A computational method for the identification of candidate drugs for non-small cell lung cancer

Lei Chen[1,2☯], Jing Lu[3☯], Tao Huang[4], Yu-Dong Cai[1]*

1 College of Life Science, Shanghai University, Shanghai, People's Republic of China, **2** College of Information Engineering, Shanghai Maritime University, Shanghai, People's Republic of China, **3** School of Pharmacy, Key Laboratory of Molecular Pharmacology and Drug Evaluation (Yantai University), Ministry of Education, Collaborative Innovation Center of Advanced Drug Delivery System and Biotech Drugs in Universities of Shandong, Yantai University, Yantai, People's Republic of China, **4** Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China

☯ These authors contributed equally to this work.
* cai_yud@126.com

Abstract

Lung cancer causes a large number of deaths per year. Until now, a cure for this disease has not been found or developed. Finding an effective drug through traditional experimental methods invariably costs millions of dollars and takes several years. It is imperative that computational methods be developed to integrate several types of existing information to identify candidate drugs for further study, which could reduce the cost and time of development. In this study, we tried to advance this effort by proposing a computational method to identify candidate drugs for non-small cell lung cancer (NSCLC), a major type of lung cancer. The method used three steps: (1) preliminary screening, (2) screening compounds by an association test and a permutation test, (3) screening compounds using an EM clustering algorithm. In the first step, based on the chemical-chemical interaction information reported in STITCH, a well-known database that reports interactions between chemicals and proteins, and approved NSCLC drugs, compounds that can interact with at least one approved NSCLC drug were picked. In the second step, the association test selected compounds that can interact with at least one NSCLC-related chemical and at least one NSCLC-related gene, and subsequently, the permutation test was used to discard nonspecific compounds from the remaining compounds. In the final step, core compounds were selected using a powerful clustering algorithm, the EM algorithm. Six putative compounds, protoporphyrin IX, hematoporphyrin, canertinib, lapatinib, pelitinib, and dacomitinib, were identified by this method. Previously published data show that all of the selected compounds have been reported to possess anti-NSCLC activity, indicating high probabilities of these compounds being novel candidate drugs for NSCLC.

1. Introduction

Lung cancer is a major cause of cancer-related deaths worldwide [1], and the number of deaths has shown an increasing trend over the past fifteen years [2] despite improvements in research and development (R&D) and increased investments in R&D. Therefore, drug discovery for treating lung cancers important. Lung cancers comprise two major types, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for more than 85% of lung cancer cases [3], and most approved drugs, such as gefitinib, cisplatin and paclitaxel, are used to treat NSCLC.

Experimental testing during drug R&D costs millions of dollars and takes several years, and only a few drugs meet the activity and safety requirements for regulatory approval. In silico methods for early assessment are attractive for improving the success rates and reducing the costs of R&D. Many previous studies based on in silico predictions have been carried out to analyze the structure-activity relationships (SARs) of anti-NSCLC chemicals and identify promising chemicals that can act as substitutes for approved NSCLC drugs. Lang *et al.* reported that the cytotoxic activities of cucurbitacins against NSCLC A549 cells were associated with their propensity for electrophilic attack, molecular size and shape in a QSAR study [4]. Goyal *et al.* developed a 3D-QSAR model using 38 thiazolyl-pyrazoline compounds against EGFR, which is a target associated with NSCLC, and obtained two novel inhibitors by screening ZINC libraries [5]. Xiang *et al.* proposed a novel "hybrid strategy" by using a chemically reactive feature and a pharmacophore feature and identified 38 irreversible EGFR-T790M inhibitors [6]. The above methods primarily used the structures of chemicals to discover compounds that have anti-NSCLC activity. Recently, Lu *et al.* developed a novel computational model by using chemical-/protein-chemical interaction information and identified promising chemicals with potential anti-NSCLC activity that were structurally dissimilar to drugs approved for NSCLC [7]. However, the effectiveness of the method was not very high. Of the nineteen compounds identified, only six were found to have anti-NSCLC activity. Although this method needs to be improved, the concept of identifying drug candidates by integrating chemical-/protein-chemical interactions is a suitable approach. Since there are various mutations in different genes, a drug will only be effective if it can target the appropriate disease genes. From the perspective of precision medicine, different patients should receive different treatment regimens. If protein-chemical interactions are characterized during drug screening, drugs, once validated and approved, can be prescribed to patients of a certain subtype for a more precise treatment. Therefore, we tried to extend this method by using additional related information and more powerful computational tools.

In this study, we proposed an improved computational method for the identification of novel candidate drugs of NSCLC. For execution and analysis, sixteen approved NSCLC drugs, NSCLC-related chemicals, NSCLC-related genes and chemical-/protein-chemical interactions were retrieved from public websites and databases. The method consisted of three steps. In the first step, namely, preliminary screening, possible compounds were extracted by checking the chemical-chemical interactions involving approved NSCLC drugs. In the second step, these compounds were filtered by an association test and a permutation test, where the association test helps us select compounds that have associations with both NSCLC-related chemicals and NSCLC-related genes, while the permutation test can exclude nonspecific compounds that are not associated with NSCLC. Finally, the remaining compounds were analyzed using a cluster algorithm, the EM clustering algorithm, to further select core compounds. As a result, six compounds, protoporphyrin IX, hematoporphyrin, canertinib, lapatinib, pelitinib, and dacomitinib, were identified. Data from previously published reports indicate that all of these compounds have anti-NSCLC activity, implying that there is a high probability that they may

be candidate drugs for NSCLC. Moreover, canertinib, lapatinib, pelitinib, and dacomitinib were confirmed to be effective for NSCLC associated with mutations in *EGFR*, which can help formulate guidelines for the precise medical treatment of NSCLC involving these mutations.

2. Materials and methods

2.1 Approved NSCLC drugs and chemicals as well as genes related to NSCLC

2.1.1 Approved NSCLC drugs. We accessed sixteen approved NSCLC drugs from the following two websites: (1) http://www.cancer.gov/cancertopics/druginfo/lungcancer (accessed in January 2016); (2) http://www.medindia.net/drugs/medical-condition/lungcancer.htm (accessed on May 11, 2014). Detailed information about these sixteen drugs, including their mechanism [8–23], is provided in **Table 1**.

2.1.2 NSCLC-related chemicals. A total of 3,793 NSCLC-related chemicals were downloaded from the Comparative Toxicogenomics Database (CTD) (http://ctdbase.org/detail.go?type=disease&acc=MESH:D002289&view=chem, accessed in March 2015) [24], for which the disease and chemical relationships were manually extracted from the literature. After mapping

Table 1. Sixteen approved NSCLC drugs.

