A six-gene-based signature for breast cancer radiotherapy sensitivity prediction

Xing Chen
Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital
Junjie Zheng
Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital
Min ling Zhuo
Changle District Hospital
Ailong Zhang
Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital
Zhenhui You (youzhenhuifj@yeah.net)
Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital
https://orcid.org/0000-0002-2540-1682

Research article

Keywords: Breast cancer; radiotherapy sensitivity; autoencoder; LASSO Cox regression

Posted Date: April 3rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-20019/v1

A six-gene-based signature for breast cancer radiotherapy sensitivity prediction

Running title: Breast cancer radiotherapy sensitivity signature.

Xing Chen[1], Junjie Zheng[2], Min ling Zhuo[3], Ailong Zhang[4], Zhenhui You[5#]

[1]Department of general surgery, Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital, Fujian, 350000, China. E-mail: 763044127@qq.com

[2] Department of general surgery, Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital, Fujian, 350000, China. E-mail: 857748407@qq.com

[3] Department of Ultrasound, Changle District Hospital, Fujian, 350200, China. E-mail: 1251779918@qq.com

[4] Department of general surgery, Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital, Fujian, 350000, China. E-mail: ailongzh@aliyun.com

[5] Department of general surgery, Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital, Fujian, 350000, China. E-mail: youzhenhuifj@yeah.net

[#]**Correspondence:**

Zhenhui You

Department of general surgery, Fujian Medical University Provincial Clinical Medical College; Fujian Provincial Hospital, Fujian, 350000, China.

E-mail: youzhenhuifj@yeah.net

Abbreviations

BRCA	breast cancer
APBI	accelerated partial breast irradiation

IMRT	intensity-modulated radiotherapy
TCGA	the Cancer Genome Atlas
ICGC	International Cancer Genome Consortium
DEGs	differential expression genes
FC	fold change
DEG	differentially expressed gene
OS	overall survival
PLD	partial likelihood deviance
KM	Kaplan-Meier
HOXB13	Homeobox B13
BCR	biochemical recurrence
ERα	estrogen receptor α
IL	interleukin

Abstract

Background: Breast cancer (BRCA) represents the most common malignancy among women worldwide that with high mortality. Radiotherapy is a prevalent therapeutic for BRCA that with heterogeneous effectiveness among patients. **Methods:** we proposed to develop a gene expression-based signature for BRCA radiotherapy sensitivity prediction. Gene expression profiles of BRCA samples from the Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) were obtained and used as training and independent testing dataset, respectively. Differential expression genes (DEGs) in BRCA tumor samples compared with their paracancerous samples in the training set were identified by using edgeR Bioconductor package

followed by dimensionality reduction through autoencoder method and univariate Cox regression analysis to screen genes among DEGs that with significant prognosis significance in patients that were previously treated with radiation. LASSO Cox regression method was applied to screen optimal genes for constructing radiotherapy sensitivity prediction signature. **Results:** 603 DEGs were obtained in BRCA tumor samples, and seven out of which were retained after univariate cox regression analysis. LASSO Cox regression analysis finally remained six genes based on which the radiotherapy sensitivity prediction model was constructed. The signature was proved to be robust in both training and independent testing sets and an independent marker for BRCA radiotherapy sensitivity prediction. **Conclusions:** this study should be helpful for BRCA patients' therapeutics selection and clinical decision.

Key words: Breast cancer; radiotherapy sensitivity; autoencoder; LASSO Cox regression

1. Introduction

In terms of incidence, breast cancer (BRCA) is the main cancer affecting women (Ahmad Aamir, 2019). Over 2012-2016, the incidence rate of BRCA increased slightly by 0.3% per year (DeSantis C E, 2019). Mastectomy and breast-sparing surgery for radiotherapy are the most common treatment options (Caldon L J M, 2007). For the patients with early stage BRCA, accelerated partial breast irradiation (APBI) and intensity-modulated radiotherapy (IMRT) have been introduced as an alternative treatment method of mastectomy (Livi L, 2015). A population-based study showed that, the breast-conserving surgery plus radiotherapy is at least equivalent to mastectomy with respect to 10-year overall survival (Van Maaren M C, 2016). However, radiotherapy may bring side effects to patients at the same time. The estimated absolute risks from modern

radiotherapy for BRCA include lung cancer and cardiac mortality, and for long-term smokers, the absolute risks of modern radiotherapy may outweigh the benefits (Taylor C, 2017). These results would influence treatment decision making for patients with BRCA.

The decisions for radiotherapy should be tailored on the basis of patient factors, tumor biology, and the prognostic score (Sagara Y,2016). Given the predictive markers identify populations of patients who will receive substantial benefit from specific therapy, genomic assays measuring the expression of multiple genes have been development to predict response to the treatments in recent years (Munkácsy G, 2015). DeLorenzi M et al. reported a 70-gene signature for the treatment decisions in early-stage BRCA, and indicated that approximately 46% of women with BRCA with high clinical risk might not require chemotherapy (Cardoso F, 2016). Although the genotypes of 90 confirmed breast cancer risk variants was not associated with risk of radiotherapy toxicity up to 5 years following radiotherapy, the individual variants may increase risk (Leila, Dorling, 2016). Moreover, microRNA-related DNA repair/cell-cycle genes were also reported to be independently associated with relapse after radiotherapy for early BRCA (Gee H E, 2015). Currently, progress in the development of molecular markers to predict the radiation response and necessity of women with BRCA radiation remains slow (Speers C, 2017).

