Original Research

Exploring effects of DNA methylation and gene expression on pan-cancer drug response by mathematical models

Wenhua Lv[1], Xingda Zhang[2], Huili Dong[3], Qiong Wu[3], Baoqing Sun[4] and Yan Zhang[4]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China; [2]Department of Breast Surgery, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang 150081, China; [3]School of Life Science and Technology, Computational Biology Research Center, Harbin Institute of Technology, Harbin 150001, China; [4]Guangzhou Institute of Respiratory health, State Key Laboratory of Respiratory Disease, Guangzhou Medical University, Guangzhou 51000, China

Corresponding authors: Yan Zhang. Email: zhangtyo@hit.edu.cn; Baoqing Sun. Email: sunbaoqing@vip.163.com

Impact statement

Aberrant DNA methylation patterns are important factors of drug sensitivity and affect gene expression in transcriptional regulation. Therefore, we aimed to identify genes whose DNA methylation levels were related to anticancer drug response and to examine the influence of synergistic action of DNA methylation and gene expression on anticancer drug response. Digging data of epigenetics and genetics with mathematical methods helps researchers and physicians further understand the molecular mechanisms of cancer resistance. The large-scale drug screening and algorithm improvement of drug-sensitive target recognition will be beneficial to identify cancer subtypes that effectively alleviate the problem of drug resistance in cancer patients. Identification of potential drug-sensitive markers through mathematical and computer methods will be valuable for the determination of anticancer drug targets, improving the development efficiency of new anticancer drugs, and achieving accurate treatment of cancer patients.

Abstract

Since genetic alteration only accounts for 20%–30% in the drug effect-related factors, the role of epigenetic regulation mechanisms in drug response is gradually being valued. However, how epigenetic changes and abnormal gene expression affect the chemotherapy response remains unclear. Therefore, we constructed a variety of mathematical models based on the integrated DNA methylation, gene expression, and anticancer drug response data of cancer cell lines from pan-cancer levels to identify genes whose DNA methylation is associated with drug response and then to assess the impact of epigenetic regulation of gene expression on the sensitivity of anticancer drugs. The innovation of the mathematical models lies in: Linear regression model is followed by logistic regression model, which greatly shortens the calculation time and ensures the reliability of results by considering the covariates. Second, reconstruction of prediction models based on multiple dataset partition methods not only evaluates the model stability but also optimizes the drug-gene pairs. For 368,520 drug-gene pairs with $P < 0.05$ in linear models, 999 candidate pairs with both AUC ≥ 0.8 and $P < 0.05$ were obtained by logistic regression models between drug response and DNA methylation. Then 931 drug-gene pairs with 45 drugs and 491 genes were optimized by model stability assessment. Integrating both DNA methylation and gene expression markedly increased predictive power for 732 drug-gene pairs where 598 drug-gene pairs including 44 drugs and 359 genes were prioritized. Several drug target genes were enriched in the modules of the drug-gene-weighted interaction network. Besides, for cancer driver genes such as *EGFR*, *MET*, and *TET2*, synergistic effects of DNA methylation and gene expression can predict certain anticancer drugs' responses. In summary, we identified potential drug sensitivity-related markers from pan-cancer levels and concluded that synergistic regulation of DNA methylation and gene expression affect anticancer drug response.

Keywords: Drug response, DNA methylation, epigenetics, gene expression, prediction models, pan-cancer

Experimental Biology and Medicine **2021; 246: 1626–1642. DOI: 10.1177/15353702211007766**

Introduction

Recently, with the rapid development of next-generation sequencing and microarray technology, there have been many systematic studies about epigenetic landscape of cancer, which focus on transcriptional disorders caused by DNA methylation, chromatin modification, or remodeling. Abnormal DNA methylation patterns can be important markers for early cancer detection or for monitoring cancer

metastasis.[1,2] Strong evidence suggests that epigenetic regulation plays a key role not only in the early stages of cancer development, but also in disease progression, treatment response, and clinical outcomes.[3,4] DNA methylation levels of certain genes could reflect the clinical characteristics, such as disease subtypes, patient prognosis, and treatment response to cancer.

Epigenetic regulation of gene expression plays an important role in maintaining genomic stability, embryonic development, and tissue differentiation.[5] In most cases, hypermethylation of CpG islands in promoter regions results in gene silence, while hypomethylation leads to transcriptional activation. For many solid tumors, tumor suppressor genes and oncogenes are subjected to hypermethylation and hypomethylation respectively, thus leading to transcriptional repression of tumor suppressor genes, transcriptional activation of oncogenes, cell cycle dysregulation, and ultimately leading to cancer initiation and progression.[6,7] For instance, a study performed by Chen found that the growth- and plasticity-associated protein-43 (*GAP43*) was downregulated in colorectal cancer which was associated with the hypermethylation of *GAP43*.[8]

Though targeted therapy has brought hope to cancer patients, the biggest challenge is the consequent drug resistance.[9] In addition, new drug development requires long preclinical research stages and three clinical phases.[10] Therefore, investigation of drug resistance mechanisms is as important as the development of new anticancer drugs, which helps clinicians select treatment strategies and increases the efficiency of developing new drugs. The identification of genetic biomarkers provides valuable information for revealing the molecular mechanisms of cancer resistance, effectively improving the efficacy of drug therapy for some cancer patients. However, only 20%–30% of inter-individual differences associated with medication side effects or efficacy can be explained by genetic factors.[11,12] Therefore, the epigenetic regulation of protein-encoding genes related to drug absorption, distribution, metabolism, and excretion (ADME) has gradually gained attention. Epigenetic modifications and other factors in complex regulatory networks influence the expression of ADME genes through cis or trans-regulation.[13,14] DNA methylation signatures have been shown to be important markers of drug sensitivity, and aberrant DNA methylation patterns affect the sensitivity of antitumor drugs by regulating the expression levels of genes critical for drug response.[15] Küçük identified 95 silencing genes induced by promoter hypermethylation and reported that natural killer (NK) cells with reduced asparagine synthetase (ASNS) expression were more sensitive to L-asparaginase compared with cells with normal ASNS expression levels.[16] An epigenome-wide DNA methylation study of small cell lung cancer cell lines performed by Krushkal found multiple significant correlations between DNA methylation and chemosensitivity. For instance, increased DNA methylation and decreased expression of *TREX1* were associated with the sensitivity to Aurora kinase inhibitors.[17] Hypomethylation of *ADAM12* leads to increased gene expression of *ADAM12* and doxorubicin resistance for triple-negative breast cancer (TNBC) indicating that *ADAM12* is a potential therapeutic target and its hypomethylation could be a poor outcome biomarker in TNBC.[18]

However, few studies have systematically analyzed the effects of DNA methylation and gene expression regulation on anticancer drug sensitivity. Specifically, the epigenetic mechanisms of anticancer drugs at the experimental stage are rarely noticed, and these drugs are likely to bring new insights into cancer treatments in future. Therefore, here we constructed mathematical models by integrating DNA methylation, gene expression, and anticancer drug response data of cancer cell lines from pan-cancer levels to identify drug response-related biomarkers and then to examine the influence of synergistic action of DNA methylation and gene expression on anticancer drug response.

Materials and methods

An overall flow chart about the process of prediction model construction can be seen in Figure 1. Two main procedures are included in the flow chart including step 1: Identification of marker genes in which DNA methylation levels are related to the response to anticancer drugs and step 2: Prediction of anticancer drug response based on synergistic regulation of DNA methylation and gene expression.

Acquisition and quality control of drug response data of cancer cell lines

The drug response data were derived from the Genomics of Drug Sensitivity in Cancer Project (GDSC).[19] The half maximal inhibitory concentration (IC50) and binarized drug response of 265 drugs for 966 cancer cell lines (GSE68379) were obtained from the GDSC website (http://www.cancerxgene.org). The information about drug ids and names are listed in Table S1. The IC50 indicates the concentration of a drug required to inhibit cell growth by 50%, which is an important indicator for measuring whether a cell is sensitive or resistant to the drug. To ensure the analysis accuracy, the proportion of missing values for each drug was counted, and only drugs with a ratio of missing values less than 50% were retained, thus a total of 224 drugs passed quality control.

Infinium HumanMethylation450K data of cancer cell lines and quantification of DNA methylation levels of genes

Infinium HumanMethylation450K data (GSE68379) of 1001 human cancer cell lines was obtained from the gene expression omnibus (GEO) database. To be consistent with gene expression data, 619 cell lines with matching gene expression data were kept. First, the DNA methylation data at level 1 was converted to the level 3, and the quality control and normalization of DNA methylation data were performed. (1) CG sites with $P > 0.01$ in one or more samples were removed. (2) CG sites with bead count <3 in more than 5% samples were removed. (3) CG sites containing SNPs were removed. (4) CG sites mapped to multiple locations were removed. (5) CG sites on the X and Y chromosomes were removed. (6) Cell lines with a ratio of low-quality probes greater than 5% were removed; 356,276 CpG sites in 619 cell lines passed the quality

