

MANUAL

Drug Repurposing Explorer

Abhik Seal

Introduction

In this project we used the side effect data from sider which contains around 996 drugs along with 4500 side effects information. For the dataset preparation we used SIDER database containing 996 drugs and mapped it to the Drugbank ID (DBID) using pubchem identifier exchange service. We found 545 compounds correctly mapped to the Drugbank ID using pubchem CID. We then mapped the compounds by their generic name and visually inspected the drugs and formed a dataset of 746 compounds. From 746 drugs, a filter was applied in which drugs whose molecular weight was more than 1000 Da were removed along with some metals like lithium and Magnesium. The final dataset contained 727 drugs.

We collected the side effect information of 727 drugs from SIDER. The target information, ATC Codes from the Drugbank. We also manually added some of the drugs ATC codes from KEGG database. The disease information of 727 drugs was obtained from Comparative Toxicogenomics Database (CTD) [8]. CTD is a curated information resource for chemical- gene, chemical-diseases, gene-diseases, chemical –ontology interaction which is curated by hand by the CTD bio-curators. We extracted the Chemical to disease association between the compounds which are related to the therapeutic area only which will help to identify possible diseases to treat the drug and also if new diseases could be identified.

A user searching for a drug in Drug Repurposing explorer should be aware of the drug and also for what purpose it is used.

Terms in the website:

1) ECFP6- These are the ECFP6 fingerprints calculated using Accelrys pipeline pilot academic version. This type of fingerprints is used in the tool because it gave a good significant correlation value with the Proteins target similarity. This indicates that if a search is made with ECFP6 and proteins will return some top similar compounds. Tanimoto similarity is calculated for the compounds using ECFP6

2) Pubchem – It is the 881 pubchem fingerprints calculated using rcdk. Tanimoto similarity is calculated for the compounds using pubchem fingerprints.

3) Proteins – It is the compounds which shares similar proteins. The compound target information is obtained from the Drugbank data and converted to a matrix. A fast dice coefficient calculation algorithm is applied to calculate the similarity given in equation 1. It is being found that dice and tanimoto the results are proportional.

$$s(x, y) = \frac{2 \times |\{p_{x \rightarrow y} : p_{x \rightarrow y} \in P\}|}{|\{p_{x \rightarrow x} : p_{x \rightarrow x} \in P\}| + |\{p_{y \rightarrow y} : p_{y \rightarrow y} \in P\}|} \quad eq\ 1$$

4) **ROCS** – comparison of compounds based on 3D shape and color with a maximum of 400 conformers for each molecule was done to generate a 727 x 727 matrix.

5) **SIDE EFFECT**- The compound similarity via side effect was similar to proteins similarity and same algorithm was applied given in equation 1.727 drugs are mapped and a 727 x 4500 matrix was generated in order to calculate the similarity.

Using of Drug Repurposing Explorer:

a) Search via any one Option

1) A User can search via normally by side effect to list similar compound via side effect. A search cannot be made with other compounds if a drug is not in the database. The search Box automatically lists the drug names when a user tries to type a drug name. Diagram 1 given below helps to understand how to search. Other than that if a user wants to search the disease related to the search for the top similar compounds he/she can select the disease check box and it will return the drug name and disease associated with it.

Search for Drugs:

at

- atazanavir
- atenolol
- atomoxetine
- atorvastatin
- atovaquone
- atracurium
- atropine

☒ SIDE EFFECT : ☐ ECFP6 : ☐ Pubchem :

☐ Proteins: ☐ ROCS(3D Shape and Color):

☐ Disease

SEARCH NOW!

Diagram showing Side effect similarity search

Search for Drugs:

atorvastatin

☒ SIDE EFFECT : ☐ ECFP6 : ☐ Pubchem :

☐ Proteins: ☐ ROCS(3D Shape and Color):

☒ Disease

SEARCH NOW!

Diagram showing side effect and Disease search.

Diagram1

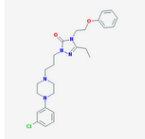
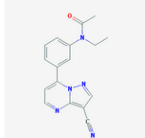
Results

For searching by one option for example only side effects the results page returns a table for the top similar drugs by that option. Diagram 2 shows the results page by selecting only option for example side effect and selecting Disease checkbox. The results page shows the top search similar drugs by side effect and it similarity calculated using equation 1, ATC Code and the wiki links of the drug .The disease table results with the top similar drug along with its disease information

Note: It might be possible that drugs which are orderly ranked by any one option, in the disease table that drug may not available. It is because the current version has 680 drugs mapped with the disease information. Later version will have more drugs and Disease information.

