

FKBP4 and MAP2K3 are promising druggable targets for treating Neoplasm Metastasis and Osteosarcoma that control activity of NR3C1, TP73 and IKZF1 transcription factors on promoters of differentially expressed genes

Demo User

geneXplain GmbH

info@genexplain.com

Data received on 07/09/2019 ; Run on 10/12/2023 ; Report generated on 10/12/2023

Genome Enhancer release 3.3 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2023.2)



Abstract

In the present study we applied the software package "Genome Enhancer" to a multiomics data set that contains *transcriptomics* and *proteomics* data. The study is done in the context of *Neoplasm Metastasis and Osteosarcoma*. 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, TP73, ELK1, IKZF1 and GTF2I. The subsequent network analysis suggested

- cyclinD1(h):Cdk6
- cyclinD1(h):Cdk6
- FKBP4
- MKK3

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: Temsirolimus, alvespimycin, (CHLOROACETYL)CARBAMIC ACID (3R,4S,5S,5R)-5-METHOXY-4-[(2R,3R)-2-METHYL-3-(3-METHYL-2-BUTENYL)OXIRANYL]-1-OXASPIRO[2.5]OCT-6-YL ESTER and Myo-Inositol.

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
Proteomics	Proteomics
RNAseq	Transcriptomics

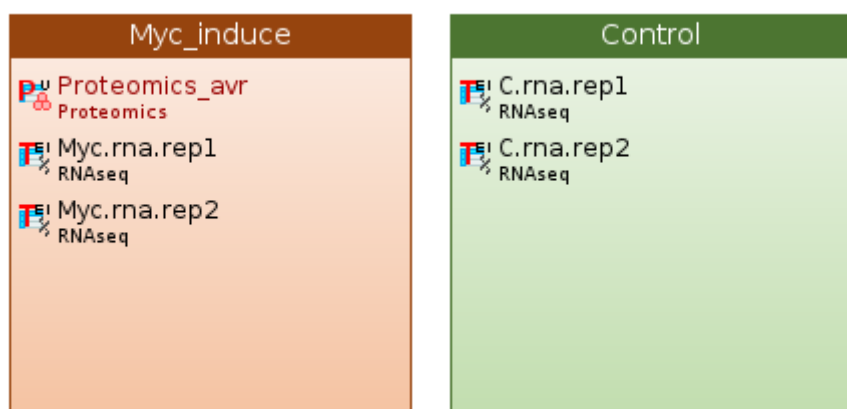


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

3. Results

We have compared the following conditions: Myc_induce *versus* Control.

3.1. Identification of target genes

In the first step of the analysis **target genes** were identified from the uploaded experimental data. We applied the Limma tool (R/Bioconductor package integrated into our pipeline) and compared gene expression in the following sets: "Myc_induce" with "Control". Limma calculated the LogFC (the logarithm to the base 2 of the fold change between different conditions), the p-value and the adjusted p-value (corrected for multiple testing) of the observed fold change. As a result, we detected 408 upregulated genes (LogFC>0.1) out of which 97 genes were found as significantly upregulated (p-value<0.1) and 389 downregulated genes (LogFC<-0.1) out of which 99 genes were significantly downregulated (p-value<0.1). See tables below for the top significant 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 Myc_induce vs. Control.

[See full table](#) →

ID	Gene symbol	Gene description	logFC	P.Value	adj.P.Val
ENSG00000136997	MYC	MYC proto-oncogene, bHLH transcription factor	5.96	9.49E-6	9.85E-3
ENSG00000120738	EGR1	early growth response 1	3.51	6.94E-4	0.17
ENSG00000198576	ARC	activity regulated cytoskeleton associated protein	2.14	1.44E-3	0.17
ENSG00000114767	RRP9	ribosomal RNA processing 9, U3 small nucleolar RNA binding protein	1.29	8.38E-3	0.28
ENSG00000176170	SPHK1	sphingosine kinase 1	1.2	9.2E-3	0.28
ENSG00000052749	RRP12	ribosomal RNA processing 12 homolog	1.19	5.07E-2	0.44
ENSG00000178053	MLF1	myeloid leukemia factor 1	1.12	6.73E-2	0.44
ENSG00000198805	PNP	purine nucleoside phosphorylase	1.11	2.69E-3	0.2
ENSG00000168003	SLC3A2	solute carrier family 3 member 2	1.11	2.45E-3	0.2
ENSG00000132768	DPH2	diphthamide biosynthesis 2	1.01	3.37E-3	0.2

Table 3. Top ten significant **down-regulated** genes in Myc_induce vs. Control.

[See full table](#) →

ID	Gene symbol	Gene description	logFC	P.Value	adj.P.Val
ENSG00000205542	TMSB4X	thymosin beta 4 X-linked	-2.58	2.38E-4	0.12
ENSG00000140416	TPM1	tropomyosin 1	-1.94	1.16E-3	0.17
ENSG00000197747	S100A10	S100 calcium binding protein A10	-1.75	9.6E-4	0.17
ENSG00000205426	KRT81	keratin 81	-1.67	2.21E-3	0.2
ENSG00000092841	MYL6	myosin light chain 6	-1.58	7.88E-4	0.17
ENSG00000100097	LGALS1	galectin 1	-1.55	1.8E-3	0.19
ENSG00000120708	TGFB1	transforming growth factor beta induced	-1.52	1.51E-2	0.35
ENSG00000114353	GNAI2	G protein subunit alpha i2	-1.47	5.43E-3	0.25
ENSG00000005884	ITGA3	integrin subunit alpha 3	-1.41	9.61E-2	0.52
ENSG00000134824	FADS2	fatty acid desaturase 2	-1.4	8.3E-4	0.17

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 Myc_induce vs. Control

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

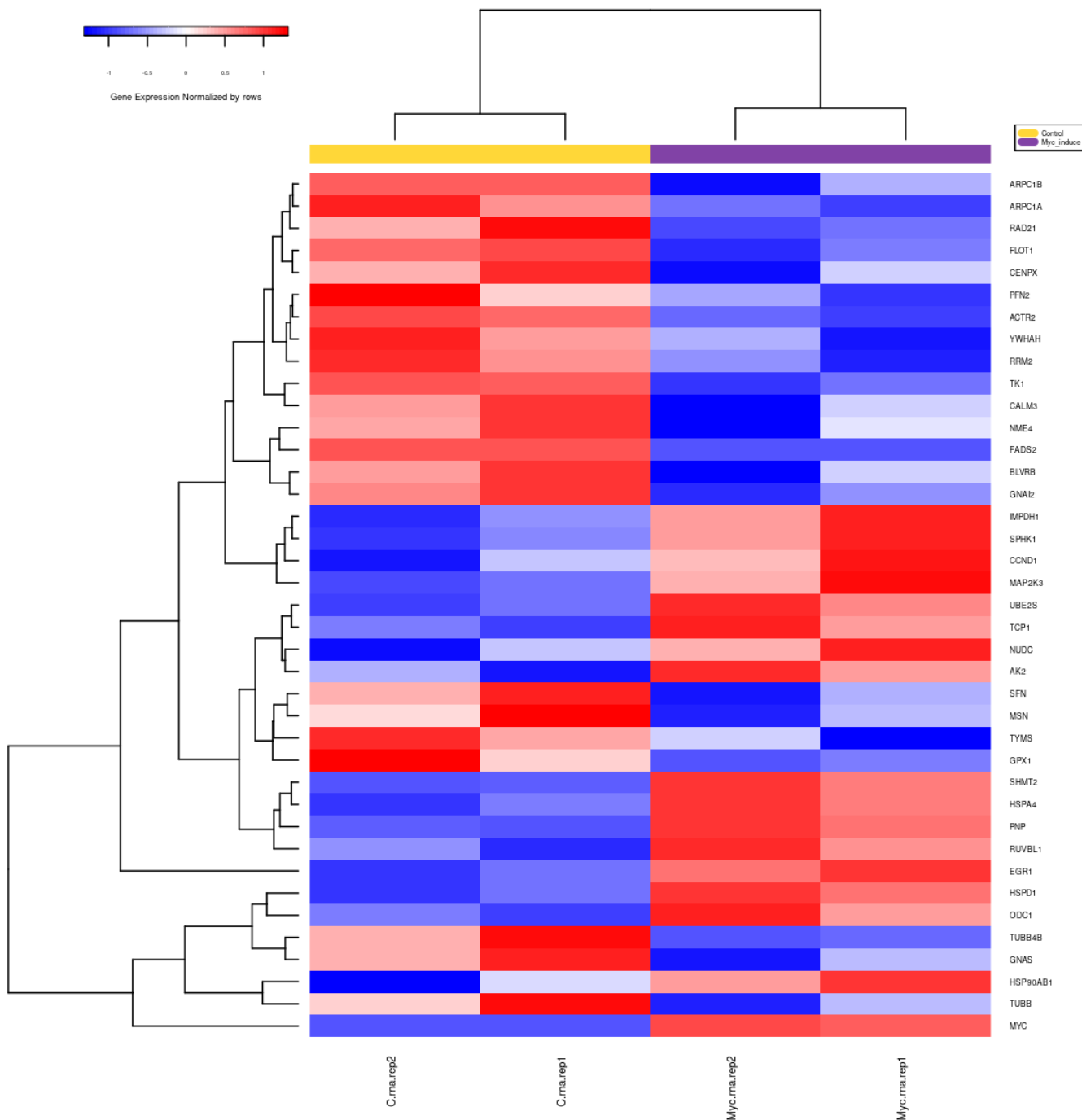


Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.

See full diagram →

Up-regulated genes in Myc_induce vs. Control:

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

GO (biological process)

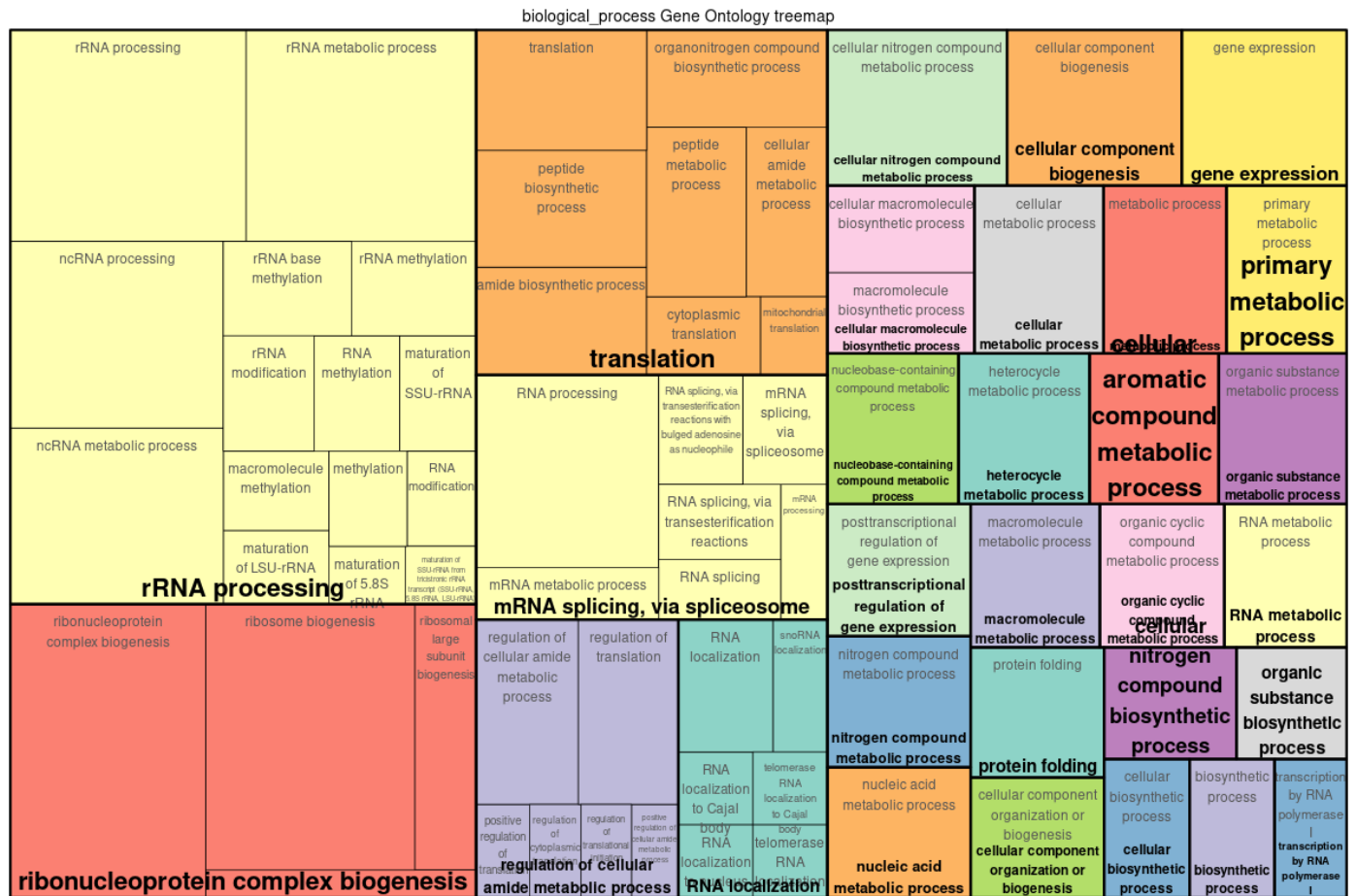


Figure 3. Enriched GO (biological process) of up-regulated genes in Myc_induce vs. Control.

