

Review Article

Open Acc<u>ess</u>

Drug Repositioning Network System Using the Power of Network Analysis and Machine Learning to Predict new Indications for the Approved Drugs "Drug Repositioning and Rate the Level of Drug Similarity

Sherief El Rweney*

Royal Holloway, University of London, United Kingdom

Abstract

Statement of the Problem: Drug discovery is a lengthy process, taking on average 12 years for the drugs to reach the market --but as Sir James Black OM once said "the best way to discover a new drug is to start with the old one". As result, this will drive to Drug repositioning concept. Drug Repurposing and repositioning is Finding a new clinical use for an approved drug. There are many factors that can be used to predict new target disease i.e., protein-protein interaction, chemical structure, gene expression and functional genomics, Phenotype and side effect, genetic variation and Machine learning. Protein-protein interaction PPI is Physical contacts with molecular docking between proteins that occur in a cell or in a living organism in vivo. There is Two Alternative Approaches PPI "Binary: yeast two hybrid (Y2H) and co-complex: (TAP-MS)". Drug Repositioning System, is a system built based on proteinprotein Binary interaction to predict new targets for the approved drugs. The system curate the data sets for human PPI, Drugs and diseases from well-known online sources (PPI from HRPD, drugs from DrugBank, Diseases from DisGeNET), Drug Repositioning System relates the 3 data sets based on genes name. Drug Repositioning Network System consisting of two interfaces: backend system where the curated data sets stored based on rational database and using Big Data tools, and frontend web interface where the end users can use many search engines to search inside the system for diseases, genes and drugs to predict and find new targets for the approved drugs based on protein interactions, from the web interface the user can make analysis based on his search result and build network between the genes, diseases and drugs and generate statistics to be able to answer his question. There are many Questions that can be answered by Drug Repositioning System and generate statistics: for example the main question is can we find new indications for existing approved drugs. Drug similarity: from the Drug Repositioning System we able to measure the percentage of drugs similarity between any pair genes interaction based on the number of shared drugs between them to rate the level of drug repositioning strength and then use the ROC analysis.

Keywords: Drug repositioning; Protein-protein interaction; Biological networks

Introduction

Many definition approaches for drug repositioning

a. Drug Repurposing or repositioning is finding a new clinical use for an approved drug. From the perspective of the repositioning drug, we are going to use the drugs that already have been approved which are the first step in the drugs discovery.

b. Drug repositioning is concept of new therapeutic indications that it is known by many names (drug repurposing, re-profiling, therapeutic switching and drug re-tasking. Drugs can be approved and used in a clinical setting by daily biases, or not succeed in clinical trials [1]. But in this Research we will just use the approved drugs.

c. Drug repositioning is the application of known drugs and compounds to treat new indications (i.e., new diseases) or by other meaning the Aim of a repositioning is to build a link between a drug and a disease [1].

Drug discovery

a. Drug discovery is a lengthy process, taking on average 12 years for the drugs to reach the market –but as Sir James Black OM once said "the best way to discover a new drug is to start with the old one [2].

b. EMBL-EBI defined the process of searching inside its databases during the phases of the drug design [3].

Basic concept

In the following Figure 1 based on the interaction between Protein A

and Protein B we can make network path from Protein A to Disease B and Drug B and network path from Protein B to Disease A and Drug A.



*Corresponding authors: Sherief El Rweney, Royal Holloway, University of London, United Kingdom, Tel: +447717317626; E-mail: shriefelrweney22@hotmail.com

Received December 04, 2017; Accepted January 23, 2018; Published January 30, 2018

Citation: Rweney SE (2018) Drug Repositioning Network System Using the Power of Network Analysis and Machine Learning to Predict new Indications for the Approved Drugs "Drug Repositioning and Rate the Level of Drug Similarity. J Proteomics Bioinform 11: 026-044. doi: 10.4172/jpb.1000463

Copyright: © 2018 Rweney SE. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Big concept

Using the power of network analysis and machine learning to predict new indications for the approved drugs (Table 1).

1. Drug Repositioning

2. Protein-Protein Interactions PPI

3. Network analysis

4. Biological Networks

Output example from our experiment

In the left side of the (Figure 2), it's clear that there are two groups of drugs target numerous diseases related to HDAC6 gene, and also on the right side we will find one group of drugs targets two groups of disease that are related to TUBB gene while there link between TUBB and HDAC6 indicates the interaction between them. As result we can make drug repositioning between the two genes.

Success story of drug repositioning, Thalidomide

Sildenafil (Viagra) and Thalidomide are instance for many drugs have been successfully repositioned [4]. A significant advantage of drug repositioning over traditional drug development is that since the repositioned drug has already passed a significant number of toxicity and other tests, its safety is known and the risk of failure for reasons of adverse toxicology are reduced. The significant cause of the high costs of pharmaceutical R&D is that 90% of drugs and more fail during development [4]. The importance of Drug repositioning increased in the last decade for many reasons, for example: Pharmaceutical companies know that many previously technologies have failed to deliver 'as advertised. Many factors used to increase the efficiency and success rates of drug repositioning: Computational approaches and other human use compounds against large numbers of protein targets [4].

Thalidomide: Was marketed to be used by pregnant women to handle morning sickness. The drug was supposed to be safe, based on an in vivo. For the pregnant women taking Thalidomide, severe skeletal birth defects have been appeared on their children born. Over 15,000 new-borns were affected, suffering from anatomical malformations. This unfortunate side-effect caused the molecule to be withdrawn as soon as possible and pushed important changes in the drug regulatory system. The story seems to be ended here, but by incidental discovery done by Jacob Sheskin who was trying to treat patients affected by erythema nodosum leprosum, a particularly painful inflammatory condition characterised by red nodules under the skin. On 1964 evening, she skin used some thalidomide, as the compound was known for its potent sleep-inductive properties to treat an affected patient could not sleep from the pain that was so strong. In the morning, the patient was very well. And it was surprise because all pain and soreness gone overnight too. After Sheskin also studied the effects of thalidomide in clinical trials he indicated that the thalidomide can treat erythema nodosum leprosum by two weeks'. Thalidomide here has a new life and also, become the first and only drug approved for this indication [5].

Basic development points included on the research

This research will go through basic points to achieve the aims from this study:

a. Curate data for drugs, proteins, diseases from online on line known resources

b. Clearing collected data and make data mining and statistics.

c. Build a large dataset containing drugs, proteins, diseases with

Curate approved drugs, Disease, Drugs	Using protein Physical interaction	Use network analysis to predict new indications	Build drug repositioning system
In this research we will focus on the first step of the drug discovery by extracting the approved drugs from known online resources related to human species.	Find new indications for existing approved drugs based on the physical interaction between genes. And based on the first step of drug design, our study is to discover new targets for Existing approved drugs, and the key point here is to join between the existing drugs and new other diseases will be based on human protein interaction	 Build network route them and make drugs and their new indications diseases. Build data set that include all of these information gene interaction, drugs, diseases and then connection between them and network analysis. 	Web based system with search engines to be published to the public to be able to query intensive data related to drug repositioning and similarity to help researchers on their work.

