PSMA7 and DUSP9 are promising druggable targets for treating Ovarian Neoplasms that control activity of REL, STAT3 and NFATC3 transcription factors on of differentially expressed genes

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 01/10/2020 ; Run on 06/09/2021 ; Report generated on 06/09/2021

Genome Enhancer release 2.4 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2021.2)

Abstract

In the present study we applied the software package "Genome Enhancer" to a multiomics data set that contains *transcriptomics and epigenomics* data. The study is done in the context of *Ovarian Neoplasms*. 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: REL, STAT3, DDIT3, NFATC3, HDAC2 and HSF1. The subsequent network analysis suggested

- IKK-gamma
- Aurora-B
- 26S proteasome
- MKP-4
- Cdk1-isoform1:cyclinB1-isoform1

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: Imatinib, AT9283 and 2,5,7-Trihydroxynaphthoquinone.

1. Introduction

Recording "-omics" data to measure gene activities, protein expression or metabolic events is becoming a standard approach to characterize the pathological state of an affected organism or tissue. Increasingly, several of these methods are applied in a combined approach leading to large "multiomics" datasets. Still the challenge remains how to reveal the underlying molecular mechanisms that render a given pathological state different from the norm. The disease-causing mechanism can be described by a re-wiring of the cellular regulatory network, for instance as a result of a genetic or epigenetic alterations influencing the activity of relevant genes. Reconstruction of the disease-specific regulatory networks can help identify potential master regulators of the respective pathological process. Knowledge about these master regulators can point to ways how to block a pathological regulatory cascade. Suppression of certain molecular targets as components of these cascades may stop the pathological process and cure the disease.

Conventional approaches of statistical "-omics" data analysis provide only very limited information about the causes of the observed phenomena and therefore contribute little to the understanding of the pathological molecular mechanism. In contrast, the "upstream analysis" method [1-4] applied here has been deviced to provide a casual interpretation of the data obtained for a pathology state. This approach comprises two major steps: (1) analysing promoters and enhancers of differentially expressed genes for the transcription factors (TFs) involved in their regulation and, thus, important for the process under study; (2) re-constructing the signaling pathways that activate these TFs and identifying master regulators at the top of such pathways. For the first step, the database TRANSFAC® [6] is employed together with the TF binding site identification algorithms Match [7] and CMA [8]. The second step involves the signal transduction database TRANSPATH® [9] and special graph search algorithms [10] implemented in the software "Genome Enhancer".

The "upstream analysis" approach has now been extended by a third step that reveals known drugs suitable to inhibit (or activate) the identified molecular targets in the context of the disease under study. This step is performed by using information from HumanPSD™ database [5]. In addition, some known drugs and investigational active chemical compounds are subsequently predicted as potential ligands for the revealed molecular targets. They are predicted using a pre-computed database of spectra of biological activities of chemical compounds of a library of 2245 known drugs and investigational chemical compounds from HumanPSD™ database. The spectra of biological activities for these compounds are computed using the program PASS on the basis of a (Q)SAR approach [11-13]. These predictions can be used for the research purposes - for further drug development and drug repurposing initiatives.

2. Data

For this study the following experimental data was used:

Table 1. Experimental datasets used in the study

File name	Data type
GSM385721.CEL	Transcriptomics
GSM385722.CEL	Transcriptomics
GSM385723.CEL	Transcriptomics
GSM385724.CEL	Transcriptomics
GSM385725.CEL	Transcriptomics
GSM385726.CEL	Transcriptomics
GSM385727.CEL	Transcriptomics
GSM385728.CEL	Transcriptomics
GSM385729.CEL	Transcriptomics
GSM385730.CEL	Transcriptomics
GSM385747_CpG_NM.fixed.hg38.top300	Epigenomics

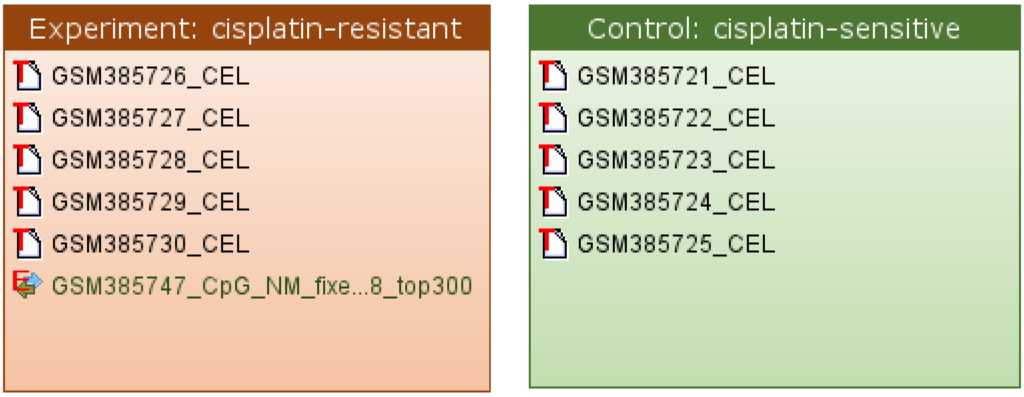


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

3. Results

We have compared the following conditions: Experiment: cisplatin-resistant *versus* Control: cisplatin-sensitive.

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: "Experiment: cisplatin-resistant" with "Control: cisplatin-sensitive". 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 4406 upregulated genes (LogFC>0) out of which 3611 genes were found as significantly upregulated (p-value<0.1) and 4457 downregulated genes (LogFC<0) out of which 3590 genes were significantly downregulated (p-value<0.1). See tables below for the top significantly up- and downregulated genes. Below we call **target genes** the full list of up- and downregulated genes revealed in our analysis (see tables in Supplementary section).

*Table 2. Top ten significant **up-regulated** genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.*

See full table →

ID	Gene symbol	Gene description	logFC	P.Value	adj.P.Val
ENSG00000123700	KCNJ2	potassium inwardly rectifying channel subfamily J member 2	5.38	8.93E-14	1.03E-10
ENSG00000064218	DMRT3	doublesex and mab-3 related transcription factor 3	4.03	1E-11	3.61E-9
ENSG00000099139	PCSK5	proprotein convertase subtilisin/kexin type 5	3.93	2.07E-14	3.4E-11
ENSG00000197705	KLHL14	kelch like family member 14	3.89	1.35E-12	6E-10
ENSG00000129038	LOXL1	lysyl oxidase like 1	3.54	3.29E-10	4.45E-8
ENSG00000133083	DCLK1	doublecortin like kinase 1	3.24	1.15E-12	5.53E-10
ENSG00000141431	ASXL3	ASXL transcriptional regulator 3	3.14	1.85E-11	5.28E-9
ENSG00000126950	TMEM35A	transmembrane protein 35A	3.05	2.24E-12	8.89E-10
ENSG00000164692	COL1A2	collagen type I alpha 2 chain	2.87	2.63E-10	4.08E-8
ENSG00000138378	STAT4	signal transducer and activator of transcription 4	2.86	4.05E-10	5E-8

*Table 3. Top ten significant **down-regulated** genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.*

See full table →

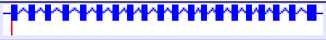
ID	Gene symbol	Gene description	logFC	P.Value	adj.P.Val
ENSG00000127324	TSPAN8	tetraspanin 8	-6.39	2.36E-15	6.78E-12
ENSG00000139292	LGR5	leucine rich repeat containing G protein-coupled receptor 5	-6.24	9.29E-18	1.07E-13
ENSG00000149968	MMP3	matrix metallopeptidase 3	-5.16	2.77E-13	2.45E-10
ENSG00000163359	COL6A3	collagen type VI alpha 3 chain	-5.08	8.79E-16	3.37E-12
ENSG00000169908	TM4SF1	transmembrane 4 L six family member 1	-4.92	2.55E-16	1.47E-12
ENSG00000153233	PTPRR	protein tyrosine phosphatase receptor type R	-4.6	8.72E-13	4.56E-10
ENSG00000166670	MMP10	matrix metallopeptidase 10	-4.45	1.46E-14	2.79E-11
ENSG00000106511	MEOX2	mesenchyme homeobox 2	-4.26	4.87E-12	1.87E-9
ENSG00000145431	PDGFC	platelet derived growth factor C	-4.14	4.94E-14	7.11E-11
ENSG00000060718	COL11A1	collagen type XI alpha 1 chain	-3.65	1.28E-10	2.42E-8

3.2. Regulatory regions of target genes

We mapped the uploaded Epigenomic peaks on the ***target genes*** and selected those peaks only that were found located in the body of the gene (in exons or introns of the genes) or in the 5000 nucleotide long flanking regions of the genes. In the tables below we demonstrate localization of such potential regulatory regions in the top up-regulated and down-regulated genes.

*Table 4. Top 5 **down-regulated** genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with epigenomic peaks.*

