

RESEARCH

Open Access



# Genetic profiling of different phenotypic subsets of breast cancer stem cells (BCSCs) in breast cancer patients

Abdel-Rahman N. Zekri<sup>1</sup>, Abeer Bahnassy<sup>2</sup>, Magda Mourad<sup>2</sup>, Ibrahim Malash<sup>3</sup>, Ola Ahmed<sup>1</sup> and Mona S. Abdellateif<sup>4\*</sup>

## Abstract

**Background:** Breast cancer stem cells (BCSCs) have a crucial role in breast carcinogenesis, development, and progression. The aim of the current study is to characterize the BCSCs through the genetic profiling of different BCSCs phenotypic subsets to determine their related genetic pathways.

**Methods:** Fresh tumor tissue samples were obtained from 31 breast cancer (BC) patients for (1) Mammosphere culture. (2) Magnetic separation of the BCSCs subsets using CD24, CD44, and CD326 Microbeads. (3) Flow cytometry (FCM) assay using CD44, CD24, and EpCAM. (4) RT-PCR profiler Arrays using stem cell (SC) panel of 84 genes for four group of cells (1) CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup> BCSCs, (2) CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> BCSCs, (3) mammospheres, and (4) normal breast tissues.

**Results:** The BCSCs (CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup>) showed significant downregulation in 13 genes and upregulation in 15, where the *CD44*, *GJB1* and *GDF3* showed the maximal expression ( $P = 0.001$ ,  $P = 0.003$  and  $P = 0.007$ ); respectively). The CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> BCSCs showed significant upregulation in 28 genes, where the *CD44*, *GDF3*, and *GJB1* showed maximal expression ( $P < 0.001$ ,  $P = 0.001$  and  $P = 0.003$ ; respectively). The mammospheres showed significant downregulation in 9 genes and a significant upregulation in 35 genes. The maximal overexpression was observed in *GJB1* and *FGF2* ( $P = 0.001$ ,  $P = 0.001$ ; respectively).

The genes which achieved significant overexpression in all SC subsets were *CD44*, *COL9A1*, *FGF1*, *FGF2*, *GDF3*, *GJA1*, *GJB1*, *GJB2*, *HSPA9*, and *KRT15*. While significant downregulation in *BMP2*, *BMP3*, *EP300*, and *KAT8*.

The genes which were differentially expressed by the mammospheres compared to the other BCSC subsets were *CCND2*, *FGF3*, *CD4*, *WNT1*, *KAT2A*, *NUMB*, *ACAN*, *COL2A1*, *TUBB3*, *ASCL2*, *FOXA2*, *ISL1*, *DTX1*, and *DVL1*.

**Conclusion:** BCSCs have specific molecular profiles that differ according to their phenotypes which could affect patients' prognosis and outcome.

**Keywords:** Stem cells, Breast cancer, CD44, CD24, EpCAM, EMT and *NOTCH*

## Background

Breast cancer (BC) is a major health problem in females worldwide, as it represents the second leading cause of cancer-related deaths in women. It had been estimated that in 2020 there were 2.261 million new cases (11.7% of all sites) and 0.685 million deaths (6.9% of all sites) from BC globally [1]. Breast cancer is a heterogeneous

\*Correspondence: mona-sayed@cu.edu.eg

<sup>4</sup> Medical Biochemistry and Molecular Biology, Cancer Biology Department, National Cancer Institute, Cairo University, Cairo, Egypt  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

disease with a multifactorial etiology. It is divided into distinct pathological subtypes including ductal, lobular, and mucinous carcinomas. Also, it has variable molecular characteristics according to estrogen receptor (ER), progesterone receptor (PR) expression, and HER2 amplification. In addition, it had been classified according to the transcriptome-based classifications into luminal and basal breast cancers [2–4].

Despite the variability of the treatment modalities and diagnostic tools available for BC patients, still there is an increased incidence of metastasis and adverse outcomes. Therefore, it is important to understand the underlying molecular mechanisms involved in the carcinogenesis and progression of BC [5].

Breast cancer stem cells (BCSCs) represent a subpopulation of tumor cells that possess the ability to self-renew, divide indefinitely, and differentiate into other types of cells according to the surrounding growth factors [6]. There is accumulating evidence proposed that BCSCs are the leading cause of cancer progression, metastasis, as well as resistance against antitumor chemo/ radio or hormonal therapy [7, 8]. The BCSCs are characterized by surface markers expression of  $CD44^+/CD24^{-/low}$ , as well as mammosphere formation [5]. The mammospheres can be developed by culturing in non-adherent non-differentiating culture conditions, which allow for the promotion of cells that are capable of survival and continuous proliferation in culture as discrete spherical clusters [9]. These mammosphere culture systems are used to identify and enrich putative BCSCs.

The CD44 is a non-kinase cell surface glycoprotein that binds to hyaluronic acid (HA) and mediates the interaction of the BCSCs and the surrounding matrix metalloprotease (MMP) and osteopontin (OPN) [10, 11]. Therefore, CD44 is important for the stemness properties of the cancer cells, as well as the regulation of cell proliferation, differentiation, and survival [12, 13].

CD24 is a glycosylphosphatidylinositol-linked cell surface glycoprotein, that inhibits chemokine receptor-4 (CXCR4), and regulates cell metastasis and proliferation [14, 15]. It is usually downregulated on the surface of BCSCs, however, its expression on BCSCs is associated with adverse outcomes in the luminal A and triple-negative BC (TNBC) subtypes [16]. The CD24 and CD326 (EpCAM) are the main surface marker expressed on the surface of the mammary stem cell (MaSCs). The MaSCs are normally present in the adult mammary gland, and they are responsible for the maintenance of the ductal architecture [17]. These cells were also identified as proliferative heterogeneous stem cells/progenitors in the luminal types of breast cancer [18]. While the  $CD24^{-/low} CD44^+$  BCSCs were more commonly enriched in the basal-like subtype and less frequent in the luminal types

[19]. The BC is characterized by a high degree of intratumor heterogeneity, as a single tumor may contain BCSCs with different phenotypes according to the molecular forms of the tumor [20, 21]. Therefore, the characterization of the BCSCs should not be relay on  $CD44^+CD24^{-/low}$  only [5]. Other markers can be used for the characterization of BCSCs including the expression of the surface markers e.g., ALDH1, Prominin-1 (CD133), and CD131, their ability to form spheroid culture, as well as the expression of different molecular markers involved in maintaining the self-renewal, differentiation and stemness properties of the BCSCs [22, 23]. These signaling pathways include Notch, Wnt/ $\beta$ -catenin, Hedgehog (Hh), TNF- $\alpha$ /NF- $\kappa$ B, transforming growth factor- $\beta$  (TGF- $\beta$ ), receptor tyrosine kinase (RTK), and Janus kinase/signal transducer and activator of transcription (JAK-STAT) pathways [24, 25].

Therefore, the aim of the current study is to characterize the BCSCs by genetic profiling of different BCSCs phenotypic subsets, and determination of their related genetic pathways. This will allow us to accurately define the possible impact of BCSCs on the development and progression of the BC, as well as their contribution to patients' response to treatment, outcomes, and survival rates. Hence it will open a new avenue for potential targeted therapy in BC patients.

## Methods

This is a retrospective cohort study included 31 patients who were histo-pathologically confirmed for BC. The study was conducted at the National Cancer Institute (NCI), Cairo University during the period from January 2019 to May 2021. All patients were subjected to full history taking, full clinical examination, complete laboratory, and radiological assessment. The normal control samples were obtained from the females who underwent reduction mammoplasty at the NCI surgical unit.

## Sample collection

Fresh tumor samples were obtained from the operation theatre in a sterile, 50 ml plastic Falcon tube containing 10 ml of Dulbecco's modified Eagle's medium (DMEM). The samples were transferred immediately to the tissue culture lab for processing. A section of the tumor was sent to the Pathology department for routine histopathology and immunohistochemistry [Estrogen receptors (ER), progesterone receptors (PR), Herceptin-2 receptors (Her-2), and Ki-67] work to confirm the diagnosis.

## Isolation of breast cancer cells

The neoplastic tissues were washed several times in Hanks Balanced Salt Solution (HBSS; Invitrogen) and minced with sterile blades into very small pieces

(0.2–0.5 mm each). The single-cell suspension was obtained by enzymatic digestion using collagenase (50–100 units/ml in HBSS; Invitrogen) according to the studied protocol. The cells were incubated for 4–18 h at 37 °C, and then filtered using a sterile stainless steel or nylon mesh. The cell suspension was washed several times by centrifugation in HBSS, and then the pellets were re-suspended in 500 µl–1 ml of Dulbecco's Modified Eagle Medium (DMEM). Using the haemocytometer, the cells were counted and divided into two parts, one part was used for mammosphere culture, and the other part was used for cellular characterization by FCM using CD44, CD24 and cytokeratin or EpCAM monoclonal antibodies.

#### **Mammosphere culture**

The mammosphere culture was performed according to the method of Dontu et al. [26] with modifications [27, 28]. Briefly, the single isolated breast cancer cells were suspended in ultra-low attachment plates at a density of  $4 \times 10^5$  viable cells/mL in primary culture and 1000 cells/mL in each passage. The cells were cultured in DMEM/Ham F-12 media (1:1) supplemented with insulin (5 mg/mL), hydrocortisone (0.5 mg/ml), and epidermal growth factor (20 ng/mL; all from Invitrogen Ltd., Paisley, Scotland). The cells were then seeded into six-well plates (2.5 ml/per plate) or T25 tissue culture flasks (5 ml per flask). The non-adherent cells were fed weekly; measured using the gridded lens. The mammospheres were enzymatically dissociated every 7 days to 2 weeks by incubation in 0.5% trypsin–EDTA solution (Invitrogen) for about 5–10 min at 37 °C, then dispersed by pipetting with a 23-gauge needle. During the mammosphere dissociation, a subset of cells from each passage was subjected to subsequent morphologic evaluation by microscopic examination. In addition to immunohistochemistry assessment using primary antibodies for pan CK, CD44, CD24, and CD133 (Abcam, UK).