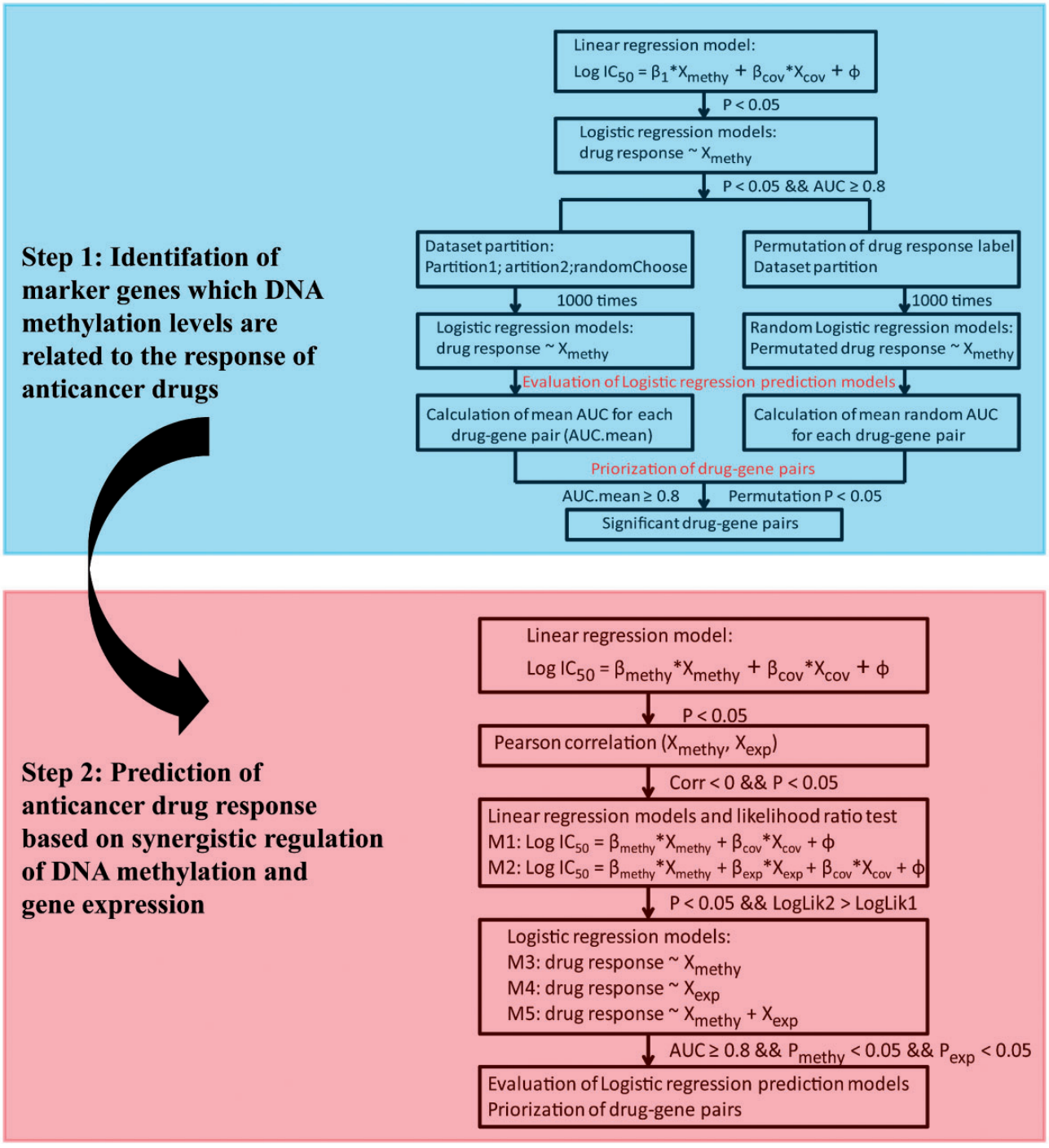

Figure 1. The overall flow chart about the process of prediction model construction. Step 1: Identification of marker genes which DNA methylation levels are related to the response to anticancer drugs. (1) Identification of significant drug-gene pairs based on linear regression models between logarithmic IC50 values and DNA methylation levels and covariates. (2) Construction of logistic regression models between drug response and DNA methylation. (3) For the candidate drug-genes pairs with $P < 0.05$ and AUC ≥ 0.8, three different dataset partition methods, permutation of drug response labels, and model reconstruction were used to evaluate the prediction models and to optimize candidate drug-gene pairs. Step 2: Prediction of anticancer drug response based on synergistic regulation of DNA methylation and gene expression. (1) Identification of significant drug-gene pairs based on linear regression models between logarithmic IC50 values and DNA methylation levels and covariates. (2) Selection of genes with negative regulation by calculating Pearson correlation coefficients between DNA methylation and gene expression for each gene. (3) Construction of linear regression models M1 and M2 and identification of drug-gene pairs whose log-likelihood ratio of model M2 is significantly greater than that of model M1. (4) Construction and comparison of logistic regression models M3 (between drug response and DNA methylation), M4 (between drug response and gene expression), and M5 (between drug response and DNA methylation, gene expression). (5) Priorization of drug-gene pairs with multiple data partition and model reconstruction methods. (A color version of this figure is available in the online journal.)

control. The data were normalized based on the BMIQ method [20]to correct the type 2 deviation. The quality control and normalization process was done by the ChAMP package,[21] and all thresholds were selected with reference to the default settings of the function.

The annotation information of the GPL13534 platform was obtained from the GEO database, and 147,017 CpG sites located on gene promoter regions were selected. The promoter region is defined as 200 bp upstream of the transcription start site (TSS200), 1500 bp upstream of the transcription start site (TSS1500), 5′ untranslated region (5′ UTR), and the first exon (1stExon).[19] The DNA methylation level of a certain gene is defined as the average DNA methylation level of CpG sites mapped to this gene.

Gene expression data of cancer cell lines and quantification of gene expression level

Raw gene expression data (E-MTAB-3610) of 619 cancer cell lines corresponding to DNA methylation samples were downloaded from the Bioinformatics Research Center of the European Molecular Biology Laboratory. The original gene expression data were read by the ReadAffy() function of the Bioconductor package affy, and the expression level was corrected and normalized by a robust multi-array average (RMA) algorithm and quantile normalization method.[22,23] Based on the annotation data of GPL13667 platform, the probes were aligned to transcripts and the expression value of a transcript is defined as the mean expression value of the probes mapped to the transcript. In order to quantify the expression level of protein coding genes, 29,751 mRNA transcripts were mapped to 19,198 genes. To analyze the regulatory relationship between gene expression and DNA methylation, we got the overlapped genes between expression and methylation data. Finally, the expression profile of 16,348 protein coding genes in 619 cancer cell lines was kept.

Construction of linear regression model between logarithmic IC50 and DNA methylation

For anticancer drugs and genes passing the control quality, we constructed linear regression models between logarithmic IC50 values and DNA methylation. Tissue types, microsatellite instability, medium of cell line, and growth property were treated as covariates which have been reported to affect drug response in previous studies.[19] The linear regression model is shown in equation (1) where logIC50 is an index for measuring drug response, X_{METHY} and X_{COV} refer to DNA methylation level and the covariates, φ is the error term, β_{METHY} and β_{COV} represent coefficients of DNA methylation and covariates, respectively. The type 2 analysis of variance (ANOVA) was used to assess whether the DNA methylation level of a gene was significantly correlated with the drug response, where $P < 0.05$ indicated a significant association. The FDR method was used to adjust the *P*-values of ANOVA by selecting FDR < 0.05 as the threshold.

$$\log IC50 = \beta_{METHY} * X_{METHY} + \beta_{COV} * X_{COV} + \varphi \quad (1)$$

Construction of logistic regression model between binarized drug response and DNA methylation

For each drug-gene pair with $P < 0.05$ in the linear regression model, the glm()function in R was used to construct the logistic regression model between binarized drug response and DNA methylation. The prediction() and performance() functions of the ROCR package were used to predict the results of logistic regression models and to compare the predicted values and observations. Then the area under curve (AUC) of receiver operating characteristic (ROC) analysis and the *P* value were used to estimate the prediction accuracy. The drug-gene pairs with $AUC \geq 0.8$ and $P < 0.05$ were identified as candidates for subsequent analysis. Three groups of drug-gene pairs are defined according to predictive capacity for drug response: (1) the high AUC group with $AUC \geq 0.8$; (2) the low AUC group with $AUC < 0.7$; (3) the moderate AUC group with $0.7 \leq AUC < 0.8$. For each of the candidate drug-gene pairs with $AUC \geq 0.8$ and $P < 0.05$, DNA methylation measurement β was divided into 10,000 aliquots, each time adding 0.0001 as a critical value, and the sensitivity and specificity corresponding to the critical value were calculated. The best classification cut-off was defined as the β value corresponding to the maximum Youden index.[24]

Evaluation of the stability of logistic regression prediction models based on multiple dataset partition methods and stochastic models

Since the number of sensitive cell lines for most drugs is generally greatly less than that of drug-resistant cell lines, three data partition methods were used to reconstruct the logistic regression prediction model to reduce the deviation possibly caused by the cell line number difference. Then the real logistic regression models based on different partition methods were compared with the random ones to verify the stability of prediction models and to prioritize important drug-gene pairs.

First, for each drug-gene pair, 50% of drug response and DNA methylation data were randomly treated as the training set to construct a logistic regression model, and the remaining 50% was used as a test set to calculate the AUC value. The processes of data set partition and model construction were repeated 1000 times, and the average AUC value was calculated as a comprehensive evaluation index. For each anticancer drug, the response label (sensitive or resistant) was randomly perturbed. Then *n* cell lines were selected as sensitive cell lines and the remaining cell lines were labeled as resistant cell lines where *n* is the number of sensitive cell lines in real cases. Similar to the above methods, the processes of data set partition and random model construction were repeated 1000 times.

Second, for each drug-gene pair, 70% of data was randomly selected as training sets to construct a logistic regression model, and the remaining 30% of the data was used as a test set. The rest of the calculation process is the same as the first data partition method. The random model was constructed similarly to that described previously, while the difference is that 70% of data was randomly selected

as the training set to construct a logistic regression model, and the remaining 30% of data was used as a test set for model prediction.

Third, for each drug-gene pair, the same number of drug-resistant cell lines and sensitive cell lines was randomly selected to construct a logistic regression model. This process was also repeated 1000 times. For each anticancer drug, the response label was randomly perturbed, and then n cell lines were randomly labeled as sensitive cell lines, and the remaining ones were labeled as resistant cell lines where n was the number of sensitive cell lines in real cases. The process of data permutation and random model construction was repeated for 1000 times.

We counted the number of AUC values in the random models greater than the average real AUC value, and defined a P value by dividing this number by 1000. A real prediction model is considered to be significant if $P < 0.05$ and the average AUC of the real model ≥ 0.8.

In this research, the linear regression model was used followed by the logistic regression model with the following reasons. First, studies have shown that tissue type, cell line culture medium, growth property, and microsatellite instability are the main factors affecting drug response. Therefore, to explore the relationship between DNA methylation and drug response, these four factors were used as covariates to correct the linear regression model and to ensure the reliability of the screened significant drug-gene pairs. Second, as is known, for all of the 224 drugs and 18,215 genes, there are 4,080,160 drug-gene pairs, which is a very time-consuming process to directly construct logistic regression models and to reconstruct both real prediction models and stochastic models by using three data set partition methods for 1000 times. Therefore, before constructing and evaluating the logistic prediction models, screening the drug gene pairs based on the linear models can effectively shorten the calculation time and improve the calculation efficiency.

Identification of the negative regulation between DNA methylation and gene expression in the promoter region

For each of the genes that DNA methylation is significantly related to anticancer drug response in linear regression models ($P < 0.05$), the Pearson correlation test between gene expression and DNA methylation in the promoter regions was performed. A gene with correlation coefficient < 0 and $P < 0.05$ was considered to have a negative regulatory relationship between DNA methylation and gene expression.

Determination of the effect of gene expression on anticancer drug response based on likelihood ratio test

For the drug-gene pairs where DNA methylation is significantly associated with drug response in the linear regression models and DNA methylation is negatively related with gene expression, two linear regression models M1 and M2 were constructed. M1 is a model between logarithmic IC50 and DNA methylation and covariates (equation (2)) and M2 is a model by adding gene expression information to M1 model (equation (3)). X_{METHY} and X_{EXP} represent DNA methylation and gene expression, β_{METHY} and β_{EXP} represent coefficients of DNA methylation and gene expression in a linear regression model, X_{COV} and β_{COV} represent the covariates and their regression coefficients. The covariates included tissue type, microsatellite instability, medium, and growth characteristics. The likelihood ratio test (LRT) was used to screen drug-gene pairs that the log likelihood value of M2 model was significantly greater than M1 model. As shown in equation (4), L2 and L1 are the maximum likelihood values of M2 and M1, respectively. LR approximately obeys Chi-square distribution.

$$M1 : \log IC50 = \beta_{METHY} * X_{METHY} + \beta_{COV} * X_{COV} + \varphi \quad (2)$$

$$M2 : \log IC50 = \beta_{METHY} * X_{METHY} + \beta_{EXP} * X_{EXP} + \beta_{COV} * X_{COV} + \varphi \quad (3)$$

$$LR = 2 * (\ln L2 - \ln L1) \quad (4)$$

Construction of logistic regression model between drug response and DNA methylation and gene expression

Three logistic regression models M3, M4, M5 were constructed for each of the significant drug-gene pairs obtained in LRTs. The dependent variables of the three types of models were binarized drug response, and independent variables of M3, M4, and M5 were DNA methylation, gene expression, and both of the two variables, respectively. The drug-gene pairs with AUC ≥ 0.8 and $P < 0.05$ in logistic regression models were identified as candidates.

Prioritization of drug-genes pairs that synergistic regulation of DNA methylation and gene expression has effect on drug response

For candidate drug-gene pairs obtained from M5, three data partition methods described previously were utilized to re-construct logistic regression models for 1000 times and to calculate the average AUC values. Then the anticancer drug response labels were randomly perturbed and the random model construction was repeated for 1000 times. We then compared the average AUC values of real models with those of random models and defined the P value by dividing the number of random AUC values greater than the real average AUC by 1000. A drug-gene pair was optimized if the real average AUC ≥ 0.8 and $P < 0.05$ for all three data partition methods.

