

# CDA: Combinatorial Drug Discovery Using Transcriptional Response Modules

Ji-Hyun Lee<sup>1,2</sup>, Dae Gyu Kim<sup>1</sup>, Tae Jeong Bae<sup>1,2</sup>, Kyoohyoung Rho<sup>1,2</sup>, Ji-Tae Kim<sup>1,2</sup>, Jong-Jun Lee<sup>1</sup>, Yeongjun Jang<sup>3</sup>, Byung Cheol Kim<sup>1</sup>, Kyoung Mii Park<sup>1,2</sup>, Sunghoon Kim<sup>1,4\*</sup>

**1** Medicinal Bioconvergence Research Center, Seoul National University, Seoul, South Korea, **2** Information Center for Bio-pharmacological Network, Seoul National University, Suwon, South Korea, **3** Korean BioInformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), Yuseong-gu, Deajeon, South Korea, **4** WCU Department of Molecular Medicine and Biopharmaceutical Sciences, Seoul National University, Seoul, South Korea

## Abstract

**Background:** Anticancer therapies that target single signal transduction pathways often fail to prevent proliferation of cancer cells because of overlapping functions and cross-talk between different signaling pathways. Recent research has identified that balanced multi-component therapies might be more efficacious than highly specific single component therapies in certain cases. Ideally, synergistic combinations can provide 1) increased efficacy of the therapeutic effect 2) reduced toxicity as a result of decreased dosage providing equivalent or increased efficacy 3) the avoidance or delayed onset of drug resistance. Therefore, the interest in combinatorial drug discovery based on systems-oriented approaches has been increasing steadily in recent years.

**Methodology:** Here we describe the development of Combinatorial Drug Assembler (CDA), a genomics and bioinformatics system, whereby using gene expression profiling, multiple signaling pathways are targeted for combinatorial drug discovery. CDA performs expression pattern matching of signaling pathway components to compare genes expressed in an input cell line (or patient sample data), with expression patterns in cell lines treated with different small molecules. Then it detects best pattern matching combinatorial drug pairs across the input gene set-related signaling pathways to detect where gene expression patterns overlap and those predicted drug pairs could likely be applied as combination therapy. We carried out *in vitro* validations on non-small cell lung cancer cells and triple-negative breast cancer (TNBC) cells. We found two combinatorial drug pairs that showed synergistic effect on lung cancer cells. Furthermore, we also observed that halofantrine and vinblastine were synergistic on TNBC cells.

**Conclusions:** CDA provides a new way for rational drug combination. Together with phExplorer, CDA also provides functional insights into combinatorial drugs. CDA is freely available at <http://cda.i-pharm.org>.

**Citation:** Lee J-H, Kim DG, Bae TJ, Rho K, Kim J-T, et al. (2012) CDA: Combinatorial Drug Discovery Using Transcriptional Response Modules. PLoS ONE 7(8): e42573. doi:10.1371/journal.pone.0042573

**Editor:** Ju-Seog Lee, University of Texas MD Anderson Cancer Center, United States of America

**Received:** March 26, 2012; **Accepted:** July 9, 2012; **Published:** August 8, 2012

**Copyright:** © 2012 Lee et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by the grants of the Global Frontier (NRF-M1AXA002-2010-0029785) and the Research Information Center Supporting Program (370C-20090004) and the WCU project (R31-2008-000-10103-0) of the Ministry of Education, Science, and Technology and Korea Healthcare Technology (A092255-0911-1110100), the Ministry of Health and Welfare Affairs, and Gyeonggi-do to Dr. Sunghoon Kim, an EU project of the 7th framework programme (METOXIA), and by the Korean Ministry of Education, Science and Technology (MEST) under grant number 20110002321. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [sungkim@snu.ac.kr](mailto:sungkim@snu.ac.kr)

## Introduction

Advances in *in vitro* test systems have shifted drug research from animal studies to target-oriented research [1]. Combining this process with genomic research, agents specifically targeting unique proteins related to specific disease have been found. Amongst these successful stories of targeted agents is the BCR-ABL kinase inhibitor imatinib (Gleevec; Novartis), which is using for the treatment of chronic myelogenous leukemia (CML). However, in such cases, drug resistance arises possibly owing to the diversity of mutations of the gene encoding BCR-ABL as well as other pathways on parallel signalling pathways [2]. Despite successes such as these, many other drug candidates targeting disease-associated gene products have been found to be inefficient or to

cause severe side effects. So the limitations of the single protein targeted agent paradigm have come to surface.

Living systems rely on complex signaling pathways to maintain their performance in the face of various perturbations [3]. This complexity appears to pose a barrier for anticancer therapies targeting single signalling pathways. Cancer cells possess compensatory mechanisms to overcome perturbations where they occur at one signalling axis and so therapies targeting only one pathway can fail in clinical trials due to lack of efficacy, or be overcome by mutations at an important receptor [4]. Recent research has identified that in some cases, balanced multi-component therapies might be better than highly specific single component therapies [5–7]. These drug combinations are pharmacodynamically synergistic, additive or antagonistic as their effects are greater than, equal to, or less than the summed effects of individual drugs,

respectively [8]. These models have garnered interest in the possibility of effective combinatorial drug discovery based on systems-oriented approaches [9–13].

Geva-Zatorsky et al. found that protein responses to combinations of drugs were described accurately by a linear superposition (weighted sum) of their responses to each drug alone [14]. With this in mind, we designed a system for multiple signaling pathways targeting combinatorial drug discovery using gene expression profile. We assumed that if there are two different drugs which regulate two different disease-associated pathways individually, combination of them might be effective unless they affect to each other in unanticipated ways. Based on this model, expression pattern matching methods should be a valuable to quantify the degree of functional similarity among genetic perturbation, disease, and drugs. However, despite current data bases of mRNA expression profiles, which contain thousands of data points, many of which are available to the public, the number of combinatorial drug discovery approaches based on expression profiles is less than might be expected.

Here we introduce the CDA, for predicting combinatorial drug candidates that target multiple signaling pathways. CDA contains 6,100 expression profiles representing 1,309 molecules which were imported from Connectivity Map [15]. When a user submits “up probe sets” and “down probe sets”, CDA starts hyper-geometric tests for signaling pathway gene set enrichment analysis. Next signaling pathway expression pattern analysis and drug set pattern analysis are performed to measure expression pattern similarity between input signatures and 6,100 expression profiles. These analyses focus on the signaling pathways which are selected in the gene set enrichment analysis (the previous step). CDA then generates lists of single drugs and combinatorial drugs showing similar expression patterns. If user input signatures are disease-related significant probe sets, high negative scoring drugs can be considered candidate drugs for treating individuals whose diseased tissues show opposite gene expression aberrations in signalling pathways as the input cell line.

We present the results in two different formats: a table view of scores and experimental details, and a network view to visualize relationships between signaling pathway entities and known drugs, proteins and diseases. phExplorer, a graphical data visualization software program, allows users to browse the complex relationships in an interactive and dynamic manner, providing clues to how chemicals work synergistically on certain signaling pathways. To validate the technique, we performed two *in vitro* combinatorial drug discovery studies, on non-small cell lung cancer cells and triple-negative breast cancer (TNBC) cells, and succeeded in each case to find combinatorial drug pairs that exerted synergistic effects in cell culture.

