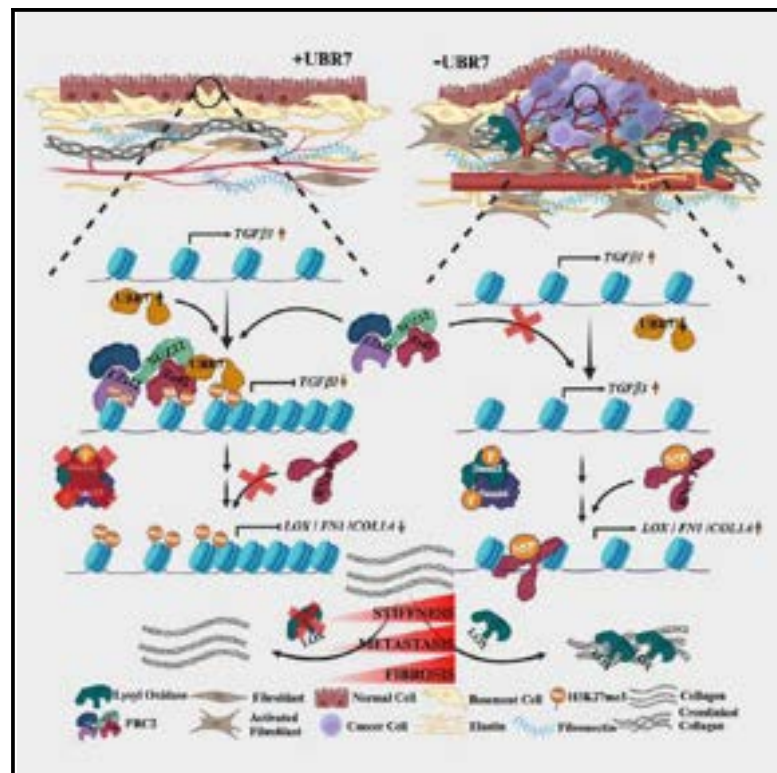**Article**

UBR7 in concert with EZH2 inhibits the TGF-β signaling leading to extracellular matrix remodeling

Graphical abstract

Authors

Swagata Adhikari, Vipin Singh, Sandhik Nandi, ..., Kundan Sengupta, Dimple Notani, Chandrima Das

Correspondence

chandrima.das@saha.ac.in

In brief

Adhikari et al. demonstrate that epigenetic regulator UBR7 functions in concert with EZH2 to modulate TGF-β signaling cascade by altering H3K27me3 landscape, thereby affecting the transcription of extracellular matrix genes. Thus, UBR7 plays a key role in stiffness-mediated metastasis suppression in triple-negative breast cancer.

Highlights

- UBR7 is a regulator of TGF-β/SMAD signaling axis

- UBR7 interacts with EZH2 to maintain H3K27me3 mark on chromatin

- UBR7 in concert with EZH2 suppresses TGFB1 gene, thereby affecting ECM gene transcription

- UBR7 impairs matrix stiffness, affecting the TNBC invasion *in vivo*

CellPress

Article

UBR7 in concert with EZH2 inhibits the TGF-β signaling leading to extracellular matrix remodeling

Swagata Adhikari,[1,3] Vipin Singh,[1,3] Sandhik Nandi,[1,3] Manorama Ghosal,[2,3,8] Nidharshan Sundar Raj,[7,8] Jayati Khanna,[6,8] Apoorva Bhattacharya,[1] Aindrila Kabiraj,[1,3] Atanu Mondal,[1,3] Madavan Vasudevan,[4] Dulal Senapati,[2,3] Himansu Roy,[5] Kundan Sengupta,[6] Dimple Notani,[7] and Chandrima Das[1,3,9,*]

[1]Biophysics and Structural Genomics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India
[2]Chemical Sciences Division, Saha Institute of Nuclear Physics, HBNI, 1/AF Bidhannagar, Kolkata 700064, India
[3]Homi Bhabha National Institute, Mumbai, India
[4]Theomics International Pvt Ltd., Bangalore, India
[5]Department of Surgery, Medical College and Hospital, Kolkata, West Bengal, India
[6]Chromosome Biology Lab (CBL), Indian Institute of Science Education and Research, Pune, Maharashtra, India
[7]National Centre for Biological Sciences, TIFR, Bangalore, India
[8]These authors contributed equally
[9]Lead contact
*Correspondence: chandrima.das@saha.ac.in
https://doi.org/10.1016/j.celrep.2024.114394

SUMMARY

The intricate interplay between resident cells and the extracellular matrix (ECM) profoundly influences cancer progression. In triple-negative breast cancer (TNBC), ECM architecture evolves due to the enrichment of lysyl oxidase, fibronectin, and collagen, promoting distant metastasis. Here we uncover a pivotal transcription regulatory mechanism involving the epigenetic regulator UBR7 and histone methyltransferase EZH2 in regulating transforming growth factor (TGF)-β/Smad signaling, affecting the expression of ECM genes. UBR7 loss leads to a dramatic reduction in facultative heterochromatin mark H3K27me3, activating ECM genes. UBR7 plays a crucial role in matrix deposition in adherent cancer cells and spheroids, altering collagen content and lysyl oxidase activity, directly affecting matrix stiffness and invasiveness. These findings are further validated *in vivo* in mice models and TNBC patients, where reduced UBR7 levels are accompanied by increased ECM component expression and activity, leading to fibrosis-mediated matrix stiffness. Thus, UBR7 is a master regulator of matrix stiffening, influencing the metastatic potential of TNBC.
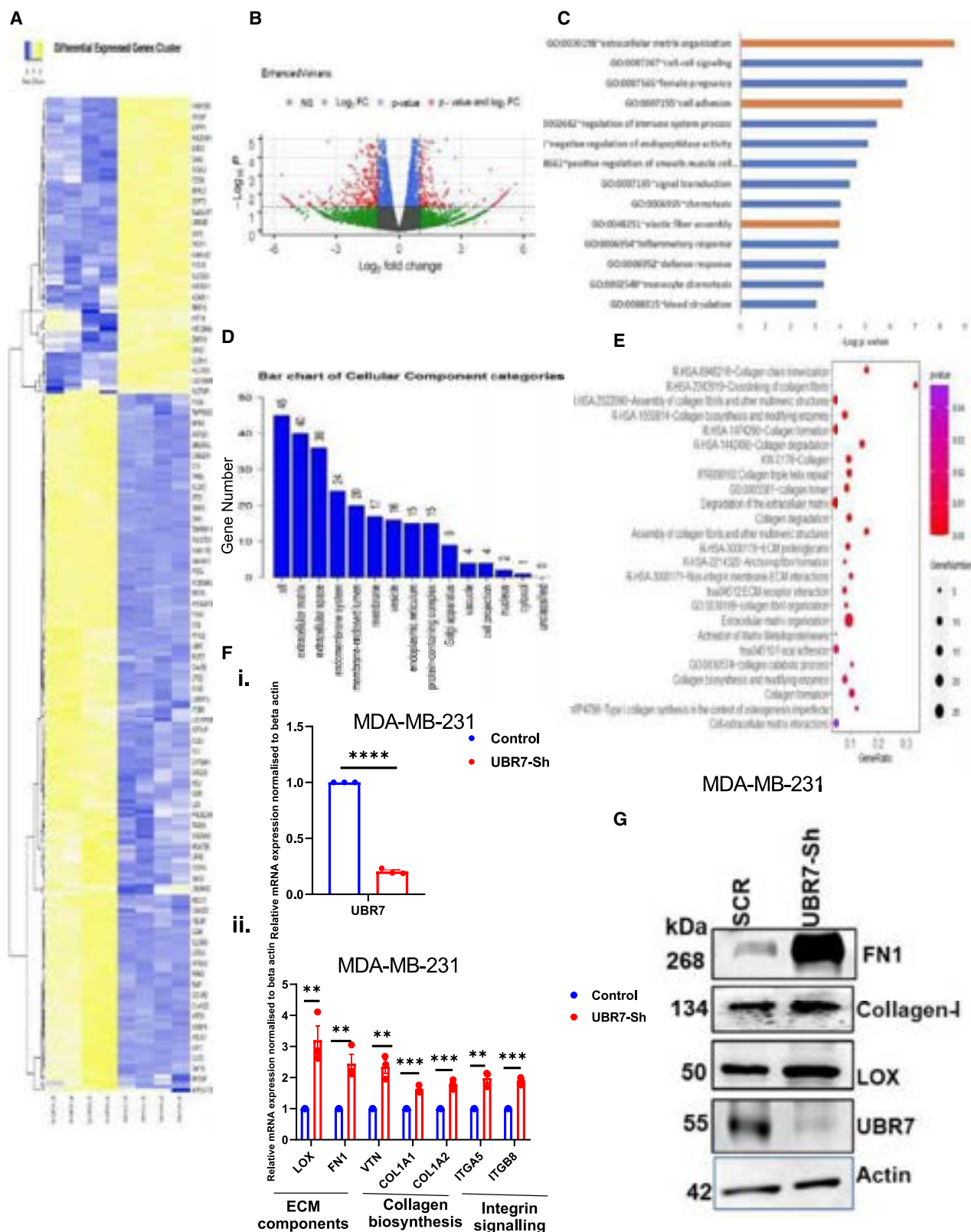
INTRODUCTION

Breast cancer continues to pose a formidable challenge as the second leading cause of cancer-related mortality among women, underscoring the pressing need for effective interventions. In the year 2023, an estimated 43,170 lives were claimed by breast cancer in the United States alone.[1] The complexity inherent in this disease, exacerbated by inter- and intra-tumor heterogeneity, present substantial challenges to the therapeutic regimen.[2] Among the different molecular subtypes, triple-negative breast cancer (TNBC) further accentuates the urgency of intervention as it exhibits a 40% recurrence rate in patients at stage I–III after conventional treatment.[3]

The intriguing disparity in tissue stiffness between tumors and healthy surroundings has emerged as a valuable tool for diagnosing breast cancer.[4,5] This mechanical property primarily stems from the remodeling of the extracellular matrix (ECM), orchestrated by intricate processes involving ECM deposition, post-translational modifications, and protease-mediated degradation.[6] Cumulatively, these alterations in ECM composition and architecture promote breast cancer malignancy by augmenting cell growth, survival, invasion, and metastasis.[7] In this context, accumulating evidence directly sheds light on the altered gene expression profiles of different ECM components, including collagen type I/type II/type III, fibronectin, and ECM-modifying enzymes such as lysyl oxidases LOX and LOXL2.[8–14] Notably, the deposition of type I collagen in the interstitial matrix directly reflects the tumor aggressiveness, as shown in the mammographic density, and elevates the risk of metastasis.[15–19] Notably, the pre-clinical breast cancer model supports mature LOX activity secreted from cancer cells that leads to collagen cross-linking augmenting the tensile strength of the matrix and preparing the pre-metastatic niche.[20] Thus, targeting LOX attenuates lung metastasis of breast cancer and sensitizes TNBC cells toward chemotherapy.[12,21] Thus, understanding the underlying mechanisms by which tumor cells reshape the ECM is pivotal in identifying targets for therapy.

A

Differential Expressed Genes Cluster

B

EnhancedVolcano

C

D

Bar chart of Cellular Component categories

E

MDA-MB-231

F

i.

MDA-MB-231

ii.

MDA-MB-231

G

MDA-MB-231

The transforming growth factor (TGF)-β signaling pathway emerges as a potent regulator of ECM deposition and remodeling in the tumor microenvironment. This pathway is a key mediator of breast cancer desmoplasia orchestrating the upregulation of the LOX enzyme and type I collagen, leading to an enrichment of fibrillar collagens.[22,23] TGF-β1, being abundant in TNBC, augments pro-metastatic collagen-I linearization.[24] Concurrently, TGF-β/SMAD signaling induces invasiveness in breast cancer spheroid models.[25] Consequently, TGF-β signaling correlates with poor overall survival in TNBC patients and drives chemoresistance by influencing dormancy and metastatic burden.[26,27] However, the role of TGF-β signaling as an autocrine regulator of ECM stiffness remains underexplored.

Increasing evidence underscores a correlation between epigenetic modifications and the progression of breast cancer. Notably, the decrease in the facultative heterochromatin mark H3K27me3 can be correlated with large tumor volumes.[28] Additionally, a propensity for reduced H3K27me3 enrichment is observed in estrogen-negative subtypes and cases with positive lymph nodes.[29] Recently, a multi-omic profiling study has shown the depletion of H3K27me3 and enrichment of H3K27ac signals in super-enhancers of oncogenes in TNBC, leading to their hyper-activation.[30] These reversible transitions between epigenetic states eventually support tumors in developing resistance to escape therapeutic pressure.[31,32] For instance, in TNBC patients, a "poised chromatin state" of the oncogenes marked by both H3K4me3 and H3K27me3 has been reported.[32,33] Notably, there is a decrease in H3K27me3 levels that triggers the transcription activation of these oncogenes following neoadjuvant chemotherapy and thereby enhancing drug tolerance that could be reversed by targeting epigenetic regulators.[32] Thus, the regulation of epigenetic transitions on various oncogenic pathway genes, including a dynamic switch between H3K27me3 and H3K27ac, needs further investigation.

Aberrant ubiquitinations are intricately associated with the etiology of malignant cancer.[34] In this context, mono ubiquitination of histone H2A and H2B plays an important role in regulating the transcription programs as well as DNA damage repair pathways and has important implications in cancer progression.[35] UBR7 belongs to the N-recognin family of mammalian E3 ubiquitin ligase,[36] which catalyze H2BK120 mono-ubiquitination by its atypical plant homeodomain (PHD) finger.[37] UBR7 has recently been shown to act as a histone chaperone promoting the reincorporation of post-nucleosomal histone H3 into chromatin in concert with NASP.[38] It acts as an epithelial to mesenchymal transition (EMT) suppressor in breast cancer,[37] a regulator of metabolic programs such as glycolysis in hepatocellular carcinoma,[39] and nucleotide biosynthesis in T cell acute lymphoblastic leukemia.[40] In the present study, we investigate the role of UBR7 in counteracting ECM stiffness and fibrosis in TNBC by targeting the TGF-β/SMAD axis. UBR7 is instrumental in recruiting the canonical polycomb repressor complex PRC2, thereby augmenting the H3K27me3 mark at the promoter and leading to the suppression of *TGFB1* and its downstream target genes of the ECM pathway, specifically in TNBC. Besides regulating the transcription mechanistically, UBR7 is intricately involved in the deposition and modifications of ECM. In this context, we further reveal that UBR7 is a key determinant of LOX activity and total collagen content from tumorspheres, *in vivo* mice model, and TNBC patient tissues. Thus, our study provides an insight into the role of UBR7 in epigenetic regulation of chromatin architecture, which has significant implications in effectively curbing matrix stiffness to reinstate an anti-metastatic phenotype.

RESULTS

UBR7 transcriptionally regulates ECM genes and TGF-β signaling cascade in TNBC

To gain insight into the role of UBR7 in TNBC, in the present study, we investigate the role of UBR7, an important epigenetic regulator, in the context of ECM stiffening in TNBC. We performed a comprehensive transcriptomics profiling of MDA-MB-231 control (SCR) compared to UBR7 stable knockdown MDA-MB-231 (UBR7-Sh) cells. The correlation plot is represented in Figure S1A. The differential data from the RNA sequencing (RNA-seq) are represented by heatmap analysis (Figure 1A) and volcano plot (Figure 1B). Gene Ontology (GO) analysis indicated that a significantly large subset of genes is associated with ECM organization using the Database for Annotation Visualization and Integrated Discovery (DAVID) bioinformatics tool (Figure 1C). Moreover, pathway analysis from differentially expressed genes (DEGs) is also represented by top 15 pathway hits by Kyoto Encyclopedia of Genes and Genomes (KEGG) (Figure S1B), web gene ontology annotation plot (WEGO) (Figure S1C), and a bar chart of cellular component analysis (Figure 1D). Notably, the involvement of different ECM remodeling pathways such as collagen biosynthesis, ECM-receptor interaction, etc., as identified by gene set enrichment analysis (GSEA) is represented by the dot plot (Figure 1E). These findings are further validated by qRT-PCR, confirming an increase in ECM gene

---

Figure 1. UBR7 regulates the transcription of ECM genes in TNBC

(A) Heatmap representation from RNA-seq analysis upon UBR7 knockdown (using UBR7-Sh1) shows the differentially expressed genes (log$_2$ fold change = 1) in MDA-MB-231, two biological and two technical replicates. Yellow and blue color indicate up- and downregulated, respectively.

(B) Volcano plots of differentially expressed genes scored from RNA-seq analysis in UBR7 knockdown. Green, red, and gray represent downregulated, upregulated, and non-significant, respectively.

(C) GO analysis from significant differential genes ($p \leq 0.05$) from RNA-seq reveals the most significant biological process based on $-\log$ (p value).

(D) Bar chart of cellular component categories represents the differential components along with the number of genes.

(E) Gene set enrichment analysis (GSEA), represented as a dot plot showing the number of genes and p values.

(F) qRT-PCR analysis of UBR7 (i) and ECM genes (ii) *LOX*, *FN1*, *VTN* (ECM components), *COL1A1* and *COL1A2* (collagen biosynthesis), and *ITGA5* and *ITGB8* (integrin signaling) from MDA-MB-231 expressing scrambled (SCR) and UBR7-Sh1 RNA.

(G) Immunoblots of SCR- and UBR7-Sh1-expressing cells from MDA-MB-231 to monitor the protein level changes of LOX, type I collagen, and FN1 using actin as loading control. Data information: in (Fi) and (Fii), representation is mean ± SEM (standard error of the mean), n = 3. The best three replicates are considered out of three independent experiments and p value is calculated using unpaired Student's t test. ns, non-significant; *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.
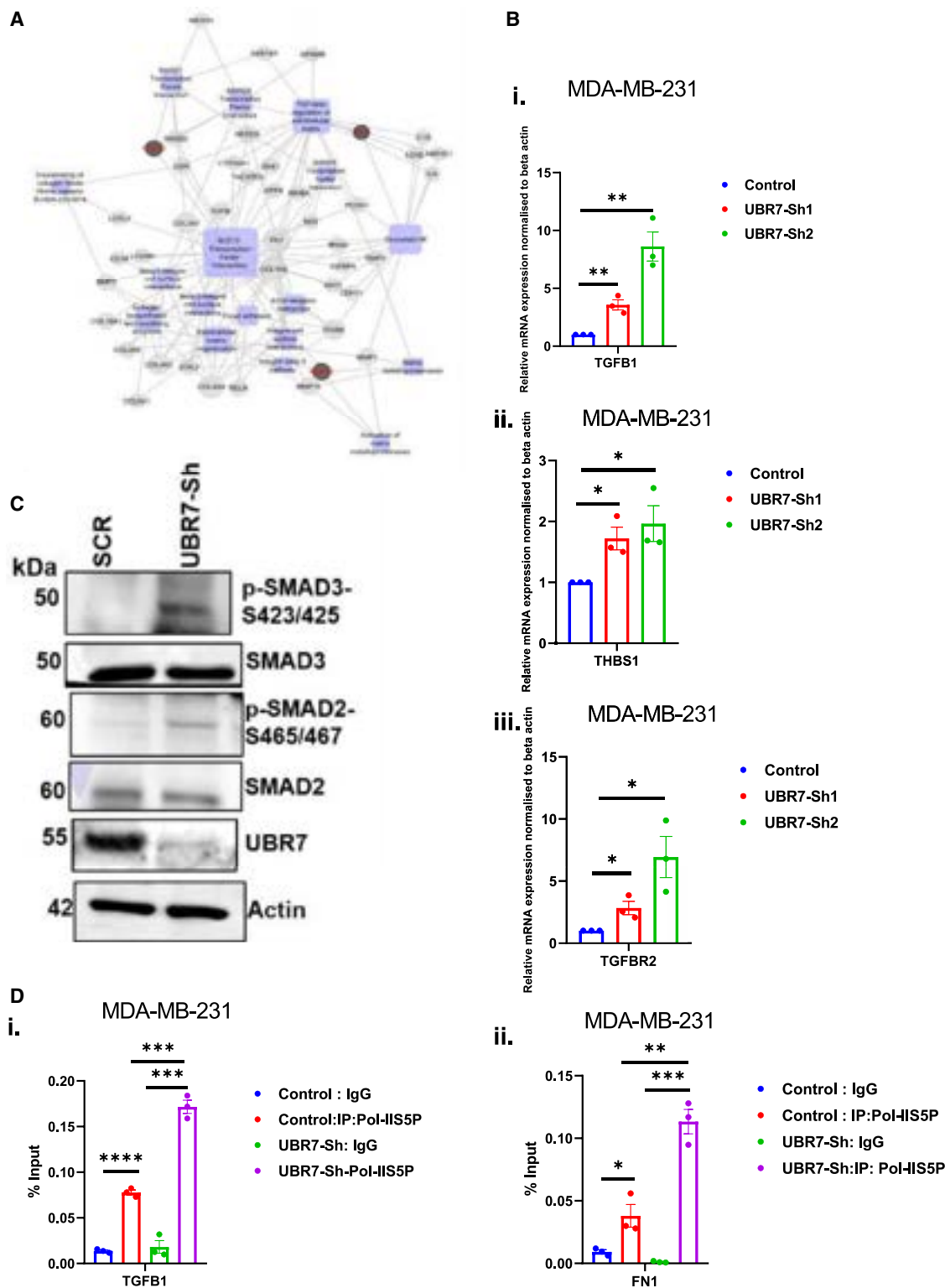
Figure 2. UBR7 transcriptionally modulates TGF-β signaling cascade in TNBC

(A) Network analysis from significant differential genes of RNA-seq using software Enrichr.

(B) qRT-PCR analysis of *TGFB1*, *TGFBR2*, and *THBS1* gene expression in MDA-MB-231 (i–iii) upon UBR7 knockdown by Sh1 and Sh2, *n* = 3, two technical replicates.

expression upon UBR7 knockdown with both short hairpin RNAs (shRNAs), and the expression of these ECM genes is rescued by UBR7-WT (Figure 1F i and ii) (Figure S1D). These genes encompass ECM components (LOX, FN1, VTN), collagen biosynthesis (COL1A1, COL1A2), and integrin signaling pathways (ITGA5 and ITGB8), all associated with matrix rigidity and metastasis in TNBC.[20,41] Western blot analysis further showed an increase in ECM protein levels (FN1, type I collagen, LOX) in UBR7-Sh cells (Figure 1G). Consistent results are observed in MDA-MB-468 cells, reaffirming the role of UBR7 in regulating ECM genes (Figure S1E i–iv, Figure S1F).

These findings prompted a deeper exploration into the upstream regulatory signaling pathways governing ECM remodeling genes. Among these, TGF-β signaling emerged as a major and enriched pathway implicated in ECM remodeling, known to promote stemness and EMT in TNBC.[42] Network analysis of the DEGs picked from RNA-seq using Enrichr demonstrates that deregulation of the TGF-β pathway is dependent upon UBR7 (Figure 2A). Similarly, GO pathway enrichment analysis represented by the dot plot also hinted toward TGF-β signaling from UBR7-knockdown RNA-seq datasets (Figure S2A). This finding aligns with previous studies indicating that dense collagen network and the production of different ECM components in the tumor microenvironment such as fibronectin and integrins are induced by the TGF-β-phospho-Smad signaling axis.[43] Upon loss of UBR7 in MDA-MB-231 cells, upregulation of positive regulators, TGFB1, THBS1, and TGFBR2 (Figures S2B and 2B i–iii), indicated the activation of TGF-β signaling. A similar regulation was observed in MDA-MB-468 cells (Figure S2C i–iii). Since the canonical SMAD signaling induced by TGF-β triggers aggressiveness in TNBC,[44] phosphorylation of SMAD-2 at S465/467 and SMAD-3 at S423/425 was examined upon UBR7 knockdown. This analysis confirms their increased phosphorylation in MDA-MB-231 cells, indicating activation of the canonical Smad-dependent TGF-β pathway (Figure 2C). A loss of UBR7 led to increased recruitment of RNA pol S5P to its target genes (Figure 2D i and ii, and Figure S2D). Significantly, the occupancy of initiating RNA polymerase II Phospho-S5 (RNA pol S5P) on TGFB1 and its downstream ECM target genes (FN1 and LOX) was found to be UBR7 dependent (Figure 2D i and ii, and Figure S2D). Thus, the increased expression of the genes upon knockdown of UBR7 could be connected to the emergence of transcription initiation events.

Collectively, these findings indicate that UBR7 is a master orchestrator of ECM genes and TGF-β signaling pathway, thereby contributing to the intricate mechanism underpinning TNBC progression.

UBR7/PRC2 co-complex is instrumental in suppressing TGF-β pathway genes

To understand the mode of regulation of the TGF-β signaling pathway by UBR7, a comparative analysis of its interactome in a normal proliferating epithelial cell line MCF10A and a TNBC

cell line MDA-MB-231 was identified by immunoprecipitation (IP) followed by mass spectrometry (MS) study. From the differential proteomics, the correlation tree and principal-component analysis (PCA) plot are represented in Figure S3A i and ii, respectively. The interacting partners of UBR7 from both MCF10A and MDA-MB-231 cells, belonging to different categories based on their molecular function and involvement in cellular pathways, were delineated through dot plot analysis and represented compositely in Figure 3A i. Employing this approach, we uncovered a distinctive set of interacting partners in MDA-MB-231 cells compared to MCF10A, depicted through the gene regulatory network in Figure 3A ii. Some of the robust interacting network hubs are epigenetic regulation of gene expression, chromatin organization, RNA polymerase I/II transcription, chromatin remodeling, etc. (Figure 3A ii). Remarkably, while in MCF10A cells several proteins involved in transcription programs (TBP, RBBP7, PolR2H, etc.) were identified as interactors of UBR7, in MDA-MB-231 cells there were no significant interactions with them (Figure 3A ii). Notably, PRC2 repressor complex (enhancer of zeste homolog 2, or EZH2, being the methyltransferase subunit) emerged as a distinct interactor of UBR7 in MDA-MB-231 cells, while NuRD repressor complex (HDAC1 deacetylase and MTA2 subunits) could be robustly detected to associate with UBR7 in MCF10A cells (Figure 3A ii). This finding has been supported as UBR7 has differential interacting partner preferences in MCF10A and MDA-MB-231, represented in the Venn diagram, which showed no common interactors between these two cell lines (Figure S3A iii). Remarkably, in MDA-MB-231 cells, EZH2 showed a high interaction score with UBR7. This prompted us to investigate further, and co-immunoprecipitation (coIP) confirmed EZH2 and embryonic ectoderm development (EED) interact with UBR7 (Figure 3B, i). Similar reciprocal coIP confirmed that UBR7 forms a complex with EZH2, EED, and, SUZ12 (Figure 3B ii).

Next, we sought to determine whether UBR7 in association with PRC2 complex goes to the TGFB1 gene promoter. It has been previously reported that TGF-β pathway genes are suppressed by H3K27me3 modification mediated by PRC2 complex.[45] We observed that UBR7 gets recruited to the TGFB1 gene promoter (Figure 3C i) and not to THBS1 and TGFBR2 (Figure S3B i and ii). Further, it was observed that the recruitment of EZH2 was notably reduced on the gene promoters upon UBR7 knockdown (Figure 3C ii). These findings unveil a coordinated interplay between UBR7 and the PRC2 complex in stabilizing the H3K27me3 mark on the chromatin of the TGFB1 gene, which was observed as the H3K27me3 mark occupancy goes down in the loss of UBR7 (Figure 3C iii), leading to the suppression of the downstream signaling pathway. These dynamic events, in turn, could exert an intriguing impact on the expression of the ECM genes, potentially linking the suppression of TGF-β signaling to ECM remodeling.

It has been previously observed that UBR7 could modulate the H2BK120Ub levels globally.[37] UBR7 knockdown in MCF10A

(C) Immunoblots represent protein levels of TGF-β pathway markers-phospho-SMAD3 S423/425, SMAD3, phospho-SMAD2 S465/467, and SMAD2 using actin as loading control from the cells expressing control (SCR) and UBR7-Sh2.