To identify the gene expression-based signature for BRCA radiotherapy sensitivity prediction, in this study, the gene expression profiles of BRCA were obtained from the Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC). The differential expression genes (DEGs) in BRCA tumor samples compared with their paracancerous samples from TCGA database were identified, and the DEGs that with

significant prognosis significance in patients have been treated with radiation were screened. Finally, A radiotherapy sensitivity prediction model was constructed with the key genes, and the accuracy of this model verified by data from both TCGA and ICGC databases.

2. Materials and methods

2.1 Study population

All the patient information was obtained from the public resource. A total of 1217 breast cancer samples were obtained from TCGA (http://tcgaportal.org/), which included 99 paired tumor and paracancerous tissue samples and 1019 unpaired tumor samples. Besides, 1019 breast cancer samples with their complete survival information were also obtained from ICGC (https://icgc.org/). There were 652 and 65 patients from TCGA and ICGC dataset were previously treated with radiation and used as training and testing set, respectively. Clinicopathological characteristics of patients from training and testing set were provided in Table 1.

2.2 Differential expression analysis

Differential expression analysis between the 99 paired samples were performed by using the edgeR package (Robinson M D, 2010) in R. The gene that with FDR ≤ 0.05 and absolute log2 fold change (FC) > 1 was defined as the differentially expressed gene (DEG).

2.3 Dimensionality reduction through autoencoder neural network algorithm

Autoencoder is an unsupervised learning technique which takes raw data without label as input and tries to reconstruct it by using fewer number of bits from the bottleneck layer. In this study, we proposed to reduce candidate genes for radiotherapy prediction

model by using autoencoder neural network algorithm through which dimensionality reduction could be carried out. H2O R package (https://cran.r-project.org/web/packages/h2o/) was adapted to perform autoencoder analysis with the gene number in bottleneck layer specified as 50.

2.4 Construction of prognostic model

Univariate Cox-regression analysis was performed to screen for genes that were significantly associated with breast cancer's overall survival (OS). LASSO Cox-regression analysis was further used to construct the prognostic model by glmnet (Engebretsen S, 2019) function package in R and calculated the risk score. The risk score was calculated on the basis of the following formula:

$$Risk\ score = \sum_{i=1}^{n} Coef_i * x_i$$

$Coef_i$ was the risk coefficient of each factor calculated by the LASSO-Cox model, and x_i was the expression value of each factor, which referred to the gene expression level in this study.

2.5 Survival analysis

The survival probability of breast cancer patient was assessed by the Kaplan-Meier survival curve using the survival package (https://cran.r-project.org/web/packages/survival/) in R language. Log-rank test was used to assess the difference of OS between different groups with the significant threshold of $P<0.05$.

2.6 Statistical analysis

Chi-square test was used to compare distributions of samples across different clinicopathological groups between training and testing set. Univariate Cox-regression analysis was adapted to screen OS-related genes. Multivariate Cox-regression analysis

was performed for the determination of significance between multiple factors and patients' OS. Statistical analysis was applied using *R version 3.5.2.* The above threshold was $p < 0.05$ unless otherwise specified.

3. Results

3.1 Differential expression genes in BRCA tumor samples

A total of 603 DEGs including 205 up- and 398 down-regulated ones were identified in BRCA tumor samples compared with their paracancerous samples. Figure 1A illustrated the differential expression pattern of all genes and Figure 1B showed the Z-score normalized mRNA expressions of the 603 DEGs in paired BRCA tumor and paracancerous tissues.

3.2 Autoencoder screened genes

For the apparent correlation among the DEGs, i.e. they were significantly up- or down-regulated in BRCA tumor samples, we applied autoencoder algorithm, which is a neural network-based learning technique for representation learning, for screening the most representative genes for following analysis. The input features were expression profiles of the 603 genes in paired BRCA tumor and paracancerous tissues, and the number of layer and gene number contained in the bottleneck layer were specified as 5 and 50, respectively. Table 2 provided the finally retained 50 genes along with their relative importance.

3.3 Radiotherapy sensitivity prediction model

Univariate Cox-regression analysis identified seven out of the 50 autoencoder remained genes including *HOXB13*, *NKX2-2*, *ADAMTS20*, *LINC00898*, *LOC284930*, *ACTL8* and *LOC101928978* were significantly correlated with BRCA patients' overall

147 survival (OS) from the training set. Figure 2A illustrated the hazard ratio (HR) and
148 significant p value of those genes. LASSO Cox-regression analysis determined six genes
149 based on which the partial likelihood deviance (PLD) had the lowest value (Figure 2B),
150 and the regression model was built as an equation: risk score = 0.0137*mRNA level of
151 *HOXB13* + 0.0928*mRNA level of *NKX2-2* + 0.0343*mRNA level of *ADAMTS20* +
152 0.103*mRNA level of *LOC284930* + 0.0419*mRNA level of *ACTL8* + 0.0871*mRNA
153 level of *LOC101928978*.

