Identification of immune-related therapeutically relevant biomarkers in breast cancer and breast cancer stem cells by transcription-wide analysis: a clinical prospective study

Linbang Wang
Chongqing Medical University First Affiliated Hospital
Yuanyuan Wang
Chongqing Medical University First Affiliated Hospital
Jingkun Liu
Honghui Hospital, Xi'an Jiaotong University
Jiaojiao Tai
Honghui Hospital, Xi'an Jiaotong University
Xuedong Yin
Chongqing Medical University First Affiliated Hospital
Wei Liu (liuweifreeman@hotmail.com)
Chongqing Medical University First Affiliated Hospital

Research

Keywords: breast cancer, cancer stem cell, tumor immune infiltration, PIMREG, MTFR2

Posted Date: April 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-21080/v1

Identification of immune-related therapeutically relevant biomarkers in breast cancer and breast cancer stem cells by transcriptome-wide analysis: a clinical prospective study

Authors: Linbang Wang[1+], Yuanyuan Wang[2+], Jingkun Liu[3], Jiaojiao Tai [3], Xuedong Yin[2], Wei Liu[1*]

Author details

[1]Department of Orthopedic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China.

[2]Department of Endocrine and Breast Surgery, The First Affliated Hospital of Chongqing Medical University, Chongqing 400016, China.

[3]Department of Orthopedics, Honghui Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710054, China

[+]These authors contributed equally to this work.

* Corresponding Author

Corresponding Author: Wei Liu, E-mail: liuweifreeman@hotmail.com.

Abstract

Background: Cancer stem cells (CSCs) represent a subset of tumor cells that are responsible for recurrence and metastasis of tumors. These cells are resistant to radiotherapy and chemotherapy. Immunotherapeutic strategies that target CSCs specifically have provided initial results; however, the mechanism of action of these strategies is unclear.

Methods: The data were requested from The Cancer Genome Atlas and Genotype-Tissue Expression, followed with the survival analysis and weighted gene co-expression network analysis to detect survival- and stemness-related genes. Patients were divided into three groups based on their immune status by applying ssGSEA with proven dependability by ESTIMATE analysis. The filtered key genes were analyzed using oncomine, qRT-PCR, and functional analysis.

Results: Patients in a group with a higher stemness and a lower immune infiltration showed a worse overall survival probability, stemness and immune infiltration characteristics of breast cancer progressed in a non-linear fashion. Thirteen key genes related to stemness and immunity were identified and the functional analysis indicated their crucial roles in cell proliferation and immune escape strategies. The qRT-PCR results showed that the expression of PIMREG and MTFR2 differed in different stages of patients.

Conclusions: Our study revealed a promising potential for CSC-target immunotherapy

in the early stage of cancer and a probable value for PIMREG and MTFR2 as biomarkers and targets for immunotherapy.

Trial registration: This study protocol registered with Chinese Clinical Trial Registry (http://www.chictr.org.cn/showproj.aspx?proj=19710; Date of registration: 25/09/2017; Registration number: ChiCTR-PDN-17012784) and was approved by the ethical committee of Affiliated Hospital of Chongqing Medical University (approval number: 2020-119).

Keywords: breast cancer, cancer stem cell, tumor immune infiltration, PIMREG, MTFR2.

Background

Breast cancer (BC) has the highest incidence rate and mortality rate among female malignant tumors, which impacts women's health significantly [1]. It is considered to be heterogeneous depending on molecular subtype and on different stages of cancer progression [2]. This heterogeneity poses challenges during treatment, even through various treatment strategies have been developed based on different pathological types [3], especially for the triple-negative BC [4]. Thus, further explorations are needed to identify new markers for guiding individualized treatment.

Growing evidence has established the presence of a subpopulation of cancer cells with stem-like properties in most human malignancies, frequently referred to as ‘cancer stem cells’ (CSCs), which possess the long-term ability to initiate and repopulate tumors

[5, 6]. Diverse mechanisms by which CSCs manage to survive through various strategies including tumor initiation, metastatic reactivation, oncogene- and immune-targeted therapy resistance have been unraveled [4, 7].

Immuno-resistance is one of the main features of tumors that helps them escape immunosurveillance and evade eradication by resisting immunosuppression [8, 9]. A myriad of strategies have been discovered in the tumor cells that allow them to circumvent the immune attack, including genetic and epigenetic alterations in the genome of tumor cells that reduce immune recognition and promote protective microenvironment [5, 10].

Evidence has emerged that CSCs have a potential role in regulating their immune characteristics [9, 10], while the molecular mechanism are unclear. In this study, we focused on CSCs in BC. To this end, differentially expressed genes (DEGs) were screened using The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) databases, mRNAsi index and WGCNA analysis were used in turns to profile the association between stemness of tumor and clinical characters and to identify genes that are related to stemness. Next, we applied immune infiltration analysis to the filtered immune-related genes from stemness and survival-related genes. Finally, the qRT-PCR analysis was used to verify that the expression of *PIMREG* and *MTFR2* was aberrant and diverse in different clinical stages.

Methods

Design and data processing of this study

The schematic flowchart of the present study is shown in Fig. 1. Gene expression and clinical information of patients and healthy individuals was collected from the common database by applying strict filters. Then the computational biology tools were applied, to evaluate stemness and immune characteristics, and biomarkers for patients with BC were revealed. The expression of key genes was evaluated by qRT-PCR.

Data acquisition

The comprehensive data of BC including 1104 tissues of patients and 113 non-tumor tissues was downloaded from the TCGA database (https://portal.gdc.cancer.gov September 26, 2019). Due to the lack of corresponding samples from non-tumor tissues, we utilized GTEx project from where 80 samples of healthy individuals were obtained. RNA-seq results of healthy tissue samples and cancer samples were combined into a matrix file using a merge script in the Perl language (http://www.perl.org/). The mRNAsi were acquired from Tathiane M. Malta et al [11]. The 29 marker gene sets for immune cell types and signaling pathway activation were obtained from Bao X et al and Liu Z et al [12, 13].

Differentially Expressed Genes

The “edgeR” R package was used to screen DEGs of normal breast and cancer samples with the following selection criteria: FDR < 0.05, and |log2 FC| > 1. The values of genes calculated by limma R package, and genes with expression of < 0.5 were deleted. The

heatmap and Kaplan-Meier (K-M) curves were drawn.

WGCNA and Module merge

Co-expression network was established using the WGCNA R package based on the DEGs. First, RNA-seq data were matched with corresponding mRNAsi and filtered by hierarchical cluster analysis to detect outliers, the Pearson correlation matrix was constructed by correlating coefficient of genes and then transforming them into weighted adjacency matrix using the power function: $apq=|cpq|\beta$ (cpq=Pearson's correlation between gene p and gene q, apq=adjacency between gene p and gene q, and β = soft threshold). The best soft threshold (soft threshold = 4) which was selected according to scale independence and mean connectivity for achieving a scale-free co-expression network. The weighted adjacency matrix was then transformed into the topological overlap matrix, DEGs were allocated into different modules for average linkage hierarchical clustering and similar genes were grouped into one module, module dendrogram was drawn with the minimum size (genome), modules with similar heights (cutoff < 0.3) were merged. As a correlation value between genes and sample traits, Gene significance (GS) was calculated based on statistical significance, which was determined using the relevant p-values in the linear regression between gene expression and clinical phenotypes (mRNAsi and EREG-mRNAsi). The significant modules related to mRNAsi were selected according to Module significance which was defined as the average GS within the module and revealed the correlation between the module and sample traits.

Identification of mRNAsi- and survival-associated genes

Module membership (MM) was defined as the correlation between the module's own genes and gene expression profiles. The key genes associated with mRNAsi were screened by GS and MM defined as cor. gene MM > 0.8 and cor. gene GS > 0.5. The interactive genes between survival-related DEGs and mRNAsi-related genes were finally identified and depicted using venn charts.

Immune infiltration grouping of BC patients

BC gene set was prepared as a gmt file for further quantitative measurements of the immune activation status. The RNA-seq data of individual cancer samples was transformed into enrichment scores of each immune-related term by Single-Sample Gene Set Enrichment Analysis (ssGSEA) in the R package gsva. Tumors samples with qualitatively different enrichment scores were divided into low, median, high infiltration clusters by using hierarchical clustering analysis in sparcl R package. Results were presented as a color dendrogram and heatmap.

Tumor microenvironment (TME) analysis of BC

The TME scores were calculated based on BC gene set and tumor purity was predicted by using estimate R package. The assessment of TME were divided into four clusters (stromal score, immune score, estimate score and tumor purity) [14, 15]. The heatmap and the violin plot were conducted to further explore the relationship between immune groups and TME.

Identification of key genes

All interactive genes associated with survival time and mRNAsi of Venn chart were enrolled for correlation analysis with three infiltration clusters. Genes with no significant difference between immune clusters and mRNAsi were exluded ($P \geq 0.05$). Correlation analysis and survival analysis of these genes were conduced. The results were shown in box plots, heatmaps, and K-M curves. Oncomine platform (http://www.oncomine.org) was used to inspect differential expression of key genes between BC and healthy tissues and between different tumors.

