Drug Repositioning for Cancer Therapy Based on Large-Scale Drug-Induced Transcriptional Signatures

Haeseung Lee, Seungmin Kang, Wankyu Kim *
Ewha Research Center for Systems Biology, Division of Molecular & Life Sciences, Ewha Womans University, Seoul, Korea
* wkim@ewha.ac.kr

Abstract

An in silico chemical genomics approach is developed to predict drug repositioning (DR) candidates for three types of cancer: glioblastoma, lung cancer, and breast cancer. It is based on a recent large-scale dataset of ~20,000 drug-induced expression profiles in multiple cancer cell lines, which provides i) a global impact of transcriptional perturbation of both known targets and unknown off-targets, and ii) rich information on drug’s mode-of-action. First, the drug-induced expression profile is shown more effective than other information, such as the drug structure or known target, using multiple HTS datasets as unbiased benchmarks. Particularly, the utility of our method was robustly demonstrated in identifying novel DR candidates. Second, we predicted 14 high-scoring DR candidates solely based on expression signatures. Eight of the fourteen drugs showed significant anti-proliferative activity against glioblastoma; i.e., ivermectin, trifluridine, astemizole, amlodipine, maprotioline, apomorphine, mometasone, and nortriptyline. Our DR score strongly correlated with that of cell-based experimental results; the top seven DR candidates were positive, corresponding to an approximately 20-fold enrichment compared with conventional HTS. Despite diverse original indications and known targets, the perturbed pathways of active DR candidates show five distinct patterns that form tight clusters together with one or more known cancer drugs, suggesting common transcriptome-level mechanisms of anti-proliferative activity.

Introduction

Drug repositioning (DR) refers to the identification of novel indications for existing drugs [1] and is considered an effective route for drug development because it reduces costs and bypasses safety concerns. However, discovering novel indications with DR is highly challenging, even with well-established high-throughput screening (HTS), because of the numerous combinations of both assays and drugs [2]. Due to these limitations, most repositioned drugs have been serendipitously developed. In silico DR is considered an alternative and efficient route to establish novel connections between diseases and existing drugs [3,4]. Advances in systems...
pharmacology approaches and the growth of drug-target information have increased the success of in silico DR [5,6]. A broad range of datasets has been utilized, such as sets related to chemical structure [7,8], drug-target relationship [9], and phenotypic information including drug side effects [10–14]. For example, Cheng et al. identified simvastatin and ketoconazole as potent anti-proliferative agents in breast cancer by analysing a drug-target network [15]. Particularly, large-scale chemical genomics data, such as the Connectivity Map or CMAP [16], have provided rich information on the modes-of-action of drugs that are reflected in the transcriptomic responses due to chemical perturbation. However, the relative utilities of different datasets were not rigorously evaluated because the compound set of CMAP was not large enough for integrative statistical analysis.

The latest version of CMAP consists of gene expression profiles of five cell lines treated with ~1,300 compounds, from which many in silico DR methods were developed using this dataset either alone or in combination with other information [17–25]. The negative correlation of gene expression with a disease led to the identification of topiramate for the treatment of inflammatory bowel disease (IBD) and cimetidine for the treatment of lung adenocarcinoma [19,20]. Iskar et al. further revealed the modes of action for multiple drugs using drug-induced expression modules conserved between humans and rats. Recently, a similar but highly expanded version of a chemical genomics dataset was publicly released by the NIH LINCS program (Library of Integrated Network-based Cellular Signatures). This dataset consists of gene expression signatures and protein binding, cellular phenotypic, and phosphoproteomic profiles due to chemical or genetic perturbation. Specifically, it produced the gene expression profiles of ~1,000 landmark genes (L1000) in response to >20,000 chemical perturbations across many cell lines. Additionally, they inferred transcriptome-level expression profiles of ~20,000 genes computationally using the 1,000 landmark genes [26].

In this study, we adopted an integrative approach for in silico DR using the expression signature (E) derived from the recent large-scale, chemical genomics dataset (LINCS) as well as chemical structure (S) and target signatures (T). Next, we applied our method to infer DR candidate anti-cancer drugs for glioblastoma, lung cancer, and breast cancer. We focused on the ability to identify novel DR candidates that are not structurally related to known anti-cancer drugs because structural analogues may be inferred easily by other structure-based methods [27,28]. The LINCS dataset covers a sufficiently large number of compounds that allowed the unbiased evaluation of the predictive power of each signature. We then predicted novel DR candidates for glioblastoma. The high-scoring candidate drugs were experimentally validated using cancer cell lines and patient-derived primary cells. The LINCS dataset also enabled us to interpret the mode of action of the validated DR candidates.

**Materials and Methods**

**Known drug set and compound-target information**

The known drug set (KD set or gold standard) was extracted from several public databases of DrugBank [29], CTD [30], PubChem [7], and KEGG DRUG [31] for glioblastoma, lung cancer, and breast cancer. Additionally, we manually curated active compounds and drugs from PubMed by searching abstracts that explicitly described anti-cancer activities for these three types of cancer (S1 Table). A typical query included several terms, such as ‘apoptosis’, ‘proliferation’, ‘cell growth’, ‘cytotoxicity’, ‘anti-cancer’, ‘cytotoxicity’, and ‘increased survival’ after treatment. The compound-target information was collected from 10 public data sources: DrugBank [32], KEGG [31], MATADOR [33], TTD [34], KiDB [35], BindingDB [36], ChEMBL [37], WOMBAT [38], CTD [30], and DCDB [39]. We included only the compound-target interactions supported by directed evidence, such as compound-target binding, activation,
inhibition, and reaction at a reasonably high affinity (e.g., Ki < 10 nM). Throughout the analysis, compounds and protein targets were mapped to NCBI PubChem Compound Identifiers (CID) and Entrez Gene ID as standard identifiers, respectively.