Construction of drug-gene-weighted interaction network

A drug-gene-weighted interaction (DGWI) network was constructed using Cytoscape software for optimized

drug-gene pairs. Nodes in the DGWI network are drugs and genes associated with drug response. Relations between drugs and genes were connected with solid lines, while genes sharing common drugs and drugs sharing common genes were connected with dashed lines. The number of common genes or common drugs was treated as edge weight. The modules of DGWI network were mined by plugin MCODE of Cytoscape by setting these following thresholds: (1) the degree threshold was set to 5, (2) haircut was used as the clustering method, (3) the node score threshold was set to 0.2, (4) the K-Core was set to 5, and (5) the maximum depth was set to 100.

Results

DNA methylation turns to be closely related to drug response across pan-cancers by linear regression models

A total of 4,080,160 linear regression models were generated for 224 anticancer drugs and 18,215 genes. DNA methylation is significantly related to drug response in 368,519 drug-gene pairs ($P < 0.05$). The number of remarkable drug-gene pairs reduced to 13,403 if FDR < 0.05 including 6013 negative pairs with $\beta_{METHY} < 0$ and 7390 positive pairs with $\beta_{METHY} > 0$. The negative relation suggests that cell lines with higher methylation level for a specified gene are prone to be more sensitive to the anticancer drug, while the positive relation has the opposite trend. The 13,403 drug-gene pairs include 113 drugs and 6958 genes where each drug has 1–1585 important genes (median: 10 and mean: 119). For NVP-BEZ235, AZD8055, AKT inhibitor VIII, RDEA119, AZD6244, BX-795, KU-55933, DNA methylation levels of most genes were positively correlated with the drug response, while for Elesclomol, EHT-1864, Bleomycin, SN-38, Nutlin-3a, PI-103, and QL-X-138, DNA methylation levels of most genes are negatively correlated with the drug response (Figure S1).

The number and percent of cell lines with hypermethylation, hypomethylation and moderate methylation levels were counted for 16 genes involved in the top 20 drug-gene pairs ordered by descending FDR (Figure 2(a)). For a certain gene, hypermethylated cell lines were defined as cell lines with $\beta \geq 0.75$, hypomethylated cell lines were defined as cell lines with $\beta \leq 0.25$, and other cell lines are defined as moderately methylated cell lines. Hierarchical clustering was performed based on the DNA methylation level of 16 genes across cancer cell lines (Figure 2(b)). For six of the top 20 drug-gene pairs with high proportion of both hypermethylated and hypomethylated cell lines, hypermethylated cell lines had higher IC50 than hypomethylated ones (Figure 2(c)) indicating that high methylation levels for these genes might be drug resistance factors.

Candidate drug-gene pairs were identified by constructing DNA methylation-based logistic regression model

A total of 363,034 significant drug-gene pairs in linear regression models were kept after removing drugs including JQ1, FK866, and Bicalutamide due to the small number of sensitive cell lines. Then we constructed logistic regression models between DNA methylation and binarized drug response. Drug-gene pairs in low AUC group, moderate AUC group, and high AUC group account for 96.7% (350,213), 3.26% (11,822), and 0.28% (999), respectively (Figure 3). Nine hundred and ninety-nine drug-gene pairs (with AUC ≥ 0.8 and $P < 0.05$) containing 46 anticancer drugs and 514 genes were selected as candidate pairs indicating DNA methylation might have potential effect on anticancer drug response (Table S2).

There are 23 drug-gene pairs with AUC ≥ 0.9, including 5 drugs and 23 genes (Table S3). The AUC of SNX-2112 and *FERMT3* is 0.945 ranking first at the list. For the drug-gene pair SNX-2112 and *FERMT3*, the optimal threshold of DNA methylation is 0.706 and the sensitivity and specificity are 1 and 0.826, respectively (Figure 4(a)). The DNA methylation level of *FERMT3* in cell lines sensitive to SNX-2112 is significantly less than that of drug-resistant cell lines (Figure 4(b)). DNA methylation of *LOC100130776* well predicts SNX-2112's response with AUC of 0.924. The optimal methylation cut-off, sensitivity, and specificity are 0.667, 0.941, and 0.852, respectively (Figure 4(c)). Cell lines sensitive to SNX-2112 have higher DNA methylation levels of *LOC100130776* than drug-resistant cell lines (Figure 4(d)).

Multiple data partition and evaluation methods prove that DNA methylation steadily predicts the response to anticancer drugs

For 999 candidate pairs identified by logistic regression models, the percentage of drug-gene pairs with AUC ≥ 0.8 and $P < 0.05$ calculated by three data partition methods are 96.6% (965/999), 94.9% (948/999), and 98.7% (986/999), respectively (Tables S4 to S6). The overlapped drug-gene pairs among three different data partition methods account for 93.2% (931/999) of candidates indicating that DNA methylation-based logistic regression model is robust (Figure 5(a)). The 931 drug-gene pairs contain 45 drugs and 491 genes and each drug has 21 significant genes on average.

To evaluate the reliability of prediction models, we compared the performance between real and random logistic regression models constructed by three different data partition methods (Figure 5(b) to (d)). The AUC values of random models range from 0.5 to 0.6 which are significantly less than those of real models ($P = 4.85\text{E-}165$ by the Wilcoxon signed rank test). These results suggest that the relation between DNA methylation and drug sensitivity across pan-cancer levels identified in this study is not a random association.

DNA methylation-mediated anticancer drug response-related genes are involved in cancer development and enriched with drug targets

In order to explain the effect of DNA methylation on anticancer drug response from functional levels, we performed an enrichment analysis for 491 genes contained in the 931 optimized drug-gene pairs by enrichR.[25] These genes are significantly annotated to multiple biological pathways

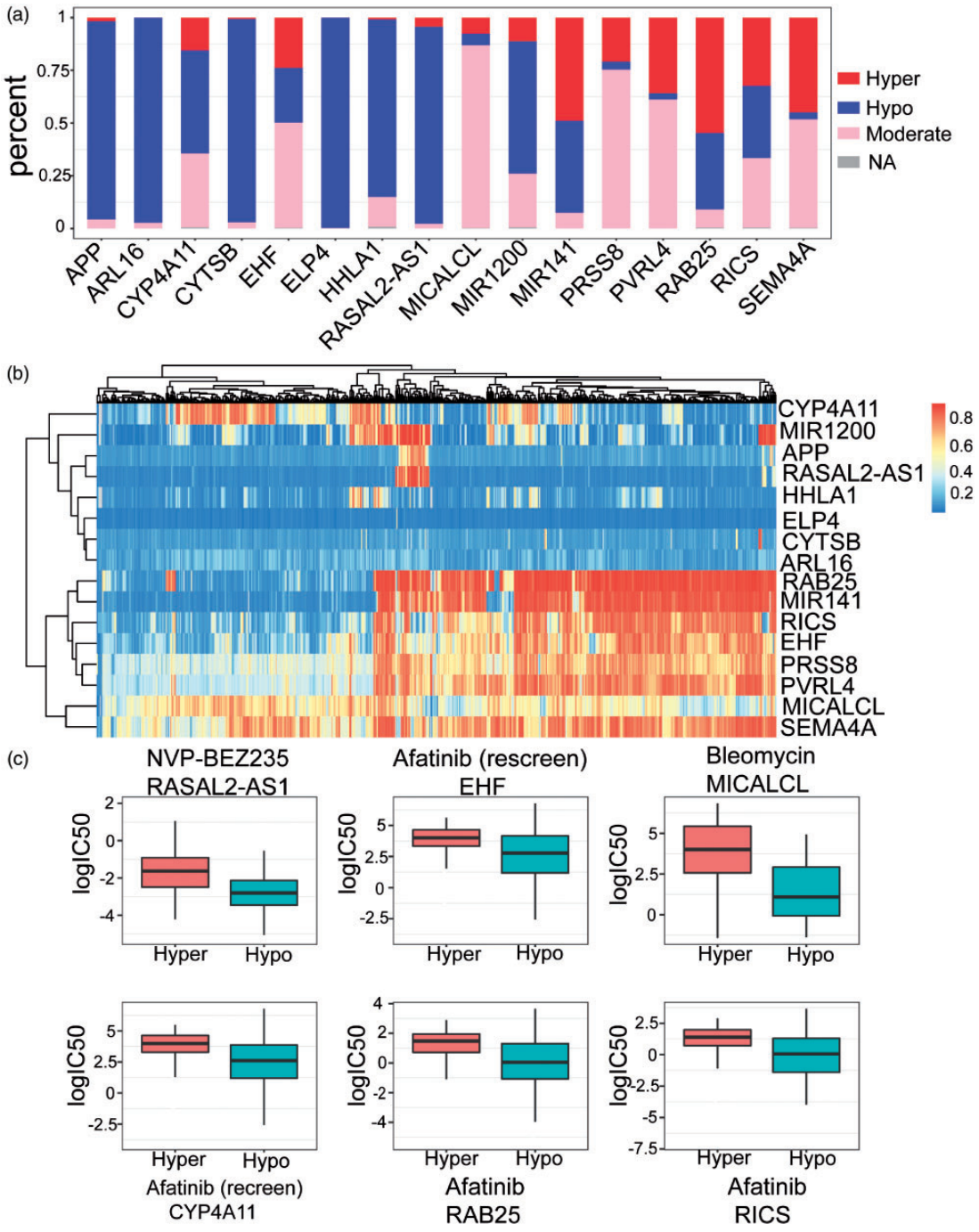

Figure 2. DNA methylation pattern of significant genes in top 20 drug-gene pairs and its influence on drug response. (a) The percent of hypermethylated, moderate, and hypomethylated cell lines; (b) Hierarchical clustering of 619 cancer cell lines and genes based on DNA methylation level of 16 genes; (c) Comparison of logarithmic IC50 values between hypermethylated and hypomethylated cell lines for a certain drug and gene pair. (A color version of this figure is available in the online journal.)

directly or indirectly related to cancer, such as the proteoglycans in cancer, pathways in cancer, PI3K-Akt signaling pathway, and Rap1 signaling pathway (Figure 6(a)). It is worth noting that four drugs namely BX912, OSI-027, PI-103, and PIK-93 target PI3K/MTOR signaling pathways which may explain the reason why DNA methylation of certain genes can predict the response to anticancer drugs. The significantly annotated biological processes, molecular functions, and cellular components are displayed in Figures S2 to S4, respectively. Then enrichmentMap, a plug-in of Cytoscape [26] was used to construct a network for the enriched pathways, biological processes, molecular functions, and cellular components based on enrichment degree and the number of genes shared by these biological terms. Some biological pathways and GO terms are closely related, such as hematopoietic lineage pathways and focal adhesion pathways, while other biological pathways are relatively independent, such as RAP1 signaling pathway and proteoglycans in cancer (Figure 6(b)).

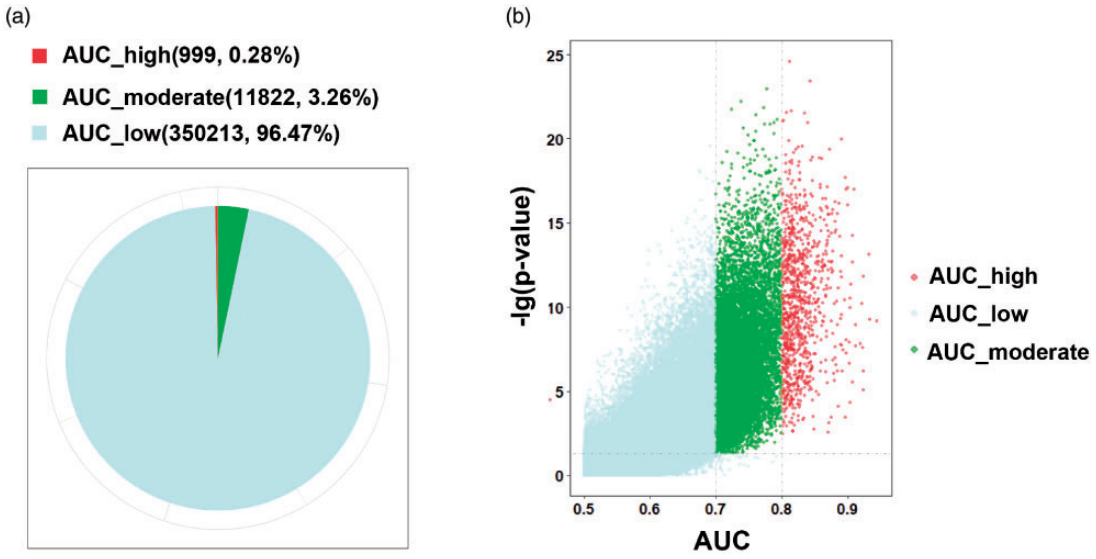

Figure 3. Distribution of AUC values of drug-gene pairs in the first logistic regression models. (a) The number and percentage of drug-gene pairs with different prediction performance; (b) Volcano plot of drug-gene pairs in logistic models showing the relationship between AUC and *P* value. Drug-gene pairs with low AUC (<0.7), moderate AUC (0.7–0.8), and high AUC (≥0.8) are represented by light cyan, green and red boxes. (A color version of this figure is available in the online journal.)

