

Research Article

# Identification of potential genes related to breast cancer brain metastasis in breast cancer patients

 Lijian Zhang<sup>1,2,3,\*</sup>, Luxuan Wang<sup>4,\*</sup>, Hua Yang<sup>5</sup>, Chunhui Li<sup>2</sup> and Chuan Fang<sup>1,2,3</sup>

<sup>1</sup>Postdoctoral Research Station of Neurosurgery, Affiliated Hospital of Hebei University, Hebei University, Baoding, Hebei, China; <sup>2</sup>Department of Neurosurgery, Affiliated Hospital of Hebei University, Hebei University, Baoding, Hebei, China; <sup>3</sup>Key Laboratory of Precise Diagnosis and Treatment of Glioma in Hebei Province, Affiliated Hospital of Hebei University, Hebei University, Baoding, Hebei, China; <sup>4</sup>Department of Neurology, Affiliated Hospital of Hebei University, Hebei University, Baoding, Hebei, China; <sup>5</sup>Department of Oncology, Affiliated Hospital of Hebei University, Hebei University, Baoding, Hebei, China

**Correspondence:** Chunhui Li (lichunhui0860312@sina.com) or Chuan Fang (chuanfang@hbu.edu.cn)



Brain metastases (BMs) usually develop in breast cancer (BC) patients. Thus, the molecular mechanisms of breast cancer brain metastasis (BCBM) are of great importance in designing therapeutic strategies to treat or prevent BCBM. The present study attempted to identify novel diagnostic and prognostic biomarkers of BCBM. Two datasets (GSE125989 and GSE100534) were obtained from the Gene Expression Omnibus (GEO) database to find differentially expressed genes (DEGs) in cases of BC with and without brain metastasis (BM). A total of 146 overlapping DEGs, including 103 up-regulated and 43 down-regulated genes, were identified. Functional enrichment analysis showed that these DEGs were mainly enriched for functions including extracellular matrix (ECM) organization and collagen catabolic fibril organization. Using protein–protein interaction (PPI) and principal component analysis (PCA) analysis, we identified ten key genes, including *LAMA4*, *COL1A1*, *COL5A2*, *COL3A1*, *COL4A1*, *COL5A1*, *COL5A3*, *COL6A3*, *COL6A2*, and *COL6A1*. Additionally, *COL5A1*, *COL4A1*, *COL1A1*, *COL6A1*, *COL6A2*, and *COL6A3* were significantly associated with the overall survival of BC patients. Furthermore, *COL6A3*, *COL5A1*, and *COL4A1* were potentially correlated with BCBM in human epidermal growth factor 2 (HER2) expression. Additionally, the miR-29 family might participate in the process of metastasis by modulating the cancer microenvironment. Based on datasets in the GEO database, several DEGs have been identified as playing potentially important roles in BCBM in BC patients.

## Background

Breast cancer (BC) is the most commonly diagnosed cancer in women [1]. The central nervous system (CNS) is one of the most common sites to which BC usually spreads, along with the bones, lungs, and liver. Metastasis of cancer to the brain has potentially devastating clinical consequences, resulting in an estimated survival time of less than 1 year, despite recent breakthroughs in neurologic therapies [2]. CNS metastases occur as a distant recurrence of BC. When cancer cells metastasize to the brain, patients' prognoses are poor, and they have few therapeutic options [3]. Surgical resection of single brain metastasis (BM) is considered a standard treatment for patients with metastases in accessible locations, good functional status, and absent/controlled extracranial disease. Overall surgical mortality is approximately 0.7–1.9%, while neurological morbidity is 3.9–6% [4]. The rate of postoperative complications, including pneumonia, urinary infections, and venous thrombosis, is approximately 13.9% [5,6]. Further, chemotherapy drugs targeting BC cells in the brain are not effective, because they cannot cross the blood–brain barrier [7,8]. In light of the poor outcomes and limited treatment options following a diagnosis of breast cancer brain metastasis (BCBM), a better understanding of the mechanisms may help improve clinical

\*These authors contributed equally to this work.

Received: 09 July 2021  
Revised: 26 August 2021  
Accepted: 08 September 2021

Accepted Manuscript online:  
20 September 2021  
Version of Record published:  
12 October 2021

decision-making. The development of novel biomarkers for BCBM would not only provide deeper insights into its pathology but would also help provide sensitive and efficient approaches for diagnosis and treatment.

Previous studies have documented that tumor BM is a complicated process, including steps such as the detachment of cells from the primary tumor, invasion of the extracellular matrix (ECM), travel through the bloodstream to arrest at a secondary site, and extravasation from the vasculature to establish metastasis in a new organ [9]. Numerous animal experiments and clinical trials appear to support Paget's 'seed and soil' hypothesis for non-random metastatic spread [10,11]. Although the initial steps for the metastasis of primary tumor cells may significantly overlap, the circulation patterns, extravasation barriers, and potential to survive in foreign tissue varies with the molecular subtype of the primary tumor [12]. Molecular cross-talk between primary BC and other cells in the surrounding microenvironment, as well as the role of primary BC cells in modulating the brain microenvironment during metastasis, are also key factors driving metastatic outgrowth [13,14]. Evidence has shown that, prior to reaching the brain, some metastatic primary BC cells already express proteins that are essential for the establishment of brain metastases, including serpins [15], matrix metalloproteases [16], and  $\alpha$ B-crystallin [17]. BCBM cells overexpress bone morphogenic proteins (BMPs) to promote the differentiation of neural progenitor cells (NPCs) into astrocytes and create a permissive microenvironment which allows cancer cells to colonize and proliferate [18]. Until now, extensive work in identifying driver genes of BCBM has provided some potential explanations for its metastasis to the brain. However, the underlying mechanisms involved in the invasion and metastatic outgrowth of BC cells into the CNS are largely unknown, and a novel option for the prevention or treatment of BM is unavailable [19].

With the development of genomics and bioinformatics, large amount of BC-related data are publicly available in several databases. Bioinformatics is a useful tool for extracting valuable information from existing data at a relatively low cost, and can help identify novel diagnostic biomarkers of BCBM [20,21]. In this study, by screening differentially expressed genes (DEGs) between BC patients with and without BM in two individual Gene Expression Omnibus (GEO) datasets, we aimed to explore potentially important genes associated with BCBM, and to investigate more information about the microenvironmental influence on BCBM development. The present study may help contribute to preventing BCBM at early stages in the future.

## Materials and methods

### Microarray data

Datasets were obtained from the National Center of Biotechnology Information (NCBI) GEO (<https://www.ncbi.nlm.nih.gov/geo/>) database. Two datasets (GSE125989 [22] and GSE100534 [23]) were used in the present study. The samples in the GSE125989 and GSE100534 datasets were detected using the Affymetrix Human Genome U133A 2.0 Array and the Affymetrix Human Gene 1.0 ST Array, respectively.

### Data processing

The GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>) based on R language limma package was used to analyze the gene expression data and to identify DEGs [24]. Those genes with  $|\log_{2}FC| > 1$  and  $P$ -values  $< 0.05$  were selected as DEGs (BC without BM vs. BCBM). There are 16 BC samples and 16 BCBM samples in the GSE125989 dataset, and there are 16 BC samples and 3 BCBM samples in the GSE100534 dataset (after removing the primary meningioma samples). To visualize the overlapping DEGs in both GSE125989 and GSE100534, a Venn diagram was generated using a freely accessible online tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

### Bioinformatics analysis of DEGs

In the Gene Ontology (GO) database, gene functions are divided into three parts, including cellular components (CCs), biological processes (BPs), and molecular functions (MFs). The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database integrating genome, chemistry, and system function information, which aims to reveal the genetic and chemical blueprint of life. It is a system function knowledge base that is based on computable experiments. We used the The Database for Annotation, Visualization and Integrated Discovery (DAVID) database for GO and KEGG enrichment analyses (<http://david.abcc.ncifcrf.gov/>). Functional protein-protein interaction (PPI) analysis is important for interpreting the potential molecular mechanisms of key cellular activities in pathogenicity. The Search Tool for the Retrieval of Interacting Genes (STRING) database (<http://string-db.org>) was used to construct PPI networks of high and low expression genes. An interaction score of 0.4 was treated as the cut-off criteria, and Cytoscape (<http://www.cytoscape.org/>) was used to visualize the PPI network. Hub genes with connection degrees  $> 9$  were selected. Then, the Molecular Complex Detection (MCODE) in the Cytoscape software was used to obtain the modules within the PPI network with MCODE scores  $> 2$  and number of nodes  $> 4$ . Functional enrichment analysis of the

genes in each module was also completed using Metascape [25]. Finally, a principal component analysis (PCA) model was visualized for clustering trends based on the variation in the phase of injury using the 'principal' function in the R software [26].

## Transcription factors and expression correlation analyses

To elucidate the selected DEGs expression regulation on a systematic level, a transcription factor (TF)–miRNA coregulatory network was generated using NetworkAnalyst based on the TF–miRNA coregulatory interaction database. The expression correlation analysis was conducted using Gene Expression Profiling Interactive Analysis (GEPIA), which was based on available The Cancer Genome Atlas (TCGA) datasets. Pearson correlation coefficients and *P*-values were calculated (<http://gepia.cancer-pku.cn/index.html>) [27].

## Survival analysis

To evaluate the effects of hub genes on the overall survival of BC patients, the publicly available Kaplan–Meier Plotter (<http://kmplot.com/analysis/>) was used. According to the median expression levels, the patients were classified into high and low expression groups [28]. Differences between the high and low expression groups were assessed using log-rank tests. *P* < 0.05 was considered to indicate statistically significant difference.