See full table →

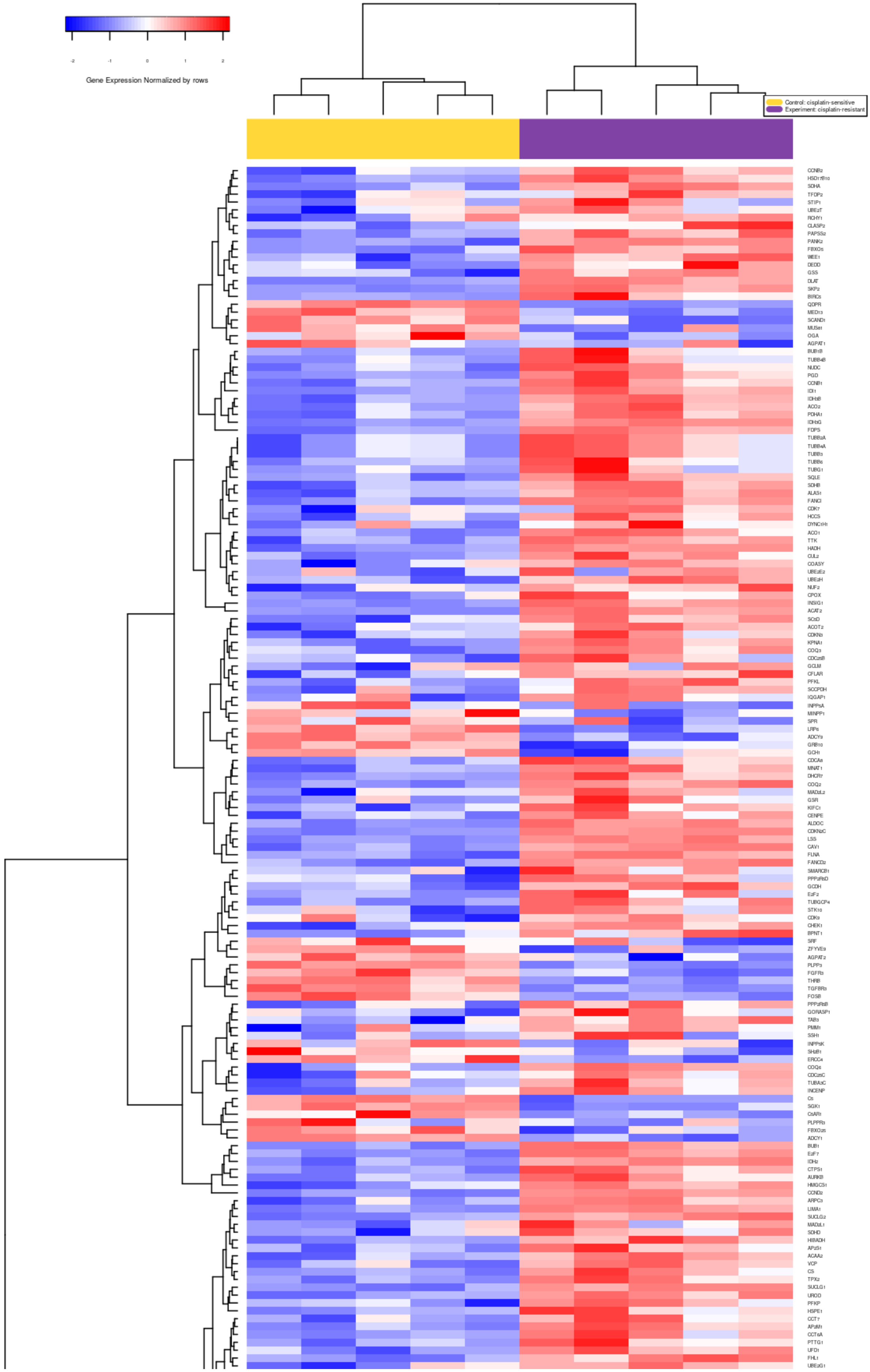
ID	Gene symbol	Gene schematic representation
ENSG00000170558	CDH2	
ENSG00000197822	OCLN	
ENSG00000118495	PLAGL1	
ENSG00000145476	CYP4V2	
ENSG00000237765	FAM200B	

3.3. 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 Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive

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



-2
-1
0
1
2
Gene Expression Normalized by rows
Control: cisplatin-sensitive
Experiment: cisplatin-resistant

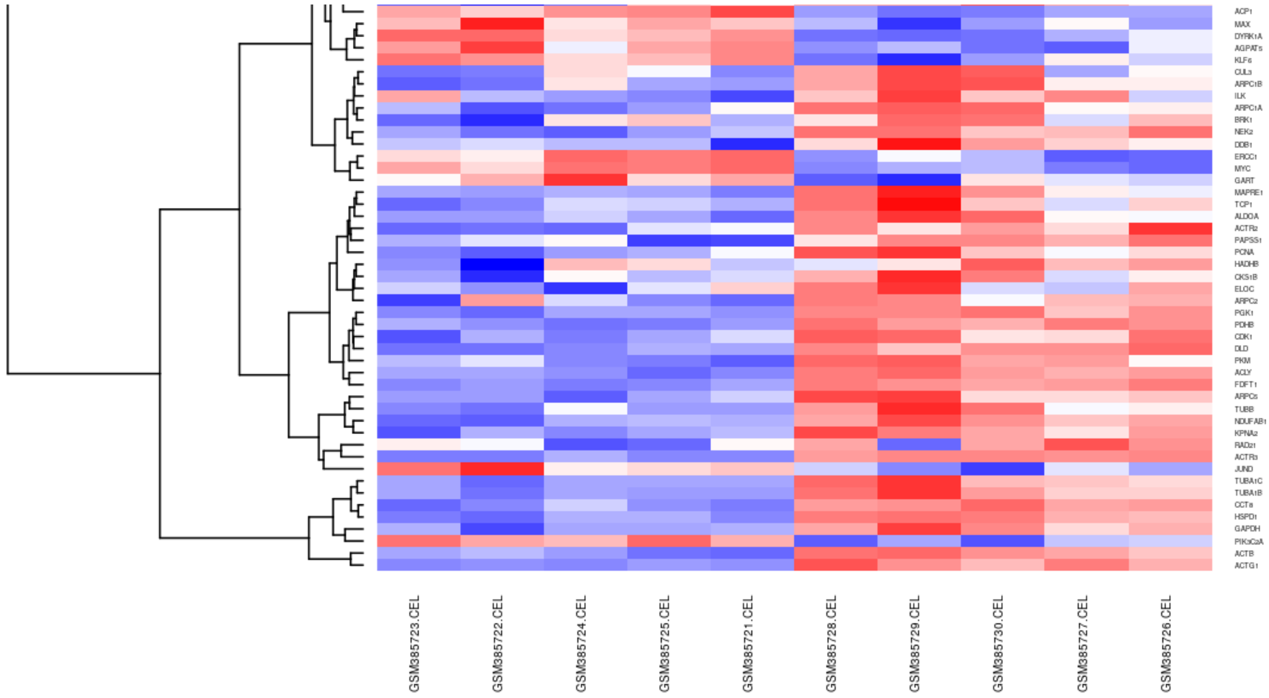


Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.

See full diagram →

Up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive:

3611 significant up-regulated genes were taken for the mapping.

GO (biological process)

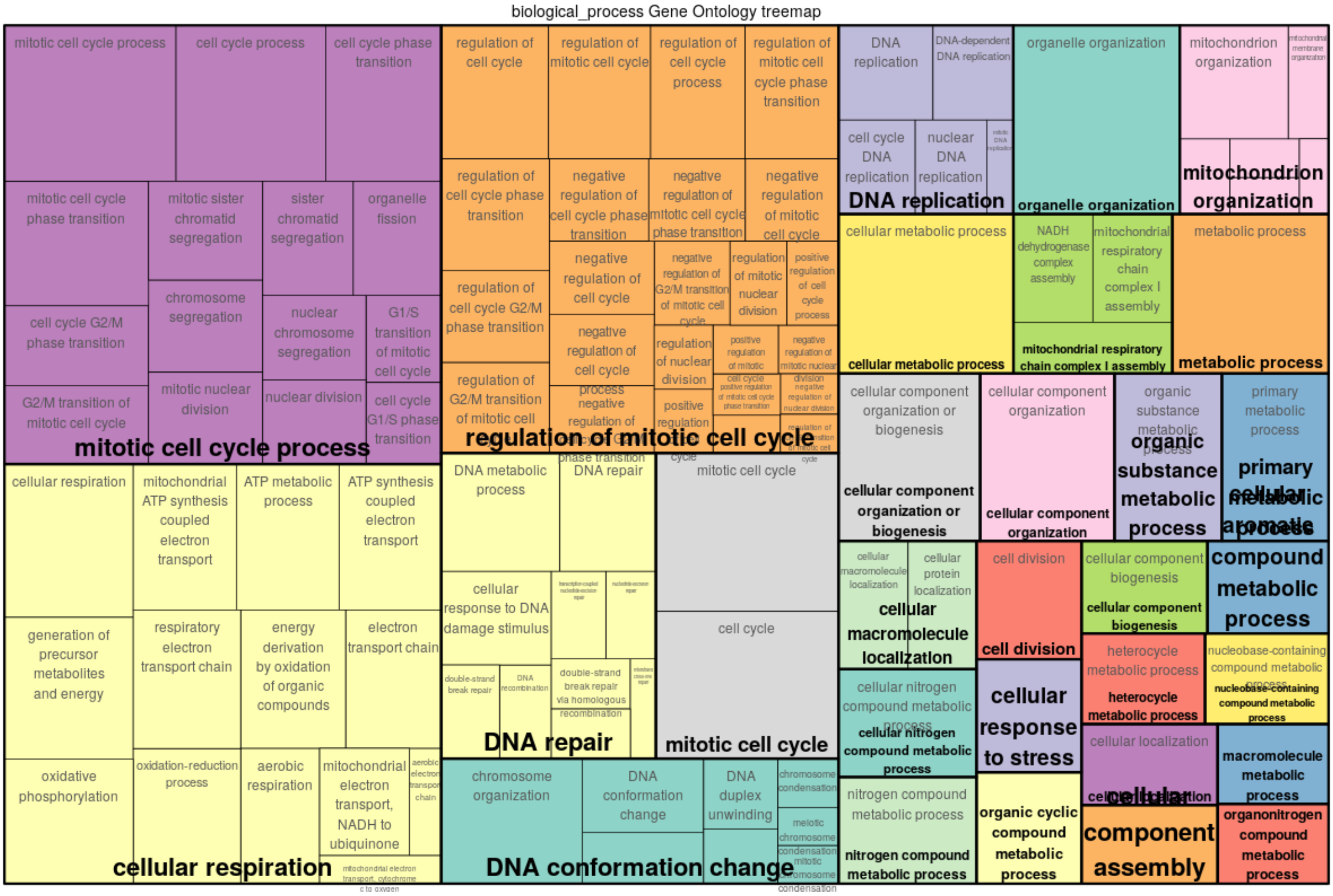


Figure 3. Enriched GO (biological process) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.

Full classification →

TRANSPATH® Pathways (2021.2)

