EPHA2 and CREBBP are promising druggable targets for treating Colorectal Neoplasms that control activity of EP300, ESR1 and RXRA transcription factor on promoters of genes carrying sequence variations

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 30/10/2024 ; Run on 18/11/2024 ; Report generated on 20/11/2024

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

From the data set analyzed in this study, we found the following TFs to be potentially involved in the regulation of the genes carrying sequence variations: EP300, ESR1 and RXRA. The subsequent network analysis suggested

- Eck
- PML
- Cdk8:cyclinC:hN1:SNW1:RBP-Jkappa:MAML:PCAF:p300:CREBBP

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: Sunitinib, seliciclib, LE-SN38 and 2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One.

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 genes carrying sequence variations 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
CRC_variants	Genomics

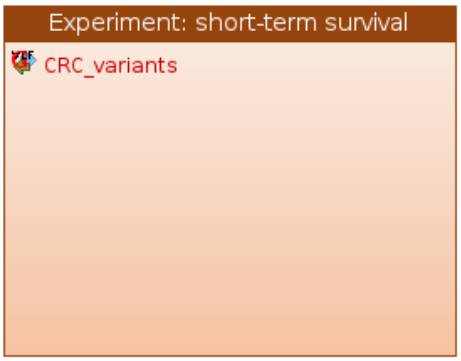


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

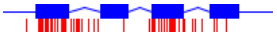
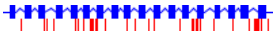
3. Results

We have analyzed the following condition: Experiment: short-term survival.

3.1. Identification of target genes

In the first step of the analysis ***target genes*** were identified from the uploaded experimental data. The most frequently mutated genes were used as target genes.

Table 2. Top ten the most frequently mutated genes in Experiment: short-term survival.
See full table →

ID	Gene description	Gene symbol	Gene schematic representation	Number of variations	Gene weight	Weighted score
ENSG00000132570	pterin-4 alpha-carbinolamine dehydratase 2	PCBD2		172	455.67	683.5
ENSG00000169894	mucin 3A, cell surface associated	MUC3A		68	184.38	553.15
ENSG00000206503	major histocompatibility complex, class I, A	HLA-A		73	182.66	547.98
ENSG00000204525	major histocompatibility complex, class I, C	HLA-C		71	168.72	506.16
ENSG00000196735	major histocompatibility complex, class II, DQ alpha 1	HLA-DQA1		63	146.15	438.45
ENSG00000228716	dihydrofolate reductase	DHFR		56	138.67	416.02
ENSG00000008710	polycystin 1, transient receptor potential channel interacting	PKD1		50	134.06	402.17
ENSG00000111700	solute carrier organic anion transporter family member 1B3	SLCO1B3		51	126.82	380.46
ENSG00000242086	MUC20 overlapping transcript	MUC20-OT1		147	373.28	373.28
ENSG00000234745	major histocompatibility complex, class I, B	HLA-B		51	124.41	373.22

3.2. Functional classification of genes

A functional analysis of genes carrying sequence variations was done by mapping the genes to several known ontologies, such as Gene Ontology (GO), disease ontology (based on HumanPSD™ database) and the ontology of signal transduction and metabolic pathways from the TRANSPATH® database. Statistical significance was computed using a binomial test.
Figures 2-4 show the most significant categories.

The most frequently mutated genes in Experiment: short-term survival:

300 top mutated genes were taken for the mapping.

GO (biological process)

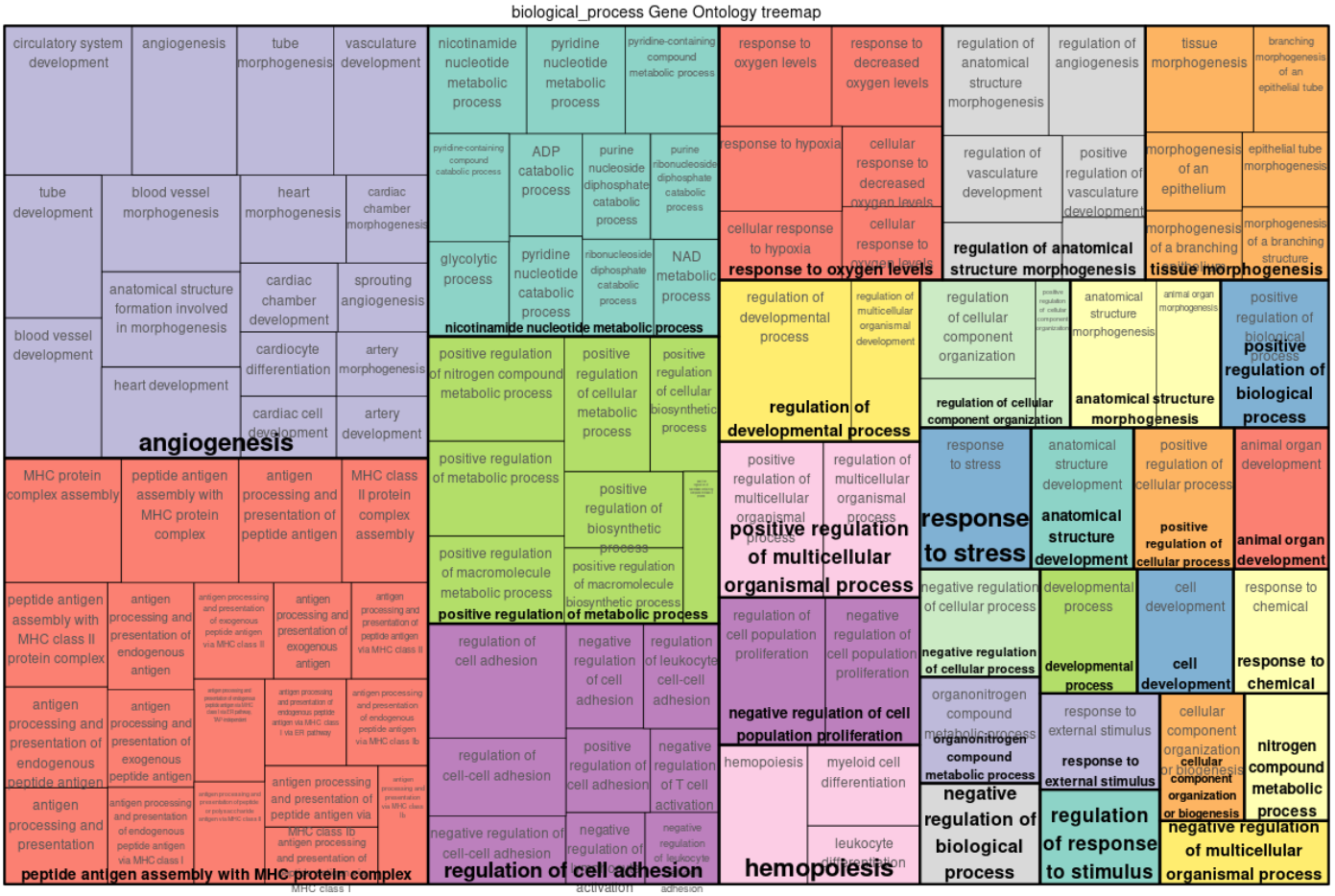


Figure 2. Enriched GO (biological process) of the most frequently mutated genes in Experiment: short-term survival.

