EGFR and MET are promising druggable targets for treating Squamous Cell Carcinoma that control activity of CTNNB1, SMAD4 and SP1 transcription factors on promoters of differentially expressed genes

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 08/12/2025 ; Run on 12/07/2026 ; Report generated on 12/07/2026

Genome Enhancer release 3.8 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2026.1)

Abstract

In the present study we applied the software package "Genome Enhancer" to a data set that contains *transcriptomics* data. The study is done in the context of *Squamous Cell Carcinoma*. The goal of this pipeline is to identify potential drug targets in the molecular network that governs the studied pathological process. In the first step of analysis pipeline discovers transcription factors (TFs) that regulate genes activities in the pathological state. The activities of these TFs are controlled by so-called master regulators, which are identified in the second step of analysis. After a subsequent druggability checkup, the most promising master regulators are chosen as potential drug targets for the analyzed pathology. At the end the pipeline comes up with (a) a list of known drugs and (b) investigational active chemical compounds with the potential to interact with selected drug targets.

From the data set analyzed in this study, we found the following TFs to be potentially involved in the regulation of the differentially expressed genes: CTNNB1, SMAD4, JUND, SP1, NR3C1 and ZNF462. The subsequent network analysis suggested

- EGFR
- c-Met
- EGFR

as the most promising molecular targets for further research, drug development and drug repurposing initiatives on the basis of identified molecular mechanism of the studied pathology.

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® [5] is employed together with the TF binding site identification algorithms Match [6] and CMA [7]. The second step involves the signal transduction database TRANSPATH® [8] and special graph search algorithms [10-11] 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 [11]. In addition, some known drugs and investigational active chemical compounds are subsequently predicted as potential ligands for the revealed molecular targets. They are predicted using a pre-computed database of spectra of biological activities of chemical compounds of a library of 2245 known drugs and investigational chemical compounds from HumanPSD™ database. The spectra of biological activities for these compounds are computed using the program PASS on the basis of a (Q)SAR approach [12-14]. These predictions can be used for the research purposes - for further drug development and drug repurposing initiatives.

2. Data

For this study the following experimental data was used:

Table 1. Experimental datasets used in the study

File name	Data type
SRR349741_e.fastq	Transcriptomics
SRR349742.fastq	Transcriptomics
SRR349748.fastq	Transcriptomics
SRR349749.fastq	Transcriptomics

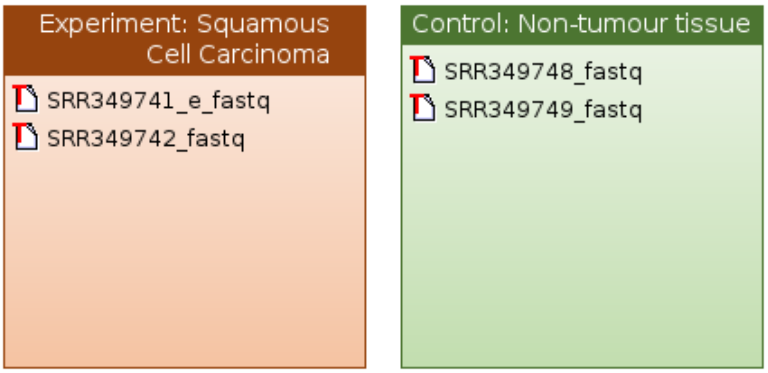


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

3. Results

We have compared the following conditions: Experiment: Squamous Cell Carcinoma *versus* Control: Non-tumour tissue.

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 edgeR tool (R/Bioconductor package integrated into our pipeline) and compared gene expression in the following sets: "Experiment: Squamous Cell Carcinoma" with "Control: Non-tumour tissue". edgeR 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 560 upregulated genes (LogFC>0.5) out of which 277 genes were found as significantly upregulated (p-value<0.01) and 526 downregulated genes (LogFC<-0.5) out of which 314 genes were significantly downregulated (p-value<0.01).

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: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.*
See full table →

ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000115758	ODC1	ornithine decarboxylase 1	6.73	10.37	6.78E-9	6.85E-7
ENSG00000148053	NTRK2	neurotrophic receptor tyrosine kinase 2	5.99	9.33	1.47E-9	1.99E-7
ENSG00000113140	SPARC	secreted protein acidic and cysteine rich	5.74	10.75	1.44E-7	9.8E-6
ENSG00000163359	COL6A3	collagen type VI alpha 3 chain	5.19	9.2	1.54E-5	4.36E-4
ENSG00000120708	TGFBI	transforming growth factor beta induced	4.81	8.83	1.53E-9	2.01E-7
ENSG00000134871	COL4A2	collagen type IV alpha 2 chain	4.69	8.02	9.35E-10	1.36E-7
ENSG00000186340	THBS2	thrombospondin 2	4.67	8.54	6.35E-5	1.34E-3
ENSG00000146648	EGFR	epidermal growth factor receptor	4.44	9.65	3.25E-4	4.84E-3
ENSG00000145824	CXCL14	C-X-C motif chemokine ligand 14	4.43	8.61	2.44E-5	6.33E-4
ENSG00000187134	AKR1C1	aldo-keto reductase family 1 member C1	4.41	9.04	1.06E-10	2.88E-8

*Table 3. Top ten significant **down-regulated** genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.*
See full table →

ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000136155	SCEL	sciellin	-7.72	11.12	2.73E-15	5.38E-12
ENSG00000163209	SPRR3	small proline rich protein 3	-6.69	14.45	8.44E-4	1.1E-2
ENSG00000143369	ECM1	extracellular matrix protein 1	-6.38	11.04	4.35E-10	7.45E-8
ENSG00000189334	S100A14	S100 calcium binding protein A14	-6.37	10.46	1.1E-10	2.88E-8
ENSG00000229732		novel transcript	-6.27	12.97	4.93E-12	2.42E-9
ENSG00000086548	CEACAM6	CEA cell adhesion molecule 6	-6.2	10.31	5.18E-14	4.37E-11
ENSG00000171401	KRT13	keratin 13	-6.15	14.93	8.06E-11	2.44E-8
ENSG00000087128	TMPRSS11E	transmembrane serine protease 11E	-5.98	10.11	6.26E-9	6.48E-7
ENSG00000197632	SERPINB2	serpin family B member 2	-5.86	8.73	5.56E-14	4.37E-11
ENSG00000165272	AQP3	aquaporin 3 (Gill blood group)	-5.81	11.35	5.75E-5	1.23E-3

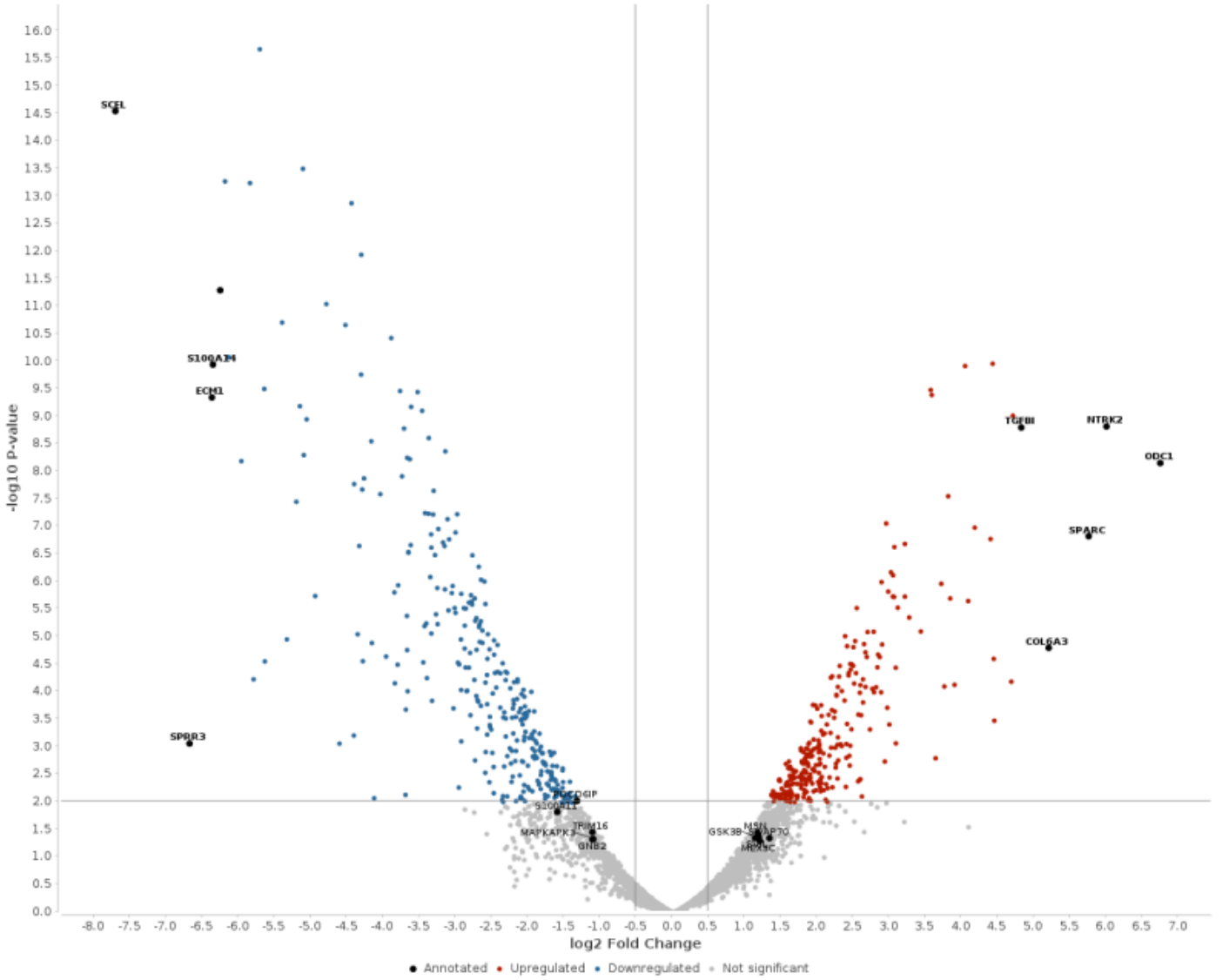


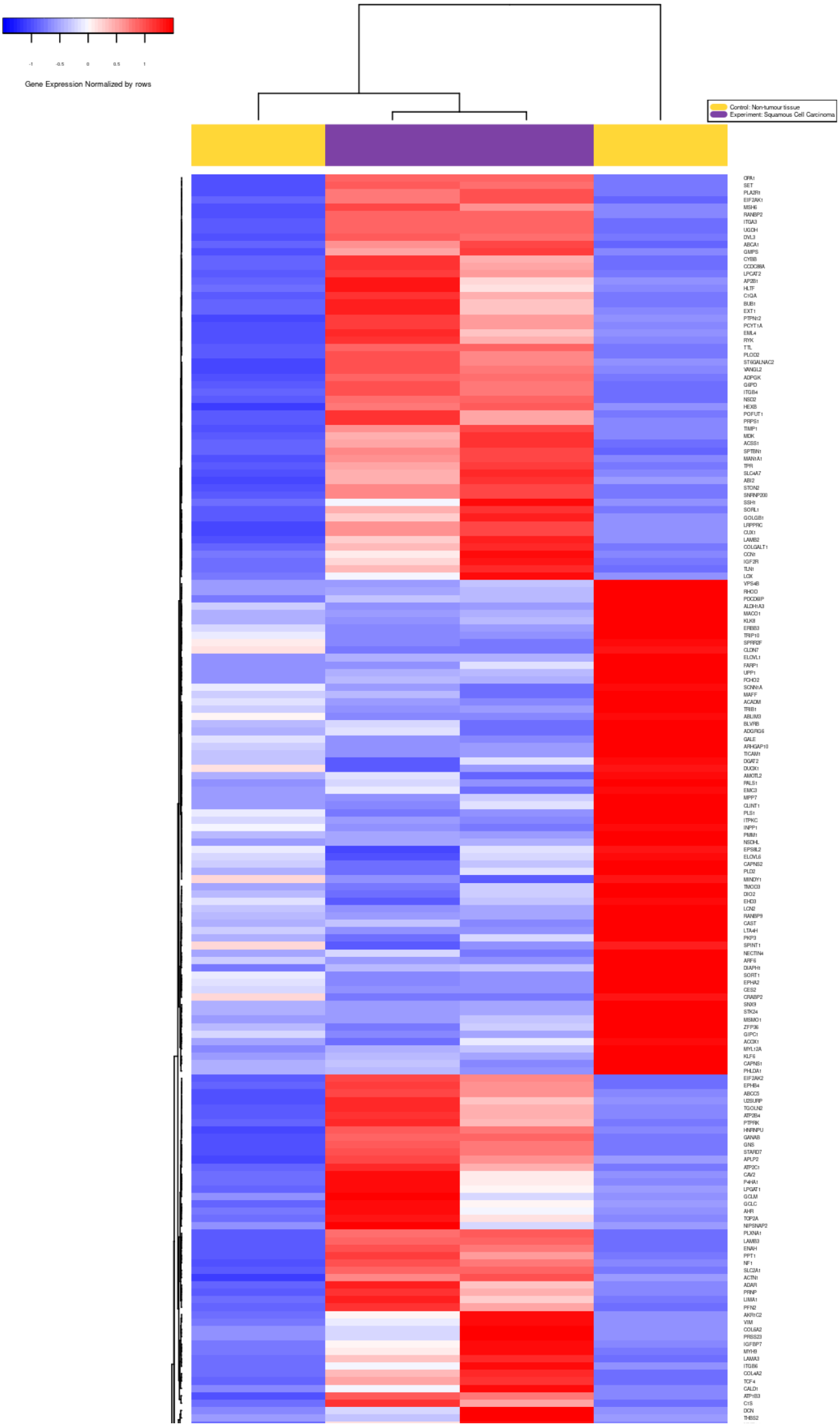
Figure 3. Volcano plot. Each dot represents one gene. Gray dots represent genes with no significant expression between conditions analyzed, the blue dots represent downregulated genes, and the red dots represent upregulated genes. Annotated are the top DEGs (bold) and the genes that encode top 10 up- and down-regulated master-regulators identified in the current study.

3.2. Functional classification of genes

A functional analysis of differentially expressed genes was done by mapping the significant up-regulated and significant down-regulated genes to several known ontologies, such as Gene Ontology (GO), disease ontology (based on HumanPSD™ database) and the ontology of signal transduction and metabolic pathways from the TRANSPATH® database. Statistical significance was computed using a binomial test.
Figures 4-9 show the most significant categories.

Heatmap of differentially expressed genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue

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


-1
-0.5
0
0.5
1
Gene Expression Normalized by rows
Control: Non-tumour tissue
Experiment: Squamous Cell Carcinoma
OPA1
SET
PLA2R1
EIF2AK1
MSH6
RANBP2
ITGA3
UGDH
DVL3
ABCA1
GMPS
CYBB
CCDC88A
LPCAT2
AP2B1
HLTF
C1QA
BUB1
EXT1
PTPN12
PCYT1A
EML4
RYK
TTL
PLOD2
ST6GALNAC2
VANGL2
ADIPOR
GIPD
ITGB4
NSD2
HEXB
POFUT1
PRPS1
TIMP1
MDK
ACSS1
SPTBN1
MAN1A1
TPR
SLC4A7
ABI2
STON2
SNRNP200
SSH1
SORL1
GOLGB1
LRPPRC
CUX1
LAMB2
COL5A1T1
CCN1
IGF2R
TLN1
LOX
VPS4B
RHOD
PDCD6IP
ALDH1A3
MACC1
KLK8
ERBB3
TRIP10
SPRR2F
CLDN7
ELOVL1
FARP1
UPP1
PCHD2
SCNN1A
MAFF
ACADM
TRIB1
ABLIM3
BLVRB
ADGRG6
GALE
ARHGAP10
TICAM1
DGAT2
DUOX1
ANKTL2
PALS1
EMC3
MPP7
CLINT1
PLS1
ITPKC
INPP1
PMM1
NSDHL
EPS8L2
ELOVL6
CAPNS2
PLD2
MINDY1
TMOD3
DIO2
EHD3
LCN2
RANBP9
CAST
LTA4H
PKP3
SPINT1
NECTIN4
ARF6
DIAPH1
SORT1
EPHA2
CES2
CRABP2
SNX9
STK24
MSMO1
ZFP36
GIPC1
ACOX1
MYL12A
KLF6
CAPNS1
PHLDA1
EIF2AK2
EPHB4
ABCC5
LOSURP
TGOLN2
ATP2B4
PTPRK
HNRNPU
GANAB
GNS
STARD7
APLP2
ATP2C1
CAV2
P4HA1
LPGAT1
GCLM
GCLC
AHR
TOP2A
NIPSNAP2
PLXNA1
LAMB3
ENAH
PPT1
NF1
SLC2A1
ACTN1
ADAR
PRNP
LIMA1
PFN2
AKR1C2
VIM
COL6A2
PRSS23
IGFBP7
MYH9
LAMA3
ITGB6
COL4A2
TCF4
CALD1
ATP1B3
CYS
DCN
THBS2

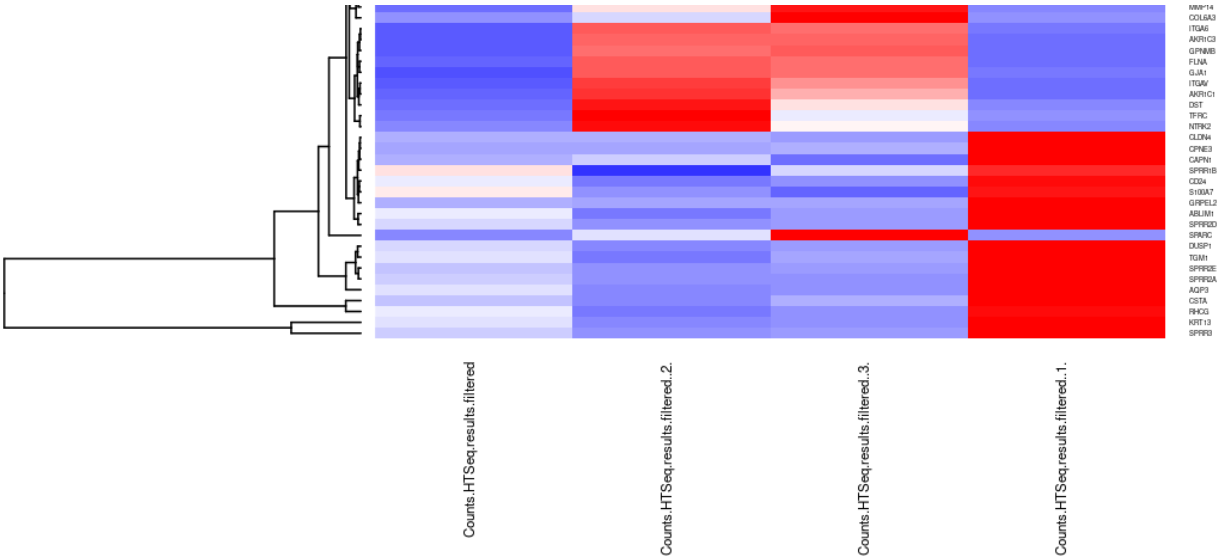


Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.
See full diagram →

Up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue:

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

GO (biological process)

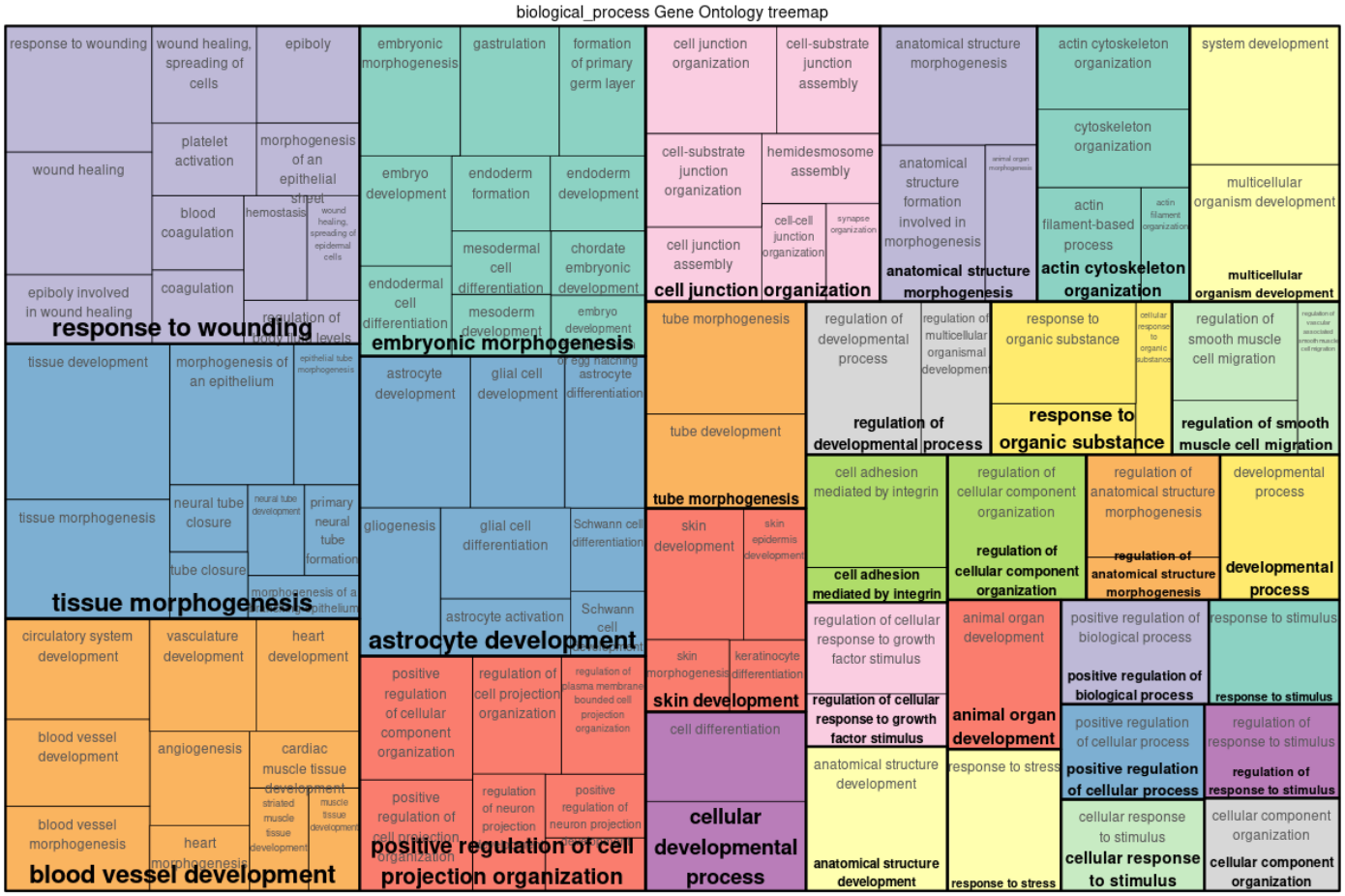


Figure 4. Enriched GO (biological process) of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.
Full classification →

