Identification of The Prognostic Genes For Early Basal-Like Breast Cancer With Weighted Gene Co-Expression Network Analysis

Keyu Yuan
Capital Medical University
Min Wu
Capital Medical University
Xue Yu
Peking University International Hospital
Xia Zhao
Capital Medical University
Yu Feng
Capital Medical University
Yanping Li (lyp671102@bjsjth.cn)
Capital Medical University
Shuzhen Lv
Capital Medical University

Research Article

Keywords: Breast cancer, basal-like, early stage, prognosis

Posted Date: November 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1110962/v1

Abstract

Background: Breast cancer (BC) has become the leading cause of death for women's malignancies and increasingly threatens the health of women worldwide. However, the basal-like BC is lack of effective targeted drugs. Therefore, biomarkers that related to the prognosis of early breast cancer need to be found.

Methods: The RNA-seq data of 87 cases of early basal-like BC and 111 cases of normal breast tissue from The Cancer Genome Atlas (TCGA) were explored by Weighted Gene Co-Expression Network Analysis (WGCNA)method and Limma package. Then intersected genes (IGs) were identified and hub genes were selected by Maximal Clique Centrality method. The prognostic effect of the hub genes was also evaluated in early basal-like BC.

Results: A total of 601 IGs were identified in this study. APPI network was constructed and top 10 hub genes were selected, namely cyclin B1 (CCNB1), cyclin A2 (CCNA2), cyclin dependent kinase 1 (CDK1), cell division cycle 20 (CDC20), DNA topoisomerase II alpha (TOP2A), BUB1 mitotic checkpoint serine/threonine kinase (BUB1), aurora kinase B (AURKB), cyclin B2 (CCNB2), kinesin family member 11 (KIF11), and assembly factor for spindle microtubules (ASPM). Only AURKB was found to be significant with the overall prognosis of early basal-like BC. The immune cells infiltration analysis displayed that the infiltration numbers of CD4+ T cell and naïve CD8+ T cell were positively correlated with AURKB expression level, while that of naïve B cell and macrophage M2 cell were negatively correlated with AURKB expression level in basal-like BC.

Conclusion: AURKB might be a potential prognostic indicator in early basal-like BC.

Background

BC is the most common malignant tumor in females and seriously undermined the health of women worldwide. In 2018, the number of new BC cases was about 2.1 million, while it raised to 2.3 million 2 years later. These are more pronounced in developing countries[1-4]. The potential risk factors of BC have been proven to be followings: family susceptibility, hormone exposure, aging, and unhealthy lifestyle[5,6]. For early diagnosis and treatment, auxiliary examinations have developed rapidly. The most frequently used methods in clinical diagnosis are mammography, ultrasound, and MRI examination[7]. The common treatments of BC are including surgical resection, chemotherapy, radiotherapy, and endocrine therapy. Furthermore, molecular targeted therapy and immunotherapy are also obtained marked improvement[8-10]. The above treatments are very significantly important in clinic.

The prognosis and survival of BC is improved with the continuous improvement of medicine. In 2000, according to Perou's research results, BC is basically divided into 4 types: Luminal subtype, HER-2 over-expression subtype, basal-like subtype, and normal breast like subtype[11]. Therefore, the therapeutic guideline and consensus of BC are mainly based on molecular classification. The basal-like BC was high

invasiveness and lacking effective target drugs. Currently, some biomarkers are used in BC, such as CEA, CA125, CA153, etc., but effective prognostic indicator is still deficiency in basal-like BC.

Exploring the potential prognostic factors, which could predict the survival of the basal-like BC in advance and give higher-intensity necessary treatments, is extremely important[12]. The change of genomes has been commonly found in basal-like BCs and it may affect the effectiveness of target therapy and long-term prognosis of basal-like BCs. The improvements of genomes testing are helpful for identifying the genes related to the prognosis and promoting development of new drugs. Compared to pathological inspection, genome detection is more direct to the root of the disease and the biological characteristics, and it handles the problem that clinical parameters could not perfectly reflect. Furthermore, compared to pathological biopsy, more information could be provided in a shorter time by the new second-generation sequencing technology. This study intends to explore the biomarkers that affect the prognosis of basal-like BC from the perspective of biometric analysis.

Methods

Differentially expressed genes identifying

The RNA-seq data of early basal-like BC and normal breast tissue were downloaded and then selected from the TCGA database. The clinical data of early basal-like BC was also obtained from the TCGA database and only early basal-like BC with prognostic data were included in the following analysis. The early basal-like BC was identified as basal-like BC with American Joint Committee on Cancer TNM Staging System (AJCC TNM, 2018 Edition) stage I and ⅡA in this study. The differentially expressed genes （DEGs）were identified by using Limma package on R software. The criteria was defined as adjusted *P* value≦ 0.05 and $|Log_2FC|≧1$. The consent was un-needed for the TCGA data used in this study.

