ceRNA Network Development and Tumour-infiltrating Immune Cell Analysis in Metastatic Breast Cancer of Bone

Shuzhong Liu
Peking Union Medical College Hospital
An Song
Peking Union Medical College Hospital
Xi Zhou
Peking Union Medical College Hospital
Zhen Huo
Peking Union Medical College Hospital
Siyuan Yao
Peking Union Medical College Hospital
Yong Liu
Peking Union Medical College Hospital
Yipeng Wang (ypwang_pumch@163.com)
Peking Union Medical College Hospital

Research

Keywords: breast cancer, bone metastasis, ceRNA network, immune infiltration, prognosis, nomogram

Posted Date: May 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-31393/v1

Abstract

Background: Advanced breast cancer commonly metastasises to the bone and the molecular mechanism explaining the bone affinity of breast cancer cells is unclear. Thus, we developed nomograms based on a competing endogenous RNA (ceRNA) network and analysed tumour-infiltrating immunecells to elucidate the molecular pathways that may predict the prognosis of breast cancer patients.

Methods: We obtained the RNA expression profile of 1091 primary breast cancer samples from The Cancer Genome Atlas database, 58 of which had bone metastasis. We analysed differential RNA expression patterns between breast cancer with and without bone metastasis and developed a ceRNA network. Cibersort was employed to differentiate between immune cell types based on tumour transcripts. Nomograms were then established using the ceRNA network and immune cell analysis. The value of prognostic factors was evaluated by Kaplan-Meier survival analysis and Cox proportional risk model.

Results: There were significant differences in lncRNAs, 18 miRNAs, and 20mRNAs between breast cancer with and without bone metastasis, which were used to construct a ceRNA network. We found that the protein-coding genes gjb3, cammv, ptprz1,and fbn3 were significant in our Kaplan-Meier analysis. We also observed significant differences in plasma cell and follicular helper T cell populations between the two groups. In addition, the proportions of mast cells, gamma delta T cells, and plasma cells differed depending on disease location and stage. Our analysis revealed that a high proportion of follicular helper T cells and a low proportion of eosinophils promoted survival and that dlx6-as1, wnt6,and gabbr2expression may be related to bone metastasis of breast cancer.

Conclusions: We provided a bioinformatic basis for exploring the molecular mechanism of bone metastasis in breast cancer patients and identified factors that may predict this.

Background

Breast cancer has the highest prevalence amongst all cancers and is leading the cause of death due to cancer in women [1]. Advanced stages of breast cancer often metastasises to the bone and studies have shown that 47–85% of patients with breast cancer will experience bone metastasis [2]. The lowest rate of bone metastasis has been found in tumours that were negative for estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) [3, 4]. Conversely, increased bone metastasis rates are observed in $HER2^{+}$, $ER^{+}/HER2^{-}/Ki67^{Hi}$, and $ER+/HER2^{-}/Ki67^{Low}$ tumours [3, 4]. Frequently, bone metastasis occurs in the spine, rib, pelvis, proximal femur, and skull. Bone deterioration in these areas lead to serious complications such as pain, fractures, hypercalcaemia, spinal cord compression, and other nerve compression symptoms, which may be fatal and significantly reduce the quality of life of patients [1–5]. The process of bone metastasis is complex and has multiple stages. It requires the separation of breast cancer cells from the primary tumour site, transport to the bone via blood or lymph, and survival and proliferation in the target bone tissue [6–8]. Genomics research has shown that every

step of metastasis is related to a series of sub-events. However, molecular mechanisms of bone metastasis in breast cancer is not fully understood [9–11].

In this study, we developed a complete protein interaction network by constructing a ceRNA network and gene expression profile of breast cancer patients with and without bone metastasis to examine the underlying molecular mechanisms. Prediction of bone metastasis and subsequent prognosis of breast cancer may be related to both the ceRNA network and the type of tumour-infiltrating immune cells, thus we obtained the gene expression profiles of breast cancer patients from The Cancer Genome Atlas (TCGA) and applied the Cibersort algorithm. This was then used to establish a predictive nomogram. In addition, we evaluated the internal relationship between tumour-infiltrating immune cells and ceRNA networks, which provided clues to the molecular mechanisms and clinical predictors of breast cancer metastasis.

2. Materials And Methods

2.1 Data collection

Breast cancer RNA-seq raw count and fragments per kilobase per million mapped reads (FPKM) data and miRNA-seq data were downloaded from the database of TCGA. Annotation information was downloaded from genecode. The microRNA (miRNA) expression profile included 503 genes and 1078 samples (bone metastasis: 58; non metastasis: 1020). The messenger RNA (mRNA) expression profile included 1925 genes and 1091 samples (bone metastasis: 58; non metastasis: 1033). The long non-coding RNA (lncRNA) expression profile includes 1925 genes and 1091 samples (bone metastasis: 58; non metastasis: 1033).

2.2 Differential Gene Acquisition, Expression Analysis And Functional Annotation

We collected the RNA-seq raw count and fpkm fragments of 1091 primary breast cancer samples, of which 58 had bone metastasis. We also retrieved demographic information and survival endpoints of patients. We filtered out the genes that were not specific to breast cancer and analysed the differential of RNA expression between bone metastatic breast cancer and non-bone metastasis breast cancer by Differentially expressed seq2 (DEseq2). Up or downregulated genes were defined as FDR-adjusted $P < 0.05$ and log fold change > 1.0 or < -1.0. DEseq2 package was specially designed for RNA-seq raw count differential expression analysis of R package and used DEseq2 package to calculate the different genes of bone metastasis samples and non-bone metastasis samples, threshold: logFC = 1, P = 0.05. Relevant data provided by TCGA is public. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of DEmRNAs were carried out with David online tool. Terms and pathways with $P < 0.05$ were selected.

2.3 Construction of ceRNA network in breast cancer bone metastasis

miRNA-mRNA and lncRNA-miRNA interaction were predicted using miRTarBase and lncbase v.2 experimental module, respectively. Search for miRNA with DElncRNA as target in lncbase database, and then search for target genes of these miRNAs in miRTarBase, and intersect these target genes with DEmRNA. Then, the Pearson correlation coefficient (PCC) was calculated and the correlation between lncRNA and mRNA was screened. We used Cytoscape v.3.5.1 to select the miRNAs that regulate lncRNAs and mRNAs and were significant in the hypergeometric test and correlation analysis, which was used for the visualising the ceRNA network.

2.4 Clinical significance of ceRNA network in bone metastasis of breast cancer

Survival analysis and Cox risk regression model analysis were performed for the genes in the ceRNA network to obtain the key genes related to prognosis. The differences of gene expression in different TNM stage samples were analyzed.

2.5 Model Building

We randomised 1091 breast cancer samples into two groups. 75% of the samples were used as training set and 25% as test set. Lasso regression model was constructed with training set, and the best lambda value and gene set were obtained. Then, the life time of the test set is predicted by the constructed model, and receiver operating characteristic curve (ROC curve) and calibration curve are drawn. Lasso regression model is constructed with glmnet package. ROC curve is drawn with timeROC package.

The Cox risk regression model was constructed with these genes, and the nomogram was drawn with RMS package. In order to evaluate the prediction effect of nomogram, the calibration curve is drawn.

2.6 Cibersort Estimation

Cibersort software (http://cibersort.stanford.edu/) estimates the abundance of 22 different types of immune cells in the sample based on RNA-seq count data. We calculated the immune cell components of 1091 breast cancer samples with RNA-seq data. In order to explore the clinical significance of different samples with different proportion of immune cells, for a certain cell, all samples are sorted according to the cell proportion, with the median as the dividing line. The samples were categorised as high- or low-proportion and samples with $P < 0.05$ were included in the Kaplan-Meier survival analysis.

2.7 Survival analysis and nomogram of central players in tumour immunity

The impact of the immune cell types on the prognosis of breast cancer patients was examined by Kaplan-Meier survival analysis and Cox regression. Furthermore, the Wilcoxon rank-sum test evaluated the relationship between immune cell subtypes, the occurrence of metastasis, and TNM stage. In the original Cox model, immune cells showed a significant correlation with prognosis and were selected to establish nomogram. We quantified the ROC and Area under curve (AUC) to evaluate the sensitivity and specificity of our diagnostic and prognostic models. The accuracy of the predictive capabilities of the nomogram was evaluated by the calibration curve and the consistency index. The relationship between ceRNAs and the 22 types of immune cells were investigated using Pearson's correlation coefficient. Statistical significance was determined by a two-sided $P < 0.05$. R version 3.5.1 (Institute of Statistics and Mathematics, Austria) with the following packages was used for statistical analyses: GDCRNATools, edgeR, ggplot2, rms, glmnet, preprocessCore, survminer, and timeROC.

3. Results

3.1 Identification and expression analysis of genes with significant differences

We identified differentially expressed genes with a log fold change of > 1.0 or < -1.0 and FDR of < 0.05. The differential expression of lncRNA, miRNA and mRNA (Fig. 1A-C) in breast cancer samples with and without bone metastasis was calculated by DESeq2 package. There are 183 up-regulated lncRNAs, 13 down regulated lncRNAs; 17 up-regulated miRNAs, 0 down regulated miRNAs; 520 up-regulated mRNAs, 26 down regulated mRNAs (Fig. 1G). Figure 1D-F shows the expression profiles of differentially expressed genes.

