ICT Focus

http://www.ictfocus.org

JOURNAL ARTICLE OPEN ACCESS

# Biomolecular Network-based Study of a Parasitic Disease and Therapeutic Drugs

Altankhuyag Avirmed[1], Uranchimeg Erdenedalai[1], Selenge Erdenechimeg[2], Yansen Su[3], Tseren-Onolt Ishdorj[1]

[1]School of Information and Communication Technology, Mongolian University of Science and Technology, Ulaanbaatar – 13345, Mongolia
[2]Mongolian University of Pharmaceutical Science, Ulaanbaatar – 14250, Mongolia
[3]School of Artificial Intelligence, Anhui University, Anhui – Hefei, China

e-mail: tseren-onolt@must.edu.mn

## *Abstract*

Computational drug repurposing methods, particularly biomolecular network-based disease-drug-target interaction models, are essential tools for integrating large-scale heterogenous molecular information and revealing functional mechanisms, as well as for main regulatory modules of interactants which can be useful in developing new drugs. In the present study, a drug-centric network for a parasitic disease (*Echinococcosis*) and therapeutic drugs have been considered. A complex network with more than 12,000 vertices and more than 33,000 edges representing interactions of 84 echinococcosis-related genes with associated proteins was built and analyzed. The networks of disease similarity and drug similarity were constructed based on the complex network. As a result, three drugs (D08356, D00701, and D00506) associated with three candidate diseases through three pathways and a protein complex have been extracted. This effort tries to predict the anti-echinococcosis effects of the drugs' combinations with benzimidazole.

***Keywords:*** **drug repurposing, interaction network, machine-learning, parasitic disease.**

Received: August 25, 2022.
Accepted: September 24, 2022.
Published: September 29, 2022.

Corresponding author: Tseren-Onolt Ishdorj

## 1. INTRODUCTION

In the past decades, the experimental approaches of drug development have been used heavily. However, such processes can be quite time-consuming and costly as it would take more than a decade and more than a billion dollars to introduce a new drug to the market, [35].

In order to improve the situation, it is imperative to develop fast and reliable computational methods, such as virtual screening for drug development, [22, 35, 36]. On the other hand, instead of traditional drug discovery methods, computational drug repurposing approach based on the already existing FDA-approved (Food and Drug Administration) drugs has been used effectively. This accelerates drug development time, lowers the cost and failure rates, and increases the drug-target accuracy and safety of humans, [15, 28, 34, 35]. Other diseases and drug side effects cause poor drug sensitivity and less therapeutic efficacy while time progresses. As a result, the patients need to increase drug doses, dosage regimen, as well as the combination of other drugs, etc., in efforts to be cured of their diseases. This, in turn, would cause many negative effects on various organs of the human body. A fast way to adapt to the new situation is drug repurposing, [32]. In general, computer-aided models of drug design and repurposing, simulation, and computer vision technologies have come to play an important role in systems biology and medicine research to understand complicated molecular interaction mechanisms and complex regulatory functions. A variety of modelling tools have been developed to simulate biochemical interactions, gene transcription kinetics, metabolic control, and drug delivery pathway mechanisms, which helps us to systematically test, and experimentally verify knowledge of biological and medicinal processes, [8, 31, 40, 43]. Network-based modelling technique is a suitable tool for drug and disease-related studies, [11, 20, 30, 39, 41].

We conducted here a research finding and exploring possible drug combinations against *Echinococcosis*, [23]. In this regard, a multi-type dataset of echinococcosis-related genes, proteins, and drugs were collected from various public sources, [1, 3, 12]; different heterogeneous interaction networks based on collected information were built; a qualitative analysis of interaction networks suggested that a certain principle of genes, proteins, pathways of drugs are important in network regulations. Different modelling platforms were deployed in the construction and analysis of molecular interaction networks - to mention some: `Cytoscape, STICTH, GeneCard, KEGG, STRINGS`, [1, 4, 6, 25, 26, 37].

The mathematical and biological modelling foundations and data sources, on which the article was based, are presented in the next section.

## 2. PRELIMINARY AND METHODS

### 2.1. Biological network analysis

Let us introduce some notions and notations of graph and network characteristics used in the background of biological network construction and analysis.

A biological network is used not only in representing biological processes, but also in analytic and hypothesis formulation, [11]. The analysis and visualization of biologically relevant networks representing metabolic, regulatory, or signaling pathways, protein-protein or genetic interactions, or connections between similar ligands has become commonplace. With the advent of high-throughput methods that generate vast amounts of data from diverse measurement sources, biological networks have become increasingly important as an integrating context for data, [20]. Network tools give functionality for studying complex processes. We can analyze the global characteristics of the data via metrics such as degree, clustering, shortest path, centrality, and density, [19]. We can

identify key elements, and important subnets which could help us explore interaction mechanisms, modularity, etc.

For instance, the stress of a node in a biological network such as a protein-signaling network indicates the relevance of a protein as functionally capable of holding together communicating nodes. The higher the value, the higher the relevance of the protein in connecting regulatory molecules. A network *motif* is a pattern of connectivity that occurs more frequently than might be expected by a random connection of nodes. As might be expected by the reuse we often see in biology, biological networks tend to have a small set of network motifs that act like components in a larger circuit, [5]. Biological networks are classified into certain categories such as pathways, similarity networks, regulatory networks, and interaction networks, [24, 29].

In mathematical terms, a biological network is a graph written as $G = (V(G), E(G), \phi_G)$ where $V(G)$ is the set of vertices (nodes) and $E(G)$ is the set of edges in the graph. $\phi_G$ is the set of incidence functions that define which edge goes with which vertices. The edges between nodes can either be directed or undirected. A few measurement attributes related to network structure and its vertices and edges were considered as follows.

*Node degree* `deg(v)` of node $v$ is the number of edges connected to this node. In a directed network, the node *indegree* is the number of edges directed towards this node, and the node *outdegree* is the number of edges directed away from this node. The length of a path is the number of edges forming it, and there can be many paths connecting two given vertices. An edge-weighted directed network is a digraph where weights associated with each edge.

*A shortest path/distance* from vertex $s$ to vertex $t$ is a directed path such that no other paths exist with a lower weight. The length of the shortest path is denoted as $S(s,t)$. Vertex connectivity refers to the number of its neighbors. For a vertex $n$, the neighborhood connectivity is determined by the average of all its neighbor connections.