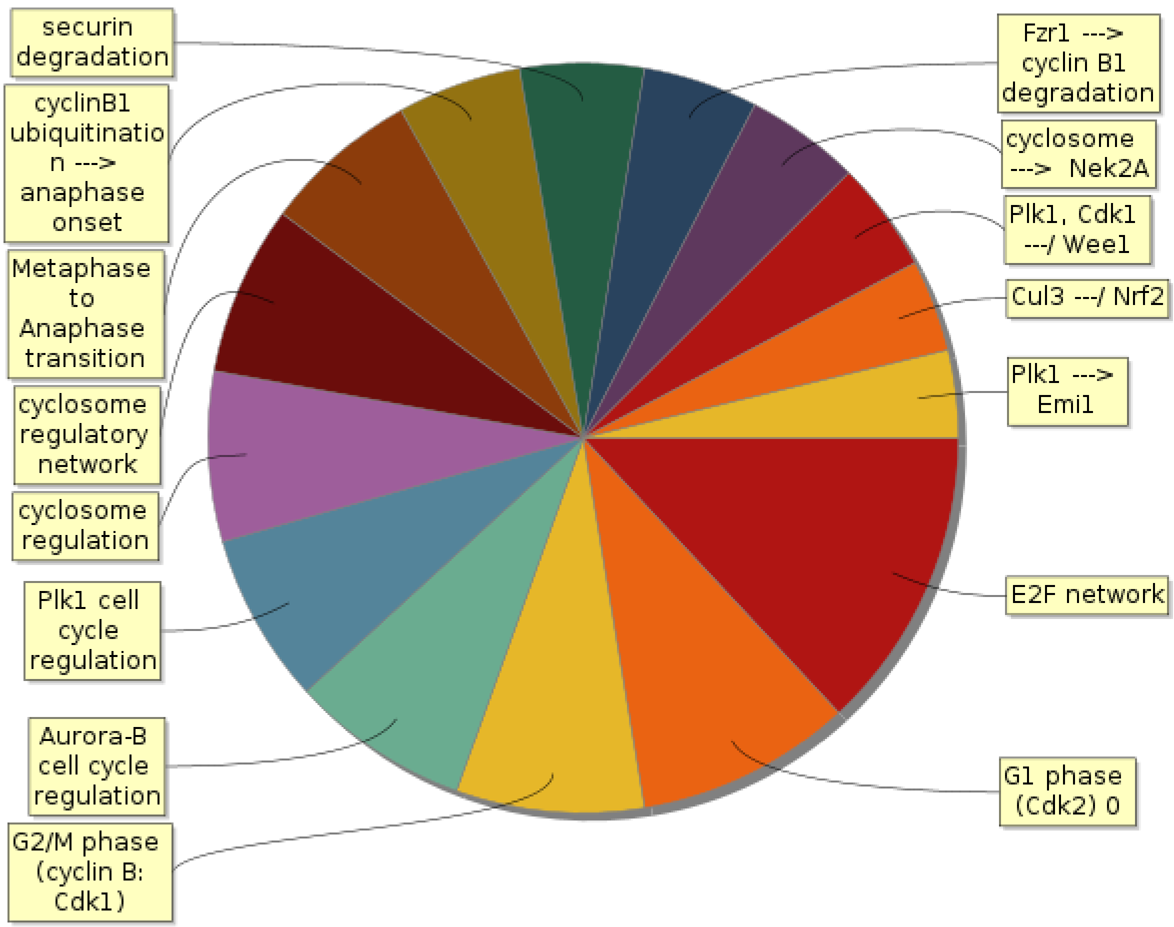


Figure 4. Enriched TRANSPATH® Pathways (2021.2) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.

Full classification →

HumanPSD(TM) disease (2021.2)

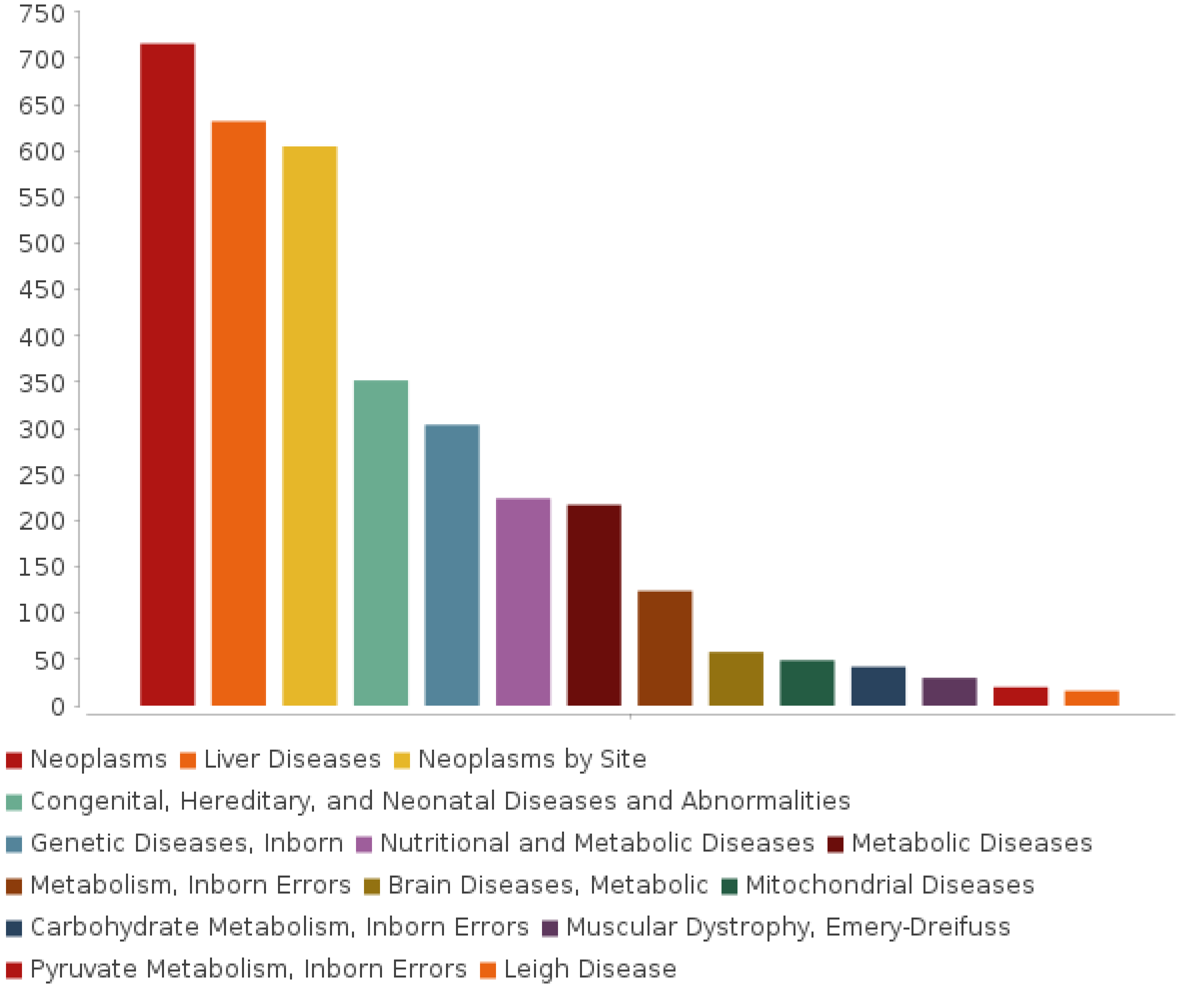


Figure 5. Enriched HumanPSD(TM) disease (2021.2) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. The size of the bars correspond to the number of bio-markers of the given disease found among the input set.

Full classification →

Down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive:

3590 significant down-regulated genes were taken for the mapping.

GO (biological process)

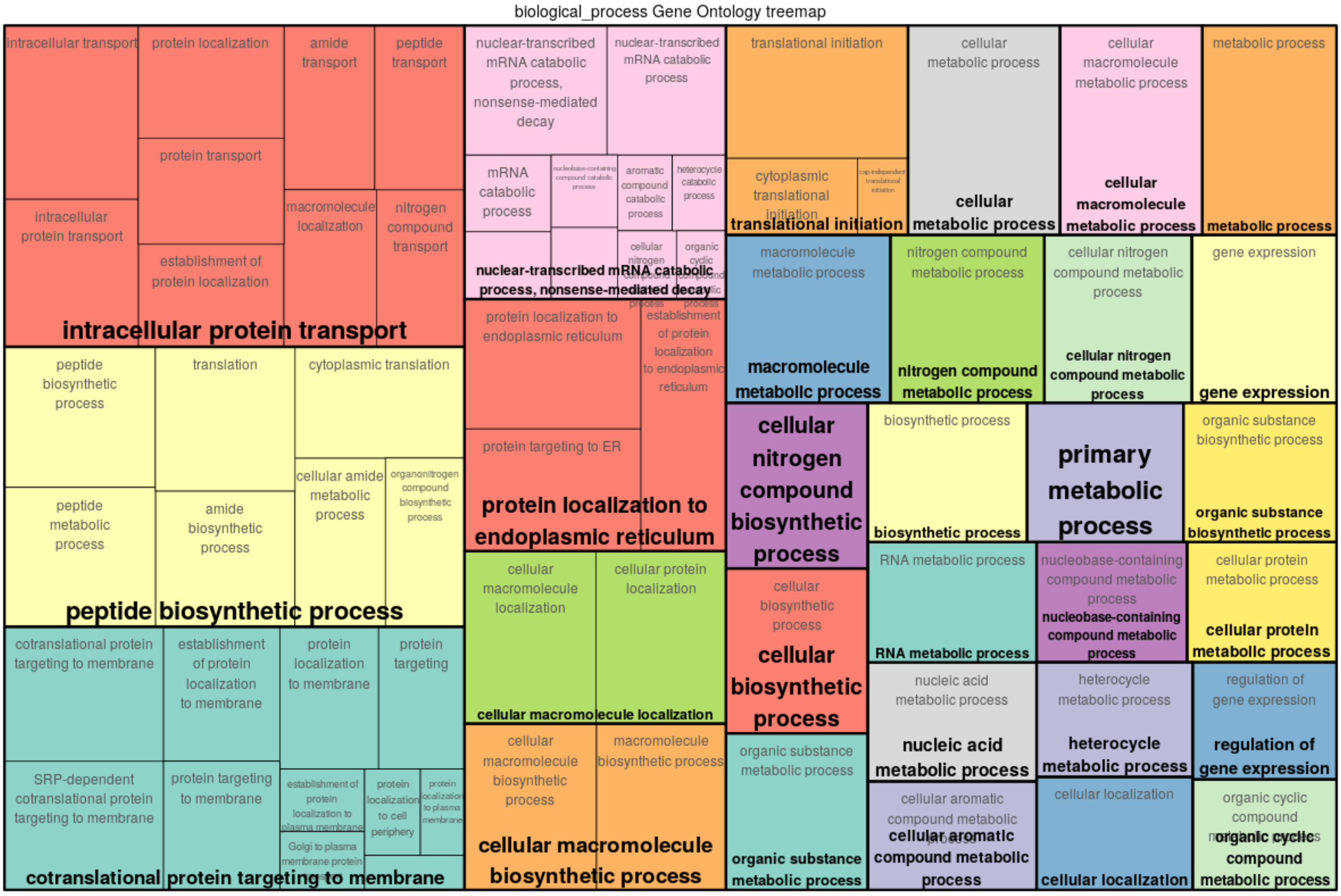


Figure 6. Enriched GO (biological process) of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.

