Mathematical Biosciences and Engineering

MBE, 21(1): 1590–1609.
DOI: 10.3934/mbe.2024069
Received: 01 November 2023
Revised: 06 December 2023
Accepted: 13 December 2023
Published: 02 January 2024

http://www.aimspress.com/journal/MBE

Research article

Anoikis-related mRNA-lncRNA and DNA methylation profiles for overall survival prediction in breast cancer patients

Huili Yang, Wangren Qiu* and Zi Liu*

Computer Department, Jingdezhen Ceramic University, Jingdezhen 333403, China

* **Correspondence:** Email: qiuone@163.com, liuzi@jcu.edu.cn.

Abstract: As a type of programmed cell death, anoikis resistance plays an essential role in tumor metastasis, allowing cancer cells to survive in the systemic circulation and as a key pathway for regulating critical biological processes. We conducted an exploratory analysis to improve risk stratification and optimize adjuvant treatment choices for patients with breast cancer, and identify multigene features in mRNA and lncRNA transcriptome profiles associated with anoikis. First, the variance selection method filters low information content genes in RNA sequence and then extracts the mRNA and lncRNA expression data base on annotation files. Then, the top ten key mRNAs are screened out through the PPI network. Pearson analysis has been employed to identify lncRNAs related to anoikis, and the prognosis-related lncRNAs are selected using Univariate Cox regression and machine learning. Finally, we identified a group of RNAs (including ten mRNAs and six lncRNAs) and integrated the expression data of 16 genes to construct a risk-scoring system for BRCA prognosis and drug sensitivity analysis. The risk score's validity has been evaluated with the ROC curve, Kaplan-Meier survival curve analysis and decision curve analysis (DCA). For the methylation data, we have obtained 169 anoikis-related prognostic methylation sites, integrated these sites with 16 RNA features and further used the deep learning model to evaluate and predict the survival risk of patients. The developed anoikis feature is demonstrated a consistency index (C-index) of 0.778, indicating its potential to predict the survival probability of breast cancer patients using deep learning methods.

Keywords: anoikis; prognosis; BRCA; deep learning; survival analysis

1. Introduction

Breast cancer is the most prevalent type of cancer that spreads to lymph nodes and distant organs, posing significant challenges in its treatment due to its metastatic nature [1]. Based on the most recent survey conducted by the International Agency for Research on Cancer (IARC) in 2018, breast cancer emerges as the predominant neoplasm affecting women globally, constituting 24.2% of the total reported instances, as indicated by the latest data. The frequency of breast cancer exhibits a notable surge during the reproductive years, displaying a pronounced acceleration with advancing age. However, this upward trend adopts a more gradual pace after reaching the approximate threshold of 50 years, coinciding with the average age at which menopause typically occurs [2]. Despite substantial progress in medical management, breast cancer remains a formidable disease that endangers human lives.

Anoikis refers to a distinct form of cellular demise instigated by the disruption of cellular interaction with the neighboring extracellular matrices (ECM) [3–5]. Anoikis functions as a protective mechanism that counteracts the onset of abnormal cellular growth or adhesion onto deviant extracellular matrices, serving to prevent pathological disturbances. Thus, anoikis assumes a critical role in the regulation and preservation of the cellular state, rendering it indispensable for essential processes such as body development, tissue homeostasis, disease onset and the intricate process of tumor metastasis [6]. The ability to resist anoikis is a major factor in cancer metastasis, as it allows tumor cells to spread through the circulatory system to other distant organs [7]. Moreover, tumor cells exhibit diverse mechanisms to counteract anoikis [8]. Hence, anoikis resistance assumes paramount importance in driving tumor advancement and metastasis. Notably, within the context of breast cancer, the development of anoikis resistance represents a malignant phenotype associated with cancer cell dissemination to distant sites [9].

Novel prognostic indicators hold promise in discerning patients at elevated risk, facilitating precise identification. By leveraging gene expression profiles of primary breast cancers, it becomes feasible to pinpoint individuals who are most prone to developing metastatic disease, consequently fostering the exploration of novel therapeutic targets [10]. Notably, variations in the expression levels of genes situated downstream of the anoikis pathway profoundly influence the proliferative and metastatic capacity of breast cancer cells. Some studies have indicated that the expression levels of protein kinases, such as Akt and FAK, can affect the anoikis status of breast cancer cells [11]. The proliferative and metastatic behavior of breast cancer cells is subject to regulation by specific mRNAs associated with the anoikis pathway. Therefore, studying these proteins shows promise for treating early breast cancer. Furthermore, evidence suggests that dysregulated expression of Long non-coding RNAs (lncRNAs) contributes to the initiation and progression of breast cancer [12]. Nevertheless, the precise association between distinct lncRNAs and the phenomenon of anoikis resistance, as well as the process of distant metastasis in cancer cells, remains elusive.

In summary, identifying anoikis-related genes is of great significance for deciphering the underlying mechanism of breast cancer and exploring new therapeutic targets. In this study, we constructed a novel mRNA/lncRNA signature associated with anoikis for breast cancer patients using methods such as the PPI network, Lasso Cox and random survival forest. Subsequently, we evaluated the prognostic value of anoikis-associated features and investigated anoikis-associated breast cancer tumor subtype, immune microenvironment analysis and drug sensitivity prediction. Furthermore, anoikis-related prognostic methylation sites were identified, and anoikis-related methylation sites and mRNA/lncRNA features were integrated as input to the DeepSurv model. The C-index was

utilized to evaluate the value of these features for predicting the probability of patient survival and the model's performance.

2. Materials and methods

In this study, we focus on mine anoikis-related prognostic signatures in BRCA and evaluates the prognostic value of these signatures. The study's schematic representation is depicted in Figure 1.

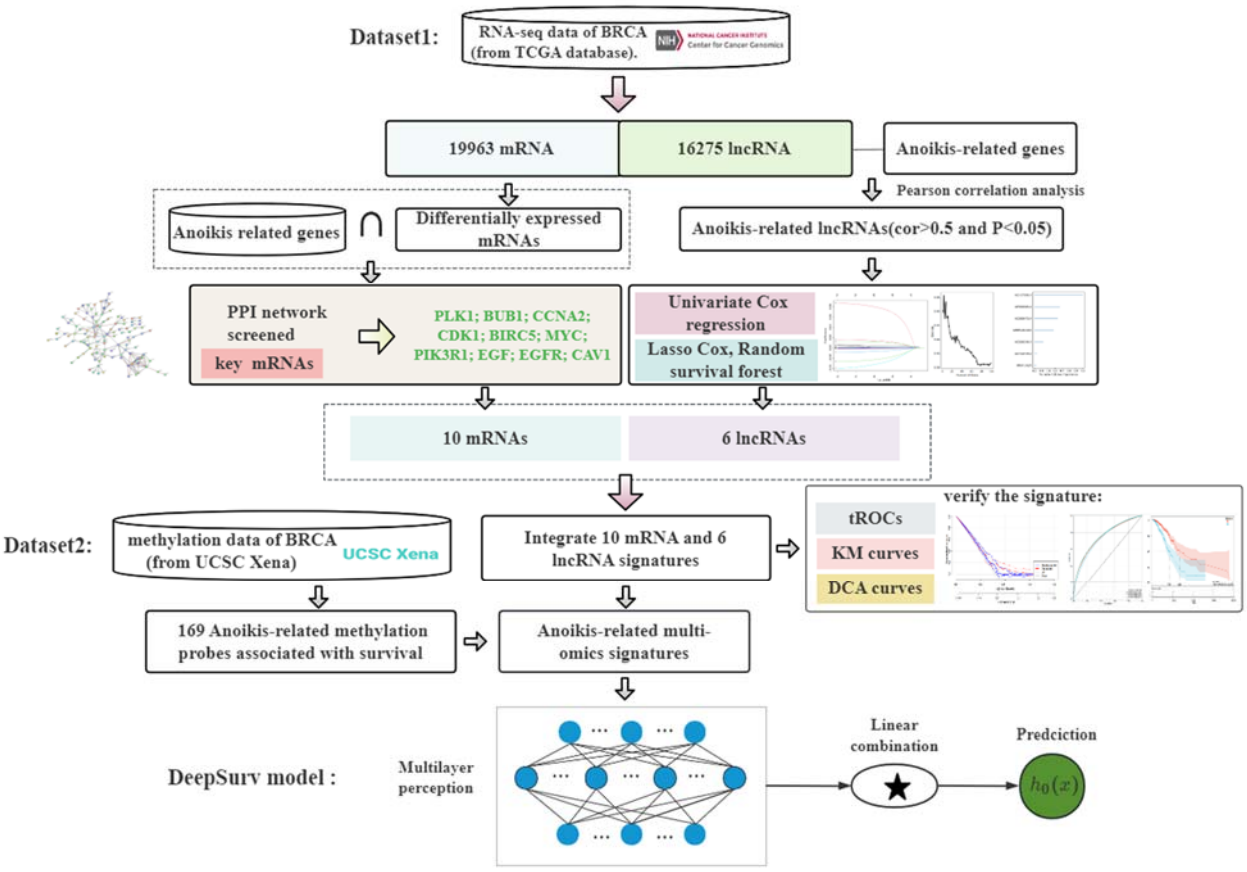

Figure 1. Flowchart of the overall study design.

