25. A. A. Beg, D. Baltimore, *Science* **274**, 782 (1996).
26. D. W. Nicholson *et al.*, *Nature* **376**, 37 (1995).
27. B. Beutler, I. W. Milsark, A. C. Cerami, *Science* **229**, 869 (1985).
28. E. K. Shiner, K. P. Rumbaugh, S. C. Williams, *FEMS Microbiol. Rev.* **29**, 935 (2005).
29. L. Hagberg, D. E. Briles, C. S. Eden, *J. Immunol.* **134**, 4118 (1985).

30. We thank B. Beutler, M. Karin, and G. Nolan for critically reading the manuscript and helpful comments. This work was supported by The Skaggs Institute of Chemical Biology (K.D.J.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1156499/DC1
Materials and Methods
Figs. S1 to S12
Table S1
References

14 February 2008; accepted 6 June 2008
Published online 19 June 2008;
10.1126/science.1156499
Include this information when citing this paper.

Drug Target Identification Using Side-Effect Similarity

Monica Campillos,[1]* Michael Kuhn,[1]* Anne-Claude Gavin,[1] Lars Juhl Jensen,[1,2] Peer Bork[1,3]†

Targets for drugs have so far been predicted on the basis of molecular or cellular features, for example, by exploiting similarity in chemical structure or in activity across cell lines. We used phenotypic side-effect similarities to infer whether two drugs share a target. Applied to 746 marketed drugs, a network of 1018 side effect–driven drug-drug relations became apparent, 261 of which are formed by chemically dissimilar drugs from different therapeutic indications. We experimentally tested 20 of these unexpected drug-drug relations and validated 13 implied drug-target relations by in vitro binding assays, of which 11 reveal inhibition constants equal to less than 10 micromolar. Nine of these were tested and confirmed in cell assays, documenting the feasibility of using phenotypic information to infer molecular interactions and hinting at new uses of marketed drugs.

The treatment of human diseases with carefully selected drugs provides a long-lasting controlled chemical perturbation experiment in a complex organism. Its readout includes the regulated recording of side effects summarized in the package inserts (also known as patient information leaflets or drug labels). Drug side effects are complex phenomenological observations that have been attributed to a number of molecular scenarios including the interaction with the primary or additional targets (off-targets hereafter), downstream pathway perturbations, kinetic and dosage effects, drug-drug interference, insufficient metabolization, effects of active metabolites, and aggregation or irreversible target binding of the drug (*1*). Of these, direct interaction with proteins seems to be one of the most important scenarios (*2*, *3*).

Although unexpected activities derived from off-targets are usually unwanted and harmful, they can sometimes be beneficial and have led to new therapeutic indications for drugs. For instance, sildenafil (Viagra, Pfizer Incorporated, New York, New York) was developed to treat angina, but its side effect of prolonged penile erections in human volunteers led to a change in the therapeutic area of the drug (*4*).

Similar side effects of unrelated drugs can be caused by their common off-targets. For example, the two dissimilar drugs cisapride and astemizole both cause cardiac arrhythmias because they inhibit the cardiac ion channel hERG in addition to their primary targets (serotonin and histamine receptors, respectively) (*5*). In general, drugs with similar in vitro protein binding profiles tend to cause similar side effects (*6*, *7*), implying a direct correlation between target binding and side-effect similarity and hence a possibility to predict off-target binding. So far, additional targets for known drugs have been systematically identified through phenotypic assays, by exploiting chemical similarity measures, and through docking strategies [e.g., (*8*–*12*), reviewed in (*13*)]. All of these discoveries had their bases in the analysis of cell assays or binding studies and did not consider the entire human system.

