The relationship between rational drug design and drug side effects

Juan Wang, Zhi-xin Li, Cheng-xiang Qiu, Dong Wang and Qing-hua Cui

Abstract
Previous analysis of systems pharmacology has revealed a tendency of rational drug design in the pharmaceutical industry. The targets of new drugs tend to be close with the corresponding disease genes in the biological networks. However, it remains unclear whether the rational drug design introduces disadvantages, i.e. side effects. Therefore, it is important to dissect the relationship between rational drug design and drug side effects. Based on a recently released drug side effect database, SIDER, here we analyzed the relationship between drug side effects and the rational drug design. We revealed that the incidence drug side effect is significantly associated with the network distance of drug targets and disease genes. Drugs with the distances of three or four have the smallest incidence of side effects, whereas drugs with the distances of more than four or smaller than three show significantly greater incidence of side effects. Furthermore, protein drugs and small molecule drugs show significant differences. Drugs hitting membrane targets and drugs hitting cytoplasm targets also show differences. Failure drugs because of severe side effects show smaller network distances than approved drugs. These results suggest that researchers should be prudent on rationalizing the drug design. Too small distances between drug targets and disease genes may not always be advantageous for rational design for drug discovery.

Keywords: rational drug design; systems pharmacology; network distance; side effects; SIDER

INTRODUCTION
The discovery of effective and safe drugs is the ultimate goal of medicine and the pharmaceutical industry. In recent years, systems pharmacology, an emerging field of pharmacology, has greatly helped researchers in the identification of new drugs and drug targets, and in understanding the mechanisms of drugs [1, 2]. For example, by analyzing drug–target networks, Yildirim et al. [3] observed that the shortest path between drug targets and the disease genes in the biological networks for drugs approved since 1996 is significantly shorter than that for drugs before 1996. This indicates that the pharmaceutical industry moves toward the rationalization of drug design, that is, shortening the network distance between drug targets and disease genes in the biological networks. This shift suggests a tendency of more specific drugs for specific diseases [3].
However, whether rational drug design introduces disadvantages such as drug side effects remains largely unknown.

Side effects are human responses to drug treatment at the phenotypic level, and are increasingly important in pharmacology [4]. The side effects of drugs may provide some important information for researchers in pharmaceutical science and technology. For example, through a multiple-step network biology approach, the relationship between biological process and drug side effects can be discovered [5]. The similarity of side effects between two drugs could be an important metric to identify novel drug targets [6]. Similar with the above study, a recent report has revealed that neighbors of drug targets in a biological network contribute to drug side effect similarity [7]. These studies provide helps in the understanding of drug, target and side effect. Because pharmaceutical industry tend to rationalize the drug design, it therefore becomes important to investigate whether such move is associated with drug side effects, and to study whether there are optimal network distances between the drug targets and the disease genes in biological networks, based on the observation of the rational drug design and drug side effects.

In this study, we investigated the relationship between the network distance of drug targets and the disease genes in a human signaling network and drug side effects.

METHODS

Data of drug targets, disease genes and side effects

In this study, we obtained the data of drug targets and disease genes from the study of Yildirim et al. [3]. We downloaded the drug side effect data from SIDER, an online side effect resource (http://sideeffects.embl.de) [4]. Two files in SIDER were considered. One file contains side effects, SE1, extracted from the drug labels based on the Thesaurus of Adverse Reaction Terms. The other file contains euphoria-related side effects, SE2, extracted from Unified Medical Language System. The drugs that do not have recorded side effects in SIDER are called drugs without side effects in this study. The approved time of drug was obtained from DrugBank [8], through which we also discriminate protein drugs from small molecule drugs.

Human cellular signaling network and network analysis

In this study, we used the human signaling network to analyze the association of the drug target–disease gene network distance and drug side effects. We obtained the human cellular signaling network from the studies of Cui et al. [9, 10], which contains 1635 nodes and 5089 links. This human cellular signaling network is a general cellular signaling network. Cui et al. originally studied cancer gene mutation patterns based on this network. We implemented the Dijkstra algorithm for the shortest path identification in the human signaling network and calculated network distances based on the identified paths. Here, network distance is defined as the number of steps from the target to the corresponding disease gene. We further calculated the network distance for each pair of drug target and disease gene.

