PhIN: A Protein Pharmacology Interaction Network Database

Z Wang, J Li, R Dang, L Liang and J Lin

Due to the intrinsic complexity of molecular interactions, besides target proteins many drugs interact with several additional targets, which may lead to unwanted side effects. The traditional drug design paradigm, a one-target–one-drug mindset, which has been proven successfully in drug development for decades, is now challenged by the clinical attrition figures and several studies indicating polypharmacology (which is that drugs interact with multiple targets). Especially, single-target drug intervention has been shown not effective in combating the complex systemic diseases like cancers, AIDS, cardiovascular diseases, and neurodegenerative disorders. As a consequence, network analysis, due to its capability of investigating complex relationships, is more and more attractive and network-based approaches such as network pharmacology, network medicine, diseases network, etc., are emerging. Compared with the traditional approach to drug discovery, systems-oriented computational approaches apply and leverage the parallelism and high-dimensionality of the existing molecular data to construct molecular models to model broader bimolecular systems. Recently, various databases concerning drug–protein, molecule–protein, and protein–protein interactions were developed, e.g., VisANT, a network platform integrating genes, drugs, diseases, and therapies; ChemProt, a disease chemical biology database; DINIES, a web interface for drug–target interaction prediction; and VNP, a database used for visualizing the disease–target–drug interaction network.

Recently, Paolini et al. mapped the human pharmacological interaction network using the definition that two proteins interact if they bind at least 10% of shared compounds with a difference in potency of only 10-fold below an activity cutoff of 10 μM, and applied it in network analysis of drug–target interactions associated with asthma. Using a similar idea, Hu et al. performed a systematic selectivity-centric analysis of target–ligand interactions and identified more than 200 molecular scaffolds that are selective for established target families.

Although a large amount of ligand–target and target–target bioactivity data were provided from public data sources, including PubChem, ChEMBL and BindingDB, to our knowledge there are no databases or tools that utilize such data to provide comprehensive pharmacology interaction network analysis based on a user-defined interaction. In this article we report a protein pharmacology interaction network database (PhIN), aiming to assist multitarget drug discovery by providing comprehensive and flexible network pharmacology analysis. Using PhIN, users can obtain interacting target networks within or across human pathways, between human and virus, by defining the number of shared compounds or scaffolds under an activity cutoff. We expect PhIN to be a useful tool for multitarget drug development. PhIN is freely available at http://cadd.pharmacy.nankai.edu.cn/phlin/.

METHODS

Data sources and statistics

In consideration of bioactivity data standardization and data update cycle, we use ChEMBL as the bioactivity data source. In the current version of PhIN, ChEMBL (v. 18) was used, which contains about 1,350,000 compound entries and about 9,400 target entries, with more than 12,400,000 activity measurements, among which about 2,800,000 have valid ChEMBL-converted values (hereafter referred to as the pChEMBL value, which is a negative logarithmic of published activity; see detail description at https://www.ebi.ac.uk/chembl/faq#faq87). Only data with pChEMBL values were used in PhIN. Pathways from SMPDB and viral data from ICTV (http://www.ictvonline.org/) and ViralZone were integrated. For a certain target-compound entry, if multiple potency measurements were reported, their geometric means are averaged.
mean value was used (hereafter generally referred to as pChEMBL value). For example, for neostigmine bromide (CHEMBL54126) and its target acetylcholinesterase (CHEMBL220), there are 12 corresponding bioactivities, among which 11 have valid pChEMBL values: 7.66, 8.70, 6.22, 8.04, 7.35, 7.51, 7.39, 7.00, 10.37, and 6.70; therefore, active data used in PhIN is 7.63. Figure 1 shows the data workflow for PhIN.

**Scaffold generation**

The scaffold in this study is the “Bemis and Murcko” (BM) scaffolds generation. For example, for neostigmine bromide (CHEMBL54126) and its target acetylcholinesterase (CHEMBL220), there are 12 corresponding bioactivities, among which 11 have valid pChEMBL values: 7.66, 8.70, 6.22, 8.04, 7.35, 7.51, 7.39, 7.00, 10.37, and 6.70; therefore, active data used in PhIN is 7.63. Figure 1 shows the data workflow for PhIN.

