

# SCIENTIFIC REPORTS



OPEN

## Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations

Received: 19 January 2015

Accepted: 01 April 2015

Published: 11 May 2015

Aron R. Perez-Lopez<sup>1,\*</sup>, Kristóf Z. Szalay<sup>1</sup>, Dénes Türei<sup>2,†</sup>, Dezső Módos<sup>2,3</sup>, Katalin Lenti<sup>3</sup>, Tamás Korcsmáros<sup>2,4,5</sup> & Peter Csermely<sup>1</sup>

Network-based methods are playing an increasingly important role in drug design. Our main question in this paper was whether the efficiency of drug target proteins to spread perturbations in the human interactome is larger if the binding drugs have side effects, as compared to those which have no reported side effects. Our results showed that in general, drug targets were better spreaders of perturbations than non-target proteins, and in particular, targets of drugs with side effects were also better spreaders of perturbations than targets of drugs having no reported side effects in human protein-protein interaction networks. Colorectal cancer-related proteins were good spreaders and had a high centrality, while type 2 diabetes-related proteins showed an average spreading efficiency and had an average centrality in the human interactome. Moreover, the interactome-distance between drug targets and disease-related proteins was higher in diabetes than in colorectal cancer. Our results may help a better understanding of the network position and dynamics of drug targets and disease-related proteins, and may contribute to develop additional, network-based tests to increase the potential safety of drug candidates.

Due to the “curse of attrition” drug side effects are subjects of increasing concerns<sup>1–4</sup>. In recent years a growing number of side effect databases helped pharmacovigilance efforts<sup>2,5–10</sup>. In addition, the prediction of drug side effects was a subject of several excellent network studies. These contributions constructed and analyzed drug—side effect networks<sup>1,8,11</sup>, side effect similarity-based drug—drug networks<sup>12–14</sup>, drug target—side effect networks (including correlated drug binding profiles and side effect profiles and protein domain networks)<sup>3,5,7,15,16</sup>, as well as drug—side effect—biological pathway multi-layer networks<sup>9,10,17,18</sup>.

Parallel with the sequencing of the human genome, the pharmaceutical industry increasingly turned towards rational drug design, where drug target candidates are selected on the basis of known disease-related genes. In recent years, however, it became apparent that drug action often extends beyond its primary target, and also affects the neighbourhood of the primary target in molecular networks<sup>4,19–23</sup>. The influence on network neighbourhood can be efficiently modelled as a spreading process. Indeed, network spreading efficiency became increasingly used to characterize the dynamics of a wide variety of networks, such as the propagation of infections and computer viruses<sup>24–26</sup>, as well as the spread

<sup>1</sup>Department of Medical Chemistry, Semmelweis University, P.O. Box 260, H-1444 Budapest 8, Hungary.

<sup>2</sup>Department of Genetics, Eötvös Loránd University, Pázmány P. s. 1C, H-1117 Budapest, Hungary. <sup>3</sup>Department of Morphology and Physiology, Faculty of Health Sciences, Semmelweis University, Vas u. 17, H-1088 Budapest, Hungary. <sup>4</sup>TGAC, The Genome Analysis Centre, Norwich, UK. <sup>5</sup>Gut Health and Food Safety Programme, Institute of Food Research, Norwich, UK. <sup>\*</sup>Current address: Apáczai Csere János High School of the Eötvös Loránd University, Papnövelde u. 4-6, H-1053 Budapest, Hungary. <sup>†</sup>Current address: European Bioinformatics Institute(EMBL-EBI), Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. Correspondence and requests for materials should be addressed to P.C. (email: csermely.peter@med.semmelweis-univ.hu)

of information, innovations and social influence<sup>27–30</sup>. Long-range spread of conformational changes *via* protein-protein interaction networks is supported by several pieces of experimental evidence<sup>31,32</sup>. Moreover, recent studies extended the use of information-spread to molecular networks highlighting the usefulness of this approach in finding key amino acids of protein structure networks, biologically relevant changes of cellular functions upon stress, reprogramming biological networks, and uncovering the attractor changes in malignant transformation<sup>33–36</sup>. However, network spreading efficiency has been used to characterize drug targets neither in general, nor restricted to targets of drugs having side effects.

In this study we investigated, whether the efficiency of drug target proteins to spread perturbations in the human interactome is larger, if drugs targeting them have side effects, as compared to the spreading efficiency of targets of those drugs, which have no reported side effects. Encouraged by our findings that drug targets in general, and targets of drugs having side effects in particular, spread perturbation better in the human interactome than other proteins, we specifically examined two diseases, colorectal cancer and diabetes. These two, wide-spread diseases were selected, since they represent target groups of different drug design strategies<sup>4</sup>, and they had been the subjects of several former network-related studies<sup>37–45</sup>. We found that colorectal cancer-related proteins were good spreaders and had a high centrality in the human protein-protein interaction network. On the contrary, type 2 diabetes-related proteins showed an average spreading efficiency, and had an average centrality. Additionally, network shortest path (geodesic distance) between drug targets and disease-related proteins was higher in diabetes than in colorectal cancer. Our results give novel details on the network topology and dynamics of disease-related and drug target proteins, and may initiate the development of novel, network-based pharmacovigilance methods increasing the potential safety of drug candidates.

