

RESEARCH ARTICLE

NTD-DR: Nonnegative tensor decomposition for drug repositioning

Ali Akbar Jamali¹, Yuting Tan^{1,2}, Anthony Kusalik^{1,3*}, Fang-Xiang Wu^{1,3,4*}

1 Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK, Canada, **2** School of Mathematics and Statistics, Huazhong Normal University, Wuhan, China, **3** Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada, **4** Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK, Canada

* kusalik@cs.usask.ca (AK); faw341@mail.usask.ca (FXW)



Abstract

Computational drug repositioning aims to identify potential applications of existing drugs for the treatment of diseases for which they were not designed. This approach can considerably accelerate the traditional drug discovery process by decreasing the required time and costs of drug development. Tensor decomposition enables us to integrate multiple drug- and disease-related data to boost the performance of prediction. In this study, a nonnegative tensor decomposition for drug repositioning, NTD-DR, is proposed. In order to capture the hidden information in drug-target, drug-disease, and target-disease networks, NTD-DR uses these pairwise associations to construct a three-dimensional tensor representing drug-target-disease triplet associations and integrates them with similarity information of drugs, targets, and disease to make a prediction. We compare NTD-DR with recent state-of-the-art methods in terms of the area under the receiver operating characteristic (ROC) curve (AUC) and the area under the precision and recall curve (AUPR) and find that our method outperforms competing methods. Moreover, case studies with five diseases also confirm the reliability of predictions made by NTD-DR. Our proposed method identifies more known associations among the top 50 predictions than other methods. In addition, novel associations identified by NTD-DR are validated by literature analyses.

OPEN ACCESS

Citation: Jamali AA, Tan Y, Kusalik A, Wu F-X (2022) NTD-DR: Nonnegative tensor decomposition for drug repositioning. PLoS ONE 17(7): e0270852. <https://doi.org/10.1371/journal.pone.0270852>

Editor: Jinyan Li, University of Technology Sydney, AUSTRALIA

Received: March 9, 2022

Accepted: June 20, 2022

Published: July 21, 2022

Copyright: © 2022 Jamali et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from Github: <https://github.com/AliJam82/NTD-DR>.

Funding: FXW received the fund from Natural Sciences and Engineering Research Council of Canada (NSERC).

Competing interests: The authors have declared that no competing interests exist.

Introduction

Developing a new drug is a costly process in terms of time, risk, and financial resources. For one drug from an initial idea to a product in the market it requires 17–20 years and ~USD 2 billion [1]. Fortunately, complementary approaches can hasten the process of drug discovery. Drug repositioning, wherein an existing approved drug is used to treat a disease other than the one it is designed for, is an opportunity to decrease the relative expense of drug discovery.

Both experimental and computational methods can be used for drug repositioning. Experimental drug repositioning includes screening drugs across a set of targets (i.e. proteins, nucleic acids, etc. which interact with drugs) and diseases that requires facilities and procedures which are expensive and tedious. On the other hand, computational approaches try to avoid these limitations by predicting associations between existing drugs and diseases. The later methods are promising because they are efficient in terms of time, expenses, and results.

Computational drug repositioning methods can be classified into different categories including machine learning, network-based, text mining, and signature matching methods. Most machine learning methods solve the question of drug repositioning as a classification task. Among the many classification algorithms available, support vector machines (SVMs) [2, 3] and random forest [4] have been used for drug repositioning. The most widely used methods for drug repositioning belong to the network-based category, which relies on interaction network data. These methods include random walk [5, 6], matrix factorization [7–11], and network inference [12–15]. Text mining approaches [16, 17] rely on biomedical data and entities, and the co-occurrence of similar/same keywords. Signature matching-based methods for drug repositioning first identify the disease gene signature (drug signature) based on the gene expression profile of the corresponding disease (drug). Then the identified disease gene (drug) signature is connected with drugs (diseases) through the drug-induced gene expression analysis. In this way, the association between drugs and diseases can be identified [18–20].

Regardless of the category, computational methods use experimentally validated associations of drugs and diseases to identify new associations between drugs and diseases. Most of these methods are based on the assumption that if drug C1 is associated with disease D1, and drug C2 is similar to drug C1, disease D1 can be associated with drug C2. Li and Pan [12] used similarity side-information and a hybrid neural network to develop a recommendation system to identify existing and novel associations between drugs and diseases to be used for drug repositioning. Luo *et al.* [21] developed a recommendation system for drug-disease association identification called Drug Repositioning Recommendation System (DDRS). Their method uses drug-disease, disease-condition, and drug-condition networks to determine whether a drug can be used to treat a disease. Xuan *et al.* [8] proposed a nonnegative matrix factorization method called DisDrugPred. They constructed similarity matrices for drugs and diseases based on drug-disease associations. They then used these similarities to predict new associations between drugs and disease. He *et al.* [22] introduced a unique method called PIMD that predicts drug therapeutics using multi-dimensional data. In their method, clusters of drugs that can be used for the treatment of a particular disease are constructed based on multiple drug similarities. Jin *et al.* [13] proposed an approach that integrates multiple heterogeneous networks to predict association scores of drugs and diseases. Their method combines drug and disease features retrieved from multiple drug networks and known drug-disease association networks, respectively.

Despite the advantages of these methods, they usually do not consider drug-disease associations at the molecular level. Any association between drugs and diseases is facilitated by a hidden component or target (gene, protein, etc.) which is typically ignored in drug repositioning methods. To integrate these hidden components in drug repositioning, a method is required that can handle different types of data. A tensor, as an n -dimensional array, can store multiple types of association information where each dimension represents one type of data. Tensor decomposition has been used as an effective way to study tensors in various fields of bioinformatics for novel association identification. Huang *et al.* [23] proposed a tensor decomposition method named tensor decomposition with rational constraints (TDRC) to identify the multiple types of associations between microRNAs and diseases. Luo *et al.* [24] developed a tensor decomposition method to identify microRNA-disease associations. In their work, the association information of microRNAs, genes, and diseases was used. To boost the performance of their method, the authors used biological similarity information of microRNAs, genes, and diseases as auxiliary information. Chen *et al.* [25] proposed a tensor decomposition approach to identify drug-target-disease associations for drug discovery. Moreover, Wang *et al.* [26] developed a tensor decomposition method to identify new drug-target-disease triplet

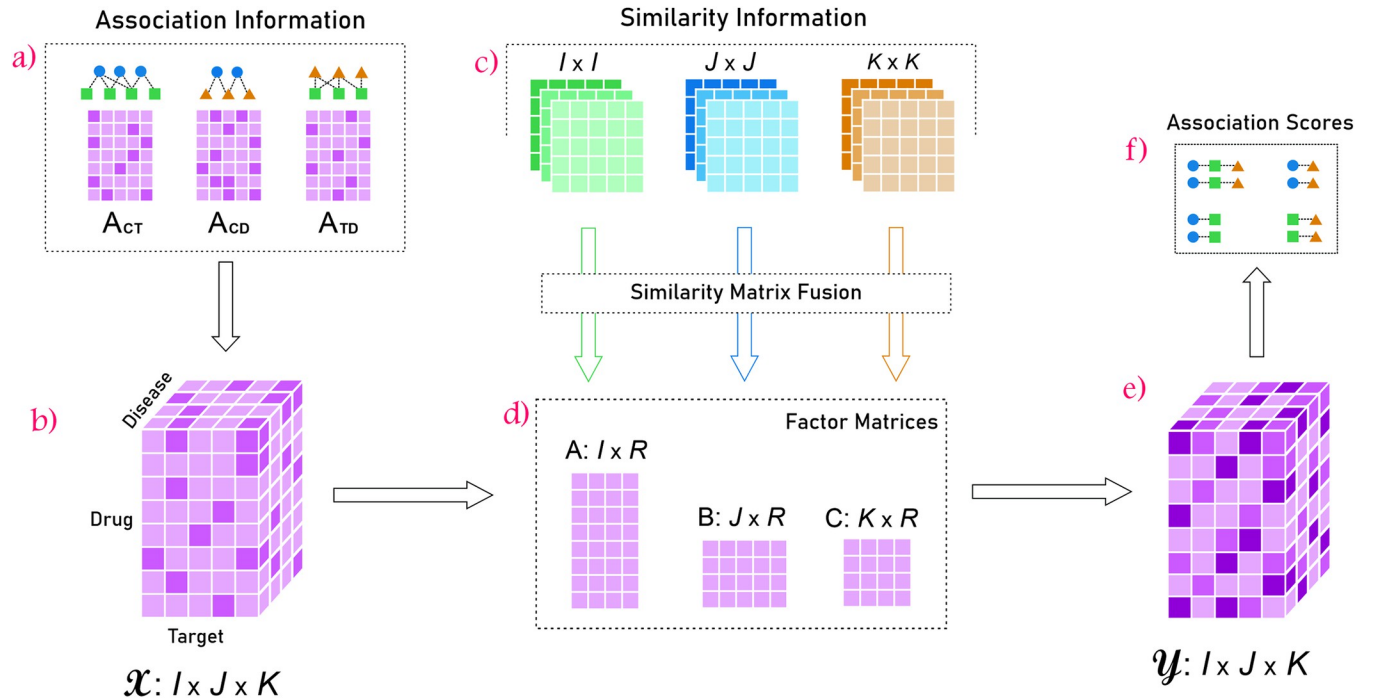


Fig 1. The workflow of NTD-DR. a) Known drug-target (A_{CT}), drug-disease (A_{CD}), and target-disease (A_{TD}) pairwise associations are collected. b) Drug-target-disease association tensor \mathcal{X} is constructed based on A_{CT} , A_{CD} , and A_{TD} . c) Multiple similarity measures for drugs, targets, and diseases are collected and are fused to build a single similarity matrix for each of drugs, targets, and diseases. d) Drug-target-disease association tensor is factorized into three factor matrices A, B, and C. e) Tensor \mathcal{Y} is reconstructed using similarity matrices upon the convergence of the factor matrices (see Section “Optimization process”). f) The pairwise or triplet association scores are computed.

