The histone demethylase HR suppresses breast cancer development through enhanced CELF2 tumor suppressor activity

Yao Shen
Jasvinder Singh
Bindeshwar Sah
Zhongming Chen
Wootae Ha
Christine Henzler
Tao Su
Columbia University Irving Medical Center
Lilian Xie
Yibin Deng
Gen Li
Hua Guo
Carol Lange
University of Minnesota https://orcid.org/0000-0003-2751-3976
Hanina Hibshoosh
Columbia University
Liang Liu (liu00965@umn.edu)
University of Minnesota https://orcid.org/0000-0001-7245-355X

Article

Keywords: Hairless, breast cancer, histone methylation, CELF2, epigenetic therapy

Posted Date: May 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1639413/v1

The histone demethylase HR suppresses breast cancer development through enhanced CELF2 tumor suppressor activity

Yao Shen[1#], Jasvinder Singh[2#], Bindeshwar Sah[2#], Zhongming Chen[2], Wootae Ha[2], Christine Henzler[3], Tao Su[4], Lillian Xie[2], Yibin Deng[5,6], Gen Li[7], Hua Guo[4], Carol Lange[6,8], Hanina Hibshoosh[4], and Liang Liu[2,6*]

[1]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032
[2]The Hormel Institute, University of Minnesota, Austin, MN
[3]Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455
[4]Department of Pathology, Columbia University Irving Medical Center, New York, NY 10032
[5] Department of Urology, University of Minnesota Medical School, Minneapolis, MN 55455
[6]Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455
[7]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, 48109
[8]Department of Medicine. University of Minnesota, Minneapolis, MN 55455

[#]These authors contributed equally.

[*]To whom correspondence should be addressed:
The Hormel Institute, University of Minnesota
Austin, MN
55912

Phone: 507-437-9627
Email: LIU00965@umn.edu

ABSTRACT

The *hairless* (*HR*) gene encodes a transcription factor with histone demethylase activity that is essential for development and tissue homeostasis. Previous studies suggest that mutational inactivation of HR promotes tumorigenesis. To investigate HR mutations in breast cancer, we performed targeted next-gen sequencing using DNA isolated from primary breast cancer tissues. We identified *HR* somatic mutations in approximately 15% of the patient cohort (n=85), compared to 23% for *BRCA2*, 13% for *GATA3,* 7% for *BRCA1*, and 3% for *PTEN* in the same patient cohort. We also found an average 23% *HR* copy number loss in breast cancers. In support of HR's antitumor functions, HR reconstitution in *HR*-deficient human breast cancer cells significantly suppressed tumor growth in orthotopic xenograft mouse models. We further demonstrated that HR's antitumor activity was at least partly mediated by transcriptional activation of *CELF2*, a tumor suppressor with RNA-binding activity. Consistent with HR's histone demethylase activity, pharmacologic inhibition of histone methylation suppressed HR-deficient breast cancer cell proliferation, migration and tumor growth. Taken together, we identified HR as a novel tumor suppressor that is frequently mutated in breast cancer. We also showed that pharmacologic inhibition of histone methylation is effective in suppressing HR-deficient breast tumor growth and progression.

Keywords Hairless, breast cancer, histone methylation, CELF2, epigenetic therapy

Breast cancer (BC) is the most commonly diagnosed malignancy (excluding non-melanoma skin cancer) and the second leading cause of cancer-related death among women in the United States[1]. It is a heterogeneous disease, encompassing various subtypes with distinct origins, behavior, prognoses, and response to therapy[2]. Cancer genomics studies have revealed high-risk predisposing germline mutations, such as *BRCA1*/*BRCA2*, and other medium- to low-risk mutations in familial BC cases, including *CHEK2*, *TP53*, *PTEN*, and *CDH1*[3,4]. Despite decades of intense research into the genetic mechanisms involved in the initiation and progression of BC, only about 10% of cases are linked to pathogenic alterations in known risk genes[3]. The causes of the majority of breast malignancies remain elusive. It has now been established that the initiation and progression of BC, traditionally considered a genetic disease, involve epigenetic abnormalities in conjunction to genetic alterations[5]. Epigenetic regulators, such as DNA methyltransferases and histone-modifying enzymes, play essential roles in gene regulation and cancer development[6-9]. Genetic alterations affecting the function or enzymatic activity of these epigenetic regulators have been linked to cancer and other developmental disorders[10-12]. Histone methyltransferase dysregulation, for example, has been causally linked with BC pathogenesis[12-15].

The *hairless* (*HR*) gene encodes a transcription factor that is essential for development and tissue homeostasis[16,17]. Major functional domains of the HR protein are conserved between humans, rats, and mice, including a zinc finger domain, multiple LXXLL motifs that mediate protein-protein interactions, and a Jumonji C (JmjC) domain at the C-terminus[15]. Previous studies have demonstrated that the role of HR in transcriptional regulation is partly mediated via its binding with the thyroid hormone receptor, retinoic acid receptor-related orphan receptors, and vitamin D receptor[15,18]. The JmjC domain is a signature motif found among histone demethylases[8,19]. We recently reported that HR is a histone demethylase that

targets mono- or di-methylated histone H3 lysine 9 (H3K9me1 or me2)[20]. While H3K9me1 is found in both actively expressed or repressed genes, H3K9me2 is generally associated with repressed genes and heterochromatin[8,14,21]. By demethylating H3K9me1 or me2, HR may epigenetically suppress or activate gene expression via chromatin remodeling.

Mice with loss-of-function *Hr* mutations are hairless and highly susceptible to UV-induced skin tumorigenesis[22], which is suggestive of HR's antitumor function in skin tumorigenesis. We previously identified several HR target genes that are important regulators of cell proliferation and apoptosis[20,23-27]. Analysis of the cBioPortal for Cancer Genomics database revealed over 200 missense *HR* mutations in various human cancer types[15,28]. Notably, mutations affecting HR JmjC domain are reported in BC[15]. Herein, we provide evidence showing *HR* mutations are highly prevalent in breast tumors based on targeted deep-sequencing. *In vitro* and xenograft mouse experiments further supported the antitumor function of HR that is at least partly mediated via CELF2 in BC development. We also show that inhibition of H3K9 methylation abrogates the growth and migration of HR-deficient BC cells.

Results

Frequent HR mutations in BC tissues and cell lines.

HR exhibits antitumor activity during skin tumorigenesis[15,22], but its tumor suppressor function remains to be defined in other settings. Analysis of BC genomics datasets revealed several recurrent HR mutations such as G337D, S366C, E448G, R747H, R927C, P1046L, and G368fs (frame-shift insertion)[15,28]. To better define *HR* mutation prevalence and spectrum in BC, we performed targeted next-gen sequencing using genomic DNA sample from primary breast tumor tissues. DNA samples were pre-extracted and stored in the Macromolecular Bank of the

Molecular Pathology Shared Resource at Columbia University Irving Medical Center (CUIMC). Samples from patients with clinical and tumor registry data in the Database Shared Resource (CUIMC) were selected and included in this study. DNA samples from at least 20 luminal A, luminal B, HER2-enriched, or triple-negative breast cancer (TBNC) subtypes (defined based on receptor expression status) were obtained through the CUIMC Macromolecular Bank and used in the targeted sequencing analysis.

In addition to *HR*, mutations of other genes such as *BRCA1*, *BRCA2*, *PTEN*, *ESR1* and *TP53*, were also included in the targeted sequencing panel in order to compare the prevalence of *HR* mutations relative to these established BC susceptibility genes in the same patient cohort (**Supplementary Table 1**). We used Agilent XTHS2 hybrid capture-based target enrichment method to prepare the sequencing libraries. Libraries from 18 luminal A, 21 luminal B, 25 HER2-enriched, 19 TNBC, and six BC cell line samples passed sequencing library quality control and proceeded to next-gen sequencing.

As summarized in **Fig. 1A**, the prevalence of *TP53* somatic mutations (36%) and their enrichment in HER2-enriched/TNBC subtypes (**Table 1**) were consistent with previous findings[29-32], which validated the accuracy and reliability of our targeted sequencing results. Interestingly, we found that the mutation frequency in *HR* coding regions (15%) is lower than that of *BRCA2* (23%), but higher than those of other BC susceptibility genes including *BRCA1* (7%), *ESR1* (3%) and *PTEN* (3%) (**Fig. 1A**). Unlike higher *TP53* mutations in HER2-enriched/TNBC subtypes, *HR* mutations had a relatively even distribution among different subtypes (**Table 1**). Notably, multiple *HR* mutation types were detected in the same patient in approximately 50% patients with *HR* mutations (indicated by black bars in **Fig. 1A**), whereas the other genes were affected predominantly by missense mutations (indicated by green bars in **Fig. 1A**), indicating a higher genetic instability of the *HR* gene locus in a subset of patients.

Additionally, we found that *HR* mutations co-occurred with those of *TP53* and *BRCA2* in a subset of specimens (**Fig. 1B-C**), although the clinical significance of these co-occurring mutations is not clear. Survival analysis revealed no significant association between *HR* mutations and patient survival. However, we found that patients with *HR* and *TP53* co-mutations had a significantly increased risk of tumor recurrence (**Fig. 1D**), suggesting a possible collaborative function between the loss of HR and p53 in promoting breast tumorigenesis. Detailed mutation profiling of the *HR* locus (including non-coding regions) in each tumor sample (**Supplementary Fig. 1A**) suggested that some tumors harbored more mutations than others. Furthermore, we found three mutations in the HR JmjC domain and several recurrent missense *HR* mutations that are present in other cancer types, including P63R, G337D, and G823S (**Supplementary Fig. 1B**). The impact of these recurrent mutations on HR function remains to be determined in future functional studies.

HR copy number (CN) loss is associated with an increased risk of tumor recurrence.