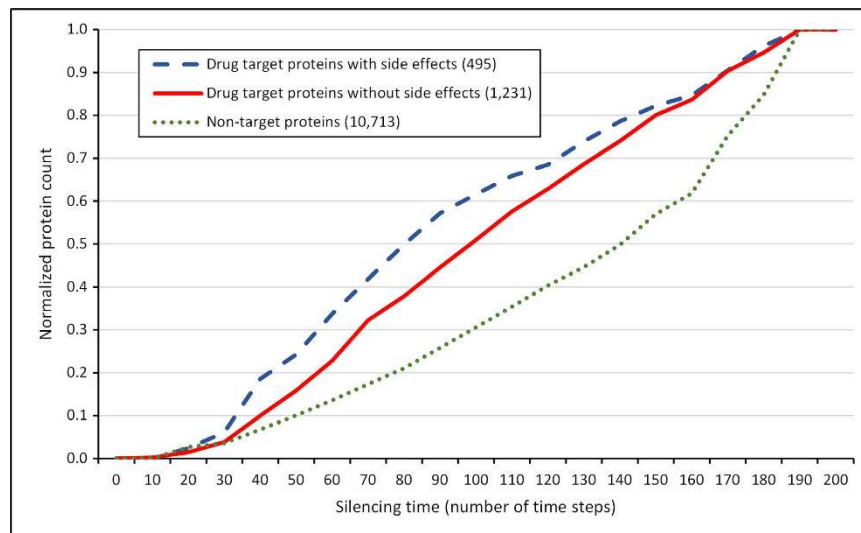
## Results

**Targets of drugs with side effects spread perturbations better in the human interactome than targets of drugs without side effects.** The initial working hypothesis of our research was that drugs having protein targets that better propagate changes in the human interactome may have a higher probability of causing side effects. This hypothesis is in agreement with earlier findings showing that the interactome neighbourhood contributed to drug side-effect similarity<sup>20</sup>. In order to test our hypothesis, we compared the propagation of perturbations started from drug targets with and without known side effect, as well as that of non-target proteins in the human protein-protein interaction network using the Turbine network dynamics software package developed earlier in our group<sup>35</sup>.

To compare the spreading efficiency of drug target proteins with and without side effects we ran a series of perturbation simulations on the human interactome using the Turbine programme<sup>35</sup>. We assembled a human interactome containing 12,439 proteins and 174,666 edges using the STRING database<sup>46</sup>, out of which 1,726 were target proteins of 3,626 human drugs obtained from the DrugBank database<sup>47</sup> and a total of 99,423 drug-side effect pairs from the SIDER database<sup>2</sup> were analysed as described in Methods in detail. Simulations were based on the communicating vessels network dynamics model tested earlier<sup>35</sup>, where changes from one protein to its neighbours ‘flow’ in proportion with the energy differences between the ‘source’ and the ‘target’ proteins. We examined a total of 495 target proteins of 597 drugs (Suppl. Table 1), which were reported to have side effects according to the SIDER database<sup>2</sup>. As control groups, we have also examined the 1,231 target proteins of the remaining 3,029 drugs having no reported side effects in the SIDER database<sup>2</sup>, as well as the remaining 10,713 proteins in our human interactome, which were not listed as drug targets in DrugBank<sup>47</sup>. For each selected protein target we calculated the silencing time, which is the number of time steps in the simulation needed for the initial perturbation to disappear completely due to dissipation. Small silencing time values were shown to be an efficient measure of large spreading efficiency of network nodes earlier<sup>35</sup>, since in this case the initial perturbation efficiently spreads in the network and it becomes dissipated fast.

Figure 1 shows the cumulative distribution of the normalized number of proteins having an increasing silencing time (thus decreasing perturbation efficiency). Targets of drugs with side effects had a significantly larger proportion of small silencing times (i.e. large spreading efficiency) than targets of drugs having no side effects (Mann-Whitney-Wilcoxon test,  $p = 1.677e-5$ ). Similarly, the proportion of targets of drugs without side effects having a small silencing time (i.e. large spreading efficiency) was significantly larger than that of human interactome proteins, which have not been reported as drug targets in DrugBank<sup>47</sup> (Mann-Whitney-Wilcoxon test,  $p = 2.2e-16$ ). Thus targets of drugs with side effects were found to be better spreaders of perturbations than targets of drugs having no reported side effects. Importantly, drug targets were also better spreaders of perturbations than non-target proteins.

