Drug–target network

Muhammed A Yıldırım[1,2,3], Kwang-Il Goh[1,4,5], Michael E Cusick[1,2], Albert-László Barabási[1,4,6] & Marc Vidal[1,2]

The global set of relationships between protein targets of all drugs and all disease-gene products in the human protein–protein interaction or 'interactome' network remains uncharacterized. We built a bipartite graph composed of US Food and Drug Administration–approved drugs and proteins linked by drug–target binary associations. The resulting network connects most drugs into a highly interlinked giant component, with strong local clustering of drugs of similar types according to Anatomical Therapeutic Chemical classification. Topological analyses of this network quantitatively showed an overabundance of 'follow-on' drugs, that is, drugs that target already targeted proteins. By including drugs currently under investigation, we identified a trend toward more functionally diverse targets improving polypharmacology. To analyze the relationships between drug targets and disease-gene products, we measured the shortest distance between both sets of proteins in current models of the human interactome network. Significant differences in distance were found between etiological and palliative drugs. A recent trend toward more rational drug design was observed.

The pharmaceutical industry has historically relied upon particular families of 'druggable' proteins against which chemists attempt to develop compounds with desired actions[1–4]. Most drugs act by binding to specific proteins, thereby changing their biochemical and/or biophysical activities, with multiple consequences on various functions. Yet most US Food and Drug Administration (FDA)-approved drugs currently used by clinicians were developed without knowledge of the molecular mechanisms responsible for their indicated diseases[5].

Proteins rarely function in isolation in and outside the cell; rather, proteins operate as part of highly interconnected cellular networks referred to as interactome networks[6–9]. Our goal here is to understand drug targets in the context of cellular and disease networks. Molecular and genetic studies of disease over recent decades have produced an impressive list of gene–disease associations[10–12]. We combined the tools of network biology with systematic information about drugs and their targets to (i) analyze properties of drug-target networks as part of cellular networks, (ii) assess retrospectively and prospectively network-based relationships between drugs and their targets, quantifying ongoing trends and shifts in drug discovery, and (iii) quantify interrelationships between drug targets and disease-gene products.

[1]Center for Cancer Systems Biology (CCSB) and [2]Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, 44 Binney St., Boston, Massachusetts 02115, USA. [3]Division of Engineering and Applied Sciences, Harvard University, 9 Oxford St., Cambridge, Massachusetts 02138, USA. [4]Center for Complex Network Research and Department of Physics, University of Notre Dame, 225 Nieuwland Science Hall, Notre Dame, IN 46556, USA. [5]Department of Physics, Korea University, Seoul 136-713, Korea. [6]Center for Complex Network Research, and the Departments of Physics, Computer Science and Biology, Northeastern University, 110 Forsyth Street, 111 Dana Research Center, Boston, MA 02115. Correspondence should be addressed to: A.-L.B. (alb@nd.edu) or M.V. (marc_vidal@dfci.harvard.edu).

Published online 5 October 2007; doi:10.1038/nbt1338

RESULTS

Lists of drugs and corresponding targets were obtained from the DrugBank database[13]. As of March 29, 2006, DrugBank contained 4,252 drug entries, including 1,178 FDA-approved drugs (1,065 small molecules and 113 proteins/peptides) and 3,074 drugs under investigation ("experimental drugs")(**Fig. 1**). The FDA-approved drugs target 394 human proteins in total. Most drugs target only a few proteins, but some have many targets (**Fig. 1a**). The average number of target proteins per drug is 1.8. Likewise, many proteins are targeted by more than one drug (**Fig. 1b**). An analysis of the chemical similarity between drugs targeting these proteins reveals that most of the drugs have a distinct chemical structure (**Supplementary Notes** and **Supplementary Fig. 1** online).

Generating a drug–target network

First we used all known FDA-approved drugs and their targets to generate a bipartite graph of drug–protein interactions in which a drug and a protein are connected to each other if the protein is a known target of the drug, giving rise to a 'drug–target network' (DT network) (**Fig. 2**). From the bipartite DT network graph, we generated two biologically relevant network projections. In the 'drug network', nodes represent drugs, and two drugs are connected to each other if they share at least one target protein (**Supplementary Fig. 2** online). In the complementary 'target-protein network' (TP network), nodes are proteins, and two proteins are connected if they are both targeted by at least one common drug (**Fig. 3a**). By itself, network visualization of drug–protein associations provides an important survey of the current status of drug discovery. We next used quantitative tools to discern global trends encoded in these maps.

If most drugs specifically targeted a single protein, then the drug network would consist of isolated nodes with few or no edges between them. Instead, the drug network displays many connections between different drugs and drug classes. Out of 890 approved drugs with known human protein targets, 788 have at least one link to other drugs, that is, they share targets with other drugs. There are 476 drugs in the giant component, the largest connected component of the network (**Supplementary Fig. 2**). We colored drug nodes according to the Anatomical Therapeutic Chemical (ATC) classification. Although the drug-network layout was generated independently of any knowledge about drug classes, the resulting network is naturally and visibly clustered by major therapeutic classes. The most obvious example of clustering is a large tightly

