DRUG DISCOVERY

Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data

Marina Sirota,[1,2,3]* Joel T. Dudley,[1,2,3]* Jeewon Kim,[4] Annie P. Chiang,[1,2,3] Alex A. Morgan,[1,2,3] Alejandro Sweet-Cordero,[1,5] Julien Sage,[1,5,6] Atul J. Butte[1,3,5]†

Published 17 August 2011; revised 28 September 2011

The application of established drug compounds to new therapeutic indications, known as drug repositioning, offers several advantages over traditional drug development, including reduced development costs and shorter paths to approval. Recent approaches to drug repositioning use high-throughput experimental approaches to assess a compound's potential therapeutic qualities. Here, we present a systematic computational approach to predict novel therapeutic indications on the basis of comprehensive testing of molecular signatures in drug-disease pairs. We integrated gene expression measurements from 100 diseases and gene expression measurements on 164 drug compounds, yielding predicted therapeutic potentials for these drugs. We recovered many known drug and disease relationships using computationally derived therapeutic potentials and also predict many new indications for these 164 drugs. We experimentally validated a prediction for the antiulcer drug cimetidine as a candidate therapeutic in the treatment of lung adenocarcinoma, and demonstrate its efficacy both in vitro and in vivo using mouse xenograft models. This computational method provides a systematic approach for repositioning established drugs to treat a wide range of human diseases.

INTRODUCTION

The identification of new disease indications for approved drugs, or drug repositioning, offers several advantages over traditional drug development (1). The traditional paradigm of drug discovery is generally regarded as protracted and costly, with studies showing that it takes about 15 years and more than $1 billion to develop and bring a new drug to market (2). A substantial portion of drug development costs are incurred during early development and toxicity testing, with more than 90% of drugs failing to move beyond these early stages (3). The repositioning of drugs already approved for human use mitigates the costs and risks associated with early stages of drug development and offers shorter routes to approval for therapeutic indications. Successful examples of drug repositioning include the indication of sildenafil for erectile dysfunction and pulmonary hypertension, thalidomide for severe erythema nodosum leprosum, and retinoic acid for acute promyelocytic leukemia (4).

The prevailing approach to drug repositioning is based on established tools and techniques developed for screening libraries of lead compounds against biological targets of interest in early-stage drug discovery. In the case of drug repositioning, libraries of approved drugs are screened across a broad set of putative biological targets with high-throughput screening (HTS) technologies that incorporate multiplex biological assays (1, 4). Computational approaches to the discovery of novel biological targets for approved drugs have been developed to overcome the high costs and other logistical limitations associated with experimental HTS approaches, allowing for greatly expanded searches for novel repositioning opportunities (5–10). Although these approaches provide comprehensive and systematic means to discover new biological targets for existing drugs, they do not address the challenge of moving from in silico or in vitro binding of a drug to a biological target to realized therapeutic use in affected patients.

Gene expression microarrays (11, 12) enable the measurement of genome-wide expression, and they are regularly and broadly applied in clinical studies of human diseases. Comparative gene expression analysis of primary affected, peripheral and secondary organs, tissues, and biofluids is used to study the molecular pathophysiology of a disease condition or to identify expression patterns that serve as prognostic or diagnostic indicators. By examining the sets of genes that are up- and down-regulated in a disease state compared to a normal state, it is possible to create a gene expression profile, or signature, of a disease (13–17). When derived from clinical samples, these signatures represent the accurate and consistent gene expression state imparted by immunological, metabolic, and other complex factors that make up the broad physiological manifestation of a disease (18, 19).

Microarrays are also used to discover gene expression patterns that signify the perturbation of biological systems by drug compounds (20–24). A collection of genome-wide transcriptional expression data from cultured human cells treated with a broad range of Food and Drug Administration (FDA)–approved bioactive small molecules (25) has been used to study the molecular basis of drug effects and particular diseases of interest like colon cancer (26) or prostate cancer (27).

Here, we performed a large-scale integration of expression signatures of human diseases from the public data with previously described drug-effect signatures (25). Using this approach, we built a compendium of disease-drug relationship predictions based on matching genome-wide signatures of disease pathophysiology and drug effect. Instead of examining a single drug-disease pair or even looking at reactions of a large set of drugs on a single disease, we focused on discovering connections between drugs and diseases across all the available gene measurements. Here, we present evidence that this strategy can confirm already known therapeutic uses for drugs and uncover new uses for FDA-approved drugs.

[1]Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, 251 Campus Drive, Stanford, CA 94305–5415, USA. [2]Training Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA 94305–5479, USA. [3]Lucile Packard Children's Hospital, 725 Welch Road, Palo Alto, CA 94304, USA. [4]Stanford Comprehensive Cancer Center, Stanford School of Medicine, Stanford, CA 94305, USA. [5]Cancer Biology Program, Stanford University School of Medicine, Stanford, CA 94305, USA. [6]Department of Genetics, Stanford University, Stanford, CA 94305, USA.
*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: abutte@stanford.edu
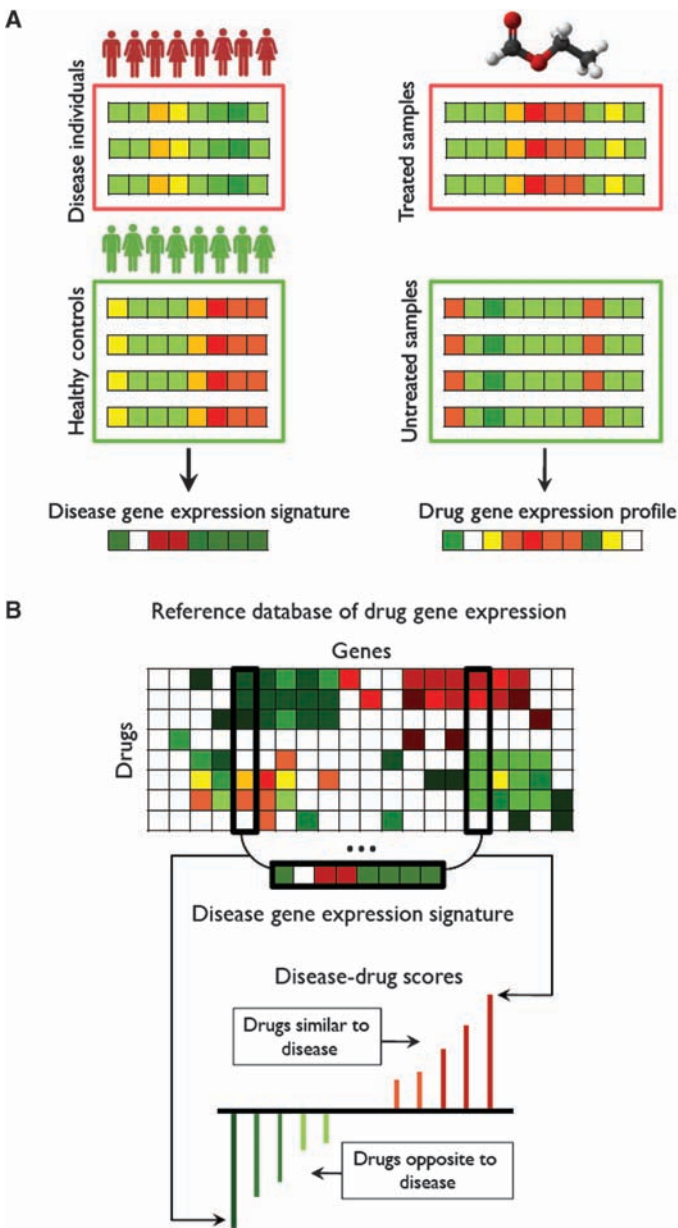
Fig. 1. Analytic workflow. (**A**) Two gene expression collections are used: a set of disease-associated gene expression data with corresponding controls and a set of gene expression data from tissue treated with drugs and small molecules with corresponding controls. SAM is used to obtain a signature of significantly up- and down-regulated genes for each disease. Rank normalization and the pre-processing procedure previously described (25) are used to create a reference database of drug gene expression. (**B**) A modification to the Connectivity Map method (25) is used to query the disease signature against the drug reference expression set to assign a drug-disease score to each drug-disease pair based on profile similarity. These scores are interpreted, resulting in a list of candidate therapeutics for each disease of interest.

RESULTS

We identified and combined data from publicly available microarray data sets from National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) representing 100 diseases with gene expression data from several human cancer cell lines treated with 164 drugs or small molecules to predict previously undescribed therapeutic drug-disease relationships. We then defined a disease signature for each of the 100 diseases as a set of genes that are significantly up- and down-regulated for that disease compared to normal values using significance analysis of microarrays (SAM) (19, 28). Next, we statistically compared each of the disease signatures to each of the reference drug expression signatures from the Connectivity Map (25) to infer possible therapeutic indications. A schematic overview of our methodology is shown in Fig. 1. We calculate a similarity score for every pairing of a drug and disease reflecting the similarity of the drug and disease signatures, ranging from +1, indicating a perfect correlation of signatures, to −1, indicating exactly opposite signatures. Our hypothesis is that if a disease state is signified by a specific set of genome-wide expression changes, and if exposure to a particular drug causes the reverse set of changes in a model cell line (that is, negative similarity scores indicating an opposing relationship), then that drug has the potential to have a therapeutic effect on that disease.