(D) ChIP studies show the enrichment of RNA polymerase II-S5P on the promoter of TGFB1 and FN1 genes from the cells expressing control and UBR7-Sh1 (i and ii); $n = 3$, two technical replicates. Data information: in (B) and (D), error bars indicate SEM. Unpaired Student's t tests. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.
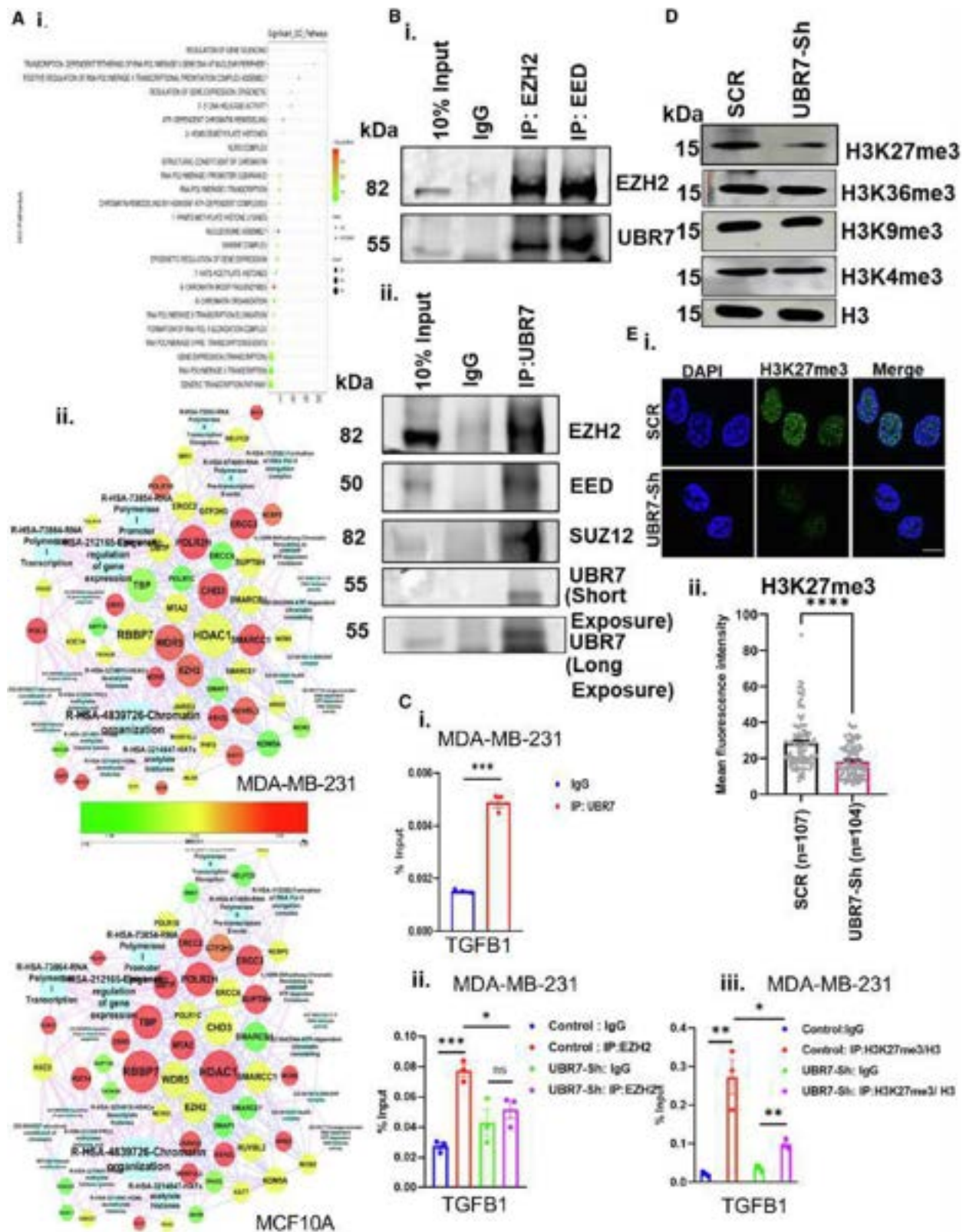
(legend on next page)

cells reduced H2BK120Ub, while overexpression in MDA-MB-231 cells elevated H2BK120Ub in chromatin.[37] While screening the global alterations in other histone modification status, we serendipitously observed that UBR7-Sh cells showed a significant global reduction in the facultative heterochromatin mark H3K27me3 as compared to the control (SCR) cells. Remarkably, some of the other methylation marks, such as H3K4me3, H3K9me3, and H3K36me3, did not show significant alteration in UBR7-Sh cells (Figure 3D). These results could be connected in retrospect as UBR7 emerged as a robust interactor of PRC2 in MDA-MB-231 cells (Figure 3B i and ii) that maintain H3K27me3 levels. Interestingly, we observed a concomitant increase in the activation mark H3K27ac in UBR7-Sh cells (Figure S3C). In order to authenticate the findings further, immunofluorescence of H3K27me3 (Figure 3E i and ii) and H3K27ac (Figure S3D i and ii) was performed in control (SCR) and UBR7-Sh cells, suggesting a nuanced mechanism of methylation-acetylation switch regulated by UBR7 in TNBC.

Taken together, the above findings converge on the fact that UBR7 interacts with the PRC2 complex that alters the global H3K27me3 status. Hence, we hypothesized that the absence of UBR7 could affect the H3K27me3 occupancy on chromatin genome-wide, which was sought next.

UBR7 orchestrates epigenetic alterations to suppress ECM gene transcription

Being a concordant regulator of TGF-β pathway and ECM gene transcription, UBR7 simultaneously leads to a dramatic transition in H3K27me3 and H3K27ac marks globally. In this context, based on our above findings, we hypothesized UBR7-dependent epigenetic transitions as a key mediator in the transcription regulation of ECM genes via TGF-β signaling. In order to understand how UBR7 could affect the global chromatin landscape, chromatin IP (ChIP) sequencing (ChIP-seq) on both MDA-MB-231 control and UBR7-Sh cells was performed. Replicate reproducibility of the ChIP-seq results using deep tools has been represented in Figure S4A. Profile plot, quantitative peak analysis summary, recruitment ratio across different genomic elements, and distribution of peaks based on gene types are represented in Figures S4B–S4E respectively. Remarkably, a drastic reduction of H3K27me3 level on chromatin in UBR7-Sh cells as compared to MDA-MB-231 control was observed as represented by an occupancy heatmap (Figure 4A). H3K27ac levels on chromatin also showed partial alteration in the absence of

UBR7 (Figure S4F), although the changes in the H3K27me3 status were more prominent. In order to appreciate the alteration in H3K27me3 status across the different chromosomes, we analyzed the chromosomal coverage plots from MDA-MB-231 expressing SCR and UBR7-Sh cells. The results indicated that UBR7 positively regulates the enrichment of the H3K27me3 mark across all chromosomes (Figure 4B).

In a hunt to identify the different subset of genes that are epigenetically regulated by UBR7, based on the H3K27me3 and H3K27ac enrichment, seven classes (encompassing both the post-translational modifications [PTMs] being up-, down-, or reciprocally regulated or unaffected upon knockdown) were annotated to categorize the chromatin states. Among these seven classes, three emerged as a focal point in the study based on the changes in the chromatin landscape, regulated by UBR7. In the absence of UBR7, class I encoded genes (protein-coding gene no. 1066) with a reduction in H3K27me3 mark, class II comprised genes (protein-coding gene no. 1586) with solely H3K27ac induction independent of H3K27me3, and class III encompassed genes (protein-coding gene no. 30) that showed reduced H3K27me3 coupled with H3K27ac enrichment and are represented in the Venn diagram (Figure 4C). Class I–III genes were subsequently organized into different biological pathways (Figures 4D and S4G i and ii), respectively. Remarkably, an extensive subset within these gene clusters, responsive to UBR7-dependent alterations in epigenetic marks, emerged as integral components of various ECM remodeling pathways, as highlighted in the dot plot analysis (Figures 4D and S4G i and ii). Notably, class I and II genes had pathways associated with the cell surface, extracellular region, cell-cell adhesion, extracellular space, cell adhesion, ECM, and its organization among the top 10 hits. Class III genes were fewer and could be grouped into the pathways associated with DNA-templated transcription, extracellular exosome, signaling by Wnt, RNA Pol-II transcription, and generic transcription pathway. Next, the functional gene network analysis with the ECM pathway genes of the class I subset was performed as it had the maximum representation including that of cell adhesion, ECM, membrane, cell surface, ECM-receptor interaction, cell migration, and metastasis (Figure 4E).

Following the comprehensive network analysis, peak profiles/tracks on the ECM pathway genes that were identified to be exclusively H3K27me3 enriched and UBR7 dependent (belonging to the class I category) were determined (Figures 4F


Figure 3. UBR7 interacts with PRC2 complex

(A) Representative dot plot using IP-MS data from MCF10A and MDA-MB-231 cells showing the number of interactors belonging to different molecular functions and cellular process categories (i). Representative bubble plot from gene regulatory network analysis of IP-MS data in MDA-MB-231 and MCF10A using Cytoscape software (ii). Interactions are scored through gradient. Green to red indicates lesser to robust bonding interactions.

(B) CoIP of UBR7 from MDA-MB-231 lysate showing UBR7 for pull-down of EZH2 and EED through immunoblots (i) and reverse pull-down of UBR7 showing UBR7, SUZ12, EED, and EZH2 through immunoblots (ii).

(C) Bar plot represents qPCR enrichment of UBR7 on *TGFB1* gene in MDA-MB-231 cells (i), EZH2 on the *TGFB1* gene in presence and absence of UBR7 (UBR7-Sh1) (ii), and H3K27me3 on *TGFB1* gene in presence and absence of UBR7 in MDA-MB-231 cells (iii) scored through ChIP. Three biological replicates and two technical replicates.

(D) Immunoblot analysis of H3K27me3, H3K36me3, H3K9me3, and H3K4me3 using H3 as loading control from MDA-MB-231 cells stably expressing SCR and UBR7-Sh1.

(E) Representative images of immunofluorescence from MDA-MB-231 cells stably transfected with SCR (*n* = 107) and UBR7-Sh1 (*n* = 104) show nucleus via DAPI staining and H3K27me3 via green fluorescence (i). Scale bar, 10μm. Mean fluorescence intensity is quantified (ii). Number of independent biological replicates is two. Data information: in (C) and (E), unpaired Student's t tests. Error bars indicate SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$."
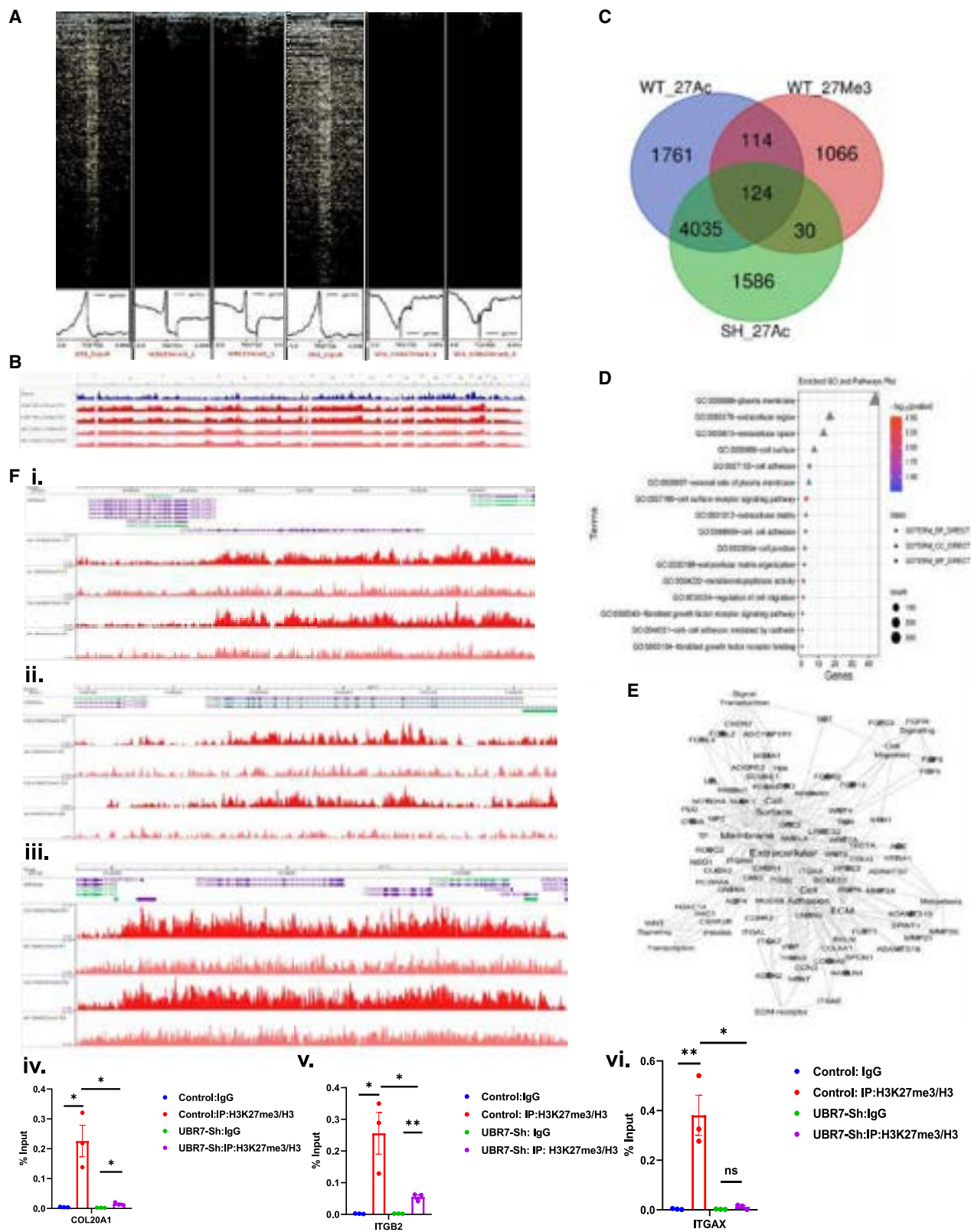
i–iii, S5A, and S5B) and subsequently validated by ChIP assays (Figure 4F iv–vi). Indeed, we observed that these class I genes, including *COL6A5*, *COL20A1*, *ITGB2*, *ITGAX*, *WNT4*, *TGFB1*, and *TGFB3*, were enriched in H3K27me3, which gets substantially reduced in the absence of UBR7 (Figures 4F i–vi, and S5A–S5D). The reduction in the H3K27me3 mark at the gene promoters could be connected with increased RNA Pol S5P recruitment, as observed previously.

These observations indicate that UBR7 is a versatile transcription regulatory protein that is instrumental in maintaining the H3K27me3 status of the ECM genes, thereby leading to transcription repression. This could directly manifest in the alteration of the matrix stiffness, thereby affecting the metastatic potential of the cancer cells.

UBR7 regulates the functional modification of ECM proteins

UBR7 emerges as an important regulator of ECM genes, which translates into discernible biochemical changes within the ECM milieu. It is well known that collagen fibers in the ECM get post-translationally modified, which alters its elasticity, thereby affecting the metastatic potential of the cancer cells.[46] Lysyl oxidase is one such enzyme that modifies collagen fibers.[46] In order to investigate the role of UBR7 in the biochemical changes in ECM, we established 3D-spheroid culture from control (SCR) and UBR7-knockdown (UBR7-Sh) cells. The transition of cultured adherent cells to 3D spheroids led to an intriguing tapestry of molecular changes where canonical cancer stem cell marker ALDH1 and drug-resistance determinant MRP1 were expressed aligning seamlessly with existing literature[47,48] (Figure S6A). Subsequently, the expression of the ECM protein (FN1 and LOX) was compared between SCR and UBR7-Sh spheres. Remarkably, like the adherent cells, expression of FN1 and Pro-LOX was significantly induced in UBR7-Sh as compared to SCR spheres, as observed by western blotting (Figure 5A i and ii). The percentage of FN1-positive cell population also showed a heightened expression by flow cytometry in UBR7-Sh as compared to SCR cells (Figure 5B i–iii). Significantly, immunofluorescence using antibodies of collagen-I and FN1 further confirmed an induction of these ECM markers in UBR7-Sh as compared to SCR spheres and was quantified by correlated total cell fluorescence (CTCF) measurements (Figure 5C i–iii and S6B i and ii).

With UBR7 identified as a crucial regulator of LOX gene transcription, coupled with established links between elevated LOX activity and unfavorable prognosis in TNBC,[17] we investigated the role of UBR7 in functional ECM modifications. Remarkably, the levels of Pro-LOX and glycosylated LOX were reduced upon UBR7 wild-type overexpression in MDA-MB-231 cells (Figure 5D). In order to monitor the functional consequence of the induction of LOX upon UBR7 knockdown, LOX activity assay was performed in MDA-MB-231 and MDA-MB-468 adherent cells as well as MDA-MB-231 spheroids. Importantly, in the extracellular milieu, secreted Pro-LOX is proteolytically processed by bone morphogenetic protein-1 (BMP1) to form a mature, enzymatically active form of LOX that catalyzes the cross-linking of collagen and elastins.[49] As a result of increased accumulation, an elevated activity of LOX enzyme can be expected from the sphere supernatant in UBR7 knockdown condition. The LOX activity measurements were performed from both cell lysates and supernatants in SCR control and UBR7-Sh cells. A schematic representation of the assay protocol from the MDA-MB-231 spheres is represented in Figure 5E i. Remarkably, in the spheres (Figure 5E ii and iii) as well as adherent cells (Figures S6C and S6D), the LOX activity was significantly enhanced in both cell lysates and supernatants from UBR7-Sh as compared to SCR in MDA-MB-231 as well as in MDA-MB-468 cells, whereas the overexpression of UBR7 diminished the activity of lysyl oxidase (Figure S6E).

With LOX activity intricately linked to collagen cross-linking within the ECM, UBR7 emerges as a seminal regulator of ECM modifications, thereby inhibiting LOX activity. Thus, the absence of UBR7 augments the LOX activity that could be implicated in alteration in the matrix stiffening, thereby having an important consequence in the context of TNBC metastasis.

UBR7 regulates ECM stiffness and suppresses TNBC invasion

UBR7 is an important regulator of ECM gene expression as well as the biochemical modification of ECM proteins, such as LOX. Intriguingly, LOX, a renowned ECM remodeler, stands as a harbinger of cancer cell invasiveness and metastasis.[20,21] Thus, a compelling hypothesis is UBR7 depletion might trigger ECM architectural changes, particularly an upsurge in ECM stiffness, thereby leading to a favorable environment for TNBC invasion and metastasis.

To delve into the altered ECM mechanics, we cultured MDA-MB-231 cells on ECM-coated gels for a span of 10 days, followed by matrix decellularization, as represented in Figure 6A. The stiffness of the decellularized matrix was evaluated by atomic force microscopic (AFM) studies. AFM height, peak force imaging, and Derjaguin Muller (DM) modulus analyses vividly captured

Figure 4. UBR7 stabilizes the facultative heterochromatin mark H3K27me3 on ECM gene promoters

(A) Occupancy heatmap of H3K27me3 from ChIP-sequencing input and IP in MDA-MB-231 SCR and UBR7-Sh1 cells, *n* = 2, technical replicates two.

(B) Coverage plot showing the global distribution of H3K27me3 from ChIP-sequencing in MDA-MB-231 expressing SCR and UBR7-Sh1 cells.

(C) Venn diagram from the global analysis of ChIP-seq data, categorized into class I- Reduction of H3K27me3 in loss of UBR7, class II- Reduction of H3K27me3 and enrichment of H3K27ac in loss of UBR7, and Class III- Enrichment of H3K27ac in absence of UBR7.

(D) Dot plot representation from the subset of genes selected through the Venn diagram that undergo UBR7-dependent enrichment of H3K27me3 (class I), showing their cellular function categories.

(E) Network analysis of class I gene subsets. Black encodes the gene name where enrichment is affected and white represents the genes with no change, the white box indicates the functional category.

(F) Integrative Genomics Viewer (IGV) representation of H3K27me3 CHIP-seq tracks on the genes named *COL20A1*, *ITGB2*, and *ITGAX* (i, ii, and iii) along with the validation of ChIP assays (iv, v, and vi) from control and UBR7-Sh1. *n* = 3, two technical replicates. Data information: in (F), error bars indicate SEM. *p* values were calculated by unpaired Student's t tests. *p < 0.05 and **p < 0.01.
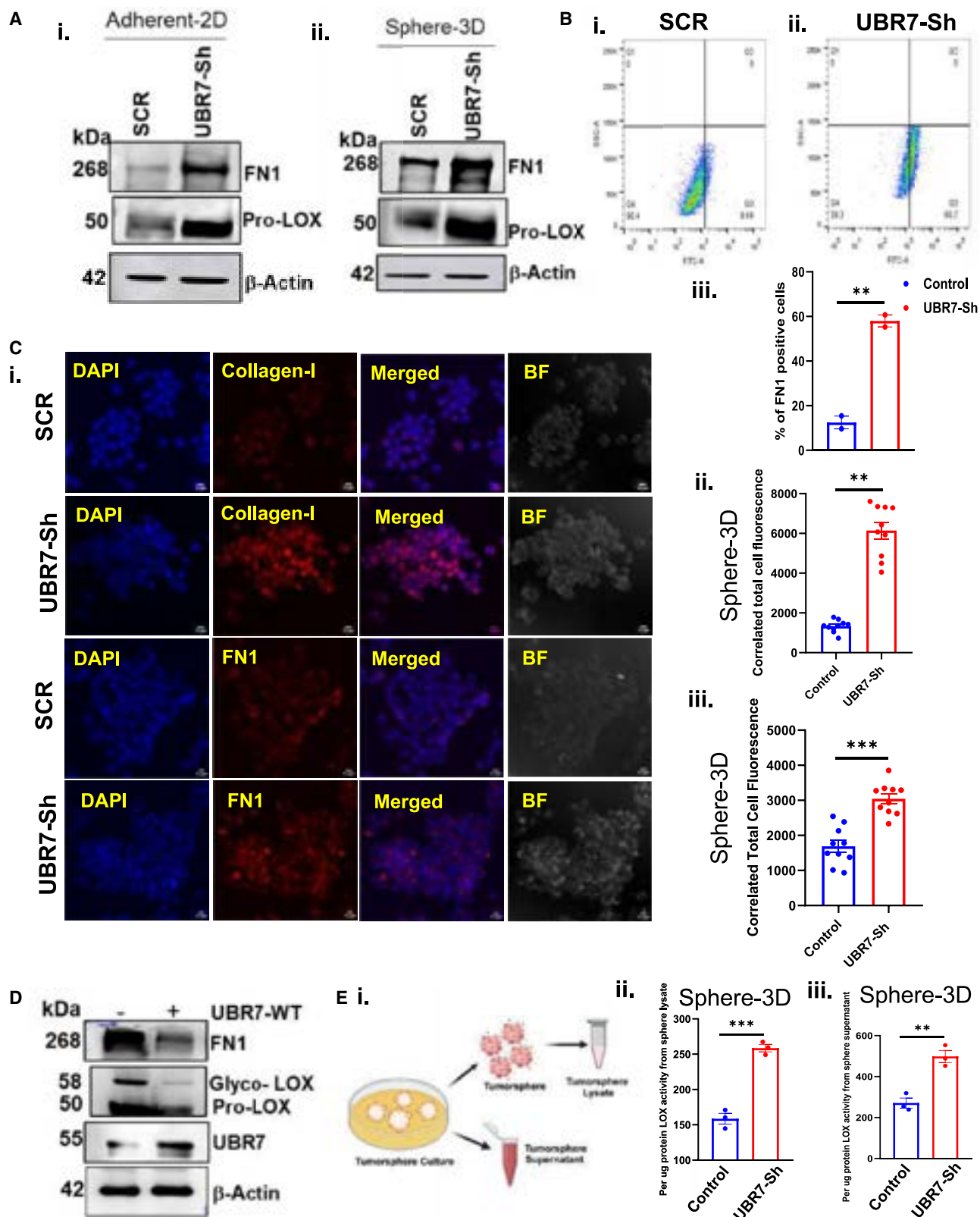
(legend on next page)

the escalating stiffness in the ECM generated by UBR7-Sh as compared to SCR cells (Figure 6B i–iv and S7A i–iv). AFM topographic height images of MDA-MB-231 cells overexpressing SCR and UBR7-Sh are represented in Figure S7A (i and ii) and their corresponding peak force images are shown in Figure S7A (iii and iv). The alteration in the Derjaguin Muller Toropov (DMT) modulus of UBR7-Sh as compared to SCR cells is represented in Figure 6B (i and iii). The corresponding depth histograms of DMT modulus from quantitative nanomechanics (QNM) analysis are represented in Figure 6B (ii and iv). The overall differences in DMT modulus between SCR and UBR7-Sh cells are reflected in a tabular format (Figure 6C). However, another AFM study reinstated that UBR7-mediated stiffness modulation of ECM is a direct result of the deregulation of TGF-β signaling axis. In this context, AFM height, DMT modulus, and DMT histogram of the UBR7-Sh cells and UBR7-Sh cells treated with TGF-β inhibitor were analyzed (Figures S8Ai–iii and S8Bi–iii). There is a reduction of the stiffness (approximately 25–30 kPa to 5–15 kPa) in the case of UBR7-Sh cells while inhibiting TGF-β signaling, which clearly indicates the regulatory axis of UBR7 is TGF-β signaling (Figures S8Ai-iii and S8Bi-iii). A stiff matrix directly activates EMT and promotes intravasation of cancer cells and metastasis.[6] Remarkably, the expression of the mesenchymal marker Snail could be partly compensated by UBR7 overexpression in MDA-MB-231 cells with heightened LOX expression (Figure S7B i). Concomitant treatment with DZNep (EZH2 inhibitor) led to an alleviation of this repression and Snail expression was restored (Figure S7B i). Similarly, DZNep treatment can promote the expression of Snail even in UBR7-overexpressed conditions only (Figure S7B ii). Subsequently, the invasion assay elucidated that UBR7 could rescue the enhanced invasiveness of LOX-overexpressed MDA-MB-231 cells (Figure 6D i–iv). Further, DZNep treatment in a UBR7-overexpressed background promoted the invasiveness of MDA-MB-231 cells (Figure 6E i–iv).

These results collectively indicate that UBR7 efficiently regulates matrix stiffness and the consequent invasive potential of TNBC. Remarkably, UBR7 can alleviate the LOX-mediated cellular invasiveness. Inhibition of EZH2 function can antagonistically affect UBR7-mediated suppression of the TNBC invasion. Thus, a crosstalk between UBR7/LOX, as well as UBR7/PRC2 complex, culminates in TNBC metastatic potential.

UBR7 suppresses ECM components and regulates the matrix stiffness *in vivo* as well as in TNBC patients

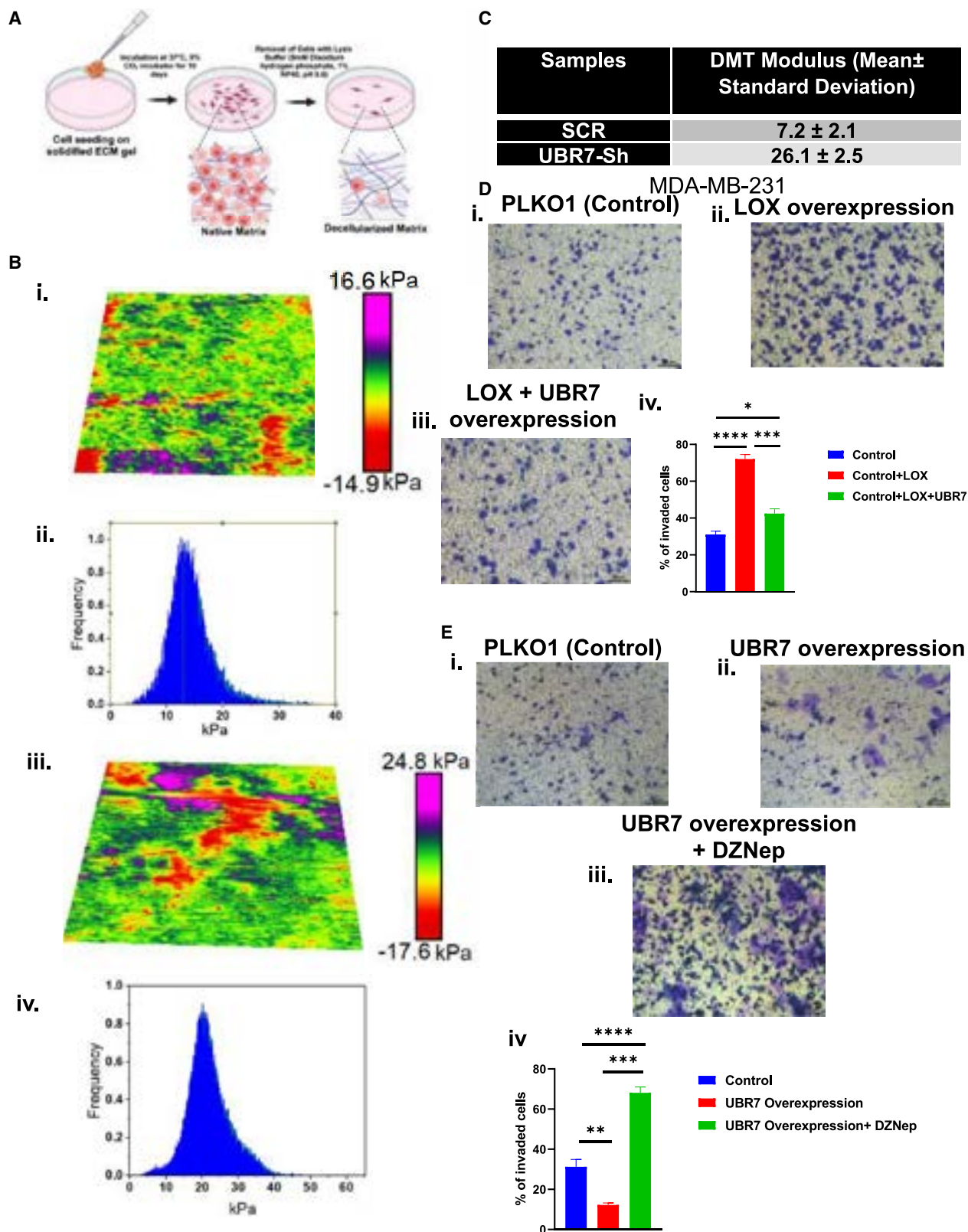Next, we validated our findings in the *in vivo* syngeneic mice model where we observed that UBR7 loss induces tumor growth as compared to control (empty vector) (Figures S9A and S9B). Importantly, both LOX activity and total collagen content seemed to be significantly increased in the tumors derived from UBR7-Sh-expressing 4T1 cells in comparison to control, whereas UBR7-WT-overexpressing tumors show a significant drop for the same (Figures S9C, S9D, and S10A–S10D). Furthermore, picrosirius staining confirmed that collagen content is significantly higher in UBR7-Sh as compared to control tumors, represented in Figures S10E and S10F. We finally aimed to unveil the correlation of UBR7 expression with ECM reshaping, in particular with ECM deposition and modifications in TNBC patients.

