Identification of Metastasis-Associated Genes in Triple-Negative Breast Cancer Using Weighted Gene Co-expression Network Analysis

Evolutionary Bioinformatics
Volume 16: 1–8

DOI: 10.1177/1176934320954868

Wenting Xie, Zhongshi Du, Yijie Chen, Naxiang Liu, Zhaoming Zhong, Youhong Shen and Lina Tang

Department of Ultrasound, Fujian Cancer Hospital and Fujian Medical University Cancer Hospital, Fujian Province, China.

ABSTRACT: Triple-negative breast cancer (TNBC) is the most aggressive and fatal sub-type of breast cancer. This study aimed to identify metastasis-associated genes that could serve as biomarkers for TNBC diagnosis and prognosis. RNA-seq data and clinical information on TNBC from the Cancer Genome Atlas were used to conduct analyses. Expression data were used to establish co-expression modules using average linkage hierarchical clustering. We used weighted gene co-expression network analysis to explore the associations between gene sets and clinical features and to identify metastasis-associated candidate biomarkers. The K-M plotter website was used to explore the association between the expression of candidate biomarkers and patient survival. In addition, receiver operating characteristic curve analysis was used to illustrate the diagnostic performance of candidate genes. The pale turquoise module was significantly associated with the occurrence of metastasis. In this module, 64 genes were identified, and its functional enrichment analysis revealed that they were mainly associated with transcriptional misregulation in cancer, microRNAs in cancer, and negative regulation of angiogenesis. Further, 4 genes, *IGSF10, RUNX1T1, XIST*, and *TSHZ2*, which were negatively associated with relapse-free survival and have seldom been reported before in TNBC, were selected. In addition, the mRNA expression levels of the 4 candidate genes were significantly lower in TNBC tumor tissues compared with healthy tissues. Based on the K-M plotter, these 4 genes were correlated with poor prognosis of TNBC. The area under the curve of *IGSF10, RUNX1T1, TSHZ2*, and *XIST* was 0.918, 0.957, 0.977, and 0.749. These findings provide new insight into TNBC metastasis. *IGSF10, RUNX1T1, TSHZ2*, and *XIST* could be used as candidate biomarkers for the diagnosis and prognosis of TNBC metastasis.

KEYWORDS: Triple negative breast neoplasms, WGCNA, neoplasm metastasis, genes, biomarker

RECEIVED: June 15, 2020. **ACCEPTED:** August 10, 2020.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The present study was supported by the Startup Fund for scientific research, Fujian Medical University (grant number: 2019QH1194) and Fujian Province Science and Technology Plan Guided Project fund (grant number: 2017Y0021).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Lina Tang, Department of Ultrasound, Fujian Cancer Hospital and Fujian Medical University Cancer Hospital, Fuma road 420, Fuzhou 350014, Fujian Province, China. Email: tanglina@fjzlhospital.com

Introduction

Breast cancer is the second most commonly diagnosed cancer, and it accounts for ~11.6% of all cancer cases.[1] According to the expression of receptor proteins and genes, breast cancer can be categorized into 4 subtypes. Among these subtypes, triple-negative breast cancer (TNBC),[2] which comprises ~15% to 20% of breast cancer is generally defined as being ER-negative, PR-negative, or HER2 negative. A previous study reported that TNBC has a high risk of metastases and typically behaves more aggressively, which increases the poor prognosis of patients.[3] Its metastases at distant sites are also the primary cause of cancer-related death in patients. Due to the lack of gene targets for metastases, available therapies are largely unsuccessful in treating metastases.[4] The treatment of metastases is more dependent on the estrogen, progesterone, and human epidermal growth factor receptor 2 status of the patient. Since TNBC lacks these molecular targets, few new agents have been approved for treating the subset of patients with metastases.[5]

Metastasis is an evolutionary process.[6] Multiple competing subclones can emerge in primary tumors, culminating in the formation of metastases.[7] Genetic and epigenetic alterations in primary tumor cells contribute to the evolutionary process.[6] Nearly 12% of breast cancer cases eventually become metastatic cases, and after diagnosis with metastatic breast cancer, the 5-year survival rate is 26%.[8] When compared with their metastatic breast cancer counterparts, patients with metastatic TNBC have a higher death rate.[9] Since metastatic breast cancer is incurable, especially metastatic TNBC, there has been substantial interest in understanding changes in metastasis-associated genes. In addition, it is necessary to explore new metastasis-associated biomarkers to determine their utility in diagnosis and predicting prognosis. Also, biomarkers detected need robustness and stability. A recent study reported that a novel network-based approach can identify the biomarkers of breast cancer survivability.[10]

We identified genes involved in metastatic TNBC through a comprehensive analysis of the Cancer Genome Atlas (TCGA) gene expression data. Weighted gene co-expression network analysis (WGCNA) is a systems biology method to describe the patterns in correlations of genes among samples. It has been proven to be a reliable tool for identifying candidate biomarkers,[11] and it has been used to identify biologically meaningful modules related to metastatic breast cancer in our study. Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) analysis was performed to explore

