Citation: CPT Pharmacometrics Syst. Pharmacol. (2015) 4, e25; doi:10.1002/psp4.25

ORIGINAL ARTICLE

PhIN: A Protein Pharmacology Interaction Network Database

Z Wang[1], J Li[1,2], R Dang[1], L Liang[1,2] and J Lin[1,2]*

Network pharmacology is a new and hot concept in drug discovery for its ability to investigate the complexity of polypharmacology, and becomes more and more important in drug development. Here we report a protein pharmacology interaction network database (PhIN), aiming to assist multitarget drug discovery by providing comprehensive and flexible network pharmacology analysis. Overall, PhIN contains 1,126,060 target–target interaction pairs in terms of shared compounds and 3,428,020 pairs in terms of shared scaffolds, which involve 12,419,700 activity data, 9,414 targets, 314 viral targets, 652 pathways, 1,359,400 compounds, and 309,556 scaffolds. 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/phin/.

CPT Pharmacometrics Syst. Pharmacol. (2015) **4**, e25; doi:10.1002/psp4.25; published online on 18 March 2015

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.[1–3] 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[4] and several studies indicating polypharmacology (which is that drugs interact with multiple targets).[5,6] Especially, single-target drug intervention has been shown not effective in combating the complex systemic diseases like cancers, AIDS, cardiovascular diseases, and neurodegenerative disorders.[7–10] As a consequence, network analysis, due to its capability of investigating complex relationships, is more and more attractive[8,11–15] and network-based approaches such as network pharmacology,[4] network medicine,[7] diseases network,[16] 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.[17] Recently, various databases concerning drug–protein, molecule–protein, and protein–protein interactions were developed, e.g., VisANT,[18] a network platform integrating genes, drugs, diseases, and therapies; ChemProt,[19] a disease chemical biology database; DINIES,[20] a web interface for drug–target interaction prediction; and VNP,[21] a database used for visualizing the disease–target–drug interaction network.

Recently, Paolini *et al.*[6] 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.*[22] 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,[23] ChEMBL,[24] and BindingDB,[25] 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, and virus and virus by defining a number of shared compounds or scaffolds under an activity cutoff. We expect PhIN to be a useful tool and data source for multitarget drug development.

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#faq67). 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

[1]State Key Laboratory of Medicinal Chemical Biology and College of Pharmacy, Nankai University, Tianjin, China; [2]High-Throughput Molecular Drug Discovery Center, Tianjin Joint Academy of Biomedicine and Technology, Tianjin, China. *Correspondence: J Lin (jianpinglin@nankai.edu.cn)
Received 2 December 2014; accepted 11 February 2015; published online on 18 March 2015. doi:10.1002/psp4.25

