

## Research Article

# Prediction of Response to Radiotherapy by Characterizing the Transcriptomic Features in Clinical Tumor Samples across 15 Cancer Types

Yu Xu,<sup>1</sup> Chao Tang,<sup>2</sup> Yan Wu,<sup>2</sup> Ling Luo,<sup>2</sup> Ying Wang,<sup>2</sup> Yongzhong Wu ,<sup>2</sup> and Xiaolong Shi <sup>2</sup>

<sup>1</sup>College of Bioengineering, Chongqing University, Chongqing, China

<sup>2</sup>Radiation and Cancer Biology Laboratory, Radiation Oncology Center, Chongqing Key Laboratory of Translational Research for Cancer Metastasis and Individualized Treatment, Institute and Chongqing Cancer Hospital, Chongqing University Cancer Hospital and Chongqing Cancer, Chongqing 400030, China

Correspondence should be addressed to Yongzhong Wu; [cqmdwyz@163.com](mailto:cqmdwyz@163.com) and Xiaolong Shi; [xshi.bear@cqu.edu.cn](mailto:xshi.bear@cqu.edu.cn)

Received 13 March 2022; Revised 12 April 2022; Accepted 13 April 2022; Published 9 May 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Yu Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Purpose.** Radiotherapy (RT) is one of the major cancer treatments. However, the responses to RT vary among individual patients, partly due to the differences of the status of gene expression and mutation in tumors of patients. Identification of patients who will benefit from RT will improve the efficacy of RT. However, only a few clinical biomarkers were currently used to predict RT response. Our aim is to obtain gene signatures that can be used to predict RT response by analyzing the transcriptome differences between RT responder and nonresponder groups. **Materials and Methods.** We obtained transcriptome data of 1664 patients treated with RT from the TCGA database across 15 cancer types. First, the genes with a significant difference between RT responder (R group) and nonresponder groups (PD group) were identified, and the top 100 genes were used to build the gene signatures. Then, we developed the predictive model based on binary logistic regression to predict patient response to RT. **Results.** We identified a series of differentially expressed genes between the two groups, which are involved in cell proliferation, migration, invasion, EMT, and DNA damage repair pathway. Among them, MDC1, UCP2, and RBM45 have been demonstrated to be involved in DNA damage repair and radiosensitivity. Our analysis revealed that the predictive model was highly specific for distinguishing the R and PD patients in different cancer types with an area under the curve (AUC) ranging from 0.772 to 0.972. It also provided a more accurate prediction than that from a single-gene signature for the overall survival (OS) of patients. **Conclusion.** The predictive model has a potential clinical application as a biomarker to help physicians create optimal treatment plans. Furthermore, some of the genes identified here may be directly involved in radioresistance, providing clues for further studies on the mechanism of radioresistance.

## 1. Introduction

Cancer is the leading cause of morbidity and mortality in the world, regardless of the level of human development, which is accounting for over 9.9 million deaths worldwide annually [1, 2]. Radiation therapy (RT) has played a major role in cancer therapeutics, with approximately 50% of cancer patients using RT [3]. RT mainly induces apoptosis of cancer cells by causing DNA double-strand breaks (DSBs) [4]. RT is

often curative, particularly in some head and neck cancer, prostate cancer, and cervical cancer. However, the inherent radioresistance of tumor cells and the acquired radioresistance can reduce the therapeutic effect and ultimately lead to poor outcomes of patients such as tumor recurrence, metastasis, and patient death [5]. In the era of precision medicine, there is a growing need for precision radiotherapy. This requires the plans of individualized treatment by considering patients' multiple information, including the

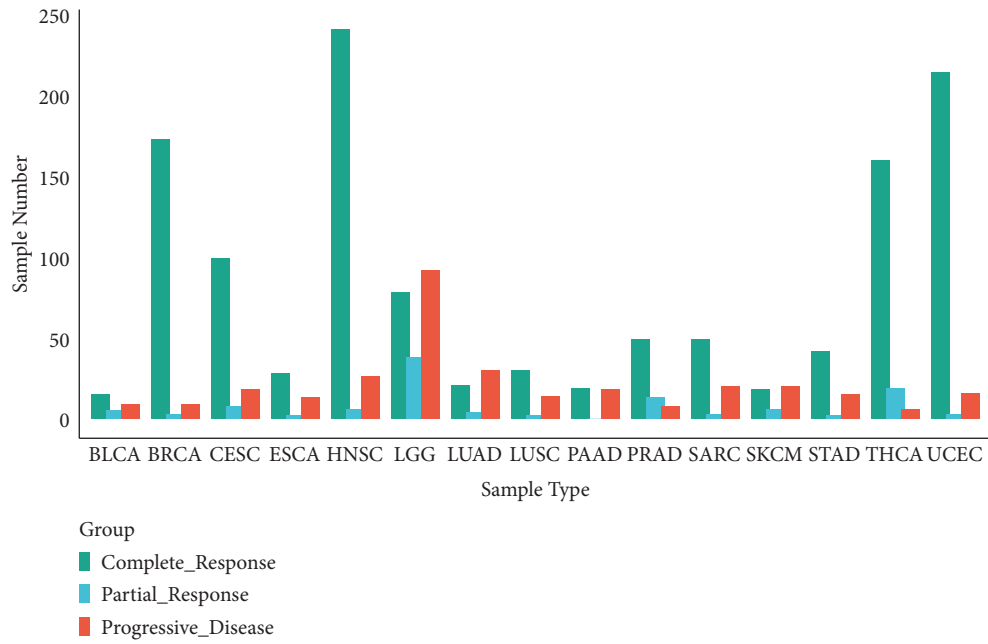


FIGURE 1: Statistical histogram of cancer types of samples and grouping.

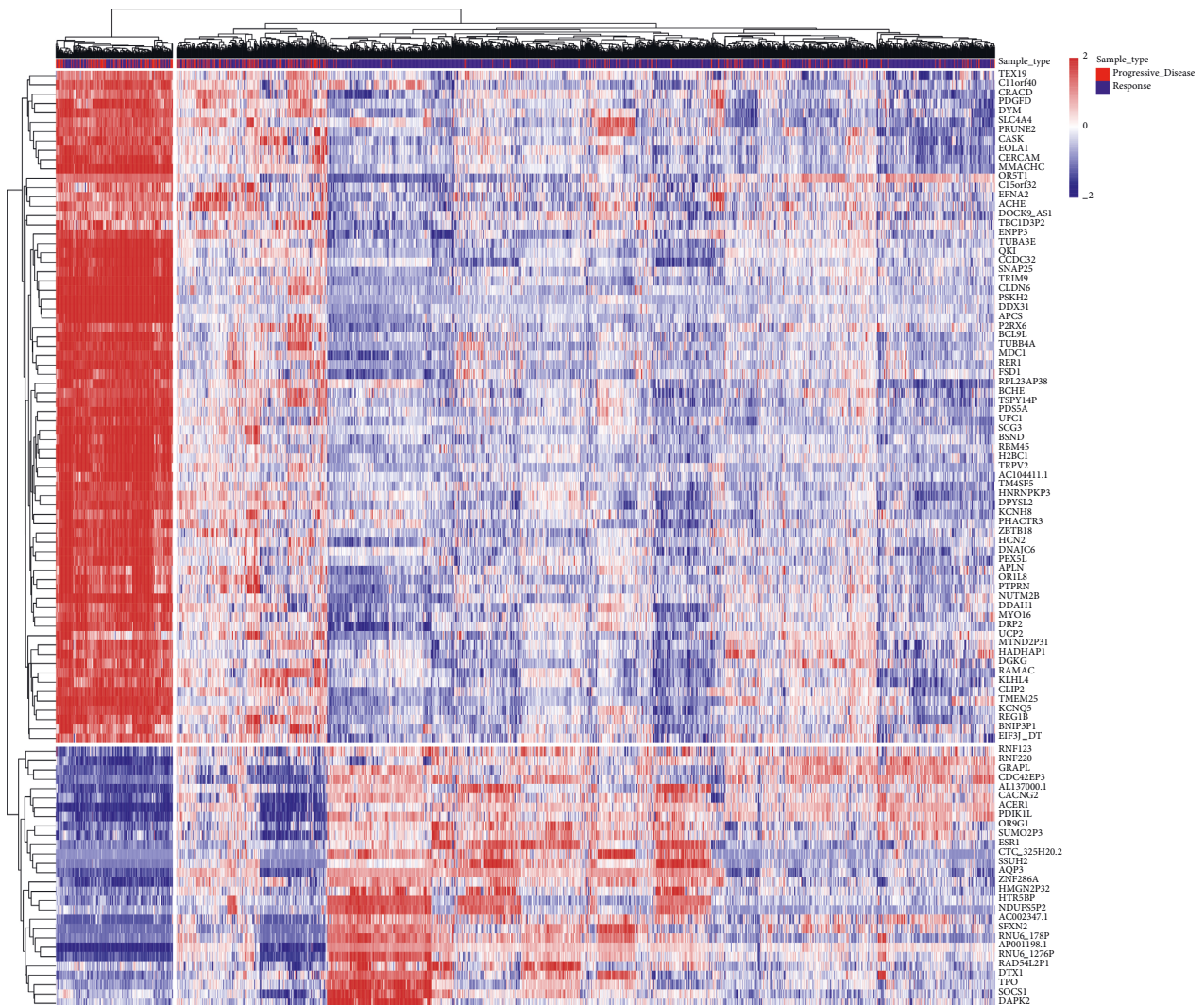


FIGURE 2: Heatmap of the top 100 differentially genes for all cancer species analyzed. Both samples and genes were clustered with average linkage.

