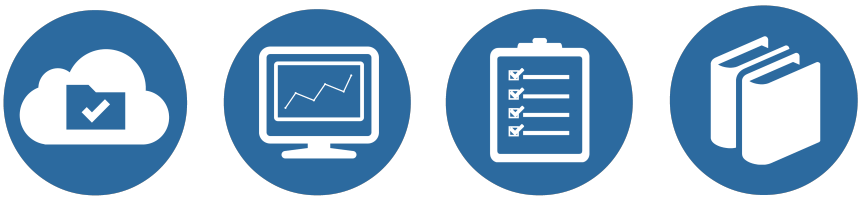
PDGFRA and MTOR are promising druggable targets for treating Non-Small-Cell Lung Carcinoma and Lung Neoplasms that control activity of EP300, ESR2 and POU5F1 transcription factor on promoters of genes carrying sequence variations

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 10/04/2022 ; Run on 26/11/2024 ; Report generated on 26/11/2024

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 *Non-Small-Cell Lung Carcinoma and Lung 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, ESR2 and POU5F1. The subsequent network analysis suggested

- PDGFRalpha
- mTOR

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, ruboxistaurin, 6,7,12,13-tetrahydro-5H-indolo[2,3-a]pyrrolo[3,4-c]carbazol-5-one and 3-[1-(3-Aminopropyl)-1h-Indol-3-Yl]-4-(1-Methyl-1h-Indol-3-Yl)-1h-Pyrrole-2,5-Dione.

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
NCI-H1975	Genomics

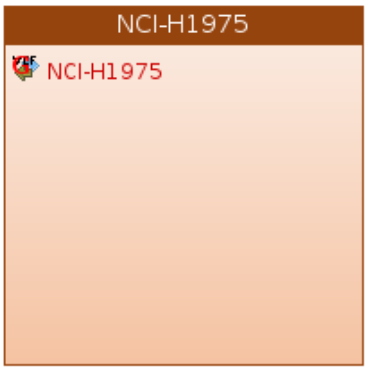


Figure 1. Annotation diagram of experimental data used in this study. With the colored boxes we show those sub-categories of the data that are compared in our analysis.

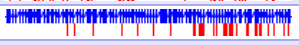
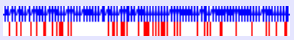
3. Results

We have analyzed the following condition: NCI-H1975.

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 NCI-H1975.
See full table →

ID	Gene description	Gene symbol	Gene schematic representation	Number of variations	Gene weight	Weighted score
ENSG00000034152	mitogen-activated protein kinase kinase 3	MAP2K3		105	262.28	786.84
ENSG00000178104	phosphodiesterase 4D interacting protein	PDE4DIP		128	295.9	591.81
ENSG00000081479	LDL receptor related protein 2	LRP2		54	134.41	403.22
ENSG00000101680	laminin subunit alpha 1	LAMA1		47	117.9	353.7
ENSG00000107611	cubilin	CUBN		46	105.43	316.29
ENSG00000160145	kalirin RhoGEF kinase	KALRN		42	105.43	316.29
ENSG00000168702	LDL receptor related protein 1B	LRP1B		63	152.34	304.67
ENSG00000123384	LDL receptor related protein 1	LRP1		36	92.9	278.71
ENSG00000115414	fibronectin 1	FN1		36	88.85	266.55
ENSG00000095777	myosin IIIA	MYO3A		40	86.23	258.68

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 NCI-H1975:

300 top mutated genes were taken for the mapping.

GO (biological process)

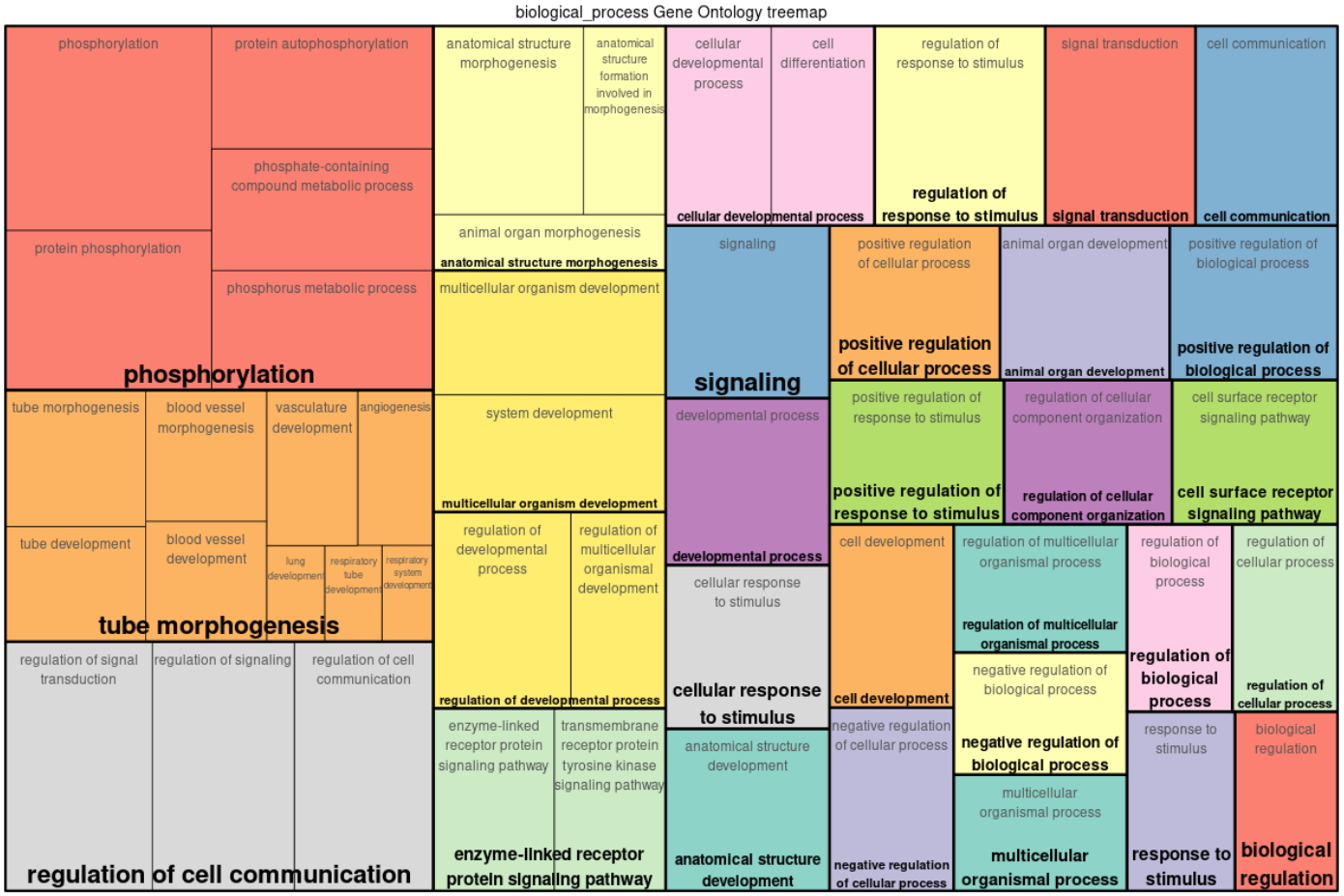


Figure 2. Enriched GO (biological process) of the most frequently mutated genes in NCI-H1975.

Full classification →

TRANSPATH® Pathways (2024.2)

