A Gene Family-led Meta-Analysis of Drug-Target Interactions

Qiong Cheng
Department of Computer Science
University of North Carolina at Charlotte
NC, USA
qcheng1@uncc.edu

Saurabh Mehta
Department of Applied Chemistry
Delhi Technological University Delhi,
India
saurabh.dtu@gmail.com

Stephan Schürer
Department of Molecular and Cellular Pharmacology
University of Miami
FL, USA
sschuerer@med.miami.edu

***Abstract*— The shift from "magic bullets" to "magic shotguns" has prospered the pharmaceutical industry, where a belief is that a drug that "hits" multiple sensitive nodes belonging to a network of interacting targets offers the potential for higher efficacy and may limit drawbacks arising from the use of a single-target drug or a combination of multiple drugs. More high-quality drug-target interactions, profiled by the "magic shotguns" design paradigm, have emerged from diverse labs. However, the rapid production at scale and in variety challenges the efficacy of data usage. A creative and systematical approach is in demand to comprehensively analyze and integrate drug-target interactions for better prediction.**

In the project, taking the diversity into consideration, we carried out a gene family-led, meta-analysis, investigated diverse properties that drug-target pharmacological promiscuity relies on, and examined the consistency or integrity of drug-target data. The novel approach can facilitate the identification of cohesive multi-target combinations and revealed the interconnected properties of determining and rationalizing the promiscuity of profiled drugs. The approach can be further expanded to coordinate other experimental drug-target data and set a stage for the analysis of the mechanism of action of biological therapies.

Keywords*— *polypharmocology, ontology, drug-target interactions, information content, meta-analysis

I. INTRODUCTION

A mainstream drug discovery process is an assembly pipeline starting with target identification, following with lead identification and optimization, and completing with preclinical and clinical assessment [62]. In the process, target selection is a key factor for a success drug; drug efficacy and safety are the main consideration.

Numerous target-based drug profiling datasets have been generated through the efforts of a panel of protein targets against drug compound libraries. They are publically available via different databases such as NIMH Psychoactive Drug Screening Program (PDSP) data [1, 2], Kinome Knowledgebase (KKB) [19], KINOMEscan, and ChEMBL [22].

It is challenging for a data coordinator center to systematically collect and manage the data as well as execute data governance with diversity in consideration, especially when a drug-target experimental data may be biasedly towards a subset of proteins being selected to screen against a chemical library [68, 44]. Unpredicted drug toxicity happens when candidate compound binds to and influences not only the desired target but also other unintended proteins, which cause side effects [63-66]. Identifying and prioritizing on/off targets is challenging and plays a critical role in drug discovery.

We propose a knowledge-driven approach and software infrastructure to facilitate data governance. In the paper, we focus on cross validation of the efficacy of Drug Target Ontology (DTO) and examine the consistency or integrity of drug-target interaction data at diverse levels for better data governance.

First, we used the druggable genome described in the IDG project [51, 59-62, 64] as the reference to examine the number of selected targets in a drug target data set. We further visualized its coverage in the classification tree of drug targets presented in the DTO [59], which underlying gene family classification structure provides a global functional view of target classes. We also employed the statistical method to measure whether compounds interacting with targets within the same protein category of DTO are more similar to those across different categories. We expanded the knowledge driven approach to explore drug promiscuity and calculate the relatedness in terms of the chemical structure similarity of related drug sets.

II. MATERIALS, METHODS, AND RESULTS

A. Data

Our interest in the integrative analysis has been inspired by the early successful replacement of "magic bullets" with "magic shortguns" in treating psychiatric diseases [7,9]. Among various databases, PDSP (NIMH Psychoactive Drug Screening Program) database is a database of binding affinities (Ki) of drugs and chemical compounds for receptors, neurotransmitter transporters, ion channels, and enzymes, which is profiled in diverse species [1-2]. It includes ~55.4k ligand-receptor combinations in the version (kidb110121), where a ligand is defined as active against a protein target if and only if the binding affinity ki<10 M. After adding organisms from UniProt and also adding Human orthologs, we

obtained 3934 unique ligands with SMILES (MW > 50) modulating 428 unique UniProt IDs.