WGCNA

The gene expressing dataset of the early basal-like BC and normal breast tissue were further explored using WGCNA method. The genes in the WGCNA module with highest correlation value and most significance were further intersected with the DEGs. The genes those were overlapped in DEGs and WGCNA module were defined as IGs.

Functional analysis of IGs

The gene ontology (GO) and Kyoto encyclopedia of genes and genomes （KEGG） pathway enrichment analysis of IGs were carried out on The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/). The GO functional analysis inclued three categories: biological process (BP), cellule components (CC) and molecule functional (MF). The terms with a *P* value≦ 0.05 were considered as significant.

Identifying the hub IGs and evaluating their prognostic effect

Before identifying the hub IGs, the protein-protein internetwork of IGs was constructed with the String website (https://www.string-db.org/) and Cytoscape V3.7.2. Then the top 10 hub IGs were selected by applying the Maximal Clique Centrality (MCC) method in Cytoscape software. The hub IGs significantly associated with the prognosis were screened out by using the survival package on R software with the criteria of *P* value 0.05 and Hazard ratio 1. Furthermore, the expression levels of prognosis associated IGs were validated in TIMER dataset (http://cistrome.dfci.harvard.edu/TIMER/) and their relationship with immune cells infiltration in basal-like BC was also analyzed.

Results

IGs identified in early basal-like BC

The heatmap of gene expression difference between 87 cases of early basal-like BC and 111 case of normal breast tissues was presented in **Figure 1A.** Then total of 1465 DEGs were identified in the early basal-like BC as shown in **Figure 1B**. After analyzed by WGCNA, 16 modules were obtained (**Figure 2**). The turquoise module with highest correlation coefficient (cor.=0.98, *P v*alue < 1.0e-200) and most significance was selected for the following analysis (**Figure 2**). By intersecting the DEGs and the genes in turquoise module, 601 IGs were identified (**Figure 3B**).

Functional analysis of IGs

The functional analysis result of IGs was presented in **Table 1**. In the BP category, IGs were mainly enriched in cell division, mitotic nuclear division and cell proliferation. In the CC category, IGs were mainly enriched in nucleus, cytoplasm and cytosol. In the MF category, IGs were mainly enriched in protein binding, ATP binding and poly(A) RNA binding. IGs were mostly enriched in the pathways including cell cycle, p53 signaling pathway and DNA replication.

Hub IGs selecting and PPI network building

A PPI network with 519 nodes and 520 edges was built (**Suppl Figure 1**). The PPI enrichment *P* value and average local clustering co-efficient was < 1.0e-16, 0.338, respectively. Then top ten hub genes were selected by MCC method, namely CCNB1, CCNA2, CDK1, CDC20, TOP2A, BUB1, AURKB, CCNB2, KIF11 and ASPM (**Figure 3B**).

Overall survival analysis of hub genes and validation

The prognostic affection of the above hub genes in early basal-like BC were explored by applying the TCGA clinical data. The overall survival (OS) analysis revealed that only AURKB was found to be significant (*P* value =0.029) (**Figure 4**). The TIMR dataset demonstrated that comparing with the normal breast tissues, the expression level of AURKB was higher in basal-like BC (**Figure 5A**). The immune cells infiltration analysis displayed that in basal-like BC tissue, the infiltration numbers of CD4+ T cell (Rho=0.162, *P* value =0.0322) and naïve CD8+ T cell(Rho=0.163, *P* value =0.0314) were positively correlated with AURKB expression level, while the infiltration numbers of naïve B cell(Rho=-0.159, *P* value

=0.0362) and macrophage M2 cell(Rho=-0.191, *P* value =0.0116) were negatively correlated with AURKB expression level (**Figure 5B-E**).

Discussion

Although the 5-year survival rate of BC has exceeded 80%, that of basal-like BC is far lower, which is the worst type of BC and accounts for 15% of BCs[2]. The basal-like BC represents a group that is based on the PAM50 classification showing extremely low expression rates of hormone receptors and HER-2 as well as CK5/6 expressing[13]. To a certain extent, it has an intersection with another concept triple-negative breast cancer (TNBC). Reports shows that a higher risk of recurrence and metastasis has been found in basal-like BC in the first few years of being diagnosed than other types. Nevertheless, the early intervention applied to basal-like BC could result in meaningful impacts on the prognosis of basal-like BC[14-16].

With the advancement of precision medicine, targeted drugs have gradually emerged, which are commonly used in the breast malignment tumor. For example, BRCA1 mutations are more common in breast tumors[16]. As the result, studies about poly ADP-ribose polymerase (PARP) inhibitor are constantly carried out and eventually the medicine approved by the Food and Drug Administration (FDA) and guidelines[17]. However, the basal-like BC is lack of effective targeted drugs. Therefore, it is eager to find markers that can be related to the prognosis of early breast cancer at the genetic level and explore novel treatment targets.