We performed total collagen measurement as well as estimated LOX activity from TNBC patients as represented in Figure S11A. LOX family oxidases are known to catalyze cross-linking of collagen and elastins and are highly expressed in invasive breast tumors.[50] Inhibition of LOX by β-Aminopropionitrile (BAPN) causes a reduction in the deposition of total collagen along with a stark decrease in collagen cross-liking in breast tumor, which consequently demonstrates the role of LOX in mediating tumor fibrosis.[18,21] Since UBR7 is a suppressor of *LOX* and *COL1A1/COL1A2* genes, we hypothesized that, in TNBC patients, the expression of UBR7 could be inversely regulated with LOX and total collagen content. Intriguingly, we observed an anticorrelation between UBR7 expression and total collagen content as well as LOX activity from TNBC patients (Figures S11B and S11C). Notably, solid tumors evolve as they enrich their ECM with collagen and other ECM components, thereby exerting a drastic change in their microenvironment as compared to normal tissue. In this context, ECM remodelers such as LOX determine the matrix stiffness by promoting fibrillar collagen and fibronectin content, thereby leading to altered mechanosignaling assisted by integrins, resulting in cancer progression.[6,12] We further monitored by immunohistochemistry (IHC) the expression of UBR7 and ECM markers from the parallel sections obtained from the same TNBC patients. We observed that, in low-UBR7-expressing patient tissues, collagen-1 and FN1 were elevated in expression in multiple TNBC patients (Figures S11D, S11E, and S12A–S12C). Furthermore, *in silico* expression analysis of UBR7 with different ECM components (COL1A1 and COL1A2) was performed using the R2 Genomics Analysis and Visualisation Platform database. In a cohort of 201 TNBC patients (GSE202203), a low UBR7 expression showed a concomitant high expression of COL1A1 and COL1A2 (Figure S12D i and ii).

Further, we wanted to investigate whether the changes in expression of UBR7 and the ECM markers are specific to clinical

---

Figure 5. UBR7 modifies ECM in TNBC primary spheres

(A) Immunoblot analysis for Pro-LOX, collagen-I, and FN1 for adherent MDA-MB-231 (i) and tumorsphere from SCR and UBR7-Sh1 cells (ii).

(B) Fluorescence-activated cell sorting (FACS) analysis showing FN1+ cells in (i) MDA-MB-231 expressing SCR and (ii) MDA-MB-231 expressing UBR7-Sh1 conditions. (iii) Percentage of FN1+ cells are quantified and represented graphically. *n* = 2.

(C) (i) Representative immunofluorescence images of tumorspheres from SCR-expressing and UBR7 stable knockdown cells (UBR7-Sh1) represent collagen-I and FN1 staining in red fluorescence. Correlated total cell fluorescence (CTCF) is quantified for collagen-I (ii) and FN1 staining (iii). For quantification, *n* = 10.

(D) Representative immunoblot analysis of FN1, glycosylated LOX, and Pro-LOX from MDA-MB-231 cells expressing PLKO1 and UBR7 wild type using actin as a loading control.

(E) Pictorial representation of collecting tumor spheres and their supernatant for measuring LOX activity (i). Performance of LOX activity assay from lysate and supernatant of tumorsphere expressing SCR and UBR7-Sh1 (ii and iii). *n* = 3, two technical replicates. Data information: in (B), (C), and (E), error bars represent SEM. Unpaired Student's t tests. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

A

B

i.

16.6 kPa

-14.9 kPa

ii.

iii.

24.8 kPa

-17.6 kPa

iv.

C

Samples	DMT Modulus (Mean± Standard Deviation)
SCR	7.2 ± 2.1
UBR7-Sh	26.1 ± 2.5

D MDA-MB-231

i. **PLKO1 (Control)**

ii. **LOX overexpression**

iii. **LOX + UBR7 overexpression**

iv.

% of invaded cells

*
**** ***

■ Control
■ Control+LOX
■ Control+LOX+UBR7

E

i. **PLKO1 (Control)**

ii. **UBR7 overexpression**

iii. **UBR7 overexpression + DZNep**

iv

% of invaded cells

****
***
**

■ Control
■ UBR7 Overexpression
■ UBR7 Overexpression+ DZNep

(legend on next page)

grades of TNBC or not. In order to understand that, commercial tissue microarray slides for TNBC patients were subjected to IHC studies where grade 2 and grade 3 categories of patient tissues were stained for three individual proteins UBR7, fibronectin 1, and type I collagen. Depending upon the staining intensities of UBR7 and fibronectin 1, we demarcated the patients (belonging to grade 2 and grade 3) into three categories, weak positive (<20% positive area), moderate positive (20%–25% positive area), and strong positive (>25% positive area) (Figure 7A). Similarly, type I collagens from grade 2 and grade 3 tumors are shown along with the quantification in Figures S13A and S13B. Remarkably, the weak positive staining of UBR7 is around 39% in grade 2, which translates into 58% in grade 3, whereas the weak positive staining for fibronectin 1 decreases from grade 2 to grade 3 from around 46% to 30%. Further, the moderate positive staining of fibronectin 1 also elevates from 30% to 49% from grade 2 to grade 3 tumors. This clearly establishes that the expression of fibronectin 1 is significantly elevated in grade 3 tumors, where UBR7 is low (Figures 7A and 7B). Representative IHC images showed fibronectin 1 and UBR7 staining from grade 2 and grade 3 tumor sections (Figures 7C and 7D, respectively). The percentage positive area calculation re-established that a significant difference is observed in FN1 and UBR7 distribution only in grade 3 tumors (Figures 7E and 7F). These revelations position UBR7 as a crucial therapeutic target; augmenting its expression could potentially mitigate metastasis in TNBC.

DISCUSSION

The tumor microenvironment undergoes a profound alteration during cancer progression, prominently manifested by modifications in the ECM architecture. The shift toward increased elastic modulus, with breast tumor tissues exhibiting a remarkable rise over 1,000 Pa compared to the normal range of 500–1,000 Pa,[22,51] is characteristic of this phenomenon. This transformation is intricately linked to an array of changes encompassing collagen and elastin cross-linking, enzymatic modifications by lysyl oxidase and lysyl hydroxylase family, and molecular compositions such as glycosylation—a composite matrix that stands in stark contrast to their non-malignant counterparts.[51] In this backdrop, the tumor cells, through autocrine and paracrine signaling, invoke matrix stiffening, which in turn modulates the phenotypic plasticity of cancer cells. This bidirectional interplay fuels oncogenic signaling pathways that align with the hallmarks of cancer, including inducing angiogenesis and activating invasion and metastasis.[19,52] While various transcription factors

have been identified in orchestrating ECM remodeling, the epigenetic landscape governing these changes remains a fascinating, albeit uncharted, territory.

Recently, UBR7 has been established as a histone H2B mono-ubiquitin ligase that activates CDH4 expression leading to EMT and metastasis suppression.[37] The present study unveils a hitherto unexplored role of UBR7 in orchestrating ECM remodeling—an intrinsic driver of EMT and metastasis. Cooperation between tumor cell plasticity and ECM confinement primes the movement of the cancer cells through the matrix, ultimately leading to the initiation of the metastatic cascade.[53] Emerging as master regulators, Wnt and TGF-β signaling pathways drive EMT as well as ECM restructuring.[53,54] Notably, while previous reports link UBR7 loss to Wnt signaling and subsequent EMT induction,[37] the present study highlights its role in suppressing the TGF-β pathway, thereby regulating the ECM gene expression. Importantly, the ECM genes that are regulated by UBR7 belong to different pathways such as ECM-receptor interaction, collagen biosynthesis, focal adhesion, and integrin signaling and are plausible targets to impede cancer cell migration.[6]

A rich tapestry of chromatin modifications is differentially regulated in TNBC by different epigenetic regulators. In addition to its established role in H2BK120 mono ubiquitination,[37] it has recently been shown that UBR7 is involved in the recognition of the post-nucleosomal histone marks, including H3K4me3 and H3K9me3.[38] Notably, in TNBC, we have observed a global reduction of the facultative heterochromatin mark H3K27me3 upon loss of UBR7. It has previously been established that EZH2 is involved in the deposition of the H3K27me3 mark onto *TGFBRII* and *THBS1* genes during the early stages of breast cancer.[45] Remarkably, we identify that UBR7 interacts with the PRC2 complex and plays an important role in the recruitment of EZH2 to the *TGFB1* gene promoter to stabilize the H3K27me3 mark. Thus, the absence of UBR7 alleviates this repression, leading to the invoking of TGF-β signaling, which translates into the induction of ECM genes, including *FN1*, collagen, and *LOX*. A similar PRC2-mediated regulatory mechanism leading to suppression of ECM gene expression has previously been reported. For instance, a histone variant mH2A1.2 attenuates bone metastasis of breast cancer in coordination with EZH2, which induces the H3K27me3 mark on the *LOX* gene promoter, thereby keeping it inactive.[55] Thus, the present study substantiates that epigenetic regulators such as UBR7 play a fundamental role in reprogramming the epigenetic landscape of chromatin. Interestingly, UBR7 gets recruited on the *TGFB1* gene promoter to suppress H3K27me3 occupancy and that can be translated to regulate
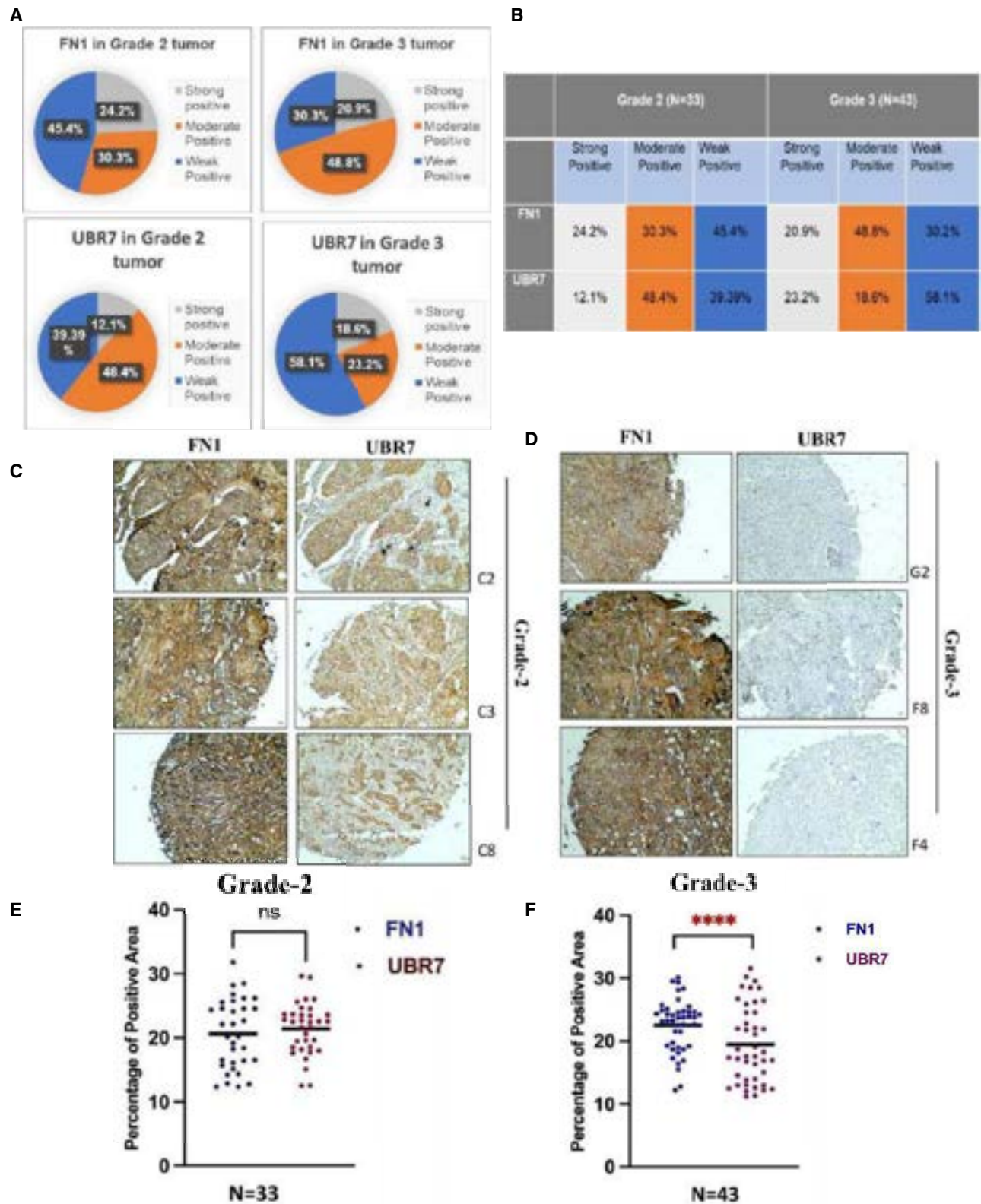

Figure 6. UBR7 affects metastasis in TNBC by modulating ECM

(A) Schematic representation showing the steps of cell seeding on solidified ECM matrix, maintenance, and final decellularization using lysis buffer (8 mM disodium hydrogen phosphate, 1% NP-40, pH 9.6) to prepare for AFM studies.

(B) DMT young modulus image of MDA-MB-231 expressing SCR (i) and UBR7-Sh1 (iii). Corresponding depth histogram from QNM analysis representing DMT young modulus in (ii) MDA-MB-231 SCR and (iv) UBR7-Sh1 cell-generated matrix. The experiment was performed using three biological replicates ($n = 3$). Three different fields from each replicate were taken for calculation.

(C) DMT young modulus calculation represented as mean ± standard deviation.

(D and E) Invasion of MDA-MB-231 cells expressing PLKO1 (i), LOX (ii), and LOX overexpression complemented with UBR7-WT (iii) in one set (D) and PLKO1 (i), UBR7-WT (ii), and UBR7-WT in combination with 1 μM DZNep (iii) in another set (E) was imaged in a Matrigel chamber. The percentage of invaded cells was quantified using ImageJ software for both the first set (D, iv) and the second set (E, iv). $n = 3$. Data information: in (D) and (E), error bars represent SEM. p values are calculated by using unpaired Student's t tests. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$.

A

FN1 in Grade 2 tumor
- Strong positive: 24.2%
- Moderate Positive: 30.3%
- Weak Positive: 45.4%

FN1 in Grade 3 tumor
- Strong positive: 20.9%
- Moderate Positive: 48.8%
- Weak Positive: 30.3%

UBR7 in Grade 2 tumor
- Strong positive: 12.1%
- Moderate Positive: 48.4%
- Weak Positive: 39.39%

UBR7 in Grade 3 tumor
- Strong positive: 18.6%
- Moderate Positive: 23.2%
- Weak Positive: 58.1%

B

	Grade 2 (N=33)			Grade 3 (N=43)		
	Strong Positive	Moderate Positive	Weak Positive	Strong Positive	Moderate Positive	Weak Positive
FN1	24.2%	30.3%	45.4%	20.9%	48.8%	30.2%
UBR7	12.1%	48.4%	39.39%	23.2%	18.6%	58.1%

C

FN1   UBR7
Grade-2
C2
C3
C8
Grade-2

D

FN1   UBR7
Grade-3
G2
F8
F4
Grade-3

E

ns
- FN1
- UBR7
Percentage of Positive Area
N=33

F

****
- FN1
- UBR7
Percentage of Positive Area
N=43

(legend on next page)

the expression of the downstream ECM genes. Notably, UBR7 silencing can also affect the H327me3 enrichment on ECM genes, although the molecular mechanism is yet to be elucidated. Such dramatic alterations in H3K27me3 occupancy in chromatin, mediated by UBR7, lead to TNBC suppression.

Notably, our investigation extends beyond transcriptional regulation, delving into the functional implications of UBR7 on ECM mechanics. It is well known that, apart from ECM deposition by the cancer cells, ECM modifications mediate an additional level of complexity that augments matrix stiffening and fibrosis.[6] Notably, the absence of UBR7 led to an alteration in the activity of LOX as well as the expression of ECM proteins like collagen and FN1 that could also be well scored in simple 3D spheroid cultures and *in vivo* syngeneic mice models. In the present study, we further establish that UBR7 is involved in glycosylated LOX production, as well as its activity, which has direct implications for collagen deposition and cross-linking, an observation substantiated by the inverse correlation between UBR7 expression and lox activity within breast tumor tissues. The alteration in matrix stiffness upon UBR7 knockdown in MDA-MB-231 cells could be connected to the increased invasive property and consequently a heightened metastatic potential of the cells in the absence of UBR7. Remarkably, the cooperative function of PRC2 and UBR7 reflects a multi-layered transcription regulation mechanism affecting the ECM stiffness, which in turn plays a crucial role in curbing the invasive property of TNBC. This association was not observed in normal proliferating MCF10A cells. This indicates that UBR7 has preferred interacting partners in fine-tuning the gene expression program differently in normal vs. cancer cells. Remarkably, the role of UBR7 as a metastasis suppressor through the regulation of EMT programs has already been established.[37] In the present study, we establish its intriguing role in ECM remodeling, which is another major contributor to the increased metastatic potential of TNBC.

Overall, the present study demonstrates that UBR7 plays a pivotal role in the ECM remodeling of TNBC. Employing multipronged strategies, it has been established that UBR7 emerges as a master regulator, orchestrating TGF-β signaling and a diverse array of ECM genes, ultimately affecting ECM mechanics and metastatic potential of TNBC. In this context, the functional crosstalk established between UBR7 and PRC2 complex is instrumental in the maintenance of the epigenetic landscape of chromatin, including that of the ECM genes. Therefore, stabilizing UBR7 possibly implicates in alteration of ECM scaffolding by suppressing the enzymatic activity of ECM remodeler LOX and hampers total collagen deposition. These findings highlight a potential role of UBR7 in reshaping the tumor-associated matrix in terms of its stiffness and rigidity to overcome the challenge of matrix hardening (desmoplasia), which directly affects cell survival, metastasis, and recurrence in TNBC.

Limitations of the study

In this study, we have elucidated the role of UBR7 in concert with EZH2 in TGF-β-mediated transcription regulation of ECM genes. Interestingly, we have observed that UBR7 is an important regulator of the epigenetic landscape of heterochromatin, thereby suppressing stiffness-mediated metastasis of TNBC. However, the ECM component is majorly contributed by the stromal cells. In this context, the crosstalk between the tumor and stromal cells of the microenvironment plays a pivotal role in ECM deposition and modifications. Hence, the future perspective of the study will be to understand these complex regulatory mechanisms that have direct consequences in regulating stiffness-mediated metastasis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS
  - Human subjects
  - Mice
  - Cell lines
- METHOD DETAILS
  - Cell culture, treatment, and preparation of stable cell line
  - RNA-sequencing
  - Gene overexpression
  - Quantitative real-time PCR
  - Immunoblotting
  - Flow cytometry
  - Mass spectrometry
  - Analysis of mass spectrometry data
  - Identification of differentially expressed proteins (DEPs) and their biological analysis
  - Biological pathway network analysis
  - Co-immunoprecipitation
  - Chromatin immunoprecipitation
  - ChIP-sequencing assay
  - ChIP-seq data processing
  - Quantitative and qualitative gene expression profiling
  - Estimation of replicate reproducibility, global occupancy profile analysis, and differentially enriched peaks analysis
  - Sphere formation
  - Immunofluorescence
  - LOX activity assay
  - Total collagen assay
  - Cell invasion assay
  - Cell embedding in ECM matrix
  - Atomic force microscopy
  - *In vivo* tumorigenicity assays
  - Immunohistochemistry (IHC)

Figure 7. Inverse correlation between UBR7 and ECM markers in grade-specific TNBC
(A) Pie chart showing quantification of UBR7 and fibronectin categorized into weak positive (percentage area positive <20%), moderately positive (percentage area positive (20%–25%), and strong positive (percentage area positive >30%) among clinical grade 2 (n = 33) and grade 3 (n = 43) TNBC patients.
(B) Tabular representation highlighting quantification of UBR7 and FN1 in grade 2 and grade 3 TNBC patients.
(C and D) Immunohistochemistry images showing the distribution of fibronectin and UBR7 in grade 2 TNBC patients and grade 3 TNBC patients simultaneously.
(E and F) Quantification scatterplot of grade 2 and grade 3 TNBC patients showing significant difference in UBR7 and FN1 in grade 3. Data information: in (E) and (F), using unpaired Student's t tests, p values are calculated. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

- ○ Picrosirius Red staining
- ○ Collection and processing of breast cancer patient tissue samples
- ○ IHC of tissue microarray slides
- ○ In silico analysis
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

AUTHOR CONTRIBUTIONS

C.D. conceived the study and designed the experiments. S.A., V.S., S.N., M.G., J.K., A.B., and A.K. designed and performed the experiments. S.A., S.N., M.G., J.K., A.B., A.M., D.S., K.S., and D.N. analyzed the results. N.S.R. and M.V. performed advanced bioinformatics and computational analysis. S.N. and H.R. contributed to the molecular subtyping of the patient cohort. S.A. and C.D. wrote the paper. All the authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Siegel, R.L., Miller, K.D., Wagle, N.S., and Jemal, A. (2023). Cancer statistics, 2023. CA. Cancer J. Clin. *73*, 17–48. https://doi.org/10.3322/CAAC.21763.

2. Polyak, K. (2011). Heterogeneity in breast cancer. J. Clin. Invest. *121*, 3786–3788. https://doi.org/10.1172/JCI60534.

3. Stewart, R.L., Updike, K.L., Factor, R.E., Henry, N.L., Boucher, K.M., Bernard, P.S., and Varley, K.E. (2019). A Multigene Assay Determines Risk of Recurrence in Patients with Triple-Negative Breast Cancer. Cancer Res. *79*, 3466–3478. https://doi.org/10.1158/0008-5472.CAN-18-3014.

4. Janmey, P.A., and Miller, R.T. (2011). Mechanisms of mechanical signaling in development and disease. J. Cell Sci. *124*, 9–18. https://doi.org/10.1242/JCS.071001.

5. Butcher, D.T., Alliston, T., and Weaver, V.M. (2009). A tense situation: forcing tumour progression. Nat. Rev. Cancer *9*, 108–122. https://doi.org/10.1038/NRC2544.

6. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K.J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat. Commun. *11*, 5120. https://doi.org/10.1038/S41467-020-18794-X.

7. Levental, K.R., Yu, H., Kass, L., Lakins, J.N., Egeblad, M., Erler, J.T., Fong, S.F.T., Csiszar, K., Giaccia, A., Weninger, W., et al. (2009). Matrix Cross-linking Forces Tumor Progression by Enhancing Integrin signaling. Cell *139*, 891–906. https://doi.org/10.1016/J.CELL.2009.10.027.

8. Choi, S., Friedrichs, J., Song, Y.H., Werner, C., Estroff, L.A., and Fischbach, C. (2019). Intrafibrillar, bone-mimetic collagen mineralization regulates breast cancer cell adhesion and migration. Biomaterials *198*, 95–106. https://doi.org/10.1016/J.BIOMATERIALS.2018.05.002.

9. Kauppila, S., Stenbäck, F., Risteli, J., Jukkola, A., and Risteli, L. (1998). Aberrant type I and type III collagen gene expression in human breast cancer in vivo. J. Pathol. *186*, 262–268. https://doi.org/10.1002/(SICI)1096-9896(1998110)186:3.

10. Xiong, G.-F., and Xu, R. (2016). Function of cancer cell-derived extracellular matrix in tumor progression. J. Cancer Metastasis Treat. *2*, 357. https://doi.org/10.20517/2394-4722.2016.08.

11. Zhang, H., Fredericks, T., Xiong, G., Qi, Y., Rychahou, P.G., Li, J.D., Pih-lajaniemi, T., Xu, W., and Xu, R. (2018). Membrane associated collagen XIII promotes cancer metastasis and enhances anoikis resistance. Breast Cancer Res. *20*, 116. https://doi.org/10.1186/S13058-018-1030-Y.

12. Saatci, O., Kaymak, A., Raza, U., Ersan, P.G., Akbulut, O., Banister, C.E., Sikirzhytski, V., Tokat, U.M., Aykut, G., Ansari, S.A., et al. (2020). Targeting lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer. Nat. Commun. *111*, 2416–2417. https://doi.org/10.1038/s41467-020-16199-4.

13. Dinca, S.C., Greiner, D., Weidenfeld, K., Bond, L., Barkan, D., and Jorcyk, C.L. (2021). Novel mechanism for OSM-promoted extracellular matrix remodeling in breast cancer: LOXL2 upregulation and subsequent ECM alignment. Breast Cancer Res. *23*, 56. https://doi.org/10.1186/S13058-021-01430-X/FIGURES/7.

14. Barker, H.E., Chang, J., Cox, T.R., Lang, G., Bird, D., Nicolau, M., Evans, H.R., Gartland, A., and Erler, J.T. (2011). LOXL2-mediated matrix remodeling in metastasis and mammary gland involution. Cancer Res. *71*, 1561–1572. https://doi.org/10.1158/0008-5472.CAN-10-2868.

15. Conklin, M.W., Eickhoff, J.C., Riching, K.M., Pehlke, C.A., Eliceiri, K.W., Provenzano, P.P., Friedl, A., and Keely, P.J. (2011). Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am. J. Pathol. *178*, 1221–1232. https://doi.org/10.1016/J.AJPATH.2010.11.076.

16. Ramaswamy, S., Ross, K.N., Lander, E.S., and Golub, T.R. (2003). A molecular signature of metastasis in primary solid tumors. Nat. Genet. *33*, 49–54. https://doi.org/10.1038/NG1060.

17. Leo, C., Cotic, C., Pomp, V., Fink, D., and Varga, Z. (2018). Overexpression of Lox in triple-negative breast cancer. Ann. Diagn. Pathol. *34*, 98–102. https://doi.org/10.1016/J.ANNDIAGPATH.2018.03.009.

18. Cox, T.R., Bird, D., Baker, A.M., Barker, H.E., Ho, M.W.Y., Lang, G., and Erler, J.T. (2013). LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis. Cancer Res. *73*, 1721–1732. https://doi.org/10.1158/0008-5472.CAN-12-2233.

19. Huang, J., Zhang, L., Wan, D., Zhou, L., Zheng, S., Lin, S., and Qiao, Y. (2021). Extracellular matrix and its therapeutic potential for cancer treatment. Signal Transduct. Target. Ther. *6*, 153. https://doi.org/10.1038/S41392-021-00544-0.

20. Erler, J.T., Bennewith, K.L., Cox, T.R., Lang, G., Bird, D., Koong, A., Le, Q.T., and Giaccia, A.J. (2009). Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche. Cancer Cell *15*, 35–44. https://doi.org/10.1016/J.CCR.2008.11.012.

21. Rachman-Tzemah, C., Zaffryar-Eilot, S., Grossman, M., Ribero, D., Timaner, M., Mäki, J.M., Myllyharju, J., Bertolini, F., Hershkovitz, D., Sagi, I., et al. (2017). Blocking Surgically Induced Lysyl Oxidase Activity Reduces the Risk of Lung Metastases. Cell Rep. *19*, 774–784. https://doi.org/10.1016/J.CELREP.2017.04.005.

22. Mohammadi, H., and Sahai, E. (2018). Mechanisms and impact of altered tumour mechanics. Nat. Cell Biol. *20*, 766–774. https://doi.org/10.1038/s41556-018-0131-2.

23. Meng, C., He, Y., Wei, Z., Lu, Y., Du, F., Ou, G., Wang, N., Luo, X.G., Ma, W., Zhang, T.C., and He, H. (2018). MRTF-A mediates the activation of COL1A1 expression stimulated by multiple signaling pathways in human breast cancer cells. Biomed. Pharmacother. *104*, 718–728. https://doi.org/10.1016/J.BIOPHA.2018.05.092.

24. Jia, H., Janjanam, J., Wu, S.C., Wang, R., Pano, G., Celestine, M., Martinot, O., Breeze-Jones, H., Clayton, G., Garcin, C., et al. (2019). The tumor cell-secreted matricellular protein WISP1 drives pro-metastatic collagen linearization. EMBO J. *38*, e101302. https://doi.org/10.15252/EMBJ.2018101302.

25. Wiercinska, E., Naber, H.P.H., Pardali, E., Van Der Pluijm, G., Van Dam, H., and Ten Dijke, P. (2011). The TGF-β/Smad pathway induces breast cancer cell invasion through the up-regulation of matrix metalloproteinase 2 and 9 in a spheroid invasion model system. Breast Cancer Res. Treat. *128*, 657–666. https://doi.org/10.1007/S10549-010-1147-X/METRICS.

26. Bhola, N.E., Balko, J.M., Dugger, T.C., Kuba, M.G., Sánchez, V., Sanders, M., Stanford, J., Cook, R.S., and Arteaga, C.L. (2013). TGF-β inhibition enhances chemotherapy action against triple-negative breast cancer. J. Clin. Invest. *123*, 1348–1358. https://doi.org/10.1172/JCI65416.

27. Jalalirad, M., Haddad, T.C., Salisbury, J.L., Radisky, D., Zhang, M., Schroeder, M., Tuma, A., Leof, E., Carter, J.M., Degnim, A.C., et al. (2021). Aurora-A kinase oncogenic signaling mediates TGF-β-induced triple-negative breast cancer plasticity and chemoresistance. Oncogene *40*, 2509–2523. https://doi.org/10.1038/s41388-021-01711-x.