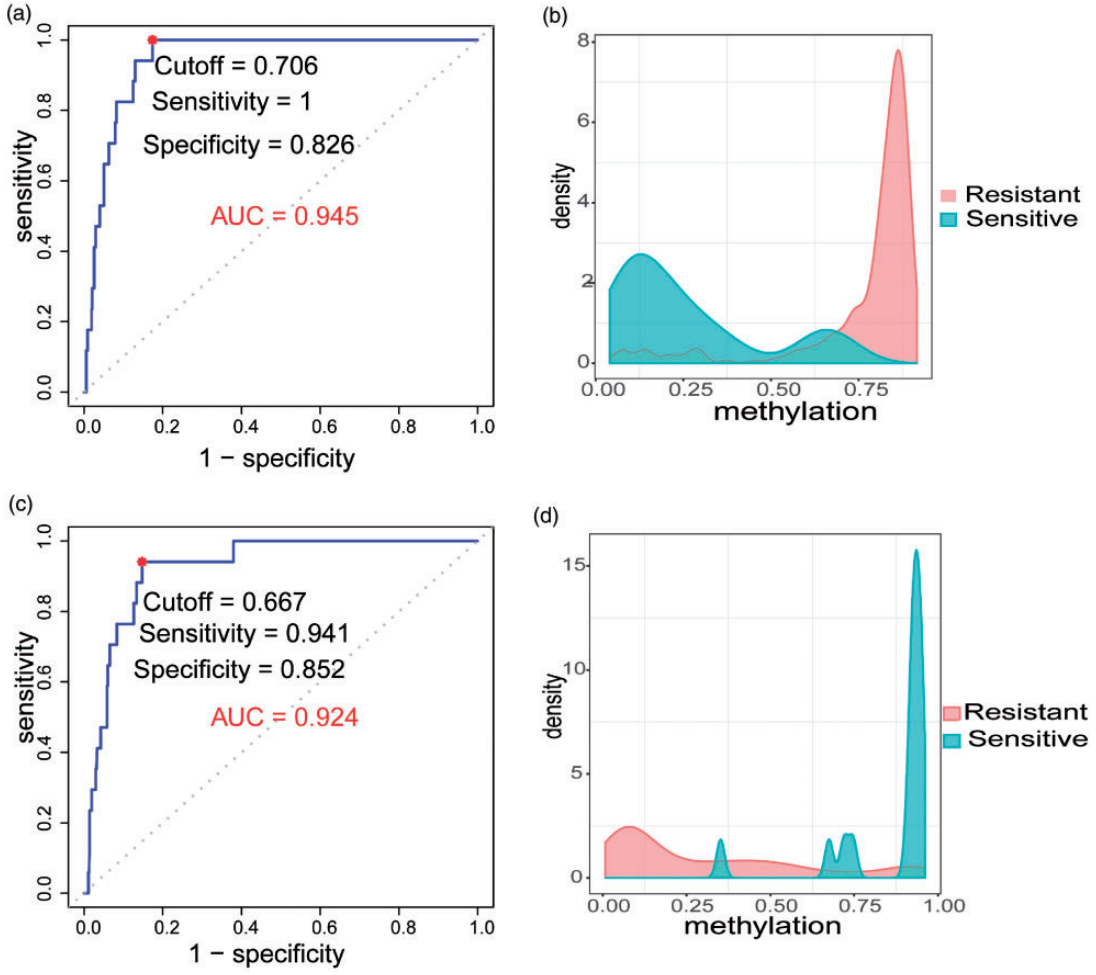

Figure 4. Evaluation of performance of DNA methylation for predicting drug response. (a) The ROC plot to show the performance for predicting SNX-2112's drug response based on *FERMT3*'s DNA methylation values; (b) Density plot for comparing the DNA methylation distribution of *FERMT3* between cell lines that are sensitive to SNX-2112 and those resistant ones; (c) The ROC plot to show the performance for predicting SNX-2112's drug response based on *AGAP2-AS1*'s DNA methylation values; (d) Density plot for comparing the DNA methylation distribution of *AGAP2-AS1* between cell lines that are sensitive to SNX-2112 and those resistant ones. (A color version of this figure is available in the online journal.)

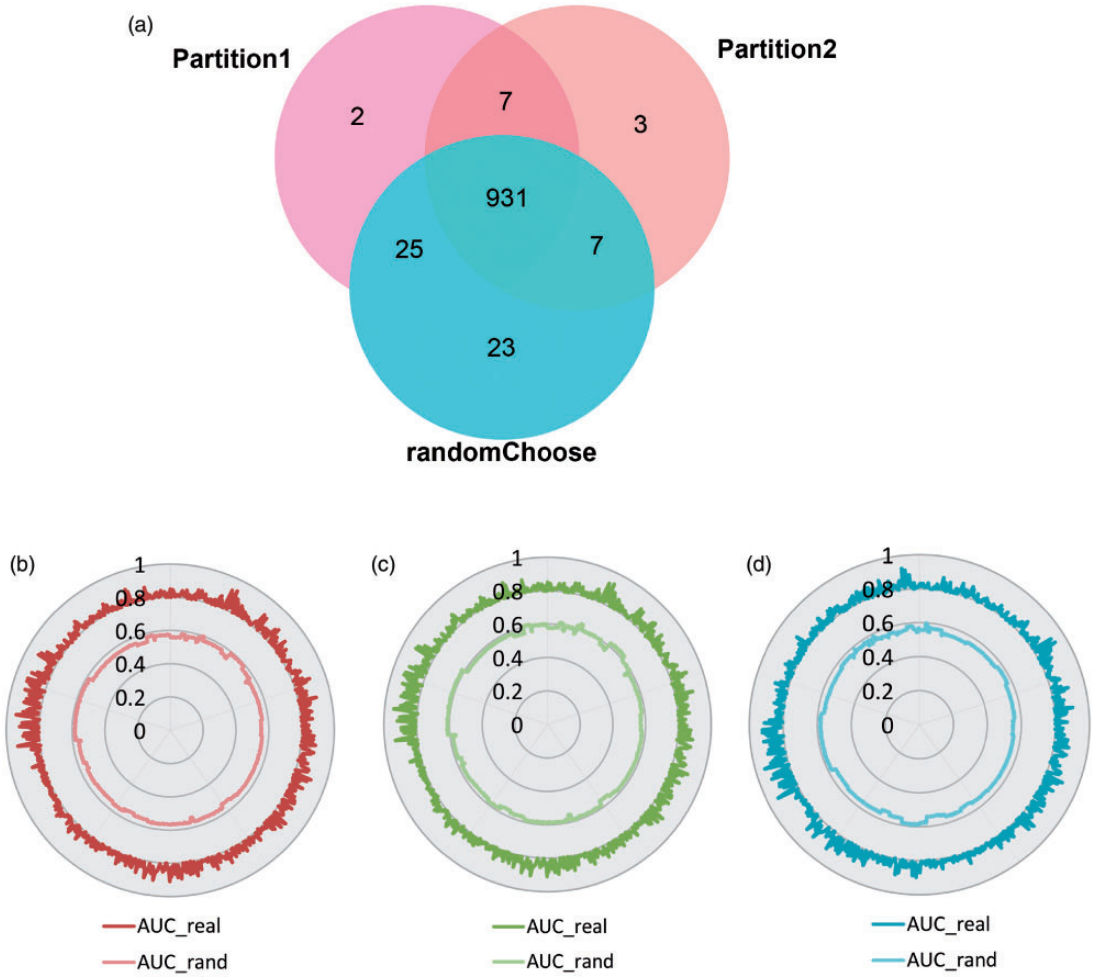

Figure 5. Priorization of drug-gene pairs by prediction models based on three data partition methods. (a) Venn plot of significant drug-gene pairs obtained by three different data partition methods; (b) (c) (d) Comparison of prediction performance between real and random Logistic regression models based on three data partition methods. Partition1 and Partiton2 represent the partition methods using 50% and 70% data as training set, respectively, while randomChoose means the partition method randomly selecting the same number of resistant and sensitive cell lines. (A color version of this figure is available in the online journal.)

To explain the potential mechanisms by which the preferred 931 drug-gene pairs influence drug response, 491 genes identified in this study were compared with 2242 FDA-approved human drug targets obtained from the DrugBank database. Forty-nine of the 491 genes related with response to anticancer drugs are FDA-approved drug targets, three of which are known oncogenes, namely *EGFR*, *ERBB2*, and *MET*.[27] The drug response to WZ3105 and SNX-2112 was predicted by DNA methylation of *EGFR* with an average AUC of 0.81 (Figure 7(a)). Although there is no direct evidence suggesting a relationship between DNA methylation of *EGFR* and these two anticancer drugs, based on the role of *EGFR* in cancer and the relationship between DNA methylation of *EGFR* and colorectal cancer resistance reported in the literature, we speculate *EGFR* is likely to be a potential predictor of drug response to WZ3105 and SNX-2112. The DNA methylation of *ERBB2* is significantly correlated with the drug response to PI-103 and Tivozanib with average AUC values of 0.84 and 0.81, respectively (Figure 7(b)). Cancer cell lines with higher DNA methylation levels of *ERBB2* tend to be more sensitive to the anticancer drug, suggesting that *ERBB2* may be a sensitive marker of Tivozanib. DNA methylation of *MET* can well predict the response to a variety of anticancer drugs, such as BIX02189 targeting the ERK MAPK signaling pathway, PIK-93 targeting the PI3K/MTOR signaling pathway, THZ-2–101-1 and NPK76-II-72–1 targeting the cell cycle pathway, and TL-1–85 targeting other kinase pathways (Figure 7(c)).

Synergistic regulation of DNA methylation and gene expression affects sensitivity of anticancer drugs

DNA methylation of the promoter regions of 8909 genes was negatively related with gene expression with the average Pearson correlation coefficient of −0.28 (Figure S5). Eight thousand nine hundred and nine genes are contained in 180,073 drug-genes pairs where DNA methylation is significantly related to the response to anticancer drugs. The expression levels of 6937 genes contained in 27,161 drug-gene pairs can improve the predictive power of the M1 model based on the likelihood ratio tests.