3.2 Functional Enrichment` Analysis Of Significantly Different Genes

Figure 2 shows the enrichment of the top 10 GO terms and pathways of DEmRNA. We can find that the main functional areas of gene enrichment are: go: 0007267 ~ cell signalling, go: 0070098 ~ chemokine mediated signalling, go: 0010628 ~ position regulation of gene expression, go: 0010942 ~ position regulation of cell death, go: 0010737 ~ protein kinase A signalling, and go: 0008284 ~ position regulation of cell promotion. The main pathways of gene enrichment are: hsa04080: neuroactive living receptor interaction, hsa04062: chemokine signalling, hsa05204: chemical carcinogenesis, hsa00982: drug metabolism - cytochrome P450, hsa00350: tyrosine metabolism and other tumour related signalling pathways.

3.3 Construction Of Protein Interaction Network Of Differential Genes

We input the differential gene into the string protein database, and the score threshold is 0.9 (the highest confidence). The final protein interaction network has 183 nodes, 491 edges, PPI enrichment P < 1.0e-16. See Fig. 3 for details. The average degree of each node is 5.37. This suggests that there is a complex regulatory relationship between the different genes involved in breast cancer with bone metastasis.

3.4 Construction And Survival Analysis Of Cerna Network

The shared miRNAs of differentially expressed mRNA and lncRNA were obtained in the LncBase and miRTarBase databases, and then the Pearson's correlation coefficient of lncRNA-mRNA pairs with shared miRNA was calculated, and the positive expressed lncRNA-mRNA pairs were screened (Table 1). Using Cytoscape software to construct the ceRNA network (Fig. 4A,B), we got the ceRNA network related to breast cancer with bone metastasis, including 20 mRNAs, 18 miRNAs, 2 lncRNAs, 48 edges. We analysed the gene batch survival in the ceRNA network. The results showed that jgb3, camgv, ptprz1, fbn3 mRNA were related to prognosis (Fig. 4C-F).

Table 1
lncRNA-miRNA-mRNA pairs

miRNA	lncRNA	mRNA	P value	PCC
hsa-let-7b-5p	DLX6-AS1	IGF2BP2	3.62E-32	0.347
hsa-let-7b-5p	DLX6-AS1	IGF2BP3	2.00E-61	0.471
hsa-let-7f-5p	DLX6-AS1	KLK6	2.52E-27	0.32
hsa-mir-1-3p	DLX6-AS1	GJB3	7.91E-41	0.389
hsa-mir-1-3p	DLX6-AS1	SOX6	1.75E-85	0.545
hsa-mir-124-3p	DLX6-AS1	OCA2	1.92E-46	0.414
hsa-mir-124-3p	DLX6-AS1	GABBR2	3.18E-52	0.437
hsa-mir-124-3p	DLX6-AS1	BARX1	1.49E-26	0.315
hsa-mir-124-3p	DLX6-AS1	PRDM13	4.99E-29	0.329
hsa-mir-124-3p	DLX6-AS1	PTPRZ1	1.09E-31	0.344
hsa-mir-132-3p	DLX6-AS1	FBN3	7.32E-56	0.451
hsa-mir-148b-3p	DLX6-AS1	CCKBR	1.28E-31	0.344
hsa-mir-148b-3p	DLX6-AS1	DLX6	0	0.884
hsa-mir-155-5p	AFAP1-AS1	SOX6	2.52E-25	0.308
hsa-mir-16-5p	DLX6-AS1	ALKAL2	7.95E-44	0.403
hsa-mir-16-5p	DLX6-AS1	SOX6	1.75E-85	0.545
hsa-mir-16-5p	DLX6-AS1	KLHL34	2.04E-28	0.326
hsa-mir-16-5p	DLX6-AS1	CAMKV	4.92E-25	0.306
hsa-mir-181a-5p	DLX6-AS1	PTPRZ1	1.09E-31	0.344
hsa-mir-181a-5p	DLX6-AS1	OCA2	1.92E-46	0.414
hsa-mir-195-5p	DLX6-AS1	CAMKV	4.92E-25	0.306
hsa-mir-221-3p	DLX6-AS1	FBN3	7.32E-56	0.451
hsa-mir-26a-5p	DLX6-AS1	CAMKV	4.92E-25	0.306
hsa-mir-27a-3p	DLX6-AS1	RNF182	2.35E-44	0.405
hsa-mir-30b-5p	DLX6-AS1	CAMKV	4.92E-25	0.306
hsa-mir-30c-5p	DLX6-AS1	CAMKV	4.92E-25	0.306
hsa-mir-320a	DLX6-AS1	POLR2F	2.15E-47	0.418

miRNA	lncRNA	mRNA	P value	PCC
hsa-mir-320a	DLX6-AS1	IGF2BP3	2.00E-61	0.471
hsa-mir-7-5p	DLX6-AS1	IL12RB2	6.13E-50	0.428
hsa-mir-9-5p	DLX6-AS1	WNT6	2.16E-32	0.348

3.5 Lasso Regression Analysis And Nomogram Construction

LncRNA and mRNA in the ceRNA network of bone metastasis were used to calculate the optimal lambda value (Fig. 5A, B) of lasso regression model, and the lambda = 0.0133 was obtained. Six genes (camgv, alkal2, gabbr2, Barx1, fbn3, and Wnt6) obtained by lambda = 0.0133 were used to construct a multiple Cox risk regression model (Fig. 5C). The six genes were used to construct a graph based on Cox regression (Fig. 5D), which predicted the 1-, 3-, and 5-year survival status. To evaluate the prediction effect of the nomogram model, the 1-, 3-, 5-year calibration curve (Fig. 5E) was drawn, and the results showed that nomogram performed well. The ROC curve (Fig. 5F) shows that the average ROC of the nomogram is 0.686, which has a high prediction accuracy. Lasso regression results show that all six genes are necessary for modelling. In addition, ROC and calibration curves showed acceptable accuracy (1-, 3-, and 5-year survival AUC were 0.746, 0.686, and 0.642, respectively) and the distinction of the nomogram.

3.6 Relationship Between Gene Expression Level And Tnm Stage

The relationship between mRNA and different TNM stages were analysed by t-test. We observed that fbn3 expression inversely correlated with T and N stages (Fig. 6A, B). Furthermore, high expression of alkal2 correlated with low T stage (Fig. 6C). The expression of gabbr2 also correlated with T and N stages, the higher the T and N stages, the lower the gene expression (Fig. 6D, E) and the expression of camgv significantly correlated with N stages (Fig. 6F).

3.7 Distribution of tumour-infiltrating immune cells in breast cancer with bone metastasis

In order to visualize, Fig. 7A only shows the immune cell components of 58 bone metastasis samples. The split violin diagram reveals the different immune cell proportions between breast cancer with and without bone metastasis (Fig. 7B). Among them, plasma cell and follicular helper T cell have significant differences in the two different samples.

3.8 Clinical Significance Of Immune Cell Components

We analysed the clinical correlation and prognosis of 22 kinds of immune cells. The results showed that there were significantly more mast cells present in M1/M0 (Fig. 8A, P *= 0.002*), stage IV/stage I (Fig. 8B, P *= 0.002*), and T4/T1 samples (Fig. 8C, P *= 0.008*). Similarly, gamma delta T cells were significantly overrepresented in M1/M0 (Fig. 8D, P *= 0.035*), and stage IV/stage I samples (Fig. 8E, P *= 0.05*). Survival analysis revealed that the samples with a low proportion of eosinophilic had better survival status (Fig. 8F, log-rank test, *P = 0.01*); the samples with a high proportion of follicular helper T cells (the proportion was higher than the median) had better survival status (Fig. 8G, log-rank test, *P = 0.025*).

3.9 Clinical correlation of immune cells and nomogram multiple Cox risk regression analysis

We made multivariate Cox regression on 22 kinds of immune cells, and finally showed that the model composed of activated mast cell, gamma delta T cell, activated dendritic cell, follicular helper T cell, Eosinophils and Neutrophils had the smallest Akaike information criterion (AIC) (Fig. 9A). According to the above clinical correlation analysis, six kinds of immune cells (activated mast cell, gamma delta T cell, activated dendritic cell, follicular helper T cell, eosinophils, neutrophils) were selected to construct multiple Cox risk regression and calculate the risk ratio. Then, the nomogram (Fig. 9B) of 1, 3 and 5 years was developed. The calibration curve (Fig. 9C) showed that our nomograms had an adequate predictive capacity and the product under the average ROC curve was 0.616 (Fig. 9D). Selection of the best threshold of ROC curve for 5-year survival status and division of the samples into a high-risk group and low-risk group showed a significant difference in survival status (*P < 0.001*, Fig. 9E, F) ROC curve and calibration curve showed that the nomogram was consistent and had good accuracy with 1-, 3-, and 5-year survival AUC of 0.549, 0.616, and 0.644. Multiple Cox regression analysis showed that the high-risk group differed significantly from the low-risk group.

3.10 Co-expression Analysis Of Immune Cells And Key Genes

Pearson's correlation coefficient indicated the correlation between the thermogram of RNAs (Fig. 10A) and 22 types of lymphocytes in breast cancer samples (Fig. 10B). Here we specifically analyzed the correlation between T-regulatory cells and Wnt6, Barx1, and found that T-regulatory cells and Wnt6 (Fig. 10C, R *= 0.11, P = 0.0027*), T-regulatory cells and Barx1 (Fig. 10D, R *= 0.072, P = 0.0017*) were positively correlated. We found that Wnt6 was positively correlated with KLK6 and gjb3. Wnt6 was positively correlated with fbn3 and gabbr2; dlx6 and dlx6-AS1 were positively correlated. Naïve $CD4^+$ T cells and gamma delta T cells positively correlated. Memory B cells negatively correlated with naïve B cells and plasma cells. $CD8^+$ T cells positively correlated with monocytes. M1 macrophages positively correlated with M2 macrophages.

