CILP Inhibits Brain Metastasis by Meditating Mast Cells via the MAPK Signaling Pathway in Breast Cancer

Xiaolin Sun
Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University
Xingguo Zhou
The Second Hospital, Cheeloo College of Medicine, Shandong University
Alei Feng
Shandong Provincial Hospital Affiliated to Shandong First Medical University
Gongwen Xu
Shandong Jianzhu University
Qiang Wang
Shandong Provincial Hospital Affiliated to Shandong First Medical University
Qiang Li
Shandong Provincial Hospital Affiliated to Shandong First Medical University
Linzong Xu
Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University
Zhanyu Zhang
Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University
Yingchao Liu
Shandong Provincial Hospital Affiliated to Shandong First Medical University
Zhe Yang
Shandong Provincial Hospital Affiliated to Shandong First Medical University
Xiaomei Li (sdulixiaomei@163.com)
Shandong Provincial Hospital https://orcid.org/0000-0002-3221-5888

Primary research

Keywords: breast cancer, brain metastases, tumor-infiltrating immune cells, CILP, mast cells resting, mast cells, MAPK Signaling Pathway

Posted Date: March 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-328026/v1

Abstract

Background: Approximately 15–30% of patients with breast cancer (BRCA) eventually develop brain metastases (BMs) with high morbidity and mortality. Herein, we aimed to identify genes specific to breast cancer brain metastases (BCBM) from an immune infiltration perspective.

Methods: GSE100534 and GSE125989 were obtained from the NCBI Gene Expression Omnibus (GEO), then performed normalization using Rstudio and perl 5. We constructed a Weighted Gene Co-Expression Network Analysis (WGCNA) and obtained differentially expressed genes (DEGs) in BMs sample compared with primary BRCA tissue. Then we performed GO and KEGG pathway analysis. The LinkedOmics and UALCAN analysis showed the expression of gene in BRCA. The Kaplan-Meier plotter database was used to evaluate the prognosis. The composition of significant tumor-infiltrating immune cells was assessed using the CIBERSORT algorithm. Spearman's correlation analysis revealed the correlation between CILP gene and immune cells in TCGA cohort and Timer database. Using GSEA analysis, we conducted to identify the potential pathways in BCBM.

Results: The cartilage intermediate layer protein (CILP) was a late event in BRCA (stage III to IV) with poor prognosis (P< 0.05). LinkedOmics showed that the mRNA expression of CILP was down-regulated in advanced cancer (P< 0.05). Besides, UALCAN analysis showed that CILP expression was downregulated in HER2-positive and triple-negative breast cancer which were more prone to BMs (P< 0.05). CILP was the hub gene which was significantly associated with BCBM identified by WGCNA (R2=−0.6, P=3e-06). We found that the resting infiltration of mast cells in the BCBM group was significantly lower than that in the primary BRCA group (P= 0.01). In addition, Spearman's correlation analysis revealed that the expression of CILP positively correlated with that of mast cells (P< 0.05). Finally, the FCERI-mediated MAPK activation (NES=2.1847, P=0, FDR=0.0031), which could regulate mast cell activity, were enriched in BCBM.

Conclusions: CILP can influence the progression of BRCA favored for BMs through meditating mast cells via the MAPK signaling pathway.

Background

Of all the solid tumors, breast cancer (BRCA) is the second most common cause of brain metastases (BMs) (*1-3*). Approximately 15–30% of patients with BRCA eventually develop BMs with high morbidity and mortality (*3, 4*). BMs occur in 30–55% of HER2-positive (HER2+) metastatic BRCA patients and up to half die from intracranial progression (*5*), whereas the median survival rate is only 6 months in triple-negative BRCA (TNBC) with BMs (*6, 7*). Unfortunately, effective treatment is not available since the central nervous system (CNS) is traditionally considered to be a privileged site due to blood-brain barrier (*2, 8*). As a result, identification of genetic and epigenetic alterations is essential to the development of BMs targeted therapies (*9*).

Pursuing of the mechanism involved in BMs has never stopped although it still remains largely unclear. Bos et al. showed that ST6GALNAC5 acted as a specific mediator of BMs through enhancing its adhesion to brain endothelial cells and facilitating crossing the blood-brain barrier by BRCA cells (*10*). Highly sialylated N-glycans which were up-regulated in the brain-seeking cell line 231BR likely play a crucial part in breast cancer brain metastasis (BCBM) evaluated by integrated transcriptomics, glycomics, and proteomics (*11*). GABAA receptor alpha3 (Gabra3), normally exclusively expressed in the adult brain, was inversely correlated with BRCA survival and promoted BRCA cell invasion and BMs by activating the AKT pathway (*12*). Recently, it was also reported that YTHDF3 could enhance the translation of m6A-enriched transcripts for ST6GALNAC5, EGFR, and GJA1 thus prompting BCBM (*13*).

The tumor microenvironment, composed of diverse immune cells, tumor cells, and cytokines, has both an adverse and beneficial impact on tumorigenesis (*14*). The is growing evidence indicating that an ineffective immune response influences the behavior of BRCA, suggesting that it is an immunogenic cancer type as well (*15*). Berghoff et al. demonstrated that the immune reaction to BMs was mostly characterized by activation of microglia/macrophages and upregulation of biomarkers involved in phagocytosis (*16*). However, an understanding of the microenvironment and its underlying molecular mechanisms contributing to the BCBM are limited (*17*).

Herein, especially from an immune infiltration perspective and using large TCGA and the GEO datasets of BRCA samples, we used a series of bioinformatics tools, such as weighted gene co-expression network analysis (WGCNA), GSEA analysis, and CIBERSORT estimation to identify genes specific to BCBM.

Methods

Data acquisition and processing

The invasive carcinoma (BRCA) datasets GSE100534 and GSE125989 were obtained from the NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/gds). GSE125989 consists of 32 BRCA samples with matched primary BRCAs and BM from 16 patients analyzed using the GPL571 ([HG-U133A_2 Affymetrix] Human Genome U133A 2.0 Array) platform. As for the GSE100534, we obtained 19 BRCA samples including 3 BCBM samples and 16 BRCA samples analyzed using the GPL6244 ([HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]) platform. The R packages of “GEOquery” and “oligo” were used to access and process and normalize data, using the (Robust Multichip Average (RMA) algorithm) and normalization of the raw data (.CEL files) and matrix construction. The R packages of “hugene10sttranscriptcluster” (GPL6244) and “hgu133a2” (GPL571) were respectively used to match the probe to their gene symbol, and the probes matching several genes were removed. For genes matched by multiple probes, we selected the probe with the highest expression average in the samples. We combined the GSE100534 and GSE125989 datasets by the perl script, then used the “sva” package (under the R environment, version 3.6.3) to preprocess and remove the batch effect and eventually obtained a merged dataset. The Fragments Per Kilobase Million (FPKM) data of TCGA RNA-Seq were downloaded from TCGA (https://portal.gdc.cancer.gov/) and contained 1208 samples (normal: 112;

tumor: 1096). The clinical follow-up information contained 1097 samples. For our study, samples with no clinical information were excluded.

