CANCER

Human iPSC-based breast cancer model identifies S100P-dependent cancer stemness induced by *BRCA1* mutation

Jingxin Liu[1,2,3,4,5,6,7]†, Cai Zhao[1,2,3,4,7,8]†, Jiahao Chen[1,2,3,4,5,7]†, Pengguihang Zeng[1,2,3,4,7]†, Qingjian Li[9], Ranran Dai[1,2,3,4], Xingqiang Lai[10], Wenqian Song[11,12], Jianing Chen[13,14], Xixi Zhu[15], Xinyi Liu[1,2,3,4], Jun Sun[1,2,3,4], Jia Wang[1,3,4], Peihang Fang[1,2,3], Tengfei Wang[1,2,3], Wenjie Chen[5], Diana Guallar[16], Nan Cao[10], Jianli Zhao[9], Shicheng Su[13,14], Andy Peng Xiang[10], Yi Arial Zeng[11,12], Jie Li[17], Junchao Cai[18], Dung-Fang Lee[19,20], Jinxin Bei[21], Yongliang Huo[15], Hai Hu[13,14], Shengbao Suo[6,22]*, Dong-feng Huang[1]*, Jin Bai[2]*, Junjun Ding[1,2,3,4,7]*

Breast cancer is the most common malignancy in females and remains the leading cause of cancer-related deaths for women worldwide. The cellular and molecular basis of breast tumorigenesis is not completely understood partly due to the lack of human research models which simulate the development of breast cancer. Here, we developed a method for generating functional mammary-like cells (MCs) from human-induced pluripotent stem cells (iPSCs). The iPSC-MCs closely resemble human primary MCs at cellular, transcriptional, and functional levels. Using this method, a breast cancer model was generated using patient-derived iPSCs harboring germline *BRCA1* mutation. The patient iPSC-MCs recapitulated the transcriptome, clinical genomic alteration, and tumorigenic ability of breast cancer cells. We also identified S100P as an oncogene downstream of mutated *BRCA1* that promotes cancer cell stemness and tumorigenesis. Our study establishes a promising system of breast cancer for studying the mechanism of tumorigenesis and identifying potential therapeutic targets.

INTRODUCTION

The high incidence of tumor occurrence and metastasis is an important issues in the field of breast cancer research (*1*, *2*). Breast cancer metastasis has been extensively explored relying on patient-derived cancer models, such as patient-derived cell lines, patient-derived organoids, and patient-derived xenografts (*3*–*6*). The investigation on early occurrence of breast tumors also accumulated a large amount of clinical data in terms of genetic or nongenetic factors associated with the carcinogenesis thanks to the advance of omic technologies (*7*–*10*). Although many gene mutations have been found to be correlated with the occurrence of breast cancer, few mutated genes have been validated for genotype-phenotype relationship, and the mechanism of carcinogenesis caused by these mutations is even less explored in human tumorigenesis system (*10*). One reason is the lack of human-derived models for early stage of genesis of breast cancer (*4*). Therefore, a human-derived system model that can simulate breast tumorigenesis is urgently needed.

BRCA1 mutation is strongly associated with the concurrence of triple-negative breast cancer (TNBC) (*11*), which is the most severe clinical subtype and is characterized by the related pathological features including an excess of stemness activation, higher histologic grade, and a medullary histopathology (*11*–*13*). Although new therapeutic strategies including poly(ADP-ribose) polymerase inhibitors and immunotherapy have been used to treat TNBC in recent years, the survival rate for patient, especially in the case of metastatic relapse, is still low (*14*). The study of the molecular mechanism of tumorigenesis for TNBC is critical for the development of early diagnostic and therapeutic targets for breast cancer.

[1]Department of Rehabilitation Medicine, The Seventh Affiliated Hospital, Zhongshan School of Medicine, Sun Yat-sen University, Shenzhen, Guangdong, 518107, China. [2]Cancer Institute, Xuzhou Medical University, 221002, Jiangsu, China. [3]Frontiers Medical Center, Tianfu Jincheng Laboratory, Department of Gynecology and Obstetrics, West China Second Hospital, West China Biomedical Big Data Center, West China Hospital/West China School of Medicine, Sichuan University, Chengdu 610041, China. [4]School of Basic Medical Sciences, Guangdong Medical University, Advanced Medical Technology Center, The First Affiliated Hospital, Center for Stem Cell Biology and Tissue Engineering, Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong, China. [5]Vaccine Research Institute, The Third Affiliated Hospital of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, 510630, China. [6]Guangzhou National Laboratory, Guangzhou, 510005, China. [7]Shenzhen Eye Hospital, Shenzhen Eye Medical Center, Southern Medical University, 18 Zetian Road, Futian District, Shenzhen, 518040, China. [8]Sunshine Lake Pharma Co. Ltd, Dongguan, 523871, China. [9]Department of Oncology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120, China. [10]Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510080, China. [11]State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Institute of Biochemistry and Cell Biology, Shanghai, 200031, China. [12]Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, 200031, China. [13]Breast Tumor Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120, China. [14]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, 510120, China. [15]School of Basic Medical Sciences, Guangzhou Medical University, Guangzhou, 511436, China. [16]Center for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela (USC)-Health Research Institute (IDIS), Santiago de Compostela, 15782, Spain. [17]Department of Breast and Thyroid Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510275, China. [18]Department of Immunology, Sun Yat-sen University Zhongshan School of Medicine, Guangzhou, 510080, China. [19]Department of Integrative Biology and Pharmacology, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA. [20]The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA. [21]Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060 China. [22]State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, 510120, China.
*Corresponding author. Email: suo_shengbao@gzlab.ac.cn (S.S.); huangdf@mail.sysu.edu.cn (D.-f.H.); bj@xzhmu.edu.cn (J.B.); dingjunj@mail.sysu.edu.cn (J.D.)
†These authors contributed equally to this work.

Human-induced pluripotent stem cells (iPSCs) are suitable for modeling of various human diseases, since they have the patient-specific genetic background and can be indefinitely expanded as differentiated to particular lineages in vitro, which may at least partly replicate organ development and disease pathogenesis (*15*, *16*). Thus, iPSCs can be used to construct systems that mimic tumor initiation and progression and provide a large number of samples in a repertoire of human genetic backgrounds (*17*). However, the application of iPSCs in breast cancer research is limited due to the lack of feasible methods to obtain functional mammary cells from the iPSCs (*18*). The human iPSC-based breast cancer model simulating the occurrence of tumor has not been established.

In the present study, we developed a method to differentiate human iPSCs to the functional mammary-like cells and used this approach to establishing a *BRCA1* mutation–associated breast cancer model. The *BRCA1*-mutant iPSC-derived mammary cells recapitulated the transcription profile, clinical genomic alteration, and tumorigenesis of breast cancer. By integrating global transcriptional and computational analyses, we demonstrated that *BRCA1* mutation enhanced the stemness of mammary cells through S100P, which is critical for the tumorigenic capacity induced by *BRCA1* mutation, and could be a promising therapeutic target in breast cancer. This iPSC-derived breast cancer model can be used for further mechanistic studies, discovery of novel therapeutic targets, and screening for drug candidates.

RESULTS

Development of a method for generating iPSC-derived mammary cells with in vitro and in vivo functions

We developed a protocol to differentiate human iPSCs into functional mammary cells (iPSC-MCs) based on the knowledge that mammary gland originates from epidermal ectoderm in human embryonic development (Fig. 1A) (*19*). The iPSCs were first differentiated to non-neural ectoderm cells, which were then induced to iPSC-MCs that exhibited the morphology of primary mammary cells (primary-MCs) (Fig. 1B), and expressed the specific markers associated with the different stages of mammary differentiation (fig. S1, A and B). Furthermore, the iPSC-MCs also expressed the myoepithelial cell marker and the luminal epithelial cell marker that are characteristic of primary-MCs (Fig. 1C, and fig. S1, C and D). Single-cell RNA sequencing (RNA-seq) further validated the existence of these two subpopulations and revealed that the proportion of these iPSCs differentiating into mammary epithelial cells was ~92.0% (fig. S2, A to F). Principal component analysis (PCA) of the transcriptomic profiles showed that the iPSC-MCs and primary-MCs clustered together and were distinct from other cell types (Fig. 1D). Tissue-specific gene ontology (GO) enrichment analysis further indicated that the expression profile of iPSC-MCs at day 40 (D40) was close to that of the mammary gland (fig. S1E). These results suggest that the iPSC-MCs are transcriptionally similar to the primary-MCs. To functionally characterize the iPSC-MCs, the cells were subjected to three-dimensional (3D) culture to generate mammary-like organoids (*20*). As shown in fig. S1F, iPSC-MCs formed branched mammary-like structures after 15 days of 3D culture. These organoids expressed the milk protein β-casein in response to prolactin treatment (Fig. 1, E and F), indicating that the iPSC-MCs simulated primary-MCs functionally as well. Together, we successfully differentiated human iPSCs into mammary-like cells, which are transcriptionally and in vitro functionally similar to primary-MCs.

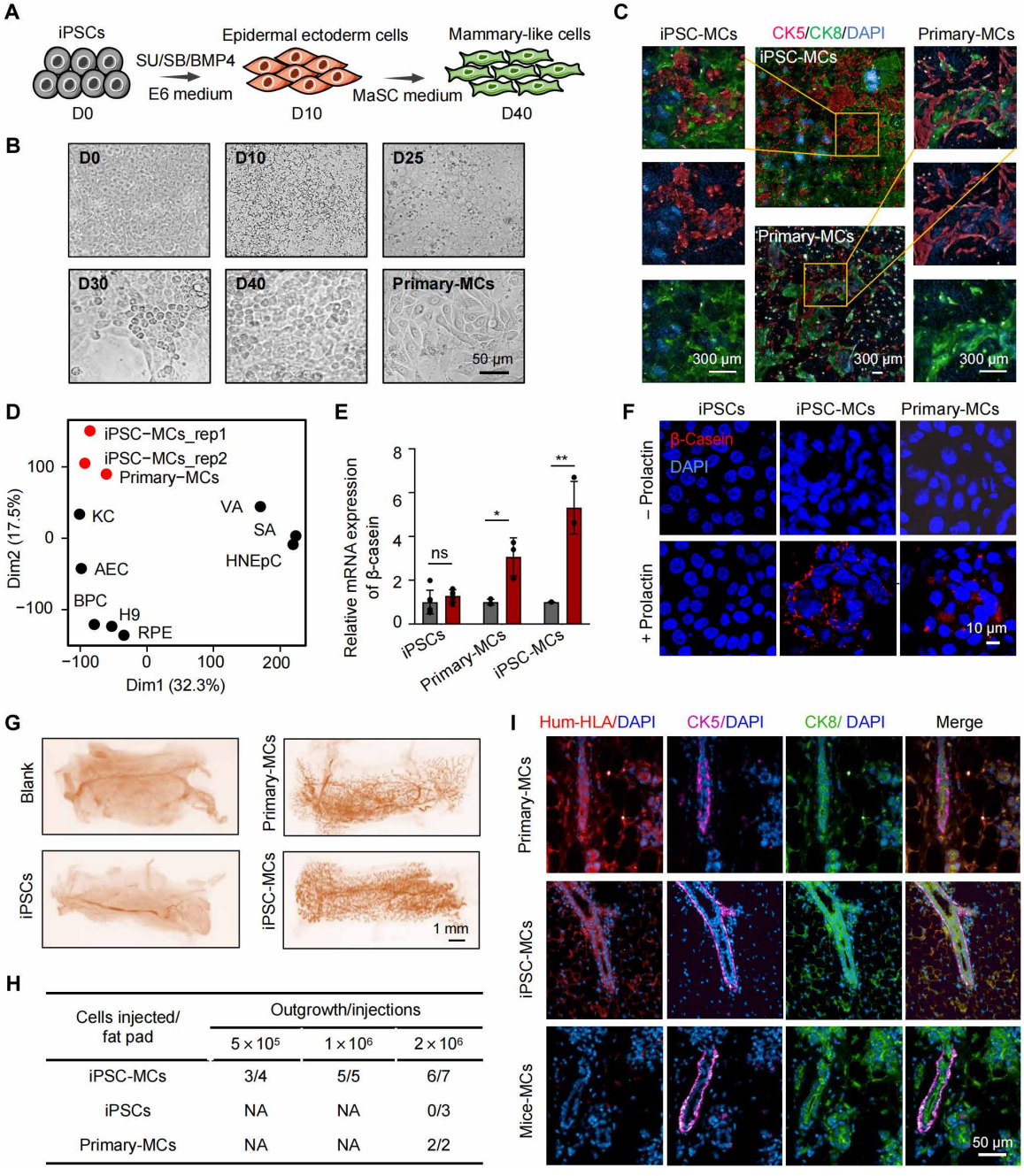
To further evaluate the in vivo function of iPSC-MCs, we performed mammary gland reconstruction in mice, which can be used for analyzing the in vivo regenerative potential of human primary-MCs (*21*). Briefly, the endogenous rudimentary mammary epithelium was removed from 4-week-old nonobese diabetic (NOD)/severe combined immunodeficient (SCID) female mice, and human breast fibroblasts were injected into the cleared fat pads for 2 weeks. Last, the iPSC-MCs and primary-MCs were implanted into the humanized fat pad to regenerate human mammary tissues. Both iPSC-MCs and primary-MCs successfully formed ductal outgrowths with side branching, while the iPSCs did not form any outgrowths or formed teratoma (Fig. 1, G and H). Furthermore, the two-layer mammary ductal structures expressed specific human leukocyte antigen–ABC (HLA-ABC) and consisted of an outer layer of myoepithelial cells (CK5) and an inner layer of luminal epithelial cells (CK8) (Fig. 1I). Together, the iPSC-MCs formed human mammary-like structure in vivo similar to the primary-MCs.

BRCA1 mutation induces tumorigenesis in the iPSC-based disease model

To establish a *BRCA1* mutation–associated breast cancer model, we first generated iPSCs from the monocytes of family with hereditary breast cancer, one of whom had developed breast cancer, and the other is healthy nonmutation carrier [wild type (WT)] (fig. S3A). This patient has a heterozygous c.2253_2254delGT mutation (Fig. 2B), which is a previously reported hotspot mutation that causes BRCA1 truncation (fig. S3B) and is responsible for the triple-negative phenotype of these patients (*22*). The *BRCA1*-mutant and WT iPSC clones were generated by delivering *OCT4*, *SOX2*, *KLF4*, and *MYC* (OSKM) using nonintegrating Sendai virus (SeV) (*15*). The endogenous pluripotency genes were activated, while the exogenous transgenes were silenced in the iPSCs (figs. S3, C and E, and S4A). In addition, the iPSC lines were karyotypically normal (fig. S4B) and were able to differentiate into all three germ layers in teratomas (fig. S4C). This indicates that somatic cells from the *BRCA1*-mutant patient and healthy member can be reprogrammed into iPSCs that maintain pluripotency and differentiation potential.

To obtain a genetically identical control, the heterozygous *BRCA1* mutation in these patient-derived iPSCs was reversed using the CRISPR-Cas9 tool (fig. S3D) and confirmed by genome sequencing (Fig. 2B). *BRCA1*-mutant correction did not affect the pluripotency and differentiation potential of iPSCs (figs. S3E and S4, A to C).