RESULTS AND DISCUSSION

The incidence of drug side effects is significantly associated with the distance between the drug targets and the disease genes in the human cellular signaling network

We first investigated the association of the drug side effects with the rational drug design. We initially calculated the network distances between the drug targets and disease genes based on the identified shortest paths in the human signaling network for all drugs. We then classified drugs with targets and disease genes into different groups according to network distance between target and disease genes. The total number of drugs whose targets and disease genes can be found in the network is 213. The number of drug target and disease gene pairs is 1199. We next clustered pairs of the drug targets and the diseases genes into different groups according to their network distances. Through drug–target mapping data, we know the drugs in each group and then can count corresponding drug numbers. By mapping side effect data to drugs, the number of drugs with at least one reported side effect in each group can be counted. An incidence of drug side effects in group $i$ (distance is $i$) was then calculated by formula (1) for the two types of side effects:

$$\text{Incidence } i = \frac{n_i}{N_i}$$

where $n_i$ is the number of drugs that have at least one reported side effect for all drugs in group $i$, and $N_i$ is
the total number of drugs in group $i$. As a result, at distances of three and four, the drugs show minimal incidence of side effects (Figure 1). Drugs with distances between their targets and disease genes smaller than three or bigger than four show significantly increasing incidence of side effects. We tested the significance of the lower incidence of drugs with distances of three and four by randomization test ($P = 0.0022$ for SE1; $P = 0.0004$ for SE2), which is performed by randomly permuting the original network distances and counting the number of network distances of three and four in the random case. For SE1, we further found that drugs with network distances of three and four have significantly smaller number of side effects than drugs with distances smaller than three (median number 69 versus 102, $P = 0.002$, Wilcoxon test) and have a little bit smaller but not significant number of side effects than drugs with distances greater than four (median 69 versus 71). For SE2, we found a smaller number of SE2 in drugs with distances of three and four than drugs with distances smaller than three and distances greater than four (median number 4 versus 4 versus 4, mean number 5.48 versus 5.76 versus 7.04, $P = 0.36$; $P = 0.02$, Wilcoxon test). From a large to small network distance between drug targets and disease genes, the incidence of drug side effects decreases with the decreasing distance from the origin, suggesting that rationalizing the drug design is advantageous considering the side effects. The incidence of drug side effects achieves the optimal point at the distances of three and four, and then dramatically increases as the network distance decreases smaller than three. SE1 and SE2 showed highly consistent patterns, with Spearman’s correlation between incidences of side effects at $R = 0.97$, $P = 6.55 \times 10^{-5}$. This suggests that the rational drug design can decrease the side effects when the network distance is big, but when distance is smaller than a cutoff distance (i.e. 3), a dramatically increasing incidence of side effect occurs. This should be considered by pharmacology researchers. Although the close distance between the drug targets and disease genes seems to improve the efficiency of drugs, this procedure may also introduce other problems, i.e. significantly increasing incidence of side effects for drugs with distances smaller than three. Although the reason why distance less than 3 leads to enrichment of incidences of drug side effects remains unclear, the location of disease genes in biological networks may provide clues. It is known that normally disease genes have more interactions than nondisease genes in biological networks [11]. Therefore, drugs with smaller distance of targets and disease genes will have greater effects on disease genes and therefore have more downstream effects, which may result in higher incidence of side effects.

The distribution analysis of network distance between the drug targets and the disease genes also shows that the network distances of three and four are the optimal distances in rational drug design

In the previous section, we revealed that drugs with a network distance of three or four between the drug targets and the disease genes show the optimal incidence of drug side effects. Drugs with distances smaller than three show dramatically higher incidence of side effects and drugs with distances larger than four also show a similar trend. To address this in detail, we analyzed the statistical distribution of the network.
distances for drugs with reported side effects (SE1/SE2) and drugs without reported side effects (non-SE1/non-SE2), respectively. The numbers of drug target and disease pairs for drugs with SE1 and SE2 are 691 and 608, respectively. The result shows that drugs have enriched network distances of three and four (Figure 2). We further tested the statistical significance of the lower incidence of drugs with distances of three and four by randomization test by comparing distribution densities at network distances of three and four for drugs with side effects and drugs without side effects in random cases ($P = 0.0002$ for SE1; $P = 0.0002$ for SE2). The density of SE1 (SE2) is less than that of non-SE1 (non-SE2) at the network distances of three and four, but higher than non-SE1 (non-SE2) at a network distance smaller than three. Two groups did not show differences at network distances larger than four. Both data sets show highly consistent results (Figure 2A and 2B), which suggests that the optimal distance for the rational drug design are the network distances of three and four based on the analysis of the human signaling network. Interestingly, drugs also show significantly higher fraction of distance of three and four in the human protein–protein interaction network [3]. We next investigated whether drugs approved recently (2002–11) show higher fraction of distances of three and four than drugs approved before 2002 (the starting year for the drugs approved before 2002 used in this study is 1992). As a result, both groups of drugs show significant higher fraction of distances of three and four. Furthermore, we did not find differences when considering the target is in upstream or downstream disease genes. In addition, we have investigated the relationship between network distance and side effect considering that drug targets are located in membrane and cytoplasm and found significant differences. For drugs hitting membrane proteins, the optimal distances seem to be two and three; whereas for drugs hitting cytoplasm proteins, the optimal distances seem to be three and four. We also examined the etiological and palliative drugs. We found that palliative drugs have greater network distances than etiological drugs ($P = 6.8 \times 10^{-5}$, Wilcoxon test), which is consistent with the result by Yildirim et al. [3]. Although two groups of drugs have the same median distance of 4, palliative drugs have a higher fraction of network distance four than etiological drugs. Therefore, two groups of drugs seem to show different optimal network distances.

