

Dissecting the heterogeneity and tumorigenesis of BRCA1 deficient mammary tumors via single cell RNA sequencing

Heng Sun^{1,2,3,4}, Jianming Zeng^{1,2}, Zhengqiang Miao^{1,2}, Kuan Cheok Lei^{1,2}, Chen Huang^{1,2}, Lingling Hu^{1,2}, Sek Man Su^{1,2}, Un In Chan^{1,2}, Kai Miao^{1,2,3}, Xu Zhang^{1,2}, Aiping Zhang^{1,2}, Sen Guo^{1,2}, Si Chen^{1,2}, Ya Meng^{1,2,5}, Min Deng^{1,2}, Wenhui Hao^{1,2}, Haipeng Lei^{1,2}, Ying Lin⁶, Zhonglin Yang^{7,8}, Dongjiang Tang^{7,8}, Koon Ho Wong^{1,2}, Xiaohua Douglas Zhang^{1,2,3}, Xiaoling Xu^{1,2,3}, Chu-Xia Deng^{1,2,3}✉

1. Cancer Center, Faculty of Health Sciences, University of Macau, Macau SAR, China.

2. Centre for Precision Medicine Research and Training, Faculty of Health Sciences, University of Macau, Macau SAR, China.

3. MOE Frontiers Science Center for Precision Oncology, University of Macau, Macau SAR, China.

4. Zhuhai Research Institute, University of Macau, China.

5. Zhuhai Precision Medical Center, Zhuhai People's Hospital, Jinan University, Zhuhai, China.

6. Breast Disease Center, the First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China.

7. Zhuhai SanMed Biotech Ltd, Zhuhai, China.

8. Joint Research Center of Liquid Biopsy in Guangdong, Hong Kong and Macao, Zhuhai, China.

✉ Corresponding author: Chu-Xia Deng, PhD, Phone: (853) 8822-4997; E-mail: cx.deng@um.edu.mo.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2021.06.17; Accepted: 2021.10.08; Published: 2021.10.25

Abstract

Background: BRCA1 plays critical roles in mammary gland development and mammary tumorigenesis. And loss of BRCA1 induces mammary tumors in a stochastic manner. These tumors present great heterogeneity at both intertumor and intratumor levels.

Methods: To comprehensively elucidate the heterogeneity of BRCA1 deficient mammary tumors and the underlying mechanisms for tumor initiation and progression, we conducted bulk and single cell RNA sequencing (scRNA-seq) on both mammary gland cells and mammary tumor cells isolated from *Brcal* knockout mice.

Results: We found the BRCA1 deficient tumors could be classified into four subtypes with distinct molecular features and different sensitivities to anti-cancer drugs at the intertumor level. Whereas within the tumors, heterogeneous subgroups were classified mainly due to the different activities of cell proliferation, DNA damage response/repair and epithelial-to-mesenchymal transition (EMT). Besides, we reconstructed the BRCA1 related mammary tumorigenesis to uncover the transcriptomes alterations during this process via pseudo-temporal analysis of the scRNA-seq data. Furthermore, from candidate markers for BRCA1 mutant tumors, we discovered and validated one oncogene *Mrc2*, whose loss could reduce mammary tumor growth *in vitro* and *in vivo*.

Conclusion: Our study provides a useful resource for better understanding of mammary tumorigenesis induced by BRCA1 deficiency.

Key words: *Brcal/BRCA1*; single cell RNA-seq; tumor heterogeneity; mammary tumor; *Mrc2*

Introduction

Germline mutations of *BRCA1* predispose women to early onset of breast and ovarian cancers [1]. And majority of *BRCA1*-related breast cancers belong to the most refractory triple negative breast cancer (TNBC) subtype, bearing basal-like feature [2], though the cancers may mainly originate from the

luminal mammary epithelia [3, 4]. Meanwhile, mutations of *BRCA1* have also been identified in other subtypes of breast cancers [2]. Through its diverse functions in DNA damage repair, cell cycle control, transcription regulation, ubiquitination and so on, *BRCA1* acts as a very significant tumor suppressor

and genomic safeguard [5, 6]. BRCA1 deficiency induces severe cellular stress; when occurring in the mammary glands, it impairs the regular developmental process and eventually causes tumorigenesis due to accumulation of genome instability and other alterations [6-9].

We have previously established a mouse model carrying mammary specific knockout of *Brca1* (MMTV-Cre; *Brca1^{fl/fl}*), which spontaneously develops mammary tumors in a stochastic manner [10]. The tumors display variable histopathological features, which could even be detected at different intratumor regions [11]. Such inter- and intra-tumor heterogeneity could also be observed in human breast cancers carrying *BRCA1* mutations. According to previous studies, around 40% of *BRCA1*-related breast cancers belong to ductal carcinoma, not otherwise specified (ductal NOS) type, while typical medullary carcinoma (MC) feature can be observed in 7.8% to 13% of *BRCA1*-associated breast carcinomas [12-14]. More than half of *BRCA1*-related breast cancers harbor *TP53* mutation and *MYC* amplification [15]. Besides, about 78% of breast cancers arising in *BRCA1* mutation carriers are ER-negative, 79% are PR-negative, 90% are HER2-negative, and 69% are Triple-negative [16]. Such heterogeneity of histopathology, genetic alterations and gene expression patterns not only reflects the complexity of mammary tumorigenesis induced by *BRCA1* deficiency, but also leads to the difficulty of the efficient treatment on such refractory cancers. Meanwhile, more details are still badly lacking to figure out the origin and evolution of the heterogeneity among and within *BRCA1*-deficient mammary tumors.

scRNA-seq has been widely harnessed to uncover the diverse cell types within heterogeneous cell populations, like many well differentiated organs [17] and many types of cancers [18]. In this study, we performed bulk RNA-seq and other experiments to elucidate the intertumor heterogeneity among *BRCA1* deficient mammary tumors, and scRNA-seq of more than 20000 mammary cells and tumor cells to reveal the intratumor heterogeneity and gene expression variations during the mammary tumorigenesis process. We also identified and validated one oncogene *Mrc2*, whose loss could block the tumor cell growth both *in vitro* and *in vivo*.

Materials and methods

Experimental mice

All animals were housed and handled in accordance with the guidelines of the Institutional Animal Care and Use Committee of the University

of Macau. The MMTV-Cre; *Brca1^{fl/fl}*, and MMTV-Cre; *Brca1^{fl/fl}*; *Trp53^{fl/+}* mice were described previously [10] and were crossed with *ROSA^{mT/mG}* mice [19] to generate the MMTV-Cre; *Brca1^{fl/fl}*; *ROSA^{mT/mG}*, MMTV-Cre; *Trp53^{fl/+}*; *ROSA^{mT/mG}*, and MMTV-Cre; *Brca1^{fl/fl}*; *Trp53^{fl/+}*; *ROSA^{mT/mG}* strains for monitoring the CRE-active cells in mammary gland. Mice were in a C57BL/6 and 129/Sv mixed background. The presence of floxed, deleted *Trp53* and *Brca1* alleles, *mTmG* alleles and of the *Cre* recombinase allele was confirmed by PCR with tail DNA as template [10, 19].

Cell lines

The immortalized mouse *Brca1*-WT epithelial cell line B477 and *BRCA1*-mutant epithelial cell line G600 were derived from the mammary gland of *Brca1*-WT (*Brca1^{+/+}*; *Trp53^{+/+}*) and *Brca1*-mutant (*Brca1^{Δ/Δ}*; *Trp53^{+/+}*) mice, respectively [20]. Several *BRCA1* deficient mouse mammary tumor cell lines were derived from the mammary tumors of the MMTV-Cre; *Brca1^{fl/fl}*, or MMTV-Cre; *Brca1^{fl/fl}*; *Trp53^{fl/+}* mice. The MCF7, MDA-MB-231, MDA-MB-436 cell lines, and the 293FT cell line for lentiviral production were obtained from American Tissue Culture Collection (ATCC). All cells were cultured in Dulbecco's modified Eagle medium (DMEM) (Gibco Life Technologies) containing 10% FBS with or without 100 IU/mL penicillin, and 100 µg/mL streptomycin.

Histology and antibody staining

For histology, tissues were fixed in the 10% formalin, blocked in paraffin, sectioned as 5 µm thick, stained with hematoxylin and eosin, and examined by light microscopy. Antibodies against ERalpha (sc-542; Santa Cruz), PR (sc-166169; Santa Cruz), ERBB2 (sc-33684; Santa Cruz), and MRC2 (sc-271148; Santa Cruz) were used for immunohistochemical staining, by using a Histostain® Plus Broad Spectrum kit (859043; Life Technologies) as per the manufacturer's instruction. Antibodies against Keratin14 (PRB-155P; Covance), Keratin18 (sc-53256; Santa Cruz), E-Cadherin (3195; Cell Signaling Technology) and Vimentin (5741; Cell Signaling Technology) were used for immunofluorescent staining, and nuclei were stained with DAPI (62248; ThermoFisher Scientific).

Analysis of public human breast cancer data

We reanalyzed molecular data on 6113 breast cancers from previous studies collected in cBioPortal, including data from METABRIC and TCGA, etc. [21-28]. Among these cancer samples, 169 (2.76%) harbored the genetic alteration on *BRCA1*. While some alterations of *BRCA1*, like amplification, may not be oncogenic as *BRCA1* function as a well-known tumor suppressor, others including deep deletion,

truncating and missense mutations (putative driver) were considered as *BRCA1* deficient alterations here. Within the 63 *BRCA1* genetically deficient cases, matched RNA-seq data were only available from 13 samples in TCGA database. Meanwhile, 16 TCGA breast cancers with germline pathogenic *BRCA1* mutation [29] were also identified. On these 29 *BRCA1* deficient breast cancers with RNA-seq, we performed the hierarchy clustering analysis based on overall transcriptome and molecular subtyping analysis based on PAM50 according to previous study [30]. Besides, a heatmap describing the expression patterns of several mammary basal/luminal lineage markers and epithelial/mesenchymal markers were drawn based on the RNA-seq data by using R [31].

To determine the protein level of MRC2 in human breast tissue and breast cancers, IHC results are cited from human protein atlas (<https://www.proteinatlas.org/>); and the relative protein levels of MRC2 are collected from Clinical Proteomic Tumor Analysis Consortium (CPTAC) Confirmatory/Discovery dataset [32].

Bulk tissue RNA sequencing

The mammary tumors from *Brca1*-deficient mice were cut into small pieces and homogenized in TRIzol reagent (15596018; ThermoFisher scientific) by using the Precellys®24 homogenizer (P000669-PR240-A; Bertin Technologies) and tissue homogenizing CKmix (KTO3961-1-009.2; Bertin Technologies) as per the manufacturer's instruction. About 5×10^5 freshly FACS sorted DAPI-CD24^{Hi}CD29^{Lo} luminal cells were re-suspended in TRIzol reagent after washing with PBS. After isolated from the TRIzol lysate, the RNA samples were constructed into RNA sequencing libraries by using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (E7530L, New England Biolabs) according to the manufacturer's instruction. Sequencing was performed on Illumina HiSeq2500 or HiSeq X Ten platforms by multiplexed paired end run with 100 or 150 cycles.

The RNA sequencing reads were aligned to the reference mm10 mouse genome with TopHat 2.1.1 by using default parameters [33]. Gene expression was quantified as FPKM using Cufflinks 2.2.1 [34]. To cluster the 23 *Brca1*-deficient mouse mammary tumors into subgroups, hierarchical clustering was performed using R with default parameters. Differentially expressed genes among sub-clusters were identified by using DESeq2 [35]. And, the following cellular function analysis of differentially expressed genes was performed using DAVID [36].

Single cells suspension preparation

Three-to-four month-old female virgin or

pregnant at day 12.5 mice were sacrificed to isolate the fourth pair of mammary glands for single-cell preparation as previously described [37]. In brief, the mammary glands were minced, washed in PBS and digested in DMEM/F12 containing 300 U/mL collagenase III (S4M7602S, Worthington, Lakewood, NJ), 100 U/mL hyaluronidase (H3506; Sigma-Aldrich), 5% FBS (Gibco), 5 µg/mL insulin (350-020; Biosource, Rockville, MD), 10 ng/mL EGF (13247-051, Invitrogen), and 500 ng/mL hydrocortisone (H0888; Sigma-Aldrich) for about 1 hour at 37 °C, 5% CO₂. The resultant organoid suspension was sequentially re-suspended in DMEM/F12 supplemented with 5 mg/mL dispase II (825-001; Roche Diagnostics, Indianapolis, IN) and 0.1 mg/mL DNase I (58C10349; Worthington) for 5 min at 37 °C, and then digested with 0.25% trypsin-EDTA for 1-2 min, and treated with RBC lysis buffer (00433-57; eBioscience, San Diego, CA) for 3 min to remove the red blood cells before filtration through a 40-µm cell strainer (352340, BD Falcon) to obtain single-cell suspension. Enrichment of Lin⁻ epithelial cells was achieved through selective depletion of hematopoietic and endothelial cells by using the EasySeq mouse epithelial cell enrichment kit according to the manufacturer (19758; Stemcell). When the mammary tumors grew to around 1 cm in diameter, they were removed from the *Brca1*-deficient mice for single-cell preparation as per the same method.