To determine whether the mammary cells obtained from patient-derived iPSCs can recapitulate the tumorigenic phenotype, we transplanted the WT, *BRCA1*-mutant, and *BRCA1*-corrected iPSC-MCs in NOD/SCID mice at D40 (Fig. 2A and fig. S5A). Only the *BRCA1*-mutant cells exhibited subcutaneous tumorigenic ability in vivo (Fig. 2, C to E, and fig. S5B), and the tumors derived from mutant iPSC-MCs were histologically similar to primary breast tumors rather than normal breast tissues (Fig. 2F). Furthermore, the iPSC-MCs-tumor expressed the cell proliferation markers P120 and Ki67 (Fig. 2G), and their transcriptomic profile was similar to primary breast tumors compared to other tumor types (Fig. 2H). Last, we found the *BRCA1*-mutant iPSC-MCs-tumors lackedestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), which is characteristic of TNBC (Fig. 2I). Together, *BRCA1*-mutant iPSC-MCs can give rise to neoplasms of the breast lineage that are histopathologically similar to the TNBC subtype and thus can be used to recapitulate breast tumorigenesis.



Cells injected/ fat pad	Outgrowth/injections		
	5×10^5	1×10^6	2×10^6
iPSC-MCs	3/4	5/5	6/7
iPSCs	NA	NA	0/3
Primary-MCs	NA	NA	2/2

Fig. 1. Establishment of method to generate iPSC-derived mammary cells with function in vitro and in vivo. (**A**) Schematic representation of the differentiation protocol from D0 to D40. MaSC, mammary stem cells. (**B**) Representative images of the cell morphology during mammary differentiation. Scale bar, 50 μm. (**C**) Immunofluorescence staining of basal cell marker (CK5, red), luminal cell marker (CK8, green), and 4′,6-diamidino-2-phenylindole (DAPI; blue) in iPSC-MCs and primary-MCs. Scale bars, 300 μm. (**D**) PCA was performed to compare the expression profiles of iPSC-MCs derived from normal human iPSCs and cells of other tissues of ectodermal origin, including primary-MCs, keratinocytes (KC), visceral adipose (VA), subcutaneous adipose (SA), retinal pigment epithelial (RPE), airway epithelial cells (AEC), primary brain pericytes (BPC), primary human nasal epithelial cells (HNEpC), and human embryonic stem cell line (H9). Data used in this analysis are shown in table S6. (**E**) RT-qPCR assay for expression of β-casein in iPSCs, primary-MCs, and iPSC-MCs with and without prolactin treatment. Error bars represent ±SD; *t* test, two-tailed; $n = 6$ (iPSCs), $n = 3$ (iPSC-MCs and primary-MCs) distinct samples. *P* values (left to right): $P = 0.6046$, $P = 0.0379$, and $P = 0.0033$. (**F**) Immunofluorescence staining for β-casein (red) and DAPI (blue) in iPSCs (left) and organoids derived from iPSC-MCs (middle) and primary-MCs (right) with and without prolactin treatment. Scale bar, 10 μm. (**G**) Carmine-stained whole-mount examination of outgrowth from xenografts in humanized fat pads inoculated with iPSC-MCs, iPSCs, and primary-MCs. Scale bar, 1 mm. (**H**) Number of outgrowths generated in NOD-SCID mouse fat pads inoculated with cells from different subpopulations. (**I**) Immunofluorescence staining for human leukocyte antigen–ABC (HLA-ABC, red), myoepithelial cell marker (CK5, purple), luminal epithelial cell marker (CK8, green), and DAPI (blue) in the two-layer ducts generated by primary-MCs (top) and iPSC-MCs (middle) and in mouse mammary tissue (bottom). Scale bar, 50 μm. ns, not significant; NA, not available.

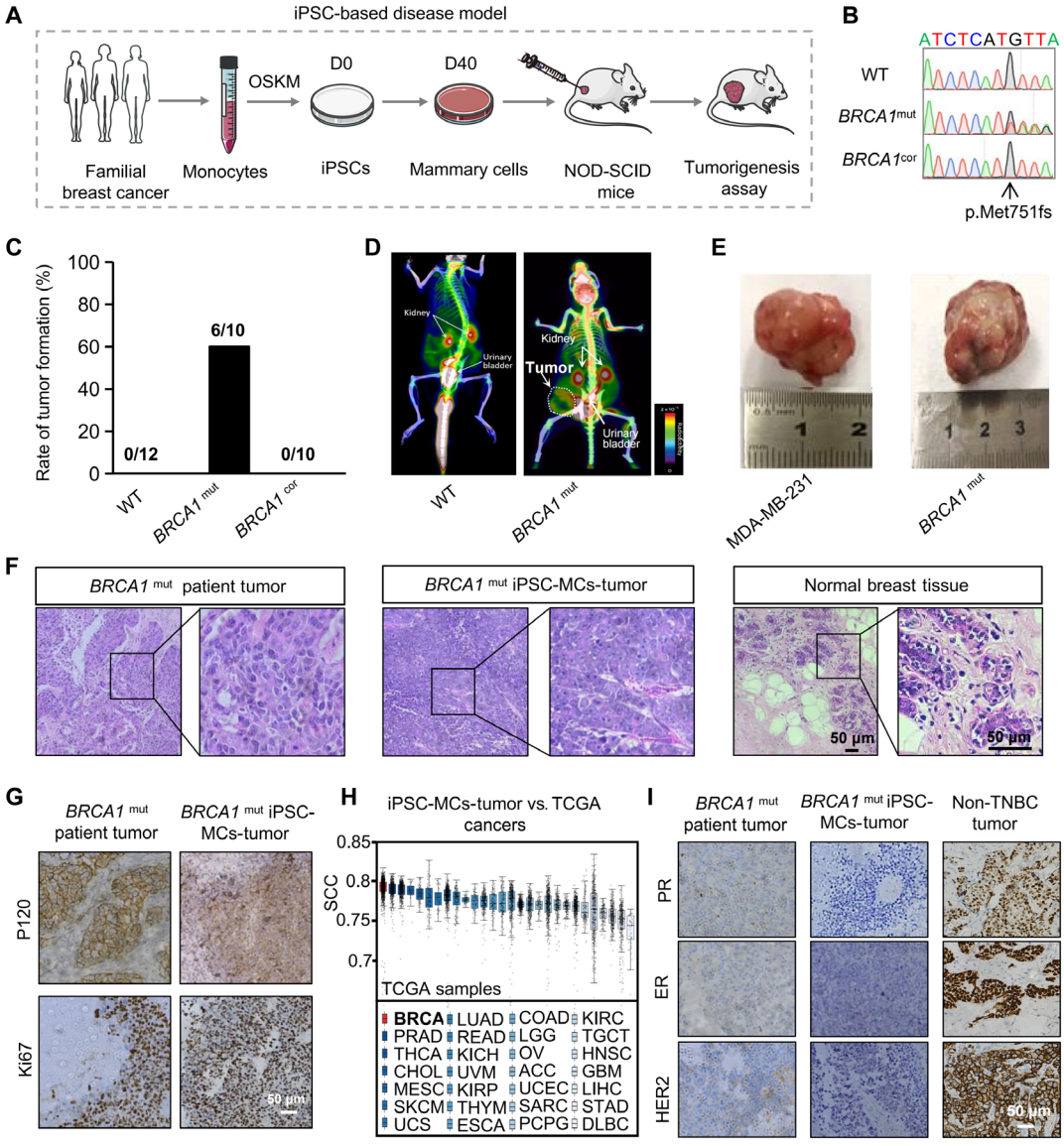


Fig. 2. *BRCA1* mutation induces tumorigenesis in the iPSC-based disease model. (**A**) Schematic overview of patient-derived iPSC-mammary cell subcutaneous injection into NOD-SCID mice. OSKM, *OCT4, SOX2, KLF4, and MYC*. (**B**) Sequencing map showing the exact site of *BRCA1* mutation (*BRCA1*^{mut}) and the correction of the *BRCA1* mutation (*BRCA1*^{cor}). (**C**) Tumor xenograft experiments by subcutaneous injection into mice demonstrate that patient-derived iPSC-MC but not WT and mutant-corrected iPSC-MC recapture in vivo tumorigenic ability. (**D**) Representative 18F-FDG PET/CT images of mice injected with WT iPSC-MC and patient-derived iPSC-MC. (**E**) Representative images of actual tumors formed by MAB-231 and patient-derived iPSC-MC. (**F**) Representative hematoxylin and eosin (H&E) staining of patient tumor tissue, tumor tissue formed by patient-derived iPSC-MC, and normal breast tissue. Scale bar, 50 μm. The right images show higher magnification of the cell. Scale bar, 50 μm. (**G**) Representative Ki67 and P120 staining of patient tumor tissue and tumor tissue formed by patient-derived iPSC-MC. Scale bar, 50 μm. (**H**) PCCs were measured between the iPSC-MCs-tumor and TCGA cancer samples. TPMs of all genes for iPSC-MCs-tumor and top expressed gene for TCGA cancer samples are used for analysis. Cancer types abbreviated per TCGA Cancer Codes. (**I**) Representative ER, PR, and HER2 staining of patient tumor tissue and tumor tissue formed by patient-derived iPSC-MC. Scale bar, 50 μm.

BRCA1-mutant iPSC-MCs exhibit the *BRCA1*-related breast cancer gene signature

To further validate this model and track the tumorigenesis process, we performed global transcriptome analysis of the iPSCs of two *BRCA1*-mutant clones, one mutant-corrected and one WT clones at the distinct stages of mammary lineage differentiation by mRNA-seq. Pearson correlation coefficients (PCCs) showed that the gene expression profiles of *BRCA1*-mutant and mutant-corrected or WT iPSCs were mainly clustered together at D0 to D20 (Fig. 3A), indicating a similar transcriptome profile at the initial stage of differentiation. However, their profiles started to diverge in the differentiation stages of D30 and D40 (Fig. 3A), indicating that *BRCA1* mutation either interrupted the normal differentiation process or activated an oncogenic signature, which is further suggestive of a correlation between mammary development and tumorigenesis. The iPSC-MCs-tumors showed a greater PCC with *BRCA1*-mutant compared to mutant-corrected or WT iPSC-MCs (fig. S6A). The above results also indicated that the transcriptional variations between clones showed comparatively minor.

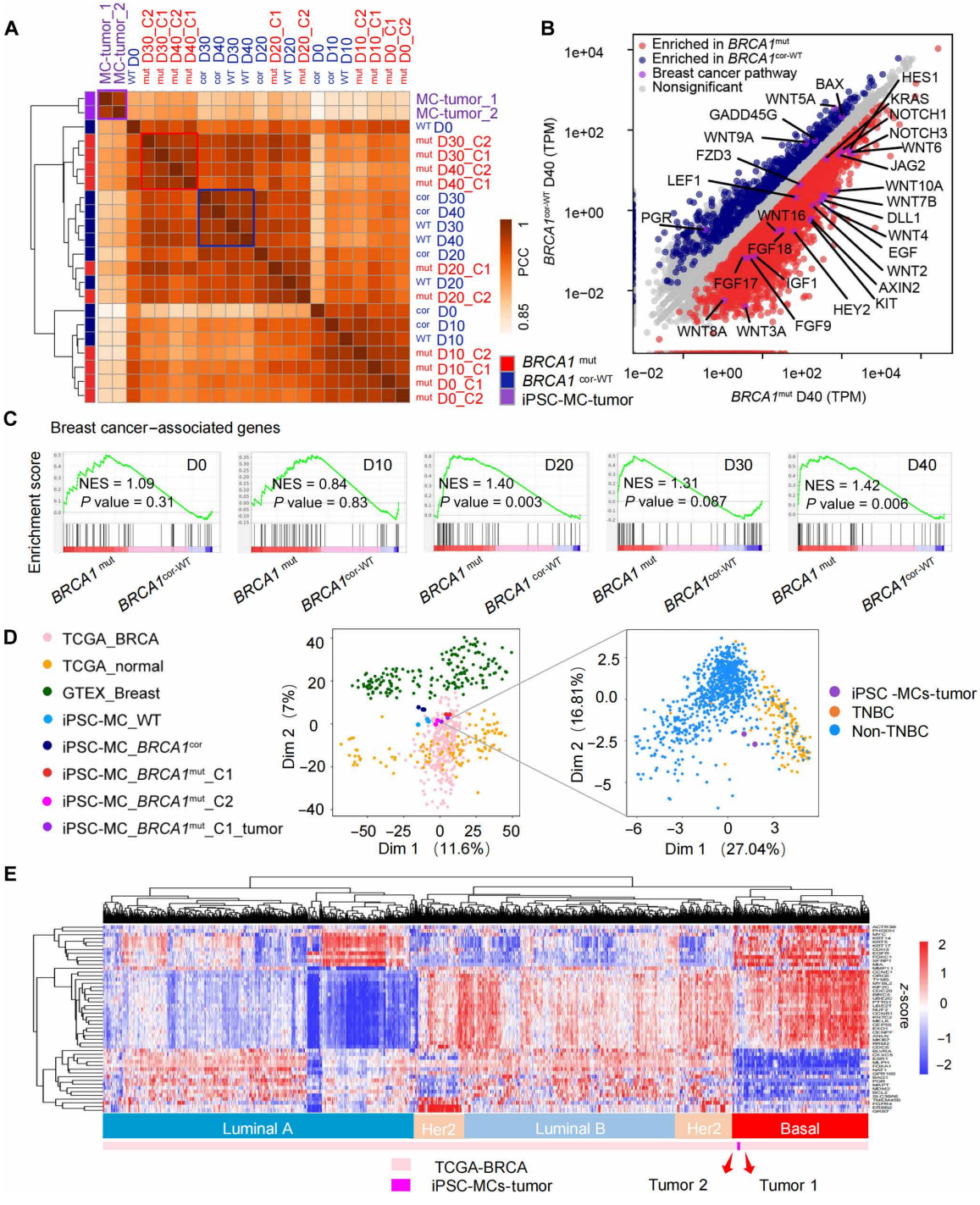


Fig. 3. *BRCA1*-mutant iPSC-MCs exhibit the *BRCA1*-related breast cancer gene signature. (**A**) Correlation matrix of *BRCA1*-mutant, *BRCA1*-mutant corrected, and WT iPSC mammary differentiation time course based on the TPM of all genes expressed. (**B**) Scatterplot presenting the values of TPM for each gene in the *BRCA1*-mutant samples (*x* axis) versus the control samples (*y* axis). Purple dots mark genes associated with breast cancer pathways ($n = 4$ or 6, two clones, two to three replicates each clone). (**C**) Gene set enrichment analysis (GSEA) results showing the enrichment of breast cancer–associated genes in *BRCA1*-mutant samples versus control samples during mammary differentiation time course ($n = 4$ or 6, two to three clones, two replicates each clone). NES, normalized enrichment score. (**D**) PCA was performed to compare the expression profiles of iPSC-MCs, iPSC-MCs-tumor, breast tumor (TCGA), para-carcinoma tissue (from Genotype-Tissue Expression Project, GTEx), and normal breast tissue (GTEx). (**E**) Hierarchical clustering of TCGA-BRCA samples and iPSC-MCs-tumor on the basis of PAM50 gene expression.

