

ITGA6 and ITGA1 are promising druggable targets for treating Parkinson Disease that control activity of NR3C1, YY1 and SMAD2 transcription factors on promoters of differentially expressed genes

Demo User

geneXplain GmbH

info@genexplain.com

Data received on 02/07/2020 ; Run on 12/04/2022 ; Report generated on 12/04/2022

Genome Enhancer release 3.0 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2022.1)



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: NR3C1, YY1, ZEB1, SMAD2, TFCP2 and VDR. The subsequent network analysis suggested

- integrins
- tec
- SNCA
- SNCA

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 devised to provide a casual interpretation of the data obtained for a pathology state. This approach comprises two major steps: (1) analysing promoters and enhancers of differentially expressed genes for the transcription factors (TFs) involved in their regulation and, thus, important for the process under study; (2) re-constructing the signaling pathways that activate these TFs and identifying master regulators at the top of such pathways. For the first step, the database TRANSFAC® [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 |

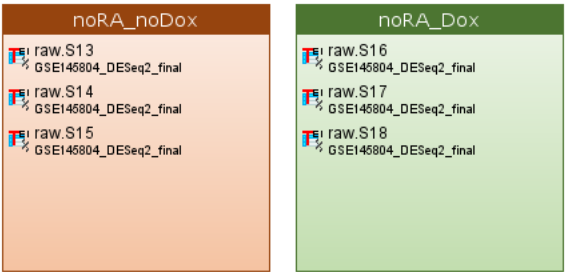


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

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 compared gene expression in "noRA_Dox" with "noRA_noDox". Genes were ranked according to the fold-change and 300 most up-regulated genes (see Table 2) and 300 most downregulated genes (see Table 3) were selected for further analysis.

Table 2. Top ten **up-regulated** genes in noRA_Dox vs. noRA_noDox.

See full table →

| ID | Gene description | Gene symbol | logFC | logCPM | PValue | FDR |
|-----------------|--|-------------|-------|----------|----------|----------|
| ENSG00000145335 | synuclein alpha | SNCA | 4.38 | 10.36 | 2.71E-88 | 4.59E-84 |
| ENSG00000118785 | secreted phosphoprotein 1 | SPP1 | 2.98 | -0.28 | 6.23E-7 | 2.11E-3 |
| ENSG00000169282 | potassium voltage-gated channel subfamily A member regulatory beta subunit 1 | KCNAB1 | 2.91 | 0.49 | 1.32E-7 | 5.61E-4 |
| ENSG00000162692 | vascular cell adhesion molecule 1 | VCAM1 | 2.67 | -4.05E-2 | 1.1E-5 | 2.34E-2 |
| ENSG00000237280 | novel transcript | | 2.15 | 0.2 | 1.42E-4 | 0.16 |
| ENSG00000214892 | USP8 pseudogene 1 | USP8P1 | 1.81 | 0.19 | 4.4E-3 | 0.83 |
| ENSG00000224837 | glycine cleavage system protein H pseudogene 5 | GCSHP5 | 1.72 | -0.1 | 1.84E-3 | 0.6 |
| ENSG00000243300 | | | 1.65 | 0.44 | 2.8E-3 | 0.68 |
| ENSG00000229474 | PAT1 homolog 2 | PATL2 | 1.53 | -0.14 | 7.97E-3 | 0.98 |
| ENSG00000236813 | basic transcription factor 3 pseudogene 8 | BTF3P8 | 1.53 | -0.48 | 9.95E-3 | 0.98 |

Table 3. Top ten **down-regulated** genes in noRA_Dox vs. noRA_noDox.

See full table →

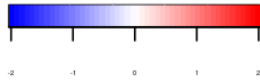
| ID | Gene description | Gene symbol | logFC | logCPM | PValue | FDR |
|-----------------|--|-------------|--------|----------|---------|---------|
| ENSG00000186081 | keratin 5 | KRT5 | -10.36 | 2.57 | 9.35E-7 | 2.64E-3 |
| ENSG00000257594 | polypeptide N-acetylgalactosaminyltransferase 4 | GALNT4 | -2.37 | 1.39E-2 | 4.31E-5 | 6.08E-2 |
| ENSG00000167244 | insulin like growth factor 2 | IGF2 | -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 | solute carrier family 37 member 2 | SLC37A2 | -1.9 | -0.42 | 9.67E-3 | 0.98 |
| ENSG00000111679 | protein tyrosine phosphatase non-receptor type 6 | PTPN6 | -1.9 | -0.13 | 4.43E-3 | 0.83 |
| ENSG00000179846 | NTPase KAP family P-loop domain containing 1 | NKPD1 | -1.87 | -0.33 | 3.5E-3 | 0.79 |
| ENSG00000137801 | thrombospondin 1 | THBS1 | -1.83 | 0.49 | 4.4E-3 | 0.83 |
| ENSG00000269054 | ZNF497 antisense RNA 1 | ZNF497-AS1 | -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

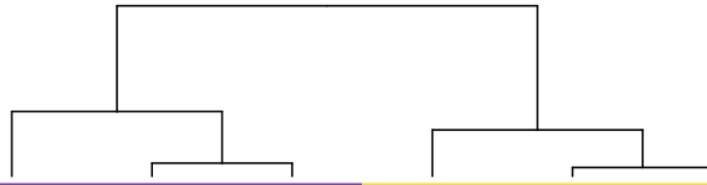
A functional analysis of differentially expressed genes was done by mapping the significant up-regulated and significant down-regulated genes to several known ontologies, such as Gene Ontology (GO), disease ontology (based on HumanPSD™ database) and the ontology of signal transduction and metabolic pathways 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

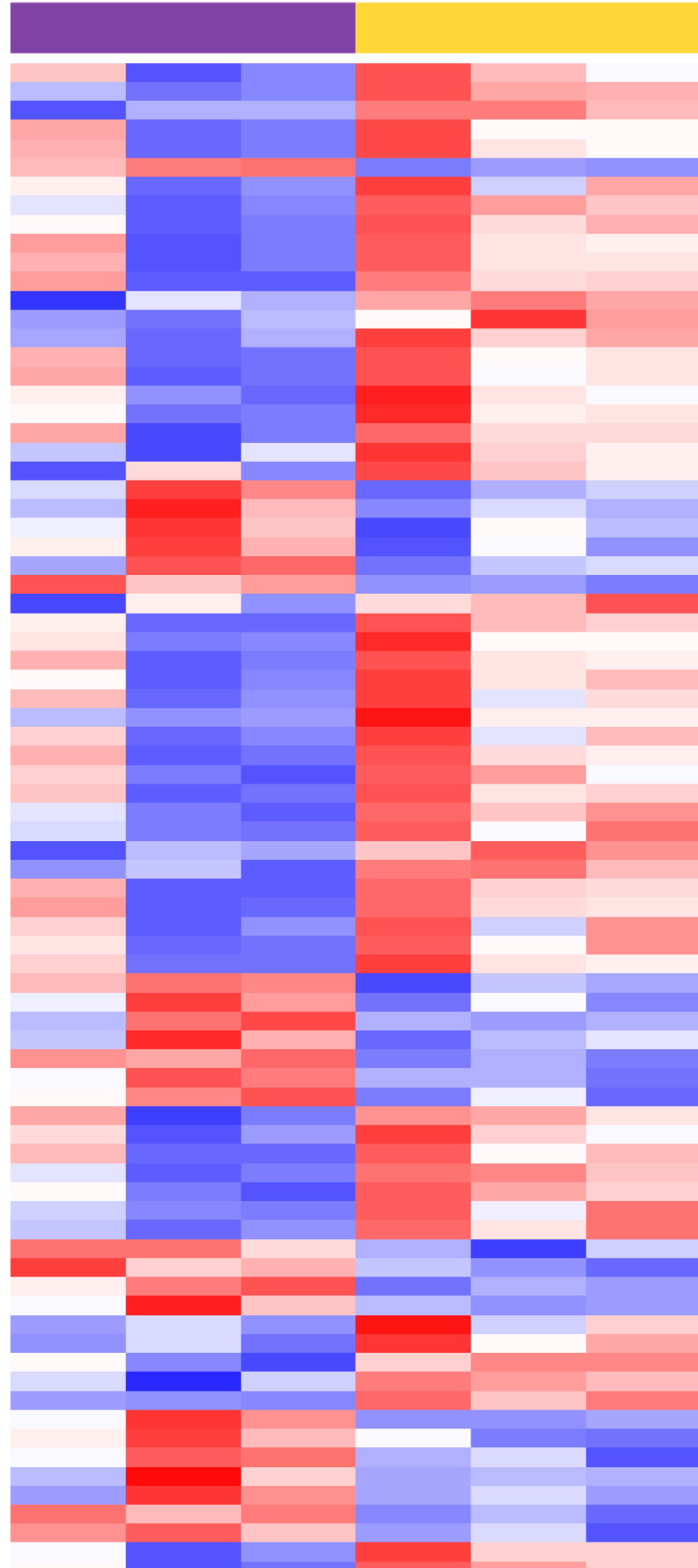
A heatmap of all differentially expressed genes playing a potential regulatory role in the system (enriched in TRANSPATH® pathways) is presented in Figure 2.