Table 1: Aims from research



known interaction between them with programing interface to able to query the dataset to answer questions.

d. Build network medicine to analyses the new targets for the approved drugs.

e. Strength the success of drug repositioning hypothesis between genes pairs by Apply the machine learning to find and calculate the percentage of the drug similarity.

Research Backgrounds will be Discussing within this Research

a. Protein-Protein Interactions Essentials

- b. Basic terminologies of networks and networks analysis
- c. Biological Networks
- d. Elements and principles of network theory
- e. The principles of Network Medicine

Protein-protein interactions essentials: Key concepts to building and analyzing interactome

PPI definition: Physical contacts with molecular docking between proteins that occur in a cell or in a living organism *in vivo*.

Definition considerations:

(1) Surely not all the proteins in the particular complex interact. But

by other hand, proteins share functional contact can be interact with each other and this is apply for Any protein in the ribosome or in the basal transcriptional apparatus in the complex [6].

(2) The interaction interface should be intentional and not accidental [6].

(3) The interaction interface should be no generic [6].

(4) That PPIs imply physical contact between proteins does not mean that such contacts are static or permanent [6].

(5) Not all possible interactions will occur in any cell at any time.

Two Alternative Approaches PPI Binary and Co-Complex:

Interactions between proteins are done at either a large or small scale by using two techniques:

1. Binary: yeast two-hybrid (Y2H). Measure direct interactions between proteins [6].

2. Co-complex: tandem affinity purification coupled to mass spectrometry (TAP-MS) [6].

Measure both direct and indirect interactions between proteins. Both are widely applied in large-scale investigations (Figure 3).

The main databases and repositories that include PPIs: As we mentioned before we are going to build our drugs re-poisoning network based on protein interaction, because of this we have to know the available types of interaction databases, the differences, the most comprehensive and the stable repositories. Indeed The number of known protein–protein interactions has been increased later. This information already become available, many publicly databases have founded to collect and store protein–protein interaction data. Protein– protein interactions have been can be collected from six basic databases, integrated and the results compared The six databases (the Biological



Figure 3: Binary methods and co complex methods: two approaches to determine PPIs. The interactions shown in the left panel (green links) correspond to the true interactions existing between two groups of proteins (set A with four proteins and set B with three proteins). The interactions shown in the right panels correspond to the networks derived from the experimentally measured interactions existing between the six proteins analyzed: the network in the top right panel (blue links) presents the interactions obtained using a binary method; the network in the bottom right panel (red links) presents the interactions obtained using a applying the spoke model To the TAP-MS experimental data, but three of the interactions deduced (links with an X) do not occur [7].

General Repository for Interaction Datasets [BioGRID], the Molecular INTeraction database [MINT], the Bimolecular Interaction Network Database [BIND], the Database of Interacting Proteins [DIP], the IntAct molecular interaction database [IntAct] and the Human Protein Reference Database [HPRD]) [6]. With respect to human protein-protein interaction data, HPRD seems to be the most comprehensive. To obtain a complete dataset, however, interactions from all six databases such as the Agile Protein Interaction Database (APID) offer access to integrated protein–protein interactions for datasets, although these also currently have certain restrictions [6]. A comparison of the main databases and repositories that include protein interactions is shown in (Table 2).

Analysis of coverage and ways to improve ppi reliability:

1. A first obstacle to evaluate the reliability of PPIs is the low coverage of the databases for each specific interactome.

2. One way to increase coverage is to integrate data reported by different primary databases The following (Figure 4) is example, the data on human PPIs coming from six different primary databases show a small overlap [6].

Basic terminologies of networks and networks analysis

To able to build our study and understand the networks we have to go through Basic terminologies of networks and networks analysis [7,8].

Network science:

Main definition: Network analysis is a recently new area of data analysis. From a data science point of view, a network is a collection of interconnected Objects. We may call the network objects "nodes," "vertexes," or "actors," and call the connections between them "arcs," "edges," "links," or "ties." We may represent networks graphically and mathematically as graphs. Mathematically speaking, a graph is a set of nodes connected with edges [7,8].

J Proteomics Bioinform, an open access journal ISSN: 0974-276X

Acronym	Database Full Name and URL	PPI Sources	Type of MI	Species	<i>n</i> Proteins (Dec. 2009)	<i>n</i> Interactions (Dec. 2009)
Primary Data	bases: PPI experimental data (curated from s studies)	pecific SSc & LSc published				
BIND	Biomolecular Interaction Network Database, http://bond. unleashedinformatics.com/	Ssc & Lsc published studies (literature-curated)	PPIs & others	All	[31,972]	[58,266]
BioGRID	Biological General Repository for Interaction Datasets, http://www.thebiogrid.org/	Ssc & Lsc published studies (literature-curated)	PPIs & others	All	[28,717]	[108,691]
DIP	Database of Interacting Proteins, http://dip. doe-mbi.ucla.edu/dip/	Ssc & Lsc published studies (literature-curated)	Only PPIs	All	20,728	57,683
HPRD	Human Protein Reference Database, http:// www.hprd.org/	Ssc & Lsc published studies (literature-curated)	Only PPIs	Human	27,081	38,806
IntAct	IntAct Molecular Interaction Database, http:// www.ebi.ac.uk/intact/	Ssc & Lsc published studies (literature-curated)	PPIs & others	All	[60,504]	[202,826]
MINT	Molecular INTeraction database, http://mint. bio.uniroma2.it/mint/	Ssc & Lsc published studies (literature-curated)	Only PPIs	All	30,089	83,744
MIPS-MPact	MIPS protein interaction resource on yeast, http://mips.gsf.de/ genre/proj/mpact/	Derived from CYGD	Only PPIs	Yeast	1,500	4,300
MIPS-MPPI	MIPS Mammalian Protein-Protein Interaction Database, http://mips.gsf.de/proj/ppi	Ssc published studies (literature-curated)	Only PPIs	Mammalian	982	937
Meta-Databa	ses: PPI experimental data (integrated and u repositories)	nified from different public				
APID	Agile Protein Interaction DataAnalyzer, http:// bioinfow.dep.usal.es/apid/	BIND, BioGRID, DIP, HPRD, IntAct, MINT	Only PPIs	All	56,460	3,22,579
MPIDB	The Microbial Protein Interaction Database, http://www.jcvi.org/mpidb/	BIND, DIP, IntAct, MINT, other sets (exp & litcurated)	Only PPIs	Microbial	7,810	24,295
PINA	Protein Interaction Network Analysis platform, http://csbi.ltdk. helsinki.fi/pina/	BioGRID, DIP, HPRD, IntAct, MINT, MPact	Only PPIs	All	[?]	1,88,823
Prediction Dat	abases: PPI experimental and predicted data lato sensu derived from differen	("functional interactions", i.e. t types of data)	, interactions			
MiMI	Michigan Molecular Interactions, http://mimi. ncibi.org/MimiWeb/	BIND, BioGRID, DIP, HPRD, IntAct, & nonPPI data	PPIs & others	All	[45,452]	[391,386]
PIPs	Human PPI Prediction database, http://www. compbio.dundee. ac.uk/www-pips/	BIND, DIP, HPRD, OPHID, & nonPPI data	PPIs & others	Human	[?]	[37,606]
OPHID	Online Predicted Human Interaction Database, http://ophid. utoronto.ca/	BIND, BioGRID, HPRD, IntAct, MINT, MPact, & nonPPI data	PPIs & others	Human	[?]	[424,066]
STRING	Known and Predicted Protein-Protein Interactions, http://string. embl.de/	BIND, BioGRID, DIP, HPRD, IntAct, MINT, & nonPPI data	PPIs & others	All	[2,590,259]	[88,633,860]
UniHI	Unified Human Interactome, http://www.mdc- berlin.de/unihi/	BIND, BioGRID, DIP, HPRD, IntAct, MINT, & nonPPI data	PPIs & others	Human	[22,307]	[200,473]

