

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

4 **Authors: Linbang Wang¹⁺, Yuanyuan Wang²⁺, Jingkun Liu³, Jiaojiao Tai³,**
5 **Xuedong Yin², Wei Liu^{1*}**

6 **Author details**

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

9 ²Department of Endocrine and Breast Surgery, The First Affiliated Hospital of
10 Chongqing Medical University, Chongqing 400016, China.

11 ³Department of Orthopedics, Honghui Hospital, Xi'an Jiaotong University, Xi'an,
12 Shaanxi 710054, China

13 ⁺These authors contributed equally to this work.

14 * Corresponding Author

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

16

17 **Abstract**

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

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

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

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

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

39 **Trial registration:** This study protocol registered with Chinese Clinical Trial Registry
40 (<http://www.chictr.org.cn/showproj.aspx?proj=19710>; Date of registration: 25/09/2017;
41 Registration number: ChiCTR-PDN-17012784) and was approved by the ethical
42 committee of Affiliated Hospital of Chongqing Medical University (approval number:
43 2020-119).

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

46 **Background**

47 Breast cancer (BC) has the highest incidence rate and mortality rate among female
48 malignant tumors, which impacts women's health significantly [1]. It is considered to
49 be heterogeneous depending on molecular subtype and on different stages of cancer
50 progression [2]. This heterogeneity poses challenges during treatment, even through
51 various treatment strategies have been developed based on different pathological types
52 [3], especially for the triple-negative BC [4]. Thus, further explorations are needed to
53 identify new markers for guiding individualized treatment.

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

57 [5, 6]. Diverse mechanisms by which CSCs manage to survive through various
58 strategies including tumor initiation, metastatic reactivation, oncogene- and immune-
59 targeted therapy resistance have been unraveled [4, 7].

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

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

76 **Methods**

77 **Design and data processing of this study**

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

83 **Data acquisition**

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

93 **Differentially Expressed Genes**

94 The “edgeR” R package was used to screen DEGs of normal breast and cancer samples
95 with the following selection criteria: $FDR < 0.05$, and $|\log_2 FC| > 1$. The values of genes
96 calculated by limma R package, and genes with expression of < 0.5 were deleted. The

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

98 **WGCNA and Module merge**

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

118 **Identification of mRNAsi- and survival-associated genes**

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

124 **Immune infiltration grouping of BC patients**

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

132 **Tumor microenvironment (TME) analysis of BC**

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

138 **Identification of key genes**

139 All interactive genes associated with survival time and mRNA_{si} of Venn chart were
140 enrolled for correlation analysis with three infiltration clusters. Genes with no
141 significant difference between immune clusters and mRNA_{si} were excluded ($P \geq 0.05$).
142 Correlation analysis and survival analysis of these genes were conducted. The results
143 were shown in box plots, heatmaps, and K-M curves. Oncomine platform
144 (<http://www.oncomine.org>) was used to inspect differential expression of key genes
145 between BC and healthy tissues and between different tumors.

146 **Functional analysis of key genes**

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

156 **Patients**

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

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

174 **Total RNA extraction**

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

179 **Reverse transcription and qRT-PCR quantification**

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

191 **Statistical analysis**

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

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

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

202 **Results**

203 **Screening for survival-related DEGs**

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

209 **Correlation analysis of mRNAsi and clinical characteristics**

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

219 **Identification of significant modules and mRNAsi-associated genes**

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

234 **Immune phenotype landscape in the TME of BC**

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

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

253 **Key genes identification and correspondence functional analysis**

254 Of 18 mRNAsi-related genes that were analyzed for correlation with three immune cell
255 infiltration clusters, 13 key genes were finally enrolled (Fig. S4). There was statistical
256 significance of the expression of ANLN, CCNB2, CEP55, FAM83D, MELK, MTFR2,
257 PIMREG, RAD54L, RAD51, SHCBP1, SKA1, UBE2T, and SKA3 within three
258 immune cell infiltration clusters. As shown in Fig. S5a-r, the expression levels for all
259 13 key genes were higher in tumor samples than in the normal samples. The K-M curves
260 showed that patients with higher level of expression of these genes had a poor overall

