

RESEARCH

Open Access



Drug repurposing and prediction of multiple interaction types via graph embedding

E. Amiri Sourì¹, A. Chenoweth^{2,3}, S. N. Karagiannis^{2,3} 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

² St. John's Institute
of Dermatology, School of Basic
and Medical Biosciences, Guy's
Hospital, King's College London,
London SE1 9RT, UK

³ Breast Cancer Now Research
Unit, School of Cancer
and Pharmaceutical Sciences,
Guy's Cancer Centre, King's
College London, London SE1
9RT, UK

Abstract

Background: Finding drugs that can interact with a specific target to induce a desired therapeutic outcome is key deliverable in drug discovery for targeted treatment. Therefore, both identifying new drug–target links, as well as delineating the type of drug interaction, are important in drug repurposing studies.

Results: A computational drug repurposing approach was proposed to predict novel drug–target interactions (DTIs), as well as to predict the type of interaction induced.

The methodology is based on mining a heterogeneous graph that integrates drug–drug and protein–protein similarity networks, together with verified drug–disease and protein–disease associations. In order to extract appropriate features, the three-layer heterogeneous graph was mapped to low dimensional vectors using node embedding principles. The DTI prediction problem was formulated as a multi-label, multi-class classification task, aiming to determine drug modes of action. DTIs were defined by concatenating pairs of drug and target vectors extracted from graph embedding, which were used as input to classification via gradient boosted trees, where a model is trained to predict the type of interaction. After validating the prediction ability of DT2Vec+, a comprehensive analysis of all unknown DTIs was conducted to predict the degree and type of interaction. Finally, the model was applied to propose potential approved drugs to target cancer-specific biomarkers.

Conclusion: DT2Vec+ showed promising results in predicting type of DTI, which was achieved via integrating and mapping triplet drug–target–disease association graphs into low-dimensional dense vectors. To our knowledge, this is the first approach that addresses prediction between drugs and targets across six interaction types.

Keywords: Drug–target interaction, Drug discovery, Drug repurposing, Network embedding, Machine learning

Introduction

Drug discovery remains a time-consuming and costly process, with low success rate [1, 2]. Early studies on drug effects through trial-and-error procedures have been superseded by advances in chemical genomic research allowing complex analyses and insight in mechanisms of drug actions [3]. Due to the intrinsic complexity of molecular structures, the vast majority of drugs tend to interact with multiple targets either in a unique disease pathway or across multiple pathways [4], referred to as polypharmacology [5].



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

This observation has changed the drug design philosophy of “one gene, one drug, one disease” paradigm and fuelled consecutive work in drug discovery [5]. Reaping the effects of polypharmacological approaches to discover unknown off-targets for approved drugs and allow existing drugs to be applied for the treatment of another indication, drug repurposing has emerged as a promising route to drug discovery [6].

Drug repurposing can offer a shorter path in drug discovery, bypassing several steps in drug development [2, 7], and can improve understanding of drug mechanisms of action and drug side effects [8]. Drug repurposing strategies have shown successful applications in the past [1], with potential advantages and approved examples outlined in relevant reviews (for example see papers by Pushpakom et al. [9] and Ashburn et al. [10]). The first main step is finding possible drug–target interactions (DTIs) outside the original scope of a drug [9], which it is a crucial task in drug discovery. In addition to predicting the presence of DTIs, it is also important to determine the type of interaction, as different types of interactions can have varying therapeutic outcomes by increasing or decreasing expression and reaction [11]. Drug development relies heavily upon targeting “druggable” genes, which include more than 4500 genes in the human genome [12, 13]. However, only a small number of these genes are targeted by approved drugs and many interactions between genomic and chemical spaces remain unknown [14]. Although some DTIs have been identified via serendipitous or rational observations, they are not efficient in practice due to the vast search space that needs to be covered in lab experimentation [15]. These limitations strengthen the case for computational investigation of DTIs, as means of narrowing the search space and proposing the most promising cases on which to focus drug development.

Moreover, in pathological processes of complex diseases, multiple genes and pathways may be implicated [16, 17]. Targeting druggable genes (known as targeted therapy) is a key strategy in treating complex diseases, including cancer [18]. Recently, research on drug repurposing has shown that disease-specific gene biomarkers, as well as associations between drugs and diseases, can be used to accurately predict drugs for new diseases through identifying novel off-target interactions [11, 19–21]. However, most of the existing studies focused on either drug-disease or drug–target interactions as two isolated tasks and did not take account of the relationships between these [18]. Therefore, systematic integration of relationships represented in gene, disease, and drug networks can lead to more reliable new DTI prediction [2, 22]. Additionally, these developments can significantly reduce and refine experimental laboratory costs and processes, as well as the risk of failure in drug development [11].

In terms of methodological avenues to predicting DTIs, traditionally molecular docking and ligand-based methods have been used [23]. However, docking-based methods rely on the availability of protein 3D structural information, which can be challenging for large-scale prediction [7, 22], while ligand-based methods are not accurate in cases where only a small number of known binding ligands are available [24]. Recently, machine learning-based (ML) methods have attracted much interest in drug repurposing, due to their ability to analyze large numbers of DTIs efficiently by extracting latent association patterns [19, 25]. A wide variety of ML methods in drug repurposing have been proposed and summarized previously [26–28].

Similarity-based methods rely on the key underlying assumption that similar drugs developed for similar diseases tend to target similar proteins and constitute the most widely applied strategy to integrate biological networks [7, 29]. These methods can integrate different large-scale associations of genes, diseases, and drugs into a heterogeneous graph [30, 31], and then formulate DTI prediction as a **link prediction** problem in graph analysis [14, 18]. The “2vec” (short for “to vector”) methods form an important category of embedding methods (e.g. “graph2vec”, “node2vec” etc.) and have shown promising potential in representing input features in ML tasks by mapping graph structural properties to low dimensional vectors [32]. Recently, different embedding methods have been proposed for predicting new DTIs [3, 18, 25, 32, 33]. Although these methods achieved promising results, they can only predict binary interactions between drugs and target proteins and cannot identify the type of interaction (i.e. *action* type such as activation, expression, reaction, etc.) or the effect of a drug on its target (i.e. *degree* type such as increase/decrease in expression, etc.), which would be essential in the process of targeted treatment and in understanding drug action for drug repurposing [11, 21].

In targeted therapy for cancer [34], the current therapeutic debacle is discovering and targeting cancer biomarkers. Carcinogenesis is the result of mutations in oncogenes and/or tumour suppressor genes [35]. Molecular targeted therapies that inhibit oncogenes and/or activate tumour suppressor gene products can limit or stop tumour progression [36]. Despite efforts in pharmaceutical research to develop new drugs to target these genes, the implementation of discovered drugs in clinical practice has lagged far behind expectations. Therefore, drug repurposing should be considered as a promising strategy to address the unmet need for efficacious cancer therapies. One of the main advantages of this approach is the availability of toxicity, pharmacodynamic, pharmacokinetic profiles for these drugs have already been established [37].

