Title: SALL4 promotes tumor progression in breast cancer by targeting EMT

Authors: Teng Chen[1], Julia YS Tsang[2], Xiao-Chun Su[1], Peng Li[1], Wen-Qin Sun[1], Iris L.K. Wong[1], Kit-Ying Choy[1], Qing Yang[1], Gary MK Tse[2], Tak Hang Chan[1,3§], Larry M.C. Chow[1§].

Affiliations:

[1]Department of Applied Biology and Chemical Technology and State Key Laboratory of Chemical Biology and Drug Discovery, The Hong Kong Polytechnic University, Hong Kong SAR, China.

[2]Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Ngan Shing Street, Shatin, NT, Hong Kong

[3]Department of Chemistry, McGill University, 801 Sherbrooke St. West, Montreal, Quebec, H3A 2K6 Canada.

§Corresponding authors:

Larry M.C. Chow: Y813, Lee Shau Kee Building, Department of Applied Biology and Chemical Technology and State Key Laboratory of Chemical Biology and Drug Discovery, The Hong Kong Polytechnic University, Hong Kong SAR, China. E-mail: larry.chow@polyu.edu.hk Fax: +852-23649932; Tel: +852-34008662;

Tak Hang Chan: Department of Chemistry, McGill University, 801 Sherbrooke St. West, Montreal, Québec H3A 0B8 Canada.

E-mail: tak-hang.chan@polyu.edu.hk Fax: 1-514-398-3797; Tel:1-514-398-6999;

Key words: breast cancer; lung metastasis, tumorigenicity, SALL4, epithelial-mesenchymal transition

The total number of figures: 5

The total number of tables: 1

Number of Supporting Information: 5

Abstract

Sal-like protein 4 (SALL4) is overexpressed in breast cancer and might contribute to breast cancer progression, but the molecular mechanism remains unknown. Here we found that within a group of 371 ethnic Chinese breast cancer patients, SALL4 was associated with lower grade (p=0.002) and Progesterone Receptor (PR) positivity (p=0.004) for overall cases; lower Ki67 (p=0.045) and high vimentin (p=0.007) for luminal cases. Patients with high SALL4 expression in lymph node (LN) metastasis showed a significant worse survival than those with low expression. Knockout of SALL4 in a triple negative breast cancer cell line MDA-MB-231-Red-FLuc-GFP led to suppressed ability in proliferation, clonogenic formation, migration and mammosphere formation *in vitro,* tumorigenicity and metastasis *in vivo*. On the other hand, overexpression of SALL4 enhanced migration and mammosphere formation *in vitro* and tumorigenicity *in vivo.* Mechanistically, there was a positive correlation between SALL4 expression and mesenchymal markers including Zinc finger E-box binding homeobox 1 (ZEB1), vimentin, Slug, and Snail *in vivo*. ChIP experiment indicated that SALL4 can bind to the promoter region of vimentin (-778 to -550 bp). Taken together, we hypothesize that SALL4 promotes tumor progression in breast cancer by inducing the mesenchymal markers like vimentin through directly binding to its promoter. Increased SALL4 level in metastatic lymph node relative to primary site is an important poor survival marker in breast cancer.

Introduction

Sal-like protein 4 (SALL4) belongs to a group of zinc finger transcription factors featured by multiple cys (2) his (2) (C2H2)-type zinc finger domains. It is expressed in fetal cells for maintaining pluripotency and self-renewal[1,2] partly through physical and functional interaction with Oct4, Sox2, and Nanog[3-5]. SALL4 is normally downregulated or absent in most adult tissues. However, it is dysregulated and aberrantly expressed in hematological[6] and solid malignancies like lung cancer[7], colorectal carcinoma[8,9], endometrial cancer[10], glioma[11], esophageal squamous cell carcinoma[12], gastrointestinal carcinoma[13] and hepatocellular carcinoma (HCC)[14]. SALL4 can interact with HDAC complex to exert suppressive effect on PTEN in hematological cancer and hepatocellular carcinoma, leading to cell proliferation and tumor growth. A 12-amino acid peptide from the N-terminus of SALL4 was able to specifically block the interaction of SALL4 and its epigenetic partner HDAC complex to achieve therapeutic effects through the reactivation of PTEN.[15,16] This peptide was optimized and proved to antagonize the SALL4–NURD interaction, thereby releasing the transcription-repression of SALL4 on massive transcripts that are beneficial to patient survival.[17]

Breast cancer is the second most common malignancy behind lung cancer with more than 2.1 million cases and 627 thousand deaths each year worldwide. Among various types of breast cancer, triple-negative breast cancer (TNBC) has poorer prognosis due to higher metastatic potential and lack of specific therapeutic targets. SALL4 was reported to be overexpressed in invasive breast cancer and defined as a dispersion factor in basal-like breast cancer due to suppression of the expression of E-cadherin.[18] It was reported to promote breast cancer cell migration[19] and stemness[20] possibly through regulation of integrin α6β1 and modulating alternative splicing of CD44 mRNA. SALL4 was also essential for breast cancer cell proliferation[21] and knockdown of SALL4 led to cell cycle arrest and reversal of chemo-resistance of breast cancer cells by down regulation of breast cancer resistance protein

(BCRP)[22]. The investigation of SALL4 in breast cancer, however, is limited and majority of these studies remain at cellular level. In this work, systematic *in vitro* and *in vivo* experiments together with clinical investigation have been conducted to explore the function of SALL4 in different oncogenic processes including cell survival, chemoresistance, metastasis and tumorigenesis. Our results demonstrated that genetic knockout of *SALL4* suppressed tumorigenicity and metastasis of breast cancer cells through inhibition of epithelial-mesenchymal transition (EMT).

Material and Methods

Patients and tissues

The study was approved by Joint Chinese University of Hong Kong-New Territories East Cluster clinical research ethics committee. The cohort consisted of breast cancer patients from Prince of Wales Hospital between 2002 and 2008. Clinical data were retrieved from medical records. All H&E slides were reviewed for histological diagnosis and grade. Outcome data included date of initial diagnosis and time to first tumor relapse (DFS) and breast cancer specific survival (BCSS).

Cell lines

MDA-MB-231-Red-FLuc-GFP (authenticated by Perkin Elmer), MDA-MB-468, MCF7 cell lines were cultured in RPMI 1640 medium. $CD44^{+}CD24^{-}$ MCF7 cancer stem cells was cultured in MammoCult™ Basal Medium (Human) medium with 10% MammoCult™ Proliferation Supplement, heparin, hydrocortisone 21-hemisuccinate (sodium salt), 100 U/mL penicillin and 100 μg/mL of streptomycin.