2.1. Data collection and processing

We obtained transcriptome data (RNA-seq, HTSeq-Counts/FPKMs) from the TCGA database, encompassing 1113 breast cancer patients and 113 normal controls, along with their corresponding clinical information. The Ensemble gene ID was then converted into gene symbols based on the GTF annotation file. Unimportant genes were filtered using the variance selection method, and mRNA and lncRNA gene expressions were extracted based on gene types specified in the annotation file. Methylome data (Illumina Human Methylation 450, HM450) and associated clinical information were obtained from the TCGA database using UCSC Xena (https//xena.ucsc.edu/). The anoikis-related genes obtained from the GeneCards database [13] (https://www.genecards.org/, accessed on 17 October 2022) using the search term "Anoikis.", and 794 anoikis-related genes were obtained.

2.2. Identification of anoikis-related signature

We analyzed the differentially expressed mRNAs in breast cancer tissues and normal tissues using the mRNA expression data. This analysis was performed using linear models and Bayesian statistical methods [14]. We applied a threshold of adjusted P-values $\text{adj.PVal} < 0.05$ and $|\text{logFC}| > 1$ to identify differentially expressed genes. The differentially expressed mRNAs were intersected with aforementioned those above 794 anoikis-related genes, and the differentially expressed anoikis-related mRNAs in the intersection were retained.

The expression data of 794 anoikis genes were obtained from breast cancer gene expression data. We used R software to pair each lncRNA with all the anoikis genes and calculate the Pearson correlation coefficient for each gene expression pair. The correlation coefficient greater than 0.5 and the P value less than 0.05 were used as thresholds to identify lncRNAs associated with anoikis.

2.3. PPI network screening key mRNA

The STRING database (https://string-db.org/) can depict the potential interactions between proteins, including physical contacts and indirect correlations relating to protein functionality, while also constructing a representation of protein-protein interaction networks [15]. The aim is to elucidate the nature of the interactions that exist between these genes or proteins. This used the STRING database to search for the interaction between differentially expressed anoikis-related mRNAs, and a protein interaction network representation was constructed. Following the export of gene interaction results, the Cytoscape software [16] was utilized to visualize the protein-protein interaction (PPI) network diagram. Additionally, the 'cytoHubba' plug-in [17] was employed to calculate the node scores of genes in the PPI network, enabling the identification of the top ten key mRNAs.

2.4. Screening of prognostic-related lncRNAs by univariate Cox regression and a machine learning survival model

The expression data of anoikis-related lncRNAs and survival information (including survival time and survival status) were integrated according to the samples. Samples with missing survival data were eliminated for subsequent survival and prognosis analysis. Then, differential analysis was performed on the expression of lncRNAs between breast cancer patients and normal tissues. A threshold of $\text{adj.PVal} < 0.05$ and $|\text{logFC}| > 1$ were set to screen for differentially expressed lncRNAs. LncRNAs associated with survival were identified using univariate Cox regression analysis [18].

The Least Absolute Shrinkage and Selection Operator (LASSO) realizes the dimensionality reduction of genetic data by constructing a penalty function to compress and control the regression coefficients of independent variables. We applied LASSO penalty Cox regression to find significantly correlated genes [19,20]. Random Survival Forest (RSF) is a computational methodology based on the principles of random forest, tailored to analyze survival data that incorporates right-censored observations [21] and its effectiveness has been demonstrated in different fields [22]. In addition, RSF is also a non-parametric tree-based ensemble learning method that can automatically handle the difficulties of the Cox model and can also be used to select and rank variables [21,22].

We implemented the two algorithms using the R packages 'glmnet' and 'randomForestSRC', respectively, and incorporated the overlapping genes from the screening results of the two models as

prognostic anoikis-related lncRNAs into the prognostic scoring model for breast cancer patients.

2.5. Construction of anoikis risk score for breast cancer

We calculated the risk score for each patient according to the calculation method in the literature [23] to evaluate the risk rate of different populations. The calculation formula is as follows:

$$\text{Riskscore} = \mathrm{h}_0(t) \cdot \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n) \quad (1)$$

where β is the coefficient of multivariate cox regression, $\mathrm{h}_0(\mathrm{t})$ is the baseline hazard function at time t = 0, and x is the gene expression level. The predictive effect of Riskscore was evaluated using KM survival curves and time-dependent receiver operating characteristic (ROC) [24] curves. Subsequently, patients were segregated into high-risk and low-risk categories based on the optimal threshold.

2.6. Consensus clustering identifies distinct anoikis patterns

Consensus clustering is a common cluster analysis method [25], which is often used to find sample subsets with similar phenotypes. In cancer subtype analysis, consensus clustering classifies tumor samples with similar gene expression profiles into the same subtype by the k-Means method. For this, we can use consensus clustering to identify distinct anoikis patterns associated with anoikis gene expressions, classify samples into different clusters and apply survival analysis to study the difference in survival between different clusters. On this basis, the reliability of the clustering is verified using the Unified Manifold Approximation and Projection (UMAP) and tSNE methods.

2.7. Exploring the significance of anoikis prognostic genes on cancer treatment response based on the GDCS database

GDSC (Genomics of Drug Sensitivity in Cancer) can be used to guide the optimal clinical application of anticancer drugs and has a significant impact on the design, cost and ultimate success of new anticancer drug development [26]. OncoPredict is a tool designed to predict drug response in cancer patients based on cellular screening data [27]. It utilizes large-scale gene expression and drug screening data (training dataset) to build a ridge regression model. Applying newly acquired gene expression data (test dataset) into the model, oncoPredict generates predictions regarding drug sensitivity. Based on this method, we can use the cell line drug sensitivity data from the GDSC database as the training set and our expression data as a new dataset for oncoPredict to predict the half-maximal inhibitory concentration (IC50) for each drug and each patient [26]. IC50 is a measurement parameter used to describe the potency or toxicity of a drug. Regarding cellular activity, the lower the IC50 value, the higher the drug's potency. Therefore, by comparing IC50 values, we can assess the differential effects of drugs among patients with different risks and evaluate the sensitivity differences to drugs in different risk groups. This helps in understanding the response of different patients to anticancer drugs. Differences in IC50 values of chemotherapeutic drugs in the high and low Riskscore groups will be compared using the Wilcoxon test.

2.8. Selection of anoikis-related methylation sites

Methylation data obtained from UCSC are Beta values ranging from 0 to 1, and Beta values are obtained by calculating the intensity between methylated and unmethylated alleles. $\text{Beta} \geq 0.6$ means complete methylation, $0.2 \leq \text{Beta} \leq 0.6$ is considered partially methylated and $\text{Beta} \leq 0.2$ means completely unmethylated.

In this paper, we filter the methylation data of breast cancer according to the following criteria: (a) Filter probes with detection p-value (default > 0.01); (b) exclude probes with < 3 beads in at least 5% of samples per probe; (c) filter out all non-CpG probes contained in your dataset; (d) filter all SNP-related probes; (e) filter all multi-hit probes; and (f) filter out all probes located in chromosome X and Y. The data was subjected to champ filtering, followed by differential methylation analysis. Differential methylation sites were then identified using $P < 0.05$ and $|\text{delta beta}| > 0.2$ as the threshold values. The next step involved identifying the genes corresponding to the differentially methylated sites through the annotation file. Subsequently, the sites without corresponding genes were excluded. According to the crossover between the gene corresponding to the methylation site and the anoikis-related gene, the methylation site corresponding to the crossover gene is retained, that is, the anoikis-related methylation site, the survival-related anoikis methylation sites were screened by univariate cox regression.

2.9. Deep learning model: DeepSurv

DeepSurv is a model that applies deep learning techniques to the Cox proportional hazards loss function [28], it nonlinearizes the linear function in the Cox proportional hazards model, thereby learning and predicting the complex nonlinear relationship between characteristics and individual mortality risk. DeepSurv is a deep feed-forward neural network, which can predict the influence of patient covariates on the hazard rate by setting the network weight θ. The fundamental components of DeepSurv are depicted in Figure 2. The baseline data of patients serves as the input to the network, while the output is represented by a single node with a linear activation function. This linear activation estimates the logarithmic hazard function in the Cox proportional hazards model. The network is trained by setting the objective function as Eq (2) [28]:

$$l(\theta) \coloneqq -\frac{1}{N_{E=1}} \sum_{i:E_i=1} \left(\hat{h}_\theta(x_i) - \log \sum_{j \in \Re(T_i)} e^{\hat{h}_\theta(x_j)} \right) + \lambda \cdot \|\theta\|_2^2 \tag{2}$$

It is worth noting that many machine learning algorithms do not contain time-to-event data and only treat the results as discrete, such as prognosis and treatment response outcomes [29]. However, the DeepSurv model has applicability in survival analysis and has some advantages in handling survival data and predicting risk. It can automatically learn higher-level representations from input features and help the model capture complex interactions and nonlinear relationships, thereby improving predictive performance. Experimental results from Feng Zhu et al. [30] show that the DeepSurv model outperforms traditional Cox models and is more robust in the case of missing data, making it more suitable for revealing real-world situations. As a more common deep learning model for prognosis, DeepSurv has been widely promoted to different tasks, proving its performance stability.