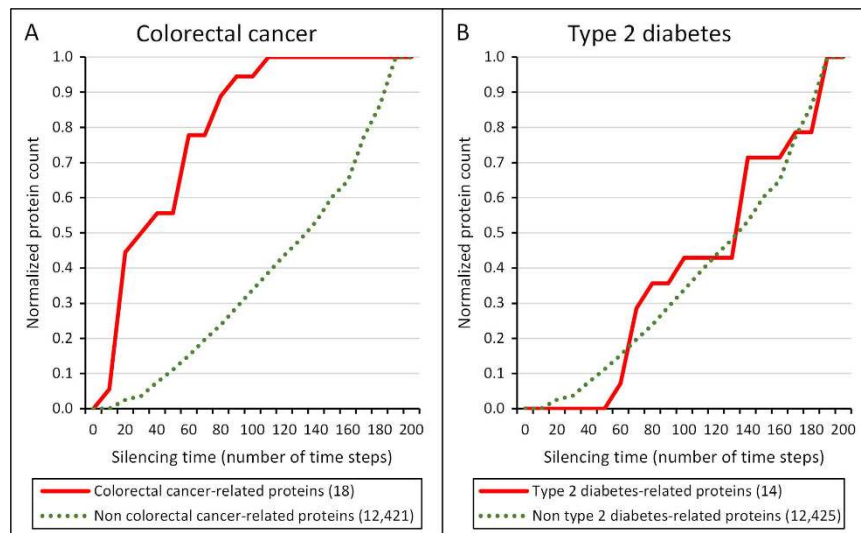
Simulations shown on Fig. 1 were run with a starting energy of 1,000 units and a dissipation value of 5 units. Being curious whether our result is robust for the variations of simulation parameters, we repeated these simulations using a starting energy of 10,000 and a dissipation of 1 or 5 units. Under these conditions we obtained very similar results (Suppl. Figs. 1 and 2) to those shown on Fig. 1. When we split the starting energy of 1,000 units equally among targets of multi-target drugs instead of examining each target protein alone as the source of perturbations, we were able to reproduce the same pattern (Suppl. Fig. 3) as that of Fig. 1. Furthermore, to test the robustness of the results against the choice of protein-protein interaction network, we randomly deleted 50% of the 12,439 proteins in our human interactome. Examining the spreading efficiency in the giant component of this truncated interactome we obtained very similar results (Suppl. Fig. 4) to those shown in Fig. 1.



**Figure 1. Cumulative silencing time distribution of drug targets and non-target proteins.** The diagram shows the cumulative distribution of the normalized number of proteins with given silencing times, which are drug targets with known side effects (blue dashed line), which are drug targets without known side effects (red solid line) and which are not drug targets (green dotted line). The number of proteins was normalized by dividing the number of proteins in each silencing time range by the total number of proteins allowing a better comparison. The total number of drug targets with and without side effects and non-target proteins was 495, 1,231 and 10,713, respectively. The human interactome containing 12,439 proteins and 174,666 edges was built from the STRING database<sup>46</sup>, 1,726 human drug targets were obtained from the DrugBank database<sup>47</sup> and 99,423 drug-side effect pairs were taken from the SIDER database<sup>2</sup>. Silencing times were calculated separately for every protein/drug target with the Turbine program<sup>35</sup> as described in the Methods section using a starting energy of 1,000 and a dissipation value of 5 units. Statistical analysis was performed using the Mann-Whitney (Wilcoxon rank sum) test function of the R package<sup>56</sup>. There was a statistically significant difference ( $p = 1.677e-5$ ) between the silencing times of drug targets with known side effects and the silencing times of drug targets without reported side effects. The difference between the silencing times of drug targets and non-target proteins was also statistically significant ( $p = 2.2e-16$ ).

Next we were curious whether the larger spreading efficiency of drug targets with side effects, as compared to drug targets without side effects or proteins having no reported drugs bound to them, is also shown by examining perturbation reach values. Perturbation reach values show the number of proteins, which received the perturbation from the initial perturbation source protein until the perturbation was dissipated from the system. Small perturbation reach values were shown to characterize small spreading efficiency in earlier studies<sup>35</sup>, since in this case the original perturbation reached only a small number of proteins before it became dissipated. Targets of drugs with side effects had a significantly smaller proportion of small perturbation reach values (i.e. small spreading efficiency) than that of targets of drugs having no side effects (Mann-Whitney-Wilcoxon test,  $p = 1.663e-5$ ; Suppl. Fig. 5). Similarly, the proportion of targets of drugs without side effects having a small perturbation reach value (i.e. small spreading efficiency) was significantly smaller than that of human interactome proteins, which have not been reported as drug targets in DrugBank<sup>47</sup> (Mann-Whitney-Wilcoxon test,  $p = 2.2e-16$ ; Suppl. Fig. 5). Using a starting energy of 10,000 but a dissipation of 1 instead of 5 units, or splitting this starting energy equally among targets of multi-target drugs, we obtained very similar results (Suppl. Figs. 6 and 7). These studies confirmed that drug targets are better spreaders of perturbations than non-target proteins, and also that targets of drugs with side effects are better spreaders of perturbations than targets of drugs having no reported side effects.

A qualitatively similar picture emerged, when we examined the spreading efficiency of target proteins of drugs against two diseases, colorectal cancer and type 2 diabetes (Suppl. Tables 2-6). We chose these two diseases, because they represent very well the target groups of different drug design strategies<sup>4</sup>, and they had been the subjects of several former network-related studies<sup>37-45</sup>. Drug targets of both diseases were found to be better spreaders of perturbations than non-target proteins (Suppl. Fig. 8;  $p = 3.367e-5$  and  $p = 5.88e-5$  for colorectal cancer and diabetes, respectively). There was a tendency showing that targets of drugs with side effects were better spreaders of perturbations than targets of drugs having no reported side effects both in colorectal cancer and in diabetes. However, due to the low number of identified drug targets having side effects (3 and 25, respectively), these latter differences were not statistically significant ( $p = 1$  and  $p = 0.2593$ , respectively).