To evaluate the significance of our predictions, we used a permutation approach under which random drug signatures were generated and the analysis was repeated 100 times for each disease. We computed the false discovery rate (FDR) for individual drug-disease similarity score values, or q value, by calculating the expected number of false positives, given the actual distribution of similarity scores and the distribution of scores after randomization for each disease signature.

Of more than 16,000 possible drug-disease pairings, 2664 were statistically significant ($q < 0.05$), more than half of which suggest a therapeutic (negative or opposite) relationship. Overall, our method provided significant candidate therapeutic drug-disease relationships for 53 of

Table 1. Drugs and diseases with the most indications.

Drugs with most indications		Diseases with most indications	
Vorinostat	21	Transitional cell carcinoma	95
Gefitinib	18	Melanoma	79
HC toxin	18	Cardiomyopathy	73
Colforsin	17	Adenocarcinoma of lung	72
17-Dimethylamino-geldanamycin	16	Multiple benign melanocytic nevi	68
Trichostatin A	16	Squamous cell carcinoma of lung	67
3-Hydroxy-DL-kynurenine	15	Malignant neoplasm of stomach	66
5114445	15	Dermatomyositis	63
Dexverapamil	15	Malignant mesothelioma of pleura	53
Prochlorperazine	15	Primary cardiomyopathy	48

the 100 diseases at an FDR of 5%. The other 47 diseases could not be significantly associated with any of the 164 drugs we tried and were excluded from further consideration. Each of the 164 drugs was significantly associated with at least 1 of the 53 remaining diseases.
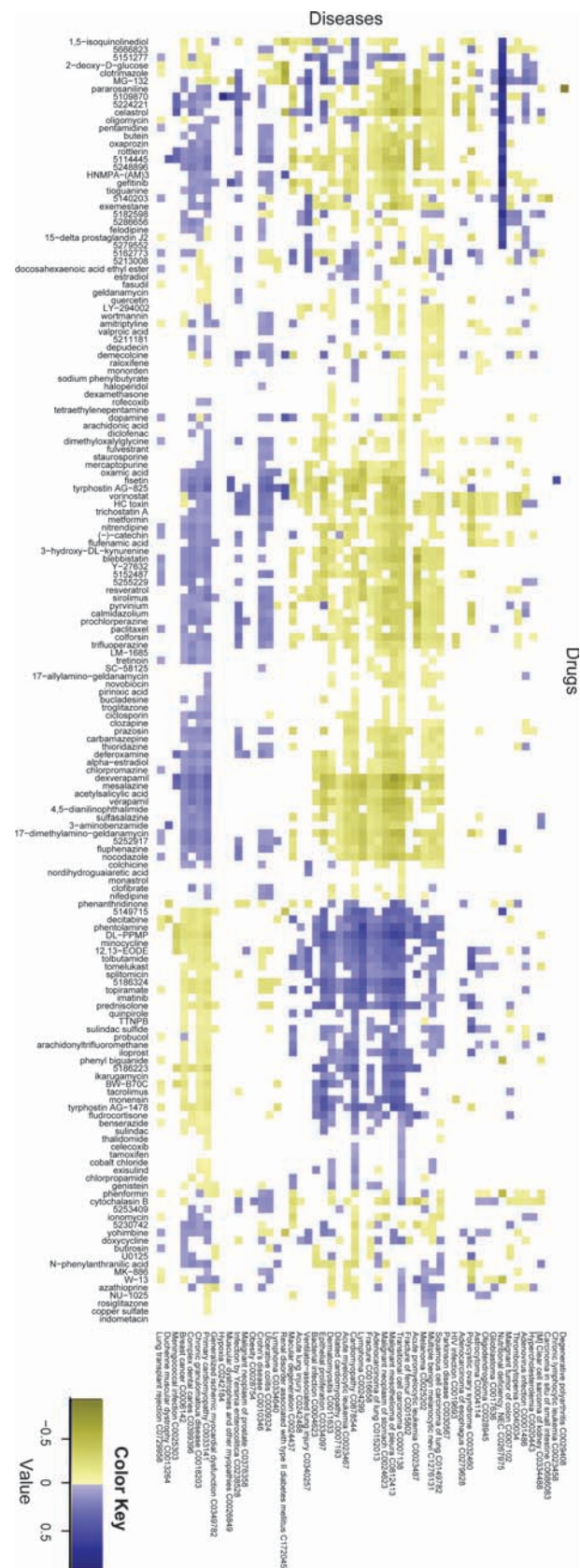
Many cancers, including gastric cancer, melanoma, and transitional cell carcinoma, showed the highest number of significant matches to therapies (Table 1). The drug predicted to be efficacious for the largest set of diseases was vorinostat, a histone deacetylase (HDAC) inhibitor with significant therapeutic association with 21 of the 100 diseases in our data set. Other HDAC inhibitors, such as trichostatin A and *Helminthosporium carbonum* (HC) toxin, as well as gefitinib, an epidermal growth factor receptor (EGFR) inhibitor used in cancer treatment, had more than 20 significant predicted therapeutic indications (Table 1).

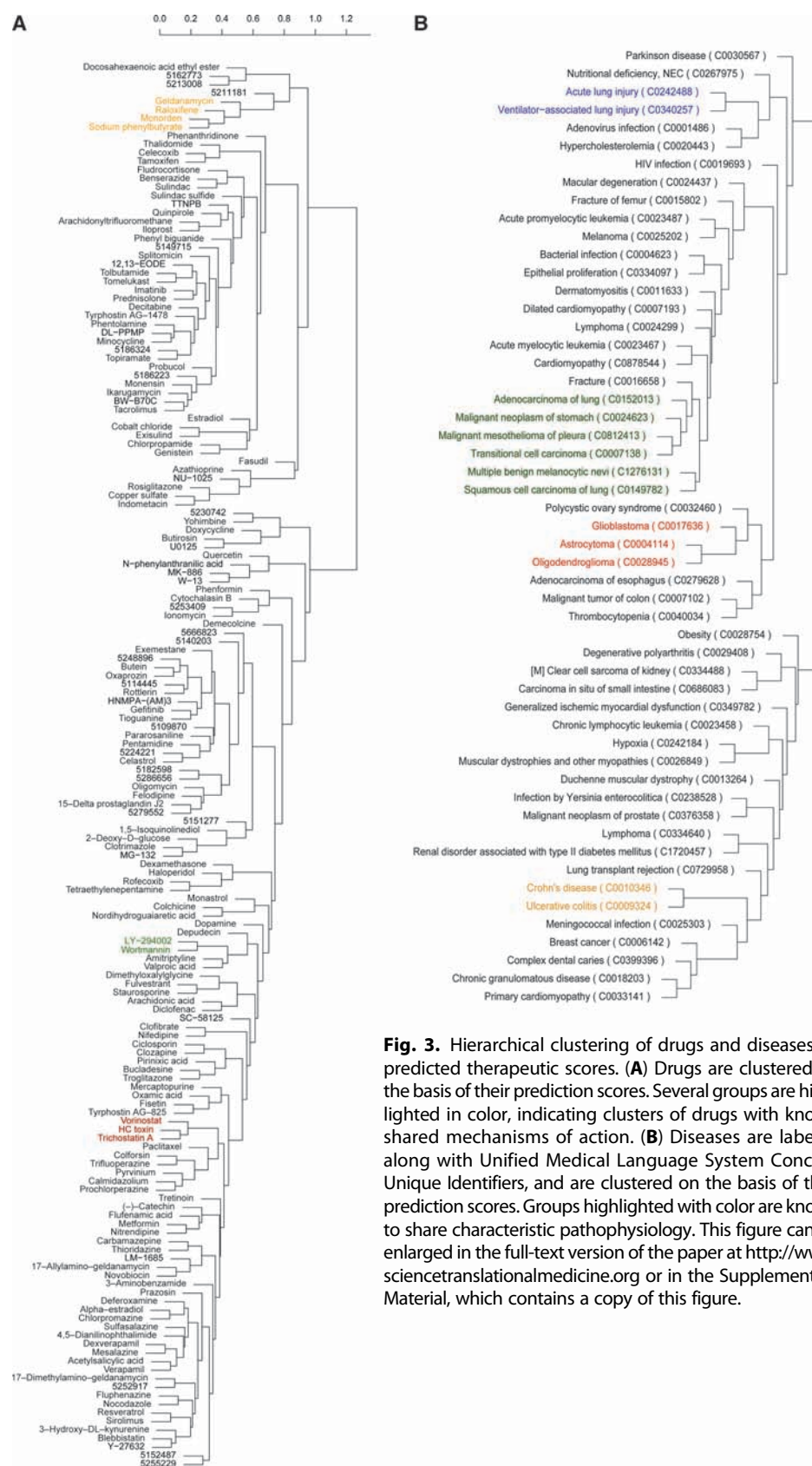## Global properties of drug-disease connections

By examining the relationships between the 164 drugs and 53 diseases based on predicted therapeutic scores, we identified major clusters of predicted therapeutic relationships between drugs and diseases (Fig. 2). We then looked separately at the clustering of drugs on the basis of their similarity scores across diseases (Fig. 3A). Drugs with similar mechanisms of action clustered together. For example, the HDAC inhibitors vorinostat, HC toxin, and trichostatin A formed a cluster, as did the phosphatidylinositol 3-kinase (PI3K) inhibitors LY-294002 and wortmannin, and the salicylate anti-inflammatory drugs sulfasalazine, mesalazine, and acetylsalicylic acid (Table 2). We also found clusters of drugs that are known to affect different aspects of the same biological process. For example, geldanamycin, raloxifene, monorden, and sodium phenylbutyrate are all known to perturb the activity of the heat shock protein 90 (HSP90) chaperone complex, either through direct inhibition or through perturbation of upstream factors (*29*), and these agents also clustered together (Table 2).