Functional analysis of key genes

Protein-protein interaction (PPI) network analysis was applied to show the relationship between different proteins of these key genes by using String(https://string-db.org/). The functional enrichment analysis was also conducted using clusterProfiler R package to investigate the biological functions of key genes. Gene ontology (GO) functional annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were used in the study with the threshold values: $P < 0.05$, and FDR < 0.05. The results were shown in PPI network graph and dot plots. The relation of co-expressed key genes and clinical characters are analyzed by houyunhuang/ggcor R package.

Patients

BC samples and corresponding para-cancerous tissues were collected from patients

who underwent modified radical mastectomy from January 2020 to March 2020 from the department of endocrine and breast surgery in the First Affiliated Hospital of Chongqing Medical University. Tissues were excised and immediately transferred into liquid nitrogen. All patients were informed and written informed consent were provided. The study was conducted according to the clinical practice guidelines of the International Conference on Harmonization and the Declaration of Helsinki. All patients were diagnosed with triple-negative BC with no evidence of distant metastasis, only patients with Tumor Node Metastasis (TNM) stages II-III were included in the study, certified by two pathologists. BC was divided into two subtypes based on their TNM stages. Requisite clinical data were acquired from hospital records and pathology reports. This study protocol was approved by the ethical committee of Affiliated Hospital of Chongqing Medical University. This study protocol registered with Chinese Clinical Trial Registry (http://www.chictr.org.cn/showproj.aspx?proj=19710; Date of registration: 25/09/2017; Registration number: ChiCTR-PDN-17012784) and was approved by the ethical committee of Affiliated Hospital of Chongqing Medical University (approval number: 2020-119).

Total RNA extraction

Total RNA from BC tissues was extracted using an UNIQ-10 column Total RNA Extractio Kit (Sangon Biotech). The RNA concentration and purity were assessed using a SMA4000 microspectrophotometer (Merinton Instrument, Inc) and by RNA electrophoresis with DYY-6C electrophoresis apparatus (Liuyi. Beijing).

Reverse transcription and qRT-PCR quantification

RNA with concentration ranging from 91.84 ng/μl to 1325.94 ng/μl from human BC tumor and para-carcinoma tissues and were reversed-transcribed using a RR047A cDNA synthesis kit (TaKaRa, China). Quantitative PCR was performed for CEP55, MTFR2, PIMREG using a 2X SG Fast qPCR Master Mix (High Rox, B639273, BBI) with the Step One Plus fluorescence quantitative PCR instrument (ABI, Foster, CA, USA), GAPDH is used for internal control gene. The primers designed by Primer Premier 5.0 are listed as below: MTFR2-F: CTCCTCCACCACTTCCTCCTCAG; MTFR2-R: CGCTCAATTGCACGAAGCTTAACC; PIMREG-F: GAGTGCTTTGGGTGCCGTGTC; PIMREG-R: CCGCCTTGATCGCCGTAATGG; CEP55-F: GTGGGGATCGAAGCCTAGTA; CEP55-R: TCATACACGAGCCACTGCTG.

Statistical analysis

The K-M analysis was applied to show the survival difference of mRNAsi groups and immune clusters. Wilcox analysis was utilized to analyze the difference of mRNAsi between normal and cancer groups. Krudkal-Wallis H test was conducted to reveal the relationship between mRNAsi, infiltration clusters, TME scores, tumor purity and clinical information like TNM stage ($P < 0.05$). The beeswarm plots and violin plots were made to visualize this results.

The differences in gene expression between tumor and normal tissues, and para

comparison between subgroups were expressed as mean with standard deviations and compared through Student's paired *t*-test and Student's *t*-test accordingly by applying GraphPad Prism 8, $P < 0.05$ indicated statistically significant difference.

Results

Screening for survival-related DEGs

Samples for which with RNA-seq and comprehensive information was available were enrolled. A total of 2261 different expression genes were screened with 989 up-regulated genes and 1272 down-regulated genes and compared with normal samples. The results were visualized by heat map (Fig. S1), in which 212 survival-related DEGs were filtered (including 89 upregulated genes and 123 down-regulated genes).

Correlation analysis of mRNAsi and clinical characteristics

mRNAsi was revealed by Tathiane M. Malta et al. as an index for evaluating the stemness of tumor cells, thus it is also considered as a quantitative description of CSCs. Wilcox analysis showed that the mRNAsi level in cancer group was significantly higher than that of normal group (Fig. 2a). The K-M curve showed that patients in low mRNAsi group had a longer overall-survival time within 5 year-follow-up (Fig. S2a). As shown in Fig. 2c-d, the score for mRNAsi had a rising tendency in a non-linear manner in the progression of BC, to be noted, the dynamic change of stemness of BC tissue from stage II to stage III were inverse comparing to the overall tendency (stage I to stage IV).

Identification of significant modules and mRNAsi-associated genes

In order to identify biologically significant genes that were associated with BC stemness, a gene co-expression network was conducted by performing WGCNA. DEGs with similar expression patterns were allocated into one module by WGCNA (Fig. 3a), soft threshold in this study was set at 4 in order to ensure a scale-free network (Fig. 3c-d), and nine significant gene modules were obtained for subsequent analysis as shown in Fig. 3b. Next, we explored the correlation between these nine modules and the clinical traits (including mRNAsi and EREG-mRNAsi). The result showed that three gene modules were correlated with mRNAsi (the value of correlation > 0.6), among these, the turquoise module with a positive correlation at 0.83 was most significantly related to mRNAsi, the blue and brown module had significantly negative correlation with mRNAsi (Fig. 3e). Next, 100 key genes were screened from turquoise module with a threshold defined as cor.MM > 0.8 and cor.GS > 0.5 (Fig. 3c). Finally, 18 interactive genes were selected from the intersection of 212 survival-related DEGs as well as 100 mRNAsi-related genes.

Immune phenotype landscape in the TME of BC

The immune cell infiltration status of each sample was assessed by ssGSEA and transformed into enrichment scores based on its transcriptomes according to 29 marker gene sets. All 1104 BC samples were allocated into three hierarchical clusters (low infiltration: 11 patients; intermediate infiltration: 618 patients; and high infiltration: 475 patients) (Fig. 4a). ESTIMATE method was then applied to testify the reliability of this

model, which evaluated the conditions of cellular composition in each sample by multiple indicators, including stromal score, immune score, tumor purity, and an ESTIMATE score that is regarded as sum of the first three scores. The results showed that patients in high infiltration cluster had higher stromal score, higher immune score, lower tumor purity, and higher overall ESTIMATE score (Fig. 4b-e). Survival analysis was also applied to assess the of prognosis value of ESTIMATE scores. As shown in Fig. S2a-e, high immune score, low tumor purity, and high ESTAMATE score had a longer overall-survival time. As shown in Figure S3a-e, stromal score and estimate score showed a downward tendency in a non-linear manner in the tumor progression, while estimate score showed a rising tendency in a non-linear manner in the progression of Lymph Node. To be noted, the change of stomal score from stage II to stage III and ESTAMATE score from T2 to T3 showed inverse correlation compared with the global trend. The Stromal score of N2 reaches the top.

Key genes identification and correspondence functional analysis

Of 18 mRNAsi-related genes that were analyzed for correlation with three immune cell infiltration clusters, 13 key genes were finally enrolled (Fig. S4). There was statistical significance of the expression of ANLN, CCNB2, CEP55, FAM83D, MELK, MTFR2, PIMREG, RAD54L, RAD51, SHCBP1, SKA1, UBE2T, and SKA3 within three immune cell infiltration clusters. As shown in Fig. S5a-r, the expression levels for all 13 key genes were higher in tumor samples than in the normal samples. The K-M curves showed that patients with higher level of expression of these genes had a poor overall

survival time (Fig. S6a-m). The results indicated that these 13 key genes were associated with mRNAsi and immunization as good prognostic factors.

In order to further explore the molecular function of the 13 key genes identified, PPI and enrichment analysis were conducted. As shown in Fig. 5b-c, CEP55 proteins had one of the highest node number (node number is 12). The clinical significance of these co-expressed key genes is shown in Fig. 5, key genes were mainly associated with cancer immune and TME. GO functional enrichment analysis indicated these genes were related to cell proliferation, such as organelle fission, chromosome segregation, spindle, midbody, and chromosomal region (Fig. S7a). The results of KEGG showed the top enriched terms were homologous recombination and Fanconi anemia pathway and indicated a close correlation with DNA damage and repair (Fig. S7b). The result of TIMER showed that most of the key genes had strong correlation with B lymphocyte and dendritic cells' function (Fig. 2e). Through differentially expression analysis of pan-cancer and normal samples by Oncomine, it was found that PIMREG, MTFR2, and CEP55 were overexpressed in BC and also in many other cancers (Fig. 6a).

Validation of key genes expression using qRT-PCR

The expression levels of PIMREG, MTFR2, and CEP55 were assessed by qRT-PCR, which showed a significantly higher level of expression for these genes in BC samples compared with that in corresponding adjacent breast samples. We also investigated the relationship between gene expressions and TNM stage. Significantly different expression levels of PIMREG and MTFR2 were identified between TNM stage II group

and stage III group as shown in Fig. 6b-g.