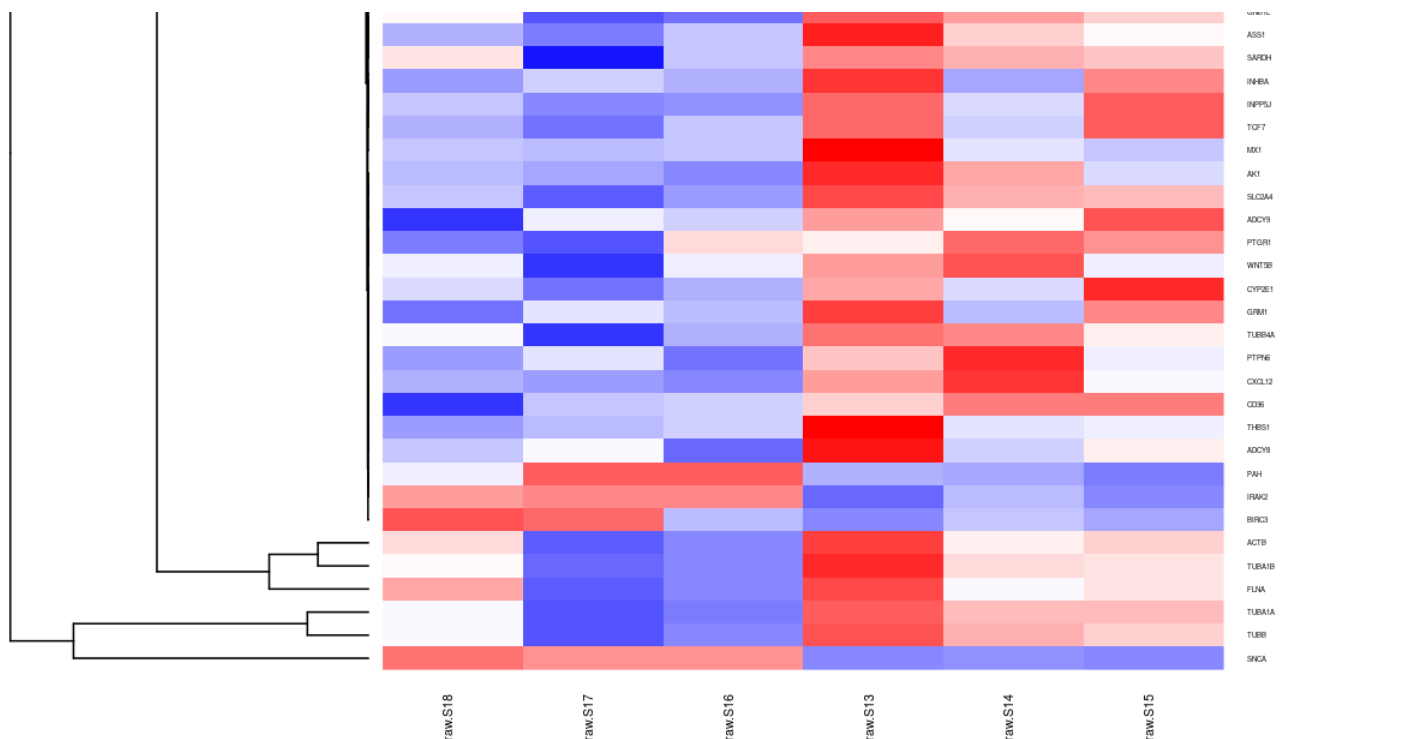
Gene Expression Normalized by rows



noRA noDox
noRA_Dox



APH1A
FBN1
SCD3
TUBB4B
FADS2
TXNIP
IGF1R
PGK1
ACTG1
PFN1
PDGFBR
PPKP
MAPK9
USP34
HPR2
HIC4
MEG3
BIRC5
HIC3
SHGOL1
FZD7
INSIG1
HEXB
OS
GRI1
ADAM9
ATAD2
ADAMTS1
ACR2
AVR1B1
PGAM1
MMP2
HIC3
FASN
DUSP6
HIC3
HIC1
DHX27
RORA
TGFB3
ADCY1
TIAM1
FANCD
MVD
MVK
HIC2
ALDOA
PC
CDC25A
NTSE
ARHGAP10
TEC
COL2
CCH2
PLP1
TAB1
DOK1
GALX1
DFFB
SLF1
CDKOR1
DBP1
IL13RB
RFP1
EFB
BST1
BAGLNT1
MAPT
AHPR
PLK3
EPAS1
AKA
PTGS1
SORBS1
HBBT
CASP8
LAT32
ASCL1
ENG



Up-regulated genes in noRA_Dox vs. noRA_noDox:

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

GO (biological process)



TRANSPATH® Pathways (2022.1)

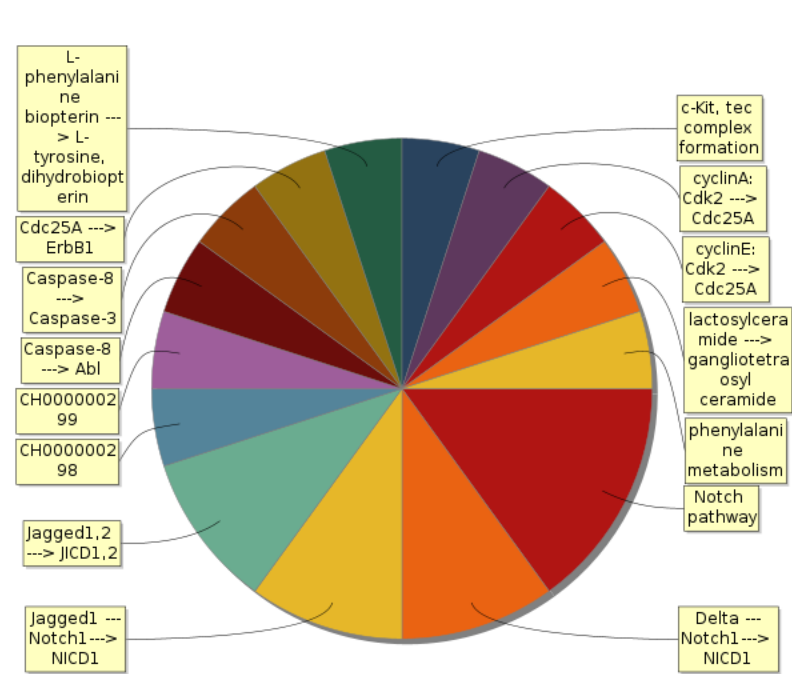


Figure 4. Enriched TRANSPATH® Pathways (2022.1) of up-regulated genes in noRA_Dox vs. noRA_noDox. [Full classification →](#)

HumanPSD(TM) disease (2022.1)

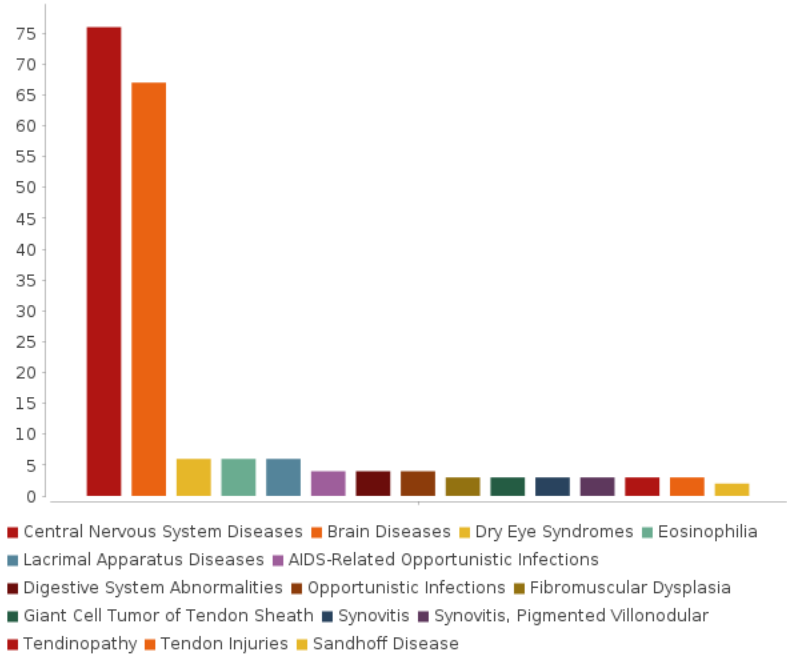


Figure 5. Enriched HumanPSD(TM) disease (2022.1) of up-regulated genes in noRA_Dox vs. noRA_noDox. The size of the bars correspond to the number of biomarkers of the given disease found among the input set. [Full classification →](#)

