

# PDGFRA and PSMD5 are promising druggable targets for treating Ovarian Neoplasms that control activity of SMAD3, STAT3 and MEF2C transcription factors on of differentially expressed genes

Demo User

geneXplain GmbH

info@genexplain.com

Data received on 10/12/2021 ; Run on 06/07/2023 ; Report generated on 06/07/2023

Genome Enhancer release 3.2 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2023.1)

---



## 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: SMAD3, STAT3, YBX1, MEF2C, HSF1 and MEF2D. The subsequent network analysis suggested

- TGFbetaR-II
- plk4
- PDGFRalpha
- 26S proteasome

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: Erlotinib, seliciclib and Bortezomib.

## 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
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

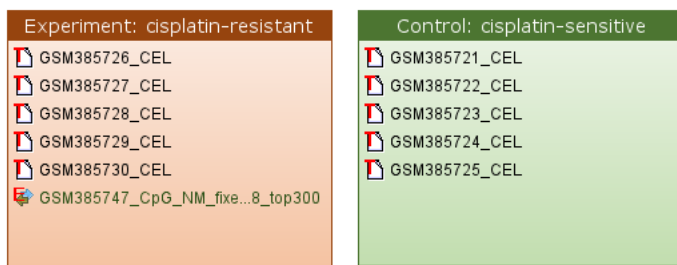


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

### 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 4060 upregulated genes (LogFC>0.1) out of which 3350 genes were found as significantly upregulated (p-value<0.1) and 4162 downregulated genes (LogFC<-0.1) out of which 3351 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	5.37	6.79E-14	7.28E-11
ENSG00000064218	DMRT3	doublesex and mab-3 related transcription factor 3	4.03	7.48E-12	2.59E-9
ENSG00000099139	PCSK5	proprotein convertase subtilisin/kexin type 5	3.93	1.35E-14	2.07E-11
ENSG00000197705	KLHL14	kelch like family member 14	3.89	9.84E-13	4.22E-10
ENSG00000129038	LOXL1	lysyl oxidase like 1	3.54	2.27E-10	3.24E-8
ENSG00000133083	DCLK1	doublecortin like kinase 1	3.24	8.07E-13	3.76E-10
ENSG00000141431	ASXL3	ASXL transcriptional regulator 3	3.14	1.36E-11	3.64E-9
ENSG00000126950	TMEM35A	transmembrane protein 35A	3.05	1.6E-12	6.15E-10
ENSG00000164692	COL1A2	collagen type I alpha 2 chain	2.86	2.21E-10	3.24E-8
ENSG00000138378	STAT4	signal transducer and activator of transcription 4	2.86	3.04E-10	3.75E-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	1.5E-15	4.04E-12
ENSG00000139292	LGR5	leucine rich repeat containing G protein-coupled receptor 5	-6.24	5.76E-18	6.18E-14
ENSG00000149968	MMP3	matrix metalloproteinase 3	-5.16	2E-13	1.65E-10
ENSG00000163359	COL6A3	collagen type VI alpha 3 chain	-5.08	5.33E-16	1.9E-12
ENSG00000169908	TM4SF1	transmembrane 4 L six family member 1	-4.94	1.59E-16	8.54E-13
ENSG00000153233	PTPRR	protein tyrosine phosphatase receptor type R	-4.6	5.98E-13	3.21E-10
ENSG00000166670	MMP10	matrix metalloproteinase 10	-4.45	9.28E-15	1.66E-11
ENSG00000106511	MEOX2	mesenchyme homeobox 2	-4.26	3.66E-12	1.35E-9
ENSG00000145431	PDGFC	platelet derived growth factor C	-4.14	3.26E-14	4.37E-11
ENSG00000060718	COL11A1	collagen type XI alpha 1 chain	-3.65	9.63E-11	1.75E-8



### **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.

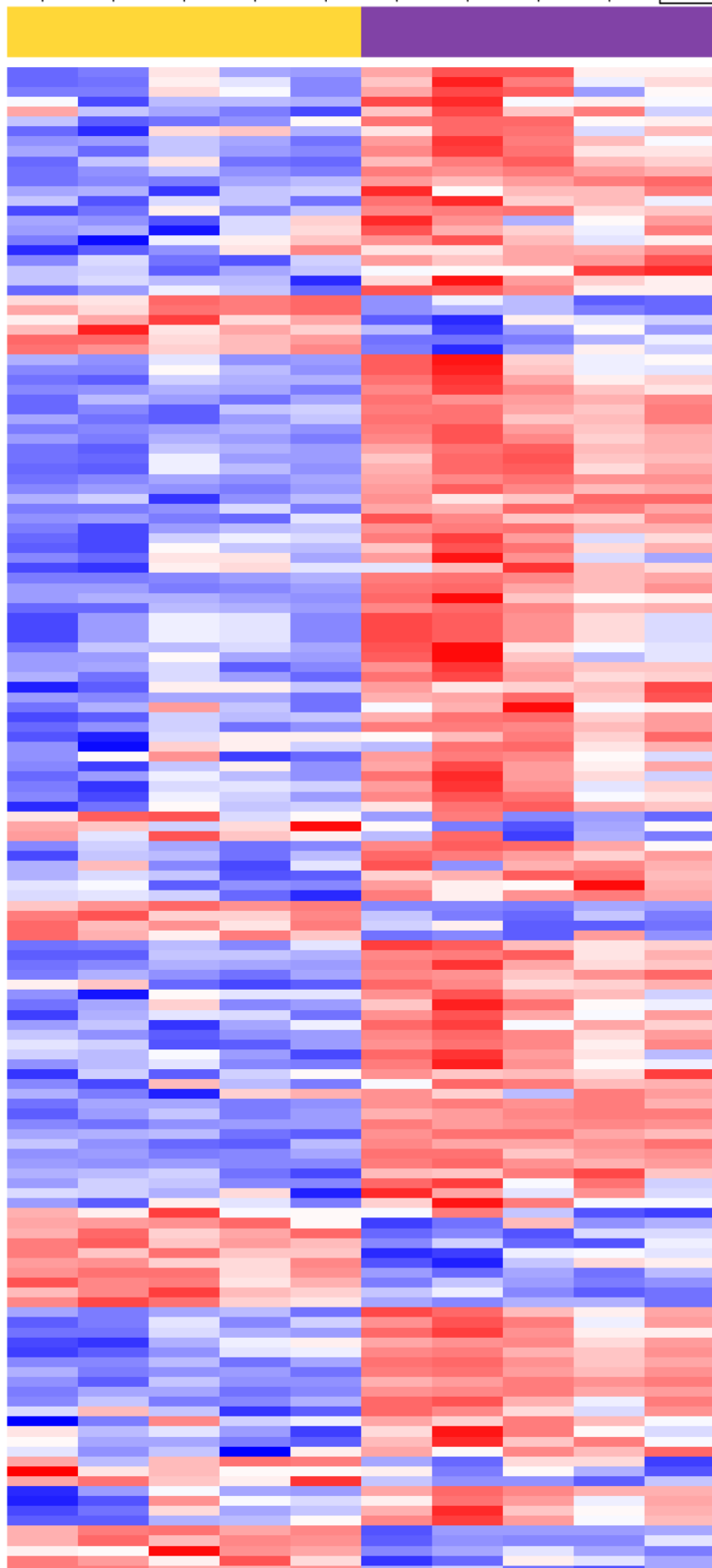
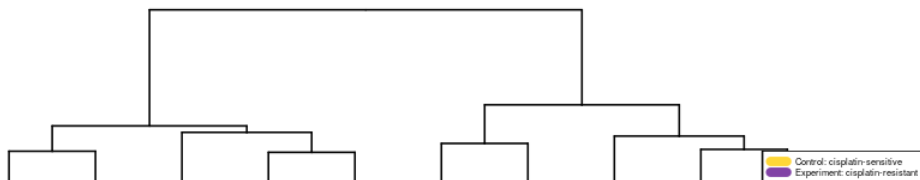
#### **Heatmap of differentially expressed genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive**

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



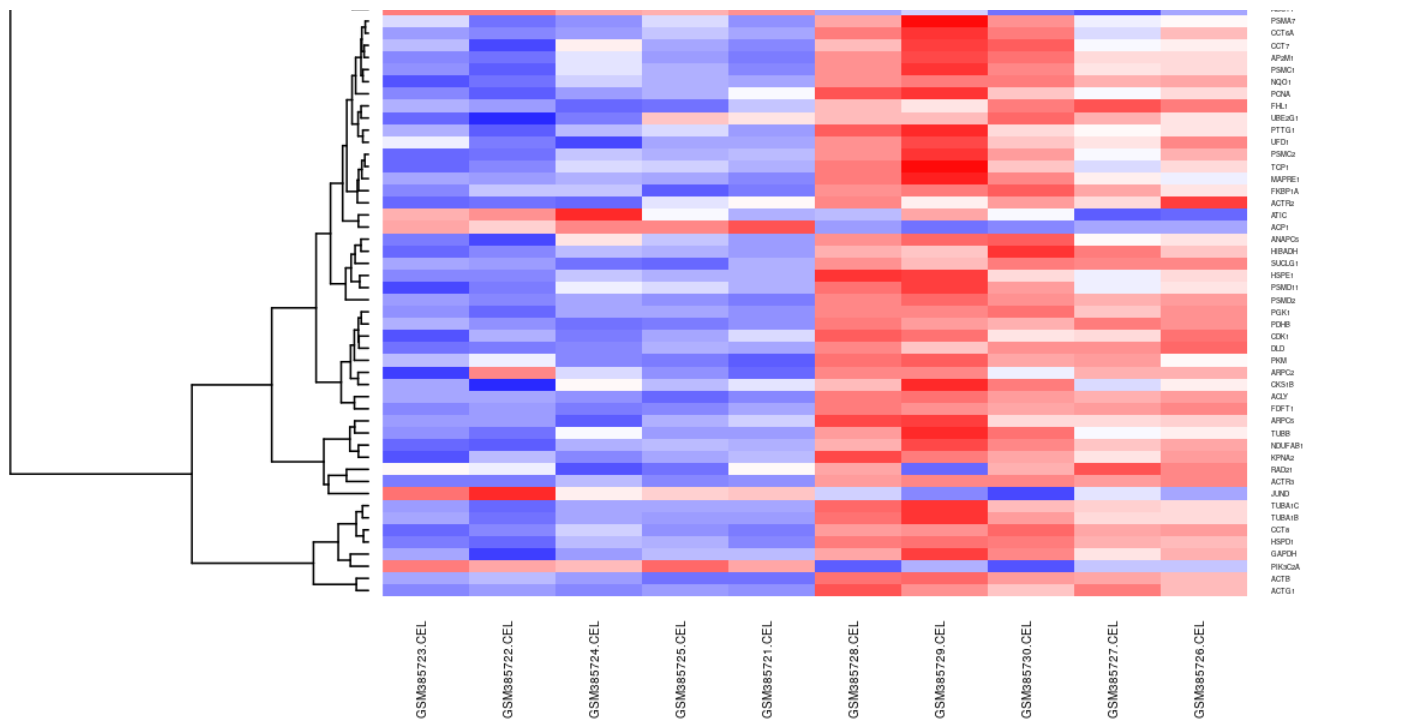


Gene Expression Normalized by rows



APPC8  
PSMG5  
GJA3  
ANAPC11  
ILK  
APPC1A  
BRK1  
OS  
TFPI2  
VOP  
LIMA1  
SUCLG2  
PSMG7  
AP51  
APPC3  
MAD4L1  
SDH  
URE5T  
RCHY1  
CDC29  
GLASP2  
DOH1  
NUDC  
ERCC1  
MPC  
QART  
MAX  
DYRK1A  
KLF6  
RUB3  
TUBB4B  
COB1  
FGD  
CHP1A1  
PSMG6  
NEK2  
ID1  
PSMG1  
IDH8  
AOC2  
PCNA1  
IDH5  
PSMG5  
WEE1  
SOX4  
FRS3  
PSMG4  
PSMG12  
COB2  
STP1  
TFPI2  
DLAT  
SKP2  
BRCA2  
FDP5  
TUBB3A  
TUBB4A  
TUBB1  
TUBB6  
TUBG1  
SQLE  
PSMG13  
NUP2  
ANAPC4  
DYRK1H  
SOX8  
FANG2  
CDC27  
CDK7  
IGFAP1  
HCS3  
ANAPC7  
CDKN3  
PSMG3  
ACD12  
INPP5A  
MINPP1  
SPR1  
ACD1  
TTK  
URE52  
URE5H  
DEDD  
GSS  
GDF8  
MCD13  
SCAND1  
MUS81  
CDC4A  
MAY1  
DNK1  
CDG2  
GGCK  
MAD4L2  
GSR  
GENPE  
KIFC1  
KIF1A  
CDG1  
CDC4B  
PSMG3  
OFLAR  
SQDPMH  
GCM1  
LSS  
CAV1  
CDKNAC  
FLNA  
FANG2b  
INGG1  
AOF2  
GCD1  
EIF2  
SMARCB1  
KIF22  
SIF  
ZF-YVE9  
URP6  
ACD79  
GRIH3  
GCH1  
THRB  
TGFB3  
FOXP3  
FOXP8  
CTPS1  
KIF23  
ALPNS  
OHEK1  
HMOG31  
BUB1  
EIF7  
IDH2  
CDKN2  
TUBGCP4  
STX13  
PMM1  
GORASP1  
ANAPC2  
TBR1  
INPP5K  
SHB1  
ERCC4  
CDG6  
CDKNAC  
TUBA3C  
INCENP  
CS  
SGK1  
GSK3  
FOXO25  
ATP11C