**Web interface**

The PhIN web interface is built with the Python and Django framework (v. 1.6.2) (https://www.djangoproject.com/) and deployed using Nginx (http://nginx.org). Tools used in PhIN are: JSME, a JavaScript-based molecular editor, used for structure display; Cytoscape.js (http://cytoscape.github.io/cytoscape.js/), used for network representation; PostgreSQL (v. 9.3) together with MongoDB (http://www.mongodb.org), used as a database server (http://www.postgresql.org/); and RDKit, used for small molecule manipulation.

**RESULTS**

**Polypharmacology interactions network**

In PhIN, one of two criteria, compounds criteria or scaffolds criteria, is used to define an interacting target pair. When using compounds criteria, two targets are defined as interacting if they share a certain number of bound ligands above a certain pChEMBL value cutoff. Paolini et al. define an interaction target pair if there is at least a ligand binding to both targets under a certain activity range, while Hu et al. employ a simple definition that two targets are related to each other if they share at least five active compounds. Based on the above two studies, we set the default threshold for activity (pChEMBL value) and the number of shared ligands to 6 (1 μM) and 5, respectively. For example, in the current database there are 34 and 48 ligands with pChEMBL above 6 against sucrase-isomaltase (CHEMBL2748) and maltase-glucoamylase (CHEMBL2074), respectively, and 23 of them are in common, so these two targets are defined as interactive or related. To give the user a global view of target–target interactions in PhIN and enable the user to conduct large-scale analysis, three matrices that contain shared compounds/scaffolds of each target pair can be found in Supplementary Tables S1, S2, and S3. Overall, the current database contains 1,126,060 target–target interaction pairs in terms of
shared compounds and 3,428,020 pairs in terms of shared scaffolds.

**Pathway integration**
Pathway data were collected from SMPDB (v. 2.0)\(^3\) which contains 618 pathways of human and were classified into metabolic pathway, physiological pathway, signaling pathway, drug metabolism pathway, drug action pathway, and disease pathway; proteins that are involved in each pathway were mapped to ChEMBL targets through Uniprot ID, as in Table 1.

**Viral–viral and viral–human target–target interaction network**
Nowadays, viruses are mainly classified by phenotypic characteristics including morphology, nucleic acid type, mode of replication, host organisms, and the type of disease they cause etc. (http://en.wikipedia.org/wiki/Virus_classification). The ICTV viral system and Baltimore

### Table 1 Number of ChEMBL targets mapped to SMPDB

| Pathway class       | Number of pathways | Number of targets | Number of compounds | Number of BM scaffold |
|---------------------|--------------------|-------------------|---------------------|-----------------------|
| Disease pathway     | 215                | 294               | 25,437              | 8,785                 |
| Drug action pathway | 238                | 346               | 63,342              | 21,839                |
| Drug metabolism pathways | 62            | 162               | 25,419              | 7,901                 |
| Metabolic pathways  | 85                 | 350               | 32,498              | 11,353                |
| Physiological pathways | 6              | 105               | 18,583              | 7,328                 |
| Signaling pathways  | 15                 | 107               | 18,995              | 7,166                 |

For compounds and BM scaffolds, pChEMBL value cutoff is 6.

![PhIN: A Pharmacology Interaction Network database.](image)

**Figure 2** Search and results interface of PhIN. (a) Query options, including text areas for thresholds of shared number and pChEMBL value, toggle for network type—compound or scaffold, text area for ChEMBL ID and three tree-like menus. (b) Result network graph of excitatory neural signaling through 5-HTR 4 and serotonin; external targets of this pathway are also shown. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
classification system\textsuperscript{33} are the two main schemes and both of them were integrated into PhIN. Moreover, targets belonging to viruses also link to ViralZone,\textsuperscript{34} which was developed by the Swiss-Prot virus annotation team for providing comprehensive viral information. PhIN contains 314 viral information. Using PhIN, users can not only obtain a target–target interaction network within a virus by defining the shared ligand/scaffold number and pChEMBL value cut-off, but also the target–target interaction network between the virus and human. For example, the interaction network of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (\textsuperscript{Supplementary Figure S2}) shows that besides targets of HIV, there are also connections with cell lines such as T-cells. Therefore, we hope that PhIN could be useful in antivirus drug discovery.