Expression signatures for the target diseases

The expression profiles for the target diseases were downloaded from either the TCGA data portal [https://tcga-data.nci.nih.gov] or NCBI GEO [40] that included both normal and disease conditions. The glioblastoma datasets of 200 patients was obtained from TCGA that were divided into four canonical subtypes—classical, mesenchymal, proneural, and neural. Four distinct expression signatures were extracted for the corresponding subtypes. For lung and breast cancer, 11 and 16 microarray datasets were obtained, respectively. The 11 lung cancer datasets consisted of GDS1761, GDS1312, GDS2771, GSE5364, GSE7670, GSE10072, GSE10799, GSE1987, GSE2088, GSE1037, and GSE11969. The 16 breast cancer datasets consisted of GDS817, GDS820, GDS823, GDS1761, GDS1925, GDS2250, GDS2617, GDS2618, GDS2635, GDS2739, GSE3564, GSE10780, GSE15852, GSE16443, GSE17072, and GSE20266. All microarray datasets were processed and normalized using the SAM package [41]. The DEGs were extracted as expression signatures using FDR<0.05 as a cutoff.

Expression signatures from the LINCS dataset

The LINCS dataset included an extensive catalog of gene-expression profiles generated by the Library of Integrated Network-based Cellular Signatures (LINCS) project from 59 human cancer cells in response to ~20,000 chemical perturbations. The LINCS team has produced the expression profiles of 1,000 landmark genes using a high-throughput Luminex-based assay [42]. The whole transcriptomic profiles for ~20,000 genes were deduced from the measured expression values of the 1,000 landmark genes. The expression signatures for the compounds were downloaded from the LINCS project page (www.lincsproject.org) that consisted of the 100 most up- and down-regulated genes in response to each compound.

Cell Viability Assay

The compounds used for the cell viability assay were purchased from Sigma-Aldrich and Enzo Life Science. Both the glioblastoma cell lines (A172, T98G, U251, and U87) and primary cells (GBL cells) were cultured in DMEM high-glucose cell culture medium (HyClone; SH30243.01) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin solution, and 30 μg/ml Plasmocure (InvivoGen, ant-pc) reagent to prevent mycoplasma contamination.

To measure the cytotoxic activity of the candidate drugs, the metabolic activity of the viable cells was measured using WST reagent (EZ-Cytox, DoGEN) in 96-well plates. The number of seeded cells was adjusted according to the growth rate of each cell type (100–500 cells/well). Twenty-four hours after seeding, the cells were treated with drugs at a concentration of 10 μM and further cultured for 72 h at 37°C. One-tenth of the medium volume of WST reagent was added to the cells, and the absorbance was measured at 450 nm after 2 h using a SpectraMax 190 microplate reader (Molecular Devices). The experiments were repeated five times.

Pathway enrichment analysis

KEGG pathways were downloaded from the MSigDB database (http://www.broadinstitute.org/gsea/msigdb), which consist of 186 gene sets representing various biological processes. Since many disease-related pathways are redundant with other signalling pathways (e.g. KO05200: Pathways in cancer include KO04310: Wnt, KO04210: Apoptosis, KO04115:p53 signaling, etc.),
we excluded such 28 pathways and the remaining 158 pathways were used. The significance of enrichment was calculated using the hypergeometric test. The LINCS dataset contains multiple signatures for the same drug from different cells and conditions, where the harmonic mean of the corresponding p-values was taken as its representative p-value. The adjusted q-value was then calculated using the Benjamini-Hochberg method. The distances between drugs were calculated as the cosine distance of $-\log$ p-value.

Results

Data collection and processing

First, we compiled a list of known drugs (KD set) as the benchmark for the three types of cancer (glioblastoma, lung cancer, and breast cancer) from four public databases as well as by manually curating 243 publications (S1 Table). We considered only the compounds that were explicitly stated as being active against the target disease—e.g., drug X induces apoptosis, inhibits proliferation, or shows cytotoxic activity. The KD set consisted of 132, 216, and 256 compounds for the treatment of glioblastoma, lung cancer, and breast cancer, respectively (Figure A in S1 File). Additionally, we collected an extensive list of 1,155 compounds (cancer drug set or CD set) that were reported to show the anti-proliferative activity against any other type of cancer including all the drugs in the KD set.

Next, the structural signature was extracted as 1,024 bits of FP2 chemical fingerprints implemented in Open Babel [43] for compounds (S_CD) or known drugs (S_KD). The target information was collected from 10 public databases comprising 342,311 compound-target interactions among 205,570 unique compounds (Figure A in S1 File). The sets of genes associated with each compound or drug were then generated and served as the target signatures (T_CD, T_KD). The expression signature (E_CD, E_KD) was downloaded from LINCS and consisted of the differentially expressed genes (DEGs) by chemical perturbation[26]. Multiple signatures for a single compound were common because expression signatures were generated in various cell lines of different origins. Additionally, the LINCS dataset included a significant fraction of redundant compound IDs and some generic compound names, 2D structures, and/or stereoisomers of the same molecular formulae were not distinguishable between different data sources. Therefore, the signatures of these similar compounds were merged together after converting them to the canonical SMILES format, although stereoisomers may show different pharmacological activities and some were originally assigned different compound IDs by LINCS. We note that ID mapping of a stereoisomer was very rare between two distinct stereoisomers, but was done mostly with generic name or SMILES without 3D information. The resulting number of unique compounds was reduced to 8,860. We defined a core set of 2,250 compounds for which all three types of signatures (S, T, and E) were available. The intersection of the core set and CD set was 304 drugs (Figure A in S1 File). Similarly, we also generated disease expression signatures (E_DIS) for glioblastoma (4 sets), lung cancer (11 sets), and breast cancer (16 sets) from TCGA [44] or public microarray datasets from GEO. The detailed procedure is described in the Materials and Methods section.