3.11 Flow Chart

In order to make it easier for readers to understand all the analysis process of the article, we have drawn a flow chart (Fig. 11).

4. Discussion

Breast cancer originates from mammary epithelial cells, which is an aggressive tumour type with better prognosis [1–3]. Advanced breast cancer patients often develop distant metastasis in locations such as the bone. However, the underlying molecular mechanisms are still unknown. In the process of tumour occurrence and metastasis, key determining factors are the tumour's molecular and cellular features, which are often used in the clinic to determine the prognosis. Bone metastasis are detrimental to the quality of life and life span of breast cancer patients and brings serious complications including pain, fracture, spinal cord compression, and malignant hypercalcemia [1–4]. Due to the frequent occurrence of bone metastasis in patients with advanced breast cancer, the pathogenesis and clinical management of bone metastasis are important and challenging topics in basic research and clinical practice. The ceRNAs network, including mRNA, miRNA, and lncRNA, and immune cells that infiltrate the tumour may be critical to further understand this phenomenon [12–15]. We observed that tumour samples with bone metastasis had significantly altered proportions of infiltrating cells compared with breast cancer without bone metastasis. We then developed two models to predict the prognosis of breast cancer patients with bone metastasis. The resulting AUC of the two nomograms demonstrated their value in a clinical setting.

Lncrna is a type of ncRNA with over 200 nucleotides, which has nothing to do with protein-coding [12–16]. New evidence suggests that lncRNA imbalance occurs frequently in many malignant tumours and are vital for carcinogenesis through post-transcriptional regulation and epigenetic modification [17, 18]. Here, we bioinformatically analysed the ceRNA network regulating bone metastasis of breast cancer with 20 protein encoded mRNAs, 2 lncRNAs and 18 miRNAs. We found a significant correlation between four protein-encoding genes (gjb3, camgv, ptprz1, fbn3) and their associated miRNAs and lncRNAs with the survival of breast cancer patients with bone metastasis. The AUC values of 1-, 3-, and 5-year survival were 0.746, 0.686, and 0.642 respectively. Using a hypergeometric test and correlation analysis, we found a significant correlation between gjb3 (dlx6-as1, hsa-mir-1-3p), fbn3 (dlx6-as1, hsa-mir-132-3p), camkv (dlx6-as1, hsa-mir-16-5p), ptprz1 (dlx6-as1, hsa-mir-181a-5p), and the total survival rate of breast cancer patients with bone metastasis. These results indicated that dlx6-as1 may occur and be a key player for bone metastasis in advanced breast cancer patients. We speculate that dlx6-as1 may regulate the occurrence and progression of metastasis by interacting with Wnt/β-catenin signalling.

Recently, an increasing number of studies show that aberrant lncrnas expression leads to the development of many kinds of malignant tumours, including breast cancer [12–18]. As a kind of lncrnas, dlx6-as1 is believed to be carcinogenic by regulating the progression of renal cell carcinoma, liver cell carcinoma, glioma, pancreatic cancer and lung adenocarcinoma [19–26]. Normal brain tissue has a high expression of DLX6-AS1, which is involved in the development regulation [24–26]. In recent years, it has

been found that dlx6-as1 is abnormally expressed in a variety of tumour tissues and is closely related to a poor clinical outcome [19–26]. However, the molecular mechanisms of dlx6-as1 and how it contributes to the pathogenesis of breast cancer are still unclear. Zhao et al. demonstrated significant upregulation of dlx6-as1 in breast cancer tissues and cell lines [27]. Furthermore, high dlx6-as1 expression is linked to poor outcome regarding tumour size, lymph node metastasis, TNM stage, and survival of breast cancer patients [19–27]. SiRNA knockout of dlx6-as1 showed a reduction in proliferation, apoptosis, invasion, migration, and epithelial-mesenchymal transition (EMT) of breast cancer cells [25–27]. These findings suggest that the progression of breast cancer could be partly due to an overexpression of dlx6-as1. Studies have indicated that mRNA expression may be regulated by lncRNAs through competitive communication between cernas and miRNAs [28, 29]. For example, dlx6-as1 silencing inhibits cell proliferation, migration, and invasion in non-small cell lung cancer by interacting with mir-144. In addition, dlx6-as1 is important in the carcinogenesis of glioma by competing with mir-197-5p. In pancreatic cancer, knocking out the dlx6-as1 gene lead to inhibition of proliferation and metastasis of cancer cells through enhancement of mir-181b's endogenous effects. In this study, dlx6-as1 may be important for the expression and regulation of miRNAs in breast cancer bone metastasis. Zhao et al. evaluated the expression of dlx6-as1 in breast cancer and analysis of the correlation between dlx6-as1 expression and clinicopathological parameters showed increased dlx6-as1 expression in tumour tissue compared with normal tissue, which was linked to poor prognosis of breast cancer patients [27]. Similar to pancreatic cancer, Dlx6-as1 gene knockout reduces the proliferation, invasion, and migration capacity of breast cancer cells and promoted apoptosis [27]. Furthermore, luciferase analysis confirmed that dlx6-as1 is an endogenous mediator of mir-505-3p and negatively regulates its expression. In addition, mir-505-3p inhibits runt related transcription factor 2 (Runx2) expression by binding directly to the 3' untranslated region. Partial reversal of mir-505-3p's carcinogenic effects can be achieved by overpressing Runx2. Wang et al. additionally showed that dlx6-as1 triggers its downstream effects on breast cancer cells by downregulating Fus [23]. These findings indicate that dlx6-as1 promotes breast cancer progression and is consistent with our results. Targeting Dlx6-as1 may be beneficial for the treatment of breast cancer.

The regulatory mechanism between lncrnas and miRNAs is extraordinarily complex. Among them, lncrna can be used as the combination of ceRNA and miRNA to compete and share mRNA, thus forming a complex lncRNA-miRNA-mRNA network [30]. Wnt/β-catenin signalling is critical in numerous cellular processes such as proliferation, invasion, and migration [31]. The involvement of Wnt/β-catenin signalling has been demonstrated in several malignant tumours including breast cancer and Wnt/β-catenin activation promotes tumour cell growth and metastasis [31, 32]. Previous studies have shown that Wnt activation inhibits memory T cells by reducing key transcription factors produced by these cells [33]. In our study, we found the activation of mast cells was correlated with Wnt6 expression. Zhang et al proved that Wnt signal is often activated under the action of dlx6-as1 [34]. Therefore, we speculate that Wnt pathway may be important for the impact of dlx6-as1 on the composition of immune cells. However, the molecular mechanisms of dlx6-as1-induced Wnt activation remains to be elucidated.

Guo et al. demonstrated dlx6-as1 upregulation in cell lines and tissues from bladder cancer patients, which augmented the proliferation, invasion, and migration of these cells by regulating EMT process and

the activity of Wnt/β-catenin signalling [35]. dlx6-as1's impact on the characteristics of cancer stem cells in osteosarcoma has also been investigated and was found to positively correlate with more advanced disease and poorer survival [36]. Similar findings have been demonstrated in prostate cancer patients. Recently, several studies have shown the participation of Wnt6, Barx1, ptprz1, and other genes in the regulation of Wnt/β-catenin signalling and their involvement in the development of malignant tumours [31–41]. These findings support our hypothesis that dlx6-as1 may affect the mRNA expression of Wnt6, Barx1, ptprz1 and other genes through ceRNA network to regulate Wnt/β-catenin signalling and participate in the distant bone metastasis of breast cancer.

In this study, we used six genes (camgv, alkal2, gabbr2, Barx1, fbn3, and Wnt6) to construct a multiple Cox risk regression model in the process of Lasso regression. Using these six genes to construct the nomogram based on Cox regression can predict the survival status of one year, three years and five years, and the prediction effect of the model is satisfactory.

In the study of immune infiltration of breast cancer bone metastasis, we found that plasma cell and follicular helper T cell were significantly different in two different samples. We also observed that the proportion of mast cells, gamma delta T cells, plasma cells, follicular helper T cells, and eosinophils varied depending on disease stage and progression. These results suggest that plasma cell, follicular helper T cell, mast cell and gamma delta T cell have potential biological prediction value for predicting bone metastasis, disease grading and staging of breast cancer, which is expected to be further verified and applied in clinical diagnosis and treatment. Subsequently, our correlation analysis showed that Wnt6 and gabbr2 positively correlated with mast cell activation. Therefore, our research focuses on the above six genes and their related pathways.

Our research inevitably has several limitations. The data included in our study are from Western countries and may not be directly extrapolated to patients in Asian countries. We were unable to comprehensively analyse the clinical and pathological parameters due to limited public information, which may impact our analyses. Therefore, we minimised bias by investigating the gene and protein expression of key biomarkers at the cell and tissue level.

Tumour microenvironment often affects the process of tumour invasion and migration [25, 26, 42]. Invasion of tumour cells is largely dependent on the composition of the extracellular matrix as well as growth factors that are secreted by the surrounding cells [25, 26]. Furthermore, metastasis of the tumour may have occurred in the early stage of the tumour development and is not directly related to proliferation. Therefore, it is necessary to determine the molecular mechanism leading to aggressive breast cancer with bone metastasis. We developed a ceRNA network based on breast cancer samples and the nomogram of tumour-infiltrating immune cells predicted the prognosis of breast cancer patients with and without bone metastasis with high accuracy. The predictive nomogram can provide more comprehensive clinical data for individual treatment of breast cancer bone metastasis. Further studies should be performed to show the interactions and communications between cancer cells and immune

cells. In particular, the exosomes secreted by tumour cells contain ceRNAs and may also play a role mediating breast cancer metastasis.