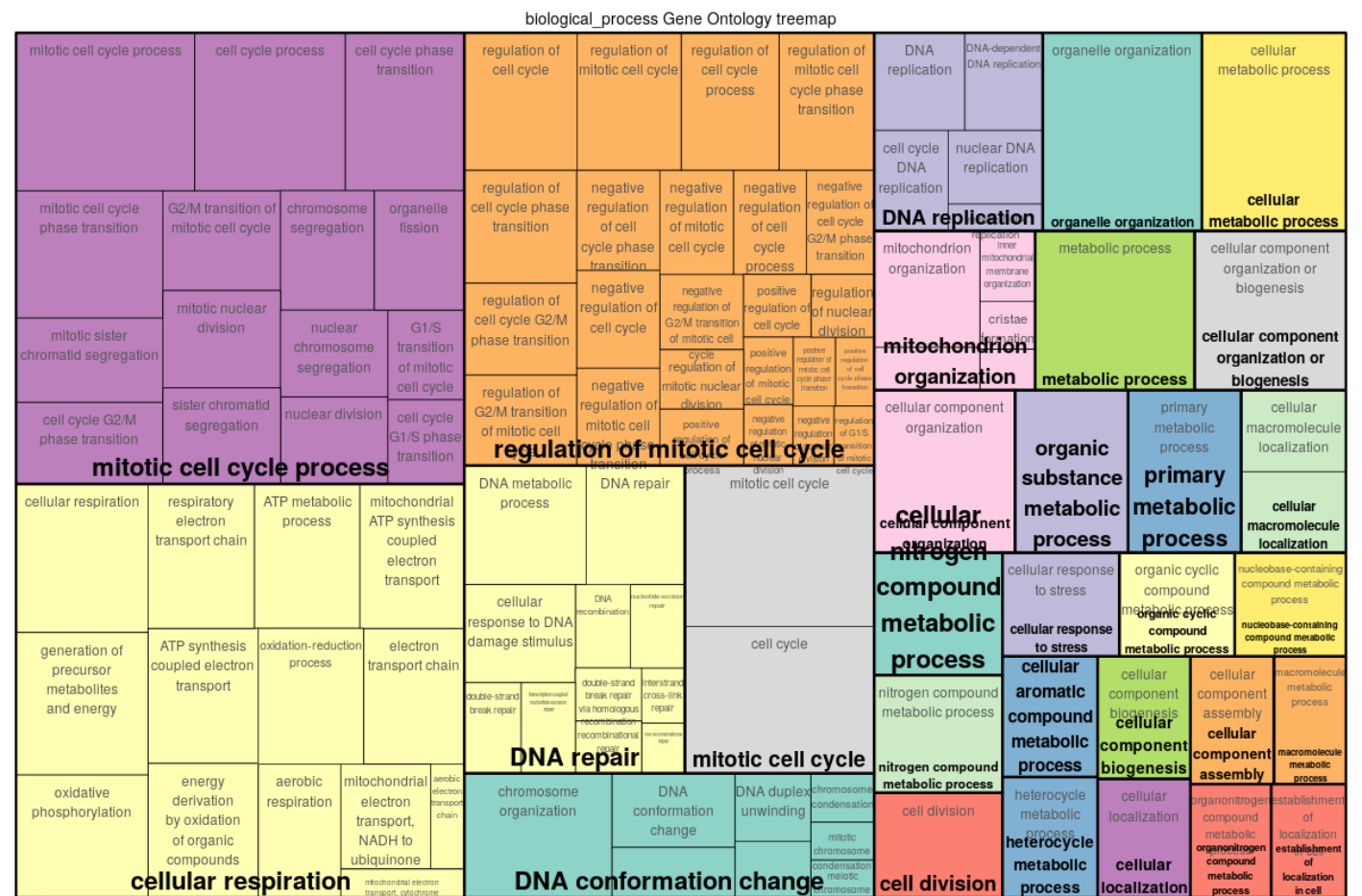




### Up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive:

3350 significant up-regulated genes were taken for the mapping.

### GO (biological process)





TRANSPATH® Pathways (2023.1)

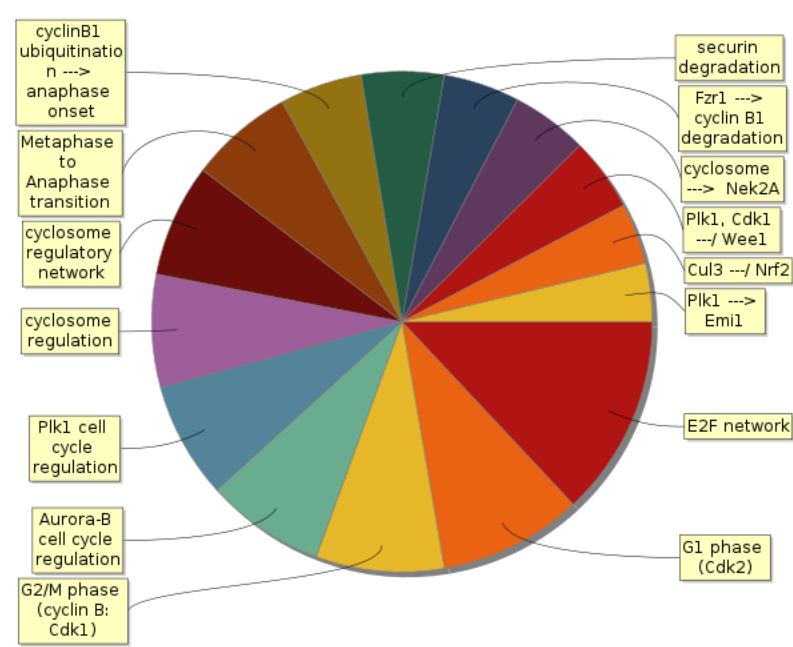


Figure 4. Enriched TRANSPATH® Pathways (2023.1) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. [Full classification →](#)

HumanPSD(TM) disease (2023.1)

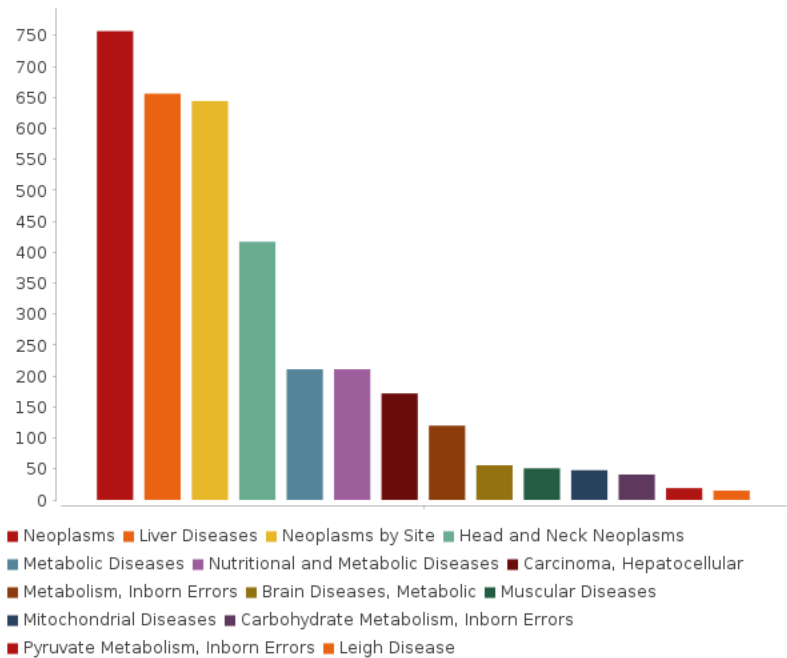


Figure 5. Enriched HumanPSD(TM) disease (2023.1) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. The size of the bars correspond to the number of biomarkers of the given disease found among the input set. [Full classification →](#)

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

3351 significant down-regulated genes were taken for the mapping.

GO (biological process)



intracellular transport		protein transport		amide transport		nuclear-transcribed mRNA catabolic process, nonsense-mediated decay		nuclear-transcribed mRNA catabolic process		translational initiation		cellular metabolic process		cellular macromolecule metabolic process		metabolic process			
intracellular protein transport		establishment of protein localization		macromolecule localization		mRNA catabolic process		mRNA metabolic process		cytoplasmic translational initiation		cellular metabolic process		cellular macromolecule metabolic process		metabolic process			
protein localization		peptide transport		nitrogen compound transport		nuclear-transcribed mRNA catabolic process, nonsense-mediated decay		heterocycle catabolic process		translational initiation		cellular metabolic process		cellular nitrogen compound metabolic process		gene expression			
intracellular protein transport						protein localization to endoplasmic reticulum		establishment of protein localization to endoplasmic reticulum		macromolecule metabolic process		nitrogen compound metabolic process		cellular nitrogen compound metabolic process		gene expression			
peptide biosynthetic process		translation		cytoplasmic translation		protein targeting to ER protein localization to endoplasmic reticulum						macromolecule metabolic process		nitrogen compound metabolic process		gene expression			
peptide metabolic process		amide biosynthetic process		cellular amide metabolic process		protein targeting to ER protein localization to endoplasmic reticulum						organic substance metabolic process		cellular nitrogen compound biosynthetic process		nucleic acid metabolic process			
peptide biosynthetic process						cellular macromolecule biosynthetic process		macromolecule biosynthetic process		organic substance metabolic process		cellular biosynthetic process		RNA metabolic process		nucleic acid metabolic process			
cotranslational protein targeting to membrane		establishment of protein localization to membrane		protein localization to membrane		protein targeting		cellular macromolecule biosynthetic process						cellular biosynthetic process		cellular protein metabolic process		heterocycle metabolic process	
SRP-dependent cotranslational protein targeting to membrane		protein targeting to membrane		establishment of protein localization to plasma membrane		protein localization to cell periphery		protein localization to plasma membrane		primary metabolic process		organic substance biosynthetic process		nucleobase-containing compound metabolic process		organic cyclic compound metabolic process			
cotranslational protein targeting to membrane						cellular macromolecule localization		cellular protein localization		primary metabolic process		organic substance biosynthetic process		regulation of gene expression		cellular localization			

### Full classification →

A pie chart illustrating the distribution of 15 biological pathways. The chart is divided into 15 colored slices, each connected by a thin grey line to a corresponding text label in a yellow box. The labels are arranged around the chart, with some on the left and some on the right. The pathways and their approximate proportions are as follows:

- TGFbeta1 --> Smad2/3** (Red, largest slice, ~35%)
- insulin pathway** (Orange, ~15%)
- ErbB3 --> survival** (Yellow, ~10%)
- IL-3 signaling** (Green, ~8%)
- insulin --> AKT-1 pathway** (Blue, ~7%)
- NRG --> Akt-1** (Purple, ~5%)
- TGFbetaR1 --> pak2, ERK1 --> SMAD7, SERPINE1** (Brown, ~4%)
- PRL --> PI3K --> AKT** (Dark Blue, ~3%)
- MIC2-isoform2 --> FosB --> MMP9** (Dark Purple, ~3%)
- MEK --> EZR** (Dark Green, ~3%)
- MIC2-isoform2 --> JNK, JunD --> MMP9** (Dark Red, ~3%)
- TGFbetaR1 --> ERK** (Dark Brown, ~3%)
- tuberlin --> beta-catenin** (Grey, ~3%)
- TGFbeta pathway** (Light Grey, ~3%)