28. Nandy, D., Rajam, S.M., and Dutta, D. (2020). A three layered histone epigenetics in breast cancer metastasis. Cell Biosci. *10*, 52. https://doi.org/10.1186/S13578-020-00415-1.

29. Wei, Y., Xia, W., Zhang, Z., Liu, J., Wang, H., Adsay, N.V., Albarracin, C., Yu, D., Abbruzzese, J.L., Mills, G.B., et al. (2008). Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. Mol. Carcinog. *47*, 701–706. https://doi.org/10.1002/MC.20413.

30. Huang, H., Hu, J., Maryam, A., Huang, Q., Zhang, Y., Ramakrishnan, S., Li, J., Ma, H., Ma, V.W.S., Cheuk, W., et al. (2021). Defining super-enhancer landscape in triple-negative breast cancer by multiomic profiling. Nat. Commun. *12*, 2242–2316. https://doi.org/10.1038/s41467-021-22445-0.

31. Knoechel, B., Roderick, J.E., Williamson, K.E., Zhu, J., Lohr, J.G., Cotton, M.J., Gillespie, S.M., Fernandez, D., Ku, M., Wang, H., et al. (2014). An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia. Nat. Genet. *46*, 364–370. https://doi.org/10.1038/NG.2913.

32. Marsolier, J., Prompsy, P., Durand, A., Lyne, A.M., Landragin, C., Trouchet, A., Bento, S.T., Eisele, A., Foulon, S., Baudre, L., et al. (2022). H3K27me3 conditions chemotolerance in triple-negative breast cancer. Nat. Genet. *54*, 459–468. https://doi.org/10.1038/S41588-022-01047-6.

33. Mukherjee, S., Adhikary, S., Gadad, S.S., Mondal, P., Sen, S., Choudhari, R., Singh, V., Adhikari, S., Mandal, P., Chaudhuri, S., et al. (2020). Suppression of poised oncogenes by ZMYND8 promotes chemo-sensitization. Cell Death Dis. *11*, 1073. https://doi.org/10.1038/S41419-020-03129-X.

34. Deng, L., Meng, T., Chen, L., Wei, W., and Wang, P. (2020). The role of ubiquitination in tumorigenesis and targeted drug discovery. Signal Transduct. Target. Ther. *5*, 11. https://doi.org/10.1038/S41392-020-0107-0.

35. Mattiroli, F., and Penengo, L. (2021). Histone Ubiquitination: An Integrative Signaling Platform in Genome Stability. Trends Genet. *37*, 566–581. https://doi.org/10.1016/J.TIG.2020.12.005.

36. Tasaki, T., Mulder, L.C.F., Iwamatsu, A., Lee, M.J., Davydov, I.V., Varshavsky, A., Muesing, M., and Kwon, Y.T. (2005). A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. Mol. Cell Biol. *25*, 7120–7136. https://doi.org/10.1128/MCB.25.16.7120-7136.2005.

37. Adhikary, S., Chakravarti, D., Terranova, C., Sengupta, I., Maitituoheti, M., Dasgupta, A., Srivastava, D.K., Ma, J., Raman, A.T., Tarco, E., et al. (2019). Atypical plant homeodomain of UBR7 functions as an H2BK120Ub ligase and breast tumor suppressor. Nat. Commun. *10*, 1398. https://doi.org/10.1038/S41467-019-08986-5.

38. Hogan, A.K., Sathyan, K.M., Willis, A.B., Khurana, S., Srivastava, S., Zasadzińska, E., Lee, A.S., Bailey, A.O., Gaynes, M.N., Huang, J., et al. (2021). UBR7 acts as a histone chaperone for post-nucleosomal histone H3. EMBO J. *40*, e108307. https://doi.org/10.15252/EMBJ.2021108307.

39. Zhao, L., Kang, M., Liu, X., Wang, Z., Wang, Y., Chen, H., Liu, W., Liu, S., Li, B., Li, C., et al. (2022). UBR7 inhibits HCC tumorigenesis by targeting Keap1/Nrf2/Bach1/HK2 and glycolysis. J. Exp. Clin. Cancer Res. *41*, 330–418. https://doi.org/10.1186/S13046-022-02528-6/FIGURES/8.

40. Srivastava, S., Sahu, U., Zhou, Y., Hogan, A.K., Sathyan, K.M., Bodner, J., Huang, J., Wong, K.A., Khalatyan, N., Savas, J.N., et al. (2021). NOTCH1-driven UBR7 stimulates nucleotide biosynthesis to promote T cell acute lymphoblastic leukemia. Sci. Adv. *7*, 9781–9808. https://doi.org/10.1126/SCIADV.ABC9781.

41. Kai, F., Drain, A.P., and Weaver, V.M. (2019). The Extracellular Matrix Modulates the Metastatic Journey. Dev. Cell *49*, 332–346. https://doi.org/10.1016/J.DEVCEL.2019.03.026.

42. Xu, X., Zhang, L., He, X., Zhang, P., Sun, C., Xu, X., Lu, Y., and Li, F. (2018). TGF-β plays a vital role in triple-negative breast cancer (TNBC) drug-resistance through regulating stemness, EMT and apoptosis. Biochem. Biophys. Res. Commun. *502*, 160–165. https://doi.org/10.1016/J.BBRC.2018.05.139.

43. Furler, R.L., Nixon, D.F., Brantner, C.A., Popratiloff, A., and Uittenbogaart, C.H. (2018). TGF-β Sustains Tumor Progression through Biochemical and Mechanical Signal Transduction. Cancers *10*, 199. https://doi.org/10.3390/CANCERS10060199.

44. Liu, S., Ren, J., and ten Dijke, P. (2021). Targeting TGFβ signal transduction for cancer therapy. Signal Transduct. Target. Ther. *6*, 8–20. https://doi.org/10.1038/s41392-020-00436-9.

45. Hinshelwood, R.A., Huschtscha, L.I., Melki, J., Stirzaker, C., Abdipranoto, A., Vissel, B., Ravasi, T., Wells, C.A., Hume, D.A., Reddel, R.R., and Clark, S.J. (2007). Concordant epigenetic silencing of transforming growth factor-beta signaling pathway genes occurs early in breast carcinogenesis. Cancer Res. *67*, 11517–11527. https://doi.org/10.1158/0008-5472.CAN-07-1284.

46. Kagan, H.M., and Li, W. (2003). Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell. J. Cell. Biochem. *88*, 660–672. https://doi.org/10.1002/JCB.10413.

47. Liu, S.Y., and Zheng, P.S. (2013). High aldehyde dehydrogenase activity identifies cancer stem cells in human cervical cancer. Oncotarget *4*, 2462–2475. https://doi.org/10.18632/ONCOTARGET.1578.

48. Sun, F.F., Hu, Y.H., Xiong, L.P., Tu, X.Y., Zhao, J.H., Chen, S.S., Song, J., and Ye, X.Q. (2015). Enhanced expression of stem cell markers and drug resistance in sphere-forming non-small cell lung cancer cells. Int. J. Clin. Exp. Pathol. *8*, 6287–6300.

49. Grimsby, J.L., Lucero, H.A., Trackman, P.C., Ravid, K., and Kagan, H.M. (2010). Role of lysyl oxidase propeptide in secretion and enzyme activity. J. Cell. Biochem. *111*, 1231–1243. https://doi.org/10.1002/JCB.22845.

50. Xiao, Q., and Ge, G. (2012). Lysyl oxidase, extracellular matrix remodeling and cancer metastasis. Cancer Microenviron. *5*, 261–273. https://doi.org/10.1007/S12307-012-0105-Z.

51. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov. *12*, 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059.

52. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674. https://doi.org/10.1016/J.CELL.2011.02.013.

53. Coban, B., Bergonzini, C., Zweemer, A.J.M., and Danen, E.H.J. (2021). Metastasis: crosstalk between tissue mechanics and tumour cell

plasticity. Br. J. Cancer *124*, 49–57. https://doi.org/10.1038/S41416-020-01150-7.

54. Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. Nat. Rev. Mol. Cell Biol. *15*, 786–801. https://doi.org/10.1038/NRM3904.

55. Kim, J., Shin, Y., Lee, S., Kim, M., Punj, V., Lu, J.F., Shin, H., Kim, K., Ulmer, T.S., Koh, J., et al. (2018). Regulation of Breast Cancer-Induced Osteoclastogenesis by MacroH2A1.2 Involving EZH2-Mediated H3K27me3. Cell Rep. *24*, 224–237. https://doi.org/10.1016/J.CELREP.2018.06.020.

56. Adhikary, S., Sanyal, S., Basu, M., Sengupta, I., Sen, S., Srivastava, D.K., Roy, S., and Das, C. (2016). Selective Recognition of H3.1K36 Dimethylation/H4K16 Acetylation Facilitates the Regulation of All-trans-retinoic Acid (ATRA)-responsive Genes by Putative Chromatin Reader ZMYND8. J. Biol. Chem. *291*, 2664–2681. https://doi.org/10.1074/JBC.M115.679985.

57. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57. https://doi.org/10.1038/NPROT.2008.211.

58. Yeung, N., Cline, M.S., Kuchinsky, A., Smoot, M.E., and Bader, G.D. (2008). Exploring biological networks with Cytoscape software. Curr. Protoc. Bioinforma *Chapter 8*, 8.13.1–8.13.20, *Chapter 8*. https://doi.org/10.1002/0471250953.BI0813S23.

59. Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D., and Amit, I. (2013). High-throughput chromatin immuno-precipitation for genome-wide mapping of in vivo protein-DNA interac-tions and epigenomic states. Nat. Protoc. *8*, 539–554. https://doi.org/10.1038/NPROT.2013.023.

60. Adhikary, S., Chakravarti, D., Terranova, C., Sengupta, I., Maitituoheti, M., Dasgupta, A., Srivastava, D.K., Ma, J., Raman, A.T., Tarco, E., et al. (2019). Atypical plant homeodomain of UBR7 functions as an H2BK120Ub ligase and breast tumor suppressor. Nat. Commun. *101*, 1398–1415. https://doi.org/10.1038/s41467-019-08986-5.

61. Sengupta, I., Das, D., Singh, S.P., Chakravarty, R., and Das, C. (2017). Host transcription factor Speckled 110 kDa (Sp110), a nuclear body pro-tein, is hijacked by hepatitis B virus protein X for viral persistence. J. Biol. Chem. *292*, 20379–20393. https://doi.org/10.1074/JBC.M117.796839.

62. Maja, M., Mohammed, D., Dumitru, A.C., Verstraeten, S., Lingurski, M., Mingeot-Leclercq, M.P., Alsteens, D., and Tyteca, D. (2022). Surface cholesterol-enriched domains specifically promote invasion of breast can-cer cell lines by controlling invadopodia and extracellular matrix degrada-tion. Cell. Mol. Life Sci. *79*, 417. https://doi.org/10.1007/S00018-022-04426-8.

63. Harris, G.M., Raitman, I., and Schwarzbauer, J.E. (2018). Cell-derived de-cellularized extracellular matrices. Methods Cell Biol. *143*, 97–114. https://doi.org/10.1016/BS.MCB.2017.08.007.

64. Viji Babu, P.K., Rianna, C., Mirastschijski, U., and Radmacher, M. (2019). Nano-mechanical mapping of interdependent cell and ECM mechanics by AFM force spectroscopy. Sci. Rep. *9*, 12317. https://doi.org/10.1038/S41598-019-48566-7.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Fibronectin monoclonal antibody	Invitrogen	MA5-17075;Cat#MA5-17075; RRID:AB_2538546
Anti-Fibronectin antibody	Abcam	ab2413;Cat#ab2413;RRID:AB_2262874
Anti-Collagen I	Abcam	ab34710;Cat#ab34710;RRID:AB_731684
Recombinant Anti-LOX antibody	Abcam	ab174316;Cat#ab174316; RRID:AB_2630343
Phospho-SMAD2 (Ser 465/467)mAb	CST	3108;Cat#3108;RRID:AB_490941
Recombinant Anti-Smad2 antibody	Abcam	ab33875;Cat#ab33875;RRID:AB_777976
Phospho-SMAD3 (Ser 423/425)mAb	CST	9520;Cat#9520;RRID:AB_2193207
Anti-Smad3 antibody	Abcam	ab227223;Cat#ab227223
Anti-Histone H3 (tri methyl K27) antibody	Abcam	ab6002;Cat#ab6002;RRID:AB_305237
Anti-Histone H3 (tri methyl K27) antibody	Abcam	Cat#ab192985;RRID:AB_2650559
Anti-Histone H3 (tri methyl K4) antibody	Abcam	ab8580;Cat#ab8580;RRID:AB_306649
Anti-Histone H3 antibody	Abcam	ab1791;Cat#ab1791;RRID:AB_302613
Anti-Histone H3 (tri methyl K36) antibody	Abcam	ab9050;Cat#ab9050;RRID:AB_306966
Anti-Histone H3 (tri methyl K9) antibody	Abcam	ab8898;Cat#ab8898;RRID:AB_306848
Anti-Histone H3 (Acetyl K27) antibody	Abcam	ab4729;Cat#ab4729;RRID:AB_2118291
Anti-Histone H3 (Acetyl K27) antibody	Abcam	Cat#ab177178;RRID:AB_2828007
Recombinant Anti-KMT6/EZH2 antibody	Abcam	ab191250;Cat#ab191250
SUZ12 (D39F6) mAb	CST	3737;Cat#3737;RRID:AB_2196850
Recombinant Anti-EED antibody	Abcam	ab240650;Cat#ab240650
Anti-ALDH1A1 antibody	Abcam	ab227964;Cat#ab227964
Recombinant Anti-MRP1 antibody	Abcam	ab233383;Cat#ab233383
Anti-UBR7 Antibody	Abcam	ab228911;Cat#ab228911
UBR7 polyclonal antibody	Bethyl laboratories	A304-131A;Cat#A304-131A
Anti-beta Actin antibody	Abcam	ab8226;Cat#ab8226;RRID:AB_306371
Anti-actin antibody	Abcam	Cat#3280;RRID:AB_303668
Goat Anti-Rabbit IgG HRP	Abcam	ab97051;Cat#ab97051; RRID:AB_10679369
Goat Anti-Mouse HRP secondary	Advansta	05071–500;Cat#05071-500
Anti-Rabbit Alexa Fluor 594	Invitrogen	A11037;Cat#A11037;RRID:AB_2534095
Bacterial and virus strains		
E.coli. DH5α, STBL3 competent cells	Lab generated	N/A
Biological samples		
Mouse breast tumor tissue and Human tumor tissue (control cohort and triple negative breast cancer patients)	For mouse tissue: IISER Pune, Animal house; for patient tissue: Department of Surgery, R G Kar Medical College and Hospital, Kolkata India ECR/322/inst/WB/2013/RR-20	N/A
Chemicals, peptides, and recombinant proteins		
Protease inhibitor cocktail complete EDTA free	Roche	04693132001;Cat#04693132001
PhoSTOP	Roche	49068450;Cat#49068450
ECM Gel	MilliporeSigma	E6909;Cat#E6909
Dyna Beads	Invitrogen	10002D;Cat#10002D

Continued

REAGENT or RESOURCE		SOURCE	IDENTIFIER
TGF beta Ligand		MedChemExpress	HY-P7118;Cat#HY-P7118
TGFB RI kinase inhibitor VI		Sigma Aldrich	SB431542;Cat#SB431542
Sodium Deoxycholate		Sigma Aldrich	S1827;Cat#S1827
Nonidet P-40		Abcam	142227;Cat#142227
Tris		Sigma Aldrich	154563;Cat#154563
Glycine		BioRad	1610718;Cat#1610718
NaCl		Sigma Aldrich	S7653;Cat#S7653
HEPES		Sigma Aldrich	H3375;Cat#H3375
EDTA		Sigma Aldrich	E9884;Cat#E9884
DTT		Sigma Aldrich	D0632;Cat#D0632
Tween 20		Sigma Aldrich	P9416;Cat#P9416
SDS		Sigma Aldrich	L3771;Cat#L3771
BSA		Sigma Aldrich	A7906;Cat#A7906
$MgCl_2$		Sigma Aldrich	M1028;Cat#M1028
DMEM-F12		Gibco	D6421;Cat#D6421
bFGF		Gibco	13256–029;Cat#13256-029
EGF		Sigma Aldrich	SRP3027;Cat#SRP3027
Insulin		Sigma Aldrich	I2643;Cat#I2643
Fat free BSA		Sigma Aldrich	A7030;Cat#A7030
B27		Gibco	17504–044;Cat#17504-044
Critical commercial assays			
Picrosirius Red Stain Kit		Polysciences	24901;Cat#24901
cDNA synthesis kit		Thermo Fischer	K16315;Cat#K16315
Lox activity assay kit		Abcam	ab112139;Cat#ab112139
Total collagen assay kit		Abcam	ab222942;Cat#ab222942
Deposited data			
RNA-sequencing		This paper	GSE262488
ChIP-sequencing		This paper	GSE262487
Experimental models: Cell lines			
MDA-MB-231		ATCC	HTB-26;Cat#HTB-26
MDA-MB-468		ATCC	HTB-132;Cat#HTB-132
NIH3T3		ATCC	CRL-1658;Cat#CRL-1658
HEK293T		ATCC	CRL-3216;Cat#CRL-3216
Experimental models: Organisms/strains			
BALB/c female mice syngeneic		Animal House facility, IISER, Pune, India	N/A
Oligonucleotides			
UBR7 Fw	CCAGAACAGGGAAAGGATGA	This Paper	N/A
UBR7 Rev	TCCTGAAGTTTGCAGCCAG	This Paper	N/A
Beta Actin Fw	AGGCACCAGGGCGTGAT	This Paper	N/A
Beta actin Rev	GCCCACATAGGAATCCTTCTGAC	This Paper	N/A
LOX Fw	GGATACGGCACTGGCTACTT	This Paper	N/A
LOX Rev	GACGCCTGGATGTAGTAGGG	This Paper	N/A
FN1 Fw	CTGGCCGAAAATACATTGTAAA	This Paper	N/A
FN1 Rev	CCACAGTCGGGTCAGGAG	This Paper	N/A
VTN Fw	AGACAAGTACTACCGAGTCA	This Paper	N/A
VTN Rev	CTAAGGGACCTTTATTGGGC	This Paper	N/A
COL1A1 Fw	GTGGCAGTGATGGAAGTGTG	This Paper	N/A
COL1A1 Rev	AGGACCAGCGTTACCAACAG	This Paper	N/A

Continued

REAGENT or RESOURCE		SOURCE	IDENTIFIER
COL1A2 Fw	GAGACCCTTCTTACTCCTGAA	This Paper	N/A
COL1A2 Rev	GTGATGTTCTGAGAGGCATAG	This Paper	N/A
ITGB8 Fw	TGTGTGCTGGGCATGGAGAGTGT	This Paper	N/A
ITGB8 Rev	CAGTGCTGGGCTGCTGCTGAA	This Paper	N/A
ITGA5 Fw	GTCGGGGGCTTCAACTTAGAC	This Paper	N/A
ITGA5 Rev	CCTGGCTGGCTGGTATTAGC	This Paper	N/A
TGFB1 Fw	TCGCCAGAGTGGTTATCTT	This Paper	N/A
TGFB1 Rev	TAGTGAACCCGTTGATGTCC	This Paper	N/A
THBS1 Fw	TCAGTACAGAAATAACGAGGAATGGAC	This Paper	N/A
THBS1 Rev	CATTCTCCAATCAGGAACTGTGG	This Paper	N/A
TGFBR2 Fw	TGTTCCTGTAGCTCTGATGAGTGC	This Paper	N/A
TGFBR2 Rev	TCTGCTGCCGGTTAACGC	This Paper	N/A
TGFB1 ChIP Fw	AATAACCTAGATGGGCGCG	This Paper	N/A
TGFB1 ChIP Rev	CGCATCCTAGACCCTTTCTC	This Paper	N/A
THBS1 ChIP Fw	ATCAGGACACCCCAAGATTTC	This Paper	N/A
THBS1 ChIP Rev	AAGGTGAGCAAGATCAGGAAC	This Paper	N/A
TGFBR2 ChIP Fw	ACATGATTGGCAGCTACGAG	This Paper	N/A
TGFBR2 ChIP Rev	TCACTCAACTTCAACTCAGCG	This Paper	N/A
LOX ChIP Fw	CAAAGTTACACAAGCCGTTCTG	This Paper	N/A
LOX ChIP Rev	GTTCCGTTTTGTTCTTTCCCC	This Paper	N/A
FN1 ChIP Fw	ATATCCACGACCTTGCAGC	This Paper	N/A
FN1 ChIP Rev	TCCCGAGTCAGTACCCTTTAG	This Paper	N/A
ITGB2 ChIP Fw	GTGATTCTCTGGAGAGGAATGG	This Paper	N/A
ITGB2 ChIP Rev	AGCGTCAGTGTTTAGGTTCTC	This Paper	N/A
COL20A1 ChIP Fw	TGAGTCAAGCCAGGTTTCTG	This Paper	N/A
COL20A1 ChIP Rev	CCAAGGGCAGTCAGGTAAG	This Paper	N/A
ITGAX ChIP Fw	CTCTTACAACTTCCCACCCTG	This Paper	N/A
ITGAX ChIP Rev	GCTGAGCATTGAAAGCAACC	This Paper	N/A
Recombinant DNA			
LOX-ORF		Dr. Ozgur Sahin, University of South Carolina, USA	N/A
UBR7-Sh1		This Paper	N/A
UBR7-Sh2		This Paper	N/A
UBR7-WT		This Paper	N/A
PLKO1		This Paper	N/A
Software and algorithms			
Graphpad prism V.8.0		https://www.graphstats.net/graphpad-prism	N/A
ImageJ		https://imagej.nih.gov/ij/index.html	N/A
FlowJo		https://www.flowjo.com/	N/A
Cytoscape		https://cytoscape.org/	N/A
Enrichr		https://maayanlab.cloud/Enrichr/	N/A

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Prof. Chandrima Das (Chandrima.das@saha.ac.in).

Material availability

This study did not generate new unique reagents.

Data and code availability

- RNA-seq and ChIP-seq data have been deposited at the GEO database with accession IDs GSE262488 and GSE262487, respectively. Proteomics data is available on request.
- This paper doesn't report original code.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS

Human subjects

Human breast cancer patient tissues were collected with the consent of the patients from the multidisciplinary research unit (MRU), R G Kar Medical College and Hospital, Kolkata, India for this prospective study. Molecular subtyping has been done for the patients and characterized as triple-negative breast cancer. Clinical details of the patients were recorded including stage, grade, TNM. The research work along with the analysis was performed at Saha Institute of Nuclear Physics, Kolkata, India as this study was approved by the Institutional Ethics Committee (ECR/322/inst/WB/2013/RR-20).

Mice

In vivo, BALB/c syngeneic female mice experiments were performed in the Animal house facility of IISER Pune, India in collaboration with Prof. Kundan Sengupta. Six to eight weeks aged mice group was taken for the injections. The weight of the nice was noted. 5 million 4T1 cells were injected into a single flank of each mouse. On the 21st day, the tumors were evicted, and proceeded for further experiments.

Cell lines

HEK 293T, MDA-MB-468, NIH3T3cell lines (ATCC) were grown in DMEM (Sigma) and 10% fetal calf serum (Sigma) supplemented with 1% penicillin/streptomycin. The MDA-MB-231 cell line was maintained in RPMI (Gibco) and 12% fetal bovine serum (Sigma) supplemented with 1% penicillin/streptomycin.

METHOD DETAILS

Cell culture, treatment, and preparation of stable cell line

HEK293T cells were grown in DMEM (Gibco) with 10% fetal bovine serum (FBS; Gibco), 1% antibiotic-antimycotic (Gibco) at 37°C and 5% (v/v) CO_2 incubator. MCF10A cells were cultured in DMEM/Ham's F12 medium which is supplemented with 5% horse serum (Gibco), insulin, epidermal growth factor (EGF), hydrocortisone, cholera toxin (Sigma), and 1% antibiotic-antimycotic. Additionally, MDA-MB-231 and MDA-MB-468 cell lines were maintained in RPMI-1460 (Gibco), supplemented with 12% FBS (Gibco), and 1% antibiotic-antimycotic (Gibco). All cell lines were purchased from ATCC, USA.

UBR7 sh-RNA constructs with the sequences UBR7-sh1 5′-CAGTGCACCCAGGGTTATTTG-3′ and UBR7-sh2 5′-GCTTAAAGC TAAGCAGCTTAT-3′ were cloned in PLKO.1-puro vector backbone (Sigma-Aldrich) which targets 3′-UTR of the gene. Scrambled RNA (SCR) was stably transfected to make control cells. Recombinant lentivirus was produced following the previous protocol.[37] Concisely, HEK293T cells were seeded at a density of 3 × 10^5 cells in 60 mm dishes which was followed by transfection with UBR7 sh-RNA, packaging vector (pPAX2), and envelope vector (pMD2.G) together using lipofectamine 2000 as per the protocol of the manufacturer. After transfection, the collection of viral supernatants was done at 48 and 72 h. After that MDA-MB-231 cells were infected with the addition of viral supernatant containing 10 μg/ml polybrene, three times in 48 h time intervals. The transduced cells were selected by the addition of Puromycin (4 μg/mL) (Sigma) in media and maintained for consecutive three days.

DZNep treatment was performed in MDA-MB-231 cells with a 1 μM dose for 24 h.

RNA-sequencing

RNA was extracted using RNeasy kit. The size of RNA fragments was selected by SPRI bead-based size selection and library preparation was prepared using NEBNext Ultra II directional RNA library preparation kit according to the protocol of the manufacturer. Generation of 100 bp length paired end reads was done on the Illumina platform. The data has been submitted and has an accession id GSE262488.

The raw sequencing reads were undergone quality-based filtering and adapter trimming by using fastp (v0.20.1) and quality reads were aligned with the Homo sapiens GRch38 genome using the splice aware aligner hisat (v2.2.1). Next, by using stringtie assembler (v2.1.5), the alignments were assembled into transcripts. Then the computation of the read counts for the genes and normalisation of expression values with the Fragment per kilobase per million (FPKM) metric was performed by stringtie.

Using DEseq2 R package, differential expression was analyzed where read counts were taken as input, and genes with a *p*-value less than 0.05 were considered to be significantly changed in expression between the control and UBR7 knockdown in MDA-MB-231

cells. A log2 fold change higher than 1 and a log2 fold change lower than −1 were considered as upregulated and downregulated genes respectively. Gene ontology and pathway enrichment analysis were performed to identify the differentially regulated pathways by using DAVID functional annotation tool (v6.8). Gene ontology plots were generated using WEGO. Additionally, GSEA (Gene Set Enrichment Analysis) of DEG's was performed using WEBGSEALT tool and key enriched gene sets were visualized using Bubble Plot representation. The Gene Ratio represents the NES (Normalized Enrichment Score) obtained from GSEA analysis.[37] The enrichment results are provided with an enrichment ratio in ORA or to the normalized enrichment score (NES) in GSEA. The false discovery rate (FDR) for the categories is ≤ 0.05 to include all significantly enriched sets. The RNA-sequencing analysis has been detailed in Table S1.

Gene overexpression
UBR7 wild-type vector (UBR7-WT) was cloned into the pHAGE-CMV-fullEF1a-IRES-ZsGreen and LOX-ORF constructs were overexpressed in MDA-MB-231 cells using lipofectamine 2000 invitrogen. LOX-ORF construct was a gift from Dr. Ozgur Sahin, University of South Carolina, USA.

Quantitative real-time PCR
Isolation of total RNA from cells was performed by using TRIzol (Invitrogen) which was followed by reverse transcription of 2 μg of RNA by cDNA synthesis kit (Thermo Scientific) as per the manufacturer's protocol. cDNA was diluted 1:1 and qRT-PCR was performed using ABI-SYBR GREEN mix (Applied Biosystems) and the run was performed in StepONE plus FAST Real-time PCR machine. Relative expression of mRNA was determined by $2^{-\Delta Ct}$ method for real value. The analysis of each sample was done independently three times and the results from one representative experiment were shown with its technical triplicates. All the primer details are listed in the Key resources table.

Immunoblotting
Whole-cell lysate preparation was done by RIPA lysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, and 1 mM EDTA) as elaborated previously.[56] The lysates were run on 11% or 15% SDS-polyacrylamide gel electrophoresis. Then the blots were probed with specific primary antibodies which was followed by blocking of the membrane with 5% bovine serum albumin (BSA) or non-fat dry milk made in TBST (Tris-buffered saline and 0.1% Tween 20). All the antibodies are listed in the Key resources table.

Flow cytometry
TNBC extracellular matrix protein fibronectin 1 was assessed by flow cytometric analysis from MDA-MB-231 and UBR7 stable knockdown cells using Fibronectin monoclonal antibody (Thermofisher Scientific), followed by incubation with FITC-conjugated secondary antibody. Flow cytometric studies were carried out on BD FACS Aria III with appropriate filters. Each sample was analyzed for 10000 events. Data were analyzed using the FlowJo software. The changes in mean intensity were analyzed as dot plots of SSC against the specific dye (FITC) using quadrant gates with appropriate population distribution.

