Replication stress inducing ELF3 upregulation promotes BRCA1-deficient breast tumorigenesis in luminal progenitors

Jiadong Zhou[1,7], Xiao Albert Zhou[1,7], Li Hu[2,7], Yujie Ma[1], Jun Zhan[3], Zhanzhan Xu[1], Mei Zhou[1], Qinjian Shen[1], Zhaofei Liu[4], Shaohua Ma[5*], Yuntao Xie[2*], Jiadong Wang[1,6*]

Authors and affiliations

[1]Department of Radiation Medicine, School of Basic Medical Sciences, Peking University International Cancer Institute, Institute of Advanced Clinical Medicine, State Key Laboratory of Molecular Oncology, Peking University Health Science Center, Beijing 100191, China

[2]Familial & Hereditary Cancer Center, Peking University Cancer Hospital & Institute, Beijing 100142, China.

[3]Department of Anatomy, Histology and Embryology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China.

[4]Medical Isotopes Research Center, Department of Radiation Medicine, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China.

[5]State Key Laboratory of Molecular Oncology, Beijing, Key Laboratory of Carcinogenesis and Translational Research, Department of Thoracic Surgery I, Peking University Cancer Hospital and Institute, Beijing, China.

[6]Department of Gastrointestinal Translational Research, Peking University Cancer Hospital, Beijing 100142, China

[7]These authors contributed equally.

*Corresponding authors:

Jiadong Wang, Department of Radiation Medicine, Institute of Systems Biomedicine, School of Basic Medical Sciences, Peking University Health Science Center, Xueyuan Road 38, Haidian District, Beijing 100191, China. Phone: 8610-8280-2450; Email: wangjd@bjmu.edu.cn; Yuntao Xie, Familial & Hereditary Cancer Center, Peking University Cancer Hospital& Institute, Beijing, 100142, China. Phone: 8610-8819-6362; E-mail: zlxyt2@bjmu.edu.cn ; Shaohua Ma, Department of Thoracic Surgery, Peking University Third Hospital, Beijing, China. Email: doctor_msh@bjmu.edu.cn

The authors declare no potential conflicts of interest.

Abstract

BRCA1 is a critical tumor suppressor, mutations in which greatly increase risks for many tumors in carriers, most notably breast cancer. Luminal progenitor cells (LPs) are the currently recognized cells origin of BRCA1-deficient breast cancers. However, the reason why LPs are prone to transform with BRCA1 deficiency has not been elucidated. Here, using single-cell sequencing of human BRCA1 mutant breast cancers and RNA sequencing (RNA-seq) of BRCA1-deficient normal mammary cells, we reveal that replication stress is a feature of LPs and a driving factor during BRCA1-associated tumorigenesis. Mechanistically, replication stress and BRCA1 deficiency lead to significant upregulation of ELF3 expression. ELF3 can help suppress excessive genomic instability and promote LP transformation with BRCA1 deficiency. Moreover, ELF3 emerged as a core transcription factor regulating LP genes, leading to LP expansion. Our findings suggest that replication stress is a driving factor during BRCA1-associated tumorigenesis in luminal progenitor cells and elucidates the key role of ELF3 during this process.

Introduction

BRCA1 is one of the most important tumor suppressor genes. Carriers of BRCA1 germline mutations are at greatly increased risks for developing various tumors, including breast, ovarian, prostate, colon, pancreatic, and gastric cancers (*1*, *2*).Notably, BRCA1 is most closely associated with breast cancer. However, why BRCA1-associated tumors exhibit significant tissue specificity is unknown, and the effects of BRCA1 defects in normal human mammary cells are largely obscure.

In addition to the tissue specificity of tumors caused by BRCA1 deficiency, breast cancers associated with BRCA1 deficiency exhibit a set of other distinctive features. Most BRCA1 mutation carriers develop triple-negative breast cancers (TNBCs), in which tumors show negative expression of estrogen receptor (ER), progesterone receptor (PR), and Her2 (human epidermal growth factor receptor 2), and these cancers are usually also classified as basal-like breast cancers. The currently recognized cells of origin of BRCA1-deficient breast cancers are the luminal progenitor cell (LPs) in the mammary gland (*3*-*6*). LPs show abnormal expansion in the normal mammary tissue of BRCA1 mutation carriers (*3*, *7*), which might be caused by a differentiation blockade (*8*). Investigating the traits of LPs is important for tumor prevention and treatment in BRCA1 mutation carriers and a wider range of patients with TNBCs and basal-like breast cancer. However, the reason LPs are prone to transform into cancer cells when BRCA1 is deficient is unclear.

ELF3 is a member of the ETS family of transcription factors. Early reports suggested that ELF3 may regulate the development and differentiation of mammary tissue (*9*). Intriguingly, ELF3 has opposite functions in tumorigenesis in different tissues. In breast cancers, ELF3 shows high expression (*10*), plays an oncogenic role by promoting EMT (*11*) and forms a positive feedback loop with Her2 to maintain the transformation phenotype of Her2+ breast cancers (*12*). However, the mechanism of high ELF3 expression in certain types of breast cancer and its association with breast cancers beyond the Her2+ subtype is currently unknown.

In this study, utilizing single-cell sequencing of BRCA1 mutant breast cancers and time-resolved RNA-seq of BRCA1-deficient normal mammary cells, we revealed that LPs have a trait of higher replication stress compared to other cell populations in the normal mammary tissue, endowing them with the potential for transformation when BRCA1 is deficient, and ELF3 plays a key role in the process. Mechanistically,

ELF3 expression is significantly upregulated by replication stress and BRCA1 deficiency, dependent on the ATR-Chk1-E2F axis and the transcriptional regulation function of BRCA1, respectively. In addition, ELF3 can help suppress excessive genomic instability, promoting the tolerance of LPs to BRCA1 deficiency. Notably, BRCA1 deficiency causes cells to display LP transcriptional profile characteristics, leading to dedifferentiation and expansion of LPs, and ELF3 is a core transcription factor during this process. Our study reveals why BRCA1 deficiency is prone to result in tumorigenesis in LPs and elucidates the key role of replication stress and ELF3 during this process, suggesting that ELF3 could be a promising target for BRCA1-associated breast cancers.

Results

1. LPs have higher levels of replication stress compared with other mammary cell populations

LPs in the mammary gland are the currently recognized cells of origin of BRCA1-deficient breast cancers. However, the reason for LPs are prone to transform with BRCA1 deficiency has not been elucidated. Firstly, we considered that LPs, which are stimulated by upstream progesterone paracrine signals through RANKL (*5*), may have higher levels of replication stress, which can trigger tumorigenesis by increasing genomic instability. We utilized both published data and our published single-cell sequencing data (*8*) to address this question. We evaluated the replication stress levels in different mammary cell populations from the dataset of Lim et al. (*3*) by scoring multiple DNA replication pathway gene expression as previously studied (*13*, *14*). The result showed that LPs have the highest level of replication stress among all subgroups of normal human mammary cells (Fig. 1A, Supplemental Table S1). We further analyze our previously obtained single-cell sequencing data of four BRCA1 mutation carriers (case #1-4) and normal mammary cells of three matched non-carriers (case #5-7). According to our previous analysis, case #3 BRCA1 mutant carrier developed basal-like breast cancer, which was derived from LPs. Importantly, in this case of BRCA1 mutant carrier, single-cell sequencing data revealed that tumor cells had the highest levels of replication stress, and LPs had the second-highest levels (Fig. 1B, Supplemental Fig. S1A and Supplemental Table S2). Moreover, we applied Monocle to reconstruct the transformation trajectories from normal mammary cells to tumor cells in the BRCA1 mutation carrier and found that the levels of replication stress showed an increasing trend during the process of LP transformation into tumor cells (Fig. 1C and Supplemental Fig. S1B). These data demonstrate that replication stress is an important trait of LPs and might be a key driver of BRCA1-deficient tumorigenesis, which is consistent with the previous report that BRCA1 mutant mammary epithelial cells are defective in replication stress suppression (*15*).

Next, we aimed to find other key regulators during BRCA1-deficient tumorigenesis in LPs. In addition to inducing replication stress, BRCA1 deficiency also profoundly affects the transcriptional profile of normal mammary cells by disrupting the transcriptional function of BRCA1, which may further boost the process of tumorigenesis. To investigate the transcriptional impact of BRCA1 deficiency in normal human mammary epithelial cells, we took advantage of the Tet-on system to build an inducible knockdown model in MCF10A cells. By adjusting the treatment condition of DOX (doxycycline), different time points of

BRCA1 deficiency can be controlled. Treatment with DOX causes a constant and efficient knockdown of BRCA1, verified at mRNA and protein levels (Supplemental Fig. S1C). We showed that DOX treatment induces significant growth arrest and decreased cell survival (Supplemental Fig. S1D) as well as an increase in genomic instability (Supplemental Fig. S1E), consistent with the phenotypes of BRCA1-deficient cells (*16*), confirming that our system functioned as expected. We then performed time-resolved RNA-seq with different durations of DOX treatment (2, 5 and 10 days; Supplemental Fig. S1F). Notably, long-term DOX treatment resulted in decreased number of colony formation, indicating impacted cell survival, thus cells collected for sequencing at the end of 10-days DOX treatment may have gained survival advantage compared to cells that did not survive throughout the DOX treatment. Differential gene expression analysis revealed that BRCA1 was efficiently knocked down during the whole experiment (Fig. 1D). Importantly, the principal component analysis (PCA) revealed that samples showed a time-course distribution in the dimension of PC1, which has the dominant power of explanation (Fig. 1E and Supplemental Fig. S1G). These data demonstrate that the variations between samples can be largely attributed to time. To further investigate the modes of differentially expressed genes (DEGs) that change over time, we carried out the DEG cluster analysis using Mfuzz (*17*). Analysis of the Day 10 DEGs showed that genes could be divided into 5 clusters with different expression change patterns (Fig. 1F). We mainly focused on Cluster 1, for these sharply upregulated genes may play more important roles in cells gaining survival advantage with BRCA1 deficiency at Day 10, mimicking the initiation of BRCA1 deficiency associated tumorigenesis. To find out the most potent regulator upregulated both by replication stress and BRCA1 deficiency, we performed RNA-seq in HU-treated MCF10A cells and overlapped the upregulated genes in HU RNA-seq and Cluster1 genes in BRCA1 deficiency RNA-seq (Fig. 1G). The results showed that ELF3, a transcription factor, is one of the most dominant genes in this analysis. ELF3 has been reported to be involved in functions including EMT in human breast cancer and promotes the transformation of normal human mammary cells (*11, 18*). Therefore, we speculated that ELF3 could be an important driving factor during the tumorigenesis of BRCA1-deficient and basal-like breast cancer.

2. ELF3 is upregulated in BRCA1-associated breast cancer and is related to a worse prognosis.

We next explored the expression of ELF3 in breast cancers in the TCGA and METABRIC databases to

further investigate the association between ELF3 and BRCA1 and validate our previous analysis. In both clinical databases, ELF3 expression is significantly higher in BRCA1-associated breast tumors than in non-BRCA1-associated breast tumors (Fig. 2A). Moreover, the expression levels of BRCA1 and ELF3 showed a significant negative correlation, consistent with our findings that ELF3 was upregulated under BRCA1 deficiency (Fig. 2B). Since most BRCA1 mutation carriers develop basal-like breast cancers, we next investigated the association between ELF3 and basal-like breast cancers. We exploited the TCGA and METABRIC databases to compare ELF3 expression levels among different breast cancer subtypes. As we expected, ELF3 showed the highest expression levels in basal-like breast cancer in both databases (Fig. 2C). We took advantage of novel subgroups identified by the METABRIC (*19*) and further explored ELF3 expression in ten IntClusts of the METABRIC database. Among all subgroups, IntClust 5 was enriched with Her2+ cancers, and IntClust 10 was enriched with basal-like cancers. These two subgroups showed the highest expression levels of ELF3, with no significant difference between them (Fig. 2D). The higher levels of ELF3 expression in Her2 subtype breast cancer are consistent with previous studies showing that ELF3 may be a downstream transcription target of Her2 and that there might be a positive feedback loop between them (*12, 20, 21*). On the other hand, IntClust 4 shows a favorable outcome and has a 'CNA-devoid' feature, which has the lowest level of ELF3 expression. These data are consistent with our findings that ELF3 is upregulated by BRCA1 loss, indicate that ELF3 may play an important role in basal-like breast tumorigenesis, and moreover, the expression of ELF3 may be related to genomic stability status. Furthermore, among human breast cancer tissue samples, ELF3 expression levels were significantly higher in TNBCs than in other subtypes (Fig. 2E and F). Additionally, TNBCs have higher ELF3 expression levels than the Her2 subtype (Fig. 2E and F), revealing a tighter connection between ELF3 and BRCA1-associated breast cancers.

ELF3 shows significant copy number amplification in breast cancer (*22*). Thus, we wondered that whether the high expression levels of ELF3 in BRCA1-associated breast cancers are due to copy number variation. Therefore, we analyzed the proportion of ELF3 copy number amplification in different subtypes (Supplement Fig. S2) and found that this was not the case. In TCGA basal-like breast cancer, the ELF3 amplification proportion was not higher than that in other subtypes but was only 3.5%, which was lower than the total level (7.3%). These data demonstrate that the upregulation of ELF3 in BRCA1-associated

breast cancer is not caused by copy number amplification.

Furthermore, we investigated whether ELF3 expression could be a biomarker for breast cancer prognosis. We found that higher ELF3 expression is correlated with worse prognosis in TNBC and basal-like breast cancer (Fig. 2G). In the Her2 subtype, higher ELF3 expression is also related to shorter overall survival, consistent with previous reports that as a Her2 target gene, ELF3 may help promote tumor invasion (*21*). Interestingly, ELF3 expression shows less or no prognostic prediction value in luminal A and luminal B subtypes, suggesting that ELF3 plays a more powerful role in TNBC and basal-like breast cancer. These data imply that in BRCA1-associated breast cancer, ELF3 tends to be upregulated and plays a “driver” role, promoting tumor progression and leading to a worse prognosis.

3. ELF3 upregulation is induced by replication stress via the ATR-Chk1-E2F axis and by BRCA1 deficiency via GATA3 transcription.

We next focused on the mechanisms of ELF3 upregulation in the presence of replication stress and BRCA1 deficiency. Consistent with the HU RNA-seq analysis, our single-cell RNA-seq data showed that in both BRCA1 WT and mutant carriers, the levels of replication stress and ELF3 expression showed a significant positive correlation (Fig. 3A). ELF3 was also upregulated by HU-induced replication stress in MCF10A cells (Fig 3B). Since ATR is the key kinase that responds to replication stress, we treated cells with ATRi and found that it could efficiently block ELF3 upregulation induced by HU, while ATMi had no such effect (Fig. 3B). Moreover, downstream of ATR, ELF3 expression upregulation depended on Chk1 activation but not Chk2 activation (Fig. 3C). It has been reported that the transcriptional activation induced by replication stress-ATR-Chk1 pathway largely depends on E2F transcription factor activation (*23*). Under replication stress, ATR-Chk1-mediated phosphorylation of E2F6 blocks its inhibitory function on E2F and activates E2F transcription. Thus, we overexpressed E2F6 and found that E2F6 largely blocks the upregulation of ELF3 under HU treatment conditions (Fig. 3D), validating that ELF3 upregulation relies on E2F6 inactivation and E2F de-suppression. Moreover, utilizing ENCODE ChIP-seq data, we found that the ELF3 promoter is indeed bound by E2F1 (Fig. 3E). These data demonstrate that ELF3 is upregulated by the ATR-Chk1-E2F axis under replication stress conditions.

