Biochemical and Biophysical Research Communications 470 (2016) 137–143

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Association analysis of the perturbation of interactions in biological pathways and anticancer drug activity

Junehawk Lee [a, b], Doheon Lee [a, c, *]

[a] Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[b] Department of Convergence Technology Research, Korea Institute of Science and Technology Information, Daejeon, Republic of Korea
[c] Bio-Synergy Research Center, Daejeon, Republic of Korea

ARTICLE INFO

Article history:
Received 29 December 2015
Accepted 3 January 2016
Available online 6 January 2016

Keywords:
Drug response prediction
Signaling pathway
Domain–domain interaction
Precision medicine

ABSTRACT

Understanding how different genomic mutational landscapes in patients with cancer lead to different responses to anticancer drugs is an important challenge for realizing precision medicine for cancer. Many studies have analyzed the comprehensive anticancer drug-response profiles and genomic profiles of cancer cell lines to identify the relationship between the anticancer drug response and genomic alternations. However, few studies have focused on interpreting these profiles with a network perspective. In this work, we analyzed genomic alterations in cancer cell lines by considering which interactions in the signaling pathway were perturbed by mutations. With our interaction-centric approach, we identified novel interaction/drug response associations for two drugs (afatinib and ixabepilone) for which no gene-centric association could be found. When we compared the performance of classifiers for predicting the responses to 164 drugs, the classifiers trained with interaction-centric features outperformed the classifiers trained with gene-centric features, despite the smaller number of features (p-value $= 2.0 \times 10^{-3}$). By incorporating the interaction information from signaling pathways, we revealed associations between genomic alterations and drug responses that could be missed when using a gene-centric approach.

1. Introduction

Recent developments in massively parallel experiment techniques have enabled the production of large amounts of data at an unprecedented scale, facilitating studies of the cancer cell response to anticancer drugs [1–4]. These studies have aimed to characterize genomic variation in cancer cell lines and identify genomic variations correlated with the anticancer drug response to realize precision medicine. For example, in the Cancer Cell Line Encyclopedia (CCLE) [1], comprehensive biological data of gene expression, targeted sequencing, and copy numbers from multiple human cancer cell lines were used to identify novel drug response predictors (e.g., SLFN11 expression and irinotecan). The NCI-60 cell panel [2] has also been used to identify correlations between variants in genes and responses to anticancer agents (e.g., BRAF V600 variants and vemurafenib) by analyzing exome sequencing data.

Although these studies successfully identified relationships between the expression or variants of a gene and drug sensitivity by using gene/protein-centric approach, these studies could not distinguish among differences caused by different positions of variations in a gene, different functional impacts of variation, and different locations of a gene/protein within a biological pathway. Thus, substantial effects of genomic variation on drug sensitivity may not have been observed [5,6]. For example, Wang et al. [6] showed that mutations in different domains in one gene may cause different diseases by perturbing different interactions within the biological network. Moreover, the mutations in each interacting domain used by two interacting proteins were found to potentially cause the same disease.

Accordingly, in this study, we aimed to identify perturbed interactions in biological pathways by considering the position of variation in a gene and the resulting functional impact using structurally resolved domain–domain interaction information. Additionally, we explored the effects of perturbation of interactions on anticancer drug sensitivity by comparing cells with and without the perturbation. Our results provide important insights into individualized precision medicine approaches using interaction-

Abbreviations: DRA, Drug response-associated.

* Corresponding author. Bio-Synergy Research Center, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea.
E-mail address: dhlee@kaist.ac.kr (D. Lee).

http://dx.doi.org/10.1016/j.bbrc.2016.01.009
0006-291X/

based information.

2. Materials and methods

2.1. Drug response data

We downloaded raw drug response data for the NCI-60 cell panel from CellMiner and performed quality control as previously described [7]. We selected 196 drugs that are approved by the US Food and Drug Administration (FDA) or are currently being used in clinical test.

For CCLE drug-response profile data, we downloaded pharmacological profiling drug data for 24 anticancer drugs across 504 cell lines from the CCLE portal. We used the activity area (the area above the dose–response curve) for our study because this measurement provides a comprehensive representation of drug activity according to CCLE.

2.2. Cancer cell line mutational profiles

We downloaded NCI-60 exome sequencing data from Ingenuity and selected type II variants annotated by Abaan et al. [2] (a total of 115,260 variants). Hybrid capture sequencing data from CCLE, including mutational profiles of 1651 genes for 905 cancer cell lines, were downloaded from the CCLE portal. We used the dataset recommended by CCLE, with filtering of common or neutral variants.

