Submitted: 03 July, 2023 Accepted: 07 August, 2023 Published: 15 April, 2025 DOI:10.22514/ejgo.2025.050

ORIGINAL RESEARCH

European Journal of Gynaecological Oncology

Development of a random survival forest model based on cuproptosis-related genes (CRGs) for predicting overall survival in patients with invasive breast carcinoma

Fangfang Zhu[1,*], Dafang Xu[1,*]

[1] First affiliated Hospital of Huzhou University, 313204 Huzhou, Zhejiang, China

***Correspondence**
dr_zhuff@163.com;
dr_zhuff163@outlook.com
(Fangfang Zhu);
lengyue19860426@163.com
(Dafang Xu)

Abstract

Background: The clinical decision-making of invasive cancer (BRCA) depends on the prediction of overall survival rate. To predict the overall survival rate of BRCA patients, a random survival forest (RSF) model based on copper poisoning related genes (CRGs) was established. **Methods**: We analyzed the expression level of CRG using cell lines. The Cancer Genome Map (TCGA)-BRCA data is used to develop and evaluate RSF models. We analyzed the relationships between various clinical parameters, functional enrichment, immune cell ratio and RSF scores, as well as the IC_{50} of various drugs in the Cancer Drug Sensitivity Genome 2 (GDSC2) database. **Results**: Compared with normal control cell lines, CRG in BRCA cell lines is upregulated. The RSF model performs well in predicting the overall survival rate of BRCA patients. There were significant differences in RSF scores among BRCA patients in terms of age, radiation status, staging, T staging, and N staging (p-value < 0.05). In BRCA samples with higher RSF scores, hypoxia, glycolysis, mechanism target of rapamycin complex 1 (mTORC1) signaling, DNA replication, and cell cycle were all enhanced; On the contrary, inflammatory responses, natural killer cells, mature B cell differentiation, mediated cytotoxicity, and autophagy regulation are all inhibited. The proportion of immature B cells, activated dendritic cells, resting memory differentiation cluster 4 (CD4) T cells, and follicle helper T cells was significantly correlated with RSF scores (p-value < 0.05), while M2 macrophages, neutrophils, and immature CD4 T cells were negatively correlated. Higher RSF scores were associated with increased resistance to VX-11e_2096 and ERK_6604_1714 but greater sensitivity to Acetalax_1804, WEHI-539_1997 and AZD5991_1720. **Conclusions**: The RSF score is related to various clinical features, immune cell ratio, and drug sensitivity. It is an effective tool for predicting the overall survival rate of BRCA patients.

Keywords

Breast invasive carcinoma; Random survival forest model; Cuproptosis-related genes; Clinical characteristics; Drug sensitivity

1. Introduction

Invasive breast carcinoma (BRCA) is a significant health concern that affects women worldwide [1, 2]. According to data published by the World Health Organization (WHO), the 5-year survival rate of BRCA patients is approximately 40–60% [3]. Therefore, accurately predicting overall survival in BRCA patients is crucial if we are to determine appropriate treatment strategies and manage patient care in an effective manner [4]. Several methods have been developed to predict the overall survival of BRCA patients; however, these approaches are associated with numerous limitations [5]. These include challenges related to small sample sizes, heterogeneity within BRCA subtypes, and variations in treatment regimens. Additionally, the reliance on clinical parameters alone may not fully capture the complexities of BRCA progression and response to therapy. Therefore, there is an urgent need to develop new prognostic tools with greater levels of accuracy.

The prognosis of BRCA patients is highly dependent on the stage at which the cancer is diagnosed [6]. Early detection and timely intervention can significantly improve the chances of successful treatment and long-term survival [7–10]. However, accurately predicting the overall survival of BRCA patients remains a significant challenge due to the complex biological mechanisms underlying this disease [11]. Several factors, including genetic variations, lifestyle choices and environmental variables, all play a significant role in the growth and

progression of BRCA [12, 13]. Therefore, it is essential that we develop accurate prognostic tools that can optimize clinical decision-making and improve patient outcomes.

Over recent years, researchers have focused on identifying novel biomarkers and pathways that could assist in predicting the overall survival of patients with BRCA [14, 15]. Cuproptosis, a recently discovered mechanism of cell death, is known to be associated with the onset and spread of cancer [16]. Recent studies [17, 18] identified an association between cuproptosis-related genes (CRGs) and a multitude of biological events, including oxidative stress, inflammatory processes and genomic instability. In a variety of cancers, particularly lung cancer [19], glioblastoma [20] and colon cancer [21], CRGs have been identified as potential biomarkers for predicting overall survival. Because of this, we can hypothesize that a CRG-based approach could provide a more accurate and trustworthy tool for calculating the overall survival of patients with BRCA.

Although various techniques have been developed to predict the overall survival of patients with BRCA, these strategies have limitations. These techniques include traditional clinical models that utilize patient demographics, tumor characteristics and treatment history [22]. However, these models often fail to consider molecular and genetic factors, thus limiting their accuracy and personalized predictions. In addition, genomic-based approaches that analyze gene expression profiles and somatic mutations have shown promise but may not fully capture the full complexity of BRCA heterogeneity [23, 24]; furthermore, clinical parameters may not accurately reflect the complex biology of the disease itself [25]. Therefore, machine learning algorithms, particularly the random survival forest (RSF) algorithm, have gained significant recognition as an alternative method for predicting the overall survival of cancer patients [26–28]. As compared to survival support vector machine (SVM) and other machine learning algorithms, the superiority of RSF lies in its ability to handle censored survival data in an effect manner, a common challenge in survival analysis, by incorporating censored status during tree construction [29]. In addition, RSF collates decision trees to capture complex non-linear relationships between predictors and survival times more efficiently than SVM [29, 30]. This advantage is substantiated by studies demonstrating the improved predictive performance of RSF [31]. Moreover, RSF provides variable importance measures, thus facilitating the identification of key prognostic factors [32]. The robustness of the model against overfitting, and its simplicity in terms of implementation, renders it highly accessible to researchers and practitioners with varying levels of expertise [33]. The purpose of this study was to create and assess an RSF model that utilizes CRGs for predicting the overall survival of patients with BRCA.

Our specific goals were as follows: (1) utilize information from The Cancer Genome Atlas (TCGA) to construct and assess the efficacy of an RSF model that utilizes CRGs to forecast the overall survival of individuals with BRCA; (2) investigate the correlation between RSF scores and other characteristics, such as clinical traits, functional enrichment, immune cell proportions and drug sensitivity, and (3) provide insights into the biological mechanisms underlying BRCA. Achieving these objectives will provide valuable information for personalized treatment strategies and the development of new therapeutic targets for BRCA patients. In addition, our findings will contribute to the growing body of knowledge related to cuproptosis and its potential role in cancer prognosis; this information could also have implications for other types of cancer.

2. Methods

2.1 Identifying and selecting CRGs

Some CRGs were identified by analyzing previous studies [16, 34]. Based on our literature search, we selected the following as CRGs for analysis: *FDX1*, *LIAS*, *LIPT1*, *DLD*, *DLAT*, *PDHA1*, *PDHB*, *MTF1*, *GLS* and *CDKN2A*.

2.2 *In vitro* expression analysis of CRGs

2.2.1 Cell lines

One BRCA cell line (BT 20) and one normal mammary gland cell line (HMEC) were purchased from the American Type Culture Collection (ATCC, USA) and cultured in accordance with the manufacturer's instructions.

