2024 Volume 2, Issue 1
Big Data Analytics for Healthcare
ISSN: 2630-4767 (Online)

Bioinformatics-Based Identification of Key Metabolic Genes in Breast Cancer and Survival Prognosis Analysis

Yehao Luo[1], Xiusong Tang[1], Donghan Xu[2], Ting Lyu[3], Xianghua You[4], Yuzhou Pang[1], Renfeng Li[5]*

[1]School of Zhuang Medicine, Guangxi University of Chinese Medicine, Nanning 530200, Guangxi Province, China

[2]Faculty of Chinese Medicine, Macau University of Science and Technology, Macau 999078, China

[3]Hubei Minzu University, Enshi 445000, Hubei Province, China

[4]Zhuhai Integrated Traditional Chinese and Western Medicine Hospital, Zhuhai 519000, Guangdong Province, China

[5]The First Affiliated Hospital of Guangxi University of Chinese Medicine, Nanning 530023, Guangxi Province, China

***Corresponding author:** Renfeng Li, lirenfeng@163.com

Abstract:

Objective: To analyze key metabolic genes in breast cancer using bioinformatics methods and conduct survival prognosis analysis. *Methods:* Transcriptome data for breast cancer was obtained from the Cancer Genome Atlas (TCGA) database. Relevant metabolic genes were identified using the GSEA database and matched with genes in the TCGA database to determine the final metabolic genes. The Lasso model was constructed to obtain survival prognosis analysis results. *Results:* Three metabolic genes related to breast cancer were identified: *POLR2K*, *NMNAT2*, and *SUCLA2*. Survival analysis showed that the maximum survival time for both the high-risk and low-risk groups was 24 years. Age, status, and tumor stage were identified as independent prognostic factors. *Conclusion:* The *POLR2K* gene is the most significantly overexpressed and shows a preliminary correlation with the occurrence, development, and prognosis of breast cancer. However, further experimental validation is needed to confirm these findings.

Keywords:

Bioinformatics
Breast cancer
Metabolic genes
Survival analysis
Prognosis

Online publication: June 28, 2024

1. Introduction

Breast cancer (BC) is one of the most common malignancies among women. In China, the incidence of this disease accounts for 7% to 10% of the total number of malignant tumors, with 169,000 new cases of female breast cancer reported each year, representing 12.25% of the global total incidence [1,2]. Currently, the exact pathogenesis of breast cancer is not fully understood, but it is generally believed to be primarily related to genetic factors, gene mutations, decreased immune function, and neurological abnormalities in the medical community [3,4]. Regarding the survival and prognostic status of breast cancer, studies [5] have shown that for patients aged less than or equal to 35 years old, the 5-year survival rates without distant metastasis and the overall 5-year survival rates are 83.4% and 88.1%, respectively. For patients aged

greater than or equal to 35 years old, these rates are 88.3% and 88.2%, respectively. The prognosis of breast cancer is influenced by a combination of factors including patient-specific factors (such as tumor size, lymph node metastasis, pathological grading), environmental factors (such as poor dietary and lifestyle habits, psychosocial factors), genetic factors (such as family heredity, immune deficiencies), and molecular biology factors such as estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER-2), nucleus-related antigen Ki-67 (Ki-67), and tumor protein P53 [6-8].

Bioinformatics is a science that utilizes computers as tools to store, retrieve, and analyze biological information in life science research [9,10]. As one of the current focal points in natural science research, bioinformatics encompasses emerging and important research areas such as gene expression profiling, metabolic network analysis, gene chip design, and proteomics data analysis [11]. Therefore, it is significant to utilize bioinformatics methods to identify key metabolic genes in breast cancer and conduct survival prognosis analysis, providing valuable insights and a basis for breast cancer research.

2. Materials and methods

2.1. Data acquisition

2.1.1. Acquisition of breast cancer transcriptome data

The transcriptome data for breast cancer was retrieved from the TCGA database (https://portal.gdc.cancer.gov/). In the TCGA database, case screening was performed first, selecting “breast” as the tissue type in the Case ID option, “TCGA” as the study type in the Program option, and “TCGA-BRCA” in the Project option. The remaining options were set to the TCGA database defaults. Next, file screening was conducted, selecting “transcriptome profiling” in the Data Category option, “Gene Expression Quantification” in the Data Type option, and “HTSeq-FPKM” in the Workflow Type option. After selecting the data as described above, it was downloaded to obtain the transcriptome data for breast cancer.

2.1.2. Acquisition of clinical data for breast cancer

The acquisition of clinical data for breast cancer also relied on the TCGA database for screening. Following the steps outlined in “1.1.1,” we first performed case screening in the TCGA database. In the Case ID option, we selected “breast” as the tissue type, “TCGA” as the research type in the Program option, and “TCGA-BRCA” in the Project option. The remaining options were set to the TCGA database defaults. Next, we proceeded to file screening, selecting “clinical” as the Data Category and “bcr xml” as the Data Format in the File options. After making these selections, the clinical data for breast cancer were downloaded.

2.1.3. Acquisition of expression data for breast cancer metabolism-related genes

Firstly, we searched for relevant metabolic genes through the GSEA database (http://software.broadinstitute.org/gsea/index.jsp), downloaded the “KEGG gene sets, gene symbols” file, and matched it with breast cancer genes obtained from the TCGA database. Finally, we obtained the metabolic genes of breast cancer.