Table 2: The table divided in three sections: primary databases, which include PPIs from large- and small-scale (Lsc & Ssc) experimental data that are usually obtained from curation of research articles (8 resources included: BIND, BioGRID, DIP, HPRD, IntAct, MINT, MIPS-MPACT, MIPS-MPPI); meta-databases, which include PPIs derived from integration and unification of several primary repositories (3 resources: APID, MPIDB, PINA); prediction databases, which include PPIs from experimental analyses together with predicted PPIs obtained from the analyses of heterogenous biological data (5 resources: MiMI, PIPS, OPHID, STRING, UniHI). The table shows the total number of proteins and interactions that were reported by each repository in December 2009 (as far as we could see in the respective Web site). The numbers are in brackets [] when the repository includes PPIs and other types of interactions (e.g., protein-ligand interactions or for the case of prediction databases nonPPI data). The question mark [?] indicates that the number of distinct proteins included is such repository could not be found in the Web. doi:10.1371/journal.pcbi.1000807.t001._

Graph elements, types, and density:

1. Digraph: Is directed graph that for example connects node A to node B but not the other way around.

2. Multigraph: parallel edges in a graph while node A may be connected to node B by more than one edge.

3. Simple graph: graph without loops and parallel edges.

4. Weighted graph: graph with weighted edges while weights assigned to graph edges. Weight is usually (but not necessarily) a number between 0 and 1, inclusive. The larger the weight, the stronger the connection between the nodes.

5. Degree: number of edges connected (incident) to the node (Figure 5).

6. In degree: the number of edges coming into the node and this is for directed graph.

7. Out degree: (the number of edges going out of the node).

8. Graph density d ($0 \le d \le 1$): how close the graph is to a complete graph Density for directed graph with **e** edges and **n** nodes:

Density for undirected graph:
$$d = \frac{e}{n(n-1)}$$

$$d = \frac{2e}{n(n-1)}$$

9. Diameter: The largest distance between any two nodes in a graph A.



Analysis of human interactome PPI data showing the coverage of six major primary Databases (BIND, BioGRID, DIP, HPRD, IntAct, and MINT). C) Each database lacks a substantial proportion of the total reported PPIs.



10. Connected component: set of all nodes in a graph such that there is a path from each node in the set to each other node in the set.

11. A clique: set of nodes such that each node is directly connected to each other node in the set.

12. Neighbourhood (G (A): A set of nodes directly connected to a node A.

13. The local clustering coefficient of a node A: density of the neighborhood of A without the node A itself. The clustering coefficient of any node in a star is 0. The clustering coefficient of any node in a complete graph is 1.

14. Network community: set of nodes such that the number of edges interconnecting these nodes is much larger than the number of edges crossing the community boundary measure of quality of community structure.

15. Centralities: measure of the importance of a node in a network. Often scaled to the range between 0 (an unimportant, peripheral node) and 1 (an important, central node).

Type of Centralities:

1-Degree centrality Closeness

2-Betweenness centrality

3-Eigenvector centrality

Network analysis sequence

1- Identifying discrete entities and the relations between them. The entities become the network nodes, and the relations become edges network.

2- Measures are calculated: density, number of components, GCC size, diameter, centralities, clustering coefficients, and so on.

3- Network communities are identified. If the network ends up being modular. Finally, results are interpreted, and a report with a lot of appealing pictures is produced Using net Using networkx tools for example.

Tools to exploring and analyzing a network

Here we will explorer some known tools as example:

networkx: The networkx module included in Python contains essential tools for creating, modifying, exploring, plotting, exporting, and importing networks. It supports simple and directed graphs and multigraphs. By looking into the values inside the graph we have to call many functions connected to networkx:

To get the number of nodes:

len(borders)

To list nodes: borders.nodes()

To list edges: borders.edges()

List of neighbours: borders.neighbors("Germany")

• ['Czech Republic', 'France', 'Netherlands, Kingdom of the', 'Denmark', 'Switzerland', 'Belgium', 'Netherlands', 'Luxembourg', 'Poland', 'Austria']

- To calculate the number of degree related to specific node borders.degree("Poland")

•

- To list the number of degrees for all nodes: borders.degree()

• {'Iran': 8, 'Nigeria': 4, 'Chad': 6, 'Bulgaria': 5, 'France': 14, 'Lebanon': 2, 'Namibia': 4, «...»}

- List dictionary of all clustering coefficient for all nodes: nx.clustering(borders)

{'Iran': 0.2857142857142857, 'Nigeria': 0.5, 'Chad': 0.4, 'Bulgaria': 0.4, 'France': 0.12087912087912088, 'Lebanon': 1.0, 'Namibia': 0.5, «...»} nx.clustering(borders, "Lithuania")

0.833333333333333334

- List the connected component to create subgraph: list(nx. connected_components(borders))

[{'Iran', 'Chad', 'Bulgaria', 'Latvia', 'France', 'Western Sahara', «...»}]

- To calculate which node has the highest centrality:

nx.degree_centralit(borders) # People's Republic of China

nx.in_degree_centrality(borders)nx.out_degree_centrality(borders)nx.closeness_centrality(borders)#Francenx.betweenness_centrality(borders)#Francenx.betweenness_

- Functions detect maximal cliques:

find_cliques() and isolates() to find (zero-degree nodes).

list(nx.find_cliques(borders))

[['Iran', 'Nagorno-Karabakh Republic', 'Armenia', 'Azerbaijan'],

• ['Iran', 'Afghanistan', 'Pakistan'], «...»]

nx.isolates(borders)

• ['Penguinia']

2. NetworKit: Is highly efficient and parallelizable network analysis toolkit suitable for the large network analysis rather than networkx. NetworKit developers claim that "community detection in a 3 billion edge web graph can be performed on a 16-core server in a matter of minutes". NetworKit is integrated with matplotlib, scipy, numpy, pandas, and Networkx.

3. Gephi: An interactive visualization and exploration platform for all kinds of networks and complex systems.

Biological networks

In this research, we go to use the network concept as one of the data science tool analysis to make the primary analysis of our data set.

Biological network types and characteristics: Networks of coupled dynamical systems have been used to model biological oscillators, Josephson junction arrays, excitable media, neural networks, spatial games, genetic control networks and many other self-organizing systems. Ordinarily, the connection topology is assumed there are two kind of networks that will either completely regular or completely random, But many biological, technological and social networks located somewhere between these two [9].

