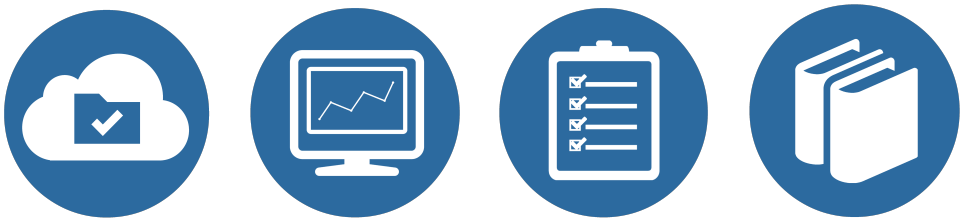
EGFR and ITGA3 are promising druggable targets for treating Squamous Cell Carcinoma that control activity of SMAD1, ILF3 and JUN transcription factors on promoters of differentially expressed genes

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 13/08/2019 ; Run on 20/11/2024 ; Report generated on 21/11/2024

Genome Enhancer release 3.5 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2024.2)



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: SMAD1, ILF3, TCF3, JUN, RELA and FOS. The subsequent network analysis suggested

- integrins
- 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. Having checked the actual druggability potential of the full list of identified targets, both, via information available in medical literature and via cheminformatics analysis of drug compounds, we have identified the following drugs as the most promising treatment candidates for the studied pathology: Erlotinib, Curcumin and {(2Z)-4-AMINO-2-[(4-METHOXYPHENYL)IMINO]-2,3-DIHYDRO-1,3-THIAZOL-5-YL}(4-METHOXYPHENYL)METHANONE.

1. Introduction

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

Conventional approaches of statistical "-omics" data analysis provide only very limited information about the causes of the observed phenomena and therefore contribute little to the understanding of the pathological molecular mechanism. In contrast, the "upstream analysis" method [1-4] applied here has been deviced to provide a casual interpretation of the data obtained for a pathology state. This approach comprises two major steps: (1) analysing promoters and enhancers of differentially expressed genes for the transcription factors (TFs) involved in their regulation and, thus, important for the process under study; (2) re-constructing the signaling pathways that activate these TFs and identifying master regulators at the top of such pathways. For the first step, the database TRANSFAC® [6] is employed together with the TF binding site identification algorithms Match [7] and CMA [8]. The second step involves the signal transduction database TRANSPATH® [9] and special graph search algorithms [10-11] 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 [5]. In addition, some known drugs and investigational active chemical compounds are subsequently predicted as potential ligands for the revealed molecular targets. They are predicted using a pre-computed database of spectra of biological activities of chemical compounds of a library of 2245 known drugs and investigational chemical compounds from HumanPSD™ database. The spectra of biological activities for these compounds are computed using the program PASS on the basis of a (Q)SAR approach [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.fastq	Transcriptomics
SRR349742.fastq	Transcriptomics
SRR349748.fastq	Transcriptomics
SRR349749.fastq	Transcriptomics

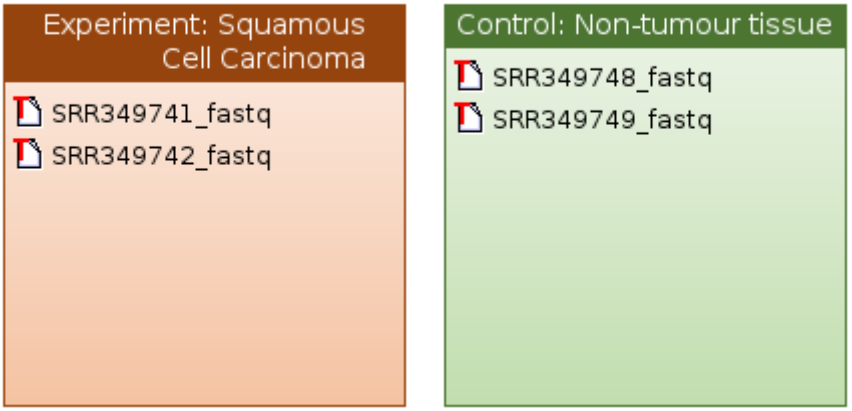


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

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 1955 upregulated genes (LogFC>0.1) out of which 768 genes were found as significantly upregulated (p-value<0.1) and 1739 downregulated genes (LogFC<-0.1) out of which 650 genes were significantly downregulated (p-value<0.1). See tables below for the top significantly up- and downregulated genes. Below we call **target genes** the full list of up- and downregulated genes revealed in our analysis (see tables in Supplementary section).

*Table 2. Top ten significant **up-regulated** genes in Experiment: 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 →

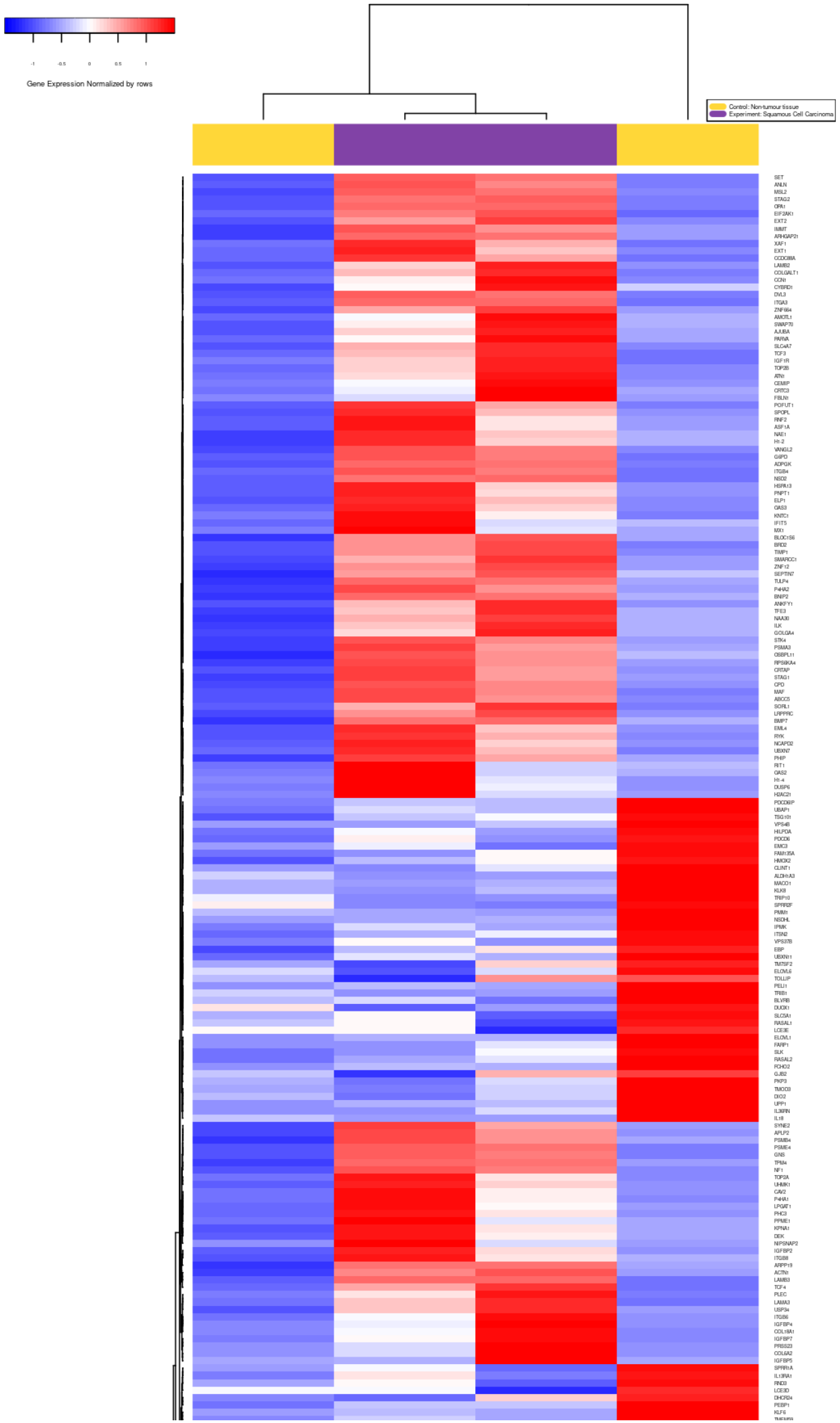
ID	Gene symbol	Gene description	logFC	logCPM	PValue	FDR
ENSG00000136155	SCEL	sciellin	-7.72	11.12	2.73E-15	5.38E-12
ENSG00000163209	SPRR3	small proline rich protein 3	-6.69	14.45	8.44E-4	1.1E-2
ENSG00000143369	ECM1	extracellular matrix protein 1	-6.38	11.04	4.35E-10	7.45E-8
ENSG00000189334	S100A14	S100 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

3.2. Functional classification of genes

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

Heatmap of differentially expressed genes in Experiment: 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
SET
ANLN
MSL2
STAG2
OPA1
EIF2AK1
EXT2
IMMT
ARHGAP21
XAF1
EXT1
CCDC88A
LAMB2
COLGALT1
CCNI
CYBRD1
DVL3
ITGA3
ZNF664
ANKRD11
SWAP70
AJUBA
PARVA
SLC4A7
TCF3
IGF1R
TOP2B
ATN1
CEMIP
CRTC3
FBLN1
POFUT1
SPOPL
RNF2
ASF1A
NAE1
H1-2
VANGL2
G6PD
ADPGK
ITGB4
NSD2
HSPA13
PNPT1
ELP1
OAS3
KNTC1
IFIT5
MXI
BLOC1S6
BRD2
TIMP1
SMARCC1
ZNF12
SEPTIN7
TULP4
P4HA2
BNIP2
ANKFY1
TFE3
NAA30
ILK
GOLGA4
STK4
PSMA3
OSBPL11
RPS6KA4
CRTAP
STAG1
CPD
MAF
ABCC5
SORL1
LRPPRC
BMP7
EML4
RYK
NCAPD2
UBXN7
PHIP
RIT1
OAS2
H1-4
DUSP6
H2AC21
PDCD6IP
UBAP1
TSG101
VPS4B
HILPDA
PDCD6
EMC3
FAM135A
HMOX2
CLINT1
ALDH1A3
MACC1
KLK8
TRIP10
SPRR2F
PMM1
NSDHL
IPMK
ITSN2
VPS37B
EBP
UBXN11
TM7SF2
ELOVL6
TOLLIP
PELI1
TRIB1
BLVRB
DUOX1
SLC5A1
RASAL1
LCE3E
ELOVL1
FARP1
SLK
RASAL2
FOXO3
GJB2
PKP3
TMOD3
DIO2
UPP1
IL36RN
IL1B
SYNE2
APLP2
PSMB4
PSME4
GNS
TPM4
NF1
TOP3A
UHMK1
CAV2
P4HA1
LPGAT1
PHC3
PPME1
KPNA1
DEK
NIPSNAP2
IGFBP2
ITGB8
ARPP19
ACTN1
LAMB3
TCF4
PLEC
LAMA3
USP34
ITGB6
IGFBP4
COL18A1
IGFBP7
PRSS23
COL6A2
IGFBP5
SPRR1A
IL13RA1
RND3
LCE3D
DHCR24
PEBP1
KLF6
TMEM59

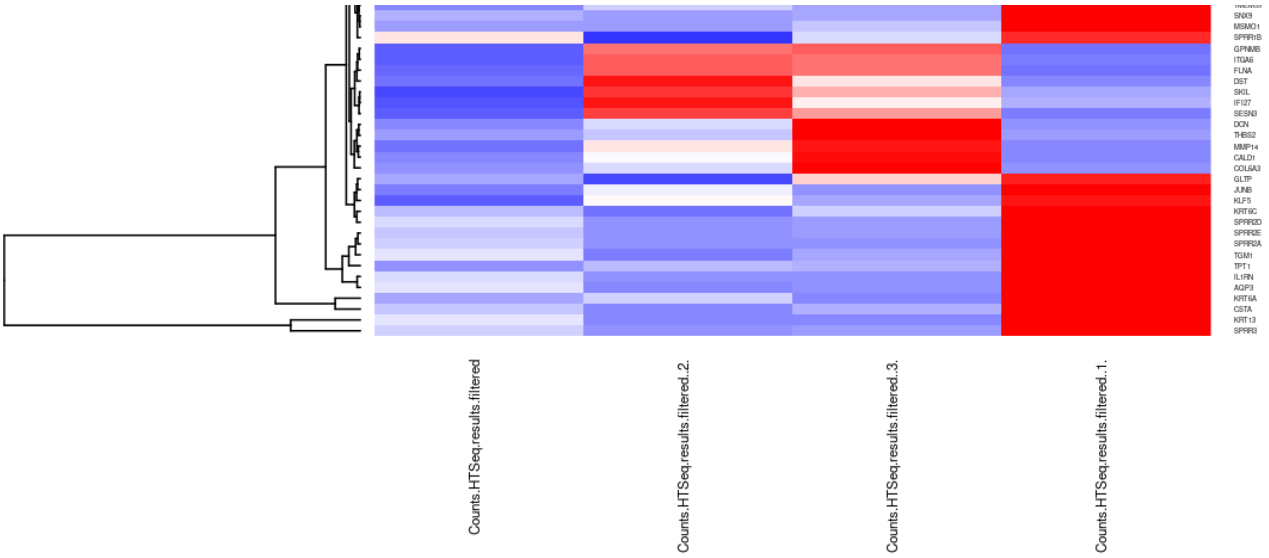


Figure 2. Heatmap of genes enriched in Transpath categories. The colored bar at the top shows the types of the samples according to the legend in the upper right corner.