In addition to missense mutations, analysis of the Cancer Cell Line Encyclopedia (CCLE) mutation database revealed *HR* gene CN loss among human BC cell lines[33,34], highlighting a potential role of HR deficiency in BC development. To determine the frequency of *HR* CN changes in primary tumor tissues, we performed droplet digital PCR (ddPCR) assays to assess the prevalence of *HR* CN status in different BC subtypes (n=94). The ddPCR results summarized in **Fig. 2A** identified *HR* CN loss in all four subtypes, with a relative low frequency in the luminal A subtype and a high frequency in the HER2-enriched subtype, largely consistent with the observation of *HR* CN loss in human BC cell lines. Among the BC cell lines, we found *HR* CN loss in T47D and MDA-MB-231, but not in MCF10A and MCF7 cells (data not shown).

To determine whether *HR* CN loss was associated with clinical outcomes, we retrieved patient clinical information from the CUIMC tumor registry, including survival status, tumor recurrence, and remission. We conducted Fisher's exact test to determine the significance of differences in the association between *HR* CN loss and survival, tumor recurrence, or remission with that from *HR* wild-type tumors. As shown in **Fig. 2B** and **Fig. 2C**, *HR* CN loss was significantly associated with tumor recurrence (**Fig. 2B**, p = 0.01663, n=63), but no significant association with tumor remission was observed (**Fig. 2C**, p = 0.5519, n=54). Similarly, *HR* CN loss was associated with reduced patient survival albeit not statistically significant (data not shown). Taken together, these analyses suggested that *HR* deficiency may serve as a novel biomarker for an increased risk of tumor recurrence.

HR re-expression reduced histone H3K9 methylation in BC cells as well as tumor growth in vivo.

While HR mutation and CN loss were detected in clinical specimen, only HR CN loss was detected in two of the BC cell lines (MDA-MB-231 and T47D). HR expression in MDA-MB-231 and T47D was undetectable via qRT-PCR or western blotting (**Fig. 3A**), consistent with our ddPCR analysis. To test if HR reconstitution in these HR-deficient BC cells could suppress H3K9 methylation, we utilized a lentivirus-based expression system to restore HR expression in T47D and MDA-MB-231 cells. Puromycin-resistant HR-expressing (HR) and control lentivirus-transduced (Ctrl) cells were analyzed via qRT-PCR and western blotting to confirm HR expression (**Fig. 3A**). Consistent with HR's H3K9 demethylase activity, HR re-expression led to a drastic loss of histone H3K9 methylation, as confirmed by immunofluorescence (IF) staining (**Fig. 3B**). Based on *in vitro* cell proliferation assays using the Incucyte system, we

found no significant changes between HR-expressing and control MDA-MB-231 and T47D BC cells (data not shown).

To test if HR restoration reduces tumor growth by these HR-deficient cells *in vivo*, we injected control and HR-expressing cells orthotopically into mammary fat pads in NOD/SCID mice. Approximately 8 weeks after injection, mice were euthanized for necropsy to examine tumor growth at the injected mammary pads and metastasis to other organs. As shown in **Fig. 3C**, HR re-expression substantially suppressed both MDA-MB-231 and T47D tumor growth. Necropsy analysis found tumor metastasis to the skin and lung in mice injected with control MDA-MB-231 cells but not HR-expressing cells, whereas no metastatic lesions were found in mice injected with either control or HR-expressing T47D cells (data not shown). These xenograft experiments suggested that HR effectively inhibited the *in vivo* tumor growth and metastasis of MDA-MB-231 cells.

UNC0642 is a potent and selective small-molecule inhibitor of H3K9 methylation, which targets H3K9 methyltransferases G9a and GLP[35]. Based on the H3K9 demethylase activity of HR, we postulated that inhibition of H3K9 methylation via UNC0642 can suppress the growth and viability of HR-deficient cancer cells. To test this, we treated MDA-MB-231 cells with different concentrations of UNC0642 and measured their proliferation over a period of seven days. As shown in **Fig. 3D**, UNC0642 at both 5 and 10 μM significantly inhibited MDA-MB-231 growth (**Fig. 3D**, left panel). Intriguingly, HR re-expression in MDA-MB-231 cells conferred resistance to UNC0642 treatment (**Fig. 3D**, right panel), probably because these cells may have adapted to low genomic H3K9 methylation following HR reconstitution. UNC0642-induced inhibition of H3K9 methylation was confirmed by immunofluorescence (IF) staining and western blotting (**Fig. 3E**). In addition to growth inhibition, *in vitro* clonogenic assays and wound-healing assays demonstrated that UNC0642 significantly reduced the clonogenicity and

migration of MDA-MB-231 cells, respectively (**Fig. 3F&G**, and **Supplementary Fig. 2A&B**). We also demonstrated that UNC0642 significantly inhibited MDA-MD-231 xenograft tumor growth in NOD SCID mice that was coupled with marked losses of H3K9me1/m2 in comparion with DMSO-treated mice (**Supplementary Fig. 2C**), highlighting the *in vivo* antitumor efficacy of UNC0642 and its potential as an epigenetic anticancer drug.

Identification of target genes mediating the antitumor function of HR in MDA-MB-231 BC cells.

To identify HR-regulated genes that might facilitate its tumor suppressor function in BC cells, we performed RNA-seq to determine differential gene expression between HR-expressing and control MDA-MB-231 cells. Differentially expressed genes between four pairs of *HR*-deleted breast tumors and matched normal breast tissues were also identified via RNA-seq. Among the top-ranked differentially expressed genes between these two RNA-seq datasets, *CELF2* and *NR2F1* were consistently decreased in *HR*-deleted breast tumors (**Fig. 4A**), while upregulated in HR-expressing MDA-MB-231 cells (**Fig. 4B**). In contrast, we found that the expression of *CDK15*, *CA9*, and *ARSI* was elevated in tumors but repressed in HR-expressing MDA-MB-231 cells (**Fig. 4A&B**). As shown in **Fig. 4C**, we validated the differential expression of *CELF2*, *NR2F1*, *CDK15*, and *CA9* between HR-expressing and control MDA-MB-231 cells via qRT-PCR.

To assess the clinical relevance of HR and its target genes, we compared TCGA RNA-seq data from breast tumors and matched normal mammary tissues (n=112 pairs) via paired gene expression analysis. We found that *HR* expression is significantly decreased in tumors, in parallel to significant changes in the expression of these newly identified HR target genes (**Fig. 4D**). We have performed IF staining in matched pairs of *HR*-deleted breast tumors and

matched normal tissue, which revealed an increase in H3K9 methylation in the tumor tissue, coupled with increased expression of CDK15 and CA9 but decreased expression of CELF2 (**Fig. 4E**), consistent with the qRT-PCR and RNA-seq results. These HR-regulated genes may contribute to BC pathogenesis and could also serve as novel biomarkers for the identification of patients who may respond to BC treatment via H3K9 methylation inhibitors.

CELF2 exhibits potent anti-tumor activity in MDA-MB-231 cells.

Among the newly identified HR target genes, *CELF2* encodes an RNA-binding protein. CELF2 has been shown to suppress lung cancer cell proliferation by repressing AKT phosphorylation in a PTEN-dependent manner[36]. CELF2 activity is known to be epigenetically regulated via DNA methylation[37,38]. Notably, loss of HR expression in human BC specimens is coupled with CELF2 downregulation (**Fig. 4C**). Consistent with the positive correlation between HR and CELF2 expression, HR reconstitution reactivated CELF2 expression in MDA-MB-231 cells (**Fig. 5A**). Additionally, we showed that transient expression of CELF2 in MDA-MB-231 cells inhibited cell proliferation dose-dependently *in vitro* (**Fig. 5B**). To further explore the antitumor activity of CELF2 in BC cells, we generated stable CELF2-expressing MDA-MB-231 cells. Compared with vector-transfected control MDA-MB-231 cells, ectopic CELF2 expression significantly suppressed MDA-MB-231 cell proliferation and migration (**Supplemental Figure 2D**). Following injection into the mammary fat pads in NOD SCID mice, tumor growth by CELF2-expressing MDA-MB-231 cells was markedly reduced compared to the control MDA-MB-231 cells (**Fig. 5C&D**). Taken together, these experiments provide compelling evidence that CELF2 is upregulated by HR and exerts a strong tumor-suppressive effect on BC cells.

Discussion

The human *HR* gene is located on chromosome 8p, a region that is frequently lost in breast and other cancer types[39]. Analysis of genes near the *HR* locus on 8p reveals no other known tumor suppressor genes except for *HR* (unpublished observations), suggesting that HR loss might contribute to breast tumorigenesis. Few genetic studies of BC in the past have reported any significant HR mutations in BC patients. In this study, we performed targeted next-gen sequencing to define HR mutation frequency and profile in a cohort of BC samples collected between 2001-2013. Approximately 20-30 samples representing each BC subtype (luminal A, luminal B, HER2-enriched, and TNBC) were included in this study. Subtypes were defined by classical immunohistochemistry markers (ER, PR and HER2) together with KI67 index and clinicopathologic variables (tumor size, grade, stage and nodal involvement)[40,41]. The mutation rates of the established BC susceptiblity genes analyzed in this study are largely consistent with those reported in previous studies. Notably, we found *HR* mutations in approximately 15% of the cohort, which is lower than that of *TP53* and *BRCA2* but higher than that of *BRCA1*, *PTEN* and *ESR1* (**Fig. 1A**). After additional mutation filtering to remove variants labeled as "benign or likely benign" in the ClinVar database, potential pathogenic *HR* mutations remain prevalent compared to *BRCA2/PTEN/BRCA1* mutations (**Suppl. Fig. 1C**), suggesting that *HR* mutation might function as a risk factor for BC.