ChEMBL is a database covering a wide range of curated and annotated data, manually extracted from the primary published literature and databases. The database collects the rich information of binding, functional and ADMET (absorption, distribution, metabolism, excretion and toxicity) for a large number of drug-like bioactive compounds [22] and covers targets ranging from single proteins, to protein complexes, then tissues and finally whole organism in vivo data. It includes ~5.4 million bioactivity measurements for more than 1 million compounds and ~5200 protein targets. We extracted all binding bioassay measurements for the LINCS compounds [23], which show the confidence level greater than or equal to 7. There are three binding types, such as Ki, Kd, or IC50. We aggregated all those measurements with the negative logarithm value greater than or equal to 5, along with mapping the target ChEMBL ID to UniprotID and then gene symbol (homo sapiens).

B. Visualization of studied drug targets under the reference of drug target classification showing that studied drug targets are highly biased

In order to effectively analyze and illustrate the coverage of small molecule that bind to the human druggable genome, we employ Drug Target Ontology (DTO) [59] as the reference of druggable target knowledgebase (version DTO_merged_v09_b3101), where a relatively complete collection of disease-linked human proteins and their belonging classification schema are enclosed. The classification of druggable proteins has been represented as in a tree-like hierarchical structure (also called subsumption tree), where leaf nodes represent as protein targets, non-leaf nodes as classes (also called categories), and edges adjacent to two nodes as the "is-a" subsumption relationship [59]. For example, ADRA1A is an alpha adrenoceptor (adrenergic receptor), which is a member of the GPCR superfamily that recognizes the endogenous ligand adrenaline and noradrenaline and which are amines and thus is also classified as aminergic receptor [58]. In the subsumption "is-a" hierarchy, upper-level target class subsumes its lower-level target classes.

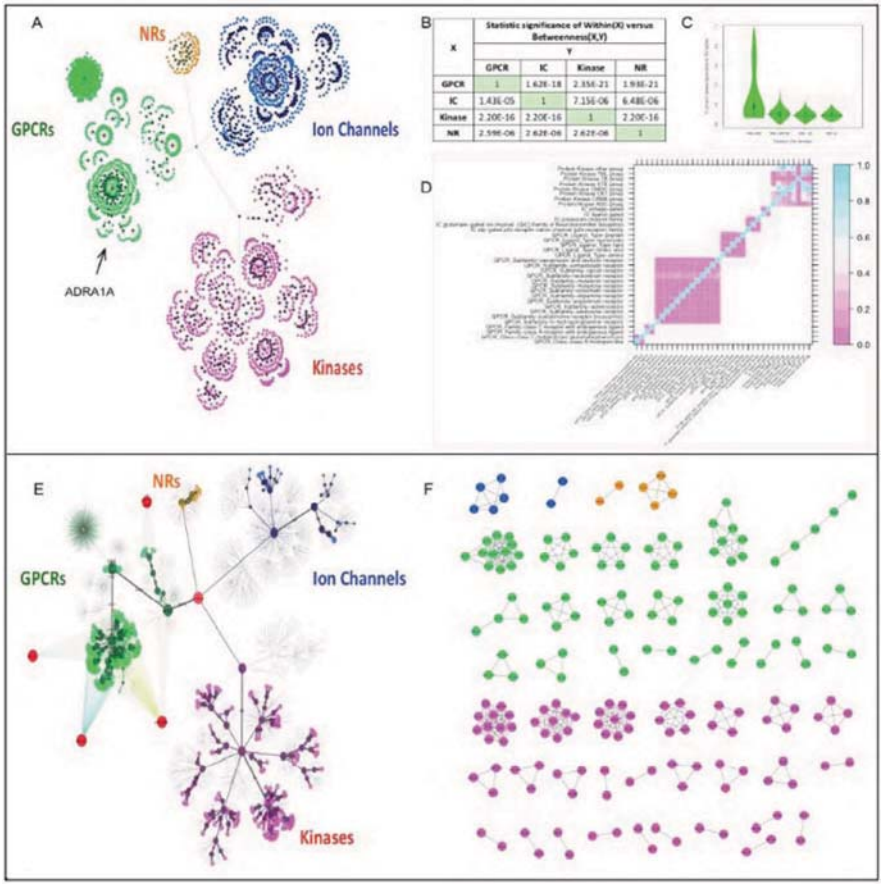

Figure 1. DTO drug target classifications facilitated the integrative analysis of PDSP binding bioactivity data. A. Visualization of DTO drug targets and its classification tree. **B.** Target-driven active PDSP ligands distinguish IDG families. We calculated the similarity of active compounds driven by targets in the same IDG superfamily and also in different superfamilies. We further measured the statistical significance of population difference by t-test. It turns out the active compounds against targets in the same protein family are more similar than in different families. **C.** Distribution of the similarity scores of pair-wise active compound profiled against targets in the same group against in different groups as an example. **D.** Statistical significance of the specificity of DTO classification. We measured the statistical significance for families or subfamilies of interest by using the same method as B. **E.** The coverage of active protein targets selected by PDSP compound library on DTO (Diverse GPCRs' annotations have been stressed, for example, ligand-binding receptor, amino acid derived receptor, amine associated receptor, peptide-binding receptor). **F.** Concordant target-to-target relatedness network, taking account of both related ligand set similarity and DTO target classification hierarchical structure similarity. Red nodes represent kinases, green GPCRs, blue ion channels, and orange nuclear receptors.