Overview of the analysis

We developed a series of classifiers to predict DR candidate drugs for the treatment of glioblastoma, lung cancer, and breast cancer. Our method utilizes three types of signatures that are derived from chemical structure (S), drug-target relation (T), and gene expression data (E). DR candidates were predicted based on the similarity of these signatures between the compounds and disease (or its known drugs). The prediction performance was thoroughly inspected in an unbiased manner using i) a conventional cross-validation scheme that utilizes known drugs
The work described herein consisted of three stages: 1) building association signatures, 2) constructing a series of classifiers, and 3) evaluating the prediction performance. The aim of the first stage was to associate compounds and a target disease (or its known anti-cancer drugs) based on the similarity of the three signature types (Fig 1A). In total, seven distinct types of associations that were independent of each other were established. First, a compound was predicted as a DR candidate based on its structural similarity to the known drugs (SCPD-SKD). The expression (E) and target (T) signatures essentially are a list of genes that could be associated by any method for gene set enrichment analysis. Therefore, we could generate six additional types of associations between the compounds (TCPD,E CPD) and disease (TKD,E KD,E DIS).

We used the Tanimoto coefficient as a measure of signature similarity because all signatures can be represented as a binary vector of 0s and 1s—i.e., the presence or absence of genes or structural fingerprints. Because multiple signatures are allowed for a single compound and disease, we calculated the mean of the Tanimoto coefficient values for a given compound-disease pair. (KD set) as a benchmark, ii) the 29 anti-cancer HTS datasets for 11,000–41,000 compounds, and iii) assays based on glioblastoma cancer cell lines and patient-derived primary cells.

The work described herein consisted of three stages: 1) building association signatures, 2) constructing a series of classifiers, and 3) evaluating the prediction performance. The aim of the first stage was to associate compounds and a target disease (or its known anti-cancer drugs) based on the similarity of the three signature types (Fig 1A). In total, seven distinct types of associations that were independent of each other were established. First, a compound was predicted as a DR candidate based on its structural similarity to the known drugs (SCPD-SKD). The expression (E) and target (T) signatures essentially are a list of genes that could be associated by any method for gene set enrichment analysis. Therefore, we could generate six additional types of associations between the compounds (TCPD,E CPD) and disease (TKD,E KD,E DIS). We used the Tanimoto coefficient as a measure of signature similarity because all signatures can be represented as a binary vector of 0s and 1s—i.e., the presence or absence of genes or structural fingerprints. Because multiple signatures are allowed for a single compound and disease, we calculated the mean of the Tanimoto coefficient values for a given compound-disease pair.
The second stage involved the construction of a series of classifiers to predict DR candidates (Fig 1B). To compare the performance of each signature, seven classifiers were constructed using a logistic regression. The resulting classifiers used a single (S, T, E), combination of the two (ST, SE, TE), or all three types (STE) of compound signatures. Finally, the prediction performance was thoroughly inspected in an unbiased manner using i) a conventional cross-validation scheme that utilizes known drugs as the gold standard, ii) the 29 large-scale HTS datasets for anti-proliferative activity, and iii) assays that use both established glioblastoma cell lines and patient-derived glioblastoma primary cells. We further interpreted the potential modes of action of drugs using pathway enrichment analysis followed by comparisons with other cancer drugs.

Performance evaluation using the known drug (KD) dataset

To ensure unbiased evaluation, we limited our analysis to the core set of 2,250 compounds so that different classifiers can be compared using exactly the same benchmark. The core set included 79, 100, and 132 known anti-cancer drugs (KD set) as benchmarks for glioblastoma, lung cancer, and breast cancer, respectively. The negative set (1,946 compounds) was prepared by excluding the CD set from the core set.

As stated in the previous section, we constructed seven different classifiers based on a single signature (S-, T-, and E-classifier) or combination of multiple signatures (ST-, SE-, TE-, and STE-classifier). Their relative performances were evaluated using 100 rounds of three-fold cross-validation schemes. The classifiers were based on a logistic regression that automatically weights the component features to yield a unified prediction score—the DR score—scaled from zero to one. The target-based classifiers (T) perform better than the other classifiers (S, E) in identifying known drugs (KD core set) in all three types of cancer (Fig 2). This trend was consistently observed in combination with other signatures (ST, TE > SE). We reasoned that the targets of known drugs (KD) tend to be better characterized than other compounds, and target-based classifiers may consequently favor known drugs. Therefore, we further evaluated the classifiers in a more unbiased manner in the following sections.

Evaluation in comparison with public anti-cancer HTS

To avoid potential bias, we collected dozens of large-scale anti-cancer HTS assays from PubChem [45]. They consisted of eight, thirteen, and eight HTS assays for GBM, lung cancer, and breast cancer cell lines, respectively (Table A in S1 File). These assays all used different cell
lines and were reasonably distinct from each other in terms of both assayed and hit compounds—e.g., only a 20% hit overlap was observed between the two assays for GBM (AID57, AID59) (Figure B in S1 File). Again, we limited our evaluation to the assayed compounds that also belonged to the core set. To focus on the ability to predict novel DR candidates, known drugs (KD) were excluded from the hit compounds. The remaining hits of each assay were divided into two categories: i) anti-cancer hits that belonged to the CD set or had two or more structural analogs in the CD set with a Tanimoto coefficient $>$ 0.7, and ii) novel hits that were not included in the former set. Therefore, the anti-cancer hits represented those that could be readily predicted based on structural similarity and the novel hits, which were the difficult cases and are the focus of this study.