Therefore, we explored side-effect information generated from the use of marketed drugs to infer molecular activities of drugs that are not implicit by their chemical similarity or the sequence similarity of their known targets. We developed a measure for side-effect similarity

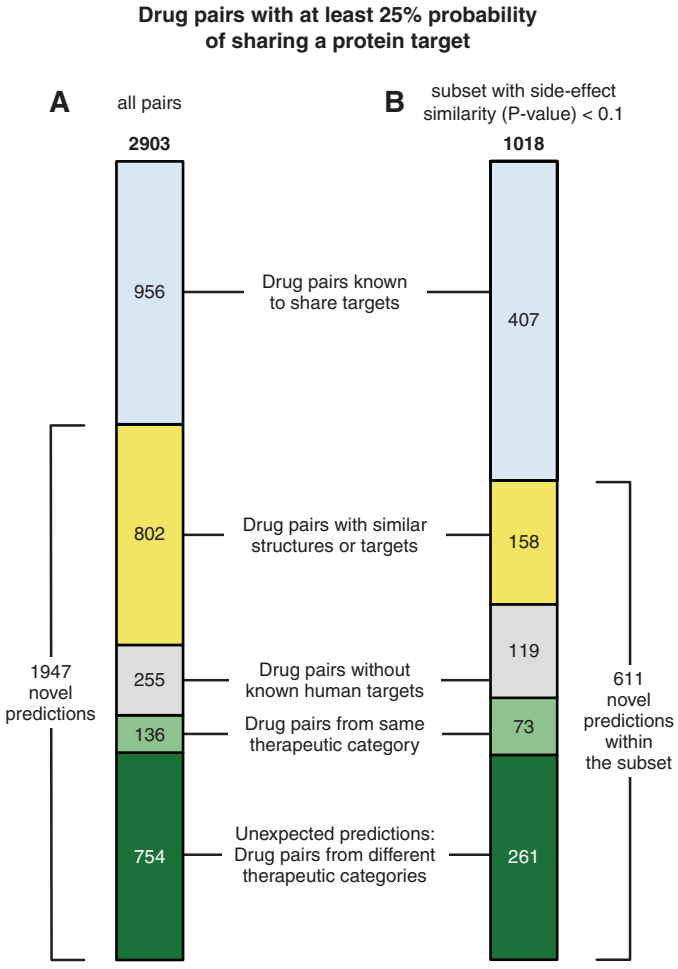


Fig. 1. Breakdown of drug pairs predicted to share a target. (**A**) We subjected the initial set of 2903 to a series of stringent filters, leaving 754 pairs that imply unexpected drug-target relations. In particular, we filtered out pairs of structurally similar drugs (Tanimoto 2D similarity > 0.6) and drug pairs with similar known targets [normalized bitscore > 0.12, which corresponds to ~28% sequence identity (fig. S6)] because both molecular features can be used independently to infer targets (*13*, *20*–*22*). Thus, the unexpected relations contain combined contributions from weak chemical similarities and a range of side-effect similarities. (**B**) The subset of drug pairs that are predominantly based on strong side-effect similarity [*P* value < 0.1 (*15*)] was used for network analysis (Fig. 2).

[1]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany. [2]Novo Nordisk Foundation Centre for Protein Research, Panum Institute, Blegdamsvej 3b, 2200 Copenhagen, Denmark. [3]Max-Delbrück-Centre for Molecular Medicine, Robert-Rössle-Strasse 10, 13092 Berlin, Germany.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bork@embl.de

(fig. S1), analyzed the likelihood of sharing protein targets for 277,885 pairs of 746 marketed drugs, and confirmed experimentally that side-effect similarity indeed indicates common protein targets of unrelated drugs. Thus, we are able to propose additional targets for many existing drugs, often implicated in different therapeutic categories.

To classify side effects, we used the Unified Medical Language System (UMLS) ontology for medical symptoms (*14*) and extracted relevant terms from drug package inserts (*15*). We used the relations between terms in the ontology to capture also similarities between drugs annotated with distinct but closely related terms. Not all side effects are independent of each other; for example, most drugs that cause nausea also cause vomiting. We corrected for this redundancy by weighting side effects in a manner analogous to the down-weighting of similar protein sequences within multiple alignments (*15*, *16*).