## Construction of BCBM prediction model

Here, we classified all samples into two sets, namely the samples with BM and samples without BM, adopted the expression of genes as a continuous predictor variable by glm package in R software [29]. We set the expression level of the genes as an independent variable, and whether it developed BM as the dependent variable to construct the logistic regression model. Firstly, the expression levels of COL4A1, COL5A1, and COL6A3 were extracted from GSE125989 and GSE100534 datasets. In each dataset, z-score normalization ( $\sim N(0,1)$  normal distribution) was conducted, and then we merged the two datasets. Subsequently, the z-scores of these genes were used for building a logistic regression prediction model to predict whether BM would develop in BC patients. And the accuracy of the model was determined by receiver operating characteristic (ROC) curve.

## Results

### Identification of DEGs related to BCBM in BC patients

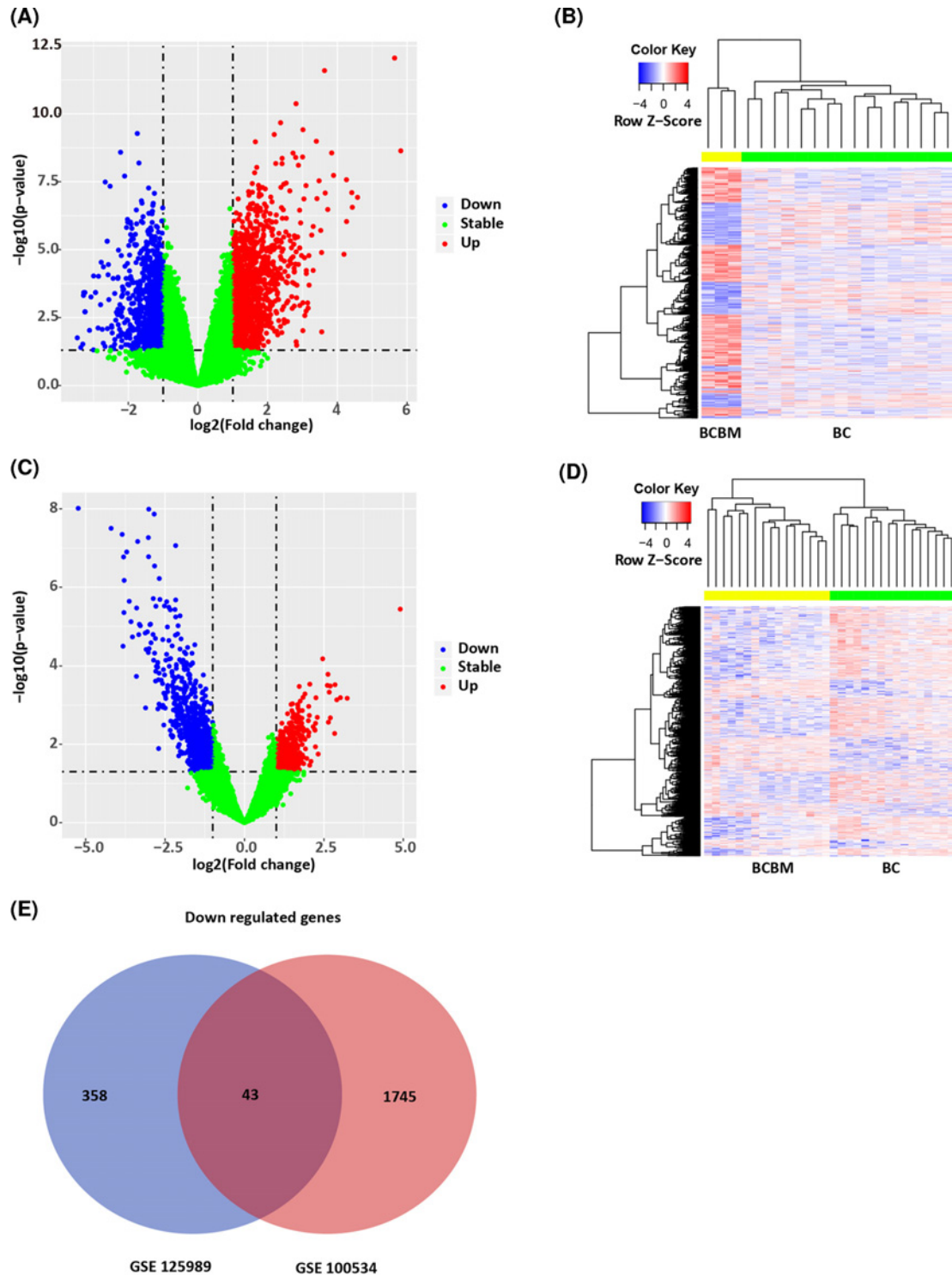
In order to screen the genes potentially associated with BCBM in BC patients, DEG analysis was performed between BC specimens (without BM) and BCBM specimens, which were conducted in GSE100534 and GSE125989, respectively. In the GSE100534 dataset, BCBM specimens were compared with BC specimens. A total of 2485 DEGs were obtained, including 971 up-regulated and 1593 down-regulated (Figure 1A,B). In the GSE125989 dataset, BCBM specimens were compared with BC specimens. A total of 1417 DEGs were obtained, including 984 up-regulated and 433 down-regulated (Figure 1C,D). A total of 146 overlapping genes (103 up-regulated and 43 down-regulated) were obtained between the two datasets (Figure 1E), which we further analyzed in our research.

### PPI network construction and Hub gene selection

We next constructed a PPI network based on the 146 overlapping DEGs using the STRING database, and visualized it using Cytoscape software. There were 64 nodes and 174 edges in the PPI network (Figure 2A). Among these, the top ten genes with the highest interaction degrees were then identified, including Collagen Type I  $\alpha$  1 Chains (COL1A1), Collagen Type III  $\alpha$  1 Chains (COL3A1), Collagen Type V  $\alpha$  2 Chains (COL5A2), Collagen Type V  $\alpha$  3 Chains (COL5A3), Collagen Type VI  $\alpha$  1 Chains (COL6A1), Collagen Type VI  $\alpha$  2 Chains (COL6A2), Collagen Type VI  $\alpha$  3 Chains (COL6A3), Collagen Type XIV  $\alpha$  1 Chains (COL14A1), Collagen Type XV,  $\alpha$  1 (COL15A1), and Laminin Subunit  $\alpha$  4 (LAMA4) (Figure 2B).

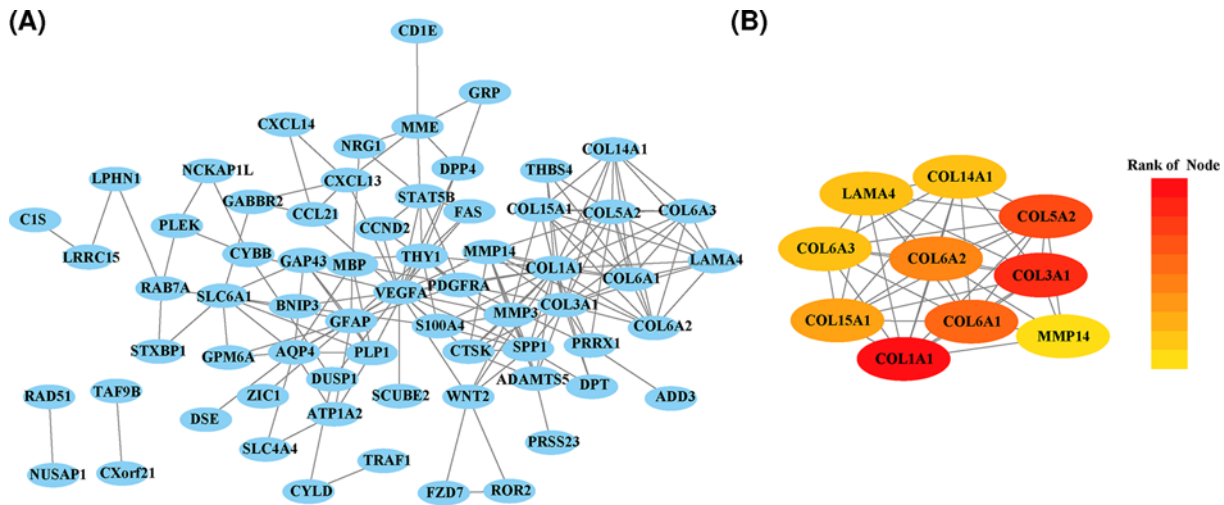
### PCA of the 146 DEGs

We performed PCA based on the expression profiles of DEGs to find DEGs that were highly related to the development of BCBM (Figure 3). Results showed that the first principal component (PC1) accounted for 60.8% of the total variance. The second principal component (PC2) accounted for an additional 11.4% of the total variance. Finally, changes in COL6A1, COL5A3, COL5A1, COL5A2, COL6A1, COL3A, COL6A1, COL14A1, and LAMA4 contributed the most to the separation between the BCBM and BC samples, which were also within top modules of the 146 DEGs. All of these genes are associated with ECM organization, the collagen catabolic process, and cellular adhesion.