PubChem ID	Name	Mechanism	NSCLC-related Genes
CID4033	Mechlorethamine	Agents directly acting on DNA	*CASP9, ERBB2, ERCC1, TP53*
CID36314	Paclitaxel	Antimitotic agent	*ABCB1, ABCB11, AKT1, BIRC5, CSF3, EGFR, ERBB2, KIF5B, TP53, VEGFA*
CID38904	Carboplatin	Agents directly acting on DNA	*ABCB1, AKT1, ALK, AREG, AREGB, BIRC5, CASP8, CASP9, CCND1, CDKN2A, CSF3, EGFR, EML4, ERBB2, ERCC1, GSTP1, HRAS, KRAS, MDM2, MET, NQO1, NRAS, PIK3CA, PIK3CB, RASSF1, RRM1, STAT3, STK4, TGFA, TP53, VEGFA, VEGFC*
CID57166	Porfimer Sodium	Photosensitizing agent	*PRDX1*
CID60750	Gemcitabine	Agents interfering with DNA synthesis	-
CID60780	Vinorelbine	Antimitotic agent	-
CID60843	Pemetrexed	Agents interfering with DNA synthesis	*AKT1, ALK, EGFR, EML4, ERBB2, ERCC1, HRAS, KRAS, NRAS, RRM1, TP53, VEGFA*
CID123631	Gefitinib	EGFR inhibitor	*ABCB1, ABCB11, AKT1, AKT2, ALK, AREG, AREGB, AXL, BIRC5, CASP8, CCND1, CDK4, CDKN2A, E2F1, EGFR, EML4, ERBB2, ERCC1, FASLG, FGFR1, FOXM1, FOXO3, GRB2, HBEGF, HRAS, KRAS, MAPK1, MAPK10, MAPK13, MAPK14, MAPK3, MDM2, MET, MMP9, NRAS, PIK3CA, PIK3CB, PTHLH, RASSF1, RB1, STAT3, STK4, TERT, TGFA, TP53, VEGFA, VEGFC*
CID126941	Methotrexate	Agents interfering with DNA synthesis	*ABCB1, ABCB11, AKT1, ALK, CASP8, CCND1, CD93, CDKN2A, CSF3, DNMT3L, EGFR, ENO1, ERBB2, FASLG, GRB2, HRAS, IL10, IL6R, LILRB1, MAPK14, MDM2, NCOA3, PROS1, RB1, TP53, TRAF1, VEGFA*
CID148124	Docetaxel	Antimitotic agent	-
CID176870	Erlotinib	EGFR inhibitor	*ABCB1, AKT1, ALK, AREG, AREGB, AXL, BIRC5, CCND1, CDKN2A, EGF, EGFR, EML4, ERBB2, ERCC1, FASLG, GRB2, HBEGF, HRAS, KRAS, MAPK1, MAPK3, MET, NRAS, PIK3CA, PIK3CB, STAT3, STK4, TGFA, TP53, VEGFA, VEGFC*
CID441203	Cisplatin	Agents directly acting on DNA	-
CID5360373	Bleomycin	Agents directly acting on DNA	-
CID10184653	Afatinib	EGFR/HER2/HER4 inhibitor	*ALK, EGF, EGFR, EML4, ERBB2, HRAS, IL6R, KRAS, MET, NRAS, PIK3CA*
CID11626560	Crizotinib	ALK/HGFR inhibitor	*AKT1, ALK, AXL, BIRC5, EGF, EGFR, EML4, ERBB2, HBEGF, HRAS, KIF5B, KRAS, MAPK1, MAPK3, MET, NRAS, PIK3CA, ROS1, STAT3, STK4, TGFA*
CID57379345	Ceritinib	ALK antagonist	-

https://doi.org/10.1371/journal.pone.0183411.t001

these chemicals to their PubChem IDs, 3,085 chemicals were retained; these chemicals comprised the dataset S_c and are listed in S1 Table.

2.1.3 NSCLC-related genes. We identified NSCLC-related genes using the following two public databases: (1) Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) [25, 26]; (2) CTD [24]. More specifically, from KEGG, 54 genes associated with NSCLC-related pathways were retrieved (accessed in February 2014), and from CTD, we identified 104 NSCLC-related genes for which there was direct evidence of association with NSCLC (accessed in March 2015). After combining these two sets of NSCLC-related genes, 148 genes were obtained; these genes comprised the dataset S_g and are listed in S2 Table.

2.2 Chemical-/protein-chemical interaction

The basis of our method for identifying candidate drugs for NSCLC is to discover compounds that have similar functions as approved NSCLC drugs and close relationships with NSCLC-related chemicals and genes. To implement the method, we mined databases for chemical-chemical interactions and protein-chemical interactions. This section provides a brief description of our approach.

2.2.1 Chemical-chemical interaction. This information was retrieved from the Search Tool for Interactions of Chemicals (STITCH, http://stitch.embl.de/) [27], a well-known public database that catalogs large numbers of interactions between chemicals and proteins. Chemicals are linked to other chemicals according to the evidence derived from experiments, databases and the literature. This type of chemical-chemical interaction information is widely used to investigate several biological problems [7, 28–36]. We downloaded a file, named "chemical_chemical.links.detailed.v4.0.tsv.gz", from STITCH (Version 4.0), which lists large numbers of chemical-chemical interactions. For each interaction, there are two PubChem IDs and five scores labeled "Similarity", "Experimental", "Database", "Textmining" and "Combined_score", respectively. The "Similarity", "Experimental", "Database", and "Textmining" scores are obtained by examining the structures, activities, reactions and co-occurrence in the literature of chemicals, respectively. Finally, the "Combined_score" was determined by integrating all of the aforementioned scores. To formulate this mathematically, let us denote the above five scores for chemicals c_1 and c_2 using $Q_S^{cc}(c_1, c_2)$, $Q_E^{cc}(c_1, c_2)$, $Q_D^{cc}(c_1, c_2)$, $Q_T^{cc}(c_1, c_2)$ and $Q_C^{cc}(c_1, c_2)$. Because the "Combined_score" can widely indicate associations between chemicals, it was used here to indicate the interactiveness of two chemicals, *i.e.*, two chemicals were deemed to interact with each other if and only if the "Combined_score" between them was greater than zero.