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

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

276 **Validation of key genes expression using qRT-PCR**

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

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

283 **Discussion**

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

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

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

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

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

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

345 13 stemness-related genes were found to expressed differently, expression level of these
346 genes were significantly upregulated in tumor samples and associated with poorer
347 overall survival and progression in BC. Among these, RAD51 is a protein encoding
348 gene that is recruited to the perturbed replication DSBs and forks sites and respectively
349 blocks the exonuclease activity of MRE11 on DSB repair and on the replicated genome
350 and eventually limit self-DNA accumulation in the cytosol, this process also prevents
351 the initiation of innate immune signaling mediated by STING [26], which brought an
352 insight to the potential relationship between stemness and immune escape mechanisms
353 in the tumor cell.

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

366 pancreatic ductal adenocarcinoma [29]. PIMREG is known to positively regulate
367 STAT3 activity to promote cell differentiation and shown to be associated with poor
368 survival in the BC and pancreatic cancer [29-32]. It has been proven that both anti- and
369 proinflammatory cytokines signal through associated receptor/JAK complexes and
370 result in the phosphorylation of STAT3 [33]. CEP55 protein is required for membrane
371 fusion of cytokinesis through the inhibitory of cyclin-dependent kinase 1
372 phosphorylation [34, 35], CEP55 was found to inhibit apoptosis of human glioma cells
373 via the PI3K/Akt signaling pathway, the PI3K/Akt pathway is a crucial pathway in the
374 immune escape, it promotes the progression of BC growth by suppressing NK cell
375 cytotoxicity through eIF2b [36], also, the inhibition of PI3K in lung cancer
376 downregulated PD-L1 expression [37]. MTFR2 is known as a protein that belongs to
377 the MTFR family, it has been found that the MTFR2 was upregulated in BC and
378 associated with poor survival of BC patients [38]. MTFR2 alters glucose metabolism
379 through activating HIF1 α and HIF2 α (EPAS1) in the BC cell lines [38, 39], the stability
380 of which is also controlled by PI3K/AKT signaling axis [40]. MTFR2 was also
381 identified as an activator of TTK promoter in glioma stem-like cells [41]. One of the
382 strengths of the immune-therapies is the ability to target multiple antigens, which makes
383 these approaches perfectly suited for targeting heterogenous CSC populations [42].
384 PIMREG and MTFR2 are considered as potential target genes in CSC population that
385 are required for effective immune targeting, the pan-cancer analysis implies that these
386 genes may also play important roles in other tumors.

387 **Conclusion**

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

393 **List of abbreviations**

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

400 **Declarations**

401 **Ethics approval and consent to participate**

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

407 (<http://www.chictr.org.cn/showproj.aspx?proj=19710>; Date of registration: 25/09/2017;
408 Registration number: ChiCTR-PDN-17012784) and was approved by the ethical
409 committee of Affiliated Hospital of Chongqing Medical University (approval number:
410 2020-119).

411 **Consent for publication**

412 Not applicable

413 **Availability of data and materials**

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

416 **Competing interests**

417 The authors declare that they have no competing interest.

418 **Funding**

419 Not applicable

420 **Authors' contributions**

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

425 **Acknowledgements**

426 Not applicable

427 **Authors' information (optional)**

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

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

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

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

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

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

440 **Figure legends**

441 Fig. 1 Overall flowchart of key gene identification

442 Fig. 2 Differentially analysis of mRNAsi in breast cancer

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

445 normal samples.

446 (b) The interactive genes between survival-related DEGs and mRNAsi-related genes.

447 The Venn chart shows 193 survival-related DEGs (blue), 82 mRNAsi-related genes (red)
448 and 18 interactive genes.

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

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

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

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

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

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

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

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

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

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

474 Fig. 5 The interaction between key genes.

475 (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
477 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
479 strength of the relationship.

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

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

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

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

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

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

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

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

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

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

501 (c) The K-M curve shows significant difference between low (blue) and high (red)
502 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.

522 While the left genes (ANLN, CCNB2, CEP55, FAM83D, MELK, MTFR2, PIMREG,
523 RAD54L, RAD51, SHCBP1, SKA1, UBE2T, and SKA3) are analyzed with a
524 difference between infiltration clusters. (ns, $p > 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, p
525 < 0.001).

