



DRIM: A Web-Based System for Investigating Drug Response at the Molecular Level by Condition-Specific Multi-Omics Data Integration

Minsik Oh¹, Sungjoon Park¹, Sangseon Lee², Dohoon Lee³, Sangsoo Lim², Dabin Jeong³, Kyuri Jo⁴, Inuk Jung⁵ and Sun Kim^{2,3,6*}

¹ Department of Computer Science and Engineering, Seoul National University, Seoul, South Korea, ² Bioinformatics Institute, Seoul National University, Seoul, South Korea, ³ Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea, ⁴ Department of Computer Engineering, Chungbuk National University, Cheongju, South Korea, ⁵ Department of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea, ⁶ Department of Computer Science and Engineering, Institute of Engineering Research, Seoul National University, Seoul, South Korea

OPEN ACCESS

Edited by:

Xiaoyong Pan,
Shanghai Jiao Tong University, China

Reviewed by:

Hamed Bostan,
North Carolina State University,
United States
Tao Zeng,
Shanghai Research Center for Brain
Science and Brain-Inspired
Intelligence, China
Fuhai Li,
Washington University in St. Louis,
United States

*Correspondence:

Sun Kim
sunkim.bioinfo@snu.ac.kr

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 22 May 2020

Accepted: 14 August 2020

Published: 12 November 2020

Citation:

Oh M, Park S, Lee S, Lee D, Lim S,
Jeong D, Jo K, Jung I and Kim S
(2020) DRIM: A Web-Based System
for Investigating Drug Response at the
Molecular Level by Condition-Specific
Multi-Omics Data Integration.
Front. Genet. 11:564792.
doi: 10.3389/fgene.2020.564792

Pharmacogenomics is the study of how genes affect a person's response to drugs. Thus, understanding the effect of drug at the molecular level can be helpful in both drug discovery and personalized medicine. Over the years, transcriptome data upon drug treatment has been collected and several databases compiled before drug treatment cancer cell multi-omics data with drug sensitivity (IC_{50} , AUC) or time-series transcriptomic data after drug treatment. However, analyzing transcriptome data upon drug treatment is challenging since more than 20,000 genes interact in complex ways. In addition, due to the difficulty of both time-series analysis and multi-omics integration, current methods can hardly perform analysis of databases with different data characteristics. One effective way is to interpret transcriptome data in terms of well-characterized biological pathways. Another way is to leverage state-of-the-art methods for multi-omics data integration. In this paper, we developed Drug Response analysis Integrating Multi-omics and time-series data (DRIM), an integrative multi-omics and time-series data analysis framework that identifies perturbed sub-pathways and regulation mechanisms upon drug treatment. The system takes drug name and cell line identification numbers or user's drug control/treat time-series gene expression data as input. Then, analysis of multi-omics data upon drug treatment is performed in two perspectives. For the multi-omics perspective analysis, IC_{50} -related multi-omics potential mediator genes are determined by embedding multi-omics data to gene-centric vector space using a tensor decomposition method and an autoencoder deep learning model. Then, perturbed pathway analysis of potential mediator genes is performed. For the time-series perspective analysis, time-varying perturbed sub-pathways upon drug treatment are constructed. Additionally, a network involving transcription factors (TFs), multi-omics potential mediator genes, and perturbed sub-pathways is constructed, and paths to

perturbed pathways from TFs are determined by an influence maximization method. To demonstrate the utility of our system, we provide analysis results of sub-pathway regulatory mechanisms in breast cancer cell lines of different drug sensitivity. DRIM is available at: <http://biohealth.snu.ac.kr/software/DRIM/>.

Keywords: multi-omics, drug-response, time-series, perturbed pathway, web-system, pharmacogenomics

1. INTRODUCTION

The variability in drug responses among cells is a major challenge in cancer drug therapy, thus personalized drug response research is much needed (Sweeney, 1983). With the recent advances in instrument technologies, drug response analysis at the molecular level has become possible, thus we have an opportunity to investigate relationship between drug response phenotypes and corresponding molecular data, for example, multi-omics data upon drug treatment. Large-scale drug response genomics data help identify molecular markers related with therapeutic response (Garnett et al., 2012). Furthermore, more than 100 US Food and Drug Administration (FDA)-approved drugs have been developed from rapidly growing pharmacogenomics studies. This shows that pharmacogenomics data could be used for drug development at various stages, from drug targets to patient therapeutics. Moreover, genomics data of the patient can be regarded as a predictive factor for drug response. It can be thought of as an early response signal before the phenotypic change of cells by drug (Surendiran et al., 2008).

Current pharmacogenomics data analysis can be extended in two directions to broaden the understanding of drug response. The first direction is to perform a pathway-level analysis. Analyzing drug responses at the individual gene level is difficult to explain biological variability and also difficult to interpret gene-drug associations (Wang et al., 2019). Thus, focus of pharmacogenomics research is changing to investigate multiple gene products at the biological pathway level (Weinshilboum and Wang, 2004). A recent study shows that analysis of transcriptome data can be effectively done at the pathway level, which facilitates clear biological interpretation (Lim et al., 2020). The second direction is to perform multi-omics level analysis. Recently, precision medicine studies have been conducted at the multi-omics level, which is called “pharmaco-omics” beyond pharmacogenomics by integrating genomics, proteomics, epigenomics, and metabolomics data (Adam and Aliferis, 2019; Ginsburg et al., 2019). Many studies have shown that multi-omics integration helps unravel complex biological mechanisms (Subramanian et al., 2020). Integrative analysis of multi-omics data can help understand cell line-specific gene regulation mechanisms for pathway activation (Kim et al., 2016; Oh et al., 2020) and it can be used as a signature for drug response sub-pathway identification (Xu et al., 2019). Single omics analysis can detect only a smaller subset, but multi-omics analysis can detect more comprehensive pathways that respond to chemical exposure (Canzler et al., 2020).

There are several pharmacogenomics databases such as Genomics of Drug Sensitivity in Cancer (GDSC) (Iorio et al.,

2016), Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012), Patient-Derived Xenograft (PDX) mice models (Gao et al., 2015), and NCI-60 Human Tumor Cell Lines Screen (Abaan et al., 2013). These databases can be used for cell line-specific drug sensitivity analysis with multi-omics signature at the molecular level. In addition, data from after drug treatment time-series experiments can be used to capture time-varying cell line-specific drug response as signature of cell death, proliferation, and drug resistance. The Library of Integrated Network-based Cellular Signatures (LINCS) L-1000 (Subramanian et al., 2017) project measures cell viability upon genetic and chemical perturbations by 978 landmark genes. Another database compiled time-series transcriptome data using the NCI-60 cell line upon anti-cancer drug treatment (Monks et al., 2018).

There are several databases that enable computational pharmacogenomics study. GDSC measured the response of 988 cell lines to 518 drug compounds (Iorio et al., 2016). It provides mutation, copy number variation, DNA methylation, and gene expression data of cell lines before drug treatment. CCLE (Barretina et al., 2012) measured genomics profiles and response to 24 anticancer drugs in 947 cell lines. A recent study (Ghandi et al., 2019) performed RNA sequencing (RNA-seq), whole-exome sequencing (WES), whole-genome sequencing (WGS), reverse-phase protein array (RPPA), reduced representation bisulfite sequencing (RRBS), microRNA expression profiling, histone modification profiling, metabolites profiling (Li et al., 2019), and 1,448 drugs response (Corsello et al., 2020) for CCLE cell lines. NCI-60 cell lines are the most widely studied cell lines in human cancer research. CellMiner (Reinhold et al., 2012) is a website that provides 20,503 chemical compounds response of NCI-60 cells and also genomics data before drug treatment as mutation, DNA methylation, microRNA expression, gene expression, and protein data. The NCI Transcriptional Pharmacodynamics Workbench (NCI TPW) (Monks et al., 2018) provides time-series pharmacogenomics data and a web page that allows data exploration. They measured that gene expression changes the NCI-60 cell line after drug exposure of 2, 6, and 24 h to 15 anticancer drugs. NCI-DREAM community (Bansal et al., 2014) measured that gene expression changes the OCI-LY3 cell line after 14 anticancer drug treatment for 6, 12, and 24 h to predict the activity of pairs of compounds.

By utilizing pharmacogenomics data in various databases, a number of studies have been performed to analyze pharmacogenomics data in terms of IC50 prediction and drug response gene/pathway identification. **Table 1** summarizes pharmacogenomics data analysis methods. Multi-Omics Late Integration (MOLI) (Sharifi-Noghabi et al., 2019) is an end-to-end deep neural network-based drug response prediction

TABLE 1 | Pharmacogenomics data analysis methods, their input, output, and algorithms.