Mass spectrometry
Sample preparation for mass spectrometry was followed by the previous protocol (Mondal et al.). Briefly, MCF10A and MDA-MB-231 cells were taken in ice-cold PBS. The cell pellet was then resuspended in isotonic buffer (ice-cold) containing 10mM Tris-HCl pH 7.5, 3mM $CaCl_2$, 2mM $MgCl_2$, and 0.3M sucrose) and incubated for 15 min on ice for swelling of the cells. By the addition of 0.65% NP-40 (v/v), the nuclei were released from cells, followed by vortexing for 10 s and immediate centrifugation. After removing the supernatant, the nuclei pellet was dissolved in extraction buffer (20mM HEPES pH-7.9, 0.42M NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 1% NP-40, 25% Glycerol, 1mM DTT and 1X protease inhibitor cocktail PIC) for 30 min with time to time mixing to lyse the nuclei. The nuclei extract was then collected after high-speed centrifugation at 4°C. Then the supernatant was diluted 1:1 with dilution buffer containing 20mM HEPES pH 7.9, 1.5mM $MgCl_2$, 0.2mM EDTA, 1% NP-40, 1mM DTT, and 1X PIC. The diluted supernatant was pre-cleared by incubating with IgG and Dynabeads for 30 min at 4°C consecutively. Then the pre-cleared nuclear extract was incubated with IgG and respective antibodies overnight at 4°C on a nutator. Following 2 h of bead binding, the dynabeads were washed five times thoroughly with wash buffer composed of 20mM HEPES pH-7.9, 200mM NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 1% NP-40, 1mM DTT, and 1X PIC. After that, the protein complexes were eluted from the beads using elution buffer composed of 28% Glycerol, 10% SDS, 120mM Tris pH 6.8, and 50mM DTT and proceeded for LC-MS/MS which was followed by analysis.

Analysis of mass spectrometry data
The raw data of mass spectrometry were processed with the help of the MaxQuant software20 (version 1.5.5.1) and searched with a query of UniProt Homo Sapiens in Andromeda search engine53. Setting the precursor mass tolerances 20 ppm and 4.5 ppm, first and main searches were done, respectively. Next, the MS/MS tolerance was set to 20 ppm. The minimum length of the peptide was fixed to six amino acids. As the specificity for trypsin cleavage was required, it was allowed up to two missed cleavage sites. The peptide, protein along with the site false discovery rate (FDR) was fixed to 1%. The setting of modifications by light as well as heavy iodoacetamide on cysteine residues (carbamidomethylation) was performed by selecting label type modification in Andromeda. The

abundance of Protein was then estimated by MaxQuant. From this, the median value of the ratios between light and heavy dimethyl modifications was calculated which were measured from each protein for all unique peptides. The reproducibility of the replicates was estimated by performing PCA (Principal Component Analysis) and Correlation Condition Tree based on Pearson Correlation and Average linkage rule.

Identification of differentially expressed proteins (DEPs) and their biological analysis
Proteins that are differentially expressed in tumor samples as compared to normal with a minimum of 2-fold and above with FDR score of ≤ 0.05 were considered. Using Database for Annotation Visualization and Integrated Discovery (DAVID), gene Ontology (GO) along with pathway analysis was performed. A significant enrichment test was then performed along with the Bonferroni correction method to obtain the most significantly enriched biological processes from the set of differentially expressed genes in DAVID. Additionally, Fisher's exact test was done in DAVID-Protein Ontology and pathways. Here the calculation of p-value significance was performed based on the ratio of proteins obtained to the expected number of proteins (O/E) while the total number of proteins was considered for the respective pathway in Homo sapiens. The hierarchical clustering method was done by employing Cluster 3.0. Protein expression data or the FPKM for all samples was taken and was transformed to log2 while the proteins with low expression (FPKM <0.05) and invariant were removed. After that the proteins were centered. The clustering was done on the basis of the differential expression pattern of proteins and their fold change. Ultimately, the heatmap was visualized in Java TreeView (Version 3.0). Student's t-test was performed along with a fold change cutoff of 2 or above and -2 or below satisfying an FDR score of ≤ 0.05 and a corrected p-value of ≤ 0.01 to qualify differentially expressed genes.

Biological pathway network analysis
Significant differentially expressed transcripts were subjected to GO and Pathway enrichment analysis by employing DAVID tool.[57] The GO and pathways having an FDR score of ≤ 0.05, were filtered for further downstream analysis. To perform protein regulatory network modeling, the key biological dysregulated GO and pathways with the differentially expressed proteins was considered as an input to the Pathreg algorithm. The key nodes and edges which represent variations of the protein regulation in different cell lines, were identified with the help of the nodes and edges obtained from Pathreg algorithm provided as an input to Cytoscape v2.8.2.[58]

Identification of differentially expressed proteins was performed by applying a Fold Change of ≥ 1.5 for both up and down-regulated proteins and a student's t-test p Value of ≤ 0.05. Protein Ontology and Pathway Enrichment of differentially expressed proteins were performed by applying Fisher's exact test which was based on the ratio of obtained number of proteins to the expected number of proteins (O/E) with a consideration of the total number of proteins for the respective pathway in Homo sapiens with an FDR of <0.05. Fisher's exact test was performed based on the ratio of obtained number of proteins to the expected number of proteins (O/E) with a consideration of the total number of proteins for the respective pathway in Homo sapiens with an FDR of <0.05.

Co-immunoprecipitation
Cells were crosslinked by using 1% formaldehyde, followed by neutralization using 125 mM glycine (dissolved in 1X PBS). The cell lysate was prepared by sonication in RIPA buffer (50 mM Tris pH8.0, 0.1% SDS, 0.5% Sodium Deoxycholate, 150 mM NaCl, 1mM EDTA, and 1% NP-40) supplemented with 1X Phosphatase and protease blocker. Then the lysate was centrifugated at 13000 rpm at 4°C for 10min and the supernatant was collected and precleared by using IgG antibody. The IP was performed with respective antibodies to pull the protein by using Dyna Beads. The protein-bound beads were washed thrice before eluting the protein from the beads to analyze by western blotting.

Chromatin immunoprecipitation
ChIP assays were performed as mentioned elsewhere.[59] Briefly, cells were crosslinked with 1% formaldehyde, followed by neutralisation with 125 mM glycine. The cells were lysed by cell lysis buffer (5mM PIPES pH 8.0, 85mM KCl, 0.5% NP-40, 1mM DTT, and 1X PIC) and centrifuged to obtain the nuclear pellet. Then the nuclear lysis buffer (Tris pH 8.0, 10mM EDTA, 1% SDS, 1mM DTT, and 1X PIC) was added and the chromatin was sheared by sonication cycles and immunoprecipitated with the respective antibodies or with IgG as negative control. The next day, following the bead binding, beads were washed once with low salt, high salt, and lithium chloride wash buffer and twice with TE buffer. The protein complex was eluted from the beads at room temperature using elution buffer (10mM NaHCO$_3$, 1% SDS) and de-crosslinked with 5M NaCl at 65°C overnight. Following RNase A and proteinase K digestion, ChIP DNA was extracted by the phenol-chloroform method. ChIP DNA was analyzed using primers for specific genes by quantitative real-time PCR (qPCR). Each ChIP experiment was performed three times independently with three technical triplicates. Primers used in ChIP qPCR are listed in the Key resources table.

ChIP-sequencing assay
ChIP assay was performed with a little modification as described previously.[60] The samples were subjected to library preparation at the NGS facility NCBS Bangalore. NEBNext Ultra II DNA Library Prep with Sample Purification Beads (Catalog no-E7103L) kit protocol was used for library preparation of ChIP samples for Illumina sequencing. All ChIP-Seq libraries with fragment length

(∼250–500 bp) were sequenced on the NovaSeq 6000 platform using an SP flow cell with a 2x100 sequencing read length. The data has been submitted and has an accession id GSE262487. The ChIP-sequencing analysis details have been provided in Table S2.

ChIP-seq data processing

The Input (Whole Cell Extract) and the Ip (ChIP DNA) were QCed for integrity using Agilent Bioanalyzer 2100 and samples with clean bands only were used for further investigation. Libraries for ChIP-seq were prepared according to the TruSeq ChIP Library Preparation Kit (Illumina) system. The quantity and quality of the final library were analyzed by Life Technologies Qubit3.0 Fluorometer and Agilent Bioanalyzer 2100 respectively. 150 bp Paired-end sequencing was performed on Illumina HiSeq X (Illumina Inc., San Diego, CA, USA). The sequencing was planned to yield >20 million reads per library with a read length of 150 bp in the Paired End module.

Quantitative and qualitative gene expression profiling

The raw data obtained in FASTq format was subjected to stringent quality control using the FastQC tool kit. Scanning, identification, and removal of the adapter were done using the Trim Galore tool. The most recent Homo sapiens genome build (GrCh38) was obtained from iGenomes, and indexing was carried out using the BWA build with the default parameters. The raw Fastq files were checked for PCR duplication and were deleted using 'rmdup' option of the Samtools 1.3.1. Then the GrCh38 genome assembly was used to align each raw file using BWA with the default settings for paired-end sequencing, as previously reported. The following alignment, duplicate sequences were identified using the Picard program. Preseq was used to provide alignment-level quality control and library complexity assessment. Furthermore, normalised bigWig files with a scale of 1 million mapped reads were produced. Deeptools was used to perform genome-wide IP enrichment with reference to Input in addition to gene body meta profile. The phantompeakqual tools were used to calculate the strand cross-correlation peak and ChIP-seq quality measurements such as NSC and RSC. Peak calling in IP was conducted using the MACS2 process, and detected peaks were annotated using the GFF3 file for GrCh38 genome assembly. As part of the computing pipeline, several plots such as profile plots, correlation matrix, PCA, peak score calculation, peak visualisation files, and other input files required by tools such as IGV were obtained.

Estimation of replicate reproducibility, global occupancy profile analysis, and differentially enriched peaks analysis

Since the IP libraries were sequenced in replicates, estimation of reproducibility was done at various levels including PCA, Correlation Tree, Plot profiles, and Profile Heatmaps way of representation. Further, the consensus peaks in H3K27ac and H3K27Me3 were compared in the presence and absence (Silenced) of UBR7 by plotting the peaks in a Venn diagram to understand the transition of epigenetic mark enrichment on various genes. The results of such analysis led to the identification of a statistically significant list of genes that could be indicative of a positive or negative change in gene regulation under the influence of UBR7.

The DAVID tool and major Gene Ontologies, pathways, and associated biological components were used to conduct biological analyses of differentially enriched genes. To obtain the best result of substantially enriched biological processes, a significant enrichment test was performed in DAVID using the list of differentially expressed genes and the Bonferroni correction procedure. Fisher's exact test was used in DAVID-Gene Ontology and pathways, and p-value significance was computed based on the ratio of the obtained number of genes to the expected number of genes (O/E) taking into account the total number of genes in Homo sapiens. Cluster 3.0 was used to do hierarchical clustering. Peak data was collected and converted to log2. Low-expressed genes (FPKM ≤ 0.05) and invariant genes were eliminated. The genes were then centered, and clustering was done based on their differential histone modification enrichment pattern and fold change. The heatmap was finally visualised in Java TreeView 3.0. For further downstream analysis, only GO and pathways with an FDR value of ≤ 0.05 were evaluated. Pathreg algorithm for gene regulatory network modeling was fed with key physiologically dysregulated GO and pathways, as well as differentially recruited genes. The Pathreg algorithm result (nodes and edges) was sent into Cytoscape v2.8.2[58] to identify important nodes and edges that might be reflective of the gene regulatory alterations caused by UBR7 knockdown.

A Peak score of ≥ 1.5 and a student's t-test p-value of ≤ 0.05 were used to identify differentially recruited genes. Gene Ontology and Pathway Enrichment of differentially recruited genes were performed with an FDR of 0.05 using Fisher's exact test based on the ratio obtained to the expected number of genes (O/E) considering the total number of genes for the respective pathway in Homo Sapiens.

Sphere formation

MDA-MB-231 and UBR7 stable knockdown cells were seeded in the 6 well plates with ultra-low attachment property (Sigma Corning) at a density of 2.4×10^4 in serum-free DMEM-F12 media which is supplemented with bFGF (Gibco), EGF (Gibco), insulin (Sigma), BSA, and B27 (BD Biosciences). Following 7 days of maturation, tumorspheres, and the media supernatant were collected and processed for LOX activity assay. Some spheres were processed for immunofluorescence.

Immunofluorescence

Immunofluorescence in cells was performed following the previous protocol.[61] Briefly, the cells were fixed using 4% paraformaldehyde for 10 min, before being permeabilized with 1% Triton X-100 in PBS and blocked with a blocking solution containing 3% BSA in PBS at room temperature. After that, the cells were incubated for an hour with the primary antibody which was followed by washing

with PBST (PBS+0.05% Tween 20). Then Alexa conjugated secondary antibodies were added and incubated for 1 h in the dark. The coverslips were mounted using DAPI containing mounting media. The images were taken in a Zeiss 710 Anisotropy microscope.

The tumorspheres were fixed using chilled methanol for 15 min and then blocking was done using 3% BSA for 1 h at ambient temperature. After that, the remaining spheres were incubated with primary antibodies overnight at a 4-degree temperature. The next day, secondary antibodies were added and mounting was done. Finally, the images were captured in a Nikon Ti-E confocal microscope.

LOX activity assay

Lysyl oxidase activity from the cell, patient tissue lysate, and cell culture medium was measured by LOX activity assay kit (fluorometric) (ab112139) following the manufacturer's protocol.[12] Briefly, 5×10^5 cells were harvested and washed with cold PBS. Then cells were homogenised quickly with RIPA buffer without DTT or detergents, followed by centrifugation at 13,000 g in cold for 10 min to remove any insoluble material. The supernatant was collected and kept on ice. For patient tissue, a 10mg tissue sample was harvested and washed with ice-cold PBS. Then extraction buffer (6M Urea, 10 mM Tris pH 7.4, and protease inhibitors) was added and homogenised by tissue homogeniser multiple times for 90 s keeping it in ice. The sample was centrifuged at 13,000 g for 10 min in cold and the supernatant was kept in ice. In the case of cell media, it was centrifuged similarly and kept in ice. Protein concentrations were determined for cell and tissue lysates. Multiple dilutions of samples in PBS +0.1% BSA were prepared. The reaction mix containing the sample treated as unknown, HRP substrate stock solution, HRP stock solution, and assay buffer were prepared and a background control was made without the sample, followed by incubation at 37°C for 40 min in dark. Finally, the fluorescence was measured on a microplate reader at Ex/Em 540/590 nm.

Total collagen assay

Total collagen was measured from patient tissue by total collagen assay kit (Percholate free) (ab222942) as per the manufacturer's protocol.[12] 100mg tissue sample was harvested for each assay. Distilled water was added to the tissue and homogenised properly using a tissue homogeniser. Then 10 N NaOH was added to the tissue homogenate and kept at 120°C for 1 h in a pressure-tight polypropylene vial. Following alkaline hydrolysis, the vial was placed in ice and 10 N HCl was slowly added to neutralize. The samples were then centrifuged at 13,000 g for 5 min and supernatant was collected. Blank and samples were added to the wells of a microplate reader and allowed to evaporate at 65°C on a heat block. A crystalline residue formed was dissolved by an oxidation reagent mix, followed by the addition of developer and DMAB concentrate solution to the reaction well consecutively. The plate was then sealed and incubated at 65°C for 45 min. Finally, the absorbance was measured at OD 560nm.

Cell invasion assay

ECM gel (Engelbreth-Holm-Swarm murine sarcoma) was diluted in serum-free RPMI-1460 to 0.5 mg/mL and added 140 μL to the 8.0 μm cell culture inserts (Thermofisher Scientific) as previously mentioned.[62] The gel was allowed to polymerize and then 0.8×10^5 cells were seeded in the upper chamber of the cassette with serum-free media whereas there was the complete media with 10% FBS filled the lower chamber. The cells were then allowed to migrate for 24 h and after 24 h, the invaded cells were fixed using 4% paraformaldehyde, followed by staining with crystal violet stain for 10 min and then washed with PBS to remove the excess stain. After that, the non-invaded cells were wiped from the ECM gel of the upper chamber. Image acquisition of the migrated cells was done and images were captured from three independent fields for each sample. Thereafter the images were analyzed by ImageJ software (National institutes of health; NIH).

Cell embedding in ECM matrix

The cells were seeded on ECM gel-coated coverslips and incubated for 10 days to allow the deposition of ECM secreted by cells. After 10 days, matrix decellularization was performed using extraction protocol followed by ref.[63,64] and then proceeded to the AFM study.

Atomic force microscopy

AFM imaging was conducted using the Biocatalyst – Bruker AFM instrument to image the matrices without cells. A suitable AFM probe was used for PeakForce Tapping in QNM mode i.e., Scanasyst Fluid Cantilever (Bruker, spring constant 0.7 N/m). The deflection sensitivity of the AFM cantilever was calibrated against a clean silicon surface and the spring constant was calibrated using the thermal tune method.[64] For each sample, the elastic modulus and topology images were captured from different fields over the sizes of 5×5 μm² with a 256 pixels $\times$ 256 pixels digital resolution. The tip radius of the cantilever was quantified using the tip quantification function available in the nanoscope analysis software.

Approaching the AFM probe to the ECM surface of the system automatically adjusts the probe-sample distance to maintain a constant, low peak force throughout the imaging process. The scan rate was 0.5 Hz and the peak force setpoint was below 1 nN. As the probe starts scanning the sample, the AFM system records the topographic data along with other corresponding mechanical properties, such as DMT modulus and adhesion. An optical microscope attached to the AFM can monitor the tip in our area of interest. Acquisition of images was done by the Nanoscope working software (version 9.1) and image processing along with analysis was performed using Nanoscope Analysis software (version 1.5).

In vivo tumorigenicity assays

In the central animal house facility of IISER Pune, six to eight-week-old female BALB/c mice were maintained in advanced ventilated cages with food and water. 5 million 4T1 cells transfected separately with control, UBR7-Sh and UBR7-WT, resuspended in 100 μL PBS, were injected into the single flank side of the mice subcutaneously. After the tumor growth had been observed, on the 21st day, the mice were sacrificed and the tumors were excised. Tumors were weighed and then measured using slide calipers. The volumes of the tumors were estimated by the formula, $V = (Width^2 \times length)/2$. Post eviction the tumors ($n = 3$) were processed for the functional assays- LOX activity and total collagen content measurement, immunohistochemistry (IHC), and picrosirius staining.

Immunohistochemistry (IHC)

The tissue was evicted and put into a 10% neutral buffered formalin solution. Paraffin-embedded tissue blocks were prepared by an automated tissue processor. Tissue blocks were trimmed in the trim mode of semiautomated microtome to expose the whole tissue. Tissue sections were cut into 3-μm thick sections and placed onto a floating water bath. After that, the sections were mounted on slides coated with poly-L-lysine, and sequentially slides were deparaffinized by placing the slides in a microwave oven at 50°C for 15 min. Tissue sections were rehydrated by placing the slides in a xylene bath for 3 min and then excess liquid was wiped off. Slides were placed serially in 90%, 80%, and then 70% ethanol for 3 min in each bath while removing the excess liquid in every step. Slides were then placed in a staining rack submerged in tris-EDTA (pH-8.0) and antigen retrieval was done by heating the slides at 95°C for 15 min. Then the slides were left to cool at room temperature for 1 h. The rest of the staining was followed according to the manufacturer's protocol (Mouse and Rabbit Specific HRP/DAB IHC Detection Kit - Micro-polymer ab236466). Briefly, sufficient drops of Hydrogen Peroxide Blocking solution were added to cover the sections and incubated at room temperature for 10 min, and washed twice in phosphate buffer. After that, a protein Block was applied and incubated at room temperature for 10 min to block nonspecific background staining before being washed in phosphate buffer. After that, the primary antibody was given in appropriate dilution and incubated according to the manufacturer's protocol. Slides were washed vigorously in phosphate buffer thrice. The secondary antibody was applied according to kit protocol and incubated for 15 min. Slides were washed 4 times in phosphate buffer. DAB Chromogen was added to DAB Substrate, mixed by swirling, and applied to the tissue then slides were Incubated for 1–10 min (empirically determined until the brownish color was observed and then rinsed four times in buffer. Slides were stained with Hematoxylin and counterstained with eosin stain. Slides were then gradually dehydrated by two changes in 70%, 80% then 90% ethanol in a stepwise manner for 3 min in each bath while removing the excess liquid in every step. Slides were put in a Xylene bath for 3 min and excess liquid was washed off. Slides were mounted with coverslips with DPX mounting media and visualized under a Light microscope for observation.

Picrosirius Red staining

Tumor tissues derived from mice were cut into pieces and then fixed in a 10% neutral buffered formalin solution. Tumor tissue blocks were prepared by embedding in parafilm. The protocol followed here was as recommended by Picrosirius Red Stain Kit (Polysciences, catalog no. 24901).[12] Briefly, the slides containing tissue sections were deparaffinised by incubating at 65°C for 20 min which was followed by xylene wash twice. The slides were then rehydrated using decreasing alcohol percentages. After that, the slides were stained with Picrosirius Red as per the manufacturer's protocol. Dehydration of the slides was then followed. After mounting, the samples were analyzed under a brightfield microscope. The red staining intensity was measured using ImageJ software.

Collection and processing of breast cancer patient tissue samples

Human breast cancer patient samples were obtained with the consent of all the patients from the Department of Surgery, R G Kar Medical College and Hospital, Kolkata, and research along with further analysis were done in Saha Institute of Nuclear Physics in accordance with the Institutional Ethics Committee (ECR/322/inst/WB/2013/RR-20). From primary site tumor tissues, $n = 4$ samples, LOX activity assay, and Total collagen assay were performed along with gene expression study for UBR7 by qRT-PCR. For the western blot, the tissues were homogenised properly after resuspending in RIPA buffer with 1X PIC, using tissue homogeniser and then centrifuged at 13,000 rpm for 10 min. For RNA isolation, the tissues were homogenised after the addition of Trizol and kept in ice. For IHC studies, $n = 5$ patient samples were analyzed where expression of UBR7 and ECM markers was quantified using ImageJ software.

IHC of tissue microarray slides

TNBC tissue microarray slides including clinical details were procured from tissuearray.com. UBR7 and FN1 staining were performed from the slides with the same patient array (code no. BRE1121) and collagen staining was done from another TNBC array slide (code no. BR1301a) following the above-mentioned IHC protocol. From the former slide, $n = 33$ patient tissues from grade 2 and $n = 43$ patient tissues from grade 3 were quantified. From the later slide, $n = 41$ for grade 2 and $n = 64$ for grade 3 patients were quantified.

In silico analysis

In silico studies was performed from r2 genomics analysis and visualisation platform. From the GSE202203 dataset of 201 TNBC patients, the log2 transformed values of UBR7 and different ECM markers were obtained (R2: Genomics Analysis and Visualization Platform (http://r2.amc.nl)).

QUANTIFICATION AND STATISTICAL ANALYSIS

An unpaired students' t-test was performed to calculate the significance statistically. Graphpad prism V.8.0 was used. All data have been represented as mean ± SE with error bars and *p* value ≤ 0.05 was granted as statistically significant. The data points were displayed for each graph. Individual experiments were carried out at least three times in triplicate.

All the patient tumor IHC images (including tissue microarray) were quantified using ImageJ open-access software. The mean threshold intensities are measured, and the percentage of the positive area (area fraction) and statistical analysis were done using GraphPad Prism software. Unpaired students' t-test was used for statistical analysis between the UBR7 and FN1 percentage positive area and a scatterplot was done by plotting the mean with SEM.

All the IHC images from mouse tumor tissue and Picrosirius Red staining images from the same were quantified by ImageJ open-access software. Statistical analyses were performed in a similar way as the patients. For LOX activity and total collagen content assay from patient tissue, Pearson correlation was used for statistical analysis. For the total collagen content, a significant inverse correlation with UBR7 expression has been found where R value is −0.98 and the *p*-value is 0.010. Similarly, the anti-correlation between UBR7 expression and LOX activity from patient tissues was confirmed by the R value −0.92 and the *p*-value is 0.039. Notably, unpaired students' t-test has been performed for in silico analysis of UBR7, COL1A1, and COL1A2 expression from the TNBC patient database from r2 genomics analysis and visualization platform.

ARTICLE

Atypical plant homeodomain of UBR7 functions as an H2BK120Ub ligase and breast tumor suppressor

Santanu Adhikary[1,2], Deepavali Chakravarti[3,4], Christopher Terranova[3], Isha Sengupta[1], Mayinuer Maitituoheti[3], Anirban Dasgupta[2], Dushyant Kumar Srivastava[2], Junsheng Ma[5], Ayush T. Raman[3], Emily Tarco[6], Aysegul A. Sahin[7], Roland Bassett[5], Fei Yang[6], Coya Tapia[7], Siddhartha Roy[2], Kunal Rai[3] & Chandrima Das[1]

The roles of Plant Homeodomain (PHD) fingers in catalysis of histone modifications are unknown. We demonstrated that the PHD finger of Ubiquitin Protein Ligase E3 Component N-Recognin7 (UBR7) harbors E3 ubiquitin ligase activity toward monoubiquitination of histone H2B at lysine120 (H2BK120Ub). Purified PHD finger or full-length UBR7 monoubiquitinated H2BK120 in vitro, and loss of UBR7 drastically reduced H2BK120Ub genome-wide binding sites in MCF10A cells. Low UBR7 expression was correlated with occurrence of triple-negative breast cancer and metastatic tumors. Consistently, UBR7 knockdown enhanced the invasiveness, induced epithelial-to-mesenchymal transition and promoted metastasis. Conversely, ectopic expression of UBR7 restored these cellular phenotypes and reduced tumor growth. Mechanistically, UBR7 loss reduced H2BK120Ub levels on cell adhesion genes, including CDH4, and upregulated the Wnt/β-Catenin signaling pathway. CDH4 overexpression could partially revert UBR7-dependent cellular phenotypes. Collectively, our results established UBR7 as a histone H2B monoubiquitin ligase that suppresses tumorigenesis and metastasis of triple-negative breast cancer.

[1] Biophysics and Structural Genomics Division, Saha Institute of Nuclear Physics, 1/AF Bidhan Nagar, Kolkata 700064, India. [2] Structural Biology and Bio-Informatics Division, CSIR-Indian Institute of Chemical Biology, 4 Raja S.C. Mullick Road, Kolkata 700032, India. [3] Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [4] Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [5] Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [6] Department of Translational Molecular Pathology and Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [7] Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. These authors contributed equally: Santanu Adhikary, Deepavali Chakravarti, Christopher Terranova. Correspondence and requests for materials should be addressed to S.R. (email: roysiddhartha@iicb.res.in) or to K.R. (email: krai@mdanderson.org) or to C.D. (email: chandrima.das@saha.ac.in)

Breast cancer is the most common cause of cancer mortality in female individuals. The heterogeneity of the disease poses immense challenges in deciphering therapeutic strategies[1]. The hormone receptor-negative or triple-negative subtype has the worst prognoses due to the lack of targeted therapies[2,3]. Although accumulation of genetic defects has been involved in the development of oncogenesis, epigenetic abnormalities play a significant role in the initiation, progression, and metastasis of the disease[4]. Specifically, epithelial-to-mesenchymal transition (EMT), which preludes the onset of metastasis[5,6], is thought to be driven by epigenetic alterations[7,8]. Primarily, histone modifications, which include methylation, acetylation, and ubiquitination, play crucial roles in maintaining homeostasis, failure of which may lead to disease initiation or progression[9]. Importantly, breast cancer cases with worse prognoses have lower levels of H3K18Ac, H4K12Ac, H3K4Me2, H4K20Me3, and H4R3Me2 marks[9,10].

Monoubiquitination at lysines on histones H2A and H2B have an antagonistic relationship in oncogenesis[11–14]. In contrast with polyubiquitination, which marks the protein for its proteasome-mediated degradation, monoubiquitination of histone H2B plays key roles in transcription memory and elongation, DNA damage response, viral infection, stem cell differentiation, and oncogenesis[15,16]. The E3 ligases for H2B monoubiquitination, RNF20 and RNF40, are reported to act as potent tumor suppressors, regulate DNA double-stranded break repair, and modulate stem cell differentiation[14,17,18].

The ubiquitin protein ligase E3 component N-recognin (UBR) family of mammalian E3 ligases containing seven members (UBR1–UBR7) is characterized by a 70-residue zinc finger-type UBR-box domain, which is essential for recognition of the N-degrons[19–21]. Despite harboring a UBR-box, UBR3, UBR6, and UBR7 do not bind to N-degrons. Although the members of the UBR family of proteins are generally heterogeneous in size and sequence, they harbor specific signatures unique to ubiquitin ligases or a substrate-recognition subunit of the E3 complex like the RING/HECT (really interesting new gene/homologous to the E6AP carboxyl terminus) domain or F-box[21,22]. A RING domain is present in UBR1, UBR2, and UBR3, a HECT domain is present in UBR5, and an F-box is present in UBR6. Of note, UBR7 has evolved with a plant homeodomain (PHD) finger, which is a putative chromatin-binding module, not present in any other UBR family proteins. Although the PHD finger is well characterized as a reader of methylated, acetylated, or unmodified histones[23], its role in enzymatic catalysis is not known. Furthermore, little is known about the role of UBR7 in carcinogenesis. In the present study, we demonstrated that the UBR7-PHD finger is an H2BK120 monoubiquitin ligase and a tumor suppressor in triple-negative breast cancer cases.