Regular network (large world) and Random network (small world): Regular networks 'rewired' to present increasing amounts of disorder. These systems can be highly clustered, we call them 'small-world' networks, and simple 'majority-rule' running on a small-world graph can outperform all known human and genetic algorithm-generated rules running on a ring lattice [9]. In this graph is indication of the Random rewiring procedure for interpolating between a regular ring, lattice and a random network, Start with a ring of n vertices, each connected to its k nearest neighbors by undirected edges in a clockwise sense [9]. With probability p, we reconnect this edge to a vertex chosen uniformly at random over the entire ring, Different values of p. For p = 0, the original ring is unchanged Regular network; as p increases, the graph becomes increasingly disordered until for p = 1, all edges are rewired randomly random network (Figure 6).

Characteristic path length L(p) and clustering coefficient C(p):

The family of randomly rewired graphs described as following:

- 1. Characteristic path length L(p): number of edges in the shortest path Between two vertices, averaged over all pairs of vertices (Figure 7) [9].
- 2. Clustering coefficient C(p): C measures the cliquishness of a typical Friendship circle [9]. The ratio of number of Ei of edges that exist among neighbors, over the number of edges that could exist. Suppose that a vertex v has kv neighbors then

J Proteomics Bioinform, an open access journal ISSN: 0974-276X at most kv(Kv-1)/2 edges can exist between them (this occurs when every neighbor of v is connected to every other neighbor of v). Define C as the average of Cv over all v (Figure 8) [9].

- **3.** Characteristics of network whereas intermediate values of **p**: The graph is a small-world network: Highly clustered like a regular graph, yet with small characteristic path length, like a random graph (Figure 9).
- **4.** Characteristics of regular network whereas p close to 0: Highly clustering coefficient C(p), Highly Characteristic path length L(p) and large world (Figure 10).
- 5. Characteristics of random network whereas P close to 1: Low clustering coefficient C(p), Low Characteristic path length L(p) and Small world (Figure 11).

Graph theory of Erdo"s and Re'nyi (ER): Networks of complex topology have been described with the random graph theory of Erdo"s and Re'nyi (ER) but in the absence of data on large networks, the predictions of the ER theory were rarely tested in the real world [10]. To build the graph: start with N nodes then with probability of p connects each pair of nodes. Here graph is created with approximately pN(N-1)/2 randomly placed links. The node degrees follow a Poisson distribution which indicates:









1. That most nodes have approximately the same number of links (close to the average degree <k>).

2. The tail (high k region) of the degree distribution P(k) decreases exponentially, which indicates that nodes that significantly deviate from the average are extremely rare Independent of the system [10] and the identity of its constituents, the probability P(k) that a vertex in the network interacts with k other vertices decays as a power law: P(k) \sim k -yAs we mentioned before In the small-world model that introduced by Watts and Strogatz (WS). N vertices form a one-dimensional lattice, each vertex being connected to its two nearest and next-nearest neighbours. With probability p, each edge is reconnected to a vertex chosen at random. The long-range connections generated by this process decrease the distance between the vertices, leading to a small-world phenomenon. A common feature of the ER and WS models is that the probability of finding a highly connected vertex (that is, a large k) decreases exponentially with k; thus, vertices with large connectivity are practically absent. In contrast, the power-law tail characterizing P(k) for the networks studied indicates that highly connected (large k) vertices have a large chance of occurring, dominating the connectivity [10].

Network medicine: a network-based approach to human disease: Most cellular components exert their functions through interactions with other cellular components, which can be located either in the same cell or across cells, and even across organs. Which resulting complex network in the human intercom. Network-based approaches to human disease have multiple potential biological and clinical applications. A better understanding of the effects of cellular interconnectedness on disease progression may lead to the identification of disease genes and disease pathways, which, in turn, may offer better targets for drug development [11].

Biological network maps

In the following (Figure 12) and (Figure 13), we briefly discuss the most studied network maps and their limitations [11,12].

Regulatory networks

Nodes are either proteins or a putative DNA regulatory element and directed edges represent:

Techniques: ChIP-chip and ChIP-seq

Databases: UniPROBE, JASPAR, TRANSFAC, BCI

Databases: PhosphoELM, PhosphoSite, PHOSIDA (Figure 14) [11,12].

RNA networks

They capture the interactions between RNAs and DNA in regulating gene expression.

Nodes represent small non-coding RNAs (miRNAs) or small interfering RNAs (siRNAs) and DNAregulatory elements. Links represent regulation.

Databases:

1. Predicted microRNA targets: TargetScan, PicTar, microRNA, miRBase, miRDB

2. Experimentally supported targets: TarBase, miRecords

Elements and principles of network theory

In the following, we summarize the aspects of network theory that pertain to biological networks.

Modules: Topological modules are a way that represents highly interlinked local regions in the high degree of clustering inside most networks. Although the identification of such modules can be computationally challenging, a wide array of network-clustering tools have emerged over the past few years [11,12].

Degree distribution and hubs:

- 1. In case the most nodes have approximately the same number of links, and highly connected nodes (hubs) are rare, so we talking about here about random network.
- 2. While the degree distribution follows the well-known Poisson distribution so here will find the fraction of links with a given degree [11].
- 3. In real networks, human protein–protein interaction and metabolic networks are example of scale free, which means that the degree distribution has a power-law tail; that is, the degree distribution P(k), with degree k, follows $P(k) \sim ky$ where y is called the degree exponent. Highly connected hubs that hold the whole network together [11].

Small-world phenomena: If there are relatively short paths between any pair of nodes, so that exactly the small-world property which means that most proteins (or metabolites) are only a few interactions (or reactions) from any other proteins (metabolites Motifs a group of nodes that link to each other, forming a small sub network within a network) [11].

Betweenness centrality: Nodes with a high betweenness centrality definition is (a measure of the number of shortest paths that go through each node) are often called bottlenecks. Regulatory networks which are

J Proteomics Bioinform, an open access journal ISSN: 0974-276X





example of networks with directed edges, bottlenecks tend to correlate with essentiality hypotheses and organizing principles that link network structure to biological function and disease [11].

The principles of network medicine

Hubs: Disease genes tend to avoid hubs and segregate at the functional periphery of the interactome. In human's essential genes, not disease genes are encoded in hubs [11].

Local hypothesis: If a gene or molecule is involved in a specific biochemical process or disease, its direct interactors might also be suspected to have some role in the same biochemical process and this is exactly similar to hypothesis of this research. Proteins involved in the same disease have an increased tendency to interact with each other [11].

Corollary of the local hypothesis: Mutations in interacting proteins often lead to similar disease phenotypes [11].

Disease module hypothesis: Cellular components associated with a specific disease phenotype show a tendency to cluster in the same network neighborhood. Network parsimony principle Causal molecular pathways often coincide with the shortest molecular paths between known disease-associated components [12].

Shared components hypothesis: Diseases that share diseaseassociated cellular components (genes, proteins, metabolites or microRNAs) show phenotypic similarity and comorbidity. Local clustering of disease genes: disease modules Based on network science principle Modules that we mentioned before and network medicine principle local' hypothesis the disease modules concept comes.

