RESEARCH

Novel drug-target interactions via link prediction and network embedding

E. Amiri Souri[1], R. Laddach[1,2], S. N. Karagiannis[2,3], L. G. Papageorgiou[4] and S. Tsoka[1*]

*Correspondence:
sophia.tsoka@kcl.ac.uk
[1] Department
of Informatics, Faculty
of Natural, Mathematical
and Engineering Sciences,
King's College London, Bush
House, London WC2B 4BG,
UK
Full list of author information
is available at the end of the
article

Abstract

Background: As many interactions between the chemical and genomic space remain undiscovered, computational methods able to identify potential drug-target interactions (DTIs) are employed to accelerate drug discovery and reduce the required cost. Predicting new DTIs can leverage drug repurposing by identifying new targets for approved drugs. However, developing an accurate computational framework that can efficiently incorporate chemical and genomic spaces remains extremely demanding. A key issue is that most DTI predictions suffer from the lack of experimentally validated negative interactions or limited availability of target 3D structures.

Results: We report DT2Vec, a pipeline for DTI prediction based on graph embedding and gradient boosted tree classification. It maps drug-drug and protein–protein similarity networks to low-dimensional features and the DTI prediction is formulated as binary classification based on a strategy of concatenating the drug and target embedding vectors as input features. DT2Vec was compared with three top-performing graph similarity-based algorithms on a standard benchmark dataset and achieved competitive results. In order to explore credible novel DTIs, the model was applied to data from the ChEMBL repository that contain experimentally validated positive and negative interactions which yield a strong predictive model. Then, the developed model was applied to all possible unknown DTIs to predict new interactions. The applicability of DT2Vec as an effective method for drug repurposing is discussed through case studies and evaluation of some novel DTI predictions is undertaken using molecular docking.

Conclusions: The proposed method was able to integrate and map chemical and genomic space into low-dimensional dense vectors and showed promising results in predicting novel DTIs.

Keywords: Machine learning, Graph-embedding, Drug repurposing, Link prediction

Background

Discovering a new drug is a high-risk, time-consuming, and expensive process [1], a process that typically takes more than 15 years, costs $2.6 billion and is limited by less than 10% success rates [2, 3]. Therefore, there is strong interest to develop new efficient methods able to discover previously unknown activities of existing drugs to uncover new medical purposes outside the original scope, a process known as *drug repositioning* or *repurposing* [4]. Drug repurposing is a promising alternative to traditional drug

Amiri Souri *et al. BMC Bioinformatics*   (2022) 23:121

Page 2 of 16

discovery approaches, offering a shorter route to clinical development, bypassing several stages of clinical development [5] which have already been completed for the original target and reducing major risks, expense and time by several years [6]. The first major step in repositioning a drug is identifying possible potential target proteins by predicting valid DTIs [7].

Due to expensive and time-consuming laboratory experiments, limited availability of physical resources, and the complexity of integrating chemical and genomic spaces, just a small number of DTIs are experimentally validated. Therefore, accurate prediction of DTIs at large scale remains a challenge [8]. Recently, computational prediction methods based on machine learning (ML) principles have received increasing attention [9] because of their ability to integrate different types of biological data and analyse large numbers of possible interactions efficiently, leading to faster and cheaper assessments [10]. There is an urgent need for sophisticated computational modelling approaches [11] to limit the number of potential interactions which can be reasonably verifiable by in vitro screening [12].

Traditional computational methods in DTI prediction are mainly categorised in two types of strategies: molecular docking simulations and ligand-based approaches [13]. However, the applicability of docking simulations is limited by the availability of 3D crystal structures of target proteins which is still unknown for the majority of membrane proteins especially G-protein-coupled receptors (GPCR) [6]. Ligand-based approaches also suffer from low prediction rates when the number of known binding ligands is small [14]. To avoid the limitations of traditional methods, several ML-based models have been developed which achieved considerable success by translating large scale chemogenomics data to a set of features and extracting latent patterns in DTIs [15]. The most popular group of ML methods for DTI prediction is similarity methods which incorporate target-target and drug-drug similarity metrics. These rely on a key underlying assumption that similar drugs may tend to target similar proteins and vice versa [6]. Similarity-based approaches have several advantages which include the ability to connect chemical and genomic spaces and the availability of well-defined similarity measures between the chemical structure of drugs and genomic sequences [16]. By integrating multiple types of similarities into a heterogeneous network, determining new DTIs can be formulated as a link prediction problem in graph analysis [8].

Graph embedding algorithms are popular methods recently used in graph analytics to represent graph structural properties as a set of low dimensional vectors [17] which can be introduced into ML models as input features. The use of embedding methods to infer different biological interactions such as drug-drug [18], protein–protein [19], and drug-target [10] outperforms current state-of-the-art methods [18]. The "2vec" (short for "to vector") models like "graph2vec", "node2vec", etc. are an important category of embedding algorithms inspired by the "word2vec model" [20] a popular word embedding algorithm in natural language processing, which used neural networks to learn word vectors from sentences [20]. The benefits conferred by representing DTI as a vector stem from its capability to incorporate heterogeneous chemical and genomic data into a unified space, in addition to the fact that different ML algorithms can handle numerical input features well. Researchers in recent years have developed different embedding methods for predicting DTIs, such as TriModel [21], DTI-HeNE [9] and DTiGEMS+ [10] based

on drug chemical structure and protein sequence to build similarity networks (details in Additional file 1: Related work). Another embedding-based method for DTI prediction has looked into integrating multi-molecular associations such as protein, drug, disease, lncRNA, and miRNA from multiple databases into a heterogeneous network [22]. Besides embedding methods, a wide variety of computational algorithms have been proposed and summarized recently [6, 10, 23].

Although these models achieved promising performance, the lack of experimentally validated negative interactions is a common limitation of almost all supervised learning methods [8]. To this end, usually non-interacting drug-target pairs are assumed as negative samples. This affects the efficiency of developed models in real-life applications because it could include some positive interactions that have not been tested yet [21]. Moreover, as discussed previously, since the number of known DTIs is considerably smaller than unknown drug-target pairs (labelled as negative samples), it leads to imbalanced classification and skewed results. Therefore, selecting realistic negative interactions was highlighted as one of the important tasks in future developments of DTI prediction [24].

We have previously reported mathematical optimization as means of predicting the affinity of DTI [25, 26]. In this paper, we developed a computational framework that employs data from multiple DTI sources and formulates the problem of deriving new DTIs as link prediction. We used two datasets separately to evaluate and extract novel DTIs; the Golden Standard and ChEMBL datasets. The link prediction methodology employs the following stages: (1) creation of a drug-drug similarity network (2) creation of a protein–protein similarity network, (3) feature extraction using graph embedding of networks in (1) and (2) via node2vec, and defining interactions by concatenating pairs of drug and target features, (4) a classification scheme that employs the embedding features and gradient boosted trees to predict new DTIs. Method performance is based on an external validation metric of the classification model, evaluation of prediction is assessed via molecular docking of drugs predicted to bind the protein target and different case studies of newly predicted DTIs are discussed.

Materials and methods

The overall computational framework is shown in Fig. 1 and includes data pre-processing, implementation of the proposed methodology and evaluation. As described below, first the performance of our approach is compared to similar methods from the literature on a standard benchmark dataset, and then a dataset containing experimentally verified positive and negative interactions was collected to detect more realistic DTIs and drug repurposing.