Abbreviations

ceRNA
competing endogenous RNA; TCGA:The Cancer Genome Atlas; lncRNA:long non-coding RNA; miRNA:microRNA; mRNA:messenger RNA; ER:estrogen receptor; HER2:human epidermal growth factor receptor 2; FPKM:fragments per kilobase per million mapped reads; DE:Differentially expressed; GO:Gene ontology and KEGG:Kyoto Encyclopedia of Genes and Genomes; PCC:Pearson correlation coefficient; ROC curve:receiver operating characteristic curve; AUC:Area under curve; AIC:Akaike information criterion; EMT:epithelial-mesenchymal transition; Runx2:runt related transcription factor 2;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by National Natural Science Foundation of China (Project No. 81871746; Grant recipient: Y.W.), and Peking Union Medical College Graduate Student Innovation Fund (2018) (Project No. 2018-1002-02-08; Grant recipient: S.L.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors’ contributions

Conceptualization, Shuzhong Liu, An Song, Yong Liu and Yipeng Wang; Formal analysis, Shuzhong Liu, An Song, Xi Zhou, Zhen Huo, Siyuan Yao and Yong Liu; Funding acquisition, Shuzhong Liu and Yipeng Wang; Methodology, Shuzhong Liu, Siyuan Yao and Yipeng Wang; Software, Xi Zhou, Siyuan Yao and Yong Liu; Supervision, Yong Liu and Yipeng Wang; Writing – original draft, Shuzhong Liu and An Song; Writing – review & editing, Shuzhong Liu and Yipeng Wang.

Acknowledgements

We would like to thank our colleagues at the Department of Orthopaedic Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College.

Author details

[1]Department of Orthopaedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China.

[2]Department of Endocrinology, Key Laboratory of Endocrinology, National Health and Family Planning Commission, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Beijing, People's Republic of China.

[3]Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Beijing, China

References

1. Forouzanfar MH, Foreman KJ, Delossantos AM. **et al**. Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. Lancet. 2011;378:1461–84.
2. Nielsen OS, Munro AJ, Tannock IF. Bone metastases: pathophysiology and management policy. J Clin Oncol. 1991;9:509–24.
3. Hayashi N, Iwamoto T, Qi Y. **et al**. Bone metastasis-related signaling pathways in breast cancers stratified by estrogen receptor status. J Cancer. 2017;8(6):1045–52.
4. Arciero CA, Guo Y, Jiang R. **et al**. ER+/HER2 + Breast Cancer Has Different Metastatic Patterns and Better Survival Than ER-/HER2 + Breast Cancer. Clin Breast Cancer. 2019;19(4):236–45.
5. Kawaguchi T, Yan L, Qi Q. **et al**. Novel MicroRNA-Based Risk Score Identified by Integrated Analyses to Predict Metastasis and Poor Prognosis in Breast Cancer. Ann Surg Oncol. 2018;25(13):4037–46.
6. Chen X, Pei Z, Peng H. **et al**. Exploring the molecular mechanism associated with breast cancer bone metastasis using bioinformatic analysis and microarray genetic interaction network. Med (Baltim).

2018;97(37):e12032.

7. Johnstone CN, Pattison AD, Gorringe KL, **et al**. **Functional and genomic characterisation of a xenograft model system for the study of metastasis in triple-negative breast cancer**. **Dis Model Mech** 2018; **11**(**5**). **pii**: **dmm032250**.
8. Erin N, Ogan N, Yerlikaya A. Secretomes reveal several novel proteins as well as TGF-β1 as the top upstream regulator of metastatic process in breast cancer. Breast Cancer Res Treat. 2018;170(2):235–50.
9. Gangoda L, Liem M, Ang CS. **et al**. Proteomic Profiling of Exosomes Secreted by Breast Cancer Cells with Varying Metastatic Potential. Proteomics. 2017;17:23–4.
10. Lawler K, Papouli E, Naceur-Lombardelli C. **et al**. Gene expression modules in primary breast cancers as risk factors for organotropic patterns of first metastatic spread: a case control study. Breast Cancer Res. 2017;19(1):113.
11. Li JN, Zhong R, Zhou XH. **Prediction of Bone Metastasis in Breast Cancer Based on Minimal Driver Gene Set in Gene Dependency Network**. **Genes** (**Basel**) 2019; **10**(**6**). **pii**: **E466**.
12. Zhang Q, Jin X, Shi W. **et al**. A long non-coding RNA LINC00461-dependent mechanism underlying breast cancer invasion and migration via the miR-144-3p/KPNA2 axis. Cancer Cell Int. 2020;20:137.
13. Ni C, Fang QQ, Chen WZ. **et al**. Breast cancer-derived exosomes transmit lncRNA SNHG16 to induce CD73 + γδ1 Treg cells. Signal Transduct Target Ther. 2020;5(1):41.
14. Wang B, Zhang Y, Zhang H. **et al**. Long intergenic non-protein coding RNA 324 prevents breast cancer progression by modulating miR-10b-5p. Aging. 2020;12(8):6680–99.
15. Wen X, Gao L, Hu Y. LAceModule. Identification of Competing Endogenous RNA Modules by Integrating Dynamic Correlation. Front Genet. 2020;11:235.
16. Hua K, Deng X, Hu J. **et al**. Long noncoding RNA HOST2, working as a competitive endogenous RNA, promotes STAT3-mediated cell proliferation and migration via decoying of let-7b in triple-negative breast cancer. J Exp Clin Cancer Res. 2020;39(1):58.
17. Yan P, Tang L, Liu L. **et al**. Identification of candidate RNA signatures in triple-negative breast cancer by the construction of a competing endogenous RNA network with integrative analyses of Gene Expression Omnibus and The Cancer Genome Atlas data. Oncol Lett. 2020;19(3):1915–27.
18. Lou W, Ding B, Fu P. Pseudogene-Derived lncRNAs and Their miRNA Sponging Mechanism in Human Cancer. Front Cell Dev Biol. 2020;8:85.
19. Li X, Zhang H, Wu X. Long noncoding RNA DLX6-AS1 accelerates the glioma carcinogenesis by competing endogenous sponging miR-197-5p to relieve E2F1. Gene. 2019;686:1–7.
20. Tian W, Jiang C, Huang Z. **et al**. Comprehensive analysis of dysregulated lncRNAs, miRNAs and mRNAs with associated ceRNA network in esophageal squamous cell carcinoma. Gene. 2019;696:206–18.
21. Zhang JJ, Xu WR, Chen B. **et al**. The up-regulated lncRNA DLX6-AS1 in colorectal cancer promotes cell proliferation, invasion and migration via modulating PI3K/AKT/mTOR pathway. Eur Rev Med

Pharmacol Sci. 2019;23(19):8321–31.

22. Wu DM, Zheng ZH, Zhang YB. **et al**. Down-regulated lncRNA DLX6-AS1 inhibits tumorigenesis through STAT3 signaling pathway by suppressing CADM1 promoter methylation in liver cancer stem cells. J Exp Clin Cancer Res. 2019;38(1):237.
23. Wang M, Li Y, Yang Y. **et al**. Long non-coding RNA DLX6-AS1 is associated with malignant progression and promotes proliferation and invasion in esophageal squamous cell carcinoma. Mol Med Rep. 2019;19(3):1942–50.
24. Alcaraz-Sanabria A, Baliu-Piqué M, Saiz-Ladera C. **et al**. Genomic Signatures of Immune Activation Predict Outcome in Advanced Stages of Ovarian Cancer and Basal-Like Breast Tumors. Front Oncol. 2020;9:1486.
25. Liu Z, Mi M, Li X. **et al**. lncRNA OSTN-AS1 May Represent a Novel Immune-Related Prognostic Marker for Triple-Negative Breast Cancer Based on Integrated Analysis of a ceRNA Network. Front Genet. 2019;10:850.
26. Reddy SM, Reuben A, Barua S. **et al**. Poor Response to Neoadjuvant Chemotherapy Correlates with Mast Cell Infiltration in Inflammatory Breast Cancer. Cancer Immunol Res. 2019;7(6):1025–35.
27. Zhao P, Guan H, Dai Z. **et al**. Long noncoding RNA DLX6-AS1 promotes breast cancer progression via miR-505-3p/RUNX2 axis. Eur J Pharmacol. 2019;865:172778.
28. Xiong H, Shen J, Chen Z. **et al**. H19/let-7/Lin28 ceRNA network mediates autophagy inhibiting epithelial-mesenchymal transition in breast cancer. Int J Oncol. 2020;56(3):794–806.
29. Yang SJ, Wang DD, Zhou SY. **et al**. Identification of circRNA-miRNA networks for exploring an underlying prognosis strategy for breast cancer. Epigenomics. 2020;12(2):101–25.
30. Yao Y, Zhang T, Qi L. **et al**. Integrated analysis of co-expression and ceRNA network identifies five lncRNAs as prognostic markers for breast cancer. J Cell Mol Med. 2019;23(12):8410–9.
31. Shimomura T, Kawakami M, Tatsumi K. **et al**. The Role of the Wnt Signaling Pathway in Upper Jaw Development of Chick Embryo. Acta Histochem Cytochem. 2019;52(1):19–26.
32. Matthews BG, Roguljic H, Franceschetti T. **et al**. Gene-expression analysis of cementoblasts and osteoblasts. J Periodontal Res. 2016;51(3):304 – 12.
33. Woo J, Miletich I, Kim BM. **et al**. Barx1-mediated inhibition of Wnt signaling in the mouse thoracic foregut controls tracheo-esophageal septation and epithelial differentiation. PLoS One. 2011;6(7):e22493.
34. Zhang RM, Tang T, Yu HM. **et al**. LncRNA DLX6-AS1/miR-129-5p/DLK1 axis aggravates stemness of osteosarcoma through Wnt signaling. Biochem Biophys Res Commun. 2018;507(1–4):260–6.
35. Guo J, Chen Z, Jiang H. **et al**. The lncRNA DLX6-AS1 promoted cell proliferation, invasion, migration and epithelial-to-mesenchymal transition in bladder cancer via modulating Wnt/β-catenin signaling pathway. Cancer Cell Int. 2019;19:312.
36. Zhang RM, Tang T, Yu HM. **et al**. LncRNA DLX6-AS1/miR-129-5p/DLK1 axis aggravates stemness of osteosarcoma through Wnt signaling. Biochem Biophys Res Commun. 2018;507(1–4):260–6.