Full classification →

TRANSPATH® Pathways (2024.2)

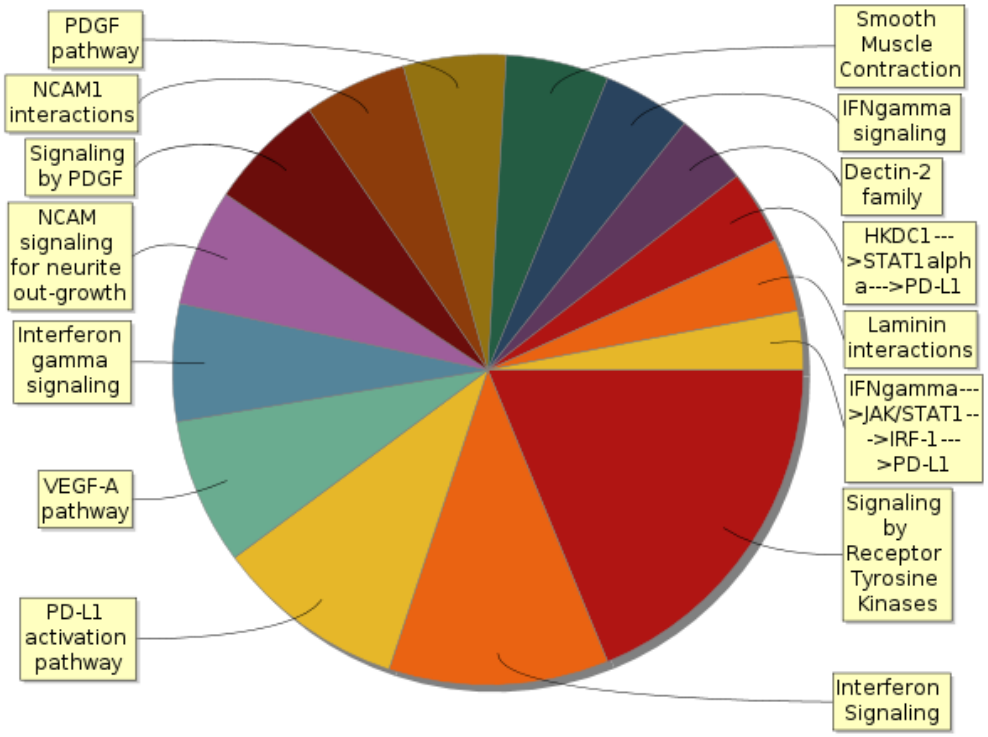


Figure 3. Enriched TRANSPATH® Pathways (2024.2) of the most frequently mutated genes in Experiment: short-term survival.
Full classification →

HumanPSD(TM) disease (2024.2)

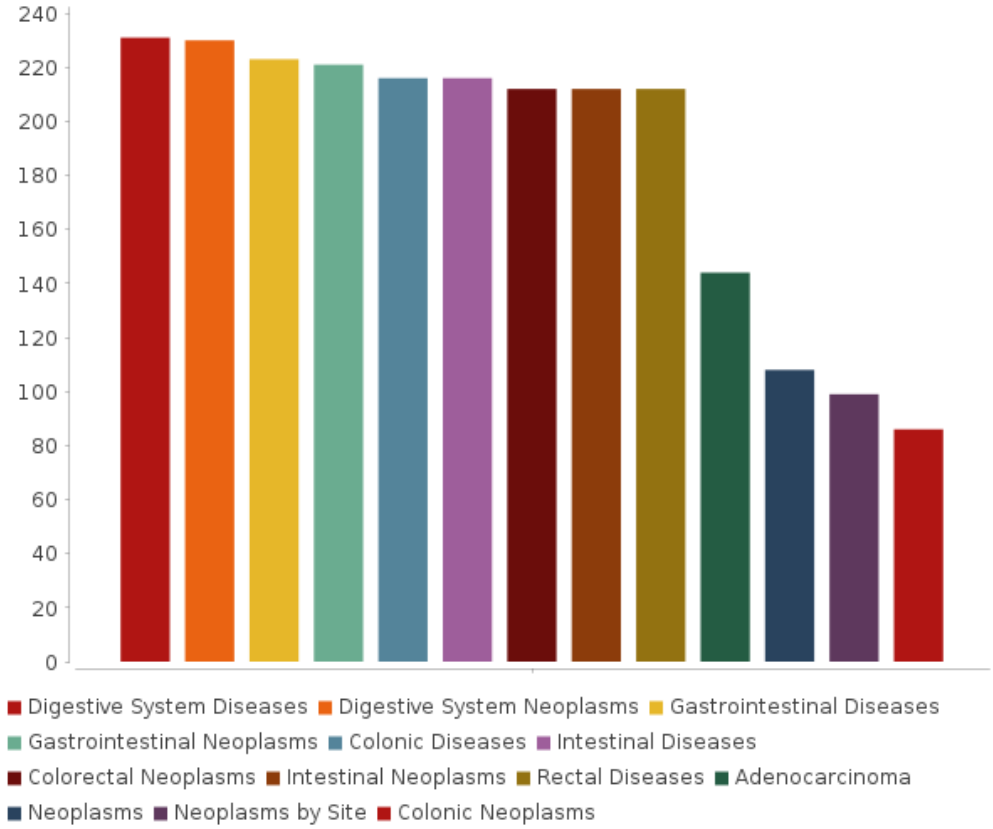
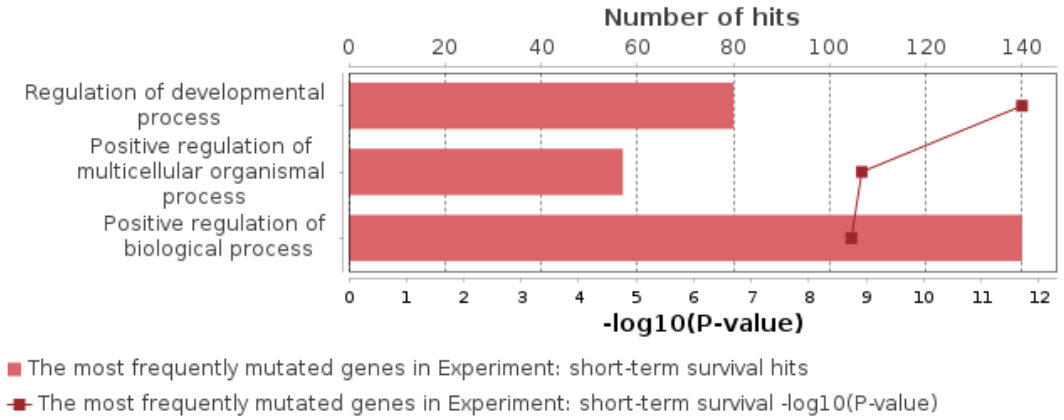


Figure 4. Enriched HumanPSD(TM) disease (2024.2) of the most frequently mutated genes in Experiment: short-term survival. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

Full classification →

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



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 3). We identified 24673 mutations potentially affecting gene regulation. Table 4 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 3. Mutations revealed in the most frequently mutated genes

See full table →

ID	Gene symbol	Gene schematic representation	Number of variations
ENSG00000132570	PCBD2		660
ENSG00000248923	MTND5P11		459
ENSG00000247627	MTND4P12		374
ENSG00000293331	ENSG00000293331		360
ENSG00000249119	MTND6P4		279
ENSG00000242086	MUC20-OT1		252
ENSG00000198868	MTND4LP30		245
ENSG00000263963	ENSG00000263963		245
ENSG00000154237	LRRK1		230
ENSG00000244921	MTCYBP18		225

Table 4. PWMs whose sites were lost or gained due to mutations in the most frequently mutated genes

See full table →

ID	P-value (gains)	P-value (losses)	yesCount (gains)	yesCount (losses)
V$E2F1_Q3_01	4.7E-2	0	8	2397
V$ETV5CLOCK_01	3.21E-2	6.41E-153	27	15591
V$HIF2A_Q6	2.05E-2	5.23E-202	4	2084
V$E2F8_03	7.81E-3	0	3	11880
V$ARNT_06	6.15E-5	6.13E-221	119	2399
V$HIF1A_Q6	1.73E-18	4.49E-181	199	1710
V$ZBED1_01	1.34E-134	3.87E-246	3373	3191
V$MECP2_01	8.87E-138	2.68E-172	1520	1480
V$ERFETV7_01	7.43E-138	7.52E-202	10975	11613
V$E2F7_01	2.66E-150	2.19E-152	1180	1242
V$E2F6_05	3.44E-215		5855	
V$FOXO4_05	8.8E-218		11900	
V$GCM_Q2	8.99E-237		7376	
V$SREBP_Q3	3.85E-274		32742	
V$GCM1TBX21_01	1.38E-275	6.02E-17	6914	1446
V$TFDP1_03	3.19E-276	3.67E-206	6170	10943
V$SP4_Q5	1.44E-276	4.58E-2	3600	2
V$TEAD4MAX_02	6.17E-300		22355	
V$SP2_05	3.24E-308	4.58E-2	3341	2
V$E2F1_Q6	0	5.25E-9	17364	33

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 (the most frequently mutated genes in Experiment: short-term survival).

To build the most specific composite modules we choose top mutated genes as the input of CMA algorithm. The obtained CMA model is then applied to compute CMA score for all the most frequently mutated genes in Experiment: short-term survival.