We therefore treated each clone as biological replicates in subsequent analyses, with WT/corrected clones serving as controls (*BRCA1*$^{cor\text{-}WT}$) and two patient-derived clones constituting the mutant group (*BRCA1*^{mut}). Building on this experimental framework, we specifically detected enriched breast cancer–associated gene expression in differentiating *BRCA1*-mutant iPSC-MCs compared to controls (Fig. 3C). Notably, this oncogenic signature emerged exclusively during mammary differentiation after D20 but not earlier stages, suggesting that *BRCA1* mutation leads to the expression of tumor-associated genes during mammary differentiation. As shown in Fig. 3B, oncogenes such as *KRAS*, *HES1*, *NOTCH3*, *WNT10A*, etc. were up-regulated in the *BRCA1*-mutant iPSC-MCs. Further analysis revealed that the up-regulated genes in mutant group relative to the control group were functionally enriched in the breast cancer–associated Wnt and Hippo signaling pathways (fig. S6B). Last, PCA showed that the iPSC-MCs-tumors as well as iPSC-MCs of *BRCA1* mutant clones were clustered with The Cancer Genome Atlas (TCGA) breast tumors while diverging from normal breast or para-cancerous tissue, and the iPSC-MC tumors tended to be clustered with TNBC rather than non-TNBC (Fig. 3D). Besides, the iPSC-MC tumors are transcriptionally similar to the basal-like subtype of breast cancer based on PAM50 (prediction analysis of microarray 50-gene classifier) classification (Fig. 3E), which is consistent with the pathological phenotype of TNBC and the fact that the patient had TNBC. Together, these results demonstrate that patient iPSC-derived MCs gain an oncogenic transcriptional signature during mammary differentiation.

Clinically correlated mutations can be mimicked in iPSC-based disease model

Genome instability caused by *BRCA1* mutation is one of the phenotypes of clinical *BRCA1*-related tumors and probably the factor driving malignant transformation (*23*, *24*). We thus tested whether the iPSC-based model can also simulate this genome alteration using whole-exome sequencing (WES). Samples at different differentiation stages (D0, D10, D20, D30, and D40) of *BRCA1*-mutant iPSCs and mutant corrected iPSCs, as well as tumors formed by *BRCA1*-mutant iPSC-MCs, were collected for WES (Fig. 4A). We first analyzed the mutation in the in vitro–differentiated mammary cell and found more mutations occurring and more mutated genes were accumulated in patient iPSC-MCs compared to the *BRCA1*-mutant corrected iPSC-MCs during differentiation (Fig. 4B and fig. S7A). However, the two groups of samples at D40 appear to contain a considerable number of mutations, which may be attributed to a screening mechanism that eliminates cells that have accumulated an excessive number of potentially harmful mutations, while cells having fewer mutation may more easily pass through various stages of differentiation and predominate in the final culture (Fig. 4B). Further analysis showed an increase in mutations for iPSC-MCs-tumors relative to the early stage in vitro–differentiated iPSC-MCs, and these mutations in iPSC-MCs-tumors had a genomic distribution similar to that of clinical *BRCA1*-mutant breast tumor (Fig. 4, C and D). In addition, clinical nonrecurrent mutation could be reproduced in both differentiated iPSC-MCs and iPSC-MCs-tumors, while the clinical recurrent mutation including *KMT2C*, *PIK3C2G*, and *AHNAK2* only significantly enriched in iPSC-MCs-tumors rather than the differentiated iPSC-MC (Fig. 4E), which indicates an accumulation of clinical mutations in the process of tumorigenesis. Notably, the mutation of P53-R280k was also detected in the iPSC-MCs-tumor, which is a mutation frequently occurring in clinical *BRCA1*-mutant breast tumor and was also found in the TNBC cell line MDA-MB-231 (*25*). Those results demonstrate that *BRCA1* mutation might induce clinically correlated genetic alteration and partly simulate the accumulation of clinical mutation in the iPSC-based disease model. Together, the results above demonstrate that the iPSC-based disease model mimics the accumulation of clinical genetic alteration during tumorigenesis, which can serve as a promising system for studying the breast cancer initiation and progression.

BRCA1 mutation promotes tumorigenesis by activating S100P

To further explore the application of the iPSC model, we attempted to dissect the mechanisms underlying *BRCA1* mutation–driven mammary tumorigenesis. We first screened for activated genes common to the late stage of mammary differentiation and primary breast tumors to identify overlapping genes that are potentially involved in the early stage of tumorigenesis. As shown in Fig. 5 (A to C), there were 26 overlapping genes, of which *S100P* was the top overexpressed gene in primary breast tumor and was significantly up-regulated in primary breast tumors harboring germline *BRCA1* mutations compared to normal breast tissues (Fig. 5D). The rest of the candidates contain some known oncogenes associated with metastasis and prognosis of breast cancer and some novel genes that might have oncogenic potential during tumorigenesis (Fig. 5C).

We then analyzed the role of S100P in tumor promotion by RNA interference–mediated knockdown in iPSC-based breast cancer model. Notably, S100P knockdown reduced the in vivo tumorigenic ability of *BRCA1*-mutant iPS-MCs in mouse mammary fat pad (Fig. 5, E to G, and fig. S8, A and B). Also, a significant reduction of tumorigenic ability of breast cancer cells in the xenograft mouse model was caused by S100P knockdown (Fig. 5, H and I, and fig. S8, C to G). In addition, depletion of S100P significantly decreased the in vitro proliferation ability of the cancer cell line (fig. S8, H to K). These results indicate that S100P might contribute to the *BRCA1*-related tumor initiation and progression.

Next, we investigated the role of BRCA1 in S100P regulation in mammary cells. The expression of *S100P* was inversely correlated to that of *BRCA1* during the differentiation of iPSCs into mammary epithelial cells (Fig. 5J), We then wondered whether BRCA1 has a repressive regulation on the expression of *S100P*. Consistent with the up-regulation of *S100P* in *BRCA1*-mutant iPSC-MCs, BRCA1 knockdown increased *S100P* expression (Fig. 5K and fig. S8L), while BRCA1 overexpression reduced *S100P* expression (Fig. 5L) in normal breast cells (MCF10A). We next performed chromatin immunoprecipitation (ChIP)–quantitative polymerase chain reaction (qPCR) analysis and found that BRCA1 was recruited to the endogenous *S100P* promoter (Fig. 5M). Since BRCA1 is an established transcription factor (*26*–*29*), it is likely that BRCA1 binds to the *S100P* promoter region and inhibits its expression. Together, we identified *S100P* as a novel oncogene in *BRCA1*-mutant breast cancer that is up-regulated in response to *BRCA1* mutation.

S100P activates stemness in a BRCA1-deficient iPSC-based disease model

Cancer stem cells (CSCs) are crucial to tumorigenesis (*30*), and activation of embryonic stemness has been reported in *BRCA1* mutation–associated breast tumors (*31*, *32*). However, it is unclear how *BRCA1* mutation activates stemness and whether stemness activation is involved in the early stage of malignant transformation. We detected an activation of signature genes associated with human embryonic

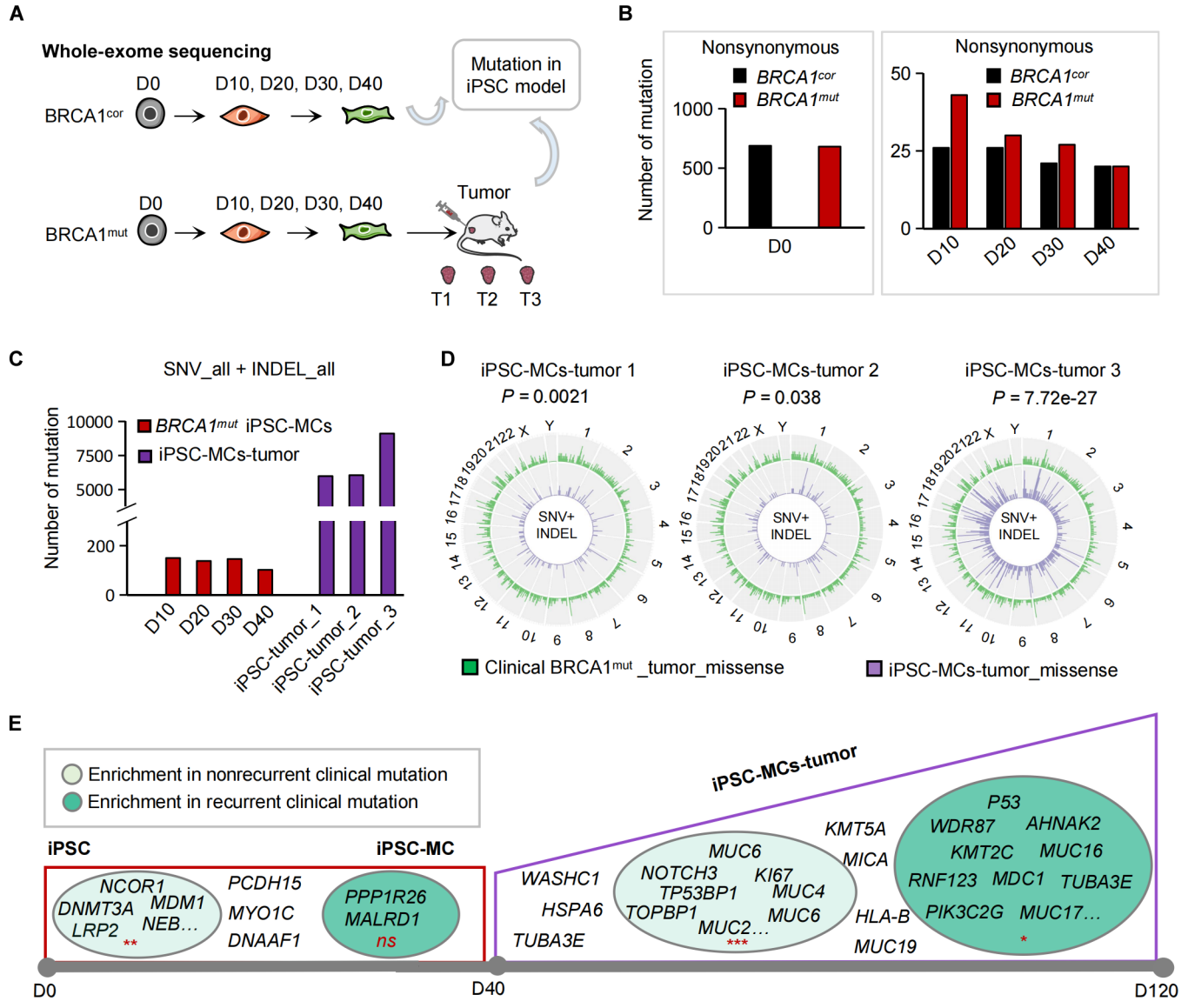


Fig. 4. Clinically correlated mutations can be mimicked in the iPSC-based disease model. (**A**) Schematic overview of WES in iPSC-based disease model. (**B**) Number of somatic mutations as nonsynonymous single-nucleotide variant (SNV) detected in iPSCs (left) and iPSC-MCs (right) with *BRCA1*-mutant and *BRCA1*-mutant corrected. (**C**) Number of SNV and insertion/deletion mutation (INDEL) detected in *BRCA1*-mutant iPSC-MCs and iPSC-MCs-tumor. (**D**) Circos plot showing the chromosome-wide codistribution of mutations in iPSC-MCs-tumor (purple) and clinical *BRCA1*-mutant breast tumor (green). The clinical mutation data were from cBioPortal. Hypergeometric test, *P* values (left to right): $P = 0.0021$, $P = 0.038$, and $P = 7.72 \times 10^{-27}$. (**E**) Landscape showing enrichment of nonrecurrent and recurrent clinical mutation in the iPSC-based disease model. Hypergeometric test, *P* values (left to right): $P = 0.0064$, $P = 0.561$, $P = 5.84 \times 10^{-06}$, and $P = 0.033$. The clinical mutation data were from cBioPortal.

stem cells (hESCs) as well as breast CSCs in the *BRCA1*-mutant iPSC-MCs during mammary differentiation (Fig. 6, A and B). In addition, the key pluripotent transcriptional factors OCT4, SOX2, NANOG, and KLF4 and the reprogramming factor MYC were also highly expressed in *BRCA1*-mutant iPSC-MCs (Fig. 6C). These results indicate that stemness increased in the early stage and the whole process of differentiation in the *BRCA1*-mutant group.

We then asked whether S100P regulates stemness. Ectopic expression of S100P (S100P-OE) up-regulated *OCT4*, *SOX2*, and *NANOG* (Fig. 6D, and fig. S9, A and B), as well as the breast tumor–related stemness marker aldehyde dehydrogenase 1 (ALDH1) (Fig. 6E and fig. S9C) in both normal mammary cells and breast cancer cells. Consistently, overexpression of S100P significantly increased the number of mammospheres and colonies in vitro compared to the control (Fig. 6F and fig. S9, D and E). Gene set enrichment analysis (GSEA) analysis further revealed an enrichment of pluripotency genes in control MCF10A cells compared to S100P-knockdown cells (fig. S9F). In addition, treatment with the S100P inhibitor reduced the stemness in mammary-related cells (fig. S9, J to M). Together, these results indicate that S100P promotes stemness in mammary cells.