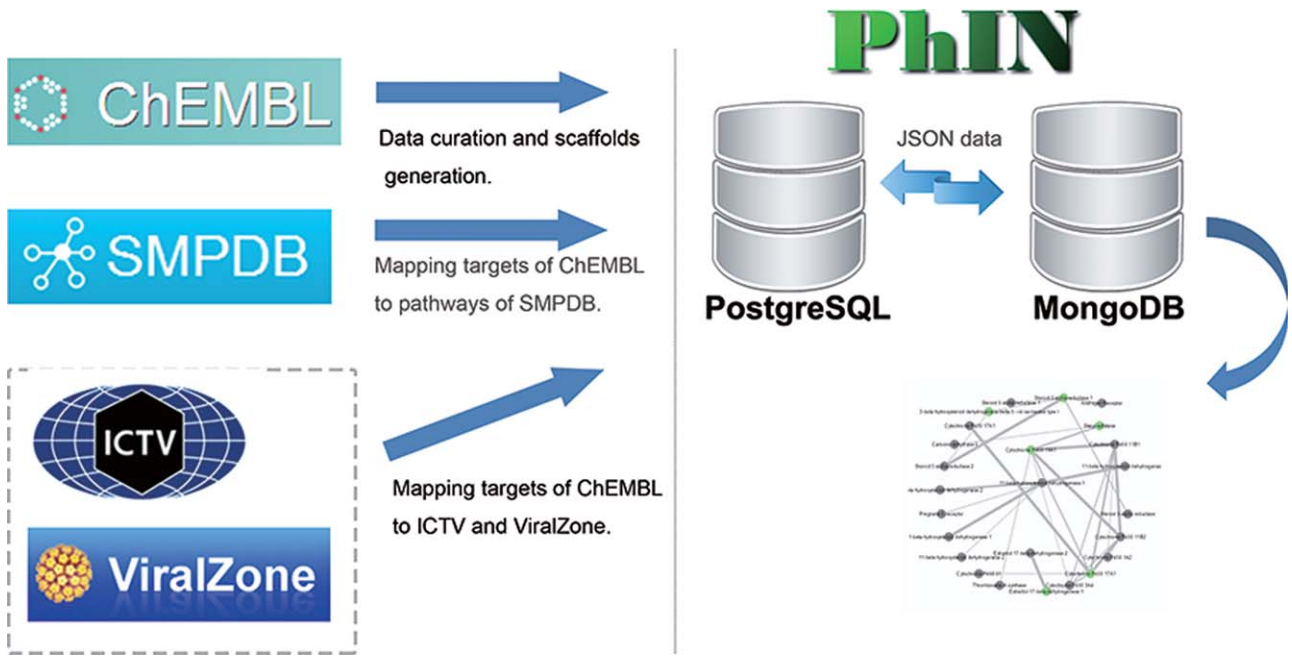


Figure 1 Overall data workflow of PhIN database. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

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, 7.04, 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)[26] scaffold, which was widely accepted for scaffold representation and used in molecular scaffold analysis[27,28] and was employed for scaffold representation. First, the scaffolds of a target were collected by isolating BM scaffolds of its ligands; then for each scaffold the maximum pChEMBL value of a ligand underlying the scaffold was selected as its activity against the target and the maximum or mean pChEMBL value of a ligand underlying the scaffold can be selected as its activity against the target; finally, target pairs were created depending on their shared scaffolds number and an activity threshold. For example, the sucrase-isomaltase (CHEMBL3114) and maltase-glucoamylase (CHEMBL2074) pair share 11 scaffolds with default searching parameters (shared scaffold 5, pChEMBL value 6) (**Supplementary Figure S1**). The RDKit (http://rdkit.org/) package was used for BM scaffolds generation and there are 309,556 unique scaffolds, which is downloadable at the webserver, in current release.

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,[29] 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*.[4] 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*.[30] 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)[31] 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

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.

shared compounds and 3,428,020 pairs in terms of shared scaffolds.

Pathway integration

Pathway data were collected from SMPDB (v. 2.0)[32] 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

