

# Predicting radiotherapy response for patients with soft tissue sarcoma by developing a molecular signature

ZAXIANG TANG<sup>1-3</sup>, QINGHUA ZENG<sup>4</sup>, YAN LI<sup>3</sup>, XINYAN ZHANG<sup>3</sup>,  
MARK J. SUTO<sup>4</sup>, BO XU<sup>4</sup> and NENGJUN YI<sup>3</sup>

<sup>1</sup>Department of Biostatistics, School of Public Health, Medical College of Soochow University, Suzhou, Jiangsu 215123; <sup>2</sup>Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, and Center for Genetic Epidemiology and Genomics, Medical College of Soochow University, Suzhou, Jiangsu 215123, P.R. China; <sup>3</sup>Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294; <sup>4</sup>Drug Discovery Division, Southern Research Institute, Birmingham, AL 35294, USA

Received March 23, 2017; Accepted August 28, 2017

DOI: 10.3892/or.2017.5999

**Abstract.** Soft tissue sarcomas are rare and aggressive tumors arising from connective tissues. Adjuvant radiotherapy is a commonly used treatment approach for the majority of sarcomas. We attempted to identify a gene signature that can predict radiosensitive patients who are most likely to have a better treatment response from radiotherapy, compared with disease progression. Using the publicly available data of soft tissue sarcoma from The Cancer Genome Atlas, we developed a cross-validation procedure to identify a predictive gene signature for radiosensitivity. The results showed that the predicted radiosensitive patients who received radiotherapy had significantly improved treatment response. We further provide supportive evidence to validate our sensitivity prediction. Results showed that the predicted radiosensitive patients who received radiotherapy had significantly improved survival than patients who did not. ROC analysis showed that the developed gene signature had a powerful prediction on treatment response. We further found that predicted radiosensitive patients who received radiotherapy had a significantly reduced rate of new tumor events. Finally, we validated our gene signature using a hierarchical cluster analysis, and found that the predicted sensitivities were well-matched with results from the cluster analysis. These results are consistent with our expectation, suggesting that the identified gene signature and radiosensitivity prediction are effective. The genes involved in the signature may provide

a molecular basis for prognostic studies and radiotherapy target discovery.

## Introduction

Soft tissue sarcomas are rare and aggressive malignancies that arise in any of the mesodermal tissues of the extremities, trunk and retroperitoneum, or head and neck. It includes approximately 50 histological types and account for approximately 1% of all adult cancers. The reported incidence rates range from 1.8 to 5 per 100,000 per year (1). Estimated new cases and deaths from soft tissue sarcoma in the USA in 2016 are 12,310 and 4,990, respectively (2). Sarcoma presenting at an advanced stage has a dismal prognosis. The patients survival has slightly improved over the last 20 years. A recent report from the National Cancer Institute shows that the 5-year relative survival rate for soft tissue sarcoma is approximately 65% (3). Optimal systemic therapy options for sarcoma remain limited.

Adjuvant radiotherapy, as one of the major modalities of cancer therapy, playing an important role in integrated multimodality treatment of sarcoma. Radiotherapy technology for soft tissue sarcoma has significantly advanced over the past 50 years (4). However, due to the complex heterogeneity of sarcoma, the treatment response of radiotherapy is different in patients. A retrospective study shows that preoperative radiotherapy might be not suitable for all patients with primary soft tissue sarcoma of the limbs (5). Late and chronic toxicities of radiotherapy, such as severe induration, loss of subcutaneous tissue, subcutaneous fibrosis, hair loss, fractures, difficulty in thinking and headaches are often concerns (4,6-8). The clinical benefit of radiotherapy on soft tissue sarcomas has not been observed significantly (9-11). We argue that if we can identify a molecular signature predictive of radiotherapeutic sensitivity and response, we might be able to improve local control, side-effects and overall survival in patients with soft tissue sarcoma.

Over the last 15 years, personalized medicine has evolved whereby clinical decisions are adjusted according to a patient's molecular profile. Genomic technologies such as microarrays provide powerful tools for identifying a genetic signature for

---

*Correspondence to:* Dr Nengjun Yi, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

E-mail: nyi@uab.edu

Dr Bo Xu, Department of Oncology, Southern Research Institute, Birmingham, AL 35205, USA

E-mail: bxu@southernresearch.org

**Key words:** gene signature, prediction, radiosensitivity, treatment response, sarcoma

patients who are most likely to benefit from radiotherapy. Many gene signatures with specificity in terms of prognosis or prediction of response have been developed and validated in the past decades for radiotherapy indication. The gene signature has been used to predict sensitive patients in many cancer types including glioblastoma, cervical, breast, colorectal, head and neck cancer (12-18), but very few in sarcoma.

In the present study, RNAseq data for soft tissue sarcoma from The Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov/>) was used to develop radiosensitive gene signature for predicting radiosensitive patients, defined as patients with relatively better response (e.g., complete response, partial response, or stable disease) after treatment, compared with disease progression. Since sarcoma is a rare disease, it is difficult, if not impossible, to find another independent dataset and enough RNAseq data to do independent-sample validation. In addition, due to the limitation of the small sample size in TCGA dataset, it is not ideal to do split-sample validation. To overcome these difficulties, we performed a cross-validated adaptive signature design that combined the gene signature development and the validation test in a single dataset, as introduced by Freidlin and Simon (19) and Freidlin *et al.* (20). Following this novel framework, we improved this procedure and model to develop a radiosensitive gene signature for predicting radiosensitive patients in the rare sarcoma data.

## Material and methods

**Study samples.** All data including clinical information and RNAseq expression were downloaded from TCGA (update at March 2016). Firstly, we combined the clinical information, including treatment outcome, radiotherapy, chemotherapy, new tumor event, survival time, and other clinical information available for 261 patients downloaded from TCGA. Secondly, we filtered the records with missing treatment response and radiotherapy information, which reduced our sample to 219 patients. Expression data included 259 patients for 20,502 genes with clear gene names after removing duplicate patients from raw data with 265 samples. We calculated the variance of expression for each gene, and kept the genes with variance of >20% quantile. We filtered the genes with a maximum expression value of 10 as they showed almost no expression. Furthermore, genes with proportion of zero expression >50% were also removed. Thirdly, we merged standardized expression data with clinical information, then obtained 218 patients with 15,702 genes expression profiles for final analysis. Finally, missing values in clinical data were filled by multiple imputation using the R package mice. The cleaned clinical data are summarized in Table I.

### Methods

**Gene signature development.** In the present study, the radiosensitive patients are defined as a group of patients who have relatively better treatment response if they receive radiotherapy, including complete response, partial response and stable disease, compared with disease progression. To predict radiosensitive patients by molecular signature, we used the following modeling assumptions: there is a subset of  $S$  predictive (sensitive) genes that significantly interact with radiotherapy. The treatment response of radiotherapy is influ-

enced by these predictive genes through the following logistic model:

$$\log(p/(1-p)) = \mu + t\lambda + x_1b_1 + x_2b_2 + \dots + x_sb_s + rx_1i_1 + rx_2i_2 + \dots + rx_si_s$$

where  $\mu$  is the intercept;  $\lambda$  is the effect of radiotherapy that all patients experience regardless of their gene expression levels;  $t$  is an indicator for radiotherapy with 1 indicating radiotherapy and 0 otherwise;  $b_1$  to  $b_s$  are the main effects for these sensitive genes;  $x_1$  to  $x_s$  denote the expression level of these sensitive genes;  $i_1$  to  $i_s$  are radiotherapy-gene interaction effects that reflect the degree to which the effect of radiotherapy on treatment response is influenced by the expression levels of sensitive genes.

If the effects of radiotherapy-gene interaction are negative, patients who overexpress the sensitive genes will have a positive response after radiotherapy compared with no radiotherapy. We assume that a subgroup of patients overexpresses some (but not necessarily all) of the sensitivity genes. Then, if the total odds ratio (OR) is less than a preset threshold value (such as <1), this suggests that radiotherapy would help reduce the risk of disease progression for these patients, and thus, these patients are called radiosensitive patients. The above logistic model is different from our previous method (21), where a Cox model was used for predicting survival.

**Cross-validation procedure.** The procedure, a novel cross-validated adaptive signature design, was developed to identify sensitive patients in one clinical trial (19-21). Following their framework, we improved this approach to develop a radiosensitive gene signature for the current sarcoma data, described by the following two-step procedure.