Then we investigate the mechanism of ELF3 upregulation induced by BRCA1 deficiency. ELF3

expression gradually increased during 10 days of DOX treatment (Fig. 3F). ELF3 upregulation was further confirmed by siRNA-mediated BRCA1 knockdown in MCF10A cells to strictly exclude the unknown influence of the Tet-on system (Fig. 3G). BRCA1 has transcription regulation functions by interacting with other transcription factors to bind to gene promoters and regulate gene expression (*24*). Therefore, we explored transcription factors associated with BRCA1, including GATA3 (*25*) and Myc (*26*), for whether the binding motifs of these transcription factors appear in the ELF3 promoter sequence utilizing the JASPAR database (Supplement Fig. S3A). We found that E2F1 was present in the prediction list with a high relative score, consistent with our results that E2F1 can bind ELF3 promoter. The prediction results show that the motif of GATA3 appeared in the ELF3 promoter multiple times with an overall high relative score. Indeed, knockdown of GATA3 in MCF10A cells also strongly induced ELF3 expression (Fig. 3H). Similarly, we found multiple GATA3 binding peaks at the promoter of ELF3 in two independent ChIP-seq datasets from the ENCODE database, consistent with two of the GATA3 binding sequences predicted in the JASPAR database (Fig. 3I). These data suggest that ELF3 is a GATA3 downstream target gene. Importantly, when both BRCA1 and GATA3 were knocked down, there was no significant additive effect on ELF3 upregulation (Fig. 3J and Supplement Fig. S3B), implying that BRCA1 and GATA3 mainly act in the same pathway for ELF3 induction.

Notably, we found that and BRCA1 knockdown and HU treatment have a superimposed effect on upregulating ELF3 expression levels (Fig. 3K and Supplement Fig. S3C). Furthermore, GATA3 knockdown and HU treatment also had a significant superimposed effect in upregulating ELF3 expression (Fig. 3L and Supplement Fig. S3D). These data validate that the BRCA1-GATA3 axis and replication stress regulates ELF3 expression independently.

4. ELF3 helps cells tolerate replication stress and sustain cell survival

Next, we investigated the impact of ELF3 upregulation under replication stress conditions and BRCA1 deficiency. Based on the association between ELF3 and the worse prognosis in BRCA1-associated breast cancer patients (Fig. 2G), we hypothesized that upregulation of ELF3 could help cells deal with replication stress and suppress excessive genomic instability, facilitating cancer initiation and cancer evolution. As expected, ELF3 knockdown greatly increased the sensitivity of MCF10A cells to HU and cisplatin (Fig.

4A and B, Supplement Fig. S4A). This finding suggests that ELF3 can help normal cells deal with replication stress. Furthermore, ELF3 knockdown in HCC1937 and SUM149PT cells, both of which are human breast cancer cell lines with BRCA1 mutations and allelic loss, also caused a significant decline in cell proliferation (Fig. 4C and D, Supplement Fig. S4B and C), suggesting that ELF3 is essential for BRCA1 mutant breast cancer cell survival. Moreover, HCC1937 and SUM149PT cells became more sensitive to HU and cisplatin when ELF3 was deficient (Fig. 4E and F, Supplemental Fig. S4D and E). Since PARP inhibitors can induce replication fork stalling by PARP1 trapping and lead to replication stress in BRCA1-deficient cells, we also tested olaparib sensitivity in HCC1937 cells and obtained similar results (Supplement Fig. S4F). These results demonstrate that in BRCA1 mutant breast cancer cells, ELF3 can help cells tolerate replication stress and sustain cell survival.

In addition, utilizing the CellMiner database, we found that among nearly sixty diverse human cancer cell lines, lower ELF3 expression is correlated with higher drug sensitivity to cisplatin and etoposide (Fig. 4G). Consistent results were obtained using the GDSC database (Fig. 4H, Supplement Fig. 4G). Moreover, in nude mice, SUM149PT tumors with ELF3 deficiency were significantly smaller than those in the control group (Fig. 4I). These data indicate that ELF3 may help normal cells tolerate replication stress and subsequent DNA damage, allowing transformation into malignant tumors. During cancer progression, BRCA1 mutant cancer cells gradually become reliant on high ELF3 expression to preserve a certain degree of genomic stability, making ELF3 a promising therapeutic target and biomarker in BRCA1-associated breast cancers.

5. ELF3 helps maintain the stability of DNA replication.

We then investigated the mechanism by which ELF3 helps tolerate replication stress. ELF3 deficiency leads to more endogenous DNA damage in MCF10A cells (Supplement Fig. S5A). In BRCA1-deficient breast cancer cell line HCC1937, knockdown of ELF3 resulted in a significant increase in spontaneous γH2AX foci (Fig. 5A), demonstrating that ELF3 deficiency leads to more endogenous DNA damage in BRCA1-deficient breast cancer cells. Moreover, with HU treatment and release for repair, ELF3 deficiency resulted in a significant increase in the number of γH2AX and 53BP1 foci per cell and the proportion of cells with damage foci (Fig. 5B-D). Under replication stress conditions, γH2AX and 53BP1 foci represent DNA double-strand breaks formed by collapsed replication forks (*27*). These data suggest

that ELF3 deficiency results in a defective ability to repair stalled replication forks; thus, more DNA damage is generated. The upregulation of ELF3 expression during replication stress helps cells better manage stalled replication forks and prevents excessive DNA damage.

To further explore how ELF3 contributes to replication stress management, RNA-seq on ELF3-knockdown MCF10A cells was performed. Gene set enrichment analysis (GSEA) showed that multiple pathways associated with DNA replication were significantly downregulated in ELF3-deficient cells (Fig. 5E), which was further verified by RT–qPCR (Fig. 5F and Supplement Fig. S5B). These data indicate that DNA replication pathways are affected by ELF3 deficiency.

DNA fiber assay was performed to more closely investigate the effects of ELF3 deficiency on DNA replication. ELF3 deficiency significantly slowed down the DNA replication progression (Fig. 5G), with significantly reduced symmetry of replication forks (Fig. 5H), suggesting that more replication forks were stalled or unstable. Moreover, with HU treatment, ELF3 knockdown undermined the stability of stalled replication forks (Fig. 5I). Notably, double-knockdown of ELF3 and BRCA1 led to a further decrease in replication fork stability (Fig. 5I). These data reveal that ELF3 expression upregulation when BRCA1 is deficient can help cells deal with replication stress, while simultaneous ELF3 deficiency leads to intolerable levels of genomic stability.

6. ELF3 is a core transcription factor inducing LP genes expression under BRCA1 deficiency

Finally, we explored the possible reasons for the tissue specificity of BRCA1-associated tumors. It has been reported that BRCA1 deficiency disturbs the development and differentiation of mammary tissue and causes abnormal expansion of the LP population (*3, 8, 28*). We hypothesized that BRCA1 deficiency might affect the normal differentiation process of LPs by disturbing the transcriptional profile, thus leading to abnormal expansion and transformation of LPs. We analyzed the time-course RNA-seq data using a published signature gene set from LPs (*3*) and found that over time with BRCA1 deficiency, normal mammary cells gradually show a higher signature score of LP gene expression (Fig. 6A). Similarly, using the sets of genes up-and downregulated in the LP population compared to in other mammary cell populations obtained from the data of Pellacani et al. (*29*), we found that genes upregulated in LPs also tended to be upregulated in BRCA1-deficient conditions, whereas genes downregulated in LPs tended to be downregulated (Fig. 6B). These results suggest that BRCA1 may broadly affect the transcription profile

of mammary cells, and therefore, when BRCA1 is deficient, LPs may tend to maintain a dedifferentiated state, leading to the abnormal expansion and tumorigenic properties of these cells in BRCA1 mutant carriers.

Multiple ETS family transcription factors were shown to be involved in human mammary development and differentiation in previous reports. However, the function of ELF3 in LPs has not yet been elucidated. We used GSEA to explore the transcriptional profiles of BRCA1-deficient mammary cells, ELF3-overexpressing mammary cells and human LPs from previous reports (*29*). Surprisingly, these three transcriptional profiles have significant correlations (Fig. 6C), indicating that ELF3 is not only a core transcription node under BRCA1 deficiency conditions but also, more importantly, that ELF3 plays significant roles in the transcriptional regulation of human LPs. In addition, in our single-cell sequencing data, the expression of ELF3 in LPs is the highest among all mammary cell populations (Fig. 6D and Supplement Fig. S6A). Same results were obtained by analyzing the published data from Pellacani et al. (*29*) (Supplement Fig. S6B). To investigate ELF3 functions in regulating LP genes transcription, we used SEA (simple enrichment analysis) of MEME to perform motif enrichment of ESE and ELF subfamilies transcription factors in promoters of LP genes, as these transcription factors are more closely related to ELF3 in the ETS family, and several, such as ELF5 and EHF, have been reported to regulate LPs (*29-31*). ELF3 was at the top of the list (Fig. 6E), demonstrating that ELF3 has higher affinity for the promoters of LP genes. Moreover, our single-cell data revealed that ELF3 expression levels were the highest in LPs compared to other members of the ESE and ELF subfamilies (Fig. 6F), including ELF5, which has been reported to play vital roles in LP differentiation. In addition, we analyzed the ATAC-seq data in Brca1-deficient mouse LPs from Bach et al. (*32*). The ATAC-seq compared the chromatin accessibility of LPs in Brca1/p53-deficient mice with WT mice, and motif enrichment analysis of transcription factors was performed in regions with increased accessibility of Brca1-deficient mouse LPs. ELF3 was among the top genes (Supplement Fig. S6C) and is significantly enriched among all ETS transcription factors (Fig. 6G). These data further confirm our hypothesis that ELF3 plays an important role in the transcriptional regulation of LPs.

Discussion

In this study, we revealed that LPs have the trait of higher replication stress, an important factor for their potential to transformation. The difference of LPs from other cell populations in normal mammary tissue has received considerable attention in the field. Research in mouse mammary tissue has revealed that the overall luminal epithelial cell population is enriched in mitotic, cell cycle and DNA repair-related pathways under estrogen-and progestogen-stimulated conditions relative to the basal and stromal cell populations (*33*). Subsequently, Casey et al. (*34*) found that LPs of mouse mammary tissue are enriched in DNA repair and cell division pathways. A recent preprint from the same group (*35*) continued the analysis of mouse tissue sequencing data and found that in LPs, DNA repair and replication pathways are significantly upregulated, indicating that LPs are more capable of DNA damage repair and thus more tolerant to BRCA1 deficiency. Furthermore, Nolan E et al. (*5*) found that RANK+ LPs are highly proliferative and prone to DNA damage. These studies did not reveal the reason for LP tendency to transform. In our study, we focus on the characteristic of LPs compared to other cell populations and using sequencing data from normal human mammary tissues (*3*), we reveal that the DNA replication pathways and the cellular response to replication stress pathways are significantly upregulated in LPs compared to those in other cell populations. We argue that high levels of replication stress may be the major source of DNA damage in LPs, resulting in DDR gene upregulation in these cells, and the high proliferation feature of LP could be the main reason for the high replication stress.

Replication stress is a hallmark of cancer, fueling cancer evolution by increasing genomic instability (*36*). LPs have inherently higher replication stress levels, and when BRCA1 is deficient, replication stress is further elevated during the transformation from normal LPs to tumor cells, suggesting that replication stress is an important driver of BRCA1-deficient tumorigenesis and the suppression of replication stress could be a new approach for preventive intervention in BRCA1 mutation carriers with the advantages of being less damaging than prophylactic resection. It was reported haploinsufficiency in breast epithelial cell cultures from BRCA1 heterozygotes, which could lead to premature failure through replication stress(*37*). Moreover, we found that both replication stress and BRCA1 deficiency itself can induce high expression of ELF3, which also plays an important role in the tendency of LPs to transform with BRCA1 deficiency. First, ELF3 can help suppress excessive genomic instability, which is crucial for cell tolerance and survival with BRCA1 deficiency. We speculate that high ELF3 expression can help LPs maintain

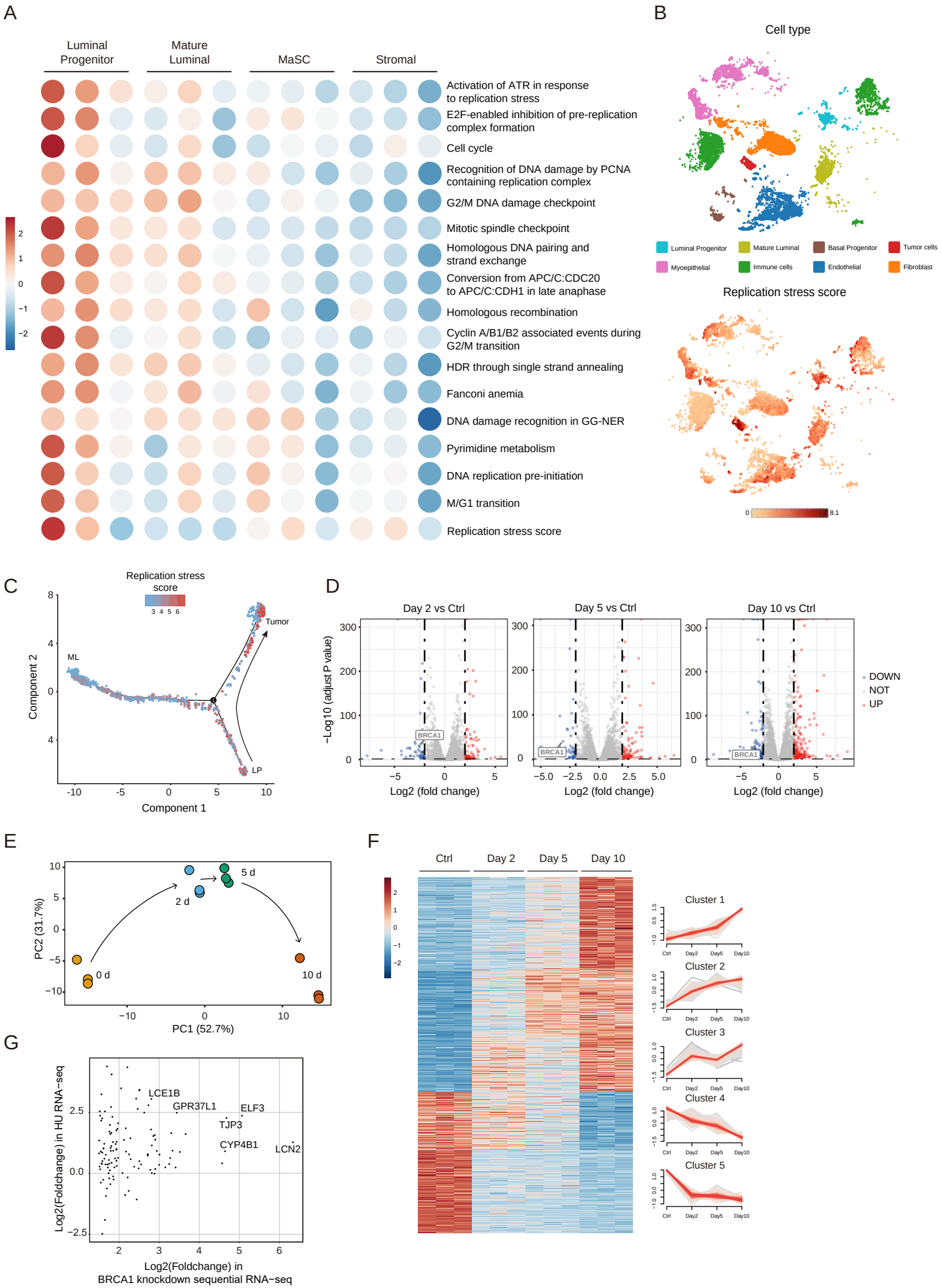
moderate levels of genomic instability that allow cells to proliferate and is sufficient to fuel cancer evolution, including subsequent events such as p53 mutation and BRCA1 loss of heterozygosity (LOH). Second, ELF3 upregulation can boost LP gene transcription, leading to LP dedifferentiation and expansion. Finally, as an EMT promoter in mammary tissue, ELF3 can accelerate the transformation of LPs to malignant cells. Therefore, high levels of replication stress and ELF3 expression are important factors resulting in the high tendency for LPs to transform (Fig. 7).