TRANSPATH® Pathways (2026.1)

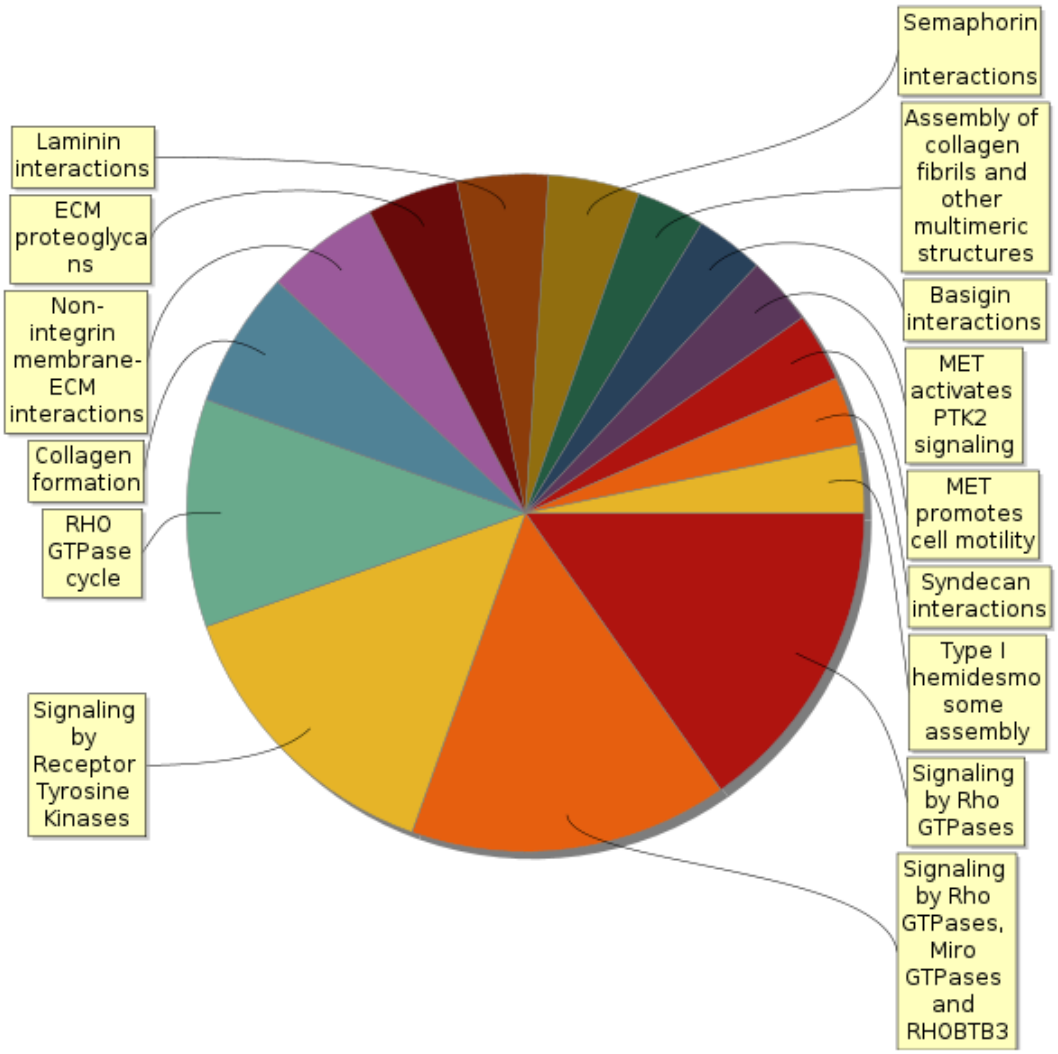


Figure 5. Enriched TRANSPATH® Pathways (2026.1) of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

Full classification →

HumanPSD(TM) disease (2026.1)

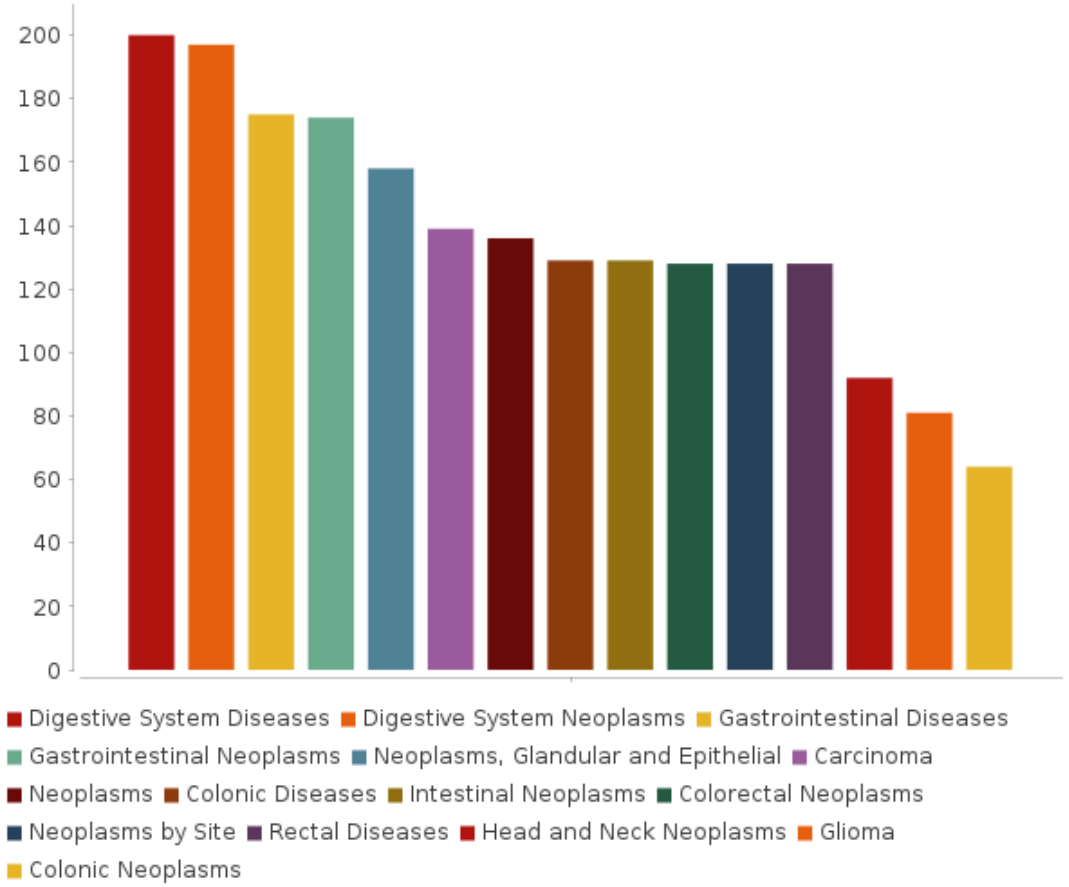


Figure 6. Enriched HumanPSD(TM) disease (2026.1) of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

Full classification →

Down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue:

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

GO (biological process)

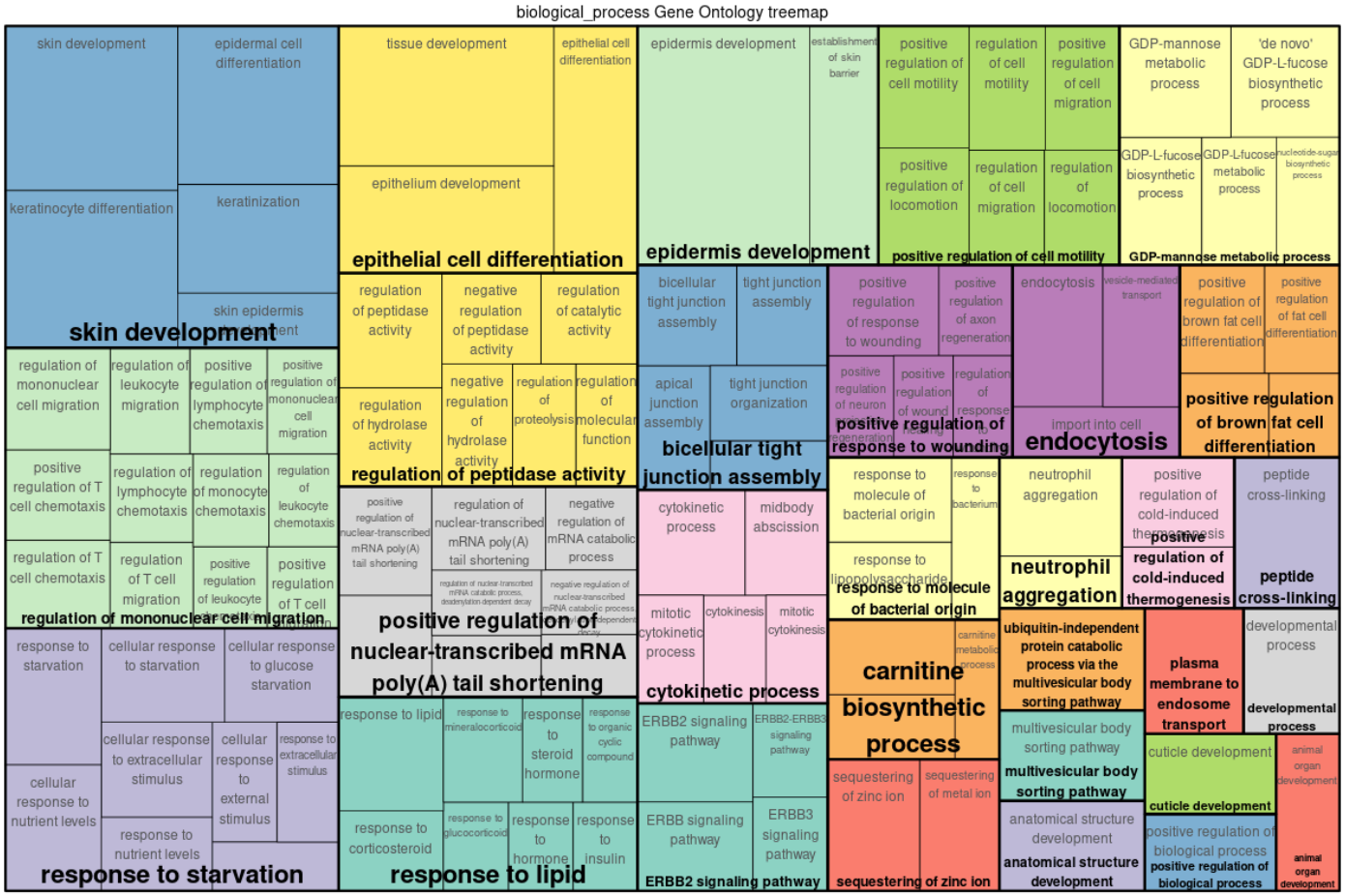


Figure 7. Enriched GO (biological process) of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.
Full classification →

TRANSPATH® Pathways (2026.1)

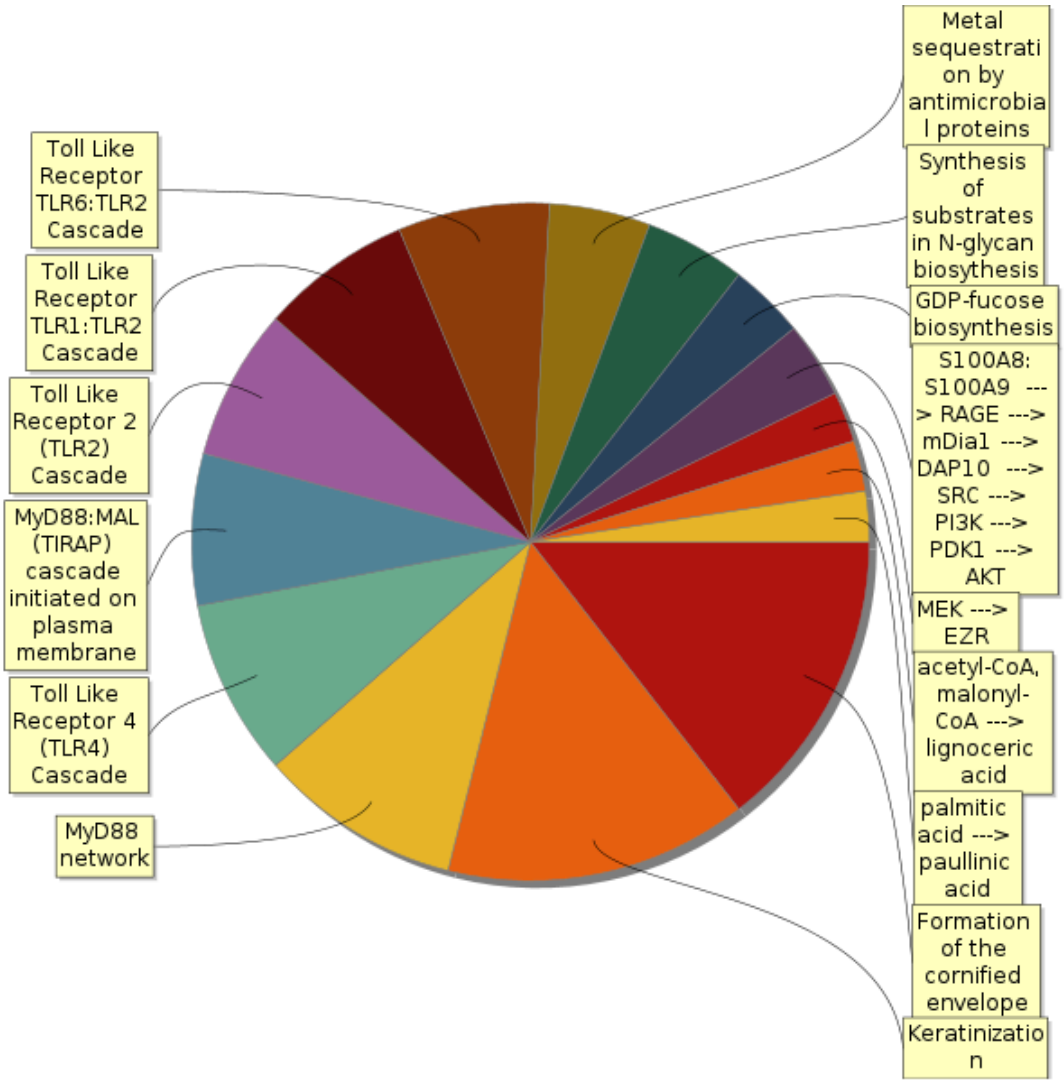


Figure 8. Enriched TRANSPATH® Pathways (2026.1) of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

Full classification →

HumanPSD(TM) disease (2026.1)

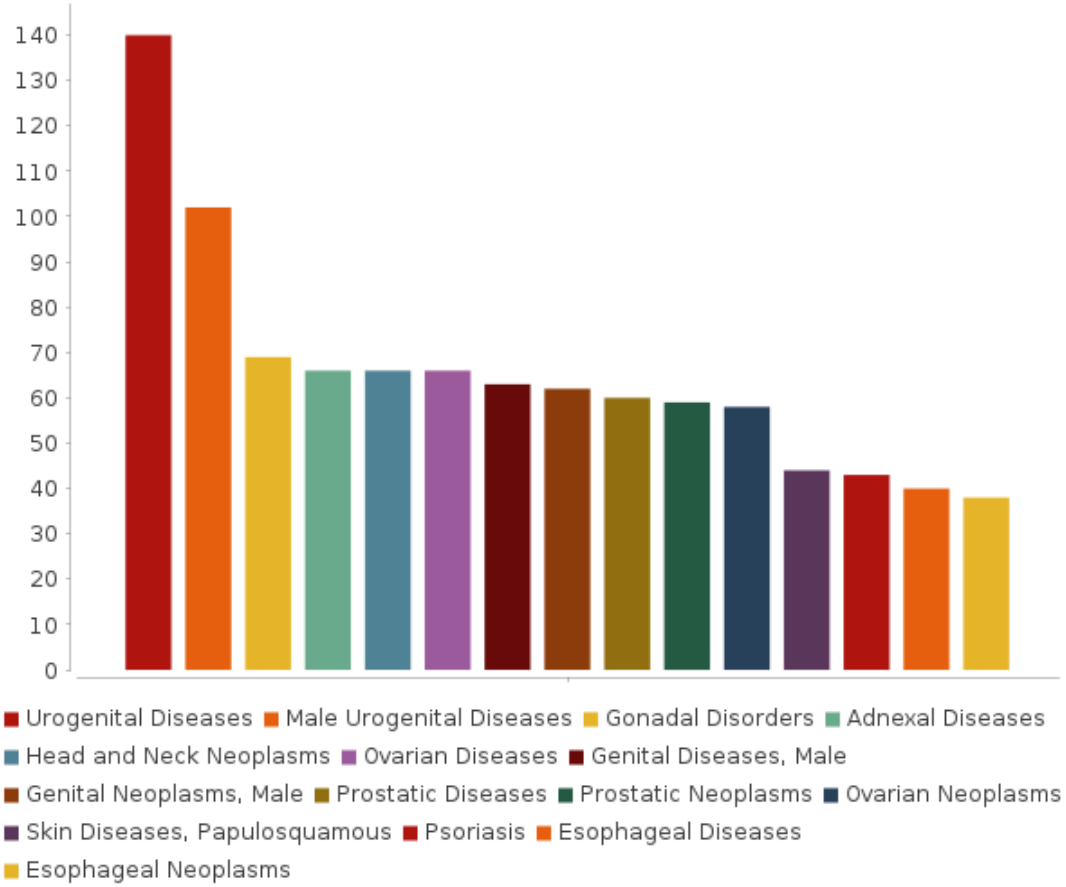
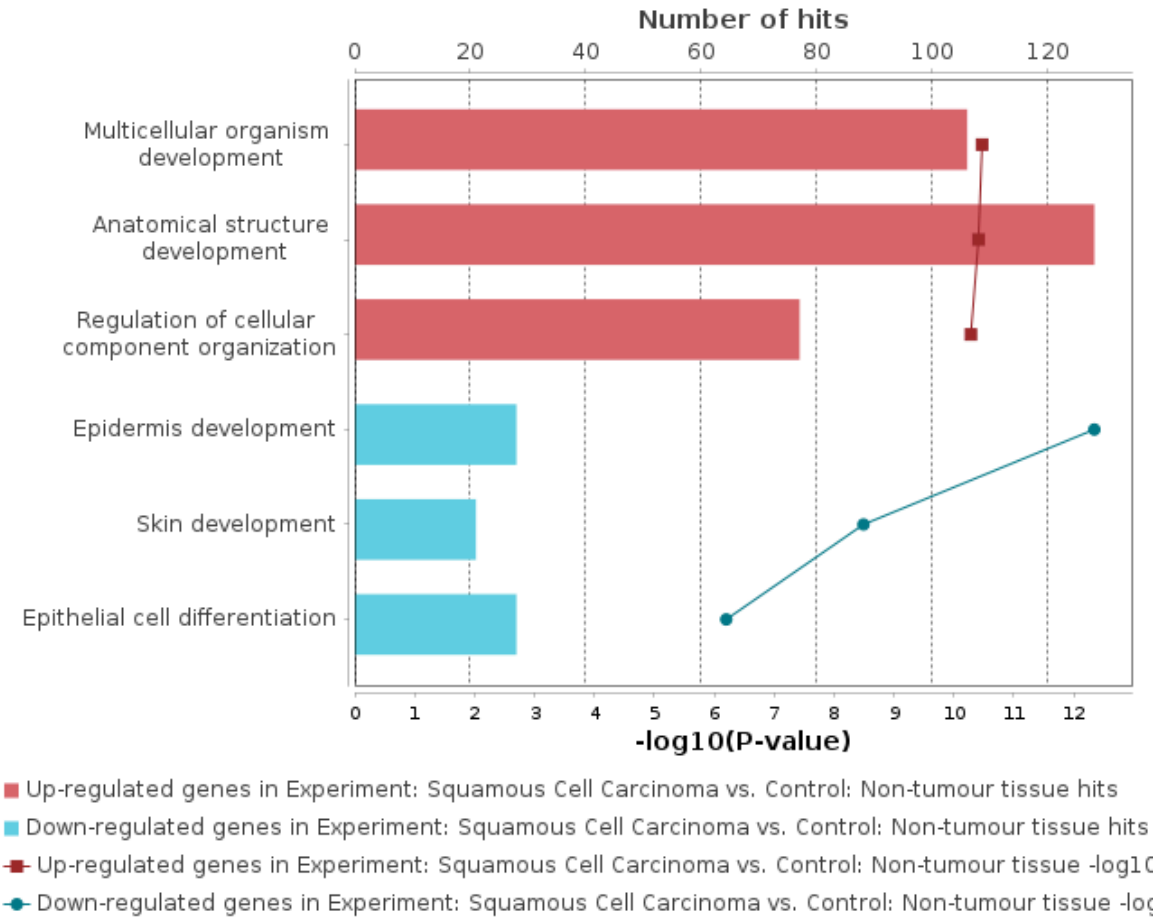


Figure 9. Enriched HumanPSD(TM) disease (2026.1) of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

Full classification →

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



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

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

Classically, **enhancers** are defined as regions in the genome that increase transcription of one or several genes when inserted in either orientation at various distances upstream or downstream of the gene [7]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [8].