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$EP300_08	V$FOXO4_05	V$NEUROD_02	V$LXRA_02	V$OSF2_Q6	V$RXRRAR_01
0.79; N=3	0.85; N=2	0.96; N=3	0.78; N=3	0.89; N=3	0.76; N=2

Module width: 138

Module 2:

V$SREBP_Q3	V$AML3_Q3	V$HAIRYLIKE_Q3	V$NR3C1_03	V$ESR1_12
0.85; N=3	0.92; N=3	0.97; N=3	0.80; N=2	0.72; N=2

Module width: 165

Model score (-p*log10(pval)): 18.85
Wilcoxon p-value (pval): 2.00e-39
Penalty (p): 0.487
Average yes-set score: 8.64
Average no-set score: 6.59
AUC: 0.74
Separation point: 8.26
False-positive: 19.72%
False-negative: 42.00%

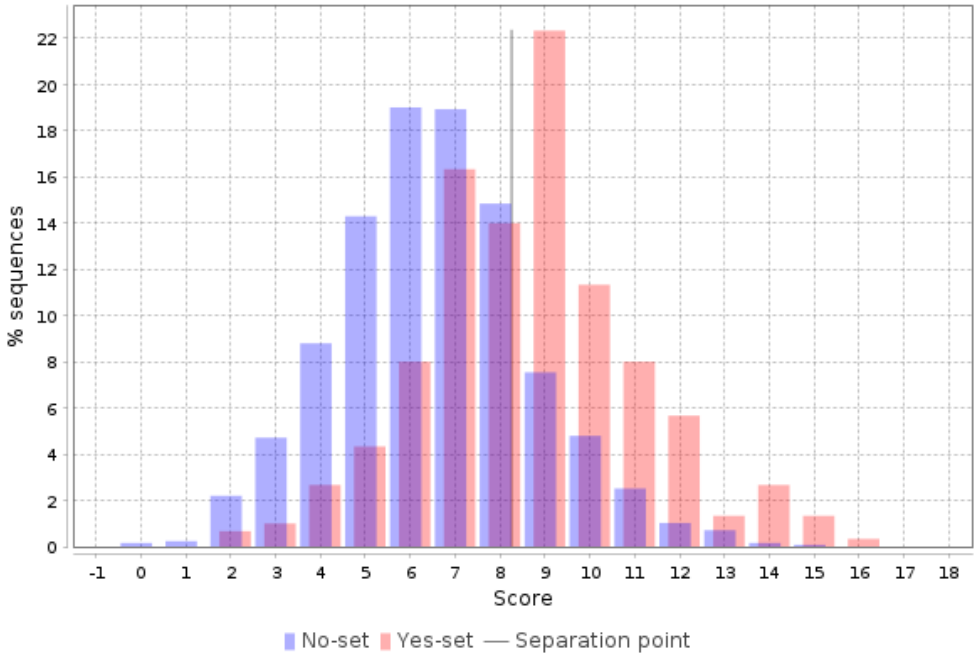


Table 5. List of top ten the most frequently mutated genes in Experiment: short-term survival with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000008300	CELSR3	cadherin EGF LAG seven-pass G-type receptor 3	17.46	Runx2(h), NeuroD1(h), foxo4(h), LXRalpha(h), p300(h), RAR-alpha(h),RXRalpha(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h)...
ENSG00000284575	MIR4793	microRNA 4793	17.46	Runx2(h), NeuroD1(h), foxo4(h), LXRalpha(h), p300(h), RAR-alpha(h),RXRalpha(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h)...
ENSG00000205869	KRTAP5-1	keratin associated protein 5-1	17.35	GR(h), SREBP-1(h),SREBP-2(h), ER-alpha(h), Runx2(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h), RAR-alpha(h),RXRalpha(h), foxo4(h)...
ENSG00000131584	ACAP3	ArfGAP with coiled-coil, ankyrin repeat and PH domains 3	17.12	LXRalpha(h), foxo4(h), SREBP-1(h),SREBP-2(h), Runx2(h), NeuroD1(h), p300(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h)...
ENSG00000172889	EGFL7	EGF like domain multiple 7	16.5	p300(h), RAR-alpha(h),RXRalpha(h), SREBP-1(h),SREBP-2(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h), GR(h), ER-alpha(h), LXRalpha(h)...
ENSG00000169692	AGPAT2	1-acylglycerol-3-phosphate O-acyltransferase 2	16.5	p300(h), RAR-alpha(h),RXRalpha(h), SREBP-1(h),SREBP-2(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h), GR(h), ER-alpha(h), LXRalpha(h)...
ENSG00000125826	RBCK1	RANBP2-type and C3HC4-type zinc finger containing 1	16.2	NeuroD1(h), LXRalpha(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h), Runx2(h), p300(h), SREBP-1(h),SREBP-2(h), RAR-alpha(h),RXRalpha(h)...
ENSG00000276580	MIR8055	microRNA 8055	16.09	foxo4(h), NeuroD1(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h), p300(h), Runx2(h), SREBP-1(h),SREBP-2(h), ER-alpha(h)...
ENSG00000148120	AOPEP	aminopeptidase O (putative)	16.08	foxo4(h), SREBP-1(h),SREBP-2(h), DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h), Runx2(h), GR(h), p300(h), ER-alpha(h)...
ENSG00000286159		novel transcript, antisense to PREX1	15.82	DEC1(h),DEC2(h),HELT(h),HEY-1(h),HEY-2(h), SREBP-1(h),SREBP-2(h), p300(h), ER-alpha(h), GR(h), LXRalpha(h), Runx2(h)...

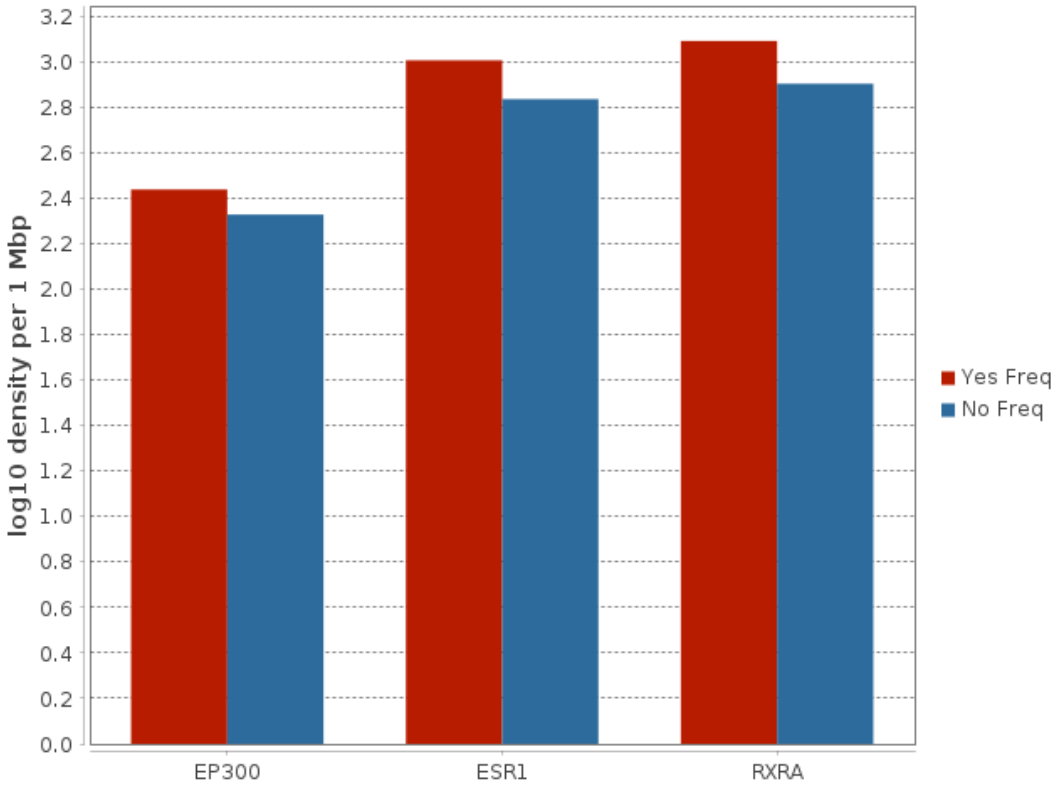
On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 16 transcription factors controlling expression of the genes associated with genomic variations (see Table 6).

