Shah, Mansi and Cardenas, Ryan and Wang, Belinda and Persson, Jenny and Mongan, Nigel P. and Grabowska, Anna M. and Allegrucci, Cinzia (2017) HOXC8 regulates self-renewal, differentiation and transformation of breast cancer stem cells. Molecular Cancer, 16 (38). pp. 1-19. ISSN 1476-4598

CrossMark

HOXC8 regulates self-renewal, differentiation and transformation of breast cancer stem cells

Mansi Shah[1], Ryan Cardenas[1], Belinda Wang[1], Jenny Persson[3,4], Nigel P. Mongan[1,2], Anna Grabowska[5] and Cinzia Allegrucci[1*] iD

Abstract

Background: Homeobox genes are master regulators of cell fate during embryonic development and their expression is altered in cancer. By regulating the balance between cell proliferation and differentiation, they maintain homeostasis of normal tissues. Here, we screened the expression of homeobox genes in mammary stem cells to establish their role in stem cells transformation in breast cancer.

Methods: Using a Homeobox Genes PCR array, we screened 83 homeobox genes in normal cancer breast stem/progenitor cells isolated by flow cytometry. The candidate gene *HOXC8* epigenetic regulation was studied by DNA methylation and miRNA expression analyses. Self-renewal and differentiation of HOXC8-overexpressing or knockdown cells were assessed by flow cytometry and mammosphere, 3D matrigel and soft agar assays. Clinical relevance of in vitro findings were validated by bioinformatics analysis of patient datasets from TCGA and METABRIC studies.

Results: In this study we demonstrate altered expression of homeobox genes in breast cancer stem/progenitor cells. *HOXC8* was consistently downregulated in stem/progenitor cells of all breast molecular subtypes, thus representing an interesting tumour suppressor candidate. We show that downregulated expression of *HOXC8* is associated with DNA methylation at the gene promoter and expression of *miR196* family members. Functional studies demonstrated that HOXC8 gain of function induces a decrease in the $CD44^+/CD24^{-/low}$ cancer stem cell population and proportion of chemoresistant cells, with a concomitant increase in $CD24^+$ differentiated cells. Increased HOXC8 levels also decrease the ability of cancer cells to form mammospheres and to grow in anchorage-independent conditions. Furthermore, loss of HOXC8 in non-tumorigenic mammary epithelial cells expands the cancer stem/progenitor cells pool, increases stem cell self-renewal, prevents differentiation induced by retinoic acid and induces a transformed phenotype.

Conclusions: Taken together, our study points to an important role of homeobox genes in breast cancer stem/ progenitor cell function and establishes HOXC8 as a suppressor of stemness and transformation in the mammary gland lineage.

Keywords: Breast cancer, Cancer stem cells, Self-renewal, Differentiation, Homeobox genes, HOXC8

* Correspondence: cinzia.allegrucci@nottingham.ac.uk
[1]SVMS, University of Nottingham, Sutton Bonington Campus, Loughborough
LE12 5RD, UK
Full list of author information is available at the end of the article

Shah *et al. Molecular Cancer* (2017) 16:38

Page 2 of 19

Background

Cancer stem/progenitor cell cells (CSC) are considered to play an important role in breast cancer complexity as they retain the fundamental features of normal stem cells, being able to self-renew and differentiate into the different cell types that comprise the heterogeneous tumour mass [1, 2]. Breast CSC have a degree of cellular plasticity and it is unclear whether they originate from transformation of normal mammary stem cells (MaSC) or by reprogramming of differentiated cells to a malignant phenotype [1].

Genes that are required for tissue development and normal stem cell plasticity may contribute to the molecular blueprint of breast CSC and tumorigenesis. Indeed, it is well established that many molecular pathways involved in normal development are altered during cancer progression [1]. Amongst the developmental pathways, homeobox genes function as "master" gene regulators. They are characterised by a homeodomain that binds specific DNA elements and encode transcription factors involved in the regulation of cell growth and differentiation, thus establishing and maintaining cell identity and fate during development [3]. Given their role, it has been postulated that a balanced expression of these genes is critical for stem cell function, especially for self-renewal, differentiation and quiescence [4]. Within homeobox genes, the HOX gene family is important in early tissue differentiation. For instance, the HOXA gene cluster is silenced in pluripotent human embryonic stem cells and epigenetically reprogrammed and expressed when cells are induced to differentiate [5]. They also regulate neural and hematopoietic stem cells function [6, 7].

Homeobox genes are expressed during mammary gland development and they appear to be regulated both by hormones and extracellular matrix remodelling [8]. Examples include *Hoxc6* and *Msx2* being down-regulated and up-regulated by estrogen and progesterone, respectively. *Msx2* is also regulated by the interaction between epithelial and stromal cells in the gland [8]. Several homeobox genes are involved in proliferation and differentiation of mammary stem/progenitor cells. For instance *Prrx1, Six1, Lbx1, Sox9, Msx, Zeb1, Dlx4* have been shown to regulate epithelial-to-mesenchymal transition (EMT) in mammary cells, thus promoting stem cell-like characteristics [9–15].

Aberrant expression of homeobox genes has been reported in several malignancies [4]. Mis-expression of homeobox genes can lead to abnormal differentiation and proliferation, leading to a change in cell identity or homeotic transformation, therefore playing an important role in carcinogenesis [16]. In cancer, homeobox genes function as "tumour modulators" as their deregulation normally involve either up-regulation of genes expressed in undifferentiated cells or down-regulation of genes expressed in differentiated tissue, thus acting either as oncogenes or tumour suppressor genes [17]. Abnormal expression of homeobox genes leading to loss of differentiation is observed in breast cancer where expression of *Hoxb6, Hoxb7, Hoxc6, Hoxc8, Hoxd4, Hoxd8, Hoxd9, Hoxd10, Hoxa5* is lost in mouse mammary tumours and *Hoxa1, Hoxd3, Hoxd12, Msx1, Six1* and *Oct3* are instead expressed in cancer, but absent or lowly expressed in the normal differentiated gland [8]. This deregulated gene expression has been shown to be involved in neoplastic transformation by regulating cell cycle, apoptosis, angiogenesis, metastasis and cell adhesion [8]. Epigenetic mechanisms, such as DNA methylation, histone modification and silencing by non-coding RNA are involved in the regulation of homeobox gene expression [4]. Moreover, epigenetic silencing of these genes is an early event in breast carcinogenesis [18–20] and DNA methylation of homeobox genes is associated with specific breast cancer cell types and cancer molecular subtypes [21–23].

In this study, we screened the expression of homeobox genes in breast CSC and demonstrate that a large number of genes are differentially expressed in immortalised CSC compared to normal MaSC, in agreement with evidence suggesting an involvement of developmentally-regulated genes in early stages of cancer transformation. We also reported expression of homebox genes uniquely altered in different breast cancer subtypes, which could represent novel CSC biomarkers for patient stratification. Of the genes commonly altered across different breast cancer molecular subtypes, *HOXC8* was chosen as a putative novel tumour suppressor gene. We show that *HOXC8* downregulation in breast CSC is associated with epigenetic silencing. Downregulation of *HOXC8* in breast cancer was also found by meta-analysis of breast cancer data from large cohort studies. Gain of function of *HOXC8* reduced CSC self-renewal and the ability of cancer cells to grow in anchorage-independent conditions. Conversely, loss of function of *HOXC8* in mammary normal cells induced CSC proliferation and colony formation. In addition, reduced *HOXC8* expression impaired cell differentiation and response to retinoic acid. Taken together, our study shows that homeobox genes represent novel biomarkers of breast CSC and that *HOXC8* functions as a novel tumour suppressor gene by regulating breast CSC proliferation and differentiation.

Methods

Cells and materials

Human mammary epithelial cells (HMEC) were obtained from Invitrogen and cultured with the proprietary HuMEC Ready Medium. The immortalised human mammary epithelial cell line MCF10A (ATCC) was cultured in HuMEC Ready Medium supplemented with 100 ng/ml cholera toxin. Breast cancer cell lines MCF-7, HCC1954,

Shah *et al. Molecular Cancer* (2017) 16:38

Page 3 of 19

HCC1428 (all from ATCC), MDA-MB-468, MDA-MB-231, BT549 and Hs578T (all from the NCI-60 cell collection, CRN cell bank, University of Nottingham) were grown in RPMI medium supplemented with 10% foetal calf serum (FCS), 1% Penicillin/Streptomycin (Pen/Strep), 1% L-Glutamine, 1% sodium pyruvate, 1% non-essential amino acids (NEAA). HEK 293 T (ATCC) cells were grown in DMEM medium containing 10% FCS supplemented, 1% Pen/Strep, 1% L-Glutamine, 1% sodium pyruvate, 1% NEAA. Cell lines obtained from ATCC were used within few passages from the original stocks, whereas other cell lines were authenticated by STR profiling (Eurofins Genomics, Germany) using the Promega PowerPlex 21 PCR kit and matched against the ATCC STR database. All cell lines were tested form mycoplasma contamination using the EZ-PCR Mycoplasma Test Kit (Geneflow). All cell culture materials were from Invitrogen and chemicals from Sigma-Aldrich, unless otherwise stated.

Cell treatments
For chemoresistance assay, cells were incubated with Paclitaxel (10nM), doxorubicin (500 nM) and 5-Fluorouracil (5-FU) (500 µM) for 5 days. Viable cells were trypsinised, collected and the CD44$^+$/CD24$^{-/low}$ profile was analysed by flow cytometry.

For differentiation experiments, cells were treated with 1 µM all-*trans*-retinoic acid (ATRA) for 7 days.

Flow cytometry
Stem/progenitor cell populations were identified and isolated as CD44$^+$/CD24$^{-/low}$ by FACS (Beckman Coulter MoFlo XDP sorter). The following antibodies and isotype controls were used: FITC-anti-CD44 (eBioscience 11-0441, 1:100 dilution), PE-anti-CD24 (eBioscience 12-0247, 1:20 dilution), APC-anti-CD44 (eBioscience 17-0441, 1:167 dilution), FITC-anti-CD24 (eBioscience 11-0247, 1:20 dilution), FITC-rat isotype control (eBioscience 11-4031, 1:100 dilution), PE-mouse isotype control (eBioscience 9012-4714, 1:20 dilution), APC-rat isotype control (eBioscience 17-4031, 1:167 dilution), FITC-mouse isotype control (eBioscience 11-4714, 1:20 dilution). Briefly, 5x10^5 cells were incubated with antibodies in the dark at 4 °C for 1 h. Cells were washed and re-suspended in complete medium for analysis. For cell sorting, 10^6 cells in 100 µL of complete medium were stained as described above. Cells were then treated with DNase I (Qiagen) in RDD buffer (10 mM Tris-HCL, 2.5 mM MgCl2, 0.5 mM CaCl2, pH7.6) for at least 15 min at room temperature, and passed through a 40 µm mesh (Becton Dickinson LTD) to avoid cell clumping. Data were analysed with the Weasel software.

Homeobox gene expression array
Total RNA was extracted with the RNeasy Mini Kit (Qiagen) and cDNA transcribed with the RT^2 First Strand Kit (Qiagen). Expression of homeobox genes was analysed by real-time PCR (qRT-PCR) using the Homeobox (HOX) Genes RT^2 Profiler PCR Array (Qiagen) on a Roche LightCycler® 480 System. Real-time PCR data analysis was performed by using the RT^2 Profiler™ PCR Array Data Analysis (Qiagen) and based on the $\Delta\Delta C_T$ method with normalization of the raw data to the housekeeping gene *RPLP0* after analysis with the BestKeeper software [24].