The number of drug-gene pairs with AUC $\geq$ 0.8 and $P < 0.05$ in DNA methylation-based M3 model, gene expression-based M4 model, and M5 model combining

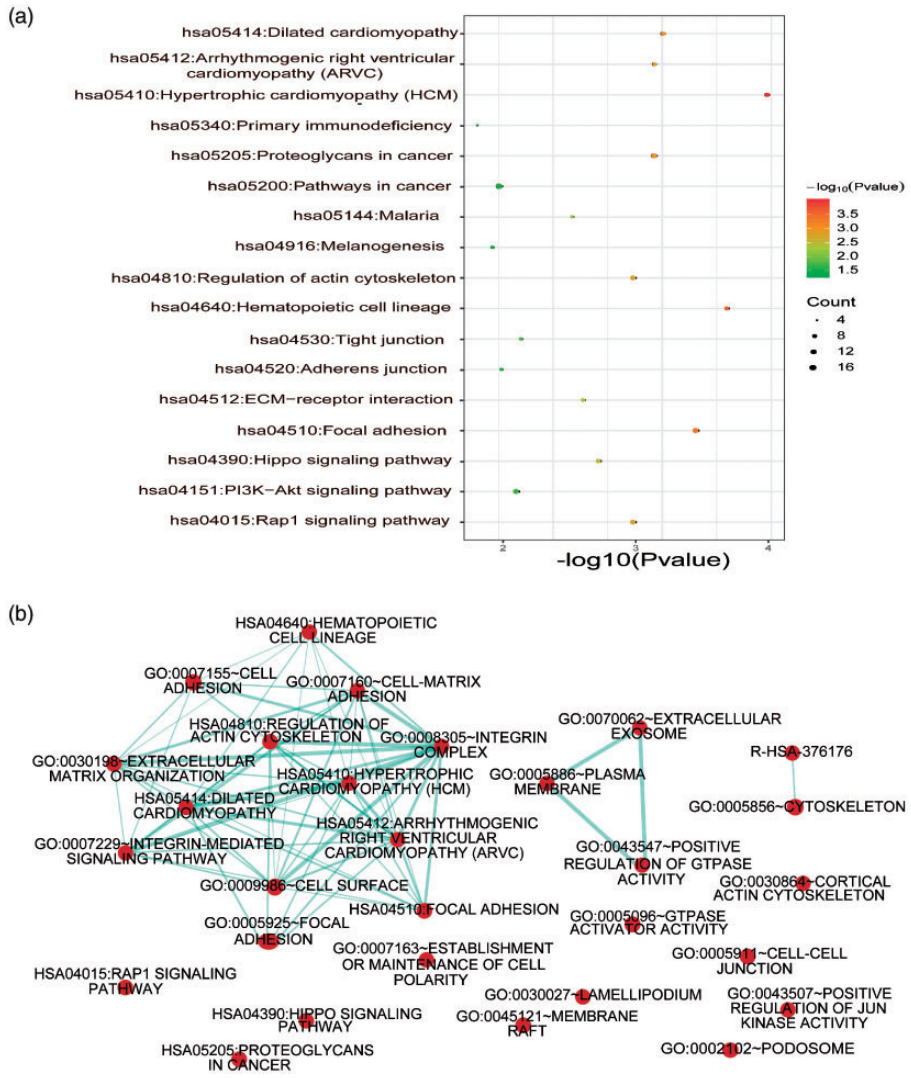

Figure 6. Enrichment analysis result of 491 genes in 931 drug-gene pairs. (a) Significant biological pathways to which 491 genes in 931 drug-gene pairs are enriched. X-axis and y-axis represent the negative logarithmic *P* value of enrichment analysis and significant pathways. The redder a node is, the higher degree of enrichment the pathway has. The node size is in proportion to the number of overlapped genes between interesting genes and those in a certain pathway. (b) The network of significantly enriched biological pathways and GO terms. A pathway or a GO term is represented by a node. A line connecting two nodes means there are common genes between significant terms. The width of a line is proportion to the size of common genes. (A color version of this figure is available in the online journal.)

DNA methylation and gene expression is 372, 1299, and 732, respectively (Tables S7 to S9). M3 model shares 148 and 156 drug-gene pairs with M4 and M5, respectively. M4 model shares 367 drug-gene pairs with M5, and the three models have 88 overlapped drug-gene pairs in common (Figure 8(a)). For 68 drug-gene pairs in M5 model, DNA methylation makes more contribution to predicting drug response than gene expression, and for 279 drug-gene pairs, gene expression makes more contribution to predicting drug response than DNA methylation. In addition, for 297 drug-gene pairs (Table S10), integration of DNA methylation and gene expression observably improve the performance of prediction compared with DNA methylation-based M3 models or gene expression-based M4 models. It is worth noting that 7 of the 297 drug-gene pairs have AUC greater than 0.8 in M5 models, while AUC values are less than 0.7 in both M3 and M4 models indicating the important influence of negative regulation between DNA methylation and gene expression on the response to anticancer drugs (Figure 8(b) to (i)).

There are 81.7% (598/732) drug-gene pairs with an average AUC ≥ 0.8 in all of the three data partition methods (Table S11). And the predictive power of the real model is significantly greater than the random model ($P < 0.05$). As shown in Figure 9(a), Partition1 and Partition2 represent the methods using 50% and 70% data as training set respectively, and randomChoose represents the method selecting the same number of sensitive and resistant cell lines. The real prediction models have higher AUC values than random models for three different data partitioning methods (Figure 9(b) to (d)).

DNA methylation and gene expression regulation affect drug response revealed by DGWI network

A DGWI network was constructed for the 598 optimized drug-gene pairs. The DGWI network contains 403 nodes

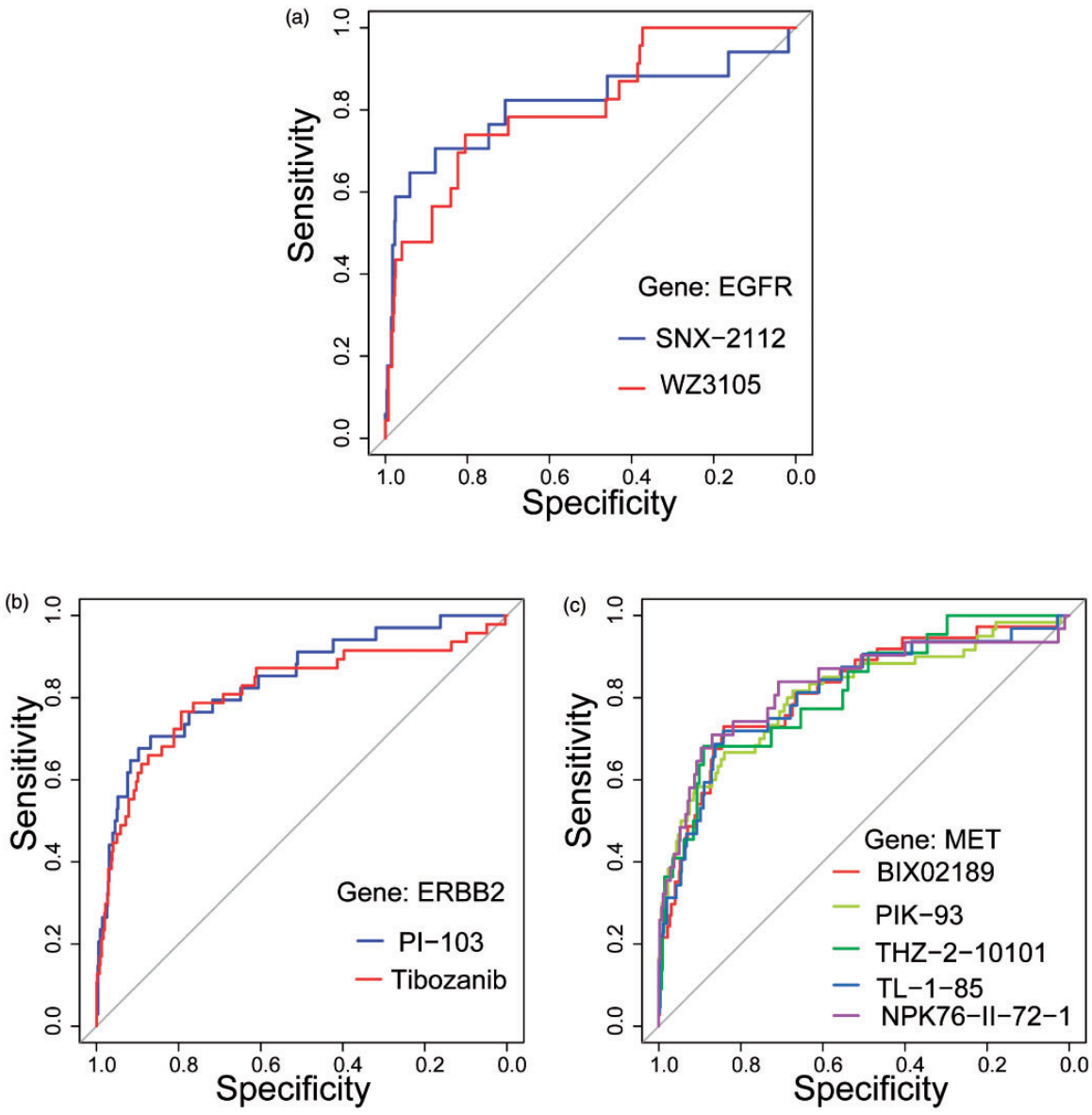

Figure 7. The performance for prediction of anticancer drug response based on DNA methylation levels of three driver genes. (a) ROC curves for drug response prediction based on DNA methylation of *EGFR*; (b) ROC curves for drug response prediction based on DNA methylation of *ERBB2*; (c) ROC curves for drug response prediction based on DNA methylation of *MET*. X-axis and y-axis refer to specificity and sensitivity, respectively. (A color version of this figure is available in the online journal.)

including 44 anticancer drugs, 359 drug response-related genes and 10,003 edges including 598 edges between drugs and genes, 269 edges among drugs and 9136 edges among genes. The network has a clustering coefficient of 0.823, a network diameter of 4, and an average of 49.6 neighbor nodes for each node. Seven clusters (Table S12) were obtained by network module analysis with scores ranging from 6.0 to 66.6, number of nodes ranging from 6 to 117, and number of edges ranging from 15 to 3870.

Module 1 contains 2 drugs namely AT-751 and NPK76-II-72–1, targeting cell cycle, and 115 genes including 16 approved drug targets and 3 known cancer driver genes namely *EGFR*, *MET*, and *ARID2* (Figure 10(a)). Protein coding products of *MCM6* and *CDC7* are involved in the cell cycle pathway [28,29]affecting DNA synthesis. Our study showed that DNA methylation and gene expression levels of these two genes could well predict the drug response to NPK76-II-72–1 with average AUC values of 0.87 and 0.8, respectively, suggesting that these two genes may be potential drug targets. In addition, the cancer driver genes *EGFR* and *MET* play a central role in the cytokine receptor interaction process of the cancer pathway (hsa05200), and *RASSF5* and *RAC2* are involved in the downstream MAPK signaling pathway which affects cancer development (Figure 10(b)). *EGFR*, *MET*, and *RAC2* are drug target genes in which *EGFR* and *MET* predict the drug response to NPK76-II-72–1 with an average AUC $\geq$ 0.85; *RASSF5* and *RAC2* predict the drug response to AT-7519 with an average AUC $\geq$ 0.85. In summary, multiple genes in Module 1 are closely related to cancer, and synergy of DNA