## Results

### Drug Combination Suggestion through Transcription Response Module Analysis

CDA uses gene expression data in cellular models to pinpoint combinatorial drug pairs that can regulate multiple signaling pathways that potentially synergize to cause disease states, or which through alternate pathways compensate to reduce the efficacy of a drug targeting only one pathway. The combinatorial drug possibility is predicted by gene expression pattern comparison within the selected disease-related signaling pathways. The possibility is scored using Kolmogorov-Smirnov statistics. CDA is composed of four steps; 1) Preparing input signatures and gene set enrichment analysis of signaling pathways 2) Pathway expression pattern analysis 3) Drug set pattern analysis 4) Counting of the number of pathways which show positive/negative correlations with input signatures for drug ranking (See methods for more

details and Figure 1). To validate the technique, we performed one *in silico* single drug discovery study and two *in vitro* combinatorial drug discovery studies. In an *in silico* validation for single drug analysis, CDA successfully identified a molecule having similar function (Case one). As the discovered combinatorial drug pairs were mostly novel, we carried out *in vitro* validations on non-small cell lung cancer cells and triple-negative breast cancer (TNBC) cells (Case two and three).

### Case One: Molecules Function as Estrogen Antagonist

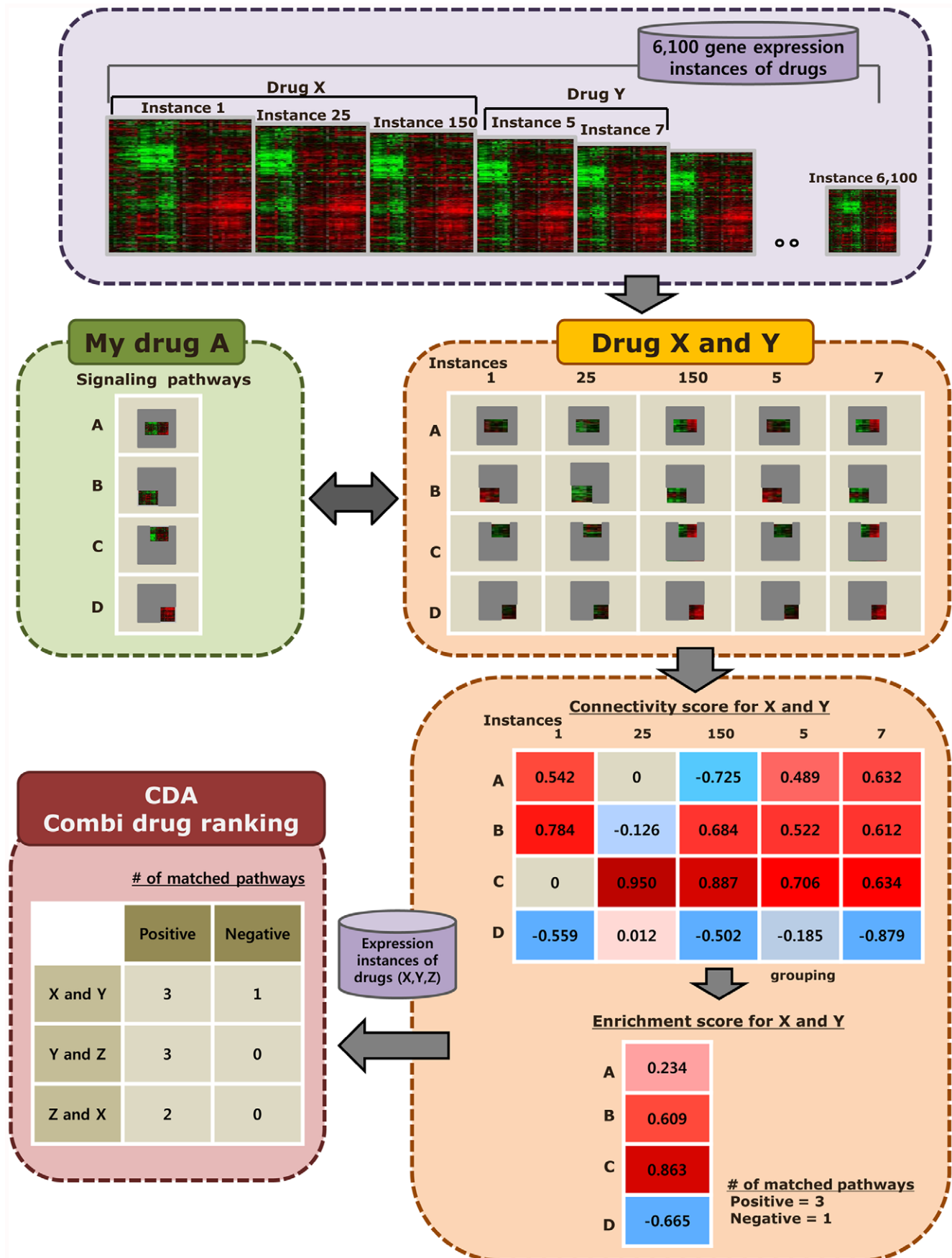
Elevated blood levels of estrogen is associated with an increased risk of breast cancer [16]. Gene expression signatures in breast cancer cells treated with Letrozole (fifty eight untreated tumors and fifty eight letrozole-treated tumors, GDS3116) were used to search the molecules function as estrogen antagonist [17]. Letrozole inhibits, aromatase, an enzyme that participates in estrogen biosynthesis. By inhibiting estrogen synthesis, letrozole slows the proliferations of breast cancer cells. Table 1 shows that cells treated with fulvestrant share a very similar expression pattern to those treated with letrozole. Fulvestrant is an estrogen receptor antagonist with no agonist effects. Fulvestrant not only down-regulates transcriptional activities of estrogen receptor but also induce its degradation. Fulvestrant was approved by the FDA for the treatment of postmenopausal women with hormone receptor-positive metastatic breast cancer [18]. The gene expression signatures of cells treated with fulvestrant in 6 different signaling pathways resembled those of letrozole. Not surprisingly, they show similar patterns of gene expression on the plasma membrane estrogen receptor signaling pathway as well as on LPA receptor mediated events pathway and stabilization, expansion of the E-cadherin adherens junction pathway, and Reelin signaling pathway.

These results are illuminating in light of the connections in the literature which show these pathways are regulated by estrogen and/or involved in cancer progression. E-cadherin is a cell-cell adhesion protein, and has been shown to play a crucial role in tumor suppression [19]. A recent study by Oesterrich et al. showed that estrogen caused down-regulation of E-cadherin levels in breast cancer cells [20]. Lysophosphatidic acid (LPA; 1-acyl-glycerol 3-phosphate), which is also regulated by estrogen [21,22] is one of the simplest natural phospholipids that mediates multiple processes including neurogenesis, angiogenesis, wound healing, and cancer progression [23,24]. Reelin is a secreted signaling protein associated with regulation of neuronal cell positioning and migration. Its down-regulation is associated with increased migratory ability and reduced survival in breast cancer [25]. The relationship between reelin and estrogen/breast cancer is not fully understood.

Letrozole inhibits estrogen synthesis, whereas fulvestrant blocks the estrogen receptor. Although the mechanisms of those two compounds are different, the signaling cascades they affect would be expected to be similar in their down-regulation of transcriptional activity in downstream pathways. As levels of estrogen are decreased after the treatment of letrozole, signaling pathways related to E-cadherin and LPA are affected, and this perturbation in these pathways are also observed in cells treated with fulvestrant. They both regulate reelin signaling pathway to induce apoptosis in cancer cells through as yet unknown mechanisms.

