ITGAL and MAPK12 are promising druggable targets for treating Colorectal Neoplasms that control activity of TP53, AR and NR3C1 transcription factor on promoters of genes carrying sequence variations

Demo User
geneXplain GmbH
info@genexplain.com

Data received on 10/12/2021 ; Run on 05/07/2023 ; Report generated on 05/07/2023

Genome Enhancer release 3.2 (TRANSFAC®, TRANSPATH® and HumanPSD™ release 2023.1)

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: TP53, AR and NR3C1. The subsequent network analysis suggested

- p38gamma
- integrins
- PML

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, seliciclib, Camptothecin 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] 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 [11-13]. 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

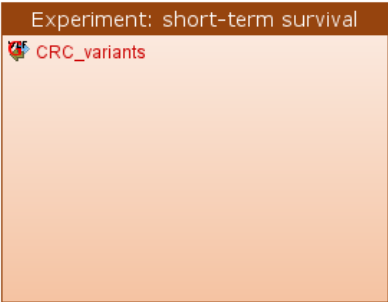| File name	Data type
CRC_variants	Genomics

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	419.42	629.12
ENSG00000228716	dihydrofolate reductase	DHFR		56	131.75	395.25
ENSG00000067057	phosphofructokinase, platelet	PFKP		92	223.43	335.15
ENSG00000003402	CASP8 and FADD like apoptosis regulator	CFLAR		43	111.41	334.23
ENSG00000176890	thymidylate synthetase	TYMS		44	110.62	331.86
ENSG00000169894	mucin 3A, cell surface associated	MUC3A		68	162.02	324.05
ENSG00000242086	MUC20 overlapping transcript	MUC20-OT1		147	323.51	323.51
ENSG00000204525	major histocompatibility complex, class I, C	HLA-C		71	154.6	309.21
ENSG00000248923	MT-ND5 pseudogene 11	MTND5P11		126	306.95	306.95
ENSG00000134086	von Hippel-Lindau tumor suppressor	VHL		43	98.24	294.71

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)

biological_process Gene Ontology treemap

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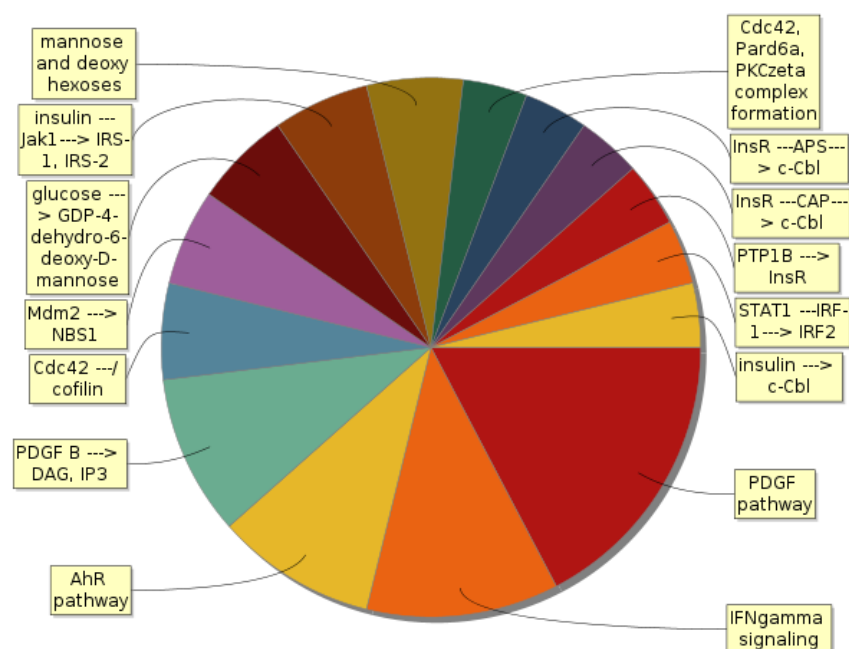**Full classification →**

TRANSPATH® Pathways (2023.1)

Figure 3. Enriched TRANSPATH® Pathways (2023.1) of the most frequently mutated genes in Experiment: short-term survival.
Full classification →

HumanPSD(TM) disease (2023.1)

- Digestive System Diseases
- Digestive System Neoplasms
- Gastrointestinal Diseases
- Gastrointestinal Neoplasms
- Colonic Diseases
- Intestinal Diseases
- Colorectal Neoplasms
- Intestinal Neoplasms
- Rectal Diseases
- Adenocarcinoma
- Neoplasms by Site
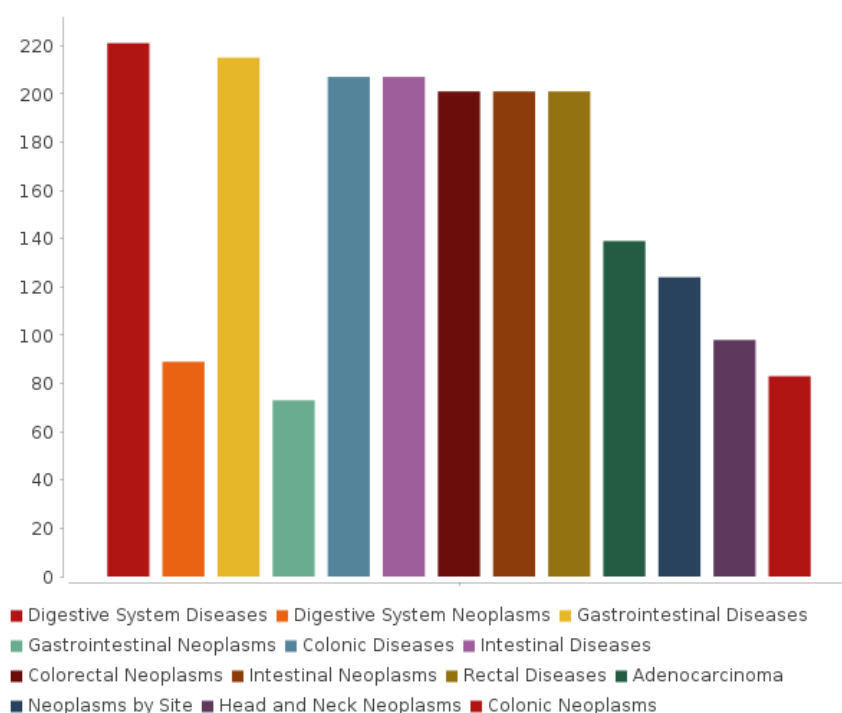- Head and Neck Neoplasms
- Colonic Neoplasms

Figure 4. Enriched HumanPSD(TM) disease (2023.1) 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):

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

The most frequently mutated genes in Experiment: short-term survival -log10(P-value)
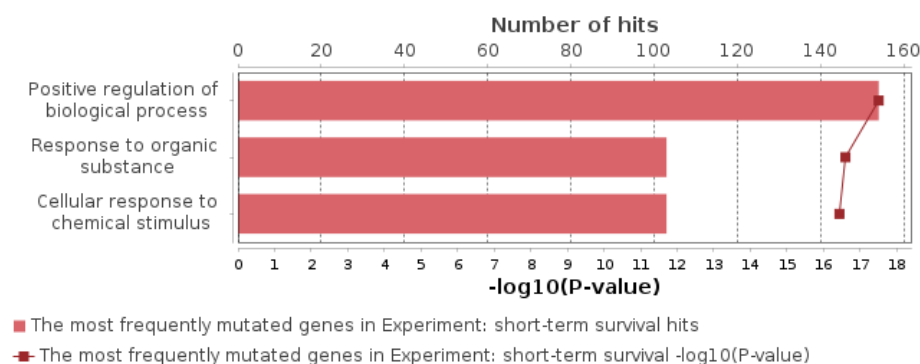
3.3. Analysis of enriched transcription factor binding sites and composite modules

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 23645 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
ENSG00000230021	ENSG00000230021		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 →

