A Web Tool for Ranking Candidate Drugs Against a Selected Disease Based on a Combination of Functional and Structural Criteria

Evangelos Karatzas [a,b], George Minadakis [b,d], George Kolios [c], Alex Delis [a], George M. Spyrou [b,d,*]

[a] Department of Informatics and Telecommunications, University of Athens, 15703 Athens, Greece
[b] The Cyprus Institute of Neurology and Genetics, 6 International Airport Avenue, Nicosia, 2370, Cyprus
[c] Laboratory of Pharmacology, Department of Medicine, Democritus University of Thrace, Greece
[d] The Cyprus School of Molecular Medicine, 6 International Airport Avenue, Nicosia 2370, Cyprus

A R T I C L E   I N F O

A B S T R A C T

Drug repurposing techniques allow existing drugs to be tested against diseases outside their initial spectrum, resulting in reduced cost and eliminating the long time-frames of new drug development. In silico drug repurposing further speeds up the process either by proposing drugs suitable to invert the transcriptomic profile of a disease or by indicating drugs based on their common targets or structural similarity with other drugs with similar mode of action.

Such methods usually return a number of potential repurposed drugs that need to be tested against the disease in in vitro, pre-clinical and clinical studies. Thus, it is crucial to have a more sophisticated candidate drug ranking in order to start testing from the most promising chemical substances. As a means to enhance the above decision process, we present CoDReS (Composite Drug Reranking Scoring), a drug (re-)ranking web-based tool, which combines an initial drug ranking (i.e. repurposing score or hypothesis/potentiality score) with a functional score of each drug considered in conjunction with the disease under study as well as with a structural score derived from potential drugability violations. Furthermore, a structural similarity clustering is applied on the considered drugs and a handful of structural exemplars are suggested for further in vitro and in vivo validation. The user is able to filter the results further, through structural similarity examination of the candidate drugs with drugs that have failed against the queried disease where related clinical trials have been carried out.

CoDReS is publicly available online at http://bioinformatics.cing.ac.cy/codres.

1. Introduction

Transcriptomic-based computational drug repurposing (DR) tools, such as Connectivity Map [1] and L1000CDS² [2], compare a disease-related gene expression profile with a number of stored existing expression profiles corresponding to cellular responses against a number of perturbations. Existing tools return lists of candidate repurposed drugs, which can be ordered by their inhibition score. The inhibition score describes the potentiality of a chemical substance to alter the perturbed gene signature state of a disease back to its "normal-healthy" values. Although the inhibition score might give insight onto the potency of a drug against a disease, it alone cannot guarantee success in a clinical trial. On the other hand, cheminformatics tools, such as ChemMine Tools [3] and programming packages such as Rcpi [4] and ChemmineR [5] can suggest drugs with similar structure and possibly similar mode of action to drugs with a-priori knowledge regarding their effectiveness either against a specific disease-related mechanism or against diseases with phenotypic similarity to the targeted disease. However, the derived similarity score is often not enough to deem a drug an appropriate candidate against a disease.

Other types of drug information are ought to be examined, like the candidate drug's functional relation to the disease and its binding affinity to any related-to-the-disease gene target as well as its drug-likeness evaluation based on structural rules that might categorize the drug inappropriate for clinical trials. In order to attain both the scoring implementation for these different drug aspects and provide a more meaningful ranking of the candidate repurposed drugs, we have developed the CoDReS (Composite Drug Reranking Score) web-based tool based on- and extending the initial methodology introduced in [6] in the following ways; CoDReS integrates information from updated biological databases, incorporates binding

Fig. 1. CoDReS summary figure.

queried disease in clinical trials. A summary figure of the CoDReS pipeline is depicted in Fig. 1.

2. Tool Description

2.1. Scoring Scheme

A composite score (from here on referred to as CoDReS) is calculated, for each drug, as the normalized weighted sum of the initial a-priori score (aS) with a functional (FS) and a structural score (StS) as introduced below:

$$CoDReS_i = \frac{w_{aS} * aS_i + w_{FS} * FS_i + w_{StS} * StS_i}{\max CoDReS}, i = 1, \ldots, N \ drugs$$