Results

UBR7-PHD monoubiquitinates histone H2B lysine 120 in vitro. UBR7, a protein with an unknown function, contains a UBR-box domain, which is essential for the recognition of N-degrons[20–22], and a PHD finger (Fig. 1a), which is highly conserved across species (Supplementary Fig. 1a). Although the sequence alignment of the UBR7-PHD finger (which is stabilized by zinc ion coordination in a cross-braced topology) with other well-characterized H3K4Me3 or H3K4Me0 binders exhibited several conserved residues, it displayed weak interaction with trimethylated histone peptides (Supplementary Fig. 1b–f). Although full-length UBR7 protein interacted with all recombinant histones in vitro, the PHD finger preferentially interacted with recombinant histone H2B (Fig. 1b) and could also immunoprecipitate them from MCF10A cells (Supplementary Fig. 1g). Sequence alignment of the UBR7-PHD with other classical RING

finger-E3 ubiquitin ligases indicated that zinc-coordinating His163 and His166 are unique in contrast to the other RING fingers (Supplementary Fig. 1b). Site-directed mutagenesis of H163S/H166S of the UBR7-PHD did not significantly compromise its association with histone H2B at various levels of organization (Fig. 1c). We observed similar results during immunoprecipitation assays from HEK293T cells (Fig. 1d). However, mutation of lysine 120 to arginine (K120R) in histone H2B abrogated its binding preference for UBR7 as observed through immunoprecipitation assays (Fig. 1e). Based on the zinc coordination fold similarity between the RING and the PHD finger, we hypothesized that UBR-PHD function as an E3 ubiquitin ligase for histone H2B substrate. Purified recombinant UBR7 full-length wild-type protein (UBR7-WT), full-length H163S/H166S catalytic-mutant (UBR7-CM), or individual domains (UBR or PHD) were incubated in the presence of an E1 ubiquitin-activating enzyme, an E2 ubiquitin-conjugating enzyme (UbcH6), ATP, inorganic pyrophosphatase, and biotin-tagged ubiquitin. UBR7-WT and the PHD finger alone could monoubiquitinate purified H2B, H2A/H2B dimer, core histone octamers, and purified nucleosomes (Fig. 1f, g and Supplementary Fig. 1h–k), in contrast with the other E3 ubiquitin ligases, which usually act in complex[24–26]. On the other hand, UBR7-CM failed to promote H2B ubiquitination (Fig. 1f, g and Supplementary Fig. 1j–o). Thus, our results demonstrated the E3 ubiquitin ligase function of UBR7 toward monoubiquitination of histone H2B in vitro.

UBR7 regulates H2BK120Ub levels ex vivo. Next, we sought to determine whether UBR7 regulates H2BK120Ub in mammalian cells. We observed lower UBR7 levels in human and murine breast cancer cells than in their "normal" counterparts. For example, UBR7 messenger RNA (mRNA) transcript and protein levels were lower in MCF7, T47D, MDA-MB-231, and MDA-MB-468 cells than in MCF10A and MCF12A cells (Fig. 2a and Supplementary Fig. 2a). We observed similar patterns in 21PT and 21MT2 compared to 16N (human) and 4T07 compared to 4T1 (murine) cells (Supplementary Fig. 2b-e). Genetic depletion of UBR7 by two different short hairpin RNAs (shRNAs) in MCF10A and MCF12A cells led to a dramatic reduction in global H2BK120Ub levels (Fig. 2b and Supplementary Fig. 2f, g). Importantly, the reduction caused by UBR7-shRNA was rescued by UBR7-WT, but not by UBR7-CM (Fig. 2c and Supplementary Fig. 2h). Consistently, UBR7-WT, but not UBR7-CM, increased H2BK120Ub levels in the MDA-MB-231 and MDA-MB-468 breast cancer cells (Fig. 2d, e and Supplementary Fig. 2h).

To examine changes in H2BK120Ub genome wide, we performed chromatin immunoprecipitation sequencing (ChIP-seq) of H2BK120Ub in control and UBR7-knockdown MCF10A cells. UBR7-knockdown cells had a drastically lower number of H2BK120Ub binding sites (1079) compared to control cells (8401) (Fig. 2f and Supplementary Data 1), which was verified by ChIP-quantitative polymerase chain reaction (qPCR) (Fig. 2g). Similarly, the intensity of H2BK120Ub enrichment was drastically reduced as demonstrated by average density plots and verified by ChIP-qPCR for selected genes (Fig. 2h, i and Supplementary Fig. 2i, j).

To analyze the impact of UBR7-mediated H2BK120Ub on the chromatin landscape, we also performed ChIP-seq for histone modifications H3K79Me2 (transcription), H3K4Me3 (promoters), H3K4Me1 (enhancers), H3K27Ac (active enhancers), H3K27Me3 (polycomb-repressed), and H3K9Me3 (heterochromatin)[27]. Consistent with prior reports, we noted a loss of H3K79Me2 on H2BK120Ub gene targets upon UBR7 knockdown (Fig. 2j). Identification of chromatin state transitions between control and UBR7-depleted cells using the ChromHMM
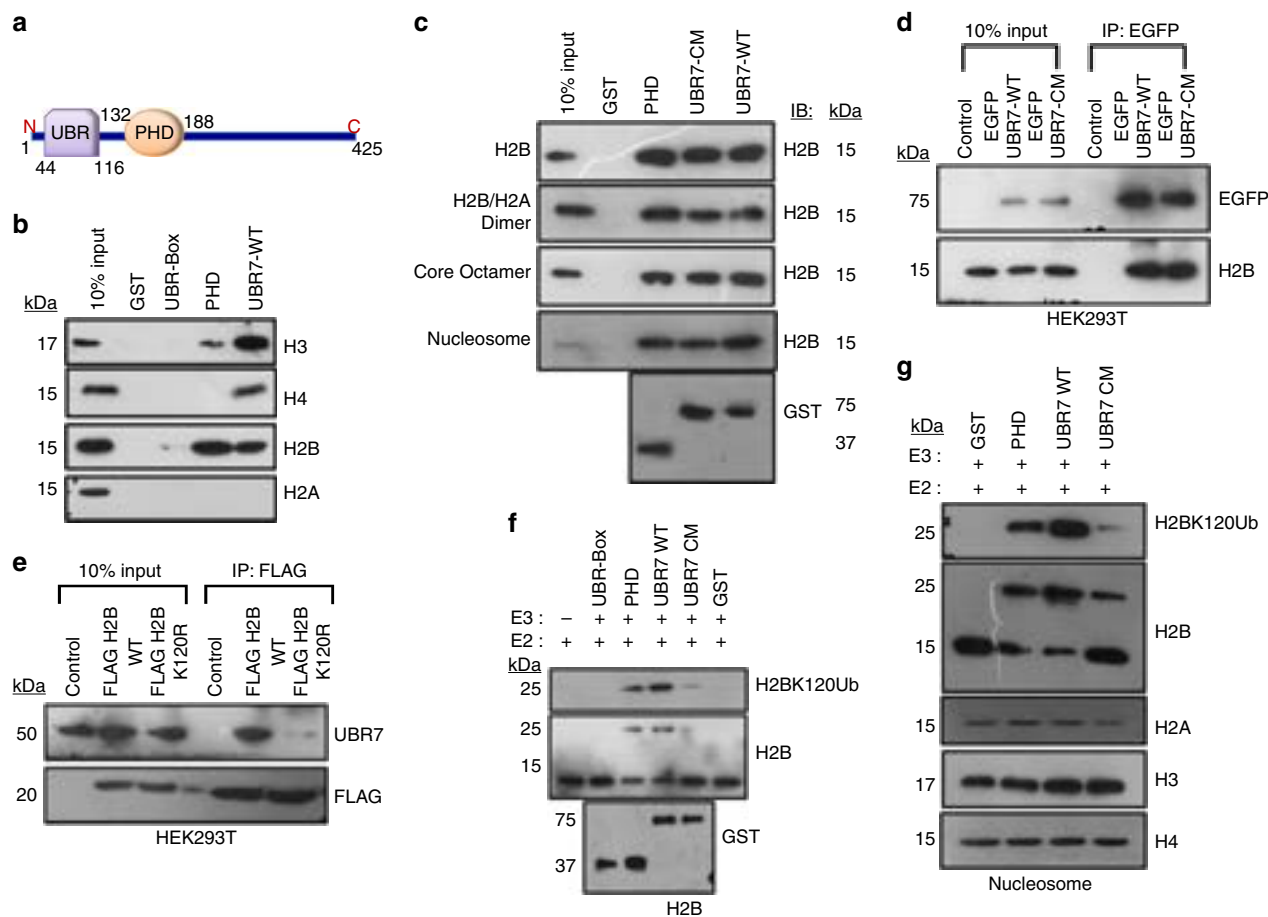
Fig. 1 UBR7 is a histone H2B ubiquitin ligase. **a** Schematic of the domain organization of UBR7. **b** Immunoblots showing the in vitro interaction of purified UBR7 with recombinant histones H3, H4, H2B, and H2A. **c** Interaction of UBR7 full-length wild-type (WT) or catalytic-mutant (CM) with recombinant H2B, H2A/H2B dimer, core octamer, or purified nucleosomes from HeLa cells. **d** Ex vivo interaction of H2B with UBR7 WT or CM in HEK293T cells. **e** Ex vivo interaction of UBR7 with H2B WT or H2B mutant (K120R) in HEK293T cells. **f, g** In vitro ubiquitination assay with recombinant H2B (**f**), or purified nucleosomes from HeLa cells (**g**)

algorithm in a 10-state model (Fig. 2k; see Methods) indicated that H2BK120Ub was primarily present in conjunction with H3K79Me2. Also, the most potent transitions were from highly transcribed states in control cells to low/non-transcribed states in UBR7-depleted cells (States 1 to 5, States 2 to 4 or 1, States 3 to 5, and States 4 to 6) (Fig. 2k, l). Overall, these experiments identified UBR7 as an E3 ubiquitin ligase in vivo and demonstrated the importance of UBR7 in maintaining specific chromatin patterns in cells.

Next, we defined the relative impact of UBR7 and other two known H2BK120Ub E3 ligases, RNF20 and RNF40, on its levels and genomic distribution. Knockdown of all the three E3s, UBR7, RNF20, and RNF40, separately in MCF10A cells led to significant reduction of total H2BK120Ub levels (Supplementary Fig. 3a). Direct comparison between ChIP-seq profiles for H2BK120Ub showed drastic losses at 10,919 sites in UBR7-deficient, 11,005 sites in RNF20-deficient, and 11,069 sites in RNF40-deficient cells (Supplementary Fig. 3b–g). Sites that lose H2BK120Ub upon individual knockdown of each of these enzymes significantly overlapped (Supplementary Fig. 3h–m), suggesting cooperative interaction between UBR7, RNF20, and RNF40 and requirement of all three proteins for maintenance of H2BK120Ub levels in the cell.

Low UBR7 is associated with triple-negative breast cancer. Next, to further explore UBR7's biological function, we examined its expression in The Cancer Genome Atlas Group (TCGA) breast cancer mRNA expression data[28]. This analysis revealed reduced UBR7 expression in triple-negative and basal-like breast tumors (Fig. 3a), which was confirmed in an independent cohort of breast cancer tissue microarrays (TMAs) containing 371 breast tumors using a UBR7-specific antibody (Supplementary Fig. 4a and Supplementary Data 2). We found that lower levels of UBR7 correlated significantly with triple-negative status (Fig. 3b) and individual estrogen receptor (ER) status and progesterone receptor (PR) status (Fig. 3c, d and Supplementary Data 2). Also, metastatic intraductal carcinomas harbored considerably lower levels of UBR7 than did primary tumors (Fig. 3e, f). Furthermore, UBR7 was associated with a better metastasis-free survival rate in the aggressive basal-type breast cancer (mesenchymal subtype tumors without endocrine therapy) (Supplementary Fig. 4b).

UBR7 acts as tumor metastasis suppressor gene. A series of loss-of-function and gain-of-function experiments established the role of UBR7 as a tumor and metastasis suppressor gene. Reduction of UBR7 in MCF10A and MCF12A cells increased two-dimensional (2D) proliferation drastically (Fig. 3g and Supplementary Fig. 4c), as well as expression of the proliferation markers Ki-67, PCNA, and MCM2[29] (Fig. 3h–j and Supplementary Fig. 4d). Overexpression of UBR7-WT and not UBR7-CM reversed the increased proliferation observed in UBR7-knockdown MCF10A cells (UBR7-sh1) (Fig. 3k) and the
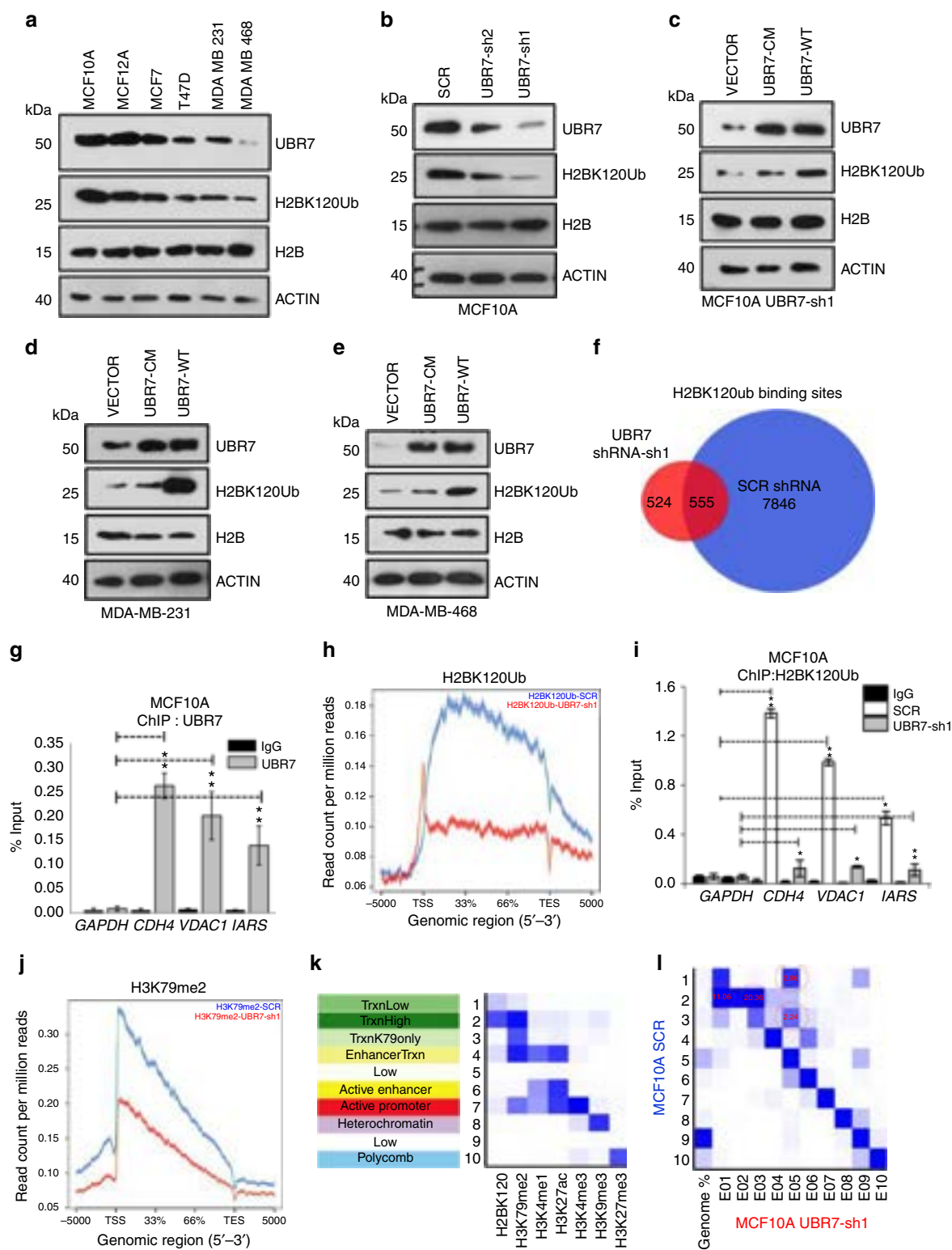
expression of associated proliferation markers (Fig. 3l and Supplementary Fig. 4e, f). Similarly, overexpression of UBR7-WT, but not UBR7-CM, in MDA-MB-231 and MDA-MB-468, two basal-like breast cancer cells that express UBR7 at low levels, dramatically reduced 2D proliferation (Supplementary Fig. 4g, h). Next, we examined three-dimensional (3D) soft agar colony growth of UBR7-depleted MCF10A cells in vitro and found

substantially higher numbers of colonies as well as bigger colonies in UBR7-depleted cells than in control (Fig. 4a, b). Importantly, the impact of wild-type derivatives in the 3D growth assay in MDA-MB-231 cells was abrogated by UBR7-CM (Fig. 4c, d). Overexpression of UBR7-WT and not UBR7-CM abrogated mammary fat pad tumor formation in vivo (Fig. 4e, f). Additionally, immunohistochemical analysis of the proliferation

Fig. 2 UBR7 is downregulated in invasive breast cancer cells. **a** Immunoblots of MCF10A, MCF7, T47D, MDA-MB-231, and MDA-MB-468 cells to monitor expression of UBR7, H2BK120Ub, and H2B. ACTIN was used as a loading control. **b** Immunoblots for UBR7, H2BK120Ub, H2B, and ACTIN (loading control) in MCF10A cells expressing scrambled (SCR), UBR7-sh1, or UBR7-sh2 shRNA. **c–e** Immunoblots for UBR7, H2BK120Ub, H2B, and ACTIN (loading control) in MCF10A UBR7-sh1 (**c**), MDA-MB-231 (**d**), and MDA-MB-468 (**e**) cells expressing wild-type (UBR7-WT) and catalytic-mutant (UBR7-CM) UBR7. ACTIN was used as a loading control. **f** Venn diagram showing overlap of total H2BK120Ub binding sites in Control (SCR) and UBR7-sh1-expressing MCF10A cells. **g** Bar plot for quantitative PCR (qPCR) enrichment of UBR7 chromatin immunoprecipitation (ChIP) in MCF10A cells for selected genes. *GAPDH* was used as a negative control. **h** Average genebody density plot for H2BK120Ub binding sites in Control (SCR) and UBR7-sh1-expressing MCF10A cells. **i** Bar plot for qPCR enrichment of H2BK120Ub ChIP in MCF10A cells expressing SCR or UBR7-sh1. *GAPDH* was used as a negative control. **j** Average genebody density plot for H3K79me2 binding sites in Control (SCR) and UBR7-sh1-expressing MCF10A cells. **k** Emission parameter for a 10-state chromatin state model called by the default parameters of ChromHMM. States in the left column were annotated based on their closeness to the nearest transcription start sites (TSS) and nature of constituent marks. **l** Overlap enrichment analysis displaying chromatin state transitions between MCF10A control (SCR) cells (*Y*-axis) and MCF10A UBR7-sh1 cells (*X*-axis). The most significant state transitions include losses of H2BK120Ub/H3K79me2 low (States 1 to 5), H2BK120Ub/H3K79me2 high (States 2 to 1 or 3) and H3K79me2 only (States 3 to 5), which are highlighted by red circles. In **g**, **i**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). *P* values were calculated using two-tailed *t* tests. *$P < 0.05$; **$P < 0.001$

Table 1 Five chosen GO terms from DAVID on differentially expressed genes in UBR7-knockdown MCF10A cells (FDR <0.01; FC >2)

GO TERM	FDR
GO:0006414~ translational elongation	2.0E−43
GO:0007049~ cell cycle	5.3E−16
GO:0007155~ cell adhesion	1.3E−05
GO:0031497~ chromatin assembly	5.8E−05
GO:0016126~ sterol biosynthesis process	1.5E−03

DAVID Database for Annotation, Visualization and Integrated Discovery, *GO* gene ontology, *FC* fold change, *FDR* false discovery rate

---

marker Ki-67, tumors derived from mice, confirmed the anti-proliferative role of UBR7-WT, but not UBR7-CM (Fig. 4g). These results supported a tumor-suppressive role for UBR7 in breast cancer.

Consistent with a metastasis-suppressive role for UBR7 and its lower levels in metastatic tumors, UBR7-WT overexpressing MDA-MB-231 cells, but not UBR7-CM-overexpressing MDA-MB-231 cells, were unable to seed to the lung upon intravenous injection, whereas control cells formed overt lung metastases (Fig. 4h). Consistently, UBR7 loss in MCF10A cells enhanced invasion in a Matrigel chamber (Fig. 4i, j) and migration in a scratch assay (Fig. 4k, l and Supplementary Fig. 5a, b). Importantly, these phenotypes were rescued by overexpression of UBR7-WT, but not UBR7-CM, in MCF10A UBR7-sh1 cells (Fig. 4m–p), as well as in MDA-MB-231 and MDA-MB-468 cells (Supplementary Fig. 5c–h). Overall, these results establish the tumor- and metastasis-suppressive functions of UBR7 in triple-negative breast cancer.

UBR7 suppresses EMT. To gain insight into the molecular mechanism of these observations, we performed transcriptomic profiling of MCF10A cells harboring control and UBR7-specific shRNA using RNA-sequencing (RNA-seq) and identified 2348 up-regulated and 2576 downregulated genes (Fig. 5a, Supplementary Fig. 6a, b and Supplementary Data 3). We verified a subset of these genes using individual qPCR experiments (Supplementary Fig. 6c, d). Consistent with the cellular phenotypes, the misregulated genes exhibited enrichment in Cadherin and invasive breast cancer signatures (Fig. 5b, Supplementary Data 4 and Table 1). Importantly, we noted that UBR7-low (shRNA harboring) cells had characteristics of epithelial-to-mesenchymal

transition (EMT), a cellular process typically associated with breast cancer metastasis[5], as judged by loss of expression of epithelial markers (*CDH1*, *CLDN1*, *CLDN7*, and *CYTK18*) and gain of expression of mesenchymal markers (*CDH2*, *ZEB1*, *SNAI1*, *SNAI2*, *TWIST*, *VIM*)[6,30] in qPCR, western blot, and immunofluorescent analyses (Fig. 5c–f and Supplementary Fig. 7a–c). Overexpression of UBR7-WT in MCF10A UBR7-shRNA harboring cells (UBR7-sh1) rescued the expression of these markers, whereas UBR7-CM did not (Fig. 5g–i). Similarly, overexpression of UBR7-WT, but not UBR7-CM, reduced the expression of mesenchymal markers and induced the expression of epithelial markers in MDA-MB-231 and MDA-MB-468 cells (Fig. 5j, k and Supplementary Fig. 7d–g). Therefore, our results indicated that loss of UBR7 promotes EMT, a phenomenon thought to precede metastasis.

UBR7 suppresses EMT through activation of CDH4. To define the UBR7 gene targets that may drive the observed phenotypes, we overlapped sites that harbored loss of H2BK120Ub in UBR7-knockdown cells using ChIP-seq (7846, $p < 1e{-}8$) with differentially expressed genes (4924; $p < 0.01$; fold change >1.5) and found that H2BK120Ub targeted 318 downregulated and 117 up-regulated genes (Fig. 6a, Supplementary Fig. 8a, b, and Supplementary Data 5). Several such genes were enriched in cell–cell adhesion processes (Supplementary Data 6) including cadherins such as *CDH4* and *CDH13* (Fig. 6b and Supplementary Fig. 8c–e). *CDH4* (or R-cadherin) is suggested to play important roles in suppressing invasion of basal-type breast cancer[31]. Furthermore, *CDH4* expression exhibited consistent patterns to UBR7 in matched "normal" and "malignant" breast cancer cell lines (Fig. 6c and Supplementary Fig. 8f). Importantly, we detected UBR7 occupancy in *CDH4* locus in control MCF10A cells using ChIP-qPCR as well as rescue of H2BK120Ub levels in MCF10A UBR7-sh1, MDA-MB-231, and MDA-MB-468 cells by overexpression of UBR7-WT, but not UBR7-CM (Figs 2g, 6d–f). Next, we sought to determine whether CDH4 was in part responsible for anti-invasive phenotypes of UBR7. Overexpression of CDH4 in MCF10A and MCF12A UBR7-sh1 cells drastically reduced cellular invasion, migration, proliferation, and suppressing EMT, and established an epistatic relationship between UBR7 and CDH4 (Fig. 6g–m and Supplementary Figs 8g–j, 9a,b). Similarly, CDH4 overexpression drastically reduced the invasive properties of both MDA-MB-231 and MDA-MB-468 cells (Supplementary Fig. 9c–i). Overall, our results indicated that functional loss of UBR7 can be restored by cell–cell adhesion genes like *CDH4*.
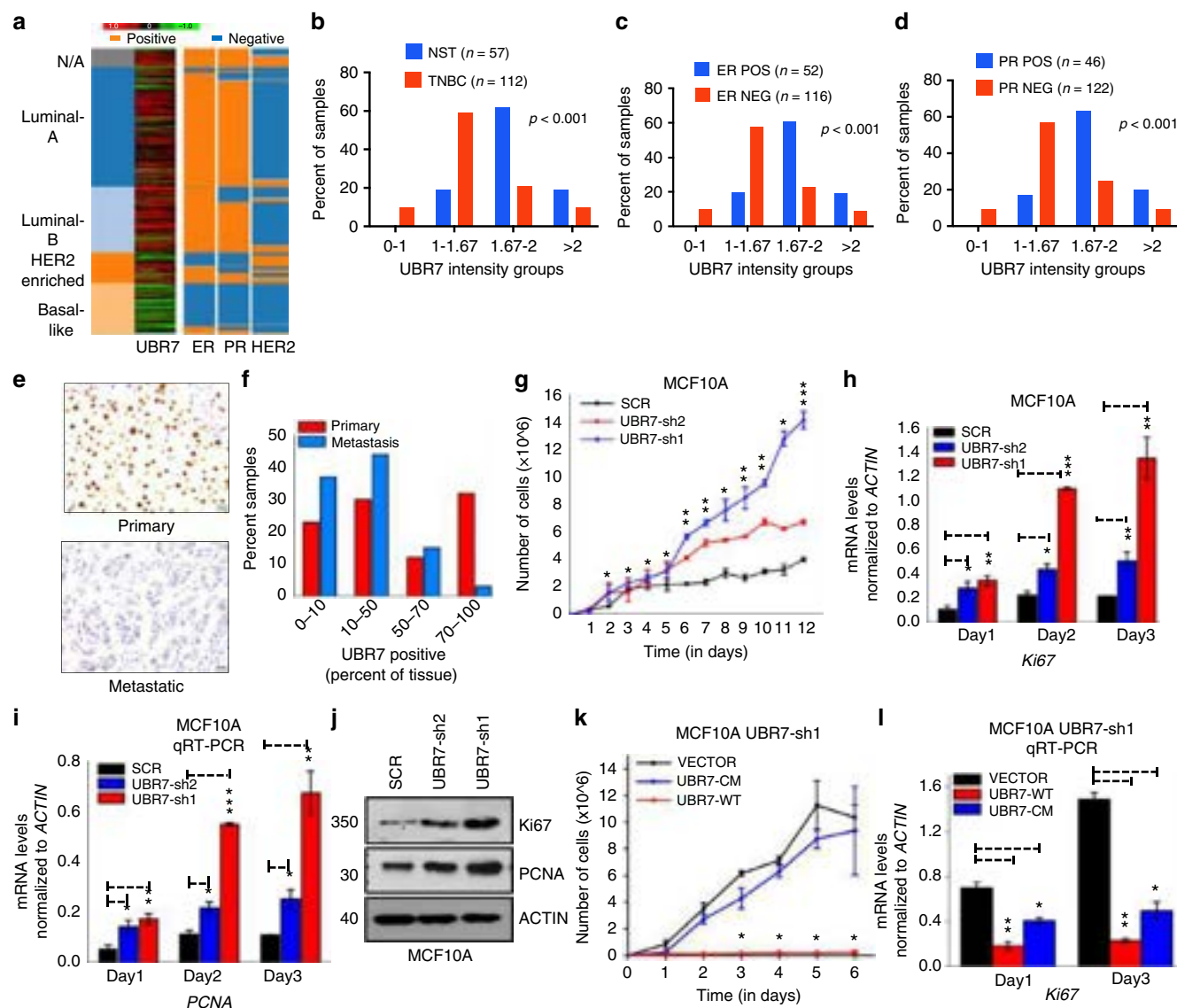
Fig. 3 UBR7 inhibits breast cancer cell proliferation. **a** Heatmap showing that *UBR7* expression is downregulated in triple-negative or basal-type breast cancer (analysis of data from TCGA database). Green and red indicates down- and up-regulated, respectively. **b–d** Immunohistochemical analysis representing low UBR7 expression analyzed in triple-negative (**b**), estrogen receptor (ER)-negative (**c**), and progesterone receptor (PR)-negative (**d**) breast tumors. **e, f** Representative images of primary and metastatic invasive breast carcinoma showing UBR7 expression as analyzed immunohistochemically. Scale bar indicates 20 μm. **g** Proliferation of MCF10A cells expressing scrambled (SCR) or UBR7 (UBR7-sh1/UBR7-sh2) short hairpin RNAs (shRNAs). **h, i** Quantitative real-time PCR (qRT-PCR) analysis of *Ki-67* (**h**) and *PCNA* (**i**) from cultured MCF10A cells expressing SCR or UBR7 (UBR7-sh1/UBR7-sh2) shRNAs in a time-dependent manner. **j** Immunoblots of MCF10A cells expressing SCR or UBR7-shRNAs to monitor expression of Ki-67 and PCNA. ACTIN was used as a loading control. **k** Proliferation of cultured MCF10A UBR7-sh1 cells expressing a vector (VECTOR), WT (UBR7-WT), or catalytic-mutant (UBR7-CM). **l** qRT-PCR analysis of *Ki-67* in MCF10A UBR7-sh1 cells expressing vector (VECTOR), wild-type (UBR7-WT), or CM (UBR7-CM) in a time-dependent manner. In all panels, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). *P* values were calculated using two-tailed *t* tests. *$P < 0.05$; **$P < 0.001$; ***$P < 0.0001$