Real-time PCR
Gene expression was analysed by qRT-PCR using TaqMan® chemistry (Applied Biosystems). Data were analysed by the $\Delta\Delta C_T$ method with normalization to the housekeeping genes *RPLP0*. *HOXC8* copy number was measured by qRT-PCR using genomic DNA (extracted with DNeasy Blood &Tissue kit, Qiagen) and LuminoCt® SYBR® Green qPCR ReadyMix. miRNAs were purified from cell pellets using the miRNeasy Mini Kit (Qiagen) and reverse transcribed with miScript II RT kit (Qiagen). qRT-PCR was performed using the LuminoCt® SYBR® Green qPCR ReadyMix. Data were analysed by the $\Delta\Delta CT$ method with normalization to the endogenous control *RNU6B*. For primers and assay used see Additional file 1: Table S1.

Bisulfite sequencing
Genomic DNA was isolated using DNeasy Tissue kit (Qiagen). Bisulfite conversion (200 ng genomic DNA) was achieved using the EZ DNA methylation kit (Zymo Research) and 3 µl of bisulfite converted DNA was used for PCR reaction using the HotStarTaq masterm mix kit (Qiagen). Primers spanning the CpG island sequence were designed using the Methprimer software (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) (Additional file 1: Table S1). Bisulfire converted and purified PCR products were directly sequenced.

Lentiviral plasmids for HOXC8 overexpression and knockdown
The HOXC8 plasmid was prepared by amplifying *HOXC8* cDNA from HMEC cells using Phusion® high-fidelity DNA polymerase PCR master mix (New England Biolabs) and cloning between EcoRI and SpeI sites of a modified pSIN-EF2-Puro plasmid (pSin-EF2-Nanog-Pur was a gift from James Thomson, Addgene plasmid # 16578). *HOXC8* and scrambled shRNA primers were constructed using previously reported sequences [25]. shRNA harpins were ligated into AgeI/EcoRI-digested pLKO-Tet-On plasmid (Tet-pLKO-puro was a gift from Dmitri Wiederschain, Addgene plasmid # 21915). Plasmids were transformed in Stabl3 cells (Invitrogen) and purified using Qiagen plasmid purification kits (Qiagen). Viral particles were produced by transfecting HEK

Shah *et al. Molecular Cancer* (2017) 16:38

Page 4 of 19

293 T cells with shRNA or pSIN plasmids, together with psPAX2 and pMD2.G plamids (gift from Didier Trono, Addgene plasmids # 12260 and 12259) using FuGENE®6 (Promega). Virus-containing supernatant was collected at 72 h post-transfection, filtered through 0.45 μm filter and added of 6 μg/mL hexadimethrine bromide (polybrene) (Sigma-Aldrich). Stable transgenic cells were selected after 72 h using Puromycin (0.75–1 μg/ml). Tet-inducible shRNA was induced by treatment with 2ug/ml doxycycline.

Western blotting

Nuclear proteins were extracted with NucBuster™ Protein Extraction Kit (Calbiochem). Extracted proteins were loaded into a 12% or 15% Acrylamide gel (20 μg/lane), separated by SDS-PAGE electrophoresis and blotted onto a PVDF membrane. Membranes were blocked with 5% skimmed milk and then probed overnight at 4 °C with a rabbit anti-HOXC8 antibody (1:1,000, Sigma H1791) in the presence of 0.1% Tween 20 and 5% milk. Membranes were stripped and the incubated with a mouse anti-Lamin A/C antibody (1: 2,000, Cell Signalling 4777) overnight at 4 °C. Peroxidase conjugated donkey anti-rabbit and sheep anti-mouse (1:10,000; GE Healthcare NA934 and NA931, respectively) antibodies were incubated for 1 h at RT. ECL prime kit (GE Healthcare) was used to detect chemiluminescence.

Mammospheres culture

Single cells were plated in ultra-low adherent flasks coated with Poly 2-hydroxyethyl methacrylate (Poly-HEMA) (Sigma-Aldrich) at a density of $2x10^4$ cells/mL. Cells were grown in mammosphere medium consisting of Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12) supplemented with 20 ng/mL epidermal growth factor (EGF, R&D Systems) 20 ng/mL basic fibroblast growth factor (bFGF) (R&D Systems), 5 μg/mL insulin, and 1% pen/strep, B27, 4 μg/mL heparin, and 5 μg/mL hydrocortisone. Cells were incubated at 37 °C and 5% CO_2 for 7 days to attain first generation mammospheres. Formed mammospheres were collected by gentle centrifugation at $200 \times g$ for 4 min, dissociated enzymatically with trypsin and mechanically by pipetting. Serial passaging to attain consequent generations of mammospheres (up to four) was done by re-plating dissociated mammosphere-derived cells at a density of $5x10^3$ cells/ml and culture for 7–10 days in mammosphere medium. For PKH26 labelled mammospheres, $2x10^6$ cells were stained with 1 μM PKH26 for 5 min, washed and then and plated in mammosphere medium for the first generation. After 7 days, mammospheres were collected, dissociated as described above and sorted to isolate the PKH26 positive[high] population. Sorted cells

were cultured again in mammosphere medium to obtain the second generation.

Soft agar assay

Cancer cells ($1\text{-}3x10^4$ cells/6 well) were seeded in 1 ml of 0.5% noble agar in complete RPMI medium overlaying 2 ml 1% agar in the same medium. MCF10A cells were instead seeded in HuMEC medium, as described above. After 2 weeks culture, cell colonies were stained with 0.05% crystal violet in 10% ethanol/PBS for 1 h and colonies ≥ of 100 μm counted under a MZ125 Leica stereomicroscope.

3D Matrigel assay

The 3D on-top Matrigel assay was performed according to Lee et al. [26]. Cells ($1.5x10^4$) were plated into a 4 well plate covered with 120 μl of growth factor reduced Matrigel (BD Biosciences) and cells allowed to attached for 30 min. Cells were then cultured in 500 μl HuMEC Ready Medium supplemented with 100 ng/ml cholera toxin and 10% Matrigel for up to 7 days, with medium changed every other day.

Bioinformatics and statistics

For the analysis of the Homeobox gene array, hierarchical cluster analysis was performed by using the software ArrayMining (http://www.arraymining.net/). Venn diagrams were generated with the online tool Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

The analysis of copy number variation in clinical datasets was performed with the software OASIS (OASIS: A Web-based Platform for Exploratory Analysis of Cancer Genome and Transcriptome data, www.oasis-genomics.org). Expression analysis of homeobox genes in TCGA (RNASeq and PAM50) and METABRIC patient datasets was performed using the UCSC Cancer Browser (https://genome-cancer.ucsc.edu/) and the OASIS portal, respectively [27, 28]. For the analysis of *HOXC8* in the TCGA dataset, clinical parameters and next-generation RNA sequencing data (RNAseqv2) were obtained from TCGA ccRCC and normal non-tumour breast tissue sample set [29, 30]. Patients were sub-divided depending on clinic-pathological parameters provided by TCGA. Normalized gene counts (gene counts rescaled according to library size) from each patient were compiled into a tab-delimited file for downstream analysis by RobinA implementation of the Bioconductor-edgeR software [31, 32]. Expression analysis of *HOXC8* in the METABRIC dataset [33, 34] was also obtained from the OASIS portal. Methylation analysis of the TCGA dataset was performed by using the UCSC Cancer Browser and the MethHC browser (http://methhc.mbc.nctu.edu.tw/php/index.php) [35].

All data are expressed as mean ± standard deviation. Comparisons between two samples were done using

Shah et al. Molecular Cancer (2017) 16:38

Page 5 of 19

Student's *t*-test. One-way or Two-way analysis of variance (ANOVA) analyses followed by Bonferroni's multiple comparison tests were used for multiple group comparisons. Statistical analyses were performed with GraphPad Prism with significance levels set at *P < 0.05, **P < 0.01, ***P < 0.001.

Results

Expression of homeobox genes in breast CSC

In this study we investigated the expression of 83 different homeobox genes involved in morphogenesis, body pattern formation, embryonic development, and cellular differentiation in normal MaSC and breast CSC. We used cell lines representing normal mammary epithelium, immortalised mammary epithelium (IMM) and the breast cancer molecular subtypes luminal (LUM: ER⁺, PR⁺, HER2⁻), HER2-enriched (HER2-E: ER⁻, PR⁻, HER2⁺) and basal/triple negative (TN: ER⁻, PR⁻, HER2⁻). $CD44^+/CD24^{-/low}$ stem cell/progenitor cells were isolated from normal and immortalised mammary cells (HMEC and MCF10A, respectively), LUM (MCF-7, HCC1428), HER2-E (MDA-MB-468, HCC1954) and TN (BT549, Hs578T and MDA-MB-231) cancer cells. Expression of the majority (77.11%) of the homeobox genes analysed was altered in CSC compared to MaSC ($CD44^+/CD24^{-/low}$ in HMEC) across different cancer subtypes (≥2-fold change, with genes considered differentially expressed only if either upregulated or down-regulated in at least 2 cell lines of the same molecular subtype).
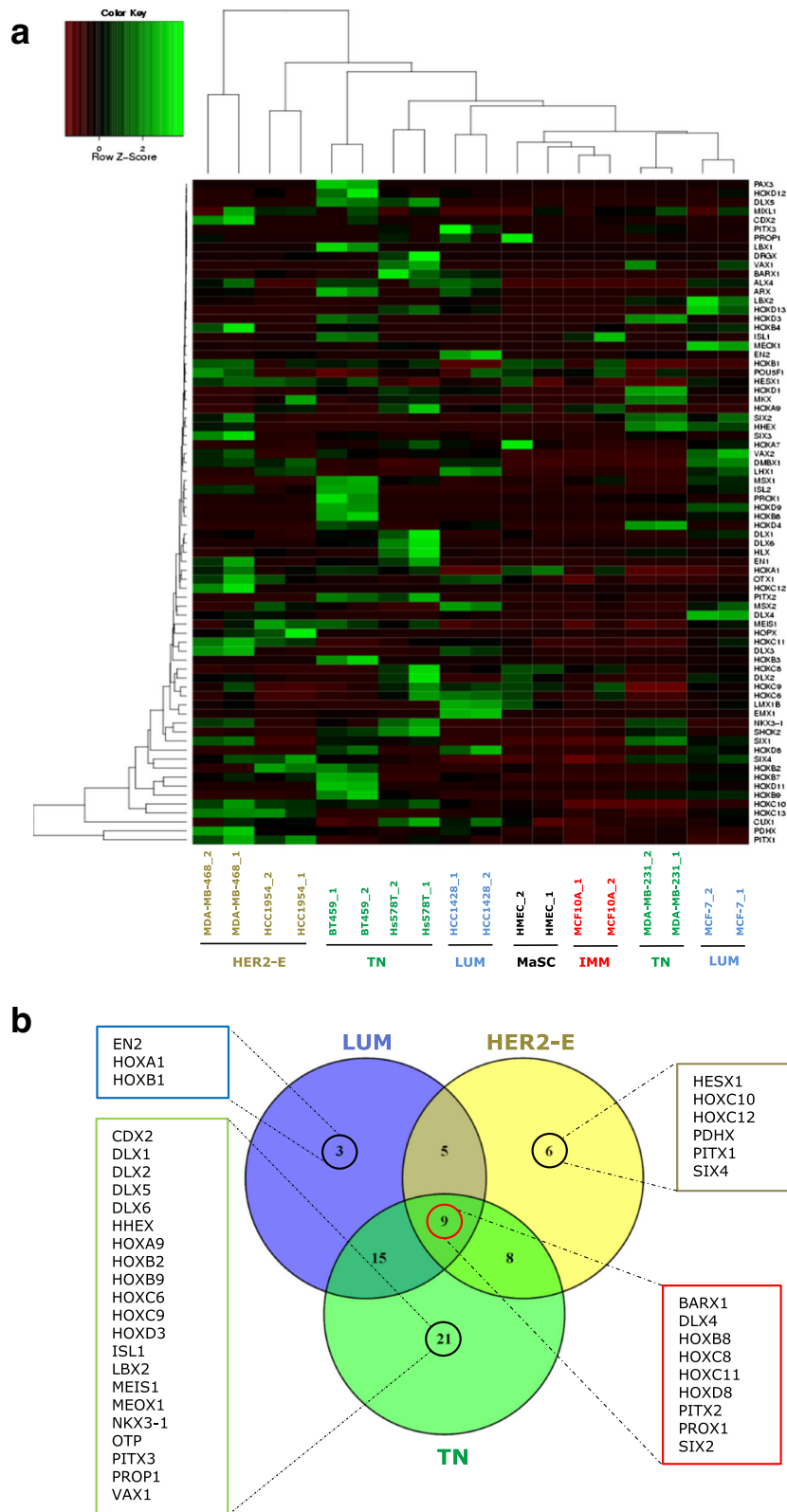
Unsupervised hierarchical cluster analysis showed that HMEC and MCF10A clustered together, showing a similar expression of homeobox genes in non-tumorigenic stem cells. Normal stem cells clustered closer to luminal CSC, whereas HER2-E and TN CSC clustered separately (Fig. 1a). Surprisingly, MDA-MB-231 cells clustered with luminal CSC and not TN CSC, suggesting that this cell line may have fundamental genetic and epigenetic differences compared to the other TN cell lines analysed.

