PDGFRA and DTL are promising druggable targets for treating ovarian neoplasms that control activity of E2F1, MYOD1 and SMAD3 transcription factors on of differentially expressed genes in ovary tissue

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 14/08/2019 ; Run on 18/02/2020 ; Report generated on 18/02/2020

Genome Enhancer release 1.9 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2020.1)

Abstract

In the present study we applied the software package "Genome Enhancer" to a multiomics data set that contains *transcriptomics and epigenomics* data obtained from *ovary* tissue. 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) novel biologically 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: E2F1, MYOD1, STAT5A, SMAD3, FOXO1 and CEBPA. The subsequent network analysis suggested PDGFRA, PSMA7, KAT2B, DTL and PSMC5 as the most promising and druggable molecular targets. Finally, the following drugs were identified as the most promising treatment candidates: Pazopanib, Coenzyme A, 2-ACETYLAMINO-4-METHYL-PENTANOIC ACID [1-(1-FORMYL-PENTYLCARBAMOYL)-3-METHYL-BUTYL]-AMIDE, Enprofylline, 1-3 Sugar Ring of Pentamannosyl 6-Phosphate and 4-Nitrocatechol.

1. Introduction

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

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

The "upstream analysis" approach has now been extended by a third step that reveals known drugs suitable to inhibit (or activate) the identified molecular targets in the context of the disease under study. This step is performed by using information from HumanPSD™ database [5]. In addition, new potential small molecular ligands are subsequently predicted for the revealed targets. A general druggability check is performed using a precomputed database of biologcal activities of chemical compounds from a library of about 13000 pharmaceutically most active compounds. The spectra of biological activities are computed using the program PASS on the basis of a (Q)SAR approach [11-13].

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 12520 upregulated genes (LogFC>0) out of which 8443 genes were found as significantly upregulated (p-value<0.1) and 12394 downregulated genes (LogFC<0) out of which 8285 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 voltage-gated channel subfamily J member 2	5.3	1.93E-15	3.77E-12
ENSG00000064218	DMRT3	doublesex and mab-3 related transcription factor 3	5.17	2.21E-16	9.14E-13
ENSG00000099139	PCSK5	proprotein convertase subtilisin/kexin type 5	4.46	8.58E-13	3.17E-10
ENSG00000196507	TCEAL3	transcription elongation factor A like 3	3.99	4.42E-16	1.22E-12
ENSG00000197705	KLHL14	kelch like family member 14	3.67	4.09E-15	5.9E-12
ENSG00000103449	SALL1	spalt like transcription factor 1	3.4	3.73E-12	8.39E-10
ENSG00000138378	STAT4	signal transducer and activator of transcription 4	3.38	8.43E-12	1.64E-9
ENSG00000164692	COL1A2	collagen type I alpha 2 chain	3.29	6.87E-15	7.44E-12
ENSG00000133083	DCLK1	doublecortin like kinase 1	3.29	5.68E-15	6.74E-12
ENSG00000126950	TMEM35A	transmembrane protein 35A	3.16	3.53E-15	5.5E-12

*Table 4. 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
ENSG00000149968	MMP3	matrix metallopeptidase 3	-6.61	1.61E-18	3.03E-14
ENSG00000127324	TSPAN8	tetraspanin 8	-6.09	1.74E-14	1.61E-11
ENSG00000139292	LGR5	leucine rich repeat containing G protein-coupled receptor 5	-5.53	1.25E-16	7.78E-13
ENSG00000153233	PTPRR	protein tyrosine phosphatase, receptor type R	-5.29	2.09E-16	9.14E-13
ENSG00000169908	TM4SF1	transmembrane 4 L six family member 1	-4.66	2.44E-18	3.03E-14
ENSG00000106511	MEOX2	mesenchyme homeobox 2	-4.63	9.26E-16	2.31E-12
ENSG00000163359	COL6A3	collagen type VI alpha 3 chain	-4.54	1.82E-17	1.51E-13
ENSG00000060718	COL11A1	collagen type XI alpha 1 chain	-4.53	2.8E-14	2.21E-11
ENSG00000166670	MMP10	matrix metallopeptidase 10	-4.29	1.28E-15	2.89E-12
ENSG00000145431	PDGFC	platelet derived growth factor C	-4.1	3.79E-16	1.18E-12

3.2. Regulatory regions of target genes

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

Table 3. Top 3 ***up-regulated*** *genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with epigenomic peaks.*
See full table →

ID	Gene symbol	Gene schematic representation
ENSG00000260774	CTD-2083E4.4	
ENSG00000027075	PRKCH	
ENSG00000186684	CYP27C1	

Table 5. Top 7 ***down-regulated*** *genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with epigenomic peaks.*
See full table →

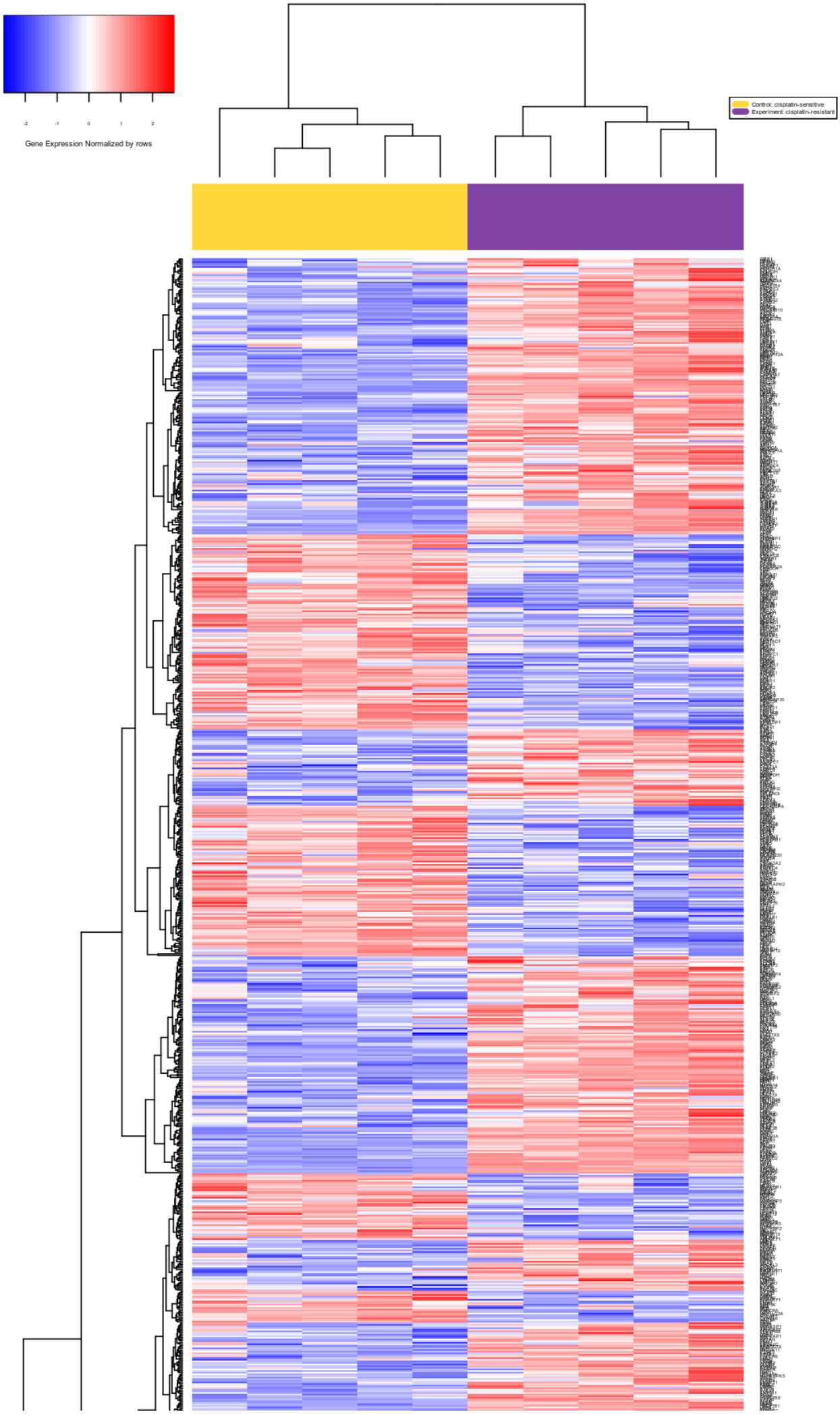
ID	Gene symbol	Gene schematic representation
ENSG00000170558	CDH2	
ENSG00000197921	HES5	
ENSG00000197822	OCLN	
ENSG00000146648	EGFR	
ENSG00000145476	CYP4V2	
ENSG00000237765	FAM200B	
ENSG00000118495	PLAGL1	

3.3. Functional classification of genes

A functional analysis of differentially expressed genes was done by mapping the significant up-regulated and significant down-regulated genes to several known ontologies, such as Gene Ontology (GO), disease ontology (based on HumanPSD™ database) and the ontology of signal transduction and metabolic pathways from the TRANSPATH® database. Statistical significance was computed using a binomial test.
Figures 3-8 show the most significant categories.