2.2.2 Protein-chemical interaction. In addition to chemical-chemical interactions, STITCH also contains information on protein-chemical interactions. This information has also been applied to investigate many biological problems [7, 36–41]. We downloaded the file "protein_chemical.links.detailed.v4.0.tsv.gz" from this database (Version 4.0), in which the interactions between chemicals and proteins from 1,133 organisms were collected. From the obtained file, we extracted the interactions involving human proteins by selecting lines containing "9606" that is the code of Homo sapiens in STITCH. For each extracted interaction, there is one chemical, represented by a PubChem ID; one protein, represented by an Ensembl ID; and five scores, labeled "Experimental", "Prediction", "Database", "Textmining" and "Combined_score", respectively. To formulate this mathematically, we used $Q_E^{pc}(p, c)$, $Q_P^{pc}(p, c)$, $Q_D^{pc}(p, c)$, $Q_T^{pc}(p, c)$ and $Q_C^{pc}(p, c)$ to denote the scores between protein p and chemical c. As above, the "Combined_score" was used to define the interactiveness of chemicals and proteins, *i.e.*, a chemical and a protein were deemed to interact with each other if and only if their "Combined_score" was greater than zero.
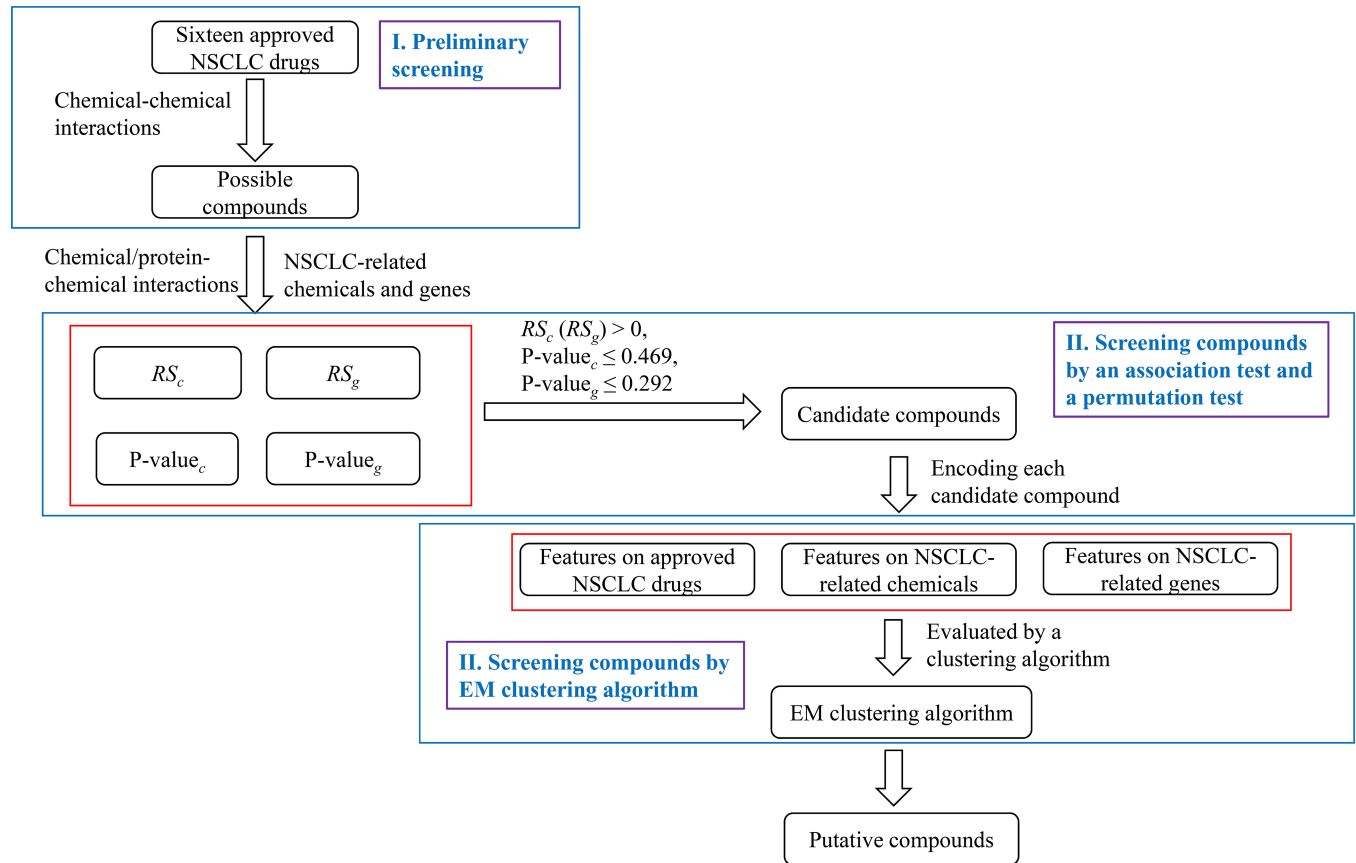
Fig 1. The flow chart of the computational method.

2.3 Method for identification of novel candidate drugs for NSCLC

This section provides a detailed description of the computational method for the identification of candidate drugs for NSCLC. This method consisted of three steps: (1) preliminary screening, (2) screening compounds by an association test and a permutation test, and (3) screening compounds by an EM clustering algorithm. A flow chart of the method is illustrated in **Fig 1** and the pseudo codes are provided in **Table 2**.

2.3.1 Preliminary screening. Many studies have reported that compounds that can interact with each other invariably share similar functions [7, 28–35]. As mentioned in Section 2.1.1, sixteen approved NSCLC drugs were retrieved from public websites. The compounds that can interact with these drugs are more likely to be potential drugs for NSCLC than those that cannot interact with any of them. Therefore, we obtained lists of compounds that can interact with at least one approved NSCLC drug by mining the chemical-chemical interactions database as described in Section 2.2.1. This list of compounds comprised a compound set denoted by P_c for convenience.

2.3.2 Screening compounds by an association test and a permutation test. After the preliminary screening, several possible compounds were obtained. It was almost impossible to test them individually through traditional experiments, and therefore, further screening was required. It was easy to determine if a candidate drug was closely related to the biological processes associated with NSCLC, including its related genes and chemicals. Here, we built an association test and a permutation test to screen for relevant compounds in P_c.

Table 2. The pseudo codes of the method for identification of novel candidate drugs for NSCLC.

Input: Approved NSCLC drugs, NSCLC-related chemicals S_c, NSCLC-related genes S_g, chemical/protein-chemical interactions retrieved from STITCH
Output: A list of putative compounds
1. Preliminary screening
1.1 Using the approved NSCLC drugs and chemical-chemical interactions, compounds that can interact with at least one approved NSCLC drug were obtained and comprised the set P_c.
2. Screening compounds by an association test and a permutation test
2.1 For each compound c in P_c, if it can interact with at least one NSCLC-related chemical and one NSCLC-related gene (i.e., $RS_c(c)>0$ and $RS_g(c)>0$), it was selected.
2.2 For each compound c obtained in step 2.1, do the permutation test and calculate the P-value for NSCLC-related chemicals $P-value_c(c)$ (cf. Eq 1) as well as P-value for NSCLC-related genes $P-value_g(c)$ (cf. Eq 2)
2.3 If a compound c was assigned $P-value_c(c)$ and $P-value_g(c)$ no more than the corresponding threshold, it was selected.
3. Screening compounds by the EM clustering algorithm
3.1 For each compound obtained in step 2, it was represented by fifteen features based on the interaction scores of protein/chemical-chemical interactions.
3.2 For each approved drug, it was also represented by fifteen features used in step 3.1.
3.3 The remaining compounds and approved drugs were fed into the EM clustering algorithm.
3.4 According to the clustering results, compounds in the same category as approved drugs were picked and outputted.

In the association test, compounds that had associations with NSCLC-related chemicals/genes were selected. Each compound c in P_c was linked to all NSCLC-related chemicals in S_c using chemical-chemical interactions as described in Section 2.2.1. If c can interact with at least one NSCLC-related chemical, then it was selected because it is highly related to at least one NSLC-related chemical. On the other hand, we also screened compounds in P_c using NSCLC-related genes in S_g. More specifically, each compound c in P_c was linked to all of the NSCLC-related genes in S_g using protein-chemical interactions as described in Section 2.2.2. The compounds that could interact with at least one gene in S_g were selected. By considering both the NSCLC-related chemicals and genes, we selected compounds in P_c that could interact with at least one NSCLC-related chemical and gene. Additionally, for a later evaluation of the importance of the selected compound c, we extracted all of the chemical-chemical interactions between c and NSCLC-related chemicals. The mean value of the "Combined_score" of these interactions was calculated, which was called the rating score of c for NSCLC-related chemicals and denoted by $RS_c(c)$. Similarly, we also extracted all of the protein-chemical interactions between c and NSCLC-related genes. The mean value of their "Combined_score" was calculated. This value was called the rating score of c for NSCLC-related genes and denoted by $RS_g(c)$. In a formalization way, we selected compounds in P_c with $RS_c>0$ and $RS_g>0$ in the association test.