\textbf{Molecule–molecule functional similarity}

Drugs with overlapping drug targets are likely to be side effect similar.\textsuperscript{35,36} Therefore, PhIN introduces molecule–molecule functional similarity by defining that functional similarity of two molecules is determined by their interactive targets in common (the default thresholds for target numbers and pChEMBL values are 3 and 6, respectively). For example, ziprasidone (CHEMBL708) and clozapine (CHEMBL42), which are FDA-approved drugs for the treatment of schizophrenia, act against 18 targets in common, so they are defined as functionally similar although structurally different. Functional similarity might also help to develop more structurally diversified leads during drug development. This information represented as a table in each compound page (\textsuperscript{Supplementary Figure S3}). Users can enter the ChEMBL ID of a compound, such as CHEMBL708, at the ChEMBL ID input area at the search panel to jump to the page that contains a table of functionally similar compounds.

\textbf{Case study}

By generating the polypharmacology interactions network of a signaling pathway, excitatory neural signaling through 5-HTR 4 and serotonin (ENST5S), we will give a step-by-step example of using PhIN. The 5-HT receptor, a member of the seven transmembrane spanning the G-protein-coupled family of receptors, is widely distributed in the central nervous system and peripheral tissues.\textsuperscript{37,38} By searching and browsing pathway trees, check the excitatory neural signaling through 5-HTR 4 and serotonin and leave the other options as their defaults, as shown in Figure 2a. Then click the submit button, and a network will be created in seconds (Figure 2b) (however, this depends on the user selection: the more targets selection, the longer the generation time). More options for manipulating the network can be found at the option tab, including node labeling, node filtering by

\textbf{Figure 3} Pharmacology interaction network of serotonin 4 (5-HT 4) receptor. Nodes is highlighted (green) by pathway (excitatory neural signaling through 5-HTR 4 and serotonin). Edge width is proportional to number of shared compounds (scaffolds). Interaction information between HERG and serotonin 4 (5-HT 4) receptor is shown in the left panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
pathway, protein type, or viral type, a toggle for showing a full network or internal network (in this example, only targets involved in ENST5S will be shown) and a button for exporting the current network as a picture. Clicking on the node and edge between nodes will display target information and target-pair information in the left panel (Figure 3). Moreover, molecular analysis inside the network, including a fingerprint-based similarity search (an ECFP-like fingerprint implemented in RDKit) and substructure search, can be performed in the “utility” panel. If one of the ligands in a ligand set of a target, similar to (the Tanimoto coefficient is employed as threshold) or contains the user-input-structure, the target will be highlighted in yellow. This will help the user to find potential targets for the user-input-structure among the constructed pharmacology interaction network. Full data underlying the network is at the data tab and can be downloaded as a PDF or CSV (comma separated value).