We visualized the DTO subsumption tree by the force-directed graph-drawing algorithm plugged in Cytoscape [31, 67] (Figure 1A), where we illuminate different druggable protein families by colors. There are four colors, and each color stands for each superfamily (also called an IDG family, for example, G protein-coupled receptors (GPCRs), kinases, ion channels (ICs), and nuclear receptors (NRs)). Protein nodes are colored in light and protein class nodes in dark. In the reference graph, each node has the uniform size. In addition to the "is-a" relationship, we also included other relationships in order to characterize both chemical structural, and pattern-based or ligand-bound annotation of protein targets. For example, ADRA1A binds an amine endogenous ligand;GRM5 is bound to an amino-acid ligand.

We downloaded the PDSP data, curated, performed a cross-reference to Uniprot IDs, and extracted the active drug-target combinations with a cutoff of Ki < 10uM, and then aggregated active compounds-hitting targets in each category. In order to demonstrate the distribution of those compounds-hit targets in druggable proteome, we use the above-mentioned reference graph as the background with opacity = 0.5 and have highlighted those nodes of compound-interacting targets in PDSP and its belonging classes (see Figure 1E). For each active target or category, we showed the number of active compounds on the edge adjacent to its node. We emphasized the hierarchical structure and active compounds in different target classes by designing the weighted scoring scheme to scale the node size and edge width, which takes account of the size of the category, the level of the category, and the logarithm of the number of active compounds hitting targets.

The knowledge-driven approach to visualize the experimental coverage, as shown in Figure 1A, and 1E, illustrates areas with very limited experimental exploration, which may provide opportunities for new drug development efforts. For example, all 424 olfactory receptor proteins are not profiled in both screening data (see Discussion).

C. Ligand drug target specificity and selection

We also analyzed the number of protein targets profiled against each PDSP compound. Out of 16,969 aggregated human small molecule-target combinations, we observed that 71.2% drugs have been tested against more than one target and 20.2% compounds have been screened by different laboratories, designating them as "compounds of interest". We note that a large number of compounds have been tested biasedly towards a subset of druggable proteome. The observed selectivity bias may reasonably arise from the technical expertise or prior knowledge about the druggable proteome.

D. Relationship between target class and molecular properties

We calculated a set of Tanimoto (Tc) scores as a measure of molecular properties for all pairwise active compounds in the drug-target data in order to investigate the relationship between target class and molecular properties of related active compounds. Tanimoto (Tc) scores are measured under Atom type Extended Connectivity Fingerprints with a diameter of 4 (ECFP4). Distinct differences are observed when comparing the distribution of Tc scores between sets of compounds against same IDG subfamilies with those against different IDG subfamilies (see Figure 1B, 1C). Then we measured the statistical significance of the distribution difference, namely, the occurrence possibility of the mean Tc score of active compounds of intra-target class greater than inter-target classes by chance. Figure 1C takes nuclear receptors as an example, showing the distribution of intra-class and inter-class, where the interquartile range, median, min, max, and density are shown. Through performing the unpaired two-sample one-tailed t-test, we examined that the population mean of the Tc scores calculated within target class is larger than the population mean across different classes. The distribution greatness indicates that active compounds hitting targets in the same class are significantly more similar than the ones hitting targets from different classes. Figure 1B demonstrates the statistical significance of the distribution greatness of intra-class against inter-class.