Full classification →

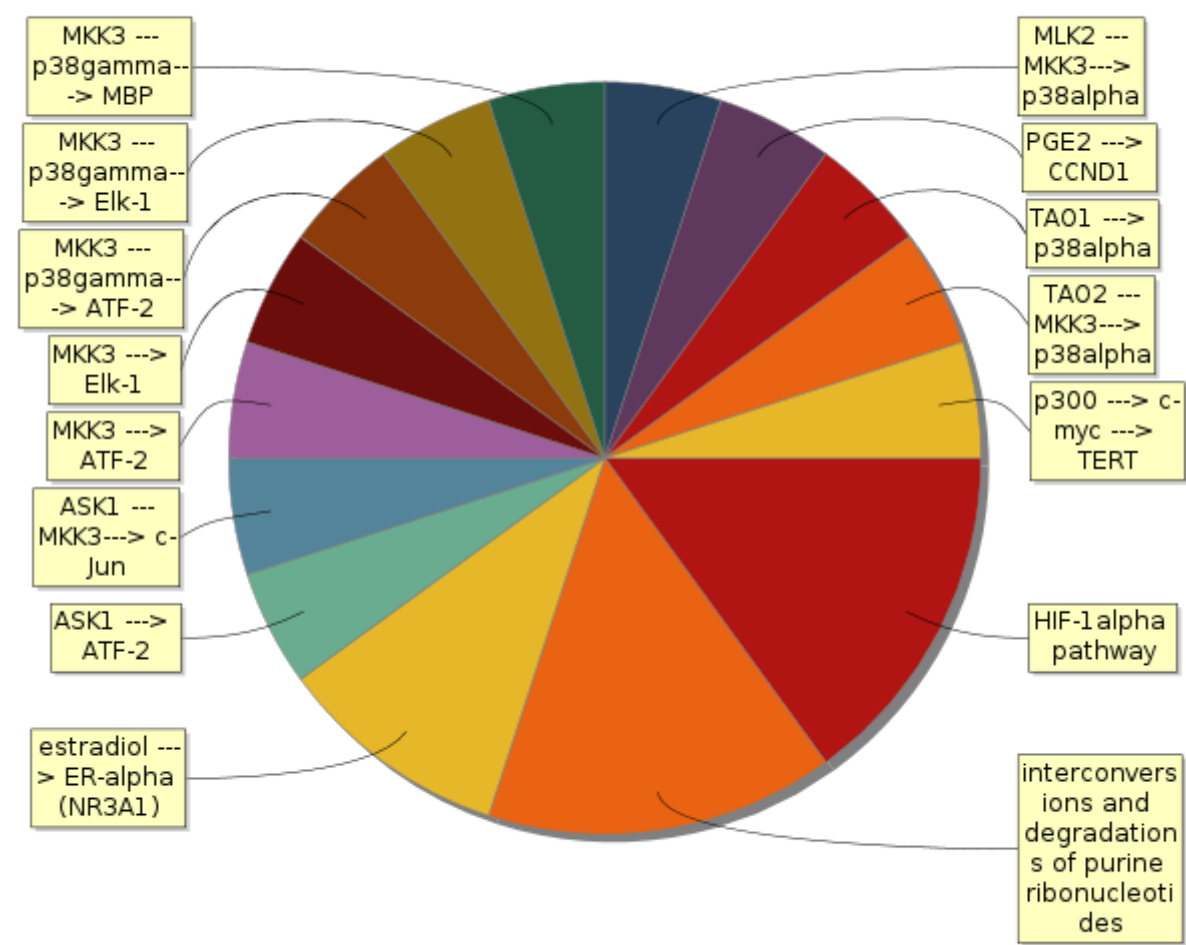


Figure 4. Enriched TRANSPATH® Pathways (2023.2) of up-regulated genes in Myc_induce vs. Control.
[Full classification →](#)

HumanPSD(TM) disease (2023.2)

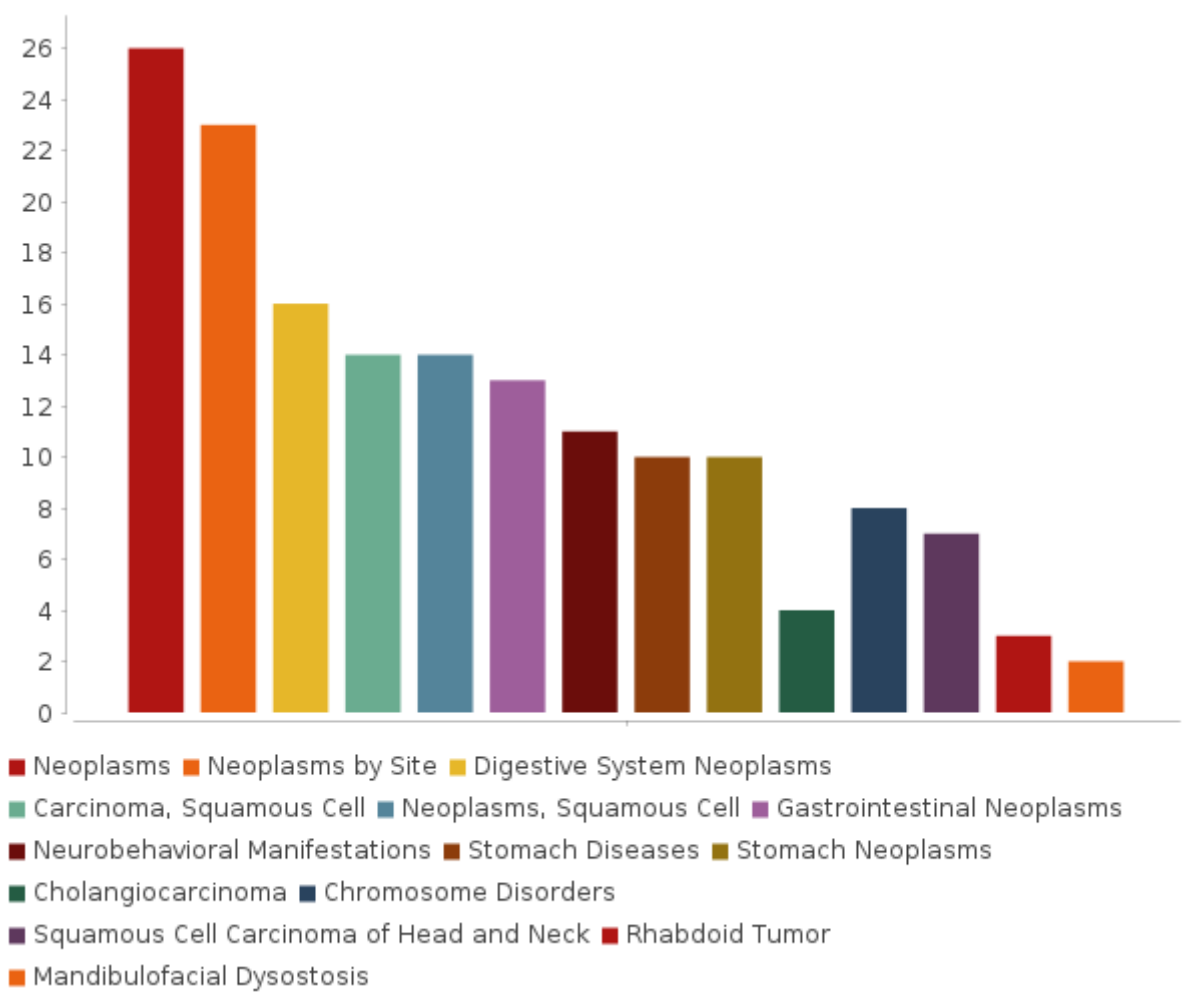


Figure 5. Enriched HumanPSD(TM) disease (2023.2) of up-regulated genes in Myc_induce vs. Control. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

[Full classification](#) →

Down-regulated genes in Myc_induce vs. Control:

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

GO (biological process)



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

Full classification →

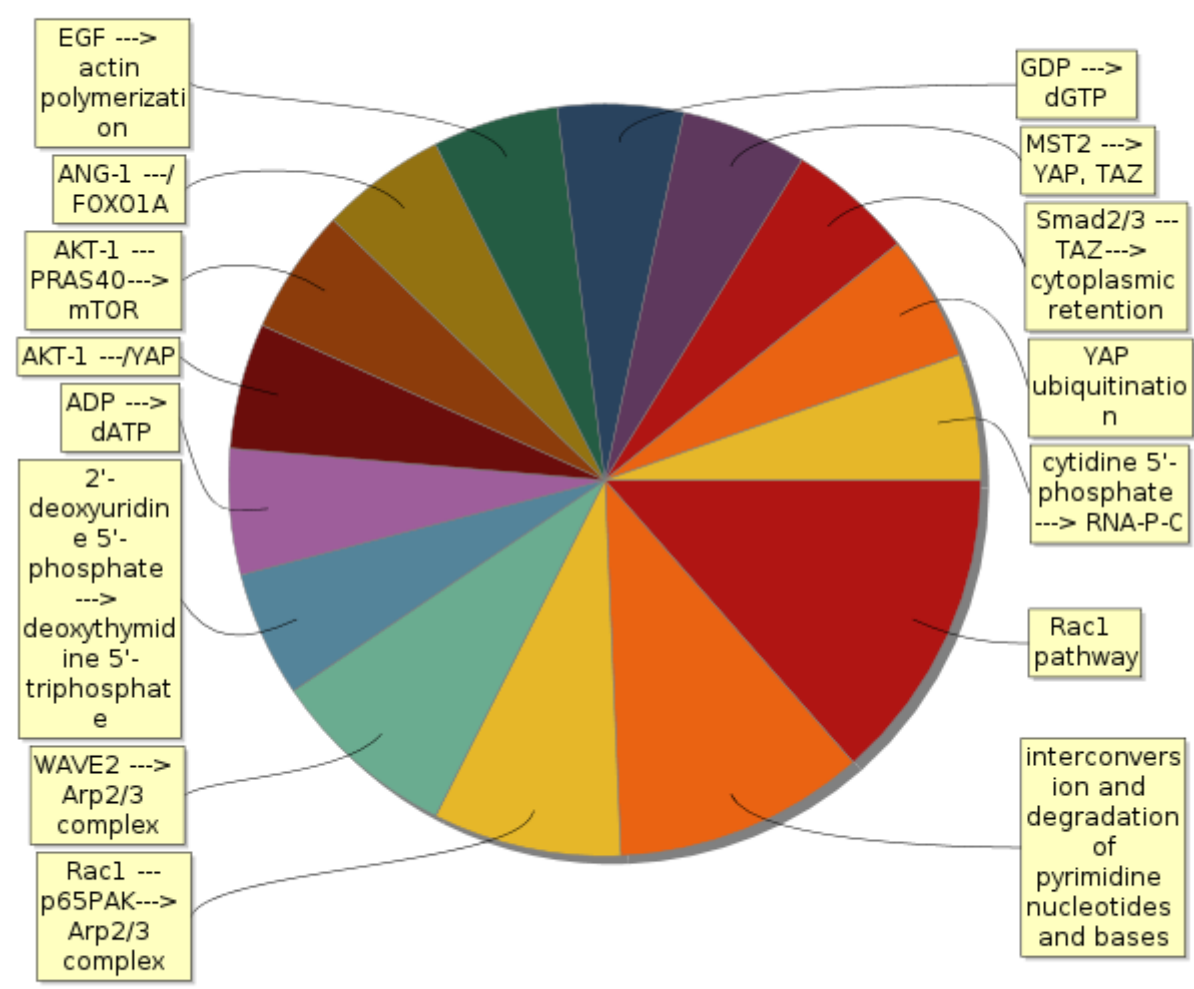


Figure 7. Enriched TRANSPATH® Pathways (2023.2) of down-regulated genes in Myc_induce vs. Control.
[Full classification →](#)

HumanPSD(TM) disease (2023.2)

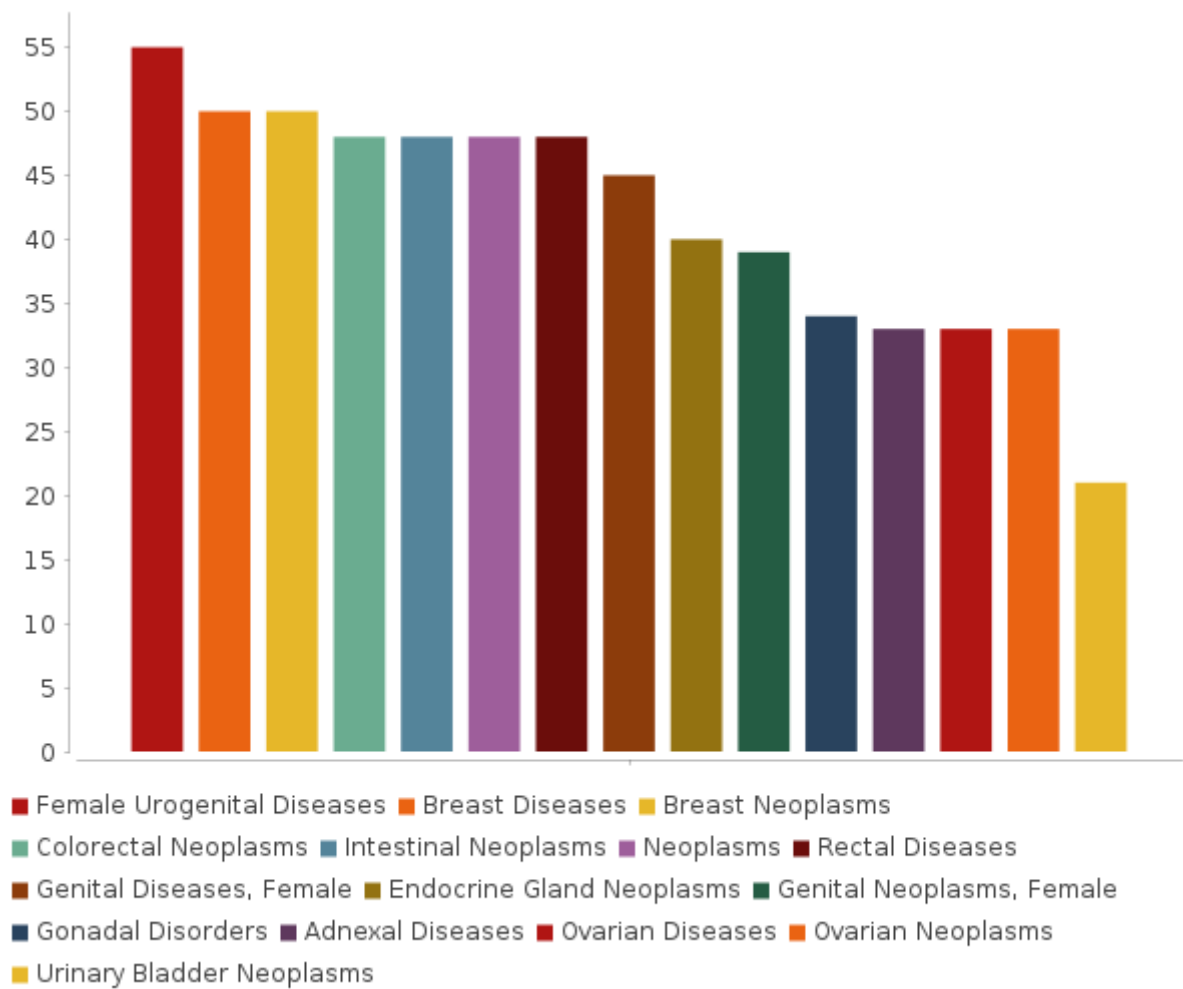


Figure 8. Enriched HumanPSD(TM) disease (2023.2) of down-regulated genes in Myc_induce vs. Control. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

[Full classification](#) →

3.3. Identification of proteins

In the first step of the proteome data analysis target proteins were identified from the uploaded experimental data (the list of 4665 proteins) and were converted to corresponding genes. These genes were used in the further steps of analysis.

Table 4. Top ten the list of genes provided as input in Myc_induce.