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

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


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

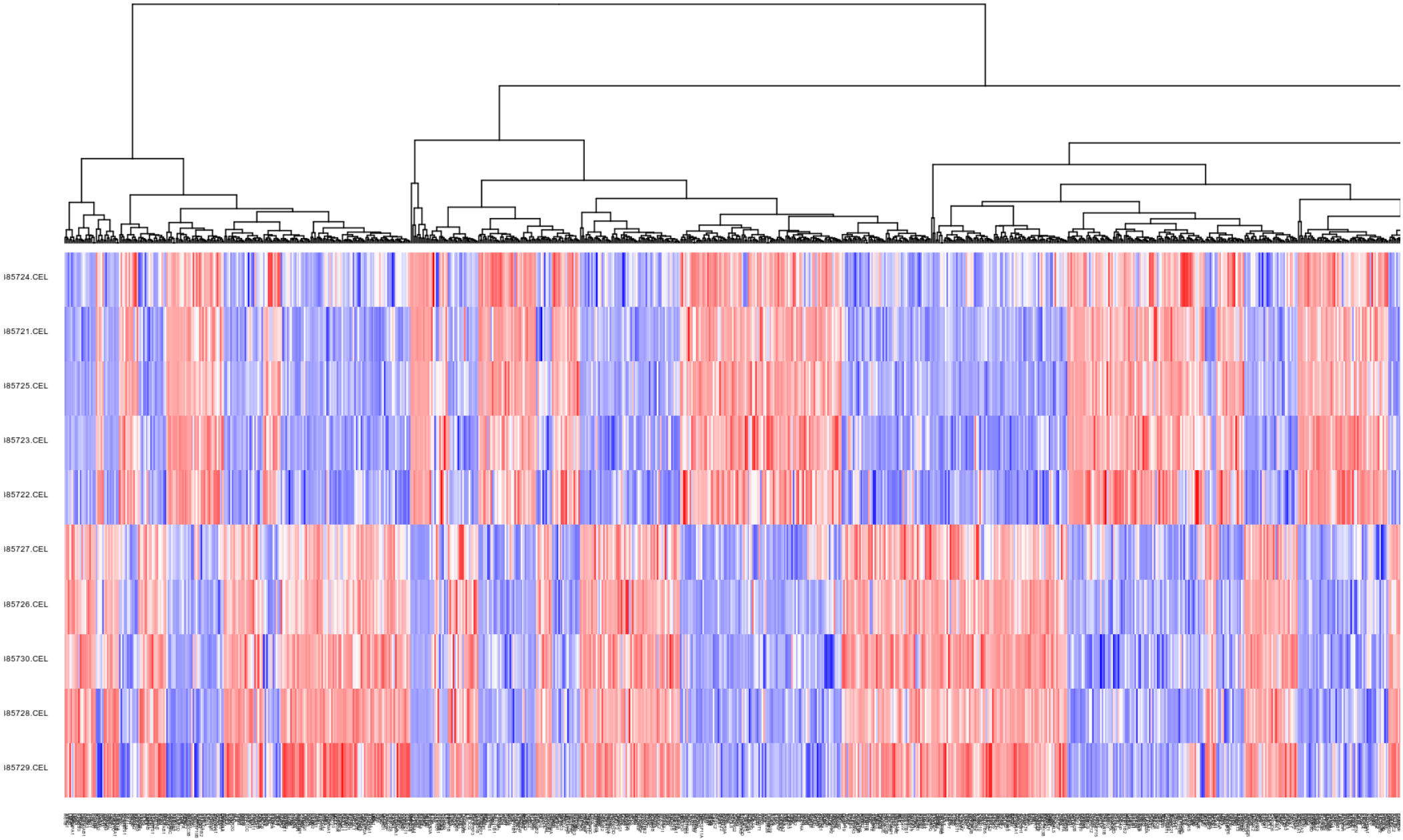

85724.CEL
85721.CEL
85725.CEL
85723.CEL
85722.CEL
85727.CEL
85726.CEL
85730.CEL
85728.CEL
85729.CEL

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

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

8443 significant up-regulated genes were taken for the mapping.

GO (biological process)

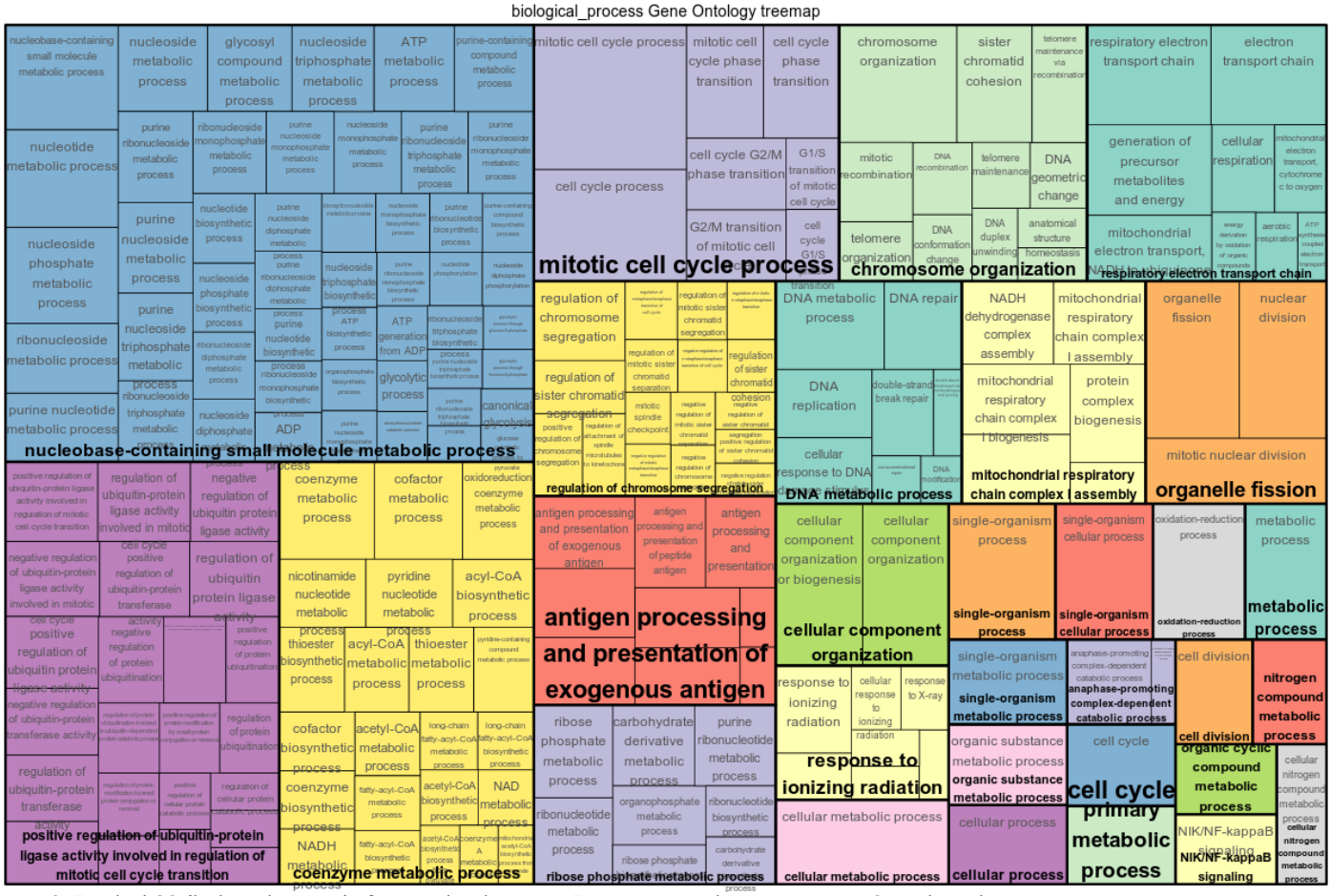


Figure 3. Enriched GO (biological process) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.
Full classification →

TRANSPATH® Pathways (2020.1)

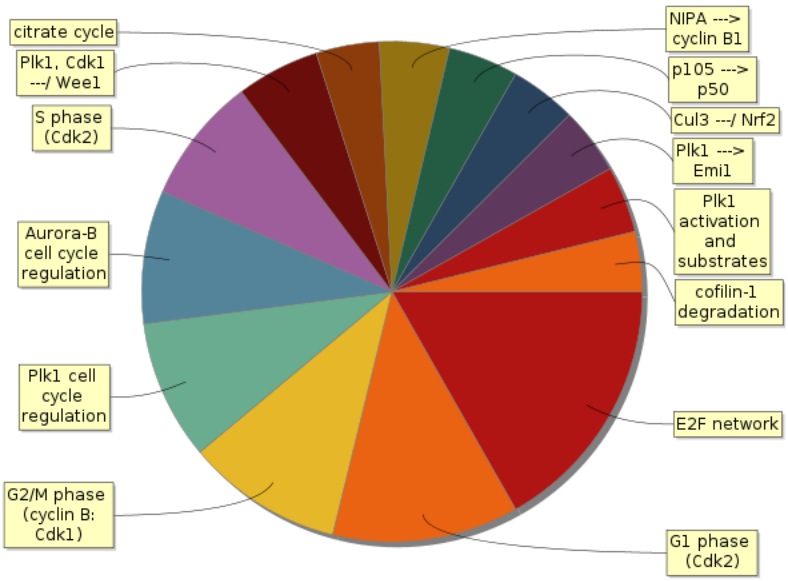


Figure 4. Enriched TRANSPATH® Pathways (2020.1) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.
Full classification →

HumanPSD(TM) disease (2020.1)

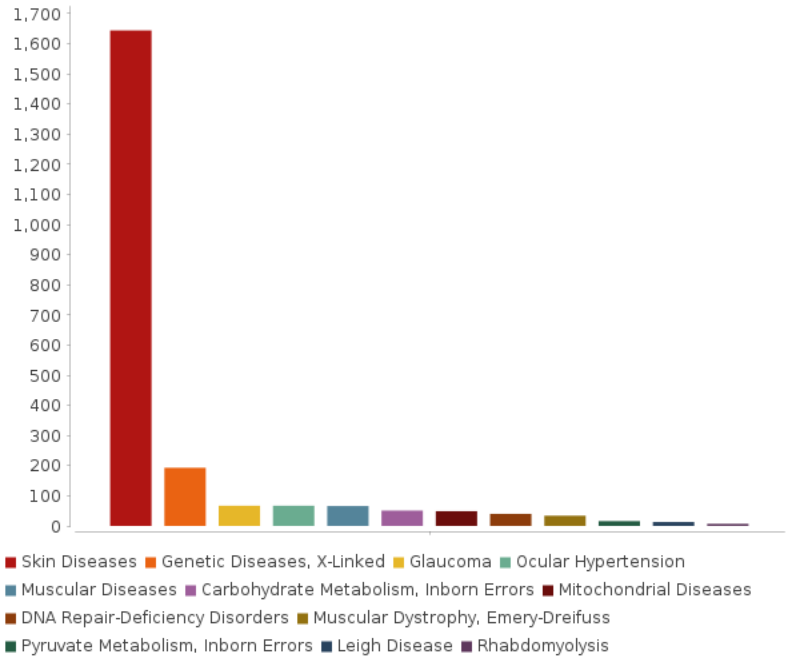


Figure 5. Enriched HumanPSD(TM) disease (2020.1) of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. The size of the bars correspond to the number of bio-markers of the given disease found among the input set.
Full classification →

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

8285 significant down-regulated genes were taken for the mapping.