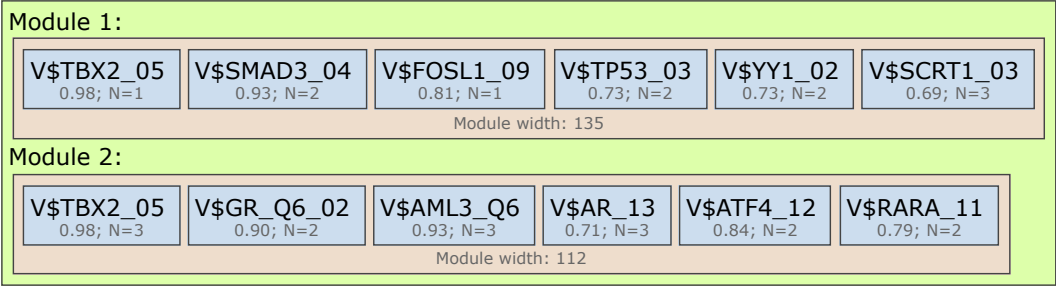| ID	P-value (gains)	P-value (losses)	yesCount (gains)	yesCount (losses)
V$OTX2_06	3.65E-2	8.72E-14	25	70
V$GSC_04	3.12E-2	3.07E-11	5	82
V$ZBTB33_07	1.06E-2	8.27E-12	13	611
V$ZBTB33_05	6.47E-3	6.92E-10	12	535
V$MATH1_Q2	1.99E-3	3.66E-12	204	389
V$E2A_03	1.6E-3	1.14E-12	231	387
V$MAFG_12	3.44E-5	3.47E-10	34	1295
V$E2F8_03	1.22E-6	1.32E-9	63	1618
V$MYBL2_04	1.81E-12	1.1E-10	2161	2211
V$E2F1DP1RB_01	1.88E-25	1.47E-9	3568	3831
V$E2F4DP1_01	1.88E-25	1.14E-10	3568	4126
V$SP1_02	1.18E-26	1.56E-2	250	6
V$KLF8_Q5	6.06E-27	6.79E-3	313	214
V$PAX5_Q6	4.25E-27	1.83E-3	1935	2567
V$E2F1_Q6	3.25E-28	3.18E-2	2073	14
V$E2F1DP2_01	2.9E-31	1.81E-11	3552	3615
V$E2F3_12	1.1E-31	3.12E-2	1493	5
V$E2F_Q4_02	1.46E-35	2.93E-2	2015	55
V$E2F1_04	2.31E-36		903	
V$E2F3_09	1.41E-41		3966	

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$TBX2_05	V$SMAD3_04	V$FOSL1_09	V$TP53_03	V$YY1_02	V$SCRT1_03
0.98; N=1	0.93; N=2	0.81; N=1	0.73; N=2	0.73; N=2	0.69; N=3

Module width: 135

Module 2:

V$TBX2_05	V$GR_Q6_02	V$AML3_Q6	V$AR_13	V$ATF4_12	V$RARA_11
0.98; N=3	0.90; N=2	0.93; N=3	0.71; N=3	0.84; N=2	0.79; N=2

Module width: 112

Model score (-p*log10(pval)): 23.26
Wilcoxon p-value (pval): 9.59e-50
Penalty (p): 0.475
Average yes-set score: 11.98
Average no-set score: 10.16
AUC: 0.77
Separation point: 10.98
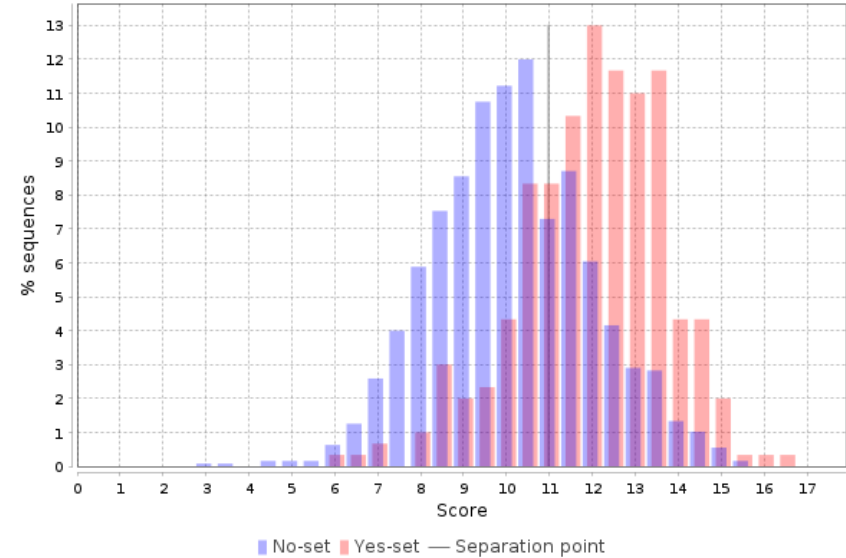**False-positive:** 31.22%
False-negative: 24.00%

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
ENSG00000197043	ANXA6	annexin A6	17.5	AR(h), SCRT1(h), Fra-1(h), YY1(h), Runx2(h), TBX2(h), p53(h)...
ENSG00000122133	PAEP	progestagen associated endometrial protein	17.31	Fra-1(h), SCRT1(h), TBX2(h), YY1(h), SMAD3(h), p53(h), GR(h)...
ENSG00000269807		novel transcript, antisense to MPND	17.31	AR(h), TBX2(h), SCRT1(h), GR(h), ATF-4(h), YY1(h), Fra-1(h)...
ENSG00000214176	PLEKHM1P1	pleckstrin homology and RUN domain containing M1 pseudogene 1	17.14	YY1(h), SMAD3(h), p53(h), SCRT1(h), TBX2(h), Fra-1(h), RAR-alpha(h)...
ENSG00000225671	FCF1P6	FCF1 pseudogene 6	17.02	GR(h), Runx2(h), RAR-alpha(h), TBX2(h), AR(h), ATF-4(h), SCRT1(h)...
ENSG00000107281	NPDC1	neural proliferation, differentiation and control 1	17.02	TBX2(h), GR(h), RAR-alpha(h), AR(h), Runx2(h), SMAD3(h), YY1(h)...
ENSG00000274928	KRT89P	keratin 89 pseudogene	17	ATF-4(h), p53(h), Runx2(h), SMAD3(h), YY1(h), AR(h), RAR-alpha(h)...
ENSG00000160179	ABCG1	ATP binding cassette subfamily G member 1	16.99	SMAD3(h), YY1(h), SCRT1(h), p53(h), TBX2(h), Fra-1(h), RAR-alpha(h)...
ENSG00000131944	FAAP24	FA core complex associated protein 24	16.95	GR(h), TBX2(h), RAR-alpha(h), AR(h), Runx2(h), ATF-4(h), SMAD3(h)...
ENSG00000119041	GTF3C3	general transcription factor IIIC subunit 3	16.85	RAR-alpha(h), GR(h), TBX2(h), Runx2(h), SCRT1(h), YY1(h), Fra-1(h)...

On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 11 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
MO000019548	TP53	tumor protein p53	8.49	1.48
MO000021454	AR	androgen receptor	6.91	1.48
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	6.85	1.41
MO000025684	FOSL1	FOS like 1, AP-1 transcription factor subunit	6.8	1.51
MO000033904	RARA	retinoic acid receptor alpha	6.77	1.62
MO000057832	SMAD3	SMAD family member 3	6.6	1.78
MO000019140	ATF4	activating transcription factor 4	5.58	2.15
MO000078913	YY1	YY1 transcription factor	5.54	1.24
MO000026285	RUNX2	RUNX family transcription factor 2	4.51	2.26
MO000028209	TBX2	T-box transcription factor 2	3.86	1.93