A large number of genes were differentially expressed in immortalised stem cells, suggesting altered expression of homeobox genes is an early event during CSC transformation. The largest number of differentially expressed genes was observed in TN CSC. Differential expression of a large number of genes was consistent across cell lines of the same molecular subtypes, representing 54.8% of the genes in LUM, 41.8% in HER2-E, and 72.2% in TN. A minority of genes changed expression in either only one cell line within molecular subtypes or in different directions (either down- or upregulated). The majority of genes in the immortalised stem cells were downregulated, whereas upregulated genes were predominant in CSC across the different molecular subtypes (Additional file 2: Figure S1A). Of the downregulated genes in the immortal stem cells, the majority were shared by TN CSC, whereas

upregulated genes were shared similarly by all CSC subtypes. Interestingly, the TN CSC showed the largest number of uniquely downregulated and upregulated homeobox genes (Additional file 2: Figure S1B, Figure S1C). When considering the cancer subtype-specific homeobox genes, three genes were unique to LUM CSC, 6 to HER2-E CSC, and 21 to TN CSC. Nine genes were instead common to all cancer subtypes (considering at least 2 cell lines) (Fig. 1b). Of these, six genes were either consistently upregulated (BARX1, DLX4, HOXB8, PITX2, SIX2) or downregulated (HOXC8).

The expression of those genes that were differentially expressed either in all or specific cancer subtypes (Fig. 1b) was further validated by analysis of large cohort patient data from the publically available TCGA and METABRIC datasets [29, 30, 33, 34]. Figure 2 summarises the validation using TCGA RNASeq (all samples or samples classified according to their PAM50/subtype specific signature [30]) and METABRIC data, with genes upregulated in tumours shown as red and downregulated in tumours (upregulated in normal tissue) as green (see also Additional file 3: Figure S2). The majority of homeobox genes altered in breast CSC showed a good overlap of expression with patient tumours (76.3%; 29/38 genes matching at least one dataset), confirming the clinical relevance of our findings. Of the genes that were altered in all cancer subtypes, the expression of five matched all datasets (HOXC8, HOXC11, HOXD8, PROX1, SIX2), although the overlap was found when considering the expression in the majority of the subtypes as HOXC11 and PROX1 were downregulated in TN CSC. DLX4 matched two datasets (TCGCA and PAM50) and BARX1 and PITX1 only the METABRIC dataset. In contrast, no match was found for HOXB8 expression.

For the luminal-specific genes, HOXA1 was found downregulated in all datasets, EN2 upregulated in the METABRIC, and HOXB1 expression showed no overlap. The genes in the HER2-E subtype corresponded well with patient data, with HOXC10, HOXC12, PITX1, SIX4 matching all datasets and HESX1 and PDHX the METABRIC and TCGA sets, respectively. Finally, In the TN-specific gene set, DLX5, HHEX, HOXA9 expression corresponded to all datasets, whereas HOXB9, HOXC6, ISL1, LBX2, VAX1 only to two datasets (TCGA/PAM50 or TCGA/METABRIC or METABRIC/PAM50). Interestingly, of the genes that overlapped with only one dataset (CDX2, DLX1, DLX6, HOXB2, OTP, PROP1), three matched with their PAM50 subtype (CDX2, DLX6 and HOXB2). However, the expression of a number of gene did not correspond to any dataset (DLX2, HOXC9, HOXD3, MEIS1, MEOX1, NKX3-1, PITX3), suggesting that the expression of these genes might possibly be specific to the particular cell lines analysed in this study.

Shah *et al. Molecular Cancer* (2017) 16:38

Page 6 of 19

Fig. 1 (See legend on next page.)

Shah et al. Molecular Cancer (2017) 16:38

Page 7 of 19

(See figure on previous page.)

Fig. 1 Differential expression of homeobox genes in CSC compared to normal MaSC. **a** Unsupervised hierarchical clustering of 83 homeobox genes based on expression in CSC isolated by CD44+/CD24low/- cell sorting of different cell lines representing breast cancer molecular subtypes (LUM, HER2-E, TN), immortalised mammary epithelial cells (IMM) and normal mammary cells. Homeobox gene expression was determined by qRT-PCR using using the Homeobox Genes RT2 Profiler PCR Array (Qiagen) with two independent biological replicates. **b** Venn diagram representing differentially expressed genes unique or in common among CSC of different breast cancer molecular subtypes compared to normal MaSC (CD44+/CD24-/low cells from HMEC). Subtype-specific and common differentially expressed genes are indicated in the highlighted boxes

HOXC8 is downregulated in breast CSC and tumour tissues

Our array data showed *HOXC8* downregulation in CSC compared to normal MaSC, suggesting a possible role as tumour suppressor gene that regulates normal function of mammary stem cells. Microarray data was validated by qRT-PCR, with HER2-E and TN CSC showing most significant downregulation (Fig. 3a). This expression profile was mirrored by the unsorted cell lines for HER2-E and TN subtypes, but not for LUM subtype (Fig. 3b). In order to extend our analysis of *HOXC8* to clinical samples, we next analysed the expression of *HOXC8* in the latest TCGA [29, 30] and METABRIC [33, 34] datasets. Analysis of RNA-Seq data of the TCGA patient tumour samples ($n = 1218$) revealed *HOXC8* expression is lower in primary breast tumours compared to non-tumour tissue, whereas no difference was observed in metastatic tumours (Fig. 3c). Significant low expression was also found in TN (ER−/PR−/HER2−) compared with ER+ tumours (ER+/PR+/HER2+ or ER+/PR+/HER2−), which is consistent with our findings in CSC (Fig. 3d). The same profile of expression between normal and tumour tissue was also shown by the analysis of the METABRIC dataset ($n = 1321$) (Fig. 3e, f).

HOXC8 downregulation in breast CSC is associated with epigenetic silencing

In order to study the regulation of *HOXC8* expression in breast CSC, we next analysed whether downregulation could be due to copy number variation or epigenetic regulation. Analysis of the TCGA [29, 30] and METABRIC [33, 34] breast cancer datasets showed *HOXC8* is rarely deleted in breast cancer and its expression does not correlate with copy number in breast cancer patients (Additional file 4: Figure S3). Similarly, cell lines analysed in this study showed no *HOXC8* deletion when compared to normal mammary epithelial cells (HMEC) or MCF10A cells (Fig. 4a).

Therefore we focused our attention on the epigenetic regulation of *HOXC8*. Expression of *miR196* has been previously linked to regulation of *HOXC8* [36] and therefore we measured the expression of two *miR196* family members in HER2-E and TN CSC, as these showed the most downregulated expression of *HOXC8*. The expression of *miR196-a* and *miR196-b*

was significantly increased in CSC isolated from MDA-MB-231, but not in CSC isolated from other cell lines (Fig. 4b). Direct bisulphite sequencing analysis of the CpG island in the *HOXC8* promoter upstream of the transcription start site revealed high levels of DNA methylation in all CSC compared to normal MaSC (Fig. 4c and Additional file 5: Figure S4). A significant increase in DNA methylation at the promoter region of *HOXC8* was also observed in the breast TCGA tumour samples compared to normal tissue (Fig. 4d), thus demonstrating an important role of epigenetic modifications in the regulation of *HOXC8* expression and the clinical relevance of the in vitro findings.

Gain of function of HOXC8 reduces CSC self-renewal and chemoresistance

We next assessed the functional role of *HOXC8* in breast CSC by inducing its expression in cancer cell lines. Constitutive overexpression of *HOXC8* in low-expressing cell lines (BT549, MDA-MB-231, MDA-MB-468) resulted in high levels of expression both at the RNA and protein level (Additional file 6: Figure S5). Induced HOXC8 expression caused a significant reduction in the CD44+/CD24-/low population (Fig. 5a). The reduction was due to an induction of differentiation reflected by an increased double positive epithelial cell fraction (CD44+/CD24+) in both BT549 and MDA-MB-468 cells, but not in MDA-MB-231. This effect on CSC fate was further validated by the increased expression of *CD24* and reduced expression of the CSC marker *ALDH1* (Fig. 5b). To explore whether the increased differentiation induced by HOXC8 would affect the resistance of cells to chemotherapy, we treated control BT549 and MDA-MB-468 cells with the standard of care anti-cancer drugs Paclitaxel, Doxorubicin and 5- FU and evaluated the effect of these drugs on the CSC population. Paclitaxel and 5-FU increased the number of CD44+/CD24-/low cells after 5 days of treatment, whereas no effect was seen with doxorubicin (Fig. 5c). Consistent with the significant decrease in CSC and increased differentiation in these cell lines, overexpression of *HOXC8* caused a significant reduction in chemoresistant cells (Fig. 5d).

Shah *et al. Molecular Cancer* (2017) 16:38

Page 8 of 19

Fig. 2 (See legend on next page.)

Shah *et al. Molecular Cancer* (2017) 16:38

Page 9 of 19

(See figure on previous page.)

Fig. 2 Data heatmaps and histograms showing expression of homeobox genes in the TCGA (RNASeq and PAM50) and METABRIC datasets. Genomic heatmaps compare expression of genes between normal and tumour samples, with *yellow* and *blue* colour intensity indicating high and low expression, respectively. The *red* and *green* colours under the genomic heatmaps indicate the statistics track showing the logarithmic plot of *P*-values for each gene using Student' *t*-test followed by Bonferroni's correction. Bars above the *line* indicates that the *red* subgroup (Tumour sample) is greater than the *green* subgroup; a bar below the *line* indicates that the *green* subgroup (Normal sample) is greater than the *red* subgroup. Bars are coloured in *red* or *green* when *P* < 0.05. The same colour code summarises the same information obtained from analysis of the METABRIC and PAM50 datasets (Additional file 3: Figure S2) and the homeobox PCR array (Fig. 1) in the tables below the heatmaps. **a**. Homeobox genes differentially expressed in CSC of all cancer subtypes. **b**. Homeobox genes differentially expressed in LUM CSC. **c**. Homeobox genes differentially expressed in HER2-E CSC. **d**. Homeobox genes differentially expressed in TN CSC

To further evaluate the effect of *HOXC8* in the regulation of CSC self-renewal, we performed serial passaging of mammospheres at clonal density. The reduced CSC population induced by HOXC8 expression resulted in a significant reduction in mammosphere formation over 4 generations for all the cell lines tested (Fig. 6a), confirming a role of HOXC8 in determining stem cell fate. As the CD44$^+$/CD24$^-$ population can identify both CSC and early progenitor cells, we also stained HOXC8-expressing cancer cells with the PKH26 dye and studied the ability of HOXC8 to regulated the more primitive and quiescent CSC population which is able to retain the dye upon cell proliferation. Therefore after labelling, the first generation mammospheres were sorted for PKH26$^{positive/high}$ CSC and new mammospheres formed. Overexpression of HOXC8 reduced the ability of quiescent PKH26$^{positive/high}$ CSC to form second generation mammospheres (Fig. 6b). Finally, HOXC8- overexpressing cells displayed a reduced growth in soft agar (Fig. 6c), demonstrating a diminished transformed phenotype.