Immunohistochemistry and scoring

Tissue microarray (TMA) containing representative tumor areas from primary site and nodal metastasis were constructed with 0.6 mm tissue cores as previously described[23]. 4-μm TMA sections were stained for SALL4 and vimentin on Ventatna BenchMark sutostainer using Ultraview Universal DAB Detection Kit (Ventana, Tucson, AZ). Biomarker expression was assessed for the staining intensity (graded 1-3) and percentage of positively stained tumor cells. An immunoscore for SALL4 was obtained by multiplying staining intensity with percentage of staining nuclei and the median score was used as cutoff. Positive vimentin was defined by ≥ 5% of tumor cells with strong/ moderate cytoplasmic staining. The expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67, were retrieved from our database. Further details on IHC staining were listed in Table

S1.

CRISPR-cas9-based SALL4 knockout and Lentivector-based SALL4-overexpression in MDA-MB-231-Red-FLuc-GFP cells

SALL4 sgRNA CRISPR/Cas9 All-in-One Lentivectors and the negative control vector were purchased in ABM Inc. SALL4 overexpression plasmid was purchased from DNASU (clone: HsCD00443167). Lentivirus particles were produced in HEK293FT cells using the third generation packaging system by transfecting pLP1, pLP2, pVSVG and the SALL4 sgRNA CRISPR/Cas9 plasmids using calcium phosphate method. Lentivirus particles was mixed with Polybrene (8 μg/mL) and added to MDA-MB-231-Red-FLuc-GFP cells (MOI = 1) for cell infection. Serial dilution in 96-well plate was used to isolate single clones. Sequence of sgRNA targeting SALL4: CTGTGAGAAATGCTGTG.

Clonogenic assay and mammosphere formation assay

For clonogenic assay, three thousand MDA-MB-231-Red-FLuc-GFP cells were seeded in each well of a 6-well plate. After 7 to 8 days, the medium was aspirated and the cells were fixed with 100% methanol and stained by crystal violet before scanning for colony number counting (Image J).

For mammosphere formation assay, fifteen thousand MDA-MB-231-Red-FLuc-GFP cells were resuspended in the cancer stem cell medium. After incubation of 7 (for SALL4-knockout clones) or 5 days (for SALL4-overexpression clones), mammospheres with a diameter greater than 100 μm were counted in 15 microscopic fields.

Wound healing and trans-well assay

Two hundred thousand cells for SALL4-knockout clones (or one hundred and fifty thousand cells for SALL4-overexpression clones) were seeded into each well of a 24-well plate containing RPMI medium with 5% FBS. Cells were treated with 10 μg/ml mitomycin C (Sigma) for 2 h before a scratch was made on the surface of the well using a pipette tip. Pictures were

taken to record the healing status every 8 hours.

For trans-well assay, sixty to eighty thousand cells were seeded in the upper chamber with 200 μL of FBS-free RPMI medium. The lower chamber was added with 750 μL of RPMI medium with 10 % FBS. After incubation for 16 h, the upper chamber was fixed with 100% methanol and stained with 750 μL of 0.2% crystal violet. The chamber was washed and air-dried for picture taking.

Antibody raised against SALL4

The human SALL4 peptide KDCHRENGGSSEDM (amino acids 130-143) was chosen and used to prepare an antipeptide antibody by Shine gene company (Shang Hai, China). Serum was obtained after three rounds of immunizations in New Zealand rabbits. SALL4 antibody was obtained by antigen affinity purification. The antibody specificity was verified and the results were included in the supplementary Figure S1.

Chromatin Immunoprecipitation (ChIP)

Cells were crosslinked with formaldehyde (1% high purity grade) and lysed with complete SDS lysis buffer. The lysate was sonicated to shear DNA to lengths between 200 and 1000 bp and pre-cleared with 60 μL salmon sperm DNA /Protein G Agarose-50% Slurry (Merck Millipore) to reduce nonspecific background. SALL4 antibody (or rabbit IgG antibody) was added to the supernatant and incubated overnight at 4 °C with rotation. A volume of 60 μL salmon sperm DNA /Protein G Agarose-50% Slurry was added and incubated for 1 h at 4 °C to collect the antibody/histone complex. The agarose was then pelleted, washed and eluted by 600 μL of freshly prepared elution buffer. After reverse crosslinking, targeted-chromatin DNA was isolated using conventional phenol–chloroform method for real-time PCR.

Primer sequence of vimentin promoter was as follows: VIM1 F: 5'-CCAAGTAACCTGCAGTACCCC-3', R: 5'-GCTGAGTACTTACCCGCCAA-3'; VIM2 F: 5'- CCCAAGGTCAATTGCACGAA-3', R: 5'- AATGACAGAATCTTTGGCGGC-3'; VIM3

F: 5'- GGCCCAGCTGTAAGTTGGTA-3', R: 5'- TCTGTCGAGGGACCTAACGG-3'.

Western blot

Proteins of cell or tumor were extracted with RIPA buffer. Equal amount of proteins was loaded onto an 8% polyacrylamide gel for separation and then electroblotted onto PVDF Immobilon P membrane (MilliPore). The membrane was blocked by TBST (10 mM Tris-HCl; pH 7.5, 150 mM NaCl, 0.1% Tween 20) containing 5% powdered skim milk for 1 hour at room temperature with agitation. After blocking, the membrane was washed with TBST and incubated with 1:2000 SALL4 antibody (dilution ratio of other antibodies was referring to the antibody datasheets) for overnight at 4°C with agitation. The membrane was washed and then incubated with 1:3000 secondary antibody (Santa Cruz Biotechnology Inc.) for at least 1 hour at room temperature with agitation. After washing with TBST, the chemiluminescence signal of the membrane was detected after incubation in the SuperSignal Substrate Western Blot kit (Pierce). Sources of primary antibodies were: β-actin (Santa Cruz # sc-47778), NF90 (Santa Cruz # sc-377406), Lamin B (Cell signaling # 12586), V5 (Thermo Fisher # R960-25), ZEB1 (Cell signaling # 3396), vimentin (Cell signaling # 5741), Slug (Cell signaling # 9585), Snail (Cell signaling # 3879), E-cadherin (Cell signaling # 3195).

RNA extraction, reverse transcription and real-time qPCR

Total RNA was extracted from cells using Trizol Reagent (Invitrogen) according to the manufacturer's instructions. RNA (2 μg) was reverse-transcribed in 20 μL with the use of High Capacity Reverse Transcription Kit (Applied Biosystems). Quantitative PCR was performed using SsoFast TM EvaGreen Supermix with Low ROX (Bio-rad). Amplification was done using the following parameters: 95 °C (30 s), 40 cycles of 95 °C (10 s), 60 °C (30 s). β-actin was used as an internal control. The following primers were used: VIM F: 5'-GACAATGCGTCTCTGGCACGTCTT -3', R: 5'- TCCTCCGCCTCCTGCAGGTTCTT-3';
ACTB F: 5'-CTCTTCCAGCCTTCCTTCCT-3', R: 5'-AGCACTGTGTTGGCGTACAG-3'.