The following diagram represents the key transcription factors, which were predicted to be potentially regulating genes carrying sequence variations in the analyzed pathology: TP53, AR and NR3C1.
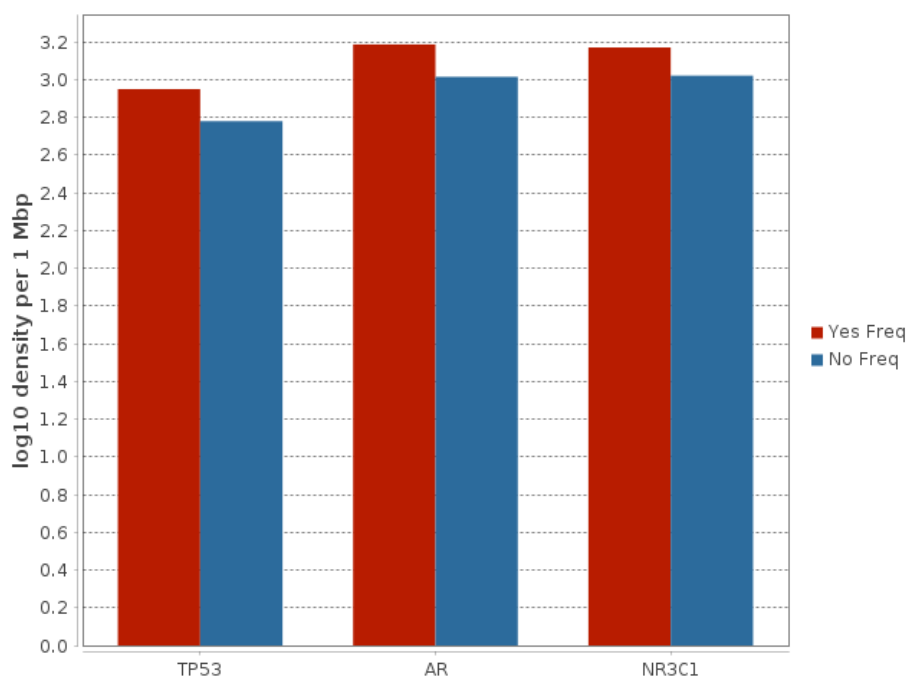
3.4. Finding master regulators in networks

In the second step of the upstream analysis common regulators of the revealed TFs were identified. We identified 174 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
MO000138949	Drp1(h)	13	NMD_transcript_variant,stop_gained	Gaa/Taa
MO000206935	C11orf74(h)	12	stop_gained	Gaa/Taa
MO000211774	DPAGT1(h)	10	NMD_transcript_variant,frameshift_variant	ttc/ttTc
MO000019673	p85alpha(h)	9	stop_gained	Cga/Tga
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
MO000214698	MS4A6A(h)	8	NMD_transcript_variant,frameshift_variant	-/T,tta/ttTa
MO000035319	kinectin(h)	7	NMD_transcript_variant,frameshift_variant	-/A

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
MO000079043	PML-4(h)	PML	PML nuclear body scaffold	163	136.46
MO000045014	PML(h)	PML	PML nuclear body scaffold	249	136.46
MO000004689	p38gamma(h)	MAPK12	mitogen-activated protein kinase 12	259	114.67
MO000021383	PTP1B(h)	PTPN1	protein tyrosine phosphatase non-receptor type 1	389	112.08
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...	422	191.78
MO000021981	BRCA1(h)	BRCA1	BRCA1 DNA repair associated	455	120.89
MO000078913	YY1(h)	YY1	YY1 transcription factor	466	66.77
MO000033757	p38gamma(h){p}	MAPK12	mitogen-activated protein kinase 12	479	114.67
MO000031189	PKCdelta(h)	PRKCD	protein kinase C delta	489	67.22
MO000042124	MKK3:Dyrk1B:PHS2:HNF-1alpha	DYRK1B, HNF1A, MAP2K3, PCBD2	HNF1 homeobox A, dual specificity tyrosine phosphorylation regulated kinase 1B, mitogen-activated pr...	497	629.12

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.
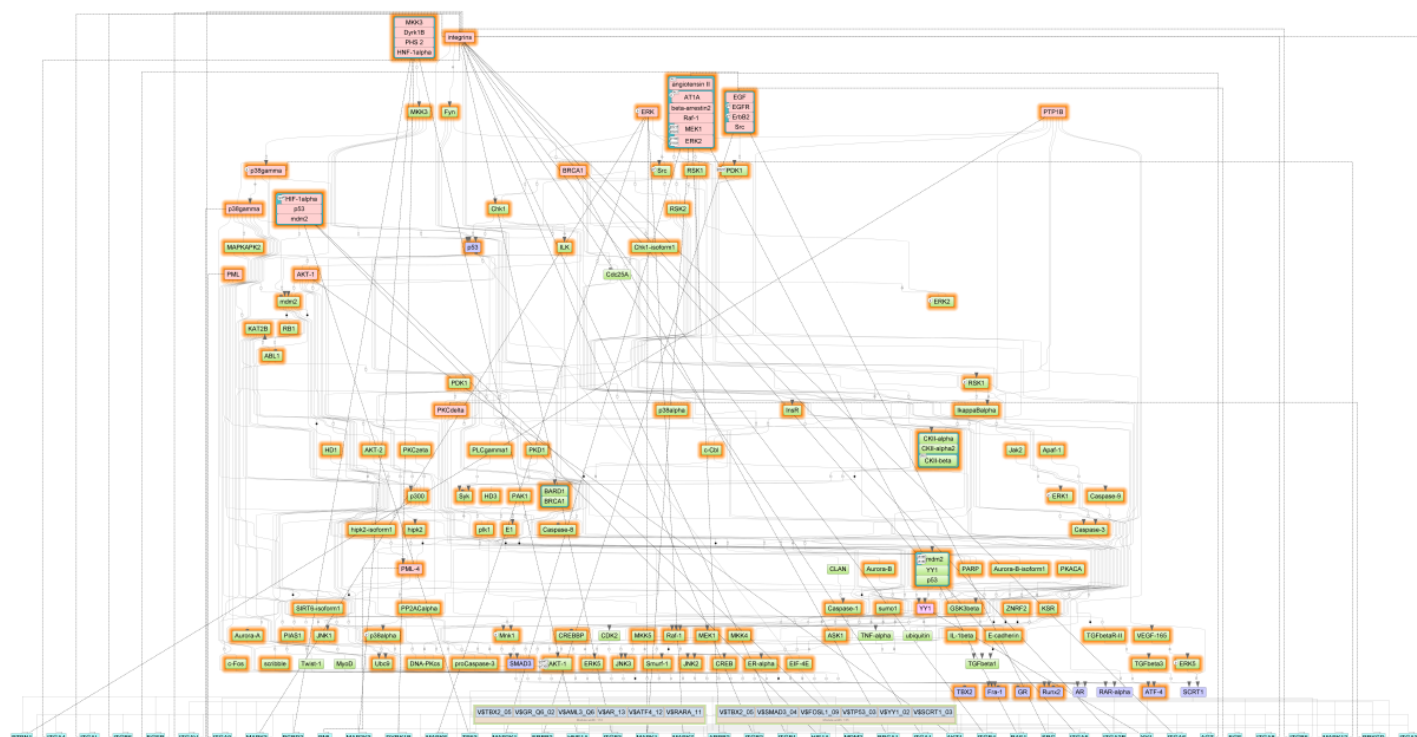
Figure 5. Diagram of intracellular regulatory signal transduction pathways of the most frequently mutated genes in Experiment: short-term survival. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have 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 [11-13] 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 [11-13] 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
ITGAL	integrin subunit alpha L	14	422	191.78
ITGA3	integrin subunit alpha 3	2	422	191.78
ITGB5	integrin subunit beta 5	2	422	191.78
ITGA6	integrin subunit alpha 6	1	422	191.78
ITGA4	integrin subunit alpha 4	5	422	191.78
MAPK12	mitogen-activated protein kinase 12	60	671	114.67

*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
ITGAL	integrin subunit alpha L	0.64	422	191.78
MAPK12	mitogen-activated protein kinase 12	36.68	671	114.67
PCBD2	pterin-4 alpha-carbinolamine dehydratase 2	0.35	958	629.12
PSMC5	proteasome 26S subunit, ATPase 5	1.28	1023	91.75
PSMD5	proteasome 26S subunit, non-ATPase 5	1.28	1023	91.75
PSMA7	proteasome 20S subunit alpha 7	2.48	1023	91.75

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:

- p38gamma
- integrins
- PML

This result allows us to suggest the following schema of affecting the molecular mechanism of the studied pathology:
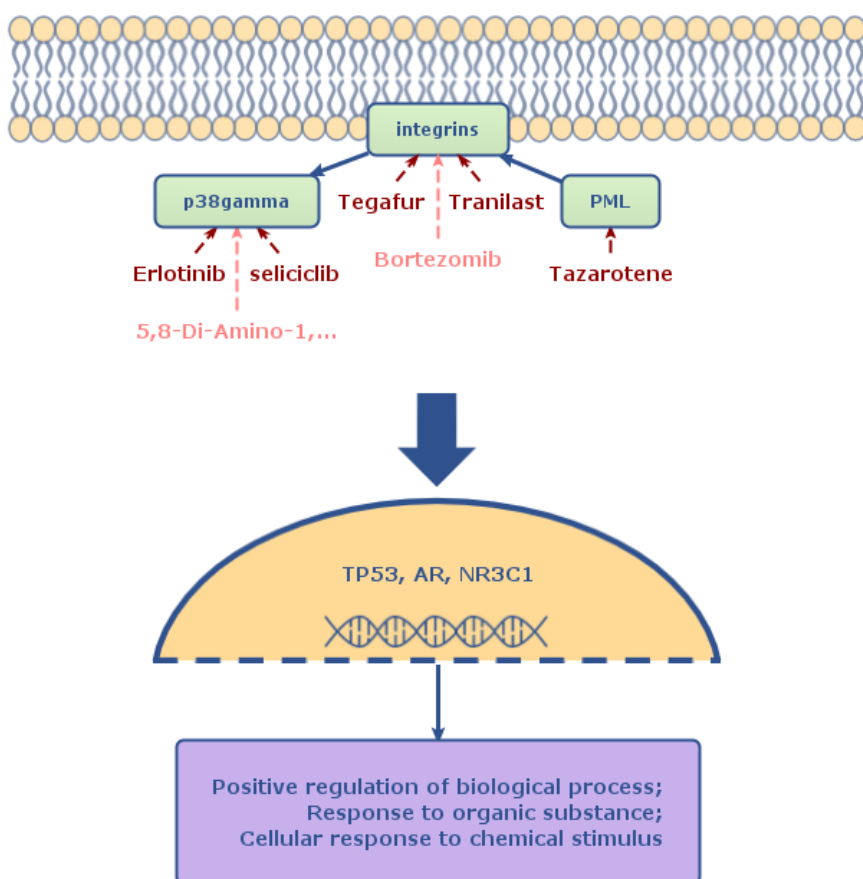
Drugs which are shown on this schema: Tranilast, Bortezomib, Erlotinib, Tazarotene, Tegafur, seliciclib and 5,8-Di-Amino-1,4-Dihydroxy-Anthraquinone, should be considered as a prospective research initiative for further drug repurposing and drug development. These drugs were selected as top matching 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, PDGFRB, FGFR1, BRAF, PDGFRA, RAF1, FRS2, MAPK11, FLT1, AKT1, TEK, MAPK3, RET, AKT2	95	12	Phase 4: Colorectal Neoplasms, Neoplasms, Rectal Neoplasms	Colorectal Neoplasms (FDA)
encorafenib	MAPK10, MAPK1, MAPK8, MAPK9, MAPK7, MAPK6, LIMK1, BRAF, MAPK3, MAP2K4, RAF1	90	7	Phase 3: Colorectal Neoplasms, Melanoma, Neoplasms, Rectal Neoplasms	Colorectal Neoplasms (ClinicalTrials, ClinicalTrials, FDA)
Fluorouracil	CASP8, CASP3, BAX, ITGAL, FAS, BIRC5, CDKN1A	85	12	Phase 4: Colorectal Neoplasms, Adenocarcinoma, Bowen's Disease, Breast Neoplasms, Carcinoma, Carcinoma, Basal Cell, Carcinoma, Squamous Cell, Cicatrix, Colonic Neoplasms, Digestive System Neoplasms, Esophageal Neoplasms, Foot Diseases, Gastrointestinal Neoplasms, Glaucoma, Glaucoma, Open-Angle, Head and Neck Neoplasms, Hypopigmentation, Intestinal Neoplasms, Keloid, Keratosis, Keratosis, Actinic, Liver Neoplasms, Neoplasm Metastasis, Neoplasms, Neoplasms, Basal Cell, Neoplasms, Second Primary, Neoplasms, Squamous Cell, Pancreatic Neoplasms, Photosensitivity Disorders, Postoperative Complications, Pterygium, Rectal Neoplasms, Recurrence, Skin Diseases, Skin Neoplasms, Squamous Cell Carcinoma of Head and Neck, Stomach Neoplasms, Vitiligo, Warts	Colorectal Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials)
Oxaliplatin	CTNNB1, IGF2, CDH1, AKT1, AKT2	83	12	Phase 4: Colorectal Neoplasms, Adenocarcinoma, Colonic Neoplasms, Digestive System Neoplasms, Esophageal Neoplasms, Gastrointestinal Neoplasms, Intestinal Neoplasms, Neoplasm Metastasis, Neoplasms, Neurotoxicity Syndromes, Postoperative Complications, Rectal Neoplasms, Stomach Neoplasms	Colorectal Neoplasms (FDA, PUBMED)
Aflibercept	VEGFA, PGF	67	7	Phase 3: Colorectal Neoplasms, COVID-19, Choroidal Neovascularization, Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetes Mellitus, Type 2, Edema, Eye Diseases, Glaucoma, Glaucoma, Neovascular, Macular Degeneration, Macular Edema, Melanoma, Myopia, Myopia, Degenerative, Neoplasm Metastasis, Neoplasms, Neovascularization, Pathologic, Premature Birth, Prostatic Neoplasms, Rectal Neoplasms, Retinal Degeneration, Retinal Diseases, Retinal Vein Occlusion, Retinopathy of Prematurity, Telangiectasis, Vision Disorders, Vision, Low, Wet Macular Degeneration	Colorectal Neoplasms (FDA)
Irinotecan	MAPK10, BIRC5, CDKN1A	63	12	Phase 4: Colorectal Neoplasms, Adenocarcinoma, Carcinoma, Small Cell, Colonic Neoplasms, Esophageal Neoplasms, Gastrointestinal Neoplasms, Lung Neoplasms, Neoplasm Metastasis, Neoplasms, Neuroblastoma, Rectal Neoplasms, Small Cell Lung Carcinoma, Stomach Neoplasms	Colorectal Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, FDA, FDA)
Bevacizumab	FCGR2A, VEGFA	62	12	Phase 4: Colorectal Neoplasms, Breast Neoplasms, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Cataract, Cicatrix, Colonic Neoplasms, Corneal Neovascularization, Diabetic Retinopathy, Dilatation, Pathologic, Edema, Epistaxis, Fallopian Tube Neoplasms, Glaucoma, Hemorrhage, Ischemia, Liver Neoplasms, Lung Neoplasms, Macular Degeneration, Macular Edema, Neoplasm Metastasis, Neoplasms, Neovascularization, Pathologic, Optic Nerve Diseases, Ovarian Neoplasms, Peritoneal Neoplasms, Pterygium, Rectal Neoplasms, Recurrence, Retinal Detachment, Retinal Diseases, Retinal Vein Occlusion, Telangiectasia, Hereditary Hemorrhagic, Telangiectasis, Triple Negative Breast Neoplasms, Vitreous Hemorrhage	Colorectal Neoplasms (FDA, FDA)
Panitumumab	EGFR	47	12	Phase 4: Colorectal Neoplasms, Colonic Neoplasms, Neoplasms, Rectal Neoplasms	Colorectal Neoplasms (FDA)
Cetuximab	FCGR2A, EGFR	47	12	Phase 4: Colorectal Neoplasms, Carcinoma, Carcinoma, Squamous Cell, Colonic Neoplasms, Head and Neck Neoplasms, Liver Neoplasms, Neoplasm Metastasis, Neoplasms, Rectal Neoplasms, Squamous Cell Carcinoma of Head and Neck	Colorectal Neoplasms (FDA, FDA)
Capecitabine	CDKN2A	27	12	Phase 4: Colorectal Neoplasms, Adenocarcinoma, Brain Neoplasms, Breast Neoplasms, Carcinoma, Colonic Neoplasms, Gastrointestinal Neoplasms,	Colorectal Neoplasms