GO (biological process)

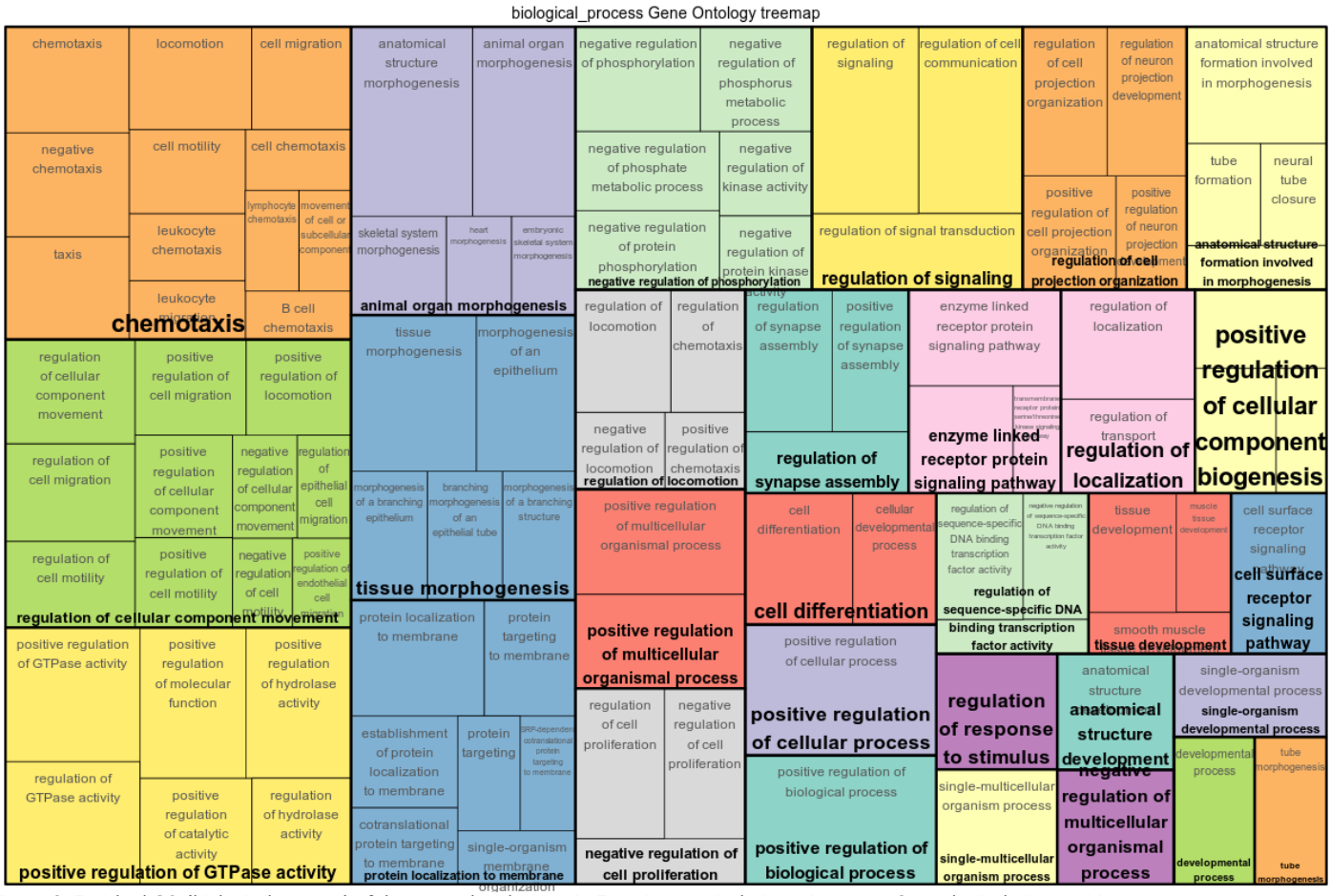


Figure 6. Enriched GO (biological process) of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.
Full classification →

TRANSPATH® Pathways (2020.1)

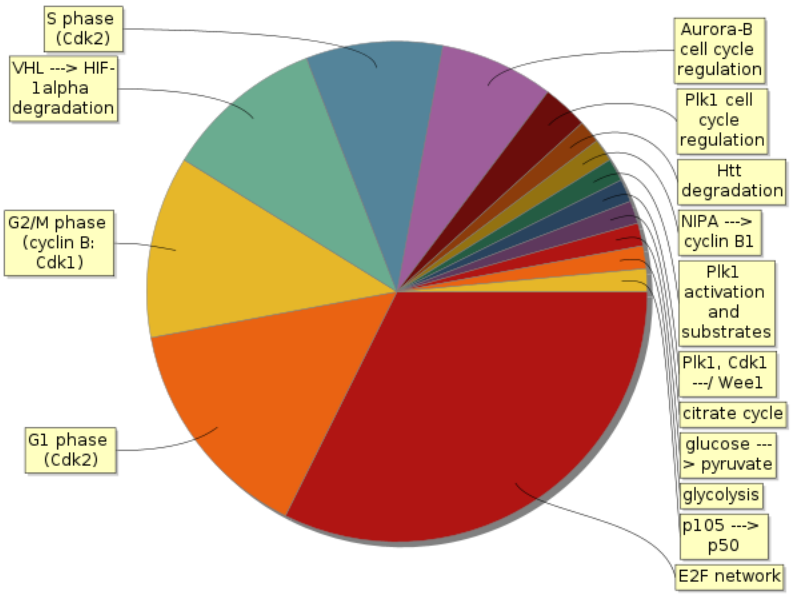


Figure 7. Enriched TRANSPATH® Pathways (2020.1) of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive.
Full classification →

HumanPSD(TM) disease (2020.1)

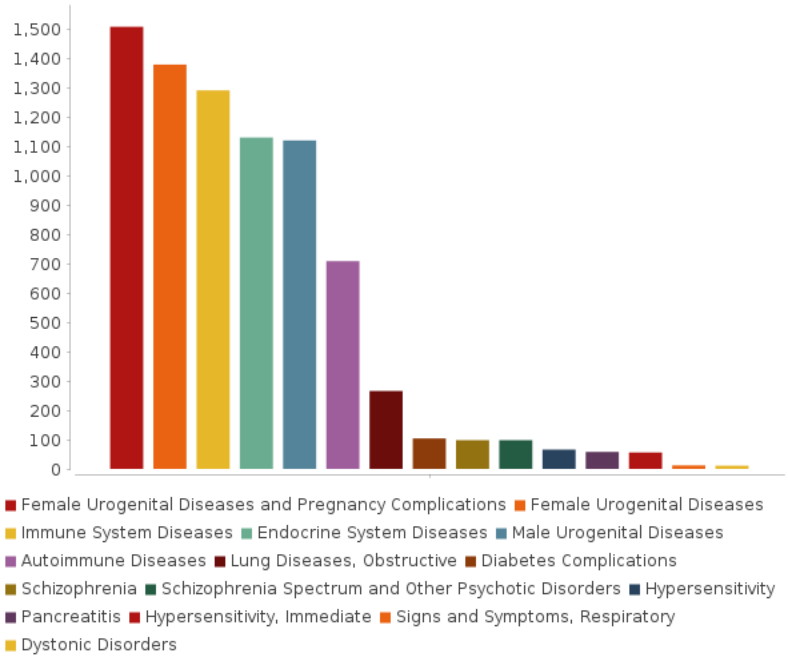


Figure 8. Enriched HumanPSD(TM) disease (2020.1) of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. The size of the bars correspond to the number of bio-markers of the given disease found among the input set.

Full classification →

3.4. Analysis of enriched transcription factor binding sites and composite modules

In the next step a search for transcription factors binding sites (TFBS) was performed in the regulatory regions of the ***target genes*** by using the TF binding motif library of the TRANSFAC® database. We searched for so called **composite-modules** that act as potential condition-specific **enhancers** of the ***target genes*** in their upstream regulatory regions (-1000 bp upstream of transcription start site (TSS)) and identify transcription factors regulating activity of the genes through such **enhancers**.
Classically, **enhancers** are defined as regions in the genome that increase transcription of one or several genes when inserted in either orientation at various distances upstream or downstream of the gene [8]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [9].
In the current work we use the Epigenomics data from the track(s) "GSM385747_CpG_NM.fixed.hg38.top300" to predict positions of potential ***enhancers*** regulating the differentially expressed genes revealed by comparative transcriptomics analysis. We took genomic regions -550bp upstream and 550bp downstream from the middle point of each interval of the track and check if these regions are located inside the 5kb flanking arias of the differentially expressed genes (or inside the body of the genes). In such cases, these genomic regions are used for the search for potential condition-specific enhancers. In all other cases when the differentially expressed genes did not contain epigenomic peaks in their body or in the 5kb flanking regions we used the upstream regulatory regions of these genes (-1000bp upstream and 100bp downstream of TSS) for the search for condition-specific enhancers.