154 3.4 Risk score is an independent marker for radiotherapy sensitivity prediction

155 Samples that have been treated with radiation from both training and testing set were
156 assigned scores according to the risk score equation. Kaplan-Meier (KM) survival
157 analysis stratified by the median risk score uncovered its unfavorable survival role for
158 BRCA patients in training set (Figure 3A), which should indicate that high risk score
159 correlated with insensitive response to radiotherapy. Besides, there were 65 patients had
160 radiotherapy information out of the 1019 breast cancer samples from ICGC, which were
161 also assigned risk scores according to the risk score equation. Kaplan-Meier analysis
162 indicated significantly inferior OS of samples with higher risk score. To further explore if
163 risk score was independent of other common clinicopathological features in predicting
164 radiotherapy sensitivity, we performed multivariate Cox regression analysis by
165 simultaneously taking age, stage and risk score into account. As a result, risk score was
166 proved to significantly correlate with the overall survival of BRCA patients that have
167 treated with radiation in both TCGA (Figure 3C) and ICGC sets (Figure 3D).

168 **4. Discussion**

Deep learning models have been widely used in the area of bioinformatics, including biomedical signal processing, biomedical imaging, omics and tumor classification (Min S, 2017), (Lee D, 2017), (Liu J, 2017). Some deep neural network models are capable of learning a meaningful latent space, which could be used to explore and generate hypothetical gene expression profiles under various types of molecular and genetic perturbation, or to predict a tumor's response to specific therapies (Way G P, 2018). In this study, we identified 603 genes that differently expressed in BRCA tumor samples compared with paracancerous samples. Among them, the most representative genes were screened using a neural network-based learning technique, which is mainly used for data dimensionality reduction or feature extraction. 50 genes were considered relatively important, and their associations with BRCA patients' OS was further investigated by univariate Cox-regression analysis. Finally, we reduce the number of candidate key genes to 6 by using LASSO Cox-regression analysis, including *HOXB13*, *NKX2-2*, *ADAMTS20*, *LOC284930*, *ACTL8* and *LOC101928978*.

Homeobox B13 (*HOXB13*) is a member of human ANTP class homeobox gene and is located at 17q21.32 of chromosome (Holland P W H, 2007). The expression of *HOXB13* have been associated with the development of several cancers. For example, *HOXB13* was reported to be able to mediate NF-κB/p65 pathway, and regulate the proliferation and metastasis of esophageal squamous cell carcinoma (Li R, 2018). Besides, *HOXB13* expression was significantly associated with prostate ductal type adenocarcinoma and biochemical recurrence (BCR) as well as shorter BCR-free survival (Cho Y A, 2018) In BRCA, *HOXB13* has long been identified as a prognostic biomarker (Xiao-Jun M, 2006; Wang Z, 2007). *HOXB13* can confer tamoxifen resistance by directly

192 downregulating the transcription of estrogen receptor α (ERα), and transcriptionally
193 upregulated interleukin (IL)-6, activating the mTOR pathway via STAT3
194 phosphorylation to promote the proliferation of BRCA tumor cells and the recruitment of
195 fibroblast, leading to disease progression and recurrence (Shah N, 2013).

196 A disintegrin and metalloprotease domains with thrombospondins motifs
197 (ADAMTSs) are complex extracellular proteases that have been related to both
198 oncogenic and tumor-protective functions (Cal S, 2015). Multiple subtypes of ADAMTSs
199 was proved to play a role in the development of BRCA. For example, *ADAMTS1*
200 expression was decreased in BRCA, which can stimulate the migration and invasion of
201 breast cancer cells *in vitro*. It can also response to VEGF, and implicate in tissue
202 remodeling events observed in cancer development (Freitas V M, 2013; Suély V, 2016).
203 *ADAMTS6* suppressed tumor progression via the ERK signaling pathway and might serve
204 as a prognostic marker in BRCA (Xie Y, 2014). The expression level of *ADAMTS20* was
205 also significantly associated with histological grade of breast invasive ductal carcinoma
206 (Guo X, 2018). Meanwhile, the actin-like protein 8 (ACTL8) protein was reported to be
207 highly expressed in BRCA specimens and is closely correlated with the
208 clinicopathological features and prognosis (He L C, 2016)

209 *NKX2-2* is an oligodendroglial and astrocytic lineage marker, and also a useful
210 immunohistochemical marker for Ewing sarcoma (Rousseau A, 2006; Yoshida A, 2012).
211 Yang Y et al. indicated that, the high expression of *NKX2-2* was significantly correlated
212 with poor OS for all invasive breast cancer patients (Yang Y, 2018). *NKX2-2* is one of the
213 downstream target genes of GLI1 in the Sonic hedgehog (Shh) signaling pathway, and
214 impairment of this pathway can result in both birth defects and cancer (Shahi M H, 2010).

However, the role of *NKX2-2* in breast cancer has not been revealed. Unfortunately, little research has been done on the functions of *LOC284930* and *LOC101928978*. Only one report suggested that, the expression of *LOC284930* was positively correlated with ERG overexpression, which is the most frequent genomic rearrangement in prostate cancer (Boormans J L, 2013).

Overall, 3 of the 6 key genes have been confirmed to be involved in the development of BRCA and are expected to serve as prognostic indicators. To our knowledge, this is the first time these genes was linked to the radiotherapy sensitivity of BRCA.