**Step 1. Training step.** Split the data into  $K$  ( $K=10$  used in the present study) parts with the same sample size randomly. Then,  $(K-1)$  parts are used to fit model as training data. The radiosensitive patients are predicted the in the  $k$ -th part (validation data). In the training data, for each gene  $j$ , fit a logistic model:  $\log(p/(1-p)) = \mu + t\lambda + x_jb_j + rx_ji_j$ . Then, the P-values for  $i_j$  were used to rank the genes. The top significant  $g$  genes were used to build a gene signature, by calculating an index, called nominal OR (nOR) using the following equation:

$$\exp[t\lambda + \sum_{j=1}^g (x_jb_j + rx_ji_j)]$$

for patients in the validation data ( $k$ -th part). Here,  $\lambda$  could be the value averaged over the estimates from  $g$  single gene model. Patient in the validation set who has nOR lower than a specified threshold  $R$  ( $R<1$ ) will be predicted as radiosensitive patient.

**Step 2. Validation step.** Cycling the above procedures from  $k=1$  to 10 in turn. Each patient only appears once in one of the validation data. After the cross-validation, each patient is predicted as either radiosensitive or not. For radiosensitive patients, univariate logistic analysis is then performed to test the association between radiotherapy and treatment outcome. A significant result with a small OR (<1) will indicate that radiotherapy is beneficial for predicted radiosensitive patients, then the  $g$  genes signature and the sensitivity prediction is considered effective and reasonable.

In the above procedure, there are two key tuning parameters:  $g$  and  $R$  in the Training step. The optimal values of the

Table I. Association among treatment response and clinical factors.

| Characteristics                                       | Treatment response <sup>a</sup> |      | Univariate analysis |         | Multivariable analysis |         |
|---|---------------------------------|------|---------------------|---------|------------------------|---------|
|   | CR/PR/SD                        | PD   | OR (95% CI)         | P-value | OR (95% CI)            | P-value |
| Sex   |                                 |      |                     |         |                        |         |
| Female  | 84                              | 32   | 1.00                |         | 1.00                   |         |
| Male  | 60                              | 42   | 1.146 (1.011-1.298) | 0.0346  | 1.105 (0.967-1.263)    | 0.1440  |
| Age (median, 60;<br>interquartile range, 52-70 years) |                                 |      |                     |         |                        |         |
| <60   | 64                              | 36   | 1.00                |         | 1.00                   |         |
| ≥60   | 80                              | 38   | 0.963 (0.848-1.093) | 0.5570  | 0.998 (0.993-1.003)    | 0.4235  |
| Race <sup>b</sup>                                     |                                 |      |                     |         |                        |         |
| White   | 124                             | 67   | 1.00                |         | 1.00                   |         |
| Non-white   | 16                              | 6    | 0.925 (0.750-1.141) | 0.4670  | 0.980 (0.791-1.214)    | 0.8555  |
| History of malignancy                                 |                                 |      |                     |         |                        |         |
| No  | 119                             | 70   |                     |         |                        |         |
| Yes   | 25                              | 4    | 0.793 (0.660-0.952) | 0.0137  | 0.854 (0.707-1.032)    | 0.1048  |
| Histologic diagnosis                                  |                                 |      |                     |         |                        |         |
| LMS   | 48                              | 36   | 1.00                |         | 1.00                   |         |
| DLS   | 27                              | 21   | 1.009 (0.855-1.190) | 0.9157  | 0.985 (0.817-1.189)    | 0.8784  |
| DT  | 2                               | 0    | 0.651 (0.339-1.251) | 0.1996  | 0.594 (0.306-1.152)    | 0.1247  |
| MPNST   | 6                               | 3    | 0.909 (0.660-1.252) | 0.5603  | 0.751 (0.520-1.084)    | 0.1279  |
| MFS   | 19                              | 5    | 0.802 (0.650-0.991) | 0.0422  | 0.818 (0.646-1.037)    | 0.0984  |
| SS  | 7                               | 1    | 0.738 (0.527-1.035) | 0.0795  | 0.668 (0.465-0.958)    | 0.0295  |
| UPS   | 35                              | 8    | 0.785 (0.661-0.931) | 0.0060  | 0.795 (0.655-0.964)    | 0.0205  |
| Residual tumor  |                                 |      |                     |         |                        |         |
| R0  | 98                              | 39   | 1.00                |         | 1.00                   |         |
| R1+R2   | 31+2                            | 23+2 | 1.158 (1.002-1.338) | 0.0487  | 1.137 (0.973-1.328)    | 0.1071  |
| RX  | 13                              | 10   | 1.162 (0.944-1.431) | 0.1590  | 1.196 (0.949-1.507)    | 0.1314  |
| Chemotherapy  |                                 |      |                     |         |                        |         |
| No  | 113                             | 48   | 1.00                |         | 1.00                   |         |
| Yes   | 31                              | 26   | 1.171 (1.016-1.350) | 0.0304  | 1.142 (0.980-1.332)    | 0.0911  |
| Radiotherapy  |                                 |      |                     |         |                        |         |
| No  | 89                              | 49   | 1.00                |         | 1.00                   |         |
| Yes   | 55                              | 25   | 0.958 (0.841-1.092) | 0.5250  | 1.060 (0.910-1.235)    | 0.4547  |

<sup>a</sup>CR/PR/SD denote complete response (129 patients), partial response (5 patients) and stable disease (10 patients); PD denotes progressive disease. <sup>b</sup>Non-white group includes African American (16 patients) and Asian (6 patients). There are 5 missing data. LMS, leiomyosarcoma; DLS, dedifferentiated liposarcoma; UPS, undifferentiated pleomorphic sarcoma; MFS, myxofibrosarcoma; DT, desmoid tumor; SS, synovial sarcoma; MPNT, malignant peripheral nerve sheath tumors.

tuning parameters  $g$  and  $R$  are usually not known in advance. One can use a nested inner loop of  $K$ -fold cross-validation approach on the training data to select the optimal tuning parameter value without affecting statistical validity of the procedure. A similar procedure and a work flow plot can be found in (21), which adopt a cox model. A more general description about the procedure is also available (19,20). In the present study, the 10-fold cross-validation is recommended and used, because it permits the maximization of the portion of study patients contributing to the development of the diagnostic signature and the minimization of prediction error (22).

## Results

*Clinical information summaries.* Table I summarizes the results of clinical information. There were 74 patients with progressive disease. There are 129, 5 and 15 patients with complete response, partial response and stable disease, respectively, thus, considered as positive responders. Univariate and multivariate analyses indicate that radiotherapy is not a significant benefit therapeutic treatment on the current dataset. We further performed a 10-fold cross-validation, showing that the clinical factors had a poor prediction on treatment response, with AUC=0.6.

Table II. The 65 genes included in the radiosensitive gene signature and their interaction effects with radiotherapy.