Fluorescence-activated cell sorting (FACS)

The enriched Lin⁻ mouse mammary epithelial cells and tumor cells were stained at a concentration of 1×10^6 cells per 100 µl of FACS buffer (HBSS with 0.5% BSA and 2 mM EDTA). Antibodies against mouse CD24 (anti CD24-PE-Cy™7, 560536, BD), and CD29 (anti CD29-APC, 102216, Biolegend) were used to stain the cells. DAPI (1 µg/mL) was also used as an indicator for cell viability. Antibody incubations were performed at 4 °C for 30 min. Then the cells were washed with PBS for twice before sorting by using FACSAria III Cell Sorter (Becton Dickinson). DAPI-CD24^{Hi}CD29^{Lo}mGFP⁺ mammary luminal cells and DAPI-CD24⁺ tumor cells were sorted out for following single cell RNA sequencing.

Fluidigm C1 platform based single cell capture and RNA sequencing

The sorted mammary luminal cells or tumor cells re-suspended in FACS buffer (300-500 cells/µL) were mixed (3:2 ratio) with C1 Cell Suspension Reagent (100-6201, Fluidigm) before loading onto a 10 to 17-µm-diameter C1 Integrated Fluidic Circuit (IFC 100-5760, Fluidigm) [38]. Each capture site was carefully examined and recorded under a ZEISS

microscope in bright field channel for cell doublets. Cell lysing, reverse transcription and cDNA amplification were performed on the C1 Single-Cell Auto Prep IFC, as specified by the manufacturer (protocol 100-7168 E1). The SMARTer Ultra Low RNA Kit (634936, Clontech) was used for cDNA synthesis from the single cells. Illumina NGS library was constructed with Nextera XT DNA Sample Prep kit (FC-131-1096, Illumina), according to the manufacturer's recommendations (protocol 100-7168 E1). Sequencing was performed on Illumina HiSeq2500 (Illumina) high output run mode by multiplexed paired end run with 60, 100 or 125 cycles.

The single cell RNA-seq reads were filtered with sickle for quality control following the parameter "sickle pe -q 30 -l 36 -g", then trimmed using trim_galore as per parameter "trim_galore -q 25 --phred33 --length 50 -e 0.1 --stringency 3 --paired". The trimmed sequences were aligned to the reference mm10 mouse genome with STAR-2.5.3a by using default parameters. Gene expression was quantified as raw counts using subread. To cluster the single cells into subgroups, PCA analysis with top 2000 highly variable genes was performed by using R package Seurat and projected on tSNE plots. Default parameters of the pipeline within Seurat were used to find the main clusters of single cells (FindClusters) with a customized setting for resolution = 0.7. Cluster marker genes were calculated using Wilcoxon rank-sum test contrasting each cluster against the rest and selected with the cutoff p value < 0.01. And, the following GO analysis of the marker genes was performed using DAVID [36]. To get the CNVs information by using scRNA-seq data, R package inferCNV was used with the default parameters. And the scRNA-seq data of wild type mammary gland cells from our previous study was used as normal control [37]. Pipeline M3Drop [39] was utilized for cluster and marker genes analysis of tumor cells within each tumor following the default parameter except the customized setting k = 3 for the number of sub-clusters. Top 20 marker genes for each sub-clustered were displayed in the heatmaps, while all the marker genes selected with AUC > 0.7 and p value < 0.05 were used for GO analysis via DAVID. The analysis on activation levels of hallmark pathways of cancer was performed by using GSVA based on the gene sets from MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/>).

Dropseq of single mammary cells and tumor cells

The single cell suspension dissociated from mouse mammary glands or mammary tumors (after epithelia enrichment) were loaded to the microfluidic

droplet generator for dropseq as per the protocol described by McCarroll's lab [40]. The dropseq libraries were constructed according to the same protocol and sequenced on Illumina HiSeq X Ten platform with paired-end 150 bp.

Dropseq datasets were processed with Dropseq tools v.1.13. Cell and molecular barcode tagging, primer and poly-A trimming, read alignment, barcode recovering, gene/exon annotations and barcode repairs were sequentially accomplished according to the online protocol (<https://github.com/broadinstitute/Dropseq/files/2425535/DropseqAlignmentCookbookv1.2Jan2016.pdf>). Read alignment was done using STAR aligner against Ensembl Mus musculus GRCm38.p6 assembly. Counts from all libraries were aggregated into MG1, MG2, MT1 and MT2 which represented mammary cells from mouse 1 and mouse 2, tumor cells from mouse 1 and mouse 2 respectively.

Cluster analysis was done with R package Seurat v. 3.0.2. For each count table, cells with less than 100 expressed genes and genes present in less than 4 cells were removed. The remaining counts were further filtered to remove 1% cells with excessively large or small numbers of genes, as well as 1% cells with excessively high percentages of mitochondrial genes. Cell cycle scores were calculated from 74 G1S genes and 43 G2M genes selected according to their GO annotations. Scaling and variable regression were then performed with SCTtransform using negative binomial model to regress out the effects of percentage of mitochondrial genes and difference of cell cycle score from counts and then select the top 3000 highly variable genes for PCA. Truncated PCA was performed to estimate the top 30 principle components, of which the first N components were selected where N represented the turning point of elbow plot of standard deviations of principle components. The selected principle components were used to calculate the Euclidean distances of cells, which was then used to construct an SNN graph with k = 20, and clustering was done using standard modularity function, SLM algorithm for modularity optimization and resolution = 0.6. Clustering was visualized in t-SNE embeddings. Cluster marker genes were calculated using Wilcoxon rank-sum test contrasting each cluster against the rest. Constraints of a maximum of 500 cells in either group per contrast, only testing genes present in a minimum of 25% cells in either group and exhibiting log fold change higher than 0.25, and only including genes with positive fold changes, were applied. Genes with corrected p-value < 0.05 were retained as markers, adjusted by Bonferroni correction. A stricter marker identification was also performed with ROC method, with the same

constraints except that genes with power > 0.8 were retained.

Combined analyses with more than one individual count table were performed by merging datasets from individual analyses. Filtering and cell cycle scoring were omitted, and the rest was done as previously described.

Brca1-deficient mammary tumorigenesis was reconstructed by trajectory analysis with R package monocle v. 2.10.1. As the mammary tumors were generated from the mammary epithelial cells, we removed the libraries of non-epithelial cells, including endothelial, immune cells and fibroblasts, before performing the monocle analysis. Combined count tables (filtered MG1+MT1, and filtered MG2+MT2) with cluster identities were imported from Seurat results. Lower detection limit was set as 0.5, and negative binomial size model was specified for size and dispersion estimations. The number of cells each gene was detectably expressed was calculated, genes which were expressed in less than 10% of cells were considered unexpressed. Differential expression analysis on expressed genes was done to identify genes differentially expressed between groups, genes with q-values < 0.01 were included as ordering genes for dimensionality reduction. Dimensionality reduction was done with DDRTree algorithm and maximum components = 2. Cells were ordered along pseudo-time trajectory and were assigned states according to their positions along the trajectory. Visual inspection and adjustment ensured mammary cells were located at the beginning. State markers, group markers and cluster markers were calculated using monocle differential gene testing routine with q < 0.01.

CRISPR/Cas9 based gene knockout

For targeting murine *Mrc2* and other candidate genes, the relevant oligos (Table S1) were cloned into the lenti-CRISPR/Cas9-v2 vector (Addgene, #52961) following the Zhang lab protocol [41]. The lentiviral plasmid, envelope plasmid (pMD2.G), and packaging plasmid (psPAX2) were co-transfected into 293FT cells with PEI to produce the lentiviruses. The culture medium was collected and filtered through a 0.45-μm filter at 72 h after transfection. The viral media were then 100× concentrated via PEG precipitation, resuspended with PBS, and harvested for infection. Target cells were infected with the enriched virus for 48 hours. Positive cells were selected for 5-7 days with puromycin (4 μg/mL and 8 μg/mL for B477 and G600 cells respectively) after infection.

For validation of target modification, genomic DNA was isolated from the single clones of targeted cell lines. Following PCR amplification of murine

Mrc2 (Forward: tcttcctcatttcagccag, and Reverse: agtaaggctcgagcacatagg), the PCR products were sequenced. Allele modifications were determined by using the control cell as a reference sequence.

Cell proliferation assay and drug sensitivity test

Cell proliferation was assessed by using alamarBlue cell viability reagent (DAL1100, Invitrogen) according to the manufacturer's instruction. Briefly, 3000 cells were seeded into 96-well-plate containing 100 μl cell culture medium for 24-72 hours. The alamarBlue reagent was diluted as 1:10 with cell culture medium. Then the cells were incubated with the diluted alamarBlue reagent to replace the previous culture medium, for 3 hours in a cell culture incubator, protected from light. Then the fluorescence was measured and recorded by using a microplate fluorescence reader (SpectraMax® M5, Molecular Devices) with an excitation wavelength of 560 nm and an emission of 590 nm. All alamarBlue assays were performed at least 3 times in triplicate.

To test the sensitivity of mammary tumor cell lines to anti-cancer drugs, 3000 cells were seeded into 96-well-plate containing 100 μl cell culture medium. 24 hours later, the culture medium were changed to cisplatin (P4394, Sigma) or olaparib (AZD2281, Selleckchem) containing medium at given concentrations. Then the cell viability was tested 72 after drug treatment by using alamarBlue cell viability reagent as mentioned above.

Mammary fat pad allografts formation

Eight-week-old female nude mice used as recipients were obtained from the Animal Facility of the Faculty of Health Science, University of Macau. 2 × 10⁵ B477 or G600 cells suspended in 100 μl of 50% Matrigel (356235, Corning) plus 50% PBS were implanted into the fat pad of the nude mice. For each mouse, both flanks of the fourth pair of mammary fat pads were implanted. The mice were monitored for allografts growth every other day. And the allograft volume was calculated using the formula $V = \pi / 6 \times L \times W \times H$, where L and W were the length and width of the allografts respectively.

siRNA knockdown

siRNA oligos were designed by and purchased from Gene Pharma (Shanghai, China). The sequences of the siRNA oligos were listed in Table S2. The siRNA oligos were transfected into cells by using Lipofectamine 2000 (11668019, Invitrogen) as per the manufacturer's instruction.

Real-time PCR

Total RNA was extracted from the cells 48 after

siRNAs transfection by using TRIzol reagent according to the manufacturer's instructions (15596026, Life Technologies). Reverse transcription to cDNA was initiated with 1 µg of each RNA sample using a QuantiTect Reverse Transcription Kit (205313, Qiagen Inc.) following the manufacturer's instructions. Quantitative real-time PCR was performed using the Applied Biosystems™ QuantStudio™ 7 Flex Real-Time PCR System (ThermoFisher scientific). The reaction mixture contained 6 µl 2 × SYBR Green (04707516001, Roche), 4 µl PCR-qualified H₂O, 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), and 1 µl of cDNA from the 100 µl diluted (1:5) stock solution prepared above. Each sample was analyzed in a 384-well PCR plate. The data were analyzed initially using SPSS 13.0 software (SPSS Company) included with the PCR machine. The results were analyzed statistically and graphed using Microsoft Excel 2010. The primers used for real-time reverse transcription PCR are shown in Table S1. *Actb* was used as an internal control.

Statistics and reproducibility

Statistical analysis was performed with R software. An error bar represents mean ± s.e.m. Each experiment was repeated at least three times independently with similar results. Two-tailed Student's t-test was used to compare differences between groups. Differences in compared groups were considered statistically significantly different with *p* values lower than 0.05.