A key feature of BRCA1-associated tumors is the tissue specificity, the mechanism of which has long been a research focus. The transcriptional regulation functions of BRCA1 are critical for the development and differentiation of normal mammary tissue. LPs with abnormal expansion have been observed in BRCA1 mutation carriers. Several important differentiation pathways, such as WNT, Notch and Hedgehog, were downregulated in these abnormal LPs (*8*), suggesting that BRCA1 deficiency may lead to a differentiation blockade in LPs and, therefore, abnormal expansion. Our RNA-seq data revealed that the transcriptional profile of BRCA1-deficient MCF10A cells has a striking similarity to that of LPs in normal human mammary tissue, and ELF3 might be a core mediating transcription factor. This finding suggests that BRCA1 has a broad regulatory effect on gene expression in normal mammary cells and has specific functions in LPs; thus, its deficiency causes LPs to maintain a dedifferentiated state, leading to abnormal expansion of this cell population, and ELF3 may be a core transcription factor inducing these changes. In addition, our study reveals another layer of the mechanism underlying the tissue specificity of BRCA1-associated tumors—the capability of BRCA1 to regulate ELF3 expression. ELF3 is an epithelial-specific transcription factor, but the functions of ELF3 in different epithelial tissues vary. For example, ELF3 promotes tumor development in breast, prostate, and lung cancers (*10, 38, 39*), whereas it suppresses tumorigenesis in bladder and oral squamous cancers (*40, 41*). In addition, ELF3 itself plays an important role in the development and differentiation of normal mammary tissue; thus, dysregulation of ELF3 may disturb these processes. Therefore, we suggest that replication stress caused by BRCA1 deficiency leads to upregulation of ELF3 expression in LPs, thus promoting the tissue-specific oncogenic functions of ELF3 and breast tumorigenesis.

This study also has some limitations. First, we used the MCF10A cell line, which is widely used for investigations into normal human mammary epithelial cells with the advantage of being easy to manipulate

and modify, and compared to patient-derived cells, unknown genetic backgrounds in addition to BRCA1 mutations are avoided. However, the transcriptional profile and epigenetic characteristics of this cell line are distinct from those of normal human mammary cell populations and LPs (*29*), thus there are certain gaps between the cell model and real LP cells. Although our data show that MCF10A cells can reflect the actual functions of BRCA1 in terms of maintaining genomic stability (Fig. 1B and C) and our RNA-seq data demonstrate enrichment in the known transcriptional target genes of BRCA1, the effect of BRCA1 defects and replication stress in LPs needs to be further verified in primary human mammary cell populations. And analysis of TNBC as well as Basel like tumor data needs to be further confirmed in BRCA1-deficient breast cancer. Second, the higher replication stress levels in LPs may lead to altered expression or modifications of molecules other than ELF3, affecting the development and progression of BRCA1-associated breast cancer, which also needs to be further explored. In addition, there are likely other DEGs in our BRCA1-deficient RNA-seq that play important roles in the progression of BRCA1-associated breast cancer, and these genes also have potential research value.

Taken together, we revealed that LPs have a trait of higher replication stress, endowing them with the potential for transformation with BRCA1 deficiency. Moreover, the levels of replication stress gradually increase during tumorigenesis, indicating that replication stress is a driving factor of BRCA1-associated malignant transformation. Mechanistically, ELF3 is significantly upregulated upon replication stress and BRCA1 deficiency, which can help suppress excessive genomic instability by stabilizing the DNA replication process, empowering LPs to tolerate and to transform with BRCA1 deficiency. In cells with inherently low replicative stress, such as other non-LP mammary cells, the ELF3-associated mechanism might help cells endure the high replicative stress caused by BRCA1 deficiency without leading to cancerous changes. However, in LP cells, which naturally experience higher replicative stress, this ELF3-related mechanism may make them more susceptible to transformation into cancer cells. Notably, BRCA1 deficiency causes a tendency for cells to display characteristics of the LP transcriptional profile, and ELF3 is a core transcription factor mediating this process. Our study reveals why BRCA1 deficiency is prone to result in tumorigenesis in LPs and elucidates the key role of replication stress and ELF3 during this process, and suggests promising targets for BRCA1-associated breast cancers.


A
Luminal Progenitor
Mature Luminal
MaSC
Stromal
Activation of ATR in response to replication stress
E2F-enabled inhibition of pre-replication complex formation
Cell cycle
Recognition of DNA damage by PCNA containing replication complex
G2/M DNA damage checkpoint
Mitotic spindle checkpoint
Homologous DNA pairing and strand exchange
Conversion from APC/C:CDC20 to APC/C:CDH1 in late anaphase
Homologous recombination
Cyclin A/B1/B2 associated events during G2/M transition
HDR through single strand annealing
Fanconi anemia
DNA damage recognition in GG-NER
Pyrimidine metabolism
DNA replication pre-initiation
M/G1 transition
Replication stress score
B
Cell type
Luminal Progenitor
Mature Luminal
Basal Progenitor
Tumor cells
Myoepithelial
Immune cells
Endothelial
Fibroblast
Replication stress score
C
Replication stress score
Tumor
ML
LP
Component 2
Component 1
D
Day 2 vs Ctrl
Day 5 vs Ctrl
Day 10 vs Ctrl
−Log10 (adjust P value)
Log2 (fold change)
BRCA1
DOWN
NOT
UP
E
PC2 (31.7%)
PC1 (52.7%)
0 d
2 d
5 d
10 d
F
Ctrl
Day 2
Day 5
Day 10
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
G
Log2(Foldchange) in HU RNA−seq
Log2(Foldchange) in BRCA1 knockdown sequential RNA−seq
LCE1B
GPR37L1
ELF3
TJP3
CYP4B1
LCN2

Figures

Figure 1. LPs have higher levels of replication stress compared with other mammary cell populations.

(A) Replication stress pathway scores of different human normal mammary cell populations, scaled by row, transformed to mean=0 and variance=1. The gene expression data were from Lim et al. (3). The gene expression signatures were from Dreyer et al. (13) and Takahashi et al. (14).

(B) T-distributed statistical neighbor embedding (t-SNE) plots visualizing the replication stress scores and ELF3 expression levels in breast cancer and normal mammary tissue from a BRCA1-mutated triple-negative breast cancer patient. The cells are colored by cell type (top) and replication stress score (bottom).

(C) Monocle-generated cell trajectories visualizing the change in replication stress scores during tumorigenesis. The tumor cells and compared normal luminal cells were obtained from a triple-negative breast cancer patient (case #3) carrying a BRCA1 germline mutation. LP, luminal progenitor cells; ML, mature luminal cells.

(D) Volcano plots showing DEGs in MCF10A-shBRCA1-Tet-on cell RNA-seq results after 1 μg/mL DOX treatment for 2 (left), 5 (middle) and 10 (right) days versus the Ctrl group. DEGs are genes with |log2-fold change| > 2 and FDR < 0.001.

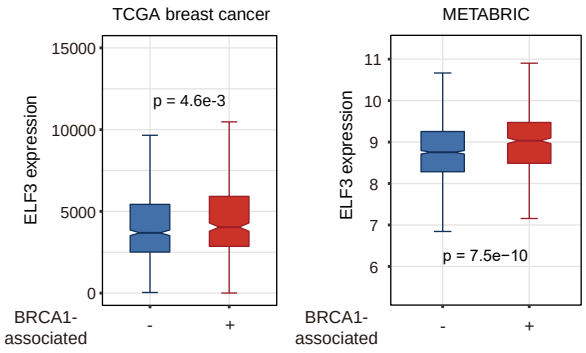
(E) PCA plot of MCF10A-shBRCA1-Tet-on cell RNA-seq results.

(F) Mfuzz clustering analysis of Day 10 versus Ctrl DEGs. The heatmap (left) shows the differential expression mode of all genes in each cluster. The line chart (right) shows the global changing mode of each cluster.

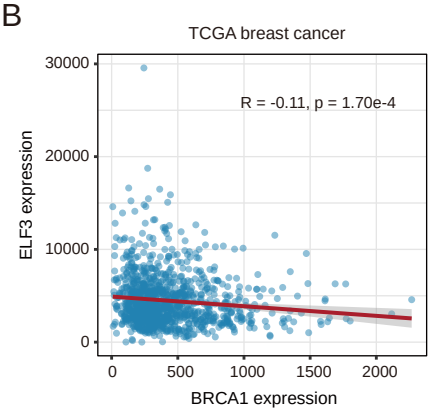
(G) Overlap of HU treated MCF10A RNA-seq results and MCF10A-shBRCA1-Tet-on cell RNA-seq Cluster1 genes.

Figure 2

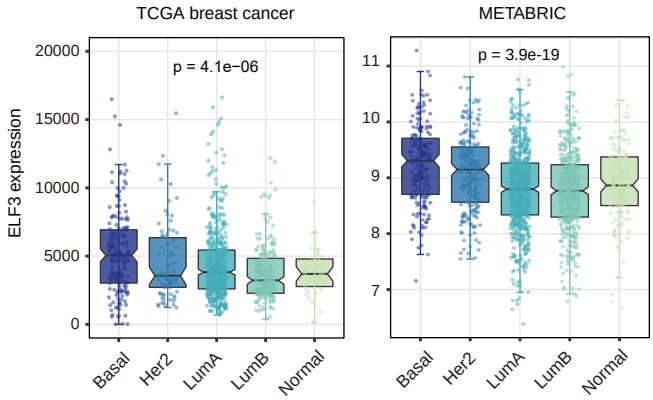
A



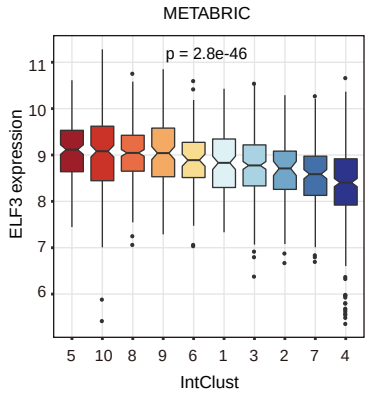
B



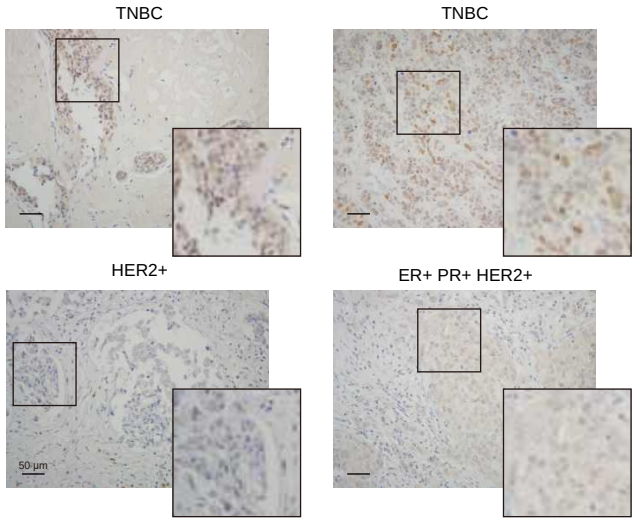
C



D



E



F

	0	1	2	3	total
non - TNBC	26 24.8%	54 51.4%	21 20%	4 3.8%	105
TNBC	3 11.5%	13 50%	6 23.1%	4 15.4%	26
Her2+	15 35.7%	18 42.9%	6 14.3%	3 7.1%	42

G

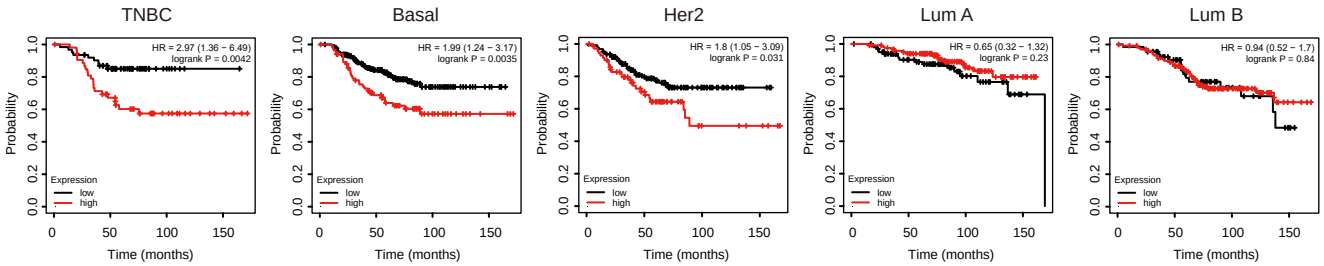


Figure 2. ELF3 is upregulated in BRCA1-associated breast cancers and is related to a worse prognosis.

(A) ELF3 expression levels in BRCA1-associated breast cancers and BRCA1-nonassociated breast cancers in the TCGA and METABRIC databases. BRCA1-associated breast cancers comprise samples with BRCA1 mutations, heterozygous loss and homozygous deletion (Mann–Whitney U test)(TCGA: BRCA1-associated, N=362; BRCA1-nonassociated, N=720; METABRIC: BRCA1-associated, N=540, BRCA1-nonassociated, N=1364).

(B) Correlation of ELF3 and BRCA1 expression levels in TCGA breast cancer datasets.

(C) ELF3 expression levels of different molecular subtypes of breast cancers in the TCGA and METABRIC databases (one-way ANOVA).

(D) ELF3 expression levels of different IntClusts of breast cancers in the METABRIC database (one-way ANOVA; for IntClust 5 and 10, Mann–Whitney U test, ns, no significance).

(E) Immunohistochemistry images of human triple-negative, Her2+ and ER+ PR+ Her2+ breast cancer samples stained with the ELF3 antibody.