Full classification →

TRANSPATH® Pathways (2021.2)

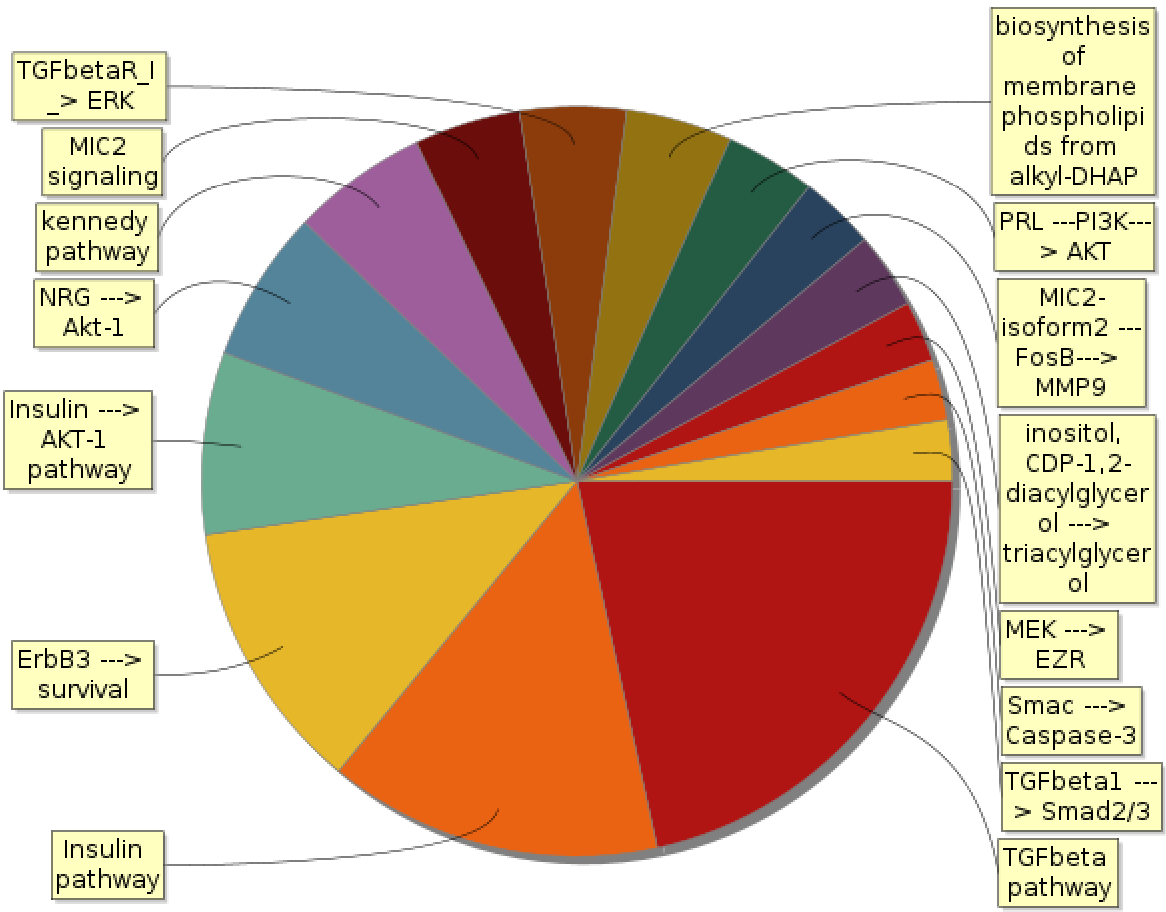


Figure 7. Enriched TRANSPATH® Pathways (2021.2) of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.

Full classification →

HumanPSD(TM) disease (2021.2)

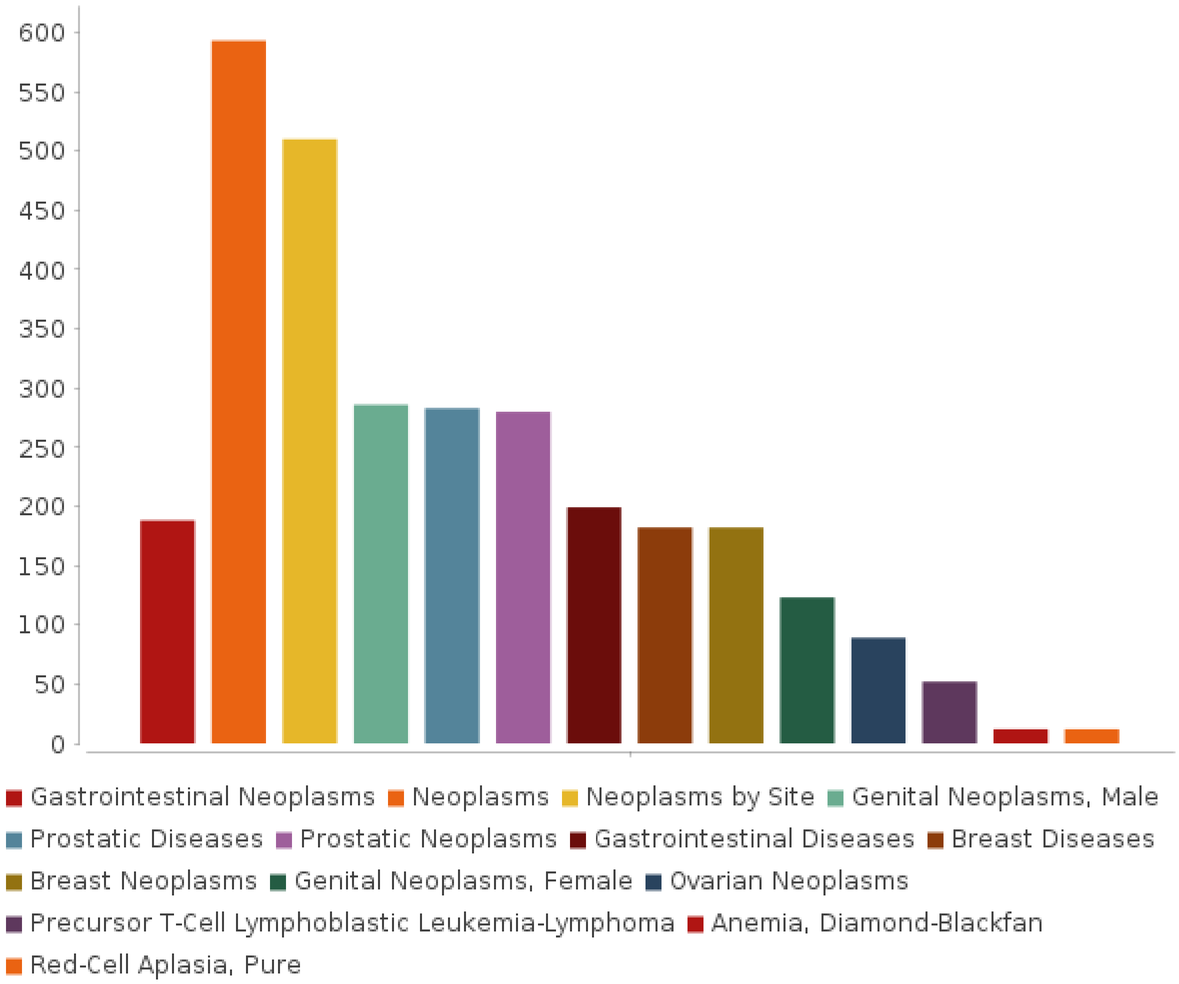
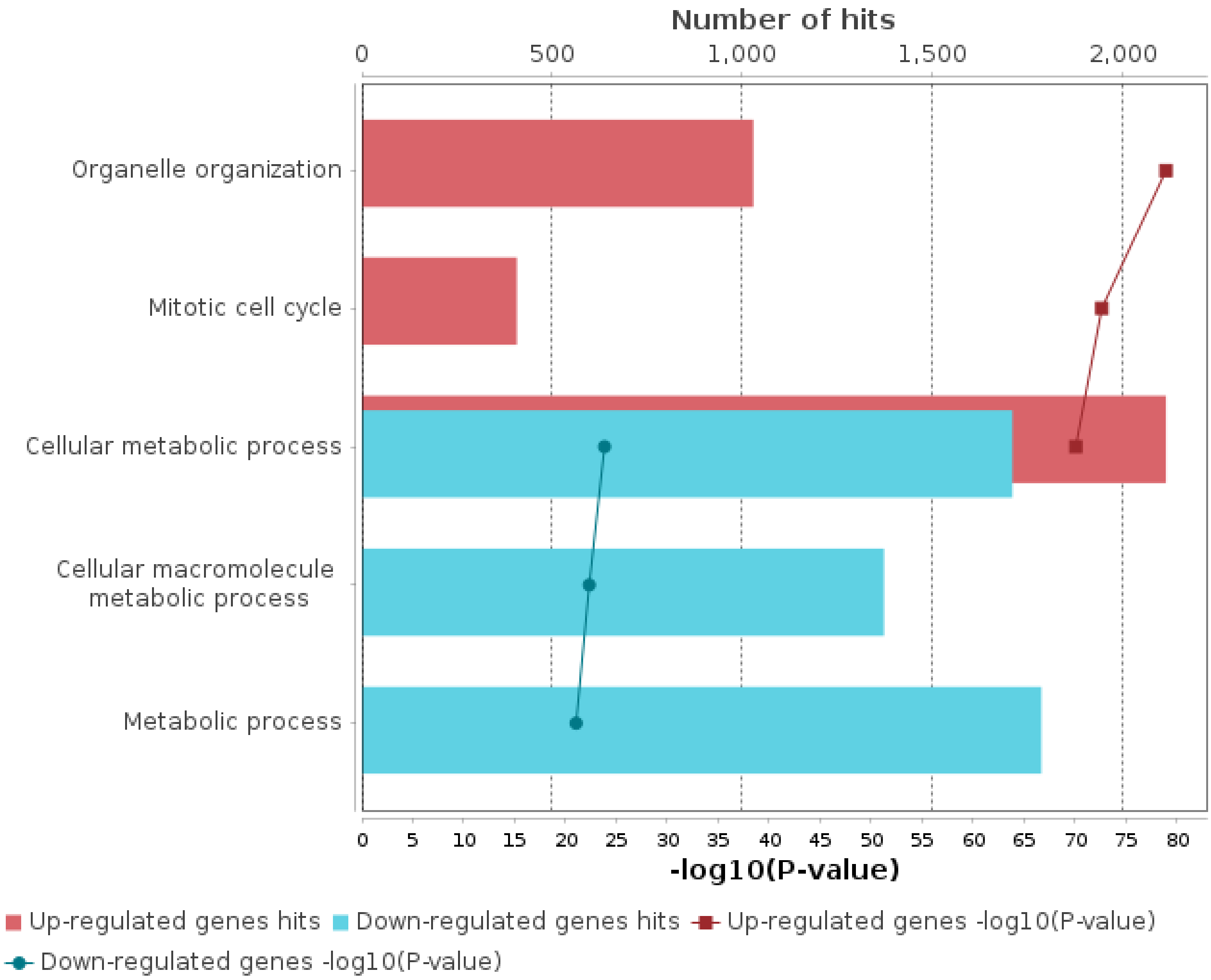


Figure 8. Enriched HumanPSD(TM) disease (2021.2) of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. The size of the bars correspond to the number of bio-markers of the given disease found among the input set.

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):



3.4. 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].