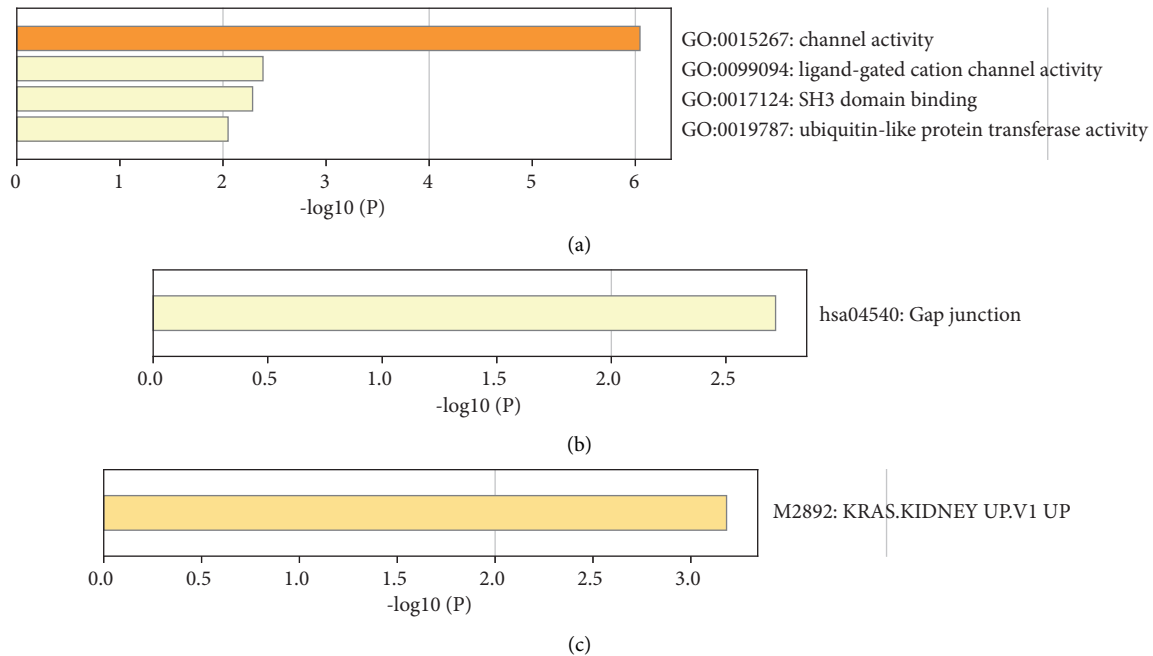


FIGURE 3: Histogram of pathway enrichment of top 100 differential genes.

status of gene mutation and profile of gene expression, in order to achieve the best therapeutic efficacy. Radioresistance is partly due to the differences of the status of gene expression and mutation in tumors of patients. Thus, identification of patients who will benefit from RT through analyzing gene mutation or gene expression will improve the efficacy of RT.

The previous studies have demonstrated that the mutations of several genes influenced the radiosensitivity. Ataxia-telangiectasia mutated (ATM) is a central regulator of DNA damage response [6]. Mutation and inactivation of ATM can lead to increased gene instability and impaired repair of DNA double-strand breaks. Tumors harboring ATM mutation are extremely sensitive to radiation. Mutations of ATM have been found in several tumor types. Targeted next-generation sequencing has revealed an 8% incidence of ATM mutations in prostate cancer [7]. ATM is also one of the commonly mutated genes in mantle cell lymphoma (MCL) and lung adenocarcinoma [7, 8]. Recent experimental evidence demonstrated that the mutational status of ATM can be used as a biomarker for radiotherapy [9]. The correlation between KEAP1/NRF2 mutation status and radioresistance has been investigated. Keap1-Nrf2 is a master regulator of cellular response to oxidative and radiation stress [10]. The KEAP1-NRF2 pathway is involved in protection of cells from oxidative and toxic stresses. Recent studies indicated that the activating mutation of the KEAP1-NRF2 pathway induces radioresistance, and KEAP1/NRF2 mutation status is a strong predictor of RT outcome in patients with NSCLC [11].

Although the mutations of these genes could serve as predictive biomarkers for personalization of therapeutic strategies, their widespread use would be limited due to the limited number of ATM or KEAP1-NRF2 mutations carriers

among cancer patients. The previous studies revealed that the difference in the status of gene expression (including lncRNA and miRNA) and DNA methylation on tumor suppressor gene may also contribute to resistance to RT. For example, radioresistance in NSCLC has been associated with overexpression of antioxidant enzymes such as Mn-superoxide dismutase (Mn-SOD) [12]. Several studies revealed many miRNAs, including miR-95 [13], miR-221, miR-222 [14], and miR-106b [15], enhanced radioresistance in cancer cells, while miR-30a [16], miR-16 [17], miR-449 [18], miR-17 [19], and miR-100 [20] enhance the radiosensitivity. The role of epigenetic modifications, especially DNA methylation, has been explored in radioresistance in malignant tumors. The changes in radiosensitivity caused by DNA methylation in the promoter region of genes associated with cell damage repair, cell proliferation, and cell cycle have been demonstrated [21]. For example, Liu et al. [22] found the promoter region of the ERCC1 gene in two radiosensitive cell lines was hypermethylated, while it was hypomethylated in the other two radioresistant cell lines. These studies provided potential biomarkers for stratification of RT patients.

In previous studies, cancer cell lines usually were used as a model for radioresistance study. We speculated that radioresistance probably involves a complex transcriptional coexpression network within tumor cells. In tumor tissue, the factors involved in the response to radiation might be more complex than in cell lines. Therefore, biomarkers obtained from the analysis of actual clinical tumor samples may be closer to clinical application. The development of high-throughput sequencing methods for mRNA (RNA-Seq) has provided a very powerful tool to analyze the transcriptome of tumors, and a huge amount of data are available [23]. In this study, we performed a large-scale analysis of transcriptomic data collected from The Cancer

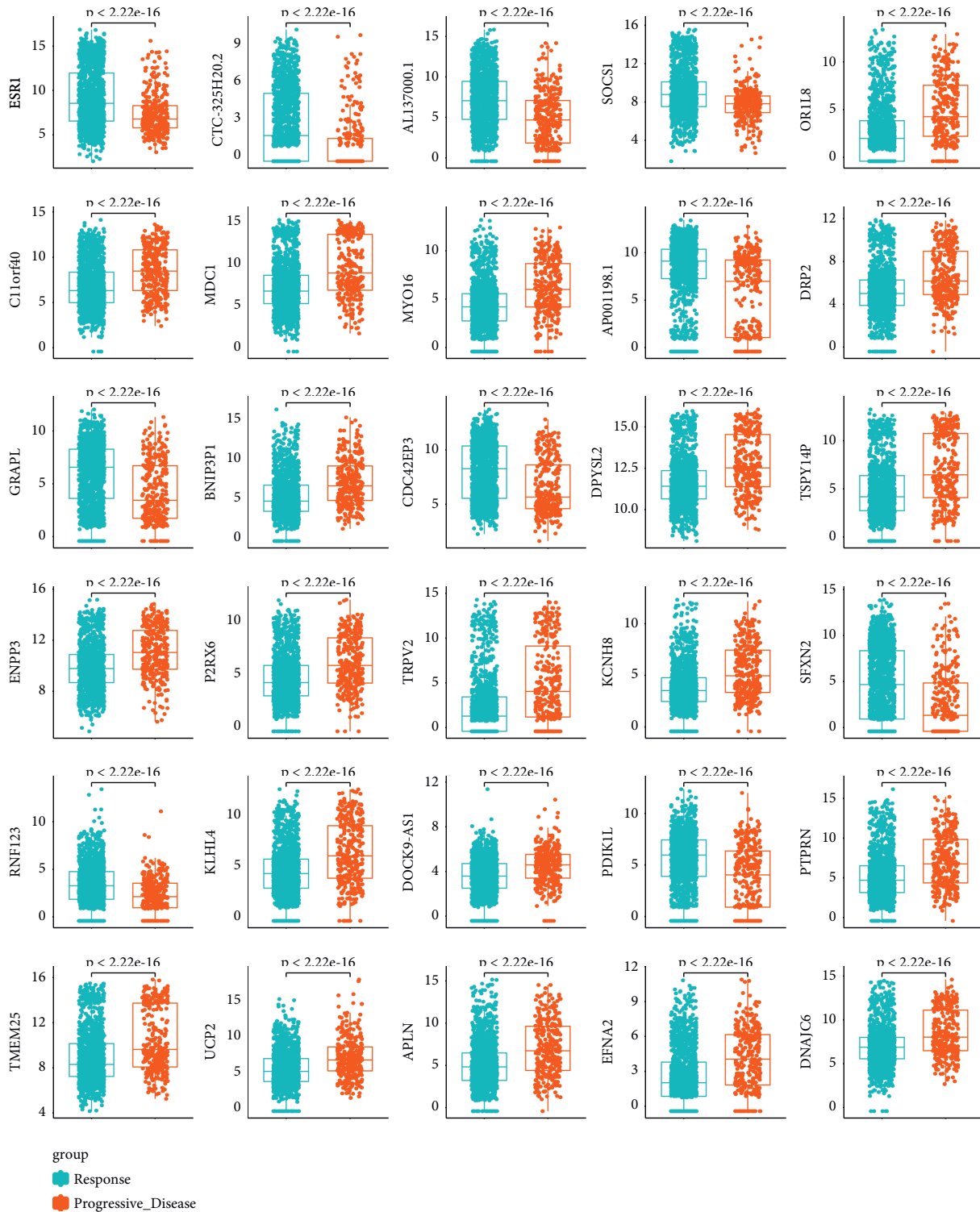


FIGURE 4: mRNA levels of differentially expressed genes between response and progressive disease tumors. The distribution of gene expression values in R or PD samples was drawn through the boxplot, and the  $p$  value is marked ( $t$ -test).

Genome Atlas (TCGA) from those who were treated with RT across 15 different cancer types. The gene signatures for predicting RT response were obtained by analyzing the transcriptome differences between RT responder and non-responder groups, and the performance of gene signatures was estimated in this study.

## 2. Materials and Methods

**2.1. Data Acquisition.** The data for this study were obtained from the TCGA website (<https://portal.gdc.cancer.gov/>). This study downloaded 15 types of cancer clinical data and gene expression data. Fifteen types of cancers include

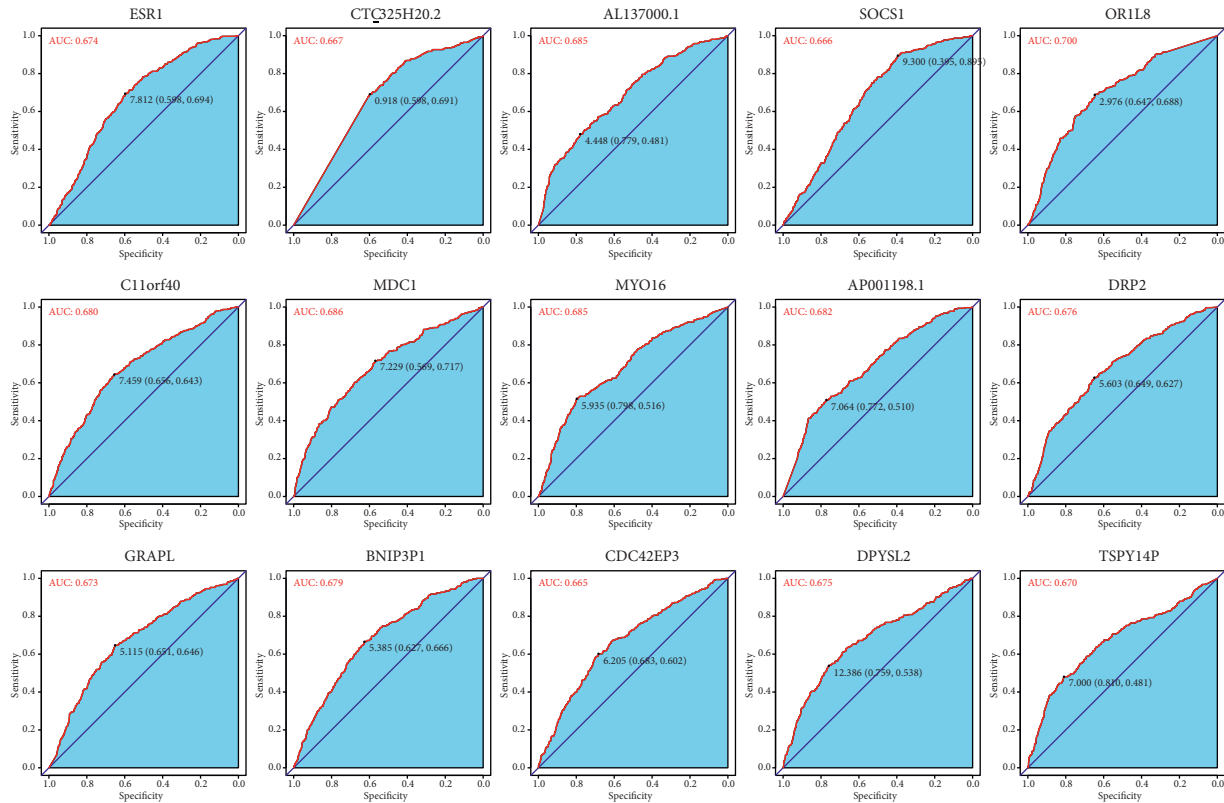


FIGURE 5: ROC curve of differentially expressed genes between response tumors and progressive disease tumors.