37. Kim BM, Buchner G, Miletich I. **et al**. The stomach mesenchymal transcription factor Barx1 specifies gastric epithelial identity through inhibition of transient Wnt signaling. Dev Cell. 2005;8(4):611 – 22.
38. Yang J, Ye Z, Mei D. **et al**. Long noncoding RNA DLX6-AS1 promotes tumorigenesis by modulating miR-497-5p/FZD4/FZD6/Wnt/β-catenin pathway in pancreatic cancer. Cancer Manag Res. 2019;11:4209–21.
39. Wang P, Mokhtari R, Pedrosa E. **et al**. CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in cerebral organoids derived from iPS cells. Mol Autism. 2017;8:11.
40. Xue J, Zhu W, Song J. **et al**. Activation of PPARα by clofibrate sensitizes pancreatic cancer cells to radiation through the Wnt/β-catenin pathway. Oncogene. 2018;37(7):953–62.
41. Liu YT, Shang D, Akatsuka S. **et al**. Chronic oxidative stress causes amplification and overexpression of ptprz1 protein tyrosine phosphatase to activate beta-catenin pathway. Am J Pathol. 2007;171(6):1978-88.
42. Xiao Y, Ma D, Zhao S. **et al**. Multi-Omics Profiling Reveals Distinct Microenvironment Characterization and Suggests Immune Escape Mechanisms of Triple-Negative Breast Cancer. Clin Cancer Res. 2019;25(16):5002–14.

Figures

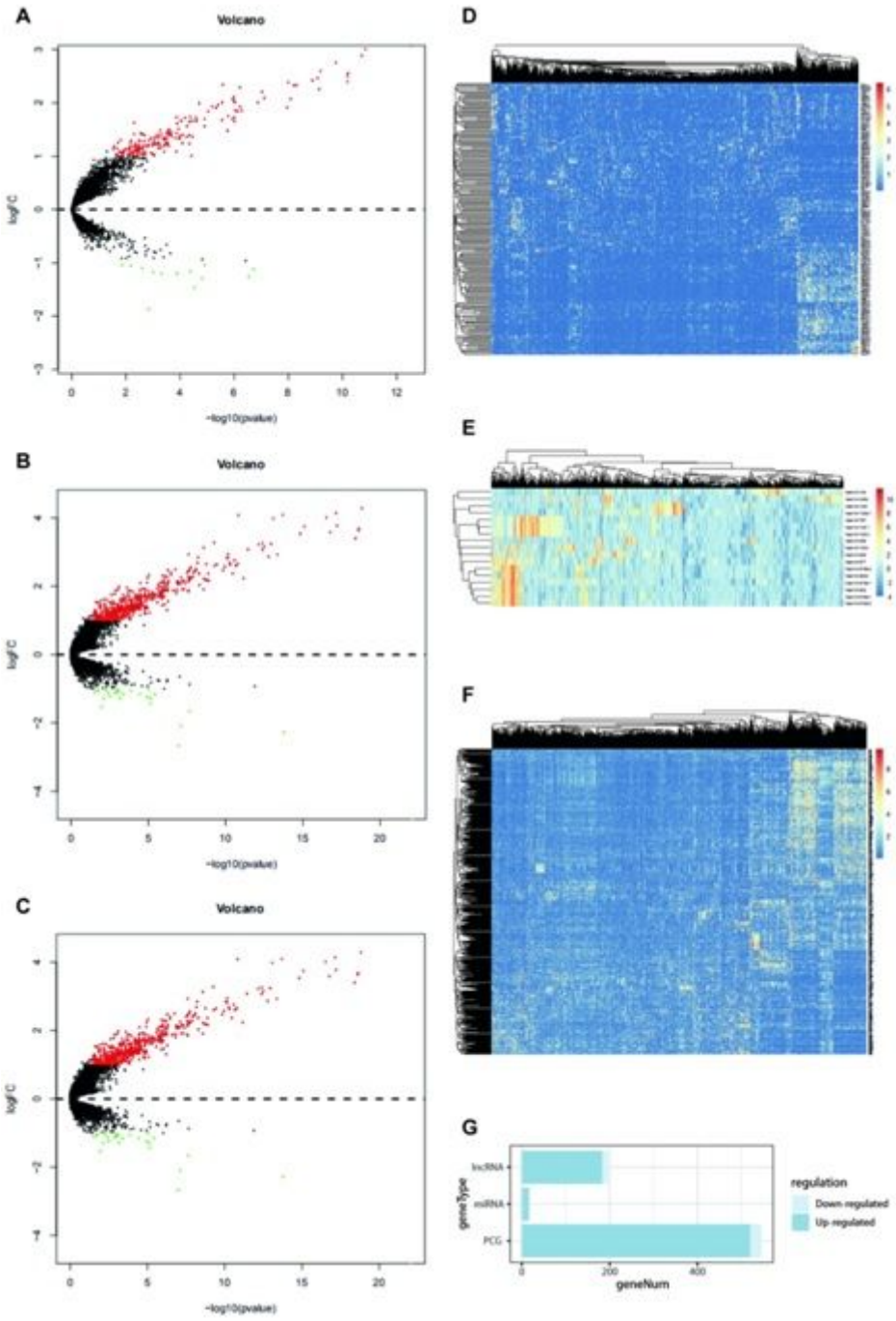

Figure 1

(A-C) Volcano map of differential expression of lncRNA, miRNA and mRNA. Red dot indicates genes significantly up-regulated in bone metastasis samples, green indicates genes significantly down-regulated, and black dot indicates genes without significant difference.(D-F) Thermogram of differential gene expression of lncRNA, miRNA and mRNA. (G) Statistical map of differential genes.

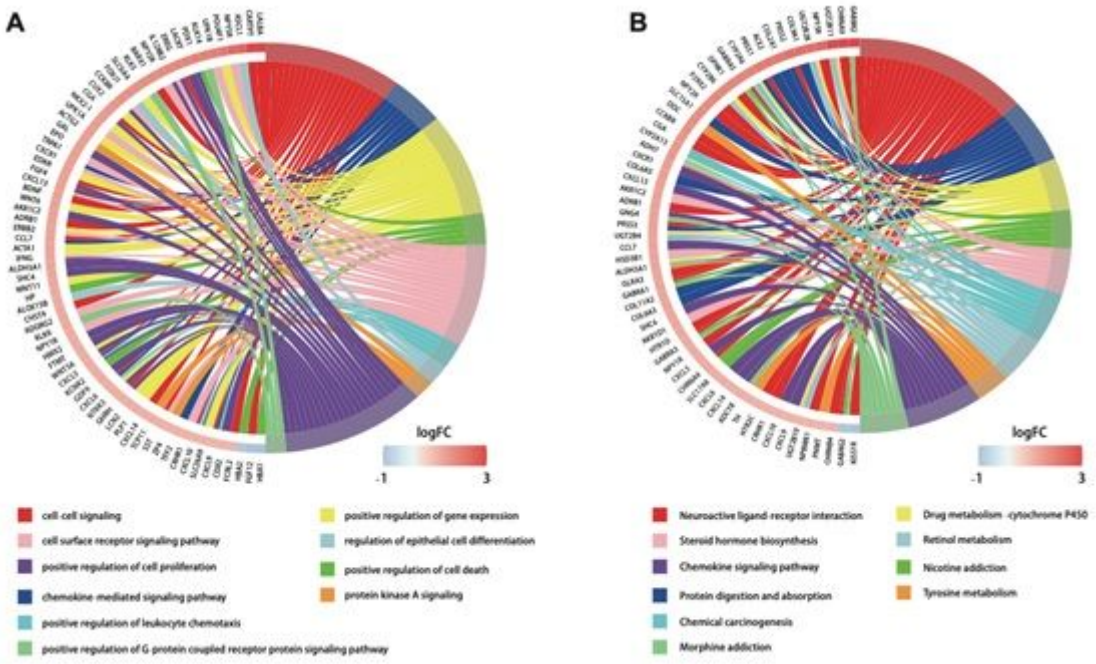

Figure 2

(A)GO function enrichment circle (B) Pathway enrichment cycle.