Method	Input	Output	Algorithm
MOLI	Multi-omics data	Drug response (IC50)	Deep learning
DSPLMF	Multi-omics data, chemical structures	Drug response (IC50)	Logistic Matrix Factorization
CancerDAP	Multi-omics data	Sub-pathway signatures for drug response	Random forest, logistic regression
DryNetMC	Drug treatment time-series gene expression data	Clinically relevant genes	Differential network analysis

method. MOLI takes mutation, copy number, and gene expression as input, and predicts drug response using each omics type-specific encoder. Drug Sensitivity Prediction using a novel regularization approach in Logistic Matrix Factorization (DSPLMF) (Emdadi and Eslahchi, 2020) is a drug response prediction method based on recommender systems. DSPLMF takes cell line similarity matrix consisted of gene, copy number, mutation, and IC50 and drug similarity matrix as input, and predicts drug response using matrix factorization and nearest neighbor algorithm. CancerDAP (Xu et al., 2019) is a pipeline that integrates gene expression, copy number variation, and DNA methylation to identify sub-pathway signature of anticancer drug response. The user can browse drug active sub-pathway using CancerDAP web page. Differential regulatory Network-based Modeling and Characterization (DryNetMC) (Zhang et al., 2019) is a network-based algorithm to detect key cancer resistance genes based on time-series RNA-seq data. DryNetMC uses time-series RNA-seq data after drug treatment as input. From the data, it constructs drug-sensitive network and drug-resistant network utilizing ordinary differential equations and extracts differential network. Using differential network, a node importance is measured by topology, entropy, and gene expression changes to prioritize genes of clinical relevance.

Lv et al. (2018) performed an analysis of differentially expressed genes (DEGs) on hepatocellular carcinoma (HCC) patients for drug discovery from gene expression data. They divided HCC patients into two groups: high/low-PKM2 to investigate the effect of pyruvate kinase isozymes M2 (PKM2) gene expression on HCC patients in terms of metabolic changes and prognosis. The study identified metabolic genes related to poor HCC patient survival and screened drugs that target metabolic enzymes associated with poor survival. Some of the screened drugs have been used in antitumor clinical studies. Another study proposed a tensor decomposition-based drug discovery method for neurological disorder from gene expression data (Taguchi and Turki, 2019). They selected genes through tensor decomposition-based feature extraction using mouse Alzheimer's single-cell RNA-seq data. These genes are significantly overlapped with the target genes of Alzheimer's disease drugs. Recently, a deep learning-based generative model

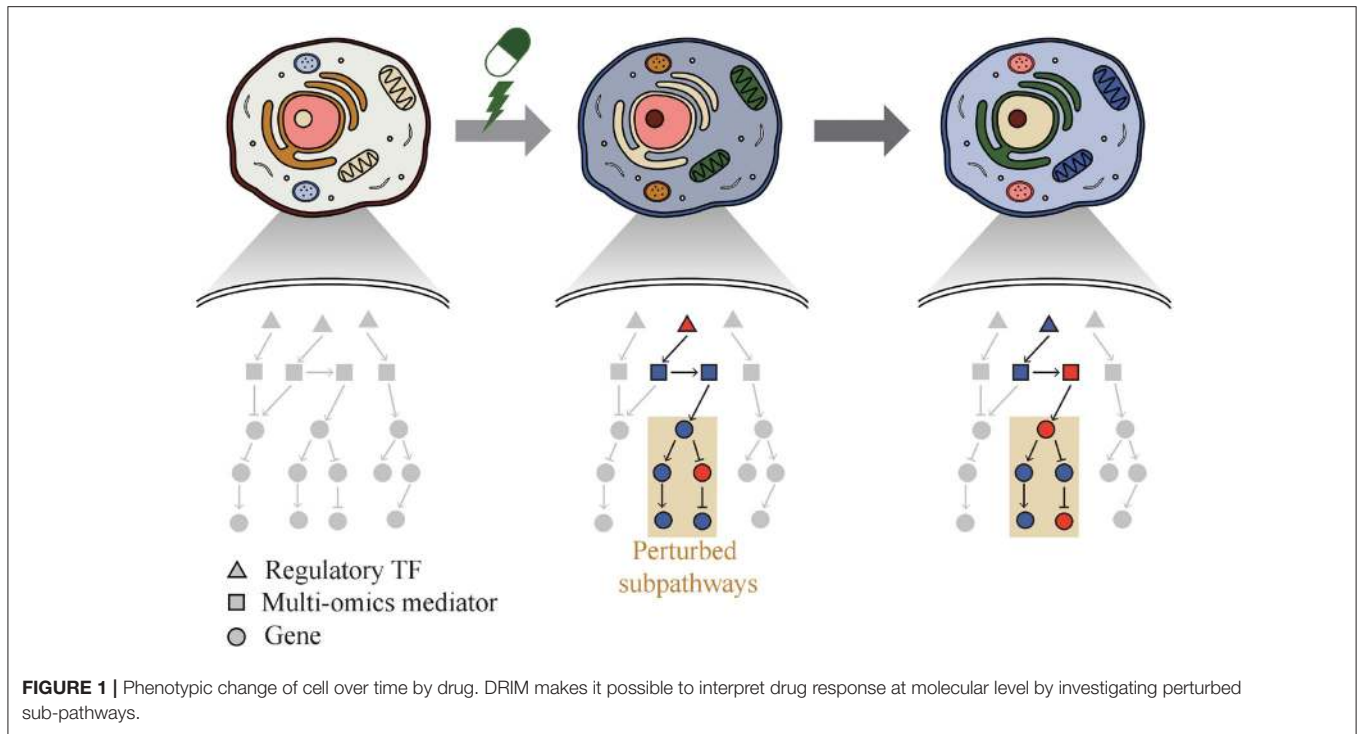
(Méndez-Lucio et al., 2020) proposed to design active-like molecules from gene expression signatures. The generative model takes the desired gene expression profile induced by drug treatment or gene knock-out experiment as input. The study generates a molecular representation that is likely to have caused a change in gene expression.

2. MOTIVATION

To utilize rapidly accumulating drug response omics data, many computational methods for drug response prediction have been developed. Machine learning methods are often used to process high-dimensional genomics data, such as matrix-factorization models (Brouwer and Lió, 2017; Wang et al., 2017), network-based models (Zhang et al., 2015, 2018), and deep learning models (Sharifi-Noghabi et al., 2019; Baptista et al., 2020). Moreover, analysis methods for time-series omics data have been developed (Jo et al., 2016; Ahn et al., 2019; Kang et al., 2019; Kim et al., 2019). However, utilizing these tools for the analysis of pharmacogenomics databases requires expert-level bioinformatics skill.

Thus, a web-based system called Drug Response analysis Integrating Multi-omics and time-series data (DRIM) was developed and presented in this paper by integrating condition-specific multi-omics data to investigate temporal drug response at the molecular level like **Figure 1**. The condition of the sample can be defined as a combination of three variables that are cell line type, drug type, and drug dose. DRIM aims to identify perturbed sub-pathways and regulatory mechanisms upon drug treatment using an integrative analysis framework on both multi-omics and time-series data. By simply taking drug name and cell line ID or user's drug control/treat time-series gene expression data as input, DRIM performs the analysis in two perspectives. First, IC_{50} -related multi-omics potential mediator genes are chosen by embedding multi-omics data into gene-centric vector space using either a tensor decomposition or an autoencoder deep learning model. The tensor decomposition does not require pre-training to determine relationship among different omics components. Feature space from tensor decomposition is linear combination of input features, thus it is easy to interpret how the feature space combines input features. On the other hand, the autoencoder can learn nonlinear relationship of multi-omics data. Autoencoder requires pre-training but it can generate a feature space dynamically for new incoming multi-omics data. In terms of computation time, tensor decomposition is faster than the autoencoder. Then, the potential mediator genes are extended to the identification of perturbed pathways upon drug treatment over time. This time-series analysis construct a network containing transcription factors (TFs), multi-omics mediator genes, and perturbed sub-pathways by an influence maximization-based method.

To demonstrate the utility of our system, we provide analysis results of sub-pathway regulatory mechanisms in breast cancer cell lines of different breast cancer drug sensitivity.



3. METHODS

The system workflow is illustrated in **Figure 2**. In Step 1, the user selects a drug and cell lines to be analyzed for perturbed pathway analysis or uploads their drug control/treat time-series gene expression data. In Step 2, through time-series gene expression data analysis after drug treatment, perturbed sub-pathways are identified. In Step 3, multi-omics potential mediator genes are selected by multi-omics integration methods. In Step 4, a time-bounded network is constructed and the most regulatory path is identified by influence maximization. In Step 5, the system visualizes networks involving TF, mediator genes, and perturbed sub-pathways that change over time upon drug treatment. A detailed description of each step in the workflow is as follows.

3.1. Step 1: Input

The user selects a drug and cell lines to be analyzed for perturbed pathway analysis or uploads their own drug control/treat time-series gene expression data. The system uses two time-series gene expression after drug treatment databases LINCS L-1000 and NCI-60. In both databases, there are control and treated data for drugs per cell line. For each condition, the gene expression was measured at each time point. These databases are available as GSE70138 and GSE116438 in GEO.