Table 6. Transcription factors of the predicted enhancer model potentially regulating the genes carrying sequence variations (the most frequently mutated genes in Experiment: short-term survival). ***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
MO000056654	EP300	E1A binding protein p300	7.7	1.29
MO000019431	ESR1	estrogen receptor 1	7.61	1.48
MO000019619	RXRA	retinoid X receptor alpha	7.5	1.54
MO000026285	RUNX2	RUNX family transcription factor 2	6.73	1.62
MO000025765	SREBF2	sterol regulatory element binding transcription factor 2	6.34	1.28
MO000000904	FOXO4	forkhead box O4	6.31	1.75
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	6.25	1.87
MO000033904	RARA	retinoic acid receptor alpha	6.16	1.67
MO000028384	NEUROD1	neuronal differentiation 1	5.92	1.56
MO000056029	SREBF1	sterol regulatory element binding transcription factor 1	5.4	1.5

The following diagram represents the key transcription factors, which were predicted to be potentially regulating genes carrying sequence variations in the analyzed pathology: EP300, ESR1 and RXRA.



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 172 signaling proteins whose structure and function is highly damaged by the mutations (see Table 7).

Table 7. Signaling proteins whose structure and function are damaged by the mutations in the most frequently mutated genes

See full table →

ID	Title	Mutation count	Consequence	Codons
MO000019673	p85alpha(h)	22	stop_gained	Cga/Tga
MO000138949	Drp1(h)	22	NMD_transcript_variant,stop_gained	Gaa/Taa
MO000206935	C11orf74(h)	12	stop_gained	Gaa/Taa
MO000167580	Sur-8(h)	11	NMD_transcript_variant,stop_gained	Gaa/Taa
MO000168719	GIPN(h)	10	NMD_transcript_variant,frameshift_variant	aat/aaAt
MO000211774	DPAGT1(h)	10	NMD_transcript_variant,frameshift_variant	ttc/ttTc
MO000190658	GPSM2(h)	9	stop_gained	Gaa/Taa
MO000093071	chd8(h)	8	stop_gained	taC/taA
MO000113258	MYPT1(h)	8	NMD_transcript_variant,frameshift_variant	aga/aAga
MO000127741	SMC4L1(h)	8	stop_gained	Cga/Tga

Top 100 mutated proteins for the most frequently mutated genes 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 Table 8.

Table 8. Master regulators that may govern the regulation of the most frequently mutated genes in Experiment: short-term survival. ***Total rank*** *is the sum of the ranks of the master molecules sorted by keynode score, CMA score, genomics data.*
See full table →

ID	Master molecule name	Gene symbol	Gene description	Total rank	Weighted score
MO000019174	Eck(h)	EPHA2	EPH receptor A2	387	146.98
MO000045014	PML(h)	PML	PML nuclear body scaffold	442	164.01
MO000018003	PP2A(h)	PPP2CA, PPP2R3A, PPP2R3B, PPP2R5A, PPP2R5B, PPP2R5C, PPP2R5D	protein phosphatase 2 catalytic subunit alpha, protein phosphatase 2 regulatory subunit B"alpha, pr...	526	142.73
MO001087323	PP2A-B56-beta,gamma	PPP2CA, PPP2CB, PPP2R1A, PPP2R1B, PPP2R5B, PPP2R5C	protein phosphatase 2 catalytic subunit alpha, protein phosphatase 2 catalytic subunit beta, protein...	563	142.73
MO000334531	Eck-isoform2(h)	EPHA2	EPH receptor A2	592	146.98
MO000137320	Eck-isoform1(h)	EPHA2	EPH receptor A2	593	146.98
MO001079984	Ubc9(h){sumoC93}:sumo1(h) {clCG97,93}	SUMO1, UBE2I	small ubiquitin like modifier 1, ubiquitin conjugating enzyme E2 I	721	214.35
MO001080002	Ubc9(h) {sumo2C93}:SUMO2(h) {clCG93,93}	SUMO2, UBE2I	small ubiquitin like modifier 2, ubiquitin conjugating enzyme E2 I	723	214.35
MO000019975	Ubc9(h)	UBE2I	ubiquitin conjugating enzyme E2 I	724	214.35
MO001080010	Ubc9(h){sumo3C93}:sumo3(h) {clCG92,93}	SUMO3, UBE2I	small ubiquitin like modifier 3, ubiquitin conjugating enzyme E2 I	725	214.35

The intracellular regulatory pathways controlled by the above-mentioned master regulators are depicted in Figure 5. This diagram displays the connections between identified transcription factors, which play important roles in the regulation of genes carrying sequence variations, and selected master regulators, which are responsible for the regulation of these TFs.

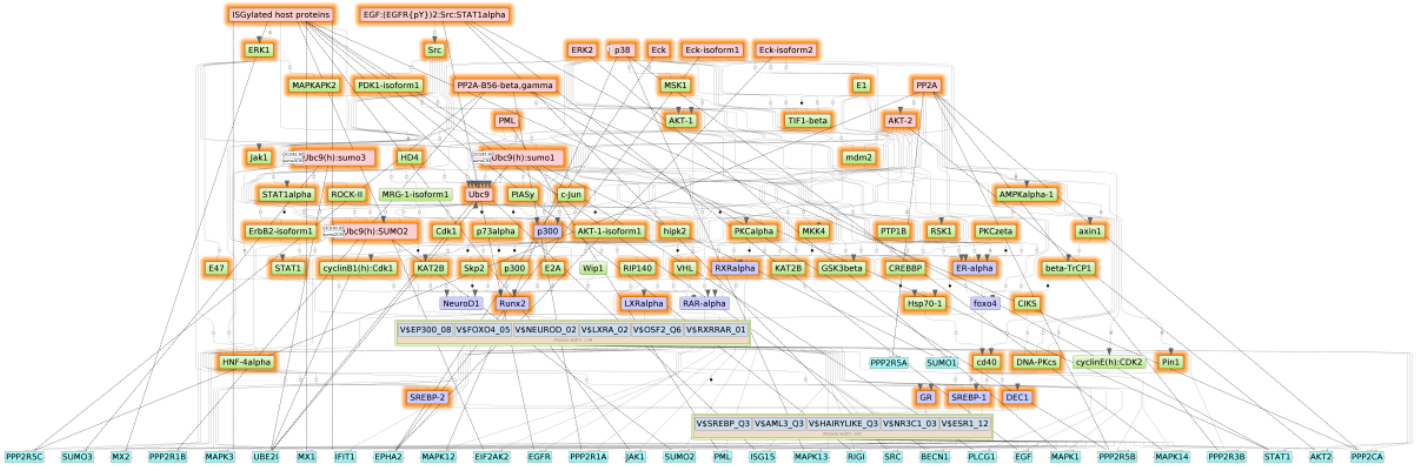


Figure 5. Diagram of intracellular regulatory signal transduction pathways of the most frequently mutated genes in Experiment: short-term survival. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network during the search for master regulators from selected TFs. Orange frames highlight molecules presented in original mapping.
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 9. 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	Total rank	Weighted score
EPHA2	EPH receptor A2	33	387	146.98
PML	PML nuclear body scaffold	1	442	164.01
PPP2CA	protein phosphatase 2 catalytic subunit alpha	3	526	142.73
CREBBP	CREB binding protein	1	968	117.67
EP300	E1A binding protein p300	3	968	117.67
KAT2A	lysine acetyltransferase 2A	1	968	117.67