bladder urothelial carcinoma (BLCA), invasive breast cancer (BRCA), cervical squamous cell carcinoma (CESC), esophageal cancer (ESCA), head and neck squamous cell carcinoma (HNSC), low-grade glioma (LGG), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pancreatic cancer (PAAD), prostate adenocarcinoma (PRAD), sarcoma (SARC), cutaneous melanoma (SKCM), gastric adenocarcinoma (STAD), thyroid carcinoma (THCA), and endometrial carcinoma (UCEC). First, this study extracted the clinical treatment information and overall survival information of patients. Next, tumor tissue samples from those patients treated with radiotherapy were selected according to clinical data, and then the samples were divided into two categories according to RECIST response results: response group (including complete response and partial response) and disease progression group. Finally, 1664 tumor samples were used for analysis in this study, including 1350 response (*R*) samples and 314 progressive disease (PD) samples.

**2.2. Expression Analysis and Differential Expression Gene Identification.** Data of the mRNA profile of samples selected from the TCGA dataset were analyzed using the DESeq2 package [24] in R language. In the gene expression data, the genes will not be analyzed further if the read counts of these genes were less than 10 in 80% of samples. In this study, samples were divided into response (*R*) and progressive disease (PD) samples. The differentially expressed genes

(DEGs) between *R* and PD samples were identified at the criteria of  $p < 0.05$  and  $|\log_{2}FC| > 1$ . Wald test was used to calculate the  $p$  value. The  $t$ -test was also used to calculate  $p$  value in this study. The top 100 genes with the significant difference were obtained for follow-up analysis. For individual tumor types, the tumor type-specific gene signature was obtained from the top 100 genes according to the  $p$  value of each gene expression difference in individual tumor types.

**2.3. Binary Logistic Regression Analysis.** Logistic regression analysis was used to establish a prediction model based on risk factors and used one or more explanatory variables to predict a class of response variables [25]. The standardized expression value of the top 100 genes was used as the variable; the *R* sample was used as the reference group, and the value was set to 1; the value of progressive sample was set to 0. The RT response was predicted by logistic regression modeling and quantified using a prediction index. The prediction index is the regression coefficient obtained by multiplying the expression level of the selected gene with the logistic regression model ( $\beta$ ). The prediction index for each patient is as follows: predictive index =  $\text{Expr gene1} * \beta_{\text{gene1}} + \text{Expr gene2} * \beta_{\text{gene2}} + \text{Expr gene3} * \beta_{\text{gene3}} + \dots + \text{Expr geneN} * \beta_{\text{geneN}} + \text{intercept}$ .

If the prediction index is less than the threshold (e.g.,  $< 0.5$ ), the sample is more likely to be a progressive disease, and if the prediction index is more than the threshold, the sample is more likely to be radiosensitive.

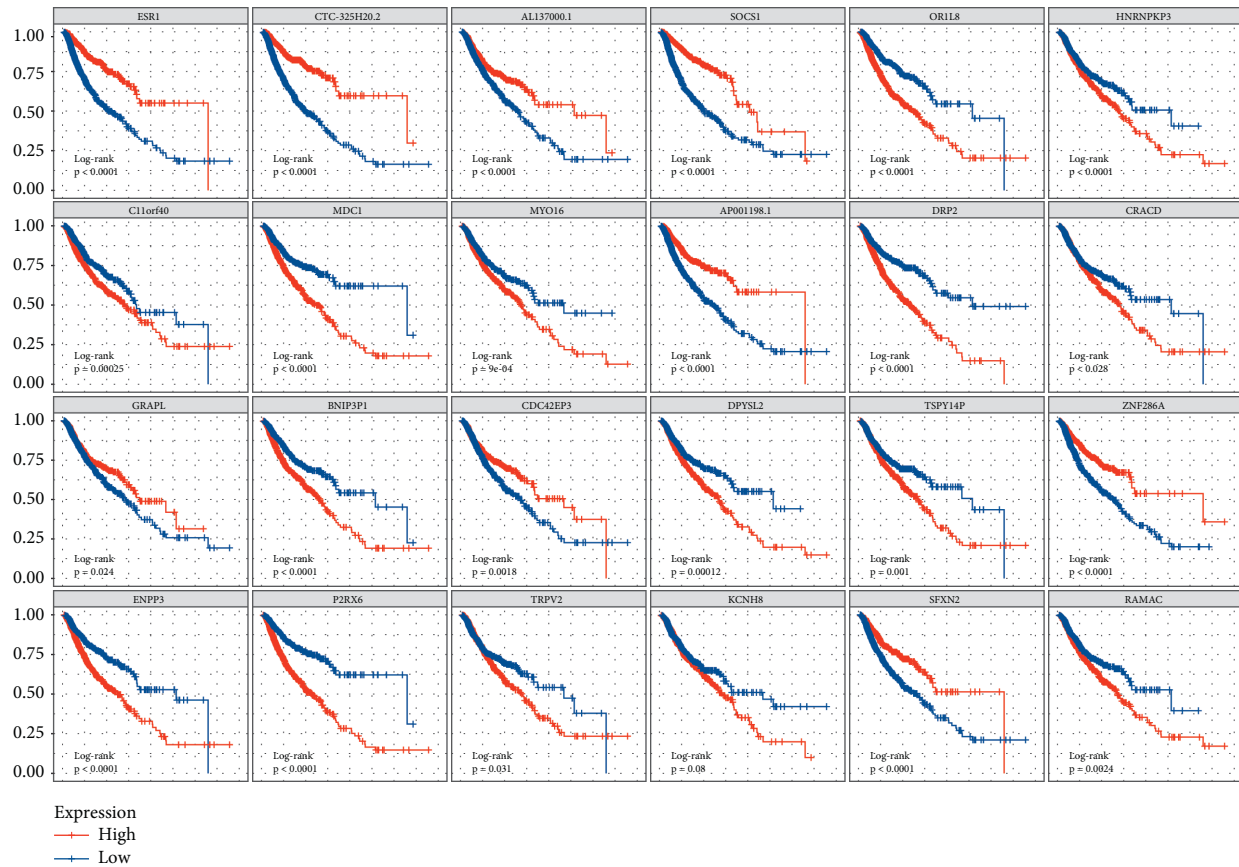


FIGURE 6: Overall survival curve of differential expression between reactive tumor and progressive disease tumor. The abscissa of the survival curve is the observation time and the ordinate is the survival rate. The median expression level of each gene was taken as the threshold, the high expression level group was higher than the median, and the low expression level group was lower than the median. The log-rank test was used to test the statistical significance of the two groups of data.

**2.4. K-Fold Cross Validation.** In the logistic regression model, the data set was divided into a training set and a validation set. In this study, the tumor expression level data set was divided into  $k$  subsets ( $k = 10$ ) by using  $k$ -fold cross validation, and the method was repeated ten times. In each repetition, one subset was randomly selected as the validation set, and the remaining  $k - 1$  subset was used as the training set. The training set data were used for establishing a logistic regression model, and then validation set data were brought into the model for evaluation. In each repetition, each evaluation score was retained, while the model was discarded. After ten repetitions, the model evaluation score was used to summarize the ability of logistic regression model. AUC (area under the curve) was used to evaluate a degree or measure of separability.

**2.5. Visualization of Differential Expression Gene.** The graphics of differentially expressed genes were drawn on the basis of R. Survminer package was used to analyze and visualize the survival curve of differentially expressed genes, and log-rank test was used to calculate the  $p$  value. The overall survival curve was divided into high expression group and low expression group according to the gene expression level. A box plot was drawn using the ggpubr

package to show the distribution of DEG in  $R$  and  $PD$  samples, and the  $p$  value was calculated using the  $t$ -test. The ROC curve of DEGs was drawn by using the pROC software package, and the best threshold to distinguish between  $R$  and  $PD$  samples was marked. Heatmap of DEGs was drawn by using pheatmap package. In the regression model, the samples were sorted from small to large according to the prediction index, and figures of prediction index were drawn by using the ggplot2 package. The best threshold point of AUC was used as a dividing line between  $R$  and  $PD$  samples.

**2.6. Enrichment Pathway Analysis.** Enrichment analysis of the top 100 differential genes was performed by the online tool Metascape (<https://metascape.org/>).