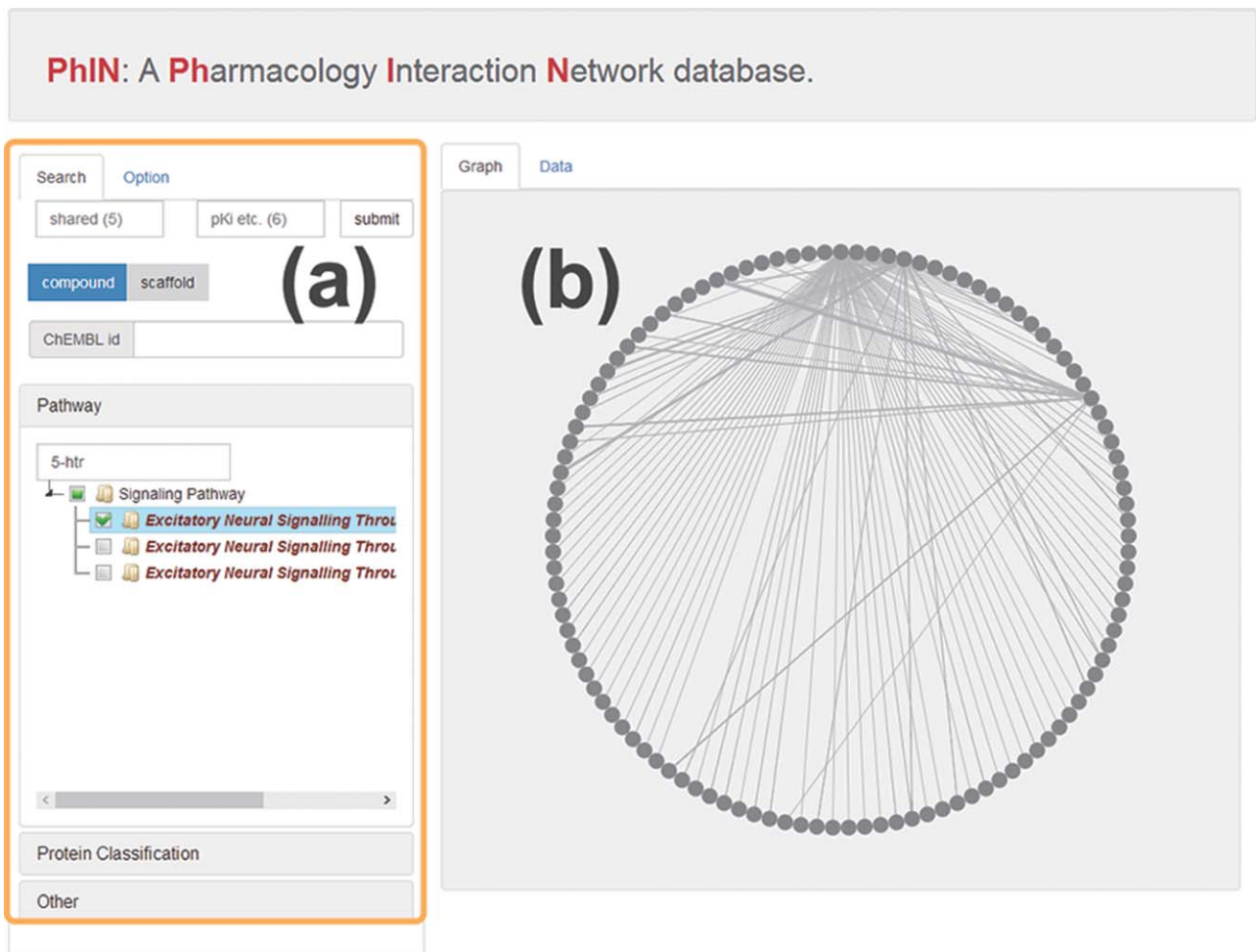


Figure 2 Search and results interface of PhIN. (**a**) Query options, including text areas for thresholds of shared number and pChEMBL value, toggle for network type—compound or scaffold, text area for ChEMBL ID and three tree-like menus. (**b**) Result network graph of excitatory neural signaling through 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.]