We applied the Composite Module Analyst (CMA) [8] method to detect such potential enhancers, as targets of multiple TFs bound in a cooperative manner to the regulatory regions of the genes of interest. CMA applies a genetic algorithm to construct a generalized model of the enhancers by specifying combinations of TF motifs (from TRANSFAC®) whose sites are most frequently clustered together in the regulatory regions of the studied genes. CMA identifies the transcription factors that through their cooperation provide a synergistic effect and thus have a great influence on the gene regulation process.

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

To build the most specific composite modules we choose top 300 significant up-regulated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all up-regulated genes.

The model consists of 2 module(s). Below, for each module the following information is shown:
- PWMs producing matches,
- number of individual matches for each PWM,
- score of the best match.

Module 1:

V$EGR1_11	V$MZF1_Q5_01	V$FOXM1_01	V$STAT5A_03	V$MYOD_Q6_02	V$SP3_Q3
0.89; N=3	0.98; N=2	0.94; N=2	0.00; N=3	0.99; N=3	0.84; N=1

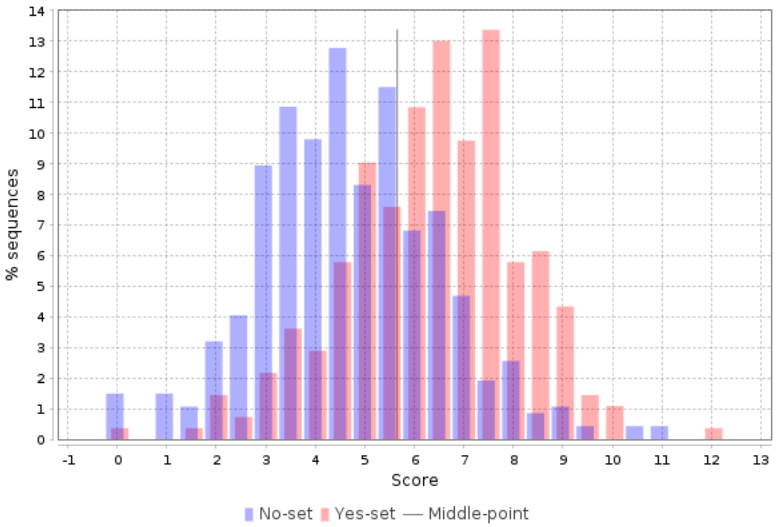
Module width: 148

Module 2:

V$E2F1_09	V$ZNF219_01	V$HLF_01	V$NFY_Q6	V$IRF7_01	V$POU6F1_02
0.96; N=2	0.91; N=2	0.00; N=2	0.00; N=2	0.00; N=3	0.79; N=3

Module width: 139

Model score (-p*log10(pval)): 14.14
Wilcoxon p-value (pval): 1.62e-30
Penalty (p): 0.475
Average yes-set score: 6.34
Average no-set score: 4.73
AUC: 0.75
Middle-point: 5.66
False-positive: 28.09%
False-negative: 31.41%



See model visualization table →

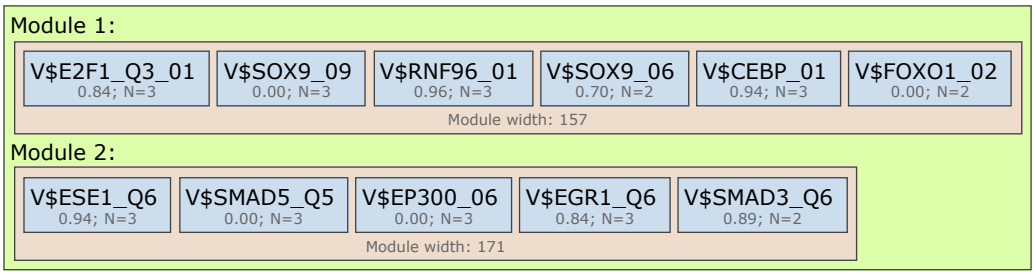
Table 6. List of top ten up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*
See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000168916	ZNF608	zinc finger protein 608	14.09	NF-YA(h),NF-YB(h),NF-YC(h), Egr-1(h), MZF-1(h), zfp219(h), Sp3(h), IRF-7(h), POU6F1(h)...
ENSG00000257261	RP11-96H19.1		14.01	POU6F1(h), Hlf(h), IRF-7(h), NF-YA(h),NF-YB(h),NF-YC(h), STAT5A(h), Sp3(h), Egr-1(h)...
ENSG00000136044	APPL2	adaptor protein, phosphotyrosine interacting with PH domain and leucine zipper 2	13.99	Sp3(h), STAT5A(h), Egr-1(h), MZF-1(h), POU6F1(h), NF-YA(h),NF-YB(h),NF-YC(h), IRF-7(h)...
ENSG00000177868	SVBP	small vasohibin binding protein	13.76	MZF-1(h), Egr-1(h), STAT5A(h), Sp3(h), NF-YA(h),NF-YB(h),NF-YC(h), IRF-7(h), Hlf(h)...
ENSG00000125648	SLC25A23	solute carrier family 25 member 23	13.71	NF-YA(h),NF-YB(h),NF-YC(h), IRF-7(h), Hlf(h), MZF-1(h), foxm1(h), STAT5A(h), Egr-1(h)...
ENSG00000083123	BCKDHB	branched chain keto acid dehydrogenase E1 subunit beta	13.61	foxm1(h), Hlf(h), IRF-7(h), POU6F1(h), NF-YA(h),NF-YB(h),NF-YC(h), STAT5A(h), Egr-1(h)...
ENSG00000170191	NANP	N-acetylneuraminic acid phosphatase	13.59	Egr-1(h), Sp3(h), STAT5A(h), E2F-1(h), POU6F1(h), IRF-7(h), NF-YA(h),NF-YB(h),NF-YC(h)...
ENSG00000184205	TSPYL2	TSPY like 2	13.48	E2F-1(h), Sp3(h), STAT5A(h), MZF-1(h), zfp219(h), POU6F1(h), IRF-7(h)...
ENSG00000146049	KAAG1	kidney associated antigen 1	13.27	foxm1(h), NF-YA(h),NF-YB(h),NF-YC(h), STAT5A(h), E2F-1(h), POU6F1(h), Hlf(h), IRF-7(h)...
ENSG00000157741	UBN2	ubinuclein 2	13.26	IRF-7(h), NF-YA(h),NF-YB(h),NF-YC(h), Hlf(h), POU6F1(h), Sp3(h), STAT5A(h), Egr-1(h)...

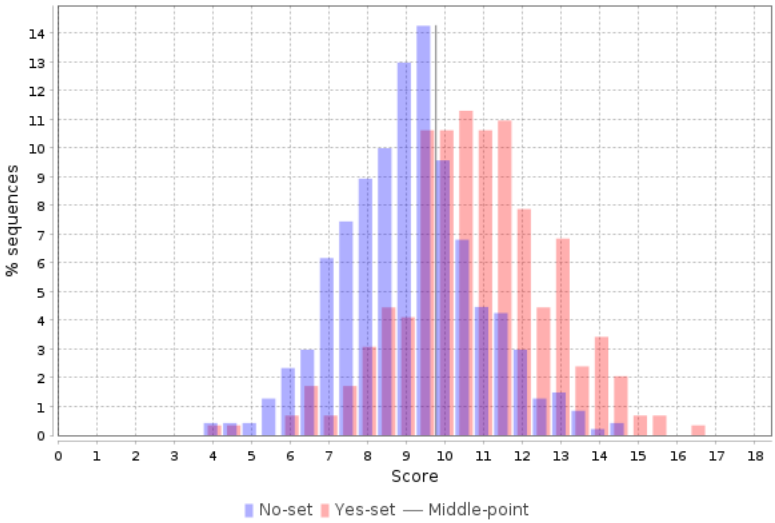
Enhancer model potentially involved in regulation of target genes (down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).

To build the most specific composite modules we choose top 300 significant down-regulated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all down-regulated genes.

The model consists of 2 module(s). Below, for each module the following information is shown:
- PWMs producing matches,
- number of individual matches for each PWM,
- score of the best match.