#### **Characterization of breast cancer cells by flow cytometry**

A portion of the cell suspension was used for flow cytometric characterization of breast cancer stem cells using the Cell Quest program for the following conjugated antibodies: CD45-FITC (lymphocyte marker), CD24-PE (cancer stem cell & epithelial marker), CD44-FITC (cancer stem cell marker), pan cytokeratin or cytokeratin 19-PE (epithelial marker) or EpCAM-PE (epithelial marker) according to manufacturers' instructions (Becton & Dickinson, R&D, Miltenyi). Appropriate isotype controls were included in all cases to determine the areas of non-specific staining and unstained cells from each sample were also analyzed as a negative control. Accordingly, five subsets of cells were identified in each stained

cases: CD44<sup>+</sup>/CK<sup>-</sup> or EpCAM<sup>-</sup> cells, CD44<sup>+</sup>/CK<sup>+</sup> or EpCAM<sup>+</sup> cells, CD44<sup>-</sup>/CK<sup>+</sup> or EpCAM<sup>+</sup> cells, CD44<sup>+</sup>/CD24<sup>-/low</sup> cells and CD24<sup>+</sup> cells.

#### **Separation of the breast cancer stem cells (BCSCs)**

After the preparation of single-cell suspension, magnetic separation of the BCSCs subsets was done using the LS separation column (Miltenyi Biotec B.V. & Co. KG) according to the manufacturers' instructions. The cells were separated by magnetic selection after staining with CD24, CD44 and CD326 (EpCAM) Micro Beads labeled with monoclonal antibodies. Finally, different subsets of cells were collected by magnetic separation and stored for subsequent RNA extraction. Accordingly, the cells were divided into three groups including G1: CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup> cells; G2: CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> cells; and G3: mammospheres,

#### **Gene profiling array**

**RNA extraction and quantitative real-time PCR (qRT-PCR)** RNA was extracted and purified from the different groups of cells after magnetic separation using RNeasy Midi Kit (Cat. No. 74104, Qiagen) according to manufacturers' instructions. The qRT-PCR was done using the RT2 profiler array (Cat. No. 330401, Qiagen). As for the stem cells (SCs) profiling assay, the SABiosciences RT2 qPCR Master Mixes (Cat. No. 330522, Qiagen) was used to obtain the most accurate results from the PCR Array.

The PCR was performed in the MaxPro3000 real time PCR (Startagen). Regarding the 96 well plate array the following reagents were mixed in a 5-ml tube at the recommended concentrations using a multi-channel pipette: 2X SABiosciences RT2 qPCR Master Mix, diluted First Strand cDNA Synthesis Reaction (using 500 ng RNA) and H<sub>2</sub>O. The amplification cycles were formed of an initiation step at 95 °C for 10 m, followed by 40 cycles at 95 °C for 15 s and, 55 °C for 90 s. The cycle threshold (Ct) for each well was determined and the  $\Delta\Delta C_t$  method was used for data analysis by the instrument's software. The Ct values of the control wells were determined including the Ct value of genomic DNA Control (GDC) and if it was greater than 35, the level of genomic DNA contamination was considered too low to affect gene expression profiling results. The studied genes were illustrated in Table 1.

#### **Data analysis**

Data management and analysis were performed using statistical software package SPSS, version 22 (IBM, Armonk, Ny, USA). The flow cytometry data were presented as median and interquartile ranges (IQR) according to the performed normality test. Comparison between data

**Table 1** The selected genes and their functions

Function	Genes
Cell cycle regulators	<i>APC, AXIN1, CCNA2, CCND1, CCND2, CCNE1, CDK1, CDC42, EP300, FGF1, FGF2, FGF3, FGF4, MYC, NOTCH2, PARD6A, RBI</i>
Chromosome and chromatin modulators	<i>KAT2A, HDAC2, KAT8, KAT7, RB1, TERT</i>
Genes regulating symmetric/asymmetric cell division	<i>DHH, NOTCH1, NOTCH2, NUMB, PARD6A</i>
Self-renewal markers	<i>HSPA9, KAT8, KAT7, NEUROG2, SOX1, SOX2</i>
Cytokines and growth factors	<i>BMP1, BMP2, BMP3, CXCL12, FGF1, FGF2, FGF3, FGF4, GDF2, GDF3, IGF1, JAG1</i>
Genes regulating cell–cell communication	<i>DHH, DLL1, GJA1, GJB1, GJB2, JAG1</i>
Cell adhesion molecules	<i>APC, BGLAP, CD4, CD44, CDH1, CDH2, COL9A1, CTNNA1, CXCL12, NCAM1</i>
Metabolic markers	<i>ABCG2, ALDH1A1, ALDH2, FGFR1</i>
Stem cell differentiation markers	Embryonic Cell Lineage markers: <i>ACTC1, ASCL2, FOXA2, PDX1 (IPF1), ISL1, KRT15, MSX1, MYOD1, T</i> Hematopoietic Cell Lineage Markers: <i>CD3D, CD4, CD8A, CD8B, MME</i> Mesenchymal Cell Lineage Markers: <i>ACAN (AGC1), ALP1, BGLAP, COL1A1, COL2A1, COL9A1, PPARG</i> Neural Cell Lineage Markers: <i>CD44, NCAM1, SIGMAR1, S100B, TUBB3</i>
Signaling pathways-portal or stem cell maintenance	Notch Pathway: <i>DLL1, DLL3, DTX1, DTX2, DVL1, EP300, KAT2A, HDAC2, JAG1, NOTCH1, NOTCH2, NUMB</i> Wnt Pathway: <i>ADAR, APC, AXIN1, BTRC, CCND1, FRAT1, FZD1, MYC, PPARD, WNT1</i>

were analyzed using Mann–Whitney. The PCR Array Data Analysis Web Portal presents the results in a tabular format, a scatter plot, a three-dimensional profile, and a volcano plot. All tests of hypotheses were performed at the alpha level of 0.05, with a 95% confidence interval.

## Results

### Patients' characteristics

The median age of the recruited BC females was 47 (range; 22–68) years, and the mean age was  $48.1 \pm 11.4$  years. There were 15 (48.4%) patients with grade 2 tumor, and 16 (51.6%) patients with grade 3. Lymph node (LN) involvement was encountered in 22 (71%) patients, and capsular invasion was detected in 18 (58.1%) BC patients. The ER and PR were expressed in the tumor tissue of 14 (45.2%) BC females, while HER2 was expressed in the tumor tissue of 7 (22.6%) patients. In addition, there were 8 (25.8%) patients positive for distant metastasis (Table 2).

### Mammosphere culture of the primary breast cancer cells

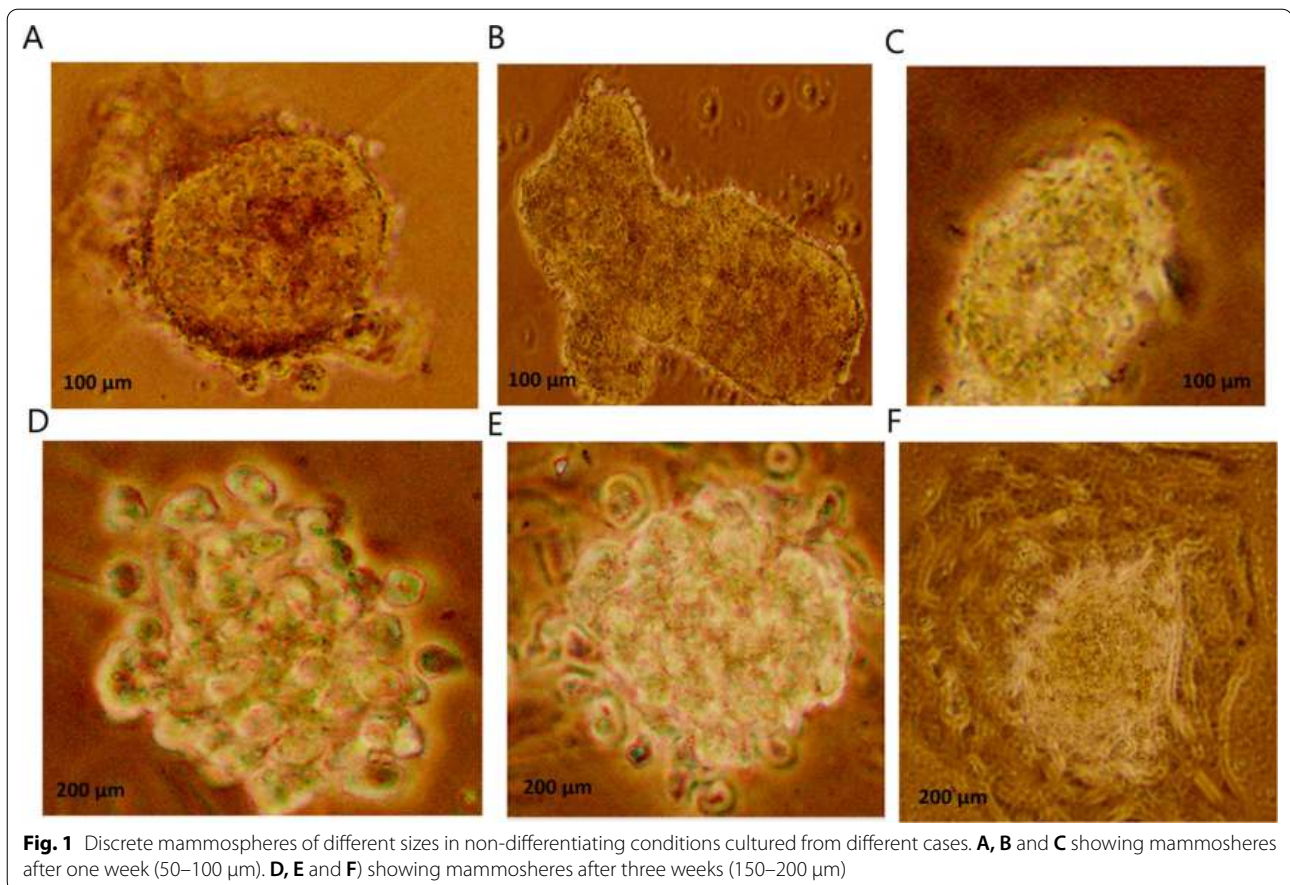
The mammospheres formed of cells that are capable of surviving and proliferating as discrete clusters in non-adherent, non-differentiating culture conditions. Such spheroids, which are enriched in progenitor cells capable of differentiating along multiple lineages. The size of the mammospheres depends upon the proliferation of the cells which varied according to the severity and the aggressiveness of the disease.