2.2.2 Total RNA extraction

Total RNA was extracted from both the BRCA and normal cell lines by isopycnic centrifugation, as described previously [35]. The extracted RNA was then incubated with RNase-free DNase I (Roche, Germany) at 37 °C for 15 minutes. The quality of the extracted total RNA was then assessed by considering the 260/280 ratio *via* 2100 Bioanalyzer (Agilent Technologies, Germany).

2.2.3 RNA-seq analysis

RNA samples were sent to Macrogen, Seoul, South Korea, for RNA-seq analysis. Following RNA-seq analysis, we identified the normalized gene expression values of the CRGs in reads per kilo base million reads (RPKM) and fragments per kilo base million reads (FPKM). The obtained FPKM values against CRGs in BRCA cells and the normal control cell line were then compared to identify differences in expression level.

2.3 Analysis of genetic alterations

cBioPortal (https://www.cbioportal.org/) is a widely used online platform for the analysis of genetic alterations in cancer research [36]; this software allows researchers to investigate and visualize genomic data from various cancer studies, thus providing valuable insights into the genetic landscape of different tumor types. cBioPortal also offers user-friendly tools for analyzing and interpreting genetic alterations, facilitating the discovery of potential therapeutic targets and biomarkers for precision medicine approaches. In the present research, we performed mutational analysis of CRGs in BRCA samples.

2.4 Data sources and preprocessing for RSF construction

First, we searched the TCGA database (https://www.cancer.gov/ccg/research/genome-sequencing/tcga) and retrieved a TCGA-BRCA dataset [37]. This dataset

contained RNA-seq information as well as statistics relating to the overall survival of patients with BRCA. Samples with incomplete information relating to overall survival or low levels of CRG expression were excluded. Our analysis included a total of 1057 BRCA patients (Table 1).

2.5 Construction of the RSF model

The RSF model is a powerful machine-learning algorithm that is used to predict survival. In this study, the RSF model was developed using the "randomForestSRC" package in R software (https://www.r-project.org/). The randomForestSRC package generates decision curves by evaluating the net benefit of using the RSF model across a range of probability thresholds. The tool plots the proportion of true positives against the proportion of false positives at various threshold levels. Moreover, the randomForestSRC package calculates RSF scores by aggregating predictions from multiple decision trees. To develop our new RSF model, we used the CRGs as predictor variables, and used overall survival time and censoring status as response variables. During analysis, the number of trees was set at 500; this allowed the accurate calculation of RSF scores. The optimal tuning parameters for the RSF model were selected using the "tune.randomForestSRC" function in the "randomForestSRC" package. The tune.randomForestSRC function employs cross-validation and grid search techniques to identify the best combination of hyperparameters by considering data and a range of hyperparameter values. The function then performs cross-validation by splitting the data into subsets (folds), and for each combination of hyperparameters, the package trains the RSF model using a subset of data and tests it on the remaining data. The performance of the model was evaluated using a selected evaluation metric (*e.g.*, the concordance index or the log-likelihood); the hyperparameters that yield the best performance were then selected. Thus, the tune.randomForestSRC package plays a crucial role in pinpointing the optimal hyperparameters for the RSF model, thereby enhancing the accuracy and precision of the survival prediction model. Ultimately, the discriminatory power of the RSF model was evaluated using time-dependent receiver operating characteristic curve (ROC) curves. The decision curves were used to compare the overall advantage of various prediction criteria in order to assess the clinical applicability of the RSF model.

2.6 Clinical characteristics, functional enrichment, immune cell proportions, and drug sensitivity analysis

The Wilcoxon rank-sum test and Kruskal-Wallis test were used to analyze the relationship between RSF scores with additional clinical variables. The "clusterProfiler" module in the R package was used to carry out gene set enrichment analysis (GSEA) [38]. The GSEA analysis often includes the prediction or identification of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG).

In addition, the "CIBERSORT" tool in the R package was used to calculate immune cell proportions, and Spearman's correlation coefficient was used to investigate the relationship between RSF scores and immune cell proportions [39]. The Genomics of Drug Sensitivity in Cancer 2 (GDSC2) dataset, which provides data relating to the drug sensitivity of cell lines with cancer, along with pRRophetic software, were used to analyze drug sensitivity [40, 41]. Spearman's correlation coefficient was used to analyze the relationship between RSF grades and sensitivity to drugs.

TABLE 1. Clinical data of the samples included in the TCGA-BRCA dataset.

Sr. No	Clinicopathological feature		No. Samples	Number of excluded samples with missing information	Number of total included samples
1	Age distribution	Above 50	562	0	1057
		Below 50	495		
	Radiotherapy	Yes	47	10	1057
		No	1000		
	BRCA stage distribution	Stage 1	183	12	1057
		Stage 2	595		
		Stage 3	247		
		Stage 4	20		
6	Tumor stage distribution	T1	400	30	1057
		T2	365		
		T3	200		
		T4	54		
	Nodal metastasis distribution	N0	300	157	1057
		N1	200		
		N2	200		
		N3	200		

2.7 Statistical analysis

R software was used to perform all statistical analyses. All tests were two-sided; $p < 0.05$ was considered to be statistically significant. The Benjamini-Hochberg approach was used to compensate for multiple testing in order to reduce the rate of false discovery.

3. Results

3.1 *In vitro* expression analysis of CRGs expression

In this study, we conducted RNA-seq analysis of two cell lines; the first cell line was a BRCA cell line (BT 20 cells); the other was a normal control cell line (HMEC cells). Our objective was to validate the expression levels of CRGs in these two cell lines. For this validation, we utilized FPKM, a widely used expression quantitative value that is commonly used in gene expression analysis. As shown in Fig. 1, RNA-seq results revealed that CRGs were expressed in both the normal (HMEC) and BRCA (BT 20) cell lines. Notably, the FPKM values of CRGs were significantly (p-value < 0.05) higher in the BRCA cell line (BT 20) when compared to the normal cell line (HMEC).

3.2 Genetic alteration in CRGs

Next, we analyzed the genetic alterations of CRGs across the TCGA-BRCA dataset using cBioPortal, including amplification, deletion, mutation and gene fusion. Analysis revealed that all CRGs (FDX1, LIAS, LIPT1, DLD, DLAT, PDHA1, PDHB, MTF1, GLS and CDKN2A) exhibited genetic alterations in a relatively small proportion of the analyzed BRCA samples. Specifically, these alterations were observed in 0.3%, 1.1%, 0.4%, 1.2%, 0.2%, 0.9%, 0.3%, 1.1%, 0.9% and 5% proportions of the BRCA samples, respectively (Fig. 2). These findings highlight the diversity of genetic alterations in CRGs within the TCGA-BRCA dataset and provide insights into their potential relevance with regards to the development and progression of BRCA.

3.3 Development and performance of the RSF model

In terms of forecasting the overall survival of patients with BRCA, we found that the RSF model based on CRGs performed well. Individuals with high RSF scores had a substantially lower overall survival than individuals with low RSF values, as illustrated in Fig. 3A (log-rank test, $p < 0.0001$; Kaplan-Meier curve). Furthermore, when compared to individuals with low RSF values, individuals with high RSF values had significant (p-value = 0.05) worse rates of disease-specific survival, progress-free survival and relapse-free survival (**Supplementary material**). The overall survival calibration curves for 1, 3 and 5 years showed good agreement with the expected and actual survival probabilities (Fig. 3B). In addition, at 1, 3 & 5 years, the areas under the ROC curves (AUC) for predicting overall survival were 0.978, 0.985 and 0.991, respectively (Fig. 3C). Finally, the decision curves indicated that the RSF model had higher net benefits than both the treat-all-patients strategy and the treat-none strategy across a wide range of threshold probabilities for 1-, 3- and 5-year overall survival prediction (Fig. 3D–F). Our findings suggest that the RSF model could be used as a prognostic tool to stratify patients based on their overall survival and guide personalized treatment strategies accordingly.

