

Integrating Systems Biology Sources Illuminates Drug Action

A Gottlieb¹ and RB Altman^{1,2}

There are significant gaps in our understanding of the pathways by which drugs act. This incomplete knowledge limits our ability to use mechanistic molecular information rationally to repurpose drugs, understand their side effects, and predict their interactions with other drugs. Here, we present DrugRouter, a novel method for generating drug-specific pathways of action by linking target genes, disease genes, and pharmacogenes using gene interaction networks. We construct pathways for more than a hundred drugs and show that the genes included in our pathways (i) co-occur with the query drug in the literature, (ii) significantly overlap or are adjacent to known drug-response pathways, and (iii) are adjacent to genes that are hits in genome-wide association studies assessing drug response. Finally, these computed pathways suggest novel drug-repositioning opportunities (e.g., statins for follicular thyroid cancer), gene–side effect associations, and gene–drug interactions. Thus, DrugRouter generates hypotheses about drug actions using systems biology data.

Pathways form the basis of our understanding of how cellular processes occur and provide a framework for inferring cellular phenotypes. Drug research and development has provided powerful medications during the past several decades.¹ However, our understanding of the therapeutic effects of the drugs, their side effects (SEs), and drug interactions is still limited by incomplete knowledge of the underlying cellular pathways through which drugs act. For many applications, including drug discovery, drug repurposing, and the definition of pharmacogenomic modulators, we need a molecular-level understanding of drug effects, and this is often either missing or incomplete.

We focus here on inferring the pathways of interacting biological macromolecules that modulate response to a drug. By generating hypotheses for drug-specific pathways, we reduce the search space and enable researchers to focus their experimental efforts on the most promising directions. The primary challenge for building accurate pathways is our inadequate understanding of gene interactions, both their locations and temporal dependencies. Thus, straightforward network algorithms applied to gene interaction data sets yield a very high rate of false positives when they are used to connect drug targets to the genes that produce the final phenotypes. Previous methods manage network noise (in the context of drug response) by using only curated cellular pathways from public databases^{2–4} or by constructing pathways using only short paths prevalent in multiple drugs.⁵ These methods tend to ignore cross-talk

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

- ✓ Although the targets of many drugs are known, as are the key genes that modulate drug response, we generally have incomplete knowledge of the molecular pathways by which drugs act.

WHAT QUESTION DID THIS STUDY ADDRESS?

- ✓ This study interrogates the pharmacodynamic mode of action of a drug. It further explores the applicability of this knowledge to applications such as drug repositioning and associating genes with side effects and with drug interactions.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

- ✓ This study generates testable hypotheses about pharmacodynamic pathways of drugs. Using these pathways, we suggest alternative indications (drug repositioning) and associate proteins with drug side effects and with drug–drug interactions.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS

- ✓ The pathways and gene associations produced in this study provide leads for new drug targets that may drive drug development. Pathway genes may also be candidates for novel pharmacogenes (genes modulating drug response). Finally, we suggest alternative therapeutic indications for approved drugs.

¹Department of Genetics, Stanford University, Stanford, California, USA.; ²Department of Bioengineering, Stanford University, Stanford, California, USA
Correspondence: A Gottlieb (assafigo@stanford.edu)

Received 4 December 2013; accepted 18 February 2014; advance online publication 19 March 2014. doi:10.1038/clpt.2014.51

between pathways or focus only on pathways that are common to multiple drugs.

Here, we borrow an analogy from roads and traffic in which gene interactions (protein–protein, metabolic, and transcriptional interactions) are roads, and traversing the network is akin to finding the quickest route between points of interest. Network interactions that are part of a curated biological pathway have higher credibility than other gene interactions and are considered “highways.” The less-reliable and uncurated connections are viewed as “side roads.” Our method, DrugRouter, adopts a conservative strategy that assembles drug-specific pathways in which “highways” are used preferentially and “side roads” are used only when the highways do not connect the desired starting and ending points. The inputs to our method are genes and gene products (henceforth called genes for brevity) of three classes related to a particular drug of interest: (i) the drug’s target genes, (ii) the drug’s pharmacogenes that are known to modulate its mechanism of action (i.e., genes whose variation influences drug response), and (iii) the genes associated with the drug’s therapeutic effect or disease target. DrugRouter selects robust paths that connect these three sets of genes to one another; the genes that are visited during this “tour” are then assumed to be relevant to the molecular response to a drug. We focus on the action of drugs (pharmacodynamics (PD)) and not their metabolism (pharmacokinetics, which is also an important area) by excluding pharmacokinetic genes before applying our algorithms. **Figure 1** illustrates the steps in our method.

We show that the pathways that we construct are useful for four applications: (i) elucidating drug-specific PD pathways, (ii) suggesting alternative indications for a drug (drug repositioning), (iii) associating genes with drug SEs, and (iv) associating genes with drug–drug interactions (DDIs). We validate each of these applications independently.

RESULTS

Drug PD pathways as perturbed cellular pathways

A key assumption of our method is that drug-related pathways of action are chiefly drawn from existing knowledge of biology

and do not represent uncharacterized cellular interactions. Existing pathway databases reflect current knowledge of cellular mechanisms that are studied for many reasons, including their relevance to basic metabolism and disease processes. It is important for us to demonstrate that these databases contain useful and relevant knowledge for inferring the mechanism of action of drugs. Indeed, known PD pathways display significant overlap with other curated cellular pathways (**Supplementary Material** online, section 1).