*Network diameter* is the longest of the shortest paths between two vertices. If the network is disconnected, its diameter is determined by the longest of the diameters of each connected component.

*Clustering coefficient* is a measure of the degree to which nodes form a complete graph. The local clustering coefficient of undirected graphs is defined as $C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)}$. The average clustering coefficient is $\overline{C} = \frac{1}{n}\Sigma_{i=1}^{n} C_i$

Another important network measurement is *centrality*, it indicates which node takes up critical position in one whole network. Most typical degree centrality measures are *degree centrality, betweenness centrality and closeness centrality.*

*Degree centrality* of a node $v$ is calculated $C_D(v) = \frac{deg(v)}{(n-1)(n-2)}$, where $n$ is the number of nodes in the network.

*Stress centrality* of vertex $v$ is the number of all shortest paths passing through it: $C_S(v) = \Sigma_{s\neq v\neq t\in V}\rho_{st}(v)$, where $\rho_{st}(v)$ is the number of shortest paths passing through $v$. *Stress centrality* of edge $e$ is $C_S(e) = \Sigma_{s\in V}\Sigma_{t\in V}\rho_{st}(e)$, where $\rho_{st}(e)$ is the number of shortest paths containing edge $e$.

In both cases, stress centrality measures the amount of communication that passes an element in an all-to-all scenario.

*Betweenness centrality* for a node $v$ is calculated as $C_B(v) = \Sigma_{s\neq v\neq t\in V}\frac{\sigma_{st}(v)}{\sigma_{st}}$, where $\sigma_{st}$ is the number of shortest paths from $s$ to $t$, $\sigma_{st}(v)$ is the number of shortest paths from $s$ to $t$ pass through node $v$. Betweenness centrality is to measure one node undertaking a mediation role in a network. If one node locates in the only way that others nodes have to go through, such as communication, then this node should be important and have a high betweenness centrality.

*Closeness centrality* of vertex $v$ is calculated based on shortest paths as $C_c(v) = \frac{n-1}{\Sigma_j d(i,j))}$ where $i \neq j$, $n$ is the number of nodes, $d(i,j)$ is the length of the shortest path from vertex $i$ to vertices $j$ in the network. Closeness centrality measures how short the shortest paths are from node $i$ to all nodes. Closeness centrality is a useful measure that estimates how fast the flow of information would be through a given node to other nodes.

For the readers' convenience note that one could distinguish between centralities as degree centrality to measure activity of transferring and communication, betweenness centrality to calculate mediation or control of interest, and closeness centrality to estimate the level of efficiency and convenience, [9].

*Connected Components.* In an undirected network, all pairwise connected vertices form a connected component. The number of connected components represents the connectivity of the network, and the fewer connected components, the more strongly connected the network. The number of vertices in each component is called the size of the component. A separate network can be created and topology analyzed by selecting a large connected component of the network.

*Clustering* methods are commonly used in biological network analysis to group network elements based on certain metrics, [19]. A variety of algorithms are used for cluster analysis, all based on the similarity measure of elements. For example, the similarity of two points in n-dimensional space is measured by the Euclidean distance between them:

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}.$$

For the reader's reference, we mention here a few common clustering approaches used in systems medicine modelling: *Hierarchical clustering, centroid linkage, k-Means clustering, MCL clustering, spectral clustering*, etc., [13].

**TABLE I**
Echinococcosis genes

| No. | Gene Symbol | Description | Category | GC Id | Scores |
|---|---|---|---|---|---|
| 1 | HLA-DRB1 | Major Histocompatibility Complex, Class II, DR Beta 1 | Protein Coding | GC06M032578 | 11.58690453 |
| 2 | MT-CO1 | Mitochondrially Encoded Cytochrome C Oxidase I | Protein Coding | GCMTP005906 | 10.81720829 |
| 3 | TAP2 | Transporter 2, ATP Binding Cassette Subfamily B Member | Protein Coding | GC06M032821 | 10.69569111 |
| 4 | TLR2 | Toll Like Receptor 2 | Protein Coding | GC04P153684 | 9.982717514 |
| 5 | TLR4 | Toll Like Receptor 4 | Protein Coding | GC09P117704 | 9.931704521 |
| 6 | HLA-B | Major Histocompatibility Complex, Class I, B | Protein Coding | GC06M061261 | 9.87610817 |
| 7 | HLA-DQB1 | Major Histocompatibility Complex, Class II, DQ Beta 1 | Protein Coding | GC06M061358 | 9.757335663 |
| 8 | IL6 | Interleukin 6 | Protein Coding | GC07P022725 | 8.776535988 |
| 9 | MT-ND1 | Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 1 | Protein Coding | GCMTP003309 | 8.54378891 |
| 10 | IFNG | Interferon Gamma | Protein Coding | GC12M068154 | 8.49133873 |
| 11 | IL10 | Interleukin 10 | Protein Coding | GC01M206767 | 8.444591522 |
| – | – | – | – | – | – |
| 81 | INS | Insulin | Protein Coding | GC11M002159 | 0.269246519 |
| 82 | MB | Myoglobin | Protein Coding | GC22M035606 | 0.190386042 |

## 2.2. A parasitic disease and drug combination

There are two forms of echinococcosis in humans: cystic echinococcosis (CE; aka hydatidosis) and alveolar echinococcosis (AE). They are caused by the tapeworms *Echinococcus granulosus* and *E. multilocularis*, respectively, [23]. Currently, anti-parasitic or anti-cancer drugs and compound classes are mainly used for chemotherapeutic treatment against echinococcosis. In particular, benzimidazoles (albendazole, ABZ; mebendazole, MBZ), given either alone or combined with praziquantel (PZ) show better efficacy, [10]. But studies have shown that benzimidazoles are not therapeutic enough and the chance of cure with ABZ treatment in CE cases ranged from 11.8% to 35.2% only, [14, 18]. Also, some patients may not be able to use ABZ/MBZ, and it is common for helminths to be resistant to benzimidazole drugs, [17]. Therefore, there is an important need for new

and improved drugs against echinococcosis, [38]. Recently, interactions and efficacies of some old drugs against echinococcosis have been explored but have not yet been officially recommended for treatment, [33].

Drug-drug interaction has three types of effects: doubling, taking two drugs with the same effect can intensify their side effects; antagonism, two drugs with opposite effects interact, thereby reducing the effectiveness of one or both; alteration, one drug can alter the way another drug is absorbed, distributed, metabolized, or excreted in the body. For drug-disease interaction, sometimes a drug that is good for one disease can have a negative effect on another disease, [7].