178 ***In vivo* tumorigenicity and metastasis study**

179 All animal experiments were conducted with the permission from the Department of Health,
180 Hong Kong Government and the Animal Subjects Ethics Sub-Committee of the Hong Kong
181 Polytechnic University. A total of $2x10^5$ to 10^7 cells in 1:1 PBS/Matrigel (Corning) were
182 injected subcutaneously in the right flank of BALB/c nude mice (4-6 weeks old female, Charles
183 River, China). All mice were weighed and tumor volume was measured using microcalipers
184 every other day. Tumor volume was calculated using the formula (A) (B^2) /2, where A was the
185 length of the longest aspect of the tumor, and B was the length of the tumor perpendicular to
186 A. Tumor initiating cell (TIC) frequency was calculated by
187 http://bioinf.wehi.edu.au/software/elda/[24] based on the number of mice with tumor on day 6
188 after inoculation.
189 For metastasis study, SALL4-knockout MDA-MB-231-Red-FLuc-GFP cells (2×10^6) or its
190 vector control cells in 100 µL of PBS were injected intravenously into the tail vein of 6 to 8-
191 week-old nude mice. Metastasis was monitored by *in vivo* animal imaging system (IVIS
192 Lumina series III, Perkin-Elmer). Mice were administrated with D-luciferin (150 mg/kg) by
193 intraperitoneal injection and anesthetized with isoflurane. Data were quantified and expressed
194 as average radiance (total photon/s/cm^2/sr).

195 **Statistical analysis**

196 For clinical results, chi-square analysis, Fisher's exact test, Mann-Whitney U test and Wilcoxon
197 test were used, when appropriate, to compare SALL4 expression with different relevant
198 parameters using SPSS version 23. Survival data were evaluated with Kaplan Meier analysis.
199 Results of *in vitro* and *in vivo* experiments were expressed as the mean ± SEM of at least 3
200 independent experiments. The criterion for significance was $p > 0.05$ (ns), $p < 0.05$ (∗), $p <$
201 0.01 (∗∗) and $p < 0.001$ (∗∗∗) for all comparisons.

Results

SALL4 was upregulated in breast cancer and may associate with lymph node metastasis in breast cancer patients.

Data analysis of the biggest breast cancer study[25] from the cBioPortal for Cancer Genomics[26,27] revealed that SALL4 was altered in 10% of 1904 cases of breast cancer. Amplification and mRNA high were the major types of alterations, accounting for 5.62% and 3.63% respectively (Figure 1A). Analysis from Oncomine database also demonstrated that the SALL4 mRNA level was 2.6-fold higher in invasive breast carcinoma than that in paired normal breast tissues (n=593, p<0.001, Figure 1B). Kaplan-Meier Plotter analysis[28] revealed that overexpression of SALL4 was correlated with poor overall survival (n=1402, HR=1.52, 95% CI=1.1 to 2.09, logrank p=0.01, Figure 1C).

In addition, we have examined SALL4 expression in a local cohort of 371 breast cancers by immunohistochemistry. Features of the cohort were shown in Table 1. Mean immunoscore was used as cutoff to define low SALL4 cases and high SALL4 cases (49%). Example images of low and high SALL4 staining were presented in Figure 1D. SALL4 expression was associated with lower grade (p=0.002), PR positivity (p=0.004) (Table 1). Interestingly, significant associations of SALL4 with lower Ki67 (p=0.045) and high vimentin (p=0.007) in luminal cancers were observed. For the 89 cases with paired nodal metastasis for SALL4 analysis, a significant upregulation of SALL4 was found in nodal metastasis compared to primary tumor (p=0.04) (Table S2). In line, the mean/median SALL4 score was higher in nodal metastasis than primary tumor (p=0.05) (Table S2). In this subset of cases, SALL4 expression in primary tumor consistently had no correlation with survival (Figure 1E; left). However, those patients with high SALL4 expression in nodal metastasis was associated with a poor survival (DFS: log-rank=3.605, p=0.058; BCSS: log-rank=7.261, p=0.007; Figure 1E; middle). The worst survival was observed especially in those with SALL4 upregulation in nodal metastasis (Tumor

lo LN hi group) (Figure 1E; right). To summarize, SALL4 was overexpressed in breast cancer and may correlate with lymph node metastasis.

Knockout of SALL4 in MDA-MB-231-Red-FLuc-GFP cells resulted in lowered proliferation, colony formation, mammosphere formation and migration.

SALL4 expression level in different breast cancer cell lines was measured including triple negative MDA-MB-468 and MDA-MB-231-Red-FLuc-GFP, ER positive MCF7 and $CD44^+/CD24^-$ MCF7 cancer stem cells (CSCs). Relatively higher protein expression was detected in MDA-MB-231-Red-FLuc-GFP cells (Figure 2A). We performed CRISPR-Cas9 mediated loss-of function study of SALL4 using this cell line. Two SALL4 deficient clones, KO#1 and KO#2, displayed significantly reduced level of SALL4 protein (Figure 2B). Of note, SALL4 in KO#2 was almost depleted, resulting in a slight but significant growth inhibition ($p<0.001$, Figure 2C). Knockdown of SALL4 also inhibited cell proliferation of MCF7 cells (supplementary Figure S2).

Clonogenic assay indicated that KO#1 and KO#2 had fewer and smaller colonies compared with wild type MDA-MB-231-Red-FLuc-GFP or vector infected cells (Figure 2D left). The colony formation percentage in KO#1 and KO#2 was reduced to 80% and 50% of vector control group, respectively ($p<0.001$, Figure 2D right). SALL4 was suggested as a stem cell biomarker for liver cancer[29] and it regulated the metastatic potential and stemness of gastric cancer cells[13]. We found that the stemness in terms of mammospheres from the two SALL4 knockout clones (KO#1 and KO#2) were reduced compared to wild-type or vector control (Figure 2E top). In addition, their mammosphere formation capability was reduced from 150±13 mammospheres (with diameter > 100 μm) per 15 microscopic views in vector control to 70±11 in KO#1 ($p<0.001$) and 60±5 in KO#2 ($p <0.001$) (Figure 2E bottom), representing a drop of 54% and 62% respectively.