5. Conclusions

In this study, a signature for BRCA radiotherapy sensitivity prediction was development based on the expression of 6 characteristic DEGs, including *HOXB13*, *NKX2-2*, *ADAMTS20*, *LOC284930*, *ACTL8* and *LOC101928978,* in BRCA tumor samples compared with their paracancerous samples for the first time. A radiotherapy sensitivity prediction signature was constructed with these characteristic DEGs, and this signature was proved to be reliable.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The dataset(s) supporting the conclusions of this article is(are) available in the TCGA (http://tcgaportal.org/) and ICGC (https://icgc.org/).

Conflicts of Interest: The authors declare that they have no competing interests.

Funding: Sponsored by Fujian Provincial Health Technology Project (No. 2018-CX-1).

Authors' contributions: Xing Chen, Junjie Zheng put forward the ideas of this article, written this article and analysed the data. Ailong Zhang helped for acquisition of data and analysis and interpretation of data. Xing Chen, Zhenhui You helped for revising the manuscript All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

References

Ahmad Aamir. Breast Cancer Statistics: Recent Trends. Advances in experimental medicine and biology, 2019; 1152.

Boormans JL, Korsten H, Ziel-van der Made AJ, van Leenders GJ, de Vos CV, Jenster G. et al. Identification of TDRD1 as a direct target gene of ERG in primary prostate cancer. International Journal of Cancer, 2013; 133: 335-345.

Cal S, López-Otín, Carlos. ADAMTS proteases and cancer. Matrix Biology, 2015; 44-46,77-85.

Caldon L J M, Walters S J, Ratcliffe J, Reed MW. What influences clinicians' operative preferences for women with breast cancer? An application of the discrete choice experiment. European Journal of Cancer, 2007; 43: 0-1669.

Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. Ned Tijdschr Geneeskd, 2016; 375: 717-729.

Cho Y A, Park C K, Cho N H. 217P HoxB13 expression in prostate cancer: Clinicopathologic characteristics and role as potential diagnostic marker. Annals of Oncology, 2018; 29(suppl_9): mdy434. 005.

DeSantis C E, Ma J. Gaudet M M, et al. Breast cancer statistics. CA: a cancer journal for clinicians, 2019; 69: 438-451.

Engebretsen S, Bohlin J. Statistical predictions with glmnet. Clin Epigenetics, 2019; 11: 123.

Freitas VM, do Amaral JB, Silva TA, Santos ES, Mangone FR, Pinheiro Jde J et al. Decreased expression of ADAMTS-1 in human breast tumors stimulates migration and invasion. Molecular Cancer, 2013; 12: 2.

Gee HE, Buffa FM, Harris AL, Toohey JM, Carroll SL, Cooper CL. et al. microRNA-related DNA repair/cell-cycle genes independently associated with relapse following radiotherapy for early breast cancer. International journal of radiation oncology, biology, physics, 2015; 93: 1104-1114.

Xuhui Guo, Juntao Li, Hengwei Zhang, Hui Liu, Zhenzhen Liu, Xudong Wei. Relationship Between ADAMTS8, ADAMTS18, and ADAMTS20 (A Disintegrin and Metalloproteinase with Thrombospondin Motifs) Expressions and Tumor Molecular Classification, Clinical Pathological Parameters, and Prognosis in Breast Invasive Ductal Carcinoma. Medical Science Monitor. 2018; 24: 3726-3735.

He L C, Zhu Z F, Zheng Q Q, et al. ACTL8 expression and its correlation with clinicopathological features and prognosis in breast cancer. Chin J Pathophysiol, 2016; 32: 1138-1141.

Holland P W H, Booth H A F, Bruford E A. Classification and nomenclature of all human homeobox genes. BMC Biology, 2007; 5: 47. DOI: 10.1186/1741-7007-5-47

281 Lee D, Yoo J, Ye J C. Deep residual learning for compressed sensing MRI[C]//2017
282 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE,
283 2017; 15-18.
284 Leila, Dorling, Gillian, et al. Patients with a High Polygenic Risk of Breast Cancer do not
285 have An Increased Risk of Radiotherapy Toxicity. Clinical cancer research: an
286 official journal of the American Association for Cancer Research. 2016.
287 Jian Liu, Xuesong Wang, Yuhu Cheng, and Lin Zhang. Tumor gene expression data
288 classification via sample expansion- based deep learning. Oncotarget, 2017; 8:
289 109646-109660.
290 Li RC, Ke S, Meng FK, Lu J, Zou XJ, He ZG, et al. CiRS-7 promotes growth and
291 metastasis of esophageal squamous cell carcinoma via regulation of miR-7/HOXB13.
292 Cell Death & Disease, 2018; 9.
293 Meattini I, Saieva C, Marrazzo L, Di Brina L, Pallotta S, Mangoni M, et al. Accelerated
294 partial breast irradiation using intensity-modulated radiotherapy versus whole breast
295 irradiation: 5-year survival analysis of a phase 3 randomised controlled trial.
296 European Journal of Cancer, 2015; 51: 451-463.
297 Min S, Lee B, Yoon S. Deep learning in bioinformatics. Briefings in bioinformatics, 2017;
298 18: 851-869.
299 Munkácsy G, Szász MA, Menyhárt O. Gene expression-based prognostic and predictive
300 tools in breast cancer. Breast Cancer, 2015; 22: 245-252.
301 Robinson M D, Mccarthy D J, Smyth G K. edgeR: a Bioconductor package for
302 differential expression analysis of digital gene expression data. Bioinformatics, 2010;
303 26: 139-140.