<https://doi.org/10.1371/journal.pone.0270852.g001>

associations. Chen *et al.* [27] described a tensor decomposition method named neural tensor network (NeurTN) for personalized medicine. This method combines the concept of tensor and deep neural networks to find the associations among drugs, targets, and diseases. Existing methods suffer from two shortcomings: they consider only triplet associations of drugs, targets, and diseases which ignores valuable pairwise associations; or they use single similarity for drugs, targets, and diseases which ignores the impact of various similarity information. Inspired by this, we use a tensor to integrate triplet association information of drugs, targets, and diseases. In contrast to previous works, we propose a nonnegative tensor decomposition for drug repositioning (NTD-DR) which applies drug-target, drug-disease, and target-disease pairwise associations and combines them to make predictions using multiple types of similarities for drugs, targets, and diseases. Moreover, NTD-DR not only can identify the triplet drug-target-disease associations, but also it can predict the pairwise associations between drugs, targets, and diseases.

NTD-DR is outlined as follows. First, we collect drug-disease, drug-target, and target-disease pairwise associations to construct a three-dimensional association tensor. Second, we formulate an objective function to decompose the constructed tensor into three factor matrices and integrate them with similarity information of drugs, targets, and diseases. Then we reconstruct the tensor, based on the factor matrices. Finally, we retrieve the prediction score for triplet and pairwise associations from the reconstructed tensor. We evaluate the performance of our method using cross-validation and separate data. Fig 1 schematically illustrates our algorithm.

Materials and methods

Data

Association information. We retrieve data from various public sources. To construct a drug-target association matrix, we download data from DrugBank [28], UniProt [29], and SuperTarget [30]. After removing redundant entries, the drug-target association matrix ($A_{CT} \in \mathbb{R}^{I \times J}$) consists of 13898 validated associations. Drug-disease associations are downloaded from Online Mendelian Inheritance in Man (OMIM) [31] and the Comparative Toxicogenomics Database (CTD) [32], and $A_{CD} \in \mathbb{R}^{I \times K}$ is constructed with 550319 known associations. Moreover, 14730 associations between targets and diseases are retrieved from the Comparative Toxicogenomics Database, OMIM, Uniprot, DisGeNET [33], and GAD [34] to construct $A_{TD} \in \mathbb{R}^{J \times K}$.

After the associations between drugs, targets, and diseases are retrieved, we construct the drug-target-disease tensor ($\mathcal{X} \in \mathbb{R}^{I \times J \times K}$) with 114319 triplet associations. Not surprisingly, our constructed tensor is very sparse (the ratio of known to unknown interactions is 1:119884). The sparsity of the drug-target-disease tensor can lead to the inability of our association models to learn robust feature representations of drugs, targets, and diseases, making them more vulnerable to the cold-start problem, which results in low generalization performance of the models. To mitigate the sparsity of the tensor, in this study we filter out those drugs, targets, and diseases with less than five interactions. After filtration the tensor includes $I = 810$ drugs, $J = 302$ targets and $K = 542$ diseases. This filtration step results in a significant reduction in the sparsity to a ratio of known to unknown interactions of 1:1708. In the final constructed tensor, we randomly divide the validated interactions (known as positive samples) into three subsets: 90% for training and testing our method (dataset P), 5% to set parameters (dataset S), and the remaining 5% as a separate validation data (dataset I) for case studies. All subsets include randomly chosen negative samples equal in number to the positive samples.

Similarity information. To boost the performance of prediction we employ different types of similarities for drugs, targets, and diseases to construct multiple similarity matrices for each category. For drugs, we construct five types of similarities, including 1) chemical structure-based and 2) ATC-based similarities using DrugBank data, 3) target-based and 4) gene ontology (GO)-based similarities using UniProt data, and 5) pathway-based similarity using Comparative Toxicogenomics Database data. For targets, we construct three types of similarity matrices: 1) sequence-based similarity, which is computed using sequence structure information of targets retrieved from DrugBank and UniProt; 2) protein-protein interaction (PPI) network similarity of targets, which is calculated based on the data retrieved from InAct [35], BioGrid [36], MINT [37], STRING [38], and HPRD [39]; and 3) gene ontology (GO) semantic similarity of targets, which is computed using data from UniProt. For diseases, we construct four types of similarity matrices, including 1) drug-disease association-based, 2) gene ontology (GO)-based, 3) disease-gene association-based, and 4) PPI-based similarities. These last four similarities are computed using data retrieved from DisGeNET and Comparative Toxicogenomics Database. The construction of these similarity matrices is described in detail in our previous work [40]. Finally, after multiple similarity matrices for drugs, targets, and diseases are constructed, we combine the similarity matrices of each category (drugs, targets, and diseases) via the Similarity Network Fusion (SNF) method described by Wang *et al.* [41] to construct final fused similarity matrices for each category of drugs, targets, and diseases (S_C , S_T and S_D).

Problem formulation

This study aims to identify new drug-target-disease, drug-target, drug-disease, and target-disease associations. The identification of these associations can be formulated as a tensor completion problem. Consider a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ to describe known associations among I drugs, J targets, and K diseases, where $\mathcal{X}_{ijk} = 1$ if the association among drug i , target j , and disease k is known, and 0 otherwise. Assume that the matrix (second-order tensor) $A_{CT} \in \mathbb{R}^{I \times J}$ describes known associations among I drugs and J targets, where $A_{CTij} = 1$ if the association between drug i and target j is known and 0 otherwise. Similarly, the matrix $A_{CD} \in \mathbb{R}^{I \times K}$ describes known associations among I drugs and K diseases, while the matrix $A_{TD} \in \mathbb{R}^{J \times K}$ describes known associations among J targets and K diseases. Furthermore, matrices $S_C \in \mathbb{R}^{I \times I}$, $S_T \in \mathbb{R}^{J \times J}$, and $S_D \in \mathbb{R}^{K \times K}$ describe the similarity of I drugs, J targets, and K diseases, respectively. We use the pairwise associations to construct tensor \mathcal{X} and after decomposition, we use similarity information of drugs, targets and diseases to update three factor matrices and reconstruct tensor \mathcal{Y} . The rank of tensor \mathcal{Y} is the minimum number of rank-1 tensors needed to produce \mathcal{Y} as their summation. Therefore, a third-order tensor $\mathcal{Y} \in \mathbb{R}^{I \times J \times K}$ of rank at most R can be written as:

$$\mathcal{Y} = \sum_{r=1}^R a_r \circ b_r \circ c_r \tag{1}$$

where $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$, $c_r \in \mathbb{R}^K$ for $r = 1, \dots, R$. Elementwise, Eq (1) can be written as:

$$\mathcal{Y}_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} \tag{2}$$

for $i = 1, \dots, I$; $j = 1, \dots, J$; and $k = 1, \dots, K$.

The factor matrices refer to the combination of the column vectors from the rank-1 components; i.e., $A = [a_1, \dots, a_R]$ and likewise, for B and C . With this notation, the above third-order tensor can be denoted by $\mathcal{Y} = [A, B, C]$.

Optimization process

Now we consider estimating A , B , and C from data in a third-order tensor \mathcal{X} with the constraints that the elements of A , B , and C are nonnegatives. Adopting a least square criterion, we have the following optimization problem:

$$\begin{aligned} \min_{A,B,C} \mathcal{L}_T(A, B, C) &= \min_{A,B,C} \alpha \|\mathcal{X} - \mathcal{Y}\|_F^2 \\ &= \min_{A,B,C} \alpha \|\mathcal{X} - \sum_{r=1}^R a_r \circ b_r \circ c_r\|_F^2 \\ \text{s.t. } &A \geq 0, B \geq 0, C \geq 0 \end{aligned} \tag{3}$$

where α is a positive regularization coefficient to regulate the tensor decomposition, $\|\mathcal{Y}\|_F^2$ is the squared Frobenius norm of the tensor, and the sum of square error is the objective function computed as follows:

$$\|\mathcal{Y}\|_F^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \mathcal{Y}_{ijk}^2 \tag{4}$$