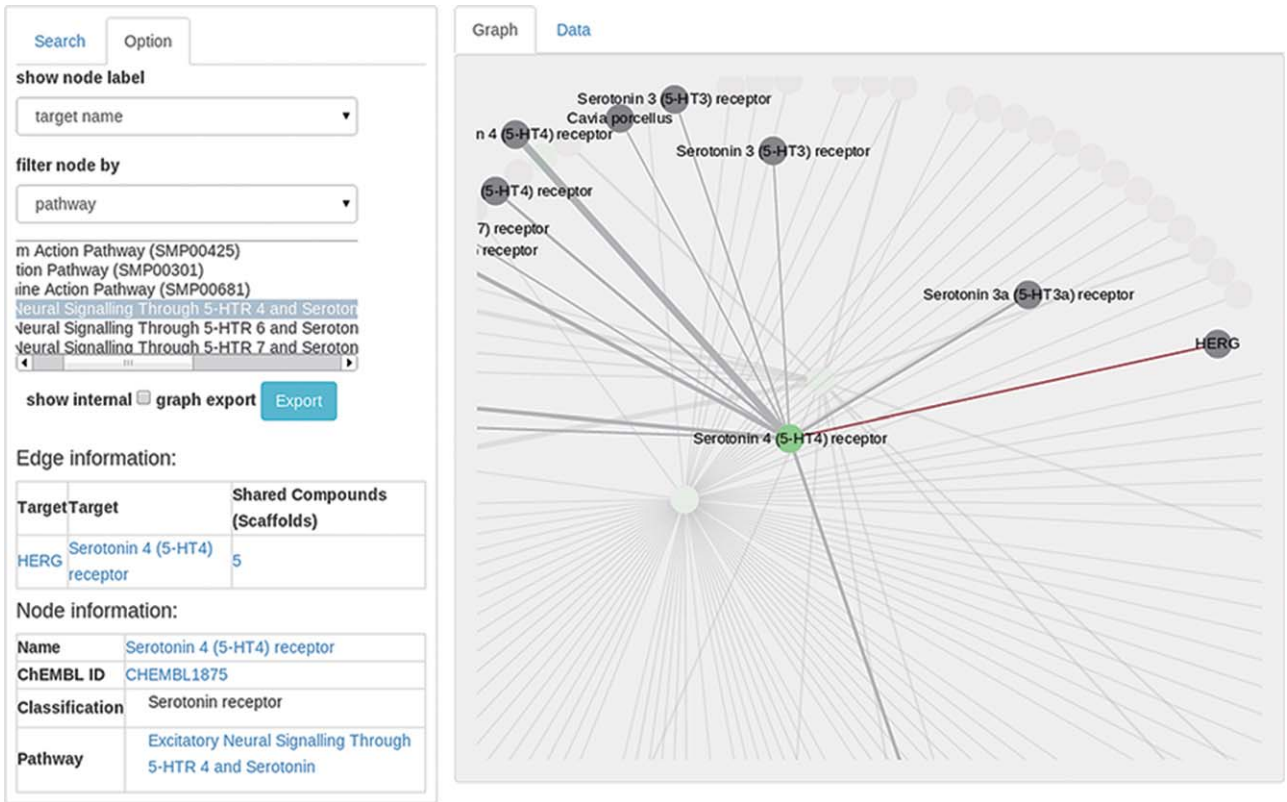


Figure 3 Pharmacology interaction network of serotonin 4 (5-HT 4) receptor. Nodes is highlighted (green) by pathway (excitatory neural signaling through 5-HTR 4 and serotonin). Edge width is proportional to number of shared compounds (scaffolds). Interaction information between HERG and serotonin 4 (5-HT 4) receptor is shown in the left panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

classification system[33] are the two main schemes and both of them were integrated into PhIN. Moreover, targets belonging to viruses also link to ViralZone,[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 cutoff, 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 (**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.

Molecule–molecule functional similarity

Drugs with overlapping drug targets are likely to be side effect similar.[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 (**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.

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.[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 2**a. Then click the submit button, and a network will be created in seconds (**Figure 2**b) (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

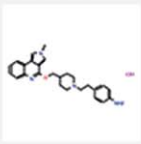
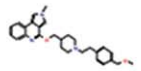
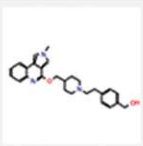
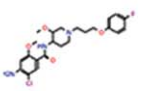
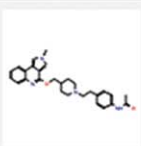
#	Name(ID)	ChEMBL ID	Structure	HERG	Serotonin 4 (5-HT4) receptor
1	1274049	CHEMBL2179675		7.3	8.3
2	1274045	CHEMBL2179671		7.7	7.7
3	1274046	CHEMBL2179672		8.0	8.0
4	CISAPRIDE	CHEMBL1729		7.64615384615	7.54
5	1274044	CHEMBL2179670		7.7	8.7