We screened the gene expression data of early basal-like BC in the TCGA database and identified 1465 DEGs. Then, the WGCNA method was subsequently used to explore the difference in gene expression between the early basal-like breast cancer tissue and the normal control tissue. After taking the intersection with the DEGs, 601 IGs were obtained which were mostly related to the process of cell division and proliferation. Then we had identified the top ten hub genes, namely CCNB1, CCNA2, CDK1, CDC20, TOP2A, BUB1, AURKB, CCNB2, KIF11 and ASPM. However, only AURKB was found to be associated with OS in early basal-like breast cancer. Moreover, we also found that the expression level of AURKB was related to the infiltration state of immune cells in the basal-like BC microenvironment.

AURKB as a serine/threonine kinase is located at 17p13.1. It belongs to the same family together with AURKA and AURKC. The AURKB participates in the regulation of cell mitosis including the regulation of spindle function and the aggregation and separation of chromosomes. Because of its unstable position at the chromosome, abnormal expression is found frequently. The AURKB was reported to be overexpressed and amplificated in various types of malignant tumor tissue. The overexpression of AURKB might resulted in the low expression of pro-apoptotic proteins and promote cell proliferation[18]. A series of studies in osteosarcoma showed that AURKB might activate intracellular signaling pathways and lead to disease progression[19]. Furthermore, the high expression of AURKB was found to be associated with poor prognosis in renal clear cell carcinoma[20]. Similar results also appeared in gastric cancer and colorectal cancer[20-22]. In addition, in non-small cell lung cancer, Ahmed not only proved that

AURKB was associated with prognosis, but also discovered the inverse correlation between the expression of AURKB mRNA and cancer drug resistance[23].

In breast cancer, researchers had applied the immunohistochemical methods to explore the association between AURKB and clinicopathological parameters. AURKB was also reported to be related to chemotherapy resistance in patients undergoing neoadjuvant chemotherapy[24]. In addition, a study revealed that the single-nucleotide polymorphisms in AURKB interrelated with tumor risk and survival in TNBC[25]. The basal-like BC is characterized by high mortality, earlier recurrence and metastasis[14]. It is particularly important to intervene and evaluate the basal-like BC in the early stage[26]. Thus, in this study, we had specifically focused on investigate the AURKB in early stage basal-like BC.

Conclusions

Although we have analyzed and proved the relationship between the AURKB and the prognosis of early basal-like breast cancer through the TCGA database, the detailed and precise molecular biological mechanism still needs to be verified.

Abbreviations

BC: Breast Cancer

TCGA: The Cancer Genome Atlas

WGCNA: Weighted Gene Co-Expression Network Analysis

IGs: Intersected Genes

CCNB1: Cyclin B1

CCNA2: Cyclin A2

CDK1: Cyclin dependent kinase 1

CDC20: Cell division cycle 20

TOP2A: DNA topoisomerase II alpha

BUB1 mitotic checkpoint serine/threonine kinase (BUB1),

AURKB: Aurora kinase B

CCNB2: Cyclin B2

KIF11: Kinesin family member 11

ASPM: Assembly factor for spindle microtubules

DEGs: Differentially expressed genes

GO: Gene ontology

KEGG: Kyoto encyclopedia of genes and genomes

DAVID: Database for Annotation, Visualization and Integrated Discovery

BP: Biological process

CC: Cellule components

MF: Molecule functional

MCC: Maximal Clique Centrality

OS: Overall survival

TNBC: Triple-negative breast cancer

PARP: Poly ADP-ribose polymerase

FDA: Food and Drug Administration

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files

Competing interests

The authors declare that they have no competing interests

Funding

None

Authors' contributions

Keyu Yuan conducted a preliminary analysis and was a major contributor in writing the manuscript. Min Wu acquired and analyzed the data. Xue Yu completed further analysis of the data and participated in manuscript writing. Xia Zhao occupied a position in data analysis. Yu Feng played a role in data collection and article editing. Yanping Li designed the study. Shuzhen Lv put forward the entry point for breast cancer research and supervise the editing of the entire article. All authors read and approved the final manuscript

Acknowledgements

Not applicable

References

1. Li N, Deng Y, Zhou L, et al. Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to 2017: results from the Global Burden of Disease Study 2017. J Hematol Oncol. 2019;12(1):140.

2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.

3. Global Burden of Disease Cancer C, Fitzmaurice C, Akinyemiju TF, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol. 2018;4(11):1553-1568.

4. Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2020;70(4):313.

5. Samavat H, Kurzer MS. Estrogen metabolism and breast cancer. Cancer Lett. 2015;356(2 Pt A):231-243.

6. Sun YS, Zhao Z, Yang ZN, et al. Risk Factors and Preventions of Breast Cancer. Int J Biol Sci. 2017;13(11):1387-1397.

7. Houssami N, Turner RM, Morrow M. Meta-analysis of pre-operative magnetic resonance imaging (MRI) and surgical treatment for breast cancer. Breast Cancer Res Treat. 2017;165(2):273-283.