With the concept of the tensor matricization [42], the above objective functions are equivalent to any one of the following functions:

$$\begin{aligned}
 \mathcal{L}_T(A, B, C) &= \alpha \| X_{(1)} - A(C \otimes B)^T \|^2_F \\
 \mathcal{L}_T(A, B, C) &= \alpha \| X_{(2)} - B(C \otimes A)^T \|^2_F \\
 \mathcal{L}_T(A, B, C) &= \alpha \| X_{(3)} - C(B \otimes A)^T \|^2_F
 \end{aligned}
 \tag{5}$$

where $X_{(n)}$ is the mode- n matricization of tensor \mathcal{X} , and \otimes is the Khatri-Rao product of two matrices. Drug-target, drug-disease, and target-disease pairwise associations are taken into consideration by the following optimization equation:

$$\begin{aligned}
 \min_{A,B,C} \mathcal{L}_A(A, B, C) &= \min_{A,B,C} \{ \lambda_{CT} \| A_{CT} - \sum_{r=1}^R a_r \circ b_r \|^2_F + \lambda_{CD} \| A_{CD} - \sum_{r=1}^R a_r \circ c_r \|^2_F + \lambda_{TD} \| A_{TD} - \sum_{r=1}^R b_r \circ c_r \|^2_F \} \\
 &= \min_{A,B,C} \{ \lambda_{CT} \| A_{CT} - AB^T \|^2_F + \lambda_{CD} \| A_{CD} - AC^T \|^2_F + \lambda_{TD} \| A_{TD} - BC^T \|^2_F \} \\
 &\quad s.t. A \geq 0, B \geq 0, C \geq 0
 \end{aligned}
 \tag{6}$$

where λ_{CT} , λ_{CD} , and λ_{TD} are positive regularization coefficients to regulate the importance of their corresponding associations. The similarity of drugs, targets, and diseases are taken into consideration by the following optimization equation:

$$\begin{aligned}
 \min_{A,B,C} \mathcal{L}_S(A, B, C) &= \min_{A,B,C} \{ \gamma_C \sum_{ij=1}^I S_{Cij} \| a_i - a_j \|^2_2 + \gamma_T \sum_{ij=1}^J S_{Tij} \| b_i - b_j \|^2_2 + \gamma_D \sum_{ij=1}^K S_{Dij} \| c_i - c_j \|^2_2 \} \\
 &= \min_{A,B,C} \{ \gamma_C tr(A^T L_C A) + \gamma_T tr(B^T L_T B) + \gamma_D tr(C^T L_D C) \} \\
 &\quad s.t. A \geq 0, B \geq 0, C \geq 0
 \end{aligned}
 \tag{7}$$

where a_i is the i -th row of matrix A , $tr(\cdot)$ is the trace of a matrix, $L_C = D_C - S_C$ is the Laplacian matrix of drugs, D_C is the diagonal matrix whose i -th diagonal element is the summation of the i -th column of the drug similarity matrix S_C , and likewise for L_T and L_D . Variables γ_C , γ_T , and γ_D are positive regularization coefficients to regulate the importance of their corresponding similarities. Then the integrated optimization problem is as follows:

$$\begin{aligned}
 \min_{A,B,C} \mathcal{L}(A, B, C) &= \min_{A,B,C} \{ \mathcal{L}_T(A, B, C) + \mathcal{L}_S(A, B, C) + \mathcal{L}_A(A, B, C) \} \\
 &\quad s.t. A \geq 0, B \geq 0, C \geq 0
 \end{aligned}
 \tag{8}$$

The Karush–Kuhn–Tucker (KKT) conditions [43] for the above optimization problems are similar to those in the work of Tian *et al.* [44] as follows:

$$\begin{aligned}
 A \geq 0, \frac{\partial \mathcal{L}(A, B, C)}{\partial A} &\geq 0, A \star \frac{\partial \mathcal{L}(A, B, C)}{\partial A} = 0 \\
 B \geq 0, \frac{\partial \mathcal{L}(A, B, C)}{\partial B} &\geq 0, B \star \frac{\partial \mathcal{L}(A, B, C)}{\partial B} = 0 \\
 C \geq 0, \frac{\partial \mathcal{L}(A, B, C)}{\partial C} &\geq 0, C \star \frac{\partial \mathcal{L}(A, B, C)}{\partial C} = 0
 \end{aligned}
 \tag{9}$$

Taking the derivatives of $\mathcal{L}(A, B, C)$ with respect to A yields:

$$\begin{aligned} \frac{\partial \mathcal{L}(A, B, C)}{\partial A} &= \frac{\partial \mathcal{L}_T(A, B, C)}{\partial A} + \frac{\partial \mathcal{L}_S(A, B, C)}{\partial A} + \frac{\partial \mathcal{L}_A(A, B, C)}{\partial A} \\ &= \alpha(-2X_{(1)}(C \otimes B) + 2A(C \otimes B)^T(C \otimes B)) + 2\gamma_C L_C A \\ &\quad + \lambda_{CT}(-2A_{CT}B + 2A(B^T B)) + \lambda_{CD}(-2A_{CD}C + 2A(C^T C)) \\ &= -2(\alpha X_{(1)}(C \otimes B) + \gamma_C S_C A + \lambda_{CT} A_{CT} B + \lambda_{CD} A_{CD} C) \\ &\quad + 2A(\alpha(C^T C) \star (B^T B) + \lambda_{CT}(B^T B) + \lambda_{CD}(C^T C)) + 2\gamma_C D_C A \end{aligned} \tag{10}$$

Similarly, the derivatives of $\mathcal{L}(A, B, C)$ with respect to B and C are as follows:

$$\begin{aligned} \frac{\partial \mathcal{L}(A, B, C)}{\partial B} &= -2\left(\alpha X_{(2)}(C \otimes A) + \gamma_T S_T B + \lambda_{CT} A_{CT}^T A + \lambda_{TD} A_{TD} C\right) \\ &\quad + 2B(\alpha(C^T C) \star (A^T A) + \lambda_{CT}(A^T A) + \lambda_{TD}(C^T C)) + 2\gamma_T D_T B \\ \frac{\partial \mathcal{L}(A, B, C)}{\partial C} &= -2\left(\alpha X_{(3)}(B \otimes A) + \gamma_D S_D C + \lambda_{CD} A_{CD}^T A + \lambda_{TD} A_{TD}^T B\right) \\ &\quad + 2C(\alpha(B^T B) \star (A^T A) + \lambda_{CD}(A^T A) + \lambda_{TD}(B^T B)) + 2\gamma_D D_D C \end{aligned} \tag{11}$$

Therefore, we can have the following updating rules:

$$\begin{aligned} A &\leftarrow A \star \left[\frac{\alpha X_{(1)}(C \otimes B) + \gamma_C S_C A + \lambda_{CT} A_{CT} B + \lambda_{CD} A_{CD} C}{A(\alpha(C^T C) \star (B^T B) + \lambda_{CT}(B^T B) + \lambda_{CD}(C^T C)) + \gamma_C D_C A} \right] \\ B &\leftarrow B \star \left[\frac{\alpha X_{(2)}(C \otimes A) + \gamma_T S_T B + \lambda_{CT} A_{CT}^T A + \lambda_{TD} A_{TD} C}{B(\alpha(C^T C) \star (A^T A) + \lambda_{CT}(A^T A) + \lambda_{TD}(C^T C)) + \gamma_T D_T B} \right] \\ C &\leftarrow C \star \left[\frac{\alpha X_{(3)}(B \otimes A) + \gamma_D S_D C + \lambda_{CD} A_{CD}^T A + \lambda_{TD} A_{TD}^T B}{C(\alpha(B^T B) \star (A^T A) + \lambda_{CD}(A^T A) + \lambda_{TD}(B^T B)) + \gamma_D D_D C} \right] \end{aligned} \tag{12}$$

where \star is the Hadamard product of two matrices and $\left[\frac{\cdot}{\cdot} \right]$ is the element-wise division of two matrices. When matrices A , B , and C converge, tensor association matrices A_{CT}^* (drug-target association prediction matrix), A_{CD}^* (drug-disease association prediction matrix), and A_{TD}^* (target-disease association prediction matrix) can be constructed based on Eq (2). The aim of this study is to identify the associations between drugs and diseases through the construction of matrix A_{CD}^* , although our method can also be applied to identify the associations between targets and drugs through A_{CT}^* and between targets and diseases through A_{TD}^* by appropriately adjusting the parameters in the objective function (8).

Performance evaluation

In this study, 10-fold cross-validation (CV) is performed in two steps. First, dataset S is used to set up parameters, then dataset P is used to evaluate the performance of the prediction. In the association matrix where the rows (columns) include drugs (diseases), we use experimentally verified associations as positive samples. In most studies, undiscovered associations are used as negative samples. Because these associations can be potentially associations that are not yet discovered, we use the method developed by Luo *et al.* [45] to filter undiscovered associations. Then a number equal to the positive samples in datasets S and P are chosen randomly from filtered undiscovered pairs to be negative samples. Both positive and negative samples are randomly divided into 10 equal subsets. Nine subsets are used in turn as training sets and the remaining subset as test set. CV is performed under three scenarios including:

CV_{pairwise} : cross-validation on the pairs within the association matrices, which evaluates the prediction of new pairs by the method.

$CV_{\text{column-wise}}$: cross-validation on the columns within the association matrices, which evaluates the prediction of new column (disease) entries. The aim of this CV is to evaluate the performance of our proposed method in detecting associations between existing drugs and diseases that have not been previously treated with those drugs.

$CV_{\text{row-wise}}$: cross-validation on the rows within the association matrices, which aims to evaluate the prediction of new row (drug) entries. The aim of this CV is to evaluate the performance of our proposed method in detecting associations between existing diseases and new drugs or drugs that have not been previously used to treat the diseases.