Drug repurposing of approved drugs provides an effective method for rapid identification of new therapeutic agents to treat diseases that have drug-resistant bacteria and other emerging infectious diseases, [28]. Many active compounds identified from phenotypic screens have weak activities and cannot be directly applied in humans as a single agent, [32, 42]. Therefore, synergistic drug combination is particularly used in drug repurposing.

## 3. RESULTS

**Molecular network analysis.** The objective of this work was to predict possible drug combinations that treat the considered parasitic diseases in a direct or indirect way and reveal potential drug target molecules. To do so, basic qualitative analysis on the molecular interaction networks of disease and drug-related data have been carried out.

82 different types of genes of echinococcosis and its pathogens were extracted from a public library `genecards.org` [12] as shown in Table I. In the next step, those proteins that interact with the 82 genes were filtered from 13 different libraries as shown in Table II, and corresponding gene-protein, protein-protein interaction networks were constructed, where Cytoscape platform [26] was utilized. Attributes of each network and merged ones were presented in Table III. In order to reduce the size of merged large network with 12,173 vertices and 33,171 edges and then identify critical components, ranking of vertices and edges by their attributes, *betweenness centrality, closeness centrality, eccentricity, stress* for vertex, while *weight, confidence-score, edge-betweenness* for edge, respectively, have been performed. By eliminating the low scored vertices and edges out from the network, a molecular network composed of strongly linked key protein-protein was obtained.

**TABLE II**
Disease-protein association network

| No. | Database | # | Description | #vertex | #edges |
|---|---|---|---|---|---|
| 1 | iRefIndex | 30106 | protein-protein, imported, bipartite expansion, evidence | 6241 | 30106 |
| 2 | mentha | 7325 | protein-protein, imported, clustered, spoke expansion | 2 | 1 |
| 3 | InnateDB-All | 5462 | protein-protein, internally-curated, spoke expansion, | 1681 | 5462 |
| 4 | BioGrid | 5145 | protein-protein, internally-curated, rapid curation, spoke | 3257 | 5145 |
| 5 | tfact2gene | 4249 | text-mining, internally-curated, imported, | 619 | 4249 |
| 6 | Reactome-FIs | 3924 | protein-protein, predicted, imported, clustered | 23 | 33 |
| 7 | UniProt | 2776 | protein-protein, nucleicacid-protein, smallmolecule-protein, | 4 | 5 |
| 8 | InnateDB | 1590 | protein-protein, internally-curated, spoke expansion, | 593 | 1590 |
| 9 | EBI-GOA-nonIntAct | 908 | protein-protein, nucleicacid-protein, rapid curation, | 504 | 908 |
| 10 | MatrixDB | 649 | protein-protein, smallmolecule-protein, internally-curated, | 183 | 536 |
| 11 | bhf-ucl | 536 | protein-protein, smallmolecule-protein, nucleicacid-protein | 183 | 536 |
| 12 | BAR | 99 | protein-protein, imported, spoke expansion, predicted | 90 | 99 |
| 13 | MPIDB | 1 | protein-protein, internally-curated, imex curation, spoke | 2 | 1 |
| | | | Merged network | 12,173 | 33,171 |

The heat map in Figure 1.A shows the node attributes of the gene protein association network as an example.

We analyzed topological and functional properties of the network utilizing PCA (principle component analysis) [16] and clustering methods of `jActiveModules app`, then

**TABLE III**
Network statistics

| No. | Database | Vertex | Edge | Avg neighbors | Network diameter | Radius | Network density | Heterogeneity | Centrality | Connected components |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | iRefIndex | 6241 | 30106 | 2.779 | 13 | 7 | 4.344 | 6.926 | 0.133 | 31 |
| 2 | Mentha | | | | | | | | | |
| 3 | InnateDB-All | 1681 | 5462 | 2.924 | 10 | 5 | 0.002 | 3.546 | 0.131 | 8 |
| 4 | BioGrid | 3257 | 5145 | 2.461 | 14 | 7 | 0.001 | 5.735 | 0.166 | 26 |
| 5 | tfact2gene | 619 | 4249 | 4.305 | 7 | 4 | 0.007 | 3.205 | 0.342 | 2 |
| 6 | Reactome-FIs | 23 | 33 | 3.048 | 4 | 2 | 0.152 | 1.121 | 0.661 | 2 |
| 7 | UniProt | | | | | | | | | |
| 8 | InnateDB | 593 | 1590 | 2.709 | 13 | 7 | 0.005 | 2.432 | 0.161 | 16 |
| 9 | EBI-GOA-nonIntAct | 504 | 908 | 2.351 | 10 | 5 | 0.024 | 1.840 | 0.326 | 74 |
| 10 | MatrixDB | | | | | | | | | |
| 11 | bhf-ucl | 183 | 538 | 7.423 | 5 | 3 | 0.146 | 0.947 | 0.460 | 23 |
| 12 | BAR | 90 | 99 | 2.278 | 6 | 3 | 0.032 | 2.379 | 0.460 | 4 |