2.2. Data processing

We used Perl to process the acquired breast cancer transcriptome data and clinical data using relevant script files such as merge.pl, symbol.pl, and getClinical.pl. This allowed us to obtain a transcriptome data matrix and a clinical data matrix. We combined the metabolic genes obtained from the GSEA database with the relevant genes from TCGA breast cancer and used relevant script files to extract the expression levels of metabolic genes in breast cancer. Differential analysis was performed using R×64 3.6.1 software. After acquiring differential genes, we combined the TCGA metabolic genes with survival time using relevant script files and organized clinical data. We then used R×64 3.6.1 software again for prognostic-related metabolic gene screening, Lasso regression model construction, survival analysis, risk curve plotting, and independent prognostic analysis.

3. Results

3.1. Breast cancer transcriptome data processing results

After processing with Perl, we obtained breast cancer transcriptome data consisting of 1,222 samples, including 113 normal samples and 1,109 tumor samples. After converting gene IDs, we obtained 56,753 relevant genes

(**Table 1**).

3.2. Breast cancer clinical data processing results

After downloading the breast cancer clinical data from the TCGA database, we organized the data using relevant Perl scripts, resulting in a total of 443 clinical data points. The results are shown in **Table 2**.

3.3. Breast cancer metabolic gene expression processing results

By processing the metabolic genes obtained from the GSEA database and the breast cancer genes obtained from the TCGA database using relevant script files, we obtained the expression levels of metabolic genes in breast cancer, with a total of 944 data points (**Table 3**). Differential analysis was performed using R×64 3.6.1 software, and the results are presented in **Table 4** and **Figure 1**.

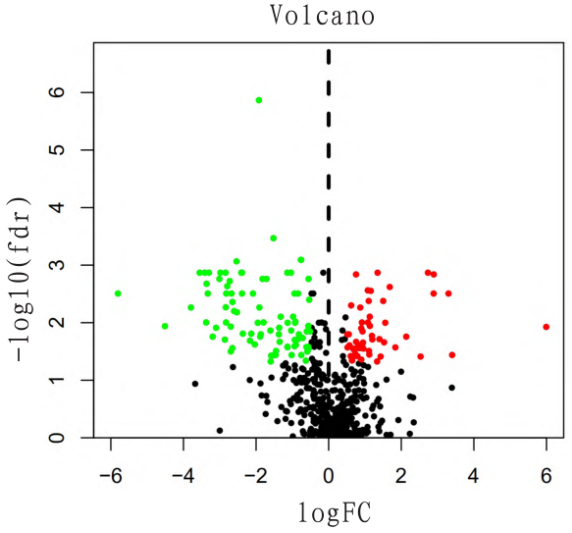

Figure 1. Volcano plot of differential analysis of metabolic genes. Note: In the figure, black represents no difference in gene expression between normal and tumor samples, green represents downregulation of genes in tumor samples, and red represents upregulation of genes in tumor samples.

3.4. Merging results of TCGA metabolic genes and survival time

Firstly, the clinical data of breast cancer was organized [12], removing samples with missing survival time (futime) or survival time (futime) less than 30 days. Samples with missing survival status (fustat) were also excluded. The differential metabolic genes obtained from **3.3.** were then merged with survival time using relevant script files, resulting in 119 relevant genes (**Appendix 1**).

3.5. Lasso model construction results

The prognostic-related metabolic genes obtained were used to construct a Lasso model using the glmnet and survival packages in R×64 3.6.1 software [13,14]. The risk score was calculated using the formula: risk score = (Coefficient-$Gene_1$ × expression of $Gene_1$) + (Coefficient-$Gene_2$ × expression of $Gene_2$) + ⋯ + (Coefficient-$Gene_n$ × expression of $Gene_n$) (**Appendix 2**).

3.6. Screening results of prognostic-related metabolic genes

The differential metabolic genes obtained from **3.4.** were screened using univariate Cox regression with a *P*-value < 0.05. A hazard ratio (HR) value greater than 1 indicates that the gene is a high-risk gene, and $P < 0.05$ indicates that the gene is associated with prognosis. The prognostic-related metabolic genes were screened using the survival package in R×64 3.6.1 software [15,16]. The results are presented in **Table 5** and **Figure 2**.

Table 5. Screening results of prognostic-related metabolic genes

Gene names	HR	HR.95L	HR.95H	*P*
POLR2K	1.01	1.00	1.02	0.01
NMNAT2	1.08	1.01	1.16	0.03
SUCLA2	1.05	1.01	1.09	0.01