|    | Gene names   | Main effects<br>of genes (SE) | P-value | Main effects of<br>radiotherapy (SE) | P-values | Interaction<br>effects (SE) | P-value  |
|----|--------------|-------------------------------|---------|--------------------------------------|----------|-----------------------------|----------|
| 1  | CDC37L1      | -0.027 (0.034)                | 0.419   | -0.002 (0.065)                       | 0.976    | 0.375 (0.089)               | 3.49E-05 |
| 2  | OSBPL2       | -0.091 (0.038)                | 0.017   | -0.037 (0.065)                       | 0.570    | 0.259 (0.067)               | 1.65E-04 |
| 3  | ZNF880       | -0.149 (0.047)                | 0.002   | -0.043 (0.065)                       | 0.514    | 0.235 (0.063)               | 2.46E-04 |
| 4  | AGFG1        | 0.083 (0.034)                 | 0.017   | -0.062 (0.065)                       | 0.343    | -0.318 (0.086)              | 2.68E-04 |
| 5  | ZYG11B       | -0.009 (0.032)                | 0.778   | -0.023 (0.065)                       | 0.720    | 0.495 (0.139)               | 4.64E-04 |
| 6  | NAT1         | 0.089 (0.035)                 | 0.012   | -0.052 (0.065)                       | 0.431    | -0.278 (0.079)              | 5.55E-04 |
| 7  | NEXN         | -0.028 (0.039)                | 0.463   | -0.023 (0.065)                       | 0.728    | 0.231 (0.066)               | 6.10E-04 |
| 8  | PCNT         | -0.027 (0.032)                | 0.397   | -0.005 (0.066)                       | 0.943    | 0.844 (0.243)               | 6.36E-04 |
| 9  | TBC1D13      | -0.202 (0.041)                | 0.000   | -0.019 (0.065)                       | 0.775    | 0.214 (0.063)               | 8.05E-04 |
| 10 | KPNA6        | -0.062 (0.041)                | 0.132   | -0.052 (0.065)                       | 0.429    | 0.217 (0.064)               | 8.59E-04 |
| 11 | GOLGA7       | 0.107 (0.041)                 | 0.010   | -0.041 (0.065)                       | 0.526    | -0.215 (0.064)              | 9.27E-04 |
| 12 | TRAM1        | 0.074 (0.038)                 | 0.049   | -0.040 (0.065)                       | 0.543    | -0.233 (0.069)              | 9.31E-04 |
| 13 | BICD2        | -0.045 (0.040)                | 0.256   | -0.056 (0.065)                       | 0.396    | 0.221 (0.066)               | 9.34E-04 |
| 14 | CD302        | 0.016 (0.036)                 | 0.655   | -0.082 (0.066)                       | 0.215    | -0.251 (0.075)              | 9.47E-04 |
| 15 | ZNF285       | -0.066 (0.040)                | 0.100   | -0.038 (0.065)                       | 0.560    | 0.217 (0.065)               | 1.09E-03 |
| 16 | C14orf180    | -0.038 (0.034)                | 0.261   | -0.004 (0.067)                       | 0.950    | 0.345 (0.105)               | 1.13E-03 |
| 17 | TPM1         | -0.051 (0.035)                | 0.150   | -0.005 (0.067)                       | 0.941    | 0.274 (0.084)               | 1.23E-03 |
| 18 | ARTN         | -0.015 (0.034)                | 0.651   | -0.006 (0.066)                       | 0.931    | 0.324 (0.099)               | 1.24E-03 |
| 19 | NBPF14       | -0.026 (0.038)                | 0.489   | -0.058 (0.065)                       | 0.373    | 0.223 (0.068)               | 1.27E-03 |
| 20 | SGPP1        | 0.089 (0.041)                 | 0.029   | -0.036 (0.066)                       | 0.581    | -0.211 (0.065)              | 1.39E-03 |
| 21 | ZNF2         | -0.123 (0.048)                | 0.011   | -0.044 (0.066)                       | 0.503    | 0.207 (0.064)               | 1.44E-03 |
| 22 | WDFY2        | 0.056 (0.038)                 | 0.142   | -0.034 (0.066)                       | 0.605    | -0.220 (0.069)              | 1.50E-03 |
| 23 | ACVR1C       | -0.026 (0.036)                | 0.472   | -0.037 (0.065)                       | 0.572    | 0.234 (0.073)               | 1.56E-03 |
| 24 | PHKG1        | -0.017 (0.036)                | 0.628   | -0.025 (0.066)                       | 0.704    | 0.241 (0.076)               | 1.80E-03 |
| 25 | TSSK3        | 0.015 (0.033)                 | 0.652   | -0.016 (0.066)                       | 0.808    | 0.324 (0.103)               | 1.84E-03 |
| 26 | C5orf15      | 0.060 (0.036)                 | 0.093   | -0.045 (0.066)                       | 0.492    | -0.246 (0.078)              | 1.86E-03 |
| 27 | TGFBR2       | 0.094 (0.039)                 | 0.016   | -0.044 (0.066)                       | 0.498    | -0.213 (0.068)              | 1.89E-03 |
| 28 | LSM14B       | -0.042 (0.038)                | 0.268   | -0.061 (0.066)                       | 0.358    | 0.217 (0.069)               | 2.00E-03 |
| 29 | DMPK         | -0.060 (0.038)                | 0.113   | -0.025 (0.066)                       | 0.704    | 0.220 (0.070)               | 2.01E-03 |
| 30 | IPO13        | -0.084 (0.045)                | 0.063   | -0.047 (0.066)                       | 0.475    | 0.198 (0.063)               | 2.04E-03 |
| 31 | ADPGK        | 0.019 (0.038)                 | 0.617   | -0.002 (0.067)                       | 0.974    | -0.221 (0.071)              | 2.12E-03 |
| 32 | FLJ35390     | -0.075 (0.037)                | 0.044   | -0.038 (0.066)                       | 0.561    | 0.222 (0.072)               | 2.16E-03 |
| 33 | LPHN3        | -0.028 (0.035)                | 0.418   | -0.011 (0.067)                       | 0.874    | 0.261 (0.084)               | 2.27E-03 |
| 34 | SLC16A8      | -0.019 (0.037)                | 0.607   | -0.020 (0.066)                       | 0.765    | 0.221 (0.072)               | 2.33E-03 |
| 35 | RNF141       | 0.059 (0.038)                 | 0.124   | -0.049 (0.066)                       | 0.457    | -0.210 (0.068)              | 2.40E-03 |
| 36 | CRYZ         | -0.082 (0.044)                | 0.062   | -0.050 (0.066)                       | 0.452    | 0.195 (0.064)               | 2.53E-03 |
| 37 | BTF3L4       | -0.028 (0.037)                | 0.447   | -0.061 (0.066)                       | 0.357    | 0.220 (0.072)               | 2.56E-03 |
| 38 | AHDC1        | -0.016 (0.043)                | 0.719   | -0.046 (0.065)                       | 0.479    | 0.190 (0.063)               | 2.75E-03 |
| 39 | DTNA         | -0.025 (0.037)                | 0.493   | -0.016 (0.066)                       | 0.807    | 0.220 (0.073)               | 2.75E-03 |
| 40 | CAMK2G       | -0.055 (0.035)                | 0.122   | 0.019 (0.071)                        | 0.794    | 0.278 (0.092)               | 2.77E-03 |
| 41 | RILPL2       | 0.017 (0.034)                 | 0.620   | -0.051 (0.066)                       | 0.437    | -0.295 (0.097)              | 2.77E-03 |
| 42 | PDLIM3       | -0.056 (0.041)                | 0.171   | -0.024 (0.066)                       | 0.714    | 0.199 (0.066)               | 2.79E-03 |
| 43 | DKFZp761E198 | 0.028 (0.036)                 | 0.441   | -0.034 (0.066)                       | 0.601    | -0.236 (0.078)              | 2.83E-03 |
| 44 | IL1RAP       | 0.256 (0.091)                 | 0.006   | -0.054 (0.066)                       | 0.411    | -0.294 (0.097)              | 2.85E-03 |
| 45 | CLASP1       | -0.019 (0.038)                | 0.620   | -0.022 (0.066)                       | 0.740    | 0.211 (0.070)               | 2.86E-03 |
| 46 | ZNF830       | -0.031 (0.043)                | 0.470   | -0.043 (0.065)                       | 0.509    | 0.191 (0.063)               | 2.90E-03 |
| 47 | DRAM1        | 0.061 (0.032)                 | 0.060   | -0.071 (0.067)                       | 0.289    | -0.453 (0.150)              | 2.91E-03 |
| 48 | CDC5L        | 0.000 (0.035)                 | 0.993   | -0.052 (0.066)                       | 0.429    | 0.253 (0.084)               | 2.96E-03 |
| 49 | ZNF330       | -0.066 (0.039)                | 0.091   | -0.047 (0.066)                       | 0.474    | 0.202 (0.067)               | 2.97E-03 |
| 50 | ABCA8        | 0.129 (0.051)                 | 0.012   | -0.043 (0.066)                       | 0.512    | -0.196 (0.065)              | 2.99E-03 |
| 51 | SNAPC3       | 0.042 (0.034)                 | 0.220   | 0.023 (0.067)                        | 0.726    | 0.263 (0.088)               | 3.11E-03 |

Table II. The 65 genes included in the radiosensitive gene signature and their interaction effects with radiotherapy.

| Gene names   | Main effects of genes (SE) | P-value | Main effects of radiotherapy (SE) | P-values | Interaction effects (SE) | P-value  |
|--------------|----------------------------|---------|-----------------------------------|----------|--------------------------|----------|
| 52 ARF3      | -0.045 (0.033)             | 0.174   | -0.065 (0.065)                    | 0.319    | -0.300 (0.101)           | 3.19E-03 |
| 53 RPL23AP7  | -0.095 (0.049)             | 0.054   | -0.048 (0.066)                    | 0.468    | 0.192 (0.065)            | 3.26E-03 |
| 54 LOC642852 | -0.066 (0.042)             | 0.117   | -0.050 (0.066)                    | 0.447    | 0.191 (0.064)            | 3.33E-03 |
| 55 S100PBP   | 0.009 (0.037)              | 0.819   | -0.037 (0.065)                    | 0.570    | 0.206 (0.069)            | 3.36E-03 |
| 56 LETM2     | 0.164 (0.065)              | 0.012   | -0.050 (0.067)                    | 0.450    | -0.222 (0.075)           | 3.40E-03 |
| 57 LYN       | 0.065 (0.040)              | 0.101   | -0.044 (0.066)                    | 0.500    | -0.197 (0.066)           | 3.42E-03 |
| 58 SGCA      | -0.045 (0.037)             | 0.226   | -0.018 (0.067)                    | 0.787    | 0.221 (0.075)            | 3.51E-03 |
| 59 CTNND2    | 0.015 (0.032)              | 0.639   | 0.055 (0.073)                     | 0.451    | 0.926 (0.314)            | 3.56E-03 |
| 60 SNTA1     | 0.016 (0.035)              | 0.643   | 0.002 (0.067)                     | 0.977    | 0.256 (0.087)            | 3.70E-03 |
| 61 PAPSS2    | 0.076 (0.035)              | 0.031   | -0.053 (0.066)                    | 0.419    | -0.243 (0.083)           | 3.73E-03 |
| 62 NUMBL     | -0.094 (0.042)             | 0.026   | -0.046 (0.066)                    | 0.484    | 0.188 (0.064)            | 3.81E-03 |
| 63 SLC25A4   | -0.061 (0.041)             | 0.137   | -0.024 (0.067)                    | 0.714    | 0.193 (0.066)            | 3.87E-03 |
| 64 STK25     | -0.177 (0.041)             | 0.000   | -0.034 (0.065)                    | 0.603    | 0.183 (0.063)            | 4.01E-03 |
| 65 KLHL30    | -0.018(0.036)              | 0.616   | -0.018 (0.066)                    | 0.786    | 0.225 (0.077)            | 4.03E-03 |