[See full table](#) →

ID	Gene description	Gene symbol	Proteomics_avr
ENSG00000173598	nudix hydrolase 4	NUDT4	4.36
ENSG00000100335	mitochondrial elongation factor 1	MIEF1	3.8
ENSG00000115884	syndecan 1	SDC1	3.62
ENSG00000102910	lon peptidase 2, peroxisomal	LONP2	3.3
ENSG00000179046	tripartite motif family like 2	TRIML2	2.87
ENSG00000114648	kelch like family member 18	KLHL18	2.76
ENSG00000170525	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	PFKFB3	2.69
ENSG00000120949	TNF receptor superfamily member 8	TNFRSF8	2.46
ENSG00000188158	NHS actin remodeling regulator	NHS	2.46
ENSG00000119599	DDB1 and CUL4 associated factor 4	DCAF4	2.42

3.4. Functional classification of expressed proteins

A functional analysis of expressed proteins was done by mapping the protein IDs 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 9-11 show the most significant categories.

The list of proteins provided as input in Myc_induce:

4660 the list of genes provided as input in Myc_induce genes were taken for the mapping.

GO (biological process)

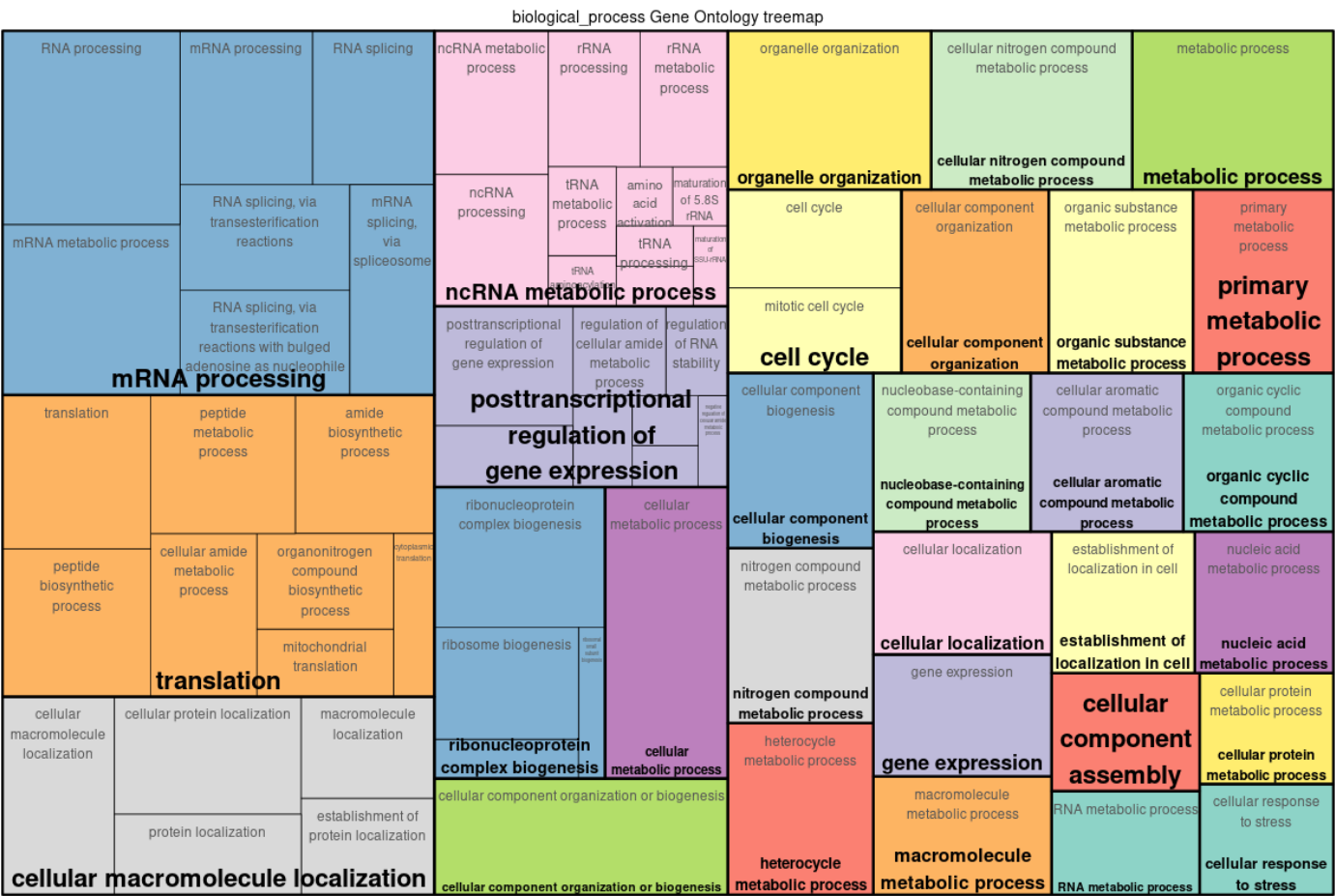


Figure 9. Enriched GO (biological process) of the list of proteins provided as input in Myc_induce.

Full classification →

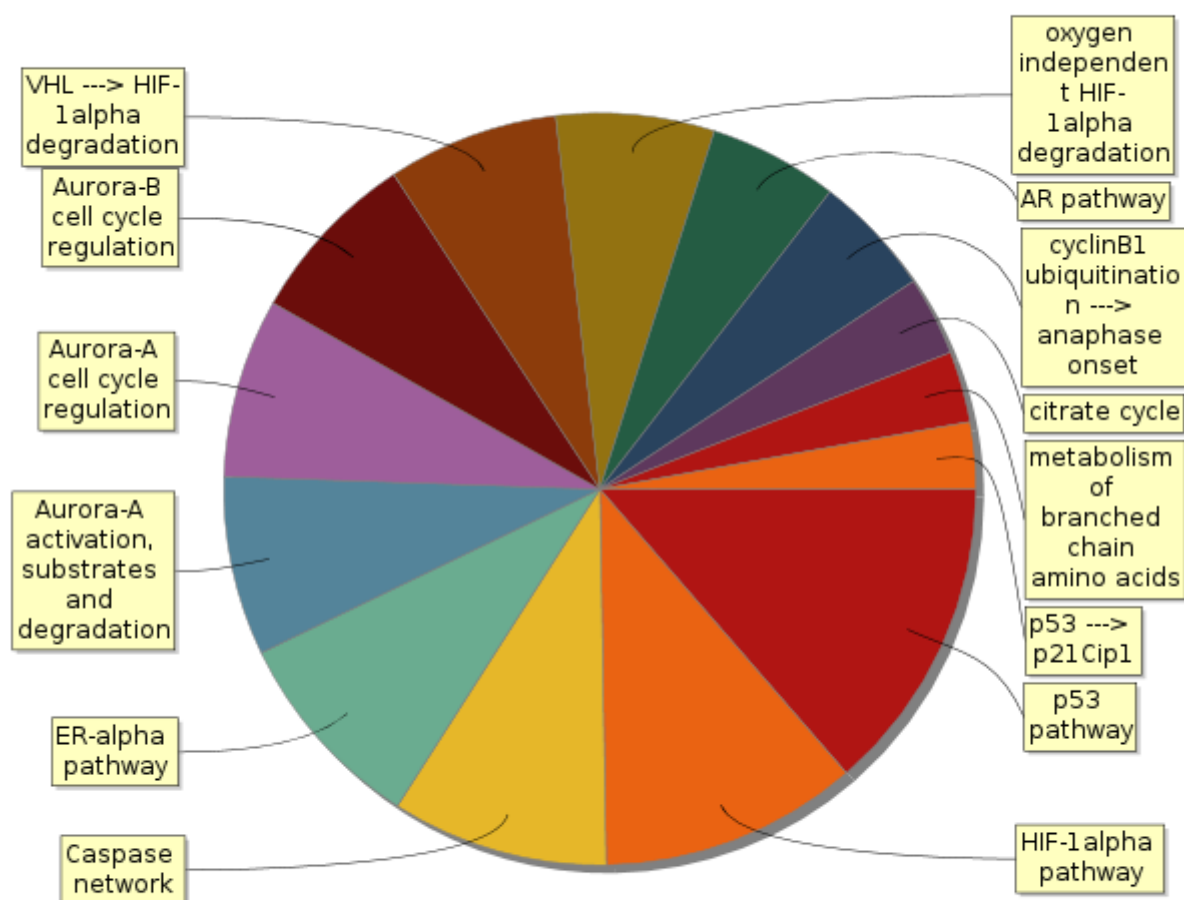


Figure 10. Enriched TRANSPATH® Pathways (2023.2) of the list of proteins provided as input in Myc_induce.
[Full classification →](#)

HumanPSD(TM) disease (2023.2)

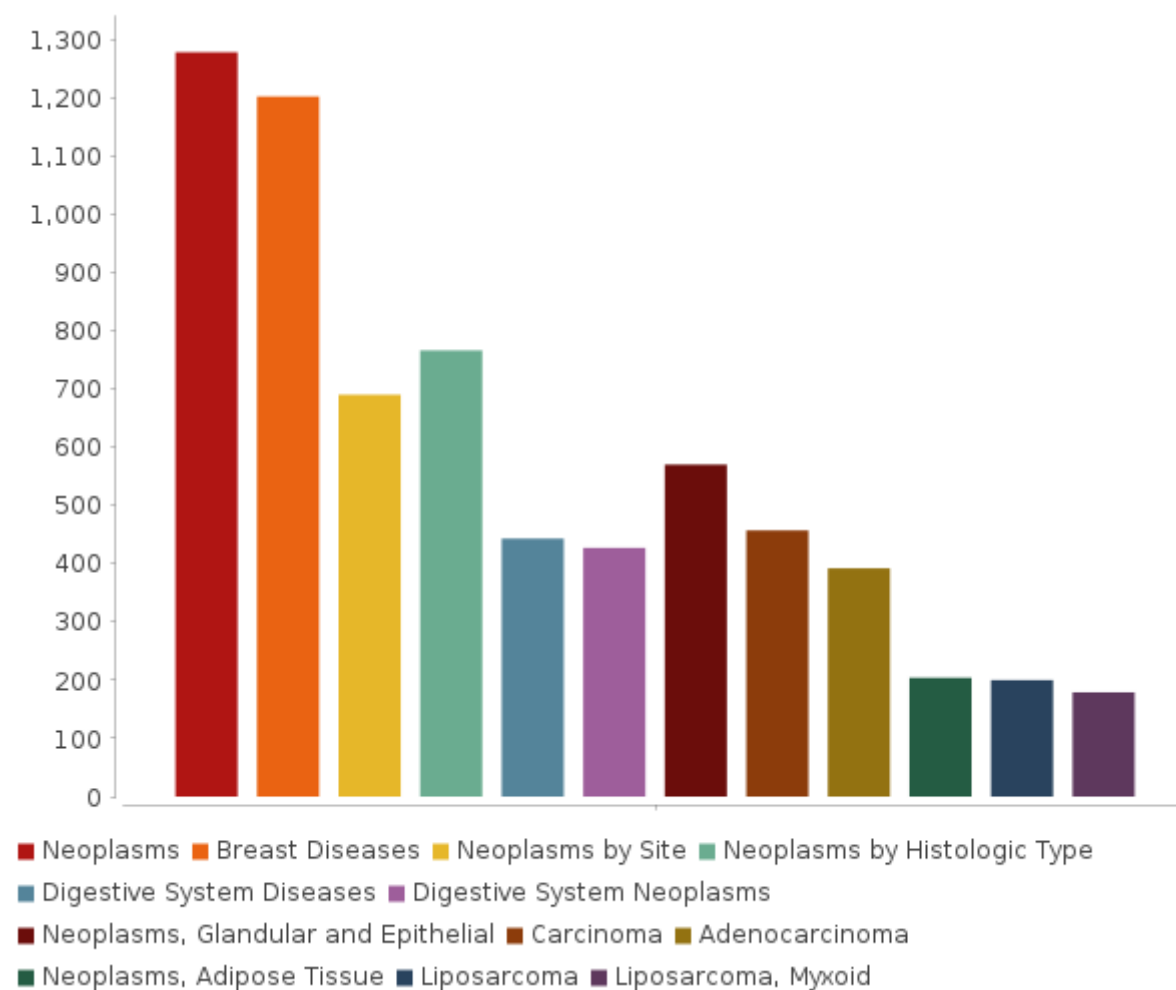


Figure 11. Enriched HumanPSD(TM) disease (2023.2) of the list of proteins provided as input in Myc_induce. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

[Full classification](#) →

3.5. Comparison plot of transcriptome and proteome

After the analysis of transcriptome and proteome data they were compared with each other. Below we plot 1037 genes and 4660 proteins.

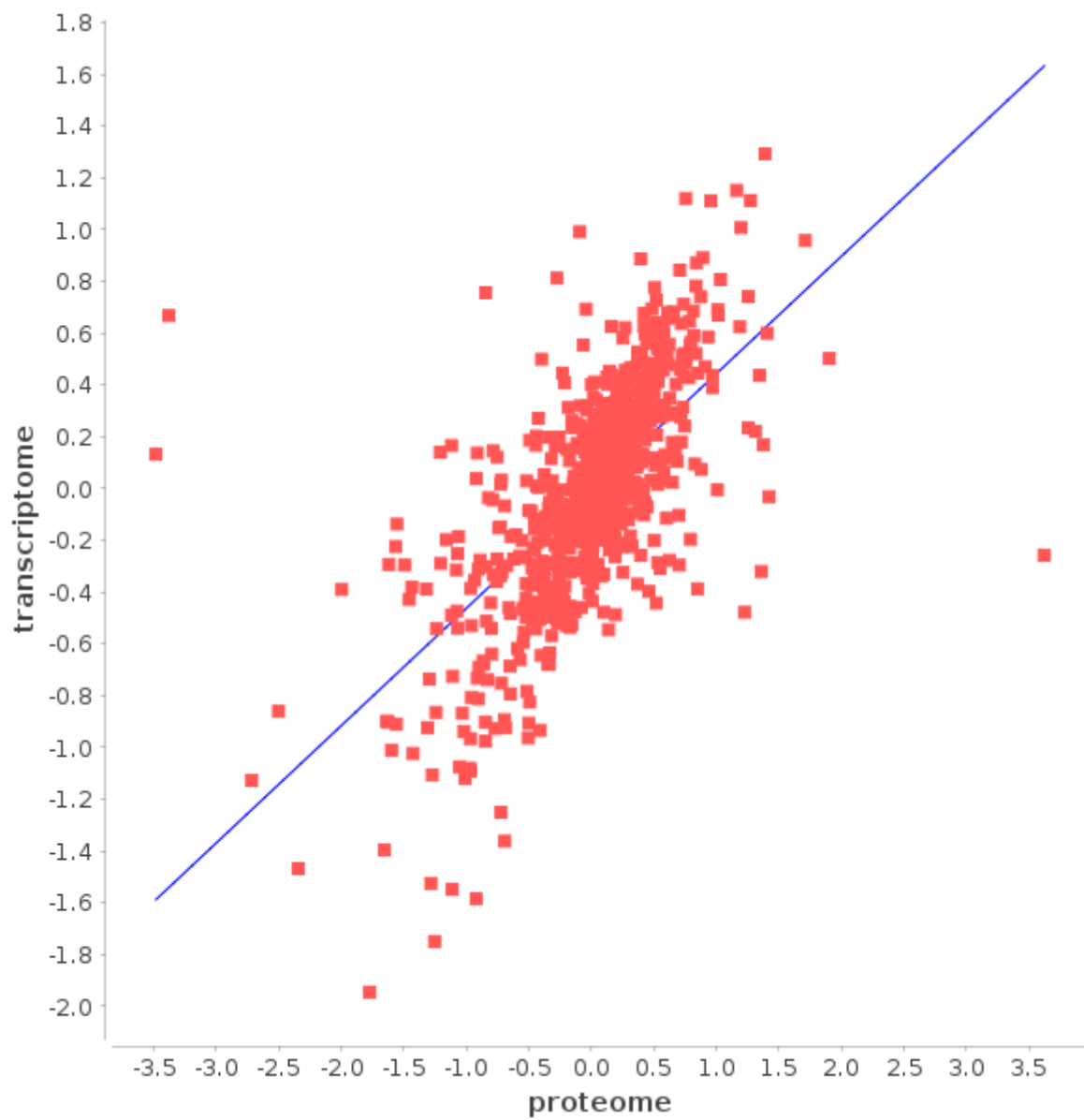


Figure 12. Comparison plot of comparison proteome vs transcriptome. X axis: protein expression value - Proteomics_avr. Y axis: LogFC of differential gene expression.

