TLR4 and CCND3 are promising druggable targets for treating hepatitis c that control activity of IRF3, IRF7 and E2F1 transcription factors on promoters of differentially expressed genes in liver tissue

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 13/08/2019 ; Run on 17/02/2020 ; Report generated on 17/02/2020

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 data set that contains *transcriptomics* data obtained from *liver* tissue. The study is done in the context of *hepatitis c*. 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: IRF3, IRF7, RUNX3, E2F1, E2F4 and E2F3. The subsequent network analysis suggested TLR4, LY96, IL1R1, CCND3 and TLR3 as the most promising and druggable molecular targets. Finally, the following drugs were identified as the most promising treatment candidates: Anakinra, Naloxone, 3-Hydroxy-Myristic Acid, Uridine, 2,3-Dihydroxy-Benzoic Acid and N-Hexadecanoylglycine.

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
E01_Transcriptomics_LogFC-Table	Transcriptomics

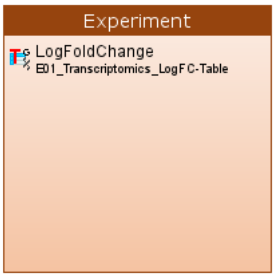


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

3. Results

We have analysed the following condition: Experiment.

3.1. Identification of target genes

In the first step of the analysis ***target genes*** were identified from the uploaded experimental data. Genes were ranked according to the expression value and 300 genes with highest value (see Table 2) and 300 genes with lowest value (see Table 3) were selected for further analysis.

Table 2. Top ten high expressed genes in Experiment.
See full table →

ID	Gene description	Gene symbol	LogFoldChange
ENSG00000137959	interferon induced protein 44 like	IFI44L	6.19
ENSG00000169245	C-X-C motif chemokine ligand 10	CXCL10	6.02
ENSG00000134321	radical S-adenosyl methionine domain containing 2	RSAD2	5.97
ENSG00000137965	interferon induced protein 44	IFI44	3.78
ENSG00000133106	epithelial stromal interaction 1	EPSTI1	3.77
ENSG00000185745	interferon induced protein with tetratricopeptide repeats 1	IFIT1	3.71
ENSG00000187608	ISG15 ubiquitin-like modifier	ISG15	3.63
ENSG00000185201	interferon induced transmembrane protein 2	IFITM2	3.54
ENSG00000185885	interferon induced transmembrane protein 1	IFITM1	3.54
ENSG00000135114	2'-5'-oligoadenylate synthetase like	OASL	3.48

Table 3. Top ten low expressed genes in Experiment.
See full table →

ID	Gene description	Gene symbol	LogFoldChange
ENSG00000167910	cytochrome P450 family 7 subfamily A member 1	CYP7A1	-1.09
ENSG00000169282	potassium voltage-gated channel subfamily A member regulatory beta subunit 1	KCNAB1	-1.04
ENSG00000171560	fibrinogen alpha chain	FGA	-0.98
ENSG00000152133	G-patch domain containing 11	GPATCH11	-0.96
ENSG00000182372	ceroid-lipofuscinosis, neuronal 8	CLN8	-0.91
ENSG00000130649	cytochrome P450 family 2 subfamily E member 1	CYP2E1	-0.88
ENSG00000253327	RAD21 antisense RNA 1	RAD21-AS1	-0.88
ENSG00000170323	fatty acid binding protein 4	FABP4	-0.87
ENSG00000175390	eukaryotic translation initiation factor 3 subunit F	EIF3F	-0.86
ENSG00000261609	gigaxonin	GAN	-0.8

3.2. Functional classification of genes

A functional analysis of differentially expressed genes was done by mapping the top high expressed and top low expressed 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 2-7 show the most significant categories.

High expressed genes in Experiment:

300 top high expressed genes were taken for the mapping.

GO (biological process)

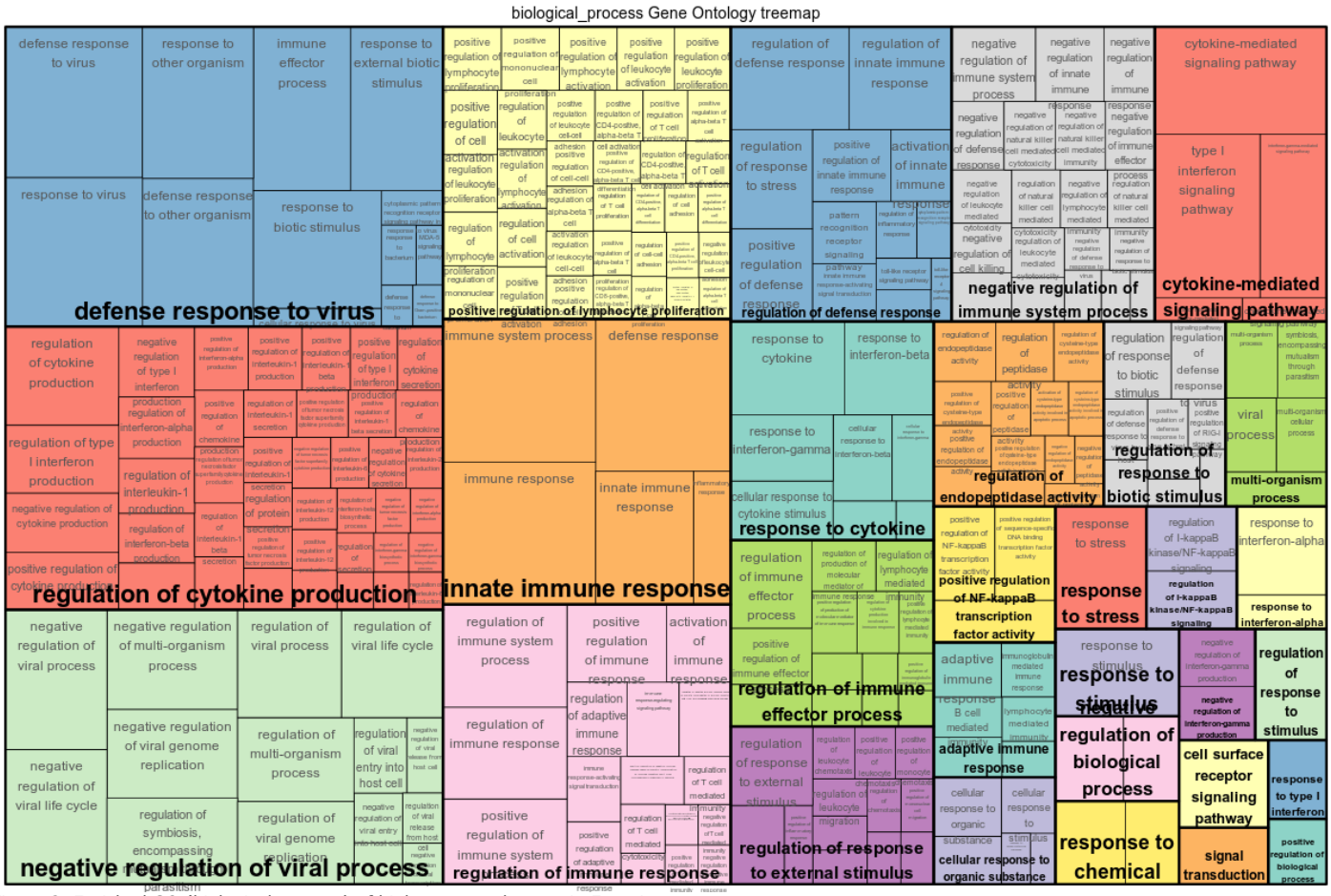


Figure 2. Enriched GO (biological process) of high expressed genes in Experiment.
Full classification →