Loss of function of HOXC8 increases stem cell self-renewal and transformation

Because of the proposed role of HOXC8 in the regulation of breast stem cell fate, we next evaluated whether HOXC8 loss of function in non-tumorigenic mammary epithelial cells would be responsible for the acquisition of stemness by differentiated cells. We therefore knocked down *HOXC8* in the immortal but not-tumorigenic MCF10A cells, as stem cells in this cell line expressed similar levels of *HOXC8* to those isolated from normal HMEC cells (Fig. 3a). *HOXC8* shRNA resulted in about 80% knockdown in gene expression and no detectable levels of HOXC8 protein (Fig. 7a). HOXC8 knockdown induced a significant increase in CD44$^+$/CD24$^{-/low}$ cells with a concomitant decrease in CD44$^+$/CD24$^+$ and CD44$^-$/CD24$^+$ cells, confirming its role in maintaining a differentiated state in the mammary epithelial lineage (Fig. 7b). HOXC8 downregulation also increased the number of mammospheres over several generations (Fig. 7c), indicating that silencing of *HOXC8* is associated with a switch of cell fate and acquisition of stem cell characteristics.

Reduced HOXC8 expression also resulted in a loss of cell organisation as cells grown in 3D matrigel appeared to lose their dome-shaped morphology and acquire a mesenchymal/basal phenotype (Fig. 7d). Importantly, HOXC8 knockdown induced a ~ 2.5-fold increase in the number and size of colonies of MCF10A cells grown in anchorage-independent conditions compared to the scrambled control (Fig. 7e, f), indicating an augmented transformed phenotype.

Expression of HOXC8 is required for breast stem cell differentiation

Retinoic acid signalling and HOX genes are important regulators of embryo development and tissue differentiation [5, 37] Because our findings revealed a role of HOXC8 in the regulation of breast CSC self-renewal and differentiation, we next evaluated whether HOXC8 could be involved in the differentiation of mammary cells induced by retinoic acid. For this experiment, the MCF10A-Tet-HOXC8 shRNA cell line was treated with retinoic acid for 7 days in presence or absence of doxycyclin to induce HOXC8 knockdown. As previously indicated [38, 39], *HOXC8* expression can be regulated by retinoids and therefore treatment of mammary cells with retinoic acid induced *HOXC8* expression with concomitant increase in the proportion of CD24$^+$ cells after 4 days treatment or CD44$^-$/CD24$^-$ differentiated cells after 7 days (Fig. 8, Additional file 7: Figure S6). Importantly, HOXC8 knockdown induced by doxycycline in these cells reduced cell differentiation induced by retinoic acid and both CD24$^+$ and CD44$^-$/CD24$^-$ cell populations, suggesting a role for HOXC8 in the regulation of stem cell fate during differentiation by retinoic acid signalling in the mammary gland.

Discussion

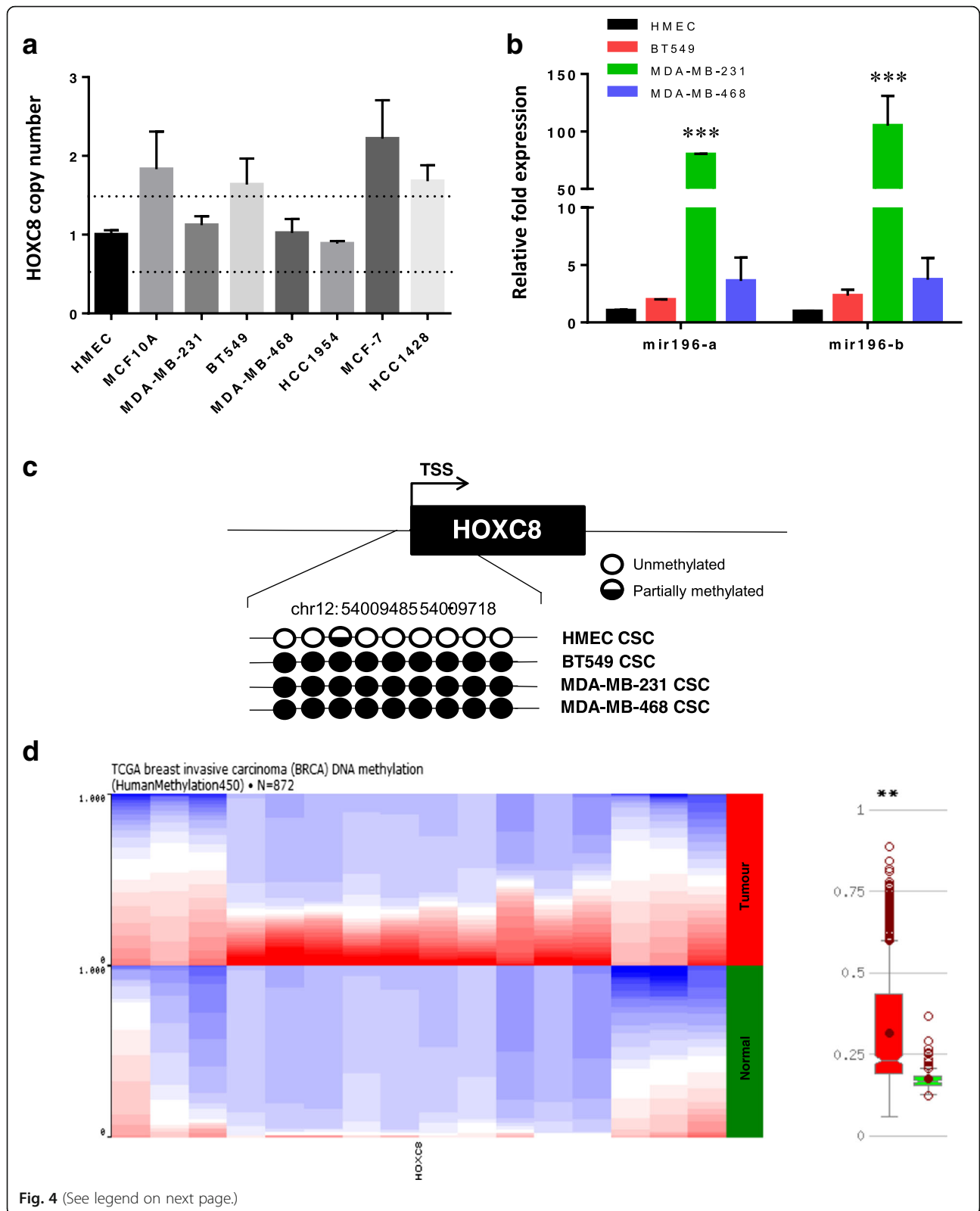The importance of developmental genes in the regulation of cancer stem cell self-renewal and plasticity is well established [1, 40]. This study investigated for the first time the expression of a panel of developmental regulatorsof the homeobox family in breast cancer stem/progenitor cells. We found that homeobox genes are deregulated at an early stage of stem cell transformation as well in different

Shah *et al. Molecular Cancer* (2017) 16:38

Page 10 of 19
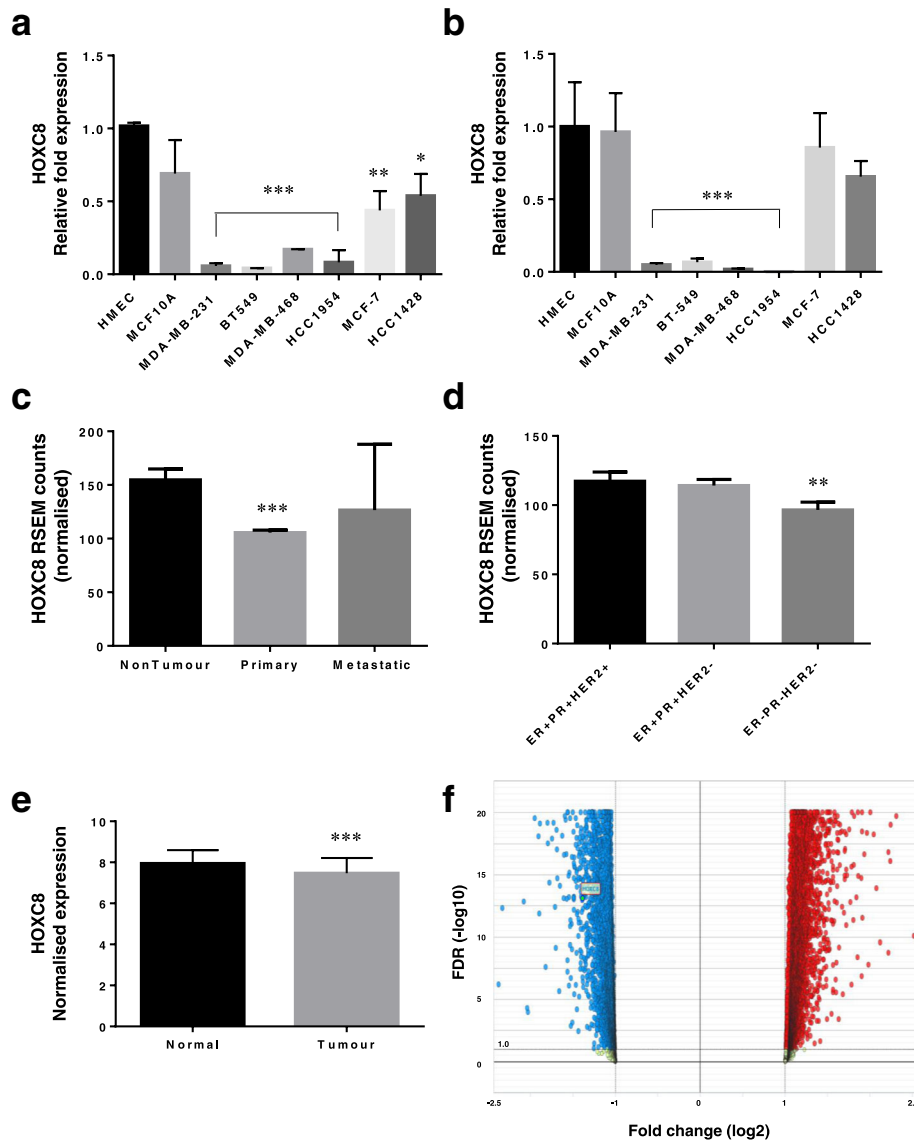
Fig. 3 *HOXC8* expression is reduced in breast CSC and breast cancer clinical samples. **a** Expression of *HOXC8* in CSC sorted as CD44$^+$/CD24$^{low/-}$ cell population as determined by TaqMan® qRT-PCR. Results are presented as relative fold expression relative to *RPLP0* and HMEC used as calibrator ($n = 3$–4). Relative fold expression levels were analysed by One-way ANOVA followed by Bonferroni's multiple comparisons test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. **b** Expression of *HOXC8* in unsorted cell lines as determined by TaqMan® qRT-PCR. Results are presented as relative fold expression relative to *RPLP0* and HMEC used as calibrator ($n = 3$–4). Relative fold expression levels were analysed by One-way ANOVA followed by Bonferroni's multiple comparisons test. ***$P < 0.001$. **c** Expression of *HOXC8* in patient tumour tissue determined by RNA-Seq in the breast TCGA dataset [29, 30]. Statistical analysis was performed by Bioconductor-edgeR. ***$P < 0.001$. **d** Expression of *HOXC8* in patient tumour tissue classified according to hormone receptor status as determined by RNA-Seq in the breast TCGA dataset. Statistical analysis was performed by Bioconductor-edgeR. **$P < 0.01$, ER$^+$PR$^+$HER2$^+$ ($n = 115$), ER$^+$PR$^+$HER2$^-$ ($n = 408$), ER$^-$PR$^-$HER2$^-$ ($n = 128$). **e** Expression of *HOXC8* in patient tumour tissue determined by microarray analysis of the breast cancer METABRIC dataset access through the OASIS software [33, 34]. Statistical analysis was performed by Unpaired Student's *t*-test. ***$P < 0.001$. **f** Volcano plot of the METABRIC dataset showing differentially expressed genes in breast cancer patients compared to normal tissue as visualised in the OASIS software. The significant downregulated *HOXC8* expression is highlighted

cancer molecular subtypes. We found that the majority of homeobox genes analysed were differentially expressed in transformed/non tumorigenic breast stem cells, with the majority being downregulated compared to normal MaSC.