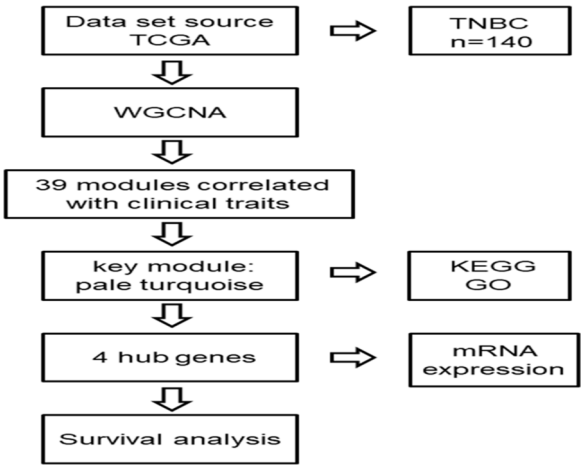

Figure 1. Study workflow.
Abbreviations: TCGA, the Cancer Genome Atlas; TNBC, triple-negative breast cancer; WGCNA, weighted gene co-expression network analysis; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

functions and pathways related to genes within key modules and identify their biological meaning. We identified 4 candidate genes, *IGSF10, RUNX1T1, XIST*, and *TSHZ2*, from the related module that were associated with TNBC prognosis. These candidate genes could serve as candidate biomarkers of metastatic TNBC and contribute to understanding TNBC progression.

Materials and Methods

Data analyses were performed, as indicated in Figure 1.

Data sources

Breast tissue RNA sequence data (HTseq-counts) and clinical features were accessed from TCGA (https://cancergenome.nih.gov/; accession date: 14 September 2018). Of the available data, 140 TNBC and 13 adjacent healthy tissue samples were selected for analysis. These data are publicly accessible, and no further ethical approval was required from the Ethics Committee.

Weighted gene co-expression network analysis

The WGCNA package in R was used to construct the gene co-expression network. Data were normalized using edgeR,[12] and then we screened the genes in the top 25% of the variance. All 13 845 genes from the 140 TNBC samples were used to establish co-expression modules. An interaction coefficient was calculated between genes. The adjacency matrix was converted to a topological overlap matrix (TOM), and then genes were divided into different gene modules according to the TOM-based heterogeneity measure. The soft-thresholding power was 8. Gene modules were constructed using a dynamic tree cut algorithm, and the minimum number of genes was set as 30 to obtain more reliable results. A module eigenvalue distance threshold was set as 0.25 to merge highly similar modules. The module that had the highest correlation and was significantly related to metastasis was selected and used for further analysis. Modules with a *P*-value $<.05$ were identified as clinical trait-related modules.

Function-enrichment analyses of metastasis-associated modules

Gene Ontology (GO) annotation and KEGG pathway enrichment were used to analyze genes using the Database for Annotation, Visualization, and Integration Discovery (http://david.abcc.ncifcrf.gov/) to explore the biological functions of genes in metastasis-associated modules.[13] The threshold of significance was set as $P<.05$.

Survival analysis of candidate genes

The K-M plotter website (http://www.kmplot.com) was used to analyze the association between the expression of hub genes and the survival of patients.[14] The threshold was adjusted to $P<.05$.

Expression of genes critical in triple-negative breast cancer and healthy tissues

Samples of 140 TNBC tissues and 13 corresponding healthy tissues were used to explore the expression of candidate genes. Differences in gene expression between the 2 groups were analyzed using the Mann-Whitney *U* test in SPSS Statistics version 20.0 software (IBM Corp.). Data were visualized using GraphPad Prism 7.0 (GraphPad Software Inc, CA, USA). A *P*-value of $<.05$ was considered to be statistically significant.

Statistical analysis of clinical covariates

We separated the TNBC cases into metastatic and non-metastatic cases. Then, we calculated the significant differences between different ages, T, N, and stages between groups. A Chi-square test was used to compare binary variables, and continuous variables were analyzed using t-tests. A *P*-value of $<.05$ was considered statistically significant.

Results

Expression value analysis of mRNA-seq data of triple-negative breast cancer

A total of 140 TNBC samples were obtained from TCGA. Clinical information for TNBC patients is shown in Table 1. We transformed the RNA-seq data to gene expression information. Genes with missing and negative values were eliminated. As a result, 25% of the genes before the variance were obtained. A total of 13 845 expression values of genes were selected for analysis using the WGCNA package.

Table 1. Tumor characteristics for TNBC patients in the present study.

CHARACTERISTICS	INFORMATION	SAMPLE NUMBER N = 140
Age (years)		55.90 ± 12.50
T		
	T1-T2	122 (87.1%)
	T3-T4	18 (12.9%)
N		
	No	88 (62.8%)
	Yes	52 (37.1%)
M		
	No	128 (91.4%)
	Yes	12 (8.6%)
Stage		
	Stage I-II	117 (83.6%)
	Stage III-IV	23 (16.4%)

Construction of the co-expression module and identification of the key module of triple-negative breast cancer

The WGCNA algorithm was used to construct the co-expression modules most associated with the TNBC clinical traits (Figure 2). Clinical information of the TNBC samples such as age, TNM, and the stage was retrieved from TCGA (Figure 2A). Age was expressed as the mean ± standard deviation and T, N, and stage were binary variables, for which a Chi-square test was used to analyze differences. Our results show that there was no difference in the mean age of the metastatic and non-metastatic groups, while there were significant differences in T, N, and stage between the 2 groups (Supplemental Table 1). We set the soft-thresholding power as 8 for further analysis and set the cut height as 0.25, and we eventually constructed 39 modules (Figure 2B-D).