Our targeted sequencing results confirmed frequent *TP53* mutations in BC with subtype-dependent variations (higher in HER2-enriched and TNBC tumors than in luminal A and B tumors) (**Fig. 1A**). *TP53* mutation is also detected in MDA-MB-231, MDA-MB-468 and T47D BC cell lines, consistent with previous findings[42,43]. Unlike *TP53*, *HR* mutation does not show any subtype-dependent variations and is not detected in the BC cell lines examined in this study. Intriguingly, approximately 50% of HR-mutant tumors harbor two or more types of

mutations in HR, indicating a possible genetic instability at the HR locus in these patients. Survival analysis revealed no significant association between *HR* mutation and patient survival in this relatively small patient cohort, but *HR* and *TP53* co-mutations are associated with a significant risk of tumor recurrence (**Fig. 1D**). The clinical significance of *HR* mutations and their collaborative function with *TP53* mutations await further exploration in large patient cohorts.

In support of the antitumor function of HR in BC development, xenograft experiments clearly demonstrated that HR significantly inhibited xenograft tumor growth by HR-deficient MDA-MB-231 and 47D cells (**Fig. 3**). Through RNA-seq analysis, we identified several new HR target genes in BC cells, including *CELF2* and *CDK15*, which are important regulators of cell proliferation and apoptosis. *CDK15* encodes a cyclin-dependent kinase, but its role in cancer is not clearly understood. While CDK15 exhibits anti-apoptotic as well as tumorigenic activity and confers resistance to TRAIL-induced apoptosis[44], inhibition of CDK15 paradoxically enhances BC cell invasion and metastasis[45]. CELF2 functions as an RNA-binding protein to modulate its target genes post-transcriptionally via pre-mRNA alternative splicing to influences mRNA translation and stability. *CELF2* is suggested to be a tumor suppressor gene and can be epigenetically silenced via DNA methylation in human cancers[37]. Further, CELF2 can suppress lung cancer cell proliferation by repressing AKT phosphorylation in a PTEN-dependent manner[36]. Using the MDA-MB-231 cell culture and xenograft models, we demonstrated that ectopic CELF2 expression exhibits potent antitumor activity in MDA-MB-231 cells. The molecular targets and pathways mediating CELF2 tumor suppressive activity in BC cells await further studies.

In light of its demethylase activity, the tumor-suppressive function of HR might be mediated epigenetically through H3K9 methylation. HR can demethylate H3K9me1 and

H3K9me2 to activate or repress its target gene expression in a context-dependent manner[15,20,46,47]. Mutational inactivation of HR may lead to epigenetic activation of cancer-promoting genes or repression of tumor suppressor genes to promote BC development. Thus, HR-dependent alterations in H3K9 methylation offers an exciting opportunity to examine the mechanistic link between histone methylation and BC development. As a proof of concept, we demonstrated that HR-deficient MDA-MB-231 BC cells are highly sensitive to the treatment of a specific H3K9 methylation inhibitor UNC0642 (**Fig. 3**). Notably, MDA-MB-231 cells with HR expression became resistant to UNC0642 treatment, suggesting that HR-dependent H3K9 demethylation rendered MDA-MB-231 cells insensitive to H3K9 methylation inhibition. These findings highlight the potential use of UNC0642 or other methylation inhibitors in epigenetic BC therapy.

In mammalian cells, G9A is a major histone methyltransferase that catalyzes H3K9me1 and H3K9me2. Consistent with the opposite roles between HR (a histone demethylase) and G9A (a histone methyltransferase) in histone H3K9 methylation, *G9A* is suggested to function as an oncogene, in contrast to the antitumor function of HR. G9A is aberrantly upregulated in various human cancers and is associated with poor patient prognosis[48-51]. Thus, HR loss may contribute to tumorigenesis via increased H3K9 methylation. It is unknown, however, whether HR and G9A share common target genes in the genome. In addition to its role in H3K9 methylation, G9A can also methylate the p53 protein at Lys^{373} to inhibit p53 activity[50]. Intriguingly, HR has been shown to interact with p53 and/or the p53 DNA response element to influence p53-regulated pathways[52]. The interaction between HR and p53 may contribute, at least in part, to the antitumor function of HR. Whether HR-p53 interaction is mediated via the demethylase activity of HR remains to be determined.

In conclusion, the present study demonstrated that HR exerts a tumor-suppressive effect on BC cells. This antitumor function is partly mediated via the upregulation of CELF2. Ectopic expression of CELF2 alone is sufficient to significantly inhibit BC cell growth and migration. Furthermore, we demonstrated that the pharmacological inhibition of H3K9 methylation effectively suppressed the growth and survival of *HR*-deficient BC cells. This study paves the way for future studies to explore specific inhibitors of the H3K9 methylation pathway as epigenetic therapeutics to abrogate the growth and progression of HR-deficient tumors.

Methods

Cell culture and transfection. Human BC cell lines (ATCC) were cultured in RPMI1640 medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin (Life Technologies, Carlsbad, CA, USA). All cell lines were grown in a humidified atmosphere with 5% CO_2 at 37 °C, authenticated, and tested for mycoplasma infection (Lonza, Rockland, ME, USA). *HR*-expressing and control lentiviral vectors were designed and prepared by Vector Builder (VectorBuilder Inc., Chicago, IL, USA). *CELF2*-expressing and control DNA plasmids were obtained from Addgene (#96900). Cell transfection was performed using the jetOPTIMUS® reagent (Polyplus-transfection®, Illkirch-Graffenstaden, France). Following lentiviral transduction or plasmid DNA transfection, cells were selected using puromycin (1 μg/mL) or neomycin (0.7-1 mg/mL) to generate stable *HR*-expressing or *CELF2*-expressing cell lines and respective control cell lines for subsequent *in vitro* and xenograft experiments.

Clinical sample information and characteristics. DNA, RNA and primary breast tumor tissues or normal breast tissues were collected and stored in the Molecular Pathology Shared

Resource at CUIMC (IRB protocol #AAAB2667, approved by the Institutional Reviewer Board of Columbia University). Median patient age at the time of diagnosis was 51 years, with a range of 29–92 years. Over 90% of the patients had ductal carcinoma while others had carcinoma of another histological type, such as lobular carcinoma, medullary carcinoma, or mixed-ductal and lobular carcinoma. Median breast tumor tissue purity was 75%, with a range of 50-100%. Tumor subtypes were determined based on receptor (ER, PR, HER2) expression. At least 20 samples represent each tumor subtype were obtained for targeted sequencing analysis. Primary BC tissue sections with matched normal breast tissue sections were obtained through the CUIMC Tissue Bank for immunofluorescence experiments. Patient clinical information and tumor registry data were obtained through the CUIMC Database Shared Resource (IRB protocol #AAAL5871).

Targeted sequencing and CN analysis. Genomic DNA from BC tissues and cell lines was used for *HR* mutation analysis via targeted sequencing using custom probes from Agilent to capture the DNA fragments of *HR* and 10 other BC-related genes (**Table 1**). Captured DNA was used for library construction and sequencing following the manufacturer instructions. Similarly, genomic DNA from BC tissues and cell lines was subjected to *HR* CN analysis by ddPCR assays using a predesigned HR CN assay (dHsaCNS987944070, Bio-Rad, Hercules, CA, USA) and *RPP30* as a reference control gene. RNA samples from selected HR-deficient BC tissues were used for transcriptome profiling via RNA-seq as previously described[53].

HR reconstitution in HR-deficient BC cells. MDA-MB-231 and T47D are two HR-deficient BC cell lines with *HR* CN loss and no *HR* gene expression. MDA-MB-231 and T47D cells were transduced with a pLenti6.3/V5™-based *HR*-expressing lentivirus or with vector control

lentivirus, and puromycin was used for selection (Vector Builder). Puromycin-resistant HR-expressing MDA-MB-231 and T47D cells and the respective vector-transduced control cells were subjected to cell proliferation analysis, H3K9 methylation analysis, as well as qRT-PCR and western blotting.

Cell proliferation and wound-healing assays. Cell proliferation was assessed using the IncuCyte® live cell analysis system (Sartorius, Goettingen Germany) as previously described[54]. Briefly, 4×10^3 cells were seeded in a 96-well plate and treated with either UNC0642 (1, 5, and 10 μM, Cayman Chemicals, Ann Arbor, MI, USA), an H3K9 methylation inhibitor, or DMSO as control. Images were acquired over seven days and analyzed using IncuCyte ZOOM. Cell proliferation was expressed as the percentage of cell density observed during this period. For wound-healing experiments, cells were seeded in triplicates in 96-well plates and grown to confluence. Wounds were then created using the IncuCyte® Wound Maker. Thereafter, cells were washed with PBS to remove cellular debris. Fresh media with UNC0642 (5 μM and 10 μM) was added. The plate was incubated and wound healing was recorded every 12 h in the IncuCyte ZOOM time-lapse imaging system for a total of seven days. The wound healing rate was expressed as the width of wounds at specific time points. Statistical significance was determined using the Graphpad Prism software (GraphPad Software, San Diego, CA).

Clonogenic assay. To assess the colony-forming ability of BC cells following UNC0642 treatment, cells were seeded in six-well plates and treated with UNC0642 (5 and 10 μM) or DMSO (as control) for 48 h. Treated and control cells were split and re-seeded at 1,000 cells/well in a new six-well plate with UNC0642 or DMSO for another 48 h. Cells were then

incubated in fresh medium without UNC0642 or DMSO for six additional days before being fixed in 4% formaldehyde and stained with 0.5% crystal violet. The colony number was counted using the ImageJ software.

Mouse xenograft experiments. NSG mice (NOD SCID IL2Rg^{null}) were purchased from The Jackson Laboratory (Stock #005557). Animal care and use were conducted in accordance with NIH guidelines, and the mice were housed in an Association for Assessment and Accreditation of Laboratory Animal Care-accredited facility. All animal experiments were performed in strict accordance with protocols approved by the Institutional Animal Care and Use Committee of the University of Minnesota (2111-39576A). Mouse xenograft experiments were performed as previously described[55]. Briefly, MDA-MB-231 and T47D cells with reconstituted *HR* or *CELF2* expression and their respective controls were orthotopically injected into the mammary fat pad of NOD SCID mice. For each injection, 1×10^6 cells were resuspended in 100 µl of a serum-free media and Matrigel mixture (1:1 ratio). For UNC0642 treatment, mice bearing MDA-MB-231 cells were administered with either UNC0642 (5 mg/kg body weight, in 0.2% DMSO) or DMSO via i.p. injection three times a week for a total of 4 weeks. Tumor growth was monitored and recorded weekly for up to eight weeks or until the tumor reached 2 cm in diameter, after which the animals were euthanized for tumor collection and analysis.