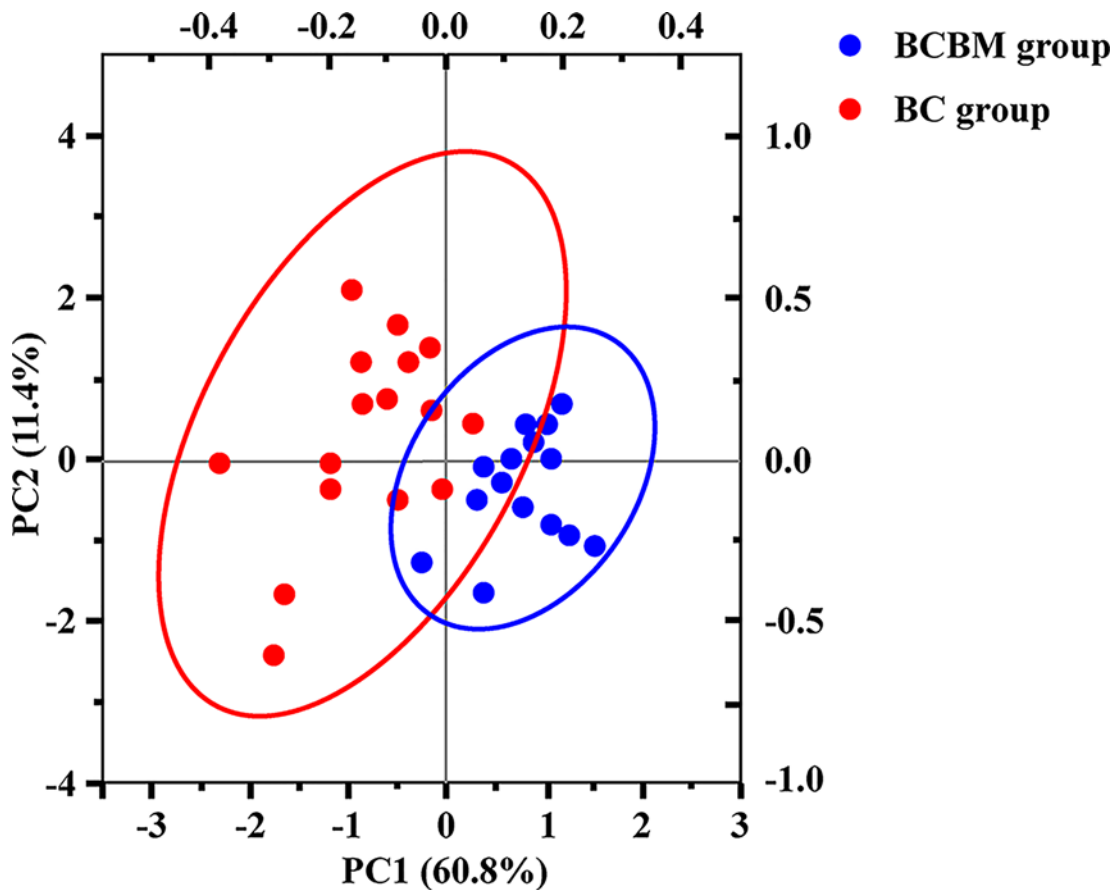


**Figure 1. Identification of DEGs related to BCM in BC patients**

(A) Volcano map of DEGs in GSE100534. X-axis:  $\log_2\text{FC}$ ; Y-axis:  $-\log_{10}$  (FDR). Blue: down-regulated genes; red: up-regulated genes; green: non-significant. (B) Heat map of DEGs in GSE100534. X-axis: sample; Y-axis: different genes. Red: high expression; blue: low expression. BCM and BC samples could be obviously clustered into two clusters. (C) Volcano map of DEGs in GSE125989. X-axis:  $\log_2\text{FC}$ ; Y-axis:  $-\log_{10}$  (FDR). Blue: down-regulated genes; red: up-regulated genes; green: non-significant. (D) Heat map of DEGs in GSE125989. X-axis: sample; Y-axis: different genes. Red: high expression; blue: low expression. BCM and BC samples could be obviously clustered into two clusters. (E) Venn diagram of overlapped DEGs between GSE100534 and GSE125989.



**Figure 2. PPI network and top module of 146 DEGs**  
 (A) PPI network of DEGs in light blue and top one module in orange. (B) Top ten hub genes.



**Figure 3. PCA of DEGs**  
 The variance in gene expression between BCBM group and BC group.

## KEGG pathway and GO term enrichment analyses

We next performed functional enrichment analyses on the DEGs to obtain more information about the genes. Detailed results of GO enrichment analysis are shown in Table 1. The most significantly enriched MF terms were ECM



**Table 1 GO functional enrichment analysis for the DEGs**

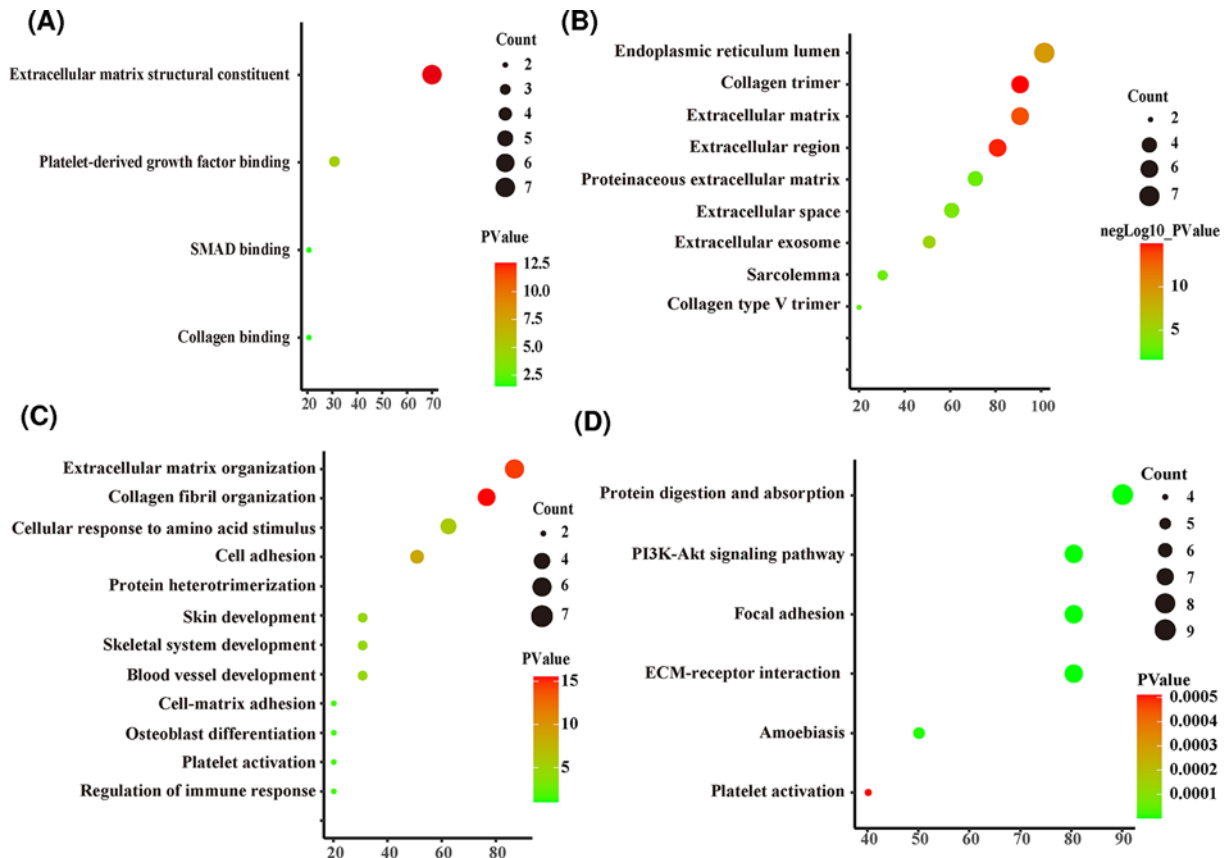
Term	Category	P-value	UniProt ID
<b>BP</b>			
Extracellular matrix organization	GO:0030198	2.66E-15	COL1A1, COL3A1, COL14A1, LAMA4, COL6A2, COL5A3, COL5A2, COL6A1, COL6A3
Collagen fibril organization	GO:0030199	3.10E-09	COL1A1, COL3A1, COL14A1, COL5A3, COL5A2
Cellular response to amino acid stimulus	GO:0071230	1.71E-06	COL1A1, COL3A1, COL5A2, COL6A1
Cell adhesion	GO:0007155	1.72E-06	COL1A1, COL15A1, LAMA4, COL6A2, COL6A1, COL6A3
Protein heterotrimerization	GO:0070208	2.32E-05	COL1A1, COL6A2, COL6A1
Skin development	GO:0043588	1.42E-04	COL3A1, COL5A3, COL5A2
Skeletal system development	GO:0001501	0.00229126	COL1A1, COL3A1, COL5A2
Blood vessel development	GO:0001568	0.020188218	COL1A1, LAMA4
Cell–matrix adhesion	GO:0007160	0.047226948	COL3A1, COL5A3
Osteoblast differentiation	GO:0001649	0.054392332	COL1A1, COL6A1
Platelet activation	GO:0030168	0.059988631	COL1A1, COL3A1
Regulation of immune response	GO:0050776	0.091476749	COL1A1, COL3A1
<b>MF</b>			
Extracellular matrix structural constituent	GO:0005201	2.59E-13	COL1A1, COL3A1, COL15A1, COL14A1, LAMA4, COL5A3, COL5A2
Platelet-derived growth factor binding	GO:0048407	1.39E-05	COL1A1, COL3A1, COL6A1
SMAD binding	GO:0046332	0.022698305	COL3A1, COL5A2
Collagen binding	GO:0005518	0.03154496	COL14A1, COL5A3
<b>CC</b>			
Endoplasmic reticulum lumen	GO:0005788	2.66E-15	COL1A1, COL3A1, COL15A1, COL6A2, COL5A3, COL5A2, COL6A1, COL6A3
Collagen trimer	GO:0005581	3.10E-09	COL1A1, COL3A1, COL14A1, LAMA4, COL6A2, COL5A3, COL5A2, COL6A1, COL6A3
Extracellular matrix	GO:0031012	1.71E-06	COL1A1, COL3A1, COL14A1, COL5A3, COL5A2
Extracellular region	GO:0005576	1.72E-06	COL1A1, COL3A1, COL5A2, COL6A1
Proteinaceous extracellular matrix	GO:0005578	2.32E-05	COL1A1, COL15A1, LAMA4, COL6A2, COL6A1, COL6A3
Extracellular space	GO:0005615	1.42E-04	COL1A1, COL6A2, COL6A1
Extracellular exosome	GO:0070062	0.00229126	COL3A1, COL5A3, COL5A2
Sarcolemma	GO:0042383	0.020188218	COL1A1, COL3A1, COL5A2
Collagen type V trimer	GO:0005588	0.047226948	COL1A1, LAMA4
Collagen type VI trimer	GO:0005589	2.98E-16	COL3A1, COL5A3

Category refers to the GO functional categories.

organization, collagen catabolic process, and cellular adhesion (Figure 4A). The most noteworthy enriched CC terms were extracellular region and endoplasmic reticulum lumen ECM (Figure 4B). The most significantly enriched BP terms were ECM structural constituent, platelet-derived growth factor binding, and SMAD binding (Figure 4C). Detailed results of KEGG enrichment analysis are also shown in Table 2. The top six KEGG pathways are displayed in Figure 4D, including protein digestion and absorption, the PI3K–Akt signaling pathway, the Focal adhesion pathway, and ECM–receptor interactions.