### 3. Result

**3.1. Patient Characteristics.** In order to describe the transcriptome characteristics of human cancer response to radiotherapy, the transcriptome of 1664 TCGA clinical samples from 15 cancer types was analyzed, including bladder urothelial carcinoma (BLCA,  $n = 29$ ), breast invasive carcinoma (BRCA,  $n = 185$ ), cervical squamous cell carcinoma (CESC,  $n = 125$ ), esophageal carcinoma (ESCA,

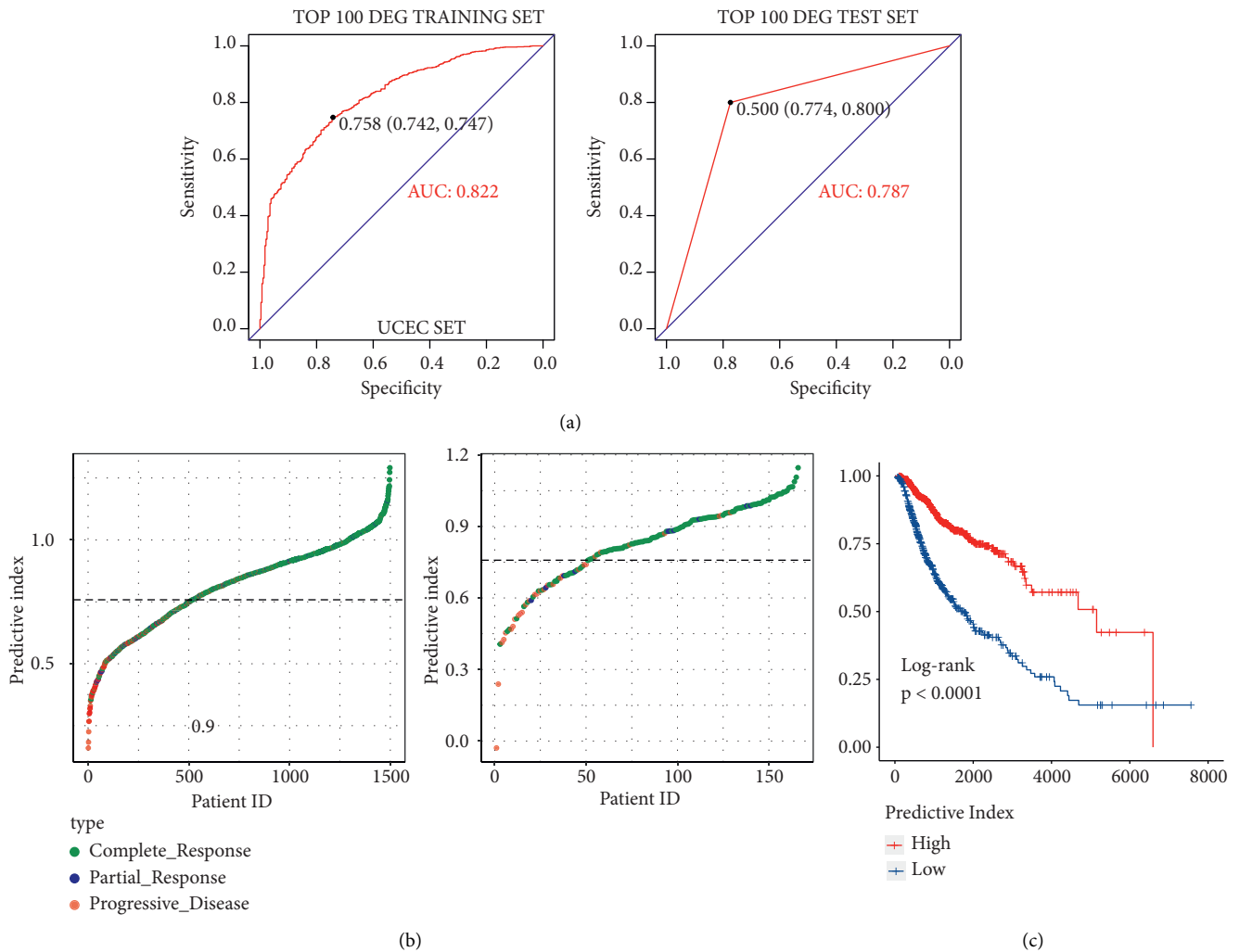


FIGURE 7: ROC curve, predictor chart, and overall survival rate chart of reactive tumor and progressive disease tumor after logistic regression. (a) ROC curve. (b) The ordinate is sorted from small to large according to the prediction index, and the threshold is 0.758. (c) Overall survival curve.

$n = 43$ ), head and neck squamous cell carcinoma (HNSC,  $n = 273$ ), low-grade glioma (LGG,  $n = 208$ ), lung adenocarcinoma (LUAD,  $n = 55$ ), LUSC ( $n = 46$ ), pancreatic cancer (PAAD,  $n = 37$ ), prostate adenocarcinoma (PRAD,  $n = 70$ ), sarcoma (SARC,  $n = 72$ ), cutaneous melanoma (SKCM,  $n = 59$ ), gastric adenocarcinoma (STAD,  $n = 59$ ), thyroid carcinoma (THCA,  $n = 44$ ), and endometrial carcinoma (UCEC,  $n = 233$ ). The response group ( $R$ ) was defined as patients with a relatively good response after RT (such as complete or partial response,  $n = 1350$ ). The progressive disease group ( $PD$ ) was defined as patients with imaging progressive disease after RT ( $n = 314$ ). The cancer types and groups of samples were shown in Figure 1.

**3.2. Identification of Differentially Expressed Genes in  $R$  Group and  $PD$  Group.** First, according to the DEGs selection criteria  $|\log_2 fc| > 1$ ,  $p < 0.05$ , the differentially expressed genes between  $R$  and  $PD$  were identified in 15 analyzed cancer types. Then, the  $t$ -test is carried out using these differential

genes to identify genes associated with  $R$  or  $PD$  (cutoff threshold  $p < 0.05$ ). The heatmap was drawn using genes with the top 100 differential genes (Figure 2). Among the top 100 differential genes, the high expression of 72 genes and the low expression of 28 genes were highly correlated with  $PD$  samples. The functional annotation for these genes (Table S1) provides insight into the underlying biological mechanism leading to  $RT$  resistance. Genes involved in migration, cell proliferation, cell invasion, tumor metastasis, and EMT were significantly upregulated in the  $PD$  group (e.g., MDC1, UCP2, RBM45, BCL9L, P2RX6, RER1, EFNA2, CASK, CERCAM, and PTPRN). Interestingly, MDC1 has been reported as a key regulator of the DNA damage response in higher eukaryotes [26], and UCP2 and RBM45 have been implicated in  $RT$  resistance [27, 28]. Among the top 100 genes, there was a class of genes related to a ubiquitination proteasome hydrolysis system, such as RBM45, TRIM9, PTPRN, RNF123, RNF220, and DTX1. Ubiquitination is an important means of regulating target genes at the protein level. Several noncoding RNAs were also

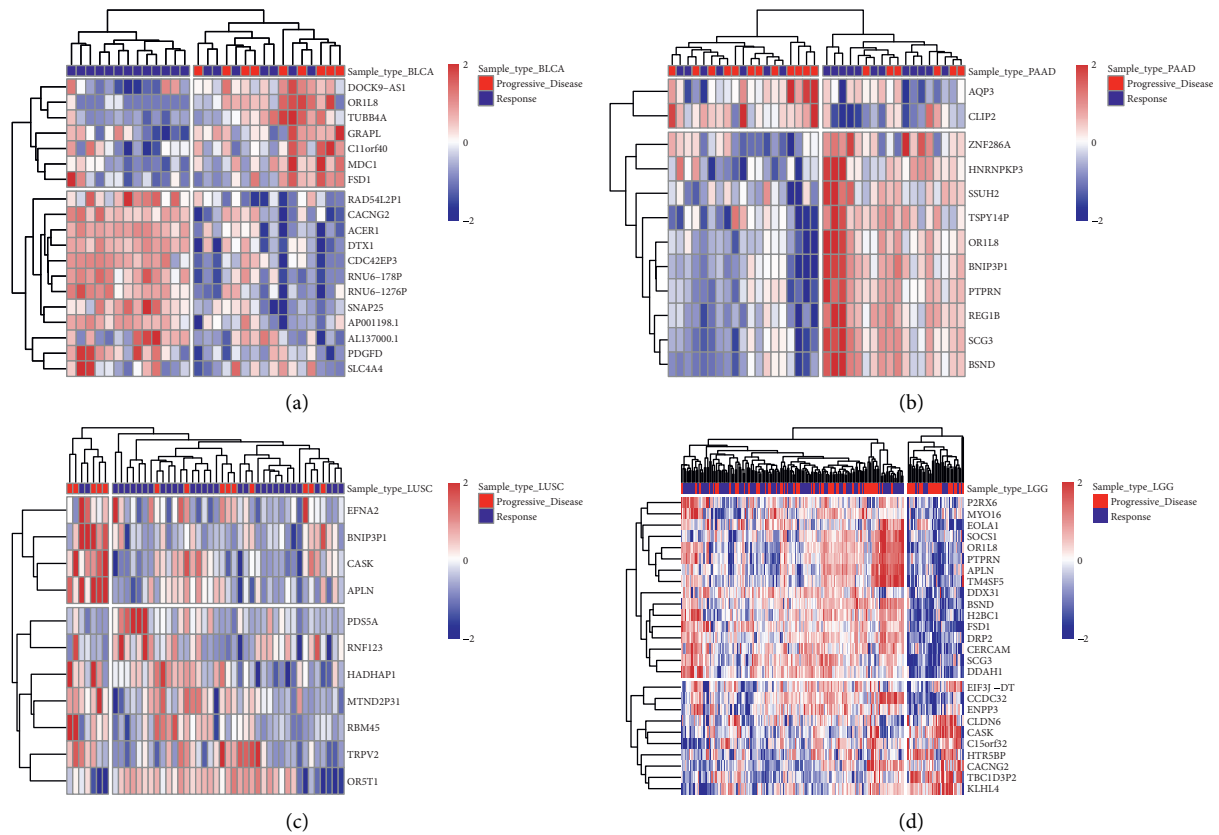


FIGURE 8: Heatmap of differentially expressed genes in different tumor types.

found in this study, including AC104411.1, CTC-325H20.2, AP001198.1, C15ORF32, C11ORF40, MTND2P31, RNU6-1276P, AC002347.1, RNU6-178P, HMG2P32, and EIF3J-DT.

GO analysis of the top 100 differential genes including biological process (BP), molecular function (MF), and oncogenic signatures (OncSig) analysis was conducted. In terms of MF, these genes are significantly enriched in channel activity, ligand-gated cation channel activity, SH3 domain binding, and ubiquitin-like protein transferase activity. BP analysis showed that the gap junction was significantly enriched. OncSig analysis showed that four genes (TRIM9, SCG3, SNAP25, and TUBB4A) were related to overexpression of KRAS. Our analysis revealed that the four genes tend to be highly expressed in PD samples. Their roles in cancer have been studied. For example, TRIM9 [29] and SCG3 [30] have been reported to promote the proliferation of cancer cell, and SNAP25 and TUBB4A were identified as potential prognostic biomarkers for prostate cancer [31] or lung adenocarcinoma [32] (Figure 3)

**3.3. Assessing the Discriminative Power with a Single-Gene Signature.** In this study, we identified the differentially expressed genes. We visualized the differential genes using the boxplot to estimate the discriminative power for R and PD samples with the single-gene signature (Figure 4). These genes show that there is a statistically significant differential

expression between R and PD samples. We further used the receiver operating characteristic (ROC) curve to evaluate the specificity of single-gene signature. As shown in Figure 5, the AUC value of single-gene discrimination model was between 0.615 and 0.7, and OR1L8 gene has the highest specificity (AUC = 0.700), implying that the discriminative power of single-gene signature is not very strong. However, the expression patterns of many genes may reflect the underlying molecular mechanisms of tumor radioresistance. For example, the high expression of MDC1 will increase the ability of DNA damage repair and eventually lead to an increase in radiotherapy resistance. Therefore, we believe that the expression patterns of some of these genes still have clinical predictive power.