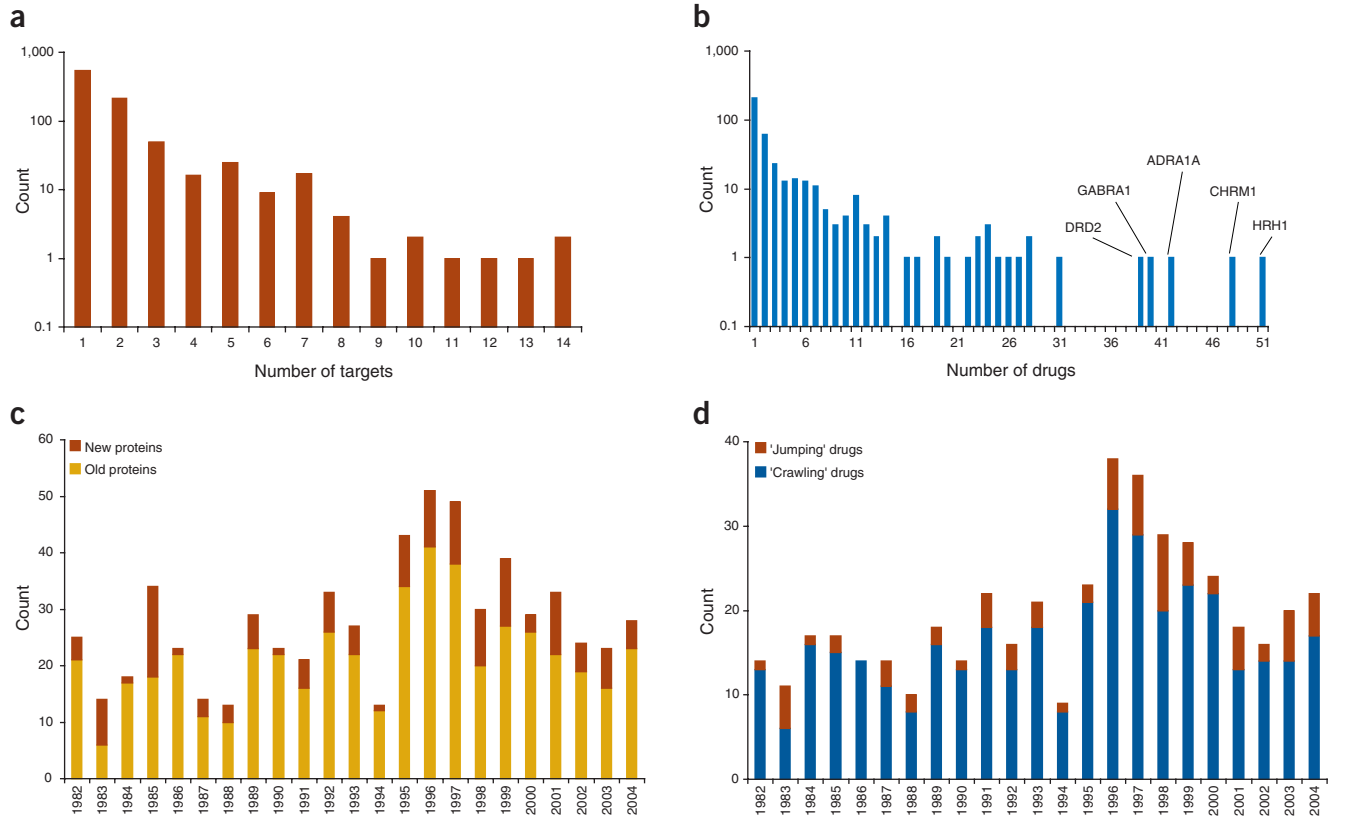

Figure 1 Distribution of drugs and drug targets. (**a**) Distribution of drugs with respect to number of their targets. The FDA-approved drugs target 394 human proteins in total. Most drugs target only a few proteins, but some have many targets; for example, propiomazine (Largon) and promazine (Sparine) have 14 targets each, and olanzapine (Zyprexa, Zydis) and ziprasidone (Geodon) have 11 targets each. (**b**) Distribution of target proteins with respect to number of times the target protein is targeted by a distinct drug. The most-targeted proteins are the histamine H1 receptor (HRH1) (targeted by 51 drugs), the muscarinic 1 cholinergic receptor (CHRM1) (48 drugs), the α1A adrenergic receptor (ADRA1A) (42 drugs) and the dopamine receptor D_2 (DRD2) (40 drugs). (**c**) Number of distinct proteins targeted each year. 'New proteins' are the newly introduced proteins. (**d**) New drugs introduced each year. 'Jumping drugs' are the drugs with totally new sets of target proteins, whereas 'Crawling drugs' target at least one already targeted protein. The FDA approved 19.5 new chemical entities on average each year in the last 25 years, of which 6.3 act on novel targets.

interconnected neurological drug cluster (**Fig. 2**). In contrast, antineoplastic drugs and drugs for metabolic diseases do not form a single distinct cluster. These classes are underrepresented in the giant component and overrepresented in the smaller components (quantified in the **Supplementary Notes**), representing the least-connected drug classes relative to their class sizes.

The TP network provides a complementary, protein-centered view of pharmacological space[14]. In the TP network, 305 out of 394 target proteins are connected to other target proteins (**Fig. 3a**). Drugs with multiple targets are responsible for the high interconnectedness of the TP network. Although such promiscuous drugs were once thought to be undesirable in favor of more target-specific drugs, the recent success of anticancer drugs like imatinib (Gleevec) and sunitinib (Sutent) and of nonselective drugs for mood disorders and schizophrenia[15] seems to be shifting the industry toward such polypharmacology[16,17].

Historically the drug industry has relied upon a small number of targets. Topological features of the drug network and the TP network, such as giant component size, degree distribution and clustering coefficient (**Supplementary Figs. 3**,**4** online), quantitatively confirm this bias. Giant component size is the largest connected component of a network and measures local functional clustering when compared to other networks of similar topological properties[10,18]. We find that the actual size of the giant component of the drug network (476) is significantly smaller than the average giant component of 10^4 randomized networks generated by randomly shuffling the associations between drugs and proteins while keeping the number of links per drug and target protein unchanged (788 ± 9; empirical $P < 10^{-4}$) (**Supplementary Fig. 3a**). Similarly, the size of the giant component of the TP network (122) is significantly smaller than the average size for randomized target-protein networks (302 ± 8 gene products; empirical $P < 10^{-4}$) (**Supplementary Fig. 3b**). This result suggests that the polypharmacology acting on the target proteins is particularly enriched for highly targeted proteins. Higher-degree nodes in the drug network and TP network are preferentially connected to each other rather than being distributed homogeneously throughout the network, leading to a much smaller giant component size than expected[18]. Therefore, the DT network represents an intermediate structure between a completely random network with a very large giant component and a functionally fully segregated network broken into isolated clusters. The pharmaceutical industry shows a tendency to target already validated target proteins, causing an abundance of 'follow-on' drugs, and our analysis confirms this tendency.

The DT network is ever expanding with the continuous introduction of new FDA-approved drugs. Between 1982 and 2004, new drugs targeted on average 27.7 proteins each year, with 6.3 new therapeutic target proteins (**Fig. 1c**). If the target(s) of a novel drug correspond(s) to previously untargeted proteins, this drug would form a disconnected

component rather than increase the size of the already existing components of the DT network. Such drugs, which act discontinuously in the network, are referred to as 'jumping' drugs[19]. If one or more target(s) of a new drug is a known drug target, then this drug connects to one of the already connected components in the DT network, as if it were continuously 'crawling' in the DT network. In the same 1982–2004 period, 19.6 drugs were approved per year, but only 17% of these were jumping drugs (**Fig. 1d**). Targets of the jumping drugs constitute 67% of the newly introduced targets during this period. Hence, new drugs tend to bind known target proteins, attaching to islands of highly interconnected known targets in the network.