See full diagram →

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

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

GO (biological process)

Figure 3. Enriched GO (biological process) of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

Full classification →

TRANSPATH® Pathways (2024.2)

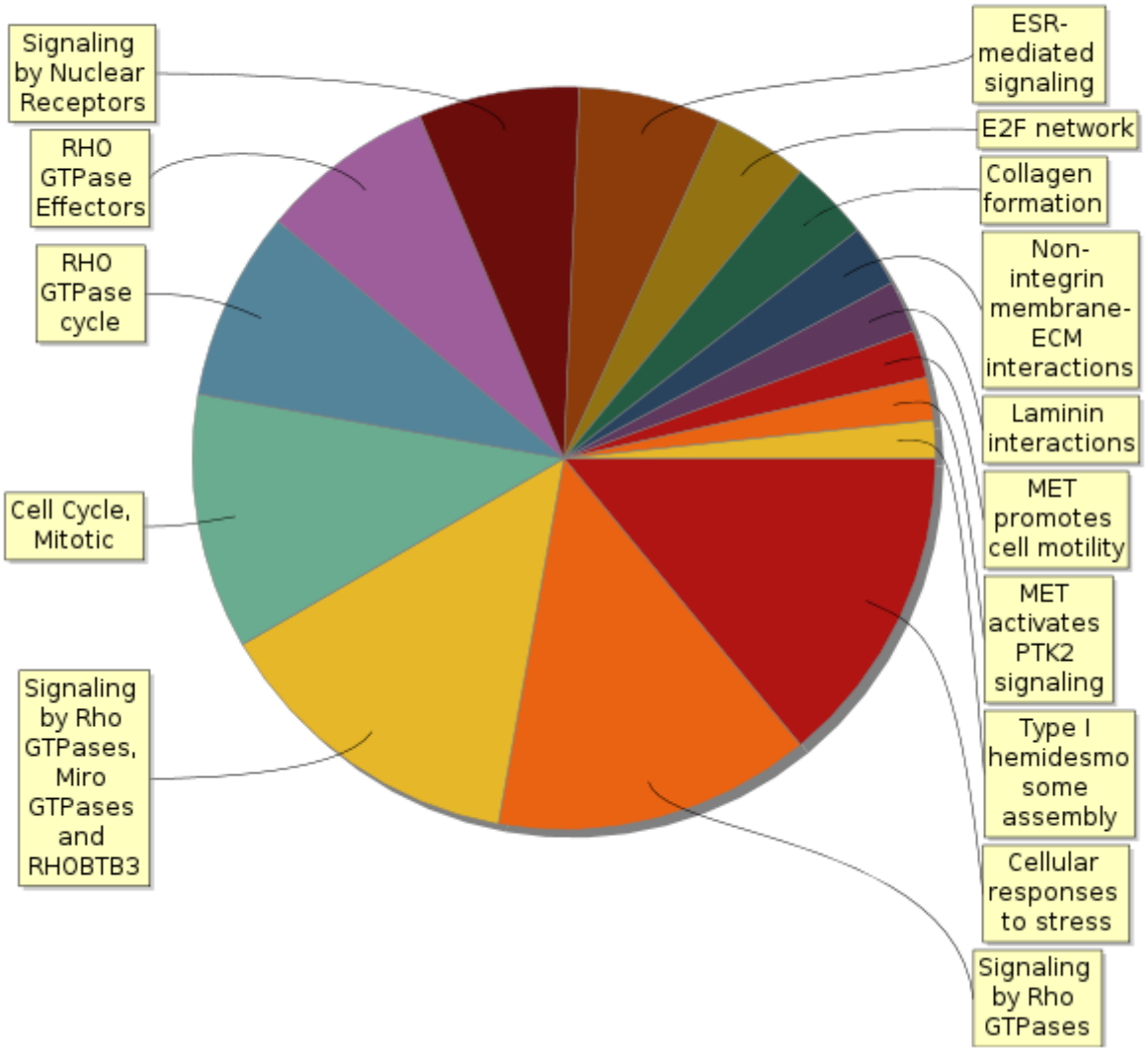


Figure 4. Enriched TRANSPATH® Pathways (2024.2) of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

Full classification →

HumanPSD(TM) disease (2024.2)

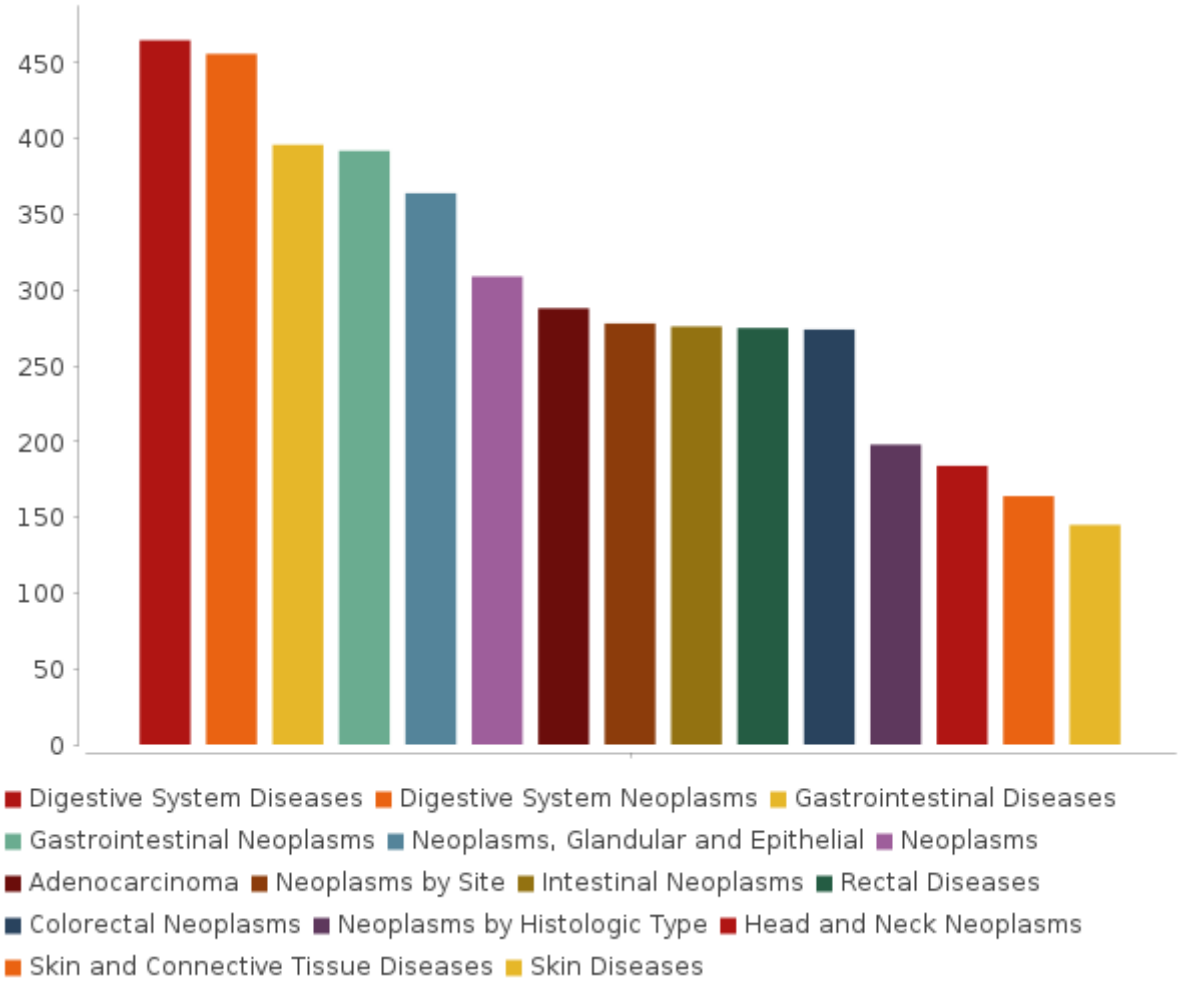


Figure 5. Enriched HumanPSD(TM) disease (2024.2) of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

Full classification →

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

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

GO (biological process)

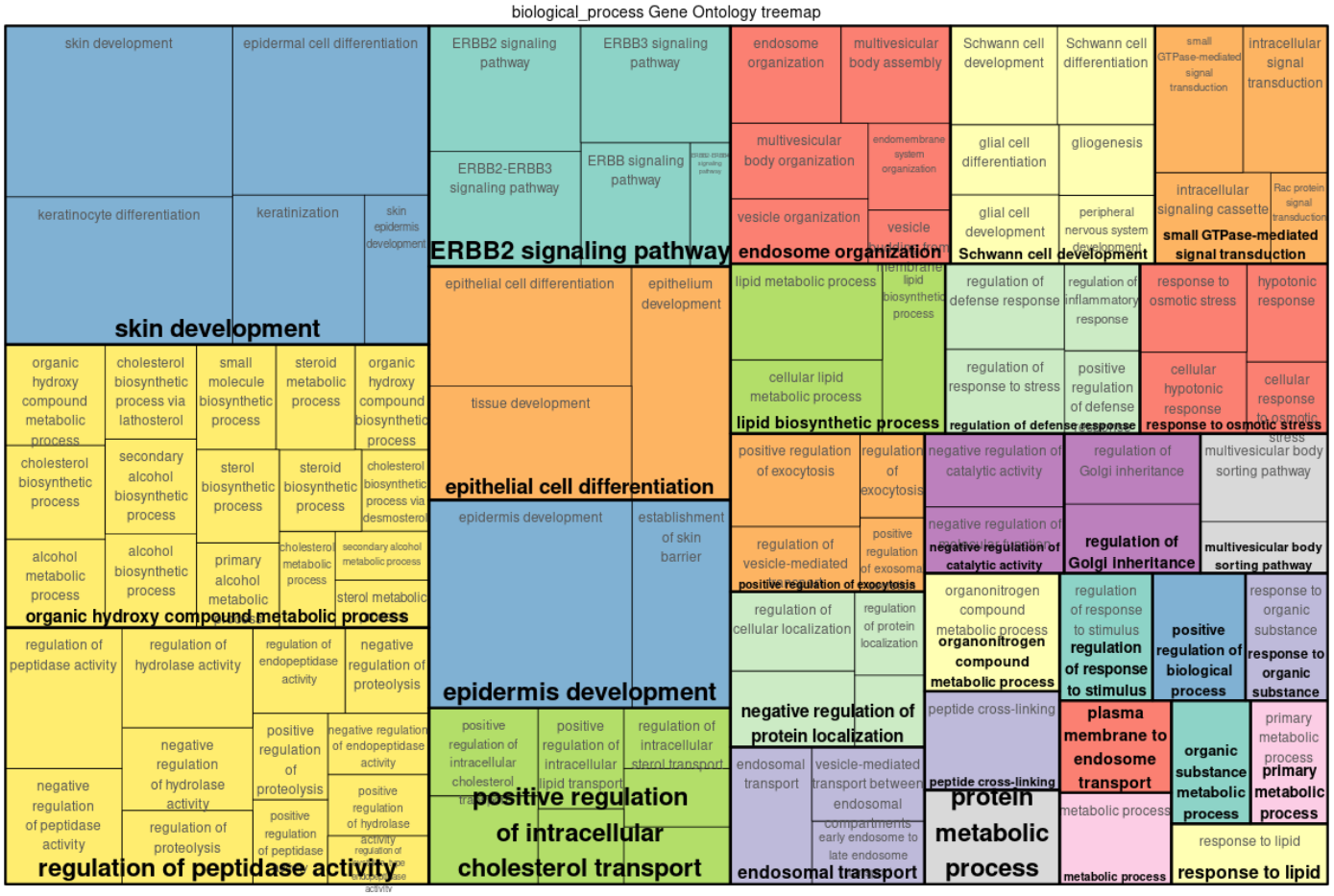


Figure 6. Enriched GO (biological process) of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