*Table 10. 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.*

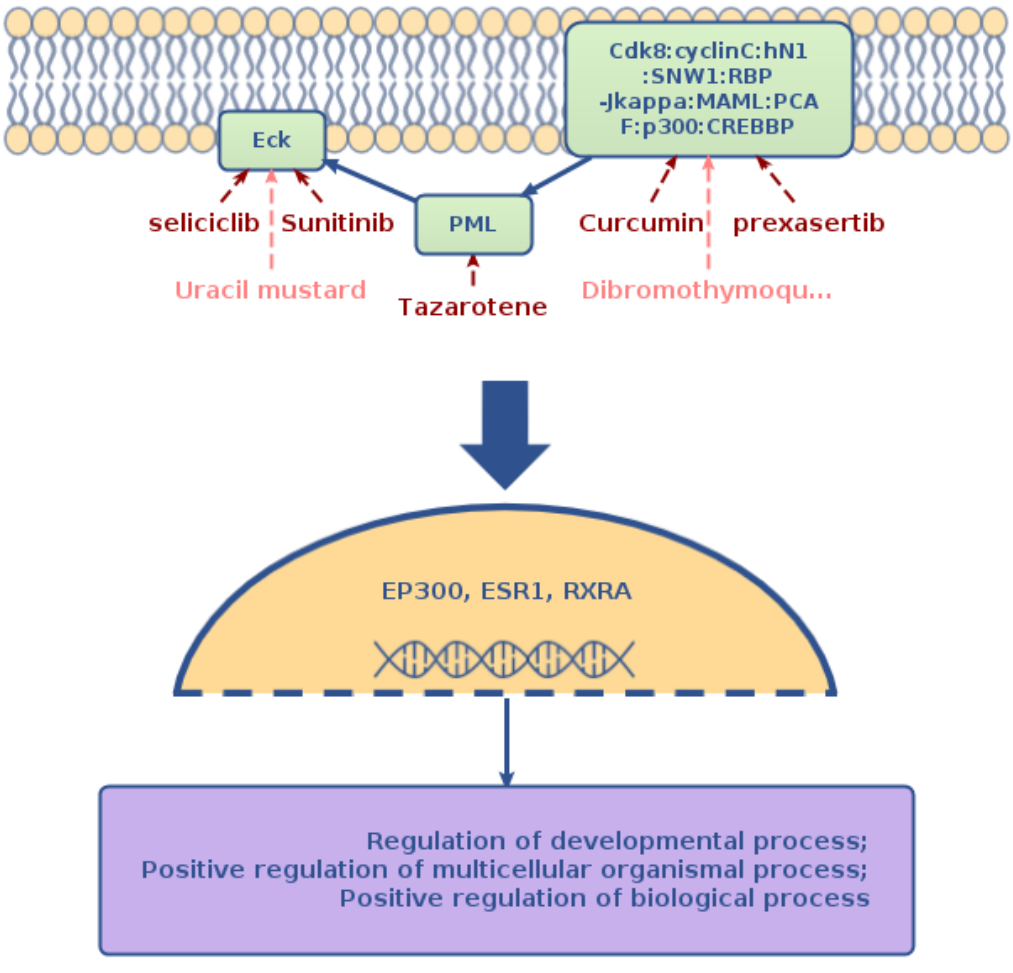
See full table →

Gene symbol	Gene Description	Druggability score	Total rank	Weighted score
EPHA2	EPH receptor A2	4.97	387	146.98
CREBBP	CREB binding protein	4.72	968	117.67
EP300	E1A binding protein p300	3.49	968	117.67
MAPK12	mitogen-activated protein kinase 12	36.68	970	132.22
MAP2K3	mitogen-activated protein kinase kinase 3	5.86	1027	683.5
DYRK1B	dual specificity tyrosine phosphorylation regulated kinase 1B	16.35	1027	683.5

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:

- Eck
- PML
- Cdk8:cyclinC:hN1:SNW1:RBP-Jkappa:MAML:PCAF:p300:CREBBP

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: Sunitinib, Tazarotene, Uracil mustard, prexasertib, seliciclib, Dibromothymoquinone and Curcumin, 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 12 and 13), 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.

If sufficient information regarding the known associations between predicted drugs and variants identified in the studied pathology was found, this will be reflected in the ***Somatic variants*** column of the FDA approved and repurposed drugs used in clinical trials tables. Details on these variant-drug associations can be found in the Molecular Tumor Board (MTB) report generated for the studied pathology.

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

Drugs approved in clinical trials for Oncology

Table 11. 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
Regorafenib	KIT, MAPK1, ABL1, KDR, FGFR2, PDGFRB, FGFR1, BRAF, PDGFRA, RAF1, FLT4, FRS2, MAPK11, EPHA2, FLT1, AKT1, DDR2, TEK, MAPK3, RET, AKT2	96	3	small molecule,approved	Colorectal Neoplasms (FDA)
encorafenib	MAPK10, MAPK1, MAPK8, MAPK9, MAPK7, MAPK6, LIMK1, BRAF, MAPK3, MAP2K4, RAF1	90	1	small molecule, approved	Colorectal Neoplasms (ClinicalTrials, ClinicalTrials, FDA)
Fluorouracil	IL11, IL6R, ERCC1, CASP8, CASP3, BAX, ITGAL, FAS, BIRC5, CDKN1A	86	6	small molecule,approved	Colorectal Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials)
Oxaliplatin	CTNNB1, IGF2, MMP9, CDH1, AKT1, AKT2	82	6	small molecule,approved,investigational	Colorectal Neoplasms (FDA, PUBMED)
fruquintinib	KDR, FLT1, FLT4	81	1	small molecule, approved	Colorectal Neoplasms (FDA)
Aflibercept	VEGFA, NOS3, PGF, LGALS1	78	3	biotech,approved	Colorectal Neoplasms (FDA)
Irinotecan	MAPK10, BIRC5, TOP1, RUNX3, CDKN1A	71	5	small molecule,approved,investigational	Colorectal Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, FDA, FDA)
Bevacizumab	FCGR2A, VEGFA, FCGR2B	63	5	biotech,approved,investigational	Colorectal Neoplasms (FDA, FDA)
Panitumumab	EGFR	60	2	biotech,approved,investigational	Colorectal Neoplasms (FDA)
Cetuximab	FCGR2A, EGFR, FCGR2B	60	5	biotech,approved	Colorectal Neoplasms (FDA, FDA)
Capecitabine	CDKN2A	36	5	small molecule,approved,investigational	Colorectal Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, FDA, FDA)

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 12. 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
Sunitinib	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, SIRT3, PRKCA, CSNK1E, SIRT2, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, CAMKK1, EPHB1, SIRT1, TGFBR2, FYN, PRKD2	98	2	small molecule,approved,investigational
Sirolimus	ROCK2, MARK3, NEK7, PRKACA, GSK3B, HIPK3, NEK2, NFE2L2, SGK1, CAMK1, NEK6, BAX, RPS6KA5, CSK, IL15, MAPKAPK5, PRKD1, MAPK3, MELK, MAPK8, MAPKAPK3, CAMKK2, DYRK1A, MAP2K1, FGF2, EIF4E, PRKCZ, MAPK12, MAPK11, MAPKAPK2, CHEK2, AKT2, STK3, PDPK1, IKBKB, MAPK10, RPS6KA3, MAPK15, PAK6, ITGAL, PRKCA, IL10, PRKAA1, AURKB, RPS6KA1, FKBP4, HSP90AA1, MAPK14, CSNK1D, PLK1, PGR, FKBP1A, EEF2K, AKT1, CHEK1, RBL2, TP53, RPS6KB1, MAPK1, IL7, HIPK2, PIM2, MAPK9, MTOR, PAK4, PIM3, MAPK13, CAMKK1, CTLA4, DCN	97	2	small molecule,approved,investigational
Sorafenib	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, HIPK3, MAP3K11, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, DYRK1A, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RIPK3, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, EEF2K, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	97	1	small molecule,approved,investigational
Erlotinib	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BIRC5, BMX, ERBB3, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5,	97	1	small molecule,approved,investigational

	PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, MAP3K1, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, ILK, CDC42BPA, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, ERBB4, PAK3, BRAF, FER, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2, NR1I2			
Vandetanib	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, VEGFA, TGFBR1, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, MST1R, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, CDK1, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	97	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 13. Repurposed drugs used in clinical trials for other pathologies (prospective drugs against the identified drug targets on the basis of literature curation in HumanPSD™ database)

See full table →

Name	Target names	Drug score	Maximum trial phase
seliciclib	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, HIPK3, MAP3K11, CDK4, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, DYRK1A, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, CDK1, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, EEF2K, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	93	PHASE2: Cystic Fibrosis, Cysts, Fibrosis
1-(5-Tert-Butyl-2-P-Tolyl-2h-Pyrazol-3-Yl)-3-[4-(2-Morpholin-4-Yl-Ethoxy)-Naphthalen-1-Yl]-Urea	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, HIPK3, MAP3K11, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, DYRK1A, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, EEF2K, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	93	PHASE2: Arthritis, Arthritis, Rheumatoid, Psoriasis
pi-103	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, HIPK3, MAP3K11, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, DYRK1A, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, EEF2K, PRKCH, ACVR2A, PTK2B, AKT1,	93	N/A

	AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2		
ruboxistaurin	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, HIPK3, MAP3K11, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, DYRK1A, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, EEF2K, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	93	PHASE1: Diabetes Mellitus, Diabetes Mellitus, Type 2, Heart Failure
Flavopiridol	BMPR1A, IKBKE, ABL1, FGFR2, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, CDK4, SYK, EPHA1, CAMK1, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, AXL, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, PIK3CB, WEE1, MAPK7, MAPK11, EPHA5, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, CDK6, KDR, MAPK6, MAPK15, PAK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, EPHB3, ACVR1, DDR2, CHEK1, PTK6, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, XIAP, MAP3K4, PAK1, CLK1, CDK5, NLK, PRKACB, RET, ABL2, CSF1R, STK10, TXK, MAP4K4, MARK3, IRAK3, MAP3K10, SLK, ACVRL1, INSR, NEK2, MAP2K2, MAP2K3, EPHA2, CAMK2B, RPS6KA5, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, TNIK, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, DDR1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, FGFR4, ACVR1B, CAMK2G, MET, PRKCQ, EPHA4, CDK1, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, CDC42BPA, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, BRAF, FER, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	93	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

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

Table 14. Prospective drugs, predicted by PASS software to be active against the identified drug targets with 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
LE-SN38	HIF1A, TOP2A, CASP3, NFKB1, TOP1, TOP2B, LGALS1	93	4.69
Camptothecin	HIF1A, TOP2A, CASP3, NFKB1, TOP1, TOP2B, LGALS1	93	4.67
Topotecan	HIF1A, TOP2A, CASP3, NFKB1, TOP1, TOP2B, LGALS1	92	4.41
BNP 1350	TOP2A, NFKB1, TOP1, TOP2B, LGALS1	89	3.49
Irinotecan	HIF1A, TOP2A, TOP1, TOP2B, LGALS1	88	3.17

Table 15. Prospective drugs, predicted by PASS software to be active against the identified drug targets, though without cheminformatically predicted activity against the studied disease(s) (drug candidates predicted with the cheminformatics tool PASS)
See full table →

Name	Target names	Drug score	Target activity score
2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One	IRAK3, PAK2, SENP8, CDK4, SGK1, EPHA1, MAP2K3, CAMK1, CAMK2B, LIMK1, PKN1, MAP2K6, MAPK8, GRK6, MAPKAPK3, PRKD3, CSNK1G2, DYRK1A, HSP90AB1, CLK4, CSNK1G1, CAMK2A, EPHA5, IRAK1, CHEK2, FES, AKT2, DAPK3, MAPK10, RPS6KA3, IRAK4, CDK6, CAMK2G, CSNK1A1, STAT1, PAK6, EPHA4, CSNK1E, CDK1, MAP2K4, AURKB, CDK7, PRKAA1, HSP90AA1, CSNK1D, CDC42BPA, EEF2K, PTK2B, AKT1, AURKA, CHEK1, MAPK9, EPHB2, DYRK1B, GSK3A, PAK4, EPHA3, PAK3, CAMK2D, EPHB1, PAK1, CLK1, PKN3, CDK5	96	22.53
Iodophenyl	EIF2AK3, NEK7, PAK2, GSK3B, HIPK3, VRK1, NEK3, NEK6, EIF2AK2, LIMK1, MELK, UHMK1, MAPKAPK3, RPS6KB2, CSNK1G2, MAP4K1, VRK2, TRIO, STK11, OBSCN, CSNK1G1, NEK10, KALRN, IRAK1, CSNK2A1, CHEK2, AKT2, PDPK1, STK3, IRAK4, PINK1, CSNK1A1, PAK6, CSNK1E, RAF1, AURKB, CSNK1D, CHEK1, TAF1, TAOK1, ATM, RIPK2, PIM2, HIPK2, PRPF4B, RPS6KA2, PAK4, PIM3, PRKDC, NUAK1, PASK, PAK1, CDK5, TRPM7, STK10, ROCK2, ULK1, BCR, MAK, MAP4K4, CIITA, MARK3, IRAK3, SLK, NEK2, SGK1, TAOK2, RPS6KA5, MAPKAPK5, TNIK, BUB1B, CSNK2A2, DCAF1, TNK2, MAP4K3, RIPK1, STK4, MAP4K2, MAPKAPK2, DAPK3, RPS6KA3, RIPK3, PIK3CG, LATS1, PRKAA1, RPS6KA1, ILK, CDC42BPA, MASTL, HIPK1, AKT1, AURKA, LMTK2, PKMYT1, RPS6KB1, TBK1, ATR, MTOR, GSK3A, LATS2, ERN1, PAK3, BRAF, MINK1	96	41.32
Rbt205 Inhibitor	RPS6KA3, CDK6, CAMK2G, GRK2, MAP3K10, PRKCQ, PRKACA, GSK3B, PRKCA, CDK1, MAP2K4, CDK4, CDK7, RPS6KA1, PRKAA1, SGK1, PRKCH, EEF2K, CAMK1, CAMK2B, MAPKAPK5, PDP1, PRKD1, PKN1, MAP2K6, GRK6, RPS6KA2, MAPKAPK3, GSK3A, GRK5, PRKD3, PRKAR1B, PDP2, PRKCE, PRKCZ, CAMK2D, PTK2, CAMK2A, SIRT1, PRKCD, PRKCI, CDK5, GRK3, PKN3, PRKACB, DAPK3	95	31.32
7-[4-(Dimethylamino)Phenyl]-N-Hydroxy-4,6-Dimethyl-7-Oxo-2,4-Heptadienamide	HDAC4, HDAC2, HDAC6, HDAC7, HDAC9, HDAC3, HDAC5, HDAC1, LGALS1	94	11.63
Uracil mustard	TXK, ABL1, FGFR2, JAK3, INSR, PEAK1, SYK, EPHA1, EIF2AK2, EPHA2, ERBB2, TEK, CSK, AXL, TYK2, MELK, JAK1, IGF1R, SRC, PDGFRB, TNK2, DDR1, WEE1, PTK2, EPHA5, FLT1, PRKCD, LYN, B4GALT1, FES, FGFR4, MET, KDR, EPHA4, PDGFRA, FLT4, MST1R, EGFR, EPHB3, PTK2B, DDR2, PTK6, JAK2, KIT, BMX, RIPK2, ERBB3, EPHB2, FGFR3, MTOR, NTRK2, TYRO3, MERTK, EPHB4, EPHA3, YES1, FGFR1, ERBB4, FER, ITK, EPHB1, FYN, RET, CSF1R, ABL2	93	12.74

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Sunitinib, seliciclib, LE-SN38 and 2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One. These drugs were selected for acting on the following targets: EPHA2, HIF1A and MAP2K3, which were predicted to be active in the molecular mechanism of the studied pathology.

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

Supplementary drug info

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

Table 16. Supplementary drug info: extended list of drugs used for treatment of neoplasms with respective drug scores predicted for the studied pathology.

Drug	Disease	Predicted Drug Score	Somatic variants
Abarelix	Prostatic Neoplasms	7	
Abemaciclib	Breast Neoplasms	73	
Abiraterone	Prostatic Neoplasms, Castration-Resistant	-	
Abiraterone acetate	Prostatic Neoplasms, Castration-Resistant	-	
Acalabrutinib	Lymphoma, Mantle-Cell	40	
Acitretin	Psoriasis	73	
Ado-trastuzumab emtansine	Breast Neoplasms|Neoplasms	86	
Afatinib	Carcinoma, Non-Small-Cell Lung	85	

