HUMAN BIOLOGY
2025, VOL. 95, ISSUE 3
ORIGINAL ARTICLE

OPEN ACCESS

Nomogram For Predicting Survival in Breast Cancer Patients Based on Lncrna Expression and Clinical Characteristics

Alon Ben-Arie[1], Hagit Deutsch[2], Jolene Seibe[1]-Seamon[2], John F Visintine[1], Benjamin E Leiby[1]*

ABSTRACT

Objective: Current prognostic models for breast cancer (BC) are largely dependent on clinical factors and immunohistochemical markers, or in some cases, a limited number of gene signatures. These approaches have certain limitations in accuracy and clinical applicability. This study aimed to construct a predictive nomogram that integrates molecular signatures of long non-coding RNAs (lncRNAs) with conventional clinical factors, thereby providing a more comprehensive and individualized tool for survival prediction in BC patients.

Methods: Using The Cancer Genome Atlas (TCGA) database, RNA-sequencing data and clinical information of breast cancer patients were retrieved. Differentially expressed genes were identified with the DESeq2 R package, followed by univariate and multivariate Cox regression analyses to identify prognostic lncRNA biomarkers. A 9-lncRNA risk score model was then established and validated. Independent prognostic factors were further integrated with clinical variables, and a predictive nomogram was constructed. Model performance was evaluated using the concordance index (C-index), Kaplan–Meier survival analysis, ROC curves, and calibration plots.

Results: A total of 1208 transcriptome profiles were analysed, including 1096 breast cancer and 112 normal tissue samples. From these, 2100 differentially expressed genes were identified. Nine lncRNAs (AC068858.1, AC000067.1, LINC00460, LINC02408, AC136475.5, AC023043.4, AC073359.1, AC244502.1, and COL4A2-AS1) were significantly associated with overall survival (OS). Four acted as risk factors (HR > 1), whereas five served as protective factors (HR < 1). The 9-lncRNA signature stratified patients into high- and low-risk groups with significant prognostic differences ($p < 0.001$). Time-dependent ROC curves demonstrated strong predictive accuracy, with AUC values ranging from 0.72–0.92 across different datasets and follow-up periods. Multivariate Cox analysis confirmed that age and the lncRNA model were independent prognostic predictors. A nomogram combining these two factors was constructed, achieving a C-index of 0.81 and demonstrating excellent calibration for 1-, 3-, and 5-year OS predictions.

Conclusion: The 9-lncRNA-based prognostic model, integrated with clinical risk factors such as age, provides a robust and individualized tool for predicting breast cancer survival. This nomogram may serve as a valuable reference for clinical decision-making and personalized management strategies in breast cancer patients.

INTRODUCTION

Breast cancer (BC) is the most common malignancy among women worldwide, with rising incidence rates in many regions. In China alone, approximately 416,000 new cases and 120,000 deaths occur annually. ALLEMANI C et al. (2018). Despite advances in early detection and treatment, 30–40% of patients with early-stage BC eventually progress to advanced disease, with a median survival of only three years and markedly reduced long-term survival. Given the heterogeneous nature of BC—characterized by substantial molecular and clinical variability—accurate prognostic prediction remains a major clinical challenge. Yue Gong et al. (2021). Traditionally, prognostic models have relied on clinical and pathological variables, including tumor stage,

[1]Department of Obstetrics and Gynecology, Division of Maternal–Fetal Medicine, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania
[2]Maternal–Fetal Medicine Division, Department of Obstetrics and Gynecology, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania

Correspondence to: Prof. Hagit Deutsch, Maternal–Fetal Medicine Division, Department of Obstetrics and Gynecology, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania, Email: hagit.deutsch@jefferson.edu

Keywords: Breast Cancer; Long Non-Coding RNA; Prognosis; Nomogram; Survival

histological subtype, hormone receptor status, and HER2 expression. Lang GT et al. (2020)- Rinn JL et al. (2012). Although these models have improved treatment stratification, they fail to capture the molecular complexity underlying tumor progression. More advanced gene expression–based models, such as the 70-gene MammaPrint and 21-gene Oncotype DX signatures, have provided important insights but are limited by platform dependence, high cost, and restricted clinical utility in diverse populations.

Long non-coding RNAs (lncRNAs)—defined as transcripts longer than 200 nucleotides without protein-coding capacity—have emerged as crucial regulators of gene expression, chromatin remodeling, and tumor biology. Mounting evidence shows that lncRNAs are involved in cell proliferation, apoptosis, invasion, and metastasis, and their dysregulation is closely linked to cancer prognosis. Morris K Vet al. (2014). Unlike protein-coding genes, lncRNAs exhibit tissue- and cancer-type specificity, making them attractive biomarkers for prognosis prediction. Yan X et al. (2015).