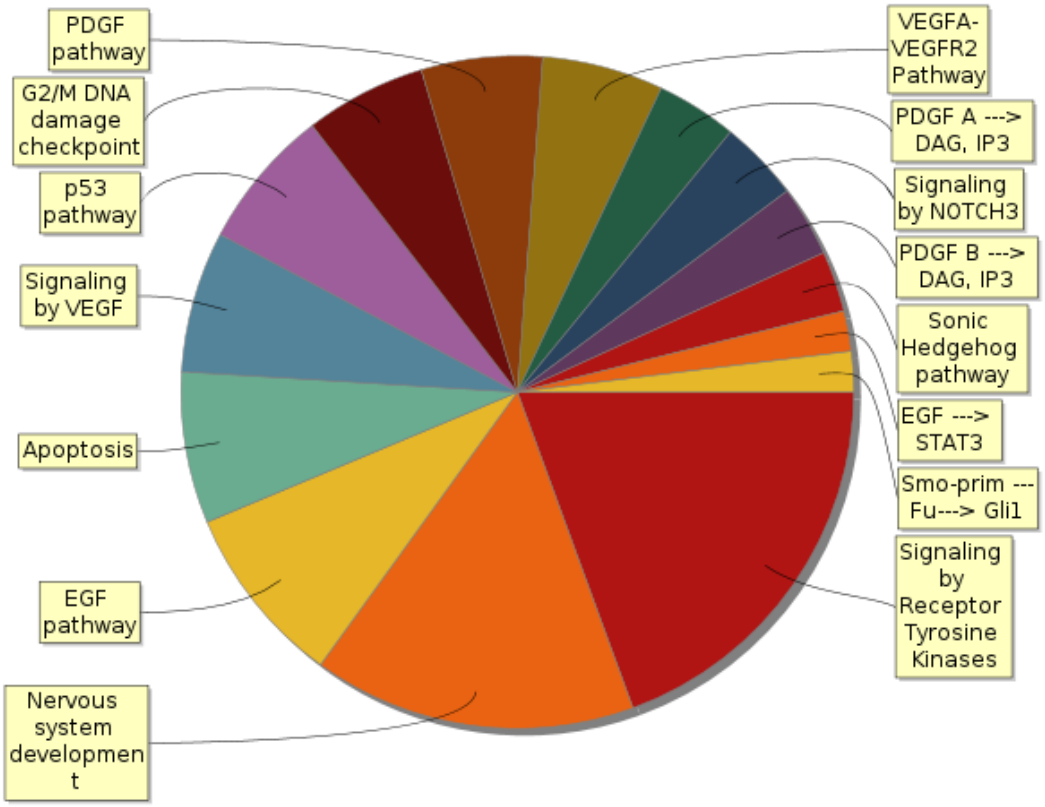


Figure 3. Enriched TRANSPATH® Pathways (2024.2) of the most frequently mutated genes in NCI-H1975.
Full classification →

HumanPSD(TM) disease (2024.2)

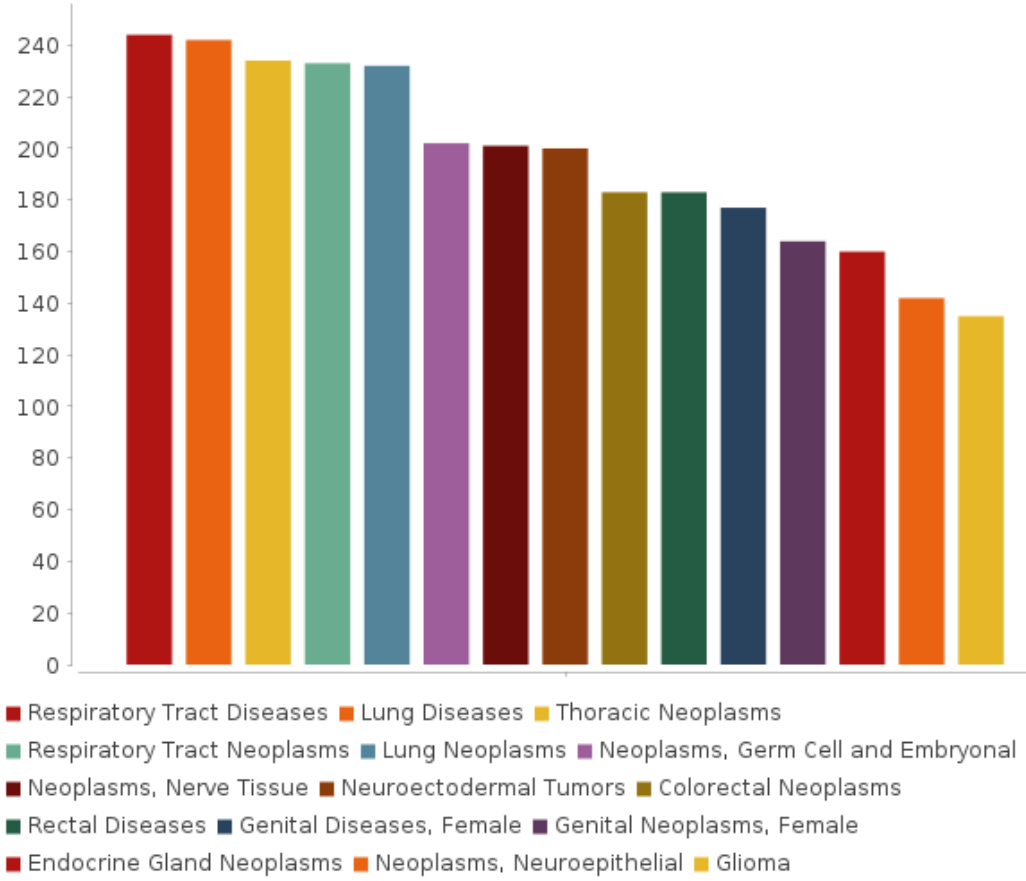
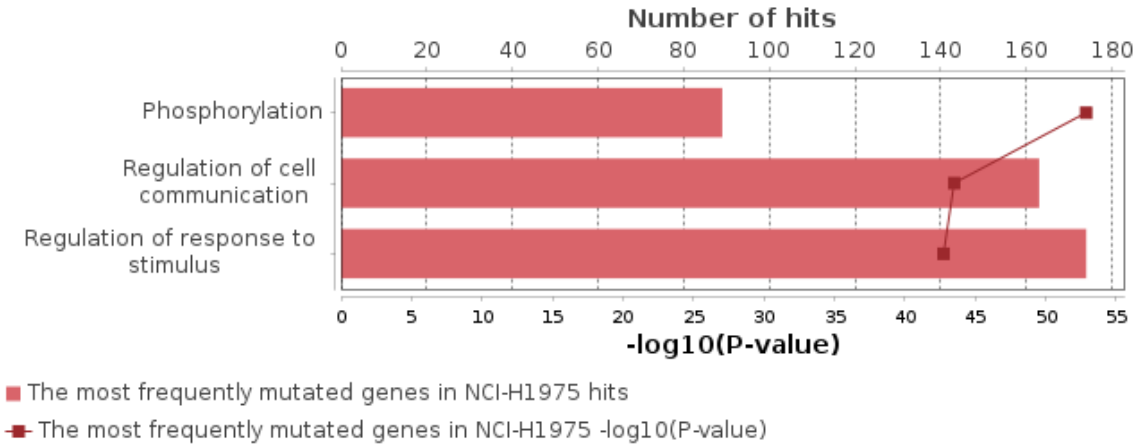


Figure 4. Enriched HumanPSD(TM) disease (2024.2) of the most frequently mutated genes in NCI-H1975. The size of the bars correspond to the number of biomarkers of the given disease found among the input set.

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 7060 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 →

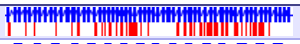
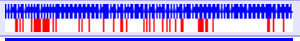
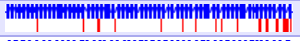
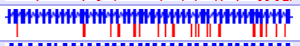
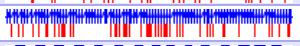
ID	Gene symbol	Gene schematic representation	Number of variations
ENSG00000178104	PDE4DIP		155
ENSG00000034152	MAP2K3		105
ENSG00000154358	OBSCN		82
ENSG00000155657	TTN		79
ENSG00000237298	TTN-AS1		77
ENSG00000197943	PLCG2		66
ENSG00000289733	ENSG00000289733		66
ENSG00000168702	LRP1B		64
ENSG00000008128	CDK11A		62
ENSG00000268575	ENSG00000268575		62

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$ZBTB33_07	4.34E-2	4.45E-8	23	499
V$CGBP_01	1.98E-2	2.45E-18	15	1579
V$CREB_02	8.05E-3	1.28E-8	967	772
V$CREB_Q2	6.15E-3	2.52E-9	973	1063
V$CREB_Q4	6.15E-3	2.52E-9	973	1063
V$CREM_Q6	5.91E-3	3.63E-9	486	976
V$ATF7_01	5.34E-3	2.16E-10	936	1108
V$CREB1_17	3.23E-3	1.88E-7	12	1190
V$CREB_Q3	2.97E-3	1.37E-8	622	701
V$SALL2_01	2.95E-4	5.78E-8	25	74
V$HIF1A_Q5	1.79E-4	7.69E-9	124	298
V$PAX3_05	2.35E-5	6.41E-8	1994	1272
V$ZBTB33_05	1.06E-5	1.06E-7	137	450
V$ELK1_03	4.65E-6	5.15E-10	3372	4616
V$ELK1_04	4.65E-6	1.32E-8	3372	4386
V$NRF1_Q5	1.98E-17	1.29E-2	453	132
V$KLF3_04	1.64E-17	2.91E-3	171	99
V$E2F1_Q6	2.33E-18	6E-3	1709	16
V$E2F2_06	7.72E-20	1.28E-3	621	736
V$E2F4DP2_01	3.83E-20	4.78E-2	2536	4

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 NCI-H1975).

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 NCI-H1975.