Gene associated with a specific disease tend to cluster in the same neighborhood: As we said before the process or disease, its direct interactors might also be suspected to have some role in the same biochemical process. In line with this 'local' hypothesis, proteins that are involved in the same disease show a high propensity to interact with each other see (Figure 15).

Based on this principle we have 3 modules

1. **'Topological module':** a locally dense neighborhood in a network, such that nodes have a higher tendency to link to nodes within the same local neighborhood than to nodes outside it.

2. **'Functional module':** nodes of similar or related function in the same network neighborhood, where function captures the role of a gene in defining detectable phenotypes.

3. 'Disease module': a group of network components that together

J Proteomics Bioinform, an open access journal ISSN: 0974-276X



The Drug Bank database is a unique bioinformatics and cheminformatics resource that combines detailed drug.	The data on human PPIs coming from six different primary databases. Or by other words, Analysis of human interactome PPI data showing the coverage of six major primary databases (BIND, BioGRID, DIP, HPRD, IntAct, and MINT).	DisGeNET is a discovery platform containing one of the largest publicly available collections of genes and variants associated to human diseases. The original data sources for the human included in DisGeNET are from the following:
Drug Bank is database while the users can search or download complete data set for the drugs.	With respect to human protein-protein interaction data, HPRD seems to be the most comprehensive, so we will extract PPI data from HPRD database to use on our research.	GDAs from UniProt, PsyGeNET, ClinVar, Orphanet, the GWAS Catalog, CTD (human data), and Human Phenotype Ontology.

Table 3: Data for drugs, proteins, diseases from online known resources.

resources

(Table 3) [13-16].



Figure 15: Proteins that are involved in the same disease show a high propensity to interact with each other.

contribute to a cellular function and disruption of which results in a particular disease phenotype [11].

Drug Repositioning Network System Process

There are different kinds of similarity that can be used to predict another indication to approved target; my system will depend on protein interaction to find another target for existing drugs (Figure 16).

species and calculated 1474 gene names that mapped to their drugs. The following Chart indicates the distribution of the connections between the drugs and each genes inside the drug database. After filtering the

Raw data filtering and initial statistics

the drugs and each genes inside the drug database. After filtering the Drug database based on human species, we found that there are 1474 Genes. Each gene has connection with group of drugs whilst the total number of connections between all genes and their drugs groups are 1611. This means that some genes have more than one connections with more than different group of drugs. For example: chart connections axe is the number of connections for each gene and the distribution has tail indicates that the majority of Genes connected with one group of drugs. Form original data we will find that the Gene name "HTR1B" has 3 connections with different group of drugs (Figure 17).

Curate data for drugs, proteins, diseases from online known

Drug data set: Raw drugs data set has been filtered by using human

Diseases data set: The following Chart indicates the distribution of the connections between the diseases and each genes inside the Diseases database. After filtering the diseases database based in comparison with human drug genes, we found that there are 1219 Genes. Each gene has connection with group of diseases or one disease while the total number connections between all genes and their diseases groups are 31928. This means that some genes have more than one connections with more than different group of diseases. For example: chart connections axe is the number of connections for each gene, the distribution has tail indicates that the majority of Genes have less than 50 connections





with diseases. Note: the connection relation between the Genes and diseases is many to many, this means that it can be one gene connects to many diseases and it can be one disease connects to many genes. Form original data we will find that the Gene name "HRAS" has 230 connections with the diseases while "CFB" has 10 connections with the diseases (Figure 18).

PPI data set: The following Chart indicates the distribution of the binary Interactions between Proteins.

After filtering the protein interaction database based in comparison with human drug genes, we found that there are 1472 Genes. Each gene has interaction with another pair of genes. The total number of interactions between all genes is 10952. This means that some genes have more than one interaction with more than one another genes. For example: chart connections axe is the number of interactions for each gene, the distribution has tail indicates that the majority of Genes have less than 20 interactions with other genes. Note: the gene interaction relation between the pair Genes is many to many, this means that it can be one gene interacts with many other genes and vice versa. Form original data we will find that the Gene name "SMAD" has 90 interactions with other genes while "PYGM" has 3 interactions with other genes (Figure 19).

Drug Repositioning Network System Manual

On this part we will explorer the first prototype of the Drug Repositioning Network System and how to use to predict and discover new targets. The system has two main parts backend and frontend where the backend Back end of the system uses Big Data tools and network analysis to provide any queries and questions coming from



Chart indicates the distribution of the binary Interactions between the genes
The gene interaction relation between the pair Genes is many to many, this means that it can be one gene interacts with many other genes and vice versa.
The distribution has tail indicates that the majority of Genes have less than 20 interactions with other genes



users using the frontend web interface and search engines. Backend of the system is supply chain for many search engines used by end user to extract many kinds of data, make predictions and analysis and answer questions. Fronded system where contains search engines where the users can "Find new indications for approved drugs based on Gene interaction". The purpose from using this part is achieving the main aims form that research and answer questions like "are there any new diseases can the approved drugs target ?" or "are there any approved drugs can target existing disease rather than its original drugs". By

answer this questions, we trying to predict a new indications for the approved drug. Also to make network analysis after building route path between the pair genes and their corresponding drugs and diseases. The Snapshots figures show example how to use the web interface of the Drug Repositioning system.

Case 1

At first snapshot

For example search for "Obesity" disease on the first snapshot indicates how the data appeared on the disease table columns. The "Obesity" disease might be appeared on the diseaseName1 column or on diseaseName2 and the both genes interactors will appear on the Gene1 column and Gene2 column. If the "Obesity" appeared on diseaseName1 column then the gene that will appear on Gene1 column is the original gene of the "Obesity", for example "TUB" gene. And the gene that appears on Gene2 column, for example "LCK" is the interactor pair of the "TUB", and "LCK" gene also is the original of the disease that will appear on diseaseName2 column for example "Diarrhea" disease. It is noticeable that "LCK" gene connected to 12 diseases at the same time. As result, our hypothesis predicts that all drugs originally target all the 12 diseases of "LCK" gene can also target "Obesity" disease (Figure 20).

At the second snapshot

We use the gene of "Obesity" disease "TUB" to search on "Drug search engine to know the original drugs related to "Obesity" disease and another predicable groups of drugs that can target "Obesity" disease For example: "DB02028" is original drug for "Obesity disease", while "DB01064; DB05210; DB08059" are drugs group that connected to "PIK3R1" gene and new predictable drugs target "Obesity disease" (Figure 21).

At the snapshot 3

We search by the drug ID on "**Drug search engine**" and use the gene of the drug that will appear on the result to search on "**Disease search engine**" to know the diseases that originally related to the drug and the other predictable diseases that can be targeted by the drug. For example: we can search on "**Drug search engine**" for "**Aspirin**" drug

by ID, we can get the Aspirin ID from **drugBank** site. Here Aspirin ID is (**DB00945**) and "**NFKB1**"gene is connected to Aspirin which we can use to search on "**Disease search engine**" to know which diseases targeted by "**Aspirin**" and other predictable diseases can be targeted by "**Aspirin**". We can make this exercise many times on each gene connected to "**Aspirin**" (Figure 22).