In the current work we use the Epigenomics data from the track(s) "GSM385747_CpG_NM.fixed.hg38.top300" to predict positions of potential ***enhancers*** regulating the differentially expressed genes revealed by comparative transcriptomics analysis. We took genomic regions -550bp upstream and 550bp downstream from the middle point of each interval of the track and check if these regions are located inside the 5kb flanking arias of the differentially expressed genes (or inside the body of the genes). In such cases, these genomic regions are used for the search for potential condition-specific enhancers. In all other cases when the differentially expressed genes did not contain epigenomic peaks in their body or in the 5kb flanking regions we used the upstream regulatory regions of these genes (-1000bp upstream and 100bp downstream of TSS) for the search for condition-specific enhancers.

We applied the Composite Module Analyst (CMA) [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 Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).

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.

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$EGR3_Q6
0.89; N=3

V$LEF1_Q5_01
0.96; N=2

V$CREL_01
0.85; N=2

V$RARA_16
0.83; N=3

V$CHOP_01
0.78; N=3

V$STAT3_01
0.77; N=2

Module width: 96

Module 2:

V$SLUG_Q6_01
0.98; N=3

V$AML2_Q3
0.97; N=2

V$NFYA_07
0.93; N=2

V$IRF8_Q6
0.97; N=2

V$FOXO3_05
0.78; N=1

V$ISL1_05
0.99; N=2

Module width: 116

Model score (-p*log10(pval)): 16.20
Wilcoxon p-value (pval): 7.39e-35
Penalty (p): 0.475
Average yes-set score: 5.77
Average no-set score: 4.57
AUC: 0.76
Separation point: 5.00
False-positive: 34.60%
False-negative: 25.33%

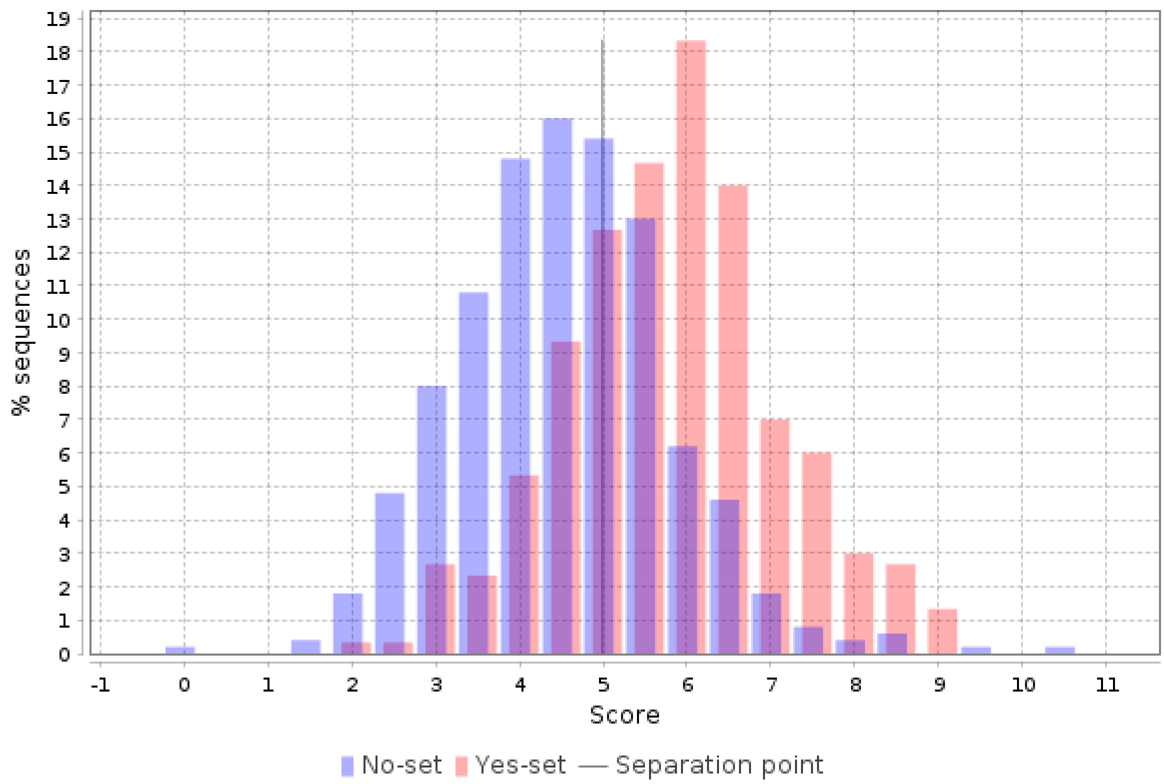


Table 5. List of top ten up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000131844	MCCC2	methylcrotonoyl-CoA carboxylase 2	10.92	egr-3(h), C/EBPalpha(h),CHOP-10(h), LEF-1(h), STAT3(h), c-Rel(h), NF-YA(h), IRF-8(h)...
ENSG00000130766	SESN2	sestrin 2	10.66	C/EBPalpha(h),CHOP-10(h), NR1B1(h), c-Rel(h), egr-3(h), LEF-1(h), IRF-8(h), FOXO3a(h)...
ENSG00000186399	GOLGA8R	golgin A8 family member R	9.87	IRF-8(h), FOXO3a(h), NF-YA(h), egr-3(h), LEF-1(h), C/EBPalpha(h),CHOP-10(h), STAT3(h)...
ENSG00000178115	GOLGA8Q	golgin A8 family member Q	9.86	NR1B1(h), egr-3(h), STAT3(h), LEF-1(h), C/EBPalpha(h),CHOP-10(h), NF-YA(h), FOXO3a(h)...
ENSG00000261247	GOLGA8T	golgin A8 family member T	9.86	NR1B1(h), egr-3(h), STAT3(h), LEF-1(h), C/EBPalpha(h),CHOP-10(h), NF-YA(h), FOXO3a(h)...
ENSG00000179938	GOLGA8J	golgin A8 family member J	9.85	NR1B1(h), egr-3(h), STAT3(h), LEF-1(h), C/EBPalpha(h),CHOP-10(h), NF-YA(h), FOXO3a(h)...
ENSG00000261794	GOLGA8H	golgin A8 family member H	9.84	NR1B1(h), egr-3(h), STAT3(h), LEF-1(h), C/EBPalpha(h),CHOP-10(h), NF-YA(h), FOXO3a(h)...
ENSG00000114126	TFDP2	transcription factor Dp-2	9.43	NF-YA(h), FOXO3a(h), egr-3(h), LEF-1(h), C/EBPalpha(h),CHOP-10(h), NR1B1(h), c-Rel(h)...
ENSG00000171735	CAMTA1	calmodulin binding transcription activator 1	9.3	NR1B1(h), egr-3(h), c-Rel(h), STAT3(h), LEF-1(h), NF-YA(h), FOXO3a(h)...
ENSG00000163939	PBRM1	polybromo 1	9.27	islet1(h), NF-YA(h), LEF-1(h), egr-3(h), NR1B1(h), STAT3(h), c-Rel(h)...

Enhancer model potentially involved in regulation of target genes (down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).

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.

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$MSX1_11
0.94; N=3

V$E2F2_11
0.88; N=3

V$NFAT4_Q5
1.00; N=1

V$HDAC2_02
0.83; N=2

V$HOXD13_03
0.94; N=1

Module width: 91

Module 2:

V$TFCP2_08
0.81; N=2

V$RSRFC4_01
0.88; N=2

V$HSF1_01
0.94; N=3

V$JUND_02
0.89; N=2

V$CMYB_Q5
0.99; N=3

V$AHRARNT_01
0.92; N=3

Module width: 135

Model score (-p*log10(pval)): 15.55
Wilcoxon p-value (pval): 1.18e-32
Penalty (p): 0.487
Average yes-set score: 4.20
Average no-set score: 2.66
AUC: 0.75
Separation point: 3.50
False-positive: 29.00%
False-negative: 28.67%