In the current work, we use the Genomics data from the "Yes VCF track" track to predict positions of potential ***enhancers*** where the observed sequence variations may influence the gene expression in the pathology under study. We scan 5kb flanking regions and the body of all genes caring the variations, with a sliding window of 1100bp size and find the position of the window with the maximal sum of the mutation weights, where we then perform the search for potential condition-specific enhancers (CMA model search).

We analyzed mutations that were revealed in the potential enhancers located upstream, downstream or inside the ***target genes*** (see Table 4). We identified 595 mutations potentially affecting gene regulation. Table 5 shows the following lists of PWMs whose sites were lost or gained due to these mutations. Weighting of mutations was done in respect to the significance of the change in TF affinity binding to the sequence. Mutations that maximally affected the change of binding affinity received higher weights. These PWMs were put in focus of the CMA algorithm that constructs the model of the enhancers by specifying combinations of TF motifs (see more details of the algorithm in the Methods section).

Table 4. Mutations revealed in Experiment: Squamous Cell Carcinoma versus Control: Non-tumour tissue

See full table →

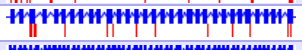
ID	Gene symbol	Gene schematic representation	Number of variations
ENSG00000146648	EGFR		21
ENSG00000083857	FAT1		16
ENSG00000134871	COL4A2		13
ENSG00000186340	THBS2		10
ENSG00000226445	THBS2-AS1		9
ENSG00000145012	LPP		8
ENSG00000114999	TTL		7
ENSG00000142173	COL6A2		7
ENSG00000152291	TGOLN2		7
ENSG00000157214	STEAP2		7

Table 5. PWMs whose sites were lost or gained due to mutations in Experiment: Squamous Cell Carcinoma and Control: Non-tumour tissue

See full table →

ID	P-value (gains)	P-value (losses)	yesCount (gains)	yesCount (losses)
V$EGR1_07	4.62E-2	1.48E-24	5	1134
V$E2F7_04	3.9E-2	5.96E-23	11	744
V$GLI2_05	2.5E-2	1.36E-22	11	2807
V$E2F3_05	1.58E-2	3.85E-25	27	1467
V$E2F1_Q4_01	1.5E-2	1.98E-27	11	1490
V$RUNX3_01	5.83E-6	3.04E-24	151	1895
V$E2F1_05	3.16E-7	6.77E-27	39	1042
V$TEF_05	2.04E-7	1.43E-18	452	538
V$MEIS1ELF1_01	2.29E-11	1.37E-16	2061	1805
V$TFDP1_03	1.12E-12	6.17E-24	275	1398
V$GLI1_Q3	1.34E-19		833	
V$MECP2_02	3.65E-20	1.39E-3	738	39
V$SP1_09	3.12E-20	4.61E-2	342	4
V$E2F1DP2_01	1.1E-20	1.22E-16	2155	2222
V$E2F1EOMES_02	8.12E-21	5.89E-4	705	366
V$SP2_09	1.61E-22	1.16E-2	236	25
V$E2F3_10	7.25E-24	4.61E-2	1850	4
V$RBAK_01	1.65E-24	5.74E-22	454	471
V$E2F4_14	8.56E-25		1845	
V$E2F2_08	3.25E-26		1119	

We applied the Composite Module Analyst (CMA) [7] 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: Squamous Cell Carcinoma vs. Control: Non-tumour tissue).

To build the most specific composite modules we choose all significant up-regulated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

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$ZFP281_01	V$SMAD4_07	V$MEIS1_09	V$JUND_Q6	V$BETACATENIN_Q3	V$SF1_Q5_01
0.94; N=3	0.93; N=2	0.93; N=2	0.93; N=3	0.91; N=3	0.95; N=2

V$TGIF1_04
0.81; N=2

Module width: 142

Module 2:

V$ZFP281_01	V$GATA3_11	V$COUPTF2_01	V$MEF2C_06	V$PR_03	V$AP2_Q6
0.94; N=1	0.98; N=3	0.95; N=1	0.91; N=3	0.79; N=2	0.94; N=2

Module width: 136

Model score (-p*log10(pval)): 15.10
Wilcoxon p-value (pval): 2.51e-33
Penalty (p): 0.463
Average yes-set score: 3.56
Average no-set score: 2.11
AUC: 0.76
Separation point: 2.61
False-positive: 36.62%
False-negative: 22.83%

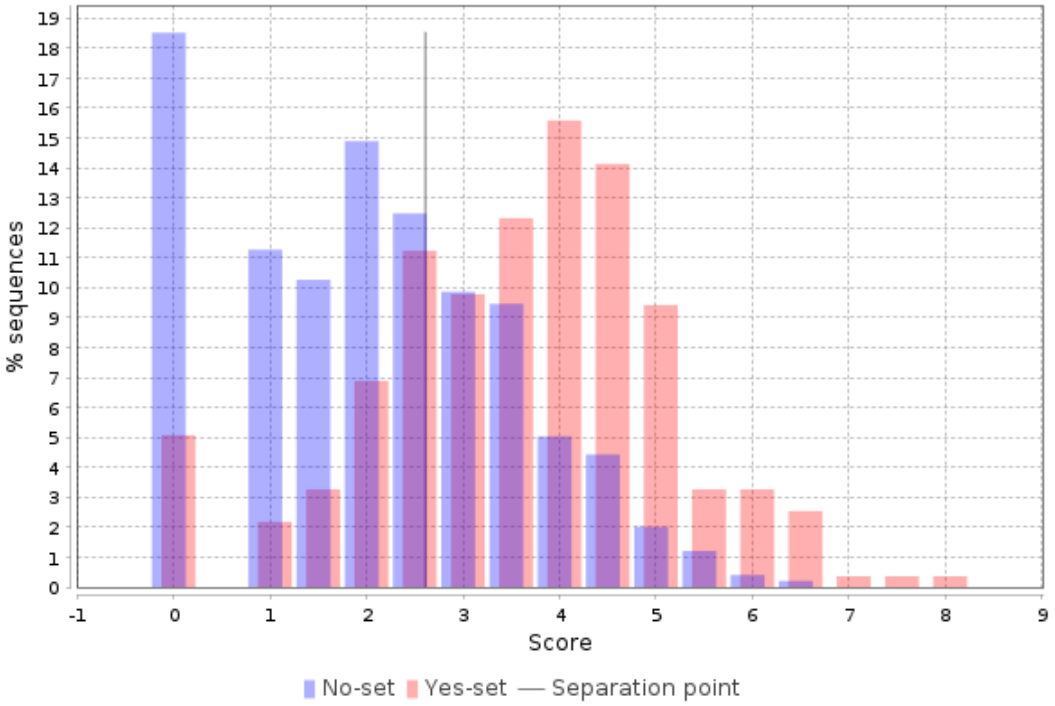


Table 6. List of top ten up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*
See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000059588	TARBP1	TAR (HIV-1) RNA binding protein 1	7.78	MEF-2C(h), GATA-3(h), PR(h), beta-catenin(h), TGIF-1(h), Meis1(h), JunD(h)
ENSG00000150093	ITGB1	integrin subunit beta 1	7.43	MEF-2C(h), JunD(h), GATA-3(h), PR(h), beta-catenin(h), COUP-TFII(h)
ENSG00000162521	RBBP4	RB binding protein 4, chromatin remodeling factor	7.27	PR(h), MEF-2C(h), JunD(h), GATA-3(h), TGIF-1(h), SMAD4(h), beta-catenin(h)
ENSG00000138018	SELENOI	selenoprotein I	7.02	Meis1(h), JunD(h), beta-catenin(h), MEF-2C(h), TGIF-1(h), GATA-3(h)
ENSG00000146242	TPBG	trophoblast glycoprotein	6.76	MEF-2C(h), GATA-3(h), beta-catenin(h), JunD(h), Meis1(h), ZNF281(h)
ENSG00000115306	SPTBN1	spectrin beta, non-erythrocytic 1	6.73	GATA-3(h), beta-catenin(h), MEF-2C(h), JunD(h), ZNF281(h), Meis1(h)
ENSG00000140848	CPNE2	copine 2	6.71	FTZ-F1(h), ZNF281(h), TGIF-1(h), GATA-3(h), JunD(h)
ENSG00000146247	PHIP	pleckstrin homology domain interacting protein	6.67	JunD(h), GATA-3(h), MEF-2C(h), COUP-TFII(h), beta-catenin(h)
ENSG00000151694	ADAM17	ADAM metallopeptidase domain 17	6.54	MEF-2C(h), GATA-3(h), beta-catenin(h), COUP-TFII(h), JunD(h), Meis1(h)
ENSG00000134330	IAH1	isoamyl acetate hydrolyzing esterase 1 (putative)	6.54	MEF-2C(h), GATA-3(h), beta-catenin(h), COUP-TFII(h), JunD(h), Meis1(h)

Enhancer model potentially involved in regulation of target genes (down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue).

To build the most specific composite modules we choose top 300 significant down-regulated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

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$SP1_09	V$GATA1_06	V$ZNF462_01	V$IK3_01	V$HOXA10_04	V$RARG_09
0.93; N=2	0.94; N=2	0.92; N=2	0.84; N=1	0.93; N=2	0.72; N=1

Module width: 121

Module 2:

V$SP1_09	V$GR_Q6_01	V$FOXA1_07	V$NKX25_Q5	V$ESRRA_10	V$GLI1_Q3
0.94; N=2	0.99; N=2	0.90; N=1	0.95; N=2	0.90; N=2	0.94; N=1

Module width: 97

Model score (-p*log10(pval)): 13.85
Wilcoxon p-value (pval): 6.58e-30
Penalty (p): 0.475
Average yes-set score: 5.09
Average no-set score: 3.90
AUC: 0.74
Separation point: 4.36
False-positive: 34.41%
False-negative: 24.75%

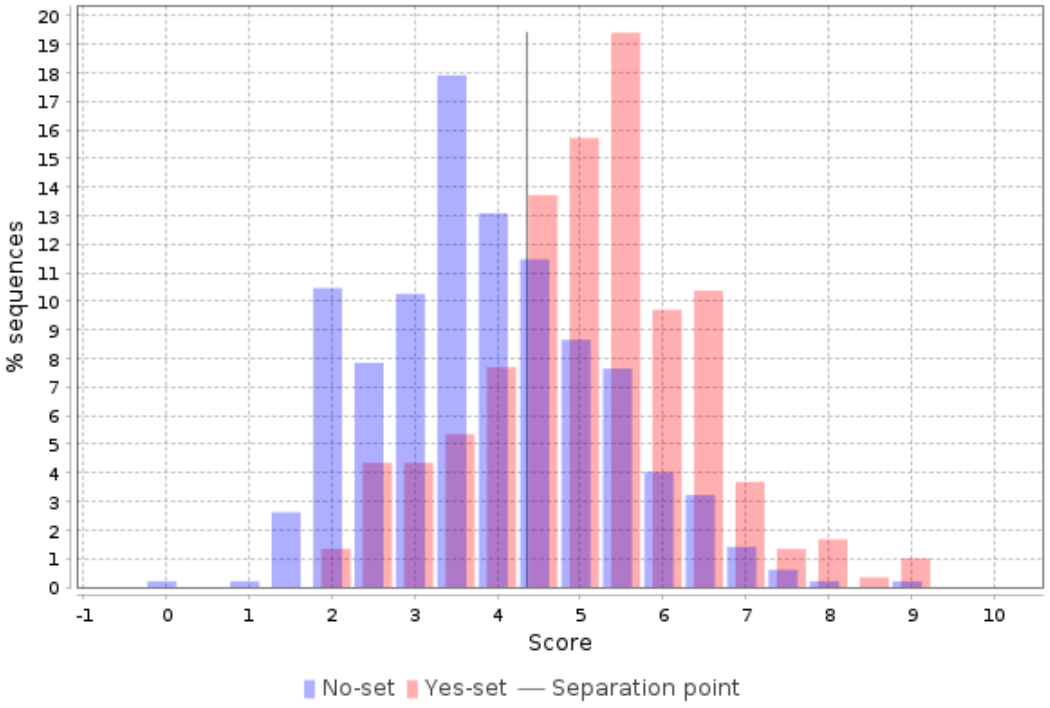


Table 7. List of top ten down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000197324	LRP10	LDL receptor related protein 10	9.15	Hox-A10(h), GATA-1(h), GR(h), ZNF462(h), RAR-gamma(h), FOXA1(h), Sp1(h)...
ENSG00000167656	LY6D	lymphocyte antigen 6 family member D	9.15	RAR-gamma(h), ZNF462(h), NKX-2.5(h), GR(h), Sp1(h), GLI1(h), IKZF1(h)
ENSG00000052344	PRSS8	serine protease 8	8.85	ERR1(h), IKZF1(h), Sp1(h), RAR-gamma(h), GR(h), ZNF462(h), NKX-2.5(h)...
ENSG00000101439	CST3	cystatin C	8.59	IKZF1(h), RAR-gamma(h), ZNF462(h), ERR1(h), NKX-2.5(h), Sp1(h)
ENSG00000186187	ZNRF1	zinc and ring finger 1	8.39	Hox-A10(h), IKZF1(h), ZNF462(h), RAR-gamma(h), Sp1(h), NKX-2.5(h), FOXA1(h)
ENSG00000173846	PLK3	polo like kinase 3	8.18	ZNF462(h), GLI1(h), Sp1(h), RAR-gamma(h), GATA-1(h), Hox-A10(h)
ENSG00000160685	ZBTB7B	zinc finger and BTB domain containing 7B	8.15	RAR-gamma(h), Sp1(h), ZNF462(h), NKX-2.5(h), GLI1(h)
ENSG00000052802	MSMO1	methylsterol monooxygenase 1	8.02	Hox-A10(h), ZNF462(h), NKX-2.5(h), RAR-gamma(h), Sp1(h), ERR1(h), IKZF1(h)
ENSG00000185499	MUC1	mucin 1, cell surface associated	7.96	GR(h), GLI1(h), Sp1(h), NKX-2.5(h), IKZF1(h), ZNF462(h)
ENSG00000171223	JUNB	JunB proto-oncogene, AP-1 transcription factor subunit	7.91	NKX-2.5(h), IKZF1(h), GLI1(h), Sp1(h), ZNF462(h), RAR-gamma(h), FOXA1(h)...

On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 14 and 11 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: Squamous Cell Carcinoma vs. Control: Non-tumour tissue). ***Yes-No ratio*** *is the ratio between frequencies of the sites in Yes sequences versus No sequences. It describes the level of the enrichment of binding sites for the indicated TF in the regulatory target regions.* ***Regulatory score*** *is the measure of involvement of the given TF in the controlling of expression of genes that encode master regulators presented below (through positive feedback loops).*

See full table →

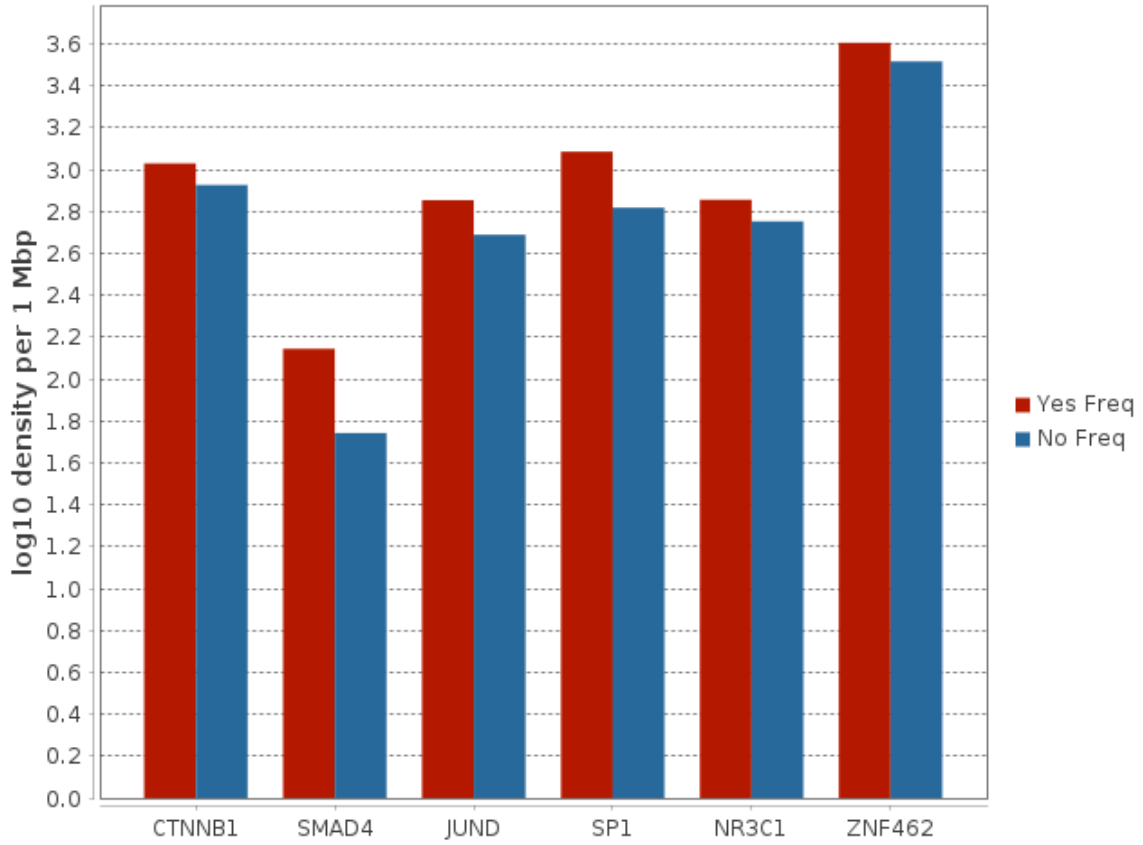
ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000017049	CTNNB1	catenin beta 1	2.9	1.27
MO000020402	SMAD4	SMAD family member 4	2.29	2.52
MO000007834	JUND	JunD proto-oncogene, AP-1 transcription factor subunit	2.08	1.46
MO000054297	PGR	progesterone receptor	2.06	1.59
MO000031322	MEF2C	myocyte enhancer factor 2C	1.97	4.21
MO000092625	GATA3	GATA binding protein 3	1.9	1.55
MO000028758	ZNF281	zinc finger protein 281	1.8	10.82
MO000001275	TFAP2A	transcription factor AP-2 alpha	1.75	3.97
MO000019620	NR5A1	nuclear receptor subfamily 5 group A member 1	1.74	1.37
MO000013620	TGIF1	TGFB induced factor homeobox 1	1.66	1.75

Table 9. Transcription factors of the predicted enhancer model potentially regulating the differentially expressed genes (down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue). ***Yes-No ratio*** *is the ratio between frequencies of the sites in Yes sequences versus No sequences. It describes the level of the enrichment of binding sites for the indicated TF in the regulatory target regions.* ***Regulatory score*** *is the measure of involvement of the given TF in the controlling of expression of genes that encode master regulators presented below (through positive feedback loops).*

See full table →

ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000033308	SP1	Sp1 transcription factor	1.52	1.85
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	1.35	1.27
MO000092587	ZNF462	zinc finger protein 462	1.28	1.23
MO000019117	GLI1	GLI family zinc finger 1	1.23	1.78
MO000026678	IKZF1	IKAROS family zinc finger 1	1.18	1.24
MO000046001	GATA1	GATA binding protein 1	1.15	3.74
MO000089495	HOXA10	homeobox A10	1.14	
MO000026492	FOXA1	forkhead box A1	1.1	4.04
MO000026738	ESRRA	estrogen related receptor alpha	1.1	1.81
MO000025130	RARG	retinoic acid receptor gamma	1.06	3.88

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: CTNNB1, SMAD4, JUND, SP1, NR3C1 and ZNF462.