## Survival analysis

To investigate the impacts of those ten hub genes on the survival rates of BC patients, we used Kaplan–Meier curve analysis. The median expression level was regarded as the cut-off value for dividing the patients into two groups (high and low expression groups). Results showed that high expression of COL14A1 (HR = 0.56 [0.4–0.79],  $P=0.00081$ ), COL6A2 (HR = 0.74 [0.6–0.93],  $P=0.0082$ ), COL6A1 (HR = 0.59 [0.45–0.78],  $P=0.00012$ ), COL5A1 (HR = 0.76 [0.59–0.98],  $P=0.034$ ), and collagen type I  $\alpha$  1 chain (COL1A1) (HR = 0.8 [0.65–0.99],  $P=0.044$ ) were associated with poor overall survival rates for BC patients. LAMA4, COL3A1, and COL5A1 expression had no significant influence



**Figure 4.** GO enrichment and KEGG pathway analysis of DEGs in BC and BCBM group  
 (A) GO categories of MF. (B) GO categories of CC. (C) GO categories of BP. (D) KEGG pathway analysis of the DEGs.

**Table 2** Pathway enrichment analysis for the DEGs

Term	Category	P-value	UniProt ID
Protein digestion and absorption	hsa04974	4.62E-15	COL1A1, COL3A1, COL15A1, COL14A1, COL6A2, COL5A3, COL5A2, COL6A1, COL6A3
ECM-receptor interaction	hsa04512	1.43E-12	COL1A1, COL3A1, LAMA4, COL6A2, COL5A3, COL5A2, COL6A1, COL6A3
Focal adhesion	hsa04510	6.68E-10	COL1A1, COL3A1, LAMA4, COL6A2, COL5A3, COL5A2, COL6A1, COL6A3
PI3K-Akt signaling pathway	hsa04151	2.48E-08	COL1A1, COL3A1, LAMA4, COL6A2, COL5A3, COL5A2, COL6A1, COL6A3
Amoebiasis	hsa05146	6.33E-06	COL1A1, COL3A1, LAMA4, COL5A3, COL5A2
Platelet activation	hsa04611	5.10E-04	COL1A1, COL3A1, COL5A3, COL5A2

Category refers to the pathway functional categories.

on the overall survival rates of BC patients ( $P > 0.05$ ) (Figure 5). Moreover, correlation analysis results indicated that the expression of COL4A1, COL5A1, and COL6A3 was positively correlated with HER2 positive expression (Figure 6).

### TF regulatory network analysis of ten genes

We constructed a gene–TF regulatory network based on the ten identified genes—which included 56 interaction pairs between 10 genes and 45 TFs (Figure 7A). LAMA4 was found to be regulated by 11 TFs, COL1A1 was regulated by 10 TFs, COL5A2 was regulated by 8 TFs, COL3A1 was regulated by 5 TFs, COL4A1 was regulated by 4 TFs, COL5A1

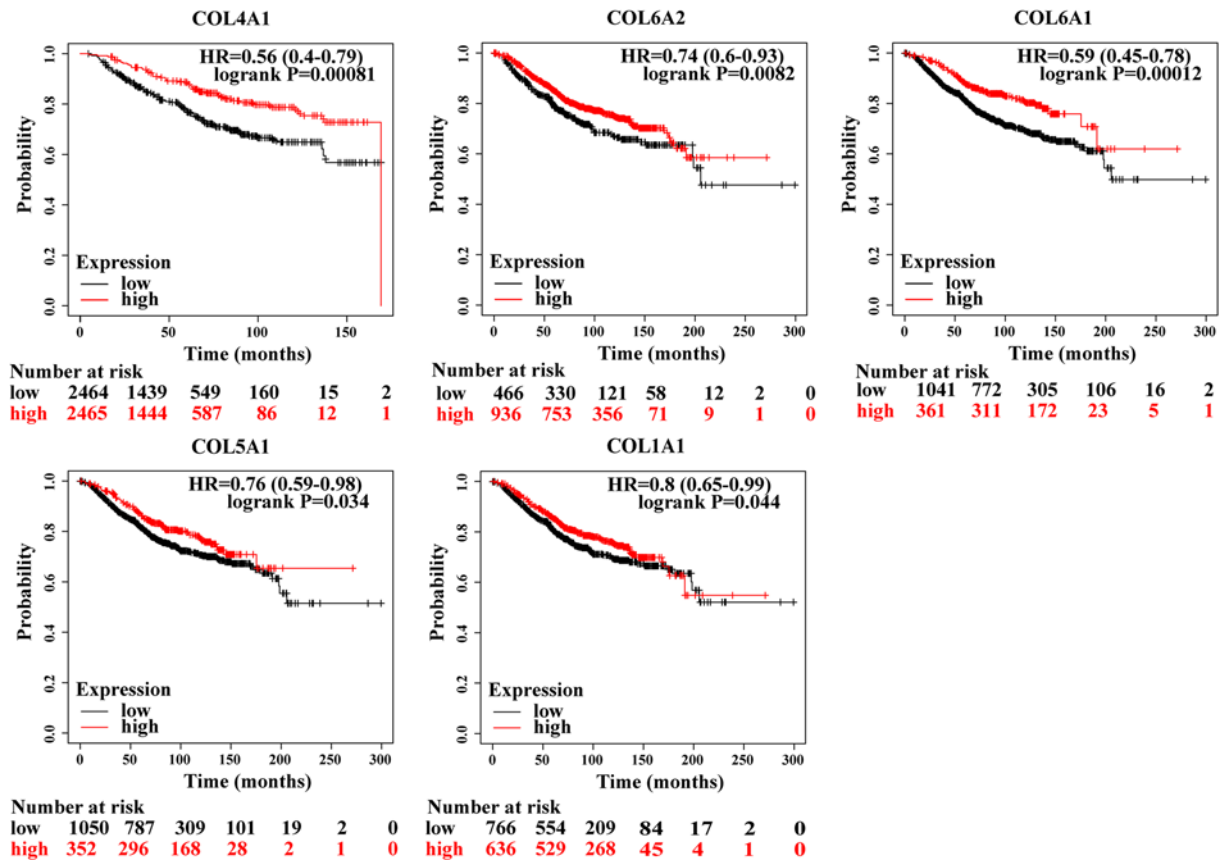


Figure 5. Prognostic value of ten genes

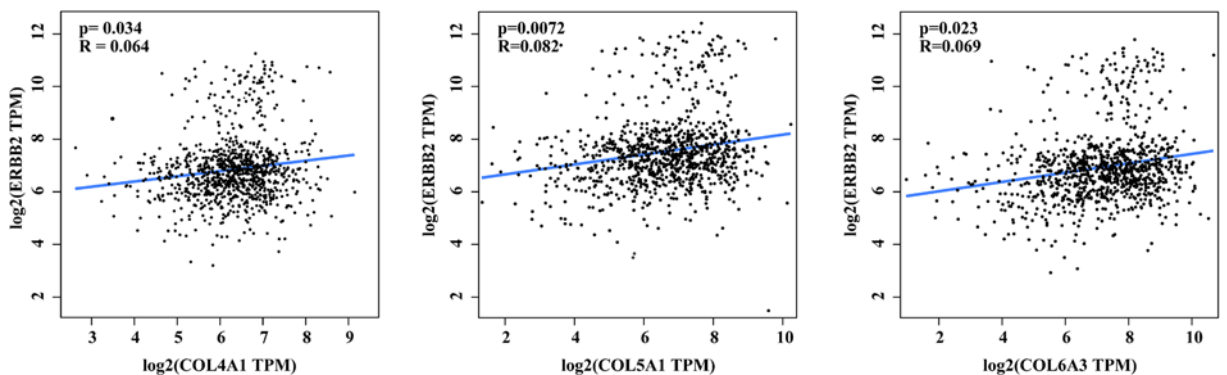


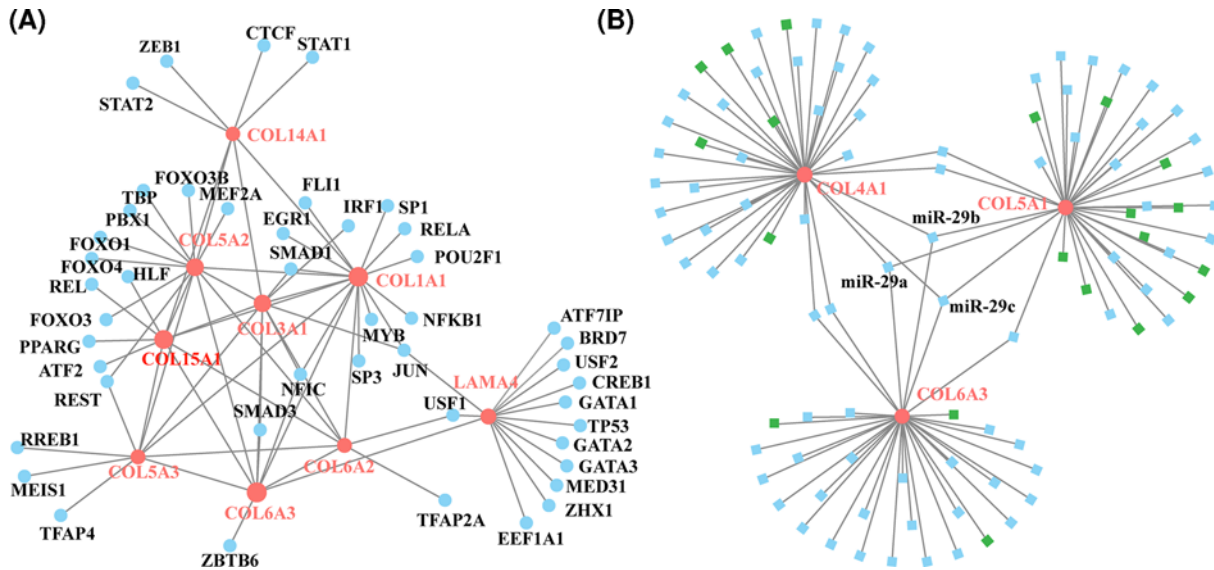
Figure 6. Correlation between HER2 and genes

was regulated by 4 TFs, COL5A3 was regulated by 4 TFs, COL6A3 was regulated by 3 TFs, and COL6A2 was regulated by 2 TFs. We further investigated the miRNAs that could regulate COL4A1, COL5A1, and COL6A3 (Table 3). Our findings suggested that the expression of these three genes was significantly correlated with each other, and that they were all also under the regulation of miRNA-29 family, including miR-29a, miR-29b, and miR-29c (Figure 7B).