Standard benchmark dataset

A *"Golden-standard dataset"* was introduced by Yamanishi et al. [27, 28] and has been used previously as a reference for predicting DTIs and comparing the performance of different models [1, 14, 21, 23, 29]. It consists of a binary drug-target edge-list M_{DTI} (Fig. 1: a3) and two similarity matrices in chemical and genomic spaces M_{DDS} (Fig. 1: a1) and M_{PPS} (Fig. 1: a2) respectively. Since experimentally validated negative interactions are not available, interactions between all possible drug-target pairs where a known
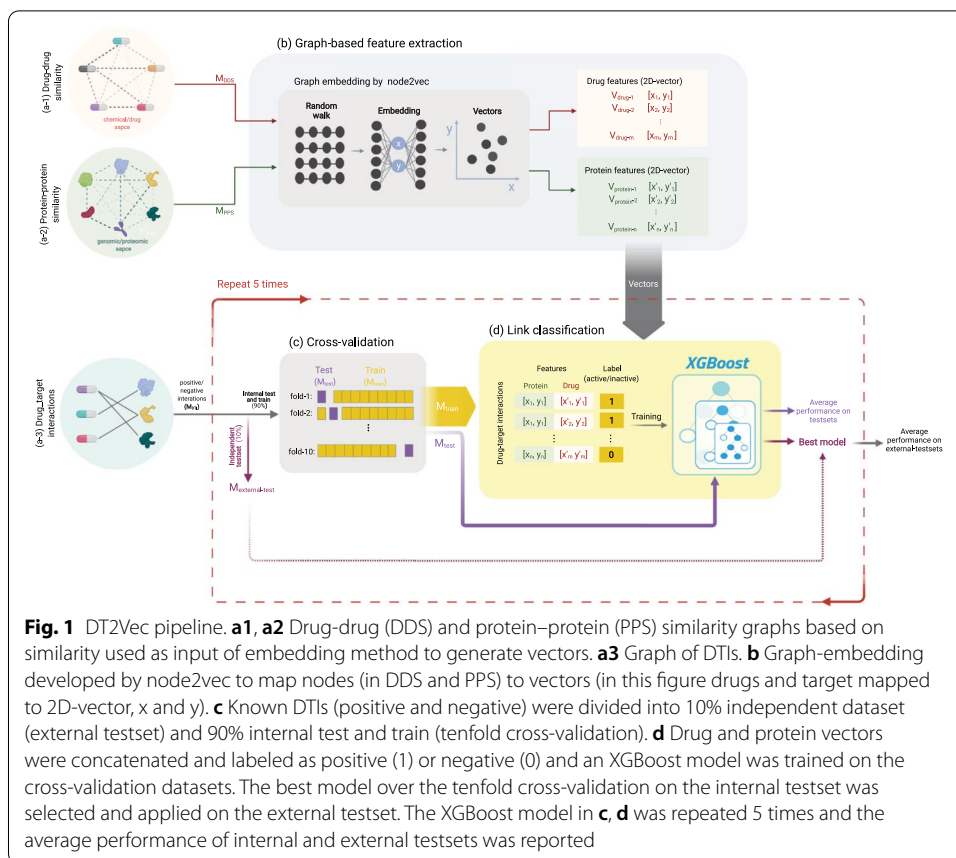
Fig. 1 DT2Vec pipeline. **a1**, **a2** Drug-drug (DDS) and protein–protein (PPS) similarity graphs based on similarity used as input of embedding method to generate vectors. **a3** Graph of DTIs. **b** Graph-embedding developed by node2vec to map nodes (in DDS and PPS) to vectors (in this figure drugs and target mapped to 2D-vector, x and y). **c** Known DTIs (positive and negative) were divided into 10% independent dataset (external testset) and 90% internal test and train (tenfold cross-validation). **d** Drug and protein vectors were concatenated and labeled as positive (1) or negative (0) and an XGBoost model was trained on the cross-validation datasets. The best model over the tenfold cross-validation on the internal testset was selected and applied on the external testset. The XGBoost model in **c**, **d** was repeated 5 times and the average performance of internal and external testsets was reported

Table 1 Dataset details (a) Golden-standard dataset (b) ChEMBL-based dataset

(a) Golden_standarad_dataset	# Drug	# Target	# Negative/unknown DTI	# Positive interaction	Class ratio
Enzyme (E)	445	664	292,554	2926	0.01
Ion channel (IC)	210	204	41,364	1476	0.04
G-protein-coupled receptors (GPCR)	223	95	20,550	635	0.03
Nuclear receptor (NR)	54	26	1314	90	0.07
Total	791	989	777,172	5127	0.01
(b) ChEMBL_dataset	# Drug	# Target	# Real negative DTI[†]	# Positive interaction*[,†]	Class ratio
CheMBL	548	556	2057	1721	0.84
	# Weak DTI**	# Unknown DTI[‡]			
	532	300,378			

*pChEMBL value $\geq$ 5.5, **pChEMBL value > 0, [†]Development-dataset, [‡]Experimental-dataset

interaction does not exist, were considered as the set of negative interactions in most studies [21, 27, 30]. Details of the "*Golden-standard dataset*" are given in Table 1a and Additional file 2.

In order to obtain a balanced dataset of positive and negative interactions, in most studies unknown interactions were selected randomly and labelled negative [6, 10, 21]. However, this set of randomly selected cases may include some real positive interactions that are yet unknown, which may lead to artefacts in DTI prediction [1, 23]. Therefore, in this case, the Recall metric (i.e. true positives/(true positives + false negatives)) can better reflect the performance of the models, as it is calculated on validated positive labels. In detecting new DTIs, false-positive predictions (i.e. negative interactions that were predicted to be positive) may be reported as newly discovered interactions [10, 14, 21, 28], thereby creating problems in distinguishing between false-positives caused by model error and newly discovered interactions. Therefore, although the Golden-standard dataset is convenient as a common benchmark dataset for comparing the performance of different developed models, its limitation is that it only contains experimentally validated positive interactions which are not suitable for training a realistic DTI prediction model [15]. Screening reliable true negative interactions was recently highlighted as one of the critical steps in improving the prediction accuracy of developed methods [24, 29] and this point is better addressed through the use of the dataset described below.

A realistic dataset for drug repurposing extracted from the ChEMBL database

To address these limitations, a DTI dataset was collected from the ChEMBL repository [31] that contains experimentally validated negative and positive interactions. In literature, an activity threshold of pChEMBL of 5.0 is typically used to label an interaction as active. In chemical assay experiments, the acceptable model results should be higher than 10 μM affinity (or pChEMBL = 5) [32]. To ensure that the positive interactions are strong and consequently offer more accurate prediction results, activity greater than 5.5 was chosen [33]. Positive (pChEMBL $\geq$ 5.5) and negative (labelled as inactive interactions in ChEMBL repository) interactions form a binary edge-list M_{DTI} (Fig. 1: a3).