Experimental drugs

DrugBank also contains information about experimental drugs (drugs in the pipeline or not yet approved by the FDA). Although these experimental drugs have some biases, statistically significant trends in drug discovery might still be observed. There are currently more than 3,000 experimental drugs, 808 of which have at least one identified human protein target. The total number of drug targets increases to 1,011 when adding experimental drug targets. Inclusion of experimental drugs increases the size of the TP network giant component to 725, still significantly smaller than the average size of the giant component of 10^4 randomly generated graphs of identical node and degree distribution (782 ± 11) (**Supplementary Fig. 3d**). In contrast, the network formed by protein targets that are targeted by experimental drugs has a giant component size of 596, which is significantly larger than randomized networks (551 ± 10; $P < 10^{-4}$) (**Supplementary Fig. 3f**). These results indicate a trend toward more diversified target proteins. Experimental drugs show an increased tendency to be more promiscuous and introduce more associations between already existing proteins in the network, a phenomenon validated by the changes of the degree distribution and clustering coefficient following the addition of these drugs (**Supplementary Fig. 4c–e**).

Inclusion of experimental drugs changes the cellular component profile of target proteins. About 60% of approved drug targets are membrane proteins, down to ~40% when experimental targets are added (**Fig. 3b**). We also looked at the targets of drugs approved in the last 10 years (1996–2006). Despite the increase in diversity of experimental drug targets, targets of recently FDA-approved drugs are still mostly membrane proteins. Membrane proteins are easier to target, not least because getting drugs across membranes is challenging. Whereas experimental drugs clearly tend to target proteins localized in other cellular compartments, these efforts have not yet yielded significant changes in the cellular compartment distribution of approved drug targets.

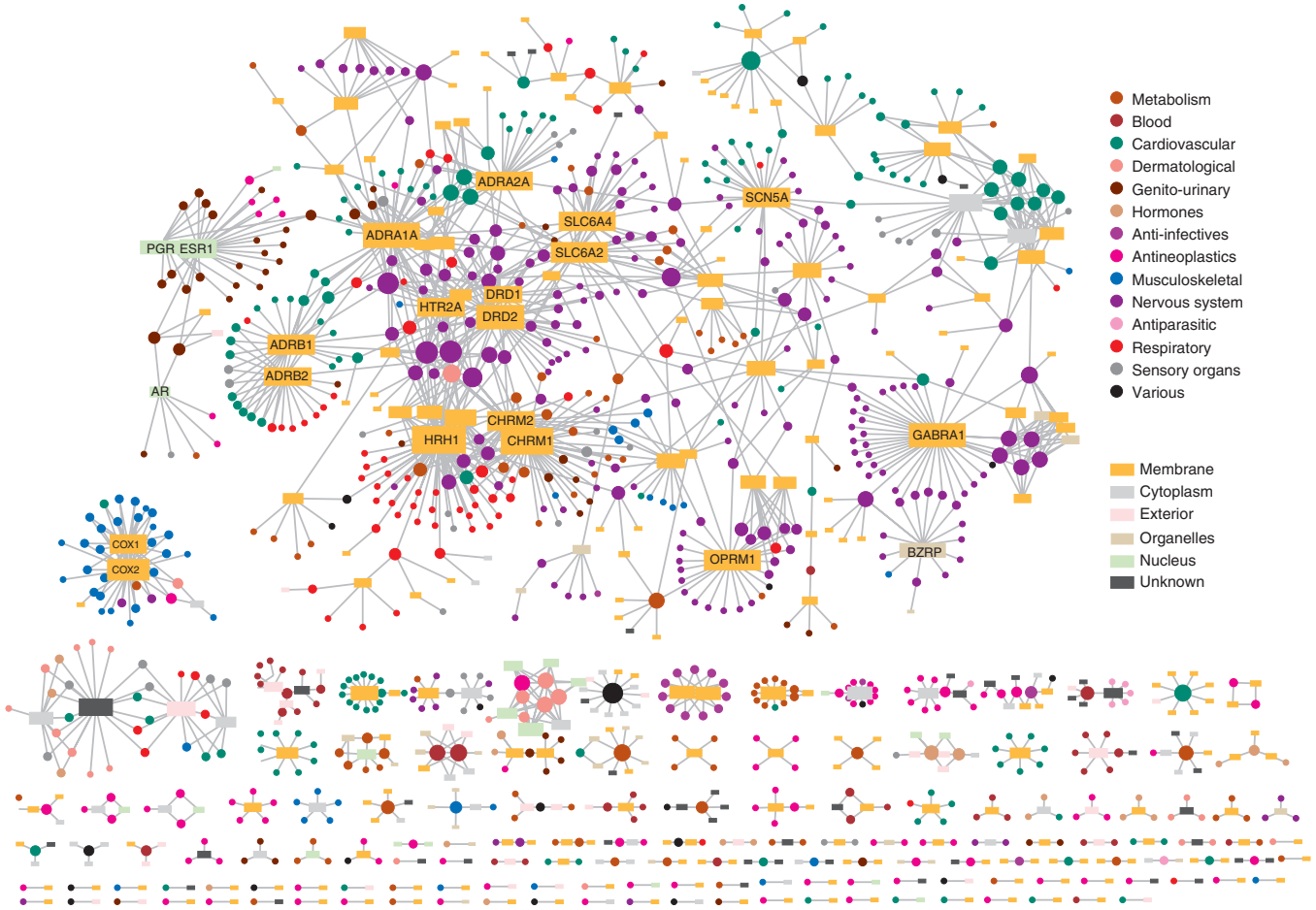

Figure 2 Drug–target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug has (the number of drugs targeting the protein). Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their cellular component obtained from the Gene Ontology database.