Model score (-p*log10(pval)): 15.37
Wilcoxon p-value (pval): 2.78e-32
Penalty (p): 0.487
Average yes-set score: 10.80
Average no-set score: 9.10
AUC: 0.75
Middle-point: 9.77
False-positive: 30.64%
False-negative: 27.74%



See model visualization table →

Table 7. 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
ENSG00000176209	SMIM19	small integral membrane protein 19	18.48	C/EBPalpha(h), FOXO1A(h), Sox-9(h), E2F-1(h), Egr-1(h), Smad5(h), p300(h)...
ENSG00000262500	MAPK8IP1P1	mitogen-activated protein kinase 8 interacting protein 1 pseudogene 1	18.1	C/EBPalpha(h), Sox-9(h), FOXO1A(h), E2F-1(h), ESE-1(h), p300(h), Egr-1(h)...
ENSG00000263503	MAPK8IP1P2	mitogen-activated protein kinase 8 interacting protein 1 pseudogene 2	17.77	Smad5(h), Egr-1(h), p300(h), ESE-1(h), E2F-1(h), Sox-9(h), FOXO1A(h)...
ENSG00000171608	PIK3CD	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta	17.63	C/EBPalpha(h), ESE-1(h), Sox-9(h), Egr-1(h), RNF96(h), FOXO1A(h), p300(h)...
ENSG00000165379	LRFN5	leucine rich repeat and fibronectin type III domain containing 5	17.57	p300(h), ESE-1(h), Smad5(h), Egr-1(h), Smad3(h), E2F-1(h), Sox-9(h)...
ENSG00000167526	RPL13	ribosomal protein L13	17.47	C/EBPalpha(h), Sox-9(h), p300(h), RNF96(h), E2F-1(h), FOXO1A(h), Egr-1(h)...
ENSG00000200084	SNORD68	small nucleolar RNA, C/D box 68	17.47	C/EBPalpha(h), Sox-9(h), p300(h), RNF96(h), E2F-1(h), FOXO1A(h), Egr-1(h)...
ENSG00000131981	LGALS3	galectin 3	17.42	Sox-9(h), FOXO1A(h), p300(h), ESE-1(h), C/EBPalpha(h), Egr-1(h), E2F-1(h)...
ENSG00000147533	GOLGA7	golgin A7	17.4	C/EBPalpha(h), Smad3(h), Sox-9(h), FOXO1A(h), p300(h), Smad5(h), ESE-1(h)...
ENSG00000156052	GNAQ	G protein subunit alpha q	17.35	Egr-1(h), RNF96(h), Smad3(h), E2F-1(h), Smad5(h), Sox-9(h), FOXO1A(h)...

On the basis of the enhancer models we identified the following transcription factors potentially regulating the ***target genes*** of our interest. We found 14 and 10 transcription factors controlling expression of up- and down-regulated genes respectively (see Tables 8-9).

*Table 8. 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 TFin 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
MO0000004274	E2F1	E2F transcription factor 1	4.49	1.76
MO0000019612	MYOD1	myogenic differentiation 1	4.45	1.45
MO0000013125	STAT5A	signal transducer and activator of transcription 5A	4.31	3.97
MO0000025939	NFYA	nuclear transcription factor Y subunit alpha	4.21	3.83
MO0000017914	EGR1	early growth response 1	3.89	1.37
MO0000002361	NFYB	nuclear transcription factor Y subunit beta	3.84	3.83
MO0000007703	IRF7	interferon regulatory factor 7	3.79	1.37
MO0000088314	FOXM1	forkhead box M1	3.55	1.38
MO0000028320	null	null	3.17	1.25
MO0000046079	SP3	Sp3 transcription factor	2.61	1.19

*Table 9. 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 TFin 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
MO0000057832	SMAD3	SMAD family member 3	2.57	2.79
MO0000034454	FOXO1	forkhead box O1	2.48	1.36
MO0000019418	CEBPA	CCAAT/enhancer binding protein alpha	2.41	1.8
MO0000056654	EP300	E1A binding protein p300	2.31	1.35
MO0000017914	EGR1	early growth response 1	2.15	1.7
MO0000004274	E2F1	E2F transcription factor 1	2.06	2.47
MO0000054232	ELF3	E74 like ETS transcription factor 3	1.98	1.19
MO0000069886	TRIM28	tripartite motif containing 28	1.86	1.29
MO0000018993	SOX9	SRY-box 9	1.76	1.57
MO0000020635	SMAD5	SMAD family member 5	1.67	2.79

3.5. Finding master regulators in networks

In the second step of the upstream analysis common regulators of the revealed TFs were identified. These master regulators appear to be the key candidates for therapeutic targets as they have a master effect on regulation of intracellular pathways that activate the pathological process of our study. The identified master regulators are shown in Tables 10-11.

*Table 10. 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
MO0000041170	EAC(h)	CYLD	CYLD lysine 63 deubiquitinase	1.05	216
MO0000162677	PHLPP(h)	PHLPP1	PH domain and leucine rich repeat protein phosphatase 1	0.78	233
MO0000010977	PDGFRalpha(h)	PDGFRA	platelet derived growth factor receptor alpha	2.93	255
MO0000129050	EAC-isoform1(h)	CYLD	CYLD lysine 63 deubiquitinase	1.05	279
MO0000038943	cyclinA:Cdk2{pY15}	CCNA1, CCNA2, CDK2	cyclin A1, cyclin A2, cyclin dependent kinase 2	0.94	353
MO0000030895	Chk2(h)	CHEK2	checkpoint kinase 2	0.88	360
MO0000021739	cyclinA2(h)	CCNA2	cyclin A2	0.79	362
MO0000056491	p/CAF(h)	KAT2B	lysine acetyltransferase 2B	0.62	370
MO0000081890	Chk2-isoform1(h)	CHEK2	checkpoint kinase 2	0.88	379
MO0000044269	cyclosome(h):Fzr(h)	ANAPC1, ANAPC10, ANAPC11, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC26, CDC27, FZR1	anaphase promoting complex subunit 1, anaphase promoting complex subunit 10, anaphase promoting comp...	0.47	387

*Table 11. Master regulators that may govern the regulation of **down-regulated** genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics and epigenomics data.*

See full table →

ID	Master molecule name	Gene symbol	Gene description	logFC	Total rank
MO0000016677	ErbB1(h)	EGFR	epidermal growth factor receptor	-1.43	186
MO0000022222	MKP-1(h)	DUSP1	dual specificity phosphatase 1	-1.22	216
MO0000118076	EGF:ErbB1{pY}:ErbB2{pY}:Src	EGF, EGFR, ERBB2, SRC	SRC proto-oncogene, non-receptor tyrosine kinase, epidermal growth factor, epidermal growth factor r...	-1.43	240
MO0000020219	Caspase-8(h)	CASP8	caspase 8	-0.49	316
MO0000083769	MKP-1(h)	DUSP1	dual specificity phosphatase 1	-1.22	329
MO0000101468	LRRK2(h)	LRRK2	leucine rich repeat kinase 2	-1.02	331
MO0000101469	LRRK2(h)	LRRK2	leucine rich repeat kinase 2	-1.02	335
MO0000019070	XIAP(h)	XIAP	X-linked inhibitor of apoptosis	-0.58	393
MO0000031202	Cdc14A(h)	CDC14A	cell division cycle 14A	-0.49	396
MO0000117508	TC-PTP(h)	PTPN2	protein tyrosine phosphatase, non-receptor type 2	-0.75	402

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.

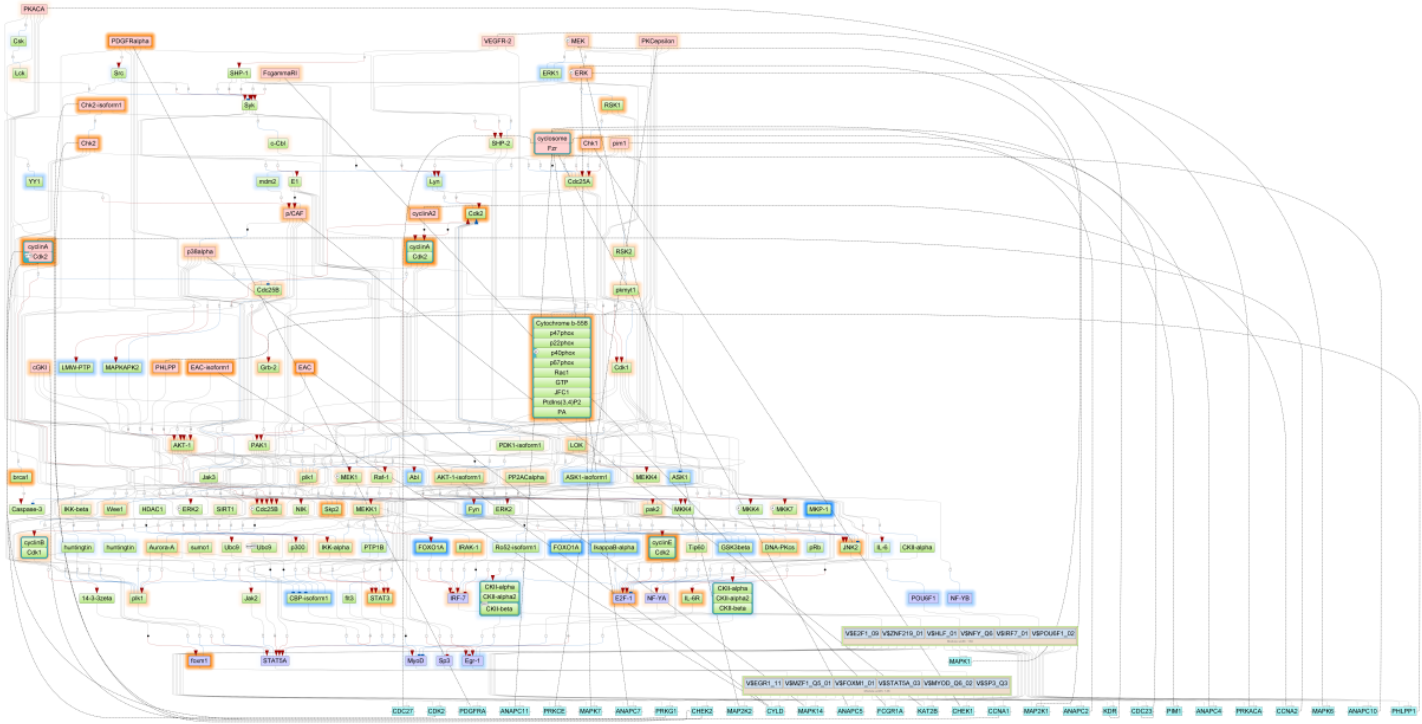


Figure 9. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules,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