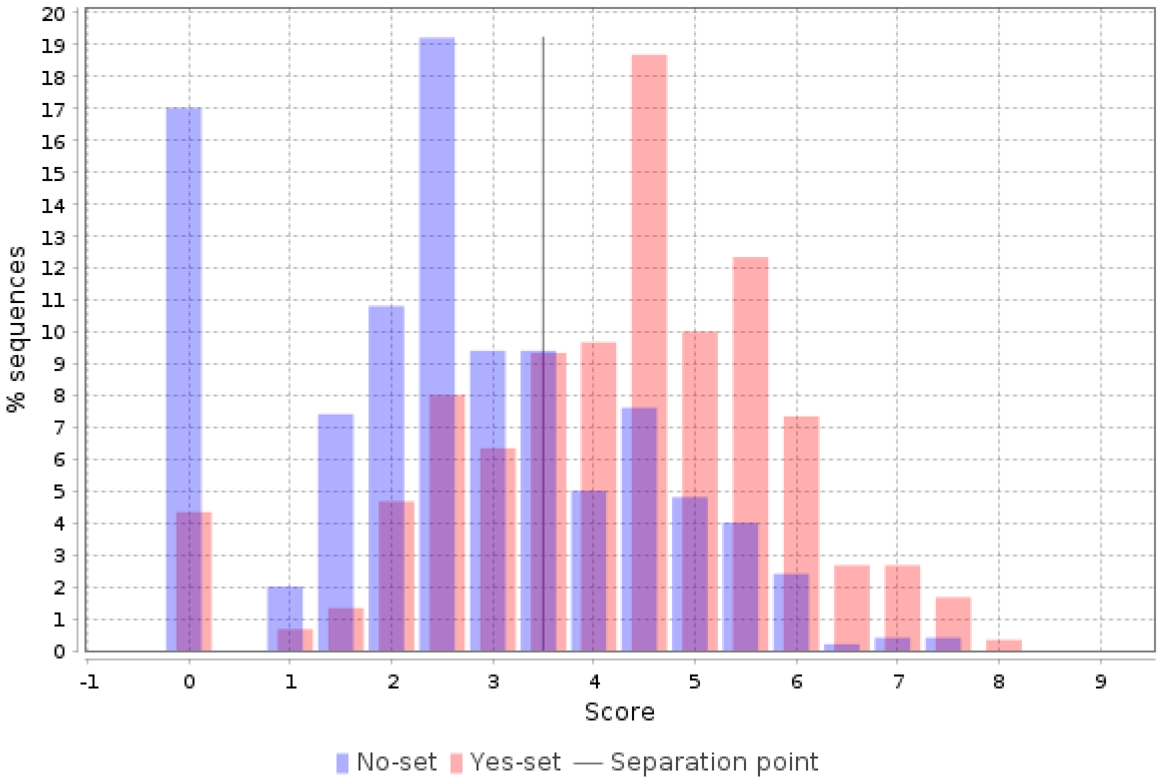


Table 6. List of top ten down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000152782	PANK1	pantothenate kinase 1	9.27	E2F-2(h), hdac2(h), Msx-1(h), Mef-2a(h), HSF1(h), AhR(h),arnt(h), CP2(h)...
ENSG00000144867	SRPRB	SRP receptor subunit beta	8.61	JunD(h), Msx-1(h), NFATc3(h), hdac2(h), HOXD13(h), CP2(h), HSF1(h)...
ENSG00000227077	AC107983.1	ribosomal protein S28 (RPS28) pseudogene	8.54	NFATc3(h), hdac2(h), HSF1(h), Mef-2a(h), CP2(h), AhR(h),arnt(h)
ENSG00000163697	APBB2	amyloid beta precursor protein binding family B member 2	8.15	CP2(h), E2F-2(h), Msx-1(h), hdac2(h), JunD(h), NFATc3(h), HSF1(h)...
ENSG00000104408	EIF3E	eukaryotic translation initiation factor 3 subunit E	8.14	hdac2(h), Msx-1(h), NFATc3(h), HOXD13(h), HSF1(h), JunD(h)
ENSG00000106105	GARS1	glycyl-tRNA synthetase 1	8.09	HSF1(h), hdac2(h), CP2(h), AhR(h),arnt(h), E2F-2(h)
ENSG00000159247	TUBBP5	tubulin beta pseudogene 5	8.06	JunD(h), HSF1(h), CP2(h), hdac2(h), NFATc3(h), E2F-2(h)
ENSG00000147416	ATP6V1B2	ATPase H+ transporting V1 subunit B2	8.05	HSF1(h), NFATc3(h), AhR(h),arnt(h), hdac2(h), E2F-2(h), CP2(h)
ENSG00000111110	PPM1H	protein phosphatase, Mg2+/Mn2+ dependent 1H	8.01	CP2(h), HSF1(h), HOXD13(h), hdac2(h), Msx-1(h), JunD(h)
ENSG00000163785	RYK	receptor like tyrosine kinase	7.93	E2F-2(h), hdac2(h), Mef-2a(h), NFATc3(h), HSF1(h), AhR(h),arnt(h), CP2(h)

On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 13 and 12 transcription factors controlling expression of up- and down-regulated genes respectively (see Tables 7-8).

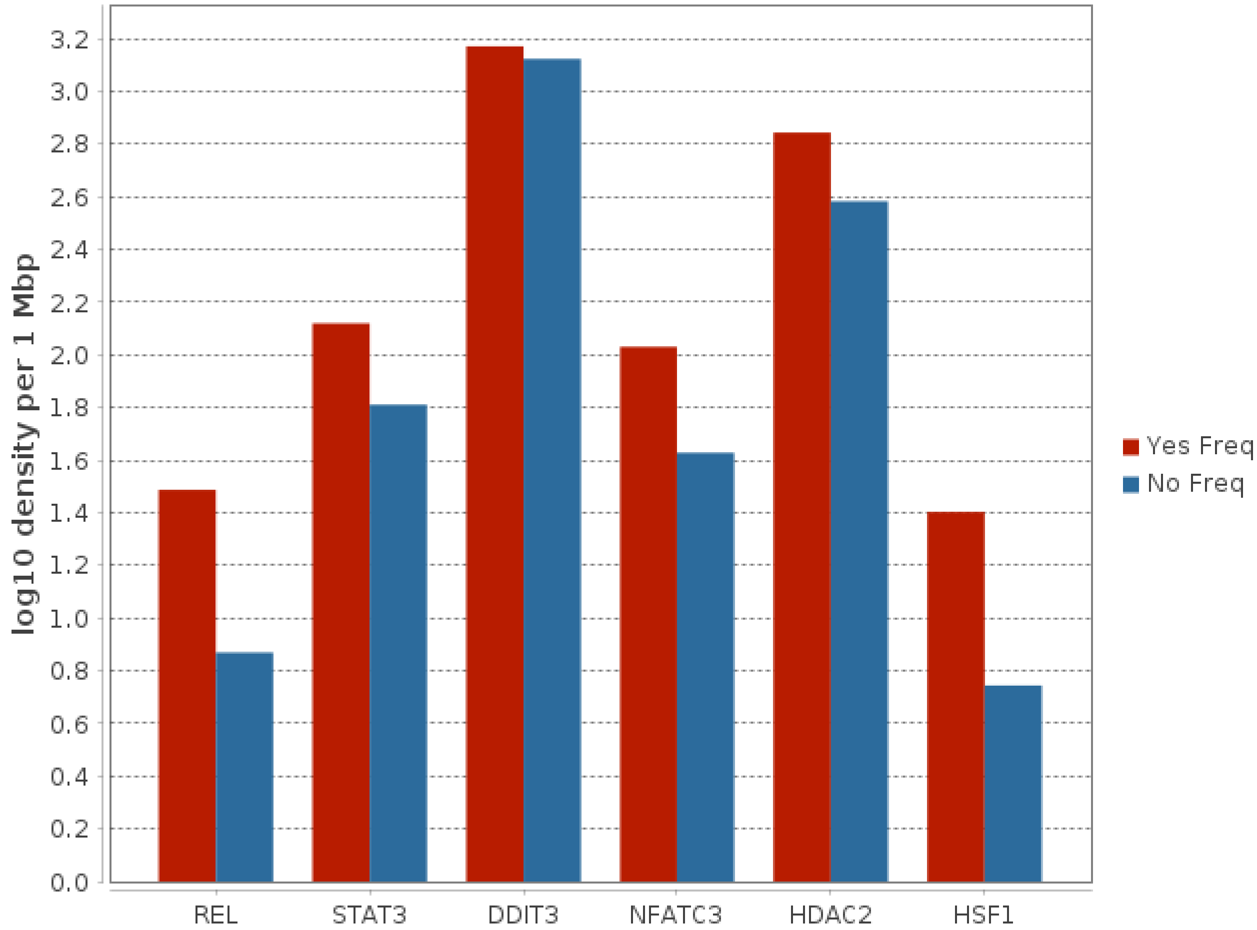
Table 7. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive). ***Yes-No ratio*** *is the ratio between frequencies of the sites in Yes sequences versus 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
MO000019368	REL	REL proto-oncogene, NF-kB subunit	3.08	4.14
MO000013123	STAT3	signal transducer and activator of transcription 3	3.03	2.04
MO000020832	DDIT3	DNA damage inducible transcript 3	2.7	1.12
MO000028767	SNAI2	snail family transcriptional repressor 2	2.55	1.6
MO000019418	CEBPA	CCAAT enhancer binding protein alpha	2.5	1.13
MO000025939	NFYA	nuclear transcription factor Y subunit alpha	2.43	2.28
MO000033904	RARA	retinoic acid receptor alpha	2.31	1.62
MO000020701	FOXO3	forkhead box O3	2.26	1.17
MO000026238	RUNX3	RUNX family transcription factor 3	2.04	3.31
MO000159782	LEF1	lymphoid enhancer binding factor 1	2.02	1.6

Table 8. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive). ***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
MO000020739	NFATC3	nuclear factor of activated T cells 3	1.71	2.52
MO000058923	HDAC2	histone deacetylase 2	1.55	1.82
MO000033378	HSF1	heat shock transcription factor 1	1.51	4.55
MO000009619	MYB	MYB proto-oncogene, transcription factor	1.46	2.42
MO000114191	ARNT	aryl hydrocarbon receptor nuclear translocator	1.43	1.49
MO000007834	JUND	JunD proto-oncogene, AP-1 transcription factor subunit	1.4	3.41
MO000004278	E2F2	E2F transcription factor 2	1.35	2.56
MO000084966	MEF2A	myocyte enhancer factor 2A	1.28	10.24
MO000117988	TFCP2	transcription factor CP2	1.13	1.49
MO000025932	AHR	aryl hydrocarbon receptor	0.91	1.49

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: REL, STAT3, DDIT3, NFATC3, HDAC2 and HSF1.



3.5. 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 9-10.

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

See full table →

ID	Master molecule name	Gene symbol	Gene description	logFC	Total rank
MO000019309	IKK-gamma(h)	IKBKG	inhibitor of nuclear factor kappa B kinase regulatory subunit gamma	0.9	150
MO000092591	Cdk1-isoform1(h):cyclinB1-isoform1(h)	CCNB1, CDK1	cyclin B1, cyclin dependent kinase 1	0.83	199
MO000200699	IKK-gamma-isoform3(h)	IKBKG	inhibitor of nuclear factor kappa B kinase regulatory subunit gamma	0.9	224
MO000200698	IKK-gamma-isoform2(h)	IKBKG	inhibitor of nuclear factor kappa B kinase regulatory subunit gamma	0.9	225
MO000150044	IKK(h)	CHUK, IKBKB, IKBKG	component of inhibitor of nuclear factor kappa B kinase complex, inhibitor of nuclear factor kappa B...	0.9	246
MO000022448	cyclinB1(h)	CCNB1	cyclin B1	0.83	274
MO000032712	MKP-4(h)	DUSP9	dual specificity phosphatase 9	0.75	274
MO000032484	Aurora-B(h)	AURKB	aurora kinase B	1.04	332
MO000023615	cyclinB1(h):Cdk1(h)	CCNB1, CDK1	cyclin B1, cyclin dependent kinase 1	0.83	333
MO000021736	Cdk2(h)	CDK2	cyclin dependent kinase 2	0.8	347

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

See full table →

ID	Master molecule name	Gene symbol	Gene description	logFC	Total rank
MO000129772	PTP-SL(h)	PTPRR	protein tyrosine phosphatase receptor type R	-4.6	49
MO000210517	FBXO25(h)	FBXO25	F-box protein 25	-0.47	147
MO000022315	PKCiota(h)	PRKCI	protein kinase C iota	-0.82	148
MO000033272	SGK-1(h)	SGK1	serum/glucocorticoid regulated kinase 1	-0.99	193
MO000022222	MKP-1(h)	DUSP1	dual specificity phosphatase 1	-1.38	199
MO000137752	PAK3(h)	PAK3	p21 (RAC1) activated kinase 3	-0.54	206
MO000137751	PAK3-isoform1(h)	PAK3	p21 (RAC1) activated kinase 3	-0.54	212
MO000256617	PAK3-isoform3(h)	PAK3	p21 (RAC1) activated kinase 3	-0.54	212
MO000256618	PAK3-isoform4(h)	PAK3	p21 (RAC1) activated kinase 3	-0.54	212
MO000137753	PAK3-isoform2(h)	PAK3	p21 (RAC1) activated kinase 3	-0.54	215

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.