#### 3.4. Evaluation of Survival Rate by the Single-Gene Signature.

In order to investigate whether the single-gene signature has the ability to predict the survival rate of patients, we analyzed the overall survival data of the patients. The overall survival curve is shown in Figure 6. Statistically significant correlations were observed between the survival time and the expression in many chosen genes. For example, MDC1 ( $p < 0.0001$ ), EFNA2 ( $p < 0.0001$ ), BCL9L ( $p < 0.0001$ ), RER1 ( $p < 0.0001$ ), and P2RX6 ( $p < 0.0001$ ) tended to be highly expressed in PD samples, and the overall survival rate of the high expression group was significantly lower than that of the low expression group. Conversely, ESR1 ( $p < 0.0001$ ),



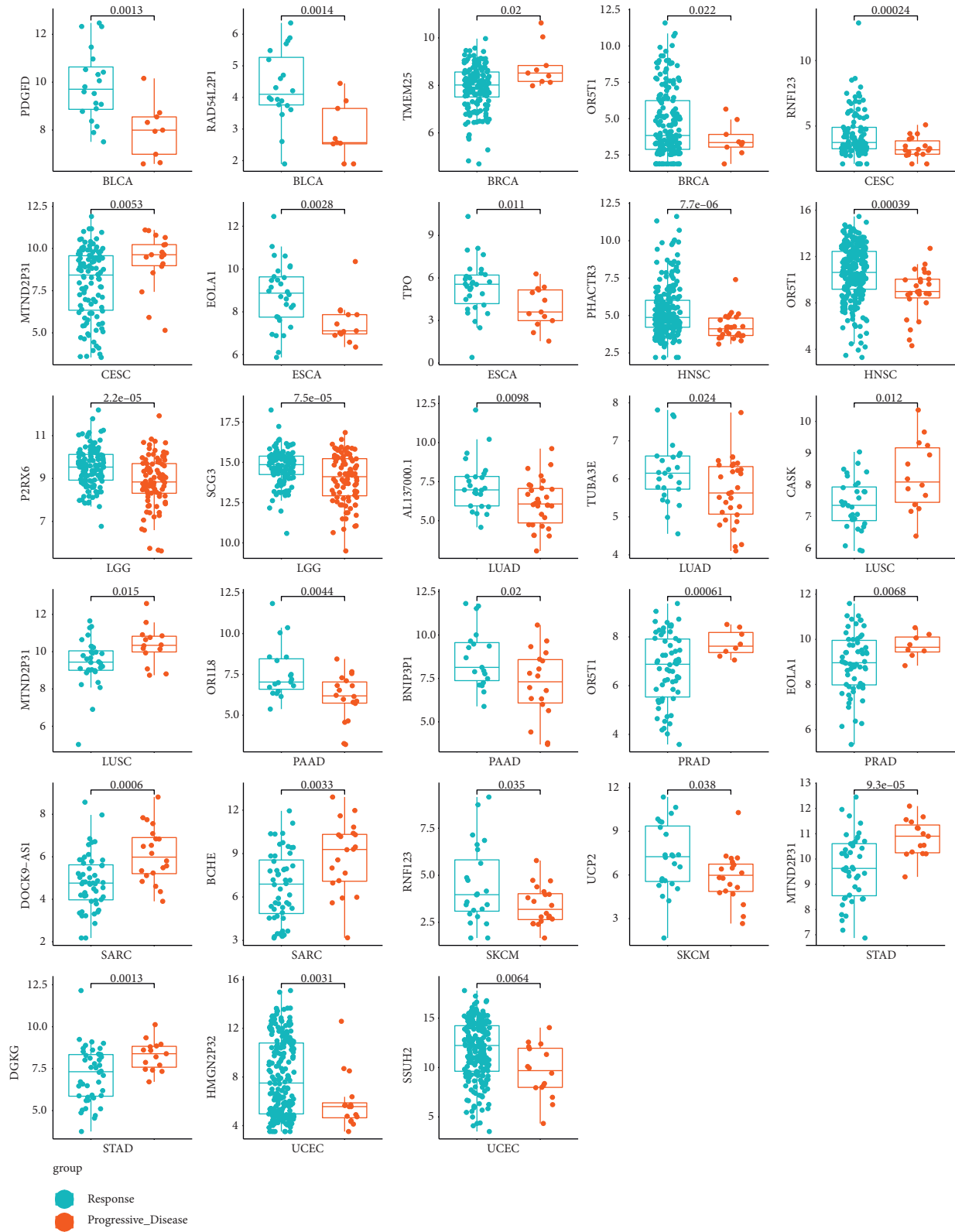


FIGURE 9: Boxplot of differentially expressed genes in different tumor types.

RNF123 ( $p < 0.0001$ ), DTX1 ( $p < 0.0001$ ), and CACNG2 ( $p < 0.0001$ ) tended to be highly expressed in the *R* sample, and the survival rate of the high expression group was

significantly higher than that of the low expression group. These results revealed that *R* group patients had a better survival rate than PD patients.

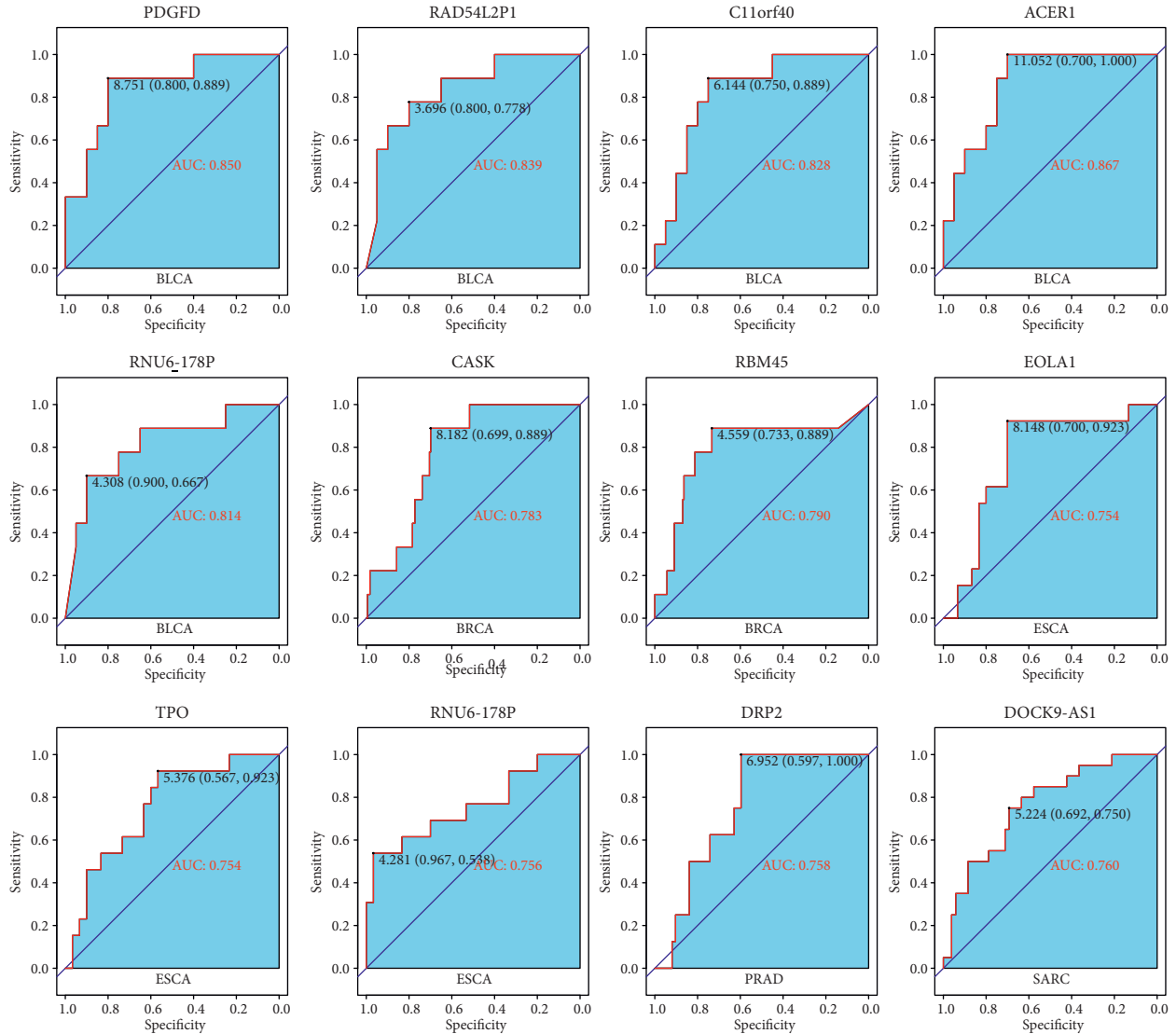


FIGURE 10: ROC map of differentially expressed genes in different tumor types.

**3.5. Using Logistic Regression Model to Distinguish R and PD Samples.** As mentioned above, the discriminative power of single-gene signature is not very strong. Therefore, we developed a binary logistic regression model to classify the samples. The prediction index was used to quantify each sample. We used 90% of the data set as the training set and 10% as the validation set. The ROC curve was drawn according to the prediction index of the regression model. Results are shown in Figure 7. The AUC value reached 0.822 and 0.787 in the training set and validation set at the best threshold score of 0.758, respectively (Figure 7(a)). Compared with the best AUC value of single-gene signature (OR1L8, AUC=0.700), the logistic regression model has better performance for distinguishing R and PD samples. The performance of classifying patients by prediction index is shown by visualizing predicted and actual classifications (Figure 7(b)). The result showed that PD samples were enriched on the left side of the prediction index curve and R samples are enriched on the right side of the prediction

index curve, implying that most of the samples can be classified correctly. We also assessed whether the prediction index can reach a better prediction for the survival rate of patients. As shown in Figure 7(c), the patients with a high prediction index had better survival than those patients with a low prediction index ( $p < 0.0001$ ). This also implied that the RT responding patients had a better survival rate.