**Figure 2. Cumulative silencing time distribution of colorectal cancer- and type 2 diabetes mellitus-related proteins, as well as proteins, which are not related to these diseases.** The diagram shows the cumulative distribution of the normalized number of proteins with given silencing times, which are related to the disease (red line), as well as those, which are not related to the disease (green dotted line); for colorectal cancer (Panel A) and type 2 diabetes (Panel B). The number of proteins was normalized by dividing the number of proteins in each silencing time range by the total number of proteins allowing a better comparison. The total number of colorectal cancer-related proteins and type 2 diabetes-related proteins in the human interactome was 18 and 14, respectively. The human interactome containing 12,439 proteins and 174,666 edges was built from the STRING database<sup>46</sup>. Colorectal cancer- and type 2 diabetes-related proteins were obtained from the Cancer Gene Census database<sup>48</sup> and from the article of Parchwani *et al.*<sup>49</sup>, respectively. Silencing times were calculated separately for every protein with the Turbine program<sup>35</sup> as described in the Methods section using a starting energy of 1,000 and a dissipation value of 5 units. Statistical analysis was performed using the Mann-Whitney (Wilcoxon rank sum) test function of the R package<sup>56</sup>. There was a statistically significant difference between the silencing times of disease-related and non-related proteins in case of colorectal cancer ( $p=2.329e-9$ ) and but there was none in case of type 2 diabetes ( $p=0.8343$ ).

**Colorectal cancer-related proteins are good spreaders of perturbations and have a high centrality, while type-2 diabetes-related proteins show an average spreading efficiency and average centrality.** Very importantly, a rather interesting difference emerged, when we examined the spreading efficiency of proteins related to colorectal cancer and diabetes. Mutated genes and their corresponding proteins in colorectal cancer and in type-2 diabetes were obtained from the Cancer Gene Census database<sup>48</sup> (Suppl. Table 7) and from the article of Parchwani *et al.*<sup>49</sup> (Suppl. Table 8), respectively. In case of colorectal cancer, disease-associated proteins were found to be significantly better spreaders than the residual proteins of the human interactome. On the contrary, diabetes-related proteins showed indistinguishable spreading properties to the rest of human proteins, which were not associated with the onset of diabetes (Fig. 2). To test the robustness of the results against the choice of protein-protein interaction network, we randomly deleted 50% of the 12,439 proteins in our human interactome. Here again, colorectal cancer-associated proteins were found to be significantly better spreaders than the residual proteins of the human interactome (data not shown;  $p=0.00021$  in Mann-Whitney test) and spreading efficiency of diabetes-related proteins showed no significant difference as compared to the rest of human proteins (data not shown;  $p=0.095$  in Mann-Whitney test).

These findings are in agreement with earlier results showing that cancer-associated proteins are enriched in proteins having a high centrality in the human interactome<sup>37,38,40,42-45</sup>. Indeed, in our human interactome, cancer-related proteins had a significantly higher degree, closeness and betweenness centralities than diabetes-related proteins, having a 9.6-, 1.2- and 54-fold increase, respectively (Table 1). In agreement with their similar silencing time values (Suppl. Fig. 8), drug targets without or with side effects showed no significant centrality differences in the human interactome (Suppl. Table 9).

**The interactome distance between drug targets and disease-related proteins is higher in diabetes than in colorectal cancer.** Encouraged by the results showing an increased centrality of cancer-related, but not of diabetes-related proteins in the human interactome, we examined the interactome geodesic distance (i.e. shortest path) between drug targets and disease related proteins in both diseases using the neighbourhood matrices of related proteins. Our data show that the geodesic distance

Centrality type	Disease-related proteins			Proteins, which are not related to any of the two diseases		
	Colorectal cancer	Type 2 diabetes	Statistical difference between cancer- and diabetes-related proteins	Centrality value	Statistical difference from values of cancer-related proteins	Statistical difference from values of diabetes-related proteins
Degree (number of neighbours)	159.5	9.000	7.09e-5	9.000	2.58e-9	0.830
Closeness centrality (1/edge)	0.357	0.294	3.46e-5	0.277	1.90e-10	0.122
Betweenness centrality (fraction of shortest paths passing through the node)	2.55e-3	1.16e-5	1.24e-4	1.34e-5	3.23e-9	0.922

**Table 1.** Average human interactome centralities of proteins related to colorectal cancer and type 2 diabetes. The table shows the medians of the centralities of proteins related to colorectal cancer and type 2 diabetes (results were very similar, if instead of medians we used their arithmetic means; data not shown). The total number of colorectal cancer- and type 2 diabetes-related proteins was 18 and 14, respectively. Centrality values were calculated with the Pajek programme<sup>57</sup>. The human interactome containing 12,439 proteins and 174,666 edges was built from the STRING database<sup>46</sup>. Colorectal cancer-related proteins were obtained from the Cancer Gene Census database<sup>48</sup>, type 2 diabetes-related proteins were obtained from the article of Parchwani *et al.*<sup>49</sup>. Statistical analysis was performed using the Wilcoxon rank sum (Mann-Whitney) test function of the R package<sup>56</sup>.