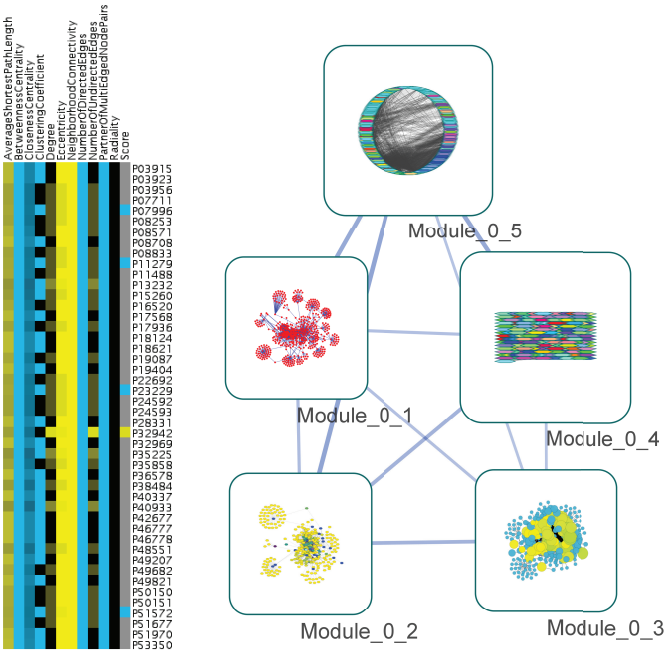



**Figure 1**. A. Protein node attributes (heat map) in the network. B. The main five modules.

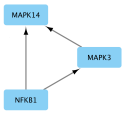
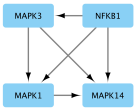
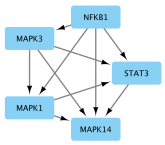
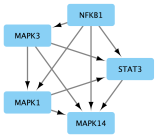
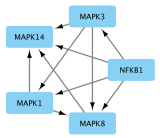
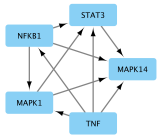
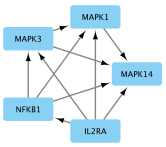
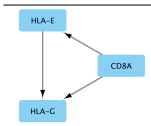
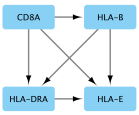
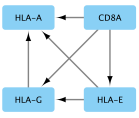
*averageShortestPathlength, betweenessCentrality, clusteringCoefficient, degree, neighborhoodConnectivity dimensions* and *similarity* were computed resulting 10 key sub-networks (motifs) were identified in the network, see Table IV. These networks were further considered as main regulatory units of the network. Moreover, as shown in Table V and Figure 1.B, the entire network was divided into 5 sub-modules linked to each other through the principle components (proteins) of each module.

**Drug-disease interaction network.** WHO (World Health Organization) suggests two drugs, namely Albendazole and Mebendazole, for the treatment of echinococcosis, [10,27]. The target molecules, other interacting drugs, transporters, and enzymes of mentioned drugs listed in Table VI were obtained from DrugBank [1]. Using the online search engine STITCH [4], interaction network of each Albendazole and Mebendazole with some other proteins and drug compounds are constructed as depicted in Figure 2.

**Heterogeneous network for drug-disease interactions.** Looking for potential drugs and their combinations having anti-echinococcosis effects, 14 candidate drugs were found

**TABLE IV**

10 motifs (subnet) of the gene-protein association network

| SubNetworks | Molecules | Av.Shortest PathLength | Betweenness Centrality | Clustering Coefficient | Degree | Neighborh. Connectivity |
|---|---|---|---|---|---|---|
|  | MAPK3 | 2.55 | 0.01 | 0.46 | 16.00 | 16.75 |
| | NFKB1 | 2.24 | 0.04 | 0.27 | 26.00 | 17.82 |
| | MAPK14 | 2.33 | 0.02 | 0.34 | 23.00 | 17.43 |
|  | MAPK1 | 2.50 | 0.01 | 0.39 | 19.00 | 15.63 |
| | MAPK14 | 2.33 | 0.02 | 0.34 | 23.00 | 17.43 |
| | NFKB1 | 2.24 | 0.04 | 0.27 | 26.00 | 17.82 |
| | MAPK3 | 2.55 | 0.01 | 0.46 | 16.00 | 16.75 |
|  | STAT3 | 2.48 | 0.02 | 0.31 | 25.00 | 14.44 |
| | MAPK3 | 2.55 | 0.01 | 0.46 | 16.00 | 16.75 |
| | MAPK14 | 2.33 | 0.02 | 0.34 | 23.00 | 17.43 |
| | NFKB1 | 2.24 | 0.04 | 0.27 | 26.00 | 17.82 |
| | MAPK1 | 2.50 | 0.01 | 0.39 | 19.00 | 15.63 |
|  | MAPK1 | 2.50 | 0.01 | 0.39 | 19.00 | 15.63 |
| | MAPK3 | 2.55 | 0.01 | 0.46 | 16.00 | 16.75 |
| | MAPK14 | 2.33 | 0.02 | 0.34 | 23.00 | 17.43 |
| | NFKB1 | 2.24 | 0.04 | 0.27 | 26.00 | 17.82 |
| | STAT3 | 2.48 | 0.02 | 0.31 | 25.00 | 14.44 |
|  | MAPK1 | 2.50 | 0.01 | 0.39 | 19.00 | 15.63 |
| | MAPK3 | 2.55 | 0.01 | 0.46 | 16.00 | 16.75 |
| | MAPK14 | 2.33 | 0.02 | 0.34 | 23.00 | 17.43 |
| | NFKB1 | 2.24 | 0.04 | 0.27 | 26.00 | 17.82 |
| | MAPK8 | 2.24 | 0.02 | 0.30 | 24.00 | 34.92 |
|  | STAT3 | 2.48 | 0.02 | 0.31 | 25.00 | 14.44 |
| | MAPK1 | 2.50 | 0.01 | 0.39 | 19.00 | 15.63 |
| | MAPK14 | 2.33 | 0.02 | 0.34 | 23.00 | 17.43 |
| | NFKB1 | 2.24 | 0.04 | 0.27 | 26.00 | 17.82 |
| | TNF | 2.70 | 0.00 | 0.49 | 10.00 | 16.50 |
|  | MAPK1 | 2.50 | 0.01 | 0.39 | 19.00 | 15.63 |
| | MAPK3 | 2.55 | 0.01 | 0.46 | 16.00 | 16.75 |
| | MAPK14 | 2.33 | 0.02 | 0.34 | 23.00 | 17.43 |
| | NFKB1 | 2.24 | 0.04 | 0.27 | 26.00 | 17.82 |
| | IL2RA | 2.24 | 0.01 | 0.31 | 26.00 | 33.69 |
|  | HLA-E | 2.04 | 0.02 | 0.33 | 57.00 | 30.96 |
| | CD8A | 1.84 | 0.07 | 0.25 | 83.00 | 26.63 |
| | HLA-G | 2.20 | 0.02 | 0.35 | 51.00 | 30.27 |
|  | HLA-B | 2.14 | 0.01 | 0.30 | 37.00 | 37.38 |
| | HLA-E | 2.04 | 0.02 | 0.33 | 57.00 | 30.96 |
| | CD8A | 1.84 | 0.07 | 0.25 | 83.00 | 26.63 |
| | HLA-DRA | 1.69 | 0.16 | 0.17 | 74.00 | 25.69 |
|  | HLA-A | 2.22 | 0.01 | 0.42 | 47.00 | 32.68 |
| | HLA-G | 2.20 | 0.02 | 0.35 | 51.00 | 30.27 |
| | CD8A | 1.84 | 0.07 | 0.25 | 83.00 | 26.63 |
| | HLA-E | 2.04 | 0.02 | 0.33 | 57.00 | 30.96 |

(see Table VII) from KEGG, GeneCard, and DrugBank [1, 2, 12] databases by entering the disease name as a keyword. These 14 drugs were widely examined by using the HDR app of Cytoscape [26]. First, a drug similarity network of 7,838 drugs with 887,883 interactions and a disease similarity network of 5,080 diseases with 19,729 interactions, respectively, was built; 1,933 number of associations between the two similarity networks were detected.