In this work, we propose an ML-based computational pipeline for drug repurposing, DT2Vec+, that integrates the triplet associations of drug–target–disease data to a heterogeneous graph by incorporating drug–drug and protein–protein similarity networks, together with verified drug–disease and protein–disease associations [19, 38]. To extract features from the network, the three-layer heterogeneous graph was mapped to low dimensional vectors using **node2vec**. The DTI prediction problem was formulated as a multi-label, multi-class classification to determine drug modes of action, defined as “increases^expression”, “decreases^expression”, “decreases^reaction”, “increases^reaction”, “increases^activity”, “decreases^activity”. In previous work, Wang and Zeng [11] proposed an ML-based model using Boltzmann machines to predict three types of DTIs (binding, activation and inhibition), so to our knowledge DTVec+ is the first method investigating six drug–target interaction degrees and types. DTIs were defined by concatenating pairs of drug and target vectors extracted from graph embedding, which were used as input to gradient boosted trees (XGBoost) to train a model for interaction type prediction. Cross-validation was used to evaluate performance and a comprehensive analysis of unknown DTIs was conducted to evaluate results in terms of the degree and type of interaction. Our results have also been appraised in terms of case studies for potential drugs that target important oncogenes and their medical potential is discussed.

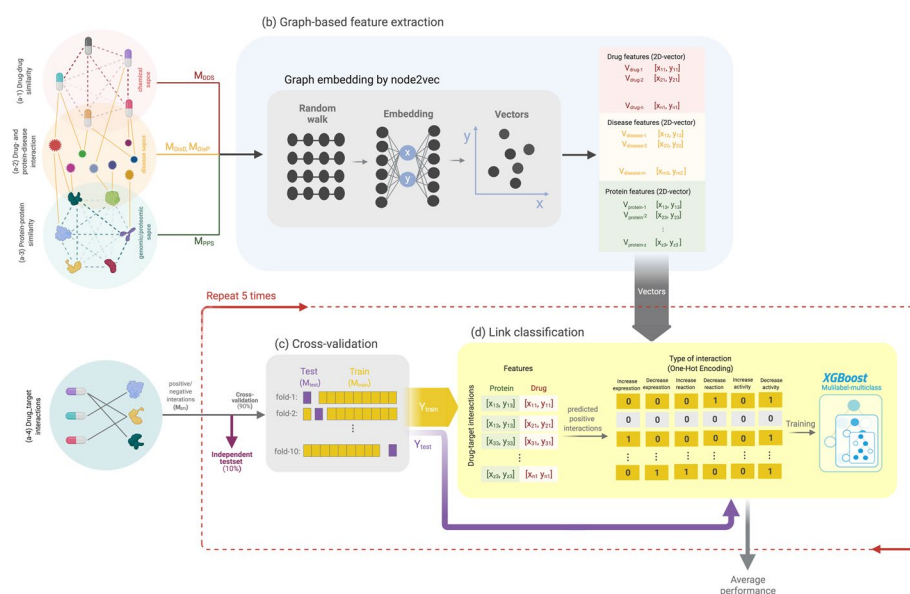


Fig. 1 DT2Vec+ pipeline. **(a-1,2,3)** integrating drug–drug (DDS) and protein–protein (PPS) similarity graphs with drug–disease (DDis) and disease–protein (DisP) association graph as input of embedding method to low dimensional vectors. **(a-4)** Drug–target interaction graph with different edge types. **b** Graph-embedding developed by node2vec to map nodes to vectors (in this figure, nodes are shown mapped to 2D-vector, x and y). **c** Known drug–target interactions (six types of interactions) 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 labelled using one-hot encoding and an XGBoost model was trained on each label using cross-validation. 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

Materials and methods

Figure 1 presents an overview of the computational framework in DT2Vec+, which includes network integration, feature extraction, implementation of the proposed methodology and evaluation using cross-validation. Categorical labels for degree and type of interaction were converted to binary vectors through one-hot encoding, and one-vs-rest strategies were used to train the model against each label. The average performance of all models was measured on external test-sets.

Dataset and similarity networks

Drug–target interactions (M_{DTI} , Fig. 1: a-4), disease–protein (M_{DisP}), and disease–drug (M_{DisD}) associations (Fig. 1: a-2) were extracted from the Comparative Toxicogenomics Database (CTD) [38]. Since only drug MeSHid was provided in the Comparative Toxicogenomics Database (CTD), we used DrugBank [39] and ChEMBL [40] datasets to find drug SMILES and phase which shows molecule connectivity and chirality, and trials testing stage of drug respectively. Additional file 1: Fig. S1 shows steps performed to collect the dataset. Only approved phase-4 drugs with therapeutic evidence were selected for further repurposing analysis. Table 1 shows details of the selected dataset.

The CTD dataset categorised DTIs into different degrees and types of interaction based on published references as positive interactions. Selecting experimentally validated negative DTIs is an important point in developing an accurate model for DTI

Table 1 Dataset details

Dataset	Frequency
Protein	1141
Drug	280
Disease	589
Disease-protein interaction	1305
Disease-drug interaction	301
Drug–target (protein) interaction	4086
Drug–target pairs	3502

prediction. Therefore, the ChEMBL dataset was used to select 348 validated negative DTIs labelled as ‘inactive’ [41]. The known DTIs named ‘development dataset’ which is used for developing the ML model, comprise 4086 interactions between 3502 drug–target pairs assigned to six categories as labels, namely “increases^expression” (1392), “decreases^expression” (708), “decreases^reaction” (158), “increases^reaction” (1017), “increases^activity” (396), and “decreases^activity” (415). DTIs have one (2922), two (576), or three (4) types of interactions. Importantly, cases with more than one interaction type were ‘multi-label’ and among these DTIs that have two interaction degrees with the same action type (e.g. increases^expression and decreases^expression) were considered as ambiguous, and removed from further analysis. All possible drug–target pairs without known interactions were defined as an ‘experimental dataset’ and used for drug repurposing analysis.

To calculate protein–protein similarity (PPS), M_{PPS} (Fig. 1: a-3) by sequence alignment [42], the sequences of target proteins were extracted from UniProt [43]. The parallelised version of protein similarity calculation was implemented using the “protr” package in R 3.3 [44]. Drug similarity measures were defined based on 166 structural fingerprints from canonical SMILES using MACCS [45]. Then, the drug–drug similarity (DDS) network, M_{DDS} (Fig. 1: a-1) was calculated based on the Tanimoto coefficient in the range of 0 to 1 [46] which was implemented using Open Babel [47] in Python 3.7.3 [48]. A triplet drug–target–disease association graph was generated by integrating four networks of M_{DisP} , M_{DisD} , M_{PPS} , and M_{DDS} by matching similar proteins and drugs on M_{PPS} and M_{DisP} and M_{DisP} and M_{DDS} respectively (Fig. 1a). The new graph consists of 692,177 edges between 2011 nodes (1141 proteins, 280 approved drugs and 589 diseases).