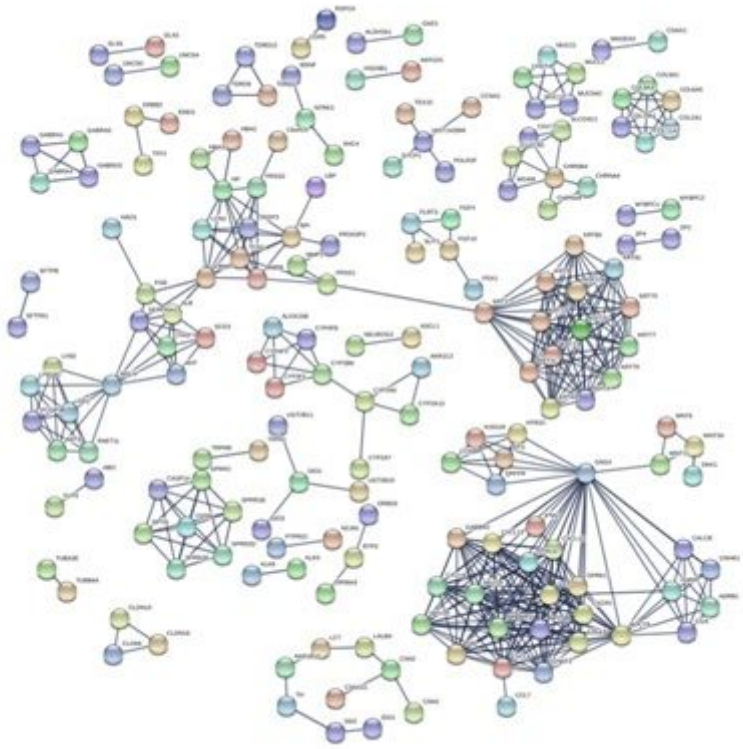

Figure 3

Protein interaction network of different genes.

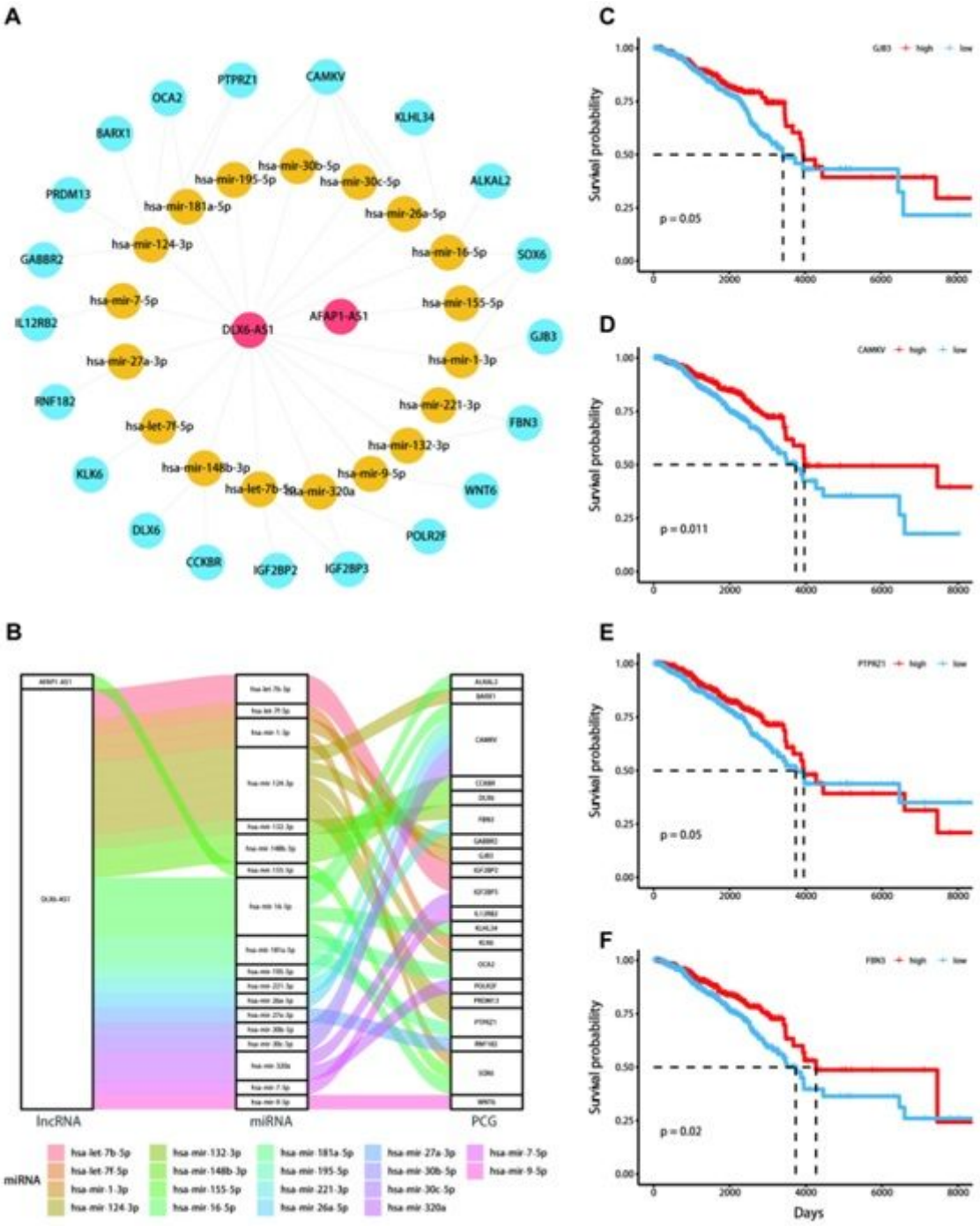

Figure 4

Construction of ceRNA network and prognosis analysis. (A) ceRNA network: The pink node represents lncRNA, the Yellow node represents miRNA, and the blue node represents PCG (B)Sankey plot of ceRNA network (C-F) Significant survival curve in ceRNA network

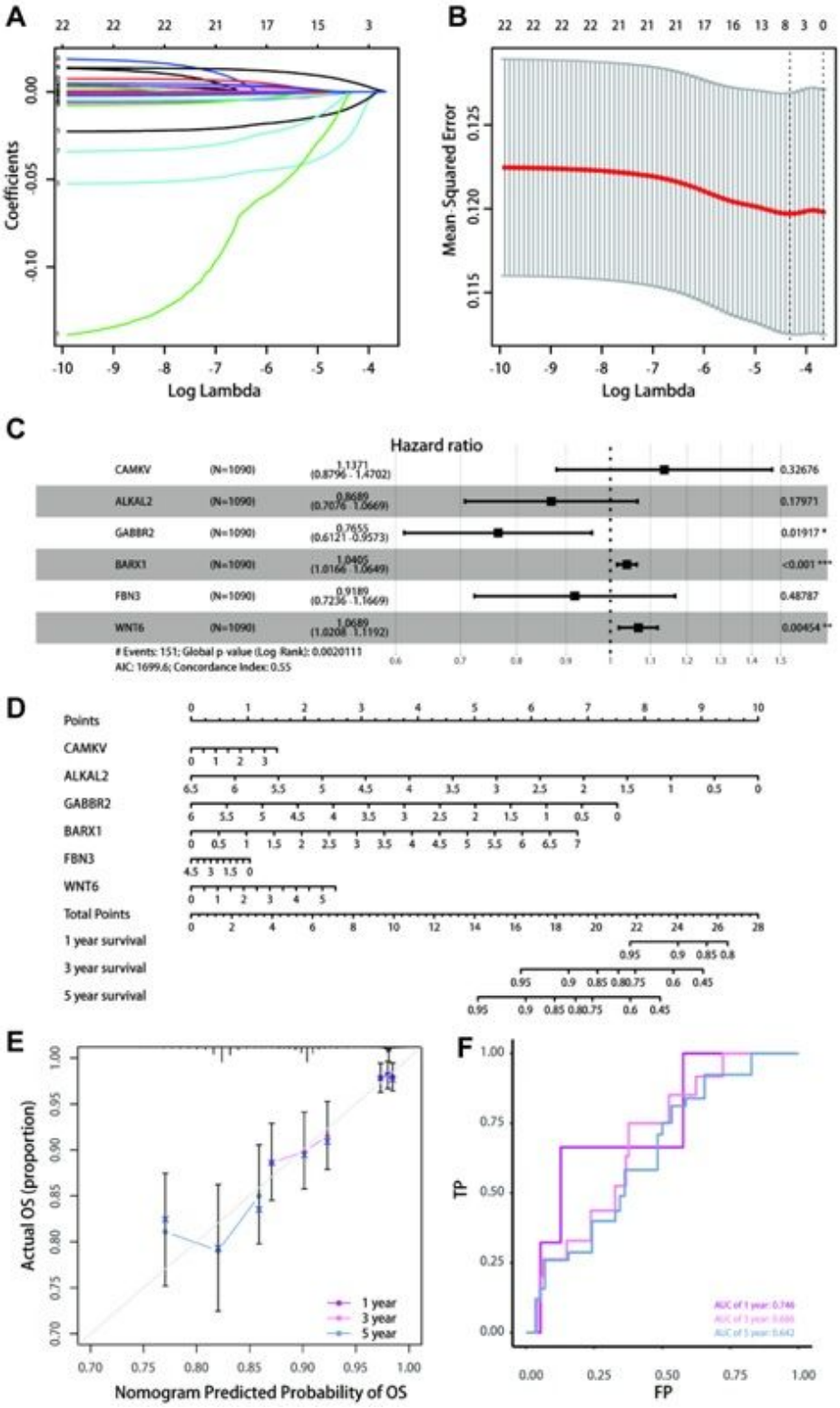

Figure 5

Lasso regression analysis and nomogram construction. (A,B) Selection of important coefficient lambda in lasso regression