Metastasis-associated module analysis

Using the module-trait correlations heatmap, we identified that the pale turquoise module was the most highly related to the characteristic of metastasis (correlation coefficient = 0.26,

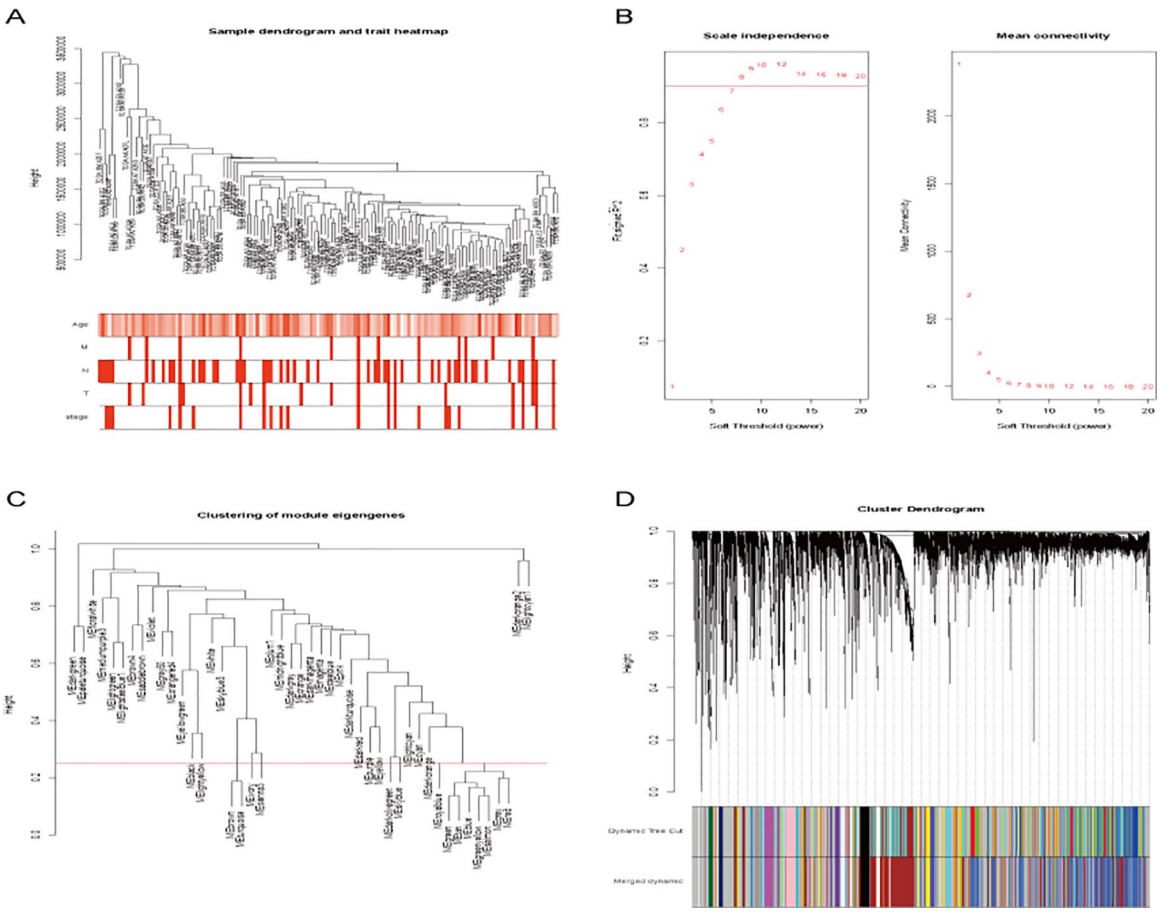

Figure 2. Weighted gene co-expression network analysis to construct the co-expression module network. (A) Clustering dendrograms of genes. Color intensity varies with age, T, N, M, and stage. (B) Scale-free fit index (left) and the mean connectivity (right) for soft-thresholding powers. (C) Clustering of module eigengenes. The height of the red line is 0.25. (D) Clustering dendrograms of all genes. As a result, 39 co-expression modules were constructed and highlighted with different colors.