Network-based feature extraction using node2vec

Changing the format or shape of raw data to extract informative and discriminative features, known as feature extraction, is an important step for an effective ML model. Graph embedding methods can create powerful representation structural information by converting the topological properties of a heterogeneous network to a set of features in low-dimensional space [25, 49] that can be used as input to a predictive model. In this work, we used node2vec [50], a neural network-based node embedding method to automatically map nodes in the drug–target–disease association graph into a 100-dimensional vector which was reported as the best threshold to accurately preserve graph information [25, 49–51]. Recent machine learning research has shown that node2vec is a

superior method for node embedding compared to other existing state-of-the-art methods [51, 52]. In a recent study, node2vec demonstrated encouraging outcomes in predicting drug–target interactions (DTIs) by converting drug, protein, disease, lncRNA and miRNA association networks into vectors [49, 50]. Figure 1b illustrates the embedding process, where drug (V_{drug}), protein (V_{protein}) and diseases (V_{disease}) nodes are mapped to two-dimensional vectors, as an example. Node2vec was implemented in python 2 using publicly available GitHub source code (<https://github.com/aditya-grover/node2vec>) [50].

Data-splitting, cross-validation, and performance evaluation metrics

Various validation methods may be used for drug–target interaction prediction models [53]. Among these, cross-Validation (CV) methods are preferred due to their robustness in estimating how a model generalizes. Validation was conducted on internal and external testing as follows. All known DTIs (M_{DTI}) in the ‘*development dataset*’ were split into 80% training and 10% validation set to train and select the best model. The external testset (i.e. the remaining 10% of the data) was used to evaluate the performance of the model and was blind to the process of developing the model. We applied a tenfold CV (Fig. 1c) which involved randomly splitting the data into 10 partitions and iteratively selecting each partition as the testing data and training the model on the remaining partitions and this procedure was repeated five times.

Choosing the right metrics for evaluating the performance of the model is important, but can be challenging depending on the underlying assumptions. The aim of the DTI prediction model is to report positive interactions among all unknown drug–target pairs, so that highly positive interactions can be validated experimentally, therefore a low false-positive rate is desirable. In this case, the Precision metric (true positive/(true positive + false positive)) can better reflect performance. However, in general a DTI prediction model should have high precision without sacrificing other metrics. Therefore, to evaluate the performance of the model, average Precision, f1-score, and accuracy across all cross-validation sets on external test sets were calculated.

Multi-label, multi-class ML-based link classification

DTIs were represented as a 200-dimensional vector based on a strategy of concatenating the drug and target embedding features [32] and labelled based on six types of interaction, in terms of affecting expression, reaction rate or binding affinity (increases^{expression}, decreases^{expression}, decreases^{reaction}, increases^{reaction}, increases^{activity}, and decreases^{activity}). The DTI prediction problem was formulated as a multi-label, multi-class link classification built on XGBoost [54]. 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 [12]. Moreover, by taking advantage of XGBoost in returning the prediction probability score, we were able to rank DTIs based on the confidence of the model. We used one-hot encoding to create six binary vector labels and different models were trained based on one-vs-rest strategies for each label type (Fig. 1d). Training data was balanced by over-sampling via Synthetic Minority Over-sampling Technique (SMOTE) [55]. Over-sampling is a technique used to balance sample numbers in imbalanced datasets by generating synthetic data for the minority class, and the number of samples increased in

each class depends on the specific implementation and the degree of imbalance [55]. 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 = r , subsample ratio = 1, gamma = 1, minimum child weight = 2, early stopping = 20 and learning-gamma rate = 0.01.

Novel DTIs extraction and drug repurposing

After validating the performance of the proposed method using cross-validation, in order to detect novel DTIs a two-step prediction was applied to unknown interactions (named as ‘experimental dataset’). First, experimentally validated negative interactions and DT2Vec [56] were used to predict highly positive interactions with a probability score $\geq 0.95\%$. DT2Vec is a machine learning pipeline that formulates the problem of deriving new drug–target interactions as binary (positive or negative) link prediction [56]. Then DT2Vec+ was applied on positive interactions to identify the six types of DTIs based on the triple association graph. A small number of the novel DTIs was analysed for target treatment. In this work, we also focus on several cancer gene biomarkers and establish a connection between our repurposing prediction results and potential role of the predicted DTI in cancer drug development.

Results

DT2Vec+ was developed based on concatenating drug–target pair vectors extracted from a heterogeneous association graph consisting of three node types of drug, protein and diseases, connected through four edge types (DDS, PPS, DDis, and DisP associations) as shown in Fig. 1a,b. Figure 2a shows Principal Component Analysis of mapped vectors associated with the three node types. DTIs were defined by concatenating each drug–target pair vector and labelling based on the type of interaction. Six models were trained for each type of DTI based on one-vs-rest strategy. Performance was assessed via tenfold cross-validation repeated five times as described previously, and average results for each label are shown in Table 2. When applied to external test sets, the proposed

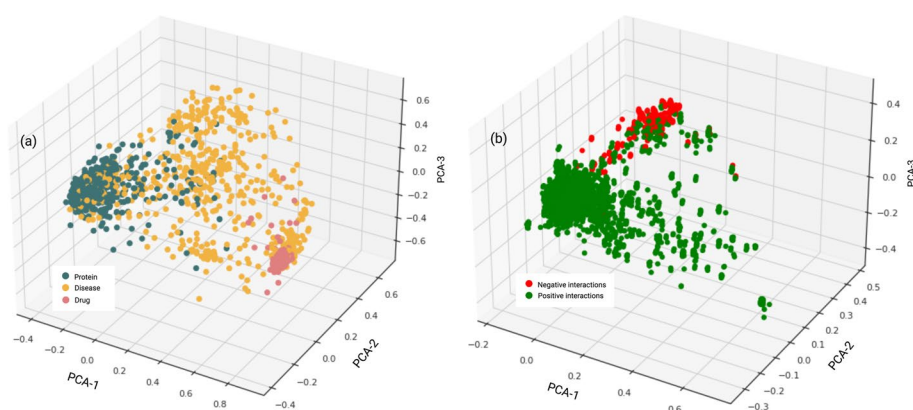


Fig. 2 Drugs, diseases, proteins and DTI visualisation. **a** PCA of drugs, diseases, and proteins vectors extracted from heterogeneous association graph. **b** PCA of DTIs mapped vectors by concatenating drug–target vectors

Table 2 DT2Vec+ performance metrics (variance shown in parentheses)

	Mean accuracy	Mean f1_score	Mean precision
increases^expression	73.35% (0.01)	73.34% (0.02)	73.27% (0.01)
decreases^expression	74.68% (0.01)	72.08% (0.02)	80.23% (0.01)
increases^reaction	77.2% (0.03)	71.84% (0.03)	91.78% (0.02)
decreases^reaction	77.8% (0.02)	76.68% (0.02)	80.79% (0.02)
increases^activity	75.63% (0.01)	70.26% (0.03)	89.95% (0.01)
decreases^activity	83.88% (0.02)	82.17% (0.03)	91.46% (0.01)
Total	77.09% (0.02)	74.39% (0.02)	84.58% (0.01)
Binary (active/inactive)	95.81% (0.02)	96% (0.02)	92.54% (0.03)

method achieved average Accuracy, f1-score, and Precision of 77.09% (0.02), 74.39% (0.02), and 84.58% (0.01), respectively.