The effect of SALL4 genetic deficiency on cell migration was investigated by wound healing

assay. Wild type and vector control cells can completely heal the scratched wound after 36 hours (Figure 2F). However, the wound healing percentage of KO#1 and KO#2 was significantly reduced to 62% and 48% ($p<0.001$) of that in vector control (Figure 2F). Similarly, trans-well assay demonstrated that the cell migration ability was reduced from 111±3% in vector control cells to 51±3% in KO#1 and 19±2% in KO#2 ($p<0.001$ for both, Figure 2G). Since SALL4 was reported to regulate ATP-binding cassette drug transporter in leukemia[30], the effect of SALL4 deficiency on doxorubicin (DOX) sensitivity in breast cancer cell lines was tested. There was no significant difference in IC_{50} of DOX (WT: 0.40±0.10 μM, -Vec: 0.36±0.11 μM, KO#1: 0.22±0.07 μM, KO#2: 0.21±0.05 μM; $p>0.05$). To conclude, genetic-deficiency of SALL4 resulted in reduced cell proliferation, colony formation, mammosphere formation and migration but did not have significant effect on chemosensitivity in MDA-MB231-Red-FLuc-GFP cells.

Upregulation of SALL4 enhanced cell migration and mammosphere formation.

In parallel, gain-of-function study of SALL4 in MDA-MB231-Red-FLuc-GFP cells was conducted. Lentivirus system was used to overexpress V5-tagged SALL4. Single clones were picked after blasticidin selection. Two single clones B8 and G3 showed a higher level of V5-tagged SALL4 protein expression compared to parental control or vector control (Figure 3A). There were no significant changes in cell proliferation (Figure 3B) or clonogenicity (Figure 3C) in B8 and G3 compared to vector control. Overexpression of SALL4 in MCF7 also had no effect on cell proliferation (supplementary Figure S2). There has been only one previous study investigating the overexpression of SALL4 in breast cancer and it indirectly demonstrated that overexpression of SALL4 did not have significant enhancement on cell proliferation.[31] There was no significant difference in IC_{50} of DOX in B8 or G3 either (WT: 0.36±0.05 μM, +Vec: 0.28±0.07 μM, B8: 0.39±0.08 μM, G3: 0.45±0.12 μM; $p>0.05$). The mammospheres formed by B8 and G3 were significantly bigger (Figure 3D left) and the number of mammospheres

(diameter more than 100 μm in 15 views) in B8 (226±12) and G3 (205±13) was about three-fold higher than that of vector control group (62±8) and parental cells (50±14) (Figure 3D). Migration ability were also enhanced after SALL4 was overexpressed (Figure 3E/F). The wound healing percentage was increased from 58±4% in vector control cells to 102±2% of B8 and 100±2% of G3 ($p<0.01$ for both, Figure 3E right). The relative cell migration ability was enhanced from 79±6% in vector control cells to 135±5% of B8 and 123±3% of G3 ($p<0.05$, G3 vs +Vec; $p<0.01$, B8 vs +Vec, Figure 3F).

Genetic-deficiency of SALL4 suppressed tumorigenicity, tumor growth and lung metastasis.

We investigated the tumorigenicity of SALL4-knockout (KO#1 and KO#2) and SALL4-overexpressed (B8 and G3) MDA-MB231-Red-FLuc-GFP cells. Cells with different inoculation density were injected subcutaneously in the right flank of female Balb/c nude mice. The tumor incidence was significantly lower in KO#1 and KO#2 compared to that of vector control group (Figure 4A). Tumor initiating cell (TIC) frequency was reduced from 100% in vector control (1/881,373) to 36% in KO#1 (1/2,455,616) and 11% in KO#2 (1/7,718,519), suggesting that knockout of SALL4 resulted in a reduction of TIC population. This result is consistent with the *in vitro* observation that genetic-deficiency of SALL4 resulted in reduced rate of mammosphere formation (Figure 2E). The tumor growth rate of SALL4-deficient groups was significantly reduced, especially in KO#2 group where almost no SALL4 protein was found ($p<0.001$, Figure 4B). The tumor volume of KO#1 and KO#2 was also significantly smaller than that of the control group at three different inoculation cell densities (Figure 4C and D). These results suggested that genetic deficiency of SALL4 in MDA-MB231-Red-FLuc-GFP cells reduced the tumorigenicity and inhibited the growth of the tumors.

In parallel, effect of SALL4 overexpression on tumorigenicity was also investigated. While TIC frequency of vector control (1/753,624) was low, overexpression of SALL4 increased TIC

frequency significantly in B8 (1) and G3 (1/176,307) (Figure 4E). This result was consistent with the mammosphere formation assay (Figure 3D). Tumor growth in nude mice was monitored when $5X10^6$ cells were inoculated. B8, but not G3, showed a small but significant increase in growth rate compared to vector control ($p<0.01$, Figure 4F). This was reflected in the tumor volume after tumors were excised on day 24 ($p<0.05$, Figure 4G and H). The difference between B8 and G3 might be due to the fact that B8 had more exogenous SALL4 than G3 (Figure 3A). The above result suggested that overexpression of SALL4 enhanced tumorigenicity and slightly promoted tumor growth in nude mice.

As deficiency of SALL4 suppressed *in vitro* cell migration (Figure 2F/2G), we measured the effect of genetic deficiency of SALL4 on *in vivo* metastasis. The vector control (-Vec) and KO#2 cells were injected intravenously into female Balb/c nude mice. While metastasis was observed in 3 out 6 mice in vector control group, none of 5 mice of KO#2 group (BLI $>2.0x10^4$) had metastasis signal (Figure 4I left). Quantification of bioluminescence signal in KO#2 was significantly lower compared to vector control (Figure 4I, right, $p<0.001$). *Ex vivo* imaging of different organs was also conducted to localize the metastatic site when the animals were sacrificed. Bioluminescence signal was only found in lung but not in brain, liver and spleen (Figure 4J, top). Quantification of bioluminescence signal in vector control lungs ($6.9x10^5$) was 20–fold of that in KO#2 ($3.4x10^4$) (Figure 4J, bottom). Lung metastatic nodules were visible in vector control group but not in KO#2 group (Figure 4K). Lung metastatic nodules were examined in paraffin-embedded sections stained with human leukocyte antigen (HLA) and H & E solution to locate the metastatic foci (Figure 4L). The size of each foci and average number of foci for each mouse were plotted in Figure 4M. It demonstrated that there was less metastatic foci (Figure 4M, right, $p<0.05$) and metastatic area indicated by size of each foci (Figure 4M, left, $p<0.001$) was smaller after SALL4 was knockout. Together, these results suggested that

genetic-deficiency of SALL4 reduced lung metastasis of MDA-MB-231-Red-FLuc-GFP cells *in vivo*.