Full comparison →

Comparison of up-regulated genes (transcriptome data) and the list of proteins provided as input (proteome data)

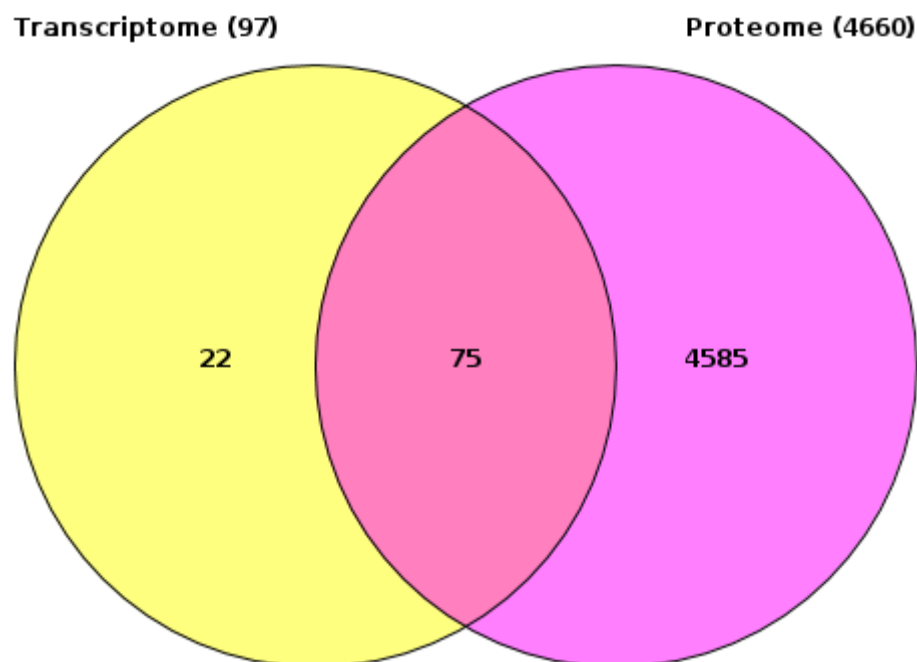


Figure 13. Intersection of up-regulated genes and the list of proteins provided as input
[See full diagram](#) →

Comparison of down-regulated genes (transcriptome data) and the list of proteins provided as input (proteome data)

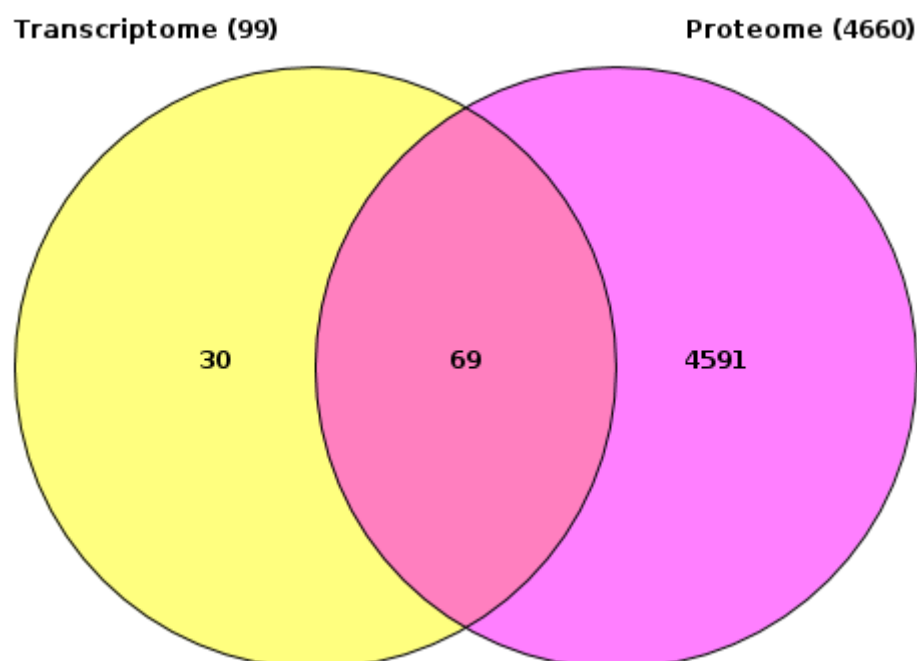
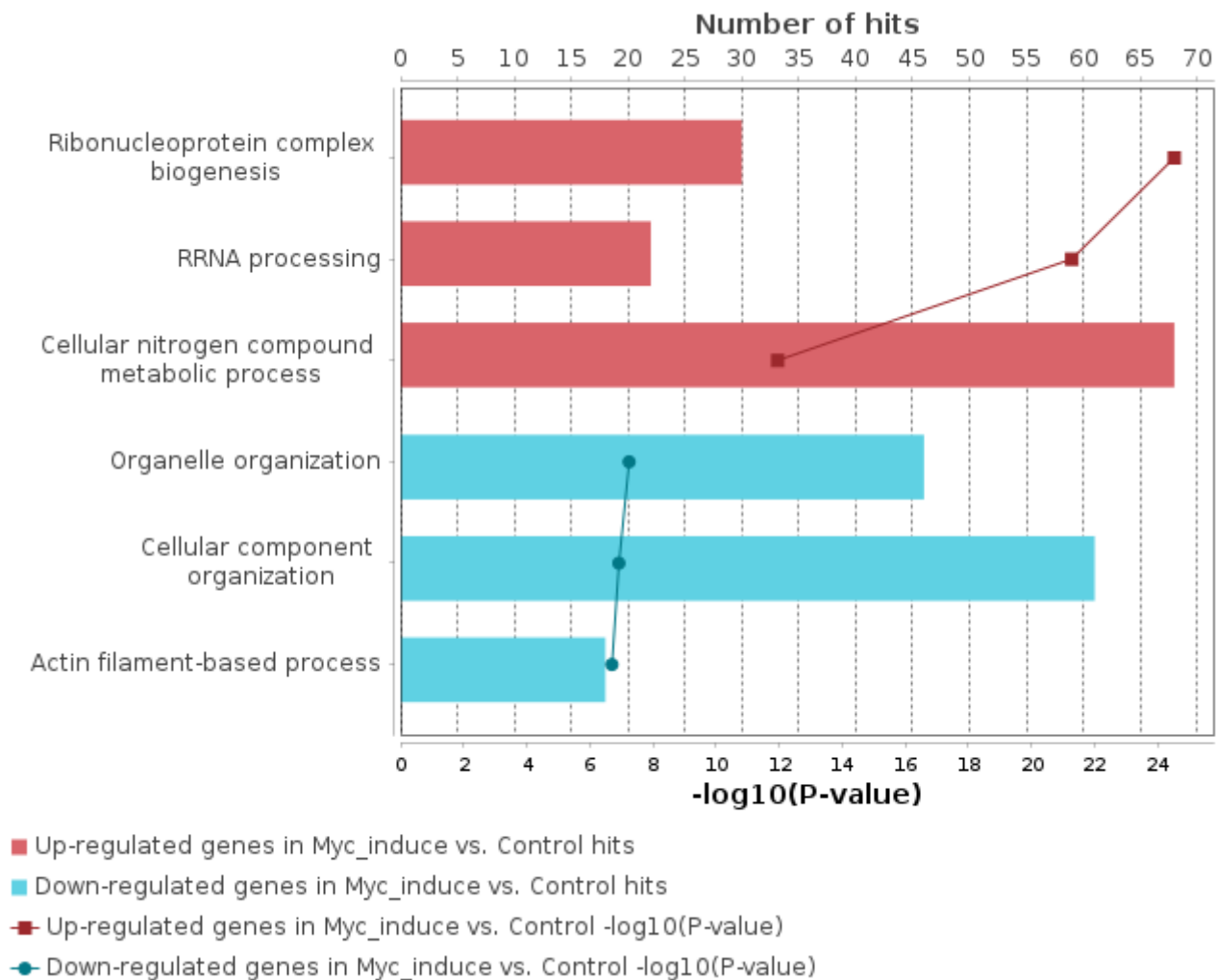


Figure 14. Intersection of down-regulated genes and the list of proteins provided as input
[See full diagram](#) →

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):



3.6. 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 Myc_induce vs. Control).

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

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\$THRB_08 0.69; N=3	V\$GR_Q4 0.96; N=2	V\$AREB6_01 0.95; N=1	V\$MRF4_Q3 0.99; N=1	V\$GABPA_08 0.86; N=2
-------------------------	-----------------------	--------------------------	-------------------------	--------------------------

Module width: 90

Module 2:

V\$FXRRXRA_01 0.72; N=3	V\$P73_02 0.85; N=1	V\$MECP2_Q4 0.85; N=1	V\$ELK1_09 0.68; N=3	V\$SP1_Q6_01 0.92; N=2	V\$TFAP4_06 0.92; N=3
----------------------------	------------------------	--------------------------	-------------------------	---------------------------	--------------------------

Module width: 121

Model score (-p*log10(pval)): 9.69

Wilcoxon p-value (pval): 1.25e-20

Penalty (p): 0.487

Average yes-set score: 7.69

Average no-set score: 5.84

AUC: 0.83

Separation point: 6.54

False-positive: 30.24%

False-negative: 17.53%

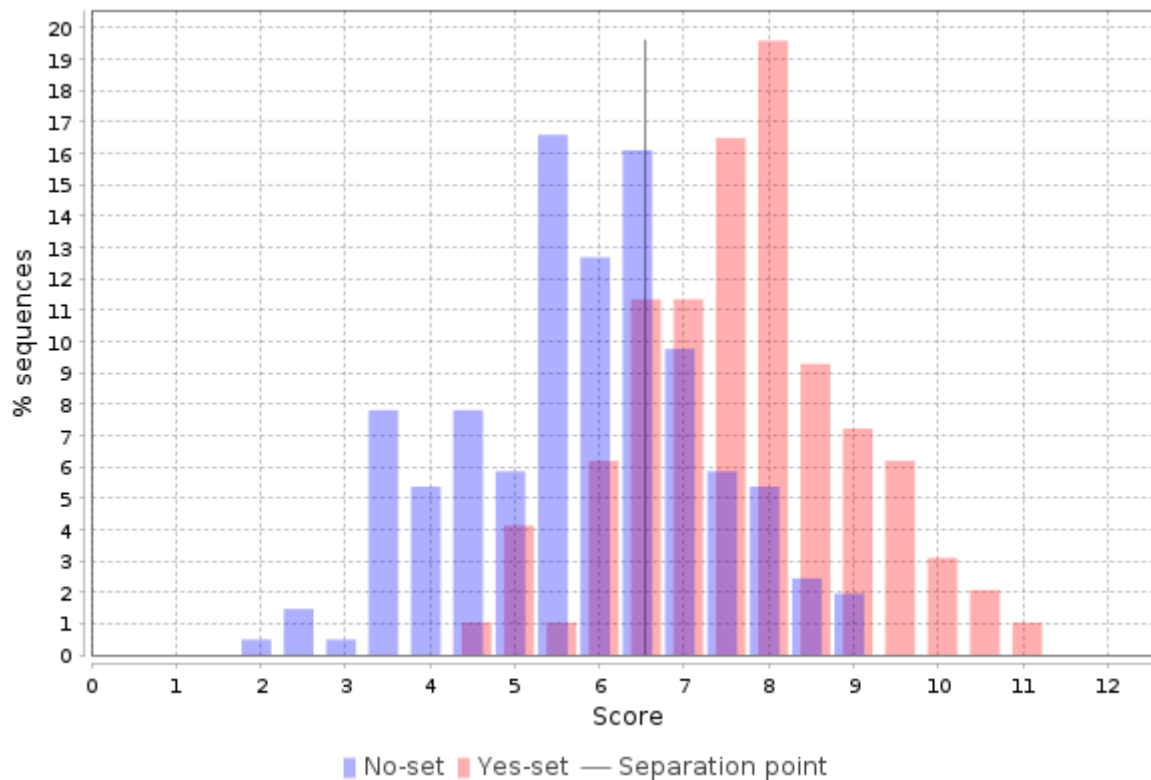


Table 5. List of top ten up-regulated genes in *Myc_induce* vs. *Control* with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.

[See full table](#) →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000101210	EEF1A2	eukaryotic translation elongation factor 1 alpha 2	12.34	GABP-alpha(h), Myf-6(h), T3R-beta(h), p73(h), GR(h), ELK-1(h), Sp1(h),Sp3(h),Sp4(h)...
ENSG00000126088	UROD	uroporphyrinogen decarboxylase	11.05	Sp1(h),Sp3(h),Sp4(h), FXR(h),RXRalpha(h), ELK-1(h), AP-4(h), p73(h), ZEB1(h), MeCp2(h)...
ENSG00000120705	ETF1	eukaryotic translation termination factor 1	11	T3R-beta(h), GR(h), Myf-6(h), AP-4(h), ELK-1(h), Sp1(h),Sp3(h),Sp4(h), FXR(h),RXRalpha(h)...
ENSG00000100142	POLR2F	RNA polymerase II, I and III subunit F	10.55	p73(h), GR(h), T3R-beta(h), GABP-alpha(h), AP-4(h), ELK-1(h), FXR(h),RXRalpha(h)...
ENSG00000104522	GFUS	GDP-L-fucose synthase	10.43	ELK-1(h), Sp1(h),Sp3(h),Sp4(h), MeCp2(h), FXR(h),RXRalpha(h), AP-4(h), p73(h), T3R-beta(h)...
ENSG00000184428	TOP1MT	DNA topoisomerase I mitochondrial	10.38	Sp1(h),Sp3(h),Sp4(h), p73(h), FXR(h),RXRalpha(h), AP-4(h), ZEB1(h), Myf-6(h), T3R-beta(h)...
ENSG00000204463	BAG6	BAG cochaperone 6	10.3	Sp1(h),Sp3(h),Sp4(h), p73(h), AP-4(h), ELK-1(h), FXR(h),RXRalpha(h), MeCp2(h), T3R-beta(h)...
ENSG00000100029	PES1	pescadillo ribosomal biogenesis factor 1	10.18	MeCp2(h), FXR(h),RXRalpha(h), AP-4(h), p73(h), GR(h), Sp1(h),Sp3(h),Sp4(h), Myf-6(h)...
ENSG00000128283	CDC42EP1	CDC42 effector protein 1	9.95	AP-4(h), FXR(h),RXRalpha(h), Sp1(h),Sp3(h),Sp4(h), MeCp2(h), ZEB1(h), GR(h), GABP-alpha(h)...
ENSG00000071462	BUD23	BUD23 rRNA methyltransferase and ribosome maturation factor	9.93	FXR(h),RXRalpha(h), T3R-beta(h), ZEB1(h), Myf-6(h), AP-4(h), MeCp2(h), Sp1(h),Sp3(h),Sp4(h)...