3.4 Relationships between RSF scores and clinical features

Next, we investigated the association between RSF scores and various clinical traits in individuals with BRCA. As depicted in Fig. 4, we observed notable age-related differences in the RSF scores of BRCA patients, with those over 50 years-of-age exhibiting lower RSF scores when compared to those under 50 years-of-age (Fig. 4A). Furthermore, RSF scores were higher in patients who underwent radiation therapy when compared to those who did not (Fig. 4B).

In addition, we investigated the RSF scores with respect to the cancer stage, T stage and N stage. Notably, individuals with advanced cancer stages, higher T stages or positive N stages, exhibited significantly lower RSF scores when compared to those with early stages of cancer, lower T stages or negative N stages (Fig. 4C–E). These findings suggest that the RSF model has the potential to stratify patients based on their prognosis when utilizing these clinical characteristics as important predictive factors. The utilization of RSF scores in combination with these clinical traits could facilitate individualized patient management and treatment decision-making in the context of BRCA.

3.5 Functional enrichment analysis

Next, we employed GSEA to investigate the enrichment of signature gene sets, GO gene sets, and KEGG gene sets in BRCA samples with high RSF scores. Our aim was to gain a deeper understanding of the molecular processes underlying the RSF model and its association with CRGs.

As illustrated in Fig. 5, the CRGs were associated with various GO terms in BRCA, including cytoplasmic translation, the unwinding of DNA in DNA replication, and the attachment of spindle microtubules to the kinetochore. Furthermore, Fig. 6 shows that CRGs were associated with several KEGG pathways in BRCA, including ribosomes, DNA replication, RNA polymerase and cell cycle pathways.

These findings shed light on the biological mechanisms implicated in BRCA and provide potential therapeutic targets for personalized treatment strategies. By uncovering the enriched gene sets and pathways in BRCA patients with high RSF scores, we gained valuable insights that may aid in the development of targeted therapies and approaches for precision medicine.

3.6 Relationships between CRGs, immune cell proportions and drug sensitivity

Next, we investigated the association between CRGs and immune cell proportions using the CIBERSORT algorithm. Fig. 7 shows that CRGs exhibited a positive correlation with the percentages of naive B cells, resting memory CD4 T

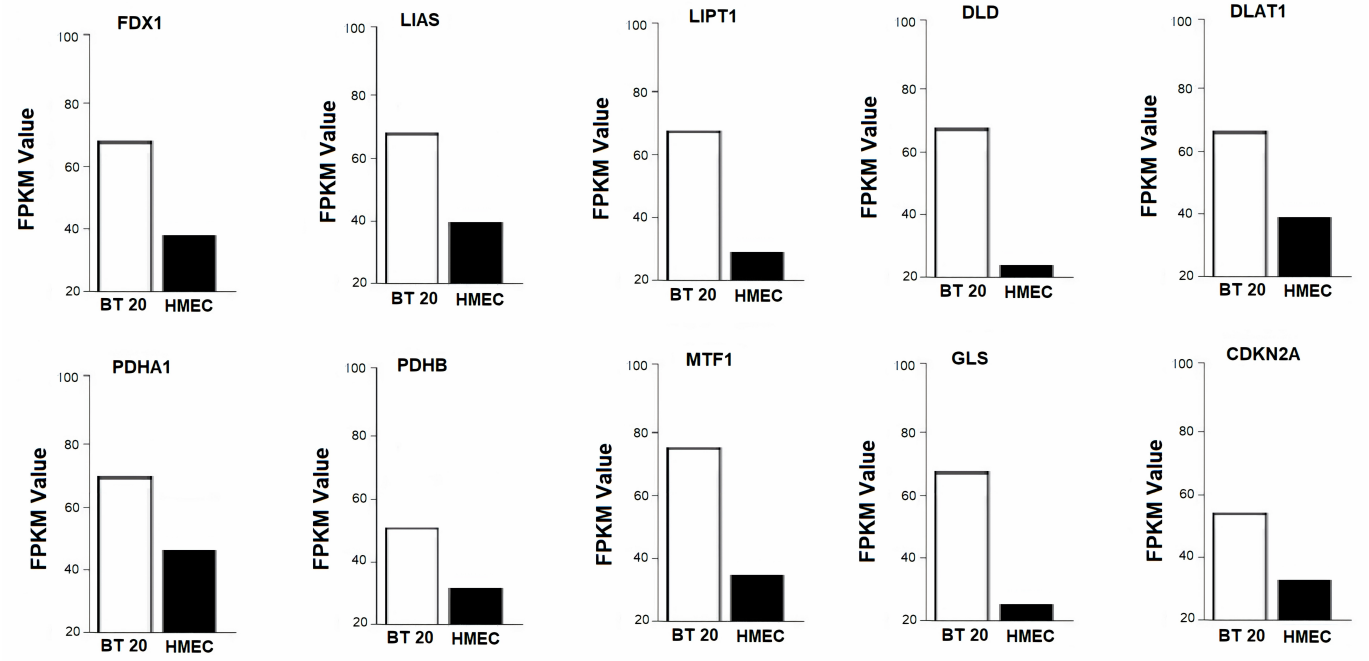

FIGURE 1. Expression analysis of CRGs using BT 20 and HMEC cell lines *via* RNA-seq analysis. CRGs: Cuproptosis-Related Genes; RNA-seq: RNA sequencing.

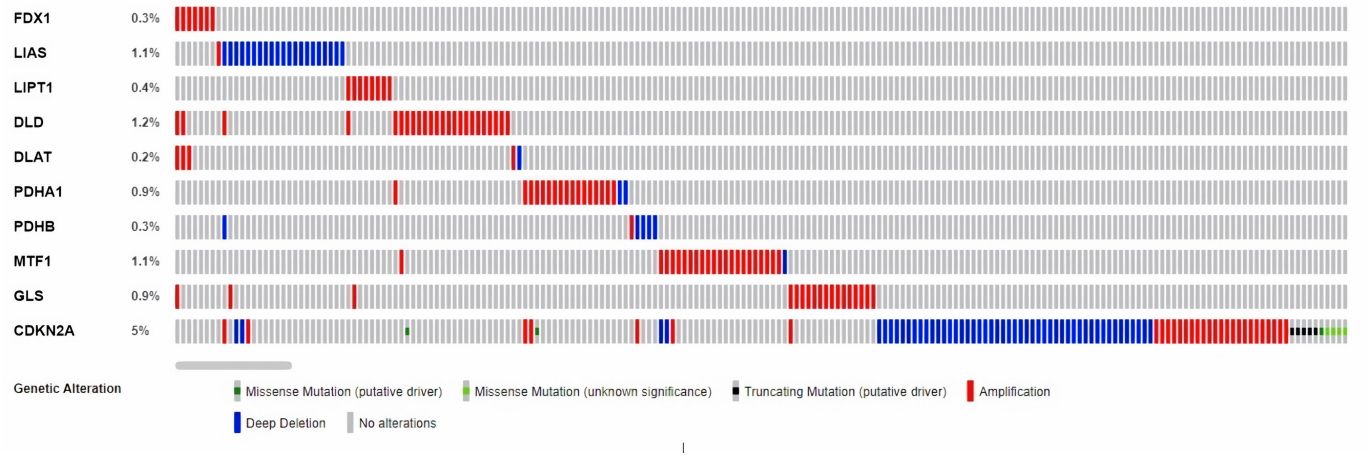

FIGURE 2. The analysis of genetic alterations of CRGs in BRCA samples *via* cBioPortal software.

cells, activated dendritic cells, and follicular helper T cells. Conversely, we identified negative correlations with the percentages of M2 macrophages, neutrophils and naive CD4 T cells.