The association test helps us select compounds that have associations with NSCLC-related chemicals/genes. However, some compounds may have unusual properties and may interact nonspecifically with several compounds and genes. However, they may not be linked to NSCLC. Therefore, we built a permutation test to evaluate each compound that passed the association test and excluded these types of compounds. Let c be a compound that passed the association test. For the NSCLC-related chemical set S_c, we randomly produced 1,000 sets with the same size as S_c, which were denoted as $S_c^1, S_c^2, \ldots, S_c^{1000}$. For each set S_c^i ($i = 1,2,\ldots,1000$), we calculated the rating score of c using the procedures mentioned above. Thus, c was assigned one rating score based on its associations in S_c and an additional 1,000 rating scores based on

the sets $S_c^1, S_c^2, \ldots, S_c^{1000}$. A parameter, namely, the P-value for NSCLC-related chemicals, was calculated for c by using the formula

$$\text{P} - \text{value}_c(c) = \frac{W_c}{1000} \qquad (1)$$

where W_c is the number of sets for which the rating scores were larger than $RS_c(c)$. Furthermore, we also evaluated the importance of compounds that passed the association test using the NSCLC-related gene set S_g. One thousand sets, $S_g^1, S_g^2, \ldots, S_g^{1000}$, were randomly constructed, and each of them had the same size as S_g. As above, we also calculated the rating score of compound c for each set and computed the P-value of c for NSCLC-related genes, denoted by P – value$_g(c)$, as

$$\text{P} - \text{value}_g(c) = \frac{W_g}{1000} \qquad (2)$$

where W_g represents the number of sets for which the rating scores were larger than $RS_g(c)$. Generally, we should select compounds with low P-values for NSCLC-related chemicals/genes. However, it is quite difficult to determine the thresholds of these two P-values. On the other hand, the P-values of approved NSCLC drugs were important indicators, which helped us select proper thresholds. Therefore, we computed the P-values of approved NSCLC drugs for NSCLC-related chemicals and genes. To retain as many candidate compounds as possible and avoid missing potential compounds, we selected the maximum values from the P-values of approved NSCLC drugs as the thresholds of the two P-values.

2.3.3 Screening compounds by the EM clustering algorithm. Some candidate compounds were able to pass the association test and permutation test. These compounds have many or few associations with NSCLC. A procedure was built to screen core candidate compounds that have extensive associations with approved NSCLC drugs, NSCLC-related chemicals and NSCLC-related genes from this set.

2.3.3.1 Feature extraction: As described in Section 2.2, five scores for chemical-chemical interactions and five scores for protein-chemical interactions were introduced. However, the first two procedures used only the last score. Here, all scores were used to extract useful features, which can accurately measure the associations between candidate compounds and approved NSCLC drugs, NSCLC-related chemicals or NSCLC-related genes. For each candidate compound c, fifteen features were extracted, of which five features were from the five scores for chemical-chemical interactions between c and approved NSCLC drugs, five features were from the five scores for chemical-chemical interactions between c and NSCLC-related chemicals and the last five features were from the five scores for protein-chemical interactions between c and NSCLC-related genes. Of the five features derived from the five scores of chemical-chemical interactions between c and approved NSCLC drugs, we only described how to extract a feature from the "Similarity" score; the others can be obtained in a similar way. Let $d_1, d_2, \ldots, d_l$ be approved NSCLC drugs such that $Q_S^{cc}(c, d_i) > 0$ $(i = 1, 2, \ldots, l)$. The mean value of these scores was taken as a feature. Particularly, if $l = 0$, then this feature was set to zero. For the five features derived from the five scores of chemical-chemical interactions between c and NSCLC-related chemicals, each of them can be obtained in a similar fashion as the feature mentioned above. Finally, of the five features derived from protein-chemical interactions between c and NSCLC-related genes, we only provide a description of features derived from the "Experimental" score; others can be constructed in a similar way. Let $g_1, g_2, \ldots, g_k$ be NSCLC-related genes with $Q_E^{pc}(g_i, c) > 0$ $(i = 1, 2, \ldots, k)$. Then, the mean value of these scores was counted as a feature. Additionally, it was set to zero if $k = 0$. Furthermore, each approved

NSCLC drug was also encoded by these fifteen features described above. All of the candidate compounds and approved NSCLC drugs were subsequently fed into a clustering algorithm.

2.3.3.2 EM clustering algorithm: The EM algorithm, proposed by Dempster *et al.* [42], is an iterative method to find the maximum likelihood of parameters in statistical models. The iteration procedure alternates between executing an expectation (E) step and a maximization (M) step. Its steps are listed in Table 3. If the dataset obeys a distribution that can be approximated by a mixture of Gaussian distributions, the EM algorithm can be extended to clustering. The unobserved data set Z represents which Gaussian the datum in observed data set Y comes from. By utilizing the EM algorithm, the parameters of each Gaussian can be estimated, which helps to assign each datum to a particular one.

Weka [43] is a suite of software collecting several popular state-of-the-art machine learning algorithms and data preprocessing tools. The "EM" tool implements the EM clustering algorithm described above. For convenience, it was directly employed in this study to cluster the candidate compounds and approved NSCLC drugs. The default parameters were used to execute "EM" in which the class number can be automatically determined. Based on the cluster results, candidate compounds in the same category as approved NSCLC drugs were picked, and these were called putative compounds for convenience.

3. Results and discussion

3.1 Results of the preliminary screening

Sixteen approved NSCLC drugs were used in this study. In the preliminary screening procedure, we extracted all compounds that can interact with at least one approved NSCLC drug, obtaining 3,261 possible compounds. These compounds are listed in S3 Table.

3.2 Results of the association test and permutation test

Several possible compounds were identified in the preliminary screening procedure. Clearly, not all of them have anti-NSCLC activity. In the association test, they were linked to NSCLC-related chemicals and NSCLC-related genes. Those that can interact with at least one NSCLC-related chemical and one NSCLC-related gene were kept, resulting in 1,281 compounds. In addition, we calculated the rating scores for NSCLC-related chemicals (cf. RS_c) and NSCLC-related genes (cf. RS_g) for each of the 1,281 compounds. These scores are available in S3 Table. It is necessary to note that the sixteen approved NSCLC drugs were also examined in the association test. The results show that ten of them can interact with at least one NSCLC-related chemical and one NSCLC-related gene. They are listed in Table 4. Additionally, the two rating

Table 3. The procedures of the EM algorithm.

EM algorithm
Input: an observed dataset Y, an unobserved latent data set Z, the joint distribution $P(Y,Z\|\theta)$ and the conditional distribution $P(Z\|Y,\theta)$
Output: parameter $\theta*$
(1) Set an initial parameter $\theta^{(0)}$ and $i = 0$
(2) E-step: for the current estimated value $\theta^{(i)}$ of parameter θ, calculate the objective function: $Q(\theta,\theta^{(i)}) = E_{Z\|Y,\theta^{(i)}}[\log P(Y,Z\|\theta)] = \sum_{Z}\log P(Y,Z\|\theta)P(Z\|Y,\theta^{(i)})$
(3) M-step: the new estimated value of parameter θ, say $\theta^{(i+1)}$, by maximizing $Q(\theta,\theta^{(i)})$, i.e., $\theta^{(i+1)} = \arg\max_{\theta} Q(\theta,\theta^{(i)})$
(4) Check the condition $\|\theta^{i+1}-\theta^{i}\|_2 < \varepsilon$. If the condition is not satisfied, $i = i + 1$, go to (2); else set $\theta* = \theta^{(i+1)}$ and output $\theta*$.

Table 4. The measurements of ten approved NSCLC drugs yielded by the computational method.

PubChem ID	Name	Rating score on NSCLC-related chemicals	Rating score on NSCLC-related genes	P-value on NSCLC-related chemicals	P-value on NSCLC-related genes
CID4033	Mechlorethamine	296.299	358.250	0.248	0.115
CID36314	Paclitaxel	306.485	462.889	0.346	0.221
CID38904	Carboplatin	307.897	332.267	0.196	0.238
CID57166	Porfimer Sodium	307.231	218.000	0.312	0.032
CID60843	Pemetrexed	278.798	233.750	0.337	0.124
CID123631	Gefitinib	303.047	493.275	0.289	0.292
CID126941	Methotrexate	330.201	362.250	0.469	0.150
CID176870	Erlotinib	320.090	579.929	0.273	0.175
CID10184653	Afatinib	365.038	451.091	0.128	0.063
CID11626560	Crizotinib	293.407	579.500	0.263	0.187

scores were also calculated and are listed in Table 4. These ten drugs helped us to further screen important compounds.