Out of 645 target classes we collected in the classification of druggable human proteome, we elicited 66 classes of interest, which come from 11 upper-level classes. We focused on the target class pairs across which there are at least a finite number k of pairwise active compounds (k=4). By employing the same validation approach, we examined the relationship between molecular properties and target classes (see Figure 1D). For example, we evaluated the categorization of 1) three members of the GPCR superfamily, such as the rhodopsin-like GPCRs, the secretin-like GPCRs, and the metabotropic glutamate receptor family, 2) five ligand type-associated receptors, such as amine-associated receptors, amino acid receptors, lipid G protein–coupled receptors, nucleoside receptors, and peptide receptors, 3) the 13 GPCR sub families, 4) eight kinase groups, and 5) three ion channel families. A significant p-value reported in the unpaired two-sample one-tailed t-test tested our hypothesis. The assessment of five GPCR ligand types in DTO demonstrated that the categorization has been distinguished unambiguously (see Figure 1D); the assessment of 13 GPCR subfamily classification showed the same merit, except the neurotensin receptors and endothelin receptors; except the kinase TK group, other groups have not been differentiated consistently (see Figure 1D). The confusion may be caused by the innate structures of protein kinases since a large number of protein kinase enzymes share a common cofactor and similar three-dimensional structure of the catalytic site. Thus, active compounds that are screened may not effectively differentiate kinases. Another few of non-proved agreements may arise from the possible biased experimental selectivity of compounds, e.g., with the limited number of compound-GPCR combinations.

We further examined the relationship between target class and molecular properties of the related ligand sets, not individual compound but overall active compounds against targets. Our hypothesis is that overall related ligand sets active against intra-class are more similar than inter-class. Here we measured the target-to-target similarity in terms of the related ligand set similarity using the SEA method (see Methods). We implemented the SEA method in [5] and calculated the Z-score derived from the SEA underlying statistical model, which is

the normalized raw score summarized over the related ligand set(s) (see Methods). By employing unpaired two-sample one-tailed t-test, we measured the statistical significance of the greatness of the population mean of the Z-scores within target class against across different classes. For example, we have a total of 12 receptors in the 5-hydroxytryptamine receptor (5-HT receptors) GPCR subfamily since the diversity of metabotropic 5-HT receptors is increased by alternative splicing that produces isoforms of the 5-HT2A, 5-HT2C, 5-HT4, 5-HT6 and 5-HT7 receptors. We calculated the similarity of pairwise receptors in the 5-HT receptor subfamily by SEA. We also calculated the pairwise target similarity across the 5-HT receptor subfamily and one of any other 11 subfamilies of interest. After the t-test examination, we observed that targets in 5-HT receptor subfamily are more discriminatively different from other 10 GPCR subfamilies, but less distinctly different from dopamine receptors. Authors in [21] demonstrated that 5-HT receptors may modulate the release of striatal dopamine receptors in acute psychoses, which indicates that both subfamilies may function similar, especially in the acute psychotic states.

E. Promiscuity across different target families or a number of target subtypes

The existing and emerging bioassay activity/inhibition data offers a chance of exploring the pharmacology target space through the chemistry of their ligands. The SEA method used multivariate regression model to investigate pairwise target-to-target associations in terms of the related ligand set chemical structural similarity [5,6, 24-30]. We focused on 237 PDSP targets, and through the SEA method we calculated the similarity score of 56,169 target pairs on the basis of the similarity of related ligand sets, where no single ligand need be shared between pairwise targets, but overall. Through the method, we obtained 5870 pairs under the threshold E-value < 10-10 and built target-to-target relatedness network (see Figure 2 and Method). The smaller E-value, the higher confidence the pairwise targets could be related. In the network, each node is a drug target active against at least four compound and nodes are linked by an edge if their expected similarity E-value is satisfied with the specified threshold (see methods), indicating the relatedness of the associated targets. Connected components are formed and targets linked in a component are similar. The target-to-target relatedness network may be varied with the change of the threshold. The smaller the threshold, the more isolated components would be formed. However the target-to-target relatedness network built from the bioactivity data can indicate the diversity of targets inhibited by compounds through multiple distinct mechanisms, e.g., same or different subtypes or remotely related targets. For example, the NECA drug acts through blocking a number of ADORA1, ADORA2A, ADORA2B, and ADORA3 adenosine receptors.