Down-regulated genes in noRA_Dox vs. noRA_noDox:

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

GO (biological process)

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|---|--|---|--|---|--|---|--|---|--|--|--|---|--|---|--|--|--|--|--|---|--|--|--|--|--|
| regulation of cell development | | | regulation of neurogenesis | | regulation of axonogenesis | | regulation of cell morphogenesis | | axonogenesis | | plasma membrane bounded cell projection morphogenesis | | negative regulation of phosphorylation | | negative regulation of protein phosphorylation | | regulation of developmental process | | regulation of differentiation | | cell junction organization | | synapse organization | | | | | |
| regulation of cell morphogenesis involved in differentiation | | | regulation of anatomical structure morphogenesis | | negative regulation of nervous system | | positive regulation of cell development | | negative regulation of neurogenesis | | cell projection morphogenesis | | cell part morphogenesis | | negative regulation of phosphate metabolic process | | negative regulation of MAPK cascade | | regulation of multicellular organismal development | | positive regulation of developmental process | | cell-substrate junction organization | | cell junction assembly | | cell-substrate junction assembly | |
| regulation of nervous system development | | | regulation of neuron projection development | | development negative regulation of neuron projection | | negative regulation of cell morphogenesis involved in differentiation | | negative regulation of neuron differentiation | | neuron projection morphogenesis | | cellular component morphogenesis | | negative regulation of protein tyrosine phosphorylation process | | negative regulation of protein tyrosine phosphorylation process | | regulation of developmental process | | positive regulation of developmental process | | cell junction organization | | cell junction assembly | | cell-substrate junction assembly | |
| regulation of neuron differentiation | | | positive regulation of axonogenesis | | development positive regulation of nervous system development | | positive regulation of axon | | negative regulation of cell projection organization | | axon development | | neuron projection development | | neurogenesis | | generation of neurons | | nervous system development | | system development | | multicellular organism development | | multicellular organism development | | multicellular organism development | |
| regulation of cell morphogenesis involved in differentiation | | | positive regulation of cell morphogenesis involved in differentiation | | positive regulation of dendrite development | | extension regulation of dendrite | | positive regulation of neurogenesis | | cell projection organization | | plasma membrane bounded cell projection organization | | dendrite development | | embryonic morphogenesis | | gastrulation | | anatomical structure development | | anatomical structure morphogenesis | | negative regulation of nitrogen compound metabolic process | | negative regulation of nitrogen compound metabolic process | |
| cholesterol biosynthetic process | | | secondary alcohol biosynthetic process | | sterol biosynthetic process | | protein localization to membrane | | SRP-dependent cotranslational protein targeting to membrane | | cell projection organization | | plasma membrane bounded cell projection organization | | dendrite development | | embryonic morphogenesis | | gastrulation | | anatomical structure development | | anatomical structure morphogenesis | | negative regulation of nitrogen compound metabolic process | | negative regulation of nitrogen compound metabolic process | |
| cholesterol biosynthetic process via desmosterol | | | cholesterol biosynthetic process via lathosterol | | alcohol biosynthetic process | | cotranslational protein targeting to membrane | | protein localization to endoplasmic reticulum | | cell morphogenesis involved in differentiation | | cell morphogenesis | | cell morphogenesis | | embryonic organ morphogenesis | | inner ear morphogenesis | | developmental process | | neuron differentiation | | cellular component organization | | cellular component organization | |
| cholesterol metabolic process | | | organic hydroxy compound metabolic process | | alcohol metabolic process | | protein targeting to membrane | | protein targeting to ER | | cell morphogenesis involved in neuron differentiation | | cell morphogenesis involved in differentiation | | cell morphogenesis involved in differentiation | | embryonic organ morphogenesis | | neuron development | | developmental process | | neuron differentiation | | cellular component organization or biogenesis | | regulation of multicellular organismal process | |
| secondary alcohol metabolic process | | | phenol-containing compound metabolic process | | norepinephrine metabolic process | | establishment of protein localization to endoplasmic reticulum | | cellular protein localization | | multicellular organismal homeostasis | | energy homeostasis | | renal system process | | cell differentiation | | cellular developmental process | | negative regulation of biological process | | cellular component organization or biogenesis | | regulation of multicellular organismal process | | regulation of biological process | |
| sterol metabolic process | | | organic hydroxy compound biosynthetic process | | small molecule biosynthetic process | | protein localization to membrane | | protein localization | | multicellular organismal homeostasis | | energy homeostasis | | renal system process | | cell differentiation | | cellular developmental process | | negative regulation of biological process | | regulation of biological quality | | regulation of biological quality | | regulation of metabolic process | |

Full classification →

p53 ---> p21Cip1
 p53 ---> PUMA
 glucose ---> pyruvate
 glycolysis
 estradiol ---> PRLR
 VDR network
 VDR ---> RXR-alpha ---> transcriptional activation
 tubulin ---> restin, IQGAP1 ---> actin
 Rac1 pathway
 acetyl-CoA, malonyl-CoA ---> lignoceric acid
 LXR network
 release of active TGFbeta
 cAMP ---> PPAR-alpha
 biosynthesis of saturated and n - 9 series of MUFA and PUFA
 pyruvate ---> D-glucose

Full classification →

HumanPSD(TM) disease (2022.1)

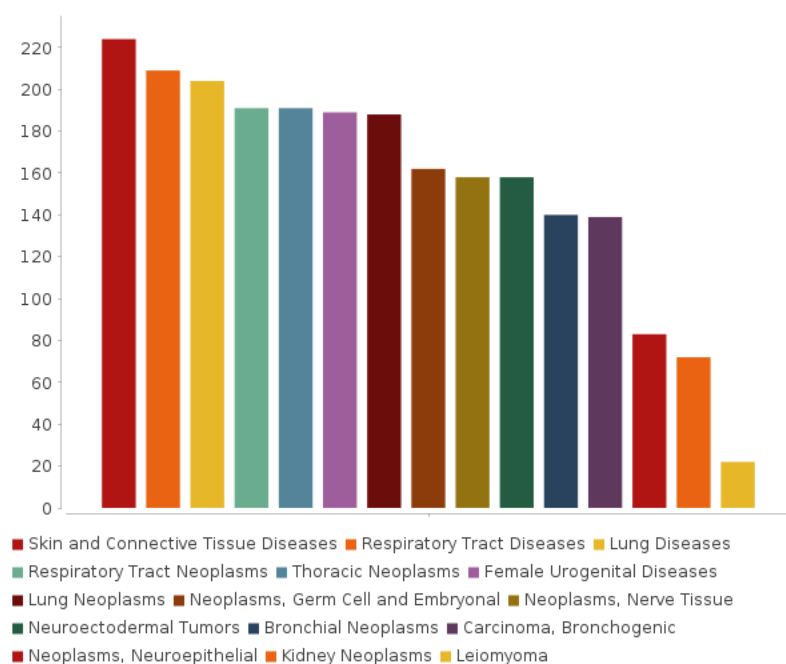
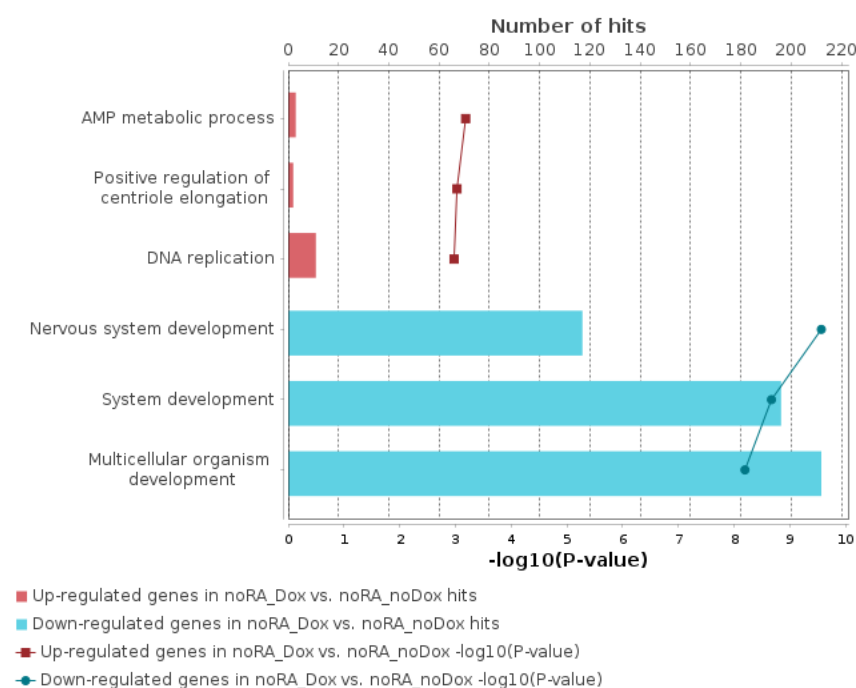


Figure 8. Enriched HumanPSD(TM) disease (2022.1) of down-regulated genes in noRA_Dox vs. noRA_noDox. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

[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].