Figure 4 Shared compounds of serotonin 4 (5-HT 4) receptor and HERG at pChEMBL value above 6. [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.[4] This approach has proven to be useful in target identifications,[5,39] multitarget drug design,[40] and side effect prediction.[36] 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.[41–43] 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.[44] 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

pharmacophore that already provides some of the nascent activity profile rather than attempting to merge and integrate pharmacophores.[4,45]

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 multitargets, 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,[46] DrugBank,[47] TTD,[48] and NCI/Nature Pathway Interaction Database (http://pid.nci.nih.gov), and analytical tools, such as the Similarity Ensemble Approach (SEA)[49] and machine learning-based predictive models,[6] 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).

Conflict of Interest. The authors declare no conflicts of interest.

Author Contributions. J.L. and Z.W. wrote the article. J.L., Z.W., and J.L. designed the research. Z.W., L.L., and R.D. performed the research.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

✓ Analysis of target-pair interaction in terms of shared compounds has proven to be valuable in drug discovery. However, to our knowledge, no comprehensive database that contains such information has been reported.

WHAT QUESTION DID THIS STUDY ADDRESS?

✓ This study addressed applications of an interacting target network, which was generated by shared compounds or a scaffold of a target pair for polypharmacology.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

✓ This study reports a Web-based tool that provides an interacting target network within or across human pathways, between human and virus, virus and virus, using a user-defined target-pair interaction.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS

✓ This Web-based tool can be used to assist polypharmacology, including target identification, side effect prediction, and multitarget drug design.

1. Csermely, P., Korcsmáros, T., Kiss, H.J.M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: A comprehensive review. *Pharmacol. Ther.* **138**, 333–408 (2013).
2. Csermely, P., Agoston, V. & Pongor, S. The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol. Sci.* **26**, 178–182 (2005).
3. Pujol, A., Mosca, R., Farrés, J. & Aloy, P. Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol. Sci.* **31**, 115–123 (2010).
4. Hopkins, A.L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682–690 (2008).
5. Keiser, M.J. *et al.* Predicting new molecular targets for known drugs. *Nature* **462**, 175–181 (2009).
6. Paolini, G.V, Shapland, R.H.B., Hoorn, W.P. van, Mason, J.S. & Hopkins, A.L. Global mapping of pharmacological space. *Nat. Biotechnol.* **24**, 805–815 (2006).
7. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
8. Azmi, A.S. Network pharmacology for cancer drug discovery: are we there yet? *Future Med. Chem.* **4**, 939–941 (2012).
9. Espinoza-Fonseca, L.M. The benefits of the multi-target approach in drug design and discovery. *Bioorg. Med. Chem.* **14**, 896–897 (2006).
10. Zimmermann, G.R., Lehár, J. & Keith, C.T. Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov. Today* **12**, 34–42 (2007).
11. Chandra, N. & Padiadpu, J. Network approaches to drug discovery. *Expert Opin. Drug Discov.* **8**, 7–20 (2013).
12. Leung, E.L., Cao, Z.-W., Jiang, Z.-H., Zhou, H. & Liu, L. Network-based drug discovery by integrating systems biology and computational technologies. *Brief. Bioinform.* **14**, 491–505 (2013).
13. Rask-Andersen, M., Almén, M.S. & Schiöth, H.B. Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.* **10**, 579–590 (2011).