Epigenetic deregulation of homeobox genes is a frequent and early event in breast cancer. DNA methylation of the *HOX* cluster has been reported to be associated with mammary epithelial cell immortalisation, suggesting that

Shah et al. Molecular Cancer (2017) 16:38

Page 11 of 19

Fig. 4 (See legend on next page.)

Shah *et al. Molecular Cancer* (2017) 16:38

Page 12 of 19

(See figure on previous page.)
Fig. 4 *HOXC8* expression is downregulated in TN and HER2-E CSC by epigenetic silencing. **a** *HOXC8* genomic copy number in CSC sorted as CD44⁺/CD24^low/- cell population as determined by SYBR® qRT-PCR. Results are presented as fold difference relative to HMEC used as calibrator ($n = 4$). *Dotted lines* represent the range of normal copy number 1 ± 95% confidence interval. **b** Expression of *miR196-a* and *miR196-b* in CSC sorted as CD44⁺/CD24^low/- cell population as determined by SYBR® qRT-PCR. Results are presented as relative fold expression relative to *RNU6B* and HMEC used as calibrator ($n = 3$). Relative fold expression levels were analysed by One-way ANOVA followed by Bonferroni's multiple comparisons test. ***$P < 0.001$. **c** DNA methylation profile of *HOXC8* promoter CpG Island. DNA methylation was analysed by direct PCR bisulfite sequencing. *Black circles* indicated methylated CG dinucleotides, *white circles* indicate unmethylated CG dinucleotides, *white/black* semicircles indicate partially methyalted CG dinucleotides. **d** DNA methylation profile of *HOXC8* promoter of TCGA clinical samples profiled by Illumina Infinium HumanMethylation 450 BeadChip array ($n = 872$). *Left panel* shows DNA methylation normalised proportions in normal and tumour samples as visualised in UCSC Cancer Browser (*blue* colour represents lower methylation, *red* colour represents higher methylation). *Right panel* shows DNA methylation levels in tumour samples (*red*) compared to normal (*green*) as visualised by MethHC browser. Statistical analysis was performed by Student's *t*-test. **$P < 0.01$

modulation of early developmental genes plays a role in escaping cell senescence [19]. In addition, alterations in DNA methylation of several homeobox genes is found in ductal carcinoma in situ (DCIS), and low-grade breast tumours [18] [41], with those found in DCIS being associated with disease progression [42]. These findings suggest a model whereby stem cells could be the cell of origin of breast tumours, as developmental genes that are normally reversibly silenced by polycomb complexes in these cells [43] become permanently silenced via DNA methylation during carcinogenesis [40, 44, 45]. Our data support this hypothesis as the majority of downregulated homeobox genes in immortal stem cells were also found not to be expressed or expressed a low level in TN breast cancer, a molecular subtype particularly enriched in CSC. A number of homeobox genes were differentially expressed only in specific breast molecular subtypes. Particularly important is the finding that the highest number of unique differentially expressed homeobox genes were found in TN CSC as this molecular subtype of breast cancer is still poorly understood and does not respond to current conventional or targeted therapies [46, 47]. DNA methylation profile of different molecular subtypes identified high enrichment of homebox genes being hypermethylated specifically in ER⁻ which were associated with high rate of tumour recurrence and therefore likely contributing to poor patient outcome [48]. Expression of homeobox genes have been also found to be specific of TN breast cancer subtypes, with the MSL (mesenchymal-stem-like) subtype being particularly enriched [49]. Therefore these newly identified CSC homeobox genes could represent novel biomarkers that could be used for patient stratification in precision medicine and warrant further investigation.

Among the differentially expressed genes identified in this study, *HOXC8* was found consistently downregulated in CSC and therefore represented a novel putative tumour suppressor candidate. *HOXC8* was especially downregulated in TN and HER2-E CSC and it was also found to be downregulated in clinical samples of TCGA and METAB-RIC datasets, strengthening its significance as a potential biomarker. In line with previous studies reporting the importance of epigenetic regulation of homeobox genes during development and carcinogenesis [17], we also found that *HOXC8* silencing is associated with the expression of *miR196* and DNA methylation at the gene promoter region in CSC. Regulation of *HOXC8* via miRNA 196 has been reported in breast cancer whereby the ration of *miR196* and *HOXC8* correlates with cell migration and metastasis [36] and methylation of the *HOXC8* promoter can induce gene silencing by inducing polymerase II stalling during transcription [50].

Our results demonstrate that HOXC8 is an important regulator of stem cell self-renewal and that HOXC8 can act as modulator of cell differentiation in the breast. Using a combination of strategies and cell culture models, we show that silencing of *HOXC8* confers mammary cells with stemness potential and this can be reversed by its re-expression with induction of differentiation. The enhanced CSC population induced by downregulation of HOXC8 in mammary cells is also associated with an increase in transformed phenotype and this observation supports the notion that breast cell tumorigenicity is enhanced through acquisition of an undifferentiated state [51]. HOX proteins are transcription factors that regulate cell proliferation and differentiation during embryonic development. In normal adult tissues, they maintain tissue homeostasis by defining cell fate and lineage commitment [4, 16, 17]. Several cancers demonstrate altered expression of HOX genes but their role in transformation of stem cells and establishment of CSC is not well-known. Recently, a role in CSC regulation has been strongly established for HOXA5 in breast and colorectal cancer, whereby HOXA5 can maintain homeostasis by suppressing stemness in these tissues [52, 53].

During development, *HOXC8* is expressed in the neural tube and somatic mesoderm and its expression is essential for skeletal development [54, 55]. *HOXC8* is also expressed in the normal mammary gland [8, 56, 57] and it participates in the initiation of mammary morphogenesis [58]. Previous studies have reported an increased

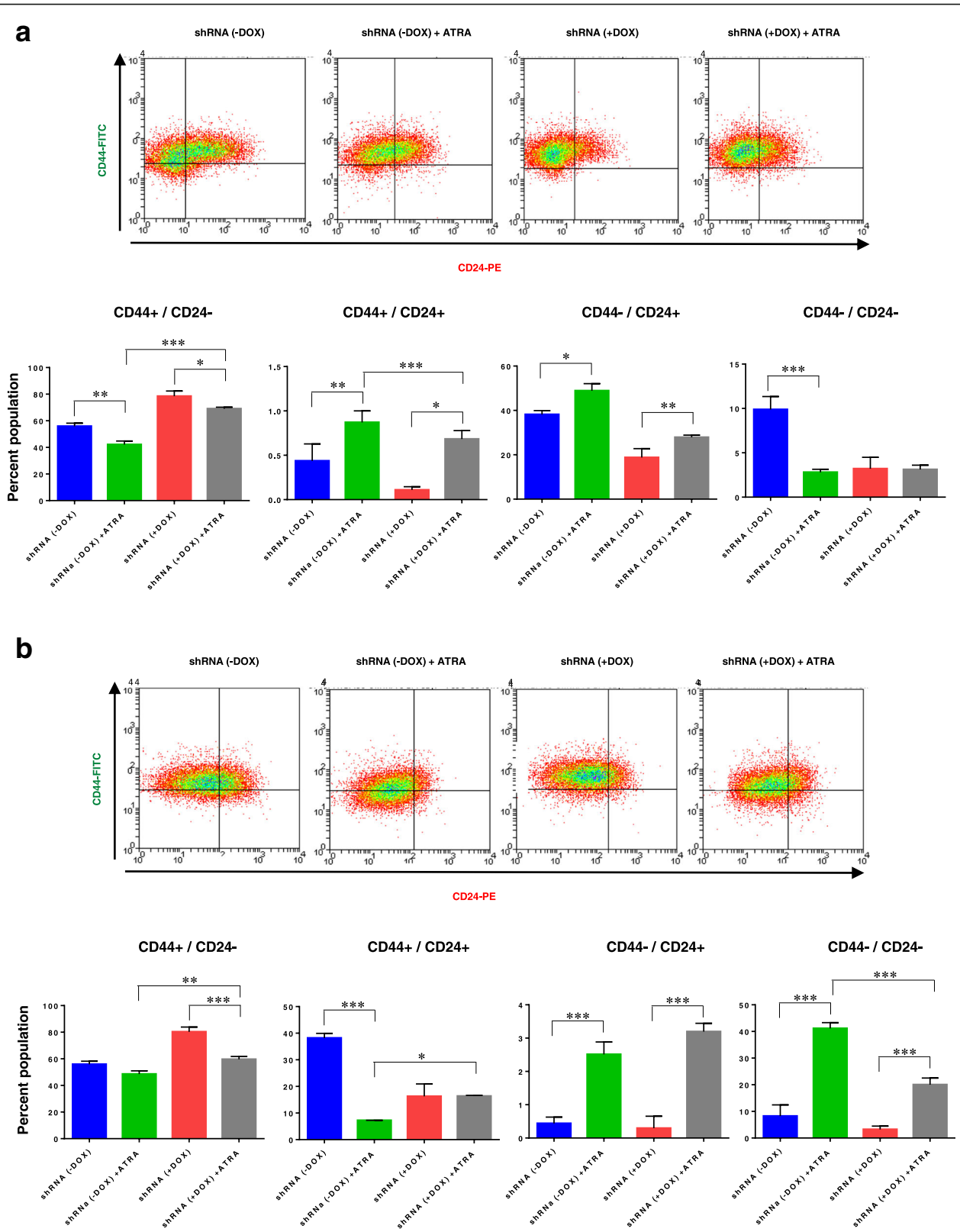Shah *et al. Molecular Cancer* (2017) 16:38

