

- CLK4
- SNCA
- FIGF
- SNCA
- Osteopontin

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 2022.2 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac).
The master regulator search uses the TRANSPATH® database (BIOBASE), release 2022.2 (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 2022.2 (https://genexplain.com/humanpsd).
The Ensembl database release Human104.38 (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-score$_{PSD}$*),
  2. ranking by "Disease activity score" (*D-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
0. 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
1. 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.
2. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal.* **2006**;50(2):66-75 (russ)
3. 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

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.