RNA extraction, qRT-PCR, and RNA-seq analysis. Total RNA was isolated from cultured BC cells using the RNeasy kit (Qiagen, Hilden, Germany). For qRT-PCR, 2 µg of RNA was reverse-transcribed into cDNA. qRT-PCR was performed using RT^2 SYBR Green qPCR Mastermix (Qiagen), and the results were analyzed using the Bio-Rad CFX 96 Touch system (Bio-Rad). RNA-seq analysis was performed as previously described[53]. Total RNA (500 ng)

from each sample was subjected to poly-A pull-down to enrich mRNAs for library preparation using the Illumina TruSeq RNA prep kit (Illumina, San Diego, CA, USA). The resulting libraries were sequenced on an Illumina HiSeq4000 at the University of Minnesota Genomics Center. Sequencing reads were mapped to the human reference genome (NCBI/build37.2) using Tophat (version 2.0.4). Differentially expressed genes (DEGs) between HR-expressing and control cells or between primary breast tumors and adjacent normal mammary tissues were determined using the DESeq2 software package[56], with a fold change (FC) cutoff set at >2 or <0.5.

Western blotting and immunofluorescence staining. Primary antibodies were purchased from Abcam (Cambridge, MA, USA; HR, ab202840), Boster (Pleasanton, CA, USA; Hr, A04817), Aviva (Aviva Systems Biology, San Diego, CA; CELF2, ARP40323), Sigma (St. Louis, MO, USA; FLAG, F1804; β-actin, A1978), and ABClonal (Woburn, MA, USA; H3K9me1, A2355; H3K9me2, A2359; CDK15, A13163; NR2F1, A16437; CA9, A1658). For western blotting, approximately 30 μg of total cell lysate from each sample was loaded. IF staining was performed as previously described[20]. Briefly, cultured cells on glass coverslips or frozen tissue sections (6 μm in thickness) were fixed in 4% paraformaldehyde for 10 min or in cold ethanol for 20 min. Fixed cells or tissue sections were then washed 3 times with PBS and then incubated with blocking buffer (0.1% Triton X-100 and 10% normal serum in PBS) for 1 h, followed by incubation with a specific primary antibody (1:100 dilution) overnight at 4 ºC in a humidified chamber. After 3 consecutive 5-min washes with PBS, cell or tissue sections were incubated with secondary antibodies Alexa Fluor 488-goat anti-mouse Ig G or Alexa Fluor 568-goat anti-rabbit Ig G (Thermo Fisher, 1:100 dilution) at room temperature for 1 h. Negative controls were stained with secondary antibodies alone. Sections were mounted with gelvatol

mounting media containing 4,6-diamidino-2-phenylindole dihydrochloride (DAPI). Images were acquired using a fluorescence confocal microscope (Zeiss, Thornwood, NY, USA). All IF experiments were performed in at least three independent experiments.

Mutation calling and bioinformatics. Each pair of 2x151 MiSeq reads had significant overlap (median insert size: 164bp, mean insert size: 185 bp), so reads were merged using pear (v. 0.9.11) enforcing a minimum 50 bp overlap. Merged reads were used for all further analyses. Umitools (v.1.1.1) was used to remove molecular barcodes and demultiplex merged reads. BWA (v.0.7.17) was then employed to align reads, and base quality score recalibration was performed following GATK best practices. Variants were called with freebayes (v.1.3.1), mutect2 (v.4.1.9.0) and vardict (v.1.8.2). The resulting vcf files were then processed to split multiple-alternate-allele calls into separate lines, left align and normalize variants (bcftools v.1.9), and decompose multinucleotide polymorphisms (MNPs were generated by freebayes) into their component SNPs and indels (using vt decompose, v.0.5). Variants that were called by all three of the callers and annotated as “PASS” by mutect2 and vardict were subjected to further analysis. Variants were annotated with ANNOVAR[57] to add ExAC (exac03) and Gnomad Exome (gnomad211_exome) population frequencies and ClinVar clinical significance (clinvar_20200316), and were filtered to remove variants with an overall population frequency >1% in either ExAC or Gnomad, or a clinical significance of “Benign” or “Likely Benign” in ClinVar. Two samples (986 and 1256) that had median coverage <20x were excluded from further analysis. Variants for the remaining samples were compiled using a custom R script (R v. 4.0.0) to filter variants. Multiple criteria were applied to filter variants, i.e., only variants with > 5% variant allele frequency with at least three alternative reads supporting each variant were retained. To remove technical artifacts as well as common germline mutations, the Mutect2

Panel of Normals variants were used to filter the calls. Figures were generated using the R packages maftools (v.2.4.12)[58] and ggplot2 (v.3.3.3)[59].

Statistics and reproducibility. A proportion Z-test was performed to determine whether the proportion of BC tissues with CN loss was the same between different BC subtypes. Fisher's exact test was used to assess whether there was a significant association between CN loss and tumor recurrence or remission. To explore the co-occurrence of mutations within the same samples, we performed Fisher's exact test to determine whether there is a significant overlap between sample lists with specific mutations. P-values from the tests were FDR-corrected to control false positives. Breast cancer RNA-seq raw count data were obtained from the TCGA database. Reads were mapped to the pre-built Hg38 human genome, and the overlap between the reads and the genes was determined using STAR. Gene expression was modeled by negative binomial distribution using the DESeq2 package. Differential gene expression analysis was conducted by comparing the tumor samples versus the matching normal samples via the paired Student's t-test for a total of 112 pairs. To identify whether there are significant associations between genetic variants and time to tumor recurrence and survival in cancer patients, survival analysis was performed and the log-rank test is used to compare Kaplan-Meier curves between individual mutations as well as co-mutations.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All pertinent data are available within the manuscript or upon request. The sequencing datasets generated and analyzed during the current study will be available in the NCBI Sequence Read Archive database (accession code PRJNA799398). All code used to analyze sequencing data as described above are from publicly available resources.

References

1. Siegel RL, Miller KD, Fuchs HE, et al: Cancer statistics, 2021 (vol 71, pg 7, 2021). Ca-a Cancer Journal for Clinicians 71:359-359, 2021
2. Dai X, Cheng H, Bai Z, et al: Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. J Cancer 8:3131-3141, 2017
3. Ellis MJ, Perou CM: The Genomic Landscape of Breast Cancer as a Therapeutic Roadmap. Cancer Discovery 3:27-34, 2013
4. Apostolou P, Fostira F: Hereditary Breast Cancer: The Era of New Susceptibility Genes. Biomed Research International 2013, 2013
5. Flavahan WA, Gaskell E, Bernstein BE: Epigenetic plasticity and the hallmarks of cancer. Science 357, 2017
6. Bernstein BE, Meissner A, Lander ES: The mammalian epigenome. Cell 128:669-81, 2007
7. Chi P, Allis CD, Wang GG: Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer 10:457-69, 2010
8. Pedersen MT, Helin K: Histone demethylases in development and disease. Trends in Cell Biology 20:662-671, 2010
9. Feinberg AP, Koldobskiy MA, Gondor A: Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nat Rev Genet 17:284-99, 2016
10. Albert M, Helin K: Histone methyltransferases in cancer. Semin Cell Dev Biol 21:209-20, 2010
11. Pedersen MT, Helin K: Histone demethylases in development and disease. Trends Cell Biol 20:662-71, 2010
12. D'Oto A, Tian Q-W, Davidoff AM, et al: Histone demethylases and their roles in cancer epigenetics. Journal of medical oncology and therapeutics 1:34-40, 2016
13. Michalak EM, Visvader JE: Dysregulation of histone methyltransferases in breast cancer - Opportunities for new targeted therapies? Mol Oncol 10:1497-1515, 2016
14. Shinkai Y, Tachibana M: H3K9 methyltransferase G9a and the related molecule GLP. Genes Dev 25:781-8, 2011
15. Trager MH, Sah B, Chen Z, et al: Control of Breast Cancer Pathogenesis by Histone Methylation and the Hairless Histone Demethylase. Endocrinology 162, 2021
16. Potter GB, Beaudoin GMJ, DeRenzo CL, et al: The hairless gene mutated in congenital hair loss disorders encodes a novel nuclear receptor corepressor. Genes & Development 15:2687-2701, 2001
17. Ahmad W, Haque WFU, Brancolini V, et al: Alopecia universalis associated with a mutation in the human hairless gene. Science 279:720-724, 1998