Drug targets and essentiality

To examine the global relationships between drug-target proteins in the human interactome network, we overlaid the TP network onto a network of physical protein-protein interactions (PPIs) derived from high-quality systematic interactome mapping[20,21] and from literature curation[20]. There are 262 drug-target proteins present in the PPI network. The drug-target proteins have 42% more interacting proteins (degree) on average than any protein in the PPI network ($P < 10^{-6}$, Wilcoxon rank-sum test) (**Fig. 4a** and **Supplementary Fig. 5** online). In PPI networks, the degree of a protein correlates with the essentiality of the protein[8]. To investigate the role of drug-target protein essentiality, we compared the degree of drug targets to the predicted human essential proteins, that is, proteins whose orthologous mouse protein is encoded by a gene found essential in knockout

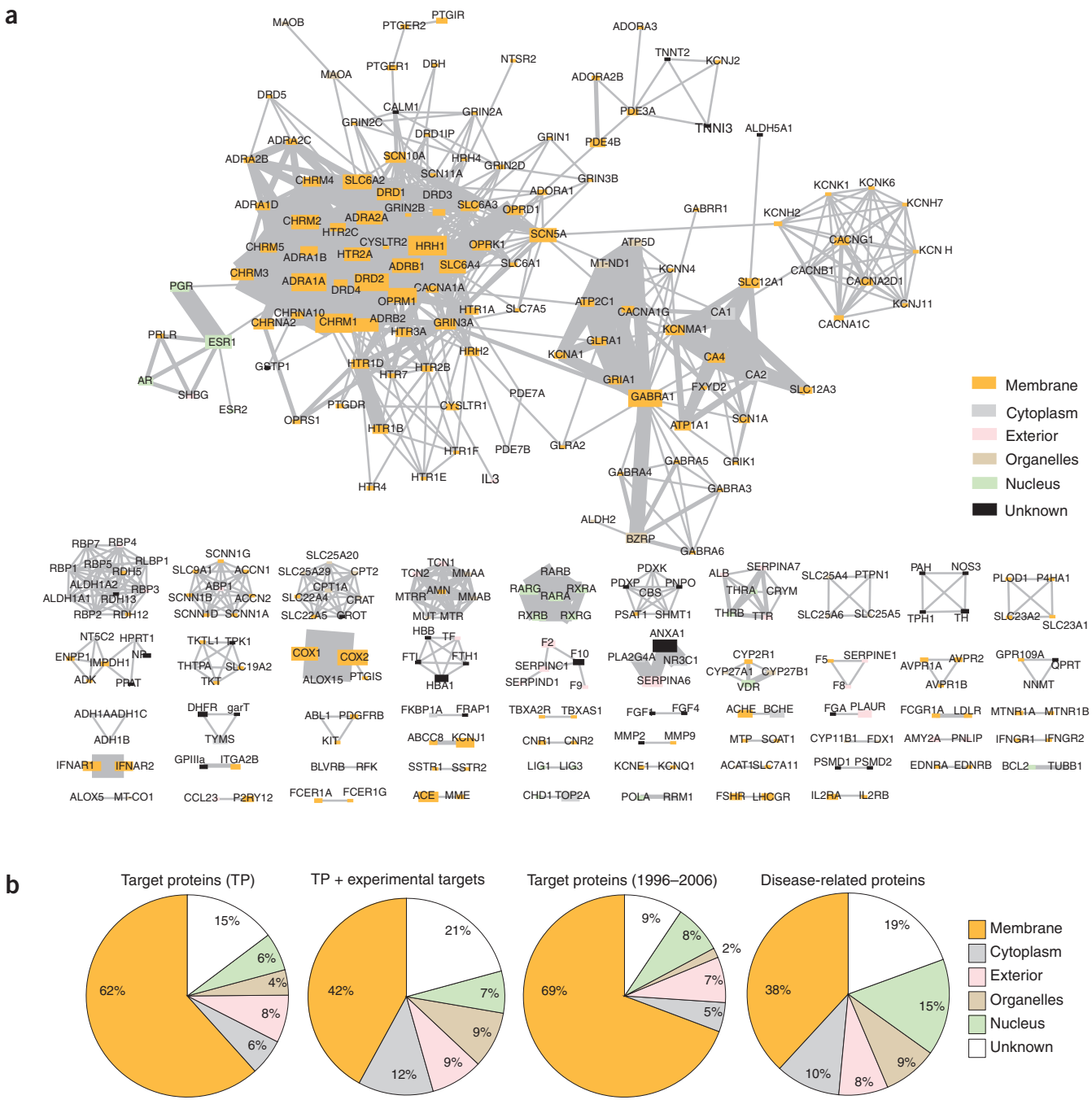

Figure 3 Target-protein network (TP network) and cellular component profiles. (**a**) In the TP network, each node is a protein, two proteins being connected if they are targeted by the same drug. The size of each node is proportional to the number of drugs targeting the gene. The nodes are colored according to their cellular component obtained from the Gene Ontology database. The thickness of the edge connecting two nodes is proportional to the number of drugs targeting both proteins at once. The TP network can be overlapped with cellular networks generated based on PPIs[20,21], transcription factor–promoter interactions[34–36] and metabolic reactions[37] to reveal correlations between targets of the same drug and other cellular functions. (**b**) Cellular component distribution of several classes of drug targets and disease genes.

experiments[22]. The average degree of essential proteins was significantly higher than the average degree of target proteins ($P < 10^{-2}$ for approved drug targets, $P < 10^{-10}$ with addition of experimental drug targets, Wilcoxon rank-sum test). Although drug-target proteins in general have more interactors, they do not necessarily show a trend towards greater essentiality (**Supplementary Notes** examines other topological features).

Essential proteins tend to coordinate the activity of diverse biological processes or 'modules'[23] and so tend to be coexpressed with other genes[10]. To assess coexpression, we used expression data from 36 different human tissue microarray experiments[24]. Genes encoding drug-target proteins show less coexpression with other genes compared with essential genes (**Fig. 4b**), indicating that drugs more likely act within modules than between modules. We also measured the average number of different tissues in which target proteins are expressed. The tissue count is lower than the average (**Fig. 4c**), showing high tissue specificity.

Drug targets and human disease genes

Currently the Online Mendelian Inheritance in Man (OMIM) reports on more than 1,284 disorders and 1,777 disease-related genes[25]. A map of disorder–disease gene associations in the OMIM Morbid Map was generated recently[10] (**Fig. 5a** and **Supplementary Fig. 6** online). Similarly to the TP network, a gene-centered human disease-gene (HDG) network was generated from the human disease map, where two genes are connected if they are associated with the same disease. A small portion of validated disease genes (166 genes) encodes drug-target proteins, with 71 genes (43%) associated with two or more diseases. Drugs approved recently (1996–2006) show the same proportion (43%). However, experimental drugs target 210 proteins in the HDG network, only 54 (26%) of which are involved in multiple diseases. The trend in experimental drug discovery seems to be toward more specific targets for a disease. Moreover, drug targets have significantly lower degrees ($P < 10^{-2}$ for approved drug targets, $P < 10^{-4}$ with addition of experimental drug targets using Welch's approximate *t*-test) compared with the network average (**Fig. 5b**), indicating an ongoing shift of drug development toward diseases with associated genes that were not prior drug targets.