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$AP2GAMMA_Q5	V$HIC1_09	V$NMYC_02	V$EP300_08	V$POU5F1_05
2.00; N=3	0.96; N=2	1.00; N=3	0.82; N=3	0.84; N=3

Module width: 150

Module 2:

V$AP2GAMMA_Q5	V$PBX1_04	V$GR_Q6_01	V$ERBETA_Q5	V$OCT4_01
1.00; N=1	0.84; N=3	0.95; N=3	0.86; N=3	0.79; N=3

Module width: 158

Model score (-p*log10(pval)): 31.82
Wilcoxon p-value (pval): 3.23e-64
Penalty (p): 0.501
Average yes-set score: 9.43
Average no-set score: 6.98
AUC: 0.81
Separation point: 7.97
False-positive: 32.22%
False-negative: 18.00%
The AUC of the model achieves value significantly higher than expected for a random set of regulatory regions
Z-score = 4.01

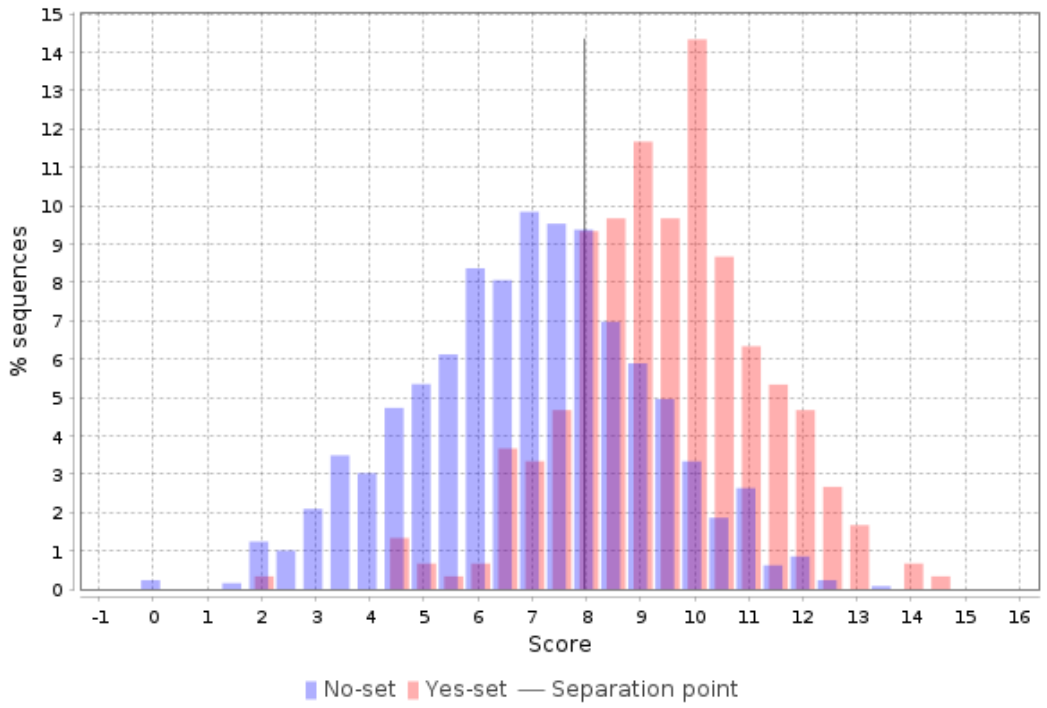


Table 5. List of top ten the most frequently mutated genes in NCI-H1975 with identified enhancers in their regulatory regions. ***CMA score*** *- the score of the CMA model of the enhancer identified in the regulatory region.*

See full table →

Ensembl IDs	Gene symbol	Gene description	CMA score	Factor names
ENSG00000042832	TG	thyroglobulin	14.7	POU5F1(h), HIC1(h), ER-beta(h), GR(h), PBX-1(h), p300(h), N-Myc(h)...
ENSG00000135250	SRPK2	SRSF protein kinase 2	14.41	PBX-1(h), POU5F1(h), N-Myc(h), ER-beta(h), HIC1(h), GR(h), p300(h)...
ENSG00000101464	PIGU	phosphatidylinositol glycan anchor biosynthesis class U	14.26	POU5F1(h), GR(h), PBX-1(h), ER-beta(h), N-Myc(h), HIC1(h), AP-2gamma(h)...
ENSG00000108443	RPS6KB1	ribosomal protein S6 kinase B1	14.14	ER-beta(h), N-Myc(h), p300(h), POU5F1(h), GR(h), PBX-1(h)
ENSG00000267318		novel protein	14.14	ER-beta(h), N-Myc(h), p300(h), POU5F1(h), GR(h), PBX-1(h)
ENSG00000264475		novel transcript, antisense to LAMA1	14.09	POU5F1(h), N-Myc(h), HIC1(h), p300(h), AP-2gamma(h), ER-beta(h), GR(h)...
ENSG00000163545	NUAK2	NUAK family kinase 2	14	p300(h), POU5F1(h), HIC1(h), GR(h), PBX-1(h), N-Myc(h), AP-2gamma(h)
ENSG00000107779	BMPR1A	bone morphogenetic protein receptor type 1A	13.85	HIC1(h), N-Myc(h), PBX-1(h), POU5F1(h), GR(h), AP-2gamma(h), p300(h)...
ENSG00000070371	CLTCL1	clathrin heavy chain like 1	13.81	HIC1(h), p300(h), POU5F1(h), PBX-1(h), N-Myc(h), GR(h), AP-2gamma(h)
ENSG00000286367		novel transcript, antisense to CLTCL1	13.81	HIC1(h), p300(h), POU5F1(h), PBX-1(h), N-Myc(h), GR(h), AP-2gamma(h)

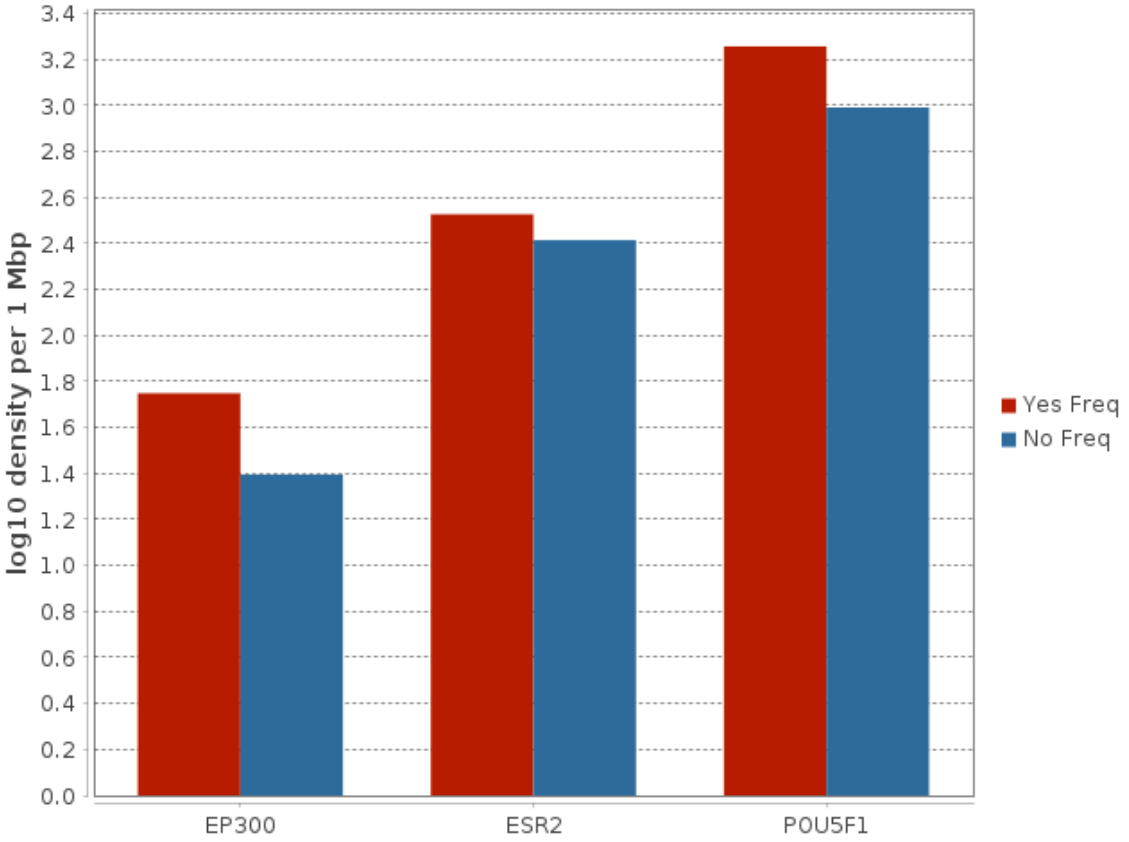
On the basis of the enhancer models we identified transcription factors potentially regulating the ***target genes*** of our interest. We found 8 transcription factors controlling expression of the genes associated with genomic variations (see Table 6).