Discussion

Typical treatments for BC besides surgery include endocrine therapy, chemical therapy, and radiotherapy; these have led to increased survival rates in majority of patients since most of the tumor cell express and respond to receptors for estrogen and progesterone receptor. However, as the most aggressive subtype, triple-negative BC have limited treatment options and worse prognosis. The presence of CSCs was discovered as one of the main causes for treatment resistance and there is lack of targets for CSCs in advanced cancers. In fact, one of the main manifestations in the cancer progression is the gradual loss of the differentiated phenotype and gaining stem cell-like characteristics [5]. Evidence showed that the frequency of CSC is increasing with tumor progression in multiple solid cancer types [16]; however, this process is non-linear in BC due to the dynamic negative feedback control of tumor cell populations and high heterogeneity within cells in BC tissue [17]. Therefore, it is crucial to observe the dynamic change of the stemness characteristics in the progression of BC. In this study, we discovered that the mRNAsi score was the lowest in stage I and reaches the highest in stage IV, which matched with the data published by earlier researchers. To be noted, tumor reached the first peak in stage II (and T2) and decreased in stage III (and T3) of the stem cell characteristics, this provided an insight into a specific timeline for detecting the vitality of breast CSC [17].

Diverse mechanisms have been unraveled that could help CSCs to survive under

hostile environment, including enhancing DNA-repair capacity [18], among which are overexpression of multifunctional efflux transporters and aberrant activation of developmental pathways [19, 20]. We applied mRNAsi-based WGCNA and identified the module that was associated with BCSC characteristics. Functional annotations of the module were primarily associated with cell cycle control, cell proliferation characteristics, and DNA-damage repair.

Immunotherapy has been developed rapidly in recent years due to the deepened acknowledgement of the immune system evasion by tumors. Developments, such as immune checkpoint inhibitors and receptors, have been made for the treatment of aggressive cancer. In order to observe the potential immune target in the CSC, we depicted the immune landscape of BC by clustering all patients into three clusters of different immune infiltration degrees. Our immune landscape revealed that T cell mediated immune response and innate immune response may lead the anti-tumor effect, further results from TIMER have confirmed these findings and provided a more specific profile of immune include dendritic cell and B lymphocyte. This provided a prospect and reminder in the direction of CSC-target immunotherapy since there have been successful attempts of generating CSC-primed T cells in vitro that showed targeting of CSCs after adoptive transfer *in vivo* [21]. It has also been demonstrated that significant anti-CSC immunity was induced by dendritic cell vaccine basing on CSC both in vitro and in immune-competent hosts [8, 22]. B lymphocytes were recently found to play a pivotal role in the cancer immune by being an assistant for T lymphocyte [23].

We also testified the reliability of the immune landscape by applying ESTIMATE analysis, as shown by someone, the relevance of breast CSC markers may vary according to the heterogeneity of the TME [24], where contains multiple types of cells, such as fibroblasts, endothelial cells, and mesenchymal stem cells, they are thought to have an affection to breast CSCs survival by secreting signaling molecules. Our results showed that patients in the high-infiltration cluster had the highest immune cell, highest stromal cell proportion, and lowest tumor purity compared with those in patients in the low- and intermediate-infiltration clusters, which is in agreement with previous studies. The change the stromal score from stage II to stage III and the estimate score from T2 to T3 also showed inversion, which synchronized with the stemness character. We found samples in stage IIIA tend to have highest stromal scores (N2), higher estimate scores (T3) and lower mRNAsi while samples in stage II tended to have lower stromal scores (N0-1), lower estimate scores (T2) and higher mRNAsi. These results indicated that the immune-like activity enhanced and likely validly regulated the stemness of BC cells in stage IIIA compared to stage II. The relationship between immune characteristics and stemness needs to be explored and the genes related to both stemness and immune properties are promising research targets.

As Pan Q et al. stated, identification of specific antigens or genetic alterations in CSCs may provide more specific targets for immunotherapy [25]. In order to uncover the epigenetic regulations of BCSC have utilized to survive under immune surveillance, the stemness-related genes were assessed under different immune infiltration clusters,

13 stemness-related genes were found to expressed differently, expression level of these genes were significantly upregulated in tumor samples and associated with poorer overall survival and progression in BC. Among these, RAD51 is a protein encoding gene that is recruited to the perturbed replication DSBs and forks sites and respectively blocks the exonuclease activity of MRE11 on DSB repair and on the replicated genome and eventually limit self-DNA accumulation in the cytosol, this process also prevents the initiation of innate immune signaling mediated by STING [26], which brought an insight to the potential relationship between stemness and immune escape mechanisms in the tumor cell.

In order to observe the relationship between the expression of key genes and clinical characteristics of patients, we applied qRT-PCR to evaluate the expression of PIMREG, MTFR2, and CEP55 (Fig.7). Results have shown that these genes were highly expressed in BC tissue, these results indicated that the selected key genes may become therapeutic targets of BC, to be noted, PIMREG, MTFR2 were shown to have higher expression in TNM stage II compared with stage III, this indicate that the selection of preferable timing is rather crucial, for both the stemness of the tumor cells and immune infiltration characters of tumor tissue are in the phases of competition. PIMREG is commonly known as a cell cycle promoter of hypoxic fetal cardiomyocytes [27]. It has been discovered that PIMREG promoted the aggressiveness of BC by disrupting the NF-κB/IκBα feedback loop [28]. To be noted, NF-κB activation was proven to be involved in the tumor associated macrophages-mediated tumor growth in human

pancreatic ductal adenocarcinoma [29]. PIMREG is known to positively regulate STAT3 activity to promote cell differentiation and shown to be associated with poor survival in the BC and pancreatic cancer [29-32]. It has been proven that both anti- and proinflammatory cytokines signal through associated receptor/JAK complexes and result in the phosphorylation of STAT3 [33]. CEP55 protein is required for membrane fusion of cytokinesis through the inhibitory of cyclin-dependent kinase 1 phosphorylation [34, 35], CEP55 was found to inhibit apoptosis of human glioma cells via the PI3K/Akt signaling pathway, the PI3K/Akt pathway is a crucial pathway in the immune escape, it promotes the progression of BC growth by suppressing NK cell cytotoxicity through eIF2b [36], also, the inhibition of PI3K in lung cancer downregulated PD-L1 expression [37]. MTFR2 is known as a protein that belongs to the MTFR family, it has been found that the MTFR2 was upregulated in BC and associated with poor survival of BC patients [38]. MTFR2 alters glucose metabolism through activating HIF1α and HIF2α (EPAS1) in the BC cell lines [38, 39], the stability of which is also controlled by PI3K/AKT signaling axis [40]. MTFR2 was also identified as an activator of TTK promoter in glioma stem-like cells [41]. One of the strengths of the immune-therapies is the ability to target multiple antigens, which makes these approaches perfectly suited for targeting heterogenous CSC populations [42]. PIMREG and MTFR2 are considered as potential target genes in CSC population that are required for effective immune targeting, the pan-cancer analysis implies that these genes may also play important roles in other tumors.

Conclusion

We identified 13 key genes related to stemness and immune escape and the stemness and immune escape degree of BC were increased non-linearly in the early stage of BC progression. *PIMREG* and *MTFR2* are considered as effective diagnostic markers and potential targets for therapy. The development of stemness is involved in the regulation of both innate- and acquired-immune microenvironment.

List of abbreviations

BC: Breast cancer; CSCs: Cancer stem cells; DEGs: Differentially expressed genes; TCGA: The Cancer Genome Atlas; GTEx: Genotype-Tissue Expression; K-M: Kaplan-Meier; GS: Gene significance; MM: Module membership; ssGSEA: Single-Sample Gene Set Enrichment Analysis; TME: Tumor microenvironment; PPI: Protein-protein interaction; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; TNM: Tumor Node Metastasis.

Declarations

Ethics approval and consent to participate

All patients were informed and written informed consent were provided. The study was conducted according to the clinical practice guidelines of the International Conference on Harmonization and the Declaration of Helsinki. This study protocol was approved by the ethical committee of Affiliated Hospital of Chongqing Medical University. This study protocol registered with Chinese Clinical Trial Registry

(http://www.chictr.org.cn/showproj.aspx?proj=19710; Date of registration: 25/09/2017; Registration number: ChiCTR-PDN-17012784) and was approved by the ethical committee of Affiliated Hospital of Chongqing Medical University (approval number: 2020-119).

Consent for publication

Not applicable

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interest.

Funding

Not applicable

Authors' contributions

Conception and design: LW. Acquisition of data: WYY. Analysis and interpretation of: WLB. Writing, review and/or revision of manuscript: WLB, LJK, TJJ. Administrative, technical or material support: YXD. Study supervision: LW, LJK. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Authors' information (optional)

Linbang Wang, Department of Orthopedic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China.

Yuanyuan Wang, Department of Endocrine and Breast Surgery, The First Affliated Hospital of Chongqing Medical University, Chongqing 400016, China.

Jingkun Liu, Department of Orthopedics, Honghui Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710054, China

Jiaojiao Tai, Department of Orthopedics, Honghui Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710054, China

Xuedong Yin, Department of Endocrine and Breast Surgery, The First Affliated Hospital of Chongqing Medical University, Chongqing 400016, China.

Wei Liu, Department of Orthopedic Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China.

Figure legends

Fig. 1 Overall flowchart of key gene identification

Fig. 2 Differentially analysis of mRNAsi in breast cancer

(a) The beeswarm plot shows significant difference of mRNAsi between normal (blue dot) and tumor (red dot) samples. The mRNAsi of most tumor samples is higher than

normal samples.