Drug similarity measures were calculated through MACCS [34] based on structural information which can codify 166 structural fingerprints in bit positions. Then, we measured the similarity between drug pairs using the Tanimoto coefficient in the range of 0 to 1 [35]. Open Babel [36] in Python 3.7.3 was used to generate the drug-drug similarity network M_{DDS} (Fig. 1: a1) [37]. The protein similarity network, M_{PPS} (Fig. 1: a2), was computed using sequence alignment [38], implemented through the parallelised version of protein similarity calculation using the "protr" package in R 4.0.2 [39]. Additional file 3: Figure S1 and Additional file 2 describe steps performed to collect the ChEMBL-based dataset. Known positive and negative DTIs comprise the *development-dataset* (Table 1), a set of interactions used to build the ML model. All possible drug-target pairs with no known (active or inactive) interactions in ChEMBL (Table 1) were defined as the *'experimental-dataset'* and were used for predicting interactions and performing drug repurposing.

Development and evaluation of the DT2Vec model

We report DT2Vec, an ML method for drug-target interaction prediction, trained on features extracted using graph embeddings. DT2Vec was implemented and evaluated on the Golden-standard dataset as well as experimentally validated datasets extracted from ChEMBL. The first dataset was used as a benchmark to validate the performance of the

developed model through comparison with three state-of-the-art open-source chemog-enomic algorithms that employed the Golden-standard database, namely DNILMF [40], DT-Hybrid [41], DDR [14]. These methods are based on graph similarity algorithms and are reported as top-performing methods in DTI prediction [1, 10]. Details of these methods are in Additional file 1: Related work. After benchmarking, the DT2Vec model was implemented on the ChEMBL based dataset which included first training the model on known positive and negative interactions and then applying the model on unknown DTIs to detect novel interactions.

Graph-based feature generation using node2vec

The performance of ML algorithms is highly dependent on choosing a set of informative and discriminative features. Many ML methods benefit from semantically meaningful features, automatically extracted from highly structured objects like graphs which not only reduce the manual feature engineering effort but also enhance the predictive capa-bility of the model. As shown in Fig. 1b, in order to extract features, node2vec [42], a semi-supervised feature learning algorithm was applied to the weighted graph of drug-drug and protein–protein similarities separately to embed drug and target nodes into a continuous vector space with n-dimensions (as an example, in Fig. 1b the embedded vectors, V_{drug} and $V_{protein}$ are two-dimensional). Based on recent ML research, node-2vec outperforms other existing state-of-the-art methods in node embedding [17, 42, 43]. Recently, node2vec showed promising results on DTI prediction by mapping drug, protein, disease, lncRNA and miRNA association networks to vectors [44]. We used node2vec implemented in Python 2 using the source code available in GitHub [43]. DTIs were defined by concatenating the embeddings of the drug and protein similarity net-works and then were used as input features for an ML classifier. The drug-drug and pro-tein–protein similarity networks (Fig. 1a, b) were clustered using Louvain [45] to obtain a topological characterisation of the structure of the networks and networkX [46] was used to visualize the networks. Then the drug and target embedded vectors are visual-ized based on two principal components using PCA [47] to illustrate how the embed-ding vectors represent the communities in the networks.

Data-partitioning and cross-validation

There are several strategies that can be used in validating DTI prediction models [48]. Cross-validation (CV) schemes are a robust strategy to estimate how a model general-izes, whereby data is split multiple times to increase the variation in the training and testing data. The developing processes employed internal and external testing whereby the drug-target edge lists, M_{DTI}, were split into 90% internal training (80% train, 10% internal-test) and 10% external testset, repeated 5-times tenfold CV (Fig. 1c). The best model was selected based on the internal test set and assessed on the external testset which is blind to the process of developing the model, to obtain a more realistic rep-resentation of generalised performance [49]. DT2Vec was trained and tested on the Golden-standard (positive and randomly selected unknown interactions) and ChEMBL (positive interactions with pChEMBL $\geq$ 5.5 and experimentally validated negative inter-actions, named *'development_dataset'*) datasets. After validating the performance of the method using CV, the final model is built on the whole data. Then the final model on

the ChEMBL dataset was applied to the unknown ChEMBL interactions (named '*experimental-dataset*') to detect novel DTIs. Details are shown in Additional file 4: Fig. S2.

Machine learning-based link classification

DTIs were represented as a 2n-dimension vector (n = 100, Fig. 1d shows n = 2 as an example) by concatenating the drug and target embedding features and labelled as "active" or "inactive" as described previously [17]. The DTI prediction problem was formulated as a binary classification problem built on XGBoost [8] (Fig. 1d). XGBoost is a stochastic gradient boosting algorithm which combines weak ensemble decision trees and was selected due to its high speed, accuracy, and ability to handle imbalanced datasets [4]. Moreover, by taking advantage of XGBoost returning the prediction probability score, we were able to rank DTIs based on the confidence score that the model provides. Grid-search was performed on training set samples within each cross-validation fold to find the best set of hyperparameters. The model was implemented in Python 3.7.3, using XGBoost 0.90 with hyperparameters of maximum tree depth = 4, subsample ratio = 1, minimum child weight = 2, and learning-gamma rate = 0.8. To evaluate the performance of the model, the average Precision, Recall, and f_β-score across all cross-validation sets are calculated.

Extracting new DTIs

In order to demonstrate the use of the DT2Vec for drug repurposing, novel DTIs were predicted from a dataset of all drug-target pairs in ChEMBL where interaction is not known (*experimental-dataset*). After benchmarking the model through cross-validation, DT2Vec was built on known DTIs in the *development-dataset*, before being applied to the *experimental-dataset*. Selected newly predicted DTIs by our method were assessed by performing docking. First, predicted DTIs were selected by two criterial: (1) a probability score by XGBoost ≥ 0.99%, and (2) DTIs having drugs in phase-4 clinical trials. The amino-acid sequence of protein targets of interest was used to obtain PDB structures for docking (https://www.rcsb.org/). Chains attributed to homo sapiens were used to calculate the similarity score based on sequence alignment using protr [39] with default settings in R 4.0.5. SwissDock [50] was used with default parameters to perform drug-protein docking.

Results

Implementation of DT2Vec model on the benchmark Golden standard dataset and comparison with other methods

We developed and evaluated our proposed embedding-based DTI model (DT2Vec) on the benchmark Golden-standard dataset (Table 1a) [27]. Node2vec was used on DDS and PPS to map them to 100-dimension vectors which reported as the best vector size for preserving network neighborhoods of nodes [10, 17, 42, 43]. To obtain a topological characterisation of the networks, the drug and protein similarity networks were clustered which consisted of 4 (with 360, 203,181, and 47 drug members) and 7 (with 484, 204, 120, 95, 42, 26, and 18 protein members) drug and protein communities respectively as shown Additional file 5: Fig. S3 a-1, a-3. Additional file 5: Figure S3 a-2, a-4 show PCA of drugs and targets based on embedded vectors

as features and colours indicating cluster membership in DDS and PPS networks. It was observed (Additional file 55: Fig. S3a) that the embedded vectors based on the Golden-standard dataset can represent the topological features of networks well. In Additional file 6: Fig. S4a, the PCA of embedded vectors in the protein target similarity dataset is shown, coloured according to protein type (i.e. enzymes, GPC receptors, nuclear receptors, and ion channels).