Protein group	Average network distance from disease-related proteins(edges)
24 drug targets without known side effects used in the treatment of colorectal cancer	2.528
3 drug targets with known side effects used in the treatment of colorectal cancer	2.389
14 drug targets without known side effects used in the treatment of type 2 diabetes	3.250*
25 drug targets with known side effects used in the treatment of type 2 diabetes	3.234**

**Table 2.** Average network distance of drug targets without and with known side effects used in the treatment of colorectal cancer and type 2 diabetes from the disease-associated proteins. \*This value is significantly greater than the average network distance of drug targets without known side effects in colorectal cancer ( $p = 1.062e-05$ ). Statistical analysis was performed using the Welch (Student's) two sample t-test function of the R package<sup>56</sup>. \*\*This value is significantly greater than the average network distance of drug targets with known side effects in colorectal cancer ( $p = 0.005441$ ). Statistical analysis was performed using the Welch (Student's) two sample t-test function of the R package<sup>56</sup>. The table shows the arithmetic mean of the average network distance between drug targets (with and without known side effects used in the treatment of colorectal cancer and type 2 diabetes) and the proteins related to the respective disease (results were very similar, if instead of arithmetic means we used the medians; data not shown). The total number of colorectal cancer- and diabetes-related proteins in the human interactome were 18 and 14, respectively. Average network distances were calculated as shortest paths using the Pajek programme<sup>58</sup>. Proteins were labelled by their UniProt ID<sup>54</sup>. Human interactome containing 12,439 proteins and 174,666 edges was built from the STRING database<sup>46</sup>, 1,726 human drug targets were obtained from the DrugBank database<sup>47</sup> and 99,423 drug-side effect pairs were taken from the SIDER database<sup>2</sup>. Colorectal cancer- and type 2 diabetes-related proteins were obtained from the Cancer Gene Census database<sup>48</sup> and from the article of Parchwani *et al.*<sup>49</sup>, respectively. We used the mean values and the t-test because of the near-normal distribution of the average network distances.

in the human interactome between drug targets and disease-related proteins is significantly larger in case of type-2 diabetes than in colorectal cancer (targets without side effects:  $p = 1.062e-5$ ; targets with side effects:  $p = 5.441e-3$ ). (Table 2; Suppl. Tables 10-13 and Suppl. Fig. 9) This finding is supported by the visual representation of the human sub-interactome of drug target and disease-related proteins of these two diseases (Suppl. Fig. 10), where drug targets and disease-related proteins of colorectal cancer



are intertwined, while these two groups of proteins remain rather separated in type-2 diabetes. This observation is further substantiated by the fact, that only 1 of the 18 colorectal cancer-related proteins (6%) is not connected to the giant component of the sub-interactome, while 10 of the 14 diabetes-related proteins (71%) are missing from the same giant component (Suppl. Fig. 10).

## Discussion

The most important finding of our study is that 1.) drug targets are better spreaders of perturbations in the human interactome than non-target proteins in general; and in particular, 2.) targets of drugs with side effects are also better spreaders of perturbations than targets of drugs having no reported side effects (Fig. 1). These findings were robust, since they could be reproduced when we used different perturbation parameters (Suppl. Figs. 1,2 and 3), different measures of perturbation spread (Suppl. Figs. 5, 6 and 7), and reduced the size (coverage) of the human interactome to half of the original (Suppl. Fig. 4). These results are in agreement with those of a previous study showing that the interactome neighbourhood contributed to side-effect similarity<sup>20</sup>.

Importantly, colorectal cancer-related proteins are good spreaders of perturbations and had a high centrality, while type-2 diabetes-related proteins showed an average spreading efficiency and had an average centrality in the human interactome (Fig. 2 and Table 1). These findings are in agreement with earlier results showing that cancer-associated proteins are enriched in hubs, bottlenecks and bridges all having a high centrality in the human interactome<sup>37,38,40,42–45</sup>.

Furthermore, the interactome-distance between drug targets and disease-related proteins was higher in diabetes than in colorectal cancer (Table 2; Suppl. Tables 10–13 and Suppl. Fig. 9). This finding is in agreement with both the results of previous studies and intuitive insights on the classification of drug target strategies<sup>4</sup>. Most drug targets are 3 or 4 steps away in the human interactome from proteins involved in the same disease<sup>50</sup>. Moreover, cancer-related and metabolic disease-related proteins were shown to have an average network distance to the related drug targets of 2.3 and ~5 network edges, which are smaller and higher than the most abundant distance values, respectively, forming the two extremes of the distance-spectrum<sup>50</sup>. The former value is in the range we found in our study (Table 2). The latter value of a disease group containing diabetes is much larger than that related to cancer, which is again in agreement with our findings. As a general trend, rapidly proliferating cells, like those in cancer, are attacked at their central proteins, while differentiated cells, such as those involved in type-2 diabetes, are attacked at the neighbours of central proteins<sup>4</sup>. These assumptions are also in agreement with a smaller network distance of centrally positioned cancer-related proteins from centrally positioned cancer drug targets than the distance between the more peripheral diabetes-related proteins and drug targets.

Analysis of perturbation spread in molecular networks may be used to develop additional, network-based tests to increase the potential safety of drug candidates. Assessment of perturbation spread in weighted networks (where the edges are weighted according to the abundance of their end-node proteins of relevant tissues, e.g. the endothelial cell in colorectal cancer, as well as hepatocyte and myocyte in diabetes, as described in our earlier study for the yeast interactome<sup>51</sup>), directed networks (such as signalling networks<sup>4,52</sup>), or networks considering the subcellular localization of participating proteins<sup>53</sup>, as well as using quantitative measures of side-effect severity and abundance may provide additional information and will be subjects of later studies.