## TABLE V

The modular structure of the merged network from Table II

| Module | Active path Score | #Nodes | #Edges | Module | Node | Edge | Score |
|---|---|---|---|---|---|---|---|
| Module_1 | 20.4565164 | 611 | 2323 | | 34 | 433 | 26.242 |
| Module_2 | 15.1492831 | 342 | 2656 | | 49 | 210 | 8.750 |
| Module_3 | 13.8364176 | 339 | 2744 | Module_1 | 21 | 60 | 5.900 |
| Module_4 | 9.50119515 | 211 | 2225 | | 13 | 25 | 4.167 |
| Module_5 | 9.39945342 | 209 | 1659 | | 6 | 9 | 3.600 |

## TABLE VI

Targets of each drug ALBENDAZOLE and MEBENDAZOLE

| No. | Drug | Targets | Enzymes | #Interact. drugs | Transporter |
|---|---|---|---|---|---|
| 1 | ALBENDAZOL, DB00518 | NTubulin alpha-1A chain, ATubulin beta-2 chain, NTubulin beta-4B chain, UFumarate reductase flavoprotein subunit | Cytochrome P450 1A1, CYP1A1 Cytochrome P450 1A2, CYP1A2 Cytochrome P450 3A4, CYP3A4 Cytochrome P450 2C19, CYP2C19 | 581/719 | P-glycoprotein 1, ABCB1 |
| 2 | MEBENDAZOLE, DB00643 | Tubulin alpha-1A chain, TUBA1A | Cytochrome P450 1A1, CYP1A1 | Cimetidine | |

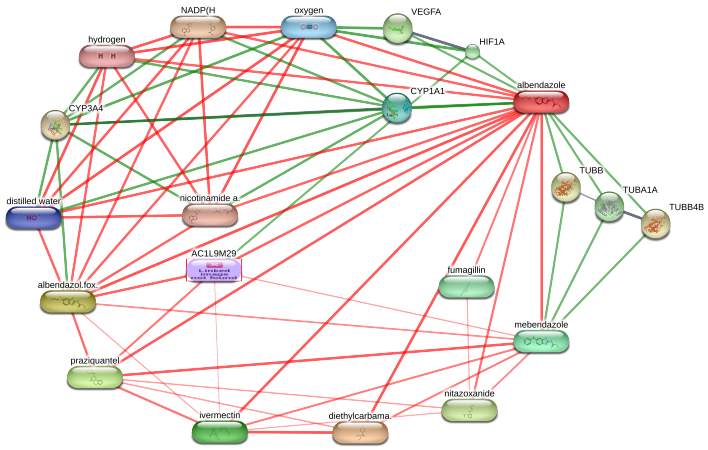



**Figure 2**. Drug(Albendazole, Mebendazole)-protein interaction network. Stronger associations are represented by thicker lines. Protein-protein interactions are shown in grey, chemical-protein interactions in green and interactions between chemicals in red. Chemical-chemical links are used to extend the network. Small nodes: protein of unknown 3D structure, large protein nodes: some 3D structure is known or predicted. (STITCH [4]).

In the second step, the 14 drugs with its associated diseases were extracted from the drug-disease interaction network constructed in the previous step, Table VII.

Third, we checked new diseases that are potentially treated and/or affected by those 14 drugs among the disease similarity network. As a result, 3,222 candidate diseases out of 5,080 were predicted, 7 of them had direct links to the 14 drugs, see Table VIII. Since the drug-disease interaction network could be massive, for a simplicity purpose, the 3,222 diseases were ranked by RWRH algorithm [21] with parameters of *Back probability 0.5, Jumping probability 0.6, and Sub-Data (drug&disease) importance weight 0.7*. Then, the first 60 candidate diseases with higher ranks were selected for the further study. As an example, six diseases with ranking scores are listed in Table IX. A drug-disease interaction network composed of 80 vertices and 292 edges was constructed from the selected 60 diseases.

The heterogeneous network depicted in Figure 3 contains vertices representing 1 protein complex, 14 drugs, 7 known diseases, 53 unknown candidate diseases, 5 pathways, and 5 types of edges representing disease-disease, disease-pathway, disease-protein complex, drug-disease, drug-pathway, and drug-protein functional interactions. The thickness of the edges represents the strength (weight) of the link or interaction in the network. From

## TABLE VII

14 Anti-echinococcosis drug information

| DrugBank ID | Drug name | Target genes |
|---|---|---|
| D00134 | Albendazole (JAN/USP/INN), Albenza (TN) | |
| D00252 | Carbamazepine (JP17/USP/INN), Equetro (TN), Tegretol (TN) | 11280, 6323, 6326, 6328, 6329, 6331, 6334, 6335, 6336 |
| D00368 | Mebendazole (JAN/USP/INN), Vermox (TN) | |
| D00427 | Norvir (TN), Ritonavir (JAN/USAN/INN) | |
| D00471 | Biltricide (TN), Praziquantel (JAN/USP/INN) | |
| D00506 | Luminal (TN), Phenobarbital (JP17/USP/INN) | 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 55879 |
| D00512 | Dilantin (TN), Phenytoin (JP17/USP/INN) | 11280, 6323, 6326, 6328, 6329, 6331, 6334, 6335, 6336 |
| D00701 | Luminal sodium (TN), Phenobarbital sodium (JAN/USP/INN) | 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 55879 |
| D02103 | Aleviatin (TN), Dilantin (TN), Phenytoin sodium (USP), Phenytoin sodium for injection (JP17) | 11280, 6323, 6326, 6328, 6329, 6331, 6334, 6335, 6336 |
| D05017 | Metronidazole phosphate (USAN) | |
| D07106 | Albendazole oxide (INN) | |
| D07595 | Fosphenytoin sodium hydrate (JAN), Fostoin (TN) | 11280, 6323, 6326, 6328, 6329, 6331, 6334, 6335, 6336 |
| D08356 | Gratusminal (TN), Phenobarbital diethylamine | 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 55879 |
| D10005 | Ensulizole (USP/INN), Phenylbenzimidazole sulfonic acid | |