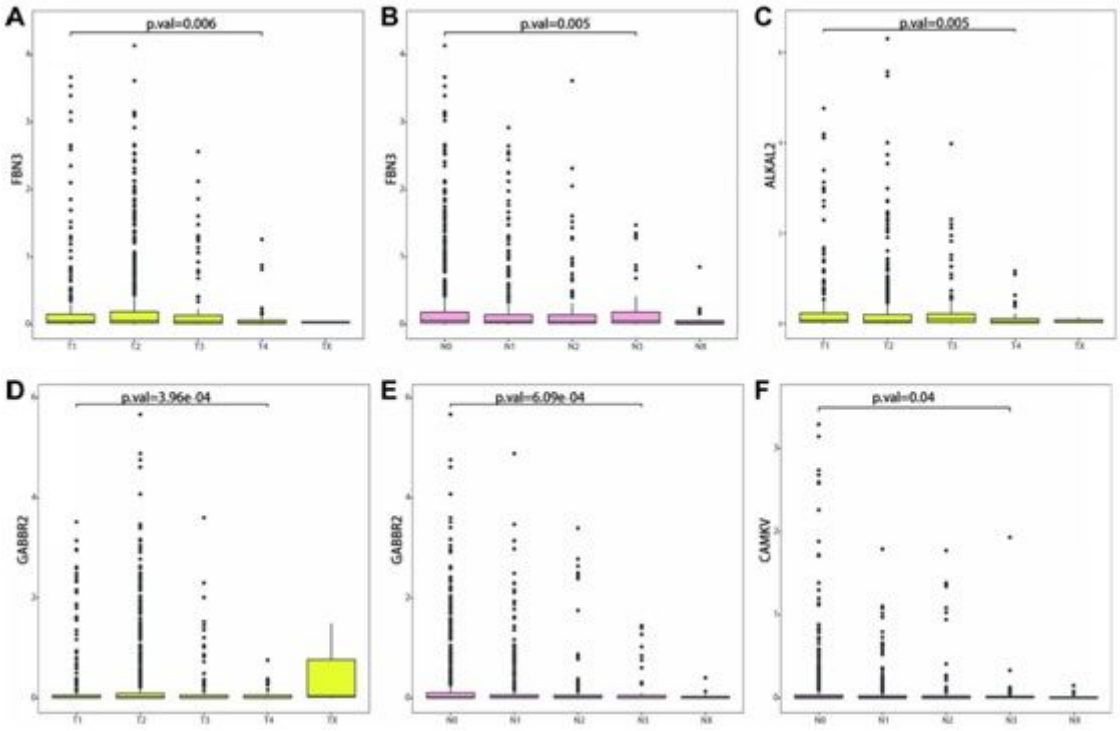

Figure 6

Relationship between gene expression level and TNM stage.

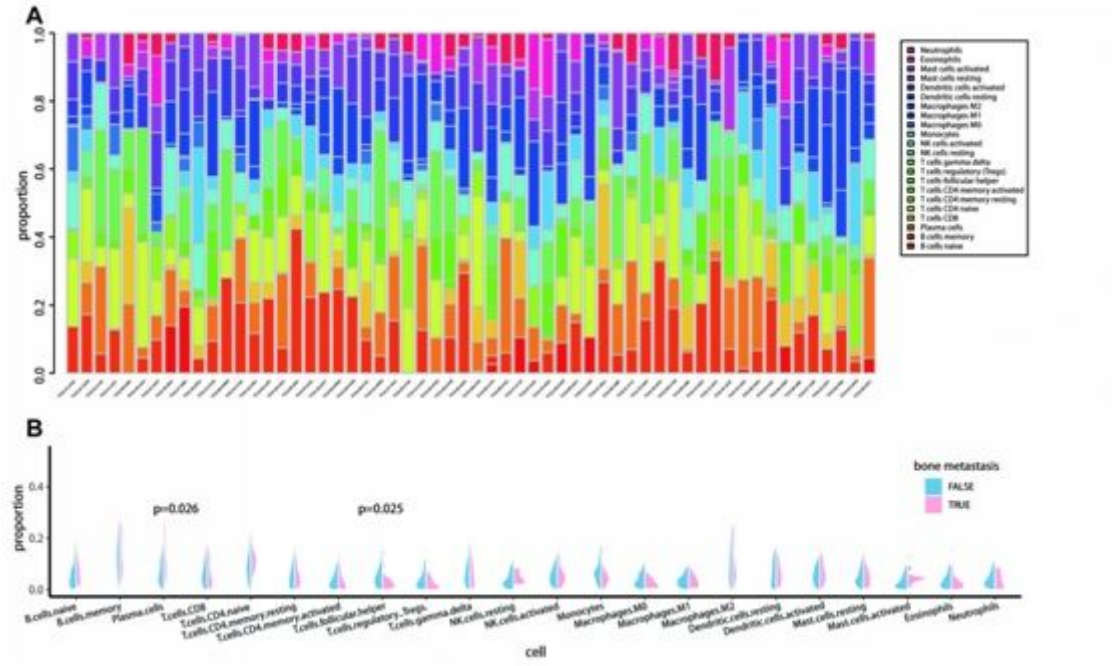

Figure 7

Component analysis of immune cells. (A)Proportion of lymphocytes in 58 bone metastasis samples (B)Proportion difference of 22 kinds of immune cells in bone metastasis and non bone metastasis

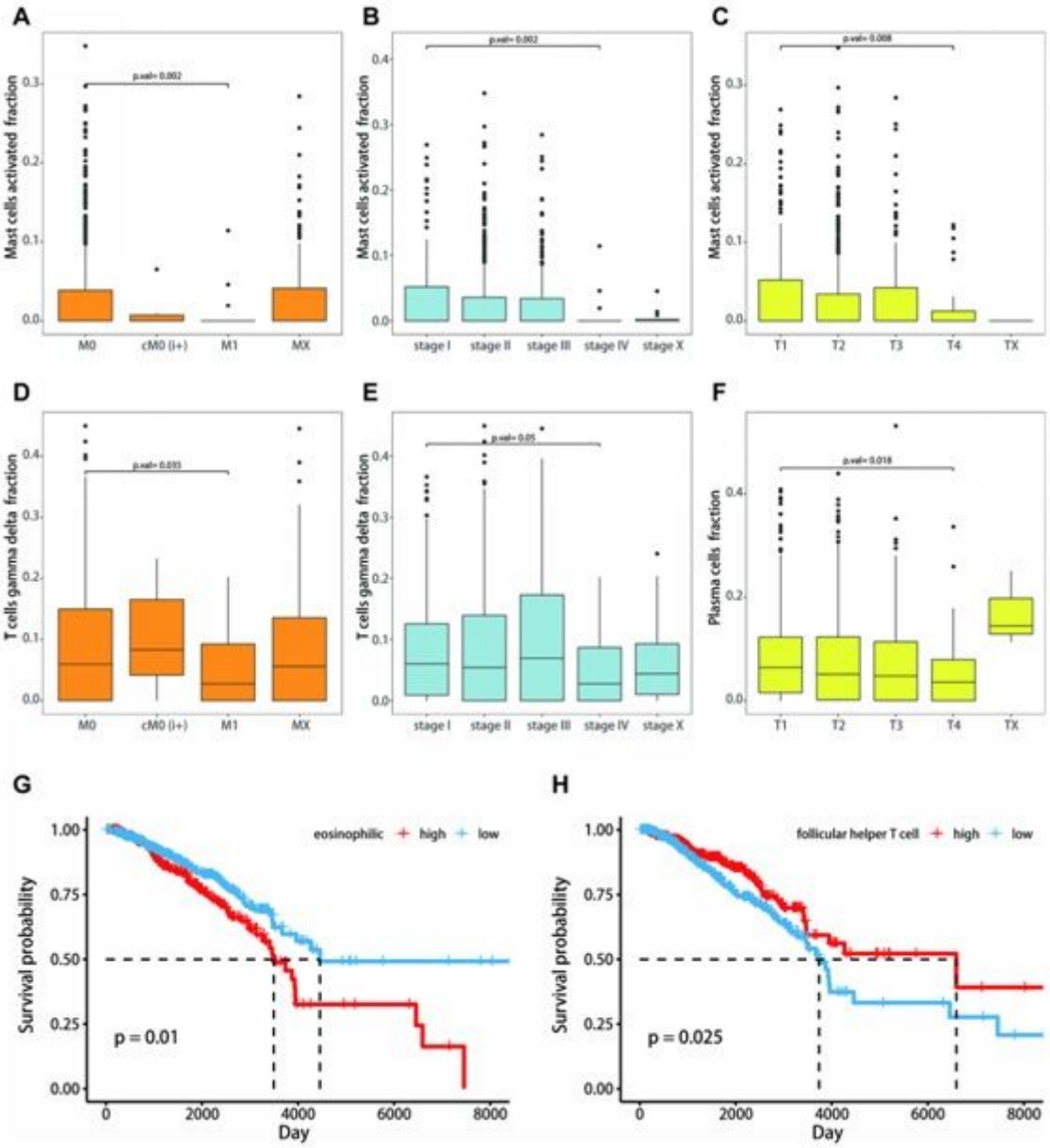

Figure 8

(A-E) Different cell contents in different stages. (F) There are differences between the survival states of the samples with high eosinophilic content and those with low eosinophilic content. (G) There are differences in survival status between the high-content samples and the low-content samples.