Table 6. Transcription factors of the predicted enhancer model potentially regulating the genes carrying sequence variations (the most frequently mutated genes in NCI-H1975). ***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	5.42	2.26
MO000059335	ESR2	estrogen receptor 2	4.45	1.3
MO000056618	POU5F1	POU class 5 homeobox 1	4.09	1.84
MO000031266	NR3C1	nuclear receptor subfamily 3 group C member 1	3.8	1.65
MO000026369	MYCN	MYCN proto-oncogene, bHLH transcription factor	3.66	1.45
MO000117395	HIC1	HIC ZBTB transcriptional repressor 1	3.17	1.48
MO000026449	TFAP2C	transcription factor AP-2 gamma	0	1.53
MO000042273	PBX1	PBX homeobox 1	0	4.77

The following diagram represents the key transcription factors, which were predicted to be potentially regulating genes carrying sequence variations in the analyzed pathology: EP300, ESR2 and POU5F1.



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 13 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
MO000139573	Myomegalin(h)	20	NMD_transcript_variant,stop_gained	Cga/Tga,tGg/tAg,tgG/tgA
MO000009403	MKK3(h)	5	NMD_transcript_variant,stop_gained	Cag/Tag
MO000018990	BMP4(h)	1	stop_lost	Tga/Cga
MO000019237	TNF-alpha(h)	1	stop_gained	Cga/Tga
MO000032335	RSK1(h)	1	NMD_transcript_variant,stop_lost	Tga/Cga
MO000032374	raptor(h)	1	stop_gained	CTg/TAg
MO000035011	SRPK1(h)	1	stop_gained	tCa/tGa
MO000036592	TAO1(h)	1	stop_gained	Cga/Tga
MO000059823	PDI(h)	1	NMD_transcript_variant,stop_lost	Tga/Cga
MO000085385	SHARP(h)	1	stop_lost	tgA/tgG

Top 13 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 NCI-H1975. ***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
MO000010977	PDGFRalpha(h)	PDGFRA	platelet derived growth factor receptor alpha	258	207.19
MO000032379	mTOR(h)	MTOR	mechanistic target of rapamycin kinase	333	121.37
MO000005217	flt3(h)	FLT3	fms related receptor tyrosine kinase 3	401	132.46
MO000054152	mTOR(h):rictor(h)	MTOR, RICTOR	RPTOR independent companion of MTOR complex 2, mechanistic target of rapamycin kinase	454	121.37
MO000202967	Tie2(h)	TEK	TEK receptor tyrosine kinase	472	78.36
MO000122463	mTOR(h):rictor(h):mLST8(h):SIN1(h)	MAPKAP1, MLST8, MTOR, RICTOR	MAPK associated protein 1, MTOR associated protein, LST8 homolog, RPTOR independent companion of MTO...	477	121.37
MO000280531	rictor-isoform1(h):mTOR(h):SIN1(h):mLST8(h)	MAPKAP1, MLST8, MTOR, RICTOR	MAPK associated protein 1, MTOR associated protein, LST8 homolog, RPTOR independent companion of MTO...	483	121.37
MO000165201	mTOR(h):rictor(h):mLST8(h):SIN1(h):Protor-1(h)	MAPKAP1, MLST8, MTOR, PRR5, RICTOR	MAPK associated protein 1, MTOR associated protein, LST8 homolog, RPTOR independent companion of MTO...	484	121.37
MO000122429	SIN1(h):rictor(h):mTOR(h)	MAPKAP1, MTOR, RICTOR	MAPK associated protein 1, RPTOR independent companion of MTOR complex 2, mechanistic target of rapa...	487	121.37
MO000112248	PDGFRalpha-isoform1(h)	PDGFRA	platelet derived growth factor receptor alpha	508	207.19

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.

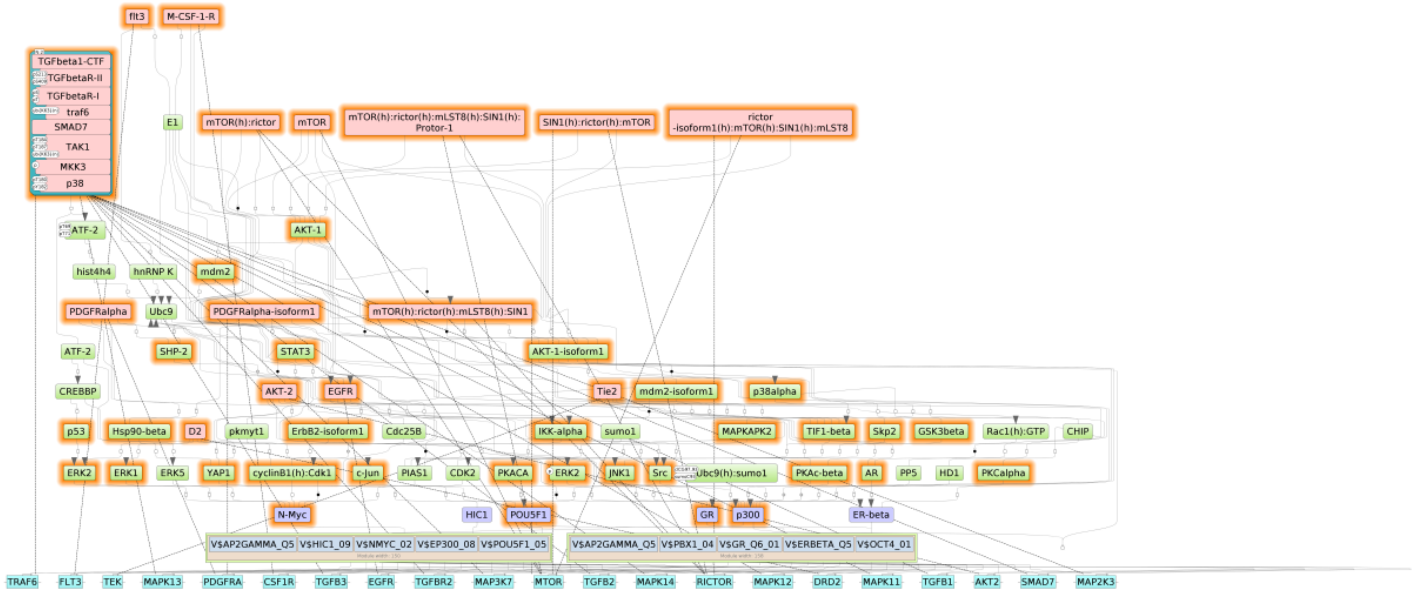


Figure 5. Diagram of intracellular regulatory signal transduction pathways of the most frequently mutated genes in NCI-H1975. Master regulators are indicated by red rectangles, transcription factors are blue rectangles, and green rectangles are intermediate molecules, which have been added to the network 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
PDGFRA	platelet derived growth factor receptor alpha	55	258	207.19
MTOR	mechanistic target of rapamycin kinase	44	333	121.37
FLT3	fms related receptor tyrosine kinase 3	53	401	132.46
RICTOR	RPTOR independent companion of MTOR complex 2	3	454	121.37
TEK	TEK receptor tyrosine kinase	39	472	78.36
INSR	insulin receptor	52	533	208.44