Results

Intertumor heterogeneity of BRCA1-deficient mammary tumors

Both human individuals with heredofamilial BRCA1 mutations and BRCA1-deficient mouse models stochastically develop mammary tumors [8, 42-44], indicating multiple factors might contribute to the oncogenic process and lead to great heterogeneity of the tumors. To investigate the heterogeneity of BRCA1-deficient breast cancers, we firstly referred the human breast cancer cases from public data. Among 6113 breast cancer samples from different cohorts collected in cBioPortal, 169 (2.76%) harbored the genetic alterations on BRCA1 (Figure S1A). As BRCA1 functions as a tumor suppressor, we only selected the cancer samples with loss-of-function mutations of BRCA1, including deep deletion, truncating mutation and missense mutation (putative driver) as BRCA1-deficient cancers for further analysis. Within 63 BRCA1-deficient cancers, related RNA-seq data were available from 13 samples in TCGA database (Figure S1 and Table S3). As these samples harbored

somatic mutations of *BRCA1*, for comparison we included 16 TCGA breast cancer samples with germline *BRCA1* mutations, which were identified by Inagaki-Kawata et al. [29]. Based on PAM50 calculation, these cancers are assessed as distinct molecular subtypes, with majority belonging to the basal-like subtype (Figure S1B and Table S3). We further checked expression levels of *ESR1*, *PGR* and *ERBB2*, and several marker genes for mammary basal (*KRT14*, *KRT17*) and luminal (*KRT8*, *KRT18*) lineages, epithelial (*CDH1*, *EPCAM*) and mesenchymal (*VIM*, *FN1*) cells. In line with the molecular subtyping result, different samples display distinct expression patterns of these genes, indicating great intertumor heterogeneity among both *BRCA1* germline and somatic mutant breast cancers.

To further study the heterogeneity of *BRCA1*-deficient mammary tumors, we collected 23 tumors from two previously established mouse models, where *MMTV-Cre* driven loss of *BRCA1* (*MMTV-Cre; Brca1^{fl/fl}*) specifically develops mammary tumors (BrTs) in 1-1.5 year, while combination of *Trp53* mutation (*MMTV-Cre; Brca1^{fl/fl};Trp53^{fl/+}*) could greatly accelerate the tumorigenesis (Figure 1A). Among the tumors, 10 were BrTs and the other 13 were collected from double mutant mice (termed as Br53Ts). Molecular subtyping of the tumors demonstrated that majority of the tumors are basal-like, while other subtypes also exist (Figure 1A), which is similar to human *BRCA1*-deficient cancers (Figure S1). Besides, we examined the histological features of these tumors, and found overt intertumor diversity. Some tumors are characterized with mesenchymal/fibroblast-like feature, while others with glandular structure, medullary feature or net-like feature (Figure 1B). Majority of human *BRCA1*-related breast cancers are negative for ERα, PR and/or ERBB2. So, we checked the expression levels of these receptors in histological sections of mouse tumors (Figure 1C). About 78.2% of the *BRCA1*-deficient mammary tumors do not express ERα or PR, though about 60% of the tumors showed detectable level of ERBB2 (Figure S2A, Table S4). Taken together, these results manifested that the *BRCA1*-deficient mammary tumors bear great intertumor heterogeneity.

Molecular subtyping of *BRCA1*-deficient mammary tumors

Next, we wanted to classify the heterogeneous mammary tumors with their distinct molecular characteristics. Based on the whole transcriptomes of the 23 *BRCA1*-deficient mammary tumors, four subgroups are identified via hierachal clustering, and termed as mesenchymal like, luminal like I and II (Lum I and Lum II), and mixed types respectively

because of their differential expression patterns of mammary basal/luminal lineage markers and epithelial/mesenchymal cell markers (Figure 2A-B). For example, the mesenchymal like tumors highly express mesenchymal markers *Vim* and *Fn1*, compared with other tumors. The tumors of luminal like subgroups highly express luminal markers *Krt8* and *Krt18*, while displaying low expression levels of

basal markers *Krt14* and *krt17*. Moreover, the mixed type tumors express both the basal and luminal markers. We also carefully examined the histopathological features of the mammary tumors, and found the mesenchymal like tumors display low differentiation level and typical fibroblast like morphology, while the luminal like tumors are in relatively high differentiation level (Figure S2B).

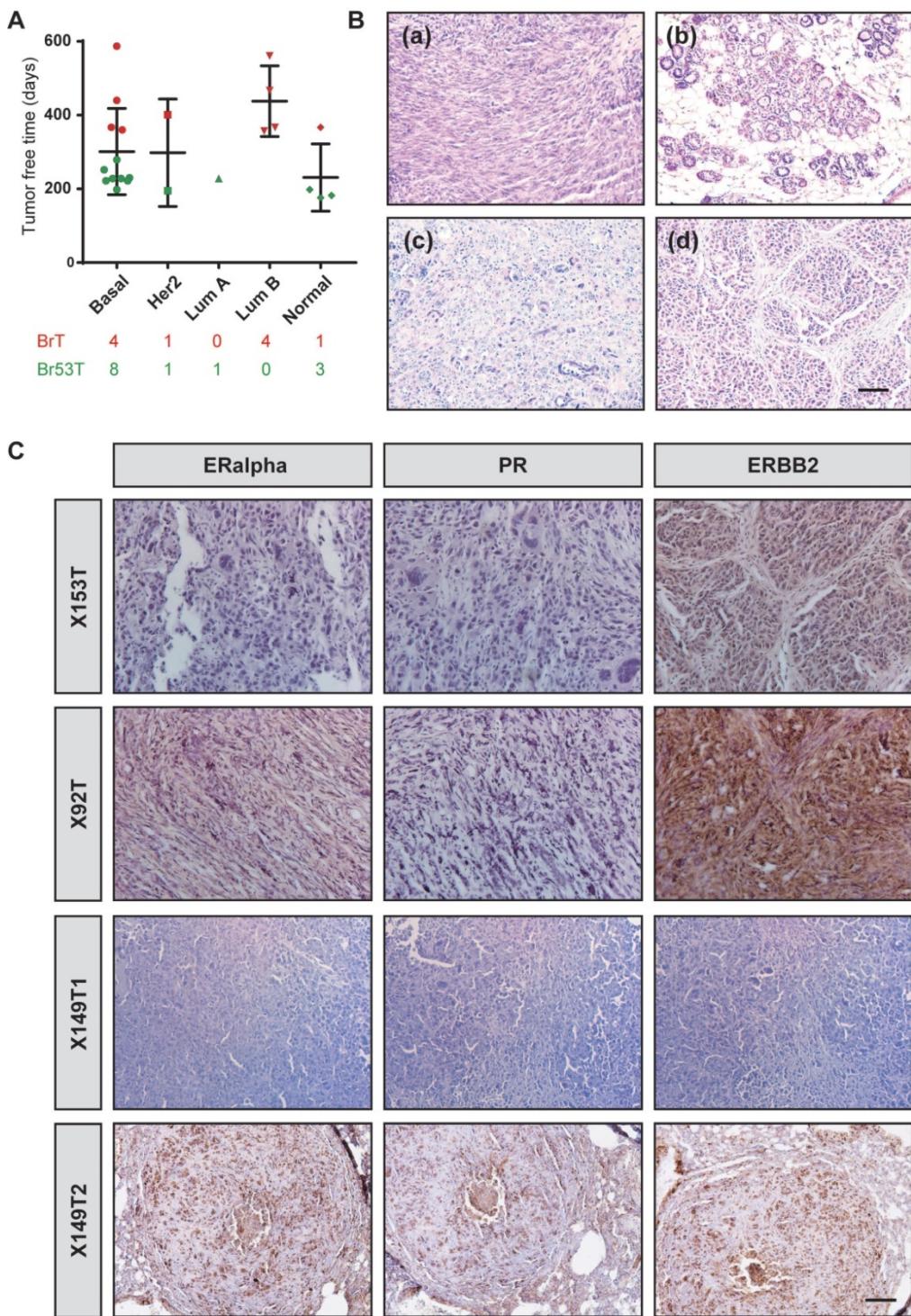


Figure 1. Intertumor heterogeneity of BRCA1-deficient mammary tumors. A. Summary of the tumor free time of 23 BRCA1-deficient mouse mammary tumors. The tumors are grouped as their PAM50 subtyping and marked according to their genotype. BrT (red), mammary tumors from MMTV-Cre; *Brca1^{fl/fl}* mice; Br53T (green), mammary tumors from MMTV-Cre; *Brca1^{fl/fl}*; *Trp53^{fl/fl}* mice. **B.** Distinct histopathological features of BRCA1-deficient mammary tumors are shown by H&E stained sections: (a), one tumor showing basal-like morphology; (b), one tumor showing luminal-like morphology; (c), one tumor showing mesenchymal-like morphology; (d), one tumor showing intermediate morphology. **C.** Immunohistochemical staining for ERalpha, PR, and ERBB2 across four tumors (X153T, X92T, X149T1, X149T2). The columns represent ERalpha, PR, and ERBB2 staining, respectively. The rows represent the four tumors. ERalpha staining shows varying degrees of nuclear staining. PR staining shows varying degrees of nuclear staining. ERBB2 staining shows varying degrees of cytoplasmic and nuclear staining, with some tumors showing strong positive staining.

with mesenchymal/fibroblast-like feature; (b), one tumor with glandular structure; (c), one tumor with medullary feature; (d), one tumor with net-like feature and typical pushing margins. **C.** Distinct expression patterns of ERAlpha, PR and ERBB2 in BRCA1-deficient mammary tumors are shown by IHC stained sections. Scale bar, 50 μ m.

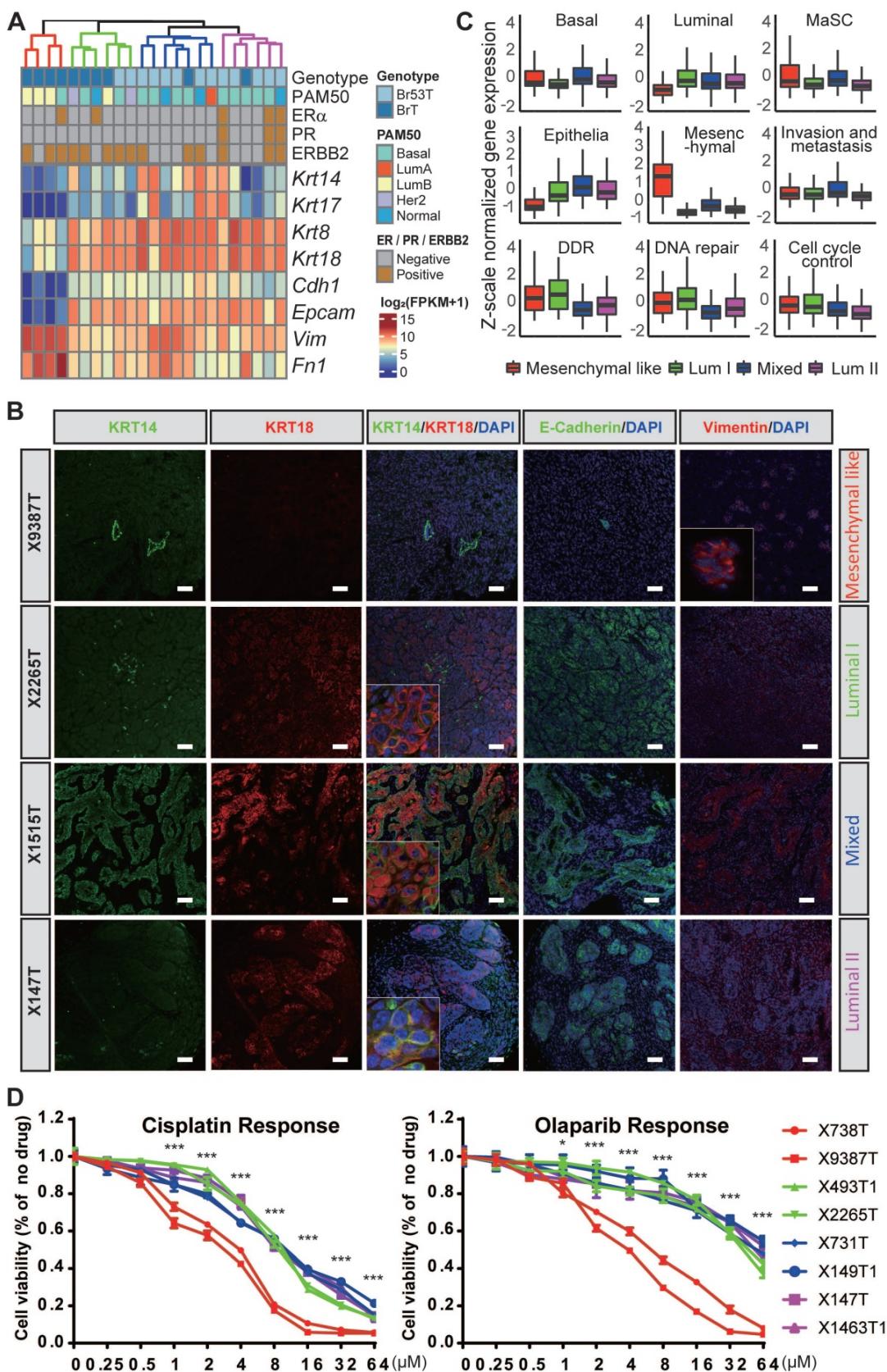


Figure 2. Molecular subtyping of BRCA1-deficient mammary tumors. **A.** 23 BRCA1-deficient mammary tumors are hierarchically clustered into four subgroups based on whole transcriptome. The genotype, intrinsic cancer subtype based on PAM50, ER α /PR/ERBB2 expression patterns based on IHC staining as well as mRNA levels of several selected genes are shown. **B.** The expression patterns of KRT14, KRT18, E-Cadherin and Vimentin in 4 BRCA1-deficient mammary tumors are shown by IHC stained sections. Scale bar, 50 μ m.

immune-fluorescence stained sections. The 4 tumors were from 4 distinct subgroups shown in panel **A**. Scale bar, 50 μ m. **C**. Boxplots representing Z-scale normalized gene expression values from 4 subgroups of BRCA1-deficient mammary tumors show expression levels of different groups of genes (Figure S3). The box represents the interquartile range and the line is the median. **D**. The sensitivities of 8 *Brcal*-mutant mouse mammary cell lines to cisplatin (left panel) and olaparib (right panel). The colors of the lines represent the subtypes of the mammary tumors where the cell lines were derived. Wilcoxon rank sum test was conducted for the analysis of drug sensitivity between mesenchymal-like tumors and the other 3 groups. ns, no significance; *, P < 0.05; **, P < 0.01, and ***, P < 0.001.