Diseases similarly clustered on the basis of their computed therapeutic scores across the panel of drugs (Table 3 and Fig. 3B). We found a large cluster of cancers, including lung, stomach, and other cancers, pointing to potential commonality of predicted therapeutic response between these conditions (Table 3 and Fig. 3B, green). The inflammatory bowel diseases (IBDs) ulcerative colitis (UC) and Crohn's disease (CD) appeared together as part of a larger cluster of diseases for which corticosteroids and other immunosuppressive drugs are broadly indicated (Table 3 and Fig. 3B, yellow). IBD was clustered near infection by *Yersinia enterocolitica*, which can present clinical symptoms similar to those of IBD (*30*). The reported clusters were determined to be significant by bootstrap analysis (*31*).

Fig. 2. Heat map of drug-disease scores. Most of the heat map is white, indicating that most of the drug and gene expression profiles are not significantly concordant. Yellow indicates a negative (therapeutic) drug-disease score, meaning the expression profiles of the two are opposing, and that the drug might be a potential treatment option of the disease. Blue indicates a positive drug-disease score, meaning the expression profiles of the two are similar, and therefore, the drug may exacerbate the disease. This figure can be enlarged in the full-text version of the paper at http://www.sciencetranslationalmedicine.org or in the Supplementary Material, which contains a copy of this figure.

A

0.0 0.2 0.4 0.6 0.8 1.0 1.2

Docosahexaenoic acid ethyl ester
5162773
5213008
5211181
Geldanamycin
Raloxifene
Monorden
Sodium phenylbutyrate
Phenanthridinone
Thalidomide
Celecoxib
Tamoxifen
Fludrocortisone
Benserazide
Sulindac
Sulindac sulfide
TTNPB
Quinpirole
Arachidonyltrifluoromethane
Iloprost
Phenyl biguanide
5149715
Splitomicin
12,13-EODE
Tolbutamide
Tomelukast
Imatinib
Prednisolone
Decitabine
Tyrphostin AG-1478
Phentolamine
DL-PPMP
Minocycline
5186324
Topiramate
Probucol
5186223
Monensin
Ikarugamycin
BW-B70C
Tacrolimus
Estradiol
Cobalt chloride
Exisulind
Chlorpropamide
Genistein
Fasudil
Azathioprine
NU-1025
Rosiglitazone
Copper sulfate
Indometacin
5230742
Yohimbine
Doxycycline
Butirosin
U0125
Quercetin
N-phenylanthranilic acid
MK-886
W-13
Phenformin
Cytochalasin B
5253409
Ionomycin
Demecolcine
5666823
5140203
Obesity
Exemestane
5248896
Butein
Oxaprozin
5114445
Rottlerin
HNMPA-(AM)3
Gefitinib
5109870
Tioguanine
Pararosaniline
Pentamidine
5224221
Celastrol
5182598
5286656
Oligomycin
Felodipine
15-Delta prostaglandin J2
5279552
5151277
1,5-Isoquinolinediol
2-Deoxy-D-glucose
Clotrimazole
MG-132
Dexamethasone
Haloperidol
Rofecoxib
Tetraethylenepentamine
Monastrol
Colchicine
Nordihydroguaiaretic acid
Dopamine
Depudecin
LY-294002
Wortmannin
Amitriptyline
Valproic acid
Dimethyloxalylglycine
Fulvestrant
Staurosporine
Arachidonic acid
Diclofenac
SC-58125
Clofibrate
Nifedipine
Ciclosporin
Clozapine
Pirinixic acid
Bucladesine
Troglitazone
Mercaptopurine
Oxamic acid
Fisetin
Tyrphostin AG-825
Vorinostat
HC toxin
Trichostatin A
Paclitaxel
Colforsin
Trifluoperazine
Pyrvinium
Calmidazolium
Prochlorperazine
Tretinoin
(–)-Catechin
Flufenamic acid
Metformin
Nitrendipine
Carbamazepine
Thioridazine
LM-1685
17-Allylamino-geldanamycin
Novobiocin
3-Aminobenzamide
Prazosin
Deferoxamine
Alpha-estradiol
Chlorpromazine
Sulfasalazine
4,5-Dianilinophthalimide
Dexverapamil
Mesalazine
Acetylsalicylic acid
Verapamil
17-Dimethylamino-geldanamycin
5252917
Fluphenazine
Nocodazole
Resveratrol
Sirolimus
3-Hydroxy-DL-kynurenine
Blebbistatin
Y-27632
5152487
5255229

B

Parkinson disease (C0030567)
Nutritional deficiency, NEC (C0267975)
Acute lung injury (C0242488)
Ventilator-associated lung injury (C0340257)
Adenovirus infection (C0001486)
Hypercholesterolemia (C0020443)
HIV infection (C0019693)
Macular degeneration (C0024437)
Fracture of femur (C0015802)
Acute promyelocytic leukemia (C0023487)
Melanoma (C0025202)
Bacterial infection (C0004623)
Epithelial proliferation (C0334097)
Dermatomyositis (C0011633)
Dilated cardiomyopathy (C0007193)
Lymphoma (C0024299)
Acute myelocytic leukemia (C0023467)
Cardiomyopathy (C0878544)
Fracture (C0016658)
Adenocarcinoma of lung (C0152013)
Malignant neoplasm of stomach (C0024623)
Malignant mesothelioma of pleura (C0812413)
Transitional cell carcinoma (C0007138)
Multiple benign melanocytic nevi (C1276131)
Squamous cell carcinoma of lung (C0149782)
Polycystic ovary syndrome (C0032460)
Glioblastoma (C0017636)
Astrocytoma (C0004114)
Oligodendroglioma (C0028945)
Adenocarcinoma of esophagus (C0279628)
Malignant tumor of colon (C0007102)
Thrombocytopenia (C0040034)
Obesity (C0028754)
Degenerative polyarthritis (C0029408)
[M] Clear cell sarcoma of kidney (C0334488)
Carcinoma in situ of small intestine (C0686083)
Generalized ischemic myocardial dysfunction (C0349782)
Chronic lymphocytic leukemia (C0023458)
Hypoxia (C0242184)
Muscular dystrophies and other myopathies (C0026849)
Duchenne muscular dystrophy (C0013264)
Infection by Yersinia enterocolitica (C0238528)
Malignant neoplasm of prostate (C0376358)
Lymphoma (C0334640)
Renal disorder associated with type II diabetes mellitus (C1720457)
Lung transplant rejection (C0729958)
Crohn's disease (C0010346)
Ulcerative colitis (C0009324)
Meningococcal infection (C0025303)
Breast cancer (C0006142)
Complex dental caries (C0399396)
Chronic granulomatous disease (C0018203)
Primary cardiomyopathy (C0033141)

Fig. 3. Hierarchical clustering of drugs and diseases by predicted therapeutic scores. (**A**) Drugs are clustered on the basis of their prediction scores. Several groups are highlighted in color, indicating clusters of drugs with known shared mechanisms of action. (**B**) Diseases are labeled along with Unified Medical Language System Concept Unique Identifiers, and are clustered on the basis of their prediction scores. Groups highlighted with color are known to share characteristic pathophysiology. This figure can be enlarged in the full-text version of the paper at http://www.sciencetranslationalmedicine.org or in the Supplementary Material, which contains a copy of this figure.

Prediction of known and previously unassociated drugs for individual diseases

In addition to examining overarching relationships between groups of drugs and diseases, we also determined individual therapeutic predictions based on the hypothesis that if a drug has a gene expression signature that is opposite to a disease signature, that drug could potentially be used as a treatment for that disease. One of the strongest therapeutic predictions for CD and UC was the corticosteroid prednisolone, a well-known treatment for these conditions (*32*), with a score of −0.216. We found that topiramate, an anticonvulsant drug currently used to treat epilepsy, had a stronger therapeutic score for CD than the established therapeutic prednisolone, with a score of −0.220, as shown by the comparison of the CD signature to the gene expression profile of topiramate. Genes that were up-regulated in disease were down-regulated by the drug and vice versa, reflecting the significant score. Topiramate was also one of the strongest predicted therapies for UC, based on our analysis. We established the efficacy of topiramate in ameliorating an IBD phenotype in an independent study using a trinitrobenzenesulfonic acid (TNBS)–treated rat model of colitis and reported in an accompanying paper (*33*).