Wikipedia Link

Similarity of atorvastatin by sideEffect

RANK	DRUG NAME	CHEMICAL STRUCTURE	SIMILARITY	ATC CODE
1	nefazodone		0.579055441	N06AX06
2	zaleplon		0.546906188	N05CF03

Similarity of atorvastatin by sideEffect and its associated diseases

DRUG NAME	SIMILARITY	DISEASE NAME	PUBMED ID
nefazodone	0.579055441	Akathisia, Drug-Induced	11488371
nefazodone	0.579055441	Anxiety Disorders	8666573
nefazodone	0.579055441	Basal Ganglia Diseases	12766931
nefazodone	0.579055441	Cocaine-Related Disorders	15730347
nefazodone	0.579055441	Depressive Disorder	11235924;11488371;16410371;18083778;7855217;8626361;8666573;8986316
nefazodone	0.579055441	Depressive Disorder, Major	9184611
zaleplon	0.546906188	Seizures	11060714;8905326
zaleplon	0.546906188	Sleep Initiation and Maintenance Disorders	10682226;11060714;17557470
acamprosate	0.546666667	Alcoholic Intoxication	15608650
acamprosate	0.546666667	Alcoholism	11772933;15608650;19523047;20585342

Diagram 2

b) Search via any two or more Options

Apart from selecting one option a user can select two options for searching for example ECFP6 or Pubchem with Side Effect. Following with it there are dropdown boxes given, where a user can do a weighted combo search. This type of search is very useful because in drug discovery and repurposing area varied structures can have similar side effect and a protein target can also have varied drug compounds. Along with it clicking the disease checkbox it will return the disease information to.

For searching the database a query should like the following:

a) Search compounds which have less structural similarity but high side effect similarity (application of weights example (**side effect= 0.9 ECFP6 = 0.1**))

Or

b) Search compounds which have high protein sharing similarity and less 3D shape similarity (ROCS)

Or

c) Search compounds which have equal side effect and structural similarity weights (0.5, 0.5) or (0.3, 0.3). you can give any value searching with equal weights.

Case study 1:

Searching for Similar group of compounds.

In this case study I want to have the entire statins top ranked using a combo search option. The statins fluvastatin, rosuvastatin, cervastatin, pravastatin, lovastatin, simvastatin. It is not possible for a single search option to get all the statins ranked in order because they have varied structures so as share different proteins and some of them have different side effects. But when a combo search is made with ECFP fingerprints, Proteins and Side effect with weights **0.2, 0.3, 0.5** respectively then we can rank the statins in top search. This type of combo search is very powerful in drug discovery research which could be useful for drug repurposing.

The major criteria for combo search are the weights, a user should be careful in selecting the weight and performing the search. Table 1 gives the comparisons of different search. The number in the brackets represent the rank.

ECFP	Proteins	SideEffect	ROCS	ECFP + Proteins + SideEffect (0.2,0.3,0.5)
fluvastatin	cerivastatin	nefazodone	ciclesonide	Cerivastatin
cerivastatin	fluvastatin	zaleplon	fluvastatin	Fluvastatin
rosuvastatin	pravastatin	acamprosate	cerivastatin	Rosuvastatin
flurbiprofen	rosuvastatin	Rosuvastatin(82)	Pravastatin(8)	pravastatin
pravastatin	Sitagliptin	cerivastatin(114)	Rosuvastatin(22)	sitagliptin
Lovastatin(365)	Simvastatin (11)	Fluvastatin(226)	Lovastatin(66)	Simvastatin(7)
Simvastatin(499)	Lovastatin(12)	Lovastatin(301)	Simvastatin(86)	Lovastatin(8)

Note: If any user using this tool found something interesting please mail your results to abseal@indiana.edu or djwild@indiana.edu. It will help to make a good manual for the user while searching the database.

If you are using Drug repurposing explorer for your research and you have good results don't forget to cite the website <http://cheminfov.informatics.indiana.edu/chem2disease/>. The article is in process.

A comment link is made for your valuable comments.