Viable mammospheres were produced in 13 out of the 31 cases included in the study (cultures were feasible for 15 cases only). In these cases, the mammospheres ranged

**Table 2** Clinico-pathological characteristics of the assessed breast cancer patients

		Frequency (N = 31)	Percent (%)
Age		$48.1 \pm 11.4$ 47 (22–68)	
Grade	2	15	48.4
	3	16	51.6
Tumor type	IDC	26	83.9
	ILC	3	9.7
	IDC + signet ring	1	3.2
	IDC + IBC	1	3.2
LN	Negative	9	29.0
	Positive	22	71.0
Distant metastasis	Negative	23	74.2
	Positive	8	25.8
ER	Negative	17	54.8
	Positive	14	45.2
PR	Negative	17	54.8
	Positive	14	45.2
Her2	Negative	24	77.4
	Positive	7	22.6
DCIS	Negative	22	71.0
	Positive	9	29.0
Capsular invasion	Negative	13	41.9
	Positive	18	58.1

in size from 20 to 180  $\mu\text{m}$  and were successfully cultured past the third passage (Fig. 1). Immunohistochemistry using lineage markers were performed to cell blocks



**Fig. 1** Discrete mammospheres of different sizes in non-differentiating conditions cultured from different cases. **A, B** and **C** showing mammospheres after one week (50–100 μm). **D, E** and **F** showing mammospheres after three weeks (150–200 μm)

obtained from mammospheres after the second passage as well as after full differentiation.

#### Characterization of breast cancer stem cells by flow cytometry

The  $CD44^{+}/CD24^{low/-}$  population has been isolated from the primary breast tumors. These cells were found to be enriched with tumorigenic cells. Though, cells expressing CD44 have been mentioned as breast cancer stem cells; however, a portion of neoplastic breast epithelium is expressing CD44 too. Therefore, we used a panel of markers that enabled us to identify and enumerate  $CD44^{+}/CD24^{low/-}$  cells (tumorigenic stem cells),  $CD44^{+}/CK^{-}$  or  $CD44^{+}/EpCAM^{-}$  cells (stem cells),  $CD44^{+}/CK^{+}$  or  $CD44^{+}/EpCAM^{+}$  cells (neoplastic breast epithelium).

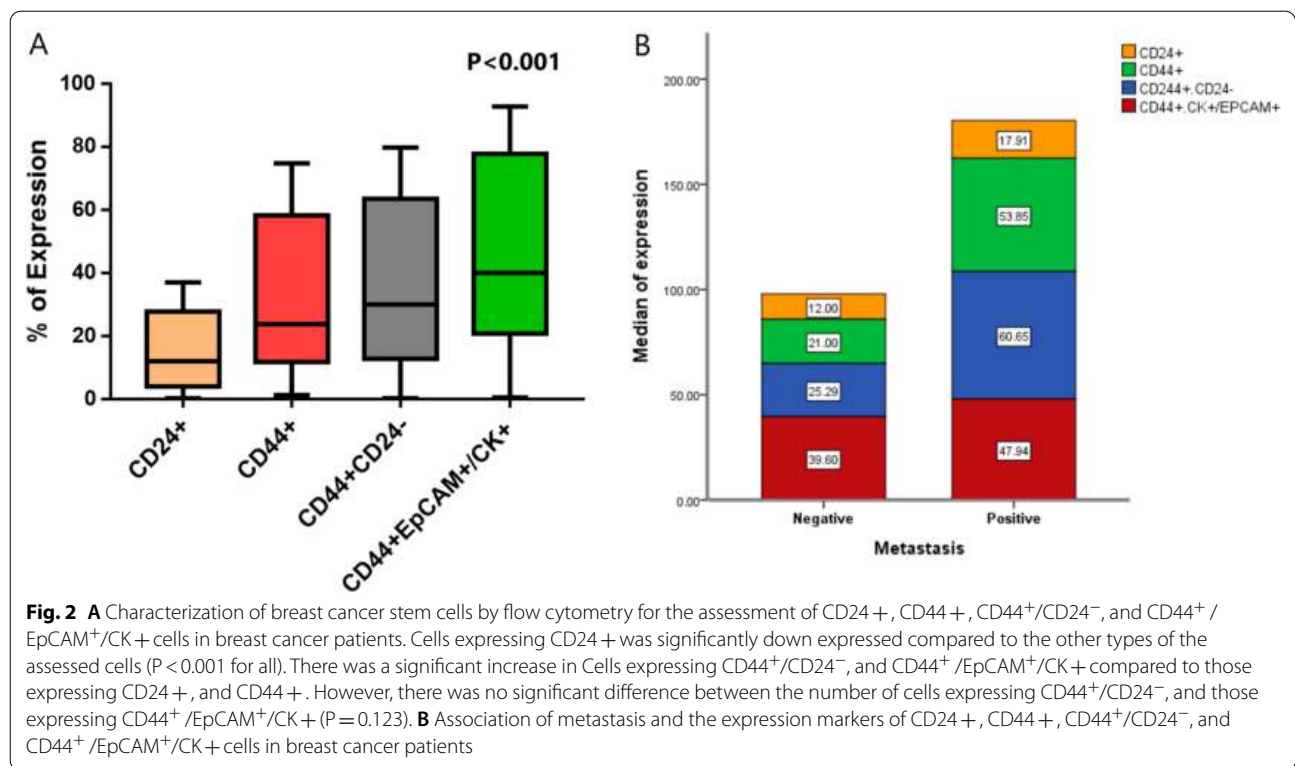
The neoplastic breast epithelium which expressed both CD44 and EpCAM or CD44 and CK was found to be 39.9% (range: 0.5–92.8%). Cells expressed CD44 only were 23.8% (range: 1.3–74.7%), and those expressed CD24+ only were 12% (range: 0.1–37%). While cells expressed  $CD44^{+}/CD24^{-}$  were 29.9% (range: 0.1–79.7%, Fig. 2).

There was a significant association between  $CD44^{+}/CD24^{low/-}$  expression and the ability of the tumor to metastasize, as the expression of  $CD44^{+}/CD24^{low/-}$  in patients with distant metastasis was 61% (range: 7–80%), compared to 25.3% (range: 0.1–68%) in those who had not metastasize ( $P=0.038$ ). Notably, patients with increased CD44+ expression, showed increase incidence of metastasis, ER and PR expression, though it did not reach a significant level ( $P=0.071$ , 0.059 and 0.059; respectively, Table 3).

#### Data of the profiling array

The profiler Arrays was performed using SC panel of 84 genes which included SC specific markers, SC differentiation markers, and signaling pathway markers for SC (Fig. 3). The profiling genes were assessed in four groups including G1: BCSCs ( $CD44^{+}/CD24^{low/-}/EpCAM^{-}$ ), G2: BCSCs ( $CD44^{+}/CD24^{low/-}/EpCAM^{+}$ ), and G3: cultured mammospheres, compared to G4: the control normal group (Fig. 4).

Out of the 84 studied genes, the BCSCs ( $CD44^{+}/CD24^{-}/EpCAM^{-}$ ) showed that there were 13 significantly down-regulated genes which are involved in