In this study, we employ the final set of mRNA-lncRNA and DNA methylation features related to anoikis as the input for DeepSurv. A single output node is utilized to compute the patients' survival risk by applying the negative logarithmic partial likelihood function. Furthermore, we evaluate the prediction performance using the consistency index. The C-index value assesses the predictive accuracy, where a score of 0.5 indicates random predictions and higher values represent improved performance. Additionally, a high C-index suggests that the selected anoikis-related features have significant predictive value for deep survival analysis.

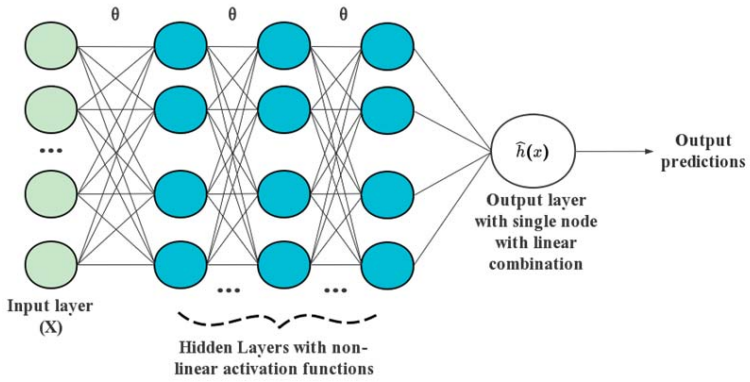

Figure 2. DeepSurv model frame diagram.

3. Results

The study mainly included transcriptome data of 1113 breast cancer samples and 113 normal samples from TCGA, and methylation data of 891 samples. First, the variance filtering method was used to exclude 3716 genes with a variance of 0. Subsequently, a total of 19,962 mRNAs and 16,275 lncRNAs were extracted based on the annotation files.

3.1. Identification of prognostic mRNAs associated with anoikis

A total of 4514 differentially expressed mRNAs were identified in both normal and tumor tissues, with 2119 being upregulated and 2395 downregulated. By intersecting with 794 anoikis-related genes, we obtained 249 overlapping genes (Figure 3), which are the differentially expressed mRNAs associated with anoikis.

To explore the interactions between the 249 anoikis-related mRNAs, we created a protein-protein interaction (PPI) network using the String database and Cytoscape software (Figure 4). We focused our analysis on individual networks comprising ten or more nodes while excluding networks with fewer than ten nodes. Additionally, we quantified the connectivity of each node within these networks. Based on the node scores, we identified the top ten key mRNAs (Figure 4): PLK1, BUB1, CCN1 (CCNA2), CDK1, BIRC5, MYC, PIK3R1, EGF, EGFR and CAV1.

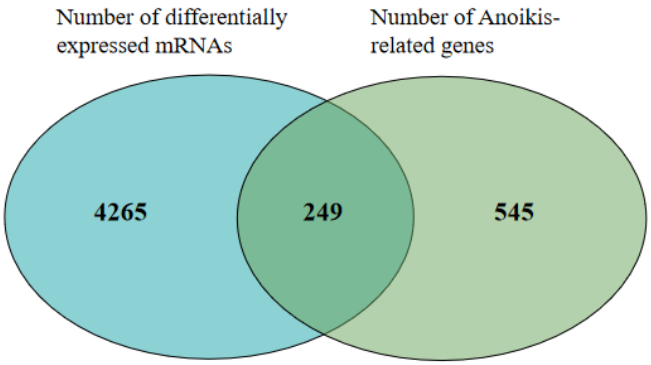

Figure 3. Venn diagram of mRNAs and anoikis genes.

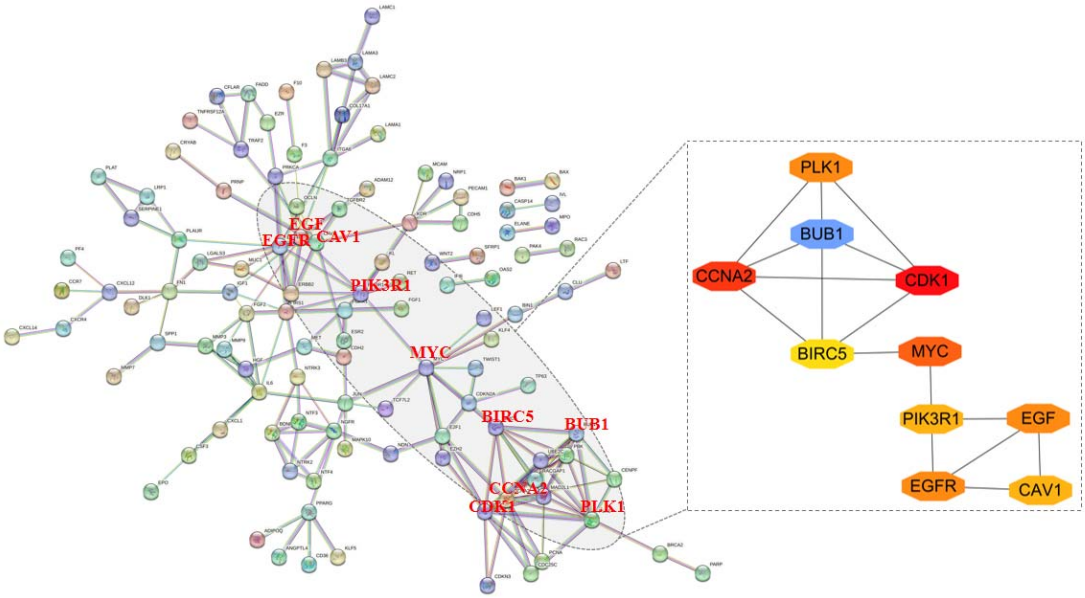

Figure 4. PPI network and top ten key mRNAs.

3.2. Identification of prognostic lncRNAs associated with anoikis

An investigation into the correlation of its expression data with the anoikis genes was carried out through a meticulous Pearson correlation analysis for the lncRNA expression matrix. Following the rigorous assessment, a selection of 1481 lncRNAs that exhibited a Pearson correlation coefficient surpassing 0.5 and an adjusted P value less than 0.05 were retained for further analysis. Then 548 differentially expressed lncRNAs were screened out according to $\text{adj.PVal} < 0.05$ and $|\text{logFC}| > 1$ (Figure 5(a)). Last, employing a univariate Cox regression analysis at a significance level of $P < 0.05$, a compelling ensemble of 21 lncRNAs with considerable impact on patient survival was attained.

The least absolute shrinkage and selection operator (LASSO) regression analysis identified 15 lncRNAs with non-zero coefficients. Figure 5(b),(c) represent the variation characteristics of the Lasso Cox regression variable coefficients and the process of selecting the optimal value of parameter i in the Lasso regression model through the cross-validation method, respectively. Concurrently, the random survival forest was employed to screen lncRNAs based on their feature importance.

Specifically, the analysis was configured with three trees, and a threshold of 0.3 was applied to determine the importance of features, leading to the acquisition of seven lncRNAs. Figure 5(d),(e) show the error rate curve and variable importance of RSF, respectively. By intersecting the feature selection results of Lasso Cox regression and random survival forest, six prognostic lncRNAs related to anoikis were obtained: AC068473.4, AC104109.2, MRPL20.AS1, AC026310.1, AC137056.1 and AP006545.2.