3.2. Step 2: Identifying Perturbed Sub-pathway With Time Series

Step 2 is for identifying perturbed sub-pathways of DEGs that are defined using a time-series data analysis tool, TimeTP (Jo et al., 2016). First, each pathway is represented as a directed graph from the KEGG pathway database. For each node in the

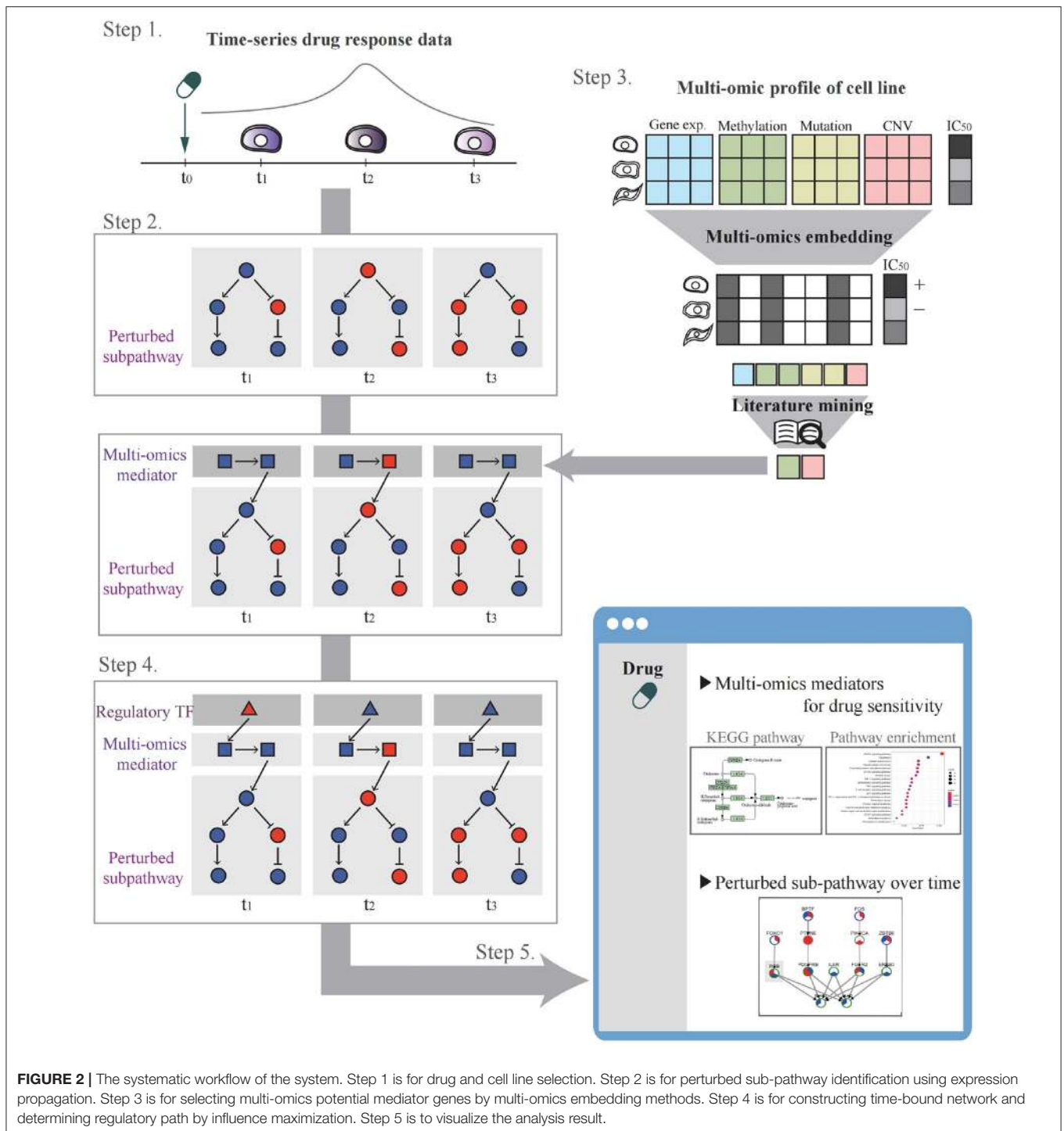
pathway, the system assigns a time vector \vec{v} of 1 (overexpressed) or -1 (underexpressed) and 0 (unchanged) that are defined by comparing gene expression levels, treated vs. control. Limma (Smyth, 2005; Ritchie et al., 2015) was used to define DEGs at each time point with robust multiarray average (RMA) normalization (Kupfer et al., 2012). When there is no control sample, differential expression genes are defined by comparing either to the expression level of the previous time point or to the expression level of initial time point. Second, a perturbed sub-pathway is determined by choosing valid edges in the pathway graph. Assume that there is an edge $N1 \rightarrow N2$ between two genes, $N1$ and $N2$, that have differential time vectors \vec{v}_1 and \vec{v}_2 . To measure the direction of propagation and the number of delayed time points between two vectors, cross-correlation is defined as

$$(\vec{v}_1 \star \vec{v}_2)(n) = \sum_{t=-\infty}^{\infty} \vec{v}_1(t)\vec{v}_2(t+n) \quad (1)$$

where $\vec{v}(t) = 0$ for $t \leq 0$ or $t > T$ (this happens at the preceding or trailing entries of two vectors). Cross-correlation is maximized when the two vectors overlap most with n delay.

$$d(\vec{v}_1, \vec{v}_2) = \operatorname{argmax}_n (\vec{v}_1 \star \vec{v}_2)(n) \quad (2)$$

If $d(\vec{v}_1, \vec{v}_2)$ is negative, it means that the propagation direction is opposite to the given direction. The opposite edge is considered as invalid and excluded from the perturbed sub-pathway. When delay n is larger than a threshold value, the edge is filtered out. After choosing valid edges, a sub-graph that has more than two valid edges is determined



as a perturbed sub-pathway. *P*-value of a perturbed sub-pathway is determined by permutation test. The null distribution is generated by randomly re-assigning differential expression vector for each gene in the sub-pathways. A sum of cross-correlations of edges is used as a pathway-level statistics and *P*-value for a perturbed sub-pathway is calculated from the null distribution.

3.3. Step 3: Embedding Multi-Omics for Selecting Potential Mediator Genes

Step 3 determines potential mediator genes related to drug sensitivity from the multi-omics regulation perspective. The system integrates four multi-omics data such as gene expression, copy number variation, DNA methylation, and mutation from the CCLC database. Each omics data processed to a gene-centric

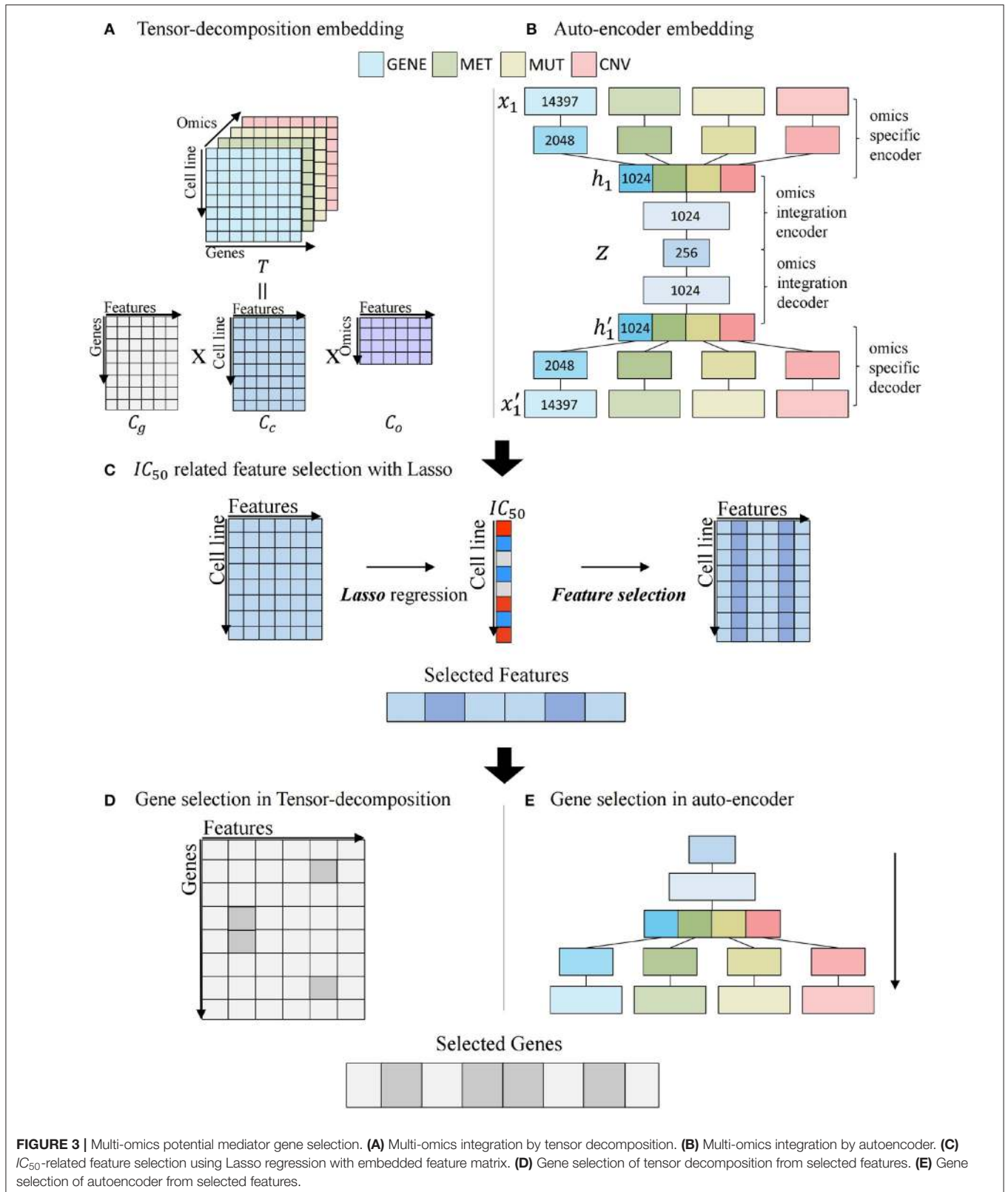


FIGURE 3 | Multi-omics potential mediator gene selection. **(A)** Multi-omics integration by tensor decomposition. **(B)** Multi-omics integration by autoencoder. **(C)** IC_{50} -related feature selection using Lasso regression with embedded feature matrix. **(D)** Gene selection of tensor decomposition from selected features. **(E)** Gene selection of autoencoder from selected features.