Aflibercept	Colorectal Neoplasms|Diabetic Retinopathy|Edema|Vascular Diseases|Wet Macular Degeneration	78	
Alectinib	Carcinoma, Non-Small-Cell Lung	77	
Alemtuzumab	Brain Abscess|Leukemia, Lymphocytic, Chronic, B-Cell|Multiple Sclerosis|Multiple Sclerosis, Relapsing-Remitting|Sclerosis	-	
Alitretinoin	Sarcoma, Kaposi	64	
Alpelisib	Breast Neoplasms	-	
Altretamine	Ovarian Neoplasms	-	
Aminolevulinic acid	Keratosis|Keratosis, Actinic	24	
Anagrelide	Thrombocythemia, Essential|Thrombocytosis	-	
Anastrozole	Breast Neoplasms|Hypersensitivity|Obesity|Obesity, Morbid|Recurrence|Weight Loss	64	
Apalutamide	Prostatic Neoplasms, Castration-Resistant	12	
Aprepitant	Nausea|Neoplasms|Postoperative Nausea and Vomiting	-	
Arsenic trioxide	Leukemia, Promyelocytic, Acute	84	
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	94	
Azacitidine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes|Preleukemia|Syndrome	74	
Belinostat	Lymphoma, T-Cell, Peripheral	84	
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	63	
Bexarotene	Lymphoma, T-Cell|Lymphoma, T-Cell, Cutaneous	71	
Bicalutamide	Prostatic Neoplasms	67	
Binimetinib	Melanoma	75	
Blinatumomab	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-	
Bortezomib	Brain Abscess|Glomerulonephritis|Glomerulonephritis, IGA|Kidney Diseases|Multiple Myeloma|Neoplasms, Plasma Cell|Nephritis|Renal Insufficiency	74	
Bosutinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	86	
Brentuximab vedotin	Hodgkin Disease|Lymphoma|Lymphoma, Large-Cell, Anaplastic|Lymphoma, T-Cell, Peripheral	-	
Brigatinib	Carcinoma, Non-Small-Cell Lung	79	
Buserelin	Prostatic Neoplasms	-	
Cabazitaxel	Prostatic Neoplasms, Castration-Resistant	81	
Cabergoline	Drug-Related Side Effects and Adverse Reactions|Pituitary Neoplasms	18	
Cabozantinib	Thyroid Neoplasms	85	
Capecitabine	Breast Neoplasms|Colonic Neoplasms|Colorectal Neoplasms	36	
Carboplatin	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms|Neuroendocrine Tumors|Ovarian Neoplasms|Retinoblastoma	-	XRCC1:Q399R:sensitivity/response:B2
Carfilzomib	Multiple Myeloma	82	
Carmustine	Astrocytoma|Glioblastoma|Hodgkin Disease|Medulloblastoma|Multiple Myeloma|Neoplasms	17	
Ceritinib	Carcinoma, Non-Small-Cell Lung	89	
Cetuximab	Colorectal Neoplasms	60	
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	80	XRCC1:Q399R:sensitivity/response:B2

Cladribine	Leukemia, Hairy Cell	76
Clofarabine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	75
Cobimetinib	Melanoma	84
Copanlisib	Lymphoma, Follicular	87
Crizotinib	Carcinoma, Non-Small-Cell Lung	93
Cyproterone acetate	Prostatic Neoplasms	12
Dabrafenib	Melanoma	77
Dacomitinib	Carcinoma, Non-Small-Cell Lung	86
Daratumumab	Multiple Myeloma	22
Dasatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase|Precursor Cell Lymphoblastic Leukemia-Lymphoma	96
Decitabine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes	70
Degarelix	Cardiovascular Diseases|Prostatic Neoplasms|Vascular Diseases	74
Denosumab	Arthritis, Rheumatoid|Bone Diseases|Bone Diseases, Metabolic|Breast Neoplasms|Hyperparathyroidism|Hyperparathyroidism, Primary|Metabolic Diseases|Neoplasm Metastasis|Neoplasms|Osteoporosis|Osteoporosis, Postmenopausal|Prostatic Neoplasms	50
Dexrazoxane	Breast Neoplasms|Cardiomyopathies	59
Dienogest	Menorrhagia	59
Dinutuximab	Neuroblastoma	-
Docetaxel	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Prostatic Neoplasms|Squamous Cell Carcinoma of Head and Neck|Stomach Neoplasms	77
Doxorubicin	Neoplasms|Multiple Myeloma|Carcinoma, Ovarian Epithelial|Ovarian Neoplasms|Leukemia, Lymphoid|Breast Neoplasms|Lymphoma, Follicular|Thyroid Neoplasms|Triple Negative Breast Neoplasms|Glioma	93
Durvalumab	Carcinoma, Non-Small-Cell Lung|Carcinoma, Transitional Cell	-
Dutasteride	Alcoholism|Hyperplasia|Hypertrophy|Neoplasms|Prostatic Hyperplasia	-
Duvelisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	79
Elotuzumab	Multiple Myeloma	80
Enasidenib	Leukemia, Myeloid, Acute	-
Encorafenib	Colorectal Neoplasms|Melanoma	90
Enfortumab vedotin	Carcinoma, Transitional Cell|Neoplasms	-
Entrectinib	Carcinoma, Non-Small-Cell Lung	61
Enzalutamide	Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	16
Epirubicin	Breast Neoplasms	72
Erdafitinib	Urinary Bladder Neoplasms	93
Eribulin	Breast Neoplasms|Drug-Related Side Effects and Adverse Reactions|Neoplasms	-
Erlotinib	Carcinoma, Non-Small-Cell Lung|Neoplasms|Pancreatic Neoplasms	97
Erlotinib hydrochloride	Carcinoma, Non-Small-Cell Lung|Gastrointestinal Stromal Tumors	-
Estramustine	Prostatic Neoplasms	68
Ethinyl Estradiol	Acne Vulgaris|Neoplasms	65
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	86
Exemestane	Breast Neoplasms	24
Fedratinib	Primary Myelofibrosis	66
Finasteride	Hyperplasia|Neoplasms|Prostatic Hyperplasia	40
Flavopiridol	Leukemia, Lymphocytic, Chronic, B-Cell	93
Fluorouracil	Skin Neoplasms|Neoplasms, Basal Cell|Neoplasms, Second Primary|Neoplasms, Squamous Cell|Neoplasms|Colorectal Neoplasms|Pancreatic Neoplasms	86
Fluoxymesterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	57
Flutamide	Premenstrual Dysphoric Disorder|Premenstrual Syndrome|Prostatic Neoplasms	66
Fulvestrant	Breast Neoplasms	73

Gefitinib	Carcinoma, Non-Small-Cell Lung	92
Gemcitabine	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Ovarian Neoplasms|Pancreatic Neoplasms	83
Gemtuzumab ozogamicin	Leukemia, Myeloid, Acute	31
Gilteritinib	Leukemia, Myeloid, Acute	81
Glasdegib	Leukemia, Myeloid, Acute	36
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	84
Ibritumomab	Lymphoma, B-Cell|Lymphoma, Follicular	31
Ibrutinib	Graft vs Host Disease|Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, B-Cell, Marginal Zone|Lymphoma, Mantle-Cell|Waldenstrom Macroglobulinemia	89
Idarubicin	Leukemia, Myeloid, Acute	61
Idelalisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	76
Ifosfamide	Neoplasms	45
Imatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Mastocytosis, Systemic|Neoplasms	92
Inotuzumab ozogamicin	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Ipilimumab	Carcinoma, Renal Cell|Melanoma	53
Irinotecan	Colorectal Neoplasms	71
Ivosidenib	Leukemia, Myeloid, Acute	-
Ixabepilone	Breast Neoplasms	-
Ixazomib	Multiple Myeloma	-
Lapatinib	Breast Neoplasms	93
Larotrectinib	Neoplasm Metastasis	59
Lenalidomide	Brain Abscess|Lupus Erythematosus, Cutaneous|Myelodysplastic Syndromes|Neoplasms, Plasma Cell	65
Lenvatinib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	93
Letrozole	Breast Neoplasms|Cysts|Fibroma|Myofibroma|Myoma|Ovarian Cysts|Syndrome	45
Leuprolide	Hot Flashes|Ovarian Hyperstimulation Syndrome|Prostatic Neoplasms|Puberty, Precocious	-
Levamisole	Ascariasis|Colonic Neoplasms|Helminthiasis	3
Levonorgestrel	Epilepsy|Hyperplasia|Menorrhagia	50
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	62
Lorlatinib	Carcinoma, Non-Small-Cell Lung	77
Masoprocol	Keratosis, Actinic	64
Medroxyprogesterone Acetate	Depression|Depression, Postpartum|Depressive Disorder|Metrorrhagia|Neoplasms|Uterine Hemorrhage	76
Megestrol acetate	Acquired Immunodeficiency Syndrome|Bites and Stings|Breast Neoplasms|Pain|Wasting Syndrome	65
Methotrexate	Neoplasms|Breast Neoplasms|Head and Neck Neoplasms|Ovarian Neoplasms|Lymphoma, T-Cell, Peripheral|Brain Neoplasms|Colorectal Neoplasms|Neuroblastoma|Carcinoma, Squamous Cell	74
Methyltestosterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	-
Midostaurin	Leukemia, Mast-Cell|Leukemia, Myeloid, Acute|Mastocytosis, Systemic	92
Mitotane	Adrenocortical Carcinoma	47
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	74
Mogamulizumab	Mycosis Fungoides|Neoplasms|Sezary Syndrome	-
Moxetumomab pasudotox	Leukemia, Hairy Cell|Neoplasms	13
Necitumumab	Carcinoma, Non-Small-Cell Lung|Neoplasms	-