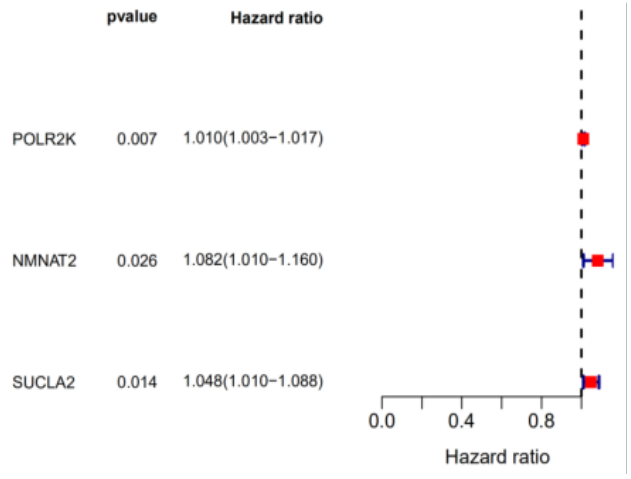

Figure 2. Forest plot of prognostic-related metabolic genes

Table 1. Relevant genes in breast cancer transcriptome samples

Gene name	TCGA-HU-A4GH-11A-11R-A36D-31	TCGA-BR-6454-11A-01R-1802-13	TCGA-CG-5730-11A-01R-1602-13	TCGA-HU-A4GN-11A-12R-A251-31	TCGA-HU-A4GY-11A-11R-A36D-31
TTYH3	5.74	16.94	22.84	4.88	4.80
LRRC34	0.64	0.13	0.14	0.51	0.79
FSTL4	0.04	0.11	0.04	0.01	0.06
KCNK2	0.68	0.28	0.08	0.30	1.16
DNLZ	0.13	0.09	0.05	0.13	0.07
DCLRE1C	1.34	2.12	0.95	1.60	1.72
TRDN	0.05	0.00	0.00	0.04	0.05
COX5BP8	0.51	0.00	0.00	0.00	0.20
GNG5P1	0.00	0.00	0.00	0.00	0.00
PLSCR1	10.56	14.46	9.24	10.92	11.58
GNG5P5	0.00	0.10	0.00	0.00	0.00

Note: The rows in the table represent sample names; within the sample names, they are split by "-," where the fourth digit starting with 1 represents para-cancerous or normal samples, and those starting with 0 represent tumor samples; the columns represent gene names; the numbers represent the expression level of the gene in that sample, similarly hereinafter.

Table 2. Breast cancer clinical data

Case IDs	Survival time (d)	Survival status	Age (years)	Gender	Grade	Stage	T	M	N
TCGA-VQ-A8PS	406	1	76	Male	G2	IIIA	T3	M0	N1
TCGA-CG-4301	92	0	75	Female	G3	IV	T4	M1	N1
TCGA-R5-A7ZR	185	1	70	Female	G2	III	T3	M0	NX
TCGA-VQ-AA64	560	1	68	Male	G2	IIIB	T3	M0	N2
TCGA-BR-8686	477	0	69	Male	G3	IIIB	T4b	M0	N1
TCGA-B7-A5TJ	335	0	74	Male	G1	IIB	T4a	M0	NX
TCGA-BR-A4QL	352	0	75	Female	G2	IIIB	T3	M0	N3a
TCGA-D7-6817	389	0	63	Male	G3	IIIA	T2b	M0	N3
TCGA-MX-A5UG	113	1	78	Male	G3	IIIA	T3	M0	N1
TCGA-IN-A6RJ	379	0	64	Male	G3	IA	T1b	M0	N0
TCGA-CG-4475	699	0	76	Male	G3	IIB	T3	M0	N1
TCGA-BR-8591	409	0	79	Male	G3	IIIC	T4a	M0	N3a
TCGA-BR-8676	229	0	59	Male	G3	IIIB	T3	M0	N3a
TCGA-D7-A4YV	180	0	69	Female	G3	IIB	T3	M0	N1
TCGA-BR-8285	17	1	57	Female	G3	IIIC	T4a	M0	N3a
TCGA-CD-5798	86	0	82	Male	G2	II	T3	M0	N0
TCGA-FP-A4BF	168	1	68	Male	G3	IIIA	T3	M0	N2

Note: In survival status, 0 represents death, and 1 represents alive.

Table 3. Breast cancer metabolic gene expression levels

Gene names	TCGA-BH-A0AU-11A-11R-A12P-07	TCGA-BH-A208-11A-51R-A157-07	TCGA-BH-A0H7-11A-13R-A089-07	TCGA-BH-A209-11A-42R-A157-07	TCGA-BH-A0AZ-11A-22R-A12P-07	TCGA-BH-A0B5-11A-23R-A12P-07	TCGA-BH-A1FR-11B-42R-A13Q-07	TCGA-E9-A1ND-11A-43R-A144-07	TCGA-GI-A2C9-11A-22R-A21T-07
ACSM2A	0.01	0.01	0.02	0.01	0.00	0.08	0.03	0.01	0.03
PHGDH	19.44	19.24	27.43	12.11	12.63	22.25	10.88	20.97	12.95
ALDH1A1	19.54	22.09	24.01	42.27	23.52	63.85	44.21	64.19	46.02
HAGHL	0.34	0.29	1.10	0.14	0.31	0.04	0.33	0.67	0.24
GYS2	0.06	0.44	0.50	0.14	0.03	5.29	0.07	0.15	0.06
HMGCS1	10.97	14.02	15.62	10.20	12.58	10.39	25.80	4.51	10.27
GUCY1B1	3.85	3.71	3.31	6.90	4.92	3.46	9.06	3.23	6.36
CYP3A7	0.14	0.11	0.08	0.16	0.14	0.04	0.12	0.11	0.07
PDE4B	7.70	17.06	7.53	6.10	9.02	0.44	7.92	2.90	14.33
ACADVL	79.51	63.54	45.37	50.67	63.26	91.12	61.85	111.95	74.71
MARS2	4.08	3.77	3.67	3.49	2.93	1.21	3.61	1.73	3.67
CAT	46.05	54.39	47.72	101.72	39.79	184.48	51.54	85.97	61.86
GSTM3	17.96	9.92	34.01	24.72	15.66	10.87	13.98	12.52	21.26
POLR2D	7.02	6.99	5.50	7.34	6.56	4.22	8.22	5.41	6.45
APIP	4.01	4.32	4.06	2.83	3.70	7.87	4.36	3.95	2.98
ACSS2	16.25	15.54	19.83	17.64	9.75	97.55	14.47	22.78	10.76
DGKA	2.36	2.29	2.01	2.87	2.97	0.53	2.74	2.44	3.19
ADSL	9.72	10.89	10.26	8.48	7.78	6.16	7.29	9.22	7.72
MARS	9.41	9.97	10.57	11.68	8.90	8.24	10.75	10.11	8.86