Interestingly, the performance of each signature was in contrast to the result of the previous evaluation based on the KD set. The expression-based classifier (E) best predicted anti-cancer hits in all three cancers (Fig 3A). By contrast, the target-based classifier (T) performed worst. This trend was more obvious in the prediction of novel hits (Fig 3B). The classifiers based on structure (S) or target signature (T) were essentially unable to predict novel hits (median AUC = 0.41~0.62), whereas the expression signature (E) performed reasonably well (median AUC = 0.73~0.79, Fig 3B). This trend was consistently observed in combination with other
signatures (SE > TE > ST). Notably, the single signature (E) consistently performed better alone than in combination with other signatures (SE, TE) for all three cancer types. Overall, the expression signature (E) was more informative than the structure (S) or target (T) signatures among the HTS datasets for the three different cancers. The data also suggested that previous evaluations of any in silico DR could be biased if known drugs were used as the only benchmark set, particularly if the method was highly dependent on target information. Notably, focusing on the core set retains only a small set of compounds for evaluation, but this approach remained sufficient for unbiased statistical tests using the same set of compounds (Table A in S1 File). The evaluation results for all the compounds in the HTS dataset showed essentially the same trend (Figure C in S1 File).

**Experimental validation of DR candidate drugs for glioblastoma**

Glioblastoma multiforme (GBM) is the most common and aggressive form of brain cancer showing a highly poor prognosis despite concurrent or sequential chemo-radiotherapy [46]. Temozolomide (TMZ) is a first-line chemotherapeutic agent for the treatment of GBM. As an alkylating agent, TMZ transfers methyl groups to the purine bases of DNA, causing single- and double-strand DNA breaks and subsequent apoptotic cell death. Most studies reported that chemotherapy conferred a limited overall survival benefit to GBM patients. The median time to recurrence was only 6.9 months after standard treatment [47]. Therefore, the need for more effective drugs is clear and unmet.

Here, we listed DR candidates for GBM and experimentally tested their anti-tumor activities solely based on the expression signature. First, the 8,860 compounds were ordered by the DR score, i.e. the percentile rank of the predictions by the same classifier (E) as described in the previous section. Next, we applied several filters: a) FDA-approved; b) DR score >0.9; and c) indication to pass brain-blood barrier (BBB) either by literature or by in silico prediction at http://www.cbligand.org/BBB. One high-scoring candidate (ivermectin, DR score = 0.98) was also included despite being indicated not to pass BBB. The final 14 DR candidates were selected, and their anti-tumor activities were tested using four GBM cell lines and eight patient-derived primary cells. Our DR scores correlated well with the anti-tumor activities among the cell lines tested in terms of both anti-proliferative activity and the fraction of significant growth inhibition (Fig 4A and 4B). Eight of the 14 candidates were significant hits in three or more GBM cells at 10 μM. Notably, the top seven candidates showed strong anti-tumor activity, corresponding to an approximately 20 fold hit enrichment from ~5% hit rate of conventional high-throughput screening.

We surveyed the literature for reports concerning the anti-cancer activity of the eight active DR candidates, as summarized in Table 1. We found that four drugs (ivermectin, astemizole, nortriptyline, and apomorphine) were previously reported to show anti-tumor activity in brain cancer. Three drugs (trifluridine, amlodipine, and maprotiline) demonstrated anti-tumor activity in other cancers but not in glioblastoma. Mometasone is a corticosteroid and is considered a novel drug for cancer because it has not been reported to show anti-tumor activity in any type of cancer. Corticosteroids have been used to reduce peritumoral edema and chemotherapy-associated side effects, such as pain, nausea, and vomiting [48]. All active DR candidates except ivermectin have been shown to be able to pass the BBB in either literature reports or via in silico prediction.

**Interpretation of drug modes of action for the active DR candidates**

Drug-induced transcriptomic profiles reflect direct and indirect changes of cellular physiology in omics scale and provide rich information on the pharmacological mechanism of drugs.
order to interpret their modes of action, we performed pathway enrichment analysis of up-
and down-regulated genes using KEGG pathways. Collectively, 32 and 17 pathways were

Fig 4. The high-scoring DR candidates for glioblastoma among the FDA-approved drugs that were
predicted based only on the expression signatures. (A) DR scores, (B) the fraction of significantly
inhibited cells summarizing the results of (C), (C) the anti-proliferative activities (% growth inhibition) for the
four glioblastoma cell lines (four cell lines of TG98, A172, U251MG, and U87MG) and the eight patient-
derived primary cells (the GBLs) at 10 μM, (D) in silico prediction scores for BBB transport based on http://
www.cbligand.org/BBB. The red asterisk indicates experimental support for passing the BBB according to the
literature. Overall, anti-proliferative activities across glioblastoma cells strongly correlated with the rankings
by the DR score. Most DR candidates were shown to be able to pass the BBB.

doi:10.1371/journal.pone.0150460.g004
significantly enriched in the up- and down-regulated genes, respectively, at a cut-off q-value < 0.1 (Fig 5). Overall, the data suggest plausible mechanisms for anti-tumor activities that are shared among the DR candidates. Up-regulated pathways included apoptosis, amino acid and lipid metabolism, and tumor-suppressive P53/MAPK/WNT signaling, which suggested that the activities of anti-proliferative processes were increased. Conversely, cell proliferative pathways, such as the cell cycle, DNA replication, DNA repair, and ribosome assembly, were down-regulated.