To determine whether S100P mediates the stemness arising from *BRCA1* mutation, we knocked down S100P in *BRCA1*-silenced normal

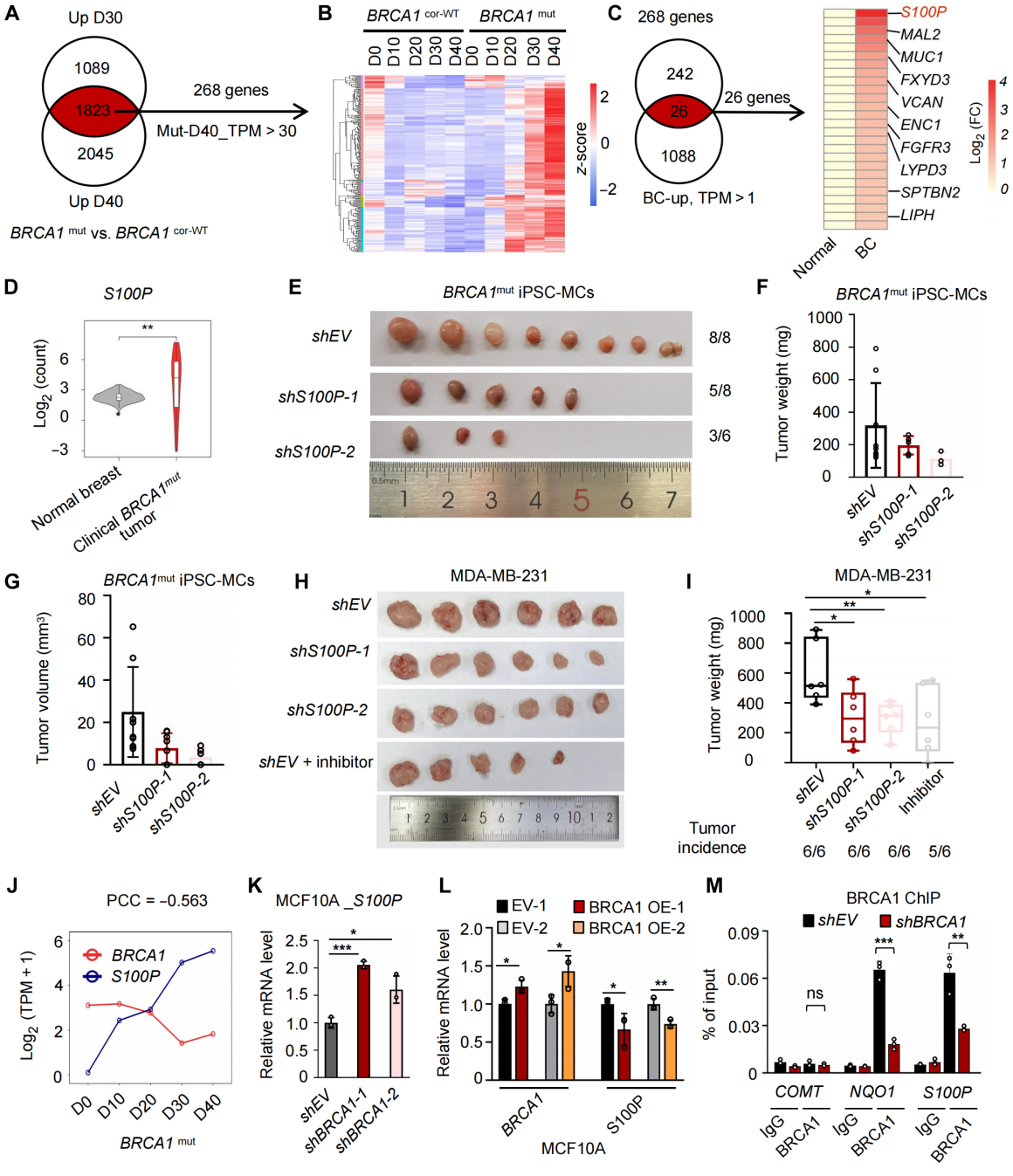


Fig. 5. *BRCA1* mutation promotes tumorigenesis by activating S100P. (**A**) Overlap of genes up-regulated at D30 and D40 in *BRCA1*-mutant iPSC-MCs compared to control iPSC-MCs [fold change (FC) > 2, $P < 0.05$]; ($n = 4$ or 6, two clones, two to three replicates each clone). (**B**) Heatmap of mRNA expression of genes identified in (A). (**C**) Overlap of the 268 genes identified in (A) with genes overexpressed in breast cancer (BC; left). Heatmap of mRNA expression of 26 genes plotted by $\log_2$ (TPM + 1) in breast tumor relative to normal breast tissue (right), data from GEPIA2 [FC > 2, q value < 0.01, analysis of variance (ANOVA) test]. (**D**) *S100P* mRNA level in normal and *BRCA1*-mutant breast tumor. Data from TCGA; Wilcox test; $n = 678$ (normal), $n = 300$ (tumors); $P = 0.002$. (**E**) Representative pictures demonstrating tumors formed by *BRCA1*-mutant iPSC-MCs in mouse mammary fat pad with/without S100P knockdown. EV, empty vector. (**F** and **G**) Histogram showing the weight (F) and volume (G) of formed tumors in (E). Error bars represent ±SD; $n = 6$ to 8 samples. (**H**) Images of subcutaneous MDA-MB-231 tumors with/without S100P knockdown and inhibitor treatment. (**I**) Histogram showing the weight of formed tumors in (H) (±SD; $n = 5$ to 6 tumors). *t* test, two-tailed; *P* values (left to right): $P = 0.0267$, $P = 0.0099$, and $P = 0.0289$. (**J**) Negative correlation between *S100P* and *BRCA1* mRNA expression during mammary differentiation ($n = 4$, two clones, two replicates each clone). (**K**) RT-qPCR showing an increase of *S100P* expression induced by BRCA1 knockdown in MCF10A cells (±SD; $n = 3$ distinct samples). *t* test, two-tailed; *P* values (left to right): $P = 0.00009$ and $P = 0.0166$. (**L**) RT-qPCR showing *BRCA1* and *S100P* expression in MCF10A cells transiently transfected with *BRCA1* and empty vectors (±SD; $n = 3$ distinct samples). *t* test, one-tailed; *P* values (left to right): $P = 0.0101$, $P = 0.0167$, $P = 0.0311$, and $P = 0.0041$. (**M**) ChIP-qPCR showing BRCA1 binding on *S100P* promoter after shEV or sh*BRCA1* (±SD; $n = 3$ distinct samples). *COMT* and *NQO1* were negative and positive controls, respectively. *t* test, two-tailed; *P* values (left to right): $P = 0.4871$, $P = 0.0002$, and $P = 0.0070$. IgG, immunoglobulin G.

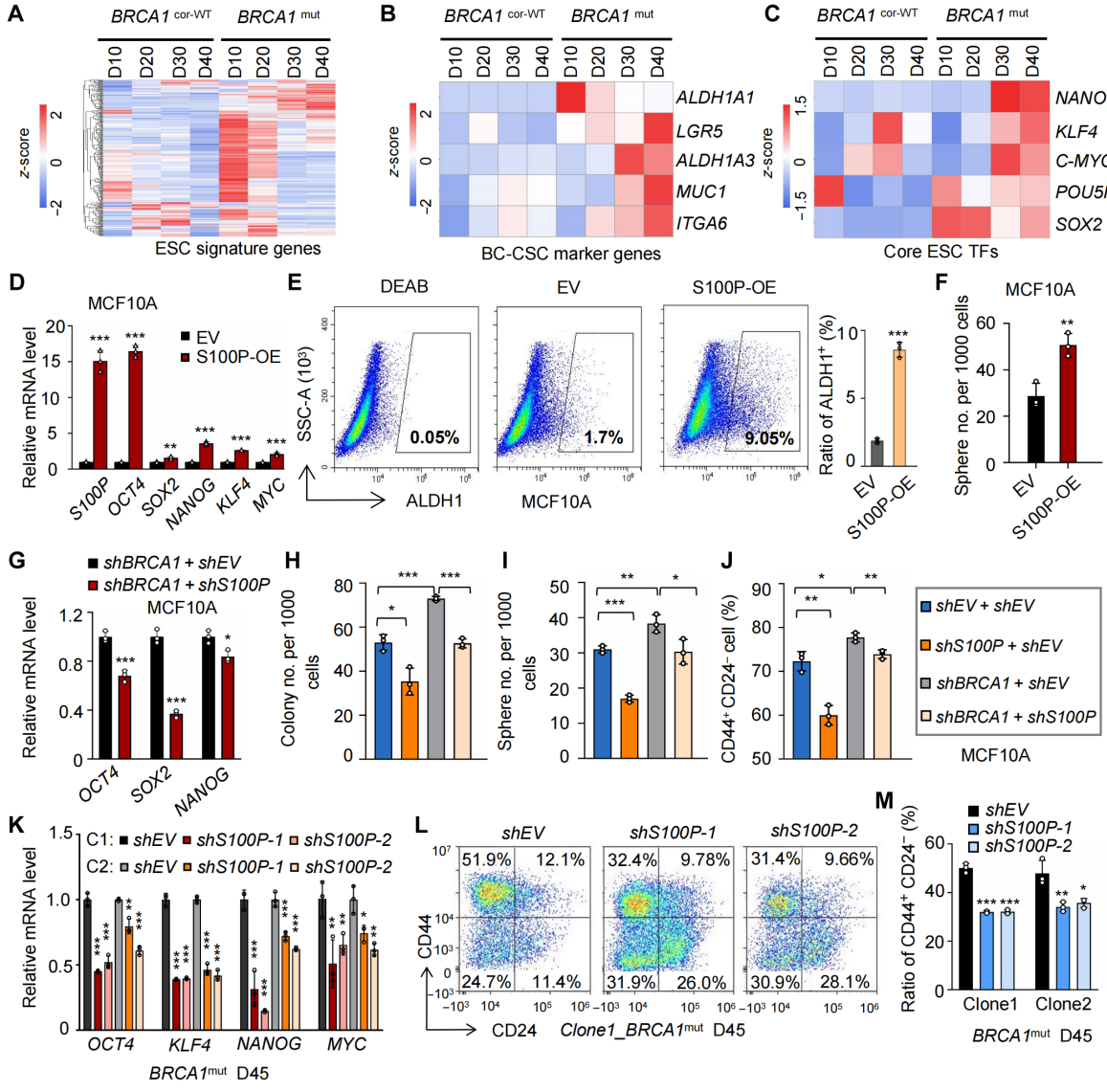


Fig. 6. S100P activates stemness in BRCA1-deficient iPSC-based disease model. (**A** to **C**) Enrichment pattern of hESC-associated genes (A), breast CSC-associated genes (B), and pluripotent transcriptional factors (TFs) (C) during mammary differentiation. ($n = 4$ or 6, two clones, two to three replicates each clone). (**D**) RT-qPCR showing expression of stemness genes in ectopic expression of S100P (S100P-OE) and EV (empty vector) MCF10A cells (±SD; $n = 3$ distinct samples). *t* test, two-tailed; *P* values (left to right): $P = 0.00008$, $P = 4.715 \times 10^{-06}$, $P = 0.0052$, $P = 0.00003$, $P = 1.507 \times 10^{-07}$, and $P = 0.0004$. (**E**) Proportion of ALDH1$^+$ mammary cells induced by S100P-OE in MCF10A cells (±SD; $n = 3$ distinct samples). *t* test, two-tailed; $P = 0.00004$. DEAB, N,N-diethylaminobenzaldehyde. (**F**) Number of mammospheres formed by EV or S100P-OE MCF10A cells (±SD; $n = 3$ distinct samples). *t* test, two-tailed; $P = 0.0070$. (**G**) RT-qPCR assay showing the expression of *OCT4*, *SOX2*, and *NANOG* in sh*BRCA1* and sh*BRCA1* + sh*S100P* MCF10A cells (±SD; $n = 3$ distinct samples). *t* test, two-tailed; *P* values (left to right): $P = 0.0009$, $P = 0.00007$, and $P = 0.0174$. (**H** and **I**) Number of colonies (H) and mammospheres (I) formed by MCF10A cells with shEV, sh*BRCA1*, sh*S100P*, and sh*BRCA1* + sh*S100P*. Error bars represent ±SD; $n = 3$ distinct samples. *t* test, two-tailed; *P* values (left to right): $P = 0.0125$, $P = 0.0008$, $P = 0.0001$, $P = 0.0001$, $P = 0.0094$, and $P = 0.0205$. (**J**) Proportion of CD44$^+$ CD24$^-$ mammary cells in MCF10A cells with shEV, sh*BRCA1*, *shS100P*, and sh*BRCA1* + *shS100P* (±SD; $n = 3$ distinct samples). *P* values (left to right): $P = 0.0024$, $P = 0.0179$, and $P = 0.0097$. (**K**) RT-qPCR assay showing *OCT4*, *NANOG*, *KLF4*, and *c-MYC* expression in *BRCA1*-mutant iPSC-MCs with or without S100P knockdown (±SD; $n = 3$ distinct samples). C1 denotes clone 1, and C2 denotes clone 2. *t* test, one-tailed; *P* values (left to right): $P = 2.321 \times 10^{-05}$, $P = 0.00014$, $P = 0.0016$, $P = 1.8270 \times 10^{-05}$, $P = 1.2077 \times 10^{-05}$, $P = 1.2883 \times 10^{-05}$, $P = 2.230 \times 10^{-05}$, $P = 1.3269 \times 10^{-05}$, $P = 0.0007$, $P = 1.6747 \times 10^{-05}$, $P = 0.0009$, $P = 0.0002$, $P = 0.0098$, $P = 0.0091$, $P = 0.0108$, and $P = 0.0023$. (**L**) Representative pictures demonstrating the CD44$^+$ CD24$^-$ population in *BRCA1*-mutant iPSC-MCs with or without S100P knockdown (S100P-KD). (**M**) Proportion of CD44$^+$ CD24$^-$ cell was reduced by S100P-KD in *BRCA1*-mutant iPSC-MCs (±SD; $n = 3$ distinct samples). *t* test, one-tailed; *P* values (left to right): $P = 4.3983 \times 10^{-05}$, $P = 5.4594 \times 10^{-05}$, $P = 0.0019$, and $P = 0.0203$.

breast cells MCF10A. *BRCA1* knockdown increased the stem-like characteristics of MCF10A cells, which were reversed by the additional inactivation of S100P (Fig. 6, G to J). As for *BRCA1*-null cancer cells, which exhibit a high ALDH activity, knockdown of S100P substantially reduced the cell population expressing ALDH1 (fig. S9G), Furthermore, S100P depletion also reversed the increased proliferation ability of normal breast cells lacking *BRCA1* (fig. S9, H and I). We found that S100P knockdown reduced the stemness characteristics in the *BRCA1*-mutant iPSC-MCs (Fig. 6, K to M). The above observations reveal that S100P functions downstream of *BRCA1* mutation to induce cancer stemness in mammary cells.