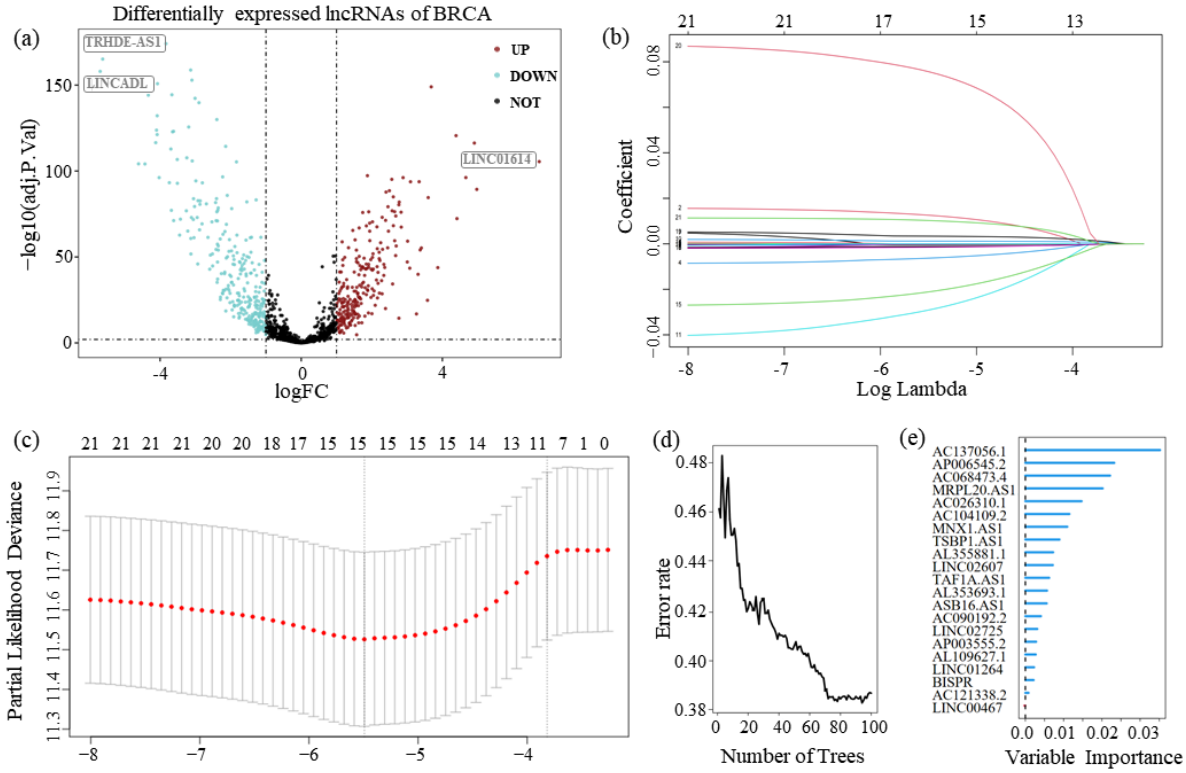

Figure 5. Differential analysis, Lasso regression and random survival forest used for screening of anoikis-related lncRNAs. (a) Volcano plot for differential analysis of anoikis-related lncRNAs; (b) Least absolute shrinkage and selection operator (LASSO) coefficients; (c) Tenfold cross-validation for tuning parameter selection in the LASSO model; (d) Error rates for different numbers of trees in a random survival forest; and (e) Corresponding importance of different genes.

3.3. Constructing a risk score prognostic model

The expression data of the last screened ten mRNAs and six lncRNAs were merged, and the risk score of each patient was calculated using Eq (1). Time-dependent ROC curves were utilized to evaluate the performance, showing area under the curve (AUC) values of 0.75, 0.74 and 0.77 for 1-, 3- and 5-year overall survival, respectively (Figure 6(a)). We calculate the Maxstat (Maximally Selected Rank Statistics) statistic [31] for each possible cutoff value of Riskscore, and select the cutoff value with the maximum Maxstat statistic as the optimal cutoff value. Patients were stratified into high-risk and low-risk groups based on the optimal cut-off value. The Kaplan-Meier was applied to analyze the differences in prognosis between the two groups. The analysis of the Kaplan-Meier curve unveiled a distinct divergence in survival probability and prognosis between the high-risk group and the low-risk

group (Figure 6(b)). Notably, the high-risk group displayed a considerably lower survival probability and worse prognosis, highlighting the potential prognostic significance of the risk stratification. A decision curve analysis was performed to determine the clinical applicability and value. The decision curve in Figure 6(c) displays the prediction curves of the two models along with their corresponding confidence intervals. It demonstrates that when compared to the model solely based on clinical information (Base model), the inclusion of the risk score (Full model) yielded a more favorable net benefit and enhanced instructiveness for clinical applications. Therefore, incorporating risk scores provides clinicians with a more precise means of assessing patient outcomes.

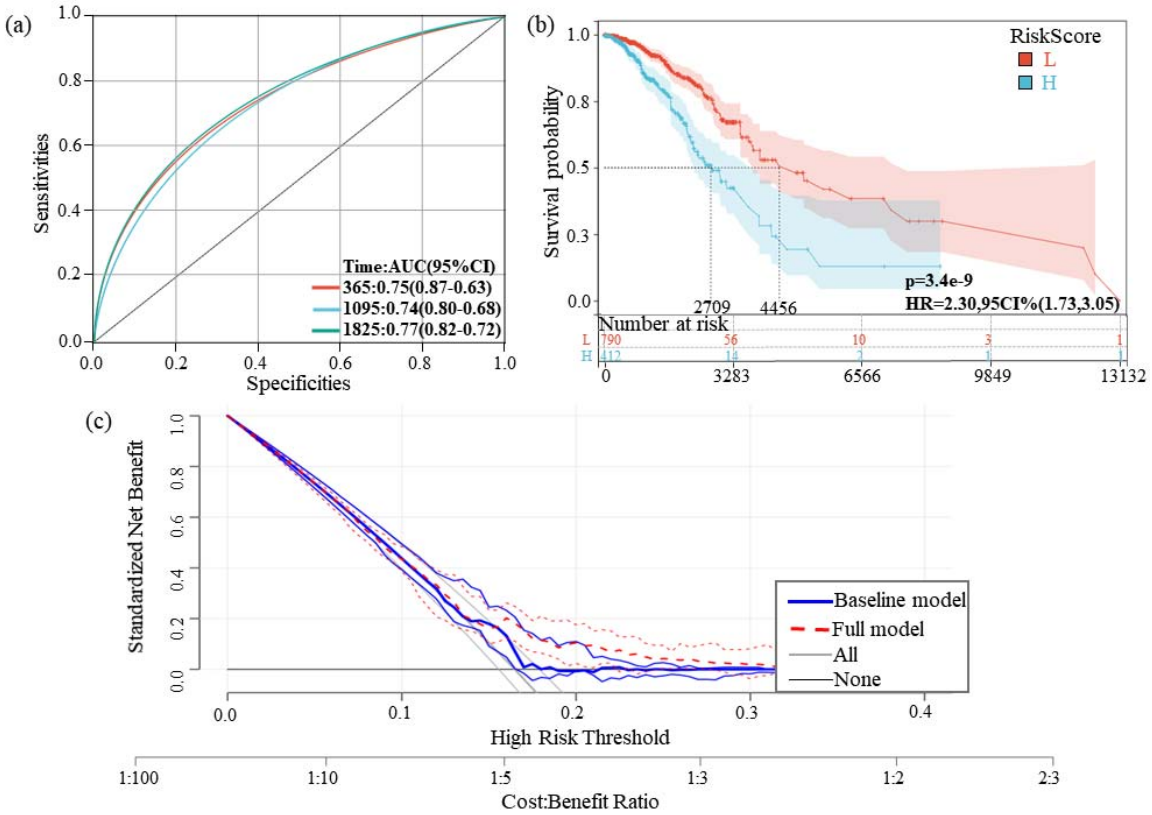

Figure 6. Validation analysis of the Riskscore. (a) Receiver operating characteristic (ROC) reflects the role of Riskscore in predicting overall survival; (b) Survival curve plot for high RiskScore and low RiskScore; and (c) Decision curve analysis (DCA) curves to evaluate the predictive effect of risk scores from the perspective of clinical benefit, the X-axis represents the threshold probability, while the Y-axis represents the net benefit rate.

3.4. 16 anoikis-related genes were used to analyze breast cancer subtypes

In order to identify different anoikis-related gene expression patterns and thereby better understand the survival differences between different subtypes, we used the expression data of 16 prognosis-related genes to perform consistent clustering on breast cancer patients. Remarkably, when the optimal value of k was set at 4, the cohort could be effectively stratified into four distinct subtypes (Figure 7(a),(b)). Furthermore, subsequent analysis of overall survival demonstrated significant

variations in prognosis among these four subtypes (Figure 7(c)), indicating the potential clinical relevance and prognostic significance of these subtypes in breast cancer. To validate the accuracy of this clustering, UMAP and tSNE techniques were employed. The results depicted that when the parameter was set to 4, the four cluster subtypes could be reliably identified (Figure 7(d),(e)).

