Perspectives in Science (2016) 9, 49—52

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/pisc

Computational approaches for the study of the role of small molecules in diseases☆

Monica Campillos[a,b,*]

[a] *Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany*
[b] *German Center for Diabetes Research, Helmholtz Zentrum München, 85764 Neuherberg, Germany*

Received 21 March 2016; accepted 16 May 2016
Available online 26 August 2016

KEYWORDS
Drug discovery;
Drug target prediction;
Drug—disease connections;
Computational approaches;
Systems biology;
Bioinformatics

Summary An enormous amount of molecular and phenotypic information of drugs as well as diseases is now available in public repositories. Computational analysis of these datasets is facilitating the acquisition of a systems view of how drugs act on our human organism and interfere with diseases. Here, I highlight recent approaches integrating large-scale information of drugs and diseases that are contributing to change our current view on how drugs interfere with human diseases.

Introduction

Small molecules are the substances most often used as therapeutic agents. However, despite the huge investment of pharmaceutical companies in the development of new drugs, only few novel compounds are approved annually for medical treatment (Emanuel, 2015). The high drug attrition rate is due to a lack of efficacy and unexpected toxicity of drugs (Waring et al., 2015), indicating that our understanding on how compounds affect human biological circuits and interfere with diseases is far from complete.

The recent explosion of biological information of drugs in the public domain is facilitating the study of drug action on the human organism in an unprecedented scale. Over the last two decades, several databases storing molecular and phenotypic information of drugs have appeared on the public domain. Examples of drug target databases are DrugBank (Wishart et al., 2006), ChEMBL (Gaulton et al., 2011) and Matador (Gunther et al., 2008). Repositories of *in vivo* and *in vivo* phenotypic effects of drugs include SIDER (Kuhn et al., 2010, 2016), a database of side effects of marketed drugs, warehouses of high-throughput chemical

* Correspondence to: Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany.
E-mail address: monica.campillos@helmholtz-muenchen.de