8. von Minckwitz G, Procter M, de Azambuja E, et al. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. N Engl J Med. 2017;377(2):122-131.

9. McDonald ES, Clark AS, Tchou J, Zhang P, Freedman GM. Clinical Diagnosis and Management of Breast Cancer. J Nucl Med. 2016;57 Suppl 1:9S-16S.

10. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.

11. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-752.

12. Sporikova Z, Koudelakova V, Trojanec R, Hajduch M. Genetic Markers in Triple-Negative Breast Cancer. Clin Breast Cancer. 2018;18(5):e841-e850.

13. Prat A, Karginova O, Parker JS, et al. Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. Breast Cancer Res Treat. 2013;142(2):237-255.

14. Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast. 2015;24 Suppl 2:S26-35.

15. Alexandrou S, George SM, Ormandy CJ, Lim E, Oakes SR, Caldon CE. The Proliferative and Apoptotic Landscape of Basal-like Breast Cancer. Int J Mol Sci. 2019;20(3).

16. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010;363(20):1938-1948.

17. Griguolo G, Dieci MV, Guarneri V, Conte P. Olaparib for the treatment of breast cancer. Expert Rev Anticancer Ther. 2018;18(6):519-530.

18. Bertran-Alamillo J, Cattan V, Schoumacher M, et al. AURKB as a target in non-small cell lung cancer with acquired resistance to anti-EGFR therapy. Nat Commun. 2019;10(1):1812.

19. Pi WS, Cao ZY, Liu JM, et al. Potential Molecular Mechanisms of AURKB in the Oncogenesis and Progression of Osteosarcoma Cells: A Label-Free Quantitative Proteomics Analysis. Technol Cancer Res Treat. 2018;18:1533033819853262.

20. Liu Q, Zhang X, Tang H, et al. Bioinformatics Analysis Suggests the Combined Expression of AURKB and KIF18B Being an Important Event in the Development of Clear Cell Renal Cell Carcinoma. Pathol Oncol Res. 2020;26(3):1583-1594.

21. Enjoji M, Iida S, Sugita H, et al. BubR1 and AURKB overexpression are associated with a favorable prognosis in gastric cancer. Mol Med Rep. 2009;2(4):589-596.

22. Pohl A, Azuma M, Zhang W, et al. Pharmacogenetic profiling of Aurora kinase B is associated with overall survival in metastatic colorectal cancer. Pharmacogenomics J. 2011;11(2):93-99.

23. Al-Khafaji AS, Davies MP, Risk JM, et al. Aurora B expression modulates paclitaxel response in non-small cell lung cancer. Br J Cancer. 2017;116(5):592-599.

24. Zhang Y, Jiang C, Li H, et al. Elevated Aurora B expression contributes to chemoresistance and poor prognosis in breast cancer. Int J Clin Exp Pathol. 2015;8(1):751-757.

25. Liao Y, Liao Y, Li J, Li J, Fan Y, Xu B. Polymorphisms in AURKA and AURKB are associated with the survival of triple-negative breast cancer patients treated with taxane-based adjuvant chemotherapy. Cancer Manag Res. 2018;10:3801-3808.

26. Valastyan S, Weinberg RA. Tumor metastasis: molecular insights and evolving paradigms. Cell. 2011;147(2):275-292.

Table

Table 1. The functional analysis of the intersected genes in early stage of basal-like breast cancer

Category	Term	Count	%	*P* Value
BP	GO:0051301~cell division	65	11.16838	3.48E-30
	GO:0007067~mitotic nuclear division	48	8.247423	4.53E-23
	GO:0000082~G1/S transition of mitotic cell cycle	28	4.810997	1.19E-17
	GO:0006260~DNA replication	32	5.498282	2.57E-16
	GO:0007062~sister chromatid cohesion	25	4.295533	1.93E-14
	GO:0007059~chromosome segregation	20	3.436426	3.00E-13
	GO:0008283~cell proliferation	42	7.216495	4.71E-12
	GO:0000070~mitotic sister chromatid segregation	12	2.061856	1.15E-10
	GO:0006270~DNA replication initiation	13	2.233677	1.55E-10
CC	GO:0005654~nucleoplasm	179	30.75601	1.46E-23
	GO:0005829~cytosol	187	32.13058	2.44E-18
	GO:0005634~nucleus	255	43.81443	1.99E-15
	GO:0000775~chromosome, centromeric region	19	3.264605	4.79E-14
	GO:0030496~midbody	25	4.295533	1.17E-12
	GO:0000777~condensed chromosome kinetochore	20	3.436426	1.40E-11
	GO:0005819~spindle	22	3.780069	1.20E-10
	GO:0000922~spindle pole	20	3.436426	8.53E-10
	GO:0005737~cytoplasm	227	39.00344	1.30E-09
MF	GO:0005515~protein binding	383	65.80756	1.72E-19
	GO:0005524~ATP binding	86	14.77663	1.20E-07
	GO:0044822~poly(A) RNA binding	70	12.02749	1.58E-07
	GO:0008017~microtubule binding	24	4.123711	2.45E-07
	GO:0003682~chromatin binding	33	5.670103	1.27E-06
	GO:0019901~protein kinase binding	30	5.154639	1.25E-05
	GO:0042393~histone binding	15	2.57732	3.64E-05
	GO:0042802~identical protein binding	46	7.90378	3.90E-05
	GO:0042803~protein homodimerization activity	44	7.560137	9.25E-05
	GO:0046978~TAP1 binding	4	0.687285	1.28E-04