Full classification →

TRANSPATH® Pathways (2024.2)

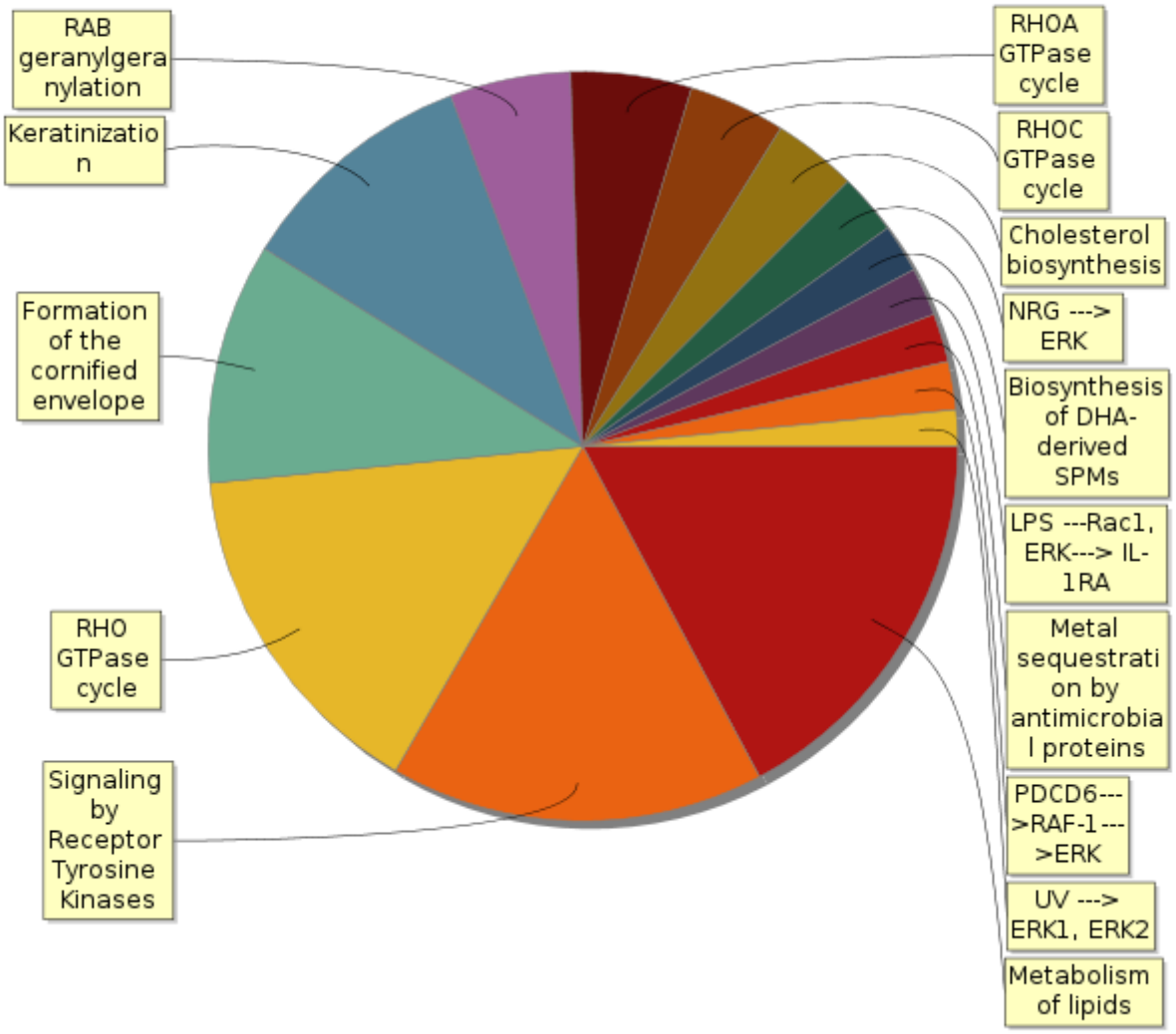


Figure 7. Enriched TRANSPATH® Pathways (2024.2) of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue.

Full classification →

HumanPSD(TM) disease (2024.2)

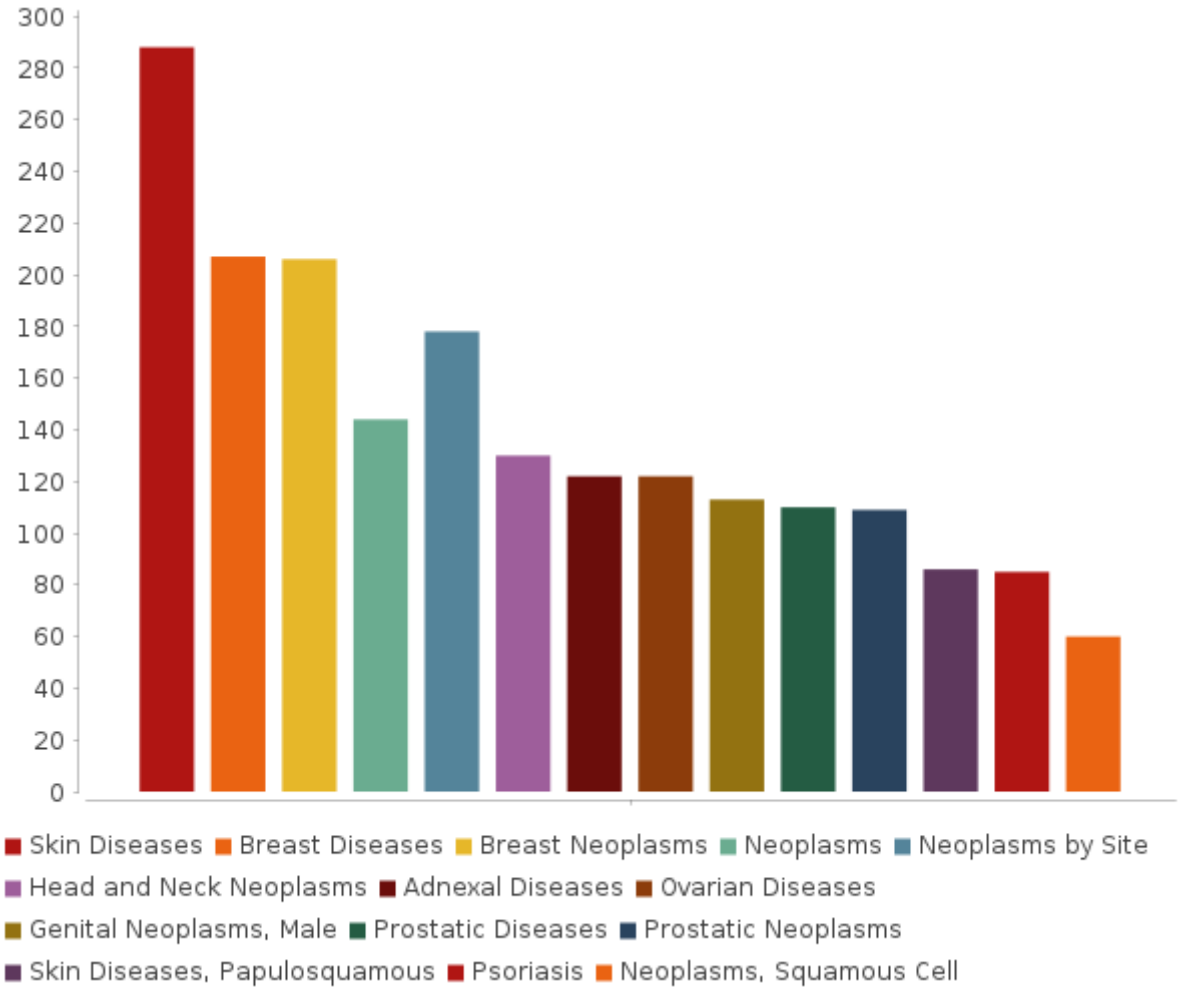
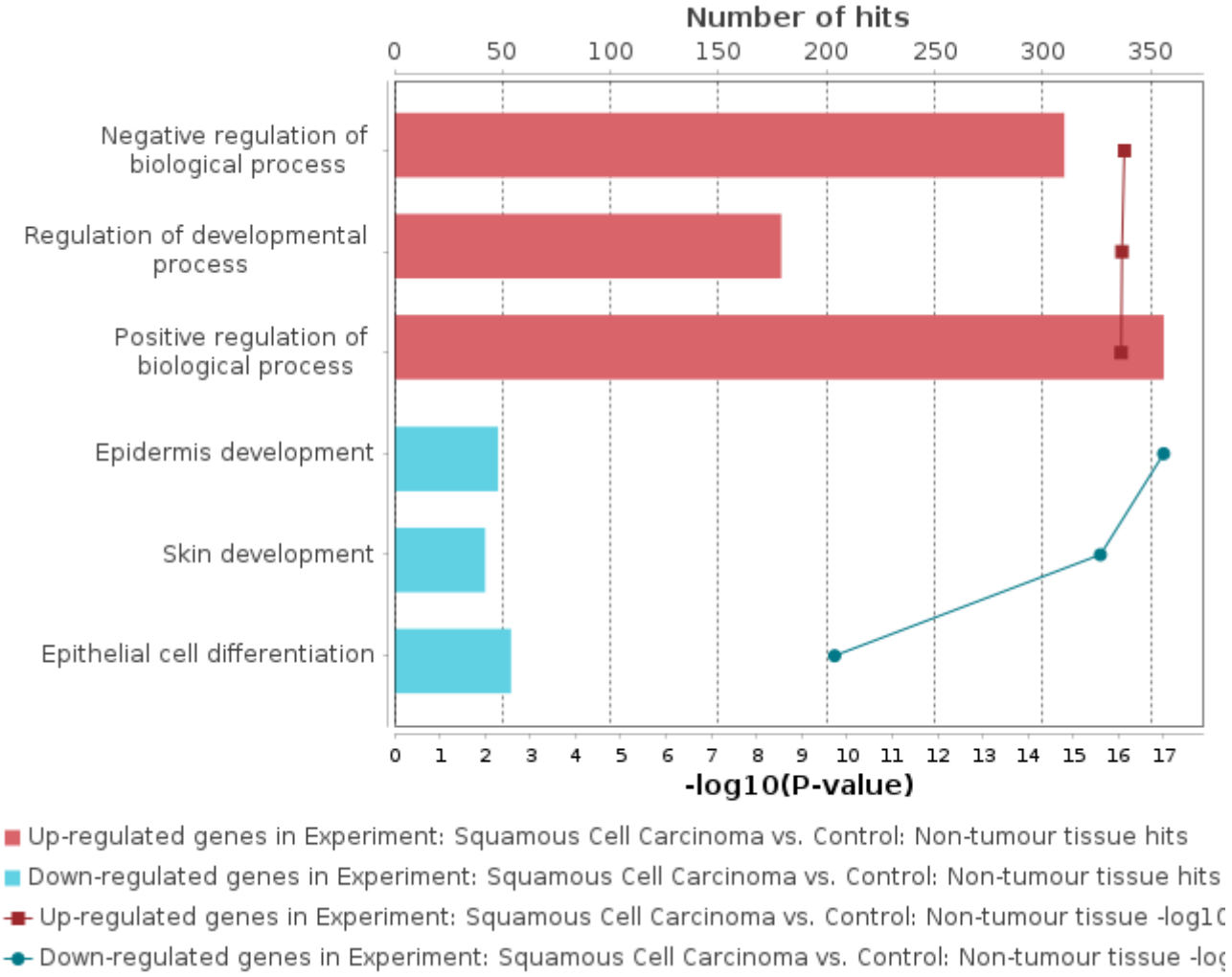


Figure 8. Enriched HumanPSD(TM) disease (2024.2) of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

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 [8]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [9].