Prediction and validation of cimetidine as a previously undescribed indication for lung adenocarcinoma

To perform an initial experimental evaluation of our approach, we chose to evaluate one of the therapeutic predictions for lung adenocarcinoma (LA), because lung cancer contributes the greatest burden of cancer mortality and incidence in Europe and the United States (*34*). Although our methodology predicted multiple new therapeutic relationships for LA, we chose to test cimetidine because it is an off-patent and inexpensive drug available over the counter in the United States and has a favorable side effect profile (*35*). Our prediction score of −0.088 for cimetidine was moderate, but still more significant than the score of −0.075 for gefitinib, a well-known therapy for LA.

Table 2. Significant drug clusters. Drugs were clustered in an unsupervised manner by their predicted therapeutic score across 100 diseases (see Fig. 3A).

Drug group	Drugs
PI3K inhibitors	LY-294002 and wortmannin
HSP90 inhibitors	Geldanamycin, raloxifene, monorden, and sodium phenylbutyrate
HDAC inhibitors	Vorinostat, HC toxin, and trichostatin A
Salicylate anti-inflammatory agents	Sulfasalazine, mesalazine, and acetylsalicylic acid

Table 3. Significant disease clusters. Diseases were clustered in an unsupervised manner according to their predicted therapeutic response across 164 drug compounds (see Fig. 3B).

Canonical	Noncanonical
Cancers	Crohn's disease and lung transplant
Ulcerative colitis and Crohn's disease	Polycystic ovary and glioblastoma
	Cardiomyopathy and cancer

To evaluate the efficacy of cimetidine against LA in vitro, we performed an MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide] colorimetric assay to assess growth and proliferation of LA cells (A549) after exposure to cimetidine, and assessed apoptosis through detection of DNA fragmentation by TUNEL (terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate nick end labeling) assay. LA cells exposed to increasing concentrations of cimetidine exhibited a dose-dependent reduction in growth and proliferation when compared to cells treated with phosphate-buffered saline (PBS) vehicle (Fig. 4A). LA cells also exhibited extensive apoptosis 1 day after direct exposure to 2000 μM cimetidine (Fig. 4B, right; green, TUNEL-positive), whereas apoptosis was not detectable by TUNEL fluorescence microscopy in LA cells from the same line after exposure to the PBS vehicle solution.

To further evaluate the efficacy of cimetidine against LA in vivo, we tested three doses of cimetidine in mice implanted with a human A549 LA cell line and followed the growth of these tumors for 12 days. Tumors in mice treated with vehicle (Fig. 4C, red line) grew to 3.25 times their original volume, whereas tumors in mice treated with our positive control dosing with the chemotherapeutic doxorubicin grew to twice the original volume (Fig. 4C, orange line). Cimetidine caused a statistically significant reduction in growth compared to control (100 mg/kg per day group versus PBS control on day 12; $P < 0.001$, t test), with the highest dose of cimetidine approaching the efficacy of the positive control therapy doxorubicin. Cimetidine exhibited its antineoplastic properties in a dose-dependent manner, with tumors in mice treated with 50 mg/kg per day (Fig. 4C, purple line) growing to 2.8 times the original size, and those treated with 100 mg/kg per day (Fig. 4C, light blue line) growing only 2.3 times as large [$P < 0.05$, Tukey's honest significant difference (HSD)]. Figure 4D shows representative examples of tumors treated with cimetidine (100 mg/kg per day) (left) in comparison to control (right).

To test the specificity of our predictions, we performed a similar experiment in mice using the ACHN renal cell carcinoma cell line as the origin of the implants (fig. S1). We chose ACHN as a control for this experiment because it did not have any significant association with cimetidine according to our analysis (score 0, $P = 1.0$). Concordant with our computational prediction, we found that cimetidine had no effect on renal cell carcinoma xenograft tumors in a mouse model (fig. S1).

DISCUSSION

Here, we sought to discover previously undescribed therapeutic relationships between drug compounds and diseases through the large-scale integration of public gene expression data for drugs and disease.

Our hypothesis was that a comprehensive and systematic comparison of drug and disease gene expression signatures could be used to build a global space of drug-disease relationships, which could be explored to identify previously unknown therapeutic relationships between drugs and disease conditions. The global clusterings of drugs and diseases by predicted therapeutic scores revealed known therapeutic clusters of drugs and diseases and also provided pathophysiological context to support interpretation of novel predicted therapeutic relationships.

Drugs that clustered on the basis of their gene signature similarity across diseases exhibited shared mechanisms of action (for example, HDAC inhibition) or shared physiological processes (for example, immunomodulation). In some cases, the axis of commonality is not apparent within drug clusters, and these situations may inform our understanding of a drug's unknown mechanism of action or even suggest new biology. For example, we found a cluster containing the salicylate drugs sulfasalazine, mesalazine (a metabolite of sulfasalazine), and acetylsalicylic acid grouped with the calcium channel blocker verapamil and its R-enantiomer, dexverapamil. The two distinct drug types represented in this cluster are not known to act via the same mechanism and are not generally considered to be interchangeable or complementary therapeutic options for a common disease indication. Although the underlying mechanism is unclear, both mesalazine and sulfasalazine are used as maintenance drugs in the treatment of the inflammatory bowel disorders CD and UC. A possible therapeutic role for verapamil in IBD is indicated by several studies that show verapamil to reduce inflammation in UC through leukotriene inhibition (36) and promote mucosal healing in experimental colitis (37). Further study into the genes and potential additional drug targets driving this clustering could reveal new information about IBD biology and new therapeutic directions.

The diseases that clustered on the basis of their gene signature profiles showed within-cluster consistency of established therapeutic conventions and likely shared pathophysiology. The major cluster containing CD and lung transplant rejection also harbored a number of diseases for which corticosteroids and other immunosuppressants are regularly prescribed. The large cluster of cancers shares many of the same antineoplastic and immunomodulatory therapeutics. Less evident groupings, such as that between polycystic ovary syndrome (PCOS) and glioblastoma, may present opportunities for drug repositioning within a cluster. However, in some cases, such groupings may have alternative explanations. For example, the peculiar clustering of cardiomyopathies with cancers could be a reflection of the fact that many chemotherapeutics induce cardiomyopathies as a side effect (38).

In addition to examining broader relationships among groups of drugs and diseases, we also discovered individual therapeutic predictions for 53 of the 100 diseases in our data set. It is not surprising that
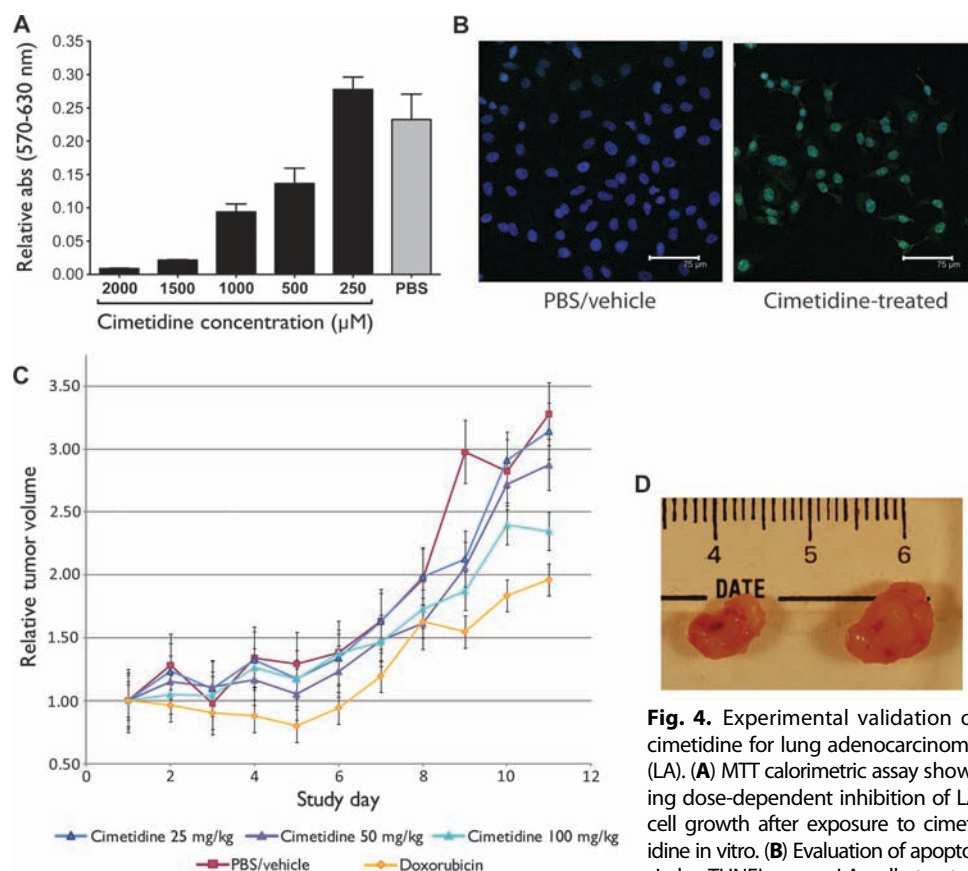
Fig. 4. Experimental validation of cimetidine for lung adenocarcinoma (LA). (**A**) MTT calorimetric assay showing dose-dependent inhibition of LA cell growth after exposure to cimetidine in vitro. (**B**) Evaluation of apoptosis by TUNEL assay. LA cells treated with 2000 μM cimetidine exhibit a significant increase in TUNEL-positive (green) nuclei compared to vehicle (PBS)–treated control. The scale bar is 75 μM. (**C**) Results from a tumor xenograft experiment testing the efficacy of H2 agonist cimetidine in inhibiting the growth of A549 LA cell line tumors in SCID mice. Three treatment groups (25, 50, or 100 mg/kg per injection) and one control group (PBS) were used. Another group was treated with doxorubicin as a positive control. (**D**) Representative images of tumors treated with high dose of cimetidine (left) and control (right).