Our observation is as follows. Some of them are consistently from the same IDG family and others are observed as the "cross-over" evidence that targets of interest are related to other targets coming from different types or subtypes of IDG families (see Figure 2). For instance, the glutamate receptor, ionotropic, AMPA 3 (GRIA3) is an ion channel and there are two GPCR targets (GRM2 and GRM3) and three other ion channels (GRIA1, GRIA2, and GRIK4) related to it. Among all those GPCR targets of endogenous ligand-binding mode, somatostatin receptors, prostaglandin receptors, galanin receptors, and adenosine receptors are discovered in the case of the mutual relationships, though coming from different GPCR families or sub-families. We found a list of 24 GPCR targets of different subtypes related to 5-HT3A receptor which is a kind of the serotonin receptors found in the central and peripheral nervous systems. The 5-HT3A receptor is closely related to other serotonin G protein-coupled receptors such as the 5-hydroxytryptamine subtypes 1, 2 and 4–7 (5-HT1,2,4–7) [5,7,8]. It is also related to targets that appear far from the serotonin receptors as shown in many bioinformatics metrics reflected from the ontology, but all of them have endogenous ligand-binding mode. These evidences may suggest that the promiscuity that small molecule compounds possessed is possibly across target boundaries. The polypharmacology across different target families or a number of target subtypes has been also pointed out in [5, 27, 32]. It turns out that local binding site may be more important than global structural or function information.

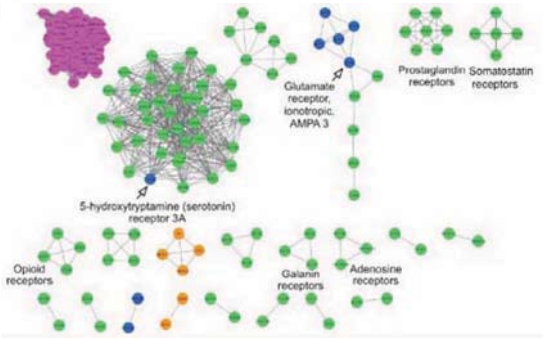

Figure 2. **Target-to-target relatedness network extracted from PDSP**, showing that these functional related ligand-focused targets are grouped in terms of ligand set similarity. Coloring represents different IDG (illuminating druggable genome) families. Red represents kinases, green GPCRs, blue ion channels, and orange nuclear receptors. We observed that completely connected graph have very similar gene symbols, for example, somatostatin receptors, prostaglandin receptors, opioid receptors, galanin receptors, and adenosine receptors. The completely connected components may be varied with the changing of the threshold (the lower the threshold, the more the isolated compounds would be formed). Most of connected components possess a unique color, which indicates that these targets in the component are from the same IDG family. We observed the "cross-over" evidence that interesting targets are related to targets of different types or subtypes of IDG families (see the arrowed nodes). For instance, the glutamate receptor, ionotropic, AMPA 3 (GRIA3) is an ion channel and there are two GPCR targets (GRM2 and GRM3) and three other ion channels (GRIA1, GRIA2, and GRIK4) related to it. 5-HT_{3A} receptor is an ion channel and a list of 24 GPCR targets is related to it.

III. DISCUSSION

We studied the pharmacology space problem in drug discovery by designing a gene family-led meta-level knowledge-driven approach to analyze drug-target data and examine the bias of target selection in experimental design. We analyzed the target promiscuity of compounds in PDSP drug screening data, designed and implemented a uniform way to visualize and compare the coverage of drug targets and to ensemble target-to-target similarity on the basis of associated ligand set similarity, in conjunction with knowledge represented in drug target ontology (DTO). Our ultimate goal

is to integrate and coordinate heterogeneous metadata for understanding dynamic biological systems and to develop improved computational tools that can quantify context-dependent disease-drugs-targets combinations for personalized medication. Such tools can potentially inform the prevention, diagnosis and treatment of diseases.

During this analysis, we note that different types of bias may exist in a binding bioactivity data. One possible concern is the chemical space bias arising from the studied ligands or screening libraries toward biogenic molecules [30]. Another concern exists in the selection of a panel of protein targets against a given compound. We note that some proteins are never profiled against a drug in both screening data, for example, all 424 olfactory receptor proteins. It seems experimentally reasonable since there is the lack of comprehensive understanding of olfactory mechanism in relation to diverse chronic psychiatric diseases (the sense of smells is neglected by clinicians in routine work) [57]. However, it is noticed that olfaction plays an important role in protecting us from environmental dangers. Olfactory dysfunction is common in later life and early warning of neurodegenerative disease, which is receiving more attention as a sign in the early detection and differential diagnosis of neurological and mental diseases. The timely effective treatment of olfactory loss is facilitative to cure related complications of neurodegenerative disease.