3.4. Finding master regulators in networks

In the second step of the upstream analysis common regulators of the revealed TFs were identified. We identified 11 signaling proteins whose structure and function is highly damaged by the mutations (see Table 10).

Table 10. Signaling proteins whose structure and function are damaged by the mutations in Experiment: Squamous Cell Carcinoma and Control: Non-tumour tissue

See full table →

ID	Title	Mutation count	Consequence	Codons
MO000144222	APT2(h)	6	stop_lost	Tag/Cag
MO000208420	GJB3(h)	2	stop_gained	tGg/tAg
MO000036793	ada(h)	1	NMD_transcript_variant,splice_region_variant,stop_gained	tGg/tAg
MO000109306	PSMA4(h)	1	stop_lost	Tga/Cga
MO000119197	wolframin(h)	1	stop_gained	Caa/Taa
MO000172130	c3orf1(h)	1	NMD_transcript_variant,stop_lost	tGa/tCa
MO000175986	oas2(h)	1	stop_lost	tAg/tGg
MO000189841	ZSWIM1(h)	1	stop_gained	tGg/tAg
MO000212738	EMC10(h)	1	stop_lost	taG/taT
MO000219203	PSMG1(h)	1	NMD_transcript_variant,stop_lost	Taa/Caa

Top 11 mutated proteins for Experiment: Squamous Cell Carcinoma and Control: Non-tumour tissue were used in the algorithm of master regulator search as a list of nodes of the signal transduction network that are removed from the network during the search of master regulators (see more details about the algorithm in the Methods section). 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 11-12.

*Table 11. Master regulators that may govern the regulation of **up-regulated** genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. **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	logFC	Total rank
MO0000016677	EGFR(h)	EGFR	epidermal growth factor receptor	4.44	46
MO0000082228	EGFR-p60(h)	EGFR	epidermal growth factor receptor	4.44	52
MO0000082277	EGFR-p170(h)	EGFR	epidermal growth factor receptor	4.44	53
MO0000042551	EGFR(h){pY}	EGFR	epidermal growth factor receptor	4.44	55
MO0000125420	EGFR(h){ub}n	EGFR	epidermal growth factor receptor	4.44	57
MO0000199635	EGFR(h){pY869}{pY1016}{pY1092}	EGFR	epidermal growth factor receptor	4.44	59
MO0000097590	EGFR(h){pY1016}	EGFR	epidermal growth factor receptor	4.44	60
MO0000114808	EGFR(h){pY1197}	EGFR	epidermal growth factor receptor	4.44	60
MO0000325070	EGFR(h){pY1069}{pY1092}{pY1197}	EGFR	epidermal growth factor receptor	4.44	60
MO0000097989	EGFR(h){pY1092}{pY1197}{pY1016}	EGFR	epidermal growth factor receptor	4.44	63

*Table 12. Master regulators that may govern the regulation of **down-regulated** genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. **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	logFC	Total rank
MO0000033396	DUSP5(h)	DUSP5	dual specificity phosphatase 5	-4.8	26
MO0000112152	MRG-1(h)	CITED2	Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2	-3.42	50
MO0000137304	DUSP5(h)	DUSP5	dual specificity phosphatase 5	-4.8	51
MO0000031101	plk3(h)	PLK3	polo like kinase 3	-2.83	54
MO0000138699	plk3(h)	PLK3	polo like kinase 3	-2.83	64
MO0000019174	Eck(h)	EPHA2	EPH receptor A2	-3.32	87
MO0000021356	EGFR(h){pY}	EGFR, ERBB2, ERBB3, ERBB4	epidermal growth factor receptor, erb-b2 receptor tyrosine kinase 2, erb-b2 receptor tyrosine kinase...	-2.6	88
MO0000112151	MRG-1-isoform1(h)	CITED2	Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2	-3.42	92
MO0000035359	MRG-1-isoform2(h)	CITED2	Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2	-3.42	93
MO0000004672	ERK1(h)	MAPK3	mitogen-activated protein kinase 3	-2.26	95

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figures 10 and 11. 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.

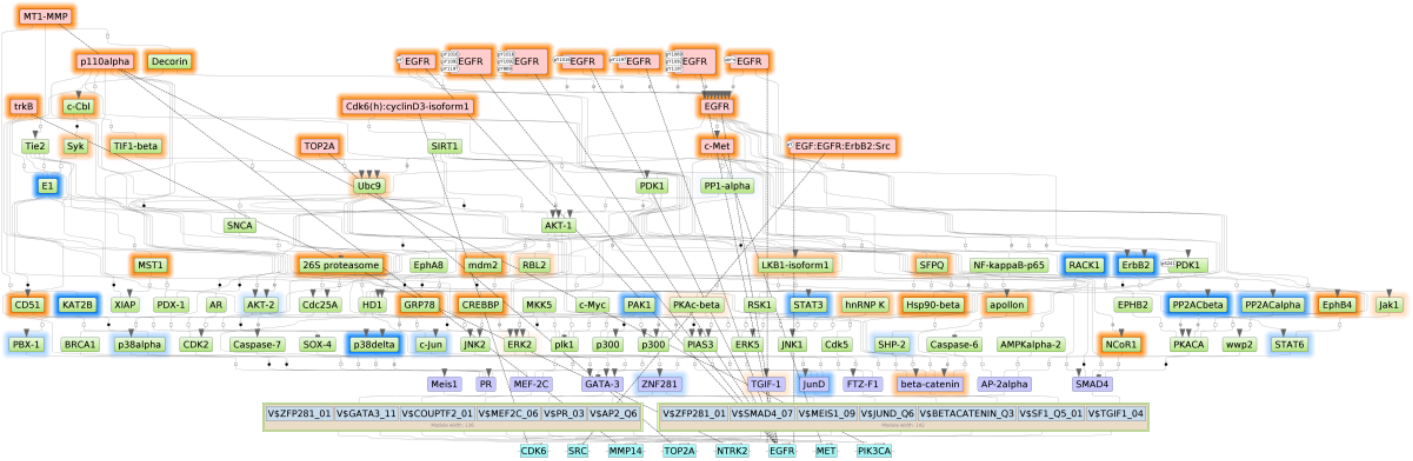


Figure 10. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate 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 →

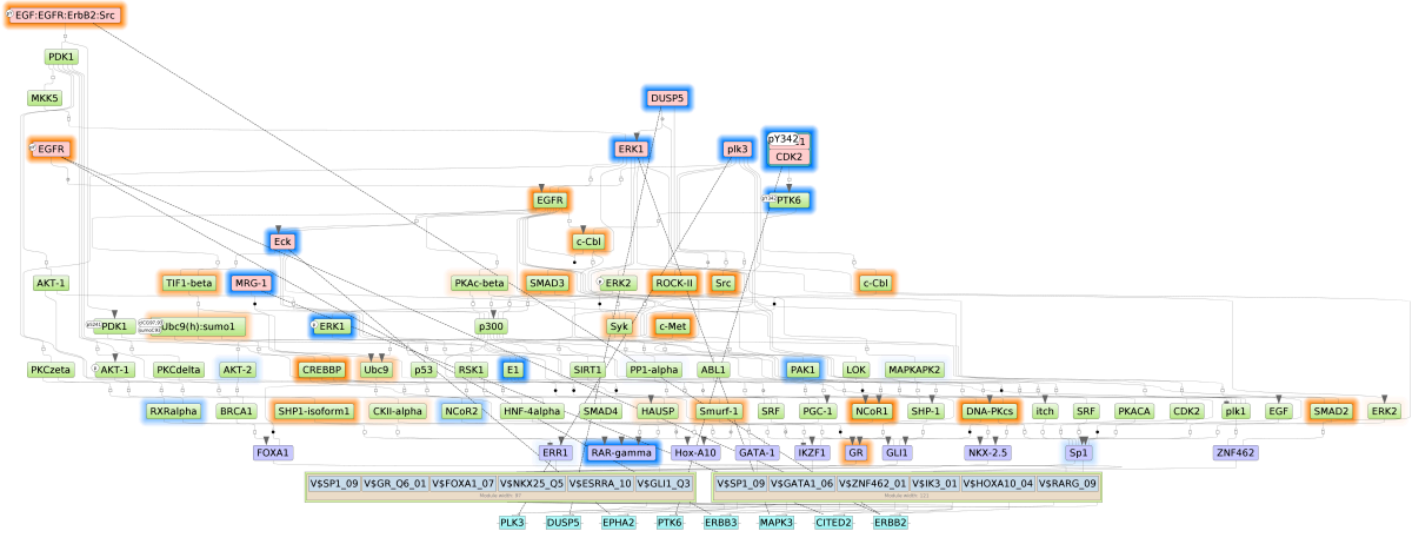


Figure 11. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during the search for master regulators from selected TFs. Orange and blue frames highlight molecules that are encoded by up- and downregulated genes, resp.

See full diagram →

4. Finding prospective drug targets

The identified master regulators that may govern pathology associated genes were checked for druggability potential using HumanPSD™ [11] database of gene-disease-drug assignments and PASS [12-14] software for prediction of biological activities of chemical compounds on the basis of a (Q)SAR approach. Respectively, for each master regulator protein we have computed two Druggability scores: HumanPSD Druggability score and PASS Druggability score. Where Druggability score represents the number of drugs that are potentially suitable for inhibition (or activation) of the corresponding target either according to the information extracted from medical literature (from HumanPSD™ database) or according to cheminformatics predictions of compounds activity against the examined target (from PASS software).

The cheminformatics druggability check is done using a pre-computed database of spectra of biological activities of chemical compounds from a library of all small molecular drugs from HumanPSD™ database, 2507 pharmaceutically active known chemical compounds in total. The spectra of biological activities has been computed using the program PASS [12-14] on the basis of a (Q)SAR approach.

If both Druggability scores were below defined thresholds (see Methods section for the details) such master regulator proteins were not used in further analysis of drug prediction.

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

*Table 13. Prospective drug targets selected from full list of identified master regulators filtered by Druggability score from HumanPSD™ database. **Druggability score** contains the number of drugs that are potentially suitable for inhibition (or activation) of the target. The drug targets are sorted according to the **Total rank** which is the sum of three ranks computed on the basis of the three scores: keynode score, CMA score and expression change score (logFC, if present). See Methods section for details.*

See full table →

Gene symbol	Gene Description	Druggability score	logFC	Total rank
EGFR	epidermal growth factor receptor	95	4.44	17
MET	MET proto-oncogene, receptor tyrosine kinase	55	3.42	35
NTRK2	neurotrophic receptor tyrosine kinase 2	46	5.99	48
ODC1	ornithine decarboxylase 1	4	6.73	62
ITGB1	integrin subunit beta 1	6	2.3	67
ITGAV	integrin subunit alpha V	3	3.04	108

*Table 14. Prospective drug targets selected from full list of identified master regulators filtered by Druggability score predicted by PASS software. Here, the **Druggability score** for master regulator proteins is computed as a sum of PASS calculated probabilities to be active as a target for various small molecular compounds. The drug targets are sorted according to the **Total rank** which is the sum of three ranks computed on the basis of the three scores: keynode score, CMA score and expression change score (logFC, if present). See Methods section for details.*

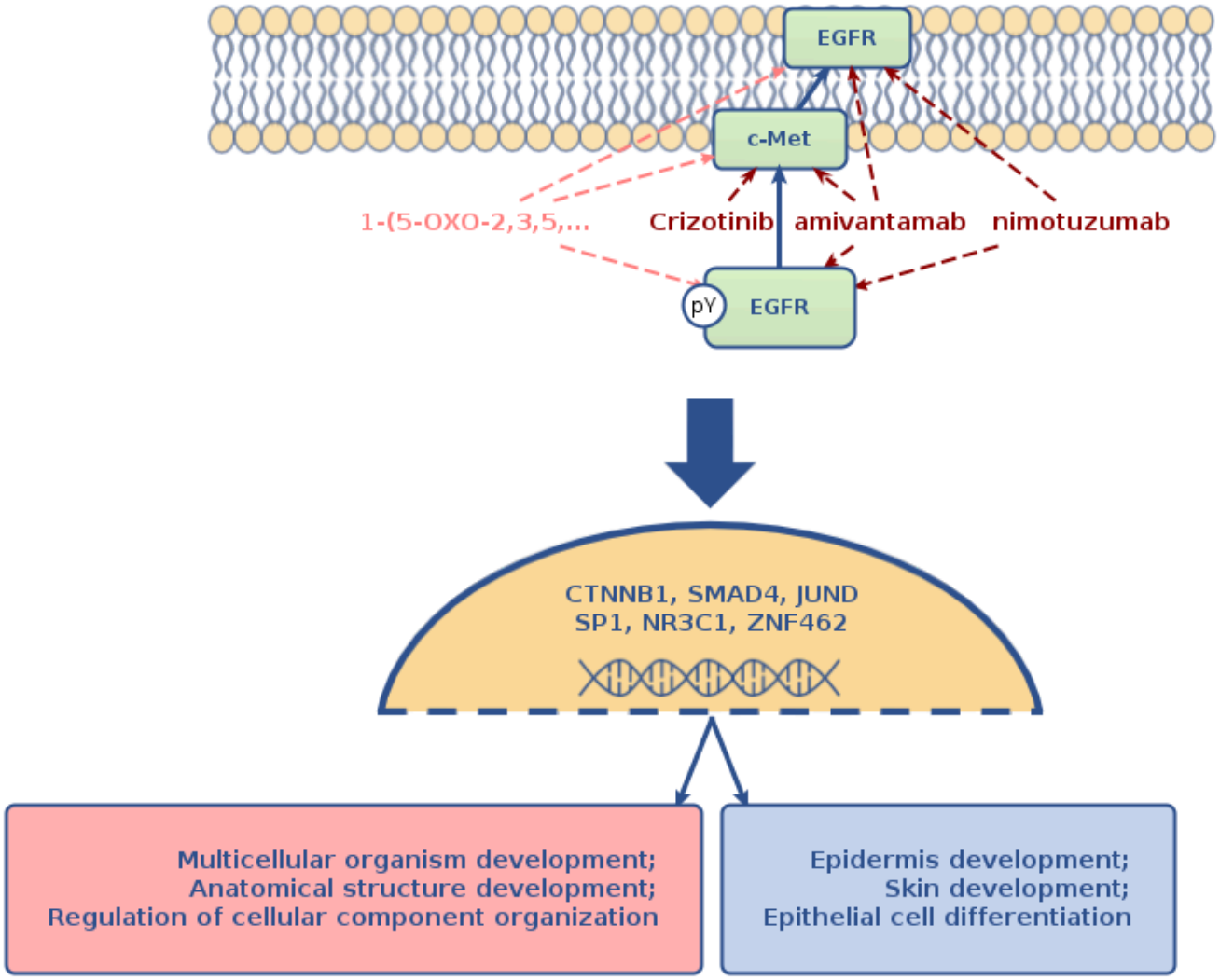
See full table →

Gene symbol	Gene Description	Druggability score	logFC	Total rank
EGFR	epidermal growth factor receptor	38.45	4.44	17
MET	MET proto-oncogene, receptor tyrosine kinase	12.03	3.42	35
NTRK2	neurotrophic receptor tyrosine kinase 2	12.4	5.99	48
ODC1	ornithine decarboxylase 1	0.41	6.73	62
ITGB1	integrin subunit beta 1	10.03	2.3	67
ITGAV	integrin subunit alpha V	7.12	3.04	108

Below we represent schematically the main mechanism of the studied pathology. In the schema we considered the top two drug targets of each of the two categories computed above. In addition we have added two top identified master regulators for which no drugs may be identified yet, but that are playing the crucial role in the molecular mechanism of the studied pathology. Thus the molecular mechanism of the studied pathology was predicted to be mainly based on the following key master regulators:

- EGFR
- c-Met
- EGFR

This result allows us to suggest the following schema of affecting the molecular mechanism of the studied pathology:



Drugs which are shown on this schema: Crizotinib, amivantamab, 1-(5-OXO-2,3,5,9B-TETRAHYDRO-1H-PYRROLO[2,1-A]ISOINDOL-9-YL)-3-(5-PYRROLIDIN-2-YL-1H-PYRAZOL-3-YL)-UREA and nimotuzumab, should be considered as a prospective research initiative for further drug repurposing and drug development. These drugs were selected as top matching treatments to the most prospective drug targets of the studied pathology, however, these results should be considered with special caution and are to be used for research purposes only, as there is not enough clinical information for adapting these results towards immediate treatment of patients.
The drugs given in dark red color on the schema are FDA approved drugs or drugs which have gone through various phases of clinical trials as active treatments against the selected targets.
The drugs given in pink color on the schema are drugs, which were cheminformatically predicted to be active against the selected targets.

5. Identification of potential drugs

In the last step of the analysis we strived to identify known activities as well as drugs with cheminformatically predicted activities that are potentially suitable for inhibition (or activation) of the identified molecular targets in the context of specified human diseases(s).

Proposed drugs are top ranked drug candidates, that were found to be active on the identified targets and were selected from 4 categories:

1. FDA approved drugs or used in clinical trials drugs for the studied pathology;
2. Repurposing drugs used in clinical trials for other pathologies;
3. Drugs, predicted by PASS to be active against identified drug targets and against the studied pathology;
4. Drugs, predicted by PASS to be active against identified drug targets but for other pathologies.

Proposed drugs were selected on the basis of Drug rank which was computed from the ranks sum based on the individual ranks of the following scores:

- Target activity score (depends on ranks of all targets that were found for the selected drug);
- Disease activity score (weighted sum of number of clinical trials on disease(s) under study where the selected drug is known to be applied or PASS Disease activity score - cheminformatically predicted property of the compound to be active against the studied disease(s));
- Clinical validity score (applicable only for drugs predicted on the basis of literature curation in HumanPSD™ database (Tables 16 and 17), reflects the number of the highest clinical trials phase on which the drug was tested for any pathology).

You can refer to the Methods section for more details on drug ranking procedure.

Based on the Drug rank, a numerical value of Drug score was calculated, which reflects the potential activity of the respective drug on the overall molecular mechanism of the studied pathology. Drug score values belong to the range from 1 to 100 and are calculated as a quotient of maximum drug rank and the drug rank of the given drug multiplied by 100.