modest single-agent activity (*46*). Recent studies have also shown that HC toxin effectively suppresses the malignant phenotype of neuroblastoma cells (*47*, *48*). Our prediction that celastrol, an antioxidant and anti-inflammatory drug, would be effective against melanoma is also supported by previous research, which shows that the compound resembles the well-characterized ATF2 (activating transcription factor 2) peptide and may therefore offer new approaches for treatment (*49*).

After computationally screening our large set of compounds, we predicted and validated cimetidine as a possible therapeutic agent for LA. Cimetidine is a histamine H2 receptor (*H2R*) antagonist that acts as an inhibitor of acid production in the stomach. It is used and prescribed most often as an over-the-counter drug for heartburn and peptic ulcers. Cimetidine has previously demonstrated efficacy in other adenocarcinoma models (*50*–*53*). A molecular mechanism or other molecular basis for the observed efficacy of cimetidine against LA has not been previously reported or otherwise described. Initial reports of cimetidine's antitumor activity proposed that cimetidine inhibits tumor growth via immunomodulation (*54*–*57*). However, as demonstrated by the in vitro and in vivo experimental results of this study, cimetidine inhibits tumor growth and cellular proliferation in immune-deficient hosts and by direct exposure to LA cells. More recent studies have demonstrated that cimetidine can inhibit vascular development in tumors, suggesting that cimetidine might work to inhibit tumor-associated angiogenesis (*52*, *58*). Additional studies into the antineoplastic properties of cimetidine have revealed putative roles in the inhibition of tumor cell adhesion and induction of apoptosis (*59*, *60*). It is also possible that the efficacy of cimetidine for LA is mediated by some yet unknown interaction between cimetidine and a target outside of its canonical H2R target. The characteristics of the gene signatures can be used to learn more about the potential mechanism of action of the drug.

Despite computational and experimental validation of our approach, there are several caveats. One potential drawback to the approach described here is the need to have gene expression profile measurements on the candidate drugs being evaluated. Such data are publicly available for more than 1000 compounds in the second release of the Connectivity Map; however, there are numerous compounds that are not part of the database. In addition, it is not clear whether drug performance in a breast cancer cell line, used extensively to measure transcriptional response of drugs in Connectivity Map, is relevant to all types of diseases. However, the number of publicly available data reporting the effects of drugs on gene expression across disease tissues continues to

a large proportion of the diseases we studied did not have any significant therapeutic predictions, because the set of drugs we examined was limited, and the baseline assumption was that most drugs will not promiscuously exhibit therapeutic potential for many diseases.

Many of our therapeutic predictions recapitulate established therapeutic knowledge. For instance, we predict that HDAC inhibitors (trichostatin A, valproic acid, vorinostat, and HC toxin) have therapeutic effects in treatment of different types of brain tumors (astrocytoma, glioblastoma, and oligodendroglioma) as well as other cancers (esophagus, lung, and colon). HDAC inhibitors are anticancer agents that act by inhibiting cancer cell proliferation and inducing apoptosis (*39*–*41*). Trichostatin A was recently shown to restore the loss of *NECL1*, a tumor suppressor in glioma (*42*). Valproic acid is a drug used to treat epilepsy and bipolar disorder but has been more recently reported to be effective in the treatment of various cancers including glioma (*43*), myeloma (*44*), and melanoma (*45*), often in combination with other therapies. Vorinostat is currently used for the treatment of several lymphomas but has recently undergone successful phase II trials for treatment of recurrent glioblastoma multiforme (GBM), which showed that vorinostat is well tolerated in patients with recurrent GBM and has

grow (*61*). Furthermore, disease-related microarray data could be combined with other types of knowledge on drugs (*62, 63*) for computational prediction of drug side effects.

Therapeutic efficacy is always more complex than just a simple matching of expression profiles. For example, compounds have to reach the desired or appropriate tissue to have an effect. However, the tissue-agnostic methodology used in this analysis (the disease and drug gene expression was not measured on the same tissues) might be suitable to find both direct and distant effects of drugs. Although our findings for cimetidine will need further preclinical testing and demonstration in larger randomized prospective clinical trials, our results validate the concept of computational analysis of public gene expression databases as a potentially useful approach to drug discovery that may uncover additional uses for approved drugs.

MATERIALS AND METHODS

We combined data from publicly available microarray data sets representing 100 diseases and gene expression data from human cell lines treated with 164 drugs or small molecules to predict therapeutic drug-disease interactions.

Drug and disease gene expression data

Disease expression data were obtained from the NCBI GEO (*64, 65*), with methods previously described (*66–68*). We used 176 gene expression microarray data sets, each with samples reflecting disease tissue and control nondisease tissue (table S1). These data sets covered 100 diseases, totaling 3113 microarrays (table S1). A data set with measurements for LA and adjacent normal tissues measured from human patients (GSE2514) was among the 176 data sets used for the study.

Rank normalization was applied to the disease expression data to carry out robust cross-platform analysis (*69, 70*). Because experiments were run on different platforms, we standardize gene identifiers from chip-specific probe identifiers to NCBI GeneID identifiers using AILUN (*71*) to be able to reason across multiple experiments. In cases where multiple microarray probes mapped to the same NCBI GeneID, we averaged across individual probe expression values. On each data set, we performed SAM (*28*), comparing control samples to disease samples to generate a list of up- and down-regulated genes for each disease state. We used an FDR threshold of 0.05 for *q* values. The list of significantly up- and down-regulated genes for each disease-control comparison constituted a disease signature (Fig. 1).

Drug-exposure gene expression microarray data measured on several types of cancer cell lines for 164 distinct small molecules were obtained from Lamb *et al.* (*25*). Most of the experiments were carried out on the breast cancer epithelial cell line MCF7, but a subset of the compounds was also profiled in the prostate cancer epithelial cell line PC3 and the nonepithelial lines HL60 (leukemia) and SKMEL5 (melanoma). The original experiments were carried out on two platforms: GEO Platform (GPL) 96 Affymetrix GeneChip Human Genome U133 Array Set HG-U133A and GPL3921 Affymetrix GeneChip HT-HG_U133A Arrays. To be able to reason between drug and disease expression data, we standardized gene identifiers from microarray-specific probe identifiers to NCBI GeneID identifiers, averaging across individual probe expression values (Fig. 1). We followed the preprocessing and normalization steps described in the supplementary material in Lamb *et al.* (*25*).

Computational pipeline

We used the drug-exposure expression data as a reference database and queried this database with each individual disease signature by applying a nonparametric, rank-based, pattern-matching strategy based on the methodology originally introduced by Lamb *et al.* (*25*) to generate a ranked list of potential treatments for each of the diseases of interest (Fig. 1). Given a disease signature, we evaluated its similarity to each of the reference expression drug profiles in the drug expression set by computing an enrichment score for the up- and down-regulated disease genes. If the up-regulated disease genes appear near the top (up-regulated) of the rank-ordered drug gene expression list and the down-regulated disease genes fall near the bottom (down-regulated) of the rank-ordered drug gene expression list, we can conclude that the drug and the disease expression profiles are similar, and thus the drug might cause a change in tissue expression similar to having the disease. More interestingly, if the up-regulated disease genes fall near the bottom of the rank-ordered drug gene expression list and the down-regulated disease genes are near the top of the rank-ordered drug gene expression, then the given drug and disease have complimentary expression profiles and the drug might be a possible treatment option for the disease of interest.