Analysis of differentially expressed genes

We performed a differential analysis comparing primary BRCA tissues and BCBM tissues using the merging GEO dataset. Differentially expressed genes (DEGs) were displayed in a volcano plot and heatmap, screened, and then matrixes were constructed using the “limma” package in RStudio using log2FoldChange (FC)≥ 1 and adjusted P-value<0.05 as cut-off values.

Construction of weighted gene co-expression network analysis

The data used were from the merging dataset of GSE100534 and GSE125989 as described above from the GEO database. We chose all 11,786 genes for the weighted gene co-expression network analysis (WGCNA), then the R package WGCNA (*18*) was conducted, and the power parameter was pre-calculated using the pickSoftThreshold function. By calculating scale-free topology fit exponentials for several powers, it is possible to provide appropriate soft threshold power for network construction. After choosing appropriate soft-thresholding power, the adjacency was transformed into topological overlap matrix (TOM), which measured the network connectivity of a gene (*19*). In order to classify genes with similar expression profiles into gene modules, an average linkage hierarchical clustering was carried out according to the TOM-based dissimilarity measure with a minimum module size of 30 for the gene dendrogram (*20*). To further analyze the module, we merged highly similar modules with the dissimilarity of<0.25 by implementing clustering of module eigengenes. Then we calculated the correlation between module eigengenes (MEs) and clinical features of BRCA. As a representative of all the genes in each module, MEs were defined as the first principal component of each gene module. Gene significance (GS) was defined as the $\log_{10}$ transformation of the P-value (GS=lgP) in the linear regression between gene expression of the module and clinical features. In addition, module significance was defined as the mean GS for all genes in the module. The visualization in gene network of eigengenes was also carried out and displayed by Cytoscape 3.8.0.

Biological function and pathway enrichment analysis

In gene networks that conform to scale-free distributions, genes with similar expression patterns could be synergistically regulated, pathway shared, or functionally related. We selected the module of the most relevant to clinical characteristics, and then we performed gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis via the “clusterprofiler” package in RStudio software to acquire the enriched biological process and KEGG pathways for further analysis (*21*).

Kaplan-Meier Plotter Database analysis

The correlation between CILP expression and survival in BRCA was analyzed by Kaplan-Meier plotter (http://kmplot.com/analysis/). The hazard ratio (HR) with 95% confidence intervals (CI) and log-rank P-

values were computed.

Analysis of CILP alterations in breast cancer samples

The LinkedOmics database (http://www.linkedomics.org/login.php), which is applied to analyze 32 TCGA cancer-related multi-dimensional datasets, is a Web-based platform. The data types include microRNA, SNP, methylation status, gene mutations, and clinical data (*22*). We performed a non-parametric analysis of the clinical data in TCGA BRCA cohort, the transcriptional variation in the levels of CILP expression in different pathologic stages (Kruskal-Wallis Test) and pathology M stage (Wilcox Test) was evaluated using this platform. Next, we applied UALCAN (http://ualcan.path.uab.edu/) (*23*) to analyze the transcriptional levels of CILP in breast invasive carcinoma and their association with intrinsic BRCA subclasses.

GSEA analysis

Samples from the TCGA were divided into two groups based on the expression of CILP by the surv_cutpoint function (res.cut value) and gene set enrichment analysis (GSEA) software (http://software.broadinstitute.org/gsea/index.jsp) was applied in the two groups to identify the pathways associated with change in CILP. The cut-off standard for the GSEA were P-values <0.05 and false discovery rate (FDR)<0.25.

TIMER Database Analysis

TIMER is a comprehensive resource for systematic analysis of immune infiltrates across multiple cancer types (http://timer.comp-genomics.org/) (*24*). Statistical analyses that correlated the expression of CILP and the presence of mast cells using Spearman's statistical analysis. The results an indication of the purity-adjusted partial Spearman's rho value as a degree of their correlation.

CIBERSORT Estimation

We assessed 22 types of immune cell utilizing the CIBERSORT algorithm. Only samples with a CIBERSORT output of P<0.05 were considered worthy of further analysis. Significant differences in the proportion of immune infiltrating cells between the primary BRCA tissue and BM tissue were determined using the Wilcoxon rank-sum test. Infiltration of immune cells from adjacent normal tissue and BRCA tissues were also analyzed. The level of immune cell infiltration in each sample was combined with the pathological stage using EXCEL, and samples lacking clinical information were deleted; thus, 800 BRCA samples were ultimately obtained for further analysis.

Statistical analysis

All statistical analyses were carried out using Rstudio 3.6.3 and perl 5 (v5.30.0). We performed survival analysis using the "survminer" and "survival" packages in RStudio software. We then used surv_cutpoint function to get the best cutoff dividing genes into high and low expression groups for clinical analysis.

The strength of the Spearman's correlation was determined using the following guide for the absolute value: 0.00–0.19, "very weak;" 0.20–0.39, "weak"; 0.40–0.59, "moderate"; 0.60–0.79, "strong"; and 0.80–1.0, "very strong". The statistical significance was described as follows: ns, not significant; *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001; ****P ≤ 0.0001.

Results

CILP acted as a late event in BRCA with poor prognosis

The Function module of LinkedOmics was used to analyze mRNA sequencing data from 1093 BRCA patients in TCGA. Generally, CILP mRNA levels were reduced with higher grade of clinical stage (Figure 1a). Thus, it was lower in the M1 stage than in the M0 stage (Figure 1b). As shown in Figure 1c, the mRNA expression of CILP different across the normal and Luminal, Her2 positive, and TN intrinsic subclasses of BRCA. The expression of CILP was significantly lower in HER2 positive and TNBC than luminal subclass. To further examine the prognostic potential of CILP in BRCAs, Kaplan-Meier plotter database was used to evaluate the CILP prognostic value. Interestingly, the poor prognosis in BRCA was shown to correlate with lower CILP expression (Figure 1d). However, CILP expression had no effects on survival in early BRCA (stage I to II), and showed a worse overall survival (OS) in advanced BRCA (stage III to IV) of RNA-seq data (Figure 1e).

From the above findings, CILP may serve as a late event occurring in BRCA with poor prognosis.

Identification of markedly different expressed genes between breast cancer tissue and brain metastases tissue from breast cancer

To identify DEGs between BRCA tissue and BM tissue from BRCA, data from 51 samples from two independent mRNA expression arrays of GSE100534 and GSE125989 datasets (*25, 26*) were downloaded from the GEO database and normalized and merged using the "oligo" and "sva" packages in RStudio (Figure 2a). According to the thresholds set (adjusted P-value<0.05 and log2FC ≥ 1), 90 DEGs (62 downregulated and 28 upregulated) were identified (Figure 2b). Subsequently, we visualized the 90 DEGs in the form of heatmap including all samples (Figure 2c).