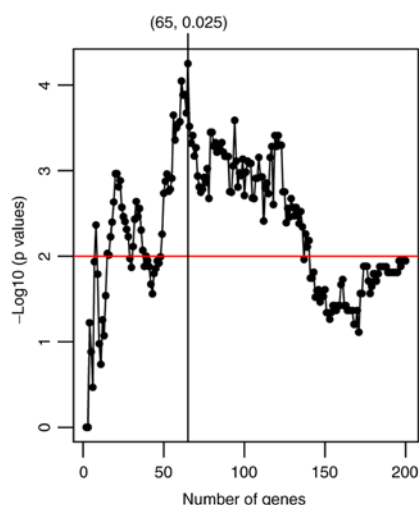


Figure 1. The  $-\log_{10}(p\text{-values})$  profile by univariate logistic tests between radiotherapy and non-radiotherapy groups for predicted RS patients. It was found that gene signatures including top 65 significant genes with a threshold  $nOR=0.025$  can provide a powerful prediction.

**Development of radiosensitive gene signature.** Following the proposed procedure, we analyzed the current dataset to obtain the tuning parameters by 10-fold cross-validation. The 10 loops produced 10 combinations of tuning parameters, of which the gene signature might be different with each loop. Theoretically, the reselection of the significant genes for different loops of the cross-validation is essential to the validity of the approach (23). However, it does not mean that the classifications and selection are unstable or that the classifier will not predict accurately for independent data. Genomic signatures are generally not unique (20,24). As suggested by Freidlin *et al* (20), the first cross-validation subset could be used to select the tuning parameter. We found that the gene signatures including the 65 significant genes with a threshold  $nOR=0.025$  could provide a powerful prediction of response with the smallest P-values. Then,  $g=65$  and  $R=0.025$  were chosen as the tuning parameters.

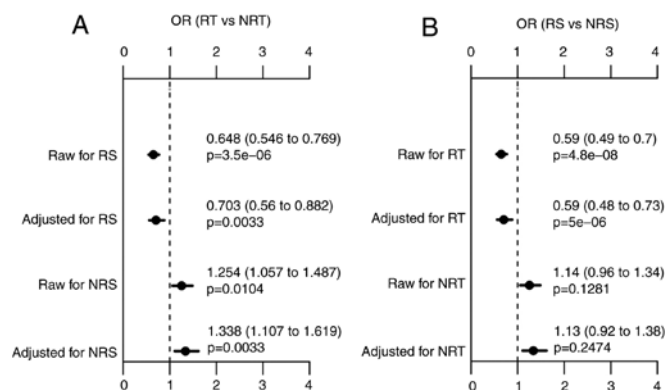


Figure 2 (A, B). The OR estimation for radiotherapy (RT) vs. non-radiotherapy (NRT) and predicted radiosensitive (RS) vs. non-radiosensitive (NRS). The adjusted factors are sex, age, race, history of other malignancy, histologic diagnosis, residual tumor and chemotherapy.

Fig. 1 shows the corresponding  $p$ -value profile by comparing the risk of disease progression between radiotherapy and non-radiotherapy groups for predicted radiosensitive patients. Table II shows the 65 genes included in the radiosensitive gene signature and their interaction effects with radiotherapy.

**Validation of radiosensitive gene signature.** Following the standard validation procedure described above, 86 patients were predicted as radiosensitive. We performed univariate logistic regression analysis to assess the effect of radiotherapy. Furthermore, multivariable analysis was also performed by adjusting gender, age, race, history of other malignancy, histologic diagnosis, residual tumor and chemotherapy. Fig. 2A shows that radiotherapy strongly associates with the improved treatment response for radiosensitive patients. While for non-radiosensitive patients, radiotherapy increased the risk of disease progression. When comparing radiosensitive and non-radiosensitive patients when they all received radiotherapy, the risk of disease progression for radiosensitive patients was significantly reduced, as shown in Fig. 2B. These results are

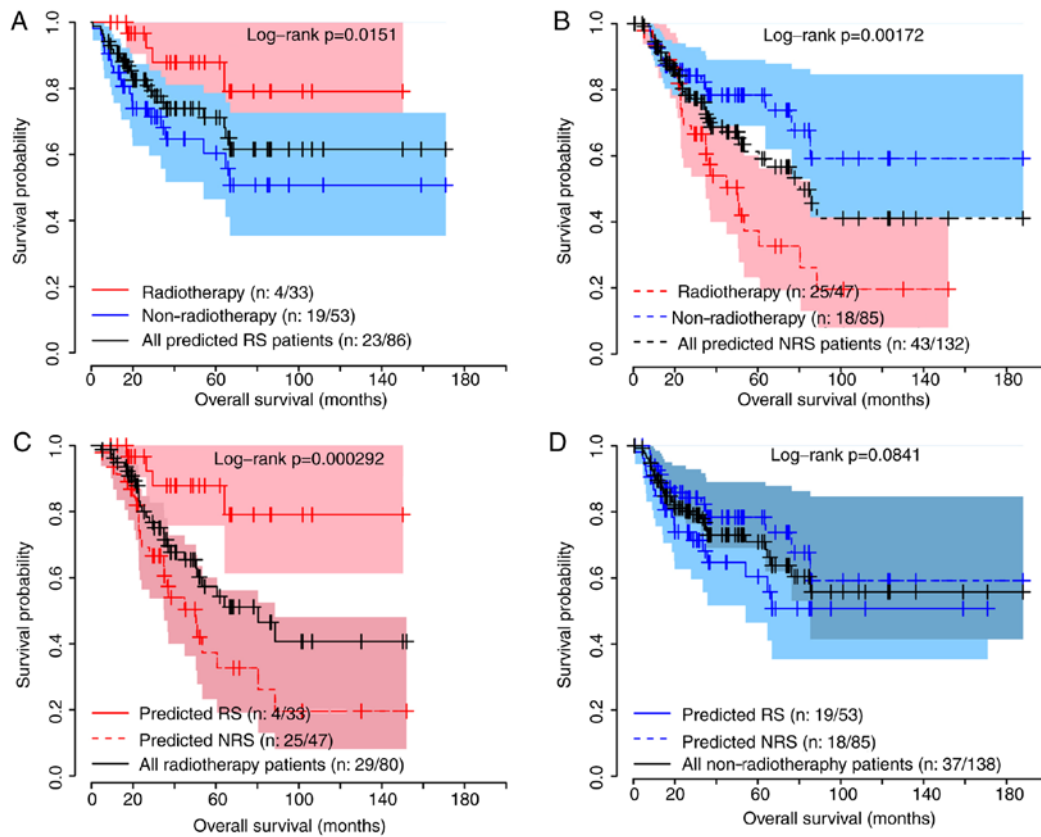


Figure 3 (A-D). The survival curves under radiotherapy and non-radiotherapy for both predicted radiosensitive (RS) and non-radiosensitive (NRS) patients. The colored areas denote the 95% confidence intervals for survival rate.

reasonable and consistent with general clinical expectation, and suggest that the predicted radiosensitive patients are accurate, and the positive radiotherapy effect on predicted radiosensitive patient is validated effectively.