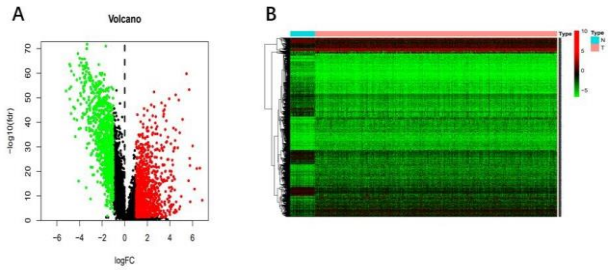
Given these considerations, this study aimed to identify lncRNAs associated with breast cancer survival using large-scale TCGA datasets and to integrate these molecular biomarkers with clinical risk factors to construct a predictive nomogram. Miyamoto et al, (2018) - Quinn JJ et al. (2014). Such a model could provide clinicians with a practical and individualized tool for risk assessment and treatment planning.

MATERIALS AND METHODS

Data acquisition

RNA-sequencing data (lncRNA and mRNAexpression profiles) and clinical information for 1096. BC patients and 112 normal controls were downloaded from the TCGA-BRCA project Engreitz JM et al. (2016)- Guo et al. (2016). (https://portal.gdc.cancer)

Figure 1: Differential expression analysis of lncRNAs in breast cancer.



(A) Volcano plot displaying differentially expressed lncRNAs between breast cancer and normal tissue samples.

(B) Heatmap showing the expression patterns of significantly differentially expressed lncRNAs in breast cancer patients.

Differential expression analysis

The DESeq2 R package was used to identify differentially expressed lncRNAs (DELs) and mRNAs (DEMs) between tumor and normal samples. Lin TY et al. (2016), Beermann J et al. (2016), Pandey GK et al. (2014). Thresholds were set at $|\log2FC| > 1$ and adjusted $p < 0.05$. Results were visualized using volcano plots (ggplot2) and heatmaps (pheatmap).

Prognostic lncRNA identification and risk score model

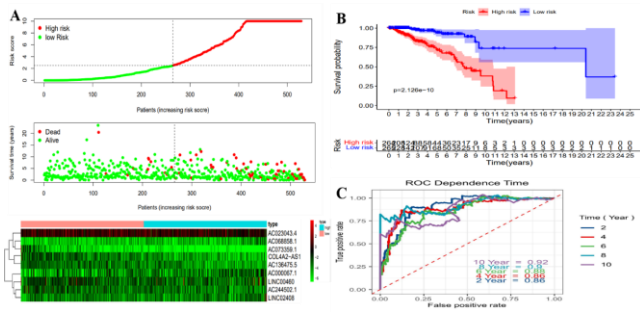
Univariate Cox regression was performed to identify lncRNAs significantly associated with overall survival (OS). Boon RA et al. (2016)- Qu L et al. (2018).

Table 1: Clinical characteristics of breast cancer patients in the TCGA cohort

Variable	Primary Dataset (n=524)	Entire Dataset (n=1053)	p-value
Age (<60 / ≥60)	285 / 239	573 / 480	0.64
Stage (I–II / III–IV)	388 / 136	781 / 272	0.57
T stage (T1–T2 / T3–T4)	421 / 103	846 / 207	0.42
N stage (N0–N1 / N2 N3)	352 / 172	691 / 362	0.39
M stage (M0 / M1)	501 / 23	1008 / 45	0.71

Candidate lncRNAs were further validated using multivariate Cox analysis. White NM, et al. (2017), Simpson PT et al. (2010), Park YH et al. (2011). A prognostic risk score was then calculated for each patient as a weighted sum of expression values multiplied by their corresponding Cox regression coefficients. Rakha EA, et al. (2010), Chowdhury N et al. (2006). Patients were stratified into high- and low-risk groups based on the median risk score.

Figure 2: Validation of the 9-lncRNA signature in the primary dataset.



(A) Distribution of risk scores, overall survival status, and heatmap of the 9-lncRNA expression profiles. (B) Kaplan–Meier survival curves comparing high-risk and low-risk groups. (C) Time-dependent ROC curves for predicting 2-, 4-, 6-, 8-, and 10-year overall survival.