To create DTI labels in the Golden-standard dataset, as known negative samples are not available, unknown interactions were randomly selected to create negatively labelled samples. This assumption leads to unreliable false-positive predictions (and therefore Precision, i.e. true-positives/(true-positives + false-positives)), so the f_2-score, which weighs Recall higher than Precision, was deemed suitable. However, since a trade-off between Precision and Recall exists, the goal of the model should be high Recall without sacrificing Precision. The performance was measured based on tenfold cross-validation which was repeated five times and compared with three methods DNILMF [40], DT-Hybrid [41], DDR [14]. The DT2Vec model achieved f2-score, Recall, Precision average of 91.69% (1.5), 92.63% (0.82), 88.13% (0.5) which was better than DNILMF with 87.92% (1.4), 87.84% (1.63), 88.27% (1.84), DT-Hybrid with 72.7% (1.2), 70.76% (1.55), 81.72% (0.95) and DDR with 89.87% (1.25), 89.83% (1.4), 90.08% (1.07) respectively.

Development of DT2Vec on ChEMBL interactions

A DTI dataset was collected from ChEMBL containing experimentally validated negative and positive interactions in order to offer a more realistic interaction set (Table 1b). ChEMBL DDS and PPS networks were clustered to 4 (with 193, 150, 105, and 100 drugs) and 5 (with 304, 153, 86, 8, and 5 proteins) communities respectively (Additional file 5: Fig. S3 b-1, b-3). Additional file 5: Figure S3 b-2, b-4 shows PCA of drugs and targets using the embedding vectors colored based on the cluster membership, showing that node2vec vectors can represent the topological properties of the original networks well. For reference, Additional file 6: Fig. S4b shows PCA of protein target embedding vectors colored according to protein type.
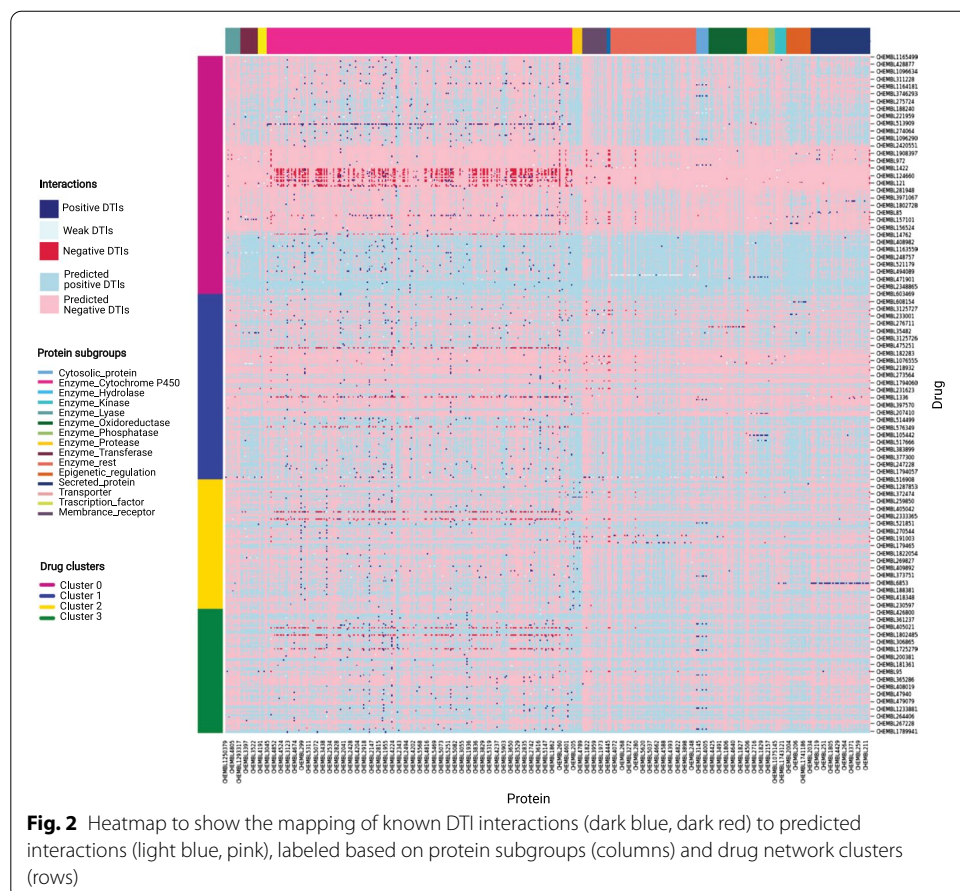
Interactions in the ChEMBL dataset were divided into (1) a *development-dataset* comprising 2057 negative and 1721 positive (pChEMBL value $\geq$ 5.5) interactions, and were used to train and test DT2Vec, (2) an *experimental-datase*t, comprising all unknown interactions (a total of 300,378), which were used to predict and extract novel interactions and evaluate performance on independent datasets. In contrast to the benchmark Golden-standard dataset, in ChEMBL data the DT2Vec model was trained with experimentally verified negative interactions, therefore false-positive predictions and the Precision metric are more realistic, and the f_1-score (which weights Precision and Recall equally) was used. The average performance through 5 times tenfold CV on the external test sets was calculated and the model achieved high Precision of 92.79% (0.02) showing low false-positive predictions and indicating that the model can accurately predict novel DTIs. The model also demonstrated promising results on Recall, f1-score, AUPR and AUC with 92.88% (0.02), 92.82% (0.01), 89.42% (0.02), and 94.09% (0.01) respectively (Additional file 7: Fig. S5 shows ROC plot across 5 runs).
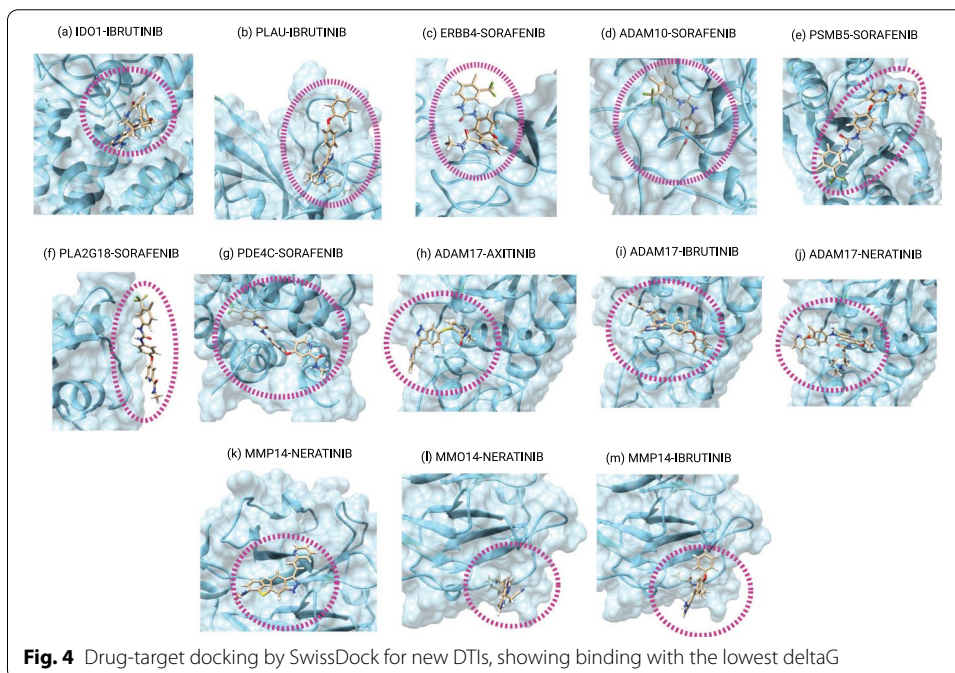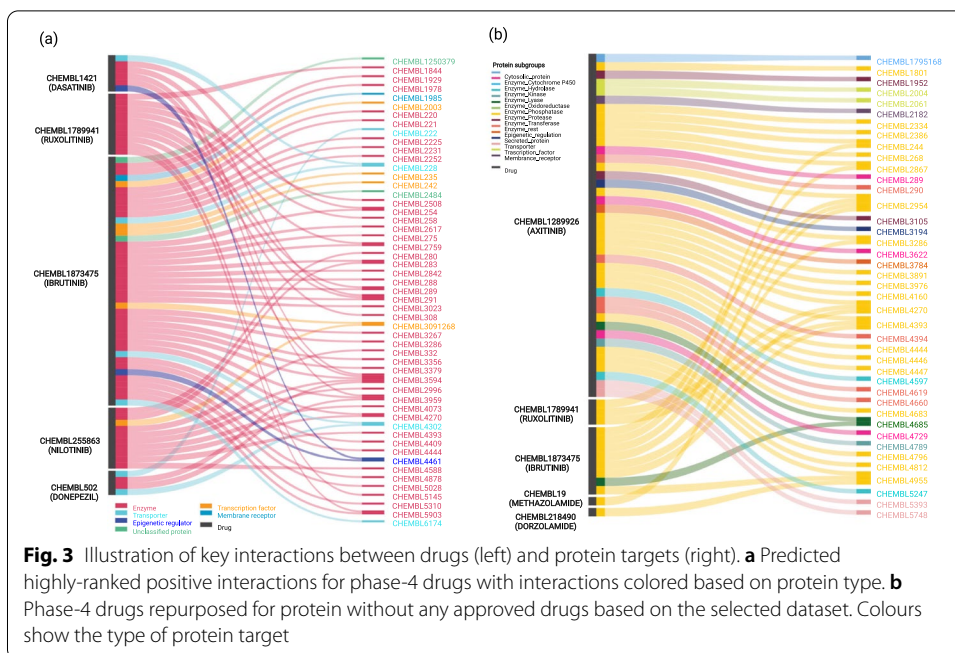
Extracting and evaluating DTIs