Building drug-specific PD pathways

We constructed pathways for 113 drugs with at least one known drug target, pharmacogene, and disease-associated gene (Methods; **Supplementary Material** online, **Supplementary Data File S1** online). The pathways display high variability in size (90 ± 75 genes on average). Our pathway computations are robust: when we removed individual pharmacogenes or disease-associated genes, we found little overall change in the inferred pathways (see Building Drug-Specific PD Pathways section of the Methods and **Figure 2**).

We clustered the drugs according to the overlap of the genes in their computed pathways (Building Drug-Specific PD Pathways section of the Methods). **Figure 3** displays clusters of minimal size four, spanning 67 drugs. As expected, the clusters reflect similar or related therapeutic families. We provide an illustration and analysis of a small-sized pathway of tiludronate in the **Supplementary Material** online (section 2 and **Supplementary Figure S1** online).

We demonstrate the validity of our pathways with three observations made regarding the inferred genes of the pathway (genes in the drug-specific pathway, excluding the drug targets, pharmacogenes, and disease genes): (i) genes in our drug-specific pathway have significant co-occurrence in the literature; (ii) our drug-specific pathways have significant overlap with available gold standard PD pathways, and (iii) our pathways are enriched for genes that are hits in drug-based genome-wide association studies of warfarin, paclitaxel, and gemcitabine. To assess the

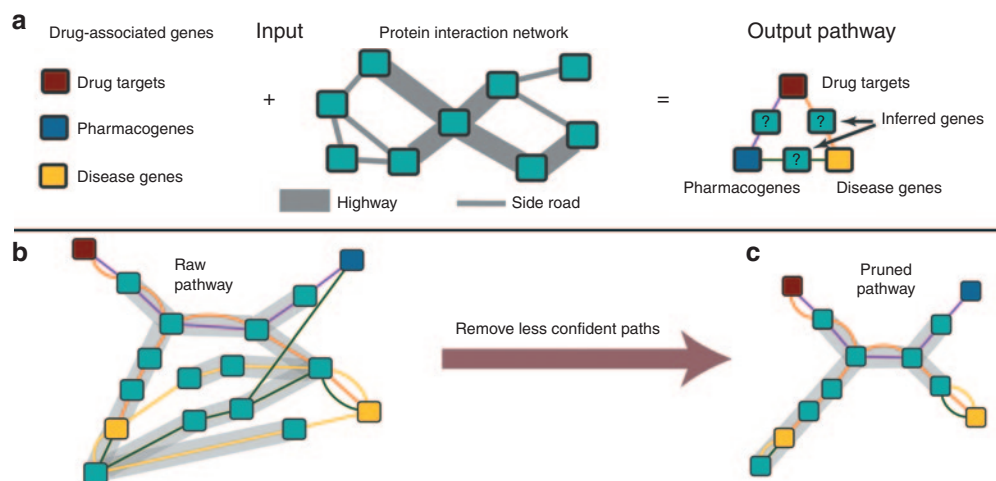


Figure 1 Illustration of the DrugRouter method. **(a)** Method input, including drug targets, pharmacogenes, and disease genes, and the network constructed of highways and side roads; **(b)** building the raw pathway, connecting drug targets, pharmacogenes, and disease genes; and **(c)** pruning the raw pathway by retaining only paths traversed by more than one color.

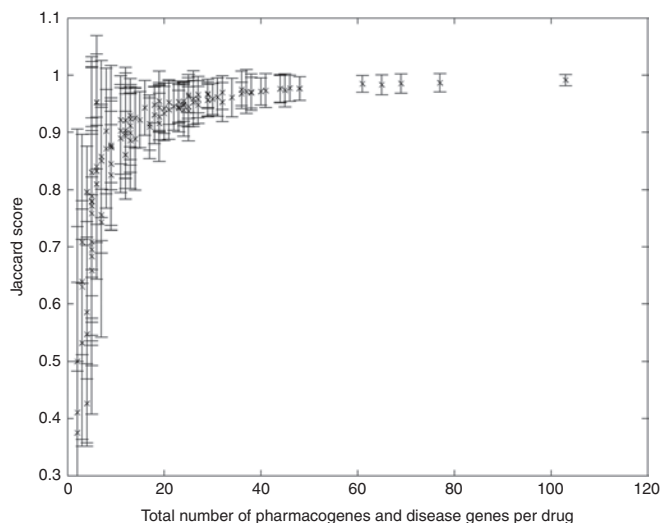


Figure 2 Robustness of inferred pathways to removal of pharmacogenes and disease genes. The mean and SD of the Jaccard score between the genes in the inferred pathway and in the leave-one-out pathway (intersection size divided by union size) as a function of the total number of pharmacogenes and disease genes per drug.

significance of these measures, we compared the drug-specific pathways against a set of randomly constructed control pathways (Building Drug-Specific PD Pathways section of the Methods).

For obtaining the support of literature co-occurrences, we queried PubMed and PubMed Central for textual co-occurrences of drugs and genes (Building Drug-Specific PD Pathways section of the Methods). We found that the inferred genes significantly co-occur with the corresponding drug in comparison with the unrelated control genes (Wilcoxon rank-sum test, 109/113 drugs—a false-discovery rate (FDR) < 0.05). The obtained P values were also lower than the P values obtained for other pathways (88% of the pathways) and random pathways (Wilcoxon rank-sum test, $P < e^{-120}$), in which 87% of the individual pathways performed better than random pathways and the remaining 15 pathways were all small (14 ± 10 inferred pathway genes, as compared with 75 ± 59 for other drug-specific pathways ($P < 2e^{-7}$); see also Discussion).