(b) The interactive genes between survival-related DEGs and mRNAsi-related genes. The Venn chat shows 193 survival-related DEGs (blue), 82 mRNAsi-related genes (red) and 18 interactive genes.

(c) The beeswarm plot shows that TNM stages IV and II have higher mRNAsi. The order of average mRNAsi from high to low is IV, II, III, and I.

(d) The beeswarm plot shows that T4 and T2 have higher mRNAsi. The order of average mRNAsi from high to low is T4, T2, T3, and T1.

(e) TIMER revealed the correlation of key gene expressions with immune cells (including macrophages, neutrophils, dendritic cells, B cells, CD8+T cells, and CD4+T cells) infiltration level in BC.

Fig. 3 Weighted gene co-expression network of breast cancer.

(a) Confirmation of hub modules. The branches of the cluster dendrogram represent nine different gene modules with different colors. Every leaf corresponds to a gene. Dynamic Tree Cut corresponds to the original module and Merged Dynamic corresponds to the final module.

(b) The correlation coefficient shows the relationship between the gene module and the clinical traits (mRNAsi and EREG-mRNAsi). Red corresponds to a positive correlation and blue corresponds to a negative correlation. The corresponding P-value is also annotated.

(c) and (d) Respectively represent for the cluster dendrogram and module-trait relationships of the soft threshold.

(e) Scatter plot of module eigengenes in turquoise module.

Fig. 4 Three infiltration clusters of BC patients after using ssGSEA according to 29 marker gene sets.

(a) 1069 BC samples are divided into high infiltration (red band), median infiltration (green band), and low infiltration clusters (blue band). The assessment of infiltration clusters conducted by TME are clustered into stromal score (b), immune score (c), and tumor purity (e) estimate score (d).

Fig. 5 The interaction between key genes.

(a) Mantel analysis of co-expression of 13 key genes and clinical information, Class01 to Class05, respectively, represent the survival time and survival status, stage and TNM staging, tumor microenvironment, mRNAsi and EREG-mRNAsi, and immune group.

(b) PPI network analysis of 13 key genes. Thickness of the solid line represents the strength of the relationship.

(c) Histogram shows the number of nodes of 13 key genes in PPI network.

Fig. 6 The mRNA expression of key genes in multiple cancers and in BC patients

(a) The expression of key genes in pan-cancer scale is evaluated by Oncomine, the number in the table cell is determined by the number of data that meet thresholds. The

color depth represents the gene rank. The red cells suggest that the expression levels of key genes are relatively higher in tumor tissues, while blue cells imply the opposite.

(b) The result of qRT-PCR compared the differential expression of key gene in tumor and para-carcinoma tissue of patients, the expression between patients in stage II and stage III group is also evaluated.

Fig. 7 Schematic diagram of immuno-resistance profile of the CSCs.

Fig. S1 Screening of differentially expressed genes. The heat map of 2261 genes' expression shows significant difference between normal and tumor samples. Green to red means the gene expression from low to high. The blue band in the top shows normal samples from GTEx and TCGA database and the red band shows cancer samples from TCGA database. (x: samples, y: genes)

Fig. S2. Analysis of prognosis value of mRNAsi and TME

(a) The K-M curve shows significant difference between low (blue) and high (red) mRNAsi groups within 5 year-follow-up.

(b) The K-M curve shows significant differences between low infiltration (green), median infiltration (blue) and high infiltration (red) clusters.

(c) The K-M curve shows significant difference between low (blue) and high (red) immune score cluster.

503 (d) The K-M curve shows significant difference between low (blue) and high (red)
504 tumor purity cluster.

505 (e) The K-M curve shows significant difference between low (blue) and high (red)
506 estimate score cluster.

507 Fig. S3. Analysis of clinicopathological characteristics of mRNAsi and TME

508 (a) The beeswarm plot shows TNM stage I and stage III have higher stromal score. The
509 order of average stromal score from high to low is TNM stage I, TNM stage III, TNM
510 stage II, and TNM stage IV.

511 (b) The beeswarm plot shows that T1 has a higher stromal score. The order of average
512 stromal score from high to low is T1, T2, T3, and T4.

513 (c) The beeswarm plot shows that N2 and N3 have higher stromal score. The order of
514 average stromal score from low to high is N1, N0, N3, and N2.

515 (d) The beeswarm plot shows that T1 and T3 have higher Estimate score. The order of
516 average Estimate score from high to low is T1, T3, T2, and T4.

517 (e) The beeswarm plot shows that T1 and T3 have lower tumor purity. The order of
518 average tumor purity from low to high is T1, T3, T2, and T4.

519 Fig. S4. Correlation analysis between 18 mRNAsi-related genes and infiltration clusters

520 The boxplots show there is no statistical difference on the expression of DSCC1(D),
521 ERCC6L(F), GINS1(G), KIF4A(H) and RACGAP1(L) between 3 infiltration clusters.

While the left genes (ANLN, CCNB2, CEP55, FAM83D, MELK, MTFR2, PIMREG, RAD54L, RAD51, SHCBP1, SKA1, UBE2T, and SKA3) are analyzed with a difference between infiltration clusters. (ns, $p > 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Fig. S5 Difference analysis of the gene expression between normal and tumor samples

(a) The boxplot shows significant difference of gene expression between normal and tumor sample. The expression of these 13 genes in tumor sample are higher than in normal samples.

(b) The heatmap shows the expression change of genes from normal (blue band) to tumor (red band) samples. Green to red means gene expression from low to high.

Fig. S6 Survival analysis of 13 key genes. All the K-M curve shows significant difference between low and high expression of 13 genes ($P < 0.05$). Low expression of these genes has a longer OS.

Fig. S7 GO and KEGG pathway enrichment analysis of key genes. The size of the circle represents the number of genes and the y-axis shows the GO and KEGG pathway terms. The redder the color the higher the value of p. Molecular Function (MF); Biological Process (BP) and Cellular Component (CC).

(a) Enrichment of Gene Ontology (GO) analysis.

(b) Enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

References

1. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A: Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians* 2017, 67(6):439-448.
2. Liu Z, Zhang XS, Zhang S: Breast tumor subgroups reveal diverse clinical prognostic power. *Scientific reports* 2014, 4:4002.
3. Stingl J, Caldas C: Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nature reviews Cancer* 2007, 7(10):791-799.
4. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L: Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nature reviews Clinical oncology* 2016, 13(11):674-690.
5. Bai X, Ni J, Beretov J, Graham P, Li Y: Cancer stem cell in breast cancer therapeutic resistance. *Cancer treatment reviews* 2018, 69:152-163.
6. Sampieri K, Fodde R: Cancer stem cells and metastasis. *Seminars in cancer biology* 2012, 22(3):187-193.
7. Liu S, Wicha MS: Targeting breast cancer stem cells. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010, 28(25):4006-4012.
8. Pellegatta S, Poliani PL, Corno D, Menghi F, Ghielmetti F, Suarez-Merino B, Caldera V, Nava S, Ravanini M, Facchetti F *et al*: Neurospheres enriched in cancer stem-like cells are highly effective in eliciting a dendritic cell-mediated immune response against malignant gliomas. *Cancer research* 2006, 66(21):10247-10252.
9. Dianat-Moghadam H, Rokni M, Marofi F, Panahi Y, Yousefi M: Natural killer cell-based immunotherapy: From transplantation toward targeting cancer stem cells. *Journal of cellular physiology* 2018, 234(1):259-273.
10. Hsu JM, Xia W, Hsu YH, Chan LC, Yu WH, Cha JH, Chen CT, Liao HW, Kuo CW, Khoo KH *et al*: STT3-dependent PD-L1 accumulation on cancer stem cells promotes immune evasion. *Nature communications* 2018, 9(1):1908.
11. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, Angell H, Fredriksen T, Lafontaine L, Berger A *et al*: Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 2013, 39(4):782-795.
12. Bao X, Shi R, Zhang K, Xin S, Li X, Zhao Y, Wang Y: Immune Landscape of Invasive Ductal Carcinoma Tumor Microenvironment Identifies a Prognostic and Immunotherapeutically Relevant Gene Signature. *Frontiers in oncology* 2019, 9:903.
13. Liu Z, Li M, Jiang Z, Wang X: A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer. *Translational oncology* 2018, 11(2):311-329.
14. Jia Q, Wu W, Wang Y, Alexander PB, Sun C, Gong Z, Cheng JN, Sun H, Guan Y, Xia X *et al*: Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung cancer. *Nature communications* 2018, 9(1):5361.
15. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z: Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell reports* 2017, 18(1):248-262.
16. Fedele M, Cerchia L, Chiappetta G: The Epithelial-to-Mesenchymal Transition in Breast Cancer: Focus on Basal-Like Carcinomas. *Cancers* 2017, 9(10).