Table 4. Results of differential analysis of metabolic genes

Gene names	Con mean	Treat mean	Log FC	*P*	fdr
ALDH1A1	44.09	10.76	-2.03	0.01	0.02
GYS2	1.14	0.13	-3.20	0.00	0.02
CAT	94.95	28.95	-1.71	0.00	0.00
ACSS2	25.83	9.00	-1.52	0.00	0.00
MARS	8.81	13.80	0.65	0.01	0.05
ASPA	2.57	0.40	-2.70	0.00	0.01
ADCY5	4.61	2.40	-0.94	0.00	0.00
PFKFB3	88.49	25.53	-1.79	0.00	0.01
NEU1	11.74	20.44	0.80	0.01	0.03
CYB5R3	51.63	31.40	-0.72	0.00	0.02
POLE2	0.74	2.17	1.56	0.00	0.01
NPR1	22.94	3.20	-2.84	0.00	0.00
PMM2	1.44	2.69	0.90	0.01	0.04
PRIM2	1.92	3.67	0.93	0.00	0.01
ADH1B	182.75	17.76	-3.36	0.00	0.00
PDE1B	2.68	0.93	-1.53	0.01	0.04
B4GALT1	42.88	62.97	0.55	0.00	0.02
POLR2J	12.49	23.26	0.90	0.00	0.01
ACADS	16.55	7.51	-1.14	0.00	0.00

3.7. Survival analysis results

Based on the risk scores obtained from **3.5.**, patients were stratified into high- and low-risk groups using the median risk score for survival analysis. A *P*-value < 0.05 indicates a significant difference in survival between the high- and low-risk groups [17]. Survival curves were plotted using the survival and survminer packages in R×64 3.6.1 software (**Figure 3**).

3.8. Risk analysis results

Based on the survival risk curve obtained from section **3.5.**, the risk analysis was plotted using the pheatmap package in R×64 3.6.1 software [18]. The results showed that as the risk value increased, the number of patients who died gradually increased (**Figures 4–6**).

3.9. Independent prognostic analysis results

Based on the risk analysis results obtained from section **3.7.**, univariate and multivariate independent prognostic analyses were performed using the survival package in R×64 3.6.1 software. A risk score with $P < 0.001$ indicates that it can be used as an independent prognostic factor (**Figures 7** and **8**).

4. Discussion and conclusion

In this study, we identified three metabolism-related genes associated with breast cancer: *POLR2K*, *NMNAT2*, and *SUCLA2*. *POLR2K* refers to RNA polymerase II subunit K, a protein-coding gene involved in pathways such as RNA polymerase I promoter escape and RNA polymerase III transcription initiation. *POLR2K* encodes the smallest subunit of RNA polymerase II, which is a common