Rousseau A, Nutt CL, Betensky RA, Iafrate AJ, Han M, Ligon KL, et al. Expression of oligodendroglial and astrocytic lineage markers in diffuse gliomas: use of YKL-40, ApoE, ASCL1, and NKX2-2. Journal of Neuropathology & Experimental Neurology, 2006; 65: 1149-1156.

Sagara Y, Freedman RA, Vaz-Luis I, Mallory MA, Wong SM, Aydogan F, et al. Patient Prognostic Score and Associations With Survival Improvement Offered by Radiotherapy After Breast-Conserving Surgery for Ductal Carcinoma In Situ: A Population-Based Longitudinal Cohort Study. Journal of Clinical Oncology, 2016; 10:34:1190-6.

Speers C, Pierce LJ. Molecular Signatures of Radiation Response in Breast Cancer: Towards Personalized Decision-Making in Radiation Treatment. Int J Breast Cancer, 2017; 2017: 4279724.

Shah N, Jin K, Cruz LA, Park S, Sadik H, Cho S, et al. HOXB13 mediates tamoxifen resistance and invasiveness in human breast cancer by suppressing ERα and inducing IL-6 expression. Cancer research, 2013; 73: 5449-5458.

Shahi MH, Afzal M, Sinha S, Eberhart CG, Rey JA, Fan X, Castresana JS. Regulation of sonic hedgehog-GLI1 downstream target genes PTCH1, Cyclin D2, Plakoglobin, PAX6 and NKX2. 2 and their epigenetic status in medulloblastoma and astrocytoma. BMC cancer, 2010; 10: 614.

Suély V. Silva, Maíra A. Lima, Nathalie Cella, Ruy G. Jaeger, Vanessa M Freitas. ADAMTS-1 Is found in the nuclei of normal and tumoral breast cells. PLoS ONE, 2016; 11.

326 Taylor C, Correa C, Duane FK, Aznar MC, Anderson SJ, Bergh J, et al. Estimating the
327 Risks of Breast Cancer Radiotherapy: Evidence From Modern Radiation Doses to
328 the Lungs and Heart and From Previous Randomized Trials. Journal of Clinical
329 Oncology. 2017; 35: 1641-1649.
330 van Maaren MC, de Munck L, de Bock GH, Jobsen JJ, van Dalen T, Linn SC, et al. 10
331 year survival after breast-conserving surgery plus radiotherapy compared with
332 mastectomy in early breast cancer in the Netherlands: a population-based study. The
333 Lancet Oncology, 2016; S1470204516300675.
334 Wang Z, Dahiya S, Provencher H, Muir B, Carney E, Coser K, et al. The Prognostic
335 Biomarkers HOXB13, IL17BR, and CHDH Are Regulated by Estrogen in Breast
336 Cancer. Clinical Cancer Research, 2007; 13: 6327-6334.
337 Way G P, Greene C S. Extracting a biologically relevant latent space from cancer
338 transcriptomes with variational autoencoders. Pac Symp Biocomput, 2018; 23: 80-
339 91.
340 Ma XJ, Hilsenbeck SG, Wang W, Ding L, Sgroi DC, Bender RA, et al. The
341 HOXB13:IL17BR expression index is a prognostic factor in early-stage breast
342 cancer. Journal of Clinical Oncology Official Journal of the American Society of
343 Clinical Oncology, 2006; 24: 4611-9.
344 Xie Y, Gou Q, Xie K, Wang Z, Wang Y, Zheng H. ADAMTS6 suppresses tumor
345 progression via the ERK signaling pathway and serves as a prognostic marker in
346 human breast cancer. Oncotarget, 2014; 7: 61273-61283.
347 Yongde Yang, Yanlin Zhou, Xin Xiong, Man Huang, Xueyan Ying and Mengyuan Wang.
348 ALG3 Is Activated by Heat Shock Factor 2 and Promotes Breast Cancer Growth.

Medical science monitor: international medical journal of experimental and clinical research, 2018; 24: 3479-3487.

Yoshida A, Sekine S, Tsuta K, Fukayama M, Furuta K, Tsuda H.. NKX2. 2 is a useful immunohistochemical marker for Ewing sarcoma. The American journal of surgical pathology, 2012; 36: 993-999.

Figure legends

Figure 1 Differential gene expression analysis between paired BRCA tumor and paracancerous tissues from TCGA. (A) Volcano plot displaying gene differential expression pattern in BRCA tumor tissues compared with paracancerous tissues. Red and blue dot represents down- and up-regulated genes, respectively. Green dots are nondifferential expression genes. (B) Heatmap showing differential expression genes' (DEGs') mRNA level after Z-score normalization in paired BRCA tumor and paracancerous tissues. Horizontal and vertical axis represents genes and samples, respectively.

Figure 2 Construction of BRCA radiotherapy sensitivity prediction model. (A) Seven genes that significantly associated with overall survival (OS) of BRCA patients in radiotherapy group. Square data indicate the hazard ratios (HRs) with error bars are 95% confidence intervals (CI). (B) Selection of optimal variables, i.e. genes here, for constructing prediction model through LASSO Cox-regression method. The optimal variable number was determined by the vertical dashed line at which the Partial Likelihood Deviance (PLD) was lowest.