### Case Two: Combinatorial Drugs that Induce Apoptosis on Tumorigenic Lung Cancer Cells

This case derived from a study by Landi et al. that investigated the role of cigarette smoking in lung adenocarcinoma development and survival (forty nine normal lung tissues and fifty eight lung



**Figure 1. Analysis pipeline of CDA.** Combinatorial drug analysis process. In drug set pattern analysis step (the bottom right box), combinatorial drug analysis process treats profiles of two different molecules as a group to measure the synergistic effects of them.  
doi:10.1371/journal.pone.0042573.g001

tumor tissues, GDS3257). In our analysis, we disregarded information on smoking, disease state, and gender of the patients. In order to identify molecules that could reverse the expression pattern of lung adenocarcinoma cells, we looked for a phenotype where expression of signature genes was reversed: up-regulated genes became down-regulated, and vice versa. Signaling pathway gene set enrichment analysis of the “reversed phenotype” genes in the lung adenocarcinoma cells showed highlighted that many of the genes identified in this way are frequently associated with tumor cell growth and proliferation (Table 2). Based on this gene expression analysis, we identified, among the top 15 combinatorial drug pair candidates, two synergistic combinatorial drug pairs: alsterpaullone and scriptaid; and irinotecan and semustine. Alsterpaullone is a cyclin-dependent kinase (CDK) inhibitor that induces apoptosis [26]. Scriptaid is a class of histone deacetylase inhibitors (HDACis). HDACis are involved in cell growth, apoptosis and differentiation. Scriptaid also induces cell death in cancer cells [27,28]. Irinotecan is an anticancer drug that binds to the DNA topoisomerase 1 complex during DNA replication, preventing the resealing of single-strand breaks [29]. Semustine also known as methyl-CCNU, is another anti-cancer drug in the class of alkylating agents [30,31]. The alsterpaullone-scriptaid and irinotecan-semustine pairs showed meaningful, statistically significant expression pattern matching in seven, six lung adenocarcinoma-related pathways, respectively. Simultaneous and continuous exposure of A549 cells to different concentration of these two combinatorial drug pairs for 72 hours showed a synergism (Combination index (CI) <1 and Dose reduction index (DRI) >1; Table 3 and 4, Figure 2).

### Case Three: Combinatorial Drugs that Induce Apoptosis on Triple-negative Breast Cancer Cells

Breast cancer is the most common form of cancer in women. Human epidermal growth factor receptor 2 (HER2), also known as receptor tyrosine-protein kinase ERBB2, belongs to the epidermal growth factor receptor (EGFR) family, and it is one of the most important oncogenes in invasive breast cancer. Based on the importance of HER2 amplification on breast cancer, the HER2-targeting monoclonal antibody trastuzumab was developed [32]. Additionally, aberrant EGFR signaling is a major characteristic of a human cancer including breast cancer. Several anti-EGFR agents are currently undergoing clinical testing in breast cancer patients clinically [33]. However, triple negative breast cancer (TNBC) is a type of breast cancers that does not express the genes for estrogen receptor (ER), progesterone receptor (PR) or human epidermal growth factor receptor 2 (HER2). For that reason, novel effective therapeutic agents are needed for TNBC patients [34]. Combined treatment of general breast cancer cells with drugs that target EGFR and HER2 results in a synergistic antitumor effect [35,36]. That means that targeting EGFR family signaling pathway is a good strategy for breast cancer treatment.

To discover a synergistic combinatorial drug pair for TNBC patients, we focused on FDA approved drugs. We obtained gene expression signatures from TNBC cell lines (five normal breast cancer cell lines and five triple-negative breast cancer cell lines, GSE6569), and we selected halofantrine - vinblastine pair as a candidate pair (Figure 3). The CDA analysis indicated that the pair has opposite expression patterns compared with TNBC signatures in five different signaling pathways, including four of the EGFR family signaling pathways and one integrin pathway

(Figure 4). Aberrant activation of the EGFR family is implicated in a number of cancers and it is already the target of several antineoplastic agents [37]. A6b1- and a6b4- mediated integrin signaling is involved in apoptosis, tumour cell invasions, and cell migration.

Halofantrine is an anti-malarial agent with an unknown mode of action. Although it has cardiotoxic potential, it is safe when carefully administered [38]. Vinblastine is a microtubule-targeted anticancer drug that induces mitotic block and apoptosis by suppressing microtubule dynamics at lower concentration, and reducing microtubule polymer mass at higher concentration [39]. As shown in Figure 4B, halofantrine and vinblastine are indirectly related to EGFR family signaling pathways. Furthermore, both are also related to an integrin signaling pathway. Based on this information, we hypothesized that halofantrine and vinblastine are synergistic because they simultaneously affect the EGFR and integrin signaling pathways. Furthermore, sensitivity of HER2-positive breast cancer cells resistant to anti-HER2 therapies are related to antiapoptotic proteins MCL1 and Survivin [40]. And these two proteins commonly have protein-protein interactions with CASP3, a vinblastine-related protein [41,42]. Based on this, we hypothesized that vinblastin could be a good TNBC drug candidate. Using the steps described for all three cases, CDA users will be able to put forward testable hypotheses by combining signaling pathway expression information with known drug-protein-disease information from phExplorer.

### Discussion

Since the number of new drug has not kept pace with the enormous increase in pharma R&D spending, drug discovery researchers have become more creative in finding new uses for existing drugs [43]. Analyzing large data sets such as gene expression [15], chemical similarity [44], side-effect similarity [45], disease-drug network [46], and phenotypic disease network [47] has been applied for drug repositioning. Exploration of drug off-targets using chemical-protein interactome can also provide alternative strategy [48]. However drugs with single targets frequently show limited efficacies and drug resistance at the some point. To overcome these problems, systems-oriented drug design is now moving to multicomponent therapies and multi-targeted drugs, based on the idea that targeting drugs to act on multiple signaling pathways will maximize therapeutic efficacy [49]. With this in mind, we have designed a system for multiple signaling pathways targeting combinatorial drug discovery using gene expression profile. There are three groups of pharmacodynamically synergistic combinations; 1) anti-counteractive action group 2) complementary action group 3) facilitating action group. There are a variety of mechanism of actions represented by these combinations, arising from drug interactions with the same or different targets of the same or different pathways, and from modulations of crosstalk pathways and network robustness [8].