KEGG	hsa04110: Cell cycle	37	6.357388	3.35E-22
	hsa03030: DNA replication	11	1.890034	8.37E-07
	hsa05166: HTLV-I infection	26	4.467354	2.19E-05
	hsa04114: Oocyte meiosis	14	2.405498	4.18E-04
	hsa05203: Viral carcinogenesis	20	3.436426	4.71E-04
	hsa05200: Pathways in cancer	26	4.467354	0.012708
	hsa04914: Progesterone-mediated oocyte maturation	11	1.890034	0.002208
	hsa05168: Herpes simplex infection	17	2.920962	0.002399
	hsa04115: p53 signaling pathway	8	1.37457	0.01634

BP, biological process, CC, cellule components, MF, molecule functional, KEGG, Kyoto encyclopedia of genes and genomes

Figures

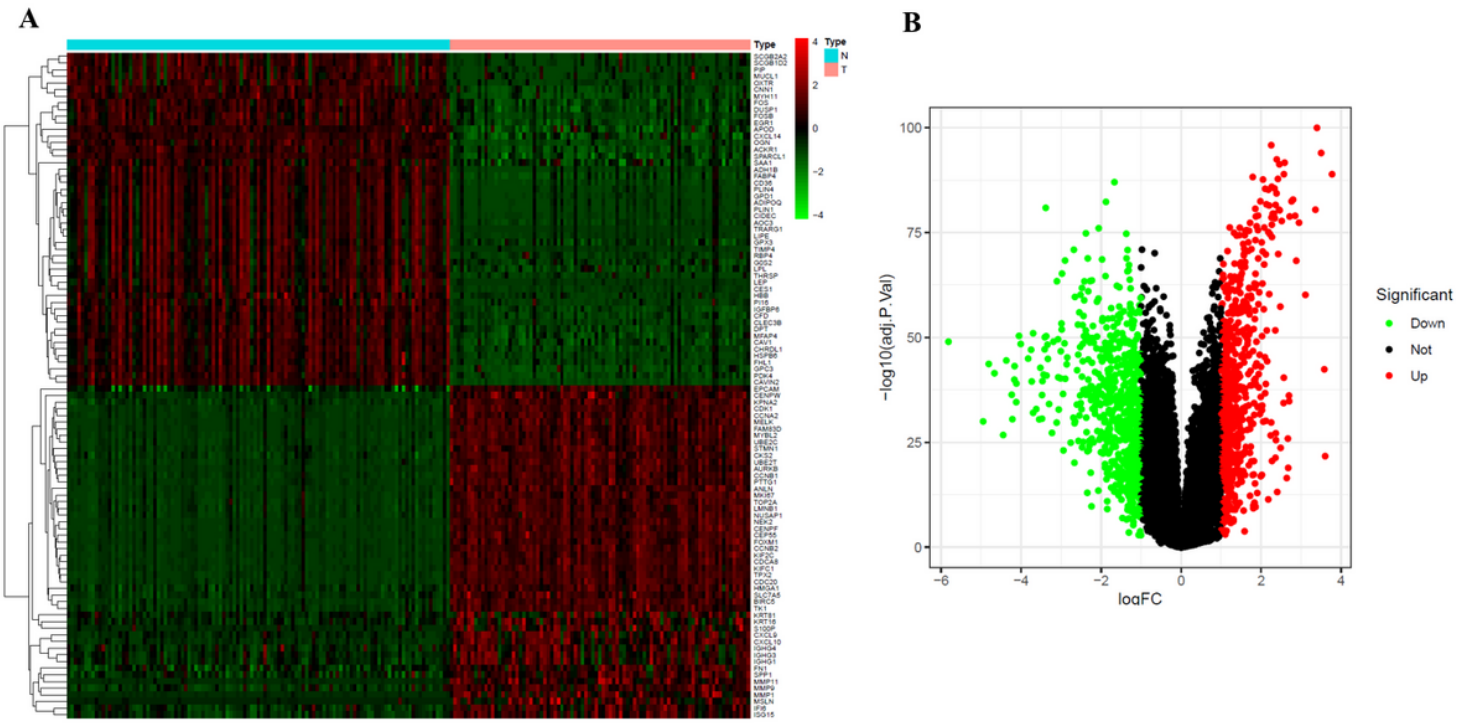

Figure 1

The heatmap and volcano plot of the differentially expressed genes in early stage basal-like breast cancer. Type: N, normal tissues, T, breast cancer. Green, low expression. Red, high expression.