17. Liu X, Johnson S, Liu S, Kanojia D, Yue W, Singh UP, Wang Q, Wang Q, Nie Q, Chen H: Nonlinear growth kinetics of breast cancer stem cells: implications for cancer stem cell targeted therapy. *Scientific reports* 2013, 3:2473.
18. Vitale I, Manic G, De Maria R, Kroemer G, Galluzzi L: DNA Damage in Stem Cells. *Molecular cell* 2017, 66(3):306-319.
19. Steinbichler TB, Dudas J, Skvortsov S, Ganswindt U, Riechelmann H, Skvortsova, II: Therapy resistance mediated by cancer stem cells. *Seminars in cancer biology* 2018, 53:156-167.
20. Choi YJ, Park JH, Han JW, Kim E, Jae-Wook O, Lee SY, Kim JH, Gurunathan S: Differential Cytotoxic Potential of Silver Nanoparticles in Human Ovarian Cancer Cells and Ovarian Cancer Stem Cells. *International journal of molecular sciences* 2016, 17(12).
21. Rapp C, Warta R, Stamova S, Nowrouzi A, Geisenberger C, Gal Z, Roesch S, Dettling S, Juenger S, Bucur M *et al*: Identification of T cell target antigens in glioblastoma stem-like cells using an integrated proteomics-based approach in patient specimens. *Acta neuropathologica* 2017, 134(2):297-316.
22. Arias-Pulido H, Cimino-Mathews A, Chaher N, Qualls C, Joste N, Colpaert C, Marotti JD, Foisey M, Prossnitz ER, Emens LA *et al*: The combined presence of CD20 + B cells and PD-L1 + tumor-infiltrating lymphocytes in inflammatory breast cancer is prognostic of improved patient outcome. *Breast cancer research and treatment* 2018, 171(2):273-282.
23. Wang H, Yang M, Lin L, Ren H, Lin C, Lin S, Shen G, Ji B, Meng C: HepG2 cells acquire stem cell-like characteristics after immune cell stimulation. *Cellular oncology (Dordrecht)* 2016, 39(1):35-45.
24. Dirkse A, Golebiewska A, Buder T, Nazarov PV, Muller A, Poovathingal S, Brons NHC, Leite S, Sauvageot N, Sarkisjan D *et al*: Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nature communications* 2019, 10(1):1787.
25. Pan Q, Li Q, Liu S, Ning N, Zhang X, Xu Y, Chang AE, Wicha MS: Concise Review: Targeting Cancer Stem Cells Using Immunologic Approaches. *Stem cells (Dayton, Ohio)* 2015, 33(7):2085-2092.
26. Mukherjee S, Abdisalaam S, Bhattacharya S, Srinivasan K, Sinha D, Asaithamby A: Mechanistic link between DNA damage sensing, repairing and signaling factors and immune signaling. *Advances in protein chemistry and structural biology* 2019, 115:297-324.
27. Hashimoto K, Kodama A, Honda T, Hanashima A, Ujihara Y, Murayama T, Nishimatsu SI, Mohri S: Fam64a is a novel cell cycle promoter of hypoxic fetal cardiomyocytes in mice. *Scientific reports* 2017, 7(1):4486.
28. Jiang L, Ren L, Zhang X, Chen H, Chen X, Lin C, Wang L, Hou N, Pan J, Zhou Z *et al*: Overexpression of PIMREG promotes breast cancer aggressiveness via constitutive activation of NF-kappaB signaling. *EBioMedicine* 2019, 43:188-200.
29. Ye H, Zhou Q, Zheng S, Li G, Lin Q, Wei L, Fu Z, Zhang B, Liu Y, Li Z *et al*: Tumor-associated macrophages promote progression and the Warburg effect via CCL18/NF-kB/VCAM-1 pathway in pancreatic ductal adenocarcinoma. *Cell death & disease* 2018, 9(5):453.
30. Yao Z, Zheng X, Lu S, He Z, Miao Y, Huang H, Chu X, Cai C, Zou F: Knockdown of FAM64A suppresses proliferation and migration of breast cancer cells. *Breast cancer (Tokyo, Japan)* 2019, 26(6):835-845.

31. Jiao Y, Fu Z, Li Y, Zhang W, Liu Y: Aberrant FAM64A mRNA expression is an independent predictor of poor survival in pancreatic cancer. *PloS one* 2019, 14(1):e0211291.
32. Li Z, Zhang Y, Zhang Z, Zhao Z, Lv Q: A four-gene signature predicts the efficacy of paclitaxel-based neoadjuvant therapy in human epidermal growth factor receptor 2-negative breast cancer. *Journal of cellular biochemistry* 2019, 120(4):6046-6056.
33. Owen KL, Brockwell NK, Parker BS: JAK-STAT Signaling: A Double-Edged Sword of Immune Regulation and Cancer Progression. *Cancers* 2019, 11(12).
34. Zhou C, Hancock JL, Khanna KK, Homer HA: First meiotic anaphase requires Cep55-dependent inhibitory cyclin-dependent kinase 1 phosphorylation. *Journal of cell science* 2019, 132(18).
35. Mondal G, Rowley M, Guidugli L, Wu J, Pankratz VS, Couch FJ: BRCA2 localization to the midbody by filamin A regulates cep55 signaling and completion of cytokinesis. *Developmental cell* 2012, 23(1):137-152.
36. Jin F, Wu Z, Hu X, Zhang J, Gao Z, Han X, Qin J, Li C, Wang Y: The PI3K/Akt/GSK-3beta/ROS/eIF2B pathway promotes breast cancer growth and metastasis via suppression of NK cell cytotoxicity and tumor cell susceptibility. *Cancer biology & medicine* 2019, 16(1):38-54.
37. Gao Y, Yang J, Cai Y, Fu S, Zhang N, Fu X, Li L: IFN-gamma-mediated inhibition of lung cancer correlates with PD-L1 expression and is regulated by PI3K-AKT signaling. *International journal of cancer* 2018, 143(4):931-943.
38. Lu G, Lai Y, Wang T, Lin W, Lu J, Ma Y, Chen Y, Ma H, Liu R, Li J: Mitochondrial fission regulator 2 (MTFR2) promotes growth, migration, invasion and tumour progression in breast cancer cells. *Aging* 2019, 11(22):10203-10219.
39. Monticone M, Panfoli I, Ravera S, Puglisi R, Jiang MM, Morello R, Candiani S, Tonachini L, Biticchi R, Fabiano A *et al*: The nuclear genes Mtfr1 and Dufd1 regulate mitochondrial dynamic and cellular respiration. *Journal of cellular physiology* 2010, 225(3):767-776.
40. Dey N, Sun Y, Carlson JH, Wu H, Lin X, Leyland-Jones B, De P: Anti-tumor efficacy of BEZ235 is complemented by its anti-angiogenic effects via downregulation of PI3K-mTOR-HIF1alpha signaling in HER2-defined breast cancers. *American journal of cancer research* 2016, 6(4):714-746.
41. Wang J, Xie Y, Bai X, Wang N, Yu H, Deng Z, Lian M, Yu S, Liu H, Xie W *et al*: Targeting dual specificity protein kinase TTK attenuates tumorigenesis of glioblastoma. *Oncotarget* 2018, 9(3):3081-3088.
42. Eltoukhy HS, Sinha G, Moore CA, Sandiford OA, Rameshwar P: Immune modulation by a cellular network of mesenchymal stem cells and breast cancer cell subsets: Implication for cancer therapy. *Cellular immunology* 2018, 326:33-41.