In the current work, we use the 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 643 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 →

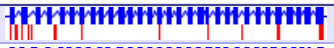
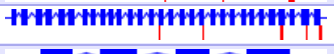
ID	Gene symbol	Gene schematic representation	Number of variations
ENSG00000146648	EGFR		21
ENSG00000083857	FAT1		16
ENSG00000134871	COL4A2		13
ENSG00000186340	THBS2		10
ENSG00000226445	THBS2-AS1		9
ENSG00000145012	LPP		8
ENSG00000114999	TTL		7
ENSG00000142173	COL6A2		7
ENSG00000152291	TGOLN2		7
ENSG00000157214	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$RUNX3_01	4.83E-2	1.15E-45	19	1895
V$CREB1_09	3.47E-2	1.24E-35	175	644
V$E2F7_02	2.67E-2	4.98E-42	16	1944
V$BHLHE41_02	2.66E-2	3.91E-43	10	278
V$E2F1_Q3_01	1.29E-4	9.25E-36	23	175
V$ARNT2_03	2.81E-6	9.04E-54	172	325
V$ARNTLPITX1_01	7.83E-14	7.37E-43	378	273
V$E2F_Q2	9.88E-16	7.73E-40	115	183
V$ERFETV7_01	1.51E-21	8.57E-42	1378	1092
V$E2F7_04	3.71E-26	3.15E-53	421	744
V$SP3_03	1.15E-32	6.72E-27	1024	1555
V$KLF4_05	3.84E-33	7.26E-18	1417	1600
V$GLI_Q2	4.53E-34		1421	
V$E2F4_09	1.35E-35		1496	
V$NF90_01	1.3E-35		2191	
V$E2F2_08	1.13E-36	1.53E-37	1674	2302
V$E2F3_05	2.56E-37	5.56E-68	1025	1505
V$E2F2_06	8.47E-38	4.38E-61	587	896
V$SP1_Q3	3.55E-38	3.09E-2	883	9
V$GCM1_06	1.64E-38	2.96E-2	1714	1

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

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

To build the most specific composite modules we choose top 300 significant up-regulated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all up-regulated genes in Experiment: 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$NF90_01	V$FOXA1_09	V$OSX_Q3	V$HIC1_10	V$HAND1E47_01	V$NR3C2_03
0.97; N=2	0.89; N=1	0.99; N=2	0.92; N=3	0.90; N=2	0.92; N=1

V$HOXA9_04
0.93; N=2

Module width: 130

Module 2:

V$NF90_01	V$FOXA1_09	V$OSX_Q3	V$TFAP2C_05	V$SMAD1_Q6	V$CUX1FOXI1_01
2.00; N=3	0.90; N=2	0.99; N=2	0.94; N=2	0.97; N=3	0.73; N=2

V$GATA4_Q5_01
1.00; N=3

Module width: 136

Model score (-p*log10(pval)): 12.52
Wilcoxon p-value (pval): 2.36e-28
Penalty (p): 0.453
Average yes-set score: 5.85
Average no-set score: 4.62
AUC: 0.73
Separation point: 5.06
False-positive: 39.03%
False-negative: 22.41%

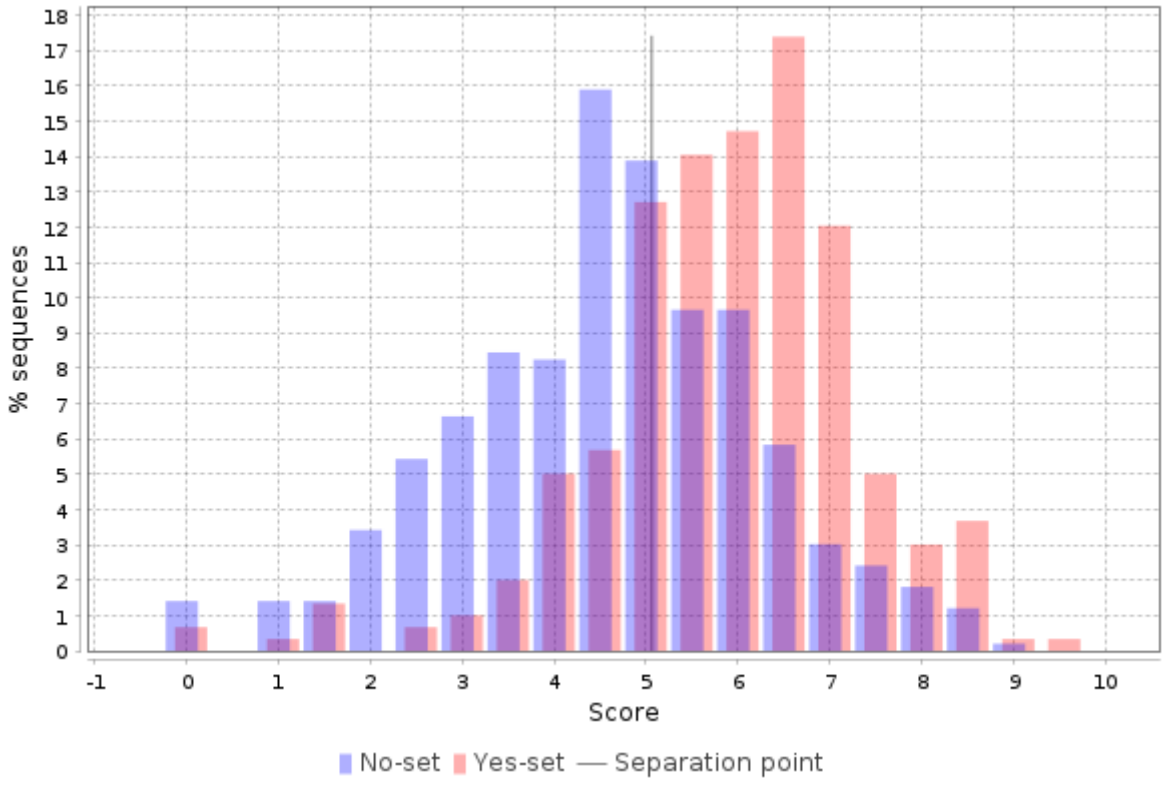


Table 6. List of top ten up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*
See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000083168	KAT6A	lysine acetyltransferase 6A	10.98	AP-2gamma(h), HIC1(h), E2A(h),HAND-1(h), SMAD1(h), FOXA1(h), ILF3(h), Sp7(h)
ENSG00000181449	SOX2	SRY-box transcription factor 2	10.89	SMAD1(h), MR(h), E2A(h),HAND-1(h), HIC1(h), CUX-1(h),FOXI1(h), FOXA1(h), Sp7(h)...
ENSG00000151923	TIAL1	TIA1 cytotoxic granule associated RNA binding protein like 1	10.78	ILF3(h), HIC1(h), E2A(h),HAND-1(h), MR(h), FOXA1(h), CUX-1(h),FOXI1(h), GATA-4(h)...
ENSG00000166225	FRS2	fibroblast growth factor receptor substrate 2	10.3	HIC1(h), E2A(h),HAND-1(h), CUX-1(h),FOXI1(h), FOXA1(h), GATA-4(h), Sp7(h)
ENSG00000114062	UBE3A	ubiquitin protein ligase E3A	10.05	SMAD1(h), AP-2gamma(h), Sp7(h), HIC1(h), FOXA1(h), E2A(h),HAND-1(h), CUX-1(h),FOXI1(h)
ENSG00000170027	YWHAG	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma	10.05	ILF3(h), Sp7(h), HIC1(h), E2A(h),HAND-1(h), SMAD1(h), FOXA1(h)
ENSG00000184640	SEPTIN9	septin 9	9.76	E2A(h),HAND-1(h), AP-2gamma(h), ILF3(h), CUX-1(h),FOXI1(h), HIC1(h), Sp7(h), GATA-4(h)
ENSG00000100243	CYB5R3	cytochrome b5 reductase 3	9.75	CUX-1(h),FOXI1(h), HIC1(h), FOXA1(h), Sp7(h), E2A(h),HAND-1(h), AP-2gamma(h)
ENSG00000166333	ILK	integrin linked kinase	9.69	E2A(h),HAND-1(h), HIC1(h), FOXA1(h), SMAD1(h), MR(h), CUX-1(h),FOXI1(h), AP-2gamma(h)
ENSG00000166337	TAF10	TATA-box binding protein associated factor 10	9.69	E2A(h),HAND-1(h), HIC1(h), FOXA1(h), SMAD1(h), MR(h), CUX-1(h),FOXI1(h), AP-2gamma(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$KLF4_05 0.92; N=3 | V$ZNF462_01 0.92; N=3 | V$RELA_03 0.95; N=2 | V$FOXA1_07 0.90; N=2 | V$PR_02 0.88; N=3 | V$NKX25_08 0.98; N=2

V$AREB6_01 0.94; N=1

Module width: 124

Module 2:

V$GLI_Q2 0.91; N=2 | V$AP1_02 0.93; N=2 | V$OCT4_02 0.77; N=3 | V$IRF4_07 0.88; N=2 | V$TEAD1_06 0.97; N=3 | V$FOXA1_07 0.90; N=2

V$TRIM28_02 0.90; N=1

Module width: 150

Model score (-p*log10(pval)): 17.22
Wilcoxon p-value (pval): 9.92e-39
Penalty (p): 0.453
Average yes-set score: 8.96
Average no-set score: 7.20
AUC: 0.77
Separation point: 7.98
False-positive: 29.58%
False-negative: 25.75%

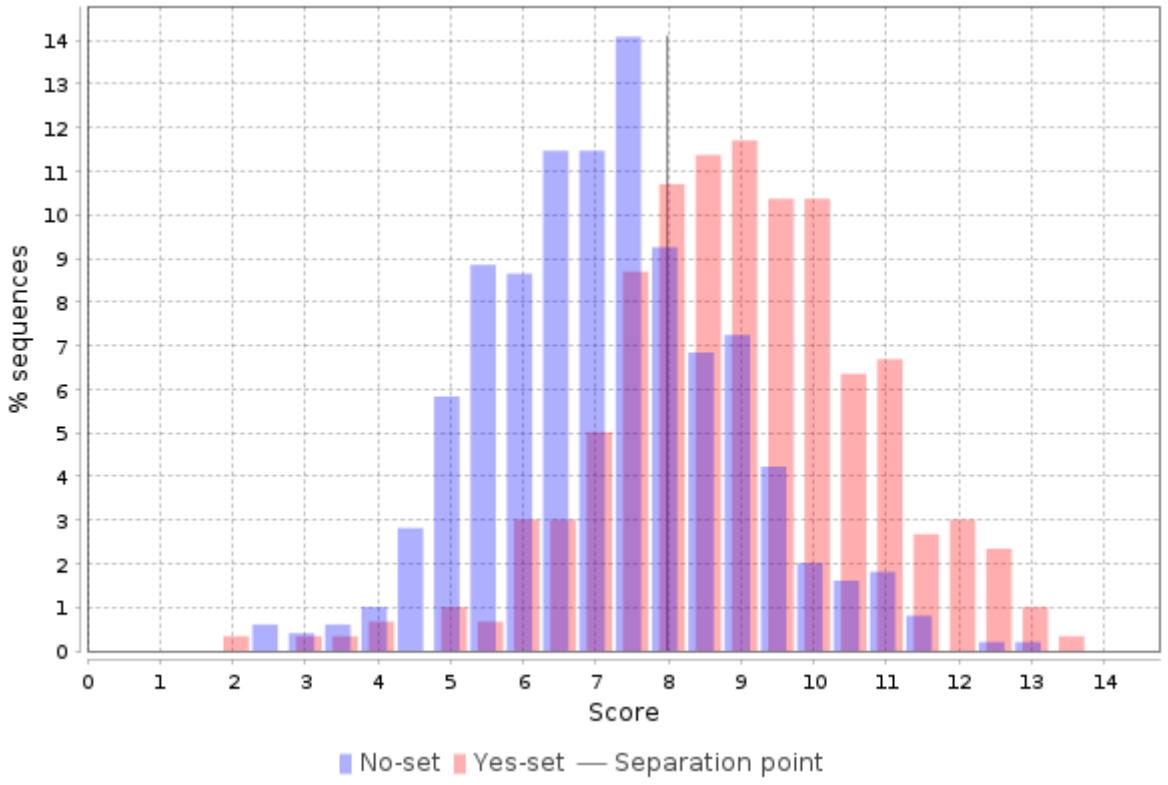


Table 7. List of top ten down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000138768	USO1	USO1 vesicle transport factor	14.24	IRF-4(h), c-Fos(h),c-Jun(h), POU5F1(h), FOXA1(h), TEF-1(h), TIF1-beta(h), PR(h)...
ENSG00000175793	SFN	stratifin	13.87	ZEB1(h), KLF4(h), PR(h), POU5F1(h), c-Fos(h),c-Jun(h), IRF-4(h), ZNF462(h)...
ENSG00000023191	RNH1	ribonuclease/angiogenin inhibitor 1	13.7	TEF-1(h), FOXA1(h), ZNF462(h), NKX-2.5(h), IRF-4(h), KLF4(h), NF-kappaB-p65(h)...
ENSG00000229732		novel transcript	13.45	PR(h), GLI1(h),GLI2(h),GLI3(h),GLIS1(h), ZNF462(h), c-Fos(h),c-Jun(h), IRF-4(h), NKX-2.5(h), TIF1-beta(h)...
ENSG00000108352	RAPGEFL1	Rap guanine nucleotide exchange factor like 1	13.45	POU5F1(h), TIF1-beta(h), IRF-4(h), NKX-2.5(h), KLF4(h), TEF-1(h), FOXA1(h)...
ENSG00000146677	RPL32P18	ribosomal protein L32 pseudogene 18	13.27	ZNF462(h), GLI1(h),GLI2(h),GLI3(h),GLIS1(h), POU5F1(h), NF-kappaB-p65(h), NKX-2.5(h), IRF-4(h), c-Fos(h),c-Jun(h)...
ENSG00000154265	ABCA5	ATP binding cassette subfamily A member 5	13.08	ZNF462(h), ZEB1(h), KLF4(h), NKX-2.5(h), NF-kappaB-p65(h), IRF-4(h), GLI1(h),GLI2(h),GLI3(h),GLIS1(h)...
ENSG00000151689	INPP1	inositol polyphosphate-1-phosphatase	13.06	NF-kappaB-p65(h), ZNF462(h), NKX-2.5(h), FOXA1(h), ZEB1(h), KLF4(h), TIF1-beta(h)...
ENSG00000146425	DYNLT1	dynein light chain Tctex-type 1	13.02	ZNF462(h), NKX-2.5(h), KLF4(h), NF-kappaB-p65(h), GLI1(h),GLI2(h),GLI3(h),GLIS1(h), IRF-4(h), POU5F1(h)...
ENSG00000137193	PIM1	Pim-1 proto-oncogene, serine/threonine kinase	13	POU5F1(h), IRF-4(h), c-Fos(h),c-Jun(h), NF-kappaB-p65(h), ZNF462(h), NKX-2.5(h), KLF4(h)