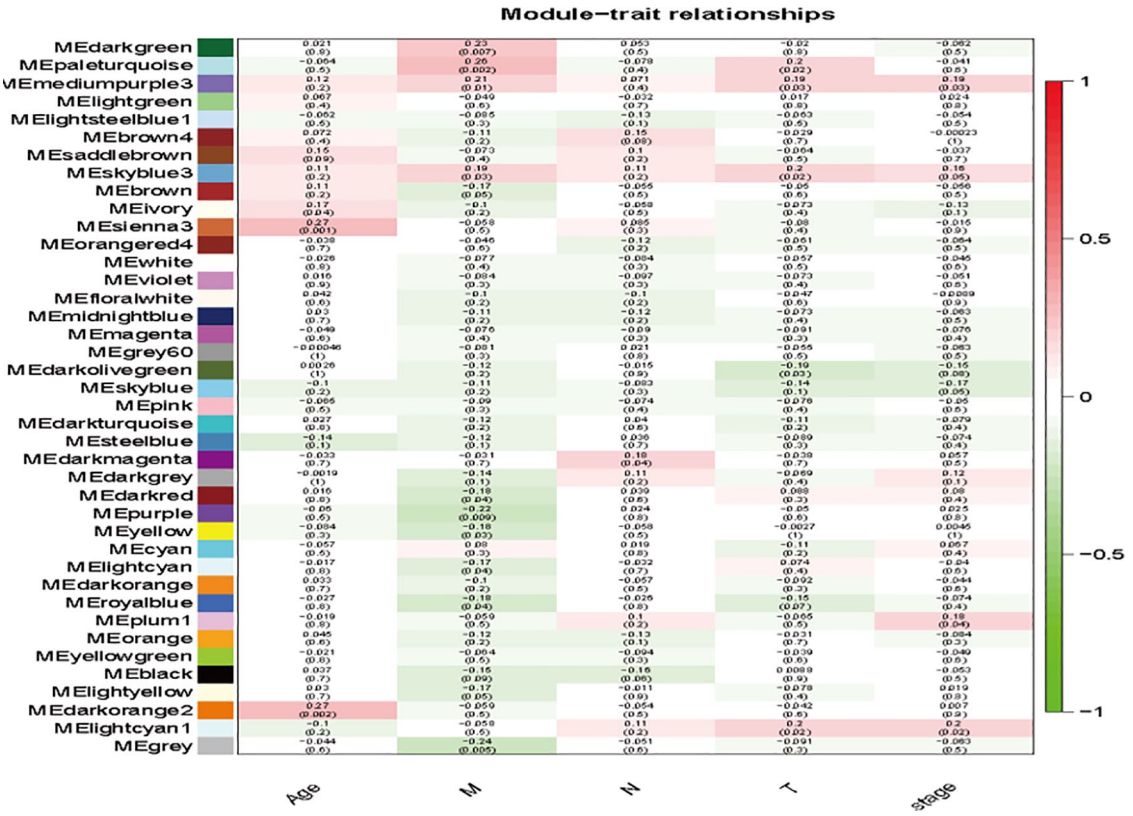

Figure 3. Heatmap of the correlation between clinical traits and eigengenes of triple-negative breast cancer. Each module contains a correlation coefficient and *P*-value.

P = 0.002; Figure 3). The pale turquoise module contained a total of 64 genes (Figure 4A, correlation coefficient = 0.67, *P* = 1.4e-09 of MM in pale turquoise). All 64 genes in the pale turquoise module were subjected to further analysis. We performed GO and KEGG enrichment analyses to reveal potential biological functions of the genes in the pale turquoise module. As presented in Figure 4B, KEGG pathway analyses showed that genes were primarily enriched in the pathways of hsa05202 (transcriptional misregulation in cancer), hsa05206 (microRNAs in cancer), and hsa04360 (axon guidance). As shown in Figure 4C, the enriched base-pair terms of GO in the pale turquoise module were mainly about the GO:0016525 (negative regulation of angiogenesis), GO:0060021 (palate development), and GO:0035988 (chondrocyte proliferation). The enriched CC terms of GO in genes were functional at GO:0031012 (extracellular matrix) and GO:0005576 (extracellular region). The enriched MF terms of GO in the key module were mainly about GO:0008201 (heparin binding) and GO:0001078 (transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding). The detailed results of the GO and KEGG analyses are illustrated in Table 2.

Novel candidate genes analysis in metastasis-associated module

Among the 30 top genes in the pale turquoise module based on intramodule connectivity and by setting MM at >0.85 and gene significance (GS) at >0.15, 26 genes with high connectivity in the pale turquoise module were identified as hub genes. Among these genes, *IGSF10, RUNX1T1, XIST*, and *TSHZ2* were negatively associated with relapse-free survival and have seldom been reported before in TNBC. As shown in Figure 5, we found that all of the genes were significantly downregulated ($P < .05$) in 140 TNBC samples compared with 13 adjacent healthy samples. Furthermore, the Kaplan–Meier curve and the log-rank test were used to assess relapse-free survival in patients. Kaplan–Meier curves showed that the lower the expression of these genes correlated significantly with poor relapse-free survival (Figure 6). Detailed HR and the log-rank *P*-values are presented in Table 3. Notably, 4 novel candidate genes in the pale turquoise module showed good prognostic values.

In addition, receiver operating characteristic (ROC) curve analysis was used to evaluate the capacity of candidate genes to the diagnosis of TNBC (Figure 7). Area under the ROC curve values for 4 novel candidate genes are presented in Table 4.

Discussion

Breast cancer is a complex and heterogeneous disease at the tumor genetics and patient's prognosis levels. Compared with other breast cancers subtypes, TNBC behaves more aggressively and those patients with TNBC have higher death rates.[15] What is worse is that metastasis is the vital tab in cancer. Patients with metastasis TNBC have an additional challenge to finding targets and treatments. In the current study, we aimed to explore the prognosis biomarker of metastasis associated TNBC using WGCNA. We downloaded gene expression