Nomogram construction

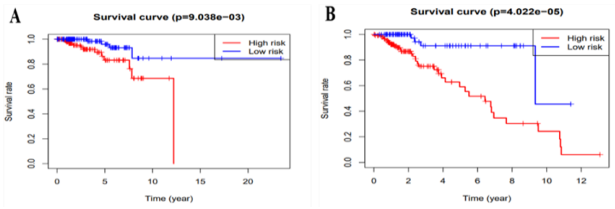
Independent prognostic factors were identified by multivariate Cox regression incorporating both clinical variables and the lncRNA risk score. Nguyen—Ngoc KV et al. (2012) A predictive nomogram was constructed to estimate 1-, 3-, and 5-year OS probabilities. Model performance was assessed using the concordance index (C-index), calibration plots, and time-dependent ROC analysis.

RESULTS

Identification of differentially expressed lncRNAs

Analysis of 1208 transcriptome samples revealed 2100 significantly differentially expressed lncRNAs (DELs) between tumor and normal breast tissues ($p < 0.05$, Perou CM et al. (2000), Blows FM et al. (2010), Colombo et al. (2011). logFC > 1).

Figure 3: 9- lncRNA risk stratification of markers for age.



A: High-risk group and low-risk Kaplan-Meier curves based on 9-lncRNA markers when age < 60; B: High-risk group and low-risk Kaplan-Meier curve based on 9-lncRNA marker when age ≥60.

Dataset partitioning and validation

After excluding patients with incomplete clinical data or survival < 0 days, 1053 BC patients were included.

Veer LJ et al. (2002), Markopoulos C et al. (2013), Knowles MA et al. (2015) Among them, 524 patients were randomly assigned to the primary dataset, while the full cohort was used as the validation dataset.

Van Batavia J et al. (2014), Warrick JI et al. (2019) No significant differences in clinical characteristics were found between groups ($p > 0.05$).

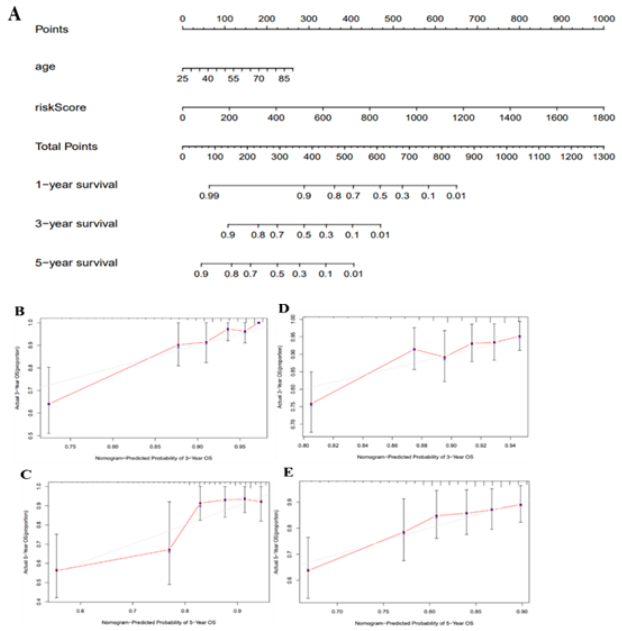
Table 2: Nine lncRNAs significantly associated with overall survival in breast cancer patients

lncRNA	HR (Hazard Ratio)	95% CI	p-value	Role
AC068858.1	3.66	1.92 – 6.82	<0.001	Risk
AC000067.1	2.58	1.41 – 4.72	0.002	Risk
LINC00460	1.18	1.07 – 1.32	0.001	Risk
LINC02408	1.87	1.25 – 2.81	0.003	Risk
AC136475.5	0.63	0.42 – 0.95	0.029	Protective
AC023043.4	0.56	0.34 – 0.92	0.021	Protective
AC073359.1	0.38	0.19 – 0.76	0.006	Protective
AC244502.1	0.41	0.22 – 0.76	0.004	Protective
COL4A2-AS1	0.52	0.31 – 0.88	0.015	Protective

Prognostic lncRNAs associated with survival

Nine lncRNAs significantly correlated with OS were identified. Beermann J et al. (2016)- Huarte M et al. (2015) Among them, AC068858.1, AC000067.1, LINC00460, and LINC02408 were associated with poor survival (HR > 1), whereas AC136475.5, AC023043.4, AC073359.1, AC244502.1, and COL4A2-AS1 were protective (HR < 1).

Figure 4: Nomogram of predicted 1/3/5-year overall survival in BC patients.