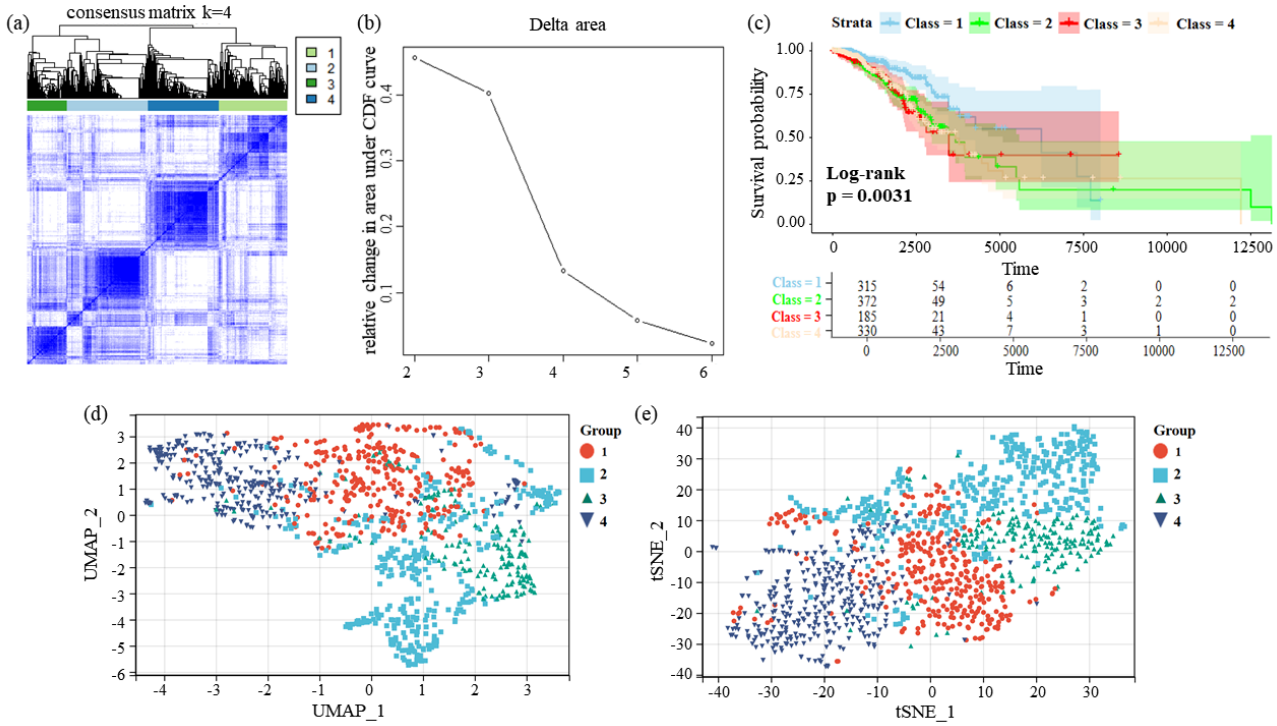

Figure 7. BRCA subsets are associated with anoikis genes. (a) Consensus matrix for k = 4 was obtained by applying consensus clustering; (b) Gravel diagram of consensus clustering; (c) Differences in the overall survival rate of the four subtypes ($P < 0.01$); (d) and (e) UMAP and tSNE distinguish four subtypes based on the expression of anoikis-related genes.

3.5. Correlation analysis of anoikis-related genes and tumor microenvironment

We utilized the BRCA_GSE176078 single-cell dataset from the TISCH database to investigate the expression patterns of 10 anoikis-related mRNAs within the tumor microenvironment (TME). In the GSE176078 dataset, TISCH integrates data from multiple single-cell samples or conditions. Based on the transcriptional features of cells, TISCH divides the cells into 39 cell clusters. Subsequently, it calculates and presents the distribution of 11 intermediate cell types based on the cell counts in each cluster (Figure 8), illustrating the relative proportions of different cell types in the samples or conditions. In Figure 9, CAV1 demonstrated high expression levels in stromal cells (including Endothelial and Fibroblasts), smooth muscle cells (SMC) and macrophages. PIK3R1 exhibited predominant expression in malignant cells, fibroblasts and immune cells (specifically B cells, CD4Tconv cells, CD8Tex cells and DC cells). Additionally, MYC exhibited elevated expression levels across various cell types, while CDK1 displayed pronounced expression in Tprolif cells. PLK1, EGFR and BUB1 were barely detected in the TME. Overall, our findings highlight the association of these anoikis-related mRNAs primarily with malignant cells and immune cells, thus offering potential

insights for the development of targeted gene therapy strategies tailored to specific cell populations.

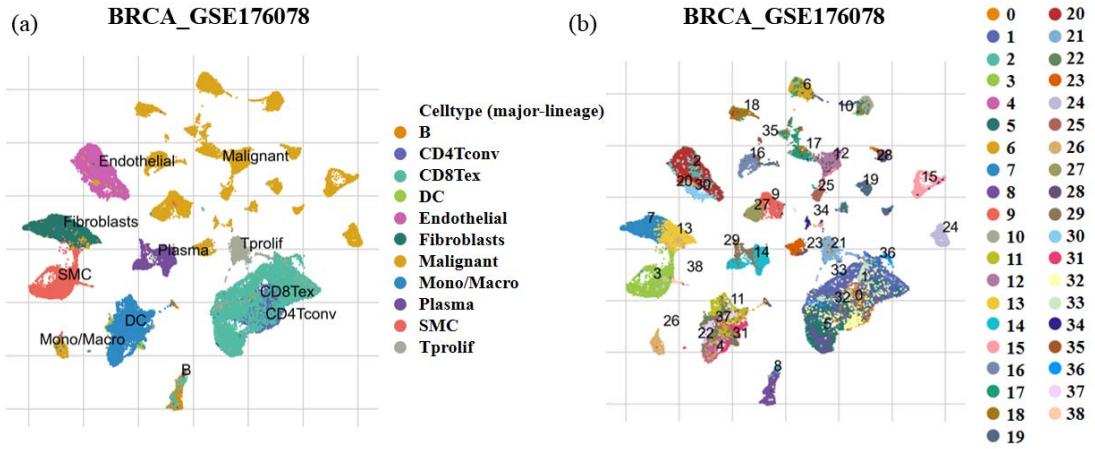

Figure 8. Annotation of all cell types and cell clusters in GSE1760782. (a) Annotation of 11 major cell types; (b) Distribution of 38 major cell clusters.

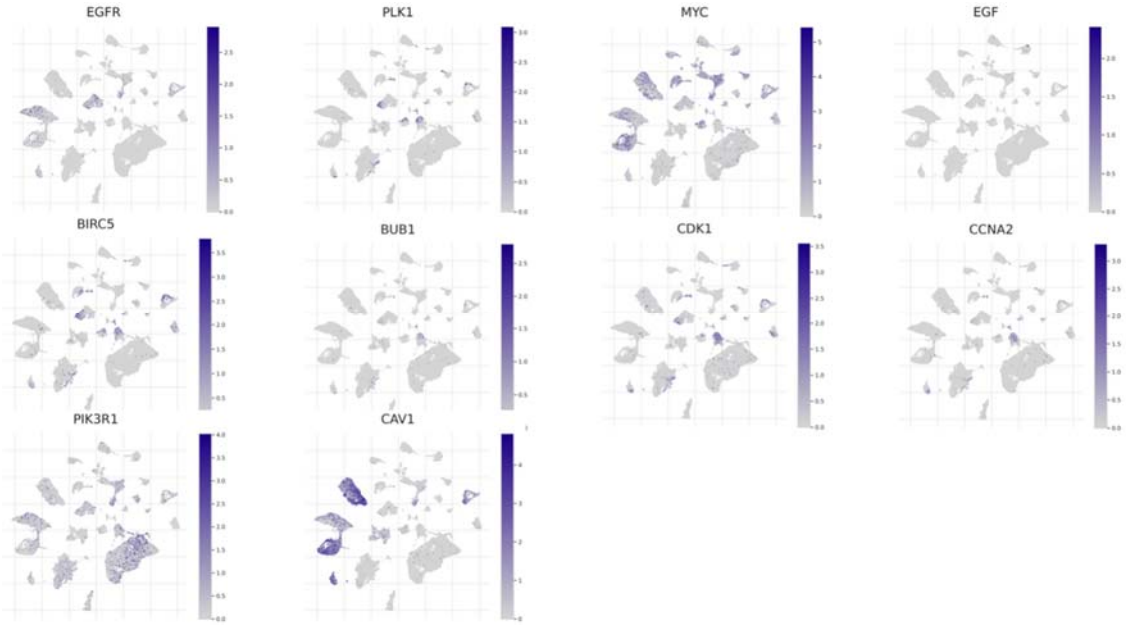

Figure 9. Expression of anoikis-related mRNAs in TME-related cells of breast cancer.

3.6. Drug sensitivity analysis

The gene expression was transformed into a drug sensitivity matrix based on the oncoPredict algorithm (Figure 10), and the IC50 levels between the two risk groups were compared. After logarithmically transforming the IC50 values, we conducted a Wilcoxon test. Based on the Wilcoxon test, it was demonstrated that 49 medications were exhibiting distinct sensitivities in the high-risk group compared to the low-risk group. Notably, within the cohort of drugs investigated, two compounds presented increased sensitivity towards the high-risk group compared to the low-risk group, as illustrated in Figure 11. This implies that patients in the high-risk group may benefit from these

medications, providing potential therapeutic options for individuals with high-risk scores.