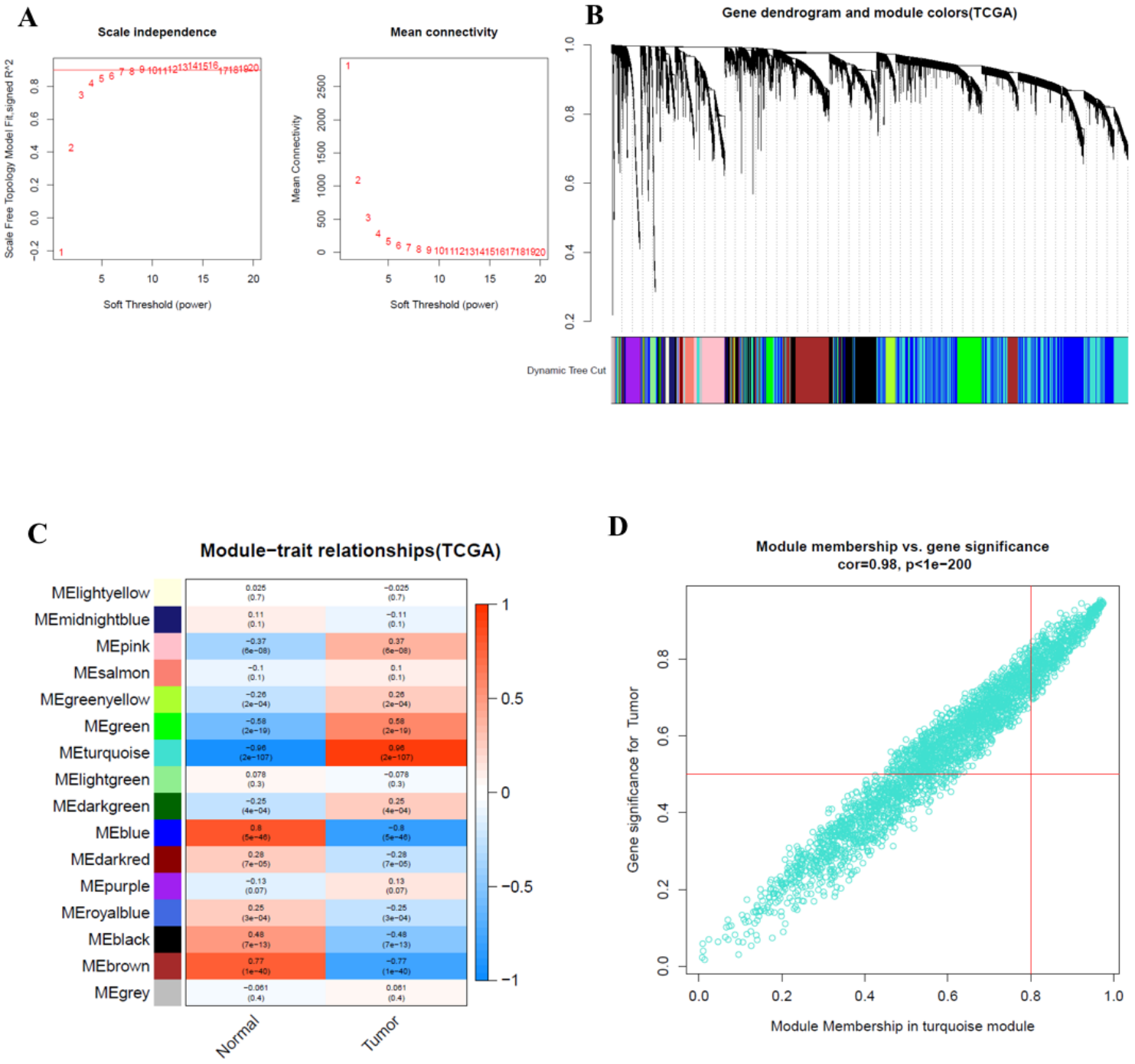

Figure 2

The weighted gene co-expression network analysis result of the early stage basal-like breast cancer. A. scale independence and mean connectivity. B. gene dendrogram. C. Trait modules. D. module membership vs gene significance in turquoise module.