UBR7 silencing activates Wnt/β-catenin signaling. The Wnt/β-catenin signaling pathway is deregulated in several cancers, including breast cancer[32]. It promotes tumor initiation, maintenance and metastasis[32]. We observed activation of the canonical Wnt/β-catenin signaling cascade upon loss of UBR7 as evident from the upregulation of key positive regulators including WNT3A, FZD2/3, LRP5/6, ROR2, and DSH2, and down-regulation of negative regulators such as DKK1 (Fig. 7a, b and Supplementary Fig. 9j). Activation of the signaling pathway was further confirmed by nuclear localization of β-catenin upon loss of UBR7 (Fig. 7c). This could serve as a mechanism downstream

of CDH4 as other cadherins, such as CDH1, are known to regulate Wnt signaling[33,34]. Indeed, restoration of CDH4 altered the nuclear localization of β-catenin to the cytoplasm and down-regulated known β-catenin target genes including *AXIN2*, *CCND1*, *C-MYC*, *COX2*, and *MMP7* (Fig. 7d–f and Supplementary Fig. 9k), thereby maintaining the epithelial state of the cell. Cytoplasmic fraction of β-catenin is phosphorylated by glycogen synthase kinase 3β (GSK3β), which marks the protein for β-transducin repeat-containing protein (β-TrCP) mediated proteasomal degradation in the absence of an activating ligand for the signaling pathway. Thus, we examined the association of
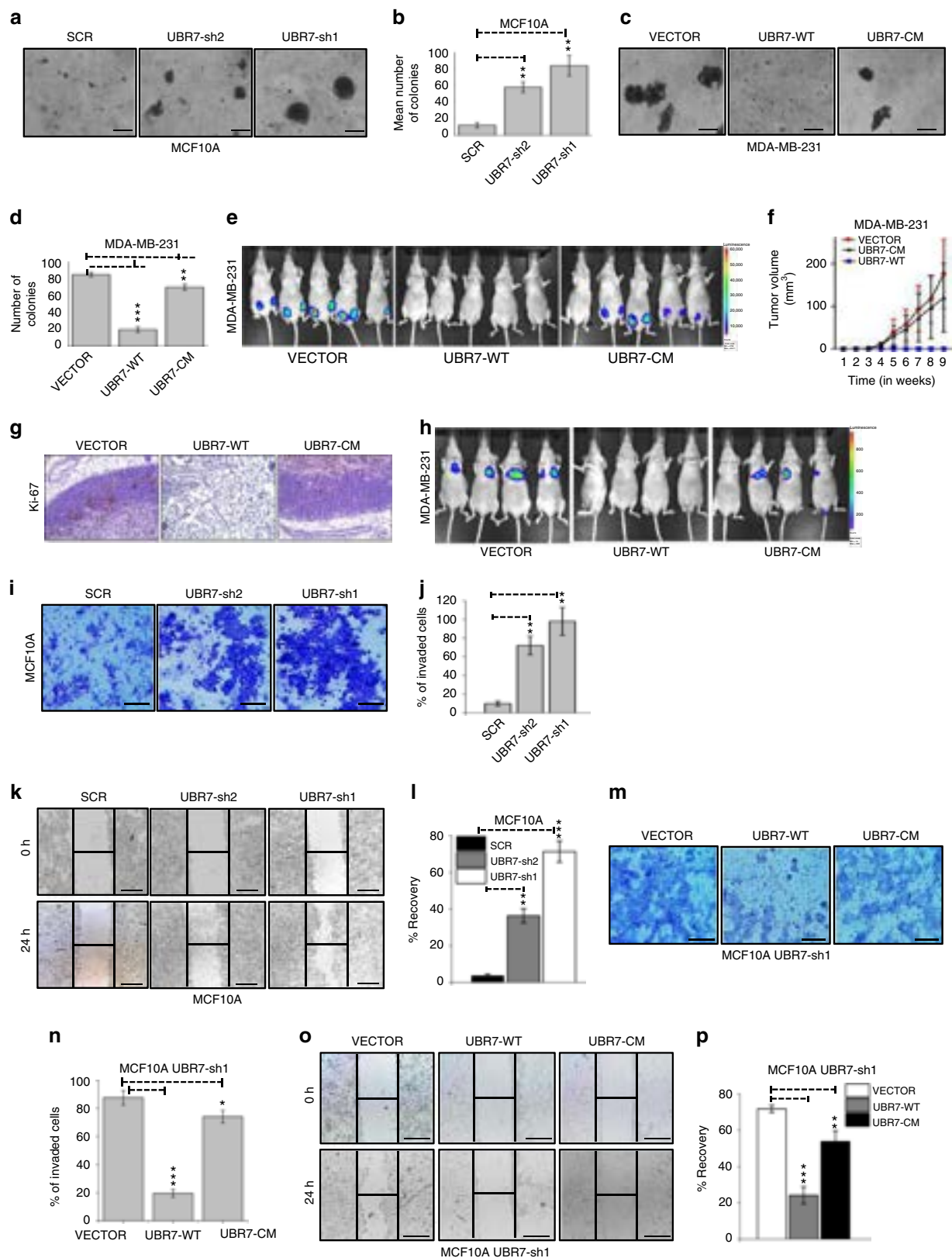
β-catenin with this degradation complex in the absence of UBR7. We found a weak association of this degradation complex, which includes GSK3β and β-TrCP with β-catenin in the absence of UBR7, indicating escape of β-catenin from the degradation pathway, and entry into the nucleus, thereby activating its target genes (Fig. 7g). Overall, our results demonstrated that UBR7 loss activates the Wnt/β-catenin signaling pathway, which is inhibited upon CDH4 restoration.

Fig. 4 UBR7 acts as tumor metastasis suppressor gene. **a–d** Soft agar assay with MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs (UBR7-sh1 or -sh2) (**a**, **b**) or with MDA-MB-231 cells expressing vector (VECTOR), wild-type (UBR7-WT), or catalytic-mutant (UBR7-CM) (**c**, **d**). **e**, **f** Tumor formation in mice injected into mammary fat pad with MDA-MB-231 cells expressing an empty vector (VECTOR) or WT (UBR7-WT) or CM (UBR7-CM). $n = 5$ mice per group. **g** Immunohistochemistry (IHC) for proliferation marker Ki-67 in tumors derived from control cells versus UBR7-WT- or UBR7-CM-overexpressing cells. Image was taken at ×20 magnification. Scale bar indicates 50 μm. **h** Tumor metastasis in mice injected into tail vein with MDA-MB-231 cells expressing empty vector (VECTOR) or WT (UBR7-WT) or CM (UBR7-CM). $n = 4$ mice per group. **i**, **j** Invaded MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs in a Matrigel chamber were photographed and counted. **k**, **l** Wound healing/migration rate of MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs was monitored. **m**, **n** Invaded MCF10A UBR7-sh1 cells expressing WT (UBR7-WT) or CM (UBR7-CM) in a Matrigel chamber were photographed and counted. **o**, **p** Wound healing/migration rate of MCF10A UBR7-sh1 cells expressing WT (UBR7-WT) or CM (UBR7-CM) was monitored. In **a**, **c**, **i**, **k**, **m**, **o** scale bar indicates 10 μm. In **b**, **d**, **j**, **l**, **n**, and **p**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). In **f**, error bars indicate standard deviation (s.d.); $n = 5$ mice per group. P values were calculated using two-tailed t tests. *$P < 0.05$; **$P < 0.001$; ***$P < 0.0001$

Discussion

PHD fingers are structurally conserved chromatin-binding modules present in proteins that are associated with chromatin and regulate gene transcription[35]. The versatile function of a PHD finger as an epigenome reader promotes the recruitment of multi-protein complexes to change the chromatin structure, thereby augmenting transcription (initiation, elongation, and termination) or its repression[35]. A PHD finger fold consists of a C-terminal α-helix and two strands of anti-parallel β-sheet, with a conserved Cys4-His-Cys3 motif in a cross-brace topology, which anchors two zinc atoms[36]. PHD fingers are well known for reading methylated H3 lysine 4 or unmethylated N-term tail of H3, with a few exceptions. The RING finger, which is well characterized for its E3 ubiquitin ligase activity, and has a Cys3-His-Cys4 motif organization, is reciprocal to the PHD finger[37]. The UBR7-PHD finger, on the other hand, has a unique motif organization (Cys4-His2-Cys2) different from that of a canonical PHD or RING finger and hence can be considered as an "atypical" PHD finger. We report here an E3 ubiquitin ligase activity by a PHD finger of UBR7 toward histone H2B at lysine 120. Of note, in comparison with other E3 ubiquitin ligases, which act in a complex[25], UBR7 can promote enzymatic catalysis in isolation. Indeed, mutating the two His residues, which is instrumental to anchoring the zinc coordination complex, abrogates the catalytic activity. Structural studies will provide more insight into the molecular mechanism of H2BK120Ub catalysis by this atypical domain. Also, a detailed investigation of its reader function could provide more insight into the role of the UBR7-PHD finger in cellular context.

UBR7 loss not only drastically reduced H2BK120Ub but also significantly reduced H3K79Me2 without much effect on H3K4Me3, H3K27Ac, H3K27Me3, and H3K9Me3. This was intriguing because several studies have suggested that H2BK120Ub marked nucleosomes act as templates for DOT1L and SET/COMPASS complexes for making H3K79Me2 and H3K4Me3, respectively[38–40]. Interestingly, in the chromatin state analyses, cells with UBR7 loss also harbored changes in transcription states that had predominance of H3K79Me2. Since H3K79Me2 mark is linked with transcriptional elongation and observed on genes that are being actively transcribed, we propose that UBR7 may play important roles in transcriptional elongation. Further biochemical studies focusing on isolation of UBR7 protein complexes or identification of interacting partners will be needed to determine its exact function in this process, if any.

Importantly, we found UBR7 loss to be highly correlated with triple-negative and basal-like breast cancer. Although the molecular features that define triple-negative breast cancer (loss of ER, PR, and Her2 expression) are clear, ambiguity exists in the definition of basal-like cancers. UBR7 loss may be a key determining feature of this aggressive subtype of breast cancer. Notably, we established that UBR7 can suppress breast tumor formation and metastasis in vivo. One possibility of such an event is through UBR7's prevention of a "self-seeding" event, in which local invading cancer cells lead to tumor formation via fusion of propagating colonies[41]. Also, it is difficult to determine the relative contribution of proliferative versus invasive role for UBR7 in its pro-metastatic function. Overall, UBR7 loss may be a predictive biomarker and provide specific vulnerabilities to epigenetic inhibitors given its drastic impact on chromatin states. Furthermore, given the substantial impact of UBR7 on maintenance of the epithelial state and inhibition of the plasticity of a cell, determining its role in suppression of other malignancies requires further investigation.

We demonstrated that CDH4/R-cadherin is a major target downstream of UBR7. Cadherins are crucial in the maintenance of cell boundary, tissue morphogenesis, and cell polarity[42], and aberrant function may lead to severe metastatic neoplasia[33]. Whereas UBR7 loss promotes breast tumor metastasis, CDH4 overexpression provides only a partial rescue of such phenotypes. CDH4 was found to be highly expressed in mammary epithelial cells, but severely down-regulated in invasive ductal carcinoma[31]. Moreover, due to heterogeneity of cancer cells, CDH4 was absent in those cells that were poorly differentiated[31]. Although we showed that UBR7 prevents metastatic colonization and represses genes that are crucial for breast cancer metastasis by targeting CDH4, other downstream pathways and target genes may also have important roles. For example, other proteins, such as integrins and cytokines, have already demonstrated roles in breast cancer etiology and metastasis[43].

Our results indicate that CDH4 may control β-catenin signaling in a manner similar to E-cadherin-mediated control of β-catenin signaling, which is well documented to play important roles in metastasis. Loss of E-cadherin leads to β-catenin release from the cell surface, resulting in escape from the β-TrCP-mediated degradation pathway, and thereby promoting its nuclear localization and activation of target genes[32,34,44]. In UBR7-depleted cells, Wnt/β-catenin signaling was one of the top misregulated pathway, and overexpression of CDH4 downregulates Wnt signaling in UBR7-depleted cells by reinstating β-catenin to the cell membrane and reducing the nuclear pool. In addition, we observed that several other Wnt signaling regulators (such as WNT3A, FZD2/3, LRP5/6, ROR2, DSH2, and DDK1) were transcriptionally controlled by UBR7, strongly suggesting that UBR7 is a key mediator of the Wnt/β-catenin signaling cascade. Collectively, our results demonstrated that UBR7 is a H2B E3 ubiquitin ligase that suppresses triple-negative subtype of breast cancer by activating CDH4/R-cadherin expression and inhibiting the canonical Wnt/β-catenin signaling pathway.
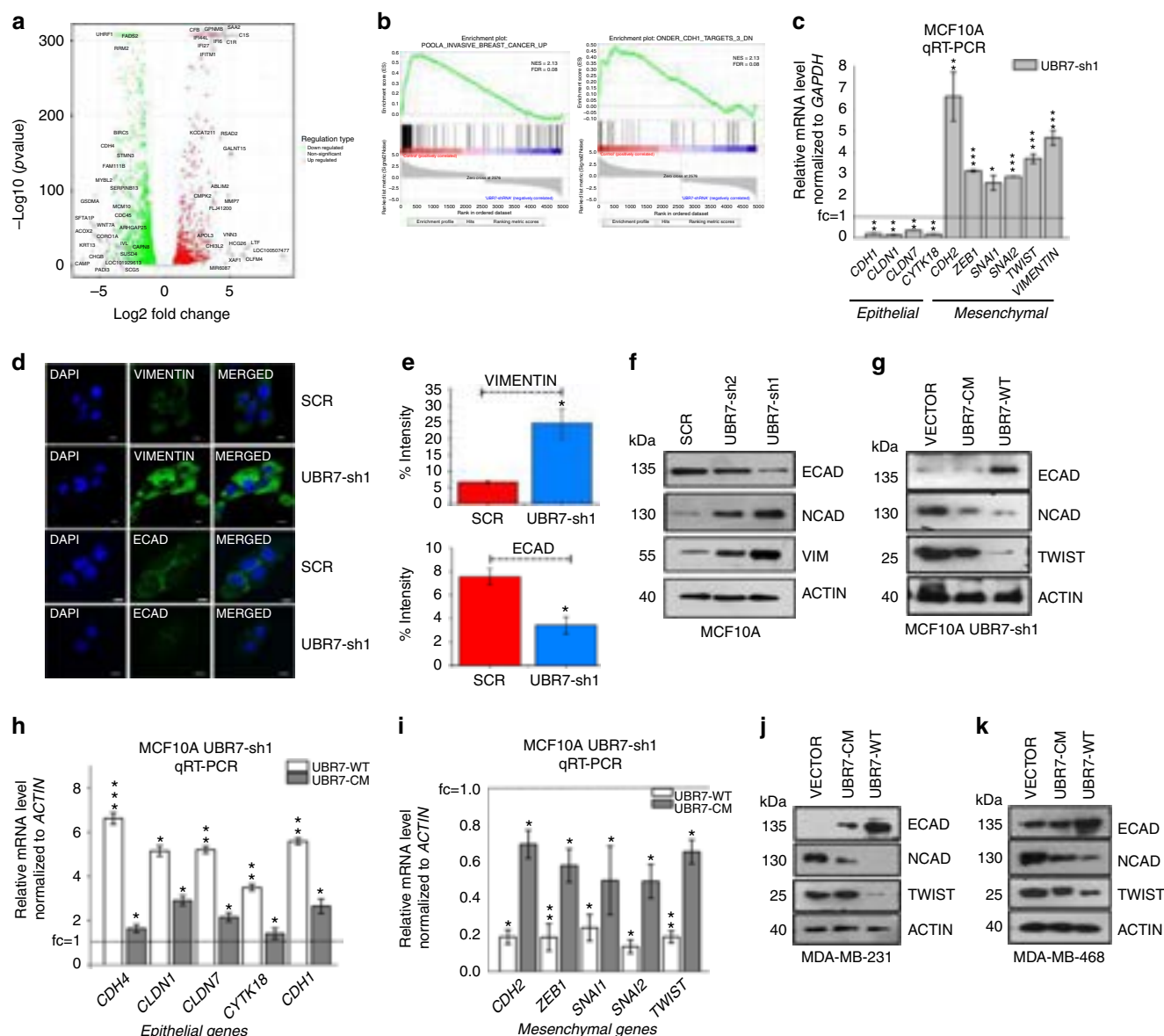
Fig. 5 UBR7 is an epithelial-to-mesenchymal transition (EMT) suppressor. **a** Volcano plots of the differentially expressed genes upon knockdown of UBR7 (using UBR7-sh1) in MCF10A cells scored via RNA-sequencing (RNA-seq) analysis. Green, red and grey indicates down-, up-regulated and non-significant, respectively. **b** Gene Set Enrichment Analysis (GSEA) output images of two chosen pathways displaying a correlation of differentially regulated genes in UBR7-knockdown MCF10A cells with the "poola_invasive_breast_cancer_upregulated" set and "Onder_CDH1_targets_downregulated" set. **c** Quantitative real-time PCR (qRT-PCR) analysis of EMT signature gene expression upon UBR7 knockdown (UBR7-sh1) in MCF10A cells. **d, e** Immunofluorescence studies showing expression of VIMENTIN and ECAD after UBR7 knockdown (UBR7-sh1) in MCF10A cells (**d**). Scale bar indicates 100 μm. Percent intensity is quantified (**e**). **f, g** Immunoblots of MCF10A cells expressing scrambled (SCR) or UBR7 short hairpin RNAs (shRNAs) (UBR7-sh1 and sh2) (**f**) or MCF10A UBR7-sh1 expressing wild-type (UBR7-WT) and catalytic-mutant (UBR7-CM) (**g**), monitoring expression of candidates regulating EMT. ACTIN was used as a loading control. **h, i** qRT-PCR analysis of epithelial (**h**) and mesenchymal (**i**) genes in MCF10A UBR7-sh1 expressing WT (UBR7-WT) and CM (UBR7-CM). **j, k** Immunoblots of MDA-MB-231 (**j**) and MDA-MB-468 (**k**) cells expressing WT (UBR7-WT) and CM (UBR7-CM) monitoring expression of candidates regulating EMT. ACTIN was used as a loading control. In **c**, **h**, **i**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. *$P < 0.05$; **$P < 0.001$; ***$P < 0.0001$

Methods

Cell lines and cell culture. HEK293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% antibiotic–antimycotic (Gibco) at 37 °C and 5% CO_2. MCF10A and MCF12A cells were maintained in DMEM/Ham's F12 supplemented with 5% horse serum (Gibco), epidermal growth factor (EGF), insulin, hydrocortisone, cholera toxin (Sigma), and 1% antibiotic–antimycotic. MDA-MB-231, MDA-MB-468, MCF7, and T47D cells were maintained in RPMI-1640 medium (Gibco) supplemented with 10% FBS, insulin, and 1% antibiotic–antimycotic. All cell lines were purchased from ATCC. 16N, 21PT, and 21MT2 (provided by R. Weinberg, Whitehead Institute, Massachusetts Institute of Technology) were maintained in DMEM supplemented with 10% FBS, insulin, hydrocortisone, EGF,

and 1% antibiotic–antimycotic. 4T1 and 4T07 (provided by R. Weinberg) were maintained in DMEM/Ham's F12 medium supplemented with 10% FBS, insulin, hydrocortisone, and 1% antibiotic–antimycotic. All cell lines used in the study were negative for mycoplasma. All cell lines were validated by MD Anderson Cancer Center Characterized Cell line core facility via DNA fingerprinting. For transient transfection, cells were counted and seeded in 12-well or 6-well or 6-cm dishes and then subjected to overexpression using Lipofectamine-2000 (Invitrogen) as per the manufacturer's protocol.

Protein purification. The full-length *UBR7* complementary DNA (cDNA) sequence, UBR domain alone, or PHD finger alone was cloned in a pDEST15

The figure panels (a–m) show molecular biology experimental data.
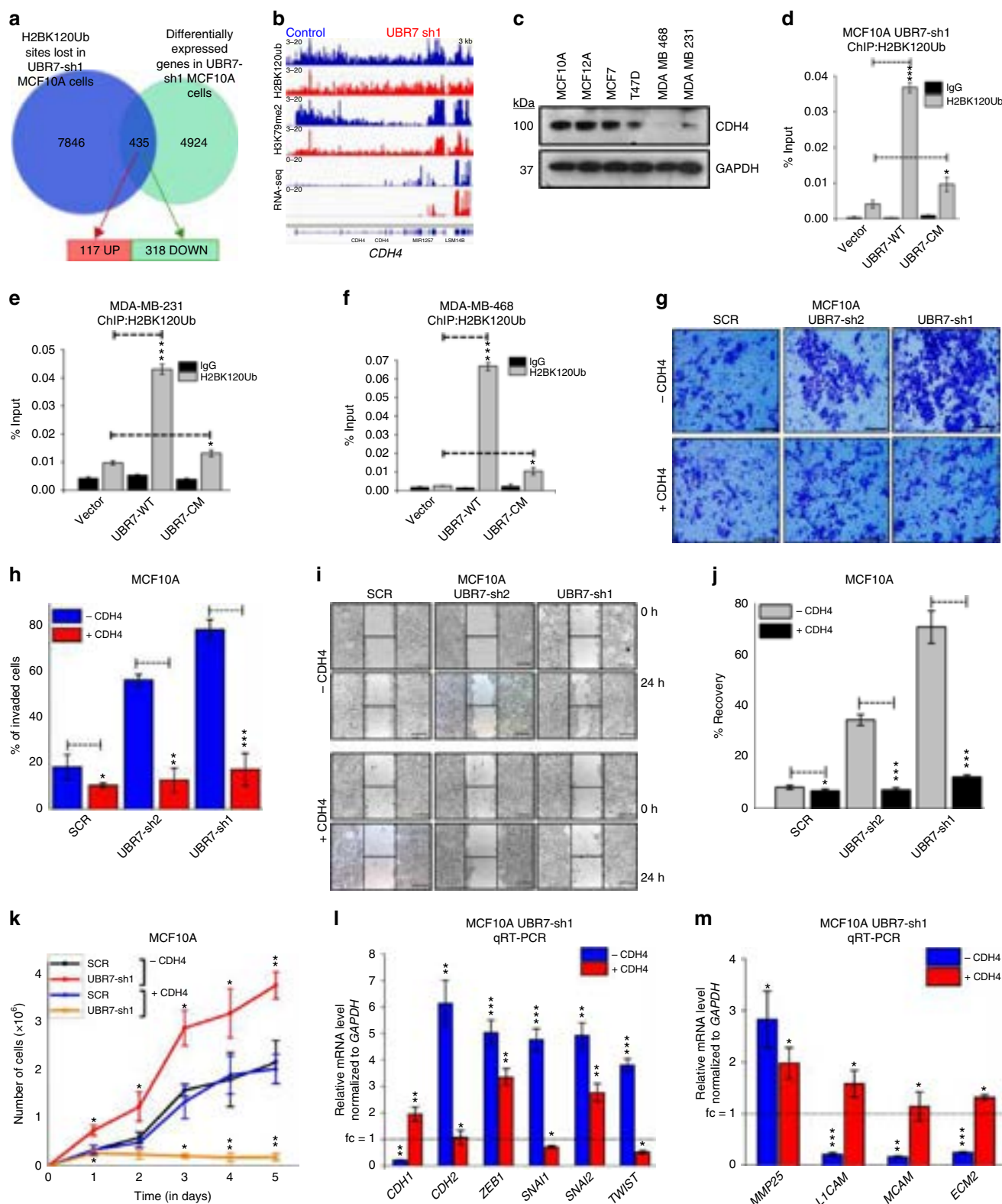
vector (GATEWAY cloning system, Invitrogen) and sequence verified. The protein used in in vitro assays were purified as described previously[45]. Briefly transformed cells were grown till 0.8 optical density (OD) and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (Sigma) at 20 °C for 16 h. The pellets were resuspended in lysis buffer and lysed mildly followed by glutathione sepharose beads binding and washing with wash buffer. The proteins were eluted and purified further to homogeneity via gel filtration chromatography using a Superdex75 column (GE Healthcare). CM derivative (H163S/H166S) were generated using a QuikChange site-directed mutagenesis kit (Stratagene) as per the standard protocol[46].

Nucleosome isolation. Nucleosomes were prepared freshly from HeLa cells as described previously[45]. Briefly, the nuclear pellet from HeLa cells was digested with MNase (0.2 Units/µl; Sigma) and extracted with TE buffer for 1 h. The mono-nucleosome was separated via sucrose gradient (5–40%) ultracentrifugation using Sorvall WXUltra100 (Thermo Fischer Scientific) with AH650 rotor for 16 h at 207,203 × g. For further analysis, these fractions were pooled and concentrated.

In vitro ubiquitination assay. In vitro ubiquitination reactions were set up with purified UBR7-PHD, UBR7-WT, or UBR7-CM (H163S/H166S) as E3 enzymes and

Fig. 6 UBR7 suppresses epithelial-to-mesenchymal transition (EMT) through activation of CDH4. **a** Venn diagram showing overlaps of H2BK120Ub enriched and differentially regulated genes after UBR7 knockdown. **b** Integrative Genomics Viewer (IGV) view of H2BK120Ub and H3K79me2 chromatin immunoprecipitation sequencing (ChIP-seq) and RNA-sequencing (RNA-seq) tracks on the *CDH4* gene in Control (SCR) or UBR7-sh1 short hairpin RNA (shRNA) expressing MCF10A cells. **c** Immunoblot showing the expression of CDH4 across different normal breast cell and breast cancer cell lines. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a loading control. **d–f** Bar plots showing H2BK120Ub ChIP in the *CDH4* gene locus in MCF10A UBR7-sh1 (**d**) or MDA-MB-231 (**e**) or MDA-MB-468 (**f**) cells expressing wild-type (UBR7-WT) and catalytic-mutant (UBR7-CM). **g, h** Invasion of MCF10A cells overexpressing CDH4 in the presence (SCR) and absence (UBR7-sh1 and UBR7-sh2) of UBR7 in a Matrigel chamber was photographed and quantitated. **i, j** Wound healing by MCF10A cells overexpressing CDH4 in the presence (SCR) and absence (UBR7-sh1 and UBR7-sh2) of UBR7 was photographed and the percent recovery was measured over time. **k** Proliferation of cultured MCF10A cells overexpressing CDH4 in the presence (SCR) and absence (UBR7-sh1) of *UBR7*. **l, m** Quantitative real-time PCR (qRT-PCR) analysis of EMT signature genes (**l**) and cell adhesion-linked genes (**m**) in UBR7-sh1 MCF10A cells upon CDH4 overexpression. In **g, i**, scale bar indicates 10 μm. In **d–f, h, j–m**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. *$P < 0.05$; **$P < 0.001$; ***$P < 0.0001$
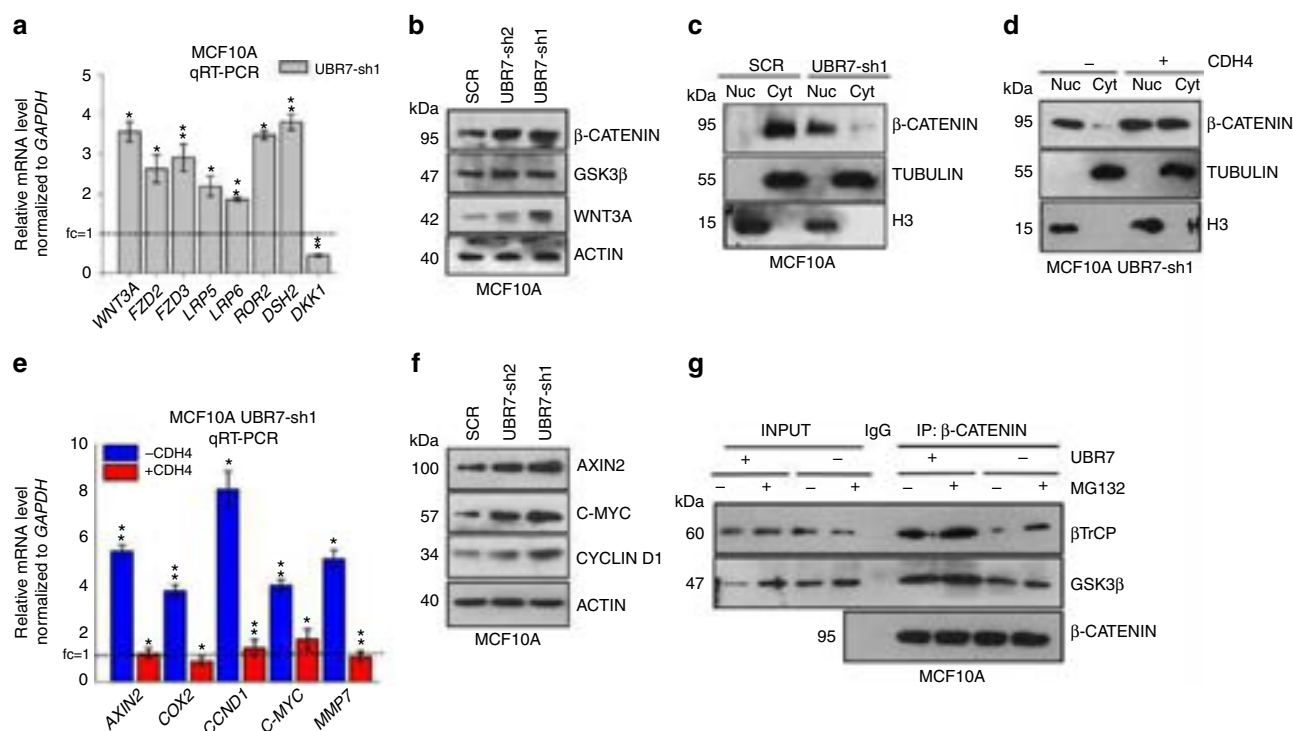
Fig. 7 UBR7 silencing activates the Wnt/β-catenin signaling pathway. **a** Quantitative real-time PCR (qRT-PCR) analysis of key regulators of the Wnt/β-catenin signaling pathway upon loss of UBR7 (UBR7-sh1). **b** Immunoblots monitoring expression of β-CATENIN, glycogen synthase kinase 3β (GSK3β), WNT3A, and ACTIN (loading control) in MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs. **c** Immunoblots monitoring expression of β-CATENIN in the nuclear and cytoplasmic fractions of MCF10A cells expressing scrambled (SCR) or UBR7-sh1shRNA. Histone H3 and TUBULIN were used as loading controls for nuclear and cytoplasmic lysates, respectively. **d** Immunoblots monitoring expression of β-CATENIN in the nuclear and cytoplasmic fractions of *CDH4*-expressing MCF10A UBR7-sh1 cells. Histone H3 and TUBULIN were used as loading controls for nuclear and cytoplasmic lysates, respectively. **e** qRT-PCR analysis of β-catenin target genes expression in UBR7-sh1 MCF10A cells upon CDH4 overexpression. **f** Immunoblots showing expression of β-catenin target genes upon UBR7 knockdown (UBR7-sh1 or -sh2) in MCF10A cells. **g** Co-immunoprecipitation of β-catenin from UBR7-knockdown MCF10A cells in the presence or absence of MG132 (20 μM for 18 h) showing association of β-catenin with GSK3β and β-transducin repeat-containing protein (β-TrCP). In **a, e**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. *$P < 0.05$; **$P < 0.001$

recombinant H2B, H2A/H2B dimer, core histones octamer, or purified nucleosome as a substrate using Ubiquitinylation kit (Cat #: BML-UW9920, Enzo Life Sciences) as per the manufacturer's protocol. Briefly, the reaction was carried out in ubiquitinylation buffer containing 100 U/ml inorganic pyrophosphatase, 1 mM dithiothreitol (DTT), and 5 mM EDTA (negative control). 2.5 μM of ubiquitin (biotinylated) was incubated with 100 nM E1, 2.5 μM E2 (for UBR7, UbcH6 acts as the E2), 5 mM Mg-ATP, and 100 nM E3 along with 1 μM substrate at 37 °C for 1 h. The reaction was stopped, and trichloroacetic acid precipitation was performed and analyzed via western blotting with antibodies against H2BK120Ub, H2B, H2A, H3, and H4 antibodies.