**Survival comparison for predicted radiosensitive patients.** We further compared the survival of these predicted radiosensitive patients under radiotherapy and non-radiotherapy. Observed significant differences partially or indirectly suggested that our sensitivity prediction was effective. Fig. 3A shows the survival curve for these predicted radiosensitive patients under radiotherapy and non-radiotherapy. The significant difference suggested that the predicted radiosensitive patients strongly benefited from radiotherapy. Fig. 3B shows the survival comparison for predicted non-radiosensitive patients. A significant difference was detected between the two groups, suggesting that radiotherapy on non-radiosensitive patients would increase the risk of death. We further compared the survival among radiosensitive and non-radiosensitive patients all under radiotherapy treatment, shown in Fig. 3C. As expected, a strong positive effect of radiotherapy was observed on radiosensitive patients. In addition, there was no significant difference in survival between radiosensitive and non-radiosensitive patients who did not receive radiotherapy treatment (Fig. 3D). Taken together, the predicted radiosensitive and non-radiosensitive patients were appropriate predictions, suggesting that gene signature is potentially a predictive biomarker for sensitivity prediction for both radiosensitive and non-radiosensitive patients.

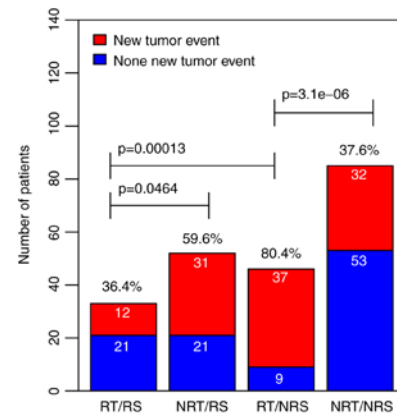


Figure 4. The comparisons of the rate of new tumor event among different groups. The rates for different groups are compared by Fisher's exact test. RT, radiotherapy; NRT, non-radiotherapy; RS, radiosensitive; NRS, non-radiosensitive.

**Comparison of new tumor event for predicted radiosensitive patients.** To further validate the predicted radiosensitive and non-radiosensitive patients, we compared the rate of new tumor event for predicted radiosensitive patients under radiotherapy and non-radiotherapy. New tumor event, as an important clinical index for prognostic outcome, was defined as metastatic, recurrent, and new primary tumor after initial treatment according to TCGA. The results by Fisher's exact test are summarized in Fig. 4. The predicted radiosensitive patients who received radiotherapy had a significantly

Table III. Association among predicted radiosensitivity and clinical factors.

| Characteristics                              | NRS | RS | Univariate analysis |         | Multivariable analysis |          |
|--|-----|----|---------------------|---------|------------------------|----------|
|  |     |    | HR (95% CI)         | P-value | HR (95% CI)            | P-value  |
| Sex  |     |    |                     |         |                        |          |
| Female                                       | 80  | 36 | 1.00                |         |                        |          |
| Male   | 52  | 50 | 1.197 (1.053-1.361) | 0.0066  | 1.064 (0.940-1.204)    | 0.329    |
| Age (median, 60; interquartile range, 52-70) |     |    |                     |         |                        |          |
| <60  | 70  | 30 |                     |         |                        |          |
| ≥60  | 62  | 56 | 1.191 (1.047-1.354) | 0.0084  | 1.047 (0.923-1.188)    | 0.478    |
| Race <sup>a</sup>                            |     |    |                     |         |                        |          |
| White  | 114 | 77 |                     |         |                        |          |
| Non-white                                    | 16  | 6  | 0.878 (0.708-1.089) | 0.2370  | 0.965 (0.789-1.181)    | 0.730    |
| History of malignancy                        |     |    |                     |         |                        |          |
| No   | 118 | 71 |                     |         |                        |          |
| Yes  | 14  | 15 | 1.152 (0.952-1.395) | 0.1480  | 1.068 (0.895-1.274)    | 0.467    |
| Histologic diagnosis <sup>a</sup>            |     |    |                     |         |                        |          |
| LMS  | 73  | 11 |                     |         |                        |          |
| DLS  | 14  | 34 | 1.781 (1.530-2.074) | 2.6e-12 | 1.763 (1.479-2.101)    | 1.64E-09 |
| DT   | 2   | 0  | 0.877 (0.481-1.602) | 0.6702  | 0.844 (0.457-1.557)    | 0.587    |
| MPNST  | 6   | 3  | 1.224 (0.911-1.645) | 0.1803  | 1.167 (0.841-1.621)    | 0.356    |
| MFS  | 10  | 14 | 1.572 (1.294-1.910) | 8.9e-06 | 1.537 (1.249-1.891)    | 6.88E-05 |
| SS   | 7   | 1  | 0.994 (0.728-1.357) | 0.9701  | 1.025 (0.748-1.406)    | 0.877    |
| UPS  | 20  | 23 | 1.498 (1.279-1.754) | 1.1e-06 | 1.477 (1.247-1.751)    | 1.14E-05 |
| Residual tumor                               |     |    |                     |         |                        |          |
| R0   | 89  | 48 |                     |         |                        |          |
| R1+R2  | 28  | 30 | 1.182 (1.018-1.372) | 0.0296  | 0.958 (0.830-1.107)    | 0.563    |
| RX   | 15  | 8  | 0.997 (0.805-1.237) | 0.9815  | 1.110 (0.895-1.377)    | 0.344    |
| Chemotherapy                                 |     |    |                     |         |                        |          |
| No   | 88  | 73 |                     |         |                        |          |
| Yes  | 44  | 13 | 0.798 (0.690-0.923) | 0.0027  | 0.944 (0.819-1.088)    | 0.426    |

<sup>a</sup>Non-white group including African American and Asian. LMS, leiomyosarcoma; DLS, dedifferentiated liposarcoma; UPS, undifferentiated pleomorphic sarcoma; MFS, myxofibrosarcoma; DT, desmoid tumor; SS, synovial sarcoma; MPNT, malignant peripheral nerve sheath tumors. NRS, non-radiosensitive; RS, radiosensitive; HR, hazard ratio; CI, confidence interval.

lower rate of new tumor events. While for non-radiosensitive patients, radiotherapy might be a risk factor to increase the rate of new tumor event. These results were consistent with the results from survival analysis for the radiosensitive and non-radiosensitive patients (Fig. 3), and further validated our prediction on radiosensitivity.

#### Associations between radiosensitivity and clinical factors.

To detect whether predicted sensitivity and clinical factors were related, we performed univariate and multivariate logistic regression analysis to find the clinical factors associated with our radiosensitivity prediction. Table III summarizes the results. Sex, age, histologic type, residual tumor and chemotherapy were significantly associated with predicted sensitivity in univariate analysis. In multivariable analysis, only histologic type was significantly associated

with prediction. We performed univariate logistic analysis to evaluate the treatment response of radiotherapy under different strata of these clinical factors. Fig. 5A shows the OR (95% CI) and P-values under different strata of clinical factors for predicted radiosensitive patients. The result suggested that the risk of disease progression was significantly reduced for the predicted radiosensitive under radiotherapy in most subgroups. For the predicted non-radiosensitive patients, there was potential risk of disease progression under radiotherapy for each subgroup, as shown in Fig. 5B. These strata analysis suggested that our prediction on radiosensitive patients was stable and reasonable.

**ROC analysis.** To show the prediction power of the prediction pattern on disease progression, we selected the 80 patients who received radiotherapy. We performed ROC analysis by using