## TABLE VIII

7 known diseases which have direct links with the 14 drugs in Table VII.

| No. | Disease ID | Name | Related genes |
|---|---|---|---|
| 1 | MIM121200 | SEIZURES, BENIGN FAMILIAL NEONATAL, 1; BFNS1 | 3785 |
| 2 | MIM121201 | SEIZURES, BENIGN FAMILIAL NEONATAL, 2; BFNS2 | 3786 |
| 3 | MIM143500 | GILBERT SYNDROME | 54658 |
| 4 | MIM181500 | SCHIZOPHRENIA; SCZD | 1116, 1312, 1610, 1814, 207, 23780, 267012, 27184, 27185, 3356, 4524, 55366, 65078, 6854, 80832, 84062 |
| 5 | MIM208085 | ARTHROGRYPOSIS, RENAL DYSFUNCTION, AND CHOLESTASIS 1; ARCS1 | 26276 |
| 6 | MIM218800 | CRIGLER-NAJJAR SYNDROME, TYPE I | 54658 |
| 7 | MIM237500 | DUBIN-JOHNSON SYNDROME; DJS | 1244 |

## TABLE IX

Candidate diseases that could have (in)direct links with the 14 drugs listed in Table VII.

| Rank | OMIM ID | Title | Score | Type |
|---|---|---|---|---|
| 8 | MIM613404 | ARTHROGRYPOSIS, RENAL DYSFUNCTION, AND CHOLESTASIS 2; ARCS2 | 0.00996769 | Disease |
| 9 | MIM602079 | TRIMETHYLAMINURIA; TMAU | 0.00442980 | Disease |
| 10 | MIM224100 | ANEMIA, DYSERYTHROPOIETIC CONGENITAL, TYPE II; CDAN2 | 0.00321845 | Disease |
| 11 | MIM606438 | HUNTINGTON DISEASE-LIKE 2; HDL2 | 0.00236688 | Disease |
| – | – | – | – | – |
| 59 | MIM613720 | EPILEPTIC ENCEPHALOPATHY, EARLY INFANTILE, 7; EIEE7 | 0.00082928 | Disease |
| 60 | MIM606437 | MOVED TO 121200 | 0.00082928 | Disease |

the network, one can see that indirect links exist between drugs D08356, D00701, and D00506 and diseases MIM611277, MIM605375, and MIM600513 through the pathways and the protein complex. Meanwhile drugs D00506 and D00252 were linked by unknown drug-drug network based on their interaction to the same diseases. The rest of the drugs have no evidence linked to unknown or predicted drugs.

A first sight, the prediction raised was that the diseases MIM605375 (epilepsy, nocturnal frontal lobe, type 3), MIM611277 (generalized epilepsy with febrile seizures plus, type 3; gefsp3), MIM600513 (epilepsy, nocturnal frontal lobe, type 1) might get treated by only or combined drugs of Luminal sodium (TN), Phenobarbital sodium (JAN/USP/INN) (D00701), Luminal (TN), Phenobarbital (JP17/USP/INN) (D00506), Gratusminal (TN), Phenobarbital diethylamine(D08356), Carbamazepine (JP17/USP/INN), Equetro (TN), Tegretol (TN) (D00252). Since these mentioned diseases were obtained from the similarity network of echinococcosis (considered above), the effective drugs just revealed could be effective anti-echinococcosis as well, which has to be investigated for further study.

The drug-disease interaction network depicted in the Figure 3 was divided into 2 main

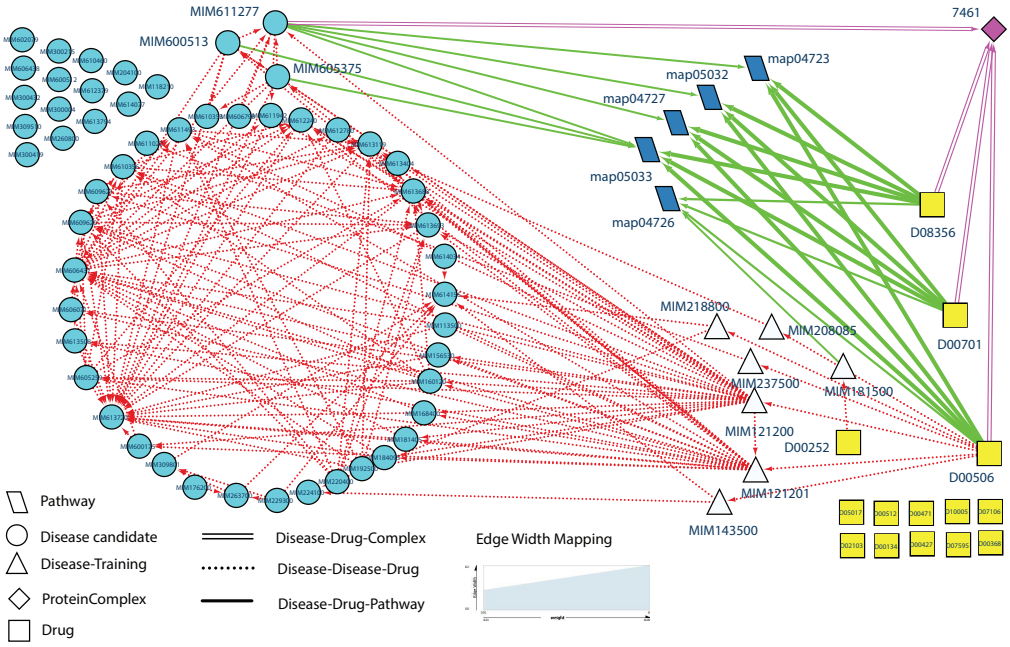



**Figure 3**. Heterogenous network of drug-disease interactions through pathways *Nicotine addiction/map05033, Morphine addiction/map05032, GABAergic synapse/map04727, Serotonergic synapse/map04726, Retrograde endocannabinoid signaling/map04723* and a protein complex *GABA-A receptor (GABRA1, GABRB2, GABRG2)*.