### Construction and validation of BCBM prediction model

Using the three most important genes (*COL4A1*, *COL5A1*, and *COL6A3*), we constructed a logistic regression prediction model ( $Y = 0.2987 \times COL5A1 + 0.9790 \times COL4A1 - 2.6080 \times COL6A3 - 0.6142$ ), using the glm function in the R language. When  $Y \geq 0.5$ , the model predicts that BM will occur in BC patients. Furthermore, we found that there was no extreme point affecting the accuracy of the model (Figure 8). All points in the model had good linear





**Figure 7. The TF regulatory network**

(A) The TF regulatory network of the top ten hub genes. Pink circle means the hub gene and blue circle means the transcription. (B) The TF regulatory network of selected DEGs (*COL4A1*, *COL5A1*, and *COL6A3*). The pink circle means the selected DEGs and blue square means the miRNAs.

**Table 3 The potential TFs of ten hub genes**

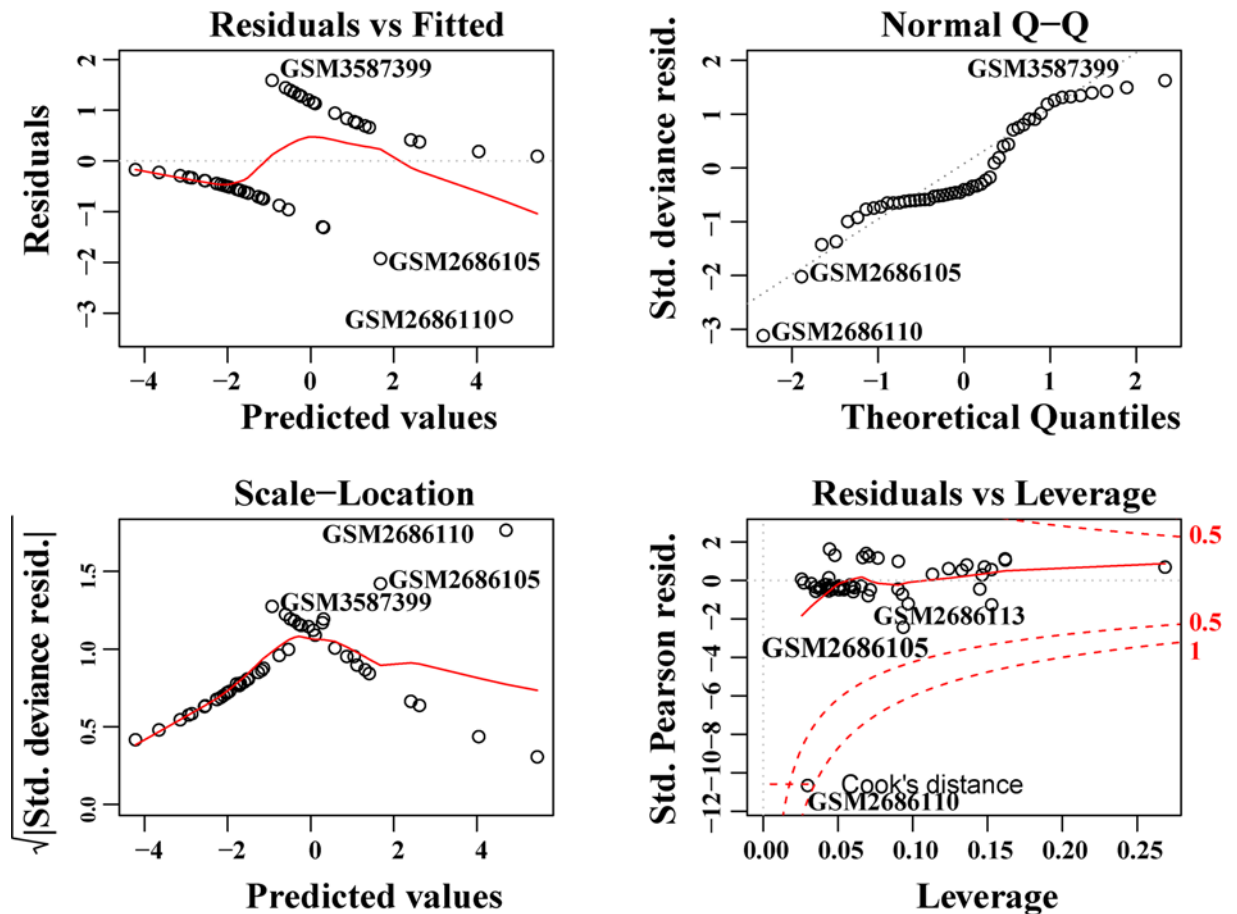
Genes	TFs	Counts
<i>LAMA4</i>	ATF71P, BRD7, USF2, CREB1, GATA1, TP53, GATA2, GATA3, MED31, ZHX1, EEF1A1	11
<i>COL1A1</i>	SP3, JUN, MYB, NFKB1, POU2F1, REL, SP1FLI1, EGR1, SMAD1, WFIC	10
<i>COL5A2</i>	MEF2A, FOXO3B, TBP, PBX1, FOXO1, FOXO4, FOXO3, REL	8
<i>COL3A1</i>	SMAD1, IRF1, NFIC, SMAD3, JUN	5
<i>COL4A1</i>	STAT2, ZEB1, CTCF, STAT1	4
<i>COL5A1</i>	HLF, REL, PPARG, ATF2	4
<i>COL5A3</i>	REST, RREB1, MEIS1, TFAP4	4
<i>COL6A3</i>	ZBTB6, SMAD3, NFIC	3
<i>COL6A2</i>	NFIC, TFAP2A	2
<i>COL6A1</i>	SMAD3	1

relationships (Figure 9). Further, the area under the curve ( $AUC = 0.91$ ) indicated that the logistic regression prediction model—which was based on the three genes—could predict whether BCM would occur in BC patients (Figure 10). Our results suggested that *COL4A1*, *COL5A1*, and *COL6A3* were potentially crucial genes for the development of BCM.

## Discussion

The development of BCM is a multistep process that is still poorly understood. Despite advances in targeted treatment, patients with BCM continue to exhibit an undesirable prognosis and impaired quality of life. Patients with a high risk of developing BCM could potentially benefit from early screening strategies, as earlier diagnoses would lead to more effective BCM treatments [30].

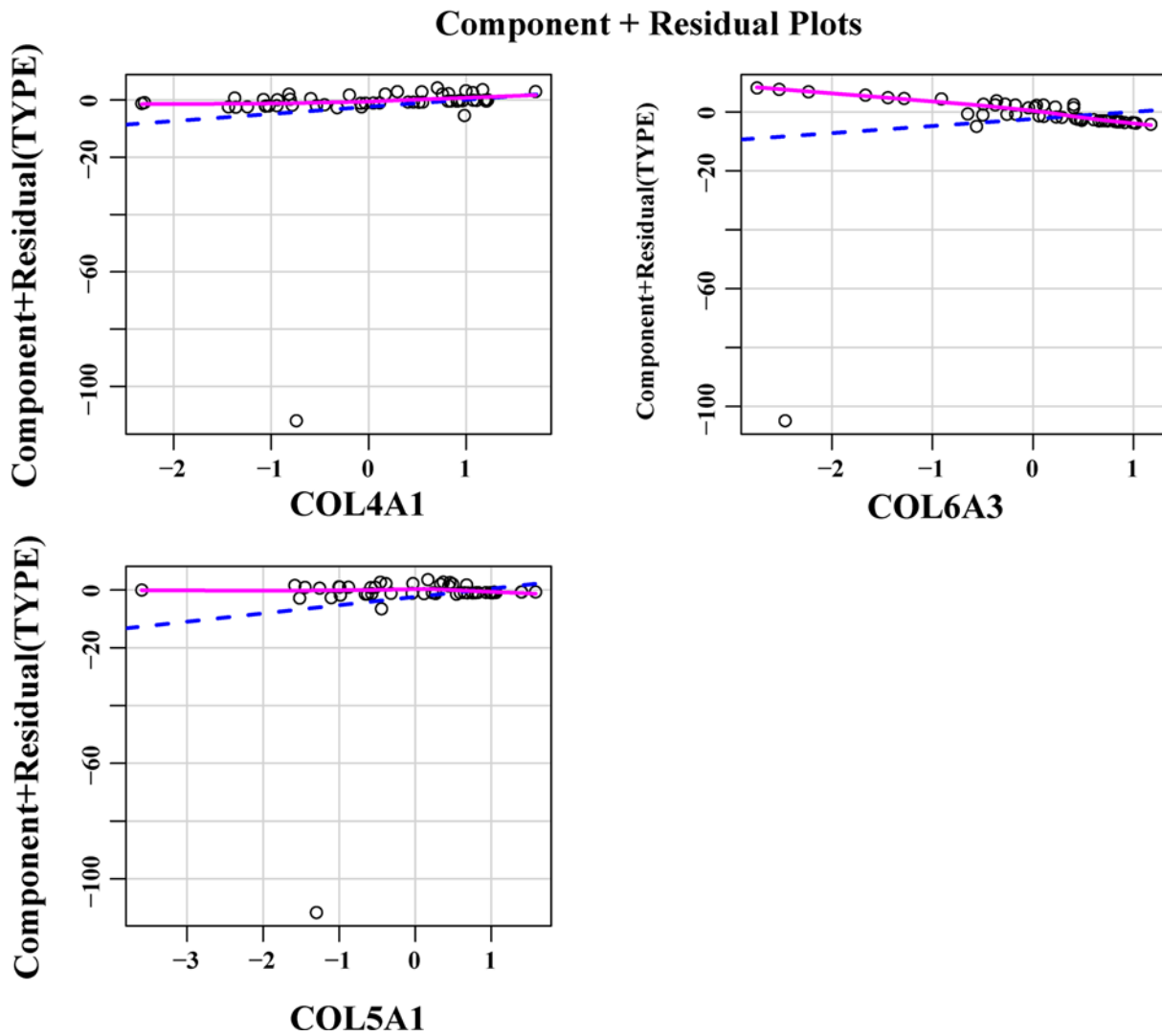
Thus, in the present study, we compared BC and BCM samples in the GSE125989 and GSE100534 datasets, and first identified 146 DEGs in total. Then, based on PPI network construction, ten DEGs were selected as the most crucial genes, including *LAMA4*, *COL1A1*, *COL5A2*, *COL3A1*, *COL4A1*, *COL5A1*, *COL5A3*, *COL6A3*, *COL6A2*, and *COL6A1*. Further PCA results indicated that these genes largely contributed to the differences between BC and BCM samples, suggesting that the ten genes were potentially important genes related to BCM. Previous studies