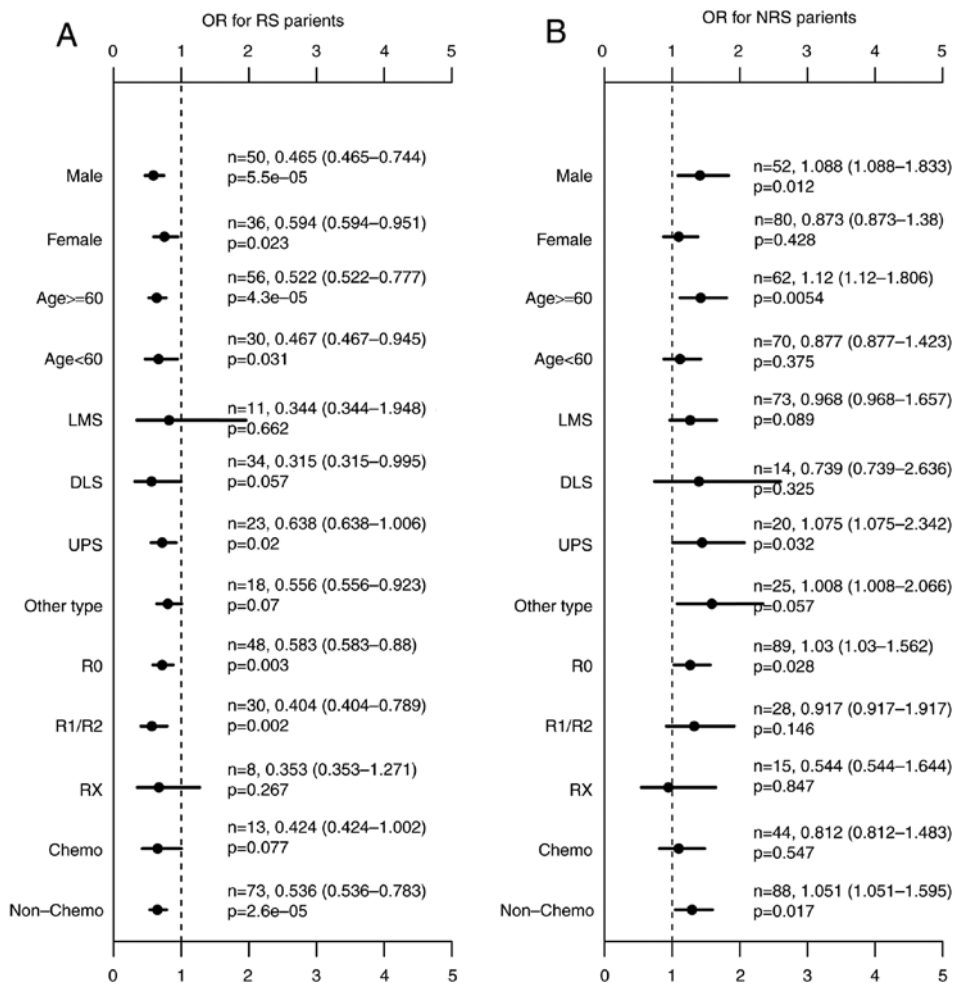


Figure 5. The strata analysis to estimate OR (radiotherapy vs. non-radiotherapy) for predicted radiosensitive (RS) (A) and non-radiosensitive (NRS) patients (B). Histologic type DT, MPNST, MFS and SS are combined together as 'Other type', because of small sample sizes. Chemo denotes chemotherapy, and non-chemo denotes no chemotherapy.

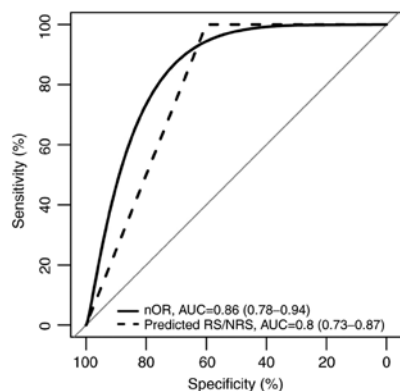


Figure 6. ROC analysis of nOR index and predicted radiosensitive/non-radiosensitive (RS/NRS) to show the power of our prediction on treatment outcome.

the nOR values calculated for each patients and their recorded clinical treatment response. Fig. 6 shows the ROC curve. The prediction power was significantly high for nOR index. After cutting off nOR values according to the threshold  $R=0.025$ , the prediction power was still significant for predicting the treatment outcome with AUC 0.80 (0.73–0.87).

**Gene signature and cluster analysis.** To provide evidence from the selected gene signature, we further extracted the expression pattern of these 65 genes to perform hierarchical clustering analysis by using R package heatmap. The result is present in Fig. 7. All patients were classified into two groups according the hierarchical cluster analysis. The blue bar below the dendrogram denoted the predicted radiosensitive patients, while the yellow bar denoted the predicted non-radiosensitive patients. Most of these predicted radiosensitive and non-radiosensitive patients were well-matched with the result of hierarchical cluster based on the selected gene signature. In our previous prediction, there were 86 patients (blue bar) predicted as radiosensitive patients. Most (78 out of 86) of them were clustered at the left branch. The right branch clustered 114, and among them, 106 patients were predicted as non-radiosensitive patients (yellow bar). This result validated the radiosensitive patients we predicted and also suggested that the radiosensitive gene signature we develop was predictive and informative.

## Discussion

Adjuvant radiotherapy is used in the treatment of soft tissue sarcomas to provide better local control. The treatment-induced



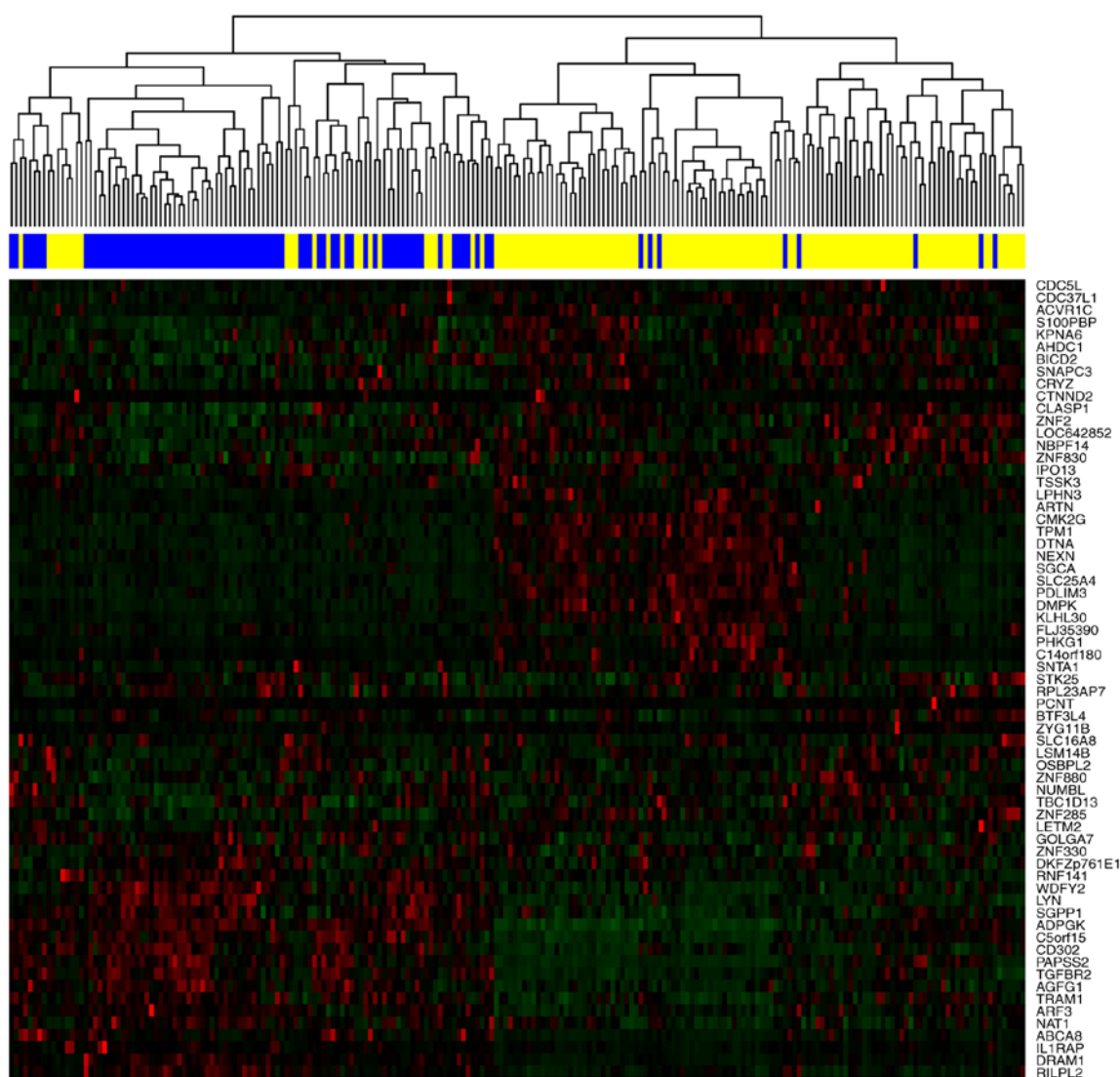


Figure 7. Hierarchical clustering analyses. Hierarchical clustering was used to determine the expression pattern of selected 65 genes. The top blue and yellow bands denote the predicted radiosensitive and non-radiosensitive patients, respectively. In total, 78 out of 86 predicted radiosensitive patients are classed at the left branch, and 106 out of 132 predicted non-radiosensitive patients are classed at the right branch.

histologic response reflects treatment effectiveness and is highly associated with the patient's prognosis. Prediction of possible outcome of a therapy treatment has important clinical significance. Clinical factors usually have a poor prediction on radiosensitivity. There is increasing evidence that patients with similar clinical characteristics have tumors that are very different with regard to genes and pathways that are driving tumor growth. In the era of precision medicine, 'omics' technologies can robustly generate large-scale molecular data, which provides extraordinary opportunities to detect biomarkers and to build more accurate gene signatures and prognostic models.