18. Thompson CC: Hairless is a nuclear receptor corepressor essential for skin function. Nucl Recept Signal 7:e010, 2009
19. Cloos PA, Christensen J, Agger K, et al: Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev 22:1115-40, 2008
20. Liu L, Kim H, Casta A, et al: Hairless is a histone H3K9 demethylase. FASEB J 28:1534-42, 2014
21. Barski A, Cuddapah S, Cui K, et al: High-resolution profiling of histone methylations in the human genome. Cell 129:823-37, 2007
22. Kim H, Casta A, Tang XW, et al: Loss of Hairless Confers Susceptibility to UVB-Induced Tumorigenesis via Disruption of NF-kappaB Signaling. Plos One 7, 2012
23. Aziz MH, Sundling KE, Dreckschmidt NE, et al: Protein kinase Cepsilon inhibits UVR-induced expression of FADD, an adaptor protein, linked to both Fas- and TNFR1-mediated apoptosis. J Invest Dermatol 129:2011-21, 2009
24. Buyru N, Baltaci E, Yavuz E, et al: Downregulation of the DLEC1 gene is not associated with promoter methylation in head and neck cancer. European Journal of Cancer 50:S93-S93, 2014
25. Ying J, Poon FF, Yu J, et al: DLEC1 is a functional 3p22.3 tumour suppressor silenced by promoter CpG methylation in colon and gastric cancers. British Journal of Cancer 100:663-669, 2009
26. Feng ZE, Li JN, Wang L, et al: The prognostic value of glycerol-3-phosphate dehydrogenase 1-like expression in head and neck squamous cell carcinoma. Histopathology 64:348-355, 2014
27. Kelly TJ, Souza AL, Clish CB, et al: A Hypoxia-Induced Positive Feedback Loop Promotes Hypoxia-Inducible Factor 1 alpha Stability through miR-210 Suppression of Glycerol-3-Phosphate Dehydrogenase 1-Like. Molecular and Cellular Biology 31:2696-2706, 2011
28. Maatough A, Whitfield GK, Brook L, et al: Human Hairless Protein Roles in Skin/Hair and Emerging Connections to Brain and Other Cancers. J Cell Biochem 119:69-80, 2018
29. Zhang Y, Xiong S, Liu B, et al: Somatic Trp53 mutations differentially drive breast cancer and evolution of metastases. Nat Commun 9:3953, 2018
30. Darb-Esfahani S, Denkert C, Stenzinger A, et al: Role of TP53 mutations in triple negative and HER2-positive breast cancer treated with neoadjuvant anthracycline/taxane-based chemotherapy. Oncotarget 7:67686-67698, 2016
31. Abubakar M, Guo CY, Koka H, et al: Clinicopathological and epidemiological significance of breast cancer subtype reclassification based on p53 immunohistochemical expression. Npj Breast Cancer 5, 2019
32. Donehower LA, Soussi T, Korkut A, et al: Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. Cell Rep 28:3010, 2019
33. Barretina J, Caponigro G, Stransky N, et al: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483:603-7, 2012
34. Jiang GL, Zhang SJ, Yazdanparast A, et al: Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. Bmc Genomics 17, 2016
35. Kaniskan HU, Jin J: Recent progress in developing selective inhibitors of protein methyltransferases. Curr Opin Chem Biol 39:100-108, 2017
36. Yeung YT, Fan S, Lu B, et al: CELF2 suppresses non-small cell lung carcinoma growth by inhibiting the PREX2-PTEN interaction. Carcinogenesis 41:377-389, 2020

37. Pique L, de Paz AM, Pineyro D, et al: Epigenetic inactivation of the splicing RNA-binding protein CELF2 in human breast cancer. Oncogene 38:7106-7112, 2019
38. Yeung YT, Fan SY, Lu BB, et al: Celf2 suppresses non-small cell lung carcinoma growth by inhibiting the prex2-pten interaction. Cancer Research 79, 2019
39. Cai Y, Crowther J, Pastor T, et al: Loss of Chromosome 8p Governs Tumor Progression and Drug Response by Altering Lipid Metabolism. Cancer Cell 29:751-766, 2016
40. Dai XF, Li T, Bai ZH, et al: Breast cancer intrinsic subtype classification, clinical use and future trends. American Journal of Cancer Research 5:2929-2943, 2015
41. Ahn HJ, Jung SJ, Kim TH, et al: Differences in Clinical Outcomes between Luminal A and B Type Breast Cancers according to the St. Gallen Consensus 2013. Journal of Breast Cancer 18:149-159, 2015
42. Huovinen M, Loikkanen J, Myllynen P, et al: Characterization of human breast cancer cell lines for the studies on p53 in chemical carcinogenesis. Toxicol In Vitro 25:1007-17, 2011
43. Silwal-Pandit L, Vollan HK, Chin SF, et al: TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. Clin Cancer Res 20:3569-80, 2014
44. Park MH, Kim SY, Kim YJ, et al: ALS2CR7 (CDK15) attenuates TRAIL induced apoptosis by inducing phosphorylation of survivin Thr34. Biochemical and Biophysical Research Communications 450:129-134, 2014
45. Li S, Dai X, Gong K, et al: PA28alpha/beta Promote Breast Cancer Cell Invasion and Metastasis via Down-Regulation of CDK15. Front Oncol 9:1283, 2019
46. Chuma M, Endo-Umeda K, Shimba S, et al: Hairless modulates ligand-dependent activation of the vitamin D receptor-retinoid X receptor heterodimer. Biol Pharm Bull 35:582-7, 2012
47. Zarach JM, Beaudoin GMJ, Coulombe PA, et al: The co-repressor hairless has a role in epithelial cell differentiation in the skin. Development 131:4189-4200, 2004
48. Tu WB, Shiah YJ, Lourenco C, et al: MYC Interacts with the G9a Histone Methyltransferase to Drive Transcriptional Repression and Tumorigenesis. Cancer Cell 34:579-+, 2018
49. Wang YF, Zhang J, Su Y, et al: G9a regulates breast cancer growth by modulating iron homeostasis through the repression of ferroxidase hephaestin. Nature Communications 8, 2017
50. Huang J, Dorsey J, Chuikov S, et al: G9a and Glp Methylate Lysine 373 in the Tumor Suppressor p53. Journal of Biological Chemistry 285:9636-9641, 2010
51. Mabe NW, Garcia NMG, Wolery SE, et al: G9a Promotes Breast Cancer Recurrence through Repression of a Pro-inflammatory Program. Cell Reports 33, 2020
52. Brook L, Palade P, Maatough A, et al: Hairless regulates p53 target genes to exert tumor suppressive functions in glioblastoma. Journal of Cellular Biochemistry 120:533-543, 2019
53. Shen Y, Kim AL, Du R, et al: Transcriptome Analysis Identifies the Dysregulation of Ultraviolet Target Genes in Human Skin Cancers. PLoS One 11:e0163054, 2016
54. Ha W, Hinde A, Xie L, et al: Biomarker function of HMGA2 in ultraviolet-induced skin cancer development. Exp Dermatol 29:1021-1026, 2020
55. Liu L, Rezvani HR, Back JH, et al: Inhibition of p38 MAPK signaling augments skin tumorigenesis via NOX2 driven ROS generation. PLoS One 9:e97245, 2014
56. Anders S, Huber W: Differential expression analysis for sequence count data. Genome Biol 11:R106, 2010

57. Wang K, Li MY, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research 38, 2010
58. Mayakonda A, Lin DC, Assenov Y, et al: Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Research 28:1747-1756, 2018
59. Wickham H: ggplot2: Elegant Graphics for Data Analysis. Ggplot2: Elegant Graphics for Data Analysis:1-212, 2009

Acknowledgements

We thank Dr. Juan Abrahante Lloréns for expert bioinformatics support, Dr. Douglas Yee for his insightful suggestions and critical review of the manuscript, and Amanda Hinde for excellent technical assistance. This work utilized the Molecular Pathology Shared Resource Tissue Bank of the Herbert Irving Comprehensive Cancer Center of Columbia University which is in part funded through the NIH/NCI Cancer Center Support Grant P30CA013696. This work was supported by NIH/NIAMS grant K01AR064315, NIH/NCI R01CA196639, Paint The Town Pink, and Prevent Cancer Foundation.

Author contributions

L.L. and H.H. designed the study. Y.S., G.L., and C.H. performed bioinformatics and statistical analyses. J.S., B.S, Z.C., W.H., and L.X. performed the experiments and analyzed the data. Y.D., T.S., H.G., and H.H. contributed research materials. L.L., C.L, and H.H. wrote and edited the manuscript. All authors contributed to the editing of the manuscript.

Competing Interests

The Authors declare no competing financial or non-financial interests.

Table 1. Mutation frequency of the genes included in the targeted deep-sequencing panel among BC subtypes (based on receptor expression status) and 5 BC cell line. Probes were designed to capture the exons of each of the genes for sequencing.

	Lum A	Lum B	HER2-enriched	TNBC	Cell lines
Total	21	16	24	20	5
HR	3	4	4	3	0
TP53	5	4	9	10	3
BRCA2	5	3	8	4	1
KMT2C	4	2	5	5	3
GATA3	3	2	3	2	1
BRCA1	1	1	1	3	0
ESR1	1	0	1	0	1
PTEN	2	0	0	0	1
SF3B1	1	0	0	1	0
WT1	0	1	0	0	0

Figure Legends

Figure 1. *HR* gene mutations are correlated with an increased recurrence rate of human BC. (**A**) Oncoplot showing mutations identified via targeted sequencing in the coding regions of *HR* and other key genes known to contribute to BC pathogenesis in 85 BC samples and 6 BC cell lines. X-axis represents each individual patient. Y-axis depicts the number of variants identified in all 11 genes in each patient. (**B**) An UpSet plot showing the co-occurrence of mutations among selected genes within each sample. Intersection size describes the number of samples that have at least one mutation in each of the genes with a dot below them. Multiple Fisher exact tests were performed to determine the significance of each pair of the co-occurring mutations among the samples. Based on the FDR-corrected p-values, *HR* and *BRCA2* had a significant co-mutation rate among samples. (**C**) Venn diagram showing the co-mutation rate of *HR*, *TP53*, and *BRCA2* among samples. (**D**) Tumor recurrence and patient survival analysis based on *HR* and *TP53* mutations or co-mutations.