Four drugs (amlodipine, astemizole, maprotiline, and nortriptyline) strongly increased lipid metabolism (steroid, terpenoid, and fatty acids). Accordingly, amlodipine, a calcium channel blocker, was previously reported to induce steroidogenesis [66]. Amlodipine and maprotiline moderately up-regulated apoptosis pathways that may be associated with PKD1 overexpression and caspase-3 activation, respectively [67,68]. Notably, apomorphine seemed to act via a distinct mechanism: genes associated with mitochondrial energy metabolism were strongly down-regulated without noticeable changes in apoptosis, metabolism, or DNA repair. These mitochondrial metabolic genes included genes that encode ATP synthases (ATP5O, ATP5D), cytochrome oxidases (COX8A, COX7), and NADH dehydrogenases (NADUFS8, NADUFB2). Mitochondrial dysfunction has been well established to modulate apoptosis and tumorigenesis [69]. The other two DR candidates (nortriptyline and maprotiline) also decreased mitochondrial energy metabolism. Down regulation of cell cycle genes was commonly observed among all eight DR candidates. Five drugs significantly decreased the expression of DNA repair genes. Overall, these active DR candidates all significantly perturbed multiple tumorigenic or tumor suppressive pathways, which may direct cancer cells toward anti-proliferative outcome.

We also performed a cluster analysis of the eight DR candidates together with the 69 cancer drugs included in the LINCS dataset using the enrichment pattern of the same 32 up- and 17 down-regulated pathways (Fig 6). These cancer drugs frequently up-regulate P53, MAPK, apoptosis and immune signaling. Many cancer drugs as well as our DR candidates down-regulate cell cycle-, DNA repair-, and p53 signaling pathways. Seven of the eight DR candidates show a highly similar enrichment pattern with other cancer drugs (cluster I–V in Fig 6). Four DR candidates (amlodipine, astemizole, nortriptyline, and maprotiline) and eight cancer drugs belong to two related clusters (I and II). Ivermectin is grouped with four other cancer drugs (cluster
IV). Mometasone shows a pattern similar to those of trametinib and exemestane (cluster III), and trifluridine shows a pattern similar to that of axitinib (cluster V). At the pathway level, the drugs in the cluster show a similar pattern of transcriptional perturbation, although their known targets or modes of action are heterogeneous. For example, ivermectin is an antiparasitic drug that selectively binds to glutamate-gated chloride ion channels. All other cancer drugs in the same cluster have different known or canonical targets (BRAF/CRAF-dabrafenib, sorafenib, vemurafenib, DNA methyltransferase-azacitidine). Similarly, the other clusters showed a high degree of heterogeneity of known targets, suggesting that the downstream effects between the DR candidates and anti-cancer drugs are shared via the same off-target or shared regulatory events. The lists of drugs in each cluster are summarized in Table B in S1 File.

**Discussion**

Although most drug development projects have been aimed to design a specific modulator for a target, polypharmacology—multiple targets for a single drug—may be more prevalent than
expected. According to a recent report by Ciceri and colleagues, an unbiased screening of 628 kinase inhibitors identified 20 hits (3.2% hit rate) that also strongly inhibited BRD4 (>90% inhibition at 50 μM) [70]. Two of these kinase-bromodomain dual inhibitors (0.32% hit rate) showed an IC50 in the nano molar range. Because both kinases and BRD4 are promising targets for cancer therapy, dual inhibitors may be an effective strategy to overcome cancer heterogeneity or resistance. Extrapolating this result to thousands of druggable targets in the entire human proteome, most known and investigational drugs may have one or more off-targets. We surveyed the number of targets per compound in our drug-target dataset collected from 10 public sources. Among the FDA approved drugs, more than 60% has one or more targets. Therefore, the idea of a ‘single drug–multiple targets’ may become a general assumption in drug development, similar to the notion ‘one gene–one polypeptide’ being replaced by ‘one gene–multiple polypeptides’. Instead of striving to avoid off-targets, active exploitation of
polypharmacology may improve treatments and benefit patients. Advances in chemical genomic approaches allow comprehensive understanding of drug modes of action at the whole transcriptome or proteome level, reflecting combined effects of multiple targets of a drug.

The efficient and systematic identification of clinically relevant off-targets and novel indications is a challenge. Although many in silico DR methods were developed, their prediction performances are not thoroughly assessed in most cases. Until recently, an in silico DR method was usually evaluated based on the AUC (area under the curve) of its receiver operating characteristic (ROC) curve using a collection of known or gold standard drugs. Experimental validation was performed in only a small number of the top-scoring DR candidates. Caution is required because the gold standard set is almost always incomplete—i.e., it includes a significant fraction of unknown false negatives or hidden hits. In fact, the very purpose of any in silico DR is to identify such hidden hits rather than to rediscover known drugs. Often, the ability to identify such novel DR candidates has not been thoroughly investigated or estimated using indirect information (e.g., overlap with clinical trial drugs). As we demonstrated, the benchmark dataset e.g., known drugs vs large-scale HTS dataset, can dramatically influence the performance evaluation.