To further dissect the molecular features of the mammary tumors, we checked the relative activation levels of several key pathways involved in mammary development and tumorigenesis, and/or biological functions of BRCA1 (Figure 2C and Figure S3A). Notably, the mammary stem cell (MaSC) markers are highly expressed in the mesenchymal like and mixed type tumors, and the activity of invasion and metastasis is also higher in these tumors than the luminal like ones. Additionally, the activity of invasion and metastasis is even higher in the mixed type tumors compared with the mesenchymal like ones, indicating the tumor cells in the intermediate state of EMT possess better capacity of invasion and migration than the cells in the complete state of EMT, which is consistent with previous reports [45, 46]. We also identified marker genes for each subgroup of the mammary tumors and summarized the enriched biological functions of the markers via gene ontology (GO) analysis using DAVID [36] (Figure S3B, and Table S4).

Further, we wondered whether different subtypes of BRCA1 deficient mammary tumors bear heterogeneous sensitivity to anti-cancer drugs. To test this, we established several mouse mammary tumor cell lines and determined their sensitivities to cisplatin and olaparib. And notably, the mesenchymal-like mammary tumor derived cell lines are more sensitive to both drugs (Figure 2D), indicating these tumors may be more sensitive to the DNA damage inducers and PARP inhibitors. We further checked the activities of anticancer drug resistance related pathways, including DNA damage response (DDR), DNA repair, ATP-binding cassette (ABC) efflux, shieldin complex and cell death regulation, based on the expression levels of relevant gene sets among tumors of different subtypes. The results indicated none of these pathways were associated with the high sensitivity of mesenchymal like tumors to cisplatin and olaparib (Figure S4). In addition, previous study reported that the decelerated proliferation and cell dormancy contribute to the multidrug resistance of tumor cells [47]. Indeed, we found the mesenchymal like tumor cells grow much faster than other subtypes (Figure S4G), which might result in their high sensitivity to the drugs.

Taken together, the BRCA1 deficient mammary tumors could be classified into four distinct subtypes, with respective markers genes, molecular features

and anti-cancer drug sensitivities.

Single cell analysis of BRCA1-deficient mammary tumor cells and luminal cells

The observation of great intertumor heterogeneity among BRCA1-deficient mammary tumors propelled us to posit whether such phenomenon is ascribed to the intratumor heterogeneity among tumor cells. To address this, we resorted to the Fluidigm C1 platform based scRNA-seq to decipher the composition of BRCA1-deficient mammary tumor cells, as well as the mammary luminal cells, in which loss of BRCA1 initiates the tumorigenesis (Figure S5, and Table S5). Totally, we performed scRNA-seq on 165 tumor cells from four mammary tumors of distinct molecular subtypes (termed as BrT1, BrT2, Br53T1, and Br53T2; Table S5), as well as 312 luminal cells from female mice with three different genotypes (*Brcal* mutant; *Trp53* mutant; and *Brcal/Trp53* double mutant, Table S5). To study whether pregnancy affects the gene expression patterns of mammary cells with different genotypes, we selected luminal cells from both 3-month-old virgin mice and age-matched pregnant mice at day 12.5 (P12). We firstly applied the principle component analysis (PCA) on all the mammary luminal cells and tumor cells with the Seurat R package [48], and identified 9 main clusters (Figure 3A). When projected into the t-distributed Stochastic Neighbor Embedding (t-SNE) plot, we found that the luminal cells with different genotypes are roughly clustered together, indicating there is little inter-individual heterogeneity (Figure 3A). However, the tumor cells appear overt intertumor heterogeneity (Figure 3A-B). Meanwhile, regardless of genotypes, the luminal cells from virgin mice are mostly grouped in C2 and C9, while the luminal cells from P12 mice are mainly grouped in C1, C3 and C7, which indicates the pregnancy vastly influences the gene expression profiles of luminal cells and overrides the effect of loss of BRCA1 and/or TRP53.

We also calculated chromosomal copy number variations (CNVs) in each cell using InferCNV [49] (Figure 3C). And a great many of CNVs are observed across almost all the chromosomes in the tumor cells. Besides, we noticed a few luminal cells also harbor some CNVs, especially on chromosomes 10, 11 and 12. These cells are perhaps at an early stage of oncogenic transformation.

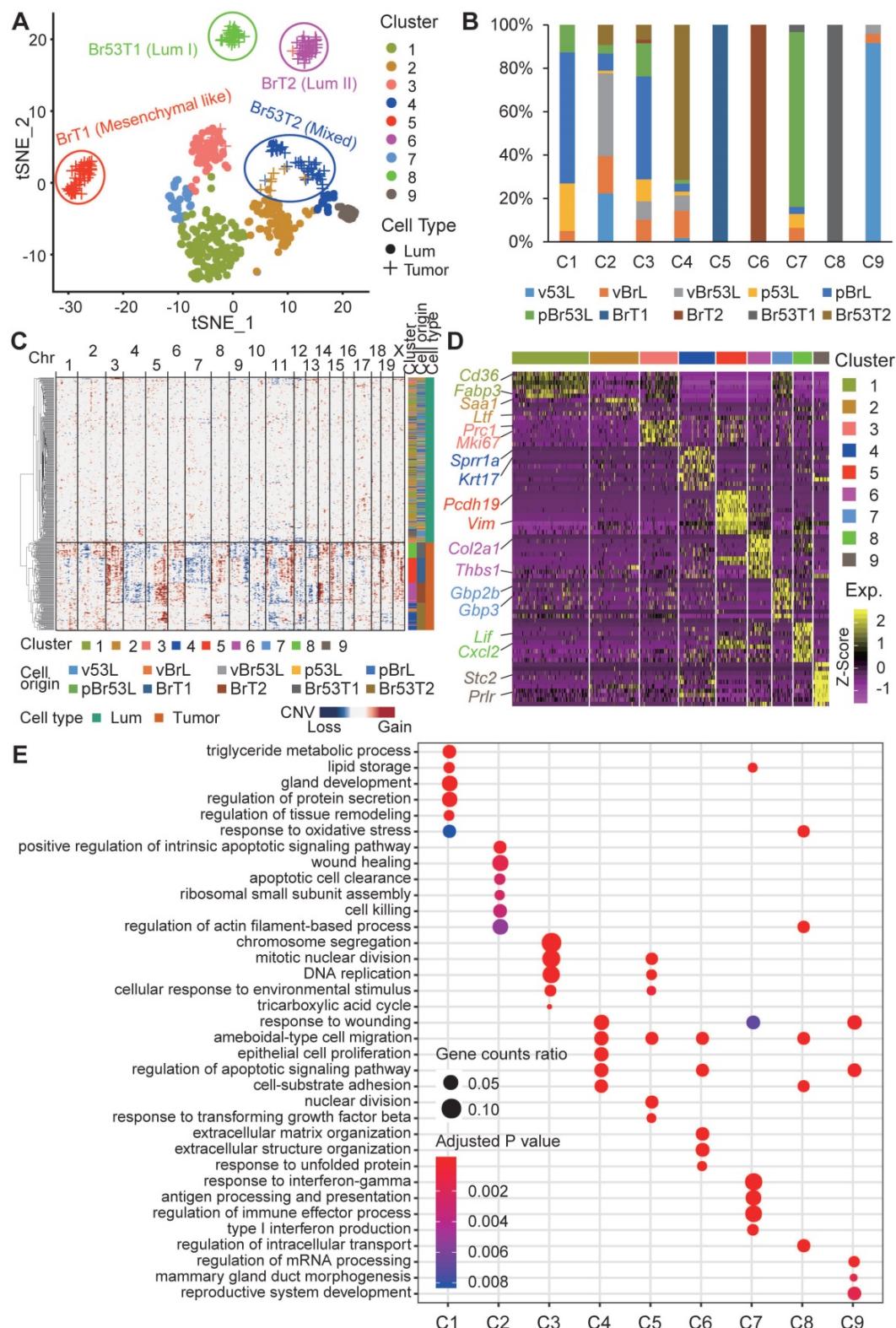


Figure 3. Single cell RNA sequencing of sorted BRCA1-deficient mammary luminal and tumor cells. **A.** tSNE plot displays the clustering of single BRCA1-deficient mammary luminal cells and tumor cells. The cells are clustered into 9 groups based on Seurat analysis and respectively labeled accordingly. **B.** Summary of the cells composition for each cluster. **C.** Heatmap shows the normalized CNV levels of single luminal and tumor cells. The red and blue colors represent copy number gain and loss, respectively. **D.** Heatmap shows the expression levels of marker genes for each cluster (Table S6). Some representative markers are highlighted on the left. **E.** Bubble diagram shows the representatives of enriched GO terms for each cluster.

To further elucidate the molecular variations among the single cells, we checked the marker genes for each cluster (Figure 3D, and Table S6). We found

that lipid and fat metabolism related genes like *Cd36* and *Fabp3* are highly expressed in C1, and also moderately expressed in C3 and C7, which is in

accordance with the pregnant situation of donor mice. Almost all the BrT1 cells (C5) highly express the mesenchymal marker *Vim*, which is in line with the bulk RNA-seq result (Figure 2A). Interestingly, *Prlr* is specifically expressed in C9 cells, most of which are luminal cells from virgin *Trp53* mutant mice, indicating both the pregnancy (P12) and deficiency of BRCA1 may block the expression of *Prlr*. Next, we examined the enriched biological functions of the cluster-specific marker genes (Figure 3E). Notably, tumor cell related clusters (C4, C5, C6 and C8) all enrich the GO term of ameboidal-type cell migration, indicating it might be a common event that tumor cells gain the capacity of ameboidal migration after oncogenic transformation.

To acquire more details of differences among the mammary luminal cells, we performed the clustering analysis only on the luminal cells (Figure S6A). These cells are divided into 5 subgroups (Lum1-5). While subgroups Lum1, Lum3 and Lum5 are mainly composed of luminal cells from P12 mice, luminal cell from virgin mice are mostly classified into Lum2 and Lum4. Every subgroup includes cells with three distinct genotypes, only except Lum5, which is dominated by luminal cells from P12 *Brca1/Trp53* double mutant mice. Marker genes and related GO analysis of each subgroup reveal that genes involved in immune system process like *Cxcl9*, *Cxcl10*, *Ifi47* are highly expressed in Lum5 cells (Figure S6B-C). This indicates loss of BRCA1 and TRP53 might induce immune response in some luminal cells.

Then, we performed the similar analysis on all the mammary tumor cells (Figure S6D-F). And we noticed that the BrT1 cells are divided into two groups, Tum4 and Tum5. Though both groups share some marker genes, Tum5 cells are in a more proliferative state, which is indicated by the induced expression of cell cycle related genes *Top2a*, *Cdk1* and so like (Figure S6D-F). Besides, we found Tum1 and Tum2 are clustered closely on the tSNE plot (Figure S6D). While all the Br53T1 cells together with a few Br53T2 cells are clustered as Tum2, the rest Br53T2 cells are classified into Tum1, which indicates there might be some similarity of transcription regulation among *Brca1/Trp53* double mutant mammary tumors. And GO analysis result of marker genes for both clusters manifests that biological function like response to virus and immune system process is significantly enriched in both clusters (Figure S6F), which has also been identified in part of double mutant luminal cells (Figure S6B-C).