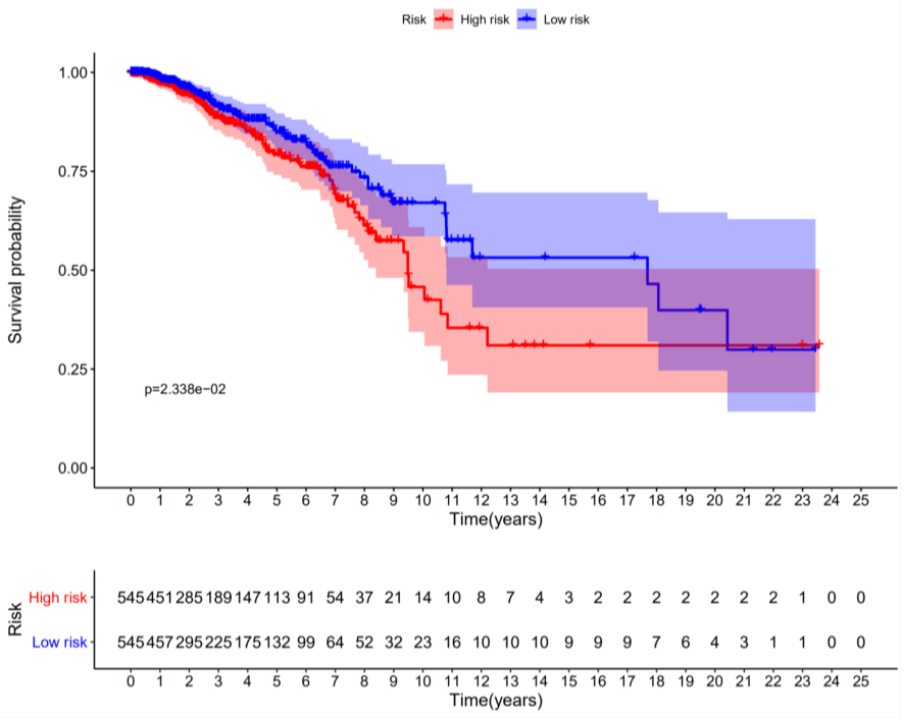

Figure 3. Survival curve diagram

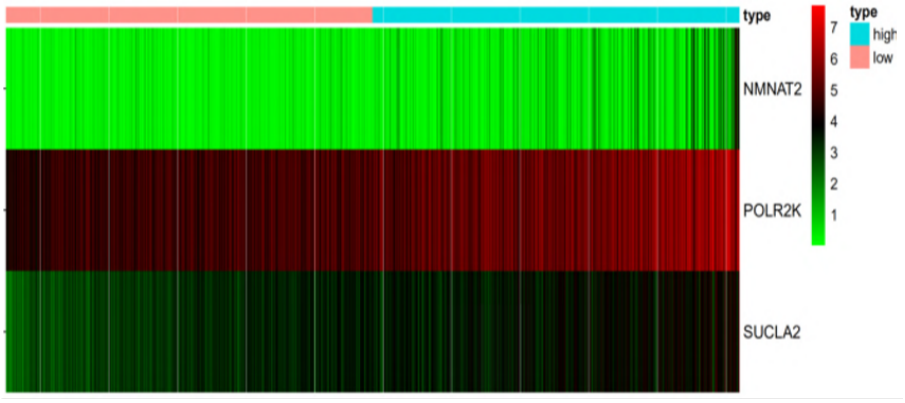

Figure 4. Risk heatmap. Note: In the figure, red represents high expression of the gene; green represents low expression of the gene.

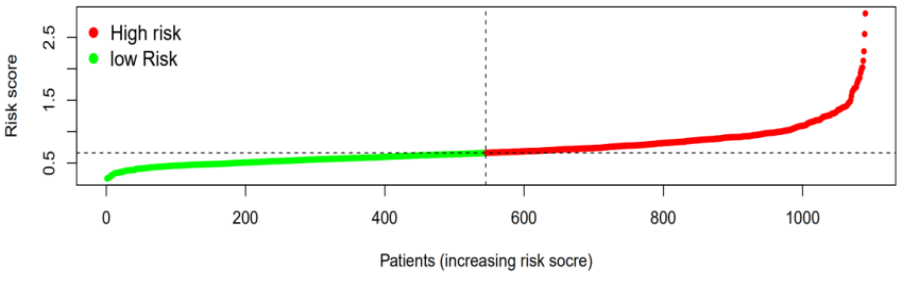

Figure 5. Risk curve graph

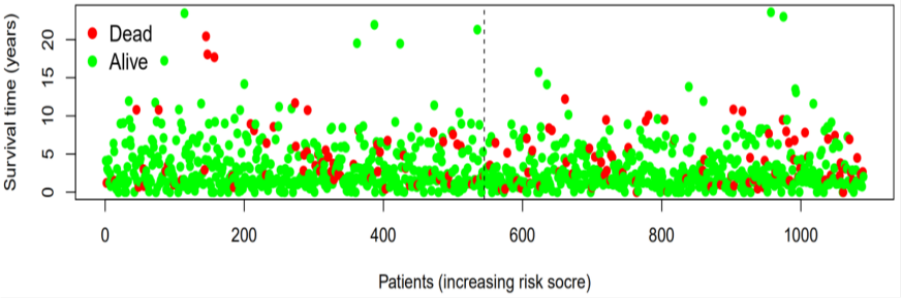

Figure 6. Survival status graph. Note: In the figure, green represents that the patient is alive; red represents that the patient has died.