Effect of SALL4 on EMT *in vivo*

Epithelial-Mesenchymal Transition (EMT) is the process where cancer cells of epithelial origin transform into cells with a mesenchymal phenotype. EMT is very important in cell migration and metastasis. Here the effect of SALL4 knockout and overexpression on EMT markers was investigated. SALL4-knockout cells (KO#2) and SALL4-overexpressing cells (B8) were inoculated subcutaneously in the right flank of Balb/c nude mice and the tumors formed were excised for analysis of EMT markers. Knockout of SALL4 led to a significant drop in mesenchymal markers including ZEB1, vimentin, Slug and Snail (Figure 5A). There was an increase in epithelial marker E-cadherin (Figure 5A). On the other hand, overexpression of SALL4 increased expression of mesenchymal markers including ZEB1, vimentin, Slug, and Snail. The level of E-cadherin was not changed significantly compared with vector control (Figure 5B). This result suggested that SALL4 was a key component in driving EMT *in vivo* and might explain why SALL4 deficiency can reduce metastasis.

SALL4 directly bound to the promoter of vimentin.

SALL4 is a transcription factor that can promote transcription of mesenchymal markers[18]. Sall4 can bind to vimentin promoter in mouse extraembryonic endoderm (XEN) cells[3]. mRNA level of vimentin was also increased after SALL4 was overexpressed (Figure 5C). Here we investigated if SALL4 can regulate the expression of vimentin in breast cancer cells by binding to its promoter. Three putative SALL4-binding sites at 5' region of vimentin gene (VIM1, -778 to -550bp; VIM2, -409 to -273bp; VIM3, +186 to +435bp) were tested (Figure 5D). ChIP results demonstrated that SALL4 bound to region I (VIM 1, -778 to -550 bp) in both vector control group (17-fold over IgG control, Figure 5E top) and B8 group (35-fold over IgG control, Figure 5E bottom). SALL4-overexpression in B8 increased SALL4 binding to VIM1 by 2.1-

fold (Figure 5E bottom and 5F). There was no binding in region II and region III in both cell lines. This suggested that SALL4 might bind to VIM1 region and drive the transcription of vimentin. In 287 of clinical luminal cases of our studied population of breast cancer patients, there was a strong association between SALL4 level and high vimentin expression (p=0.007, Table 1; p=0.043, Figure 5G). Percentage of vimentin-stained cells was higher in SALL4-high cohort (mean=4.49) than that in SALL4-low cohort (mean=1.22) (Figure 5G, p=0.043). Similar trend can be observed for the correlation between SALL4 and vimentin among the basal like breast cancers. However, the number of cases was limited for meaningful statistical analysis (Table S3). Taken together, SALL4 might bind to the promoter region (-778 to -550 bp) of vimentin gene, thereby stimulating transcription of vimentin.

Discussion

SALL4 level correlated with PR positivity.

We performed detailed analysis of SALL4 expression and other clinicopathological features of breast cancer patients. One previous study found that SALL4 mRNA level was elevated in 86.1% (31/36) of the specimens but no obvious correlations were detected between SALL4 mRNA and clinicopathological factors (tissue type, menopausal status, lymph node metastasis, ER, PR)[21]. Here we demonstrated that SALL4 protein was associated with low grade (p=0.002) and PR positive status (p=0.004) (n=371). Kobayashi et al. illustrated that higher SALL4 mRNA level was associated with early clinical stages (II rather than III and IV) of breast cancer, which was consistent with our results that SALL4 protein correlated with lower grade of the disease. PR has been linked with stemness of breast cancer cells by increasing the progenitor cell populations *in vitro*. PR induction of WNT4 and RANKL in the luminal compartment acted in a paracrine manner to enrich the basal MaSC population[32]. Together, our observation that SALL4 was associated with PR expression may suggest that SALL4 correlates with cancer cell stemness.

SALL4 may promote metastasis and tumorigenicity in breast cancer by driving EMT.

SALL4 level was highly correlated with lymph node (LN) metastasis in gastric cancer[13] and colorectal cancer[8] but there was no studies in breast cancer. A common route of metastasis of breast cancer is through axillary lymph node. Here we demonstrated, for the first time, a higher SALL4 level in LN relative to primary tumor in terms of overall and paired cases (Table S2), suggesting that SALL4 may promote breast cancer lymph node metastasis. We also demonstrated that patients with higher SALL4 expression in LN metastasis showed a significant worse survival and those cases with high SALL4 expression in LN but low in primary tumor had the worst survival. Such finding could provide a new prognosis marker for

386 breast cancer patients. Overall, our clinical data suggest that SALL4 was positively associated
387 with metastasis in breast cancer.

388 With the clinical results indicating that SALL4 might promote lymph node metastasis, our *in*
389 *vitro* and *in vivo* experiments were also consistent with such observation. Genetic deficiency
390 of SALL4 suppressed cell migration *in vitro* and lung colonization *in vivo*, whereas
391 overexpression led to the opposite, confirming that SALL4 plays a positive role in breast cancer
392 metastasis. So far, there were only two studies investigating the *in vivo* metastatic potential of
393 SALL4 (gastric[13] and endometrial cancer[33]), both demonstrating that downregulation of
394 SALL4 blocked the metastatic potential. Here we reported, for the first time, that in breast
395 cancer, knockout of SALL4 suppressed lung colonization. Our result could pave a way for
396 potential treatment of breast cancer lung metastasis by targeting SALL4.
397 Our results also shed insights on how SALL4 promotes metastasis. EMT is known to enable
398 cancer cells to acquire stem cell properties for metastasis. We propose that SALL4 promotes
399 EMT by inducing the expression of mesenchymal markers like ZEB1, Slug, Snail and vimentin
400 and suppressing epithelial marker E-cadherin. Some of these mesenchymal markers also
401 contribute to cell stemness, e.g. Snail inducing $CD44^{high}$ $CD24^{low}$ cancer stem cell population[34]
402 and ZEB1 inducing tumor-initiating capacity in pancreatic and colorectal cancer cells[35]. It is
403 possible that SALL4 promote metastasis by enhancing EMT and cancer cell stemness.
404 There are, however, limitations in our animal model of lung colonization. Some mice injected
405 with vector control plasmid did not develop lung colonization. It was uncertain if that was due
406 to unsuccessful injection because no bioluminescent signal could be detected in a similar cell
407 line (MDA-MB-231-luc-D3H1) according to the supplier, Bioware. Another limitation is that
408 such model only represented lung colonization and it may not truly represent metastasis.
409

SALL4 might directly activate vimentin expression by binding to its promoter.