These findings suggest that the RSF model may have the potential to predict the immunological state of BRCA individuals based on the interplay between CRGs and immune cell proportions. By identifying the associations between CRGs and specific immune cell types, the RSF model can offer valuable information that could be utilized to tailor treatment plans for individual patients with BRCA.

In the final phase of our study, we investigated the relationship between RSF scores and drug sensitivity using the IC_{50} values obtained from the GDSC2 database. Fig. 8 shows that higher RSF scores were associated with increased resistance to VX-11e_2096 and ERK_6604_1714, indicating reduced sensitivity to these drugs. Conversely, higher RSF scores were associated with greater sensitivity to Acetalax_1804, WEHI-539_1997 and AZD5991_1720, thus suggesting an enhanced responsiveness to these medications.

These results imply that the RSF model holds promise as a predictive tool for drug sensitivity in patients with BRCA. By assessing the correlation between RSF scores and drug responses, the RSF model can potentially inform personalized treatment plans tailored to the unique drug sensitivities of individual patients with BRCA.

4. Discussion

BRCA is the most common cancer in women and accounts for the majority of deaths due to cancer in the female population [42, 43]. In this study, we analyzed the expression levels of CRGs and created an RSF model based on these genes to predict the overall survival of patients with BRCA. The

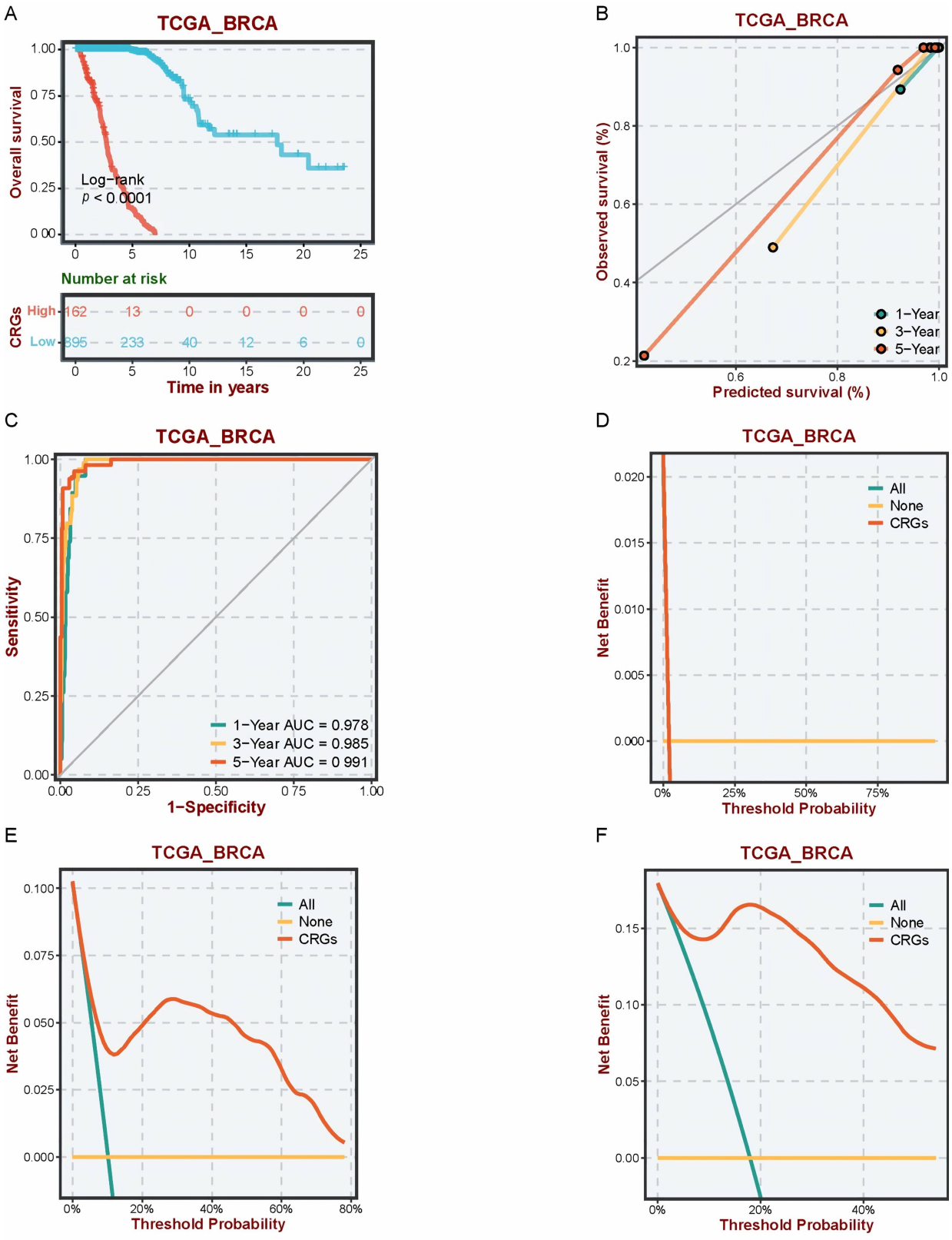

FIGURE 3. The creation of a random survival forest (RSF) model determined by CRGs to forecast overall survival in patients with BRCA. (A) Kaplan-Meier curve plotted showing the overall survival of distinct groups of BRCA patients based on the optimal threshold of their RSF score. (B) Calibration curves for the prognosis model predicting the overall survival of BRCA patients over a period of 1, 3 & 5 years. (C) Time-dependent ROC curves for RSF scores in patients with BRCA. (D) Decision curves for 1-year overall survival prediction in BRCA patients. (E) Decision curves for 3-year overall survival prediction in BRCA patients. (F) Decision curves for 5-year overall survival prediction in BRCA patients. CRGs: Cuproptosis-Related Genes; BRCA: Breast cancer; CRGs: cuproptosis-related genes; AUC: area under the curve; TCGA_BRCA: The Cancer Genome Atlas Breast Cancer Susceptibility Genes.