*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.*

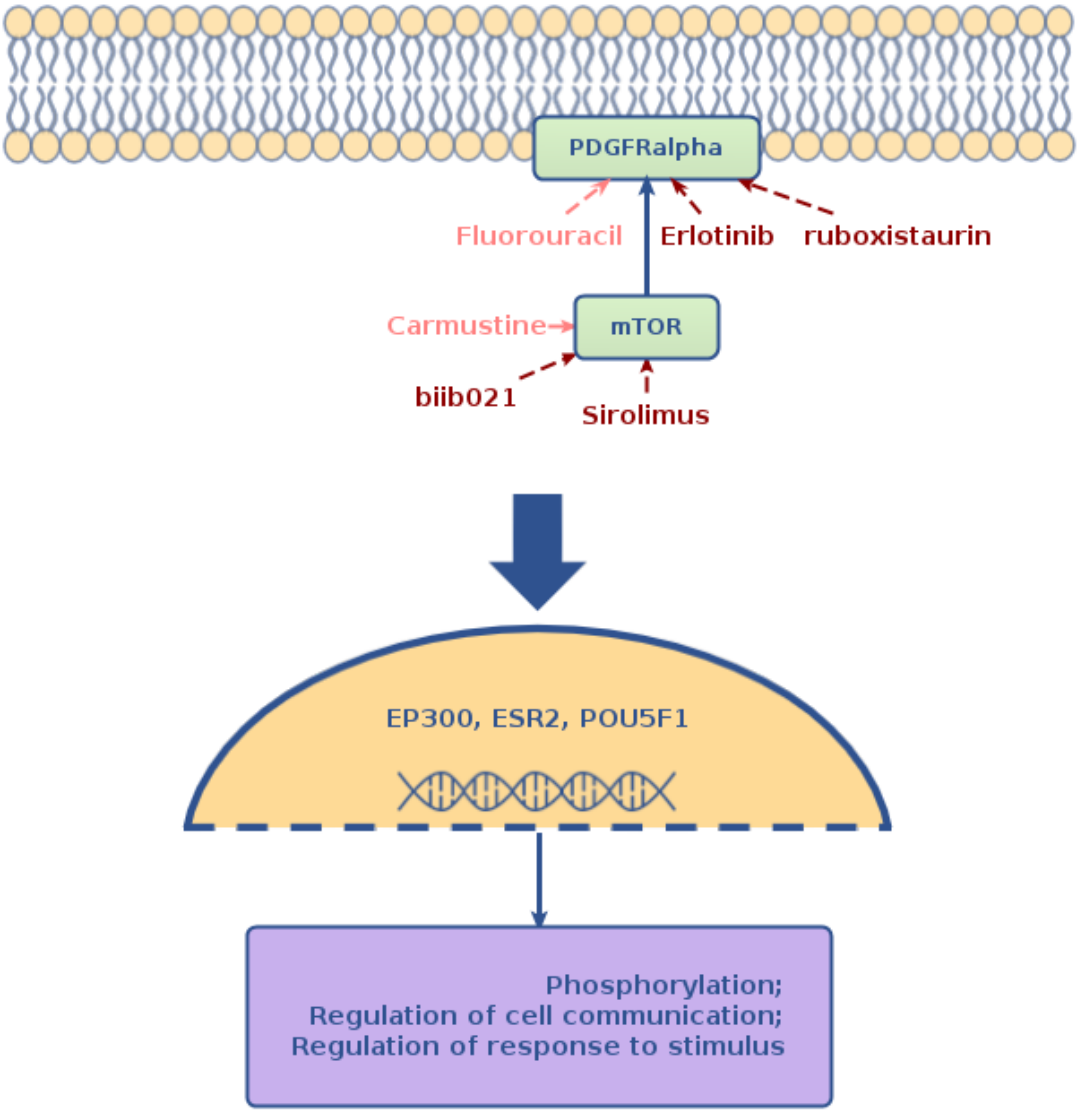
See full table →

Gene symbol	Gene Description	Druggability score	Total rank	Weighted score
PDGFRA	platelet derived growth factor receptor alpha	7.94	258	207.19
MTOR	mechanistic target of rapamycin kinase	5.37	333	121.37
FLT3	fms related receptor tyrosine kinase 3	4.02	401	132.46
TEK	TEK receptor tyrosine kinase	5.19	472	78.36
INSR	insulin receptor	3.7	533	208.44
IRAK2	interleukin 1 receptor associated kinase 2	2.81	543	76.95

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:

- PDGFRalpha
- mTOR

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: Erlotinib, Sirolimus, Carmustine, ruboxistaurin, biib021 and Fluorouracil, 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	Somatic variants	Approved
Erlotinib	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, BIRC5, NUAK2, ERBB3, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, IL2, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, MAP3K1, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1,	100	9	small molecule,approved,investigational	EGFR:T790M:resistance:A1, EGFR:L858R:response:A1	Carcinoma, Non-Small-Cell Lung (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, DailyMed, Pubmed)

	MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, ILK, EGFR, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, ERBB4, PAK3, FER, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2					
Gefitinib	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2,	98	8	small molecule,approved,investigational	EGFR:T790M:resistance:A2, EGFR:L858R:response:A1	Carcinoma, Non-Small-Cell Lung (FDA) Lung Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials)

	CSF1R, STK10, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2					
Crizotinib	MET, NTRK1, ABL1, KDR, MST1R, PRKD1, TEK, IGF1R, RPS6KB1, ALK, EPHB2, FGFR3, PKN2, SRC, NTRK2, TYRO3, PDGFRB, EPHB4, ROS1, BTK, FGR, RET, AKT2, CSF1R, PDPK1	96	7	small molecule,approved		Carcinoma, Non-Small-Cell Lung (FDA)
Osimertinib	MAPK1, ALK, ERBB3, MAPK4, TNK2, BLK, ERBB4, EGFR, AKT3, ERBB2, AKT1, PTK6, MAPK3, AKT2	94	7	small molecule,approved	EGFR:T790M:response:A1	Carcinoma, Non-Small-Cell Lung (FDA) Lung Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials)
capmatinib	MAPK1, ERBB3, GAB1, MDM2, MET, MTOR, MAPK4, PARP1, EGFR, STAT3, AKT3, PTK2,	93	4	small molecule, approved		Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA)

	AKT1, MAPK3, AKT2, TP53					
Trametinib	MAPK1, PARP1, CASP7, MAP2K2, MAP2K7, CASP3, MAP2K1, CASP9, MAPK3	91	6	small molecule,approved		Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA)
dacomitinib	MAPK1, ERBB3, SRC, ERBB4, PARP1, EGFR, AKT3, EPHA6, ERBB2, AKT1, DDR2, MAPK3, AKT2	91	3	small molecule, approved		Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA, PUBMED)
brigatinib	ALK, PARP1, EGFR, EIF2AK3, CASP3, ROS1, ERN1, FLT3, IGF1R	91	5	small molecule, approved		Carcinoma, Non-Small-Cell Lung (FDA)
lorlatinib	ALK, EGFR, PTK2, PTK2B, NTRK1, NTRK2, NTRK3, TNK2, ROS1, FES, FER	90	3	small molecule, approved, investigational		Carcinoma, Non-Small-Cell Lung (FDA, Pubmed)
Alectinib	DRD2, ALK, YAP1, AKT3, AKT1, ROS1, LATS1, RET, AKT2	89	4	small molecule,approved		Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA, Pubmed, Pubmed)
Afatinib	ERBB3, EGFR, ERBB2, ERBB4, PDPK1	88	6	small molecule,approved	EGFR:T790M:resistance:A1, EGFR:L858R:response:A1	Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA)
Paclitaxel	MAPK8, PIK3CG, TOP2A, CDK1, MYB, PIK3CA, CASP3, E2F1, MMP2, BIRC5, MAPK3, BRCA1, TP53	87	9	small molecule,approved		Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA, FDA)
Ceritinib	MAPK1, ALK, MAPK4, ROS1, INSR, PARP1, EGFR, STAT3, AKT3, CASP3, AKT1, BAX, MAPK3, AKT2, IGF1R	87	1	small molecule,approved		Carcinoma, Non-Small-Cell Lung (FDA, Pubmed)
entrectinib	ALK, NTRK1, NTRK2, NTRK3, TNK2, ROS1	84	2	small molecule, approved, investigational		Carcinoma, Non-Small-Cell Lung (FDA, Pubmed)
Etoposide	XIAP, TGFB1, TOP2A, BAX, E2F1, CASP9, TOP1, TOP2B, MYCN	84	7	small molecule,approved		Lung Neoplasms (ClinicalTrials, ClinicalTrials, DailyMed)
Gemcitabine	ERBB2, HRAS, CHEK1, BRCA1	82	9	small molecule,approved		Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA)
repotrectinib	NTRK1, NTRK2, NTRK3, ROS1	81	2	small molecule, approved		Carcinoma, Non-Small-Cell Lung (FDA)

selpercatinib	FGFR3, FGFR2, FLT1, FGFR1, RET, FLT4	80	1	small molecule, approved	Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA)
mobocertinib	EGFR, ERBB2	80	4	small molecule, approved	Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA)
amivantamab	EGFR, MET	76	3	antibody, approved	Carcinoma, Non-Small-Cell Lung (ClinicalTrials, FDA, FDA)
icotinib	EGFR	70	6	small molecule, approved	Carcinoma, Non-Small-Cell Lung (ClinicalTrials, ClinicalTrials, ClinicalTrials) Lung Neoplasms (ClinicalTrials, ClinicalTrials)
Vinorelbine	BAX, BRCA1	59	7	small molecule,approved,investigational	Carcinoma, Non-Small-Cell Lung (ClinicalTrials, ClinicalTrials, FDA) Lung Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, ClinicalTrials, PUBMED)
cemiplimab	PDCD1	39	4	antibody, approved	Lung Neoplasms (DailyMed)
Docetaxel	BAX, HRAS	29	9	small molecule,approved,investigational	Carcinoma, Non-Small-Cell Lung (FDA, FDA) Lung Neoplasms (ClinicalTrials, ClinicalTrials, ClinicalTrials)