The CMAP dataset is the first large-scale transcriptional profile and includes >1,300 compounds. Although it triggered the development of many in silico DR methods that employ chemical genomic approaches, the success of these methods was mostly anecdotal due to the limited number of transcriptome-profiled compounds. The recent LINCS dataset includes the transcriptional profiles of ~10,000 unique compounds from multiple cell lines. It not only increased the compound coverage by eight fold compared with CMAP, but allowed quantitative and unbiased evaluations via cross-comparisons with multiple benchmarks, such as known cancer drugs, public HTS dataset, and wet experiments. We also managed to compare the performance of the three most frequently used data types (structure, target, and expression) for in silico DR, an analysis that was previously unfeasible due to the small size of the CMAP dataset. Although our analysis is limited to a simple logistic regression classifier, it strongly suggests that the expression signature was the most predictive, particularly in identifying novel DR candidates. Next, we identified eight DR candidates for glioblastoma based solely on expression signatures. Our DR score positively correlated with the anti-proliferative activities in cancer cell lines and primary cells, showing an approximately 20-fold enrichment of active hits. Notably, none of the expression signatures from LINCS were generated using glioblastoma cell lines, which demonstrated that our in silico DR method remains valid across different cancer types and potentially for other diseases. We reasoned that the consensus of multiple signatures from diverse cell lines provided sufficient information to overcome the heterogeneity and noise from individual signatures. Finally, we exploited large-scale, drug-induced transcriptional profiling to interpret the modes of action of our DR candidates. These candidates showed unique patterns of pathway perturbation that were shared with multiple other cancer drugs. Common patterns were observed not only among the drugs of the same target but also among drugs of different classes, which strongly indicated many unknown off-targets or cross-talk between different targets. In the post-chemical genomic era, we may expand the scope of drug repositioning from merely seeking polypharmacology (i.e. one drug for multiple targets) to broadly considering omics-scale effects by both known targets and unknown off-targets, such as the entire transcriptome, proteome, and metabolome.

Supporting Information

S1 File. Supporting figures and tables.

(DOCX)
S1 Table. List of known drugs collected from public databases and literature. (XLSX)

Acknowledgments
We thank Prof SH Paek for providing primary cells derived from glioblastoma patients. We appreciate Prof SH Lee and DH Hwang for valuable discussions and comments.

Author Contributions
Conceived and designed the experiments: WK. Performed the experiments: HL SK. Analyzed the data: HL SK. Contributed reagents/materials/analysis tools: WK HL SK. Wrote the paper: WK HL SK.

References
1. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004; 3: 673–83. doi:10.1038/nrd1468 PMID: 15286734
2. Ekins S, Williams AJ, Krasowski MD, Freundlich JS. In silico repositioning of approved drugs for rare and neglected diseases. Drug Discov Today. Elsevier Ltd; 2011; 16: 298–310. doi:10.1016/j.drudis.2011.02.016
3. Hurle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational drug repositioning: from data to therapeutics. Clin Pharmacol Ther. 2013; 93: 335–41. doi:10.1038/clpt.2013.1 PMID: 23443757
4. Lussier Ya, Chen JL. The emergence of genome-based drug repositioning. Sci Transl Med. 2011; 3: 96ps35. doi:10.1126/scitranslmed.3001512 PMID: 21849663
5. Liu Z, Fang H, Reagan K, Xu X, Mendrick DL, Slikker W, et al. In silico drug repositioning: what we need to know. Drug Discov Today. Elsevier Ltd; 2013; 18: 110–5. doi:10.1016/j.drudis.2012.08.005
6. Ma D-L, Chan DS-H, Leung C-H. Drug repositioning by structure-based virtual screening. Chem Soc Rev. 2013; 42: 2130–41. doi:10.1039/c2cs35357a PMID: 23288298
7. Chen B, Wild D, Guha R. PubChem as a source of polypharmacology. J Chem Inf Model. 2009; 49: 2044–55. doi:10.1021/ci0901876 PMID: 19708682
8. Garcia-Sosa A, Oja M. DrugLogit: logistic discrimination between drugs and nondrugs including disease-specificity by assigning probabilities based on molecular properties. J Chem Inf Model. 2012; n/a—n/a. doi:10.2121/ci200587h
9. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. Nucleic Acids Res. Oxford University Press; 2011; 39: D1035–D1041. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3013709&tool=pmcentrez&rendertype=abstract. doi:10.1093/nar/gkq1126 PMID: 21059682
10. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. Science. 2008; 321: 263–6. doi:10.1126/science.1158140 PMID: 18621671
11. Ye H, Liu Q, Wei J. Construction of drug network based on side effects and its application for drug repositioning. PLoS One. Public Library of Science; 2014; 9: e87864. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087864. doi:10.1371/journal.pone.0087864 PMID: 24505324
12. Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. Csermely P, editor. PLoS One. Public Library of Science; 2011; 6: e28025. doi:10.1371/journal.pone.0028025 PMID: 22205936
13. Wang Y, Chen S, Deng N, Wang Y. Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data. PLoS One. Public Library of Science; 2013; 8: e78518. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0078518. doi:10.1371/journal.pone.0078518 PMID: 24244318
14. Iwata H, Sawada R, Mizutani S, Yamanishi Y. Systematic Drug Repositioning for a Wide Range of Diseases with Integrative Analyses of Phenotypic and Molecular Data. J Chem Inf Model. 2015; 55: 446–459. doi:10.1021/ci500670q PMID: 25602292
15. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol. 2012; 8: e1002503. doi:10.1371/journal.pcbi.1002503 PMID: 22589709
16. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006; 313: 1929–35. doi: 10.1126/science.1132939 PMID: 17008526

17. Iorio F, Isacchi A, di Bernardo D, Brunetti-Pierri N. Discovery of drug mode of action and drug repositioning from transcriptional responses. PNAS. 2010; 6: 1204–5. doi:10.1073/pnas.100138107

18. Hu G, Agarwal P. Human disease-drug network based on genomic expression profiles. PLoS One. 2009; 4: e6536. doi: 10.1371/journal.pone.0006536 PMID: 19657382

19. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan A a, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med. 2011; 3: 96ra77. doi:10.1126/scitranslmed.3001318 PMID: 21849665

20. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med. 2011; 3: 96ra76. doi: 10.1126/scitranslmed.3002648 PMID: 21849664