CILP was the hub gene significantly associated with BCBM identified by construction of weighted co-expression network (WGCNA)

We used the merged data to construct the gene co-expression networks, which included a total of 51 samples complete with clinical features data for the WGCNA. Raw data were preprocessed identically using RStudio for background correction and normalization. Finally, we obtained a total 11,786 genes, and all were chosen for WGCNA. WGCNA is a systematic biological approach used to analyze the expression patterns of multiple genes in different samples, which forms clusters or modules containing genes with the same expression pattern (*27*). If certain genes are located in the same module, they are likely to possess the same biological functions and the association between modules and sample

characteristics such as clinical traits can be studied (*28, 29*). The selection of the soft-thresholding power is an important step when constructing a WGCNA. We performed the network topology analysis for thresholding powers from 1 to 20 and identified the relatively balanced scale independence and mean connectivity of the WGCNA. In this study, we selected the power of β=8 (scale free R^2=0.83) as the soft-thresholding to achieve a scale-free network (Figures 3a, 3b). As a result, we set the cut height as 0.25 to merge similar modules (Figure 3c). Twenty gene-co-expression modules were identified using merged dynamic tree cut (Figure 3d). The yellow module showed the highest correlation with BM phenotype of invasive BRCA (Figure 3e). There were 399 genes in the yellow module, data for yellow module were selected for further analysis. Subsequently, we identified the module containing CILP, that is., the yellow module, as a key module accounting for the highest correlation with BRCA brain metastasis phenotype (R2=−0.6, P=3e-06) (Figure 3e). A total of 24 genes, including the CILP gene, were identified as candidate hub genes for the yellow module (Figure 3f). It is well-known that CILP is an inhibitor of transforming growth factor (TGF)-beta signaling. Interestingly, TGF-beta can enable cells to acquire a migratory and infiltrating phenotype, thus enabling the spread of cancer cells.

We further performed enrichment analysis to explore the biological processes and pathways in which the yellow module was involved. GO biological processes and the KEGG functional enrichment analyses were conducted using the R package “clusterProfiler” (*21*). GO terms such as extracellular matrix organization, collagen fibril organization, cell-substrate adhesion, epithelial cell proliferation, transmembrane receptor protein serine/threonine kinase signaling pathway, response to TGF-beta, collagen metabolic process, were mainly enriched in the “biological process” category (Figure 3g). The results of the KEGG enrichment analysis revealed that focal adhesion, complement and coagulation cascades, proteoglycans in cancer, ECM-receptor interaction, relaxin signaling pathway, TNF signaling pathway, apoptosis, TGF-beta signaling pathway, and PI3K-AKT signaling pathway, and many other related KEGG biological pathways were both important for the development of carcinoma (Figure 3h). Moreover, we constructed a network of protein-protein interaction (PPI) for all genes in the yellow module by Cytoscape, which consisted of 56 nodes and 366 edges according to the weight of the edge (≥0.07) (Figure 3i).

In summary, we may conclude that the CILP gene is significantly associated with the BM phenotype of BRCA.

CILP Expression was correlated with immune infiltration levels in breast cancer favoring BCBM

The composition of significant tumor-infiltrating immune cells was assessed using the CIBERSORT algorithm for the BRCA and BCBM tissues. For the merged GEO dataset, the heatmap (Figure 4a) and the histogram (Figure 4b) indicated that plasma cells, macrophages M2, T cells CD8, memory B cells, and follicular helper T cells accounted for a large proportion of the BCBM immune cell infiltration. While the heatmap (Figure 4c) showed M0 macrophages, macrophages M1, macrophages M2, CD8+ T cells, and resting mast cell accounted for a large proportion of BRCA immune cell infiltration. As shown in Figure 4d, the violin plot suggested that resting mast cells (P=0.010) showed obvious differences in the immune cell fractions between brain metastasis and primary BRCA via the Wilcoxon rank-sum test. Moreover, the

Wilcoxon rank-sum test suggested that memory CD4+ T cells (P<0.001), follicular helper T cells (P<0.001), regulatory T cells (Tregs) (P<0.001), resting NK cells (P=0.001), monocytes (P<0.001), M0 macrophages (P<0.001), M1 macrophages (P<0.001), M2 macrophages (P<0.001), resting mast cells (P=0.003), and activated mast cells (P=0.001) were significantly different among the immune cell fractions between BRCA tissues and adjacent normal tissues (Figure 4e). The level of infiltrating resting mast cells was low in BCBM, and the relationship between resting mast cells and BCBM is worthy of further study. BCBM occurs during stage IV of neoplasms. We further analyzed the association between infiltrating immune cells and pathological staging. We concluded that M0 macrophages (Staging: P=0.007), M2 macrophages (Staging: P=0.018), activated memory CD4+ T cells (Staging: P=0.019), resting mast cells (Staging: P=0.001), monocytes (Staging: P<0.001), and plasma cells (Staging: P=0.007) all showed significant correlations with clinical features (Figures 4f). It was obvious that resting mast cells showed low expression in stage IV of BRCA. Subsequently, we performed Kaplan-Meier survival analysis, and found that only samples enriched in M2 macrophages had a poor prognosis (P=0.005) in BRCA (Figure 4g). Subsequently, we analyzed the relationship between CILP and resting mast cells, and both were involved in BCBM (Figures 3e-f and 4d). Further studies (Figures 4h) showed that the expression of CILP positively correlated with resting mast cells (R=0.33; P<0.001), resting memory CD4+ T cells (R=0.3; P<0.001), resting dendritic cells (R=0.16; P<0.001), gamma delta T cells (R=0.14; P<0.001), and naive B cells (R=0.13; P<0.001), instead negatively correlated with activated dendritic cells (R=-0.21; P<0.001), memory B cells (R=-0.11; P=0.0014), M0 macrophages (R=-0.14; P<0.001), and follicular helper T cells (R=-0.18; P<0.001), suggesting that CILP was associated with the immune microenvironment of BRCA. Of these, resting mast cells involved in brain metastasis had a significantly weak positive correlation with the expression of CILP. In the Timer database, we also found a weak positive association between CILP and mast cells (Figure 4i). Essentially, resting mast cells showed a lower expression of stage IV of BRCA, and the expression of CILP was also found lower in stage IV of BRCA, which is consistent with the positive association between CILP and resting mast cells.

Thus, it is reasonable to speculate that CILP may influence the progression of BRCA favoring BMs through the infiltration of resting mast cells.

GSEA analysis identified the potential pathways involved in BCBM

To identify the potential function of CILP genes in BRCA samples from TCGA database, GSEA was used to identify the potential pathways involved in brain metastasis processes. Nine enriched pathways were detected, including the B cell receptor signaling pathway; Hedgehog signaling pathway; lysosome, ubiquitin mediated proteolysis; DNA double-strand break response; antigen activation of the B cell receptor (BCR) leading to generation of second messengers; CD22-mediated BCR regulation; creation of C4 and C2 activators; and FCERI-mediated MAPK activation; all of which were closely related to tumor occurrence, development, and metastasis (P-value<0.05, FDR<0.25) (Figure 5).

Discussion

CILP, namely cartilage intermediate layer protein 1, plays a crucial role in Lumbar disc disease (LDD) through negative regulation of TGF-beta signaling (*30*). A recent study reported that CILP seemed to be a promising candidate as a biomarker for cardiac fibrosis (*31*). However, to the best of our knowledge, there have been no documented reports of a possible relationship between CILP expression and BCBM.