Peptide pull-down assay. A peptide pull-down assay was performed as described previously[45]. Briefly, equivalent amounts of peptides and protein were incubated in immunoprecipitation (IP) buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.05%

NP-40, 1 mM DTT). The complex was pulled down with streptavidin beads, washed with the IP buffer, and eluted and analyzed using western blotting.

GST pull-down assay. GST and GST-fusion proteins were incubated with the recombinant histones H3, H4, H2A, and H2B at equimolar ratios in a IP buffer overnight as described previously[45]. The complex was pulled down with glutathione sepharose bead (GE Healthcare), washed with IP buffer, and analyzed using western blotting with specific antibodies. Ten percent of the histone proteins (H3, H4, H2A, and H2B) were used as inputs.

Co-immunoprecipitation. Cells were subjected to co-immunoprecipitation as delineated previously[45]. In brief, cells were lysed in lysis buffer (20 mM Tris pH 8.0,

150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 1 mM EDTA) and pulled down with specific antibodies or FLAG M2 beads, followed by washes with the lysis buffer. The immunoprecipitant was analyzed using western blotting. Ten percent of the lysate with which IP was set was used as the input.

RNA interference through lentiviral production. shRNA plasmids for UBR7 with pLKO.1-puro backbone (Sigma-Aldrich) were screened for efficient knockdown. Two of seven shRNAs were selected for subsequent experiments. Their sequences are as follows: UBR7-sh1 5′-CAGTGCACCCAGGGGTTATTTG-3′ and UBR7-sh2–5′-GCTTAAAGCTAAGCAGCTTAT-3′. UBR7-sh1 targets the 3′-UTR of the gene, so the overexpression constructs were resistant to the shRNAs. HEK293T cells were plated at a density of 3×10^5 cells in 10-cm dishes. Eight micrograms of shRNA and packaging vectors were transfected into the cells as described previously[47]. Transduced cells were selected using puromycin (10 μg/ml; Sigma) for 3 days.

UBR7 overexpression via lentiviral production. WT and H163S/H166S CM of UBR7 were cloned into the pHAGE-CMV-fullEF1a-IRES-ZsGreen (from Jeng-Shin Lee; Dana-Farber/Harvard Cancer Center). 293T cells were plated at density of 3×10^5 in 10-cm dishes. Recombinant lentiviral particles were produced using Lipofectamine 2000-mediated transient transfection in HEK293T cells. Briefly, 8 μg of overexpression vectors, a packaging vector (psPAX2), and an envelope vector (pMD2.G) were transfected into 293T cells plated in 10-cm dishes. The viral supernatant was harvested 48 and 72 h after transfection and filtered. Cells were infected three times in 48 h with the viral supernatant containing 10 μg/ml Polybrene. Green fluorescent protein-positive cells were sorted and cultured for other experiments.

Quantitative real-time PCR. Total RNA was isolated using TRIzol reagent (Invitrogen) and reverse transcribed using a Revertaid First Strand cDNA Synthesis kit (Thermo Fischer Scientific) according to the manufacturer's protocol followed by qRT-PCR using ABI-SYBR GREEN mix (Applied Biosystems). qRT-PCR was performed using StepONE plus FAST Real-time PCR machine. Each sample was analyzed independently three times and the results of one representative experiment, with technical triplicates, are shown. List of primers is provided in Supplementary Table 1.

Western blot analysis. Whole-cell lysates were prepared in lysis buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM EDTA) as described previously[45]. The lysates were electrophoresed on 7.5%, 11%, or 15% SDS-polyacrylamide gel electrophoresis gels. Blots were probed with specific antibodies. The membrane was blocked with 5% bovine serum albumin (BSA) or non-fat dry milk in TBST (Tris-buffered saline, 0.1% Tween-20) . The antibodies used are listed in Supplementary Table 2. All uncropped images of the blots are provided in Supplementary Fig. 10.

Immunofluorescence. Immunofluorescence was carried out as described previously[45]. Cells were fixed with 4% paraformaldehyde or methanol, permeabilized with 1% Triton X-100 or 100% methanol, blocked with 3% BSA, and stained with the indicated antibodies. Coverslips were mounted after staining with DAPI (4′,6-diamidino-2-phenylindole) and photographed using a Nikon T1E confocal microscope with an A1RMP Scanner Head.

Migration or wound healing assay. Migration assay was performed as described previously[48]. Cells were plated in 6-well dish in triplicates and grew to 90% confluence and the cell surface was scratched with a sterile 20 μl tip. Images were captured with Nikon T1 E100 microscope at 0 h and post 24 h after the scratch to measure the wound healing by the cells, indicating their migratory capacity and migration rate. The images were analyzed using the Image J software program (National Institutes of Health; NIH), to measure the wound recovery by the cells indicating their migratory ability.

Proliferation assay. Transduced and transfected cell lines were plated at a density of 1×10^6 cells in triplicates in a 24-well plate. MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) was added at indicated time points and MTT assay was performed as described previously[49]. The OD was measured using an Epson Plate reader. The number of proliferating cells was calculated from the standard curve.

Invasion assay. Invasion assay was performed as described previously[47]. For this assay, 1×10^6 cells were cultured in the upper well of Matrigel chambers (Corning) containing serum-free medium and allowed to migrate towards serum-enriched medium in the bottom wells. After 20 h of incubation, invading or migrating cells were stained with 0.5% Giemsa, photographed with Nikon T1 E100 microscope, and counted using the Image J software program.

Soft agar assay. Transduced human breast epithelial and cancer cell lines were resuspended in 3 ml of soft agar (medium containing 0.3% noble agar [Affymetrix]) warmed to 45 °C. The cell suspension was layered onto 3 ml of bottom agar (medium containing 0.8% noble agar) in a 6-well plate (six replicates). 2 ml of the medium was added to the top agar and changed every 3 days. Visible colonies were scored after 4–5 weeks and stained with 0.5% Giemsa. The number of colonies and mean area of colonies was calculated using the Image J software program.

Animal experiments and in vivo imaging. Animals used in these experiments were all female nude mice aged 8–10 weeks (The Jackson Laboratory). All cells used for in vivo injections were labeled with red fluorescent protein luciferase. The mice were anesthetized with isoflurane and injected with 2×10^6 MDA-MB-231 cells in 50 μl of phosphate-buffered saline (PBS) in the abdominal mammary fat pads on both sides ($n = 5$ mice). Tumor growth was monitored weekly via caliper measurement and bioluminescent imaging once every 2 weeks. Once the largest tumor diameter was reached (1.5 cm, which is the maximal tumor diameter allowed under our institutional protocol), the animals were sacrificed. For the tail vein injections, the mice ($n = 4$) were injected with 100 μl of 1×10^6 cells were injected into the medial tail vein. All mouse experiments were performed with the approval of the MD Anderson Institutional Animal Care and Use Committee.

For in vivo imaging, cells were infected with EF1-RFP-T2A-Luciferase (System Biosciences) to enable stable expression of firefly luciferase. All in vivo bioluminescent imaging was carried out at the MD Anderson Small Animal Imaging Facility. For this imaging, the animals were anesthetized with isoflurane. They were injected intraperitoneally with 3 mg of D-luciferin (Perkin Elmer) and imaged using the IVIS Spectrum Imaging System (Perkin Elmer). Analysis after acquisition was done using the Living Image software program (version 4.3; Perkin Elmer).

Tissue microarrays. Patients: Breast tumors and normal breast tissue were obtained from patients who underwent surgery at MD Anderson Cancer Center from 2005 to 2015. Information on their hormone and Her2 statuses, as well as proliferation fraction (Ki-67), was retrieved from pathological and clinical reports, ER information was available for 367 (99%) patients, PR information was available for 368 (99%) patients, Her2 information was available for 362 (98%) patients, and proliferation fraction (Ki-67) information was available for 208 (56%) patients.

Breast cancer TMAs from MD Anderson Cancer Center and a commercially available TMA BR2082a (US Biomax, Inc. Rockville, MD, USA) were used. The TMAs from MDACC included 538 breast cancers and 15 normal tissues. Breast cancers were represented in 389 (60%) cases with 3 punches of 1 mm, in 96 (17.8%) cases with 3 punches of 0.6 mm, and in 53 cases (9.9%) with 3 punches of 1 mm and 3 punches of 0.6 mm. The normal tissue samples included six punches of 0.6 mm from three normal lymph nodes, three normal breast tissues, three normal kidneys, three normal colon, and three normal lung samples. The composition of the commercial TMA is shown at the vendor's website (http://www.biomax.us/tissue-arrays/Breast/BR2082a).

Immunohistochemistry. For validation of the anti-UBR7 antibody, Western blot of UBR7 expression on the whole cell line lysate of MCF10A transfected with a control small interfering RNA (siRNA) and siRNA against UBR7, and of MCF7 transfected with siRNA against UBR7 was performed. These cell lines were paraffin embedded and the final protocol was established to perform staining on 5-μm-thick TMA sections. In brief, epitope retrieval was performed with citrate buffer at pH 6.0 for 20 min, followed by peroxidase blocking for 5 min. A polyclonal rabbit anti-UBR7 antibody (Bethyl Laboratories) was then incubated for 60 min using a 1:2000 dilution followed by polymer (goat-anti-rabbit immunoglobulin G (IgG)) and 3,3′-diaminobenzidine incubation for 8 and 10 min, respectively. Slides were counterstained with hematoxylin. The staining was performed using supplies and an autostainer from Leica Biosystems.

For evaluation of UBR7 staining, nuclear staining was scored semi-quantitative providing the percentage of stained cells and staining intensity (0 = no staining, 1 + = weak staining, 2 + = moderate staining, and 3 + = strong staining). Representative staining examples are shown in Supplementary Fig. 4a.The study was approved by the MD Anderson Institutional Review Board.

ChIP assay. ChIP assays were performed as described earlier[50]. Cells were cross-linked with 1% formaldehyde and the chromatin was sheared and immunoprecipitated with the UBR7 antibody (Bethyl Laboratories), H2BK120Ub antibody (Millipore), H2B antibody (Abcam), or as a negative control IgG. ChIP DNA was analyzed by qPCR using gene specific primers. Each ChIP experiments were performed three independent times with technical triplicates.

ChIP-seq assay. ChIP assays were performed as described previously[50] with minor modifications. Briefly, ~2×10^7 cells were harvested via cross-linking with 1% (wt/vol) formaldehyde for 10 min at 37 °C with shaking. After quenching with 150 mM glycine for 10 min at 37 °C with shaking, cells were washed twice with ice-cold PBS and frozen at −80 °C for further processing. Cross-linked pellets were thawed and lysed on ice for 30 min in ChIP harvest buffer (12 mM Tris-Cl, 1 × PBS, 6 mM EDTA, 0.5% SDS) with protease inhibitors (Sigma). Lysed cells were sonicated with

a Bioruptor (Diagenode) to obtain chromatin fragments (~200–500 bp) and centrifuged at $15,000 \times g$ for 15 min to obtain a soluble chromatin fraction. In parallel with cellular lysis and sonication, antibodies ($5 \,\mu g/3 \times 10^6$ cells) were coupled with 30 μl of magnetic protein G beads in binding/blocking buffer (PBS + 0.1% Tween + 0.2% BSA) for 2 h at 4 °C with rotation. Soluble chromatin was diluted five times using ChIP dilution buffer (10 mM Tris-Cl, 140 mM NaCl, 0.1% dissolved organic compound, 1% Triton X, 1 mM EDTA) with protease inhibitors and added to the antibody-coupled beads with rotation at 4 °C overnight. After washing, samples were treated with elution buffer (10 mM Tris-Cl, pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.5% SDS), RNase A, and Proteinase K, and cross-links were reversed overnight. ChIP DNA was purified using AMPure XP beads (Agencourt) and quantified using the Qubit 2000 (Invitrogen) and Bioanalyzer 1000 (Agilent). Libraries for Illumina sequencing were generated following the New England BioLabs (NEB) Next Ultra DNA Library Prep Kit protocol. A total of 10 cycles were used during PCR amplification for the generation of all ChIP-seq libraries. Amplified ChIP DNA was purified using double-sided AMPure XP to retain fragments (~200–500 bp) and quantified using the Qubit 2000 and Bioanalyzer 1000 before multiplexing.

ChIP-seq data processing. Raw fastq reads for all ChIP-seq experiments were processed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and quality reads were aligned to the hg19 reference genome using Bowtie version 1.1.2[51] with the following criteria: -n 1 -m 1–best–strata. Duplicate reads were marked using SAMBLASTER[52] before compression to BAM files. To directly compare Control and UBR7-shRNA ChIP-seq samples, uniquely mapped reads for each mark were normalized by total reads per condition, sorted, and indexed using samtools version 0.1.19[53].

Model-based analysis of ChIP-seq (MACS) (version 1.4.2; peak calling algorithm with a p value threshold of $1e-7$)[54] was used to identify H2BK120Ub enrichment over "input" background. Unique H2BK120Ub binding sites were identified using the concatenate, cluster, and subtract tools from the Galaxy/Cistrome web-based platform[55]. Briefly, a shared peak set was first generated by clustering intervals of H2BK120-Control peaks that directly overlapped H2BK120-UBR7-shRNA peaks by a minimum of 1 bp. Unique peaks were then identified by subtracting the total number of H2BK120 peaks in each condition by the shared peak set. Venn diagrams were generated using the Venn Diagram tool in Galaxy. To visualize ChIP-seq libraries on the IGV browser, we used deepTools version 2.4.060 to generate bigWig files by scaling the bam files to reads per kilobase per million (RPKM) using the following criteria: bamCoverage –b–normalizeUsing RPKM–smoothLength 300–binSize 30–extendReads 200 -o.

A list of known genes was obtained from the UCSC Genome browser (http://genome.ucsc.edu/). Proximal promoters were defined as ±5 kb from the transcription start site (TSS) and the genebody was defined as all genic regions outside of the +5 kb promoter region. Intergenic regions were defined as all regions outside both the proximal promoter and genebody. H2BK120 peaks were assigned to genes if they overlapped the promoter or genebody by a minimum of 1 bp. These H2BK120Ub "enriched" regions were further used for the generation of read density plots for all ChIP-seq data. All read density plots were generated using thengs.plotpackage in R[56].

ChIP-seq analysis for UBR7, RNF20, and RNF40 comparison. To directly compare Control SCR, UBR7, RNF20, and RNF40 shRNA ChIP-seq samples, uniquely mapped reads for H2BK120Ub in all conditions were normalized to ~10 million reads. For generation of Supplementary Fig. 3b–d, Control replicate1 was used for the UBR7-shRNA-1 comparison and Control replicate2 was used for the RNF20/RNF40 shRNA comparisons. Normalization of Control replicate1 and UBR7-shRNA-1 samples to ~10 million reads displayed little effect on the average density profile of H2BK120Ub (Supplementary Fig. 3b). For further UBR7, RNF20, and RNF40 comparisons, read counts for Control (SCR) H2BK120ub and Input replicates were merged together, normalized to ~10 million reads, and peaks were called using MACS (p value $1e-7$). To identify unique H2BK120 binding sites that were lost upon knockdown of either UBR7, RNF20, or RNF40, a shared peak set was first generated by clustering intervals of Control (SCR) H2BK120Ub sites that directly overlapped either UBR7, RNF20, or RNF40 shRNA H2BK120Ub peaks by a minimum of 1 bp. Unique peaks were then identified by subtracting the total number of H2BK120Ub peaks in each condition by their associated shared peak set. A final Control (SCR) shared peak set was further generated using the Control (SCR) H2BK120Ub binding sites from Supplementary Fig. 3e–g, which were lost from depletion of each factor (UBR7, RNF20, or RNF40). A final Control (SCR) unique peak set was identified by subtracting the total number of H2BK120Ub peaks in each condition by the shared peak set. H2BK120Ub peaks were assigned to genes if they overlapped the promoter (±5kbTSS) or genebody by a minimum of 1 bp and these H2BK120Ub "enriched" regions were used for the generation of average density profiles (Supplementary Fig. 3b–d, 3h) and pathway analysis (Supplementary Fig. 3i).

Chromatin state calls. ChromHMM[57] was used to identify combinatorial chromatin state patterns based on the histone modifications studied. Normalized bam files were converted into binarized data at a 1000 bp resolution using the

BinarizeBam command with a p value cut-off of $1e-5$. We specified that ChromHMM should learn a model based on 10 chromatin states. As we considered models between 8 and 20 chromatin states, we chose a 10-state model because it is large enough to identify important functional elements while still being small enough to interpret easily. Overlap enrichment was used to compute differential enrichment in each of the 10 chromatin states between Control and UBR7-shRNA samples. The ChromHMM segment files from the 10-state model contain the genomic locations of each chromatin state called in both the Control and UBR7-shRNA samples. To determine which chromatin states were enriched between conditions, we further compared the genomic locations by using the Control segments file as input for the segment directory, and by further separating the UBR7-shRNA segments file into 10 individual states and using it as input for the external coordinate directory. The UBR7-shRNA segment file was separated into individual chromatin states for the external coordinate directory with the following command:

awk -F/t '{print ≫ $4;close($4)}' ~/path_to/UBR7-sh1_segments.bed

Overlap enrichment was ran using the following command:

java -mx4000M -jar ChromHMM.jar OverlapEnrichment ~/path_to/Control_segments.bed ~/path_to/UBR7-shRNA_segments_separated OverlapEnrichment_Control_vs_UBR7

RNA-sequencing. RNA was isolated using RNeasy kit and libraries prepared using Illumina mRNA-Seq library kit. Raw FASTQ reads for all RNA-seq experiments were processed using FastQC and quality reads were aligned with the hg19 reference genome using TopHat(version 2.0.14)[58] with a Bowtie2 (version 2.2.3)[59] index based on UCSC annotations using the following criteria: -G -g 1 -r 150–mate-std-dev 50–library-type fr-unstranded. These criteria preserved only the best reads that uniquely mapped to the genome with one or fewer mismatches. To visualize RNA-seq libraries on the IGV browser, we used deepTools version 2.4.060 to generate bigWig files by scaling the bam files to RPKM using the following criteria: bamCoverage –b --normalizeUsing RPKM --smoothLength 300 --binSize 30 -o.

For identification of differentially expressed genes and gene set enrichment analysis (GSEA), raw counts were obtained by assigning reads at the gene level across the UCSC hg19 reference genome using featureCount[60] in the Rsubread package. DESeq2[61] was employed for normalization and identification of differentially expressed genes in UBR7-shRNA and Control samples. All plots were generated using the ggplot and ggrepel packages in R. GSEA[62] was run with normalized counts from all identified differentially expressed genes using the hallmark, curated, and gene ontology gene sets with default settings.

Statistical analysis. TCGA data analysis: Gene expression data for UBR7 from TCGA data were analyzed using the UCSC Xena functional genomics browser. Cancer subtypes for gene expression data were extracted from PAM50 version.

ChIP-qPCR and qRT-PCR: All qRT-PCR, ChIP, and other quantification data were collected in experiments performed in technical triplicate. Each experiment was repeated at least three times, and statistically significant results were obtained. An unpaired two-tailed Student's t test was performed using the Prism software program (GraphPad Software) to assign the significant differences between groups. Significant differences were considered when $P < 0.05$, $*P \le 0.05$, $**P \le 0.001$, and $***P \le 0.0001$. Error bars indicate the standard deviation of the mean for the technical replicates, as indicated in the legend.

TMAs: Two different TMAs were used: one for analyses of primary versus metastatic invasive ductal carcinoma tumors and the other for the association of UBR7 staining with ER, PR, and HER2 status. The intensity and percentage were measured three times for each tumor, and the means of these measures were used for data analyses (missing values were excluded). All statistical analyses were performed using R (version 3.3.1).

Data availability
ChIP-Seq and RNA-Seq data can be accessed at GEO using the accession number: GSE93759. All relevant data are available from the authors upon request.

References

1. Polyak, K. Heterogeneity in breast cancer. *J. Clin. Invest.* **121**, 3786–3788 (2011).
2. Palchik, J. D. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
3. Viale, G. The current state of breast cancer classification. *Ann. Oncol.* **23**, x207-10 (2012).

4. Stephens, P. J. et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).
5. Diepenbruck, M. & Christofori, G. Epithelial–mesenchymal transition (EMT) and metastasis: yes, no, maybe? *Curr. Opin. Cell Biol.* **43**, 7–13 (2016).
6. Felipe Lima, J., Nofech-Mozes, S., Bayani, J. & Bartlett, J. M. EMT in breast carcinoma—a review. *J. Clin. Med.* **5**, 65 (2016).
7. Taube, J. H. et al. The H3K27me3-demethylase KDM6A is suppressed in breast cancer stem-like cells, and enables the resolution of bivalency during the mesenchymal–epithelial transition. *Oncotarget* **8**, 65548–65565 (2017).
8. Malouf, G. G. et al. Architecture of epigenetic reprogramming following Twist1-mediated epithelial–mesenchymal transition. *Genome Biol.* **14**, R144 (2013).
9. Chervona, Y. & Costa, M. Histone modifications and cancer: biomarkers of prognosis? *Am. J. Cancer Res.* **2**, 589–597 (2012).
10. Elsheikh, S. E. et al. Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res.* **69**, 3802–3809 (2009).
11. Bhatnagar, S. et al. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. *Nature* **516**, 116–120 (2014).
12. Bochar, D. A. et al. BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**, 257–265 (2000).
13. Rai, K. et al. Dual roles of RNF2 in melanoma progression. *Cancer Discov.* **5**, 1314–1327 (2015).
14. Shema, E. et al. The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev.* **22**, 2664–2676 (2008).
15. Espinosa, J. M. Histone H2B ubiquitination: the cancer connection. *Genes Dev.* **22**, 2743–2749 (2008).
16. Cao, J. & Yan, Q. Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Front. Oncol.* **2**, 26 (2012).
17. Fuchs, G. et al. RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. *Mol. Cell* **46**, 662–673 (2012).
18. Kato, A. & Komatsu, K. RNF20-SNF2H pathway of chromatin relaxation in DNA double-strand break repair. *Genes (Basel)* **6**, 592–606 (2015).
19. Kwon, Y. T. et al. The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl Acad. Sci. USA* **95**, 7898–7903 (1998).
20. Tasaki, T. et al. A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol. Cell. Biol.* **25**, 7120–7136 (2005).
21. Tasaki, T. et al. The substrate recognition domains of the N-end rule pathway. *J. Biol. Chem.* **284**, 1884–1895 (2009).
22. Lee, M. J. et al. Synthetic heterovalent inhibitors targeting recognition E3 components of the N-end rule pathway. *Proc. Natl Acad. Sci. USA* **105**, 100–105 (2008).
23. Sanchez, R. & Zhou, M. M. The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **36**, 364–372 (2011).
24. Kim, J., Hake, S. B. & Roeder, R. G. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol. Cell* **20**, 759–770 (2005).
25. Mallery, D. L., Vandenberg, C. J. & Hiom, K. Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *EMBO J.* **21**, 6755–6762 (2002).
26. Xia, Y., Pao, G. M., Chen, H. W., Verma, I. M. & Hunter, T. Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. *J. Biol. Chem.* **278**, 5255–5263 (2003).
27. Maunakea, A. K., Chepelev, I. & Zhao, K. Epigenome mapping in normal and disease States. *Circ. Res.* **107**, 327–339 (2010).
28. Zhu, J. et al. The UCSC Cancer Genomics Browser. *Nat. Methods* **6**, 239–240 (2009).
29. Jurikova, M., Danihel, L., Polak, S. & Varga, I. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. *Acta Histochem.* **118**, 544–552 (2016).
30. Liu, F., Gu, L. N., Shan, B. E., Geng, C. Z. & Sang, M. X. Biomarkers for EMT and MET in breast cancer: an update. *Oncol. Lett.* **12**, 4869–4876 (2016).
31. Agiostratidou, G. et al. Loss of retinal cadherin facilitates mammary tumor progression and metastasis. *Cancer Res.* **69**, 5030–5038 (2009).
32. Pohl, S. G. et al. Wnt signaling in triple-negative breast cancer. *Oncogenesis* **6**, e310 (2017).
33. Cowin, P., Rowlands, T. M. & Hatsell, S. J. Cadherins and catenins in breast cancer. *Curr. Opin. Cell Biol.* **17**, 499–508 (2005).
34. Li, J. & Zhou, B. P. Activation of beta-catenin and Akt pathways by Twist are critical for the maintenance of EMT associated cancer stem cell-like characters. *BMC Cancer* **11**, 49 (2011).
35. Aasland, R., Gibson, T. J. & Stewart, A. F. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* **20**, 56–59 (1995).
36. Kwan, A. H. et al. Engineering a protein scaffold from a PHD finger. *Structure* **11**, 803–813 (2003).
37. Borden, K. L. & Freemont, P. S. The RING finger domain: a recent example of a sequence-structure family. *Curr. Opin. Struct. Biol.* **6**, 395–401 (1996).
38. Kim, J. et al. RAD6-mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**, 459–471 (2009).
39. Werner, M. & Ruthenburg, A. J. The United States of histone ubiquitylation and methylation. *Mol. Cell* **43**, 5–7 (2011).
40. Chatterjee, C., McGinty, R. K., Fierz, B. & Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **6**, 267–269 (2010).
41. Norton, L. & Massague, J. Is cancer a disease of self-seeding? *Nat. Med.* **12**, 875–878 (2006).
42. Halbleib, J. M. & Nelson, W. J. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Dev.* **20**, 3199–3214 (2006).
43. Glukhova, M. A. & Streuli, C. H. How integrins control breast biology. *Curr. Opin. Cell Biol.* **25**, 633–641 (2013).
44. Berx, G. & Van Roy, F. The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. *Breast Cancer Res.* **3**, 289–293 (2001).
45. Adhikary, S. et al. Selective recognition of H3.1K36 dimethylation/H4K16 acetylation facilitates the regulation of all-*trans*-retinoic acid (ATRA)-responsive genes by putative chromatin reader ZMYND8. *J. Biol. Chem.* **291**, 2664–2681 (2016).
46. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).
47. Su, X. et al. TAp63 suppresses metastasis through coordinate regulation of Dicer and miRNAs. *Nature* **467**, 986–990 (2010).
48. Liang, C. C., Park, A. Y. & Guan, J. L. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.* **2**, 329–333 (2007).
49. Venkatanarayan, A. et al. IAPP-driven metabolic reprogramming induces regression of p53-deficient tumours in vivo. *Nature* **517**, 626–630 (2015).
50. Blecher-Gonen, R. et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).
51. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
52. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
55. Liu, T. et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83 (2011).
56. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genom.* **15**, 284 (2014).
57. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
58. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
60. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
62. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Author contributions

S.R., K.R., and C.D. conceived the study, designed experiments, and analyzed the data. S.A., D.C., C.T., I.S., and M.M. designed, performed experiments, and analyzed data. E.T., F.Y., and C.T. stained and scored T.M.A.s. A.A.S. provided reagents. J.M. and R.B. performed statistical analyses related to the TMAs. A.D., D.K.S., C.T., and A.T.R. performed the bioinformatic analysis. C.D., K.R., S.R., S.A., D.C., and C.T. wrote the paper.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08986-5.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.