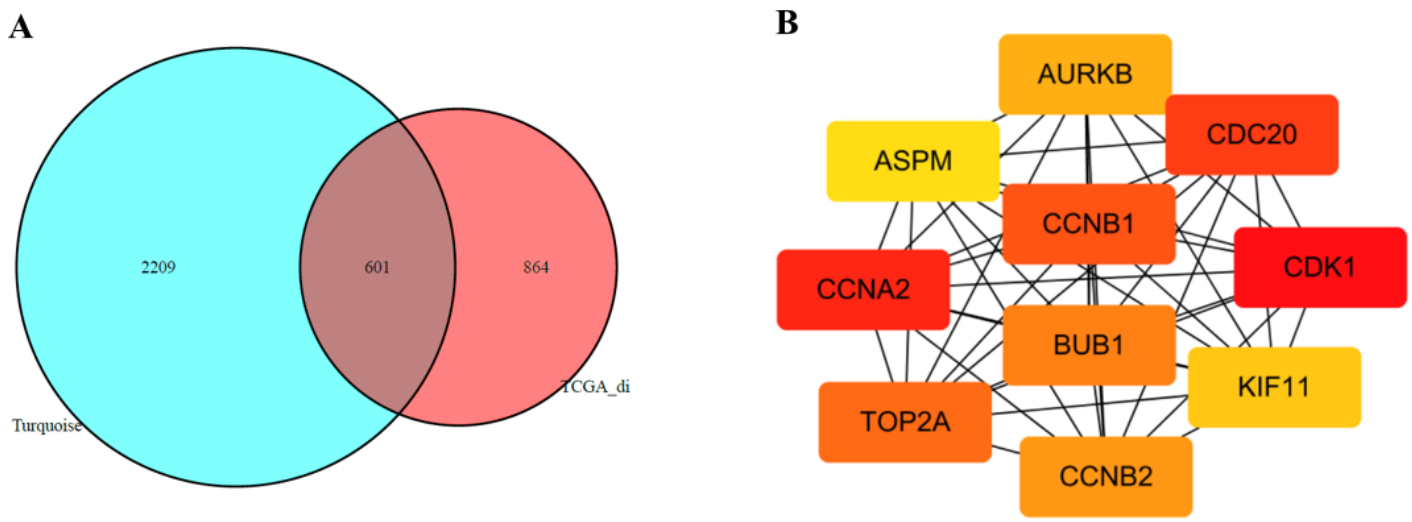

Figure 3

The intersected genes and hub genes selecting. A. Venn diagram of genes. B. hub genes.

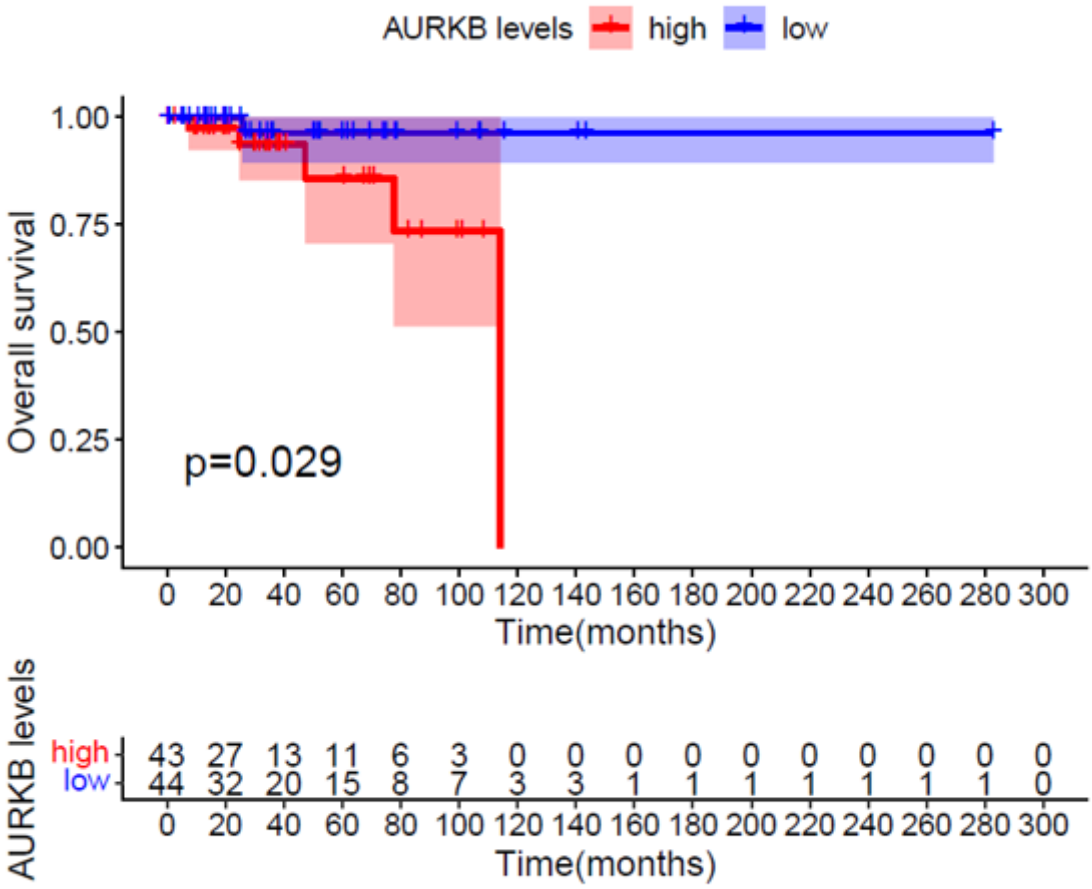

Figure 4

The overall survival analysis of AURKB in early stage basal-like breast cancer.