In our study, using WGCNA we found CILP was the hub gene significantly associated with BCBM (Fig3 e, f). Further analysis showed it might be a late event in BRCA and had obvious clinical significance (Fig1 a-b, d-e). Moreover, it was differentially expressed in TNBC and HER2+ BRCAs compared to normal and luminal type BRCA (Figure 1c), which was consist with the fact that these two subtypes were more prone to the brain metastasis of BRCA. Hence, it is reasonable to assume that it may represent a new biomarker indicative of BCBM.

Previous studies have suggested that mast cells, termed tumor-associated mast cells, infiltrate a variety of hematological and solid tumors, such as stomach, thyroid, melanoma, pancreas, prostate, and BRCA (*32-36*). Mast cells have been reported as having both antitumorigenic and pro-tumorigenic effects in cancers (*32*). To date, most studies have supported signs of mast cell CELL infiltration as a marker for favorable prognosis in BRCA, suggesting that mast cells might contribute to anti-tumoral functions of BRCA (*37, 38*). Rovere et al. documented that there is evidence of mast cell accumulation around tumors in highly hormone-receptive BRCAs, which seemed to oppose tumor activity via their cytolytic activities (*39*). It has also been demonstrated that both activated and resting mast cells and basophils were able to increase the proliferation and survival of naive and activated B cells, and promoted their differentiation into antibody-producing cells (*40*). Likewise, our study suggested that resting mast cells had lower expression levels in BCBM than in primary BRCA, and were still lower than levels expressed in BRCA than in adjacent normal tissues (Fig4 d-e). Subsequently, activated mast cells were significantly upregulated in BRCA tissues compared with adjacent normal tissues (Fig4 e), whereas there was no significant difference between primary BRCA and BCBMs (Fig4 d). In the Timer database, we found CILP expression was positively correlated with mast cells (Fig4 i). In our study, we found a positive correlation between CILP and mast cells resting (Fig4 h). Consequently, we reasonably suspect that with the development of BRCA, CILP expression becomes lower, impairing the anti-tumor function of mast cells, and thus prompting tumor invasion and metastases.

In our study, nine potential pathways were enriched in the BCBM (Fig5). Among these, the activation of B cell-related signaling pathways including the B cell receptor signaling pathway, antigen-activated B cell receptor leading to generation of second messengers, and CD22 mediated BCR regulation may play an important role in the occurrence and development of B cell-mediated tumors. The Hedgehog signaling pathway is vital for BRCA progression and metastasis (*41*). The pathways involving lysosome- and ubiquitin-mediated proteolysis can mediate the degradation of target proteins. The FCERI-mediated MAPK activation cascade can regulate mast cell activity (*42*). In combination with the study above, we hypothesis that CILP may exert an anti-tumoral role by mediating the activity of mast cells through the MAPK signaling pathway.

Our study has several limitations. First, only two datasets were included in the study which may lead to research bias for lacking of adequate data; Second, because of the limitation of online databases, we did not carry out additional omics analysis such as proteomics to strengthen our results; Third, no external datasets were used to verify the conclusion, and if there was any that would be more supportive and practical. Therefore, further research is needed to explore whether CILP can be exploited as a biomarker specific to BCBM and its potential detailed mechanism.

Conclusions

our study showed that CILP may act as an anti-tumor mediator by interacting with mast cells through the MAPK signaling pathway in BCBM and thus, has great clinical transformational significance.

Abbreviations

CILP: ; BRCA : Breast cancer; BMs: brain metastases; BCBM: breast cancer brain metastases; GO: gene ontology; KEGG : Kyoto encyclopedia of genes and genomes ; CILP: the cartilage intermediate layer protein; MAPK: mitogen-activated protein kinase; HER2+: HER2-positive; TNBC: triple-negative BRCA; CNS: the central nervous system; Gabra3: GABAA receptor alpha3; GEO: Gene Expression Omnibus; FPKM: The Fragments Per Kilobase Million; TCGA: The Cancer Genome Atlas; RNA-Seq: RNA sequencing; DEGs: Differentially expressed genes; FC: Fold Change; WGCNA: the weighted gene co-expression network analysis; TOM: topological overlap matrix; MEs: module eigengenes; GS: Gene significance; HR: The hazard ratio; CI: confidence intervals; SNP: Single Nucleotide Polymorphisms; GSEA: gene set enrichment analysis; FDR: false discovery rate; TGF-beta: transforming growth factor-beta; Tregs: regulatory T cells; BCR: the B cell receptor; LDD: Lumbar disc disease.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study are available from Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), LinkedOmics database, UALCAN, Kaplan-Meier plotter and TIMER. The materials and figures supporting the conclusions of this article are included within the article.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This research was supported by the Taishan Scholars Program (NO. TSQN20161070), Shandong Provincial Key Research and Development Program (2016GGX101035 and 2019GGX101004).

Authors' contributions

Xiaomei Li conceived and designed the study. Xiaolin Sun analyzed GEO and TCGA data, and wrote the manuscript. Xingguo Zhou assisted with data analysis. Zhe Yang and Yingchao Liu helped to revise the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgements

We would like to thank GEO and TCGA project.

References

1. A. Pellerino *et al.*, Management of Brain and Leptomeningeal Metastases from Breast Cancer. *International Journal of Molecular Sciences***21**, (2020).
2. I. Witzel, L. Oliveira-Ferrer, K. Pantel, V. Mueller, H. Wikman, Breast cancer brain metastases: biology and new clinical perspectives. *Breast Cancer Research***18**, (2016).
3. L. Nayak, E. Q. Lee, P. Y. Wen, Epidemiology of Brain Metastases. *Current Oncology Reports***14**, 48-54 (2012).
4. E. Tabouret *et al.*, Recent Trends in Epidemiology of Brain Metastases: An Overview. *Anticancer Research***32**, 4655-4662 (2012).
5. C. C. O'Sullivan, N. N. Davarpanah, J. Abraham, S. E. Bates, Current challenges in the management of breast cancer brain metastases. *Semin. Oncol.***44**, 85-100 (2017).
6. S. Dawood *et al.*, Survival among women with triple receptor-negative breast cancer and brain metastases. *Annals of Oncology***20**, 621-627 (2009).
7. N. U. Lin *et al.*, Sites of Distant Recurrence and Clinical Outcomes in Patients With Metastatic Triple-negative Breast Cancer High Incidence of Central Nervous System Metastases. *Cancer***113**, 2638-2645 (2008).
8. N. Kotecki, F. Lefranc, D. Devriendt, A. Awada, Therapy of breast cancer brain metastases: challenges, emerging treatments and perspectives. *Ther Adv Med Oncol***10**, 1758835918780312 (2018).
9. R. M. S. M. Pedrosa, D. A. Mustafa, R. Soffietti, J. M. Kros, Breast cancer brain metastasis: molecular mechanisms and directions for treatment. *Neuro-Oncology***20**, 1439-1449 (2018).
10. P. D. Bos *et al.*, Genes that mediate breast cancer metastasis to the brain. *Nature***459**, 1005-U1137 (2009).