A: A nomogram used to predict overall survival; B: Calibration chart of 3-year overall survival predicted by the nomogram in the primary dataset; C: Calibrated chart of 5-year overall survival predicted by the nomogram in the primary dataset; D: Calibration chart of 3-year overall survival predicted by the nomogram in the entire dataset; E: Calibration chart of 5-year overall survival predicted by the nomogram in the entire dataset.

Construction of the 9-lncRNA risk score model

A prognostic risk score formula was generated from the 9 lncRNAs. Kopp F et al. (2018) Kaplan–Meier analysis demonstrated significantly worse survival in the high-risk group compared with the low-risk group ($p < 0.001$). ROC curves confirmed strong predictive accuracy, with AUC values up to 0.92 at 8 years.

Independence of the 9-lncRNA model from clinical variables

Multivariate Cox regression revealed that age and the 9-lncRNA risk score were independent predictors of OS ($p < 0.001$). Stratified analyses confirmed the prognostic utility of the 9-lncRNA signature in both younger (<60 years) and older (≥60 years) subgroups.

Development and validation of the predictive nomogram

A nomogram integrating age and the 9-lncRNA signature was constructed to predict individual survival outcomes. Calibration plots demonstrated excellent agreement between predicted and observed survival rates. The model achieved a C-index of 0.81, indicating strong predictive performance.

Table 3: Multivariate Cox regression analysis in the primary dataset

Variable	HR	95% CI	p-value
Age	1.059	1.032 – 1.087	<0.001
Stage	1.114	0.982 – 1.263	0.091
T stage	1.082	0.944 – 1.238	0.245
N stage	1.071	0.988 – 1.161	0.089
M stage	1.129	0.946 – 1.348	0.174
9-lncRNA model	1.035	1.018 – 1.052	<0.001

Table 4: Univariate and multivariate Cox proportional risk regression analysis of each risk factor in the entire dataset

Clinical features	Univariate analysis		multivariate analysis	
	HR	p value	HR	p value
Age	1.036	<0.001	1.035	<0.001
Stage	2.123	<0.001	1.531	0.110
T	1.562	<0.001	1.042	0.790
N	1.705	<0.001	1.248	0.140
M	5.907	<0.001	1.460	0.380
Risk score	1.005	<0.001	1.004	0.0004

DISCUSSION

Breast cancer prognosis remains difficult to predict due to its marked heterogeneity. Current clinical models, although useful, are limited in capturing molecular complexity and patient variability. This study highlights the clinical potential of lncRNAs as robust biomarkers for survival prediction.

By analyzing TCGA data, we identified a panel of 9 lncRNAs that effectively stratified patients into distinct prognostic groups. Importantly, the lncRNA-based model retained predictive power independent of conventional clinical factors such as TNM stage, confirming its robustness. Integration of age and lncRNA risk scores into a nomogram further improved prognostic accuracy, providing clinicians with a simple and quantitative tool for individualized prediction.

Compared with existing models such as Oncotype DX or MammaPrint, our lncRNA-based nomogram is advantageous because it integrates molecular and clinical information, potentially enhancing clinical applicability in diverse patient populations.

Nevertheless, validation in external cohorts and prospective clinical studies is essential before clinical implementation. Future studies should also explore the biological functions of these 9 lncRNAs in breast cancer pathogenesis, which may uncover novel therapeutic targets.

CONCLUSION

This study identified a novel 9-lncRNA signature with strong prognostic value for breast cancer patients. By integrating this molecular signature with age, we developed a predictive nomogram that demonstrated excellent accuracy and reliability in estimating survival. This tool holds promise for guiding personalized treatment decisions and improving patient outcomes in breast cancer management.

REFERENCES

1. Global Cancer Observatory: December, 2020.

2. ALLEMANI C, MATSUDA T, DI CARLO V, et al. 2018. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37513025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet. 391(10125):1023-75.

3. Yue Gong, Peng Ji, Yun-Song Yang, et al. 2021. Metabolic-Pathway-Based Subtyping of Triple- Negative Breast Cancer Reveals Potential Therapeutic Targets. Cell Metab. 33(1):51-64.

4. Lang GT, Jiang YZ, Shi JX, et al. 2020. Characterization of the genomic landscape and actionable mutations in Chinese breast cancers by clinical sequencing. Nat Commun. 11(1):5679.

5. Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 81:145–66.

6. Morris KV, Mattick JS. 2014. The rise of regulatory RNA. Nat Rev Genet. 15(6):423-37.

7.Yan X, Hu Z, Feng Y, et al. 2015. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. Cancer Cell. 28:529–40.