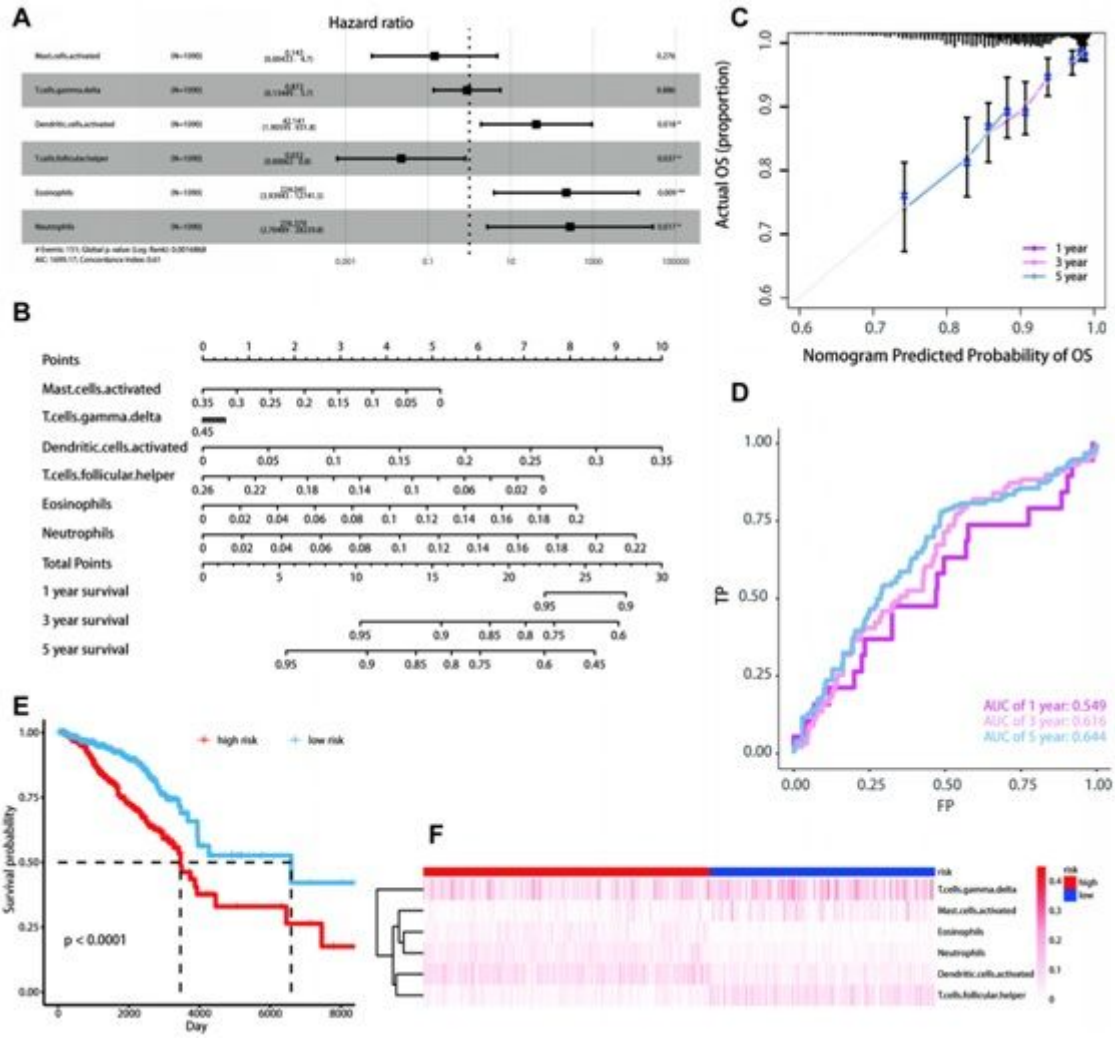

Figure 9

(A) Forest map of multiple Cox regression results (B) Nomogram based on multiple Cox regression (C) Calibration curve (D)ROC curve (E) Survival curve of high risk group and low risk group (F)Thermogram of six lymphocyte contents in high risk group and low risk group

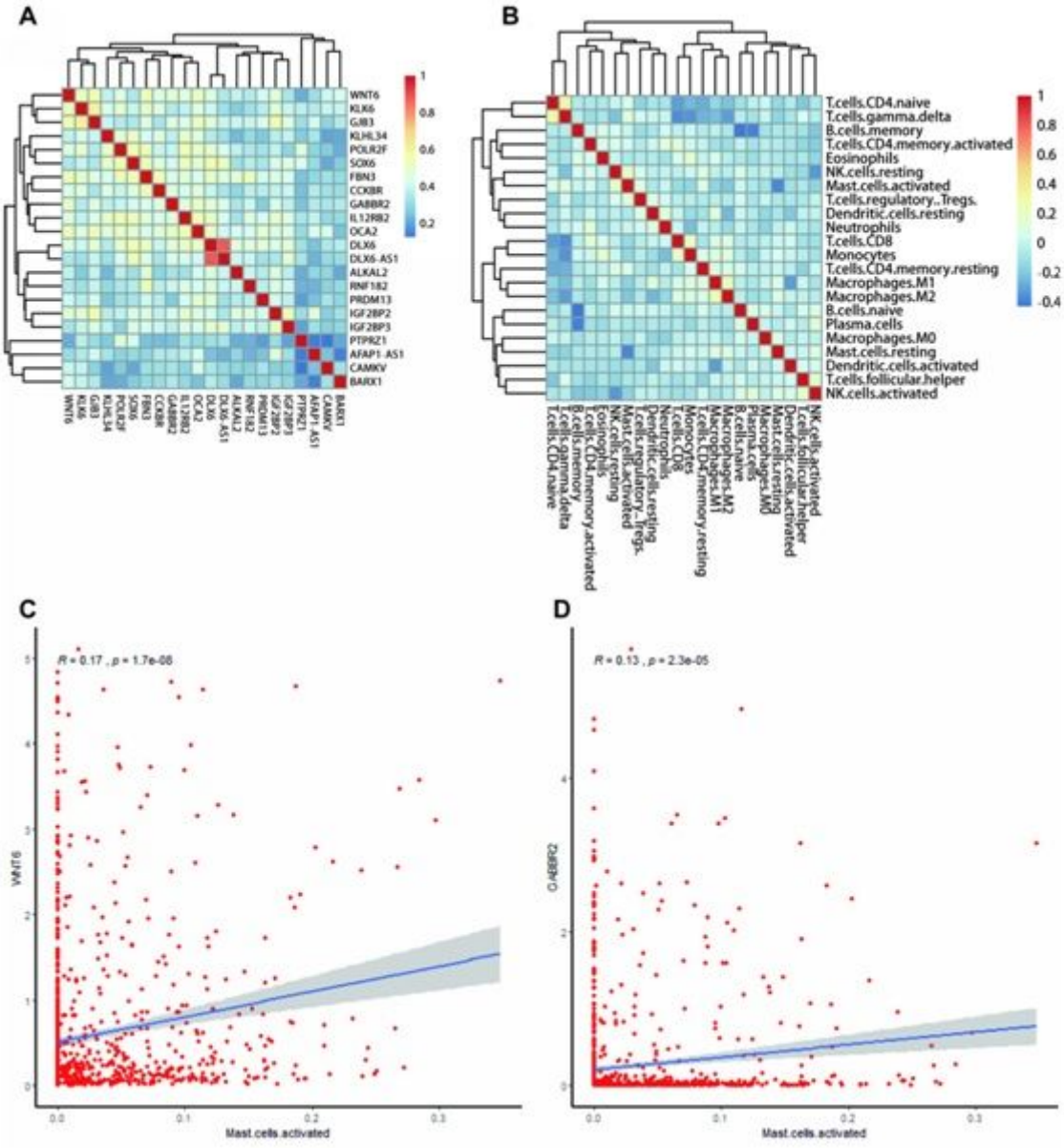

Figure 10

(A) Correlation thermogram of lncRNA and PCG in prognosis related ceRNA network (B) Correlation thermogram of 22 kinds of lymphocytes in breast cancer samples (C) Correlation between T.cells.regulatory.Tregs content and Wnt6 expression value (D) Correlation between T.cell.regulatory.tregs content and BARX1 expression value

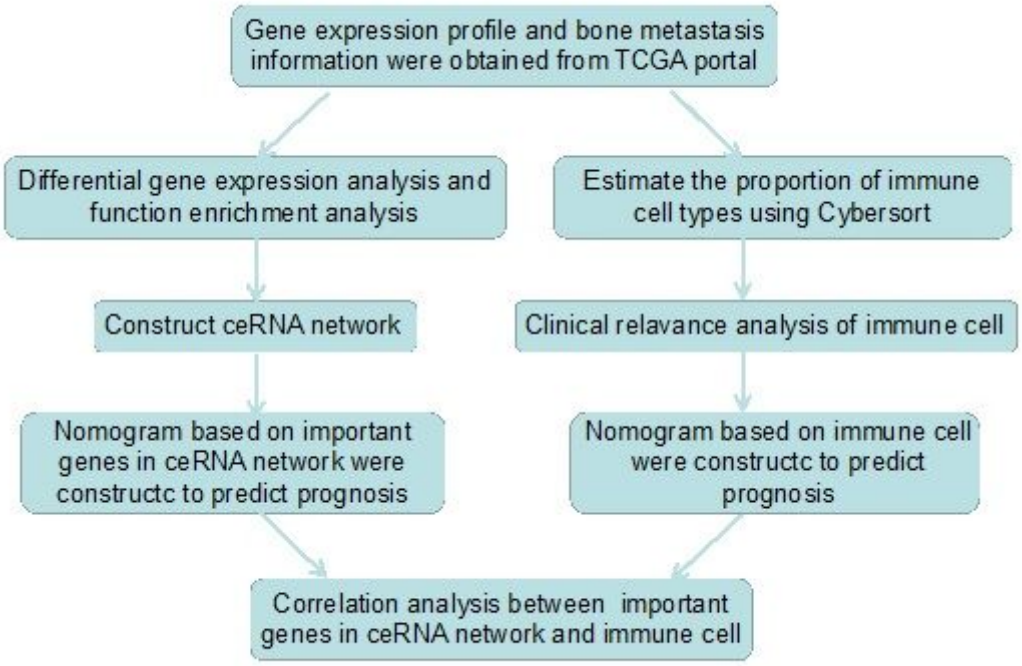

Figure 11

Article flowchart