On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 13 and 17 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 →

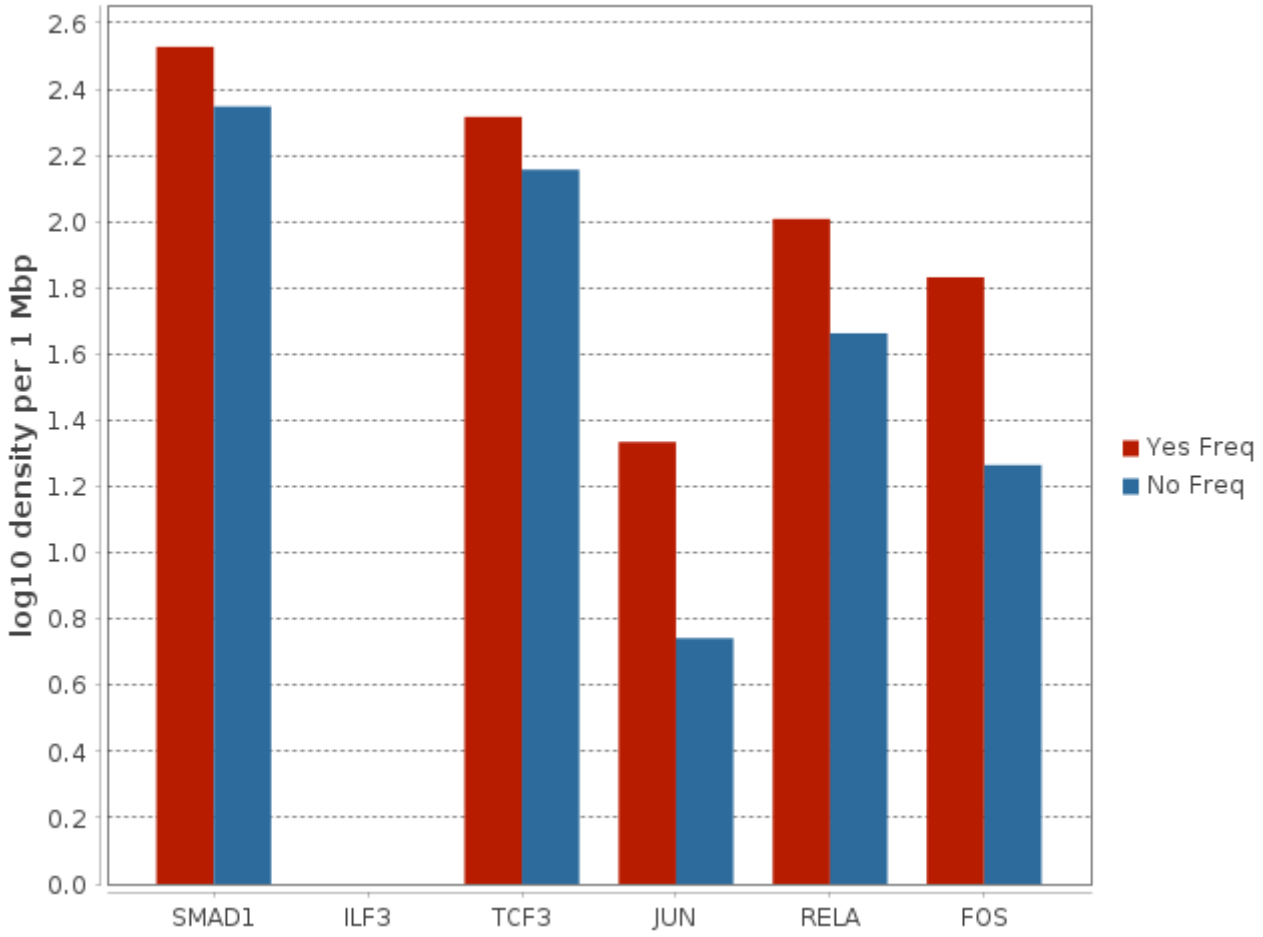
ID	Gene symbol	Gene description	Regulatory score	Yes-No ratio
MO000019609	SMAD1	SMAD family member 1	3.14	1.51
MO000137286	ILF3	interleukin enhancer binding factor 3	3.1	
MO000032492	TCF3	transcription factor 3	3.07	1.44
MO000024708	CUX1	cut like homeobox 1	2.88	3.02
MO000026663	GATA4	GATA binding protein 4	2.82	1.7
MO000026492	FOXA1	forkhead box A1	2.5	1.56
MO000119037	HOXA9	homeobox A9	2.5	2.25
MO000028230	HAND1	heart and neural crest derivatives expressed 1	2.43	3.02
MO000176655	SP7	Sp7 transcription factor	2.31	1.43
MO000021449	NR3C2	nuclear receptor subfamily 3 group C member 2	2.27	1.21

*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
MO000019469	JUN	Jun proto-oncogene, AP-1 transcription factor subunit	3.64	3.91
MO000079319	RELA	RELA proto-oncogene, NF-kB subunit	3.57	2.21
MO000018137	FOS	Fos proto-oncogene, AP-1 transcription factor subunit	3.42	3.69
MO000054297	PGR	progesterone receptor	3.27	1.52
MO000056618	POU5F1	POU class 5 homeobox 1	2.91	2.64
MO000139677	ZEB1	zinc finger E-box binding homeobox 1	2.88	2.19
MO000069886	TRIM28	tripartite motif containing 28	2.87	1.11
MO000092587	ZNF462	zinc finger protein 462	2.79	1.28
MO000125561	KLF4	KLF transcription factor 4	2.77	1.32
MO000019117	GLI1	GLI family zinc finger 1	2.72	1.74

The following diagram represents the key transcription factors, which were predicted to be potentially regulating differentially expressed genes in the analyzed pathology: SMAD1, ILF3, TCF3, JUN, RELA and FOS.



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
MO000016677	EGFR(h)	EGFR	epidermal growth factor receptor	4.44	173
MO000017291	integrins	ITGA1, ITGA2B, ITGA3, ITGA4, ITGA5, ITGA6, ITGA8, ITGA9, ITGAL, ITGAV, ITGB1, ITGB2, ITGB3, ITGB4, I...	integrin subunit alpha 1, integrin subunit alpha 2b, integrin subunit alpha 3, integrin subunit alph...	3.05	185
MO000082228	EGFR-p60(h)	EGFR	epidermal growth factor receptor	4.44	191
MO000082230	EGFR-p110(h)	EGFR	epidermal growth factor receptor	4.44	191
MO000087397	EGFR-isoform4(h)	EGFR	epidermal growth factor receptor	4.44	191
MO000082277	EGFR-p170(h)	EGFR	epidermal growth factor receptor	4.44	195
MO000125420	EGFR(h){ub}n	EGFR	epidermal growth factor receptor	4.44	206
MO000023529	EphB4(h)	EPHB4	EPH receptor B4	2.66	213
MO000042551	EGFR(h){pY}	EGFR	epidermal growth factor receptor	4.44	215
MO000093763	EphB4-isoform1(h)	EPHB4	EPH receptor B4	2.66	217

*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
MO000056491	KAT2B(h)	KAT2B	lysine acetyltransferase 2B	-3.15	112
MO000033396	DUSP5(h)	DUSP5	dual specificity phosphatase 5	-4.8	114
MO000137304	DUSP5(h)	DUSP5	dual specificity phosphatase 5	-4.8	123
MO000033299	pim1(h)	PIM1	Pim-1 proto-oncogene, serine/threonine kinase	-3.01	127
MO000031101	plk3(h)	PLK3	polo like kinase 3	-2.83	130
MO001092411	Cdk8:cyclinC:hN1:SNW1:RBP-Jkappa:MAML:PCAF:p300:CREBBP	CCNC, CDK8, CREBBP, EP300, KAT2A, KAT2B, MAML1, MAML2, MAML3, MAMLD1, NOTCH1, RBPJ, SNW1	CREB binding protein, E1A binding protein p300, SNW domain containing 1, cyclin C, cyclin dependent ...	-3.15	140
MO000021356	EGFR(h){pY}	EGFR, ERBB2, ERBB3, ERBB4	epidermal growth factor receptor, erb-b2 receptor tyrosine kinase 2, erb-b2 receptor tyrosine kinase...	-2.6	230
MO000112152	MRG-1(h)	CITED2	Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2	-3.42	309
MO000004672	ERK1(h)	MAPK3	mitogen-activated protein kinase 3	-2.26	320
MO000003497	Csk(h)	CSK	C-terminal Src kinase	-1.57	327

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figures 9 and 10. These diagrams display the connections between identified transcription factors, which play important roles in the regulation of differentially expressed genes, and selected master regulators, which are responsible for the regulation of these TFs.

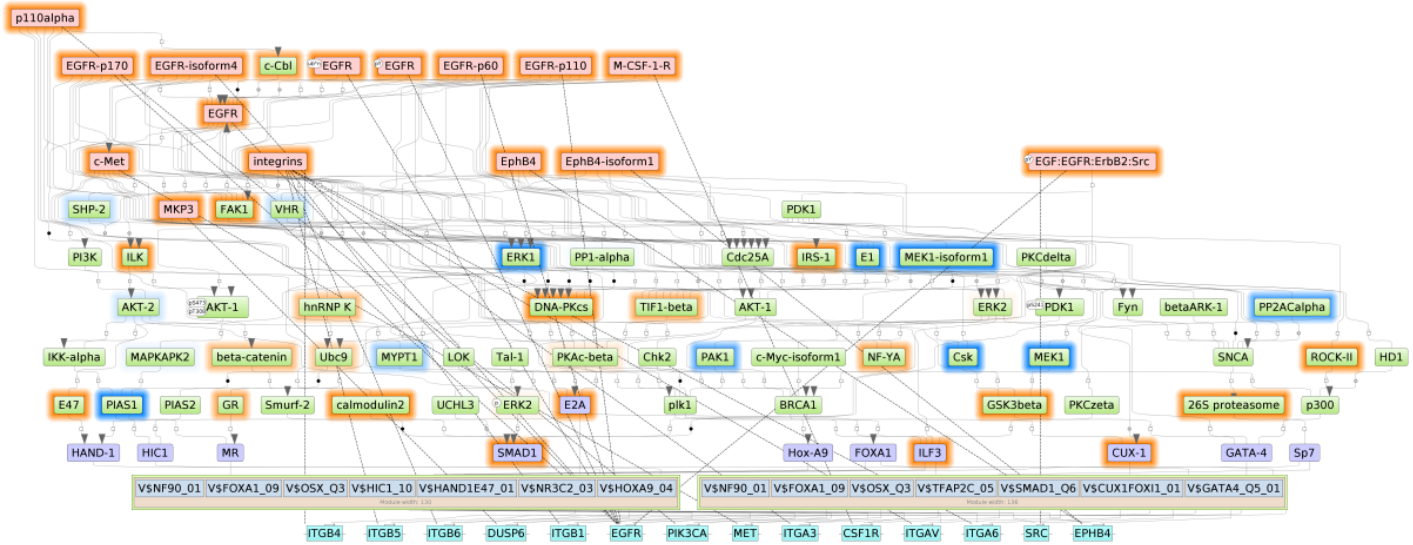


Figure 9. Diagram of intracellular regulatory signal transduction pathways of up-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, 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 →