(F) ELF3 expression levels of the indicated types of human breast cancers (Mann-Whitney U test, *, $p < 0.05$).

(G) Survival analysis of the indicated subtypes of human breast cancer in the KMplot database. For the TNBC group, samples with negative expression of all hormone receptors (ER, PR and Her2) are included. In the TNBC group, high expression (n=54), low expression (n=62); in the basal group, high expression (n=85), low expression (n=211); in the Lum A group, high expression (n=121), low expression (n=101); in the Lum B group, high expression (n=133), low expression (n=67); and in the Her2 group, high expression (n=72), low expression (n=126).

Figure 3

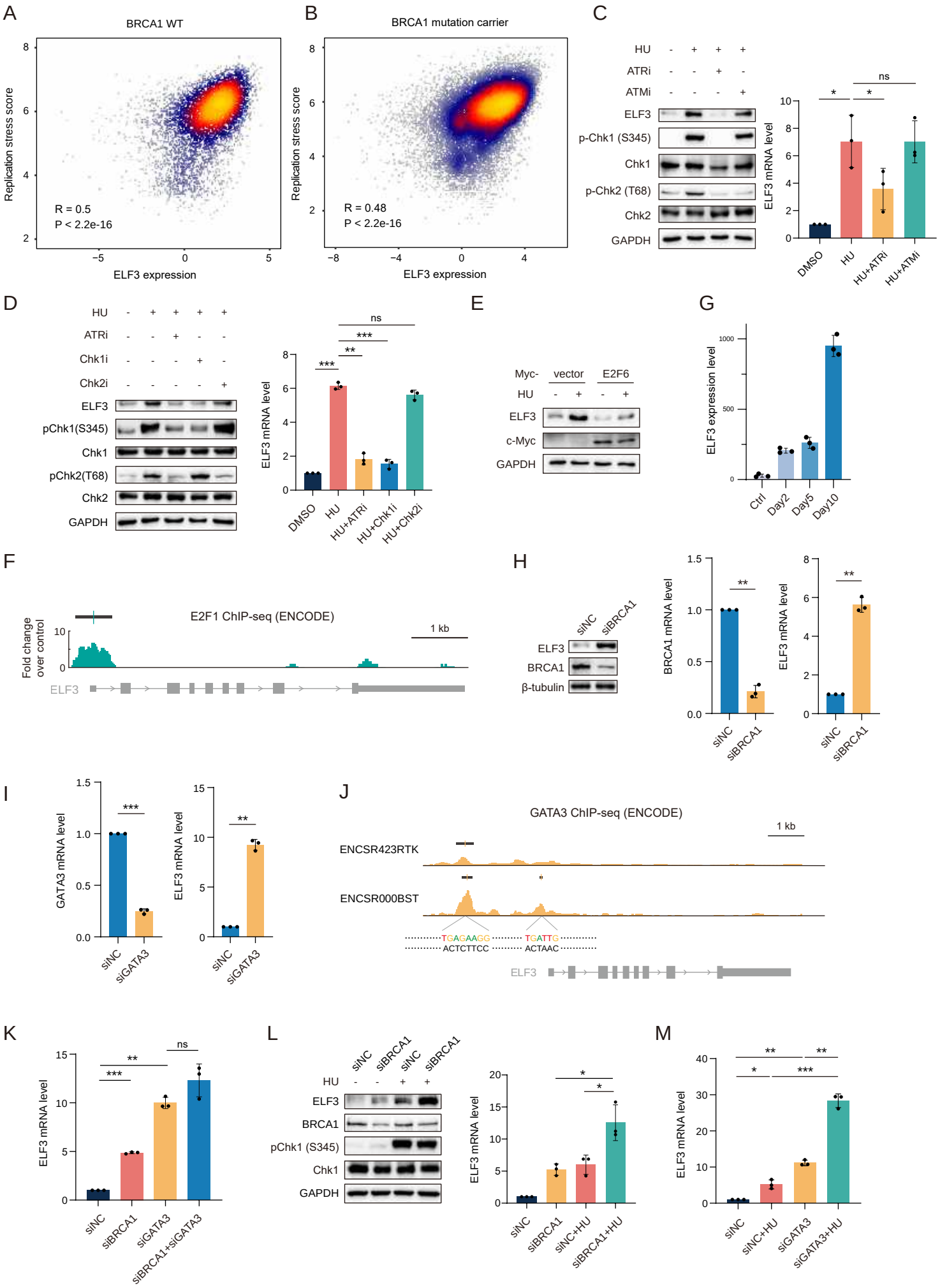

A
BRCA1 WT
Replication stress score
ELF3 expression
R = 0.5
P < 2.2e-16
B
BRCA1 mutation carrier
Replication stress score
ELF3 expression
R = 0.48
P < 2.2e-16
C
HU
ATRi
ATMi
ELF3
p-Chk1 (S345)
Chk1
p-Chk2 (T68)
Chk2
GAPDH
ELF3 mRNA level
DMSO
HU
HU+ATRi
HU+ATMi
ns
D
HU
ATRi
Chk1i
Chk2i
ELF3
pChk1(S345)
Chk1
pChk2(T68)
Chk2
GAPDH
ELF3 mRNA level
DMSO
HU
HU+ATRi
HU+Chk1i
HU+Chk2i
E
Myc-
vector
E2F6
HU
ELF3
c-Myc
GAPDH
G
ELF3 expression level
Ctrl
Day2
Day5
Day10
F
E2F1 ChIP-seq (ENCODE)
Fold change over control
1 kb
ELF3
H
siNC
siBRCA1
ELF3
BRCA1
β-tubulin
BRCA1 mRNA level
ELF3 mRNA level
I
GATA3 mRNA level
ELF3 mRNA level
siNC
siGATA3
J
GATA3 ChIP-seq (ENCODE)
1 kb
ENCSR423RTK
ENCSR000BST
TGAGAAGG
ACTCTTCC
TGATTG
ACTAAC
ELF3
K
ELF3 mRNA level
siNC
siBRCA1
siGATA3
siBRCA1+siGATA3
ns
L
siNC
siBRCA1
siNC
siBRCA1
HU
ELF3
BRCA1
pChk1 (S345)
Chk1
GAPDH
ELF3 mRNA level
siNC
siBRCA1
siNC+HU
siBRCA1+HU
M
ELF3 mRNA level
siNC
siNC+HU
siGATA3
siGATA3+HU

Figure 3. ELF3 upregulation is induced by replication stress via the ATR-Chk1-E2F axis and by BRCA1 deficiency via GATA3 transcription.

(A) Spearman correlation analysis of ELF3 expression levels and replication stress scores in normal mammary tissue from noncarriers (left) and BRCA1 mutation carriers (right). Expression data is from single-cell RNAseq data, including 3 BRCA1 WT patients and 4 BRCA1 mutant patients. All cells (normal cells and tumor cells) are involving, and ELF3 expression was normalized by reads in each cell.

(B) Western blot and RT–qPCR results of MCF10A cells treated with the indicated drugs for 12 h. Drug concentration: HU 1 mM, ATRi (VE821) 10 μM, ATMi (KU5593) 10 μM.

(C) Western blot and RT–qPCR results of MCF10A cells treated with the indicated drugs for 12 h. Drug concentrations: HU 1 mM, ATRi (VE821) 10 μM, Chk1i (GDC-0575) 50 μM, Chk2i (BML-277) 10 μM.

(D) Western blot results of MCF10A cells transfected with the empty vector or Myc-E2F6 and treated with DMSO or 1 mM HU for 12 h.

(E) ChIP-Seq enrichment signal (green plot, ENCODE ENCFF858GLM; data are representative of two independent experiments) and the location of detected peaks (black line, ENCODE ENCFF692OYJ; irreproducible discovery rate (IDR) thresholded peaks) of E2F1 in MCF7 cells from the ENCODE database.

(F) ELF3 expression changes as BRCA1 deficiency duration increases. ELF3 expression levels are indicated by the CPM of the RNA-Seq results (means ± SD, n=3).

(G) Western blot and RT–qPCR results of MCF10A cells transfected with siNC or siBRCA1 for 48 h.

(H) RT–qPCR results of MCF10A cells transfected with siNC or siGATA3 for 48 h.

(I) ChIP-Seq enrichment signal (yellow plot up, ENCODE ENCFF384CPN; yellow plot down, ENCODE ENCFF342GNN. Data are representative of two independent experiments) and the location of detected peaks (black line up, ENCODE ENCFF352QVM; black line down, ENCODE ENCFF437NQS; both are irreproducible

discovery rate (IDR) thresholded peaks) of GATA3 in MCF7 cells from two independent studies (ENCSR423RTK and ENCSR000BST) from the ENCODE database. The motif sequences were obtained from the JASPAR database.

(J) RT–qPCR results of MCF10A cells transfected with the indicated siRNA for 48 h.

(K) Western blot and RT–qPCR results of MCF10A cells transfected with siNC or siBRCA1 for 24 h followed by treatment with DMSO or 1 mM HU for 12 h.

(L) RT–qPCR results of MCF10A cells transfected with siNC or siGATA3 for 24 h followed by treatment with 1 mM HU for 12 h. In all panels of RT–qPCR results, data are presented as the mean ± SD, n=3, *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, no significance, by paired two-tailed Student’s t test.

Figure 4

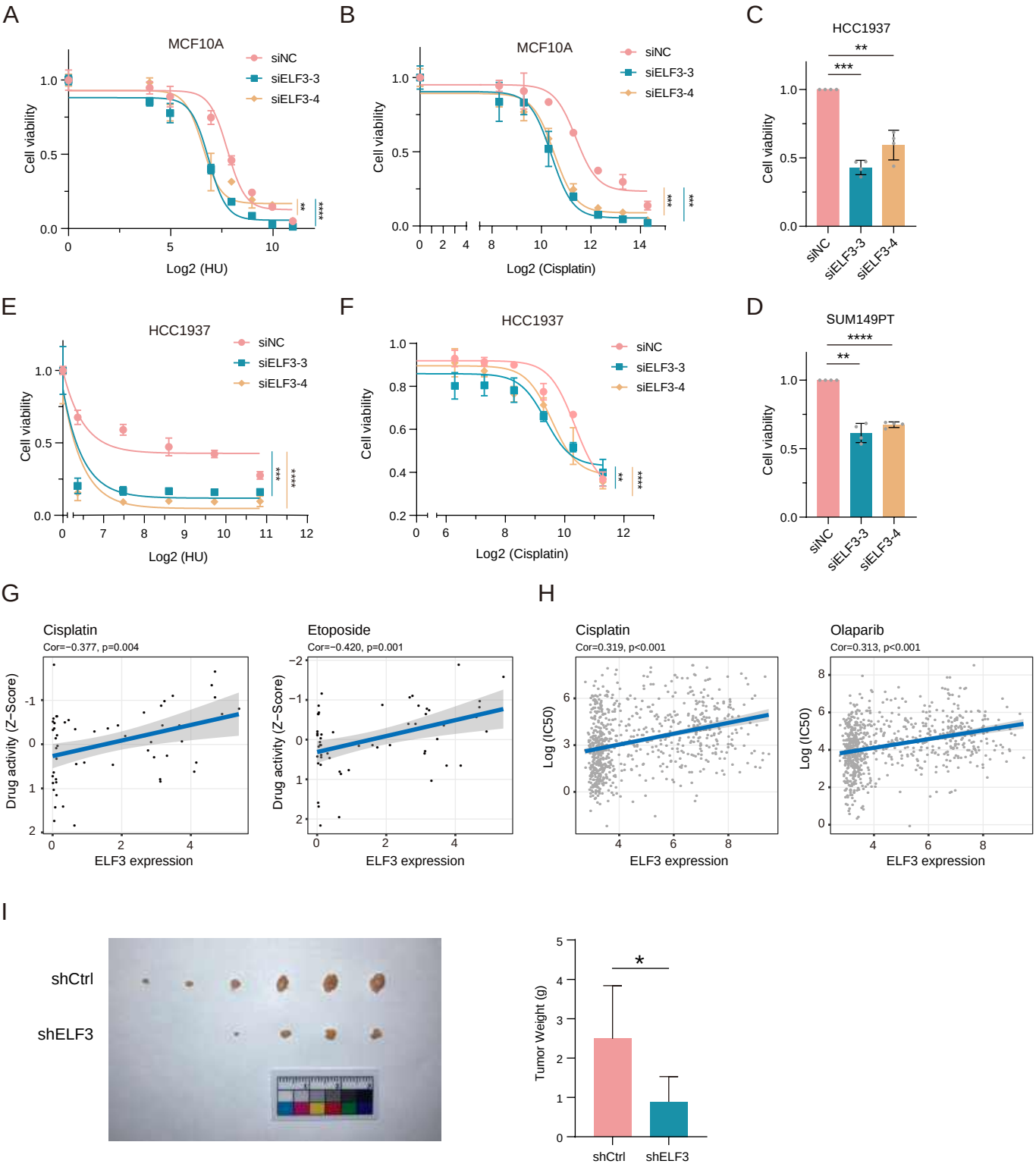

A
MCF10A
siNC
siELF3-3
siELF3-4
Cell viability
Log2 (HU)
B
MCF10A
Log2 (Cisplatin)
C
HCC1937
siELF3-3
siELF3-4
D
SUM149PT
E
HCC1937
F
HCC1937
G
Cisplatin
Cor=−0.377, p=0.004
Etoposide
Cor=−0.420, p=0.001
Drug activity (Z−Score)
ELF3 expression
H
Cisplatin
Cor=0.319, p<0.001
Olaparib
Cor=0.313, p<0.001
Log (IC50)
ELF3 expression
I
shCtrl
shELF3
Tumor Weight (g)

Figure 4. ELF3 helps suppress excessive genomic instability.

(A and B) Cell viability curve of MCF10A cells transfected with siNC or siELF3 for 48 h, treated with the indicated drugs and measured by the CCK-8 assay (mean ± SD, n=3, two-way ANOVA, **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$).

(C and D) Cell viability of HCC1937 (C) and SUM149PT (D) cells transfected with siNC or siELF3 for 5 days detected by the CCK-8 assay (mean ± SD, n=4, paired two-tailed Student's t test, **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$).

(E and F) Cell viability curve of HCC1937 cells transfected with siNC or siELF3 for 48 h, treated with the indicated drugs and measured by the CCK-8 assay (mean ± SD, n=3, two-way ANOVA, **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$).

(G and H) Correlation of ELF3 expression levels and cell sensitivity of the indicated drugs in the CellMiner (G) and GDSC (H) databases.

(I) Nude mouse tumors of SUM149PT cells infected with shCtrl of shELF3 lentivirus (mean ± SD, unpaired two-tailed Student's t test, *, $p < 0.05$).