In order to demonstrate the use of the DT2Vec for drug repurposing, novel DTIs were predicted from a dataset of all drug-target pairs in ChEMBL where interaction is not known (*experimental-dataset*). Figure 2 shows DTIs as a heatmap where all *known* drug-target interactions (dark blue for positive or red for negative) and *predicted* DTIs (light blue for positive or red for negative) are mapped, with proteins in columns coloured according to subgroup and drugs in rows coloured by chemical similarity. By comparing known and predicted DTIs in Fig. 2, we illustrate that prediction via DT2Vec can extend beyond the 'similar drug for similar target' principle, which has traditionally been the basis of various drug repurposing efforts.

To further illustrate the nature of predicted DTIs via DT2vec, the top novel positive interactions where a drug has been approved at phase-4 clinical trial are shown (Fig. 3a). In the *development-dataset* (known DTIs), 387 (out of 556) proteins have known positive interactions and only 162 can be associated with phase-4 drugs (394 without any approved drugs). Figure 3b shows the top predicted positive interactions for proteins without any phase-4 drugs in this dataset, which represent cases where repurposing may be highly promising.

Drug-target docking was used to investigate the validity of some newly predicted DTIs and shed light on the relevant molecular interaction. Figure 4 and Additional file 8: Fig.

Fig. 2 Heatmap to show the mapping of known DTI interactions (dark blue, dark red) to predicted interactions (light blue, pink), labeled based on protein subgroups (columns) and drug network clusters (rows)

Fig. 3 Illustration of key interactions between drugs (left) and protein targets (right). **a** Predicted highly-ranked positive interactions for phase-4 drugs with interactions colored based on protein type. **b** Phase-4 drugs repurposed for protein without any approved drugs based on the selected dataset. Colours show the type of protein target

Fig. 4 Drug-target docking by SwissDock for new DTIs, showing binding with the lowest deltaG

S6 show docking results and visualised with UCSF Chimera for new predicted DTIs. The free energy of binding (deltaG) for the first 10 docking groups is shown in Additional file 8: Fig. S6a. In general, negative deltaG indicates favourable binding of drg to the respective protein. Among 556 target proteins in the ChEMBL dataset, 77 proteins cannot be associated to a known 3D structure, indicating challenging cases where ML-based DTI prediction can be particularly advantageous for drug discovery. Additional

file 10: Table S1 shows predicted phase-4 drugs (approved drugs) for the set of proteins with no known 3D structure. The next section discusses the evaluation of drug-target interactions in more detail.

Discussion

We report a novel methodology for drug repurposing based on graph embedding and below we evaluate our method by focusing on case studies of newly predicted DTIs that we identified. In this section, we provide extensive discussion of novel DTIs that are associated with approved drugs with multiple highly positive new interactions, cancer-related targets [51], and proteins without 3D-structure data [6] where drug discovery is more challenging and therefore of particular interest.

As demonstrated in Fig. 3a, based on DT2Vec prediction, IBRUTINIB (CHEMBL1873475) targeted multiple proteins which are important in repurposing. IBRUTINIB is a known inhibitor of Bruton's tyrosine kinase (BTK). By acting downstream of the B cell receptor (BCR), IBRUTINIB can block malignant B cell signalling and activation and lead to apoptosis [52, 53]. Based on the known interactions in *development-dataset*, IBRUTINIB can bind to eleven proteins (BTK, BLK, PSCTK4, BMX, EGFR, ERBB2, ERBB4, PTK4, CDHF12, ITK, and ERG). In predictions obtained by our method, it is indicated that it targets eight proteins that have no approved drugs based on our dataset (Fig. 3b). Emerging evidence shows that some of these genes, such as CHEMBL4685 (IDO1) and CHEMBL3286 (PLAU), are linked to cancer development. Specifically, IDO1, a metabolic enzyme involved in tryptophan metabolism and an interferon-induced checkpoint molecule associated with immune suppression, has been linked to many types of cancer, such as acute myeloid leukaemia, ovarian cancer or colorectal cancer. It is indicated that IDO1 is part of the malignant transformation process, helping malignant cells escape eradication by the immune system. Inhibiting IDO1 could increase the effect of chemotherapy as well as other immunotherapeutic protocols [54–57]. In the case of PLAU, elevated expression levels are found to be correlated with malignancy, it is more commonly associated with cancer progression than the tissue plasminogen activator (tPA) [58] and inhibitors to this target have been sought as anticancer agents. It is noted that clinical evaluation of these agents is hampered by incompatibilities between human and murine biology. Moreover, urokinase is used by normal cells for tissue remodelling and vessel growth, which necessitates distinguishing cancer-associated urokinase features for specific targeting [59, 60]. Molecular docking was used to validate the interactions between IBRUTINIB and these two protein targets, showing favourable interactions where IBRUTINIB can bind to one region in PLAU and three regions in IDO1 (Additional file 8: Fig. S6 ab,1-2).