TRANSPATH® Pathways (2020.1)

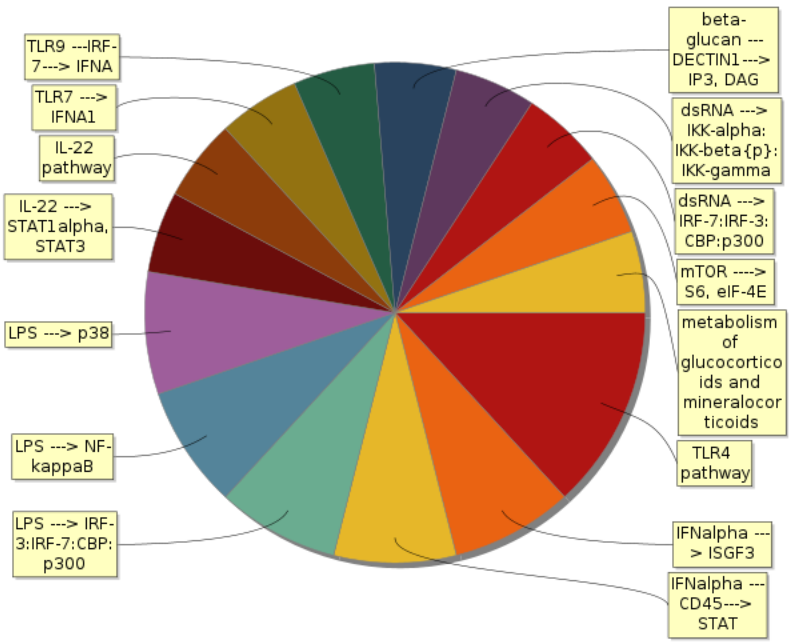


Figure 3. Enriched TRANSPATH® Pathways (2020.1) of high expressed genes in Experiment.
Full classification →

HumanPSD(TM) disease (2020.1)

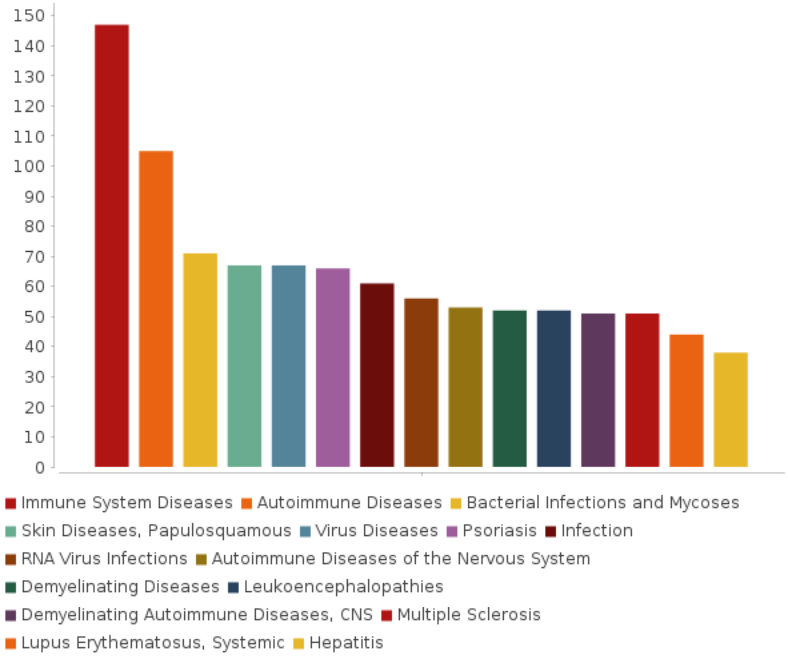


Figure 4. Enriched HumanPSD(TM) disease (2020.1) of high expressed genes in Experiment. The size of the bars correspond to the number of bio-markers of the given disease found among the input set.

Full classification →

Low expressed genes in Experiment:

300 top low expressed genes were taken for the mapping.

GO (biological process)