(*cell line* × *gene*) matrix to discover potential mediator genes from the perspective of multi-omics regulation. The gene expression and copy number variation values were normalized by min-max normalization. The mutation data were binarized to 1 if mutations exist in the gene or 0 otherwise. To process methylation data, methylation levels of probes located within the transcription start site and 1 KB upstream of promoter regions were averaged per gene. The IC_{50} value measured for each cell line is used as the drug response phenotype.

The system uses two machine learning algorithms, a tensor-decomposition method and an autoencoder method, to embed high-dimensional multi-omics data to low-dimensional feature space. The embedding of the multi-omics data is to create a “gene-centric” feature space, which means that regulation information, such as copy number variation, DNA methylation, and mutation, is tied to a gene while embedding multi-omics data.

Figures 3A,B illustrates the process of embedding gene-centric multi-omics data with two algorithms. For tensor decomposition, we used the PARAFAC model that decomposes a tensor into three two-dimensional matrices (Rabanser et al., 2017). As shown in **Figure 3A**, tensor T with elements x_{ijk} composed of *cell line* × *gene* × *omics* matrix and is factorized into three matrix C_g , C_c , and C_o with g_{if} , c_{jf} , and o_{kf} . C_g , C_c , and C_o are defined as gene, cell line, and omics components, respectively. $f = 1, \dots, R$, R is the number of features.

$$x_{ijk} = \sum_{f=1}^R g_{if} c_{jf} o_{kf} + e_{ijk} \quad (3)$$

We used C_c matrix that embeds cell line-specific multi-omics relationship.

Figure 3B describes the process of autoencoder embedding that is unsupervised artificial neural network to learn efficient encoded representation of data (Kramer, 1991). We constructed a late-integration autoencoder that encodes gene-centric multi-omics data. An input vector is represented as $x = (x_1, \dots, x_n, x_{n+1}, \dots, x_{2n}, x_{2n+1}, \dots, x_{3n}, x_{3n+1}, \dots, x_{4n})$ that is a concatenation of four multi-omics values and n is the number of genes. An autoencoder is to reconstruct x' as output for an input vector x . For each layer l , we used *relu* as activation function between input layer x and output layer y .

$$y = f_l(x) = \text{relu}(W_l x + b_l) \quad (4)$$

The autoencoder consists of four system components: an omics-specific encoder, an omics-integration encoder, an omics-integration decoder, and an omics-specific decoder. In the omics-specific encoder, features are learned individually for each omics data. For each omics data of x_i with $i = (1, 2, 3, 4)$, x_i is encoded to h_i .

$$h_i = F^k(x) = f_k \circ \dots \circ f_1(x) \quad (5)$$

where k is the number of layer, and $f_k \circ f_{k-1}(x) = f_k(f_{k-1}(x))$ is the composition function of f . The omics-integration encoder learns relationship among multi-omics data using concatenated

omics features $h = (h_1, h_2, h_3, h_4)$ and encodes h to z in a similar way to Equation (5). z is an embedding vector that learns the regulation of multi-omics relationship. The omics-integration decoder decodes z to h' . The omics-specific decoder decodes omics specific h'_i to x'_i and reconstruct input $x' = (x'_1, x'_2, x'_3, x'_4)$ in the opposite way to the encoder. For each encoder and decoder, we used 2 layers and 2,048, 1,024 hidden neurons in the omics specific layers, and 1,024, 256 hidden neurons in the omics integration layers. We used *mean squared error (MSE) loss* as a loss function with $L2$ regularization on the weight vector such as Equation (6).

$$\text{Loss} = \sum_{i=1}^N \frac{1}{N} (x_i - x'_i)^2 + \lambda * \sum_{i=1}^P |w_i| \quad (6)$$

N is the number of data, P is the number of layer, and w_i is the weight of i th layer. **Figure 3C** illustrates the feature selection process, using C_c matrix by tensor decomposition or z vector by autoencoder multi-omics embedding matrix. Least Absolute Shrinkage and Selection Operator (LASSO) regression model (Tibshirani, 1996) is constructed using IC_{50} as a target value. Features with non-zero coefficients in regression are considered as features that are significantly associated with the IC_{50} value.

Figures 3D,E depicts the gene selection step related to associated features from **Figure 3C**. In **Figure 3D**, tensor-decomposition method using C_g matrix is for gene selection. For each gene, the row-wise *argmax* operation can be used to obtain the feature most related to the gene, and if the feature is among the IC_{50} -related features obtained in the previous step (features whose coefficients are large in Lasso regression), the gene is selected. The product of $C_g(g, f)$ and *coef*(f) is defined as the omics score of the gene, where *coef*(f) is the coefficient of f' th feature in Lasso regression.

The autoencoder method uses decoder part for gene selection in **Figure 3E**. To evaluate features of a gene in terms of multi-omics, a process-selected feature in the decoder is activated and propagated to the omics data layer. Activation of the final layer is measured through the gene-wise summation and the omics score is computed. The significant genes related with the features are selected.

Selection of multi-omics potential mediator genes is done by combining the two scores, a condition-specific omics score and a literature-based score using BEST, a biomedical entity search tool (Lee et al., 2016). When a drug name is submitted to the BEST system, genes that are known to be related to the drug are selected in a ranked list in the order of relevance to the drug. Combining the two scores is done by a method that was developed for microRNA and target gene interaction (Oh et al., 2017).

3.4. Step 4: Construct TF-Regulatory Time-Bounded Network and Identify Regulatory Path

Step 4 is for constructing TF-regulatory time-bounded network and determining regulatory paths. First, two networks are

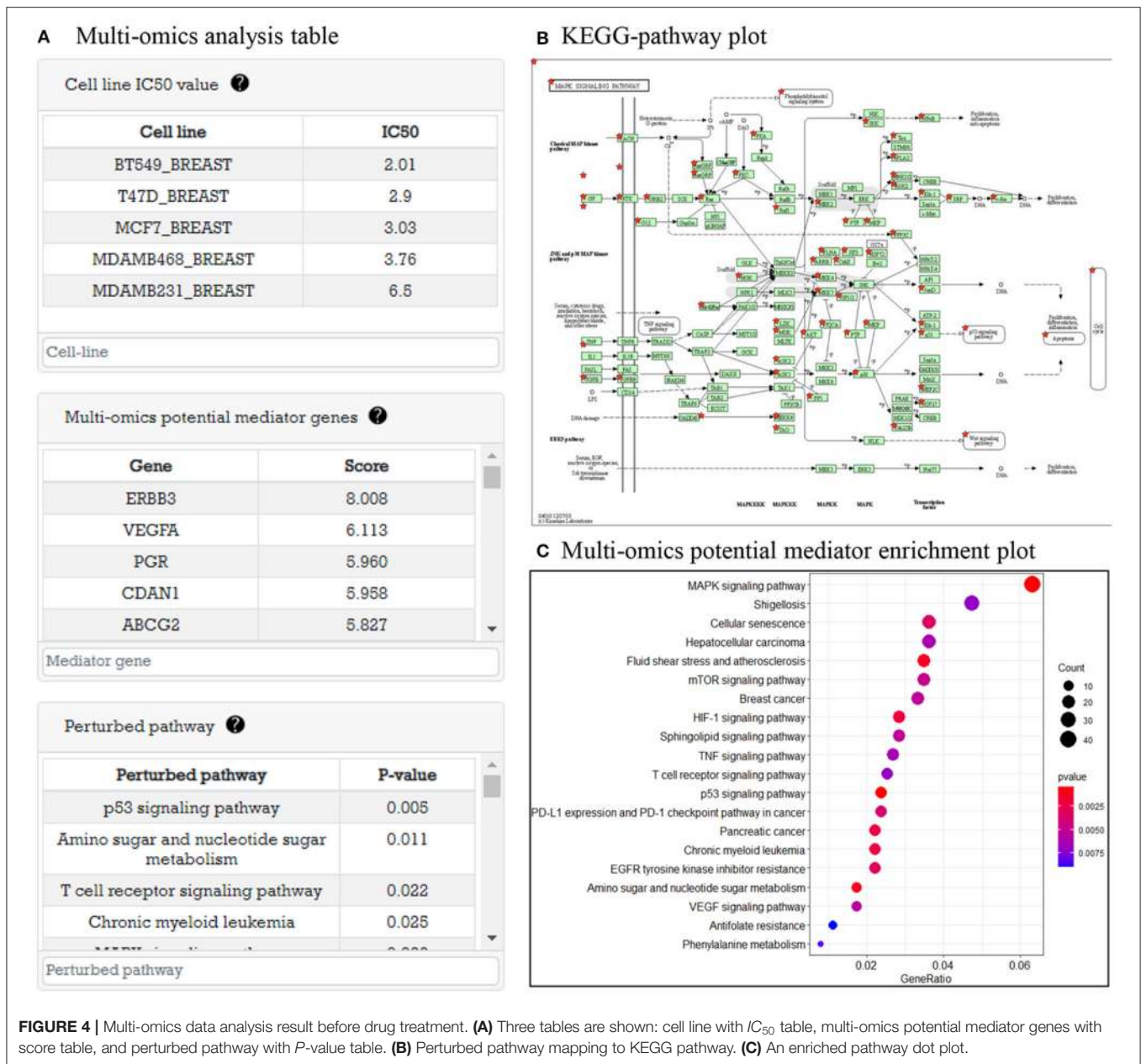


FIGURE 4 | Multi-omics data analysis result before drug treatment. **(A)** Three tables are shown: cell line with IC_{50} table, multi-omics potential mediator genes with score table, and perturbed pathway with P -value table. **(B)** Perturbed pathway mapping to KEGG pathway. **(C)** An enriched pathway dot plot.