Page 13 of 19
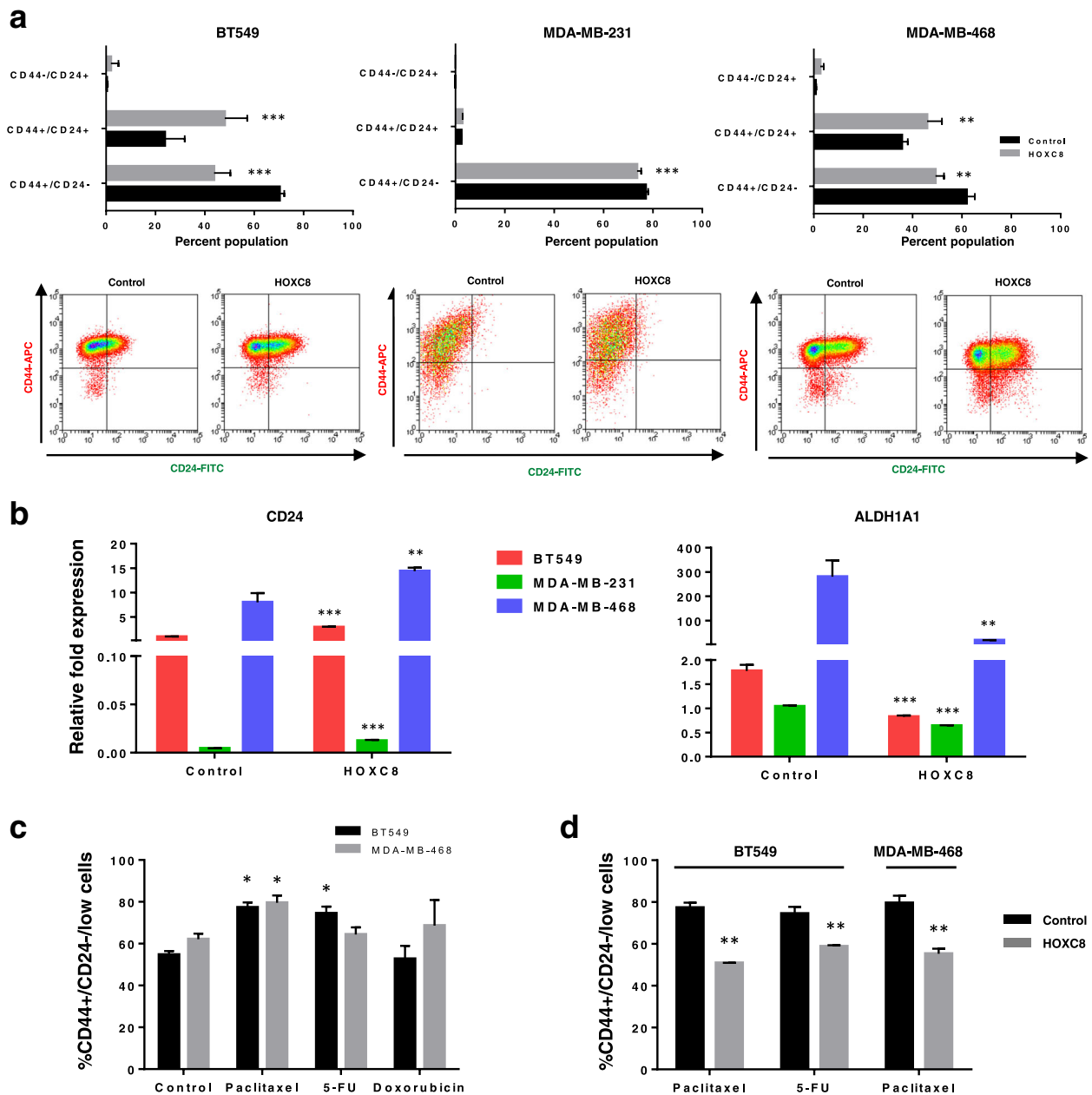
Fig. 5 HOXC8 expression negatively regulates CSC and chemoresistance. **a** Representative FACS profiles of cells transduced with pSIN-HOXC8 or pSIN empty control vector and quantification of CSC and non-CSC populations stained with CD44-APC and CD244-FITC. Results were analysed by comparing levels of CD44$^+$/CD24$^-$, CD44$^+$/CD24$^+$, CD44$^-$/CD24$^+$ cells in pSIN-HOXC8 transduced cells compared to control vector for each cell line ($n = 3$). Statistical analysis was performed by Two-way ANOVA followed by Bonferroni's multiple comparisons test. **$P < 0.01$, ***$P < 0.001$. **b** Expression of *CD24* and *ALDH1A1* in Control and HOXC8-overexpressing cells. Gene expression was measured by TaqMan® qRT-PCR. Results are presented as relative fold expression relative to *RPLP0* and control (pSIN vector) used as calibrator ($n = 3$). Relative fold expression levels were analysed by Unpaired Student's *t*-test. **$P < 0.01$, ***$P < 0.001$. **c** Percentage of CD44$^+$/CD24$^{low/-}$ cells in cell lines treated with 10nM Paclitaxel, 500 μM 5-FU and 500nM Doxorubicin for 5 days as measured by FACS analysis. Results were analysed by comparing levels of CD44$^+$/CD24$^{low/-}$ cells in untreated versus treated cells by One-way ANOVA followed by Bonferroni's multiple comparisons test for each cell line ($n = 3$). *$P < 0.05$. **d** Percentage of CD44$^+$/CD24$^{low/-}$ cells in cell lines treated with 10nM Paclitaxel, 500 μM 5-FU and 500nM Doxorubicin for 5 days as measured by FACS analysis. Results were analysed by comparing levels of CD44$^+$/CD24$^{low/-}$ cells in pSIN-HOXC8 or pSIN empty control vector transduced cells by Unpaired Student's *t*-test for each cell line ($n = 3$). **$P < 0.01$

Shah *et al. Molecular Cancer* (2017) 16:38

Page 14 of 19
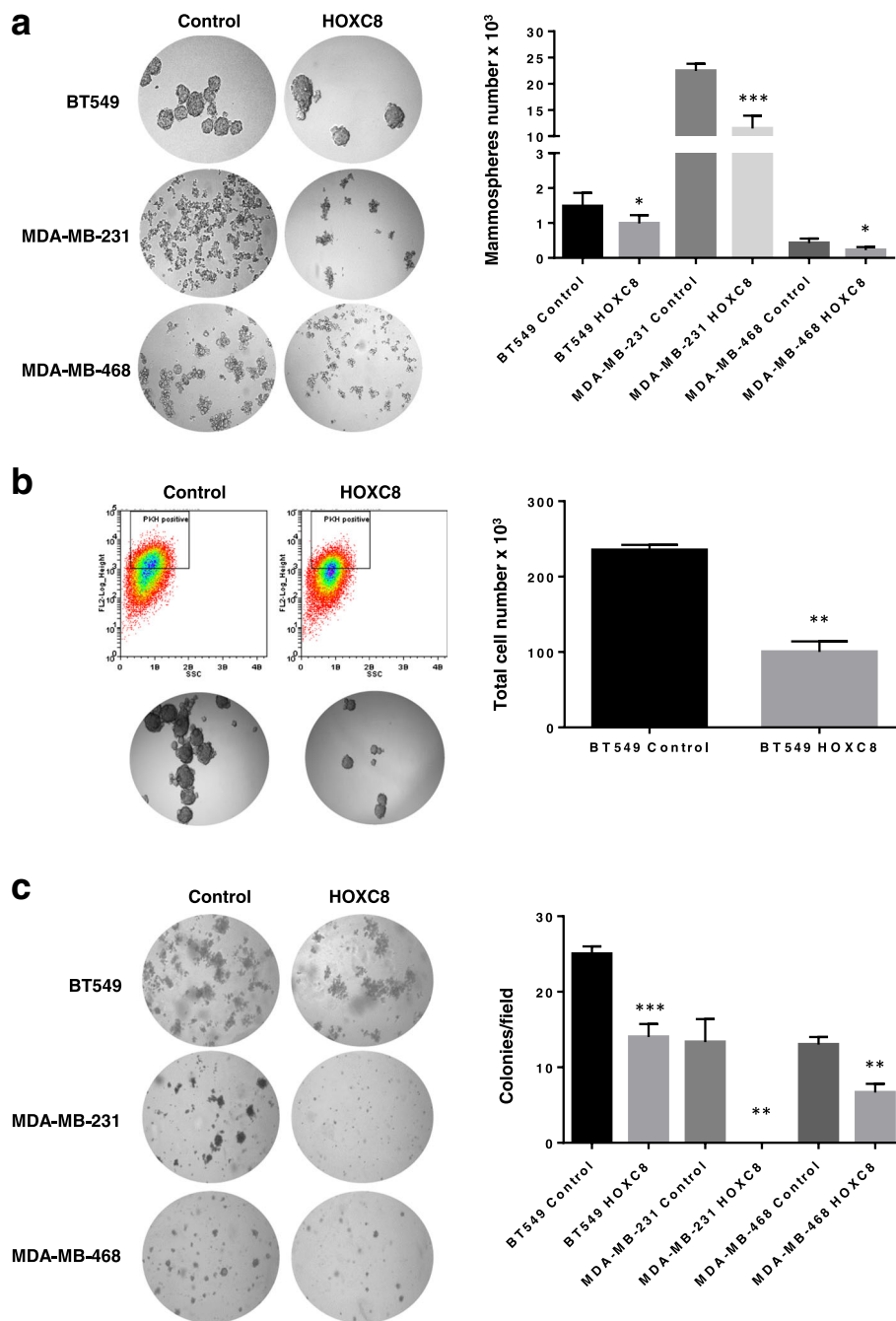
Fig. 6 HOXC8 expression reduced CSC self-renewal and anchorage independent cell growth. **a** Mammospheres morphology and number obtained by self-renewal of cell overexpressing HOXC8 or control vector after 4 generations. Mammospheres numbers were compared by Unpaired Student's *t*-test for each cell line (*n* = 4). *$P < 0.05$, ***$P < 0.001$. Only spheres ≥ 100 μM were counted, images were taken at 10X magnification. **b** Second generation mammospheres obtained after labelling with the PKH26 dye and sorting of the PKH26 $^{positive/high}$ population from first generation mammospheres. Results show the morphology of mammospheres overexpressing HOXC8 or control vector and total mammospheres cell number. Mammospheres numbers were compared by Unaired Student's *t*-test (*n* = 3). **$P < 0.01$. Only spheres ≥ 100 μM were counted, images were taken at 10X magnification. **c** Colonies of cells overexpressing HOXC8 or control vector grown in soft agar for 2 weeks. The number of colonies were calculated by counting 10 fields of view. Results were analysed by Unpaired Student's *t*-test for each cell line (*n* = 3). **$P < 0.01$, ***$P < 0.001$. Images were taken at 10X magnification