Drugs do not target diseases equally, but are clearly enriched in some regions of the human disease network (**Fig. 5a**). To quantify this effect, we investigated the distribution of drug targets in the HDG network. Starting from a node in the network, we looked at the ratio of drug targets with respect to the distance from the origin, and we took the average of such ratios. If the drug targets were not clustered in a region, starting from a drug target would not be different than starting from a random node. Instead, we see a strong enrichment in the first and the second neighbors (**Fig. 5c**), showing a bias toward clustering of drug targets in the HDG network.

Cellular network–based relationships between drug targets and disease genes

The increase in new drug targets has been relatively slow in the aftermath of the sequencing of the human genome[1–4]. Effective usage of genomic information depends on finding systems connections between genetic variation, disease processes and drugs[26–29]. We undertook a systems-based investigation to explore whether drugs, their corresponding target proteins and disease-gene products might relate to each other at a higher level of organization.

Drugs act by exploiting two principal mechanisms: etiology-specific or palliative[30]. Etiology-specific drugs target the actual cause of the disease or etiologically related factors. Palliative drugs target proteins that are not the actual cause of the disease, but whose activity can be perturbed to counteract the symptoms of disease-causing proteins.

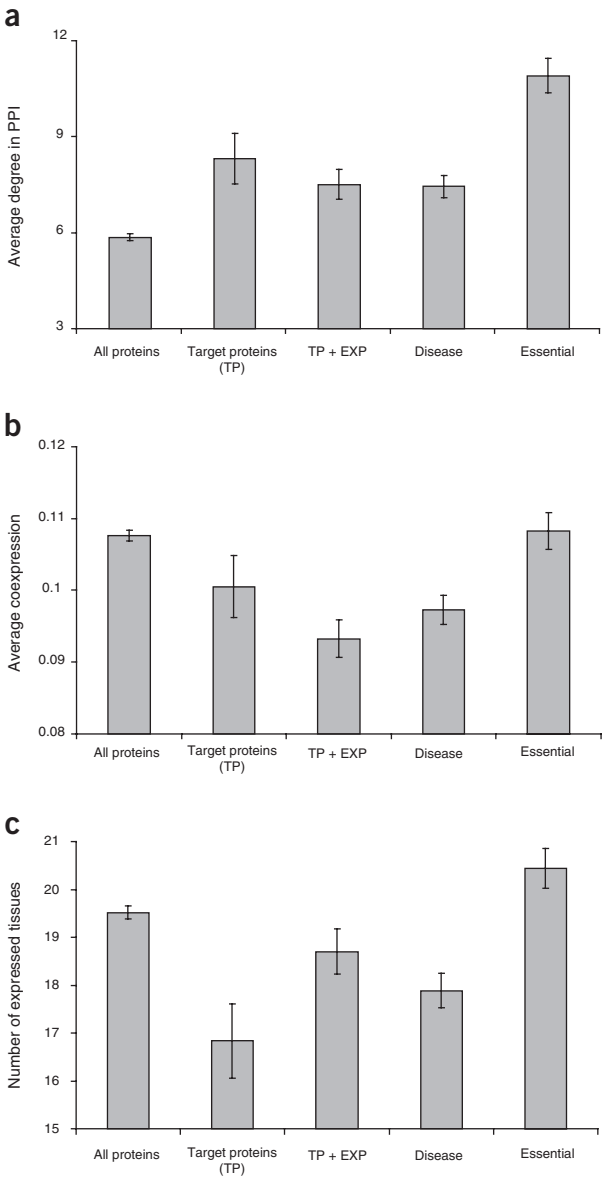

Figure 4 Drug targets, protein interactions and coexpression. (**a**) Average degree of different classes of proteins in PPI network. (**b**) Average Pearson correlation coefficients of a gene in a particular class with the rest of the cellular genes. (**c**) Average number of tissues in which each class is expressed.

We examined etiological and palliative drugs in the DT network by quantifying the relations between drug targets and disease–gene products in the human PPI set. We measured the minimum shortest distances between drug targets and disease-gene products implicated in their common disorder. The actual mechanism through which a drug acts may be unknown, but the shortest distance estimates the number of molecular steps between a drug target and the corresponding disease cause. In the combined human interactome network[20,21] there were 922 drug–disease pairs in which at least one drug-target protein and one corresponding disease-gene product were present. We observed a clear enrichment in the region of lower shortest distances compared with the randomized gene groups of similar size