After validating the performance of DT2Vec+, in order to extract new interaction types in unknown DTIs ('experimental dataset'), we performed a two-step prediction of binary (activation or inactive) and multi-class, multi-label (interaction type) classification. Figure 2b shows the PCA of positive and negative DTIs defined based on embedded vectors. In the first step, a model was trained to stratify positive and negative DTIs. The performance of the model on the external-test set is summarised in Table 2. This achieved higher than 90% on all metrics. Then, the model was applied to unknown DTIs to find positive interactions. Interactions with positive probability score of 95% or higher were selected as highly positive interactions which then were used as input of DTI2Vec+ to identify the type of interaction.

Figure 3 summarises labels of all drug–target pairs which take the form of either *known* (coloured red for positive and purple for negative) or *predicted* (coloured blue for positive and lilac for negative) interactions. DTIs with multiple labels were marked darker compared with the interactions with one label type. We predicted 18,736 and 787 DTIs with two and three label types, respectively. The top 20 new predicted DTIs for each interaction type are shown in Fig. 4. However, there were some highly positive interactions that did not belong to any of our six labels or which had more than three labels (ambiguous/other interactions), which we excluded from further analysis. For example, Amikacin, Desogestrel, and Astemizole may interact with multiple proteins and decrease the activity, expression, and reaction respectively while Carfilzomib, Hydralazine hydrochloride, and Butamben may increase the activity, expression, and reaction of multiple proteins. The novel high-scoring DTIs proposed by this method can narrow down the search space in a wet-laboratory experiment towards finding drugs able to target a specific protein.

Discussion

Drug repurposing holds promising potential in reducing failure risks and costs of developing new drugs. Finding appropriate drugs that can interact with a specific target is a pivotal step in drug repurposing strategies. However, due to the massive number of potential interactions, it is almost impossible to base drug discovery solely on wet-lab experiments without the help of computational methods and virtual screening which are

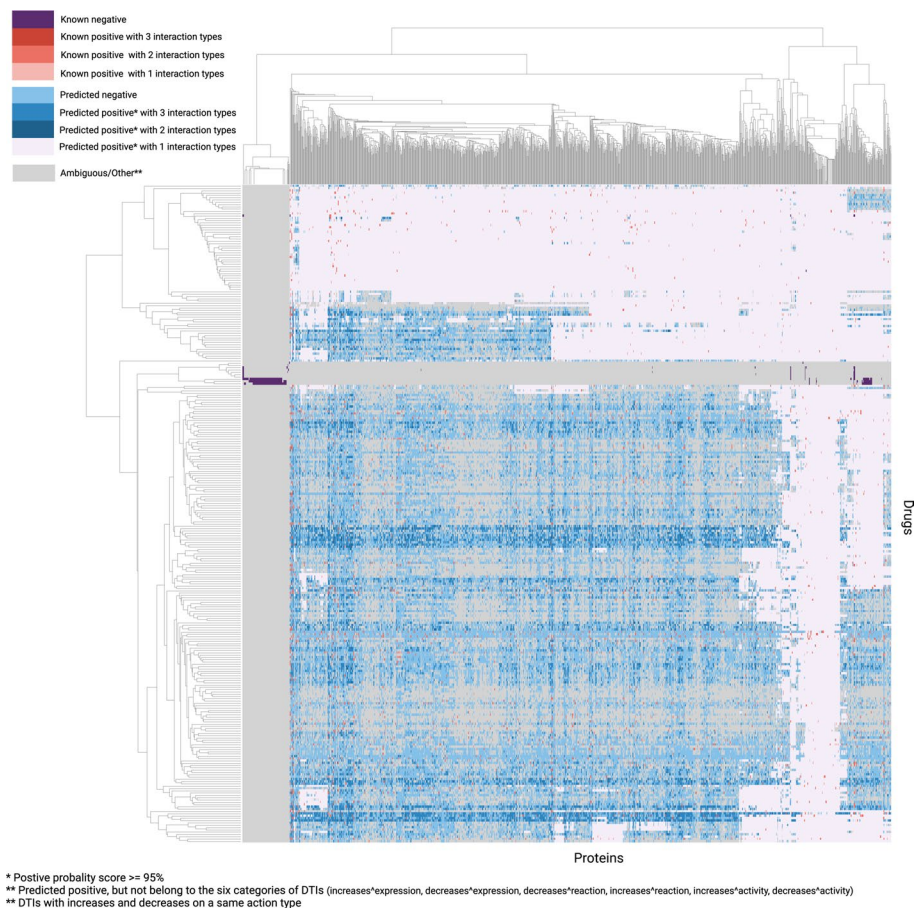


Fig. 3 Interaction type of all drug–target pairs. The heatmap shows the mapping of known DTI interactions (red and purple) and predicted interactions (blue and lilac). Each interaction can have multiple types of interactions, which were coloured darker

able to reduce the number of potential interactions for downstream experimental validation. In this work, we report an ML model, DT2Vec+, to predict the type of drug–target interaction based on features extracted from a heterogeneous drug–disease–target graph using an embedding approach. The problem of drug–target link prediction was formulated as multi-label, multi-class classification and the method was able to stratify DTIs into six different interaction types, with performance higher than 75% on average in test sets across validation metrics. Our pipeline was used as a tool for the identification of targeted treatments by selecting potentially targetable oncogenes and predicting some drug candidates that affect the activity of proteins associated with breast and other cancers, as discussed next.

BIRC5 (Survivin) has been reported to be an important biomarker for breast cancer [17, 57], with high expression of BIRC5 correlated with worse survival, and is a promising target for drug discovery and breast cancer therapeutics. BIRC5 encoding survivin is involved in carcinogenesis by influencing cell division and proliferation and inhibiting apoptosis [58]. Therefore, downregulation of BIRC5 can act as an inhibitor of tumour cell migration and invasion through the PI3K/Akt signaling pathway. In the CTD dataset, there are seven known DTIs where the drugs can decrease expression of BIRC5,