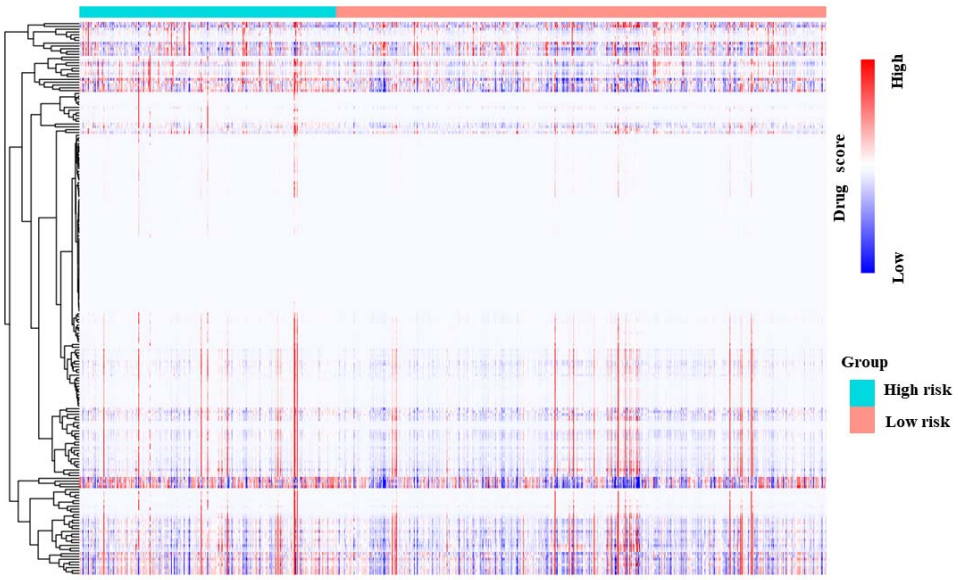

Figure 10. OncoPredict was used to convert gene expression profiles into a drug susceptibility matrix for 198 drugs.

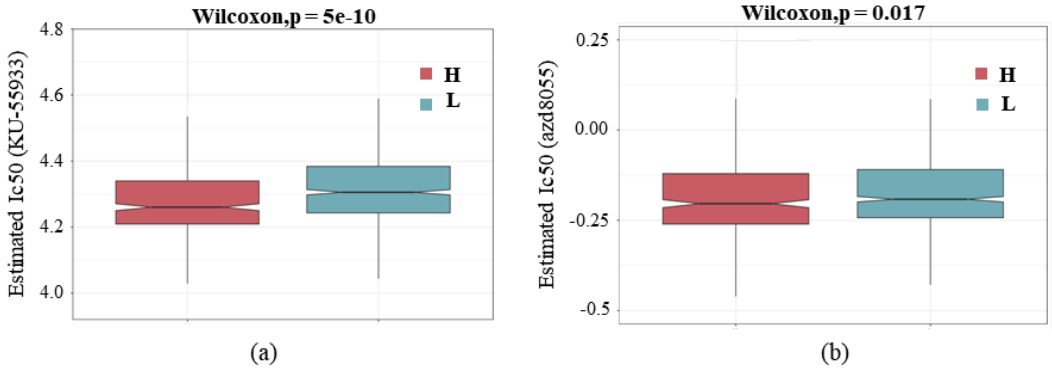

Figure 11. Sensitive drugs found in high-risk groups. (a) KU-55933, a selective DNA-dependent protein kinase (ATM) inhibitor; and (b) AZD8055, the class of mammalian target of rapamycin (mTOR) inhibitors.

3.7. Deep learning predicts patient risk

We applied the deep learning model DeepSurv to our screened gene and methylation probes to assess and predict patient risk. The original DeepSurv model, developed by Shaham et al., relies on the Theano and Lasagne libraries. However, considering the possibility of errors within these libraries, we opted to implement DeepSurv using the widely adopted Keras framework. In the DeepSurv model, we use the gradient descent optimization method to find the network weights that minimize the loss function. To optimize network training, we utilized modern deep learning techniques, including input normalization, the Scaled Exponential Linear Unit (SELU) as an activation function, the Adaptive Estimation of Moments (ADAM) algorithm for gradient descent, Nesterov momentum and learning

rate decay. We predefine the range of the number of layers, number of nodes per layer, learning rate, decay, activation function, dropout probability and L2 regularization for the network. Then, we used grid hyperparameter search to obtain the optimal parameters within these predefined ranges. The screened anoikis-related RNA signature genes and methylation signatures are used as the input of DeepSurv. The results demonstrated that the integration of multi-omics data, specifically combining mRNA-lncRNA signatures with methylation data, yielded improved prognostic accuracy for breast cancer patients, as indicated by a C-index of 0.7779 (Table 1). It's important to highlight that cancer datasets often exhibit a relatively limited number of samples. Consequently, when fine-tuning parameters for larger networks, the risk of overfitting becomes a notable concern. The critical consideration here is selecting an appropriately sized network to balance model complexity and the dataset's intrinsic limitations. Once an optimal network size is determined, it becomes apparent that variations in other hyperparameters have minimal impact on the model's performance. This underscores the sensitivity of deep learning models to dataset size and the necessity for a reasonable approach in parameter adjustments for effective and robust performance in the context of cancer datasets.

Bichindaritz et al. [32] used the weighted mining algorithm of local maximum quasi-clique merging (lmQCM) [33] as a feature extraction method. They extracted 17 methylation features and 116 mRNA features from breast cancer data in TCGA, which were then applied to multiple models for performance comparison. Notably, the DeepSurv model achieved a C-index of 0.6523. In a study conducted by Zhang et al. [34], four methods were compared across six cancer datasets. Specifically, DeepSurv yielded an average C-index of approximately 0.65 when applied to the 20,502-dimensional transcriptome genes of TCGA-BRCA (refer to Figure 3 in the original publication). Based on TCGA-based breast cancer data, the identification of dimensions related to anoikis-related genes and methylation signatures not only circumvents the challenges posed by high-dimensionality but also holds the potential for more effective prediction of patient survival probability. This underscores the significance of including anoikis-related prognostic features in the prognosis assessment of breast cancer patients. Moreover, the fusion of multi-omics features exhibits superior prognostic efficacy compared to the utilization of single-omics alone.

Table 1. DeepSurv model evaluation results.

BRCA from TCGA	C-index
mRNAs+lncRNAs	0.6761
mRNAs+lncRNAs+meth	0.7779

4. Discussion

Anoikis is a crucial defense of organisms, which prevents detached cells from reattaching to new substrates in the wrong place and prevents their growth arrest [35]. Unlike apoptosis, anoikis is specifically triggered by the loss of cell adhesion, representing a distinct programmed cell death mode. Under physiological conditions, the significance of anoikis in normal tissues lies in its ability to prevent ectopic cell colonization, promoting proper bodily development and maintaining tissue homeostasis. However, tumor cells exhibit resistance to anoikis under pathological conditions. In the process of tumor cells detaching from the primary site, passing through the lymphatic and circulatory system, invading and implanting into the secondary site for proliferation and growth, the tolerance to anoikis becomes a prerequisite for tumor metastasis and drug resistance [36]. The resistance of

cancer cells to anoikis contributes to the invasion, migration and development of drug resistance in tumors [37]. Numerous studies have demonstrated the presence of anoikis resistance in various types of tumors, including colon cancer, breast cancer, lung cancer, melanoma, prostate cancer, gastric cancer, among others.

In this study, we aim to provide a comprehensive perspective on the investigation of anoikis-associated multi-omic signatures. Developing a multigene signature prognostic model based on transcriptome profiles of mRNA and lncRNA can effectively divide BRCA patients into high and low risk groups, with the DCA curve demonstrating favorable clinical utility. Furthermore, we investigated anoikis-related methylation probes. The results of the DeepSurv model showed that integrated anoikis-related methylation and mRNA-lncRNA multi-omic signatures could serve as a more effective prognostic tool for breast cancer patients.

Most of the genes we screened have been experimentally shown to be associated with cancer. These include CAV1, PLK1, BUB1, CCNA2, MYC, EGFR and EGF. CAV-1 mediates the resistance of H460 cells to anoikis [38]. The research indicates that CAV1 plays a crucial role in the progression of breast cancer. Upregulating CAV-1 expression can effectively inhibit the growth of primary breast cancer and its metastasis to the brain, which has potential therapeutic implications for preventing tumor invasion and distant metastasis [39]. PLK1 is a serine/threonine kinase that is highly expressed during the G2 phase of the cell cycle [40–42]. The overactivation of PLK1 is associated with the dysregulation of the normal ER signaling pathway, resulting in hormone resistance in breast cancer cells [43]. Therefore, inhibiting PLK1 to overcome the issue of hormone resistance in treatment may provide a new direction for breast cancer therapy.