It would be ideal to perform a clinical phase II trial for developing biomarkers for patients most likely to respond to radiotherapy. However, such biomarkers are difficult to identify and are often not available by the time phase III trials are initialized. The main concerns are: i) signature development in high dimensional gene expression data requires large sample sizes, and ii) when the proportion of sensitive patients is low, a large number of patients need to be enrolled to ensure

sufficient number of sensitive patients and achieve acceptable power. This may not be feasible for sarcoma and other rare cancers. The proposed design and model described here may be useful in such situations, especially for rare cancer types, like sarcoma.

In this study, we extend and propose a statistical approach to develop a radiosensitive gene signature and predict radio-sensitive patients with sarcoma. A nested inner loop 10-fold cross-validation was used to develop and validate the gene signature. This is a standard internal validation procedure. The 10-fold cross-validation permits the maximization of the portion of study patients contributing to the development of the diagnostic signature and the minimization of prediction error (22). We verified our gene signature and sensitivity prediction by several aspects: i) we developed a novel cross-validation procedure. The result showed that the predicted radiosensitive patients under radiotherapy had significantly better treatment response than both predicted radiosensitive patients under non-radiotherapy, and predicted non-radiosensitive patients who receive radiotherapy; ii) after adjusting

chemotherapy and other clinical factors, the risk of disease progression was still significantly reduced under radiotherapy for these predicted radiosensitive patients; iii) we also validated our prediction by comparing the overall survival. The results showed that the predicted radiosensitive patients who received radiotherapy had significantly better survival than both predicted radiosensitive patients under non-radiotherapy, and predicted non-radiosensitive patients who receive radiotherapy; iv) the predicted radiosensitive patients who received radiotherapy had a significantly reduced rate of new tumor events. v) ROC analysis suggested a powerful prediction of our model with AUC=0.8; vi) Strata analysis showed that the predicted radiosensitive patients still had significantly low risk of progression under radiotherapy, irrespective of subgroup; vii) we finally validated our prediction using a hierarchical cluster analysis on the selected gene signature. We found that more than 80% predicted radiosensitive patients and non-radiosensitive patients were well-clustered into two branches of the dendrogram. All these direct and indirect results confirmed our prediction on radiosensitivity and the gene signature.

The developed gene signature is easy to apply for predicting new patients. According to the estimation of each gene in gene signature (Table II), calculating the OR for each gene using the standardized expression value of RNAseq, and comparing product of these OR (nOR) with the threshold 0.025 can predict patients who are radiosensitive (based on whether their nOR value is less than the threshold). This simple model can easily be applied in clinical practice, because gene expression data are readily available for patients.

The genes included in the signature not only associate with the radiosensitivity, but also may contribute to the molecular basis of sarcoma. Results from pathway analysis show that these involved genes associated with a wide range of pathways and functional annotation, like calcium signaling pathway, inflammatory mediator regulation of TRP channels, wnt signaling pathway, microRNAs and proteoglycans in cancer, transcriptional and misregulation in cancer, colorectal and pancreatic cancer. In the current gene signature list, CDC5L is correlated with advanced TNM stage of osteosarcoma (25) and responds poorly to neo-adjuvant chemotherapy (26). Overexpression of CDC5L favors cell cycle progression of hepatocellular carcinoma cells (27) and also plays an important role in glioma (28). TGFBR2 might be a key gene associated with sarcoma (29). It also acts as a driver of colorectal cancer (30,31), and is strongly associated with gastric (32), oral carcinoma (33), ovarian (34) and breast cancer (35,36). ABCA8 and WDFY2 contribute to the progression and development of ovarian cancer (37,38). A meta-analysis suggests that NAT1 polymorphisms significantly associate with the risk of pancreatic cancer (39). A previous study also suggests that NAT1 might be a possible prognostic biomarker for lymph node-positive breast cancer. Other genes, like CDC37L1, ARTN, SGPP1, S100BP, LETM2, CTNND2 and KLHL30, were found to be related with several different types of cancer (40-46). These genes involved in current radiosensitive gene signature strongly suggest that they might be responsible for molecular basis of sarcoma and radiosensitivity, and also supported our development of this signature.

Although this study has much strength, there are some drawbacks, such as the small sample size and the lack of suitable external evidence from molecular, cellular, and importantly clinical level, although we provided several related results to support our prediction. In addition, the signature identified in this study has no overlap with that of our previous one on a survival outcome (21). This may be due to the outcomes being different, and the small sample size. Although, these are some drawbacks, we still provided some useful model framework for rare cancer to develop radiosensitive gene signatures. We also suggested a potential useful gene signature for predicting radiosensitivity for patients with soft tissue sarcoma.

### Acknowledgements

We acknowledge the contributions of the TCGA Research Network. We also thank the Lynae Hanks for editing the manuscript. The present study was supported by grants from the China Scholarship Council, the National Natural Science Foundation of China (nos. 81573253 and 81773541), project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions at Soochow University to ZXT, the National Natural Science Foundation of China (no. 81672743), the fund (no. 2016YFC0904600) from the Chinese Ministry of Science and Technology, NIH R01CA133093, the Alabama Innovation Fund to BX, and was also supported by the research grant: NIH 2 R01GM069430 to NJY.

### References

1. Wibmer C, Leithner A, Zielonke N, Sperl M and Windhager R: Increasing incidence rates of soft tissue sarcomas? A population-based epidemiologic study and literature review. *Ann Oncol* 21: 1106-1111, 2010.
2. American Cancer Society: Cancer Facts and Figures 2016. In: Atlanta, Ga: American Cancer Society (Available online, <http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf>), 2016.
3. von Mehren M, Randall RL, Benjamin RS, Boles S, Bui MM, Conrad EU III, Ganjoo KN, George S, Gonzalez RJ, Heslin MJ, *et al*: Soft Tissue Sarcoma, Version 2.2016, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 14: 758-786, 2016.
4. Wang D and Abrams RA: Radiotherapy for soft tissue sarcoma: 50 years of change and improvement. *Am Soc Clin Oncol Educ Book* 34: 244-251, 2014.
5. Levy A, Bonvalot S, Bellefqih S, Vilcot L, Rimareix F, Terrier P, Belemsagha D, Cascales A, Domont J, Mir O, *et al*: Is preoperative radiotherapy suitable for all patients with primary soft tissue sarcoma of the limbs? *Eur J Surg Oncol* 40: 1648-1654, 2014.
6. El-Bared N, Wong P and Wang D: Soft tissue sarcoma and radiation therapy advances, impact on toxicity. *Curr Treat Options Oncol* 16: 19, 2015.
7. Davis AM, O'Sullivan B, Turcotte R, Bell R, Catton C, Chabot P, Wunder J, Hammond A, Benk V, Kandel R, *et al*: Late radiation morbidity following randomization to preoperative versus postoperative radiotherapy in extremity soft tissue sarcoma. *Radiother Oncol* 75: 48-53, 2005.
8. Dickie CI, Parent AL, Griffin AM, Fung S, Chung PW, Catton CN, Ferguson PC, Wunder JS, Bell RS, Sharpe MB, *et al*: Bone fractures following external beam radiotherapy and limb-preservation surgery for lower extremity soft tissue sarcoma: Relationship to irradiated bone length, volume, tumor location and dose. *Int J Radiat Oncol Biol Phys* 75: 1119-1124, 2009.
9. Pisters PW, O'Sullivan B and Maki RG: Evidence-based recommendations for local therapy for soft tissue sarcomas. *J Clin Oncol* 25: 1003-1008, 2007.