14. Harrold, J.M., Ramanathan, M. & Mager, D.E. Network-based approaches in drug discovery and early development. *Clin. Pharmacol. Ther.* **94**, 651–658 (2013).
15. 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).
16. Goh, K., Cusick, M.E., Valle, D., Childs, B., Vidal, M. & Barabási, A.-L. The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8685–8690 (2007).
17. 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**, 438–447 (2010).
18. Hu, Z. *et al.* VisANT 4.0: integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res.* **41**, W225–231 (2013).
19. Kim Kjærulff, S. *et al.* ChemProt-2.0: visual navigation in a disease chemical biology database. *Nucleic Acids Res.* **41**, D464–469 (2013).
20. Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M. & Goto, S. DINIES: drug-target interaction network inference engine based on supervised analysis. *Nucleic Acids Res.* **42**, W39–45 (2014).
21. Hu, Q.-N. *et al.* VNP: interactive visual network pharmacology of diseases, targets, and drugs. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e105 (2014).
22. Hu, Y., Wassermann, A.M., Lounkine, E. & Bajorath, J. Systematic analysis of public domain compound potency data identifies selective molecular scaffolds across druggable target families. *J. Med. Chem.* **53**, 752–758 (2010).
23. Wang, Y. *et al.* PubChem's bioassay database. *Nucleic Acids Res.* **40**, D400–412 (2012).
24. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–1107 (2012).
25. Liu, T., Lin, Y., Wen, X., Jorissen, R.N. & Gilson, M.K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **35**, D198–201 (2007).
26. Bemis, G.W. & Murcko, M. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **39**, 2887–2893 (1996).
27. Hu, Y., Stumpfe, D. & Bajorath, J. Lessons learned from molecular scaffold analysis. *J. Chem. Inf. Model.* **51**, 1742–1753 (2011).
28. Jiang, Y.-Y., Kong, D.-X., Qin, T., Li, X., Caetano-Anollés, G. & Zhang, H.-Y. The impact of oxygen on metabolic evolution: a chemoinformatic investigation. *PLoS Comput. Biol.* **8**, e1002426 (2012).
29. Bienfait, B. & Ertl, P. JSME: a free molecule editor in JavaScript. *J. Cheminform.* **5**, 24 (2013).
30. Hu, Y. & Bajorath, 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. Mysinger, M.M., Carchia, 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.* ViralZone: 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. van & Bork, P. Network neighbors of drug targets contribute to drug side-effect similarity. *PLoS One* **6**, e22187 (2011).
37. Gershon, M. Review article: serotonin receptors and transporters—roles in normal and abnormal gastrointestinal motility. *Aliment. Pharmacol. Ther.* **20**, 3–14 (2004).
38. Hegde, S.S. & Eglen, R.M. Peripheral 5-HT4 receptors. *FASEB J.* **10**, 1398–1407 (1996).
39. Atreya, R.V., Sun, J. & Zhao, Z. Exploring drug-target interaction networks of illicit drugs. *BMC Genomics* **14** Suppl 4, S1 (2013).
40. Knight, Z.A., Lin, H. & Shokat, K.M. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **10**, 130–137 (2010).
41. Finlayson, K., Witchel, H.J., McCulloch, J. & Sharkey, J. Acquired QT interval prolongation and HERG: implications for drug discovery and development. *Eur. J. Pharmacol.* **500**, 129–142 (2004).
42. Hoffmann, P. & Warner, B. Are hERG channel inhibition and QT interval prolongation all there is in drug-induced torsadogenesis? A review of emerging trends. *J. Pharmacol. Toxicol. Methods* **53**, 87–105 (2006).
43. Roden, D.M. Drug-induced prolongation of the QT interval. *N. Engl. J. Med.* **350**, 1013–1022 (2004).
44. Safety Alerts for Human Medical Products — Propulsid (cisapride) Dear Healthcare Professional Letter Apr 2000. <http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm175000.htm> (2000).
45. Wermuth, C.G. Selective optimization of side activities: another way for drug discovery. *J. Med. Chem.* **47**, 1303–1314 (2004).
46. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–360 (2010).
47. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–1097 (2014).
48. Zhu, F. *et al.* Update of TTD: therapeutic target database. *Database* **38**, 787–791 (2010).
49. Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J. & Shoichet, B.K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **25**, 197–206 (2007).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://www.wileyonlinelibrary.com/psp4)