For the permutation test, we calculated the P-values for NSCLC-related chemicals (cf. Eq 1) and NSCLC-related genes (cf. Eq 2) for each of the 1,281 compounds that passed the association test; these are provided in S3 Table. Furthermore, these two P-values were also computed for the ten approved NSCLC drugs and are listed in Table 4. The maximum P-value of the ten approved NSCLC drugs for NSCLC-related chemicals was 0.469, and the maximum P-value of the ten approved NSCLC drugs for NSCLC-related genes was 0.292. Accordingly, 0.469 and 0.292 were set as the thresholds for the P-values for NSCLC-related chemicals and NSCLC-related genes, respectively, *i.e.*, we selected the compounds with P-values less than or equal to 0.469 for NSCLC-related chemicals and P-values less than or equal to 0.292 for NSCLC-related genes. Based on these thresholds, 1,007 compounds were retained, which are listed in S4 Table.

3.3 Results of the EM clustering algorithm

To further select core candidate compounds from the 1,007 compounds obtained after the permutation test, they were represented by fifteen features, as described in Section 2.3.3. In addition, the ten approved NSCLC drugs listed in Table 4 were also encoded in the same way. Next, the EM clustering algorithm was used to cluster these 1,017 compounds (1007 candidate compounds and ten approved NSCLC drugs). The results are provided in S5 Table. Four categories were built by the EM clustering algorithm. Notably, the ten approved NSCLC drugs were clustered in the same category (cluster3). Clearly, candidate compounds in this category are more likely to be novel drugs for NSCLC than other candidate compounds. Therefore, they were extracted, resulting in 98 candidate compounds, which are listed in S6 Table.

However, 98 candidate compounds are still too many to screen for potential drugs for NSCLC. Therefore, these compounds and the ten approved NSCLC drugs were again input into the EM clustering algorithm. The clustering results are available in S6 Table and show that five categories were used to cluster these compounds. Interestingly, the ten approved NSCLC drugs were still clustered in the same category (cluster3). Another six candidate compounds were also in this category and are listed in Table 5. These six putative compounds were deemed to be significant for further investigation.

3.4 Analysis of significant candidate drugs

In this study, six putative compounds for NSCLC were identified by our method, which are listed in Table 5. To give their associations with approved drugs and NSCLC-related genes, a

Table 5. Detailed information of six putative compounds.

PubChem ID	Name	Rating score on NSCLC-related chemicals	P-value on NSCLC-related chemicals	Rating score on NSCLC-related genes	P-value on NSCLC-related genes	NSCLC-related Genes
CID4971	Protoporphyrin IX	347.889	0.326	437.000	0.115	*ABCB1, ABCB10, CASP8, CAT, SOD2, TP53*
CID11103	Hematoporphyrin	309.920	0.225	224.333	0.029	*MAP2K1, MAPK1, MAPK3*
CID156413	Canertinib	331.293	0.155	781.000	0.054	*AKT1, AREG, AREGB, EGF, EGFR, ERBB2, MAPK10, MMP9, VEGFA*
CID208908	Lapatinib	307.412	0.303	504.750	0.119	*ABCB1, ABCB11, AKT1, AREG, AREGB, BIRC5, CCND1, CLPTM1L, EGFR, ERBB2, FOXM1, HRAS, KRAS, MAPK1, MAPK3, MET, NRAS, PIK3CA, STAT3, TGFA, TP53, VEGFA*
CID6445562	Pelitinib	322.081	0.153	605.833	0.112	*AKT1, BIRC5, EGF, EGFR, ERBB2, STK4*
CID11511120	Dacomitinib	302.176	0.131	576.000	0.025	*EGFR, EML4, ERBB2, ERCC1, NRAS*

https://doi.org/10.1371/journal.pone.0183411.t005

network consisting of the interactions among putative compounds, approved drugs and NSCLC-related genes was plotted in Fig 2. It can be observed that each putative compound is closely related to at least one approved drugs (see Fig 2(D)) and one NSCLC-related genes (see Fig 2(C)), suggesting that these putative compounds can be novel candidate drugs for NSCLC. In addition, the interactions between four putative compounds: Pelitinib, Dacomitinib, Canertinib and Lapatinib comprise a clique (see Fig 2(B)), a graph such that each pair of nodes is connected by an edge, implying they are highly related with each other. If one can be validated to be a novel drug for NSCLC, the rest putative compounds can be novel drugs with high probabilities. For other two putative compounds: Hematoporphyrin and Protoporphyrin IX, they can interact with each other, inducing the same results mentioned above. To give a more convincing explanation, a summary of the extensive data in the literature that support the anti-NSCLC activity of these compounds is presented below.

3.4.1 Protoporphyrin IX (CID4971). Photodynamic therapy (PDT) can be used for the treatment of different tumors [44]. 5-Aminolevulinic acid (ALA) is a pro-drug of the photosensitizer protoporphyrin IX (PPIX). ALA-mediated PDT showed photo-cytotoxicities towards H460 cell lines by activating p38 MAPK and JNK signal pathways [45]. Postiglione *et al.* reported that gefitinib combined with 5-ALA/PDT improved the response of NSCLC cell lines H1299 (p53-/-) and A549 (p53+/+) without EGFR mutations [46]. Zn PPIX by inhibiting heme oxygenase 1 reduced tumor growth of LL/2 mouse lung cancer cells [47]. Moreover, Zn PPIX increased the radiosensitivity of human NSCLC A549 cells and the cell apoptotic index when combined with irradiation [48].

3.4.2 Hematoporphyrin (CID11103). Hematoporphyrin and its derivatives can lead to induct DNA damage [49, 50]. Hematoporphyrin derivatives (HPD) are used for photodynamic therapy by selectively destroying malignant tumors, such as cancers of lung, digestive tract, and genitourinary tract [51]. LoCicero *et al.* reported that HPD decreased some symptoms of NSCLC patients, especially coughing [52]. Moreover, Edell *et al.* reported that 93% patients with early superficial squamous cell carcinoma achieved a complete response to HPD phototherapy and indicated that it may be an efficient alternative to surgical resection [53].