In another example, among phase-4 clinical trial drugs in the *development-dataset*, SORAFENIB (CHEMBL1336) was linked to having one of the highest positive interactions. SORAFENIB is a kinase inhibitor approved for treating patients with inoperable liver cancer [61] and metastatic renal cell carcinoma. DT2Vec predicted five new targets for this drug (probability score ≥ 0.99%): ERBB4 (CHEMBL3009), ADAM10 (CHEMBL5028), PSMB5 (CHEMBL4662), PLA2G1B (CHEMBL4426) and PDE4C (CHEMBL291). Specifically, ERBB4 has been recently found to be expressed in several tumours and tumour cell lines and its inhibition can slow tumour growth [62]. ADAM10

is likely to be involved in breast cancer progression, especially in the basal subtype [63]. PSMB5 is associated with proliferation and drug resistance in triple-negative breast cancer [64]. Recent studies point to a relationship between PLA2G1B [65] and PDE4C [66] with cancer. The docking results validating the interactions are shown in Fig. 4 (details in Additional file 7: Fig. S5 a,b3-7). Therefore, there is considerable evidence to support repurposing SORAFENIB to these new targets.

Immune evasion in cancer is an unsolved problem affecting the efficacy of immunotherapies and decreasing patient survival [67]. We selected the cases of ADAM17 and MMP14 targets to illustrate the potential of our methodology, as they have been implicated in tumour evasion through metalloproteinase function and catalysis of cleavage of endogenous MHC class I-related chain molecule (MIC) A and B [68]. NK cells recognise and become activated by interacting with MIC via the NKG2D receptor. The soluble form of MIC (sMIC) can also bind to NKG2D, which is internalised and subsequently reduces NK anti-tumoural functions [69, 70]. Two PDB entries with the same similarity score have been found for ADAM17 (2I47 and 3G42); 2I47 was selected due to better resolution and chain A was used for docking [71]. Three PDB entries were identified for MMP14 with the same similarity score (3C7X, 6CLZ, 6CM1) and 3C7X was selected [72]. There were 17 drugs predicted to interact with ADAM17 (Additional file 11: Table S2a) and 28 for MMP14 (Additional file 11: Table S2b). The DDS was used to assess the similarity between structures predicted to bind to each of the targets. This was in order to identify any bias towards a common core structure shared between drugs, which could indicate that drugs were identified only based on structural similarity. We note that some drugs predicted by the method are structurally different according to fingerprint similarity, which increases the number of potential therapeutic options (Additional file 9: Fig. S7). CHEMBL1289926, CHEMBL1873475, and CHEMBL18002 were identified as having the highest probability of interaction with ADAM17. Similarly, CHEMBL1289926, CHEMBL1873475 and CHEMBL1789941 were predicted to have positive interaction with MMP14. The docking results of these new interactions were shown in Fig. 4 (details in Additional file 8: Fig. S6 a,b8-13). The deltaG for the first 10 groups of molecules indicates favourable binding.

Other interesting examples are proteins without known 3D-structure for which drug repurposing can have a significant impact. For example, EPHA6 (CHEMBL4526), plays an important role in the formation of breast cancer and poses a new therapeutic target for patients with ER-negative and HER2 positive [73]. Based on DT2Vec prediction, four approved drugs AXITINIB (CHEMBL1289926), IBRUTINIB (CHEMBL1873475), DORZOLAMIDE (CHEMBL218490), AFATINIB (CHEMBL1173655), and DONEPEZIL (CHEMBL502) can bind to EPHA6 with the probability score ≥ 0.99%. AXITINIB has been shown to offer promising results on inhibiting the growth of breast cancer in animal models [74], renal cell carcinoma in clinical trials [75] and several other tumour types [76]. As mentioned before, IBRUTINIB is known as a cancer growth inhibitor and inducer of apoptosis [52, 53]. AFATINIB is also approved as a treatment for lung cancer [77], breast cancer and other cancer types [78]. Finally, DORZOLAMIDE has shown antitumor activity which affects TXNIP-dependent tumour suppression pathways and also causes downregulation in the level of bcl-2 in cancer cells. A previous study also provided evidence for synergistic antitumor activities of DORZOLAMIDE and

mitomycinC against Ehrlich ascites carcinoma tumour growth in vivo, and this might offer a potential combination to evaluate in future clinical studies [79]. It shows that the method was able to suggest new tumour inhibitor drugs for a protein of unknown 3D structure with a crucial role in cancer [80].

Conclusion

In overview, drug repurposing is a promising avenue in drug discovery, supporting the discovery of new protein targets for an approved drug. The availability of large scale chemogenomic data, coupled with the efficiency of ML methodologies, can support drug discovery in a time-efficient and cost-effective manner. In this study, we present a machine learning pipeline that combines network embedding and gradient boosted tree classification, to cast a link prediction strategy for detecting new DTIs. The model was implemented and validated via mining two different drug-target datasets, and evaluations of DTIs included molecular docking simulations and reviewing the literature. A key advantage of our method is that it does not require a priori 3D structure information, and relies solely on drug chemical structures and protein sequences for proposing promising repurposing cases. We note that predicting new and previously unknown DTIs is not only important for drug repositioning purposes, but can also improve our understanding of drug side effects which are usually caused by unexpected interactions of drugs with off-target proteins.

Abbreviations
CV: Cross-validation; DDS: Drug-drug similarity; DTI: Drug target interaction; GPCR: G-protein-coupled receptors; ML: Machine learning; PCA: Principal component analysis; PPS: Protein–protein similarity; tPA: Tissue plasminogen activator.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-022-04650-w.

Additional file 1: Related work. Related methods from literature. Summary of different embedding and similarity based methods for DTI prediction.

Additional file 2: Datasets. Details about datasets used in this study are provided.

Additional file 3: Figure S1. Data collection procedure. (a) Workflow to collect data from the ChEMBL database, (b) Scatter plot of the pChEMBL value of collected DTIs on a boxplot.

Additional file 4: Figure S2. Datasets and data splitting. Details of data types and data splitting in cross-validation.

Additional file 5: Figure S3. Networks of drug and protein similarities. Topological representation of the DDS and PPS networks and PCA of drugs and targets based on embedded vectors.

Additional file 6: Figure S4. PCA plots. PCA of embedded vectors of proteins coloured according to protein type.

Additional file 7: Figure S5. ROC plots over the ten-fold five times cross-validation.

Additional file 8: Figure S6. DTIs through docking. Molecular docking performed via SwissDock for novel predicted DTIs. (a) deltaG for the first 10 groups of molecules clustered by conformers similarity. (b) Binding locations with the lowest deltaG and all groups of conformers.

Additional file 9: Figure S7. Drug similarity heatmaps. Drug-drug similarity for drugs predicted to interact with (a) ADAM17 and (b) MMP14.

Additional file 10: Table S1. Predicted phase-4 drugs for proteins of unknown 3D structure. List of potential phase-4 drugs (approved drugs) that can interact with unknown 3D structure proteins.

Additional file 11: Table S2. Predicted drugs for MMP14 and ADAM1. List of potential approved drugs that can target MMP14 and ADAM17.