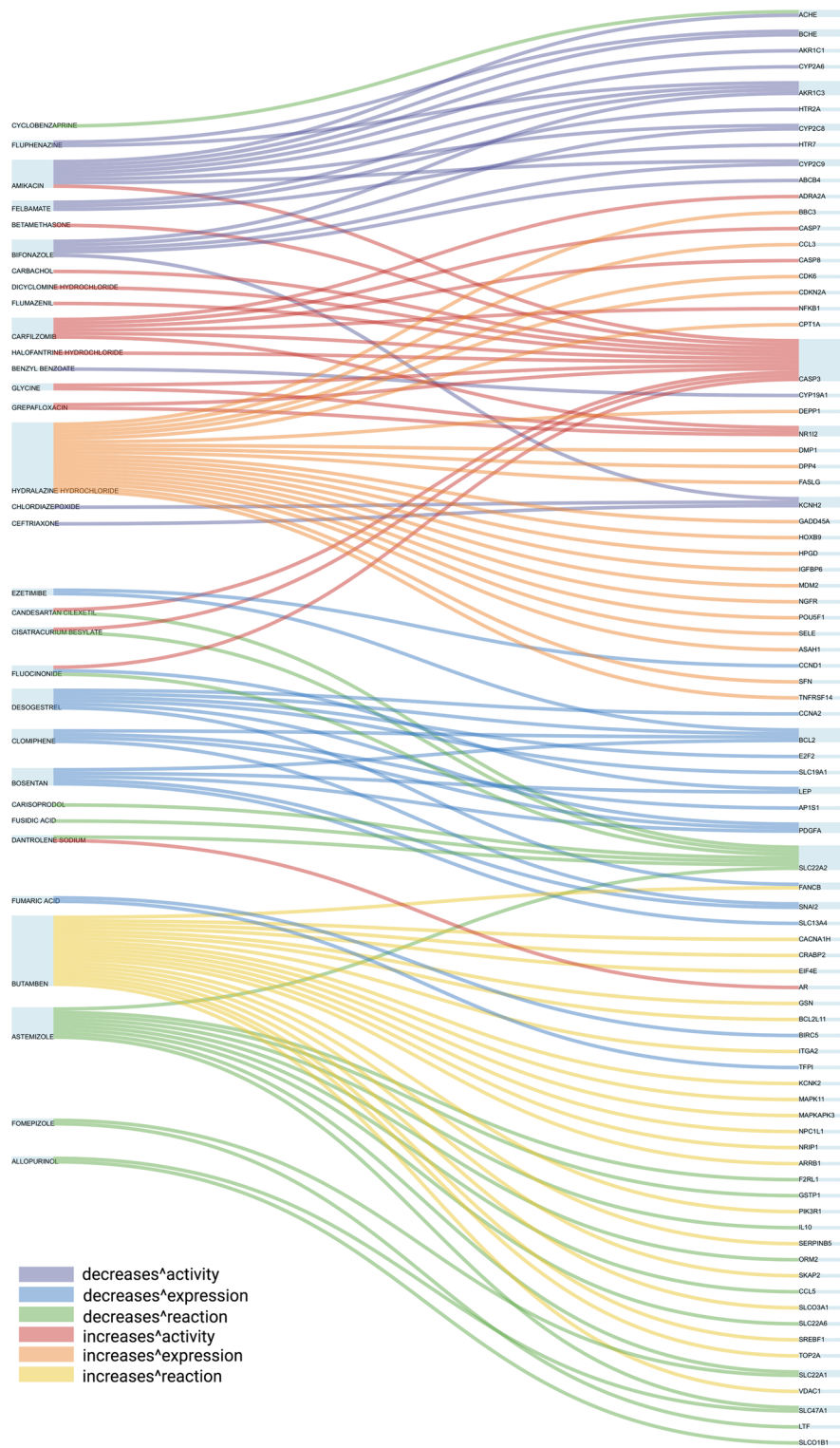


Fig. 4 Top novel drug–target interactions. Top 20 DTIs predicted using DT2Vec++ for each type of interaction coloured based on the type of interaction

namely Digitoxin (ChEMBL254219), Phenylbutanoic acid (ChEMBL1469), Bexarotene (ChEMBL1023), Ciclopirox (ChEMBL1413), Danthron (ChEMBL53418), Etodolac (ChEMBL622), Ethacrynic acid (ChEMBL456). Based on DT2Vec+ prediction results, Aminolevulinic acid (ChEMBL601), Cladribine (ChEMBL1619), Fumaric acid (ChEMBL503160), and Bosentan (ChEMBL957) were among the top drugs predicted to decrease the expression of BIRC5. Aminolevulinic acid and Cladribine have been investigated for cancer treatment in several clinical trials. Recently, a study investigating the relationship between survivin expression and melanoma after using Aminolevulinic acid treatment showed inhibition of melanoma growth in mice by downregulating survivin expression, which prolonged the survival of melanoma-bearing mice [59]. Aminolevulinic acid photodynamic therapy also induced apoptosis in cervical cancer cells *in vitro* and *in vivo* by decreasing survivin expression [60]. Cladribine also inhibited cell proliferation and induced apoptosis in multiple myeloma cells *in vitro* [61]. Cladribine also inhibited cell proliferation and induces apoptosis in multiple myeloma cells *in vitro* [62, 63]. Cladribine has been shown to increase progression-free survival and median time to second treatment in chronic lymphocytic leukaemia patients [61]. Finally, Bosentan has been reported to inhibit breast carcinoma metastasis to bone tissue in a skinfold chamber model [64].

BCL-2 is a key protein regulator of apoptosis and is overexpressed in many cancer types [65]. BCL-2 has been reported to be frequently expressed in breast cancer [66] and can reduce the effectiveness of chemotherapy [67]. There is an inverse correlation between the expression of BCL-2 and mutated p53, an important tumour suppressor gene, which in turn leads to increased cell proliferation rates and poor outcomes [66]. Targeted therapy against BCL-2 may improve the effects of chemotherapy in breast cancer patients [66]. Clioquinol (ChEMBL497), Adenosine (ChEMBL477), Diacerein (ChEMBL41286), Azelaic Acid (ChEMBL1238), Dequalinium (ChEMBL333826), Azelastine (ChEMBL639) and Bazedoxifene (ChEMBL46740) are DT2Vec+ predicted drugs that might be able to decrease the expression of BCL-2. Research showed that Clioquinol reduced the viability of eight different human cancer cell lines by inducing cell death through apoptotic pathways [68]. It also induced autophagy in leukaemia and multiple myeloma cell lines. Downregulated expression of BCL-2 can inhibit the interaction between Beclin 1 and BCL-2 and stimulate autophagy [69]. Adenosine induces cell cycle arrest and apoptosis in ovarian cancer cell lines by down-regulating BCL-2 [70]. Adenosine also induced apoptosis in different cancer types such as breast [71], leukaemia [72], gastric [73], colon [74], melanoma [75], and head and neck cancer cell lines, and suppressed BCL-2 expression [75].

Diacerein is another approved drug that has been reported to exert anti-proliferative effects on breast cancer cell lines, induced apoptosis and decreased the expression of BCL-2 [76]. Azelaic acid has cytotoxic action on many tumour cells and antileukemic activity in different types of acute myeloid leukaemia cells. It increases Notch expression which leads to the loss of BCL-2 expression [77]. Recently, Dequalinium showed promising results *in vitro* and *in vivo*, inhibiting the growth and proliferation of human glioma cells by decreasing BCL-2 expression [78]. Dequalinium injection into tumour-bearing mice inhibited the growth of human colon cancer cells [79]. The drug was able to prolong the survival of mice with bladder carcinoma cells [80], and inhibit the growth,

migration and invasion of melanoma cells *in vitro* [81]. Dequalinium also reduced acute myeloid leukaemia cell activity, proliferation, induced apoptosis and increased survival of rats with ovarian cancer [82]. *In vitro* and *in vivo* analyses showed that Azelastine could decrease levels of BCL-2 and inhibit colorectal cancer cell proliferation [83]. Bazedoxifene has been used in clinical trials for treating pancreatic and breast cancer. Bazedoxifene is now being repositioned as a new strategy for treating multiple cancer types (such as breast cancer, pancreatic cancer, colon cancer, etc.) by downregulating anti-apoptotic proteins such as BCL-2 [84].