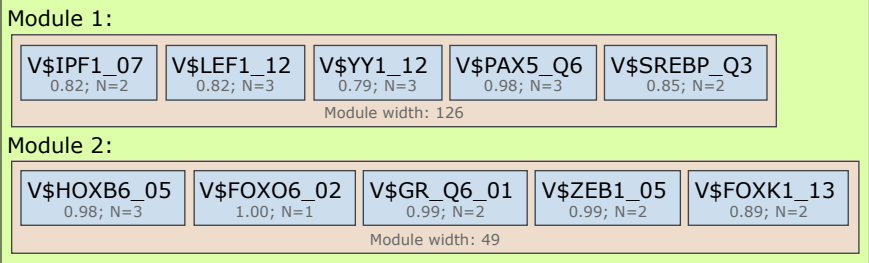
We applied the Composite Module Analyst (CMA) [8] method to detect such potential enhancers, as targets of multiple TFs bound in a cooperative manner to the regulatory regions of the genes of interest. CMA applies a genetic algorithm to construct a generalized model of the enhancers by specifying combinations 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.



Model score (-p*log10(pval)): 15.37
Wilcoxon p-value (pval): 2.20e-31
Penalty (p): 0.501
Average yes-set score: 6.30
Average no-set score: 4.77
AUC: 0.75
Separation point: 5.43
False-positive: 33.67%
False-negative: 23.65%

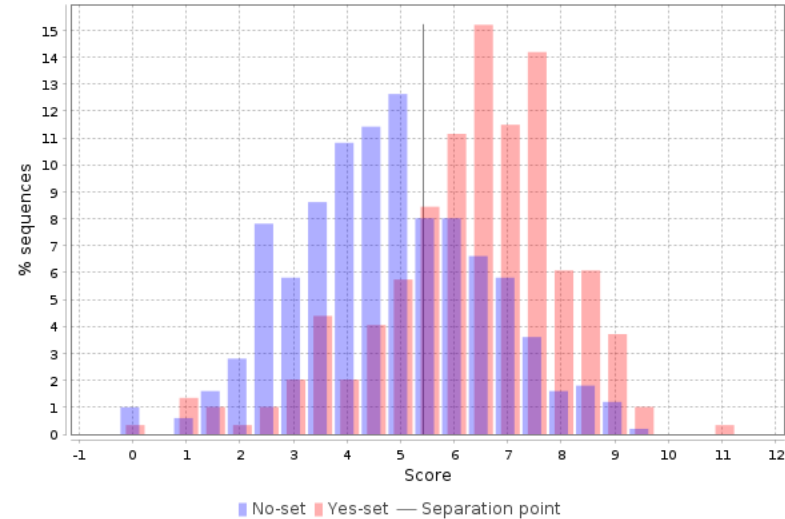


Table 4. List of top ten up-regulated genes in noRA_Dox vs. noRA_noDox with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.

[See full table →](#)

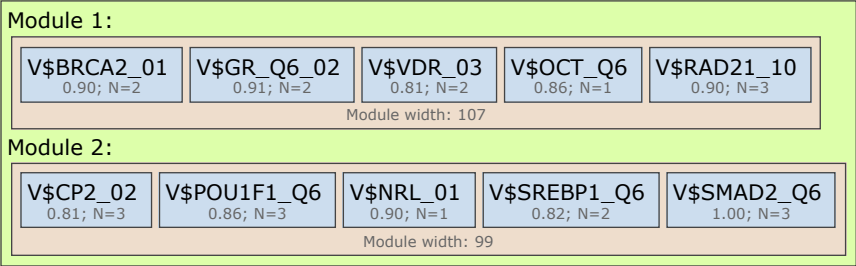
| Ensembl IDs | Gene symbol | Gene description | CMA score | Factor names |
|-----------------|-------------|---|-----------|--|
| ENSG00000198695 | MT-ND6 | mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6 | 11.47 | LEF-1(h), PDX-1(h), SREBP-1(h),SREBP-2(h), YY1(h), Pax-5(h), FOXK1(h), GR(h)... |
| ENSG00000225093 | RPL3P7 | ribosomal protein L3 pseudogene 7 | 11.28 | SREBP-1(h),SREBP-2(h), YY1(h), LEF-1(h), PDX-1(h), Pax-5(h), FOXK1(h), FOXO6(h)... |
| ENSG00000271383 | NBPF19 | NBPF member 19 | 10.99 | Hox-B6(h), FOXK1(h), LEF-1(h), PDX-1(h), SREBP-1(h),SREBP-2(h), YY1(h) |
| ENSG00000261101 | | novel transcript, overlapping to ARMCX3 | 10.98 | LEF-1(h), YY1(h), Pax-5(h), PDX-1(h), SREBP-1(h),SREBP-2(h), FOXK1(h), GR(h)... |
| ENSG00000210195 | MT-TT | mitochondrially encoded tRNA-Thr (ACN) | 10.97 | LEF-1(h), PDX-1(h), SREBP-1(h),SREBP-2(h), YY1(h), Pax-5(h), FOXK1(h), GR(h)... |
| ENSG00000200494 | Y_RNA | Y RNA | 10.79 | Pax-5(h), PDX-1(h), FOXO6(h), LEF-1(h), YY1(h), GR(h), FOXK1(h)... |
| ENSG00000145780 | FEM1C | fem-1 homolog C | 10.77 | LEF-1(h), SREBP-1(h),SREBP-2(h), YY1(h), PDX-1(h), FOXK1(h), Hox-B6(h), GR(h) |
| ENSG00000242457 | RBBP4P2 | RBBP4 pseudogene 2 | 10.69 | YY1(h), LEF-1(h), Pax-5(h), PDX-1(h), SREBP-1(h),SREBP-2(h), FOXK1(h), GR(h) |
| ENSG00000214223 | HNRNPA1P10 | heterogeneous nuclear ribonucleoprotein A1 pseudogene 10 | 10.66 | LEF-1(h), Pax-5(h), SREBP-1(h),SREBP-2(h), PDX-1(h), GR(h), YY1(h), FOXK1(h) |
| ENSG00000259045 | MTCO1P2 | MT-CO1 pseudogene 2 | 10.58 | YY1(h), PDX-1(h), Pax-5(h), LEF-1(h), SREBP-1(h),SREBP-2(h), FOXK1(h), Hox-B6(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.



Model score (-p*log10(pval)): 17.59
Wilcoxon p-value (pval): 7.93e-36
Penalty (p): 0.501
Average yes-set score: 10.08
Average no-set score: 8.56
AUC: 0.76
Separation point: 9.72
False-positive: 20.24%
False-negative: 37.29%

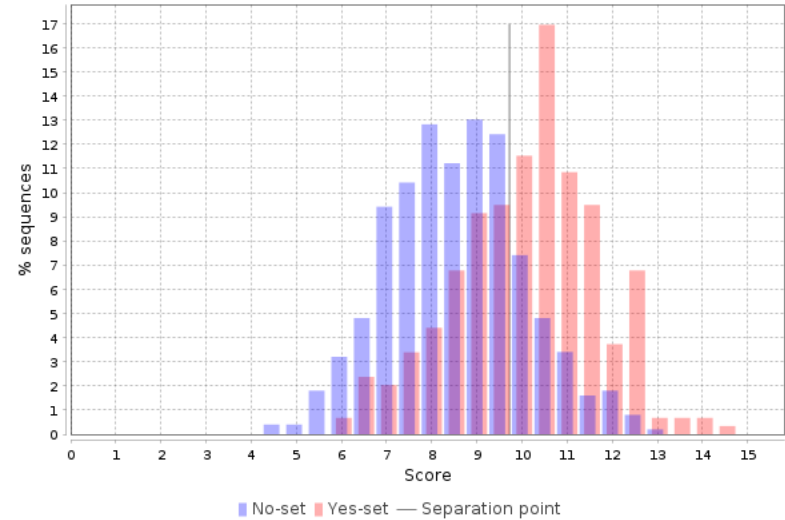


Table 5. List of top ten down-regulated genes in noRA_Dox vs. noRA_noDox with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.
[See full table →](#)