As discussed before, tensor decomposition has the potential to identify novel associations between drug-target-disease, drug-target, drug-disease, and target-disease. As the aim of this study is to focus on the identification of drug and disease associations, we set $\lambda_{CD} = 1$ and $\gamma_C = \gamma_D = 1$. Also, several parameters introduced to our objective functions in Eqs 3–7 are set so as to boost the performance of prediction. Grid search is used to set and select the optimum value for the other parameters. These parameters are as follows: R , the rank of the tensor is set from $\{\frac{1}{2}, \frac{1}{2}, \frac{K}{2}, I, J, K\}$; α , a positive regularization coefficient from $\{0.1, 0.3, 0.4, 0.7, 0.9\}$; λ_{CT} and λ_{TD} from $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$; γ_T from $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$.

To either set parameters or investigate the performance of the proposed method, the area under the receiver operating characteristic (ROC) curve (AUC) and the area under the precision and recall curve (AUPR) are used as evaluation metrics.

Implementation

The NTD-DR algorithm is implemented in Python 3.6 programming language using standard modules numpy [46], pandas [47], matplotlib [48], scipy [49], networkx [50], scikit-learn [51]. Comparator methods are also programmed in Python. The source codes and data are available at <https://github.com/AliJam82/NTD-DR>.

Case study with a separate dataset

To further investigate the reliability of our method, a separate dataset I as described in Section “Data” is used for case studies. We select a case study disease in dataset I and set its interaction profiles equal to 0. After prediction, we retrieve the association scores of validated associations originally in dataset I. An optimal model should be able to predict greater association scores for validated associations. Since the algorithm is completely blind to this dataset, the accuracy of the predictions made by the method is a reliable measure of the performance of prediction.

Results

In this section, a comprehensive investigation of the performance of the proposed method is discussed. First, we discuss parameter tuning for our method to obtain an optimal combination of parameters based on the CV on dataset S. Then, we discuss the superiority of the proposed method on recovering missing drug-disease associations by comparing with baseline methods in three CV scenarios using dataset P in terms of AUC and AUPR. Moreover, we present case studies based on the results of an evaluation with a separate dataset, dataset I.

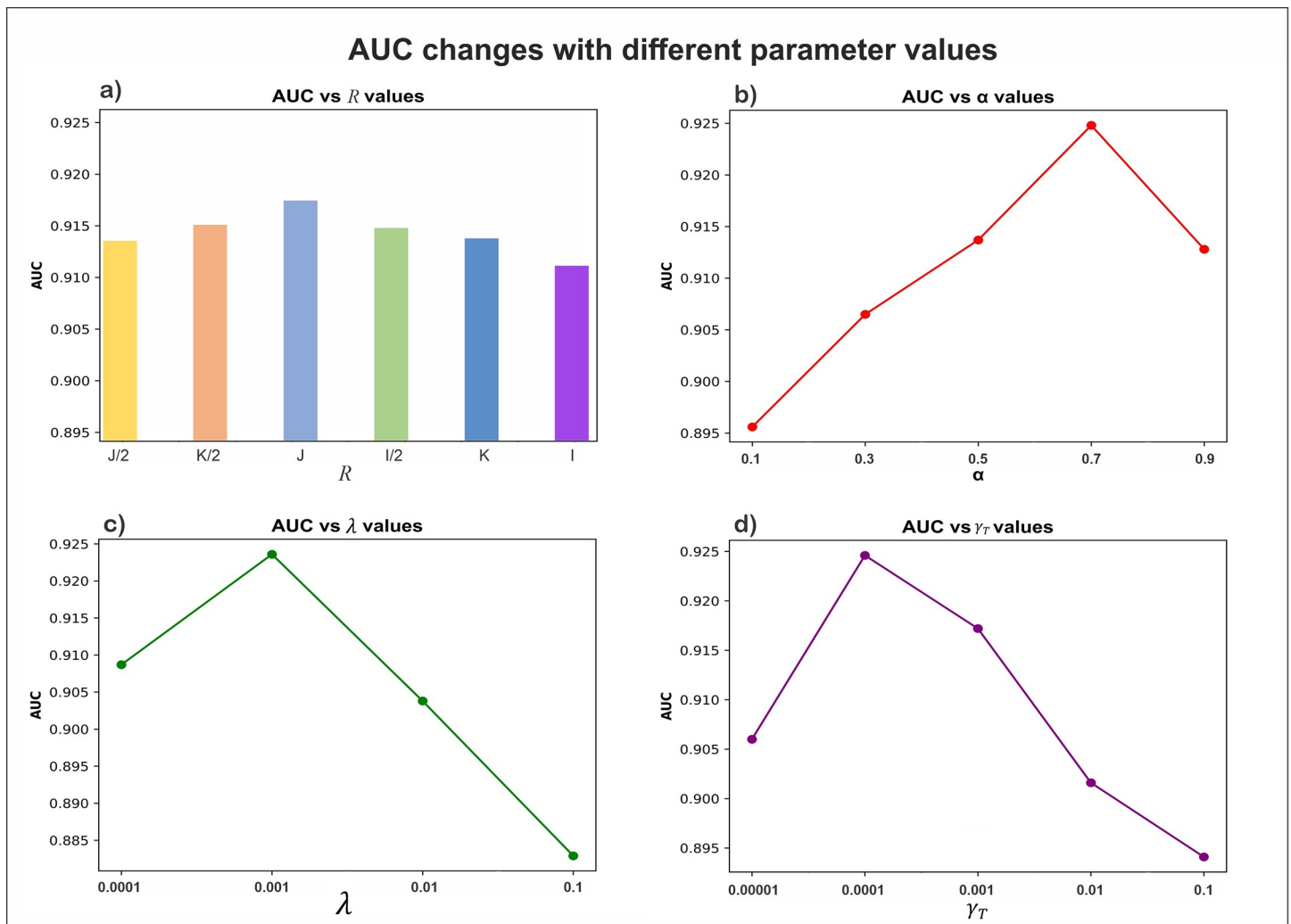


Fig 2. The effect of different values of four parameters on the performance of NTD-DR. The panels show the AUC changes with the increase of parameter a) R b) α , c) λ (the same value is set to both λ_{CT} and λ_{TD}), and d) γ_T , respectively.

<https://doi.org/10.1371/journal.pone.0270852.g002>

Parameter optimization

There are several parameters in our proposed method. We analyze each in turn with Scenario CV_{pairwise} yielding the results shown in Fig 2. While Fig 2 only shows AUC, the performance measured AUPR is similar.

Impact of R . The results (Fig 2A) show that the values for AUC and AUPR generally increase with an increase of R . However, the AUC and AUPR start to decrease when R exceeds a certain value. The greater values of R might lead to an expensive computation and only result in insignificant performance improvement. The best performance is achieved when we set $R = J$ (the number of targets). Thus, we adopt $R = J$ to be the optimal rank of our tensor, which not only increases the performance of prediction but also reduces the cost of computing resources.

Impact of α . We fix $R = J$ as the optimal rank of our tensor and vary α . As confirmed by the results (Fig 2B), the performance of our method initially improves with increasing α and then declines when it is greater than 0.7. The optimal performance achieved when $\alpha = 0.7$.

AUC (on the left) and AUPR (on the right) curves under CV_{pairwise}

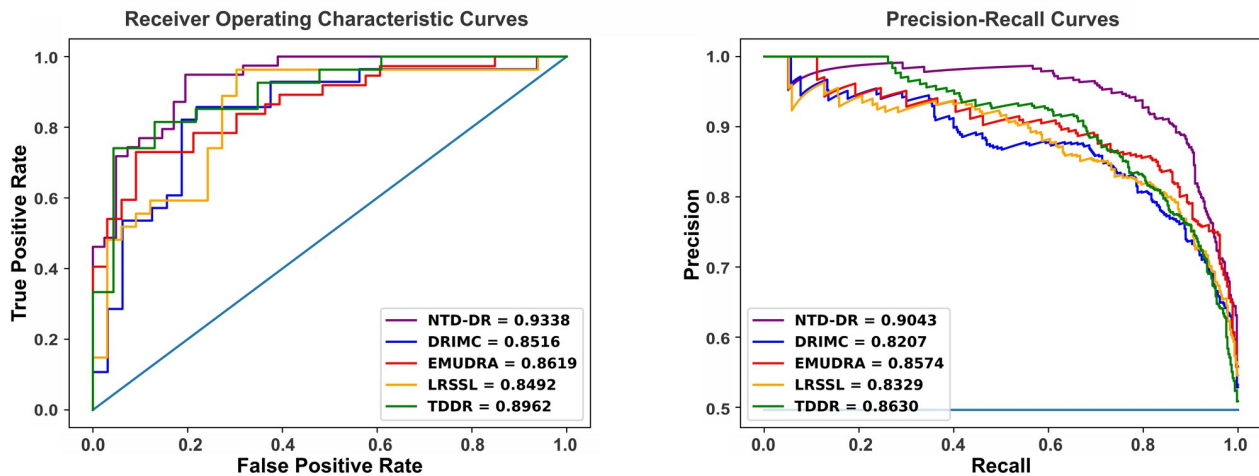


Fig 3. ROC and precision-recall curves of NTD-DR and the existing methods for Scenario CV_{pairwise} using dataset P.