Case 2

Imagine that we need to make test and experiment between two of genes, for example we need to know if there is protein interaction between them or not, and if there is connection between them we can build network route path between the drugs and diseases for gene1 and the drugs and diseases for gene2. From the system we search for gene1 to generate its related drugs and disease and search for gene2 to generate its drugs and disease and then generate network path between them and do the drug repositioning between the two genes. For example at the snapshot 4 (Figure 23). We search for HDAC6 as gene1 and generate all its related drugs and disease. And then at snapshot 5 (Figure 24) we search for TUBB as gene2 and generate all its related drugs and diseases.





J Proteomics Bioinform, an open access journal ISSN: 0974-276X



Search for HDAC6 gene		Drugs and Diseases relat	ted to HDAC6 gene	Genes have interaction with HDAC6 gene		
← → ○ loca HD406	ihost, newi Johp net-Giosexe (-Drug) search engine GD>>		4	Image: The second se		
Drug1	Diseas1		Interactor_1_Gene_symb	ol Interactor_2_Gene_symbol		
D802546; D805223; D806176	Prostatic Neoplasms	/	HDAC6	TUBB		
D802546; D805223; D8063	Prostatic Neoplasms		HDAC6	TUBA18		
D802546; D805223; D806176	Prostatic Neoplasms		HDAC6	TUBAAA		
D802546; D805223; D806176	Prostatic Neoplasms		HD4C6	EP300		
D802546; D805223; D806176	Prostatic Neoplasms		HDAC6	RELB		
D802546; D805223; D806176	Prostatic Neoplasms		HD4C6	UEB		
D802546; D805223; D806176	Prostatic Neoplasms		HDAC6	UBC		
D805015; D806603	Prostatic Neoplasms		HDAC6	TU88		
D805015; D806603	Prostatic Neoplasms		HDAC6	TUBA18		
D805015; D806603	Prostatic Neoplasms		HDAC6	TUBAAA		
D805015; D806603	Prostatic Neoplasms		HD4C6	EPSOD		
D805015; D806603	Prostatic Neoplasms		HD4C6	RELB		
D805015; D805603	Prostatic Neoplasms		HDAC6	U68		
			0.00			

Discussions, Experiments, Questions and Answers

Question 1: Can we search for specific disease to know its original drugs and new predicted drugs can use or in the original genes for disease and their pair interactors?

Answer: Based on Case 1 on the Drug repositioning network System manual section, yes we can.

Question 2: Can we find new indications for existing approved drugs?

Answer: Based on Case 1 on the Drug repositioning network System manual section, yes we can.

Question 3: Can we search to know if there is any relation between one drug and disease or between two genes?

Answer: Based on case2 Drug repositioning network System manual section, yes we can.

Question 4: if we know that there is interaction between the gene of drug and the gene of disease, can we make basic analysis and plot?

Answer: Based on case2 Drug repositioning network System manual section, yes we can, For example, the following TUBB (Figure 25) is network plot for search done for TUBB gene that shows each diseases and drugs connected to TUBB gene with some initial statistics statistics like number of Nodes and connected edges included in the network And HDAC6 (Figure 26) is network plot for search done for HDAC6 gene that shows each diseases and drugs connected to HDAC6 gene with some initial statistics like number of Nodes and connected edges included in the network. In the following network (Figure 27) is example of network analysis between TUBB and HDAC6 and this after our system found connection between the two genes because there is interaction between their proteins. In the left side of the picture, it's clear that there are two groups of drugs target numerous diseases related to "HDAC6" gene, and also on the right side we will find one group of drugs targets two groups of disease that are related to "TUBB" gene while there is link between "TUBB" and "HDAC6" indicates the interaction between them. Under our predication no we can predict that the 2 groups of the drugs related to "HDAC6" gene can target the two groups of diseases related to "TUBB" and vice versa. We can study this network example that generated from the drug repositioning system from the perspective of Protein-protein interaction networks map that we discussed on section" Biological network maps" we will find that the of Protein-protein interaction networks map is a part form our network where is nodes represent proteins and edges represent a physical interaction between two proteins ("TUBB" and "HDAC6") but also each protein connect to its drugs and diseases. Reference to section Elements and principles of network theory" our network here has two main hubs ("TUBB" and "HDAC6") that connect all the network groups with each other but the question here, what if there is more than one genes have interactions? And another question, what if there are huge number of genes of the same network that have interaction and there is more than one drug and disease connected to more than one gene as the same time, the answer here will drive us to the drug similarity hypotheses or disease similarity and this is what we will discuss on the next section. Also, the network follow Degree distribution and hubs which means if we analyze the second picture

		ed domestical (m.a. = 1	
	← → 0 kcanogreestrop						
Search for 1 drugs searc Engine	FUB on nding New Indications for th	Approved Drugs Based on Gene Inte	eraction				
/	diseaseName1	Gene1	Gene2		diseaseName2		
	Drug1		Gen	1 Genei	Drug2		
	D800541; D801179; D801873; D803010; D805147; D8	25281; D806772	708	44 MYC	D808813		
	D801873; D803010; D805147		708	ac Mic	D808813		
	D802518; D800570; D800642; D801873; D803010; D8	15147; D807574	1,8	IA MYC	D608813		
	D801873; D803010; D805147; D807574		7.6	18 M/C	D608813		
	D801873; D803010		7.8	8 MiC	D606813		PIK3R1
	D801873; D803010; D805147; D807574		TUB	18 N9(8)	D800945		Gene2
	D800570		7.8	n Pige	D801064; D805210; D808059		
TUB	D801873; D803010; D805147; D807574		708	18 PKGR	D801064 D805210 D808055		Predictable
genel	D802028		1.8	PIGR	D801064; D805210; D808059		drugs can target
	DB02028		708	SRC	D801254; D801670; D801866; D801893; D	801908; D801947; D801962; D802175;	"Obesity

J Proteomics Bioinform, an open access journal ISSN: 0974-276X





"Sample of Drug repositioning network degree distribution" and the statistics (Figure 28). We will find the network average in degree and average out degree is the same 2.6875 but the number of nodes and edges not the same but still close,. And From the graph our network like human protein–protein interaction and metabolic networks are scale free, which means that the degree distribution has a power-law tail, and this is because we highly connected hubs that hold the whole network together.

Current Experiment, Drug Similarity

Main hypothesis

My experiment developed by test each pair genes have interaction to find shared drugs between their related drugs groups, and based on the number of shared drugs exist between the two groups to the total number of the drugs on the two groups we can calculate the percentage of the similarity. The percentage of the drug similarity indicates to what extent the level of drug repositioning hypothesis is high or low. For example: We know that there is interaction between the proteins "TUBB" as gene interactor2 and "HDAC6" as interactor1.

Gene "TUBB" has drug group (Drug A, Drug B, Drug C, Drug D)

Gene "HDAC6" has drug group (Drug E, Drug A, Drug B)

The shared drugs between the two groups are (Drug A, Drug B)

The percentage of drug similarity between two interactos=2/8*100=25%

The hypothesis of drug repositioning level is 25%

This means that drugs for each gene can target and suitable for the diseases for the other gene by 25%

At the following the actual experiment picture is explanation of Calculate the highest level of drug similarity similarity=100%, Calculate the similarity less than 100%, Applying ROC Curve (Table 4).