11. W. Peng, R. Zhu, S. Zhou, P. Mirzaei, Y. Mechref, Integrated Transcriptomics, Proteomics, and Glycomics Reveals the Association between Upregulation of Sialylated N-glycans/Integrin and Breast Cancer Brain Metastasis. *Scientific Reports***9**, (2019).
12. K. Gumireddy *et al.*, The mRNA-edited form of GABRA3 suppresses GABRA3-mediated Akt activation and breast cancer metastasis. *Nature Communications***7**, (2016).
13. G. Chang *et al.*, YTHDF3 Induces the Translation of m6A-Enriched Gene Transcripts to Promote Breast Cancer Brain Metastasis. *Cancer cell***38**, 857-871.e857 (2020).
14. J. Kim, J.-S. Bae, Tumor-Associated Macrophages and Neutrophils in Tumor Microenvironment. *Mediators of Inflammation***2016**, (2016).
15. R. Duchnowska *et al.*, Immune response in breast cancer brain metastases and their microenvironment: the role of the PD-1/PD-L axis. *Breast Cancer Research***18**, (2016).
16. A. S. Berghoff, H. Lassmann, M. Preusser, R. Hoeftberger, Characterization of the inflammatory response to solid cancer metastases in the human brain. *Clinical & Experimental Metastasis***30**, 69-81 (2013).
17. M. J. Sambade *et al.*, Examination and prognostic implications of the unique microenvironment of breast cancer brain metastases. *Breast Cancer Research and Treatment***176**, 321-328 (2019).
18. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics***9**, (2008).
19. J. A. Botia *et al.*, An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *Bmc Systems Biology***11**, (2017).
20. B. Rahmani *et al.*, Recursive Indirect-Paths Modularity (RIP-M) for Detecting Community Structure in RNA-Seq Co-expression Networks. *Frontiers in Genetics***7**, (2016).
21. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics-a Journal of Integrative Biology***16**, 284-287 (2012).
22. S. V. Vasaikar, P. Straub, J. Wang, B. Zhang, LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Research***46**, D956-D963 (2018).
23. D. S. Chandrashekar *et al.*, UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia***19**, 649-658 (2017).
24. T. Li *et al.*, TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Research***48**, W509-W514 (2020).
25. T. Iwamoto *et al.*, Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. *Scientific Reports***9**, (2019).
26. H.-J. Schulten *et al.*, Comprehensive molecular biomarker identification in breast cancer brain metastases. *Journal of Translational Medicine***15**, (2017).
27. P. Beckerman *et al.*, Human Kidney Tubule-Specific Gene Expression Based Dissection of Chronic Kidney Disease Traits. *Ebiomedicine***24**, 267-276 (2017).

28. J. Tang *et al.*, Prognostic Genes of Breast Cancer Identified by Gene Co-expression Network Analysis. *Frontiers in Oncology***8**, (2018).
29. E. Y. Rosen *et al.*, Functional Genomic Analyses Identify Pathways Dysregulated by Progranulin Deficiency, Implicating Wnt Signaling. *Neuron***71**, 1030-1042 (2011).
30. S. Seki *et al.*, A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. *Nature Genetics***37**, 607-612 (2005).
31. S. GroSs, T. Thum, TGF-beta Inhibitor CILP as a Novel Biomarker for Cardiac Fibrosis. *JACC. Basic to translational science***5**, 444-446 (2020).
32. G. Varricchi *et al.*, Are Mast Cells MASTers in Cancer? *Frontiers in Immunology***8**, (2017).
33. A. Johansson *et al.*, Mast Cells Are Novel Independent Prognostic Markers in Prostate Cancer and Represent a Target for Therapy. *American Journal of Pathology***177**, 1031-1041 (2010).
34. E. Dundar *et al.*, The significance and relationship between mast cells and tumour angiogenesis in non-small cell lung carcinoma. *Journal of International Medical Research***36**, 88-95 (2008).
35. B. Tuna, K. Yorukoglu, M. Unlu, M. U. Mungan, Z. Kirkali, Association of mast cells with microvessel density in renal cell carcinomas. *European Urology***50**, 530-534 (2006).
36. S. Ch'ng, R. A. Wallis, L. Yuan, P. F. Davis, S. T. Tan, Mast cells and cutaneous malignancies. *Modern Pathology***19**, 149-159 (2006).
37. L. Gou, G. G.-L. Yue, P. T. Puno, C. B.-S. Lau, A review on the relationship of mast cells and macrophages in breast cancer - Can herbs or natural products facilitate their anti-tumor effects? *Pharmacological research*, 105321-105321 (2020).
38. A. Aponte-Lopez, J. Enciso, S. Munoz-Cruz, E. M. Fuentes-Panana, An In Vitro Model of Mast Cell Recruitment and Activation by Breast Cancer Cells Supports Anti-Tumoral Responses. *International Journal of Molecular Sciences***21**, (2020).
39. F. Della Rovere *et al.*, Mast cells in invasive ductal breast cancer: Different behavior in high and minimum hormone-receptive cancers. *Anticancer Research***27**, 2465-2471 (2007).
40. S. Merluzzi *et al.*, Mast cells, basophils and B cell connection network. *Molecular Immunology***63**, 94-103 (2015).
41. N. A. Riobo-Del Galdo, A. Lara Montero, E. V. Wertheimer, Role of Hedgehog Signaling in Breast Cancer: Pathogenesis and Therapeutics. *Cells***8**, (2019).
42. Y. Shaik, A. Caraffa, G. Ronconi, G. Lessiani, P. Conti, Impact of polyphenols on mast cells with special emphasis on the effect of quercetin and luteolin. *Central European Journal of Immunology***43**, 476-481 (2018).