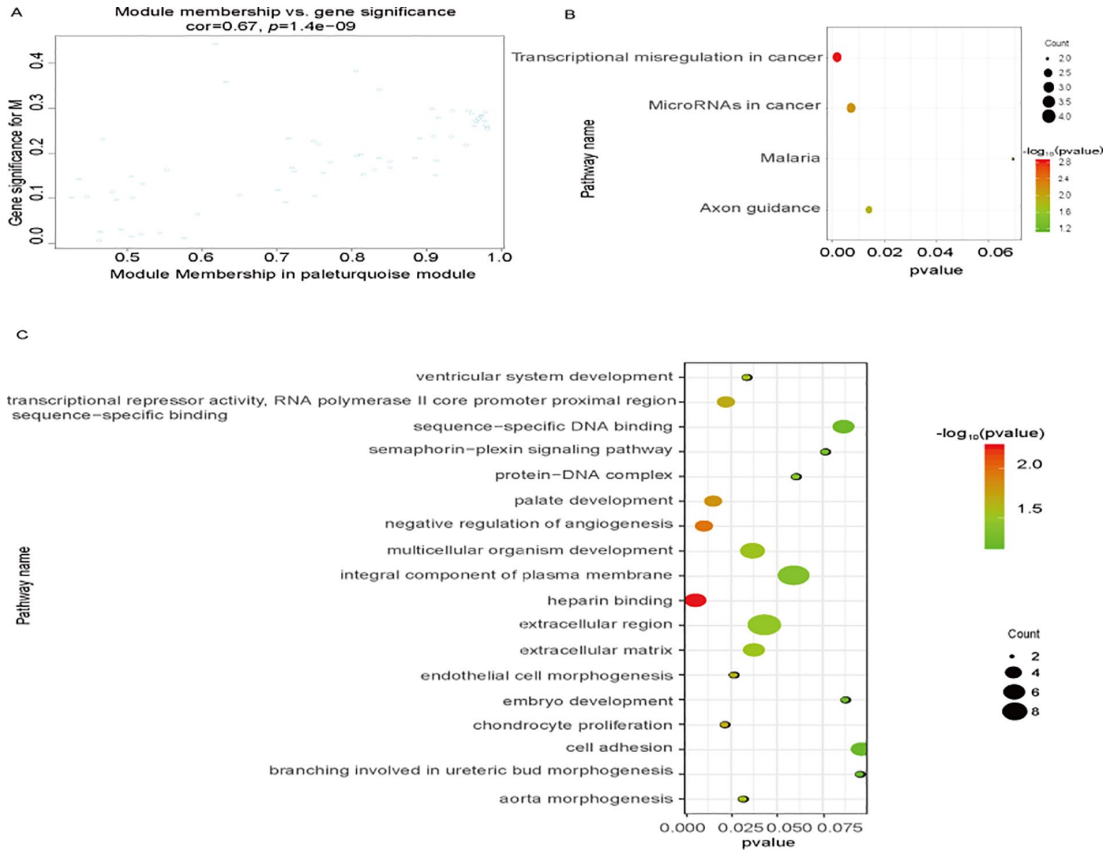

Figure 4. Metastasis-related module analyses and functional annotation. (A) Scatterplot of genes in the pale turquoise module. (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis for genes in the pale turquoise module. (C) Gene ontology (GO) analysis for genes in the pale turquoise module.

Table 2. The GO and KEGG analysis of key module.

CATEGORY	TERM	INVOLVED IN	*P*
GO			
GOTERM_BP_DIRECT	GO:0016525	Negative regulation of angiogenesis	.009
GOTERM_BP_DIRECT	GO:0060021	Palate development	.014
GOTERM_BP_DIRECT	GO:0035988	Chondrocyte proliferation	.021
GOTERM_BP_DIRECT	GO:0001886	Endothelial cell morphogenesis	.025
GOTERM_BP_DIRECT	GO:0035909	Aorta morphogenesis	.030
GOTERM_BP_DIRECT	GO:0021591	Ventricular system development	.032
GOTERM_BP_DIRECT	GO:0007275	Multicellular organism development	.034
GOTERM_CC_DIRECT	GO:0031012	Extracellular matrix	.036
GOTERM_CC_DIRECT	GO:0005576	Extracellular region	.041
GOTERM_MF_DIRECT	GO:0008201	Heparin binding	.004
GOTERM_MF_DIRECT	GO:0001078	Transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	.02
KEGG			
KEGG_PATHWAY	hsa05202	Transcriptional misregulation in cancer	.001
KEGG_PATHWAY	hsa05206	MicroRNAs in cancer	.006
KEGG_PATHWAY	hsa04360	Axon guidance	.013