The weights w_{aS}, w_{FS} and w_{StS} are user-defined parameters that determine the desired influence of each part (a-priori, functional and structural scores respectively) to the final score and have equal default values. The a-priori scores can be uploaded by the user and are
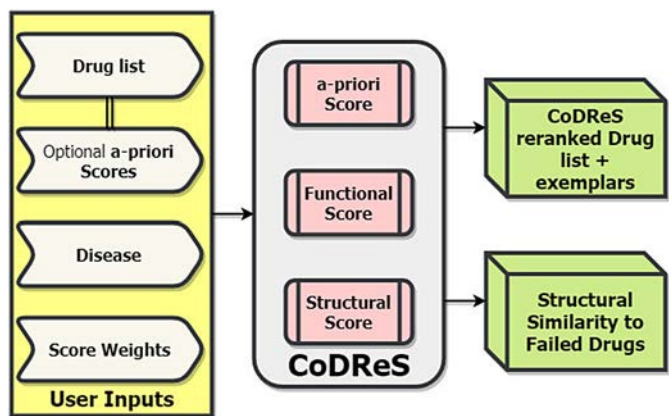
affinity scores between ligands and proteins, evaluates drug-likeness and presents structural similarities between input drugs and possible failed drugs that have already been tested against the

Table 1
Information regarding resources integrated to CoDReS.

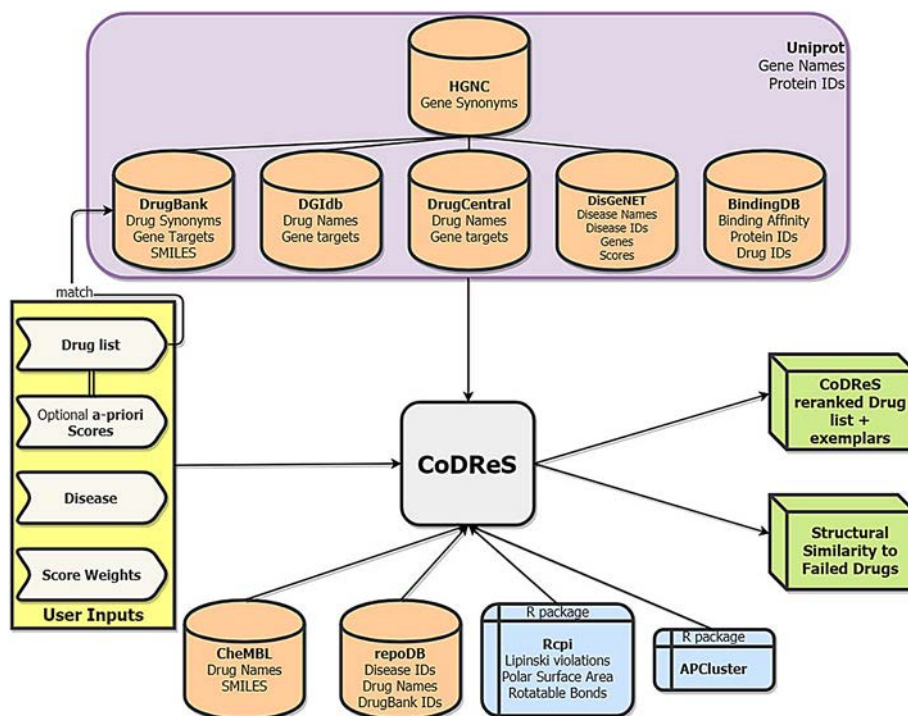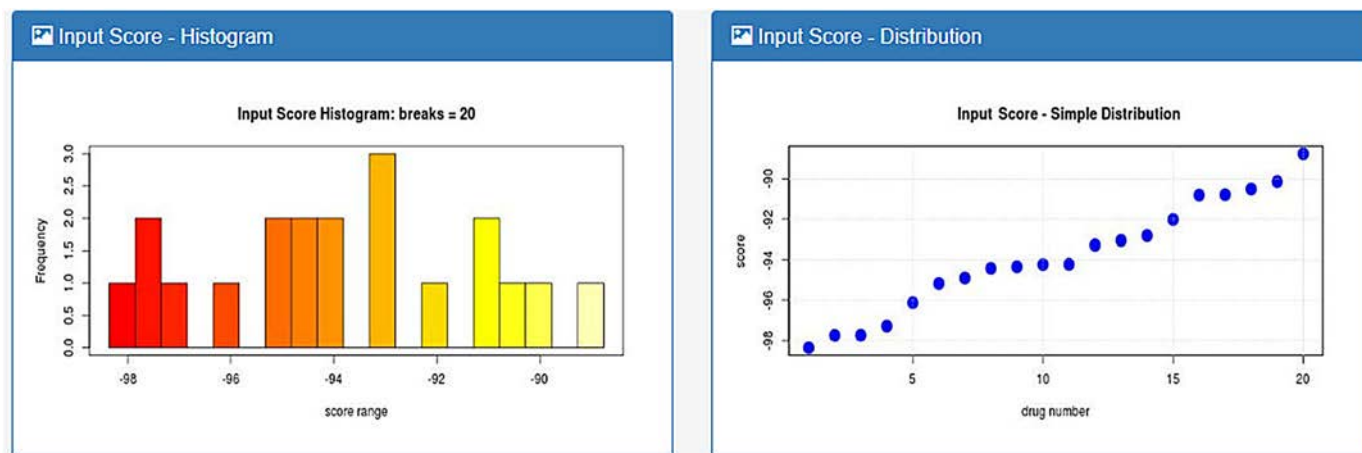| Database Name	Link	File	Current Version	Last Update
BindingDB	https://www.bindingdb.org/bind/chemsearch/marvin/SDFdownload.jsp?all_download = yes	BindingDB_All_2019m1.tsv.zip	–	2019/02
CheMBL	ftp://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/latest/	chembl_24_1_mysql.tar.gz	24.1	2018/06
DGIdb	http://www.dgidb.org/downloads	Interactions TSV	3.0.2	2018/01
DisGeNET	http://www.disgenet.org/downloads	ALL gene-disease associations	6.0	2019/02
DrugBank	https://www.drugbank.ca/releases/latest#open-data	DrugBank Vocabulary	5.1.2	2018/12
	https://www.drugbank.ca/releases/latest#protein-identifiers	Drug Target Identifiers - All	5.1.2	2018/12
	https://www.drugbank.ca/releases/latest#structures	Structural External Links - All	5.1.2	2018/12
DrugCentral	http://drugcentral.org/download	Drug-target interaction	10.4	2018/08
HGNC	https://www.genenames.org/cgi-bin/download	Approved Symbol, Synonyms	–	2019/02
repoDB	http://apps.chiragjpgroup.org/repoDB/	full repoDB dataset	1.2	2017/07
Uniprot	ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/by_organism	HUMAN_9606_idmapping.dat	–	2019/02