<https://doi.org/10.1371/journal.pone.0270852.g003>

Impact of λ_{CT} and λ_{TD} . We fix $\alpha = 0.7$ and $R = J$ and vary λ_{CT} and λ_{TD} . We find that (Fig 2C) the performance of our method shows an increasing trend with the decrease of λ_{CT} and λ_{TD} , but it starts to trend down when λ_{CT} and λ_{TD} decrease beyond 10^{-3} . The best performance is obtained when $\lambda_{CT} = \lambda_{TD} = 10^{-3}$.

Impact of γ_T . We fix other parameters to the above values and vary γ_T . The results show (Fig 2D) that both AUC and AUPR values improve with lower values of γ_T . However, when the value of γ_T is less than 10^{-4} , the performance decreases. The best performance is obtained when γ_T is set to 10^{-4} .

Comparison with other methods

To investigate the performance of our proposed method, we compare its performance in terms of AUC and AUPR with that of existing methods including DRIMC [52], EMUDRA [53], LRSSL [54], and a tensor decomposition method [55] that we refer to as TDDR in this paper. Each method is configured with its defined settings and best parameter values as reported in its original study. Each method is then run with the same data (dataset P) described in Section “Data” using the three cross-validation scenarios. Based on the result and as can be seen in Figs 3–5, our proposed method outperforms all competing methods. Under the CV_{pairwise} scenario, NTD-DR obtains the best performance (AUC = 0.9338, AUPR = 0.9043) compared to that of other competing methods: DRIMC (AUC = 0.8516, AUPR = 0.8207), EMUDRA (AUC = 0.8619, AUPR = 0.8574), LRSSL (AUC = 0.8492, AUPR = 0.8329), and TDDR (AUC = 0.8962, AUPR = 0.8630) (Fig 3). TDDR is another tensor decomposition method and it achieves the second-best results in terms of AUC and AUPR.

The results under the $CV_{\text{column-wise}}$ scenario are presented in Fig 4. To summarize, NTD-DR achieves the best performance (AUC = 0.9139, AUPR = 0.9076), outperforming other competing methods: DRIMC (AUC = 0.8398, AUPR = 0.8148), EMUDRA (AUC = 0.8549, AUPR = 0.8328), LRSSL (AUC = 0.8329, AUPR = 0.8003), and TDDR (AUC = 0.8762, AUPR = 0.8513). Under this scenario, TDDR, an alternative tensor decomposition method, shows the second-best performance in terms of AUC and AUPR.

AUC (on the left) and AUPR (on the right) curves under $CV_{\text{column-wise}}$

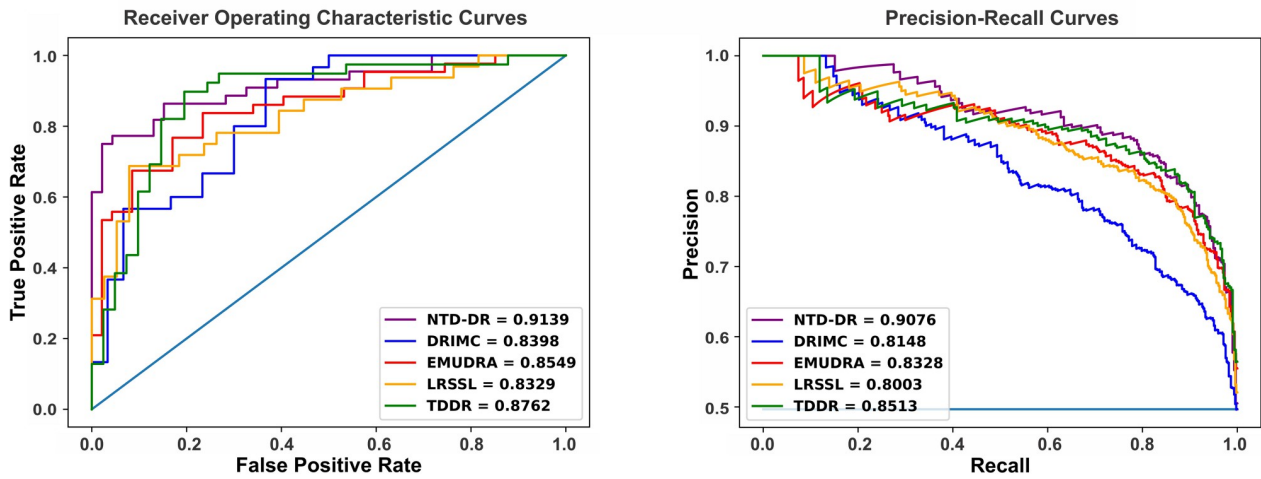


Fig 4. ROC and precision-recall curves of NTD-DR and the existing methods for Scenario $CV_{\text{column-wise}}$ using dataset P.

<https://doi.org/10.1371/journal.pone.0270852.g004>

Fig 5 shows the AUC and AUPR values for all competing methods under the the $CV_{\text{row-wise}}$. As shown, NTD-DR again obtains the best values for AUC = 0.9273 and AUPR = 0.8978 compared to the other methods: DRIMC (AUC = 0.8327, AUPR = 0.8269), EMUDRA (AUC = 0.8614, AUPR = 0.8548), LRSSL (AUC = 0.8219, AUPR = 0.8074), and TDDR (AUC = 0.8869, AUPR = 0.8653). Similar to the other scenarios, TDDR obtains the second-best results in terms of AUC and AUPR.

Case study

To further demonstrate the reliability of NTD-DR in discovering novel drug-disease associations, the case studies on breast ductal carcinoma, prostate cancer, pancreatic neoplasms,

AUC (on the left) and AUPR (on the right) curves under $CV_{\text{row-wise}}$

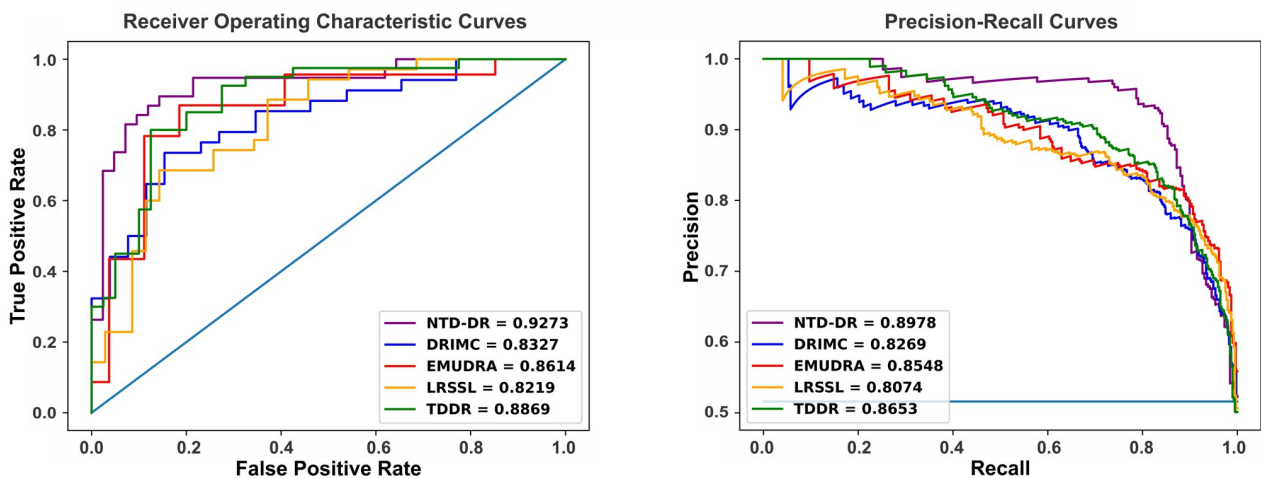


Fig 5. ROC and precision-recall curves of NTD-DR and the existing methods for scenario $CV_{\text{row-wise}}$ using dataset P.

<https://doi.org/10.1371/journal.pone.0270852.g005>

colorectal neoplasms, and small cell lung carcinoma within dataset I (as described in Section “Case study with an independent dataset”) are performed using the optimal parameter combination determined in Section “Performance evaluation” (i.e., $R = J$, $\alpha = 0.7$, $\lambda_{CT} = \lambda_{TD} = 10^{-3}$, and $\gamma_T = 10^{-4}$). The top 50 predictions according to the predicted association scores, which are drug candidates for the corresponding disease, are shown in S1–S5 Tables. In this study, we assume that if drug C interacts with target T and target T is associated with disease D, then drug C can be associated with the disease D (predicted association) and can be used for the treatment of disease D. We check whether the predictions were in the original dataset I. Those that are not, are deemed novel, hypothesized associations. We then perform a literature search for evidence to support the novel associations.

For breast ductal carcinoma, NTD-DR predicts 46 experimentally verified associations within its top 50 while the other competing methods predict fewer verified associations. TDDR is the second-best prediction method predicting 37 known associations out of its top 50 predictions. NTD-DR predicts four novel associations with lumiracoxib, etoricoxib, thimerosal, and cisplatin in its top 50 predictions. There is evidence in the literature that the first three drugs can be associated with breast carcinoma via a mutual protein.