Liver Neoplasms, Neoplasm Metastasis, Neoplasms, Neoplasms, Second Primary, Rectal Neoplasms, Stomach 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
Erlotinib	BMPR1A, IKBKE, ABL1, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BIRC5, BMX, ERBB3, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K1, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, MAP2K4, PRKAA1, RPS6KA1, ILK, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, ERBB4, PAK3, BRAF, FER, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	98	7	Phase 3: Colorectal Neoplasms, Adenocarcinoma, Adenocarcinoma of Lung, Brain Neoplasms, Carcinoma, Carcinoma, Acinar Cell, Carcinoma, Adenosquamous, Carcinoma, Hepatocellular, Carcinoma, Large Cell, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Disease Progression, Esophageal Neoplasms, Fallopian Tube Neoplasms, Head and Neck Neoplasms, Lip Neoplasms, Lung Neoplasms, Mouth Neoplasms, Neoplasm Metastasis, Neoplasms, Ovarian Neoplasms, Pancreatic Intraductal Neoplasms, Pancreatic Neoplasms, Peritoneal Neoplasms, Rectal Neoplasms, Small Cell Lung Carcinoma, Squamous Cell Carcinoma of Head and Neck, Thoracic Neoplasms
Sunitinib	BMPR1A, IKBKE, ABL1, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, CAMKK1, EPHB1, SIRT1, TGFBR2, FYN, PRKD2	98	7	Phase 3: Colorectal Neoplasms, Adenoma, Adenoma, Islet Cell, Brain Neoplasms, Breast Neoplasms, Carcinoma, Carcinoma, Hepatocellular, Carcinoma, Islet Cell, Carcinoma, Non-Small-Cell Lung, Carcinoma, Renal Cell, Gastrointestinal Stromal Tumors, Glioblastoma, Kidney Diseases, Kidney Neoplasms, Liver Neoplasms, Lung Neoplasms, Neoplasms, Neoplasms by Site, Pancreatic Neoplasms, Prostatic Neoplasms, Rectal Neoplasms, Recurrence, Urogenital Neoplasms, Urologic Neoplasms
Gefitinib	BMPR1A, IKBKE, ABL1, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, IGF2, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK,	98	7	Phase 3: Colorectal Neoplasms, Adenocarcinoma, Adenocarcinoma of Lung, Adenocarcinoma, Bronchiolo-Alveolar, Brain Neoplasms, Breast Neoplasms, Carcinoma, Carcinoma, Non-Small-Cell Lung, Carcinoma, Squamous Cell, Head and Neck Neoplasms, Lung Neoplasms, Neoplasm Metastasis, Neoplasms, Rectal Neoplasms, Squamous Cell Carcinoma of Head and Neck, Thoracic Neoplasms, Urinary Bladder Neoplasms

	LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2			
Tegafur	ITGA5, VEGFA, ITGB1, MTOR, ITGB4, FGF2, ITGA3, ITGA6, ITGB5, EGFR, PTK2, CASP3, AKT1, ITGB3, AKT2	96	7	Phase 3: Colorectal Neoplasms, Adenocarcinoma, Ascites, Breast Neoplasms, Carcinoma, Cholangiocarcinoma, Colonic Neoplasms, Esophageal Neoplasms, Head and Neck Neoplasms, Liver Neoplasms, Nasopharyngeal Carcinoma, Neoplasm Metastasis, Neoplasms, Pancreatic Neoplasms, Rectal Neoplasms, Recurrence, Stomach Neoplasms, Urinary Bladder Neoplasms
Lapatinib	BMPR1A, IKBKE, ABL1, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BIRC5, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	95	5	Phase 2: Colorectal Neoplasms, Adenocarcinoma, Adenoma, Astrocytoma, Bile Duct Neoplasms, Brain Neoplasms, Breast Diseases, Breast Neoplasms, Breast Neoplasms, Male, Carcinoma, Carcinoma, Acinar Cell, Carcinoma, Adenoid Cystic, Carcinoma, Ductal, Carcinoma, Ductal, Breast, Carcinoma, Hepatocellular, Carcinoma, Mucoepidermoid, Carcinoma, Non-Small-Cell Lung, Carcinoma, Ovarian Epithelial, Carcinoma, Small Cell, Carcinoma, Squamous Cell, Carcinoma, Transitional Cell, Carcinoma, Verrucous, Central Nervous System Neoplasms, Cholangiocarcinoma, Cysts, Dermoid Cyst, Diarrhea, Digestive System Neoplasms, Endometrial Neoplasms, Ependymoma, Esophageal Neoplasms, Esophageal Squamous Cell Carcinoma, Fibroma, Gallbladder Neoplasms, Gastrointestinal Neoplasms, Glioblastoma, Glioma, Gliosarcoma, Head and Neck Neoplasms, Hypopharyngeal Neoplasms, Inflammatory Breast Neoplasms, Intestinal Neoplasms, Laryngeal Diseases, Laryngeal Neoplasms, Liver Neoplasms, Lung Neoplasms, Medulloblastoma, Melanoma, Mucoepidermoid Tumor, Nasopharyngeal Carcinoma, Neoplasm Metastasis, Neoplasms, Neoplasms, Second Primary, Neoplasms, Unknown Primary, Neoplastic Cells, Circulating, Nervous System Neoplasms, Neurilemmoma, Neuroblastoma, Neurofibroma, Neurofibromatoses, Neurofibromatosis 1, Neurofibromatosis 2, Neuroma, Neuroma, Acoustic, Oligodendroglioma, Oropharyngeal Neoplasms, Ovarian Neoplasms, Pancreatic Neoplasms, Paranasal Sinus Neoplasms, Peritoneal Neoplasms, Pharyngeal Neoplasms, Pituitary Diseases, Pituitary Neoplasms, Prolactinoma, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Rectal Neoplasms, Recurrence, Salivary Gland Neoplasms, Sarcoma, Small Cell Lung Carcinoma, Spinal Cord Neoplasms, Squamous Cell Carcinoma of Head and Neck, Stomach Neoplasms, Thymoma, Tongue Neoplasms, Urinary Bladder Neoplasms, Uterine Cervical Neoplasms, Uterine Neoplasms

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

Repurposing drugs

Table 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, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, CDK4, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, ROCK2, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, SGK1, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, CDK1, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	82	Phase 2: ACTH-Secreting Pituitary Adenoma, Adenoma, Carcinoma, Non-Small-Cell Lung, Cystic Fibrosis, Cysts, Fibrosis, Pituitary ACTH Hypersecretion, Pituitary Neoplasms
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, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, ROCK2, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, SGK1, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	82	Phase 2: Arthritis, Arthritis, Rheumatoid, Psoriasis
ruboxistaurin	BMPR1A, IKBKE, ABL1, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, CHEK2, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BMX, HIPK2, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, ROCK2, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, SGK1, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, IKBKB, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	82	Phase 3: Diabetes Mellitus, Diabetes Mellitus, Type 1, Diabetes Mellitus, Type 2, Diabetic Neuropathies, Diabetic Retinopathy, Edema, Macular Edema, Nervous System Diseases, Peripheral Nervous System Diseases, Retinal Diseases
Tofacitinib	BMPR1A, IKBKE, ABL1, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14,	82	Phase 4: Alopecia, Alopecia Areata, Aortic Arch Syndromes, Arteritis, Arthritis, Arthritis, Psoriatic, Arthritis, Rheumatoid, COVID-19, Colitis, Colitis, Ulcerative, Disease, Embolism, Granuloma, Granulomatosis with Polyangiitis, Infections, Lung Diseases, Lung Diseases, Interstitial, Necrosis, Rheumatic Fever, ST Elevation Myocardial Infarction, Spondylarthritis, Spondylitis, Spondylitis, Ankylosing, Systemic Vasculitis, Takayasu Arteritis, Thromboembolism, Ulcer, Vasculitis

CSNK1D, ACVR1, CHEK1, PTK6, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, ROCK2, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, SGK1, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, RPS6KB1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, MAPK13, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2