**Full classification →**

**HumanPSD(TM) disease (2023.1)**



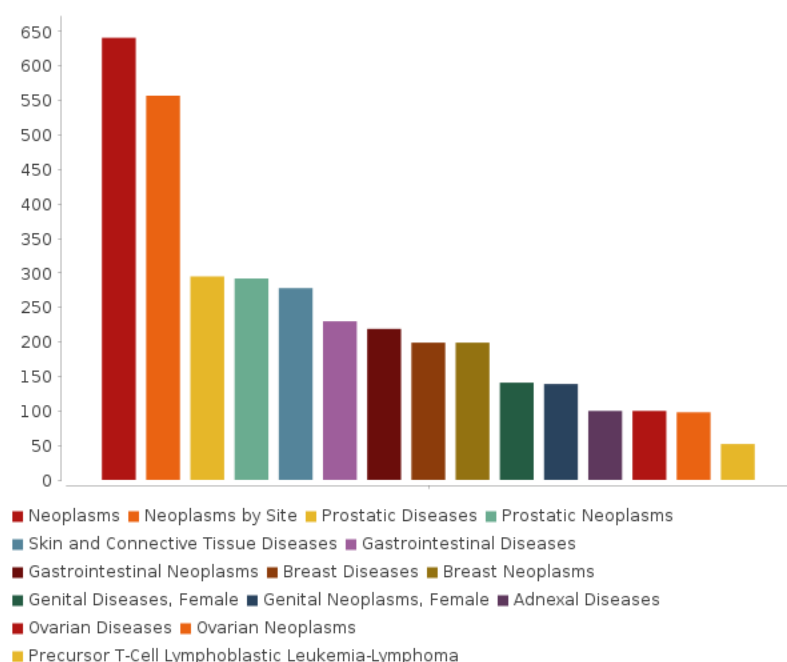
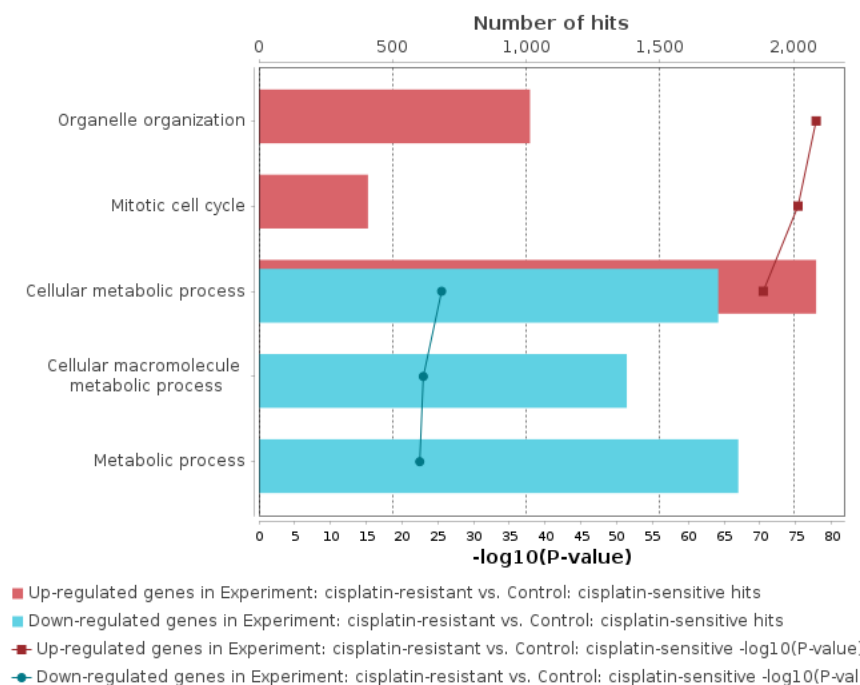


Figure 8. Enriched HumanPSD(TM) disease (2023.1) of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

[Full classification →](#)

The result of overall Gene Ontology (GO) analysis of the differentially expressed genes of the studied pathology can be summarized by the following diagram, revealing the most significant functional categories overrepresented among the observed (differentially expressed genes):



### 3.3. Analysis of enriched transcription factor binding sites and composite modules

In the next step a search for transcription factors binding sites (TFBS) was performed in the regulatory regions of the **target genes** by using the TF binding motif library of the [TRANSFAC®](#) database. We searched for so called **composite modules** that act as potential condition-specific **enhancers** of the **target genes** in their upstream regulatory regions (~1000 bp upstream of transcription start site (TSS)) and identify transcription factors regulating activity of the genes through such **enhancers**.

Classically, **enhancers** are defined as regions in the genome that increase transcription of one or several genes when inserted in either orientation at various distances upstream or downstream of the gene [8]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [9].

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 epigenomics 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 areas 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



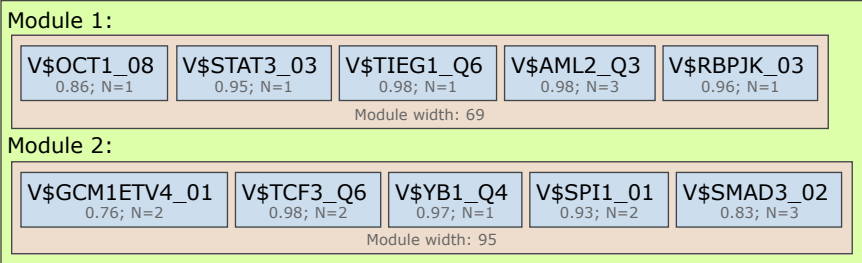
enhancers by specifying combinations of TF motifs (from [TRANSFAC®](#)) whose sites are most frequently clustered together in the regulatory regions of the studied genes. CMA identifies the transcription factors that through their cooperation provide a synergistic effect and thus have a great influence 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 in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.

The model consists of 2 module(s). Below, for each module the following information is shown:

- PWMs producing matches,
- number of individual matches for each PWM,
- score of the best match.



**Model score (-p\*log10(pval)):** 14.26  
**Wilcoxon p-value (pval):** 3.50e-29  
**Penalty (p):** 0.501  
**Average yes-set score:** 3.34  
**Average no-set score:** 2.45  
**AUC:** 0.74  
**Separation point:** 2.62  
**False-positive:** 41.00%  
**False-negative:** 20.33%

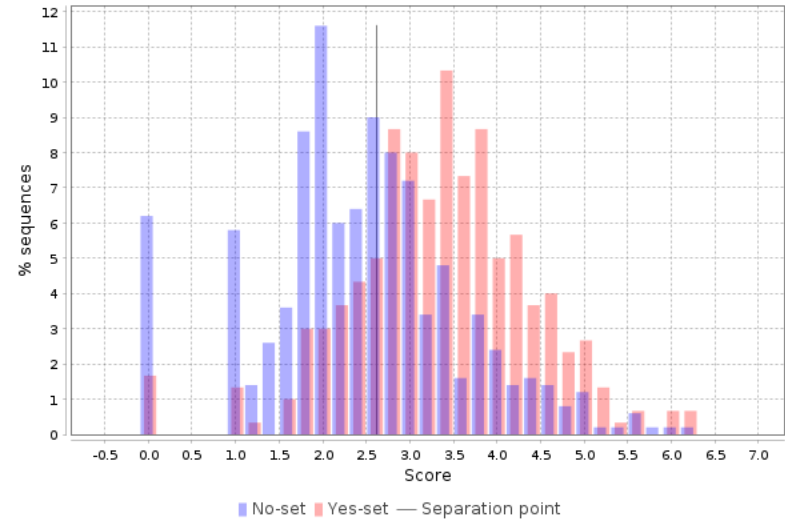


Table 4. List of top ten up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.

[See full table](#) →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000158796	DEDD	death effector domain containing	7.58	YB-1(h), TCF-7L1(h), SMAD3(h), ETV4(h),GCMa(h), KLF10(h), Runx3(h), STAT3(h)
ENSG00000173540	GMPPB	GDP-mannose pyrophosphorylase B	7.11	YB-1(h), ETV4(h),GCMa(h), TCF-7L1(h), PU.1(h), KLF10(h), STAT3(h), RBP-Jkappa(h)...
ENSG00000100124	ANKRD54	ankyrin repeat domain 54	6.39	STAT3(h), ETV4(h),GCMa(h), KLF10(h), YB-1(h), PU.1(h), SMAD3(h), TCF-7L1(h)
ENSG00000131089	ARHGEF9	Cdc42 guanine nucleotide exchange factor 9	6.38	RBP-Jkappa(h), POU2F1(h), TCF-7L1(h), YB-1(h), SMAD3(h)
ENSG00000213965	NUDT19	nudix hydrolase 19	6.32	YB-1(h), PU.1(h), SMAD3(h), TCF-7L1(h), Runx3(h), ETV4(h),GCMa(h), RBP-Jkappa(h)
ENSG00000175756	AURKAIP1	aurora kinase A interacting protein 1	6.26	RBP-Jkappa(h), KLF10(h), ETV4(h),GCMa(h), PU.1(h), SMAD3(h), TCF-7L1(h), YB-1(h)
ENSG00000254986	DPP3	dipeptidyl peptidase 3	6.21	TCF-7L1(h), SMAD3(h), YB-1(h), STAT3(h), KLF10(h), RBP-Jkappa(h)
ENSG00000277791	PSMB3	proteasome 20S subunit beta 3	6.21	RBP-Jkappa(h), POU2F1(h), ETV4(h),GCMa(h), PU.1(h), TCF-7L1(h), YB-1(h), SMAD3(h)
ENSG00000253293	HOXA10	homeobox A10	6.19	SMAD3(h), ETV4(h),GCMa(h), YB-1(h), TCF-7L1(h), PU.1(h), RBP-Jkappa(h), KLF10(h)
ENSG00000186185	KIF18B	kinesin family member 18B	6.1	Runx3(h), YB-1(h), TCF-7L1(h), PU.1(h), RBP-Jkappa(h), KLF10(h), ETV4(h),GCMa(h)...

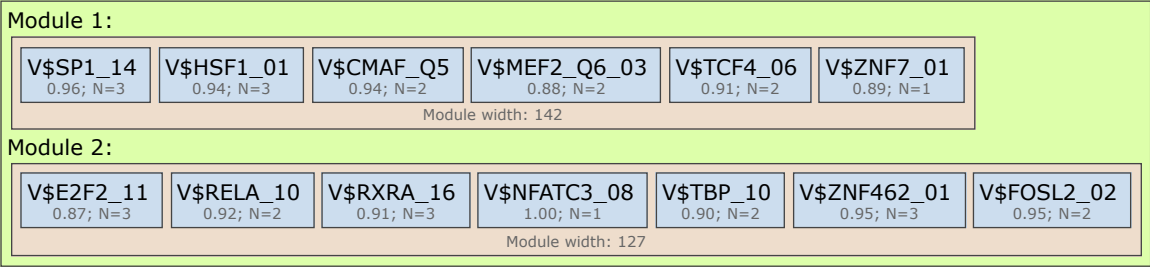
**Enhancer model potentially involved in regulation of target genes (down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).**



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

The model consists of 2 module(s). Below, for each module the following information is shown:

- PWMs producing matches,
- number of individual matches for each PWM,
- score of the best match.



**Model score (-p\*log10(pval)):** 16.64  
**Wilcoxon p-value (pval):** 1.21e-36  
**Penalty (p):** 0.463  
**Average yes-set score:** 4.06  
**Average no-set score:** 2.44  
**AUC:** 0.77  
**Separation point:** 3.38  
**False-positive:** 26.80%  
**False-negative:** 28.33%

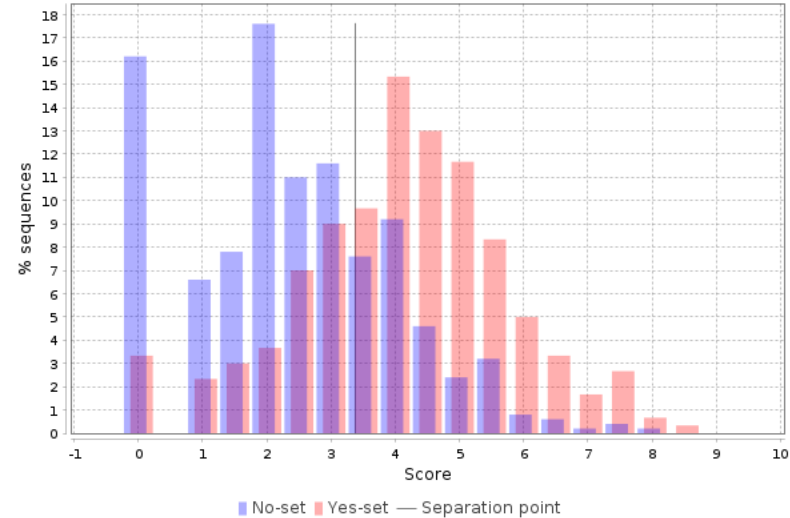


Table 5. List of top ten down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with identified enhancers in their regulatory regions. **CMA score** - the score of the CMA model of the enhancer identified in the regulatory region.