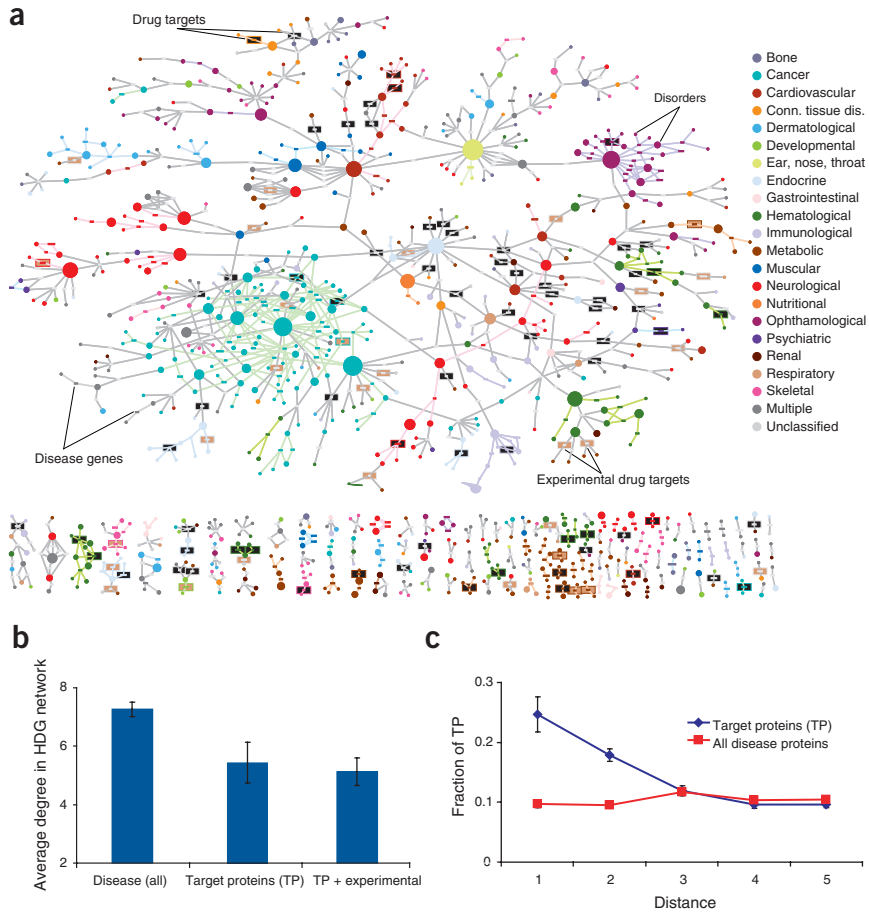

Figure 5 Human Disease Network and drug targets. (**a**) The Human Disease Network was generated from OMIM-based disorder–disease gene associations[10], where circles and rectangles correspond to disorders and disease genes, respectively. A link is placed between a disorder and a disease gene if mutations in that gene lead to the specific disorder. Disease genes associated only with a single disease are not shown for clarity. The drug targets are marked by larger boxes in the map, and they are colored either black or tan corresponding to targets of approved and experimental drugs, respectively. Disease categories that are preferentially associated with genes and that are also frequent drug targets are endocrine, hematological, cardiovascular and psychiatric. In contrast, cancer, muscular, skeletal, gastrointestinal and dermatological disease categories are associated with fewer drug targets than average (**Supplementary Fig. 6c**). (**b**) Average degree of several gene classes in the human disease gene HDG network[10]. (**c**) Fraction of target proteins while applying a breadth-first search starting from either a target protein or a random protein in the HDG network with respect to distance.

and pairings (**Fig. 6a**). Most drugs closely matched the randomized distances, suggesting a preponderance of palliative drugs. Most drugs found in earlier stages show little selection toward the genetic cause of the disease, which is consistent with the fact that they were found through traditional chemical screening. A similar analysis can also be performed by gene co-expressions (**Supplementary Fig. 7**).

Given recent increased knowledge about the cause of disease, more-rational drug design is expected to become more frequent. To test this hypothesis, we repeated the distance analysis above for the drugs approved since 1996. There is a significant shift toward higher weights in the shorter distances compared with the drugs approved before 1996 ($P < 10^{-2}$, Kolmogorov-Smirnov test), consistent with a move toward rational drug design (**Fig. 6b**).

Drug target to disease gene relations are more direct than randomized expectation for the cancer, endocrine, psychiatric and respiratory disease classes (**Fig. 6c**). With the possible exception of cancer, many diseases in these classes arise from aberrant activity of drug-accessible plasma membrane receptors. Successful drugs for those disease classes generally mimic the natural interactors of those receptors[15,31]. In contrast, distances are longer than randomized expectation for the developmental, muscular and ophthalmologic disease classes. Metabolic disease–gene products show the farthest distance to the corresponding drug targets, although this distance might be shorter if metabolic networks were used instead.

Cancer is a genetic disease caused by combinations of hyperactive oncogenes and defective tumor suppressor genes[32]. Modern anticancer drugs are rationally designed to alter or suppress aberrant oncogene activity. For instance imatinib[33] directly targets tyrosine kinase receptors (*KIT*, *BCR-ABL1* and *PDGFRB*) and thus shows a distance of zero in our measurement. Compared with other disease classes, we see a greater proportion of drugs with distance 1 or 2 in the cancer class (**Fig. 6c**,**d**). The 'farthest' cancer drugs, mainly used in advanced cases of cancer, such as abaralix (Plenaxis), carmustine (BiCNU) and zoledronate (Zometa, Reclast), are palliative drugs prescribed to counteract the debilitating effects of radiation therapy or to target all cells of a specific tissue instead of targeting cancer cells specifically.

DISCUSSION

We used the concepts of network biology to integrate data from DrugBank and OMIM with information on gene expression and PPIs, allowing us to address three questions regarding drug development: (i) What are the industry trends? (ii) What are the properties of drug targets in the context of cellular networks? (iii) How do drug targets relate to disease-gene products? The results indicate that well-known targets remain the preponderant targets of new drugs, with recent slow diversification of protein targets. The drug targets occupy certain regions in the interactome networks, and their topological signatures are different compared with essential proteins, an observation also supported by the expression profile analysis. The novel distance metric nicely shows that most drugs are palliative and do not directly perturb the protein(s) corresponding to the underlying cause of disease. With improving understanding of the genetic basis of disease, drug targets are becoming more related to disease-gene products.

Although these data sets are far from complete, our network analyses still provide statistically significant characteristics of drug targets. The list of drug targets is updated frequently, and we could not compare the drugs currently under investigation with drugs that were under investigation in the past to predict the reasons behind success or failure of a particular drug. The characteristics of drugs under investigation will not be exactly the same as those that will be validated. However, using the available