In summary, our results contributed to a better understanding of the network position and dynamics of disease-related and drug target proteins. The findings may help the future development of novel, network-based pharmacovigilance methods increasing the potential safety of drug candidates.

## Methods

**Construction of the human protein-protein interaction network.** In this paper, we examined the propagation of perturbations in the human protein-protein interaction network (interactome). The choice of this type of network was driven by the fact that it contains the most proteins and the greatest number of connections (as opposed to signalling networks or regulatory networks). Human interactome data were downloaded from the STRING database<sup>46</sup> on 8 February, 2013. STRING contains interaction data based on a vast number of data collection principles. We have only used manually collected ('database' column) or experimental ('experiments' column) data having higher reliability than e.g. predicted data. Only human protein-protein interactions were included in the interactome. In order to facilitate the comparison with drug targets, the STRING Ensemble Protein ID (ENSP) protein codes were translated to UniProt ID<sup>54</sup> using the UniProt translator. From the original 13,484 ENSP IDs we managed to translate 12,493 to UniProt IDs, but only 12,439 proteins were connected to other proteins. The database contained a total of 377,920 human protein-protein interactions, out of which 350,528 remained after translating the protein IDs to UniProt IDs using the UniProt translator, which were further reduced to 174,666 after eliminating multiple links and loops (self-links). The original STRING database also contained edge weights indicating the reliability of data. Since we only worked with manually collected and experimental data, our interactome contained no edge weights.

**Measurement of the propagation of perturbations in the human interactome.** The propagation of perturbations in the human interactome was measured with the network perturbation analysis software for simulating network dynamics called Turbine<sup>35</sup>. For the simulation experiments we chose the

software's communicating vessels model<sup>35</sup>, where changes from one protein to its neighbours 'flow' in proportion with the energy differences between the 'source' and the 'target' proteins. The communicating vessels model<sup>35</sup> contains a starting energy ( $E$ ) and a dissipation parameter ( $D$ ), where the starting energy is distributed equally among the proteins of the human interactome specified at the individual simulations, while in each step of the simulation the program subtracts  $D$  units of energy from each protein of the interactome. In most simulations  $E$  and  $D$  were set to 1000 and 5 units, respectively. Having these starting energy and dissipation parameters it was possible to trace the propagation of perturbations in the network rather easily. However, all the key simulations were also examined using different  $E$  and  $D$  values to examine the robustness of the results. To characterise the propagation efficiency of the starting node(s), the measure of silencing time<sup>35</sup> was used, which is the time elapsed from the start of the simulation until the energy of all nodes reaches the minimum threshold of less than 1 unit. We also calculated perturbation reach values<sup>35</sup>, which show the number of proteins receiving the perturbation from the initial perturbation source protein until the perturbation was dissipated from the system.

**Characterisation of drug side effects.** Drug side effects were collected from the SIDER database<sup>2</sup>. This database contains information about drug side effects and their frequencies from public documentation and package inserts, with the help of drug labels and terms from MedDRA (Medical Dictionary for Regulatory Activities). SIDER data were downloaded from the version of 17 October, 2012. This version of the SIDER database<sup>2</sup> contained 996 drugs, 4,192 unique side effects and 215,850 drug-side effect pairs. After eliminating the duplicates, 99,423 drug-side effect pairs remained. In order to be able to compare data, we converted drug IDs in the SIDER database<sup>2</sup> into IDs of the DrugBank database<sup>47</sup> by matching the drug names.

**Characterisation of drug targets.** We collected drug targets from the DrugBank database<sup>47</sup> version last updated on 10 February, 2013. The XML version of the database was used, including the drug names, indications and target list. The proteins in the target list were identified by their UniProt IDs<sup>54</sup> with the help of the external reference table available in the database. From the drug target list only those drugs that targeted human proteins were selected. From the original 6,718 drugs 3,926 such drugs were found, of which 3,626 had target proteins contained in our human interactome.

After comparison with the drug—side effect data from the SIDER database<sup>2</sup>, we found that 597 drugs (with a total of 495 target proteins) had known side effects, while the remaining 3,029 drugs (with 1,231 target proteins) had no reported side effects to date.

**Protein and drug target data related to the two examined diseases: colorectal cancer and type 2 diabetes.** Genes involved in colorectal cancer were collected from the Cancer Gene Census<sup>48</sup> database, by selecting those proteins in the entire database that contained the word 'colorectal' in their 'Tumour Types' column. Genes related to type 2 diabetes were obtained from the article of Parchwani *et al.*<sup>49</sup>. The 18 genes involved in colorectal cancer and the 46 genes related to type 2 diabetes were then mapped to proteins marked by UniProt ID<sup>54</sup> with the help of the Protein Identifier Cross-Reference (PICR)<sup>55</sup> application. See Suppl. Tables 7 and 8 for the genes and their respective proteins involved in the two diseases. From these proteins, all 18 colorectal cancer-related but only 14 type 2 diabetes-related were contained in our interactome. Drugs used in treatment of colorectal cancer and diabetes and their drug targets were collected based on the drug indications in the DrugBank database<sup>47</sup>. See Suppl. Table 2 for the relevant keywords used. We found 11 drugs against colorectal cancer and 36 against type 2 diabetes, which all had valid targets. Drugs against colorectal cancer and type 2 diabetes had 33 and 42 target proteins, respectively, out of which 27 and 39, respectively, were contained in our human interactome.