[See full table →](#)

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000052126	PLEKHA5	pleckstrin homology domain containing A5	8.99	MEF-2A(h),MEF-2B(h),MEF-2C(h),MEF-2D(h), TCF-7L2(h), RXRalpha(h), NF-kappaB-p65(h), HSF1(h), Sp1(h), E2F-2(h)...
ENSG00000087269	NOP14	NOP14 nucleolar protein	8.71	RXRalpha(h), NF-kappaB-p65(h), E2F-2(h), ZNF462(h), c-Maf(h), Sp1(h)
ENSG00000185404	SP140L	SP140 nuclear body protein like	8.69	HSF1(h), c-Maf(h), NF-kappaB-p65(h), ZNF7(h), RXRalpha(h), ZNF462(h), Fra-2(h)
ENSG00000233927	RPS28	ribosomal protein S28	8.63	HSF1(h), TCF-7L2(h), RXRalpha(h), ZNF462(h), ZNF7(h), NF-kappaB-p65(h)
ENSG00000099875	MKNK2	MAPK interacting serine/threonine kinase 2	8.55	TBP(h), E2F-2(h), RXRalpha(h), NF-kappaB-p65(h), HSF1(h), TCF-7L2(h), ZNF462(h)
ENSG00000163697	APBB2	amyloid beta precursor protein binding family B member 2	8.27	ZNF462(h), NF-kappaB-p65(h), RXRalpha(h), E2F-2(h), TBP(h), MEF-2A(h),MEF-2B(h),MEF-2C(h),MEF-2D(h), Fra-2(h)...
ENSG00000152782	PANK1	pantothenate kinase 1	8.27	ZNF462(h), E2F-2(h), MEF-2A(h),MEF-2B(h),MEF-2C(h),MEF-2D(h), TBP(h), HSF1(h), c-Maf(h), NF-kappaB-p65(h)
ENSG00000137221	TJAP1	tight junction associated protein 1	8.23	TCF-7L2(h), c-Maf(h), ZNF7(h), HSF1(h), ZNF462(h), RXRalpha(h), NF-kappaB-p65(h)
ENSG00000112139	MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1	8.22	TCF-7L2(h), TBP(h), MEF-2A(h),MEF-2B(h),MEF-2C(h),MEF-2D(h), ZNF462(h), E2F-2(h), NF-kappaB-p65(h), HSF1(h)...
ENSG00000148634	HERC4	HECT and RLD domain containing E3 ubiquitin protein ligase 4	8.11	ZNF7(h), ZNF462(h), E2F-2(h), HSF1(h), Sp1(h), MEF-2A(h),MEF-2B(h),MEF-2C(h),MEF-2D(h)

On the basis of the enhancer models we identified transcription factors potentially regulating the **target genes** of our interest. We found 11 and 16 transcription factors controlling expression of up- and down-regulated genes respectively (see Tables 6-7).



Table 6. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (up-regulated genes in 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 →

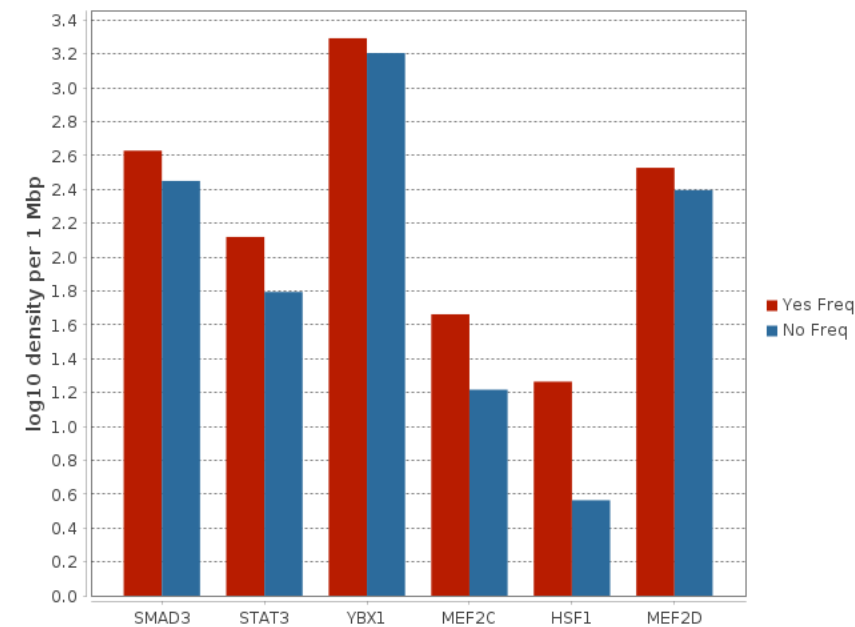
ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000057832	SMAD3	SMAD family member 3	3.25	1.51
MO000013123	STAT3	signal transducer and activator of transcription 3	2.51	2.11
MO000083480	YBX1	Y-box binding protein 1	2.47	1.22
MO000085616	SPI1	Spi-1 proto-oncogene	2.24	2.07
MO000046009	ETV4	ETS variant transcription factor 4	2.02	4.18
MO000026306	GCM1	glial cells missing transcription factor 1	2.01	1.37
MO000026238	RUNX3	RUNX family transcription factor 3	1.67	4.6
MO000026845	TCF7L1	transcription factor 7 like 1	1.66	1.98
MO000025003	POU2F1	POU class 2 homeobox 1	1.58	1.45
MO000030964	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	1.54	10.03

Table 7. 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
MO000031322	MEF2C	myocyte enhancer factor 2C	2.77	2.79
MO000033378	HSF1	heat shock transcription factor 1	2.57	5.02
MO000085555	MEF2D	myocyte enhancer factor 2D	2.46	1.35
MO000033308	SP1	Sp1 transcription factor	2.34	1.43
MO000089762	ZNF7	zinc finger protein 7	2.01	1.67
MO000079319	RELA	RELA proto-oncogene, NF-kB subunit	1.99	1.57
MO000084966	MEF2A	myocyte enhancer factor 2A	1.97	4.06
MO000019619	RXRA	retinoid X receptor alpha	1.86	1.71
MO000026074	FOSL2	FOS like 2, AP-1 transcription factor subunit	1.83	2.05
MO000026882	TCF7L2	transcription factor 7 like 2	1.8	2.51

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: SMAD3, STAT3, YBX1, MEF2C, HSF1 and MEF2D.



### 3.4. Finding master regulators in networks

In the second step of the upstream analysis common regulators of the revealed TFs were identified. These master regulators appear to be the key candidates for therapeutic targets as they have a master effect on regulation of intracellular pathways that activate the pathological process of our study. The identified master regulators are shown in Tables 8-9.



Table 8. Master regulators that may govern the regulation of **up-regulated** genes in 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
MO000017448	TGFbetaR-II(h)	TGFBR2	transforming growth factor beta receptor 2	2.79	80
MO000010977	PDGFRalpha(h)	PDGFRA	platelet derived growth factor receptor alpha	2.83	100
MO000083753	TGFbetaR-II-isoform1(h)	TGFBR2	transforming growth factor beta receptor 2	2.79	123
MO000041170	EAC(h)	CYLD	CYLD lysine 63 deubiquitinase	0.95	161
MO000202927	TGFbetaR-II-isoform2(h)	TGFBR2	transforming growth factor beta receptor 2	2.79	198
MO000962884	TGFbetaR-II-isoform3(h)	TGFBR2	transforming growth factor beta receptor 2	2.79	198
MO000092591	Cdk1-isoform1(h):cyclinB1-isoform1(h)	CCNB1, CDK1	cyclin B1, cyclin dependent kinase 1	0.83	209
MO000022448	cyclinB1(h)	CCNB1	cyclin B1	0.83	214
MO000023445	Cdc25A(h)	CDC25A	cell division cycle 25A	0.78	228
MO000023615	cyclinB1(h):Cdk1(h)	CCNB1, CDK1	cyclin B1, cyclin dependent kinase 1	0.83	246

Table 9. 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
MO000022222	MKP-1(h)	DUSP1	dual specificity phosphatase 1	-1.38
MO000118076	EGF:EGFR{pY}:ErbB2{pY}:Src	EGF, EGFR, ERBB2, SRC	SRC proto-oncogene, non-receptor tyrosine kinase, epidermal growth factor, epidermal growth factor r...	-0.85
MO000337356	Cdc23(h):APC7(h):Cdc16(h):APC5(h):APC4(h):Cdc27(h):ANAPC2(h):APC1(h):ANAPC16(h)	ANAPC1, ANAPC16, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC27	anaphase promoting complex subunit 1, anaphase promoting complex subunit 16, anaphase promoting comp...	-0.89
MO000129772	PTP-SL(h)	PTPRR	protein tyrosine phosphatase receptor type R	-4.6
MO000273494	APC1(h):ANAPC2(h):Cdc27(h):APC4(h):APC5(h):Cdc16(h):APC7(h):Cdc23(h):APC10(h):ANAPC16(h)	ANAPC1, ANAPC10, ANAPC16, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC27	anaphase promoting complex subunit 1, anaphase promoting complex subunit 10, anaphase promoting comp...	-0.89
MO000016677	EGFR(h)	EGFR	epidermal growth factor receptor	-0.85
MO000103285	MKP-7-isoform1(h)	DUSP16	dual specificity phosphatase 16	-0.36
MO000036550	MKP-7(h)	DUSP16	dual specificity phosphatase 16	-0.36
MO000092896	p38beta(h)	MAPK11	mitogen-activated protein kinase 11	-0.41
MO000170234	RNF4(h)	RNF4	ring finger protein 4	-0.89

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.



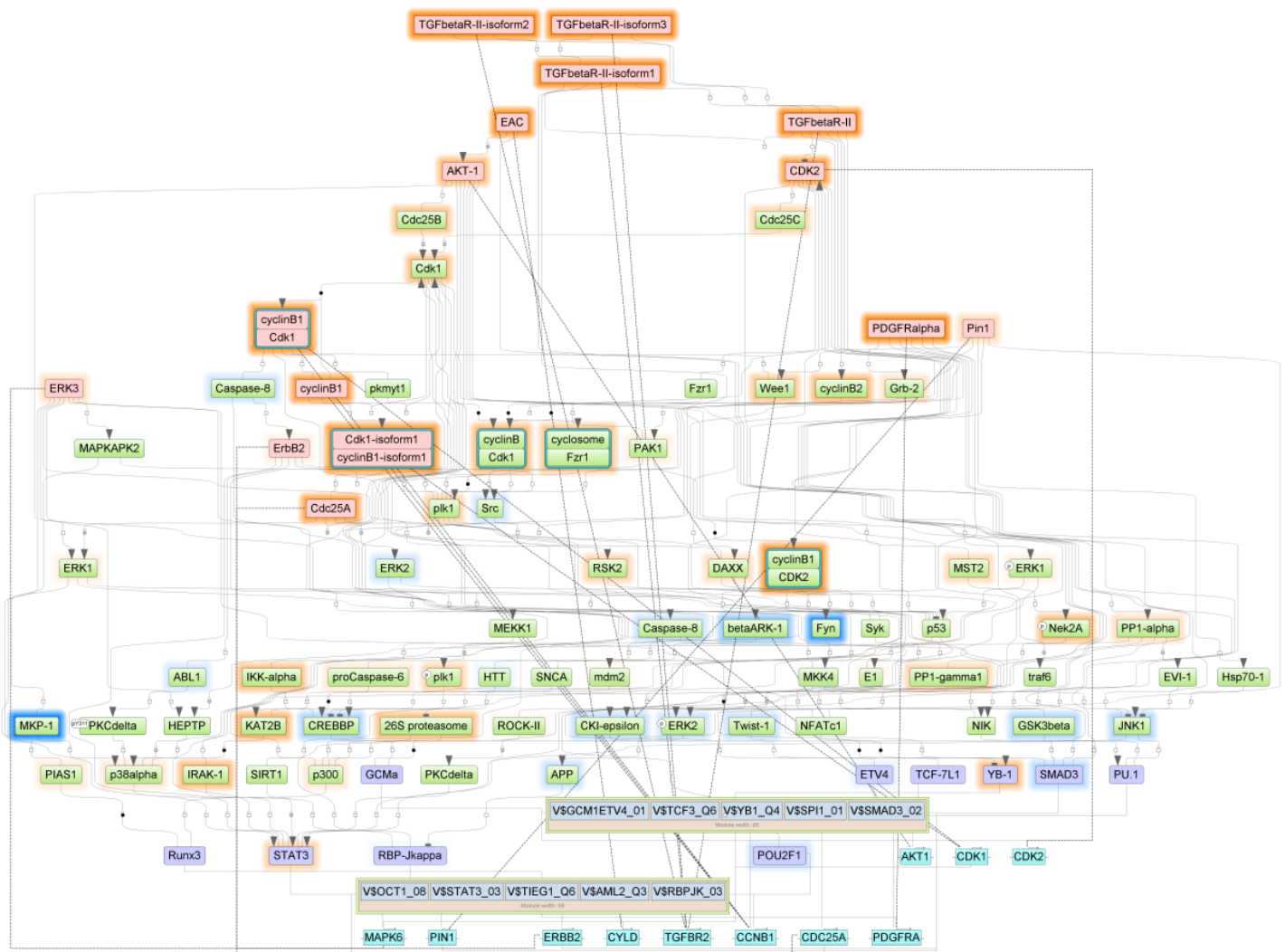


Figure 9. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, 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 →](#)