Top drugs of each category are given in the tables below:

Drugs approved in clinical trials for Oncology

Table 15. Clinically approved (FDA, ENA, etc.) drugs for the studied pathology (most promising and clinically approved treatment candidates selected for the identified drug targets on the basis of literature curation in HumanPSD™ database)
See full table →

Name	Target names	Drug score	Disease activity score	Disease trial phase	Approved
Fluorouracil	BIRC5	19	12	Phase 4: Carcinoma, Squamous Cell, Bowen's Disease, Breast Neoplasms, Carcinoma, Carcinoma, Basal Cell, Colorectal Neoplasms, Dermatitis, Dermatitis, Atopic, Gastrointestinal Neoplasms, Glaucoma, Head and Neck Neoplasms, Intestinal Neoplasms, Keratosis, Keratosis, Actinic, Liver Neoplasms, Neoplasm Metastasis, Neoplasms, Neoplasms by Site, Neoplasms, Basal Cell, Neoplasms, Squamous Cell, Photosensitivity Disorders, Postoperative Complications, Psoriasis, Pterygium, Rectal Neoplasms, Skin Diseases, Skin Neoplasms, Vitiligo	Carcinoma, Squamous Cell (ClinicalTrials, ClinicalTrials, ClinicalTrials)

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

Drugs approved in clinical trials

Table 16. Drugs used in clinical trials for the studied pathology (most promising treatment candidates selected for the identified drug targets on the basis of literature curation in HumanPSD™ database)
See full table →

Name	Target names	Drug score	Disease activity score	Disease trial phase
nimotuzumab	EGFR	99	5	Phase 3: Carcinoma, Squamous Cell, Adenocarcinoma, Carcinoma, Esophageal Neoplasms, Esophageal Squamous Cell Carcinoma, Head and Neck Neoplasms, Nasopharyngeal Carcinoma, Nasopharyngeal Diseases, Nasopharyngeal Neoplasms, Neoplasms, Pharyngeal Diseases, Pharyngeal Neoplasms
Tegafur	ITGA6, EGFR, ITGB1, PTK2, ITGB4, ITGA3	95	2	Phase 2: Carcinoma, Squamous Cell, Adenocarcinoma, Biliary Tract Neoplasms, Breast Neoplasms, Colorectal Neoplasms, Esophageal Neoplasms, Esophageal Squamous Cell Carcinoma, Head and Neck Neoplasms, Nasopharyngeal Carcinoma, Nasopharyngeal Neoplasms, Neoplasm Metastasis, Neoplasms, Pancreatic Neoplasms, Rectal Neoplasms, Stomach Neoplasms
Panitumumab	EGFR	95	2	Phase 2: Carcinoma, Squamous Cell, Adenocarcinoma, Adenoma, Adenoma, Pleomorphic, Brain Neoplasms, Carcinoma, Carcinoma, Non-Small-Cell Lung, Cholangiocarcinoma, Colonic Neoplasms, Colorectal Neoplasms, Esophageal Neoplasms, Head and Neck Neoplasms, Histiocytosis, Lung Neoplasms, Neoplasm Metastasis, Neoplasms, Rectal Neoplasms, Recurrence, Squamous Cell Carcinoma of Head and Neck, Thyroid Neoplasms
icotinib	EGFR	95	2	Phase 2: Carcinoma, Squamous Cell, Adenocarcinoma, Adenocarcinoma of Lung, Bronchial Neoplasms, Carcinoma, Non-Small-Cell Lung, Esophageal Neoplasms, Esophageal Squamous Cell Carcinoma, Nasopharyngeal Carcinoma, Nasopharyngeal Neoplasms
Crizotinib	MET, SRC, NTRK2, EPHB4, CSF1R	94	5	Phase 3: Carcinoma, Squamous Cell, Adenocarcinoma, Brain Abscess, Carcinoma, Carcinoma, Large Cell, Carcinoma, Non-Small-Cell Lung, Lung Neoplasms, Multiple Endocrine Neoplasia, Multiple Endocrine Neoplasia Type 2a, Neoplasms

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

Repurposing drugs

Table 17. Repurposed drugs used in clinical trials for other pathologies (prospective drugs against the identified drug targets on the basis of literature curation in HumanPSD™ database)
See full table →

Name	Target names	Drug score	Maximum trial phase
amivantamab	EGFR, MET	90	Phase 3: Carcinoma, Non-Small-Cell Lung
sotrastaurin	ROCK2, EGFR, MET, PTK2, PIK3CA, EPHB4, PAK2, GSK3B	90	Phase 2: Edema, Macular Edema, Melanoma, Panuveitis, Uveitis, Uveitis, Posterior
falnidamol	EGFR	90	N/A
olmutinib	EGFR	90	N/A
hematoxylin	EGFR, MET, SRC	89	NULL: Colorectal Neoplasms, Neoplasms, Rectal Neoplasms

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

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

Table 18. Prospective drugs, predicted by PASS software to be active against the identified drug targets, though without cheminformatically predicted activity against the studied disease(s) (drug candidates predicted with the cheminformatics tool PASS)
See full table →

Name	Target names	Drug score	Target activity score
1-(5-OXO-2,3,5,9B-TETRAHYDRO-1H-PYRROLO[2,1-A]ISOINDOL-9-YL)-3-(5-PYRROLIDIN-2-YL-1H-PYRAZOL-3-YL)-UREA	EGFR, CDK6, MET, NTRK2, PAK2, EPHB4, RYK, CDK4, CSF1R	100	0.96
Carmustine	GSR, TXNRD1, TOP2A	100	0.53
CEP-1347	EGFR, MET, PRKD3, NTRK2, EPHB4, RYK, CSF1R	100	0.52
(2S)-1-{4-[(4-ANILINO-5-BROMOPYRIMIDIN-2-YL)AMINO]PHENOXY}-3-(DIMETHYLAMINO)PROPAN-2-OL	EGFR, CDK6, MET, NTRK2, PTK2, EPHB4, PAK2, RYK, CDK4, CSF1R	100	0.52
3-Bromo-7-Nitroindazole	CDK6, HSPD1, GSK3B, CDK4	99	0.49

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: nimotuzumab, amivantamab and 1-(5-OXO-2,3,5,9B-TETRAHYDRO-1H-PYRROLO[2,1-A]ISOINDOL-9-YL)-3-(5-PYRROLIDIN-2-YL-1H-PYRAZOL-3-YL)-UREA. These drugs were selected for acting on the following targets: EGFR, which were predicted to be active in the molecular mechanism of the studied pathology.

The selected drugs are top ranked drug candidates from each of the four categories of drugs: (1) FDA approved drugs or used in clinical trials drugs for the studied pathology; (2) repurposing drugs used in clinical trials for other pathologies; (3) drugs, predicted by PASS software to be active against the studied pathology; (4) drugs, predicted by PASS software to be repurposed from other pathologies.

Supplementary drug info

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

Table 19. Supplementary drug info: extended list of drugs used for treatment of neoplasms with respective drug scores predicted for the studied pathology.

Drug	Disease	Predicted Drug Score
Abarelix	Prostatic Neoplasms	-
abemaciclib	Breast Neoplasms	90
Abiraterone	Prostatic Neoplasms, Castration-Resistant	-
abiraterone acetate	Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	-
acalabrutinib	Lymphoma, Mantle-Cell	-
Acetaminophen	Acute Pain|Arthritis|Back Pain|Brain Injuries|Common Cold|Dysmenorrhea|Fever|Headache|Hernia, Abdominal|Hypertension|Infections|Influenza, Human|Meningitis, Bacterial|Migraine Disorders|Muscle Cramp|Nausea|Neoplasms|Obesity|Obesity, Morbid|Osteoarthritis|Osteoarthritis, Knee|Pain|Pharyngitis|Postoperative Nausea and Vomiting|Streptococcal Infections|Thrombotic Stroke|Tonsillitis|Toothache|Vomiting|Wounds and Injuries	-
Acetylsalicylic acid	Acute Coronary Syndrome|Arthritis|Asthma|Atherosclerosis|Atrial Fibrillation|Back Pain|Cardiovascular Diseases|Cerebral Infarction|Common Cold|Coronary Artery Disease|Coronary Disease|Diabetes Mellitus|Diabetes Mellitus, Type 2|Dysmenorrhea|Fever|Headache|Heart Diseases|Heart Failure|Hypertension|Internal Hernia|Ischemia|Migraine Disorders|Muscle Cramp|Myalgia|Myocardial Infarction|Myocardial Ischemia|Neoplasms|Pain|Peptic Ulcer|Periodontitis|Placental Insufficiency|Pre-Eclampsia|Scleroderma, Diffuse|Scleroderma, Systemic|Sclerosis|Spasm|ST Elevation Myocardial Infarction|Stroke|Thrombosis|Toothache|Ulcer|Vascular Diseases	12
ado-trastuzumab emtansine	Breast Neoplasms	24
Afatinib	Carcinoma, Non-Small-Cell Lung	83

afatinib dimaleate	Carcinoma, Non-Small-Cell Lung	-
Aflibercept	Colorectal Neoplasms|Macular Degeneration|Macular Edema|Retinal Detachment|Retinal Diseases|Retinal Vein Occlusion|Wet Macular Degeneration	-
Aldesleukin	Carcinoma|Carcinoma, Renal Cell|Melanoma	-
Alectinib	Carcinoma, Non-Small-Cell Lung	-
Alemtuzumab	Leukemia, Lymphocytic, Chronic, B-Cell|Nerve Degeneration	-
Alfuzosin	Prostatic Hyperplasia	-
Alitretinoin	Sarcoma, Kaposi	-
Allopurinol	Bipolar Disorder|Colitis-Associated Neoplasms|Diabetes Mellitus|Gout|Heart Failure|Hypertension|Hyperuricemia|Internal Hernia|Ischemia|Leukemia|Lymphoma|Neoplasms|Response Evaluation Criteria in Solid Tumors|Stroke	-
alpelisib	Breast Neoplasms	84
Altretamine	Ovarian Neoplasms	-
Amifostine	Head and Neck Neoplasms|Neoplasms|Ovarian Neoplasms|Xerostomia	31
Aminoglutethimide	Neoplasms	-
amivantamab	Carcinoma, Non-Small-Cell Lung	90
Amphotericin B	Aspergillosis|Candidiasis|Candidiasis, Invasive|Cryptococcosis|Histoplasmosis|Infections|Leishmaniasis|Leishmaniasis, Visceral|Leukemia|Meningitis, Cryptococcal|Mucormycosis|Mycoses|Sporotrichosis|Zygomycosis	-
Amsacrine	Neoplasms	55
Anastrozole	Breast Neoplasms|Endometriosis|Hypogonadism|Neoplasms	-
apalutamide	Prostatic Neoplasms, Castration-Resistant	-
Aprepitant	Breast Neoplasms|Nausea|Neoplasms|Postoperative Nausea and Vomiting|Vomiting	-
Aripiprazole	Alcoholism|Alzheimer Disease|Anxiety|Attention Deficit Disorder with Hyperactivity|Autistic Disorder|Bipolar Disorder|Child Development Disorders, Pervasive|Colitis-Associated Neoplasms|Depressive Disorder|Depressive Disorder, Major|Psychotic Disorders|Schizophrenia|Tourette Syndrome|Weight Loss	-
Arsenic trioxide	Leukemia, Promyelocytic, Acute	59
arzoxifene	Endometrial Neoplasms	-
asciminib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	-
Asparaginase Erwinia chrysanthemi	Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
atezolizumab	Carcinoma, Non-Small-Cell Lung|Carcinoma, Transitional Cell|Triple Negative Breast Neoplasms	-
avapritinib	Gastrointestinal Stromal Tumors|Neoplasms	-
avelumab	Carcinoma, Merkel Cell|Carcinoma, Renal Cell|Carcinoma, Transitional Cell	-
Avobenzone	Melanosis|Skin Neoplasms|Sunburn	-
Axitinib	Carcinoma, Renal Cell	-
Azacitidine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes	-
Azithromycin	Asthma|Blepharitis|Bronchiectasis|Bronchiolitis|Bronchiolitis Obliterans|Bronchopulmonary Dysplasia|Chlamydia Infections|COVID-19|Cystic Fibrosis|Diarrhea|Endometritis|Eye Diseases|Fever|Gonorrhea|Infections|Lung Diseases|Malnutrition|Mycobacterium avium-intracellulare Infection|Otitis Media|Pharyngitis|Pneumonia|Premature Birth|Pulmonary Disease, Chronic Obstructive|Sepsis|Sinusitis|Skin Diseases, Infectious|Tonsillitis|Urethritis|Uterine Cervicitis|Yaws	-
belantamab mafodotin	Multiple Myeloma|Neoplasms	-
Belinostat	Lymphoma, T-Cell, Peripheral	22
belzutifan	Carcinoma, Renal Cell|Hemangioblastoma|Neuroendocrine Tumors|von Hippel-Lindau Disease	-
Bendamustine	Leukemia|Leukemia, Lymphocytic, Chronic, B-Cell	-
Bevacizumab	Colorectal Neoplasms	-
Bexarotene	Lymphoma|Lymphoma, T-Cell, Cutaneous	-
Bicalutamide	Neoplasms|Prostatic Neoplasms	-
binimetinib	Melanoma	-
Blinatumomab	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Bortezomib	Multiple Myeloma|Neoplasms|Renal Insufficiency, Chronic	1
Bosutinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	40
Brentuximab vedotin	Hodgkin Disease|Lymphoma, Large-Cell, Anaplastic|Lymphoma, T-Cell, Peripheral	-
brigatinib	Carcinoma, Non-Small-Cell Lung	79

Bupivacaine	Anesthesia|Ankle Fractures|Appendicitis|Brain Neoplasms|Dwarfism|Fractures, Bone|Hallux Valgus|Headache|Hemorrhage|Hemorrhoids|Hernia|Hernia, Inguinal|Hypotension|Migraine Disorders|Neoplasms|Neuralgia|Obesity|Obesity, Morbid|Opioid-Related Disorders|Osteoarthritis|Osteoarthritis, Hip|Osteoarthritis, Knee|Pain|Rib Fractures|Rotator Cuff Injuries|Thoracic Injuries|Thrombotic Stroke|Wounds and Injuries	-
Bupropion	Alcoholism|Attention Deficit Disorder with Hyperactivity|Depressive Disorder|Depressive Disorder, Major|Lung Neoplasms|Mood Disorders|Neoplasms|Schizophrenia|Substance-Related Disorders|Tobacco Use Disorder	-
Buserelin	Prostatic Neoplasms	-
Busulfan	Leukemia|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Neoplasms	-
Cabazitaxel	Neoplasms|Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	26
Cabergoline	Infertility|Infertility, Female|Pituitary Neoplasms	-
Cabozantinib	Thyroid Neoplasms	76
Capecitabine	Adenocarcinoma|Breast Neoplasms|Colonic Neoplasms|Colorectal Neoplasms|Neoplasms|Rectal Neoplasms	-
capivasertib	Breast Neoplasms	82
capmatinib	Carcinoma, Non-Small-Cell Lung	88
Carbamazepine	Bipolar Disorder|Cocaine-Related Disorders|Colitis-Associated Neoplasms|Depressive Disorder|Epilepsies, Partial|Epilepsy|Glossopharyngeal Nerve Diseases|Seizures|Trigeminal Neuralgia	-
Carboplatin	Breast Neoplasms|Ovarian Neoplasms	-
Carfilzomib	Brain Abscess|Multiple Myeloma|Neoplasms|Neoplasms, Plasma Cell	20
Carmustine	Astrocytoma|Ependymoma|Glioblastoma|Glioma|Hodgkin Disease|Medulloblastoma|Multiple Myeloma|Neoplasms	28
Celecoxib	Arthritis, Rheumatoid|Hypertension|Inflammation|Low Back Pain|Neoplasms|Osteoarthritis|Osteoarthritis, Hip|Osteoarthritis, Knee|Rectal Neoplasms|Spondylitis, Ankylosing|Wounds and Injuries	-
cemiplimab	Carcinoma, Basal Cell|Carcinoma, Squamous Cell|Lung Neoplasms|Neoplasms|Skin Neoplasms	-
Ceritinib	Carcinoma, Non-Small-Cell Lung	63
Cetuximab	Colorectal Neoplasms	72
Chlorambucil	Hodgkin Disease|Neoplasms|Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
Chlorotrianisene	Carcinoma, Hepatocellular	-
Choline C 11	Prostatic Neoplasms	-
Cinacalcet	Hyperparathyroidism, Primary|Hyperparathyroidism, Secondary|Parathyroid Neoplasms|Renal Insufficiency, Chronic	-
Cladribine	Brain Abscess|Leukemia|Leukemia, Hairy Cell|Multiple Sclerosis|Sclerosis	-
Clarithromycin	Bacterial Infections|Bronchitis, Chronic|Duodenal Ulcer|Helicobacter Infections|Infections|Mycobacterium avium-intracellulare Infection|Neoplasms|Otitis Media|Periodontitis|Pharyngitis|Pneumonia|Sinusitis|Skin Diseases, Bacterial|Stomach Neoplasms|Tonsillitis|Tuberculosis	-
Clofarabine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
Clonidine	Attention Deficit Disorder with Hyperactivity|Cardiovascular Diseases|Glaucoma|Hypertension|Migraine Disorders|Neoplasms|Neuralgia	-
Cobimetinib	Melanoma	-
copanlisib	Lymphoma, Follicular	64
Cortisone acetate	Addison Disease|Adrenal Hyperplasia, Congenital|Anemia, Aplastic|Anemia, Hemolytic|Arthritis, Gouty|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Asthma|Bursitis|Dermatitis Herpetiformis|Dermatitis, Atopic|Dermatitis, Contact|Dermatitis, Exfoliative|Dermatitis, Seborrheic|Dermatomyositis|Enteritis|Erythema Multiforme|Eye Diseases|Herpes Zoster Ophthalmicus|Hypercalcemia|Hypersensitivity|Iridocyclitis|Iritis|Leukemia|Lupus Erythematosus, Systemic|Lymphoma|Nephrotic Syndrome|Ophthalmia, Sympathetic|Osteoarthritis|Peptic Ulcer|Pneumonia|Polymyositis|Purpura, Thrombocytopenic, Idiopathic|Stevens-Johnson Syndrome|Synovitis|Tennis Elbow|Tenosynovitis|Thrombocytopenia|Thyroiditis, Subacute|Tuberculosis, Meningeal|Tuberculosis, Pulmonary|Uveitis, Posterior	-
CP-4055	Leukemia, Myeloid, Acute	-
Crizotinib	Carcinoma, Non-Small-Cell Lung	94
Cyclophosphamide	Breast Neoplasms|Hodgkin Disease|Lupus Nephritis|Lymphoma|Lymphoma, Non-Hodgkin|Multiple Myeloma|Neoplasms|Nephrotic Syndrome	-
Cyclosporine	Acute Kidney Injury|Anemia|Anemia, Aplastic|Arthritis, Rheumatoid|Conjunctivitis, Allergic|COVID-19|Dermatitis, Atopic|Diabetes Mellitus|Dry Eye Syndromes|Eczema|Eye Diseases|Graft vs Host Disease|Immune System	14

	Diseases|Infections|Inflammation|Keratoconjunctivitis|Keratoconjunctivitis Sicca|Leukemia|Meibomian Gland Dysfunction|Myelodysplastic Syndromes|Neoplasms|Psoriasis|Pterygium|Renal Insufficiency|Renal Insufficiency, Chronic|Stevens-Johnson Syndrome|Uveitis	
Cyproterone acetate	Prostatic Neoplasms	-
Cytarabine	Leukemia|Leukemia, Lymphoid|Leukemia, Myeloid, Acute|Leukemia, Promyelocytic, Acute|Meningeal Carcinomatosis|Neoplasms|Precursor Cell Lymphoblastic Leukemia-Lymphoma	60
Dabrafenib	Melanoma	-
Dacarbazine	Hodgkin Disease|Melanoma|Neoplasms	-
dacomitinib	Carcinoma, Non-Small-Cell Lung	83
Dactinomycin	Neoplasms|Wilms Tumor	46
Daratumumab	Multiple Myeloma	-
darolutamide	Neoplasms|Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	-
Dasatinib	Leukemia|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid|Precursor Cell Lymphoblastic Leukemia-Lymphoma	79
Daunorubicin	Leukemia|Leukemia, Lymphoid|Leukemia, Myeloid, Acute|Neoplasms	50
Decitabine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes	-
Degarelix	Neoplasms|Prostatic Neoplasms	-
Dehydroepiandrosterone	Arthritis, Rheumatoid|Dysmenorrhea|Hypogonadism|Infertility|Klinefelter Syndrome|Neoplasms|Orchitis|Osteoarthritis|Pain	38
Denileukin diftitox	Lymphoma, T-Cell, Cutaneous	-
Dexamethasone	Acne Vulgaris|Addison Disease|Adrenal Hyperplasia, Congenital|Adrenal Insufficiency|Adrenocortical Hyperfunction|Anemia, Aplastic|Anemia, Hemolytic|Arthritis, Gouty|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Asthma|Brain Edema|Bursitis|Cataract|Chorioretinitis|Choroiditis|Colitis, Ulcerative|Conjunctivitis, Allergic|Corneal Injuries|COVID-19|Crohn Disease|Dermatitis|Dermatitis Herpetiformis|Dermatitis, Atopic|Dermatitis, Contact|Dermatitis, Exfoliative|Dermatitis, Seborrheic|Diabetic Retinopathy|Erythema Multiforme|Eye Diseases|Hemorrhoids|Herpes Simplex|Herpes Zoster Ophthalmicus|Hypercalcemia|Infections|Inflammation|Iridocyclitis|Iritis|Keratitis|Laryngeal Edema|Leukemia|Lupus Erythematosus, Systemic|Lymphoma|Macular Edema|Multiple Myeloma|Multiple Sclerosis|Mycosis Fungoides|Myocarditis|Nasal Obstruction|Nausea|Neoplasms|Nephrotic Syndrome|Ophthalmia, Sympathetic|Optic Neuritis|Osteoarthritis|Pemphigus|Peptic Ulcer|Pneumonia|Postoperative Nausea and Vomiting|Psoriasis|Pulmonary Eosinophilia|Purpura, Thrombocytopenic, Idiopathic|Respiratory Insufficiency|Retinal Vein Occlusion|Rhinitis, Allergic|Rhinitis, Allergic, Seasonal|Sarcoidosis|Serum Sickness|Spondylitis, Ankylosing|Stevens-Johnson Syndrome|Synovitis|Tennis Elbow|Tenosynovitis|Thrombocytopenia|Thyroiditis, Subacute|Transfusion Reaction|Trichinellosis|Tuberculosis, Meningeal|Tuberculosis, Pulmonary|Uveitis, Intermediate|Uveitis, Posterior|Vomiting	19
dexamethasone sodium phosphate	Acne Vulgaris|Addison Disease|Adrenal Hyperplasia, Congenital|Adrenal Insufficiency|Anemia, Aplastic|Anemia, Hemolytic|Arthritis, Gouty|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Asthma|Brain Edema|Bursitis|Chorioretinitis|Choroiditis|Colitis, Ulcerative|Conjunctivitis, Allergic|Corneal Injuries|Crohn Disease|Dermatitis|Dermatitis Herpetiformis|Dermatitis, Atopic|Dermatitis, Contact|Dermatitis, Exfoliative|Dermatitis, Seborrheic|Erythema Multiforme|Eye Diseases|Hemorrhoids|Herpes Simplex|Herpes Zoster Ophthalmicus|Hypercalcemia|Inflammation|Iridocyclitis|Iritis|Keratitis|Laryngeal Edema|Leukemia|Lupus Erythematosus, Systemic|Lymphoma|Macular Edema|Multiple Myeloma|Mycosis Fungoides|Myocarditis|Nasal Obstruction|Neoplasms|Nephrotic Syndrome|Ophthalmia, Sympathetic|Optic Neuritis|Osteoarthritis|Pemphigus|Peptic Ulcer|Pneumonia|Psoriasis|Pulmonary Eosinophilia|Purpura, Thrombocytopenic, Idiopathic|Rhinitis, Allergic|Rhinitis, Allergic, Seasonal|Sarcoidosis|Serum Sickness|Spondylitis, Ankylosing|Stevens-Johnson Syndrome|Synovitis|Tennis Elbow|Tenosynovitis|Thrombocytopenia|Thyroiditis, Subacute|Transfusion Reaction|Trichinellosis|Tuberculosis, Meningeal|Tuberculosis, Pulmonary|Uveitis, Posterior	-
Dexmedetomidine	Acute Kidney Injury|Brain Neoplasms|Cognitive Dysfunction|Cough|Delirium|Heart Diseases|Heart Septal Defects, Atrial|Hypertension|Inflammation|Neoplasms|Pain|Postoperative Cognitive Complications|Psychomotor Agitation|Psychotic Disorders|Respiratory Insufficiency|Shock, Septic|Stomach Neoplasms|Wounds and Injuries	-
Dexrazoxane	Breast Neoplasms|Cardiomyopathies	50
Docetaxel	Breast Neoplasms|Carcinoma|Carcinoma, Non-Small-Cell Lung|Carcinoma, Squamous Cell|Head and Neck Neoplasms|Lung Neoplasms|Neoplasms|Prostatic Neoplasms|Squamous Cell Carcinoma of Head and Neck|Stomach Neoplasms	-