Intratumor heterogeneity of BRCA1-deficient mammary tumors

As the intertumor heterogeneity was much

higher than the intratumor one, we next hierarchically clustered the single tumor cells separately according to their tumor identity via another R program M3drop [39]. As shown in Figure 4A, the single cells within each tumor could be further divided into three subgroups, with individual marker genes, which enrich distinct biological functions (Figure S7, and Table S7). Notably, the GO terms enriched by the marker genes for subgroups demonstrate that the main diversity among three subgroups within individual tumor are concerned with cell cycle control, DNA repair, cell migration, apoptosis, and mammary gland development (Figure 4B, and Figure S7). Next, in each individual tumor cell, we checked the molecular subtyping via calculation of PAM50 (Figure 4C). In line with previous studies [50, 51], the result showed that various subtypes of breast cancer could be identified on the single cell level, though the majority are normal-like, which accounts for about 46% (76/165) of all the tumor cells (Table S5). In addition, we evaluated the activation levels of biological hallmark pathways via GSVA analysis [52] (Figure 4C). Most of the pathways are activated in part of the single tumor cells with different levels. Especially, cell proliferation related pathways like E2F_targets and MYC_targets are highly activated in some of the tumor cells within all four tumors. And immune response related pathways such as TNF α _signaling via NFKB, TGF β _signaling, IL6_JAK_STAT5_signaling also display high activity, especially in the cells with low proliferation (Figure 4C). In sum, there is great intratumor heterogeneity among single cells based on analysis of the gene expression patterns, PAM50 subtyping and the activity of biological hallmark pathways.

Dropseq to reveal molecular changes during BRCA1-deficient mammary tumorigenesis

After analysis of inter- and intra-tumor heterogeneity of the BRCA1-deficient mammary tumors, we wanted to further explore how the mutant mammary cells transform into tumor cells. Here, we applied dropseq to analyze pair-matched mammary gland cells and mammary tumor cells from two *Brca1* mutant mice (termed as MT1 and MT2 hereafter). Totally, 21157 cells including 3153 mammary cells and 18004 mammary tumor cells with high-quality sequencing data were analyzed and classified into 15 main clusters (Figure 5A). Then, we identified the cell type of each cluster according to the cell origin and marker genes (Figure 5A-C, and Table S8). The representative marker genes of each cluster are shown in Figure 5D-E (Table S9).

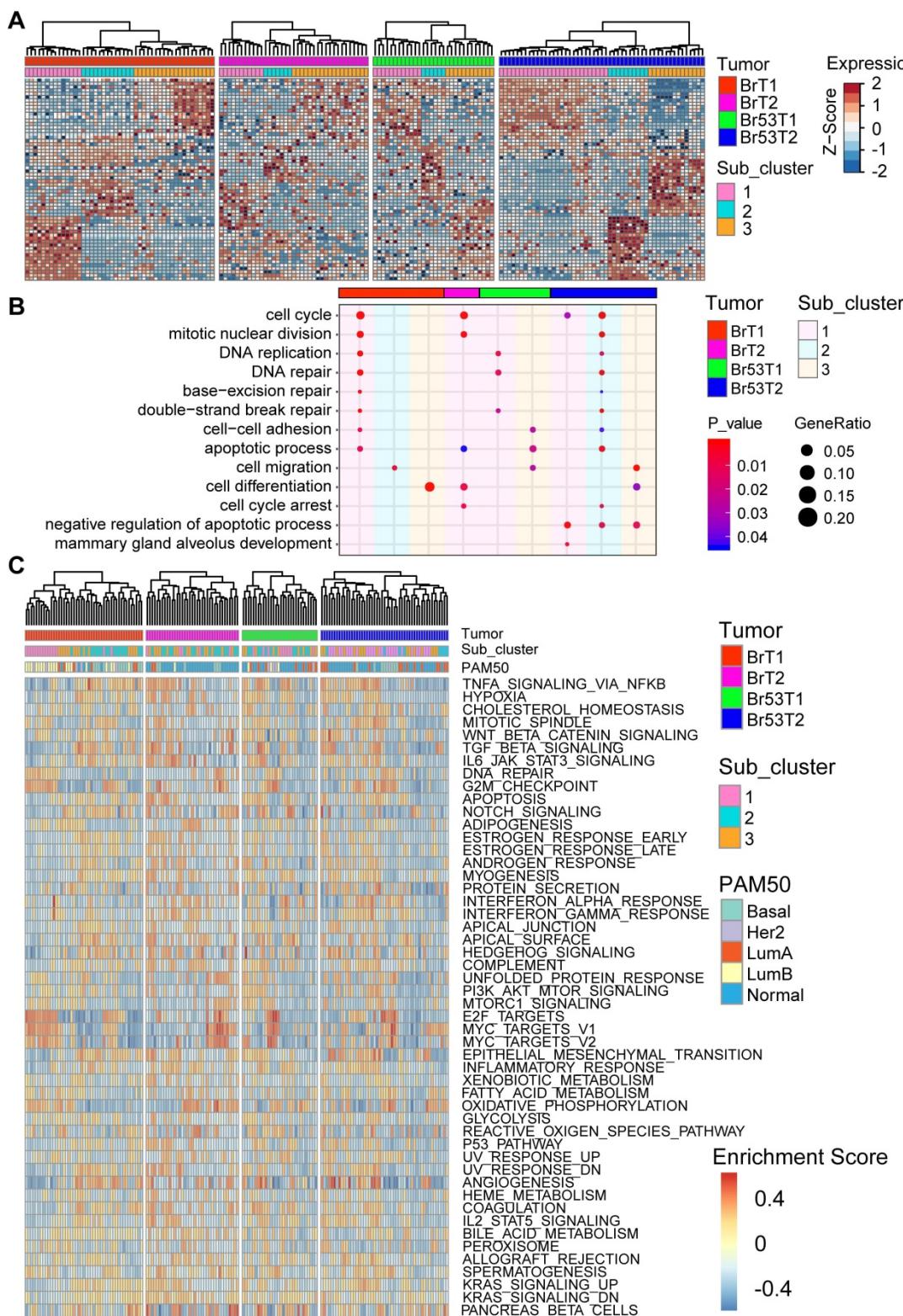


Figure 4. Intratumor heterogeneity of BRCA1-deficient mammary tumors revealed by single cell RNA sequencing. **A.** Heatmap shows the expression patterns of marker genes for each sub-cluster within individual mammary tumor. Single cells of each tumor are divided into three sub-clusters based on M3Drop analysis (Table S7). **B.** Bubble diagram shows the representatives of enriched GO terms for each sub-cluster of individual tumor. **C.** Heatmap shows the heterogeneous activation of biological hallmarks pathways among single mammary tumor cells. The intrinsic cancer subtype of the single tumor cells based on PAM50 scaling is shown (above) as well.

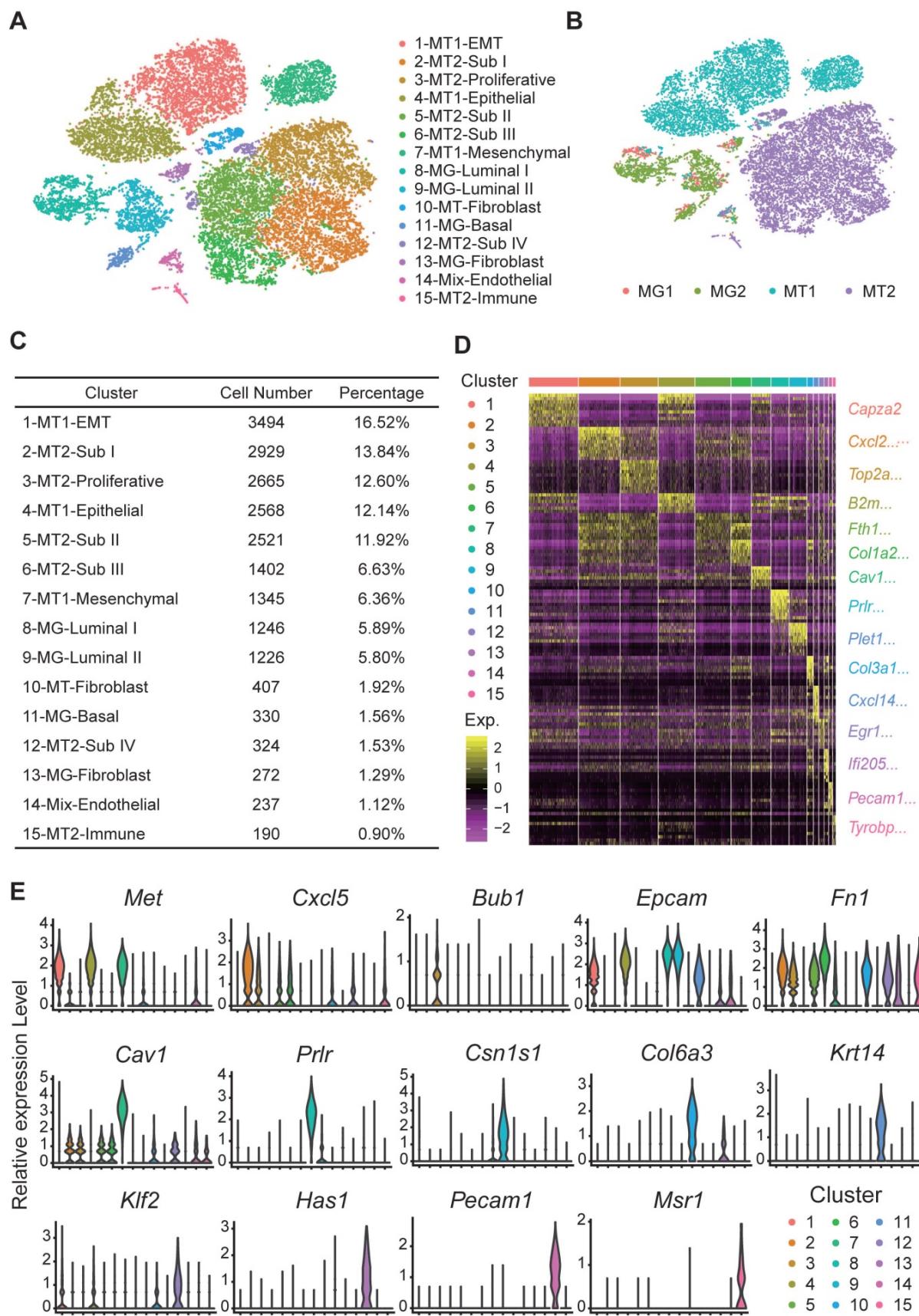


Figure 5. Dropseq of paired mammary gland (MG) cells and mammary tumor (MT) cells from 2 *Brca1* mutant mice. **A** and **B**. tSNE plots demonstrate the cell types and groups of the MG and MT cells (Table S8). The cells are divided into 15 groups, which are separately clustered and respectively labeled with different colors in panel **A**. The cells origin is shown in panel **B**. **C**. Summary of cell number and percentage of each group of cells. **D**. Heatmap displays the expression patterns of marker genes for each group of cells (Table S9). **E**. Violin plots show the expression patterns of representative markers across the cell groups. The y-axis indicates the normalized read count.