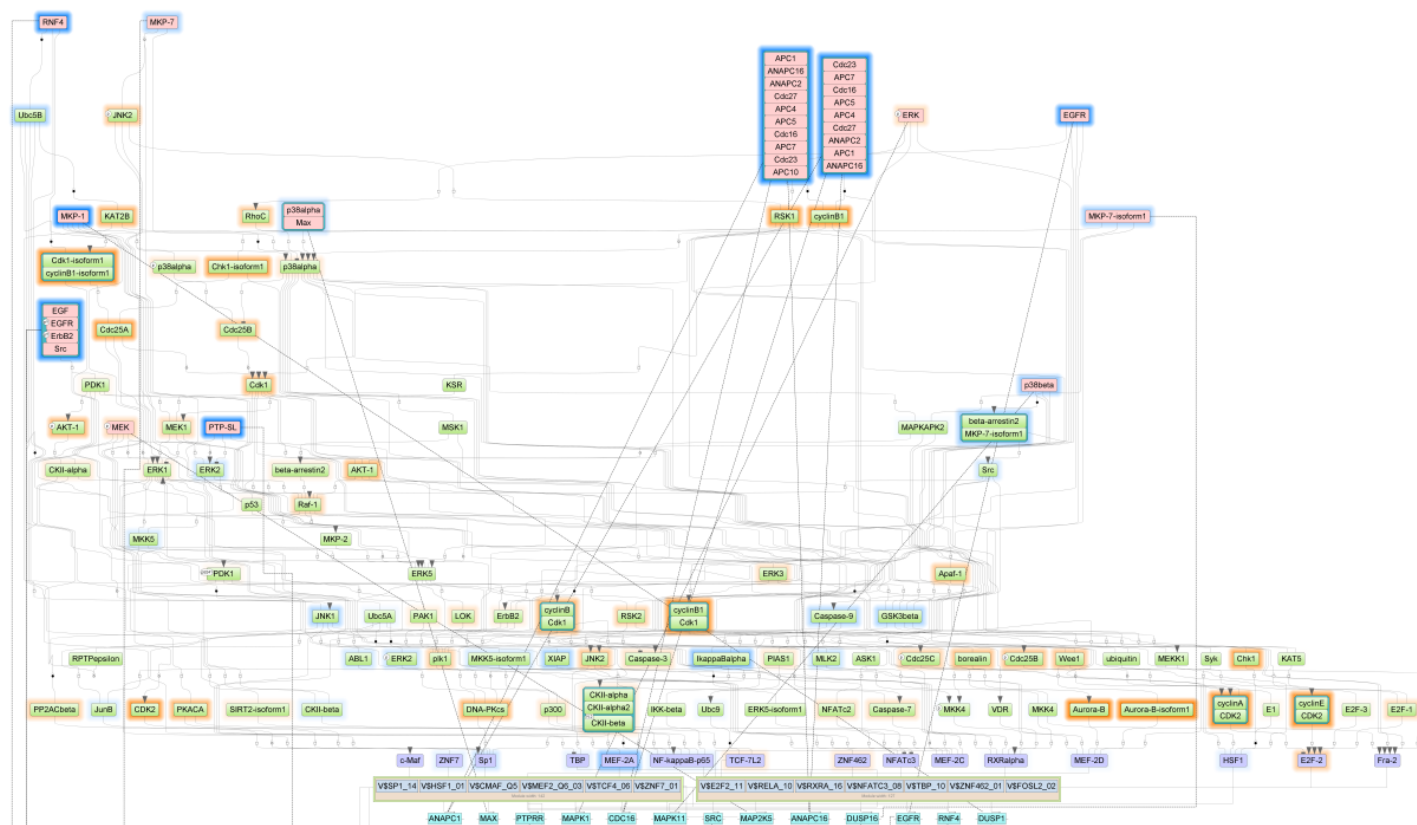


Figure 10. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate 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 Methods section for the details) such master regulator proteins were not used in further analysis of drug prediction.

As a result we created the following two tables of prospective drug targets (top targets are shown here):

 Table 10. Prospective drug targets selected from full list of identified master regulators filtered by Druggability score from HumanPSD™ database. **Druggability score** contains the number of drugs that are potentially suitable for inhibition (or activation) of the target. The drug targets are sorted according to the **Total rank** which is the sum of three ranks computed on the basis of the three scores: keynode score, CMA score and expression change score (logFC, if present). See Methods section for details.

[See full table →](#)

Gene symbol	Gene Description	Druggability score	logFC	Total rank
PDGFRA	platelet derived growth factor receptor alpha	53	2.83	310
PLK4	polo like kinase 4	28	0.59	315
PSMA7	proteasome 20S subunit alpha 7	2	0.52	384
TGFBR2	transforming growth factor beta receptor 2	28	2.79	431
AURKB	aurora kinase B	45	1.03	463
PPP1CC	protein phosphatase 1 catalytic subunit gamma	1	0.33	601

 Table 11. Prospective drug targets selected from full list of identified master regulators filtered by Druggability score predicted by PASS software. Here, the **Druggability score** for master regulator proteins is computed as a sum of PASS calculated probabilities to be active as a target for various small molecular compounds. The drug targets are sorted according to the **Total rank** which is the sum of three ranks computed on the basis of the three scores: keynode score, CMA score and expression change score (logFC, if present). See Methods section for details.

[See full table →](#)

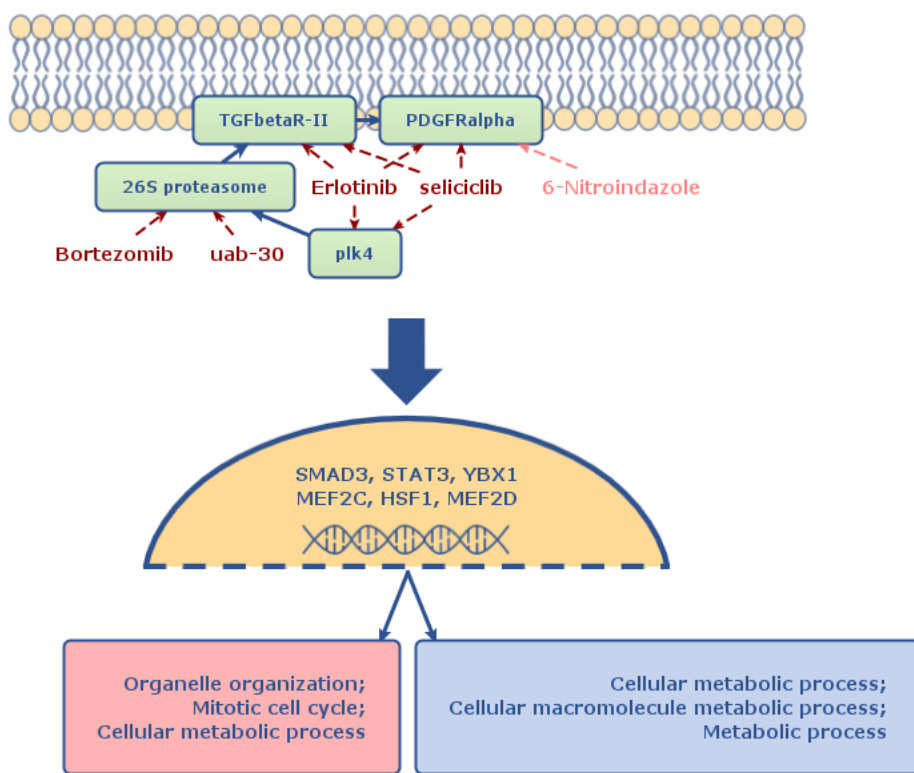
Gene symbol	Gene Description	Druggability score	logFC	Total rank
PDGFRA	platelet derived growth factor receptor alpha	6.48	2.83	310
PSMD5	proteasome 26S subunit, non-ATPase 5	1.28	0.52	384
PSMA7	proteasome 20S subunit alpha 7	2.17	0.52	384
DUSP9	dual specificity phosphatase 9	4.91	0.75	459
AURKB	aurora kinase B	1.99	1.03	463
AURKA	aurora kinase A	1.06	0.65	610



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:

- TGFbetaR-II
- plk4
- PDGFRalpha
- 26S proteasome

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: Bortezomib, 6-Nitroindazole, Erlotinib, seliciclib and uab-30, 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 for Oncology



Table 12. Clinically approved (FDA, EMA, etc.) drugs for the studied pathology (most promising and clinically approved 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	Approved
Paclitaxel	PIK3CA, CASP3, E2F1, BIRC5, CDK1, CDK2, MAPK3, BRCA1	85	12	Phase 4: Ovarian Neoplasms, Adenocarcinoma, Breast Neoplasms, Carcinoma, Carcinoma, Large Cell, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Squamous Cell, Lung Neoplasms, Neoplasms, Neuroendocrine Tumors, Squamous Cell Carcinoma of Head and Neck, Triple Negative Breast Neoplasms	Ovarian Neoplasms ( <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">FDA</a> , <a href="#">FDA</a> )
Gemcitabine	ERBB2, HRAS, CHEK1, BRCA1	72	7	Phase 3: Ovarian Neoplasms, Adenocarcinoma, Adenocarcinoma, Clear Cell, Adenocarcinoma, Mucinous, Adenomyoepithelioma, Adenoviridae Infections, Anxiety, Bile Duct Neoplasms, Biliary Tract Neoplasms, Brain Neoplasms, Breast Neoplasms, Brenner Tumor, Bronchial Diseases, Bronchial Neoplasms, Carcinoma, Carcinoma in Situ, Carcinoma, Acinar Cell, Carcinoma, Adenosquamous, Carcinoma, Bronchogenic, Carcinoma, Ductal, Carcinoma, Endometrioid, Carcinoma, Hepatocellular, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Pancreatic Ductal, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Carcinoma, Transitional Cell, Cholangiocarcinoma, Cystadenocarcinoma, Cystadenocarcinoma, Serous, Cysts, Digestive System Diseases, Digestive System Neoplasms, Drug-Related Side Effects and Adverse Reactions, Embolism, Endocrine Gland Neoplasms, Endocrine System Diseases, Epstein-Barr Virus Infections, Fallopian Tube Neoplasms, Fever, Fibrosis, Gallbladder Neoplasms, Gastrointestinal Diseases, Gastrointestinal Neoplasms, Genital Neoplasms, Female, Granuloma, Lethal Midline, Head and Neck Neoplasms, Hepatoblastoma, Hernia, Hernia, Ventral, Hodgkin Disease, Hydrothorax, Hypersensitivity, Hyperthermia, Immunoblastic Lymphadenopathy, Infections, Intestinal Diseases, Intestinal Neoplasms, Kidney Neoplasms, Klatzkin Tumor, Leiomyosarcoma, Liver Cirrhosis, Liver Neoplasms, Lung Diseases, Lung Neoplasms, Lymphadenopathy, Lymphatic Diseases, Lymphoma, Lymphoma, B-Cell, Lymphoma, Extranodal NK-T-Cell, Lymphoma, Follicular, Lymphoma, Large B-Cell, Diffuse, Lymphoma, Mantle-Cell, Lymphoma, Non-Hodgkin, Lymphoma, T-Cell, Lymphoma, T-Cell, Peripheral, Mesothelioma, Mesothelioma, Malignant, Muscle Neoplasms, Myoepithelioma, Myosarcoma, Nasopharyngeal Carcinoma, Nasopharyngeal Diseases, Nasopharyngeal Neoplasms, Neoplasm Metastasis, Neoplasm Recurrence, Local, Neoplasms, Neoplasms by Histologic Type, Neoplasms by Site, Neoplasms, Germ Cell and Embryonal, Neoplasms, Glandular and Epithelial, Neoplasms, Nerve Tissue, Neoplasms, Unknown Primary, Osteosarcoma, Pancreatic Cyst, Pancreatic Diseases, Pancreatic Intraductal Neoplasms, Pancreatic Neoplasms, Pelvic Neoplasms, Peritoneal Neoplasms, Pharyngeal Diseases, Pharyngeal Neoplasms, Prostatic Neoplasms, Recurrence, Renal Insufficiency, Respiratory Tract Diseases, Respiratory Tract Neoplasms, Sarcoma, Small Cell Lung Carcinoma, Thoracic Neoplasms, Thromboembolism, Triple Negative Breast Neoplasms, Ureteral Neoplasms, Urethral Neoplasms, Urinary Bladder Neoplasms, Urogenital Neoplasms, Uterine Cervical Neoplasms, Virus Diseases	Ovarian Neoplasms ( <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">FDA</a> )
Doxorubicin	MAPK14, PIK3CA, BAX, BIRC5, BRCA1, CDKN1B	71	12	Phase 4: Ovarian Neoplasms, Brain Abscess, Breast Neoplasms, Burkitt Lymphoma, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Ovarian Epithelial, Enteropathy-Associated T-Cell Lymphoma, Immunoblastic Lymphadenopathy, Leukemia, Leukemia, Lymphoid, Lymphadenopathy, Lymphatic Diseases, Lymphoma, Lymphoma, B-Cell, Lymphoma, Follicular, Lymphoma, Large B-Cell, Diffuse, Lymphoma, Large-Cell, Anaplastic, Lymphoma, Mantle-Cell, Lymphoma, Non-Hodgkin, Lymphoma, T-Cell, Lymphoma, T-Cell, Peripheral, Lymphoproliferative Disorders, Mediastinal Neoplasms, Multiple Myeloma, Myosarcoma, Neoplasms, Neoplasms, Plasma Cell, Obesity, Panniculitis, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Rhabdomyosarcoma, Sarcoma, Urinary Bladder Neoplasms, Wilms Tumor	Ovarian Neoplasms ( <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">ClinicalTrials</a> , <a href="#">DailyMed</a> )
Olaparib	PARP1	61	12	Phase 4: Ovarian Neoplasms, Breast Neoplasms, Carcinoma, Ovarian Epithelial, Neoplasms, Prostatic Neoplasms	Ovarian Neoplasms ( <a href="#">FDA</a> , <a href="#">FDA</a> )