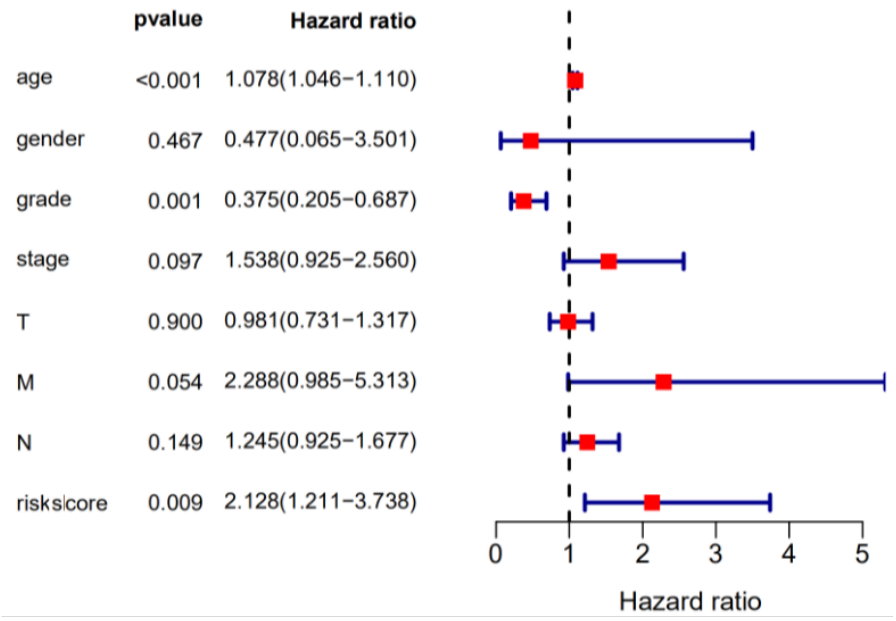

Figure 7. Forest plot of multivariate independent prognostic analysis

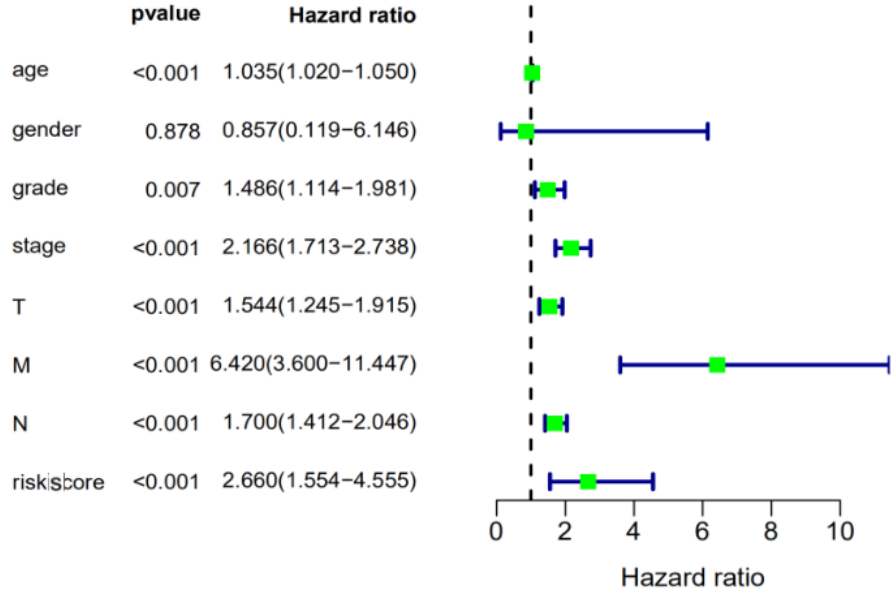

Figure 8. Forest plot of univariate independent prognostic analysis

subunit of three RNA polymerases [19]. Studies have speculated that the upregulation of *POLR2K* may promote the assembly of polymerase III, thereby facilitating cell proliferation and cancer development, and it can serve as a risk factor for breast cancer [20-22].

NMNAT2 stands for nicotinamide nucleotide adenylyltransferase 2, a protein-coding gene whose primary function is to catalyze the synthesis of nicotinamide adenine dinucleotide from β-nicotinamide mononucleotide and adenosine-triphosphate (ATP) [23,24]. Research has indicated that the expression of *NMNAT2* is associated with the depth of tumor invasion and tumor staging (tumor node metastasis, TNM) [25].

SUCLA2 is a protein-coding gene that plays a crucial role in the citric acid cycle. SCS-A hydrolyzes ATP, converting succinate into succinyl-CoA [26]. Current studies have found that mutations in this gene are the main cause of mitochondrial DNA depletion syndrome, primarily affecting the brain and skeletal muscles [27]. Regrettably, there is currently no relevant literature reporting its association with cancer or breast cancer, providing a direction for our future research.

In the survival analysis (**Figure 3**), 545 samples were included in the study. The results revealed that as time (years) progressed, the number of surviving patients gradually decreased, with a maximum survival time of 24 years for both the high-risk and low-risk groups. In the survival risk heatmap (**Figure 5**), we observed that the *POLR2K* gene was the most significantly overexpressed gene, suggesting a close association between this gene and the survival prognosis of breast cancer.

In the independent prognostic factor analysis, univariate analysis (**Figure 8**) identified age, stage, and TNM staging as independent prognostic factors. In the multivariate analysis, age was closely related to the multivariate prognosis.

Although we have used bioinformatics techniques to identify potential metabolic genes that may affect the prognosis of breast cancer in a large sample, there are still some limitations to this study. Firstly, some clinical follow-up information was missing for certain samples. Secondly, relying solely on bioinformatics analysis is insufficient, and experimental validation is necessary to confirm the results. Therefore, further genetic and experimental studies on larger samples, along with experimental validation, are required. In conclusion, this study employed bioinformatics to explore relevant metabolic genes in breast cancer and conducted a survival prognosis analysis of key metabolic genes, providing a reference method and approach for breast cancer treatment research.

Funding

National Natural Science Foundation of China (81973976); Guangxi Zhuang and Yao Medicine and Medical Care Combined Talent Small Highland Project (Guangxi Party Committee Office [2017] No. 44); The Sixth Batch of National Traditional Chinese Medicine Expert Academic Experience Inheritance Project of the National Administration of Traditional Chinese Medicine (National Traditional Chinese Medicine Human Resources and Education Development [2017] No. 29).

Disclosure statement

The authors declare no conflict of interest.

References

[1] Chen X, Wang J, 2013, Surgery, People's Health Publishing House, Beijing, 256.