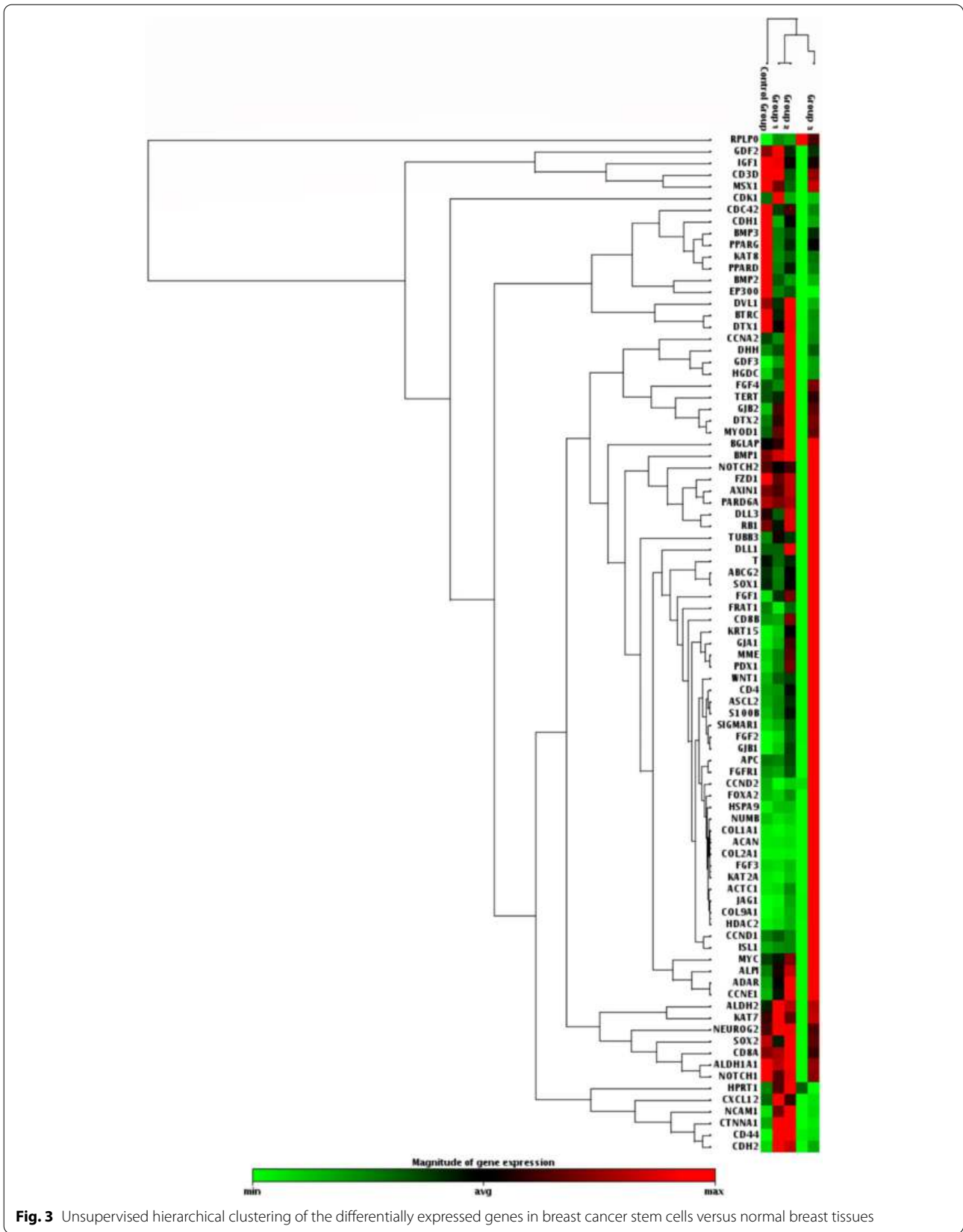


**Table 3** Association between the breast cancer cell subsets and patients' clinical features

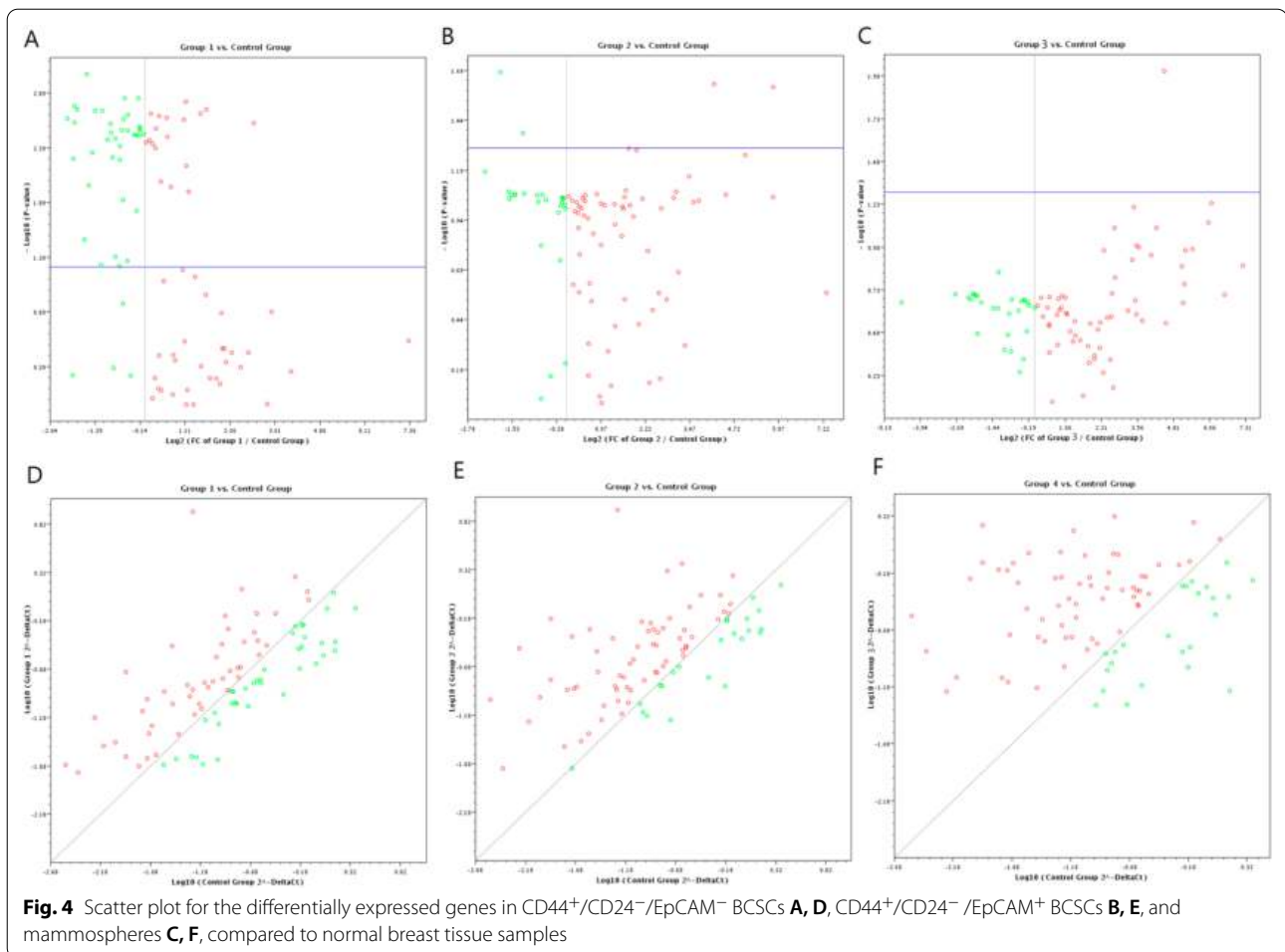
Clinical parameters	CD24 <sup>+</sup>	P value	CD44 <sup>+</sup>	P value	CD44 <sup>+</sup> CD24 <sup>-</sup>	P value	CD44 <sup>+</sup> CK <sup>+</sup> /EPCAM <sup>+</sup>	P value	
Grade	2	13.8 (0.1–37)	0.353	28 (3.6–66)	0.906	30 (0.1–68)	0.502	40 (6–88.7)	0.937
	3	11.7 (0.9–34)		22 (1.3–75)		34 (4.7–80)		41 (0.5–93)	
Tumor type	IDC	12.3 (0.9–37)	0.342	23.7 (1.3–75)	0.166	26 (2–80)	0.076	53 (0.5–93)	0.390
	ILC	12.6 (0.9–34.7)		59 (16–63)		63.5 (44.6–71)		40 (13–42)	
LN	– ve	11.5 (0.1–34)	0.273	21 (1–63)	0.507	22.5 (0.1–71)	0.654	65 (5.7–93)	0.219
	+ ve	12.5 (0.9–37)		26 (3.6–75)		33.5 (2–80)		36 (0.5–93)	
Metastasis	– ve	12 (0.1–37)	0.355	21 (1.3–66)	0.071	25.3 (0.1–68)	0.038	40 (0.5–93)	0.821
	+ ve	18 (4–34)		53.8 (6.5–75)		61 (7–80)		48 (3.5–79)	
ER	– ve	12 (0.1–34)	0.953	16.5 (1.3–66)	0.059	22.5 (0.1–71)	0.23	56 (3.5–93)	0.336
	+ ve	12 (0.9–37)		38.6 (3.6–75)		48 (2–79.7)		37.7 (0.5–87)	
PR	– ve	12 (0.1–34)	0.953	16.5 (1.3–66)	0.059	22.5 (0.1–71)	0.23	56 (3.5–93)	0.336
	+ ve	12 (0.9–37)		38.6 (3.6–75)		48 (2–79.7)		37.7 (0.5–87)	
HER2	– ve	12 (0.1–37)	1.00	28 (1.3–74.7)	0.473	33.5 (0.1–80)	0.695	41 (3.5–93)	0.872
	+ ve	12.5 (0.9–34)		23.7 (3.6–58)		30 (4–64.8)		40 (0.5–87)	
Capsular invasion	– ve	11.6 (0.1–43)	0.708	23.8 (1.3–66)	0.953	37 (0.1–71)	0.798	40 (5.7–93)	0.828
	+ ve	12.3 (0.9–37)		26 (3.6–74.7)		28 (2–79.7)		49 (0.5–93)	

Cytokines and Growth Factors (*BMP2*, *BMP3*, *CCND2*, *CDC42*, *EP300*), *Wnt* Pathway (*BTRC*, *FRAT1*, *PPARD*), Cell Adhesion Molecules (*CDH1*), Mesenchymal Cell Lineage Markers (*COL1A1*, *PPARG*), and Chromosome and Chromatin Modulators (*KAT2A*, *KAT8*). While there were 15 genes that significantly up-regulated which

were involved in cell adhesion (*CD44*, *CDH2*, *COL9A1*, *CTNNA1*, *CXCL12*, *NCAM1*), cytokines and growth factors (*FGF1*, *FGF2*, *CXCL12*, *GDF3*), cell–cell communication (*GJA1*, *GJB1*, *GJB2*), cell cycle regulation (*CDK1*), self-renewal (*HSPA9*), and embryonic Cell Lineage markers (*KRT15*). Where the *CD44* showed the maximal



**Fig. 3** Unsupervised hierarchical clustering of the differentially expressed genes in breast cancer stem cells versus normal breast tissues



expression (fold regulation: 161.48,  $P=0.001$ ), followed by *GJB1* and *GDF3* (fold regulation: 16.56 ( $P=0.003$ ) and 11.37 ( $P=0.007$ ); respectively) compared to the control group.