3.4.3 Canertinib (CID156413). Canertinib (Cl-1033) is a selective tyrosine kinase inhibitor (TKI) that blocks signal transduction through EGFRs [54]. Slichenmyer *et al.* reported that canertinib significantly suppressed the tumor growth of H125 NSCLC carcinoma [55]. Jänne *et al.* reported that canertinib had modest activity in advanced-stage NSCLC patients [56]. Moreover, canertinib was confirmed to be more effective than erlotinib and gefitinib against
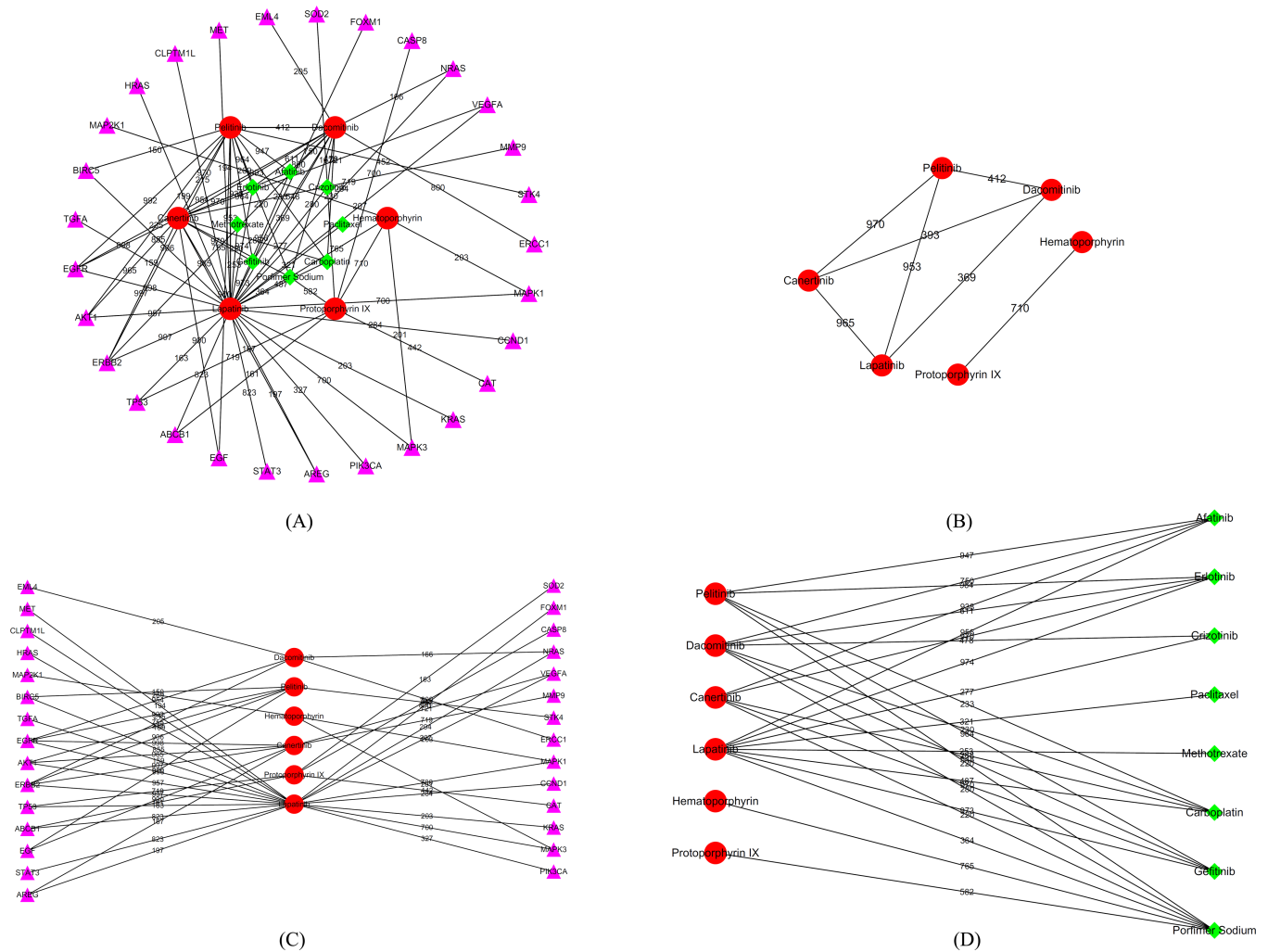
Fig 2. The interaction sub-network of putative compounds, NSCLC-related genes and approved drugs. Red circles represent putative compounds, pink triangles represent NSCLC-related genes, green diamonds represent approved drugs. Weights on edges are "Combined_score" of corresponding chemical-chemical interactions or protein-chemical interactions. (A) the whole sub-network; (B) the sub-network of putative compounds; (C) the sub-network of putative compounds and NSCLC-related genes; (D) the sub-network of putative compounds and approved drugs.

NSCLC cell lines with the EGFR L858R mutation and the EGFR L858R/T790M double mutations [57].

3.4.4 Lapatinib (CID208908). Lapatinib is a dual TKI of EGFR and the human epidermal receptor type 2 (ErbB2) receptor for treating advanced or metastatic breast cancer with the overexpression of ErbB2 protein [58]. Diaz *et al.* reported that lapatinib significantly reduced cell proliferation, DNA synthesis and colony formation in NSCLC A549 cells and inhibited the angiogenesis of tumors in mice [59]. Moreover, Kim *et al.* reported that the combination of lapatinib and cetuximab had enhanced cytotoxicity against gefitinib-resistant NSCLC cells [60]. Lapatinib also showed an inhibitory effect against the NSCLC cell line H3255 with the EGFR L858R mutation [57].

3.4.5 Pelitinib (CID6445562). Pelitinib (EKB-569) is a selective and irreversible inhibitor of EGFR. It showed clinical activity in two NSCLC patients with EGFR mutations and gefitinib resistance [61] and stabilized the disease in another NSCLC patient for 33 weeks [62].

Yoshimura *et al.* reported that pelitinib decreased multiple pulmonary metastases in two advanced NSCLC patients with EGFR mutations [61].

3.4.6 Dacomitinib (CID11511120). Dacomitinib (PF-00299804) is an irreversible pan-HER TKI that targets EGFRs. Ramalingam *et al.* reported that dacomitinib significantly improved progression-free survival compared with erlotinib in some clinical and molecular subsets, such as *KRAS* wild-type/*EGFR* wild-type and *EGFR* mutants [63]. However, the side effects of dacomitinib occurred more frequently and with greater intensity compared with erlotinib or gefitinib [64].

Of the above six putative compounds, canertinib, lapatinib, pelitinib, and dacomitinib may be promising for the treatment of NSCLC with *EGFR* mutations. Notably, only canertinib was identified in a previous study [7]. Therefore, the other newly identified compounds could be useful in future studies. Additionally, all of the compounds identified by the proposed method have been shown to possess anti-NSCLC activity. In a previous study, only 31.58% (6/19) of the identified compounds had anti-NSCLC activity. Therefore, it can be concluded that our method is effective at identifying candidate drugs for NSCLC.

4. Conclusions

This study used a computational method for identifying novel putative compounds of NSCLC, which were deemed to have anti-NSCLC activity. Several related materials, including chemical-chemical interactions, protein-chemical interactions, and the EM clustering algorithm were used for its implementation. Six compounds were identified, and further the analysis of the results indicated that all of them have anti-NSCLC activity. We hope that these newly identified compounds will be further validated by experimental data, which could lead to new therapies for treating NSCLC.

Supporting information

S1 Table. 3085 chemicals related to NSCLC.
(DOCX)

S2 Table. 148 genes related to NSCLC.
(DOCX)

S3 Table. 3261 possible compounds after preliminary screening.
(PDF)

S4 Table. 1007 candidate compounds filtered by the association test and permutation test.
(PDF)

S5 Table. Clustering results by the EM algorithm on 1007 candidate compounds and ten approved NSCLC drugs.
(DOCX)

S6 Table. Clustering results by the EM algorithm on 98 candidate compounds and ten approved NSCLC drugs.
(DOCX)

Author Contributions

Conceptualization: Lei Chen, Yu-Dong Cai.

Data curation: Lei Chen.

Formal analysis: Jing Lu, Tao Huang.

Funding acquisition: Lei Chen, Yu-Dong Cai.

Methodology: Lei Chen, Jing Lu.

Supervision: Yu-Dong Cai.

Validation: Jing Lu, Tao Huang, Yu-Dong Cai.

Writing – original draft: Lei Chen, Jing Lu.

Writing – review & editing: Tao Huang, Yu-Dong Cai.

References

1. A guide for journalists on Non-Small Cell Lung Cancer (NSCLC) and its treatment. http://www.roche.com/med-lung-cancer.pdf (accessed 20 April 2016).