| Ensembl IDs | Gene symbol | Gene description | CMA score | Factor names |
|-----------------|-------------|---|-----------|--|
| ENSG00000142208 | AKT1 | AKT serine/threonine kinase 1 | 14.93 | VDR(h), CP2(h), SMAD2(h), SREBP-1(h), BRCA2(h), Nrl(h), hHR21(h)... |
| ENSG00000105732 | ZNF574 | zinc finger protein 574 | 14.7 | SREBP-1(h), SMAD2(h), CP2(h), Nrl(h), VDR(h), GR(h), BRCA2(h)... |
| ENSG00000134594 | RAB33A | RAB33A, member RAS oncogene family | 14.63 | SREBP-1(h), CP2(h), Nrl(h), SMAD2(h), hHR21(h), GR(h), VDR(h)... |
| ENSG00000107819 | SFXN3 | sideroflexin 3 | 14.59 | VDR(h), SMAD2(h), POU1F1(h), hHR21(h), BRCA2(h), GR(h), Octa-factor(h),POU2AF1(h),POU2F1(h),POU2F2(h),POU2F3(h),POU3F1(h),POU3F2(h),POU3F3(h),POU3B2(h),oct-B3(h)... |
| ENSG00000203392 | | novel transcript, antisense to CSPG4 | 14.43 | VDR(h), hHR21(h), BRCA2(h), Nrl(h), GR(h), SMAD2(h), SREBP-1(h)... |
| ENSG00000110651 | CD81 | CD81 molecule | 14.36 | VDR(h), hHR21(h), GR(h), BRCA2(h), CP2(h), SMAD2(h), SREBP-1(h)... |
| ENSG00000141002 | TCF25 | transcription factor 25 | 14.27 | hHR21(h), SMAD2(h), VDR(h), GR(h), SREBP-1(h), CP2(h), BRCA2(h)... |
| ENSG00000172346 | CSDC2 | cold shock domain containing C2 | 14.25 | Octa-factor(h),POU2AF1(h),POU2F1(h),POU2F2(h),POU2F3(h),POU3F1(h),POU3F2(h),POU3F3(h),POU3B2(h),oct-B3(h), SMAD2(h), CP2(h), SREBP-1(h), Nrl(h), BRCA2(h), VDR(h)... |
| ENSG00000105339 | DENND3 | DENN domain containing 3 | 14.19 | SREBP-1(h), CP2(h), Nrl(h), SMAD2(h), hHR21(h), GR(h), BRCA2(h)... |
| ENSG00000225746 | MEG8 | maternally expressed 8, small nucleolar RNA host gene | 14.17 | POU1F1(h), SREBP-1(h), CP2(h), Nrl(h), Octa-factor(h),POU2AF1(h),POU2F1(h),POU2F2(h),POU2F3(h),POU3F1(h),POU3F2(h),POU3F3(h),POU3B2(h),oct-B3(h), GR(h), hHR21(h)... |

On the basis of the enhancer models we identified transcription factors potentially regulating the **target genes** of our interest. We found 11 and 23 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 →

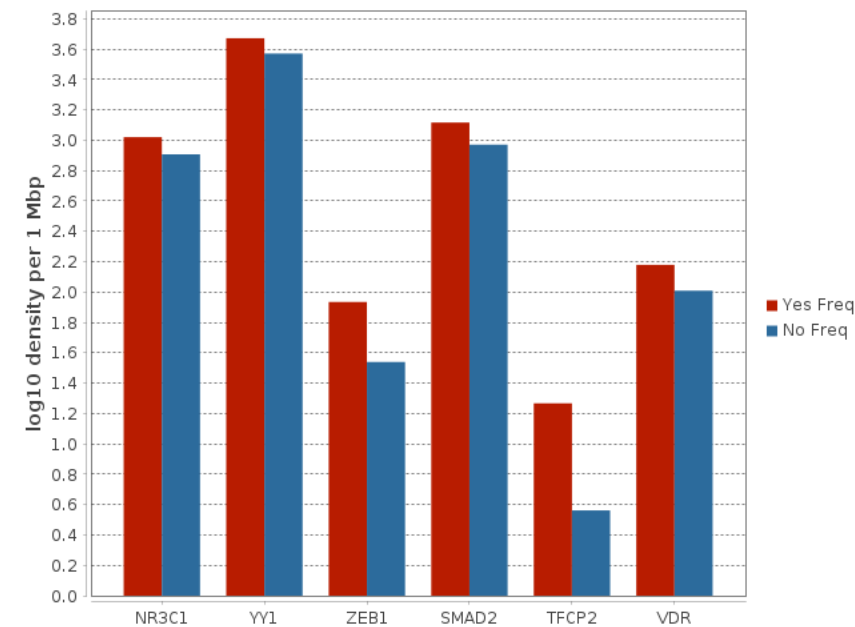
| ID | Gene symbol | Gene description | Regulatory score | Yes-No ratio |
|-------------|-------------|--|------------------|--------------|
| MO000031266 | NR3C1 | nuclear receptor subfamily 3 group C member 1 | 2.31 | 1.3 |
| MO000078913 | YY1 | YY1 transcription factor | 2.12 | 1.26 |
| MO000139677 | ZEB1 | zinc finger E-box binding homeobox 1 | 2.06 | 2.48 |
| MO000007664 | PDX1 | pancreatic and duodenal homeobox 1 | 2.05 | 2.48 |
| MO000056029 | SREBF1 | sterol regulatory element binding transcription factor 1 | 1.93 | 3.37 |
| MO000025765 | SREBF2 | sterol regulatory element binding transcription factor 2 | 1.78 | 2.44 |
| MO000159782 | LEF1 | lymphoid enhancer binding factor 1 | 1.68 | 1.51 |
| MO000046117 | FOXK1 | forkhead box K1 | 1.66 | 2.33 |
| MO000024683 | PAX5 | paired box 5 | 1.64 | 1.18 |
| MO000046018 | HOXB6 | homeobox B6 | 1.31 | 2 |

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 |
|-------------|-------------|--|------------------|--------------|
| MO000057829 | SMAD2 | SMAD family member 2 | 3.03 | 1.4 |
| MO000117988 | TFCP2 | transcription factor CP2 | 2.85 | 5.07 |
| MO000021495 | VDR | vitamin D receptor | 2.47 | 1.48 |
| MO000056029 | SREBF1 | sterol regulatory element binding transcription factor 1 | 2.42 | 3.62 |
| MO000031266 | NR3C1 | nuclear receptor subfamily 3 group C member 1 | 2.36 | 4.17 |
| MO000042938 | RAD21 | RAD21 cohesin complex component | 2.34 | 1.33 |
| MO000010120 | NRL | neural retina leucine zipper | 2.3 | 1.24 |
| MO000056618 | POU5F1 | POU class 5 homeobox 1 | 2.27 | 3.62 |
| MO000045487 | BRCA2 | BRCA2 DNA repair associated | 2.17 | 1.32 |
| MO000025003 | POU2F1 | POU class 2 homeobox 1 | 2.04 | 1.92 |

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: NR3C1, YY1, ZEB1, SMAD2, TFCP2 and VDR.



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 | 55 |
| MO000044264 | SNCA(h) | SNCA | synuclein alpha | 4.38 | 56 |
| MO000103362 | SNCA-isoform3(h) | SNCA | synuclein alpha | 4.38 | 57 |
| MO000044265 | SNCA-isoform1(h) | SNCA | synuclein alpha | 4.38 | 58 |
| MO000103359 | SNCA-isoform2(h) | SNCA | synuclein alpha | 4.38 | 59 |
| MO000044272 | SNCA(h){gly} | SNCA | synuclein alpha | 4.38 | 78 |
| MO000032073 | cIAP-2(h) | BIRC3 | baculoviral IAP repeat containing 3 | 0.62 | 84 |
| MO000020219 | Caspase-8(h) | CASP8 | caspase 8 | 0.55 | 87 |
| MO000021036 | Caspase-8(h) | CASP8 | caspase 8 | 0.55 | 119 |
| MO000043060 | (Caspase-8(h))2 | CASP8 | caspase 8 | 0.55 | 122 |

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 |
|-------------|--|------------------------------------|---|-------|------------|
| MO000004685 | ERK4(h) | MAPK4 | mitogen-activated protein kinase 4 | -0.63 | 183 |
| MO000017421 | AKT-1(h) | AKT1 | AKT serine/threonine kinase 1 | -0.41 | 218 |
| MO000165201 | mTOR(h):rictor(h):mLST8(h):SIN1(h):Protor-1(h) | MAPKAP1, MLST8, MTOR, PRR5, RICTOR | MAPK associated protein 1, MTOR associated protein, LST8 homolog, RPTOR independent companion of MTO... | -0.66 | 244 |
| 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 | 259 |
| MO000117489 | MMP15(h) | MMP15 | matrix metalloproteinase 15 | -0.54 | 266 |
| MO000032657 | DUSP8(h) | DUSP8 | dual specificity phosphatase 8 | -0.65 | 273 |
| MO000034546 | AKT-1(h){pT308}{pS473} | AKT1 | AKT serine/threonine kinase 1 | -0.41 | 291 |
| MO000044885 | PP1-alpha(h) | PPP1CA | protein phosphatase 1 catalytic subunit alpha | -0.52 | 292 |
| MO000036746 | Cdk5(h):CDK5R1p25(h) | CDK5, CDK5R1 | cyclin dependent kinase 5, cyclin dependent kinase 5 regulatory subunit 1 | -0.53 | 294 |
| MO000058481 | AKT-1-isoform1(h) | AKT1 | AKT serine/threonine kinase 1 | -0.41 | 295 |

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.