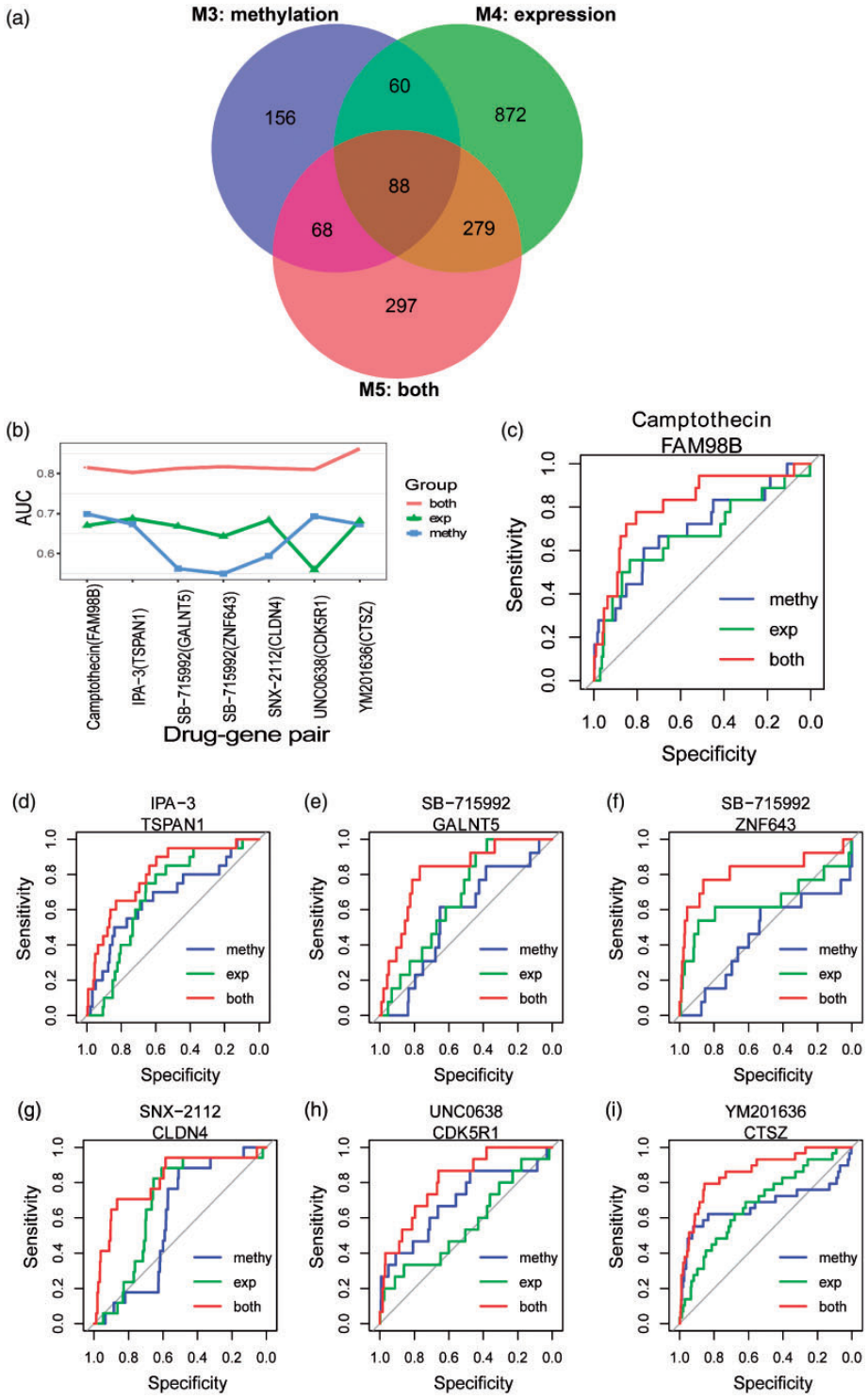

Figure 8. Synergistic effects of DNA methylation and gene expression enhance predictive performance of anticancer drugs' response. (a) Significant drug-gene pairs with both AUC ≥ 0.8 and $P < 0.05$ in logistic regression models M3, M4, and M5. Results of three models are marked with blue, green, and red. (b) Comparison of AUC values of three models for seven drug-gene pairs; (c) Drug response to Camptothecin predicted by *FAM98B*; (d) Drug response to IPA-3 predicted by *TSPAN1*; (e) Drug response to SB-715992 predicted by *GALNT5*; (f) Drug response to SB-71599 predicted by *ZNF643*; (g) Drug response to SNX-2112 predicted by *CLDN4*; (h) Drug response to UNC0638 predicted by *CDK5R1*; (i) Drug response to YM201636 predicted by *CTSZ*. (A color version of this figure is available in the online journal.)

methylation and expression of these genes affects drug response, indicating that these genes may be a hot spot for anticancer drug researches in future.

Module 2 contains five drugs and 133 genes, of which SNX-2112 is a drug targeting protein stability and degradation pathways. Heat shock protein 90 (HSP90) is the main target of SNX-2112. Nineteen drug target genes and one cancer-driver gene namely *TET2* are included in Module 2 (Figure 10(c)). SNX-2112, a HSP90 inhibitor, has been reported to induce differentiation and apoptosis of

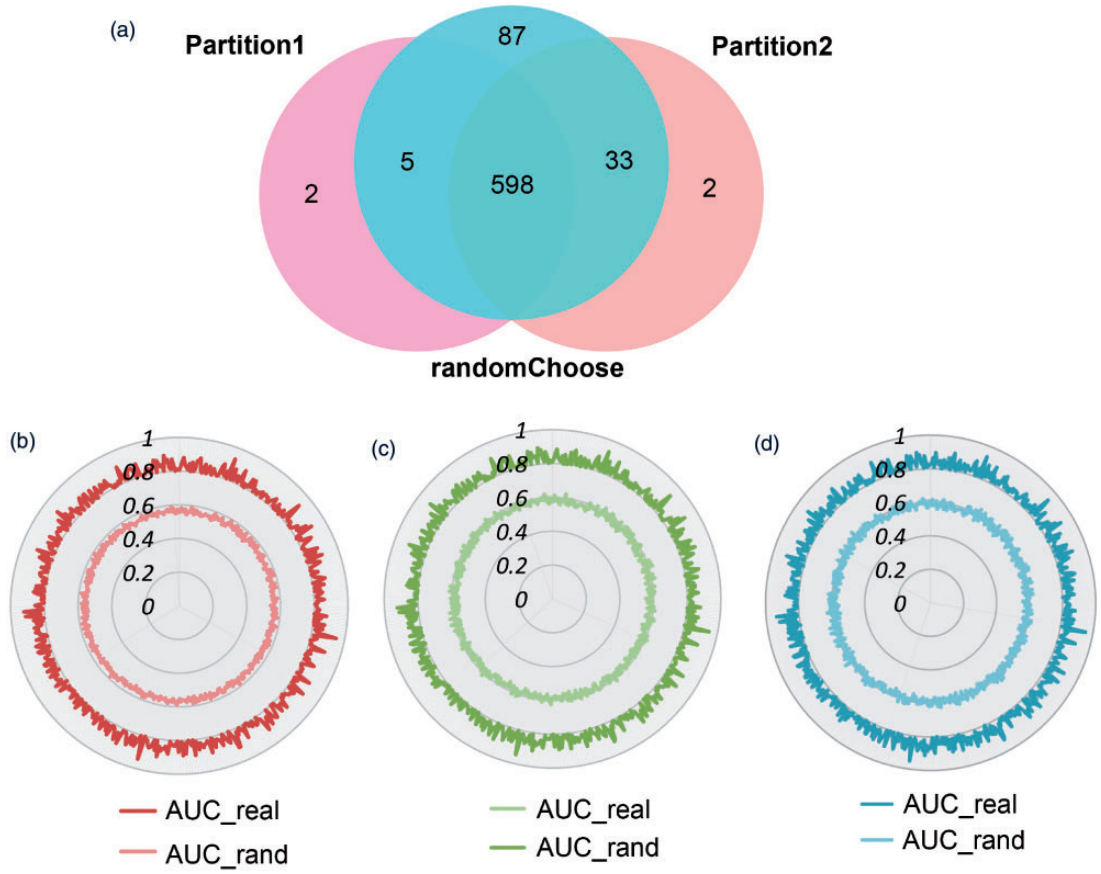

Figure 9. Comparison of prediction results of logistic regression models based on three different data partition methods. (a) Overlap of significant drug-gene pairs obtained by three different data partition methods; (b) Comparison of real and random AUC values based on the first data partition method; (c) Comparison of real and random AUC values based on the second data partition method; (d) Comparison of real and random AUC values based on the third data partition method. Partition1 and Partition2 represent the methods using 50% and 70% data as training set, respectively, while randomChoose represents the method using the same number of sensitive and resistant cell lines. (A color version of this figure is available in the online journal.)

human acute myeloid leukemia cells by regulating Akt/NF-kB signaling pathway, and thus may be a drug for treatment of acute myeloid leukemia and other malignant blood diseases.[30,31] Protein coding products of some genes play an important role in the upstream of PI3K-AKT signaling pathway, and HSP90, a target of SNX-2112, is also involved in this pathway (Figure 10(d)). Therefore, the effect of DNA methylation mediated gene expression regulation of these genes on the drug response to SNX-2112 can be explained to some extent. *TET2* is a known tumor suppressor gene oxidizing 5mC to 5hmC, which promotes DNA demethylation, participates in chromatin-modifying pathways, and determines cell fate.[32] *TET2* is closely related to the occurrence, chemotherapy resistance, and poor prognosis of acute myeloid leukemia.[33] Besides, decreased expression of *TET2* and increased levels of 5hmC have become important markers for a variety of solid tumors and increased expression of *TET2* inhibits cancer cell proliferation and metastasis.[34] In our research, the regulation between DNA methylation and gene expression of *TET2* well predicts the drug response to SNX-2112 with an average AUC of 0.93 indicating that *TET2* is not only a tumor suppressor gene, but also has the potential to be an important marker for anticancer drugs.

Discussion

The drug response to NVP-BEZ235 was correlated with the DNA methylation levels of 1585 genes based on linear regression models. NVP-BEZ235 is a dual inhibitor of phosphatidylinositol kinase (PI3K) and mammalian target of rapamycin (TOR) that inhibits both mTOR1 and mTOR2. A number of clinical studies have evaluated the NVP-BEZ235's therapeutic effect on advanced solid tumors, metastatic breast cancer, and various types of leukemia. Although there is no direct evidence demonstrating the relationship between DNA methylation and NVP-BEZ235, NVP-BEZ235 is a drug targeting PI3K-AKT signaling pathway. PI3K-AKT signaling pathway regulates breast cancer genome by controlling H3K4 methylation, and the subcellular localization of KDM5A, a H3K4 demethylase, may be a pharmacodynamic marker of PI3K-AKT inhibitor.[35] The protein-coding products of some genes related with NVP-BEZ235's response are involved in the PI3K-AKT signaling pathway (Figure S6), suggesting that DNA methylation patterns of these genes are likely to be related to the PI3K-AKT signaling pathway, and then affect drug response. In addition, these genes are also annotated into multiple cancer-associated or drug-related signaling