Figures

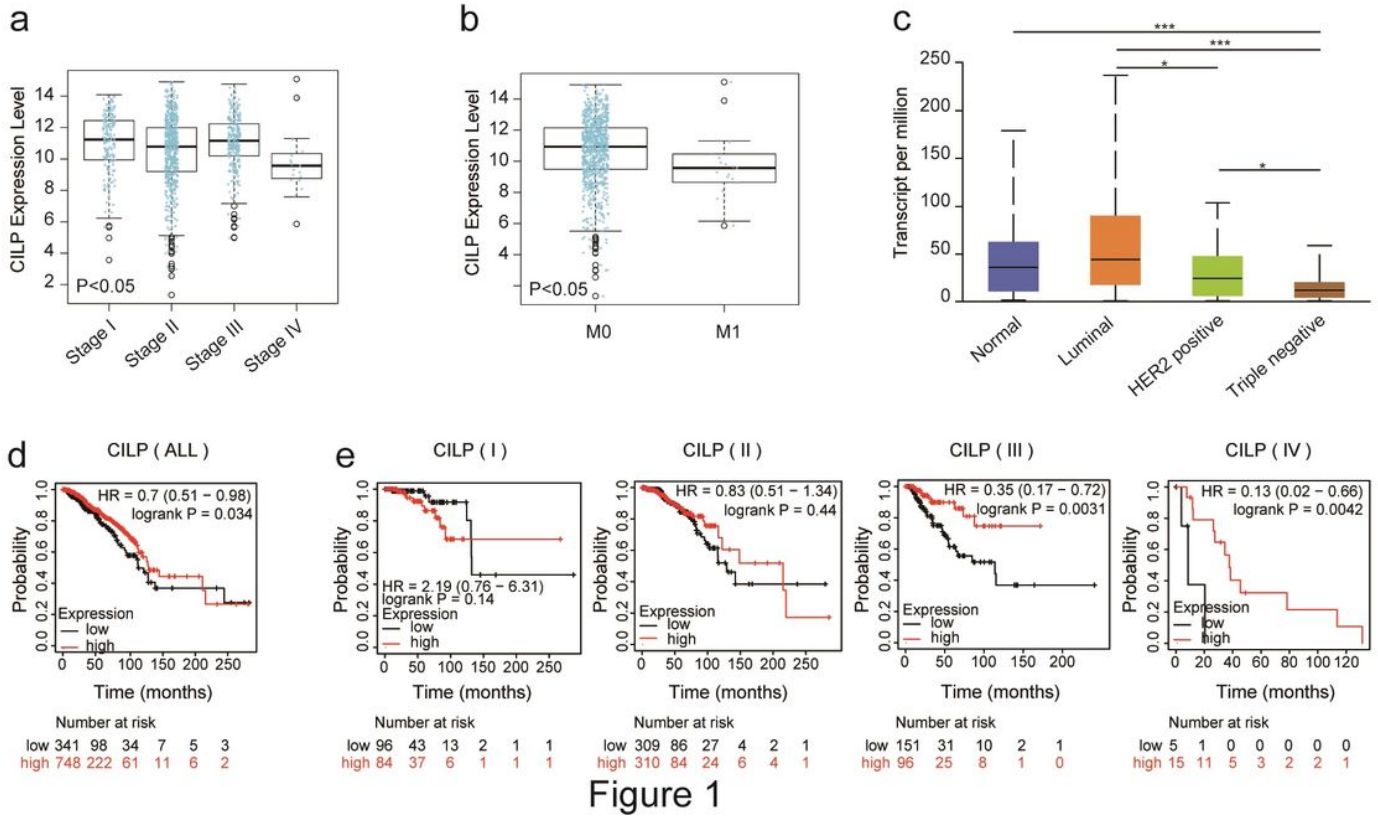


Figure 1

Figure 1

lower CILP expression is correlated with decreased overall survival in advanced breast cancer patients (a and b) LinkedOmics database showing the RNA expression levels of CILP between different pathologic stages (a) and M stage (b) of BRCA patients. n=1071, Kruskal-Wallis Test (a); n=931, Wilcox Test (b). (c) UALCAN analysis showing the association for CILP transcripts of intrinsic subclasses of BRCA. Student's t test, *p<0.05, ***p<0.001. (d) Results of Kaplan-Meier survival analyses for BRCA patients with primary tumors expressing high or low levels of CILP mRNA as assessed using the Kaplan Meier plotter database. (e) Kaplan-Meier survival analysis of CILP mRNA expression of BRCA patients with differential pathologic stages (I, II, III and IV) determined by the Kaplan Meier plotter database. The P-value of data (d) and (e) calculated by the log rank test.

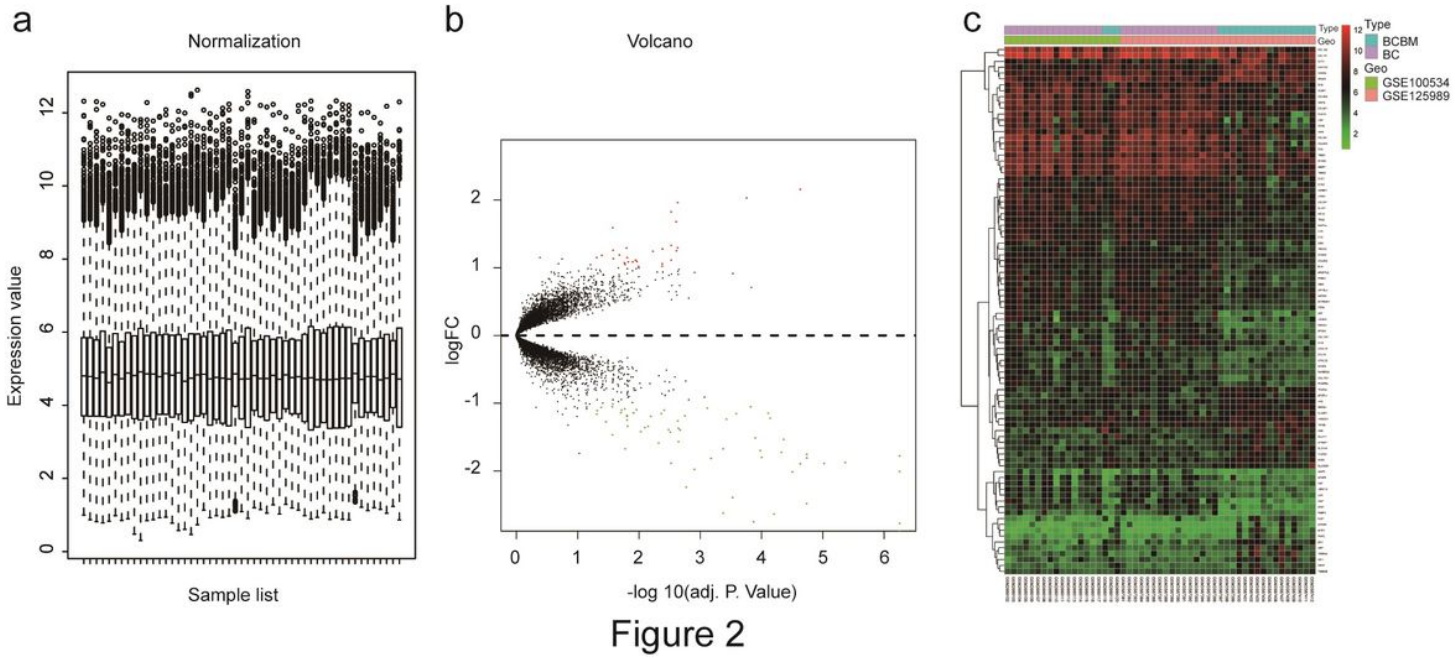


Figure 2

Figure 2

Identification of the markedly different expressed Genes between breast cancer tissue and brain metastases tissue from breast cancer (a) Box-plot of gene expression of the merged GSE125989 and GSE100534 datasets after normalization by “sva” package. (b) The Volcano plot for differently expressed genes in merged datasets. Red/green circles classify the upregulated/downregulated genes according to the criteria: log2FC ≥1 and adjusted P-value <0.05. (c) Heatmap of 90 DEGs (28 up- and 62 downregulated genes) in merged datasets. Green, black, and red respectively represent the differences in lower, medium, and high expression levels for different genes. Type: Blue and purple respectively represents BRCA brain metastases tissues (BCBM) and BC tissues.