Fig. 2. CoDReS integration scheme.

Fig. 3. Input score diagrams are drawn after the user uploads a drug list with their respective scores as returned by any drug repurposing tool.

automatically normalized in the unit interval [0, 1] by dividing with the absolute maximum a-priori score.

The functional score requires the calculation of two different parameters:

(i) the Confidence Score, which reflects the gene-disease association and

(ii) the Ki, which is an inhibitory constant, measured in nM, and represents the reciprocal of the binding affinity between the inhibitor (drug) and the enzyme (target) [7]. The smaller the Ki, the greater the binding affinity.

The FS for each drug is calculated as the sum of the products of Confidence Score with the inverse value of Ki, for each gene target of the drug that has been related to the queried disease. Each drug's FS is finally normalized in [0, 1] by dividing with the maximum FS.

$$FS = \frac{\sum_{j=1}^{nGenes} ConfidenceScore_j \cdot \frac{1}{Ki_j}}{maxFS}$$

The structural score calculates a substance's drug-likeness based on the Lipinski "rules of 5" [8] and Veber's rule [9]. According to the Lipinski rules, in order for a drug to be orally active in humans, it should conform to the following rules: (i) have ≤5 hydrogen bond donors, (ii) have ≤10 hydrogen bond acceptors, (iii) weigh <500 Da and (iv) have an octanol-water partition coefficient (log P) ≤5. The Veber's rule further requires that the chemical substance (v) contains ≤ 10 rotatable bonds and (vi) its polar surface area does not exceed 140 Å^2 (angstrom^2). The final StS for each drug is a value within the range [0, 1] calculated in the following way:

$$StS = 1 - \frac{numViolations}{6}$$

where "6" is the maximum number of structural rules that a drug might violate.

2.2. Development

The static components of the user interface (UI) of the CoDReS web-based application are developed in php, html, css (bootstrap) and javascript (ajax), while the dynamic components of the UI are refreshed via php and back-end R scripts. Several data-repositories have been downloaded, parsed and integrated into a MySQL database, which in turn serves the CoDReS web-based application. Information regarding the database releases, versions and links can be found on Table 1.