Nelarabine	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma	-
Neratinib	Breast Neoplasms	87
Nilotinib	Blast Crisis|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase	87
Nilutamide	Prostatic Neoplasms	12
Nintedanib	Fibrosis|Idiopathic Pulmonary Fibrosis	94
Niraparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms	40
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	70
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	53
Olaratumab	Sarcoma	-
Osimertinib	Carcinoma, Non-Small-Cell Lung	92
Oxaliplatin	Colonic Neoplasms|Colorectal Neoplasms|Neoplasms|Rectal Neoplasms	82
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	91
Palbociclib	Breast Neoplasms	62
Panitumumab	Colorectal Neoplasms	60
Panobinostat	Multiple Myeloma	82
Pazopanib	Carcinoma|Carcinoma, Renal Cell|Sarcoma	92
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	23
Pertuzumab	Breast Neoplasms	90
Pomalidomide	Multiple Myeloma	25
Ponatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Precursor Cell Lymphoblastic Leukemia-Lymphoma	91
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	96
Relugolix	Prostatic Neoplasms	-
Ribociclib	Breast Neoplasms	72
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	31
Romidepsin	Lymphoma, T-Cell, Cutaneous	88
Rucaparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms|Prostatic Neoplasms, Castration-Resistant	78
Ruxolitinib	Graft vs Host Disease|Polycythemia|Polycythemia Vera|Primary Myelofibrosis|Thrombocytosis	79
Selinexor	Multiple Myeloma	70
Selumetinib	Neurofibromatosis 1	86
Siltuximab	Giant Lymph Node Hyperplasia	-
Sirolimus	Angiomyolipoma|Constriction, Pathologic|Coronary Restenosis|Eye Diseases|Immune System Diseases|Kidney Failure, Chronic|Lipoma|Tuberous Sclerosis	97
Sonidegib	Carcinoma, Basal Cell	46

Sorafenib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	97
Sunitinib	Adenoma|Carcinoma, Renal Cell|Digestive System Neoplasms|Gastrointestinal Neoplasms|Gastrointestinal Stromal Tumors|Intestinal Neoplasms	98
Talazoparib	Breast Neoplasms	56
Tamoxifen	Breast Diseases|Cystic Fibrosis|Cysts|Fibroadenoma|Fibrocystic Breast Disease|Hemorrhage|Menorrhagia|Menstruation Disturbances|Metrorrhagia|Neoplasms	85
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	46
Temsirolimus	Carcinoma, Renal Cell	88
Teniposide	Precursor Cell Lymphoblastic Leukemia-Lymphoma	76
Thalidomide	Brain Abscess|Immune System Diseases|Multiple Myeloma|Neoplasms, Plasma Cell	84
Tivozanib	Carcinoma, Renal Cell	88
Tocilizumab	Arthritis|Arthritis, Juvenile|Arthritis, Rheumatoid|Behavior|Cytokine Release Syndrome|Giant Cell Arteritis|Neurobehavioral Manifestations|Oral Manifestations|Psychotic Disorders|Schizophrenia|Tic Disorders	39
Topotecan	Small Cell Lung Carcinoma	53
Toremifene	Breast Neoplasms	80
Trabectedin	Leiomyosarcoma|Liposarcoma	-
Trametinib	Carcinoma, Non-Small-Cell Lung|Melanoma	93
Trastuzumab	Breast Neoplasms|Neoplasms	81
Tretinoin	Lentigo	89
Triptorelin	Fatty Liver|Hypogonadism|Infertility, Female|Prostatic Neoplasms	84
Tucatinib	Breast Neoplasms	91
Valrubicin	Urinary Bladder Neoplasms	50
Vandetanib	Thyroid Neoplasms	97
Vemurafenib	Melanoma	87
Venetoclax	Leukemia, Lymphocytic, Chronic, B-Cell|Leukemia, Myeloid, Acute	53
Vinblastine	Glioma	60
Vincristine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	53
Vinorelbine	Carcinoma, Non-Small-Cell Lung	54
Vismodegib	Carcinoma, Basal Cell	32
Vorinostat	Lymphoma, T-Cell, Cutaneous	86
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	8

6. Conclusion

We applied the software package "Genome Enhancer" to a data set that contains *genomics* data. The study is done in the context of *Colorectal Neoplasms*. The data were pre-processed, statistically analyzed and genes carrying sequence variations 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:

Sunitinib, seliciclib, LE-SN38 and 2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One

These drugs were selected for acting on the following targets: EPHA2, HIF1A and MAP2K3, 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:

Eck, PML and Cdk8:cyclinC:hN1:SNW1:RBP-Jkappa:MAML:PCAF:p300:CREBBP

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: Sunitinib, Tazarotene, Uracil mustard, prexasertib, seliciclib, Dibromothymoquinone and Curcumin. 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 genes carrying sequence variations 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.

- Eck
- PML
- Cdk8:cyclinC:hN1:SNW1:RBP-Jkappa:MAML:PCAF:p300:CREBBP

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 genes carrying sequence variations 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.

Genomic data processing

When analyzing a list of genomic variations (from input vcf file or computed by Genome Enhancer from SNP list or from fastq files), first of all, we compute a specific mutation weight (w_1) for each variation depending on it's location in gene body and gene flanking regions (-1000 upstream and +1000 downstream of the gene body).

w_1 = 0.7 for variations in exon area
w_1 = 1.3 for variations in promoter region (-1000bp upstream and 100bp downstream of TSS),
w_1 = 1.0 for variations in other locations.

Next, VCF track (Yes track), provided as input or created by Genome Enhancer from SNP list or fastq files, is compared to Random VCF track (No track) of 10000 random human variations. On both tracks we calculate the score delta values (differences between PWM score values of the TF sites with the reference or with the alternative allele of the considered variation). For each variation we find then the maximal score delta values at each PWM leading either to the gain or to the loss of TF site (with the alternative allele). For selecting the maximum score delta values we consider both directions of DNA strand. Next, by going through all variations we compute two p-values for each PWM – the p-value of site losses and p-value of site gains. The p-values are computed using cumulative Binomial distribution estimating the random chances to observe the found high number of lost or gained TF sites in Yes track in the comparison to the No track. The PWM cut-offs are optimized to obtain the most extreme p-values. We further take top 20 best matrices by p-value from each: gained and lost sites and calculate the mutation weights on the Yes track on the basis of the obtained 40 matrices. Each mutation is assigned with a respective matrix that got the maximum delta value either for the site gain or for the site loss (changed the binding affinity most significantly). This delta is then compared to other delta values that were computed for the respective matrix on the No track. The eventual weight that reflects the transcription factor binding affinity change caused by the mutation is calculated as follows:

w_2 = -log10(NoGr / NoAll), if NoGr > 0
w_2 = -log10(1.0 / (2.0 * NoAll), if NoGr = 0

where NoGr is the number of deltas from the No track that appeared to be greater than the inspected delta and NoAll is the total number of deltas in the No track. The resulting track is then constructed that contains all sites of the initial Yes track together with the additional weights reflecting the transcription factor binding affinity change caused by the mutation.
The list of 40 matrices most affected by variations will be further used in composite modules search described in the next section.

Total Gene mutation weight is the sum of the weights w_1 of all variations located inside the gene body and in the gene flanking regions summed up with the weight w_2 that reflects the transcription factor binding affinity change caused by the mutation. This weight is calculated by estimating the importance of a certain mutation in terms of gains or losses of binding sites caused by it.
Next, a weighted score is calculated for all genes with the following formula:
Weighted score = In_disease * In_transpath * Gene mutation weight, where

In_disease = 2.0 for genes assigned to selected diseases,
In_transpath = 1.5 for genes mapped to Transpath pathways,
and In_disease = In_transpath = 1.0 in all other cases.

At the next step, 300 genes with highest weighted score are selected for further CMA model search.
The mutation weights ($w = w_1+w_2$) are also used to find the regulatory regions of the genes most affected by the variations/SNP. A sliding window of 1100 bp is used to scan through the intronic, 5' and 3' regions of the genes and a region is selected with the highest sum of the mutation weights.

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 enhancers under study as compared to a background set of promoters of housekeeping genes. 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). Each composite module is forced to include at least one matrix that was identified as matrix causing the significant change in the transcription factor binding affinity as the result of the observed mutation.

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 - Detailed report. Composite modules and master regulators (the most frequently mutated genes in Experiment: short-term survival).
2. Supplementary table 2 - 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.