Figure 5

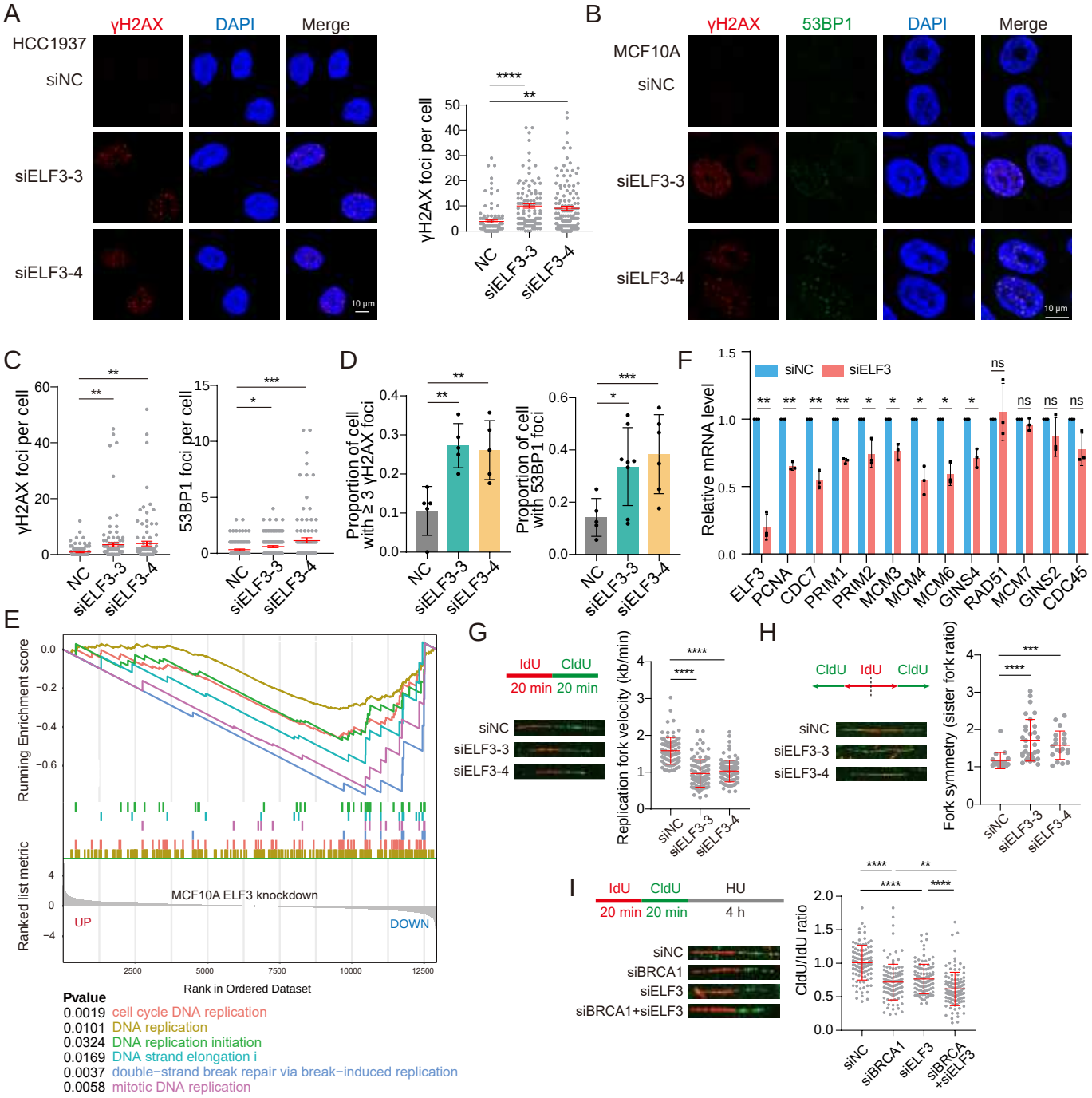

A
γH2AX
DAPI
Merge
HCC1937
siNC
siELF3-3
siELF3-4
10 μm
γH2AX foci per cell
NC
siELF3-3
siELF3-4
B
γH2AX
53BP1
DAPI
Merge
MCF10A
siNC
siELF3-3
siELF3-4
10 μm
C
γH2AX foci per cell
53BP1 foci per cell
D
Proportion of cell with ≥ 3 γH2AX foci
Proportion of cell with 53BP1 foci
F
siNC
siELF3
Relative mRNA level
ELF3
PCNA
CDC7
PRIM1
PRIM2
MCM3
MCM4
MCM6
GINS4
RAD51
MCM7
GINS2
CDC45
E
Running Enrichment score
Ranked list metric
MCF10A ELF3 knockdown
UP
DOWN
Rank in Ordered Dataset
Pvalue
0.0019 cell cycle DNA replication
0.0101 DNA replication
0.0324 DNA replication initiation
0.0169 DNA strand elongation i
0.0037 double−strand break repair via break−induced replication
0.0058 mitotic DNA replication
G
IdU
CldU
20 min 20 min
siNC
siELF3-3
siELF3-4
Replication fork velocity (kb/min)
H
CldU
IdU
CldU
Fork symmetry (sister fork ratio)
I
IdU
CldU
HU
20 min 20 min
4 h
siNC
siBRCA1
siELF3
siBRCA1+siELF3
CldU/IdU ratio

Figure 5. ELF3 helps maintain the stability of DNA replication.

(A) Spontaneous γH2AX foci of HCC1937 cells transfected with siNC or siELF3 for 72 h (mean ± SEM, $n > 100$, unpaired two-tailed Student's t test, **, $p < 0.01$; ****, $p < 0.0001$).

(B and C) γH2AX and 53BP1 foci of MCF10A cells transfected with siNC or siELF3 for 48 h and treated with 2 mM HU for 4 h followed by release for 20 h (mean ± SEM, $n > 100$, unpaired two-tailed Student's t test, *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

(D) Proportions of cells with γH2AX (≥ 3) and 53BP1 foci from (B and C) (mean ± SD, unpaired two-tailed Student's t test, *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

(E) Enrichment plot of DNA replication pathways in MCF10A ELF3 knockdown RNA seq.

(F) RT–qPCR results of MCF10A cells transfected with siNC or siELF3 (#4) for 72 h (mean ± SD, n=3, paired two-tailed Student's t test, *, $p < 0.05$; **, $p < 0.01$; ns, no significance).

(G) Replication fork velocity indicated by DNA fiber assay of MCF10A cells transfected with the indicated siRNA for 72 h (mean ± SD, $n > 80$, unpaired two-tailed Student's t test, ****, $p < 0.0001$).

(H) Replication fork symmetry indicated by the DNA fiber assay of MCF10A cells transfected with the indicated siRNA for 72 h (mean ± SD, Mann–Whitney test, ***, $p < 0.001$; ****, $p < 0.0001$).

(I) Replication fork stability of MCF10A cells transfected with the indicated siRNA for 72 h (mean ± SD, $n > 100$, unpaired two-tailed Student's t test, **, $p < 0.01$; ****, $p < 0.0001$).

Figure 6

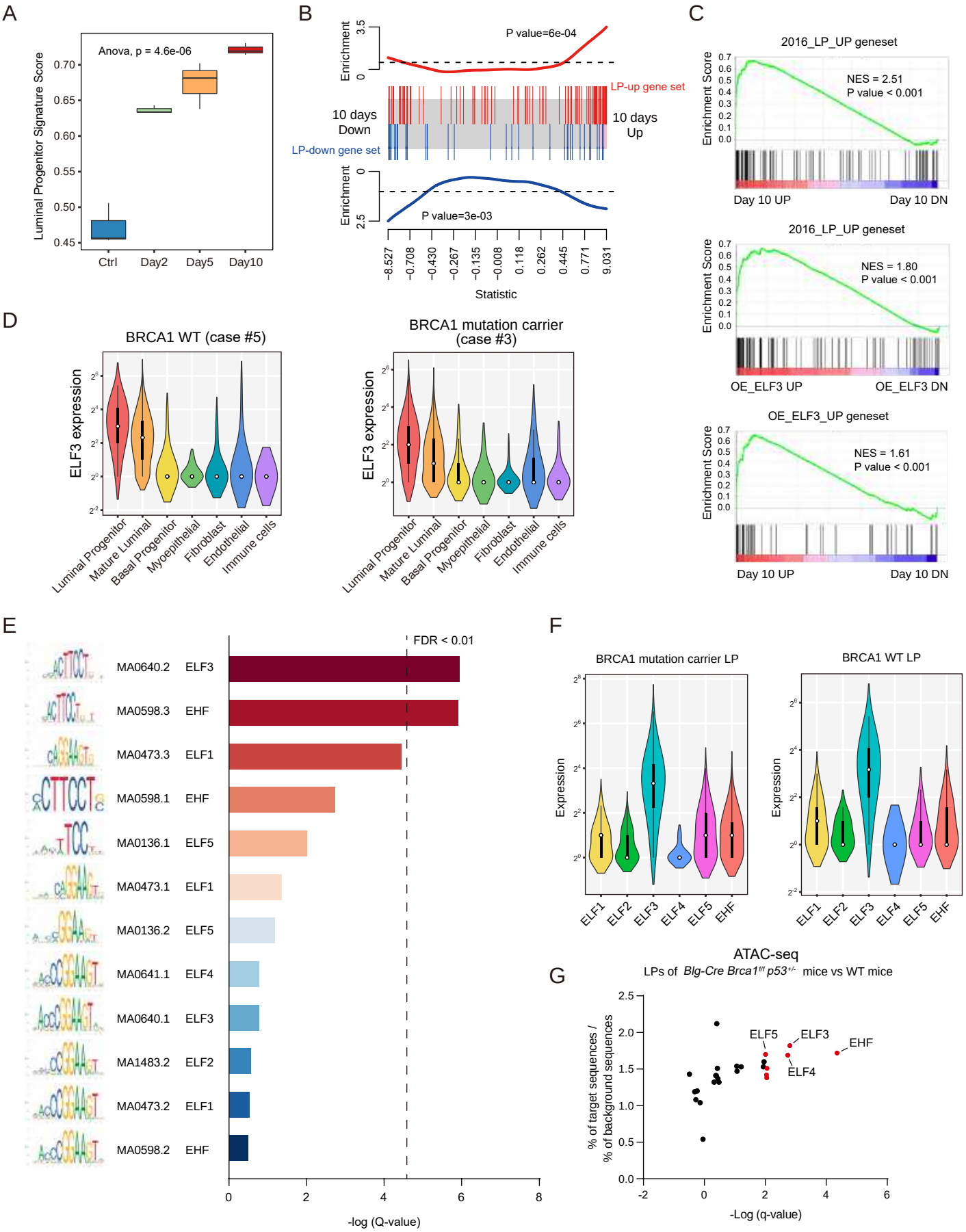

A
Anova, p = 4.6e-06
Luminal Progenitor Signature Score
Ctrl
Day2
Day5
Day10
B
P value=6e-04
Enrichment
LP-up gene set
10 days Down
10 days Up
LP-down gene set
P value=3e-03
Statistic
C
2016_LP_UP geneset
NES = 2.51
P value < 0.001
Enrichment Score
Day 10 UP
Day 10 DN
2016_LP_UP geneset
NES = 1.80
P value < 0.001
OE_ELF3 UP
OE_ELF3 DN
OE_ELF3_UP geneset
NES = 1.61
P value < 0.001
D
BRCA1 WT (case #5)
BRCA1 mutation carrier (case #3)
ELF3 expression
Luminal Progenitor
Mature Luminal
Basal Progenitor
Myoepithelial
Fibroblast
Endothelial
Immune cells
E
FDR < 0.01
MA0640.2 ELF3
MA0598.3 EHF
MA0473.3 ELF1
MA0598.1 EHF
MA0136.1 ELF5
MA0473.1 ELF1
MA0136.2 ELF5
MA0641.1 ELF4
MA0640.1 ELF3
MA1483.2 ELF2
MA0473.2 ELF1
MA0598.2 EHF
-log (Q-value)
F
BRCA1 mutation carrier LP
BRCA1 WT LP
Expression
ELF1
ELF2
ELF3
ELF4
ELF5
EHF
G
ATAC-seq
LPs of Blg-Cre Brca1f/f p53+/- mice vs WT mice
% of target sequences / % of background sequences
ELF5
ELF3
EHF
ELF4
-Log (q-value)

Figure 6. BRCA1 deficiency disturbs the differentiation of LPs.

(A) LP signature score of samples of MCF10A-shBRCA1-Tet-on RNA-seq. The signature score was generated using data from Lim et al. (3).

(B) Barcode plot of LP gene set enrichment in Day 10 vs. Ctrl. LP gene sets were obtained from data from Pellacani et al. (33).

(C) GSEA plots of the transcriptional profile of MCF10A cells with BRCA1 deficiency, MCF10A cells with ELF3 overexpression and normal human LPs. The LP gene set was obtained from the data of Pellacani et al. (33).

(D) Violin plots displaying the distribution of ELF3 expression levels in distinct cell types in normal mammary tissue from noncarriers (left) and BRCA1 mutation carrier (right).

(E) ETS subfamily transcription factor motif enrichment results of LP gene promoters.

(F) Expression levels of ETS subfamily transcription factors in LPs from BRCA1 mutation carrier and noncarrier of the single-cell sequencing data.

(G) ETS subfamily transcription factor enrichment of ATAC-Seq of LPs of BRCA1 deficient mice and WT mice from Bach et al. (37).

Figure 7

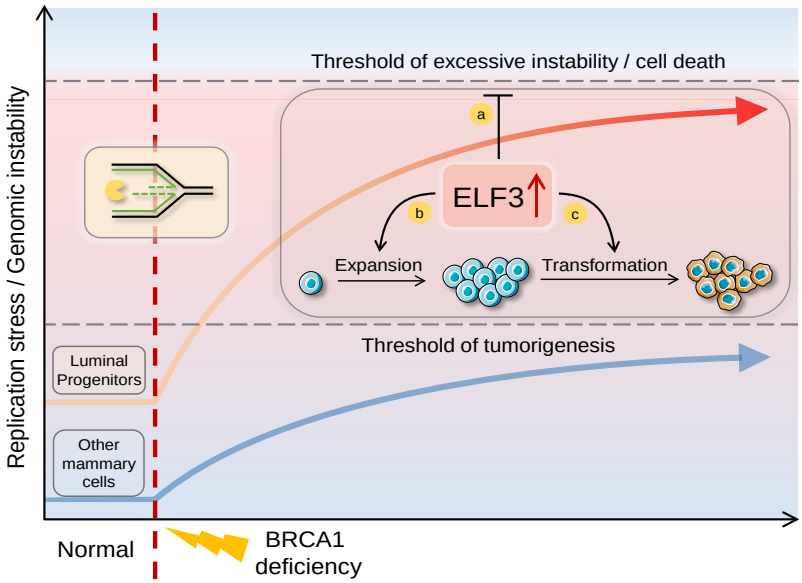


Figure 7. Replication stress inducing ELF3 upregulation promotes BRCA1-deficient breast tumorigenesis. LPs have higher replication stress levels under normal conditions (left) and thus are more likely to exceed the tumorigenesis threshold under BRCA1 deficiency (right). Replication stress and BRCA1 deficiency results in further ELF3 upregulation (top right). ELF3 can suppress excessive replication stress and help LPs maintain moderate levels of genomic instability, which is tolerable for cells to proliferate and enough to fuel cancer evolution (a). In non-LP mammary cells, which exhibit inherently low replicative stress, the ELF3 upregulation may enable cells to endure the increased replicative stress caused by BRCA1 deficiency without leading to malignancy. However, in LP cells, which naturally experience higher levels of replicative stress, this ELF3-mediated mechanism may increase susceptibility to cancerous transformation. In addition, ELF3 upregulation can boost LP gene transcription, leading to LP dedifferentiation and expansion (b). Finally, as an EMT promoter in the mammary tissue, ELF3 can accelerate the transformation process from LPs to malignant cells (c).