**DISCUSSION**

The objective of this work was to provide an informative interaction network for polypharmacology. The definition of interaction between target pairs is by the shared compounds, which was derived from previous studies. We also extended the definition of interaction between target pairs using shared scaffolds. As stated in the case study part, we gave a simple example (ENST5S) of typical use of the PhIN database. From the network, we can find that 5-HT4 is interacting with the human ether-a-go-go related gene (HERG), since the two proteins share five compounds with pChEMBL above 6 (Figure 4). Because HERG codes for a protein of the alpha subunit of a potassium ion channel, which is essential for the maintenance of normal cardiac function, inhibition of HERG is implicated in long QT syndrome, which is a potentially life-threatening arrhythmia. Since the edge connection between ENST5S and HERG indicates that active compounds against ENST5S may also bind to HERG, extra attention should be paid when selecting hit compounds for ENST5S. An inside view of shared compounds between ENST5S and HERG shows that one of the five compounds is a withdrawn drug, Cisapride, in many countries because of the side effect long QT syndrome. This example shows that PhIN might help researchers to identify a side target of drugs and avoid the side effect in drug development. Besides side effect prediction, a selective optimization of side activities analysis (SOSA) of the pharmacology network from PhIN can provide a wealth of opportunities for identifying lead series that already exhibit interesting mixtures of pharmacology, which follow the idea of initiating polypharmacology from an integrated...
PhIN can also assist in target identification and multitarget drug design. For example, from the interactive network of the serotonin transporter (5-HT) (CHEMBL228), with default search parameters, we can find that it interacts with the adrenergic receptor beta protein family (ADRB) (Supplementary Table S4). Of these shared compounds, five are approved drugs, and further, one is an approved drug (carvedilol, CHEMBL723) for ADRB, which indicates that 5-HT is a potential target for drugs of the ADRB protein family, and also proteins of the ADRB protein family are potential targets of drugs for 5-HT. A related study has shown that paxil (CHEMBL490), one of the 5-HT reuptake inhibitors, exhibited good binding affinity for ADRB. Multitarget therapeutics, including drug combinations and drugs with multiligands, are important for complex diseases such as HIV infection and cancer. The interaction network created by PhIN may assist such multitarget drug design. For example, with default search parameters, the interaction network of HIV-1 shows that HIV-1 protease and HIV-1 reverse transcriptase shared 15 compounds. Moreover, a molecular search (similarity or substructure) inside the network can help users to evaluate the structure. For example, if a structure, 2-(1H-imidazol-2-yl)-1H-indole (Supplementary Figure S4), is drawn, and a substructure search inside the network is performed, only HIV-1 reverse transcriptase and HIV-1 protease were highlighted in yellow, which indicates that this structure may be a good starting point for designing a multitarget drug against HIV-1 protease and HIV-1 reverse transcriptase.

In summary, we constructed a PhIN database by integrating bioactivity data from ChEMBL (a public database) in terms of shared compounds and scaffolds. Using shared scaffolds to identify a pharmacology interaction network is introduced. We also incorporated virus taxonomy data from ICTV and ViralZone and human pathway data from SMPDB, which could help readers to study protein pharmacology interactions for disease pathways, biological pathways, and viral infection. We hope that PhIN will be a useful source for system pharmacology researchers and multitarget drug design.

We have released the initial version of PhIN. In the future, more biological entities, such as gene ontology terms and disease, as well as more pharmacology databases such as KEGG, DrugBank, TTD, and NCI/Nature Pathway Interaction Database (http://pid.nci.nih.gov), and analytical tools, such as the Similarity Ensemble Approach (SEA) and machine learning-based predictive models, will be integrated. In addition, we will update PhIN according to the databases to keep pace with the ever-increasing publication data.

Acknowledgments. This work was supported by the National Basic Research Program (973 Program, No. 2011CBA00800 and No. 2013CB911100).
16. Chavali, A.K., D’Auria, K.M., Hewlett, E.L., Pearson, R.D. & Papin, J.A. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. Trends Microbiol. 20, 113–123 (2012).

17. Liu, T., Lin, Y., Wen, X., Jorissen, R.N. & Gilson, M.K. BindingDB: a web-accessible database. Nucleic Acids Res. 41, D464–469 (2013).

18. Bi, E., Wang, Y.-Y., Kong, D.-X., Qin, T., Li, X., Caetano-Anollés, G. & Wang, G. H. Lessons learned from molecular scaffold analysis. J. Med. Chem. 53, 752–758 (2010).

19. Kim, K., Choi, J., Lee, H., Lee, J. & Park, K. Polypharmacology directed compound data mining: identification of promiscuous chemotypes with different activity profiles and comparison to approved drugs. J. Chem. Inf. Model. 50, 2112–2118 (2010).

20. Myślinger, M.M., Chirciu, M., Irwin, J.J. & Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J. Med. Chem. 55, 6582–6594 (2012).

21. Jewison, T. et al. SMPDB 2.0: big improvements to the small molecule pathway database. Nucleic Acids Res. 42, 478–484 (2013).