Flavopiridol	BMPR1A, IKBKE, ABL1, NEK7, JAK3, PAK2, PRKACA, GSK3B, MAP3K11, CDK4, SYK, NEK6, EIF2AK2, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, TGFBR1, SRC, CAMKK2, PRKD3, CSNK1G2, MAP4K1, STK11, CSNK1G1, CLK4, MAPK12, PLK3, WEE1, MAPK7, MAPK11, FLT1, PRKCD, LYN, CSNK2A1, AKT2, STK3, PDPK1, MAPK10, CDK6, KDR, MAPK6, PRKCA, CSNK1E, PDGFRA, RAF1, AURKB, CDK7, MAPK14, CSNK1D, ACVR1, CHEK1, PTK6, BMX, PIM2, RIPK2, FGFR3, RPS6KA2, PAK4, NTRK2, PIM3, YES1, NUAK1, CAMK2D, ITK, XIAP, MAP3K4, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, STK10, MAP4K4, MARK3, MAP3K10, SLK, INSR, NEK2, MAP2K2, MAP2K3, RPS6KA5, CSK, PRKD1, TEK, PKN1, TYK2, IGF1R, MAP3K5, CSNK2A2, TNK2, PDGFRB, MAP2K1, MAP4K3, STK4, CAMK2A, PTK2, MAPKAPK2, FES, PLK4, DAPK3, RPS6KA3, CAMK2G, MET, PRKCQ, EPHA4, CDK1, MAP2K4, PRKAA1, RPS6KA1, EGFR, PLK1, PRKCH, ACVR2A, PTK2B, AKT1, AURKA, JAK2, MAP3K20, PKMYT1, KIT, MAPK1, DYRK1B, EPHB2, MAPK9, GSK3A, MERTK, LATS2, PRKCE, FGFR1, PAK3, BRAF, FER, CAMKK1, EPHB1, TGFBR2, FYN, PRKD2	82	Phase 2: Adenocarcinoma, Brain Abscess, Breast Neoplasms, Carcinoma, Hepatocellular, Carcinoma, Ovarian Epithelial, Carcinoma, Renal Cell, Embolism, Endometrial Neoplasms, Esophageal Neoplasms, Germinoma, Granuloma, Head and Neck Neoplasms, Hodgkin Disease, Hypereosinophilic Syndrome, Immunoblastic Lymphadenopathy, Kidney Neoplasms, Leukemia, Leukemia, Basophilic, Acute, Leukemia, Eosinophilic, Acute, Leukemia, Erythroblastic, Acute, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, Leukemia, Megakaryoblastic, Acute, Leukemia, Monocytic, Acute, Leukemia, Myeloid, Leukemia, Myeloid, Acute, Leukemia, Myelomonocytic, Acute, Leukemia, Prolymphocytic, Leukemia, T-Cell, Leukemia-Lymphoma, Adult T-Cell, Liver Neoplasms, Lymphadenopathy, Lymphatic Diseases, Lymphoma, Lymphoma, B-Cell, Lymphoma, B-Cell, Marginal Zone, Lymphoma, Follicular, Lymphoma, Large B-Cell, Diffuse, Lymphoma, Large-Cell, Anaplastic, Lymphoma, Mantle-Cell, Lymphoma, Non-Hodgkin, Lymphoma, T-Cell, Lymphoma, T-Cell, Cutaneous, Lymphomatoid Granulomatosis, Melanoma, Multiple Myeloma, Mycoses, Mycosis Fungoides, Myelodysplastic Syndromes, Neoplasms, Neoplasms, Germ Cell and Embryonal, Neoplasms, Plasma Cell, Ovarian Neoplasms, Pancreatic Neoplasms, Peritoneal Neoplasms, Precursor Cell Lymphoblastic Leukemia-Lymphoma, Preleukemia, Prostatic Neoplasms, Prostatic Neoplasms, Castration-Resistant, Recurrence, Sarcoma, Seminoma, Sezary Syndrome, Stomach Neoplasms, Testicular Neoplasms, Thromboembolism, 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
Camptothecin	HIF1A, CASP3, NFKB1	82	0.69
Topotecan	HIF1A, CASP3, NFKB1	82	0.69
LE-SN38	HIF1A, CASP3, NFKB1	82	0.65
Ouabain	NFKB2, STAT3, CASP8, HIF1A, CASP3, STAT1, NFKB1, RELA	79	0.44
MGI-114	GDNF, CASP8, CASP9	74	0.33

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	PAK2, SENP8, CDK4, SGK1, MAP2K3, LIMK1, PKN1, MAP2K6, MAPK8, PRKD3, CSNK1G2, CLK4, CSNK1G1, CAMK2A, IRAK1, CHEK2, FES, AKT2, DAPK3, MAPK10, RPS6KA3, IRAK4, CDK6, CAMK2G, CSNK1A1, STAT1, EPHA4, CSNK1E, CDK1, MAP2K4, AURKB, CDK7, PRKAA1, HSP90AA1, CSNK1D, PTK2B, AKT1, AURKA, CHEK1, MAPK9, EPHB2, DYRK1B, GSK3A, PAK4, PAK3, CAMK2D, EPHB1, PAK1, CLK1, CDK5	96	13.49
Iodophenyl	STK10, ROCK2, MAP4K4, MARK3, NEK7, PAK2, GSK3B, SLK, VRK1, NEK2, SGK1, NEK6, EIF2AK2, TAOK2, RPS6KA5, LIMK1, MELK, CSNK2A2, TNK2, CSNK1G2, MAP4K1, TRIO, STK11, MAP4K3, CSNK1G1, STK4, NEK10, MAP4K2, MAPKAPK2, IRAK1, CSNK2A1, CHEK2, AKT2, PDPK1, DAPK3, STK3, RPS6KA3, PIK3CG, IRAK4, PINK1, CSNK1A1, CSNK1E, RAF1, PRKAA1, RPS6KA1, AURKB, ILK, CSNK1D, MASTL, HIPK1, AKT1, AURKA, CHEK1, TAOK1, ATM, LMTK2, PKMYT1, RPS6KB1, TBK1, ATR, RIPK2, PIM2, HIPK2, MTOR, GSK3A, RPS6KA2, PAK4, PIM3, LATS2, PRKDC, NUAK1, PAK3, PASK, BRAF, PAK1, CDK5	96	22.06
Rbt205 Inhibitor	RPS6KA3, CDK6, CAMK2G, GRK2, MAP3K10, PRKCQ, PRKACA, GSK3B, PRKCA, CDK1, MAP2K4, CDK4, CDK7, RPS6KA1, PRKAA1, SGK1, PRKCH, PRKD1, PKN1, MAP2K6, RPS6KA2, GSK3A, GRK5, PRKD3, PRKCE, PRKCZ, CAMK2D, PTK2, CAMK2A, SIRT1, PRKCD, PRKCI, CDK5, PRKACB, DAPK3	95	20.5
7-[4-(Dimethylamino)Phenyl]-N-Hydroxy-4,6-Dimethyl-7-Oxo-2,4-Heptadienamide	HDAC4, HDAC2, HDAC6, HDAC7, HDAC3, HDAC5, HDAC1	95	7.96
Uracil mustard	MET, KDR, ABL1, JAK3, EPHA4, PDGFRA, INSR, MST1R, EGFR, SYK, PTK2B, EIF2AK2, ERBB2, PTK6, JAK2, TEK, CSK, TYK2, MELK, JAK1, IGF1R, KIT, BMX, RIPK2, ERBB3, EPHB2, FGFR3, SRC, MTOR, NTRK2, MERTK, PDGFRB, TNK2, YES1, FGFR1, ERBB4, FER, ITK, EPHB1, WEE1, PTK2, FLT1, PRKCD, LYN, FYN, RET, FES, ABL2	93	6.94