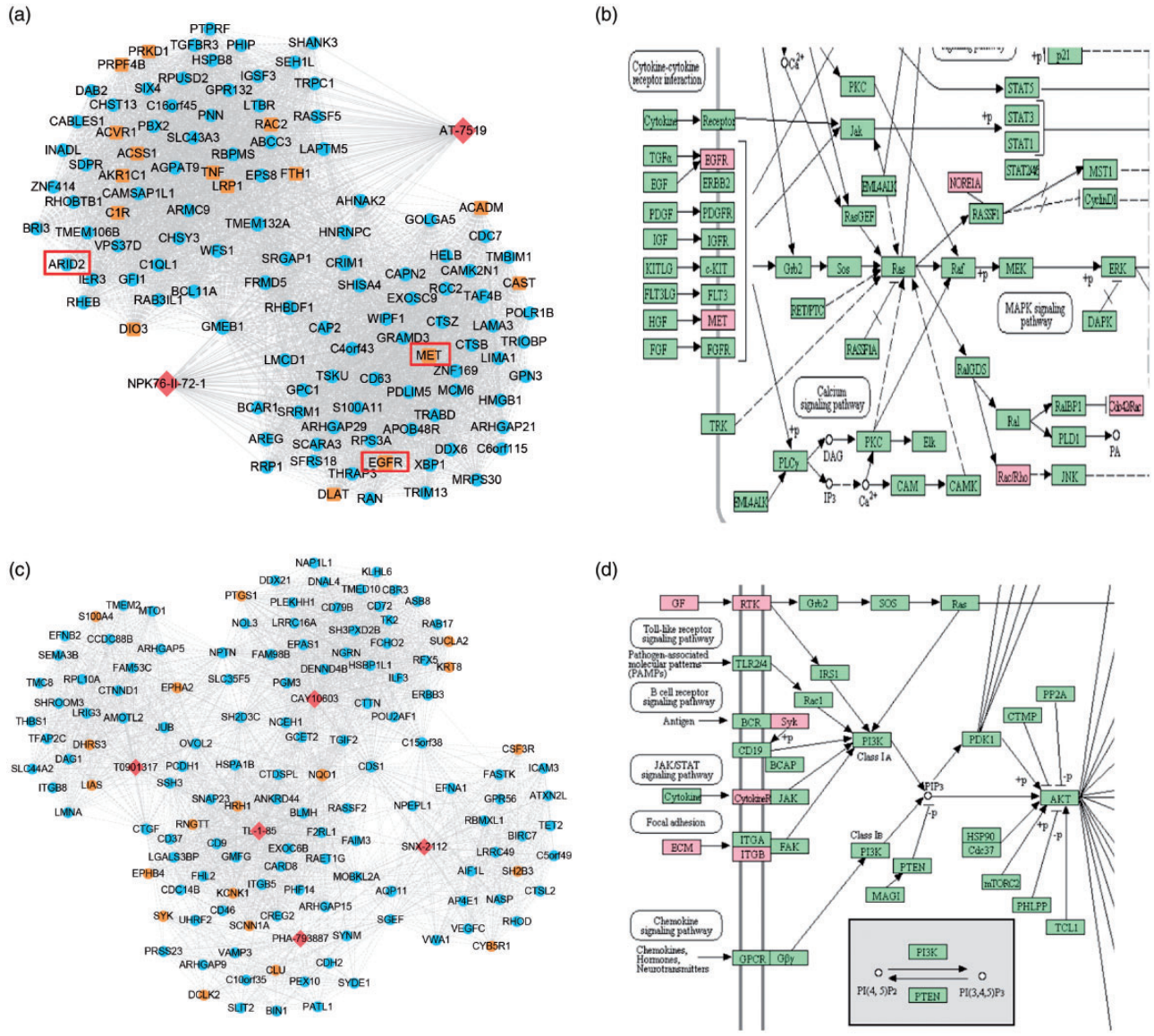

Figure 10. Target genes in modules of DGWI network are involved in cancer pathways and PI3K-AKT signaling pathway. (a) Drug targets and cancer driver genes in cluster 1; (b) Genes in cluster 1 are involved in human cancer pathway. (c) Drug targets and cancer driver genes in cluster 2; (d) Genes in cluster 2 are involved in PI3K-AKT signaling pathway. Drugs and gene are marked with diamonds and circles, orange circles represent drug targets. Cancer driver genes are marked by rectangles. Protein coding products of genes in clusters are marked by pink rectangles in pathways. (A color version of this figure is available in the online journal.)

pathways, such as the cancer signaling pathway, the MAPK signaling pathway, and the drug metabolism cytochrome P450 pathway (Figures S7 to S9), reflecting the effect of DNA methylation on cancer and drug response.

For the top-ranked drug-gene pairs sorted by FDR values of linear regression models in descending order, DNA methylation is generally positively correlated with IC50 of drugs. For instance, for the top 10 drug-gene pairs, there is only one negative correlation pair (Figure S10), for the top 100 drug-gene pairs, positive correlation pairs account for 90%, and for the top 200 drug-gene pairs, the number of negative correlation pairs is only 27. It is worth noting that the regression coefficient of DNA methylation is −0.64 if the linear regression model between logarithmic IC50 of ABT-869 and DNA methylation of *LOC100302401* is directly constructed. However, DNA methylation of *LOC100302401* turns to be positively correlated with logarithmic IC50 of ABT-869 with regression coefficient of 2.5 once the tissue type, microsatelite instability, media, growth characteristics are added as covariates to the linear regression model. The introduction of covariates increases the coefficient R^2 and improves the linear fit suggesting the necessity of adding covariates when building linear regression models.

We identified 999 candidate drug-gene pairs by constructing logistic regression models between the binarized drug response and DNA methylation. Among the top 20 drugs with the most candidate genes, SNX-2112, a selective inhibitor of HSP90, inhibits tumor cell proliferation and angiogenesis by inhibiting cytokine-induced Akt and extracellular signal-related kinase activity in a variety of myeloma and hematological tumors.[31] Treatment of cancer cell lines with HSP90 inhibitors results in decreased mRNA and protein expression levels of DNA methyltransferases

DNMT1, DNMT3A, and DNMT3B, which thereby inhibits tumor cell growth and induces tumor cell apoptosis.[36] UNC0638 is an inhibitor of histone methyltransferase G9A, which is selective for a variety of epigenetic targets and selectively target genes with H3K9me2 modification. Genes regulated by UNC0638 usually have higher DNA methylation levels.[37] For the human breast cancer cell line MDA-MB-231, UNC0638 decreased the H3K9me2 levels in a concentration-dependent manner, and significantly inhibited the cloning of the human breast cancer cell line MCF7 by reducing the H3K9me2 abundance of the G9A-regulated endogenous gene promoter. Combined with the results of our research and the support of previous studies, it may help to develop and select future epigenetic treatment strategies for cancer.

FERMT3 is involved in protein interaction during integrin activation and plays an important role in cell adhesion, migration, differentiation, and proliferation. Early studies mainly focus on gene mutations of *FERMT3*, which may lead to autosomal recessive adhesion defect syndrome.[38] Recent researches have shown that *FERMT3* is involved in cancer occurrence and progression. For instance, Lu reported that *FERMT3* could promote cell proliferation of high-grade glioma and become resistant to temozolomide by activating integrin-mediated Wnt signaling pathway.[39] *FERMT3* interacts with ribosomes to regulate c-Myc protein expression, thereby promoting cell proliferation of chronic myeloid leukemia.[40] However, few studies concentrate on the relationship between *FERMT3* and DNA methylation. DNA methylation level of *FERMT3* well distinguishes SNX-212 sensitive and resistant cell lines in our study. *LOC100130776* with official gene name *AGAP2-AS1* is an lncRNA located on chromosome 12. Previous studies have confirmed that *AGAP2-AS1* is closely related to the development of various cancers and has recently been identified as an oncogenic lncRNA. *AGAP2-AS1* is highly expressed in gastric cancer tissues and cell lines, and gastric cancer patients with high expression levels of *AGAP2-AS1* have a worse prognosis and a shorter overall survival compared to patients with low expression levels of *AGAP2-AS1*. The transcription factor SP1 activates *AGAP2-AS1* and thereby promotes the growth and invasion of gastric cancer cells, indicating that *AGAP2-AS1* is a potential target for the diagnosis and treatment of gastric cancer.[41] Up-regulated *AGAP2-AS1* inhibits tumor suppressor gene *LAST2* and *KLF2* transcription by interacting with *EZH2* and *LSD1*, and plays an oncogenic role in non-small cell lung cancer.[42] Combined with the role of *AGAP2-AS1* in cancer development and drug resistance of various cancers reported in previous studies, we infer that *AGAP2-AS1* may be a potential signature for cancer prognosis and a target for pharmacotherapy.

To explore the effect of gene expression on the response to anticancer drugs, the most significant 20 drug-gene pairs obtained by likelihood ratio tests are listed in Table S13, of which Afatinib, a tyrosine kinase inhibitor, is included in 10 drug-gene pairs. Afatinib irreversibly inhibits ERBB2 and EGFR kinase and is commonly used to treat *EGFR* mutant non-small cell lung cancer.[43] Due to the inhibition of *ERBB2* activity, Afatinib has a therapeutic effect on certain breast cancer subtypes and other cancers driven by *EGFR* and *ERBB2*.[44] The expression of *ERBB2*, *GRB7*, *LAD1* and *SSH3* is significantly related to the response to Afatinib in two screening data sets. Expression levels of *ERBB2* significantly improve the performance of linear model by increasing the log likelihood from -1100.81 to -1045.59 ($P = 7.85\text{E-}26$). Sanchez-Vega[45] reported that for esophageal cancer and gastric cancer patients who were resistant to Trastuzumab, the sensitivity of Afatinib was related to copy number amplification of *EGFR* and *ERBB2*, offering support for the relationship between expression levels of *ERBB2* and the response to Afatinib identified in our research. Despite no direct evidence demonstrating the relationship between *GRB7* and Afatinib, the high expression of *GRB7* is strongly associated with poor survival in patients with primary breast cancer, indicating *GRB7* is likely to be a prognostic marker and drug target.[46] The expression of *GRB7* is associated with the risk of recurrence in triple-negative breast cancer patients receiving chemotherapy, suggesting that *GRB7* and related signaling pathway may be a potential biological target for cancer therapy.[47] A study has reported that DNA methylation in the promoter region of *LAD1* is significantly associated with poor survival of renal clear cell carcinoma, and can predict the prognosis of patients.[48] In our research, there is a strong negative correlation between DNA methylation and expression of *LAD1* (Pearson correlation coefficient $= -0.71$). Therefore, it is speculated that the negative regulation may have an impact on the growth and metastasis of cancer cells and the prognosis of cancer patients, and thus *LAD1* may be a potential target for cancer drug treatment.

The AUC values of M5 models are greater than DNA methylation-based M3 models and gene expression-based M4 models (Figure S11(a)). Then the comparison was performed respectively in three groups to further investigate the effect of gene expression regulation on the response to anticancer drugs (Figure S11(b) to (d)). For some drug-gene pairs, the introduction of gene expression significantly optimizes the original model. In the first comparison group, M3 model has higher AUC values than M4 model indicating that DNA methylation plays a leading role in predicting drug response for these drug-gene pairs. In the second and third comparison groups, M4 model has higher AUC values than M3 model indicating that gene expression plays a leading role in predicting drug response for these drug-gene pairs.

In conclusion, our study systematically identifies the molecular characteristics of anticancer drug responses at pan-cancer levels by integrating DNA methylation data, gene expression data, and drug response data from large-scale cancer cell lines, and explores the effect of epigenetic regulation on drug resistance. Although we have got some meaningful results, this study still has some limitations. First, these results are obtained based on cancer cell line models, and it still requires further validation by in vivo studies. Second, due to the limited amount of data, we conducted drug sensitivity studies on pan-cancer levels. In the future, as cell lines and clinical data increase,

cancer type specific analysis will be performed to further explore the molecular mechanism of drug resistance.

AUTHORS' CONTRIBUTIONS

YZ and BQS contributed to the study concept and design, study supervision, and manuscript review. WHL contributed to data analyses, interpretation of data, and manuscript writing. XDZ, QW, and HLD contributed to data preparation, data acquisition, data review, and figure plotting. All authors have read and approved the final manuscript.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

This work is supported by the National Natural Science Foundation of China (grant numbers 31771601, 61972116).