The recorded side effects vary greatly in abundance: Some, like megaloblastic anemia, are caused by only a few drugs, whereas others, like dizziness, occur for most. Within a reference set of 502 drugs with 4857 known human drug-target relations (*15*) from the Matador (*17*), DrugBank (*18*), and PDSP K_i (Psychoactive Drug Screening Program inhibition constant) databases (*19*), we observed an inverse correla-

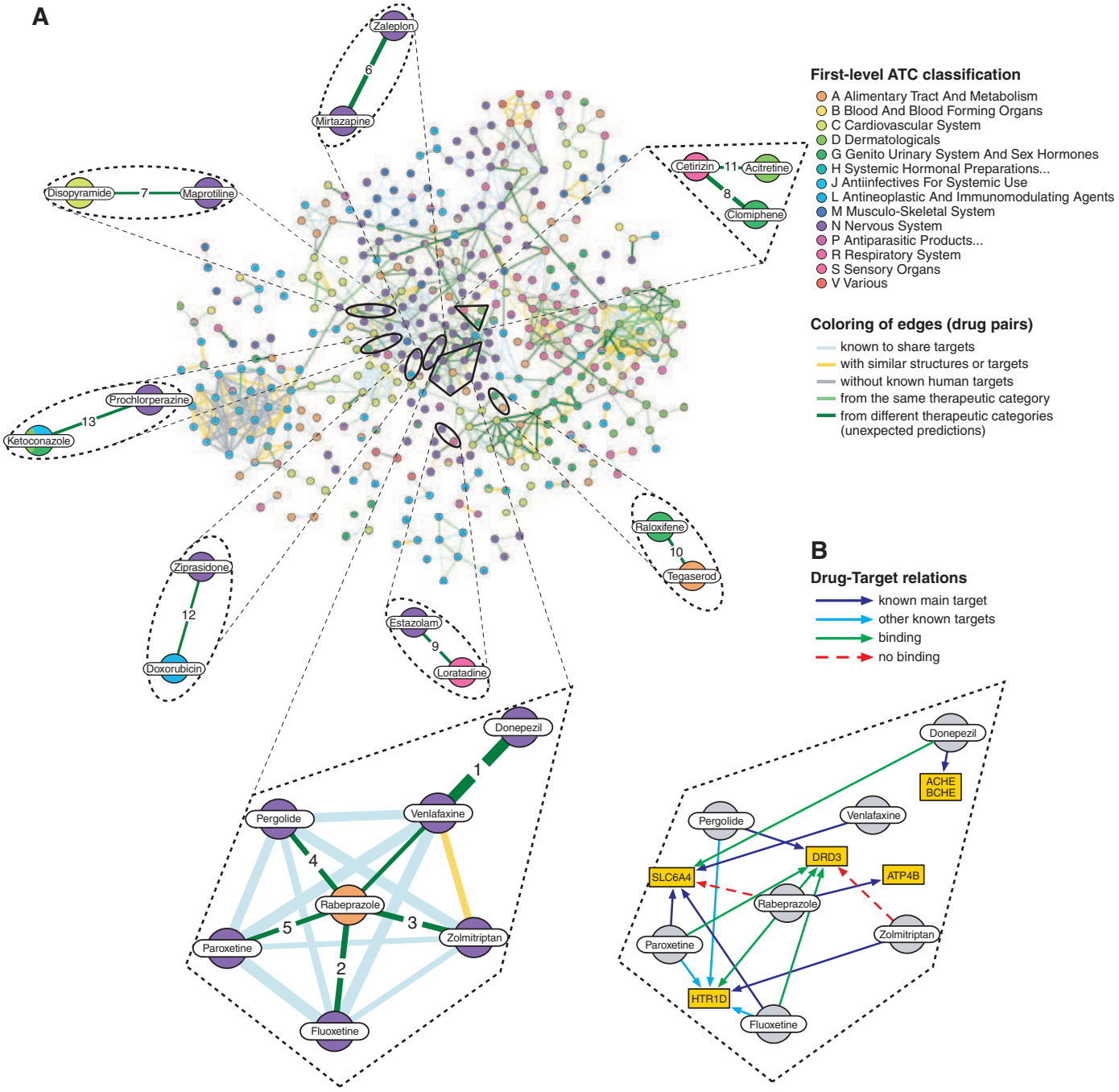


Fig. 2. Network of drugs predicted to have common protein targets. (**A**) 424 drugs (nodes) form 1018 pairs with strong side-effect similarity and above 25% probability of sharing a target (edges, width proportional to probability). Drug subnetworks around the antiulcer drug rabeprazole and other experimentally confirmed predictions are magnified. (**B**) Selected drug-target relations in the subnetwork around rabeprazole (see fig. S10 for other drug-target pairs). Predicted drug-target relations that were experimentally validated (Fig. 3) are shown with green arrows; dashed red arrows indicate that the predicted targets could not be confirmed. The confirmed relations are sufficient to prove the predicted drug-drug relations in the rabeprazole subnetwork.

likelihood of two drugs to share a protein target, and we weighted side effects accordingly (fig. S1D).