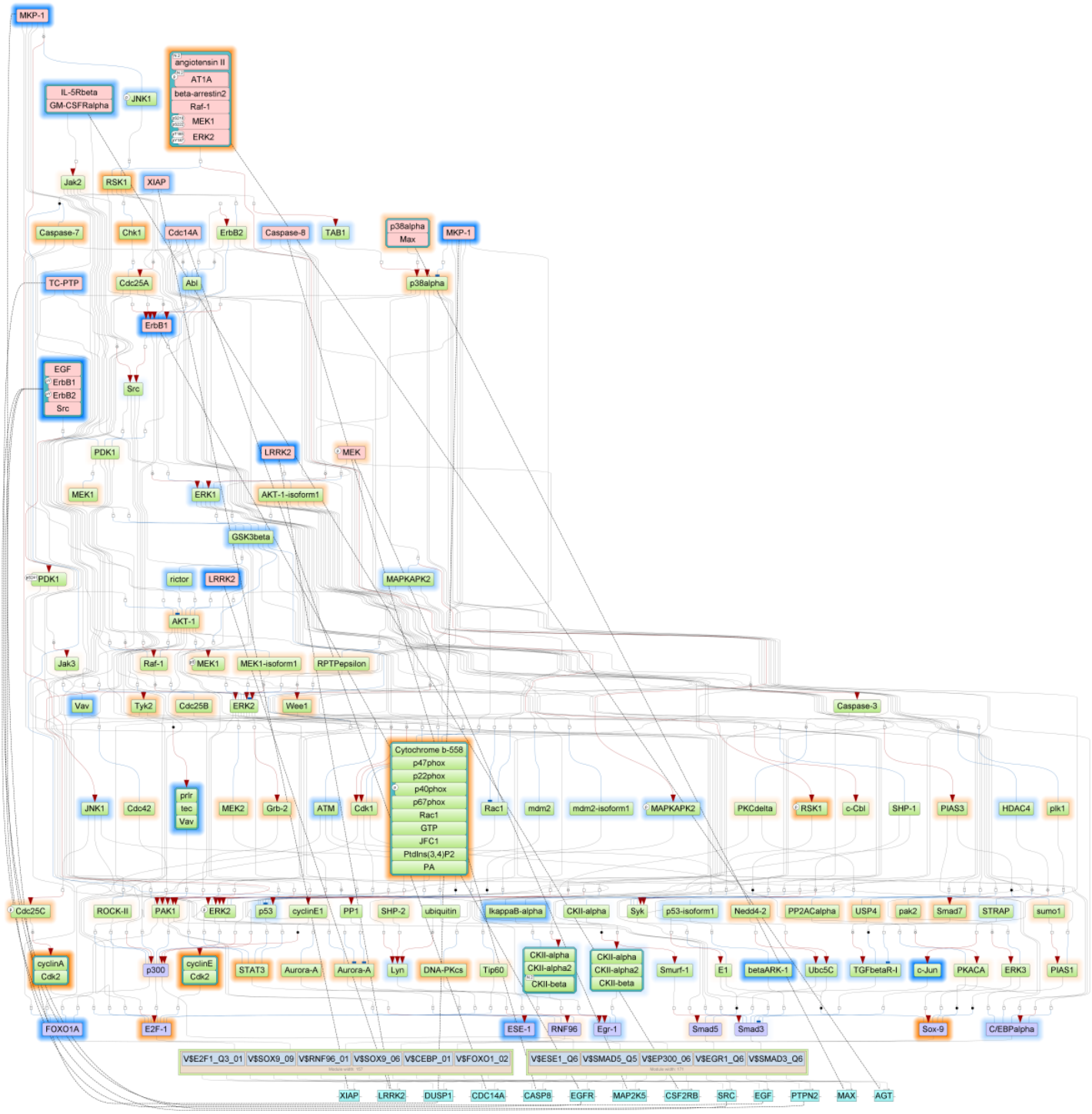


Figure 10. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate 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. Identification of potential drugs

In the last step of the analysis we strived to identify known drugs as well as new potentially active chemical compounds that are potentially suitable for inhibition (or activation) of the identified molecular targets in the context of specified human disease.

First, we identify known drugs using information from HumanPSD™ database [5] about their targets and about clinical trials where the drugs have been tested for the treatment of various human diseases. Table 12 shows the resulting list of druggable master regulators that represent the predicted drug targets of the studied pathology. Table 13 lists chemical compounds and known drugs (from the HumanPSD™ database) potentially acting on corresponding master regulators.

*Table 12. Known drug targets for known drugs revealed in this study.The column **Druggability score** contains the number of drugs that are potentially suitable for inhibition (or activation) of the target. **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	Gene symbol	Gene description	Druggability score	logFC	Total rank
ENSG00000134853	PDGFRA	platelet derived growth factor receptor alpha	8	2.93	255
ENSG00000101182	PSMA7	proteasome subunit alpha 7	3	0.44	446
ENSG00000114166	KAT2B	lysine acetyltransferase 2B	3	0.62	512
ENSG00000177885	GRB2	growth factor receptor bound protein 2	2	0.39	634
ENSG00000186298	PPP1CC	protein phosphatase 1 catalytic subunit gamma	4	0.41	649
ENSG00000163513	TGFBR2	transforming growth factor beta receptor 2	1	3.08	672
ENSG00000150337	FCGR1A	Fc fragment of IgG receptor Ia	21	0.37	729
ENSG00000145386	CCNA2	cyclin A2	33	0.94	743
ENSG00000178999	AURKB	aurora kinase B	3	0.83	754
ENSG00000113575	PPP2CA	protein phosphatase 2 catalytic subunit alpha	3	0.7	808

*Table 13. The list of drugs (from Human PSD) approved or used in clinical trials for the application in ovarian neoplasms and acting on master regulators revealed in our study. The column **Target activity score** contains the value of numeric function that depends on ranks of all targets that were found for the drug. The column **Disease activity score** contains the weighted sum of user selected diseases where the drug is known to be applied. We use sum of clinical trials phases as the weight of the disease. **Drug rank** column contains total rank of given drug among all found. See Methods section for details.*

See full table →

ID	Name	Target names	Target activity score	NA	Phase 1	Phase 2	Phase 3	Phase 4	Dis act sco
DB06589	Pazopanib	ITK, KDR, PDGFRB, PDGFRA	1.25	Ovarian Neoplasms, Adenocarcinoma, Astrocytoma, Bites and Stings, Breast Neoplasms, Carcinoma, Merkel Cell, Carcinoma, Renal Cell...	Ovarian Neoplasms, Brain Neoplasms, Breast Neoplasms, Breast Neoplasms, Male, Carcinoma, Hepatocellular, Carcinoma, Islet Cell, Carcinoma, Non-Small-Cell Lung...	Ovarian Neoplasms, Adenocarcinoma, Adenoma, Oxyphilic, Brain Abscess, Brain Neoplasms, Breast Neoplasms, Breast Neoplasms, Male...	Ovarian Neoplasms, Carcinoma, Non-Small-Cell Lung, Carcinoma, Renal Cell, Chondrosarcoma, Chondrosarcoma, Mesenchymal, Fibrosarcoma, Glomus Tumor...	Carcinoma, Renal Cell, Neoplasms, Noma	7
DB01268	Sunitinib	KDR, PDGFRB, PDGFRA	1.05	Adenocarcinoma, Breast Neoplasms, Carcinoid Tumor, Carcinoma, Hepatocellular, Carcinoma, Islet Cell, Carcinoma, Non-Small-Cell Lung, Carcinoma, Renal Cell...	Acquired Immunodeficiency Syndrome, Adenocarcinoma, Amyloidosis, Anger, Brain Abscess, Brain Neoplasms, Breast Neoplasms...	Ovarian Neoplasms, Acquired Immunodeficiency Syndrome, Adenocarcinoma, Adenocarcinoma, Clear Cell, Adenoma, Islet Cell, Adrenocortical Carcinoma, Ascites...	Brain Neoplasms, Breast Neoplasms, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Islet Cell, Carcinoma, Non-Small-Cell Lung, Carcinoma, Renal Cell...	Carcinoma, Renal Cell, Gastrointestinal Neoplasms, Gastrointestinal Stromal Tumors, Intestinal Neoplasms, Lung Neoplasms, Neoplasms, Neuroendocrine Tumors...	2
DB08896	Regorafenib	KDR, PDGFRB, PDGFRA, RAF1	0.96	Adenocarcinoma, Carcinoma, Hepatocellular, Cholangiocarcinoma, Colorectal Neoplasms, Gastrointestinal Stromal Tumors, Glioblastoma, Liver Neoplasms...	Carcinoma, Hepatocellular, Carcinoma, Small Cell, Colorectal Neoplasms, Esophageal Neoplasms, Gastrointestinal Neoplasms, Gastrointestinal Stromal Tumors, Intestinal Neoplasms...	Ovarian Neoplasms, Adenocarcinoma, Bile Duct Neoplasms, Brain Abscess, Breast Neoplasms, Carcinoid Tumor, Carcinoma, Adenoid Cystic...	Carcinoma, Hepatocellular, Colonic Neoplasms, Colorectal Neoplasms, Esophageal Neoplasms, Gastrointestinal Stromal Tumors, Neoplasms, Noma...	Colorectal Neoplasms, Gastrointestinal Stromal Tumors, Neoplasms, Rectal Neoplasms	2
DB00619	Imatinib	PDGFRB, PDGFRA	0.7	Adenocarcinoma, Central Nervous System Neoplasms, Chordoma, Colonic Neoplasms, Desmoplastic Small Round Cell Tumor, Gastrointestinal Neoplasms, Gastrointestinal Stromal Tumors...	Ovarian Neoplasms, Acquired Immunodeficiency Syndrome, Acute Lung Injury, Amyloidosis, Angiomyoma, Brain Abscess, Brain Neoplasms...	Ovarian Neoplasms, Angiomyoma, Arthritis, Arthritis, Rheumatoid, Asthma, Blast Crisis, Brain Neoplasms...	Astrocytoma, Bone Marrow Diseases, Gastrointestinal Stromal Tumors, Glioblastoma, Graft vs Host Disease, Hematologic Diseases, Hypertension...	Breast Neoplasms, Gastrointestinal Stromal Tumors, Leukemia, Leukemia, Lymphoid, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Mastocytosis...	3
DB06616	Bosutinib	CAMK2G, MAP2K1, CDK2	0.5	Breast Neoplasms, Leukemia, Leukemia, Lymphoid, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid, Neoplasms, Precursor Cell Lymphoblastic Leukemia-Lymphoma	Ovarian Neoplasms, Acute Kidney Injury, Breast Neoplasms, Carcinoma, Non-Small-Cell Lung, Cholangiocarcinoma, Cognitive Dysfunction, Colorectal Neoplasms...	Brain Abscess, Breast Neoplasms, Cholangiocarcinoma, Colorectal Neoplasms, Cysts, Glioblastoma, Kidney Diseases, Cystic...	Leukemia, Leukemia, Myelogenous, Chronic, BCR-ABL Positive, Leukemia, Myeloid	Leukemia, Myeloid	1