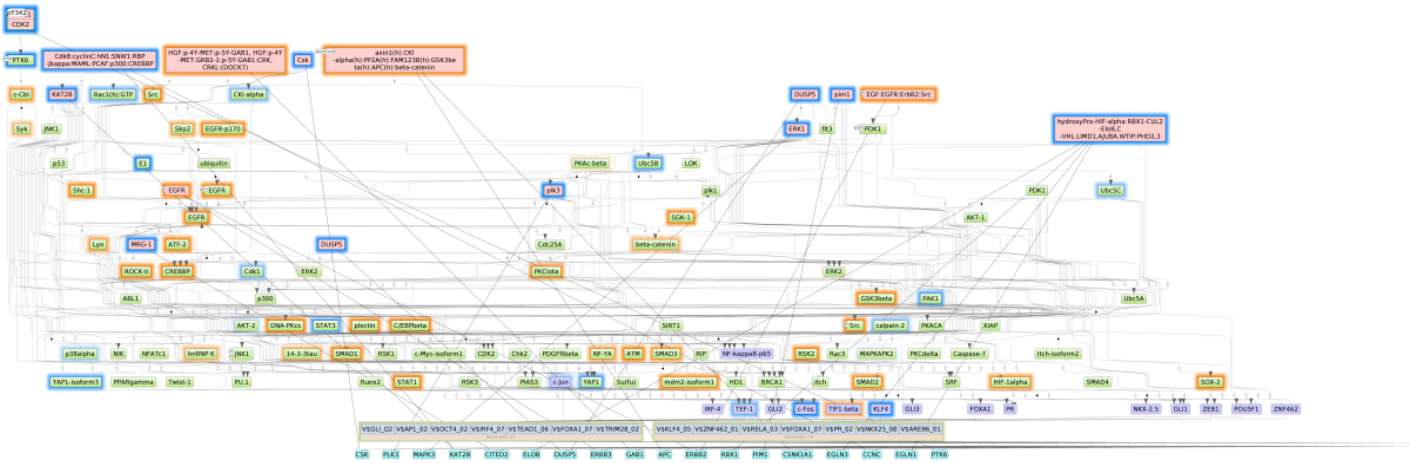


Figure 10. Diagram of intracellular regulatory signal transduction pathways of down-regulated genes in Experiment: Squamous Cell Carcinoma vs. Control: Non-tumour tissue. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate 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™ [5] 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	173
ITGA3	integrin subunit alpha 3	2	3.05	185
ITGB5	integrin subunit beta 5	2	3.05	185
ITGA6	integrin subunit alpha 6	1	3.05	185
ITGB4	integrin subunit beta 4	2	3.05	185
ITGAV	integrin subunit alpha V	3	3.05	185

*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.*

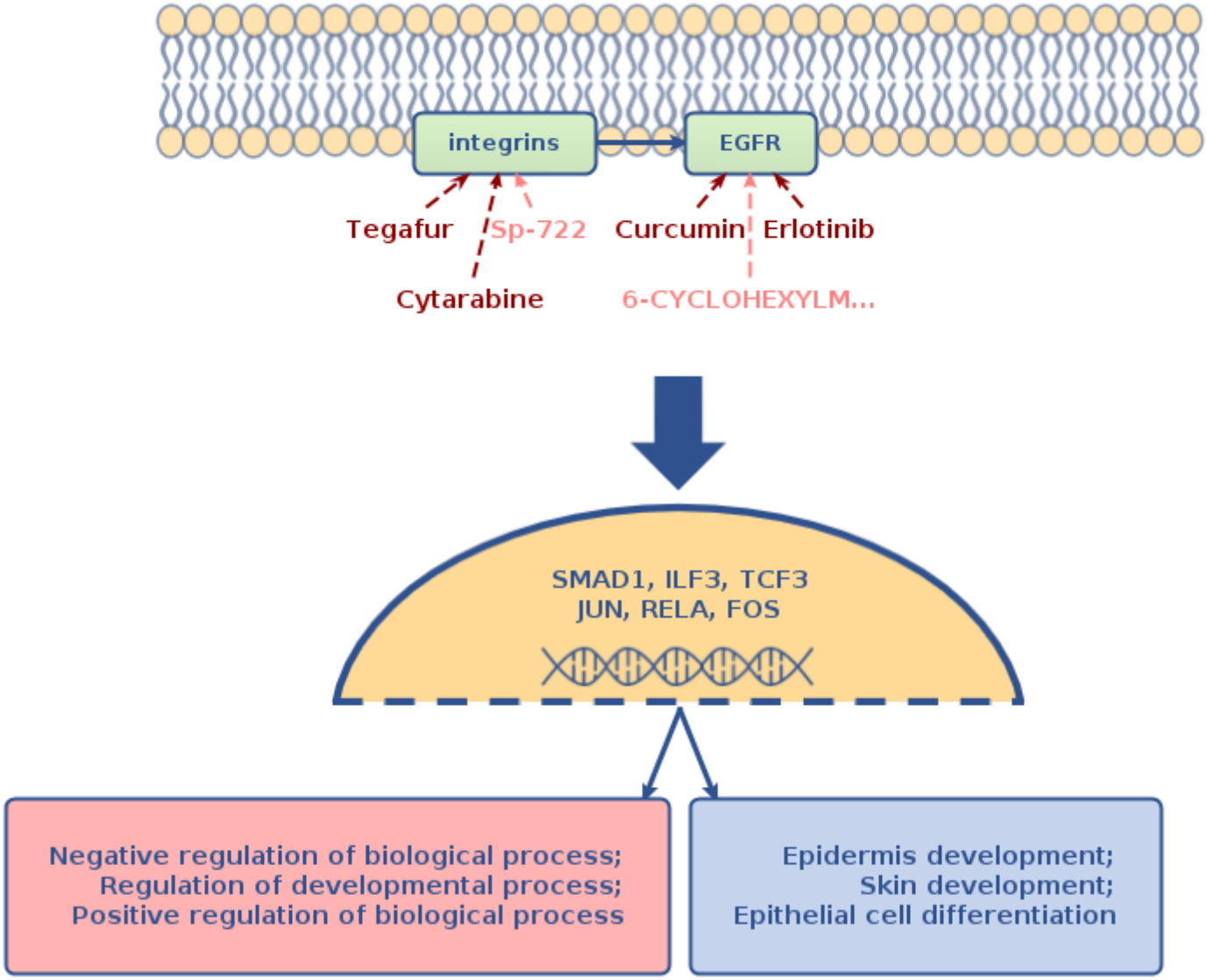
See full table →

Gene symbol	Gene Description	Druggability score	logFC	Total rank
EGFR	epidermal growth factor receptor	40.43	4.44	173
ITGA3	integrin subunit alpha 3	6.21	3.05	185
ITGB5	integrin subunit beta 5	6.21	3.05	185
ITGA6	integrin subunit alpha 6	6.21	3.05	185
ITGB6	integrin subunit beta 6	6.21	3.05	185
ITGB4	integrin subunit beta 4	6.21	3.05	185

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:

- integrins
- 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: Cytarabine, Erlotinib, Tegafur, Sp-722, Curcumin and 6-CYCLOHEXYLMETHOXY-2-(3'-CHLOROANILINO) PURINE, 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	PTPRC, BIRC5, CDKN1A	46	5	small molecule,approved	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
Erlotinib	RPS6KA3, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, PRKAA1, BMPR2, ILK, EGFR, SYK, EIF2AK2, BIRC5, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2, CSNK1G2, EPHB4, YES1, BRAF, STK11, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R	99	4	small molecule,approved,investigational
Tegafur	ITGA6, VEGFA, ITGB5, EGFR, ITGB1, PTK2, ITGB4, ITGA3	99	2	small molecule,approved
Gefitinib	RPS6KA3, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, PRKAA1, BMPR2, EGFR, SYK, EIF2AK2, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2, CSNK1G2, EPHB4, YES1, BRAF, STK11, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R	95	2	small molecule,approved,investigational
Lapatinib	RPS6KA3, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, PRKAA1, BMPR2, EGFR, SYK, EIF2AK2, BIRC5, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2, CSNK1G2, EPHB4, YES1, BRAF, STK11, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R	94	1	small molecule,approved,investigational
Imatinib	RPS6KA3, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, PRKAA1, BMPR2, EGFR, SYK, EIF2AK2, BIRC5, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2, CSNK1G2, EPHB4, CRK, YES1, BRAF, STK11, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R	94	1	small molecule,approved

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
Curcumin	CD44, CDK6, MET, SMAD2, SMAD3, HSPA5, HK2, CCNB1, GSK3B, VIM, CDK4, PTEN, PRKAA1, CDC20, PARP1, EGFR, CTNNB1, SLC2A1, HIF1A, ATM, BIRC5, SUZ12, JAK1, ATR, VEGFA, CEBPA, PCNA, YWHAE, APP, NEDD4, EPAS1, PSEN1, BIRC2, MMP14, CCND1, SKP2, IGFBP5, SMO, CCNA2, CLTC, CDKN1A, JAG1	91	EARLY_PHASE1: Chronic Periodontitis, Hematoma, Hematoma, Subdural, Hematoma, Subdural, Chronic, Hematoma, Subdural, Intracranial, Oral Ulcer, Periodontal Pocket, Periodontitis, Recurrence, Ulcer
seliciclib	RPS6KA3, ROCK2, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, CDK4, PRKAA1, BMPR2, EGFR, SYK, SGK1, EIF2AK2, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2, CSNK1G2, EPHB4, YES1, BRAF, STK11, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R	91	PHASE2: Cystic Fibrosis, Cysts, Fibrosis
Flavopiridol	RPS6KA3, CDK6, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, CDK4, PRKAA1, BMPR2, EGFR, SYK, EIF2AK2, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2, CSNK1G2, EPHB4, YES1, BRAF, STK11, PIK3CB, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R	91	PHASE1: Brain Abscess, Carcinoma, Carcinoma, Non-Small-Cell Lung, Carcinoma, Small Cell, Cytopenia, Esophageal Neoplasms, Intestinal Neoplasms, Leukemia, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, Leukemia, Prolymphocytic, Lymphoma, Lymphoma, B-Cell, Lymphoma, B-Cell, Marginal Zone, Lymphoma, Follicular, Lymphoma, Large B-Cell, Diffuse, Lymphoma, Mantle-Cell, Lymphoma, Non-Hodgkin, Mesothelioma, Mesothelioma, Malignant, Multiple Myeloma, Neoplasms, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Recurrence, Thrombocytopenia, Waldenstrom Macroglobulinemia
Tofacitinib	RPS6KA3, ROCK2, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, PRKAA1, BMPR2, EGFR, SYK, SGK1, EIF2AK2, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2, CSNK1G2, EPHB4, YES1, BRAF, STK11, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R	90	EARLY_PHASE1: Vitiligo
1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea	RPS6KA3, ROCK2, MAP4K4, MET, NEK7, PAK2, GSK3B, LATS1, MAP3K11, PRKAA1, BMPR2, EGFR, SYK, SGK1, EIF2AK2, IGF1R, JAK1, TGFBR1, SRC, CSNK2A2, PRKD3, PIP4K2B, NTRK2,	90	PHASE2: Arthritis, Arthritis, Rheumatoid, Psoriasis

	CSNK1G2, EPHB4, YES1, BRAF, STK11, STK4, PTK2, PIK3CA, LYN, CSNK2A1, PRKACB, CSF1R

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
{(2Z)-4-AMINO-2-[(4-METHOXYPHENYL)IMINO]-2,3-DIHYDRO-1,3-THIAZOL-5-YL}(4-METHOXYPHENYL)METHANONE	CCND1, CDK6, CCND3, CCNB1, CCNA2, CCNT1, CCNB2, CDK4	100	4.63
3-Bromo-7-Nitroindazole	RPS6KA3, CCND1, CDK6, HSPD1, CCND3, CCNB1, GSK3B, CCNA2, CCNT1, CCNB2, CDK4	100	3.29
O6-CYCLOHEXYLMETHOXY-2-(4'-SULPHAMOYLANILINO) PURINE	CCND1, CDK6, CCND3, CCNB1, CCNA2, CCNT1, CDK4, CCNB2	100	3.03
2-ANILINO-6-CYCLOHEXYLMETHOXYPURINE	CCND1, CDK6, CCND3, CCNB1, CCNA2, CCNT1, CDK4, CCNB2	100	2.85
2-(2-HYDROXYETHYLAMINO)-6-(3-CHLOROANILINO)-9-ISOPROPYLPURINE	CCND1, CDK6, SRC, CCND3, CCNB1, CCNA2, CCNT1, CDK4, CCNB2	100	2.79