CoDReS works with drug synonyms that have been downloaded from DrugBank [10]. DrugBank is a drug-centric online database that provides detailed information on drugs and their gene targets. The

A/A	Initial Position	Input Drug	DrugBank Name	DrugBank ID	Input Score	Functional Score	Structural Score	CoDReS
1	2	cefotaxime	✔ cefotaxime	DB00493	0.993798	1	0.833333	1.000000
2	15	vorinostat	✔ vorinostat	DB02546	0.935543	0.566667	0.833333	0.826118
3	19	daunorubicin	✔ daunorubicin	DB00694	0.916429	0.433333	0.666667	0.713242
4	11	palonosetron	palonosetron	DB00377	0.95791	0	1	0.692543
5	7	pidorubicine	epirubicin	DB00445	0.964823	0.433333	0.5	0.671407
6	5	pirarubicin	pirarubicin	DB11616	0.977227	0	0.666667	0.581471
7	18	dactinomycin	dactinomycin	DB00970	0.920191	0.2	0.5	0.573087
8	1	ly-303511	ly-303511	unassigned	1	0	0	0.353715
9	3	cgp-60474	cgp-60474	unassigned	0.993697	0	0	0.351486
10	4	pik-75	pik-75	unassigned	0.989122	0	0	0.349868

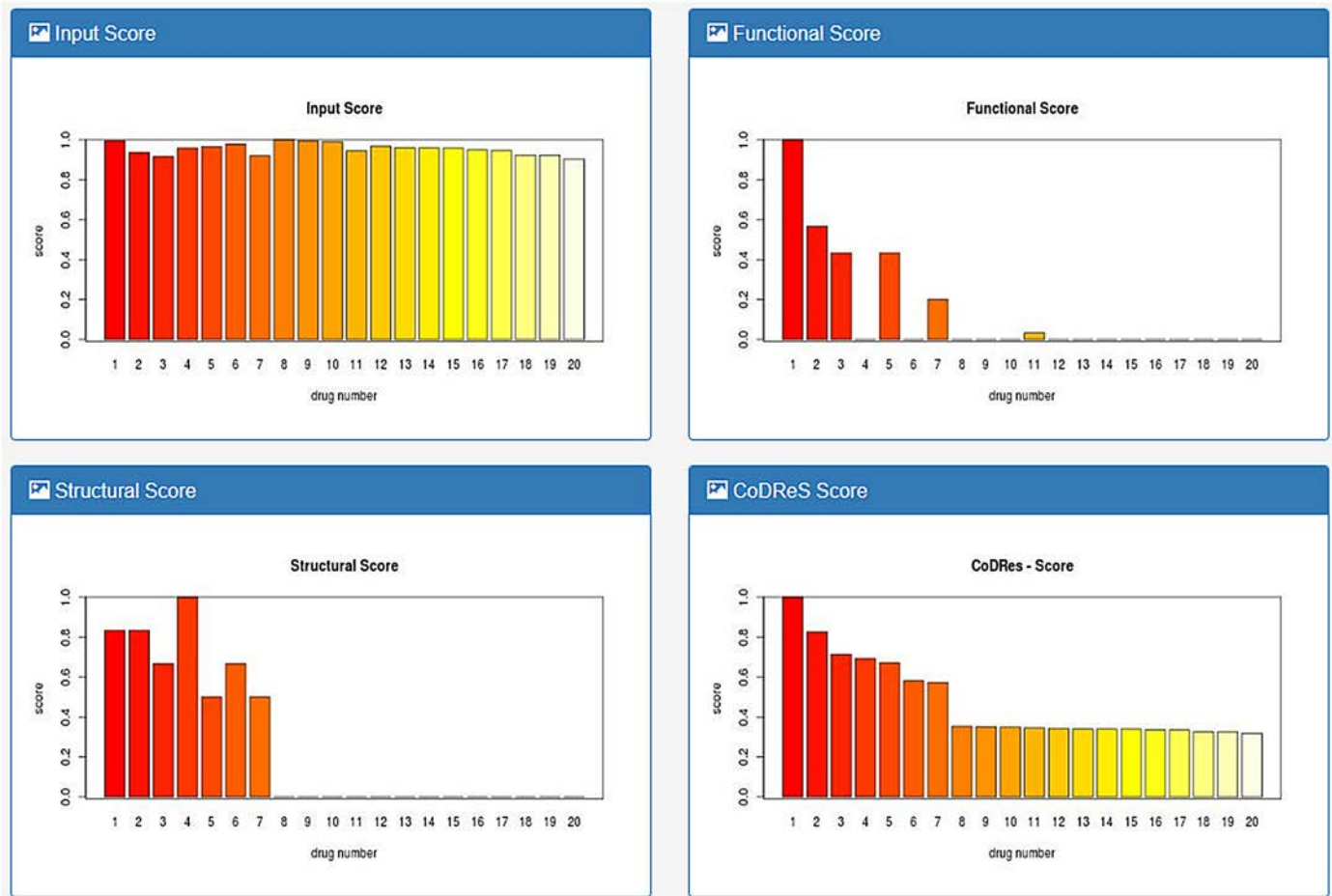**Fig. 4.** Main CoDReS output matrix.