It is observed that the selectivity of most drugs is usually profiled against a panel of same kinase families or GPCRs families. It is reasonable since "binding sites within members of conserved and evolutionarily related targets are generally conserved and prone to multi-target inhibition" [4]. However false negatives may be expected since not all targets have been profiled for a given compound. Especially, it is rarely profiled for those targets belonging to different gene families, distantly related or unrelated. Those unintended off-targets may cause unpredicted side effects or unforeseen serious toxicity. It is actually impossible to specify all multi-targets by prior knowledge if the mechanism of drug action is not clear. It is also currently infeasible to profile all drugs against a panel of all proteins for diseases at different phrases, disease progression, and traits that affects specific tissues, tissue compartments, or cells. Computational tools therefore offer an attractive alternative, especially if they can be shown to reduce the search space or leverage the priority of potential drug-specific target combinations in the context since the search space for multi-targets and multi-drugs is huge.

ACKNOWLEDGMENT

We especially thank the investigators and their colleagues who generated and generously deposited drug-screening data. This research was supported by grant U54HL127624 awarded by the National Heart, Lung, and Blood Institute through funds provided by the trans-NIH Library of Integrated Network-based Cellular Signatures (LINCS) Program (http://www.lincsproject.org/) and the trans-NIH Big Data to Knowledge (BD2K) initiative (http://www.bd2k.nih.gov).

REFERENCES

[1] K., W. K., S., M. F., Huang, X. P., Lansu, K., McCorvy, J. D., Giguère, P. M., ... & Roth, B. L. (2015). PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. Nature structural & molecular biology, 22(5), 362-369.

[2] Besnard J, Ruda GF, Setola V, Abecassis K, Rodriguiz RM, Huang XP, Norval S, Sassano MF, Shin AI, Webster LA, Simeons FR, Stojanovski L, Prat A, Seidah NG, Constam DB, Bickerton GR, Read KD, Wetsel WC, Gilbert IH, Roth BL, Hopkins AL. (2012) Automated design of ligands to polypharmacological profiles. Nature. 2012 Dec 13;492(7428):215-20. doi: 10.1038/nature11691.

[3] Anton Yuryev (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Pathway Analysis for Drug Discovery. ISBN 978-0-470-10705-8

[4] Guilio Rastelli and Luca Pinzi. (2015) Computational polypharmacology comes of age. Frontiers in Pharmacology. Volume 6 | article 157. doi: 10.3389/fphar.2015.00157

[5] Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. (2007) Relating protein pharmacology by ligand chemistry. Nat Biotech 25 (2), 197-206.

[6] Hert, J., Keiser, M. J., Irwin, J. J., Oprea, T. I., & Shoichet, B. K. (2008). Quantifying the relationships among drug classes. Journal of chemical information and modeling, 48(4), 755-765.

[7] Roth, B.L., Sheffler, D.J. & Kroeze, W.K. (2004) Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat. Rev. Drug Discov. 3, 353–359.

[8] Kroeze, W.K., Kristiansen, K. & Roth, B.L. (2002) Molecular biology of serotonin receptors structure and function at the molecular level. Curr. Top. Med. Chem. 2, 507–528.

[9] Wesley K. Kroeze and Bryan L. Roth. (2012) Polypharmacological Drugs: " Magic Shotguns " for Psychiatric Diseases. Peters, Jens-Uwe, ed. Polypharmacology in Drug Discovery. Somerset, NJ, USA: John Wiley & Sons.

[10] Uhlén M et al, 2015. Tissue-based map of the human proteome. Science

[11] Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., ... & Furlong, L. I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database, 2015, bav028.

[12] Pfam database.

[13] Mestres, Jordi, et al. (2008) Data completeness—the Achilles heel of drug-target networks. Nature biotechnology 26.9: 983-984.

[14] GO and related tools.

[15] Zhidian Du, Lin Li, Chin-Fu Chen, Philip. S. Yu, and James Z. Wang. (2009) G-sesame: web tools for go-term-based gene similarity analysis and knowledge discovery. Nucleic Acids Research, 37:W345–W349

[16] Li H, Pellegrini M, Eisenberg D. (2005) Detection of parallel functional modules by comparative analysis of genome sequences. Nat Biotechnol 2005, 23(2):253-260.

[17] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D668-72. 16381955

[18] Uhlén et al (2015). Tissue-based map of the human proteome. Science PubMed: 25613900 DOI: 10.1126/science.1260419

[19] Stephan C. Schürer and Steven M. Muskal (2012) Kinome-wide Activity Modeling from Diverse Public High-Quality Data Sets. J Chem Inf Model. 2013; 53(1): 27–38 DOI: 10.1021/ci300403k

[20] Law V et al. (2014) DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014 Jan 1;42(1):D1091-7. PubMed ID: 24203711

[21] Vollenweider, F. X., Vontobel, P., Hell, D., & Leenders, K. L. (1999) 5-HT modulation of dopamine release in basal ganglia in psilocybin-induced psychosis in man—a PET study with [11C] raclopride. Neuropsychopharmacology, 20(5), 424-433.