http://dx.doi.org/10.1016/j.pisc.2016.08.001

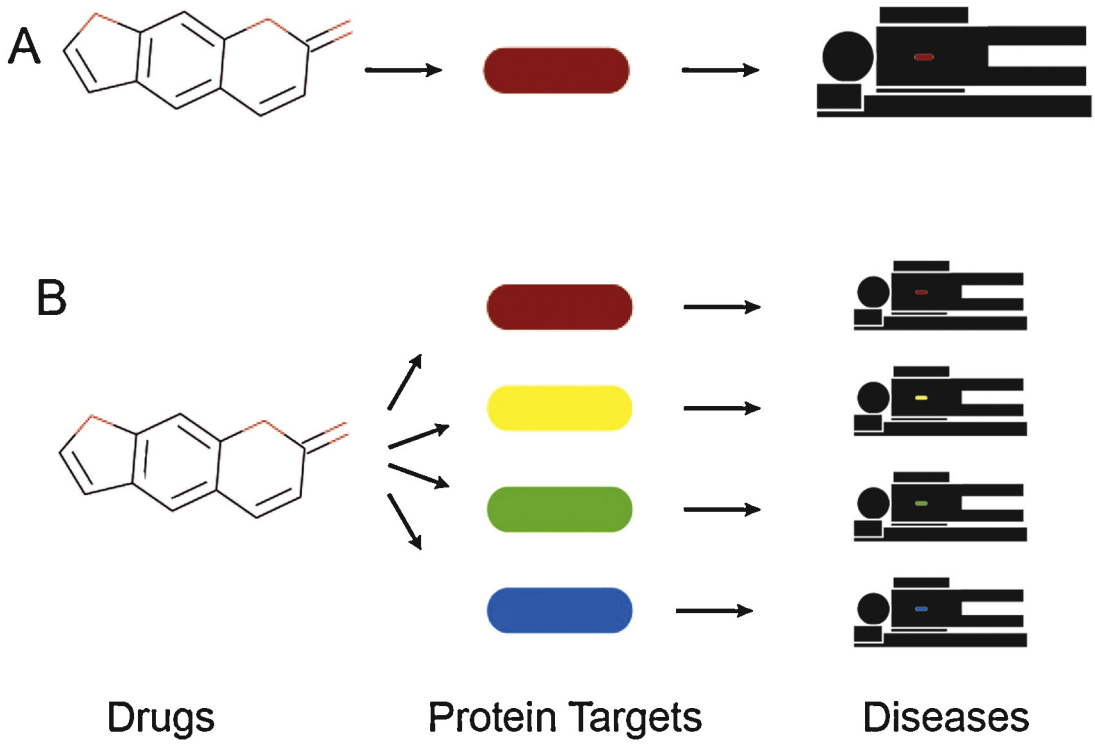

Figure 1 The classical pharmacology view of ''one drug, one target, one disease'' (A) is changing to a more complex scenario of ''many drugs, many targets and many diseases'' (B).

genetics experiments such as ChemBank (Seiler et al., 2008) and PubChem Biassay (Wang et al., 2009) and repositories of gene expression profiles after drug perturbation in cancer cell lines (Lamb et al., 2006).

Resources containing large-scale information of diseases have existed since more than three decades. The first database collecting clinical as well as molecular information of inherited diseases was the 'Online Mendelian Inheritance in Man (OMIM)' (http://omim.org/). More recently, dedicated databases storing genome-wide association disease studies (GWAS), such as the NHGRI GWAS Catalog (Welter et al., 2014) and molecular information of diseases (Pinero et al., 2015) as well as resources offering clinical phenotypes of more than 5000 common and rare diseases (Kohler et al., 2014; Vogt et al., 2014a,b) such as Orphanet and Decipher (Firth et al., 2009) have been released in the public domain.

The integrative analysis of chemical and disease information is changing our view on drug mechanisms of action as well as how drugs interfere with disease mechanisms. The analysis of large-scale drug target information soon evidenced the polypharmacological activity of drugs, that is, the property of drugs to interfere with many protein targets (Anighoro et al., 2014; Jalencas and Mestres, 2013; Peters, 2013). The classical view of ''one drug, one target, one disease'' (Imming et al., 2006) is evolving to a more complex scenario of ''many drugs, many targets and many diseases'' (Mestres et al., 2008; Yildirim et al., 2007) (Fig. 1). Here, I will highlight recent computational efforts that have contributed to enhance our knowledge of drug modes of action and disease relationships.

Results

Elucidation of drug targets

Due to the medical and biological relevance of the discovery of novel drug targets, uncovering new targets of drugs has been an active area on drug discovery research in the last years. Diverse chemo and bio-informatics approaches have been developed to predict drug targets. Chemo-informatics approaches exploit similarities on two and three dimensional structural features of compounds to assign novel targets to compounds (Keiser et al., 2007; Liu et al., 2013; Paolini et al., 2006; Xia et al., 2004), while bioinformatics approaches rely on the analysis of biological properties of drugs. These properties include side effects (Campillos et al., 2008), gene expression profiles after drug perturbation (Lamb et al., 2006; Xia et al., 2004), cytotoxicity profiles of chemicals across a panel of cancer cell lines (Shoemaker, 2006) and bioactivity profiles of chemicals on chemical genetics screens (Petrone et al., 2012).

Biological and chemical properties of compounds have also been exploited in combination to uncover molecular information of compounds, for example in docking approaches where the interaction between compounds and proteins are modeled based on the compound and protein structures (Laird and Blake, 2004) and machine-learning methods that incorporate chemical structure and protein target information (Li et al., 2015).

Drug—disease relationships

Recent computational approaches exploit large scale information on drugs and diseases to infer novel drug—disease connections (Guney et al., 2016; Vogt et al., 2014a; Hopkins and Groom, 2002; Yildirim et al., 2007). Profiling methods where signatures of phenotypic features such as gene expression profiles of drugs and diseases are compared is a common approach to link drugs and diseases. In this context, anticorrelations of gene expression profiles after drug perturbation in cancer cell lines with gene expression signatures of diseases have revealed novel drug—disease connections (Lamb et al., 2006).