[2] Bai S, 2017, Epidemiological Status and Disease Characteristics of Breast Cancer in China. World Latest Medicine Information, 17(41): 253–256.

[3] Li W, Zhao T, Jiang X, et al., 2022, Visual Analysis of Knowledge Map of Traditional Chinese Medicine Treatment of Breast Cancer Based on CNKI. Western Journal of Traditional Chinese Medicine, 35(3): 79–83.

[4] Ge J, Jia Y, Li J, et al., 2020, Discussion on the Pathway Related to the Onset of Breast Cancer Based on Enrichment Analysis. Systems Medicine, 5(11): 29–33.

[5] Kan B, Sun M, Yan G, et al., 2016, Clinicopathological Features and Prognostic Analysis of Breast-Conserving Surgery Patients with Breast Cancer in Different Age Groups. Progress in Modern Biomedicine, 16(34): 6733–6737.

[6] Song C, Meng Y, Shang P, et al., 2018, Related Influencing Factors of the Occurrence, Development, and Prognosis of Breast Cancer. Journal of Hebei North University (Natural Science Edition), 34(8): 54–56.

[7] Yu C, Pei J, 2020, Prediction of the Proportion of Infiltrating Immune Cells in Breast Cancer and Prognostic Analysis of Subtypes Based on CIBERSORT. Journal of Yunnan Minzu University (Natural Science Edition), 29(1): 78–83.

[8] Duan Y, 2013, Clinicopathological Features and Prognostic Analysis of Different Molecular Subtypes of Breast Cancer, dissertation, Ningxia Medical University.

[9] Wu X, Zhao M, 2005, Bioinformatics - A Guide to Basic and Clinical Medical Applications, Science Press, Beijing, 3–5.

[10] Chu H, 2018, Application of Data Mining in Bioinformatics. Digital Technology and Application, 36(10): 123–124.

[11] Li J, Huo F, Li Z, et al., 2023, Exploring the Mechanism of Qianghuoshenshi Decoction in the Treatment of Ankylosing Spondylitis Based on Bioinformatics and Molecular Docking Technology. Western Journal of Traditional Chinese Medicine, 36(2): 72–79.

[12] Liu GM, Xie WX, Zhang CY, et al., 2020, Identification of a Four-Gene Metabolic Signature Predicting Overall Survival for Hepatocellular Carcinoma. Cell Physiol, 235(2): 1624–1636.

[13] Li Z, Li F, Peng Y, et al., 2020, Identification of Three m6A-Related mRNAs Signature and Risk Score for the Prognostication of Hepatocellular Carcinoma. Cancer Med, 9(5): 1877–1889.

[14] Zhang M, Zhu K, Pu H, et al., 2019, An Immune-Related Signature Predicts Survival in Patients with Lung Adenocarcinoma. Front Oncol, 10(9): 1314–1327.

[15] Zhang C, Wang F, Guo F, et al., 2020, A 13-Gene Risk Score System and a Nomogram Survival Model for Predicting the Prognosis of Clear Cell Renal Cell Carcinoma. Urol Oncol, 38(3): 71–74.

[16] Ma N, Tie C, Yu B, et al., 2020, Identifying lncRNA-miRNA-mRNA Networks to Investigate Alzheimer's Disease Pathogenesis and Therapy Strategy. Aging (Albany NY), 12(3): 2897–2920.

[17] George B, Seals S, Aban I, 2014, Survival Analysis and Regression Models. J Nucl Cardiol, 21(4): 686–694.

[18] Wang J, Chen X, Tian Y, et al., 2020, Six-Gene Signature for Predicting Survival in Patients with Head and Neck Squamous Cell Carcinoma. Aging (Albany NY), 12(1): 767–783.

[19] Jung CR, Choi S, Im DS, 2007, The NS5A Protein of Hepatitis C Virus Represses Gene Expression of hRPB10alpha, a Common Subunit of Host RNA Polymerases, Through Interferon Regulatory Factor-1 Binding Site. Virus Res, 129(2): 155–165.

[20] Lin Y, Li Z, Ozsolak F, et al., 2012, An In-Depth Map of Polyadenylation Sites in Cancer. Nucleic Acids Res, 40(17): 8460–8471.

[21] Liu Y, Zhu X, Zhu J, et al., 2007, Identification of Differential Expression of Genes in Hepatocellular Carcinoma by Suppression Subtractive Hybridization Combined cDNA Microarray. Oncol Rep, 18(4): 943–951.

[22] Kontorovich T, Levy A, Korostishevsky M, et al., 2010, Single Nucleotide Polymorphisms in miRNA Binding Sites and miRNA Genes as Breast/Ovarian Cancer Risk Modifiers in Jewish High-Risk Women. Int J Cancer, 127(3): 589–597.

[23] Pottorf T, Mann A, Fross S, et al., 2018, Nicotinamide Mononucleotide Adenylyltransferase 2 Maintains Neuronal Structural Integrity Through the Maintenance of Golgi Structure. Neurochem Int, 12(1): 86–97.

[24] Mayer PR, Huang N, Dewey CM, et al., 2010, Expression, Localization, and Biochemical Characterization of Nicotinamide Mononucleotide Adenylyltransferase 2. J Biol Chem, 285(51): 40387–40396.