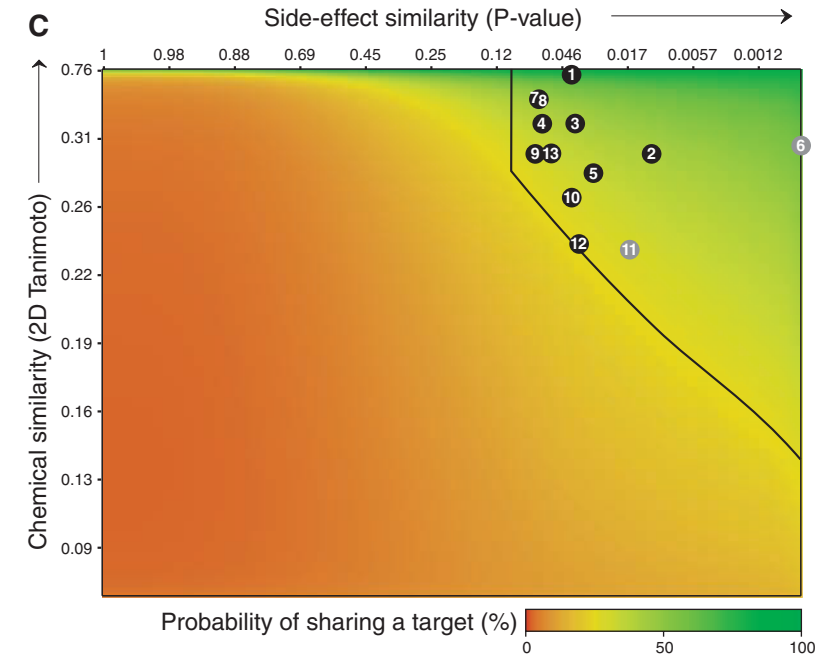
A measure for side-effect similarity was established by using these weighting schemes and by incorporating statistical significance assessments (*15*). We tested the predictive power of this side-effect similarity measure on our reference set of 502 drugs with known human targets and observed a clear correlation between side-effect similarity and the likelihood that two drugs share a protein target (fig. S1H). Side-effect similarity can thus be used to predict new targets for old drugs.

Consistent with previous studies [e.g., (*20*–*22*)], we observed in our reference set that chemically similar drugs [according to the two-dimensional (2D) Tanimoto chemical similarity score (*15*)] are likely to have the same targets (fig. S1H). The corresponding predictions showed only a small overlap with those based on side effects: In the reference set, only 35 drug pairs are in common between the 198 and 301 pairs with more than 50% probability of sharing targets according to their side-effect similarity and chemical similarity, respectively. Consequently, we combined side-effect similarity and chemical

A

Pair	Prob.	Tested Drug	Reference Drug	Target	K_i(μM)	Cellular Act. Antag./Inhib.	Ago.
1	75	donepezil	venlafaxine	5HTT	9 ❶	**100**	
2	40	fluoxetine	rabeprazole	DRD3	2 ❷	27	**60**
3	40	rabeprazole	zolmitriptan	HTR1D	7.6 ❸	*Assay not available*	
4	36	rabeprazole	pergolide	DRD3	1.6 ❹	**87**	<20
5	33	paroxetine	rabeprazole	DRD3	3.8 ❺	<20	*46*
6	57	zaleplon	mirtazapine	HRH1	26 ⑥	**74**	<20
7	44	disopyramide	maprotiline	HRH1	2.7 ❼	**106**	<20
8	43	clomiphene	cetirizine	HRH1	6.5 ❽	*44*	<20
9	30	loratodine	estazolam	BZRP	5 ❾	*Assay not available*	
10	29	raloxifene	tegaserod	HTR1D	0.3 ❿	*Assay not available*	
11	29	acitretin	cetirizine	HRH1	15 ⑪	**59**	<20
12	25	doxorubicin	ziprasidone	HRH1	10 ⑫	*Compound interference*	
*13	30	ketoconazole	prochlorperazine	Serot. Rec.	2.8 ⑬	*45*	<20