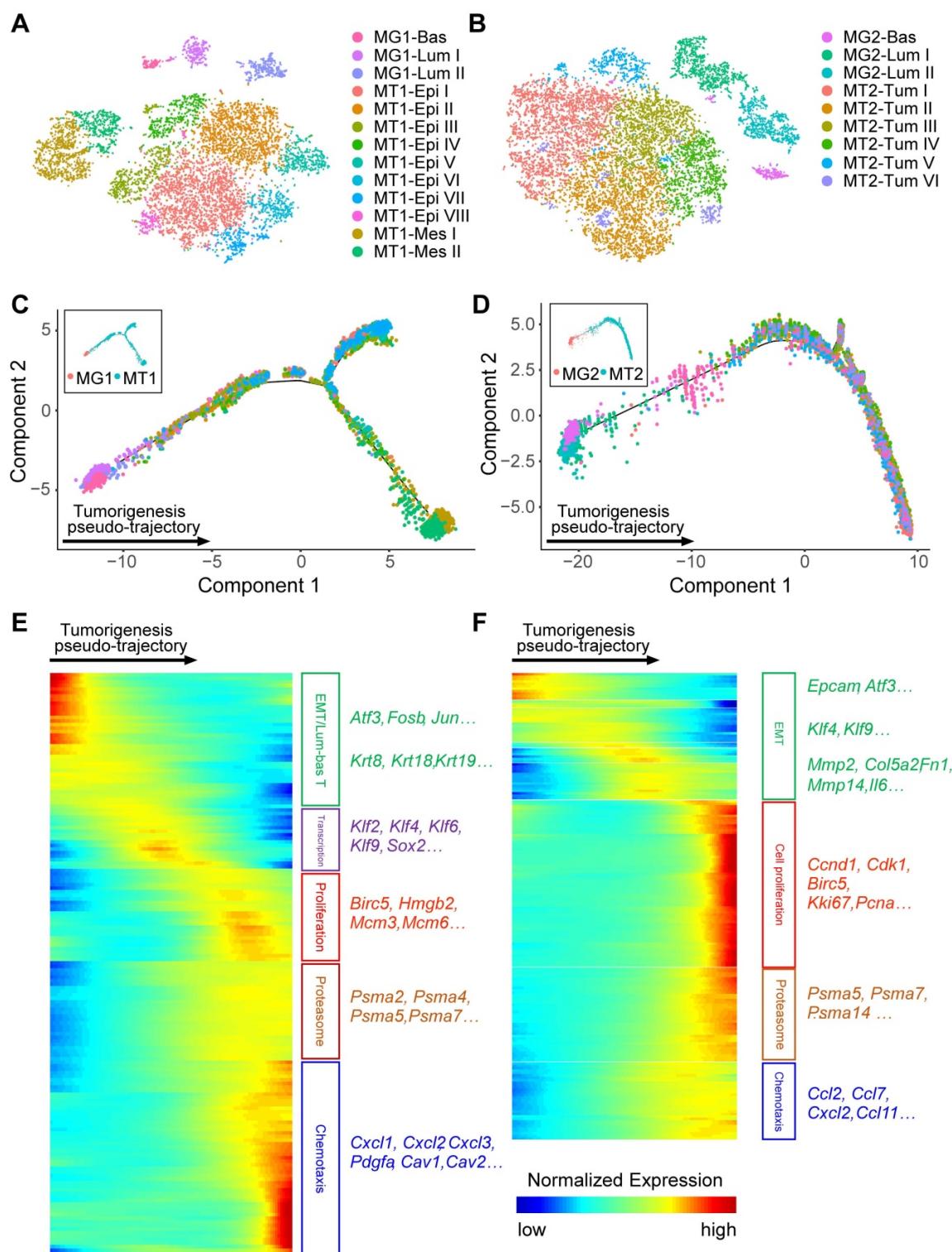


Figure 6. Molecular changes during BRCA1-deficiency induced tumorigenesis. **A** and **B**. tSNE plots demonstrate the cell types and groups of the filtered MG and MT cells from two individual mice. The single MG and MT cells from mouse 1 and 2 are divided into 13 (**A**) and 9 sub-groups (**B**) respectively. **C** and **D**. Monocle analysis reveals the pseudo-temporal trajectories of tumorigenesis in mouse 1 (**C**) and 2 (**D**). **E** and **F**. Heatmaps show the expression patterns of differentially expressed genes (rows) along the pseudo-time (columns) of tumorigenesis (left panels, Table S10). The enriched biological functions and representative markers for gene clusters with similar expression pattern are shown in the right panels.

To uncover the molecular changes during BRCA1-deficiency induced tumorigenesis, we separated the single cells (only mammary epithelial and tumor cells) as per their mouse origin because the tumorigenesis process may be different among

individuals. Among the single cells from mouse 1, 13 subgroups including 1 MG basal, 2 MG luminal, 8 epithelial MT and 2 mesenchymal MT clusters are identified (Figure 6A). Similarly, 1 MG basal, 2 MG luminal, and 6 MT subgroups are identified in mouse

2 (Figure 6B). Then, we performed the monocle analysis to reconstruct the pseudo-temporal trajectories of the tumorigenesis [53] (Figure 6C-D). And three continuous states along the pseudo-temporal trajectories are classified (Figure S8A-B, and Table S10). Next, we further examined the differentially expressed genes (DEGs) with distinct variation tendencies along the pseudo-temporal trajectories (Figure 6E-F). Notably, we found a consistent EMT process during tumorigenesis in two mice, with expression of epithelial markers like *Epcam* and *Atf3* are gradually reduced. And at the end stage of tumorigenesis, genes related to cell proliferation, proteasome and chemotaxis are highly expressed in tumor cells of both mice.

Besides, we further checked some important signaling pathways related to mammary development and tumor initiation and progression, to determine whether their activities are changed during the tumorigenesis process (Figure S8C). We noticed that expression of the epithelial markers was instantly shut down at the mammary cells-tumor cells joint along the pseudo-trajectory, which might be the oncogenic transformation turning point, while the mesenchymal markers are gently induced across this turning point (Figure S8C). Interestingly, expression of markers for basal cells and MaSCs shows a down-and-up tendency, with the lowest activity at the transformation turning point. Meanwhile, expression of markers for luminal cells is greatly reduced during the tumor initiation stage. All the results indicated the mammary cells would shut down the lineage specific genes and loss their transcription control of intrinsic features once the tumorigenesis initiates. However, the expression of basal and MaSCs marker genes may be required for the survival and/or expansion of the tumor cells, and so gradually induced in the late stages.

In addition, signaling pathways regarding cell cycle control, DNA damage response, survival-cell death regulation, invasion and migration, PI3K-AKT, Notch, and MAPK signaling all appear the similar down-and-up variation tendency (Figure S8C), indicating the overall transcription activity might be reduced to a very low level at the oncogenic transformation turning point. Taken together, our analysis of dropseq data provides a panorama of the molecular changes during BRCA1 deficiency induced mammary tumorigenesis process.

Identification of DEGs between BRCA1-deficient luminal and tumor cells

In our mouse models, the *Brca1* mutation in luminal cells results in mammary tumors formation [10, 54]. So, we wondered whether any of the DEGs

between *Brca1* mutant luminal and tumor cells could serve as driver and/or marker for the tumorigenesis. Therefore, we compared the whole transcriptomes of luminal cells and tumors in single cell and bulk RNA sequencing data (Figure S9A-B, and Table S11). We found more than half of the DEGs discovered in bulk RNA-seq are overlapped with the DEGs identified in scRNA-seq (Figure S9C). Next we checked the biological functions of these common DEGs via GO analysis, and found fatty acid and lipid metabolism related pathways are significantly enriched in the luminal cells while cell adhesion, migration and immune response etc. are activated in the tumor cells (Figure S9D). Then, we ranked the DEGs by the fold-change of expression levels, and compared the top 100 DEGs for luminal and tumor cells separately, in both single cell and bulk RNA-seq data. We found 41 and 25 common genes are highly expressed in luminal cells and tumor cells, respectively (Figure S9E-G). Notably, 5 collagen-encoding genes show up as tumor DEGs. We just picked two of them as well as other 15 tumor DEGs for further functional validation. Meanwhile we chose 10 luminal DEGs for the parallel validation. Some tumor DEGs were excluded for further investigation because their functions in mammary tumorigenesis have been well studied, such as *H19* [55], *S100a4* [56, 57], *Rspo1* [58, 59], etc.

Accelerated proliferation is a very significant and general hallmark of tumor cells. To test whether the candidate genes are involved in the regulation of cell proliferation, we knocked out them separately via CRISPR/Cas9 system in both *Brca1* wild type *Trp53* mutant B477 mammary epithelial cells and *Brca1* mutant G600 mammary epithelial cells [20]. Compared with the non-targeting control cells, proliferation level is altered after knockout of several genes, among which loss of *Mrc2* could reduce the cell growth in both lines (Figure 7A-B).

***Mrc2* serves as an oncogene in mammary tumorigenesis**

As knockout of *Mrc2* inhibits cell growth in both B477 and G600 cells, we wanted to further study the function of MRC2 in mammary tumors. Previous study reported that MRC2 is mainly expressed in the stromal cells but not the epithelial cells of mammary glands [60] and only expressed in some breast tumor cells [61]. By carefully checking the expression patterns of MRC2 in histological sections of *Brca1* mutant mouse mammary glands and tumors, we found MRC2 is expressed at low level in the basal cells and highly expressed in the tumor cells (Figure S10A). Bulk RNA-seq data confirmed such expression patterns (Figure S10B). We further checked the expression level of *Mrc2* in the 4 subtypes of

BRCA1-deficient mouse mammary tumors, and found *Mrc2* was highly expressed in the mesenchymal-like tumors when compared to other 3 subtypes (Figure S10C). We also found the expression of MRC2 could be detected in some normal and cancer tissues of human breast samples collected in public database,

the human protein atlas (Figure S10D). And the overall protein level of MRC2 is significantly higher in the breast cancers compared with normal tissues among the human samples collected in the Clinical Proteomic Tumor Analysis Consortium (CPTAC) dataset [32] (Figure S10E).

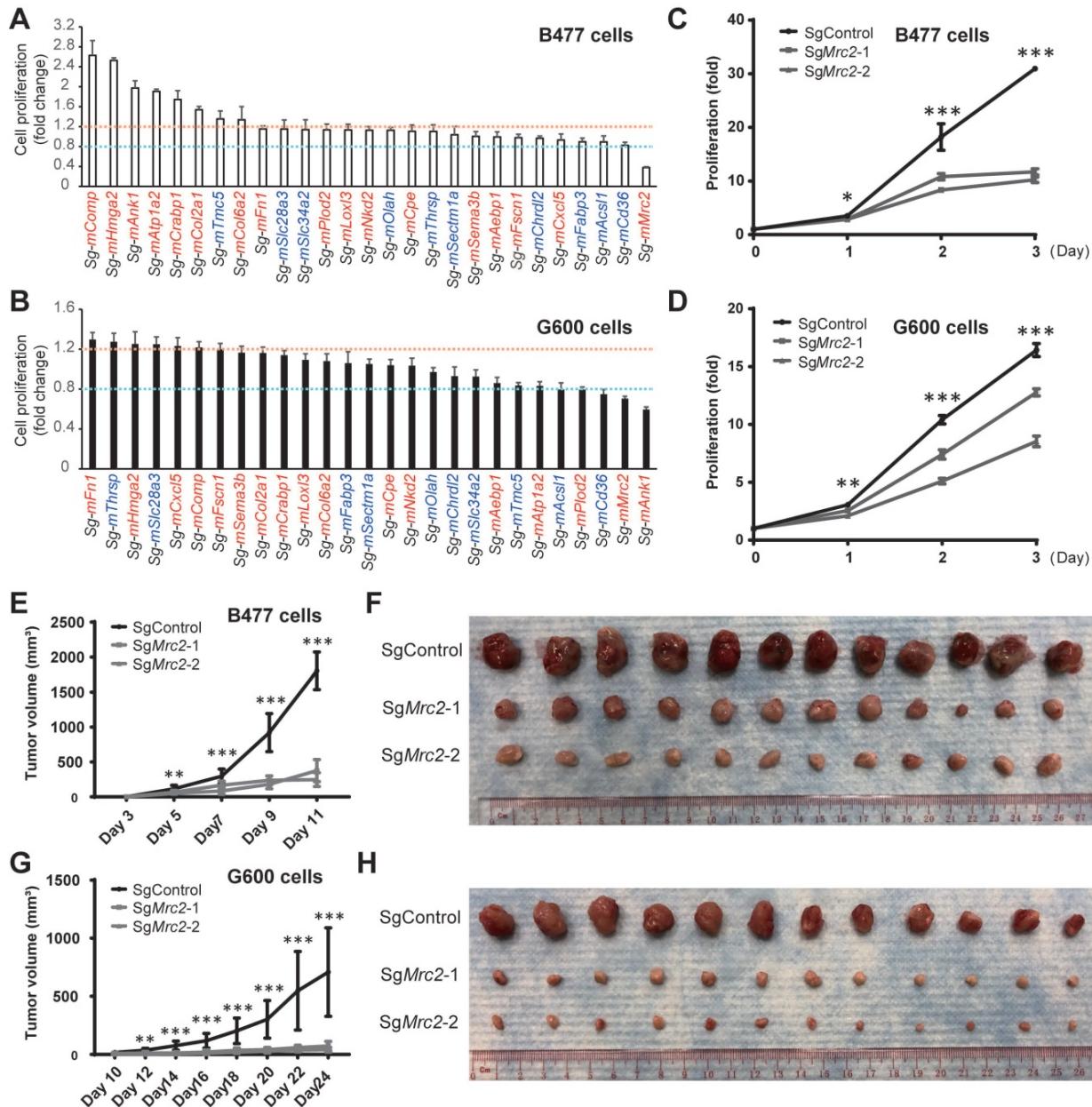


Figure 7. *Mrc2* serves as an oncogene in mammary tumorigenesis. **A** and **B**, Cell proliferation screening of candidate driver genes identifies a pro-proliferation role of MRC2 in mammary tumor cells. 27 candidate genes were selected and knocked out individually in B477 cells (**A**) and G600 cells (**B**). The genes with red color or blue color indicate their up-regulation or down-regulation in the mammary tumor cells compared with luminal cells (Table S11). The cell proliferation was measured with alamarBlue assay 72 hours after cells were seeded. The cells with non-target sgRNA were used as control for cell proliferation fold change calculation. The fold change > 1.2 or < 0.8 was used as cutoff for cell proliferation induction (orange line) or reduction (cyan line), respectively. **C** and **D**, Cell growth curves indicate the proliferation was significantly inhibited after *Mrc2* knockout. Two different sgRNAs were used to target *Mrc2* in B477 cells (**C**) and G600 cells (**D**). **E-H**, Knockout of *Mrc2* blocked tumor cells growth *in vivo*. The tumor growth of implanted *Mrc2* knockout B477 (**E** and **F**) or G600 (**G** and **H**) cells was significantly blocked in nude mice. The tumor volume changes (**E** and **G**) and photos of allografts (**F** and **H**) are shown. *, p value < 0.05 ; **, p value < 0.01 ; ***, p value < 0.001 .