constructed to search upstream regulators of perturbed sub-pathway. A gene regulatory network (GRN) is constructed from HTRIdb (Bovolenta et al., 2012) for interaction information between TF and multi-omics potential mediator. A protein-interaction network (PIN) is instantiated from STRING (Szklarczyk et al., 2015) database for gene-gene interaction. To combine GRN, PIN, and perturbed sub-pathways as TF-regulatory time-bounded networks, we used the method described in Step 2.

Next, the most likely regulatory paths are identified by the influence maximization method that has been widely used to select marketing targets in the social network to maximize the spread of influence (Kempe et al., 2003). Our system uses a

labeled influence maximization algorithm (Li, 2011) to the time-bounded network to identify most influential regulatory path from TF to perturbed sub-pathway (Jo et al., 2016).

3.5. Step 5: Analysis Result on the Web

The system provides analysis results on the web from two perspectives: multi-omics data before drug treatment and time-series gene expression data after drug treatment.

3.5.1. Multi-Omics Analysis Result Before Drug Treatment

In this part, system provides analysis results of multi-omics data before drug treatment. As an example, in **Figure 4A**, there are

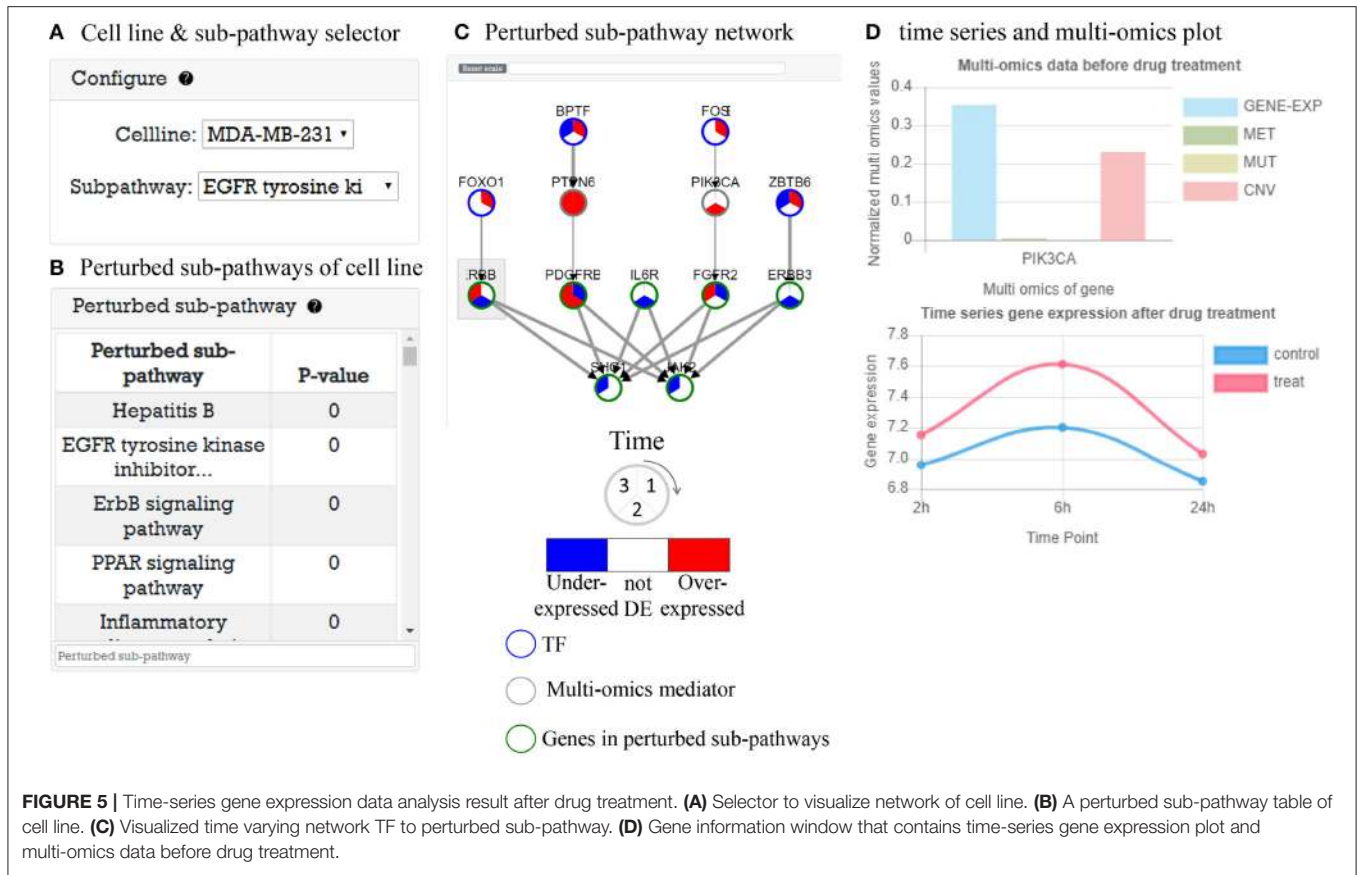


FIGURE 5 | Time-series gene expression data analysis result after drug treatment. **(A)** Selector to visualize network of cell line. **(B)** A perturbed sub-pathway table of cell line. **(C)** Visualized time varying network TF to perturbed sub-pathway. **(D)** Gene information window that contains time-series gene expression plot and multi-omics data before drug treatment.

tables representing cell line IC_{50} , multi-omics potential mediator genes related to IC_{50} value, and perturbed pathways that are enriched by potential mediator genes. In **Figure 4B**, when the user clicks on the pathway in the pathway table, a KEGG pathway plot is created. **Figure 4C** is GO enrichment analysis plot of potential mediator genes to show the biological functions of the multi-omics potential mediator gene set in relation to drug sensitivity.

3.5.2. Time-Series Gene Expression Analysis Result After Drug Treatment

This part provides time-series gene expression data after drug treatment analysis results with perturbed sub-pathways. As an example, in **Figure 5A**, user can select cell line and perturbed sub-pathway to explore. When the user select a cell line, a perturbed sub-pathway table (**Figure 5B**) is generated with P -value. **Figure 5C** shows a TF-pathway network in time clock. When user clicks the gene node, a popup window appears to display multi-omics measurement of gene and expression plot of gene over time like **Figure 5D**. Furthermore, the user can search genes in the network. The user can control the network size by choosing a cut-off value for DEGs to identify perturbed pathway. If the cutoff is low, the number of nodes edges increases, which may cause false positive problems. In the opposite case, there may be a false negative problem. In either case, predicted perturbed

TABLE 2 | Five breast cancer cell lines that are available multi-omics data before drug treatment with lapatinib sensitivity and time-series gene expression data after drug treatment.

Cell line	Molecular sub-subtype	$IC_{50}(\mu M)$
BT-549	Basal B	2.02
T-47D	Luminal	2.90
MCF7	Luminal	3.04
MDA-MB-468	Basal A	3.77
MDA-MB-231	Basal B	6.50

pathways are computationally predicted, thus the user may need to further investigate perturbed pathways.

4. CASE STUDY: COMPARATIVE ANALYSIS OF BREAST CANCER CELL LINES THAT HAVE DIFFERENT SENSITIVITY WITH LAPATINIB

To demonstrate the usefulness of DRIM, we conducted an analysis on breast cancer cell lines in response to lapatinib administration. The lapatinib is a dual inhibitor on both targets epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2) tyrosine kinases

(Medina and Goodin, 2008). It was approved by FDA in combination therapy for HER2-positive/overexpressed breast cancer patients. We chose five representative breast cancer cell lines that have distinct sensitivity/resistance on lapatinib (Table 2). These cell lines are all available on both multi-omics and time-series data to fully utilize the nature of DRIM.