Enhancer model potentially involved in regulation of target genes (down-regulated genes in *Myc_induce* vs. *Control*).

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

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\$NR3C1_03 0.71; N=3	V\$SP3_Q3 0.82; N=2	V\$LYF1_01 0.83; N=2	V\$TCF12_08 0.98; N=1
--------------------------	------------------------	-------------------------	--------------------------

Module width: 133

Module 2:

V\$TFIII_Q6 0.96; N=1	V\$CTF_01 0.99; N=1	V\$AML2_01 0.95; N=2	V\$PROX1_02 0.84; N=1	V\$RELA_10 0.97; N=2
--------------------------	------------------------	-------------------------	--------------------------	-------------------------

Module width: 175

Model score (-p*log10(pval)): 12.00

Wilcoxon p-value (pval): 6.40e-24

Penalty (p): 0.517

Average yes-set score: 5.74

Average no-set score: 4.02

AUC: 0.85

Separation point: 5.20

False-positive: 15.61%

False-negative: 28.28%

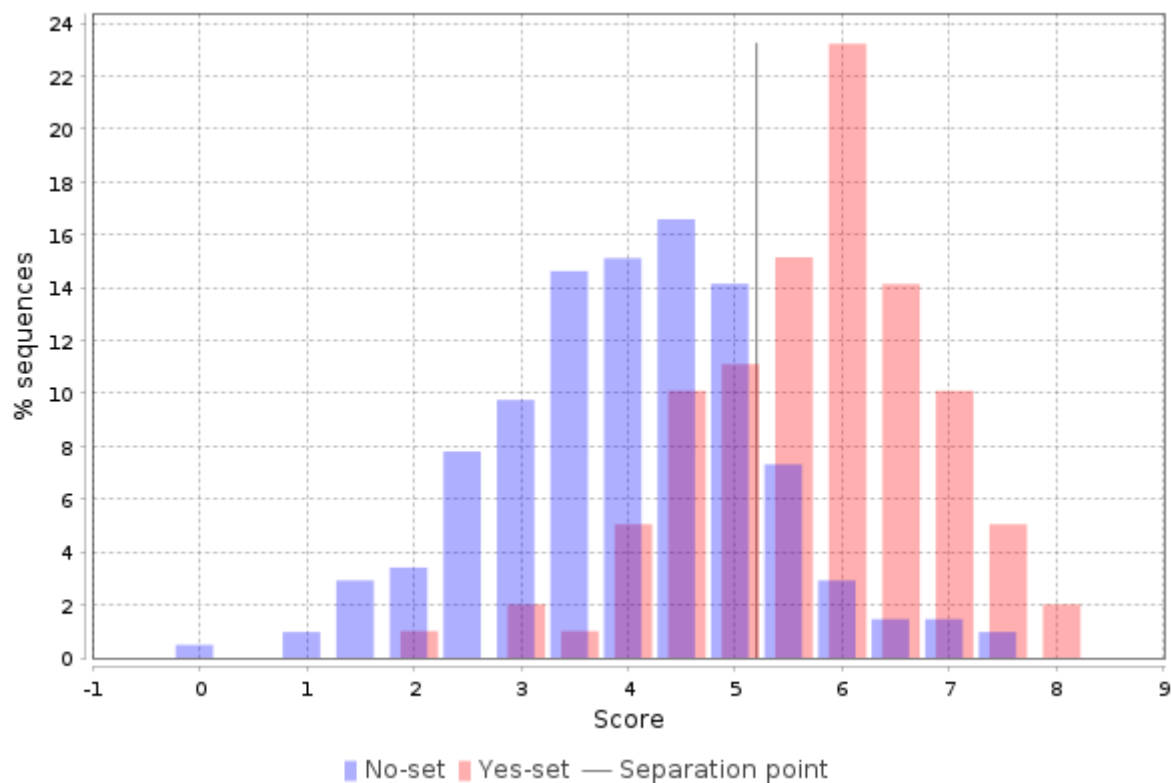


Table 6. List of top ten down-regulated genes in *Myc_induce* vs. *Control* with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.

[See full table](#) →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000142945	KIF2C	kinesin family member 2C	7.92	HTF-4(h), Runx3(h), GR(h), IKZF1(h), Sp3(h), NF-1C(h), PROX-1(h)
ENSG00000125534	PPDPF	pancreatic progenitor cell differentiation and proliferation factor	7.92	TFII-I(h), PROX-1(h), NF-1C(h), GR(h), IKZF1(h), Sp3(h)
ENSG00000161011	SQSTM1	sequestosome 1	7.9	GR(h), Sp3(h), IKZF1(h), HTF-4(h), NF-1C(h), PROX-1(h), TFII-I(h)
ENSG00000140988	RPS2	ribosomal protein S2	7.7	GR(h), PROX-1(h), NF-1C(h), TFII-I(h), Sp3(h), IKZF1(h), HTF-4(h)
ENSG00000196230	TUBB	tubulin beta class I	7.61	GR(h), Sp3(h), IKZF1(h), HTF-4(h), NF-1C(h), Runx3(h)
ENSG00000010292	NCAPD2	non-SMC condensin I complex subunit D2	7.43	PROX-1(h), Runx3(h), NF-1C(h), GR(h), Sp3(h), IKZF1(h), HTF-4(h)
ENSG00000169715	MT1E	metallothionein 1E	7.42	Runx3(h), HTF-4(h), IKZF1(h), Sp3(h), NF-1C(h), GR(h)
ENSG00000113140	SPARC	secreted protein acidic and cysteine rich	7.34	Sp3(h), TFII-I(h), IKZF1(h), NF-1C(h), PROX-1(h), Runx3(h), GR(h)...
ENSG00000196576	PLXNB2	plexin B2	7.33	NF-1C(h), TFII-I(h), IKZF1(h), HTF-4(h), Sp3(h), PROX-1(h), GR(h)
ENSG00000164754	RAD21	RAD21 cohesin complex component	7.28	Sp3(h), Runx3(h), NF-1C(h), GR(h), PROX-1(h), IKZF1(h)

On the basis of the enhancer models we identified transcription factors potentially regulating the **target genes** of our interest. We found 14 and 9 transcription factors controlling expression of up- and down-regulated genes respectively (see Tables 7-8).

Table 7. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (up-regulated genes in *Myc_induce* vs. *Control*). **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](#) →

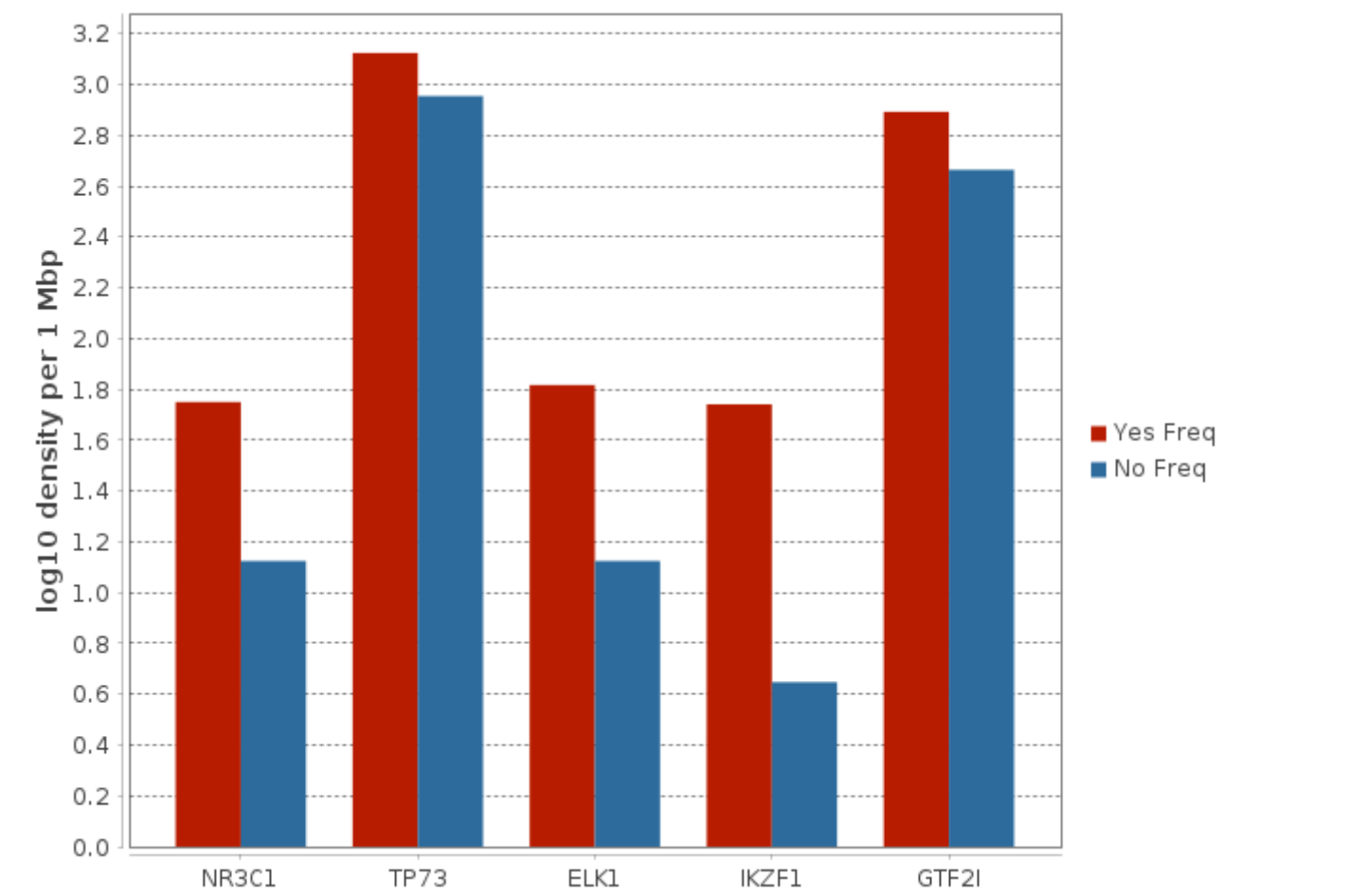
ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	3.15	4.22
MO000028707	TP73	tumor protein p73	2.81	1.48
MO000019544	ELK1	ETS transcription factor ELK1	2.78	4.93
MO000024986	MYF6	myogenic factor 6	2.59	1.68
MO000019619	RXRA	retinoid X receptor alpha	2.51	2.9
MO000033308	SP1	Sp1 transcription factor	2.34	1.66
MO000139677	ZEB1	zinc finger E-box binding homeobox 1	2.04	7.74
MO000028711	MECP2	methyl-CpG binding protein 2	1.92	1.15
MO000045528	GABPA	GA binding protein transcription factor subunit alpha	1.83	1.58
MO000088742	NR1H4	nuclear receptor subfamily 1 group H member 4	1.75	5.91

Table 8. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (down-regulated genes in Myc_induce vs. Control). **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
MO000026678	IKZF1	IKAROS family zinc finger 1	2.31	12.41
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	1.84	1.84
MO000019622	GTF2I	general transcription factor Iii	1.52	1.69
MO000079319	RELA	RELA proto-oncogene, NF-kB subunit	1.46	10.34
MO000026238	RUNX3	RUNX family transcription factor 3	1.36	8.27
MO000025717	TCF12	transcription factor 12	1.33	3.38
MO000046079	SP3	Sp3 transcription factor	1.05	1.29
MO000024750	NFIC	nuclear factor I C	0	2.43
MO000028703	PROX1	prospero homeobox 1	0	1.62

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: NR3C1, TP73, ELK1, IKZF1 and GTF2I.



3.7. Finding master regulators in networks

In the second step of the upstream analysis common regulators of the revealed TFs were identified. Using proteomics data we selected differentially expressed proteins that are involved in signal transduction pathways and used these proteins as "context set" [4] in the algorithm of identification of master regulators. 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 9-10.

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

[See full table](#) →

ID	Master molecule name	Gene symbol	Gene description	Contained in proteome set	Total rank	logFC (transcriptome)
MO000023504	cyclinD1(h):Cdk6(h)	CCND1, CDK6	cyclin D1, cyclin dependent kinase 6	1	87	0.68
MO000023505	cyclinD1(h):Cdk6(h) {pT177}	CCND1, CDK6	cyclin D1, cyclin dependent kinase 6	1	88	0.68
MO000032096	proCaspase-3(h):Hsp10(h):Hsp60(h)	CASP3, HSPD1, HSPE1	caspase 3, heat shock protein family D (Hsp60) member 1, heat shock protein family E (Hsp10) member ...	1	90	0.53
MO000031196	cyclinD1(h):Cdk4(h)	CCND1, CDK4	cyclin D1, cyclin dependent kinase 4	1	98	0.68
MO000009403	MKK3(h)	MAP2K3	mitogen-activated protein kinase kinase 3	1	113	0.45
MO000041399	MKK3(h){pS189} {pT193}	MAP2K3	mitogen-activated protein kinase kinase 3	1	116	0.45
MO000022217	MKK3(h){p}	MAP2K3	mitogen-activated protein kinase kinase 3	1	117	0.45
MO000161221	MKK3-isoform2(h)	MAP2K3	mitogen-activated protein kinase kinase 3	1	123	0.45
MO000161222	MKK3-isoform3(h)	MAP2K3	mitogen-activated protein kinase kinase 3	1	123	0.45
MO000078695	MKK3-isoform1(h)	MAP2K3	mitogen-activated protein kinase kinase 3	1	125	0.45

Table 10. Master regulators that may govern the regulation of **down-regulated** genes in *Myc_induce* vs. *Control*. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics and proteomics data.