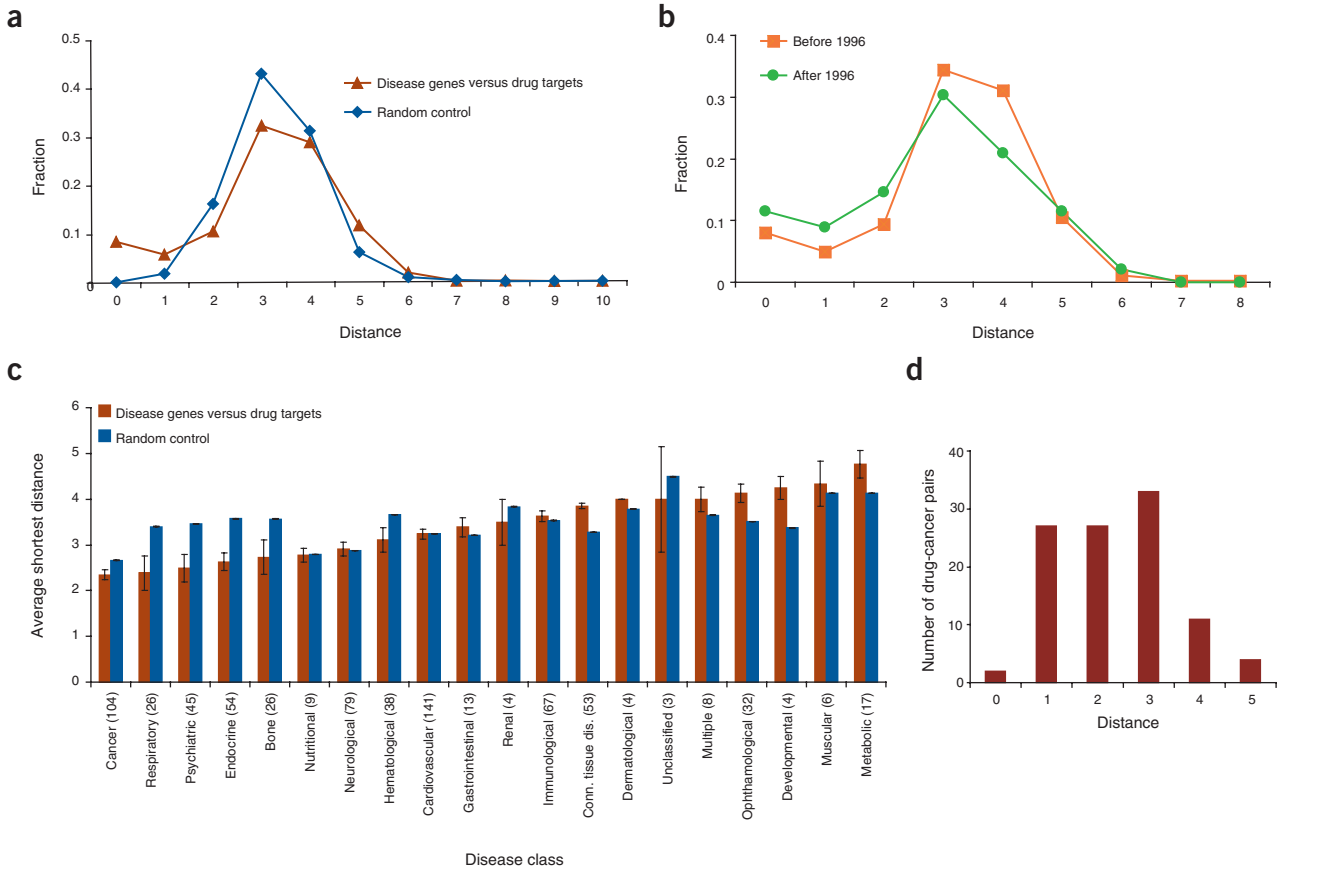

Figure 6 Drug targets and disease genes on human PPI network. (**a**) Distribution of the shortest distances for the actual data (red) and random groups of proteins (blue). There is an enhancement at the distances 0 and 1. (**b**) Shortest distance comparison of targets of drugs approved in last 10 years and before 1996. (**c**) Shortest distance for the different disease classes. The numbers inside the parentheses show how many such drug target–disease gene pairs are present in the corresponding disease class. (**d**) Distribution of shortest distances for the cancer-disease class of drugs.

information about experimental drugs we mapped the current test space to the fullest extent possible, finding quantifiable differences in the topological characteristics of approved drugs and experimental drugs.

Our analysis of the DT network suggests a need to update the single drug–single target paradigm, just as single protein–single function relations are somewhat limited to accurately describe the reality of cellular processes[16]. Future attempts at rational drug design will eventually take into account the 'systems' effects of a drug on the greater network upstream and downstream of the actual drug target, which could pave the way to more specific drugs for diseases.

METHODS

Drug databases. We downloaded the DrugBank database[13] as of March 29, 2006. DrugBank combines detailed drug data (chemical, pharmacological and pharmaceutical) with comprehensive drug-target information (primary sequence, three-dimensional structure and pathway involvement). The database contains 4,252 drug entries including 1,178 FDA-approved small molecule drugs (including 113 FDA-approved biotech (protein/peptide) drugs) and 3,074 experimental drugs. We selected the drugs which are known to have human target proteins. All associations between drugs and known target proteins can be found online as **Supplementary Tables 1** and **2**.

OMIM database and diseasome map. We used the OMIM Morbid Map[25], which contains the most complete known disorder–gene associations, as of December 21, 2005. We used the same disorder classification as used previously [10]. There are 1,284 disorders, which are grouped into 22 disorder classes, and 1,777 disease genes. The Human Diseasome Network was constructed by linking disorders to genes if mutations in a gene are implicated in formation of a disorder (see **Supplementary Notes** for a more detailed description).

Topological features of a network. The 'degree' of a node is the number of edges connecting to the node. The 'giant component' is the largest connected component of the network. The 'clustering coefficient' is defined as $C_i = 2n/k_i(k_i - 1)$, where n denotes the number of direct links connecting the k_i nearest neighbors of node i. If clustering coefficient of a node equals 1, then the node is at the center of a fully interlinked cluster. If the clustering coefficient is close to 0, then the node is part of a loosely connected group. The average of C_i over all nodes of a network assesses network modularity.

Randomization of drug–target protein associations. To obtain random controls for the topological features of DT network, drug network and TP network, we first generated a randomized DT network by randomly shuffling the drug–target protein associations, while keeping unchanged both the number of proteins that a drug targets and the number of drugs that a protein is targeted by. From this network, we created the randomized drug network and TP network by projecting onto drug and protein spaces, respectively. We generated 10^4 independent randomized samples.

Time-stamping the drugs. We downloaded the Drugs@FDA database as of November 21, 2006 (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/). This database includes the history of actions taken for each drug. We used the first date that the drug (that is, new chemical entity) was approved as our time stamp. We also used the Encyclopedia of Molecular Targets (http://emot.mit.edu/) database

to validate our time stamps. Drugs for which we could not identify any approval date were excluded from the time-stamp analysis.

Protein–protein interaction data. To obtain detailed human PPI data, we used two high-quality systematic yeast two-hybrid experiments[20,21] and PPIs obtained from literature by manual curation[20]. The integrated set of PPIs contains 22,052 non-self-interacting, nonredundant interactions between 7,533 genes. The giant component of the PPI network contains 7,279 proteins, of which 253 are targets of approved drugs and 1,159 are associated with diseases.

Gene expression microarray data. We used microarray data available for 36 normal human tissues[24]. There were 293 approved drug targets and 808 of all drug targets that have expression information. A gene is considered to be 'expressed' if the *P* value associated with its transcript abundance is less than the threshold ($P < 0.02$)[24]. Genes that are not expressed in any examined tissue are excluded from the analysis.