$K_i \leq 10$ 11 ● **Activity>50** 6
$10 < K_i < 30$ 2 ● *40<Activity<50* 3



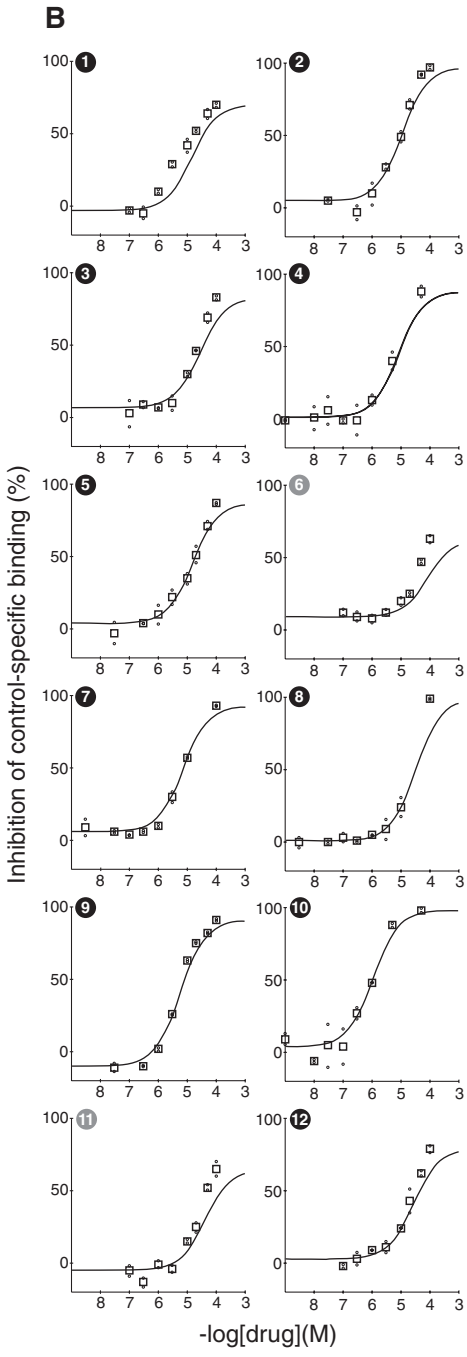


Fig. 3. Novel drug-target relations. (**A**) Values of K_i for the 13 drug-target relations were measured for those drugs that showed an in vitro binding activity higher than 40% at 50 μM. When possible, drug-target relations were validated in cell assays by measuring the activity of the compounds at 50 μM. The asterisk denotes a candidate that was partially insoluble (fig. S8B). (**B**) Concentration curves from competition assays for the novel drug-target relations. (**C**) By using our reference set of 4857 drug-target relations (*15*), we assigned probabilities on the basis of a combination of side-effect similarity and chemical similarity. The line delimits the area used to construct the network in Fig. 2 with shared target probability >25% and side-effect similarity *P* value < 0.1. Drug pairs that were experimentally confirmed to share a target are denoted by black and gray dots according to K_i value [see (A)].

similarity and benchmarked the result against our reference set to obtain the final probabilities for any two drugs to share a protein target (*15*). Both specificity and sensitivity improved considerably (fig. S3).

We next applied our target prediction method to a larger set of 746 human-marketed drugs for which side-effect information is available (table S1), including 244 drugs that have no annotated human targets in our reference set (for example, antibiotics). After exclusion of 44 drugs with less than seven side effects (too few to make specific predictions) (fig. S4), we predicted 2903 pairs of drugs to share a target with over 25% probability (Fig. 1A and fig. S5). We use this arbitrary 25% cutoff in the following because, above this value, the combined method was more sensitive than chemical similarity or side-effect similarity alone (fig. S3D). The actual chance of sharing a target is likely to be higher than our scoring scheme indicates because many binding partners for known drugs are not known yet and were thus counted as false negatives in our benchmarks. Among the 2903 predicted pairs, 956 were known to have common targets (Fig. 1A), which is fivefold higher than what would be expected by random chance (*15*). From the remaining 1947 drug pairs that imply novel predicted interactions, we removed 1193 drug pairs that could be expected to share targets because the drugs are in related indication areas, are chemically similar, or have similar targets (Fig. 1A and fig. S6) (*13*, *15*). Thus, we predicted unexpected, shared targets for 754 drug pairs.