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



## Drugs approved in clinical trials



Table 13. 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
Erlotinib	STK10, IKBKE, PAK2, PRKACA, MAP3K11, NEK2, NEK6, ERBB2, MAPK3, MAP2K6, TYK2, MELK, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, CSNK2A1, PLK4, STK3, RPS6KA3, CAMK2G, MET, MAPK6, TTK, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, ILK, MAPK14, PLK1, AKT1, AURKA, CHEK1, MAP3K20, BIRC5, PIM2, CDK8, MAPK9, CDK9, PIK3CA, TGFB2, PAK1, CDK5, CDK2, PRKD2	97	6	Phase 3: Ovarian Neoplasms, Adenocarcinoma, Adenocarcinoma of Lung, Brain Neoplasms, Carcinoma, Carcinoma, Acinar Cell, Carcinoma, Adenosquamous, Carcinoma, Hepatocellular, Carcinoma, Large Cell, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Colorectal Neoplasms, Disease Progression, Esophageal Neoplasms, Fallopian Tube Neoplasms, Head and Neck Neoplasms, Lip Neoplasms, Lung Neoplasms, Mouth Neoplasms, Neoplasm Metastasis, Neoplasms, Pancreatic Intraductal Neoplasms, Pancreatic Neoplasms, Peritoneal Neoplasms, Rectal Neoplasms, Small Cell Lung Carcinoma, Squamous Cell Carcinoma of Head and Neck, Thoracic Neoplasms
Pazopanib	STK10, RPS6KA3, CAMK2G, MET, IKBKE, MAPK6, PAK2, PRKACA, TTK, MAP3K11, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, NEK2, PLK1, NEK6, ERBB2, AKT1, AURKA, CHEK1, MAP3K20, MAPK3, MAP2K6, TYK2, MELK, PIM2, CDK8, MAPK9, CDK9, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1,	97	7	Phase 3: Ovarian Neoplasms, Anemia, Carcinoma, Carcinoma, Non-Small-Cell Lung, Carcinoma, Renal Cell, Chondrosarcoma, Chondrosarcoma, Mesenchymal, Dilatation, Pathologic, Epistaxis, Fibrosarcoma, Glomus Tumor, Granular Cell Tumor, Hemangiosarcoma, Histiocytoma, Histiocytoma, Benign Fibrous, Histiocytoma, Malignant Fibrous, Leiomyosarcoma, Liposarcoma, Lung Neoplasms, Myosarcoma, Neoplasms, Nerve Sheath Neoplasms, Neurofibrosarcoma, Osteosarcoma, Ranula, Sarcoma, Sarcoma, Alveolar Soft Part, Sarcoma, Clear Cell, Sarcoma, Synovial, Telangiectasia, Hereditary Hemorrhagic, Telangiectasis



	PTK2, PIK3CA, TGFB2, PAK1, CDK5, CSNK2A1, CDK2, PRKD2, PLK4, STK3			
Sorafenib	STK10, IKBKE, PAK2, PRKACA, MAP3K11, NEK2, NEK6, ERBB2, MAPK3, MAP2K6, TYK2, MELK, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, CSNK2A1, CHEK2, PLK4, STK3, RPS6KA3, CAMK2G, MET, MAPK6, TTK, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, PLK1, AKT1, AURKA, CHEK1, MAP3K20, HIPK2, PIM2, CDK8, MAPK9, CDK9, PIK3CA, TGFB2, PAK1, CDK5, CDK2, PRKD2	94	4	Phase 2: Ovarian Neoplasms, Acute Disease, Adenocarcinoma, Adenocarcinoma of Lung, Adenocarcinoma, Follicular, Adenoma, Adenoma, Liver Cell, Adrenal Cortex Neoplasms, Adrenocortical Carcinoma, Astrocytoma, Bile Duct Neoplasms, Biliary Tract Neoplasms, Brain Abscess, Brain Neoplasms, Breast Neoplasms, Breast Neoplasms, Male, Carcinoid Tumor, Carcinoma, Carcinoma, Ductal, Carcinoma, Hepatocellular, Carcinoma, Islet Cell, Carcinoma, Medullary, Carcinoma, Neuroendocrine, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Renal Cell, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Carcinoma, Transitional Cell, Carcinoma, Verrucous, Carcinosarcoma, Central Nervous System Neoplasms, Cholangiocarcinoma, Colonic Neoplasms, Colorectal Neoplasms, Desmoplastic Small Round Cell Tumor, Digestive System Neoplasms, Disease Progression, Endocrine Gland Neoplasms, Esophageal Neoplasms, Fallopian Tube Neoplasms, Fibroma, Fibrosarcoma, Fibrosis, Gallbladder Neoplasms, Gastrinoma, Gastrointestinal Neoplasms, Gastrointestinal Stromal Tumors, Glioblastoma, Glioma, Gliosarcoma, Glucagonoma, Head and Neck Neoplasms, Hemangiosarcoma, Hepatitis, Hepatitis A, Hepatitis B, Hepatitis C, Hepatoblastoma, Hepatopulmonary Syndrome, Histiocytoma, Histiocytoma, Benign Fibrous, Histiocytoma, Malignant Fibrous, Hypertension, Hypertension, Portal, Hypopharyngeal Neoplasms, Immunoblastic Lymphadenopathy, Insulinoma, Intestinal Neoplasms, Keloid, Kidney Diseases, Kidney Neoplasms, Klatkin Tumor, Laryngeal Diseases, Laryngeal Neoplasms, Leiomyosarcoma, Leukemia, Leukemia, Biphenotypic, Acute, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, Leukemia, Monocytic, Acute, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Leukemia, Myeloid, Acute, Leukemia, Myelomonocytic, Chronic, Leukemia, Myelomonocytic, Juvenile, Leukemia, Promyelocytic, Acute, Leukemia, T-Cell, Leukemia-Lymphoma, Adult T-Cell, Liver Cirrhosis, Liver Diseases, Liver Neoplasms, Lung Neoplasms, Lymphadenopathy, Lymphatic Diseases, Lymphoma, 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, Malignant Carcinoid Syndrome, Melanoma, Mesothelioma, Mesothelioma, Malignant, Metaplasia, Mixed Tumor, Mullerian, Multiple Endocrine Neoplasia, Multiple Endocrine Neoplasia Type 2a, Multiple Endocrine Neoplasia Type 2b, Multiple Myeloma, Myelodysplastic Syndromes, Myeloproliferative Disorders, Myosarcoma, Nasopharyngeal Carcinoma, Nasopharyngeal Neoplasms, Neoplasm Metastasis, Neoplasms, Neoplasms by Histologic Type, Neoplasms by Site, Neoplasms, Glandular and Epithelial, Neoplasms, Plasma Cell, Neoplasms, Second Primary, Neoplasms, Squamous Cell, Neoplasms, Unknown Primary, Nerve Sheath Neoplasms, Nervous System Neoplasms, Neuroblastoma, Neuroectodermal Tumors, Neuroectodermal Tumors, Primitive, Neuroectodermal Tumors, Primitive, Peripheral, Neuroendocrine Tumors, Neurofibroma, Neurofibromatosis 1, Neurofibrosarcoma, Oropharyngeal Neoplasms, Osteosarcoma, Pancreatic Neoplasms, Paranasal Sinus Neoplasms, Peritoneal Neoplasms, Pharyngeal Neoplasms, Plasmablastic Lymphoma, Plasmacytoma, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Preleukemia, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Rectal Neoplasms, Recurrence, Retroviridae Infections, Rhabdomyosarcoma, Rhabdomyosarcoma, Embryonal, Salivary Gland Neoplasms, Sarcoma, Sarcoma, Ewing, Sarcoma, Synovial, Skin Neoplasms, Small Cell Lung Carcinoma, Somatostatinoma, Squamous Cell Carcinoma of Head and Neck, Stomach Neoplasms, Syndrome, Testicular Neoplasms, Thrombosis, Thyroid Cancer, Papillary, Thyroid Carcinoma, Anaplastic, Thyroid Diseases, Thyroid Neoplasms, Tongue Neoplasms, Triple Negative Breast Neoplasms, Ureteral Neoplasms, Urethral Neoplasms, Urinary Bladder Neoplasms, Uterine Cervical Neoplasms, Uveal Neoplasms, Vaccinia, Vipoma, Wilms Tumor
Lapatinib	STK10, RPS6KA3, CAMK2G, MET, IKBKE, MAPK6, PAK2, PRKACA, TTK, MAP3K11, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, PLK1, NEK6, ERBB2, AKT1, AURKA, CHEK1, MAP3K20, BIRC5, MAPK3, MAP2K6, TYK2, MELK, PIM2, CDK8, MAPK9, CDK9, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1,	93	3	Phase 2: Ovarian Neoplasms, Adenocarcinoma, Adenoma, Astrocytoma, Bile Duct Neoplasms, Brain Neoplasms, Breast Diseases, Breast Neoplasms, Breast Neoplasms, Male, Carcinoma, Carcinoma, Acinar Cell, Carcinoma, Adenoid Cystic, Carcinoma, Ductal, Carcinoma, Ductal, Breast, Carcinoma, Hepatocellular, Carcinoma, Mucoepidermoid, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Carcinoma, Transitional Cell, Carcinoma, Verrucous, Central Nervous System Neoplasms, Cholangiocarcinoma, Colorectal Neoplasms, Cysts, Dermoid Cyst, Diarrhea, Digestive System Neoplasms, Endometrial Neoplasms, Ependymoma, Esophageal Neoplasms, Esophageal Squamous Cell Carcinoma, Fibroma, Gallbladder Neoplasms, Gastrointestinal Neoplasms, Glioblastoma, Glioma, Gliosarcoma, Head and Neck Neoplasms, Hypopharyngeal Neoplasms, Inflammatory Breast Neoplasms, Intestinal Neoplasms, Laryngeal Diseases, Laryngeal Neoplasms, Liver Neoplasms, Lung Neoplasms, Medulloblastoma, Melanoma, Mucoepidermoid Tumor, Nasopharyngeal Carcinoma, Neoplasm Metastasis, Neoplasms, Neoplasms, Second Primary, Neoplasms, Unknown Primary, Neoplastic Cells, Circulating, Nervous System Neoplasms, Neurilemmoma, Neuroblastoma, Neurofibroma, Neurofibromatosis, Neurofibromatosis 1, Neurofibromatosis 2, Neuroma, Neuroma, Acoustic, Oligodendroglioma, Oropharyngeal Neoplasms, Pancreatic Neoplasms, Paranasal Sinus Neoplasms, Peritoneal Neoplasms, Pharyngeal Neoplasms, Pituitary Diseases, Pituitary Neoplasms, Prolactinoma, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Rectal Neoplasms, Recurrence, Salivary Gland Neoplasms, Sarcoma, Small Cell Lung Carcinoma, Spinal Cord Neoplasms, Squamous Cell Carcinoma of Head and Neck, Stomach Neoplasms, Thymoma, Tongue Neoplasms, Urinary Bladder Neoplasms, Uterine Cervical Neoplasms, Uterine Neoplasms