Regarding the CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> BCSCs, the significantly downregulated genes were involved in cytokines and growth factors (*BMP2*, *BMP3*, *EP300*), embryonic cell lineage markers (*MSX1*), chromosome and chromatin modulators (*KAT8*). While there were 28 genes that showed significant differential upregulation compared to the control group. These genes were involved in cell cycle regulation (*CCNA2*, *CCNE1*, *FGF2*, *FGF4*), cell adhesion molecules (*CD44*, *CDH1*, *CDH2*, *COL9A1*, *NCAM1*), cytokines and growth factors (*FGF1*, *FGF2*, *FGF4*, *GDF3*), cell–cell communication (*GJA1*, *GJB1*, *GJB2*, *JAG1*, *DHH*), self-renewal markers (*HSPA9*), *Wnt* pathway (*ADAR*), Notch pathway (*DTX2*, *HDAC2*, *JAG1*), mesenchymal cell lineage markers (*ALPI*), neural cell lineage markers (*NCAM1*, *S100B*, *CD44*, *SIGMAR1*), hematopoietic cell lineage markers (*CD8B*, *MME*), and embryonic cell lineage markers (*ACTC1*, *KRT15*,

*MYOD1*, *PDX1*), Where the *CD44* showed the maximal expression (fold regulation: 158.5,  $P<0.001$ ), followed by *GDF3* and *GJB1* (fold regulation: 55.9 ( $P=0.001$ ) and 55.8 ( $P=0.003$ ); respectively) compared to the control group.

Regarding the differential gene expression in the mammospheres, there were a significant downregulation in 9 genes which involved in cell cycle regulation (*CDC42*, *EP300*), chromosome and chromatin modulators (*KAT8*), *Wnt* Pathway (*PPARD*, *BTRC*), cytokines and growth factors (*BMP2*, *CXCL12*), Notch pathway (*DTX1*, *DVL1*). While a significant upregulation in 35 genes which involved in cell cycle regulation (*CCND1*, *CCND2*, *CCNE1*, *CDC42*, *FGF1*, *FGF2*, *FGF3*), cell adhesion molecules (*CD4*, *CD44*, *COL9A1*), cytokines and growth factors (*CXCL12*, *FGF1*, *FGF2*, *FGF3*, *GDF3*, *JAG1*), cell–cell communication (*GJA1*, *GJB1*, *GJB2*, *JAG1*), self-renewal markers (*HSPA9*), *Wnt* pathway (*ADAR*, *CCND1*, *WNT1*), Notch pathway (*HDAC2*, *JAG1*, *KAT2A*, *NUMB*), mesenchymal cell lineage markers (*ACAN*, *COL1A1*, *COL2A1*, *COL9A1*), neural cell lineage



**Table 4** Gene profiling expression in BCSCs and Mammospheres in relation to the normal breast tissues

	Gene symbol	BCSCs (CD44 <sup>+</sup> /CD24 <sup>-</sup> /EpCAM <sup>+</sup> )		BCSCs (CD44 <sup>+</sup> /CD24 <sup>-</sup> /EpCAM <sup>+</sup> )		Mammospheres	
		Fold Regulation	p-value	Fold Regulation	p-value	Fold Regulation	p-value
1	<a href="#">ABCG2</a>	-1.6281	0.051	1.3	660.0	2.5	0.847
2	<a href="#">ACAN</a>	1.1471	0.269	1.5	5370.	23.2	0.002
3	<a href="#">ACTC1</a>	1.3732	0.378	3.4	0.015	13.3	0.007
4	<a href="#">ADAR</a>	2.6098	0.432	5.3	0.011	5.2	0.036
5	<a href="#">ALDH1A1</a>	-1.1745	0.074	1.0	0.088	-1.3	0.648
6	<a href="#">ALDH2</a>	2.1901	0.280	1.8	0.085	1.8	0.361
7	<a href="#">ALPI</a>	2.0707	0.241	3.1	0.028	3.7	0.171
8	<a href="#">APC</a>	-1.1125	0.061	1.3	0.101	3.6	0.360
9	<a href="#">ASCL2</a>	1.4347	0.095	2.5	0.095	6.2	0.015
10	<a href="#">AXIN1</a>	-1.1082	0.059	1.2	0.660.	1.4	0.622
11	<a href="#">BGLAP</a>	1.1489	0.055	2.0	0.095	1.8	0.632
12	<a href="#">BMP1</a>	1.2193	0.055	1.4	0.109	1.4	0.467
13	<a href="#">BMP2*</a>	-3.0931	0.037	-4.9	240.0	-6.7	0.018
14	<a href="#">BMP3*</a>	-3.9218	0.033	-2.7	0.018	-2.4	0.492
15	<a href="#">BTRC</a>	-2.2991	0.012	-1.1	0.091	-4.3	0.021
16	<a href="#">CCNA2</a>	-1.4797	0.066	2.6	0.037	-1.5	0.567
17	<a href="#">CCND1</a>	1.2887	0.366	-1.0	0.243	3.8	0.028
18	<a href="#">CCND2</a>	-4.0323	0.002	-1.4	0.151	5.2	0.021
19	<a href="#">CCNE1</a>	2.8992	0.135	6.3	0.010	6.3	0.016
20	<a href="#">CD3D</a>	-1.0256	0.058	-1.5	0.091	-1.1	0.253
21	<a href="#">CD4</a>	1.1127	0.139	2.1	0.097	4.2	0.023
22	<a href="#">CD44*</a>	161.4889	0.001	158.5	P<0.001	11.1	0.009
23	<a href="#">CD8A</a>	1.0867	0.074	1.4	1530.	-1.2	0.512
24	<a href="#">CD8B</a>	-1.1373	0.055	3.4	60.03	4.8	0.036
25	<a href="#">CDC42</a>	-2.632	0.018	-1.6	660.0	-4.1	0.027
26	<a href="#">CDH1</a>	-2.7698	0.014	-1.6	0.660.	-2.7	0.493
27	<a href="#">CDH2</a>	10.5735	0.007	8.8	0.012	1.9	0.896
28	<a href="#">CDK1</a>	3.5049	0.020	-1.7	6110.	-2.4	0.907
29	<a href="#">COL1A1</a>	-3.2339	0.028	1.3	0.265	22.3	0.040
30	<a href="#">COL2A1</a>	1.56	0.327	2.0	3600.	44.2	0.003
31	<a href="#">COL9A1*</a>	3.9564	0.014	10.9	0.60.0	64.6	0.040
32	<a href="#">CTNNA1</a>	4.3456	0.016	4.1	0.382	-2.1	0.251
33	<a href="#">CXCL12</a>	3.2167	0.019	2.0	0.154	-4.4	0.029
34	<a href="#">DHH</a>	1.51	0.345	4.4	40.02	1.4	0.263

**Table 4** (continued)

35	<a href="#">DLL1</a>	1.01	0.40	2.9	0.139	3.2	0.302
36	<a href="#">DLL3</a>	-1.78	0.058	1.5	0.112	1.7	0.491
37	<a href="#">DTX1</a>	-1.91	0.051	6-1.	460.1	-4.6	0.017
38	<a href="#">DTX2</a>	2.204	0.162	3.7	0.021	2.6	0.361
39	<a href="#">DVL1</a>	-1.7788	0.057	1.3	0.171	-3.9	0.035
40	<a href="#">EP300*</a>	-3.7119	0.023	-2.9	170.0	-24.3	0.003
41	<a href="#">FGF1*</a>	4.7446	0.010	8.0	0.011	11.4	0.026
42	<a href="#">FGF2*</a>	7.3431	0.007	32.6	070.0	94.9	0.001
43	<a href="#">FGF3</a>	-1.3168	0.054	1.5	0.479	11.5	0.002
44	<a href="#">FGF4</a>	-1.4022	0.058	3.0	0.025	2.1	0.648
45	<a href="#">FGFR1</a>	-1.1833	0.053	1.6	0.293	4.8	0.022
46	<a href="#">FOXA2</a>	-1.5262	0.063	1.6	930.2	6.9	0.019
47	<a href="#">FRAT1</a>	-2.3417	0.018	1.1	0.243	3.0	0.322
48	<a href="#">FZD1</a>	-1.4016	0.057	-1.1	50.06	1.1	0.271
49	<a href="#">GDF2</a>	1.3577	0.308	-1.7	0.093	-1.8	0.459
50	<a href="#">GDF3*</a>	11.3746	0.007	55.9	10.00	10.4	0.022
51	<a href="#">GJAI*</a>	5.3162	0.018	22.4	50.00	36.4	0.012
52	<a href="#">GJB1*</a>	16.5655	0.003	55.8	0.003	147.5	0.001
53	<a href="#">GJB2*</a>	4.4462	0.012	7.0	0170.	4.6	0.043
54	<a href="#">HDAC2</a>	2.522	0.170	6.0	1600.	37.4	0.001
55	<a href="#">HSPA9*</a>	4.5628	0.014	4.8	4200.	34.9	0.001
56	<a href="#">IGF1</a>	-1.1042	0.054	-1.8	0.086	-1.7	0.271
57	<a href="#">ISL1</a>	1.3139	0.063	1.4	5400.	5.8	0.0371
58	<a href="#">JAG1</a>	2.2006	0.153	13.1	210.0	69.7	0.001
59	<a href="#">KAT2A</a>	-2.0661	0.121	2.2	0.523	18.7	0.004
60	<a href="#">KAT7</a>	1.6486	0.121	1.1	1170.	1.4	0.522
61	<a href="#">KAT8*</a>	-4.4894	0.005	-3.1	300.0	-3.6	0.037
62	<a href="#">KRT15*</a>	4.2551	0.013	17.7	0.012	34.3	0.001
63	<a href="#">MME</a>	2.2591	0.670	6.2	0880.	9.4	0.016
64	<a href="#">MSX1</a>	-1.4198	0.063	-3.1	350.0	-1.2	0.574
65	<a href="#">MYC</a>	1.2105	0.087	1.9	4530.	2.7	0.054