The correlation between S100P and BRCA1 was lastly determined in clinically relevant samples. First, compared to normal breast tissue, *S100P* specifically highly expressed in the mRNA level in patient tumor–containing *BRCA1* mutation (Fig. 7A). Second, *S100P* is more likely to be highly expressed in the mRNA level in tumor with high histologic grade and heavy tumor mutation burden, which are pathological features associated with *BRCA1* mutation (Fig. 7, B to D). Third,

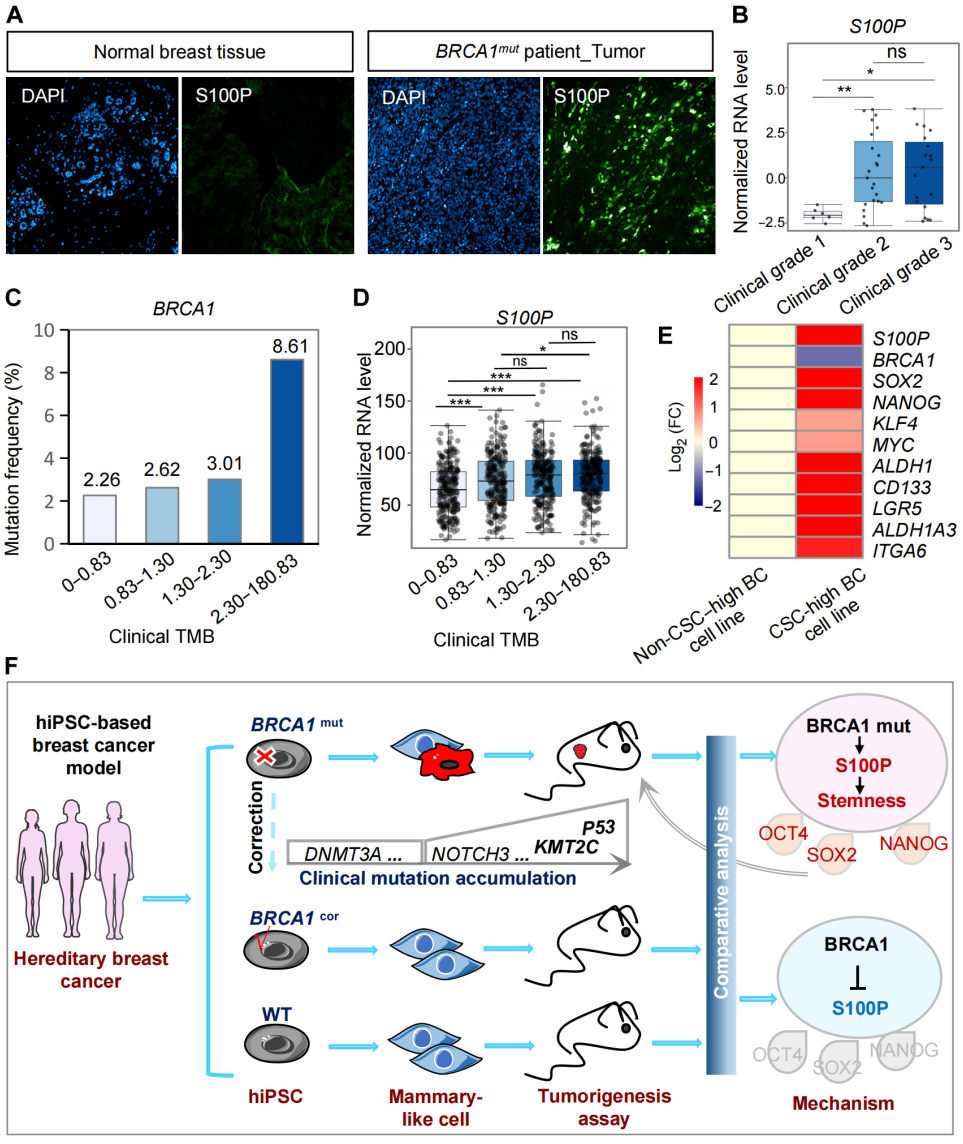


Fig. 7. Clinical analysis reveals an association of S100P and BRCA1 in tumorigenesis. (**A**) S100P expression in clinical normal breast tissue and breast tumor with *BRCA1* mutation. (**B**) Boxplot showing S100P expression in breast tumors with different histologic grades. $n = 6$ grade 1 samples, $n = 23$ grade 2 samples, and $n = 19$ grade 3 samples. *t* test, two-tailed; *P* values (left to right): $P = 0.006$, $P = 0.042$, and $P = 1.000$. Analysis was made on the basis of published data (*64*). (**C**) Histogram showing mutation frequency of *BRCA1* in breast tumors with different tumor mutation burden (TMB). $n = 247$, $n = 235$, $n = 243$, and $n = 236$ (left to right). Data from cBioPortal. (**D**) Boxplot showing S100P expression in breast tumors with different TMB. $n = 247$, $n = 235$, $n = 243$, and $n = 236$ (left to right). *t* test, two-tailed; *P* values (left to right): $P = 0.00097$, $P = 1.56 \times 10^{-06}$, $P = 8.76 \times 10^{-09}$, $P = 0.174$, $P = 0.030$, and $P = 0.4310$. Data from cBioPortal. (**E**) Heatmap showing the expression of *S100P* and *BRCA1* in non–CSC-high (HMLER) or CSC-high (BPLER) breast cancer cell lines. Data from GSE131631. (**F**) iPSCs from breast cancer family were directed to differentiation toward mammary cells, during which the *BRCA1* mutation–induced breast tumorigenesis process can be simulated. In terms of mechanism, we found that *BRCA1* mutation–induced carcinogenesis is related to the activation of S100P. In WT iPSC-MCs, BRCA1 inhibits the expression of *S100P* in a targeted manner. In the *BRCA1* mutation group, the increased expression of S100P promotes breast cell carcinogenesis by enhancing stemness gene expression. hiPSC, human-induced pluripotent stem cell.

the *S100P* was highly expressed in CSC-high breast cancer cells, but *BRCA1* was conversely highly expressed in non-CSC–high cells (Fig. 7E). In addition, a stronger positive correlation between high-level *S100P* expression with the CSC surface marker *MUC1* was observed in the basal-like subtype compared to the other three subtypes (fig. S9N), which is consistent with the fact that *BRCA1*-mutant breast cancer is mainly basal like and harbors more cancer stem cells (*31*, *32*). These results confirmed that *S100P* expression was positively correlated with the stemness level and was clinically associated with *BRCA1* mutation. Together, *BRCA1* mutation enhances stemness in both normal mammary epithelial cells and breast cancer cells through S100P, which is likely the underlying factor associated with malignant transformation and tumor progression (Fig. 7F).

DISCUSSION

The iPSCs are a promising tool for cancer research and drug discovery. However, its application in breast cancer research has not been explored, partly due to the lack of efficient protocols for generating functional mammary cells from iPSCs. In this study, we developed a method to differentiate iPSCs to the mammary lineage and used the iPSCs-MCs to model *BRCA1* mutation–induced tumorigenesis and genetic alteration. With the help of this model, we were able to identify S100P as a mammary oncogene that is activated as a result of *BRCA1* mutation and likely promotes tumorigenesis by inducing cancer stem cells. Our findings highlight the potential role of S100P inhibitor as preventive and therapeutic drugs for breast cancer.

iPSC-MCs provide a suitable model for studying the cellular hierarchy and molecular events in embryonic mammary development as well as investigating the relationship between mammary differentiation and tumorigenesis. Many pieces of evidences have indicated that dysregulation in mammary development is directly or indirectly related to the occurrence of tumors (*33*–*36*). The human breast undergoes a series of changes from intrauterine state to senescence (*37*). The initial intrauterine stages of mammary gland development are not well documented in humans, and the available information is mainly based on studies on mouse models (*38*). The directed differentiation of iPSCs toward specific lineages can simulate the embryonic development of the corresponding organs as well as disease pathogenesis in vitro. Therefore, the iPSC-derived mammary cells can be an alternative model to study the embryonic mammary development and breast tumorigenesis. BRCA1 plays an important role in regulating mammary gland development. A previous study on 15- and 33-week-old human fetuses showed that *BRCA1* expression decreases with fetal age (*39*). Consistently, we found the similar phenotype in the established mammary differentiation system (fig. S1G). In addition, key genes involved in mammary development were found sequentially activated in mammary differentiation from human iPSCs (fig. S1B). These findings suggest that this established differentiation system at least partially simulates the characteristics of mammary gland development. Furthermore, we found the gene transcriptome of *BRCA1*-mutant and mutant-corrected group was similar at the iPSC stage but diverged at the differentiation stages (Fig. 3A), indicating that cancer-related molecular events might occur during differentiation or differentiation process be interrupted during tumorigenesis. Thus, our study provides a proper model for studying the correlation between mammary development and tumorigenesis.

The iPSC-based breast cancer model can be used for drug screening and identification of diagnostic markers. Here, we provide a promising drug screening platform based on the disease-associated phenotype of stemness enhancement or the ability of cell proliferation. Using an iPSC-based colorectal cancer model, Crespo *et al.* (*40*) have successfully found a novel drug which targets cancer cells with a better specificity. In addition, an iPSC based model has been used to test pharmacological interventions on myeloid (*41*). We also identified a set of candidate genes that might be involved in the early stage of breast tumorigenesis. Many of these genes, including *MUC1*, *FXYD3*, *LYPD3*, and *FGFR3*, are up-regulated in breast cancer and reported to be associated with tumor metastasis or invasion (*42*–*45*), which is suggestive of the effectiveness of using the iPSC-based model for therapeutic targets discovering. In particular, S100P could be potentially used as early diagnostic markers for *BRCA1* mutation carrier, since they were not activated in untransformed *BRCA1*-mutated breast tissue but specifically activated at the early stage of tumorigenesis and in breast tumor. By combining with the technologies of single-cell RNA-seq, cell tracking technique, and genome sequencing, this iPSC-based breast cancer model can be used to dissect the tumorigenesis stages, identify tumor initiating cell, and mapping the genetic landscape of tumorigenesis in subsequent studies. Furthermore, on the basis of these mechanistic studies, the iPSC model can be used to find more diagnostic biomarkers, therapeutical targets for cancer prevention, as well as a screen for novel drug candidates.

S100P is a promising target for the prevention and treatment of *BRCA1*-associated breast cancer. Bilateral salpingo-oophorectomy is frequently recommended for women with *BRCA1* or *BRCA2* mutation to lower cancer risk (*46*, *47*). However, recent prospective cohort studies did not report any significant association between this procedure and breast cancer risk in *BRCA1*/*2* mutation carriers (*48*–*50*). Furthermore, the existing prevention agents did not benefit women with *BRCA1* or *BRCA2* mutation in clinical trials, although observational data show that secondary prevention with tamoxifen can reduce the risk of contralateral breast cancer (*51*, *52*). Therefore, medical prevention of breast cancer in *BRCA1* mutation carriers is still controversial. Here, we demonstrated the inhibitor blocking S100P binding to receptor for advanced glycation endproducts (RAGE) could reduce the subcutaneous tumorigenesis and the stemness of breast cancer cells, which preliminarily validated the potential of S100P as a therapeutic target and its inhibitor as therapeutic drug for breast cancer.

Limitations

Similar to limitations existing in most iPSC-based disease models, the in vitro simulation of breast tumorigenesis may lack sufficient time, because the occurrence of tumor in vivo is a long-term cumulative process. Although this carcinogenic process may theoretically be accelerated by the shortened in vitro–induced mammary differentiation, some pathological characteristics of the *BRCA1*-related TNBC may not be recapitulated in the model, such as the clinical features of higher histologic grade, a medullary histopathology, and a sufficient amount of mutations as identified in patient's tumor. Also, our finding might be limited by the fact that FPS-ZM1 used in this study is not a S100P-specific inhibitor, which blocks not only the interaction between S100P and RAGE but also the binding of other proteins to RAGE, including Aβ40, HMGB1, and S100B (*53*–*55*). In addition, our finding that S100P is a potential target for the prevention and treatment of *BRCA1*-related breast cancer subtype might be somewhat limited. S100P was abnormally activated when *BRCA1* is mutated and is also overexpressed in other transformed breast cancer subtypes. Existing studies have showed that high S100P levels are related to ER/PR and HER2 overexpressing in breast tumors (*56*, *57*).

These results suggest that S100P-dependent stemness may not be responsible for the triple-negative phenotype of *BRCA1*-associated breast cancer, instead S100P overactivation might be a common mechanism underlying the malignancy of breast cancer.

MATERIALS AND METHODS

Cell lines

BRCA1-null breast cancer cell line HCC-1937 (Procell, CL-0093) was obtained from American Type Culture Collection (ATCC). The human TNBC cell line MDA-MB-231 and mammary fibroblasts were gifts from S. Su at Sun Yat-sen University. The human mammary epithelial cell line, MCF10A, was from H.H.'s laboratory at Sun Yat-sen University. The human embryonic stem cell line H9 was from N.C.'s laboratory at Sun Yat-sen University. The normal human iPSC line reprogrammed from the fibroblasts of a healthy female fetus was provided by G. Pan at the Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, and the iPSC lines reprogrammed from monocytes of the breast cancer family were established as described here in Materials and Methods. 293T cells were purchased from ATCC.

Breast cancer cells HCC-1937 were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS) (VISTECH) and 1% penicillin-streptomycin (Hyclone). MDA-MB-231 cells, mammary fibroblasts, and 293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) high-glucose media supplemented with 10% FBS (VISTECH) and 1% penicillin-streptomycin (Hyclone). MCF10A cells were cultured in DMEM/F12 (Gibco) supplemented with 5% horse serum (Biological Industries), hydrocortisone (0.5 μg/ml; STEMCELL Technologies), insulin (10 μg/ml; M9194, AbMole, USA), epidermal growth factor (EGF; 20 ng/ml; PeproTech), cholera toxin (100 ng/ml; Sigma-Aldrich), and 1% penicillin-streptomycin (Hyclone). Human iPSCs and hESC H9 were maintained in mTeSR media (STEMCELL Technologies) on tissue culture plates coated with Matrigel (BD Biosciences). All cells were cultured in 5% CO_2, 37°C incubator.

Mice

Female NOD-SCID mice of 3-week old used in mammary reconstruction experiment were purchased from Model Animal Research Center of Nanjing University. Male NOD-SCID mice subjected to vasectomy were purchased from Beijing Vital River Laboratory Animal Technology Co. Ltd., which were used to induce pseudopregnancy of female mice in the mammary reconstruction experiment. Female NOD-SCID mice (4 to 6 weeks old) used in the assay of subcutaneous tumors formation were purchased from the Model Animal Research Center of Nanjing University or from the Experimental Animal Center, Sun Yat-sen University. Female NCG mice (4 to 6 weeks old) were purchased from GuangDong GemPharmatech Co. Ltd. All mice were bred in the specific pathogen–free animal facility of the Laboratory Animal Resource Center of Sun Yat-sen University in individually ventilated cages. Animal work was approved by the Institutional Review Boards and Animal Care and Use Committees of Sun Yat-sen University (IACUC/2017-0081). Animal study followed the main principle of Animals in Research: Reporting in vivo experiments.

Mammary cell differentiation

To start mammary cell differentiation, human iPSCs were plated in 24-well plates coated with growth factor–reduced Matrigel (BD Biosciences). Single iPSCs were seeded at 6×10^5 cells/cm^2 and cultured in mTeSR media (STEMCELL Technologies) supplemented with 5 μM Y-27632 (Selleck). Mammary differentiation was started when the cells reached 100% of confluence, usually after 24 hours. For the first 11 days, cells were differentiated to non-neural ectoderm as described (*58*). Specifically, for D0 to D1 (0 to 48 hours), cells were cultured using E6 media containing bone morphogenetic protein 4 (BMP4; 10 ng/ml; PeproTech), 10 μM SB431542 (Selleck), and 10 μM SU5402 (Selleck), and the medium was changed every day. On D2 to D10, E6 medium containing BMP4 (5 ng/ml) and 10 μM SB was added to cells every other day. For D11 to D40, MammoCult media (STEMCELL Technologies) was changed to induce mammary differentiation, and the medium was changed every 48 hours. Plates were kept at 5% CO_2 in an incubator during the differentiation procedure.

Normal breast tissue dissociation

Human mammary epithelial cells were isolated from normal margin removed during breast cancer surgery for a female patient as previously described (*59*). Briefly, tissues were chopped into 1- to 3-mm squares and were dissociated in DMEM/F12, supplemented with hyaluronidase (100 U/ml; Sigma-Aldrich), collagenase (300 U/ml; Solarbio), 2% bovine serum albumin (Gibco), insulin (5 mg/ml; Sigma-Aldrich), hydrocortisone (0.5 mg/ml; STEMCELL technologies), cholera toxin (10 ng/ml; Sigma-Aldrich), and 1% penicillin-streptomycin (Hyclone) at 37°C for 16 hours. The epithelial cell–rich pellets were collected by centrifugation at 80*g* for 4 min and were subjected to one wash with DMEM/F12. The obtained pellets were further disassociated with 0.05% trypsin/EDTA (Hyclone) for 5 min to generate a single-cell suspension of mammary epithelial cells. The cell suspension was filtered through 40-μm polyethylene cell strainers (Biologix).