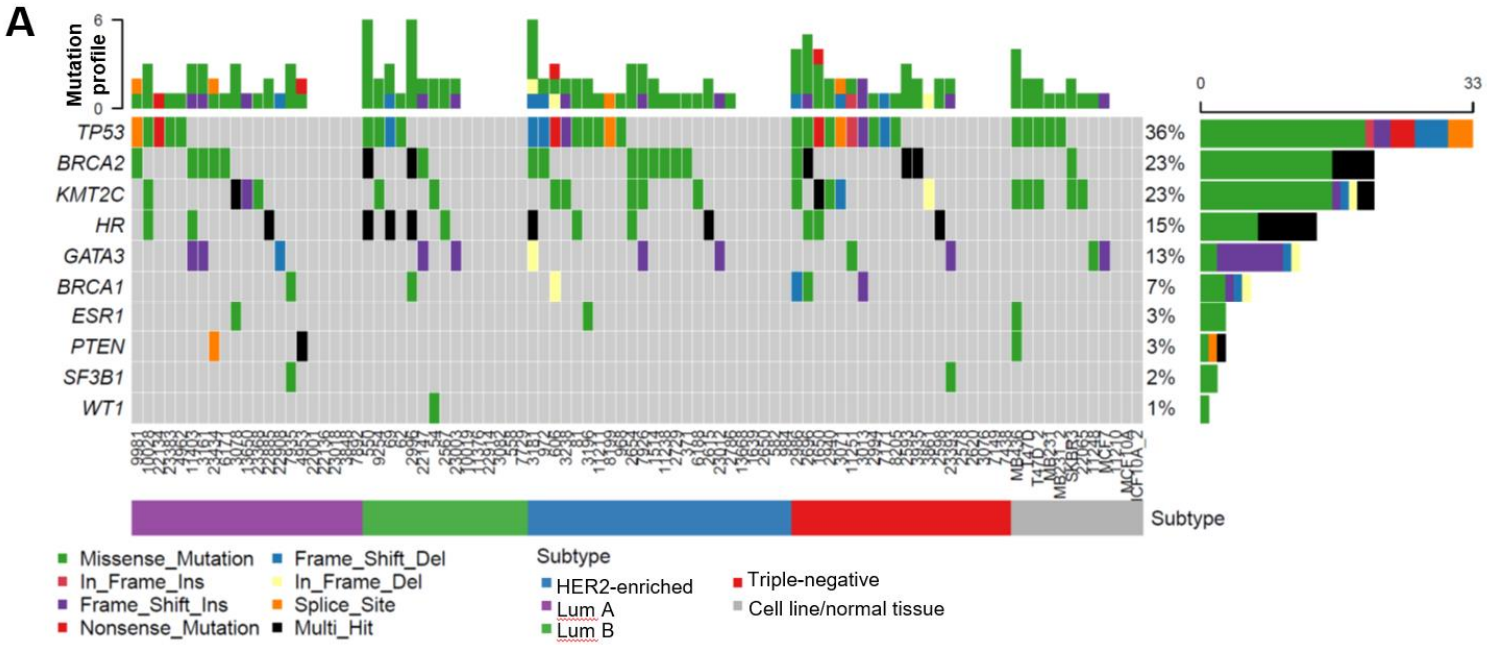
Figure 2. (**A**) Frequency of *HR* CN loss in BC specimens as determined via ddPCR analysis of 94 BC specimens encompassing different BC subtypes. Lum A group had a lower than average CN loss (23.4%) across all subtypes ($p = 0.0784$ based on the proportion Z-test), while HER2-enriched subtypes had the greatest average CN loss ($p = 0.0794$). (**B, C**): Fisher's exact tests showed a significant association between *HR* CN loss and tumor recurrence ($p = 0.01663$, **B**), but not tumor remission (**C**).

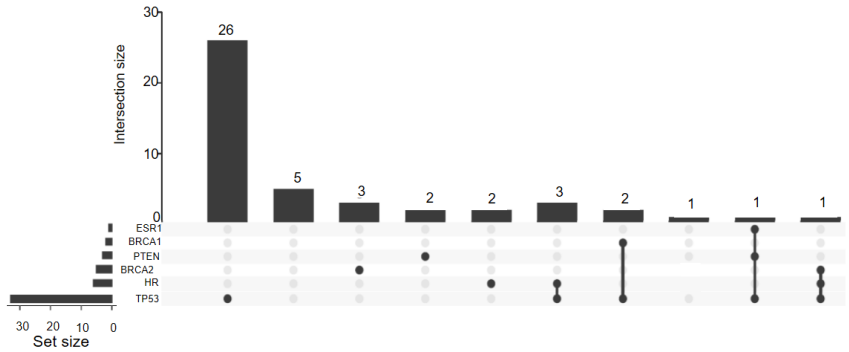
Figure 3. HR regulates breast tumor progression through H3K9 methylation. (A) HR mRNA and protein expression in MDA-MB-231 and T47D cells determined by real-time PCR and western blot. (**B**) Relative levels of mono-methylated H3K9 (H3K9me1) and di-methylated (H3K9me2) in MDA-MB-231 cells. MDA-MB-231 cells were transduced with control lentivirus (**Ctrl**) or lentivirus-HR (**HR**), followed by immunofluorescence staining with anti-H3K9me1 (red), Flag (green), and counter-staining with DAPI (blue). Scale bar: 20 μm. (**C)** Tumor growth after MDA-MB-231 and T47D cells transduced with control lentivirus (**Ctrl**) and lentivirus-HR (**HR**) were inoculated into the fat pads in NOD SCID mice. (**D**) IncuCyte ZOOM time-lapse imaging analysis for MDA-MB-231 cell proliferation after treatment with DMSO (Ctrl) or UNC0642 (1, 5, or 10 μM). Data are presented as the mean ± standard deviation (n=3). (**E**) Detection of changes in H3K9me1 and H3K9me2 in MDA-MB-231 cells after treatment with

DMSO (control) or UNC0642 by IF and western blotting. For IF experiments, cells were stained with anti-H3K9me1 or anti-H3K9me2 (red) and counter-stained with DAPI (blue). Scale bar: 50 μm. (**F**) Colonies formed by MDA-MB-231 cell after treatment with DMSO (control) or UNC0642 (5 or 10 μM). (**G**) Wound healing of the MDA-MB-231 cell monolayer after treatment with DMSO (Ctrl) or UNC0642 (5 or 10 μM). **: $p<0.01$ vs control. ***: $p<0.001$ vs control.

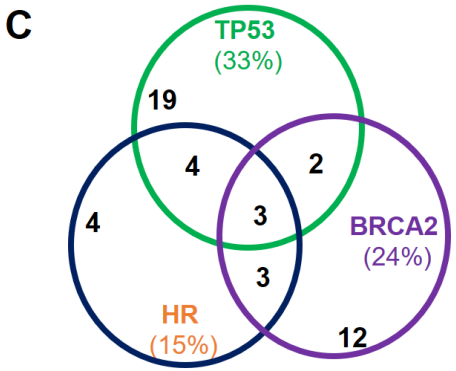
Figure 4. Identification of HR-regulated genes in BC cells and breast tumors. (**A**) RNA-seq analysis identified multiple genes whose expression was consistently increased or decreased in breast tumors compared to matched normal tissues (n=4). (**B**) RNA-seq analysis showing that the expression of the genes identified in **A** was reversed via *HR* re-expression in human MDA-MB-231 cells (n=3). (**C**) qRT-PCR validation of HR-regulated genes in HR-expressing and control MDA-MB-231 cells. **: $p<0.01$ vs control. (**D**) Comparative analysis of *HR* and target gene mRNA expression in human BC tissues compared to matched normal breast tissues from the TCGA dataset (n=112 pairs). (**E**) Representative IF images showing reduced CELF2 expression, increased H3K9me2 methylation as well as increased CDK15 and CA9 expression in HR-deficient BC tissue compared to matched normal breast tissue.

Figure 5. **CELF2 suppresses BC cell proliferation and tumor growth.** (**A**) Upregulation of CELF2 protein expression in HR-reconstituted cells. CELF2 protein expression in HR-reconstituted (HR) MDA-MB-231 cells was examined via IF (left) or western blotting (right) using a CELF2 antibody (scale bar: 50 μm). β-actin was used as a loading control. (**B**) Proliferation of MDA-MB-231 cells transfected with control or CELF2-expressingg plasmids (1 and 2 μg). Cell proliferation was expressed as the relative confluency by using the Incucyte system. *: $p<0.05$ vs the control. (**C**) Confirmation of CELF2 overexpression in a selected CELF2-transfected MDA-MB-231 cell line via IF (left) or western blotting (right) using the CELF2 antibody. (**D, E**) *In vivo* tumor growth curves of control or CELF2-overexpressing (CELF2) MDA-MB-231 cells following injection into SCID mouse fat pads. Tumor sizes measured from each mouse were illustrated with a solid (control) or dotted (CELF2 overexpression) line in the scatter plot (**D**). Average tumor sizes from control or CELF2 overexpression groups are illustrated in (**E**). ***: $p<0.001$ vs control.





Co-mutations	p-value (FDR corrected)
BRCA2_HR	0.003705461
TP53_HR	0.010243303
GATA3_HR	0.032968645
BRCA1_HR	0.036887995
KMT2C_HR	0.083540458
PTEN_HR	1
WT1_HR	1
SF3B1_HR	1
ESR1_HR	1



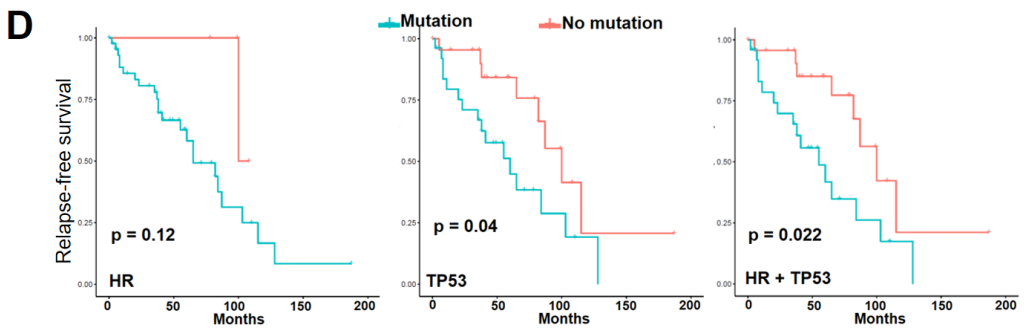


Figure 1

A Summary of *HR* CN loss in BC specimens

	Total	CN Loss	%
Luminal A	20	2	10
Luminal B	21	6	29
HER2-enriched	29	10	34
Triple negative	24	4	17