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

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

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

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

535 Fig. S7 GO and KEGG pathway enrichment analysis of key genes. The size of the circle
536 represents the number of genes and the y-axis shows the GO and KEGG pathway terms.

537 The redder the color the higher the value of p . Molecular Function (MF); Biological
538 Process (BP) and Cellular Component (CC).

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

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

542 **References**

- 543 1. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A: Breast cancer statistics, 2017,
544 racial disparity in mortality by state. *CA: a cancer journal for clinicians* 2017, 67(6):439-448.
- 545 2. Liu Z, Zhang XS, Zhang S: Breast tumor subgroups reveal diverse clinical prognostic power.
546 *Scientific reports* 2014, 4:4002.
- 547 3. Stingl J, Caldas C: Molecular heterogeneity of breast carcinomas and the cancer stem cell
548 hypothesis. *Nature reviews Cancer* 2007, 7(10):791-799.
- 549 4. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L: Triple-negative breast cancer:
550 challenges and opportunities of a heterogeneous disease. *Nature reviews Clinical oncology* 2016,
551 13(11):674-690.
- 552 5. Bai X, Ni J, Beretov J, Graham P, Li Y: Cancer stem cell in breast cancer therapeutic resistance.
553 *Cancer treatment reviews* 2018, 69:152-163.
- 554 6. Sampieri K, Fodde R: Cancer stem cells and metastasis. *Seminars in cancer biology* 2012,
555 22(3):187-193.
- 556 7. Liu S, Wicha MS: Targeting breast cancer stem cells. *Journal of clinical oncology : official
557 journal of the American Society of Clinical Oncology* 2010, 28(25):4006-4012.
- 558 8. Pellegatta S, Poliani PL, Corno D, Menghi F, Ghielmetti F, Suarez-Merino B, Caldera V, Nava
559 S, Ravanini M, Facchetti F *et al*: Neurospheres enriched in cancer stem-like cells are highly
560 effective in eliciting a dendritic cell-mediated immune response against malignant gliomas.
561 *Cancer research* 2006, 66(21):10247-10252.
- 562 9. Dianat-Moghadam H, Rokni M, Marofi F, Panahi Y, Yousefi M: Natural killer cell-based
563 immunotherapy: From transplantation toward targeting cancer stem cells. *Journal of cellular
564 physiology* 2018, 234(1):259-273.
- 565 10. Hsu JM, Xia W, Hsu YH, Chan LC, Yu WH, Cha JH, Chen CT, Liao HW, Kuo CW, Khoo KH
566 *et al*: STT3-dependent PD-L1 accumulation on cancer stem cells promotes immune evasion.
567 *Nature communications* 2018, 9(1):1908.
- 568 11. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, Angell H, Fredriksen
569 T, Lafontaine L, Berger A *et al*: Spatiotemporal dynamics of intratumoral immune cells reveal
570 the immune landscape in human cancer. *Immunity* 2013, 39(4):782-795.
- 571 12. Bao X, Shi R, Zhang K, Xin S, Li X, Zhao Y, Wang Y: Immune Landscape of Invasive Ductal
572 Carcinoma Tumor Microenvironment Identifies a Prognostic and Immunotherapeutically
573 Relevant Gene Signature. *Frontiers in oncology* 2019, 9:903.
- 574 13. Liu Z, Li M, Jiang Z, Wang X: A Comprehensive Immunologic Portrait of Triple-Negative
575 Breast Cancer. *Translational oncology* 2018, 11(2):311-329.
- 576 14. Jia Q, Wu W, Wang Y, Alexander PB, Sun C, Gong Z, Cheng JN, Sun H, Guan Y, Xia X *et al*:
577 Local mutational diversity drives intratumoral immune heterogeneity in non-small cell lung
578 cancer. *Nature communications* 2018, 9(1):5361.
- 579 15. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski
580 Z: Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships
581 and Predictors of Response to Checkpoint Blockade. *Cell reports* 2017, 18(1):248-262.
- 582 16. Fedele M, Cerchia L, Chiappetta G: The Epithelial-to-Mesenchymal Transition in Breast Cancer:
583 Focus on Basal-Like Carcinomas. *Cancers* 2017, 9(10).