In addition, consistent with the basal-like feature of *Brca1* mutant cancers, G600 cells adapt morphology more like mesenchymal feature compared with *Brca1* wild type B477 cells, which display epithelial morphology (Figure S11B). Notably, sg*Mrc2*-G600 cells undergo a morphological conversion from mesenchymal to epithelial transition (MET), as the sg*Mrc2*-G600 cells appear epithelial feature and get attached to each other (Figure S11B). In contrast, sg*Mrc2*-B477 cells exhibit very little morphological change (Figure S11B), despite reduced proliferation has been observed. Moreover, knockout of *Mrc2* with another targeting sgRNA also reduces the proliferation of B477 and G600 cells (Figure 7C-D). And knockdown of *Mrc2* with two different siRNAs confirmed the essential role of *Mrc2* in proliferation of G600 cells (Figure S12A-B). Next, we examined whether knockout of *Mrc2* could also block the tumor cell growth *in vivo* by implanting the cells into the fat pads of nude mice. The results demonstrated that tumor formation capacity of both B477 and G600 cells is obviously inhibited in the knockout groups compared with the control cells, especially in the *Brca1* mutant G600 cells (Figure 7E-H, and Figure S12C and D). In addition, we found knockdown of *Mrc2* in cell lines derived from *Brca1* mutant tumors significantly blocked cell growth and reduced the mRNA levels of proliferation markers *Ccnd1* and *Mki67* in regardless of their subtypes (Figure S12E-L). Besides, knockdown of *MRC2* in human breast cancer cell lines: MCF7, MDA-MB-231, and MDA-MB-436 could also decelerate the cell proliferation, and reduce the mRNA levels of proliferation markers *CCND1* and *MKI67* (Figure S12M-R). Taken together, *Mrc2* may serve as an oncogene in the mammary tumorigenesis due to its essential role in controlling cell proliferation.

Discussion

In this study, we explored the inter- and intra-tumor heterogeneity of BRCA1 deficient mouse mammary tumors. On the intertumor level, we uncovered great heterogeneity of histopathological features, expression patterns of ER/PR/ERBB2, molecular subtyping, and transcriptome profiles among individual tumors, and identified 4 subgroups with distinct molecular characteristics and marker genes, and also different sensitivities to anti-tumor drugs. On the intratumor level, we revealed the main gene expression variations and the differences of the enriched biological functions among subgroups of cells within individual tumors.

As the first identified breast cancer associated gene, *BRCA1* is closely associated with mammary gland development and tumorigenesis. Although the

mutation frequency of *BRCA1* is not high among breast cancer patients, women with germline mutation of *BRCA1* would reach a breast cancer risk as high as 50%-80% during their lifetime [62]. More importantly, *BRCA1* mutations are always linked to the refractory type of breast cancers, termly the basal like breast cancers, always also TNBCs [2]. In this study, we found near 80% of the *BRCA1*-deficient mouse mammary tumors are ER and PR double negative, while 60% of the tumors show detectable expression level of ERBB2 (Figure S2A), which is inconsistent with the human *BRCA1* related breast cancers [16]. One *BRCA1* associated genetic alteration 17q23 gain might lead to the amplification of *Erbb2* [63]. As the 17q23 gain identified in *BRCA1* related human breast cancers, which is far away from the *ERBB2* and *BRCA1* loci on human chromosome 17, is quite near the *Erbb2* and *Brca1* loci on mouse chromosome 11 [64]. More studies are required to pinpoint whether the expression of ERBB2 plays a driver or passenger role in the *BRCA1*-deficient mouse mammary tumors.

Another genetic alteration frequently occurred in *BRCA1* related breast cancer is the *TP53* loss, which is also observed in the *BRCA1*-deficient mouse mammary tumors [10]. Besides, synchronous one allele loss of *Trp53* not only greatly accelerates the *BRCA1* deficiency induced mammary tumorigenesis (Figure 1A), but also affects the molecular features of the tumors. We noticed that all the mesenchymal like mammary tumors in our cohort were generated from the *MMTV-Cre; Brca1^{fl/fl};Trp53^{+/+}* mice (Figure 2A), which is contradictory to previous reported role of TRP53 in EMT, where loss of TRP53 could activate EMT through regulation of microRNAs-ZEB1/2, and MEK-ERK signaling, as well as other mechanisms [65-67]. In consideration of the high mutation frequency of *Trp53* in *BRCA1* deficient mammary tumors, and the much longer period for formation of Br53Ts than BrTs (Figure 1A), the mesenchymal like BrTs may generate and accumulate *Trp53* mutation during the tumorigenesis process. This raises the hypothesis that the timing of TRP53 loss may determine the subtype of *BRCA1*-deficient mouse mammary tumors. Actually, timing of TRP53 loss has been reported to affect the subtype of mammary tumors induced by *Rb* deficiency and *Pten* deficiency [68, 69]. In addition, synergetic loss of *BRCA1* and TRP53 may greatly induce the immune response of the mammary luminal cells and tumor cells, as compared with the *Brca1* mutant cells, the *Brca1/Trp53* double mutant cells, no matter luminal or tumor cells, display higher activity of immune response related pathways (Figure S6). Hence, more studies are still needed to figure out the functions and mechanisms of

TRP53 on BRCA1 deficiency induced mammary tumorigenesis.

The BRCA1 deficiency induced mammary tumorigenesis is a continuous and dynamic process. During this process, lots of stochastic mutations and variable levels of DNA damage induced stress could be accumulated in individual cells, which might greatly affect the gene expression patterns among the cells, and lead to the intratumor heterogeneity. Meanwhile, the intratumor heterogeneity could also result from the intrinsic cell cycle differences (different cells at different cell cycle stages) and mammary hierarchical variations (different cells at different differentiation stages and mammary lineages). The above heterogeneous features could be also observed on the intertumor level. In addition, a few recently published studies provided more information about the inter- and intra-heterogeneity of BRCA1-deficient mammary tumors/breast cancers. For example, Wang, et al. (2019) performed scRNA-seq on adenovirus mediated *Brcal* knockout mammary luminal and tumor cells, and observed the continuously enhanced EMT along the tumorigenesis process [70]. Hu, et al. (2021) studied the origin and evolution of breast cancers in human BRCA1 mutation carriers [4]. They conducted scRNA-seq on breast cancer tissues and adjacent or prophylactic normal breast tissues from four BRCA1 mutation carriers. Though a continuous transcriptome change was also found from normal mammary cells to tumor cells in this study, while EMT was only overt in tumorigenesis of the basal-like/ER negative tumors but not obvious in the ER high tumors. Another study from Pal, et al. (2021) performed scRNA-seq on normal, preneoplastic and tumorigenic cells of human breast, including the BRCA1 mutation carriers [71]. For BRCA1-related tumorigenesis, the authors mainly focused on the immune microenvironment, which showed marked changes. Indeed, besides of the heterogeneous tumor cells, other types of cells especially the immune cells within the tumor microenvironment should also affect the intertumor heterogeneity. Further studies are required to uncover more details about the heterogeneity of the non-tumoral cells and their functions in BRCA1 related mammary tumorigenesis.

In this study, we identified and validated *Mrc2* could serve as a candidate oncogene for mammary tumorigenesis. MRC2 (also termed as CD280, ENDO180, urokinase-type plasminogen activator receptor-associated protein), encoded by *Mrc2*, is an endocytic mannose receptor. In normal mammary gland, expression of MRC2 is predominately restricted in the stromal fibroblasts, but rarely detected in the epithelia [60]. But interestingly, Dirk

Wienke and colleagues reported MRC2 is expressed in a small set of breast cancers, which display typical basal-like, triple negative and invasive features [61]. However, in the current study we found relatively high level of *Mrc2* could also be detected in the ER⁺ positive tumors (Table S4), and expression level of *Mrc2* is not correlated with that of *Esr1*, *Pgr* or *Erbb2* (data not shown). Notably, expression of *Mrc2* is negatively correlated with that of several epithelial markers *Cd24a*, *Cdh1*, *Epcam* and luminal markers *Krt8* and *Krt18*; meanwhile is positively correlated with that of mesenchymal/fibroblastic markers *Fn1*, *Col1a2*, *Col6a1*, *Pdgfra* and so like, in both the bulk and single cell RNA-seq data (Figure S13A). Such finding indicates the up-regulated expression of *Mrc2* may be closely associated with EMT during the tumorigenesis process. Our further gene set enrichment analysis (GSEA) of co-regulated genes of *Mrc2* indicated MRC2 may be involved in the regulation of cell cycle, PI3K-Akt/Ras/Rap1/Calcium/cGMP-PKG signaling pathways, focal adhesion, gap junction, ECM organization and other biological functions (Figure S13B-F). In our study, we found high expression of MRC2 is linked to deficiency of BRCA1, though not directly regulated by BRCA1 because transient overexpression or knockdown of BRCA1 has little effect on the expression of MRC2 (data not shown). So the increased expression of MRC2 might result from accumulated DNA damage, EMT, or other mechanisms related to BRCA1 deficiency, which needs further study to clarify.

Apart from *Mrc2*, we identified some other DEGs between the BRCA1 deficient luminal cells and tumor cells. The differential expression of these genes may promote the tumorigenesis, or functions as feedback to attempt to suppress the tumor formation. Meanwhile, we noticed that knockout of some of the DEGs did not change the proliferation of *Brcal* wild type or mutant mammary cells (Figure 7A-B). As a versatile protein, BRCA1 plays multiple roles in mammary development and tumorigenesis. So, it is interesting to further investigate the crosstalk between BRCA1 and the DEGs, and to figure out the functions and mechanisms of the DEGs in BRCA1 related mammary tumorigenesis.

In summary, our current study deciphered the inter- and intra-tumor heterogeneity of BRCA1 deficient mouse mammary tumors, and described the landscape of the mammary tumorigenesis induced by loss of BRCA1, which provides a useful resource for better understanding of BRCA1 deficiency induced mammary tumorigenesis. Besides, we identified some marker genes for BRCA1 deficient mammary tumors, which may serve as candidate targets for diagnosis, prognosis and/or treatment of BRCA1 associated

breast cancers.

Abbreviations

CPTAC: Clinical Proteomic Tumor Analysis Consortium; CNVs: copy number variations; DEGs: differentially expressed genes; DDR: DNA damage response; ductal NOS: ductal carcinoma: not otherwise specified; EMT: epithelial-to-mesenchymal transition; FACS: fluorescence-activated cell sorting; GO: gene ontology; GSEA: gene set enrichment analysis; MG: mammary gland; MaSC: mammary stem cell; MT: mammary tumor; MC: medullary carcinoma; MET: mesenchymal to epithelial transition; P12: pregnant at day 12.5; PCA: principle component analysis; SNN: shared-nearest-neighbor; scRNA-seq: single cell RNA sequencing; t-SNE: t-distributed Stochastic Neighbor Embedding; TNBC: triple negative breast cancer.

Supplementary Material

Supplementary figures and tables.

<http://www.thno.org/v11p9967s1.pdf>

Supplementary table 1.

<http://www.thno.org/v11p9967s2.xlsx>

Supplementary table 2.

<http://www.thno.org/v11p9967s3.xlsx>

Supplementary table 3.

<http://www.thno.org/v11p9967s4.xlsx>

Supplementary table 4.

<http://www.thno.org/v11p9967s5.xlsx>

Supplementary table 5.