To get an overview of the subset of predictions that are driven by side-effect similarity (1018 of the total 2903 predictions, Fig. 1B), we constructed a network of the corresponding 424 drugs with at least 25% probability of sharing a target (Fig. 2; see fig. S7 for the complete network from all predictions as depicted in Fig. 1A). Of these, 261 pairs were examined in more detail because they involved dissimilar drugs from different therapeutic indications ("unexpected relations" in Fig. 1B and table S2).

We focused on areas in the network that contain drugs from different therapeutic categories (Fig. 2). For example, there is a subnetwork of several drugs targeting the nervous system around the antiulcer drug rabeprazole, a proton pump inhibitor. Within this subnetwork, five drug pairs were predicted to share targets with a probability in the range from 30 to 75%, four of which involve rabeprazole. We validated all our predictions in this subnetwork with both in vitro and cell assays (Fig. 3). We found that rabeprazole inhibits the dopamine receptor DRD3 and binds the serotonin receptor HTR1D (Fig. 2B). The nervous system drugs pergolide, paroxetine, and fluoxetine share these targets with rabeprazole (Fig. 2B), whereas zolmitriptan seems to have only its primary target, serotonin receptor HTR1D, in common with rabeprazole (Fig. 2B). Taken together, the sharing of side effects of the proton pump inhibitor rabeprazole revealed two nervous system off-targets with affinities (Fig. 3) that have been shown to cause side effects (*23*) and should be physiologically relevant given rabeprazole's plasma concentrations (*24*). Our experimental validations also imply that all drug-drug associations in this subnetwork (Fig. 3B) are indeed caused by shared targets.

To generalize our validations, we experimentally tested predictions derived from another 15 drug pairs in addition to the five predictions around rabeprazole (Fig. 2A). All predictions involve at least one drug with a human target and are from the "unexpected" category (261 candidate pairs comprising dissimilar drugs from different indication areas in Fig. 1B). In total, for 13 of the 20 pairs tested, we confirmed binding activity to at least one of their predicted targets in vitro (Fig. 3 and figs. S8 and S9). Eleven of the observed binding affinities are strong enough to lead to side effects [median inhibitory concentrations < 50 μM (*23*)], 11 can be considered biologically active [inhibition constant (K_i) < 10 μM (*12*)], and 7 appear relevant in vivo (K_i values within one order of magnitude of the measured average drug plasma concentrations, table S3). For 9 of the 13 drug-target relations with in vitro activity, cell assays were available, and all confirmed the predicted activity (Fig. 3). Both the observed phenotypic similarity (shared side effects) that led to these predictions and the cellular activities confirmed here support the possible physiological relevance of the newly identified drug-target relations.

All verified predictions imply binding of existing drugs to proteins associated with different therapeutic categories. For example, we have found a relation between the nootropic drug donepezil and the antidepressant venlafaxine (Fig. 2B). Indeed, it has been proposed that donepezil can be used to treat depression (*25*). Although it is still unclear whether the activities we found are sufficient for direct medical applications, the respective drugs certainly can be used as leads for further optimization toward new targets (*26*–*28*).

Many aspects of the current method can be improved (*15*); for example, the inference of the shared target between drug pairs involves a manual step, and we also cannot relate the target to particular side effects. Yet, when taking into account each individual probability of sharing a drug target, the 1947 predicted drugs pairs with >25% probability (Fig. 1A) roughly translate into 860 true drug-drug relations, each implying at least one new off-target protein, more than two-thirds of them in distinct therapeutic categories. The numerous off-targets for marketed drugs suggest that many of them have a broader spectrum of targets with physiological relevance than expected.

The use of direct readouts (side effects) of a perturbed human system to reveal molecular drug-target interactions should be applicable in a number of ways. First and foremost, existing drugs could be routinely checked for additional hidden targets and potential use in different therapeutic categories. Newly uncovered off-target effects will provide insights into the molecular basis of the drug's side effects but will also increase the reference set, which, in turn, then helps improving the method. The strategy could also be used in a preclinical setting through integration of candidate drugs into the network presented here or through application to animal models.