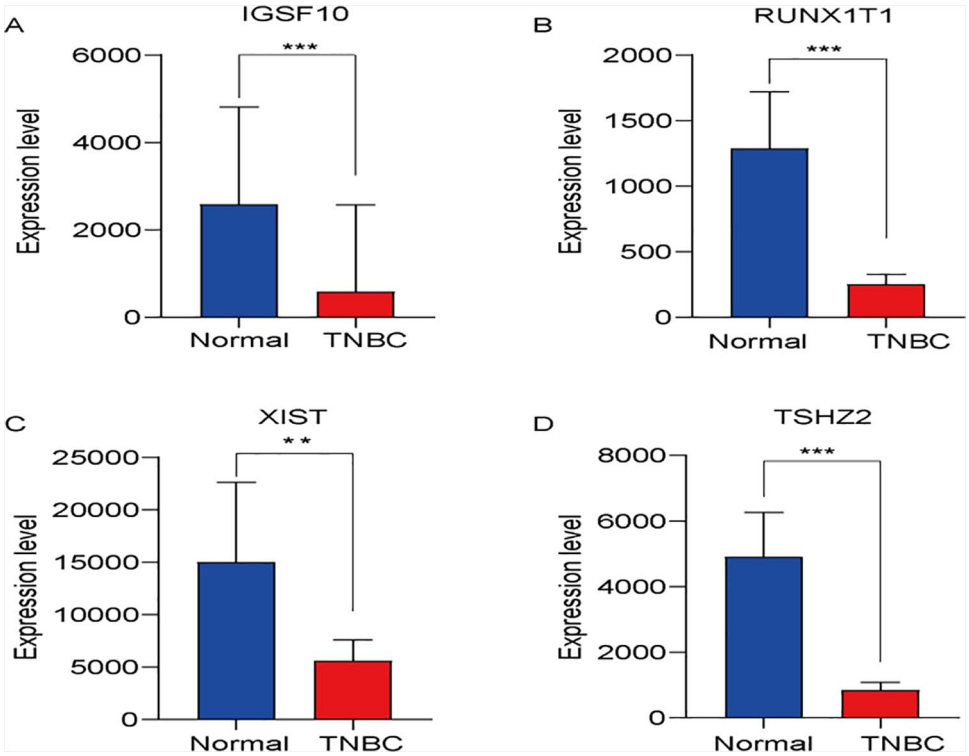

Figure 5. The mRNA expression levels of candidate genes in triple-negative breast cancer and corresponding healthy tissues based on the Cancer Genome Atlas dataset. (A) Messenger RNA expression of *IGSF10*. (B) Messenger RNA expression of *RUNX1T1*. (C) Messenger RNA expression of *XIST*, (D) Messenger RNA expression of *TSHZ2*.
*$P < .05$, **$P < .01$, and ***$P < .001$.

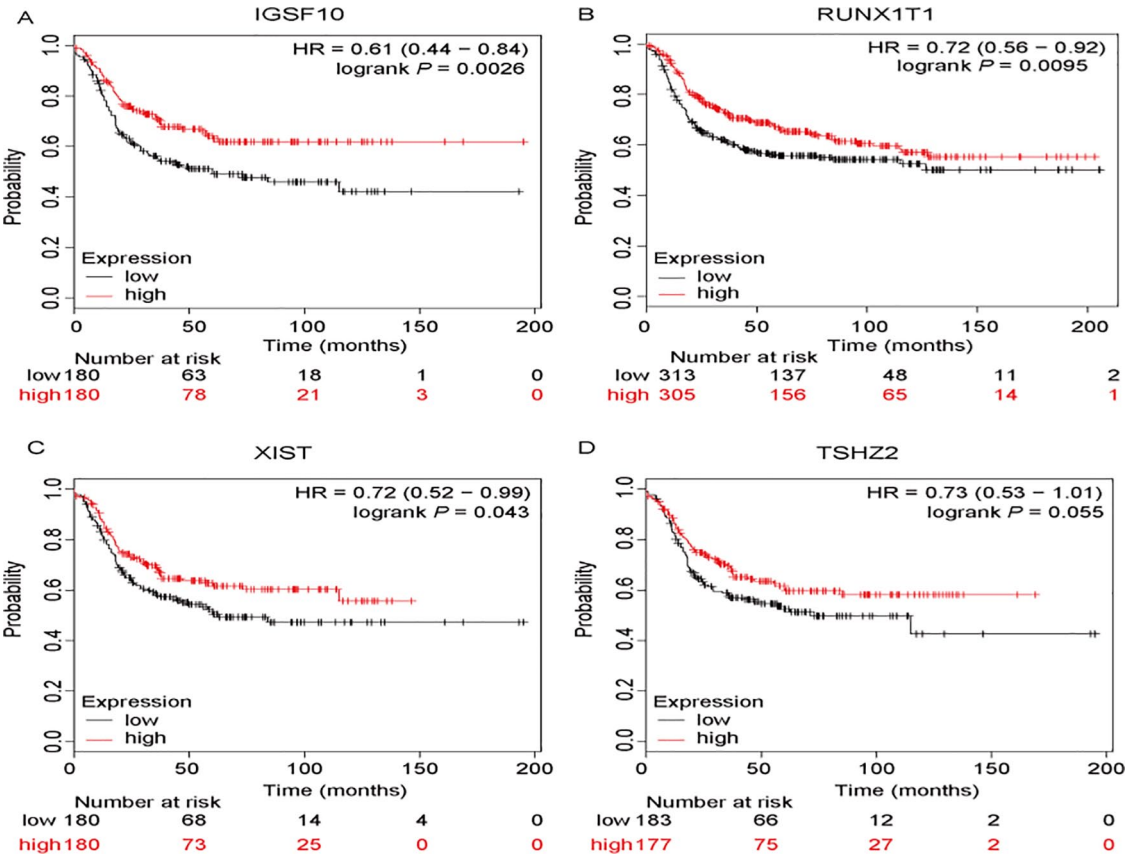

Figure 6. Associated candidate gene expression and recurrence-free survival time using the K-M plotter online platform. (A) Kaplan–Meier curves for *IGSF10*. (B) Kaplan–Meier curves for *RUNX1T1*. (C) Kaplan–Meier curves for *XIST*. (D) Kaplan–Meier curves for *TSHZ2*.