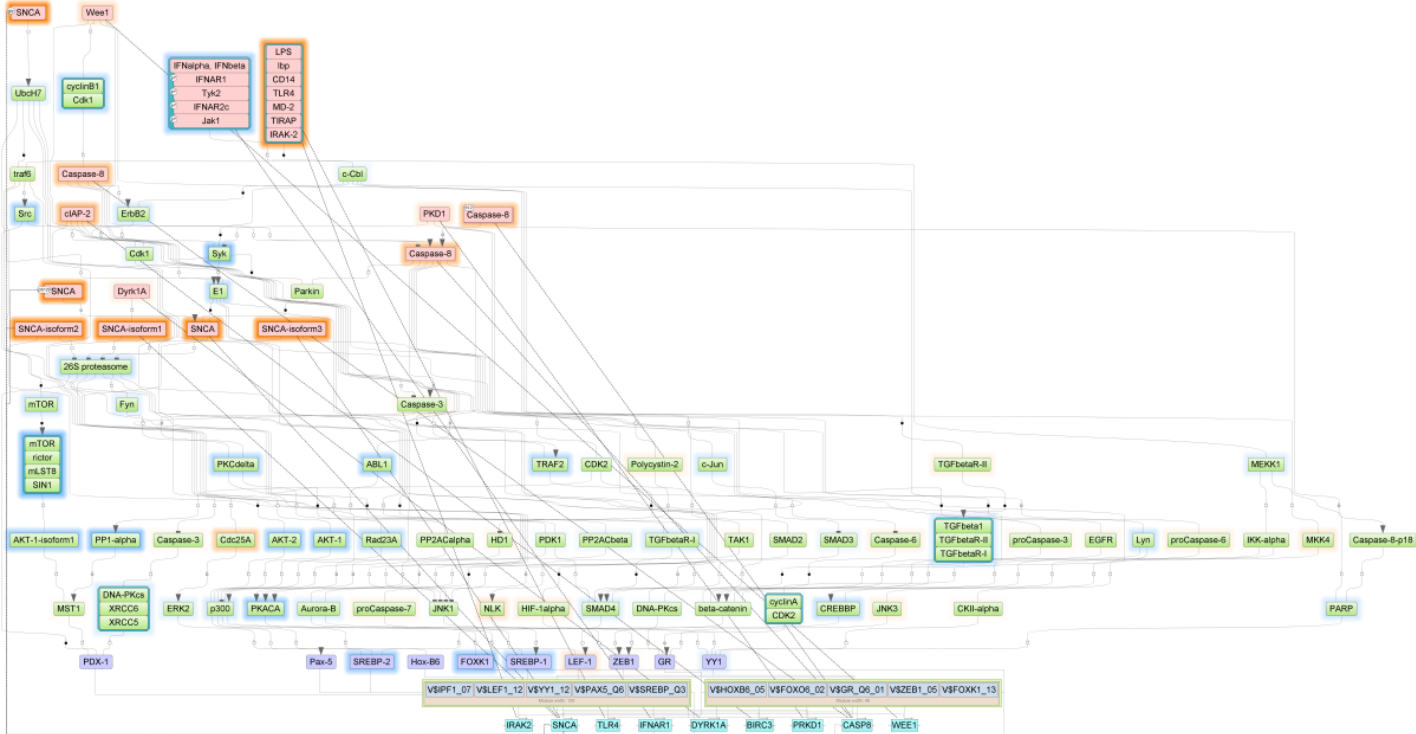



Figure 9. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in *noRA_Dox* vs. *noRA_noDox*. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the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


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. Here, the **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 |
|-------------|--|--------------------|-------|------------|
| ITGA6 | integrin subunit alpha 6 | 1 | 0.22 | 157 |
| TEC | tec protein tyrosine kinase | 28 | 0.52 | 199 |
| CLK4 | CDC like kinase 4 | 26 | 0.38 | 201 |
| PRLR | prolactin receptor | 4 | 0.52 | 231 |
| ROCK2 | Rho associated coiled-coil containing protein kinase 2 | 13 | 0.15 | 243 |
| CLK1 | CDC like kinase 1 | 27 | 0.37 | 244 |

 **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.

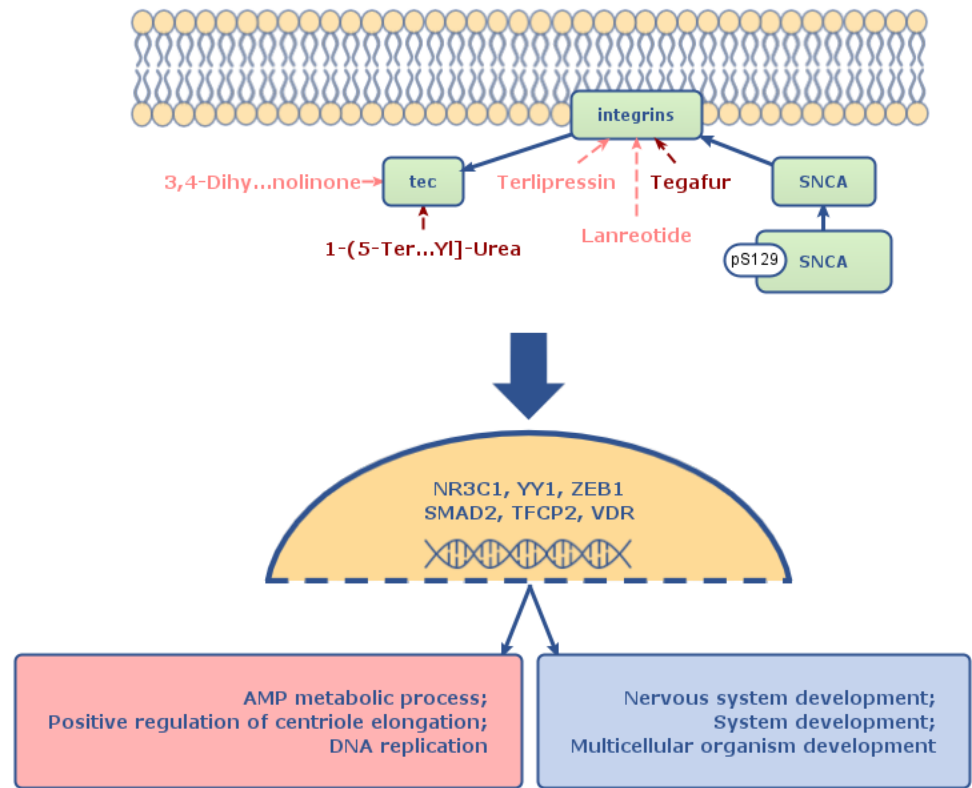
See full table →

| Gene symbol | Gene Description | Druggability score | logFC | Total rank |
|-------------|--|--------------------|-------|------------|
| ITGA6 | integrin subunit alpha 6 | 6.21 | 0.22 | 157 |
| ITGA1 | integrin subunit alpha 1 | 6.21 | 0.22 | 157 |
| CCNH | cyclin H | 2.59 | 0.23 | 168 |
| TEC | tec protein tyrosine kinase | 7.34 | 0.52 | 199 |
| CLK4 | CDC like kinase 4 | 6.44 | 0.38 | 201 |
| ROCK2 | Rho associated coiled-coil containing protein kinase 2 | 1.58 | 0.15 | 243 |

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:

- integrins
- tec
- SNCA
- SNCA

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: Terlipressin, Tegafur, Lanreotide, 1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea and 3,4-Dihydro-5-Methyl-Isoquinolinone, 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 | IKKB, MAPK10, RPS6KA3, ROCK2, TGM2, MAPK12, DYRK1A, CHEK1, FGF2, PRKD1, RPS6KB1 | 86 | 3 | Phase 2: Parkinson Disease, Acute Disease, Acute Lung Injury, Adenocarcinoma, Adenocarcinoma of Lung, Adenoma, Adenomatous Polyposis Coli, Adenomatous Polyps, 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, Angiofibroma, Angiomyolipoma, Angiomyoma, Aphasia, Aphasia, Primary Progressive, Arteriovenous Malformations, Arthrogryposis, Atrophy, Autistic Disorder, Autoimmune Lymphoproliferative Syndrome, Behavior, Blast Crisis, Bone Marrow Failure Disorders, Brain Abscess, Brain Stem Infarctions, Breast Neoplasms, Bronchitis, Burkitt Lymphoma, COVID-19, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Non-Small-Cell Lung, Carcinoma, Renal Cell, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Central Nervous System Neoplasms, Choroidal Neovascularization, Cognitive Dysfunction, Colorectal Neoplasms, Congenital Abnormalities, Constriction, Pathologic, Cysts, Deficiency Diseases, Dementia, Depression, Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetic Retinopathy, Disease, Dry Eye Syndromes, Edema, Epilepsy, Erythema, Eye Diseases, Fanconi Anemia, Fanconi Syndrome, Fibroma, Fibromatosis, Aggressive, Fibrosarcoma, Fibrosis, Frontotemporal Dementia, Gangliosidoses, Gangliosidosis, GM1, Gaucher Disease, Genetic Diseases, Inborn, Geographic Atrophy, Giant Lymph Node Hyperplasia, Glioblastoma, Glomerulonephritis, Glomerulonephritis, Membranous, Glomerulosclerosis, Focal Segmental, Glycogen Storage Disease Type VI, Graft vs Host Disease, Granuloma, Granulomatous Disease, Chronic, Graves Ophthalmopathy, Growth Disorders, HIV Infections, Hamartoma, Hamartoma Syndrome, Multiple, Head and Neck Neoplasms, Hemangioendothelioma, Hemangioma, Hematologic Diseases, Hematologic Neoplasms, Hemolysis, Hemolysis, Hodgkin Disease, Hyperinsulinism, Hypertension, Hypoglycemia, Immunologic Deficiency Syndromes, Infarction, Inflammatory Bowel Diseases, Intestinal Neoplasms, Intracranial Aneurysm, Kasabach-Merritt Syndrome, Kidney Diseases, Kidney Diseases, Cystic, Kidney Failure, Chronic, Klippel-Trenaunay-Weber Syndrome, Leiomyoma, Leiomyomatosis, 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, Lung Injury, Lung Neoplasms, Lupus Erythematosus, Systemic, Lupus Nephritis, Lymphangioma, Lymphangioma, Lymphangioma, Lymphangioma, 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, Lymphoproliferative Disorders, Macular Degeneration, Macular Edema, Metabolic Diseases, Motor Neuron Disease, Mouth Neoplasms, Multiple Myeloma, Myelodysplastic Syndromes, Myelodysplastic-Myeloproliferative Diseases, Myeloproliferative Disorders, Myoma, Myositis, Myositis, Inclusion Body, Nasopharyngeal Neoplasms, Neoplasm Metastasis, Neoplasm, Residual, Neoplasms, Neoplasms, Germ Cell and Embryonal, Neoplasms, Plasma Cell, Neovascularization, Pathologic, Nephritis, Nerve Sheath Neoplasms, Nervous System Neoplasms, Neural Tube Defects, Neuroblastoma, Neurofibroma, Neurofibroma, Plexiform, Neurofibromatosis, Neurofibromatosis 1, Neurofibrosarcoma, Neutropenia, Pancreatic Neoplasms, Pancytopenia, Peritoneal Fibrosis, 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, Prostatic Neoplasms, Pulmonary Fibrosis, Pulmonary Valve Stenosis, Purpura, Purpura, Thrombocytopenic, Purpura, Thrombocytopenic, Idiopathic, Rectal Neoplasms, Recurrence, Red-Cell Aplasia, Pure, Renal Insufficiency, Respiratory Distress Syndrome, Respiratory Distress Syndrome, Newborn, Respiratory Insufficiency, Retinal Diseases, Retroperitoneal Fibrosis, ST Elevation Myocardial Infarction, Sarcoma, Sarcoma, Kaposi, Sarcoma, Myeloid, Scleritis, Scleroderma, Diffuse, Scleroderma, Systemic, Sclerosis, Self-Injurious Behavior, Severe Combined Immunodeficiency, Skin Neoplasms, Small Cell Lung Carcinoma, Squamous Cell Carcinoma of Head and Neck, Sturge-Weber Syndrome, Syndrome, Thalassemia, Thrombocytopenia, Thyroid Diseases, Thyroid Neoplasms, Tic Disorders, Tobacco Use, Tobacco Use Disorder, Tongue Neoplasms, Tuberous Sclerosis, Urinary Bladder Neoplasms, Vascular Malformations, Waldenstrom Macroglobulinemia, Wounds and Injuries, alpha-Thalassemia, beta-Thalassemia |
| Caffeine | ATR, FGF2 | 64 | 6 | Phase 3: Parkinson Disease, Alzheimer Disease, Apnea, Asymptomatic Diseases, Atrial Fibrillation, Brain Abscess, Cellulitis, Cocaine-Related Disorders, Common Cold, Coronary Artery Disease, Coronary Disease, Diabetes Mellitus, Diabetes Mellitus, Type 2, Disorders of Excessive Somnolence, Epilepsy, Fatty Liver, Fatty Liver, Alcoholic, Headache, Heart Diseases, Hypoxia, Intermittent Claudication, Ischemia, Liver Diseases, Low Back Pain, Memory Disorders, Migraine Disorders, Migraine with Aura, Migraine without Aura, Multiple Sclerosis, Multiple Sclerosis, Relapsing-Remitting, Myocardial Ischemia, Non-alcoholic Fatty Liver Disease, Obesity, Pain, Premature Birth, Sclerosis, Sleepiness, Tension-Type Headache, Tooth Diseases |
| Acetylsalicylic acid | IKKB, FAS, FGF2 | 53 | 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, Infarction, Infections, Inflammation, Insulin Resistance, Intermittent Claudication, Intracranial Aneurysm, Intracranial Embolism, Intracranial Hemorrhages, Ischemia, Ischemic Attack, Transient, Ischemic Stroke, Kidney Diseases, Kidney Failure, Chronic, 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 |
| Cholecalciferol | FAS | 43 | 7 | Phase 4: Parkinson Disease, Affect, Amenorrhea, Anemia, Anemia, Aplastic, Anemia, Sickle Cell, Arthritis, Asthma, Avitaminosis, Benign Paroxysmal Positional Vertigo, Bone Demineralization, Pathologic, Bone Diseases, Bone Diseases, Metabolic, Brain Abscess, Bronchiectasis, Bronchiolitis, COVID-19, Calcinosis, Calcium Metabolism Disorders, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Transitional Cell, Cardio-Renal Syndrome, Cardiovascular Diseases, Colitis, Colles' Fracture, Communicable Diseases, Critical Illness, Crohn Disease, Cystic Fibrosis, Cysts, Depression, Depressive Disorder, Dermatitis, Dermatitis, Atopic, Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetes Mellitus, Type 2, Diabetic Nephropathies, Dizziness, Dysbiosis, Eclampsia, Eczema, Endometriosis, Enterocolitis, Enterocolitis, Necrotizing, Epilepsy, Fatty Liver, Fatty Liver, Alcoholic, Fibromyalgia, Fibrosis, Food Hypersensitivity, Fractures, Bone, Glucose Intolerance, HELLP Syndrome, Headache, Heart Failure, Hepatitis, Hepatitis A, Hepatitis B, Hyaline Membrane Disease, Hyperparathyroidism, Hyperparathyroidism, Primary, Hyperparathyroidism, Secondary, Hypersensitivity, Hypertension, Hypertension, Pregnancy-Induced, Hypocalcemia, Hypogonadism, Infarction, Infections, Inflammation, Inflammatory Bowel Diseases, Insulin Resistance, Irritable Bowel Syndrome, Kidney Diseases, Kidney Failure, Chronic, Leukemia, Leukemia, Lymphoid, Liver Diseases, Lung Diseases, Lung Diseases, Obstructive, Lymphoma, Magnesium Deficiency, Malnutrition, Metabolic Diseases, Mood Disorders, Multiple Sclerosis, Multiple Sclerosis, Relapsing-Remitting, Myelodysplastic Syndromes, Myocardial Infarction, Myofascial Pain Syndromes, Neoplasm Metastasis, Neoplasms, Nephrosis, Nephrotic Syndrome, Non-alcoholic Fatty Liver Disease, Obesity, Opioid-Related Disorders, Osteoarthritis, Osteoarthritis, Knee, Osteoporosis, Osteoporosis, Postmenopausal, Pain, Pancreatitis, Pancreatitis, Chronic, Peripheral Arterial Disease, Peripheral Vascular Diseases, Pneumonia, Polycystic Ovary Syndrome, Pre-Eclampsia, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Prediabetic State, Premature Birth, Proteinuria, Psychotic Disorders, Pulmonary Disease, Chronic Obstructive, Recurrence, Renal Insufficiency, Renal Insufficiency, Chronic, Respiratory Tract Infections, Rhinitis, Rhinitis, Allergic, Rhinitis, Allergic, Seasonal, Rickets, Schizophrenia, Sclerosis, Seasonal Affective Disorder, Stroke, Syndrome, Thalassemia, Tic Disorders, Toxemia, Tuberculosis, Tuberculosis, Pulmonary, Vascular Calcification, Vascular Diseases, Vertigo, Virus Diseases, Vitamin D Deficiency, alpha-Thalassemia, beta-Thalassemia |
| Melatonin | PRLR | 37 | 9 | Phase 4: Parkinson Disease, Arteriosclerosis, Attention Deficit Disorder with Hyperactivity, Autism Spectrum Disorder, Autistic Disorder, Behavior, Bipolar Disorder, Calcinosis, Carotid Stenosis, Child Development Disorders, Pervasive, Chronic Periodontitis, Colitis, Consciousness Disorders, Constriction, Pathologic, Coronary Artery Disease, Cysts, Delirium, Dementia, Disease, Dysomnias, Emergence Delirium, Endotoxemia, Enterocolitis, Enterocolitis, Necrotizing, Epilepsy, Epilepsy, Generalized, Fetal Growth Retardation, Fever, Herpes Genitalis, Hyperkinesia, Hyperthermia, Hypokinesia, Infertility, Insulin Resistance, Ischemic Stroke, Kinesia, Lennox Gastaut Syndrome, Liver Diseases, Mental Disorders, Metabolic Side Effects of Drugs and Substances, Mood Disorders, Osteoporosis, Osteoporosis, Postmenopausal, Parasomnias, Periodontal Diseases, Periodontitis, Polycystic Ovary Syndrome, Premature Birth, Psychomotor Agitation, Psychotic Disorders, REM Sleep Behavior Disorder, Schizophrenia, Seizures, Seizures, Febrile, Sleep Disorders, Circadian Rhythm, Sleep Initiation and Maintenance Disorders, Sleep Wake Disorders, Stroke, Syndrome, Tic Disorders, Vascular System Injuries, 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, PRKCQ, TTK, SLK, MAP2K4, CDK7, EIF2AK2, PRKG1, CHEK1, JAK2, PRKD1, RPS6KB1, RIPK2, MERTK, LATS2, DYRK1A, NUA1, CLK4, MAPK12, WEE1, CLK1, PLK4 | 94 | N/A |
| Sorafenib | IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, PRKCQ, TTK, SLK, MAP2K4, CDK7, EIF2AK2, PRKG1, CHEK1, JAK2, PRKD1, RPS6KB1, RIPK2, MERTK, LATS2, DYRK1A, NUA1, CLK4, MAPK12, WEE1, CLK1, PLK4 | 94 | Phase 4: Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Renal Cell, Hand-Foot Syndrome, Liver Neoplasms, Neoplasm Metastasis, Neoplasms, Thrombosis |
| seliciclib | IKBKB, MAPK10, RPS6KA3, TEC, ROCK2, PRKCQ, TTK, SLK, MAP2K4, CDK7, EIF2AK2, PRKG1, CHEK1, JAK2, PRKD1, RPS6KB1, RIPK2, MERTK, LATS2, DYRK1A, NUA1, CLK4, MAPK12, WEE1, CLK1, PLK4 | 94 | 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, PRKCQ, TTK, SLK, MAP2K4, CDK7, EIF2AK2, PRKG1, CHEK1, JAK2, PRKD1, RPS6KB1, RIPK2, MERTK, LATS2, DYRK1A, NUA1, CLK4, MAPK12, WEE1, CLK1, PLK4 | 94 | 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 |
| Tofacitinib | MAPK10, RPS6KA3, TEC, ROCK2, RIPK2, MERTK, LATS2, PRKCQ, TTK, NUA1, SLK, MAP2K4, CDK7, CLK4, MAPK12, WEE1, EIF2AK2, PRKG1, CHEK1, JAK2, CLK1, PRKD1, PLK4, RPS6KB1 | 93 | Phase 4: Alopecia, Alopecia Areata, Aortic Arch Syndromes, Arteritis, Arthritis, Arthritis, Psoriatic, Arthritis, Rheumatoid, COVID-19, Colitis, Colitis, Ulcerative, Disease, Granuloma, Granulomatosis with Polyangiitis, Infections, Lung Diseases, Lung Diseases, Interstitial, Necrosis, Rheumatic Fever, ST Elevation Myocardial Infarction, Spondylarthritis, Systemic Vasculitis, Takayasu Arteritis, 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, PTPN12, CDC25C, UBASH3B | 92 | 0.31 |
| Tiludronate | CDC25A, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, ACP, CDC25C, UBASH3B | 90 | 0.2 |
| 3-(Phosphonomethyl)Pyridine-2-Carboxylic Acid | CDC25A, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, CDC25C, ACP, UBASH3B | 89 | 0.31 |
| [[N-(Benzyloxycarbonyl)Amino]Methyl]Phosphate | CDC25A, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, CDC25C, UBASH3B | 88 | 0.21 |
| Terlipressin | ITGA6, ITGB1, ITGAV, ITGA1 | 87 | 0.18 |