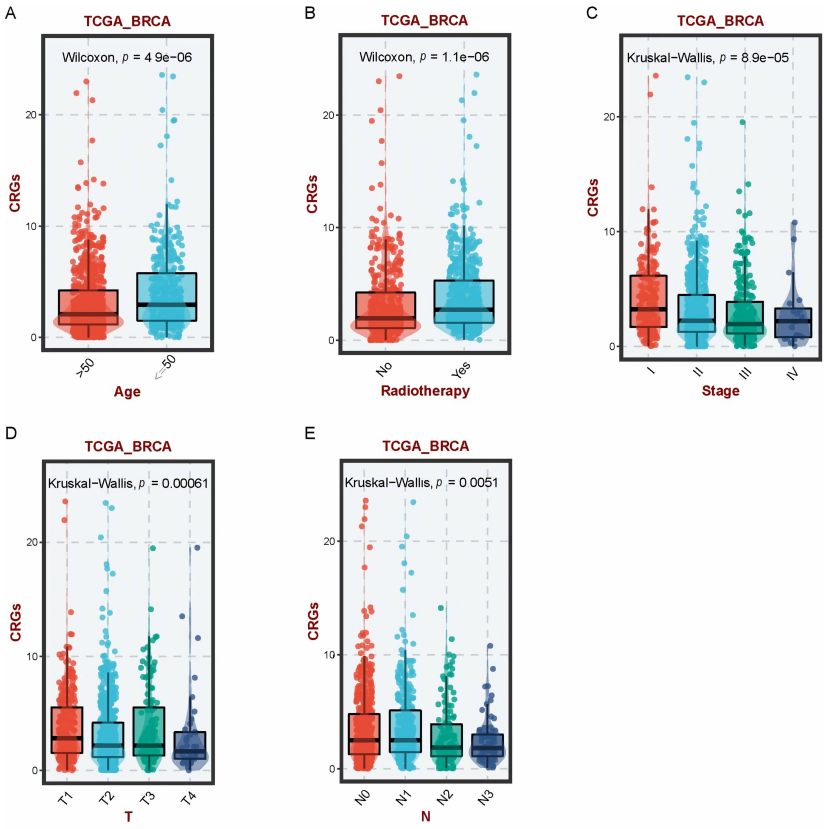

FIGURE 4. RSF scores of CRGs in BRCA patients based on different clinical characteristics. (A) RSF scores of CRGs in BRCA patients at different ages. (B) RSF scores of CRGs in BRCA patients receiving radiotherapy or not. (C) RSF scores of CRGs in BRCA patients with different stages. (D) RSF scores of CRGs in BRCA patients with different T stages. (E) RSF scores of CRGs in BRCA patients with different N stages. CRGs: Cuproptosis-Related Genes; BRCA: Breast cancer; TCGA_BRCA: The Cancer Genome Atlas Breast Cancer Susceptibility Genes.

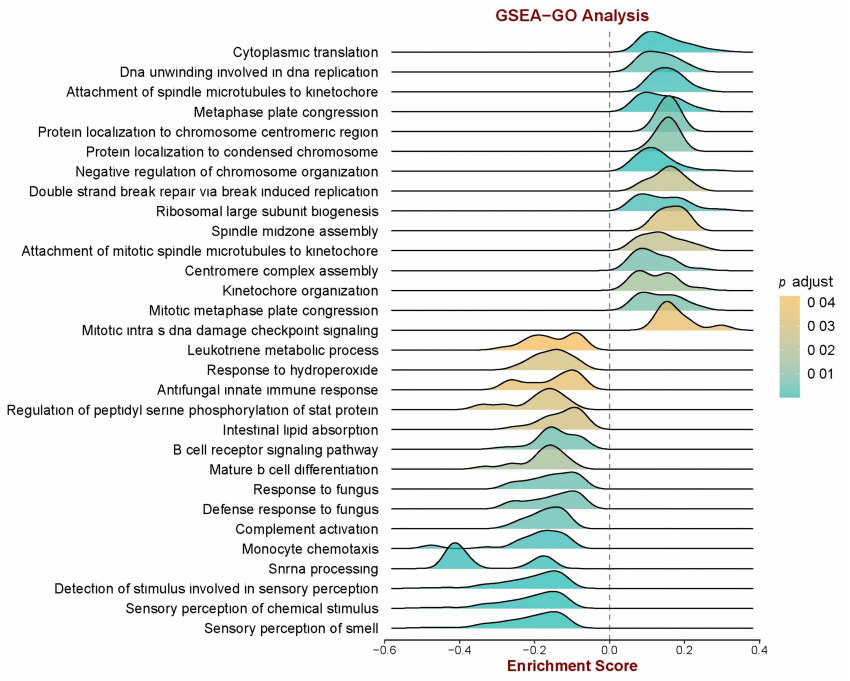

FIGURE 5. GSEA software was used to examine the GO terms of CGRs in BRCA samples with high RSF scores. GSEA: Gene Set Enrichment Analysis; GO: Gene Ontology; CRGs: Cuproptosis-Related Genes; BRCA: Breast cancer; RSF: Random Survival Forest.

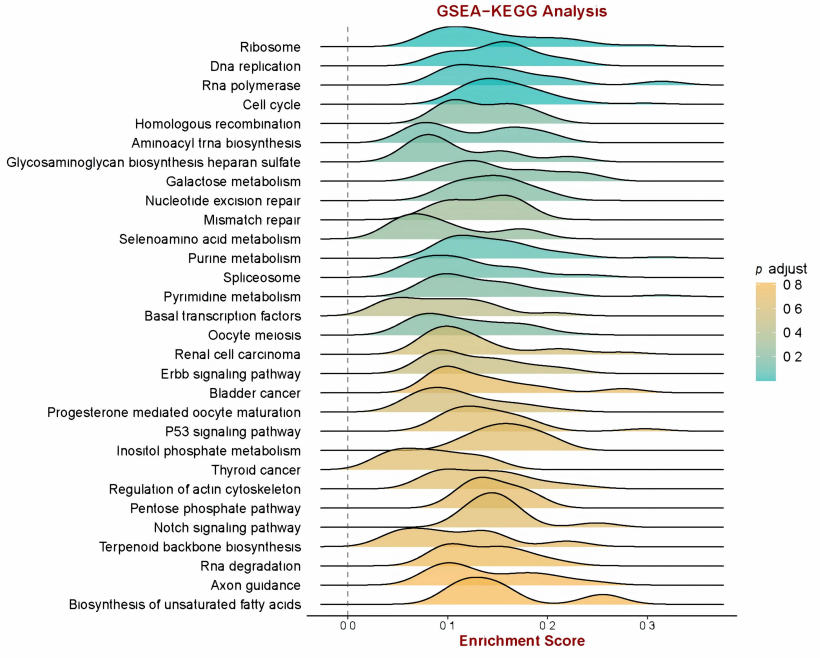

FIGURE 6. The analysis of CGRs with high RSF scores in relation to KEGG terms in patients with BRCA. KEGG: Kyoto Encyclopedia of Genes and Genomes; CRGs: Cuproptosis-Related Genes; BRCA: Breast cancer; RSF: Random Survival Forest; GSEA: Gene Set Enrichment Analysis.