Figures

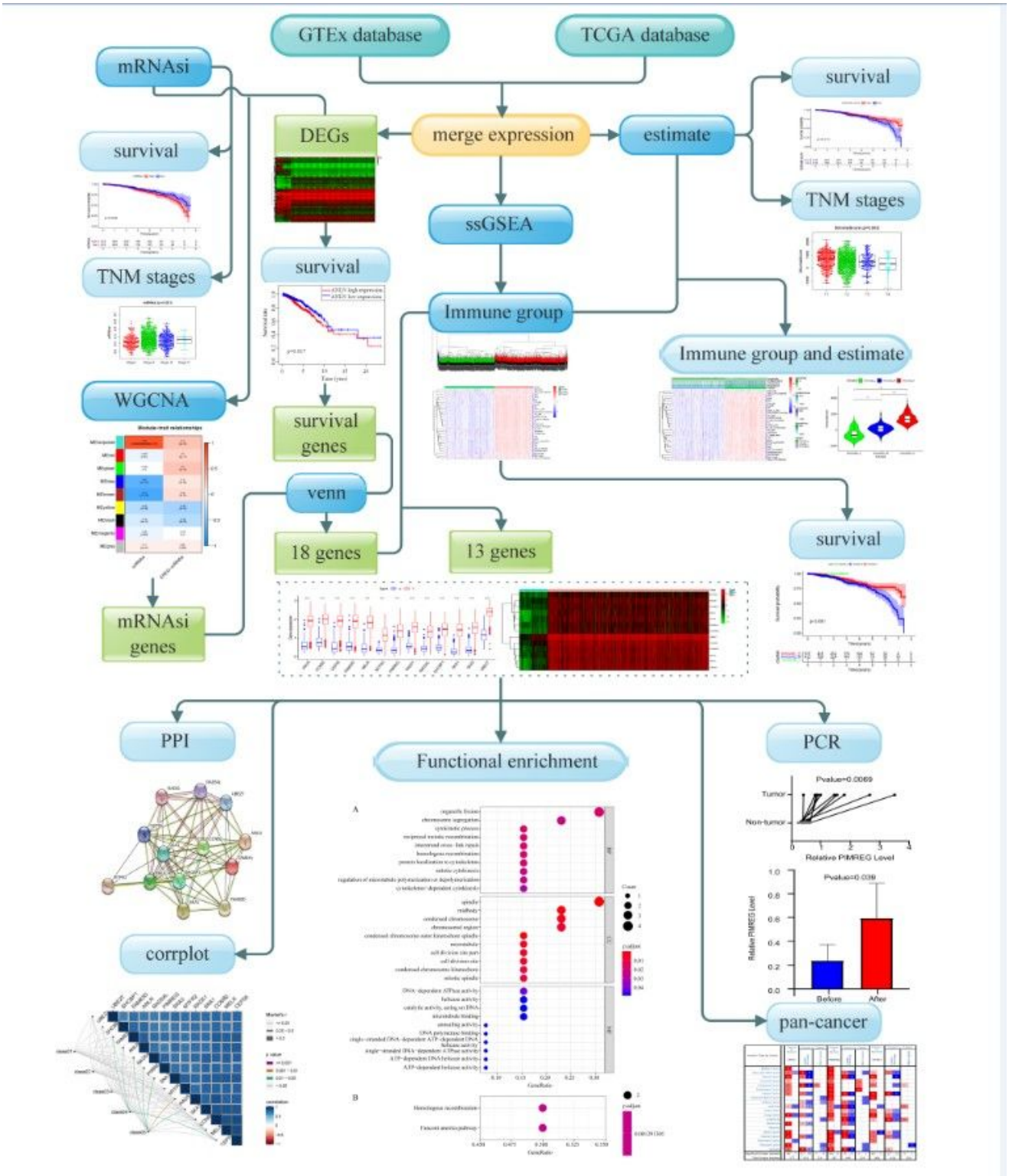

Figure 1

Overall flowchart of key gene identification

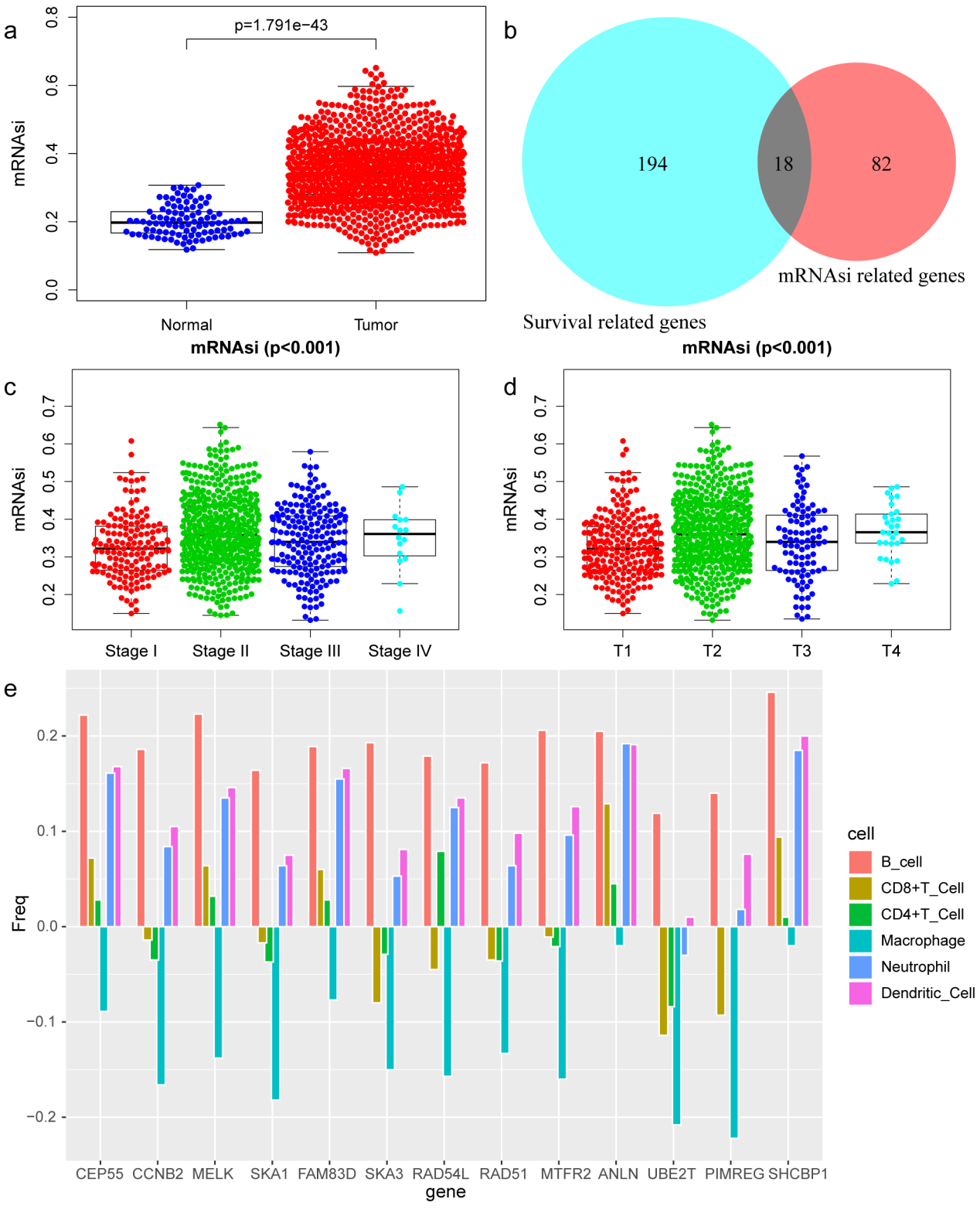

Figure 2

Differentially analysis of mRNAsi in breast cancer (a) The beeswarm plot shows significant difference of mRNAsi between normal (bluedot) and tumor (red dot) samples. The mRNAsi of most tumor samples is higher than normal samples. (b) The interactive genes between survival-related DEGs and mRNAsi-related genes. The Venn chat shows 193 survival-related DEGs (blue), 82 mRNAsi-related genes (red) and 18 interactive genes. (c) The beeswarm plot shows that TNM stages IV and II have higher mRNAsi. The order

of average mRNAsi from high to low is IV, II, III, and I. (d) The beeswarm plot shows that T4 and T2 have higher mRNAsi. The order of average mRNAsi from high to low is T4, T2, T3, and T1. (e) TIMER revealed the correlation of key gene expressions with immune cells (including macrophages, neutrophils, dendritic cells, B cells, CD8+T cells, and CD4+T cells) infiltration level in BC.