Figure 3 Risk score is an independent factor for BRCA radiotherapy sensitivity. (A) Kaplan-Meier (KM) plot of BRCA patients in radiotherapy group from TCGA stratified

by the median risk score. P value calculated using log-rank test was provided. (B) KM plot of BRCA patients in radiotherapy group from ICGC stratified by the median risk score. P value calculated using log-rank test was provided. (C) Forest plot of multivariate Cox regression analysis of the training set indicated the risk score as an independent marker for BRCA radiotherapy sensitivity. (D) Forest plot of multivariate Cox regression analysis of the testing set indicated the risk score as an independent marker for BRCA radiotherapy sensitivity.

Tables

Table 1 Clinicopathological characteristics of BRCA radiotherapy samples from TCGA database.

Characteristics	Radiotherapy Patients		χ^2	P value
	Training cohort (N = 652)	Testing cohort (N = 65)		
Age (Mean±SD)	57.60±9.22	56.97±10.94	0.0035	0.95
Pathologic stage				
I	99 (15.18%)	26 (40%)		
II	310 (47.55%)	30 (46.15%)	7.1183	0.076
III	174 (26.69%)	9 (13.85%)		
IV	9 (1.38%)	0 (0%)		
Survival time				
＜ Median	341 (52.3%)	26 (40%)	2.57	0.11
＞ Median	311 (47.7%)	39 (60%)		
Receptor status				

Non triple-negative	567 (86.9%)	/	/	/
triple-negative	85 (13.1%)	/		
OS status				
Dead	53 (8.13%)	9 (13.85%)		
Alive	539 (82.67%)	56 (86.15%)	0.69	0.41
Unknown	60 (9.20%)	0 (0%)		

382 **Table 2** The 50 genes remained through autoencoder algorithm.

Gene	Score	Gene	Score
HOXB13	1	FCRL4	0.796
CXADRP3	0.973	LINC00898	0.795
GNGT1	0.941	LOC101928932	0.793
NKX2_2	0.918	TLX1	0.790
POTEC	0.909	MS4A15	0.790
PYDC1	0.897	GOLGA8T	0.787
DSCAM.AS1	0.886	LOC284930	0.781
MAGEA6	0.867	MIR8071_1	0.778
LINC01644	0.866	UBE2E2.AS1	0.775
ABCC13	0.860	EIF4E1B	0.775
VSTM2A_OT1	0.856	ASCL1	0.771
CLPS	0.843	WT1_AS	0.765
C5orf66_AS1	0.842	LINC00052	0.762
EPHA8	0.841	GOLGA6L3	0.762
C8orf34_AS1	0.839	LINC00628	0.761

KISS1R	0.837	OPRPN	0.760
ADAMTS20	0.828	GAL3ST2	0.760
LINC01844	0.827	LINC00466	0.759
DLX2_DT	0.820	ACTL8	0.757
FOXD3_AS1	0.816	LINC01344	0.755
LOC339685	0.814	GPR139	0.750
WT1	0.809	RTBDN	0.750
MAGEA3	0.807	TMEM270	0.748
TACR3	0.803	LOC101928978	0.747
LHFPL5	0.803	PRAC2	0.744