Fig. 5. Drug score diagrams for each scoring parameter.

rest of the databases that were used in CoDReS that include drug names have been parsed and have had their drug names translated to DrugBank's usual names. DrugBank identifiers were also assigned to each input drug where applicable or an "unassigned" value was given otherwise. At the gene level, CoDReS works on gene synonyms that derive from the HGNC [11] database and every other database that contains gene identifiers, is parsed and translated according to the HGNC gene synonyms.

The backend of CoDReS is developed in R. The FS's first parameter, namely Confidence Score, is taken from DisGeNET [12] which is an online database linking genes to diseases by integrating information of various biological databases and giving a score to each interaction. The Ki value of a drug-protein pair is queried from BindingDB [13] which is another online database that contains data from experimentally validated binding affinities between proteins and ligands. To achieve the proper linking between the databases, we convert genes to proteins through Uniprot [14] and by querying BindingDB the proteins are linked

drug name	pregabalin	topiramate	zonisamide	sertraline	ezogabine	eslicarbazepine acetate	piracetam
clobazam	0.113	0.186	0.269	0.412	0.354	0.526	0.348
methylprednisolone	0.333	0.329	0.241	0.268	0.279	0.303	0.275
prednisolone	0.333	0.312	0.241	0.268	0.279	0.303	0.294
lamotrigine	0.088	0.159	0.26	0.354	0.397	0.27	0.188
valproic acid	0.448	0.154	0.059	0.098	0.173	0.132	0.216
acetazolamide	0.132	0.392	0.551	0.096	0.231	0.284	0.293
carbamazepine	0.128	0.162	0.277	0.273	0.333	0.69	0.281
clonazepam	0.141	0.226	0.423	0.333	0.351	0.414	0.384
ethosuximide	0.326	0.268	0.243	0.224	0.246	0.293	0.426
felbamate	0.31	0.333	0.288	0.208	0.574	0.431	0.271

Fig. 6. Structural similarity of input drugs (rows) to clinically failed drugs of input disease (columns) as found in repoDB.

to drug identifiers either from DrugBank or CheMBL [15]. If no Ki value is found for a drug-protein pair, the application uses the median Ki value (184) of BindingDB, instead of the average value (120523.41) which results from the database's outliers. Uniprot is an online knowledgebase that hosts annotated sequences of over 120 million proteins as well as provides protein visualization methods. CheMBL is another drug related database similar to DrugBank. The gene targets of the input drugs are found in the parsed DrugBank, DrugCentral [16] and DGIdb [17] databases.

The StS of each drug is calculated via the Rcpi package of R and requires a drug's molecular structure as input. Each molecular structure is extracted by its respective simplified molecular-input line-entry system (SMILES) type, which is a specification in form of a line notation for describing the structure of chemical species using short ASCII strings. CoDReS tries to either map DrugBank identifiers or CheMBL drug names to SMILES. For every violated rule (aforementioned Lipinski and Veber rules), a drug receives a "plus 1" violation score, with the lowest possible score being six violations. In case there was no SMILES for a specific drug, the candidate drug is assigned a "zero" StS score acting in a

conservative manner by adopting the worst-case scenario presenting the max number of violations.

Another important aspect of the CoDReS tool, is that it highlights the highest ranked drugs of structural clusters, as exemplars, by applying an affinity propagation clustering via the R package APCluster [18] on the similarity matrix of the fingerprints of the input drugs. Specifically, the calcDrugFPSim function of the Rcpi package is used in order to calculate the similarity matrix with a compact E-State fragments fingerprint type and a tanimoto metric as arguments. The structural exemplars are presented as a good choice of disease inhibitors for further investigation, since different structural properties might target different biological mechanisms of a disease phenotype.