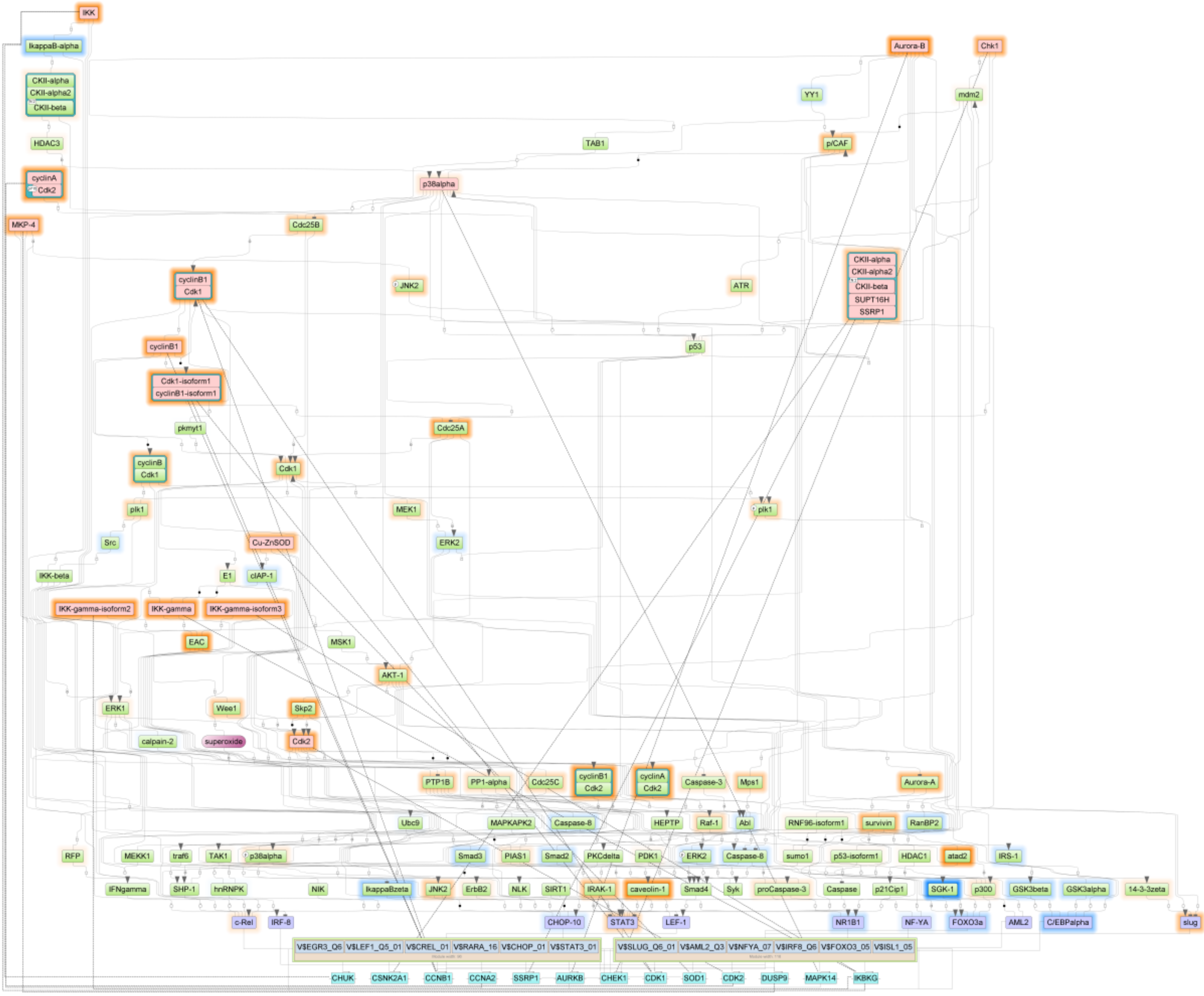


Figure 9. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules,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

See full diagram →

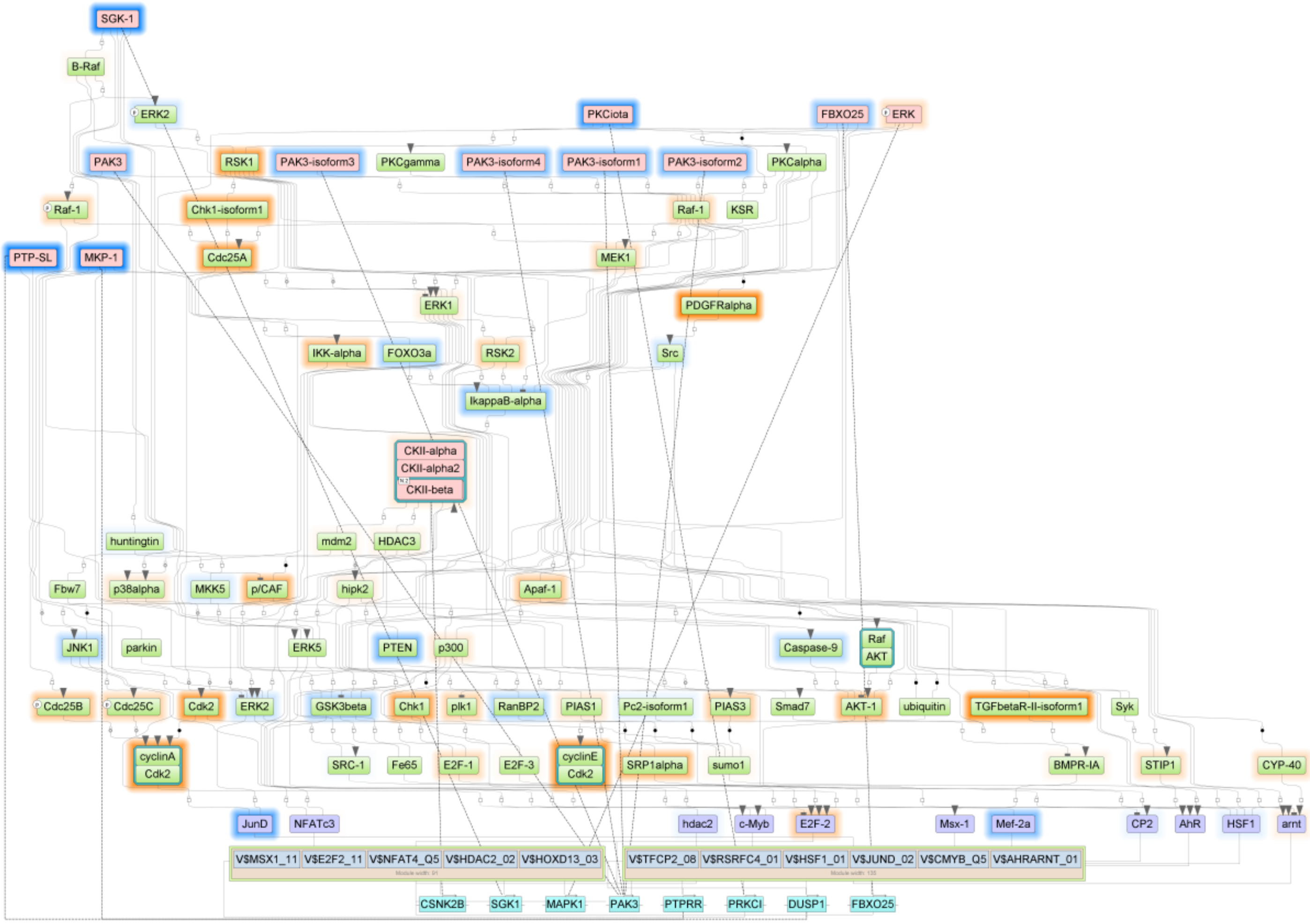


Figure 10. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during the search for master regulators from selected TFs. Orange and blue frames highlight molecules that are encoded by up- and downregulated genes, resp.

See full diagram →

4. Finding prospective drug targets

The identified master regulators that may govern pathology associated genes were checked for druggability potential using HumanPSD™ [5] database of gene-disease-drug assignments and PASS [11-13] software for prediction of biological activities of chemical compounds on the basis of a (Q)SAR approach. Respectively, for each master regulator protein we have computed two Druggability scores: HumanPSD Druggability score and PASS Druggability score. Where Druggability score represents the number of drugs that are potentially suitable for inhibition (or activation) of the corresponding target either according to the information extracted from medical literature (from HumanPSD™ database) or according to cheminformatics predictions of compounds activity against the examined target (from PASS software).

The cheminformatics druggability check is done using a pre-computed database of spectra of biological activities of chemical compounds from a library of all small molecular drugs from HumanPSD™ database, 2507 pharmaceutically active known chemical compounds in total. The spectra of biological activities has been computed using the program PASS [11-13] on the basis of a (Q)SAR approach.

If both Druggability scores were below defined thresholds (see Method 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	logFC	Total rank
PSMA7	proteasome 20S subunit alpha 7	3	0.53	352
AURKB	aurora kinase B	3	1.04	497
PDGFRA	platelet derived growth factor receptor alpha	8	2.83	591
GLRX	glutaredoxin	1	0.85	803
ME1	malic enzyme 1	2	0.98	829
PPP1CC	protein phosphatase 1 catalytic subunit gamma	4	0.35	863