21. Iskar M, Zeller G, Blattmann P, Campillos M, Kuhn M, Kaminska KH, et al. Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. Mol Syst Biol. EMBO and Macmillan Publishers Limited; 2013; 9: 662. doi:10.1038/msb.2013.20 PMID: 23632384

22. Wang K, Sun J, Zhou S, Wan C, Qin S, Li C, et al. Prediction of Drug-Target Interactions for Drug Repositioning Only Based on Genomic Expression Similarity. Markel S, editor. PLoS Comput Biol. Public Library of Science; 2013; 9: e1003315. Available:http://dx.plos.org/10.1371/journal.pcbi.1003315. doi: 10.1371/journal.pcbi.1003315 PMID: 24244130

23. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol. Nature Publishing Group; 2011; 7: 496. doi: 10.1038/msb.2011.26

24. Iorio F, Shrestha RL, Levin N, Boilot V, Garnett MJ, Saez-Rodriguez J, et al. A Semi-Supervised Approach for Refining Transcriptional Signatures of Drug Response and Repositioning Predictions. PLoS One. 2015; 10: e0139446. doi: 10.1371/journal.pone.0139446 PMID: 26452147

25. Shin E, Lee YC, Kim SR, Kim SH, Park J. Drug Signature-based Finding of Additional Clinical Use of LC28-0126 for Neutrophilic Bronchial Asthma. Sci Rep. Nature Publishing Group; 2015; 5: 17784. doi: 10.1038/srep17784 PMID: 26626943

26. Vidović D, Koleti A, Schürer SC. Large-scale integration of small molecule-induced genome-wide transcriptional responses, Kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing systems-level drug action. Front Genet. 2014; 5: 342. doi: 10.3389/fgene.2014.00342 PMID: 25324859

27. Keiser MJ, Setola V, Irwin JJ, Siddiqi MI. Virtual screening strategies: Recent advances in the identification and design of anti-cancer agents. Methods. Elsevier Inc.; 2015; 71: 64–70. doi: 10.1016/j.ymeth.2014.08.010

28. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. Nucleic Acids Res. Oxford University Press; 2011; 39: D1035–D1041. doi: 10.1093/nar/gkr1126 PMID: 21059682

29. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, et al. The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Res. 2013; 41: D1104–14. doi: 10.1093/nar/gks994 PMID: 23093600

30. Kaniehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012; 40: D1128–36. Available:http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3245130&tool=pmcentrez&rendertype=abstract. doi: 10.1093/nar/gkr797 PMID: 21948793
35. Roth BL, Lopez E, Patel S, Kroeze WK. The Multiplicity of Serotonin Receptors: Uselessly Diverse Molecules or an Embarrassment of Riches? Neurosci. 2000; 6: 252–262. doi: 10.1177/107385840000600408

36. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res. 2007; 35: D198–201. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1751547&tool=pmcentrez&rendertype=abstract. PMID: 17145705

37. Bento a P, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. Nucleic Acids Res. 2014; 42: D1083–90. doi: 10.1093/nar/gkt1031 PMID: 24214965

38. Olah M, Rad R, Ostopovic L, Bora A, Hadaruga N, Hadaruga D, et al. WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery [Internet]. Schreiber SL, Kapoor TM, Wess G, editors. Weinheim, Germany: Wiley-VCH Verlag GmbH; 2007. doi: 10.1002/9783527619375

39. Liu Y, Wei Q, Yu G, Gai W, Li Y, Chen X. DCDB 2.0: a major update of the drug combination database. Database (Oxford). 2014; 2014: bau124. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4275564&tool=pmcentrez&rendertype=abstract.

40. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013; 41: D991–5. Available: http://www.ncbi.nlm.nih.gov/geo/.

41. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001; 98: 5116–21. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=331738&tool=pmcentrez&rendertype=abstract. doi: 10.1093/nar/gks1193 PMID: 23193258

42. Peck D, Crawford ED, Ross KN, Stegmaier K, Golub TR, Lamb J. A method for high-throughput gene expression signature analysis. Genome Biol. 2006; 7: R61. Available: http://genomebiology.com/2006/7/7/R61. PMID: 16859521

43. O’Boyle NM, Banch M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011; 3: 33. Available: http://www.jcheminf.com/content/3/1/33. doi: 10.1186/1758-2946-3-33 PMID: 21982300

44. Brennan CW, Verhaak RGW, McKenna A, Campos B, Nourshmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. Cell. 2013; 155: 462–77. doi: 10.1016/j.cell.2013.09.034 PMID: 24120142

45. Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO, et al. An overview of the PubChem BioAssay resource. Nucleic Acids Res. Oxford University Press; 2010; 38: D255–D266. Available: http://www.ncbi.nlm.nih.gov/geo/.

46. Meir EG Van, Hadjipanayis CG, Norden AD, Shu H, Wen PY, Olson JJ. Exciting New Advances in Neuro-Oncology: The Avenue to a Cure for Malignant Glioma. 2010; 60: 166–193. doi: 10.3322/caac.20069.Exciting PMID: 20445000

47. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med. United States; 2005; 352: 10–19. doi: 10.1056/NEJMoa043330 PMID: 15758009

48. Markman M, Sheidler V, Ettinger DS, Quaskey SA, Mellitis ED. Antiemetic efficacy of dexamethasone. Neuro-Oncology: The Avenue to a Cure for Malignant Glioma. 2010; 60: 589–599. doi: 10.3233/JAD-2010-1262

49. Holmberg H, Sjöberg M, Almqvist B, Lindgren M. TAS-102 confers high trifluridine incorporation into DNA and sustained antitumor activity in mouse xenograft models. Oncol Rep. 2014; 31: 243–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/22495656. PMID: 22495656