**Table 4** (continued)

6	<i>MYOD1</i>	2.113	0.121	3.1	0300.	2.1	0.361
6	<i>NCAM1</i>	8.1533	0.002	11.8	160.0	1.2	0.278
6	<i>NEUROG2</i>	1.5407	0.117	1.4	0.087	-1.0	0.645
6	<i>NOTCH1</i>	-1.5608	0.068	-1.1	0.089	-1.4	0.498
7	<i>NOTCH2</i>	-1.2399	0.053 1	-1.1	00.09	1.5	0.624
7	<i>NUMB</i>	-1.6512	0.350	-1.1	0890.	6.4	0.021
7	<i>PARD6A</i>	-1.1228	0.054	-1.0	0.101	1.2	0.590
7	<i>PDX1</i>	2.9293	0.352	8.4	180.0	12.2	0.001
7	<i>PPARD</i>	-3.9006	0.013	-2.3	5110.	-3.9	0.036
7	<i>PPARG</i>	-3.9828	0.030	-2.3	6110.	-1.9	0.167
7	<i>RBI</i>	-1.5377	0.064	1.3	0.126	1.4	0.493
7	<i>SI00B</i>	1.7234	0.352	3.2	250.0	6.8	0.022
7	<i>SIGMARI</i>	1.7859	0.215	3.9	120.0	10.8	0.001
7	<i>SOX1</i>	-1.6281	0.085	1.2	0.106	2.4	0.170
8	<i>SOX2</i>	-1.8858	0.121	1.2	0.094	-1.3	0.068
8	<i>T</i>	-1.5148	0.117	-1.1	0.089	2.2	0.647
8	<i>TERT</i>	1.2309	0.524	2.9	0.099	1.7	0.493
8	<i>TUBB3</i>	2.3106	0.535	1.7	0.135	4.2	0.035
8	<i>WNT1</i>	2.1246	0.554	2.6	0.390	6.6	0.022

Cells in orange color denoted up-regulated genes, cells in green color denoted downregulated genes, cells in blue color denoted differentially significant expression, p-value is significant if <0.05.

\*indicated the significantly expressed genes in the assessed three groups of cells.

markers (*SI00B*, *CD44*, *SIGMARI*, *TUBB3*), hematopoietic cell lineage markers (*CD8B*, *MME*), embryonic cell lineage markers (*ACTC1*, *ASCL2*, *FOXA2*, *ISL1*, *KRT15*, *PDX1*), and a metabolic marker (*FGFR1*). The maximal over expression was observed in *GJB1* (fold regulation: 147.5, P=0.001), followed by *FGF2*, *JAG1* and *COL9A1* (fold regulation: 94.9 (P=0.001), 69.7 (P=0.001) and 64.6 (P=0.040); respectively) compared to the control group (Table 4, Figs. 5, 6).

The differentially overexpressed markers in all BCSCs subsets compared to the control group were those mostly involved in cell cycle regulation (*CCNE1*, *EP300*, *FGF1*, *FGF2*), cell adhesion molecules (*CD44*, *CDH2*, *COL9A1*, *CTNNA1*), cytokines and growth factors (*FGF1*, *FGF2*, *GDF3*, *JAG1*), cell-cell communication (*GJA1*, *GJB1*, *GJB2*, *JAG1*), self-renewal markers (*HSPA9*), *Wnt* pathway (*ADAR*, *WNT1*), Notch pathway (*DTX2*, *HDAC2*, *JAG1*), mesenchymal cell lineage markers (*ALPI*, *COL9A1*), neural cell lineage markers (*SI00B*, *CD44*, *SIGMARI*), hematopoietic cell lineage markers (*MME*), and embryonic cell lineage markers (*KRT15*, *MYOD1*,

*PDX1*). While the most commonly down regulated genes in all studied groups compared to the control were *CDC42* (cell cycle regulation), *BMP2*, *BMP3*, (cytokines and growth factors), *KAT8* (chromosome and chromatin modulators), *PPARD*, *BTRC* (*Wnt* Pathway), *EP300* (*Notch* pathway), and *PPARG* (mesenchymal cell lineage markers).

However, the genes which achieved significant over expression in all studied SC subsets were *CD44*, *COL9A1*, *FGF1*, *FGF2*, *GDF3*, *GJA1*, *GJB1*, *GJB2*, *HSPA9*, and *KRT15*. While those which significantly downregulated were *BMP2*, *BMP3*, *EP300*, and *KAT8*.

## Discussion

Breast cancer is a heterogenous disease characterized by variable genetical and phenotypical subtypes, which accordingly leads to diverse outcomes in BC patients. This heterogeneity of the BC cells is mainly linked to the cell of origin [29]. An increasing body of evidence reported that BC could be developed from dysregulation of the mammary stem cells. In the current study

we tried to investigate the underlying molecular pathways involved in BC development and progression. We assessed three different subsets of BCSCs included (1) BCSCs (CD44<sup>+</sup>/CD24<sup>low/-</sup>/EpCAM<sup>-</sup>), (2) BCSCs (CD44<sup>+</sup>/CD24<sup>low/</sup>/EpCAM<sup>+</sup>), and (3) cultured mammospheres, through molecular profiling of 84 genes involved in the stemness properties of BCSCs, all were compared to normal breast tissue control.

The present data showed that the three subtypes of BCSCs shared over expression of common genes responsible for the self-renewal (*HSPA9*), cell cycle regulation (*CCNE1*, *EP300*, *FGF1*, *FGF2*), cell adhesion (*CD44*, *CDH2*, *COL9A1*, *CTNNA1*), cell–cell communication (*GJA1*, *GJB1*, *GJB2*, *JAG1*), expression of some cytokines and growth factors (*FGF1*, *FGF2*, *GDF3*, *JAG1*), and expression of differentiating cell lineage including mesenchymal markers (*ALPI*, *COL9A1*), neural markers (*S100B*, *CD44*, *SIGMAR1*), hematopoietic markers (*MME*), and embryonic cell lineage markers (*KRT15*, *MYOD1*, *PDX1*). In addition, they express some genes required for the stemness properties of the recruited cells included those involved in *Wnt* pathway (*ADAR*, *WNT1*) and Notch pathway (*DTX2*, *HDAC2*, *JAG1*). However, each type of cells has its own characteristically differential gene expression according to its phenotype, which later will formulate the tumor behavior and patients' outcomes. In consistent with these data, many published series illustrated the importance of *Wnt* pathway, hedgehog pathway, and notch pathway for maintain the tumorigenicity, self-renewal and epithelial-mesenchymal transition (EMT) properties of the cancer stem cells [30–32].

The current study demonstrated that the CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> BCSCs, showed significant differential upregulation in some genes when compared to the CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup> BCSCs.