3D culture

3D culture was performed as previously reported (*20*). A volume of 120 μl of growth factor–reduced Matrigel (BD Biosciences) was added into each 24-well plate (Thermo Fisher Scientific) and allowed to solidify for 15 min in a 37°C, 5% CO_2 incubator. After primary mammary epithelial cells or iPSC-derived mammary cells were trypsinzed and counted, 12,000 cells per well in 1 ml of DMEM/F12 supplemented with 2% Matrigel, 2% horse serum (BI), hydrocortisone (0.5 μg/ml), insulin (10 μg/ml), EGF (5 ng/ml), cholera toxin (100 ng/ml), and 1% penicillin-streptomycin were seeded on the top of the solidified Matrigel layer. To induce milk protein expression, cells were cultured in the presence of prolactin (5 μg/ml) for 3 days. The medium was replaced every 2 to 3 days.

Mammary gland reconstruction

NOD/SCID mice (female, 4 weeks of age) were used to assess the in vivo potential of the differentiated mammary cell to reconstruct mammary structure. Fibroblasts were used to support the growth of normal epithelial cells as described previously (*60*). The fat pads were cleared before puberty and humanized by injection of a mixture of human mammary fibroblasts and growth factor–reduced Matrigel (BD Biosciences) diluted to 50% with phosphate-buffered saline (PBS) per fat pad. Primary MCs, iPSC-MCs, or iPSCs were mixed with growth factor–reduced Matrigel (BD Biosciences) diluted to 50% with PBS and injected in the cleared humanized fat pads, respectively, 2 to 4 weeks after clearing. Estrogen pellets were subcutaneously implanted at the time of the fat pad clearing, or pseudopregnancy was

induced after the second injection of cells using male mice that have undergone vasectomy to promote estrogen secretion. Injected animals were euthanized 70 to 90 days after implantation of the human mammary epithelial cells, and each injected fat pad was subjected to whole-mount analysis followed by histological staining. Mice were randomized into different groups for cell injection in a mouse mammary reconstruction experiment.

Whole-mount analysis

Carmine-alum staining was performed as described (*61*). Briefly, dissected mammary glands were spread onto glass slides, fixed with Carnoy's fixative (60% ethanol, 30% chloroform, and 10% glacial acetic acid) for overnight, hydrated, stained overnight in 0.2% carmine and 0.5% $AlK(SO_4)_2$, dehydrated in graded solutions of ethanol, cleared in Histoclear (Shandon Inc., Pittsburgh, PA, USA), and mounted.

Cell reprogramming with nonintegrating SeV

The monocyte samples from two individuals were reprogrammed, consisting of a 42-year-old female patient with TNBC and a 10-year-old female as a control subject. Human monocytes of a patient with breast cancer and the unaffected relative were isolated by Ficoll density gradient centrifugation. The obtained monocytes were cultured and maintained in complete StemPro-34 medium supplemented with stem cell factor (100 ng/ml), FLT-3 (100 ng/ml), interleukin-3 (IL-3; 20 ng/ml), and IL-6 (10 ng/ml). These monocytes were reprogrammed by transducing SeV expressing the four reprogramming factors OCT4, SOX2, KLF4, and MYC (CytoTune reprogramming kit, Invitrogen) according to the manufacturer's protocol. The reprogramming cells were maintained in mTeSR medium (STEMCELL Technologies). After 3 to 4 weeks postinduction, individual clones with hESC/iPSC morphology were picked, passaged, and examined for loss of SeV by reverse transcription (RT)–qPCR measurement of expression of the exogenous four factors. The specific qPCR primers targeting exogenous *OCT4*, *SOX2*, *KLF4*, and *MYC* are listed in table S2. The Medical Ethics Committee of Zhongshan School of Medicine of Sun Yat-sen University has approved the collection of samples from participants (SYSU-ZSMed-ETH-2021-090), from whom the informed consent of the study has been obtained.

BRCA1-mutant corrected iPSC

Patient iPSCs with *BRCA1* mutation and H9 were maintained in mTeSR (STEMCELL Technologies, 05850). To perform gene editing using a ribonucleoprotein (RNP) complex, 3 μg of Guide-it Recombinant Cas9 (Electroporation-Ready) (Clontech) and 0.4 μg of in vitro–transcripted single guide RNA (sgRNA) by the Guide-it sgRNA In Vitro Transcription Kit (Clontech) were mixed with 100 μM ssODNs and incubated for 10 min at room temperature. A total of 1×10^6 cells of iPSCs/H9 were resuspended in 100 μl of solution I of the Lonza Human Stem Cell Nucleofector Starter Kit (Lonza), followed by the addition of RNP mix. Cells and DNA were transferred to a Lonza nucleocuvette. Cells were pulsed in a Lonza 4D nucleofector using the program B-16. After electroporation, 500 μl of medium was added, and cells were returned to culture in one well of a 24-well culture plate. Twenty-four hours later, puromycin (0.5 μg/ml; Sigma-Aldrich) was added to select positive transfected cells. Cells were diluted into single cell and cultured in 96-well plates after puromycin selection. To detect the efficiency of gene editing caused by Cas9 RNP mix, genomic PCR was performed with a pair of primers flanking the mutation loci of *BRCA1*. Primers are listed in tables S1 and S4.

Histology and staining

Tissues were fixed for 24 hours, in 10% neutral-buffered formalin (Sigma-Aldrich) at room temperature, and then embedded in paraffin as follows: Briefly, tissues were processed through a graded ethanol series followed by xylene, and then embedded in paraffin, cut at 5 μm, and stained. Specific protein and marker expression was analyzed by immunostaining or immunohistochemistryas as previously described; the antibodies are listed in table S5.

Immunofluorescence

Cells were fixed in 4% formaldehyde for 15 min at room temperature, and the following steps were performed as previously described. Primary and secondary antibodies are listed in table S5. Nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI). Pictures were taken using a Eclipse Ts2R-FL (Nikon), microscope LSM780 (Zeiss), and Operetta CLS (PerkinElmer)

RT-qPCR analysis

The total RNA was isolated using RNAzol (Molecular Research Center, Inc.) and reverse-transcribed using a PrimeScript RT Master Mix (Takara) according to the manufacturer's protocol. Quantitative Real-time PCR was performed with the Roche 480 Lightcycler with SYBR qPCR Master Mix (Vazyme). Triplicate reactions were carried out for each sample. Individual gene expression was normalized to *GAPDH*. Any technical replicate with a Ct (cycle threshold) value differing by more than 0.5 from other replicates was excluded from the analysis. Primer sequences are listed in table S2.

Teratoma formation assay

iPSCs were disaggregated with Accutase solution (STEMCELL) at 37°C for 2 min. Then, 1×10^7 cells were resuspended in 50 μl of PBS (Hyclone) containing 50% Matrigel (BD Biosciences) and were subcutaneously injected into the hind legs of 5-week-old immunodeficient mice. Teratomas were isolated 2 months after injection, fixed overnight in 4% formaldehyde solution (Sigma-Aldrich), and stained with hematoxylin and eosin (H&E).

Colony formation assay

Cells were seeded in a six-well plate (200 cells per well). Culture medium was changed every day. After culturing for 2 to 3 weeks, cells were fixed with 4% paraformaldehyde (PFA; Solarbio) for 15 min and stained with 0.1% Crystal Violet formaldehyde solution (Solarbio). Individual colonies with >50 cells were counted. All experiments were performed in three independent triplicates.

AP staining

Alkaline phosphatase (AP) staining was examined using the Alkaline Phosphatase Stain Kit (YEASEN) according to the manufacturer's protocol.

Karyotype analysis

iPSCs were grown on a Matrigel-coated six-well plate before karyotyping. On the day of culture harvest, colchicine (0.2 μg/ml; DAHUIBIO) was added to the 80 to 90% confluent iPSC culture for 2.5 hours. Cells were disaggregated with Accutase solution and were resuspended in 2 ml of DMEM/F12 (Gibco).

In vitro oncosphere assay

One thousand cells were seeded into individual wells of ultralow attachment 24-well plates (Corning) and incubated with 1.5 ml of oncosphere medium (DMEM/F12 (Gibco) supplemented with 2% B27 SerumFree Supplement (Gibco), EGF (20 ng/ml; Peprotech), fibroblast growth factor (20 ng/ml; Peprotech), and 1× insulin (Thermo Fisher Scientific). Individual oncospheres with ≥50 μm in diameter were calculated after 2 weeks. All experiments were performed in three independent triplicates.

In vivo tumorigenicity

All animal work was conducted in accordance with the Yat-sen University's Institutional Animal Care and Use Committee (IACUC/2017-0081). A total of 1×10^7 cells were resuspended in 50 μl of PBS (Hyclone) containing 50% Matrigel (BD Biosciences) and were subcutaneously injected into the abdominal area near the hind legs of 4-week-old immuno-deficient NOD-SCID mice. Tumors were measured every 3 days. Tumor volumes were determined with a microcaliper for tumor length (L) and width (W) and calculated following the formula $(L \times W^2)/2$. Tumors were excised around 4 to 12 weeks after injection. Tumors were weighted, fixed overnight in 4% PFA (Solarbio), and stained with H&E and immunohistochemistry, respectively. For the experiment of tumor formation within mouse mammary fat pad, 1×10^7 cells were injected into the fourth fat pads of NCG mice, and tumor formation was assessed 70 days later. Mice were randomized into different groups for cell injection tumorigenesis experiments, and all mice that survived at the end of the experiments were included for analysis. Whenever possible, tumor measurements were taken by a technician blinded to the treatment groups, although this was not always feasible depending on the technician availability. For tumorigenicity studies, sample sizes were estimated by the number of mice necessary to demonstrate clear statistical associations based on prior work ($n \geq 5$ injections per group) (*62*).

Cell migration assay

Cells were seeded in six-well plates, and cell monolayers were scratched using sterile tips. Then, cells were washed with PBS (Hyclone) twice and cultured with 2 ml of medium. Cells were photographed after being scratched for 0, 24, and 48 hours, respectively. The distance traveled by cells in the wound area was measured. All experiments were performed in three independent triplicates.

PET/CT

All animal work was conducted in accordance with the Yat-sen University's Institutional Animal Care and Use Committee (IACUC/2017-0081). The positron emission tomography (PET)/computed tomography (CT) acquisition was performed on an Inveon system. The mice were fasted 24 hours before image acquisition and were injected with 500 μCi of 18F-FDG in ~100 μl of PBS 30 min before PET/CT imaging. The mice were maintained at 37°C after being scanned. Imaging data were analyzed using Inveon research workplace.

Cell cycle analysis

For cell cycle analysis, cells were collected at 80% confluence. After washing with cold PBS, cells were fixed in cold 70% ethanol for overnight at 4°C. Cells were spun down at 800*g* and washed twice with PBS, followed by DAPI staining (work concentration, 50 μg/ml) at 37°C for 15 min. Measurements were taken from three distinct samples and were analyzed on Beckman Coulter CytoFLEX (>10,000 cells were analyzed). Flow cytometry data were analyzed by FlowJo v.7.6 software.

Cell viability assay

Cell viability was determined using the Cell Counting Kit-8 (CCK8; Dojindo). One thousand cells were seeded into individual wells of 96-well plates and incubated with 100 μl of medium. In each well, 10 μl of CCK8 was added and incubated for 2 hours at 37°C after being seeded for 1, 2, 3, 4, and 5 days. Then, the absorbance at 450 nm was measured. All experiments were performed in three independent triplicates, and the results were presented as mean ± SD ($n = 3$).

DNA extraction and PCR

Cells were lysed in 50 μl of lysis buffer [10 mM tris-HCl (pH 8.0), 0.05% SDS, and proteinase K (25 μg/ml)] and incubated at 37°C for 1 hour followed by 80°C for 30 min. Then, the cell lysis was centrifuged at a maximum speed for 5 min. The supernatant was collected for PCR with Q5 High-Fidelity 2X Master Mix (New England Biolabs).

Protein extraction and Western blots

Whole-cell lysates were prepared using ice-cold radioimmunoprecipitation assay buffer supplemented with protease inhibitor cocktail (Roche). The samples were centrifuged for 15 min at 12,000*g* for 10 min at 4°C. The protein concentration in the supernatant was determined using the Pierce BCA Protein Assay kit (Thermo Fisher Scientific). The supernatants were mixed with 5× loading buffer and stored at −80°C after boiled at 98°C for 10 min. Samples were run on SDS–polyacrylamide gel electrophoresis protein gels and transferred to an Immobilon-P polyvinylidene difluoride membrane (Bio-Rad). Following transfer, membranes were blocked in 5% bovine serum albumin in PBS containing 0.1% Tween 20 (PBST). Blots were washed for 15 min before incubation with primary antibody at 4°C for 16 hours. For the full list of antibodies, see table S5. After being washed three times in PBST, the membranes were incubated for 1 hour at room temperature with anti-rabbit/mouse horseradish peroxidase–conjugated secondary antibodies (1:10,000; Abcam). Antibody-protein complexes were visualized using Western ECL substrate (Bio-Rad).

ChIP-qPCR

ChIP experiments were performed as previously described. Cells were fixed with 1% formaldehyde at room temperature for 10 min and subsequently quenched with 125 mM glycine for 5 min. For BRCA1 ChIPs, 4 μg of BRCA1 antibody (D-9) (Santa Cruz) was conjugated with 40 μl of Dynabeads Protein G for immunoprecipitation (Thermo Fisher Scientific). The input DNA and immunoprecipitated DNA were processed for qPCR to assess the BRCA1 binding to specific locus by using specific primers. Primers are listed in table S3.

Quality control of sequencing reads

All the Illumina sequencing reads used in the study were first quality controlled by Trim Galore with default parameters.

mRNA-seq data analysis

We aligned the RNA-seq data to the hg19 reference genome using STAR with ENCODE option bundles. Using HTSeq-count, we counted the uniquely mapped reads and transformed to RPKM (reads per kilobase per million reads) or TPM (transcripts per kilobase million) for further analyses. At least two distinct samples were used for RNA-seq experiment and data analysis. We detected the differentially expressed

genes using edgeR. Genes were considered differentially expressed when the overall false discovery rate is <0.05 and the fold change is above 2.0.

Principal components analysis

mRNA-seq samples' PCA are performed by using R package FactoMineR, based on read counts of specific gene sets. The calculation of correlation coefficients is based on TPM of specific gene sets. The visualizations are completed by using R package ggplot2 and pheatmap. The sample size for each group was determined by the number of tumors included in the TCGA database.

GO and KEGG pathway enrichment

GO and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses are mainly performed by g:profiler and DAVID. To visualize the overlaps among enriched pathways, g:profiler, Cytoscape, and Cytoscape app EnrichmentMap are used.

Gene set enrichment analysis

GSEA are performed by GSEA software Linux version.

BRCA classification

TCGA BRCA samples are classified by hierarchical clustering based on by PAM50 genes' expressions. The classification is performed on the basis of the rule described in (*63*).