Author contributions

EAS, LGP and ST conceived the study design. EAS conducted the data collection. EAS conducted computational experiments. EAS, RL, LGP, SNK and ST analysed and interpreted the data. EAS, RL, LGP, SNK and ST wrote the initial manuscript drafts. EAS, RL, LGP, SNK and ST performed critical editing. All authors reviewed the manuscript.

Availability of data and materials

The datasets analysed during the current study are available in the ChEMBL repository, https://www.ebi.ac.uk/chembl/, and Yamanishi et al. article (https://doi.org/10.1093/bioinformatics/btn162) and its supplementary information files http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/. The code is available at https://github.com/elmira-amiri/DT2Vec.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Informatics, Faculty of Natural, Mathematical and Engineering Sciences, King's College London, Bush House, London WC2B 4BG, UK. [2]St. John's Institute of Dermatology, School of Basic and Medical Biosciences, King's College London, Guy's Hospital, London SE1 9RT, UK. [3]Breast Cancer Now Research Unit, School of Cancer and Pharmaceutical Sciences, King's College London, Guy's Cancer Centre, London SE1 9RT, UK. [4]Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK.

References

1. Hao M, Bryant SH, Wang Y. Open-source chemogenomic data-driven algorithms for predicting drug–target interactions. Brief Bioinform. 2019;20:1465–74.
2. Madhukar NS, Khade PK, Huang L, Gayvert K, Galletti G, Stogniew M, et al. A Bayesian machine learning approach for drug target identification using diverse data types. Nat Commun. 2019;10:5221.
3. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3:673–83.
4. Liang S, Yu H. Revealing new therapeutic opportunities through drug target prediction: a class imbalance-tolerant machine learning approach. Bioinformatics. 2020;36:4490–7.
5. Lu Y, Guo Y, Korhonen A. Link prediction in drug-target interactions network using similarity indices. BMC Bioinform. 2017;18:39.
6. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. Brief Bioinform. 2020;22:247.
7. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18:41–58.
8. Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, et al. Drug–target interaction prediction: databases, web servers and computational models. Brief Bioinform. 2016;17:696–712.
9. Yue Y, He S. DTI-HeNE: a novel method for drug-target interaction prediction based on heterogeneous network embedding. BMC Bioinform. 2021;22:418.
10. Thafar MA, Olayan RS, Ashoor H, Albaradei S, Bajic VB, Gao X, et al. DTiGEMS+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques. J Cheminform. 2020;12:44.
11. Yao L, Evans JA, Rzhetsky A. Novel opportunities for computational biology and sociology in drug discovery: corrected paper. Trends Biotechnol. 2010;28:161–70.
12. Kaushik AC, Mehmood A, Dai X, Wei D-Q. A comparative chemogenic analysis for predicting drug-target pair via machine learning approaches. Sci Rep. 2020;10:6870.
13. Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nat Commun. 2017;8:573.
14. Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. Bioinformatics. 2018;34:1164–73.

15. Pliakos K, Vens C. Drug-target interaction prediction with tree-ensemble learning and output space reconstruction. BMC Bioinform. 2020;21:49.
16. Ding H, Takigawa I, Mamitsuka H, Zhu S. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. Brief Bioinform. 2014;15:734–47.
17. Yue X, Wang Z, Huang J, Parthasarathy S, Moosavinasab S, Huang Y, et al. Graph embedding on biomedical networks: methods, applications and evaluations. Bioinformatics. 2020;36:1241–51.
18. Celebi R, Uyar H, Yasar E, Gumus O, Dikenelli O, Dumontier M. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. BMC Bioinform. 2019;20:726.
19. Zhong X, Rajapakse JC. Graph embeddings on gene ontology annotations for protein–protein interaction prediction. BMC Bioinform. 2020;21:560.
20. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In ICLR Workshop Papers, 2013.
21. Mohamed SK, Nováček V, Nounu A. Discovering protein drug targets using knowledge graph embeddings. Bioinformatics. 2020;36:603–10.
22. Ji B-Y, You Z-H, Jiang H-J, Guo Z-H, Zheng K. Prediction of drug-target interactions from multi-molecular network based on LINE network representation method. J Transl Med. 2020;18:347.
23. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18:463–77.
24. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J, et al. Toward more realistic drug–target interaction predictions. Brief Bioinform. 2015;16:325–37.
25. Cardoso-Silva J, Papageorgiou LG, Tsoka S. Network-based piecewise linear regression for QSAR modelling. J Comput Aided Mol Des. 2019;33:831–44.
26. Cardoso-Silva J, Papadatos G, Papageorgiou LG, Tsoka S. Optimal piecewise linear regression algorithm for QSAR modelling. Mol Inform. 2019;38:e1800028.
27. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. Bioinformatics. 2008;24:i232–40.
28. Perlman L, Gottlieb A, Atias N, Ruppin E, Roded S. Combining drug and gene similarity measures for drug-target elucidation. J Comput Biol. 2011;18:133–45.
29. Liu Y, Wu M, Miao C, Zhao P, Li X-L. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLoS Comput Biol. 2016;12:e1004760.
30. Hattori M, Okuno Y, Goto S, Kanehisa M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. J Am Chem Soc. 2003;125:11853–65.
31. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucl Acids Res. 2012;40:D1100–7.
32. Lenselink EB, ten Dijke N, Bongers B, Papadatos G, van Vlijmen HWT, Kowalczyk W, et al. Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. J Cheminform. 2017;9:45.
33. Wang Z, Liang L, Yin Z, Lin J. Improving chemical similarity ensemble approach in target prediction. J Cheminform. 2016;8.
34. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci. 2002;42:1273–80.
35. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform. 2015;7:20.
36. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. J Cheminform. 2011;3:33.
37. Vilar S, Uriarte E, Santana L, Lorberbaum T, Hripcsak G, Friedman C, et al. Similarity-based modeling in large-scale prediction of drug-drug interactions. Nat Protoc. 2014;9:2147–63.
38. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008;4:682–90.
39. Xiao N, Cao D-S, Zhu M-F, Xu Q-S. protr/ProtrWeb: R package and web server for generating various numerical representation schemes of protein sequences. Bioinformatics. 2015;31:1857–9.
40. Hao M, Bryant SH, Wang Y. Predicting drug-target interactions by dual-network integrated logistic matrix factorization. Sci Rep. 2017;7:1–11.
41. Alaimo S, Pulvirenti A, Giugno R, Ferro A. Drug-target interaction prediction through domain-tuned network-based inference. Bioinformatics. 2013;29:2004–8.
42. Grover A, Leskovec J. node2vec: scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, USA, 2016, 855–864.
43. Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: a survey. Knowl Based Syst. 2018; 151:78-94.
44. Chen Z-H, You Z-H, Guo Z-H, Yi H-C, Luo G-X, Wang Y-B. Predicting drug-target interactions by Node2vec node embedding in molecular associations network. In: Intelligent computing theories and application: 16th international conference, ICIC 2020, Bari, Italy, October 2–5, 2020, Proceedings, Part II. Berlin: Springer; 2020. p. 348–58.
45. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008;2008:P10008.
46. Hagberg A, Swart P, S Chult D. Exploring network structure, dynamics, and function using NetworkX. Los Alamos: Los Alamos National Lab (LANL); 2008.
47. Maćkiewicz A, Ratajczak W. Principal components analysis (PCA). Comput Geosci. 1993;19:303–42.
48. Mathai N, Chen Y, Kirchmair J. Validation strategies for target prediction methods. Brief Bioinform. 2020;21:791–802.
49. Baumann D, Baumann K. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. J Cheminform. 2014;6:47.
50. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucl Acids Res. 2011;39(suppl_2):W270–7.