<http://www.thno.org/v11p9967s6.xlsx>

Supplementary table 6.

<http://www.thno.org/v11p9967s7.xlsx>

Supplementary table 7.

<http://www.thno.org/v11p9967s8.xlsx>

Supplementary table 8.

<http://www.thno.org/v11p9967s9.xlsx>

Supplementary table 9.

<http://www.thno.org/v11p9967s10.xlsx>

Supplementary table 10.

<http://www.thno.org/v11p9967s11.xlsx>

Supplementary table 11.

<http://www.thno.org/v11p9967s12.xlsx>

Acknowledgements

We thank members of the Deng, Xu and Wong laboratories for helpful advice and discussion. We thank Dapeng Hao, Haitao Wang and Wei Qiao for their assistance of data analysis. We thank Yixiang Shi and Tiffany Nga Teng Iu for their assistance of scRNA-seq and dropseq. This work was supported by National Natural Science Foundation of China Grant

81602587 granted to HS, the Chair Professor Grant (CPG2019-00019-FHS and CPG2021-00021-FHS) granted to CD, Macao Science and Technology Development Fund (FDCT) grant 065/2015/A2, 0011/2019/AKP, and 0034/2019/AGJ to CD and FDCT grant 0004/2019/AFJ to XDZ, and Guangdong Science and Technology Department grant 2020A050515005 to DT.

Ethics declarations

Our studies involving mice were approved by the University of Macau Animal Ethics Committee under protocols UMAEC-050-2015 and UMARE-015-2019.

Availability of data

The accession number for all sequencing data generated in this study is [GEO: GSE148614].

Author Contributions

HS designed and performed experiments, analyzed data, and wrote the manuscript. JZ, ZM, JKCL, CH, XZ, AZ, MD, CKHW, and XDZ analyzed data. HS and LH performed the dropseq. HS, SMS, UIC, KM, WH, HL, SG and XX dealt with the mice. YM performed part of immunostaining. JKCL, YL, ZY, DT, CKHW, XDZ, XX, and CD reviewed and modified the manuscript. CD conceptualized and supervised the study.

Competing Interests

The authors have declared that no competing interest exists.

References

1. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 1994; 266: 66-71.
2. Daniel G, Stover CFB, and Sara M. Tolaney. Neoadjuvant and Adjuvant Chemotherapy Considerations for Triple-Negative Breast Cancer. *The American Journal of Hematology/Oncology*. 2016; 12: 6-12.
3. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell stem cell*. 2010; 7: 403-17.
4. Hu L, Su L, Cheng H, Mo C, Ouyang T, Li J, et al. Single-Cell RNA Sequencing Reveals the Cellular Origin and Evolution of Breast Cancer in BRCA1 Mutation Carriers. *Cancer research*. 2021; 81: 2600-11.
5. Kennedy RD, Quinn JE, Johnston PG, Harkin DP. BRCA1: mechanisms of inactivation and implications for management of patients. *Lancet*. 2002; 360: 1007-14.
6. Deng CX. BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucleic acids research*. 2006; 34: 1416-26.
7. Deng CX, Scott F. Role of the tumor suppressor gene Brca1 in genetic stability and mammary gland tumor formation. *Oncogene*. 2000; 19: 1059-64.
8. Dine J, Deng CX. Mouse models of BRCA1 and their application to breast cancer research. *Cancer metastasis reviews*. 2013; 32: 25-37.
9. Xu X, Chen E, Mo L, Zhang L, Shao F, Miao K, et al. BRCA1 represses DNA replication initiation through antagonizing estrogen signaling and maintains genome stability in parallel with WEE1-MCM2 signaling during pregnancy. *Human molecular genetics*. 2019; 28: 842-57.
10. Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, et al. Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nature genetics*. 1999; 22: 37-43.
11. Brodie SG, Xu X, Qiao W, Li WM, Cao L, Deng CX. Multiple genetic changes are associated with mammary tumorigenesis in Brca1 conditional knockout mice. *Oncogene*. 2001; 20: 7514-23.

12. Marcus JN, Watson P, Page DL, Narod SA, Lenoir GM, Tonin P, et al. Hereditary breast cancer: pathobiology, prognosis, and BRCA1 and BRCA2 gene linkage. *Cancer*. 1996; 77: 697-709.
13. Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. *Breast Cancer Linkage Consortium*. *Lancet*. 1997; 349: 1505-10.
14. Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ, et al. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *Journal of the National Cancer Institute*. 1998; 90: 1138-45.
15. Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. *Nature reviews Cancer*. 2004; 4: 814-9.
16. Mavaddat N, Barrowdale D, Andrusis IL, Domchek SM, Eccles D, Nevanlinna H, et al. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *Cancer epidemiology, biomarkers & prevention*: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2012; 21: 134-47.
17. Kumar P, Tan Y, Cahan P. Understanding development and stem cells using single cell-based analyses of gene expression. *Development*. 2017; 144: 17-32.
18. Suva ML, Tirosh I. Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. *Mol Cell*. 2019; 75: 7-12.
19. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L. A global double-fluorescent Cre reporter mouse. *Genesis*. 2007; 45: 593-605.
20. Xiao C, Sharp JA, Kawahara M, Davalos AR, Diflippantonio MJ, Hu Y, et al. The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell*. 2007; 128: 977-89.
21. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012; 486: 405-9.
22. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012; 486: 395-9.
23. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012; 486: 400-4.
24. Tan J, Ong CK, Lim WK, Ng CC, Thike AA, Ng LM, et al. Genomic landscapes of breast fibroepithelial tumors. *Nature genetics*. 2015; 47: 1341-5.
25. Lefebvre C, Bacheltot T, Filleron T, Pedrero M, Campone M, Soria JC, et al. Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS medicine*. 2016; 13: e1002201.
26. Pereira B, Chin SF, Rueda OM, Volland HK, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nature communications*. 2016; 7: 11479.
27. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. 2018; 173: 291-304 e6.
28. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer cell*. 2018; 34: 427-38 e6.
29. Inagaki-Kawata Y, Yoshida K, Kawaguchi-Sakita N, Kawashima M, Nishimura T, Senda N, et al. Genetic and clinical landscape of breast cancers with germline BRCA1/2 variants. *Commun Biol*. 2020; 3: 578.
30. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2009; 27: 1160-7.
31. Team RC. R: A language and environment for statistical computing.; 2021.
32. Chen F, Chandrashekhar DS, Varambally S, Creighton CJ. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. *Nature communications*. 2019; 10: 5679.
33. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*. 2013; 14: R36.
34. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology*. 2013; 31: 46-53.
35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014; 15: 550.
36. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009; 4: 44-57.
37. Sun H, Miao Z, Zhang X, Chan UI, Su SM, Guo S, et al. Single-cell RNA-Seq reveals cell heterogeneity and hierarchy within mouse mammary epithelia. *The Journal of biological chemistry*. 2018; 293: 8315-29.
38. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature biotechnology*. 2014; 32: 1053-8.
39. Andrews TS, Hemberg M. M3Drop: dropout-based feature selection for scRNASeq. *Bioinformatics*. 2019; 35: 2865-7.
40. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. 2015; 161: 1202-14.
41. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature methods*. 2014; 11: 783-4.
42. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Bloem MJ, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *Jama*. 2017; 317: 2402-16.
43. Kotsopoulos J. BRCA Mutations and Breast Cancer Prevention. *Cancers*. 2018; 10.
44. Hawsawi YM, Al-Numair NS, Sobahy TM, Al-Ajmi AM, Al-Harbi RM, Baghdadi MA, et al. The role of BRCA1/2 in hereditary and familial breast and ovarian cancers. *Molecular genetics & genomic medicine*. 2019; 7: e879.
45. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. *Nature*. 2018; 556: 463-8.
46. Pastushenko I, Blanpain C. EMT Transition States during Tumor Progression and Metastasis. *Trends in cell biology*. 2019; 29: 212-26.
47. Rehman SK, Haynes J, Collignon E, Brown KR, Wang Y, Nixon AML, et al. Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. *Cell*. 2021; 184: 226-42 e21.
48. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*. 2015; 33: 495-502.
49. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014; 344: 1396-401.
50. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell*. 2018; 173: 879-93 e13.
51. Yeo SK, Zhu X, Okamoto T, Hao M, Wang C, Lu P, et al. Single-cell RNA-sequencing reveals distinct patterns of cell state heterogeneity in mouse models of breast cancer. *Elife*. 2020; 9.
52. Liberon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*. 2015; 1: 417-25.
53. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*. 2014; 32: 381-6.
54. Buckley NE, Mullan PB. BRCA1--conductor of the breast stem cell orchestra: the role of BRCA1 in mammary gland development and identification of cell of origin of BRCA1 mutant breast cancer. *Stem Cell Rev Rep*. 2012; 8: 982-93.
55. Lottin S, Adriaenssen E, Dupressoir T, Bertaux N, Montpellier C, Coll J, et al. Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis*. 2002; 23: 1885-95.
56. Davies MP, Rudland PS, Robertson L, Parry EW, Jolicoeur P, Barraclough R. Expression of the calcium-binding protein S100A4 (p9Ka) in MMTV-neu transgenic mice induces metastasis of mammary tumours. *Oncogene*. 1996; 13: 1631-7.
57. Andersen K, Mori H, Fata J, Bascom J, Oyjord T, Maelandsmo GM, et al. The metastasis-promoting protein S100A4 regulates mammary branching morphogenesis. *Developmental biology*. 2011; 352: 181-90.
58. Pellegrini P, Cordero A, Gallego MI, Dougall WC, Munoz P, Pujana MA, et al. Constitutive activation of RANK disrupts mammary cell fate leading to tumorigenesis. *Stem cells*. 2013; 31: 1954-65.
59. Joshi PA, Waterhouse PD, Kannan N, Narala S, Fang H, Di Grappa MA, et al. RANK Signaling Amplifies WNT-Responsive Mammary Progenitors through R-SPONDIN1. *Stem cell reports*. 2015; 5: 31-44.
60. Schnack Nielsen B, Rank F, Engelholm LH, Holm A, Dano K, Behrendt N. Urokinase receptor-associated protein (uPARAP) is expressed in connection with malignant as well as benign lesions of the human breast and occurs in specific populations of stromal cells. *International journal of cancer*. 2002; 98: 656-64.
61. Wienke D, Davies GC, Johnson DA, Sturge J, Lambros MB, Savage K, et al. The collagen receptor Endo180 (CD280) Is expressed on basal-like breast tumor cells and promotes tumor growth *in vivo*. *Cancer research*. 2007; 67: 10230-40.
62. Petruccioli N, Daly MB, Feldman GL. Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2. *Genet Med*. 2010; 12: 245-59.
63. Weaver Z, Montagna C, Xu X, Howard T, Gadina M, Brodie SG, et al. Mammary tumors in mice conditionally mutant for Brcal exhibit gross genomic instability and centrosome amplification yet display a recurring distribution of genomic imbalances that is similar to human breast cancer. *Oncogene*. 2002; 21: 5097-107.
64. Tirkkonen M, Johannsson O, Agnarsson BA, Olsson H, Ingvarsson S, Karhu R, et al. Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer research*. 1997; 57: 1222-7.
65. Chang CJ, Chao CH, Xia W, Yang JY, Xiong Y, Li CW, et al. p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nature cell biology*. 2011; 13: 317-23.
66. Kim T, Veronese A, Pichiorri F, Lee TJ, Jeon YJ, Volinia S, et al. p53 regulates epithelial-mesenchymal transition through microRNAs targeting ZEB1 and ZEB2. *The Journal of experimental medicine*. 2011; 208: 875-83.
67. Zhang J, Lei Y, Gao X, Liang Q, Li L, Feng J, et al. p53 Attenuates the oncogenic Ras-induced epithelial-mesenchymal transition in human mammary epithelial cells. *Biochemical and biophysical research communications*. 2013; 434: 606-13.
68. Jiang Z, Deng T, Jones R, Li H, Herschkowitz JI, Liu JC, et al. Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor

- subtypes depending on p53 status. *The Journal of clinical investigation.* 2010; 120: 3296-309.
- 69. Bertheau P, Lehmann-Che J, Varna M, Dumay A, Poirot B, Porcher R, et al. p53 in breast cancer subtypes and new insights into response to chemotherapy. *Breast.* 2013; 22 Suppl 2: S27-9.
 - 70. Wang H, Xiang D, Liu B, He A, Randle HJ, Zhang KX, et al. Inadequate DNA Damage Repair Promotes Mammary Transdifferentiation, Leading to BRCA1 Breast Cancer. *Cell.* 2019; 178: 135-51 e19.
 - 71. Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J.* 2021; 40: e107333.