Vimentin is an intermediate filament that forms one of the tripartite components of the cytoskeleton. We demonstrated that SALL4 can bind to vimentin promoter region VIM1 (-778 to -550 bp), possibly explaining why SALL4 induced vimentin expression in other cancers like liver cancer[36] and esophageal squamous cell carcinoma[37]. Tian group predicted, through bioinformatic analysis, that "GTGG" was one of the putative binding sites of SALL protein family[38]. We observed that VIM1 region (-778 to -550 bp) also had a "GTGG" site. This might explain why SALL4 bound preferentially to this region relative to VIM2 and VIM3 regions. As a signaling hub within the cell, vimentin has been reported to associate with many other proteins involved in EMT. Vimentin can positively regulate the transcriptional activity of Slug through the interaction with ERK.[39] The decrease of Slug after SALL4 was knockout might be partly due to reduced vimentin level. Vimentin-mediated breast cancer cell migration and lung extravasation was partly Axl (receptor tyrosine kinase Axl)-dependent.[40] This might explain how SALL4 promotes metastasis through vimentin. To complete the study, a rescue experiment would be needed in the future though.

To conclude, SALL4 promotes EMT by positive regulation of ZEB1, Slug, Snail and vimentin to enable breast cancer cells to acquire stem-cell-like and metastatic properties. This might explain why higher SALL4 level was observed in metastatic lymph node relative to primary site. The SALL4 level in metastatic lymph node relative to primary site is an important survival prognosis marker in breast cancer.

References

1. Yang J, Gao C, Chai L, Ma Y. A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. *PLoS One.* 2010;5(5):e10766.

2. Rao S, Zhen S, Roumiantsev S, McDonald LT, Yuan GC, Orkin SH. Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. *Mol Cell Biol.* 2010;30(22):5364-5380.

3. Lim CY, Tam WL, Zhang J, et al. Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. *Cell stem cell.* 2008;3(5):543-554.

4. van den Berg DL, Snoek T, Mullin NP, et al. An Oct4-centered protein interaction network in embryonic stem cells. *Cell stem cell.* 2010;6(4):369-381.

5. Wu Q, Chen X, Zhang J, et al. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem.* 2006;281(34):24090-24094.

6. Ma Y, Cui W, Yang J, et al. SALL4, a novel oncogene, is constitutively expressed in human acute myeloid leukemia (AML) and induces AML in transgenic mice. *Blood.* 2006;108(8):2726-2735.

7. Fujimoto M, Sumiyoshi S, Yoshizawa A, et al. SALL4 immunohistochemistry in non-small-cell lung carcinomas. *Histopathol.* 2014;64(2):309-311.

8. Forghanifard M, Moghbeli M, Raeisossadati R, et al. Role of SALL4 in the progression and metastasis of colorectal cancer. *J Biomed Sci.* 2013;20(1):6.

9. Habano W, Sugai T, Jiao YF, Nakamura S. Novel approach for detecting global epigenetic alterations associated with tumor cell aneuploidy. *Int J Cancer.* 2007;121(7):1487-1493.

452 10. Liu L, Zhang J, Yang X, Fang C, Xu H, Xi X. SALL4 as an Epithelial-Mesenchymal
453 Transition and Drug Resistance Inducer through the Regulation of c-Myc in
454 Endometrial Cancer. *PLoS One.* 2015;10(9):e0138515.

455 11. Zhang L, Yan Y, Jiang Y, et al. The expression of SALL4 in patients with gliomas: high
456 level of SALL4 expression is correlated with poor outcome. *J Neuro-Oncol.*
457 2015;121(2):261-268.

458 12. Singhi AD, Seethala RR, Nason K, et al. Undifferentiated carcinoma of the esophagus:
459 a clinicopathological study of 16 cases. *Hum Pathol.* 2015;46(3):366-375.

460 13. Zhang L, Xu Z, Xu X, et al. SALL4, a novel marker for human gastric carcinogenesis
461 and metastasis. *Oncogene.* 2014;33(48):5491-5500.

462 14. Zeng SS, Yamashita T, Kondo M, et al. The transcription factor SALL4 regulates
463 stemness of EpCAM-positive hepatocellular carcinoma. *J Hepatol.* 2014;60(1):127-
464 134.

465 15. Gao C, Dimitrov T, Yong KJ, et al. Targeting transcription factor SALL4 in acute
466 myeloid leukemia by interrupting its interaction with an epigenetic complex. *Blood.*
467 2013;121(8):1413-1421.

468 16. Yong KJ, Gao C, Lim JSJ, et al. Oncofetal Gene SALL4 in Aggressive Hepatocellular
469 Carcinoma. *N Engl J Med.* 2013;368(24):2266-2276.

470 17. Liu BH, Jobichen C, Chia CSB, et al. Targeting cancer addiction for SALL4 by shifting
471 its transcriptome with a pharmacologic peptide. *Proc Natl Acad Sci U S A.*
472 2018;115(30):E7119-e7128.

473 18. Itou J, Matsumoto Y, Yoshikawa K, Toi M. Sal-like 4 (SALL4) suppresses CDH1
474 expression and maintains cell dispersion in basal-like breast cancer. *FEBS letters.*
475 2013;587(18):3115-3121.

19. Itou J, Tanaka S, Li W, et al. The Sal-like 4 - integrin alpha6beta1 network promotes cell migration for metastasis via activation of focal adhesion dynamics in basal-like breast cancer cells. *BBA-Mol Cell Res.* 2017;1864(1):76-88.

20. Matsumoto Y, Itou J, Sato F, Toi M. SALL4 - KHDRBS3 network enhances stemness by modulating CD44 splicing in basal-like breast cancer. *Cancer Med-US.* 2018;7(2):454-462.

21. Kobayashi D, Kuribayshi K, Tanaka M, Watanabe N. SALL4 is essential for cancer cell proliferation and is overexpressed at early clinical stages in breast cancer. *Int J Oncol.* 2011;38(4):933-939.

22. Chen YY, Li ZZ, Ye YY, et al. Knockdown of SALL4 inhibits the proliferation and reverses the resistance of MCF-7/ADR cells to doxorubicin hydrochloride. *BMC Mol Biol.* 2016;17:6-6.

23. Tsang JYS, Lee MA, Ni YB, et al. Amyloid Precursor Protein Is Associated with Aggressive Behavior in Nonluminal Breast Cancers. *The oncologist.* 2018;23(11):1273-1281.

24. Hu Y, Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods.* 2009;347(1-2):70-78.

25. Pereira B, Chin SF, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun.* 2016;7:11479.

26. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6(269):pl1.

27. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform

for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401-404.

28. Gyorffy B, Lanczky A, Eklund AC, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat.* 2010;123(3):725-731.

29. Oikawa T, Kamiya A, Zeniya M, et al. Sal-like protein 4 (SALL4), a stem cell biomarker in liver cancers. *Hepatology (Baltimore, Md).* 2013;57(4):1469-1483.

30. Jeong HW, Cui W, Yang Y, et al. SALL4, a stem cell factor, affects the side population by regulation of the ATP-binding cassette drug transport genes. *PLoS One.* 2011;6(4):e18372.

31. !!! INVALID CITATION !!!