We computed the impact of NCI-60 exome data and CCLE sequencing data using IntOGen [8], which annotates each variant's impact on proteins using SIFT [9], Polyphen2 [10], and MutationAssessor [11]. Only the effects of a variant on the canonical form of the protein were used.

2.3. Data for interaction annotations

The domain–domain interaction data from iPfam [12] and 3did [13] were used. iPfam provides structurally resolved 11,263 domain–domain interactions (release 1.0), and 3did provides 8328 domain–domain interactions. Data from Pfam [14] (release 27.0) were also downloaded to extract protein domain information.

For signaling pathway information, 286 pathways were retrieved from the KEGG database [15], including metabolic and nonmetabolic pathways (downloaded May 20, 2015 using KEGGgraph package in R). To filter the actual protein–protein interactions (PPIs) from the KEGG signaling pathway, PPI data (release 9) were downloaded from HPRD [16].

2.4. Annotating the perturbation of interactions in biological pathways

We extracted all interactions described in the KEGG pathway database (26,870 PPIs and 11,200 protein–compound–protein interactions [PCPIs]). The KEGG ID of each interaction was converted to the Uniprot accession number using the org.Hs.eg.db package in R. Among unique PPI interactions, we selected interactions present in the HPRD database to remove indirect interactions. We also included interactions in which chemical compounds were involved. As a result, 5739 PPIs and 20,405 PCPIs were prepared. To identify the interacting domains of each interacting protein, we adopted and expanded the method described by Wang et al. [6]. We annotated each interacting protein with the domains participating in the interaction if the interaction between the domains on both partners was described in iPfam or 3did (Fig. 1A).

To annotate whether each interaction was perturbed, we used the predicted functional impact information generated by IntOGen and examined whether the mutation resided in an interacting domain. An interaction was annotated as perturbed if any interacting domain of the interacting partners had more than one coding sequence-altering mutation predicted as damaging by SIFT, Polyphen2, or MutationAssessor or if a nonsense mutation was found in one or both of the interacting proteins (Fig. 1B). PCPIs were also included by incorporating domain-ligand interaction information from iPfam, and perturbation information was annotated in the same way using domains interacting with the compound.

After annotating the perturbation status of each interaction for 60 cancer cell lines with the NCI-60 dataset, we removed interactions that were perturbed in less than 5% of cell lines. Consequently, 254 PPIs and 506 PCPIs were used during the analysis. Next, for comparison with the traditional gene-centric approach, we generated the mutation profile of each gene. Each gene was annotated as mutated if a mutation was found in the gene regardless of its location. In total, 3135 genes that were mutated in more than 5% of cell lines were selected.

2.5. Identifying drug response-associated (DRA) interactions and genes

For each drug and each interaction/gene, we calculated p-values by performing t-tests between the z-transformed -log$GI50$ values from two groups of cell lines with and without perturbation. We also calculated the difference between the mean z-transformed -log$GI50$ value in each group. We defined an interaction or a gene as DRA if the -log$GI50$ values from the two groups of NCI-60 cell lines with and without perturbation of the interaction or gene were significantly different (Bonferroni-adjusted p-value < 0.05) and the difference in the mean -log$GI50$ between the two groups was more than 0.5 (Fig. 1A).

2.6. Validating DRA interactions with other cancer cell line datasets

To validate the DRA interactions from the NCI-60 dataset using the CCLE dataset, we selected the drugs shared by NCI-60 and CCLE. For DRA interactions identified using NCI-60 data for the shared drugs, we annotated the perturbation status of each DRA interaction with CCLE sequencing data. If any interacting partners in a DRA interaction were not available in the CCLE dataset, we ignored the interaction from the validation. For the each remaining DRA interaction–drug pair, we performed t-tests to check whether the drug responses between the two groups with/without the DRA interaction perturbation differed. If the p-value was less than 0.5, we classified the DRA interaction as validated.

2.7. Training random forest classifiers for predicting drug responses

We trained a random forest classifier for each drug to examine whether the interaction perturbation could be used to predict the response to anticancer drugs more efficiently than DRA gene information (Fig. 1A). We defined the cell line as responsive to the drug if z-transformed -log$GI50 > 0.05$ and irresponsive if z-transformed -log$GI50 < 0.05$. We trained random forest classifiers using the leave-one-out cross-validation approach. Two random forest classifiers were trained for each drug, one with the perturbation information of 760 interactions and the other with the mutation information of 3135 genes. Random forest models were trained using “caret” and “randomForest” packages in R and used Cohen's kappa coefficients as a performance measure.