**Other methods.** A number of Bash shell scripts were written to automate the network simulation experiments with Turbine. Statistical analysis of the results was performed with the R software package<sup>56</sup>. The Pajek software<sup>57</sup> was used to measure geodesic distances and centralities in the human interactome, the Cytoscape software<sup>58</sup> was used to create images of the human interactome and the Inkscape software<sup>59</sup> was used to create some other images.

## References

- Fliri, A. F., Loging, W. T., Thadeio, P. F. & Volkmann, R. A. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat. Chem. Biol.* **1**, 389–397 (2005).
- Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J. & Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **6**, 343 (2010).
- Lounkine, E. *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **486**, 361–367 (2012).
- Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery. *Pharmacol. Ther.* **138**, 333–408. (2013).
- Yang, L., Luo, H., Chen, J., Xing, Q. & He, L. SePreSA: a server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a chemical-protein interactome. *Nucleic Acids Res.* **37**, W406–W412 (2009).
- Yang, L., Xu, L. & He, L. A CitationRank algorithm inheriting Google technology designed to highlight genes responsible for serious adverse drug reaction. *Bioinformatics* **25**, 2244–2250 (2009).
- Luo, H. *et al.* DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical-protein interactome. *Nucleic Acids Res.* **39**, W492–W498 (2011).

8. Oprea, T.I. *et al.* Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. *Mol. Inform.* **30**, 100–111 (2011).
9. Lopes, P. *et al.* Gathering and exploring scientific knowledge in pharmacovigilance. *PLoS ONE* **8**, e83016 (2013).
10. Oliveira, J. L. *et al.* The EU-ADR Web Platform: delivering advanced pharmacovigilance tools. *Pharmacoepidemiol. Drug Saf.* **22**, 459–467 (2013).
11. Garten, Y., Tatonetti, N. P. & Altman, R. B. Improving the prediction of pharmacogenes using text-derived drug-gene relationships. *Pac. Symp. Biocomput.* 305–314 (2010).
12. Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J. & Bork, P. Drug target identification using side-effect similarity. *Science* **321**, 263–266 (2008).
13. Yamanishi, Y., Kotera, M., Kanehisa, M., & Goto, S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* **26**, i246–i254 (2010).
14. Takarabe, M., Okuda, S., Itoh, M., Tokimatsu, T., Goto, S. & Kanehisa, M. Network analysis of adverse drug interactions. *Genome Inform.* **20**, 252–259 (2008).
15. Mizutani, S., Pauwels, E., Stoven, V., Goto, S. & Yamanishi, Y. Relating drug-protein interaction network with drug side effects. *Bioinformatics* **28**, i522–i528 (2012).
16. Iwata, H., Mizutani, S., Tabei, Y., Kotera, M., Goto, S. & Yamanishi, Y. Inferring protein domains associated with drug side effects based on drug-target interaction network. *BMC Syst. Biol.* **7**, S18 (2013).
17. Lee, S., Lee, K. H., Song, M. & Lee, D. Building the process-drug-side effect network to discover the relationship between biological processes and side effects. *BMC Bioinformatics* **12**, S2 (2011).
18. Bauer-Mehren, A. *et al.* Automatic filtering and substantiation of drug safety signals. *PLoS Comput. Biol.* **8**, e1002457 (2012).
19. Schwartz, J. M. & Nacher, J. C. Local and global modes of drug action in biochemical networks. *BMC Chem. Biol.* **9**, 4 (2009).
20. Brouwers, L., Iskar, M., Zeller, G., van Noort, V. & Bork, P. Network neighbors of drug targets contribute to drug side-effect similarity. *PLoS ONE* **6**, e22187 (2011).
21. Nussinov, R., Tsai, C.-J. & Csermely, P. Allo-network drugs: harnessing allosterity in cellular networks. *Trends Pharmacol. Sci.* **32**, 686–693 (2011).
22. Wang, J., Li, Z.-X., Qiu, C.-X., Wang, D. & Cui, Q.-H. The relationship between rational drug design and drug side effects. *Brief. Bioinform.* **13**, 377–382 (2012).
23. Nacher, J. C. & Schwartz, J. M. Modularity in protein complex and drug interactions reveals new polypharmacological properties. *PLoS ONE* **7**, e30028 (2012).
24. Hu, H., Myers, S., Colizza, V. & Vespignani, A. WiFi networks and malware epidemiology. *Proc. Natl. Acad. Sci. USA* **106**, 1318–1323 (2009).
25. Wang, P., González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding the spreading patterns of mobile phone viruses. *Science* **324**, 1071–1076 (2009).
26. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
27. Zanette, D. H. Critical behavior of propagation on small-world networks. *Phys. Rev. E* **64**, 050901 (2001).
28. Valente, T. W. Network interventions. *Science* **337**, 49–53 (2012).
29. Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. The diffusion of microfinance. *Science* **341**, 1236498 (2013).
30. Aral, S. & Walker, D. Identifying influential and susceptible members of social networks. *Science* **337**, 337–341 (2012).
31. Bray, D. & Duke, T. Conformational spread: the propagation of allosteric states in large multiprotein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 53–73 (2004).
32. Antal, M. A., Böde, C. & Csermely, P. Perturbation waves in proteins and protein networks: applications of percolation and game theories in signaling and drug design. *Curr. Protein Pept. Sci.* **10**, 161–172 (2009).
33. Stojmirović, A., Bliskovsky, A. & Yu, Y. K. CytoITMprobe: a network information flow plugin for Cytoscape. *BMC Res. Notes* **5**, 237 (2012).
34. Corneliussen, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. *Nat. Commun.* **4**, 1942 (2013).
35. Szalay, K. Z. & Csermely, P. Perturbation centrality and Turbine: A novel centrality measure obtained using a versatile network dynamics tool. *PLoS ONE* **8**, e78059 (2013).
36. Szalay, K. Z., Nussinov, R. & Csermely, P. Attractor structures of signaling networks: Consequences of different conformational barcode dynamics and their relations to network-based drug design. *Mol. Info.* **33**, 463–468 (2014).
37. Jonsson, P.F. & Bates, P.A. Global topological features of cancer proteins in the human interactome. *Bioinformatics* **22**, 2291–2297 (2006).
38. Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, 140 (2007).
39. Hase, T., Tanaka, H., Suzuki, Y., Nakagawa, S. & Kitano, H. Structure of protein interaction networks and their implications on drug design. *PLoS Comput. Biol.* **5**, e1000550 (2009).
40. Taylor, I.W. *et al.* Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnol.* **27**, 199–204 (2009).
41. Sharma, A., Chavali, S., Tabassum, R., Tandon, N. & Bharadwaj, D. Gene prioritization in type 2 diabetes using domain interactions and network analysis. *BMC Genomics* **11**, 84 (2010).
42. Sun, J. & Zhao, Z. A comparative study of cancer proteins in the human protein-protein interaction network. *BMC Genomics* **11**, S5 (2010).
43. Rosado, J. O., Henriques, J. P., & Bonatto, D. A systems pharmacology analysis of major chemotherapy combination regimens used in gastric cancer treatment: predicting potential new protein targets and drugs. *Curr. Cancer Drug Targets* **11**, 849–869 (2011).
44. Xia, J., Sun, J., Jia, P. & Zhao, Z. Do cancer proteins really interact strongly in the human protein-protein interaction network? *Comput. Biol. Chem.* **35**, 121–125 (2011).
45. Serra-Musach, J. *et al.* Cancer develops, progresses and responds to therapies through restricted perturbation of the protein-protein interaction network. *Integr. Biol.* **4**, 1038–1048 (2012).
46. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815 (2012).
47. Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for “omics” research on drugs. *Nucleic Acids Res.* **39**, D1035–D1041 (2011).
48. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2010).
49. Parchwani, D., Murthy, S., Upadhyah, A. & Patel, D. Genetic factors in the etiology of type 2 diabetes: linkage analyses, candidate gene association, and genome-wide association – still a long way to go! *Natl. J. Physiol. Pharm. Pharmacol.* **3**, 57–68 (2013).
50. Yildirim, M. A., Goh, K.-L., Cusick, M. E., Barabási, A.-L. & Vidal, M. Drug-target network. *Nat. Biotechnol.* **25**, 1119–1126 (2007).