For each disease, we computed an enrichment score (es) separately for the set of up- or down-regulated genes in the signature: es_{up} and es_{down}. We constructed a reference drug expression set by taking the difference between the gene expression values of the samples treated with a compound of interest and untreated samples. Furthermore, we carried out rank normalization on the resulting gene expression differences. Let r be the total number of genes in the reference drug expression set and let s be the number of up- or down-regulated of each gene in the disease signature. We first constructed a vector $\mathbf{V}$ of the position $(1 \ldots r)$ for each of the genes in the disease signature on the basis of the values from the reference drug expression set. Those were sorted in ascending order such that $\mathbf{V}(j)$ is the position of disease gene j, where $j = 1, 2, \ldots s$. Then, for each set of up- and down-regulated disease genes, we computed a_{up}, a_{down}, b_{up}, and b_{down} defined as follows.

If $a_{up/down} > b_{up/down}$, we set $es_{up/down}$ to $a_{up/down}$, and if $b_{up/down} > a_{up/down}$, we set $es_{up}/_{down}$ to $-b_{up/down}$. The drug-disease score dds is set to zero, where es_{up} and es_{down} have the same algebraic sign. Otherwise, we set dds = $es_{up} - es_{down}$. For more details, see the supplementary material in Lamb *et al.* (*25*).

In the original method proposed by Lamb *et al.*, the final drug-disease score was scaled from +1 to −1. In using this scaling, every disease signature was able to match some drug profile at a maximal level. Because the scores for each disease were assigned a unique scaling factor, the strength or confidence of one drug-disease association could not be compared with another drug-disease association. We thus chose not to carry out this final scaling step, instead using the resulting raw values as drug-disease scores. To present the results in a more readable fashion, we chose to average the scores by drug dose and cell line, although we realize that this might dampen some of the signal. Finally, we ranked the list of drug and small-molecule instances for each disease from most correlated to most anticorrelated, according to the computed similarity drug-disease score.

As a control, for every disease signature in our data set, we chose a random signature of the same size and recomputed the drug-disease scores for each drug. We repeated this experiment 100 times. The variance of the original scores is statistically significantly greater than the scores generated from random signatures ($P < 2.2 \times 10^{-16}$, Levene's

test), meaning that there were more significant therapeutic indications in our resulting data set than expected by chance.

To obtain a measure of significance on the drug-disease scores, we started by computing a P value for each drug-disease score by comparing the distribution of actual scores to those obtained on randomized data for each disease signature. We further used the q value (72) statistical package to compute tail area–based FDR for each drug-disease score in our result given the previously computed P values. On the basis of this analysis, we were able to establish the cutoff for significance for drug-disease scores for each disease signature, yielding an FDR of 5%.

To identify disease and drug classes, we applied hierarchical cluster analysis to the data using the computed Pearson correlation coefficients as a distance metric between disease or drug pairs. Initially, each disease or drug was assigned to its own cluster. The algorithm proceeded iteratively, at each stage joining the two most similar clusters, until there was just a single cluster left. We used the Pvclust R package (31) to compute a bootstrap analysis of the clusters. The bootstrap probability of a cluster corresponds to the frequency with which the cluster appears in bootstrap samples of the data. Approximately unbiased (AU) probability values were computed with bootstrap samples of various sizes and indicate how strongly the cluster is supported by data (AU > 95%).

Evaluation of cimetidine as a chemotherapeutic agent for non–small cell lung carcinoma

To evaluate our prediction, we performed experimental validation using tumor xenograft models of non–small cell lung carcinoma. Our experimental design was based on two human cancer cell lines that were implanted into severe combined immunodeficient (SCID) mice. Thirty mice were implanted with the A549 human adenocarcinoma cell line. Five million cells per microliter of PBS were injected into upper flanks of the animals. Thirty mice were divided into five groups of six mice each. Three groups received varying doses of cimetidine ranging from 25 to 100 mg/kg daily via intraperitoneal injection. One group was injected with doxorubicin (2 mg/kg) intraperitoneally biweekly as a positive control, and the final group received only PBS/vehicle as a control. The experiment was carried out for 12 days after the tumor implants reached a volume of 100 mm³. Tumor volumes were measured by caliper with the following equation: mm³ = 0.52 × [width (cm)]² × height (cm).

To evaluate the specificity of our approach, we performed a parallel experiment using the ACHN renal carcinoma cell line (5 million cells per microliter of PBS injection into upper flanks), because we predicted that cimetidine would not be efficacious against this cancer type. We used the same experimental design and protocol for the mice implanted with the A549 cell line as for the animals implanted with ACHN. Differences in tumor volume between groups were evaluated on the final day of dosing with Tukey's HSD test and also Student's t test between the highest dose of cimetidine and the PBS control. Mouse xenograft experiments were performed by the Transgenic Mouse Research Center core facility at the Stanford Comprehensive Cancer Center. Cell lines were verified to be free of murine pathogens by Research Animal Diagnostic and Investigative Laboratory. All chemicals and reagents were purchased from Sigma-Aldrich Corp.

Colorimetric MTT assay for cell survival and proliferation

MTT assay reagents from Millipore were used for cell survival and proliferation. The assay was carried out according to the manufacturer's instructions. A549 cells were seeded at 5000 cells per well in a 96-well

plate 15 to 18 hours before the start of the treatment. Cimetidine solution (250 mM) was prepared by adding 0.63 g of cimetidine in 8 ml of PBS and then 10 N HCl (200 µl) and 10 N NaOH (150 µl). Then, the pH was adjusted to the physiological level and total volume was adjusted to 10 ml. Cells were treated three times with cimetidine (250, 500, 1000, 1500, and 2000 µM) with 250 mM stock solution at 11 a.m., 6 p.m., and 11 a.m. the next day followed by MTT assay at 4 p.m. After the treatments, 0.01 ml of MTT (Millipore CT01-5, 50 mg/ml in PBS) solution was added to each well, and the cells were incubated for 4 hours at 37°C for the cleavage of MTT to occur. Then, color development solution (isopropanol with 0.04 N HCl, 0.1 ml each) was added and mixed thoroughly. Within an hour, absorbance was measured at 570 nm and at a reference wavelength at 630 nm. For the calculation, absorbances measured at 570 nm were subtracted from those measured at 630 nm.

TUNEL assay

TUNEL assay was carried out according to the manufacturer's instructions (Roche). A549 cells were seeded at 20,000 cells per well in Nunc four-well Lab-Tek chamber slides in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS). Cells were seeded 15 to 18 hours before the start of the treatment. Cells were treated two times with cimetidine (500 to 3000 µM with 250 mM stock solution) at 11 a.m. and 6 p.m. and were stained the next day. Medium was removed, cells were washed in cold PBS, and the cells were air-dried. Then, freshly made 4% paraformaldehyde solution was added and cells were fixed for 15 min in room temperature. Cells were rinsed with PBS and were incubated with permeabilization solution (0.1% Triton X-100) for 2 min on ice. Next, terminal deoxynucleotidyl transferase and label solution were mixed and 50 µl each of the mixture was added to each well. Cells were incubated for 1 hour in a humidified chamber at 37°C. For negative controls, no label solution was added. Slides were covered with coverslips to ensure proper staining. Cells were then washed with cold PBS three times and dried. Finally, Hoechst solution for nuclei staining (Sigma) was diluted 1:10,000 in PBS and was added to the cells for 1 to 2 min. Cells were rinsed with PBS three and two times in water. Cells were applied with aqueous mounting medium and coverslipped. Immunofluorescence was measured with a Leica confocal microscope at the Cell Sciences Imaging Center at Stanford Cancer Center.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/3/96/96ra77/DC1
Fig. S1. Effect of the H2 agonist cimetidine on the growth of ACHN renal carcinoma cell line tumors in SCID mice.
Fig. S2. Heat map of drug-disease scores.
Fig. S3A. Hierarchical clustering of drugs by predicted therapeutic scores.
Fig. S3B. Hierarchical clustering of diseases by predicted therapeutic scores.
Table S1. Datasets downloaded from the NCBI Gene Expression Omnibus for this study.

REFERENCES AND NOTES

1. T. T. Ashburn, K. B. Thor, Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).
2. J. A. DiMasi, R. W. Hansen, H. G. Grabowski, The price of innovation: New estimates of drug development costs. *J. Health Econ.* **22**, 151–185 (2003).
3. C. R. Chong, D. J. Sullivan Jr., New uses for old drugs. *Nature* **448**, 645–646 (2007).
4. J. K. Aronson, Old drugs—New uses. *Br. J. Clin. Pharmacol.* **64**, 563–565 (2007).

5. P. Y. Lum, J. M. Derry, E. E. Schadt, Integrative genomics and drug development. *Pharmacogenomics* **10**, 203–212 (2009).

6. E. E. Schadt, S. H. Friend, D. A. Shaywitz, A network view of disease and compound screening. *Nat. Rev. Drug Discov.* **8**, 286–295 (2009).

7. X. A. Qu, R. C. Gudivada, A. G. Jegga, E. K. Neumann, B. J. Aronow, Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. *BMC Bioinformatics* **10** (Suppl. 5), S4 (2009).

8. M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijer, R. C. Matos, T. B. Tran, R. Whaley, R. A. Glennon, J. Hert, K. L. Thomas, D. D. Edwards, B. K. Shoichet, B. L. Roth, Predicting new molecular targets for known drugs. *Nature* **462**, 175–181 (2009).