References and Notes

1. D. C. Liebler, F. P. Guengerich, *Nat. Rev. Drug Discov.* **4**, 410 (2005).
2. J. Blagg, *Annu. Rep. Med. Chem.* **41**, 353 (2006).
3. S. Whitebread, J. Hamon, D. Bojanic, L. Urban, *Drug Discov. Today* **10**, 1421 (2005).
4. T. T. Ashburn, K. B. Thor, *Nat. Rev. Drug Discov.* **3**, 673 (2004).
5. K. Finlayson, H. J. Witchel, J. McCulloch, J. Sharkey, *Eur. J. Pharmacol.* **500**, 129 (2004).
6. A. F. Fliri, W. T. Loging, P. F. Thadeio, R. A. Volkmann, *Nat. Chem. Biol.* **1**, 389 (2005).
7. A. F. Fliri, W. T. Loging, R. A. Volkmann, *ChemMedChem* **2**, 1774 (2007).
8. H. Gohlke, G. Klebe, *Angew. Chem. Int. Ed. Engl.* **41**, 2644 (2002).
9. M. J. Keiser *et al.*, *Nat. Biotechnol.* **25**, 197 (2007).
10. D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat. Rev. Drug Discov.* **3**, 935 (2004).
11. M. L. MacDonald *et al.*, *Nat. Chem. Biol.* **2**, 329 (2006).
12. G. V. Paolini, R. H. Shapland, W. P. van Hoorn, J. S. Mason, A. L. Hopkins, *Nat. Biotechnol.* **24**, 805 (2006).
13. M. Kuhn, M. Campillos, P. Gonzalez, L. J. Jensen, P. Bork, *FEBS Lett.* **582**, 1283 (2008).
14. O. Bodenreider, *Nucleic Acids Res.* **32**, D267 (2004).
15. Materials and methods are available as supporting material on *Science* Online.
16. M. Gerstein, E. L. Sonnhammer, C. Chothia, *J. Mol. Biol.* **236**, 1067 (1994).
17. S. Gunther *et al.*, *Nucleic Acids Res.* **36**, D919 (2008).
18. D. S. Wishart *et al.*, *Nucleic Acids Res.* **34**, D668 (2006).
19. B. L. Roth, E. Lopez, S. Patel, W. K. Kroeze, *Neuroscientist* **6**, 252 (2000).
20. Y. C. Martin, J. L. Kofron, L. M. Traphagen, *J. Med. Chem.* **45**, 4350 (2002).
21. A. Schuffenhauer, P. Floersheim, P. Acklin, E. Jacoby, *J. Chem. Inf. Comput. Sci.* **43**, 391 (2003).
22. A. F. Fliri, W. T. Loging, P. F. Thadeio, R. A. Volkmann, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 261 (2005).
23. C. M. Krejsa *et al.*, *Curr. Opin. Drug Discov. Dev.* **6**, 470 (2003).
24. C. J. Lin, J. C. Yang, Y. S. Uang, H. D. Chern, T. H. Wang, *Pharmacotherapy* **23**, 711 (2003).
25. G. Chouinard, C.-S. Peretti, U.S. Patent Trade Office application 20070082928 (2007).
26. W. Guba *et al.*, *J. Med. Chem.* **50**, 6295 (2007).
27. R. E. Martin, L. G. Green, W. Guba, N. Kratochwil, A. Christ, *J. Med. Chem.* **50**, 6291 (2007).
28. J. Tsai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3041 (2008).

29. We are grateful to M. P. Costi, R. Wade, T. Schneider, and members of the Bork group for helpful discussions and critical reading of the manuscript. This work was funded by the Bundesministerium für Bildung und Forschung QuantPro (grant no. 0313831D). M.C., M.K., A.-C.G., L.J.J., and P.B. have filed U.S. patent applications 61/043,292 and 61/043,299 based on the work in this paper.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5886/263/DC1
Materials and Methods
SOM Text
Figs. S1 to S13
Tables S1 to S4
References

21 March 2008; accepted 9 June 2008
10.1126/science.1158140