At the following (Table 4) and (Figure 29) is explanation of how to Calculate the highest level of drug similarity similarity=100% using Drug repositioning Network System (Figure 30).

Other Factors Considered to Strength Drug Repositioning Network System

Many approaches can be used to define drug similarity which is considered to be used as future development for Drug repositioning network system.

Chemical structure similarity

This approach was successfully used by implementing an unsupervised machine learning algorithm in case to cluster chemicals based on their structure. Drug repositioning using the chemical structure. Compounds sharing similar structures have similar biological activities. At the (Figure 31) Drug A shares some similarity with molecule B, indicated by the blue areas. This observation leads to the conclusion that molecule A could be active on the canonical target of molecule B, and indicated accordingly [17].





Gene expression and functional genomics similarity

Certain genes are going to be over or under expressed, identiable from the relative number of their messenger RNA (mRNA) molecules transcribed. Messenger RNA expression can reflect the activity of a drug, but it can also be used to characterize disease states [17]. This type of experiment is usually performed on a microarray [17].

At the (Figure 32) Drug repositioning using gene expression. (A) Example of result obtained from a gene expression experiment. Some of the probed genes are up-regulated (green), some of them down-regulated (red). (B) And (C) The gene expression data from the Connectivity Map provides a signature which can relate drugs on their functional aspect. For instance drug X and Y are considered similar because they share a significant amount of genes up and down related. (D) An analogue reasoning can be made with the relation drug-disease: disease signature can be treated by drugs with an anti-correlating signature [17].

Protein structure and molecular docking similarity

Many modern studies concentrated on binding sites and compared their relative similarities and that we going to use in our research. In the following picture, Drug repositioning using protein structure and binding site. It is assumed that similar binding sites can bind the same ligand. For instance at Figure 33, knowing that protein X has a similar binding site to the one in protein Y, and that molecule Z binds to protein X, one can forward the hypothesis stating that molecule Z should bind to protein Y too. Illustrations from the Protein Data Bank [17].

Phenotype and side effect-based similarity

The set of characteristics or traits attributed to an organism. Examples of phenotypes are the morphology, developmental, biochemical or physiological properties [17]. Drug repositioning using phenotype information. At the following Figure 34 is example using reported side-effects: the more side effects are commonly shared by two drugs, the more similar these two drugs are. The similarity can be used to either derive potential off targets or new indications [17].

Genetic variation-based

Genetic variations can also provide valuable insights regarding drug repositioning opportunities. In context of DNA sequencing methods and analysis pipelines, we identify genome-wide association study (GWAS) is isolate common mutations in the DNA that are significantly associated with a phenotypic trait. (GWAS) used to relate a single-nucleotide polymorphism (SNP) to a disease. The data about SNPs and their association to pathologies is indexed in databases, such as the one provided by National Human Genome Research Institute

	The actual experiment	
Calculate the highest level of drug similarity similarity=100%	Calculate the similarity less than 100%	Applying ROC curve
Highest level similarity means drug similarity 100% where the pair genes intoraclors have the same groups of drugs.	Find shared drugs between two groups of drugs related to each pair kilometers and calculate the percentage of similarity	Graphical plot that Illustrates the diagnostic ability of a binary classifier system. Plotting the true positive rate (UR) against the false positive rote (FPR) at various threshold settings and it is related in a direct and natural way of cost/benefit analysis of diagnostic decision making.
		So Based on the previous structure of our drugs similarity data which it Is generating Os if it doesn't have any shored drugs between any two groups of drugs and is If it does have shored drugs between two groups of drugs that have interaction between their pair of genes. Hero. We can use the Os as negative and is as positive to apply ROC curve.

Table 4: Explanation of calculate the highest level of drug similarity similarity=100%, Calculate the similarity less than 100%, applying ROC curve.

similarity "similarity= 100%"	FILE HOME INSERT PAGE LAYO	T FORM	MULAS DATA REVIEW VIEW					Sign in
Highest level similarity means drug similarity= 100% where the pair genes interactors have the same groups of drugs.	Čelibri · 11 · A Paste · B I U · ·	A [*] ≡ ≡ • ≡ ≡		onditional F ormat as Tal ell Styles * Style:	ormatting * ble *	Format * Cells	∑ - A Z → Z Sort & Find Filter - Selec Editing	& .t.~
1- From our repositioning system we hit empty	A1 • : X fx D	ug1						
search on "drug search engine" to generate			0			0		
all possible interactors genes with their related	1 Drugt Conol	Canal	Deug2	E	r	0	n 1	1
drugs while the search automatically will	2 DB00129: DB02159: DB02250: DB(1)7	A2M	DR00102* DR00626* DR08888					
exported the complete interaction data set	3 DB00062: DB00064 APOE	62M	DB00102; DB00626; DB08888					
for pair genes and their related drugs groups	4 DB00025: DB00031: DB00100 LRP1	A2M	DB00102; DB00626; DB08888					
to file "druggeneall.csv" on the following path	5 DB00254; DB02740; DB04464 B2M	A2M	DB00102; DB00626; DB08888					
"d:\\druggeneall.csv"	6 DB02108; DB02148; DB02685; DB(CTSB	A2M	DB00102; DB00626; DB08888					
	7 DB00786; DB01197; DB01630; DB(MMP2	A2M	DB00102; DB00626; DB08888					
2-by Run Drug Similarity tool created inside the	8 DB05066; DB05744; DB05771 IL10	A2M	DB00102; DB00626; DB08888					
drug repositioning network system will	9 DB01844; DB02114; DB02341; DB(CELA1	A2M	DB00102; DB00626; DB08888					
calculate the high level of the similarity and	10 DB00025; DB00945 HSPA5	A2M	DB00102; DB00626; DB08888					
then create file " d:\\drug\$imi.csv"	11 DB01017; DB05066; DB05133; DBCIL1B	A2M	DB00102; DB00626; DB08888					
	12 DB02405; DB03796; DB04077 PAEP	A2M	DB00102; DB00626; DB08888					
	13 DB00062; DB00064 AMBP	A2M	DB00102; DB00626; DB08888					
	14 DB01407; DB05892 NGF	A2M	DB00102; DB00626; DB08888					
	15 DB00171; DB00619; DB01254; DB(ABL1	ABL2	DB00171; DB01254; DB05184; DB07664					
	16 DB04395 EPHB2	ABL2	DB00171; DB01254; DB05184; DB07664					
	17 DB00615; DB02103; DB02935; DB(HSP908	1 MTTP	DB01094; DB04852; DB08827					
	DrugGeneall (+)			1 4				•