4.1. Multi-Omics Analysis Result Before Drug Treatment

For multi-omics analysis for before drug treatment cells, DRIM selected IC_{50} -related multi-omics potential mediator gene sets that are obtained by multi-omics integration analysis as shown in Figure 4. We carefully examined the set of candidate multi-omics potential mediator genes predictive of lapatinib sensitivity. The top 15 multi-omics potential mediator genes lapatinib are shown in Table 3 sorted by their relevance score with respect to lapatinib. Among the genes, ERBB3 (HER3) was previously known for its critical role in HER2-amplified breast cancer cells (Lee-Hoeflich et al., 2008). It is strongly associated with lapatinib sensitivity in coexpression with neuregulin-1 (NRG1) (Wilson et al., 2011). Genetic perturbations on other genes such as ABCG2, TP53, and HSF1 were also well known for lapatinib resistance (Rahko et al., 2003; Dai et al., 2008; Yallowitz et al., 2018).

4.2. Time-Series Gene Expression Analysis Result After Drug Treatment

For the temporal pharmacogenomic analysis, we investigated cell line-specific perturbed sub-pathways that may be related to different lapatinib response. The lapatinib mainly targets PI3K signaling pathway, which plays a critical role in cell growth, survival, and proliferation (Fruman et al., 2017). Conceivably, aberrant activation of PI3K signaling is known to confer resistance to drugs targeting various receptor tyrosine kinases (Eichhorn et al., 2008; Wang et al., 2011). As expected, we collectively observed a significant time-course perturbation of PI3K signaling in each of the five cell lines in Table 4.

We further examined in detail if there are differential sub-pathway level regulations among cell lines that mediate the response to the drug. Specifically, we asked whether each cell line harbors a distinct time-course regulatory path that governs the expression of a shared protein at the terminus of a pathway. To systematically identify such examples, we sought for the regulatory paths with shared terminator protein for at least two cell lines using the “overview” network generated by DRIM. To simplify the analysis, we defined the terminator proteins as the nodes without outgoing edges in the network. Moreover, for biological interpretability, we only considered the paths starting from the transcription factors, and also enforced the paths to contain at least one multi-omics mediator. Different cell lines responded to lapatinib, accompanying distinct molecular perturbations, and shared the same terminal protein at the end of the paths (Figure 6).

Interestingly, we observed that many proteins involved in PI3K signaling pathway were regulated by different signaling pathways in a cell line-specific manner. For example, vascular

TABLE 3 | Top 15 multi-omics potential mediator genes that are related to lapatinib sensitivity.

Genes	Score
ERBB3	8.01
VEGFA	6.11
PGR	5.96
CDAN1	5.96
ABCG2	5.83
ESR1	5.71
CASP8	5.64
TP53	5.63
MAP2K7	5.62
CNTN4	5.53
DCTN6	5.44
CD274	5.39
NF2	5.31
CBL	5.19
E2F1	5.14

TABLE 4 | The *P*-value of PI3K-Akt signaling pathway.

Cell line	<i>P</i> -value
BT-549	1.6e-05
T-47D	6.07e-03
MCF7	7.11e-04
MDA-MB-468	4.94e-05
MDA-MB-231	1.08e-03

endothelial growth factor A (VEGFA), a well-known effector molecule induced by PI3K signaling pathway (Karar and Maity, 2011), was shown to be activated through different signaling cascades, as shown in Figure 6A. In MDA-MB-468, VEGFA seemed to be induced by aryl hydrocarbon receptor (AhR) and aryl hydrocarbon receptor nuclear translocator (ARNT) signaling, presumably by the increased level of AhR/ARNT heterodimer as shown in Figure 6A. In BT-549 and T-47D cell lines, activation of JNK and NF- κ B signaling was shown to be associated with increased level of VEGFA, respectively. Intriguingly, the time-bounded network allows the interpretation of the temporal difference of VEGFA induction between lapatinib-treated BT-549 and T-47D cell lines, as it can be deduced that the earlier response of T-47D was due to the more rapid induction of NF- κ B than that of FOS in BT-549.

Bcl-2-like protein 1 (BCL2L1) and cyclin D3 (CCND3), overexpressed in human breast cancer, are anti-apoptotic proteins that delay cell death and increases cell survival (Simonian et al., 1997; Chi et al., 2015). In Figure 6B, in T-47D and MDA-MB-231 cell lines, BCL2L1 and CCND3 are downregulated in response to lapatinib that leads to cell death. In T-47D, overexpression of JUN throughout whole phases is a prominent characteristic. JUN is a well-known transcription regulator that induces apoptotic cell death (Bossy-Wetzel et al., 1997). It can be hypothesized that promoted cell death in response to lapatinib is attributed to the increased c-Jun.

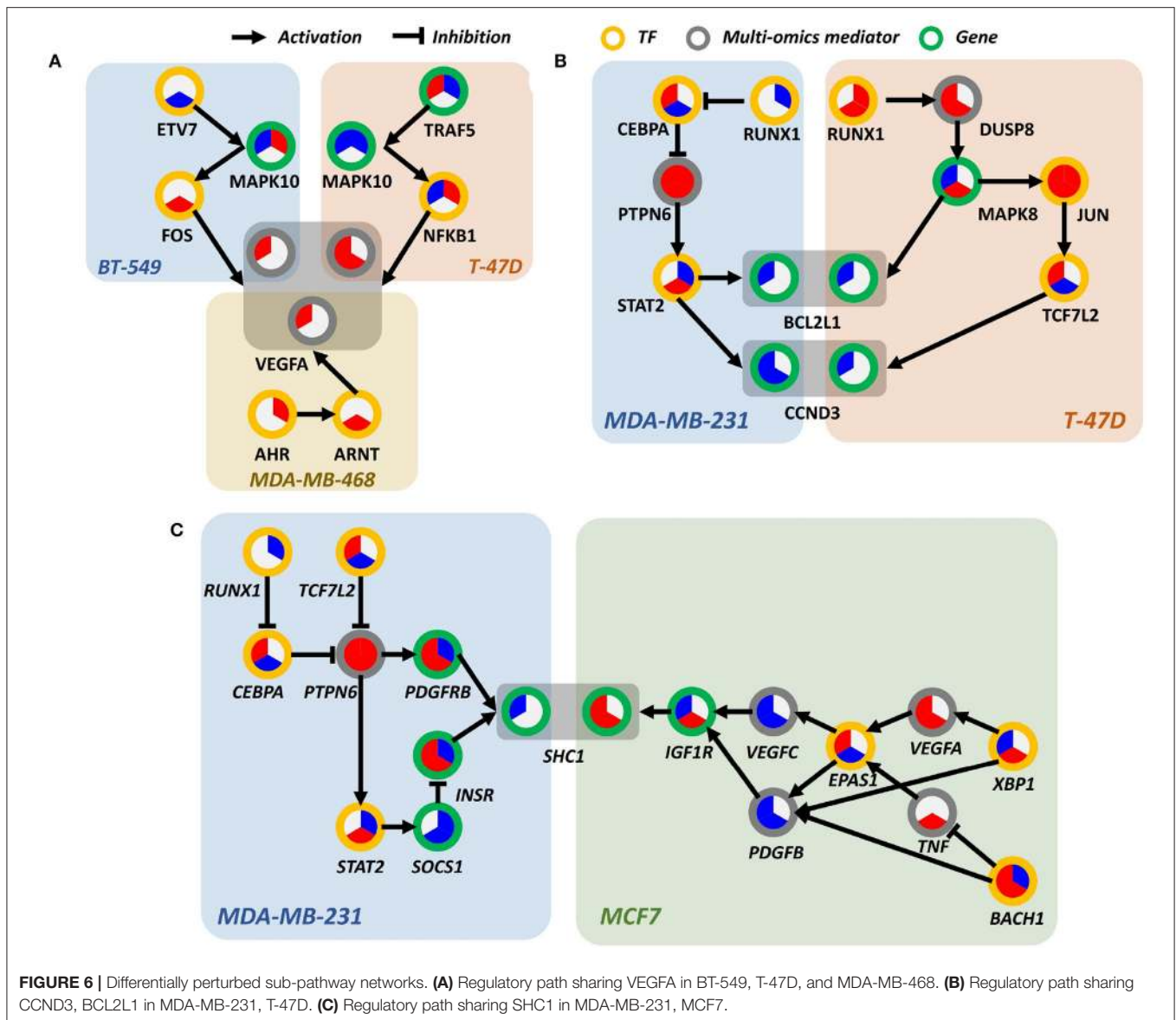


FIGURE 6 | Differentially perturbed sub-pathway networks. **(A)** Regulatory path sharing VEGFA in BT-549, T-47D, and MDA-MB-468. **(B)** Regulatory path sharing CCND3, BCL2L1 in MDA-MB-231, T-47D. **(C)** Regulatory path sharing SHC1 in MDA-MB-231, MCF7.

Another interesting characteristic is that expression of downstream molecule of JUN—transcription factor 7-like 2 (TCF7L2)—increases over time, while T-47D cell line retained a high expression level of JUN. Since activity of c-Jun is predominantly regulated through phosphorylation, expression of molecules in regulatory relations should not be necessarily correlated. In MDA-MB-231, downregulation of BCL2L1 and CCND3 is induced by signal transducer and activator of transcription 2 (STAT2) (Furth, 2014), which involved in the JAK-STAT signaling pathway that leads to oncogenesis (Thomas et al., 2015). Although temporal relations between molecules are not clear, it still gives insight into which pathways are involved in elevated cell death.