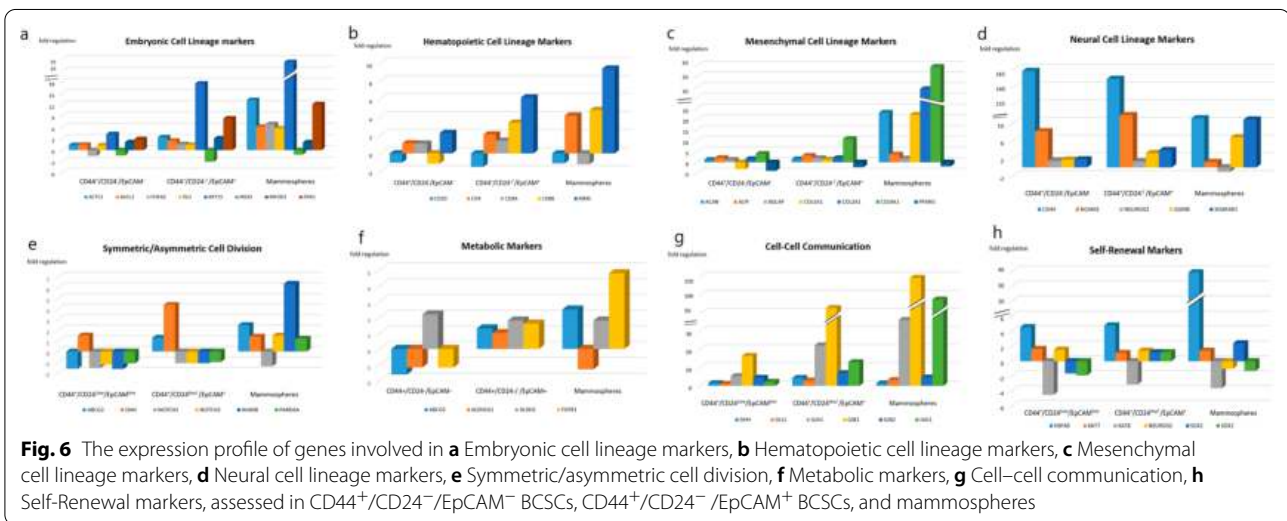
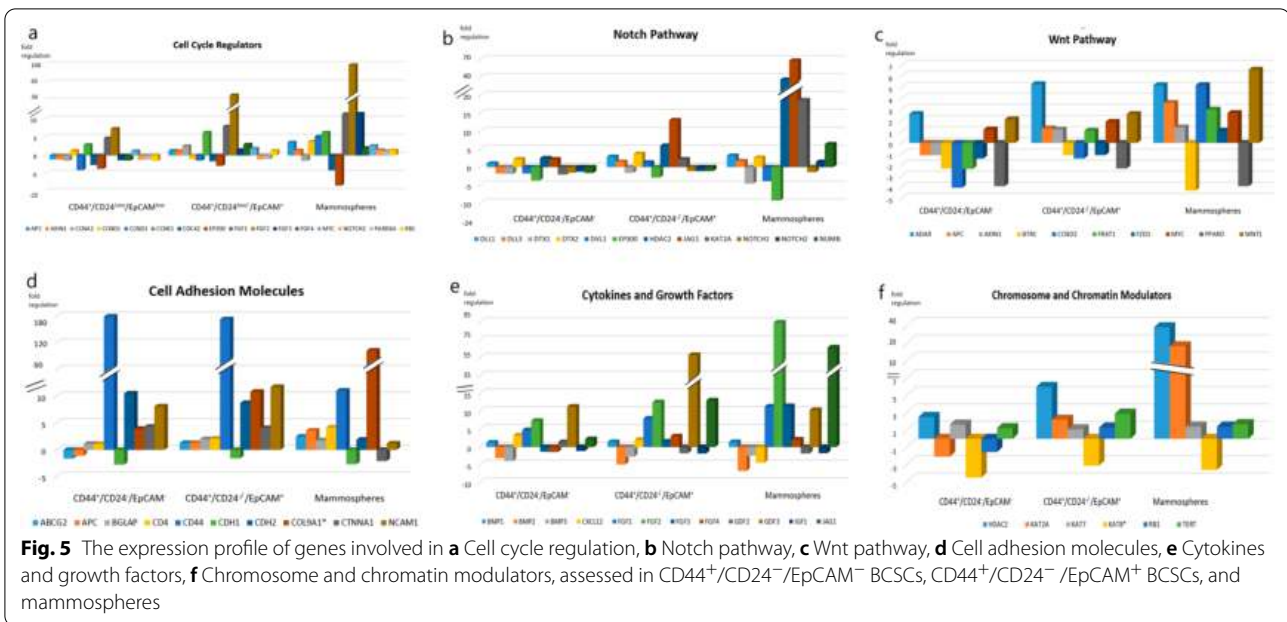
These genes were those involved in cell cycle regulation (*CCNA2*, *CCNE1*, *FGF2*, *FGF4*), cell adhesion molecules (*CDH1*, *NCAM1*), cytokines and growth factors (*FGF4*), cell–cell communication (*JAG1*, *DHH*), *Wnt* pathway (*ADAR*), *Notch* pathway (*DTX2*, *HDAC2*, *JAG1*), mesenchymal cell lineage markers (*ALPI*), neural cell lineage markers (*NCAM1*, *S100B*, *CD44*, *SIGMAR1*), hematopoietic cell lineage markers (*CD8B*, *MME*), and embryonic cell lineage markers (*ACTC1*, *MYOD1*, *PDX1*). In agreement with these data, Wu et al. [33], found that the gene expression signature associated with EMT in BCSCs were *ITGA6*, *EPCAM*, *CCND1*, *CD44*, *EGFR*, *CDH1*, and *MKI67*. Luo et al. [6], reported also that the CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> BCSCs play an important role in tumor metastasis, as it showed increased expression of genes responsible for EMT. In addition to increased expression of inflammatory cytokines and proteins associated with tumor invasion and metastasis.

Our data showed also that CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> BCSCs are more aggressive and tumorigenic than the CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup> BCSCs denoted by the differential expression of genes involved in *Wnt* and *Notch* pathway, as well as the increased expression of mesenchymal, embryonic, and neural cell lineage markers. These findings are consistent with Luo et al. [6], who reported that the BCSCs CD44<sup>+</sup>/CD24<sup>-</sup> positive for EpCAM showed increased incidence of treatment resistance and tumor recurrence. Similarly, Al-Hajj et al. [34], concluded that the BCSC (CD44<sup>+</sup>/CD24<sup>-</sup>) which express EpCAM on their cell surface, were more tumorigenic and had potent invasive properties when transplanted in immunodeficient mice in comparison to those lacking the expression of EpCAM.

Thus, the increased expression of EpCAM on the surface of BCSCs confers an aggressiveness, metastasis, drug resistance, and tumorigenic properties for the BC, which consequently associated with poorer outcomes in the BC patients.

Furthermore, both groups of BCSC subsets (CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> and CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup>) showed that the maximal expression was observed in CD44, followed by GDF3 (Growth differentiation factor-3) and GJB1 (Gap junction beta-1 protein). These data are in line with many series reported that CD44 has a fundamental role in EMT as it acts as an adhesion molecule and receptor for extracellular glycosaminoglycan hyaluronic acid, which leads to increased cell motility, and tumorigenicity [6, 35, 36]. This finding was confirmed by our results showing that CD44 expression was significantly increased in BC patients with distant metastasis.

Regarding the analysis of the gene expression profile of the cultured mammospheres, the current data demonstrated that the genes which were differentially overexpressed by the mammospheres when compared to the other BCSC subsets (CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>+</sup> and CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup>), were those involved in cell cycle regulation (*CCND2*, *FGF3*), cell adhesion molecules (*CD4*), *Wnt* pathway (*WNT1*), *Notch* pathway (*KAT2A*, *NUMB*), mesenchymal cell lineage markers (*ACAN*, *COL2A1*), neural cell lineage markers (*TUBB3*), and embryonic cell lineage markers (*ASCL2*, *FOXA2*, *ISL1*). While there was a differential downregulation in *DTX1*, *DVL1* which were involved in the *Notch* pathway. These exclusively upregulated genes in mammospheres confirmed the nature of the mammosphere derived cells which allow for the selection of highly undifferentiated and aggressive cells with marked stemness properties. In line with these data, Dontu and his colleagues performed microarray analysis of mammospheres and other differentiated BC cells. They observed a significant upregulation of genes involved in growth hormone receptors,



thrombin receptors, and Notch signaling pathway [26]. Another study done by Ramalho-Santos et al. also concluded the upregulation of Jak/Stat signaling, Notch signaling, increased transporter activity, DNA repair genes, growth hormone and thrombin receptors in mammosphere derived cells [37].

Moreover, our data revealed that the maximal overexpressed genes in the mammospheres were the *GJB1* followed by *FGF2*, *JAG1* and *COL9A1* compared to the control group. These results are consistent with many previously published studies proposed that *GJB1* and *JAG1* associated significantly with the development of distant metastasis [38], while *COL9A1* and *FGF2*

mutation associated with drug resistance in breast cancer patients [39, 40].

By comparing all studied groups, we identified 10 candidate genes which were significantly overexpressed in all BCSC subsets. These genes were important for cell cycle regulation (*FGF1*, *FGF2*), cell adhesion molecules (*CD44*, *COL9A1*), cytokines and growth factors (*FGF1*, *FGF2*, *GDF3*), cell–cell communication (*GJA1*, *GJB1*, *GJB2*), self-renewal markers (*HSPA9*), mesenchymal cell lineage markers (*COL9A1*), and embryonic cell lineage markers (*KRT15*). While those achieved significant downregulation were *BMP2*, *BMP3*, (cytokines and growth factors), *KAT8* (chromosome

and chromatin modulators), and *EP300* (Notch pathway). It had been reported that *KAT8* (K-lysine acetyltransferase 8), *BMP2*, *BMP3* (Bone morphogenetic protein), and *EP300* were significantly downregulated in breast cancer, as they have a major role in controlling tumor progression through histone acetylation, metastasis suppression and inhibition of EMT, respectively [39–41].

## Conclusions

The current study provided evidence that BCSCs have specific molecular profiles that differ according to their phenotypes, which could affect patients' prognosis and outcome. *CD44* is an important marker for characterization of BCSCs, and its co-expression with *EpCAM* provides an aggressive and metastatic properties for the BCSCs rather than those with *CD44<sup>+</sup>/CD24<sup>-</sup>/EpCAM<sup>-</sup>* phenotype. The mammospheres are the most aggressive type of breast cancer cells which exclusively had its own molecular profile including *CCND2*, *FGF3*, *CD4*, *WNT1*, *KAT2A*, *NUMB*, *ACAN*, *COL2A1*, *TUBB3*, *ASCL2*, *FOXA2*, *ISL1*, *DTX1*, and *DVL*. These genes could be considered as molecular markers for aggressiveness, metastasis, and resistance to treatment. Additionally, our data provided a panel of 14 genes (*CD44*, *COL9A1*, *FGF1*, *FGF2*, *GDF3*, *GJA1*, *GJB1*, *GJB2*, *HSPA9*, *KRT15*, *BMP2*, *BMP3*, *EP300*, and *KAT8*) which were expressed (up or downregulated) in all the assessed subsets of BCSCs which could serve as molecular markers with a potential diagnostic, prognostic and/or predictive value for breast cancer patients. However, the data of the profiling array should be validated on a larger number of patients, and also it should be correlated to the patients' clinical courses in the form of distant metastasis, response to treatment and survival rates.

## Acknowledgements

The authors of the manuscript acknowledge the National Cancer Institute and Cairo University for providing the fund and facilities required to accomplish this work.