As a second verification, comparison with the gold standard PD pathway, we focused on 43 drugs for which 21 PD pathways are available in the Pharmacogenomics Knowledge Base (PharmGKB). Twenty-three of the drug-specific pathways (corresponding to 11 known PD pathways) were enriched for

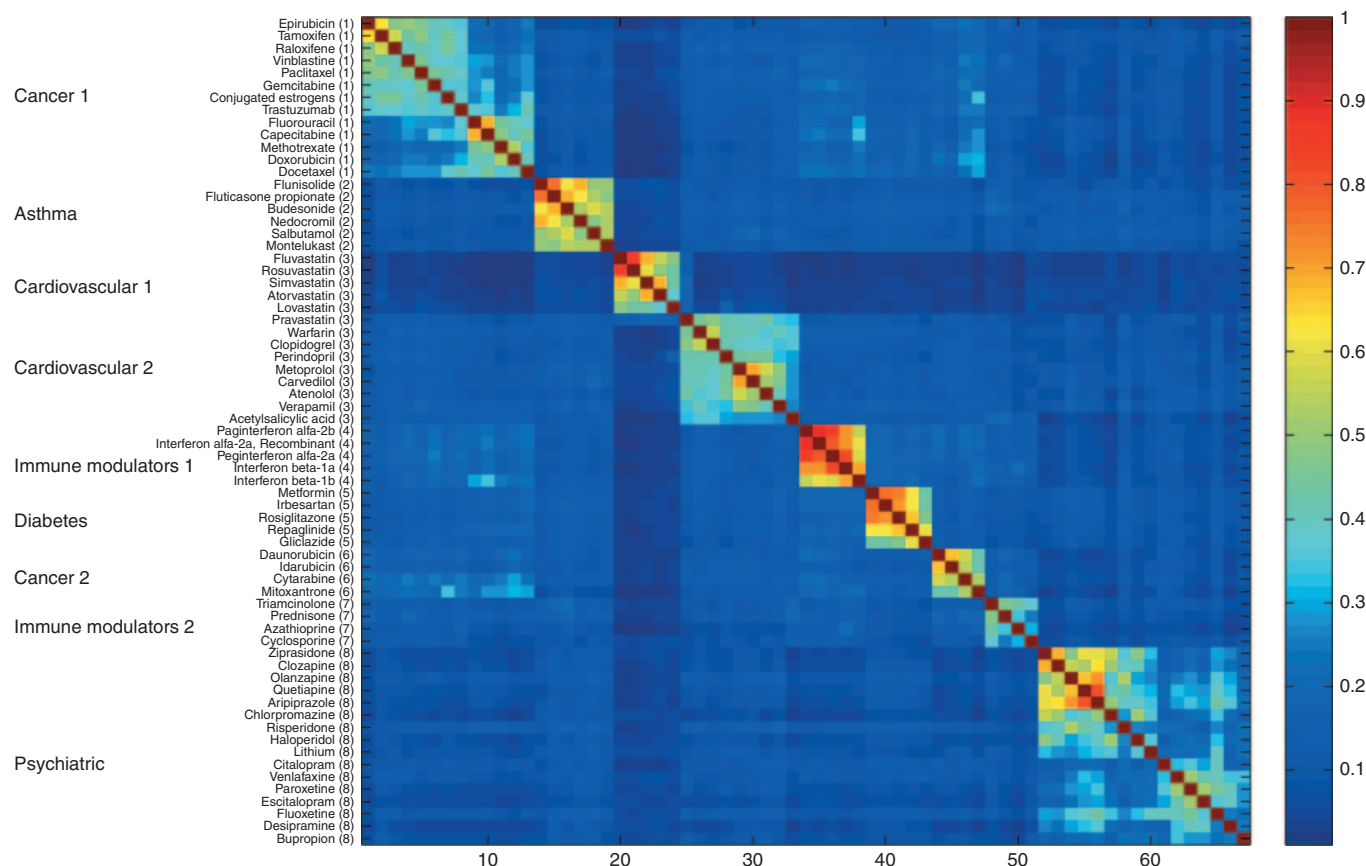


Figure 3 Clustering of drugs based on similarity of the drug-specific pathways. The x-axis and y-axis show the same 67 clustered drugs (same ordering for both axes). Colors correspond to the relative overlap of the constructed pathways for each drug pair (Jaccard score). The numbers in parentheses beside the drug names correspond to the cluster numbers. Labels on the left are “rough guides” based on the plurality of drugs in that cluster, but we acknowledge that some drugs may be in unusual clusters and this represents interesting hypotheses about their pathway connections that may deserve follow-up.

overlapping genes (hypergeometric enrichment, $FDR < 0.05$). Moreover, when we measured the distance of inferred genes in the drug-specific pathways to the known PD pathways, the remaining 20 drug-specific networks were significantly closer (in terms of network distance) to the known PD pathways than were the random control pathways ($FDR < 0.05$).

The third evaluation, focused on genome-wide association studies for warfarin,⁶ paclitaxel,⁷ and gemcitabine,⁸ is described in the **Supplementary Material** online (section 3). In addition, we describe associations between genes modulating response to doxorubicin in yeast and the doxorubicin-specific pathway genes in the **Supplementary Material** online (section 4).