**Failure drugs because of severe side effects have significantly smaller network distances than approved drugs with side effects**

The degree of severity of drug side effects is an important issue and should be considered in this study. Because the data or method to quantify the degree of severity is not available, we selected an alternative approach to address this issue. We manually curated some drugs that were failed in the clinical trial because of severe side effects from five publications (Supplementary File S1). We further curated the drug targets and disease genes as well. We found the distributions of network distances of drug targets and disease genes between the failure drugs and approved drugs with side effects used in this study are significantly different. Failure drugs have a significant shift to smaller network distances (mainly focus on network distances of two and three) than approved drugs ($P = 5.587 \times 10^{-12}$, Wilcoxon test; Supplementary Figure S1). This result suggests that
the very small network distance between the drug targets and the disease genes will not always do benefit to drug.

**Protein drugs have a higher fraction in distances three and four, and have lower incidence of side effects than small molecule drugs**

It is known that most of the drugs are small molecules and some drugs are proteins, e.g. insulin. It remains unknown whether there are differences in network distances and side effects between the two groups of drugs. In order to address this question, we first classified the drugs used in this study into two groups: protein drugs and small molecular drugs. We next investigated whether there is difference in small molecular drugs and protein drugs. We found that protein drugs have a higher fraction in distances three and four than small molecule drugs. Further analysis revealed that small molecule drugs have a significantly higher incidence of side effects than protein drugs. Over 68.9% of small molecule drugs have at least one reported side effect; whereas only 6.5% of protein drugs have at least one reported side effect ($P = 4.788 \times 10^{-15}$, Odds Ratio = 0.032, Fisher’s exact test, Figure 3). This result suggests that although small molecular drugs represent the majority of the drugs, it is necessary to increase the number of protein drugs at the viewpoint of drug side effects.

**DISCUSSION**

A previous study on systems pharmacology has revealed a trend in pharmaceutical industry to rationalize drug design, that is, the shortening of the network distance between drug targets and disease genes in the biological networks. In this study, we revealed that the incidence of drug side effects is significantly associated with the network distance. The incidence of side effects of drugs does not always decrease along with a decrease in the network distance. For network distances smaller than three, the incidence of side effects dramatically increases. Actually, the currently approved drugs indeed show enriched distribution in the network distances of three and four. This suggests that pharmacology researchers should be prudent in rationalizing the drug design. Very close distances between the drug targets and disease genes may not always be beneficial as they may even result in a higher incidence of side effects. Our study suggests two putatively optimal network distances, three and four, for the rational drug design. Finally, because of limited data, we did not perform the above analysis in the context of different drug classes. This issue should be considered when more data becomes available and more patterns are expected in the future.

**SUPPLEMENTARY DATA**

Supplementary data are available online at http://bib.oxfordjournals.org/.

**Key Points**

- The incidence of drug side effects is significantly associated with the network distance of drug targets and disease genes. Drugs with the distances of three or four have the smallest incidence of side effects, whereas drugs with the distances of more than four or smaller than three show significantly greater incidence of side effects.
- Protein drugs have a higher fraction in distances three and four, and have lower incidence of side effects than small molecular drugs.
- Failure drugs because of severe side effects have significantly smaller network distances than approved drugs.
- Too small distances between drug targets and disease genes may not always be advantageous to rational design for drug discovery.

**FUNDING**

These studies were supported by The National Natural Science Foundation of China (Grant No. 81001481).

**References**