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Erlotinib, seliciclib, Camptothecin and 2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One. These drugs were selected for acting on the following targets: MAPK12, NFKB1 and CSNK1E, 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	-	
Abemaciclib	Breast Neoplasms	58	
Abiraterone	Prostatic Neoplasms, Castration-Resistant	-	
Abiraterone acetate	Prostatic Neoplasms, Castration-Resistant	-	
Acalabrutinib	Lymphoma, Mantle-Cell	48	
Acitretin	Psoriasis	40	
Ado-trastuzumab emtansine	Breast Neoplasms|Neoplasms	78	
Afatinib	Carcinoma, Non-Small-Cell Lung	78	
Aflibercept	Colorectal Neoplasms|Diabetic Retinopathy|Edema|Vascular Diseases|Wet Macular Degeneration	67	
Alectinib	Carcinoma, Non-Small-Cell Lung	62	
Alemtuzumab	Brain Abscess|Leukemia, Lymphocytic, Chronic, B-Cell|Multiple Sclerosis|Multiple Sclerosis, Relapsing-Remitting|Sclerosis	-	
Alitretinoin	Sarcoma, Kaposi	-	
Alpelisib	Breast Neoplasms	-	
Altretamine	Ovarian Neoplasms	-	
Aminolevulinic acid	Keratosis|Keratosis, Actinic	-	
Anagrelide	Thrombocythemia, Essential|Thrombocytosis	-	
Anastrozole	Breast Neoplasms|Hypersensitivity|Obesity|Obesity, Morbid|Recurrence|Weight Loss	53	
Apalutamide	Prostatic Neoplasms, Castration-Resistant	-	
Aprepitant	Nausea|Neoplasms|Postoperative Nausea and Vomiting	-	
Arsenic trioxide	Leukemia, Promyelocytic, Acute	72	
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	89	
Azacitidine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes|Preleukemia|Syndrome	33	
Belinostat	Lymphoma, T-Cell, Peripheral	74	
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	62	
Bexarotene	Lymphoma, T-Cell|Lymphoma, T-Cell, Cutaneous	14	
Bicalutamide	Prostatic Neoplasms	42	
Binimetinib	Melanoma	76	
Blinatumomab	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-	
Bortezomib	Brain Abscess|Glomerulonephritis|Glomerulonephritis, IGA|Kidney Diseases|Multiple Myeloma|Neoplasms, Plasma Cell|Nephritis|Renal Insufficiency	66	
Bosutinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	75	
Brentuximab vedotin	Hodgkin Disease|Lymphoma|Lymphoma, Large-Cell, Anaplastic|Lymphoma, T-Cell, Peripheral	-	
Brigatinib	Carcinoma, Non-Small-Cell Lung	53	
Buserelin	Prostatic Neoplasms	-	

Cabazitaxel	Prostatic Neoplasms, Castration-Resistant	76	
Cabergoline	Drug-Related Side Effects and Adverse Reactions|Pituitary Neoplasms	-	
Cabozantinib	Thyroid Neoplasms	81	
Capecitabine	Breast Neoplasms|Colonic Neoplasms|Colorectal Neoplasms	27	
Carboplatin	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms|Neuroendocrine Tumors|Ovarian Neoplasms|Retinoblastoma	-	XRCC1:Q399R:sensitivity/response:B2
Carfilzomib	Multiple Myeloma	68	
Carmustine	Astrocytoma|Glioblastoma|Hodgkin Disease|Medulloblastoma|Multiple Myeloma|Neoplasms	34	
Ceritinib	Carcinoma, Non-Small-Cell Lung	77	
Cetuximab	Colorectal Neoplasms	47	
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	82	XRCC1:Q399R:sensitivity/response:B2
Cladribine	Leukemia, Hairy Cell	48	
Clofarabine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	42	
Cobimetinib	Melanoma	82	
Copanlisib	Lymphoma, Follicular	80	
Crizotinib	Carcinoma, Non-Small-Cell Lung	91	
Cyproterone acetate	Prostatic Neoplasms	6	
Dabrafenib	Melanoma	58	
Dacomitinib	Carcinoma, Non-Small-Cell Lung	83	
Daratumumab	Multiple Myeloma	-	
Dasatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase|Precursor Cell Lymphoblastic Leukemia-Lymphoma	93	
Decitabine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes	18	
Degarelix	Cardiovascular Diseases|Prostatic Neoplasms|Vascular Diseases	54	
Denosumab	Arthritis, Rheumatoid|Bone Diseases|Bone Diseases, Metabolic|Breast Neoplasms|Hyperparathyroidism|Hyperparathyroidism, Primary|Metabolic Diseases|Neoplasm Metastasis|Neoplasms|Osteoporosis|Osteoporosis, Postmenopausal|Prostatic Neoplasms	46	
Dexrazoxane	Breast Neoplasms|Cardiomyopathies	20	
Dienogest	Menorrhagia	35	
Dinutuximab	Neuroblastoma	-	
Docetaxel	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Prostatic Neoplasms|Squamous Cell Carcinoma of Head and Neck|Stomach Neoplasms	44	
Doxorubicin	Neoplasms|Multiple Myeloma|Carcinoma, Ovarian Epithelial|Ovarian Neoplasms|Leukemia, Lymphoid|Breast Neoplasms|Lymphoma, Follicular|Thyroid Neoplasms|Triple Negative Breast Neoplasms|Glioma	88	
Durvalumab	Carcinoma, Non-Small-Cell Lung|Carcinoma, Transitional Cell	-	
Dutasteride	Alcoholism|Hyperplasia|Hypertrophy|Neoplasms|Prostatic Hyperplasia	-	
Duvelisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	63	
Elotuzumab	Multiple Myeloma	69	
Enasidenib	Leukemia, Myeloid, Acute	-	
Encorafenib	Colorectal Neoplasms|Melanoma	90	
Enfortumab vedotin	Carcinoma, Transitional Cell|Neoplasms	-	
Entrectinib	Carcinoma, Non-Small-Cell Lung	55	
Enzalutamide	Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	-	
Epirubicin	Breast Neoplasms	57	
Erdafitinib	Urinary Bladder Neoplasms	82	
Eribulin	Breast Neoplasms|Drug-Related Side Effects and Adverse Reactions|Neoplasms	-	
Erlotinib	Carcinoma, Non-Small-Cell Lung|Neoplasms|Pancreatic Neoplasms	98	
Erlotinib hydrochloride	Carcinoma, Non-Small-Cell Lung|Gastrointestinal Stromal Tumors	-	
Estramustine	Prostatic Neoplasms	29	
Ethinyl Estradiol	Acne Vulgaris|Neoplasms	21	
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	85	
Exemestane	Breast Neoplasms	25	
Fedratinib	Primary Myelofibrosis	54	