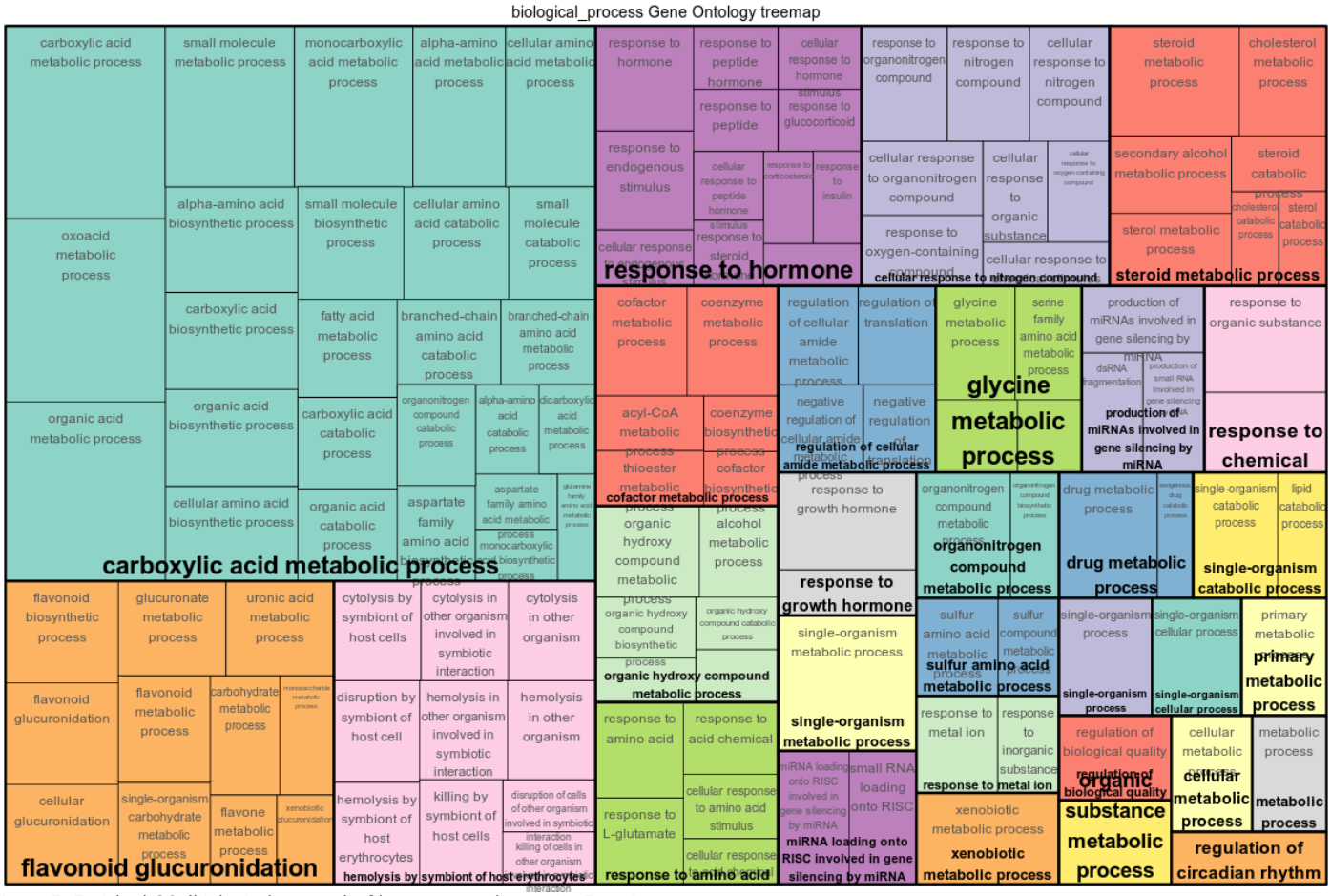


Figure 5. Enriched GO (biological process) of low expressed genes in Experiment.

Full classification →

TRANSPATH® Pathways (2020.1)

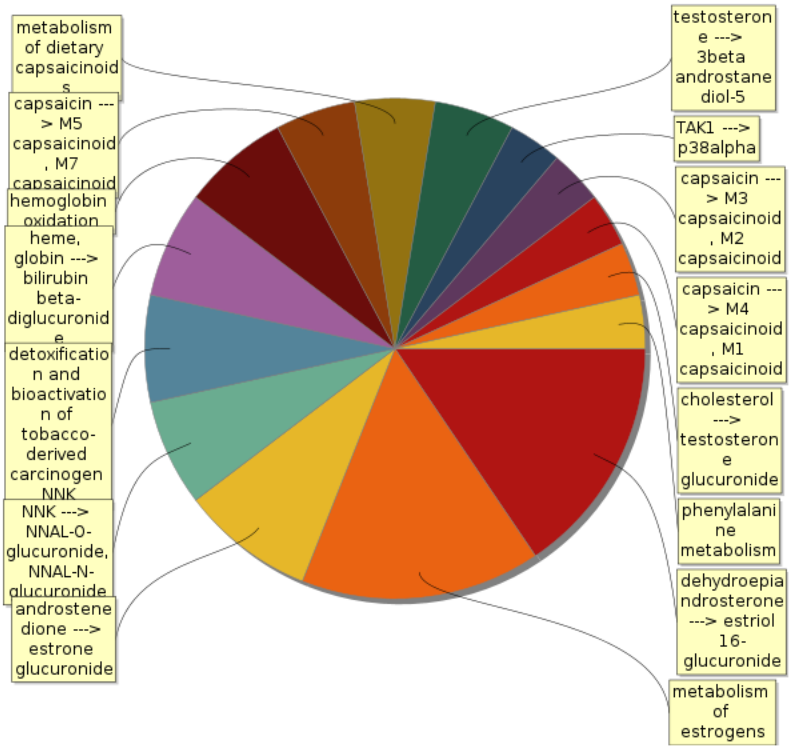


Figure 6. Enriched TRANSPATH® Pathways (2020.1) of low expressed genes in Experiment.
Full classification →

HumanPSD(TM) disease (2020.1)

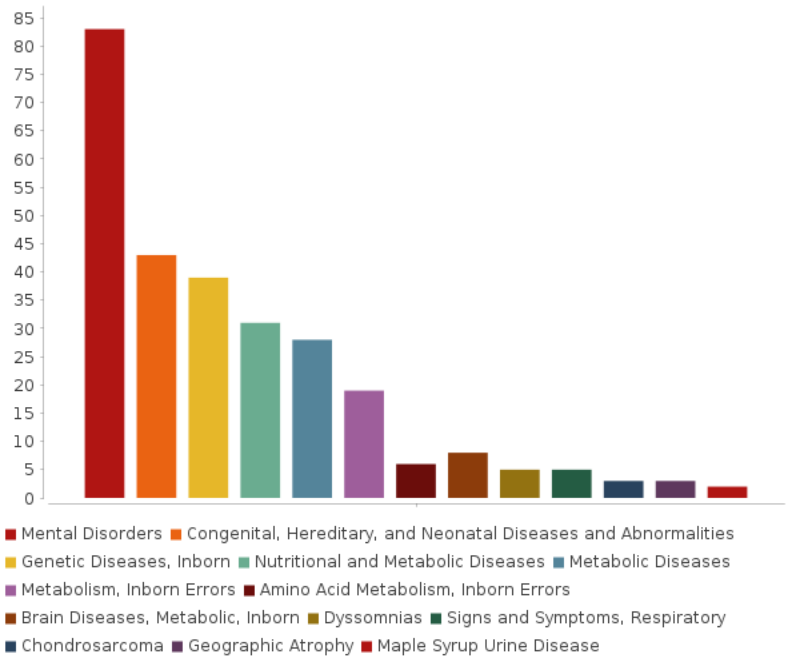


Figure 7. Enriched HumanPSD(TM) disease (2020.1) of low expressed genes in Experiment. The size of the bars correspond to the number of bio-markers of the given disease found among the input set.
Full classification →

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

In the next step a search for transcription factors binding sites (TFBS) was performed in the regulatory regions of the ***target genes*** by using the TF binding motif library of the TRANSFAC® database. We searched for so called **composite-modules** that act as potential condition-specific **enhancers** of the ***target genes*** in their upstream regulatory regions (-1000 bp upstream of transcription start site (TSS)) and identify transcription factors regulating activity of the genes through such **enhancers**.
Classically, **enhancers** are defined as regions in the genome that increase transcription of one or several genes when inserted in either orientation at various distances upstream or downstream of the gene [8]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [9].

We applied the Composite Module Analyst (CMA) [8] method to detect such potential enhancers, as targets of multiple TFs bound in a cooperative manner to the regulatory regions of the genes of interest. CMA applies a genetic algorithm to construct a generalized model of the enhancers by specifying combinations of TF motifs (from TRANSFAC®) whose sites are most frequently clustered together in the regulatory

regions of the studied genes. CMA identifies the transcription factors that through their cooperation provide a synergistic effect and thus have a great influence on the gene regulation process.

Enhancer model potentially involved in regulation of target genes (high expressed genes in Experiment).

To build the most specific composite modules we choose top high expressed genes as the input of CMA algorithm.