Shah et al. Molecular Cancer (2017) 16:38

Page 15 of 19
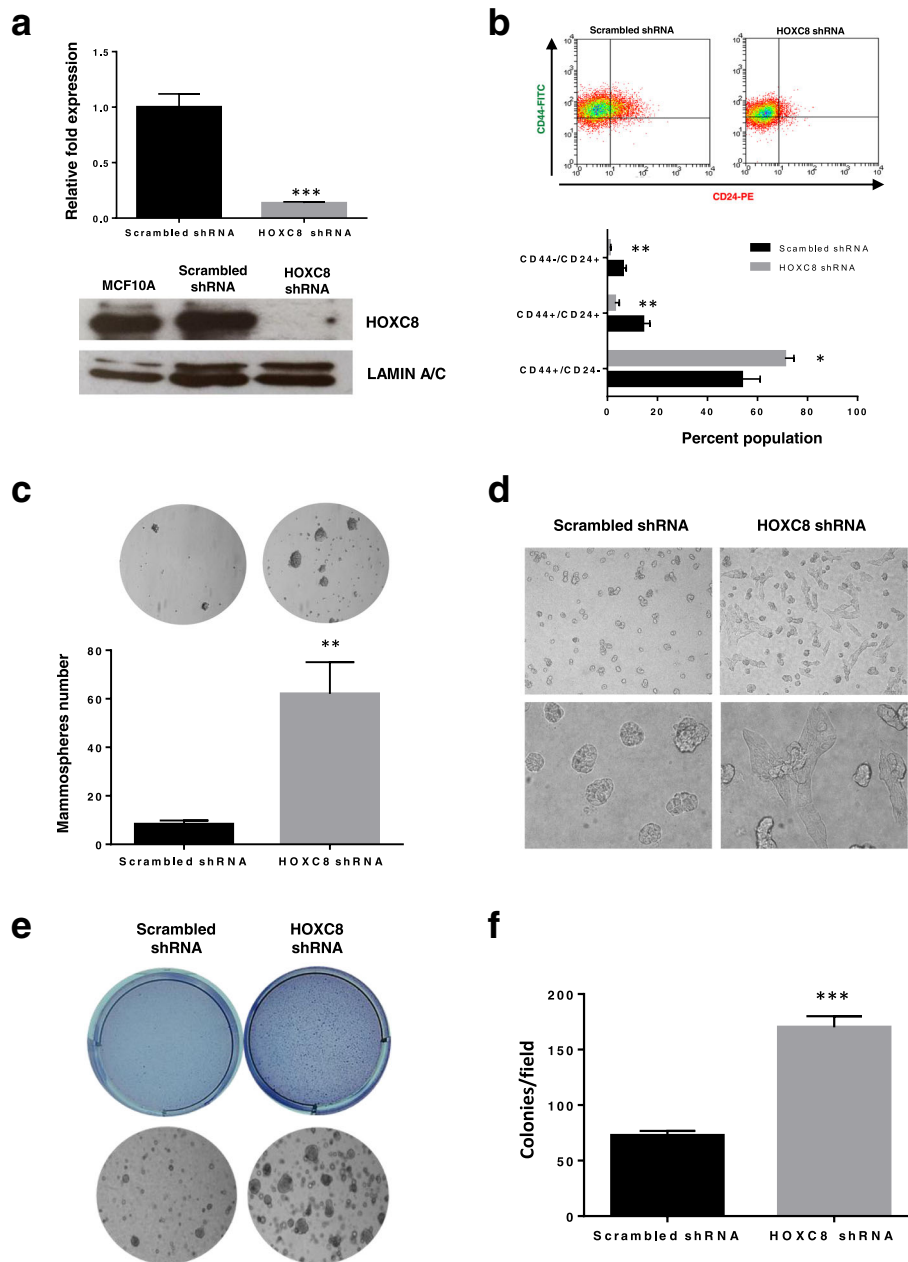
Fig. 7 HOXC8 knockdown increase CSC proliferation and transformed phenotype. **a** Knockdown of HOXC8 was induced by lentiviral transduction of MCF10A cells with pLKO-Tet-HOXC8 vector and gene expression measured by TaqMan® qRT-PCR. Results are presented as relative fold expression relative to *RPLP0* and pLKO-Tet-Scrambled vector used as calibrator (*n* = 3). Results were analysed by Unpaired Student's *t*-test (*n* = 3). ****P* < 0.001. Bottom panel shows western blotting of nuclear lysates from cells transduced with pLKO-Tet-HOXC8 vector or pLKO-Tet-Scrambled vector was conducted to detect the expression of HOXC8 (34 kDa) and LAMIN A/C (41–50 kDa) as loading control. **b** Percentage of CD44$^+$/CD24$^{low/-}$, CD44$^+$/CD24$^+$, CD44$^-$/CD24$^+$ cells after HOXC8 knockdown as measured by FACS analysis. Results were analysed by comparing cell populations levels in HOXC8 shRNA transduced cells compared to Scrambled shRNA (*n* = 3). Statistical analysis was performed by Unpaired Student's *t*-test. **P* < 0.05, ***P* < 0.01. Representative FACS profiles of cells double stained with CD44-FITC and CD244-PE to analyse the CSC and non-CSC populations are also shown. **c** Mammospheres morphology and number obtained by self-renewal of cell after HOXC8 knockdown compared to scrambled control vector after 3 generations. Mammospheres numbers were compared by Unpaired Student's *t*-test (*n* = 3). ***P* < 0.01. Only spheres ≥ 100 μM were counted, images were taken at 10X magnification. **d** Colonies of HOXC8 shRNA and Scrambled shRNA cells grown in 3D matrigel. Top panels represents 10X magnification, bottom panel represents images at 40X magnification. **e** Colonies of cells with HOXC8 knockdown compared to scrambled control grown in soft agar for 2 weeks. Top images show the whole well stained with crystal violet, bottom images show a representative field of view at 10X magnification. **f** The number of soft agar colonies were calculated by counting 10 fields of view. Results were analysed by Unpaired Student's *t*-test for each cell line (*n* = 3). ****P* < 0.001

Shah *et al. Molecular Cancer* (2017) 16:38

Page 16 of 19

Fig. 8 (See legend on next page.)

Shah *et al. Molecular Cancer* (2017) 16:38

Page 17 of 19

(See figure on previous page.)
Fig. 8 HOXC8 downregulation prevents differentiation of MCF10A cells induced by retinoic acid. **a** Representative FACS profiles and percentage of CD44$^+$/CD24$^{low/-}$, CD44$^+$/CD24$^+$, CD44$^-$/CD24$^+$ cells after retinoic acid (ATRA) treatment for 4 days and HOXC8 knockdown induced by doxycyclin (n = 3). Statistical analysis was performed by One-way ANOVA followed by Bonferroni's multiple comparisons test. *P < 0.05, **P < 0.01, ***P < 0.001. **b** Representative FACS profiles and percentage of CD44$^+$/CD24$^{low/-}$, CD44$^+$/CD24$^+$, CD44$^-$/CD24$^+$ cells after retinoic acid (ATRA) treatment for 7 days and HOXC8 knockdown induced by doxycyclin (n = 3). Statistical analysis was performed by One-way ANOVA followed by Bonferroni's multiple comparisons test. *P < 0.05, **P < 0.01, ***P < 0.001

expression of HOXC8 during breast cancer progression that can induce metastasis through direct regulation of CDH11 and EMB genes [25, 59]. This different role of HOXC8 suggests that it is possible that this gene could act as a "modulator" of carcinogenesis and therefore play either a role as suppressor or driver of transformation depending on the physiological and pathological context. Consistent with this, HOXC8 has been implicated in different cancer types, including prostate, cervical, breast, oesophageal, and pancreatic cancer [36, 60–63]. In these cancers *HOXC8* has been reported to act either as tumour suppressor gene or oncogene, suggesting that the function of *HOXC8* may depend on the activation or repression of different HOXC8 targets whose expression can be tissue-specific. Several targets of HOXC8 have been characterised [57, 64] and their functional role in the context of HOXC8 regulation may provide new insights into the role of HOXC8 in different cancer types. Although an effective approach for inhibition of the interaction between HOX proteins and co-factors PBX is available [65], targeting specific HOX proteins still remains a challenge. Therefore, a better definition of HOX transcriptional regulators and HOX targets could offer new strategies for therapeutic intervention.

We have also shown that HOXC8 downregulation in mammary cells can lead to impaired differentiation in response to retinoic acid, and that the expression of *HOXC8* can be induced by retinoids in normal breast cells. This role has also been recently demonstrated for *HOXA5*, suggesting a prominent role of HOX genes in the differentiation of the mammary gland [52]. Retinoic acid has been shown to regulate breast CSC differentiation, reduce mammospheres formation and CSC frequency in combination with chemotherapy [66, 67]. Therefore our study adds to the evidence that HOX genes are important regulators of stem cell fate and that differentiation therapy should be considered as a viable approach to target breast CSC.

Conclusions

This study reports the screening of a large number of homeobox genes in breast CSC. We show that the expression of homeobox genes is altered in early-transformed breast cells and in different breast cancer molecular subtypes. Altered expression of homeobox genes was also supported by data obtained from large cohorts of patient through bioinformatics analysis. Therefore, this screening provides novel biomarkers that can be used in precision medicine and as possible targets for CSC-directed therapies. Of the altered genes, *HOXC8* demonstrated to function as a possible novel tumour suppressor in breast CSC by regulating stem cell self-renewal, differentiation and transformation. Therefore, this study demonstrates the critical role of homeobox genes in stem cells and paves the way to larger functional studies to explore the networks involved in homeotic gene regulation in breast cancer.

Additional files

Additional file 1: Table S1. Primers and assays used in the study. (DOC 41 kb)

Additional file 2: Figure S1. Number of upregulated and downregulated homeobox genes in immortalised stem cells and cancer stem cells. a Proportion of up-and downregulated genes in immortalised mammary stem cells (IMM) and CSC of different breast cancer molecular subtypes as determined by Homeobox gene expression array. b Number of downregulated and upregulated genes which are unique or common to CSC of breast cancer molecular subtypes. c Venn diagrams representing downregulated and upregulated genes in stem cells of different breast cancer molecular subtypes. (PDF 204 kb)

Additional file 3: Figure S2. Data heatmap and histograms showing expression of homeobox genes in the PAM50 (TCGA WGA) (a,b,c,d) and METABRIC datasets (e,f,g,h). Yellow and blue colour intensity in the genomic heatmap indicate high and low expression, respectively. The red and green colours under the genomic heatmap indicate the statistics track showing the logarithmic plot of *P*-values for each gene using Student' *t*-test followed by Bonferroni's correction. Bars above the line indicates that the red subgroup (Tumour sample) is greater than the green subgroup; a bar below the line indicates that the green subgroup (Normal sample) is greater than the red subgroup. Bars are coloured in red or green when *P* < 0.05. (PDF 63 kb)

Additional file 4: Figure S3. *HOXC8* copy number variation in the breast TCGA and METABRIC studies. Bar plots and scatter plots representing the copy number variation and correlation with gene expression in patient samples from the TCGA and METABRIC datasets [29, 30, 33, 34]. (PDF 271 kb)

Additional file 5: Figure S4. Sequencing chromatograms of bisulfite converted *HOXC8* CpG island PCR products. Bisulfite converted DNA extracted from CSC was amplified by PCR and then directly sequenced. Chromatograms show the average DNA methylation of the PCR products. CG residues are circled. A TG peak indicated unmethylated C, T/G peak indicates partially methylated C, CG peak indicates methylated C. (PDF 311 kb)

Additional file 6: Figure S5. *HOXC8* overexpression in TN and HER2-E cells. Overexpression of *HOXC8* was induced by lentiviral transduction with pSIN-HOXC8 vector and gene expression measured by TaqMan®

Shah *et al. Molecular Cancer* (2017) 16:38

Page 18 of 19

qRT-PCR. Results are presented as relative fold expression relative to *RPLP0* and control (pSIN empty) vector used as calibrator ($n = 3$). Relative fold expression levels were analysed by Unpaired Student *t*-test. ***$P <$ 0.001. Bottom panel represents Western Blotting of nuclear lysates from cells transduced with pSIN-HOXC8 overexpressing vector or control vector was conducted to detect the expression of HOXC8 (34 kDa) and LAMIN A/C (41-50 kDa) as loading control. (PDF 92 kb)

Additional file 7: Figure S6. Expression of *HOXC8* and *CD24* in ATRA treated cells. Expression of *HOXC8* and *CD24* in MCF10A-Tet-HOXC8 shRNA cells measured by TaqMan® qRT-PCR. Results are presented as relative fold expression relative to *RPLP0* and control (shRNA with no doxycyclin) used as calibrator ($n = 3$). Relative fold expression levels were analysed by One-way ANOVA followed by Bonferroni's multiple comparisons test. *$P < 0.05$, **$P <$ 0.01, ***$P < 0.001$. (PDF 52 kb)

Availability of data and materials
The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Authors' contributions
MS performed experiments and contributed to the design of the study. RC performed miRNA analyses and assisted with bioinformatics analysis. BW performed gene knockdown experiments. JP critically reviewed the manuscript and assisted in study design. NM performed bioinformatics analyses and critically reviewed the manuscript. AG critically reviewed the manuscript and contributed to study design. CA performed experiments, designed the study and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
The study was approved by the University of Nottingham local ethical committee. Anonymised patient data from TGCA and METABRIC studies were available from the Genomic Data Commons Data Portal and OASIS Genomics Portal.

Author details
[1]SVMS, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK. [2]Department of Pharmacology, Weill Cornell Medicine, 1300 York Ave., New York, NY 10065, USA. [3]Department of Translational Medicine, Lund University, Malmö 205 02, Sweden. [4]Department of Molecular Biology, Umeå University, 901 87 Umeå, Sweden. [5]Cancer Biology, Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, QMC, Nottingham NG7 2UH, UK.