The BUB1 protein is a crucial mitotic checkpoint kinase that regulates the cell cycle by encoding a specific serine/threonine protein kinase [44]. The increased chromosomal instability in breast cancer cells may be closely associated with the upregulation of BUB1 expression [45]. In some aneuploid breast cancer cell lines, the BUB1 gene shows mutations and regulates its expression [46]. This suggests that BUB1 may play a significant role in breast cancer's occurrence, progression and prognosis. CCNA2 is one of the highly conserved cyclin family [47]. It regulates the G1-S and G2-M transitions of the cell cycle, is a known prognostic biomarker for survival in BRCA patients and has been associated with tamoxifen resistance [48]. MYC is a potent activator of tumorigenesis and its dysregulation has been found in various cancers [49]. The overexpression of MYC may lead to increased endoplasmic reticulum stress and cellular metabolic abnormalities and promote endocrine resistance in breast cancer cells [50,51]. Therefore, the MYC gene is considered an essential molecular target in the development and treatment of breast cancer. The abnormal expression or mutation of EGFR is associated with the occurrence and progression of various tumors. In 15–30% of breast cancer patients, overexpression of EGFR has been observed, which is associated with larger tumors and unfavorable clinical outcomes [52,53].

Several studies have shown that EGF may regulate the growth of mammary epithelium: (a) EGF promotes the growth of normal mammary tissue and rodent breast cancer [54]; (b) promotes the growth of normal human mammary epithelial cells in short-term culture [55]; and (c) it stimulates mitosis in benign breast fibroadenoma cells [56]. Anoikis-related mRNA-lncRNA risk signals include 6 lncRNAs, namely AC068473.4, AC104109.2, MRPL20.AS1, AC026310.1, AC137056.1 and AP006545.2. As far as we know, there are no research reports on the 6 lncRNAs so far.

As of now, our study stands as the inaugural endeavor to craft a set of mRNA-lncRNA gene features associated with anoikis for prognosis of breast cancer patients. This pioneering effort involves

the integration of mRNA and lncRNA expression profiles at a comprehensive whole-genome gene expression level. These intricately derived features are then harnessed within a deep survival model, a novel approach to predicting patient survival risk with heightened precision and insight. This study inevitably has some limitations. First, our study was based on data from publicly available datasets and was not tested in a prospective clinical trial, making it susceptible to bias inherent in this form of research. Second, the biological functions of integrated features still need to be further incorporated. In addition, the comparison of model results based on deep learning may ignore the differences in sample selection and experiments, and several models with better performance should be considered when used to predict the survival probability of patients.

5. Conclusions

We developed an integrated signature comprising mRNA-lncRNA associated with anoikis, which exhibits the capability to accurately classify breast cancer patients into low-risk and high-risk groups. Our findings propose that this signature holds significant potential as a robust prognostic tool for predicting survival outcomes in breast cancer patients. In addition, we used anoikis-related mRNA-lncRNA and methylation features to evaluate and predict the risk of death in breast cancer patients with a deep learning model, resulting in promising outcomes. Further validation of prospective clinical trial results can facilitate personalized treatment and consultation for breast cancer patients.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (No. 62162032), the Scientific Research Plan of the Department of Education of Jiangxi Province, China (GJJ2201004), the Jingdezhen City Biological Big Data Key Laboratory (20234PTYH001).

Conflict of interest

The authors declare no competing interests in this work.

References

1. Y. S. Sun, Z. Zhao, Z. N. Yang, F. Xu, H. J. Lu, Z. Y. Zhu, et al., Risk factors and preventions of breast cancer, *Int. J. Biol. Sci.*, **13** (2017), 1387–1397. https://doi.org/10.7150%2Fijbs.21635
2. T. J. Key, P. K. Verkasalo, E. Banks, Epidemiology of breast cancer, *Lancet Oncol.*, **2** (2001), 133–140. https://doi.org/10.1016/S1470-2045(00)00254-0
3. Y. N. Kim, K. H. Koo, J. Y. Sung, U. J. Yun, H. Kim, Anoikis resistance: an essential prerequisite for tumor metastasis, *Int. J. Cell Biol.*, **2012** (2012), e306879. https://doi.org/10.1155/2012/306879

4. S. Frisch, H. Francis, Disruption of epithelial cell-matrix interactions induces apoptosis, *J. Cell Biol.*, **124** (1994), 619–626. https://doi.org/10.1083/jcb.124.4.619
5. M. C. Guadamillas, A. Cerezo, M. A. del Pozo, Overcoming anoikis–pathways to anchorage-independent growth in cancer, *J. Cell Sci.*, **124** (2011), 3189–3197. https://doi.org/10.1242/jcs.072165
6. Y. Luo, W. Q. Tang, S. S. Xiang, J. B. Feng, X. Y. Zu, Non-coding RNAs in breast cancer: Implications for programmed cell death, *Cancer Lett.*, **550** (2022), 215929. https://doi.org/10.1016/j.canlet.2022.215929
7. D. Fanfone, Z. C. Wu, J. Mammi, K. Berthenet, D. Neves, K. Weber, et al., Confined migration promotes cancer metastasis through resistance to anoikis and increased invasiveness, *eLife*, **11** (2022), e73150. https://doi.org/10.7554/eLife.73150
8. K. Zhao, Z. Wang, T. Hackert, C. Pitzer, M. Zöller, Tspan8 and Tspan8/CD151 knockout mice unravel the contribution of tumor and host exosomes to tumor progression, *J. Exp. Clin. Cancer Res.*, **37** (2018), 312. https://doi.org/10.1186/s13046-018-0961-6
9. C. Akekawatchai, S. Roytrakul, S. Kittisenachai, P. Isarankura-Na-Ayudhya, S. Jitrapakdee, Protein profiles associated with anoikis resistance of metastatic MDA-MB-231 breast cancer cells, *Asian Pac. J. Cancer Prev.*, **17** (2016), 581–590. https://doi.org/10.7314/APJCP.2016.17.2.581
10. B. Weigelt, J. L. Peterse, L. J. van't Veer, Breast cancer metastasis: markers and models, *Nat. Rev. Cancer*, **5** (2005), 591–602. https://doi.org/10.1038/nrc1670
11. W. Q. Li, J. Lee, H. G. Vikis, S. H. Lee, G. F. Liu, J. Aurandt, et al., Activation of FAK and Src are receptor-proximal events required for netrin signaling, *Nat. Neurosci.*, **7** (2004), 1213–1221. https://doi.org/10.1038/nn1329
12. Y. Su, H. J. Wu, A. Pavlosky, L. L. Zou, X. N. Deng, Z. X. Zhang, et al., Regulatory non-coding RNA: new instruments in the orchestration of cell death, *Cell Death Dis.*, **7** (2016), e2333. https://doi.org/10.1038/cddis.2016.210
13. M. Rebhan, V. Chalifa-Caspi, J. Prilusky, D. Lance, GeneCards: integrating information about genes, proteins and diseases, *Trends Genet.*, **13** (1997), 163. https://doi.org/10.1016/s0168-9525(97)01103-7
14. M. E. Ritchie, B. Phipson, D. Wu, Y. F. Hu, C. W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.*, **43** (2015), e47. https://doi.org/10.1093/nar/gkv007
15. H. V. Cook, N.T. Doncheva, D. Szklarczyk, C. von Mering, J. L. Juhl, STRING: A virus-host protein-protein interaction database, *Viruses*, **10** (2018), 519. https://doi.org/10.3390/v10100519
16. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, et al., Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13** (2003), 2498–2504. http://www.genome.org/cgi/doi/10.1101/gr.1239303
17. C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, C. Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, *BMC Syst. Biol.*, **8** (2014), S11. https://doi.org/10.1186/1752-0509-8-S4-S11
18. D. R. Cox, Regression models and life-tables, *J. R. Stat. Soc. B.*, **34** (1972), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
19. R. Tibshirani, The lasso method for variable selection in the Cox model, *Stat. Med.*, **16** (1997), 385–395. https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