Mouse phenotype data. To predict the essentiality of a human gene, we used the phenotype information of the corresponding mouse ortholog. A human gene was defined as 'essential' if a knockout of its mouse ortholog confers lethality. We obtained the human–mouse orthology and mouse phenotype data from Mouse Genome Informatics[22] on January 3, 2006. We considered the classes of embryonic/prenatal lethality and postnatal lethality as lethal phenotypes, and the rest of the phenotypes as nonlethal ones. There were 1,267 mouse-lethal human orthologs, of which 77 are targets of approved drugs (~20% of targets of approved drugs) and 149 are targets of all drugs (including both approved and experimental drugs).

TP network and disease-gene relations on PPI. We mapped drugs to diseases by searching for disease keywords in the 'indications' field of drug information obtained from the DrugBank database[13], first automatically and then by validating resulting associations manually (**Supplementary Table 3**). For each drug–disease pair, we calculated the minimum distance on the PPI map between pairs of target proteins and disease-associated proteins. To generate randomized controls, we selected the same number of proteins from PPIs 10^4 times randomly to control for drug targets. Keeping the disease genes constant, we calculated the statistics from minimum distance values for these randomly generated drug–disease pairs.

Statistical tests. All the *t*-tests were done in Mathematica (Wolfram Research) using the HypothesisTests package. Kolmogorov-Smirnov and Wilcoxon rank-sum tests were done in Matlab (Mathworks) using the "kstest2" and "ranksum" commands, respectively. All the error terms in the text and the figures are the standard errors.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank Andrew L. Hopkins, William G. Kaelin and the members of the M.V. and A.-L.B. laboratories and the Center for Cancer Systems Biology (CCSB), especially David E. Hill, for useful discussions. This work was supported by the Dana-Farber Cancer Institute Strategic Initiative (to M.V.), the W. M. Keck Foundation (to M.V.) and an National Institutes of Health (NIH) grant 2R01-HG001715 from the National Human Genome Research Institute and the National Institute of General Medical Sciences (to M.V. and Frederick P. Roth). K.-I.G. and A.-L.B. were supported by NIH grants IH U01 A1070499-01 and U56 CA113004 and National Science Foundation Grant ITR DMR-0926737 IIS-0513650.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hopkins, A.L. & Groom, C.R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
2. Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discov.* **5**, 821–834 (2006).
3. Overington, J.P., Bissan, A. & Hopkins, A. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
4. Russ, A.P. & Lampel, S. The druggable genome: an update. *Drug Discov. Today* **10**, 1607–1610 (2005).
5. Drews, J. Stategic trends in the drug industry. *Drug Discov. Today* **8**, 411–420 (2003).
6. Barabasi, A.L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
7. Han, J.-D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
8. Jeong, H., Mason, S., Barabasi, A.-L. & Oltvai, Z. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
9. Vidal, M. Interactome modeling. *FEBS Lett.* **579**, 1834–1838 (2005).
10. Goh, K.I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685–8690 (2007).
11. Jimenez-Sanchez, G., Childs, B. & Valle, D. Human disease genes. *Nature* **409**, 853–855 (2001).
12. Peltonen, L. & McKusick, V. Dissecting human disease in the postgenomic era. *Science* **291**, 1224–1229 (2001).
13. Wishart, D.S. *et al.* Drugbank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
14. Paolini, G.V., Shapland, R.H., van Hoorn, W.P., Mason, J.S. & Hopkins, A.L. Global mapping of pharmacological space. *Nat. Biotechnol.* **24**, 805–815 (2006).
15. Roth, B.L., Sheffler, D.J. & Kroeze, W.K. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discov.* **3**, 353–359 (2004).
16. Hopkins, A.L., Mason, J.S. & Overington, J.P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **16**, 127–136 (2006).
17. Mencher, S.K. & Wang, L.G. Promiscuous drugs compared to selective drugs (promiscuity can be a virtue). *BMC Clin. Pharmacol.* **5**, 3 (2005).
18. Newman, M.E. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* **64**, 016131 (2001).
19. Cokol, M., Iossifov, I., Weinreb, C. & Rzhetsky, A. Emergent behavior of growing knowledge about molecular interactions. *Nat. Biotechnol.* **23**, 1243–1247 (2005).
20. Rual, J.-F. *et al.* Toward a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
21. Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
22. Eppig, J.T. *et al.* The Mouse Genome Database (MGD): from genes to mice–a community resource for mouse biology. *Nucleic Acids Res.* **33**, D471–D475 (2005).
23. Hartwell, L.H., Hopfield, J.J., Leibler, S. & Murray, A.W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
24. Ge, X. *et al.* Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* **86**, 127–141 (2005).
25. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. & McKusick, V.A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–D517 (2005).
26. Butcher, E.C., Berg, E. & Kunkel, E. Systems biology in drug discovery. *Nat. Biotechnol.* **22**, 1253–1259 (2004).
27. Chanda, S.K. & Caldwell, J. Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov. Today* **8**, 168–174 (2003).
28. Searls, D.B. Pharmacophylogenomics: genes, evolution and drug targets. *Nat. Rev. Drug Discov.* **2**, 613–623 (2003).
29. van der Greef, J. & McBurney, R. Rescuing drug discovery: *in vivo* systems pathology and systems pharmacology. *Nat. Rev. Drug Discov.* **4**, 961–967 (2005).
30. Lindpaintner, K. The impact of pharmacogenetics and pharmacogenomics on drug discovery. *Nat. Rev. Drug Discov.* **1**, 463–469 (2002).
31. Gershell, L.J. & Atkins, J. A brief history of novel drug discovery technologies. *Nat. Rev. Drug Discov.* **2**, 321–327 (2003).
32. Futreal, P.A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
33. Capdeville, R., Buchdunger, E., Zimmermann, J. & Matter, A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.* **1**, 493–502 (2002).
34. Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. & Teichmann, S.A. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14**, 283–291 (2004).
35. Lee, T.-I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
36. Rodriguez-Caso, C., Medina, M.A. & Sole, R.V. Topology, tinkering and evolution of the human transcription factor network. *FEBS J.* **272**, 6423–6434 (2005).
37. Wagner, A. & Fell, D.A. The small world inside large metabolic networks. *Proc. Biol. Sci.* **268**, 1803–1810 (2001).