Dolasetron	Nausea|Neoplasms|Postoperative Nausea and Vomiting	-
Donepezil	Alzheimer Disease|Amblyopia|Aphasia|Cognitive Dysfunction|Colitis-Associated Neoplasms|Delirium|Dementia|Dementia, Vascular|Multiple Sclerosis|Parkinson Disease|Stroke	14
dostarlimab	Endometrial Neoplasms|Neoplasms	-
Doxazosin	Hypertension|Prostatic Hyperplasia|Stress Disorders, Post-Traumatic	-
Doxercalciferol	Hyperparathyroidism, Secondary|Kidney Diseases|Neoplasms|Renal Insufficiency, Chronic	6
doxifluridine	Breast Neoplasms|Colonic Neoplasms|Neoplasms|Stomach Neoplasms	-
Doxorubicin	Breast Neoplasms|Carcinoma|Carcinoma, Hepatocellular|Carcinoma, Ovarian Epithelial|Heart Failure|Hodgkin Disease|Leukemia, Myelomonocytic, Acute|Lymphoma|Lymphoma, Non-Hodgkin|Multiple Myeloma|Neoplasms|Osteosarcoma|Ovarian Neoplasms|Sarcoma|Sarcoma, Kaposi|Urinary Bladder Neoplasms	82
durvalumab	Carcinoma, Non-Small-Cell Lung|Carcinoma, Transitional Cell	-
Dutasteride	Hypogonadism|Prostatic Hyperplasia|Prostatic Neoplasms	-
duvelisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	-
elacestrant	Breast Neoplasms	-
Elotuzumab	Multiple Myeloma	-
elranatamab	Multiple Myeloma	-
enasidenib	Leukemia, Myeloid, Acute	-
encorafenib	Colorectal Neoplasms|Melanoma	-
enfortumab vedotin	Carcinoma, Transitional Cell|Neoplasms	-
entrectinib	Carcinoma, Non-Small-Cell Lung	57
Enzalutamide	Neoplasms|Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	-
epcoritamab	Lymphoma, Large B-Cell, Diffuse	-
Epirubicin	Breast Neoplasms	66
erdafitinib	Urinary Bladder Neoplasms	30
Ergocalciferol	Bone Diseases, Metabolic|Crohn Disease|Cysts|Diabetes Mellitus, Type 2|Fractures, Bone|Heart Failure|Hip Fractures|Hypoparathyroidism|Hypophosphatemia, Familial|Insulin Resistance|Muscular Diseases|Neoplasms|Osteoporosis|Polycystic Ovary Syndrome|Renal Insufficiency, Chronic|Rickets, Hypophosphatemic|Syndrome|Vitamin D Deficiency	-
Eribulin	Breast Neoplasms|Neurotoxicity Syndromes	-
Erlotinib	Carcinoma, Non-Small-Cell Lung|Neoplasms	91
Esomeprazole	Arthritis, Rheumatoid|Croup|Digestive System Diseases|Duodenal Ulcer|Dyspepsia|Esophagitis|Esophagitis, Peptic|Gastritis|Gastroesophageal Reflux|Gastrointestinal Diseases|Helicobacter Infections|Hemorrhage|Intestinal Diseases|Neoplasms|Osteoarthritis|Peptic Ulcer|Spondylitis, Ankylosing|Stomach Ulcer|Ulcer|Zollinger-Ellison Syndrome	-
Estradiol	Atrophic Vaginitis|Atrophy|Breast Neoplasms|Dyspareunia|Hypogonadism|Menopause|Osteoporosis|Osteoporosis, Postmenopausal|Primary Ovarian Insufficiency	7
Estramustine	Prostatic Neoplasms	10
estramustine phosphate	Carcinoma|Neoplasms	-
Ethinyl Estradiol	Acne Vulgaris|Neoplasms	-
Etoposide	Carcinoma, Small Cell|Lung Neoplasms|Neoplasms|Small Cell Lung Carcinoma	18
Everolimus	Angiomyolipoma|Astrocytoma|Breast Neoplasms|Carcinoma|Carcinoma, Renal Cell|Kidney Diseases|Neoplasms|Neuroendocrine Tumors|Polycystic Kidney Diseases|Polycystic Kidney, Autosomal Dominant|Renal Insufficiency|Renal Insufficiency, Chronic|Tuberous Sclerosis	10
Exemestane	Breast Neoplasms|Neoplasms	-
Fentanyl	Acute Pain|Chronic Pain|Hemorrhage|Hyperalgesia|Low Back Pain|Migraine Disorders|Neoplasms|Neuralgia|Pain|Respiratory Insufficiency	-
Finasteride	Prostatic Hyperplasia|Prostatic Neoplasms	-
Flavopiridol	Leukemia, Lymphocytic, Chronic, B-Cell	90
Floxuridine	Neoplasms	-
Fludarabine	Leukemia, Lymphocytic, Chronic, B-Cell	-
fluoroestradiol f-18	Breast Neoplasms	-
Fluorouracil	Adenocarcinoma|Carcinoma|Carcinoma, Basal Cell|Carcinoma, Squamous Cell|Colonic Neoplasms|Colorectal Neoplasms|Esophageal Neoplasms|Glaucoma|Keratosis|Keratosis, Actinic|Neoplasms|Rectal Neoplasms	19
Fluoxymesterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	-
Flutamide	Prostatic Neoplasms	43
Formestane	Neoplasms	-

Fosaprepitant	Neoplasms|Postoperative Nausea and Vomiting	-
fruquintinib	Colorectal Neoplasms	-
Fulvestrant	Breast Neoplasms|Neoplasms	-
futibatinib	Cholangiocarcinoma	-
Gefitinib	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms	86
Gemcitabine	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Neoplasms|Ovarian Neoplasms|Pancreatic Neoplasms	-
Gemtuzumab ozogamicin	Leukemia, Myeloid, Acute	-
gilteritinib	Leukemia, Myeloid, Acute	53
glasdegib	Leukemia, Myeloid, Acute	60
glofitamab	Lymphoma, Large B-Cell, Diffuse	-
Goserelin	Breast Neoplasms|Endometriosis|Infertility|Leiomyoma|Neoplasms|Prostatic Neoplasms	-
Guanidine	Carcinoma, Small Cell|Lambert-Eaton Myasthenic Syndrome|Muscle Weakness	-
Hexaminolevulinate	Carcinoma	-
histrelin	Neoplasms	-
Homoharringtonine	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	47
Hydrocortisone	Adrenal Hyperplasia, Congenital|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Asthma|Bursitis|Colitis, Ulcerative|Conjunctivitis, Allergic|Crohn Disease|Dermatitis|Dermatitis, Atopic|Dermatitis, Seborrheic|Eczema|Endocrine System Diseases|Exanthema|Eye Diseases|Hemorrhoids|Herpes Labialis|Inflammation|Intertrigo|Leukemia|Lymphoma|Nasal Obstruction|Neoplasms|Osteoarthritis|Otitis Externa|Pneumonia|Pruritus|Psoriasis|Serum Sickness|Shock, Septic|Skin Diseases|Spondylitis|Tenosynovitis|Wounds and Injuries	4
Hydromorphone	Acute Pain|Low Back Pain|Neoplasms|Pain|Substance-Related Disorders	-
hydroxyflutamide	Carcinoma|Neoplasms|Prostatic Neoplasms	-
Hydroxyurea	Anemia, Sickle Cell|Carcinoma, Squamous Cell|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Melanoma|Neoplasms	-
Ibandronate	Bone Diseases|Bone Diseases, Metabolic|Bone Neoplasms|Neoplasm Metastasis|Osteoporosis|Osteoporosis, Postmenopausal	-
Ibritumomab	Lymphoma, B-Cell|Lymphoma, Follicular	-
Ibrutinib	Graft vs Host Disease|Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, B-Cell, Marginal Zone|Lymphoma, Mantle-Cell|Waldenstrom Macroglobulinemia	40
icotinib	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms	95
Idarubicin	Leukemia, Myeloid, Acute	55
Idelalisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	-
Ifosfamide	Neoplasms|Testicular Neoplasms	-
Imatinib	Gastrointestinal Stromal Tumors|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Neoplasms	83
imatinib mesylate	Breast Neoplasms|Gastrointestinal Stromal Tumors|Leukemia|Leukemia, Myelogenous, Chronic, BCR-ABL Positive	-
Imiquimod	Carcinoma, Basal Cell|Cheilitis|Condylomata Acuminata|Keratosis, Actinic|Lentigo	-
infigratinib	Cholangiocarcinoma	-
inositol	Colitis-Associated Neoplasms	-
inotuzumab ozogamicin	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Interferon Alfa-2a, Recombinant	Leukemia, Hairy Cell	-
Interferon Alfa-2b, Recombinant	Leukemia, Hairy Cell	-
Iopamidol	Glioma|Pituitary Neoplasms	-
ipilimumab	Carcinoma, Renal Cell|Melanoma	-
Irinotecan	Colorectal Neoplasms	56
isatuximab	Multiple Myeloma|Neoplasms	-
Isopropyl Alcohol	Acromegaly|Anemia|Arthritis, Psoriatic|Arthritis, Rheumatoid|Burns|Colitis, Ulcerative|Constipation|Eczema|Infections|Multiple Sclerosis|Prostatic Neoplasms|Psoriasis|Spondylitis, Ankylosing	-
ivosidenib	Leukemia, Myeloid, Acute	-
Ixabepilone	Breast Neoplasms	-
Ixazomib	Multiple Myeloma	-
Ketamine	Delirium|Depression, Postpartum|Depressive Disorder|Depressive Disorder, Major|Fractures, Bone|Hemorrhage|Lung Neoplasms|Neoplasms|Opioid-Related Disorders|Pain|Postoperative Complications|Suicidal Ideation	-
Lamotrigine	Bipolar Disorder|Cocaine-Related Disorders|Colitis-Associated Neoplasms|Depressive Disorder|Epilepsies, Partial|Epilepsy|Epilepsy, Tonic-Clonic|Mental Disorders|Mood Disorders|Seizures	-

Lanreotide	Acromegaly|Neoplasms	-
Lapatinib	Breast Neoplasms	88
larotrectinib	Neoplasm Metastasis	71
LE-SN38	Adenocarcinoma|Carcinoma|Neoplasms	-
Lenalidomide	Multiple Myeloma|Myelodysplastic Syndromes|Neoplasms	-
Lenvatinib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	-
Letrozole	Anovulation|Breast Neoplasms|Cardiac Complexes, Premature|Infertility|Leiomyoma|Neoplasms|Polycystic Ovary Syndrome|Pregnancy, Ectopic	-
Leuprolide	Endometriosis|Infertility|Neoplasms|Prostatic Neoplasms|Puberty, Precocious	-
Levamisole	Colonic Neoplasms|Helminthiasis	-
Levetiracetam	Alcoholism|Bipolar Disorder|Brain Neoplasms|Colitis-Associated Neoplasms|Epilepsies, Partial|Epilepsy|Parkinson Disease|Polyneuropathies|Seizures	-
Levobupivacaine	Muscle Weakness|Neoplasms|Paresis	-
levoleucovorin	Anemia|Colorectal Neoplasms|Osteosarcoma	-
Levonorgestrel	Endometrial Hyperplasia|Endometriosis|Hemorrhage|Menorrhagia	-
Levothyroxine	Congenital Hypothyroidism|Hypothyroidism|Thyroid Neoplasms	-
Lidocaine	Anesthesia|Arrhythmias, Cardiac|Arthritis|Back Pain|Burns|Bursitis|Hemorrhoids|Inflammation|Insect Bites and Stings|Muscle Cramp|Myalgia|Neuralgia|Neuralgia, Postherpetic|Osteoarthritis|Pain|Premature Ejaculation|Prostatic Neoplasms|Pruritus|Rectal Diseases|Sprains and Strains|Sunburn|Tendinopathy|Urethritis	53
lidocaine patch	Breast Neoplasms|Lung Diseases|Pneumonia	-
Liothyronine	Depressive Disorder|Hypothyroidism|Myxedema|Thyroid Neoplasms	-
lipegfilgrastim	Neoplasms	-
Lomustine	Brain Neoplasms|Hodgkin Disease	-
loncastuximab tesirine	Lymphoma, B-Cell|Lymphoma, Large B-Cell, Diffuse	-
lorlatinib	Carcinoma, Non-Small-Cell Lung	88
lurbinectedin	Neoplasms|Small Cell Lung Carcinoma	-
margetuximab	Breast Neoplasms	-
Mechlorethamine	Lymphoma, T-Cell, Cutaneous|Neoplasms	-
Medroxyprogesterone Acetate	Carcinoma, Renal Cell|Endometriosis|Hemorrhage|Hyperplasia|Infertility|Neoplasms|Uterine Hemorrhage	-
Megestrol acetate	Acquired Immunodeficiency Syndrome|Anorexia|Breast Neoplasms|Cachexia	-
Melphalan	Carcinoma|Multiple Myeloma|Neoplasms	-
melphalan flufenamide	Multiple Myeloma	51
Mercaptopurine	Leukemia|Leukemia, Lymphoid|Lymphoma|Neoplasms|Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
Mesna	Neoplasms|Thrombocytopenia	-
Methadone	Acute Pain|Heroin Dependence|HIV Infections|Neonatal Abstinence Syndrome|Neoplasms|Neuralgia|Opioid-Related Disorders|Pain|Scoliosis|Substance-Related Disorders	-
Methotrexate	Adenoma|Arthritis|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Breast Neoplasms|Choriocarcinoma|Communicable Diseases|Crohn Disease|Gout|Immune System Diseases|Lung Neoplasms|Lymphoma|Lymphoma, Non-Hodgkin|Meningeal Carcinomatosis|Multiple Sclerosis|Mycosis Fungoides|Neoplasms|Osteoarthritis|Precursor Cell Lymphoblastic Leukemia-Lymphoma|Pregnancy, Ectopic|Psoriasis|Uveitis	13
Methoxsalen	Lymphoma|Lymphoma, T-Cell|Mycosis Fungoides|Psoriasis	-
Methyl aminolevulinate	Glioma|Keratosis, Actinic|Neoplasms	-
Methylphenidate	Alcoholism|Alzheimer Disease|Attention Deficit Disorder with Hyperactivity|Bipolar Disorder|Brain Injuries|Brain Neoplasms|Cocaine-Related Disorders|Dementia|Depressive Disorder|Narcolepsy|Neoplasms|Parkinson Disease|Substance-Related Disorders	-
Methylprednisolone	Acne Vulgaris|Anemia|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Asthma|Bursitis|Colitis, Ulcerative|Crohn Disease|Dermatitis|Dermatitis, Atopic|Endocrine System Diseases|Fasciitis, Plantar|Inflammation|Leukemia|Lymphoma|Multiple Sclerosis|Multiple Sclerosis, Relapsing-Remitting|Osteoarthritis|Pneumonia|Purpura, Thrombocytopenic, Idiopathic|Serum Sickness|Tenosynovitis|Uveitis, Posterior	-
Methyltestosterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	-
Metoclopramide	Anorexia|Digestive System Diseases|Gastroparesis|Heartburn|Nausea|Neoplasms|Postoperative Nausea and Vomiting|Renal Insufficiency|Vomiting	-
midostaurin	Leukemia, Mast-Cell|Leukemia, Myeloid, Acute|Mastocytosis, Systemic	78

mirvetuximab soravtansine	Ovarian Neoplasms	-
Mitomycin	Adenocarcinoma|Glaucoma|Glaucoma, Neovascular|Glaucoma, Open-Angle|Myopia|Neoplasms	-
Mitotane	Adrenocortical Carcinoma	-
Mitoxantrone	Leukemia, Myeloid, Acute|Multiple Sclerosis|Prostatic Neoplasms, Castration-Resistant	33
mobocertinib	Carcinoma, Non-Small-Cell Lung	87
mogamulizumab	Mycosis Fungoides|Neoplasms|Sezary Syndrome	-
Morphine	Abdominal Neoplasms|Acute Pain|Chronic Pain|Dyspnea|Low Back Pain|Mucositis|Myocardial Infarction|Neoplasms|Pain|Pruritus|Pulmonary Disease, Chronic Obstructive|Scoliosis|Stomatitis|Substance-Related Disorders	-
mosunetuzumab	Lymphoma, Follicular	-
motixafortide	Multiple Myeloma	-
moxetumomab pasudotox	Leukemia, Hairy Cell|Neoplasms	-
mrtx-849	Carcinoma, Non-Small-Cell Lung	-
nabilone	Nausea|Neoplasms|Postoperative Nausea and Vomiting	-
Necitumumab	Carcinoma, Non-Small-Cell Lung|Neoplasms	-
Nelarabine	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma	-
neratinib	Breast Neoplasms	53
neratinib maleate	Breast Neoplasms	-
Nilotinib	Leukemia|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid	7
Nilutamide	Prostatic Neoplasms	-
niraparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms	64
Nivolumab	Melanoma	-
nogapendekin alfa	Urinary Bladder Neoplasms	-
Norepinephrine	Carcinoma|Carcinoma, Hepatocellular|Cardiovascular Diseases|Depressive Disorder|Hypotension|Hypotension, Orthostatic|Multiple System Atrophy|Poliomyelitis|Sepsis|Shock|Shock, Septic	-
Obinutuzumab	Leukemia, Lymphocytic, Chronic, B-Cell	-
Octreotide	Acromegaly|Carcinoid Tumor|Diarrhea|Neoplasms|Obesity|Vipoma	-
Ofatumumab	Leukemia, Lymphocytic, Chronic, B-Cell	-
Olaparib	Breast Neoplasms|Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Neoplasms|Ovarian Neoplasms|Pancreatic Neoplasms|Peritoneal Neoplasms|Prostatic Neoplasms, Castration-Resistant	51
olaratumab	Sarcoma	-
Ondansetron	Alcoholism|Breast Neoplasms|Gastroenteritis|Hypotension|Irritable Bowel Syndrome|Nausea|Neoplasms|Obsessive-Compulsive Disorder|Postoperative Nausea and Vomiting|Vomiting	12
Osimertinib	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms	52
Oxaliplatin	Colorectal Neoplasms|Liver Neoplasms	-
Oxycodone	Low Back Pain|Neoplasms|Pain|Substance-Related Disorders	-
Paclitaxel	Adenocarcinoma|Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Constriction, Pathologic|Coronary Restenosis|Ischemia|Neoplasms|Ovarian Neoplasms|Peripheral Arterial Disease|Peripheral Vascular Diseases	60
Palbociclib	Breast Neoplasms	79
palifosfamide	Neoplasms|Testicular Neoplasms	-
Palonosetron	Nausea|Neoplasms|Postoperative Nausea and Vomiting|Vomiting	-
Pamidronate	Bone Diseases|Breast Neoplasms|Hypercalcemia|Osteitis Deformans|Osteogenesis Imperfecta	-
Panitumumab	Colorectal Neoplasms	95
Panobinostat	Multiple Myeloma	-
Papaverine	Digestive System Diseases|Erectile Dysfunction|Kidney Neoplasms|Neoplasms	-
Paricalcitol	Diabetic Nephropathies|Hyperparathyroidism, Secondary|Kidney Diseases|Neoplasms|Renal Insufficiency, Chronic	-
Pazopanib	Carcinoma, Renal Cell|Sarcoma	80
pazopanib hydrochloride	Carcinoma, Renal Cell|Sarcoma	-
Pegaspargase	Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
Pembrolizumab	Carcinoma, Hepatocellular|Carcinoma, Merkel Cell|Carcinoma, Non-Small-Cell Lung|Carcinoma, Renal Cell|Carcinoma, Transitional Cell|Hodgkin Disease|Melanoma|Neoplasms|Stomach Neoplasms	-
Pemetrexed	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Mesothelioma|Neoplasms|Renal Insufficiency	-
pemigatinib	Cholangiocarcinoma|Neoplasms	-