*Table 14. The list of drugs (from HumanPSD) known to be acting on master regulators revealed in our study that can be proposed as a drug repurposing initiative for the treatment of ovarian neoplasms. **Target activity score** column contains value of numeric function that depends on ranks of all targets that were found for the drug. **Drug rank** column contains total rank of given drug among all found. See Methods section for details.*

ID	Name	Target names	Target activity score	NA	Phase 1	Phase 2	Phase 3	Phase 4	Dr ra
DB01017	Minocycline	CASP3, CASP1, CYCS	0.59	Acne Vulgaris, Acute Kidney Injury, Alopecia, Angelman Syndrome, Anxiety, Anxiety Disorders, Atrial Fibrillation...	Acne Vulgaris, Acute Kidney Injury, Affect, Alcohol Drinking, Alcoholism, Alopecia, Aneurysm...	Acne Vulgaris, Alcohol Drinking, Alopecia, Alzheimer Disease, Amyotrophic Lateral Sclerosis, Aneurysm, Angelman Syndrome...	Acne Vulgaris, Affect, Alopecia, Amphetamine-Related Disorders, Amyotrophic Lateral Sclerosis, Arthritis, Arthritis, Rheumatoid...	Acne Vulgaris, Affect, Alopecia, Autistic Disorder, Bacterial Infections, Bipolar Disorder, Chronic Periodontitis...	108
DB00163	Vitamin E	PPP2CB, PPP2CA	0.52	Apnea, Fatty Liver, Fatty Liver, Alcoholic, Fragile X Syndrome, HIV Infections, Hepatitis, Infection...	Alzheimer Disease, Anemia, Anemia, Iron-Deficiency, Arthritis, Arthritis, Rheumatoid, Asthma, Brain Abscess...	Acute Kidney Injury, Adrenoleukodystrophy, Anemia, Arsenic Poisoning, Arteriosclerosis, Atherosclerosis, Brain Abscess...	Alzheimer Disease, Angina, Unstable, Arterial Occlusive Diseases, Arteriosclerosis, Burns, Cardiovascular Diseases, Cataract...	Angina Pectoris, Variant, Asphyxia, Cicatrix, Cicatrix, Hypertrophic, Diabetes Mellitus, Dyslipidemias, Epilepsy...	117
DB09033	Vedolizumab	ITGA4	0.47	Cholangitis, Cholangitis, Sclerosing, Colitis, Colitis, Ulcerative, Crohn Disease, Graft vs Host Disease, Inflammatory Bowel Diseases...	Colitis, Colitis, Ulcerative, Crohn Disease, Inflammatory Bowel Diseases, Melanoma, Noma, Ulcer	Celiac Disease, Colitis, Colitis, Ulcerative, Crohn Disease, Ulcer	Cholangitis, Cholangitis, Sclerosing, Colitis, Colitis, Ulcerative, Crohn Disease, Inflammatory Bowel Diseases, Ulcer	Colitis, Colitis, Ulcerative, Crohn Disease, Ulcer	127
DB00108	Natalizumab	ITGA4, FCGR1A	0.45	Brain Abscess, Crohn Disease, Multiple Sclerosis, Multiple Sclerosis, Chronic Progressive, Multiple Sclerosis, Relapsing-Remitting, Paraneoplastic Syndromes, Nervous System	Brain Abscess, Multiple Myeloma, Multiple Sclerosis, Multiple Sclerosis, Chronic Progressive, Multiple Sclerosis, Relapsing-Remitting, Myositis, Myositis, Inclusion Body...	Arthritis, Arthritis, Rheumatoid, Brain Abscess, Crohn Disease, Epilepsies, Partial, Epilepsy, Graft vs Host Disease...	Brain Abscess, Crohn Disease, Multiple Sclerosis, Multiple Sclerosis, Chronic Progressive, Multiple Sclerosis, Relapsing-Remitting	Brain Abscess, Crohn Disease, Multiple Sclerosis, Multiple Sclerosis, Relapsing-Remitting	136
DB00707	Porfimer	FCGR1A	0.44	Carcinoma, Non-Small-Cell Lung, Carcinoma, Small Cell, Cholangiocarcinoma, Head and Neck Neoplasms, Lung Neoplasms, Mesothelioma, Neoplasms...	Adenocarcinoma, Brain Neoplasms, Carcinoma, Acinar Cell, Lung Neoplasms, Neoplasms, Noma, Pancreatic Neoplasms	Bile Duct Neoplasms, Carcinoma, Squamous Cell, Carcinoma, Verrucous, Esophageal Neoplasms, Gallbladder Neoplasms, Head and Neck Neoplasms, Liver Neoplasms...	Bile Duct Neoplasms, Central Nervous System Neoplasms, Cholangiocarcinoma, Gallbladder Neoplasms, Neoplasms, Nervous System Neoplasms, Noma...	Carcinoma, Non-Small-Cell Lung, Carcinoma, Small Cell, Lung Neoplasms, Neoplasms, Small Cell Lung Carcinoma	147

Next, new potential small molecular ligands were predicted for the revealed targets and a general druggability check was run using a pre-computed database of spectra of biological activities of chemical compounds from a library of 13040 most pharmaceutically active known compounds. The spectra of biological activities has been computed using the program PASS [11-13] on the basis of a (Q)SAR approach. Table 15 shows the resulting list of druggable master regulators, which represent the predicted drug targets of the studied pathology. Table 16 lists chemical compounds and known drugs potentially acting on the corresponding master regulators.

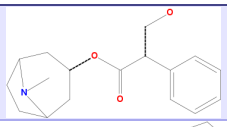
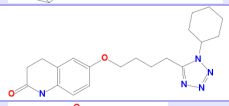
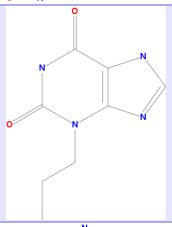
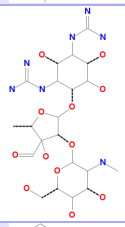
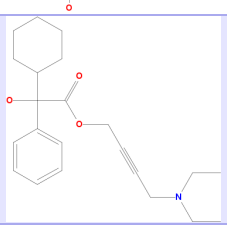
*Table 15. Extended list of drug targets revealed in this study (targets that are predicted by PASS program potentially targeted by an extended list of known drugs and pharmaceutically active chemical compounds). The column **Druggability score** contains a numeric value which indicates how suitable this target is to be inhibited (or activated) by a drug. See Methods section for details.*

See full table →

ID	Name	Gene symbol	Gene description	Druggability score	logFC	Total rank
ENSG00000134853	PDGFRA	PDGFRA	platelet derived growth factor receptor alpha	25.94	2.93	255
ENSG00000143476	DTL	DTL	denticleless E3 ubiquitin protein ligase homolog	17.43	1.12	430
ENSG00000087191	PSMC5	PSMC5	proteasome 26S subunit, ATPase 5	0.85	0.44	446
ENSG00000095261	PSMD5	PSMD5	proteasome 26S subunit, non-ATPase 5	0.85	0.44	446
ENSG00000101182	PSMA7	PSMA7	proteasome subunit alpha 7	1.7	0.44	446
ENSG00000161057	PSMC2	PSMC2	proteasome 26S subunit, ATPase 2	0.85	0.44	446
ENSG00000165916	PSMC3	PSMC3	proteasome 26S subunit, ATPase 3	0.85	0.44	446
ENSG00000020426	MNAT1	MNAT1	MNAT1, CDK activating kinase assembly factor	17.43	0.4	512
ENSG00000114166	KAT2B	KAT2B	lysine acetyltransferase 2B	32.89	0.62	512
ENSG00000134480	CCNH	CCNH	cyclin H	9.81	0.4	512

Table 16. The chemical compounds and known drugs identified by the PASS program as potentially acting on master regulators revealed in our study. Based on the revealed mechanism of action these compounds can be proposed for the treatment of ovarian neoplasms in the current pathological case. ***Disease activity score*** *column contains maximal value of probability to be active for all activities corresponding to the selected diseases for the given compound or 0 if no diseases were selected (in this case column will be hidden).* ***Target activity score*** *column contains value of numeric function which depends on all activity-mechanisms correspondent to the drug.* ***Drug rank*** *column contains total rank of given drug among all found. See Methods section for details.*