Recently, a semantic similarity method comparing signatures of organismal phenotypes of drugs and diseases has shown that drug—disease pairs sharing organismal phenotypes are often molecularly as well as clinically related (Vogt et al., 2014a), that is, phenotypically similar drug—disease pairs are enriched in drugs whose protein targets are functionally related to proteins encoded in disease genes. In addition, in these pairs, the drug is often indicated or contraindicated for the disease. Interestingly, we found that contraindicated drug—disease pairs are preferentially enriched in molecularly related associations, suggesting that by targeting the protein causally related to the disease, the drug can cause the disease and thereby produce side effects resembling disease symptoms. This finding was exploited to propose drug contraindications based on the phenotypic similarity of drugs and diseases.

Another rich source of drug—disease associations are chemical screens where the activity of a library of small molecules is tested in phenotypic assays modeling diseases. Methods that combine phenotypic high-throughput chemical screens stored in public repositories with predicted protein targets of compounds have proven useful to uncover not only drug molecular mechanisms responsible for the phenotypic activity of compounds in chemical screens disease relationships but also drug—disease relationships (Liu and Campillos, 2014; Petrone et al., 2012; Wassermann et al., 2015). High-throughput chemical screening in cancer cell lines is opening novel opportunities for the discovery of personalized treatments in cancer. Computational analyses exploiting cell cytotoxicity information of thousands of small molecules and an extensive molecular information of cell lines such as gene mutations and transcriptomics information are uncovering molecular biomarkers of drug sensitivity as well as drug mechanism of actions (Garnett et al., 2012; Rees et al., 2016; Shoemaker, 2006).

Methods that analyze drug—disease connections using molecular networks have contributed to obtain a systems view of drug therapeutic action (Yildirim et al., 2007). These analyses have revealed that only a small number of drugs target disease mechanisms directly, implying that the majority of drugs have a palliative effect, treating disease symptoms rather than disease molecular causes (Guney et al., 2016; Vogt et al., 2014a; Yildirim et al., 2007).

Concluding remarks

The studies mentioned above illustrate the possibilities of integrative computational methods for the discovery of novel insights of drugs and diseases. It is envisioned that the upcoming clinical and molecular information of patients including individual genome sequences and epigenomes, detailed clinical information on patients, histopathological features and even single-cell information will open new avenues for a deeper understanding of the individual drug response. The analysis of this amount of information will certainly require advanced methods crossing the bioinformatics, system biology and systems medicine disciplines to precisely determine the best individual treatment option.

References

Anighoro, A., Bajorath, J., Rastelli, G., 2014. Polypharmacology: challenges and opportunities in drug discovery. J. Med. Chem. 19, 7874—7887, http://dx.doi.org/10.1021/jm5006463.

Campillos, M., Kuhn, M., Gavin, A.-C., Juhl Jensen, L., Bork, P., 2008. Drug target identification using side-effect similarity. Science 5886, 263—266, http://dx.doi.org/10.1126/science.1158140.

Emanuel, E.J., 2015. The solution to drug prices. New York Times.

Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., Carter, N.P., 2009. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am. J. Hum. Genet. 4, 524—533, http://dx.doi.org/10.1016/j.ajhg.2009.03.010.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., Liu, Q., Iorio, F., Surdez, D., Chen, L., Milano, R.J., Bignell, G.R., Tam, A.T., Davies, H., Stevenson, J.A., Barthorpe, S., Lutz, S.R., Kogera, F., Lawrence, K., McLaren-Douglas, A., Mitropoulos, X., et al., 2012. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 7391, 570—575, http://dx.doi.org/10.1038/nature11005.