B *HR* CN loss and tumor recurrence

	CN loss	No CN Loss
Recurrence	10	9
No recurrence	9	35

p = 0.01663; Odds ratio = 4.205609

C *HR* CN loss and tumor remission

	CN loss	No CN Loss
Remission	10	24
No remission	8	12

p = 0.5519; Odds ratio = 0.6306341

Figure 2

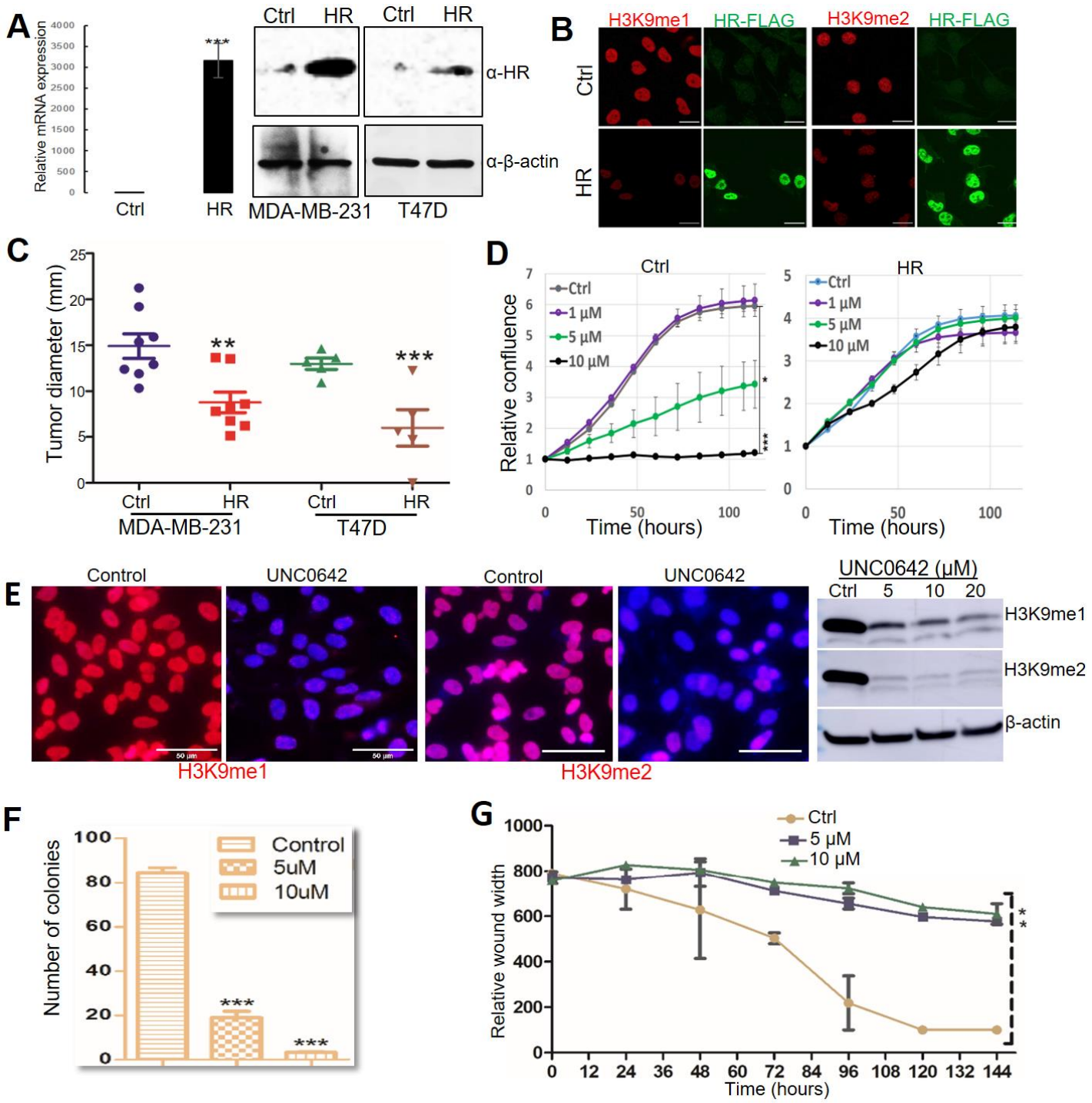

A
Relative mRNA expression
Ctrl
HR
***
Ctrl HR Ctrl HR
α-HR
α-β-actin
MDA-MB-231
T47D
B
H3K9me1
HR-FLAG
H3K9me2
HR-FLAG
Ctrl
HR
C
Tumor diameter (mm)
**
***
Ctrl HR Ctrl HR
MDA-MB-231
T47D
D
Ctrl
HR
Ctrl
1 μM
5 μM
10 μM
Relative confluence
Time (hours)
E
Control
UNC0642
Control
UNC0642
UNC0642 (μM)
Ctrl 5 10 20
H3K9me1
H3K9me2
β-actin
H3K9me1
H3K9me2
F
Number of colonies
Control
5uM
10uM
***
***
G
Ctrl
5 μM
10 μM
Relative wound width
Time (hours)

Figure 3

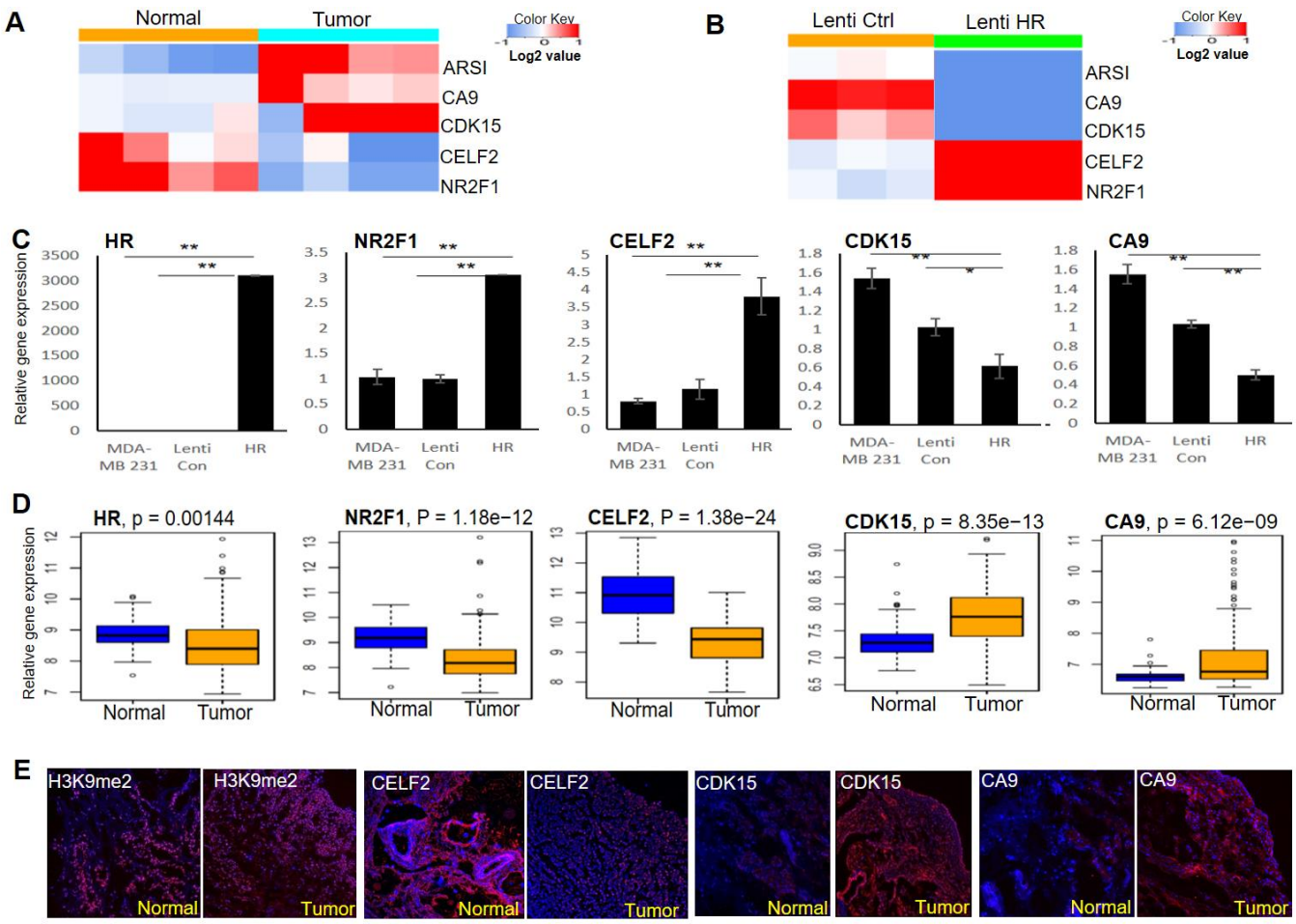

A
Normal
Tumor
Color Key
Log2 value
ARSI
CA9
CDK15
CELF2
NR2F1
B
Lenti Ctrl
Lenti HR
Color Key
Log2 value
ARSI
CA9
CDK15
CELF2
NR2F1
C
Relative gene expression
HR
NR2F1
CELF2
CDK15
CA9
MDA-MB 231
Lenti Con
HR
D
Relative gene expression
HR, p = 0.00144
NR2F1, P = 1.18e−12
CELF2, P = 1.38e−24
CDK15, p = 8.35e−13
CA9, p = 6.12e−09
Normal
Tumor
E
H3K9me2
Normal
H3K9me2
Tumor
CELF2
Normal
CELF2
Tumor
CDK15
Normal
CDK15
Tumor
CA9
Normal
CA9
Tumor