**3.6. Performance of Single-Gene Signature in Different Tumor Types.** In this study, the gene signature was obtained from 1664 TCGA clinical tumor samples. The tumors derive from 15 cancer types. However, given that there may be some degree of variation in expression patterns between different tumor types, it is necessary to select the optimal subset for each tumor type from the 100-gene signature. Therefore, we analyzed the discrimination ability of each gene for R and PD samples in each cancer type. Genes with a  $p$  value of less than 0.05 ( $p$  value less than 0.1 in few of tumor types) were

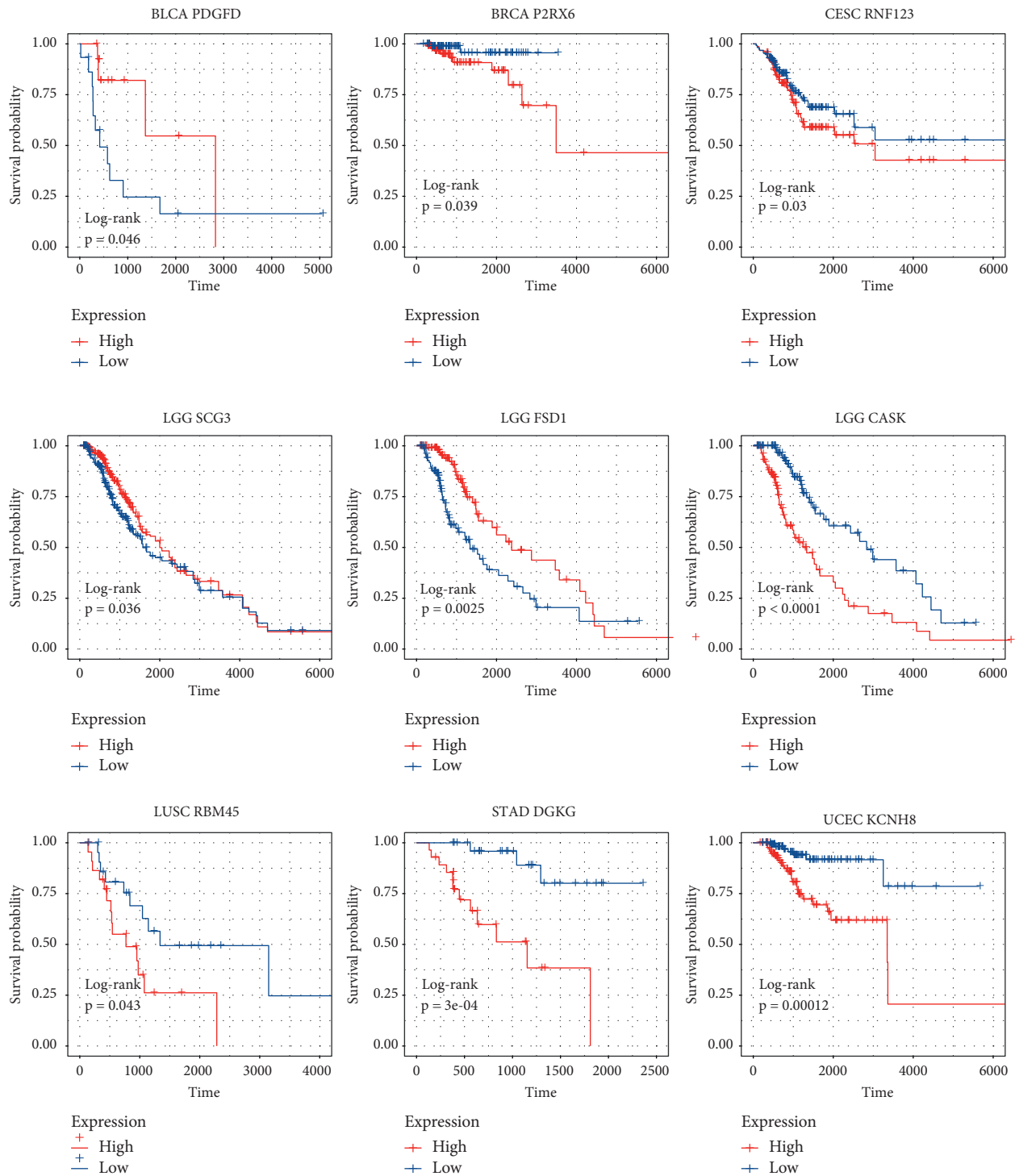


FIGURE 11: Overall survival curve of differentially expressed genes in different tumor types.

chosen to build tumor-specific gene subsets (Table S2). Cluster analysis revealed two distinct expression patterns of gene subsets in several cancer types, such as BLCA, PAAD, LUSC, and LGG (Figure 8). Many genes exhibited good ability to distinguish between *R* and *PD* groups (Figure 9). For example, RNF123 ( $p = 0.00024$ , in CESC), PHACTR3 ( $p = 7.7 E - 06$ , in HNSC), OR5T1 ( $p = 0.00039$ , in HNSC), P2RX6 ( $p = 2.2 E - 05$ , in LGG), and SCG3 ( $p = 7.5 E - 05$ , in

LGG) were expressed higher in *R* samples, while MTND2P31 ( $p = 0.0053$ , in CESC), OR5T1 ( $p = 0.00061$ , in PRAD), EOLA1 ( $p = 0.0068$ , in PRAD), BCHE ( $p = 0.0033$ , in SARC), MTND2P31 ( $p = 9.3 E - 05$ , in STAD), and DGKG ( $p = 0.0013$ , in STAD) were expressed higher in *PD* samples. Analysis by ROC curve also revealed several genes had high specificity of distinction (Figure 10), such as ACER1 (AUC = 0.867, in BLCA), RBM45 (AUC = 0.790, in BRCA),

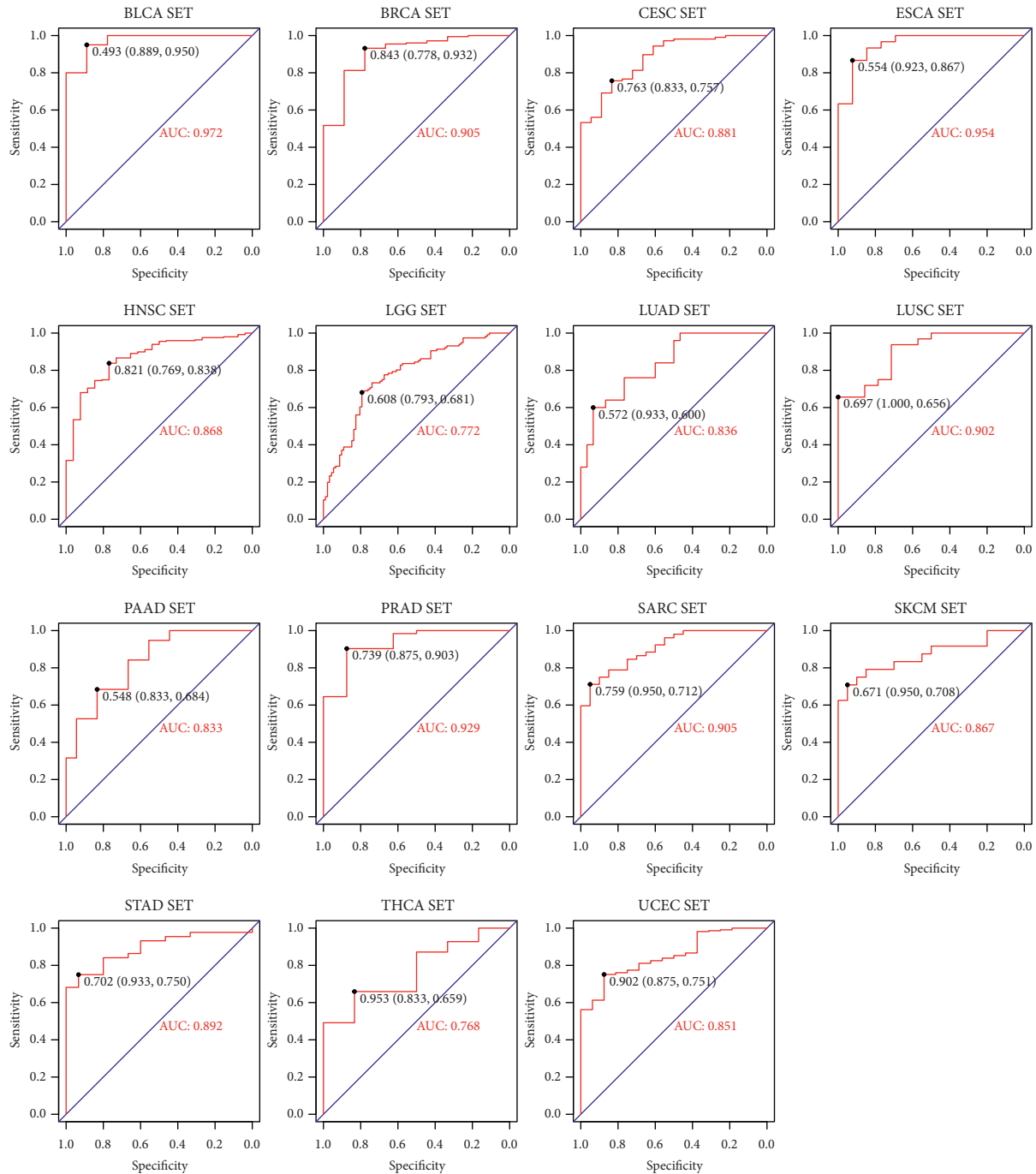


FIGURE 12: Logistic regression prediction ROC diagram of different tumor types.

RNU6 – 178P (AUC = 0.756, in ESCA), and DOCK9 – AS1 (AUC = 0.760, in SARC). Several genes also exhibited the predictive power for survival of patients (Figure 11). The higher expression of P2RX6 ( $p = 0.039$ , in BRCA), RNF123 ( $p = 0.03$ , in CESC), CASK ( $p < 0.0001$ , in LGG), DGKG ( $p = 3E-04$ , in STAD), RBM45 ( $p = 0.043$ , in LUSC), and KCN8 ( $p = 0.00012$ , in UCEC) was associated with worse survival, while the higher expression of PDGFD ( $p = 0.046$ , in BLCA), SCG3 ( $p = 0.036$ , in LGG) and FSD1 ( $p = 0.0025$ , in LGG) was associated with better survival.