Lumiracoxib and etoricoxib. The interaction of prostaglandin G/H synthase 2 with lumiracoxib and Etoricoxib were reported by Esser *et al.* [56] and Capone *et al.* [57], respectively. Also, the role of prostaglandin G/H synthase 2 in breast ductal carcinoma was reported by Saindane *et al.* [58]. Therefore, it can be hypothesized that lumiracoxib and etoricoxib are associated with breast ductal carcinoma, as predicted by NTD-DR, through prostaglandin G/H synthase 2.

Thimerosal. Stephenson *et al.* [59] found an interaction between thimerosal and superoxide dismutase. On the other hand, Kim *et al.* [60] reported the role of superoxide dismutase in breast ductal carcinoma. We can make a hypothesis that thimerosal is associated with breast ductal carcinoma, as is predicted with NTD-DR, via superoxide dismutase.

Cisplatin. There is no evidence in the literature for the association of cisplatin and breast ductal carcinoma. However, it could be that an as yet undiscovered intermediate protein connects cisplatin to this disease and experiments could be performed to find this link.

For prostate cancer, our method predicts 48 experimentally verified associations out of its top 50 predictions. It also identifies two novel associations involving docetaxel and paclitaxel that are supported by the literature.

Docetaxel and paclitaxel. Chaudhary *et al.* [61] reported the role of apoptosis regulator Bcl-2 in prostate cancer. On the other hand, interactions of apoptosis regulator Bcl-2 with docetaxel and paclitaxel were discovered by Marshall *et al.* [62] and Gan *et al.* [63], respectively. Therefore, it can be concluded that NTD-DR is justified in predicting associations of prostate cancer with docetaxel and paclitaxel. The second-best method for prostate cancer, TDDR, predicts 41 known associations within its top 50 predictions.

For pancreatic neoplasms, NTD-DR predicts 45 experimentally verified associations and five novel associations with stiripentol, pazopanib, ponatinib, sunitinib, and etoricoxib out of its top 50 predictions. All of these novel associations are supported with literature as follows.

Pazopanib, sunitinib, and ponatinib. Dineen *et al.* [64] discovered the association between pancreatic and vascular endothelial growth factor receptor 2. Moreover, Sonpavde *et al.* [65], Mendel *et al.* [66], and O’Hare *et al.* [67] reported interactions between vascular endothelial growth factor receptor 2 and pazopanib, sunitinib, and ponatinib, respectively. NTD-DR is able to identify the association between pancreatic neoplasm and these drugs through the intermediate target.

Stiripentol. Fisher *et al.* [68] reported an association between stiripentol and gamma-aminobutyric acid receptor subunit delta. On the other hand, the role of gamma-aminobutyric

Table 1. The number of known associations in the top 50 predictions made by different methods for different diseases.

	NTD-DR	DRIMC	EMUDRA	LRSSL	TDDR
Breast carcinoma	46	32	35	30	37*
Prostate cancer	48	36	38	37	41*
Pancreatic neoplasms	45	33	36*	31	36*
Colorectal neoplasms	50	37	39	36	42*
Small cell lung carcinoma	47	34	37	34	40*

The best method in the prediction of the most known associations is in **boldface**, and the second-best method is indicated with *.

<https://doi.org/10.1371/journal.pone.0270852.t001>

acid receptor subunit delta in pancreatic neoplasm was reported by Takehara *et al.* [69]. These findings support a hypothesis that stiripentol and pancreatic carcinoma is associated through an intermediate protein, gamma-aminobutyric acid receptor subunit delta.

Etoricoxib. The association between etoricoxib and prostaglandin G/H synthase 2 is reported by Capone *et al.* [57]. Moreover, Eibl *et al.* [70] found the role of prostaglandin G/H synthase 2 and pancreatic neoplasm. Based on these findings, we can hypothesize that etoricoxib is associated with pancreatic neoplasm via a mutual target, prostaglandin G/H synthase 2.

TDDR and EMUDRA, as the second-best methods, predict 36 known associations within their top 50 predictions.

For colorectal neoplasms, NTD-DR predicts 50 experimentally verified associations within its top 50 predictions while the second-best method, TDDR predicts 42 known associations within its top 50 predictions. Finally, for small cell lung carcinoma, our method predicts 47 experimentally verified associations and three novel associations with gefitinib, pazopanib, and afatinib within its top 50 predictions that are supported by literature as follows.

Gefitinib and afatinib. Sharma *et al.* [71] reported the association between epidermal growth factor receptor and small cell lung carcinoma. On the other hand, the interactions of the epidermal growth factor receptor with gefitinib and afatinib were reported by Ciardiello *et al.* [72] and Masood *et al.* [73], respectively.

Pazopanib. The interaction between pazopanib and endothelial growth factor receptor 2 and the association between endothelial growth factor receptor 2 and small cell lung carcinoma were reported by Ciardiello *et al.* [72] and Bonnesen *et al.* [74], respectively.

These findings confirm that our method can predict the associations between small cell lung carcinoma and above-mentioned drugs through corresponding targets. TDDR, as the second-best method, predicts 40 known associations within its top 50 predictions.

The results of these case studies confirm the biological and molecular hypotheses underlying NTD-DR since it can predict the most experimentally verified associations compared to the other methods (Table 1). Moreover, our method can uncover novel associations between drugs and disease implicit in the literature and which are facilitated by a mutual, experimentally verified target.

Discussion

In this study, we have proposed NTD-DR, a nonnegative tensor decomposition method, to discover drug-disease associations and enable drug repositioning using triplet associations of drugs, targets, and disease. First, NTD-DR uses pairwise drug-target, drug-disease, and target-disease associations to construct a order-three tensor. Then, to boost the performance of prediction, NTD-DR fuses multiple similarities for drugs, targets, and diseases to construct single similarity measures for drugs, targets, and diseases and later it integrates the similarities of

drugs, targets, and diseases with the decomposed tensor to make a prediction. We showed that NTD-DR outperforms existing, alternative state-of-the-art methods. Furthermore, to identify the reliability of NTD-DR, case study analyses were performed. The results confirm that our method can predict a large number of experimentally verified associations in its top 50 predictions. NTD-DR also predicted novel associations. We performed a literature search for evidence to support the novel predictions and found that most are linked together via a mutual target. Although this study focused on the identification of drug and disease associations, the proposed method can investigate associations between drugs and targets, or between targets and diseases by re-setting the value of parameters γ and λ in the model.

Wet-lab drug-disease and drug-target association identifications are time-consuming. NTD-DR can shorten the duration of these experiments. For instance, NTD-DR can reduce the search space and narrow down the set of drug-target and drug-disease trials to experimentally investigate for drug repositioning. This advantage of NTD-DR makes it a potential filtering approach not only for drug-target interactions, but also for drug-disease associations. We envision various research directions leading from this study. First, to increase the reliability of prediction in terms of biological validation, applying different types of pairwise associations can be used to construct the initial tensor. Second, the tensor decomposition requires huge computational effort for making a prediction, especially when the dimension of the tensor is large. In such a case, a paralleled tensor decomposition would increase the speed of computation. Finally, to make better use of predictions in health care and disease treatment, the predictions need to be validated biologically, experimentally, and pathologically.

In summary, NTD-DR could be effectively used as a reliable method to predict potential associations between drugs and diseases and provide a complementary tool to be used in drug discovery.

Supporting information

S1 Table. The top 50 predictions made by each method for breast ductal carcinoma.
(DOCX)

S2 Table. The top 50 predictions made by each method for prostate cancer.
(DOCX)

S3 Table. The top 50 predictions made by each method for pancreatic neoplasms.
(DOCX)

S4 Table. The top 50 predictions made by each method for colorectal neoplasms.
(DOCX)

S5 Table. The top 50 predictions made by each method for small cell lung carcinoma.
(DOCX)

Author Contributions

Conceptualization: Fang-Xiang Wu.

Formal analysis: Ali Akbar Jamali.

Funding acquisition: Anthony Kusalik, Fang-Xiang Wu.

Investigation: Ali Akbar Jamali, Anthony Kusalik, Fang-Xiang Wu.

Methodology: Ali Akbar Jamali, Yuting Tan, Fang-Xiang Wu.

Supervision: Anthony Kusalik, Fang-Xiang Wu.

Validation: Ali Akbar Jamali, Fang-Xiang Wu.

Writing – original draft: Ali Akbar Jamali.

Writing – review & editing: Ali Akbar Jamali, Yuting Tan, Anthony Kusalik, Fang-Xiang Wu.