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Erlotinib, Curcumin and {(2Z)-4-AMINO-2-[(4-METHOXYPHENYL)IMINO]-2,3-DIHYDRO-1,3-THIAZOL-5-YL}(4-METHOXYPHENYL)METHANONE. These drugs were selected for acting on the following targets: EGFR and CDK6, 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	89
Abiraterone	Prostatic Neoplasms, Castration-Resistant	-
Abiraterone acetate	Prostatic Neoplasms, Castration-Resistant	-
Acalabrutinib	Lymphoma, Mantle-Cell	-
Acitretin	Psoriasis	-
Ado-trastuzumab emtansine	Breast Neoplasms|Neoplasms	42
Afatinib	Carcinoma, Non-Small-Cell Lung	68
Aflibercept	Colorectal Neoplasms|Diabetic Retinopathy|Edema|Vascular Diseases|Wet Macular Degeneration	32
Alectinib	Carcinoma, Non-Small-Cell Lung	1
Alemtuzumab	Brain Abscess|Leukemia, Lymphocytic, Chronic, B-Cell|Multiple Sclerosis|Multiple Sclerosis, Relapsing-Remitting|Sclerosis	-
Alitretinoin	Sarcoma, Kaposi	-
Alpelisib	Breast Neoplasms	80
Altretamine	Ovarian Neoplasms	-
Aminolevulinic acid	Keratosis|Keratosis, Actinic	-
Anagrelide	Thrombocythemia, Essential|Thrombocytosis	-
Anastrozole	Breast Neoplasms|Hypersensitivity|Obesity|Obesity, Morbid|Recurrence|Weight Loss	-
Apalutamide	Prostatic Neoplasms, Castration-Resistant	-
Aprepitant	Nausea|Neoplasms|Postoperative Nausea and Vomiting	-
Arsenic trioxide	Leukemia, Promyelocytic, Acute	82
Atezolizumab	Carcinoma, Non-Small-Cell Lung|Carcinoma, Transitional Cell|Triple Negative Breast Neoplasms	-
Avelumab	Carcinoma, Merkel Cell|Carcinoma, Renal Cell|Carcinoma, Transitional Cell	-
Axitinib	Carcinoma, Renal Cell	-
Azacitidine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes|Preleukemia|Syndrome	13
Belinostat	Lymphoma, T-Cell, Peripheral	35
Bendamustine	Leukemia, Lymphocytic, Chronic, B-Cell|Leukemia, Lymphoid	-
Bevacizumab	Breast Neoplasms|Colonic Neoplasms|Colorectal Neoplasms|Corneal Neovascularization|Diabetic Retinopathy|Dilatation, Pathologic|Edema|Epistaxis|Glaucoma|Hemorrhage|Macular Degeneration|Macular Edema|Neoplasm Metastasis|Neoplasms|Neovascularization, Pathologic|Optic Nerve Diseases|Pterygium|Rectal Neoplasms|Retinal Detachment|Retinal Diseases|Retinal Vein Occlusion|Telangiectasia, Hereditary Hemorrhagic|Telangiectasis|Vitreous Hemorrhage	25
Bexarotene	Lymphoma, T-Cell|Lymphoma, T-Cell, Cutaneous	-
Bicalutamide	Prostatic Neoplasms	9
Binimetinib	Melanoma	-
Blinatumomab	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Bortezomib	Brain Abscess|Glomerulonephritis|Glomerulonephritis, IGA|Kidney Diseases|Multiple Myeloma|Neoplasms, Plasma Cell|Nephritis|Renal Insufficiency	22
Bosutinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	71
Brentuximab vedotin	Hodgkin Disease|Lymphoma|Lymphoma, Large-Cell, Anaplastic|Lymphoma, T-Cell, Peripheral	-
Brigatinib	Carcinoma, Non-Small-Cell Lung	69
Buserelin	Prostatic Neoplasms	-
Cabazitaxel	Prostatic Neoplasms, Castration-Resistant	17
Cabergoline	Drug-Related Side Effects and Adverse Reactions|Pituitary Neoplasms	-

Cabozantinib	Thyroid Neoplasms	53
Capecitabine	Breast Neoplasms|Colonic Neoplasms|Colorectal Neoplasms	-
Carboplatin	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms|Neuroendocrine Tumors|Ovarian Neoplasms|Retinoblastoma	-
Carfilzomib	Multiple Myeloma	12
Carmustine	Astrocytoma|Glioblastoma|Hodgkin Disease|Medulloblastoma|Multiple Myeloma|Neoplasms	-
Ceritinib	Carcinoma, Non-Small-Cell Lung	62
Cetuximab	Colorectal Neoplasms	53
Cinacalcet	Anemia|Calcinosis|Cardiovascular Diseases|Hyperparathyroidism|Hyperparathyroidism, Secondary|Kidney Diseases|Kidney Failure, Chronic|Neoplasm Metastasis|Neoplasms|Parathyroid Neoplasms|Renal Insufficiency|Vascular Calcification|Vascular Diseases|Vision Disorders	-
Cisplatin	Carcinoma, Squamous Cell|Neoplasms|Uterine Cervical Neoplasms|Carcinoma, Non-Small-Cell Lung|Esophageal Neoplasms|Carcinoma	-
Cladribine	Leukemia, Hairy Cell	20
Clofarabine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	51
Cobimetinib	Melanoma	-
Copanlisib	Lymphoma, Follicular	65
Crizotinib	Carcinoma, Non-Small-Cell Lung	92
Cyproterone acetate	Prostatic Neoplasms	-
Dabrafenib	Melanoma	13
Dacomitinib	Carcinoma, Non-Small-Cell Lung	87
Daratumumab	Multiple Myeloma	-
Dasatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase|Precursor Cell Lymphoblastic Leukemia-Lymphoma	88
Decitabine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes	-
Degarelix	Cardiovascular Diseases|Prostatic Neoplasms|Vascular Diseases	-
Denosumab	Arthritis, Rheumatoid|Bone Diseases|Bone Diseases, Metabolic|Breast Neoplasms|Hyperparathyroidism|Hyperparathyroidism, Primary|Metabolic Diseases|Neoplasm Metastasis|Neoplasms|Osteoporosis|Osteoporosis, Postmenopausal|Prostatic Neoplasms	-
Dexrazoxane	Breast Neoplasms|Cardiomyopathies	63
Dienogest	Menorrhagia	-
Dinutuximab	Neuroblastoma	-
Docetaxel	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Prostatic Neoplasms|Squamous Cell Carcinoma of Head and Neck|Stomach Neoplasms	-
Doxorubicin	Neoplasms|Multiple Myeloma|Carcinoma, Ovarian Epithelial|Ovarian Neoplasms|Leukemia, Lymphoid|Breast Neoplasms|Lymphoma, Follicular|Thyroid Neoplasms|Triple Negative Breast Neoplasms|Glioma	86
Durvalumab	Carcinoma, Non-Small-Cell Lung|Carcinoma, Transitional Cell	-
Dutasteride	Alcoholism|Hyperplasia|Hypertrophy|Neoplasms|Prostatic Hyperplasia	-
Duvelisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	-
Elotuzumab	Multiple Myeloma	-
Enasidenib	Leukemia, Myeloid, Acute	-
Encorafenib	Colorectal Neoplasms|Melanoma	6
Enfortumab vedotin	Carcinoma, Transitional Cell|Neoplasms	-
Entrectinib	Carcinoma, Non-Small-Cell Lung	28
Enzalutamide	Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	-
Epirubicin	Breast Neoplasms	73
Erdafitinib	Urinary Bladder Neoplasms	45
Eribulin	Breast Neoplasms|Drug-Related Side Effects and Adverse Reactions|Neoplasms	-
Erlotinib	Carcinoma, Non-Small-Cell Lung|Neoplasms|Pancreatic Neoplasms	99
Erlotinib hydrochloride	Carcinoma, Non-Small-Cell Lung|Gastrointestinal Stromal Tumors	-

Estramustine	Prostatic Neoplasms	-
Ethinyl Estradiol	Acne Vulgaris|Neoplasms	9
Everolimus	Angiomyolipoma|Arthrogryposis|Astrocytoma|Breast Neoplasms|Carcinoma, Renal Cell|Cysts|Idiopathic Pulmonary Fibrosis|Kidney Diseases, Cystic|Kidney Failure, Chronic|Lipoma|Neuroendocrine Tumors|Primary Graft Dysfunction|Sclerosis|Tuberous Sclerosis	66
Exemestane	Breast Neoplasms	-
Fedratinib	Primary Myelofibrosis	-
Finasteride	Hyperplasia|Neoplasms|Prostatic Hyperplasia	-
Flavopiridol	Leukemia, Lymphocytic, Chronic, B-Cell	91
Fluorouracil	Skin Neoplasms|Neoplasms, Basal Cell|Neoplasms, Second Primary|Neoplasms, Squamous Cell|Neoplasms|Colorectal Neoplasms|Pancreatic Neoplasms	46
Fluoxymesterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	23
Flutamide	Premenstrual Dysphoric Disorder|Premenstrual Syndrome|Prostatic Neoplasms	60
Fulvestrant	Breast Neoplasms	-
Gefitinib	Carcinoma, Non-Small-Cell Lung	95
Gemcitabine	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Ovarian Neoplasms|Pancreatic Neoplasms	16
Gemtuzumab ozogamicin	Leukemia, Myeloid, Acute	-
Gilteritinib	Leukemia, Myeloid, Acute	60
Glasdegib	Leukemia, Myeloid, Acute	64
Goserelin	Atrophy|Breast Neoplasms|Bulbo-Spinal Atrophy, X-Linked|Endometriosis|Muscular Atrophy|Myoma|Prostatic Neoplasms	-
Histrelin	Puberty, Precocious	-
Homoharringtonine	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	61
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	25
Idarubicin	Leukemia, Myeloid, Acute	37
Idelalisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	2
Ifosfamide	Neoplasms	-
Imatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Mastocytosis, Systemic|Neoplasms	94
Inotuzumab ozogamicin	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Ipilimumab	Carcinoma, Renal Cell|Melanoma	-
Irinotecan	Colorectal Neoplasms	69
Ivosidenib	Leukemia, Myeloid, Acute	-
Ixabepilone	Breast Neoplasms	-
Ixazomib	Multiple Myeloma	-
Lapatinib	Breast Neoplasms	94
Larotrectinib	Neoplasm Metastasis	39
Lenalidomide	Brain Abscess|Lupus Erythematosus, Cutaneous|Myelodysplastic Syndromes|Neoplasms, Plasma Cell	17
Lenvatinib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	-
Letrozole	Breast Neoplasms|Cysts|Fibroma|Myofibroma|Myoma|Ovarian Cysts|Syndrome	-
Leuprolide	Hot Flashes|Ovarian Hyperstimulation Syndrome|Prostatic Neoplasms|Puberty, Precocious	-
Levamisole	Ascariasis|Colonic Neoplasms|Helminthiasis	-
Levonorgestrel	Epilepsy|Hyperplasia|Menorrhagia	-
Lomustine	Brain Neoplasms|Hodgkin Disease	-
Lonafarnib	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Central Nervous System Neoplasms|Colorectal Neoplasms|Gliosarcoma|Head and Neck Neoplasms|Leukemia, Myelomonocytic, Chronic| Liver Neoplasms|Lymphoma|Myelodysplastic Syndromes|Ovarian Neoplasms|Urethral Neoplasms|Urinary Bladder Neoplasms	45