- 584 17. Liu X, Johnson S, Liu S, Kanojia D, Yue W, Singh UP, Wang Q, Wang Q, Nie Q, Chen H:
585 Nonlinear growth kinetics of breast cancer stem cells: implications for cancer stem cell targeted
586 therapy. *Scientific reports* 2013, 3:2473.
- 587 18. Vitale I, Manic G, De Maria R, Kroemer G, Galluzzi L: DNA Damage in Stem Cells. *Molecular*
588 *cell* 2017, 66(3):306-319.
- 589 19. Steinbichler TB, Dudas J, Skvortsov S, Ganswindt U, Riechelmann H, Skvortsova, II: Therapy
590 resistance mediated by cancer stem cells. *Seminars in cancer biology* 2018, 53:156-167.
- 591 20. Choi YJ, Park JH, Han JW, Kim E, Jae-Wook O, Lee SY, Kim JH, Gurunathan S: Differential
592 Cytotoxic Potential of Silver Nanoparticles in Human Ovarian Cancer Cells and Ovarian Cancer
593 Stem Cells. *International journal of molecular sciences* 2016, 17(12).
- 594 21. Rapp C, Warta R, Stamova S, Nowrouzi A, Geisenberger C, Gal Z, Roesch S, Dettling S,
595 Juenger S, Bucur M *et al*: Identification of T cell target antigens in glioblastoma stem-like cells
596 using an integrated proteomics-based approach in patient specimens. *Acta neuropathologica*
597 2017, 134(2):297-316.
- 598 22. Arias-Pulido H, Cimino-Mathews A, Chaheer N, Qualls C, Joste N, Colpaert C, Marotti JD,
599 Foisey M, Prossnitz ER, Emens LA *et al*: The combined presence of CD20 + B cells and PD-
600 L1 + tumor-infiltrating lymphocytes in inflammatory breast cancer is prognostic of improved
601 patient outcome. *Breast cancer research and treatment* 2018, 171(2):273-282.
- 602 23. Wang H, Yang M, Lin L, Ren H, Lin C, Lin S, Shen G, Ji B, Meng C: HepG2 cells acquire stem
603 cell-like characteristics after immune cell stimulation. *Cellular oncology (Dordrecht)* 2016,
604 39(1):35-45.
- 605 24. Dirkse A, Golebiewska A, Buder T, Nazarov PV, Muller A, Poovathingal S, Brons NHC, Leite
606 S, Sauvageot N, Sarkisjan D *et al*: Stem cell-associated heterogeneity in Glioblastoma results
607 from intrinsic tumor plasticity shaped by the microenvironment. *Nature communications* 2019,
608 10(1):1787.
- 609 25. Pan Q, Li Q, Liu S, Ning N, Zhang X, Xu Y, Chang AE, Wicha MS: Concise Review: Targeting
610 Cancer Stem Cells Using Immunologic Approaches. *Stem cells (Dayton, Ohio)* 2015,
611 33(7):2085-2092.
- 612 26. Mukherjee S, Abdisalaam S, Bhattacharya S, Srinivasan K, Sinha D, Asaithamby A:
613 Mechanistic link between DNA damage sensing, repairing and signaling factors and immune
614 signaling. *Advances in protein chemistry and structural biology* 2019, 115:297-324.
- 615 27. Hashimoto K, Kodama A, Honda T, Hanashima A, Ujihara Y, Murayama T, Nishimatsu SI,
616 Mohri S: Fam64a is a novel cell cycle promoter of hypoxic fetal cardiomyocytes in mice.
617 *Scientific reports* 2017, 7(1):4486.
- 618 28. Jiang L, Ren L, Zhang X, Chen H, Chen X, Lin C, Wang L, Hou N, Pan J, Zhou Z *et al*:
619 Overexpression of PIMREG promotes breast cancer aggressiveness via constitutive activation
620 of NF-kappaB signaling. *EBioMedicine* 2019, 43:188-200.
- 621 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
622 macrophages promote progression and the Warburg effect via CCL18/NF-kB/VCAM-1
623 pathway in pancreatic ductal adenocarcinoma. *Cell death & disease* 2018, 9(5):453.
- 624 30. Yao Z, Zheng X, Lu S, He Z, Miao Y, Huang H, Chu X, Cai C, Zou F: Knockdown of FAM64A
625 suppresses proliferation and migration of breast cancer cells. *Breast cancer (Tokyo, Japan)* 2019,
626 26(6):835-845.