References

1. Jamali AA, Ferdousi R, Razzaghi S, Li J, Safdari R, Ebrahimie E. DrugMiner: comparative analysis of machine learning algorithms for prediction of potential druggable proteins. *Drug Discovery Today*. 2016; 21(5):718–24. <https://doi.org/10.1016/j.drudis.2016.01.007> PMID: 26821132
2. Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, et al. Drug repositioning: a machine-learning approach through data integration. *Journal of Cheminformatics*. 2013; 5(1):30. Epub 2013/06/27. <https://doi.org/10.1186/1758-2946-5-30> PMID: 23800010; PubMed Central PMCID: PMC3704944.
3. Wang Y, Chen S, Deng N, Wang Y. Drug Repositioning by Kernel-Based Integration of Molecular Structure, Molecular Activity, and Phenotype Data. *PloS One*. 2013; 8(11):e78518. <https://doi.org/10.1371/journal.pone.0078518> PMC3823875. PMID: 24244318
4. Kim E, Choi As, Nam H. Drug repositioning of herbal compounds via a machine-learning approach. *BMC Bioinformatics*. 2019; 20(10):247. <https://doi.org/10.1186/s12859-019-2811-8> PMID: 31138103
5. Yan X-Y, Zhang S-W, He C-R. Prediction of drug-target interaction by integrating diverse heterogeneous information source with multiple kernel learning and clustering methods. *Computational Biology and Chemistry*. 2019; 78:460–7. <https://doi.org/10.1016/j.compbiolchem.2018.11.028> PMID: 30528728
6. Luo H, Wang J, Li M, Luo J, Peng X, Wu F-X, et al. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics*. 2016; 32(17):2664–71. <https://doi.org/10.1093/bioinformatics/btw228> PMID: 27153662
7. Luo H, Li M, Yang M, Wu F-X, Li Y, Wang J. Biomedical data and computational models for drug repositioning: a comprehensive review. *Briefings in Bioinformatics*. 2020:1–16. <https://doi.org/10.1093/bib/bbz176> PMID: 32043521
8. Xuan P, Cao Y, Zhang T, Wang X, Pan S, Shen T. Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics*. 2019; 35(20):4108–19. <https://doi.org/10.1093/bioinformatics/btz182> PMID: 30865257
9. Zhang W, Yue X, Lin W, Wu W, Liu R, Huang F, et al. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinformatics*. 2018; 19(1):233. <https://doi.org/10.1186/s12859-018-2220-4> PMID: 29914348
10. Dai W, Liu X, Gao Y, Chen L, Song J, Chen D, et al. Matrix Factorization-Based Prediction of Novel Drug Indications by Integrating Genomic Space. *Computational and Mathematical Methods in Medicine*. 2015; 2015:9. <https://doi.org/10.1155/2015/275045> PMID: 26078775
11. Jamali AA, Kusalik A, Wu F-X. MDIPA: a microRNA–drug interaction prediction approach based on non-negative matrix factorization. *Bioinformatics*. 2020; 36(20):5061–7. <https://doi.org/10.1093/bioinformatics/btaa577> PMID: 33212495
12. Li S, Pan X. A computational drug repositioning model based on hybrid similarity side information powered graph neural network. *Future Generation Computer Systems*. 2021; 125:24–31. <https://doi.org/10.1016/j.future.2021.06.018>.
13. Jin S, Niu Z, Jiang C, Huang W, Xia F, Jin X, et al. HeTDR: Drug repositioning based on heterogeneous networks and text mining. *Patterns*. 2021:100307. <https://doi.org/10.1016/j.patter.2021.100307> PMID: 34430926
14. Ding Y, Tang J, Guo F. Identification of Drug–Target Interactions via Dual Laplacian Regularized Least Squares with Multiple Kernel Fusion. *Knowledge-Based Systems*. 2020; 204:106254. <https://doi.org/10.1016/j.knosys.2020.106254>.
15. Lotfi Shahreza M, Ghadiri N, Mousavi SR, Varshosaz J, Green JR. A review of network-based approaches to drug repositioning. *Briefings in Bioinformatics*. 2017. Epub 2017/03/24. <https://doi.org/10.1093/bib/bbx017> PMID: 28334136.
16. Brown AS, Patel CJ. MeSHDD: Literature-based drug-drug similarity for drug repositioning. *Journal of the American Medical Informatics Association: JAMIA*. 2017; 24(3):614–8. <https://doi.org/10.1093/jamia/ocw142> PMID: 27678460.

17. Venkatasubbaiah M, Dwarakanadha Reddy P, Satyanarayana SV. Literature-based review of the drugs used for the treatment of COVID-19. *Current Medicine Research and Practice*. 2020; 10(3):100–9. <https://doi.org/10.1016/j.cmrp.2020.05.013> PMID: 32572376
18. Wang F, Ding Y, Lei X, Liao B, Wu FX. Human Protein Complex-Based Drug Signatures for Personalized Cancer Medicine. *IEEE Journal of Biomedical and Health Informatics*. 2021; 25(11):4079–88. <https://doi.org/10.1109/JBHI.2021.3120933> PMID: 34665747
19. Yang C, Zhang H, Chen M, Wang S, Qian R, Zhang L, et al. A survey of optimal strategy for signature-based drug repositioning and an application to liver cancer. *eLife*. 2022; 11:e71880. <https://doi.org/10.7554/eLife.71880> PMID: 35191375
20. Wang F, Ding Y, Lei X, Liao B, Wu FX. Identifying Gene Signatures for Cancer Drug Repositioning Based on Sample Clustering. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2022; 19(2):953–65. <https://doi.org/10.1109/TCBB.2020.3019781> PMID: 32845842
21. 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(11):1904–12. <https://doi.org/10.1093/bioinformatics/bty013> PMID: 29365057
22. He S, Wen Y, Yang X, Liu Z, Song X, Huang X, et al. PIMD: An Integrative Approach for Drug Repositioning Using Multiple Characterization Fusion. *Genomics, Proteomics & Bioinformatics*. 2020. <https://doi.org/10.1016/j.gpb.2018.10.012> PMID: 33075523
23. Huang F, Yue X, Xiong Z, Yu Z, Liu S, Zhang W. Tensor decomposition with relational constraints for predicting multiple types of microRNA-disease associations. *Briefings in Bioinformatics*. 2020; 22(3). <https://doi.org/10.1093/bib/bbaa140> PMID: 32725161
24. Luo J, Liu Y, Liu P, Lai Z, Wu H. Data Integration Using Tensor Decomposition for The Prediction of miRNA-Disease Associations. *IEEE Journal of Biomedical and Health Informatics*. 2021:1-. <https://doi.org/10.1109/JBHI.2021.3125573> PMID: 34748505
25. Chen H, Li J. Modeling Relational Drug-Target-Disease Interactions via Tensor Factorization with Multiple Web Sources. *The World Wide Web Conference*; San Francisco, CA, USA: Association for Computing Machinery; 2019. p. 218–27.
26. 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(26):628. <https://doi.org/10.1186/s12859-019-3283-6> PMID: 31839008
27. Chen H, Li J, null nE. Learning Data-Driven Drug-Target-Disease Interaction via Neural Tensor Network. *Journal Name: International Joint Conference on Artificial Intelligence (IJCAI)*. 2020:Medium: X; Size: 3452 to 8. OSTI ID: 10297213.
28. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*. 2006; 34(Database issue):D668–72. Epub 2005/12/31. <https://doi.org/10.1093/nar/gkj067> PMID: 16381955; PubMed Central PMCID: PMC1347430.
29. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*. 2018; 46(5):2699–. <https://doi.org/10.1093/nar/gky092> PMID: 29425356
30. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, et al. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Research*. 2007; 36(suppl_1):D919–D22. <https://doi.org/10.1093/nar/gkm862> PMID: 17942422
31. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Research*. 2018; 47(D1):D1038–D43. <https://doi.org/10.1093/nar/gky1151> PMID: 30445645
32. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wiegiers J, Wiegiers TC, et al. Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Research*. 2021; 49(D1):D1138–d43. Epub 2020/10/18. <https://doi.org/10.1093/nar/gkaa891> PMID: 33068428; PubMed Central PMCID: PMC7779006.
33. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*. 2019; 48(D1):D845–D55. <https://doi.org/10.1093/nar/gkz1021> PMID: 31680165
34. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nature Genetics*. 2004; 36(5):431–2. Epub 2004/05/01. <https://doi.org/10.1038/ng0504-431> PMID: 15118671.
35. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*. 2014; 42(Database issue):D358–63. <https://doi.org/10.1093/nar/gkt1115> PMID: 24234451.
36. Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Research*. 2017; 45(D1):D369–D79. Epub 12/14. <https://doi.org/10.1093/nar/gkw1102> PMID: 27980099.