3.7. Using Logistic Regression Model to Distinguish R and PD Samples in Different Tumor Types. Analysis of ROC curve revealed that the higher specificity was achieved in each cancer type by using the prediction index (Figure 12). For example, the AUC value of BLCA obtained by the prediction index was higher than from ACER1 (0.972 vs 0.867). Our results showed that most of the samples can be correctly classified in each cancer type (Figure 13). We also investigated whether the predictive model could provide better performance for the prediction of survival rate for each

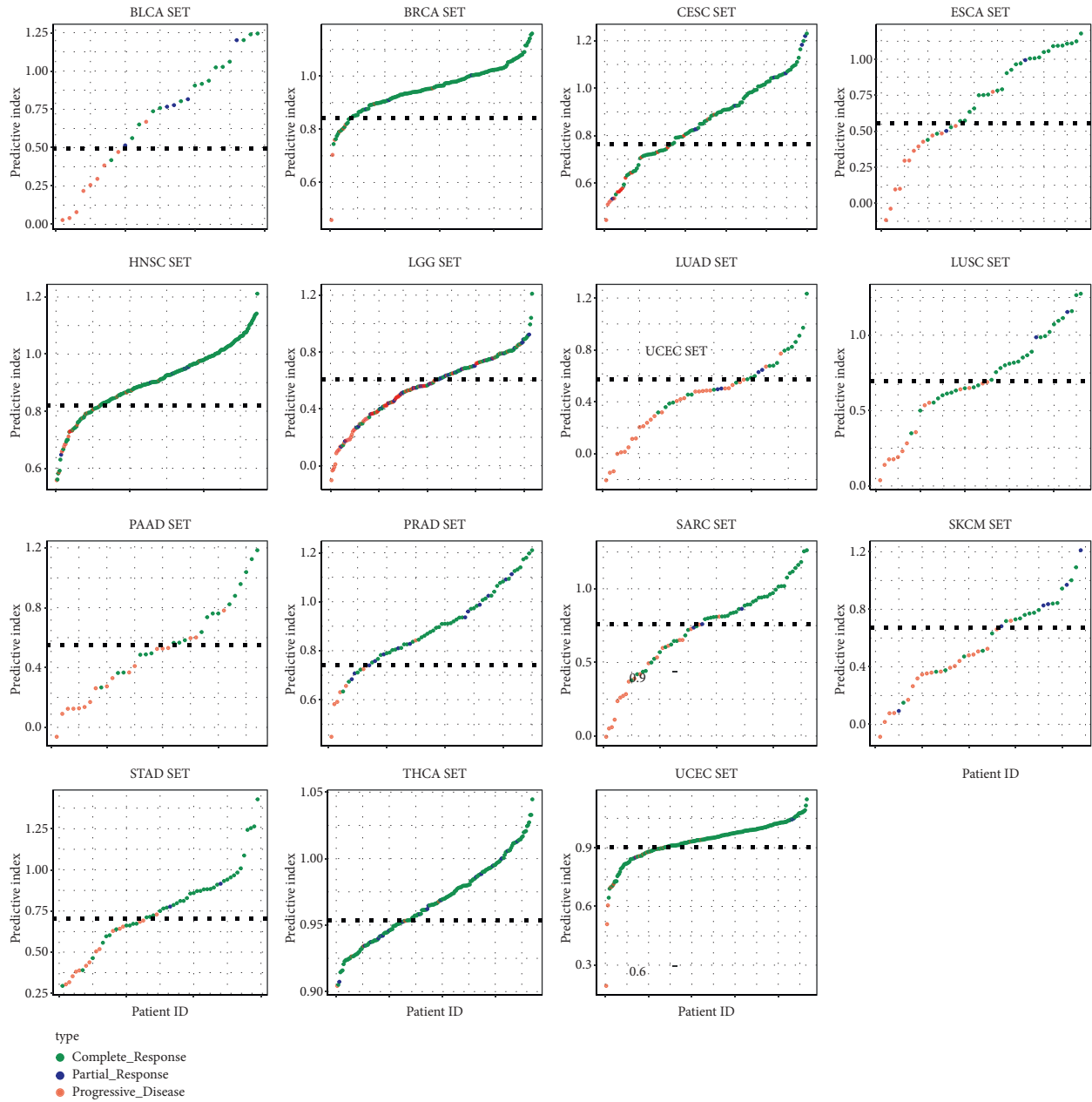


FIGURE 13: Logistic regression predictors of different tumor types.

cancer type. As shown in Figure 14, patients with a high prediction index were associated with better survival.

#### 4. Discussion

In this study, we identified the differentially expressed genes between RT responder and nonresponder. Among them, MDC1 [26], UCP2 [27], and RBM45 [28] have been demonstrated to be involved in the DNA damage pathway and radiosensitivity. MDC1 [26] was identified as a component of DNA repair complex, controlling the damage-induced cell cycle arrest checkpoint. Cells lacking MDC1 are sensitive to ionizing radiation. Our study showed that MDC1 was expressed higher in the PD group, which might lead to

stronger reparability in PD patients. UCP2 [27] is a mitochondrial transporter, which can produce proton leakage on the inner membrane of mitochondria, thus uncoupling oxidative phosphorylation and ATP synthesis. The previous studies demonstrated that irradiation treatment can increase the expression level of UCP2, and silencing of UCP2 increased the radiosensitivity of HeLa cells and led to increased apoptosis, cell cycle arrest in G2/M, and mitochondrial ROS. Our results showed that the expression of UCP2 was significantly upregulated in the PD group, supporting that UCP2 has a role in radioresistance. RBM45 [28], also named DRB1, was recently found to be a FUS-interacting RBP. The previous study demonstrated that silencing of RBM45 led to a decreased efficiency in DSBs repair. Consistent with the

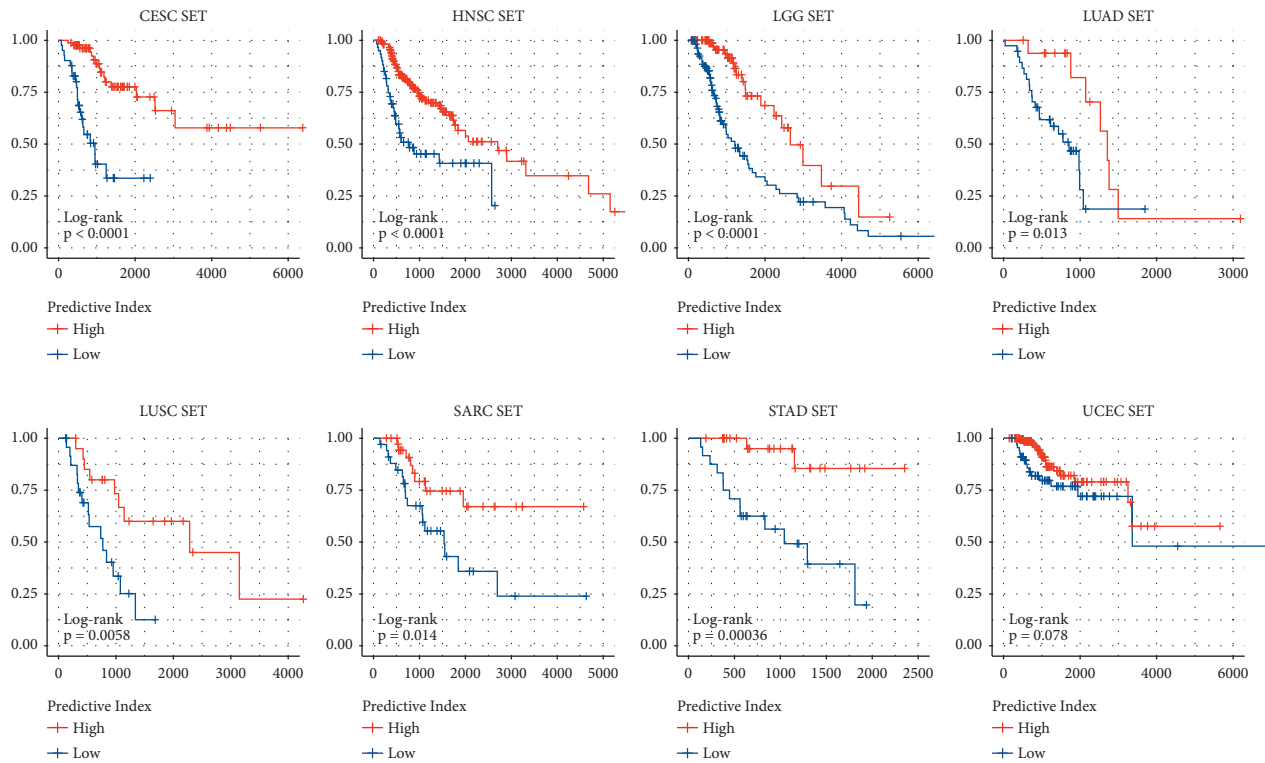


FIGURE 14: Survival curves of different tumor types.

results of the previous study, our results showed that RBM45 was expressed higher in the PD group.

Besides these three genes, several genes involving migration, cell proliferation, cell invasion, and EMT were identified in our study. For example, BCL9L [33], P2RX6 [34], RER1 [35], EFNA2 [36], CASK [37], CERCAM [38], and PTPRN [39] were demonstrated to promote EMT, metastasis, and invasion of tumor cells. Six genes involving ubiquitination proteasome process, including RBM45, TRIM9, PTPRN, RNF123, RNF220, and DTX1, were also identified. Ubiquitination plays an important role in innate and postnatal regulation of cell differentiation and cell survival. Previous studies have reported that ubiquitination-related proteins were involved in the regulation of DNA damage pathways and influenced the radiosensitivity of tumor cells [40, 41]. For example, Santra et al. found that E3 ubiquitin ligase FBXO31 mediates the degradation of cyclin D1 through ubiquitination and proteasome-mediated protein degradation pathway [40]. Knockdown of FBXO31 could prevent cells from undergoing efficient G1 arrest following  $\gamma$ -irradiation and greatly increased the sensitivity to DNA damage [41]. The differentially expressed genes also include several noncoding RNAs in our study. UFC1 is a long-chain noncoding RNA and was expressed higher in the PD group. Previous study showed that UFC1 was elevated and predicted poor prognosis of gastric cancer (GC), and knockdown of UFC1 inhibited the proliferation, migration, and invasion of GC cells [42].

Our study identified a series of genes that were differentially expressed in RT responders and nonresponders, providing useful clues for studying the molecular

mechanisms of tumor radioresistance. Those genes that were overexpressed in RT nonresponders could be potential targets for radiosensitization. The prediction models and gene signatures identified here also have the potential clinical application. A targeted mRNA sequencing technology based on the gene signatures identified in this study could be developed and used to detect mRNA levels in clinical samples. By analyzing the gene signatures, patients who will benefit from RT are identified, which could reduce the number of patients receiving unnecessary treatment and greatly reduce the cost of oncology treatment.

## Abbreviations

RT:	Radiotherapy;
EMT:	Epithelial-mesenchymal transition
AUC:	Area under the curve;
DSBs:	DNA double-strand breaks
RNA-Seq:	High-throughput sequencing methods for mRNA;
TCGA:	The Cancer Genome Atlas
BLCA:	Bladder urothelial carcinoma;
BRCA:	Invasive breast cancer
CESC:	Cervical squamous cell carcinoma;
ESCA:	Esophageal cancer
HNSC:	Head and neck squamous cell carcinoma;
LGG:	Low-grade glioma
LUAD:	Lung adenocarcinoma;
LUSC:	Lung squamous cell carcinoma
PAAD:	Pancreatic cancer;
PRAD:	Prostate adenocarcinoma

SARC: Sarcoma;  
 SKCM: Cutaneous melanoma  
 STAD: Gastric adenocarcinoma;  
 THCA: Thyroid carcinoma  
 UCEC: Endometrial carcinoma;  
 R: Response  
 PD: Progressive disease;  
 DEGs: Differentially expressed genes  
 ROC: Receiver operating characteristic curve;  
 DDR: DNA-damage response.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Yu Xu and Chao Tang contributed equally to this work.