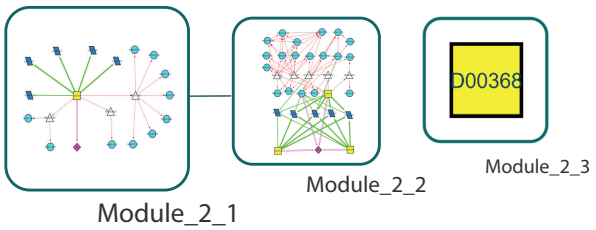



**Figure 4**. Drug-Disease interaction network (Fig. 3) is divided into 3 clusters where `Module_2_1` and `Module_2_2` are linked while `Module_2_3` is isolated.

connected clusters and an isolated one based on its active components by jActiveModules app, shown in Figure 4. As predicted above, drug D00506 is the main regulator of the first module while the next sub-network is regulated by the drugs D00701, D00506, and D08356.

## 4. CONCLUSION

Computer modelling and mathematical data analysis for a parasitic disease and possible therapeutic drugs were explored in this study. The dataset of considering disease and relevant drugs used in the investigation were mined from the public databases, i.e., NCBI, KEGG, DrugBank, PuMed, and more. By using graph and network algorithms, and automated systems (Cytoscape, STITCH, GeneCard, etc.) by utilizing obtained dataset, biological network construction of disease and drug interactions, main molecular regulations, structural, functional, and statistical analysis have been studied.

In the study of drug-disease interactions, 14 drugs against echinococcosis were selected, 7 diseases related to their direct reaction and 53 diseases that could be indirectly affected by these drugs were identified. Out of these 53 diseases, 3 diseases namely MIM611277, MIM605375, and MIM600513 (see the disease names and descriptions in Table VII and IX)

were discovered having strong interactions with the 3 candidate drugs of D08356, D00701, and D00506, or their combinations through a pathway map05033 (nicotine addiction) while these three drugs have direct interactions with a protein complex 7461 (GABA-A receptor), to which the disease MIM611277 is connected.

The drugs D08356, D00701, and D00506 have similar pharmacological effects for common targets, such as being used as neuropsychiatric agent and metabolizing enzyme inducer while also affecting the nervous system. The above-mentioned drugs also affect the sensory organs as anticonvulsant, sedative-hypnotic and anti-anxiety, [1]. Combinations of them with nicotine and morphine could be used in the same manner as pain killers.

Thus, drug-centric study against echinococcosis suggests 3 similar diseases and 3 drugs have a strong relation that could be explored for the further study to reveal the exact mechanisms if they would have anti-echinococcosis effects and whether they could be used in drug combination design. However, in this study, we have not presented any specific agents for anti-echinococcosis.

We have developed a step-by-step methodology for a) mining disease and drug-related integrated data from different public sources, b) construct heterogenous interaction networks, c) conduct analysis on the networks by grouping their structural and functional attributes, forming connected clusters of sub-networks, and detect the main controlling modules in each sub-network. For example, protein-protein interaction network of echinococcosis was divided into 10 regulator sub-networks and grouped into 5 main modules, which allows analyzis simpler for the entire network.

For the next study, network based modelling and machine learning based data analyzing methods can be combined to predict and discover considering disease development molecular mechanism, therapeutic drug compound structure, and target identification.

## REFERENCES


[1] DrugBank | Powering health insights with structured drug data.

[2] KEGG: Kyoto Encyclopedia of Genes and Genomes.

[3] NCBI - National Library of Medicine.

[4] STITCH: chemical association networks.

[5] Eric Alm and Adam P Arkin. Biological networks. *Current opinion in structural biology*, 13(2):193–202, 2003.

[6] Gary D Bader and Christopher WV Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*, 4(1):1–27, 2003.

[7] Irene V. Bijnsdorp, Elisa Giovannetti, and Godefridus J. Peters. *Analysis of Drug Interactions*, pages 421–434. Humana Press, Totowa, NJ, 2011.

[8] Aurora S Blucher, Shannon K McWeeney, Lincoln Stein, and Guanming Wu. Visualization of drug target interactions in the contexts of pathways and networks with reactomefiviz. *F1000Research*, 8, 2019.

[9] Abdul Talib Bon. Determining the size of centrality in social networks.

[10] Enrico Brunetti, Peter Kern, and Dominique Angèle Vuitton. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta Tropica*, 114(1):1–16, 2010.

[11] Feixiong Cheng, István A Kovács, and Albert-László Barabási. Network-based prediction of drug combinations. *Nature communications*, 10(1):1–11, 2019.

[12] GeneCards Human Gene Database. GeneCards - Human Genes | Gene Database | Gene Search.

[13] E. Diday and J. C. Simon. *Clustering Analysis*, pages 47–94. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976.

[14] Joseph El-On. Benzimidazole treatment of cystic echinococcosis. *Acta tropica*, 85(2):243–252, 2003.

[15] Gihanna Galindez, Julian Matschinske, Tim Daniel Rose, Sepideh Sadegh, Marisol Salgado-Albarrán, Julian Späth, Jan Baumbach, and Josch Konstantin Pauling. Lessons from the covid-19 pandemic for advancing computational drug repurposing strategies. *Nature Computational Science*, 1(1):33–41, 2021.

[16] Detlef Groth, Stefanie Hartmann, Sebastian Klie, and Joachim Selbig. *Principal Components Analysis*, pages 527–547. Humana Press, Totowa, NJ, 2013.

[17] Beate Grüner, Petra Kern, Benjamin Mayer, Tilmann Gräter, Andreas Hillenbrand, Thomas EF Barth, Rainer Muche, Doris Henne-Bruns, Wolfgang Kratzer, and Peter Kern. Comprehensive diagnosis and treatment of alveolar echinococcosis: A single-center, long-term observational study of 312 patients in germany. *GMS infectious diseases*, 5, 2017.

[18] Andrew Hemphill, Britta Stadelmann, Reto Rufener, Markus Spiliotis, Ghalia Boubaker, Joachim Müller, Norbert Müller, Daniela Gorgas, and Bruno Gottstein. Treatment of echinococcosis: albendazole and mebendazole–what else? *Parasite*, 21, 2014.

[19] Namgil Lee, Hojin Yoo, and Heejung Yang. Cluster analysis of medicinal plants and targets based on multipartite network. *Biomolecules*, 11(4), 2021.