Applying inferred pathways to suggest drug repositioning

We can use the drug-specific pathways to propose new indications for existing drugs. Using two independent criteria, we obtained 1,484 pairs of novel drug–disease pairs using the first criterion and 1,348 pairs using the second criterion (Applying Inferred Pathways to Suggest Drug Repositioning section of the Methods). Of these pairs, 195 satisfied both criteria. These predictions associate 113 drugs with 139 diseases (**Supplementary Table S1** online).

We evaluated our repositioning predictions by computing enrichment of the predictions in (i) current clinical trials in phases I–III and (ii) off-label uses extracted from electronic health records (Applying Inferred Pathways to Suggest Drug Repositioning section of the Methods). Between 15 and 27% of the experimental (i.e., not yet approved) drug and disease associations being tested in clinical trials satisfy the first and second criteria, respectively (hypergeometric test, $P < 2e^{-7}$ and $P < 2e^{-36}$, respectively). Similarly, between 19 and 21% of off-label drug uses, extracted from electronic health records,⁹ satisfy one of the two criteria (hypergeometric test, $P < 5e^{-4}$ and $P < 0.02$, respectively). Finally, our predictions were enriched in a “silver standard” set of drug–disease associations ($P < 0.02$ and $P < 0.05$ for the two criteria, respectively) (Applying Inferred Pathways to Suggest Drug Repositioning section of the Methods). Notably, our predictions were also enriched with the phenotype-driven repositioning prediction set published in ref. 10 ($P < 0.04$ and $P < 2e^{-7}$, respectively).

Another evaluation of our repositioning focused on repositioning of cancer drugs to other cancer types. We downloaded bioassays from PubChem¹¹ for 34 cancer drugs and mapped the bioassays to 30 cancer diseases (**Supplementary Methods** online, section 2). Aggregating the number of active and inactive bioassays across all studies and excluding inconclusive treatments resulted in 361 drug–disease pairs in which the drug was exclusively active or inactive. We found that only the second criterion received significant results (hypergeometric test, $P < 0.05$; odds ratio = 1.57).

Among the predicted drug-repurposing opportunities, we highlight five that were predicted using both criteria. First, conjugated estrogens and verapamil are predicted to affect type 2 diabetes mellitus. Indeed, conjugated estrogens improve glycaemic control in postmenopausal women with type 2 diabetes,¹² and short-term verapamil decreases fasting plasma glucose and

glucose turnover in non-insulin-dependent diabetic patients, possibly by inhibition of gluconeogenesis.¹³ Second, thalidomide may be useful against Alzheimer’s disease, in part through its effect on the Alzheimer’s disease–associated genes encoding for amyloid precursor protein (APP) and amyloid- β A4 precursor protein-binding family B member 2. Indeed, a recent study showed that long-term treatment with thalidomide may treat a mouse model of Alzheimer’s disease through inhibition of the β -site APP-cleaving enzyme 1.¹⁴ Third, nonsteroidal anti-inflammatory drugs, including aspirin and the withdrawn drug rofecoxib, may have a role in colon cancer. Indeed, aspirin was reported to reduce the risk of colon cancer,¹⁵ while rofecoxib plays a chemopreventive role in a mouse model of colon cancer.¹⁶ The last two examples show evidence of a drug–disease association, but the type of association (treatment or causal) is less evident at this time: (i) valproic acid may be associated with acute leukemia. Indeed, a study reported this association with a proposed mechanism involving inhibition of histone deacetylase,¹⁷ whereas a case report for three patients claimed that valproic acid causes leukemia through the same mechanism¹⁸; and (ii) statins on our list (simvastatin, lovastatin, pravastatin, atorvastatin, and fluvastatin) are associated with follicular thyroid cancer. Although some studies reported that lovastatin induces apoptosis in a different thyroid cancer—human anaplastic thyroid carcinoma,^{19,20} a single case study reported the development of thyroid follicular adenoma on simvastatin therapy²¹ and another study reported the development of thyroid neoplasms at high dosages in rats.²² A 9-year follow-up yielded inconsistent findings, according to which increased risk of thyroid cancer in men after 5 years of statin use was not supported by the 2-year-lag results or by corresponding findings in women.²³

Several drugs predicted to interact with diseases may not treat them but may instead cause them or increase their severity. Aspirin (acetylsalicylic acid) induces asthma,²⁴ and calcium or paclitaxel may increase the risk of myocardial infarction.^{25,26} These predictions, although not leading to new repositioning, may be, however, useful in exposing the potential molecular mechanisms underlying these disorders.

Inferred pathways suggest novel associations among genes, SEs, and drug interactions

In a similar manner to that used for the drug-repositioning process, the existence of a gene associated with an SE on a drug pathway may indicate that the drug induces the SE. Using a literature-curated list of gene–SE associations, 26% of the predicted SEs are known drug SEs (hypergeometric test, $P < 3e^{-15}$) (**Supplementary Material** online, section 6 and **Supplementary Table S2** online).

Because the set of known genes associated with SEs is currently limited, our drug-specific pathways can suggest novel associations between genes and potential SEs. We computed the enrichment of genes within their pathways for 764 SEs associated with the 113 drugs, accounting for similar drugs. We obtained a final list of 135 gene–SE associations, spanning 33 genes and 50 SEs (**Supplementary Material** online, section 6; **Supplementary Table S3** online; and **Supplementary**

Figure S2 online). Finally, our drug-specific pathways can associate genes with PD drug interactions. We found 15 genes enriched for co-occurrence in drug-specific pathways of severely interacting drugs (Supplementary Material online, section 7, Supplementary Table S4 online, and Supplementary Figure S3 online).