Gaulton, A., Bellis, L.J., Bento, P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michaelovich, D., Al-Lazikani, B., Overrington, J.P., 2011. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res., D1100—D1107, http://dx.doi.org/10.1093/nar/gkr777, Database issue.

Guney, E., Menche, J., Vidal, M., Barábasi, A.L., 2016. Network-based in silico drug efficacy screening. Nat. Commun. 10331, http://dx.doi.org/10.1038/ncomms10331.

Gunther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Ahmed, J., Garcia Urdiales, E., Gewiess, A., Juhl Jensen, L., Schneider, R., Skoblo, R., Russell, R.B., Bourne, P.E., Bork, P., Preissner, R., 2008. SuperTarget and Matador: resources for exploring drug-target relationships. Nucleic Acids Res., D919—D922, http://dx.doi.org/10.1093/nar/gkm862, Database issue.

Hopkins, A.L., Groom, C.R., 2002. The druggable genome. Nat. Rev. Drug Discov. 9, 727—730, http://dx.doi.org/10.1038/nrd892.

Imming, P., Sinning, C., Meyer, A., 2006. Drugs, their targets and the nature and number of drug targets. Nat. Rev. Drug Discov. 10, 821—834, http://dx.doi.org/10.1038/nrd2132.

Jalencas, X., Mestres, J., 2013. On the origins of drug polypharmacology. MedChemComm 1, 80—87, http://dx.doi.org/10.1039/c2md20242e.

Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J., Shoichet, B.K., 2007. Relating protein pharmacology by ligand chemistry. Nat. Biotechnol. 2, 197—206, http://dx.doi.org/10.1038/nbt1284.

Kohler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C.M., Brown, D.L., Brudno, M., Campbell, J., FitzPatrick, D.R., et al., 2014. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res., D966—D974, http://dx.doi.org/10.1093/nar/gkt1026, Database issue.

Kuhn, M., Campillos, M., Letunic, I., Juhl Jensen, L., Bork, P., 2010. A side effect resource to capture phenotypic effects of drugs. Mol. Syst. Biol., 343, http://dx.doi.org/10.1038/msb.2009.98.

Kuhn, M., Letunic, I., Juhl Jensen, L., Bork, P., 2016. The SIDER database of drugs and side effects. Nucleic Acids Res. D1, D1075—D1079, http://dx.doi.org/10.1093/nar/gkv1075.

Laird, E.R., Blake, J.F., 2004. Structure-based generation of viable leads from small combinatorial libraries. Curr. Opin. Drug Discov. Dev. 3, 354—359.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S.A., Haggarty, N.J., Clemons, P.A., Wei, R., Carr, S.A., Lander, E.S., GolubF T.R., 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 5795, 1929—1935, http://dx.doi.org/10.1126/science.1132939.

Li, Z.-C., Huang, M.-H., Zhong, W.-Q., Liu, Z.-Q., Xie, Y., Dai, Z., Zou, X.Y., 2015. Identification of drug-target interaction from interactome network with 'guilt-by-association' principle and topology features. Bioinformatics 32, 1057—1064, http://dx.doi.org/10.1093/bioinformatics/btv695.

Liu, X., Campillos, M.M., 2014. Unveiling new biological relationships using shared hits of chemical screening assay pairs. Bioinformatics 17, i579—i586.

Liu, X., Vogt, T., Haque, T., Campillos, M., 2013. HitPick: a web server for hit identification and target prediction of chemical screenings. Bioinformatics 15, 1910—1912, http://dx.doi.org/10.1093/bioinformatics/btt303.

Mestres, J., Gregori-Puigjané, E., Valverde, S., Solé, R.V., 2008. Data completeness — the Achilles heel of drug-target networks. Nat. Biotechnol. 9, 983—984, http://dx.doi.org/10.1038/nbt0908-983.

Paolini, G.V., Shapland, R.P., van Hoorn, W.P., Mason, J.S., Hopkins, A.L., 2006. Global mapping of pharmacological space. Nat. Biotechnol. 7, 805—815, http://dx.doi.org/10.1038/nbt1228.