2. The top 10 causes of death. http://www.who.int/mediacentre/factsheets/fs310/en/ (accessed 21 April 2016).

3. Non-Small Cell Lung Cancer Version I, 2015. http://www.nccn.org/patients/ (accessed 25 April 2016).

4. Lang KL, Silva IT, Machado VR, Zimmermann LA, Caro MS, Simoes CM, et al. Multivariate SAR and QSAR of cucurbitacin derivatives as cytotoxic compounds in a human lung adenocarcinoma cell line. J Mol Graph Model. 2014; 48:70–9. Epub 2014/01/01. https://doi.org/10.1016/j.jmgm.2013.12.004 PMID: 24378396.

5. Goyal S, Jamal S, Shanker A, Grover A. Structural investigations of T854A mutation in EGFR and identification of novel inhibitors using structure activity relationships. BMC Genomics. 2015; 16 Suppl 5:S8. Epub 2015/06/05. https://doi.org/10.1186/1471-2164-16-s5-s8 PMID: 26041145; PubMed Central PMCID: PMC4460657.

6. Xiang M, Lei K, Fan W, Lin Y, He G, Yang M, et al. In silico identification of EGFR-T790M inhibitors with novel scaffolds: start with extraction of common features. Drug Des Devel Ther. 2013; 7:789–839. Epub 2013/08/31. https://doi.org/10.2147/DDDT.S41305 PMID: 23990708; PubMed Central PMCID: PMC3748928.

7. Lu J, Chen L, Yin J, Huang T, Bi Y, Kong X, et al. Identification of new candidate drugs for lung cancer using chemical-chemical interactions, chemical-protein interactions and a K-means clustering algorithm. J Biomol Struct Dyn. 2016; 34(4):906–17. Epub 2016/02/06. https://doi.org/10.1080/07391102.2015.1060161 PMID: 26849843.

8. Mechlorethamine website. https://www.drugbank.ca/drugs/DB00888 (accessed in July 15, 2017).

9. Paclitaxel website. https://www.drugbank.ca/drugs/DB01229 (accessed in July 15, 2017).

10. Carboplatin website. https://www.drugbank.ca/drugs/DB00958 (accessed in July 15, 2017).

11. Porfimer Sodium website. https://www.drugbank.ca/drugs/DB00707 (accessed in July 15, 2017).

12. Gemcitabine website. https://www.drugbank.ca/drugs/DB00441 (accessed in July 15, 2017).

13. Vinorelbine website. https://www.drugbank.ca/drugs/DB00361 (accessed in July 15, 2017).

14. Pemetrexed website. https://www.drugbank.ca/drugs/DB00642 (accessed in July 15, 2017).

15. Gefitinib website. https://www.drugbank.ca/drugs/DB00317 (accessed in July 15, 2017).

16. Methotrexate website. https://www.drugbank.ca/drugs/DB00563 (accessed in July 15, 2017).

17. Docetaxel website. https://www.drugbank.ca/drugs/DB01248 (accessed in July 15, 2017).

18. Erlotinib website. https://www.drugbank.ca/drugs/DB00530 (accessed in July 15, 2017).

19. Cisplatin website. https://www.drugbank.ca/drugs/DB00515 (accessed in July 15, 2017).

20. Bleomycin website. https://www.drugbank.ca/drugs/DB00290 (accessed in July 15, 2017).

21. Afatinib website. https://www.drugbank.ca/drugs/DB08916 (accessed in July 15, 2017).

22. Crizotinib website. https://www.drugbank.ca/drugs/DB08865 (accessed in July 15, 2017).

23. Ceritinib website. https://www.drugbank.ca/drugs/DB09063 (accessed in July 15, 2017).

24. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, et al. The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Res. 2013; 41(Database issue): D1104–14. Epub 2012/10/25. https://doi.org/10.1093/nar/gks994 PMID: 23093600; PubMed Central PMCID: PMC3531134.

25. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 2000; 28(1):27–30. PMID: 10592173

26. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 1999; 27(1):29–34. PMID: 9847135

27. Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, Jensen LJ, et al. STITCH 4: integration of protein–chemical interactions with user data. Nucleic acids research. 2014; 42(D1):D401–D7.

28. Chen L, Zeng W-M, Cai Y-D, Feng K-Y, Chou K-C. Predicting Anatomical Therapeutic Chemical (ATC) Classification of Drugs by Integrating Chemical-Chemical Interactions and Similarities. PLoS ONE. 2012; 7(4):e35254. https://doi.org/10.1371/journal.pone.0035254 PMID: 22514724

29. Lu J, Huang G, Li HP, Feng KY, Chen L, Zheng MY, et al. Prediction of cancer drugs by chemical-chemical interactions. PLoS One. 2014; 9(2):e87791. Epub 2014/02/06. https://doi.org/10.1371/journal.pone.0087791 PMID: 24498372.

30. Chen L, Lu J, Zhang N, Huang T, Cai Y-D. A hybrid method for prediction and repositioning of drug Anatomical Therapeutic Chemical classes. Molecular BioSystems. 2014; 10(4):868–77. https://doi.org/10.1039/c3mb70490d PMID: 24492783

31. Xu R, Wang Q. Comparing a knowledge-driven approach to a supervised machine learning approach in large-scale extraction of drug-side effect relationships from free-text biomedical literature. BMC Bioinformatics. 2015; 16(5):S6. https://doi.org/10.1186/1471-2105-16-s5-s6 PMID: 25860223

32. Chen L, Chu C, Zhang Y-H, Zheng M-Y, Zhu L, Kong X, et al. Identification of Drug-Drug Interactions Using Chemical Interactions. Current Bioinformatics. 2017.

33. Re M, Valentini G. Network-Based Drug Ranking and Repositioning with Respect to DrugBank Therapeutic Categories. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2013; 10 (6):1359–71. https://doi.org/10.1109/TCBB.2013.62 PMID: 24407295

34. Re M, Valentini G. Large Scale Ranking and Repositioning of Drugs with Respect to DrugBank Therapeutic Categories. In: Bleris L, Măndoiu I, Schwartz R, Wang J, editors. Bioinformatics Research and Applications: 8th International Symposium, ISBRA 2012, Dallas, TX, USA, May 21–23, 2012 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 225–36.

35. Cheng X, Zhao S-G, Xiao X, Chou K-C. iATC-mISF: a multi-label classifier for predicting the classes of anatomical therapeutic chemicals. Bioinformatics. 2016. https://doi.org/10.1093/bioinformatics/btw644 PMID: 28172617

36. Chen L, Huang T, Zhang J, Zheng M-Y, Feng K-Y, Cai Y-D, et al. Predicting Drugs Side Effects Based on Chemical-Chemical Interactions and Protein-Chemical Interactions. BioMed Research International. 2013; 2013:485034. https://doi.org/10.1155/2013/485034 PMID: 24078917

37. Chen L, Lu J, Luo X, Feng K-Y. Prediction of drug target groups based on chemical-chemical similarities and chemical-chemical/protein connections. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2014; 1844(1):207–13.

38. Chen L, Lu J, Huang T, Yin J, Wei L, Cai Y-D. Finding Candidate Drugs for Hepatitis C Based on Chemical-Chemical and Chemical-Protein Interactions. PLoS ONE. 2014; 9(9):e107767. https://doi.org/10.1371/journal.pone.0107767 PMID: 25225900

39. Kuhn M, Al Banchaabouchi M, Campillos M, Jensen LJ, Gross C, Gavin AC, et al. Systematic identification of proteins that elicit drug side effects. Mol Syst Biol. 2013; 9:663. https://doi.org/10.1038/msb.2013.10 PMID: 23632385; PubMed Central PMCID: PMC3693830.