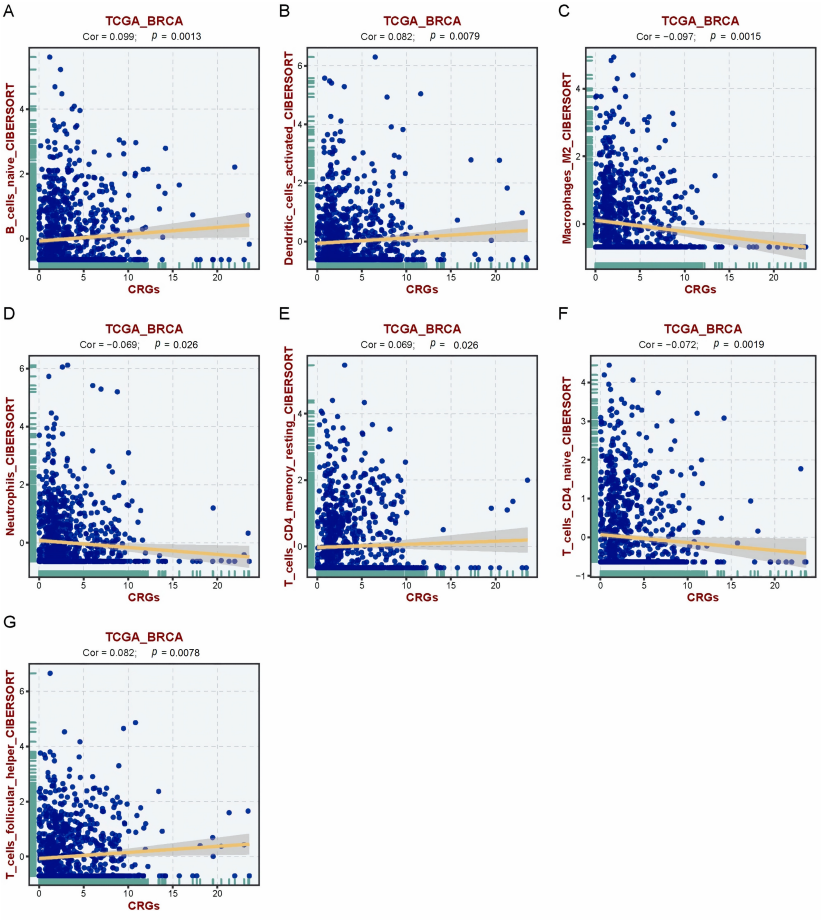

FIGURE 7. The correlation of CRGs with immune cell proportions, as characterized by the CIBERSORT algorithm. (A) naïve B cells, (B) activated dendritic cells, (C) M2 macrophages, (D) neutrophils, (E) resting memory CD4 T cells, (F) naive CD4 T cells, and (G) follicular helper T cells. CRGs: Cuproptosis-Related Genes; TCGA_BRCA: The Cancer Genome Atlas Breast Cancer Susceptibility Genes.

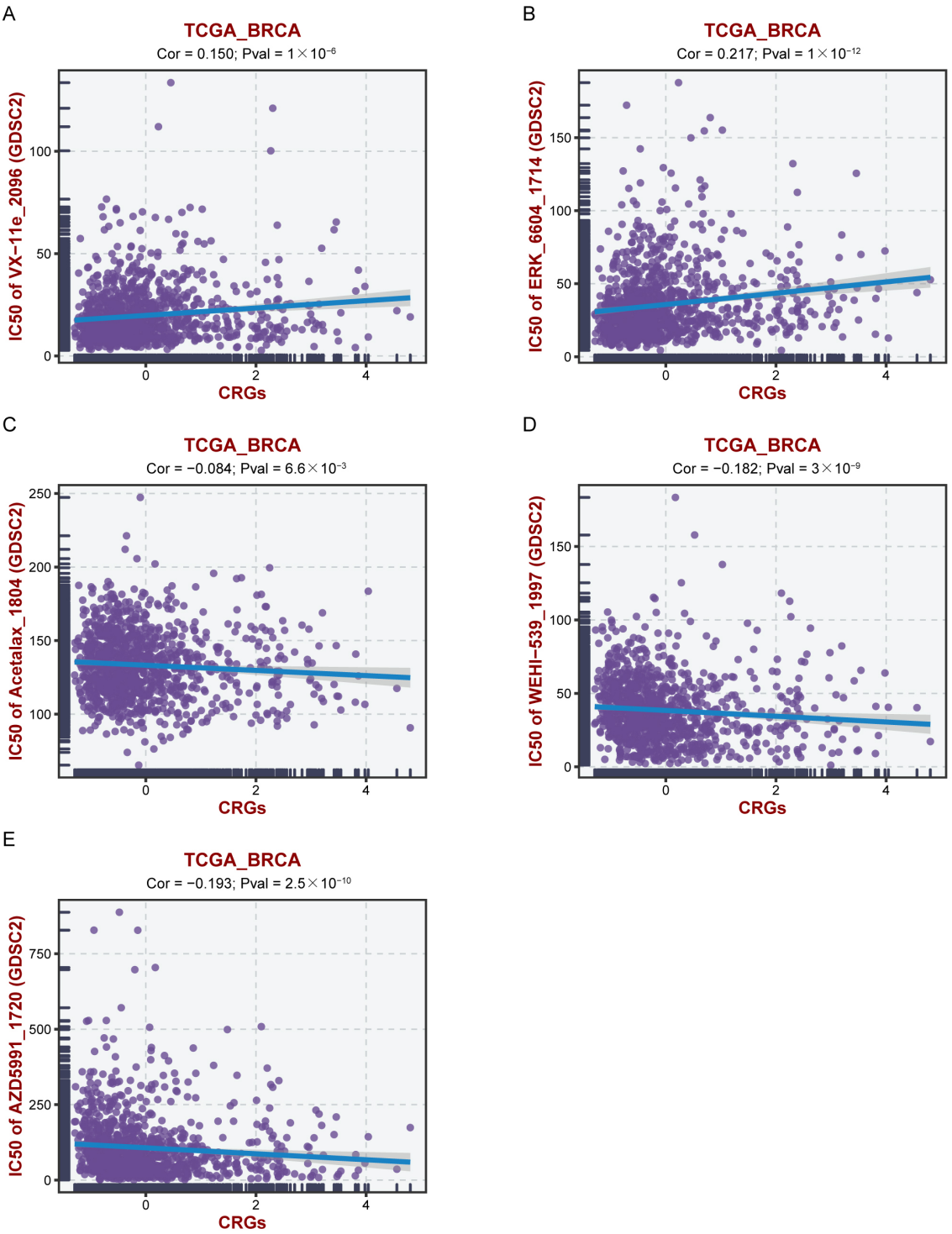

FIGURE 8. The correlation of CRGs with the IC_{50} of various drugs from the GDSC2 database. (A) VX-11e_2096, (B) ERK_6604_1714, (C), Acetalax_1804, (D) WEHI-539_1997, and (E) AZD5991_1720. CRGs: Cuproptosis-Related Genes; GDSC2: Kyoto Encyclopedia of Genes and Genomes; CRGs: cuproptosis-related genes; TCGA_BRCA: The Cancer Genome Atlas Breast Cancer Susceptibility Genes.

CRGs were significantly (p-value < 0.05) up-regulated in the BRCA cell line when compared to the normal control cell line. Moreover, the newly developed RSF model for BRCA patients performed well in terms of predicting overall survival and identified differences based on age, radiotherapy status, stage, T stage and N stage. We also found that the RSF scores were associated with various clinical characteristics, functional enrichment, immune cell proportions and drug sensitivity.

Similar studies have been conducted in the past that focused on identifying biomarkers or patient prognostic prediction using gene signatures, including BRCA patients. For instance, Wu *et al.* [44] used lncRNA signature profiling to create a molecular predictor of survival for patients with colon cancer. In patients with glioma, a separate investigation found that an endoplasmic reticulum stress-related signature was able to predict both immunological characteristics and prognosis [45].

Based on gene expression data from both experiments, these researchers were able to create models that could precisely predict patient outcomes.