Figure 4

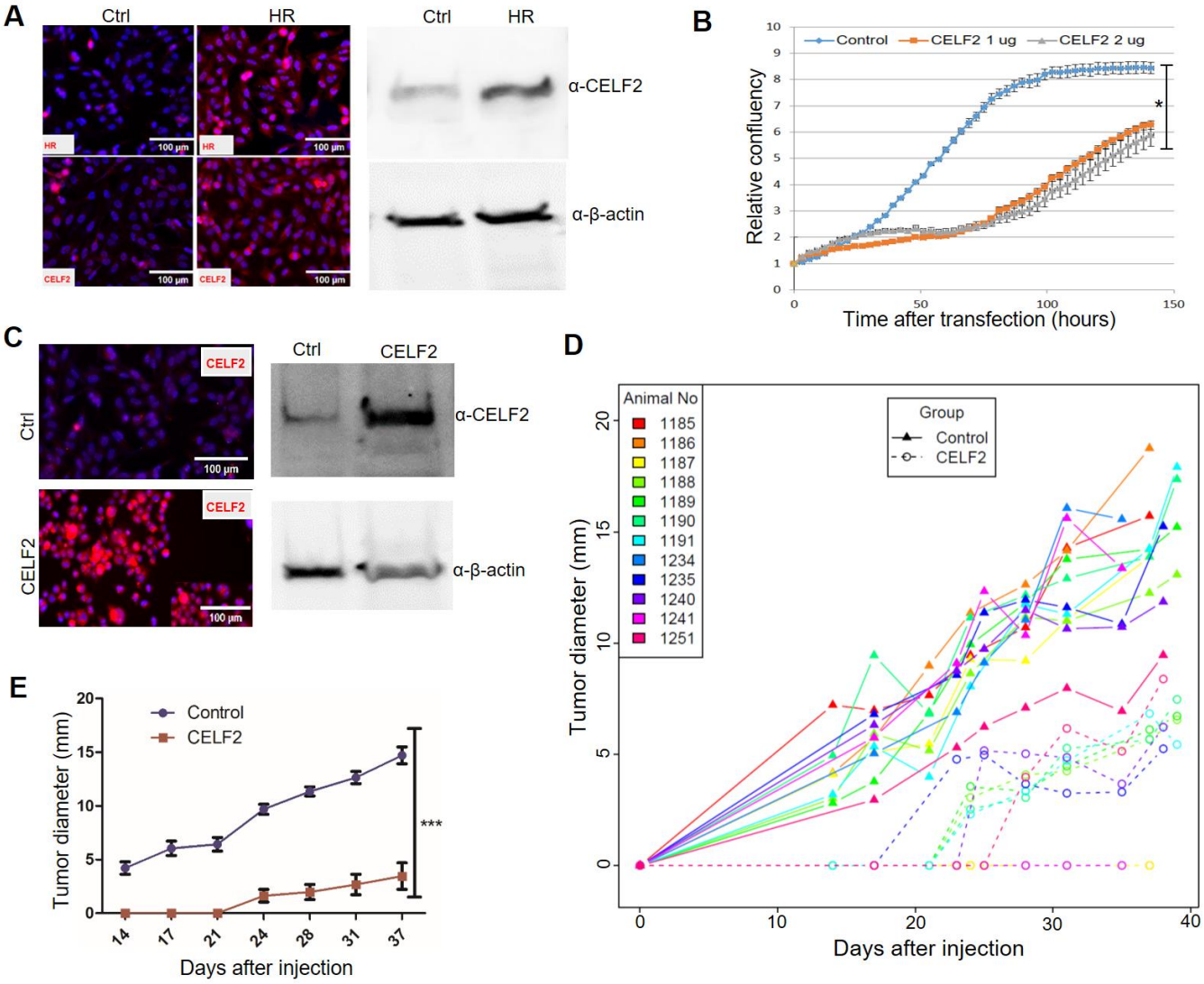
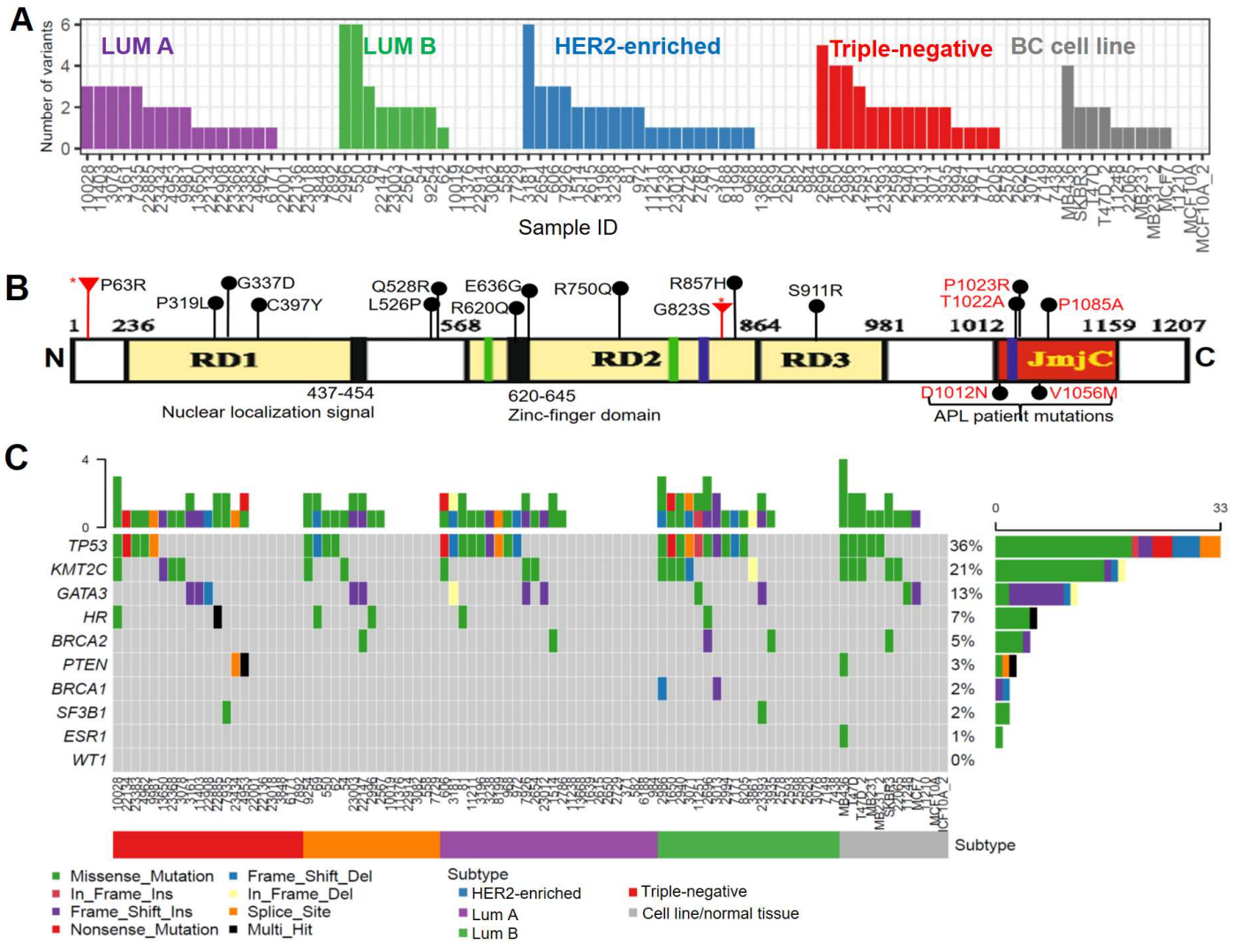

A
Ctrl
HR
Ctrl
HR
α-CELF2
α-β-actin
100 μm
B
Control
CELF2 1 ug
CELF2 2 ug
Relative confluency
Time after transfection (hours)
*
C
Ctrl
CELF2
α-CELF2
α-β-actin
D
Animal No
1185
1186
1187
1188
1189
1190
1191
1234
1235
1240
1241
1251
Group
Control
CELF2
Tumor diameter (mm)
Days after injection
E
Control
CELF2
Tumor diameter (mm)
Days after injection
***

Figure 5

Supplementary Table 1. Genes included in the targeted sequencing panel. *: Probes were designed to capture the entire HR gene locus, including both exons, introns, a 5' upstream fragment (1 kb), and a 3' downstream fragment (0.5 kb). For all other genes, probes were designed to capture the exons only.

Gene symbol	NCBI Gene ID	GenBank Accession
HR*	55806	NM_005144
TP53	7157	NM_001126118
WT1	7490	NM_006218
BRCA1	672	NM_007297
BRCA2	672	NM_000059
KMT2C (MLL3)	58508	NM_170606
GATA3	2625	NM_001002295
SF3B1	23451	NM_012433
ESR1	2099	NM_001122741

Supplementary Figure 1. HR mutation in breast cancers. (**A**) Mutation profiles of the entire HR gene locus in BC sample and selected BC cell lines. (**B**) Summary of HR missense mutations identified in BC specimens. *: **P63R** and **G823S** are two top recurrent mutations identified among the BC specimens. (**C**) Oncoplot showing mutations filtered by removing variants that are labeled as "benign/likely benign" in ClinVar database to highlight potentially consequential mutations.



Supplementary Figure 2. Impact of UNC0642 treatment on colony formation, wound healing, and in vivo tumor growth. MDA-MB-231 cells were treated with DMSO (control) or UNC0642, followed by assays of colony formation or wound healing. **(A)** Colony formation of the cells stained with violet crystal. (**B**) Representative images from wound healing assays showing that UMN0642 inhibited cell migration and wound healing in vitro. Dashed outlines indicate wound widths at the beginning (0 h) and end of the experiments (144 h). (**C**) Inhibition of MDA-MB-231 tumor growth in xenograft mouse model by UNC0642 compared to DMSO (n=10). Last UNC0642 or DMSO treatment was on day 30 after tumor cell injection (indicated by the red arrow). IF staining detected the loss of H3K9me1/m2 in UNC0642-treated tumors (**D**) Proliferation of control and two MDA-MB-231 cell clones with ectopic CELF2 expression. *: p< 0.05; **: p< 0.01; ***: p< 0.001.