[20] Elaine L. Leung, Zhi-Wei Cao, Zhi-Hong Jiang, Hua Zhou, and Liang Liu. Network-based drug discovery by integrating systems biology and computational technologies. *Briefings in Bioinformatics*, 14(4):491–505, 08 2012.

[21] Yongjin Li and Jagdish C Patra. Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network. *Bioinformatics*, 26(9):1219–1224, 2010.

[22] Stephani Joy Y Macalino, Vijayakumar Gosu, Sunhye Hong, and Sun Choi. Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research*, 38(9):1686–1701, 2015.

[23] Donald P McManus, Wenbao Zhang, Jun Li, and Paul B Bartley. Echinococcosis. *The Lancet*, 362(9392):1295–1304, 2003.

[24] Pablo Porras Millán. Network analysis of protein interaction data: an introduction. 2017.

[25] Jason Montojo, Khalid Zuberi, Harold Rodriguez, Farzana Kazi, George Wright, Sylva L Donaldson, Quaid Morris, and Gary D Bader. Genemania cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics*, 26(22):2927–2928, 2010.

[26] Keiichiro Ono. Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization.

[27] World Health Organization. Echinococcosis. https://www.who.int/en/news-room/fact-sheets/detail/echinococcosis.

[28] Kyungsoo Park. A review of computational drug repurposing. *Translational and Clinical Pharmacology*, 27(2):59–63, 2019.

[29] Georgios A Pavlopoulos, Maria Secrier, Charalampos N Moschopoulos, Theodoros G Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider, and Pantelis G Bagos. Using graph theory to analyze biological networks. *BioData mining*, 4(1):1–27, 2011.

[30] Victor-Bogdan Popescu, Krishna Kanhaiya, Dumitru Iulian Năstac, Eugen Czeizler, and Ion Petre. Network controllability solutions for computational drug repurposing using genetic algorithms. *Scientific reports*, 12(1):1–16, 2022.

[31] Victor-Bogdan Popescu, José Ángel Sánchez-Martín, Daniela Schacherer, Sadra Safadoust, Negin Majidi, Andrei Andronescu, Alexandru Nedea, Diana Ion, Eduard Mititelu, Eugen Czeizler, et al. Netcontrol4biomed: a web-based platform for controllability analysis of protein–protein interaction networks. *Bioinformatics*, 37(21):3976–3978, 2021.

[32] Sudeep Pushpakom, Francesco Iorio, Patrick A Eyers, K Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, et al. Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, 18(1):41–58, 2019.

[33] Reto Rufener, Luca Dick, Laura D'Ascoli, Dominic Ritler, Amani Hizem, Timothy NC Wells, Andrew Hemphill, and Britta Lundström-Stadelmann. Repurposing of an old drug: in vitro and in vivo efficacies of buparvaquone against echinococcus multilocularis. *International Journal for Parasitology: Drugs and Drug Resistance*, 8(3):440–450, 2018.

[34] Philippe Sanseau and Jacob Koehler. Editorial: Computational methods for drug repurposing. *Briefings in Bioinformatics*, 12(4):301–302, 07 2011.

[35] Jack W Scannell, Alex Blanckley, Helen Boldon, and Brian Warrington. Diagnosing the decline in pharmaceutical r&d efficiency. *Nature reviews Drug discovery*, 11(3):191–200, 2012.

[36] Gisbert Schneider and Uli Fechner. Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4(8):649–663, 2005.

[37] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.

[38] Mar Siles-Lucas, Adriano Casulli, Roberto Cirilli, and David Carmena. Progress in the pharmacological treatment of human cystic and alveolar echinococcosis: compounds and therapeutic targets. *PLoS neglected tropical diseases*, 12(4):e0006422, 2018.

[39] Nicoleta Siminea, Victor Popescu, Jose Angel Sanchez Martin, Daniela Florea, Georgiana Gavril, Ana-Maria Gheorghe, Corina Iţcuş, Krishna Kanhaiya, Octavian Pacioglu, Laura Ioana Popa, et al. Network analytics for drug repurposing in covid-19. *Briefings in bioinformatics*, 23(1):bbab490, 2022.

[40] Jianlin Wang, Wenxiu Wang, Chaokun Yan, Junwei Luo, and Ge Zhang. Predicting drug-disease association based on ensemble strategy. *Frontiers in Genetics*, 12:666575, 2021.

[41] Zikai Wu, Yong Wang, and Luonan Chen. Network-based drug repositioning. *Molecular BioSystems*, 9(6):1268–1281, 2013.

[42] Wei Zheng, Wei Sun, and Anton Simeonov. Drug repurposing screens and synergistic drug-combinations for infectious diseases. *British journal of pharmacology*, 175(2):181–191, 2018.

[43] Yadi Zhou, Yuan Hou, Jiayu Shen, Yin Huang, William Martin, and Feixiong Cheng. Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2. *Cell discovery*, 6(1):1–18, 2020.

## ACKNOWLEDGEMENTS

This work was supported by the Mongolian Foundation of Science and Technology, Research project SHUT-2021/350. We appreciate the two reviewers for their careful work and fruitful comments which help to improve the presentation of the paper. We also thank Mr Ider Saranchimeg, from the Mongolian University of Pharmaceutical Science, for his productive discussion and comments.

## BIOGRAPHIES

**Altankhuyag Avirmed** is doctoral candidate at the department of CS, the School of Information and Communication Technology, MUST. He received his bachelor's (2000) and master's (2018) degrees in computer science from MUST (2013). His research interests are computer modelling, simulation, and analysis of bio-molecular processes.

**Uranchimeg Erdenedalai** is a master's degree student at the department of IT of the School of Information and Communication Technology, MUST. She received her BS degree in Information systems (2013) from MUST. Her research interests are business intelligent systems and computational systems biology.

**Selenge Erdenechimeg** attained her Ph.D. degree in pharmaceutical sciences from Tohoku Pharmaceutical University, Japan (2014). Her research interests are in drug development, biological activities of medicinal plants.

**Yansen Su** attaind her Ph.D. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2014. She is currently an associate professor in the School of Artificial Intelligence, Anhui University, Hefei, China. Her main research interests include bioinformatics, systems biology, complex networks, and multi-objective optimization.

**Tseren-Onolt Ishdorj** did his Ph.D. degree in Computer Science (2007) with the University of Sevilla, Spain. He is currently a full professor at the Mongolian University of Science and Technology. His research areas are computational biology, microfluidic systems, theoretical computer science, and biocomputing.