**Figure 8. Model diagnosis diagram, Residuals vs. Leverage diagram**

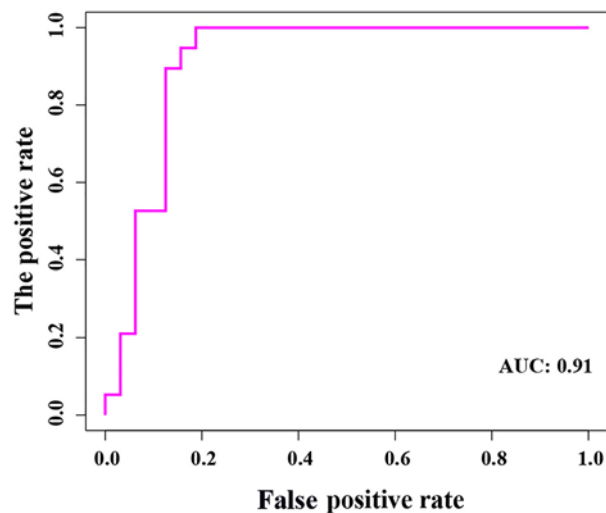
The red dotted line presented the COOK distance, and the points with COOK distance > 0.5 were considered as influential points, which would affect the reliability of the model. There were no influential points in the model.

have demonstrated that BC could regulate the surrounding stroma or tumor microenvironments, and that various components in the BC microenvironment, including suppressive immune cells, soluble factors, and alterations in the ECM, act together to effectively impede antitumor immunity, and promote BC progression and metastasis [31–34]. In this complicated process, some key genes indeed play vital roles. Many, including *BRCA1* [23], *TP53* [35], and *Sdc1* [36] have been reported widely on. Moreover, the distinct expression patterns between BC and BCM might be associated with the microenvironment surrounding the primary tumor, as well as its final metastatic sites [37]. Better understandings of the functional role of DEGs is of importance to identify metastasis-associated therapeutic biomarkers [38]. Thus, we performed GO and KEGG enrichment analyses on 146 DEGs. These DEGs were mainly enriched in GO terms including ECM structural constituent, ECM organization, collagen catabolic process, and extracellular region, which might be responsible for brain metastases. As previous studies have reported, the ECM which affects cancer cell characteristics is increasingly recognized as an important regulator in BC [39]. ECM in BC development features numerous changes in composition and organization when compared with the mammary gland under homeostasis [40]. Prior evidence supports our results to an extent. Additionally, most of the overlapping DEGs were enriched in KEGG pathways associated with cancer metastases, the including PI3K-Akt signaling pathway, the focal adhesion pathway, and ECM-receptor interaction pathways. Notably, it has been reported that mutations and activation of the PI3K signaling pathway in BC cells are linked to brain metastases [41]. Additionally, levels of PI3K pathway activity were significantly increased in brain metastases when compared with extracranial metastases [42]. PI3K/AKT/mTOR signaling has also been reported as one of the most important intracellular pathways, and can be considered a master regulator for cancer. However, the functional details of the overlapped DEGs still need to be further clarified.



**Figure 9.** The component plus residual plot of the three genes included in the model

A relatively obvious linear relationship between the horizontal axis and the vertical axis in the graph could be observed, which indicated that the independent variables were suitable for being included in the model.



**Figure 10.** The ROC analysis results of the model

Previous studies have demonstrated that cancer progression is highly associated with an increase in ECM deposition and stiffening [43–45]. As the major component of the ECM, collagen could increase the ability of cancer cells to reduce apoptosis, and to modulate angiogenesis which gradually promotes cancer progression [46,47]. Moreover, Gilkes et al. reported that intertumoral hypoxia could also lead to increased collagen deposition during breast cancer progression [48]. In our study, survival analyses also showed that increased expression of Collagen genes—including *COL1A1*, *COL14A1*, *COL5A1*, *COL6A1*, and *COL6A2*—was associated with the poor overall survival rates of BC patients. Our results support the idea that the collagen could contribute to tumor progression and result in poor prognoses. Recent studies have suggested that BCBM occurs most frequently in HER2-positive BC, with an incidence of 20–50% [49,50]. As an oncogene, HER2 may drive brain tropism, as it induces a mesenchymal state in BC cells and increases the invasiveness and metastatic potential [51]. The brain microenvironment involves several HER family ligands, including neuregulins, which result in the dimerization and activation of these receptors in metastatic brain cells [52,53]. Recent studies have demonstrated that the risk of cancer cells spreading to the brain is usually highest for women with more aggressive subtypes of BC, such as HER2 positive BC [54,55]. Thus, in this study, we analyzed the correlation between the top ten hub gene with HER2 positive expression. Our results demonstrated that *COL4A1*, *COL5A1*, and *COL6A3* were significantly correlated with HER2 positive expression.

Further, the TF–miRNA coregulatory network indicated that the gene expression patterns of *COL4A1*, *COL5A1*, and *COL6A3* were significantly correlated with each other, and they were also under the regulation of the miRNA-29 family. We speculated that those findings might provide more evidence to illustrate the underlying mechanisms of BCBM progression. Additionally, the logistic regression prediction model based on the three genes could predict whether BM could occur in BC patients, further suggesting that *COL4A1*, *COL5A1*, and *COL6A3* were potentially crucial genes for BCBM. Our results were consistent with some findings from previous studies. A recent study suggested that *COL4A1* was a prognostic biomarker for BC patients who had received neoadjuvant chemotherapy [56]. *COL5A1* was reported to be involved in the regulation of circRNA ACAP2 (circACAP2) in BC [57]. As for the miR-29 family, it has been reported to regulate tumor-related changes, including metastasis [58]. They might also contribute to the involvement of the multiple fibrosis process in many organs' loss of ECM regulation, which might also be associated with the promotion of cancer cell migration and metastasis [59]. Thus, we speculated that the miR-29 family participated in the process of metastasis by modulating the cancer microenvironment. Nevertheless, exactly how the *COL4A1*, *COL5A1*, *COL6A3*, and miR-29 family are involved in the underlying mechanisms of BCBM remains to be further investigated. Moreover, the treatment value of modifying collagen conditions in BCBM is also worth exploring. Cancer resistance has substantially hindered the ability to control cancer. Therefore, both cancer cells and tumor microenvironment must be treated, and collagen is a potential target [46]. The result obtained from Awasthi's group showed that anti-matrix metalloproteinase 9 (MMP-9) antibody therapy could decrease the COLI expression and the metastatic burden in pancreatic ductal adenocarcinoma mouse models [60]. Employing therapeutic agents directly or indirectly targeting collagen might be another promising therapeutic strategy in the treatment of cancers including BC.

There are also some limitations to our study. First, the present study was based on bioinformatics analysis of published datasets, and does not have experimental verification of the observations. Second, as we combined only two GEO datasets, the total sample size is relatively small which may result in some biasing of the results. Finally, we only analyzed the correlation between top ten hub genes with HER2 positive expression. It would also be important to investigate the correlations with other BCBM oncogenes, including estrogen receptors (ERs). Future clinical studies with larger cohorts are needed to validate our results.

## Conclusions

Here, by comparing the BC and BCBM specimens across two datasets using a series of bioinformatics analyses, we identified 146 DEGs that were potentially related to the BCBM. The DEGs were significantly enriched for ECM organization, collagen catabolic process, and cellular adhesion pathways. Among these, *COL4A1*, *COL5A1*, and *COL6A3* were potentially crucial genes that were associated with BCBM in BC patients. Our findings should provide more reference information for diagnostic and prognostic investigation of BCBM in the future.

## Data Availability

The datasets are available in the National Center of Biotechnology Information (NCBI) GEO (<https://www.ncbi.nlm.nih.gov/geo/>). Two datasets, including GSE125989 and GSE100534, were used in the present study.

## Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Funding

This work was supported by the Scientific Research Starting Foundation of Affiliated Hospital of Hebei University [grant number 31010413]; the Hebei Natural Science Foundation Precision Medicine Joint Project [grant number H2020201206]; and The Central Government Guided Local Science and Technology Development Fund Project of Hebei Province [grant number 216Z7711G].

## CRedit Author Contribution

**Lijian Zhang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft. **Luxuan Wang:** Data curation, Formal analysis, Investigation, Methodology, Writing—original draft. **Hua Yang:** Formal analysis, Investigation, Methodology, Writing—original draft. **Chunhui Li:** Conceptualization, Supervision, Validation, Investigation, Writing—review & editing. **Chuan Fang:** Conceptualization, Supervision, Investigation, Visualization, Project administration, Writing—review & editing.

## Abbreviations

BC, breast cancer; BCBM, breast cancer brain metastasis; BM, brain metastasis; BP, biological process; CC, cellular component; circACAP2, circRNA ACAP2; CNS, central nervous system; COL14A1, collagen type XIV  $\alpha$  1 chain; COL1A1, collagen type I  $\alpha$  1 chain; COL3A1, collagen type III  $\alpha$  1 chain; COL5A2, collagen type V  $\alpha$  2 chain; COL5A3, collagen type V  $\alpha$  3 chain; COL6A1, collagen type VI  $\alpha$  1 chain; COL6A2, collagen type VI  $\alpha$  2 chain; COL6A3, collagen type VI  $\alpha$  3 chain; DAVID, The Database for Annotation, Visualization and Integrated Discovery; DEG, differentially expressed gene; ECM, extracellular matrix; GEO, Gene Expression Omnibus; GEPIA, Gene Expression Profiling Interactive Analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; LAMA4, laminin subunit  $\alpha$  4; MF, molecular function; MMP-9, Matrix metalloproteinase 9; PCA, principal component analysis; ROC, receiver operating characteristic; STRING, Search Tool for the Retrieval of Interacting Genes; TCGA, The Cancer Genome Atlas; TF, transcription factor.

## References

- 1 Fahad Ullah, M. (2019) Breast cancer: current perspectives on the disease status. *Adv. Exp. Med. Biol.* **1152**, 51–64, [https://doi.org/10.1007/978-3-030-20301-6\\_4](https://doi.org/10.1007/978-3-030-20301-6_4)
- 2 Lowery, F.J. and Yu, D. (2017) Brain metastasis: unique challenges and open opportunities. *Biochim. Biophys. Acta Rev. Cancer* **1867**, 49–57, <https://doi.org/10.1016/j.bbcan.2016.12.001>
- 3 Custodio-Santos, T., Videira, M. and Brito, M.A. (2017) Brain metastasization of breast cancer. *Biochim. Biophys. Acta Rev. Cancer* **1868**, 132–147, <https://doi.org/10.1016/j.bbcan.2017.03.004>
- 4 Vogelbaum, M.A. and Suh, J.H. (2006) Resectable brain metastases. *J. Clin. Oncol.* **24**, 1289–1294, <https://doi.org/10.1200/JCO.2005.04.6235>
- 5 Narita, Y. and Shibui, S. (2009) Strategy of surgery and radiation therapy for brain metastases. *Int. J. Clin. Oncol.* **14**, 275–280, <https://doi.org/10.1007/s10147-009-0917-0>
- 6 Paek, S.H., Audu, P.B., Sperling, M.R., Cho, J. and Andrews, D.W. (2005) Reevaluation of surgery for the treatment of brain metastases: review of 208 patients with single or multiple brain metastases treated at one institution with modern neurosurgical techniques. *Neurosurgery* **56**, 1021–1034
- 7 Sawaya, R., Hammoud, M., Schoppa, D., Hess, K.R., Wu, S.Z., Shi, W.M. et al. (1998) Neurosurgical outcomes in a modern series of 400 craniotomies for treatment of parenchymal tumors. *Neurosurgery* **42**, 1044–1055, discussion 55–56, <https://doi.org/10.1097/00006123-199805000-00054>
- 8 Deeken, J.F. and Loscher, W. (2007) The blood-brain barrier and cancer: transporters, treatment, and Trojan horses. *Clin. Cancer Res.* **13**, 1663–1674, <https://doi.org/10.1158/1078-0432.CCR-06-2854>
- 9 Beasley, K.D. and Toms, S.A. (2011) The molecular pathobiology of metastasis to the brain: a review. *Neurosurg. Clin. N. Am.* **22**, 7–14, v., <https://doi.org/10.1016/j.nec.2010.08.009>
- 10 Tarin, D., Price, J.E., Kettlewell, M.G.W. et al. (1984) Mechanisms of human-tumor metastasis studied in patients with peritoneovenous shunts. *Cancer Res.* **44**, 3584–3592
- 11 Disibio, G. and French, S.W. (2008) Metastatic patterns of cancers: results from a large autopsy study. *Arch. Pathol. Lab. Med.* **132**, 931–939, <https://doi.org/10.5858/2008-132-931-MPOCRF>
- 12 Kimbung, S., Loman, N. and Hedenfalk, I. (2015) Clinical and molecular complexity of breast cancer metastases. *Semin. Cancer Biol.* **35**, 85–95, <https://doi.org/10.1016/j.semcancer.2015.08.009>
- 13 Van Swearingen, A.E., Siegel, M.B. and Anders, C.K. (2014) Breast cancer brain metastases: evidence for neuronal-like adaptation in a ‘breast-to-brain’ transition? *Breast Cancer Res.* **16**, 304, <https://doi.org/10.1186/bcr3651>
- 14 Cacho-Diaz, B., Garcia-Botello, D.R., Wegman-Ostrosky, T., Reyes-Soto, G., Ortiz-Sanchez, E. and Herrera-Montalvo, L.A. (2020) Tumor microenvironment differences between primary tumor and brain metastases. *J. Transl. Med.* **18**, 1, <https://doi.org/10.1186/s12967-019-02189-8>
- 15 Valiente, M., Obenauf, A.C., Jin, X., Chen, Q., Zhang, X.H., Lee, D.J. et al. (2014) Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* **156**, 1002–1016, <https://doi.org/10.1016/j.cell.2014.01.040>