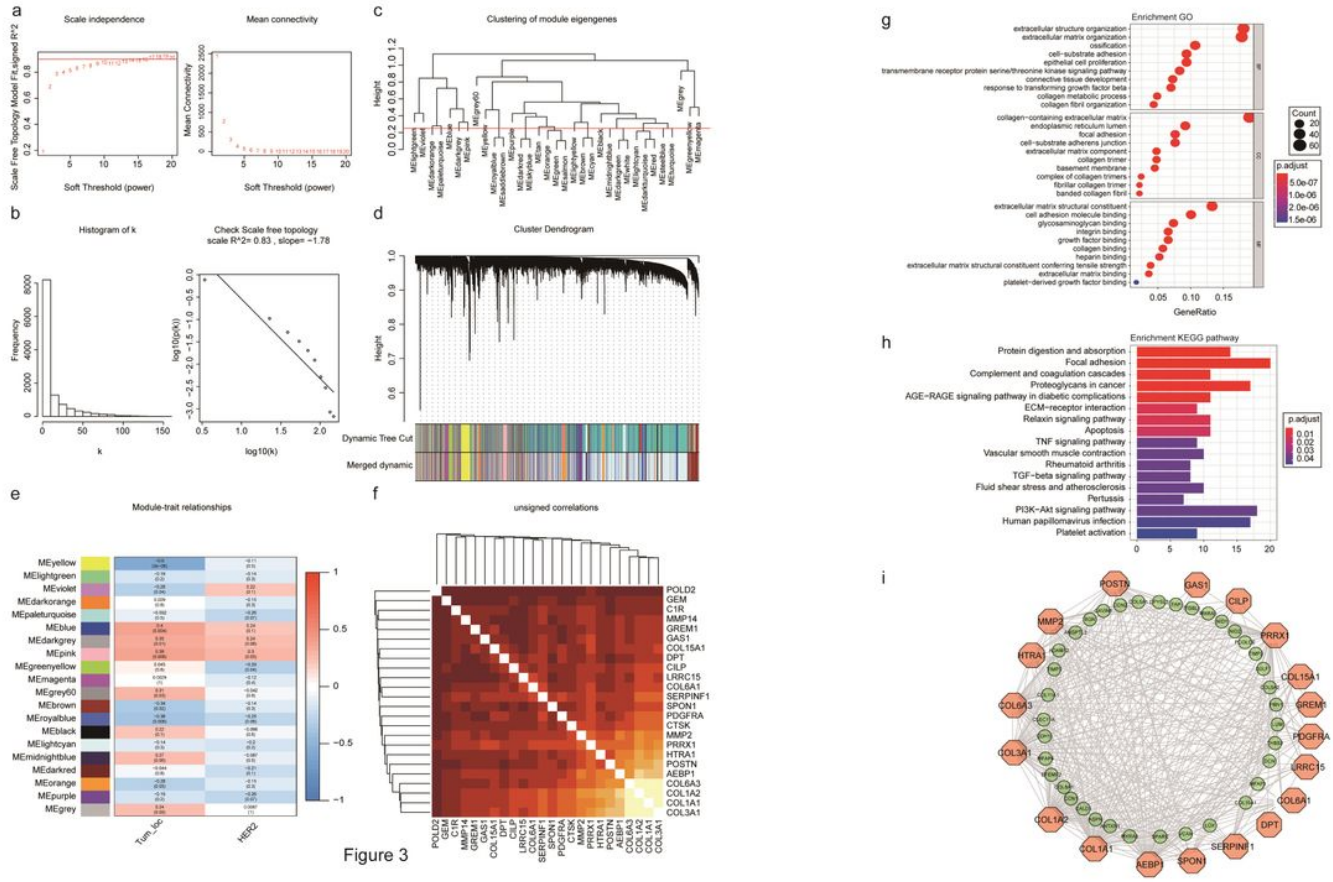
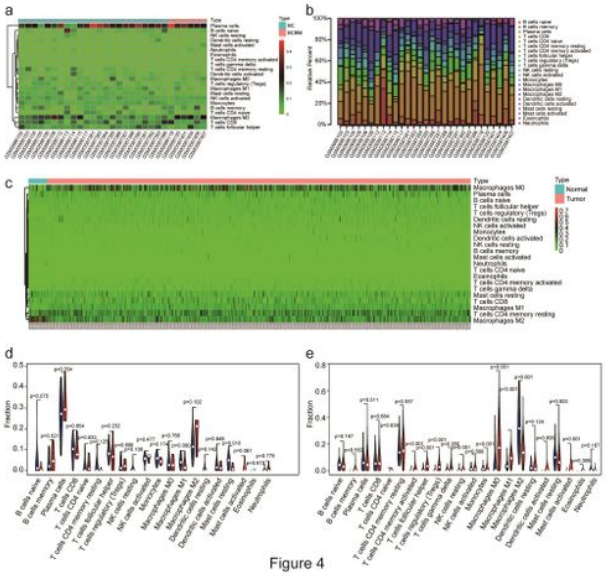


Figure 3

Figure 3

CILP is the hub gene significantly associated with BCBM analyzed by construction of weighted co-expression network (a) Scale-free fit index and the mean connectivity analysis for various soft-thresholding powers. The panel on the left shows the relationship between the soft-threshold and the scale-free R2. The right panel shows the relationship between soft-threshold and average connectivity. (b) The soft-thresholding power (β=8) in the WGCNA was determined based on a scale-free R2 (R2=0.83). (c) The cluster dendrogram of module eigengenes. Implementing clustering of module eigengenes by merging highly similar modules with the dissimilarity of <0.25. (d) A dendrogram of the all expressed genes clustered according to the different metrics. Each branch in the figure represents one gene, and each color represents one co-expression module. (e) Heatmap showing the correlation between clinical traits and module eigengenes. The yellow module containing 399 genes shows the highest negative correlation with the BCBM phenotype. Each cell contains the correlation and P-value. The correlation coefficient of each cell indicates the correlation between gene module and clinical traits, which are reduced in size from red to blue. Tum_loc represents the location of the tumor including BRCA and BCBM tissues. HER2 represents positive or negative status. (f) A heatmap showing 24 candidate hub genes of the yellow module in (e). (g) Enriched GO terms in the “cellular component (CC)”, “Biological process (BP)”, and Molecular function (MF) categories for yellow module in (e). Different sizes denote the number

of genes, while different colors denote different significances. (h) KEGG pathway was analyzed from all genes of yellow module in (e). The length of the column indicates the enrichment score, while the colors represent enrichment significance. (i) The network of protein-protein interactions (PPI) for all genes in the yellow module from (e) by Cytoscape consisting of 56 nodes and 366 edges according to the weight of edge (≥0.07).