9. L. Xie, J. Li, L. Xie, P. E. Bourne, Drug discovery using chemical systems biology: Identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput. Biol.* **5**, e1000387 (2009).

10. S. L. Kinnings, N. Liu, N. Buchmeier, P. J. Tonge, L. Xie, P. E. Bourne, Drug discovery using chemical systems biology: Repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.* **5**, e1000423 (2009).

11. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).

12. R. J. Lipshutz, D. Morris, M. Chee, E. Hubbell, M. J. Kozal, N. Shah, N. Shen, R. Yang, S. P. Fodor, Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19**, 442–447 (1995).

13. Y. L. Yap, X. W. Zhang, D. Smith, R. Soong, J. Hill, Molecular gene expression signature patterns for gastric cancer diagnosis. *Comput. Biol. Chem.* **31**, 275–287 (2007).

14. A. Milano, S. A. Pendergrass, J. L. Sargent, L. K. George, T. H. McCalmont, M. K. Connolly, M. L. Whitfield, Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS One* **3**, e2696 (2008).

15. W. Z. Wang, X. Guo, C. Duan, W. J. Ma, Y. G. Zhang, P. Xu, Z. Q. Gao, Z. F. Wang, H. Yan, Y. F. Zhang, Y. X. Yu, J. C. Chen, M. J. Lammi, Comparative analysis of gene expression profiles between the normal human cartilage and the one with endemic osteoarthritis. *Osteoarthritis Cartilage* **17**, 83–90 (2009).

16. C. S. Walsh, S. Ogawa, H. Karahashi, D. R. Scoles, J. C. Pavelka, H. Tran, C. W. Miller, N. Kawamata, C. Ginther, J. Dering, M. Sanada, Y. Nannya, D. J. Slamon, H. P. Koeffler, B. Y. Karlan, *ERCC5* is a novel biomarker of ovarian cancer prognosis. *J. Clin. Oncol.* **26**, 2952–2958 (2008).

17. G. Hu, P. Agarwal, Human disease-drug network based on genomic expression profiles. *PLoS One* **4**, e6536 (2009).

18. R. Dobrin, J. Zhu, C. Molony, C. Argman, M. L. Parrish, S. Carlson, M. F. Allan, D. Pomp, E. E. Schadt, Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol.* **10**, R55 (2009).

19. J. T. Dudley, R. Tibshirani, T. Deshpande, A. J. Butte, Disease signatures are robust across tissues and experiments. *Mol. Syst. Biol.* **5**, 307 (2009).

20. M. Del Rio, F. Molina, C. Bascoul-Mollevi, V. Copois, F. Bibeau, P. Chalbos, C. Bareil, A. Kramar, N. Salvetat, C. Fraslon, E. Conseiller, V. Granci, B. Leblanc, B. Pau, P. Martineau, M. Ychou, Gene expression signature in advanced colorectal cancer patients select drugs and response for the use of leucovorin, fluorouracil, and irinotecan. *J. Clin. Oncol.* **25**, 773–780 (2007).

21. J. Li, W. H. Wood III, K. G. Becker, A. T. Weeraratna, P. J. Morin, Gene expression response to cisplatin treatment in drug-sensitive and drug-resistant ovarian cancer cells. *Oncogene* **26**, 2860–2872 (2007).

22. I. Fichtner, W. Slisow, J. Gill, M. Becker, B. Elbe, T. Hillebrand, M. Bibby, Anticancer drug response and expression of molecular markers in early-passage xenotransplanted colon carcinomas. *Eur. J. Cancer* **40**, 298–307 (2004).

23. A. Holleman, M. H. Cheok, M. L. den Boer, W. Yang, A. J. Veerman, K. M. Kazemier, D. Pei, C. Cheng, C. H. Pui, M. V. Relling, G. E. Janka-Schaub, R. Pieters, W. E. Evans, Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment. *N. Engl. J. Med.* **351**, 533–542 (2004).

24. J. Robert, A. Vekris, P. Pourquier, J. Bonnet, Predicting drug response based on gene expression. *Crit. Rev. Oncol. Hematol.* **51**, 205–227 (2004).

25. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, T. R. Golub, The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).

26. K. S. Garman, C. R. Acharya, E. Edelman, M. Grade, J. Gaedcke, S. Sud, W. Barry, A. M. Diehl, D. Provenzale, G. S. Ginsburg, B. M. Ghadimi, T. Ried, J. R. Nevins, S. Mukherjee, D. Hsu, A. Potti, A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19432–19437 (2008).

27. S. R. Setlur, K. D. Mertz, Y. Hoshida, F. Demichelis, M. Lupien, S. Perner, A. Sboner, Y. Pawitan, O. Andrén, L. A. Johnson, J. Tang, H. O. Adami, S. Calza, A. M. Chinnaiyan, D. Rhodes, S. Tomlins, K. Fall, L. A. Mucci, P. W. Kantoff, M. J. Stampfer, S. O. Andersson, E. Varenhorst, J. E. Johansson, M. Brown, T. R. Golub, M. A. Rubin, Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *J. Natl. Cancer Inst.* **100**, 815–825 (2008).

28. V. G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5116–5121 (2001).

29. H. Okano, M. Jayachandran, A. Yoshikawa, V. M. Miller, Differential effects of chronic treatment with estrogen receptor ligands on regulation of nitric oxide synthase in porcine aortic endothelial cells. *J. Cardiovasc. Pharmacol.* **47**, 621–628 (2006).

30. K. J. Ryan, C. G. Ray, *Sherris Medical Microbiology* (McGraw-Hill, New York, 2004).

31. R. Suzuki, H. Shimodaira, Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

32. P. M. Irving, R. B. Gearry, M. P. Sparrow, P. R. Gibson, Review article: Appropriate use of corticosteroids in Crohn's disease. *Aliment. Pharmacol. Ther.* **26**, 313–329 (2007).

33. J. T. Dudley, M. Sirota, M. Shenoy, R. Pai, S. Roedder, A. P. Chiang, A. A. Morgan, P. J. Pasricha, A. J. Butte, Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* **3**, 96ra76 (2011).

34. A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, D. Forman, Global cancer statistics. *CA Cancer J. Clin.* **61**, 69–90 (2011).

35. S. M. Sabesin, Safety issues relating to long-term treatment with histamine H_2-receptor antagonists. *Aliment. Pharmacol. Ther.* **7** (Suppl. 2), 35–40 (1993).

36. D. J. Gertner, D. S. Rampton, T. R. Stevens, J. E. Lennard-Jones, Verapamil inhibits in-vitro leucotriene B4 release by rectal mucosa in active ulcerative colitis. *Aliment. Pharmacol. Ther.* **6**, 163–168 (1992).

37. R. N. Fedorak, L. R. Empey, K. Walker, Verapamil alters eicosanoid synthesis and accelerates healing during experimental colitis in rats. *Gastroenterology* **102**, 1229–1235 (1992).

38. H. Cheng, T. Force, Molecular mechanisms of cardiovascular toxicity of targeted cancer therapeutics. *Circ. Res.* **106**, 21–34 (2010).

39. D. C. Drummond, C. O. Noble, D. B. Kirpotin, Z. Guo, G. K. Scott, C. C. Benz, Clinical development of histone deacetylase inhibitors as anticancer agents. *Annu. Rev. Pharmacol. Toxicol.* **45**, 495–528 (2005).

40. R. Schneider-Stock, M. Ocker, Epigenetic therapy in cancer: Molecular background and clinical development of histone deacetylase and DNA methyltransferase inhibitors. *IDrugs* **10**, 557–561 (2007).

41. S. Shankar, R. K. Srivastava, Histone deacetylase inhibitors: Mechanisms and clinical significance in cancer: HDAC inhibitor-induced apoptosis. *Adv. Exp. Med. Biol.* **615**, 261–298 (2008).

42. J. Gao, T. Chen, J. Liu, W. Liu, G. Hu, X. Guo, B. Yin, Y. Gong, J. Zhao, B. Qiang, J. Yuan, X. Peng, Loss of NECL1, a novel tumor suppressor, can be restored in glioma by HDAC inhibitor-trichostatin A through Sp1 binding site. *Glia* **57**, 989–999 (2009).

43. K. Morita, K. Gotohda, H. Arimochi, M. S. Lee, S. Her, Histone deacetylase inhibitors promote neurosteroid-mediated cell differentiation and enhance serotonin-stimulated brain-derived neurotrophic factor gene expression in rat C6 glioma cells. *J. Neurosci. Res.* **87**, 2608–2614 (2009).

44. C. Schwartz, V. Palissot, N. Aouali, S. Wack, N. H. Brons, B. Leners, M. Bosseler, G. Berchem, Valproic acid induces non-apoptotic cell death mechanisms in multiple myeloma cell lines. *Int. J. Oncol.* **30**, 573–582 (2007).

45. A. Valentini, P. Gravina, G. Federici, S. Bernardini, Valproic acid induces apoptosis, p[16INK4A] upregulation and sensitization to chemotherapy in human melanoma cells. *Cancer Biol. Ther.* **6**, 185–191 (2007).