Genomic DNA and library preparation

Genomic DNA from differentiated cells or iPSC-MCs-tumor was prepared following the protocol for the Blood/Tissue/Cell Genome Extraction Kit (DP302). The exome sequences were efficiently enriched from 0.4 μg of genomic DNA using the Agilent liquid capture system (Agilent SureSelect Human All Exon V6) according to the manufacturer's protocol. First, qualified genomic DNA was randomly fragmented to an average size of 180 to 280 bp by a Covaris S220 sonicator. Remaining overhangs were converted into blunt ends via exonuclease polymerase activities. Second, DNA fragments were end repaired and phosphorylated, followed by A-tailing and ligation at the 3′ends with paired-end adaptors (Illumina). DNA fragments with ligated adapter molecules on both ends were selectively enriched in a PCR reaction. After PCR reaction, libraries were hybridized with liquid phase with biotin-labeled probe, and then magnetic beads with streptomycin were used to capture the exons of genes. Captured libraries were enriched in a PCR reaction to add index tags to prepare for sequencing. Products were purified using the AMPure XP system (Beckman Coulter, Beverly, USA). DNA concentration was measured using the Qubit 3.0 Flurometer (Invitrogen, USA), and libraries were analyzed for size distribution using NGS3K/caliper and quantified by real-time PCR (3 nM).

Clustering and sequencing

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using the Illumina PE Cluster Kit (Illumina, USA) according to the manufacturer's instructions. After cluster generation, the DNA libraries were sequenced on an Illumina platform, and 150-bp paired-end reads were generated.

Bulk sequence analysis

The quality of the raw FASTQ files was checked with FastQC. WES reads were mapped to hg19 using BWA. Then, SAMtools was used to rank the results by comparison; Picard was used to mark duplicate reads.

Single-cell RNA-seq library construction and data processing

The single-cell library was constructed using the Chromium Controller and the Chromium Single Cell 3′ Reagent Version 2 Kit (10x Genomics, PN-120237) according to the manufacturer's instructions. The final libraries were sequenced using the Illumina HiSeq 4000. For each sample, the cleaned data were generated by Cell Ranger (v3.0.2) (https://github.com/10XGenomics/cellranger) and filtered for the low-quality reads and unrelated sequences. The data were aligned to mouse mm10 reference genome. Data merging, thresholding, normalization, PCA, clustering analysis, visualization, and differential gene expression analysis were carried out in Seurat (v4.0.5) (https://satijalab.org/seurat) according to their recommended steps. In detail, cells were sorted on the basis of the barcodes, and the unique molecular identifiers (UMIs) were counted per gene for each cell. In total, averagely 6671 cells were captured for individual libraries, and averagely 1801 genes were detected with UMIs per cell. Cells having total mitochondria-expressed genes beyond 10% were eliminated, along with cells expressing less than 500 or greater than 5000 total genes. After this, we performed global normalization using the SCTransform function in Seurat. These preprocessed data were then analyzed to identify variable genes and PCA, and then Harmony was applied (https://github.com/immunogenomics/harmony) to integrate multiple samples. For further analysis, uniform manifold approximation and projection (UMAP) was used for dimensionality reduction. Cells were represented in a two-dimensional UMAP plane, and clusters were identified and annotated according to the previously published cell markers.

Mutation analysis

Somatic mutations in samples at different differentiation stages or iPSC-MCs-tumor were identified upon removing any mutations found in the corresponding iPSC samples or with insufficient coverage in the iPS samples. Contaminated mouse reads in iPSC-MCs-tumor sample were removed using BBSplit. Single-nucleotide polymorphism and InDel were picked by Strelka and annotated with ANNOVAR.

Circos plots

Circos plots are created by Circos following the tutorial. The clinical mutation profile was retrieved from cBioPortal. CrossMap was used to convert genome build to hg19. Genomes were divided into bins by 1 MB, and the number of how many mutations contained in each bin was counted. The 10% bins with the highest mutation numbers were identified as hotspots. Hypergeometric tests were applied to mutation hotspots of iPSC-MCs-tumor and clinical *BRCA1*-mutant breast tumor to determine the codistribution.

Enrichment in clinical mutation

Clinical mutation of *BRCA1*-mutant breast tumor was retrieved from cBioPortal. Recurrent clinical mutation was defined as mutation with a detected frequency of >50% (2 profiled samples), >25% (4 to 7 profiled samples), >15% (17 to 26 profiled samples), and >10% (31 to 42 profiled samples). Nonrecurrent clinical mutation was defined as mutation with a detected frequency of ≤50% (2 profiled samples), ≤25% (4 to 7 profiled samples), ≤15% (17 to 26 profiled samples), and ≤10% (31 to 42 profiled samples). Hypergeometric tests were used to determine the enrichment of recurrent and nonrecurrent clinical mutation in iPS-MCs and iPSC-MCs-tumor.

Standard statistical analysis

Adjusted P values were determined using an unpaired Student's t test, unless otherwise stated. Differences were considered statistically significant when $P < 0.05$ ($^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$). Data are shown as mean ± SD or SEM. Statistics were calculated using GraphPad Prism 6 software.

Supplementary Materials

The PDF file includes:

Figs. S1 to S9
Tables S1 to S7
Legends for tables S8 and S9
References

Other Supplementary Material for this manuscript includes the following:

Tables S8 and S9

REFERENCES AND NOTES

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
2. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
3. M. Greaves, Evolutionary determinants of cancer. *Cancer Discov.* **5**, 806–820 (2015).
4. U. Ben-David, R. Beroukhim, T. R. Golub, Genomic evolution of cancer models: Perils and opportunities. *Nat. Rev. Cancer* **19**, 97–109 (2019).
5. Y. Wang, F. Ye, Y. Liang, Q. Yang, Breast cancer brain metastasis: Insight into molecular mechanisms and therapeutic strategies. *Br. J. Cancer* **125**, 1056–1067 (2021).
6. L. C. Hofbauer, A. Bozec, M. Rauner, F. Jakob, S. Perner, K. Pantel, Novel approaches to target the microenvironment of bone metastasis. *Nat. Rev. Clin. Oncol.* **18**, 488–505 (2021).
7. S. Nik-Zainal, H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L. B. Alexandrov, S. Martin, D. C. Wedge, P. Van Loo, Y. S. Ju, M. Smid, A. B. Brinkman, S. Morganella, M. R. Aure, O. C. Lingjærde, A. Langerød, M. Ringnér, S.-M. Ahn, S. Boyault, J. E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J. A. Foekens, M. Gerstung, G. K. J. Hooijer, S. J. Jang, D. R. Jones, H.-Y. Kim, T. A. King, S. Krishnamurthy, H. J. Lee, J.-Y. Lee, Y. Li, S. M. Laren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporté, X. Pivot, C. A. Purdie, K. Raine, K. Ramakrishnan, F. G. Rodríguez-González, G. Romieu, A. M. Sieuwerts, P. T. Simpson, R. Shepherd, L. Stebbings, O. A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G. G. Van den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. van't Veer, A. Tutt, S. Knappskog, B. K. T. Tan, J. Jonkers, Å. Borg, N. T. Ueno, C. Sotiriou, A. Viari, P. A. Futreal, P. J. Campbell, P. N. Span, S. Van Laere, S. R. Lakhani, J. E. Eyfjord, A. M. Thompson, E. Birney, H. G. Stunnenberg, M. J. van de Vijver, J. W. M. Martens, A.-L. Børresen-Dale, A. L. Richardson, G. Kong, G. Thomas, M. R. Stratton, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
8. S. Tyanova, R. Albrechtsen, P. Kronqvist, J. Cox, M. Mann, T. Geiger, Proteomic maps of breast cancer subtypes. *Nat. Commun.* **7**, 10259 (2016).
9. B. D. Lehmann, A. Colaprico, T. C. Silva, J. Chen, H. An, Y. Ban, H. Huang, L. Wang, J. L. James, J. M. Balko, P. I. Gonzalez-Ericsson, M. E. Sanders, B. Zhang, J. A. Pietenpol, X. S. Chen, Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. *Nat. Commun.* **12**, 6276 (2021).
10. O. Ginsburg, P. Ashton-Prolla, A. Cantor, D. Mariosa, P. Brennan, The role of genomics in global cancer prevention. *Nat. Rev. Clin. Oncol.* **18**, 116–128 (2021).
11. S. S. Buys, J. F. Sandbach, A. Gammon, G. Patel, J. Kidd, K. L. Brown, L. Sharma, J. Saam, J. Lancaster, M. B. Daly, A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer* **123**, 1721–1730 (2017).
12. E. Matros, Z. C. Wang, G. Lodeiro, A. Miron, J. D. Iglehart, A. L. Richardson, BRCA1 promoter methylation in sporadic breast tumors: Relationship to gene expression profiles. *Breast Cancer Res. Treat.* **91**, 179–186 (2005).
13. M. Santarosa, R. Maestro, BRACking news on triple-negative/basal-like breast cancers: How BRCA1 deficiency may result in the development of a selective tumor subtype. *Cancer Metastasis Rev.* **31**, 131–142 (2012).
14. K. A. Won, C. Spruck, Triple-negative breast cancer therapy: Current and future perspectives (Review). *Int. J. Oncol.* **57**, 1245–1261 (2020).
15. K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
16. G. Tiscornia, E. L. Vivas, J. C. I. Belmonte, Diseases in a dish: Modeling human genetic disorders using induced pluripotent cells. *Nat. Med.* **17**, 1570–1576 (2011).
17. M. P. Chao, A. J. Gentles, S. Chatterjee, F. Lan, A. Reinisch, M. R. Corces, S. Xavy, J. Shen, D. Haag, S. Chanda, R. Sinha, R. M. Morganti, T. Nishimura, M. Ameen, H. Wu, M. Wernig, J. C. Wu, R. Majeti, Human AML-iPSCs reacquire leukemic properties after differentiation and model clonal variation of disease. *Cell Stem Cell* **20**, 329–344.e7 (2017).
18. Y. Qu, B. Han, B. Gao, S. Bose, Y. Gong, K. Wawrowsky, A. E. Giuliano, D. Sareen, X. Cui, Differentiation of human induced pluripotent stem cells to mammary-like organoids. *Stem Cell Rep.* **8**, 205–215 (2017).
19. A. Y. Propper, B. A. Howard, J. M. Veltmaat, Prenatal morphogenesis of mammary glands in mouse and rabbit. *J. Mammary Gland Biol. Neoplasia* **18**, 93–104 (2013).
20. G. Y. Lee, P. A. Kenny, E. H. Lee, M. J. Bissell, Three-dimensional culture models of normal and malignant breast epithelial cells. *Nat. Methods* **4**, 359–365 (2007).
21. P. Eirew, J. Stingl, A. Raouf, G. Turashvili, S. Aparicio, J. T. Emerman, C. J. Eaves, A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. *Nat. Med.* **14**, 1384–1389 (2008).
22. A. Kwong, J. C. W. Ho, V. Y. Shin, A. W. Kurian, E. Tai, L. J. Esserman, J. N. Weitzel, P. H. Lin, M. Field, S. M. Domchek, J. Lo, H. Y. S. Ngan, E. S. K. Ma, T. L. Chan, J. M. Ford, Rapid detection of BRCA1/2 recurrent mutations in Chinese breast and ovarian cancer patients with multiplex SNaPshot genotyping panels. *Oncotarget* **9**, 7832–7843 (2018).
23. R. Roy, J. Chun, S. N. Powell, BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12**, 68–78 (2011).
24. X. Wu, M. Guo, J. Cui, H. Cai, S. M. Wang, Heterozygotic Brca1 mutation initiates mouse genome instability at embryonic stage. *Oncogenesis* **11**, 41 (2022).
25. M. Adorno, M. Cordenonsi, M. Montagner, S. Dupont, C. Wong, B. Hann, A. Solari, S. Bobisse, M. B. Rondina, V. Guzzardo, A. R. Parenti, A. Rosato, S. Bicciato, A. Balmain, S. Piccolo, A mutant-p53/Smad complex opposes p63 to empower TGFβ-induced metastasis. *Cell* **137**, 87–98 (2009).
26. M. S. Chapman, I. M. Verma, Transcriptional activation by BRCA1. *Nature* **382**, 678–679 (1996).
27. T. Ouchi, A. N. Monteiro, A. August, S. A. Aaronson, H. Hanafusa, BRCA1 regulates p53-dependent gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2302–2306 (1998).
28. P. B. Mullan, J. E. Quinn, D. P. Harkin, The role of BRCA1 in transcriptional regulation and cell cycle control. *Oncogene* **25**, 5854–5863 (2006).
29. X. Zhang, R. Li, BRCA1-dependent transcriptional regulation: Implication in tissue-specific tumor suppression. *Cancers* **10**, 513 (2018).
30. H. David, Rudolf Virchow and modern aspects of tumor pathology. *Pathol. Res. Pract.* **183**, 356–364 (1988).
31. M. Zvelebil, E. Oliemuller, Q. Gao, O. Wansbury, A. Mackay, H. Kendrick, M. J. Smalley, J. S. Reis-Filho, B. A. Howard, Embryonic mammary signature subsets are activated in Brca1−/− and basal-like breast cancers. *Breast Cancer Res.* **15**, R25 (2013).
32. M. H. Wright, A. M. Calcagno, C. D. Salcido, M. D. Carlson, S. V. Ambudkar, L. Varticovski, Brca1 breast tumors contain distinct $CD44^{+}/CD24^{-}$ and $CD133^{+}$ cells with cancer stem cell characteristics. *Breast Cancer Res.* **10**, R10 (2008).
33. D. Pellacani, S. Tan, S. Lefort, C. J. Eaves, Transcriptional regulation of normal human mammary cell heterogeneity and its perturbation in breast cancer. *EMBO J.* **38**, e100330 (2019).
34. B. Howard, A. Ashworth, Signalling pathways implicated in early mammary gland morphogenesis and breast cancer. *PLoS Genet.* **2**, e112 (2006).
35. A. Prat, C. M. Perou, Mammary development meets cancer genomics. *Nat. Med.* **15**, 842–844 (2009).
36. M. L. Asselin-Labat, K. D. Sutherland, H. Barker, R. Thomas, M. Shackleton, N. C. Forrest, L. Hartley, L. Robb, F. G. Grosveld, J. van der Wees, G. J. Lindeman, J. E. Visvader, Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat. Cell Biol.* **9**, 201–209 (2007).
37. J. Russo, I. H. Russo, Development of the human breast. *Maturitas* **49**, 2–15 (2004).
38. J. M. Veltmaat, A. A. Mailleux, J. P. Thiery, S. Bellusci, Mouse embryonic mammogenesis as a model for the molecular regulation of pattern formation. *Differentiation* **71**, 1–17 (2003).
39. F. Magdinier, N. Dalla Venezia, G. M. Lenoir, L. Frappart, R. Dante, BRCA1 expression during prenatal development of the human mammary gland. *Oncogene* **18**, 4039–4043 (1999).
40. M. Crespo, E. Vilar, S. Y. Tsai, K. Chang, S. Amin, T. Srinivasan, T. Zhang, N. H. Pipalia, H. J. Chen, M. Witherspoon, M. Gordillo, J. Z. Xiang, F. R. Maxfield, S. Lipkin, T. Evans, S. Chen, Colonic organoids derived from human induced pluripotent stem cells for modeling colorectal cancer and drug testing. *Nat. Med.* **23**, 878–884 (2017).
41. A. G. Kotini, C. J. Chang, A. Chow, H. Yuan, T. C. Ho, T. Wang, S. Vora, A. Solovyov, C. Husser, M. Olszewska, J. Teruya-Feldstein, D. Perumal, V. M. Klimek, A. Spyridonidis, R. K. Rampal, L. Silverman, E. P. Reddy, E. Papaemmanuil, S. Parekh, B. D. Greenbaum, C. S. Leslie,