383

Figures

Figure 1

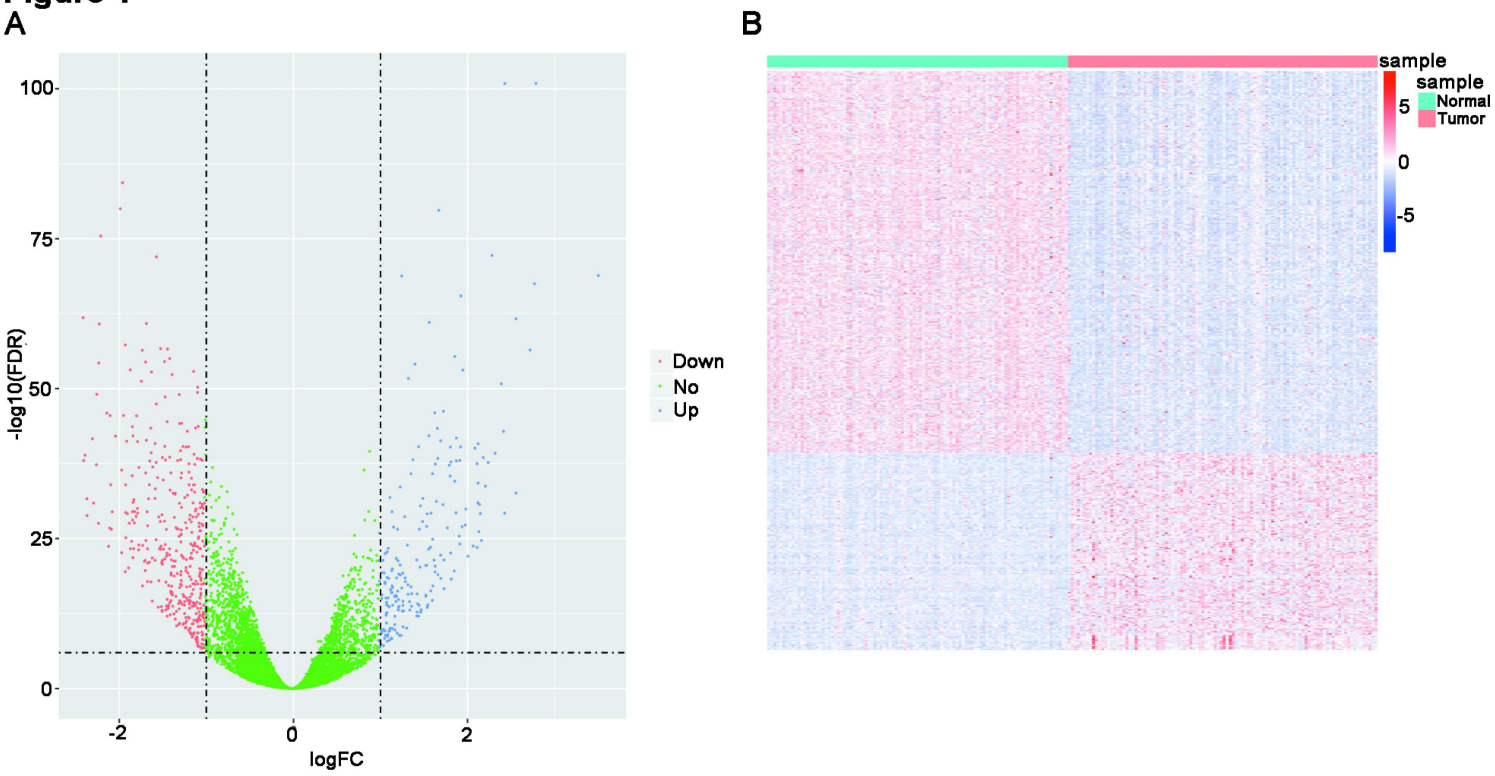

Figure 1

Differential gene expression analysis between paired BRCA tumor and paracancerous tissues from TCGA. (A) Volcano plot displaying gene differential expression pattern in BRCA tumor tissues compared with paracancerous tissues. Red and blue dot represents down- and up-regulated genes, respectively. Green dots are nondifferential expression genes. (B) Heatmap showing differential expression genes' (DEGs') mRNA level after Z-score normalization in paired BRCA tumor and paracancerous tissues. Horizontal and vertical axis represents genes and samples, respectively.

Figure 2

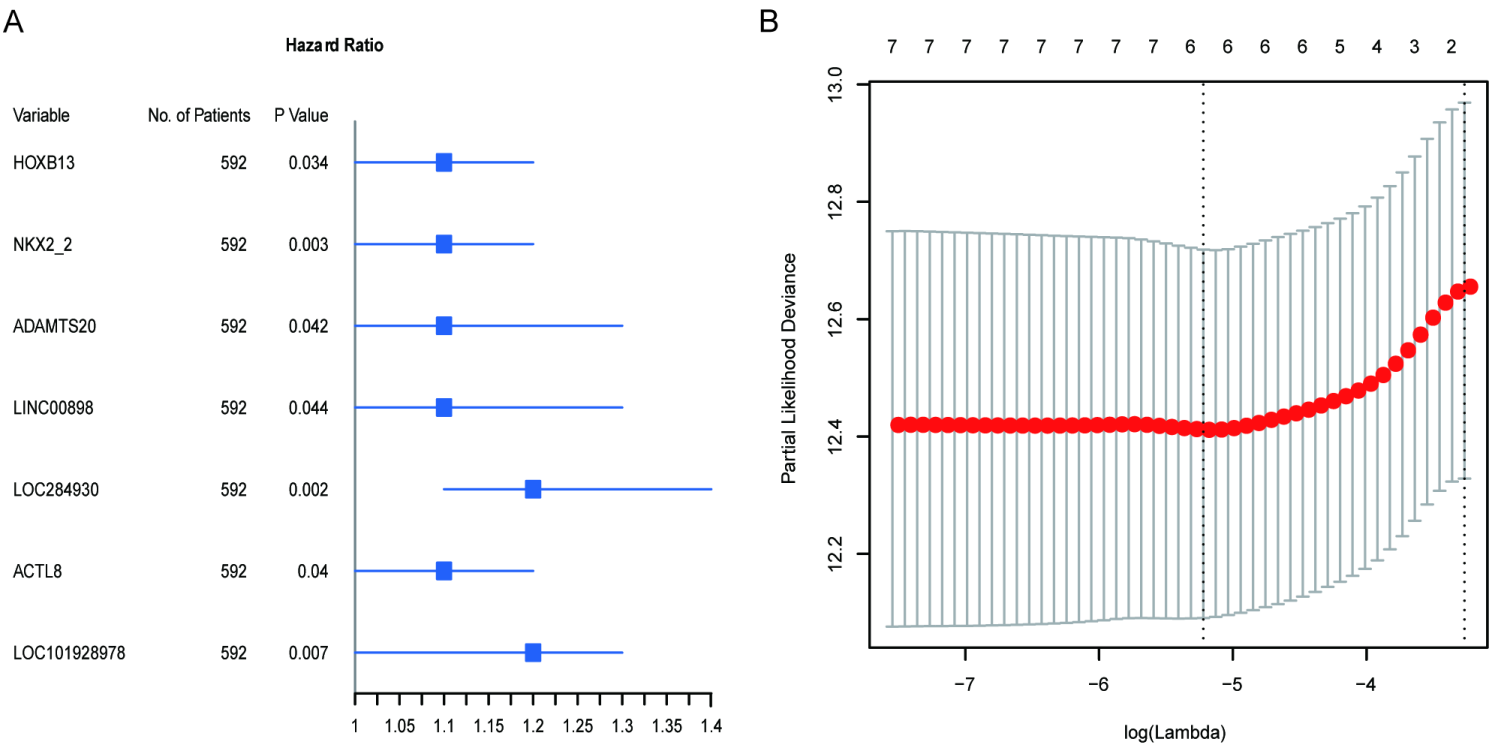

Figure 2

Construction of BRCA radiotherapy sensitivity prediction model. (A) Seven genes that significantly associated with overall survival (OS) of BRCA patients in radiotherapy group. Square data indicate the hazard ratios (HRs) with error bars are 95% confidence intervals (CI). (B) Selection of optimal variables, i.e. genes here, for constructing prediction model through LASSO Cox-regression method. The optimal variable number was determined by the vertical dashed line at which the Partial Likelihood Deviance (PLD) was lowest.