20. N. Simon, J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for Cox's proportional hazards model via coordinate descent, *J. Stat. Software*, **39** (2011), 1–13. https://doi.org/10.18637%2Fjss.v039.i05
21. H. Ishwaran, U. B. Kogalur, E. H. Blackstone, M. S. Lauer, Random survival forests, *Ann. Appl. Stat.*, **2** (2008), 841–860. https://doi.org/10.1214/08-AOAS169
22. X. Chen, H. Ishwaran, Random forests for genomic data analysis, *Genomics*, **99** (2012), 323–329. https://doi.org/10.1016/j.ygeno.2012.04.003
23. H. Q. Lin, D. Zelterman, Modeling survival data: extending the Cox model, *Technometrics*, **44** (2002), 85–86. https://doi.org/10.1198/tech.2002.s656
24. A. N. Kamarudin, T. Cox, R. Kolamunnage-Dona, Time-dependent ROC curve analysis in medical research: current methods and applications, *BMC Med. Res. Methodol.*, **17** (2017), 53. https://doi.org/10.1186/s12874-017-0332-6
25. S. Monti, P. Tamayo, J. Mesirov, T. Golub, Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data, *Mach. Learn.*, **52** (2003), 91–118. https://doi.org/10.1023/A:1023949509487
26. W. J. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, et al., Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, *Nucleic Acids Res.*, **41** (2013), D955–D961. https://doi.org/10.1093/nar/gks1111
27. D. Maeser, R. F. Gruener, R. S. Huang, OncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data, *Briefings Bioinf.*, **22** (2021), bbab260. https://doi.org/10.1093/bib/bbab260
28. J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, Y. Kluger, DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network, *BMC Med. Res. Methodol.*, **18** (2018), 1–12. https://doi.org/10.1186/s12874-018-0482-112
29. J. Adeoye, L. L. Hui, M. Koohi-Moghadam, J. Y. Tan, S. W. Choi, P. Thomson, Comparison of time-to-event machine learning models in predicting oral cavity cancer prognosis, *Int. J. Med. Inf.*, **157** (2022), 104635. https://doi.org/10.1016/j.ijmedinf.2021.104635
30. F. Zhu, R. Zhong, F. Li, C. C. Li, N. Din, H. Sweidan, et al., Development and validation of a deep transfer learning-based multivariable survival model to predict overall survival in lung cancer, *Transl. Lung Cancer Res.*, **12** (2023), 471–482. https://doi.org/10.21037%2Ftlcr-23-84
31. B. Lausen, M. Schumacher, Maximally selected rank statistics, *Biometrics*, **48** (1992), 73–85. https://doi.org/10.2307/2532740
32. I. Bichindaritz, G. H. Liu, C. Bartlett, Integrative survival analysis of breast cancer with gene expression and DNA methylation data, *Bioinformatics*, **37** (2021), 2601–2608. https://doi.org/10.1093/bioinformatics/btab140
33. J. Cheng, J. Zhang, Y. Han, X. S. Wang, X. F. Ye, Y. B. Meng, et al., Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis, *Cancer Res.*, **77** (2017), e91–e100. https://doi.org/10.1158/0008-5472.CAN-17-0313
34. Z. Y. Zhang, H. Chai, Y. Wang, Z. X. Pan, Y. D. Yang, Cancer survival prognosis with deep bayesian perturbation Cox network, *Comput. Biol. Med.*, **141** (2022), 105012. https://doi.org/10.1016/j.compbiomed.2021.105012
35. E. Kakavandi, R. Shahbahrami, H. Goudarzi, G. Eslami, E. Faghihloo, Anoikis resistance and oncoviruses, *J. Cell. Biochem.*, **119** (2018), 2484–2491. https://doi.org/10.1002/jcb.26363

36. M. J. Zou, E. Y. Baitei, R. A. Al-Rijjal, R. S. Parhar, F. A. Al-Mohanna, S. Kimura, et al., KRASG12D-mediated oncogenic transformation of thyroid follicular cells requires long-term TSH stimulation and is regulated by SPRY1, *Lab. Invest.*, **95** (2015), 1269–1277. https://doi.org/10.1038/labinvest.2015.90
37. S. Li, Y. Chen, Y. H. Zhang, X. M. Jiang, Y. Jiang, X. Qin, et al., Shear stress promotes anoikis resistance of cancer cells via caveolin-1-dependent extrinsic and intrinsic apoptotic pathways, *J. Cell. Physiol.*, **234** (2019), 3730–3743. https://doi.org/10.1002/jcp.27149
38. P. Chanvorachote, U. Nimmannit, Y. Lu, S. Talbott, B. H. Jiang, Y. Rojanasakul, Nitric oxide regulates lung carcinoma cell anoikis through inhibition of ubiquitin-proteasomal degradation of caveolin-1, *J. Biol. Chem.*, **284** (2009), 28476–28484. https://doi.org/10.1074/jbc.M109.050864
39. W. T. Chiu, H. T. Lee, F. J. Huang, K. D. Aldape, J. Yao, P. S. Steeg, et al., Caveolin-1 upregulation mediates suppression of primary breast tumor growth and brain metastases by stat3 inhibition, *Cancer Res.*, **71** (2011), 4932–4943. https://doi.org/10.1158/0008-5472.CAN-10-4249
40. K. Strebhardt, A. Ullrich, Targeting polo-like kinase 1 for cancer therapy, *Nat. Rev. Cancer*, **6** (2006), 321–330. https://doi.org/10.1038/nrc1841
41. R. M. Golsteyn, K. E. Mundt, A. M. Fry, E. A. Nigg, Cell cycle regulation of the activity and subcellular localization of Plk1, a human protein kinase implicated in mitotic spindle function, *J. Cell Biol.*, **129** (1995), 1617–1628. https://doi.org/10.1083/jcb.129.6.1617
42. F. Toyoshima-Morimoto, E. Taniguchi, N. Shinya, A. Iwamatsu, E. Nishida, Polo-like kinase 1 phosphorylates cyclin B1 and targets it to the nucleus during prophase, *Nature*, **410** (2001), 215–220. https://doi.org/10.1038/35065617
43. N. E. Bhola, V. M. Jansen, S. Bafna, J. M. Giltnane, J. M. Balko, M. V. Estrada, et al., Kinome-wide functional screen identifies role of PLK1 in hormone-independent, ER-positive breast cancer, *Cancer Res.*, **75** (2015), 405–414. https://doi.org/10.1158/0008-5472.CAN-14-2475
44. A. P. Baron, C. Schubert, F. Cubizolles, G. Siemeister, M. Hitchcock, A. Mengel, et al., Probing the catalytic functions of Bub1 kinase using the small molecule inhibitors BAY-320 and BAY-524, *Elife*, **5** (2016), e12187. https://doi.org/10.7554/eLife.12187
45. B. Yuan, Y. Xu, J. H. Woo, Y. Y. Wang, Y. K. Bae, D. S. Yoon, et al., Increased expression of mitotic checkpoint genes in breast cancer cells with chromosomal instability, *Clin. Cancer Res.*, **12** (2006), 405–410. https://doi.org/10.1158/1078-0432.CCR-05-0903
46. K. A. Myrie, M. J. Percy, J. N. Azim, C. K. Neeley, E. M. Petty, Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines, *Cancer Lett.*, **152** (2000), 193–199. https://doi.org/10.1016/S0304-3835(00)00340-2
47. M. Uhlen, P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, et al., Towards a knowledge-based human protein atlas, *Nat. Biotechnol.*, **28** (2010), 1248–1250. https://doi.org/10.1038/nbt1210-1248
48. Q. Shi, Z. Zhou, N. S. Ye, Q. L. Chen, X. X. Zheng, M. S. Fang, MiR-181a inhibits non-small cell lung cancer cell proliferation by targeting CDK1, *Cancer Biomarkers*, **20** (2017), 539–546. https://doi.org/10.3233/cbm-170350
49. C. V. Dang, K. A. O'Donnell, L. I. Zeller, T. Nguyen, R. C. Osthus, F. Li, The c-Myc target gene network, *Semin. Cancer Biol.*, **16** (2006), 253–264. https://doi.org/10.1016/j.semcancer.2006.07.014
50. A. N. Shajahan-Haq, K. L. Cook, J. L. Schwartz-Roberts, A. E. Eltayeb, D. M. Demas, A. M. Warri, et al., MYC regulates the unfolded protein response and glucose and glutamine uptake in endocrine resistant breast cancer, *Mol. Cancer*, **13** (2014), 239. https://doi.org/10.1186/1476-4598-13-239

51. C. M. McNeil, C. M. Sergio, L. R. Anderson, C. K. Inman, S. A. Eggleton, N. C. Murphy, et al., c-Myc overexpression and endocrine resistance in breast cancer, *J. Steroid Biochem. Mol. Biol.*, **102** (2006), 147–155. https://doi.org/10.1016/j.jsbmb.2006.09.028
52. S. Tsutsui, S. Ohno, S. Murakami, Y. Hachitanda, S. Oda, Prognostic value of epidermal growth factor receptor (EGFR) and its relationship to the estrogen receptor status in 1029 patients with breast cancer, *Breast Cancer Res. Treat.*, **71** (2002), 67–75. https://doi.org/10.1023/A:1013397232011
53. C. J. Witton, J. R. Reeves, J. J. Going, T. G. Cooke, J. M. Bartlett, Expression of the HER1–4 family of receptor tyrosine kinases in breast cancer, *J. Pathol.*, **200** (2003), 290–297. https://doi.org/10.1002/path.1370
54. R. W. Turkington, Stimulation of mammary carcinoma cell proliferation by epithelial growth factor in vitro1, *Cancer Res.*, **29** (1969), 1457–1458.
55. J. Taylor-Papadimitriou, M. Shearer, M. G. P. Stoker, Growth requirements of human mammary epithelial cells in culture, *Int. J. Cancer*, **20** (1977), 903–908. https://doi.org/10.1002/ijc.2910200613
56. M. G. P. Stoker, D. Pigott, J. Taylor-Papadimitriou, Response to epidermal growth factors of cultured human mammary epithelial cells from benign tumours, *Nature*, **264** (1976), 764–767. https://doi.org/10.1038/264764a0

AIMS Press