50. RESULTS / DISCUSSION. Our results indicate that TFT can indeed cross the BBB and that it can be detected in brain, along with its metabolites, as early as 15 minutes post-injection. Four major resonances were identified in the spectra of brain extract. 1988; 14: 60201.
54. Huang X, Dubuc AM, Hashizume R, Berg J, He Y, Wang J, et al. Voltage-gated potassium channel EAG2 controls mitotic entry and tumor growth in medulloblastoma via regulating cell volume dynamics. Genes Dev. 2012; 26: 1780–1796. doi:10.1101/gad.193789.112 PMID: 22855790

55. Jangi SM, Díaz-Pérez JL, Ochoa-Lizarraide B, Martin-Ruiz I, Asumendi A, Pérez-Yarza G, et al. H1 histamine receptor antagonists induce genotoxic and caspase-2-dependent apoptosis in human melanoma cells. Carcinogenesis. 2006; 27: 1787–1796. doi:10.1093/carcin/bgl021 PMID: 16569656

56. Huang BS, Leenen FHH. Sympathoinhibitory and depressor effects of amlodipine in spontaneously hypertensive rats. J Cardiovasc Pharmacol. United States; 2003; 42: 153–160. PMID: 12883316

57. Yoshida J, Ishibashi T, Nishio M. G1 cell cycle arrest by amlodipine, a dihydropyridine Ca2+ channel blocker, in human epidermoid carcinoma A431 cells. Biochem Pharmacol. England; 2007; 73: 943–953. doi: 10.1016/j.bcp.2006.12.011 PMID: 17217918

58. Taylor JM, Simpson RU. Inhibition of cancer cell growth by calcium channel antagonists in the athymic mouse. Cancer Res. 1992; 52: 2413–2418. PMID: 1533173

59. Cloonan SM, Drozgowska A,ayne D, Williams DC. The antidepressants maprotiline and fluoxetine have potent selective antiproliferative effects against Burkitt lymphoma independently of the norepinephrine and serotonin transporters. Leuk Lymphoma. England; 2010; 51: 523–539. doi: 10.3109/10428190903552112 PMID: 17217918

60. Hsu S-S, Chen W-C, Lo Y-K, Cheng J-S, Yeh J-H, Cheng H-H, et al. Effect of the antidepressant maprotiline on Ca2+ movement and proliferation in human prostate cancer cells. Clin Exp Pharmacol Physiol. Australia; 2004; 31: 444–449. doi: 10.1111/j.1440-1681.2004.04024.x

61. Neef C, Laar T Van. Relationships of Apomorphine in Patients with Parkinson’s Disease. Clin Pharmacokinet. 1999; 37: 257–271. PMID: 10511920

62. Lissoni P, Malugani F, Manganini V, Ardizzoia A, Gardani G, Bartolacelli E, et al. Psychooncology and cancer progression-related alterations advanced cancer patients. 2003; 24.

63. Ejsing TB, Linnet K. Influence of P-glycoprotein inhibition on the distribution of the tricyclic antidepressant nortriptyline over the blood-brain barrier. Hum Psychopharmacol. 2005; 20: 149–153. doi: 10.1002/hup.667 PMID: 15624117

64. Pilkington GJ, Akinwunmi J, Amar S. The role of tricyclic drugs in selective triggering of mitochondria-mediated apoptosis in neoplastic glia: a therapeutic option in malignant glioma? Radiol Oncol. 2006; 40: 73–85.

65. Parker KA, Glaysher S, Hurren J, Knight LA, McCormick D, Suovouri A, et al. The effect of tricyclic antidepressants on cutaneous melanoma cell lines and primary cell cultures. Anticancer Drugs. 2012; 23: 65–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/21897201. doi: 10.1097/CAD.0b013e32834b1894 PMID: 21897201

66. Abdou HS, Villeneuve G, Tremblay JJ. The calcium signaling pathway regulates Leydig cell steroidogenesis through a transcriptional cascade involving the nuclear receptor NR4A1 and the steroidogenic acute regulatory protein. Endocrinology. 2013; 154: 511–20. Available: http://www.ncbi.nlm.nih.gov/pubmed/23183170. doi: 10.1210/en.2012-1767 PMID: 23183170

67. Jan C-R, Su J-A, Teng C-C, Sheu M-L, Lin P-Y, Chi M-C, et al. Mechanism of maprotiline-induced apoptosis: role of [Ca2+]i, ERK, JNK and caspase-3 signaling pathways. Toxicology. 2013; 304: 1–12. Available: http://www.ncbi.nlm.nih.gov/pubmed/23219590. doi: 10.1016/j.tox.2012.11.013 PMID: 23219590

68. Ohba T, Watanabe H, Murakami M, Radovanovic M, Iino K, Ishida M, et al. Amlodipine inhibits cell proliferation via PKD1-related pathway. Biochem Biophys Res Commun. 2008; 369: 376–81. Available: http://www.ncbi.nlm.nih.gov/pubmed/18298949. doi: 10.1016/j.bbrc.2008.02.075 PMID: 18298949

69. Boland ML, Chourasia AH, Macleod KD. Mitochondrial dysfunction in cancer. Front Oncol. 2013; 3: 292. doi: 10.3389/fonc.2013.00292 PMID: 24350057

70. Ciceri P, Müller S, O’Mahony A, Fedorov O, Filippakopoulos P, Hunt JP, et al. Dual kinase-bromodomain inhibitors for rationally designed polypharmacology. Nat Chem Biol. 2014; 10: 305–12. doi: 10.1038/nchembio.1471 PMID: 24584101