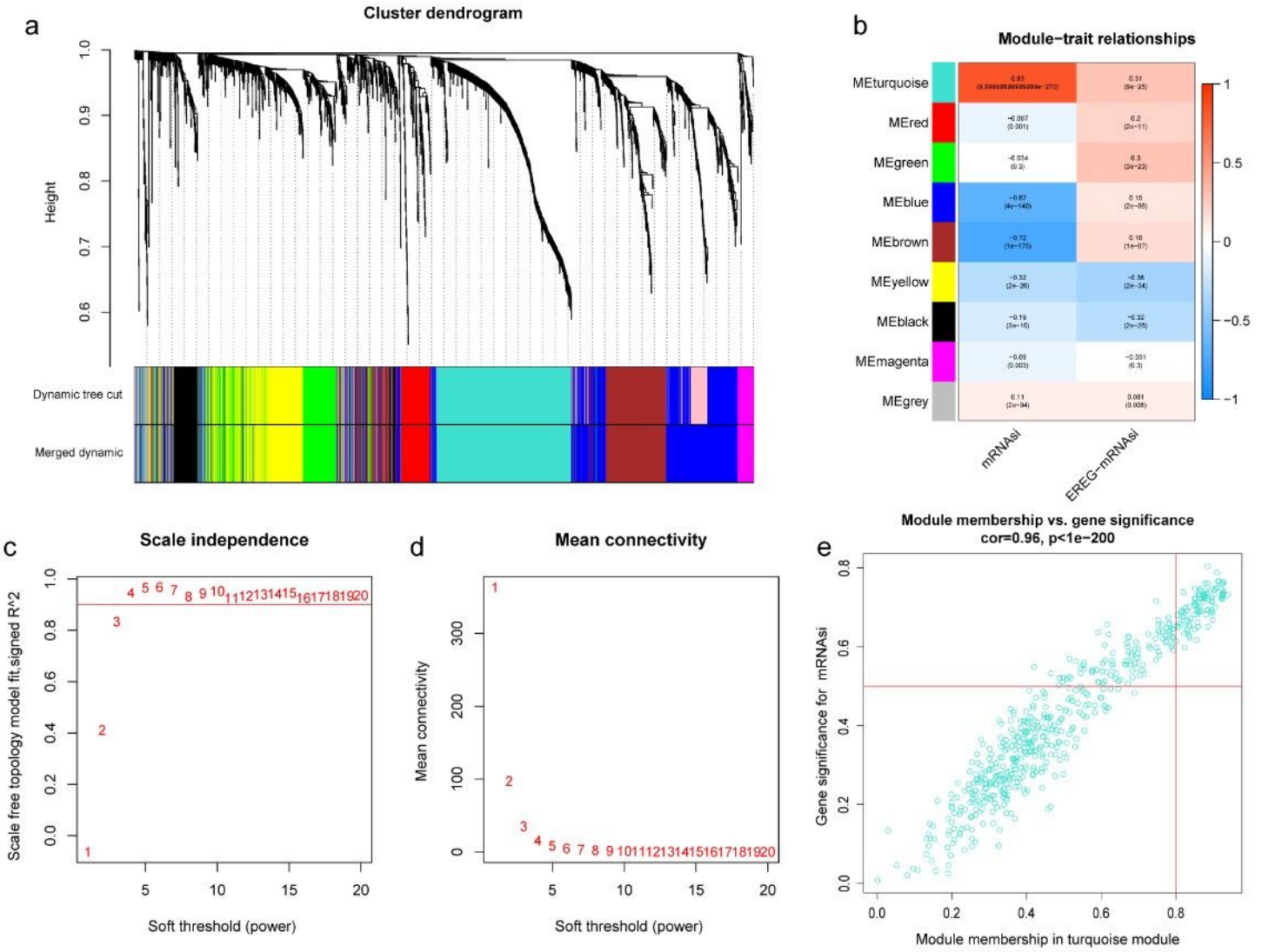

Figure 3

Weighted gene co-expression network of breast cancer. (a) Confirmation of hub modules. The branches of the cluster dendrogram represent nine different gene modules with different colors. Every leaf corresponds to a gene. Dynamic Tree Cut corresponds to the original module and Merged Dynamic corresponds to the final module. (b) The correlation coefficient shows the relationship between the gene module and the clinical traits (mRNAsi and EREG-mRNAsi). Red corresponds to a positive correlation and blue corresponds to a negative correlation. The corresponding P-value is also annotated. (c) and (d) Respectively represent for the cluster dendrogram and module-trait relationships of the soft threshold. (e) Scatter plot of module eigengenes in turquoise module.

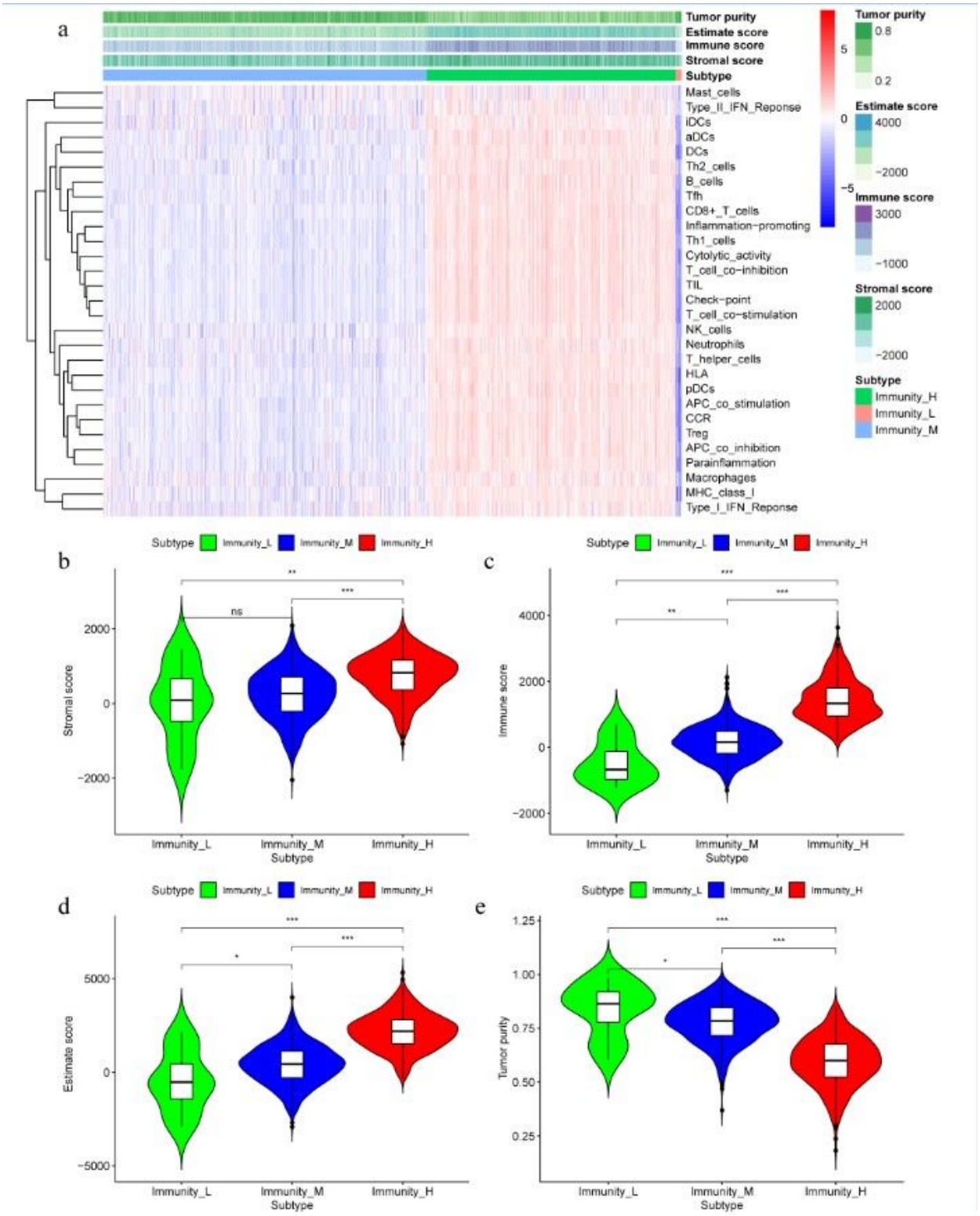

Figure 4

Three infiltration clusters of BC patients after using ssGSEA according to 29 marker gene sets. (a) 1069 BC samples are divided into high infiltration (red band), median infiltration green band), and low infiltration clusters (blue band). The assessment of infiltration clusters conducted by TME are clustered into stromal score (b), immune score (c), and tumor purity (e) estimate score (d).

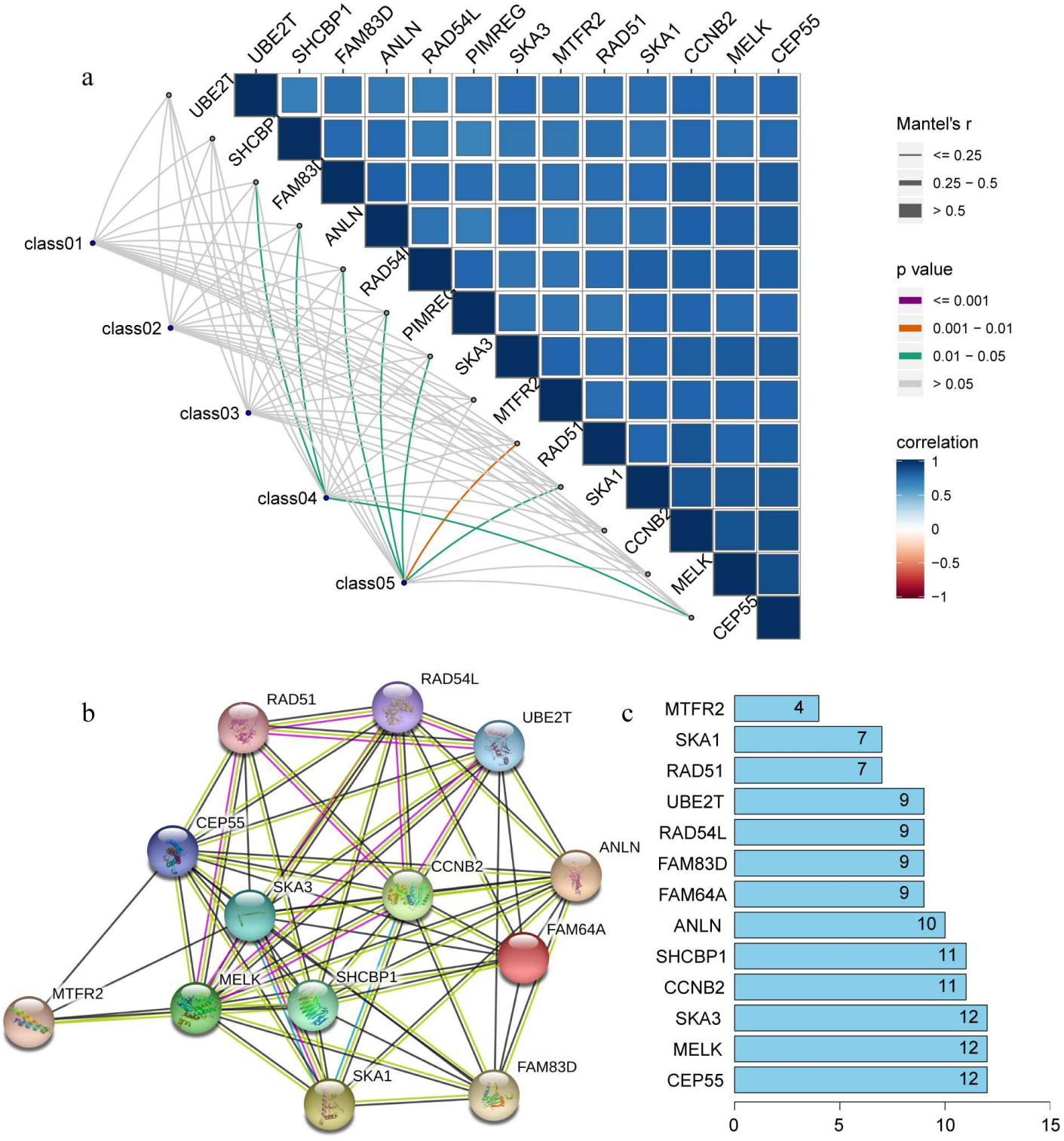

Figure 5

The interaction between key genes. (a) Mantel analysis of co-expression of 13 key genes and clinical information, Class01 476 to Class05, respectively, represent the survival time and survival status, stage and TNM staging, tumor microenvironment, mRNAsi and EREG-mRNAsi, and immune group. 478 (b) PPI network analysis of 13 key genes. Thickness of the solid line represents the strength of the relationship. (c) Histogram shows the number of nodes of 13 key genes in PPI network.