- 16 Romagnoli, M., Mineva, N.D., Polmear, M., Conrad, C., Srinivasan, S., Loussouarn, D. et al. (2014) ADAM8 expression in invasive breast cancer promotes tumor dissemination and metastasis. *EMBO Mol. Med.* **6**, 278–294, <https://doi.org/10.1002/emmm.201303373>
- 17 Malin, D., Strelakova, E., Petrovic, V., Deal, A.M., Al Ahmad, A., Adamo, B. et al. (2014) alphaB-crystallin: a novel regulator of breast cancer metastasis to the brain. *Clin. Cancer Res.* **20**, 56–67, <https://doi.org/10.1158/1078-0432.CCR-13-1255>
- 18 Neman, J., Choy, C., Kowolik, C.M., Anderson, A., Duenas, V.J., Waliy, S. et al. (2013) Co-evolution of breast-to-brain metastasis and neural progenitor cells. *Clin. Exp. Metastasis* **30**, 753–768, <https://doi.org/10.1007/s10585-013-9576-7>
- 19 Owonikoko, T.K., Arbisser, J., Zelnak, A., Shu, H.K., Shim, H., Robin, A.M. et al. (2014) Current approaches to the treatment of metastatic brain tumours. *Nat. Rev. Clin. Oncol.* **11**, 203–222, <https://doi.org/10.1038/nrclinonc.2014.25>
- 20 Qin, X.B., Zhang, W.J., Zou, L., Huang, P.J. and Sun, B.J. (2016) Identification potential biomarkers in pulmonary tuberculosis and latent infection based on bioinformatics analysis. *BMC Infect. Dis.* **16**, 500, <https://doi.org/10.1186/s12879-016-1822-6>
- 21 Lee, S.W., Wu, L.S., Huang, G.M., Huang, K.Y., Lee, T.Y. and Weng, J.T. (2016) Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis. *BMC Bioinformatics* **17**, 3, <https://doi.org/10.1186/s12859-015-0848-x>
- 22 Iwamoto, T., Niikura, N., Ogiya, R., Yasojima, H., Watanabe, K.I., Kanbayashi, C. et al. (2019) Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. *Sci. Rep.* **9**, 13343, <https://doi.org/10.1038/s41598-019-50099-y>
- 23 Schulten, H.J., Bangash, M., Karim, S., Dallol, A., Hussein, D., Merdad, A. et al. (2017) Comprehensive molecular biomarker identification in breast cancer brain metastases. *J. Transl. Med.* **15**, 269, <https://doi.org/10.1186/s12967-017-1370-x>
- 24 Suzuki, A., Horie, T. and Numabe, Y. (2019) Investigation of molecular biomarker candidates for diagnosis and prognosis of chronic periodontitis by bioinformatics analysis of pooled microarray gene expression datasets in Gene Expression Omnibus (GEO). *BMC Oral Health* **19**, 52, <https://doi.org/10.1186/s12903-019-0738-0>
- 25 Chong, X., Peng, R., Sun, Y., Zhang, L. and Zhang, Z. (2020) Identification of key genes in gastric cancer by bioinformatics analysis. *Biomed Res. Int.* **2020**, 7658230, <https://doi.org/10.1155/2020/7658230>
- 26 Parent, B.A., Seaton, M., Sood, R.F., Gu, H., Djukovic, D., Raftery, D. et al. (2016) Use of metabolomics to trend recovery and therapy after injury in critically ill trauma patients. *JAMA Surg.* **151**, e160853, <https://doi.org/10.1001/jamasurg.2016.0853>
- 27 Tang, Z., Li, C., Kang, B., Gao, G., Li, C. and Zhang, Z. (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102, <https://doi.org/10.1093/nar/gkx247>
- 28 Tang, D., Zhao, X., Zhang, L., Wang, Z. and Wang, C. (2019) Identification of hub genes to regulate breast cancer metastasis to brain by bioinformatics analyses. *J. Cell. Biochem.* **120**, 9522–9531, <https://doi.org/10.1002/jcb.28228>
- 29 Wang, L., Lu, F. and Xu, J. (2021) Identification of potential miRNA-mRNA regulatory network contributing to hypertrophic cardiomyopathy (HCM). *Front. Cardiovasc. Med.* **8**, 660372, <https://doi.org/10.3389/fcvm.2021.660372>
- 30 Bailleux, C., Eberst, L. and Bachelot, T. (2021) Treatment strategies for breast cancer brain metastases. *Br. J. Cancer* **124**, 142–155, <https://doi.org/10.1038/s41416-020-01175-y>
- 31 Soysal, S.D., Tzankov, A. and Muenst, S.E. (2015) Role of the tumor microenvironment in breast cancer. *Pathobiology* **82**, 142–152, <https://doi.org/10.1159/000430499>
- 32 Tan, W., Zhang, W., Strasner, A., Grivnennikov, S., Cheng, J.Q., Hoffman, R.M. et al. (2011) Tumour-infiltrating regulatory T cells stimulate mammary cancer metastasis through RANKL-RANK signalling. *Nature* **470**, 548–553, <https://doi.org/10.1038/nature09707>
- 33 Zhang, J., Zuo, T., Yang, J., Hu, Z., Wang, Z., Xu, R. et al. (2021) Hierarchically releasing bio-responsive nanoparticles for complete tumor microenvironment modulation via TGF- $\beta$  pathway inhibition and TAF reduction. *ACS Appl. Mater. Interfaces* **13**, 2256–2268, <https://doi.org/10.1021/acsmi.0c18545>
- 34 Wan, B., Liu, B., Huang, Y. and Lv, C. (2020) Identification of genes of prognostic value in the ccRCC microenvironment from TCGA database. *Mol. Genet. Genomic Med.* **8**, e1159, <https://doi.org/10.1002/mgg3.1159>
- 35 Lee, J.Y., Park, K., Lee, E., Ahn, T., Jung, H.H., Lim, S.H. et al. (2016) Gene expression profiling of breast cancer brain metastasis. *Sci. Rep.* **6**, 28623, <https://doi.org/10.1038/srep28623>
- 36 Sayyad, M.R., Puchalapalli, M., Vergara, N.G., Wangenstein, S.M., Moore, M., Mu, L. et al. (2019) Syndecan-1 facilitates breast cancer metastasis to the brain. *Breast Cancer Res. Treat.* **178**, 35–49, <https://doi.org/10.1007/s10549-019-05347-0>
- 37 Tomczak, K., Czerwinska, P. and Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. (Pozn.)* **19**, A68–A77, <https://doi.org/10.5114/wo.2014.47136>
- 38 Gao, F., Liang, B., Reddy, S.T., Farias-Eisner, R. and Su, X. (2014) Role of inflammation-associated microenvironment in tumorigenesis and metastasis. *Curr. Cancer Drug Targets* **14**, 30–45, <https://doi.org/10.2174/15680096113136660107>
- 39 Komemi, O., Shochet, G.E., Pomeranz, M., Fishman, A., Pasmanik-Chor, M., Drucker, L. et al. (2019) Placenta-conditioned extracellular matrix (ECM) activates breast cancer cell survival mechanisms: a key for future distant metastases. *Int. J. Cancer* **144**, 1633–1644, <https://doi.org/10.1002/ijc.31861>
- 40 Insua-Rodriguez, J. and Oskarsson, T. (2016) The extracellular matrix in breast cancer. *Adv. Drug Deliv. Rev.* **97**, 41–55, <https://doi.org/10.1016/j.addr.2015.12.017>
- 41 Blazquez, R., Wlochowitz, D., Wolff, A., Seitz, S., Wachter, A., Perera-Bel, J. et al. (2018) PI3K: a master regulator of brain metastasis-promoting macrophages/microglia. *Glia* **66**, 2438–2455, <https://doi.org/10.1002/glia.23485>
- 42 Chen, G., Chakravarti, N., Aardalen, K., Lazar, A.J., Tetzlaff, M.T., Wubbenhorst, B. et al. (2014) Molecular profiling of patient-matched brain and extracranial melanoma metastases implicates the PI3K pathway as a therapeutic target. *Clin. Cancer Res.* **20**, 5537–5546, <https://doi.org/10.1158/1078-0432.CCR-13-3003>
- 43 Provenzano, P.P., Inman, D.R., Eliceiri, K.W., Knittel, J.G., Yan, L., Rueden, C.T. et al. (2008) Collagen density promotes mammary tumor initiation and progression. *BMC Med.* **6**, 11, <https://doi.org/10.1186/1741-7015-6-11>

- 44 Lopez, J.I., Kang, I., You, W.K., McDonald, D.M. and Weaver, V.M. (2011) In situ force mapping of mammary gland transformation. *Integr. Biol. (Camb.)* **3**, 910–921, <https://doi.org/10.1039/c1ib00043h>
- 45 Junqueira, L.C., Bignolas, G. and Brentani, R.R. (1979) Picrosirius staining plus polarization microscopy, a specific method for collagen detection in tissue sections. *Histochem. J.* **11**, 447–455, <https://doi.org/10.1007/BF01002772>
- 46 Xu, S., Xu, H., Wang, W., Li, S., Li, H., Li, T. et al. (2019) The role of collagen in cancer: from bench to bedside. *J. Transl. Med.* **17**, 309, <https://doi.org/10.1186/s12967-019-2058-1>
- 47 Morris, B.A., Burkel, B., Ponik, S.M. et al. (2016) Collagen matrix density drives the metabolic shift in breast cancer cells. *EBioMedicine* **13**, 146–156, <https://doi.org/10.1016/j.ebiom.2016.10.012>
- 48 Gilkes, D.M., Chaturvedi, P., Bajpai, S., Wong, C.C., Wei, H., Pitcairn, S. et al. (2013) Collagen prolyl hydroxylases are essential for breast cancer metastasis. *Cancer Res.* **73**, 3285–3296, <https://doi.org/10.1158/0008-5472.CAN-12-3963>
- 49 Hicks, D.G., Short, S.M., Prescott, N.L., Tarr, S.M., Coleman, K.A., Yoder, B.J. et al. (2006) Breast cancers with brain metastases are more likely to be estrogen receptor negative, express the basal cytokeratin CK5/6, and overexpress HER2 or EGFR. *Am. J. Surg. Pathol.* **30**, 1097–1104, <https://doi.org/10.1097/01.pas.0000213306.05811.b9>
- 50 Witzel, I., Oliveira-Ferrer, L., Pantel, K., Muller, V. and Wikman, H. (2016) Breast cancer brain metastases: biology and new clinical perspectives. *Breast Cancer Res.* **18**, 8, <https://doi.org/10.1186/s13058-015-0665-1>
- 51 Gupta, P. and Srivastava, S.K. (2014) HER2 mediated de novo production of TGFbeta leads to SNAIL driven epithelial-to-mesenchymal transition and metastasis of breast cancer. *Mol. Oncol.* **8**, 1532–1547, <https://doi.org/10.1016/j.molonc.2014.06.006>
- 52 Da Silva, L., Simpson, P.T., Smart, C.E., Cocciardi, S., Waddell, N., Lane, A. et al. (2010) HER3 and downstream pathways are involved in colonization of brain metastases from breast cancer. *Breast Cancer Res.* **12**, R46, <https://doi.org/10.1186/bcr2603>
- 53 Momeny, M., Saunus, J.M., Marturana, F., McCart Reed, A.E., Black, D., Sala, G. et al. (2015) Heregulin-HER3-HER2 signaling promotes matrix metalloproteinase-dependent blood-brain-barrier transendothelial migration of human breast cancer cell lines. *Oncotarget* **6**, 3932–3946, <https://doi.org/10.18632/oncotarget.2846>
- 54 Evans, A.J., James, J.J., Cornford, E.J., Chan, S.Y., Burrell, H.C., Pinder, S.E. et al. (2004) Brain metastases from breast cancer: identification of a high-risk group. *Clin. Oncol.* **16**, 345–349, <https://doi.org/10.1016/j.clon.2004.03.012>
- 55 Gaedcke, J., Traub, F., Milde, S., Wilkens, L., Stan, A., Ostertag, H. et al. (2007) Predominance of the basal type and HER-2/neu type in brain metastasis from breast cancer. *Mod. Pathol.* **20**, 864–870, <https://doi.org/10.1038/modpathol.3800830>
- 56 Wang, S.M., Chen, P.M., Sung, Y.W., Huang, W.C., Huang, H.S. and Chu, P.Y. (2020) Effect of COL4A1 expression on the survival of neoadjuvant chemotherapy breast cancer patients. *J. Oncol.* **2020**, 5209695, <https://doi.org/10.1155/2020/5209695>
- 57 Zhao, B., Song, X. and Guan, H. (2020) CircACAP2 promotes breast cancer proliferation and metastasis by targeting miR-29a/b-3p-COL5A1 axis. *Life Sci.* **244**, 117179, <https://doi.org/10.1016/j.lfs.2019.117179>
- 58 Wang, Y., Zhang, X., Li, H., Yu, J. and Ren, X. (2013) The role of miRNA-29 family in cancer. *Eur. J. Cell Biol.* **92**, 123–128, <https://doi.org/10.1016/j.ejcb.2012.11.004>
- 59 Gebeshuber, C.A., Zatloukal, K. and Martinez, J. (2009) miR-29a suppresses tristetraprolin, which is a regulator of epithelial polarity and metastasis. *EMBO Rep.* **10**, 400–405, <https://doi.org/10.1038/embor.2009.9>
- 60 Awasthi, N., Mikels-Vigdal, A.J., Stefanutti, E. et al. (2019) Therapeutic efficacy of anti-MMP9 antibody in combination with nab-paclitaxel-based chemotherapy in pre-clinical models of pancreatic cancer. *J. Cell. Mol. Med.* **23**, 3878–3887, <https://doi.org/10.1111/jcmm.14242>