SHC-transforming protein 1 (SHC1), a core regulator of receptor tyrosine kinase signaling, is an essential gene for promoting immune suppression. Downstream effects of

SHC1 perturbation lead to STAT3/STAT1-related immune impairment. As previously mentioned, SHC1 can respond to EGF stimulation using multiple paths of protein phosphorylations and interactions (Zheng et al., 2013). There also exists PTPN12 as a turning point of SHC1 pTyr/Grb2 signaling that regulates cell invasion and morphogenesis. Reflecting the previous findings, our results on the perturbed sub-pathways also show multiple regulatory mechanisms that each of the breast cancer cell lines can potentially utilize favorable/possible sub-paths on the temporal flow, as shown in **Figure 6C**. This implies that upstream stimuli including EGF regulation direct multiple paths of temporal information among breast cancer cell lines (Zheng et al., 2013).

Even though all the five breast cancer cell lines were treated with the same drug, lapatinib, targeting receptors at cell surface with extreme specificity, each cell line showed different

sensitivity to the drug. This heterogeneity may occur due to the complex crosstalk between various signaling pathways, which makes the inactivation of single signaling pathway by drug treatment not enough to cause systemic dysregulation of cellular machineries. Our system allows us to dissect this phenomena by differentially characterizing the fragments of regulatory cascades toward various effector molecules for each individual cell line as shown in **Figure 6**. Furthermore, since we have intracellular mechanistic portraits of drug response for each of the cell lines, it may allow us to devise novel combination therapeutic strategies targeting additional molecules that cells depend on after the primary drug is applied.

5. DISCUSSION

For understanding the cell variability in drug response, personalized drug response analysis is demanded. In spite of increasing drug response genomics data, the interaction of high dimension multi-omics and time-series analysis are challenges for pharmacogenomics analysis. Pathway level analysis and multi-omics integration can be effective ways to interpret drug response data.

We developed an integrative multi-omics and time-series data analysis framework DRIM that finds perturbed sub-pathways and regulatory mechanisms in drug response. DRIM identifies the most likely regulatory path involving TF, multi-omics mediator gene, and perturbed sub-pathway for each cell line. DRIM provides analysis results in two perspectives. As a demonstration, we conducted an analysis of breast cancer cell lines that have different *lapatinib* sensitivity. In the multi-omics perspective result, DRIM selected multi-omics potential mediator genes that are related to *lapatinib* resistance in previous studies. In the temporal pharmacogenomic analysis result, we showed that DRIM can be used to discover distinct temporal regulatory mechanisms governing the induction of several common downstream proteins across cell lines.

REFERENCES

- Abaan, O. D., Polley, E. C., Davis, S. R., Zhu, Y. J., Bilke, S., Walker, R. L., et al. (2013). The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology. *Cancer Res.* 73, 4372–4382. doi: 10.1158/0008-5472.CAN-12-3342
- Adam, T., and Aliferis, C. (2019). *Personalized and Precision Medicine Informatics: A Workflow-Based View*. Springer Nature. doi: 10.1007/978-3-030-18626-5
- Ahn, H., Jung, I., Chae, H., Kang, D., Jung, W., and Kim, S. (2019). HTRgene: a computational method to perform the integrated analysis of multiple heterogeneous time-series data: case analysis of cold and heat stress response signaling genes in Arabidopsis. *BMC Bioinformatics* 20:588. doi: 10.1186/s12859-019-3072-2
- Bansal, M., Yang, J., Karan, C., Menden, M. P., Costello, J. C., Tang, H., et al. (2014). A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* 32, 1213–1222. doi: 10.1038/nbt.3052
- Baptista, D., Ferreira, P. G., and Rocha, M. (2020). Deep learning for drug response prediction in cancer. *Brief. Bioinformatics*. doi: 10.1093/bib/bbz171. [Epub ahead of print].
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The cancer cell line encyclopedia enables predictive modelling

DATA AVAILABILITY STATEMENT

CCLE multi-omics data with drug sensitivity analyzed for this study can be found in the depmap portal (<https://depmap.org/portal/download/>). NCI-60 cell line drug-treated time-series data analyzed for this study can be found in the GEO (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116451>). LINCS-L1000 data analyzed for this study can be found in the GEO (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92742>).

AUTHOR CONTRIBUTIONS

SK designed the project. SK, MO, SLe, and SLi designed drug response analysis algorithm. MO and IJ implemented multi-omics integration algorithm. MO, SP, and KJ implemented time-series analysis algorithm. MO and SP implemented web system. DJ designed workflow and web system. SK, SLe, SLi, DL, and DJ biologically interpreted the drug response analysis results. SK, MO, SLe, SLi, DL, DJ, and SP wrote and revised the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT (NRF-2019M3E5D307337511, NRF-2019M3E5D4065965), Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (NRF-2017M3C4A7065887), and the Collaborative Genome Program for Fostering New Post-Genome Industry of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (MSIT) (NRF-2014M3C9A3063541).

- of anticancer drug sensitivity. *Nature* 483, 603–607. doi: 10.1038/nature11003
- Bossy-Wetzell, E., Bakiri, L., and Yaniv, M. (1997). Induction of apoptosis by the transcription factor C-Jun. *EMBO J.* 16, 1695–1709. doi: 10.1093/emboj/16.7.1695
- Bovolenta, L. A., Acencio, M. L., and Lemke, N. (2012). HTRIDB: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* 13:405. doi: 10.1186/1471-2164-13-405
- Brouwer, T. and Lió, P. (2017). Bayesian hybrid matrix factorisation for data integration. *Proc. Machine Learn. Res.* 54, 557–566. doi: 10.17863/CAM.12887
- Canzler, S., Schor, J., Busch, W., Schubert, K., Rolle-Kampczyk, U. E., Seitz, H., et al. (2020). Prospects and challenges of multi-omics data integration in toxicology. *Arch. Toxicol.* 94, 371–388. doi: 10.1007/s00204-020-02656-y
- Chi, Y., Huang, S., Liu, M., Guo, L., Shen, X., and Wu, J. (2015). Cyclin d3 predicts disease-free survival in breast cancer. *Cancer Cell Int.* 15:89. doi: 10.1186/s12935-015-0245-6
- Corsello, S. M., Nagari, R. T., Spangler, R. D., Rossen, J., Kocak, M., Bryan, J. G., et al. (2020). Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* 1, 235–248. doi: 10.1038/s43018-019-0018-6