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$IRF3_07	V$IRF_Q6	V$BCL11A_02	V$IRF9_01
0.74; N=1	0.00; N=2	0.93; N=3	0.77; N=3

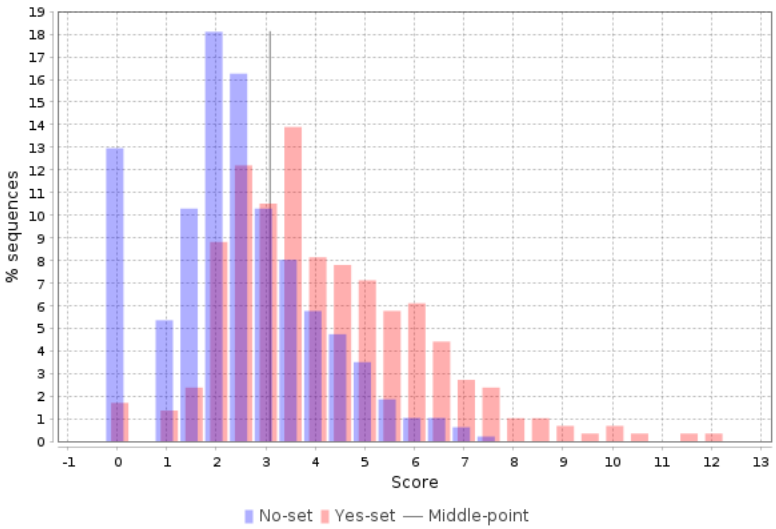
Module width: 140

Module 2:

V$IRF3_07	V$AML2_Q3_01	V$SMAD4_Q4	V$DR1_Q3
0.74; N=2	0.90; N=2	0.00; N=2	0.00; N=2

Module width: 85

Model score (-p*log10(pval)): 17.20
Wilcoxon p-value (pval): 8.15e-33
Penalty (p): 0.536
Average yes-set score: 4.15
Average no-set score: 2.46
AUC: 0.75
Middle-point: 3.09
False-positive: 29.42%
False-negative: 32.20%



See model visualization table →

Table 4. List of top ten high expressed genes in Experiment with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

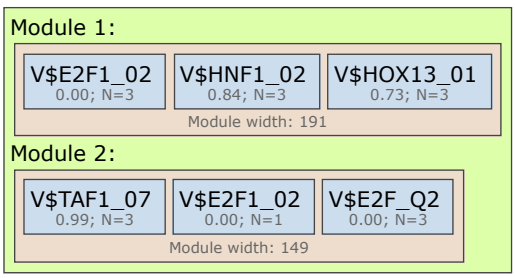
See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000130303	BST2	bone marrow stromal cell antigen 2	12.24	IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), IRF-3(h), IRF-9(h), BCL-11A(h), AML2(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), Smad4(h)
ENSG00000136514	RTP4	receptor transporter protein 4	11.97	AML2(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), Smad4(h), IRF-3(h), IRF-9(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), BCL-11A(h)
ENSG00000152778	IFIT5	interferon induced protein with tetratricopeptide repeats 5	11.4	IRF-3(h), Smad4(h), AML2(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), BCL-11A(h), IRF-9(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h)
ENSG00000213186	TRIM59	tripartite motif containing 59	10.42	IRF-3(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), IRF-9(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), AML2(h), Smad4(h), BCL-11A(h)
ENSG00000187608	ISG15	ISG15 ubiquitin-like modifier	10.15	BCL-11A(h), IRF-9(h), IRF-3(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), AML2(h)
ENSG00000141971	MVB12A	multivesicular body subunit 12A	10.08	COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), IRF-3(h), AML2(h), Smad4(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), IRF-9(h), BCL-11A(h)
ENSG00000176531	PHLDB3	pleckstrin homology like domain family B member 3	10.02	Smad4(h), AML2(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), IRF-3(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), BCL-11A(h)
ENSG00000079385	CEACAM1	carcinoembryonic antigen related cell adhesion molecule 1	9.91	COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), IRF-3(h), IRF-9(h), AML2(h), Smad4(h)
ENSG00000178685	PARP10	poly(ADP-ribose) polymerase family member 10	9.88	IRF-9(h), IRF-3(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), BCL-11A(h), AML2(h), Smad4(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h)
ENSG00000040633	PHF23	PHD finger protein 23	9.86	Smad4(h), AML2(h), COUP-TF1(h),COUP-TF2(h),HNF-4alpha(h),HNF-4gamma(h), IRF-3(h), IRF-1(h),IRF-2(h),IRF-3(h),IRF-4(h),IRF-5(h),IRF-6(h),IRF-7(h),IRF-8(h), IRF-9(h)

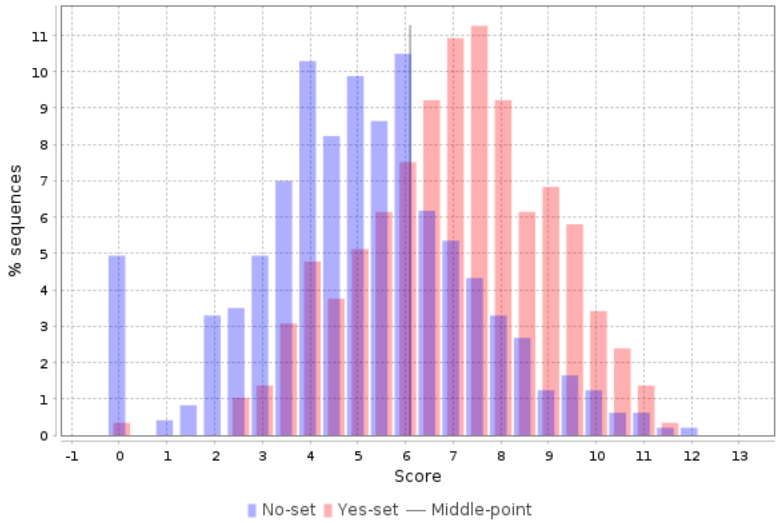
Enhancer model potentially involved in regulation of target genes (low expressed genes in Experiment).

To build the most specific composite modules we choose top low expressed genes as the input of CMA algorithm.