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

Drugs approved in clinical trials

Table 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
Sorafenib	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, CHEK2, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, HIPK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, IKBKB, RIPK3, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, EEF2K, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, RPS6KB1, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, MAPK13, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2	99	6	small molecule,approved,investigational
Sirolimus	ROCK2, MARK3, PRKACA, GSK3B, NEK2, SGK1, CAMK1, NEK6, BAX, CSK, MAPKAPK5, PRKD1, MAPK3, MELK, AURKC, MAPK8, MAPKAPK3, CAMKK2, MAP2K1, PRKCZ, MAPK12, MAPK11, MAPKAPK2, CHEK2, AKT2, PDPK1, IKBKB, MAPK10, PRKCA, PRKAA1, RPS6KA1, HSP90AA1, MAPK14, CSNK1D, PGR, FKBP1A, EEF2K, AKT1, CHEK1, RBL2, TP53, RPS6KB1, MAPK1, HIPK2, MAPK9, MTOR, PKN2, PIM3, MAPK13, CAMKK1, PIK3CA, CTLA4	96	4	small molecule,approved,investigational
Vandetanib	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, VEGFA, ACVR2B, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, MST1R, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, CDK1, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2	96	3	small molecule,approved
Nintedanib	FGFR3, SRC, KDR, FGFR2, PDGFRB, YES1, BLK, FGFR1, PDGFRA, FLT4, AKT3, FGR, HCK, CASP3, FLT1, AKT1, LYN, AKT2, FLT3, PDPK1	95	5	small molecule,approved

Pazopanib	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, MAP4K4, MARK3, IRAK3, FGF1, BLK, SLK, ACVRL1, INSR, NEK2, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2	95	3	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
ruboxistaurin	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, PRKCG, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, CHEK2, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, HIPK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, IKBKB, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, EEF2K, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, RPS6KB1, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, MAPK13, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKCB, PRKD2	86	PHASE1: Diabetes Mellitus, Diabetes Mellitus, Type 2, Heart Failure
seliciclib	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, CDK4, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, CHEK2, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, HIPK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, IKBKB, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, CDK1, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, EEF2K, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, RPS6KB1, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, MAPK13, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2	86	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	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, CHEK2, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, HIPK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, IKBKB, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, EEF2K, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, RPS6KB1, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, MAPK13, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2	86	PHASE2: Arthritis, Arthritis, Rheumatoid, Psoriasis
pi-103	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, MAPKAPK3, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7,	86	N/A

	MAP4K1, ROS1, DMPK, STK11, PRKCZ, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, CHEK2, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, HIPK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, IKBKB, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, EEF2K, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, RPS6KB1, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, MAPK13, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2		
Tofacitinib	TEC, BMPR1A, IKBKE, ABL1, FGFR2, PAK2, PRKACA, GSK3B, MAP3K11, SYK, EPHA1, CAMK1, NEK6, ERBB2, LIMK1, MAPK3, MAP2K6, JAK1, MELK, MAPK8, ACVR2B, SRC, CAMKK2, PRKD3, PIP4K2B, CSNK1G2, EPHA7, MAP4K1, ROS1, DMPK, STK11, CSNK1G1, CLK4, MAPK12, PLK3, BTK, WEE1, MAPK11, EPHA5, FLT1, PRKCD, LYN, AKT2, FLT3, MUSK, PDPK1, MAPK10, NTRK1, KDR, PRKCA, CSNK1E, TTK, PDGFRA, RAF1, CDK7, MAPK14, CSNK1D, SRMS, EPHB3, ACVR1, HCK, DDR2, CHEK1, PTK6, NUAK2, RIPK2, CDK8, FGFR3, CDK9, PKN2, RPS6KA2, NTRK2, TYRO3, PIM3, YES1, NUAK1, CAMK2D, ITK, EPHA6, MAP3K4, FGR, PAK1, CLK1, CDK5, PRKACB, RET, ABL2, CSF1R, STK10, ROCK2, MAP4K4, MARK3, IRAK3, BLK, SLK, ACVRL1, INSR, NEK2, SGK1, MAP2K2, MAP2K3, EPHA2, CSK, MAPKAPK5, PRKD1, TEK, PKN1, PIP5K1A, TYK2, IGF1R, AURKC, CAMK4, TNIK, MAP3K5, MAPK4, TNK2, PDGFRB, MAP2K1, MAP4K3, RIPK1, STK4, CAMK2A, PTK2, MAPKAPK2, EPHA8, FES, PLK4, DAPK3, FGFR4, ACVR1B, CAMK2G, MET, NTRK3, PRKAA2, PRKCQ, EPHA4, LATS1, MAP2K4, PRKAA1, RPS6KA1, FLT4, BMPR2, EGFR, PRKCH, ACVR2A, PTK2B, PRKG1, AKT1, AURKA, MAP3K20, RPS6KB1, KIT, MAPK1, ALK, DYRK1B, EPHB2, MAPK9, MERTK, LATS2, EPHA3, EPHB4, PRKCE, FGFR1, PAK3, FER, MAPK13, CAMKK1, EPHB1, AKT3, ZAP70, PIK3CA, TGFBR2, PRKD2	86	EARLY_PHASE1: Vitiligo

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
6,7,12,13-tetrahydro-5H-indolo[2,3-a]pyrrolo[3,4-c]carbazol-5-one	CAMK4, PRKCG, CAMK2G, PRKD3, PRKACA, PRKCZ, CAMK2D, PRKACG, PRKCH, EEF2K, CAMK1, CAMK2A, PRKD1, PRKACB	98	6.33
Camptothecin	HIF1A, TOP2A, CASP3, NFKB1, TOP1, TOP2B	97	3.28
LE-SN38	HIF1A, TOP2A, CASP3, NFKB1, TOP1, TOP2B	97	3.32
BNP 1350	TOP2A, NFKB1, TOP1, TOP2B	95	2.16
Etoposide	NFKB2, CASP8, HIF1A, TOP2A, CASP3, NFKB1, CASP9, TOP1, RELA, TOP2B	93	1.8

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
3-[1-(3-Aminopropyl)-1h-Indol-3-Yl]-4-(1-Methyl-1h-Indol-3-Yl)-1h-Pyrrole-2,5-Dione	CDK6, CAMK2G, GRK2, PRKAA2, PRKCQ, PRKACA, GSK3B, PRKCA, BLK, CDK1, CDK4, CDK7, RPS6KA1, PRKAA1, NEK2, SGK1, PRKCH, EEF2K, CAMK1, MAPKAPK5, PRKD1, PKN1, MAP2K6, CAMK4, LRRK2, GRK6, PRKCG, CDK9, PKN2, RPS6KA2, MAPKAPK3, GRK5, PRKD3, PRKCE, PRKCZ, CAMK2D, PRKACG, CAMK2A, SIRT1, PRKCD, PRKCI, CDK5, GRK3, PRKACB, PRKCB	97	37.83
Rbt205 Inhibitor	CDK6, CAMK2G, GRK2, PRKAA2, PRKCQ, PRKACA, GSK3B, PRKCA, BLK, CDK1, MAP2K4, CDK4, CDK7, RPS6KA1, PRKAA1, SGK1, PRKCH, EEF2K, CAMK1, MAPKAPK5, PRKD1, PKN1, MAP2K6, CAMK4, LRRK2, GRK6, PRKCG, CDK9, RPS6KA2, PKN2, MAPKAPK3, GRK5, PRKD3, PRKCE, PRKCZ, CAMK2D, PRKACG, PTK2, CAMK2A, SIRT1, PRKCD, PRKCI, CDK5, GRK3, PRKACB, PRKCB, DAPK3	96	35.66
3-[1-(3-AMINOPROPYL)-1H-INDOL-3-YL]-4-(1H-INDOL-3-YL)-1H-PYRROLE-2,5-DIONE	PRKACA, GSK3B, BLK, CDK4, SGK1, CAMK1, MAPKAPK5, PRKD1, MAP2K6, PKN1, CAMK4, GRK6, PRKCG, MAPKAPK3, PRKD3, PRKCZ, MAPK12, CAMK2A, PRKCD, CHEK2, DAPK3, CDK6, CAMK2G, GRK2, PRKAA2, PRKCQ, PRKCA, CDK1, MAP2K4, CDK7, RPS6KA1, PRKAA1, PRKCH, EEF2K, PRKG1, CHEK1, RPS6KB1, LRRK2, DYRK1B, CDK9, RPS6KA2, PKN2, GRK5, PRKCE, CAMK2D, PRKACG, SIRT1, PRKCI, CDK5, GRK3, PRKACB, PRKCB	96	40.93
Topotecan	HIF1A, TOP2A, CASP3, NFKB1, TOP1, TOP2B	96	3.08
7-[4-(Dimethylamino)Phenyl]-N-Hydroxy-4,6-Dimethyl-7-Oxo-2,4-Heptadienamide	HDAC4, HDAC2, HDAC6, HDAC3	96	2.62