Finally, if there are clinical trials carried out for a disease that have led to failed drugs against a disease, the structural similarity between these compounds and the input list of drugs is measured. For this purpose, the online dataset of repoDB [19] has been parsed, keeping the suspended, terminated and withdrawn drugs for each disease identifier. The execution pipeline together with all the integrated databases and packages are depicted in Fig. 2.

2.3. User Execution

The user is required to upload a file containing the drug names and, optionally their respective a-priori scores, as might be acquired from a drug-repurposing tool. As soon as the input file is uploaded, a histogram and a distribution diagram of the input scores are generated (Fig. 3). The weights denoting the importance of the aS, FS and StS are user-selected and have equal default values. The user must then choose a disease from a select box with auto-complete functionality that hosts all DisGeNET diseases.

The output of the CoDReS function is then rendered in tabular form and can be sorted, printed and downloaded either as plain text, csv, spreadsheet or pdf file. The main CoDReS output table consists of the CoDReS rank, the initial position of the input drugs, their input names, their DrugBank usual names and identifiers (or input name again and "unassigned" identifier respectively, if not found in DrugBank's synonyms list), their normalized score per category and their normalized CoDReS, by which they are sorted in descending order (Fig. 4). A drug-score diagram for each scoring parameter is also printed at the bottom of the page after the execution (Fig. 5).

In case there are stated failed clinical trials in repoDB for the selected disease, a similarity matrix of all input drugs against the failed drugs is returned to the user, where the column names represent the failed drugs and the row names the input drugs (Fig. 6).

3. Results and Validation

To check the validity of the CoDReS results, we considered examples disregarding a-priori scores. We chose the top forty diseases from DisGeNET with the most correlated genes that have at least twenty drug candidates in Malacards [20]. These diseases are listed in Table 2. For each disease, we created a mixture list of two hundred drugs: 95% randomly selected from DrugBank and 5% of the top drugs reported from Malacards repository as developed/used for the selected disease. After executing CoDReS for each experiment, we counted the number of the actual disease-related drugs that were found in the top 5% of the ranked drugs, based on their CoDReS along with a p-value calculated through a hypergeometric distribution test. We repeated this procedure a hundred times for each disease and then calculated the median, maximum, minimum and average p-value metrics for each disease. CoDReS ranked effectively (median p-value <.05) the input drugs in 35/40 diseases. CoDReS failed to rank drugs correctly in five out of 40 diseases but this failure can be partially explained since the top ten drugs corresponding to most of these diseases contain abstract substances or generic categories such as "Anti-Inflammatory Agents", "Cytochrome P-450 Enzyme Inhibitors", "Immunologic Factors" or drugs with close

Table 2

information on the diseases used for the validation; the two first columns present the disease's name and umls id respectively as found in disgenet, the third column the total genes that participate in the disease and the fourth column the disease's name as returned from malacards.