References
1. Shah M, Allegrucci C. Keeping an open mind: highlights and controversies of the breast cancer stem cell theory. Breast Cancer. 2012;4:155–66.
2. Valent P, Bonnet D, De Maria R, Lapidot T, Copland M, Melo JV, Chomienne C, Ishikawa F, Schuringa JJ, Stassi G, et al. Cancer stem cell definitions and terminology: the devil is in the details. Nat Rev Cancer. 2012;12:767–75.
3. Mark M, Rijli FM, Chambon P. Homeobox genes in embryogenesis and pathogenesis. Pediatr Res. 1997;42:421–9.
4. Shah N, Sukumar S. The Hox genes and their roles in oncogenesis. Nat Rev Cancer. 2010;10:361–71.
5. Atkinson SP, Koch CM, Clelland GK, Willcox S, Fowler JC, Stewart R, Lako M, Dunham I, Armstrong L. Epigenetic marking prepares the human HOXA cluster for activation during differentiation of pluripotent cells. Stem Cells. 2008;26:1174–85.
6. Barber BA, Rastegar M. Epigenetic control of Hox genes during neurogenesis, development, and disease. Ann Anat. 2010;192:261–74.
7. He H, Hua X, Yan J. Epigenetic regulations in hematopoietic Hox code. Oncogene. 2011;30:379–88.
8. Chen H, Sukumar S. Role of homeobox genes in normal mammary gland development and breast tumorigenesis. J Mammary Gland Biol Neoplasia. 2003;8:159–75.
9. Ocana OH, Corcoles R, Fabra A, Moreno-Bueno G, Acloque H, Vega S, Barrallo-Gimeno A, Cano A, Nieto MA. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. Cancer Cell. 2012;22:709–24.
10. McCoy EL, Iwanaga R, Jedlicka P, Abbey NS, Chodosh LA, Heichman KA, Welm AL, Ford HL. Six1 expands the mouse mammary epithelial stem/progenitor cell pool and induces mammary tumors that undergo epithelial-mesenchymal transition. J Clin Invest. 2009;119:2663–77.
11. Yu M, Smolen GA, Zhang J, Wittner B, Schott BJ, Brachtel E, Ramaswamy S, Maheswaran S, Haber DA. A developmentally regulated inducer of EMT, LBX1, contributes to breast cancer progression. Genes Dev. 2009;23:1737–42.
12. Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, Itzkovitz S, Noske A, Zurrer-Hardi U, Bell G, et al. Slug and Sox9 cooperatively determine the mammary stem cell state. Cell. 2012;148:1015–28.
13. Hu G, Lee H, Price SM, Shen MM, Abate-Shen C. Msx homeobox genes inhibit differentiation through upregulation of cyclin D1. Development. 2001;128:2373–84.
14. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008;10:593–601.
15. Zhang L, Yang M, Gan L, He T, Xiao X, Stewart MD, Liu X, Yang L, Zhang T, Zhao Y, Fu J. DLX4 upregulates TWIST and enhances tumor migration, invasion and metastasis. Int J Biol Sci. 2012;8:1178–87.
16. Samuel S, Naora H. Homeobox gene expression in cancer: insights from developmental regulation and deregulation. Eur J Cancer. 2005;41:2428–37.
17. Abate-Shen C. Deregulated homeobox gene expression in cancer: cause or consequence? Nat Rev Cancer. 2002;2:777–85.
18. Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. Breast Cancer Res. 2009;11:R14.
19. Novak P, Jensen TJ, Garbe JC, Stampfer MR, Futscher BW. Stepwise DNA methylation changes are linked to escape from defined proliferation barriers and mammary epithelial cell immortalization. Cancer Res. 2009;69:5251–8.
20. Novak P, Jensen T, Oshiro MM, Watts GS, Kim CJ, Futscher BW. Agglomerative epigenetic aberrations are a common event in human breast cancer. Cancer Res. 2008;68:8616–25.
21. Bloushtain-Qimron N, Yao J, Snyder EL, Shipitsin M, Campbell LL, Mani SA, Hu M, Chen H, Ustyansky V, Antosiewicz JE, et al. Cell type-specific DNA methylation patterns in the human breast. Proc Natl Acad Sci U S A. 2008; 105:14076–81.
22. Kamalakaran S, Varadan V, Giercksky Russnes HE, Levy D, Kendall J, Janevski A, Riggs M, Banerjee N, Synnestvedt M, Schlichting E, et al. DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. Mol Oncol. 2011;5:77–92.
23. Ronneberg JA, Fleischer T, Solvang HK, Nordgard SH, Edvardsen H, Potapenko I, Nebdal D, Daviaud C, Gut I, Bukholm I, et al. Methylation profiling with a panel of cancer related genes: association with estrogen receptor, TP53 mutation status and expression subtypes in sporadic breast cancer. Mol Oncol. 2011;5:61–76.
24. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotechnol Lett. 2004;26:509–15.
25. Chao F, Zhang J, Zhang Y, Liu H, Yang C, Wang J, Guo Y, Wen X, Zhang K, Huang B, et al. Embigin, regulated by HOXC8, plays a suppressive role in breast tumorigenesis. Oncotarget. 2015;6:23496–509.

Shah *et al. Molecular Cancer* (2017) 16:38

Page 19 of 19

26. Lee GY, Kenny PA, Lee EH, Bissell MJ. Three-dimensional culture models of normal and malignant breast epithelial cells. Nat Methods. 2007;4:359–65.

27. Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D, Zhu J. Exploring TCGA Pan-cancer data at the UCSC cancer genomics browser. Sci Rep. 2013;3:2652.

28. Fernandez-Banet J, Esposito A, Coffin S, Horvath IB, Estrella H, Schefzick S, Deng S, Wang K, Aching K, Ding Y, et al. OASIS: web-based platform for exploring cancer multi-omics data. Nat Meth. 2016;13:9–10.

29. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C, et al. Comprehensive molecular portraits of invasive lobular breast cancer. Cell. 2015;163:506–19.

30. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

31. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

32. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res. 2012;40:W622–627.

33. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012; 486:346–52.

34. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, Pugh M, Jones L, Russell R, Sammut SJ, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun. 2016;7:11479.

35. Huang WY, Hsu SD, Huang HY, Sun YM, Chou CH, Weng SL, Huang HD. MethHC: a database of DNA methylation and gene expression in human cancer. Nucleic Acids Res. 2015;43:D856–861.

36. Li Y, Zhang M, Chen H, Dong Z, Ganapathy V, Thangaraju M, Huang S. Ratio of miR-196s to HOXC8 messenger RNA correlates with breast cancer cell migration and metastasis. Cancer Res. 2010;70:7894–904.

37. Wang YA, Shen K, Wang Y, Brooks SC. Retinoic acid signaling is required for proper morphogenesis of mammary gland. Dev Dyn. 2005;234:892–9.

38. Vermot J, Schuhbaur B, Le Mouellic H, McCaffery P, Garnier JM, Hentsch D, Brulet P, Niederreither K, Chambon P, Dolle P, Le Roux I. Retinaldehyde dehydrogenase 2 and Hoxc8 are required in the murine brachial spinal cord for the specification of Lim1+ motoneurons and the correct distribution of Islet1+ motoneurons. Development. 2005;132:1611–21.

39. Nadauld LD, Sandoval IT, Chidester S, Yost HJ, Jones DA. Adenomatous polyposis coli control of retinoic acid biosynthesis is critical for zebrafish intestinal development and differentiation. J Biol Chem. 2004;279:51581–9.

40. Shah M, Allegrucci C. Stem cell plasticity in development and cancer: epigenetic origin of cancer stem cells. Subcell Biochem. 2013;61:545–65.

41. Faryna M, Konermann C, Aulmann S, Bermejo JL, Brugger M, Diederichs S, Rom J, Weichenhan D, Claus R, Rehli M, et al. Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis. FASEB J. 2012;26:4937–50.

42. Johnson KC, Koestler DC, Fleischer T, Chen P, Jenson EG, Marotti JD, Onega T, Kristensen VN, Christensen BC. DNA methylation in ductal carcinoma in situ related with future development of invasive breast cancer. Clin Epigenetics. 2015;7:75.

43. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006;125:301–13.

44. Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet. 2007;39:237–42.

45. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet. 2007;39:232–6.

46. Yadav BS, Chanana P, Jhamb S. Biomarkers in triple negative breast cancer: a review. World J Clin Oncol. 2015;6:252–63.

47. Eccles SA, Aboagye EO, Ali S, Anderson AS, Armes J, Berditchevski F, Blaydes JP, Brennan K, Brown NJ, Bryant HE, et al. Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. Breast Cancer Res. 2013;15:R92.

48. Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P, Visvanathan K, et al. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. Cancer Res. 2011;71:6195–207.

49. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121:2750–67.

50. Tao Y, Xi S, Briones V, Muegge K. Lsh mediated RNA polymerase II stalling at HoxC6 and HoxC8 involves DNA methylation. PLoS One. 2010;5:e9163.

51. Keller PJ, Lin AF, Arendt LM, Klebba I, Jones AD, Rudnick JA, DiMeo TA, Gilmore H, Jefferson DM, Graham RA, et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. Breast Cancer Res. 2010;12:R87.

52. Teo WW, Merino VF, Cho S, Korangath P, Liang X, Wu RC, Neumann NM, Ewald AJ, Sukumar S. HOXA5 determines cell fate transition and impedes tumor initiation and progression in breast cancer through regulation of E-cadherin and CD24. Oncogene. 2016;35(42):5539–51.

53. Ordonez-Moran P, Dafflon C, Imajo M, Nishida E, Huelsken J. HOXA5 counteracts stem cell traits by inhibiting Wnt signaling in colorectal cancer. Cancer Cell. 2015;28:815–29.

54. Shashikant CS, Ruddle FH. Combinations of closely situated cis-acting elements determine tissue-specific patterns and anterior extent of early Hoxc8 expression. Proc Natl Acad Sci U S A. 1996;93:12364–9.

55. Le Mouellic H, Lallemand Y, Brulet P. Homeosis in the mouse induced by a null mutation in the Hox-3.1 gene. Cell. 1992;69:251–64.

56. Lewis MT. Homeobox genes in mammary gland development and neoplasia. Breast Cancer Res. 2000;2:158–69.

57. Kalyani R, Lee JY, Min H, Yoon H, Kim MH. Genes frequently coexpressed with Hoxc8 provide insight into the discovery of target genes. Mol Cells. 2016;39:395–402.

58. Carroll LS, Capecchi MR. Hoxc8 initiates an ectopic mammary program by regulating Fgf10 and Tbx3 expression and Wnt/beta-catenin signaling. Development. 2015;142:4056–67.

59. Li Y, Chao F, Huang B, Liu D, Kim J, Huang S. HOXC8 promotes breast tumorigenesis by transcriptionally facilitating cadherin-11 expression. Oncotarget. 2014;5:2596–607.

60. Waltregny D, Alami Y, Clausse N, de Leval J, Castronovo V. Overexpression of the homeobox gene HOXC8 in human prostate cancer correlates with loss of tumor differentiation. Prostate. 2002;50:162–9.

61. Alami Y, Castronovo V, Belotti D, Flagiello D, Clausse N. HOXC5 and HOXC8 expression are selectively turned on in human cervical cancer cells compared to normal keratinocytes. Biochem Biophys Res Commun. 1999;257:738–45.

62. Du YB, Dong B, Shen LY, Yan WP, Dai L, Xiong HC, Liang Z, Kang XZ, Qin B, Chen KN. The survival predictive significance of HOXC6 and HOXC8 in esophageal squamous cell carcinoma. J Surg Res. 2014;188:442–50.

63. Adwan H, Zhivkova-Galunska M, Georges R, Eyol E, Kleeff J, Giese NA, Friess H, Bergmann F, Berger MR. Expression of HOXC8 is inversely related to the progression and metastasis of pancreatic ductal adenocarcinoma. Br J Cancer. 2011;105:288–95.

64. Lei H, Juan AH, Kim MS, Ruddle FH. Identification of a Hoxc8-regulated transcriptional network in mouse embryo fibroblast cells. Proc Natl Acad Sci U S A. 2006;103:10305–9.

65. Morgan R, Boxall A, Harrington KJ, Simpson GR, Gillett C, Michael A, Pandha HS. Targeting the HOX/PBX dimer in breast cancer. Breast Cancer Res Treat. 2012;136:389–98.

66. Ginestier C, Wicinski J, Cervera N, Monville F, Finetti P, Bertucci F, Wicha MS, Birnbaum D, Charafe-Jauffret E. Retinoid signaling regulates breast cancer stem cell differentiation. Cell Cycle. 2009;8:3297–302.

67. Sun R, Liu Y, Li SY, Shen S, Du XJ, Xu CF, Cao ZT, Bao Y, Zhu YH, Li YP, et al. Co-delivery of all-trans-retinoic acid and doxorubicin for cancer therapy with synergistic inhibition of cancer stem cells. Biomaterials. 2015;37:405–14.