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, BRCA1 | 81 | 0.22 |
| Lanreotide | ITGA6, ITGB1, ITGAV, ITGA1 | 80 | 0.16 |
| Thiopropine | ITGA6, ITGB1, ITGAV, ITGA1 | 79 | 0.11 |
| Ibandronate | CDC25A, PTPRO, PTPRJ, EPM2A, PTPN2, PPM1D, PTPN13, PTPN12, CDC25C, ACP, UBASH3B | 77 | 0.26 |
| 3-(4-HYDROXY-3-METHOXYPHENYL)-2-PROPENOIC ACID | MAPK10, TLR4, MAPK12, CASP8, BRCA1 | 77 | 7.52E-2 |

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, TEC, PTPN2 and BRCA1, 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, TEC, PTPN2 and BRCA1, 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:



integrins, tec, SNCA and SNCA

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: Terlipressin, Tegafur, Lanreotide, 1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea and 3,4-Dihydro-5-Methyl-Isoquinolinone. 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.

- integrins
- tec
- SNCA
- SNCA

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 2022.1 (geneXplain GmbH, Wolfenbüttel, Germany) (<https://genexplain.com/transfac>). The master regulator search uses the **TRANSPATH®** database (BIOBASE), release 2022.1 (geneXplain GmbH, Wolfenbüttel, Germany) (<https://genexplain.com/transpath>). A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in **TRANSPATH®**. The information about drugs corresponding to identified drug targets and clinical trials references were extracted from **HumanPSD™** database, release 2022.1 (<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" ($T\text{-score}_{PSD}$),
2. ranking by "Disease activity score" ($D\text{-score}_{PSD}$),
3. ranking by "Clinical validity score".

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

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

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" ($D\text{-score}_{PSD}$):

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

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

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

Method for prediction of pharmaceutical compounds

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

We selected compounds that satisfied the following conditions:

1. Toxicity below a chosen toxicity threshold (defines as Pa , probability to be active as toxic substance).

2. For all predicted pharmacological effects that correspond to a set of user selected disease(s) Pa is greater than a chosen effect threshold.
3. There are at least 2 targets (corresponding to the predicted activity-mechanisms) with predicted Pa greater than a chosen target threshold.

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

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

where $M(s)$ is the set of activity-mechanisms for the given structure (which passed the chosen threshold for activity-mechanisms Pa); $G(m)$ is the set of targets (converted to genes) that corresponds to the given activity-mechanism (m) for the given compound; $pa(m)$ is the probability to be active of the activity-mechanism (m), $IAP(g)$ is the invariant accuracy of prediction for gene from $G(m)$; $optWeight(g)$ is the additional weight multiplier for gene. T is set of all targets related to the compound intersected with input list, $|T|$ is number of elements in T , AT and $|AT|$ are set 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 TRANSCOMP: 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, Glorizova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *ChemInform*. **1999**;39(4):666-670. doi:10.1002/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 - Detailed report. Composite modules and master regulators \(up-regulated genes in noRA_Dox vs. noRA_noDox\).](#)
2. [Supplementary table 2 - Detailed report. Composite modules and master regulators \(down-regulated genes in noRA_Dox vs. noRA_noDox\).](#)
3. [Supplementary table 3 - 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.