We evaluated our gene–SE predictions by querying PubMed and PubMed Central for associations between 28 SEs from our prediction set and the genes in our network. We evaluated our gene–DDI predictions by querying PubMed and PubMed Central for the comentioning of the genes with all possible drug combinations. Nineteen SEs (of 28) were significantly comentioned with the predicted associated genes, and 10 (of 15) genes had more frequent comentioning with severely interacting drug pairs than with noninteracting drug pairs (FDR < 0.05 for both). We provide detailed analysis and highlight examples in the Supplementary Material online, sections 6 and 7 and Supplementary Table S5.

DISCUSSION

In this article, we introduce a novel method for inferring drug-specific pathways. We connect known drug-associated genes (drug targets, pharmacogenes, and genes associated with diseases treated by the drug) over protein, metabolic, and transcriptional interaction networks while preferring high-confidence interactions participating in curated cellular processes. In that sense, our method is conservative and unlikely to propose radically new pathways unless the high-throughput evidence is very strong. On evaluation of our constructed pathways, we were able to suggest novel drug repositioning, associate genes with SEs, and suggest potential causes for DDIs.

To reduce the inherent noise in biological networks, our algorithm searches for the most confident paths, using preferentially high-confidence “highways” and maintaining only highly traversed paths. Our conservative approach emphasizes precision to control for false-positive results. However, we might miss additional genes that reside either on less-traversed pathways or on “side roads” with these requirements. This property may have caused the genes on 15 small pathways to have fewer comentions with the drug than the genes in the random pathways. In addition, most protein–protein interactions (PPIs) lack directionality and sign (activation/inhibition). Incorporating such additional information into the model would enhance the pathway construction task.

Not all the drug-specific pathways were enriched in curated PD pathways, but they were significantly closer to those curated PD pathways on the interaction network. The human curation requirements enforced by PharmGKB curators lead to a very high specificity of these curated pathways but low sensitivity, potentially missing several parts with lower evidence levels.

We assumed that discovery of new disease genes or drug target genes along a pathway would imply a new drug-repositioning opportunity. However, we disregarded some factors that may prevent such an opportunity from materializing, such as the actual role of those disease or drug target genes in the treatment (e.g., activation vs. inhibition) and the expression of those genes

in a given tissue. Indeed, some of our repositioning suggestions were found to induce or elevate the disease risk.

As noted by several authors,^{27,28} drug discovery is quickly moving from the single-gene research paradigm to the systems biology analysis paradigm. DrugRouter represents a general-purpose tool to harness pathway information for multiple uses.

METHODS

Data sets

Gene interaction network. PPIs were assembled from the union of BioGrid version 3.1.94,²⁹ Database of Interacting Proteins (August 2012),³⁰ Human Protein Reference Database release 9,³¹ IntAct (October 2012),³² Molecular Interaction Database (October 2012),³³ Mammalian Protein-Protein Interaction,³⁴ and Human Integrated Protein-Protein Interaction Reference version 1.4.³⁵ Curated PPIs were extracted from Kyoto Encyclopedia of Genes and Genomes (May 2012)³⁶ human signaling pathways. Metabolic interactions between enzymes were extracted from Kyoto Encyclopedia of Genes and Genomes³⁶ human metabolic pathways. Protein–gene (transcriptional) interactions were retrieved from the Chip Enrichment Analysis database.³⁷ The network includes more than 223,000 interactions.

Pathways. Pathways were imported from the Pathway Interaction Database,³⁸ which includes the National Cancer Institute–Nature curated human pathways, and selected pathways from Reactome³⁹ and BioCarta⁴⁰ (1,331 pathways). PD pathways were downloaded from PharmGKB⁴¹ (41 pathways).

Drug-specific genes. Drug targets were downloaded from DrugBank.⁴² Drug sensitivity variants were retrieved from PharmGKB⁴¹ (Supplementary Data File S2 online).

Drug–disease genes. Genes associated with drug indications were assembled by mapping disease-associated genes retrieved from the Online Mendelian Inheritance in Man database (April 2013)⁴³ to drug indications retrieved from ref. 44. On average, 18.5 ± 18.3 disease genes were assigned to a drug (Supplementary Data File S2 online).

Drug SEs and drug interactions. Drug–SE associations were downloaded from the SIDER2 database.⁴⁵ DDIs were retrieved from DrugBank⁴² and the drugs.com website⁴⁶ as described in ref. 10.

Building the network of highways and side roads. We constructed a network by integrating three types of interactions: (i) PPIs, (ii) metabolic interactions, in which enzymes are connected via a mutual metabolite, and (iii) transcriptional interactions, in which a transcription factor is connected to a transcribed gene. Each network interaction was tagged as “highway” if it was (i) a PPI appearing in a KEGG signaling pathway, (ii) a PPI in which the two interacting genes appear in the same curated cellular pathway, or (iii) a curated metabolic interaction from KEGG metabolic pathways. The remaining interactions were tagged as “side roads.” Transcriptional interactions and “highway” interactions from KEGG are directed, and the remaining interactions are undirected (Supplementary Data File S2 online).

The DrugRouter algorithm. DrugRouter constructs pathways in two consecutive stages: (i) a construction stage and (ii) a pruning stage.