22. Baltimore, D. Expression of animal virus genomes. Bacteriol. Rev. 35, 235–241 (1971).

23. Hulo, C. et al. VirusZone: a knowledge resource to understand virus diversity. Nucleic Acids Res. 39, D576–582 (2011).

24. Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L.J. & Bork, P. Drug target identification using side-effect similarity. Science 321, 263–266 (2008).

25. Brouwers, L., Iskar, M., Zeller, G., Noort, V. & Bork, P. Network neighbors of drug targets contribute to drug side-effect similarity. Proc. Natl. Acad. Sci. U. S. A. 104, 8685–8690 (2007).

26. Goh, K., Cusick, M.E., Valle, D., Childs, B., Vidal, M. & Barabasi, A.-L. The human disease network. Proc. Natl. Acad. Sci. U. S. A. 104, 8685–8690 (2007).

27. Dudley, J.T., Schadt, E., Sirota, M., Butte, A.J. & Ashley, E. Drug discovery in a multidimensional world: systems, patterns, and networks. J. Cardiovasc. Transl. Res. 3, 439–447 (2010).

28. Hu, Z. et al. VisANT 4.0: integrative network platform to connect genes, drugs, diseases and therapies. Nucleic Acids Res. 41, W225–231 (2013).

29. Hu, Y. & Bajarath, J. Polypharmacology directed compound data mining: identification of promiscuous chemotypes with different activity profiles and comparison to approved drugs. J. Chem. Inf. Model. 50, 2112–2118 (2010).

30. Hu, Y. & Bajarath, J. Polypharmacology directed compound data mining: identification of promiscuous chemotypes with different activity profiles and comparison to approved drugs. J. Chem. Inf. Model. 50, 2112–2118 (2010).

31. Myślinger, M.M., Chirciu, M., Irwin, J.J. & Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J. Med. Chem. 55, 6582–6594 (2012).

32. Jewison, T. et al. SMPDB 2.0: big improvements to the small molecule pathway database. Nucleic Acids Res. 42, 478–484 (2013).

33. Baltimore, D. Expression of animal virus genomes. Bacteriol. Rev. 35, 235–241 (1971).

34. Hulo, C. et al. VirusZone: a knowledge resource to understand virus diversity. Nucleic Acids Res. 39, D576–582 (2011).

35. Campillos, M., Kuhn, M., Gavin, A.-C., Jensen, L.J. & Bork, P. Drug target identification using side-effect similarity. Science 321, 263–266 (2008).

36. Brouwers, L., Iskar, M., Zeller, G., Noort, V. & Bork, P. Network neighbors of drug targets contribute to drug side-effect similarity. Proc. Natl. Acad. Sci. U. S. A. 104, 8685–8690 (2007).

37. Goh, K., Cusick, M.E., Valle, D., Childs, B., Vidal, M. & Barabasi, A.-L. The human disease network. Proc. Natl. Acad. Sci. U. S. A. 104, 8685–8690 (2007).

38. Dudley, J.T., Schadt, E., Sirota, M., Butte, A.J. & Ashley, E. Drug discovery in a multidimensional world: systems, patterns, and networks. J. Cardiovasc. Transl. Res. 3, 439–447 (2010).

39. Hu, Z. et al. VisANT 4.0: integrative network platform to connect genes, drugs, diseases and therapies. Nucleic Acids Res. 41, W225–231 (2013).

40. Hu, Y. & Bajarath, J. Polypharmacology directed compound data mining: identification of promiscuous chemotypes with different activity profiles and comparison to approved drugs. J. Chem. Inf. Model. 50, 2112–2118 (2010).

41. Myślinger, M.M., Chirciu, M., Irwin, J.J. & Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J. Med. Chem. 55, 6582–6594 (2012).

42. Jewison, T. et al. SMPDB 2.0: big improvements to the small molecule pathway database. Nucleic Acids Res. 42, 478–484 (2013).

43. Baltimore, D. Expression of animal virus genomes. Bacteriol. Rev. 35, 235–241 (1971).