Materials and Methods

Cell culture

MCF10A cells were cultured in DMEM/F12 supplemented with 5% horse serum, 1% penicillin/streptomycin, 20 ng/mL EGF, 10 μg/mL insulin, 0.5 μg /mL hydrocortisone and 100 ng/mL cholera toxin. For MCF10A-shBRCA1-Tet-on cells, except for the replacement of horse serum with 8% Tet-on system approved certified fetal bovine serum, all other components remain the same. HEK293T and SUM149PT cells were cultured in DMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. HCC1937 cells were cultured in RPMI1640 supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin.

Reagents and plasmids

Reagents (antibodies and chemicals) used in this study were listed in Table S3. For the construction of Tet-on-shBRCA1-hygro plasmid, shBRCA1 sequence was subcloned into the Tet-on-pLKO-hygro vector (SinaSun). For ELF3 knockdown by lentivirus in SUM149PT, shELF3 and shCtrl sequence was subcloned into the pLKO.1-puro vector. The shRNA sequences used in this study are listed in Table S4. For ELF3 overexpression by lentivirus in MCF10A, human ELF3 was subcloned into the pCDH-CMV-MCS-EF1-puro vector, which was a kind gift from Yin lab. Myc-tagged human E2F6 was subcloned into the pDEST-N-Myc vector. PCR primers used in this study were listed in Table S5.

Transfection of siRNA and infection of shRNA

All the siRNAs were from GenePharma. For siRNA transfection, cells were seeded in 6-well plate or 35 mm^2 dish at the day before transfection. 100 μL OPTI-MEM, 40 pmol siRNA and 3 μL Lipofectamine RNAiMAX reagent were mixed and added to the cells. After 48-72 h, cells were collected for subsequent experiments. The siRNA sequences used are listed in Table S6. For shRNA infection, lentivirus was generated by co-transfection of the lentiviral vector (Tet-pLKO.1-shBRCA1, pLKO.1-shCtrl and pLKO.1-shELF3) with envelope plasmids pMD2.G and psPAX2 in HEK293T cells. Lentivirus was collected after 48 h. Cells were infected with lentivirus with polybrene. After 24 h, cells were re-infected and harvested at 72 h for subsequent experiments.

Quantitative PCR (RT-qPCR)

Total RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's guidelines. cDNA was synthesized by reverse transcription PCR using Hifair III 1st Strand cDNA Synthesis SuperMix kit (Yeasen). qPCR was performed using SYBR Green Mix (Yeasen) on ABI StepOnePlus qPCR instrument (ThermoFisher). Each group was performed in triplicate. The relative expression mRNA levels were calculated by relative quantitative method (ΔΔCt). The primers used are listed in Table S7.

Immunofluorescence

Cells were seeded on coverslips and treated as indicated. Cells were than fixed in cold 4% paraformaldehyde for 15 min and permeabilized in 0.25% Triton X-100 solution for 5 min at room temperature. After blocking with 2% BSA for 10 min, the primary antibodies were added onto the coverslips, and then incubated for 1 h at room temperature, followed by secondary antibody incubation for 30 min. Finally, the coverslips were stained with DAPI for 5 min, and images were acquired using the ZEN software.

Immunohistochemistry

All tissue microarray chips were purchased from Shanghai Outdo Biotech Co.Ltd. IHC was performed as described previously, except that the following antibodies were used. The anti-ELF3 rabbit antibody were purchased from SIGMA. After incubation with primary antibody, the HRP-conjugated rabbit secondary antibody (ZSGB-Bio) was applied and IHC slides were observed with Olympus BX51 microscope and Olympus DP73 CCD photographic system.

Colony formation assay

MCF10A-shBRCA1-Tet-on cells were seeded in 6 mm^2 plates. 1 μg/mL DOX was added for the DOX group during the whole process of cell growth to knockdown BRCA1 expression. After 14 days, the cells were washed with phosphate-buffered saline (PBS) and stained with 0.1% Coomassie brilliant blue in 10% ethanol for 30 minutes at room temperature. The stained dishes were washed with water, and the colonies were counted.

DNA fiber

The DNA fiber assay was performed as described previously with slight modifications (*42*). Briefly, cells transfected with siRNA were labeled by 25 μM IdU (Yuanye Bio-Technology) for 20 min and subsequently 200 μM CldU (Abcam) for 20 min. Then cells were treated with 4 mM HU for 4 h for replication fork

stability assay or collected for replication fork speed and symmetry assay. Cells were resuspended in lysis buffer (200 mM Tris-HCl, pH 7.5, 50 mM EDTA and 0.5% SDS) to extract DNA. Cell lysates were then dripped onto a glass slide and DNA fibers were stretched along the slide after 3 minutes incubation. The slides were air-dried and fixed in 3:1 methanol/acetic acid solution for 10 min, followed by 2.5 M HCl treatment for 1 h, then blocked in 3% BSA PBST buffer for 20 min. DNA fibers were incubated with anti-BrdU antibody (Abcam) binding CldU and anti-BrdU antibody (BD) binding IdU at 37°C for 1.5 h. After washing with PBST, the slides were incubated with Alexa Fluor 488- and Alexa Fluor 594-conjugated secondary antibody (Invitrogen) for 45 minutes. Finally, the slides were mounted in mounting reagent (Solarbio) and DNA fibers images were acquired using ZEN software.

Comet assay

The comet assay was performed as described previously with slight modifications (*43*). Briefly, cells were collected by trypsin digestion and resuspended in PBS. 0.5% normal melting agarose (NMA) was added to glass slides and quickly covered with coverslips. The slides were placed at 4 °C for 10 min. The coverslips were removed carefully. The cell suspensions were mixed gently with equal volume of 1% low melting agarose (LMA) and quickly added onto the solidified NMA, quickly covered with coverslips. The slides were placed at 4 °C for 10 min. After removing the coverslips, the slides were incubated overnight in the fresh cell lysis buffer at 4 °C overnight. Then the slides were rinsed with ddH_2O and electrophoresis at 4°C for 30 minutes at 20 V in neutral electrophoresis buffer. Then the slides were stained with 5 μg/ml of propidium iodide for 10 min. The comet was observed using a fluorescence microscope (Zeiss AxioCam 503 color). The pictures were analyzed using CASP software (http://casplab.com/).

Cell survival (CCK-8) assay

Cell survival assay was performed using CCK-8 kits (meilunbio, MA0218-3) according to the manufacturer's protocols. Cells were seeded in 96-well plates. After 12 h, medium containing indicated concentration of drugs was changed and cells were incubated for 3-5 days. DMEM containing 0.1% DMSO was used as the control group. CCK-8 solution was diluted in 10% concentration and added to wells, and the cells were incubated for 45 min-1.5 h. DMEM containing 10% CCK-8 was used as the blank group. The absorbance at 450 nm was detected by a microplate reader. Cell survival was calculated according to the formula: (ODdrug – ODblank) /(ODcontrol – ODblank).

RNA sequencing

Total RNA of cells was extracted with Trizol reagent. RNA sequencing was performed by Novogene (China). The resulting data was mapped to human reference genome (hg38) and gene counts were calculated for each sample using STAR (v2.6.1) (*44*) with the parameter "--quantMode GeneCounts". To identify differentially expressed genes (DEGs), genes with low expression levels (average counts per million reads < 0.05) were filtered out. DESeq2 was employed to detect DEGs with |log2 fold change| > 2 and FDR < 0.001. The raw RNA sequencing data have been deposited into GEO under the accession number GSExxxxxx. PCA analysis was performed using the "prcomp" function of stats (v3.6.1) R package.

Analysis of TCGA breast cancer and METABRIC data

TCGA (https://www.cancer.gov/tcga) breast cancer and METABRIC (*19*) data were downloaded using the cgdsr R package. For TCGA data, RSEM (batch normalized from Illumina HiSeq_RNASeqV2) was used as the expression levels of the genes. For METABRIC data, mRNA expression log intensity levels (Illumina Human v3 microarray) were used as the expression levels of the genes. BRCA1 associated cancers were identified as cancers with BRCA1 mutations, heterozygous loss or homozygous deletion. Pearson correlation coefficient between ELF3 expression and BRCA1 expression was calculated. Subtypes and IntClust types of breast cancers were identified according to the clinical data of TCGA and METABRIC databases.

Chip-Seq analysis of ENCODE databases

The Chip-Seq data were downloaded from the ENCODE database (https://www.encodeproject.org/) (*45, 46*) with the following identifiers: ENCFF858GLM, ENCFF692OYJ, ENCFF384CPN, ENCFF342GNN, ENCFF352QVM, and ENCFF437NQS. Images were generated using IGV (v2.7.2).

Drug sensitivity analysis of Cell Miner and GDSC data

Cell line drug sensitivity data and mRNA expression data were downloaded from the CellMiner (*47, 48*) and GDSC (*49*) databases. Pearson correlation coefficient between ELF3 expression and drug sensitivity was calculated for the indicated drugs.

Replication stress score and luminal progenitor score analysis

For replication stress score calculation, human mammary subgroups gene expression matrixes were obtained from the datasets of Lim et al. (*3*) (GSE16997). Replication stress associated pathways were

selected referring to Dreyer et al. (*13*). and gene lists were downloaded from the GSEA website (www.gsea-msigdb.org). For each pathway, replication stress scores were calculated according to the formula below, where e_g is the expression of the gene in different mammary subgroups, and n_g is the number of genes in this pathway.

$$s=\sum_g e_g/n_g$$

For luminal progenitor score calculation, the LP signature gene set with average log-fold change was obtained from data of Lim et al. (*3*), and log CPM matrix of MCF10A-shBRCA1-Tet-on DOX 10 days vs Ctrl RNA-seq data was used. For each sample, LP scores were calculated according to the formula below, which refers to Lim et al. (*3*), where x_g is the log-fold-change for the gene from LP gene set, and y_g is the log cpm for the same gene in BRCA1 knockdown RNA-seq.

$$s=\sum_g x_g y_g/\sum_g |x_g|$$

GSEA analysis and barcode plot

GSEA analysis was performed using GSEA software (v4.1.0) (*50, 51*) and clusterProfiler (v3.14.3) (*52*). For the customized gene set enrichment analysis of the transcriptional profiles of BRCA1 KD, ELF3 OE and LP, BRCA1 KD data were from MCF10A-shBRCA1-Tet-on DOX 10 days vs Ctrl RNA-Seq data and ELF3 OE data were from MCF10A OE ELF3 vs Ctrl RNA-seq data. LP gene set was obtained from 2016 cell reports, and OE ELF3 gene sets were generated using MCF10A OE ELF3 vs Ctrl RNA-seq data, where the log-fold-change cutoff was 1.3 and P value cutoff was 0.01. For the barcode plot, LP gene sets were obtained from data of Pellacani et al. (*29*), and log-fold-change of all genes from MCF10A-shBRCA1-Tet-on DOX 10 days vs Ctrl RNA-seq data was used for the enrichment analysis. P value was calculated using the “roast” function from limma (v3.42.2) (*53*).

Motif enrichment analysis

Motif matrixes of transcription factors were downloaded from the JASPAR database (*54*). LP gene list was obtained according to the data of Pellacani et al. (*29*) and LP genes are identified to significantly upregulated in LPs compared with all three other subgroups (BC, LC and SC). Promoter sequences of LP genes were extracted from the genome from 3000 bp upstream to 200 bp downstream of the TSS sites. Motif enrichment analysis was performed using SEA (Simple Enrichment Analysis) of the MEME Suite (*55*).

Single-cell sequencing data analysis

Single-cell RNA-seq (scRNA-seq) expression profiles of breast cancer tissue and normal mammary tissue from four BRCA1 mutation carriers, as well as normal mammary tissue from three non-carriers were generated using the 10x Genomics Chromium platform and NovaSeq 6000 platform. The detail information of sample processing, scRNA-seq data processing and quality control were described in our previous study (*8*). Briefly, scRNA-seq data analyses was completed utilizing the 10× Genomics Cell Ranger software suite. Principal component analysis (PCA) (twenty-five principal components) and k-means algorithm (k was set as 9.0) were used to cluster cells. We used a series of cell specific markers to identify the cell type of each cluster, as described in our previous study1. Among these, luminal progenitors cluster was identified based on the expression of KRT8/18+, CD24+, GABRP+, KIT+, and ALDH1A3+. Triple negative tumor cell population was identified based on tumor sample specific cluster, ER/PR/ HER2 expression coincident with the clinical records and PAM50 subtypes analysis. Replication stress score of each cell was calculated based on the sum expression level of genes involving in replication stress-related pathways referring to Dreyer et al. (*13*). In this study, we selected luminal lineage cells and tumor cells from a BRCA1-mutated triple-negative breast cancer patient to construct tumorigenesis trajectories with the R package Monocle 2 with standard settings.

Animal experiments

5×10^6 SUM149PT tumor cells were subcutaneously injected into the flanks of female BALB/c nude mice. Tumor growth was measured using a caliper every three days, and the tumor volume was calculated using the formula: tumor volume = length $\times$width^2/2.

Statistics Analysis

Statistical analysis was performed using GraphPad Prism (GraphPad Software). The tests used included the paired and unpaired two-tailed Student's t test, the Mann-Whitney U test, one-way and two-way ANOVA, and Pearson's correlation statistical test. A P value of less than 0.05 was considered statistically significant. Details can be found in figure legends.

Study approval

The normal human breast cancer tissue and breast cancer samples used for single-cell RNA-seq in this study were collected from Peking University Cancer Hospital. This study was carried out in accordance

with the ethical principles of the Declaration of Helsinki and was approved by the Research and Ethics Committee of Peking University Cancer Hospital. Written informed consent was obtained from all participants prior to participation. Animal experiments in this study were performed in accordance with the Guidelines of Peking University Animal Care and Use Committee.

Data availability

All the data supporting the findings of this study are available within the article and supplementary information file and from the corresponding author upon reasonable request. The RNA-sequencing raw data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (PRJCA010883) that are publicly accessible at https://ngdc.cncb.ac.cn.