Table 3. The survival analysis of candidate genes.

GENE	FULL NAME	PROBE	HIGH EXPRESSION	LOW EXPRESSION	HR	LOGRANK *P*
IGSF10	Immunoglobulin superfamily member 10	1556579_s_at	180	180	0.61 (0.44-0.84)	.0026
RUNX1T1	RUNX1 partner transcriptional co-repressor 1	205528_s_at	305	313	0.72 (0.56-0.92)	.0095
XIST	X inactive specific transcript	224589_at	180	180	0.72 (0.52-0.99)	.043
TSHZ2	Teashirt zinc finger homeobox 2	244521_at	177	183	0.73 (0.53-1.01)	.055

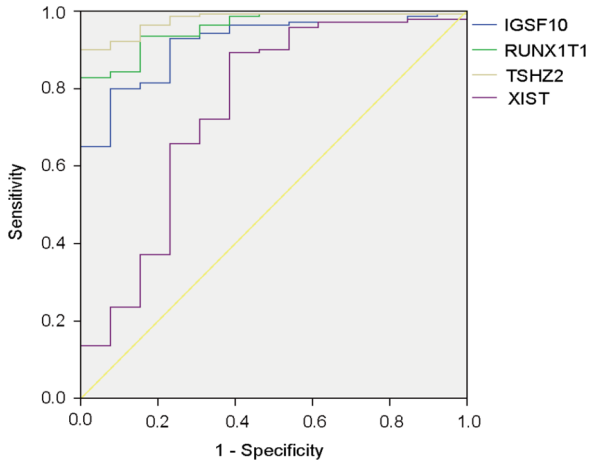

Figure 7. Receiver operating characteristic (ROC) analysis of candidate genes. These curves were used to evaluate the capacity of candidate genes in the diagnosis of triple-negative breast cancer.

Table 4. The AUC of candidate genes.

GENE	AUC	95% CI (LOWER)	95% CI (UPPER)	*P*
IGSF10	0.918	0.857	0.979	.000
RUNX1T1	0.957	0.920	0.994	.000
TSHZ2	0.977	0.954	1.000	.000
XIST	0.749	0.585	0.913	.003

Abbreviations: AUC, area under the curve; CI, confidence interval.

profiles from TCGA to construct a co-expression network and identified metastasis-associated candidate genes. After setting GS > 0.15 and MM > 0.85, we eventually obtained 26 hub genes. Some of them have been demonstrated to exert vital roles in breast cancer.[16] Among these genes, we chose 4 genes that have seldom been reported in TNBC metastasis, namely *IGSF10, RUNX1T1, XIST*, and *TSHZ2*, to further explore their prognostic and diagnosis value.

IGSF10, namely immunoglobulin superfamily member 10, is related to differentiation and developmental processes, and *IGSF10* is the genetic basis of delayed puberty and disorders of neuronal development.[17] Previous studies have reported that *IGSF10* is possibly involved in radiation-induced rat osteosarcomas.[18] However, the *IGSF10* gene has rarely been associated with cancer. One study showed that mutation in *IGSF10* might be associated with gastric and rectal cancer.[19] Whole-exome sequencing of 14 endometrial cancer tissue samples showed that *IGSF10* was the potential cancer-related gene.[20]

RUNX1T1 is a member of the transcriptional corepressors of the MTG family. It has been demonstrated that it is closely involved in the pathogenesis of acute leukemia.[21] An RNA sequencing study revealed that *RUNX1T1* was upregulated in clear renal cell carcinoma, which suggests that this gene is vital for tumorigenesis.[22] *RUNX1T1* has also been reported in other cancer types. Nasir et al[23] revealed that *RUNX1T1* might be a novel biomarker for the prediction of liver metastasis in primary pancreatic endocrine tumors. Since the dysregulated of TGFb/SMAD4 signaling may result in epigenetic silencing of *RUNX1T1*, this suggests that *RUNX1T1* is crucial for ovarian carcinogenesis.[24]

XIST is involved in the inactivation of the X chromosome, which is a non-coding RNA. It has been demonstrated that its expression has been dysregulated in numerous cancers, especially in breast cancer. BRCA1, which interacts with *XIST* RNA, takes part in the correct inactive X chromosome heterochromatin superstructure.[25] The loss of Xi might present more aggressively in breast cancer. This might suggest that *XIST* performs a vital role in the regulation of cancer-related pathways in breast cancer.[26]

TSHZ2 is a member of the TSHZ family, which includes *TSHZ1, TSHZ2*, and *TSHZ3*. It has been demonstrated that the silence of the *TSHZ2* gene may play a critical role in carcinogenesis. The expression of *TSHZ2* is downregulated in some cancers. This suggests that it might function as a tumor-suppressor gene.[27] However, the underlying molecular mechanism is not fully understood. A study reported that *TSHZ2* participated in mammary tumorigenesis via activation of GLI1.[28]

However, our present study has some limitations. First, the candidate genes should have been validated using samples from our institution via quantitative PCR or western blots. Thus, we will collect tissue samples for further investigation. Second, the biological molecular mechanisms of candidate genes in TNBC requires further exploration.