32. Joshi PA, Jackson HW, Beristain AG, et al. Progesterone induces adult mammary stem cell expansion. *Nature.* 2010;465(7299):803-807.

33. Li A, Jiao Y, Yong KJ, et al. SALL4 is a new target in endometrial cancer. *Oncogene.* 2015;34(1):63-72.

34. Mani SA, Guo W, Liao MJ, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell.* 2008;133(4):704-715.

35. Wellner U, Schubert J, Burk UC, et al. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat Cell Biol.* 2009;11(12):1487-1495.

36. Zhou W, Zou B, Liu L, et al. MicroRNA-98 acts as a tumor suppressor in hepatocellular carcinoma via targeting SALL4. *Oncotarget.* 2016;7(45):74059.

522 37. He J, Zhou M, Chen X, et al. Inhibition of SALL4 reduces tumorigenicity involving
523 epithelial-mesenchymal transition via Wnt/β-catenin pathway in esophageal squamous
524 cell carcinoma. *J Exp Clin Cancer Res.* 2016;35(1):98.

525 38. Sun C, Lan P, Han Q, et al. Oncofetal gene SALL4 reactivation by hepatitis B virus
526 counteracts miR-200c in PD-L1-induced T cell exhaustion. *Nat Commun.*
527 2018;9(1):1241.

528 39. Virtakoivu R, Mai A, Mattila E, et al. Vimentin-ERK Signaling Uncouples Slug Gene
529 Regulatory Function. *Cancer Res.* 2015;75(11):2349-2362.

530 40. Vuoriluoto K, Haugen H, Kiviluoto S, et al. Vimentin regulates EMT induction by Slug
531 and oncogenic H-Ras and migration by governing Axl expression in breast cancer.
532 *Oncogene.* 2011;30(12):1436-1448.
533

Acknowledgements

We acknowledge the cBioPortal for Cancer Genomics (http://www.cbioportal.org/), KM-plotter (http://kmplot.com/analysis) and Oncomine database (https://www.oncomine.org/resource/) for providing the breast cancer data set and analysis. We also acknowledge the University Research Facility in Life Sciences (ULS) in Hong Kong Polytechnic University for providing access to advanced instruments required for the research work in life sciences. Sincere thanks are given to PolyU studentship for their support for this project. This project was partly supported by the Area of Strategic Importance grant from the Hong Kong Polytechnic University (1-ZE22).

Disclosure Statement:

The authors declare no potential conflicts of interest.

Supporting information

Table S1 Antibodies used for IHC analysis

Table S2 Summary of SALL4 level in primary tumor and LN

Table S3 Correlation between SALL4 and vimentin in luminal and TNBC cases.

Figure S1 The specificity of antibody raised against SALL4 peptide was verified.

Figure S2 Cell proliferation assay was conducted after SALL4 was knocked down and overexpressed in MCF7.

554 **Figure legends**

555 **Figure 1 SALL4 was overexpressed in breast cancer and associated with**
556 **breast cancer patients' survival.**

557 (A) The study[25] from cBioPortal for Cancer Genomics showed the genetic alterations of
558 SALL4 in breast cancer (n=1904). Data were obtained from http://cbioportal.org (B) Oncomine
559 database demonstrated expression of SALL4 in invasive breast carcinoma and normal breast
560 tissue, n=593, p=7.18E-18. (C) Correlation between SALL4 with overall survival of global
561 breast cancer patients. n=1402, Logrank Test, p-value: 0.01, http://kmplot.com/analysis/. (D)
562 Immunohistochemical staining of Hong Kong patients' tumor section with low (left) and high
563 (right) SALL4 expression (n=371). An immunoscore for SALL4 was obtained by multiplying
564 staining intensity with percentage of staining nuclei. The median immunoscore of 70 was used
565 to divide the low (<70) and high (≥70) expression groups. Magnification: x200. (E) Kaplan-
566 Meier analysis of 89 cases with SALL4 staining results from paired nodal metastasis (LN) and
567 primary tumor (Tumor) of breast cancer patients in Hong Kong. DFS: time to first tumor
568 relapse; BCSS: breast cancer specific survival.

569 **Figure 2 Knockout of SALL4 inhibited cell proliferation, colony formation,**
570 **mammosphere formation and migration of MDA-MB-231-Red-FLuc-GFP**
571 **cells.**

572 (A) SALL4 protein level in three different breast cancer cell lines and $CD44^{+}CD24^{-}$ MCF7
573 cancer stem cells. Cells were harvested and lysed with RIPA buffer for western blot analysis.
574 (B) Western blot analysis of SALL4-knockout MDA-MB-231-Red-FLuc-GFP clones. WT:
575 wild-type. Cells infected with scrambled CRISPR-Cas9 lentivirus were marked as -Vec. KO#1

and KO#2 were SALL4-knockout single clones. NF90 was a nuclear loading control. (C) Knockout of SALL4 inhibited proliferation of KO#2 measured by MTS assay. Three thousand cells were seeded in each well of a 96-well plate in 100 μL RPMI medium. The CellTiter 96 AQueous Assay (Promega) was used to measure cell proliferation according to the manufacturer's instructions on day 1, 3, 5 after seeding. (D) Silence of SALL4 inhibited colony formation of MDA-MB-231-Red-FLuc-GFP cells. left: images of colony; right: quantification of colony numbers. (E) Mammosphere formation ability of MDA-MB-231-Red-FLuc-GFP cells was suppressed after SALL4 was knocked out. Representative images of mammosphere formation were shown (top). The number of mammospheres with diameter≥100 μm was quantified (bottom). Scale bar = 100 μm. (F) Wound healing ability was suppressed after SALL4 was knockout. Images of wounds at 0 h and 36 h were shown (top); Wound healing ability was represented as percentages of the unhealed scratch area at 36 h divided by that at 0 h (bottom). (G) Trans-well assay showed that knockout of SALL4 inhibited cell migration. Representative images of cells at 16 h after incubation were shown (left). The cell number was counted using Image J software. Cell migration ability of each group was presented as percentage normalized to wild type group (right). All experiments were repeated in three replicates and one of them was shown.

Figure 3 Overexpression of SALL4 promoted mammosphere formation and migration of MDA-MB-231-Red-FLuc-GFP cells.