- 627 31. Jiao Y, Fu Z, Li Y, Zhang W, Liu Y: Aberrant FAM64A mRNA expression is an independent
628 predictor of poor survival in pancreatic cancer. *PloS one* 2019, 14(1):e0211291.
- 629 32. Li Z, Zhang Y, Zhang Z, Zhao Z, Lv Q: A four-gene signature predicts the efficacy of paclitaxel-
630 based neoadjuvant therapy in human epidermal growth factor receptor 2-negative breast cancer.
631 *Journal of cellular biochemistry* 2019, 120(4):6046-6056.
- 632 33. Owen KL, Brockwell NK, Parker BS: JAK-STAT Signaling: A Double-Edged Sword of
633 Immune Regulation and Cancer Progression. *Cancers* 2019, 11(12).
- 634 34. Zhou C, Hancock JL, Khanna KK, Homer HA: First meiotic anaphase requires Cep55-
635 dependent inhibitory cyclin-dependent kinase 1 phosphorylation. *Journal of cell science* 2019,
636 132(18).
- 637 35. Mondal G, Rowley M, Guidugli L, Wu J, Pankratz VS, Couch FJ: BRCA2 localization to the
638 midbody by filamin A regulates cep55 signaling and completion of cytokinesis. *Developmental*
639 *cell* 2012, 23(1):137-152.
- 640 36. Jin F, Wu Z, Hu X, Zhang J, Gao Z, Han X, Qin J, Li C, Wang Y: The PI3K/Akt/GSK-
641 3beta/ROS/eIF2B pathway promotes breast cancer growth and metastasis via suppression of
642 NK cell cytotoxicity and tumor cell susceptibility. *Cancer biology & medicine* 2019, 16(1):38-
643 54.
- 644 37. Gao Y, Yang J, Cai Y, Fu S, Zhang N, Fu X, Li L: IFN-gamma-mediated inhibition of lung
645 cancer correlates with PD-L1 expression and is regulated by PI3K-AKT signaling. *International*
646 *journal of cancer* 2018, 143(4):931-943.
- 647 38. Lu G, Lai Y, Wang T, Lin W, Lu J, Ma Y, Chen Y, Ma H, Liu R, Li J: Mitochondrial fission
648 regulator 2 (MTFR2) promotes growth, migration, invasion and tumour progression in breast
649 cancer cells. *Aging* 2019, 11(22):10203-10219.
- 650 39. Monticone M, Panfoli I, Ravera S, Puglisi R, Jiang MM, Morello R, Candiani S, Tonachini L,
651 Biticchi R, Fabiano A *et al*: The nuclear genes Mtf1 and Dufd1 regulate mitochondrial dynamic
652 and cellular respiration. *Journal of cellular physiology* 2010, 225(3):767-776.
- 653 40. Dey N, Sun Y, Carlson JH, Wu H, Lin X, Leyland-Jones B, De P: Anti-tumor efficacy of
654 BEZ235 is complemented by its anti-angiogenic effects via downregulation of PI3K-mTOR-
655 HIF1alpha signaling in HER2-defined breast cancers. *American journal of cancer research*
656 2016, 6(4):714-746.
- 657 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
658 specificity protein kinase TTK attenuates tumorigenesis of glioblastoma. *Oncotarget* 2018,
659 9(3):3081-3088.
- 660 42. Eltoukhy HS, Sinha G, Moore CA, Sandiford OA, Rameshwar P: Immune modulation by a
661 cellular network of mesenchymal stem cells and breast cancer cell subsets: Implication for
662 cancer therapy. *Cellular immunology* 2018, 326:33-41.

663