Disease name	UMLS ID	Gene count	Malacards name
Malignant neoplasm of breast	C0006142	5053	Breast Cancer
Liver carcinoma	C2239176	3592	Hepatocellular Carcinoma
Colorectal Cancer	C1527249	3298	Colorectal Cancer
Malignant neoplasm of prostate	C0376358	3238	Prostate Cancer
Carcinoma of lung	C0684249	2475	Lung Cancer
melanoma	C0025202	2453	Melanoma
Malignant neoplasm of stomach	C0024623	2397	Gastric Cancer
Glioma	C0017638	2210	Glioma
Ovarian Carcinoma	C0029925	2202	Ovarian Cancer
Alzheimer's Disease	C0002395	1981	Alzheimer Disease
leukemia	C0023418	1940	Leukemia
Glioblastoma	C0017636	1936	Glioblastoma
Schizophrenia	C0036341	1922	Schizophrenia
Squamous cell carcinoma	C0007137	1875	Squamous Cell Carcinoma
Pancreatic carcinoma	C0235974	1868	Pancreatic Cancer
Rheumatoid Arthritis	C0003873	1832	Rheumatoid Arthritis
Adenocarcinoma	C0001418	1711	Adenocarcinoma
Leukemia, Myelocytic, Acute	C0023467	1702	Leukemia, Acute Myeloid
Neuroblastoma	C0027819	1698	Neuroblastoma
Diabetes Mellitus, Non-Insulin-Dependent	C0011860	1671	Diabetes Mellitus, Noninsulin-Dependent
Diabetes Mellitus	C0011849	1506	Diabetes Mellitus
Renal Cell Carcinoma	C0007134	1347	Renal Cell Carcinoma, Papillary, 1
Asthma	C0004096	1312	Asthma
Multiple Myeloma	C0026764	1311	Myeloma, Multiple
Hypertensive disease	C0020538	1309	Hypertension, Essential
Lymphoma	C0024299	1306	Lymphoma
Bladder Neoplasm	C0005695	1216	Bladder Cancer
Epilepsy	C0014544	1176	Epilepsy
Seizures	C0036572	1173	Seizure Disorder
Chronic Lymphocytic Leukemia	C0023434	1119	Leukemia, Chronic Lymphocytic
Lupus Erythematosus, Systemic	C0024141	1112	Systemic Lupus Erythematosus
Multiple Sclerosis	C0026769	1105	Multiple Sclerosis
Cervix carcinoma	C0302592	1104	Cervix carcinoma
Osteosarcoma	C0029463	1102	Osteogenic Sarcoma
Arteriosclerosis	C0003850	1086	Arteriosclerosis
Autoimmune Diseases	C0004364	1059	Autoimmune Disease
Osteosarcoma of bone	C0585442	1041	Bone Osteosarcoma
Squamous cell carcinoma of esophagus	C0279626	1022	Esophagus Squamous Cell Carcinoma
Adenoma	C0001430	999	Adenoma
Coronary Artery Disease	C1956346	980	Coronary Artery Anomaly

Table 3
The median, maximum, minimum and average *p*-value results of 100 codres executions for each disease as calculated by hypergeometric distribution tests. The median *p*-values that are above 0.05 are painted red.

Disease Name	median p-value	max p-value	min p-value	mean p-value
Adenocarcinoma	8.477E-03	7.272E-02	5.644E-04	3.483E-02
Adenoma	8.477E-03	3.268E-01	5.644E-04	2.813E-02
Alzheimer's Disease	2.196E-05	8.477E-03	6.014E-09	3.204E-04
Arteriosclerosis	2.196E-05	5.644E-04	6.014E-09	2.630E-04
Asthma	5.644E-04	8.477E-03	4.920E-07	2.297E-03
Autoimmune Diseases	2.196E-05	5.644E-04	2.196E-05	2.064E-04
Bladder Neoplasm	5.644E-04	3.268E-01	2.196E-05	7.653E-03
Carcinoma of lung	5.644E-04	7.272E-02	4.920E-07	7.107E-03
Cervix carcinoma	3.268E-01	5.915E-01	7.272E-02	3.024E-01
Chronic Lymphocytic Leukemia	8.477E-03	7.272E-02	5.644E-04	1.034E-02
Colorectal Cancer	5.644E-04	8.477E-03	2.196E-05	4.148E-03
Coronary Artery Disease	3.268E-01	3.268E-01	8.477E-03	1.874E-01
Diabetes Mellitus, Non-Insulin-Dependent	5.644E-04	7.272E-02	5.644E-04	3.027E-03
Diabetes Mellitus	4.920E-07	5.644E-04	4.920E-07	1.043E-05
Epilepsy	5.644E-04	8.477E-03	2.196E-05	1.443E-03
Glioblastoma	2.196E-05	8.477E-03	2.196E-05	4.166E-04
Glioma	5.644E-04	8.477E-03	2.196E-05	2.066E-03
Hypertensive disease	8.477E-03	7.272E-02	5.644E-04	9.818E-03
Leukemia, Myelocytic, Acute	4.920E-07	8.477E-03	6.014E-09	4.504E-04
leukemia	2.196E-05	8.477E-03	4.920E-07	1.317E-04
Liver carcinoma	8.477E-03	8.477E-03	5.644E-04	5.628E-03
Lupus Erythematosus, Systemic	6.014E-09	8.477E-03	3.599E-11	9.907E-05
Lymphoma	5.644E-04	8.477E-03	2.196E-05	1.031E-03
Malignant neoplasm of breast	5.644E-04	8.477E-03	2.196E-05	3.059E-03
Malignant neoplasm of prostate	5.644E-04	8.477E-03	5.644E-04	8.809E-04
Malignant neoplasm of stomach	5.644E-04	8.477E-03	5.644E-04	3.255E-03
melanoma	8.477E-03	7.272E-02	8.477E-03	3.160E-02
Multiple Myeloma	2.196E-05	8.477E-03	4.920E-07	5.517E-04
Multiple Sclerosis	5.644E-04	7.272E-02	2.196E-05	4.920E-03
Neuroblastoma	7.272E-02	3.268E-01	7.272E-02	7.526E-02
Osteosarcoma of bone	3.268E-01	3.268E-01	7.272E-02	3.242E-01
Osteosarcoma	3.268E-01	3.268E-01	7.272E-02	3.242E-01
Ovarian Carcinoma	2.196E-05	8.477E-03	2.196E-05	2.823E-03
Pancreatic carcinoma	6.014E-09	8.477E-03	6.014E-09	2.679E-04
Renal Cell Carcinoma	5.644E-04	5.644E-04	2.196E-05	5.101E-04
Rheumatoid Arthritis	8.477E-03	3.268E-01	8.477E-03	1.047E-01
Schizophrenia	4.920E-07	2.196E-05	8.463E-14	5.631E-06
Seizures	8.477E-03	7.272E-02	6.014E-09	5.919E-03
Squamous cell carcinoma of esophagus	6.014E-09	5.644E-04	3.599E-11	6.927E-06
Squamous cell carcinoma	5.644E-04	8.477E-03	4.920E-07	9.173E-04