M. G. Kharas, E. P. Papapetrou, Stage-specific human induced pluripotent stem cells map the progression of myeloid transformation to transplantable leukemia. *Cell Stem Cell* **20**, 315–328.e7 (2017).
42. J. A. Schroeder, M. C. Adriance, M. C. Thompson, T. D. Camenisch, S. J. Gendler, MUC1 alters β-catenin-dependent tumor formation and promotes cellular invasion. *Oncogene* **22**, 1324–1332 (2003).
43. N. J. Chew, E. V. Nguyen, S. P. Su, K. Novy, H. C. Chan, L. K. Nguyen, J. Luu, K. J. Simpson, R. S. Lee, R. J. Daly, FGFR3 signaling and function in triple negative breast cancer. *Cell Commun. Signal* **18**, 13 (2020).
44. K. J. Cocce, J. S. Jasper, T. K. Desautels, L. Everett, S. Wardell, T. Westerling, R. Baldi, T. M. Wright, K. Tavares, A. Yllanes, Y. Bae, J. T. Blitzer, C. Logsdon, D. P. Rakiec, D. A. Ruddy, T. Jiang, G. Broadwater, T. Hyslop, A. Hall, M. Laine, D. P. McDonnell, The lineage determining factor GRHL2 collaborates with FOXA1 to establish a targetable pathway in endocrine therapy-resistant breast cancer. *Cell Rep.* **29**, 889–903.e10 (2019).
45. Y. Xue, L. Lai, W. Lian, X. Tu, J. Zhou, P. Dong, D. Su, X. Wang, X. Cao, Y. Chen, Q. Wang, SOX9/FXYD3/Src axis is critical for ER^+ breast cancer stem cell function. *Mol. Cancer Res.* **17**, 238–249 (2019).
46. W. Burke, M. Daly, J. Garber, J. Botkin, M. J. Kahn, P. Lynch, A. McTiernan, K. Offit, J. Perlman, G. Petersen, E. Thomson, C. Varricchio, Recommendations for follow-up care of individuals with an inherited predisposition to cancer. II. BRCA1 and BRCA2. Cancer genetics studies consortium. *JAMA* **277**, 997–1003 (1997).
47. A. Eisen, T. R. Rebbeck, W. C. Wood, B. L. Weber, Prophylactic surgery in women with a hereditary predisposition to breast and ovarian cancer. *J. Clin. Oncol.* **18**, 1980–1995 (2000).
48. M. B. Terry, M. B. Daly, K. A. Phillips, X. Ma, N. Zeinomar, N. Leoce, G. S. Dite, R. J. MacInnis, W. K. Chung, J. A. Knight, M. C. Southey, R. L. Milne, D. Goldgar, G. G. Giles, P. C. Weideman, G. Glendon, R. Buchsbaum, I. L. Andrulis, E. M. John, S. S. Buys, J. L. Hopper, Risk-reducing oophorectomy and breast cancer risk across the spectrum of familial risk. *J. Natl. Cancer Inst.* **111**, 331–334 (2019).
49. J. Kotsopoulos, T. Huzarski, J. Gronwald, C. F. Singer, P. Moller, H. T. Lynch, S. Armel, B. Karlan, W. D. Foulkes, S. L. Neuhausen, L. Senter, N. Tung, J. N. Weitzel, A. Eisen, K. Metcalfe, C. Eng, T. Pal, G. Evans, P. Sun, J. Lubinski, S. A. Narod, Hereditary Breast Cancer Clinical Study Group, Bilateral oophorectomy and breast cancer risk in BRCA1 and BRCA2 mutation carriers. *J. Natl. Cancer Inst.* **109**, djw177 (2016).
50. P. L. Mai, A. Miller, M. H. Gail, S. Skates, K. Lu, M. E. Sherman, O. B. Ioffe, G. Rodriguez, D. E. Cohn, J. Boggess, T. Rutherford, N. D. Kauff, J. S. Rader, K. A. Phillips, P. A. DiSilvestro, A. B. Olawaiye, M. R. Ridgway, M. H. Greene, M. Piedmonte, J. L. Walker, Risk-reducing salpingo-oophorectomy and breast cancer risk reduction in the Gynecologic Oncology Group Protocol-0199 (GOG-0199). *JNCI Cancer Spectr.* **4**, pkz075 (2020).
51. L. Xu, Y. Zhao, Z. Chen, Y. Wang, L. Chen, S. Wang, Tamoxifen and risk of contralateral breast cancer among women with inherited mutations in BRCA1 and BRCA2: A meta-analysis. *Breast Cancer* **22**, 327–334 (2015).
52. A. Phillips, R. L. Milne, M. A. Rookus, M. B. Daly, A. C. Antoniou, S. Peock, D. Frost, D. F. Easton, S. Ellis, M. L. Friedlander, S. S. Buys, N. Andrieu, C. Noguès, D. Stoppa-Lyonnet, V. Bonadona, P. Pujol, S. A. McLachlan, E. M. John, M. J. Hooning, C. Seynaeve, R. A. Tollenaar, D. E. Goldgar, M. B. Terry, T. Caldes, P. C. Weideman, I. L. Andrulis, C. F. Singer, K. Birch, J. Simard, M. C. Southey, H. L. Olsson, A. Jakubowska, E. Olah, A. M. Gerdes, L. Foretova, J. L. Hopper, Tamoxifen and risk of contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *J. Clin. Oncol.* **31**, 3091–3099 (2013).
53. V. V. Giridharan, J. S. Generoso, A. Collodel, D. Dominguini, C. J. Faller, F. Tardin, G. S. Bhatti, F. Petronilho, F. Dal-Pizzol, T. Barichello, Receptor for advanced glycation end products (RAGE) mediates cognitive impairment triggered by pneumococcal meningitis. *Neurotherapeutics* **18**, 640–653 (2021).
54. G. Santos, A. Barateiro, D. Brites, A. Fernandes, S100B impairs oligodendrogenesis and myelin repair following demyelination through RAGE engagement. *Front. Cell. Neurosci.* **14**, 279 (2020).
55. J. Wang, R. Li, Z. Peng, B. Hu, X. Rao, J. Li, HMGB1 participates in LPS-induced acute lung injury by activating the AIM2 inflammasome in macrophages and inducing polarization of M1 macrophages via TLR2, TLR4, and RAGE/NF-κB signaling pathways. *Int. J. Mol. Med.* **45**, 61–80 (2020).
56. K. Kikuchi, K. M. McNamara, Y. Miki, E. Iwabuchi, A. Kanai, M. Miyashita, T. Ishida, H. Sasano, S100P and ezrin promote trans-endothelial migration of triple negative breast cancer cells. *Cell. Oncol.* **42**, 67–80 (2019).
57. Y. Cong, Y. Cui, S. Wang, L. Jiang, J. Cao, S. Zhu, E. Birkin, J. Lane, F. Ruge, W. G. Jiang, G. Qiao, Calcium-binding protein S100P promotes tumor progression but enhances chemosensitivity in breast cancer. *Front. Oncol.* **10**, 566302 (2020).
58. J. Tchieu, B. Zimmer, F. Fattahi, S. Amin, N. Zeltner, S. Chen, L. Studer, A modular platform for differentiation of human PSCs into all major ectodermal lineages. *Cell Stem Cell* **21**, 399–410.e7 (2017).
59. C. Ginestier, M. H. Hur, E. Charafe-Jauffret, F. Monville, J. Dutcher, M. Brown, J. Jacquemier, P. Viens, C. G. Kleer, S. Liu, A. Schott, D. Hayes, D. Birnbaum, M. S. Wicha, G. Dontu, ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555–567 (2007).
60. C. Kuperwasser, T. Chavarria, M. Wu, G. Magrane, J. W. Gray, L. Carey, A. Richardson, R. A. Weinberg, Reconstruction of functionally normal and malignant human breast tissues in mice. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4966–4971 (2004).
61. A. Ucar, V. Vafaizadeh, H. Jarry, J. Fiedler, P. A. Klemmt, T. Thum, B. Groner, K. Chowdhury, miR-212 and miR-132 are required for epithelial stromal interactions necessary for mouse mammary gland development. *Nat. Genet.* **42**, 1101–1108 (2010).
62. D. F. Lee, J. Su, H. S. Kim, B. Chang, D. Papatsenko, R. Zhao, Y. Yuan, J. Gingold, W. Xia, H. Darr, R. Mirzayans, M. C. Hung, C. Schaniel, I. R. Lemischka, Modeling familial cancer with induced pluripotent stem cells. *Cell* **161**, 240–254 (2015).
63. B. Wallden, J. Storhoff, T. Nielsen, N. Dowidar, C. Schaper, S. Ferree, S. Liu, S. Leung, G. Geiss, J. Snider, T. Vickery, S. R. Davies, E. R. Mardis, M. Gnant, I. Sestak, M. J. Ellis, C. M. Perou, P. S. Bernard, J. S. Parker, Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med. Genomics* **8**, 54 (2015).
64. M. Cordenonsi, F. Zanconato, L. Azzolin, M. Forcato, A. Rosato, C. Frasson, M. Montagner, M. Inui, A. R. Parenti, A. Poletti, M. G. Daidone, S. Dupont, G. Basso, S. Bicciato, S. Piccolo, The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell* **147**, 759–772 (2011).
65. M. Meng, M. Vaapil, A. Tagmount, A. Loguinov, C. Vulpe, P. Yaswen, Propagation of functional estrogen receptor positive normal human breast cells in 3D cultures. *Breast Cancer Res. Treat.* **176**, 131–140 (2019).
66. M. J. Stebbins, B. D. Gastfriend, S. G. Canfield, M. S. Lee, D. Richards, M. G. Faubion, W. J. Li, R. Daneman, S. P. Palecek, E. V. Shusta, Human pluripotent stem cell-derived brain pericyte-like cells induce blood-brain barrier properties. *Sci. Adv.* **5**, eaau7375 (2019).
67. J. G. Clark, K. H. Kim, R. S. Basom, S. A. Gharib, Plasticity of airway epithelial cell transcriptome in response to flagellin. *PLOS ONE* **10**, e0115486 (2015).
68. H. Araki, F. Miura, A. Watanabe, C. Morinaga, F. Kitaoka, Y. Kitano, N. Sakai, Y. Shibata, M. Terada, S. Goto, S. Yamanaka, M. Takahashi, T. Ito, Base-resolution methylome of retinal pigment epithelial cells used in the first trial of human induced pluripotent stem cell-based autologous transplantation. *Stem Cell Rep.* **13**, 761–774 (2019).
69. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
70. H. K. Long, M. Osterwalder, I. C. Welsh, K. Hansen, J. O. J. Davies, Y. E. Liu, M. Koska, A. T. Adams, R. Aho, N. Arora, K. Ikeda, R. M. Williams, T. Sauka-Spengler, M. H. Porteus, T. Mohun, D. E. Dickel, T. Swigut, J. R. Hughes, D. R. Higgs, A. Visel, L. Selleri, J. Wysocka, Loss of extreme long-range enhancers in human neural crest drives a craniofacial disorder. *Cell Stem Cell* **27**, 765–783.e14 (2020).
71. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
72. G. Yu, L. Wang, Y. Han, Q. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
73. S. Lê, J. Josse, F. Husson, FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
74. L. Kolberg, U. Raudvere, I. Kuzmin, P. Adler, J. Vilo, H. Peterson, g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023).
75. H. Wickham, "Toolbox" in *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016), pp. 33–74.
76. D. Merico, R. Isserlin, O. Stueker, A. Emili, G. D. Bader, Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLOS ONE* **5**, e13984 (2010).
77. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
78. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
79. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
80. G. H. Putri, S. Anders, P. T. Pyl, J. E. Pimanda, F. Zanini, Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
81. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
82. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

Acknowledgments
Funding: This work was supported by grants from the National Key R&D Program of China (2023YFA1800900), National Key Research and Development Program of China (2024YFA1106900), the National Science Foundation for Distinguished Young Scholars of China (32425022), National Natural Science Foundation of China (32170798 and 32430031), and the Guangdong Innovative and Entrepreneurial Research Team Program (2016ZT06S029) to J.D.; National Natural Science Foundation of China (82203334) and Post-doctoral Program (2023 M740821) to J.Li.; the National Natural Science Foundation of China (32470844) and the Natural Science Foundation of Guangdong Province, China (2023A1515010197); Natural Science Foundation of Guangdong Province, China (2023A1515010148) to J.S.; and the National Natural Science Foundation of China (32370972) and the Guangdong Basic and Applied Basic Research Foundation (2024B1515020052 and 2023A1515011783) to S.Su.
Author contributions: Conceptualization: J.Li., C.Z., Jiahao Chen, X.Z., W.C., D.G., J. Li, D.-F.L., J.Ba., H.H., D.-f.H., N.C., and J.D. Methodology: J.Li., C.Z., Jiahao Chen, X.La., X.Z., P.F., W.C., D.G., Junchao Cai, Y.H., J.Ba., H.H., D.-f.H., N.C., S. Su, and J.D. Software: P.Z., R.D., X.Z., and J.S. Validation: J.Li., C.Z., Jiahao Chen, P.Z., Q.L., X.La., X.Z., P.F., T.W., W.C., N.C., and J.D. Formal analysis: J.Li., C.Z., Jiahao Chen, P.Z., X.Z., X.Li., J.S., P.F., and S. Suo. Investigation: J.Li., C.Z., Jiahao Chen, Q.L., W.S., Jianing Chen, X.Z., J.W., P.F., T.W., Y.H., N.C., S. Suo, and J.D. Resources: J.Li., C.Z., P.Z., X.La., J.S., P.F., J. Li, Y.H., A.P.X., Y.A.Z., J.Z., N.C., S. Suo, S. Su, and J.D. Data curation: J.Li., C.Z., P.Z., X.Z., J.S., and J.D. Writing—original draft: J.Li., C.Z., Jiahao Chen, P.Z., J.S., J.W., D.G., D.-F.L., N.C., and J.D. Writing—review and editing: J.Li., C.Z., Jiahao Chen, J.S., D.G., J.Be., H.H., D.-f.H., N.C., and J.D. Visualization: J.Li., C.Z., Jiahao Chen, X.Z., X.Li., J.S., N.C., and J.D. Supervision: J.Li., N.C., W.C., and J.D. Project administration: S. Suo and J.D. Funding acquisition: J.Li., J.S., S. Suo, and J.D. **Competing interests:** J.D., J.Li., and C.Z. are inventors on authorized patent (ZL 2018 1 1436493.8, China) held by Sun Yat-sen University that covers the method for differentiating iPSCs toward mammary lineage. All other authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All the described RNA-seq and WES data that produced in this study have been deposited in the Sequence Read Archive under the accession number PRJNA759105 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA759105). Other publicly available data used in this study are provided in table S6, and the open-source software used in this study is provided in table S7.

Submitted 17 April 2023
Accepted 24 June 2025
Published 25 July 2025
10.1126/sciadv.adi2370