The first stage, construction, connects pairs of starting and destination points (e.g., a drug target as a starting point and a pharmacogene as the destination point). We included five different start–destination pair types: (i) Drug targets to pharmacogenes, (ii) drug targets to disease genes, (iii) pharmacogenes to pharmacogenes, (iv) pharmacogenes to disease genes, and (v) disease genes to disease genes.

For each pair type, we connected all the start and destination pairs by applying three steps. (i) Locate all the nearest highway entry points (“on-ramps”), i.e., minimal number of side roads between the start point and the highway entry point. If the distance to the nearest “on-ramp” is farther than 1 SD above the mean network path length (more than three interactions in our network), that start point is excluded from the analysis. (ii) Locate all the highway exit points (“off-ramps”) nearest to the destination

point in a similar manner to step 1. (iii) Find the shortest paths between the “on-ramps” and “off-ramps.” If none of the “off-ramps” is reachable from the “on-ramps,” we allow the use of side roads by weighting each side road as high as 10 highway interactions. We include all equidistant shortest paths. Because the effect of a pharmacogene on the drug action is less likely to stem from transcriptional regulation of the pharmacogene, we ignored transcriptional regulation of the pharmacogene when it was a destination point. The effect of this decision was negligible, in line with the robustness of the pathways (Building Drug-Specific PD Pathways section of the Methods; Figure 2). The result of the construction stage is the union of all the start–destination routes.

In the second, pruning, stage, we applied a conservative approach in which we retained only higher-confidence interactions that are traversed by tours from more than one of the five start–destination pair types (e.g., tours that connect a drug target and a pharmacogene and a tour that connects a pharmacogene to a disease gene). If, following this pruning, a pharmacogene or disease gene becomes disconnected, we also retain all the shortest routes, discovered in the construction stage, that connect the drug targets to that pharmacogene or disease gene.

Building drug-specific PD pathways. Some of the drugs in our set of 113 belong to the same drug family. However, no two drugs shared the exact set of inputs (drug targets, pharmacogenes, and disease genes). Specifically, more than 90% of the drugs share less than 80% identity in their associated gene set, and more than 90% of the drugs have a nonredundant chemical structure, with a Tanimoto coefficient lower than 0.7.

To simulate the construction of *de novo* pathways, when building pathways for drugs that have a gold standard curated PD pathway, we converted the highways that are specific to that PD pathway to side roads.

To test the robustness of the drug-specific pathways, we systematically constructed the drug-specific pathway after removal of each of the pharmacogenes or disease genes. We observed high robustness in terms of inferred genes or inferred interactions. The Jaccard scores⁴⁷ between the sets of genes or interactions in the drug-specific pathways and the leave-one-out pathways (computed as the size of the intersections between the two sets divided by the size of the union of the two sets) were 0.87 ± 0.08 and 0.85 ± 0.08 , respectively. As anticipated, the greater the number of drug-associated pharmacogenes and disease genes, the more robust is the drug-specific pathway to their removal (Figure 2).

We performed biclustering of the drugs by computing the relative overlap of the genes in the inferred pathways of each pair of drugs (Jaccard score), using a modification of the spectral coclustering algorithm from ref. 48 whereby each bicluster is further clustered by single-linkage hierarchical clustering.

For evaluation purposes, we constructed 200 random pathways per drug by connecting the known drug targets with the same number of randomly picked genes from the interaction network as the set of known pharmacogenes and disease genes. We controlled for the randomly selected genes in three ways: (i) random shuffling of the known pharmacogenes and disease genes (100 pathways); (ii) maintaining the same network degree distribution as the network degree distribution of the true pharmacogenes and disease genes of that drug (50 random pathways); and (iii) maintaining the same distribution of network distances between the random pharmacogenes and disease genes as the network distances between the true ones (50 random pathways). We required the shuffled and randomly picked genes to be distant by at least the mean network path (more than two edges) from the true pharmacogenes and disease genes. The difference in performance between each type of random pathway was negligible (1% difference in the number of significant drugs in PubMed test and no difference in the enrichment against curated PD pathways).

For the first support of literature co-occurrences, we queried PubMed abstracts and PubMed Central whole articles using the PubMed E-Utilities interface for all pairwise combinations of 1 of the 113 drugs (using generic names) and 1 of the 19,176 genes in the gene network (official gene symbols). We manually excluded 63 ambiguous gene

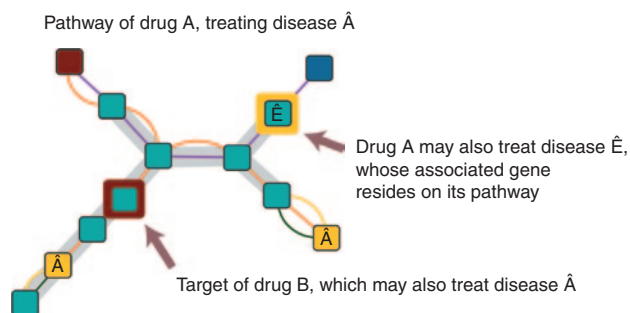


Figure 4 The drug-repositioning prediction scheme. A repositioning of drug A is suggested if (i) a drug target of drug A is found along the path between the drug target and disease genes of drug B or (ii) a disease gene unrelated to drug A is found within the pathway of drug A.

names (e.g., a known English word such as “rest” or “tag,” or a prevalent abbreviation such as “ORF” or “PDF”). Our statistics included rank-sum comparison of the number of literature co-occurrences of the drug and the drug-specific inferred genes to (i) co-occurrences of all other genes (P values for all the drugs with $FDR < 0.05$), (ii) unassociated genes involved in the curated cellular pathways (109/113 drugs with $FDR < 0.05$), (iii) all inferred genes from other pathways, or (iv) the inferred genes from the random pathways constructed for that drug. The rank-sum P values were lower than P values obtained upon shuffling of gene assignments to drug-specific pathways.