Pentostatin	Leukemia, Hairy Cell	-
Pertuzumab	Breast Neoplasms	-
Pipobroman	Neoplasms	-
pirtobrutinib	Lymphoma, Mantle-Cell	-
Pixantrone	Neoplasms	69
Plerixafor	Lymphoma|Lymphoma, Non-Hodgkin|Multiple Myeloma	-
Podofilox	Carcinoma, Squamous Cell|Condylomata Acuminata|Virus Diseases	50
polatuzumab vedotin	Lymphoma, Large B-Cell, Diffuse|Neoplasms	-
Pomalidomide	Multiple Myeloma	-
Ponatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Precursor Cell Lymphoblastic Leukemia-Lymphoma	65
Porfimer	Esophageal Neoplasms	-
Posaconazole	Aspergillosis|Candidiasis|Leukemia|Leukemia, Myeloid, Acute|Mycoses|Neutropenia	-
Potassium Citrate	Acidosis, Renal Tubular|Kidney Calculi|Neoplasms|Stomach Neoplasms	-
Povidone-iodine	Breast Neoplasms|Surgical Wound|Surgical Wound Infection|Wound Infection|Wounds and Injuries	-
Pralatrexate	Lymphoma, T-Cell, Peripheral	-
pralsetinib	Carcinoma, Medullary|Carcinoma, Non-Small-Cell Lung	-
Prednicarbate	Adrenal Hyperplasia, Congenital|Anemia|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Brain Neoplasms|Bursitis|Conjunctivitis, Allergic|Dermatitis, Atopic|Endocrine System Diseases|Inflammation|Multiple Sclerosis|Neoplasms|Osteoarthritis|Peptic Ulcer|Rhinitis, Allergic|Skin Diseases|Spondylitis|Tenosynovitis|Uveitis, Posterior	-
Prednisolone	Addison Disease|Adrenal Hyperplasia, Congenital|Anemia|Anemia, Aplastic|Anemia, Hemolytic|Arthritis|Arthritis, Gouty|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Asthma|Berylliosis|Brain Neoplasms|Bursitis|Chorioretinitis|Choroiditis|Colitis, Ulcerative|Conjunctivitis, Allergic|Crohn Disease|Dermatitis Herpetiformis|Dermatitis, Atopic|Dermatitis, Contact|Dermatitis, Exfoliative|Dermatitis, Seborrheic|Dermatomyositis|Endocrine System Diseases|Erythema Multiforme|Eye Diseases|Hemorrhoids|Herpes Zoster Ophthalmicus|Hypercalcemia|Inflammation|Iridocyclitis|Iritis|Keratitis|Leukemia|Lung Diseases|Lupus Erythematosus, Systemic|Lupus Nephritis|Lymphoma|Multiple Sclerosis|Mycosis Fungoides|Myocarditis|Nasal Obstruction|Neoplasms|Nephrotic Syndrome|Ophthalmia, Sympathetic|Optic Neuritis|Osteoarthritis|Pain|Pemphigus|Peptic Ulcer|Pneumonia|Polymyositis|Psoriasis|Pulmonary Disease, Chronic Obstructive|Pulmonary Eosinophilia|Pulpitis|Purpura, Thrombocytopenic, Idiopathic|Rhinitis, Allergic|Rhinitis, Allergic, Seasonal|Sarcoidosis|Serum Sickness|Sinusitis|Skin Diseases|Spondylitis|Spondylitis, Ankylosing|Stevens-Johnson Syndrome|Synovitis|Tennis Elbow|Tenosynovitis|Thrombocytopenia|Thyroiditis, Subacute|Trichinellosis|Tuberculosis, Meningeal|Tuberculosis, Pulmonary|Uveitis|Uveitis, Posterior	-
prednisolone phosphoric acid	Anemia|Asthma|Conjunctivitis, Allergic|Corneal Injuries|Idiopathic Pulmonary Fibrosis|Inflammation|Leukemia|Multiple Sclerosis|Nephrotic Syndrome|Pulmonary Disease, Chronic Obstructive|Pulmonary Fibrosis|Rhinitis, Allergic|Uremia|Uveitis	-
Prednisone	Addison Disease|Adrenal Hyperplasia, Congenital|Anemia, Aplastic|Anemia, Hemolytic|Arthritis|Arthritis, Gouty|Arthritis, Juvenile|Arthritis, Psoriatic|Arthritis, Rheumatoid|Asthma|Berylliosis|Bursitis|Chorioretinitis|Choroiditis|Colitis, Ulcerative|Conjunctivitis, Allergic|Crohn Disease|Dermatitis Herpetiformis|Dermatitis, Atopic|Dermatitis, Contact|Dermatitis, Exfoliative|Dermatitis, Seborrheic|Dermatomyositis|Erythema Multiforme|Giant Cell Arteritis|Heart Failure|Hematologic Diseases|Herpes Zoster Ophthalmicus|Hypercalcemia|Hypereosinophilic Syndrome|Infections|Iridocyclitis|Iritis|Keratitis|Leukemia|Lung Diseases|Lupus Erythematosus, Systemic|Lymphoma|Multiple Sclerosis|Mycosis Fungoides|Myocarditis|Nephrotic Syndrome|Ophthalmia, Sympathetic|Optic Neuritis|Osteoarthritis|Pemphigus|Pneumonia|Polymyalgia Rheumatica|Polymyositis|Psoriasis|Pulmonary Disease, Chronic Obstructive|Pulmonary Eosinophilia|Purpura, Thrombocytopenic, Idiopathic|Respiratory Tract Infections|Rhinitis, Allergic, Seasonal|Sarcoidosis|Serum Sickness|Spondylitis, Ankylosing|Stevens-Johnson Syndrome|Synovitis|Tennis Elbow|Tenosynovitis|Thrombocytopenia|Thyroiditis, Subacute|Trichinellosis|Tuberculosis, Meningeal|Tuberculosis, Pulmonary|Uveitis, Posterior|Vestibular Neuronitis	-
Procarbazine	Neoplasms	-
Progesterone	Amenorrhea|Endometrial Hyperplasia|Hemorrhage|Infertility|Premature Birth|Uterine Hemorrhage	47
Propofol	Breast Neoplasms|Coronary Artery Disease|Heart Diseases|Inflammation|Neoplasms|Obesity, Morbid|Psychomotor Agitation|Respiratory	7

	Insufficiency|Scoliosis|Wounds and Injuries	
Quetiapine	Alcoholism|Anxiety|Bipolar Disorder|Cocaine-Related Disorders|Colitis-Associated Neoplasms|Delirium|Dementia|Depressive Disorder|Depressive Disorder, Major|Mental Disorders|Mood Disorders|Obsessive-Compulsive Disorder|Parkinson Disease|Phobia, Social|Psychomotor Agitation|Psychotic Disorders|Schizophrenia|Sleep Initiation and Maintenance Disorders|Stress Disorders, Post-Traumatic	-
quizartinib	Leukemia, Myeloid, Acute	13
Radium Ra 223 Dichloride	Neoplasm Metastasis|Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	-
Raloxifene	Breast Neoplasms|Osteoporosis|Osteoporosis, Postmenopausal|Psychotic Disorders|Schizophrenia	28
Raltitrexed	Neoplasms	-
Ramucirumab	Stomach Neoplasms	-
Rasburicase	Leukemia|Lymphoma|Neoplasms	-
Regorafenib	Colorectal Neoplasms	-
relugolix	Prostatic Neoplasms	-
Remifentanil	Hypotension|Neoplasms|Respiratory Insufficiency	-
repotrectinib	Carcinoma, Non-Small-Cell Lung	71
retifanlimab	Carcinoma, Merkel Cell	-
Ribavirin	Fibrosis|Hepatitis|Hepatitis C|Hepatitis C, Chronic|HIV Infections|Infections|Kidney Diseases|Liver Cirrhosis|Neoplasms|Renal Insufficiency|Renal Insufficiency, Chronic|Respiratory Syncytial Virus Infections|Virus Diseases	-
ribociclib	Breast Neoplasms	85
ripretinib	Gastrointestinal Stromal Tumors|Neoplasms	-
Risperidone	Aggression|Alzheimer Disease|Anxiety|Attention Deficit and Disruptive Behavior Disorders|Autistic Disorder|Bipolar Disorder|Child Development Disorders, Pervasive|Colitis-Associated Neoplasms|Dementia|Developmental Disabilities|Psychotic Disorders|Schizophrenia|Stress Disorders, Post-Traumatic	-
Rituximab	Anti-Neutrophil Cytoplasmic Antibody-Associated Vasculitis|Lymphoma, B-Cell|Lymphoma, Follicular|Vasculitis	-
rivoceranib mesylate	Lymphoma|Neoplasms|Ovarian Neoplasms|Stomach Neoplasms	-
Romidepsin	Lymphoma, T-Cell, Cutaneous	-
Ropivacaine	Acute Pain|Appendicitis|Breast Neoplasms|Fractures, Bone|Hallux Valgus|Joint Diseases|Muscle Weakness|Neoplasms|Osteoarthritis|Osteoarthritis, Knee|Pain|Paralysis|Paresis|Rotator Cuff Injuries	-
Rosuvastatin	Acute Coronary Syndrome|Atherosclerosis|Cardiovascular Diseases|Coronary Artery Disease|Coronary Disease|Diabetes Mellitus|Diabetes Mellitus, Type 2|Dyslipidemias|Heart Failure|HIV Infections|Hypercholesterolemia|Hyperlipidemia, Familial Combined|Hyperlipidemias|Hyperlipoproteinemia Type II|Hyperlipoproteinemia Type IV|Hypertension|Hypertriglyceridemia|Internal Hernia|Ischemia|Metabolic Diseases|Neoplasms|Periodontitis|Prostatic Neoplasms|Renal Insufficiency|Stroke	-
rucaparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms|Prostatic Neoplasms, Castration-Resistant	20
sacituzumab govitecan	Breast Neoplasms|Triple Negative Breast Neoplasms	-
Salicylic acid	Acne Vulgaris|Callosities|Dandruff|Dermatitis, Seborrheic|Eye Diseases|Fever|Inflammation|Melanosis|Pain|Pruritus|Psoriasis|Skin Diseases|Skin Neoplasms|Sunburn|Warts	-
Sargramostim	COVID-19|Leukemia, Myeloid, Acute	-
selinexor	Multiple Myeloma	45
selpercatinib	Carcinoma, Medullary|Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Thyroid Neoplasms	-
Sildenafil	Diabetes Mellitus|Emphysema|Erectile Dysfunction|Familial Primary Pulmonary Hypertension|Heart Diseases|Heart Failure|Hypertension|Hypertension, Pulmonary|Lung Diseases|Parkinson Disease|Persistent Fetal Circulation Syndrome|Prostatic Hyperplasia|Pulmonary Arterial Hypertension|Pulmonary Edema	-
Silodosin	Prostatic Hyperplasia	-
Siltuximab	Giant Lymph Node Hyperplasia	-
Sodium Chloride	Acid-Base Imbalance|Cerebral Palsy|Common Cold|Corneal Edema|Diabetes Mellitus|Dry Eye Syndromes|Electrolytes|Eye Diseases|Fractures, Bone|Graves Ophthalmopathy|Heart Failure|Hemorrhage|Hypothermia|Kidney Diseases|Muscle Spasticity|Neoplasms|Osteoarthritis|Osteoarthritis, Knee|Paralysis|Renal Insufficiency|Tennis Elbow|Wounds and Injuries	-
Solifenacin	Prostatic Hyperplasia|Urinary Bladder, Overactive|Urinary Incontinence|Urinary Incontinence, Urge	-
Sonidegib	Carcinoma, Basal Cell	49

Sorafenib	Carcinoma|Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	82
sotorasib	Carcinoma, Non-Small-Cell Lung	-
Strontium chloride Sr-89	Neoplasm Metastasis	-
Sugammadex	Hypothermia|Neoplasms|Postoperative Complications	-
Sunitinib	Adenoma, Islet Cell|Carcinoma|Carcinoma, Renal Cell|Gastrointestinal Stromal Tumors|Neoplasms|Neuroendocrine Tumors	78
Tadalafil	Diabetes Mellitus, Type 2|Erectile Dysfunction|Hypertension|Hypertension, Pulmonary|Insulin Resistance|Lower Urinary Tract Symptoms|Muscular Dystrophies|Muscular Dystrophy, Duchenne|Prostatic Hyperplasia|Pulmonary Arterial Hypertension|Scleroderma, Systemic	-
tafasitamab	Lymphoma|Lymphoma, Large B-Cell, Diffuse|Neoplasms	-
talazoparib	Breast Neoplasms	38
talquetamab	Multiple Myeloma	-
Tamoxifen	Adenoma|Breast Neoplasms|Neoplasms	18
Tamsulosin	Hyperplasia|Nocturia|Prostatic Hyperplasia|Ureterolithiasis|Urinary Retention|Urination Disorders	-
tarlatamab	Small Cell Lung Carcinoma	-
tazemetostat	Lymphoma, Follicular|Sarcoma	-
tebentafusp	Uveal Neoplasms	-
teclistamab	Multiple Myeloma	-
Tegafur	Neoplasms	95
Temozolomide	Astrocytoma|Brain Neoplasms|Central Nervous System Neoplasms|Glioblastoma|Neoplasms	11
Temsirolimus	Carcinoma, Renal Cell|Lymphoma|Lymphoma, Mantle-Cell|Lymphoma, Non-Hodgkin	14
Teniposide	Precursor Cell Lymphoblastic Leukemia-Lymphoma	37
tepotinib	Carcinoma, Non-Small-Cell Lung|Neoplasms	53
Terazosin	Hypertension|Prostatic Hyperplasia	-
Testosterone	Brain Injuries|Brain Injuries, Traumatic|Erectile Dysfunction|Hypogonadism|Insulin Resistance|Klinefelter Syndrome|Neoplasms|Obesity|Orchitis|Sarcopenia|Wounds and Injuries	1
testosterone cypionate	Hypogonadism|Neoplasms|Orchitis	-
testosterone undecanoate	Hypogonadism|Klinefelter Syndrome|Neoplasms|Orchitis	-
Thalidomide	Immune System Diseases|Multiple Myeloma|Neoplasms	-
Thiotepa	Adenocarcinoma|Carcinoma|Hodgkin Disease|Neoplasms	-
thyrotropin alfa	Thyroid Neoplasms	-
Tioguanine	Neoplasms	-
Tipiracil	Colorectal Neoplasms	-
tirabrutinib	Lymphoma	-
tisagenlecleucel	Neoplasms	-
tislelizumab	Esophageal Squamous Cell Carcinoma	-
tisotumab vedotin	Uterine Cervical Neoplasms	-
tivozanib	Carcinoma, Renal Cell	-
Topotecan	Small Cell Lung Carcinoma	-
Toremifene	Breast Neoplasms|Neoplasms	-
toripalimab	Nasopharyngeal Carcinoma	-
Tositumomab	Lymphoma, Follicular	-
Trabectedin	Leiomyosarcoma|Liposarcoma	-
Trametinib	Carcinoma, Non-Small-Cell Lung|Melanoma	24
Tranexamic Acid	Blood Coagulation Disorders|Epistaxis|Femoral Neck Fractures|Fractures, Bone|Gastrointestinal Hemorrhage|Hematoma|Hematoma, Subdural|Hematoma, Subdural, Chronic|Hemophilia A|Hemorrhage|Hip Fractures|Joint Diseases|Menorrhagia|Neoplasms|Osteoarthritis|Osteoarthritis, Knee|Placenta Accreta|Placenta Previa|Postpartum Hemorrhage|Wounds and Injuries	-
Trastuzumab	Breast Neoplasms	61
trastuzumab deruxtecan	Breast Neoplasms|Neoplasms	54
tremelimumab	Carcinoma, Hepatocellular	-
treosulfan	Neoplasms	-
Triclosan	Breast Neoplasms|Gingivitis|Periodontal Diseases|Periodontitis	30
Trifluridine	Colorectal Neoplasms|Eye Infections|Herpes Simplex|Keratoconjunctivitis|Neoplasms	-
trilaciclib	Small Cell Lung Carcinoma	84
Triptorelin	Endometriosis|Infertility|Neoplasms|Prostatic Neoplasms	22

tucatinib	Breast Neoplasms	-
umbralisib	Lymphoma, B-Cell, Marginal Zone|Lymphoma, Follicular	1
Uracil mustard	Neoplasms	-
Valganciclovir	Acquired Immunodeficiency Syndrome|Cytomegalovirus Infections|Cytomegalovirus Retinitis|Infections|Leukemia, Lymphocytic, Chronic, B-Cell|Retinitis|Virus Diseases	-
Valproic Acid	Alzheimer Disease|Anorexia|Autistic Disorder|Bipolar Disorder|Colitis-Associated Neoplasms|Depressive Disorder|Depressive Disorder, Major|Epilepsy|Epilepsy, Absence|Epilepsy, Complex Partial|Migraine Disorders|Mitochondrial Diseases|Mood Disorders|Schizophrenia|Seizures|Vomiting	-
Valrubicin	Urinary Bladder Neoplasms	69
Vandetanib	Thyroid Neoplasms	84
Vemurafenib	Melanoma|Neoplasms	-
venetoclax	Leukemia, Lymphocytic, Chronic, B-Cell|Leukemia, Myeloid, Acute	-
Vincristine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	-
vincristine	Neoplasms	-
vinflunine	Neoplasms	-
Vinorelbine	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms	-
vipivotide tetraxetan lutetium lu-177	Prostatic Neoplasms, Castration-Resistant	-
Vismodegib	Carcinoma|Carcinoma, Basal Cell	49
Vorinostat	Lymphoma, T-Cell, Cutaneous	13
zanubrutinib	Lymphoma, Mantle-Cell	47
Ziprasidone	Anxiety|Bipolar Disorder|Colitis-Associated Neoplasms|Depressive Disorder|Depressive Disorder, Major|Mood Disorders|Psychotic Disorders|Schizophrenia	-
Zoledronate	Bone Diseases|Bone Diseases, Metabolic|Bone Neoplasms|Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Carcinoma, Renal Cell|Multiple Myeloma|Neoplasm Metastasis|Osteitis Deformans|Osteoarthritis, Hip|Osteoporosis|Osteoporosis, Postmenopausal|Pain|Prostatic Neoplasms|Spinal Cord Injuries	-

6. Conclusion

We applied the software package "Genome Enhancer" to a data set that contains *transcriptomics* data. The study is done in the context of *Squamous Cell Carcinoma*. The data were pre-processed, statistically analyzed and differentially expressed genes were identified. Also checked was the enrichment of GO or disease categories among the studied gene sets.

We propose the following drugs as most promising candidates for treating the pathology under study:

nimotuzumab, amivantamab and 1-(5-OXO-2,3,5,9B-TETRAHYDRO-1H-PYRROLO[2,1-A]ISOINDOL-9-YL)-3-(5-PYRROLIDIN-2-YL-1H-PYRAZOL-3-YL)-UREA

These drugs were selected for acting on the following targets: EGFR, which were predicted to be involved in the molecular mechanism of the pathology under study.

The identified molecular mechanism of the studied pathology was predicted to be mainly based on the following key drug targets:

EGFR, c-Met and EGFR

These potential drug targets should be considered as a prospective research initiative for further drug repurposing and drug development purposes. The following drugs were predicted as, matching those drug targets: Crizotinib, amivantamab, 1-(5-OXO-2,3,5,9B-TETRAHYDRO-1H-PYRROLO[2,1-A]ISOINDOL-9-YL)-3-(5-PYRROLIDIN-2-YL-1H-PYRAZOL-3-YL)-UREA and nimotuzumab. These drugs should be considered with special caution for research purposes only.

In this study, we came up with a detailed signal transduction network regulating differentially expressed genes in the studied pathology. In this network we have revealed the following top master regulators (signaling proteins and their complexes) that play a crucial role in the molecular mechanism of the studied pathology, which can be proposed as the most promising molecular targets for further drug repurposing and drug development initiatives.

- EGFR
- c-Met
- EGFR

Potential drug compounds which can be affecting these targets can be found in the "Finding prospective drug targets" section.

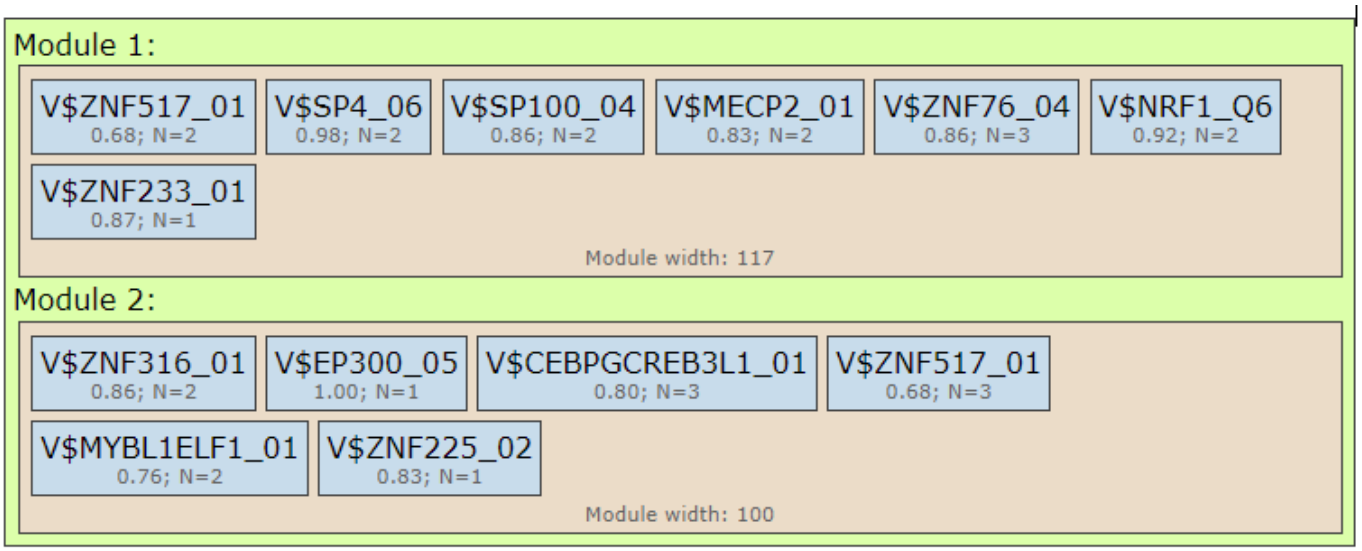
7. Methods

Databases used in the study

Transcription factor binding sites in promoters and enhancers of differentially expressed genes were analyzed using known DNA-binding motifs described in the TRANSFAC® library, release 2026.1 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac).
The master regulator search uses the TRANSPATH® database (BIOBASE), release 2026.1 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transpath). A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in TRANSPATH®.
The information about drugs corresponding to identified drug targets and clinical trials references were extracted from HumanPSD™ database, release 2026.1 (https://genexplain.com/humanpsd).
The Ensembl database release Human112.38 (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.