[22] A.P. Bento, A. Gaulton, A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Kruger, Y. Light, L.Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos and J.P. Overington (2014) The ChEMBL

[23] LINCS.

[24] Lounkine, E., Keiser, M. J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J., … Urban, L. (2012). Large Scale Prediction and Testing of Drug Activity on Side-Effect Targets. Nature, 486(7403), 361–367.

[25] Laggner, C., Kokel, D., Setola, V., Tolia, A., Lin, H., Irwin, J. J., … Shoichet, B. K. (2011). Chemical Informatics and Target Identification in a Zebrafish Phenotypic Screen. Nature Chemical Biology, 8(2), 144–146.

[26] Yadav, P. N., Abbas, A. I., Farrell, M. S., Setola, V., Sciaky, N., Huang, X.-P., … Roth, B. L. (2011). The Presynaptic Component of the Serotonergic System is Required for Clozapine's Efficacy. Neuropsychopharmacology, 36(3), 638–651.

[27] Keiser, M. J., Irwin, J. J., & Shoichet, B. K. (2010). The Chemical Basis of Pharmacology. Biochemistry, 49(48), 10267–10276.

[28] DeGraw, A. J., Keiser, M. J., Ochocki, J. D., Shoichet, B. K., & Distefano, M. D. (2010). Prediction and evaluation of protein farnesyltransferase inhibition by commercial drugs. Journal of Medicinal Chemistry, 53(6), 2464–2471.

[29] Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A., Hufeisen, S. J., … Roth, B. L. (2009). Predicting new molecular targets for known drugs. Nature, 462(7270), 175–181.

[30] Hert, J., Irwin, J. J., Laggner, C., Keiser, M. J., & Shoichet, B. K. (2009). Quantifying Biogenic Bias in Screening Libraries. Nature Chemical Biology, 5(7), 479–483.

[31] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research, 13(11), 2498–2504.

[32] Paolini, G. V., Shapland, R. H., van Hoorn, W. P., Mason, J. S., & Hopkins, A. L. (2006). Global mapping of pharmacological space. Nature biotechnology, 24(7), 805-815.

[33] Yıldırım, M. A., Goh, K. I., Cusick, M. E., Barabási, A. L., & Vidal, M. (2007). Drug—target network. Nature biotechnology, 25(10), 1119-1126.

[34] Chen B, Ding Y, Wild DJ (2012) Assessing Drug Target Association Using Semantic Linked Data. PLoS Comput Biol 8(7): e1002574. doi:10.1371/journal.pcbi.1002574

[35] A. J. Butte, I. S. Kohane (2000) Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput. 2000:418-29 PMID: 10902190

[36] Bauer-Mehren,A., Rautschka,M., Sanz,F. et al. (2010) DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics, 26, 2924–2926

[37] Kim et al.(2014) A draft map of the human proteome. Nature. 509, 575-581.

[38] Schriml, L. M., Arze, C., Nadendla, S., Chang, Y.-W. W., Mazaitis, M., Felix, V., … Kibbe, W. A. (2012) Disease Ontology: a backbone for disease semantic integration. Nucleic Acids Research, 40(Database issue), D940–D946.

[39] Bauer-Mehren,A., Bundschus,M., Rautschka,M. et al. (2011) Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. PLoS One, 6, e20284

[40] Tang J, Karhinen L, Xu T, Szwajda A, Yadav B, Wennerberg K, et al. (2013) Target Inhibition Networks: Predicting Selective Combinations of Druggable Targets to Block Cancer Survival Pathways. PLoS Comput Biol 9(9): e1003226. doi:10.1371/journal.pcbi.1003226

[41] Huang,L., Li,F., Sheng,J. et al. (2014) DrugComboRanker: drug combination discovery based on target network analysis. Bioinformatics, 30, i228–i236.

[42] Liu, Y., Wei, Q., Yu, G., Gai, W., Li, Y., & Chen, X. (2014) DCDB 2.0: a major update of the drug combination database. Database, 2014, bau124.

[43] Knight, Z. A., Lin, H., & Shokat, K. M. (2010) Targeting the cancer kinome through polypharmacology. Nature Reviews Cancer, 10(2), 130-137.