The robustness of CDA does not depend heavily on the particular bioinformatics method employed for signature extraction, thus providing a flexible analysis platform that can be adopted by a variety of users with different software tools for handling gene expression analysis. Although genome-wide expression analysis has become a routine tool in genomic research, extracting biologically meaningful information remains a major challenge. Statistically significant genes can be obtained by

**Table 1.** Top 10 molecules showing similar expression patterns of transcriptional response modules to letrozole.

|                   | FOXO1 transcription factor network | IGF1 pathway | IL4-mediated signaling events | LPA receptor mediated events | Plasma membrane estrogen receptor signaling | Reelin signaling pathway | Stabilization and expansion of the E-cadherin adherens junction |
|-------------------|------------------------------------|--------------|-------------------------------|------------------------------|---------------------------------------------|--------------------------|-----------------------------------------------------------------|
| Fulvestrant       |                                    | ○            | ○                             | ○                            | ○                                           | ○                        | ○                                                               |
| Trichostatin A    | ○                                  | ○            | ○                             |                              | ○                                           | ○                        |                                                                 |
| Irinotecan        | ○                                  | ○            |                               | ○                            | ○                                           |                          | ○                                                               |
| Ag-013608         |                                    |              |                               |                              | ○                                           | ○                        | ○                                                               |
| Tretinoin         | ○                                  | ○            |                               |                              | ○                                           |                          |                                                                 |
| Metamizole sodium |                                    | ○            |                               |                              | ○                                           |                          | ○                                                               |
| Tanespimycin      |                                    | ○            | ○                             | ○                            |                                             |                          |                                                                 |
| Vorinostat        | ○                                  | ○            |                               |                              | ○                                           |                          |                                                                 |
| Verteporfin       |                                    | ○            |                               |                              | ○                                           |                          |                                                                 |
| Daunorubicin      |                                    | ○            |                               |                              |                                             |                          | ○                                                               |

doi:10.1371/journal.pone.0042573.t001

number of different ways. Moreover, there is no standard rule to restrict the number of genes. Thus, significant gene selection is quite depending on individual researchers. Given this multiplicity of approaches, significant gene lists can be quite diverse according to extraction algorithms and research principles. This lack of standardized bioinformatics approaches brings with it a risk of insufficient information usage that can lead to inaccuracies in the final interpretation. To offset these differences, for expression analysis and interpretation, our strategy employs functionally important genes as data sets, rather than entire statistically selected gene sets. This approach was validated by an *in silico* case (Information S1). CDA provides a mechanism whereby hundreds of input signature genes will be split into signaling pathways at the first step, therefore users don't need to themselves extract a small group of significant gene sets using number of different algorithms. Through this process, CDA successfully has identified a number of molecules having similar function (Table 1). In this study, we presented case studies whereby CDA successfully predicted synergistic combinatorial drug pairs in lung cancer and triple negative breast cancer. Together with phExplorer, CDA also provides functional insights of combinatorial drugs.

Using CDA, the number of matched pathways decides the ranking of drug candidates, however, the type of matched pathways must be considered carefully. As the interpretation of result and the final decision must be made by researchers, we tried not to restrict their choice by providing strictly ordered list based on our limited pre-knowledge.

## Materials and Methods

### Data Source

Reference molecule-treated expression data was downloaded from Connectivity Map (build 02) (<http://www.broadinstitute.org/cmap/>). It contains 6,100 expression profiles representing 1,309 molecules. Molecules were selectively applied to five different human cancer cell lines for short duration. Each molecule-treated expression profile was paired with a control, and each profile was represented by a non-parametric rank-ordered list of all probe sets.

Pathway gene set data was downloaded from Pathway Interaction Database (PID) on 09/03/2010 (<http://pid.nci.nih.gov/>). Only the NCI-Nature Curated data was used. Pathway gene set information was extracted, consisting of 166 pathways comprising 2,297 genes. These genes were annotated to Affymetrix GeneChip Human Genome U133 Array Set HG-U133A probe set. The final form of pathway data consists of 166 signaling pathways and 3,726 probe sets.

Furthermore, nine public databases, EntrezGene interaction [50], MINT [51], DIP [52], CTD [53], TTD [54], ChemBank [55], PharmGKB [56], OMIM (<http://www.ncbi.nlm.nih.gov/omim/>), and GAD [57] were integrated to visualise enrich drug-protein-disease network map. For data integration in a unified format, we adopted PubChem CID for drugs, GeneID for proteins, and MeSH descriptor for diseases. The integrated database is called PharmDB, and it is available at <http://pharmdb.org/>.

### Input Signatures

Three different GDS/GSE data files were downloaded for each case study. All of them were used Affymetrix Human Genome U133A Array.

Case 1: GDS3116 - Letrozole effect on breast cancer.

Fifty eight untreated tumors vs. fifty eight letrozole-treated tumors

**Table 2.** Enriched pathway in lung adenocarcinoma.

| Pathway                            | Pathway Category                                                                                              |
|------------------------------------|---------------------------------------------------------------------------------------------------------------|
| amb2 Integrin signaling            | Integrin mediated cell-cell signaling pathways Integrin mediated cell-extracellular matrix signaling pathways |
| Aurora A signaling                 | Cell cycle pathways, mitotic                                                                                  |
| Aurora B signaling                 | Cell cycle pathways, mitotic                                                                                  |
| BMP receptor signaling             | Bone morphogenetic proteins signaling pathway                                                                 |
| Direct p53 effectors               | p53 signaling pathway                                                                                         |
| E2F transcription factor network   | Transcription factor mediated signaling pathways Cell cycle pathways, mitotic Transcription pathways          |
| Endothelins                        | Endothelin signaling pathway                                                                                  |
| FGF signaling pathway              | Fibroblast growth factor signaling pathway                                                                    |
| FOXO1 transcription factor network | Forkhead signaling pathways                                                                                   |

doi:10.1371/journal.pone.0042573.t002

Case 2: GDS3257 - Lung adenocarcinoma.

Forty nine normal lung tissues vs. fifty eight lung tumor tissues

Case 3: GSE6569 - Triple-negative breast cancer cell lines.

Five normal breast cancer cell lines: BT474, SKBR3, HCC-1419, HCC-1954, MCF7

Triple-negative breast cancer cell lines: BT20, BT549, HCC-1806, MDA-MB-231, MDA-MB-468

The expression data were normalized using RMA from the BioConductor Affy package. Then these data were analyzed using a method called empirical Bayes in limma. To extract statistically differentially expressed genes, 2-fold change and p-value <0.05 were set as default. The signatures were represented by two probe sets, “up probe sets” and “down probe sets”. With given input signatures, hyper geometric tests were performed for signaling pathway gene set enrichment analysis. Signaling pathways with p-value <0.01 were selected as it was believed that input signature genes were enriched in these pathways.

### Enrichment Analysis

Signaling pathway expression pattern analysis and drug set pattern analysis were performed based on the Kolmogorov-Smirnov statistics. To determine whether the distribution of input gene sets/or drug sets was significant, 10,000 times permutations were carried out by generating random ranking matrices. The sets with p-value <0.01 were indicated as enriched.

### Signaling Pathway Expression Pattern Analysis

6,100 molecule-treated expression profiles were rank ordered using gene set enrichment analysis for each selected pathway. As mentioned above, there were two types of input set, “up probe sets” and “down probe sets”. The expression pattern similarity is calculated for both sets. The procedure is as follows:

**Table 3.** CI values for the drug combinations at 25%, 50%, 75% levels of inhibition of A549 cell proliferation.

| CI Values                  | 25%   | 50%   | 75%   |
|----------------------------|-------|-------|-------|
| Alsterpaullone + Scriptaid | 0.887 | 0.647 | 0.483 |
| Irinotecan + Semustine     | 0.816 | 0.718 | 0.636 |

doi:10.1371/journal.pone.0042573.t003

- 1) Calculate Kolmogorov-Smirnov score for both “up probe sets” and “down probe sets”

e = an expression profile

KS<sup>e</sup> = KS (Kolmogorov – Smirnov) score for the “up probe sets” or the “down probe sets”

n = the total number of probe sets (22,283)

t = the number of probe sets in either the “up probe sets” or the “down probe sets”

j = the position of a probe set in the ordered input signature probe set lists

V(j) = the position of the jth probe set in the ordered list of all probe sets.