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)): 16.78
Wilcoxon p-value (pval): 1.87e-29
Penalty (p): 0.584
Average yes-set score: 6.99
Average no-set score: 5.11
AUC: 0.74
Middle-point: 6.10
False-positive: 29.63%
False-negative: 29.69%



See model visualization table →

Table 5. List of top ten low expressed genes in Experiment 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
ENSG00000168724	DNAJC21	DnaJ heat shock protein family (Hsp40) member C21	13.21	HoxA5(h), E2F-1(h), HNF-1alpha(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), TAFII250(h)
ENSG00000099810	MTAP	methylthioadenosine phosphorylase	13.13	DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), TAFII250(h), E2F-1(h), HoxA5(h), HNF-1alpha(h)
ENSG00000113790	EHHADH	enoyl-CoA hydratase and 3-hydroxyacyl CoA dehydrogenase	12.86	E2F-1(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), TAFII250(h), HoxA5(h), HNF-1alpha(h)
ENSG00000148730	EIF4EBP2	eukaryotic translation initiation factor 4E binding protein 2	12.86	HoxA5(h), E2F-1(h), HNF-1alpha(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), TAFII250(h)
ENSG00000178234	GALNT11	polypeptide N-acetylgalactosaminyltransferase 11	12.84	HNF-1alpha(h), HoxA5(h), E2F-1(h), TAFII250(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h)
ENSG00000112305	SMAP1	small ArfGAP 1	12.78	HoxA5(h), HNF-1alpha(h), E2F-1(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), TAFII250(h)
ENSG00000112624	GLTSCR1L	GLTSCR1 like	12.76	HoxA5(h), E2F-1(h), HNF-1alpha(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), TAFII250(h)
ENSG00000143753	DEGS1	delta 4-desaturase, sphingolipid 1	12.66	HNF-1alpha(h), HoxA5(h), E2F-1(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), TAFII250(h)
ENSG00000198947	DMD	dystrophin	12.59	TAFII250(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h), E2F-1(h), HNF-1alpha(h), HoxA5(h)
ENSG00000155903	RASA2	RAS p21 protein activator 2	12.55	HoxA5(h), HNF-1alpha(h), E2F-1(h), TAFII250(h), DP-1(h),E2F-1(h),E2F-3(h),E2F-4(h)

On the basis of the enhancer models we identified the following transcription factors potentially regulating the ***target genes*** of our interest. We found 16 and 7 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 (high expressed genes in Experiment). ***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
MO000285816	IRF3	interferon regulatory factor 3	5.19	5.87
MO000007703	IRF7	interferon regulatory factor 7	4.79	11.31
MO000026238	RUNX3	runt related transcription factor 3	4.54	1.11
MO000020402	SMAD4	SMAD family member 4	4.51	1.63
MO000027755	HNF4A	hepatocyte nuclear factor 4 alpha	4.48	1.84
MO000024736	NR2F1	nuclear receptor subfamily 2 group F member 1	3.97	1.5
MO000007686	IRF1	interferon regulatory factor 1	3.62	25.49
MO000007691	IRF2	interferon regulatory factor 2	3.14	5.87
MO000023424	IRF8	interferon regulatory factor 8	2.93	5.87
MO000007759	IRF9	interferon regulatory factor 9	2.91	5.87

Table 7. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (low expressed genes in Experiment). ***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
MO000004274	E2F1	E2F transcription factor 1	3.56	1.65
MO000023603	E2F4	E2F transcription factor 4	3.05	1.42
MO000044809	E2F3	E2F transcription factor 3	2.84	1.42
MO000013458	TFDP1	transcription factor Dp-1	2.52	1.42
MO000081793	TAF1	TATA-box binding protein associated factor 1	1.71	1.55
MO000082618	HNF1A	HNF1 homeobox A	1.69	2.02
MO000120562	HOXA5	homeobox A5	0	1.29

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 high expressed genes in Experiment. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics data.*

See full table →

ID	Master molecule name	Gene symbol	Gene description	Total rank	LogFoldChange
MO0000329204	Cdk6(h):cyclinD3-isoform1(h)	CCND3, CDK6	cyclin D3, cyclin dependent kinase 6	102	0.79
MO0000019446	Caspase-1(h)	CASP1	caspase 1	111	0.87
MO0000041437	dsRNA:TLR3:TRIF	TICAM1, TLR3	toll like receptor 3, toll like receptor adaptor molecule 1	119	0.75
MO0000038322	LPS:lbp:CD14:TLR4:MD-2:MyD88:IRAK-1{pS376}{pT387}	CD14, IRAK1, LBP, LY96, MYD88, TLR4	CD14 molecule, interleukin 1 receptor associated kinase 1, lipopolysaccharide binding protein, lymph...	121	0.62
MO0000038316	LPS:lbp:CD14:TLR4:MD-2:TIRAP:IRAK-2	CD14, IRAK2, LBP, LY96, TIRAP, TLR4	CD14 molecule, TIR domain containing adaptor protein, interleukin 1 receptor associated kinase 2, li...	123	0.61
MO0000019259	c-Cbl(h)	CBL	Cbl proto-oncogene	127	0.37
MO0000145323	uba7(h)	UBA7	ubiquitin like modifier activating enzyme 7	140	1.09
MO0000007703	IRF-7(h)	IRF7	interferon regulatory factor 7	147	2.06
MO0000079043	PML-4(h)	PML	promyelocytic leukemia	149	1.35
MO0000162702	phlpp2(h)	PHLPP2	PH domain and leucine rich repeat protein phosphatase 2	149	0.49

*Table 9. Master regulators that may govern the regulation of low expressed genes in Experiment. **Total rank** is the sum of the ranks of the master molecules sorted by keynode score, CMA score, transcriptomics data.*

See full table →

ID	Master molecule name	Gene symbol	Gene description	Total rank	LogFoldChange
MO0000030927	DNA-PKcs(h)	PRKDC	protein kinase, DNA-activated, catalytic polypeptide	110	-0.52
MO0000043414	cyclosome(h)	ANAPC1, ANAPC10, ANAPC11, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC26, CDC27	anaphase promoting complex subunit 1, anaphase promoting complex subunit 10, anaphase promoting comp...	137	-0.39
MO0000038235	itch(h)	ITCH	itchy E3 ubiquitin protein ligase	152	-0.74
MO0000031205	Cdc14B(h)	CDC14B	cell division cycle 14B	159	-0.44
MO0000018003	PP2A(h)	PPP2CA, PPP2R3A, PPP2R3B, PPP2R5A, PPP2R5B, PPP2R5C, PPP2R5D	protein phosphatase 2 catalytic subunit alpha, protein phosphatase 2 regulatory subunit B''alpha, pr...	162	-0.29
MO0000104136	cyclosome(h):Fzr1(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...	168	-0.39
MO0000082690	Itch-isoform2(h)	ITCH	itchy E3 ubiquitin protein ligase	172	-0.74
MO0000329204	Cdk6(h):cyclinD3-isoform1(h)	CCND3, CDK6	cyclin D3, cyclin dependent kinase 6	182	-0.34
MO0000022393	PKCalpha(h)	PRKCA	protein kinase C alpha	193	-0.38
MO0000089301	TBLR1(h)	TBL1XR1	transducin beta like 1 X-linked receptor 1	193	-0.55

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figures 8 and 9. 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.