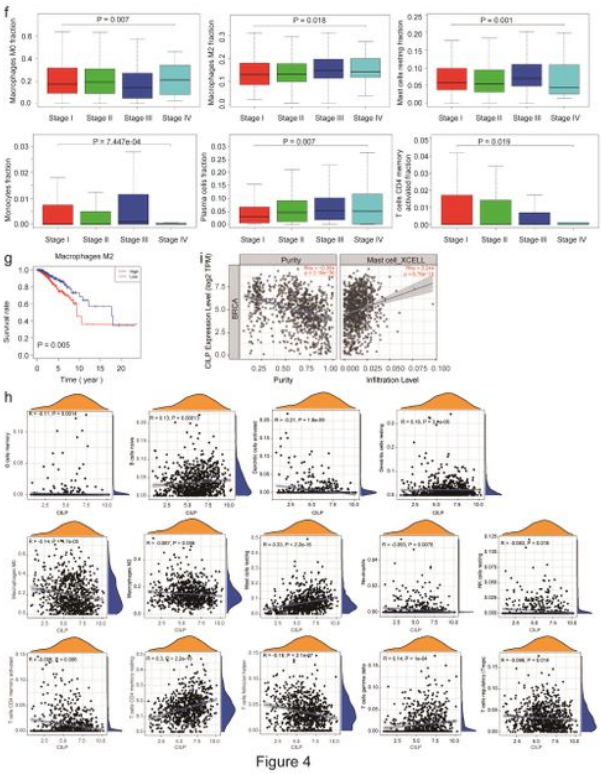


Figure 4

CILP expression is correlated with immune infiltration level in breast cancer favoring BCBM (a, b and c) CIBERSORT algorithm assessing the composition of 22 immune cells between BCBMs and primary tumors (a, b), and between breast cancer (BRCA) tissues and adjacent normal tissues (c). (d and e) The violin plot shows prominent tumor-infiltrating immune cells related to the brain metastasis of BRCA (d) and associated with occurrence of BRCA (e). Red represents the BCBM group while blue represents the BRCA group (d). Red represents the BRCA group while blue represents adjacent normal tissue group (e). The P-value of data (d) and (e) are calculated by Wilcoxon rank-sum test. (a), (b) and (d) Merged GSE cohort. (c) and (e) summarize TCGA cohort. (f) The box plots show the clinical stages of M0 macrophages, M2 macrophages, activated memory CD4+ T cells, resting mast cells, monocytes, and plasma cells. n=800, Wilcoxon rank-sum test. (g) Kaplan-Meier curves for BRCA patients enriched with M2 macrophages by “survival” package using RStudio. n=787. (h) The co-expression results of tumor infiltrating immune cells and CILP mRNA expression levels (log2) were analyzed from TCGA dataset.

n=821, Spearman's Correlation. (i) Correlation between CILP expression and mast cell infiltration of BRCA was analyzed in Timer database. Data (f)-(h): TCGA cohort.

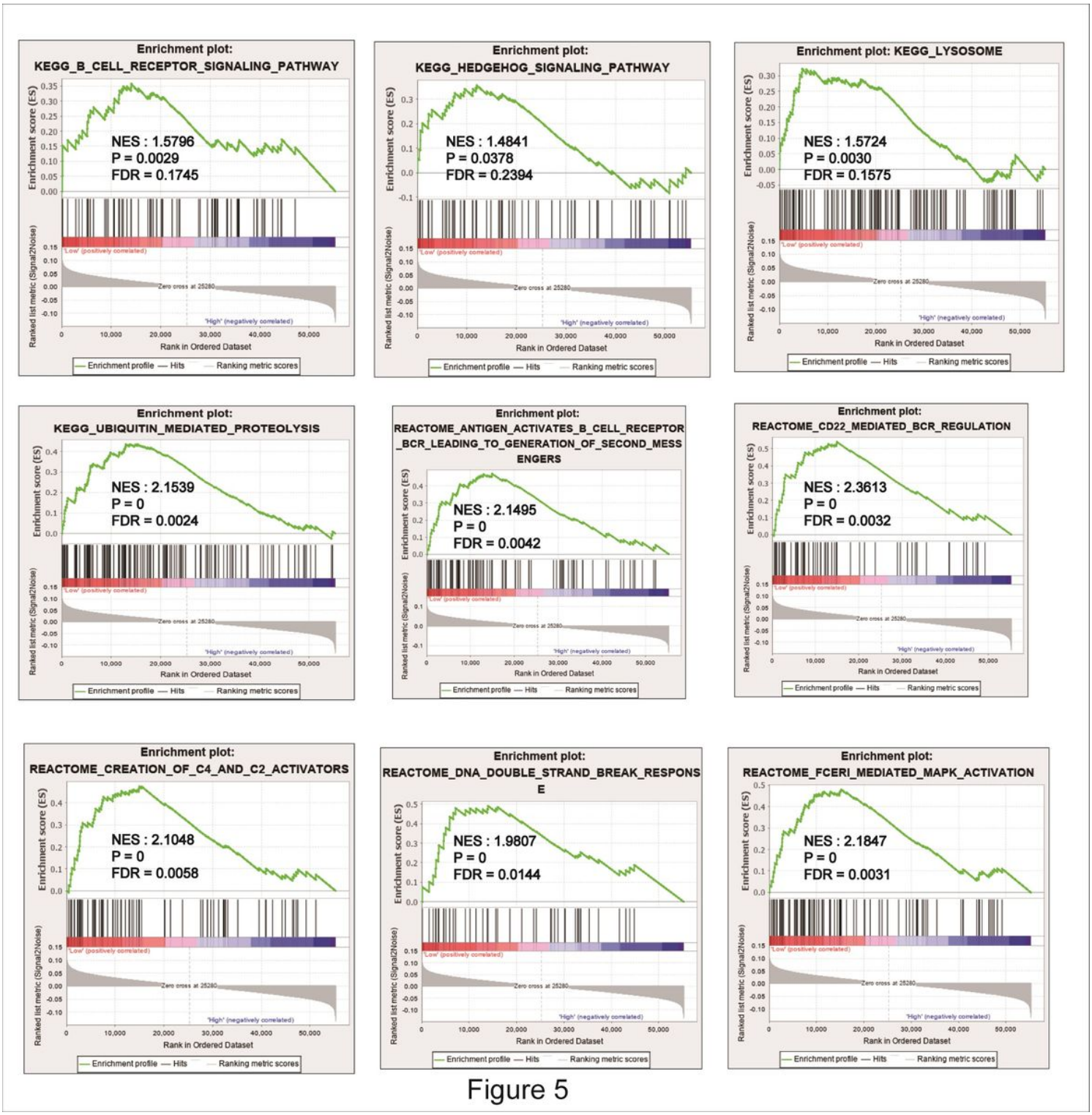


Figure 5

Figure 5

Gene set enrichment analysis using TCGA cohort Listed are only the nine most common functional pathways of KEGG and REACTOME pathway gene sets enriched in BRCA samples with low expression of the CILP gene. CILP gene expression was divided into two groups of high and low expression through the surv_cutpoint function of "survminer" package. False discovery rate (FDR), nominal P-value, and the

Normalized enrichment score (NES) are displayed in each plot. Nominal P-value<0.05, false discovery rate (FDR)<0.25.