$$a = \max_{j=1}^t \left[ \frac{j}{t} - \frac{V(j)}{n} \right]$$

$$b = \max_{j=1}^t \left[ \frac{V(j)}{n} - \frac{(j-1)}{t} \right]$$

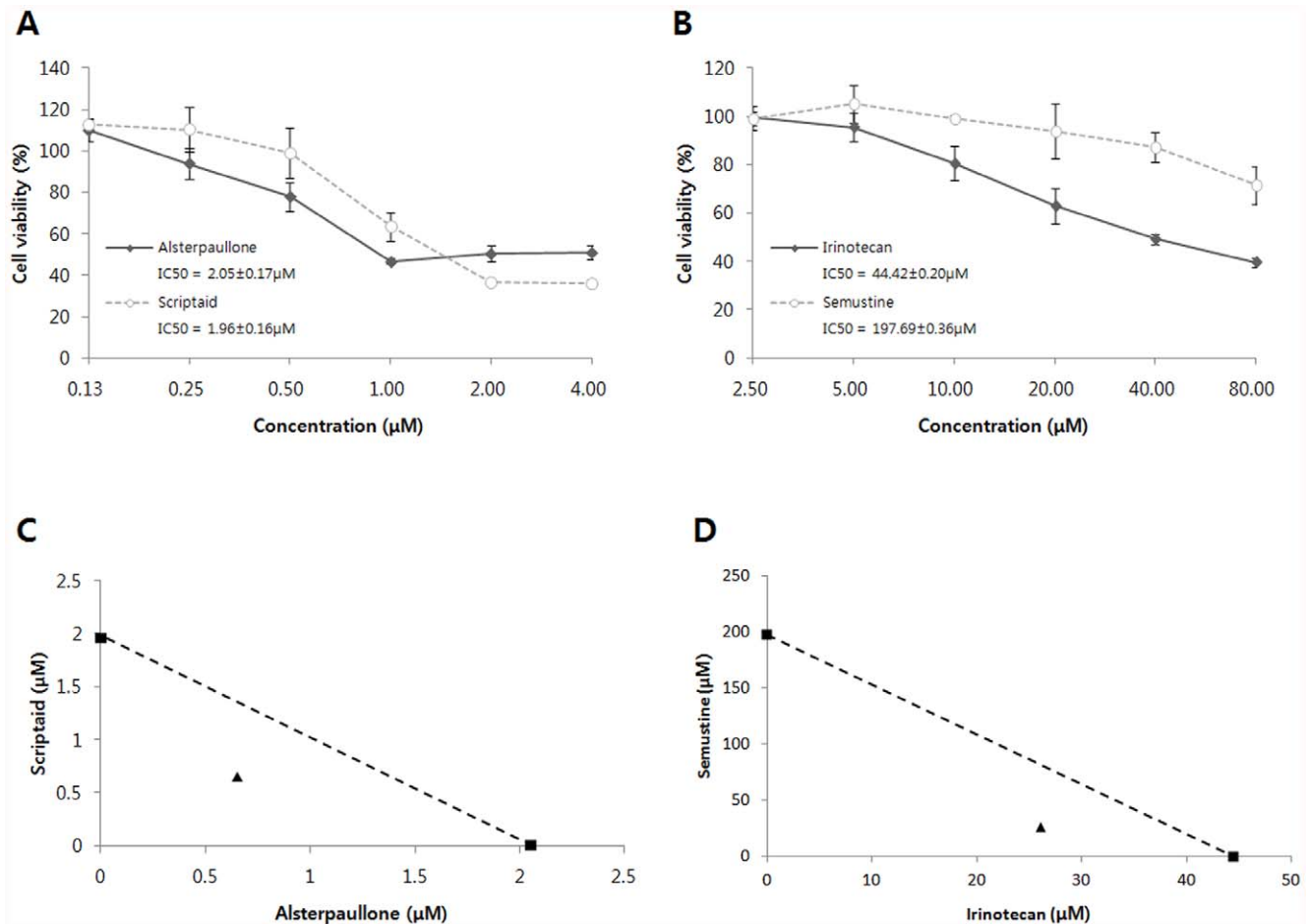
$$KS^e = \begin{cases} a & (\text{if } a > b) \\ -b & (\text{if } b > a) \end{cases}$$

**Table 4.** DRI values for the drug combinations at 25%, 50%, 75% levels of inhibition of A549 cell proliferation.

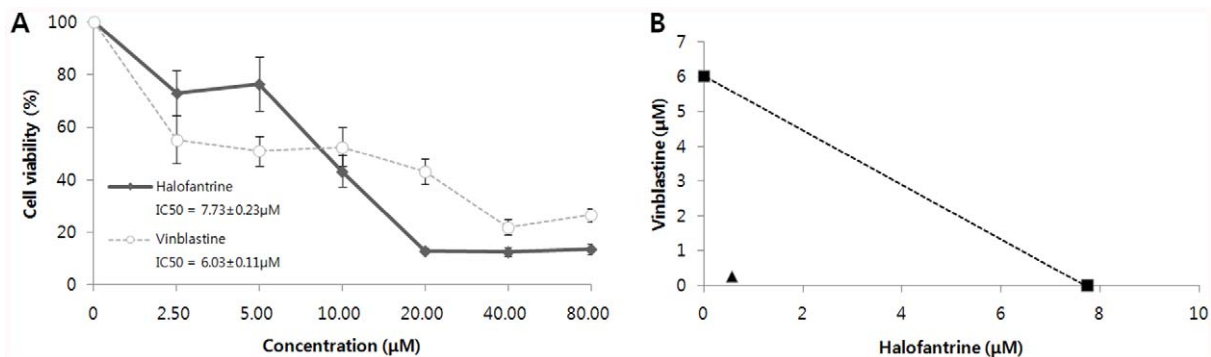
| DRI Values                 | 25%   | 50%   | 75%   |
|----------------------------|-------|-------|-------|
| Alsterpaullone + Scriptaid |       |       |       |
| Alsterpaullone             | 2.013 | 3.162 | 4.968 |
| Scriptaid                  | 2.565 | 3.020 | 3.554 |
| Irinotecan + Semustine     |       |       |       |
| Irinotecan                 | 1.452 | 1.705 | 2.002 |
| Semustine                  | 7.841 | 7.589 | 7.345 |