## Author contributions

Z AN put the idea of the study protocol, supervised the molecular work and analysis of the data; B AA and M M supervised the practical part of molecular & FCM; M I was responsible for collecting the clinical data of the patients, supervised the medical care and the follow-up of the patients; A MS and Ahmed O performed the culture and molecular work of the study; A MS analyzed the data and drafting the manuscript. All authors read and approved the final manuscript.

## Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). The manuscript was funded by the National Cancer Institute (NCI), Cairo University.

## Availability of data and materials

All data generated or analyzed during this study are included in this published article.

## Declarations

### Ethics approval and consent to participate

The study protocol was approved by the Scientific Research Committee and the Institutional Review Board of the NCI, Cairo University which was in concordance to 2011 declaration of Helsinki. Written informed consents were obtained from all patients prior to enrollment in the study.

### Consent for publication

Not applicable.

### Competing interests

All authors declare that there is no possible conflict of interest.

### Author details

<sup>1</sup>Virology and Immunology Unit, Cancer Biology Department, National Cancer Institute, Cairo University, Cairo, Egypt. <sup>2</sup>Pathology Department, National Cancer Institute, Cairo University, Cairo, Egypt. <sup>3</sup>Medical Oncology Department, National Cancer Institute, Cairo University, Cairo, Egypt. <sup>4</sup>Medical Biochemistry and Molecular Biology, Cancer Biology Department, National Cancer Institute, Cairo University, Cairo, Egypt.

Received: 18 November 2022 Accepted: 24 December 2022

Published online: 31 December 2022

## References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;10:10.
- Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol*. 2005;23(29):7350–60.
- Sharpiro C, Barsky SH. Infiltrating carcinomas of the breast: not one disease. In: Bland KI, Copeland EM III, editors. *The Breast: Comprehensive management of the benign and malignant disease*. Philadelphia: Saunders Elsevier; 2009. p. 231–47.
- Sørli E, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lønning PE, Børresen-Dale AL. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*. 2001;98(19):10869–74.
- Song K, Farzaneh M. Signaling pathways governing breast cancer stem cells behavior. *Stem Cell Res Ther*. 2021;12(1):245.
- Luo M, Brooks M, Wicha MS. Epithelial-mesenchymal plasticity of breast cancer stem cells: implications for metastasis and therapeutic resistance. *Curr Pharm Des*. 2015;21(10):1301–10.
- Olivares-Urbano MA, Griñán-Lisón C, Marchal JA, Núñez MI. CSC radioresistance: a therapeutic challenge to improve radiotherapy effectiveness in cancer. *Cells*. 2020;9(7):1651.
- De Angelis ML, Francescangeli F, Zeuner A. Breast cancer stem cells as drivers of tumor chemoresistance, dormancy and relapse: new challenges and therapeutic opportunities. *Cancers (Basel)*. 2019;11(10):1569.
- Ponti D, Costa A, Zaffaroni N, Pratesi G, Petrangolini G, Coradini D, Pilotti S, Pierotti MA, Daidone MG. Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res*. 2005;65(13):5506–11.
- Yousefnia S, Seyed Forootan F, Seyed Forootan S, Nasr Esfahani MH, Gure AO, Ghaedi K. Mechanistic pathways of malignancy in breast cancer stem cells. *Front Oncol*. 2020;10:452.
- Kim SJ, Owen SC. 2020 Hyaluronic acid binding to CD44S is indiscriminate of molecular weight. *Biochimica Biophys Acta Biomembr*. 1862. <https://doi.org/10.1016/j.bbamem.2020.183348>.

12. Tan S, Yamashita A, Gao SJ, Kurisawa M. Hyaluronic acid hydrogels with defined crosslink density for the efficient enrichment of breast cancer stem cells. *Acta Biomater.* 2019;94:320–9.
13. Shan NL, Shin Y, Yang G, Furmanski P, Suh N. Breast cancer stem cells: A review of their characteristics and the agents that affect them. *Mol Carcinog.* 2021;60(2):73–100.
14. Sin WC, Lim CL. Breast cancer stem cells—from origins to targeted therapy. *Stem Cell Investig.* 2017;4:96.
15. Schabath H, Runz S, Joumaa S, Altevogt P. CD24 affects CXCR4 function in pre-B lymphocytes and breast carcinoma cells. *J Cell Sci.* 2006;119:314–25.
16. Ortiz-Montero P, Liu-Bordes WY, Londoño-Vallejo A, Vernot JP. CD24 expression and stem-associated features define tumor cell heterogeneity and tumorigenic capacities in a model of carcinogenesis. *Cancer Manag Res.* 2018;10:5767–84.
17. Rios AC, Fu NY, Lindeman GJ, Visvader JE. In situ identification of bipotent stem cells in the mammary gland. *Nature.* 2014;506:322–7.
18. Villadsen R, Fridriksdottir AJ, Rønnov-Jessen L, Gudjonsson T, Rank F, et al. Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol.* 2007;177:87–101.
19. Vikram R, Chou WC, Hung S-C, Shen C-Y. Tumorigenic and metastatic role of CD44<sup>low</sup>/CD24<sup>low</sup> cells in luminal breast cancer. *Cancers (Basel).* 2020;12:1239.
20. Xiaobin Z, Chengxiao L, Jie Y, Haoqiang W, Guoqing W, Yingpeng L, Chen N. Breast cancer stem cells, heterogeneity, targeting therapies and therapeutic implications. *Pharmacol Res.* 2021;163:105320.
21. Strietz J, Stepputtis SS, Follo M, Bronsert P, Stickeler E, Maurer J. Human primary breast cancer stem cells are characterized by epithelial-mesenchymal plasticity. *Int J Mol Sci.* 2021;22:1808.
22. Lorico A, Rappa G. Phenotypic heterogeneity of breast cancer stem cells. *J Oncol.* 2011;2011:135039.
23. Tang DG. Understanding cancer stem cell heterogeneity and plasticity. *Cell Res.* 2012;22(3):457–72.
24. Kotiyal S, Bhattacharya S. Biochemical and Biophysical Research Communications Breast cancer stem cells, EMT and therapeutic targets. *Biochem Biophys Res Commun.* 2014;453:112–6.
25. Chung SS, Giehl N, Wu Y, Vadgama JV. STAT3 activation in HER2-overexpressing breast cancer promotes epithelial-mesenchymal transition and cancer stem cell traits. *Int J Oncol.* 2014;44:403–11.
26. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, Wicha MS. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* 2003;17(10):1253–70.
27. Wang R, et al. Comparison of mammosphere formation from breast cancer cell lines and primary breast tumors. *J Thorac Dis.* 2014;6(6):829–37.
28. El Abbass KA, Abdellateif MS, Gawish AM, Zekri AN, Malash I, Bahnassy AA. The Role of Breast Cancer Stem Cells and Some Related Molecular Biomarkers in Metastatic and Nonmetastatic Breast Cancer. *Clin Breast Cancer.* 2020;20(4):e373–84.
29. Brooks MD, Burness ML, Wicha MS. Therapeutic implications of cellular heterogeneity and plasticity in breast cancer. *Cell Stem Cell.* 2015;17:260–71.
30. Li Y, Welm B, Podsypanina K, Huang S, Chamorro M, Zhang X, Rowlands T, Egeblad M, Cowin P, Werb Z, et al. Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells. *Proc Natl Acad Sci U S A.* 2003;100:15853–8.
31. Liu S, Dontu G, Mantle ID, Patel S, Ahn NS, Jackson KW, Suri P, Wicha MS. Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res.* 2006;66:6063–71.
32. Giuli MV, Giuliani E, Screpanti I, Bellavia D, Checquolo S. Notch signaling activation as a hallmark for triple-negative breast cancer subtype. *J Oncol.* 2019;2019:8707053.
33. Wu S, Zhang H, Fouladdel S, Li H, Keller E, Wicha MS, et al. Cellular, transcriptomic and isoform heterogeneity of breast cancer cell line revealed by full-length single-cell RNA sequencing. *Comput Struct Biotechnol J.* 2020;18:676–85.
34. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci.* 2003;100:39838.
35. Ponta H, Sherman L, Herrlich PA. CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol.* 2003;4(1):33–45.
36. Hiraga T, Ito S, Nakamura H. Cancer stem-like cell marker CD44 promotes bone metastases by enhancing tumorigenicity, cell motility, and hyaluronan production. *Cancer Res.* 2013;73(13):4112–22.
37. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. “Stemness”: transcriptional profiling of embryonic and adult stem cells. *Science.* 2002;298(5593):597–600.
38. Zohny SF, Zamzami MA, Al-Malki AL, Trabulsi NH. Highly Expressed DLL4 and JAG1: Their Role in Incidence of Breast Cancer Metastasis. *Arch Med Res.* 2020;51(2):145–52.
39. Kapoor-Vazirani P, Kagey JD, Vertino MP. SUV420H2-mediated H4K20 trimethylation enforces RNA polymerase II promoter-proximal pausing by blocking hMOF-dependent H4K16 acetylation. *Mol Cell Biol.* 2011;31:1594–609.
40. Cao Y, Slaney CY, Bidwell BN, Parker BS, Johnstone CN, Rautela J, Eckhardt BL, Anderson RL. BMP4 inhibits breast cancer metastasis by blocking myeloid-derived suppressor cell activity. *Cancer Res.* 2014;74:5091–102.
41. Asaduzzaman M, Constantinou S, Min H, Gallon J, Lin ML, Singh P, Raguz S, Ali S, Shousha S, Coombes RC, Lam EW, Hu Y, Yagüe E. Tumour suppressor EP300, a modulator of paclitaxel resistance and stemness, is downregulated in metaplastic breast cancer. *Breast Cancer Res Treat.* 2017;163(3):461–74.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