References

1. D. Ford, D. F. Easton, D. T. Bishop, S. A. Narod, D. E. Goldgar, Risks of cancer in BRCA1-mutation carriers. *The Lancet* **343**, 692-695 (1994).
2. B. Friedenson, BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. *MedGenMed : Medscape general medicine* **7**, 60-60 (2005).
3. E. Lim *et al.*, Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907-913 (2009).
4. G. Molyneux *et al.*, BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* **7**, 403-417 (2010).
5. E. Nolan *et al.*, RANK ligand as a potential target for breast cancer prevention in BRCA1-mutation carriers. *Nat. Med.* **22**, 933-939 (2016).
6. H. Wang *et al.*, Inadequate DNA Damage Repair Promotes Mammary Transdifferentiation, Leading to BRCA1 Breast Cancer. *Cell* **178**, 135-151.e119 (2019).
7. T. A. Proia *et al.*, Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* **8**, 149-163 (2011).
8. L. Hu *et al.*, Single-Cell RNA Sequencing Reveals the Cellular Origin and Evolution of Breast Cancer in BRCA1 Mutation Carriers. *Cancer Res.* **81**, 2600-2611 (2021).
9. J. R. Oliver, R. Kushwah, J. Hu, Multiple roles of the epithelium-specific ETS transcription factor, ESE-1, in development and disease. *Lab. Invest.* **92**, 320-330 (2012).
10. C. H. Chang *et al.*, ESX: a structurally unique Ets overexpressed early during human breast tumorigenesis. *Oncogene* **14**, 1617-1622 (1997).
11. P. J. Schedin *et al.*, ESX induces transformation and functional epithelial to mesenchymal transition in MCF-12A mammary epithelial cells. *Oncogene* **23**, 1766-1779 (2004).
12. R. M. Neve, B. Ylstra, C. H. Chang, D. G. Albertson, C. C. Benz, ErbB2 activation of ESX gene expression. *Oncogene* **21**, 3934-3938 (2002).
13. S. B. Dreyer *et al.*, Targeting DNA Damage Response and Replication Stress in Pancreatic Cancer. *Gastroenterology* **160**, 362-377.e313 (2021).
14. N. Takahashi *et al.*, Replication stress defines distinct molecular subtypes across cancers. *Cancer Res Commun* **2**, 503-517 (2022).
15. S. Pathania *et al.*, BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nat. Commun.* **5**, 5496 (2014).
16. L. Zheng, S. Li, T. G. Boyer, W.-H. Lee, Lessons learned from BRCA1 and BRCA2. *Oncogene* **19**, 6159-6175 (2000).
17. L. Kumar, E. F. M, Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**, 5-7 (2007).
18. J. D. Prescott, K. S. Koto, M. Singh, A. Gutierrez-Hartmann, The ETS transcription factor ESE-1 transforms MCF-12A human mammary epithelial cells via a novel cytoplasmic mechanism. *Mol. Cell. Biol.* **24**, 5548-5564 (2004).
19. C. Curtis *et al.*, The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352 (2012).
20. K. L. Eckel, J. J. Tentler, G. J. Cappetta, S. E. Diamond, A. Gutierrez-Hartmann, The epithelial-

specific ETS transcription factor ESX/ESE-1/Elf-3 modulates breast cancer-associated gene expression. *DNA Cell Biol.* **22**, 79-94 (2003).

21. J.-P. Coppe *et al.*, ERBB receptor regulation of ESX/ELF3 promotes invasion in breast epithelial cells. *The Open Cancer Journal* **3**, 89-100 (2010).
22. P. Oettgen *et al.*, The novel epithelial-specific Ets transcription factor gene ESX maps to human chromosome 1q32.1. *Genomics* **45**, 456-457 (1997).
23. C. Bertoli, J. M. Skotheim, R. A. de Bruin, Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* **14**, 518-528 (2013).
24. 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).
25. D. Tkocz *et al.*, BRCA1 and GATA3 corepress FOXC1 to inhibit the pathogenesis of basal-like breast cancers. *Oncogene* **31**, 3667-3678 (2012).
26. R. D. Kennedy *et al.*, BRCA1 and c-Myc associate to transcriptionally repress psoriasin, a DNA damage-inducible gene. *Cancer Res.* **65**, 10265-10272 (2005).
27. S. Pathania *et al.*, BRCA1 Is Required for Postreplication Repair after UV-Induced DNA Damage. *Mol. Cell* **44**, 235-251 (2011).
28. S. Liu *et al.*, BRCA1 regulates human mammary stem/progenitor cell fate. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1680-1685 (2008).
29. D. Pellacani *et al.*, Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. *Cell reports* **17**, 2060-2074 (2016).
30. S. R. Oakes *et al.*, The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev.* **22**, 581-586 (2008).
31. R. Chakrabarti *et al.*, Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling. *Stem Cells* **30**, 1496-1508 (2012).
32. K. Bach *et al.*, Time-resolved single-cell analysis of Brca1 associated mammary tumourigenesis reveals aberrant differentiation of luminal progenitors. *Nat. Commun.* **12**, 1502 (2021).
33. Y. J. Shiah *et al.*, A Progesterone-CXCR4 Axis Controls Mammary Progenitor Cell Fate in the Adult Gland. *Stem cell reports* **4**, 313-322 (2015).
34. A. E. Casey *et al.*, Mammary molecular portraits reveal lineage-specific features and progenitor cell vulnerabilities. *J. Cell Biol.* **217**, 2951-2974 (2018).
35. H. Kim *et al.*, Mammary lineage dictates homologous recombination repair and PARP inhibitor vulnerability. *bioRxiv*, 2021.2005.2014.444217 (2021).
36. M. Macheret, T. D. Halazonetis, DNA replication stress as a hallmark of cancer. *Annu. Rev. Pathol.* **10**, 425-448 (2015).
37. M. Sedic *et al.*, Haploinsufficiency for BRCA1 leads to cell-type-specific genomic instability and premature senescence. *Nat Commun* **6**, 7505 (2015).
38. N. Longoni *et al.*, ETS transcription factor ESE1/ELF3 orchestrates a positive feedback loop that constitutively activates NF-κB and drives prostate cancer progression. *Cancer Res.* **73**, 4533-4547 (2013).
39. K. S. S. Enfield *et al.*, Epithelial tumor suppressor ELF3 is a lineage-specific amplified oncogene in lung adenocarcinoma. *Nat. Commun.* **10**, 5438 (2019).

40. S. Iwai *et al.*, ESE-1 inhibits the invasion of oral squamous cell carcinoma in conjunction with MMP-9 suppression. *Oral Dis.* **14**, 144-149 (2008).
41. C. G. A. R. Network, Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315-322 (2014).
42. J. Nieminuszczy, R. A. Schwab, W. Niedzwiedz, The DNA fibre technique - tracking helicases at work. *Methods* **108**, 92-98 (2016).
43. E. Rojas, M. C. Lopez, M. Valverde, Single cell gel electrophoresis assay: methodology and applications. *J. Chromatogr. B Biomed. Sci. Appl.* **722**, 225-254 (1999).
44. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
45. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
46. C. A. Davis *et al.*, The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794-d801 (2018).
47. U. T. Shankavaram *et al.*, CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics* **10**, 277 (2009).
48. W. C. Reinhold *et al.*, CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.* **72**, 3499-3511 (2012).
49. W. Yang *et al.*, Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955-961 (2013).
50. V. K. Mootha *et al.*, PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267-273 (2003).
51. A. Subramanian *et al.*, 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).
52. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).
53. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
54. J. A. Castro-Mondragon *et al.*, JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165-d173 (2022).
55. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39-49 (2015).

Acknowledgments

We thank Dr. Jianming Zeng (University of Macau), Xiaojie Sun and all the members of the bioinformatics team, biotrainee, for generously sharing their experience and codes. We thank Ence Yang, Changyu Tao and Zelin Liu for the kind sharing of computing resources and guiding on the RNA-seq data analytical methodology. We thank Lim et al., Pellacani et al. and Bach et al. for their outstanding contribution on BRCA1 associated breast tumorigenesis research and mammary cell populations sequencing data. We

thank the ENCODE Consortium and the ENCODE production laboratories for generating the ChIP-seq datasets.

Authors' Contributions

J.Zhou., X.A.Z. and J.W. designed the research. J.Zhou. performed most of the experiments and data analysis. L.H. performed the single cell sequencing data analysis. X.A.Z. performed the mouse experiments. Y.M. contributed to the RT-qPCR experiments. J.Zhan. performed the immunohistochemistry analysis. Z.X. contributed to experiment ideas and data analysis. M.Z. and Q.S. assisted the experiments. Y.X. provided the single cell sequencing data. J.Zhou. wrote the manuscript. X.A.Z. and J.W. reviewed and edited the manuscript.

Funding Statement

The National Science Fund for Distinguished Young Scholars (82125031)

National Natural Science Foundation of China (82230089)

National Key R&D Program of China (2022YFA1302800)

Young Scientists Fund of the National Natural Science Foundation of China (82102997)

The China Postdoctoral Science Foundation (2024M750126)

Postdoctoral Fellowship Program (Grade B) of China Postdoctoral Science Foundation (GZB20240037)

Figure S1

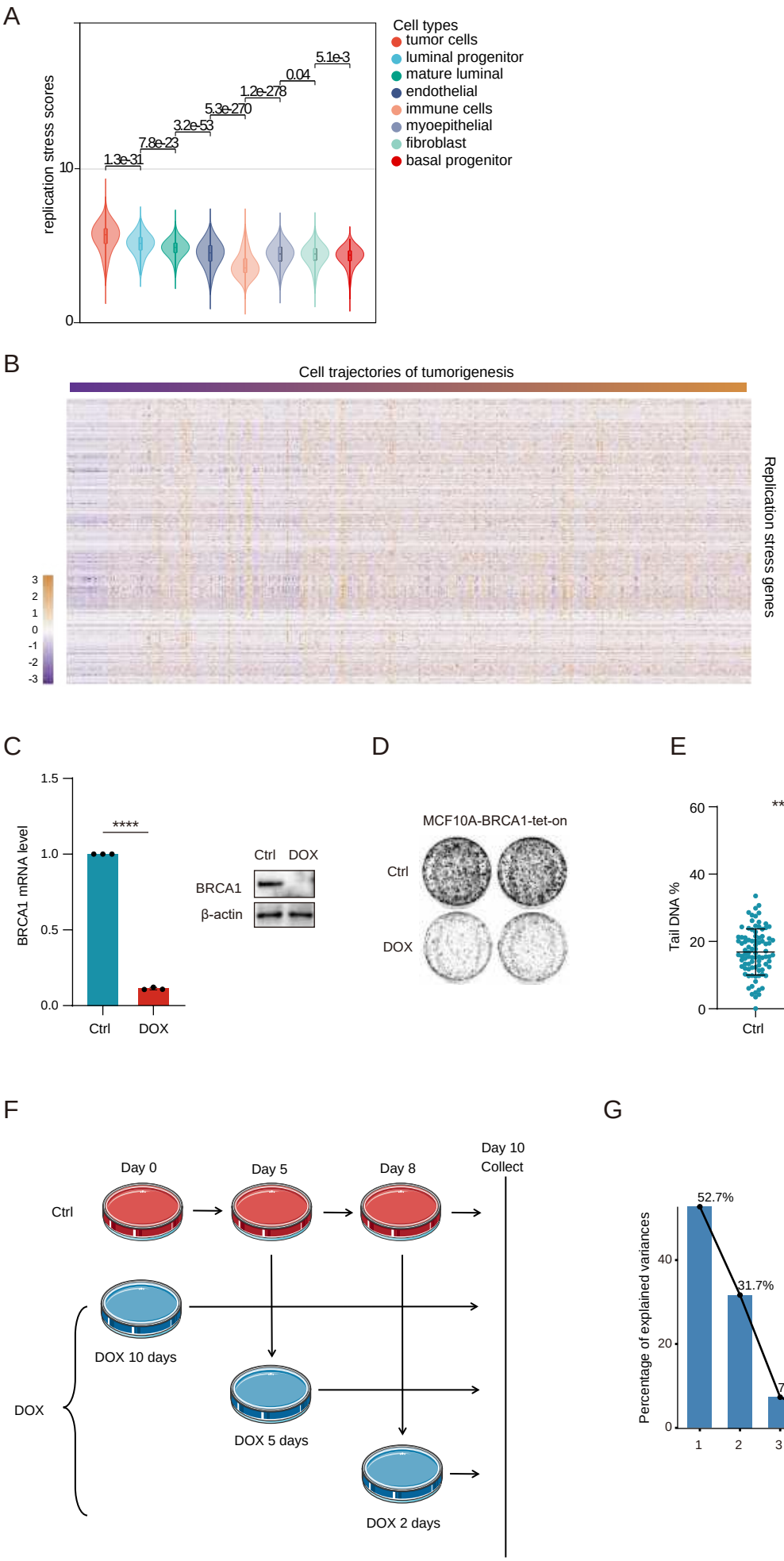

A
replication stress scores
10
0
1.3e-31
7.8e-23
3.2e-53
5.3e-270
1.2e-278
0.04
5.1e-3
Cell types
tumor cells
luminal progenitor
mature luminal
endothelial
immune cells
myoepithelial
fibroblast
basal progenitor
B
Cell trajectories of tumorigenesis
Replication stress genes
3
2
1
0
-1
-2
-3
C
BRCA1 mRNA level
1.5
1.0
0.5
0.0
****
Ctrl
DOX
Ctrl DOX
BRCA1
β-actin
D
MCF10A-BRCA1-tet-on
Ctrl
DOX
E
Tail DNA %
60
40
20
0
****
Ctrl
DOX
Ctrl
DOX
F
Day 0
Day 5
Day 8
Day 10
Collect
Ctrl
DOX 10 days
DOX
DOX 5 days
DOX 2 days
G
Percentage of explained variances
52.7%
31.7%
7.3%
3%
1.1%
0.9%
0.8%
0.7%
0.7%
0.6%
40
20
0
1
2
3
4
5
6
7
8
9
10
Dimensions

Figure S1. LPs have higher levels of replication stress compared with other mammary cell populations.

(A) Replication scores of various cell types in single cell sequencing. (Mann-Whitney U test)

(B) Heatmap showing the expression pattern of replication stress-related genes in BRCA1-mutated triple-negative breast cancer cells. The cancer cells were ordered based on Monocle-generated cell trajectories of tumorigenesis (Figure 1C).

(C) RT–qPCR (mean ± SD, n=3, paired two-tailed Student's t test, ****, $p < 0.0001$) and Western blot assay results of MCF10A-shBRCA1-Tet-on cells with or without 1 μg/mL DOX treatment for 48 h.

(D) Colony formation assay results of MCF10A-shBRCA1-Tet-on cells treated with or without 1 μg/mL DOX for 14 days.

(E) Neutral comet assay results of MCF10A-shBRCA1-Tet-on cells with or without 1 μ g/mL DOX treatment for 48 h (mean ± SD, unpaired two-tailed Student's t test, ****, $p < 0.0001$).

(F) Schematic of the experimental procedure of MCF10A-shBRCA1-Tet-on RNA-Seq.

(G) Scatter plot of the PCA results. The horizontal axis represents different principal components.

Figure S2

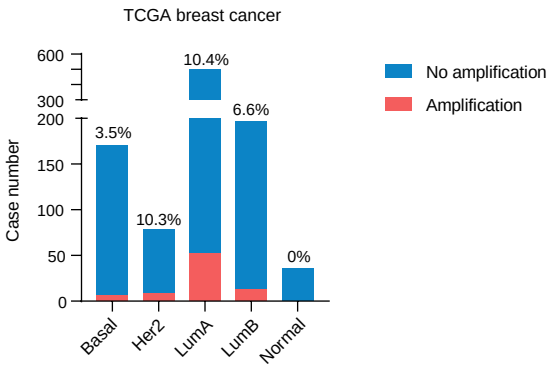

TCGA breast cancer
Case number
600
300
200
150
100
50
0
3.5%
10.3%
10.4%
6.6%
0%
Basal
Her2
LumA
LumB
Normal
No amplification
Amplification

Figure S2. ELF3 expression upregulation in basal-like breast cancers is not due to amplification.