As the result of drug search we propose the following drugs as most promising candidates for treating the pathology under study: Erlotinib, ruboxistaurin, 6,7,12,13-tetrahydro-5H-indolo[2,3-a]pyrrolo[3,4-c]carbazol-5-one and 3-[1-(3-Aminopropyl)-1h-Indol-3-Yl]-4-(1-Methyl-1h-Indol-3-Yl)-1h-Pyrrole-2,5-Dione. These drugs were selected for acting on the following targets: PDGFRA, PRKD1 and LRRK2, 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	73	
Abiraterone	Prostatic Neoplasms, Castration-Resistant	-	
Abiraterone acetate	Prostatic Neoplasms, Castration-Resistant	-	
Acalabrutinib	Lymphoma, Mantle-Cell	62	
Acitretin	Psoriasis	64	
Ado-trastuzumab emtansine	Breast Neoplasms|Neoplasms	84	
Afatinib	Carcinoma, Non-Small-Cell Lung	88	EGFR:T790M:resistance:A1, EGFR:L858R:response:A1

Aflibercept	Colorectal Neoplasms|Diabetic Retinopathy|Edema|Vascular Diseases|Wet Macular Degeneration	48
Alectinib	Carcinoma, Non-Small-Cell Lung	89
Alemtuzumab	Brain Abscess|Leukemia, Lymphocytic, Chronic, B-Cell|Multiple Sclerosis|Multiple Sclerosis, Relapsing-Remitting|Sclerosis	-
Alitretinoin	Sarcoma, Kaposi	42
Alpelisib	Breast Neoplasms	37
Altretamine	Ovarian Neoplasms	-
Aminolevulinic acid	Keratosis|Keratosis, Actinic	-
Anagrelide	Thrombocythemia, Essential|Thrombocytosis	-
Anastrozole	Breast Neoplasms|Hypersensitivity|Obesity|Obesity, Morbid|Recurrence|Weight Loss	16
Apalutamide	Prostatic Neoplasms, Castration-Resistant	21
Aprepitant	Nausea|Neoplasms|Postoperative Nausea and Vomiting	-
Arsenic trioxide	Leukemia, Promyelocytic, Acute	80
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	93
Azacitidine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes|Preleukemia|Syndrome	75
Belinostat	Lymphoma, T-Cell, Peripheral	67
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	41
Bexarotene	Lymphoma, T-Cell|Lymphoma, T-Cell, Cutaneous	63
Bicalutamide	Prostatic Neoplasms	62
Binimetinib	Melanoma	68
Blinatumomab	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-
Bortezomib	Brain Abscess|Glomerulonephritis|Glomerulonephritis, IGA|Kidney Diseases|Multiple Myeloma|Neoplasms, Plasma Cell|Nephritis|Renal Insufficiency	50
Bosutinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive	85
Brentuximab vedotin	Hodgkin Disease|Lymphoma|Lymphoma, Large-Cell, Anaplastic|Lymphoma, T-Cell, Peripheral	-
Brigatinib	Carcinoma, Non-Small-Cell Lung	91
Buserelin	Prostatic Neoplasms	-
Cabazitaxel	Prostatic Neoplasms, Castration-Resistant	74
Cabergoline	Drug-Related Side Effects and Adverse Reactions|Pituitary Neoplasms	27
Cabozantinib	Thyroid Neoplasms	94
Capecitabine	Breast Neoplasms|Colonic Neoplasms|Colorectal Neoplasms	32
Carboplatin	Carcinoma, Non-Small-Cell Lung|Lung Neoplasms|Neoplasms|Neuroendocrine Tumors|Ovarian Neoplasms|Retinoblastoma	-
Carfilzomib	Multiple Myeloma	85
Carmustine	Astrocytoma|Glioblastoma|Hodgkin Disease|Medulloblastoma|Multiple Myeloma|Neoplasms	10
Ceritinib	Carcinoma, Non-Small-Cell Lung	87
Cetuximab	Colorectal Neoplasms	46
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	81	ERCC2:K751Q:resistance:B2
Cladribine	Leukemia, Hairy Cell	62	
Clofarabine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	66	
Cobimetinib	Melanoma	75	
Copanlisib	Lymphoma, Follicular	85	
Crizotinib	Carcinoma, Non-Small-Cell Lung	96	
Cyproterone acetate	Prostatic Neoplasms	48	
Dabrafenib	Melanoma	77	
Dacomitinib	Carcinoma, Non-Small-Cell Lung	91	
Daratumumab	Multiple Myeloma	-	
Dasatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase|Precursor Cell Lymphoblastic Leukemia-Lymphoma	94	
Decitabine	Anemia, Refractory|Anemia, Refractory, with Excess of Blasts|Leukemia, Myelomonocytic, Chronic|Myelodysplastic Syndromes	64	
Degarelix	Cardiovascular Diseases|Prostatic Neoplasms|Vascular Diseases	72	
Denosumab	Arthritis, Rheumatoid|Bone Diseases|Bone Diseases, Metabolic|Breast Neoplasms|Hyperparathyroidism|Hyperparathyroidism, Primary|Metabolic Diseases|Neoplasm Metastasis|Neoplasms|Osteoporosis|Osteoporosis, Postmenopausal|Prostatic Neoplasms	42	
Dexrazoxane	Breast Neoplasms|Cardiomyopathies	61	
Dienogest	Menorrhagia	60	
Dinutuximab	Neuroblastoma	-	
Docetaxel	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Prostatic Neoplasms|Squamous Cell Carcinoma of Head and Neck|Stomach Neoplasms	29	
Doxorubicin	Neoplasms|Multiple Myeloma|Carcinoma, Ovarian Epithelial|Ovarian Neoplasms|Leukemia, Lymphoid|Breast Neoplasms|Lymphoma, Follicular|Thyroid Neoplasms|Triple Negative Breast Neoplasms|Glioma	89	
Durvalumab	Carcinoma, Non-Small-Cell Lung|Carcinoma, Transitional Cell	-	
Dutasteride	Alcoholism|Hyperplasia|Hypertrophy|Neoplasms|Prostatic Hyperplasia	-	
Duvelisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	73	
Elotuzumab	Multiple Myeloma	68	
Enasidenib	Leukemia, Myeloid, Acute	-	
Encorafenib	Colorectal Neoplasms|Melanoma	85	
Enfortumab vedotin	Carcinoma, Transitional Cell|Neoplasms	-	
Entrectinib	Carcinoma, Non-Small-Cell Lung	84	
Enzalutamide	Prostatic Neoplasms|Prostatic Neoplasms, Castration-Resistant	21	
Epirubicin	Breast Neoplasms	66	
Erdafitinib	Urinary Bladder Neoplasms	87	
Eribulin	Breast Neoplasms|Drug-Related Side Effects and Adverse Reactions|Neoplasms	-	
Erlotinib	Carcinoma, Non-Small-Cell Lung|Neoplasms|Pancreatic Neoplasms	100	EGFR:T790M:resistance:A1, EGFR:L858R:response:A1
Erlotinib hydrochloride	Carcinoma, Non-Small-Cell Lung|Gastrointestinal Stromal Tumors	-	
Estramustine	Prostatic Neoplasms	21	
Ethinyl Estradiol	Acne Vulgaris|Neoplasms	30	
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	92	
Exemestane	Breast Neoplasms	-	
Fedratinib	Primary Myelofibrosis	66	