MYC is involved broadly in many cancer types and its expression was estimated to be deregulated in up to 70% of human cancers [85]. High levels of MYC were linked to aggressive prostate cancer and triple-negative breast cancer. Two FDA approved drugs, Dihydroergotamine (ChEMBL1732) and Indinavir Sulfate (ChEMBL1735), which are predicted by the DT2Vec+, might be able to target MYC and decrease its expression. In a mouse xenograph model, Dihydroergotamine could suppress the growth of MYC-dependent human acute myeloid leukaemia, and in this study, MYC was the most statistically repressed gene by Dihydroergotamine [86]. Indinavir blocked tumour formation in an angiogenic tumour model, and this inhibition was associated with inhibition of cell invasion but not cell proliferation or cell survival. Indinavir sulfate was also effective at inhibiting the growth of various human tumour xenografts, including lung, breast, hepatocarcinoma and colon adenocarcinoma, and human tumours of haematopoietic cell origin. The drug effectively blocked the invasion of a basement membrane by lung, breast, colon adenocarcinoma [87].

Finally, STAT3 is constitutively activated in more than 40% of breast cancers and is thought to promote breast tumour progression [88]. Therefore, drugs that can reduce the activity of STAT3 have been attracting more attention. Using the DT2Vec+ methodology, four drugs were proposed to target STAT3, namely, Amsacrine (ChEMBL43), Phenylbutanoic Acid (ChEMBL1469), Doxazosin (ChEMBL707) and Capecitabine (ChEMBL1773). Amsacrine is an anti-cancer drug that showed significant activity against human acute leukaemia and it is currently approved for treatment [7]. Phenylbutanoic Acid has been reported to be able to inhibit cell proliferation by inducing apoptosis, cell cycle arrest, and senescence in colon, gastric and breast cancers. However, clinical trials with Phenylbutanoic acid in solid tumours showed no obvious benefit [89]. Doxazosin suppressed the growth of ovarian carcinoma cells and additively enhanced apoptotic cell death by IFN treatment and its effects were potentiated by reducing phosphorylation of STAT3 [90]. Doxazosin also could significantly inhibit prostate and bladder cancer cell growth *in vitro* [91, 92]. Capecitabine is a currently approved chemotherapy drug, and a meta-analysis of clinical trials of patients with triple-negative breast cancer treated with capecitabine in combination with neoadjuvant or adjuvant chemotherapy demonstrated improved survival [93]. Studies in triple-negative breast cancer patient-derived xenograft (PDX) models also have shown Capecitabine as an efficient chemotherapy agent [94].

Conclusion

In silico prediction of DTIs is an efficient approach for drug repurposing. Various methods have been proposed to predict the interactions between drugs and targets in a binary format (active or inactive), but determining drug mode of action has remained elusive. Importantly, deciphering the type of interaction in targeted treatment is an important step in developing an effective method. In this work, we reported DT2Vec+, which—to our knowledge—is the first ML-based framework to predict six types of DTIs, by integrating associations between drugs, diseases, and proteins into a heterogeneous graph consisting of DDS, PPS, DDis, and DisP edges. The triple association graph was mapped to low dimensional vectors using graph embedding, and DTIs were defined based on concatenating drug–target vectors. We show that this pipeline achieved high performance on external test sets and was applied to unknown DTIs to predict the type of interaction. DT2Vec+ can offer a means to improve and support precision targeted treatments by selecting the drug candidates that can bind to specific targets in desired action modes. Predicting potential drugs provides an alternative approach to narrow down the search space that can be investigated in follow-up laboratory experiments. This approach can significantly reduce wet-laboratory work and experimental cost, but most importantly it can refine downstream experimental validation.

We note that the benefits conferred by our methodology stem from its capability to incorporate heterogeneous chemical and genomic data into a unified space, in addition to the fact that machine learning algorithms can handle numerical input features well. Predictions generated by our method can be used for virtual screening of novel DTIs at large scale. Although we have obtained promising results in predicting different types of DTIs using the drug–target–disease association network, this first report provides the foundations for the model reported here to be expanded by integrating more biological information into the association network.

Abbreviations

CV	Cross-validation
DDS	Drug–drug similarity
DTI	Drug target interaction
ML	Machine learning
PCA	Principal component analysis
PPS	Protein–protein similarity
PDX	Patient-derived xenograft
Ddis	Drug-disease
DisP	Disease-protein in the future

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-023-05317-w>.

Additional file 1. Fig. S1: steps performed to collect the dataset.

Acknowledgements

We thank two anonymous reviewers for helpful comments and suggestions.

Author contributions

EAS and ST conceived the study design. EAS conducted the data collection. EAS conducted computational experiments. EAS, AC, SNK and ST analysed and interpreted the data. EAS, AC, SNK and ST wrote the initial manuscript drafts. EAS, AC, SNK and ST performed critical editing. All authors read and approved the final manuscript.

Funding

This work has been supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) based at Guy's and St Thomas' NHS Foundation Trust and King's College London (IS-BRC-1215-20006). The authors are solely responsible for the decision to publish, and preparation of the manuscript. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. SNK acknowledges support from the Medical Research Council (MR/L023091/1); Breast Cancer Now (147; KCL-BCN-Q3); Cancer Research UK (C30122/A11527; C30122/A15774); and the Cancer Research UK King's Health Partners Centre at King's College London (C604/A25135).

Availability of data and materials

The datasets analysed during the current study are available in the ChEMBL repository (<https://www.ebi.ac.uk/chembl/>), CTD dataset (<http://ctdbase.org/>). The code is available at <https://github.com/elmira-amiri/DT2VecPlus>.

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.

Received: 22 February 2023 Accepted: 30 April 2023

Published online: 17 May 2023

References

- Hao M, Bryant SH, Wang Y. Open-source chemogenomic data-driven algorithms for predicting drug–target interactions. *Brief Bioinform.* 2019;20:1465–74.
- Ceddia G, Pinoli P, Ceri S, Masseroli M. Matrix factorization-based technique for drug repurposing predictions. *IEEE J Biomed Health Inform.* 2020;24:3162–72.
- Mohamed SK, Nováček V, Nounu A. Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics.* 2019;66:600.
- Csermely P, Korsmáros T, Kiss HJM, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther.* 2013;138:333–408.
- 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;9:66.
- Reddy AS, Zhang S. Polypharmacology: drug discovery for the future. *Expert Rev Clin Pharmacol.* 2013. <https://doi.org/10.1586/ecp.12.74>.
- 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;66:bbz157.
- 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.
- 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.
- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov.* 2004;3:673–83.
- Wang Y, Zeng J. Predicting drug–target interactions using restricted Boltzmann machines. *Bioinformatics.* 2013;29:i126–34.
- Liang S, Yu H. Revealing new therapeutic opportunities through drug target prediction: a class imbalance-tolerant machine learning approach. *Bioinformatics.* 2020;36:4490–7.
- Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med.* 2017;9:66.
- 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.
- Mullen J, Cockell SJ, Tipney H, Woollard PM, Wipat A. Mining integrated semantic networks for drug repositioning opportunities. *PeerJ.* 2016;4: e1558.
- Yang J, Li Z, Fan X, Cheng Y. Drug–disease association and drug-repositioning predictions in complex diseases using causal inference–probabilistic matrix factorization. *ACS Publications.* 2014. <https://doi.org/10.1021/ci500340n>. Accessed 8 Nov 2021.
- Amiri Sourì E, Tsoka S. Cancer classification through a hybrid machine learning approach. In: IARIA. Lisbon, Portugal; 2020. p. 20–7.
- Chen H, Cheng F, Li J. iDrug: Integration of drug repositioning and drug–target prediction via cross-network embedding. *PLoS Comput Biol.* 2020;16: e1008040.
- Xuan P, Hu K, Cui H, Zhang T, Nakaguchi T. Learning multi-scale heterogeneous representations and global topology for drug–target interaction prediction. *IEEE J Biomed Health Inform.* 2021;66:1–1.

20. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol.* 2011;7:496.
21. Lee T, Yoon Y. Drug repositioning using drug-disease vectors based on an integrated network. *BMC Bioinform.* 2018;19:446.
22. Xuan P, Zhang Y, Cui H, Zhang T, Guo M, Nakaguchi T. Integrating multi-scale neighbouring topologies and cross-modal similarities for drug-protein interaction prediction. *Brief Bioinform.* 2021;22(66):bbab119.
23. 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.
24. 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.
25. 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.
26. Patel L, Shukla T, Huang X, Ussery DW, Wang S. Machine learning methods in drug discovery. *Molecules.* 2020;25:5277.
27. Ezzat A, Wu M, Li X-L, Kwok C-K. Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinform.* 2019;20:1337–57.
28. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. *Brief Bioinform.* 2016;17:2–12.
29. Wang R, Li S, Cheng L, Wong MH, Leung KS. Predicting associations among drugs, targets and diseases by tensor decomposition for drug repositioning. *BMC Bioinformatics.* 2019;20:628.
30. Zhao B-W, Wang L, Hu P-W, Wong L, Su X-R, Wang B-Q, et al. Fusing higher and lower-order biological information for drug repositioning via graph representation learning. *IEEE Trans Emerg Top Comput.* 2023;66:1–14.
31. Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics.* 2018;34:1904–12.
32. Yue Y, He S. DTI-HeNE: a novel method for drug-target interaction prediction based on heterogeneous network embedding. *BMC Bioinform.* 2021;22:418.
33. Zhao B-W, You Z-H, Hu L, Guo Z-H, Wang L, Chen Z-H, et al. A novel method to predict drug-target interactions based on large-scale graph representation learning. *Cancers.* 2021;13:2111.
34. Boumahdi S, de Sauvage FJ. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nat Rev Drug Discov.* 2020;19:39–56.
35. Mottini C, Napolitano F, Li Z, Gao X, Cardone L. Computer-aided drug repurposing for cancer therapy: approaches and opportunities to challenge anticancer targets. *Semin Cancer Biol.* 2021;68:59–74.
36. Morris LGT, Chan TA. Therapeutic targeting of tumor suppressor genes. *Cancer.* 2015;121:1357–68.
37. 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.
38. Davis AP, Murphy CG, Saraceni-Richards CA, Rosenstein MC, Wiegiers TC, Mattingly CJ. Comparative toxicogenomics database: a knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Res.* 2009;37 Database issue:D786–792.
39. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46:D1074–82.
40. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 2012;40:D1100–7.
41. Peng L, Liao B, Zhu W, Li Z, Li K. Predicting drug-target interactions with multi-information fusion. *IEEE J Biomed Health Inform.* 2017;21:561–72.
42. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol.* 2008;4:682–90.
43. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49:D480–9.
44. 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.
45. 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.
46. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform.* 2015;7:20.
47. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminform.* 2011;3:33.
48. 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.
49. 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.
50. Grover A, Leskovec J. node2vec: scalable feature learning for networks. 2016.
51. Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: a survey | Elsevier Enhanced Reader. 2017. <https://reader.elsevier.com/reader/sd/pii/S0950705118301540?token=088BDCA1DCA7AF61A92BCAF7AD6EF05730187BF74C5D58F0D7606C2554D5A7A38130A142A368C480C84AB1735E830BAA>. Accessed 15 Dec 2020.
52. 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.
53. Mathai N, Chen Y, Kirchmair J. Validation strategies for target prediction methods. *Brief Bioinform.* 2020;21:791–802.
54. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016:785–94.
55. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–57.