See full table →

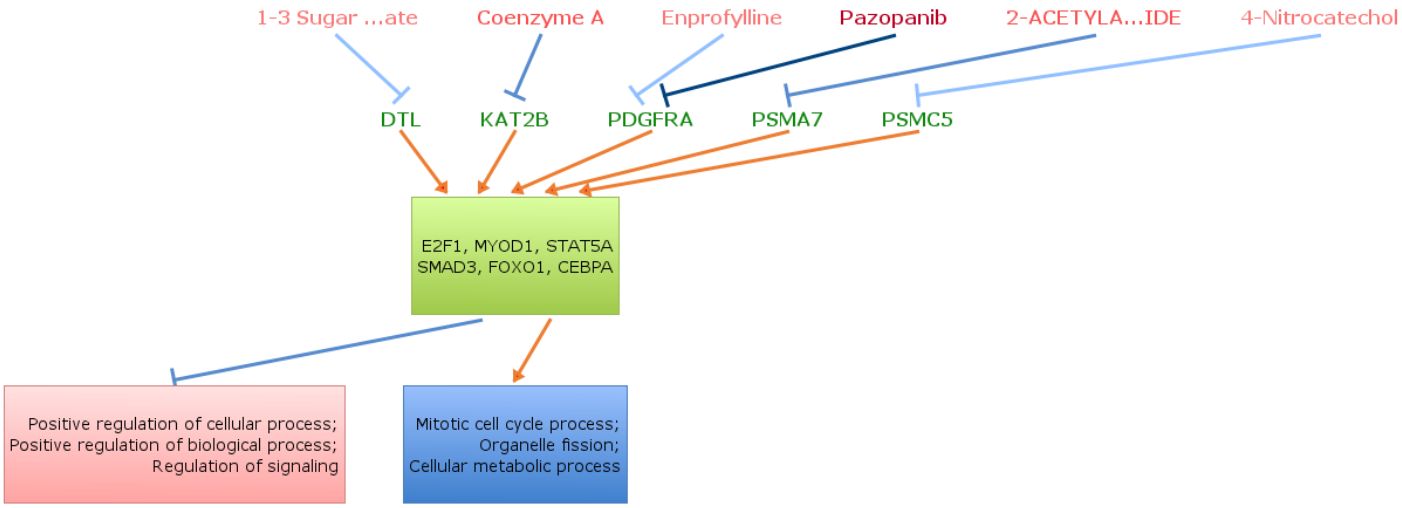
Name	Structure	Target names	Target activity score	Disease activity score	Drug rank
Hyoscyamine		BIRC3, CDK6, CDK9, CCND3, CCNB1, CDK1, CCNB2...	6.58	0	2
Cilostazol		BIRC3, CDK6, CDK9, CCND3, PRKCE, CCNB1, CDK1...	4.37	0	3
Enprofylline		TEC, KDR, JAK3, BLK, EPHA4, PDGFRA, SYK...	4.07	0	4
Streptomycin		BIRC3, CDK6, CDK9, CCND3, PRKCE, CCNB1, CDK1...	3.95	0	5
Oxybutynin		BIRC3, CDK6, CDK9, CCND3, PRKCE, CCNB1, CDK1...	3.93	0	6

As a result of the drug search we came up with two lists of chemical compounds potentially applicable to the targets of our interest. The first list is based on drugs that are known as ligands for the revealed targets in the context of the diseases in our focus as well as in other disease conditions. The second list of identified compounds is based on the prediction of their potential biological activities, which was done using the program PASS. Such computational predictions should be taken as mere suggestions and should be used with care in further experiments.

5. Conclusion

We applied the software package "Genome Enhancer" to a multi-omics data set that contains *transcriptomics and epigenomics* data obtained from *ovary* tissue. 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 schema of how the selected drugs may interfere with the identified target molecules and pathogenic processes discovered by the study reported here.



6. Methods

Databases used in the study

Transcription factor binding sites in promoters and enhancers of differentially expressed genes were analyzed using known DNA-binding motifs described in the TRANSFAC® library, release 2020.1 (geneXplain GmbH, Wolfenbüttel, Germany) (http://genexplain.com/transfac).
The master regulator search uses the TRANSPATH® database (BIOBASE), release 2020.1 (geneXplain GmbH, Wolfenbüttel, Germany) (http://genexplain.com/transpath). A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in TRANSPATH®.
The information about drugs corresponding to identified drug targets and clinical trials references were extracted from HumanPSD™ database, release 2020.1 (http://genexplain.com/humanpsd).
The Ensembl database release Human88.38 (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 sum of three other ranks:

1. ranking by "Target activity score" (*T-score*$_{PSD}$),
2. ranking by "Disease activity score" (*D-score*$_{PSD}$),
3. ranking by clinical trials phase.

To calculate clinical trials phase for the given compound we select the maximum phase of all diseases that are known to have clinical trials with this compound. "Target activity score" (*T-score*$_{PSD}$) is calculated as follows:

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

where *T* is set of all targets related to the compound intersected with input list, |*T*| is number of elements in *T*, *AT* and |*AT*| are set set of all targets related to the compound and number of elements in it, *w* is weight multiplier, *rank(t)* is rank of given target, *maxRank(T)* equals *max(rank(t))* for all targets *t* in *T*.
We use following formula to calculate "Disease activity score" (*D-score*$_{PSD}$):

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

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

Method for prediction of pharmaceutical compounds

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

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

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

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

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

where *M(s)* is the set of activity-mechanisms for the given structure (which passed the chosen threshold for activity-mechanisms *Pa*); *G(m)* is the set of targets (converted to genes) that corresponds to the given activity-mechanism (m) for the given compound; *pa(m)* is the probability to be active of the activity-mechanism (m), *IAP(g)* is the invariant accuracy of prediction for gene from *G(m)*; *optWeight(g)* is the additional weight multiplier for gene. *T* is set of all targets related to the compound intersected with input list, |*T*| is number of elements in *T*, *AT* and |*AT*| are set set of all targets related to the compound and number of elements in it, *w* is weight multiplier.

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

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

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

7. References

1. Kel A, Voss N, Jauregui R, Kel-Margoulis O, Wingender E. Beyond microarrays: Finding key transcription factors controlling signal transduction pathways. *BMC Bioinformatics.* **2006**;7(S2), S13. doi:10.1186/1471-2105-7-s2-s13
2. Stegmaier P, Voss N, Meier T, Kel A, Wingender E, Borlak J. Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer. *PLoS ONE.* **2011**;6(3):e17738. doi:10.1371/journal.pone.0017738
3. Koschmann J, Bhar A, Stegmaier P, Kel A, Wingender E. "Upstream Analysis": An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays.* **2015**;4(2):270-286. doi:10.3390/microarrays4020270.
4. Kel A, Stegmaier P, Valeev T, Koschmann J, Poroikov V, Kel-Margoulis OV, and Wingender E. Multi-omics "upstream analysis" of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer. *EuPA Open Proteom.* **2016**;13:1-13. doi:10.1016/j.euprot.2016.09.002
5. Michael H, Hogan J, Kel A et al. Building a knowledge base for systems pathology. *Brief Bioinformatics.* **2008**;9(6):518-531. doi:10.1093/bib/bbn038
6. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **2006**;34(90001):D108-D110. doi:10.1093/nar/gkj143
7. Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* **2003**;31(13):3576-3579. doi:10.1093/nar/gkg585
8. Waleev T, Shtokalo D, Konovalova T, Voss N, Cheremushkin E, Stegmaier P, Kel-Margoulis O, Wingender E, Kel A. Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res.* **2006**;34(Web Server issue):W541-5.
9. Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.* **2006**;34(90001):D546-D551. doi:10.1093/nar/gkj107
0. Boyarskikh U, Pintus S, Mandrik N, Stelmashenko D, Kiselev I, Evshin I, Sharipov R, Stegmaier P, Kolpakov F, Filipenko M, Kel A. Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3. *BMC Med Genomics.* **2018**;11(1):12. doi:10.1186/1471-2105-7-s2-s13
1. Filimonov D, Poroikov V. Probabilistic Approaches in Activity Prediction. Varnek A, Tropsha A. *Chemoinformatics Approaches to Virtual Screening.* Cambridge (UK): RSC Publishing. **2008**;:182-216.
2. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal.* **2006**;50(2):66-75 (russ)
3. Filimonov D, Poroikov V, Borodina Y, Gloriozova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *ChemInform.* **1999**;39(4):666-670. doi:10.1002/chin.199940210

Thank you for using the Genome Enhancer!

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

Supplementary material

1. Supplementary table 1 - Up-regulated genes
2. Supplementary table 2 - Down-regulated genes
3. Supplementary table 3 - Detailed report. Composite modules and master-regulators (up-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).
4. Supplementary table 4 - Detailed report. Composite modules and master-regulators (down-regulated genes in Experiment: cisplatin-resistant vs. Control: cisplatin-sensitive).
5. Supplementary table 5 - Detailed report. Pharmaceutical compounds and drug targets.

Disclaimer

Decisions regarding care and treatment of patients should be fully made by attending doctors. The predicted chemical compounds listed in the report are given only for doctor's consideration and they cannot be treated as prescribed medication. It is the physician's responsibility to independently decide whether any, none or all of the predicted compounds can be used solely or in combination for patient treatment

purposes, taking into account all applicable information regarding FDA prescribing recommendations for any therapeutic and the patient's condition, including, but not limited to, the patient's and family's medical history, physical examinations, information from various diagnostic tests, and patient preferences in accordance with the current standard of care. Whether or not a particular patient will benefit from a selected therapy is based on many factors and can vary significantly.

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

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

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

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

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