Multiple research investigations have employed gene expression profiling to forecast outcomes for patients with BRCA. For instance, Kalafi *et al*. [46] utilized clinical data with deep learning and machine learning methods for predicting the survival of patients with BRCA. Similarly, Montazeri *et al*. [47] used naive bayes, random forest trees, the one-nearest neighbour method, adaboost, support vector machines, Radial basis function (RBF) networks, and multilayer perceptrons to predict the survival of BRCA patients. In another study, Pan *et al*. [48] proposed a predictive model for patients with colorectal cancer using mRNA gene expression data. This model utilized differentially expressed genes (DEGs) profiles obtained from univariate and multivariate Cox regression analyses and compared later Tumor, Node, Metastasis (TNM) stages to investigate their predictive survival accuracy. Analysis revealed that 10 differentially expressed genes (DEGs) exerted significant impact on the survival of patients with colorectal cancer. Similarly, Yan *et al*. [49] employed random forests to identify biomarkers associated with the survival of patients with colorectal cancer based on a set of oligonucleotide microarray data. These findings highlighted four genes with the potential to predict the survival of patients with colorectal cancer. A recent study conducted by Mustafa *et al*. [50] introduced a novel prognostic model for BRCA patients, termed the ensemble model for BRCA survivability prediction (EBCSP). By deploying independent neural networks, the EBCSP model demonstrated superior performance when compared to existing benchmarks such as the multimodal DNN by integrating multi-dimensional data (MDNNMD) model. Another study, by Othman *et al*. [51], introduced a hybrid deep learning model, combining Convolutional neural network-long short-term memory (CNN-LSTM) and Convolutional neural network-gated recurrent unit (CNN-GRU) architectures, to predict the survival of patients with BRCA. The hybrid Deep Learning (DL) model produced promising results, highlighting its potential for improved survival prediction in patients with BRCA. The current investigation, in contrast to earlier studies, focused on CRGs and their relationship to overall survival in patients with BRCA. The regulation of CRGs has been shown to modulate tumor progression, as these genes have the potential to induce death in cancer cells by disrupting copper homeostasis [52].

Many research investigations have demonstrated the predictive importance of tumor-infiltrating lymphocytes (TILs) in patients with BRCA in terms of immune cell proportions. For instance, individuals with human epidermal growth factor receptor 2 (HER2)-positive BRCA, who have greater levels of TILs, were shown to have better survival rates [53]. In addition, the presence of specific subtypes of TILs, such as CD8+ T cells, is known to be a favorable prognostic factor in patients with triple-negative BRCA [53].

In contrast to existing predictive models that were built using demographic data and routine clinical examination indicators, our newly developed model utilizes molecular biomarkers extracted from gene expression analysis. This approach enables high-accuracy predictions without the need for additional experiments. Moreover, our model holds the promise of aiding clinicians to identify and treat high-risk patients in the early stages of disease. By providing more accurate, targeted, and individualized treatment plans, physicians could enhance the prognosis of patients with BRCA, ultimately improving patient outcomes.

Our study does, nevertheless, have certain limitations that need to be considered. In order to establish the validity of our findings, further validation in separate cohorts is required because all the data utilized in the present research was sourced from public sources. Furthermore, although we identified several biological pathways and processes associated with RSF scores, further research is needed to validate these relationships and fully clarify the basic processes involved. Finally, although our RSF model showed good performance in predicting overall survival, it may not apply to other types of cancer or populations; therefore, further research is required to fully determine its potential therapeutic value.

The findings of this study were used to construct and assess an RSF model to estimate the overall survival of patients with BRCA. The RSF model demonstrated good performance in predicting overall survival, and further analysis revealed that RSF scores were associated with various clinical characteristics, functional enrichment, immune cell proportions, and drug sensitivity. These findings have significant ramifications for individualized treatment plans and further research on the biological mechanisms underlying BRCA.

5. Conclusions

Our new RSF model based on CRGs is an effective tool for predicting overall survival in patients with BRCA. RSF scores were associated with various clinical characteristics, immune cell proportions and drug sensitivity. The results of this study have significant importance for both future research on the molecular mechanisms driving BRCA and personalized plans for therapy.

AVAILABILITY OF DATA AND MATERIALS

The data could be obtained by contacting the corresponding author.

AUTHOR CONTRIBUTIONS

FFZ, DFX—designed the research study. FFZ, DFX—performed the research. FFZ, DFX—analyzed the data. FFZ—wrote the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

ACKNOWLEDGMENT

Thank you to all authors and reviewers of the article, as well as all those who have contributed to the research.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary material associated with this article can be found, in the online version, at https://oss.ejgo.net/files/article/1912042245434818560/attachment/Supplementary%20material.docx.

REFERENCES

[1] Wilkinson L, Gathani T. Understanding breast cancer as a global health concern. The British Journal of Radiology. 2022; 95: 20211033.

[2] Usman M, Hameed Y, Ahmad M, Iqbal MJ, Maryam A, Mazhar A, *et al*. SHMT2 is associated with tumor purity, CD8+ T immune cells infiltration, and a novel therapeutic target in four different human cancers. Current Molecular Medicine. 2023; 23: 161–176.

[3] Mehrotra R, Yadav K. Breast cancer in India: present scenario and the challenges ahead. World Journal of Clinical Oncology. 2022; 13: 209–218.

[4] Trayes KP, Cokenakes SEH. Breast cancer treatment. American Family Physician. 2021; 104: 171–178.

[5] Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. BMC Cancer. 2019; 19: 230.

[6] Jafari SH, Saadatpour Z, Salmaninejad A, Momeni F, Mokhtari M, Nahand JS, *et al*. Breast cancer diagnosis: imaging techniques and biochemical markers. Journal of Cellular Physiology. 2018; 233: 5200–5213.

[7] Fahad Ullah M. Breast cancer: current perspectives on the disease status. Advances in Experimental Medicine and Biology. 2019; 1152: 51–64.

[8] Ahmad M, Hameed Y, Khan M, Usman M, Rehman A, Abid U, *et al*. Up-regulation of GINS1 highlighted a good diagnostic and prognostic potential of survival in three different subtypes of human cancer. Brazilian Journal of Biology. 2021; 84: e250575.

[9] Sial N, Saeed S, Ahmad M, Hameed Y, Rehman A, Abbas M, *et al*. Multi-omics analysis identified TMED2 as a shared potential biomarker in six subtypes of human cancer. International Journal of General Medicine. 2021; 14: 7025–7042.

[10] Zhu X, Tang L, Mao J, Hameed Y, Zhang J, Li N, *et al*. Decoding the mechanism behind the pathogenesis of the focal segmental glomerulosclerosis. Computational and Mathematical Methods in Medicine. 2022; 2022: 1941038.

[11] Gradishar WJ, Anderson BO, Abraham J, Aft R, Agnese D, Allison KH, *et al*. Breast cancer, version 3.2020, NCCN clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network. 2020; 18: 452–478.

[12] Britt KL, Cuzick J, Phillips KA. Key steps for effective breast cancer prevention. Nature Reviews Cancer. 2020; 20: 417–436.

[13] Usman M, Okla MK, Asif HM, AbdElgayed G, Muccee F, Ghazanfar S, *et al*. A pan-cancer analysis of GINS complex subunit 4 to identify its potential role as a biomarker in multiple human cancers. American Journal of Cancer Research. 2022; 12: 986–1008.

[14] Cocco S, Piezzo M, Calabrese A, Cianniello D, Caputo R, Lauro VD, *et al*. Biomarkers in triple-negative breast cancer: state-of-the-art and future perspectives. International Journal of Molecular Sciences. 2020; 21: 4579.

[15] Hameed Y, Usman M, Ahmad M. Does mouse mammary tumor-like virus cause human breast cancer? Applying Bradford Hill criteria postulates. Bulletin of the National Research Centre. 2020; 44: 1–13.

[16] Tang D, Chen X, Kroemer G. Cuproptosis: a copper-triggered modality of mitochondrial cell death. Cell Research. 2022; 32: 417–418.

[17] Liu J, Liu Y, Wang Y, Kang R, Tang D. HMGB1 is a mediator of cuproptosis-related sterile inflammation. Frontiers in Cell and Developmental Biology 2022; 10: 996307.