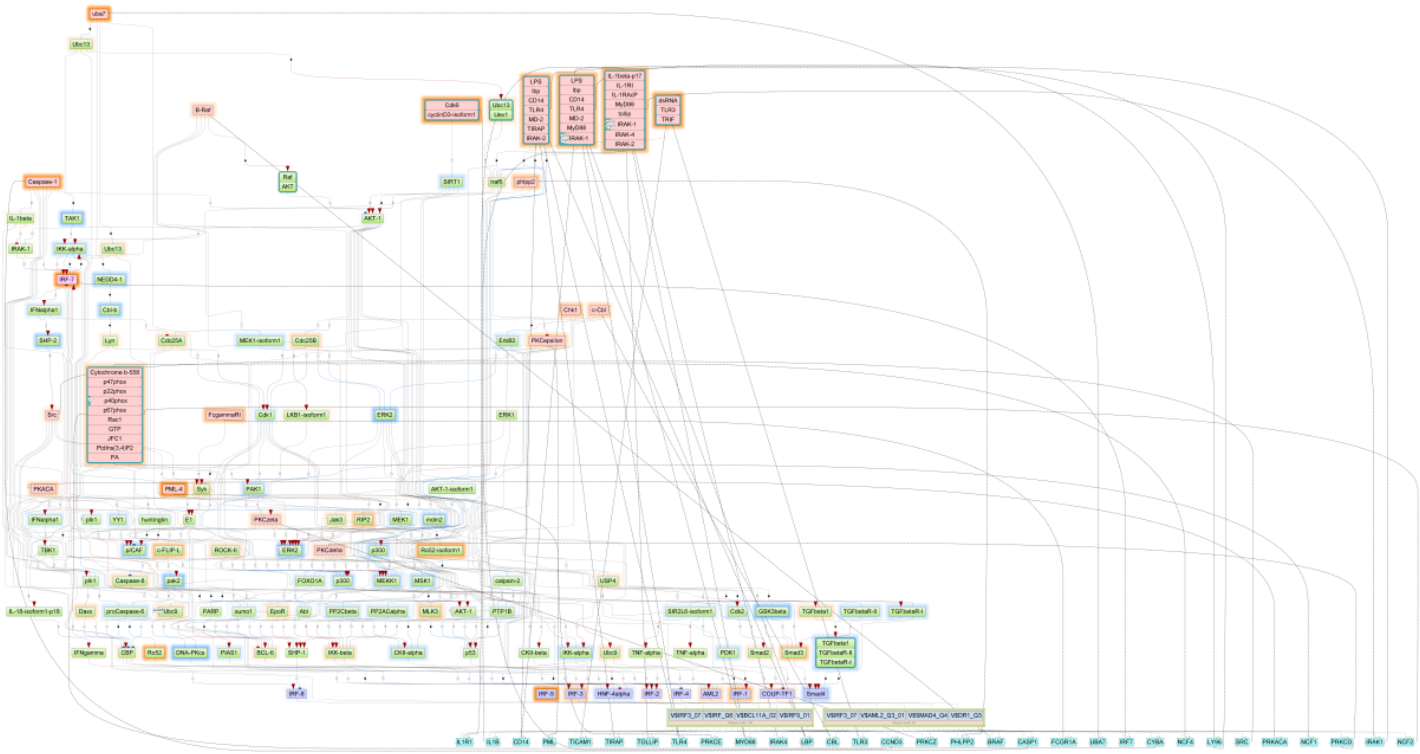


Figure 8. Diagram of intracellular regulatory signal transduction pathways of high expressed genes in Experiment. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during 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 →

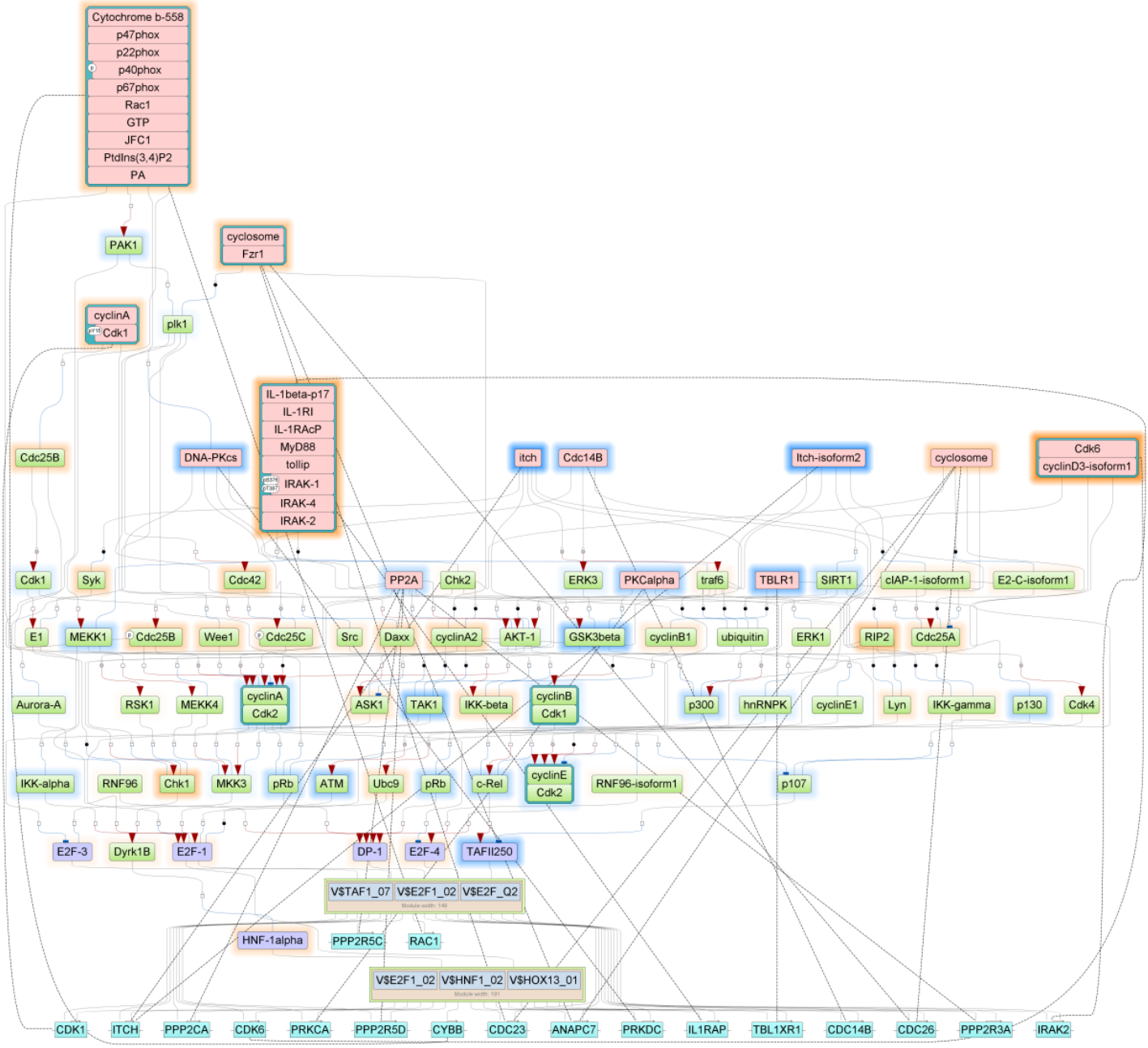


Figure 9. Diagram of intracellular regulatory signal transduction pathways of low expressed genes in Experiment. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during the 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 10 shows the resulting list of druggable master regulators that represent the predicted drug targets of the studied pathology. Table 11 lists chemical compounds and known drugs (from the HumanPSD™ database) potentially acting on corresponding master regulators.