[See full table](#) →

ID	Master molecule name	Gene symbol	Gene description	Contained in proteome set	Total rank	logFC (transcriptome)
MO000044885	PP1-alpha(h)	PPP1CA	protein phosphatase 1 catalytic subunit alpha	1	159	-0.32
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	163	-1.41
MO000031326	PP1-alpha1(h)	PPP1CA	protein phosphatase 1 catalytic subunit alpha	1	168	-0.32
MO0000312448	PP1-alpha-isoform3(h)	PPP1CA	protein phosphatase 1 catalytic subunit alpha	1	168	-0.32
MO000031324	PP1-alpha2(h)	PPP1CA	protein phosphatase 1 catalytic subunit alpha	1	170	-0.32
MO000066670	galectin-1(h)	LGALS1	galectin 1	1	171	-1.55
MO000021293	14-3-3eta(h)	YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein eta	1	199	-0.57
MO000030879	14-3-3sigma(h)	SFN	stratifin	1	202	-0.97
MO000022340	IL-8(h):CXCR2(h):G-alpha-i2(h)	CXCL8, CXCR2, GNAI2	C-X-C motif chemokine ligand 8, C-X-C motif chemokine receptor 2, G protein subunit alpha i2	1	205	-1.47
MO000022339	IL-8(h):CXCR1(h):G-alpha-i2(h)	CXCL8, CXCR1, GNAI2	C-X-C motif chemokine ligand 8, C-X-C motif chemokine receptor 1, G protein subunit alpha i2	1	206	-1.47

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figures 15 and 16. 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.

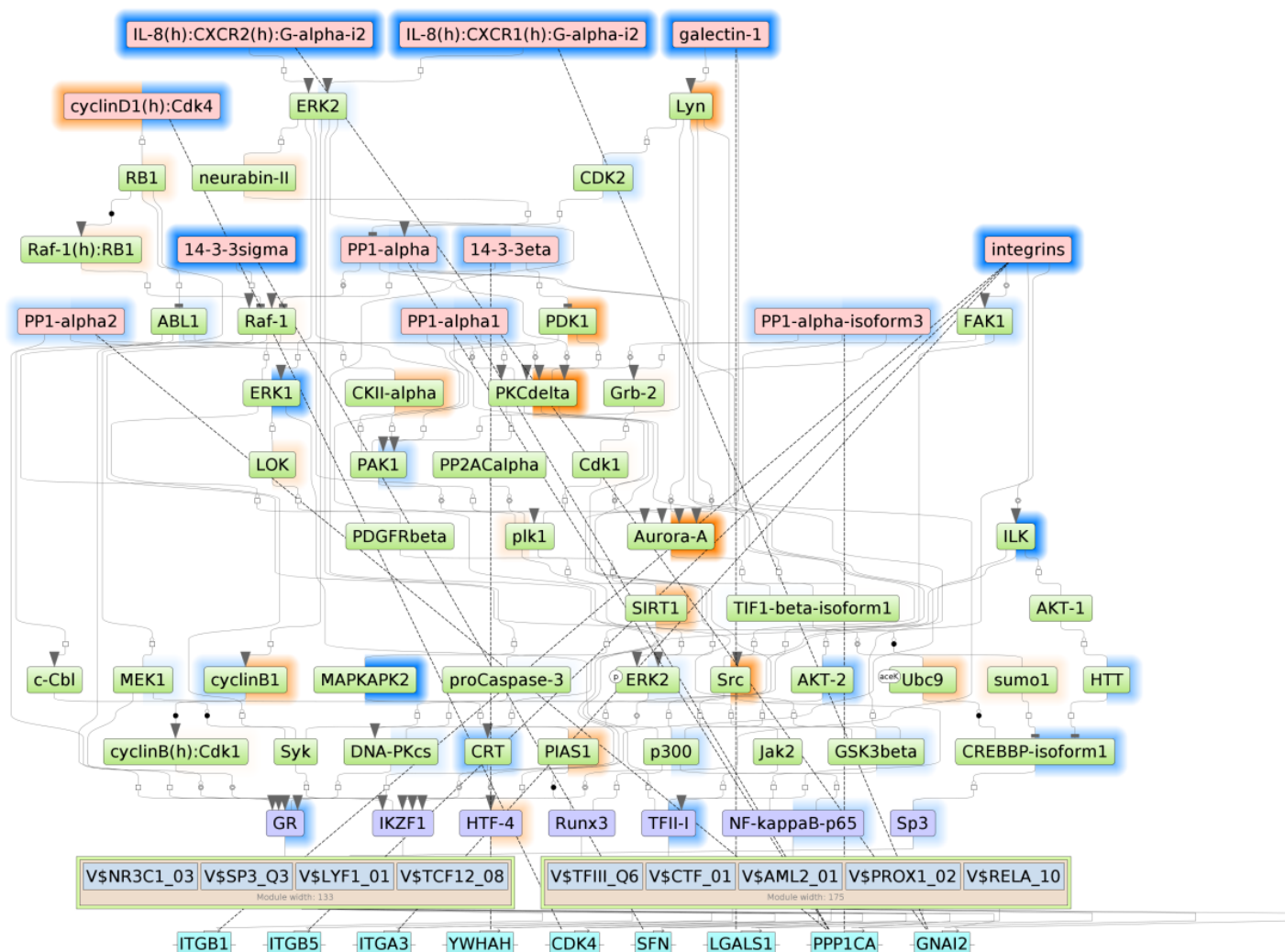


Figure 16. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in *Myc* induce vs. Control. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the 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. The left half of a highlighting frame corresponds to transcriptomic data, the right one to proteomic data.

[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 11. 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	Contained in proteome set	Total rank	logFC (transcriptome)
FKBP4	FKBP prolyl isomerase 4	1	1	221	0.87
MAP2K3	mitogen-activated protein kinase kinase 3	30	1	233	0.45
TFRC	transferrin receptor	4	1	247	0.43
PRMT1	protein arginine methyltransferase 1	2	1	262	0.56
CYCS	cytochrome c, somatic	4	1	271	0.41
HSP90AB1	heat shock protein 90 alpha family class B member 1	8	1	316	0.44



Table 12. 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.

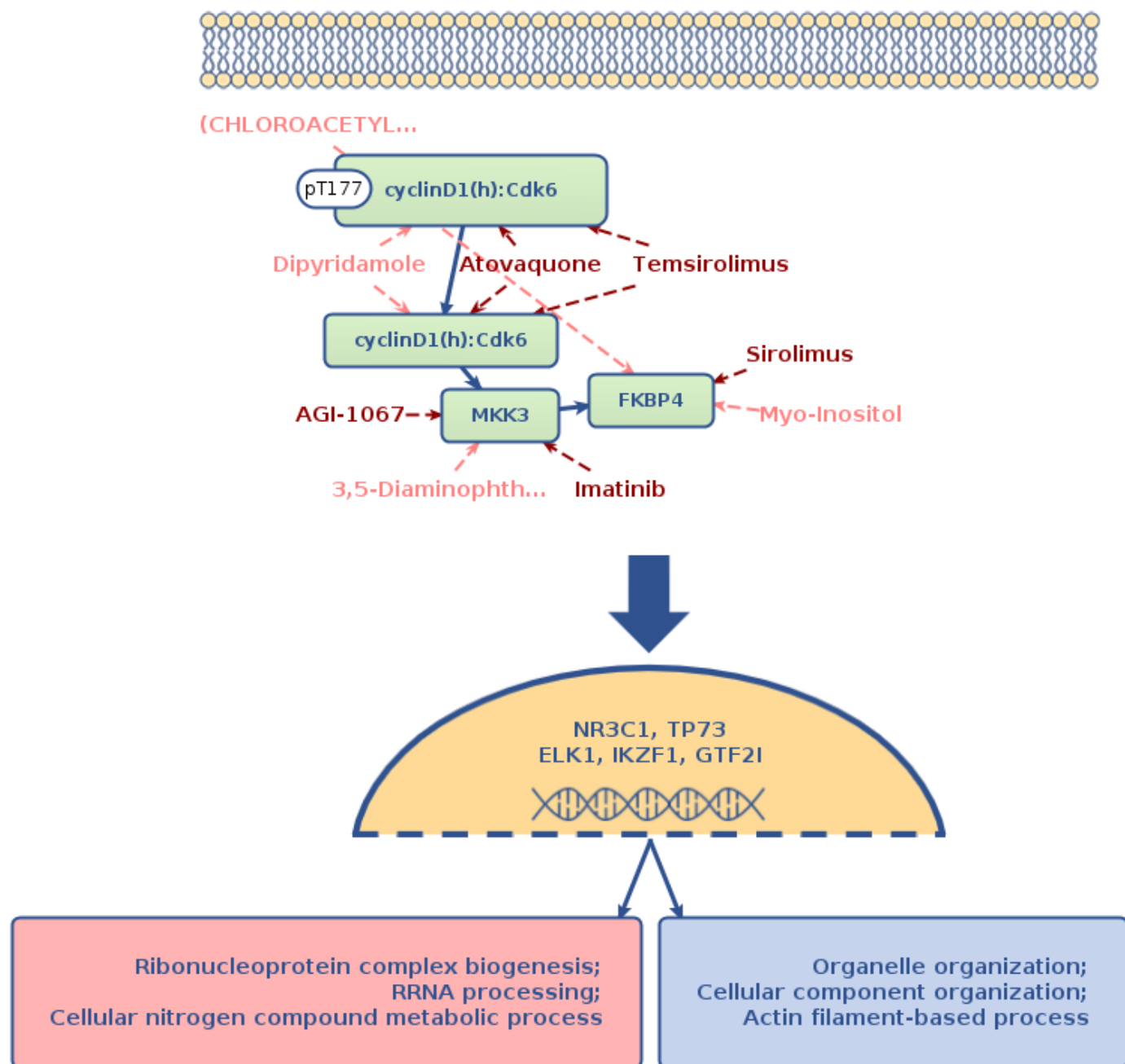
[See full table](#) →

Gene symbol	Gene Description	Druggability score	Contained in proteome set	Total rank	logFC (transcriptome)
FKBP4	FKBP prolyl isomerase 4	63.67	1	221	0.87
MAP2K3	mitogen-activated protein kinase kinase 3	6.15	1	233	0.45
PRMT1	protein arginine methyltransferase 1	1.34	1	262	0.56
HSP90AB1	heat shock protein 90 alpha family class B member 1	7.61	1	316	0.44
FASN	fatty acid synthase	313.18	1	352	0.37
NDUFAB1	NADH:ubiquinone oxidoreductase subunit AB1	134.93	1	403	0.37

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:

- cyclinD1(h):Cdk6
- cyclinD1(h):Cdk6
- FKBP4
- MKK3

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: Myo-Inositol, Imatinib, AGI-1067, Sirolimus, Atovaquone, Dipyridamole, (CHLOROACETYL)CARBAMIC ACID (3R,4S,5S,5R)-5-METHOXY-4-[(2R,3R)-2-METHYL-3-(3-METHYL-2-BUTENYL)OXIRANYL]-1-OXASPIRO[2.5]OCT-6-YL ESTER, Temsirolimus and 3,5-Diaminophthalhydrazide, 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 13 and 14), 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 13. 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
Temsirolimus	CCND1, EIF4EBP1, EIF4E	95	8	Phase 3: Neoplasm Metastasis, Angiomyoma, Arteriovenous Fistula, Breast Neoplasms, Carcinoma, Renal Cell, Constriction, Pathologic, Coronary Artery Disease, Coronary Disease, Coronary Occlusion, Coronary Stenosis, Fistula, Hypertension, Kidney Diseases, Kidney Failure, Chronic, Leiomyoma, Leiomyomatosis, Lymphangi leiomyomatosis, Lymphangiomyoma, Lymphoma, Lymphoma, Mantle-Cell, Myoma, Myosarcoma, Neoplasms, Renal Insufficiency, Renal Insufficiency, Chronic, Rhabdomyosarcoma, Rhabdomyosarcoma, Alveolar, Rhabdomyosarcoma, Embryonal, Sarcoma
Sirolimus	PLK1, FKBP1A, EIF4E, FKBP4, HSP90AA1	87	9	Phase 3: Neoplasm Metastasis, Anemia, Angiofibroma, Angiomyoma, Arteriovenous Fistula, Atherosclerosis, Brain Abscess, Brain Stem Infarctions, Breast Neoplasms, COVID-19, Cardiovascular Abnormalities, Carotid Artery Diseases, Cognitive Dysfunction, Congenital Abnormalities, Constriction, Pathologic, Coronary Artery Disease, Coronary Disease, Coronary Occlusion, Coronary Restenosis, Coronary Stenosis, Cysts, Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetic Nephropathies, Dyspnea, End Stage Liver Disease, Fanconi Anemia, Fanconi Syndrome, Fibroma, Fibrosis, Fistula, Graft vs Host Disease, Hematologic Neoplasms, Hypertension, Infarction, Ischemia, Kidney Diseases, Kidney Diseases, Cystic, Kidney Failure, Chronic, Klippel-Trenaunay-Weber Syndrome, Leiomyoma, Leiomyomatosis, Leukemia, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Leukemia, Myeloid, Acute, Leukemia, Polymorphocytic, Liver Diseases, Lupus Nephritis, Lymphangi leiomyomatosis, Lymphangiomyoma, Lymphoma, Lymphoma, Large B-Cell, Diffuse, Lymphoma, Mantle-Cell, Lymphoma, Non-Hodgkin, Multiple Myeloma, Myelodysplastic Syndromes, Myeloproliferative Disorders, Myocardial Ischemia, Myoma, Myosarcoma, Nails, Malformed, Neoplasms, Nephritis, Pachyonychia Congenita, Peritoneal Fibrosis, Pneumonia, Polycystic Kidney Diseases, Post-Acute COVID-19 Syndrome, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Pulmonary Fibrosis, Recurrence, Renal Insufficiency, Renal Insufficiency, Chronic, Retroperitoneal Fibrosis, Rhabdomyosarcoma, Rhabdomyosarcoma, Alveolar, Rhabdomyosarcoma, Embryonal, Sarcoma, Sclerosis, Sturge-Weber Syndrome, Syndrome, Tuberous Sclerosis, Vascular Malformations, Waldenstrom Macroglobulinemia
pu-h71	HSP90AB1, HSP90AA1	86	1	Phase 1: Neoplasm Metastasis, Lymphoma, Myeloproliferative Disorders, Polycythemia, Polycythemia Vera, Primary Myelofibrosis, Thrombocytopenia, Essential, Thrombocytosis
CNF1010	HSP90AB1, HSP90AA1	85	1	Phase 1: Osteosarcoma, Breast Neoplasms, Burkitt Lymphoma, Carcinoma, Renal Cell, Chondrosarcoma, Colorectal Neoplasms, Granuloma, Head and Neck Neoplasms, Hodgkin Disease, Immunoblastic Lymphadenopathy, Intraocular Lymphoma, Kidney Neoplasms, Leukemia, Leukemia, Biphenotypic, Acute, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Leukemia, Myeloid, Chronic-Phase, Leukemia, T-Cell, Leukemia-Lymphoma, Adult T-Cell, Lung Neoplasms, Lymphadenopathy, Lymphatic Diseases, Lymphoma, Lymphoma, AIDS-Related, Lymphoma, B-Cell, Lymphoma, B-Cell, Marginal Zone, Lymphoma, Extranodal NK-T-Cell, 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, Lymphomatoid Granulomatosis, Lymphoproliferative Disorders, Melanoma, Mycoses, Mycosis Fungoides, Ovarian Neoplasms, Plasmablastic Lymphoma, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Recurrence, Sarcoma, Sarcoma, Ewing, Sezary Syndrome, Stomach Neoplasms, Syndrome, Urinary Bladder Neoplasms, Uterine Neoplasms, Waldenstrom Macroglobulinemia