46. E. Galanis, K. A. Jaeckle, M. J. Maurer, J. M. Reid, M. M. Ames, J. S. Hardwick, J. F. Reilly, A. Loboda, M. Nebozhyn, V. R. Fantin, V. M. Richon, B. Scheithauer, C. Giannini, P. J. Flynn, D. F. Moore Jr., J. Zwiebel, J. C. Buckner, Phase II trial of vorinostat in recurrent glioblastoma multiforme: A North Central Cancer Treatment Group study. *J. Clin. Oncol.* **27**, 2052–2058 (2009).

47. H. E. Deubzer, V. Ehemann, A. E. Kulozik, F. Westermann, L. Savelyeva, A. Kopp-Schneider, D. Riester, M. Schwab, O. Witt, Anti-neuroblastoma activity of *Helminthosporium carbonum* (HC)-toxin is superior to that of other differentiating compounds in vitro. *Cancer Lett.* **264**, 21–28 (2008).

48. H. E. Deubzer, V. Ehemann, F. Westermann, R. Heinrich, G. Mechtersheimer, A. E. Kulozik, M. Schwab, O. Witt, Histone deacetylase inhibitor *Helminthosporium carbonum* (HC)-toxin suppresses the malignant phenotype of neuroblastoma cells. *Int. J. Cancer* **122**, 1891–1900 (2008).

49. S. Abbas, A. Bhoumik, R. Dahl, S. Vasile, S. Krajewski, N. D. Cosford, Z. A. Ronai, Preclinical studies of celastrol and acetyl isogambogic acid in melanoma. *Clin. Cancer Res.* **13**, 6769–6778 (2007).

50. D. Eaton, R. E. Hawkins, Cimetidine in colorectal cancer—Are the effects immunological or adhesion-mediated? *Br. J. Cancer* **86**, 159–160 (2002).

51. K. Furuta, S. Sato, T. Miyake, E. Okamoto, J. Ishine, S. Ishihara, Y. Amano, K. Adachi, Y. Kinoshita, Anti-tumor effects of cimetidine on hepatocellular carcinomas in diethylnitrosamine-treated rats. *Oncol. Rep.* **19**, 361–368 (2008).

52. T. Natori, M. Sata, R. Nagai, M. Makuuchi, Cimetidine inhibits angiogenesis and suppresses tumor growth. *Biomed. Pharmacother.* **59**, 56–60 (2005).

53. O. Sürücü, M. Middeke, I. Höschele, G. Kalder, S. Hennig, C. Dietz, I. Celik, Tumour growth inhibition of human pancreatic cancer xenografts in SCID mice by cimetidine. *Inflamm. Res.* **53** (Suppl. 1), S39–S40 (2004).

54. M. E. Osband, D. Hamilton, Y. J. Shen, E. Cohen, M. Shlesinger, P. Lavin, A. Brown, R. McCaffrey, Successful tumour immunotherapy with cimetidine in mice. *Lancet* **1**, 636–638 (1981).

55. R. R. Gifford, R. M. Ferguson, B. V. Voss, Cimetidine reduction of tumour formation in mice. *Lancet* **1**, 638–640 (1981).

56. R. R. Gifford, B. V. Voss, R. M. Ferguson, Cimetidine protection against lethal tumor challenge in mice. *Surgery* **90**, 344–351 (1981).

57. A. Kumar, Cimetidine: An immunomodulator. *DICP* **24**, 289–295 (1990).

58. C. G. Jiang, F. R. Liu, H. M. Xu, T. Wu, J. Gao, Effects of cimetidine on the biological behaviors of human gastric cancer cells. *Zhonghua Yi Xue Za Zhi* **86**, 1813–1816 (2006).

59. K. Kobayashi, S. Matsumoto, T. Morishima, T. Kawabe, T. Okamoto, Cimetidine inhibits cancer cell adhesion to endothelial cells and prevents metastasis by blocking E-selectin expression. *Cancer Res.* **60**, 3978–3984 (2000).

60. M. Fukuda, K. Kusama, H. Sakashita, Cimetidine inhibits salivary gland tumor cell adhesion to neural cells and induces apoptosis by blocking NCAM expression. *BMC Cancer* **8**, 376 (2008).

61. Y. A. Lin, A. P. Chiang, R. Lin, P. Yao, R. Chen, A. J. Butte, Methodologies for extracting functional pharmacogenic experiments from international repository. *AMIA Annu. Symp. Proc.* 463–467 (2007).

62. A. P. Chiang, A. J. Butte, Data-driven methods to discover molecular determinants of serious adverse drug events. *Clin. Pharmacol. Ther.* **85**, 259–268 (2009).

63. M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen, P. Bork, Drug target identification using side-effect similarity. *Science* **321**, 263–266 (2008).

64. T. Barrett, R. Edgar, Gene Expression Omnibus: Microarray data storage, submission, retrieval, and analysis. *Methods Enzymol.* **411**, 352–369 (2006).

65. R. Edgar, M. Domrachev, A. E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).

66. A. J. Butte, R. Chen, Finding disease-related genomic experiments within an international repository: first steps in translational bioinformatics. *AMIA Annu. Symp. Proc.* 106–110 (2006).

67. A. J. Butte, I. S. Kohane, Creation and implications of a phenome-genome network. *Nat. Biotechnol.* **24**, 55–62 (2006).

68. J. Dudley, A. J. Butte, Enabling integrative genomic analysis of high-impact human diseases through text mining. *Pac. Symp. Biocomput.* 580–591 (2008).

69. P. Warnat, R. Eils, B. Brors, Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC Bioinformatics* **6**, 265 (2005).

70. M. C. Abba, Y. Hu, H. Sun, J. A. Drake, S. Gaddis, K. Baggerly, A. Sahin, C. M. Aldaz, Gene expression signature of estrogen receptor α status in breast cancer. *BMC Genomics* **6**, 37 (2005).

71. R. Chen, L. Li, A. J. Butte, AILUN: Reannotating gene expression data automatically. *Nat. Methods* **4**, 879 (2007).

72. J. D. Storey, A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B* **64**, 479–498 (2002).

DRUG DISCOVERY

A Correction to the Research Article Titled: "Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data" by M. Sirota, J. T. Dudley, J. Kim, A. P. Chiang, A. A. Morgan, A. Sweet-Cordero, J. Sage, A. J. Butte

There was an error in the Supplementary Material file: table S1 was inadvertently omitted. In addition, the colors were incorrect in Fig. 2 and its caption. The corrected online version of the article and Supplementary Material is at http://stm.sciencemag.org/content/3/96/96ra77.full, and the corrected PDF version of the article is at http://stm.sciencemag.org/content/3/96/96ra77.full.pdf.

10.1126/scitranslmed.3003215

Editor's Summary

Greening Drug Discovery

Recycling is good for the environment––and for drug development too. Repurposing existing, approved drugs can speed their adoption in the clinic because they can often take advantage of the existing rigorous safety testing required by the Food and Drug Administration and other regulatory agencies. In a pair of papers, Sirota *et al.* and Dudley *et al.* examined publicly available gene expression data and determined the genes affected in 100 diseases and 164 drugs. By pairing drugs that correct abnormal gene expression in diseases, they confirm known effective drug-disease pairs and predict new indications for already approved agents. Experimental validation that an antiulcer drug and an antiepileptic can be reused for lung cancer and inflammatory bowel disease reinforces the promise of this approach.

The authors scrutinized the data in Gene Expression Omnibus and identified a disease signature for 100 diseases, which they defined as the set of mRNAs that reliably increase or decrease in patients with that disease compared to normal individuals. They compared each of these disease signatures to each of the gene expression signatures for 164 drugs from the Connectivity Map, a collection of mRNA expression data from cultured human cells treated with bioactive small molecules that is maintained at the Broad Institute at Massachusetts Institute of Technology. A similarity score calculated by the authors for every possible pair of drug and disease ranged from +1 (a perfect correlation of signatures) to −1 (exactly opposite signatures). The investigators suggested that a similarity score of −1 would predict that the drug would ameliorate the abnormalities in the disease and thus be an effective therapy.

This proved to be true for a number of drugs already on the market. The corticosteroid prednisolone, a common treatment for Crohn's disease and ulcerative colitis, showed a strong similarity score for these two diseases. The h istone deacetylase inhibitors trichostatin A, valproic acid, and vorinostat were predicted to work against brain tumors and other cancers (esophagus, lung, and colon), and there is experimental evidence that this is indeed the case. But in the ultimate test of method, the authors confirmed two new predictions in animal experiments: Cimetidine, an antiulcer drug, predicted by the authors to be effective against lung cancer, inhibited tumor cells in vitro and in vivo in mice. In addition, the antiepileptic topiramate, predicted to improve inflammatory bowel disease by similarity score, improved damage in colon tissue of rats treated with trinitrobenzenesulfonic acid, a model of the disease. These two drugs are therefore good candidates for recycling to treat two diseases in need of better therapies––lung cancer and inflammatory bowel disease––and we now have a way to mine available data for fast routes to new disease therapies.