10. Strander H, Turesson I and Cavallin-Ståhl E: A systematic overview of radiation therapy effects in soft tissue sarcomas. *Acta Oncol* 42: 516-531, 2003.
11. Yang JC, Chang AE, Baker AR, Sindelar WF, Danforth DN, Topalian SL, DeLaney T, Glatstein E, Steinberg SM, Merino MJ, *et al*: Randomized prospective study of the benefit of adjuvant radiation therapy in the treatment of soft tissue sarcomas of the extremity. *J Clin Oncol* 16: 197-203, 1998.
12. Chibon F: Cancer gene expression signatures - the rise and fall? *Eur J Cancer* 49: 2000-2009, 2013.
13. Begg AC: Predicting response to radiotherapy: Evolutions and revolutions. *Int J Radiat Biol* 85: 825-836, 2009.
14. Salendo J, Spitzner M, Kramer F, Zhang X, Jo P, Wolff HA, Kitz J, Kauffuß S, Beißbarth T, Döbelstein M, *et al*: Identification of a microRNA expression signature for chemoradiosensitivity of colorectal cancer cells, involving miRNAs-320a, -224, -132 and let7g. *Radiother Oncol* 108: 451-457, 2013.
15. Spitzner M, Emons G, Kramer F, Gaedcke J, Rave-Fränk M, Scharf JG, Burfeind P, Becker H, Beissbarth T, Ghadimi BM, *et al*: A gene expression signature for chemoradiosensitivity of colorectal cancer cells. *Int J Radiat Oncol Biol Phys* 78: 1184-1192, 2010.
16. Hall JS, Iype R, Senra J, Taylor J, Armenoult L, Oguejiofor K, Li Y, Stratford I, Stern PL, O'Connor MJ, *et al*: Investigation of radiosensitivity gene signatures in cancer cell lines. *PLoS One* 9: e86329, 2014.
17. Pramana J, Van den Brekel MW, van Velthuysen ML, Wessels LF, Nuyten DS, Hofland I, Atsma D, Pimentel N, Hoebbers FJ, Rasch CR, *et al*: Gene expression profiling to predict outcome after chemoradiation in head and neck cancer. *Int J Radiat Oncol Biol Phys* 69: 1544-1552, 2007.
18. Imadome K, Iwakawa M, Nakawatari M, Fujita H, Kato S, Ohno T, Nakamura E, Ohkubo Y, Tamaki T, Kiyohara H, *et al*: Subtypes of cervical adenosquamous carcinomas classified by EpCAM expression related to radiosensitivity. *Cancer Biol Ther* 10: 1019-1026, 2010.
19. Freidlin B and Simon R: Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res* 11: 7872-7878, 2005.
20. Freidlin B, Jiang W and Simon R: The cross-validated adaptive signature design. *Clin Cancer Res* 16: 691-698, 2010.
21. Tang Z, Zeng Q, Li Y, Zhang X, Ma J, Suto MJ, Xu B and Yi N: Development of a radiosensitivity gene signature for patients with soft tissue sarcoma. *Oncotarget* 8: 27428-27439, 2017.
22. Molinaro AM, Simon R and Pfeiffer RM: Prediction error estimation: A comparison of resampling methods. *Bioinformatics* 21: 3301-3307, 2005.
23. Simon R, Radmacher MD, Dobbin K and McShane LM: Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 95: 14-18, 2003.
24. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ and Perou CM: Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355: 560-569, 2006.
25. Mohammadi M, Goudarzi PK, Rahmani O, Kaghazian P, Yahaghi E, Taheriazam A and Ahmadi K: Evaluation of gene expression level of CDC5L and MACC1 in poor prognosis and progression of osteosarcoma. *Tumour Biol* 37: 8153-8157, 2016.
26. Martin JW, Chilton-MacNeill S, Koti M, van Wijnen AJ, Squire JA and Zielenska M: Digital expression profiling identifies RUNX2, CDC5L, MDM2, RECQL4, and CDK4 as potential predictive biomarkers for neo-adjuvant chemotherapy response in paediatric osteosarcoma. *PLoS One* 9: e95843, 2014.
27. Qiu H, Zhang X, Ni W, Shi W, Fan H, Xu J, Chen Y, Ni R and Tao T: Expression and clinical role of Cdc5L as a novel cell cycle protein in hepatocellular carcinoma. *Dig Dis Sci* 61: 795-805, 2016.
28. Chen W, Zhang L, Wang Y, Sun J, Wang D, Fan S, Ban N, Zhu J, Ji B and Wang Y: Expression of CDC5L is associated with tumor progression in gliomas. *Tumour Biol* 37: 4093-4103, 2016.
29. Camões MJ, Paulo P, Ribeiro FR, Barros-Silva JD, Almeida M, Costa VL, Cerveira N, Skotheim RI, Lothe RA, Henrique R, *et al*: Potential downstream target genes of aberrant ETS transcription factors are differentially affected in Ewing's sarcoma and prostate carcinoma. *PLoS One* 7: e49819, 2012.
30. Lee J, Katzenmaier EM, Kopitz J and Gebert J: Reconstitution of TGFBR2 in HCT116 colorectal cancer cells causes increased LFNG expression and enhanced N-acetyl-d-glucosamine incorporation into Notch1. *Cell Signal* 28: 1105-1113, 2016.
31. de Miranda NF, van Dinther M, van den Akker BE, van Wezel T, ten Dijke P and Morreau H: Transforming growth factor beta signaling in colorectal cancer cells with microsatellite instability despite biallelic mutations in TGFBR2. *Gastroenterology* 148: 1427-1437.e1428, 2015.
32. Nadauld LD, Garcia S, Natsoulis G, Bell JM, Miotke L, Hopmans ES, Xu H, Pai RK, Palm C, Regan JF, *et al*: Metastatic tumor evolution and organoid modeling implicate TGFBR2 as a cancer driver in diffuse gastric cancer. *Genome Biol* 15: 428, 2014.
33. Sivasdas VP, Gulati S, Varghese BT, Balan A and Kannan S: The early manifestation, tumor-specific occurrence and prognostic significance of TGFBR2 aberrant splicing in oral carcinoma. *Exp Cell Res* 327: 156-162, 2014.
34. Sivasdas VP, Saakshi G, Iype EM, Balan A and Kannan S: Prognostic implication of the loss of TGFBR2 expression in oral carcinoma. *Neoplasia* 62: 398-404, 2015.
35. Lei J, Rudolph A, Moysich KB, Rafiq S, Behrens S, Goode EL, Pharoah PP, Seibold P, Fasching PA, Andrulis JL, *et al*; kConFab Investigators: Assessment of variation in immunosuppressive pathway genes reveals TGFBR2 to be associated with prognosis of estrogen receptor-negative breast cancer after chemotherapy. *Breast Cancer Res* 17: 18, 2015.
36. Wei CY, Tan QX, Zhu X, Qin QH, Zhu FB, Mo QG and Yang WP: Expression of CDKN1A/p21 and TGFBR2 in breast cancer and their prognostic significance. *Int J Clin Exp Pathol* 8: 14619-14629, 2015.
37. Hedditch EL, Gao B, Russell AJ, Lu Y, Emmanuel C, Beesley J, Johnatty SE, Chen X, Harnett P, George J, *et al*; Australian Ovarian Cancer Study Group: ABCA transporter gene expression and poor outcome in epithelial ovarian cancer. *J Natl Cancer Inst* 106: 106, 2014.
38. Kannan K, Coarfa C, Rajapakshe K, Hawkins SM, Matzuk MM, Milosavljevic A and Yen L: CDKN2D-WDFY2 is a cancer-specific fusion gene recurrent in high-grade serous ovarian carcinoma. *PLoS Genet* 10: e1004216, 2014.
39. Zhang K, Gao L, Wu Y, Chen J, Lin C, Liang S, Su J, Ye J and He X: NAT1 polymorphisms and cancer risk: A systematic review and meta-analysis. *Int J Clin Exp Med* 8: 9177-9191, 2015.
40. Zhou Y, Zeng Z, Zhang W, Xiong W, Li X, Zhang B, Yi W, Xiao L, Wu M, Shen S, *et al*: Identification of candidate molecular markers of nasopharyngeal carcinoma by microarray analysis of subtracted cDNA libraries constructed by suppression subtractive hybridization. *Eur J Cancer Prev* 17: 561-571, 2008.
41. Zhang M, Zhang W, Wu Z, Liu S, Sun L, Zhong Y, Zhang X, Kong X, Qian P, Zhang H, *et al*: Artemin is hypoxia responsive and promotes oncogenicity and increased tumor initiating capacity in hepatocellular carcinoma. *Oncotarget* 7: 3267-3282, 2015.
42. Bao Y, Chen Z, Guo Y, Feng Y, Li Z, Han W, Wang J, Zhao W, Jiao Y, Li K, *et al*: Tumor suppressor microRNA-27a in colorectal carcinogenesis and progression by targeting SGPP1 and Smad2. *PLoS One* 9: e105991, 2014.
43. Lines KE, Chelala C, Dmitrovic B, Wijesuriya N, Kocher HM, Marshall JF and Crnogorac-Jurcevic T: S100P-binding protein, S100BP, mediates adhesion through regulation of cathepsin Z in pancreatic cancer cells. *Am J Pathol* 180: 1485-1494, 2012.
44. Cheng C, Zhou Y, Li H, Xiong T, Li S, Bi Y, Kong P, Wang F, Cui H, Li Y, *et al*: Whole-genome sequencing reveals diverse Models of structural variations in esophageal squamous cell carcinoma. *Am J Hum Genet* 98: 256-274, 2016.
45. Wang T, Chen YH, Hong H, Zeng Y, Zhang J, Lu JP, Jeansonne B and Lu Q: Increased nucleotide polymorphic changes in the 5'-untranslated region of delta-catenin (CTNND2) gene in prostate cancer. *Oncogene* 28: 555-564, 2009.
46. Madden MH, Anic GM, Thompson RC, Nabors LB, Olson JJ, Browning JE, Monteiro AN and Egan KM: Circadian pathway genes in relation to glioma risk and outcome. *Cancer Causes Control* 25: 25-32, 2014.