Curcumin	CCND1, PARP1, ADAM15, EIF4EBP1, EIF2S1, EGR1, TFRC, MYC, CDKN1A	84	3	Phase 2: Osteosarcoma, Aberrant Crypt Foci, Acute Kidney Injury, Adenocarcinoma, Alzheimer Disease, Amyotrophic Lateral Sclerosis, Aneurysm, Anorexia, Aortic Aneurysm, Aortic Aneurysm, Abdominal, Asthma, Atherosclerosis, Bipolar Disorder, Bites and Stings, Brain Abscess, Cachexia, Cardiovascular Diseases, Cognitive Dysfunction, Diabetes Mellitus, Diabetes Mellitus, Type 2, Diabetic Nephropathies, Diverticular Diseases, Diverticulitis, Diverticulum, Dyslipidemias, Fibrosis, Head and Neck Neoplasms, Hyperemia, Hyperlipidemias, Inflammation, Insulin Resistance, Kidney Diseases, Lung Diseases, Lung Neoplasms, Lupus Nephritis, Mental Disorders, Monoclonal Gammopathy of Undetermined Significance, Motor Neuron Disease, Mucositis, Multiple Myeloma, Multiple Sclerosis, Neoplasms, Neoplasms, Plasma Cell, Nephritis, Oral Submucous Fibrosis, Pain, Paraproteinemias, Prostatic Neoplasms, Proteinuria, Psychotic Disorders, Renal Insufficiency, Renal Insufficiency, Chronic, Sarcoma, Schizophrenia, Sclerosis, Smoldering Multiple Myeloma, Tic Disorders, Tobacco Use, Tobacco Use Disorder, Vascular Diseases, Wasting Syndrome, 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 14. 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
alvespimycin	HSP90AB1, HSP90AA1	83	Phase 1: Burkitt Lymphoma, Hodgkin Disease, Intestinal Neoplasms, Intraocular Lymphoma, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia-Lymphoma, Adult T-Cell, Lymphoma, Lymphoma, B-Cell, Marginal Zone, Lymphoma, Follicular, Lymphoma, Large B-Cell, Diffuse, Lymphoma, Large-Cell, Anaplastic, Lymphoma, Mantle-Cell, Lymphoma, Non-Hodgkin, Lymphoma, T-Cell, Lymphoma, T-Cell, Cutaneous, Mycosis Fungoides, Sezary Syndrome, Waldenstrom Macroglobulinemia
onalespib	HSP90AB1, HSP90AA1	83	N/A
retaspimycin	HSP90AB1, HSP90AA1	83	Phase 2: Neoplasms, Prostatic Neoplasms
Cerulenin	FASN	81	N/A
Atovaquone	CCND1, PARP1, EIF4EBP1, EIF2S1, MYC, CDKN1A	81	Phase 4: Malaria, Rabies

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



Table 15. 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
(CHLOROACETYL)CARBAMIC ACID (3R,4S,5S,5R)-5-METHOXY-4-[(2R,3R)-2-METHYL-3-(3-METHYL-2-BUTENYL)OXIRANYL]-1-OXASPIRO[2.5]OCT-6-YL ESTER	FKBP1A, FKBP4	91	5.63E-2
1-Anilino-8-Naphthalene Sulfonate	FASN, PRDX4	84	4.13E-2
Pentosan Polysulfate	FKBP1A, FKBP4	52	1.22E-2



Table 16. 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
Myo-Inositol	FASN, FKBP1A, PRDX4, MYC, SPHK1, FKBP4	99	0.14
Gluconolactone	FKBP1A, MYC, SPHK1, FKBP4	99	0.16
D-Galctopyranosyl-1-On	FKBP1A, MYC, SPHK1, FKBP4	99	0.16
Methyl alpha-galactoside	FKBP1A, MYC, SPHK1, FKBP4	97	9.88E-2
Methyl beta-galactoside	FKBP1A, MYC, SPHK1, FKBP4	97	9.88E-2

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Temsirolimus, alvespimycin, (CHLOROACETYL)CARBAMIC ACID (3R,4S,5S,5R)-5-METHOXY-4-[(2R,3R)-2-METHYL-3-(3-METHYL-2-BUTENYL)OXIRANYL]-1-OXASPIRO[2.5]OCT-6-YL ESTER and Myo-Inositol. These drugs were selected for acting on the following targets: EIF4E, HSP90AB1 and FKBP4, 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.

Supplementary drug info

In addition to the approved and repurposed drugs proposed by Genome Enhancer, below the **Supplementary drug info** table is given, which contains an extended list of drugs used for treatment of neoplasms. Those drugs which were predicted by Genome Enhancer as prospective treatment candidates for the studied case (both approved and repurposed) have a respective **Predicted Drug Score** assigned to them. This value on a scale from 1 to 100 reflects the potential activity of the respective drug on the overall molecular mechanism of the studied pathology. The **Predicted Drug Score** column contains the "-" (Not Identified) value in case the drug targets of the respective treatment were not found in the molecular mechanism of the studied pathology.

Table 17. Supplementary drug info: extended list of drugs used for treatment of neoplasms with respective drug scores predicted for the studied pathology.

Drug	Disease	Predicted Drug Score
Abarelix	Prostatic Neoplasms	-
Abemaciclib	Breast Neoplasms	73
Abiraterone	Prostatic Neoplasms, Castration-Resistant	-
Abiraterone acetate	Prostatic Neoplasms, Castration-Resistant	-
Acalabrutinib	Lymphoma, Mantle-Cell	-

Acitretin	Psoriasis	-
Ado-trastuzumab emtansine	Breast Neoplasms Neoplasms	46
Afatinib	Carcinoma, Non-Small-Cell Lung	-
Aflibercept	Colorectal Neoplasms Diabetic Retinopathy Edema Vascular Diseases Wet Macular Degeneration	-
Alectinib	Carcinoma, Non-Small-Cell Lung	-
Alemtuzumab	Brain Abscess Leukemia, Lymphocytic, Chronic, B-Cell Multiple Sclerosis Multiple Sclerosis, Relapsing-Remitting Sclerosis	-
Alitretinoin	Sarcoma, Kaposi	-
Alpelisib	Breast Neoplasms	-
Altretamine	Ovarian Neoplasms	-
Aminolevulinic acid	Keratosis Keratosis, Actinic	-
Anagrelide	Thrombocythemia, Essential Thrombocytosis	-
Anastrozole	Breast Neoplasms Hypersensitivity Obesity Obesity, Morbid Recurrence Weight Loss	-
Apalutamide	Prostatic Neoplasms, Castration-Resistant	-
Aprepitant	Nausea Neoplasms Postoperative Nausea and Vomiting	-
Arsenic trioxide	Leukemia, Promyelocytic, Acute	39
Atezolizumab	Carcinoma, Non-Small-Cell Lung Carcinoma, Transitional Cell Triple Negative Breast Neoplasms	-
Avelumab	Carcinoma, Merkel Cell Carcinoma, Renal Cell Carcinoma, Transitional Cell	-
Axitinib	Carcinoma, Renal Cell	-
Azacitidine	Anemia, Refractory Anemia, Refractory, with Excess of Blasts Leukemia, Myelomonocytic, Chronic Myelodysplastic Syndromes Preleukemia Syndrome	40
Belinostat	Lymphoma, T-Cell, Peripheral	54
Bendamustine	Leukemia, Lymphocytic, Chronic, B-Cell Leukemia, Lymphoid	-
Bevacizumab	Breast Neoplasms Colonic Neoplasms Colorectal Neoplasms Corneal Neovascularization Diabetic Retinopathy Dilatation, Pathologic Edema Epistaxis Glaucoma Hemorrhage Macular Degeneration Macular Edema Neoplasm Metastasis Neoplasms Neovascularization, Pathologic Optic Nerve Diseases Pterygium Rectal Neoplasms Retinal Detachment Retinal Diseases Retinal Vein Occlusion Telangiectasia, Hereditary Hemorrhagic Telangiectasis Vitreous Hemorrhage	-
Bexarotene	Lymphoma, T-Cell Lymphoma, T-Cell, Cutaneous	-
Bicalutamide	Prostatic Neoplasms	-
Binimetinib	Melanoma	-
Blinatumomab	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Bortezomib	Brain Abscess Glomerulonephritis Glomerulonephritis, IGA Kidney Diseases Multiple Myeloma Neoplasms, Plasma Cell Nephritis Renal Insufficiency	-
Bosutinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	-
Brentuximab vedotin	Hodgkin Disease Lymphoma Lymphoma, Large-Cell, Anaplastic Lymphoma, T-Cell, Peripheral	-
Brigatinib	Carcinoma, Non-Small-Cell Lung	30
Buserelin	Prostatic Neoplasms	-
Cabazitaxel	Prostatic Neoplasms, Castration-Resistant	68
Cabergoline	Drug-Related Side Effects and Adverse Reactions Pituitary Neoplasms	-
Cabozantinib	Thyroid Neoplasms	-
Capecitabine	Breast Neoplasms Colonic Neoplasms Colorectal Neoplasms	-
Carboplatin	Carcinoma, Non-Small-Cell Lung Lung Neoplasms Neoplasms Neuroendocrine Tumors Ovarian Neoplasms Retinoblastoma	-
Carfilzomib	Multiple Myeloma	33
Carmustine	Astrocytoma Glioblastoma Hodgkin Disease Medulloblastoma Multiple Myeloma Neoplasms	-
Ceritinib	Carcinoma, Non-Small-Cell Lung	30

Cetuximab	Colorectal Neoplasms	-
Cinacalcet	Anemia Calcinosis Cardiovascular Diseases Hyperparathyroidism Hyperparathyroidism, Secondary Kidney Diseases Kidney Failure, Chronic Neoplasm Metastasis Neoplasms Parathyroid Neoplasms Renal Insufficiency Vascular Calcification Vascular Diseases Vision Disorders	-
Cisplatin	Carcinoma, Squamous Cell Neoplasms Uterine Cervical Neoplasms Carcinoma, Non-Small-Cell Lung Esophageal Neoplasms Carcinoma	-
Cladribine	Leukemia, Hairy Cell	16
Clofarabine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	2
Cobimetinib	Melanoma	-
Copanlisib	Lymphoma, Follicular	43
Crizotinib	Carcinoma, Non-Small-Cell Lung	-
Cyproterone acetate	Prostatic Neoplasms	-
Dabrafenib	Melanoma	-
Dacomitinib	Carcinoma, Non-Small-Cell Lung	34
Daratumumab	Multiple Myeloma	-
Dasatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive Leukemia, Myeloid, Chronic-Phase Precursor Cell Lymphoblastic Leukemia-Lymphoma	14
Decitabine	Anemia, Refractory Anemia, Refractory, with Excess of Blasts Leukemia, Myelomonocytic, Chronic Myelodysplastic Syndromes	-
Degarelix	Cardiovascular Diseases Prostatic Neoplasms Vascular Diseases	-
Denosumab	Arthritis, Rheumatoid Bone Diseases Bone Diseases, Metabolic Breast Neoplasms Hyperparathyroidism Hyperparathyroidism, Primary Metabolic Diseases Neoplasm Metastasis Neoplasms Osteoporosis Osteoporosis, Postmenopausal Prostatic Neoplasms	-
Dexrazoxane	Breast Neoplasms Cardiomyopathies	-
Dienogest	Menorrhagia	-
Dinutuximab	Neuroblastoma	-
Docetaxel	Breast Neoplasms Carcinoma, Non-Small-Cell Lung Prostatic Neoplasms Squamous Cell Carcinoma of Head and Neck Stomach Neoplasms	-
Doxorubicin	Neoplasms Multiple Myeloma Carcinoma, Ovarian Epithelial Ovarian Neoplasms Leukemia, Lymphoid Breast Neoplasms Lymphoma, Follicular Thyroid Neoplasms Triple Negative Breast Neoplasms Glioma	-
Durvalumab	Carcinoma, Non-Small-Cell Lung Carcinoma, Transitional Cell	-
Dutasteride	Alcoholism Hyperplasia Hypertrophy Neoplasms Prostatic Hyperplasia	-
Duvelisib	Leukemia, Lymphocytic, Chronic, B-Cell Lymphoma, Follicular	-
Elotuzumab	Multiple Myeloma	-
Enasidenib	Leukemia, Myeloid, Acute	-
Encorafenib	Colorectal Neoplasms Melanoma	-
Enfortumab vedotin	Carcinoma, Transitional Cell Neoplasms	-
Entrectinib	Carcinoma, Non-Small-Cell Lung	-
Enzalutamide	Prostatic Neoplasms Prostatic Neoplasms, Castration-Resistant	-
Epirubicin	Breast Neoplasms	-
Erdafitinib	Urinary Bladder Neoplasms	67
Eribulin	Breast Neoplasms Drug-Related Side Effects and Adverse Reactions Neoplasms	-
Erlotinib	Carcinoma, Non-Small-Cell Lung Neoplasms Pancreatic Neoplasms	29
Erlotinib hydrochloride	Carcinoma, Non-Small-Cell Lung Gastrointestinal Stromal Tumors	-
Estramustine	Prostatic Neoplasms	-
Ethinyl Estradiol	Acne Vulgaris Neoplasms	-
Everolimus	Angiomyolipoma Arthrogryposis Astrocytoma Breast Neoplasms Carcinoma, Renal Cell Cysts Idiopathic Pulmonary Fibrosis Kidney Diseases, Cystic Kidney Failure, Chronic Lipoma Neuroendocrine Tumors Primary Graft Dysfunction Sclerosis Tuberous Sclerosis	-
Exemestane	Breast Neoplasms	-
Fedratinib	Primary Myelofibrosis	-