(A) Expression of SALL4 protein after overexpression assay. Lentivirus was used to overexpress V5-tagged SALL4 in MDA-MB231-Red-FLuc-GFP cells. Two single colonies (B8 and G3) were picked for further analysis. WT: wild type; +Vec: vector control. Lamin B was used as the nuclear loading control. (B) Cell growth was not affected after overexpression of SALL4 indicated by MTS assay. Three thousand cells were seeded in each well of a 96-well

plate in 100 μL RPMI medium. The CellTiter 96 AQueous Assay (Promega) was used to measure cell proliferation according to the manufacturer's instructions on day 1, 3, 5 after seeding. (C) There was no significant enhancement in colony formation ability after SALL4 was overexpressed. left: images of colony, right: quantification of colony number. (D) Overexpression of SALL4 in B8 and G3 can promote mammosphere formation. Representative images of mammospheres, scale bars, 100 μm (left). Number of mammospheres with a diameter ≥100 μm was enumerated (right). (E) Overexpression of SALL4 promoted cell migration measured in wound healing assay. Images of wounds at 0 h and 36 h were shown (left); Wound healing ability was represented as percentages of the unhealed scratch area at 36 h divided by that at 0 h (right). (F) Trans-well assay showed that the cells penetrated the insert membrane faster in SALL4-overexpressed mutants B8 and G3. Representative images of cells at 16 h after incubation were shown (left). Cell migration ability of each group was presented (right). All experiments were repeated in three replicates and one of them was shown.

Figure 4 Effect of SALL4 on tumorigenicity, tumor growth and metastasis.

(A) Tumorigenicity of SALL4-knockout clones in nude mice (n=3). On day 6 after inoculation, mice bearing tumor were recorded and TIC frequency was calculated by http://bioinf.wehi.edu.au/software/elda/ . (B) The tumor growth curve over 24 days was shown. A total of 10^6 cells (-Vec, KO#1 and KO#2) in 1:1 PBS/Matrigel were injected subcutaneously in the right flank of female Balb/c nude mice (n=3). (C) Tumors derived from -Vec, KO#1 and KO#2 cells with three different inoculation density (10^6, $5x10^6$, 10^7 cells) were excised on 24 days post implantation. (D) Calculated tumor volume from (C) was shown. Tumor volume was calculated using the formula (A) (B^2) /2, where A was the length of the longest aspect of the tumor, and B was the length of the tumor perpendicular to A. (E) Tumorigenicity of SALL4-overexpressed clones in nude mice. On day 6 after inoculation, mice bearing tumor were

recorded and TIC frequency was calculated by http://bioinf.wehi.edu.au/software/elda/ . (F) The tumor growth curve over 24 days was shown. A total of $5x10^6$ cells (+Vec, B8 and G3) in 1:1 PBS/Matrigel were injected subcutaneously in the right flank of female Balb/c nude mice (n=3). (G) Tumors derived from +Vec, B8 and G3 cells with three inoculation density (10^6, $5x10^6$, or 10^7 cells) on 24 days-post implantation were excised. (H) Calculated tumor volume from (G) was shown. (I) Metastasis represented by bioluminescence signal was shown after SALL4 was knockout. SALL4-knockout MDA-MB-231-Red-FLuc-GFP cells (KO#2, n=5) and its vector control (-Vec, n=6) were inoculated intravenously into the tail vein of female Balb/c nude mice. Bioluminescence signal was recorded by IVIS (Lumina series III, Perkin-Elmer) (left). Quantification of bioluminescence signal at indicated times was shown (right). (J) Lung metastasis was shown. *Ex vivo* imaging of lungs from -Vec and KO#2 mice was presented (top). Bioluminescence signal in lung was quantified (bottom). (K) Pictures depicting the tumor nodules in the lungs of -Vec or KO#2 mice were shown. Fresh lungs of mice were excised and fixed with neutral buffered formalin before pictures were taken. (L) Representative lung metastases indicated by the arrowheads and red circle were shown. Three layers of lung sections were sectioned and each was 60 micron apart. (M) Size of each foci (left) and average number for each group (-Vec, n=6; KO#2, n=5) (right) was presented. All the focis on the slides were counted and the size for each was measured roughly on the IHC stained sections.

Figure 5 Effect of SALL4 on EMT

(A)Expression of proteins involved in EMT after SALL4 was knockout. WT, -Vec or KO#2 cells were inoculated subcutaneously in the flank of female Balb/c nude mice. Tumors were excised and homogenized in RIPA lysis buffer for western analysis of EMT markers. (B) Expression of proteins involved in EMT after SALL4 was overexpressed. WT, +Vec and B8

cells were inoculated as shown in (A) and excised tumors were analyzed by western blot. (C) mRNA of vimentin was increased after SALL4 was overexpressed. The experiments were repeated in three replicates and one of them was shown. ** in the Figures represents $p<0.01$. (D) Three sets of primers were designed for ChIP assays of vimentin targeting 3 regions: VIM 1 (-778 to -550 bp), VIM 2 (-409 to -273bp), VIM 3 (+186 to +435bp). (E) SALL4 bound to VIM1 region in +Vec and B8 cells indicated by ChIP assay. Rabbit IgG antibody was used as negative control. (F) Fold enrichment of VIM1 region in B8 was 2.1-fold higher than that in vector control. (G) Mean of vimentin-stained cell percentage in SALL4-high and SALL4-low cases was 4.49% and 1.22% respectively in 287 cases of luminal breast cancer ($p=0.043$).

Supplementary Figure legends

Figure S1 The specificity of antibody raised against SALL4 peptide was verified.

Different subcellular components of MDA-MB-231-Red-FLuc-GFP cells were subjected for western blot analysis using peptide-derived SALL4 antibody. SALL4 was enriched in nuclear extraction with a size around 140kDa. Lamin B was used as nuclear loading control. SALL4 peptide sequence was: KDCHRENGGSSEDM.

Figure S2 Cell proliferation assay was conducted after SALL4 was knocked down and overexpressed in MCF7.

(A) SALL4 protein level was reduced after knockdown. MCF7 cells were transfected with SALL4 siRNA with lipofectamine 3000 (Thermo Fisher) according to the instructions of manufacturer. At 48h post-transfection, cells were harvested and lysed by RIPA buffer for western blot analysis. (B) Knockdown of SALL4 inhibited cell proliferation. Seven thousand

cells were seeded in each well of a 96-well plate one day before transfection. The CellTiter 96 AQueous Assay (Promega) was used to measure cell proliferation according to the manufacturer's instructions on day 1, 3, 4, 5 after transfection. (C) SALL4 protein level was increased after overexpression. MCF7 cells were transfected with SALL4 plasmid (DNASU, clone: HsCD00443167) using lipofectamine 3000 (Thermo Fisher) according to the instructions of manufacturer. At 48h post-transfection, cells were harvested and lysed by RIPA buffer for western blot analysis. (D) Overexpression of SALL4 did not affect cell proliferation. Five thousand cells were seeded in each well of a 96-well plate one day before transfection. The CellTiter 96 AQueous Assay (Promega) was used to measure cell proliferation according to the manufacturer's instructions on day 1, 3, 4, 5 after transfection. All experiments were repeated in three replicates and one of them was shown. ** and *** in the Figures represent $p<0.01$ and $p<0.001$, respectively.