*Table 10. 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 data.*

See full table →

ID	Gene symbol	Gene description	Druggability score	Total rank	LogFoldChange
ENSG00000136869	TLR4	toll like receptor 4	5	123	0.62
ENSG00000154589	LY96	lymphocyte antigen 96	2	123	0.62
ENSG00000115594	IL1R1	interleukin 1 receptor type 1	3	180	0.62
ENSG00000125538	IL1B	interleukin 1 beta	13	180	0.62
ENSG00000198001	IRAK4	interleukin 1 receptor associated kinase 4	1	180	0.62
ENSG00000162889	MAPKAPK2	mitogen-activated protein kinase-activated protein kinase 2	7	184	0.31
ENSG00000165731	RET	ret proto-oncogene	7	231	0.58
ENSG00000137752	CASP1	caspase 1	8	259	0.87
ENSG00000096968	JAK2	Janus kinase 2	8	260	0.54
ENSG00000101182	PSMA7	proteasome subunit alpha 7	3	276	0.2

*Table 11. The list of drugs (from Human PSD) approved or used in clinical trials for the application in hepatitis c 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	Disease activity score	D ra
DB00398	Sorafenib	BRAF, RET	0.46	Adenocarcinoma, Ascites, Brain Abscess, Brain Neoplasms, Breast Neoplasms, Carcinoma, Hepatocellular, Carcinoma, Non-Small-Cell Lung...	Hepatitis C, Adenocarcinoma, Adenoma, Adenoma, Liver Cell, Astrocytoma, Bile Duct Neoplasms, Biliary Tract Neoplasms...	Adenocarcinoma, Adenoma, Adenoma, Liver Cell, Adrenocortical Carcinoma, Bile Duct Neoplasms, Biliary Tract Neoplasms, Brain Abscess...	Adenocarcinoma, Breast Neoplasms, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Non-Small-Cell Lung, Carcinoma, Renal Cell, Digestive System Diseases...	Carcinoma, Hepatocellular, Carcinoma, Renal Cell, Liver Neoplasms, Neoplasms, Noma, Thrombosis	1	69
DB01183	Naloxone	TLR4	0.35	Hepatitis C, Acquired Immunodeficiency Syndrome, Arthritis, Burning Mouth Syndrome, Burns, Constipation, Cystitis...	Hepatitis C, Diabetes Mellitus, Type 1, Hepatitis, Heroin Dependence, Hyperalgesia, Hypoglycemia, Hypogonadism...	Binge-Eating Disorder, Brain Death, Bulimia, Constipation, Cystitis, Cystitis, Interstitial, Cysts...	Arthritis, Binge-Eating Disorder, Brain Death, Bulimia, Constipation, Epilepsy, Feeding and Eating Disorders...	Hepatitis C, Angina Pectoris, Angina, Unstable, Arthritis, Bursitis, Constipation, Cysts...	6	85
DB05408	IDN-6556	CASP7, CASP1	0.78		Diabetes Mellitus, Digestive System Diseases, Kidney Diseases, Liver Diseases, Renal Insufficiency	Hepatitis C, Carcinoma, Hepatocellular, Cholestasis, Diabetes Mellitus, Fatty Liver, Fatty Liver, Alcoholic, Fibrosis...			2	11
DB05475	SCV-07	TLR4	0.46	Hepatitis C, Hepatitis, Hepatitis C, Chronic, Hepatitis, Chronic		Hepatitis C, Head and Neck Neoplasms, Hepatitis, Hepatitis C, Chronic, Hepatitis, Chronic, Mucositis, Neoplasms...			3	15
DB00008	Peginterferon alfa-2a	IFNAR2	0.18	Hepatitis C, Hepatitis, Hepatitis B, Hepatitis B, Chronic, Hepatitis C, Chronic, Hepatitis D, Chronic, Hepatitis, Chronic...	Hepatitis C, Hepatitis, Hepatitis B, Hepatitis B, Chronic, Hepatitis C, Chronic, Hepatitis, Chronic	Hepatitis C, Carcinoma, Hepatocellular, Coinfection, HIV Infections, Hemophilia A, Hepatitis, Hepatitis B...	Hepatitis C, Coinfection, Depression, Fibrosis, HIV Infections, Hepatitis, Hepatitis B...	Hepatitis C, HIV Infections, Hemophilia A, Hepatitis, Hepatitis B, Hepatitis B, Chronic, Hepatitis C, Chronic...	11	19

*Table 12. 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 hepatitis c. **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	Drug rank
DB08895	Tofacitinib	JAK3, JAK2	0.83	Alopecia, Alopecia Areata, Arthritis, Arthritis, Psoriatic, Arthritis, Rheumatoid, Psoriasis	Arthritis, Arthritis, Rheumatoid, Colitis, Colitis, Ulcerative, Dermatomyositis, Erythema, Lupus Erythematosus, Systemic...	Alopecia, Alopecia Areata, Arthritis, Arthritis, Rheumatoid, Dermatitis, Dermatitis, Atopic, Erythema...	Arthritis, Arthritis, Juvenile, Arthritis, Psoriatic, Arthritis, Rheumatoid, Colitis, Colitis, Ulcerative, Ulcer	Arthritis, Arthritis, Rheumatoid	29
DB01017	Minocycline	IL1B, CASP1	0.72	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...	35
DB00026	Anakinra	IL1R1	0.71	Arthritis, Arthritis, Juvenile, Arthritis, Rheumatoid, Coronary Artery Disease, Cryopyrin-Associated Periodic Syndromes, Diabetes Mellitus, Diabetes Mellitus, Type 2...	Adenocarcinoma, Arthropathy, Neurogenic, Breast Neoplasms, Cryopyrin-Associated Periodic Syndromes, Dermatitis, Dermatitis, Atopic, Diabetes Mellitus...	Acrodermatitis, Amyotrophic Lateral Sclerosis, Anterior Cruciate Ligament Injuries, Arthritis, Arthritis, Gouty, Arthritis, Juvenile, Arthritis, Rheumatoid...	Arteritis, Arthritis, Arthritis, Juvenile, Arthritis, Rheumatoid, Dermatomyositis, Diabetes Mellitus, Type 1, Fatigue...	Arthritis, Arthritis, Rheumatoid, Diabetes Mellitus, Diabetes Mellitus, Type 2, Knee Injuries, Myocarditis, Pericarditis...	36
DB06372	Rilonacept	IL1B	0.71	Bursitis, Cardiovascular Diseases, Gout, Kidney Diseases, Renal Insufficiency, Chronic, Urticaria, Vascular Diseases	Arthritis, Arthritis, Juvenile, Diabetes Mellitus, Diabetes Mellitus, Type 1, Hearing Loss, ST Elevation Myocardial Infarction, Scleroderma, Diffuse...	Anemia, Atherosclerosis, Coronary Artery Disease, Cryopyrin-Associated Periodic Syndromes, Familial Mediterranean Fever, Hepatitis, Hepatitis, Alcoholic...	Cryopyrin-Associated Periodic Syndromes, Genetic Diseases, Inborn, Gout, Urticaria	Renal Insufficiency, Renal Insufficiency, Chronic	36
DB00775	Tirofiban	ITGB3, ITGA2B	0.66	Acute Coronary Syndrome, Coronary Artery Disease, Myocardial Infarction, Non-ST Elevated Myocardial Infarction, ST Elevation Myocardial Infarction, Stroke	Coronary Artery Disease, Renal Insufficiency	Acute Coronary Syndrome, Angina, Unstable, Myocardial Infarction, Stroke	Acute Coronary Syndrome, Angina, Unstable, Coronary Artery Disease, Myocardial Infarction, ST Elevation Myocardial Infarction, Stroke	Acute Coronary Syndrome, Coronary Artery Disease, Coronary Disease, Myocardial Infarction, No-Reflow Phenomenon, ST Elevation Myocardial Infarction	46