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.*

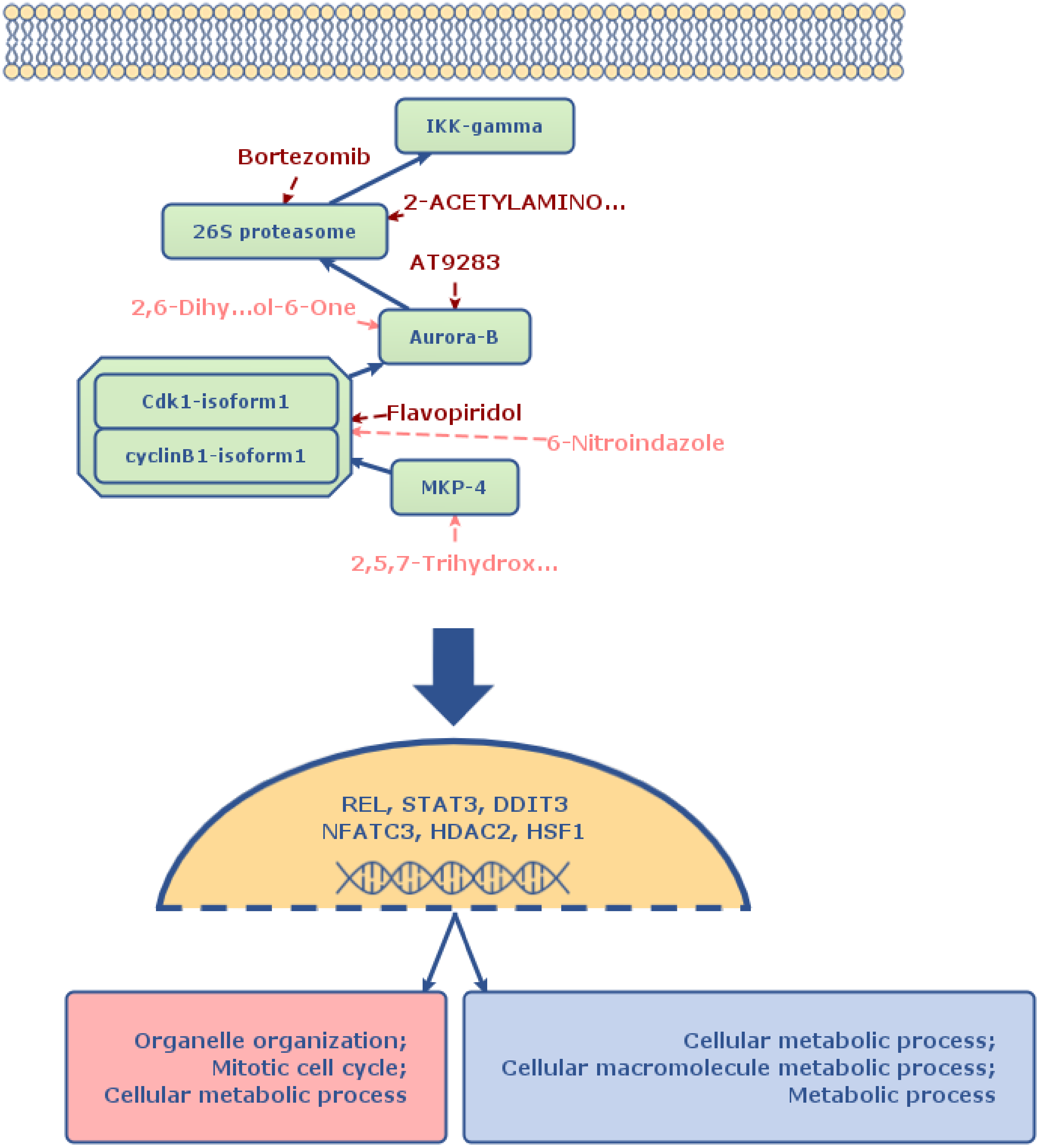
See full table →

Gene symbol	Gene Description	Druggability score	logFC	Total rank
DUSP9	dual specificity phosphatase 9	4.91	0.75	274
PSMC5	proteasome 26S subunit, ATPase 5	1.28	0.53	352
PSMD5	proteasome 26S subunit, non-ATPase 5	1.28	0.53	352
PSMA7	proteasome 20S subunit alpha 7	2.17	0.53	352
PSMC2	proteasome 26S subunit, ATPase 2	1.28	0.53	352
PSMC3	proteasome 26S subunit, ATPase 3	1.28	0.53	352

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:

- IKK-gamma
- Aurora-B
- 26S proteasome
- MKP-4
- Cdk1-isoform1:cyclinB1-isoform1

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: 2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One, 2,5,7-Trihydroxynaphthoquinone, AT9283, Bortezomib, Flavopiridol, 6-Nitroindazole and 2-ACETYLAMINO-4-METHYL-PENTANOIC ACID [1-(1-FORMYL-PENTYLCARBAMOYL)-3-METHYL-BUTYL]-AMIDE, 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.

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 rank	Disease activity score	Phase 4	Status (provided by Drugbank)
Imatinib	PDGFRB, PDGFRA	60	3	Breast Neoplasms, Gastrointestinal Stromal Tumors, Leukemia, Leukemia, Lymphoid, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Mastocytosis...	small molecule, approved
Regorafenib	PDGFRB, PDGFRA, RAF1	75	2	Colorectal Neoplasms, Gastrointestinal Stromal Tumors, Neoplasms, Rectal Neoplasms	small molecule, approved
Sunitinib	PDGFRB, PDGFRA	76	2	Carcinoma, Renal Cell, Gastrointestinal Neoplasms, Gastrointestinal Stromal Tumors, Intestinal Neoplasms, Lung Neoplasms, Neoplasms, Neuroendocrine Tumors...	small molecule, approved, investigational
Bosutinib	CAMK2G, MAP2K1, CDK2	77	1	Leukemia, Myeloid	small molecule, approved
Pazopanib	PDGFRB, PDGFRA	105	7	Carcinoma, Renal Cell, Neoplasms, Noma	small molecule, approved

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 rank	Phase 4	Status (provided by Drugbank)
AT9283	AURKA, AURKB	26	This drug was not tested on Phase 4 clinical trials yet. See full table for more details.	small molecule, investigational
Flavopiridol	CDK8, CDK9, CDK5, CDK1, CDK2, CDK7	27	This drug was not tested on Phase 4 clinical trials yet. See full table for more details.	small molecule, experimental, investigational
2-ACETYLAMINO-4-METHYL-PENTANOIC ACID [1-(1-FORMYL-PENTYLCARBAMOYL)-3-METHYL-BUTYL]-AMIDE	PSMA7	28	This drug was not tested on Phase 4 clinical trials yet. See full table for more details.	small molecule, experimental
Becaplermin	PDGFRB, PDGFRA	29	This drug was not tested on Phase 4 clinical trials yet. See full table for more details.	biotech, approved, investigational
HESPERIDIN	AURKB	30	This drug was not tested on Phase 4 clinical trials yet. See full table for more details.	small molecule, experimental

No prospective drugs were found, which would be predicted by PASS software to be active against the identified drug targets and would be predicted to have biological activity against the studied disease(s).

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 rank	Target activity score
2,5,7-Trihydroxynaphthoquinone	MAPK14, CDC25A, MAPK9, POR, CDKN3, MAPK6, CDC25B...	27	0.58
Camptothecin	HIF1A, CASP3	34	0.31
Topotecan	HIF1A, CASP3	34	0.31
LE-SN38	HIF1A, CASP3	37	0.29
6-Nitroindazole	RPS6KA3, CAMK2G, CDK9, PRKD3, GRK5, PDGFRB, PRKACA...	42	2.09

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Imatinib, AT9283 and 2,5,7-Trihydroxynaphthoquinone. These drugs were selected for acting on the following targets: PDGFRA, AURKB and DUSP9, 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 multi-omics data set that contains *transcriptomics and epigenomics* data. The study is done in the context of *Ovarian Neoplasms*. 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:

Imatinib, AT9283 and 2,5,7-Trihydroxynaphthoquinone

These drugs were selected for acting on the following targets: PDGFRA, AURKB and DUSP9, 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:

IKK-gamma, Aurora-B, 26S proteasome, MKP-4 and Cdk1-isoform1:cyclinB1-isoform1

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: 2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One, 2,5,7-Trihydroxynaphthoquinone, AT9283, Bortezomib, Flavopiridol, 6-Nitroindazole and 2-ACETYLAMINO-4-METHYL-PENTANOIC ACID [1-(1-FORMYL-PENTYLCARBAMOYL)-3-METHYL-BUTYL]-AMIDE. 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.

- IKK-gamma
- Aurora-B
- 26S proteasome
- MKP-4
- Cdk1-isoform1:cyclinB1-isoform1

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 2021.2 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac).
The master regulator search uses the TRANSPATH® database (BIOBASE), release 2021.2 (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 2021.2 (https://genexplain.com/humanpsd).
The Ensembl database release Human100.38 (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" ($D\text{-}score_{PSD}$):

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

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 set of all targets related to the compound and number of elements in it, *w* is weight multiplier.

"Druggability score" (D-score) is calculated as follows:

$$D\text{-}score(g) = IAP(g) \sum_{s \in S(g)} \sum_{m \in M(s,g)} pa(m),$$

where *S(g)* is the set of structures for which target list contains given target, *M(s,g)* is the set of activity-mechanisms (for the given structure) that corresponds to the given gene, *pa(m)* is the probability to be active of the activity-mechanism (m), *IAP(g)* is the invariant accuracy of prediction for the given gene.

8. References

1. Kel A, Voss N, Jauregui R, Kel-Margoulis O, Wingender E. Beyond microarrays: Finding key transcription factors controlling signal transduction pathways. *BMC Bioinformatics.* **2006**;7(S2), S13. doi:10.1186/1471-2105-7-s2-s13
2. Stegmaier P, Voss N, Meier T, Kel A, Wingender E, Borlak J. Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer. *PLoS ONE.* **2011**;6(3):e17738. doi:10.1371/journal.pone.0017738
3. Koschmann J, Bhar A, Stegmaier P, Kel A, Wingender E. “Upstream Analysis”: An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays.* **2015**;4(2):270-286. doi:10.3390/microarrays4020270.
4. Kel A, Stegmaier P, Valeev T, Koschmann J, Poroikov V, Kel-Margoulis OV, and Wingender E. Multi-omics “upstream analysis” of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer. *EuPA Open Proteom.* **2016**;13:1-13. doi:10.1016/j.euprot.2016.09.002
5. Michael H, Hogan J, Kel A et al. Building a knowledge base for systems pathology. *Brief Bioinformatics.* **2008**;9(6):518-531. doi:10.1093/bib/bbn038
6. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **2006**;34(90001):D108-D110. doi:10.1093/nar/gkj143
7. Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* **2003**;31(13):3576-3579. doi:10.1093/nar/gkg585
8. Waleev T, Shtokalo D, Konovalova T, Voss N, Cheremushkin E, Stegmaier P, Kel-Margoulis O, Wingender E, Kel A. Composite Module Analyst: identification of transcription factor binding

site combinations using genetic algorithm. *Nucleic Acids Res.* **2006**;34(Web Server issue):W541-5.

9. Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.* **2006**;34(90001):D546-D551. doi:10.1093/nar/gkj107
0. Boyarskikh U, Pintus S, Mandrik N, Stelmashenko D, Kiselev I, Evshin I, Sharipov R, Stegmaier P, Kolpakov F, Filipenko M, Kel A. Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3. *BMC Med Genomics.* **2018**;11(1):12. doi:10.1186/1471-2105-7-s2-s13
1. Filimonov D, Poroikov V. Probabilistic Approaches in Activity Prediction. Varnek A, Tropsha A. *Cheminformatics Approaches to Virtual Screening.* Cambridge (UK): RSC Publishing. **2008**;:182-216.
2. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal.* **2006**;50(2):66-75 (russ)
3. Filimonov D, Poroikov V, Borodina Y, Gloriozova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *ChemInform.* **1999**;39(4):666-670. doi:10.1002/chin.199940210

Thank you for using the Genome Enhancer!

In case of any questions please contact us at support@genexplain.com

Supplementary material

1. Supplementary table 1 - Up-regulated genes
2. Supplementary table 2 - Down-regulated genes
3. Supplementary table 3 - Detailed report. Composite modules and master regulators (up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).
4. Supplementary table 4 - Detailed report. Composite modules and master regulators (down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).
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.