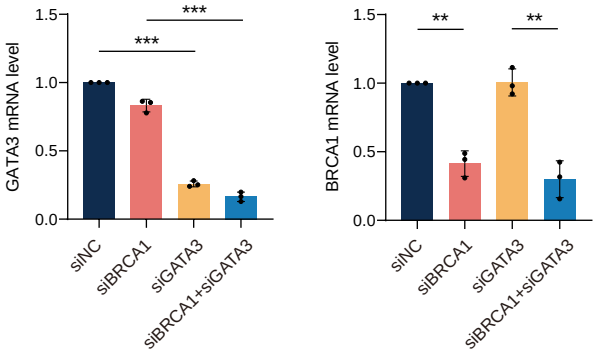
Proportions of ELF3 amplification in different molecular subtypes of TCGA breast cancers. The numbers represent the percentage of ELF3 amplification in each subtype.

Figure S3

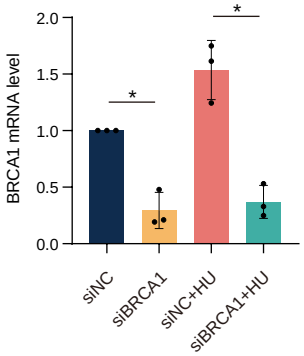
A

Name	Score	Relative score	Strand	Predicted sequence	Motif
MA0024.2.E2F1	12.77371	0.960198612	-	GTGGCGGGAAC	
MA0037.1.GATA3	7.126715	0.947082087	-	TGATTG	
MA0037.3.GATA3	8.782075	0.923144513	+	AGATTAGA	
MA0037.1.GATA3	6.422963	0.917558655	+	AGATAT	
MA0037.1.GATA3	6.215259	0.908845139	-	GGATAG	
MA0147.3.MYC	10.901	0.895428972	+	CGCCACCTGCCC	
MA0147.3.MYC	9.150411	0.858881602	+	TGCCCCGTGCCC	

B



C



D

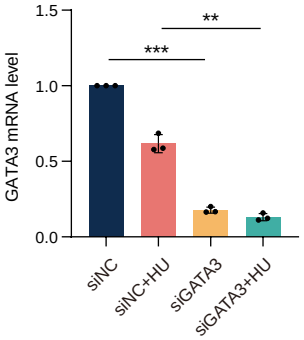


Figure S3. The replication stress response pathway is activated in BRCA1-associated breast cancers.

(A) ATR-Chk1-E2F axis gene expression levels in BRCA1-associated breast cancers and BRCA1-nonassociated breast cancers in the TCGA and METABRIC databases (Mann–Whitney U test). (B) ATR-Chk1-E2F axis gene expression levels of different molecular subtypes of breast cancers in the TCGA and METABRIC databases (one-way ANOVA). (C) RT–qPCR knockdown results related to Figure 3H. (D) Scores of E2F1, GATA3 and MYC motif enrichment in the ELF3 promoter sequence. (E) RT–qPCR knockdown results related to Figure 3K. (F) RT–qPCR knockdown results related to Figure 3L. In all panels of RT–qPCR results, data are presented as the mean ± SD, n=3, *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$, by paired two-tailed Student's t test.

Figure S4

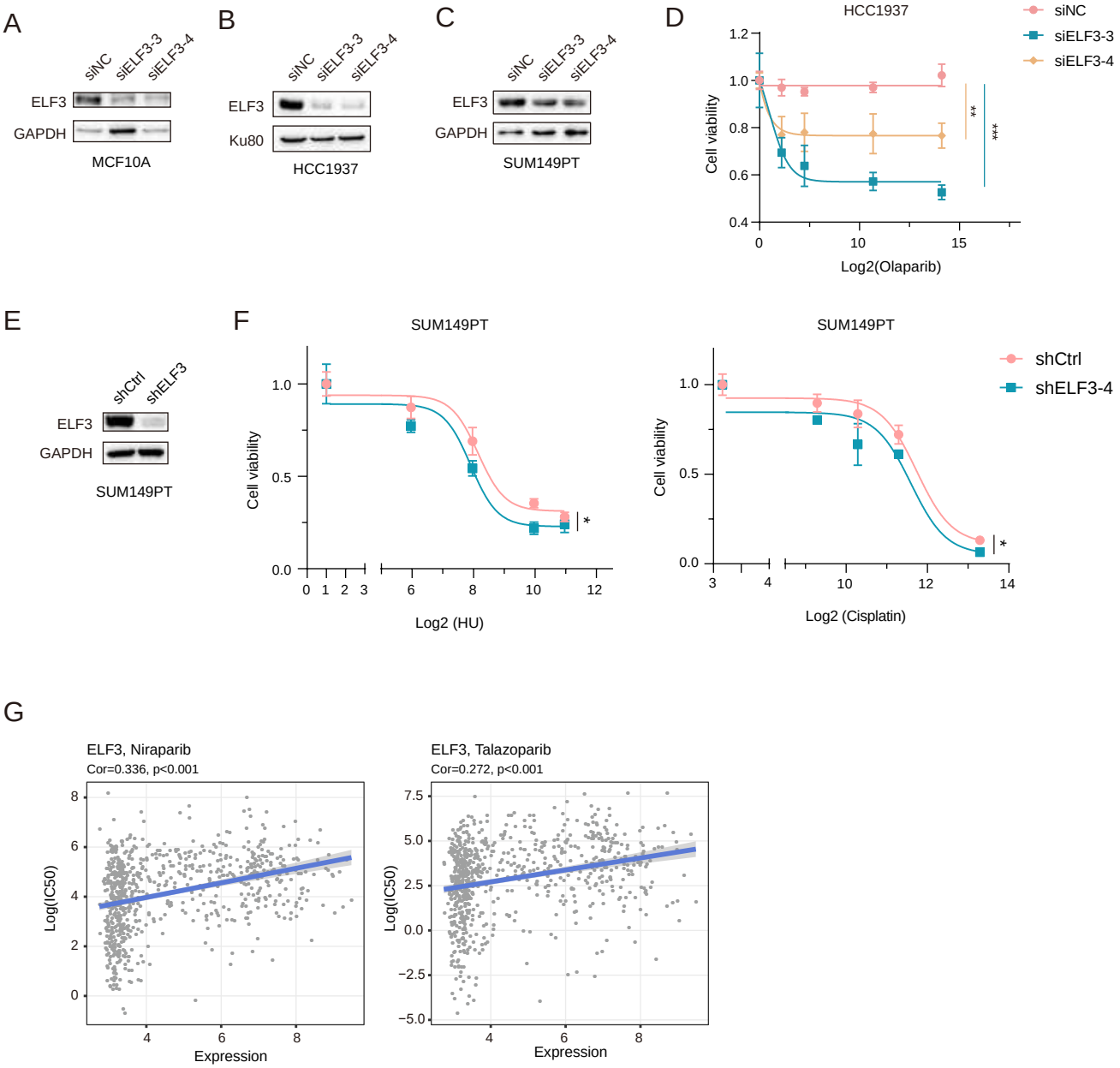

A
siNC
siELF3-3
siELF3-4
ELF3
GAPDH
MCF10A
B
siNC
siELF3-3
siELF3-4
ELF3
Ku80
HCC1937
C
siNC
siELF3-3
siELF3-4
ELF3
GAPDH
SUM149PT
D
HCC1937
siNC
siELF3-3
siELF3-4
Cell viability
Log2(Olaparib)
**
***
E
shCtrl
shELF3
ELF3
GAPDH
SUM149PT
F
SUM149PT
Cell viability
Log2 (HU)
*
SUM149PT
shCtrl
shELF3-4
Cell viability
Log2 (Cisplatin)
*
G
ELF3, Niraparib
Cor=0.336, p<0.001
Log(IC50)
Expression
ELF3, Talazoparib
Cor=0.272, p<0.001
Log(IC50)
Expression

Figure S4. ELF3 helps suppress excessive genomic instability.

(A) Western blot knockdown results related to Figure 4A and B. (B and C) Western blot knockdown results related to Figure 4C and D. (D and E) Cell viability curve of SUM149PT cells infected with shCtrl or shELF3 lentivirus and treated with the indicated concentration of drugs (HU: μM; cisplatin: nM) and detected by the CCK-8 assay (mean ± SD, n=3, paired two-tailed Student's t test, *, $p < 0.05$). (F) Cell viability curve of HCC1937 cells transfected with siNC or siELF3, treated with indicated concentration of olaparib (nM) and measured by the CCK-8 assay (mean ± SD, n=3, two-way ANOVA, **, $p < 0.01$; ***, $p < 0.001$). (G) Correlation of ELF3 expression levels and cell sensitivity to PARPi in the GDSC database.

Figure S5

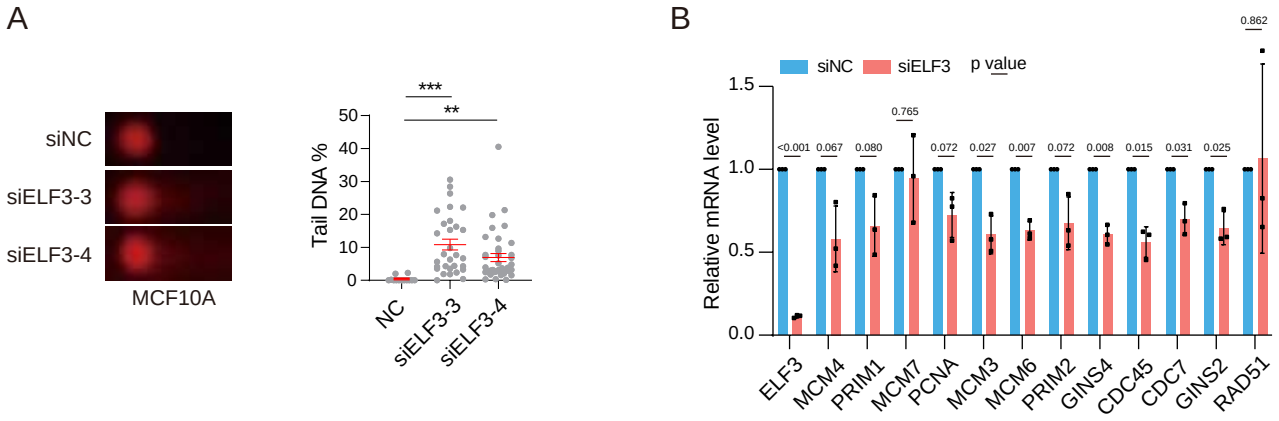


Figure S5. ELF3 helps maintain the stability of DNA replication.
(A) Neutral comet assay results of MCF10A cells transfected with siNC or siELF3 for 72 h (mean ± SEM, unpaired two-tailed Student's t test, **, $p < 0.01$; ***, $p < 0.001$).
(B) RT–qPCR results of MCF10A cells transfected with siNC or siELF3 (#5) for 72 h (mean ± SD, n=3, paired two-tailed Student's t test).

Figure S6

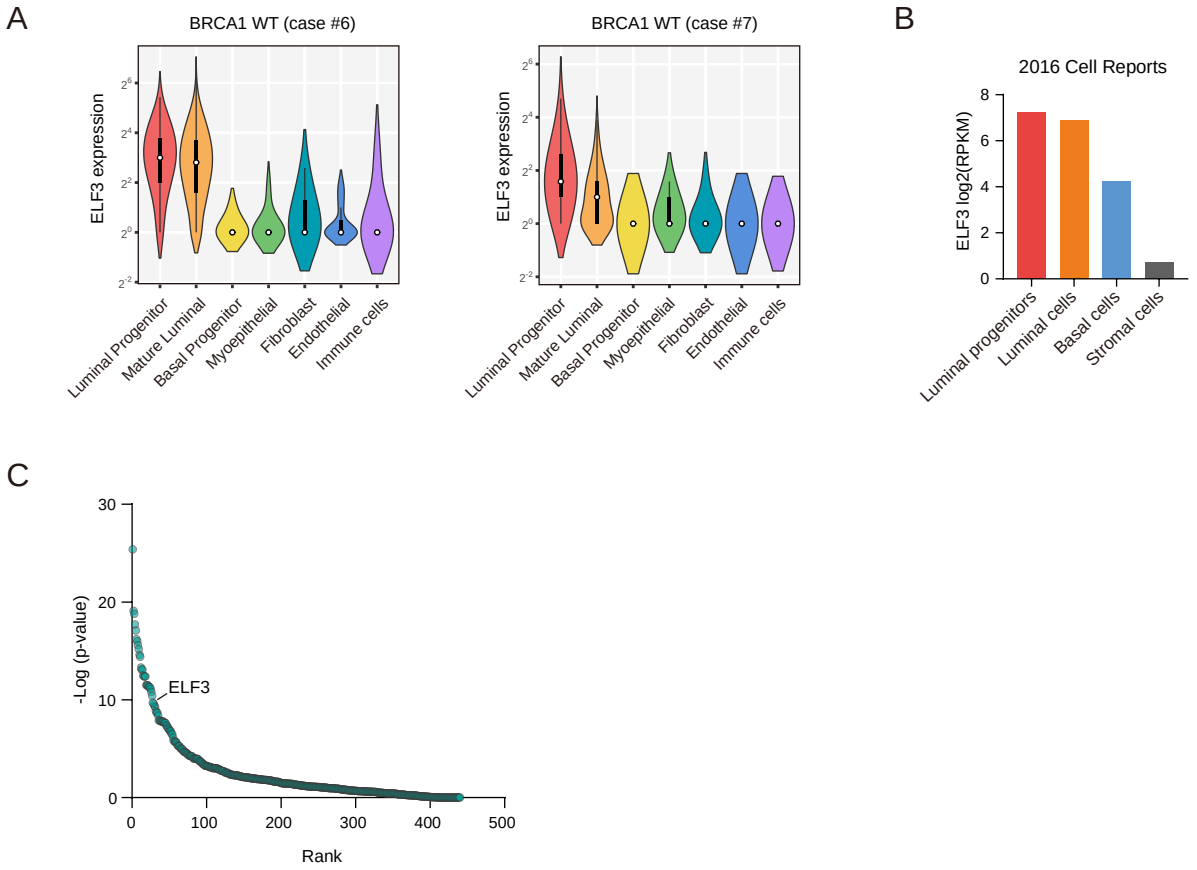

A
BRCA1 WT (case #6)
BRCA1 WT (case #7)
ELF3 expression
Luminal Progenitor
Mature Luminal
Basal Progenitor
Myoepithelial
Fibroblast
Endothelial
Immune cells
B
2016 Cell Reports
ELF3 log2(RPKM)
Luminal progenitors
Luminal cells
Basal cells
Stromal cells
C
-Log (p-value)
ELF3
Rank

Figure S6. LPs have a trait of higher replication stress, endowing them with the potential for transformation with BRCA1 deficiency.
(A) Violin plots displaying the distribution of ELF3 expression levels in distinct cell populations in normal mammary tissue from two noncarriers. (B) ELF3 expression levels (RPKM) in different cell populations from human normal mammary tissue in data of Pellacani. et al (33). (C) Heatmap showing the expression pattern of replication stress-related genes in BRCA1-mutated triple-negative breast cancer cells. The cancer cells were ordered based on Monocle-generated cell trajectories of tumorigenesis (Figure 6E).