Figure 3

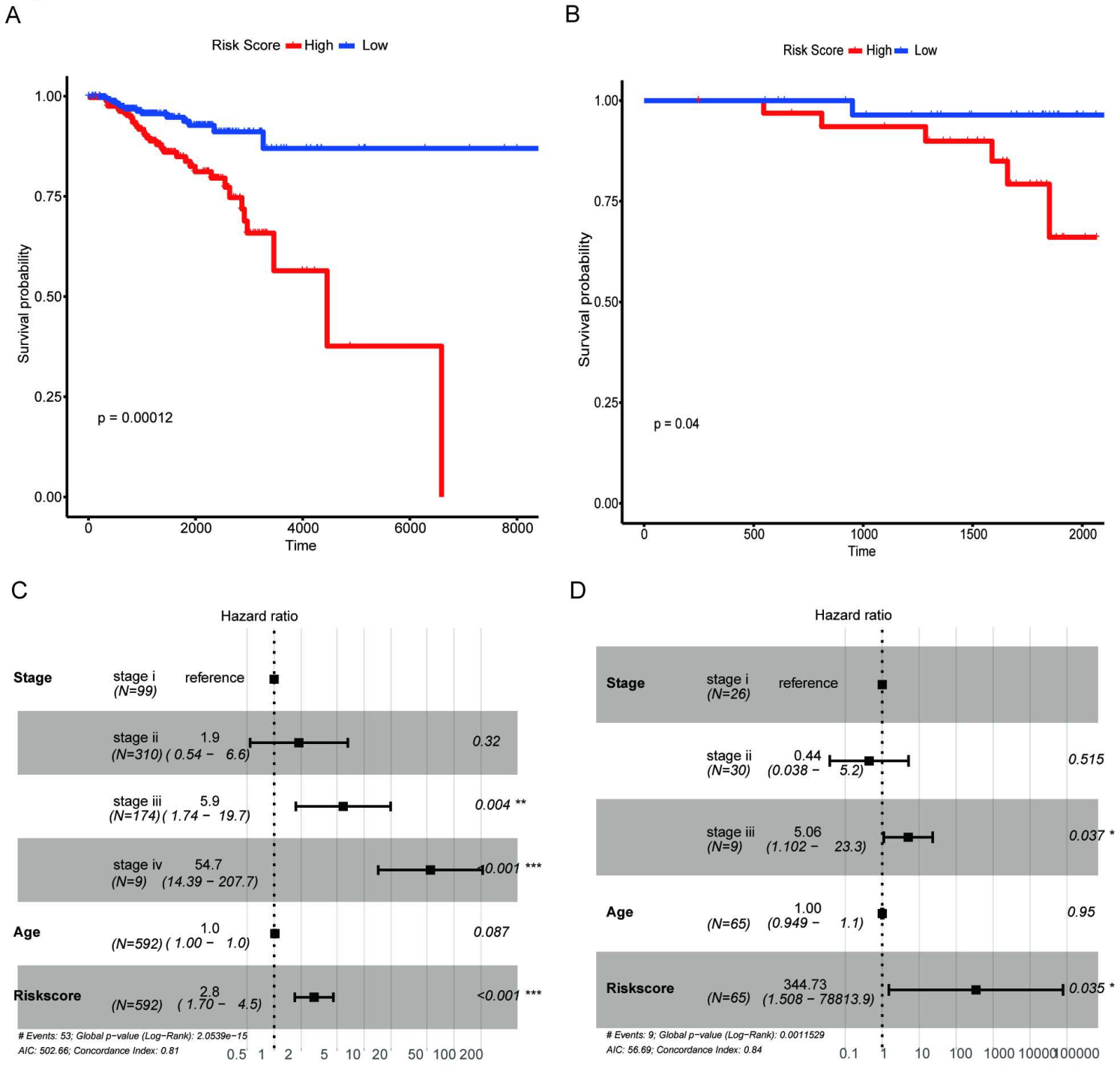

Figure 3

Risk score is an independent factor for BRCA radiotherapy sensitivity. (A) Kaplan-Meier (KM) plot of BRCA patients in radiotherapy group from TCGA stratified by the median risk score. P value calculated using log-rank test was provided. (B) KM plot of BRCA patients in radiotherapy group from ICGC stratified by the median risk score. P value calculated using log-rank test was provided. (C) Forest plot of multivariate Cox regression analysis of the training set indicated the risk score as an independent marker for BRCA

radiotherapy sensitivity. (D) Forest plot of multivariate Cox regression analysis of the testing set indicated the risk score as an independent marker for BRCA radiotherapy sensitivity.