doi:10.1371/journal.pone.0042573.t004



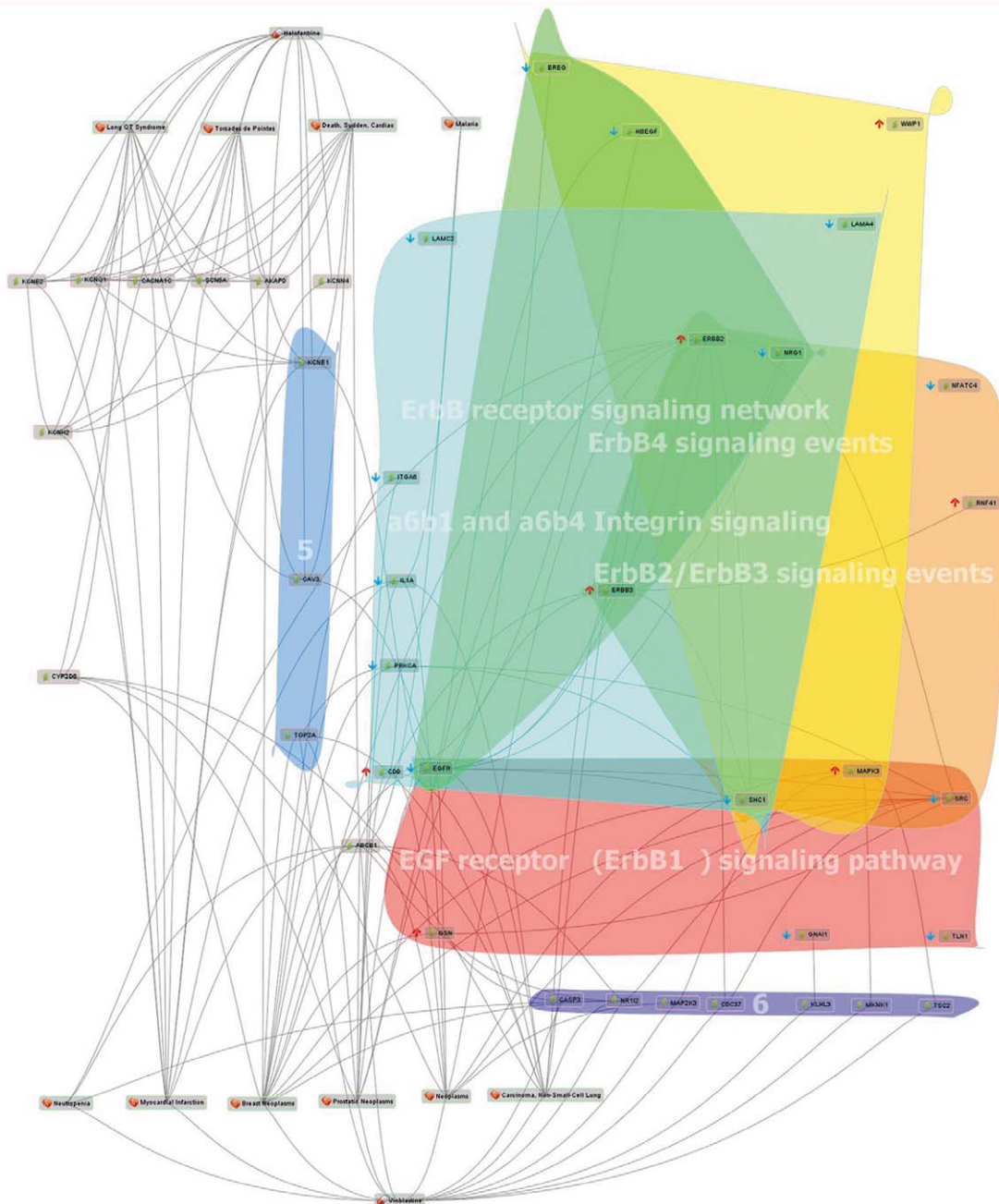


**Figure 2. Synergistic combinatorial drug pairs on lung cancer cells.** (A, B) Effects of alsterpaullone, scriptaid, irinotecan, and semustine on A549 cancer cell proliferation.  $\text{IC}_{50}$  indicates the concentration of drug that induce 50% of inhibition of cell proliferation. Error bars represent the standard deviation of six experiments. (C, D) Drug pairs were treated in 1:1 molar ratio. The  $\text{IC}_{50}$  values of each drug are plotted on the axes, and the dashed line represents additive effect. Triangle point represents the concentrations of the combinations resulting in 50% of proliferation inhibition. As the triangle points are positioned on the left of the dashed line, these combinatorial drug pairs are synergistic. The  $\text{IC}_{50}$  values of each drug in alsterpaullone-scriptaid and irinotecan-semustine combinations are  $0.65 \mu\text{M}$  and  $26.05 \mu\text{M}$ , respectively.  
doi:10.1371/journal.pone.0042573.g002

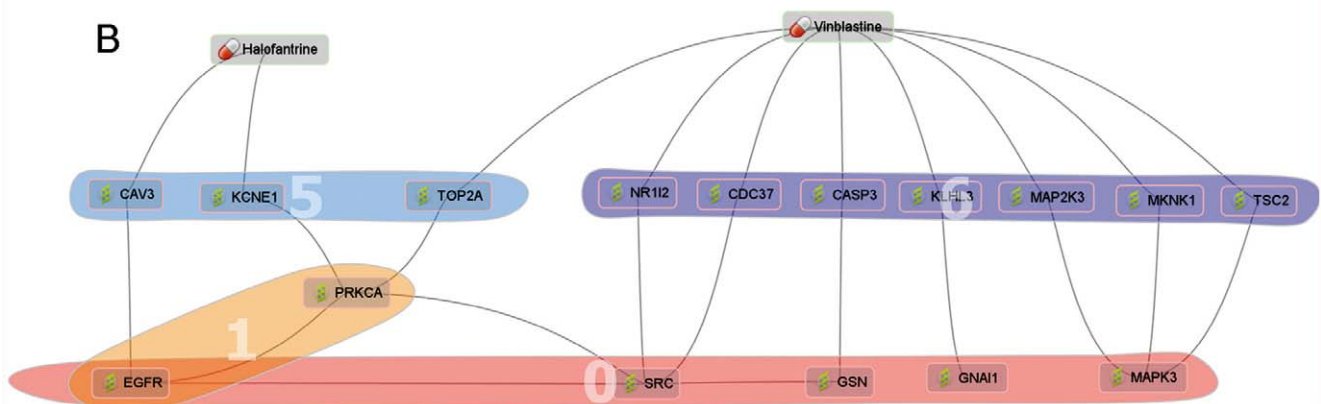


**Figure 3. In vitro validation of halofantrine and vinblastine alone and in combination in a triple-negative breast cancer cell line.** (A) Effects of halofantrine and vinblastine on MDA-MB-231 TNBC cell proliferation.  $\text{IC}_{50}$  indicates the concentration of drug that induce 50% of inhibition of cell proliferation. (B) Halofantrine and vinblastine combination was treated in 2:1 molar ratio. Halofantrine and vinblastine combination shows a strong synergistic effect. The  $\text{IC}_{50}$  values of each drug in halofantrine-vinblastine combinations are  $0.55 \mu\text{M}$  and  $0.27 \mu\text{M}$ , respectively. The combination shows a strong synergistic effect (CI value is 0.12, and DRI values for halofantrine and vinblastine are 14.17 and 22.09, respectively).  
doi:10.1371/journal.pone.0042573.g003

A



B





**Figure 4. Network map of halofantrine and vinblastine on triple-negative breast cancer using phExplorer.** (A) It seems that halofantrine and vinblastine could affect on five different signaling pathways in TNBC. Group 5: Halofantrine- or vinblastine-related proteins which are also related with proteins of A6B1 and A6B4 Integrin signaling pathway. Group 6: Proteins which are related with vinblastine as well as proteins of EGFR family signaling pathways (such as ERBB1 signaling pathway, ERBB2/ERBB3 signaling events, ERBB4 signaling events, ERBB receptor signaling network). (B) We hypothesized that halofantrine and vinblastine are synergistic because they complementary regulate integrin and EGFR signaling pathways. Group 0: A part of EGFR family signaling pathways. Group 1: A part of A6B1 and A6B4 Integrin signaling pathway. doi:10.1371/journal.pone.0042573.g004

- 2) Calculate the Enrichment Score (ES) for each profile

$$ES^e = 0 \text{ (if } KS_{up} \text{ and } KS_{down} \text{ have the same algebraic sign)}$$