Finasteride	Hyperplasia|Neoplasms|Prostatic Hyperplasia	14	
Flavopiridol	Leukemia, Lymphocytic, Chronic, B-Cell	92	
Fluorouracil	Skin Neoplasms|Neoplasms, Basal Cell|Neoplasms, Second Primary|Neoplasms, Squamous Cell|Neoplasms|Colorectal Neoplasms|Pancreatic Neoplasms	74	
Fluoxymesterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	48	
Flutamide	Premenstrual Dysphoric Disorder|Premenstrual Syndrome|Prostatic Neoplasms	14	
Fulvestrant	Breast Neoplasms	67	
Gefitinib	Carcinoma, Non-Small-Cell Lung	98	EGFR:T790M:resistance:A2, EGFR:L858R:response:A1
Gemcitabine	Breast Neoplasms|Carcinoma, Non-Small-Cell Lung|Ovarian Neoplasms|Pancreatic Neoplasms	82	
Gemtuzumab ozogamicin	Leukemia, Myeloid, Acute	1	
Gilteritinib	Leukemia, Myeloid, Acute	77	
Glasdegib	Leukemia, Myeloid, Acute	39	
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	76	
Ibritumomab	Lymphoma, B-Cell|Lymphoma, Follicular	1	
Ibrutinib	Graft vs Host Disease|Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, B-Cell, Marginal Zone|Lymphoma, Mantle-Cell|Waldenstrom Macroglobulinemia	82	
Idarubicin	Leukemia, Myeloid, Acute	59	
Idelalisib	Leukemia, Lymphocytic, Chronic, B-Cell|Lymphoma, Follicular	78	
Ifosfamide	Neoplasms	39	
Imatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Mastocytosis, Systemic|Neoplasms	85	
Inotuzumab ozogamicin	Precursor B-Cell Lymphoblastic Leukemia-Lymphoma	-	
Ipilimumab	Carcinoma, Renal Cell|Melanoma	29	
Irinotecan	Colorectal Neoplasms	61	
Ivosidenib	Leukemia, Myeloid, Acute	-	
Ixabepilone	Breast Neoplasms	-	
Ixazomib	Multiple Myeloma	-	
Lapatinib	Breast Neoplasms	91	
Larotrectinib	Neoplasm Metastasis	78	
Lenalidomide	Brain Abscess|Lupus Erythematosus, Cutaneous|Myelodysplastic Syndromes|Neoplasms, Plasma Cell	41	
Lenvatinib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	86	
Letrozole	Breast Neoplasms|Cysts|Fibroma|Myofibroma|Myoma|Ovarian Cysts|Syndrome	27	
Leuprolide	Hot Flashes|Ovarian Hyperstimulation Syndrome|Prostatic Neoplasms|Puberty, Precocious	-	
Levamisole	Ascariasis|Colonic Neoplasms|Helminthiasis	-	
Levonorgestrel	Epilepsy|Hyperplasia|Menorrhagia	61	
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	21	
Lorlatinib	Carcinoma, Non-Small-Cell Lung	90	
Masoprocol	Keratosis, Actinic	53	
Medroxyprogesterone Acetate	Depression|Depression, Postpartum|Depressive Disorder|Metrorrhagia|Neoplasms|Uterine Hemorrhage	61	
Megestrol acetate	Acquired Immunodeficiency Syndrome|Bites and Stings|Breast Neoplasms|Pain|Wasting Syndrome	57	
Methotrexate	Neoplasms|Breast Neoplasms|Head and Neck Neoplasms|Ovarian Neoplasms|Lymphoma, T-Cell, Peripheral|Brain	57	

	Neoplasms|Colorectal Neoplasms|Neuroblastoma|Carcinoma, Squamous Cell		
Methyltestosterone	Breast Neoplasms|Hypogonadism|Puberty, Delayed	24	
Midostaurin	Leukemia, Mast-Cell|Leukemia, Myeloid, Acute|Mastocytosis, Systemic	86	
Mitotane	Adrenocortical Carcinoma	10	
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	69	
Mogamulizumab	Mycosis Fungoides|Neoplasms|Sezary Syndrome	-	
Moxetumomab pasudotox	Leukemia, Hairy Cell|Neoplasms	-	
Necitumumab	Carcinoma, Non-Small-Cell Lung|Neoplasms	-	
Nelarabine	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma	-	
Neratinib	Breast Neoplasms	89	
Nilotinib	Blast Crisis|Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Leukemia, Myeloid, Chronic-Phase	78	
Nilutamide	Prostatic Neoplasms	21	
Nintedanib	Fibrosis|Idiopathic Pulmonary Fibrosis	95	
Niraparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms	50	
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	48	
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	94	EGFR:T790M:response:A1
Oxaliplatin	Colonic Neoplasms|Colorectal Neoplasms|Neoplasms|Rectal Neoplasms	80	
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	87	
Palbociclib	Breast Neoplasms	70	
Panitumumab	Colorectal Neoplasms	66	
Panobinostat	Multiple Myeloma	61	
Pazopanib	Carcinoma|Carcinoma, Renal Cell|Sarcoma	95	
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	17	
Pertuzumab	Breast Neoplasms	83	
Pomalidomide	Multiple Myeloma	28	
Ponatinib	Leukemia, Myelogenous, Chronic, BCR-ABL Positive|Precursor Cell Lymphoblastic Leukemia-Lymphoma	84	
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	93	
Relugolix	Prostatic Neoplasms	-	
Ribociclib	Breast Neoplasms	67	

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	10
Romidepsin	Lymphoma, T-Cell, Cutaneous	80
Rucaparib	Carcinoma, Ovarian Epithelial|Fallopian Tube Neoplasms|Peritoneal Neoplasms|Prostatic Neoplasms, Castration-Resistant	71
Ruxolitinib	Graft vs Host Disease|Polycythemia|Polycythemia Vera|Primary Myelofibrosis|Thrombocytosis	69
Selinexor	Multiple Myeloma	21
Selumetinib	Neurofibromatosis 1	82
Siltuximab	Giant Lymph Node Hyperplasia	-
Sirolimus	Angiomyolipoma|Constriction, Pathologic|Coronary Restenosis|Eye Diseases|Immune System Diseases|Kidney Failure, Chronic|Lipoma|Tuberous Sclerosis	96
Sonidegib	Carcinoma, Basal Cell	30
Sorafenib	Carcinoma, Hepatocellular|Carcinoma, Renal Cell|Thyroid Neoplasms	99
Sunitinib	Adenoma|Carcinoma, Renal Cell|Digestive System Neoplasms|Gastrointestinal Neoplasms|Gastrointestinal Stromal Tumors|Intestinal Neoplasms	94
Talazoparib	Breast Neoplasms	43
Tamoxifen	Breast Diseases|Cystic Fibrosis|Cysts|Fibroadenoma|Fibrocystic Breast Disease|Hemorrhage|Menorrhagia|Menstruation Disturbances|Metrorrhagia|Neoplasms	78
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	-
Temsirolimus	Carcinoma, Renal Cell	88
Teniposide	Precursor Cell Lymphoblastic Leukemia-Lymphoma	69
Thalidomide	Brain Abscess|Immune System Diseases|Multiple Myeloma|Neoplasms, Plasma Cell	61
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	57
Toremifene	Breast Neoplasms	66
Trabectedin	Leiomyosarcoma|Liposarcoma	-
Trametinib	Carcinoma, Non-Small-Cell Lung|Melanoma	91
Trastuzumab	Breast Neoplasms|Neoplasms	83
Tretinoin	Lentigo	75
Triptorelin	Fatty Liver|Hypogonadism|Infertility, Female|Prostatic Neoplasms	76
Tucatinib	Breast Neoplasms	82
Valrubicin	Urinary Bladder Neoplasms	63
Vandetanib	Thyroid Neoplasms	96
Vemurafenib	Melanoma	76
Venetoclax	Leukemia, Lymphocytic, Chronic, B-Cell|Leukemia, Myeloid, Acute	-
Vinblastine	Glioma	25
Vincristine	Precursor Cell Lymphoblastic Leukemia-Lymphoma	50
Vinorelbine	Carcinoma, Non-Small-Cell Lung	59
Vismodegib	Carcinoma, Basal Cell	35
Vorinostat	Lymphoma, T-Cell, Cutaneous	82
Zoledronate	Arthritis|Bone Marrow Diseases|Brain Abscess|Chronic Kidney Disease-Mineral and Bone Disorder|Chronic Periodontitis|HIV Infections|Hypersensitivity|Infections|Kidney Diseases|Metabolic Diseases|Multiple Myeloma|Neoplasms|Neoplasms, Plasma	-

Cell|Neoplasms, Second Primary|Osteitis|Osteoarthritis|Periodontitis|Pleural Effusion, Malignant|Prostatic Neoplasms|Renal Insufficiency, Chronic|Thalassemia|Wounds and Injuries

6. Conclusion

We applied the software package "Genome Enhancer" to a data set that contains *genomics* data. The study is done in the context of *Non-Small-Cell Lung Carcinoma and Lung 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, ruboxistaurin, 6,7,12,13-tetrahydro-5H-indolo[2,3-a]pyrrolo[3,4-c]carbazol-5-one and 3-[1-(3-Aminopropyl)-1h-Indol-3-Yl]-4-(1-Methyl-1h-Indol-3-Yl)-1h-Pyrrole-2,5-Dione

These drugs were selected for acting on the following targets: PDGFRA, PRKD1 and LRRK2, 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:

PDGFRalpha and mTOR

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: Erlotinib, Sirolimus, Carmustine, ruboxistaurin, biib021 and Fluorouracil. 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.

- PDGFRalpha
- mTOR

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 NCI-H1975).
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.