to zero gene targets participating in the disease. The statistical results are presented in Table 3.

4. Discussion

In this article we present CoDReS, a drug (re-)ranking tool that can act as a tool for post filtering drug lists generated either by conventional drug repurposing tools or by any other drug discovery pipeline. CoDReS should be used as a means of suggesting the best candidates for in vitro or clinical studies by combining a priori knowledge with functional and structural information. The in silico validation schema of CoDReS, as presented in the previous paragraph, brought the disease-related drugs to the top of the random drug pool in almost every case. Despite the promising results, this schema is only a computational validation

of the tool's capabilities. In the end, the scientists using the tool should always incorporate their knowledge, expertise and the bibliography in order to decide the best drug candidates for further experiments.

Declarations of Copmeting Interest

None declared.

Acknowledgments

References

[1] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 2017;171: 1437–52 [e1417].

[2] Duan Q, Reid SP, Clark NR, Wang Z, Fernandez NF, et al. L1000CDS 2: LINCS L1000 characteristic direction signatures search engine. NPJ Syst Biol Appl 2016;2:16015.

[3] Backman TW, Cao Y, Girke T. ChemMine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Res 2011;39:W486–91.

[4] Cao D-S, Xiao N, Xu Q-S, Chen AF. Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions. Bioinformatics 2014;31: 279–81.

[5] Cao Y, Charisi A, Cheng L-C, Jiang T, Girke T. ChemmineR: a compound mining framework for R. Bioinformatics 2008;24:1733–4.

[6] Karatzas E, Bourdakou M, Kolios G, Spyrou G. Drug repurposing in idiopathic pulmonary fibrosis filtered by a bioinformatics-derived composite score. Sci Rep 2017; 7:12569.

[7] Yung-Chi C, Prusoff WH. Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol 1973;22:3099–108.

[8] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23:3–25.

[9] Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, et al. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 2002;45: 2615–23.

[10] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2017;46:D1074–82.

[11] Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, et al. Genenames. org: the HGNC and VGNC resources in 2017. Nucleic Acids Res 2016:gkw1033.

[12] Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 2016:gkw943.

[13] Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Res 2006;35:D198–201.

[14] Consortium U. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 2018;47:D506–15.

[15] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, et al. The ChEMBL database in 2017. Nucleic Acids Res 2016;45:D945–54.

[16] Ursu O, Holmes J, Knockel J, Bologa CG, Yang JJ, et al. DrugCentral: online drug compendium. Nucleic Acids Res 2016:gkw993.

[17] Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, et al. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. Nucleic Acids Res 2017;46: D1068–73.

[18] Bodenhofer U, Kothmeier A, Hochreiter S. APCluster: an R package for affinity propagation clustering. Bioinformatics 2011;27:2463–4.

[19] Brown AS, Patel CJ. A standard database for drug repositioning. Scientific Data 2017; 4:170029.

[20] Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Res 2016;45:D877–87.