40. Duran-Frigola M, Aloy P. Analysis of Chemical and Biological Features Yields Mechanistic Insights into Drug Side Effects. Chemistry & Biology. 2013; 20(4):594–603. https://doi.org/https://doi.org/10.1016/j.chembiol.2013.03.017

41. Huang Y-F, Yeh H-Y, Soo V-W. Inferring drug-disease associations from integration of chemical, genomic and phenotype data using network propagation. BMC Medical Genomics. 2013; 6(3):S4. https://doi.org/10.1186/1755-8794-6-s3-s4 PMID: 24565337

42. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B (Methodological). 1977; 39(1):1–38.

43. Witten IH, Frank E, editors. Data Mining:Practical Machine Learning Tools and Techniques. San Francisco, Morgan, Kaufmann 2005.

44. Mitton D, Ackroyd R. History of photodynamic therapy in Great Britain. Photodiagnosis Photodyn Ther. 2005; 2(4):239–46. Epub 2005/12/01. https://doi.org/10.1016/S1572-1000(05)00111-0 PMID: 25048866.

45. Wu RW, Yow CM, Wong CK, Lam YH. Photodynamic therapy (PDT)—Initiation of apoptosis via activation of stress-activated p38 MAPK and JNK signal pathway in H460 cell lines. Photodiagnosis Photodyn Ther. 2011; 8(3):254–63. Epub 2011/08/26. https://doi.org/10.1016/j.pdpdt.2010.12.002 PMID: 21864799.

46. Postiglione I, Chiaviello A, Aloj SM, Palumbo G. 5-aminolaevulinic acid/photo-dynamic therapy and gefitinib in non-small cell lung cancer cell lines: a potential strategy to improve gefitinib therapeutic efficacy. Cell Prolif. 2013; 46(4):382–95. Epub 2013/07/23. https://doi.org/10.1111/cpr.12040 PMID: 23869760.

47. Hirai K, Sasahira T, Ohmori H, Fujii K, Kuniyasu H. Inhibition of heme oxygenase-1 by zinc protoporphyrin IX reduces tumor growth of LL/2 lung cancer in C57BL mice. International journal of cancer Journal international du cancer. 2007; 120(3):500–5. Epub 2006/10/27. https://doi.org/10.1002/ijc.22287 PMID: 17066448.

48. Zhang W, Qiao T, Zha L. Inhibition of heme oxygenase-1 enhances the radiosensitivity in human non-small cell lung cancer a549 cells. Cancer Biother Radiopharm. 2011; 26(5):639–45. Epub 2011/09/29. https://doi.org/10.1089/cbr.2010.0939 PMID: 21950555.

49. Moan J, Waksvik H, Christensen T. DNA single-strand breaks and sister chromatid exchanges induced by treatment with hematoporphyrin and light or by x-rays in human NHIK 3025 cells. Cancer Res. 1980; 40(8 Pt 1):2915–8. Epub 1980/08/01. PMID: 7388841.

50. Fiel RJ, Datta-Gupta N, Mark EH, Howard JC. Induction of DNA damage by porphyrin photosensitizers. Cancer Res. 1981; 41(9 Pt 1):3543–5. Epub 1981/09/01. PMID: 7020932.

51. Dougherty TJ, Gomer CJ, Henderson BW, Jori G, Kessel D, Korbelik M, et al. Photodynamic therapy. Journal of the National Cancer Institute. 1998; 90(12):889–905. Epub 1998/06/24. PMID: 9637138; PubMed Central PMCID: PMC4592754.

52. LoCicero J 3rd, Metzdorff M, Almgren C. Photodynamic therapy in the palliation of late stage obstructing non-small cell lung cancer. Chest. 1990; 98(1):97–100. Epub 1990/07/01. PMID: 1694475.

53. Edell ES, Cortese DA. Photodynamic therapy in the management of early superficial squamous cell carcinoma as an alternative to surgical resection. Chest. 1992; 102(5):1319–22. Epub 1992/11/01. PMID: 1424843.

54. Pipeline Report—HER2 inhibitors. http://www.bizcharts.com/product/refrows/RefRowV1_HER-2_HTML.htm. (accessed in September 12, 2015).

55. Slichenmyer WJ, Elliott WL, Fry DW. CI-1033, a pan-erbB tyrosine kinase inhibitor. Seminars in oncology. 2001; 28(5 Suppl 16):80–5. Epub 2001/11/14. doi: asonc02805n0080 [pii]. PMID: 11706399.

56. Janne PA, von Pawel J, Cohen RB, Crino L, Butts CA, Olson SS, et al. Multicenter, randomized, phase II trial of CI-1033, an irreversible pan-ERBB inhibitor, for previously treated advanced non small-cell lung cancer. J Clin Oncol. 2007; 25(25):3936–44. Epub 2007/09/01. https://doi.org/10.1200/JCO.2007.11.1336 PMID: 17761977.

57. Li D, Ambrogio L, Shimamura T, Kubo S, Takahashi M, Chirieac LR, et al. BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. Oncogene. 2008; 27(34):4702–11. https://doi.org/10.1038/onc.2008.109 PMID: 18408761.

58. Lapatnib website. http://www.drugbank.ca/drugs/DB01259 (accessed September 20, 2015).

59. Diaz R, Nguewa PA, Parrondo R, Perez-Stable C, Manrique I, Redrado M, et al. Antitumor and antiangiogenic effect of the dual EGFR and HER-2 tyrosine kinase inhibitor lapatinib in a lung cancer model. BMC Cancer. 2010; 10:188. Epub 2010/05/13. https://doi.org/10.1186/1471-2407-10-188 PMID: 20459769; PubMed Central PMCID: PMC2883966.

60. Kim HP, Han SW, Kim SH, Im SA, Oh DY, Bang YJ, et al. Combined lapatinib and cetuximab enhance cytotoxicity against gefitinib-resistant lung cancer cells. Mol Cancer Ther. 2008; 7(3):607–15. Epub 2008/03/19. https://doi.org/10.1158/1535-7163.MCT-07-2068 PMID: 18347147.

61. Yoshimura N, Kudoh S, Kimura T, Mitsuoka S, Matsuura K, Hirata K, et al. EKB-569, a new irreversible epidermal growth factor receptor tyrosine kinase inhibitor, with clinical activity in patients with non-small cell lung cancer with acquired resistance to gefitinib. Lung Cancer. 2006; 51(3):363–8. Epub 2005/12/21. https://doi.org/10.1016/j.lungcan.2005.10.006 PMID: 16364494.

62. Erlichman C, Hidalgo M, Boni JP, Martins P, Quinn SE, Zacharchuk C, et al. Phase I study of EKB-569, an irreversible inhibitor of the epidermal growth factor receptor, in patients with advanced solid tumors. J Clin Oncol. 2006; 24(15):2252–60. Epub 2006/05/20. https://doi.org/10.1200/JCO.2005.01.8960 PMID: 16710023.

63. Ramalingam SS, Blackhall F, Krzakowski M, Barrios CH, Park K, Bover I, et al. Randomized phase II study of dacomitinib (PF-00299804), an irreversible pan-human epidermal growth factor receptor inhibitor, versus erlotinib in patients with advanced non-small-cell lung cancer. J Clin Oncol. 2012; 30 (27):3337–44. Epub 2012/07/04. https://doi.org/10.1200/JCO.2011.40.9433 PMID: 22753918.

64. Brzezniak C, Carter CA, Giaccone G. Dacomitinib, a new therapy for the treatment of non-small cell lung cancer. Expert Opin Pharmacother. 2013; 14(2):247–53. Epub 2013/01/09. https://doi.org/10.1517/14656566.2013.758714 PMID: 23294134.