In summary, this study focused on metastasis-associated genes in TNBC. By combining WGCNA and other bioinformatics tools, we identified significant gene modules related to metastasis in TNBC. Four candidate genes, *IGSF10*, *RUNX1T1*, *XIST*, and *TSHZ2*, were strongly downregulated in TNBC tissues. Further survival analysis suggested that these genes have significant prognostic and diagnosis values in TNBC.

Author Contributions

LNT designed the study. WTX and ZSD wrote the main manuscript. YJC and NXL collected the data. ZMZ and YHS analyzed the data. WTX and ZSD contributed equally.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
2. Hong HC, Chuang CH, Huang WC, et al. A panel of eight microRNAs is a good predictive parameter for triple-negative breast cancer relapse. *Theranostics*. 2020;10:8771-8789.
3. Neophytou C, Boutsikos P, Papageorgis P. Molecular mechanisms and emerging therapeutic targets of triple-negative breast cancer metastasis. *Front Oncol*. 2018;8:31.
4. Hwang SY, Park S, Kwon Y. Recent therapeutic trends and promising targets in triple negative breast cancer. *Pharmacol Ther* 2019;199:30-57.
5. Gogate A, Rotter JS, Trogdon JG, et al. An updated systematic review of the cost-effectiveness of therapies for metastatic breast cancer. *Breast Cancer Res Treat*. 2019;174:343-355.
6. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science*. 2016; 352:169-175.
7. Gundem G, Van Loo P, Kremeyer B, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature*. 2015;520:353-357.
8. Peart O. Metastatic breast cancer. *Radiol Technol*. 2017;88:519m-539m.
9. Zeichner SB, Terawaki H, Gogineni K. A review of systemic treatment in metastatic triple-negative breast cancer. *Breast Cancer (Auckl)*. 2016;10:25-36.
10. Jubair S, Alkhateeb A, Tabl AA, et al. A novel approach to identify subtype-specific network biomarkers of breast cancer survivability. *Netw Model Anal Health Inform Bioinform*. 2020;9:43.
11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
12. Pizzollo J, Nielsen WJ, Shibata Y, et al. Comparative serum challenges show divergent patterns of gene expression and open chromatin in human and chimpanzee. *Genome Biol Evol* 2018;10:826-839.
13. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4: 44-57.
14. Nagy A, Lanczky A, Menyhart O, et al. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep* 2018;8:9227.
15. Li X, Yang J, Peng L, et al. Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast cancer. *Breast Cancer Res Treat*. 2017;161:279-287.
16. Zhang W, Shi S, Jiang J, et al. LncRNA MEG3 inhibits cell epithelial-mesenchymal transition by sponging miR-421 targeting E-cadherin in breast cancer. *Biomed Pharmacother*. 2017;91:312-319.
17. Howard SR. The genetic basis of delayed puberty. *Front Endocrinol (Lausanne)*. 2019;10:423.
18. Daino K, Ugolin N, Altmeyer-Morel S, et al. Gene expression profiling of alpha-radiation-induced rat osteosarcomas: identification of dysregulated genes involved in radiation-induced tumorigenesis of bone. *Int J Cancer*. 2009;125: 612-620.
19. Thutkawkorapin J, Picelli S, Kontham V, et al. Exome sequencing in one family with gastric- and rectal cancer. *BMC Genet*. 2016;17:41.
20. Chang YS, Huang HD, Yeh KT, et al. Identification of novel mutations in endometrial cancer patients by whole-exome sequencing. *Int J Oncol*. 2017;50: 1778-1784.
21. Qin YZ, Wang Y, Xu LP, et al. The dynamics of RUNX1-RUNX1T1 transcript levels after allogeneic hematopoietic stem cell transplantation predict relapse in patients with t(8;21) acute myeloid leukemia. *J Hematol Oncol*. 2017;10:44.
22. Xiong Z, Yu H, Ding Y, et al. RNA sequencing reveals upregulation of RUNX1-RUNX1T1 gene signatures in clear cell renal cell carcinoma. *Biomed Res Int*. 2014;2014:450621.
23. Nasir A, Helm J, Turner L, et al. RUNX1T1: a novel predictor of liver metastasis in primary pancreatic endocrine neoplasms. *Pancreas*. 2011;40:627-633.
24. Guo YH, Li YN, Zhao JR, et al. HBc binds to the CpG islands of HBV cccDNA and promotes an epigenetic permissive state. *Epigenetics*. 2011;6:720-726.
25. Ganesan S, Silver DP, Drapkin R, et al. Association of BRCA1 with the inactive X chromosome and XIST RNA. *Philos Trans R Soc Lond B Biol Sci*. 2004;359: 123-128.
26. Soudyab M, Iranpour M, Ghafouri-Fard S. The role of long non-coding RNAs in breast cancer. *Arch Iran Med*. 2016;19:508-517.
27. Yamamoto M, Cid E, Bru S, et al. Rare and frequent promoter methylation, respectively, of TSHZ2 and 3 genes that are both downregulated in expression in breast and prostate cancers. *PLoS ONE*. 2011;6:e17149.
28. Riku M, Inaguma S, Ito H, Tsunoda T, Ikeda H, Kasai K. Down-regulation of the zinc-finger homeobox protein TSHZ2 releases GLI1 from the nuclear repressor complex to restore its transcriptional activity during mammary tumorigenesis. *Oncotarget*. 2016;7:5690-5701.