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 13 shows the resulting list of druggable master regulators, which represent the predicted drug targets of the studied pathology. Table 14 lists chemical compounds and known drugs potentially acting on the corresponding master regulators.

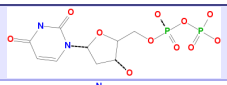
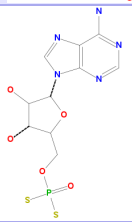
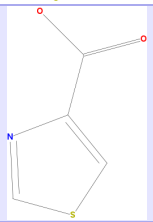
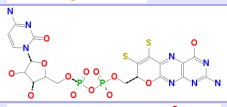
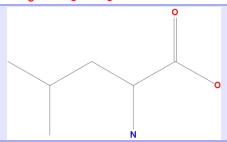
*Table 13. Extended list of drug targets revealed in this study (targets that are predicted by PASS program potentially targeted by an extended list of known drugs and pharmaceutically active chemical compounds). The column **Druggability score** contains a numeric value which indicates how suitable 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	Total rank	LogFoldChange
ENSG00000112576	CCND3	CCND3	cyclin D3	5.61	102	0.79
ENSG00000164342	TLR3	TLR3	toll like receptor 3	4.35	119	0.75
ENSG00000136869	TLR4	TLR4	toll like receptor 4	5.92	123	0.62
ENSG00000170458	CD14	CD14	CD14 molecule	1.68	123	0.62
ENSG00000140464	PML	PML	promyelocytic leukemia	10.7	149	1.35
ENSG00000104312	RIPK2	RIPK2	receptor interacting serine/threonine kinase 2	1.8	154	0.5
ENSG00000115594	IL1R1	IL1R1	interleukin 1 receptor type 1	1.68	180	0.62
ENSG00000125538	IL1B	IL1B	interleukin 1 beta	20.76	180	0.62
ENSG00000198001	IRAK4	IRAK4	interleukin 1 receptor associated kinase 4	5.45E-2	180	0.62
ENSG00000171132	PRKCE	PRKCE	protein kinase C epsilon	40.78	182	0.35

Table 14. 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 hepatitis c 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 →

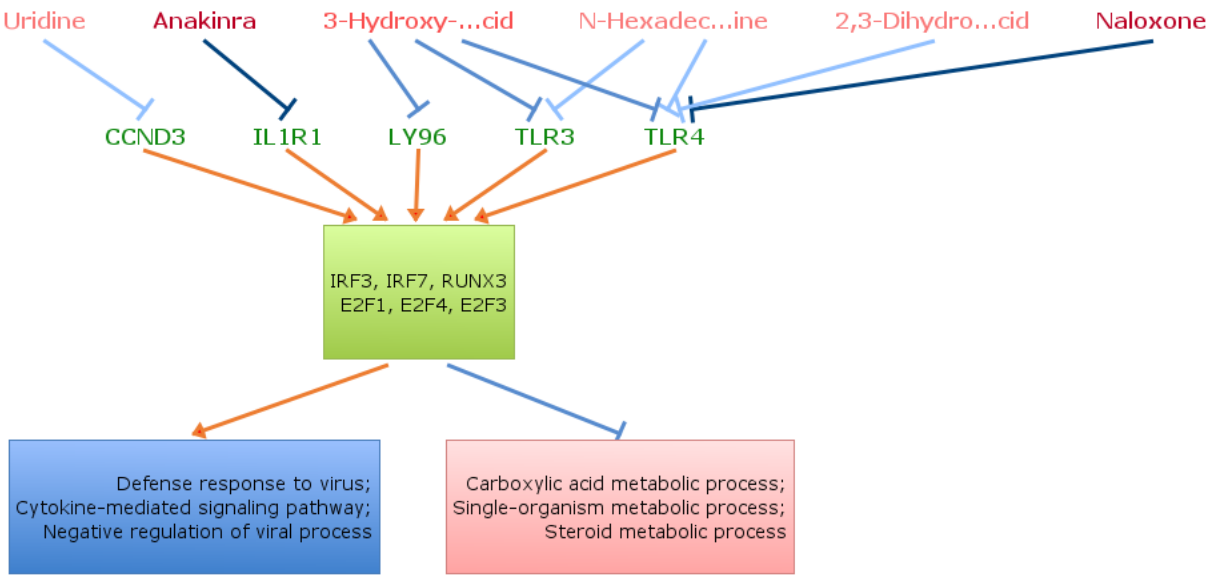
Name	Structure	Target names	Target activity score	Disease activity score	Drug rank
Deoxyuridine-5'-Diphosphate		PRKCG, PRKD3, PRKCQ, PRKCE, PRKCD, PRKCZ	1.89	0.22	27
Adenosine-5'-(Dithio)Phosphate		PRKCG, GRB2, PRKCQ, PRKCE, PRKCD, PRKCZ	1.17	0.21	46
1,3-Thiazole-4-Carboxylic Acid		PRKCG, PRKD3, PRKCQ, PRKCE, PRKCD, PRKCZ	0.92	0.25	50
Pterin Cytosine Dinucleotide		ERBB3, EPHB2, EGFR, MERTK, PRKACA, ERBB4, RET...	0.86	0.18	74
L-Leucine		NCF1, NCF2, NCF4, CYBA, PIN1	0.96	0.16	77

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 data set that contains *transcriptomics* data obtained from *liver* tissue. The study is done in the context of *hepatitis c*. 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 - Detailed report. Composite modules and master-regulators (high expressed genes in Experiment).
2. Supplementary table 2 - Detailed report. Composite modules and master-regulators (low expressed genes in Experiment).
3. Supplementary table 3 - Detailed report. Pharmaceutical compounds and drug targets.

Disclaimer

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

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

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

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

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

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