- Dai, C.-I., Tiwari, A. K., Wu, C.-P., Su, X.-D., Wang, S.-R., Liu, D.-G., et al. (2008). Lapatinib (tykerb, GW572016) reverses multidrug resistance in cancer cells by inhibiting the activity of ATP-binding cassette subfamily B member 1 and G member 2. *Cancer Res.* 68, 7905–7914. doi: 10.1158/0008-5472.CAN-08-0499
- Eichhorn, P. J., Gilli, M., Scaltriti, M., Serra, V., Guzman, M., Nijkamp, W., et al. (2008). Phosphatidylinositol 3-kinase hyperactivation results in lapatinib resistance that is reversed by the mTOR/phosphatidylinositol 3-kinase inhibitor NVP-BEZ235. *Cancer Res.* 68, 9221–9230. doi: 10.1158/0008-5472.CAN-08-1740
- Emdadi, A., and Eslahchi, C. (2020). DSPLMF: A method for cancer drug sensitivity prediction using a novel regularization approach in logistic matrix factorization. *Front. Genet.* 11:75. doi: 10.3389/fgene.2020.00075
- Fruman, D. A., Chiu, H., Hopkins, B. D., Bagrodia, S., Cantley, L. C., and Abraham, R. T. (2017). The PI3K pathway in human disease. *Cell* 170, 605–635. doi: 10.1016/j.cell.2017.07.029
- Furth, P. A. (2014). Stat signaling in different breast cancer sub-types. *Mol. Cell. Endocrinol.* 382, 612–615. doi: 10.1016/j.mce.2013.03.023
- Gao, H., Korn, J. M., Ferretti, S., Monahan, J. E., Wang, Y., Singh, M., et al. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* 21:1318. doi: 10.1038/nm.3954
- Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483, 570–575. doi: 10.1038/nature11005
- Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature* 569, 503–508. doi: 10.1038/s41586-019-1186-3
- Ginsburg, G. S., Willard, H. F., Woods, C. W., and Tsalik, E. L. (2019). *Genomic and Precision Medicine: Infectious and Inflammatory Disease*. Academic Press.
- Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., et al. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell* 166, 740–754. doi: 10.1016/j.cell.2016.06.017
- Jo, K., Jung, I., Moon, J. H., and Kim, S. (2016). Influence maximization in time bounded network identifies transcription factors regulating perturbed pathways. *Bioinformatics* 32, i128–i136. doi: 10.1093/bioinformatics/btw275
- Kang, H., Ahn, H., Jo, K., Oh, M., and Kim, S. (2019). mirTime: identifying condition-specific targets of microRNA in time-series transcript data using Gaussian process model and spherical vector clustering. *Bioinformatics*. doi: 10.1093/bioinformatics/btz306. [Epub ahead of print].
- Karar, J., and Maity, A. (2011). PI3k/akt/mTOR pathway in angiogenesis. *Front. Mol. Neurosci.* 4:51. doi: 10.3389/fnfmol.2011.00051
- Kempe, D., Kleinberg, J., and Tardos, É. (2003). “Maximizing the spread of influence through a social network,” in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 137–146. doi: 10.1145/956750.956769
- Kim, M., Rai, N., Zorraqino, V., and Tagkopoulos, I. (2016). Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nat. Commun.* 7, 1–12. doi: 10.1038/ncomms13090
- Kim, S., Jung, W., Ahn, H., Jo, K., Jung, D., Park, M., et al. (2019). Propanet: Time-varying condition-specific transcriptional network construction by network propagation. *Front. Plant Sci.* 10:698. doi: 10.3389/fpls.2019.00698
- Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* 37, 233–243. doi: 10.1002/aic.690370209
- Kupfer, P., Guthke, R., Pohlers, D., Huber, R., Koczan, D., and Kinne, R. W. (2012). Batch correction of microarray data substantially improves the identification of genes differentially expressed in rheumatoid arthritis and osteoarthritis. *BMC Med. Genomics* 5:23. doi: 10.1186/1755-8794-5-23
- Lee, S., Kim, D., Lee, K., Choi, J., Kim, S., Jeon, M., et al. (2016). Best: Next-generation biomedical entity search tool for knowledge discovery from biomedical literature. *PLoS ONE* 11:e0164680. doi: 10.1371/journal.pone.0164680
- Lee-Hoeflich, S. T., Crocker, L., Yao, E., Pham, T., Munroe, X., Hoeflich, K. P., et al. (2008). A central role for HER3 in HER2-amplified breast cancer: implications for targeted therapy. *Cancer Res.* 68, 5878–5887. doi: 10.1158/0008-5472.CAN-08-0380
- Li, F. (2011). “Privacy, security, risk and trust (PASSAT),” in *2011 IEEE Third International Conference on Social Computing (SocialCom)* (Boston, MA).
- Li, H., Ning, S., Ghandi, M., Kryukov, G. V., Gopal, S., Deik, A., et al. (2019). The landscape of cancer cell line metabolism. *Nat. Med.* 25:850. doi: 10.1038/s41591-019-0404-8
- Lim, S., Lee, S., Jung, I., Rhee, S., and Kim, S. (2020). Comprehensive and critical evaluation of individualized pathway activity measurement tools on pan-cancer data. *Brief. Bioinformatics* 21, 36–46. doi: 10.1093/bib/bby125
- Lv, W.-W., Liu, D., Liu, X.-C., Feng, T.-N., Li, L., Qian, B.-Y., et al. (2018). Effects of PKM2 on global metabolic changes and prognosis in hepatocellular carcinoma: from gene expression to drug discovery. *BMC Cancer* 18, 1–13. doi: 10.1186/s12885-018-5023-0
- Medina, P. J., and Goodin, S. (2008). Lapatinib: a dual inhibitor of human epidermal growth factor receptor tyrosine kinases. *Clin. Therap.* 30, 1426–1447. doi: 10.1016/j.clinthera.2008.08.008
- Méndez-Lucio, O., Baillif, B., Clevert, D.-A., Rouquié, D., and Wichard, J. (2020). De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat. Commun.* 11, 1–10. doi: 10.1038/s41467-019-13807-w
- Monks, A., Zhao, Y., Hose, C., Hamed, H., Krushkal, J., Fang, J., et al. (2018). The NCI transcriptional pharmacodynamics workbench: a tool to examine dynamic expression profiling of therapeutic response in the NCI-60 cell line panel. *Cancer Res.* 78, 6807–6817. doi: 10.1158/0008-5472.CAN-18-0989
- Oh, M., Park, S., Kim, S., and Chae, H. (2020). Machine learning-based analysis of multi-omics data on the cloud for investigating gene regulations. *Brief. Bioinformatics*. doi: 10.1093/bib/bbaa032. [Epub ahead of print].
- Oh, M., Rhee, S., Moon, J. H., Chae, H., Lee, S., Kang, J., et al. (2017). Literature-based condition-specific mirna-mrna target prediction. *PLoS ONE* 12:e174999. doi: 10.1371/journal.pone.0174999
- Rabanser, S., Shchur, O., and Günnemann, S. (2017). Introduction to tensor decompositions and their applications in machine learning. *arXiv. [Preprint]. arXiv:1711.10781*.
- Rahko, E., Blanco, G., Soini, Y., Bloigu, R., and Jukkola, A. (2003). A mutant TP53 gene status is associated with a poor prognosis and anthracycline-resistance in breast cancer patients. *Eur. J. Cancer* 39, 447–453. doi: 10.1016/S0959-8049(02)00499-9
- Reinhold, W. C., Sunshine, M., Liu, H., Varma, S., Kohn, K. W., Morris, J., et al. (2012). Cellminer: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.* 72, 3499–3511. doi: 10.1158/0008-5472.CAN-12-1370
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Sharifi-Noghabi, H., Zolotareva, O., Collins, C. C., and Ester, M. (2019). Moli: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 35, i501–i509. doi: 10.1093/bioinformatics/btz318
- Simonian, P. L., Grillot, D. A., and Nu nez, G. (1997). Bak can accelerate chemotherapy-induced cell death independently of its heterodimerization with Bcl-XL and Bcl-2. *Oncogene* 15, 1871–1875. doi: 10.1038/sj.onc.1201350
- Smyth, G. (2005). “Limma: linear models for microarray data,” in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, eds R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber (New York, NY: Springer), 397–420. doi: 10.1007/0-387-29362-0_23
- Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171, 1437–1452. doi: 10.1016/j.cell.2017.10.049
- Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinformatics Biol. Insights* 14:1177932219899051. doi: 10.1177/1177932219899051
- Surendiran, A., Pradhan, S., and Adithan, C. (2008). Role of pharmacogenomics in drug discovery and development. *Indian J. Pharmacol.* 40:137. doi: 10.4103/0253-7613.43158
- Sweeney, G. (1983). Variability in the human drug response. *Thromb. Res.* 29, 3–15. doi: 10.1016/0049-3848(83)90353-5
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). String v10: protein-protein interaction networks, integrated over the tree of life. *Nucl. Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003

- Taguchi, Y., and Turki, T. (2019). Neurological disorder drug discovery from gene expression with tensor decomposition. *Curr. Pharm. Design* 25, 4589–4599. doi: 10.2174/1381612825666191210160906
- Thomas, S., Snowden, J., Zeidler, M., and Danson, S. (2015). The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. *Br. J. Cancer* 113:365. doi: 10.1038/bjc.2015.233
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Wang, L., Li, X., Zhang, L., and Gao, Q. (2017). Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer* 17:513. doi: 10.1186/s12885-017-3500-5
- Wang, L., Zhang, Q., Zhang, J., Sun, S., Guo, H., Jia, Z., et al. (2011). PI3K pathway activation results in low efficacy of both trastuzumab and lapatinib. *BMC Cancer* 11:248. doi: 10.1186/1471-2407-11-248
- Wang, X., Sun, Z., Zimmermann, M. T., Bugrim, A., and Kocher, J.-P. (2019). Predict drug sensitivity of cancer cells with pathway activity inference. *BMC Med. Genomics* 12:15. doi: 10.1186/s12920-018-0449-4
- Weinshilboum, R., and Wang, L. (2004). Pharmacogenomics: bench to bedside. *Nat. Rev. Drug Discov.* 3, 739–748. doi: 10.1038/nrd1497
- Wilson, T. R., Lee, D. Y., Berry, L., Shames, D. S., and Settleman, J. (2011). Neuregulin-1-mediated autocrine signaling underlies sensitivity to HER2 kinase inhibitors in a subset of human cancers. *Cancer Cell* 20, 158–172. doi: 10.1016/j.ccr.2011.07.011
- Xu, Y., Dong, Q., Li, F., Xu, Y., Hu, C., Wang, J., et al. (2019). Identifying subpathway signatures for individualized anticancer drug response by integrating multi-omics data. *J. Transl. Med.* 17:255. doi: 10.1186/s12967-019-2010-4
- Yallowitz, A., Ghaleb, A., Garcia, L., Alexandrova, E. M., and Marchenko, N. (2018). Heat shock factor 1 confers resistance to lapatinib in ERBB2-positive breast cancer cells. *Cell Death Dis.* 9, 1–13. doi: 10.1038/s41419-018-0691-x
- Zhang, F., Wang, M., Xi, J., Yang, J., and Li, A. (2018). A novel heterogeneous network-based method for drug response prediction in cancer cell lines. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-018-21622-4
- Zhang, J., Zhu, W., Wang, Q., Gu, J., Huang, L. F., and Sun, X. (2019). Differential regulatory network-based quantification and prioritization of key genes underlying cancer drug resistance based on time-course RNA-seq data. *PLoS Comput. Biol.* 15:e1007435. doi: 10.1371/journal.pcbi.1007435
- Zhang, N., Wang, H., Fang, Y., Wang, J., Zheng, X., and Liu, X. S. (2015). Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. *PLoS Comput. Biol.* 11:e1004498. doi: 10.1371/journal.pcbi.1004498
- Zheng, Y., Zhang, C., Croucher, D. R., Soliman, M. A., St-Denis, N., Pasculescu, A., et al. (2013). Temporal regulation of EGF signalling networks by the scaffold protein SHC1. *Nature* 499, 166–171. doi: 10.1038/nature12308

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Oh, Park, Lee, Lee, Lim, Jeong, Jo, Jung and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.