[44] Cheng, Q., Ursu, O., Oprea, T. I., & Schurer, S. (2017). Learning reference-enriched approach towards large scale active ontology alignment and integration. BIBM 2017 pp. 1658-1663.

[45] Jalencas, X., & Mestres, J. (2013) Identification of Similar Binding Sites to Detect Distant Polypharmacology. Molecular Informatics, 32(11-12), 976-990.

[46] Miller, D. W., & Dill, K. A. (1997) Ligand binding to proteins: the binding landscape model. Protein science, 6(10), 2166-2179.

[47] Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., ... & Tang, Y. (2012) Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol, 8(5), e1002503.

[48] Campillos M et al. (2008) Drug target identification using side-effect similarity. Science 321:263–266

[49] Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24: i232–240.

[50] Yamanishi, Y., Kotera, M., Kanehisa, M., & Goto, S. (2010) Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. Bioinformatics, 26(12), i246-i254.

[51] Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., & Overington, J.P. A Comprehensive Map of Molecular Drug Targets. Nature Rev. Drug Discov. doi:10.1038/nrd.2016.230 (2016)

[52] Schadt, E. E., Friend, S. H., & Shaywitz, D. A. (2009) A network view of disease and compound screening. Nature reviews Drug discovery, 8(4), 286-295.

[53] Liu, W., Wu, A., Pellegrini, M., & Wang, X. (2015) Integrative analysis of human protein, function and disease networks. Scientific reports, 5.

[54] Hopkins, A. L. (2008) Network pharmacology: the next paradigm in drug discovery. Nature chemical biology, 4(11), 682-690.

[55] Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006) How many drug targets are there?. Nature reviews Drug discovery, 5(12), 993-996.

[56] Larance, M., & Lamond, A. I. (2015) Multidimensional proteomics for cell biology. Nature Reviews Molecular Cell Biology.

[57] Kivity S., Ortega-Hernandez O. D. and Shoenfeld Y. (2009) Olfaction--a window to the mind. Isr Med Assoc J. 2009 Apr;11(4):238-43. PMID: 19603599

[58] Przydzial, M. J., Bhhatarai, B., Koleti, A., Vempati, U., & Schürer, S. C. (2013). GPCR ontology: development and application of a G protein-coupled receptor pharmacology knowledge framework. Bioinformatics. 2013 Dec 15;29(24):3211-9. doi: 10.1093/bioinformatics/btt565. Epub 2013 Sep 29.

[59] Lin, Y., Mehta, S., McGinty, H. K., Turner, J. P., Vidovic, D., Forlin, M., ... & Mathias, S. L. (2017). Drug Target Ontology to Classify and Integrate Drug Discovery Data. bioRxiv, 117564

[60] Cannon, D.C., Yang, J.J., Mathias, S.L., Ursu, O., Mani, S., Waller, A., Schürer, S.C., Jensen, L.J., Sklar, L.A., Bologa, C.G. and Oprea, T.I., 2017. TIN-X: target importance and novelty explorer. Bioinformatics, p.btx200.

[61] Ursu, O., Holmes, J., Knockel, J., Bologa, C.G., Yang, J.J., Mathias, S.L., Nelson, S.J. and Oprea, T.I., 2017. DrugCentral: online drug compendium. Nucleic acids research, 45(D1), pp.D932-D939

[62] Loging WT, editor. Bioinformatics and Computational Biology in Drug Discovery and Development. Cambridge University Press; 2016 Mar 17.

[63] Acharya, A. and Garg, L.C., 2016. Drug Target Identification and Prioritization for Treatment of Ovine Foot Rot: An In Silico Approach. International journal of genomics, 2016

[64] Russ, A. P., & Lampel, S. (2005). The druggable genome: an update. Drug discovery today, 10(23), 1607-1610.

[65] Kolaja, K. (2012). Drug discovery: computer model predicts side effects. Nature, 486(7403), 326-328.

[66] Lounkine, E., Keiser, M.J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J., Lavan, P., Weber, E., Doak, A.K., Côté, S. and Shoichet, B.K., 2012. Large scale prediction and testing of drug activity on side-effect targets. Nature, 486(7403), p.361.

[67] Kobourov, S. G. (2004). Force-directed drawing algorithms.

[68] Cheng, Q., Lopez, F. A., Duran, C., Camarillo, C., Oprea, T. I., & Schurer, S. (2017). The ontology reference model for visual selectivity analysis in drug-target interactions. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 2091-2097). IEEE.