ORCID iDs

Baoqing Sun https://orcid.org/0000-0002-1671-0723
Yan Zhang https://orcid.org/0000-0002-5307-2484

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Vedeld HM, Goel A, Lind GE. Epigenetic biomarkers in gastrointestinal cancers: the current state and clinical perspectives. *Semin Cancer Biol* 2018;**51**:36–49
2. Duruisseaux M, Esteller M. Lung cancer epigenetics: from knowledge to applications. *Semin Cancer Biol* 2018;**51**:116–28
3. Assenov Y, Brocks D, Gerhauser C. Intratumor heterogeneity in epigenetic patterns. *Semin Cancer Biol* 2018;**51**:12–21
4. Chatterjee A, Rodger EJ, Eccles MR. Epigenetic drivers of tumourigenesis and cancer metastasis. *Semin Cancer Biol* 2018;**51**:149–59
5. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* 2011;**8**:686–700
6. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007;**447**:433–40
7. Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;**128**:683–92
8. Chen X, Wu H, Feng J, Li Y, Lv J, Shi W, Fan W, Xiao L, Sun D, Jiang M, Shi M. Transcriptome profiling unveils GAP43 regulates ABC transporters and EIF2 signaling in colorectal cancer cells. *BMC Cancer* 2021;**21**:24
9. Xavier CP, Pesic M, Vasconcelos MH. Understanding cancer drug resistance by developing and studying resistant cell line models. *Curr Cancer Drug Targets* 2016;**16**:226–37
10. Brodniewicz T, Grynkiewicz G. Preclinical drug development. *Acta Pol Pharm* 2010;**67**:578–85
11. Sim SC, Ingelman-Sundberg M. Pharmacogenomic biomarkers: new tools in current and future drug therapy. *Trends Pharmacol Sci* 2011;**32**:72–81
12. Shuldiner AR, Relling MV, Peterson JF, Hicks JK, Freimuth RR, Sadee W, Pereira NL, Roden DM, Johnson JA, Klein TE, Pharmacogenomics Research Network Translational Pharmacogenetics Program G, Shuldiner AR, Vesely M, Robinson SW, Ambulos N, Jr., Stass SA, Kelemen MD, Brown LA, Pollin TI, Beitelshees AL, Zhao RY, Pakyz RE, Palmer K, Alestock T, O'Neill C, Maloney K, Branham A, Sewell D, Relling MV, Crews K, Hoffman J, Cross S, Haidar C, Baker D, Hicks JK, Bell G, Greeson F, Gaur A, Reiss U, Huettel A, Cheng C, Gajjar A, Pappo A, Howard S, Hudson M, Pui CH, Jeha S, Evans WE, Broeckel U, Altman RB, Gong L, Whirl-Carrillo M, Klein TE, Sadee W, Manickam K, Sweet KM, Embi PJ, Roden D, Peterson J, Denny J, Schildcrout J, Bowton E, Pulley J, Beller M, Mitchell J, Danciu I, Price L, Pereira NL, Weinshilboum R, Wang L, Johnson JA, Nelson D, Clare-Salzler M, Elsey A, Burkley B, Langaee T, Liu F, Nessl D, Dong HJ, Lesko L, Freimuth RR, Chute CG. The pharmacogenomics research network translational pharmacogenetics program: overcoming challenges of real-world implementation. *Clin Pharmacol Ther* 2013;**94**:207–10
13. Kacevska M, Ivanov M, Ingelman-Sundberg M. Perspectives on epigenetics and its relevance to adverse drug reactions. *Clin Pharmacol Ther* 2011;**89**:902–7
14. Ivanov M, Kacevska M, Ingelman-Sundberg M. Epigenomics and interindividual differences in drug response. *Clin Pharmacol Ther* 2012;**92**:727–36
15. Evison BJ, Bilardi RA, Chiu FC, Pezzoni G, Phillips DR, Cutts SM. CpG methylation potentiates pixantrone and doxorubicin-induced DNA damage and is a marker of drug sensitivity. *Nucleic Acids Res* 2009;**37**:6355–70
16. Kucuk C, Hu X, Jiang B, Klinkebiel D, Geng H, Gong Q, Bouska A, Iqbal J, Gaulard P, McKeithan TW, Chan WC. Global promoter methylation analysis reveals novel candidate tumor suppressor genes in natural killer cell lymphoma. *Clin Cancer Res* 2015;**21**:1699–711
17. Krushkal J, Silvers T, Reinhold WC, Sonkin D, Vural S, Connelly J, Varma S, Meltzer PS, Kunkel M, Rapisarda A, Evans D, Pommier Y, Teicher BA. Epigenome-wide DNA methylation analysis of small cell lung cancer cell lines suggests potential chemotherapy targets. *Clin Epigenetics* 2020;**12**:93–120
18. Mendaza S, Ulazia-Garmendia A, Monreal-Santesteban I, Cordoba A, Azua YR, Aguiar B, Beloqui R, Armendariz P, Arriola M, Martin-Sanchez E, Guerrero-Setas D. ADAM12 is a potential therapeutic target regulated by hypomethylation in Triple-Negative breast cancer. *Int J Mol Sci* 2020;**21**:903–19
19. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, Aben N, Goncalves E, Barthorpe S, Lightfoot H, Cokelaer T, Greninger P, van Dyk E, Chang H, de Silva H, Heyn H, Deng X, Egan RK, Liu Q, Mironenko T, Mitropoulos X, Richardson L, Wang J, Zhang T, Moran S, Sayols S, Soleimani M, Tamborero D, Lopez-Bigas N, Ross-Macdonald P, Esteller M, Gray NS, Haber DA, Stratton MR, Benes CH, Wessels LF, Saez-Rodriguez J, McDermott U, Garnett MJ. A landscape of pharmacogenomic interactions in cancer. *Cell* 2016;**166**:740–54
20. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in illumina infinium 450 k DNA methylation data. *Bioinformatics* 2013;**29**:189–96
21. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S. ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* 2014;**30**:428–30
22. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003;**31**:e15
23. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;**19**:185–93
24. Wj Y. Index for rating diagnostic tests. *Cancer* 1950;**3**:32–5
25. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013;**14**:128
26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504
27. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. *Science* 2013;**339**:1546–58
28. Tudzarova S, Mulholland P, Dey A, Stoeber K, Okorokov AL, Williams GH. p53 controls CDC7 levels to reinforce G1 cell cycle arrest upon genotoxic stress. *Cell Cycle* 2016;**15**:2958–72

29. Liu Z, Li J, Chen J, Shan Q, Dai H, Xie H, Zhou L, Xu X, Zheng S. MCM family in HCC: MCM6 indicates adverse tumor features and poor outcomes and promotes S/G2 cell cycle progression. *BMC Cancer* 2018;**18**:200
30. Qin JH, Wang K, Fu XL, Zhou PJ, Liu Z, Xu DD, Wang YF, Yang DP, Xie QL, Liu QY. Hsp90 inhibitor induces KG-1a cell differentiation and apoptosis via akt/NF-kappaB signaling. *Oncol Rep* 2017;**38**:1517–24
31. Okawa Y, Hideshima T, Steed P, Vallet S, Hall S, Huang K, Rice J, Barabasz A, Foley B, Ikeda H, Raje N, Kiziltepe T, Yasui H, Enatsu S, Anderson KC. SNX-2112, a selective Hsp90 inhibitor, potently inhibits tumor cell growth, angiogenesis, and osteoclastogenesis in multiple myeloma and other hematologic tumors by abrogating signaling via akt and ERK. *Blood* 2009;**113**:846–55
32. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 2013;**502**:472–9
33. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L. Mutational landscape and significance across 12 major cancer types. *Nature* 2013;**502**:333–9
34. Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, Xu ZD, Zhu HG, Ling ZQ, Ye D, Guan KL, Xiong Y. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* 2013;**32**:663–9
35. Spangle JM, Dreijerink KM, Groner AC, Cheng H, Ohlson CE, Reyes J, Lin CY, Bradner J, Zhao JJ, Roberts TM, Brown M. PI3K/AKT signaling regulates H3K4 methylation in breast cancer. *Cell Rep* 2016;**15**:2692–704
36. Nagaraju GP, Wu C, Merchant N, Chen Z, Lesinski GB, El-Rayes BF. Epigenetic effects of inhibition of heat shock protein 90 (HSP90) in human pancreatic and colon cancer. *Cancer Lett* 2017;**402**:110–6
37. Sato T, Cesaroni M, Chung W, Panjarian S, Tran A, Madzo J, Okamoto Y, Zhang H, Chen X, Jelinek J, Issa JJ. Transcriptional selectivity of epigenetic therapy in cancer. *Cancer Res* 2017;**77**:470–81
38. McDowall A, Svensson L, Stanley P, Patzak I, Chakravarty P, Howarth K, Sabnis H, Briones M, Hogg N. Two mutations in the KINDLIN3 gene of a new leukocyte adhesion deficiency III patient reveal distinct effects on leukocyte function in vitro. *Blood* 2010;**115**:4834–42
39. Lu C, Cui C, Liu B, Zou S, Song H, Tian H, Zhao J, Li Y. FERMT3 contributes to glioblastoma cell proliferation and chemoresistance to temozolomide through integrin mediated wnt signaling. *Neurosci Lett* 2017;**657**:77–83
40. Qu J, Ero R, Feng C, Ong LT, Tan HF, Lee HS, Ismail MH, Bu WT, Nama S, Sampath P, Gao YG, Tan SM. Kindlin-3 interacts with the ribosome and regulates c-Myc expression required for proliferation of chronic myeloid leukemia cells. *Sci Rep* 2015;**5**:18491
41. Qi F, Liu X, Wu H, Yu X, Wei C, Huang X, Ji G, Nie F, Wang K. Long noncoding AGAP2-AS1 is activated by SP1 and promotes cell proliferation and invasion in gastric cancer. *J Hematol Oncol* 2017;**10**:48–61
42. Li W, Sun M, Zang C, Ma P, He J, Zhang M, Huang Z, Ding Y, Shu Y. Upregulated long non-coding RNA AGAP2-AS1 represses LATS2 and KLF2 expression through interacting with EZH2 and LSD1 in non-small-cell lung cancer cells. *Cell Death Dis* 2016;**7**:e2225
43. Li D, Ambrogio L, Shimamura T, Kubo S, Takahashi M, Chirieac LR, Padera RF, Shapiro GI, Baum A, Himmelsbach F, Rettig WJ, Meyerson M, Solca F, Greulich H, Wong KK. BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. *Oncogene* 2008;**27**:4702–11
44. Minkovsky N, Berezov A. BIBW-2992, a dual receptor tyrosine kinase inhibitor for the treatment of solid tumors. *Curr Opin Investig Drugs* 2008;**9**:1336–46
45. Sanchez-Vega F, Hechtman JF, Castel P, Ku GY, Tuvy Y, Won H, Fong CJ, Bouvier N, Nanjangud GJ, Soong J, Vakiani E, Schattner M, Kelsen DP, Lefkowitz RA, Brown K, Lacouture ME, Capanu M, Mattar M, Qeriqi B, Cecchi F, Tian Y, Hembrough T, Nagy RJ, Lanman RB, Larson SM, Pandit-Taskar N, Schoder H, Iacobuzio-Donahue CA, Ilson DH, Weber WA, Berger MF, de Stanchina E, Taylor BS, Lewis JS, Solit DB, Carrasquillo JA, Scaltriti M, Schultz N, Janjigian YY. EGFR and MET amplifications determine response to HER2 inhibition in ERBB2-amplified esophagogastric cancer. *Cancer Discov* 2019;**9**:199–209
46. Nadler Y, Gonzalez AM, Camp RL, Rimm DL, Kluger HM, Kluger Y. Growth factor receptor-bound protein-7 (Grb7) as a prognostic marker and therapeutic target in breast cancer. *Ann Oncol* 2010;**21**:466–73
47. Sparano JA, Goldstein LJ, Childs BH, Shak S, Brassard D, Badve S, Baehner FL, Bugarini R, Rowley S, Perez EA, Shulman LN, Martino S, Davidson NE, Kenny PA, Sledge GW, Jr., Gray R. Relationship between quantitative GRB7 RNA expression and recurrence after adjuvant anthracycline chemotherapy in triple-negative breast cancer. *Clin Cancer Res* 2011;**17**:7194–203
48. van Vlodrop IJH, Joosten SC, De Meyer T, Smits KM, Van Neste L, Melotte V, Baldewijns M, Schouten LJ, van den Brandt PA, Jeschke J, Yi JM, Schuebel KE, Ahuja N, Herman JG, Aarts MJ, Bosman FT, Van Criekinge W, van Engeland M. A Four-Gene promoter methylation marker panel consisting of GREM1, NEURL, LAD1, and NEFH predicts survival of clear cell renal cell cancer patients. *Clin Cancer Res* 2017;**23**:2006–18

(Received December 22, 2020, Accepted March 16, 2021)