56. Amiri Sourì E, Chenoweth A, Karagiannis SN, Tsoka S. Novel drug–target interactions via link prediction and network embedding (**submitted**) 2022.
57. Amiri Sourì E, Chenoweth A, Cheung A, Karagiannis SN, Tsoka S. Cancer Grade Model: a multi-gene machine learning-based risk classification for improving prognosis in breast cancer. *Br J Cancer*. 2021;125:748–58.
58. Wheatley SP, Altieri DC. Survivin at a glance. *J Cell Sci*. 2019;132:jcs223–826.
59. Fan H, Hu Z, Wang S, Wu W, Liu X, Geng H. 5-aminolevulinic-acid-mediated sonodynamic therapy improves the prognosis of melanoma by inhibiting survivin expression. *Cancer Biomark*. 2020;28:301–8.
60. He G-F, Bian M-L, Zhao Y-W, Xiang Q, Li H-Y, Xiao C. A study on the mechanism of 5-aminolevulinic acid photodynamic therapy in vitro and in vivo in cervical cancer. *Oncol Rep*. 2009;21:861–8.
61. Mulligan SP, Karlsson K, Strömberg M, Jönsson V, Gill D, Hammerström J, et al. Cladribine prolongs progression-free survival and time to second treatment compared to fludarabine and high-dose chlorambucil in chronic lymphocytic leukemia. *Leuk Lymphoma*. 2014;55:2769–77.
62. Chen K, Wu S, Ye S, Huang H, Zhou Y, Zhou H, et al. Dimethyl fumarate induces metabolic crisis to suppress pancreatic carcinoma. *Front Pharmacol*. 2021;12:134.
63. Saidu NEB, Kavian N, Leroy K, Jacob C, Nicco C, Batteux F, et al. Dimethyl fumarate, a two-edged drug: current status and future directions. *Med Res Rev*. 2019;39:1923–52.
64. Dréau D, Karaa A, Culberson C, Wyan H, McKillop IH, Clemens MG. Bosentan inhibits tumor vascularization and bone metastasis in an immunocompetent skin-fold chamber model of breast carcinoma cell metastasis. *Clin Exp Metastasis*. 2006;23:41–53.
65. Schenk RL, Strasser A, Dewson G. BCL-2: long and winding path from discovery to therapeutic target. *Biochem Biophys Res Commun*. 2017;482:459–69.
66. Joensuu H, Pylkkänen L, Toikkanen S. Bcl-2 protein expression and long-term survival in breast cancer. *Am J Pathol*. 1994;145:1191–8.
67. Emi M, Kim R, Tanabe K, Uchida Y, Toge T. Targeted therapy against Bcl-2-related proteins in breast cancer cells. *Breast Cancer Res*. 2005;7:R940.
68. Anticancer Activity of the Antibiotic Clotrimazole | *Cancer Research*. <https://cancerres.aacrjournals.org/content/65/8/3389.short>. Accessed 10 Dec 2021.
69. Ding W-Q, Liu B, Vaught JL, Yamauchi H, Lind SE. Anticancer activity of the antibiotic clotrimazole. *Cancer Res*. 2005;65:3389–95.
70. Shirali S, Aghaei M, Shabani M, Fathi M, Sohrabi M, Moenifard M. Adenosine induces cell cycle arrest and apoptosis via cyclinD1/Cdk4 and Bcl-2/Bax pathways in human ovarian cancer cell line OVCAR-3. *Tumour Biol*. 2013;34:1085–95.
71. Panjehpour M, Karami-Tehrani F. Adenosine modulates cell growth in the human breast cancer cells via adenosine receptors. *Oncol Res*. 2007;16:575–85.
72. Tanaka Y, Yoshihara K, Tsuyuki M, Kamiya T. Apoptosis induced by adenosine in human leukemia HL-60 cells. *Exp Cell Res*. 1994;213:242–52.
73. Saitoh M, Nagai K, Nakagawa K, Yamamura T, Yamamoto S, Nishizaki T. Adenosine induces apoptosis in the human gastric cancer cells via an intrinsic pathway relevant to activation of AMP-activated protein kinase. *Biochem Pharmacol*. 2004;67:2005–11.
74. Gessi S, Merighi S, Varani K, Cattabriga E, Benini A, Mirandola P, et al. Adenosine receptors in colon carcinoma tissues and colon tumoral cell lines: focus on the A(3) adenosine subtype. *J Cell Physiol*. 2007;211:826–36.
75. Choi MS, Moon S-M, Lee SA, Park B-R, Kim J-S, Kim DK, et al. Adenosine induces intrinsic apoptosis via the PI3K/Akt/mTOR signaling pathway in human pharyngeal squamous carcinoma FaDu cells Corrigendum in [/https://doi.org/10.3892/ol.2019.10014](https://doi.org/10.3892/ol.2019.10014). *Oncol Lett*. 2018;15:6489–96.
76. Bharti R, Dey G, Ojha PK, Rajput S, Jaganathan SK, Sen R, et al. Diacerein-mediated inhibition of IL-6/IL-6R signaling induces apoptotic effects on breast cancer. *Oncogene*. 2016;35:3965–75.
77. Dongdong Z, Jin Y, Yang T, Yang Q, Wu B, Chen Y, et al. Antiproliferative and immunoregulatory effects of azelaic acid against acute myeloid leukemia via the activation of notch signaling pathway. *Front Pharmacol*. 2019;10:1396.
78. Yu Y, Yang B, Yu J, Zhao G, Chen F. Dequalinium chloride inhibits the growth of human glioma cells in vitro and vivo: a study on molecular mechanism and potential targeted agents. *Acta Neurochir*. 2020;162:1683–90.
79. Bleday R, Weiss MJ, Salem RR, Wilson RE, Chen LB, Steele G. Inhibition of rat colon tumor isograft growth with dequalinium chloride. *Arch Surg*. 1986;121:1272–5.
80. Weiss MJ, Wong JR, Ha CS, Bleday R, Salem RR, Steele GD, et al. Dequalinium, a topical antimicrobial agent, displays anticarcinoma activity based on selective mitochondrial accumulation. *Proc Natl Acad Sci USA*. 1987;84:5444–8.
81. Helige C, Smolle J, Zellnig G, Fink-Puches R, Kerl H, Tritthart HA. Effect of dequalinium on K1735–M2 melanoma cell growth, directional migration and invasion in vitro. *Eur J Cancer*. 1992;29A:124–8.
82. Pan Y, Zhao S, Chen F. The potential value of dequalinium chloride in the treatment of cancer: focus on malignant glioma. *Clin Exp Pharmacol Physiol*. 2021;48:445–54.
83. Hu H-F, Xu WW, Li Y-J, He Y, Zhang W-X, Liao L, et al. Anti-allergic drug azelastine suppresses colon tumorigenesis by directly targeting ARF1 to inhibit IQGAP1-ERK-Drp1-mediated mitochondrial fission. *Theranostics*. 2021;11:1828–44.
84. Fu S, Chen X, Lo H-W, Lin J. Combined bazedoxifene and paclitaxel treatments inhibit cell viability, cell migration, colony formation, and tumor growth and induce apoptosis in breast cancer. *Cancer Lett*. 2019;448:11–9.
85. Dang CV. MYC on the path to cancer. *Cell*. 2012;149:22–35.
86. Call SG, Duren RP, Panigrahi AK, Nguyen L, Freire PR, Grimm SL, et al. Targeting oncogenic super enhancers in MYC-dependent AML using a small molecule activator of NR4A nuclear receptors. *Sci Rep*. 2020;10:2851.
87. Monini P, Sgadari C, Toschi E, Barillari G, Ensoli B. Antitumour effects of antiretroviral therapy. *Nat Rev Cancer*. 2004;4:861–75.
88. Banerjee K, Resat H. Constitutive activation of STAT3 in breast cancer cells: a review. *Int J Cancer*. 2016;138:2570–8.
89. Shi X, Gong L, Liu Y, Hou K, Fan Y, Li C, et al. 4-phenylbutyric acid promotes migration of gastric cancer cells by histone deacetylase inhibition-mediated IL-8 upregulation. *Epigenetics*. 2020;15:632–45.

90. Park MS, Kim B-R, Kang S, Kim D-Y, Rho SB. The antihypertension drug doxazosin suppresses JAK/STATs phosphorylation and enhances the effects of IFN- α/γ -induced apoptosis. *Genes Cancer*. 2014;5:470–9.
91. Siddiqui EJ, Shabbir M, Thompson CS, Mumtaz FH, Mikhailidis DP. Growth inhibitory effect of doxazosin on prostate and bladder cancer cells. Is the serotonin receptor pathway involved? *Anticancer Res*. 2005;25:4281–6.
92. Sullivan RM, Stone M, Marshall JF, Uberall F, Rotenberg SA. Photo-induced inactivation of protein kinase C by dequalinium inhibits motility of murine melanoma cells. 9.
93. Huo X, Li J, Zhao F, Ren D, Ahmad R, Yuan X, et al. The role of capecitabine-based neoadjuvant and adjuvant chemotherapy in early-stage triple-negative breast cancer: a systematic review and meta-analysis. *BMC Cancer*. 2021;21:78.
94. Marangoni E, Laurent C, Coussy F, El-Botty R, Château-Joubert S, Servely J-L, et al. Capecitabine efficacy is correlated with TYMP and RB1 expression in PDX established from triple-negative breast cancers. *Clin Cancer Res*. 2018;24:2605–15.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