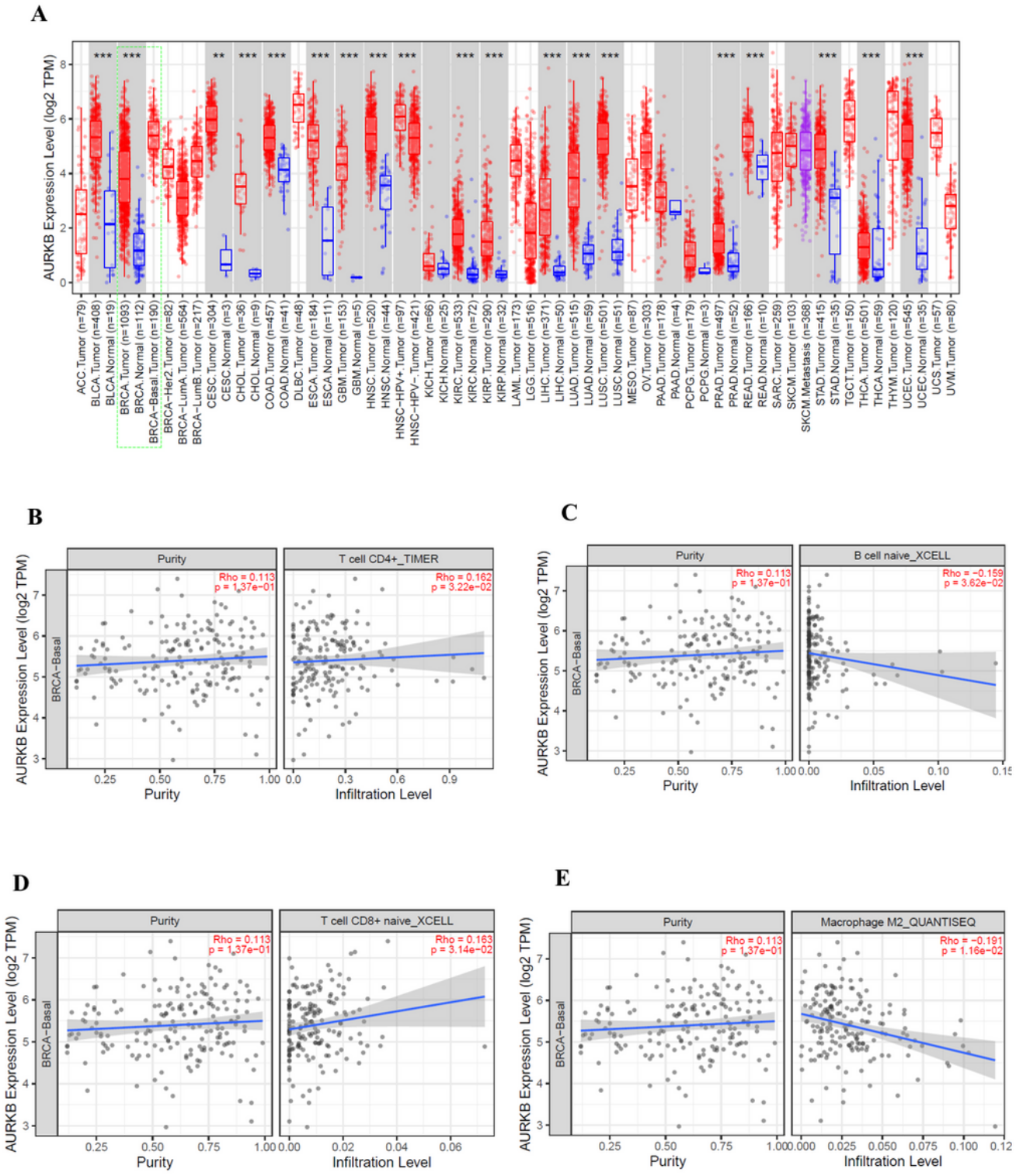

Figure 5

The expression of AURKB in cancers and immune cells infiltration analysis of AURKB in basal-like breast cancer. A. AURKB expression in various types of cancers. BRCA, breast cancer. B. CD4+ T cell. C. naïve CD8+ T cell. D. naïve B cell. E. macrophage M2 cell.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial1.docx