Otherwise, across all profiles,

$$s^e = KS_{up} - KS_{down}$$

$$p = \text{Max}(s^e)$$

$$q = \text{Mix}(s^e)$$

The ES for these profiles are:

$$ES^e = \begin{cases} \frac{s^e}{p} & (\text{if } s^e > 0) \\ -\left(\frac{s^e}{q}\right) & (\text{if } s^e < 0) \end{cases}$$

- 1) Rank the profiles in descending order of  $ES^e$

### Drug Set Pattern Analysis

Molecules were applied to different cell lines with various doses, and the ES of each molecule was calculated using the distribution of the molecule-treated profiles, using the same method as used in calculating the KS score in signaling pathway expression pattern comparison. For the case of combinatorial drug analysis, signatures of two different molecules were treated as a group. The rationale is as follows: we assume two molecules, “A” and “B” show highly similar expression pattern with the expression of signaling pathway “SP1” and “SP2”, respectively. The purpose of combinatorial drug is matching up two molecules which are synergistic or complementary. “A” and “B” are highly related with different pathways, and thus might affect to each other in unanticipated ways. For that reason, profiles of “A” and “B” are grouped as a set, then the ES (Enrichment Score) of “A and B” combination is calculated in two signaling pathways independently. So the similarity of expression pattern of “B” is now considered not only in “SP2” but also in “SP1” as a combinatorial drug partner. If “B” shows high ESs in both pathways, “B” could be a complementary partner for “A” as it covers “SP2” which “A” might not be able to regulate, and at the same time, synergistic effect could be expected in “SP1” as both of them are highly enriched in there.

Using these steps, the KS score was computed using these profiles. Then, random permutation tests (10,000 times) were carried out to estimate the significance of a distribution of those profiles. The molecules with p-value <0.01 were assumed as significant.

### Drug Ranking

At this point, we have listed single/combinatorial drugs for each disease-associated signaling pathway in our database. The goal of creating this system is to provide a means of selecting single/combinatorial drugs that can regulate disease-related signaling pathways to the greatest potential. To this end, for each drug, the number of pathways scored greater than the positive threshold was counted. The positive threshold for single drug and combinatorial drug were 0 and 0.5, respectively. The drugs were ranked in descending order of the number of pathways they appeared in. Pathways that scored less than the negative threshold were also listed. The negative threshold for single drug and combinatorial drug were 0 and -0.5, respectively. These negatively correlated pathways can be treated as negative effects.

### Cell Culture and Materials

A549 and MDA-MB-231 were purchased from American Type Culture Collection. RPMI containing 10% fetal bovine serum and 1% antibiotics were used for cell cultivation. Alsterpaullone, Scriptaid, Irinotecan hydrochloride, Semustine, Halofantrine hydrochloride, Vinblastine sulfate salt were purchased from Sigma.

### MTT Assay

A549 or MDA-MB-231 cells were seeded in the 96-well plates. After 24 h, cells were treated with indicated chemicals. After incubation for 3 days, MTT reagent (5 mg/ml) (Sigma) was added to each well, and the plate was placed at 37°C for 2 h. After aspirating the supernatant, 200 µl of dimethyl sulfoxide (Sigma) was added to each well. Colored formazan product was assayed spectrophotometrically at 570 nm using ELISA plate reader.

### Combination Index (CI) and Dose Reduction Index (DRI) Calculations

Synergism and antagonism for combinatorial drug were quantified by the combination index (CI), where  $CI < 1$ ,  $CI = 0$ ,  $CI > 0$  indicate synergism, additive, and antagonism, respectively. CI was determined by the following equation:

$$CI_{A+B} = \frac{D_{A/A+B}}{D_A} + \frac{D_{B/A+B}}{D_B}$$

$D_A$  is the concentration of drug A that induce the inhibition of cell growth.  $D_{A/A+B}$  is the concentration of drug A in the combination A+B giving the same inhibition effect. The dose reduction index (DRI) is a measure of how much the dose of each drug may be reduced in a combination for a given degree of effect compared with the concentration of each drug alone.

$$DRI_A = \frac{D_A}{D_{A/A+B}} \text{ and } DRI_B = \frac{D_B}{D_{B/A+B}}$$

CI and DRI indexes were calculated with the CalcuSyn version 2.1 software (Biosoft, Cambridge, UK).

## Supporting Information

**Information S1 A case study on acute lymphoblastic leukemia (ALL) cells.** See the ranking of rapamycin in glucocorticoid resistance ALL cells. It proved that CDA does not heavily depend on the way of the signature extraction. (DOC)