51. Zhang Z, Zhou L, Xie N, Nice EC, Zhang T, Cui Y, et al. Overcoming cancer therapeutic bottleneck by drug repurposing. Sig Transduct Target Ther. 2020;5:1–25.
52. Guha M. Imbruvica—next big drug in B-cell cancer—approved by FDA. Nat Biotechnol. 2014;32:113–4.
53. Parmar S, Patel K, Pinilla-Ibarz J. Ibrutinib (Imbruvica): a novel targeted therapy for chronic lymphocytic leukemia. Pharm Ther. 2014;39:483–519.
54. Löb S, Königsrainer A, Rammensee H-G, Opelz G, Terness P. Inhibitors of indoleamine-2,3-dioxygenase for cancer therapy: can we see the wood for the trees? Nat Rev Cancer. 2009;9:445–52.
55. Hornyák L, Dobos N, Koncz G, Karányi Z, Páll D, Szabó Z, et al. The role of indoleamine-2,3-dioxygenase in cancer development, diagnostics, and therapy. Front Immunol. 2018;9.
56. Uyttenhove C, Pilotte L, Théate I, Stroobant V, Colau D, Parmentier N, et al. Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2,3-dioxygenase. Nat Med. 2003;9:1269–74.
57. Moon YW, Hajjar J, Hwu P, Naing A. Targeting the indoleamine 2,3-dioxygenase pathway in cancer. J Immunother Cancer. 2015;3:51
58. Mahmood N, Mihalcioiu C, Rabbani SA. Multifaceted role of the urokinase-type plasminogen activator (uPA) and its receptor (uPAR): diagnostic, prognostic, and therapeutic applications. Front Oncol. 2018;8:24.
59. Tang L, Han X. The urokinase plasminogen activator system in breast cancer invasion and metastasis. Biomed Pharmacother. 2013;67:179–82.
60. Matthews H, Ranson M, Kelso MJ. Anti-tumour/metastasis effects of the potassium-sparing diuretic amiloride: an orally active anti-cancer drug waiting for its call-of-duty? Int J Cancer. 2011;129:2051–61.
61. Lang L. FDA Approves sorafenib for patients with inoperable liver cancer. Gastroenterology. 2008;134:379.
62. Segers VFM, Dugaucquier L, Feyen E, Shakeri H, De Keulenaer GW. The role of ErbB4 in cancer. Cell Oncol. 2020;43:335–52.
63. Mullooly M, McGowan PM, Kennedy SA, Madden SF, Crown J, O'Donovan N, et al. ADAM10: a new player in breast cancer progression? Br J Cancer. 2015;113:945–51.
64. Wei W, Zou Y, Jiang Q, Zhou Z, Ding H, Yan L, et al. PSMB5 is associated with proliferation and drug resistance in triple-negative breast cancer. Int J Biol Mark. 2018;33:102–8.
65. Hui DY. Group 1B phospholipase A2 in metabolic and inflammatory disease modulation. Biochim Biophys Acta Mol Cell Biol Lipids. 2019;1864:784–8.
66. Lai SH, Zervoudakis G, Chou J, Gurney ME, Quesnelle KM. PDE4 subtypes in cancer. Oncogene. 2020;39:3791–802.
67. Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, et al. Immune evasion in cancer: mechanistic basis and therapeutic strategies. Semin Cancer Biol. 2015;35:S185–98.
68. Waldhauer I, Goehlsdorf D, Gieseke F, Weinschenk T, Wittenbrink M, Ludwig A, et al. Tumor-associated MICA is shed by ADAM proteases. Cancer Res. 2008;68:6368–76.
69. Raulet DH. Roles of the NKG2D immunoreceptor and its ligands. Nat Rev Immunol. 2003;3:781–90.
70. Doubrovina ES, Doubrovin MM, Vider E, Sisson RB, O'Reilly RJ, Dupont B, et al. Evasion from NK cell immunity by MHC Class I chain-related molecules expressing colon adenocarcinoma. J Immunol. 2003;171:6891–9.
71. Condon JS, Joseph-McCarthy D, Levin JI, Lombart HG, Lovering FE, Sun L, Wang W, Xu W, Zhang Y. Identification of potent and selective TACE inhibitors via the S1 pocket. Bioorg Med Chem Lett. 2006;17:34–9.
72. Tochowicz A, Goettig P, Evans R, Visse R, Shitomi Y, Palmisano R, et al. The dimer interface of the membrane type 1 matrix metalloproteinase hemopexin domain: crystal structure and biological functions*. J Biol Chem. 2011;286:7587–600.
73. Zhou D, Ren K, Wang J, Ren H, Yang W, Wang W, et al. Erythropoietin-producing hepatocellular A6 overexpression is a novel biomarker of poor prognosis in patients with breast cancer. Oncol Lett. 2018;15:5257–63.
74. Wilmes LJ, Pallavicini MG, Fleming LM, Gibbs J, Wang D, Li K-L, et al. AG-013736, a novel inhibitor of VEGF receptor tyrosine kinases, inhibits breast cancer growth and decreases vascular permeability as detected by dynamic contrast-enhanced magnetic resonance imaging. Magn Resonan Imaging. 2007;25:319–27.
75. Rini B, Rixe O, Bukowski R, Michaelson M, Wilding G, Hudes G, et al. AG-013736, a multi-target tyrosine kinase receptor inhibitor, demonstrates anti-tumor activity in a Phase 2 study of cytokine-refractory, metastatic renal cell cancer (RCC). J Clin Oncol. 2005;23:4509–4509.
76. Rugo HS, Herbst RS, Liu G, Park JW, Kies MS, Steinfeldt HM, et al. Phase I trial of the oral antiangiogenesis agent AG-013736 in patients with advanced solid tumors: pharmacokinetic and clinical results. J Clin Oncol. 2005;23(24):5474-83.
77. Park K, Tan E-H, O'Byrne K, Zhang L, Boyer M, Mok T, et al. Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label, randomised controlled trial. Lancet Oncol. 2016;17:577–89.
78. Lin NU, Winer EP, Wheatley D, Carey LA, Houston S, Mendelson D, et al. A phase II study of afatinib (BIBW 2992), an irreversible ErbB family blocker, in patients with HER2-positive metastatic breast cancer progressing after trastuzumab. Breast Cancer Res Treat. 2012;133:1057–65.
79. Ali BM, Zaitone SA, Shouman SA, Moustafa YM. Dorzolamide synergizes the antitumor activity of mitomycin C against Ehrlich's carcinoma grown in mice: role of thioredoxin-interacting protein. Naunyn-Schmiedeberg's Arch Pharmacol. 2015;388:1271–82.
80. Boyd AW, Bartlett PF, Lackmann M. Therapeutic targeting of EPH receptors and their ligands. Nat Rev Drug Discov. 2014;13:39–62.

Publisher's Note