	PTK2, PIK3CA, TGFB2, PAK1, CDK5, CSNK2A1, CDK2, PRKD2, PLK4, STK3		
Vandetanib	STK10, RPS6KA3, CAMK2G, MET, IKBKE, MAPK6, PAK2, PRKACA, TTK, CDK1, MAP3K11, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, NEK2, PLK1, NEK6, ERBB2, AKT1, AURKA, CHEK1, MAP3K20, MAPK3, MAP2K6, TYK2, MELK, PIM2, CDK8, MAPK9, CDK9, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, PIK3CA, TGFB2, PAK1, CDK5, CSNK2A1, CDK2, PRKD2, PLK4, STK3	92	3

Phase 2: Ovarian Neoplasms, Adenocarcinoma, Adenocarcinoma, Follicular, Astrocytoma, Biliary Tract Neoplasms, Brain Abscess, Breast Neoplasms, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Medullary, Carcinoma, Neuroendocrine, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Renal Cell, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Carcinoma, Transitional Cell, Carcinoma, Verrucous, Cholangiocarcinoma, Colorectal Neoplasms, Endocrine Gland Neoplasms, Fallopian Tube Neoplasms, Gallbladder Neoplasms, Gastrointestinal Stromal Tumors, Glioblastoma, Glioma, Gliosarcoma, Head and Neck Neoplasms, Kidney Neoplasms, Leiomyoma, Leiomyomatosis, Leukemia, Leukemia, Myeloid, Leukemia, Myeloid, Acute, Lung Neoplasms, Mesothelioma, Mesothelioma, Malignant, Multiple Endocrine Neoplasia, Multiple Endocrine Neoplasia Type 2a, Multiple Endocrine Neoplasia Type 2b, Multiple Myeloma, Myoma, Neoplasm Metastasis, Neoplasms, Neoplasms, Plasma Cell, Oligodendroglioma, Pancreatic Neoplasms, Peritoneal Neoplasms, Pleural Effusion, Pleural Effusion, Malignant, Precancerous Conditions, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Rectal Neoplasms, Sarcoma, Small Cell Lung Carcinoma, Squamous Cell Carcinoma of Head and Neck, Stomach Neoplasms, Thyroid Cancer, Papillary, Thyroid Diseases, Thyroid Neoplasms, Ureteral Neoplasms, Urethral Neoplasms, Urinary Bladder Neoplasms, von Hippel-Lindau Disease

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
seliciclib	STK10, IKBKE, PAK2, PRKACA, MAP3K11, NEK2, NEK6, ERBB2, MAPK3, MAP2K6, TYK2, MELK, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, CSNK2A1, CHEK2, PLK4, STK3, RPS6KA3, CAMK2G, MET, MAPK6, TTK, CDK1, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, PLK1, AKT1, AURKA, CHEK1, MAP3K20, HIPK2, PIM2, CDK8, MAPK9, CDK9, PIK3CA, TGFBR2, PAK1, CDK5, CDK2, PRKD2	76	Phase 2: ACTH-Secreting Pituitary Adenoma, Adenoma, Carcinoma, Non-Small-Cell Lung, Cystic Fibrosis, Cysts, Fibrosis, Pituitary ACTH Hypersecretion, Pituitary Neoplasms
1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea	STK10, IKBKE, PAK2, PRKACA, MAP3K11, NEK2, NEK6, ERBB2, MAPK3, MAP2K6, TYK2, MELK, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, CSNK2A1, CHEK2, PLK4, STK3, RPS6KA3, CAMK2G, MET, MAPK6, TTK, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, PLK1, AKT1, AURKA, CHEK1, MAP3K20, HIPK2, PIM2, CDK8, MAPK9, CDK9, PIK3CA, TGFBR2, PAK1, CDK5, CDK2, PRKD2	76	Phase 2: Arthritis, Arthritis, Rheumatoid, Psoriasis
ruboxistaurin	STK10, IKBKE, PAK2, PRKACA, MAP3K11, NEK2, NEK6, ERBB2, MAPK3, MAP2K6, TYK2, MELK, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, CSNK2A1, CHEK2, PLK4, STK3, RPS6KA3, CAMK2G, MET, MAPK6, TTK, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, PLK1, AKT1, AURKA, CHEK1, MAP3K20, HIPK2, PIM2, CDK8, MAPK9, CDK9, PIK3CA, TGFBR2, PAK1, CDK5, CDK2, PRKD2	76	Phase 3: Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetes Mellitus, Type 2, Diabetic Neuropathies, Diabetic Retinopathy, Edema, Macular Edema, Nervous System Diseases, Peripheral Nervous System Diseases, Retinal Diseases
midostaurin	STK10, IKBKE, PAK2, PRKACA, MAP3K11, NEK2, NEK6, ERBB2, MAPK3, MAP2K6, TYK2, MELK, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, CASP3, CSNK2A1, PLK4, STK3, RPS6KA3, CAMK2G, MET, MAPK6, TTK, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, PLK1, AKT1, AURKA, CHEK1, MAP3K20, PIM2, CDK8, MAPK9, CASP7, CDK9, PIK3CA, TGFBR2, PAK1, CDK5, CDK2, PRKD2	76	Phase 3: Anemia, Anemia, Refractory, Anemia, Refractory, with Excess of Blasts, Leukemia, Leukemia, Myeloid, Leukemia, Myeloid, Acute, Myelodysplastic Syndromes, Preleukemia, Syndrome
CHIR-258	STK10, RPS6KA3, CAMK2G, MET, IKBKE, MAPK6, PAK2, PRKACA, TTK, MAP3K11, PDGFRA, RAF1, AURKB, CDK7, RPS6KA1, MAPK14, NEK2, PLK1, NEK6, ERBB2, AKT1, AURKA, CHEK1, MAP3K20, MAPK3, MAP2K6, TYK2, MELK, PIM2, CDK8, MAPK9, CDK9, PRKD3, PDGFRB, MAP2K1, MAP4K3, WEE1, PTK2, PIK3CA, TGFBR2, PAK1, CDK5, CSNK2A1, CDK2, PRKD2, PLK4, STK3	75	Phase 3: Carcinoma, Carcinoma, Renal Cell, Neoplasms

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



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 score	Target activity score
Bortezomib	PSMC5, PSMD13, PSMA7, PSMC3, PSMD2, PSMD1, PSMD14, PSMD5, PSMC6, CASP3, PSMD11, PSMC4, PSMC1, PSMC2, PSMD12, PSMD7, PSMD3	99	1.83
N-(4-MORPHOLINE)CARBONYL-B-(1-NAPHTHYL)-L-ALANINE-L-LEUCINE BORONIC ACID	PSMC5, PSMD13, PSMA7, PSMC3, PSMD2, PSMD1, PSMD14, PSMD5, PSMC6, PSMD11, PSMC4, PSMC1, PSMC2, PSMD12, PSMD7, PSMD3	98	1.43
6-Nitroindazole	RPS6KA3, CAMK2G, CDK9, PRKD3, GRK5, PDGFRB, PRKACA, CDK1, PDGFRA, RPS6KA1, CDK7, PAK1, CHEK1, CDK5, IRAK1, MAPK3, CDK2, CHEK2, MAP2K6	93	1.77
3,5-Diaminophthalhydrazide	RPS6KA3, NEK2, PARP1, CAMK2G, PRKD3, CHEK1, IRAK1, CHEK2, RPS6KA1	92	0.43
2,5,7-Trihydroxynaphthoquinone	MAPK14, CDC25A, MAPK9, CDKN3, MAPK6, CDC25B, DUSP14, DUSP9, MAPK3, BRCA1	91	0.32

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Erlotinib, seliciclib and Bortezomib. These drugs were selected for acting on the following targets: PDGFRA and PSMD5, 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 16. 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	34
Abiraterone	Prostatic Neoplasms, Castration-Resistant	-
Abiraterone acetate	Prostatic Neoplasms, Castration-Resistant	-
Acalabrutinib	Lymphoma, Mantle-Cell	-



Acitretin	Psoriasis	26
Ado-trastuzumab emtansine	Breast Neoplasms Neoplasms	80
Afatinib	Carcinoma, Non-Small-Cell Lung	38
Aflibercept	Colorectal Neoplasms Diabetic Retinopathy Edema Vascular Diseases Wet Macular Degeneration	-
Alectinib	Carcinoma, Non-Small-Cell Lung	24
Alemtuzumab	Brain Abscess Leukemia, Lymphocytic, Chronic, B-Cell Multiple Sclerosis Multiple Sclerosis, Relapsing-Remitting Sclerosis	-
Alitretinoin	Sarcoma, Kaposi	-
Alpelisib	Breast Neoplasms	71
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	60
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	61
Azacitidine	Anemia, Refractory Anemia, Refractory, with Excess of Blasts Leukemia, Myelomonocytic, Chronic Myelodysplastic Syndromes Preleukemia Syndrome	-
Belinostat	Lymphoma, T-Cell, Peripheral	37
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	23
Binimetinib	Melanoma	56
Blinatumomab	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Bortezomib	Brain Abscess Glomerulonephritis Glomerulonephritis, IGA Kidney Diseases Multiple Myeloma Neoplasms, Plasma Cell Nephritis Renal Insufficiency	74
Bosutinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	65
Brentuximab vedotin	Hodgkin Disease Lymphoma Lymphoma, Large-Cell, Anaplastic Lymphoma, T-Cell, Peripheral	-
Brigatinib	Carcinoma, Non-Small-Cell Lung	33
Buserelin	Prostatic Neoplasms	-
Cabazitaxel	Prostatic Neoplasms, Castration-Resistant	65
Cabergoline	Drug-Related Side Effects and Adverse Reactions Pituitary Neoplasms	-
Cabozantinib	Thyroid Neoplasms	7
Capecitabine	Breast Neoplasms Colonic Neoplasms Colorectal Neoplasms	-
Carboplatin	Carcinoma, Non-Small-Cell Lung Lung Neoplasms Neoplasms Neuroendocrine Tumors Ovarian Neoplasms Retinoblastoma	-
Carfilzomib	Multiple Myeloma	68
Carmustine	Astrocytoma Glioblastoma Hodgkin Disease Medulloblastoma Multiple Myeloma Neoplasms	33
Ceritinib	Carcinoma, Non-Small-Cell Lung	65
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	67
Cladribine	Leukemia, Hairy Cell	-
Clofarabine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
Cobimetinib	Melanoma	46
Copanlisib	Lymphoma, Follicular	72
Crizotinib	Carcinoma, Non-Small-Cell Lung	25
Cyproterone acetate	Prostatic Neoplasms	-
Dabrafenib	Melanoma	27
Dacomitinib	Carcinoma, Non-Small-Cell Lung	53
Daratumumab	Multiple Myeloma	-
Dasatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive Leukemia, Myeloid, Chronic-Phase Precursor Cell Lymphoblastic Leukemia-Lymphoma	90
Decitabine	Anemia, Refractory Anemia, Refractory, with Excess of Blasts Leukemia, Myelomonocytic, Chronic Myelodysplastic Syndromes	38
Degarelix	Cardiovascular Diseases Prostatic Neoplasms Vascular Diseases	30
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	33
Doxorubicin	Neoplasms Multiple Myeloma Carcinoma, Ovarian Epithelial Ovarian Neoplasms Leukemia, Lymphoid Breast Neoplasms Lymphoma, Follicular Thyroid Neoplasms Triple Negative Breast Neoplasms Glioma	71
Durvalumab	Carcinoma, Non-Small-Cell Lung Carcinoma, Transitional Cell	-
Dutasteride	Alcoholism Hyperplasia Hypertrophy Neoplasms Prostatic Hyperplasia	-
Duvelisib	Leukemia, Lymphocytic, Chronic, B-Cell Lymphoma, Follicular	11
Elotuzumab	Multiple Myeloma	33
Enasidenib	Leukemia, Myeloid, Acute	-
Encorafenib	Colorectal Neoplasms Melanoma	49
Enfortumab vedotin	Carcinoma, Transitional Cell Neoplasms	-