Finasteride	Hyperplasia|Neoplasms|Prostatic Hyperplasia	41
Flavopiridol	Leukemia, Lymphocytic, Chronic, B-Cell	82
Fluorouracil	Skin Neoplasms|Neoplasms, Basal Cell|Neoplasms, Second Primary|Neoplasms, Squamous Cell|Neoplasms|Colorectal Neoplasms|Pancreatic Neoplasms	85
Fluoxymesterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	39
Flutamide	Premenstrual Dysphoric Disorder|Premenstrual Syndrome|Prostatic Neoplasms	57
Fulvestrant	Breast Neoplasms	69
Gefitinib	Carcinoma, Non-Small-Cell Lung	98
Gemcitabine	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Ovarian Neoplasms|Pancreatic Neoplasms	81
Gemtuzumab ozogamicin	Leukemia, Myeloid, Acute	12
Gilteritinib	Leukemia, Myeloid, Acute	65
Glasdegib	Leukemia, Myeloid, Acute	-
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	74
Ibritumomab	Lymphoma, B-Cell|Lymphoma, Follicular	12
Ibrutinib	Graft vs Host Disease|Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, B-Cell, Marginal Zone|Lymphoma, Mantle-Cell|Waldenstrom Macroglobulinemia	88
Idarubicin	Leukemia, Myeloid, Acute	36
Idelalisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	66
Ifosfamide	Neoplasms	-
Imatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Mastocytosis, Systemic|Neoplasms	92
Inotuzumab ozogamicin	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Ipilimumab	Carcinoma, Renal Cell|Melanoma	48
Irinotecan	Colorectal Neoplasms	63
Ivosidenib	Leukemia, Myeloid, Acute	-
Ixabepilone	Breast Neoplasms	-
Ixazomib	Multiple Myeloma	-
Lapatinib	Breast Neoplasms	95
Larotrectinib	Neoplasm Metastasis	61
Lenalidomide	Brain Abscess|Lupus Erythematosus, Cutaneous|Myelodysplastic Syndromes|Neoplasms, Plasma Cell	55
Lenvatinib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	87
Letrozole	Breast Neoplasms|Cysts|Fibroma|Myofibroma|Myoma|Ovarian Cysts|Syndrome	41
Leuprolide	Hot Flashes|Ovarian Hyperstimulation Syndrome|Prostatic Neoplasms|Puberty, Precocious	-
Levamisole	Ascariasis|Colonic Neoplasms|Helminthiasis	-
Levonorgestrel	Epilepsy|Hyperplasia|Menorrhagia	39
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	40
Lorlatinib	Carcinoma, Non-Small-Cell Lung	69
Masoprocol	Keratosis, Actinic	18
Medroxyprogesterone Acetate	Depression|Depression, Postpartum|Depressive Disorder|Metrorrhagia|Neoplasms|Uterine Hemorrhage	59
Megestrol acetate	Acquired Immunodeficiency Syndrome|Bites and Stings|Breast Neoplasms|Pain|Wasting Syndrome	48
Methotrexate	Neoplasms|Breast Neoplasms|Head and Neck Neoplasms|Ovarian Neoplasms|Lymphoma, T-Cell, Peripheral|Brain Neoplasms|Colorectal Neoplasms|Neuroblastoma|Carcinoma, Squamous Cell	67
Methyltestosterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	-
Midostaurin	Leukemia, Mast-Cell|Leukemia, Myeloid, Acute|Mastocytosis, Systemic	81
Mitotane	Adrenocortical Carcinoma	5
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	41
Mogamulizumab	Mycosis Fungoides|Neoplasms|Sezary Syndrome	-
Moxetumomab pasudotox	Leukemia, Hairy Cell|Neoplasms	31
Necitumumab	Carcinoma, Non-Small-Cell Lung|Neoplasms	-
Nelarabine	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma	-

Neratinib	Breast Neoplasms	88
Nilotinib	Blast Crisis|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase	72
Nilutamide	Prostatic Neoplasms	-
Nintedanib	Fibrosis|Idiopathic Pulmonary Fibrosis	91
Niraparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms	45
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	52
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	45
Olaratumab	Sarcoma	-
Osimertinib	Carcinoma, Non-Small-Cell Lung	87
Oxaliplatin	Colonic Neoplasms|Colorectal Neoplasms|Neoplasms|Rectal Neoplasms	83
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	80
Palbociclib	Breast Neoplasms	51
Panitumumab	Colorectal Neoplasms	47
Panobinostat	Multiple Myeloma	77
Pazopanib	Carcinoma|Carcinoma, Renal Cell|Sarcoma	83
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	5
Pertuzumab	Breast Neoplasms	83
Pomalidomide	Multiple Myeloma	21
Ponatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Precursor Cell Lymphoblastic Leukemia-Lymphoma	77
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	95
Relugolix	Prostatic Neoplasms	-
Ribociclib	Breast Neoplasms	58
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	12
Romidepsin	Lymphoma, T-Cell, Cutaneous	86
Rucaparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms|Prostatic Neoplasms, Castration-Resistant	77
Ruxolitinib	Graft vs Host Disease|Polycythemia|Polycythemia Vera|Primary Myelofibrosis|Thrombocytosis	83
Selinexor	Multiple Myeloma	42
Selumetinib	Neurofibromatosis 1	83
Siltuximab	Giant Lymph Node Hyperplasia	-
Sirolimus	Angiomyolipoma|Constriction, Pathologic|Coronary Restenosis|Eye Diseases|Immune System Diseases|Kidney Failure, Chronic|Lipoma|Tuberous Sclerosis	92
Sonidegib	Carcinoma, Basal Cell	-
Sorafenib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	95
Sunitinib	Adenoma|Carcinoma, Renal Cell|Digestive System Neoplasms|Gastrointestinal Neoplasms|Gastrointestinal Stromal Tumors|Intestinal Neoplasms	98
Talazoparib	Breast Neoplasms	29
Tamoxifen	Breast Diseases|Cystic Fibrosis|Cysts|Fibroadenoma|Fibrocystic Breast Disease|Hemorrhage|Menorrhagia|Menstruation Disturbances|Metrorrhagia|Neoplasms	68
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	39
Temsirolimus	Carcinoma, Renal Cell	90

Teniposide	Precursor Cell Lymphoblastic Leukemia-Lymphoma	49
Thalidomide	Brain Abscess|Immune System Diseases|Multiple Myeloma|Neoplasms, Plasma Cell	69
Tivozanib	Carcinoma, Renal Cell	87
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	42
Toremifene	Breast Neoplasms	56
Trabectedin	Leiomyosarcoma|Liposarcoma	-
Trametinib	Carcinoma, Non-Small-Cell Lung|Melanoma	87
Trastuzumab	Breast Neoplasms|Neoplasms	77
Tretinoin	Lentigo	71
Triptorelin	Fatty Liver|Hypogonadism|Infertility, Female|Prostatic Neoplasms	71
Tucatinib	Breast Neoplasms	87
Valrubicin	Urinary Bladder Neoplasms	-
Vandetanib	Thyroid Neoplasms	95
Vemurafenib	Melanoma	78
Venetoclax	Leukemia, Lymphocytic, Chronic, B-Cell|Leukemia, Myeloid, Acute	15
Vinblastine	Glioma	43
Vincristine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	53
Vinorelbine	Carcinoma, Non-Small-Cell Lung	60
Vismodegib	Carcinoma, Basal Cell	-
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	7

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:

> **Erlotinib, seliciclib, Camptothecin and 2,6-Dihydroanthra/1,9-Cd/Pyrazol-6-One**

These drugs were selected for acting on the following targets: MAPK12, NFKB1 and CSNK1E, 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:

> **p38gamma, integrins and PML**

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: Tranilast, Bortezomib, Erlotinib, Tazarotene, Tegafur, seliciclib and 5,8-Di-Amino-1,4-Dihydroxy-Anthraquinone. 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.

- p38gamma
- integrins
- PML

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 2023.1 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac).

The master regulator search uses the TRANSPATH® database (BIOBASE), release 2023.1 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transpath). A comprehensive signal transduction network of human cells is built by the software on the basis of reactions annotated in TRANSPATH®.

The information about drugs corresponding to identified drug targets and clinical trials references were extracted from HumanPSD™ database, release 2023.1 (https://genexplain.com/humanpsd).

The Ensembl database release Human104.38 (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" ($T\text{-}score_{PSD}$),
   2. ranking by "Disease activity score" ($D\text{-}score_{PSD}$),
   3. ranking by "Clinical validity score".
"Target activity score" ($T\text{-}score_{PSD}$) 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" ($D\text{-}score_{PSD}$):

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

where D is the set of selected diseases, and if D is empty set, $D\text{-}score_{PSD}=0$. 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
0. Boyarskikh U, Pintus S, Mandrik N, Stelmashenko D, Kiselev I, Evshin I, Sharipov R, Stegmaier P, Kolpakov F, Filipenko M, Kel A. Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3. *BMC Med Genomics.* **2018**;11(1):12. doi:10.1186/1471-2105-7-s2-s13
1. Filimonov D, Poroikov V. Probabilistic Approaches in Activity Prediction. Varnek A, Tropsha A. *Cheminformatics Approaches to Virtual Screening.* Cambridge (UK): RSC Publishing. **2008**;:182-216.
2. Filimonov DA, Poroikov VV. Prognosis of specters of biological activity of organic molecules. *Russian chemical journal.* **2006**;50(2):66-75 (russ)
3. Filimonov D, Poroikov V, Borodina Y, Gloriozova T. Chemical Similarity Assessment Through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors. *ChemInform.* **1999**;39(4):666-670. doi:10.1002/chin.199940210

Thank you for using the Genome Enhancer!

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

Supplementary material

Disclaimer

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.