[25] Cui C, Qi J, Deng Q, et al., 2016, Nicotinamide Mononucleotide Adenylyl Transferase 2: A Promising Diagnostic and Therapeutic Target for Colorectal Cancer. Biomed Res Int, 29(8): 180–196.

[26] Zhao Y, Tian J, Sui S, et al., 2017, Loss of succinyl-CoA Synthase ADP-Forming β Subunit Disrupts mtDNA Stability and Mitochondrial Dynamics in Neurons. Sci Rep, 7(1): 7169–7177.

[27] Miller C, Wang L, Ostergaard E, et al., 2011, The Interplay Between SUCLA2, SUCLG2, and Mitochondrial DNA Depletion. Biochim Biophys Acta, 1812(5): 625–629.

Appendix

Appendix 1. Combined results of differential metabolic genes and survival time

Case IDs	Survival time (d)	Survival status	ALDH1A1	GYS2	CAT	ACSS2	MARS	ASPA	ADCY5
TCGA-C8-A12X	385.00	0	2.99	0.05	10.99	6.88	18.17	0.07	0.34
TCGA-AC-A2BK	2222.00	0	2.10	0.08	230.19	9.04	18.43	0.04	0.63
TCGA-AR-A24S	2976.00	0	10.57	0.05	50.49	6.25	18.20	0.40	0.27
TCGA-AR-A24K	1548.00	0	1.09	0.05	35.86	4.36	8.97	0.10	0.06
TCGA-WT-AB44	883.00	0	8.13	0.04	12.73	4.49	11.04	0.38	0.94
TCGA-AR-A24P	84.00	0	8.01	0.05	28.63	5.99	9.33	0.14	0.31
TCGA-BH-A5IZ	567.00	0	1.94	0.06	18.83	4.11	11.41	0.07	0.06
TCGA-A8-A095	1277.00	0	5.10	0.05	36.28	5.72	10.52	0.28	0.58
TCGA-BH-A0H5	1620.00	0	19.85	0.06	34.13	8.73	14.24	0.46	7.58
TCGA-E9-A24A	747.00	0	3.29	0.06	23.98	4.48	10.52	0.13	0.07
TCGA-B6-A0X4	860.00	1	0.03	0.04	14.52	7.90	11.23	0.08	0.13
TCGA-E2-A159	762.00	0	8.98	0.06	16.34	21.78	14.45	0.06	0.01
TCGA-AC-A3HN	496.00	0	14.86	0.04	34.11	7.63	12.18	0.65	5.91
TCGA-D8-A1Y1	302.00	1	6.15	0.04	17.24	7.09	14.69	0.21	0.09
TCGA-AR-A5QP	622.00	0	14.15	0.05	20.04	7.23	10.96	0.40	3.19
TCGA-OL-A66P	428.00	0	22.52	0.06	15.70	13.60	13.65	0.17	0.11
TCGA-OL-A66K	1275.00	1	12.31	0.04	21.26	9.79	8.15	0.22	0.49
TCGA-AQ-A1H2	475.00	0	1.77	0.05	63.00	17.27	15.38	0.08	5.16
TCGA-AC-A3W5	504.00	0	15.17	0.07	13.62	10.21	14.48	0.28	0.81

Note: In survival status, 0 represents alive, and 1 represents death.

Appendix 2. Lasso model construction results

Gene names	Risk score	Risk	Gene names	Risk score	Risk
TCGA-C8-A12X	0.73	high	TCGA-A2-A0YT	1.09	high
TCGA-AC-A2BK	0.71	high	TCGA-A8-A07U	0.55	low
TCGA-AR-A24S	0.76	high	TCGA-AR-A1AN	0.64	low
TCGA-AR-A24K	0.69	high	TCGA-B6-A1KC	1.25	high
TCGA-WT-AB44	0.30	low	TCGA-BH-A0AV	0.47	low
TCGA-AR-A24P	0.67	high	TCGA-D8-A1JB	1.04	high
TCGA-BH-A5IZ	0.68	high	TCGA-A2-A0T4	0.63	low
TCGA-A8-A095	0.60	low	TCGA-C8-A1HN	1.23	high
TCGA-BH-A0H5	0.56	low	TCGA-AR-A1AQ	0.96	high
TCGA-E9-A24A	0.82	high	TCGA-AO-A12C	0.63	low
TCGA-B6-A0X4	2.02	high	TCGA-EW-A2FS	0.72	high
TCGA-E2-A159	0.56	low	TCGA-AO-A03U	0.55	low

Appendix 2. (Continued)

Gene names	Risk score	Risk	Gene names	Risk score	Risk
TCGA-AC-A3HN	0.57	low	TCGA-AN-A0AR	0.38	low
TCGA-D8-A1Y1	0.70	high	TCGA-B6-A0IC	0.78	high
TCGA-AR-A5QP	0.79	high	TCGA-C8-A131	0.98	high
TCGA-OL-A66P	0.38	low	TCGA-E9-A3HO	0.47	low
TCGA-OL-A66K	0.45	low	TCGA-E9-A1N3	1.79	high
TCGA-AQ-A1H2	0.80	high	TCGA-B6-A0RI	0.58	low
TCGA-AC-A3W5	0.41	low	TCGA-E9-A1RC	0.51	low

Note: high indicates high risk, low indicates low risk