For drug-specific pathways that were not enriched with the gold standard PD pathways, we measured the average distance between the PD pathway and the drug-specific pathway over the network (ignoring the distinction between highways and side roads), in comparison with the distance between the PD pathway and the random pathways. As expected, switching roles between highways and side roads resulted in only four enriched known PD pathways.

Applying inferred pathways to suggest drug repositioning. We considered two potential criteria for applying the drug-specific pathways for drug repositioning. A drug-repositioning opportunity for drug A is found when (i) a known target of a drug A appears along the path connecting the known target and disease gene of drug B (drug A may treat the disease that drug B treats) or (ii) a known disease gene appears in the pathway computed for drug A (drug A may treat this newly implicated disease) (Figure 4).

For verification of our drug-repositioning predictions, we downloaded data on clinical trials in phases I–III up to 28 June 2013 from the clinicaltrials.gov website. Drug names were matched to generic, synonymous, and brand names in DrugBank. Condition names were converted to Online Mendelian Inheritance in Man disease names using the MetaMap tool⁴⁹ and the filtering operations described in ref. 44. When a drug–disease pair appears in more than one phase, we chose the highest phase. Overall, we obtained 410 unique drug–disease pairs that involve both drugs and diseases in our prediction set.

Off-label drug indications, as well as approved indications that do not appear in our strict gold standard, were obtained from electronic health records from Stanford Hospital.

We found additional support by comparing the results with those from a “silver standard” set of drug–disease associations. As described in ref. 10, the authors constructed drug–disease associations from four independent sources. Two of the sources, based on extraction from textual indications, were noisier and required additional evidence to be included in the gold standard. However, they used the remaining set of single-evidence associations as a “silver standard” for verification purposes.

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at <http://www.nature.com/cpt>

ACKNOWLEDGMENTS

R.B.A. and A.G. are supported by National Institutes of Health grants LM05652, GM102365, and GM61374. We thank the PharmGKB team for providing the pharmacogenomic data, Deanna Kroetz for providing data related to the paclitaxel genome-wide association study, and Ken Jung and Nigam Shah for providing the off-label uses of the drugs.

AUTHOR CONTRIBUTIONS

A.G. designed the research, performed the research, and analyzed the data. A.G. and R.B.A. wrote the manuscript.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