Finasteride	Hyperplasia Neoplasms Prostatic Hyperplasia	-
Flavopiridol	Leukemia, Lymphocytic, Chronic, B-Cell	12
Fluorouracil	Skin Neoplasms Neoplasms, Basal Cell Neoplasms, Second Primary Neoplasms, Squamous Cell Neoplasms Colorectal Neoplasms Pancreatic Neoplasms	16
Fluoxymesterone	Breast Neoplasms Hypogonadism Puberty, Delayed	-
Flutamide	Premenstrual Dysphoric Disorder Premenstrual Syndrome Prostatic Neoplasms	47
Fulvestrant	Breast Neoplasms	-
Gefitinib	Carcinoma, Non-Small-Cell Lung	24
Gemcitabine	Breast Neoplasms Carcinoma, Non-Small-Cell Lung Ovarian Neoplasms Pancreatic Neoplasms	-
Gemtuzumab ozogamicin	Leukemia, Myeloid, Acute	-
Gilteritinib	Leukemia, Myeloid, Acute	59
Glasdegib	Leukemia, Myeloid, Acute	-
Goserelin	Atrophy Breast Neoplasms Bulbo-Spinal Atrophy, X-Linked Endometriosis Muscular Atrophy Myoma Prostatic Neoplasms	-
Histrelin	Puberty, Precocious	-
Homoharringtonine	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	58
Ibritumomab	Lymphoma, B-Cell Lymphoma, Follicular	-
Ibrutinib	Graft vs Host Disease Leukemia, Lymphocytic, Chronic, B-Cell Lymphoma, B-Cell, Marginal Zone Lymphoma, Mantle-Cell Waldenstrom Macroglobulinemia	-
Idarubicin	Leukemia, Myeloid, Acute	-
Idelalisib	Leukemia, Lymphocytic, Chronic, B-Cell Lymphoma, Follicular	-
Ifosfamide	Neoplasms	-
Imatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive Mastocytosis, Systemic Neoplasms	35
Inotuzumab ozogamicin	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Ipilimumab	Carcinoma, Renal Cell Melanoma	-
Irinotecan	Colorectal Neoplasms	18
Ivosidenib	Leukemia, Myeloid, Acute	-
Ixabepilone	Breast Neoplasms	-
Ixazomib	Multiple Myeloma	-
Lapatinib	Breast Neoplasms	32
Larotrectinib	Neoplasm Metastasis	-
Lenalidomide	Brain Abscess Lupus Erythematosus, Cutaneous Myelodysplastic Syndromes Neoplasms, Plasma Cell	15
Lenvatinib	Carcinoma, Hepatocellular Carcinoma, Renal Cell Thyroid Neoplasms	-
Letrozole	Breast Neoplasms Cysts Fibroma Myofibroma Myoma Ovarian Cysts Syndrome	-
Leuprolide	Hot Flashes Ovarian Hyperstimulation Syndrome Prostatic Neoplasms Puberty, Precocious	-
Levamisole	Ascariasis Colonic Neoplasms Helminthiasis	-
Levonorgestrel	Epilepsy Hyperplasia Menorrhagia	-
Lomustine	Brain Neoplasms Hodgkin Disease	-
Lonafarnib	Breast Neoplasms Carcinoma, Non-Small-Cell Lung Central Nervous System Neoplasms Colorectal Neoplasms Gliosarcoma Head and Neck Neoplasms Leukemia, Myelomonocytic, Chronic Liver Neoplasms Lymphoma Myelodysplastic Syndromes Ovarian Neoplasms Urethral Neoplasms Urinary Bladder Neoplasms	-
Lorlatinib	Carcinoma, Non-Small-Cell Lung	-
Masoprocol	Keratosis, Actinic	-
Medroxyprogesterone Acetate	Depression Depression, Postpartum Depressive Disorder Metrorrhagia Neoplasms Uterine Hemorrhage	-
Megestrol acetate	Acquired Immunodeficiency Syndrome Bites and Stings Breast Neoplasms Pain Wasting Syndrome	-
Methotrexate	Neoplasms Breast Neoplasms Head and Neck Neoplasms Ovarian Neoplasms Lymphoma, T-Cell, Peripheral Brain Neoplasms Colorectal Neoplasms Neuroblastoma Carcinoma, Squamous Cell	38

Methyltestosterone	Breast Neoplasms Hypogonadism Puberty, Delayed	-
Midostaurin	Leukemia, Mast-Cell Leukemia, Myeloid, Acute Mastocytosis, Systemic	13
Mitotane	Adrenocortical Carcinoma	-
Mitoxantrone	Autoimmune Diseases Autoimmune Diseases of the Nervous System Demyelinating Autoimmune Diseases, CNS Immune System Diseases Leukemia, Myeloid, Acute Multiple Sclerosis Myelitis Myelitis, Transverse Nervous System Diseases Neuromyelitis Optica Prostatic Neoplasms, Castration-Resistant	-
Mogamulizumab	Mycosis Fungoides Neoplasms Sezary Syndrome	-
Moxetumomab pasudotox	Leukemia, Hairy Cell Neoplasms	-
Necitumumab	Carcinoma, Non-Small-Cell Lung Neoplasms	-
Nelarabine	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma	-
Neratinib	Breast Neoplasms	39
Nilotinib	Blast Crisis Leukemia, Myelogenous, Chronic, BCR-ABL Positive Leukemia, Myeloid, Chronic-Phase	-
Nilutamide	Prostatic Neoplasms	-
Nintedanib	Fibrosis Idiopathic Pulmonary Fibrosis	-
Niraparib	Carcinoma, Ovarian Epithelial Fallopian Tube Neoplasms Peritoneal Neoplasms	80
Nivolumab	Carcinoma, Non-Small-Cell Lung Kidney Neoplasms Neoplasms Lung Neoplasms Melanoma	-
Obinutuzumab	Leukemia, Lymphocytic, Chronic, B-Cell	-
Octreotide	Acromegaly Adenoma Ascites Carcinoid Tumor Fistula Pancreatic Fistula Pituitary Diseases Renal Insufficiency Vipoma	-
Ofatumumab	Leukemia, Lymphocytic, Chronic, B-Cell	-
Olaparib	Breast Neoplasms Carcinoma, Ovarian Epithelial Fallopian Tube Neoplasms Ovarian Neoplasms Pancreatic Neoplasms Peritoneal Neoplasms Prostatic Neoplasms, Castration-Resistant	68
Olaratumab	Sarcoma	-
Osimertinib	Carcinoma, Non-Small-Cell Lung	-
Oxaliplatin	Colonic Neoplasms Colorectal Neoplasms Neoplasms Rectal Neoplasms	-
Paclitaxel	Acute Coronary Syndrome Angina Pectoris Arteriosclerosis Breast Neoplasms Carcinoma, Non-Small-Cell Lung Cardiovascular Diseases Coronary Artery Disease Coronary Disease Coronary Stenosis Heart Diseases Myocardial Ischemia Ovarian Neoplasms Vascular Diseases	75
Palbociclib	Breast Neoplasms	-
Panitumumab	Colorectal Neoplasms	-
Panobinostat	Multiple Myeloma	-
Pazopanib	Carcinoma Carcinoma, Renal Cell Sarcoma	32
Pembrolizumab	Carcinoma, Hepatocellular Carcinoma, Merkel Cell Carcinoma, Non-Small-Cell Lung Carcinoma, Renal Cell Carcinoma, Transitional Cell Hodgkin Disease Melanoma Neoplasms Stomach Neoplasms	-
Pemetrexed	Carcinoma, Non-Small-Cell Lung Mesothelioma	-
Pentostatin	Leukemia, Hairy Cell	-
Pertuzumab	Breast Neoplasms	-
Pomalidomide	Multiple Myeloma	62
Ponatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
Pralatrexate	Lymphoma, T-Cell, Peripheral	-
Radium Ra 223 Dichloride	Prostatic Neoplasms, Castration-Resistant	-
Ramucirumab	Stomach Neoplasms	-
Rasburicase	Hyperuricemia Leukemia Lymphoma Neoplasms Syndrome Tumor Lysis Syndrome	-
Regorafenib	Colorectal Neoplasms	-
Relugolix	Prostatic Neoplasms	-
Ribociclib	Breast Neoplasms	-

Rituximab	Arthritis Arthritis, Rheumatoid Granulomatosis with Polyangiitis Leukemia Leukemia, Lymphoid Lymphoma Lymphoma, B-Cell Lymphoma, Follicular Lymphoma, Non-Hodgkin Myelitis Neuromyelitis Optica Purpura Purpura, Thrombocytopenic Purpura, Thrombocytopenic, Idiopathic Thrombocytopenia	-
Romidepsin	Lymphoma, T-Cell, Cutaneous	10
Rucaparib	Carcinoma, Ovarian Epithelial Fallopian Tube Neoplasms Peritoneal Neoplasms Prostatic Neoplasms, Castration-Resistant	33
Ruxolitinib	Graft vs Host Disease Polycythemia Polycythemia Vera Primary Myelofibrosis Thrombocytosis	-
Selinexor	Multiple Myeloma	-
Selumetinib	Neurofibromatosis 1	-
Siltuximab	Giant Lymph Node Hyperplasia	-
Sirolimus	Angiomyolipoma Constriction, Pathologic Coronary Restenosis Eye Diseases Immune System Diseases Kidney Failure, Chronic Lipoma Tuberous Sclerosis	87
Sonidegib	Carcinoma, Basal Cell	-
Sorafenib	Carcinoma, Hepatocellular Carcinoma, Renal Cell Thyroid Neoplasms	21
Sunitinib	Adenoma Carcinoma, Renal Cell Digestive System Neoplasms Gastrointestinal Neoplasms Gastrointestinal Stromal Tumors Intestinal Neoplasms	17
Talazoparib	Breast Neoplasms	54
Tamoxifen	Breast Diseases Cystic Fibrosis Cysts Fibroadenoma Fibrocystic Breast Disease Hemorrhage Menorrhagia Menstruation Disturbances Metrorrhagia Neoplasms	32
Tamsulosin	Calculi Coronary Artery Disease Heart Diseases Hernia Hernia, Inguinal Inflammation Ischemia Lithiasis Lower Urinary Tract Symptoms Myocardial Ischemia Prostatic Hyperplasia Ureteral Calculi Urinary Calculi Urolithiasis Urologic Diseases	-
Temozolomide	Astrocytoma Nervous System Neoplasms	-
Temsirolimus	Carcinoma, Renal Cell	95
Teniposide	Precursor Cell Lymphoblastic Leukemia-Lymphoma	3
Thalidomide	Brain Abscess Immune System Diseases Multiple Myeloma Neoplasms, Plasma Cell	-
Tivozanib	Carcinoma, Renal Cell	-
Tocilizumab	Arthritis Arthritis, Juvenile Arthritis, Rheumatoid Behavior Cytokine Release Syndrome Giant Cell Arteritis Neurobehavioral Manifestations Oral Manifestations Psychotic Disorders Schizophrenia Tic Disorders	-
Topotecan	Small Cell Lung Carcinoma	-
Toremifene	Breast Neoplasms	-
Trabectedin	Leiomyosarcoma Liposarcoma	-
Trametinib	Carcinoma, Non-Small-Cell Lung Melanoma	46
Trastuzumab	Breast Neoplasms Neoplasms	-
Tretinoin	Lentigo	8
Triptorelin	Fatty Liver Hypogonadism Infertility, Female Prostatic Neoplasms	47
Tucatinib	Breast Neoplasms	-
Valrubicin	Urinary Bladder Neoplasms	-
Vandetanib	Thyroid Neoplasms	22
Vemurafenib	Melanoma	-
Venetoclax	Leukemia, Lymphocytic, Chronic, B-Cell Leukemia, Myeloid, Acute	-
Vinblastine	Glioma	5
Vincristine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	60
Vinorelbine	Carcinoma, Non-Small-Cell Lung	-
Vismodegib	Carcinoma, Basal Cell	-
Vorinostat	Lymphoma, T-Cell, Cutaneous	38
Zoledronate	Arthritis Bone Marrow Diseases Brain Abscess Chronic Kidney Disease-Mineral and Bone Disorder Chronic Periodontitis HIV	-

6. Conclusion

We applied the software package "Genome Enhancer" to a multi-omics data set that contains *transcriptomics* and *proteomics* data. The study is done in the context of *Neoplasm Metastasis and Osteosarcoma*. 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:



Temsirolimus, alvespimycin, (CHLOROACETYL)CARBAMIC ACID (3R,4S,5S,5R)-5-METHOXY-4-[(2R,3R)-2-METHYL-3-(3-METHYL-2-BUTENYL)OXIRANYL]-1-OXASPIRO[2.5]OCT-6-YL ESTER and Myo-Inositol

These drugs were selected for acting on the following targets: EIF4E, HSP90AB1 and FKBP4, 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:



cyclinD1(h):Cdk6, cyclinD1(h):Cdk6, FKBP4 and MKK3

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: Myo-Inositol, Imatinib, AGI-1067, Sirolimus, Atovaquone, Dipyridamole, (CHLOROACETYL)CARBAMIC ACID (3R,4S,5S,5R)-5-METHOXY-4-[(2R,3R)-2-METHYL-3-(3-METHYL-2-BUTENYL)OXIRANYL]-1-OXASPIRO[2.5]OCT-6-YL ESTER, Temsirolimus and 3,5-Diaminophthalhydrazide. 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.

- cyclinD1(h):Cdk6
- cyclinD1(h):Cdk6
- FKBP4
- MKK3

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 2023.2 (geneXplain GmbH, Wolfenbüttel, Germany) (<https://genexplain.com/transfac>).

The master regulator search uses the **TRANSPATH®** database (BIOBASE), release 2023.2 (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 2023.2 (<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" ($T\text{-score}_{PSD}$),
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:

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

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

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

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

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, $\text{phase}(p, d)$ 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" ($T\text{-score}$) is calculated as follows:

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

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 $G(m)$; $\text{optWeight}(g)$ 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" ($D\text{-score}$) 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, Chekmenov D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. **2006**;34(90001):D108-D110. doi:10.1093/nar/gkj143
7. Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res*. **2003**;31(13):3576-3579. doi:10.1093/nar/gkg585
8. Waleev T, Shtokalo D, Konovalova T, Voss N, Cheremushkin E, Stegmaier P, Kel-Margoulis O, Wingender E, Kel A. Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res*. **2006**;34(Web Server issue):W541-5.
9. Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res*. **2006**;34(90001):D546-D551. doi:10.1093/nar/gkj107
10. Boyarskikh U, Pintus S, Mandrik N, Stelmashenko D, Kiselev I, Evshin I, Sharipov R, Stegmaier P, Kolpakov F, Filipenko M, Kel A. Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3. *BMC Med Genomics*. **2018**;11(1):12. doi:10.1186/1471-2105-7-s2-s13
11. 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.
12. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal*. **2006**;50(2):66-75 (russ)
13. Filimonov D, Poroikov V, Borodina Y, Gloriovova 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 \(up-regulated genes in Myc_induce vs. Control\).](#)
4. [Supplementary table 4 - Detailed report. Composite modules and master regulators \(down-regulated genes in Myc_induce vs. Control\).](#)
5. [Supplementary table 5 - Detailed report. Pharmaceutical compounds and drug targets.](#)

Disclaimer

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

The compounds predicted to be active against the identified drug targets in the report are not guaranteed to be active against any particular patient's condition. GeneXplain GmbH does not give any assurances or guarantees regarding the treatment information and conclusions given in the report. There is no guarantee that any third party will provide a refund for any of the treatment decisions made based on these results. None of the listed compounds was checked by Genome Enhancer for adverse side-effects or even toxic effects.

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

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

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

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