	🚺 🔒 🏷 🖓 🗧 🕴 drugSimi.csv	Excel (Product Activat	tion Failed)	TA	BLE TOOLS		? 街 🗕	- 8
	FILE HOME INSERT PAGE LAYOUT	FORMULAS DATA	REVIEW	VIEW	DESIGN			Sign in
<u>Calculation;</u> Now we found that there are 151	$ \begin{array}{c c} & & & \\ & & & \\ & & \\ Paste & & \\ & & \\ & & \\ & & \\ & & \\ & \\ & \\ $	= =	Get \$ \$ 5 N	neral · F · % · · % · · %	Conditional Formatting * Format as Table * Cell Styles * Styles	insert • 2	Sort & Find & Filter * Select * Editing	
gene interaction cases from 5656	E18 * : X √ fx 1							
the other words have the highest level								
of similarity 100% with the highest level	A A	B Court I	((Deural	D		t Devel	F
of drug repositioning performance.	20 DB00115: DB00200	MMAA	MUT	DR00115: DB	00200	<u> </u>	onesimi e Drugs	5100 - 1
5 1 51	54 DB00514	NCF4	NCF2	DB00514			0	1
	56 DB00514	RAC1	NCF2	DB00514			0	1
	57 DB00514	RAC2	NCF2	DB00514			0	1
	59 DB00514	CYBA	NCF2	DB00514			0	1
	60 DB00514	CYBB	NCF2	DB00514			0	1
	61 DB00514	NCF1	NCF2	DB00514			0	1
	62 DB00514	CYBA	NCF4	DB00514			0	1
	63 DB00514	NCF1	NCF4	DB00514			0	1
	64 DB00514	CYBB	NCF4	DB00514			0	1
	146 DB00945	NFKBIA	NFKB2	DB00945			0	1
	148 DB00945	NFKB1	NFKB2	DB00945			0	1
	164 DB00945	NFKB1	NFKBIA	DB00945			0	1
	326 DB00129	OAZ1	OAZ2	DB00129			0	1
	695 DB00242	POLE4	POLE	DB00242			0	1
	802 DB00945	PRKAB2	PRKAA1	DB00945			0	1

Figure 30: Calculation of the highest level of drug similarity similarity=100% using Drug repositioning Network System.



(http://www.genome.gov/gwastudies/). The basic idea is to use (GWAS) database to find new indication for protein target. The association between a SNP and a trait from a GWAS is the relation between a gene and a disease. Then knowing that a drug targets the given

On (A) side: Single-nucleotide poly-morphism (SNP) are associated with a phenotypic trait, here LDL cholesterol. The gene where the SNP is found (HMGCR) encodes for a protein targeted by statins (drug class). Statins are indicated as cholesterol lowering agents, which is confirmed by the trait associated with the SNP [17].

On (B) side: Sometimes the trait associated with the SNP diverges from the indication of the drug, as shown on the diagram (post-traumatic stress disorder against smoking cessation).

Disease network-based

Diseases have been grouped together, the cause of based on the pathology or infection or the biological dysfunction observed. Similar diseases are treated in a similar fashion. The relation holding between pathologies can generate drug repositioning hypotheses. In the following Figure 36 Drug repositioning using disease relationships.

- (A) The similarity calculated by looking at using disease relationships (diseasome).
- (B) The similarity calculated by looking at the commonly shared pathways.
- (C) The similarity calculated by looking at the shared drugs used for the treatment of these diseases.

Machine learning and concepts combination

Train a machine-learning Algorithm and then generate predictions out of the statistical model. First a series of biomedical heuristics is defined, then the model is trained on known data and predictions are made. In the following Figure 37 a machine learning algorithm is trained over a series of features, such as chemical similarity, shared target proteins etc. [17].







Figure 37: A machine learning algorithm is trained over a series of features, such as chemical similarity, shared target proteins. After evaluation of the model, some repositioning predictions can be generated from the statistical learning. Two recent studies address drug repositioning from Machine learning. The first method presented is called PREDICT that use drug-drug and disease- disease associations separately. The second method using Support Vector Machine (SVM) that use the 2 kinds of structural similarity, protein-protein interaction network distance and gene expression (2).

Conclusion

I can say that the main hypothesis of the drug repositioning network system is similar to local hypothesis of the network disease which proposes that the Proteins involved in the same disease have an increased tendency to interact with each other and this is because, If a gene or molecule is involved in a specific biochemical process or disease, its direct interactors might also be suspected to have some role in the same biochemical process. From this perspective, the drug repositioning system considering that the most important factor which can be used to predict high success rate of repositioning is the protein interaction and mainly depends on HPRD database which has the highest proportion of the human PPI with overlapping of the other databases, a s result, the networks that the drug repositioning system generates have the same characteristics of the network disease and the principles of networks medicine. The drug repositioning network system goes through the similarity of drugs based on the percentage of shard drugs exist on each prediction result to calculate the success rate of the drug repositioning. Undoubtedly, the drug repositioning network system will use another factors in the future alongside the protein interaction to strength the success rate of the drug repositioning, the drug repositioning network system web interface is on the phase of the development which will be Published to the public very soon, we can say that repositioning network that can the system introduces to the professional researchers is unified repository place for drugs, diseases, and genes data which are correlate to each other based on genes shared between them. Because the drug discovery in the past was taking long time to search on huge unstructured data which hard to correlate and extract accurate probability, so the predication results that can be generated from the system, can reduce the cost of time and money and probabilities for the researchers who interested in drugs discovery. Finally, this research and system are the first edition, as we mentioned the drug repositioning network system will go through many other phases that will acquire other factors besides PPI in addition using many machine learning techniques to strength the prediction results which we able then to test in the lab, this means that this research edition is one of series editions that will be published on the future but to be honest this will depend on the financial support the system will receive.

References

- Barratt MJ, Frail DE (2012) Drug repositioning: Bringing new life to shelved assets and existing drugs.
- 2. http://www.cresset-group.com/2012/12/new-from-old/

J Proteomics Bioinform, an open access journal ISSN: 0974-276X

- 3. https://www.ebi.ac.uk/training/online/course/functional-genomics-iintroduction-and-designing-e/functional-genomics-case-studies/drug-disc
- 4. https://en.wikipedia.org/wiki/Drug_repositioning
- 5. https://www.goodreads.com/book/show/1422397.Dark_Remedy
- De Las Rivas J, Fontanillo C (2010) Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. PLoS Comput Biol 6: e1000807.
- 7. Data science essential in python book.
- 8. Social Network Analysis for Startups: Finding Connections on the Social Web.
- 9. Duncan J Watts* & Steven H Strogatz.
- A.L. Barabasi, R. Albert, Science, Vol. 286, 1999. https://www.uvm.edu/ pdodds/files/papers/others/1999/barabasi1999a.pdf

- 11. Albert-László Barabási*‡§, Natali Gulbahce*‡|| and Joseph Loscalzo§ https:// www.ncbi.nlm.nih.gov/pmc/articles/PMC3140052/
- 12. © A. Paccanaro, RHUL 2016, Network Science, biological networks.
- 13. https://www.drugbank.ca
- Prasad TSK, Goel R, Kandasamy K, Kumar S, Keerthikumar S (2009) Human Protein Reference Database - 2009 Update. Nucleic Acids Res 37: D767-D772.
- Piñero J, Bravo A, Rosinach NQ, Sacristán AG, Pons JD, et al. (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucl Acids Res 45: D833-D839.
- Piñero J, Rosinach NQ, Bravo A, Pons JD, Mehren AB, et al. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database 2015: bav028.
- 17. Croset S (2014) Drug repositioning and indication discovery using description logics.