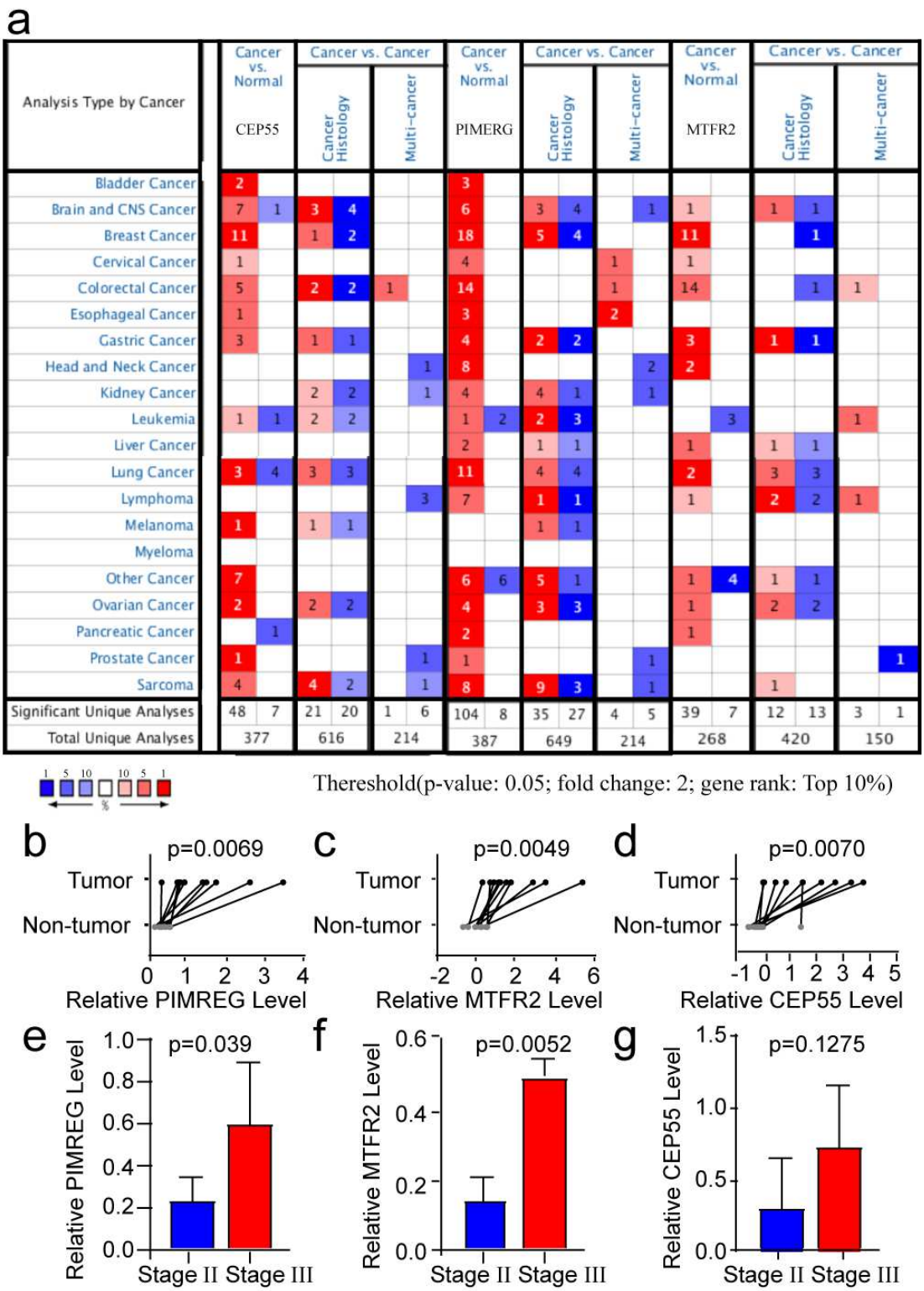

Figure 6

The mRNA expression of key genes in multiple cancers and in BC patients (a) The expression of key genes in pan-cancer scale is evaluated by Oncomine, the number in the table cell is determined by the number of data that meet thresholds. The 25 color depth represents the gene rank. The red cells suggest that the expression levels of key genes are relatively higher in tumor tissues, while blue cells imply the opposite. (b) The result of qRT-PCR compared the differential expression of key gene in tumor and para-

carcinoma tissue of patients, the expression between patients in stage II and stage III group is also evaluated.

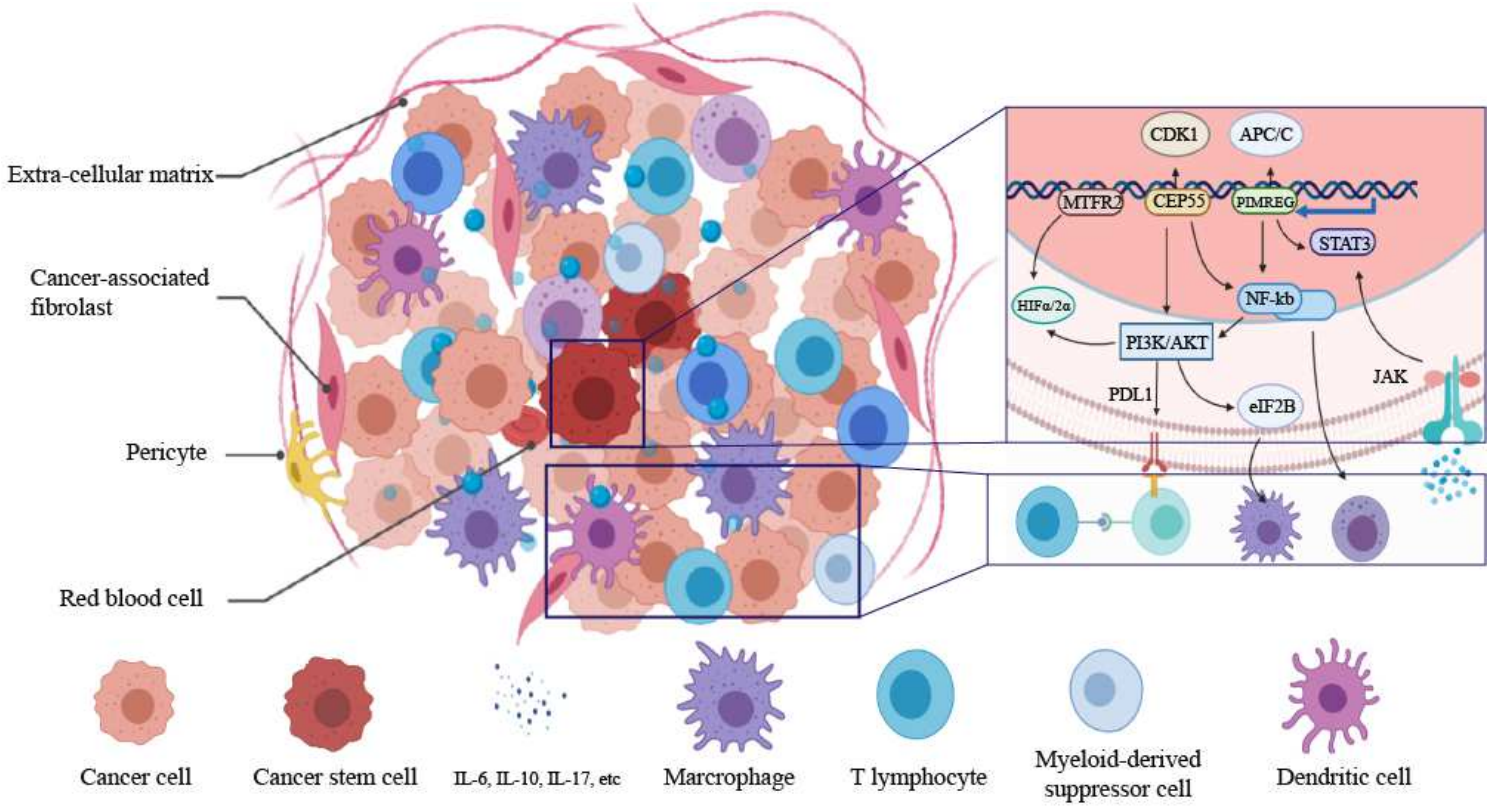

Figure 7

Schematic diagram of immuno-resistance profile of the CSCs.

Supplementary Files

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

- FigS4.tif
- FigS7.tif
- FigS5.tif
- FigS1.tif
- FigS4.tif
- FigS7.tif
- FigS3.tif
- FigS5.tif
- FigS3.tif
- FigS2.tif
- FigS6.tif

- FigS1.tif
- FigS6.tif
- FigS2.tif