© 2014 American Society for Clinical Pharmacology and Therapeutics

- Lu, J.J., Pan, W., Hu, Y.J. & Wang, Y.T. Multi-target drugs: the trend of drug research and development. *PLoS One* **7**, e40262 (2012).
- Huang, R., Wallqvist, A., Thanki, N. & Covell, D.G. Linking pathway gene expressions to the growth inhibition response from the National Cancer Institute's anticancer screen and drug mechanism of action. *Pharmacogenomics J.* **5**, 381–399 (2005).
- Chen, X. *et al.* A sub-pathway-based approach for identifying drug response principal network. *Bioinformatics* **27**, 649–654 (2011).
- Ye, H., Yang, L., Cao, Z., Tang, K. & Li, Y. A pathway profile-based method for drug repositioning. *Chinese Science Bulletin* **57**, 2106–12 (2012).
- Silberberg, Y., Gottlieb, A., Kupiec, M., Rupp, E. & Sharan, R. Large-scale elucidation of drug response pathways in humans. *J. Comput. Biol.* **19**, 163–174 (2012).
- Cooper, G.M. *et al.* A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* **112**, 1022–1027 (2008).
- Baldwin, R.M. *et al.* A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. *Clin. Cancer Res.* **18**, 5099–5109 (2012).
- Innocenti, F. *et al.* A genome-wide association study of overall survival in pancreatic cancer patients treated with gemcitabine in CALGB 80303. *Clin. Cancer Res.* **18**, 577–584 (2012).
- Jung, K. *et al.* Automated detection of off-label drug use. *PLoS One*, **9**(2), e89324 (2014).
- Gottlieb, A., Stein, G.Y., Oron, Y., Rupp, E. & Sharan, R. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol. Syst. Biol.* **8**, 592 (2012).
- Wang, Y. *et al.* An overview of the PubChem BioAssay resource. *Nucleic Acids Res.* **38**, D255–D266 (2010).
- Friday, K.E., Dong, C. & Fontenot, R.U. Conjugated equine estrogen improves glycemic control and blood lipoproteins in postmenopausal women with type 2 diabetes. *J. Clin. Endocrinol. Metab.* **86**, 48–52 (2001).
- Busch Sørensen, M., Sjøstrand, H., Sengeløv, H., Tiefenthal Thrane, M., Juul Holst, J. & Lyngsøe, J. Influence of short term verapamil treatment on glucose metabolism in patients with non-insulin dependent diabetes mellitus. *Eur. J. Clin. Pharmacol.* **41**, 401–404 (1991).
- He, P., Cheng, X., Staufienbiel, M., Li, R. & Shen, Y. Long-term treatment of thalidomide ameliorates amyloid-like pathology through inhibition of β -secretase in a mouse model of Alzheimer's disease. *PLoS One* **8**, e55091 (2013).
- Thun, M.J., Namboodiri, M.M. & Heath, C.W. Jr. Aspirin use and reduced risk of fatal colon cancer. *N. Engl. J. Med.* **325**, 1593–1596 (1991).
- Evans, J.F. Rofecoxib (Vioxx), a specific cyclooxygenase-2 inhibitor, is chemopreventive in a mouse model of colon cancer. *Am. J. Clin. Oncol.* **26**, S62–S65 (2003).
- Kuendgen, A. & Gattermann, N. Valproic acid for the treatment of myeloid malignancies. *Cancer* **110**, 943–954 (2007).
- Coyle, T.E., Bair, A.K., Stein, C., Vajpayee, N., Mehdi, S. & Wright, J. Acute leukemia associated with valproic acid treatment: a novel mechanism for leukemogenesis? *Am. J. Hematol.* **78**, 256–260 (2005).
- Wang, C.Y., Zhong, W.B., Chang, T.C., Lai, S.M. & Tsai, Y.F. Lovastatin, a 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor, induces apoptosis and differentiation in human anaplastic thyroid carcinoma cells. *J. Clin. Endocrinol. Metab.* **88**, 3021–3026 (2003).
- Zhong, W.B., Hsu, S.P., Ho, P.Y., Liang, Y.C., Chang, T.C. & Lee, W.S. Lovastatin inhibits proliferation of anaplastic thyroid cancer cells through up-regulation of p27 by interfering with the Rho/ROCK-mediated pathway. *Biochem. Pharmacol.* **82**, 1663–1672 (2011).
- McCord, E.L. & Goenka, S. Development of thyroid follicular adenoma on simvastatin therapy. *Tenn. Med.* **93**, 210–212 (2000).
- Robison, R.L., Suter, W. & Cox, R.H. Carcinogenicity and mutagenicity studies with fluvastatin, a new, entirely synthetic HMG-CoA reductase inhibitor. *Fundam. Appl. Toxicol.* **23**, 9–20 (1994).
- Friedman, G.D., Flick, E.D., Udaltsova, N., Chan, J., Quesenberry, C.P. Jr & Habel, L.A. Screening statins for possible carcinogenic risk: up to 9 years of follow-up of 361,859 recipients. *Pharmacoepidemiol. Drug Saf.* **17**, 27–36 (2008).
- Hamad, A.M., Sutcliffe, A.M. & Knox, A.J. Aspirin-induced asthma: clinical aspects, pathogenesis and management. *Drugs* **64**, 2417–2432 (2004).
- Bolland, M.J. *et al.* Effect of calcium supplements on risk of myocardial infarction and cardiovascular events: meta-analysis. *BMJ* **341**, c3691 (2010).
- Shah, K., Gupta, S., Ghosh, J., Bajpai, J. & Maheshwari, A. Acute non-ST elevation myocardial infarction following paclitaxel administration for ovarian carcinoma: a case report and review of literature. *J. Cancer Res. Ther.* **8**, 442–444 (2012).
- Kitano, H. A robustness-based approach to systems-oriented drug design. *Nature Reviews Drug Discovery* **5**, 202–210 (2007).
- Csermely, P., Korcsmáros, T., Kiss, H.J., London, G. & Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* **138**, 333–408 (2013).
- Breitkreutz, B.J. *et al.* The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.* **36**, D637–D640 (2008).
- Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M. & Eisenberg, D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**, 303–305 (2002).
- Prasad, T.S.K. *et al.* Human protein reference database—2009 update. *Nucleic Acids Res.* **37**, D767–D772 (2009).
- Kerrien, S. *et al.* The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40**, D841–D846 (2012).
- Chatr-aryamontri, A. *et al.* MINT: the Molecular INTeraction database. *Nucleic Acids Res.* **35**, D572–D574 (2007).
- Mewes, H.W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30**, 31–34 (2002).
- Schaefer, M.H., Fontaine, J.F., Vinayagam, A., Porras, P., Wanker, E.E. & Andrade-Navarro, M.A. HIPPIE: Integrating protein interaction networks with experiment based quality scores. *PLoS One* **7**, e31826 (2012).
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360 (2010).
- Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R. & Ma'ayan, A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
- Schaefer, C.F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res.* **37**, D674–D679 (2009).
- Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622 (2009).
- Nishimura, D. BioCarta. *Biotech Software & Internet Report* **2**, 117–120 (2001).
- Klein, T.E. *et al.* Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J.* **1**, 167–170 (2001).
- Wishart, D.S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36**, D901–D906 (2008).
- Hamosh, A., Scott, A.F., Amberger, J., Bocchini, C., Valle, D. & McKusick, V.A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **30**, 52–55 (2002).
- Gottlieb, A., Stein, G.Y., Rupp, E. & Sharan, R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **7**, 496 (2011).
- 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).
- Drugsite Trust. Drugs.com™: drug interactions checker. (cited 3 July 2011). <http://www.drugs.com/drug_interactions.html>.
- Jaccard, P. Nouvelles recherches sur la distribution florale. *Bul Soc Vaudoise Sci Nat* **44**, 223–270 (1908).
- Dhillon, I.S. Co-clustering documents and words using bipartite spectral graph partitioning. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, 26–29 August 2001. 269–274 (2001).
- Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* **32**, D267–D270 (2004).