51. Mihalik, Á. & Csermely, P. Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. *PLoS Comput. Biol.* **7**, e1002187 (2011).
52. Fazekas, D. *et al.* SignaLink 2 – A signaling pathway resource with multi-layered regulatory networks. *BMC Systems Biology* **7**, 7 (2013).
53. Veres, D. *et al.* (2015) ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* **43**, D485–D493 (2015).
54. The UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71–D75 (2012).
55. Wein, S. P. *et al.* Improvements in the Protein Identifier Cross-Reference service. *Nucleic Acids Res.* **40**, W276–W280 (2012).
56. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, : (R Foundation for Statistical Computing, Available: <http://www.R-project.org/>) (2013).
57. Bagatelj, V. & Mrvar, A. Pajek - Analysis and Visualization of Large Networks. in *Graph drawing software. Mathematics and visualization*. (eds Jünger, M. & Mutzel, P.) 77–103 (Springer, Berlin, 2003).
58. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
59. The Inkscape Team. Inkscape. (2014). at <http://inkscape.org>

## Acknowledgments

We thank members of the LINK-Group ([www.linkgroup.hu](http://www.linkgroup.hu)) for helpful discussions. This work was supported by the Hungarian Scientific Research Fund [OTKA K83314]. T.K. was a grantee of the János Bolyai Scholarship of the Hungarian Academy of Sciences, and is supported by a fellowship in computational biology at The Genome Analysis Centre, in partnership with the Institute of Food Research, and strategically supported by BBSRC.

## Author Contributions

P.C. initiated the project and conceived the research. A.R.P.L. performed all simulations and data analysis. D.T. and D.M. contributed in the assembly of databases. All (A.R.P.L., K.Z.S., D.T., D.M., K.L., T.K. P.C.) authors contributed to biological interpretation of the results. A.R.P.L. prepared the tables and figures. A.R.P.L. and P.C. wrote the manuscript text. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Perez-Lopez, R. *et al.* Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. *Sci. Rep.* **5**, 10182; doi: 10.1038/srep10182 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>