37. Palma A, Nardoza AP, Peluso D, Santonico E, Galeota E, Sacco F, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Research*. 2011; 40(D1):D857–D61. <https://doi.org/10.1093/nar/gkr930> PMID: 22096227
38. Gable AL, Szklarczyk D, Lyon D, Simonovic M, Wyder S, Mering Christian v, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. 2018; 47(D1):D607–D13. <https://doi.org/10.1093/nar/gky1131> PMID: 30476243
39. Venugopal A, Sebastian A, Marimuthu A, Rahiman BA, Shafreen B, Harrys Kishore CJ, et al. Human Protein Reference Database—2009 update. *Nucleic Acids Research*. 2008; 37(suppl_1):D767–D72. <https://doi.org/10.1093/nar/gkn892> PMID: 18988627
40. Jamali AA, Kusalik A, Wu F. NMTF-DTI: A Nonnegative Matrix Tri-factorization Approach with Multiple Kernel Fusion for Drug-Target Interaction Prediction. *IEEE/ACM Trans Comput Biol Bioinform*. 2021:1–. <https://doi.org/10.1109/TCBB.2021.3135978> PMID: 34914594
41. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods*. 2014; 11(3):333–7. <https://doi.org/10.1038/nmeth.2810> PMID: 24464287
42. Kolda T, Bader B. Tensor Decompositions and Applications. *SIAM Rev*. 2009; 51:455–500.
43. Li L, Wu L, Zhang H, Wu F. A Fast Algorithm for Nonnegative Matrix Factorization and Its Convergence. *IEEE Transactions on Neural Networks and Learning Systems*. 2014; 25(10):1855–63. <https://doi.org/10.1109/TNNLS.2013.2296627> PMID: 25291738
44. Tian L-P, Luo P, Wang H, Zheng H, Wu F-X. CASNMF: A Converged Algorithm for symmetrical non-negative matrix factorization. *Neurocomputing*. 2018; 275:2031–40. <https://doi.org/10.1016/j.neucom.2017.10.039>
45. Luo P, Tian L, Ruan J, Wu F. Disease Gene Prediction by Integrating PPI Networks, Clinical RNA-Seq Data and OMIM Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2019; 16(1):222–32. <https://doi.org/10.1109/TCBB.2017.2770120> PMID: 29990218
46. Svd Walt, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*. 2011; 13(2):22–30. <https://doi.org/10.1109/MCSE.2011.37>
47. McKinney W. Pandas, python data analysis library. 2015. Reference Source. 2014.
48. Hunter JD. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*. 2007; 9(3):90–5. <https://doi.org/10.1109/MCSE.2007.55>
49. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020; 17(3):261–72. <https://doi.org/10.1038/s41592-019-0686-2> PMID: 32015543
50. Hagberg AA, Schult DA, Swart PJ. Exploring Network Structure, Dynamics, and Function using NetworkX. *Proceedings of the 7th Python in Science conference (SciPy 2008)*. 2009:11–5.
51. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011; 12:2825–30. <https://doi.org/10.1093/bioinformatics/btp336>
52. Zhang W, Xu H, Li X, Gao Q, Wang L. DRIMC: an improved drug repositioning approach using Bayesian inductive matrix completion. *Bioinformatics*. 2020; 36(9):2839–47. <https://doi.org/10.1093/bioinformatics/btaa062> PMID: 31999326
53. Zhou X, Wang M, Katsyv I, Irie H, Zhang B. EMUDRA: Ensemble of Multiple Drug Repositioning Approaches to improve prediction accuracy. *Bioinformatics*. 2018; 34(18):3151–9. <https://doi.org/10.1093/bioinformatics/bty325> PMID: 29688306
54. Liang X, Zhang P, Yan L, Fu Y, Peng F, Qu L, et al. LRSSL: predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics*. 2017; 33(8):1187–96. <https://doi.org/10.1093/bioinformatics/btw770> PMID: 28096083
55. Wang R, Li S, Wong MH, Leung KS, editors. Drug-Protein-Disease Association Prediction and Drug Repositioning Based on Tensor Decomposition. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2018 3–6 Dec. 2018.
56. Esser R, Berry C, Du Z, Dawson J, Fox A, Fujimoto RA, et al. Preclinical pharmacology of lumiracoxib: a novel selective inhibitor of cyclooxygenase-2. *British Journal of Pharmacology*. 2005; 144(4):538–50. <https://doi.org/10.1038/sj.bjp.0706078> PMID: 15655513.
57. Capone ML, Tacconelli S, Di Francesco L, Sacchetti A, Sciulli MG, Patrignani P. Pharmacodynamic of cyclooxygenase inhibitors in humans. *Prostaglandins and Other Lipid Mediators*. 2007; 82(1):85–94. <https://doi.org/10.1016/j.prostaglandins.2006.05.019> PMID: 17164136
58. Saindane M, Rallabandi HR, Park KS, Heil A, Nam SE, Yoo YB, et al. Prognostic Significance of Prostaglandin-Endoperoxide Synthase-2 Expressions in Human Breast Carcinoma: A Multiomic Approach.

- Cancer Informatics. 2020; 19:1176935120969696. <https://doi.org/10.1177/1176935120969696> PMID: 33223820.
59. Stephenson GA, Stowell OG, Toma PH, Pfeiffer RR, Byrn SR. Solid-State Investigations of Erythromycin A Dihydrate: Structure, NMR Spectroscopy, and Hygroscopicity. *Journal of Pharmaceutical Sciences*. 1997; 86(11):1239–44. <https://doi.org/10.1021/js9701667> PMID: 9383733
 60. Kim YS, Gupta Vallur P, Phaëton R, Mythreye K, Hempel N. Insights into the Dichotomous Regulation of SOD2 in Cancer. *Antioxidants (Basel, Switzerland)*. 2017; 6(4):86. <https://doi.org/10.3390/antiox6040086> PMID: 29099803.
 61. Chaudhary KS, Abel PD, Lalani EN. Role of the Bcl-2 gene family in prostate cancer progression and its implications for therapeutic intervention. *Environmental Health Perspectives*. 1999; 107 Suppl 1(Suppl 1):49–57. <https://doi.org/10.1289/ehp.99107s149> PMID: 10229706.
 62. Marshall J, Chen H, Yang D, Figueira M, Bouker KB, Ling Y, et al. A phase I trial of a Bcl-2 antisense (G3139) and weekly docetaxel in patients with advanced breast cancer and other solid tumors. *Annals of Oncology*. 2004; 15(8):1274–83. <https://doi.org/10.1093/annonc/mdh317> PMID: 15277270
 63. Gan Y, Wientjes MG, Au JLS. Expression of Basic Fibroblast Growth Factor Correlates with Resistance to Paclitaxel in Human Patient Tumors. *Pharmaceutical Research*. 2006; 23(6):1324–31. <https://doi.org/10.1007/s11095-006-0136-6> PMID: 16741658
 64. Dineen SP, Lynn KD, Holloway SE, Miller AF, Sullivan JP, Shames DS, et al. Vascular Endothelial Growth Factor Receptor 2 Mediates Macrophage Infiltration into Orthotopic Pancreatic Tumors in Mice. *Cancer Research*. 2008; 68(11):4340–6. <https://doi.org/10.1158/0008-5472.CAN-07-6705> PMID: 18519694
 65. Sonpavde G, Hutson TE. Pazopanib: a novel multitargeted tyrosine kinase inhibitor. *Current Oncology Reports*. 2007; 9(2):115–9. <https://doi.org/10.1007/s11912-007-0007-2> PMID: 17288876.
 66. Mendel DB, Laird AD, Xin X, Louie SG, Christensen JG, Li G, et al. In vivo antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: determination of a pharmacokinetic/pharmacodynamic relationship. *Clinical Cancer Research*. 2003; 9(1):327–37. PMID: 12538485.
 67. O'Hare T, Shakespeare WC, Zhu X, Eide CA, Rivera VM, Wang F, et al. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell*. 2009; 16(5):401–12. <https://doi.org/10.1016/j.ccr.2009.09.028> PMID: 19878872; PubMed Central PMCID: PMC2804470.
 68. Fisher JL. The anti-convulsant stiripentol acts directly on the GABA(A) receptor as a positive allosteric modulator. *Neuropharmacology*. 2009; 56(1):190–7. Epub 2008/06/10. <https://doi.org/10.1016/j.neuropharm.2008.06.004> PMID: 18585399.
 69. Takehara A, Hosokawa M, Eguchi H, Ohigashi H, Ishikawa O, Nakamura Y, et al. Gamma-aminobutyric acid (GABA) stimulates pancreatic cancer growth through overexpressing GABAA receptor pi subunit. *Cancer Research*. 2007; 67(20):9704–12. <https://doi.org/10.1158/0008-5472.CAN-07-2099> PMID: 17942900.
 70. Eibl G, Bruemmer D, Okada Y, Duffy JP, Law RE, Reber HA, et al. PGE2 is generated by specific COX-2 activity and increases VEGF production in COX-2-expressing human pancreatic cancer cells. *Biochemical and Biophysical Research Communications*. 2003; 306(4):887–97. [https://doi.org/10.1016/s0006-291x\(03\)01079-9](https://doi.org/10.1016/s0006-291x(03)01079-9) PMID: 12821125
 71. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. *Nature Reviews: Cancer*. 2007; 7(3):169–81. <https://doi.org/10.1038/nrc2088> PMID: 17318210.
 72. Ciardiello F, Caputo R, Bianco R, Damiano V, Pomatice G, De Placido S, et al. Antitumor effect and potentiation of cytotoxic drugs activity in human cancer cells by ZD-1839 (Iressa), an epidermal growth factor receptor-selective tyrosine kinase inhibitor. *Clinical Cancer Research*. 2000; 6(5):2053–63. PMID: 10815932.
 73. Masood A, Kancha RK, Subramanian J. Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors in non-small cell lung cancer harboring uncommon EGFR mutations: Focus on afatinib. *Seminars in Oncology*. 2019; 46(3):271–83. Epub 20190911. <https://doi.org/10.1053/j.seminoncol.2019.08.004> PMID: 31558282.
 74. Bonnesen B, Pappot H, Holmstov J, Skov BG. Vascular endothelial growth factor A and vascular endothelial growth factor receptor 2 expression in non-small cell lung cancer patients: Relation to prognosis. *Lung Cancer*. 2009; 66(3):314–8. <https://doi.org/10.1016/j.lungcan.2009.02.013> PMID: 19324448