[18] Xie J, Yang Y, Gao Y, He J. Cuproptosis: mechanisms and links with cancers. Molecular Cancer. 2023; 22: 46.

[19] Hu Q, Wang R, Ma H, Zhang Z, Xue Q. Cuproptosis predicts the risk and clinical outcomes of lung adenocarcinoma. Frontiers in Oncology. 2022; 12: 922332.

[20] Zhang B, Xie L, Liu J, Liu A, He M. Construction and validation of a cuproptosis-related prognostic model for glioblastoma. Frontiers in Immunology. 2023; 14: 1082974.

[21] Huang Y, Yin D, Wu L. Identification of cuproptosis-related subtypes and development of a prognostic signature in colorectal cancer. Scientific Reports. 2022; 12: 17348.

[22] Tong Y, Cui Y, Jiang L, Pi Y, Gong Y, Zhao D. Clinical characteristics, prognostic factor and a novel dynamic prediction model for overall survival of elderly patients with chondrosarcoma: a population-based study. Frontiers in Public Health. 2022; 10: 901680.

[23] Latha NR, Rajan A, Nadhan R, Achyutuni S, Sengodan SK, Hemalatha SK, *et al*. Gene expression signatures: a tool for analysis of breast cancer prognosis and therapy. Critical Reviews in Oncology/Hematology. 2020; 151: 102964.

[24] Usman M, Hameed Y, Ahmad M, Jalil Ur Rehman, Ahmed H, Hussain MS, *et al*. Breast cancer risk and human papillomavirus infection: a Bradford Hill criteria based evaluation. Infectious Disorders-Drug Targets. 2022; 22: e200122200389.

[25] Tagliafico AS, Piana M, Schenone D, Lai R, Massone AM, Houssami N. Overview of radiomics in breast cancer diagnosis and prognostication.The Breast. 2020; 49: 74–80.

[26] Li H, Liu RB, Long CM, Teng Y, Cheng L, Liu Y. Development and validation of a new multiparametric random survival forest predictive model for breast cancer recurrence with a potential benefit to individual outcomes. Cancer Management and Research. 2022; 14: 909–923.

[27] Hameed Y, Usman M, Liang S, Ejaz S. Novel diagnostic and prognostic biomarkers of colorectal cancer: capable to overcome the heterogeneity-specific barrier and valid for global applications. PLOS ONE. 2021; 16: e0256020.

[28] Mao J, Huang X, Okla MK, Abdel-Maksoud MA, Mubarak A, Hameed Z, *et al*. Risk factors for TERT promoter mutations with papillary thyroid carcinoma patients: a meta-analysis and systematic review. Computational and Mathematical Methods in Medicine. 2022; 2022: 1721526.

[29] Iswaran H, Kogalur U, Blackstone E, Lauer M. Random survival forest. The Annals of Applied Statistics. 2008; 2: 841–860.

[30] Lee S, Lim H. Review of statistical methods for survival analysis using genomic data. Genomics & Informatics. 2019; 17: e41.

[31] Nasejje JB, Mwambi H. Application of random survival forests in understanding the determinants of under-five child mortality in Uganda in the presence of covariates that satisfy the proportional and non-proportional hazards assumption. BMC Research Notes. 2017; 10: 459.

[32] Adham D, Abbasgholizadeh N, Abazari M. Prognostic factors for survival in patients with gastric cancer using a random survival forest. Asian Pacific Journal of Cancer Prevention. 2017; 18: 129–134.

[33] Bou-Hamad I, Larocque D, Ben-Ameur H. A review of survival trees. Statistics Surveys. 2011; 5: 44–71.

[34] Tavera-Montañez C, Hainer SJ, Cangussu D, Gordon SJV, Xiao Y, Reyes-Gutierrez P, *et al*. The classic metal-sensing transcription factor MTF1 promotes myogenesis in response to copper. The FASEB Journal. 2019; 33: 14556–14574.

[35] Tong D, Schneeberger C, Leodolter S, Zeillinger R. Quantitative determination of gene expression by competitive reverse transcription-polymerase chain reaction in degraded rna samples. Analytical Biochemistry. 1997; 251: 173–177.

[36] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, *et*

al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discovery. 2012; 2: 401–404.

[37] Tomczak K, Czerwińska P, Wiznerowicz M. The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia. 2015; 19: A68–A77.

[38] Kerseviciute I, Gordevicius J. aPEAR: an R package for autonomous visualisation of pathway enrichment networks. Bioinformatics. 2023; btad672.

[39] Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods in Molecular Biology. 2018; 1711: 243–259.

[40] Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLOS ONE. 2014; 9: e107468.

[41] Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, *et al*. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Research. 2012; 41: D955–D961.

[42] Jiaxin M, Zaihua W, Jianting S, Ping W, Jing D, Jing W, *et al*. Analysis on mortality and trend of breast cancer in women in Beijing, 2010–2020. Disease Surveillance. 2022; 37: 674–678.

[43] Yuan P, Kang Y, Ma F, Fan Y, Wang J, Wang X, *et al*. Effect of epirubicin plus paclitaxel *vs* epirubicin and cyclophosphamide followed by paclitaxel on disease-free survival among patients with operable ERBB2-negative and lymph node-positive breast cancer: a randomized clinical trial. JAMA Network Open. 2023; 6: e230122.

[44] Wu Z, Lu Z, Li L, Ma M, Long F, Wu R, *et al*. Identification and validation of ferroptosis-related LncRNA signatures as a novel prognostic model for colon cancer. Frontiers in Immunology. 2022; 12: 783362.

[45] Zhang Q, Guan G, Cheng P, Cheng W, Yang L, Wu A. Characterization of an endoplasmic reticulum stress-related signature to evaluate immune features and predict prognosis in glioma. Journal of Cellular and Molecular Medicine. 2021; 25: 3870–3884.

[46] Kalafi EY, Nor NAM, Taib NA, Ganggayah MD, Town C, Dhillon SK. Machine learning and deep learning approaches in breast cancer survival prediction using clinical data. Folia Biologica. 2019; 65: 212–220.

[47] Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technology and Health Care. 2016; 24: 31–42.

[48] Pan F, Chen T, Sun X, Li K, Jiang X, Försti A, *et al*. Prognosis prediction of colorectal cancer using gene expression profiles. Frontiers in Oncology. 2019; 9: 252.

[49] Yan Z, Li J, Xiong Y, Xu W, Zheng G. Identification of candidate colon cancer biomarkers by applying a random forest approach on microarray data. Oncology Reports. 2012; 28: 1036–1042.

[50] Mustafa E, Jadoon EK, Khaliq-uz-Zaman S, Humayun MA, Maray M. An ensembled framework for human breast cancer survivability prediction using deep learning. Diagnostics. 2023; 13: 1688.

[51] Othman NA, Abdel-Fattah MA, Ali AT. A hybrid deep learning framework with decision-level fusion for breast cancer survival prediction. Big Data and Cognitive Computing. 2023; 7: 50.

[52] Chen L, Min J, Wang F. Copper homeostasis and cuproptosis in health and disease. Signal Transduction and Targeted Therapy. 2022; 7: 378.

[53] Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, *et al*. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. The Lancet Oncology. 2018; 19: 40–50.

How to cite this article: Fangfang Zhu, Dafang Xu. Development of a random survival forest model based on cuproptosis-related genes (CRGs) for predicting overall survival in patients with invasive breast carcinoma. European Journal of Gynaecological Oncology. 2025; 46(4): 46-57. doi: 10.22514/ejgo.2025.050.