Lorlatinib	Carcinoma, Non-Small-Cell Lung	76
Masoprocol	Keratosis, Actinic	-
Medroxyprogesterone Acetate	Depression|Depression, Postpartum|Depressive Disorder|Metrorrhagia|Neoplasms|Uterine Hemorrhage	14
Megestrol acetate	Acquired Immunodeficiency Syndrome|Bites and Stings|Breast Neoplasms|Pain|Wasting Syndrome	17
Methotrexate	Neoplasms|Breast Neoplasms|Head and Neck Neoplasms|Ovarian Neoplasms|Lymphoma, T-Cell, Peripheral|Brain Neoplasms|Colorectal Neoplasms|Neuroblastoma|Carcinoma, Squamous Cell	40
Methyltestosterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	-
Midostaurin	Leukemia, Mast-Cell|Leukemia, Myeloid, Acute|Mastocytosis, Systemic	89
Mitotane	Adrenocortical Carcinoma	-
Mitoxantrone	Autoimmune Diseases|Autoimmune Diseases of the Nervous System|Demyelinating Autoimmune Diseases, CNS|Immune System Diseases|Leukemia, Myeloid, Acute|Multiple Sclerosis|Myelitis|Myelitis, Transverse|Nervous System Diseases|Neuromyelitis Optica|Prostatic Neoplasms, Castration-Resistant	45
Mogamulizumab	Mycosis Fungoides|Neoplasms|Sezary Syndrome	-
Moxetumomab pasudotox	Leukemia, Hairy Cell|Neoplasms	-
Necitumumab	Carcinoma, Non-Small-Cell Lung|Neoplasms	-
Nelarabine	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma	-
Neratinib	Breast Neoplasms	73
Nilotinib	Blast Crisis|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase	12
Nilutamide	Prostatic Neoplasms	-
Nintedanib	Fibrosis|Idiopathic Pulmonary Fibrosis	62
Niraparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms	54
Nivolumab	Carcinoma, Non-Small-Cell Lung|Kidney Neoplasms|Neoplasms|Lung Neoplasms|Melanoma	-
Obinutuzumab	Leukemia, Lymphocytic, Chronic, B-Cell	-
Octreotide	Acromegaly|Adenoma|Ascites|Carcinoid Tumor|Fistula|Pancreatic Fistula|Pituitary Diseases|Renal Insufficiency|Vipoma	-
Ofatumumab	Leukemia, Lymphocytic, Chronic, B-Cell	-
Olaparib	Breast Neoplasms|Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Ovarian Neoplasms|Pancreatic Neoplasms|Peritoneal Neoplasms|Prostatic Neoplasms, Castration-Resistant	56
Olaratumab	Sarcoma	-
Osimertinib	Carcinoma, Non-Small-Cell Lung	36
Oxaliplatin	Colonic Neoplasms|Colorectal Neoplasms|Neoplasms|Rectal Neoplasms	51
Paclitaxel	Acute Coronary Syndrome|Angina Pectoris|Arteriosclerosis|Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Cardiovascular Diseases|Coronary Artery Disease|Coronary Disease|Coronary Stenosis|Heart Diseases|Myocardial Ischemia|Ovarian Neoplasms|Vascular Diseases	65
Palbociclib	Breast Neoplasms	79
Panitumumab	Colorectal Neoplasms	87
Panobinostat	Multiple Myeloma	-
Pazopanib	Carcinoma|Carcinoma, Renal Cell|Sarcoma	89
Pembrolizumab	Carcinoma, Hepatocellular|Carcinoma, Merkel Cell|Carcinoma, Non-Small-Cell Lung|Carcinoma, Renal Cell|Carcinoma, Transitional Cell|Hodgkin Disease|Melanoma|Neoplasms|Stomach Neoplasms	-
Pemetrexed	Carcinoma, Non-Small-Cell Lung|Mesothelioma	-
Pentostatin	Leukemia, Hairy Cell	27
Pertuzumab	Breast Neoplasms	-
Pomalidomide	Multiple Myeloma	20
Ponatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Precursor Cell Lymphoblastic Leukemia-Lymphoma	80

Pralatrexate	Lymphoma, T-Cell, Peripheral	-
Radium Ra 223 Dichloride	Prostatic Neoplasms, Castration-Resistant	-
Ramucirumab	Stomach Neoplasms	-
Rasburicase	Hyperuricemia|Leukemia|Lymphoma|Neoplasms|Syndrome|Tumor Lysis Syndrome	-
Regorafenib	Colorectal Neoplasms	27
Relugolix	Prostatic Neoplasms	-
Ribociclib	Breast Neoplasms	86
Rituximab	Arthritis|Arthritis, Rheumatoid|Granulomatosis with Polyangiitis|Leukemia|Leukemia, Lymphoid|Lymphoma|Lymphoma, B-Cell|Lymphoma, Follicular|Lymphoma, Non-Hodgkin|Myelitis|Neuromyelitis Optica|Purpura|Purpura, Thrombocytopenic|Purpura, Thrombocytopenic, Idiopathic|Thrombocytopenia	-
Romidepsin	Lymphoma, T-Cell, Cutaneous	12
Rucaparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms|Prostatic Neoplasms, Castration-Resistant	57
Ruxolitinib	Graft vs Host Disease|Polycythemia|Polycythemia Vera|Primary Myelofibrosis|Thrombocytosis	39
Selinexor	Multiple Myeloma	79
Selumetinib	Neurofibromatosis 1	-
Siltuximab	Giant Lymph Node Hyperplasia	-
Sirolimus	Angiomyolipoma|Constriction, Pathologic|Coronary Restenosis|Eye Diseases|Immune System Diseases|Kidney Failure, Chronic|Lipoma|Tuberous Sclerosis	82
Sonidegib	Carcinoma, Basal Cell	53
Sorafenib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	90
Sunitinib	Adenoma|Carcinoma, Renal Cell|Digestive System Neoplasms|Gastrointestinal Neoplasms|Gastrointestinal Stromal Tumors|Intestinal Neoplasms	87
Talazoparib	Breast Neoplasms	43
Tamoxifen	Breast Diseases|Cystic Fibrosis|Cysts|Fibroadenoma|Fibrocystic Breast Disease|Hemorrhage|Menorrhagia|Menstruation Disturbances|Metrorrhagia|Neoplasms	57
Tamsulosin	Calculi|Coronary Artery Disease|Heart Diseases|Hernia|Hernia, Inguinal|Inflammation|Ischemia|Lithiasis|Lower Urinary Tract Symptoms|Myocardial Ischemia|Prostatic Hyperplasia|Ureteral Calculi|Urinary Calculi|Urolithiasis|Urologic Diseases	-
Temozolomide	Astrocytoma|Nervous System Neoplasms	16
Temsirolimus	Carcinoma, Renal Cell	78
Teniposide	Precursor Cell Lymphoblastic Leukemia-Lymphoma	76
Thalidomide	Brain Abscess|Immune System Diseases|Multiple Myeloma|Neoplasms, Plasma Cell	-
Tivozanib	Carcinoma, Renal Cell	-
Tocilizumab	Arthritis|Arthritis, Juvenile|Arthritis, Rheumatoid|Behavior|Cytokine Release Syndrome|Giant Cell Arteritis|Neurobehavioral Manifestations|Oral Manifestations|Psychotic Disorders|Schizophrenia|Tic Disorders	-
Topotecan	Small Cell Lung Carcinoma	-
Toremifene	Breast Neoplasms	-
Trabectedin	Leiomyosarcoma|Liposarcoma	-
Trametinib	Carcinoma, Non-Small-Cell Lung|Melanoma	20
Trastuzumab	Breast Neoplasms|Neoplasms	42
Tretinoin	Lentigo	83
Triptorelin	Fatty Liver|Hypogonadism|Infertility, Female|Prostatic Neoplasms	13
Tucatinib	Breast Neoplasms	-
Valrubicin	Urinary Bladder Neoplasms	78
Vandetanib	Thyroid Neoplasms	94
Vemurafenib	Melanoma	7

Venetoclax	Leukemia, Lymphocytic, Chronic, B-Cell|Leukemia, Myeloid, Acute	-
Vinblastine	Glioma	16
Vincristine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	3
Vinorelbine	Carcinoma, Non-Small-Cell Lung	11
Vismodegib	Carcinoma, Basal Cell	53
Vorinostat	Lymphoma, T-Cell, Cutaneous	12
Zoledronate	Arthritis|Bone Marrow Diseases|Brain Abscess|Chronic Kidney Disease-Mineral and Bone Disorder|Chronic Periodontitis|HIV Infections|Hypersensitivity|Infections|Kidney Diseases|Metabolic Diseases|Multiple Myeloma|Neoplasms|Neoplasms, Plasma Cell|Neoplasms, Second Primary|Osteitis|Osteoarthritis|Periodontitis|Pleural Effusion, Malignant|Prostatic Neoplasms|Renal Insufficiency, Chronic|Thalassemia|Wounds and Injuries	-

6. Conclusion

We applied the software package "Genome Enhancer" to a 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:

Erlotinib, Curcumin and {(2Z)-4-AMINO-2-[(4-METHOXYPHENYL)IMINO]-2,3-DIHYDRO-1,3-THIAZOL-5-YL}(4-METHOXYPHENYL)METHANONE

These drugs were selected for acting on the following targets: EGFR and CDK6, 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:

integrins 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: Cytarabine, Erlotinib, Tegafur, Sp-722, Curcumin and 6-CYCLOHEXYLMETHOXY-2-(3'-CHLOROANILINO) PURINE. 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.

- integrins
- 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 2024.2 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac).
The master regulator search uses the TRANSPATH® database (BIOBASE), release 2024.2 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transpath). A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in TRANSPATH®.
The information about drugs corresponding to identified drug targets and clinical trials references were extracted from HumanPSD™ database, release 2024.2 (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 differentially expressed genes were analyzed using known DNA-binding motifs. The motifs are specified using position weight matrices (PWMs) that give weights to each nucleotide in each position of the DNA binding motif for a transcription factor or a group of them.
We search for transcription factor binding sites (TFBS) that are enriched in the promoters and enhancers under study as compared to a background sequence set such as promoters of genes that were not differentially regulated under the condition of the experiment. We denote study and background sets briefly as Yes and No sets. In the current work we used a workflow considering promoter sequences of a standard length of 1100 bp (-1000 to +100). The error rate in this part of the pipeline is controlled by estimating the adjusted p-value (using the Benjamini-Hochberg procedure) in comparison to the TFBS frequency found in randomly selected regions of the human genome (adj.p-value < 0.01).
We have applied the CMA algorithm (Composite Module Analyst) for searching composite modules [7] in the promoters and enhancers of the Yes and No sets. We searched for a composite module consisting of a cluster of 10 TFs in a sliding window of 200-300 bp that statistically significantly separates sequences in the Yes and No sets (minimizing Wilcoxon p-value).

Methods for finding master regulators in networks

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

Methods for analysis of pharmaceutical compounds

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

Method for analysis of known pharmaceutical compounds

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

1. ranking by "Target activity score" (*T-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. This library contains 13040 compounds along with their pre-calculated potential pharmacological activities of those substances, their possible side and toxic effects, as well as the possible mechanisms of action. All biological activities are expressed as probability values for a substance to exert this activity (*Pa*).

We selected compounds that satisfied the following conditions:

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

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

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

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

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

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

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

8. References

1. Kel A, Voss N, Jauregui R, Kel-Margoulis O, Wingender E. Beyond microarrays: Finding key transcription factors controlling signal transduction pathways. *BMC Bioinformatics.* **2006**;7(S2), S13. doi:10.1186/1471-2105-7-s2-s13
2. Stegmaier P, Voss N, Meier T, Kel A, Wingender E, Borlak J. Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer. *PLoS ONE.* **2011**;6(3):e17738. doi:10.1371/journal.pone.0017738
3. Koschmann J, Bhar A, Stegmaier P, Kel A, Wingender E. "Upstream Analysis": An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays.* **2015**;4(2):270-286. doi:10.3390/microarrays4020270.
4. Kel A, Stegmaier P, Valeev T, Koschmann J, Poroikov V, Kel-Margoulis OV, and Wingender E. Multi-omics "upstream analysis" of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer. *EuPA Open Proteom.* **2016**;13:1-13. doi:10.1016/j.euprot.2016.09.002
5. Michael H, Hogan J, Kel A et al. Building a knowledge base for systems pathology. *Brief Bioinformatics*. **2008**;9(6):518-531. doi:10.1093/bib/bbn038
6. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **2006**;34(90001):D108-D110. doi:10.1093/nar/gkj143
7. Kel AE, Gössling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* **2003**;31(13):3576-3579. doi:10.1093/nar/gkg585
8. Waleev T, Shtokalo D, Konovalova T, Voss N, Cheremushkin E, Stegmaier P, Kel-Margoulis O, Wingender E, Kel A. Composite Module Analyst: identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res.* **2006**;34(Web Server issue):W541-5.
9. Krull M, Pistor S, Voss N, Kel A, Reuter I, Kronenberg D, Michael H, Schwarzer K, Potapov A, Choi C, Kel-Margoulis O, Wingender E. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.* **2006**;34(90001):D546-D551. doi:10.1093/nar/gkj107
10. 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
11. 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
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:10.1002/chin.199940210

Thank you for using the Genome Enhancer!

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

Supplementary material

1. Supplementary table 1 - Up-regulated genes
2. Supplementary table 2 - Down-regulated genes
3. Supplementary table 3 - Detailed report. Composite modules and master regulators (up-regulated genes in Experiment: 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.