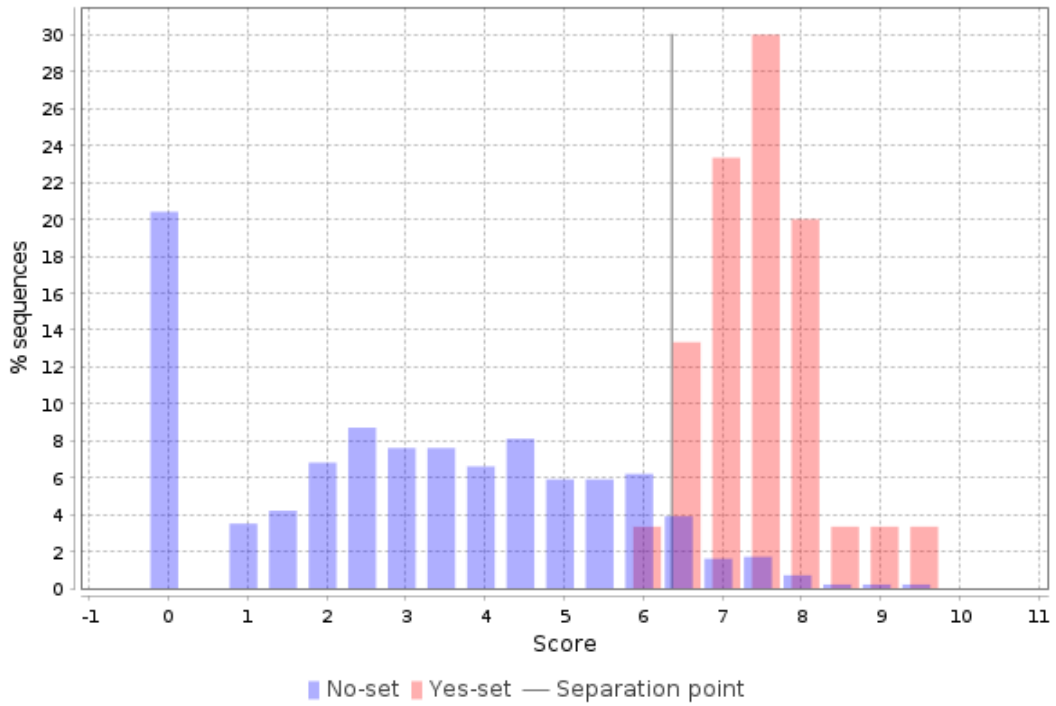
Methods for the analysis of enriched transcription factor binding sites and composite modules

Transcription factor binding sites in promoters and enhancers of disease affected 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. The matrix profile used to identify potential transcription factor binding sites is a collection of 3390 TRANSFAC® vertebrate matrices carefully selected for the purpose of enrichment analysis.
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 length 1100bp (-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 [6] 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).
CMA constructs a generalized model of the regulatory regions of the studied genes by specifying combinations of potential TFBSs that are most frequently clustered together. CMA identifies the transcription factors that through their cooperation are able to provide a synergistic effect and, thus, are likely to have a great influence on the gene regulation process.
Composite modules can have a complex hierarchical structure consisting of two levels: site models and modules. The highest hierarchical level contains two modules and corresponds to the whole promoter model. The first level, site model, corresponds to the individual site model, based on one PWM. Names of the site models are the same as the matrix names (site models are taken from the profile that was used in the site search). In the resulting schema (see example below) the site models are shown by blue boxes. Within these boxes, there are two values below each site model. The first value is the threshold value for the score of the respective site model, which is determined by the genetic algorithm during the optimization process. The second value is the maximum number of best individual matches (sites) found for this site model and taken into account for calculating the score of the module.
Promoter model example:



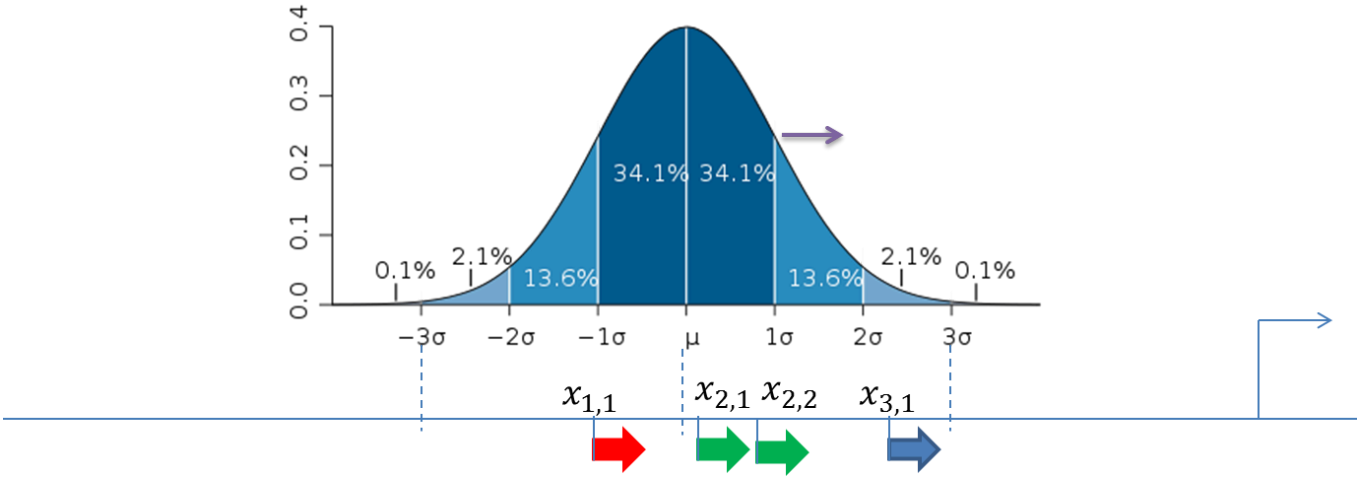
The next level, module, may contain several site models, shown within the light brown boxes. The module is characterized by its width, the average length of DNA window containing matches for the mentioned site models. In the resulting schemas modules are shown in green boxes, and they are numbered: Module 1 and Module 2. The complexity of the promoter model to be constructed is defined by the number of units of each level: number of modules, number of site models, as well as the maximum number of individual sites to be considered. In the Genome Enhancer workflow we limit the number of modules to 2, the number of site models from 6 to 8, and the maximum number of individual sites to be considered from 1 to 3.

The CMA score is calculated for each promoter depending on the number of modules, site models, sites, their scores and other statistical parameters. The higher is the CMA score of a promoter, the better is the differentiation of this promoter from the promoters of the NO set. The distribution of scores for individual promoters is shown on an example histogram below:



The CMA score of the promoters is shown on the X axis of the histogram and the percentage of promoters (% sequences) having this score is shown on the Y axis. The separation point shown in gray corresponds to the average value. Promoters from the NO set with a score above the separation point (blue bars to the right of the separation point) are referred to as false positives. Promoters from the YES set with a score below the separation point (red bars to the left of the separation point) are referred to as false negatives. The YES promoters with a score above the separation point are well separated from the NO promoters, meaning that for these promoters the constructed composite model is most suitable.

The figure below demonstrates the calculation of the score value for the composite modules in the promoter sequences. The TSS is shown as a thin arrow on the right side of the figure. Four thick arrows exemplify four sites found in this promoter. The color of the arrows exemplifies the site model which these sites belong to (three site models – red, green and blue).



A promoter model consists of K modules. The score of each module $M_{k\ (Score(M_{k}),\ k\ =\ 1,\ \ldots,\ K)}$ is calculated according to the following formula:

$$Score\ (M_k) = \max_{\mu=1,L} \sum_{t=1}^{T_k} \sum_{i=1}^{m_t} SiteScore(t,i) \times f(x_{t,i}, \mu, \sigma^2)$$

Here,

- *SiteScore (t,i)* is the site score for the sites found in the promoter, which is calculated by the MATCH algorithm.
- m_t – the number of sites of the site model t found in the promoter.
- T_k – the number of site models in the module M_k, and
- $$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The final promoter score is calculated as the sum of the module scores M_k.

Standard deviation (σ) of the normal distribution is subject to optimization by the genetic algorithm and represents the width of the module in the result of the composite module analysis.

Transcription factors that correspond to the PWMs of the composite modules are further taken to the next step of finding master regulators in networks.

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 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. [3,4]

The signal transduction network is represented as a weighted graph, which is defined as Θ = (M, E, S), where M is the set of nodes (all molecules in the database), E is the set of edges (all reactions between molecules in the database) and S : E → R+ ∪ {0} is the cost function that defines a non-negative value for every edge. In the simplest variant of the algorithm, the initial values of the cost functions for each direct reaction are defined as 1.0. So, the cost of any path through the consequent reactions will be equal to the number of reactions. In our application of the algorithm in Genome Enhancer, we used a weighted graph, which encodes the types of reactions (direct or indirect) from TRANSPATH® database into different edge weights (costs).

The core of the algorithm is the Dijkstra's shortest-path algorithm. The upstream search starts from each molecule of the input set (subset M_x of M; e.g. transcription factors found on the previous step) and constructs the shortest-path to all nodes i of M, limited by a specified radius. After evaluating all nodes of M_x the algorithm calculates the number of visits N_i for each node i of M. This corresponds to the number of molecules of the input set M_x that can receive the signal from the node i. The values N_i are required to calculate the "master-regulator" score. The higher the value N_i the better chances has molecule i to transduce its signal to the molecules of the initial list M_x. However, if to use the value of N_i as a straightforward "master-regulator" score would be too simplistic since it would rank those molecules high that are in general highly connected within the network and thus are likely to be false positives (so called "hub" molecules). So, in our score we incorporate the full number of TF nodes in the whole network M that can be reached from the molecule i. The score is computed for each potential master regulator and it reflects a certain balance between sensitivity and specificity of signal transduction from this master node to the downstream effector TFs:

$$S(k) = \sum_{k=1}^{k_{\max}} \frac{M_k}{\left(1 + \kappa \cdot \frac{N_k}{N_{\max,k}}\right) \cdot M_{\max,k}}$$

where k is the radius of pathway steps from the master node to the effector nodes, M_k is the number of input TFs reached by a signal from the master node within k steps, and N_k is the total number of all potential TFs in the database reached by a signal from the master node within k steps. $M_{max,k}$ and $N_{max,k}$ are the highest values among all possible master regulator nodes which helps to normalize the score to the (0,1)-interval. The higher this score, the more sensitive and more specific this master regulator is for the set of input TFs. The parameter k is the penalty which is set to 0.1 in the Genome Enhancer workflow.
Incorporation of proteomics data into the analysis of master regulators is done through application of the so called "Context algorithm", which is described below.

When we have additional experimental information about the activity of certain components of the signal transduction network (additional context information) we can use this information to adapt the search for master regulators. To do that, the context algorithm encodes this additional context information as modified edge costs in the signaling network. For instance, the proteomics data gives information about proteins that are expressed in the cell. We call them "context proteins". The idea of the approach is to attract the key node search (e.g. the underlying Dijkstra algortithm for shortest paths) towards context proteins by decreasing the costs of those edges that are close to the context proteins in the network. There are two major aspects of this algorithm:
1. Attraction ("gravity") of the shortest-paths towards context proteins. The gravity is achieved by decreasing the costs of the incoming and outgoing edges in the graph for the nodes corresponding to the context proteins. The zero cost of the edges gives the maximal "gravity" for the nodes. During the search the path that follows through this node will be always "cheaper" than other alternative paths. This way we push the search algorithm to construct its paths maximally through the context proteins when it is possible.
2. "Decay" factor. It is necessary to extend the attraction power of the "gravity" to a wider area of the network around the context proteins in order to attract the search algorithm to go as close as possible to the context proteins in case there is no path that goes through the context gene directly. To enable such "long distance" attraction power we distribute the gravity to edges of the next neighboring nodes. This pushes the shortest path algorithm to find alternative paths still at least close to the context proteins. The gravity strength reduces with increasing distance by a decay factor. The decay factor is set to 0.1 in the Genome Enhancer workflow.
The context algorithm is based on creating a graph G consisting of gravity factors [0..1], which reflect both aspects described above by modifying edge costs. Graph G is used to modify the costs of S of original graph Θ resulting in S' which can be then used for any subsequent key node analysis or possibly other shortest path based analysis.

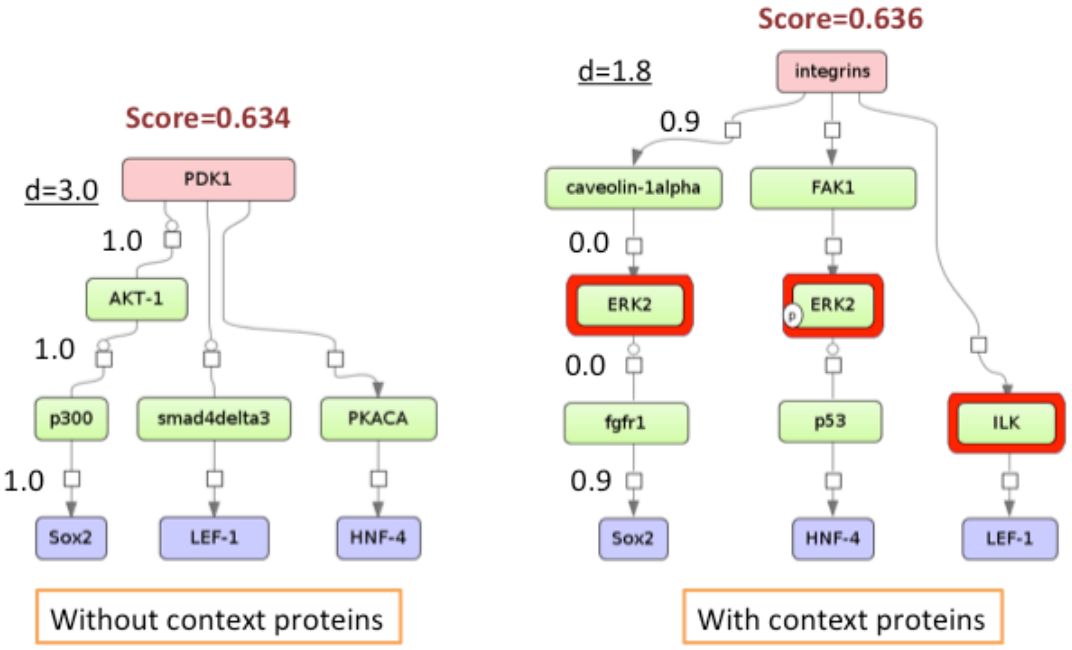
Initially the graph G is created as a copy of original signaling graph Θ, where all edges get initialized with weight zero, while the edges linked to context proteins are initialized by value 1.0. If quantitative proteomics data are available the algorithm allows to use the absolute protein expression values as the graph weights instead of 1.0. In such case the protein expression values are normalized to the range [0..1] (in such case these expression values will represent different context strengths). As the result we obtain a vector $\vec{c}$ representing the "expression" values of context proteins. The initial graph G reflects the gravity of the context proteins in the closest proximity only. The gravity is then propagated through the whole graph in downstream and upstream directions as follows. The gravity for the closest edge subsequently forwarded to the next adjacent edge by multiplying it by a user defined decay factor d.

$$D = \sum_{k=1} d^{k-1} \cdot \vec{c} \cdot S^k \qquad U = \sum_{k=1} d^{k-1} \cdot \vec{c} \cdot \left(S^T\right)^k$$

The first graph D is the result of the downstream direction and U is the result of the upstream direction of the gravity propagation through the graph G. These are two independent feed-forward runs. Both are added together, normalized to [0..1] and each weight $(D+U)_{ij}$ is replaced by $(1-(D+U)_{ij})$, such that the weights become compatible to the cost paradigm. So, the final graph G is constructed as follows:

$$G = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \cdots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} - \frac{(D+U)}{\|D+U\|_{max}}$$

At the final step, the costs of the edges of the initial signaling graph Θ are modified by the context protein information stored in the gravity graph G by simple multiplication. The factor r scales the new edge costs such that they are equal to the initial costs of graph Θ in average. $S'_{ij} = S_{ij} \cdot G_{ij} \cdot r$ for all i and j.
The Context algorithm is illustrated in the figure below.



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 [8]). We control the error rate by the FDR threshold 0.05.
To identify feed forward loops in transduction network we calculate regulatory scores for transcription factors from CMA result and keynodes analysis. The matrices from the CMA model are converted to TFs. Keynode analysis is then performed on these TFs and resulting keynodes are sorted by keynode score. Then the keynodes, which are regulated by the model, are selected.

$$S_j = \sum_{j=1}^{M_j} \frac{1}{D_{i,j}}; \; D_{i,j} = \sum_{k=1}^{C_{i,j}} d_k$$

Here:

- j – TF (Transcription Factor),
- i – MR (Master Regulator),
- S_j – Regulatory score of the TF,
- $D_{i,j}$– Cumulative distance between MR i and TF j,
- M_j – Number of MRs whose signal can reach the TF j and whose gene is regulated by TF j,
- $C_{i,j}$ – Number of steps in the network from between MR i and TF j (shortest path),
- d_k – Distance of the step k in the shortest path between MR i and TF j.

In unweighted TRANSPATH network $d_k = 1$ for each direct signaling reaction (like A+B-C = D), $d_k >= 3$ for each semantic reaction (like A → B). dk depends on the type of reaction, on basic or orthology level of its components, and on species of its components.
In weighted TRANSPATH network d_k is changed after applying the Context Algorithm as described in [4].

Methods for analysis of pharmaceutical compounds

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

Method for analysis of known pharmaceutical compounds

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

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

"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.

The clinical validity score reflects the number of the highest clinical trials phase (from 1 to 4) on which the drug was ever tested for any pathology.

Method for prediction of pharmaceutical compounds

In this study, the focus was put on compounds with high pharmacological efficiency and low toxicity. For this purpose, comprehensive library of chemical compounds and drugs was subjected to a SAR/QSAR analysis provided by the PASS software version 2020. This library contains 13040 compounds along with their pre-calculated potential pharmacological activities of those substances, their possible side and toxic effects, as well as the possible mechanisms of action. All biological activities are expressed as probability values for a substance to exert this activity (*Pa*).

We selected compounds that satisfied the following conditions:

1. Toxicity below a chosen toxicity threshold (defines as *Pa*, probability to be active as toxic substance).
2. For all predicted pharmacological effects that correspond to a set of user selected disease(s) *Pa* is greater than a chosen effect threshold.
3. There are at least 2 targets (corresponding to the predicted activity-mechanisms) with predicted *Pa* greater than a chosen target threshold.

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

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

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

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

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

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

8. References

1. Kel A, Voss N, Jauregui R, Kel-Margoulis O, Wingender E. Beyond microarrays: Finding key transcription factors controlling signal transduction pathways. *BMC Bioinformatics.* **2006**;7(S2), S13. doi:10.1186/1471-2105-7-s2-s13

2. Stegmaier P, Voss N, Meier T, Kel A, Wingender E, Borlak J. Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer. *PLoS ONE*. **2011**;6(3):e17738. doi:10.1371/journal.pone.0017738
3. Koschmann J, Bhar A, Stegmaier P, Kel A, Wingender E. "Upstream Analysis": An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays.* **2015**;4(2):270-286. doi:10.3390/microarrays4020270.
4. Kel A, Stegmaier P, Valeev T, Koschmann J, Poroikov V, Kel-Margoulis OV, and Wingender E. Multi-omics "upstream analysis" of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer. *EuPA Open Proteom.* **2016**;13:1-13. doi:10.1016/j.euprot.2016.09.002
5. 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
6. 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
7. 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.
8. 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
9. 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
10. Kel, A., Boyarskikh, U., Stegmaier, P., Leskov, L.S., Sokolov, A.V., Yevshin, I., Mandrik, N., Stelmashenko, D., Koschmann, J., Kel-Margoulis, O. and Krull, M. Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer. *BMC bioinformatics.* Cambridge (UK): RSC Publishing. **2019**;20(Suppl 4):119:1-20. doi:10.1186/s12859-019-2687-7
11. 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
12. Filimonov D, Poroikov V. Probabilistic Approaches in Activity Prediction. Varnek A, Tropsha A. *Cheminformatics Approaches to Virtual Screening.* Cambridge (UK): RSC Publishing. **2008**;:182-216.
13. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal.* **2006**;50(2):66-75 (russ)
14. 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:10.1021/ci980335o

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: Squamous Cell Carcinoma vs. Control: Non-tumour tissue).
4. Supplementary table 4 - Detailed report. Composite modules and master regulators (down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue).
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.