Entrectinib	Carcinoma, Non-Small-Cell Lung	-
Enzalutamide	Prostatic Neoplasms Prostatic Neoplasms, Castration-Resistant	-
Epirubicin	Breast Neoplasms	64
Erdafitinib	Urinary Bladder Neoplasms	77
Eribulin	Breast Neoplasms Drug-Related Side Effects and Adverse Reactions Neoplasms	-
Erlotinib	Carcinoma, Non-Small-Cell Lung Neoplasms Pancreatic Neoplasms	97
Erlotinib hydrochloride	Carcinoma, Non-Small-Cell Lung Gastrointestinal Stromal Tumors	-
Estramustine	Prostatic Neoplasms	7
Ethinyl Estradiol	Acne Vulgaris Neoplasms	20
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	63
Exemestane	Breast Neoplasms	-
Fedratinib	Primary Myelofibrosis	-
Finasteride	Hyperplasia Neoplasms Prostatic Hyperplasia	17
Flavopiridol	Leukemia, Lymphocytic, Chronic, B-Cell	84
Fluorouracil	Skin Neoplasms Neoplasms, Basal Cell Neoplasms, Second Primary Neoplasms, Squamous Cell Neoplasms Colorectal Neoplasms Pancreatic Neoplasms	76
Fluoxymesterone	Breast Neoplasms Hypogonadism Puberty, Delayed	-
Flutamide	Premenstrual Dysphoric Disorder Premenstrual Syndrome Prostatic Neoplasms	39
Fulvestrant	Breast Neoplasms	-
Gefitinib	Carcinoma, Non-Small-Cell Lung	90
Gemcitabine	Breast Neoplasms Carcinoma, Non-Small-Cell Lung Ovarian Neoplasms Pancreatic Neoplasms	72
Gemtuzumab ozogamicin	Leukemia, Myeloid, Acute	-
Gilteritinib	Leukemia, Myeloid, Acute	54
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	72
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	55
Idarubicin	Leukemia, Myeloid, Acute	-
Idelalisib	Leukemia, Lymphocytic, Chronic, B-Cell Lymphoma, Follicular	22
Ifosfamide	Neoplasms	-
Imatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive Mastocytosis, Systemic Neoplasms	90
Inotuzumab ozogamicin	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Ipilimumab	Carcinoma, Renal Cell Melanoma	-
Irinotecan	Colorectal Neoplasms	63
Ivosidenib	Leukemia, Myeloid, Acute	-
Ixabepilone	Breast Neoplasms	-
Ixazomib	Multiple Myeloma	-
Lapatinib	Breast Neoplasms	93
Larotrectinib	Neoplasm Metastasis	37
Lenalidomide	Brain Abscess Lupus Erythematosus, Cutaneous Myelodysplastic Syndromes Neoplasms, Plasma Cell	-
Lenvatinib	Carcinoma, Hepatocellular Carcinoma, Renal Cell Thyroid Neoplasms	69
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	53
Lorlatinib	Carcinoma, Non-Small-Cell Lung	11
Masoprocol	Keratosis, Actinic	-
Medroxyprogesterone Acetate	Depression Depression, Postpartum Depressive Disorder Metrorrhagia Neoplasms Uterine Hemorrhage	6
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	46
Methyltestosterone	Breast Neoplasms Hypogonadism Puberty, Delayed	-
Midostaurin	Leukemia, Mast-Cell Leukemia, Myeloid, Acute Mastocytosis, Systemic	76
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	45
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	59
Nilotinib	Blast Crisis Leukemia, Myelogenous, Chronic, BCR-ABL Positive Leukemia, Myeloid, Chronic-Phase	64
Nilutamide	Prostatic Neoplasms	-
Nintedanib	Fibrosis Idiopathic Pulmonary Fibrosis	77
Niraparib	Carcinoma, Ovarian Epithelial Fallopian Tube Neoplasms Peritoneal Neoplasms	70
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	4



	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	61
Olaratumab	Sarcoma	-
Osimertinib	Carcinoma, Non-Small-Cell Lung	58
Oxaliplatin	Colonic Neoplasms Colorectal Neoplasms Neoplasms Rectal Neoplasms	39
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	85
Palbociclib	Breast Neoplasms	-
Panitumumab	Colorectal Neoplasms	-
Panobinostat	Multiple Myeloma	-
Pazopanib	Carcinoma Carcinoma, Renal Cell Sarcoma	97
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	76
Pomalidomide	Multiple Myeloma	-
Ponatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive Precursor Cell Lymphoblastic Leukemia-Lymphoma	29
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	69
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	45
Rucaparib	Carcinoma, Ovarian Epithelial Fallopian Tube Neoplasms Peritoneal Neoplasms Prostatic Neoplasms, Castration-Resistant	87
Ruxolitinib	Graft vs Host Disease Polycythemia Polycythemia Vera Primary Myelofibrosis Thrombocytosis	51
Selinexor	Multiple Myeloma	-
Selumetinib	Neurofibromatosis 1	35
Siltuximab	Giant Lymph Node Hyperplasia	-
Sirolimus	Angiomyolipoma Constriction, Pathologic Coronary Restenosis Eye Diseases Immune System Diseases Kidney Failure, Chronic Lipoma Tuberous Sclerosis	88
Sonidegib	Carcinoma, Basal Cell	-
Sorafenib	Carcinoma, Hepatocellular Carcinoma, Renal Cell Thyroid Neoplasms	94
Sunitinib	Adenoma Carcinoma, Renal Cell Digestive System Neoplasms Gastrointestinal Neoplasms Gastrointestinal Stromal Tumors Intestinal Neoplasms	90
Talazoparib	Breast Neoplasms	57
Tamoxifen	Breast Diseases Cystic Fibrosis Cysts Fibroadenoma Fibrocystic Breast Disease Hemorrhage Menorrhagia Menstruation Disturbances Metrorrhagia Neoplasms	46
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	69
Teniposide	Precursor Cell Lymphoblastic Leukemia-Lymphoma	42
Thalidomide	Brain Abscess Immune System Diseases Multiple Myeloma Neoplasms, Plasma Cell	-
Tivozanib	Carcinoma, Renal Cell	74
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	47
Toremifene	Breast Neoplasms	29
Trabectedin	Leiomyosarcoma Liposarcoma	-
Trametinib	Carcinoma, Non-Small-Cell Lung Melanoma	81
Trastuzumab	Breast Neoplasms Neoplasms	52
Tretinoin	Lentigo	46
Triptorelin	Fatty Liver Hypogonadism Infertility, Female Prostatic Neoplasms	53
Tucatinib	Breast Neoplasms	64
Valrubicin	Urinary Bladder Neoplasms	-
Vandetanib	Thyroid Neoplasms	92
Vemurafenib	Melanoma	51
Venetoclax	Leukemia, Lymphocytic, Chronic, B-Cell Leukemia, Myeloid, Acute	-
Vinblastine	Glioma	-
Vincristine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
Vinorelbine	Carcinoma, Non-Small-Cell Lung	52
Vismodegib	Carcinoma, Basal Cell	-
Vorinostat	Lymphoma, T-Cell, Cutaneous	51
Zoledronate	Arthritis Bone Marrow Diseases Brain Abscess Chronic Kidney Disease-Mineral and Bone Disorder Chronic Periodontitis HIV Infections Hypersensitivity Infections Kidney Diseases Metabolic Diseases Multiple Myeloma Neoplasms Neoplasms, Plasma Cell Neoplasms, Second Primary Osteitis Osteoarthritis Periodontitis Pleural Effusion, Malignant Prostatic Neoplasms Renal Insufficiency, Chronic Thalassemia Wounds and Injuries	-

## 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:



**Erlotinib, seliciclib and Bortezomib**

These drugs were selected for acting on the following targets: PDGFRA and PSMD5, 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:



**TGFbetaR-II, plk4, PDGFRalpha and 26S proteasome**

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: Bortezomib, 6-Nitroindazole, Erlotinib, seliciclib and uab-30. 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.

- TGFbetaR-II
- plk4
- PDGFRalpha
- 26S proteasome

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®](https://genexplain.com/transfac) library, release 2023.1 (geneXplain GmbH, Wolfenbüttel, Germany) (<https://genexplain.com/transfac>).

The master regulator search uses the [TRANSPATH®](https://genexplain.com/transpath) database (BIOBASE), release 2023.1 (geneXplain GmbH, Wolfenbüttel, Germany) (<https://genexplain.com/transpath>). A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in [TRANSPATH®](https://genexplain.com/transpath).

The information about drugs corresponding to identified drug targets and clinical trials references were extracted from [HumanPSD™](https://genexplain.com/humanpsd) database, release 2023.1 (<https://genexplain.com/humanpsd>).

The Ensembl database release Human104.38 (hg38) (<http://www.ensembl.org>) was used for gene IDs representation and Gene Ontology (GO) (<http://geneontology.org>) was used for functional classification of the studied gene set.

### Methods for the analysis of enriched transcription factor binding sites and composite modules

Transcription factor binding sites in promoters and enhancers of differentially expressed genes were analyzed using known DNA-binding motifs. The motifs are specified using position weight matrices (PWMs) that give weights to each nucleotide in each position of the DNA binding motif for a transcription factor or a group of them.

We search for transcription factor binding sites (TFBS) that are enriched in the promoters and enhancers under study as compared to a background sequence set such as promoters of genes that were not differentially regulated under the condition of the experiment. We denote study and background sets briefly as Yes and No sets. In the current work we used a workflow considering promoter sequences of a standard length of 1100 bp (-1000 to +100). The error rate in this part of the pipeline is controlled by estimating the adjusted p-value (using the Benjamini-Hochberg procedure) in comparison to the TFBS frequency found in randomly selected regions of the human genome (adj.p-value < 0.01).

We have applied the CMA algorithm (Composite Module Analyst) for searching composite modules [7] in the promoters and enhancers of the Yes and No sets. We searched for a composite module consisting of a cluster of 10 TFs in a sliding window of 200-300 bp that statistically significantly separates sequences in the Yes and No sets (minimizing Wilcoxon p-value).

### Methods for finding master regulators in networks

We searched for master regulator molecules in signal transduction pathways upstream of the identified transcription factors. The master regulator search uses a comprehensive signal transduction network of human cells. The main algorithm of the master regulator search has been described earlier [3,4]. The goal of the algorithm is to find nodes in the global signal transduction network that may potentially regulate the activity of a set of transcription factors found at the previous step of the analysis. Such nodes are considered as most promising drug targets, since any influence on such a node may switch the transcriptional programs of hundreds of genes that are regulated by the respective TFs. In our analysis, we have run the algorithm with a maximum radius of 12 steps upstream of each TF in the input set. The error rate of this algorithm is controlled by applying it 10000 times to randomly generated sets of input transcription factors of the same set-size. Z-score and FDR value of ranks are calculated then for each potential master regulator node on the basis of such random runs (see detailed description in [9]). We control the error rate by the FDR threshold 0.05.

### Methods for analysis of pharmaceutical compounds



We seek for the optimal combination of molecular targets (key elements of the regulatory network of the cell) that potentially interact with pharmaceutical compounds from a library of known drugs and biologically active chemical compounds, using information about known drugs from HumanPSD™ and predicting potential drugs using PASS program.

### Method for analysis of known pharmaceutical compounds

We selected compounds from HumanPSD™ database that have at least one target. Next, we sort compounds using "Drug rank" that is the sum of the following ranks:

1. ranking by "Target activity score" ( $T\text{-score}_{PSD}$ ),
2. ranking by "Disease activity score" ( $D\text{-score}_{PSD}$ ),
3. ranking by "Clinical validity score".

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

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

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 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 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, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. **2006**;34(90001):D108-D110. doi:10.1093/nar/gkj143



7. Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* **2003**;31(13):3576-3579. doi:10.1093/nar/gkg585
8. Waleev T, Shtokalo D, Konovalova T, Voss N, Cheremushkin E, Stegmaier P, Kel-Margoulis O, Wingender E, Kel A. Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res.* **2006**;34(Web Server issue):W541-5.
9. Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.* **2006**;34(90001):D546-D551. doi:10.1093/nar/gkj107
0. Boyarskikh U, Pintus S, Mandrik N, Stelmashenko D, Kiselev I, Evshin I, Sharipov R, Stegmaier P, Kolpakov F, Filipenko M, Kel A. Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3. *BMC Med Genomics.* **2018**;11(1):12. doi:10.1186/1471-2105-7-s2-s13
1. Filimonov D, Poroikov V. Probabilistic Approaches in Activity Prediction. Varnek A, Tropsha A. *Cheminformatics Approaches to Virtual Screening.* Cambridge (UK): RSC Publishing. **2008**;:182-216.
2. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal.* **2006**;50(2):66-75 (russ)
3. Filimonov D, Poroikov V, Borodina Y, Glorizova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *ChemInform.* **1999**;39(4):666-670. doi:10.1002/chin.199940210

## Thank you for using the Genome Enhancer!

In case of any questions please contact us at [support@genexplain.com](mailto: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.