Peters, J.-U., 2013. Polypharmacology — foe or friend? J. Med. Chem. 22, 8955—8971, http://dx.doi.org/10.1021/jm400856t.

Petrone, P.M., Simms, B., Nigsch, F., Lounkine, E., Kutchukian, P., Cornett, A., Deng, Z., Davies, J.W., Lenkins, J.L., Glick, M., 2012. Rethinking molecular similarity: comparing compounds on the basis of biological activity. ACS Chem. Biol. 8, 1399—1409, http://dx.doi.org/10.1021/cb3001028.

Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., Furlong, L.I., 2015. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford), bav028, http://dx.doi.org/10.1093/database/bav028.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., Soule, C.K., Alexander, B., Li, A., Montgomery, P., Kotz, J.D., Hon, C.S.-Y., Munoz, B., Liefeld, T., Dančík, V., Haber, D.A., Clish, C.B., Bittker, J.A., Palmer, M., Wagner, B.K., Clemons, P.A., Shamji, A.F., Schreiber, S.L., 2016. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. 2, 109—116, http://dx.doi.org/10.1038/nchembio.1986.

Seiler, K.P., George, G.A., Happ, M.P., Bodycombe, N.E., Carrinski, H.A., Norton, S., Brudz, S., Sullivan, J.P., Muhlich, J., Serrano, M., Ferraiolo, P., Tolliday, N.J., Schreiber, S.L., Clemons, P.A., 2008. ChemBank: a small-molecule screening and cheminformatics resource database. Nucleic Acids Res., D351—D359, http://dx.doi.org/10.1093/nar/gkm843, Database issue.

Shoemaker, R.H., 2006. The NCI60 human tumour cell line anticancer drug screen. Nat. Rev. Cancer 10, 813—823, http://dx.doi.org/10.1038/nrc1951.

Vogt, I., Prinz, J., Campillos, M., 2014a. Molecularly and clinically related drugs and diseases are enriched in phenotypically similar drug-disease pairs. Genome Med. 52, http://dx.doi.org/10.1186/s13073-014-0052-z.

Vogt, I., Prinz, J., Worf, K., Campillos, M., 2014b. Systematic analysis of gene properties influencing organ system phenotypes in mammalian perturbations. Bioinformatics 21, 3093—3100, http://dx.doi.org/10.1093/bioinformatics/btu487.

Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., Byrant, S.H., 2009. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res., W623—W633, http://dx.doi.org/10.1093/nar/gkp456, Web Server issue.

Waring, M.J., Arrowsmith, J., Leach, A.R., Leeson, P.D., Mandrell, S., Owen, R.M., Pairaudeau, G., Pennie, W.D., Pickett, S.D., Wang, J., Wallace, O., Weir, A., 2015. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat. Rev. Drug Discov. 7, 475—486, http://dx.doi.org/10.1038/nrd4609.

Wassermann, A.M., Lounkine, E., Davies, J.W., Glick, M., Camargo, L.M., 2015. The opportunities of mining historical and collective data in drug discovery. Drug Discov. Today 4, 422—434, http://dx.doi.org/10.1016/j.drudis.2014.11.004.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H., 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res., D1001—D1006, http://dx.doi.org/10.1093/nar/gkt1229, Database issue.

Wishart, D.S., Knox, C., Guo, A.C., 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., D668—D672, http://dx.doi.org/10.1093/nar/gkj067, Database issue.

Xia, X., Maliski, E.G., Gallant, P., Rogers, D., 2004. Classification of kinase inhibitors using a Bayesian model. J. Med. Chem. 18, 4463—4470, http://dx.doi.org/10.1021/jm0303195.

Yildirim, M.A., Goh, K.-I., Cusick, M.E., Barabési, A.-L., Vidal, M., 2007. Drug-target network. Nat. Biotechnol. 10, 1119—1126, http://dx.doi.org/10.1038/nbt1338.