8. Miyamoto DT, Mouw KW, Feng FY, et al. 2018. Molecular biomarkers in bladder preservation therapy for muscle-invasive bladder cancer. Lancet Oncol. 19:e683–95.

9. Quinn JJ, Ilik IA, Qu K, et al. 2014. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolationby RNA purification. Nat Biotechnol. 32:933–40.

10. Engreitz JM, Haines JE, Perez EM, et al. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature. 539:452–55.

11. Guo H, Ahmed M, Zhang F, et al. 2016. Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. Nat Genet. 48:1142–50.

12. Lin TY, Li Y, Liu Q, et al. 2016. Novel theragnostic nano porphyrins for photodynamic diagnosis and trimodal therapy for bladder cancer. Biomaterials. 104:339–51.

13. Beermann J, Piccoli MT, Viereck J, et al. 2016. non-coding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. PhysiolRev. 96:1297–325.

14. Pandey GK, Mitra S, Subhash S, et al. 2014. The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. Cancer Cell.26:722–37.

15. Boon RA, Jaé N, Holdt L, et al. 2016. Long Noncoding RNAs: From Clinical Geneticsto Therapeutic Targets? J Am Coll Cardiol. 67:1214–26.

16. Qu L, Wang ZL, Chen Q, et al. 2018. Prognostic Value of a Long Non-coding RNA Signature in Localized

Clear Cell Renal Cell Carcinoma. Eur Urol. 74:756–63.

17. White NM, Zhao SG, Zhang J, et al. 2017. Multi-institutional Analysis Shows that Low PCAT-14 Expression Associates with Poor Outcomes in Prostate Cancer. Eur Urol. 71:257–66.

18. Simpson PT, ReisFilho JS, Lakhani SR. 2010. Breast pathology:beyond morphology. Semin Diagn Pathol. 27(1):91-96.

19. Park YH, Lee SJ, Cho EY, et al. 2011. Clinical relevance of TNM staging system according to breast cancer subtypes. Ann Oncol. 22(7): 1554-60.

20. Rakha EA, ReisFilho JS, Baehner F, et al. 2010. Breast cancer prognostic classification in the molecular era : the role of histological grade. Breast Cancer Res. 12(4):207.

21. Chowdhury N, Pai MR, Lobe FD, et al. 2006. Interobserver variation in breast cancer grading : a statistical modeling approach. Anal Quant Cytol Histol. 28(4):213-18.

22. Nguyen—Ngoc KV, Cheung KJ, Brenot A, et al. 2012. ECM micmenvironment regulates collective migration and local dissemination in normal and malignant mammaq epithelium. Proc Natl Acad Sci USA. 109(39) : E2595-2604.

23. Perou CM, Sorlie T, Eisen MB, et al. 2000. Molecular portraits of human breast tumours. Nature. 406(6797): 747-52.

24. Blows FM, Driver KE, Schmidt MK, et al. 2010. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short- and long-term survival: a collaborative analysis of data for of 159 eases from 12 studies. PLOS Med. 7(5): 1000279.

25. Colombo PE, Milanezi F, Weigelt B, et al. 2011. Microarrays in the 2010s: the contribution of mieroarray-based gene expression profiling to breast cancer classification, prognostication and prediction. Breast Cancer Res. 13(3): 212.

26. Veer LJ, Dai H, van de Vijver MJ, et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 415(6871): 530-36.

27. Markopoulos C. 2013. Overview of the use of Oncotype DX((R)) as an additional treatment decision tool in early breast cancer. Expert Rev Anticancer Ther. 13(2):179-94.

28. Knowles MA, Hurst CD. 2015. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat Rev Cancer.15:25–41.

29. Van Batavia J, Yamany T, Molotkov A, et al. 2014. Bladder cancers arise from distinct urothelial sub-populations. Nat Cell Biol. 16:982–91, 1–5.

30. Warrick JI, Sjödahl G, Kaag M, et al. 2019. Intratumoral Heterogeneity of Bladder Cancer by Molecular Subtypes and Histologic Variants. Eur Urol. 75:18–22.

31. Beermann J, Piccoli MT, Viereck J, et al. 2016. Noncoding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. Physiol Rev. 96:1297–325.

32. Huarte M. 2015. The emerging role of lncRNAs in cancer. Nat Med.21:1253–61.

33. Kopp F, Mendell JT. 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell. 172:393–407.