## References

- Kubinyi H (2003) Drug research: myths, hype and reality. *Nat Rev Drug Discov* 2: 665–668.
- Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, et al. (2006) Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med* 355: 2408–2417.
- Stelling J, Sauer U, Szallasi Z, Doyle FJ 3rd, Doyle J (2004) Robustness of cellular functions. *Cell* 118: 675–685.
- Kitano H (2007) A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* 6: 202–210.
- Gupta EK, Ito MK (2002) Lovastatin and extended-release niacin combination product: the first drug combination for the management of hyperlipidemia. *Heart Dis* 4: 124–137.
- Larder BA, Kemp SD, Harrigan PR (1995) Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science* 269: 696–699.
- Nelson HS (2001) Advair: combination treatment with fluticasone propionate/salmeterol in the treatment of asthma. *J Allergy Clin Immunol* 107: 398–416.
- Jia J, Zhu F, Ma X, Cao Z, Li Y, et al. (2009) Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 8: 111–128.
- Hahn CK, Ross KN, Warrington IM, Mazitschek R, Kanegai CM, et al. (2008) Expression-based screening identifies the combination of histone deacetylase inhibitors and retinoids for neuroblastoma differentiation. *Proc Natl Acad Sci U S A* 105: 9751–9756.
- Nelander S, Wang W, Nilsson B, She QB, Pratilas C, et al. (2008) Models from experiments: combinatorial drug perturbations of cancer cells. *Mol Syst Biol* 4: 216.
- Chatterjee MS, Purvis JE, Brass LF, Diamond SL (2010) Pairwise agonist scanning predicts cellular signaling responses to combinatorial stimuli. *Nat Biotechnol* 28: 727–732.
- Zhao XM, Iskar M, Zeller G, Kuhn M, van Noort V, et al. (2011) Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Comput Biol* 7: e1002323.
- Wu Z, Zhao XM, Chen L (2010) A systems biology approach to identify effective cocktail drugs. *BMC Syst Biol* 4 Suppl 2: S7.
- Geva-Zatorsky N, Dekel E, Cohen AA, Danon T, Cohen L, et al. (2010) Protein dynamics in drug combinations: a linear superposition of individual-drug responses. *Cell* 140: 643–651.
- Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313: 1929–1935.
- Clemons M, Goss P (2001) Estrogen and the risk of breast cancer. *N Engl J Med* 344: 276–285.
- Miller WR, Larionov AA, Renshaw L, Anderson TJ, White S, et al. (2007) Changes in breast cancer transcriptional profiles after treatment with the aromatase inhibitor, letrozole. *Pharmacogenet Genomics* 17: 813–826.
- Croxtall JD, McKeage K (2011) Fulvestrant: a review of its use in the management of hormone receptor-positive metastatic breast cancer in postmenopausal women. *Drugs* 71: 363–380.
- Berx G, Van Roy F (2001) The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. *Breast Cancer Res* 3: 289–293.
- Oesterreich S, Deng W, Jiang S, Cui X, Ivanova M, et al. (2003) Estrogen-mediated down-regulation of E-cadherin in breast cancer cells. *Cancer Res* 63: 5203–5208.
- Hama K, Aoki J, Bandoh K, Inoue A, Endo T, et al. (2006) Lysophosphatidic receptor, LPA3, is positively and negatively regulated by progesterone and estrogen in the mouse uterus. *Life Sci* 79: 1736–1740.
- Gonzalez-Arenas A, Avendano-Vazquez SE, Cabrera-Wrooman A, Tapia-Carrillo D, Larrea F, et al. (2008) Regulation of LPA receptor function by estrogens. *Biochim Biophys Acta* 1783: 253–262.
- Contos JJ, Ishii I, Chun J (2000) Lysophosphatidic acid receptors. *Mol Pharmacol* 58: 1188–1196.
- Moolenaar WH (1999) Bioactive lysophospholipids and their G protein-coupled receptors. *Exp Cell Res* 253: 230–238.
- Stein T, Cosimo E, Yu X, Smith PR, Simon R, et al. (2010) Loss of reelin expression in breast cancer is epigenetically controlled and associated with poor prognosis. *Am J Pathol* 177: 2323–2333.
- Lahusen T, De Siervi A, Kunick C, Senderowicz AM (2003) Alsterpaullone, a novel cyclin-dependent kinase inhibitor, induces apoptosis by activation of caspase-9 due to perturbation in mitochondrial membrane potential. *Mol Carcinog* 36: 183–194.
- Lee EJ, Lee BB, Kim SJ, Park YD, Park J, et al. (2008) Histone deacetylase inhibitor scriptaid induces cell cycle arrest and epigenetic change in colon cancer cells. *Int J Oncol* 33: 767–776.
- Brazelle W, Krehling JM, Gemmer J, Ma Y, Cress WD, et al. (2010) Histone deacetylase inhibitors downregulate checkpoint kinase 1 expression to induce cell death in non-small cell lung cancer cells. *PLoS One* 5: e14335.
- Marsh S, Hoskins JM (2010) Irinotecan pharmacogenomics. *Pharmacogenomics* 11: 1003–1010.
- Guo Y, Lu JJ, Ma X, Wang B, Hong X, et al. (2008) Combined chemoradiation for the management of nasal natural killer (NK)/T-cell lymphoma: elucidating the significance of systemic chemotherapy. *Oral Oncol* 44: 23–30.
- Zhao Z, Liu Y, He H, Chen X, Chen J, et al. (2011) Candidate genes influencing sensitivity and resistance of human glioblastoma to Semustine. *Brain Res Bull* 86: 189–194.
- Bange J, Zwick E, Ullrich A (2001) Molecular targets for breast cancer therapy and prevention. *Nat Med* 7: 548–552.
- Lo HW, Hsu SC, Hung MC (2006) EGFR signaling pathway in breast cancers: from traditional signal transduction to direct nuclear translocation. *Breast Cancer Res Treat* 95: 211–218.
- Gluz O, Liedtke C, Gottschalk N, Pusztai L, Nitz U, et al. (2009) Triple-negative breast cancer—current status and future directions. *Ann Oncol* 20: 1913–1927.
- Normanno N, Campiglio M, De LA, Somenzi G, Maiello M, et al. (2002) Cooperative inhibitory effect of ZD1839 (Iressa) in combination with trastuzumab (Herceptin) on human breast cancer cell growth. *Ann Oncol* 13: 65–72.
- Moulder SL, Yakes FM, Muthuswamy SK, Bianco R, Simpson JF, et al. (2001) Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells in vitro and in vivo. *Cancer Res* 61: 8887–8895.
- Zhang H, Berezov A, Wang Q, Zhang G, Drebin J, et al. (2007) ErbB receptors: from oncogenes to targeted cancer therapies. *J Clin Invest* 117: 2051–2058.
- Bouchaud O, Imbert P, Touze JE, Dodoo AN, Danis M, et al. (2009) Fatal cardiotoxicity related to halofantrine: a review based on a worldwide safety data base. *Malar J* 8: 289.
- Jordan MA, Wilson L (2004) Microtubules as a target for anticancer drugs. *Nat Rev Cancer* 4: 253–265.
- Valabrega G, Capellero S, Cavalloni G, Zaccarello G, Petrelli A, et al. (2011) HER2-positive breast cancer cells resistant to trastuzumab and lapatinib lose reliance upon HER2 and are sensitive to the multitargeted kinase inhibitor sorafenib. *Breast Cancer Res Treat* 130: 29–40.
- Tamm I, Wang Y, Sausville E, Scudiero DA, Vigna N, et al. (1998) IAP-family protein survivin inhibits caspase activity and apoptosis induced by Fas (CD95), Bax, caspases, and anticancer drugs. *Cancer Res* 58: 5315–5320.
- Weng C, Li Y, Xu D, Shi Y, Tang H (2005) Specific cleavage of Mcl-1 by caspase-3 in tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)-induced apoptosis in Jurkat leukemia T cells. *J Biol Chem* 280: 10491–10500.
- Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 3: 673–683.
- Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, et al. (2009) Predicting new molecular targets for known drugs. *Nature* 462: 175–181.
- Campillos M, Kuhn M, Gavin AC, Jensen IJ, Bork P (2008) Drug target identification using side-effect similarity. *Science* 321: 263–266.
- Hu G, Agarwal P (2009) Human disease-drug network based on genomic expression profiles. *PLoS One* 4: e6536.
- Hidalgo CA, Blumm N, Barabasi AL, Christakis NA (2009) A dynamic network approach for the study of human phenotypes. *PLoS Comput Biol* 5: e1000353.
- Yang L, Wang K, Chen J, Jegga AG, Luo H, et al. (2011) Exploring off-targets and off-systems for adverse drug reactions via chemical-protein interactome-clozapine-induced agranulocytosis as a case study. *PLoS Comput Biol* 7: e1002016.
- Smalley KS, Haas NK, Brafford PA, Lioni M, Flaherty KT, et al. (2006) Multiple signaling pathways must be targeted to overcome drug resistance in cell lines derived from melanoma metastases. *Mol Cancer Ther* 5: 1136–1144.
- Maglott D, Ostell J, Pruitt KD, Tatusova T (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 39: D52–57.
- Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, et al. (2010) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res* 38: D532–539.
- Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, et al. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 32: D449–451.
- Davis AP, King BL, Mockus S, Murphy CG, Saraceni-Richards C, et al. (2011) The Comparative Toxicogenomics Database: update 2011. *Nucleic Acids Res* 39: D1067–1072.
- Zhu F, Han B, Kumar P, Liu X, Ma X, et al. (2010) Update of TTD: Therapeutic Target Database. *Nucleic Acids Res* 38: D787–791.

## Author Contributions

Conceived and designed the experiments: JHL TB KR. Performed the experiments: JHL TB DGK JTK JLL KMP. Analyzed the data: JHL. Contributed reagents/materials/analysis tools: JHL YJ TB JTK BCK. Wrote the paper: JHL SK.

55. Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, et al. (2008) ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res* 36: D351–359.
56. Thorn CF, Klein TE, Altman RB (2010) Pharmacogenomics and bioinformatics: PharmGKB. *Pharmacogenomics* 11: 501–505.
57. Becker KG, Barnes KC, Bright TJ, Wang SA (2004) The genetic association database. *Nat Genet* 36: 431–432.