## Acknowledgments

This work was supported in part by projects (cstc2020jxjl130015, cstc2021jcyj-msxmX0767, cstc2019jcyj-msxmX0671, and cstc2020jcyj-msxmX1093) from the National Natural Science Foundation of China (no. 82073347) and the Integrated Innovation and Application of Key Technologies for Precise Prevention and Treatment of Primary Lung Cancer (no. 2019ZX002) from the Chongqing Municipal Health Committee.

## Supplementary Materials

Table S1: Molecular functional annotation of top 100 differential genes in all cancer species analyzed. Table S2: DEGs used by cancer species to distinguish between R and PD samples. (*Supplementary Materials*)

## References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [3] E. J. Moding, M. B. Kastan, and D. G. Kirsch, "Strategies for optimizing the response of cancer and normal tissues to radiation," *Nature Reviews Drug Discovery*, vol. 12, no. 7, pp. 526–542, 2013.
- [4] U. S. Srinivas, B. W. Q. Tan, B. A. Vellayappan, and A. D. Jeyasekharan, "ROS and the DNA damage response in cancer," *Redox Biology*, vol. 25, Article ID 101084, 2019.
- [5] L. Chang, P. Graham, J. Hao et al., "Cancer stem cells and signaling pathways in radioresistance," *Oncotarget*, vol. 7, no. 10, pp. 11002–11017, 2016.
- [6] K. Savitsky, A. Bar-Shira, S. Gilad et al., "A single ataxia telangiectasia gene with a product similar to PI-3 kinase," *Science*, vol. 268, no. 5218, pp. 1749–1753, 1995.
- [7] M. Ahmed, L. Li, C. Pinnix et al., "ATM mutation and radiosensitivity: an opportunity in the therapy of mantle cell lymphoma," *Critical Reviews In Oncology-Hematology*, vol. 107, pp. 14–19, 2016.
- [8] J. L. Bernstein, R. W. Haile, M. Stovall et al., "Radiation exposure, the ATM gene, and contralateral breast cancer in the women's environmental cancer and radiation epidemiology study," in *Journal of the National Cancer Institute: Journal of the National Cancer Institute*, P. Concannon, Ed., vol. 102, no. 7, pp. 475–483, 2010.
- [9] M. Choi, T. Kipps, and R. Kurzrock, "ATM mutations in cancer: therapeutic implications," *Molecular Cancer Therapeutics*, vol. 15, no. 8, pp. 1781–1791, 2016.
- [10] M. J. Kerins and A. Ooi, "A catalogue of somatic NRF2 gain-of-function mutations in cancer," *Scientific Reports*, vol. 8, no. 1, Article ID 12846, 2018.
- [11] Y. Jeong, N. T. Hoang, A. Lovejoy et al., "Role of KEAP1/NRF2 and TP53 mutations in lung squamous cell carcinoma development and radiation resistance," *Cancer Discovery*, vol. 7, no. 1, pp. 86–101, 2017.
- [12] Z. Gao, E. H. Sarsour, A. L. Kalen, L. Li, M. G. Kumar, and P. C. Goswami, "Late ROS accumulation and radiosensitivity in SOD1-overexpressing human glioma cells," *Free Radical Biology and Medicine*, vol. 45, no. 11, pp. 1501–1509, 2008.
- [13] X. Huang, S. Taeb, S. Jahangiri et al., "miRNA-95 mediates radioresistance in tumors by targeting the sphingolipid phosphatase SGPP1," *Cancer Research*, vol. 73, no. 23, pp. 6972–6986, 2013.
- [14] N. Mercatelli, V. Coppola, D. Bonci et al., "The inhibition of the highly expressed miR-221 and miR-222 impairs the growth of prostate carcinoma xenografts in mice," *PLoS One*, vol. 3, no. 12, Article ID e4029, 2008.
- [15] K. Hatano, B. Kumar, Y. Zhang et al., "A functional screen identifies miRNAs that inhibit DNA repair and sensitize prostate cancer cells to ionizing radiation," *Nucleic Acids Research*, vol. 43, no. 8, pp. 4075–4086, 2015.
- [16] C. G. Xu, M. F. Yang, J. X. Fan, and W. Wang, "MiR-30a and miR-205 are downregulated in hypoxia and modulate radiosensitivity of prostate cancer cells by inhibiting autophagy via TP53INP1," *European Review for Medical and Pharmacological Sciences*, vol. 20, no. 8, pp. 1501–1508, 2016.
- [17] Z. Tao, S. Xu, H. Ruan et al., "MiR-195/-16 family enhances radiotherapy via T cell activation in the tumor microenvironment by blocking the PD-L1 immune checkpoint," *Cellular Physiology and Biochemistry*, vol. 48, no. 2, pp. 801–814, 2018.
- [18] A. Mao, Y. Liu, Y. Wang et al., "miR-449a enhances radiosensitivity through modulating pRb/E2F1 in prostate cancer cells," *Tumor Biology*, vol. 37, no. 4, pp. 4831–4840, 2016.
- [19] Z. Xu, Y. Zhang, J. Ding et al., "miR-17-3p downregulates mitochondrial antioxidant enzymes and enhances the radiosensitivity of prostate cancer cells," *Molecular Therapy - Nucleic Acids*, vol. 13, pp. 64–77, 2018.
- [20] A. C. Mueller, D. Sun, and A. Dutta, "The miR-99 family regulates the DNA damage response through its target SNF2H," *Oncogene*, vol. 32, no. 9, pp. 1164–1172, 2013.

- [21] X. Zhu, Y. Wang, L. Tan, and X. Fu, "The pivotal role of DNA methylation in the radio-sensitivity of tumor radiotherapy," *Cancer Medicine*, vol. 7, no. 8, pp. 3812–3819, 2018.
- [22] Z. Liu, H. Chen, Y. Xia, A. L. Kwan, and Z. Chen, "Relationship between methylation status of ERCC1 promoter and radio-sensitivity in glioma cell lines," *Cell Biology International*, vol. 5, no. 10, pp. 156–159, 2007.
- [23] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.
- [24] J. Costa-Silva, D. Domingues, and F. M. Lopes, "RNA-Seq differential expression analysis: an extended review and a software tool," *PLoS One*, vol. 12, no. 12, Article ID e0190152, 2017.
- [25] Q. Q. Wang, S. C. Yu, X. Qi et al., "Overview of logistic regression model analysis and application," *Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine]*, vol. 53, no. 9, pp. 955–960, 2019.
- [26] G. S. Stewart, B. Wang, C. R. Bignell, A. M. R. Taylor, and S. J. Elledge, "MDC1 is a mediator of the mammalian DNA damage checkpoint," *Nature*, vol. 421, no. 6926, pp. 961–966, 2003.
- [27] C. H. Liu, Z. H. Huang, X. Y. Dong et al., "Inhibition of uncoupling protein 2 enhances the radiosensitivity of cervical cancer cells by promoting the production of reactive oxygen species," *Oxidative Medicine and Cellular Longevity*, vol. 2020, pp. 1–13, Article ID 5135893, 2020.
- [28] J. Gong, M. Huang, F. Wang et al., "RBM45 competes with HDAC1 for binding to FUS in response to DNA damage," *Nucleic Acids Research*, vol. 45, no. 22, pp. 12862–12876, 2017.
- [29] S. Venuto and G. Merla, "E3 ubiquitin ligase TRIM proteins, cell cycle and mitosis," *Cells*, vol. 8, no. 5, pp. 510–525, 2019.
- [30] Y. Wang, M. Cui, X. Cai et al., "The oncoprotein HBXIP up-regulates SCG3 through modulating E2F1 and miR-509-3p in hepatoma cells," *Cancer Letters*, vol. 352, no. 2, pp. 169–178, 2014.
- [31] L. Di, M. Gu, Y. Wu et al., "SNAP25 is a potential prognostic biomarker for prostate cancer," *Cancer Cell International*, vol. 22, no. 1, 144 pages, 2022.
- [32] S. Nakasone, A. Suzuki, H. Okazaki et al., "Predictive markers based on transcriptome modules for vinorelbine-based adjuvant chemotherapy for lung adenocarcinoma patients," *Lung Cancer*, vol. 158, pp. 115–125, 2021.
- [33] X. Wang, M. Feng, T. Xiao et al., "BCL9/BCL9L promotes tumorigenicity through immune-dependent and independent mechanisms in triple negative breast cancer," *Oncogene*, vol. 40, no. 16, pp. 2982–2997, 2021.
- [34] D. Gong, J. Zhang, Y. Chen et al., "The m6A-suppressed P2RX6 activation promotes renal cancer cells migration and invasion through ATP-induced Ca<sup>2+</sup> influx modulating ERK1/2 phosphorylation and MMP9 signaling pathway," *Journal of Experimental & Clinical Cancer Research*, vol. 38, no. 1, p. 233, 2019.
- [35] S. Chen, J. Zhang, J. Chen et al., "RER1 enhances carcinogenesis and stemness of pancreatic cancer under hypoxic environment," *Journal of Experimental & Clinical Cancer Research*, vol. 38, no. 1, 15 pages, 2019.
- [36] Y. Zhao, C. Cai, M. Zhang et al., "Ephrin-A2 promotes prostate cancer metastasis by enhancing angiogenesis and promoting EMT," *Journal of Cancer Research and Clinical Oncology*, vol. 147, no. 7, pp. 2013–2023, 2021.
- [37] J. Qu, Y. Zhou, Y. Li, J. Yu, and W. Wang, "CASK regulates Notch pathway and functions as a tumor promoter in pancreatic cancer," *Archives of Biochemistry and Biophysics*, vol. 701, Article ID 108789, 2021.
- [38] Y. Zuo, X. Xu, M. Chen, and L. Qi, "The oncogenic role of the cerebral endothelial cell adhesion molecule (CERCAM) in bladder cancer cells in vitro and in vivo," *Cancer Medicine*, vol. 10, no. 13, pp. 4437–4450, 2021.
- [39] X. Song, X. Jiao, H. Yan et al., "Overexpression of PTPRN promotes metastasis of lung adenocarcinoma and suppresses NK cell cytotoxicity," *Frontiers in Cell and Developmental Biology*, vol. 9, Article ID 622018, 2021.
- [40] M. K. Santra, N. Wajapeyee, and M. R. Green, "F-box protein FBXO31 mediates cyclin D1 degradation to induce G1 arrest after DNA damage," *Nature*, vol. 459, no. 7247, pp. 722–725, 2009.
- [41] Y. Lu, J. Li, D. Cheng et al., "The F-box protein FBXO44 mediates BRCA1 ubiquitination and degradation," *Journal of Biological Chemistry*, vol. 287, no. 49, pp. 41014–41022, 2012.
- [42] X. Zhang, W. Liang, J. Liu et al., "Long non-coding RNA UFC1 promotes gastric cancer progression by regulating miR-498/Lin28b," *Journal of Experimental & Clinical Cancer Research*, vol. 37, no. 1, p. 134, 2018.